

Analysis of oral microbiome in glaucoma patients using machine learning prediction models

Byung Woo Yoon^a, Su-Ho Lim^b, Jong Hoon Shin^c, Ji-Woong Lee^d, Young Lee^e and Je Hyun Seo^e

^aDivision of Oncology, Department of Internal Medicine, Seoul Paik Hospital, Seoul, Korea; ^bDepartment of Ophthalmology, Daegu Veterans Health Service Medical Center, Daegu, Republic of Korea; ^cDepartment of Ophthalmology, Pusan National University Yangsan Hospital, Yangsan, Republic of Korea; ^dDepartment of Ophthalmology, Pusan National University Hospital, Busan, Republic of Korea; ^eVeterans Medical Research Institute, Veterans Health Service Medical Center, Seoul, Republic of Korea

ABSTRACT

Purpose: The microbiome is considered an environmental factor that contributes to the progression of several neurodegenerative diseases. However, the association between microbiome and glaucoma remains unclear. This study investigated the features of the oral microbiome in patients with glaucoma and analyzed the microbiome biomarker candidates using a machine learning approach to predict the severity of glaucoma.

Methods: The taxonomic composition of the oral microbiome was obtained using 16S rRNA gene sequencing, operational taxonomic unit analysis, and diversity analysis. The differentially expressed gene (DEG) analysis was performed to determine the taxonomic differences between the microbiomes of patients with glaucoma and the control participants. Multinomial logistic regression and association rule mining analysis using machine learning were performed to identify the microbiome biomarker related to glaucoma severity.

Results: DEG analysis of the oral microbiome of patients with glaucoma revealed significant depletion of *Lactococcus* ($P = 3.71e^{-31}$), whereas *Faecalibacterium* was enriched ($P = 9.19e^{-14}$). The candidate rules generated from the oral microbiome, including *Lactococcus*, showed 96% accuracy for association with glaucoma.

Conclusions: Our findings indicate microbiome biomarkers for glaucoma severity with high accuracy. The relatively low oral *Lactococcus* in the glaucoma population suggests that microbial dysbiosis could play an important role in the pathophysiology of glaucoma.

ARTICLE HISTORY

Received 3 June 2021

Revised 20 July 2021

Accepted 22 July 2021

KEYWORDS



Neurodegenerative; oral microbiome; glaucoma; biomarker; dysbiosis

Introduction


Glaucoma is a progressive optic neuropathy characterized by the degeneration of the retinal ganglion cells and their axons, with a corresponding visual field defect [1]. Intraocular pressure (IOP) is a major risk factor for glaucoma development. A high IOP exerts tremendous stress and strain on the lamina cribrosa, consequently blocking the axoplasmic flow [2]. However, glaucomatous axon loss also occurs in individuals with normal IOP, and the risk of glaucoma development might differ even in individuals with the same IOP; thus, glaucoma should be assessed as a complex neurodegenerative disease. Candidate factors associated with glaucoma pathogenesis include aging, inflammation [3], and immune responses, including heat shock proteins or local inflammatory responses [4–8]. As progressive retinal ganglion cell loss is a pathophysiological feature in patients, glaucoma shares signs and symptoms with other neurodegenerative diseases, including Alzheimer's disease [9]. A population-based cohort study reported a significant association of glaucoma with Alzheimer's disease,

suggesting a potential link between neurodegenerative brain diseases and retinal neurodegenerative diseases [10,11].

Technological advances in next-generation sequencing (NGS) and bioinformatics tools have identified an association of human microbiota with neurodegenerative diseases and autoimmunity. The altered human microbiome, or microbial dysbiosis, can contribute to the pathogenesis of several diseases [12]. Although the ocular microbiome is related to eye diseases, non-ocular microbiome such as an oral microbiome could influence the pathogenesis of eye disease [3,6,13,14]. The microbiota-gut-brain axis involves a cross-talk between the gut and microbiota; enteric bacteria enters the circulation and crosses the blood-brain barrier, or to penetrates the brain via the vagus nerve [15,16]. Since the oral microbiome is a part of the gut microbiome family and 5% of oral species are found in the intestine, the microbiome-gut-retinal axis may have clinical importance. A previous study that investigated the relationship between microbiota and glaucoma reported high oral bacterial loads,

CONTACT Je Hyun Seo  jazmin2@naver.com  Veterans Medical Research Institute, Veterans Health Service Medical Center, Jinhwangdo-ro 61-gil 53, Gangdong-gu, Seoul, 05368, Republic of Korea

Byung Woo Yoon and Su-Ho Lim contributed equally to this work.

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

such as *Streptococcus* species, in patients with glaucoma [7]. To date, only a few studies have investigated the influence of the gut microbiome on glaucoma. Therefore, this study aimed to investigate the differences in the oral microbiome in patients with glaucoma based on the microbiota-gut-retinal axis, and identify the candidate microbiome biomarkers related to glaucoma pathogenesis and severity using machine learning prediction models.

Material and methods

Participants

This multicentric study was conducted by researchers from Veterans Health Service (VHS) Medical Center, Daegu Veterans Hospital, Pusan National University Yangsan Hospital, and Pusan National University Hospital. The study protocol was approved by the Institutional Review Board of each participating center. The study was performed in accordance with the tenets of the Declaration of Helsinki and written informed consent was obtained from all participants before their enrollment.

During the study period, the research team enrolled patients with glaucoma who satisfied the inclusion criteria. Glaucoma was defined as having glaucomatous optic neuropathy, such as rim thinning, notching, retinal nerve fiber layer (RNFL) defect, or glaucomatous visual field defects. The control group was defined as participants with 1) a visual acuity of 20/40 or better; and/or those without 2) a glaucomatous optic nerve head or RNFL defect, visual field defect, or other retinal diseases; or 3) a history of IOP-lowering treatment. Participants with a history of ocular surgery (except uncomplicated cataract surgery), history of ocular trauma, or other diseases affecting the visual field (e.g. diabetic retinopathy) were excluded. Participants with a dental issue, those with probiotics use within the last 1 month, those with upper respiratory infection symptoms within the last two weeks, and those taking antibiotics or using topical antibiotic eye drops within the last one month were additionally excluded.

Sample collection and DNA extraction

The oral microbiome was collected by swabbing the right and left buccal surfaces and under the tongue four times using the Gene kit (Daeil pharm company, Korea) after washing the oral cavity with clean drinking water. Metagenomic DNA was individually extracted from swabs using the GeneAll ExgeneTM Blood/Clinic/Cell SV mini kit (GeneAll, Seoul, Korea) according to the manufacturer's instructions. The DNA was stored at -80°C following DNA extraction, and the quantity and

quality of the isolated DNA were measured using Trinean Dropsense 96 (Unchained Labs, Pleasanton, CA, USA) and PicoGreen (Thermo Fisher Scientific, Waltham, MA, USA).

Sequencing of 16S rRNA gene amplicon

A 16S rRNA sequencing library was generated according to the 16S metagenomic sequencing library preparation protocol (Illumina, San Diego, CA, USA), targeting the V3 and V4 hypervariable regions of the 16S rRNA gene, which were selected based on the results reported previously [17]. The KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) and Agencourt AMPure XP system (Beckman Coulter Genomics, Brea, CA, USA) were used for polymerase chain reaction (PCR) amplification and amplicon purification, respectively. The Illumina adapter overhang nucleotide sequences were added to the gene-specific sequences. The following full-length primer sequences were used to target these adapters:

16S Amplicon PCR Forward Primer:

5' TCGTCGGCAGCGTCAGATGTGTATAAGA
GACAGCCTACGGGNGGCWGCAG

16S 3'

Amplicon PCR Reverse Primer

5' GTCTCGTGGGCTCGGAGATGTGTATAAGA
GACAGGACTACHVGGGTATCTAATCC 3'

After magnetic bead-based purification of the PCR products, a second PCR with a limited cycle was performed using primers from the Nextera XT Index Kit (Illumina). Subsequently, the purified PCR products were visualized using gel electrophoresis and quantified with a Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) on a Qubit 3.0 fluorometer. The samples were pooled and assayed using Agilent 2100 bioanalyzer (Agilent) for pre-sequencing quality analysis. The previously generated sequencing libraries were quantified by qPCR using the CFX96 Real-Time System (Bio-Rad, Hercules, CA, USA). After normalization, the libraries were sequenced on the MiSeq system (Illumina) with 300 bp paired-end reads.

