



# Accelerating genetic diagnostics in retinitis pigmentosa: implementation of a semi-automated bespoke cohort analysis workflow for Hong Kong Genome Project

Dingge Ying<sup>1</sup> · Jamie Sui Lam Kwok<sup>1</sup> · Annie Tsz Wai Chu<sup>1</sup> · Wei Ma<sup>1</sup> · Helen Ying Fung Tam<sup>1</sup> · Dicky Or<sup>1</sup> · Shirley Pik Ying Hue<sup>1</sup> · Hong Kong Genome Project<sup>1</sup> · Qing Li<sup>2</sup> · Christopher Kai Shun Leung<sup>2,3,4,5</sup> · Brian Hon Yin Chung<sup>1,6</sup>

Received: 19 September 2024 / Accepted: 13 March 2025  
© The Author(s) 2025

## Abstract

The study aims to enhance the efficiency of the genetic variant curation process at the Hong Kong Genome Institute by developing a Semi-Automated Bespoke Cohort Analysis Workflow (S-BCAW) for patients with, or suspected to have, retinitis pigmentosa (RP) in the Hong Kong Genome Project (HKGP), leveraging advances in next-generation sequencing (NGS). A comparative analysis involving 79 RP patients was conducted using both the conventional manual workflow and the novel S-BCAW, which integrates initial filtering and variant classification based on ACMG guidelines, followed by detailed manual review. The diagnostic yields from both methods were identical, but the bespoke workflow reduced analysis time by approximately 60% (1.5 h/sample). This efficiency increase resulted from automated application of ACMG rules and systematic aggregation of supportive data, including disease-specific information. The study reports 25 positive cases with a diagnostic yield of 32%, including three novel variants. The S-BCAW significantly improves efficiency, helping to end the diagnostic odyssey for patients in the HKGP. This approach facilitates rapid assessment of variant pathogenicity, enhancing the feasibility and timeliness of NGS technology for clinical applications, especially in urgent scenarios.

---

Dingge Ying and Jamie Sui Lam Kwok have contributed equally to this work.

---

✉ Christopher Kai Shun Leung  
cleung21@hku.hk

✉ Brian Hon Yin Chung  
bhychung@genomics.org.hk

- <sup>1</sup> Hong Kong Genome Institute, Hong Kong, China
- <sup>2</sup> Department of Ophthalmology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
- <sup>3</sup> Department of Ophthalmology, Queen Mary Hospital, Pok Fu Lam, Hong Kong, China
- <sup>4</sup> Hong Kong Eye Hospital, Kowloon City, Hong Kong, China
- <sup>5</sup> Grantham Hospital, Hong Kong, China
- <sup>6</sup> Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

## Introduction

Genetic diseases, characterized by DNA sequence alterations, present a diagnostic challenge due to their complex symptomatology and individual variability. The rapid advancement in genomic medicine and technologies in next-generation sequencing (NGS), particularly whole genome sequencing (GS) in the past two decades, has significantly enhanced the scope of genetic diagnostics, facilitating the identification of both known and novel genetic variants (Bagger et al. 2024; Brlek et al. 2024). It has offered great opportunities in the acceleration of accurate diagnosis, personalized treatment, and efficacious surveillance and disease prevention (Arteche-López et al. 2021), and set the stage of various population-wide GS projects around the world including the Hong Kong Genome Project (HKGP) in China (Investigators et al. 2019; Stark et al. 2019; Chu et al. 2022; Halldorsson et al. 2022; Wong et al. 2023).

Despite encouraging advancements, the tertiary analysis phase of NGS, which involves the interpretation of identified genetic variants, continues to face significant hurdles (Vintschger et al. 2023), and remains to be the bottleneck of

large-scale GS projects. Challenges are often multi-factorial, with some being region-specific, while many are commonly shared by multiple countries.

Common hurdles include the lack of standardized protocols for data management, which leads to inconsistencies in determining variant pathogenicity and delays in variant curation. Such challenges are particularly acute in clinical settings where rapid decision-making is essential for effective patient care (Bertier et al. 2016). To address these issues, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have developed comprehensive guidelines for the classification of sequence variants (Richards et al. 2015). These guidelines aim to standardize the interpretation process, enhancing the consistency and reliability of genetic diagnostics. However, the application of these guidelines can still be complex and time-consuming, underscoring the need for more efficient approaches in genetic variant analysis (Masson et al. 2022). Recognizing the need to enhance the efficiency of genetic variant curation, various technologies and computational tools have been developed to streamline the process, but these solutions often encounter limitations in handling complex genetic data and integrating nuanced clinical information effectively (Kopanov et al. 2019).

Currently, the HKGP recruits participants via three Partnering Centres and four other referring networks, all under the Hospital Authority in Hong Kong, which oversees the entire public hospital network. Previous publications have detailed description on HKGP's recruitment mechanism, operational and sequencing workflows & standards (Chu et al. 2022, 2023).

One crucial challenge for HKGP, which aims to recruit 20,000–30,000 participants for GS since July 2021, is related to the human resource infrastructure. In particular, training and building a substantial team of Genome Curators to perform genomic variant review – a niche profession specializing in tertiary genomic analysis—takes time. In addition, it is a pressing task for trained curators to complete the analysis of a minimum of 40,000–50,000 genomes.

As HKGP progresses, the demand in training more genome curators, coupled with the increasing interests of clinicians in referring patients with genetic diseases for genome sequencing, has motivated the Hong Kong Genome Institute team to further innovate, streamline, and accelerate its genome curation process. This includes transitioning from individual analysis to disease-specific cohort analysis and moving from a predominantly manual process to a more semi-automated workflow.

Additionally, when applying the ACMG/AMP guidelines for variant interpretation, challenges arise specifically in the context of cohort analysis. Cohorts, defined as groups of cases exhibiting shared disease-related phenotypes, frequently encounter challenges in the application of ACMG

criteria, especially in the absence of ClinGen Expert Panels and gene-specific interpretation guidelines (Rivera-Muñoz et al. 2018). Therefore, bespoke analysis becomes crucial to accurately reflect the unique genetic characteristics of a specific disease cohort. A tailored approach is necessary to ensure that the application of ACMG criteria effectively captures the intricacies of genetic variance within distinct groups, thereby improving diagnostic accuracy and facilitating more precise clinical decision-making. This not only aims to expedite the variant curation process but also significantly impacts patient outcomes by enabling faster and more accurate clinical diagnosis.

To further illustrate the complexities of genetic data management highlighted above, the present study adopts retinitis pigmentosa (RP), the most common form of inherited retinal dystrophies (IRD), as an example. Influenced by geographical disparities, it exhibits a global prevalence rate of approximately 1:4000, with reported range from 1:9000 to 1:750 (Nangia et al. 2012; Na et al. 2017). A study from Korea shows that the time from initial presentation of symptoms to confirmed diagnosis of RP often ranged from 10 to more than 20 years (Kim et al. 2020). The genetic landscape of RP is marked by considerable heterogeneity, with over 90 identified genes contributing to its varied phenotypic presentations across different modes of inheritance (Bhardwaj et al. 2022). This diversity complicates the diagnostic process and reduces the yield of traditional methods, which often miss variants in newly discovered or non-coding regions. NGS methods have been employed to assess the diagnostic yield of RP, revealing varied success rates across different studies and populations. Colombo et al. (2021) reported a diagnostic yield of 37% in sporadic cases and 55% in familial cases among 591 individuals using targeted gene panels (Colombo et al. 2021). In a study by Jin et al. (2023), 44% of 75 Chinese RP patients were successfully diagnosed using whole-exome sequencing (WES) (Jin et al. 2023). Additionally, González-del Pozo (2020) found a diagnostic yield of 26% in a family of 5 sequenced affected patients and 14 sequenced unaffected family members (termed complex RP) using GS (Pozo et al. 2020). These findings underscore the potential of NGS in elucidating the genetic basis of RP, although the yields vary by genetic complexity and methodology.

The primary objective of this study is to develop and evaluate a semi-automated workflow that integrates key aspects of ACMG guidelines to accelerate the process of variant classification and curation. We intend to utilize a cohort of patients tentatively diagnosed with retinitis pigmentosa (RP) as a case study to demonstrate the benefits of a bespoke analysis workflow over our current, predominately manual variant curation process. This tailored approach is designed to address the unique genetic characteristics of the cohort, thereby enhancing the precision and relevance of the genetic analysis. Expected benefits include improved efficiency and

consistency in variant interpretation across different samples, and scalable processes that facilitate the expansion of cohort sizes without a proportional increase in analysis time or resource allocation. Additionally, this method could pave the way for customized cohort analysis for other diseases.

