

# Application of Pattern Recognition Analysis to Optimize Hemifield Asymmetry Patterns for Early Detection of Glaucoma

Jack Phu<sup>1,2</sup>, Sieu K. Khuu<sup>2</sup>, Bang V. Bui<sup>3</sup>, and Michael Kalloniatis<sup>1,2</sup>

<sup>1</sup> Centre for Eye Health, University of New South Wales, Kensington, NSW, Australia

<sup>2</sup> School of Optometry and Vision Science, University of New South Wales, Kensington, NSW, Australia

<sup>3</sup> Department of Optometry and Vision Science, University of Melbourne, Parkville, VIC, Australia

**Correspondence:** Michael Kalloniatis, Centre for Eye Health and School of Optometry and Vision Science, University of New South Wales, Kensington, NSW, Australia. e-mail: m.kalloniatis@unsw.edu.au

**Received:** 5 June 2018

**Accepted:** 18 July 2018

**Published:** 4 September 2018

**Keywords:** Humphrey Field Analyzer; Glaucoma Hemifield Test; Octopus; asymmetry; visual fields; perimetry

**Citation:** Phu J, Khuu SK, Bui BV, Kalloniatis M. Application of pattern recognition analysis to optimize hemifield asymmetry patterns for early detection of glaucoma. *Trans Vis Sci Tech.* 2018;7(5):3, <https://doi.org/10.1167/tvst.7.5.3>  
Copyright 2018 The Authors

**Purpose:** To assess the diagnostic utility of a new hemifield asymmetry analysis derived using pattern recognition contrast sensitivity isocontours (CSIs) within the Humphrey Field Analyzer (HFA) 24-2 visual field (VF) test grid. The performance of an optimal CSI-derived map was compared against a commercially available clustering method (Glaucoma Hemifield Test, GHT).

**Methods:** Five hundred VF results of 116 healthy subjects were used to determine normative distribution limits for comparisons. Pattern recognition analysis was applied to HFA 24-2 sensitivity data to determine CSI theme maps delineating clusters for hemifield comparisons. Then, 1019 VF results from 228 glaucoma patients were assessed using different clustering methods to determine the true-positive rate. We also assessed additional 354 VF results of 145 healthy subjects to determine the false-positive rate.

**Results:** The optimum clustering method was the CSI-derived seven-theme class map, which identified more glaucomatous VFs compared with the GHT map. The seven-class theme map also identified more cases compared with the five-, six-, and eight-class maps, suggesting no effect of number of clusters. Integrating information regarding the location of glaucomatous defects to the CSI clusters did not improve detection rate.

**Conclusions:** A clustering map derived using CSIs improved detection of glaucomatous VFs compared with the currently available GHT. An optimized CSI-derived map may serve as an additional means to aid earlier detection of glaucoma.

**Translational Relevance:** Pattern recognition-derived theme maps provide a means for guiding test point selection for asymmetry analysis in glaucoma assessment.

## Introduction

The clinical standard of assessment of visual function in glaucoma is visual field (VF) testing using standard automated perimetry (SAP; reviewed in Phu et al.<sup>1</sup> and Jampel et al.<sup>2</sup>). A typical set of criteria for a statistically significant VF defect for glaucoma within both research and clinical settings includes three or more contiguous points flagged as statistically abnormal following the distribution of the retinal nerve fiber layer (RNFL) and abnormal global automated indices.<sup>3–8</sup>

Commercially available instruments, such as the Humphrey Field Analyzer (HFA), commonly report an asymmetry analysis as an additional index of suspicion for glaucomatous VFs (e.g., the Glaucoma Hemifield Test, GHT).<sup>9</sup> As early glaucoma typically presents with VF defects that are asymmetric about the horizontal midline, there is a role for an index that compares superior and inferior hemifield sensitivities.<sup>9,10</sup> Specifically, the GHT, which is widely used in clinical studies as a criterion for a glaucomatous VF, compares the probability scores of abnormality within mirrored zones symmetrical about the horizontal midline to determine normality.

In contrast to the fixed and mirrored zones of the GHT, alternative clusters of test locations have been suggested. The EyeSuite software of the Octopus perimeter (Haag-Streit, Mason, OH) presents asymmetric clusters that are thought to represent the paths of the RNFL bundles, which themselves are asymmetric about the horizontal midline.<sup>11</sup> However, asymmetry analysis is not performed on these clusters in current instruments. Other methods of clustering have also been previously suggested, including correlating the deviation of sensitivities of glaucoma patients.<sup>12,13</sup> Specifically, others have used clustering methods to characterize the location of VF defects in patients with glaucoma, and to apply these clustered maps to then diagnose or detect progression in others.<sup>14-17</sup> Notably, these approaches have used the VF results of patients with pre-existing disease to generate clustered maps.

More recently, contrast sensitivity isocontours (CSIs), which group the test locations within the VF with the same sensitivity signature, have been identified within SAP test grids.<sup>18</sup> These CSIs appear to correlate with kinetic perimetry isopters,<sup>18,19</sup> and, importantly, do not necessarily coincide with zones identified by the GHT and EyeSuite. The difference in clustering method may be related to the method used to derive the clusters. CSIs use a sensitivity (or functional) basis, while the GHT uses an anatomically inspired basis for grouping.

Another interesting difference between CSIs and the GHT is the number of clusters grouped about the midline. The GHT uses five clusters within each hemifield, while the number of CSIs may not necessarily be fixed or restricted. Grouping a greater number of points by limiting the number of clusters may render it more difficult to identify statistically significant asymmetries, due to distribution limits that are then defined by inclusion of spurious test locations not contributing to the same isocontour. As the purpose of a hemifield asymmetry index is to provide a means with which to identify early functional loss in glaucoma, it remains to be seen whether altering the pattern and number of clusters for comparison would result in increased detection rates. The application of such an approach allows pooling of the underlying normative data to identify deviations in sensitivity from the normal expected isocontour while making no assumptions about the type of glaucomatous defect, which may vary significantly among individual patients.<sup>20</sup>

In the present study, we sought to compare the diagnostic ability of existing clustering maps with

CSIs derived using pattern recognition analysis. Specifically, we tested three hypotheses in this study. First, we tested the hypothesis that a CSI theme map (i.e., a functional basis of test point grouping) provides greater diagnostic ability for detecting glaucomatous VFs compared with an existing map that uses a structurally inspired basis (GHT). Such a hemifield asymmetry analysis is hypothesized to yield improved diagnostic ability as it may highlight deviations from normal expected symmetries in sensitivity values. Second, we hypothesized that the difference in diagnostic ability was not solely dependent upon the number of clusters (i.e., that an arbitrary separation into multiple classes does not perform similarly to a CSI-derived map with the same number of clusters). In doing so, we also tested the hypothesis that there would be no effect of the number of clusters upon diagnostic performance for CSI-derived maps. Finally, we tested the hypothesis that the addition of glaucoma-specific defect information to describe a CSI-derived theme map would further improve diagnostic ability compared with arbitrary clusters.

## Methods

### Subjects and Patients

The medical charts of patients seen at the Centre for Eye Health (CFEH, Sydney, Australia) between January and December 2017 were retrospectively reviewed. We examined three diagnostic groups in the present study as follows: a healthy cohort for construction of the normative database, a second healthy cohort for testing the false-positive rate, and a glaucoma cohort to test the true-positive rate (TPR).

### Healthy Cohort for Normative Database

The VF results used to derive normative asymmetry data consisting of the 500 VFs of 116 healthy subjects (consisting of patients and staff of CFEH and the University of New South Wales who were deemed to be healthy) that we have recently reported.<sup>21</sup> The inclusion criteria included visual acuities 20/25 or better, normal pupil reactions, open angles on gonioscopy, intraocular pressures lower than 21 mm Hg, and normal fundus examination following pharmacologic pupillary dilation, and biomicroscopic fundus lens examination (with normal optic nerve head and macula), supplemented by optical coherence tomography (Cirrus HD-OCT; Carl Zeiss Meditec, Dublin, CA) and VF testing (HFA 24-2 Swedish

Interactive Thresholding Algorithm [SITA]-Standard; Carl Zeiss Meditec). Subjects also needed to have no cluster of VF defects (>3 contiguous points of depression at the  $P < 5\%$  level, one of which needs to be at the  $P < 1\%$ ), and no mean deviation (MD) or pattern standard deviation (PSD) flagged at a  $P < 5\%$  level, and GHT had to be within normal limits. The exclusion criteria used by Bengtsson and Heijl<sup>22</sup> also noted that suspicious VF defects must be explained by ocular status. However, we also excluded results with suspicious VF “defects” if they could be explained by other artefacts, such as lens scotomas, blepharoptosis, or inattention,<sup>23–27</sup> as these do not represent normal VF results. This cohort essentially served as the development cohort for the subsequent clustered theme maps.

### Healthy Cohort for Testing False-Positive Rates

We also performed a subgroup analysis that included healthy subjects without the requirement to meet the above VF criteria. This healthy subject test group (354 VFs of 145 subjects) consisted of subjects who were clinically healthy but could have had VF defects that were judged to not be attributable to disease, as the use of VF criteria could introduce biases that artificially reduce the false-positive rate (i.e., improve specificity). The purpose of this analysis was to validate the normative distribution: an excessive amount of false-positive results would suggest that the normative distribution is supranormal (though the biases are likely to be small when using a large database).<sup>28</sup> Note that both healthy cohorts were randomly sampled from our clinic files, and therefore consisted of patients referred to an ophthalmic center for testing who were later found to be free of pathology and also staff working within the center. As we set out to use a  $P < 0.01$  (1<sup>st</sup> percentile) as the lower limit of normality (see below section on the normative distribution), we were operating at a fixed-specificity level of 99%. This healthy cohort therefore served to validate this specificity level.