Pre-processing

The adaptor sequences were removed from the original paired-end reads using CutAdapt v1.11. Next, merged reads were generated using the adaptor-trimmed paired-end reads using FLASH v1.2.11. The low-quality merged reads were filtered according to the following criteria: reads that had ≥ 2 ambiguous nucleotides, reads with average quality score < 20 , and reads with length < 300 bp after trimming of low-

quality bases. Finally, potential chimeric reads were removed using UCHIME v4.2.40 [18].

Calculation of Operational Taxonomic Units (OTUs) and diversity analysis

The pre-processed reads from each sample were used to calculate the number of OTUs, which were determined by clustering the sequences from each sample with a 97% sequence identity cut-off using QIIME software (v.1.9.0, www.qiime.org) [19,20]. The taxonomic abundance was counted with Ribosomal Database Project (RDP) Classifier v1.1, with a confidence threshold of 0.8 derived from the pre-processed reads of each sample [21]. The microbial composition in the samples was normalized using the value calculated by a division between the taxonomic abundance count and the number of pre-processed reads for each sample. Consensus sequences were clustered using cd-hit v4.6 using the following parameters – identify, >99%; and coverage, >80% – and then, aligned with the reference sequences from the National Center for Biotechnology Information (NCBI) database using the Megablast algorithm. Finally, taxonomic profiling was performed for the assembled genome using NCBI taxonomy information. The alpha-diversity for observed OTUs and Shannon index within the samples were calculated. Principal component analysis (PCoA) was performed using the result from the beta-diversity for the glaucoma and control groups, as well as for subgroup analysis.

Machine learning-based identification of microbiome biomarker for glaucoma severity

Machine learning-based identification of glaucoma biomarkers consisted of three steps: significant taxon level identification by Random Forest (RF), identification of a significant taxon by Tag Count Comparison (TCC), and rule mining of the microbiome. RF model-based machine learning was performed with the R software (R Foundation for Statistical Computing, Vienna, Austria) using ‘ranger’ package [22]. Within each run, the number of trees (num.tree) was set to build the RF model [23–25]. The number of variables/taxa that could be selected in each splitting node (mtry) was set to a range of one-tenth of the maximum taxon number at each taxonomic level. The Gini index evaluated the importance of each taxon, and the prediction accuracy was determined by $\frac{TruePositive(TP)+TrueNegative(TN)}{TP+TN+FalsePositive+FalseNegative}$. Hundred runs were performed for each mtry parameter within each taxonomic hierarchy. We calculated the significant taxa using TCC (<http://bioconductor.org/>), which provides functions for differential expression analysis of genes (DEG) with normalization and

multi-group comparison [25]. With TCC analysis, significant oral microbiome differences were analyzed in the glaucoma group, compared to the control group, and in the primary open-angle glaucoma (POAG) subgroups, compared to the control group.

Before rule mining, a multinomial logistic regression (MLR) was performed to analyze the significance of the identified microbiome biomarker in patients with glaucoma based on the clinical factors, such as age, baseline IOP level, RNFL thickness (assessed using optical coherence tomography [OCT]), and mean deviation value in the Humphrey visual field. Patients aged <60 years comprised the younger age group, and those ≥60 years comprised the older age group. The baseline IOP cut-off was 20 mmHg; normal baseline IOP, <20 mmHg; and high baseline IOP, ≥20 mmHg. Although arbitrary, this value is acceptable for statistical processing since the IOP criterion for normal-tension glaucoma is based on a cut-off of 21 mmHg, and the mean IOP of Asians is 1–2 mmHg lower than the other ethnicities. For glaucoma severity, the glaucoma group was divided into ≥80 μm and <80 μm subgroups based on the RNFL thickness. For functional glaucoma severity, the glaucoma group was stratified according to Anderson’s glaucoma stage criteria: ≥−6.0 dB, −6.0 to −12.0 dB, and <−12.0 dB.

The classification based on the predictive association rules (CPAR) algorithm was used to generate a complete ruleset [24]. This algorithm, which uses information metrics to generate rules, is more suited for bioinformatics applications than conventional support-confidence based measures for market basket data analysis [26]. The Laplace accuracy evaluated the accuracy of the CPAR-derived rules. Using rule r , the Laplace accuracy is defined as below:

$$Laplace\ accuracy(r) = \frac{(N_g + 1)}{(N_{total} + m)}$$

where m , number of target groups; N_{total} total number of examples that satisfy the body of the rules; and N_g , number of examples that belong to the predicted target group. Only the POAG subgroup was used for subgroup analysis using MLR and CPAR because of insufficient sample size in secondary glaucoma (SG) group.

Data analyses

All the data analyses were performed using the R Statistical Package, Version 3.6.2 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was set at $P < 0.05$. For multiple comparison tests in DEG analysis, the false discovery rate was controlled using the Benjamini–Hochberg step-up procedure [27,28].

Results

Patient characteristics and metagenomic data

Ninety-six patients with glaucoma and 25 controls were recruited from March 2019 to December 2019. The glaucoma group had 62 POAG and 34 SG patients (SG cases were mostly uveitic glaucoma). There was no difference in age and sex between patients with glaucoma and the normal participants (Table 1). In the glaucoma groups, the baseline IOP was 18.49 mmHg, which was higher than that (13.76 mmHg) of the normal participants. RNFL thickness, a structural indicator of glaucoma, was low in glaucoma patients, and the visual field index, a functional indicator, was moderately advanced glaucoma. The number of the average lead counts/sample of the control group was higher than that of the glaucoma group ($P = 0.038$). Results of machine learning using the RF method showed that genus level identification was significant in 193 mtry (accuracy, 0.5714–0.7429) than those identified based on family, order, and phylum, and *Lactococcus* had the highest Gini indices (Figure 1, Table 2).

Diversity analysis in glaucoma patients and control subjects

On comparisons of the alpha-diversity of oral microbiome between glaucomatous and control subjects, the observed OTUs ($P = 0.003$ and $P = 0.011$, respectively) and the Shannon index ($P = 6.1e^{-6}$ and $P = 3.6e^{-5}$, respectively; Figure 2) were significantly higher in the control participants than that in the glaucoma group and glaucoma subgroups. PCoA of beta-diversity for the oral microbiome showed that the microbial clusters in the glaucoma and control groups differed in their taxonomy (Figure 3). On subgroup analysis, the microbial clusters in POAG and SG subgroups were similar, but that in the control group was different. The beta-distance calculated from the RDP classifier showed significant differences, but the beta-distance from the NCBI method

was not different between the glaucoma and control groups ($P < 2.2e^{-16}$, $P = 0.52$, respectively; Figure 4), whereas the beta-distance of subgroup analysis showed a significant difference in the oral microbiome ($P < 2.2e^{-16}$).

DEG analysis in glaucoma group and subgroup

For TCC-based taxon significance analysis, the MA plot revealed the different quantities of some species and the DEG analysis showed various taxa such as *Lactococcus* and *Atopobium* were significantly depleted ($P = 3.71e^{-31}$, $P = 6.69e^{-21}$, respectively), whereas *Faecalibacterium* was enriched in the glaucoma group ($P = 9.19e^{-14}$) compared with control participants (Figure 5, Supplementary Table S1). On subgroup analysis, the volcano plot of the oral microbiome showed that *Lactococcus*, *Candidatus Pelagibacter*, and *Isobaculum* were significantly depleted ($P = 1.67e^{-20}$, $P = 9.37e^{-20}$, and $P = 1.43e^{-18}$, respectively), whereas *Faecalibacterium* and *Lachnospiraceae incertae sedis* were enriched ($P = 7.21e^{-16}$ and $P = 6.02e^{-12}$, respectively) in the POAG subgroup compared with the control group (Figure 6 and Supplementary Table S2).

