The implementation of genome sequencing in rare genetic diseases diagnosis: a pilot study from the Hong Kong genome project

Wai Kei Jacky Lam,[a,b,w] Chak Sing Lau,[c,w] Ho Ming Luk,[d,w] Lisa Wing Chi Au,[e] Gary C. P. Chan,[e] Will Yap Hang Chan,[f,g] Shirley Sze Wing Cheng,[d] Timothy Hua Tse Cheng,[h] Li Li Cheung,[e] Yiu Fai Cheung,[i,j] Josephine Shuk Ching Chong,[k] Annie Tsz Wai Chu,[l] Claudia Ching Yan Chung,[l] Kin Lai Chung,[m] Cheuk Wing Fung,[i] Eva Lai Wah Fung,[k,n] Yuan Gao,[o] Stephanie Ho,[d] Shirley Pik Ying Hue,[l] Chi-Ho Lee,[c] Tsz Leung Lee,[p] Philip Hei Li,[q] Hei Man Lo,[l] Ivan Fai Man Lo,[d] Herbert Ho Fung Loong,[r,s] Becky Mingyao Ma,[t] Wei Ma,[l] Shirley Yin Yu Pang,[o] Wai-Kay Seto,[c] Steven Wai Kwan Siu,[u] Ho So,[e] Yuk Him Tam,[v] Wenshu Tang,[l] Rosanna Ming Sum Wong,[i] Desmond Yat Hin Yap,[t] Maggie Lo Yee Yau,[k,n] Brian Hon Yin Chung,[j,l,**] and Su-Vui Lo,[l,*] on behalf of Hong Kong Genome Project

[a]Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[b]Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China
[c]Department of Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
[d]Department of Clinical Genetics, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China
[e]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[f]Division of Cardiology, Department of Medicine, Queen Mary Hospital, Hong Kong Special Administrative Region, China
[g]School of Clinical Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
[h]Department of Pathology, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China
[i]Department of Paediatrics and Adolescent Medicine, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China
[j]Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
[k]Department of Paediatrics, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[l]Hong Kong Genome Institute, Hong Kong Special Administrative Region, China
[m]Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[n]Department of Paediatrics, Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[o]Division of Neurology, Department of Medicine, Queen Mary Hospital, School of Clinical Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
[p]Hospital Chief Executive Office, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China
[q]Division of Rheumatology and Clinical Immunology, Department of Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong Special Administrative Region, China
[r]Department of Clinical Oncology, The Chinese University of Hong Kong, Prince of Wales Hospital, Hong Kong Special Administrative Region, China
[s]State Key Laboratory of Translational Oncology, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

Abbreviations: aCGH, Array-based comparative genomic hybridisation; ACMG, American college of medical genetics; AD, Autosomal dominant; ADPKD, Autosomal dominant polycystic kidney disease; ALS, Amyotrophic lateral sclerosis; AR, Autosomal recessive; bp, Basepair; CI, Confidence interval; CMA, Chromosomal microarray; CNVs, Copy number variants; CoQ10, Coenzyme Q10; DNA, Deoxyribonucleic acid; EEG, Electroencephalogram; ES, Exome sequencing; FISH, Fluorescence *in situ* hybridisation; gDNA, Genomic DNA; GLUT1, Glucose transporter 1; GS, Genome sequencing; HK, Hong Kong; HKGI, Hong Kong genome institute; HKGP, Hong Kong genome project; HLA, Human leukocyte antigen; HPO, Human phenotype ontology; IGV, Integrative genomics viewer; IRB, Institutional review board; Kb, Kilobase pair; LoF, Loss-of-function; LP, Likely pathogenic; lrGS, Long read genome sequencing; MDT, Multidisciplinary team; MLPA, Multiplex ligation-dependent probe amplification; mtDNA, Mitochondrial DNA; NGS, Next-generation sequencing; NHS, National health service; ONT, Oxford nanopore technology; P, Pathogenic; PCR, Polymerase chain reaction; PKC, Paroxysmal kinesigenic choreoathetosis; Post_P, Posterior probability of pathogenicity; RDs, Rare genetic diseases; RNA, Ribonucleic acid; SCA, Spinocerebellar ataxia; SINEs, Short interspersed nuclear elements; SNP, Single nucleotide polymorphism; SNVs, Single nucleotide variants; srGS, Short read genome sequencing; STRs, Short tandem repeats; SUP, Super accuracy; SVs, Structural variants; SVI, Sequence variant interpretation; UK, United Kingdom; US, United States; UTRs, Untranslated regions; VUS, Variant of unknown significance