### Glaucoma Cohort for Testing True-Positive Rates

The patients with glaucoma were diagnosed and/or managed following clinical evaluation at CFEH<sup>29,30</sup> and were further reassessed using the following criteria for inclusion in the present study. Structural findings for glaucoma included the following, as described in previous studies<sup>4,5,8,21,31,32</sup>: enlargement

of the optic cup (i.e., cup-disk ratio), local or diffuse thinning of the neuroretinal rim, thinning of the adjacent RNFL, peripapillary atrophy, and/or the presence of disk hemorrhages. A corresponding VF defect was defined as one that corresponded retinotopically with the structural loss, and included the following criteria: a cluster of depression (at least 3 contiguous points of depression at the  $P < 0.05$  level, of which at least one is depressed at the  $P < 0.01$  level), a PSD flagged at the  $P < 0.05$  level, and a GHT result that was outside normal limits ( $P < 0.01$  level). However, glaucoma patients did not necessarily need to have a VF defect (i.e., they instead met the criteria for a “normal” VF as described above; “preperimetric glaucoma” PPG). These patients were analyzed separately to patients meeting conventional VF “fail” criteria.<sup>4,5</sup> This group was included to ensure that the VF criteria did not introduce biases into the glaucoma group, as requiring a VF defect could artificially raise the TPR. In total, 1019 VFs of 228 patients with open-angle glaucoma were analyzed. For all healthy and glaucoma subjects, they were required to have no evidence of other systemic or ocular diseases that would affect the visual pathway, nor any prior ocular surgery aside from routine and uneventful cataract surgery (must not affect the VF result) or laser trabeculectomy (for glaucoma patients).

The characteristics of all subjects are shown in Table 1. Sensitivity values (in dB) were extracted directly from the HFA printout. Left eye results were converted into the equivalent right eye orientation for analysis. In order to be included for analysis, VF results had to meet the following reliability criteria: less than 33% fixation losses, less than 33% false-positives, and less than 20% false-negatives.<sup>33</sup> For glaucoma patients, the false-negative criterion was less than 40%, to account for increasing depth of VF defect. Fixation losses were also used in conjunction with the HFA gaze tracker: if the angular deviation of fixation exceeded 3° more than 20% of the time, then the result was excluded.<sup>34,35</sup>

### Age-Correction and Pooling of Sensitivity Results

Sensitivity results from the SITA algorithm are not direct measurements of sensitivity, but are modulated as per a proprietary probabilistic function within the algorithm.<sup>36</sup> We applied our recently published methods for age-correcting SITA sensitivity measurements, which, in short, was the point-wise application of a decibel per year factor to correct sensitivities of

**Table 1.** Demographic and Clinical Characteristics of Research Participants

	Normal for Normative Database (500 VFs of 116 Subjects)	Glaucoma (1019 VFs of 228 Patients)	Additional Normal Test Group for Assessing False-Positives (354 VFs of 145 Subjects)
Age, mean $\pm$ SD, y <sup>a</sup>	59.5 $\pm$ 8.0	61.6 $\pm$ 11.9	53.5 $\pm$ 10.5
Sex, <i>n</i> , male:female <sup>b</sup>	46:70	86:142	69:76
Spherical equivalent refractive error (median and IQR, diopters) <sup>b</sup>	0.00 (−1.13, +0.75)	0.00 (−0.88, +1.00)	−0.25 (−1.66, +0.75)
Number of VFs per patient per eye (median and IQR) <sup>b</sup>			
Right eye	2 (1, 3)	2 (1, 3)	1 (1, 3)
Left eye	2 (1, 3)	2 (1, 3.8)	1 (1, 3)
MD (median and IQR, dB) <sup>a,c</sup>	0.22 (−0.61, 0.94)	−1.73 (−3.58, −0.52)	0.06 (−0.80, 0.85)
PPG <sup>d</sup>		299	
Up to −1 dB		142	
Up to −2 dB		280	
Up to −3 dB		397	
Up to −4 dB		486	
Up to −5 dB		563	
Up to −6 dB		612	
Worse than −6 dB		108	
PSD (median and IQR, dB) <sup>a</sup>	1.46 (1.31, 1.61)	2.31 (1.79, 3.75)	1.49 (1.35, 1.66)

IQR, interquartile range.

<sup>a</sup> Analyzed per individual VF.

<sup>b</sup> Analyzed per individual participant.

<sup>c</sup> Separated into MD bins for glaucoma patients.

<sup>d</sup> PPG was defined as a VF result without a cluster of three or more contiguous points of reduced sensitivity at the  $P < 0.05$  level, of which one reaches at least the  $P < 0.01$  level.

individual patients to a 50-year-old equivalent subject or patient.<sup>21</sup> This therefore allowed pooling of test results for comparison, which were then used to test the three main hypotheses in the study.

### Hypothesis 1: GHT Versus Pattern Recognition CSI Theme Maps

The first hypothesis of the study was that the diagnostic utility of a structurally inspired map (GHT) is lower compared with a functionally derived map (CSI theme classes) for detecting hemifield anomalies in glaucoma. Therefore, the first step of the present study was to characterize the pattern recognition derived CSI-theme maps of SITA-Standard sensitivities across the 24-2 test grid in the healthy cohort, as there are differences in the sensitivity measurement between SITA and full threshold, which we have recently described.<sup>18</sup>

Pattern recognition is an iterative procedure that

classifies pixel values into clusters, each with its own mean and boundary, eventually obtaining clusters forming individual theme classes that are statistically separable (also see [Supplementary Fig. S1](#)).<sup>18,35,37–39</sup>

In short, VF sensitivities at each test location within the 24-2 test grid were converted into scaled grayscale pixel values ranging from 0 (lower dB) to 255 (higher dB). The reason for converting the decibel values to grayscale pixels is because the program performs clustering using images, rather than other hierarchical clustering algorithms in which the input values are solely numeric (see supplementary material in Phu et al.<sup>18</sup>). In total, five grayscale images were generated, each representing the average sensitivities of 100 randomly allocated, age-corrected normal VF results. Five images were used as this allowed us to average 100 VF results to construct each image to overcome potential individual variability in VF result, while simultaneously providing enough inputs for the program. We have previously shown that the scaling

and number of input layers is relatively unimportant, except to provide enough dynamic range to visually appreciate the different classes (see Discussion for further details).<sup>18,35</sup> Unsupervised classification with Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) in PCI Geomatica (Version 10; PCI Geomatics, Richmond Hill, Ontario, Canada) was used for analysis (also see [Supplementary Material](#)); ISODATA is a subtype of  $k$  means clustering (a migrating means method) with automated splitting of high variance classes and merging of highly overlapping classes. Each class had a minimum transformed divergence ( $D_T$ ) value greater than 1.86, which represents a greater than 96% correct classification.<sup>40</sup> Each of these classes represented a unique CSI, guiding the generation of normative hemifield data for CSI-derived groups. The pattern recognition analysis then allows the generation of a pseudocolor theme map, in which points with the same sensitivity signature are identified by the same color (Figs. 1B–G).

Notably, the technique used in our study differs from those previously reported. Other studies have used cluster analysis to identify patterns of VF defects appearing in patients with glaucoma both cross-sectionally<sup>12,14,41</sup> and to detect progression.<sup>15–17</sup> In contrast to studies using mixture of Gaussian models,<sup>15–17</sup>  $k$  means presents a stricter class assignment for a particular datum point with the assumption that the point is highly certain to belong the assigned class, while a mixture of Gaussian incorporates a degree of uncertainty into class assignment. Their utility has been debated in the literature.<sup>42,43</sup> In the present study, as we approximately knew the anticipated number of resultant classes for VF test location assignment,<sup>18,44</sup> we applied the ISODATA algorithm. As a further point of contrast with the work described by previous authors, which identified clusters of VFs, such as patterns of VF loss,<sup>14–17</sup> application of clustering in the present study was to identify clusters of VF spatial test locations that is facilitated by the use of satellite imaging algorithms. Finally, the statistical rigidity of the classification is tested using the  $D_T$  statistic.

The GHT only considers 44 of 54 test locations of the 24-2 for its clusters: two test locations adjacent to the physiological blind spot and the eight points around those points within the temporal field are excluded (Fig. 1A). In order to compare diagnostic ability between GHT and CSI-derived maps, pattern recognition analysis was applied to the 44 test locations used by the GHT. Exclusion of the 10

points, which are not necessarily commonly affected in glaucoma,<sup>10</sup> to match the GHT test grid would assist in reducing variability for identifying abnormal VFs. A separate analysis showed no difference in the resultant CSI theme maps when using 44 or 54 locations.