MLR and association rules mining analysis for biomarkers of glaucoma

The *Lactococcus*, *Candidatus Pelagibacter*, and *Atopobium* were the significant taxa for factors age and baseline IOP. In addition, they were significantly associated with glaucoma severity based on RNFL thickness and visual field index (Table 3). Association rule generated by CPAR showed two 'glaucoma' rules and four 'normal' rules generated with a high accuracy in the 0.9–0.96 range (Table 4). 'Glaucoma' rule 1 is the most accurate, which states that if the microbial composition is *Atopobium* ≤ 0.347 , *Candidatus Pelagibacter* ≤ 0.001 , *Endobacter* ≤ 0.015 , and *Lactococcus* ≤ 0.013 , the sample patients is glaucomatous (accuracy, 96.0%).

Table 1. Demographic and baseline characteristics of glaucoma subjects.

Variables	Glaucoma n = 96	Control subject n = 25	P-value*
Age, years	61.48 ± 13.08	62.68 ± 1.73	0.385
Male/Female	66/30	18/7	0.753
BCVA, logMAR	0.21 ± 0.32	0.00 ± 0.20	<0.0001
Diagnosis, n (%)			
POAG/Secondary glaucoma	62 (64.58)/34 (35.42)	-	
Glaucoma medication period (yr)	2.74 ± 3.09	-	
Baseline IOP, mmHg	18.49 ± 6.60	13.76 ± 1.01	<0.0001
RNFL thickness, μ m	73.44 ± 14.52	105.36 ± 5.62	<0.0001
Visual field (MD), dB	-8.93 ± 8.44	0.03 ± 0.17	<0.0001
Visual field (VFI), %	76.97 ± 26.52	99.84 ± 0.47	<0.0001
Number of lead counts/sample	77,140.05 ± 23,087.21	83,310.64 ± 8,812.47	0.038

P-value* was calculated by independent T-test for continuous variables, Chi-square test was performed in case of numeric variable
IOP: intraocular pressure, BCVA: best-corrected visual acuity, logMAR: logarithm of the minimum angle of resolution, MD: mean deviation, VFI: visual field index, RNFL: retinal nerve fiber layer, POAG: primary open angle glaucoma

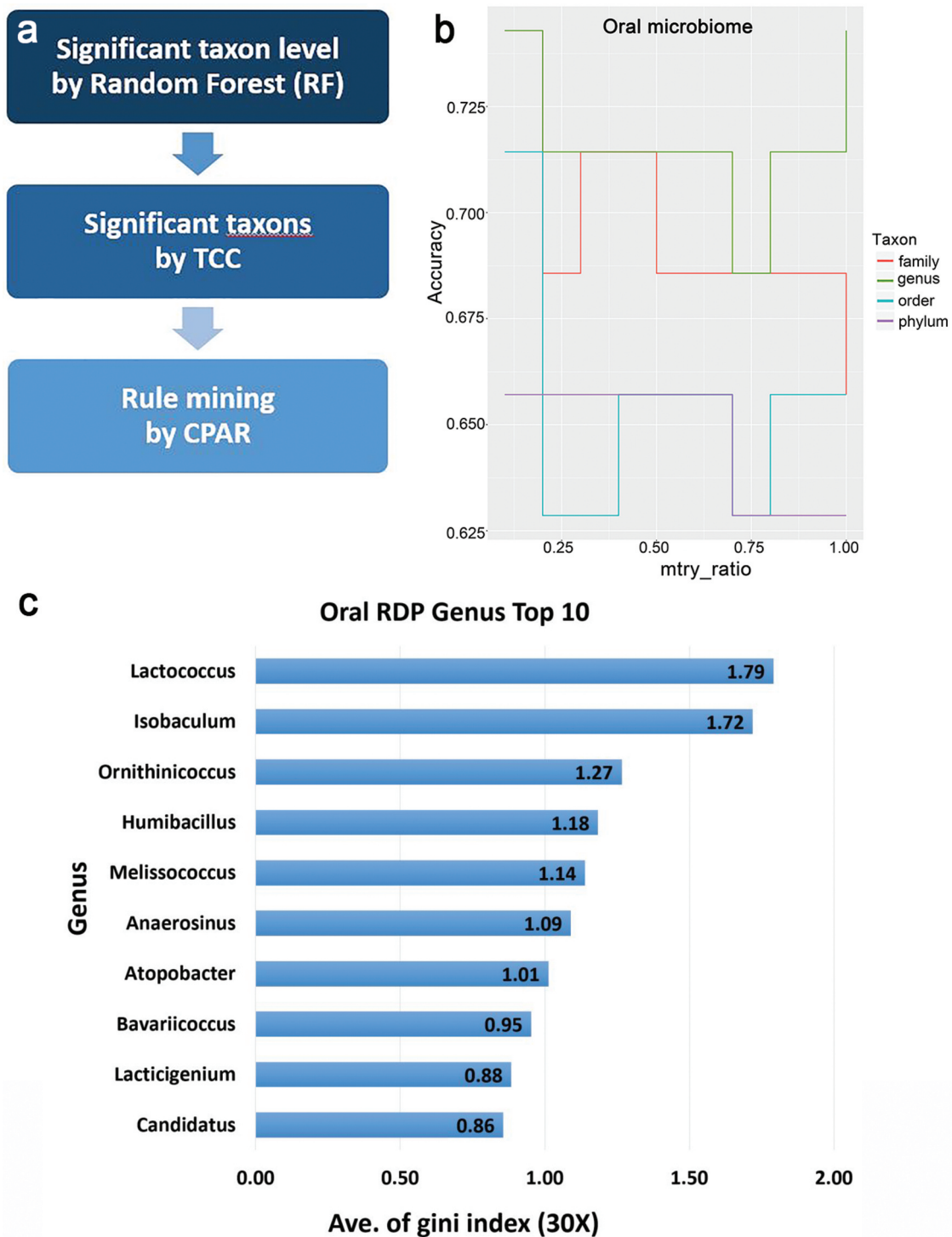


Figure 1. Metagenome analysis using machine learning and statistical models.

a: Schematic of machine learning for the study of algorithms and statistical models [23–25] TCC, tag count comparison; CPAR, Classification based on predictive association rules

b: Using Random Forest method, the significant taxa and genus levels in the oral microbiome were identified.

c: The top 10 species of the oral microbiome as identified from the Ribosomal Database Project (RDP) Data Base; *Lactococcus* was the most significant species in glaucoma patients and control subjects.

Similarly, ‘glaucoma’ rule 2 states that if the composition is *Lactococcus*, ≤ 0.013 , the patient is ‘glaucoma’ (accuracy, 90.0%). Contrarily, ‘normal’ rule 1 states that if the composition is *Candidatus Pelagibacter*, 0.001–0.014 and *Lactococcus*, 0.013–0.889, the patient is ‘normal’ (accuracy, 92.0%).