*Corresponding author. Hong Kong Genome Institute, 2/F, Building 20E, Hong Kong Science Park, Shatin, Hong Kong Special Administrative Region, China.
**Corresponding author. Hong Kong Genome Institute, 2/F, Building 20E, Hong Kong Science Park, Shatin, Hong Kong Special Administrative Region, China.
E-mail addresses: losv@genomics.org.hk (S.-V. Lo), bhychung@genomics.org.hk (B.H.Y. Chung).
[w]These authors have equal contribution.

[t]Division of Nephrology, Department of Medicine, Queen Mary Hospital, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China
[u]Department of Clinical Oncology, Queen Mary Hospital, Hong Kong Special Administrative Region, China
[v]Department of Paediatric Surgery, Hong Kong Children's Hospital, Hong Kong Special Administrative Region, China

The Lancet Regional Health - Western Pacific 2025;55: 101473
Published Online xxx
https://doi.org/10.1016/j.lanwpc.2025.101473

Summary

Background Genome sequencing (GS) has revolutionised the diagnostic odyssey of patients with rare genetic diseases (RDs) and accelerated large-scale genome projects globally. However, the impact of GS on patients with RDs is yet to be investigated among genome projects in Asia. The Hong Kong Genome Project (HKGP) was implemented to benefit patients and families with RDs in Hong Kong, and to increase the inclusiveness of Chinese genomic data. This study evaluated the impact of short read GS (srGS), complemented by long read GS (lrGS) in a subset, on individuals recruited in the pilot phase of the HKGP.

Methods GS was performed on a prospective cohort of patients with suspected genetic disease recruited by territory-wide referrals to the HKGP. All participants received srGS, while lrGS was applied to a subset to resolve technically challenging regions unclear from srGS and provide phasing information for potential compound heterozygous variants. A phenotypic-driven diagnostic workflow was implemented to filter and prioritise rare and likely disease-causing variants. The primary outcome was diagnostic yield. The impact on the diagnostic odyssey and clinical management was also assessed.

Findings A total of 1264 individuals from 520 families with a broad spectrum of RDs were recruited, with 94% of probands being Chinese. srGS was performed for all individuals and lrGS was performed in 21 individuals. The use of srGS achieved a molecular diagnosis in 24% (125/520) of probands, and an additional 4% (21/520) with the assistance from lrGS. Approximately one-third of the identified diagnostic variants being novel. Diagnostic yield was found to be significantly higher among adult probands compared to paediatric probands (32% vs 24%; $p = 0.025$). The diagnostic yield was significantly higher in probands without prior genetic testing (37%; $n = 185$) compared to those previously tested, including exome and genome sequencing (23%; $n = 335$) ($p = 0.001$). GS ended diagnostic odysseys with an average length of 15 years (0.5–59), and potentially impacted clinical management in 77% (113/146) of diagnosed probands.

Interpretation This population-based genome project shed light on the consideration of integrating srGS and lrGS in clinical workflows for RDs. The identification of unique and prevalent variants from Southeast Asia increased the inclusiveness of Chinese genomic data, contributing to greater representation and genomic diversity.