## Hypothesis 2: Optimizing the Number of Classes for Each Theme Map

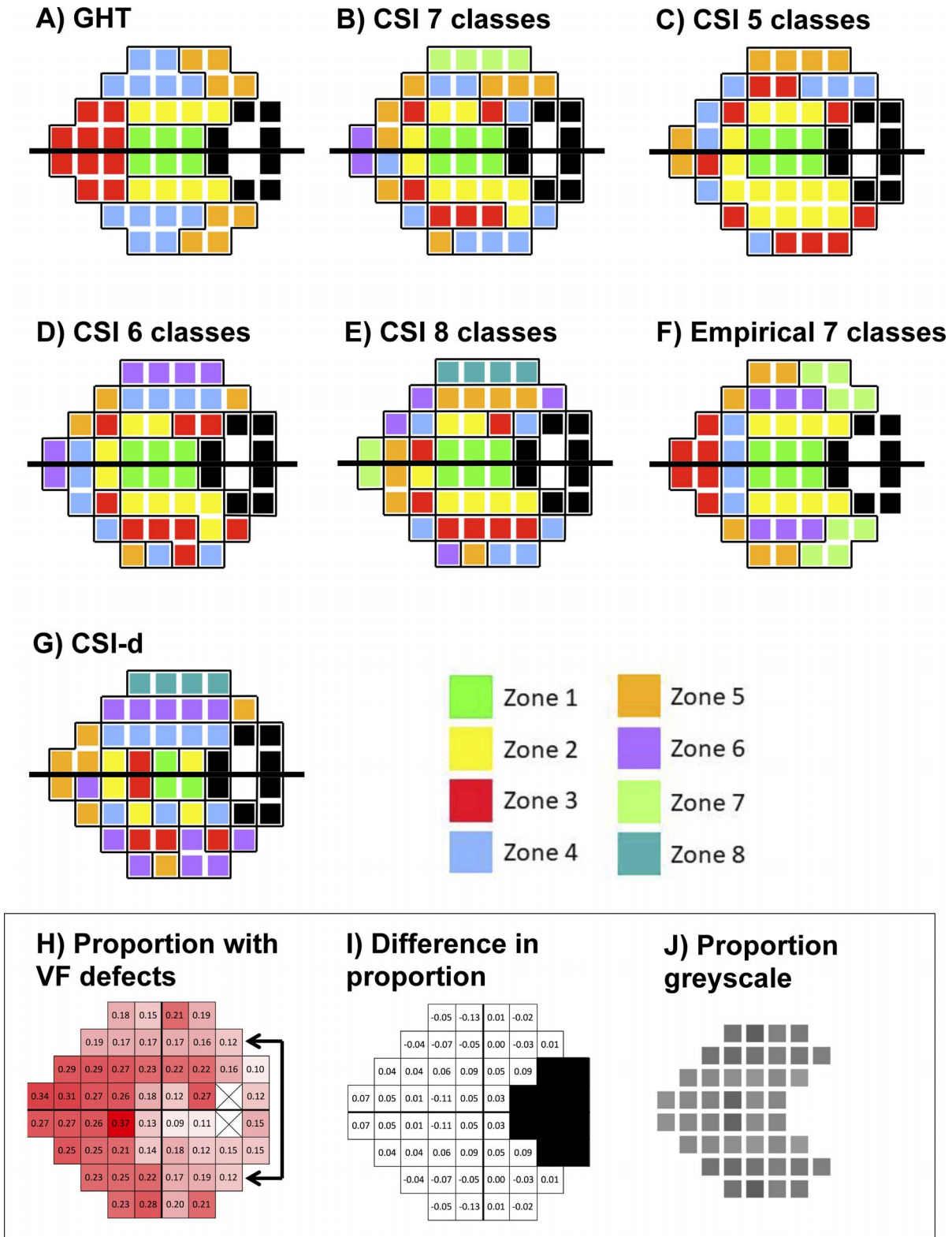
Due to the limited number of test locations within the 24-2, there was a limited number of possible zones that could be delineated within the superior and inferior hemifields before individual point pairs are compared. Therefore, the second hypothesis tested in the study was that there is an effect of number of theme classes upon diagnostic utility up to critical number of ideal classes.

Since the 24-2 lacks the “outer ring” of test locations apparent in the 30-2,<sup>18</sup> we used the seven-class theme map as the optimal pattern recognition theme map for testing the first hypothesis in the study (Fig. 1B). Due to asymmetries across the VF,<sup>45,46</sup> this theme map had an additional superior zone with no corresponding inferior zone (zone 7, Fig. 1B). This zone was analyzed by itself (a “sector” analysis) for an abnormal result (see below pass/fail criteria).

Here, we tested different numbers of functionally derived theme classes, including five (the minimum number of classes tested by the GHT), six-, and eight-class theme maps (Figs. 1C–E, respectively). To further examine whether any improvement in diagnostic utility from the theme map was due to the change in number of compared classes, we also divided the test points into seven mirrored zones, similar to the GHT. The two zones of the GHT with five or more test points (the nasal zone and the superonasal zones) were divided into smaller zones; we refer to this as the empirical seven-class map (Fig. 1F), which represented a division of test locations into groups with equal numbers of points in clustered locations across each hemifield, serving as a comparison for the seven-class CSI-derived theme map.

## Hypothesis 3: Adding Glaucoma-Specific Data to the Theme Map

Finally, we tested a combined CSI and defect map (CSI-d; Fig. 1G) to examine the effect of adding glaucoma-specific data to the theme map upon diagnostic utility. The map was derived using the sensitivities of the healthy cohort, and also information regarding the proportion of statistically signifi-



**Figure 1.** The theme maps used for hemifield asymmetry analysis. GHT (A) clusters are as per current available clusters. The CSI-derived seven-class theme map (B) was hypothesized to be the optimal map, based off our previous work.<sup>18</sup> Five- (C), six- (D), and eight- (E) class theme maps were also tested. (F) An empirically divided seven-class map based off the GHT clusters. (G) A theme map combining CSIs

← and information regarding prevalence of glaucomatous VF defects in the present cohort (H). The differences in proportions between symmetric mirrored points were obtained (I) and were normalized and converted into grayscale plots (J) to be added as another layer for pattern recognition analysis, on top of existing sensitivity grayscales.

cant VF defects ( $P < 0.05$  on the pattern deviation map) across all glaucoma patients. In addition to using the same sensitivity inputs as per the methods described above for CSI maps, it also added another layer for pattern recognition analysis describing the proportion of VF defects occurring in the glaucoma patients within the present study (Fig. 1H). As the aim was to identify asymmetries in glaucomatous VF defects, we subtracted the proportions of defects across each hemifield at mirrored test locations (e.g.,  $0.35 - 0.2 = 0.15$  proportion, Fig. 1I), normalized this value, and added it as a layer for pattern recognition analysis (Fig. 1J) alongside the standard sensitivity grayscales. In this manner, the CSI-d was a combined map of normal sensitivities further enhanced by common asymmetries in the glaucomatous VF.

### The Normative Distribution and Probability Score Calculation: Pass/Fail Criteria

The sensitivities of the 500 healthy VF results were pooled together to generate the mean, standard deviation (SD), and probability score ( $P$  score) distribution limits for each of the clustering methods. Point-wise, location-specific probability scores were calculated using a method similar to the work of Asman and Heijl.<sup>9</sup> To generate the  $P$  score, the individual sensitivity first was compared with the mean and SD at that location to generate a  $Z$  score. The  $P$  score was then determined from  $P$  value calculated from the normal distribution (Table 2).

The following two criteria for hemifield analysis were used: (1) asymmetry in the superior-inferior distributions (the “asymmetry” criterion), and (2) individual matched-sector pair distributions analyzed separately (the “sector pair” criterion) for each cluster (summarized in Fig. 2A). For the asymmetry criterion, the sum of  $P$  scores within the superior sector was determined for each healthy subject. The sum of  $P$

scores within the inferior sector was then subtracted from this to get a superior-inferior difference in  $P$  score. The upper (i.e., positive difference, where the superior cluster has a higher  $P$  score) and lower (i.e., negative difference, where the inferior cluster has a higher  $P$  score) 0.5<sup>th</sup> percentiles across the 500 normal VF results was obtained. This therefore represents the 1% two-tailed distribution limits, with  $P$  scores outside the upper and lower 0.5<sup>th</sup> percentiles were considered to be outside normal limits (ONL). For the sector pair criterion, the 0.5<sup>th</sup> percentile of each individual sector pairs (superior and inferior) was determined. This represents a 0.5<sup>th</sup> one-tailed distribution limit, whereby  $P$  scores results that were outside the 0.5<sup>th</sup> percentile for both superior and inferior sector pairs were considered ONL. We only report on within normal limits (WNL) and ONL results; borderline and other possible results such as generalized reduction were not examined.

In the GHT, normative comparisons are performed by comparing individual test locations to its matching underlying normal location (Fig. 2B). In comparison, the pattern recognition pseudocolor theme maps indicate locations with the same sensitivity signature, and thus the underlying normative data were pooled to obtain descriptive statistics from a larger overall sample (mean, SD, and distribution limits) for normative comparisons (Fig. 2C). As the GHT clusters consist of points with different sensitivity signatures (i.e., they come from different CSIs), the sensitivity data were not pooled in a similar fashion. As the sensitivities are different, they would therefore contribute different levels of variability, thus confounding the resultant distribution.

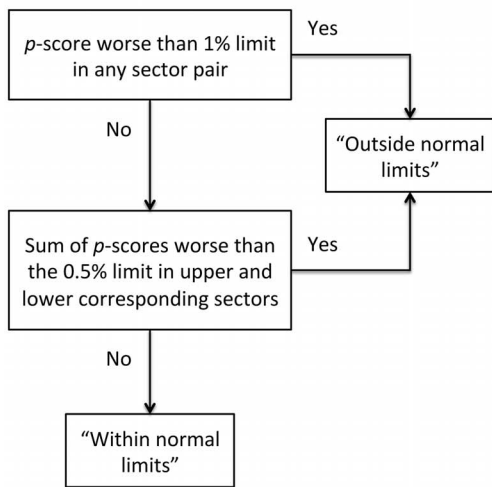
### Bootstrapping the Normative Data

To ensure that the probabilities of abnormalities were not confounded by the sample size and the unequal number of VFs contributed by each healthy subject toward the normative database, we applied a nonparametric bootstrap (resampling with replacement)<sup>21,47,48</sup> to generate the final normative data set for comparison. Applications to VF data have been described in our recent paper.<sup>21</sup> In short, we resampled (with replacement) a subset of the data (set size  $x = 500$ ) from the original cohort. This

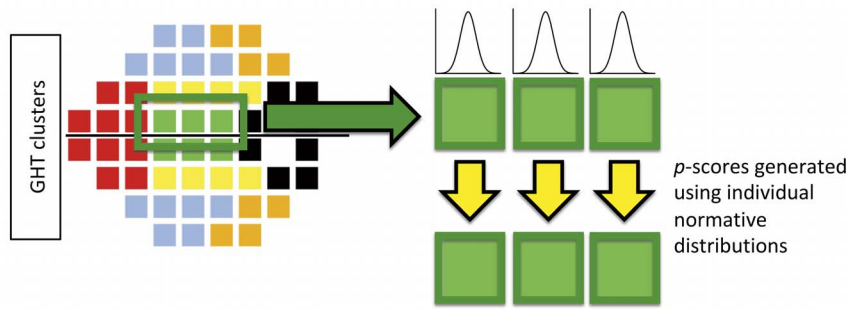
**Table 2.**  $P$  Scores According to the Calculated  $P$  Value, as Per the Methods of Asman and Heijl<sup>9</sup>