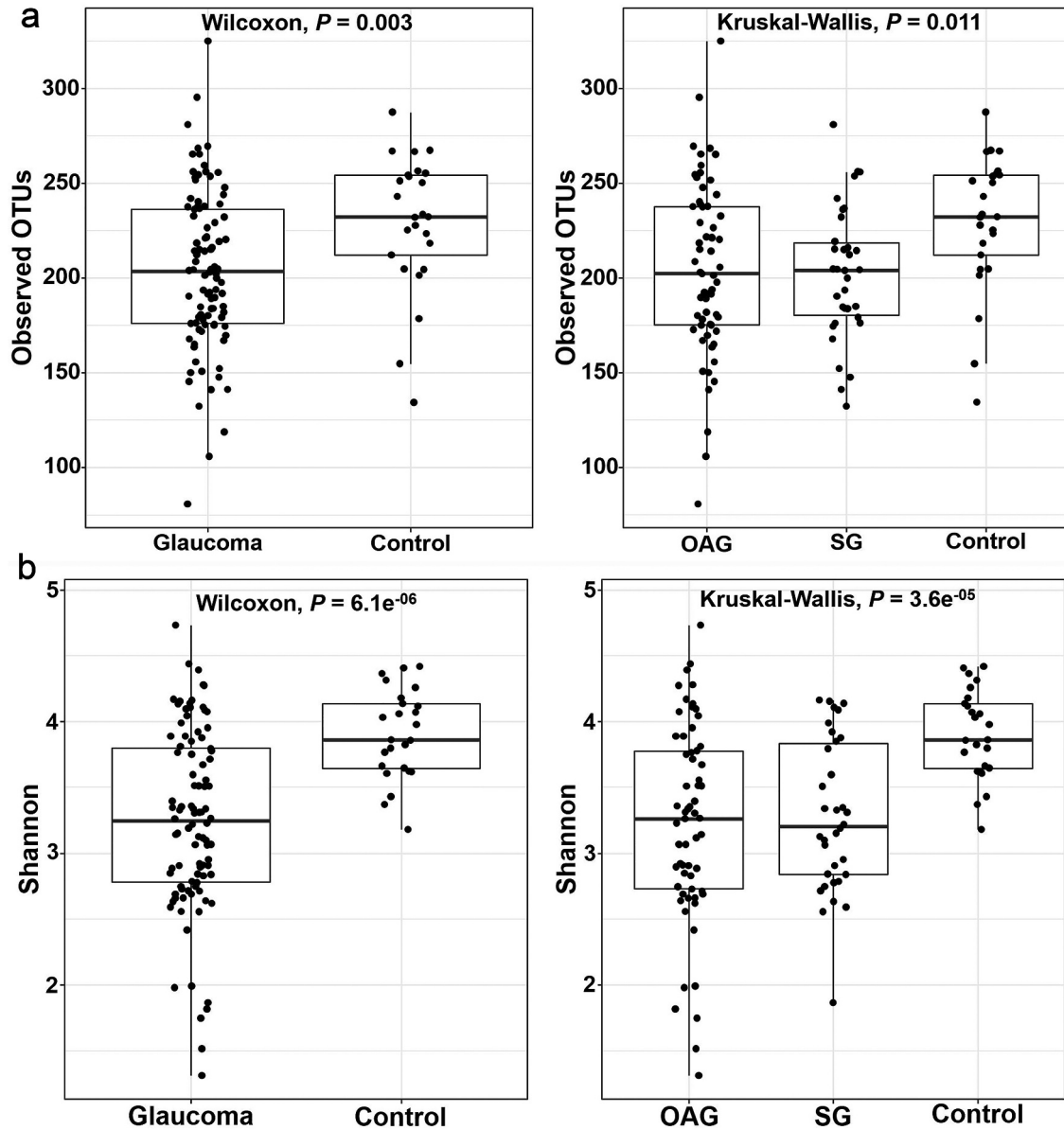
MLR and association rules mining analysis for biomarkers of POAG

In patients with POAG aged <60 years, *Lactococcus* was the significant taxon, whereas *Lactococcus*, *Candidatus Pelagibacter*, and *Atopobium* were the

Table 2. Gini index for accuracy condition in genus levels using random forest machine learning.

Taxon Level	Ntree	Ratio	mtry	Times	Accuracy	95% CI	P-Value [†]
*genus	200	0.1	193	100	0.5714 ~ 0.7429	0.3935 ~ 0.8751	$2.81e^{-08} \sim 0.0003783$
genus	200	0.2	385	100	0.5714 ~ 0.7143	0.3935 ~ 0.8536	$1.869e^{-07} \sim 0.0003783$
genus	200	0.3	578	100	0.6 ~ 0.7143	0.4211 ~ 0.8536	$1.869e^{-07} \sim 0.0001039$
genus	200	0.4	771	100	0.6 ~ 0.7143	0.4211 ~ 0.8536	$1.869e^{-07} \sim 0.0001039$
genus	200	0.5	964	100	0.5714 ~ 0.7143	0.3935 ~ 0.8536	$1.869e^{-07} \sim 0.0003783$
genus	200	0.6	1156	100	0.6 ~ 0.7143	0.4211 ~ 0.8536	$1.869e^{-07} \sim 0.0001039$
genus	200	0.7	1349	100	0.6 ~ 0.6857	0.4211 ~ 0.8315	$1.089e^{-06} \sim 0.0001039$
genus	200	0.8	1542	100	0.5714 ~ 0.7143	0.3935 ~ 0.8536	$1.869e^{-07} \sim 0.0003783$
genus	200	0.9	1734	100	0.6 ~ 0.7143	0.4211 ~ 0.8536	$1.869e^{-07} \sim 0.0001039$
genus	200	1	1927	100	0.5714 ~ 0.7429	0.3935 ~ 0.8751	$2.81e^{-08} \sim 0.0003783$

*Best accuracy condition in genus level, P-Value[†] was calculated by machine learning using random forest model
95% CI: confidential Interval

**Figure 2.** Comparison of alpha-diversity in glaucoma patients as compared with the control subjects.

(a) Observed OTUs was higher in the control group than in the glaucoma ($P = 0.003$). It was also higher in the control group than in the two subgroups ($P = 0.011$)

(b) Shannon index was higher in the control group than in the glaucoma group ($P = 6.1e^{-06}$). It was also higher in the control group than in the two subgroups ($P = 3.6e^{-05}$). POAG, primary open-angle glaucoma; SG, secondary glaucoma

significant taxa in patients with POAG aged ≥ 60 years (Table 5). In the POAG group with baseline IOP < 20 mmHg, *Lactococcus*, *Candidatus Pelagibacter*, and *Atopobium* were the significant

taxa, whereas *Lactococcus* and *Atopobium* were the significant taxa in the POAG group with baseline IOP ≥ 20 mmHg (Table 5). According to glaucoma severity based on RNFL thickness and visual field

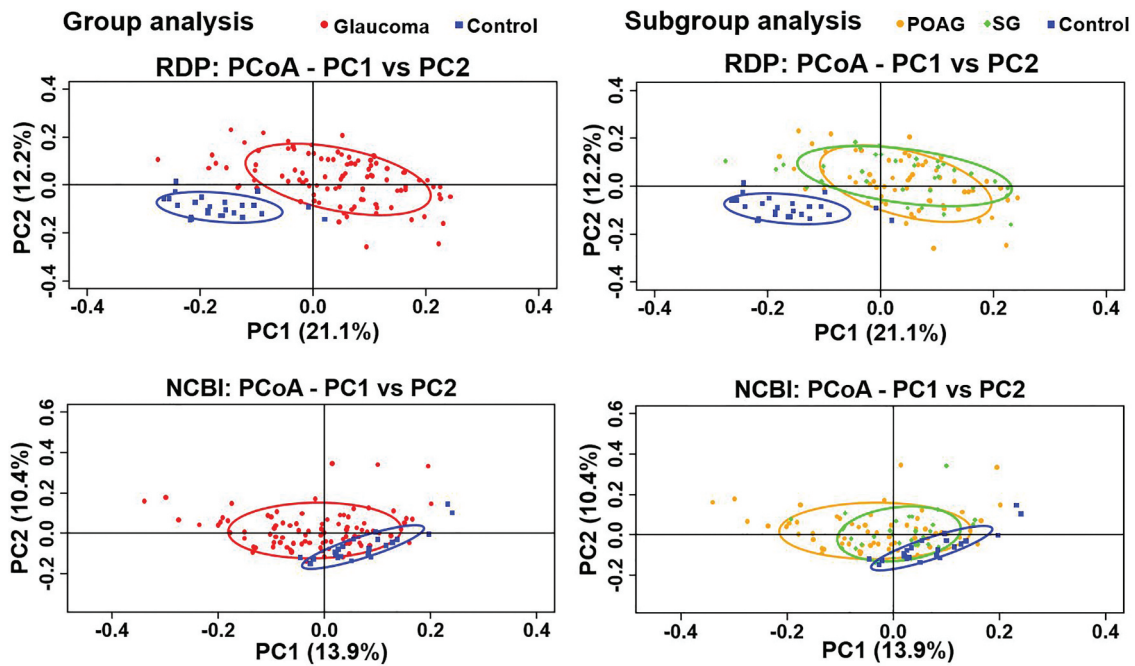


Figure 3. Differences in the oral microbiome of glaucoma patients compared with control subjects using principal component analysis for beta-diversity.