## Materials and methods

### Participants

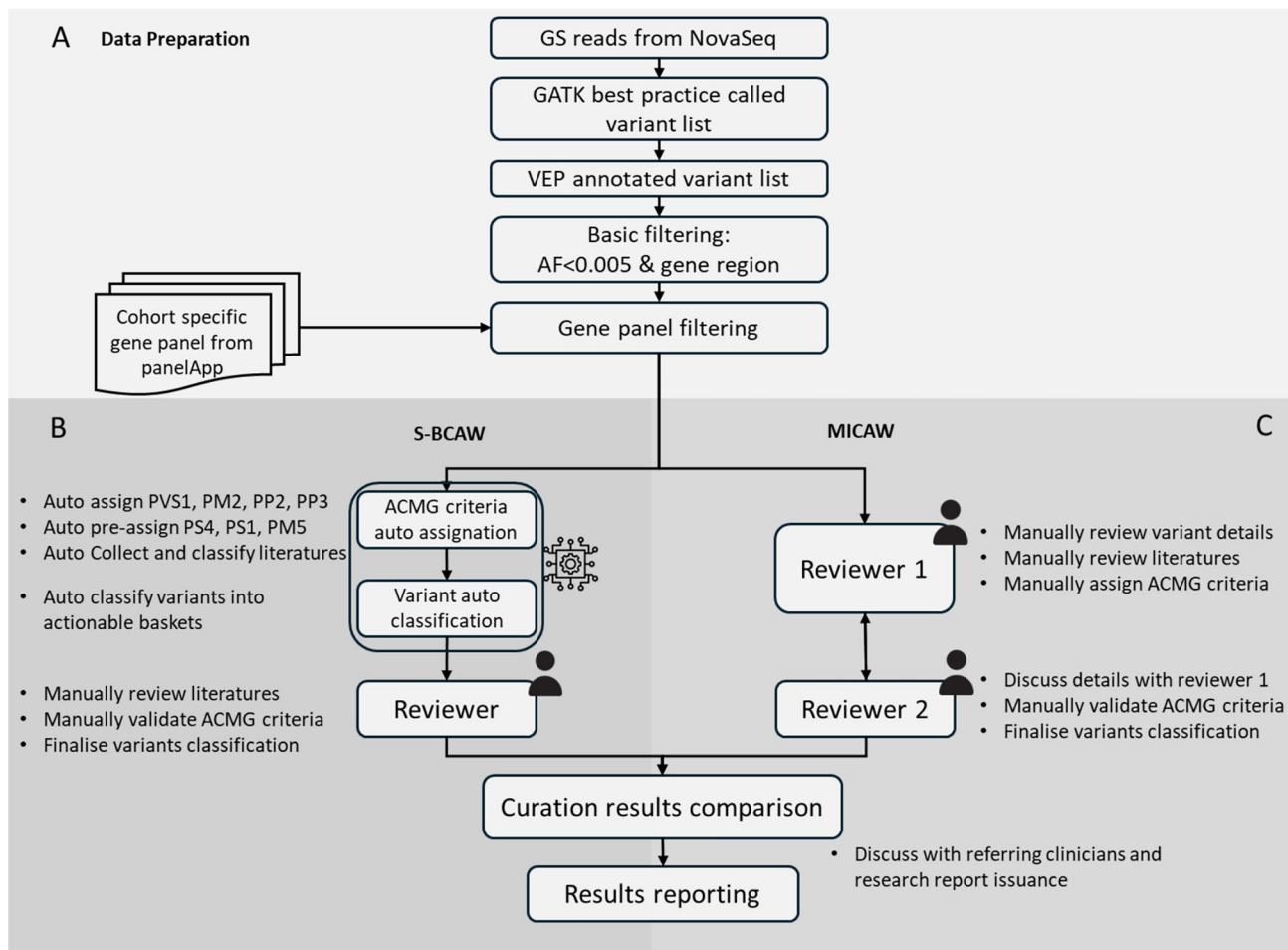
Probands with a clinical tentative diagnosis of RP together with their affected and unaffected relatives were referred from Grantham Hospital under the Hospital Authority of Hong Kong SAR, to the HKGP, with written informed consent obtained. We primarily collected whole blood samples, with buccal mucosa or saliva samples used as alternatives when whole blood collection was not feasible for the participants.

### Short read GS and data processing and filtering

Samples were prepared and sequenced by Illumina NovaSeq 6000 sequencer, followed by GATK best practice for data processing. Data was further prepared according to Fig. 1A. Selected genes were listed in Supplementary Table S1. Please refer to the Supplementary file for detailed description.

### Manual individual case analysis workflow (MICAW)

All variants in the filtered list from a single case with related family members were curated and classified by experienced reviewers in two rounds according to the standard ACMG/AMP guidelines, SVI Working Group Recommendations, along with ClinGen Variant Curation Expert Panels (VCEP) when applicable (Richards et al. 2015) (Fig. 1C). Different trained and experienced



**Fig. 1** Data processing workflow comparison flowchart of HKGP. **A** Data preparation. Filtered variant list was prepared from sequencing GS reads accordingly. **B** Details of the three steps in S-BCAW. **C**

Details of the two steps in MICAW. Curation conclusions and individual ACMG criteria assigned were compared between two workflows before reporting

genome curators, who adopted the role as “Reviewer 1” and “Reviewer 2” respectively were responsible for the tertiary analysis process.

The first round of manual variant curation by “Reviewer 1” started with assignment of variant specific ACMG criteria, namely PM2, PP2, and PP3. PM2\_supporting was only assigned to variants that are rare (gnomAD East Asian allele frequency  $\leq 0.005$ ) according to the SVI Recommendations (Chen et al. 2024). PP2 was assigned to missense variants where the gene missense constraint z-score from gnomAD database was above 3.09 and the regional missense constraint was significant ( $p < 1e-3$ ) (Chen et al. 2024). PP3 was assigned according to the computational REVEL score according to the ClinGen Recommendations for PP3/BP4 criteria (Ioannidis et al. 2016; Pejaver et al. 2022). Publication or case–control based criteria such as PS4, PS1 and PM3, were gathered from various sources including ClinVar reported cases and literature searches. PVS1 was determined by manually checking the relevant factors according to the decision tree in ClinGen PVS1 recommendation (Tayoun et al. 2018). PM4 was assigned to variants that affected protein length without altering the reading frame. Family-based criteria, PS2 and PP1, were carefully assigned by manually verifying and confirming the variant status of the parents. All criteria were applied with respect to strength modifications as recommended by ClinGen SVI working groups (Rivera-Muñoz et al. 2018).

A detailed second-round curation by “Reviewer 2” validated and confirmed the assigned ACMG criteria of candidate variants. This meticulous review process, conducted in collaboration by two reviewers, ensured precision and reliability. Additionally, this approach has proven effective in diagnosing previously undiagnosed diseases in our institute. Selected candidate variants, including compound heterozygous and homozygous pathogenic or likely pathogenic (PLP) variants in recessive genes, and single PLP variants in dominant genes, were then validated through orthogonal sequencing to confirm the diagnostic yield of the cohort.

### Semi-automated bespoke cohort analysis workflow (S-BCAW)

All filtered variants from the same cohort were merged into a single variant list, with sample ID as additional label. Selected criteria from ACMG/AMP guidelines for sequencing variant interpretation were calculated automatically, together with variant information collection for additional criteria. Variants were then classified into different actionable categories, and further reviewed by experienced reviewers to verify the details of the assigned criteria and classifications (Fig. 1B).

### S-BCAW Step 1: Automatic assignment of selected ACMG criteria

The following ACMG criteria were automatically assigned based on collected resources, or preassigned using partially collected data for each variant, to facilitate further processing.

**PVS1:** For predicted loss-of-function variants, the AutoPVS1 tool, developed in accordance with the ClinGen Sequence Variant Interpretation (SVI) Working Group guidelines for the interpretation of the loss-of-function (LoF) PVS1 ACMG/AMP variant criterion, has been incorporated into our analytical workflow (Xiang et al. 2020). This integration enables the systematic evaluation of the PVS1 criterion level across various genomic alterations, encompassing start/stop gain/loss SNVs, small frameshift indels, splicing variants, and multi-exon SVs. Furthermore, our workflow meticulously extracts detailed information based on PVS1 decision tree, thereby elucidating the foundational data supporting the determination of PVS1 criterion levels.