$P < 0.05$	2
$P < 0.02$	5
$P < 0.01$ or worse	10 multiplied by difference from normal mean

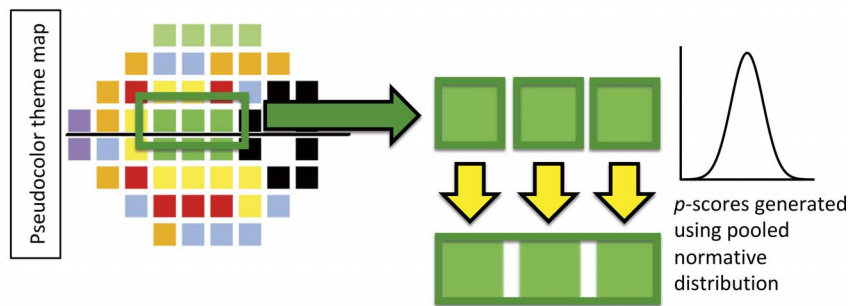
### A) Flow chart for analysis



### B) Point-wise $p$ -score comparison using the GHT grid



### C) Pooled $p$ -score generated using pattern recognition clusters



**Figure 2.** (A) The decision-making flow chart for statistical analysis, adapted from the work of Asman and Heijl,<sup>9</sup> but simplified to include only binary outcomes of “outside normal limits” and “within normal limits.” Comparison of point-wise (B) and pooled (C) methods used to derive the  $P$  scores. In (B), each point is tested against the distribution at the individual location to obtain a point-wise  $P$  value and  $P$  score. Each *color* represents points within mirrored clusters across the horizontal midline as denoted by the GHT. In (C), each point within a CSI theme class is tested against the pooled normative distribution, which is derived from all sensitivities within the class, as they share a sensitivity signature, as per Phu et al.<sup>18</sup> In contrast to (B), each *color* represents points within mirrored CSIs that share the same sensitivity signature. This further contrasts with (B), as the GHT clusters consist of points from different CSIs, which were therefore not pooled to generate a combined normative distribution.

process was repeated 200 times. The resampling technique was used to generate bootstrapped 0.5<sup>th</sup> distribution limits as described above for final normative comparison.

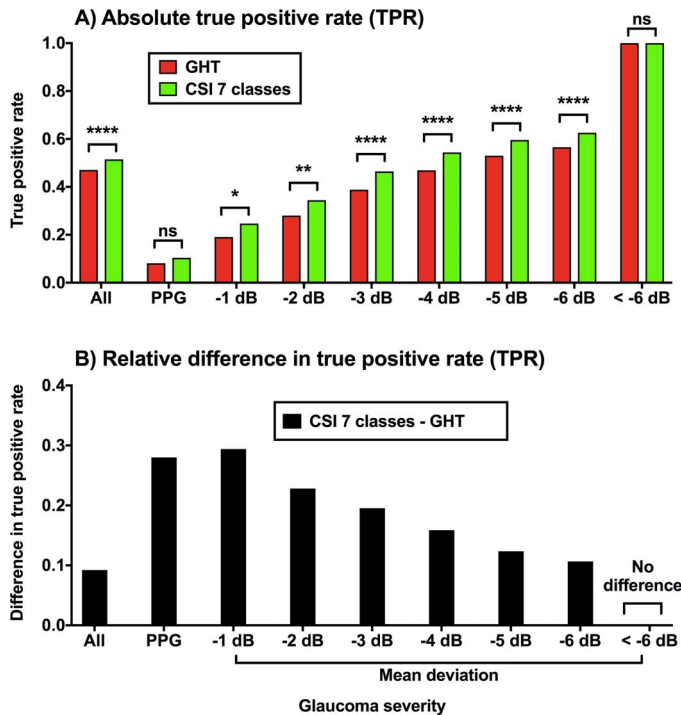
## Statistical Analysis

Diagnostic performance was compared across all clustering patterns. Because the diagnoses (WNL or ONL) were not necessarily independent as they came from the same eye, and also because each individual subject could have contributed more than one VF result, McNemar’s test was used to compare diagnostic performance of the three clustering patterns (i.e., the number of glaucoma cases detected as ONL; true-positive).<sup>49,50</sup> We were also primarily interested in the number of extra cases detected by each clustering method, rather than the cases mutually detected or missed. A  $P < 0.05$  was considered statistically significant for pairwise comparisons, but

this was adjusted for conditions where multiple comparisons were performed.

Although conventionally true-positives would be referred to as the sensitivity of the diagnostic test, we use the term TPR instead, as sensitivity in the current paper may be confused with the ability of the eye to detect a light increment (i.e., contrast sensitivity). Similarly, we also determined and compared true-negative rate (TNR) when using the test group of healthy subjects (for consistency, we do not use the term specificity). Correspondingly, this test does not mean that the specificity for the purpose of diagnostic utility was set at different levels for each test, as each test still used a 1% lower distribution limit as the cut-off for being outside normal limits (see Fig. 2). Instead, this was a test to examine whether or not the TNR is within the expected specificity set by the 1% distribution limit. Similarly, we did not assess area under the receiver operating curve because, as per the method of Asman and Heijl,<sup>9</sup> the distribution cut-offs





**Figure 3.** (A) TPR for GHT (red) and the CSI-derived seven-class theme map (green) as a function of different levels of glaucoma severity. For clarity, the upper decibel limit of each severity bin is noted (e.g.,  $-1$  dB indicates MD values “up to  $-1$  dB” and so on). The  $n$  VF results for each group were: all, 1019; PPG, 299;  $-1$  dB, 142;  $-2$  dB, 280;  $-3$  dB, 397;  $-4$  dB, 486;  $-5$  dB, 563;  $-6$  dB, 612; worse than  $-6$  dB, 108. (B) Relative increase in TPR when comparing the CSI-derived seven-class theme map with GHT (black). A positive difference indicates that the CSI theme map had a higher TPR. PPG patients were analyzed separately; the different levels of mean deviation represent patients with statistically significant VF defects, as per the Methods. The asterisks indicate significant differences between groups using McNemar’s test ( $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$ ;  $****P < 0.0001$ ).

indicating ONL on the hemifield analysis has been fixed at  $P < 0.01$ .

## Results

### Hypothesis 1: Comparison of GHT Clusters With the CSI-Derived 7 Class Theme Map

The CSI-derived seven-class theme map had a greater TPR compared with the GHT clusters (Fig. 3). The absolute increase in detection rate was small (Fig. 3A). McNemar’s test showed significant differences for all MD bins from up to  $-1$  dB to up to  $-6$  dB between the CSI map and the GHT. For PPG patients, there was no significant difference between the CSI map and the GHT ( $P = 0.0961$ ). For VFs with

a MD of worse than  $-6$  dB, both maps flagged 100% of VFs as ONL.

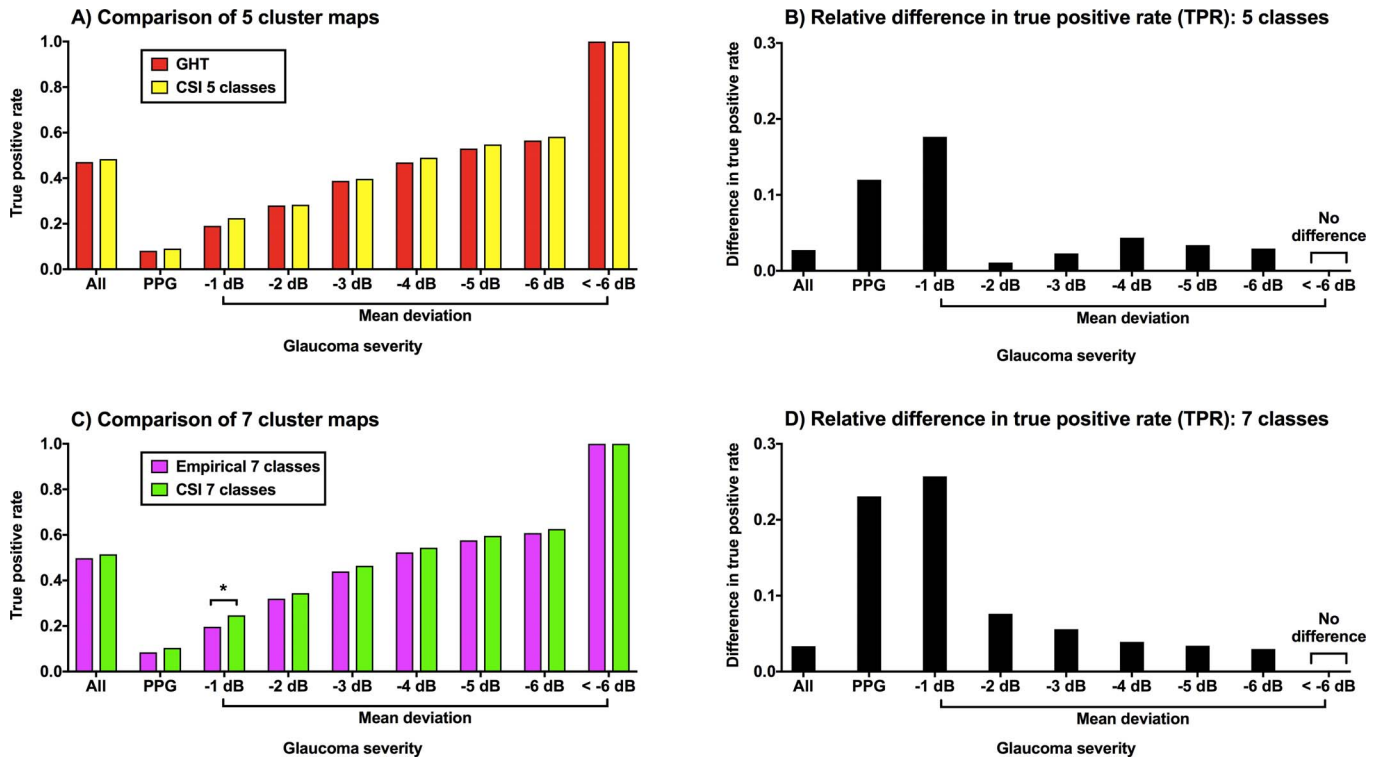
The relative change in TPR was calculated by dividing the difference in absolute number of cases identified by each technique by the number of cases identified by the GHT, respectively (Fig. 3B). Thus, although the absolute difference in TPR was similar, the relative change in TPR changed with different levels of MD, with lower MD values showing a greater relative increase in TPR when using the CSI-derived theme map.