Beta-diversity was calculated in two ways – RDP and NCBI library. PCoA shows the different clusters in glaucoma vs. control groups, and POAG and SG subgroups vs. control group.

RDP, Ribosomal Database Project Classifier; NCBI, National Center for Biotechnology Information; POAG, Primary open-angle glaucoma; SG, secondary glaucoma

index in the POAG group, *Lactococcus*, *Candidatus Pelagibacter*, and *Atopobium* were the significant taxa. CPAR-derived association rule shows one ‘POAG’ rule and four ‘normal’ rules derived with high accuracy (Table 6). The ‘POAG’ rule 1 states that if the composition is *Atopobium*, ≤ 0.433 ; *Candidatus Pelagibacter*, ≤ 0.001 ; *Endobacter*, ≤ 0.015 ; and *Lactococcus*, ≤ 0.015 , the patient has POAG (accuracy, 93.0%). The ‘normal’ rule 1 states that if the composition is *Candidatus Pelagibacter*, 0.001–0.014, the patients is ‘normal’ (accuracy, 93.0%).

Discussion

In this study, we demonstrated that the oral microbiome in patients with glaucoma was different from that of the elderly without glaucoma, and identified candidate microbiome biomarkers for glaucoma severity with high accuracy using machine learning. Our study findings are consistent with those of previous studies in that we suggested a possible relationship between microbiome and glaucoma [3–5,7,29]. This relationship can be interpreted in several ways. As an immunity problem, the microbiome acts as a possible antigen, which may activate T-helper cell development and increased the production of antibodies. The trabecular meshwork cells can recognize the bacterial antigen, a pathogen-associated molecular pattern, such as the toll-like receptor, to generate

imbalanced signaling of inflammation [30]. Although this mechanism is not fully understood, it is considered that this cascade results in the anterior chamber-associated immune deviation, which induces immune cell activation as well as homing to the ocular region [29].

A previous study on the association of glaucoma and the oral microbiome has shown that altered commensal microbiome induces changes in cytokine signaling and complement activation [7]. In addition, one possible mechanism by which the microbiome could aid the management of glaucoma is by modulating brain-derived neurotrophic factor levels, as shown in an animal model [31]. From the perspective of neuroprotection, microbiome research suggests that the pathogenesis of glaucoma may be identified. According to the study by Astafurov *et al.*, oral bacterial load, such as of *Streptococcus* spp., in glaucoma cases was increased compared with the control [7]. However, in our study, the bacterial diversity was not significantly high in glaucoma patients, but there was a significant difference in the composition of the bacteria, such as *Faecalibacterium*, which was over-expressed in patients with glaucoma. *Faecalibacterium* modulated the butyrate metabolite in the gut, which was known to be related to anti-inflammatory effect [32]. Mental health disease such as major depression negatively correlated with *Faecalibacterium*, which was different in Alzheimer's disease and Parkinson's disease [32]. However, it

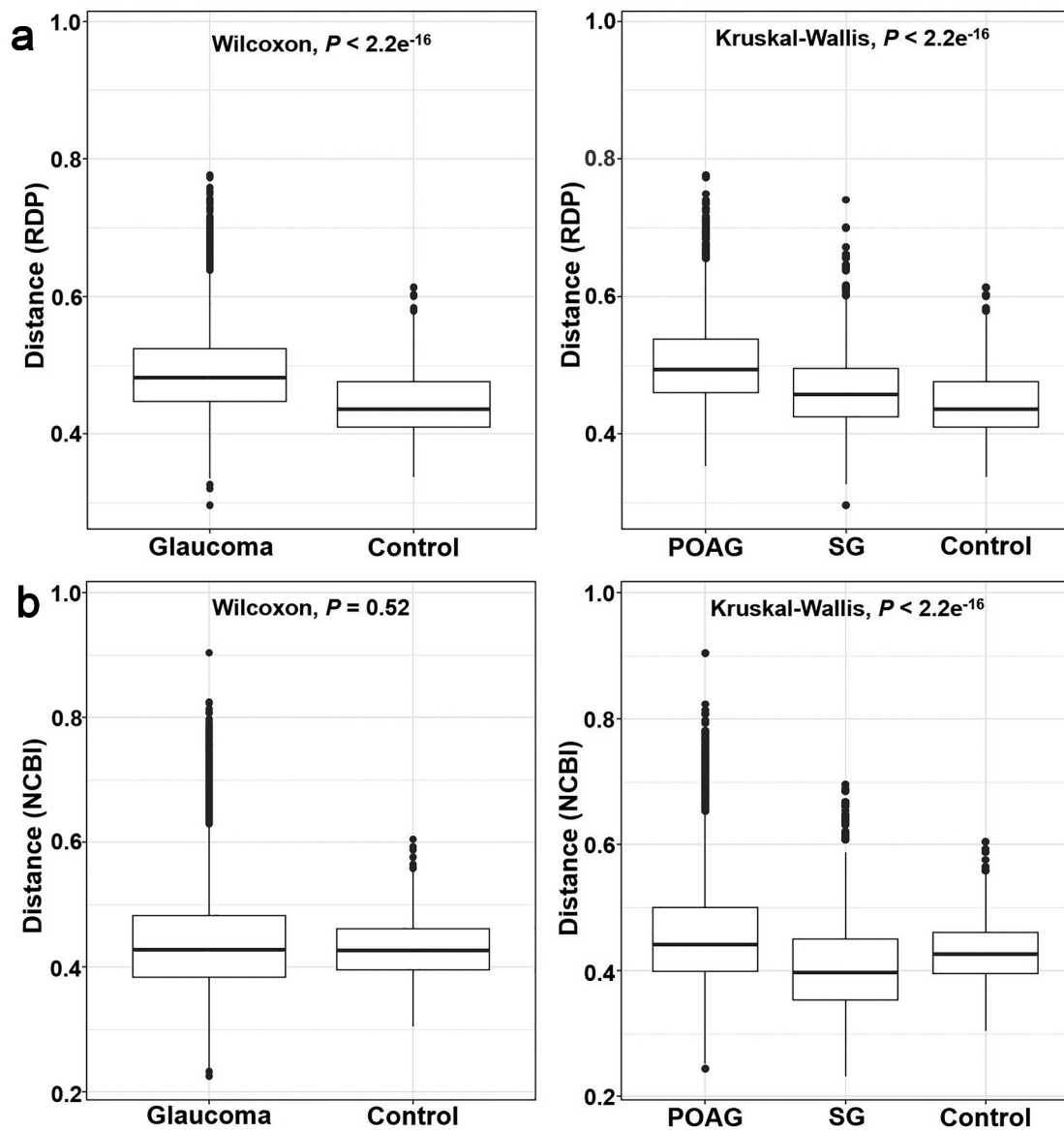


Figure 4. Beta-Diversity differences in the microbiome of glaucoma patients and control subjects.

Beta-diversity (estimated by RDP method) was significantly higher in glaucoma group than in control groups ($P < 2.2e^{-16}$); however, the beta-diversity using NCBI was not different between control and glaucoma groups ($P = 0.52$).

RDP, Ribosomal Database Project Classifier; NCBI, National Center for Biotechnology Information; POAG, Primary open-angle glaucoma; SG, secondary glaucoma

should be judged in consideration of the overall microbiome composition. Additionally, another study on the intestinal microbiota profile and POAG showed that *Megamonas* and *Bacteroides* load was decreased in POAG [33]. The study found negative correlations between visual acuity and *Blautia* and between the mean deviation of visual field and *Faecalibacterium*, and a positive correlation between RNFL thickness and *Streptococcus*. This study also showed that the gut microbiome might differ between glaucoma and control subjects, which is consistent with the results of our study, although taxa expression level was different due to differences in the targeted gut microbiome.

The reduced *Lactococcus* load in the oral microbiome suggests that dysbiosis in the microbiota

could play a role in the pathophysiology of glaucoma and can be used as a surrogate marker for glaucoma. It may be due to competition with other taxa in the oral microbiome or decreased antimicrobial activity of H_2O_2 produced by downregulation of the lactic acid bacteria. Gastrointestinal permeability is altered depending on diet and microbiota. In this regard, diet enriched with a high sugar content may cause a decrease in the number of beneficial microbes [34], which may directly or indirectly affect the development of glaucoma. The carbohydrate ingestion in POAG patients showed systemic autonomic dysregulation [35], whereas ketogenic diets and modified gut microbiome and permeability may play a role in neuroprotection in glaucoma patients [36,37].