**PM2:** To evaluate the rarity of variants in the population, disease-specific PM2 thresholds for population allele frequency, pertinent to both dominant and recessive genes within the cohort, were initially established through a manual review of the literature related to the cohort’s disease. For other genes, thresholds were adjusted to ensure they remained below the frequency of the most prevalent known PLP variants for each gene, as classified in ClinVar. For example, after an initial review of literature, we determined the PM2 threshold of RP to be 1:4000 or 0.00025. The most common known PLP variant in the RP1 gene in ClinVar associated with IRD is NM\_006269.2:c.6181del, which has an East Asian allele frequency of 0.000192. Therefore, the PM2 threshold for candidate variants in this gene was lowered to this frequency. Variants would be designated as supporting PM2 (PM2\_Supporting) if their population allele frequencies in gnomAD fall below the derived gene-specific threshold.

**PS4:** To evaluate the increase in variant prevalence among patients, two distinct sources were utilized: (1) the number of submissions for each variant in ClinVar, categorized as pathogenic, likely pathogenic, likely benign, and benign, were extracted from the most recent ClinVar FTP site; (2) the prevalence of each allele within the same cohort was quantified. Occurrences in patients were further evaluated using threshold-based rules adopted from several gene-specific ClinGen VCEP guidelines, Glaucoma VCEP and Hearing Loss VCEP. These thresholds were applied alongside the PM2 criterion to assign PS4 levels.

**PS1/PM5:** To determine if the same or different amino acid change had been reported as pathogenic, variants reported a high ( $\geq 2$ ) gold star PLP classification in ClinVar, including those vetted by the ClinGen-reviewed variant list, were incorporated into a known positive variant repository. This repository was then utilized to ascertain whether the variant under examination exhibits the same or a different amino acid substitution at the identical position within the peptide, compared to the variants in the known positive list.

**PM3:** To determine if a variant reported *in trans* with other pathogenic variants, variants located in genes co-occurring with at least one other known positive variant, or a novel PLP variant (defined in S-BCAW Step 2) would be pre-assigned for PM3. The criterion for a known PLP variant was the same as defined in PS1/PM5. According to the ClinGen Sequence Variant Interpretation Recommendation for the *in trans* Criterion (PM3) (Group 2019), if the concurrent variant in the gene is pathogenic, the PM3\_Supporting classification is assigned, indicating a stronger correlation with pathogenic potential. Conversely, if the co-occurring variant is likely pathogenic, a PM3\_Tentative classification is assigned, reflecting a lower strength of evidence, quantified as 0.25 points. This does not meet the threshold for full support but provides suggestive evidence for further review. The mode of inheritance (MOI) of the associated disease was considered in Step 2.

Some criteria that were automatically assigned simply apply the threshold-based rules used in MICAW, including **PP2** (missense intolerance gene) based on gene specific missense constraint z-score from gnomAD, and **PP3** (in-silico evidence) based on REVEL score.

In addition, Mastermind (Chunn et al. 2020), designed for searching scientific literatures by genomic variations, was included in the S-BCAW workflow. The number of scientific publications associated with the variants involved was fetched from Mastermind API, together with the variant specific link to Mastermind website for literature review. This integration helps reviewers to access related ACMG criteria that relies on scientific literatures: **PP1** (co-segregation), **PS4**, **PM3** and **PP4** (disease specific phenotype).

### S-BCAW Step 2: variant classification based on calculated ACMG criteria, gene inheritance mode and variant zygosity

After the automatic assignment of ACMG criteria, variants were classified into different actionable baskets for further evaluation (Supplementary Table S2). Firstly, combined ACMG points were calculated according to point values for ACMG/AMP strength of criterion categories, designed based on a Bayesian probability model to assess the likelihood of pathogenicity (Tavtigian et al. 2020). Subsequently, the total number of variants present within each gene in the

filtered variant list was tallied. Additionally, the gene's MOI and the zygosity of the variants were considered. Variants were allocated to the PLP classification basket if the pathogenicity assessment, combined with the count of variant(s) in the gene, substantiates a significant correlation with the disease under investigation in the cohort. Variants that contained but did not have sufficient pathogenic ACMG criteria to reach likely pathogenic classification were placed into the Possibly Pathogenic (PossP) basket. The remaining variants, which did not qualify for the categories, were designated as Variants of Uncertain Significance (VUS) and were excluded from the initial automated process to be revisited during subsequent manual curation. This systematic approach ensures a structured and rigorous evaluation of genetic variants, facilitating precise stratification based on their potential clinical relevance.

### S-BCAW Step 3: Manual curation for variants in PLP and possible pathogenic basket

For variants allocated to the PLP and PossP classification baskets, the automatically assigned ACMG criteria underwent manual review to corroborate the validity of the sources. Variants in the PLP basket were prioritized to ensure timely and accurate assessments, which was essential for robust disease causative and pathogenicity evaluations. Furthermore, ACMG criteria related to literature evidence and sample phenotypes were meticulously evaluated and additionally assigned by experts. Variants in the VUS basket were not included for manual review.

After this manual intervention, the combined ACMG points and the classification of each variant were updated to reflect the additional insights gained during the review process. This comprehensive approach allowed for a more accurate determination of variant pathogenicity.

Cases containing variants that remained in the PLP basket or promoted from PossP to PLP basket after thorough review were classified as positive cases. Concurrently, the variants in the PLP basket for these positive cases were designated as causative pathogenic (P) or likely pathogenic (LP) variants according to their reviewed ACMG classification. Conversely, samples without PLP variants, or where PLP variants could not be substantiated upon manual review, were classified as negative cases. This classification strategy facilitated a clear distinction between cases likely associated with the disease phenotype and those not linked to the observed clinical manifestations.

### Result comparison and reporting

Diagnostic outcomes derived from the two methodologies were compared prior to further analysis, encompassing the diagnostic conclusions of the cases, the identification of

disease-causing variants, and the assigned ACMG criteria. Any discrepancies observed between the results were thoroughly discussed and resolved through a joint review conducted by the reviewers from both approaches.

Additionally, the processing time for each approach, measured from the point of obtaining the filtered variant list to the completion of variant classification, was recorded and compared to evaluate efficiency.

In cases identified as compound heterozygous, the phasing status was validated using Nanopore sequencing to ensure the accuracy of the genetic interpretation. Following these validations and consultations with the referring clinicians, research reports were prepared and subsequently issued to all patients involved in the study.

This structured approach to comparison and validation ensured that the diagnostic results were both accurate and consistent, providing a reliable foundation for subsequent clinical decision-making.

## Results

### Participants

A total of 79 Chinese probands (39 male) tentatively diagnosed with RP and 30 family members (in 7 duos, 2 trios, 1 quartet and 8 families with other family structure) consented to take part in HKGP were enrolled in this study, with an average age of 52 years old (range: 6–74), respectively (Supplementary Table S3, Supplementary Figure S1). The major clinical phenotypes observed in the probands, in addition to RP, were visual impairment (HP:0000505), retinal dystrophy (HP:0000556) and rod-cone dystrophy (HP:0000510). Selected family members were also enrolled to confirm the mode of variant inheritance.

### Candidate variant list generation for workflow comparison

After GS, mapping, variant calling, and variant basic filtering based on allele frequency, location, and a list of RP-related genes, an average of 16 variants remained for each case (range: 8–24). Each variant was annotated with basic information, including genomic and gene locus, variant type and consequence, and allele frequencies in public databases. In silico deleteriousness prediction tools were also used. In addition, clinical phenotypes for each case were documented using HPO terms.

### Results from MICAW

The evaluation of candidate variants, after basic filtering and RP related gene panel filtering of all 79 cases, was

meticulously conducted by a team of skilled variant reviewers using MICAW. This rigorous process required, on average, 2.54 h per case for both reviewer 1 and reviewer 2 to assess and establish the pathogenicity of candidate variants through the application of ACMG criteria (Fig. 2B).

The analysis successfully identified 25 positive cases exhibiting PLP variants which were potentially causative for RP. Furthermore, 54 cases were determined to be negative, showing no association with PLP variants linked to RP. The diagnostic yield for our cohort is 32%.

Detailed genetic profiling of the 25 positive cases revealed that 20 cases possessed PLP variants in configurations suggestive of autosomal recessive inheritance: 14 cases were compound heterozygous, and six were homozygous. The remaining five cases displayed either heterozygous or hemizygous PLP variants in genes typically associated with autosomal dominant or X-linked inheritance patterns. The PLP variants segregate in affected family members that were sequenced.