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

Keywords: Genome sequencing; Short read genome sequencing; Long read genome sequencing; Hong Kong genome project; Population-based genome project; Rare disease; Precision medicine; Genomic diversity

Introduction

The rapid advancement in next-generation sequencing (NGS) and bioinformatic technologies over the past decade has offered huge potential to integrate genomic medicine into routine clinical care, which has revolutionised the diagnostic odyssey of patients with genetic diseases, and accelerated the execution of large-scale genome projects globally.

Patients with rare genetic diseases (RDs) often encounter challenges due to delayed or incorrect diagnoses, lifelong disabilities, and costly treatment.[1] Traditionally, making a genetic diagnosis is challenging due to the heterogeneity nature, multisystemic involvement, and pleiotropic manifestations of each of the 8000 RDs. Many remain undiagnosed despite extensive medical evaluation and standard diagnostic testing.[1–3] Over the past decade, there has been a rapid shift towards adopting exome sequencing (ES), and increasingly short read genome sequencing (srGS), owing to their significantly higher diagnostic capabilities across diverse clinical indications and populations, ranging from a diagnostic rate of 12–63%.[4] With srGS' ability to examine 90% of the genome, it provides a platform to identify a spectrum of clinically relevant disease-causing variants in both coding and non-coding regions, and the added potential to identify copy number variants (CNVs),

Research in context

Evidence before this study

Genome sequencing (GS) has revolutionised the diagnostic odyssey of patients with rare genetic diseases (RDs). As RDs have emerged as global health priorities, large-scale Genome Projects are being implemented worldwide to facilitate evidence-based clinical implementation of genomic medicine. Few Genome Projects have reported outcomes for the RD population to date. We searched for relevant studies evaluating the outcomes of GS in national Genome Projects on the PubMed database with the search terms "genome project" or "genomes project" or "initiative" or "alliance" or "flagship" or "program" or "programme" and "rare disease" and "genome sequencing". There were no language and date restrictions. We also searched from the official websites of Genome Projects to identify additional publications. Majority of the studies focused on the Western populations, with the 100,000 Genomes Projects by Genomics England, the flagship studies by Australian Genomics, and the Rare Genomes Project by the Broad Institute of MIT and Harvard contributing the most. The impact of GS on patients with RDs is yet to be investigated among Genome Projects in Asia.

Added value of this study

In the first of such studies in Asia to report preliminary results of a RD cohort within a population-wide GS project, this prospective cohort study from the Hong Kong Genome Project demonstrated the successful application of short-read (sr) and long-read (lr) GS in achieving a molecular diagnosis for 28% of the recruited families, a diagnostic yield comparable to international population-wide Genome Projects. Notably, approximately one-third of the identified diagnostic variants were novel, highlighting the value of including diverse Asian populations in genome studies. The diagnostic capability of GS was illustrated in nearly a quarter of cases that were previously undetected by a range of conventional and next-generation sequencing genetic tests, underscoring the importance of the improved genome coverage and uniformity of srGS, enabling the identification of variants often missed by genetic analyses. Furthermore, our study highlighted the complementary benefits of lrGS in resolving genetic findings that were challenging to identify by srGS, suggesting its potential for broader clinical application. The use of srGS and lrGS not only provided closure to diagnostic odysseys that had persisted for an average of 15 years, but also potentially influenced clinical management for 77% of diagnosed patients.

Implications of all the available evidence

This prospective cohort study from the Pilot Phase of the Hong Kong Genome Project provides critical insights from one of the first large-scale GS initiatives focusing on an Asian population. The identification of unique and prevalent variants from Southeast Asia offers an important complement to genome data that has primarily been derived from European and North American cohorts, contributing to increasing Asian genomic data representation and thereby promoting global genomic diversity and equity. Findings from this population-based Genome Project shed light on the consideration of integrating srGS and lrGS in clinical workflows to enable precision medicine for patients with RDs.

structural variants (SVs), and short tandem repeats (STRs) that are otherwise challenging using current genetic testing methods. According to the recent meta-analysis done by our team, with the trend of decreasing findings of variant of unknown significance (VUS), improvements in clinical management, and lower srGS cost, srGS is expected to see broader clinical adoption, slowly superseding ES and other genetic tests.[4]