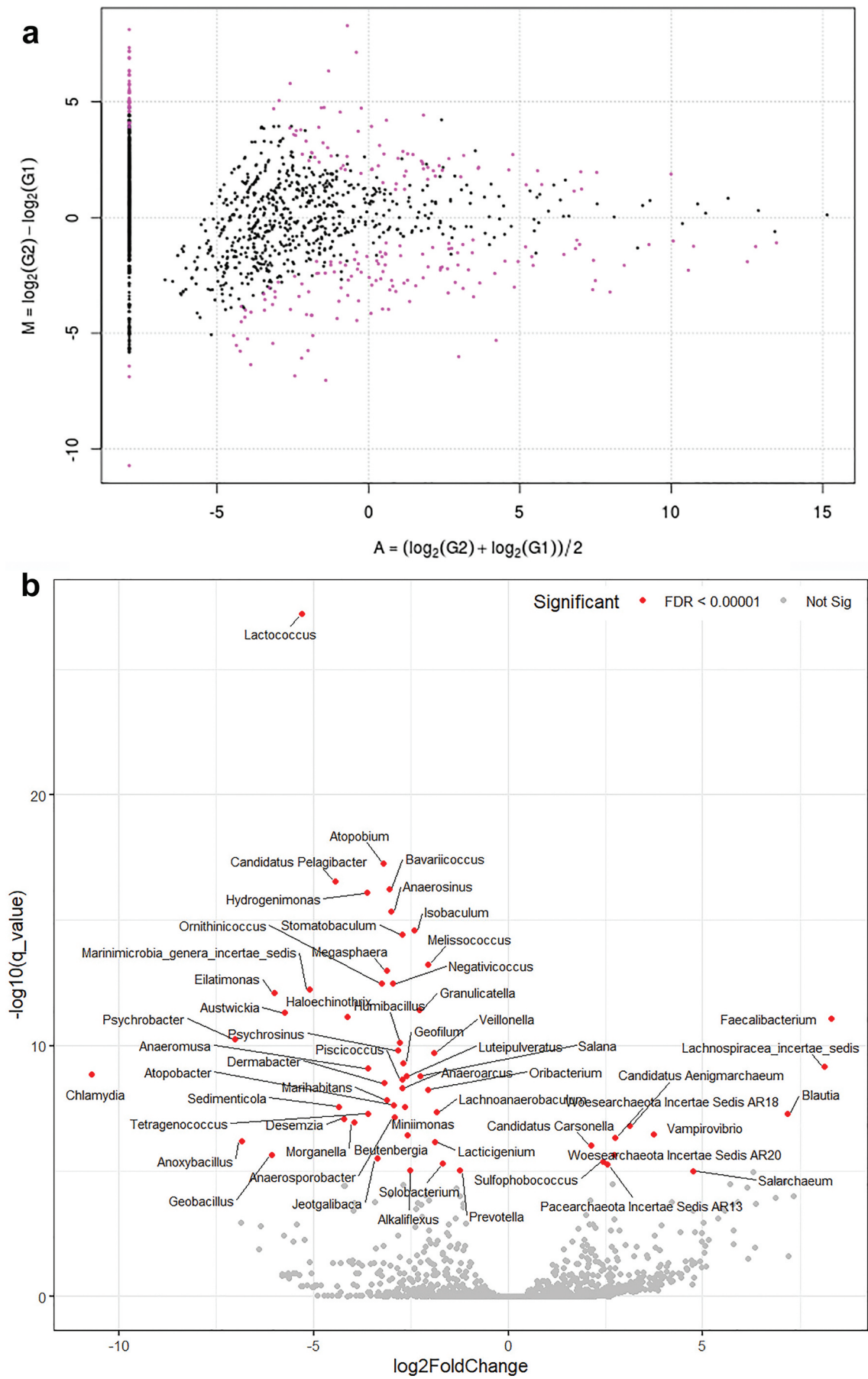


Figure 5. MA and volcano plots generated using analysis with tag count comparison for oral microbiome of glaucoma patients and control subjects.

a: MA plot of glaucoma vs control subjects (DE number of the upper panels 166 vs. that of the lower panels 171; $P < 0.05$)

b: Volcano plot of DEG in glaucoma vs control subjects *Lactococcus* and *Candidatus Pelagibacter* were downregulated in the glaucoma group
 DEG, differential expression of genes

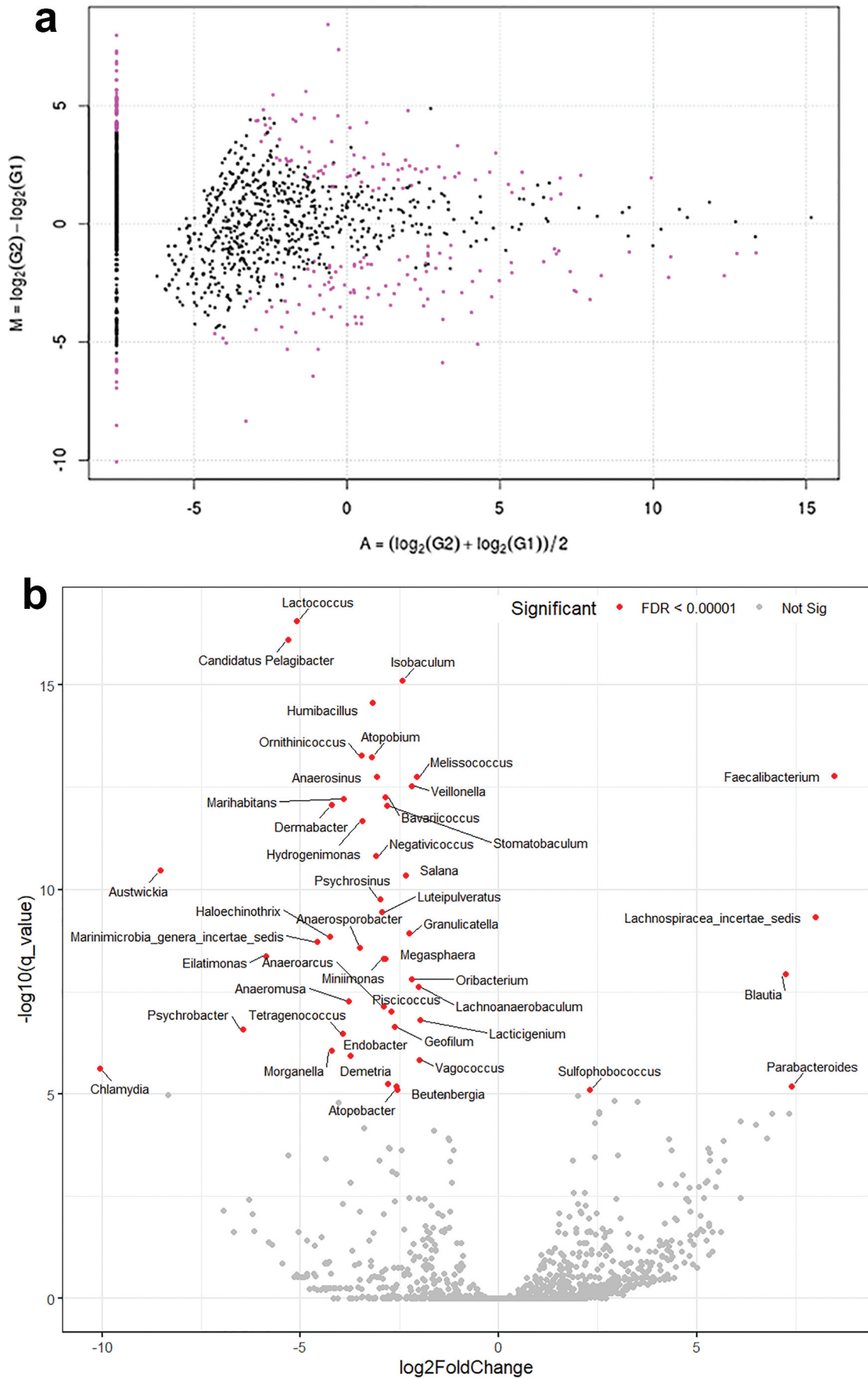


Figure 6. MA and volcano plots generated using subgroup analysis with tag count comparison for oral microbiome in the POAG subgroups compared with the control group.

a: MA plot of POAG vs control subjects (DE number of the upper panels 175 vs. that of the lower panels 142; $P < 0.05$)

b: Volcano plot of DEG in POAG vs control subjects DEG, differential expression of genes; POAG, primary open-angle glaucoma

Table 3. Multinomial logistic regression analysis of factors related taxon in glaucoma group compared with control subjects.