### Hypothesis 2: The Effect of Number of Clusters on Defect Detection

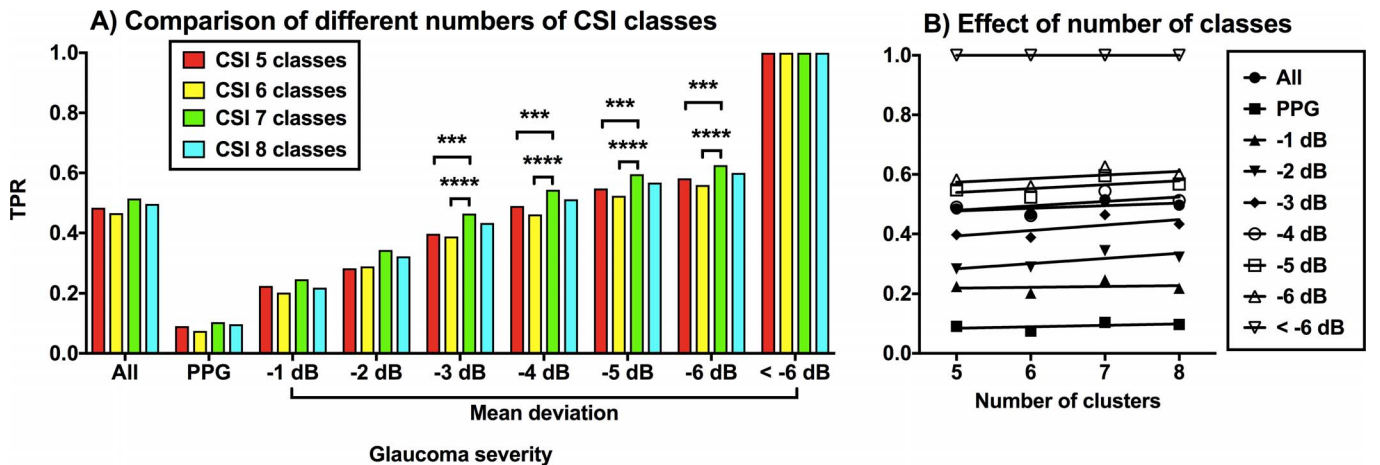
It was possible that the greater TPR using the seven-class theme map was due to the greater number of available clusters (7 vs. 5 in the GHT). We tested this hypothesis in three ways. First, we compared the GHT result with a CSI-derived five-class theme map. Differences here would indicate an effect of the assignment of test points to different classes, rather than the number of classes used. Second, we compared a seven-cluster symmetrical map further refined from the existing GHT map. This empirical seven-class map (Fig. 1F) would serve as a comparison for the CSI-derived seven-class theme map, wherein differences would be attributable to the test location cluster assignment, rather than number of classes. Third, we examined the performance of five-, six-, seven-, and eight-theme class maps derived using CSIs for a cluster number-dependent effect.

Both CSI-derived five- and seven-class theme maps had similar TPRs compared with the GHT and the empirical seven-class maps, respectively (Fig. 4). The only significant increase in TPR was when the CSI-derived map was used compared with the empirical seven-class map for the  $-1$  dB bin ( $P = 0.0265$ , Fig. 4D). Note that this difference would disappear if accounting for multiple comparisons. For all other conditions, the improvement was not statistically significant for the five (average  $P = 0.5084$ , Fig. 4A) and seven cluster conditions (average  $P = 0.2779$ , Fig. 4B). Nonetheless, there was still a high relative difference in TPR, similar to the results in Figure 3, for both five-cluster map (Fig. 4B) and seven-cluster map comparisons (Fig. 4D).

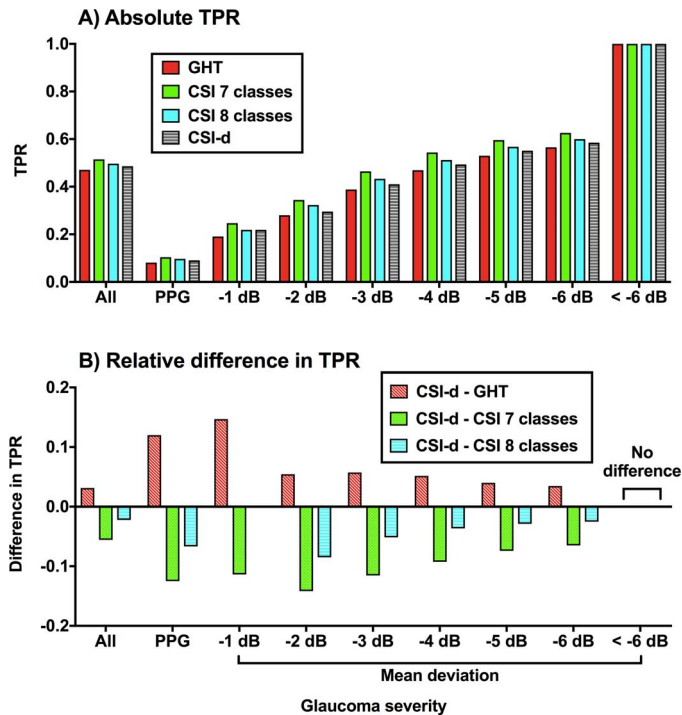
The TPRs for different numbers of CSI-derived clusters were compared. Interestingly, the seven-class theme map appeared to identify the greatest number of glaucomatous VFs, followed by the five-class theme map, the eight-class theme map, and finally



**Figure 4.** The TPR as a function of glaucoma severity, as per Figure 3 when comparing five-cluster maps ([A] GHT, red, and CSI-derived 5 class theme map, yellow) and seven-cluster maps ([C] empirical seven classes, purple, and CSI-derived seven-class theme map, green). The asterisk indicates a statistically significant difference ( $*P < 0.05$ ). Relative difference in TPR, as per Figure 3, are shown for five- and seven-cluster maps on the right hand side (B and D, respectively).



**Figure 5.** (A) Comparison of TPR found using different number of CSI classes (5, 6, 7, or 8) as a function of glaucoma severity, as per Figures 3 and 4. For clarity, only the statistical comparisons are shown between the seven-class theme map compared with all other numbers of classes. The asterisks indicate significant differences between groups using McNemar's test ( $P < 0.0083$  was considered significant to adjust for multiple comparisons;  $***P < 0.001$ ;  $****P < 0.0001$ ). (B) The TPR as a function of number of clusters used from the CSI-derived theme maps for different glaucoma severity conditions. Linear regression was performed on these data and all slopes were not significantly different to 0.



**Figure 6.** (A) Comparison of TPR found using different clustering methods as a function of glaucoma severity, as per Figure 3. For clarity, only the statistical comparisons are shown between the seven-class theme map and the CSI-d map. After adjusting for multiple comparisons ( $P < 0.0083$ ), there was no significant difference between conditions. (B) Relative increase in TPR when comparing the CSI-d map with GHT (red), CSI seven classes (green), and CSI eight classes (blue) as a function of disease severity, as per Figure 3. A positive difference indicates greater TPR by CSI-d, while a negative difference indicates a lower TPR by CSI-d.

the six-class theme map (Fig. 5A). There was no difference in TPR between the classes for PPG patient and for VF MD values up to  $-2$  dB. When the TPR was plotted as a function of number of classes within each theme map, it confirmed no effect of number of classes on detection rate (Fig. 5B, average  $P = 0.5192$  for the slope of the linear regression). Thus, in combination with the results in Figure 4, it appeared that the subtle improvements in TPR found using the seven-theme class map was more likely driven primarily by class assignment, rather than by the number of clusters used.

### Hypothesis 3: Supplementing the CSI-Derived Theme Map With Data From Glaucoma Patients

We then tested the hypothesis that adding data on the prevalence of VF defects in glaucoma would improve the diagnostic ability (Fig. 1G, CSI-d). The

CSI-d cluster map showed higher TPR compared with the GHT, but this improvement did not reach statistical significance (average  $P = 0.2997$ ; Fig. 6). The addition of glaucoma patient data showed lower TPR compared with the CSI-derived seven-class theme map for all conditions except for PPG patients ( $P = 0.5708$ ) and results with MD up to  $-1$  dB ( $P = 0.4237$ ). It also showed lower TPR compared with the CSI-derived eight-class theme map, but this did not reach statistical significance (average  $P = 0.2998$ ).

### False-Positive Rate Across all Clustering Patterns

The healthy test cohort consisting of 354 VF results of 145 subjects was used to assess the false-positive rate. These were similar across most maps as follows: GHT, 1.1%; CSI five class, 0.8%; CSI six class, 0.8%; CSI seven class, 1.1%; CSI eight class, 0.6%; empirical seven class, 0.8%; and CSI-d, 0.6%. Notably, these were all similar to the level expected by the use of the  $P < 0.01$  level cut-off for the ONL criterion.

## Discussion

Differences in superior and inferior VF loss (i.e., hemifield asymmetries) are a common feature in glaucomatous disease.<sup>51–54</sup> In the present study, we tested the hypotheses that a functional,<sup>18</sup> rather than structurally inspired basis,<sup>9,11</sup> of determining clusters improves detection rates of VF hemifield anomalies. We did this by systematically testing three hypotheses.