### Results from S-BCAW

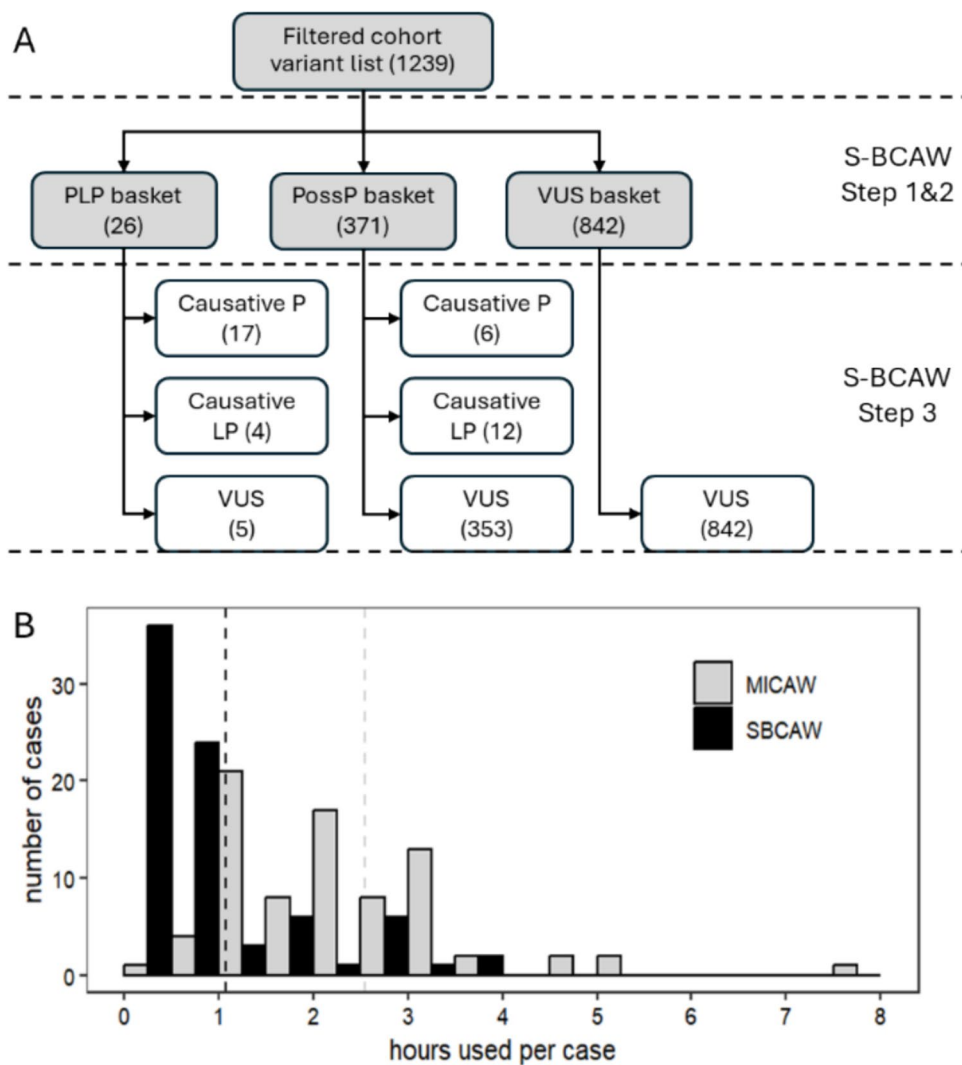
In parallel, an RP-specific analysis was conducted utilizing the S-BCAW. This analysis identified 26 variants within the PLP classification basket across 23 cases. Additionally, 371 variants were categorized within the PossP basket spanning 78 cases, following the automated assignment of ACMG criteria (Fig. 2A).

Subsequent meticulous manual review, which involved further assignment of ACMG criteria based on literature evidence and phenotype correlations, confirmed that 21 out of the 26 PLP variants were causative for RP across 16 cases. Moreover, this process led to the reclassification of 18 PossP variants in nine cases as causative for RP, following rigorous evaluation by the curation review team. The remaining cases were maintained in their respective preliminary classifications, predominantly as negative (Supplementary Table S4). The average processing time per sample was recorded at 1.08 h, including both auto and manual steps.

### Comparison of results between S-BCAW and MICAW

The processing time per case was analysed and compared between the S-BCAW and MICAW methodologies. As illustrated in Fig. 2B, the time required to curate each case demonstrates a leftward shift in the distribution for S-BCAW relative to MICAW, with a highly significant difference ( $p = 1.07E-14$ ). On average, the time reduction observed with S-BCAW compared to MICAW was 58%, with a more pronounced reduction in positively diagnosed cases (Table 1). Specifically, the proportion of cases processed in one hour or less increased dramatically, while the number of cases requiring extended processing time ( $\geq 3$  h) decreased

**Fig. 2** **A** Variant classification before and after manual review in S-BCAW. **B** Hours used per case in MICAW and S-BCAW (grouped in 0.5-h intervals). The vertical dashed lines are the average hours used of the 79 cases



**Table 1** Processing time comparison between S-BCAW and MICAW

Case groups	Average processing hour			#(%) cases ≤ 1 h		#(%) cases ≥ 3 h	
	S-BCAW	MICAW	Time saved in S-BCAW	S-BCAW	MICAW	S-BCAW	MICAW
All cases	1.08	2.54	58%	60 (76%)	5 (6%)	9 (11%)	28 (35%)
Positive cases	0.83	2.24	63%	21 (84%)	4 (16%)	1 (4%)	6 (24%)
Negative cases	1.19	2.69	56%	39 (72%)	1 (2%)	8 (15%)	22 (41%)

substantially. These findings underscore the efficiency of S-BCAW in enhancing the genetic analysis workflow, enabling more rapid and accurate variant classification, and improving the overall diagnostic procedure for RP.

A comparison of the final classifications from MICAW and S-BCAW of the 79 RP cases shows 100% consistency, with 25 positive diagnostic cases and 54 negative cases (Table 2; Supplementary Table S4). A detailed comparison of all the individual ACMG criteria assigned to disease-causing variants in positive cases also showed 100%

consistency, which included 30 very strong, 11 strong, 17 moderate and 57 supporting criteria. Within S-BCAW, the ACMG criterion assignment list was also compared before and after manual review. A total of 93 criteria were automatically assigned in S-BCAW before review, and 86 of these were kept after manual review, accounting for 75% (86/115) of the confirmed criteria list after manual review (Table 3). In addition, there are 14 negative cases with a heterozygous PLP variant in a recessive gene without a second hit, and the ACMG criterion assignment for