Taxa	coefficient	SD	Odd ratio	P-value	coefficient	SD	Odd ratio	P-value*				
Age < 60				Age ≥ 60								
<i>Lactococcus</i>	-210.481	55.386	3.88e ⁻⁹²	0.0001	-172.977	45.470	7.53e ⁻⁷⁶	0.0001				
<i>Endobacter</i>	0.223	1.750	1.25e ⁺⁰⁰	0.8984	-27.032	15.330	1.82e ⁻¹²	0.0778				
<i>Rikenella</i>	89.032	237.423	4.63e ⁺³⁸	0.7077	-80.714	338.227	8.83e ⁻³⁶	0.8114				
<i>Candidatus Pelagibacter Atopobium</i>	-560.729	205.488	3.01e ⁻²⁴⁴	0.0064	-921.479	259.819	0.00e ⁺⁰⁰	0.0004				
Baseline IOP < 20 mmHg				Baseline IOP ≥ 20 mmHg								
<i>Lactococcus</i>	-177.477	45.896	8.37e ⁻⁷⁸	0.0001	-191.253	52.588	8.70e ⁻⁸⁴	0.0003				
<i>Endobacter</i>	-13.841	12.697	9.75e ⁻⁰⁷	0.2757	0.094	1.730	1.10e ⁺⁰⁰	0.9567				
<i>Rikenella</i>	14.974	47.738	3.18e ⁺⁰⁶	0.7538	33.475	46.674	3.45e ⁺¹⁴	0.4732				
<i>Candidatus Pelagibacter Atopobium</i>	-509.789	171.543	3.99e ⁻²²²	0.0030	-319.291	145.031	2.16e ⁻¹³⁹	0.0277				
RNFL thickness ≥ 80 μm				RNFL thickness < 80 μm								
<i>Lactococcus</i>	-4.134	1.252	1.60e ⁻⁰²	0.0010	-3.561	1.310	2.84e ⁻⁰²	0.0065				
<i>Lactococcus</i>	-207.229	53.820	1.00e ⁻⁹⁰	0.0001	-173.639	45.928	3.89e ⁻⁷⁶	0.0002				
<i>Endobacter</i>	-13.957	15.484	8.68e ⁻⁰⁷	0.3674	-0.536	1.711	5.85e ⁻⁰¹	0.7542				
<i>Rikenella</i>	-38.900	295.676	1.28e ⁻¹⁷	0.8953	61.128	142.806	3.53e ⁺²⁶	0.6686				
<i>Candidatus Pelagibacter Atopobium</i>	-679.649	250.852	6.80e ⁻²⁹⁶	0.0067	-782.464	226.057	0.00e ⁺⁰⁰	0.0005				
Early glaucoma				Moderate glaucoma				Advanced glaucoma				
<i>Lactococcus</i>	-155.716	46.085	2.36e ⁻⁶⁸	0.0007	-231.783	65.006	2.18e ⁻¹⁰¹	0.0004	-218.896	60.662	8.60e ⁻⁹⁶	0.0003
<i>Endobacter</i>	-25.875	17.045	5.79e ⁻¹²	0.1290	-6.534	10.847	1.45e ⁻⁰³	0.5469	0.488	1.756	1.63e ⁺⁰⁰	0.7813
<i>Rikenella</i>	-35.096	155.035	5.73e ⁻¹⁶	0.8209	32.121	63.202	8.91e ⁺¹³	0.6113	43.848	62.576	1.10e ⁺¹⁹	0.4835
<i>Candidatus Pelagibacter Atopobium</i>	-1695.298	445.654	0.00e ⁺⁰⁰	0.0001	-1370.249	457.311	0	0.0027	-1342.329	430.780	0	0.0018
<i>Lactococcus</i>	-5.753	1.710	3.17e ⁻⁰³	0.0008	-15.354	5.126	2.15e ⁻⁰⁷	0.0027	-2.135	0.941	1.18e ⁻⁰¹	0.0232

P-value* was calculated using multinomial logistic regression analysis
 IOP: intraocular pressure, RNFL: retinal nerve fiber layer, SD: standard deviation

Table 4. Association rules generated by association rule mining for glaucoma using oral microbiome data.

Glaucoma rules					
Rule 1	<i>Atopobium</i> = [1] <i>Candidatus Pelagibacter</i> = [1] <i>Endobacter</i> = [1] <i>Lactococcus</i> = [1] → Glaucoma 0.96				
Rule 2	<i>Lactococcus</i> = [1] → Glaucoma 0.9				
Normal rules					
Rule 1	<i>Candidatus Pelagibacter</i> = [2] <i>Lactococcus</i> = [2] → Normal 0.92				
Rule 2	<i>Lactococcus</i> = [2] → Normal 0.85				
Rule 3	<i>Candidatus Pelagibacter</i> = [2] → Normal 0.7				
Rule 4	<i>Atopobium</i> = [2] → Normal 0.65				
Rule explanations (Rule 1)					
Original rule	<i>Atopobium</i> = [1] <i>Candidatus Pelagibacter</i> = [1] <i>Endobacter</i> = [1] <i>Lactococcus</i> = [1] → Glaucoma 0.96				
Rule explanations	If the compositions of <i>Atopobium</i> ≤ 0.347 <i>Candidatus Pelagibacter</i> ≤ 0.001 <i>Endobacter</i> ≤ 0.015 <i>Lactococcus</i> ≤ 0.013; the sample is Glaucoma with 96.0% accuracy				
Applied category for rule generation					
Category	<i>Atopobium</i>	<i>Candidatus Pelagibacter</i>	<i>Endobacter</i>	<i>Lactococcus</i>	<i>Rikenella</i>
[1]	≤ 0.347	≤ 0.001	≤ 0.015	≤ 0.013	≤ 0.001
[2]	0.347–2.908	0.001–0.014	0.015–1.106	0.013–0.889	0.001–0.159

CPAR (classification based on predictive Association Rules) algorithm used [24].

Several similarities may exist in glaucoma and neurodegenerative disease, as the optic nerve is a part of the central nerve system [38,39]. Hence, the microbiome study of Parkinson’s disease and Alzheimer’s disease may provide new insights into glaucoma research, suggesting that microbial dysbiosis can induce or exacerbate the disease via toxic effects from direct invasion or epigenetic changes [40]. One example of commensal pathogenic bacteria is *H. Pylori*, which has been associated with the risk of several neurodegenerative diseases, including Parkinson’s disease, Alzheimer’s disease, and POAG [41–43].

A major strength of our study is the inclusion of a relatively large number of glaucoma samples from

multiple centers using metagenomic analysis. In addition, our study used recently developed bioinformatics technology and machine learning for analysis. However, there are several limitations to this study. First, it was difficult to determine causality in this study; we could not determine whether the microbiome biomarker alone is a bystander or culprit. This study is worthwhile as an early study to suggest biomarkers that could be related to the identification of the severity of glaucoma. Additional verification is needed using cell-based and animal model studies for mechanisms. The validation study by PCR would provide for more reliable results, although the NGS technology has improved. Another limitation was the small sample size of the control group compared to that of glaucoma

Acknowledgments

We thank Theragen Bio for providing technical support for genome sequencing and metagenome analysis.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by VHS Medical Center Research Grant (grant number: VHSMC 19043), Republic of Korea.