### Hypotheses 1 and 2: Functionally Versus Structurally Inspired Clustering Methods

The initial results of study suggested that improvements in detection rate over current structurally inspired clustering methods are achievable through the use of an optimal CSI-derived map. The benefits of the CSI-derived map were greatest in early glaucoma, with an approximately 30% increase in relative detection rate. Specifically, there were increases in detection rate for patients with PPG, in which all other indices had lower rates of case detection (MD: 0%; PSD 8.4% [23.7% increase if using the CSI-derived map]; ‘event’ analysis, i.e.,  $>3$  contiguous points of significant depression: 0%). However, these benefits appeared to be related to the number of clusters used, as arbitrary organization of test locations into clusters achieved only slightly lower TPR compared with the functionally derived

maps. As expected, with advancing disease, these benefits become less appreciable. Indeed, with cumulative defects of  $-3$  dB or greater, another commonly used global index, the PSD, has a detection rate comparable with the hemifield test. Beyond  $-6$  dB, all techniques appear equally useful. We therefore suggest that the greatest applicability of our approach would be in early disease in which the functional defects may be subtler.

The small improvement in TPR using a functionally inspired theme map may be because the use of mirrored zones with equal numbers of test locations may inadvertently include locations with higher sensitivity or less likelihood of damage. This is clearly reflected in the relatively small  $P$  values in the comparisons in TPR performance across clustering methods, which may actually work against a potential inflation of Type I error when using McNemar's test for our statistical method. Specifically, the differences in sensitivity signature may effectively mask early and subtle defects because of their spurious contributions to the  $P$  score distribution limits. The advantage of the CSI seven-class theme map is the inclusion of points only relevant for that CSI. For example, the yellow zone within the CSI seven-class theme map (Fig. 1B) includes only three locations within the superior hemifield. Forcing symmetry with its six-location inferior counterpart may reduce the chance of identifying asymmetries across the midline.

The small absolute increase in TPR may be explained by the many mutual zones among all clustering methods. Even if there were subtle differences in the normative distributions as a result of the sensitivity grouping method applied to CSI maps, these would likely be small overall. Figure 1G shows clear regions where glaucomatous VF defects were more prevalent within the current glaucoma cohort, and the mutuality of the zones is illustrated clearly in Supplementary Figure S2. These patterns are similar to those reported by Schiefer et al.<sup>10</sup>

Consequently, the limited number of test locations within the 24-2 test grid (further limited by the points chosen by the GHT) means that only a finite number of plausible combinations of points used to delineate zones for asymmetry analysis are available. It would arguably not be meaningful to analyze pairs of individual points, as it would be reliant upon the accuracy of a single measurement. For this reason, the CSIs used in the present study were limited to an eight-class map, as maps with more classes would result in pairs of individual points for comparison.

As these clustering methods are limited by the

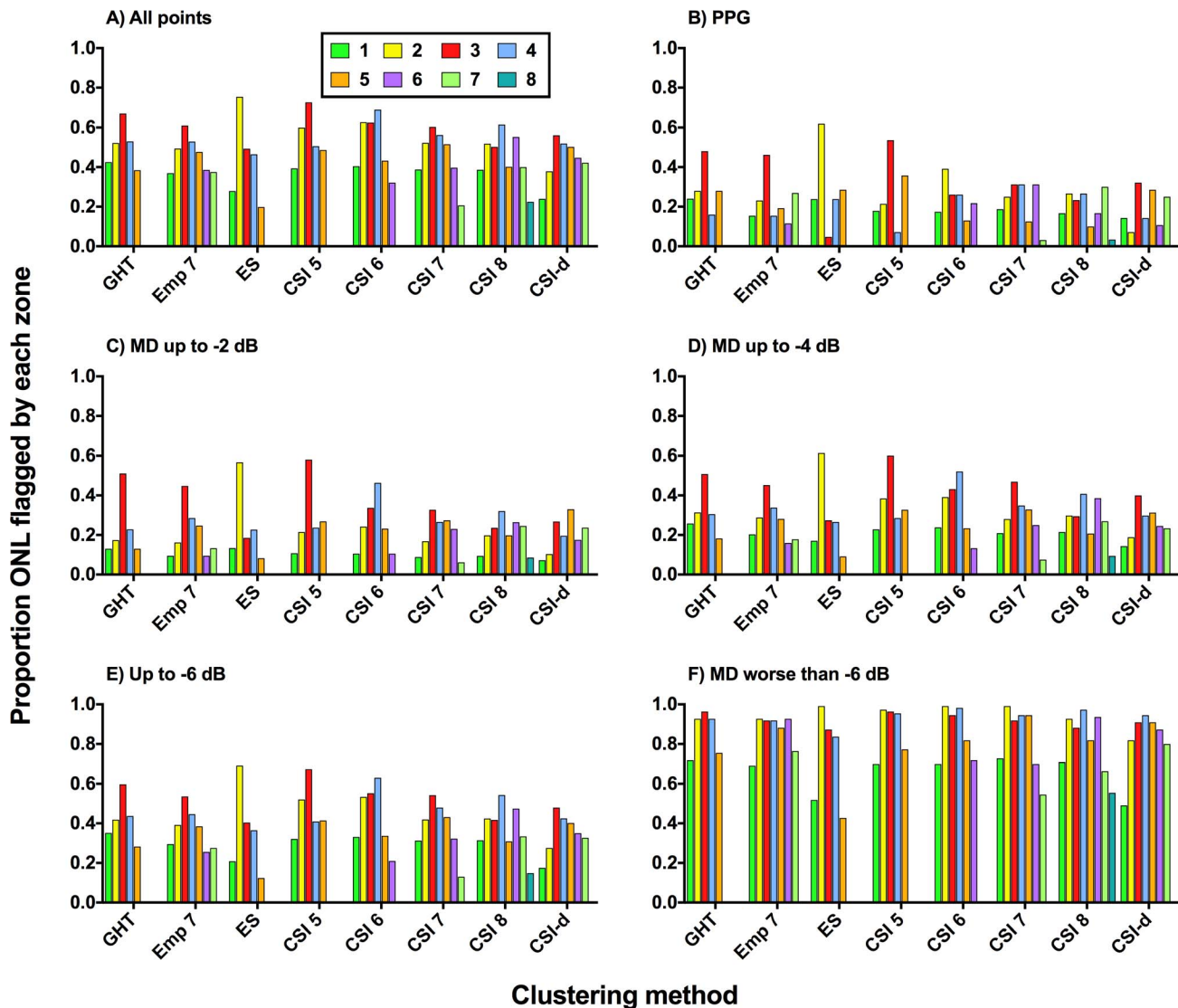
fixed points within the 24-2 test grid, further improvements to defect detection would be difficult without altering the procedure itself.<sup>55,56</sup> Recent studies have suggested that variation in fovea-to-disc angle and tilt of the horizontal raphe may account for differences in sensitivity measurements, particularly at the nasal step region, which would considerably affect any asymmetry analysis.<sup>57-59</sup> Tailoring the 24-2 test grid to test regions of interest, for example, guided by structural analyses, may provide further gains in detection not limited by the fixed grid.<sup>60</sup> Such an approach may overcome the limitations of the fixed grid of test points, as we have demonstrated in the above results, that have been commonly used to generate the RNFL pathways from the optic nerve head.<sup>61,62</sup> Altering stimulus parameters to further optimize the ability of SAP would also likely provide further improvements in diagnostic ability.<sup>8,32,63,64</sup>

### Adding Glaucomatous VF Data Did Not Improve Diagnostic Ability

We hypothesized that the use of a theme map incorporating data regarding the location of asymmetries in glaucomatous VF defects would further enhance the CSI-derived map. Specifically, the scotoma may spread across the border of the CSI, which may then lead to failure of the CSIs derived solely from healthy subjects to detect abnormalities. However, we found that resultant theme map, CSI-d, following the addition of defect information was still inferior to the optimal CSI seven-class theme map. We suggest that the reason why this data detracted from the TPR in comparison to the optimal CSI theme map is because of the heterogeneity of the location of glaucomatous VF defects.<sup>10,65,66</sup> Although there are common locations for the appearance of glaucomatous defects, no single location was sufficient for flagging all cases of glaucoma, even in cases of advanced disease. Figure 7 also highlights this, even among VF results with the same MD value, there was significant variability in terms of the location of statistically significant deficits. Thus, emphasizing specific locations for analysis would result in defects at other locations being masked.

### Limitations

The present study used cross-sectional data and only binary classification (ONL or WNL). Cross-sectional data means that the repeatability of GHT cannot be assessed. One potential confounding factor was the use of multiple VF results from some of the



**Figure 7.** The proportion of ONL flagged by each zone for each clustering method. This was calculated by dividing the number of times each zone was flagged as ONL by the total ONL flagged by the entire clustering method. Because there were instances where multiple zones were flagged ONL for each case, the total proportions for each clustering method may add up to more than 1. Each zone is identified by its individual color, as per the maps shown in Figure 2. Different stages of glaucoma severity are shown in separate panels (A–F), as per Figure 3.

glaucoma patients. We addressed this using two further subanalyses. First, we assessed the intraclass correlation among glaucoma patients who had contributed more than one VF result. We used the result from the seven-class theme map as it had the highest TPR overall. Mean ICC across the 202 patients contributing more than one result (binary variance within the patient divided by the sum of the binary variance within patient and within the total cohort)<sup>67</sup> was low at 0.22 (95% confidence interval 0.19–0.26).<sup>68</sup> This suggests little apparent bias induced by multiple contributing VF results from the

same patient on the basis of defect frequency. The second subanalysis was to randomly select one VF result for each glaucoma patient for analysis for each of the conditions seen in Figures 3 to 6. Here, the TPR results were almost identical to when multiple VF results were used from individual patients, hence we concluded that the TPR was unlikely to be affected by multiple results (see Supplementary Figs. S3–S6). For these reasons, we did not pursue further analysis using multilevel techniques.