**Table 2** List of pathogenic and likely pathogenic variants in positively diagnosed cases

Case ID	Gene	HGNC gene ID	MOI	Variant description in HGVS
HKGP003011-1	<i>ABCA4</i>	HGNC:34	AR	NM_000350.3:c.3055A > G p.(Thr1019Ala)
HKGP003011-1	<i>ABCA4</i>	HGNC:34	AR	NM_000350.3:c.1804C > T p.(Arg602Trp)
HKGP003071-1	<i>CEP290</i>	HGNC:29,021	AR	NM_025114.4:c.1616del p.(Leu539Ter)
HKGP003071-1	<i>CEP290</i>	HGNC:29,021	AR	NM_025114.4:c.6798G > A p.(Trp2266Ter)
HKGP003648-1	<i>CEP290</i>	HGNC:29,021	AR	NM_025114.4:c.6798G > A p.(Trp2266Ter)
HKGP003648-1	<i>CEP290</i>	HGNC:29,021	AR	NM_025114.4:c.3802C > T p.(Gln1268Ter)
HKGP003628-1	<i>CHM</i>	HGNC:1940	AD	NM_000390.4:c.1019C > A p.(Ser340Ter)
HKGP003466-1	<i>CNGA1</i>	HGNC:2148	AR	NM_001379270.1:c.1675G > C p.(Ala559Pro)
HKGP003466-1	<i>CNGA1</i>	HGNC:2148	AR	NM_001379270.1:c.253del p.(Leu85PhefsTer4)
HKGP003627-1	<i>CNGA1</i>	HGNC:2148	AR	NM_001379270.1:c.253del p.(Leu85PhefsTer4) Homozygous
HKGP002721-1	<i>CYP4V2</i>	HGNC:23,198	AR	NM_207352.4:c.802-8_810delinsGC p.? Homozygous
HKGP002725-1	<i>CYP4V2</i>	HGNC:23,198	AR	NM_207352.4:c.802-8_810delinsGC p.?
HKGP002725-1	<i>CYP4V2</i>	HGNC:23,198	AR	NM_207352.4:c.1091-2A > G p.?
HKGP003469-1	<i>CYP4V2</i>	HGNC:23,198	AR	NM_207352.4:c.802-8_810delinsGC p.? Homozygous
HKGP002724-1	<i>EYS</i>	HGNC:21,555	AR	NM_001142800.2:c.6557G > A p.(Gly2186Glu)
HKGP002724-1	<i>EYS</i>	HGNC:21,555	AR	NM_001142800.2:c.7492G > C p.(Ala2498Pro)
HKGP002996-1	<i>EYS</i>	HGNC:21,555	AR	NM_001142800.2:c.6416G > A p.(Cys2139Tyr)
HKGP002996-1	<i>EYS</i>	HGNC:21,555	AR	NM_001142800.2:c.2486del p.(Ile829ThrfsTer39)
HKGP003637-1	<i>EYS</i>	HGNC:21,555	AR	NM_001142800.2:c.7228 + 1G > A p.? Homozygous
HKGP003656-1	<i>MYOC</i>	HGNC:7610	AD	NM_000261.2:c.1495A > T p.(Ile499Phe)
HKGP003646-1	<i>PDE6B</i>	HGNC:8786	AR	NM_000283.4:c.1133G > A p.(Trp378Ter) Homozygous
HKGP002727-1	<i>RCBTB1</i>	HGNC:18,243	AR	NM_018191.4:c.1262_1263del p.(Tyr421SerfsTer31)
HKGP002727-1	<i>RCBTB1</i>	HGNC:18,243	AR	NM_018191.4:c.707del p.(Asn236ThrfsTer11)
HKGP002986-1	<i>RHO</i>	HGNC:10,012	AD	NM_000539.3:c.180C > A p.(Tyr60Ter)
HKGP003634-1	<i>RPGR</i>	HGNC:10,295	XLD	NM_001034853.2:c.442G > T p.(Gly148Ter) Hemizygous
HKGP002989-1	<i>TOPORS</i>	HGNC:21,653	AD	NM_005802.5:c.59dup p.(Pro21AlafsTer8)
HKGP002718-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.10637G > A p.(Gly3546Glu)
HKGP002718-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.5581G > A p.(Gly1861Ser)
HKGP002722-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.15178 T > C p.(Ser5060Pro)
HKGP002722-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.2653C > T p.(His885Tyr)
HKGP002874-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.9570 + 1G > A p.? Homozygous
HKGP002991-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.8254G > A p.(Gly2752Arg)
HKGP002991-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.5572 + 1G > A p.?
HKGP002994-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.449 T > G p.(Leu150Ter)
HKGP002994-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.9570 + 1G > A p.?
HKGP003075-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.2802 T > G p.(Cys934Trp)
HKGP003075-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.8559-2A > G p.?
HKGP003468-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.7184_7194del p.(Leu2395HisfsTer19)
HKGP003468-1	<i>USH2A</i>	HGNC:12,601	AR	NM_206933.4:c.5530C > T p.(Gln1844Ter)

these variants was also consistent between MICAW and S-BCAW (data not shown).

For the criteria that were fully auto-assigned, namely PVS1, PM2 and PP3, there was very high consistency after review. However, adjustments were made to the PP3 and PM2 criteria for certain genes, based on manual reviews guided by gene specific ClinGen VCEP Specifications, including *USH2A* in ClinGen Hearing Loss VCEP and *MYOC* in Glaucoma VCEP Specifications (Patel et al. 2021; Burdon et al.

2022). One variant NM\_206933.2(*USH2A*):c.8559-2A > G p.? was curated by ClinGen Hearing Loss VCEP, therefore the evidence assigned by the VCEP was followed instead of the Specification, with PVS1\_Moderate replaced by PM4. For the criteria that were preassigned based on partially collected variant-related literature and resources, namely PS1, PS4, PM3 and PM5, the concordance was more than 50%, indicating the effectiveness of such approaches in significantly accelerating the review efficiency of their assignment.



**Table 3** Individual ACMG criterion assigned by S-BCAW in 25 positive cases

ACMG criterion		No. of assigned to causative variants in S-BCAW		Concordance (omit strength shift)
Automation group in S-BCAW	With shifted strength	Before manual review	After manual review	
Fully auto assigned	PVS1_Moderate	2	1	96%
	PVS1	24	24	
	PM2_Supporting	37	38	
	PP2	0	0	
	PP3_Strong	3	3	
	PP3_Moderate	3	1	
	PP3	4	4	
Preassigned by partial resource	PS1	0	0	56%
	PS4	0	1	
	PS4_Supporting	0	2	
	PM3_VeryStrong	0	6	
	PM3_Strong	0	7	
	PM3	0	11	
	PM3_Supporting	13	8	
	PM3_Tentative	7	0	
	PM5	0	1	
Not assigned, only partial resource collected	PP1_Moderate	Not applicable	1	Not applicable
	PP1		2	
	PP4		3	
	PS2_Supporting		0	
	PS3_Moderate		1	
	PM4		1	
	Total		93	

The differences mainly arose from manually reviewed literature and the related phenotype of the case, which currently lack tools for accurate automated information extraction. The remaining criteria that purely rely on manual review on related literature and resources only accounted for 7% (8/115) of the total criteria assigned after review.

Collectively, a significant majority (93%, 107/115) of the individual ACMG criteria associated with the RP-causative PLP variants either fully or partially benefited from the automatic assignment process. This, coupled with the automated variant classification, contributed substantially to the time savings achieved using S-BCAW.

### S-BCAW detects novel and known RP-causing variants, resulting in ending of the diagnostic odysseys

All 39 causative PLP variants positive cases found in S-BCAW were validated for the call and/or confirmation of phasing status (for autosomal recessive conditions) using Sanger sequencing, Nanopore long read sequencing or zygosity of family members where available. The results showed 100% concordance with findings from short-read GS. Of

these variants, *USH2A* (HGNC:12,601) was the most frequently reported gene, appearing in 7 of the 25 diagnosed cases (28%), followed by *EYS* (HGNC:21,555) and *CYP4V2* (HGNC:23,198), each with 3 cases (11% each). Three variants in this cohort were reported for the first time as causative of RP, including one stop-gained variant *CEP290* (HGNC:29,021) (NM\_025114.4:c.1616del p.(Leu539Ter)) associated with *CEP290*-related ciliopathy according to ClinGen;(GCEP) one stop-gained hemizygous SNV in *RPGR* (HGNC:10,295) (NM\_001034853.2:c.442G > T p.(Gly148Ter)) associated with *RPGR*-related retinopathy according to ClinGen;(ClinGen) and one frameshift variant in *TOPORS* (HGNC:21,653) (NM\_005802.5:c.59dup p.(Pro21AlafsTer8)) associated with retinitis pigmentosa 31 (Phenotype MIM no. 609923). A founder variant in the *CYP4V2* gene NM\_207352.4:c.802-8\_810delinsGC p.? was present in homozygous form in two patients HKGP002721-1, HKGP003469-1 and *in trans* with another pathogenic variant in HKGP002725-1. LoF variants in this gene are associated with autosomal recessive Bietti's crystalline corneoretinal dystrophy (Phenotype MIM number: 210370) (Nakamura et al. 2006). In the gnomAD population database version 4.1.0, which comprises allele frequency data of