Ethical approval

The study protocol was approved by the Institutional Review Board (IRB) of each participating center—Veterans Health Service Medical Center, Korea (IRB No. 2018-12-015 and IRB No. 2020-06-025) Pusan National University Hospital (IRB No. H-1904-001-077); Daegu Veterans Health Service Medical Center (IRB No. 2019-001); and Pusan National University Yangsan Hospital (IRB No.05-2019-022).

Informed consent

Written informed consent was obtained from all subjects before their enrollment in the study.

Financial disclosure

This study was supported by a VHS Medical Center Research Grant, Republic of Korea (grant number: VHSMC 19043)

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Quigley HA. Glaucoma. *Lancet*. 2011;377(9774):1367–1377.
- [2] Fechtner RD, Khouri AS. Evolving global risk assessment of ocular hypertension to glaucoma. *Curr Opin Ophthalmol*. 2007;18(2):104–109.
- [3] Nayyar A, Gindina S, Barron A, et al. Do epigenetic changes caused by commensal microbiota contribute to development of ocular disease? A review of evidence. *Hum Genomics*. 2020;14(1):11.
- [4] Chen H, Cho KS, Vu THK, et al. Commensal microflora-induced T cell responses mediate progressive neurodegeneration in glaucoma. *Nat Commun*. 2018;9(1):3209.
- [5] Polla D, Astafurov K, Hawy E, et al. A pilot study to evaluate the oral microbiome and dental health in primary open-angle glaucoma. *J Glaucoma*. 2017;26(4):320–327.
- [6] Gupta A. Harnessing the microbiome in glaucoma and uveitis. *Med Hypotheses*. 2015;85(5):699–700.
- [7] Astafurov K, Elhawy E, Ren L, et al. Oral microbiome link to neurodegeneration in glaucoma. *PLoS One*. 2014;9(9):e104416.
- [8] Wax MB, Tezel G, Saito I, et al. Anti-Ro/SS-A positivity and heat shock protein antibodies in patients with normal-pressure glaucoma. *Am J Ophthalmol*. 1998;125(2):145–157.
- [9] Almasieh M, Wilson AM, Morquette B, et al. The molecular basis of retinal ganglion cell death in glaucoma. *Prog Retin Eye Res*. 2012;31:152–181.
- [10] Ramirez AI, de Hoz R, Salobrar-Garcia E, et al. The role of Microglia in retinal neurodegeneration: Alzheimer's disease, Parkinson, and glaucoma. *Front Aging Neurosci*. 2017;9:214.
- [11] Lin IC, Wang YH, Wang TJ, et al. Glaucoma, Alzheimer's disease, and Parkinson's disease: an 8-year population-based follow-up study. *PLoS One*. 2014;9(10):e108938.
- [12] Qin J, Li R, Raes J, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59–65.
- [13] Lu LJ, Liu J. Human microbiota and ophthalmic disease. *Yale J Biol Med*. 2016;89:325–330.
- [14] Lin P. Importance of the intestinal microbiota in ocular inflammatory diseases: a review. *Clin Exp Ophthalmol*. 2019;47(3):418–422.
- [15] Sandhu KV, Sherwin E, Schellekens H, et al. Feeding the microbiota-gut-brain axis: diet, microbiome, and neuropsychiatry. *Transl Res*. 2017;179:223–244.
- [16] van de Wouw M, Schellekens H, Dinan TG, et al. Microbiota-Gut-Brain axis: modulator of host metabolism and appetite. *J Nutr*. 2017;147(5):727–745.
- [17] Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013;41(1):e1.
- [18] Edgar RC, Haas BJ, Clemente JC, et al. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27(16):2194–2200.
- [19] Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37(8):852–857.
- [20] Lopez-Garcia A, Pineda-Quiroga C, Atxaerandio R, et al. Comparison of Mothur and QIIME for the analysis of rumen microbiota composition based on 16S rRNA amplicon sequences. *Front Microbiol*. 2018;9:3010.
- [21] Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*. 2014;42:D633–42.
- [22] Wright M, Ziegler A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw*. 2017;77:1–17.
- [23] Chang HX, Haudenschild JS, Bowen CR, et al. Metagenome-Wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Front Microbiol*. 2017;8:519.
- [24] Yin X, Cpar: HJ Classification based on predictive association rules. *Proceedings of the 2003 SIAM International Conference on Data Mining*; San Francisco, CA: SIAM; 2003. p. 331–335.
- [25] Sun J, Nishiyama T, Shimizu K, et al. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*. 2013;14(1):219.

- [26] Agrawal R, Imieliński T, Swami A Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD international conference on Management of data; Washington D.C. USA; 1993. p. 207–216.
- [27] Benjamini YHY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;57:289–300.
- [28] Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med*. 1990;9(7):811–818.
- [29] Chaiwiang N, Poyomtip T. Microbial dysbiosis and microbiota-gut-retina axis: the lesson from brain neurodegenerative diseases to primary open-angle glaucoma pathogenesis of autoimmunity. *Acta Microbiol Immunol Hung*. 2019;66(4):541–558.
- [30] Poyomtip T. Roles of Toll-Like receptor 4 for cellular pathogenesis in primary open-Angle glaucoma: a potential therapeutic strategy. *J Microbiol Immunol Infect*. 2019;52(2):201–206.
- [31] Martin KR, Quigley HA, Zack DJ, et al. Gene therapy with brain-derived neurotrophic factor as a protection: retinal ganglion cells in a rat glaucoma model. *Invest Ophthalmol Vis Sci*. 2003;44(10):4357–4365.
- [32] Leylabadlo HE, Ghotaslou R, Feizabadi MM, et al. The critical role of faecalibacterium prausnitzii in human health: an overview. *Microb Pathog*. 2020;149:104344.
- [33] Gong H, Zhang S, Li Q, et al. Gut microbiota compositional profile and serum metabolic phenotype in patients with primary open-angle glaucoma. *Exp Eye Res*. 2020;191:107921.
- [34] Beilharz JE, Kaakoush NO, Maniam J, et al. The effect of short-term exposure to energy-matched diets enriched in fat or sugar on memory, gut microbiota and markers of brain inflammation and plasticity. *Brain Behav Immun*. 2016;57:304–313.
- [35] Cao L, Graham SL, Pilowsky PM. Carbohydrate ingestion induces differential autonomic dysregulation in normal-tension glaucoma and primary open angle glaucoma. *PLoS One*. 2018;13(6):e0198432.
- [36] Ma D, Wang AC, Parikh I, et al. Ketogenic diet enhances neurovascular function with altered gut microbiome in young healthy mice. *Sci Rep*. 2018;8(1):6670.
- [37] Zarnowski T, Tulidowicz-Bielak M, Kosior-Jarecka E, et al. A ketogenic diet may offer neuroprotection in glaucoma and mitochondrial diseases of the optic nerve. *Med Hypothesis Discov Innov Ophthalmol*. 2012;1:45–49.
- [38] Gupta N, Yucel YH. Glaucoma as a neurodegenerative disease. *Curr Opin Ophthalmol*. 2007;18(2):110–114.
- [39] Danesh-Meyer HV, Levin LA. Glaucoma as a neurodegenerative disease. *J Neuroophthalmol*. 2015;35(Suppl 1):S22–8.
- [40] Spielman LJ, Gibson DL, Klegeris A. Unhealthy gut, unhealthy brain: the role of the intestinal microbiota in neurodegenerative diseases. *Neurochem Int*. 2018;120:149–163.
- [41] Kountouras J, Zavos C, Zeglinas C, et al. Helicobacter pylori-Related impact on glaucoma pathophysiology. *Invest Ophthalmol Vis Sci*. 2015;56(13):8029–8030.
- [42] Dardiotis E, Tsouris Z, Mentis AA, et al. H. pylori and Parkinson's disease: meta-analyses including clinical severity. *Clin Neurol Neurosurg*. 2018;175:16–24.
- [43] Fani L, Wolters FJ, Ikram MK, et al. Helicobacter pylori and the risk of dementia: a population-based study. *Alzheimers Dement*. 2018;14(10):1377–1382.