Stability of diagnosis (a stable diagnosis means that it is repeatable for future tests) using a particular

technique is examined to validate a disease diagnosis over time. Although we had a subset of glaucoma patients who had multiple VF results that could be used to assess stability, this was complicated by the differences in time between visits, potential disease progression (e.g., from preperimetric to perimetric glaucoma) and the effect of treatment. The overall stability across each clustering method across all patients was similar at approximately 60%, which was notably lower than the level of agreement found by others yet similar to the degree to which patients exhibit reversal of a GHT result.<sup>69,70</sup> A longitudinal study would be required to carefully test this.

We were required to develop our own normative data for analysis, as the CSI-derived maps required pooling of sensitivities to generate  $P$  score distribution limits. There were slight differences in the TPR compared with the instrument printout attributable to this reason, in addition to the limitation of binary outcomes only. Further study would be required to determine the significance of these differences. As the same normative data was used across all clustering methods, the improvement provided by CSI-derived maps was unlikely due to the underlying cohort (supported by our recent study comparing the healthy cohort with others reported in the literature<sup>21</sup>).

Finally, we pooled all healthy subject VF data together, averaging as per our previous methods for establishing CSIs, and also further compared this with an older structurally inspired map (GHT). Notably, recent studies<sup>71–73</sup> have identified the role of individual ocular anatomy on the structure-function relationship, including factors such as refraction/axial length, optic disc size and optic disc position. Averaging the VF data as a preprocessing step has both advantages and disadvantages. The advantage is that potential individualistic contributions, such as optic disc tilt, refraction and axial length, are less apparent when averaged and collapsed into five layers. For example, a patient with a tilted disc with obliquely oriented RNFL bundles is represented by its own input layer, the relative weight of this layer, equal to an average, regular bundle, may introduce a conflation to the resultant theme map. The disadvantage is that some individual information may be lost, with implications for comparisons made to patients with similar atypical configurations. It is not yet known how these may then be correlated with clustered bundles for a hemifield analysis. One cluster map that has been suggested to represent the path of the RNFL bundles is available on the Octopus perimeter (EyeSuite software). This map shares

features with both the GHT (structurally inspired, guided by RNFL bundles) and the CSI maps (asymmetric about the midline). However, we found that this map had a lower TPR compared with both the GHT (though not statistically significant) and the CSI-derived maps ( $P < 0.0001$  across all conditions except  $-1$  dB, which was  $P < 0.05$ ). This area deserves further investigation and optimization.

## Conclusions

Previous studies have used clustering analysis to identify regions and patterns of glaucomatous VF loss. To add to this field of study, we have provided a hemifield asymmetry analysis theme map derived using 24-2 VF sensitivity data from healthy subjects that is not limited to mirrored zone or fixed numbers of constituent test locations, which detects more cases of glaucoma compared with the GHT. This approach makes no assumptions regarding the type of glaucomatous defect that may occur in the individual patient, but aims to identify deviations in sensitivity from the underlying normal distribution. The CSI-derived map could prove a useful addition, but not replacement, to existing VF indices, as the implementation of such a map would be simple, requiring no modification to the physical hardware of the perimeter. This work may provide a basis for determining and optimizing further asymmetry analysis for disease detection. Identification of the CSIs may assist in guiding and tailoring the test grid to further optimize the test locations to examine for asymmetry analysis, (i.e., not restricted to the points delineated by the test grid). In current clinical practice, applications of this work include guiding methods of analysis for other test patterns (e.g., 30-2), which may not already possess a hemifield analysis (e.g., the 10-2 grid), for structural data, such as optical coherence tomography, and also for other diseases exhibiting asymmetries within the VF, such as neurological deficits.

## Acknowledgments

The authors thank Janelle Tong and Henrietta Wang for technical assistance.

Supported by grants from a PhD scholarship provided by Guide Dogs NSW/ACT and an Australian Government Research Training Program PhD Scholarship (JP). This work was supported by the National Health and Medical Research Council of Australia (NHMRC #1033224). Guide Dogs NSW/

ACT are partners in the NHMRC grant. This work was also supported by an Australian Research Council Future Fellowship (FT130100338; BVB).

M. Kalloniatis and S.K. Khuu are named inventors on a patent involving the use of different Goldmann target sizes at different visual field locations for contrast sensitivity testing (International Publication Number WO 2014/094035 A1 (USA) and European Patent Number: 13865419.9).

Disclosure: **J. Phu**, None; **S.K. Khuu**, P; **B.V. Bui**, None; **M. Kalloniatis**, P

## References

1. Phu J, Khuu SK, Yapp M, Assaad N, Hennessy MP, Kalloniatis M. The value of visual field testing in the era of advanced imaging: clinical and psychophysical perspectives. *Clin Exp Optom*. 2017;100:313–332.
2. Jampel HD, Singh K, Lin SC, et al. Assessment of visual function in glaucoma: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2011;118:986–1002.
3. Garway-Heath DF, Lascaratos G, Bunce C, Crabb DP, Russell RA, Shah A; for the United Kingdom Glaucoma Treatment Study Investigators. The United Kingdom Glaucoma Treatment Study: a multicenter, randomized, placebo-controlled clinical trial: design and methodology. *Ophthalmology*. 2013;120:68–76.
4. Jeong JH, Park KH, Jeoung JW, Kim DM. Preperimetric normal tension glaucoma study: long-term clinical course and effect of therapeutic lowering of intraocular pressure. *Acta Ophthalmol*. 2014;92:e185–e193.
5. Kim KE, Jeoung JW, Kim DM, Ahn SJ, Park KH, Kim SH. Long-term follow-up in preperimetric open-angle glaucoma: progression rates and associated factors. *Am J Ophthalmol*. 2015; 159:160–168.e1–2.
6. Leske MC, Heijl A, Hussein M, Bengtsson B, Hyman L, Komaroff E; for the Early Manifest Glaucoma Trial Group. Factors for glaucoma progression and the effect of treatment: the early manifest glaucoma trial. *Arch Ophthalmol*. 2003; 121:48–56.
7. Mills RP, Budenz DL, Lee PP, et al. Categorizing the stage of glaucoma from pre-diagnosis to end-stage disease. *Am J Ophthalmol*. 2006;141:24–30.
8. Phu J, Khuu SK, Zangerl B, Kalloniatis M. A comparison of Goldmann III, V and spatially equated test stimuli in visual field testing: the importance of complete and partial spatial summation. *Ophthalmic Physiol Opt*. 2017;37: 160–176.
9. Asman P, Heijl A. Glaucoma hemifield test. Automated visual field evaluation. *Arch Ophthalmol*. 1992;110:812–819.
10. Schiefer U, Papageorgiou E, Sample PA, et al. Spatial pattern of glaucomatous visual field loss obtained with regionally condensed stimulus arrangements. *Invest Ophthalmol Vis Sci*. 2010; 51:5685–5689.
11. Gardiner SK, Mansberger SL, Demirel S. Detection of functional change using cluster trend analysis in glaucoma. *Invest Ophthalmol Vis Sci*. 2017;58: BIO180–BIO90.
12. Mandava S, Zulauf M, Zeyen T, Caprioli J. An evaluation of clusters in the glaucomatous visual field. *Am J Ophthalmol*. 1993;116:684–691.
13. Suzuki Y, Araie M, Ohashi Y. Sectorization of the central 30 degrees visual field in glaucoma. *Ophthalmology*. 1993;100:69–75.
14. Henson DB, Spenceley SE, Bull DR. Spatial classification of glaucomatous visual field loss. *Br J Ophthalmol*. 1996;80:526–531.
15. Yousefi S, Balasubramanian M, Goldbaum MH, et al. Unsupervised Gaussian mixture-model with expectation maximization for detecting glaucomatous progression in standard automated perimetry visual fields. *Transl Vis Sci Technol*. 2016;5(3):2.
16. Yousefi S, Goldbaum MH, Balasubramanian M, et al. Learning from data: recognizing glaucomatous defect patterns and detecting progression from visual field measurements. *IEEE Trans Biomed Eng*. 2014;61:2112–2124.
17. Yousefi S, Kiwaki T, Zheng Y, et al. Detection of longitudinal visual field progression in glaucoma using machine learning [published online ahead of print Jun 8, 2018]. *Am J Ophthalmol*. doi: 10.1016/j.ajo.2018.06.007.
18. Phu J, Khuu SK, Nivison-Smith L, et al. Pattern recognition analysis reveals unique contrast sensitivity isocontours using static perimetry thresholds across the visual field. *Invest Ophthalmol Vis Sci*. 2017;58:4863–4876.
19. Phu J, Al-Saleem N, Kalloniatis M, Khuu SK. Physiologic statokinetic dissociation is eliminated by equating static and kinetic perimetry testing procedures. *J Vis*. 2016;16(14):5.
20. Cai S, Elze T, Bex PJ, Wiggs JL, Pasquale LR, Shen LQ. Clinical correlates of computationally