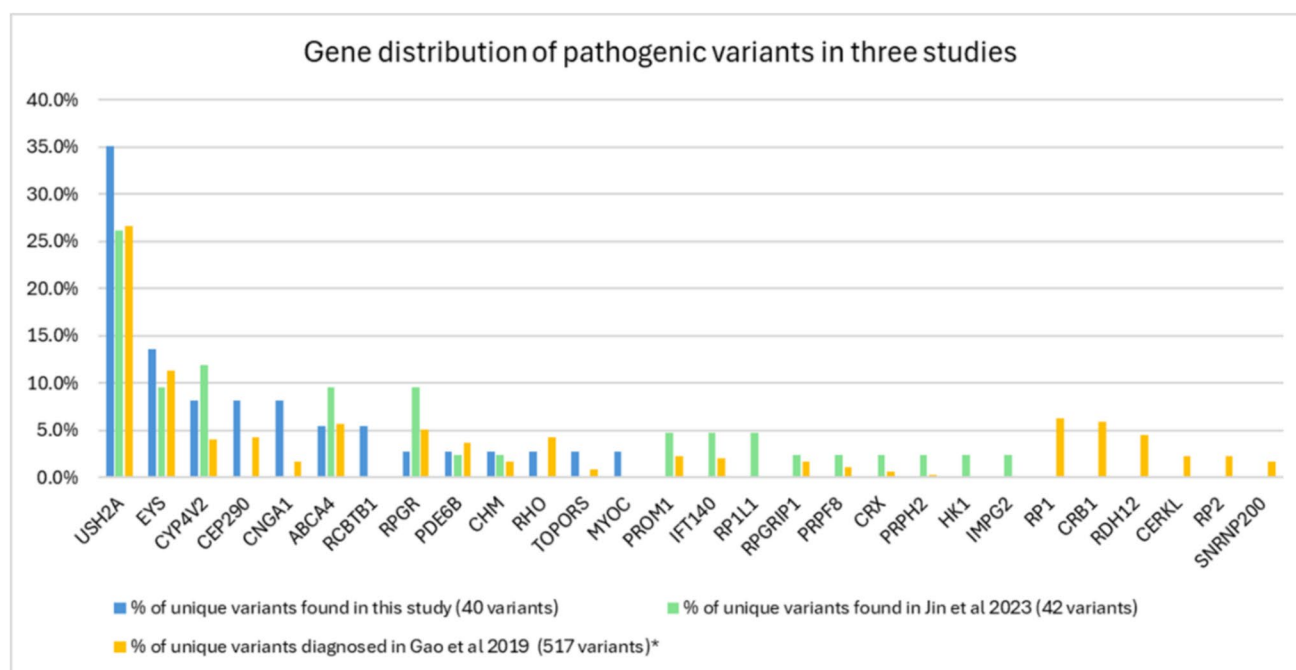
mostly ostensibly healthy individuals, this variant is found almost exclusively in East Asians with an allele frequency of 0.002832. This variant has been reported mostly in homozygous form and *in trans* with other variants in other Chinese cohort and Japanese cohorts (Nakano et al. 2012; Jarrar et al. 2020).

The positive cases exhibited an average diagnostic interval of 18.0 ( $\pm$  11.3) years from the onset of symptoms to the eventual diagnosis (Supplementary Table S5). This prolonged diagnostic journey emphasizes the significant challenges faced by individuals with RP and underscores the importance of the methodologies employed in this study. The findings highlight the critical need for earlier diagnostic interventions and the potential of molecular diagnostic approaches to substantially reduce the diagnostic odyssey for patients suffering from RP, with implications for earlier treatment and prevention.

## Discussion

This study outlines the implementation of S-BCAW and its comparison with the traditional MICAW approach in a cohort of clinical cases diagnosed with RP. The genetic variant curation phase presents substantial challenges, particularly in environments where prompt decision-making is critical. Manual variant curation is labour-intensive and prone to

delays, with the prevalence of variants of VUS and limitations in standard interpretation criteria complicating result interpretation and hindering timely clinical responses. These challenges can negatively impact effective patient care and treatment outcomes. Our proposed semi-automated bespoke cohort analysis workflow (S-BCAW) addresses these challenges by enhancing the efficiency and accuracy of handling and interpreting complex NGS data. Our findings reveal that while S-BCAW maintains the same level of accuracy as MICAW, it significantly enhances diagnostic efficiency. Specifically, S-BCAW reduces the time from variant list generation to diagnostic conclusion by around 60%. Additionally, it provides consistent performance in the assignment of individual ACMG criteria, mirroring the reliability of the conventional method. Furthermore, we successfully diagnosed 25 out of 79 cases, including the identification of 3 novel variants. We compared the genes with positive findings in our cohort versus those in a study of 75 cases by Jin et al (2023) and 1243 cases by Gao et al (2019). All three cohorts comprised patients from China. In all three studies, *USH2A* was the most reported gene, comprising 25% and 22% of patients diagnosed in our study and Jin et al.'s study respectively (Fig. 3; Supplementary Table S6); Gao et al (2019) did not provide the breakdown of number of patients by genes, but they had the most reported variants in *USH2A*. It is notable that despite the much larger cohort in Gao et al.'s study, both our and Jin et al.'s cohorts have



**Fig. 3** Distribution of reported pathogenic variants by gene. The genes with no reported variants in Gao et al.'s study are confirmed to be included in their gene panel tested. For the genes with no reported

variants in Jin et al.'s study, we assume the genes' exonic regions are covered in their WES target region (BGI exome V4 kit)

positive findings in genes not reported in Gao et al.'s cohort, such as the *MYOC* and *RCBTB1* pathogenic variants found exclusively in our study. This highlights the importance of continued genetic testing for individuals with suspected RP in a timely manner to understand the full spectrum of variants causative of RP in the Chinese population and other ethnicities.

### Enhanced efficiency and time savings

The primary factor contributing to this efficiency gain was the substantial reduction in the number of variants requiring manual review. By automatically assigning selected ACMG criteria to variants likely to be finally classified as PLP, S-BCAW concentrates manual curation efforts on variants that either require validation of the details of automatically assigned criteria or are auto-determined “hot VUS” that lack just a few more criteria for a definitive PLP classification. This approach notably reduces the time spent on variants with minimal pathogenic criteria (“cold VUS” without any pathogenic criteria automatically assigned) that are challenging to classify positively even with additional phenotypic or literature criteria. In contrast, such variants are often inconsistently handled across different reviewers in the MICAW approach.

S-BCAW integrates advanced tools specifically designed for the application of ACMG criteria, particularly for null variants, which are significant in disease mechanism and constitute approximately 30% (150 out of 498) of the variants in the ClinGen Expert curated list (excluding structural variants involving multiple genes) (Wilcox et al. 2021). The ACMG rule for the PVS1 criterion requires assessment of five distinct biological and genomic features before assigning weight-adjusted pathogenic criterion, which varies depending on the gene involved and its location within the gene. This assessment traditionally consumes a substantial amount of curation time, particularly in evaluating nonsense-mediated mRNA decay (NMD) and the functionality of affected exonic regions. Embedded within S-BCAW, the AutoPVS1 tool automates the classification of null variants according to the PVS1 criterion of the ACMG/AMP guidelines. While manual curation remains necessary to fine-tune the pathogenicity level according to specific disease mechanisms, AutoPVS1 significantly streamlines the collection and application of relevant variant information for the PVS1 decision tree, enhancing both consistency and accuracy.

Furthermore, S-BCAW automates the aggregation of scientific literature from multiple sources for further review and criterion assignment. The workflow automatically groups and counts variant classifications from ClinVar submissions, which approximates variant prevalence in cases, to assist in assessing the application of PS4. Although manual review of additional scientific reports is still required to finalize PS4

assignment, this automation substantially reduces the effort needed. Additionally, scientific publications pertinent to the variants under review are efficiently collated using the Mastermind API (Chunn et al. 2020), providing the publication number and direct links with highlighted content relevant to each variant, thereby facilitating a more streamlined review process, accelerating the assignment of ACMG criteria that highly rely on scientific studies, such as PS3 (functional study), PS4 (variant prevalence) and PM3 (*in trans* with pathogenic variant).

### Improved accuracy and consistency in assigning ACMG criteria

S-BCAW ensures that disease-specific thresholds are applied uniformly, enhancing the accuracy of associated ACMG criteria interpretations. Key to this process is the PM2 criterion, which evaluates the rarity of a variant in the context of a specific disease. For S-BCAW, an initial manual assessment of the general disease prevalence and the most frequent and known pathogenic variant of RP is performed, based on multiple review publications, establishing specific PM2 thresholds for dominant and recessive genes.

Moreover, S-BCAW leverages variant data not only on variants across all genes in individual samples, but also across different samples within the same gene, enhancing the consistency of criterion assignment. The occurrence of the same variant across multiple samples contributes to the PS4 classification, which assesses variant prevalence of the cohort under analysis. Additionally, a variant will automatically receive a tentative PM3 classification if located in a gene that already harbours any known PLP variants. This automated criterion assignment feature in S-BCAW contrasts sharply with the MICAW approach, where criterion assignment can be hindered by the need for manually maintained records of positive variants or the requirement to manually browse sample variant lists.

### Comparative insights from similar approaches

In the rapidly evolving field of genetic diagnostics, each study introduces unique methodologies to enhance the accuracy and practicality of interpreting NGS data across different disease cohorts. For example, Raquel's study develops a custom bioinformatics pipeline for cancer and cardiovascular cohorts that integrates advanced tools for variant calling, annotation, and classification, significantly increasing the diagnostic yield for these diseases (Romero et al. 2022). Schobers' research focuses on paediatric neurology, emphasizing the need for continuous re-evaluation and systematic re-analysis of cases with initially negative exome sequencing results, which has substantially improved diagnostic yield from 31 to 53% by refining bioinformatics tools and using

updated disease-gene associations (Schobers et al. 2022). Similarly, the Primary Immune Deficiency (PID) study expands gene panels and incorporates structural variant analyses, highlighting the effectiveness of continuous data re-evaluation and deeper genotype–phenotype correlations critical for disorders like PID with complex genetic backgrounds (Mørup et al. 2022). Masson et al. (2022) expanded the ACMG variant classification guidelines into a general framework that provides a broader, more nuanced classification system by integrating additional categories and specific thresholds for allele frequency and functional impacts (Masson et al. 2022). This refined system adapts the ACMG guidelines to a wider variety of genetic effects, improving precision in clinical genetic evaluations.

Our approach introduces a semi-automated system that merges innovative technologies with expert review to optimize the diagnostic workflow. By automatically assigning specific ACMG, S-BCAW streamlines the initial variant analysis process, ensuring a consistent and accurate application of these criteria. After automatic criteria assignment, as inspired by Masson's, variants are classified into actionable baskets based on calculated ACMG points, gene inheritance mode, and variant zygosity, ensuring a structured evaluation that allows for precise stratification based on clinical relevance. Variants classified into critical baskets undergo a rigorous manual review to confirm the validity of automatically assigned criteria and to incorporate additional insights from literature and phenotypic data. This dual-layer review process enhances the accuracy of the final variant classification.

While other studies predominantly focus on increasing diagnostic yield through expanding gene panels, variant types and employing new tools, our S-BCAW has expanded the clinical focus from solely improving diagnostic yield to also enhancing operational efficiency while maintaining high diagnostic standards. Across all these studies, the central theme is the drive towards tailored, innovative solutions that enhance diagnostic accuracy and yield, reflecting a commitment to adapting to the complexities of various genetic disorders and improving the precision and practicality of genetic diagnostics.

### Limitations and potential for further improvement

Several limitations of this study warrant discussion. First, while our cohort of 79 RP cases represents one of the largest GS-based studies for this condition, the sample size may not be sufficient to comprehensively validate the effectiveness and generalizability of S-BCAW across larger populations. Further evaluation with expanded cohorts would strengthen the assessment of this workflow's reliability and scalability.

Second, although S-BCAW was specifically optimized for RP variant analysis in this study, its modular architecture allows for potential adaptation to other genetic disorders.

The workflow can be customized to accommodate disease-specific requirements, including modified pathogenicity criteria from ClinGen gene-specific guidelines that deviate from standard ACMG variant interpretation guidelines. However, the broader applicability of S-BCAW across different genetic conditions requires systematic validation through additional studies.

Third, only a fraction of ACMG criteria is covered in the current S-BCAW, leaving the rest to be evaluated manually. This includes some criteria with well-defined rules that have the potential to be assigned automatically before manual review. One potential criterion is PM1, defined by mutational hotspots and functional domain. Mutational hotspots can be predicted by utilizing the number of benign missense variants in public variant databases. ClinGen Variant Curation Expert Panels Specifications need to be reviewed and summarized to establish a robust definition for the hotspots. Functional domains can be well-defined by utilizing related public knowledge-based databases. InterPro, for example, provides functional analysis of protein sequences by classifying them into families and predicting domains and important sites (Paysan-Lafosse et al. 2022). This database is particularly useful for PM1 assignment because it combines information from various sources such as Pfam, SMART, ProDom, and others, providing a broad overview of domain functions and their biological implications. However, it is important to note that the presence of a variant in a functional domain alone is insufficient for PM1 classification. To apply PM1, evidence must demonstrate that variants in the specific domain are associated with disease. This can be achieved by reviewing ClinVar or relevant publications to identify pathogenic mutations in the domain and their disease associations. Another two criteria with room for improved automatic assignment are PS1 and PS4. Currently, PS1 and PS4 are partially assigned criteria, leveraging ClinVar as variant evidence sources. Further refinement could incorporate disease matching to ensure only relevant ClinVar entries are considered. This requires developing algorithms to accurately match phenotypic descriptions within ClinVar records with the disease under investigation.

In addition, ClinGen Expert Panels continuously release and update gene-specific variant curation guidelines that modify certain ACMG criteria and their associated strength specifications. This highlights the actively evolving field of variant interpretation, which drives our commitment to future development and enhancement to accommodate these advancements. Furthermore, while S-BCAW efficiently processes certain nuclear chromosome variations, larger indels and mitochondrial DNA variants still require manual review, underscoring the need for ongoing refinement to expand its analysis capabilities.

Looking ahead, we plan to further enrich this workflow by integrating additional ACMG criteria, thereby broadening

our framework for a more comprehensive variant classification. Additionally, we will apply this enhanced approach across a wider range of disease cohorts. This expansion will not only validate the effectiveness of our methodology across different clinical contexts but also increase its utility and impact. By doing so, we anticipate improvements in diagnostic yields and the ability to uncover new genotype–phenotype correlations, which could significantly advance genetic research and therapeutic strategies.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00439-025-02737-x>.

**Acknowledgements** We thank all patients and families for participating in the study.

**Author contributions** Conceptualization: B.H.Y.C. Data curation: H.Y.F.T., J.S.L.K. Formal analysis: D.Y. Funding acquisition: H.K.G.P. Investigation: H.Y.F.T., M.W., D.O., S.P.Y.H., J.S.L.K. Methodology: J.S.L.K., D.Y. Project administration: J.S.L.K. Supervision: B.H.Y.C., A.T.W.C. Validation: M.W., Q.L. Writing-original draft: D.Y., J.S.L.K. Writing-review & editing: B.H.Y.C., A.T.W.C., C.K.S.L. All authors read and reviewed the manuscript.

**Funding** The HKGP is a publicly funded genome sequencing initiative commissioned by the Health Bureau of the HKSAR Government.

**Data availability** The data supporting this article are provided in the supplementary files available in the online version of this article at the publisher's website.

## Declarations

**Conflict of interests** The authors have no relevant financial or non-financial interests to disclose.

**Ethics approval** Ethics approvals were granted by the Central Institutional Review Board (IRB) (HKGP-2021–001, HKGP-2022–001), and the IRBs of the Department of Health (L/M 257/2021), and the University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 21–413, UW 23–289).

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent to publish** The authors affirm that human research participants provided informed consent for publication of the data in this paper. All participants consented to the submission of the case reports to the journal.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from