- derived visual field defect archetypes in patients from a glaucoma clinic. *Curr Eye Res.* 2017;42:568–574.
21. Phu J, Bui BV, Kalloniatis M, Khuu SK. How many subjects are needed for a visual field normative database? A comparison of ground truth and bootstrapped statistics. *Transl Vis Sci Technol.* 2018;7(2):1.
  22. Bengtsson B, Heijl A. Inter-subject variability and normal limits of the SITA Standard, SITA Fast, and the Humphrey Full Threshold computerized perimetry strategies, SITA STATPAC. *Acta Ophthalmol Scand.* 1999;77:125–129.
  23. Alniemi ST, Pang NK, Woog JJ, Bradley EA. Comparison of automated and manual perimetry in patients with blepharoptosis. *Ophthalm Plast Reconstr Surg.* 2013;29:361–363.
  24. Zalta AH. Lens rim artifact in automated threshold perimetry. *Ophthalmology.* 1989;96:1302–1311.
  25. Phu J, Kalloniatis M, Khuu SK. Reducing spatial uncertainty through attentional cueing improves contrast sensitivity in regions of the visual field with glaucomatous defects. *Transl Vis Sci Technol.* 2018;7(2):8.
  26. Phu J, Kalloniatis M, Khuu SK. The effect of attentional cueing and spatial uncertainty in visual field testing. *PLoS One.* 2016;11:e0150922.
  27. Wall M, Woodward KR, Brito CF. The effect of attention on conventional automated perimetry and luminance size threshold perimetry. *Invest Ophthalmol Vis Sci.* 2004;45:342–350.
  28. Anderson AJ, Johnson CA. Anatomy of a supergroup: does a criterion of normal perimetric performance generate a supernormal population? *Invest Ophthalmol Vis Sci.* 2003;44:5043–5048.
  29. Jamous KF, Kalloniatis M, Hennessy MP, Agar A, Hayen A, Zangerl B. Clinical model assisting with the collaborative care of glaucoma patients and suspects. *Clin Experiment Ophthalmol.* 2015;43:308–319.
  30. Huang J, Hennessy MP, Kalloniatis M, Zangerl B. Implementing collaborative care for glaucoma patients and suspects in Australia [published online ahead of print March 2, 2018]. *Clin Exp Ophthalmol.* doi: 10.1111/ceo.13187.
  31. Yoshioka N, Zangerl B, Phu J, et al. Consistency of structure-function correlation between spatially scaled visual field stimuli and in vivo OCT ganglion cell counts. *Invest Ophthalmol Vis Sci.* 2018;59:1693–1703.
  32. Kalloniatis M, Khuu SK. Equating spatial summation in visual field testing reveals greater loss in optic nerve disease. *Ophthalmic Physiol Opt.* 2016;36:439–452.
  33. Johnson CA, Keltner JL, Cello KE; for the Ocular Hypertension Study Group. Baseline visual field characteristics in the ocular hypertension treatment study. *Ophthalmology.* 2002;109:432–437.
  34. Ishiyama Y, Murata H, Asaoka R. The usefulness of gaze tracking as an index of visual field reliability in glaucoma patients. *Invest Ophthalmol Vis Sci.* 2015;56:6233–6236.
  35. Yoshioka N, Zangerl B, Nivison-Smith L, et al. Pattern recognition analysis of age-related retinal ganglion cell signatures in the human eye. *Invest Ophthalmol Vis Sci.* 2017;58:3086–3099.
  36. Bengtsson B, Heijl A, Olsson J. Evaluation of a new threshold visual field strategy, SITA, in normal subjects. Swedish Interactive Thresholding Algorithm. *Acta Ophthalmol Scand.* 1998;76:165–169.
  37. Chua J, Nivison-Smith L, Tan SS, Kalloniatis M. Metabolic profiling of the mouse retina using amino acid signatures: insight into developmental cell dispersion patterns. *Exp Neurol.* 2013;250:74–93.
  38. Kalloniatis M, Marc RE, Murry RF. Amino acid signatures in the primate retina. *J Neurosci.* 1996;16:6807–6829.
  39. Marc RE, Murry RF, Basinger SF. Pattern recognition of amino acid signatures in retinal neurons. *J Neurosci.* 1995;15:5106–5129.
  40. Swain PH, King RC. Two effective feature selection criteria for multispectral remote sensing. Paper presented at: 1<sup>st</sup> International Joint conference on Pattern Recognition; October 30 to November 1, 1973; Washington, D.C.
  41. Nordmann JP, Mesbah M, Berdeaux G. Scoring of visual field measured through Humphrey perimetry: principal component varimax rotation followed by validated cluster analysis. *Invest Ophthalmol Vis Sci.* 2005;46:3169–176.
  42. Vermunt JK. K-means may perform as well as mixture model clustering but may also be much worse: comment on Steinley and Brusco (2011). *Psychol Methods.* 2011;16:82–88; discussion 9–92.
  43. Steinley D, Brusco MJ. Evaluating mixture modeling for clustering: recommendations and cautions. *Psychol Methods.* 2011;16:63–79.
  44. Lachenmayr BJ, Kiermeir U, Kojetinsky S. Points of a normal visual field are not statistically independent. *Ger J Ophthalmol.* 1995;4:175–181.
  45. Heijl A, Lindgren G, Olsson J. Normal variability of static perimetric threshold values across the



- central visual field. *Arch Ophthalmol*. 1987;105:1544–1549.
46. Khuu SK, Kalloniatis M. Standard automated perimetry: determining spatial summation and its effect on contrast sensitivity across the visual field. *Invest Ophthalmol Vis Sci*. 2015;56:3565–3576.
  47. Gardiner SK, Demirel S, Goren D, Mansberger SL, Swanson WH. The effect of stimulus size on the reliable stimulus range of perimetry. *Transl Vis Sci Technol*. 2015;4(2):10.
  48. Swanson WH, Horner DG, Dul MW, Malinovsky VE. Choice of stimulus range and size can reduce test-retest variability in glaucomatous visual field defects. *Transl Vis Sci Technol*. 2014;3(5):6.
  49. Agresti A, Effects Min Y. and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. *Stat Med*. 2004;23:65–75.
  50. Fagerland MW, Lydersen S, Laake P. Recommended tests and confidence intervals for paired binomial proportions. *Stat Med*. 2014;33:2850–875.
  51. Harizman N, Oliveira C, Chiang A, et al. The ISNT rule and differentiation of normal from glaucomatous eyes. *Arch Ophthalmol*. 2006;124:1579–1583.
  52. Jonas JB, Budde WM. Diagnosis and pathogenesis of glaucomatous optic neuropathy: morphological aspects. *Prog Retin Eye Res*. 2000;19:1–40.
  53. Jonas JB, Fernandez MC, Sturmer J. Pattern of glaucomatous neuroretinal rim loss. *Ophthalmology*. 1993;100:63–68.
  54. Takagi ST, Kita Y, Yagi F, Tomita G. Macular retinal ganglion cell complex damage in the apparently normal visual field of glaucomatous eyes with hemifield defects. *J Glaucoma*. 2012;21:318–325.
  55. Ehrlich AC, Raza AS, Ritch R, Hood DC. Modifying the conventional visual field test pattern to improve the detection of early glaucomatous defects in the central 10 degrees. *Transl Vis Sci Technol*. 2014;3(6):6.
  56. Hood DC, Nguyen M, Ehrlich AC, et al. A test of a model of glaucomatous damage of the macula with high-density perimetry: implications for the locations of visual field test points. *Transl Vis Sci Technol*. 2014;3(3):5.
  57. Bedggood P, Nguyen B, Lakkis G, Turpin A, McKendrick AM. Orientation of the temporal nerve fiber raphe in healthy and in glaucomatous eyes. *Invest Ophthalmol Vis Sci*. 2017;58:4211–4217.
  58. McKendrick AM, Denniss J, Wang YX, Jonas JB, Turpin A. The proportion of individuals likely to benefit from customized optic nerve head structure-function mapping. *Ophthalmology*. 2017;124:554–561.
  59. Tanabe F, Matsumoto C, McKendrick AM, Okuyama S, Hashimoto S, Shimomura Y. The interpretation of results of 10-2 visual fields should consider individual variability in the position of the optic disc and temporal raphe. *Br J Ophthalmol*. 2018;102:323–328.
  60. Ballae Ganeshrao S, Turpin A, McKendrick AM. Sampling the visual field based on individual retinal nerve fiber layer thickness profile. *Invest Ophthalmol Vis Sci*. 2018;59:1066–1074.
  61. Leite MT, Zangwill LM, Weinreb RN, Rao HL, Alencar LM, Medeiros FA. Structure-function relationships using the Cirrus spectral domain optical coherence tomograph and standard automated perimetry. *J Glaucoma*. 2012;21:49–54.
  62. Garway-Heath DF, Holder GE, Fitzke FW, Hitchings RA. Relationship between electrophysiological, psychophysical, and anatomical measurements in glaucoma. *Invest Ophthalmol Vis Sci*. 2002;43:2213–2220.
  63. Mulholland PJ, Redmond T, Garway-Heath DF, Zlatkova MB, Anderson RS. Spatiotemporal summation of perimetric stimuli in early glaucoma. *Invest Ophthalmol Vis Sci*. 2015;56:6473–6482.
  64. Redmond T, Garway-Heath DF, Zlatkova MB, Anderson RS. Sensitivity loss in early glaucoma can be mapped to an enlargement of the area of complete spatial summation. *Invest Ophthalmol Vis Sci*. 2010;51:6540–6548.
  65. Kim JM, Kyung H, Shim SH, Azarbod P, Caprioli J. Location of initial visual field defects in glaucoma and their modes of deterioration. *Invest Ophthalmol Vis Sci*. 2015;56:7956–7962.
  66. Cho HK, Lee J, Lee M, Kee C. Initial central scotomas vs peripheral scotomas in normal-tension glaucoma: clinical characteristics and progression rates. *Eye (Lond)*. 2014;28:303–311.
  67. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Ann Fam Med*. 2004;2:204–208.
  68. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994;6:284–290.
  69. Katz J, Quigley HA, Sommer A. Repeatability of the glaucoma hemifield test in automated perim-

- etry. *Invest Ophthalmol Vis Sci.* 1995;36:1658–1664.
70. Wang M, Pasquale LR, Shen LQ, et al. Reversal of glaucoma hemifield test results and visual field features in glaucoma. *Ophthalmology.* 2018;125:352–360.
71. Denniss J, McKendrick AM, Turpin A. An anatomically customizable computational model relating the visual field to the optic nerve head in individual eyes. *Invest Ophthalmol Vis. Sci* 2012; 53:6981–6990.
72. Lamparter J, Russell RA, Zhu H, et al. The influence of intersubject variability in ocular anatomical variables on the mapping of retinal locations to the retinal nerve fiber layer and optic nerve head. *Invest Ophthalmol Vis Sci.* 2013;54:6074–6082.
73. Jansonius NM, Schiefer J, Nevalainen J, Paetzold J, Schiefer U. A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: average course, variability, and influence of refraction, optic disc size and optic disc position. *Exp Eye Res.* 2012;105:70–78.