the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Arteche-López A, Ávila-Fernández A, Romero R et al (2021) Sanger sequencing is no longer always necessary based on a single-center validation of 1109 NGS variants in 825 clinical exomes. *Sci Rep* 11:5697. <https://doi.org/10.1038/s41598-021-85182-w>
- Bagger FO, Borgwardt L, Jespersen AS et al (2024) Whole genome sequencing in clinical practice. *BMC Méd Genom* 17:39. <https://doi.org/10.1186/s12920-024-01795-w>
- Bertier G, Héту M, Joly Y (2016) Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Méd Genom* 9:52. <https://doi.org/10.1186/s12920-016-0213-6>
- Bhardwaj A, Yadav A, Yadav M, Tanwar M (2022) Genetic dissection of non-syndromic retinitis pigmentosa. *Indian J Ophthalmol* 70:2355–2385. [https://doi.org/10.4103/ijo.ijo\\_46\\_22](https://doi.org/10.4103/ijo.ijo_46_22)
- Briek P, Bulić L, Bračić M et al (2024) Implementing whole genome sequencing (WGS) in clinical practice: advantages, challenges, and future perspectives. *Cells* 13:504. <https://doi.org/10.3390/cells13060504>
- Burdon KP, Graham P, Hadler J et al (2022) Specifications of the ACMG/AMP variant curation guidelines for myocilin: recommendations from the clingen glaucoma expert panel. *Hum Mutat* 43:2170–2186. <https://doi.org/10.1002/humu.24482>
- Chen S, Francioli LC, Goodrich JK et al (2024) A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 625:92–100. <https://doi.org/10.1038/s41586-023-06045-0>
- Chu ATW, Project HKG, Fung JLF et al (2022) Potentials and challenges of launching the pilot phase of Hong Kong Genome Project. *J Transl Genet Genom*. <https://doi.org/10.2051/jtgg.2022.02>
- Chu ATW, Tong AHY, Tse DMS et al (2023) The Hong Kong genome project: building genome sequencing capacity and capability for advancing genomic science in Hong Kong. *J Transl Genet Genom* 7:196–212. <https://doi.org/10.2051/jtgg.2023.22>
- Chunn LM, Nefcy DC, Scouten RW et al (2020) Mastermind: a comprehensive genomic association search engine for empirical evidence curation and genetic variant interpretation. *Front Genet* 11:577152. <https://doi.org/10.3389/fgene.2020.577152>
- ClinGen ClinGen: RPGR-related retinopathy. <https://search.clinicalgenome.org/CCID:006026>. Accessed 1 May 2024
- Colombo L, Maltese PE, Castori M et al (2021) Molecular epidemiology in 591 Italian probands with nonsyndromic retinitis pigmentosa and usher syndrome. *Investig Ophthalmol vis Sci* 62:13. <https://doi.org/10.1167/iov.62.2.13>
- Denny JC, Rutter JL, Investigators A of URP et al (2019) The “all of us” research program. *N Engl J Med* 381:668–676. <https://doi.org/10.1056/nejmsr1809937>
- Gao F-J, Li J-K, Chen H et al (2019) Genetic and clinical findings in a large cohort of chinese patients with suspected retinitis pigmentosa. *Ophthalmology* 126:1549–1556. <https://doi.org/10.1016/j.ophtha.2019.04.038>
- GCEP R ClinGen CEP290-related ciliopathy. <https://search.clinicalgenome.org/CCID:004417>. Accessed 10 May 2024
- Group SVIW (2019) SVI Recommendation for in trans Criterion (PM3) - Version 1.0. [https://clinicalgenome.org/site/assets/files/3717/svi\\_proposal\\_for\\_pm3\\_criterion\\_-\\_version\\_1.pdf](https://clinicalgenome.org/site/assets/files/3717/svi_proposal_for_pm3_criterion_-_version_1.pdf)
- Halldorsson BV, Eggertsson HP, Moore KHS et al (2022) The sequences of 150,119 genomes in the UK biobank. *Nature* 607:732–740. <https://doi.org/10.1038/s41586-022-04965-x>

- Ioannidis NM, Rothstein JH, Pejaver V et al (2016) Revel: an ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet* 99:877–885. <https://doi.org/10.1016/j.ajhg.2016.08.016>
- Jarrar YB, Shin J, Lee S (2020) Identification and functional characterization of CYP4V2 genetic variants exhibiting decreased activity of lauric acid metabolism. *Ann Hum Genet* 84:400–411. <https://doi.org/10.1111/ahg.12388>
- Jin B, Li J, Yang Q et al (2023) Genetic characteristics of suspected retinitis pigmentosa in a cohort of Chinese patients. *Gene* 853:147087. <https://doi.org/10.1016/j.gene.2022.147087>
- Kim YN, Song JS, Oh SH et al (2020) Clinical characteristics and disease progression of retinitis pigmentosa associated with PDE6B mutations in Korean patients. *Sci Rep* 10:19540. <https://doi.org/10.1038/s41598-020-75902-z>
- Kopanos C, Tsiolkas V, Kouris A et al (2019) VarSome: the human genomic variant search engine. *Bioinformatics* 35:1978–1980. <https://doi.org/10.1093/bioinformatics/bty897>
- Masson E, Zou W-B, Génin E et al (2022) Expanding ACMG variant classification guidelines into a general framework. *Hum Genom* 16:31. <https://doi.org/10.1186/s40246-022-00407-x>
- Mørup SB, Nazaryan-Petersen L, Gabrielaite M et al (2022) Added value of reanalysis of whole exome- and whole genome sequencing data from patients suspected of primary immune deficiency using an extended gene panel and structural variation calling. *Front Immunol* 13:906328. <https://doi.org/10.3389/fimmu.2022.906328>
- Na K-H, Kim HJ, Kim KH et al (2017) Prevalence, age at diagnosis, mortality, and cause of death in retinitis pigmentosa in Korea—a nationwide population-based study. *Am J Ophthalmol* 176:157–165. <https://doi.org/10.1016/j.ajo.2017.01.014>
- Nakamura M, Lin J, Nishiguchi K et al (2006) Retinal degenerative diseases. *Adv Exp Med Biol* 572:49–53. [https://doi.org/10.1007/0-387-32442-9\\_8](https://doi.org/10.1007/0-387-32442-9_8)
- Nakano M, Kelly EJ, Wiek C et al (2012) CYP4V2 in Bietti's crystalline dystrophy: ocular localization, metabolism of  $\omega$ -3-polyunsaturated fatty acids, and functional deficit of the p. H331P variant. *Mol Pharmacol* 82:679–686. <https://doi.org/10.1124/mol.112.080085>
- Nangia V, Jonas JB, Khare A, Sinha A (2012) Prevalence of retinitis pigmentosa in India: the central india eye and medical study. *Acta Ophthalmol* 90:e649–e650. <https://doi.org/10.1111/j.1755-3768.2012.02396.x>
- Patel MJ, DiStefano MT, Oza AM et al (2021) Disease-specific ACMG/AMP guidelines improve sequence variant interpretation for hearing loss. *Genet Med : off J Am Coll Méd Genet* 23:2208–2212. <https://doi.org/10.1038/s41436-021-01254-2>
- Paysan-Lafosse T, Blum M, Chuguransky S et al (2022) InterPro in 2022. *Nucleic Acids Res* 51:D418–D427. <https://doi.org/10.1093/nar/gkac993>
- Pejaver V, Byrne AB, Feng B-J et al (2022) Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am J Hum Genet* 109:2163–2177. <https://doi.org/10.1016/j.ajhg.2022.10.013>
- Pozo MG, Fernández-Suárez E, Martín-Sánchez M et al (2020) Unmasking retinitis pigmentosa complex cases by a whole genome sequencing algorithm based on open-access tools: hidden recessive inheritance and potential oligogenic variants. *J Transl Med* 18:73. <https://doi.org/10.1186/s12967-020-02258-3>
- Richards S, Aziz N, Bale S et al (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med : off J Am Coll Méd Genet* 17:405–424. <https://doi.org/10.1038/gim.2015.30>
- Rivera-Muñoz EA, Milko LV, Harrison SM et al (2018) ClinGen variant curation expert panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. *Hum Mutat* 39:1614–1622. <https://doi.org/10.1002/humu.23645>
- Romero R, de la Fuente L, Pozo-Valero MD et al (2022) An evaluation of pipelines for DNA variant detection can guide a reanalysis protocol to increase the diagnostic ratio of genetic diseases. *npj Genom Med*. <https://doi.org/10.1038/s41525-021-00278-6>
- Schobers G, Schieving JH, Yntema HG et al (2022) Reanalysis of exome negative patients with rare disease: a pragmatic workflow for diagnostic applications. *Genome Med* 14:66. <https://doi.org/10.1186/s13073-022-01069-z>
- Stark Z, Boughtwood T, Phillips P et al (2019) Australian genomics: a federated model for integrating genomics into healthcare. *Am J Hum Genet* 105:7–14. <https://doi.org/10.1016/j.ajhg.2019.06.003>
- Tavtigian SV, Harrison SM, Boucher KM, Biesecker LG (2020) Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. *Hum Mutat* 41:1734–1737. <https://doi.org/10.1002/humu.24088>
- Tayoun ANA, Pesaran T, DiStefano MT et al (2018) Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* 39:1517–1524. <https://doi.org/10.1002/humu.23626>
- Vintschger E, Kraemer D, Joset P et al (2023) Challenges for the implementation of next generation sequencing-based expanded carrier screening: Lessons learned from the ciliopathies. *Eur J Hum Genet* 31:953–961. <https://doi.org/10.1038/s41431-022-01267-8>
- Wilcox E, Harrison SM, Lockhart E et al (2021) Creation of an expert curated variant list for clinical genomic test development and validation. *J Mol Diagn* 23:1500–1505. <https://doi.org/10.1016/j.jmoldx.2021.07.018>
- Wong E, Bertin N, Hebrard M et al (2023) The Singapore national precision medicine strategy. *Nat Genet* 55:178–186. <https://doi.org/10.1038/s41588-022-01274-x>
- Xiang J, Peng J, Baxter S, Peng Z (2020) AutoPVS1: an automatic classification tool for PVS1 interpretation of null variants. *Hum Mutat* 41:1488–1498. <https://doi.org/10.1002/humu.24051>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.