More recently, there is growing evidence to illustrate the benefits of long read genome sequencing (lrGS), alone or complementing srGS, in calling single nucleotide variants (SNVs), indels, and complex SVs, including inverted duplications and inversions, repeat expansions, and variants in difficult-to-map regions.[5–7] In addition, lrGS facilitates discovery of novel genotype-phenotype associations, and has the potential to provide variant phasing and DNA methylation changes. Latest advances in lrGS technologies have facilitated the investigation of unsolved RDs in the research setting.

Given the increasing evidence on the diagnostic and clinical benefits of srGS and lrGS globally, national Genome Projects have been launched to integrate genomic medicine into mainstream healthcare.[1,4,8,9] In particular, the government of the United Kingdom (UK) implemented the 100,000 Genomes Project in 2013, with Genomics England recently reported promising results from the pilot study, demonstrating a diagnostic rate of 25% for RDs using srGS.[10] This has led the National Health Service (NHS) England to offer srGS as part of routine clinical care for RD patients.[11,12] The project also aims to leverage lrGS to provide more accurate diagnoses and treatments for cancer patients. Evidence from Genome Projects worldwide would enhance our capability to better diagnose and manage genetic and hereditary diseases, potentially informing more evidence-based healthcare decisions. Importantly, genomic data across populations, especially those beyond Europe and North America, would contribute to generating greater representation by addressing the current gap in diversity of genomic datasets, allowing the study of population-specific genetic diseases.

With a 7.5 million population and a 92% Chinese ethnic majority in Hong Kong (HK), the Hong Kong Genome Project (HKGP) contributes representative Southern Chinese genomic data to improve global genomic diversity.[13,14] As of 2018, over 470 RDs have been identified in HK, affecting 1.5% of the population.[15] The Hong Kong Genome Institute (HKGI) was established in 2020 by the Health Bureau to spearhead the integration of genomic medicine into mainstream healthcare. HKGP, the first and largest genome project in HK, aims to sequence 40,000 to 50,000 genomes in two phases, the pilot and main phase.[8,16] The current study leveraged a large-scale genome project to evaluate the impact of genome sequencing (GS) on 1264 individuals from 520 families recruited in the HKGP pilot phase.

Through identifying a comprehensive range of causal variants, including in regions that were previously not covered by ES or conventional genetic tests, as well as assessing the utility of systematic variant prioritisation, this study also applied lrGS on a subset of patients to complement srGS to resolve technically challenging regions. More importantly, potential changes in diagnosis-predicated clinical management were assessed, which facilitated multi-disciplinary team discussions, and accelerated the decision-making process for RD patients. These insights from the pilot phase serve to provide a solid foundation for integrating GS into clinical practice to benefit RD patients and families.

Methods

Participants

Participants suspected with a genetic disease were prospectively identified and recruited across a range of medical specialities at the three Partnering Centres of HKGI (the recruitment arm of the HKGP), namely the Hong Kong Children's Hospital, The Chinese University of Hong Kong/Prince of Wales Hospital, and The University of Hong Kong/Queen Mary Hospital from July 2021 to December 2022. The recruitment criteria were deliberately inclusive, ranging from likely monogenic cause to more complex aetiology to test the broad utility of GS. Parents of the proband were invited to participate whenever feasible to allow comprehensive and accurate evaluation of the disease's mode of inheritance. Other family members of the proband might also be recruited based on the recruiting clinical practitioner's clinical judgement. Exclusion criteria include patients with a clearly recognisable condition or syndrome where genetic testing offers no additional benefit, those with a known genetic diagnosis, individuals unwilling to participate in the study, or unable to provide consent. All participants received pre-test genetic counselling and provided informed written consents following the unique three-tier consent and assent model designed by the HKGI.[8]

Detailed phenotype information, including family history and symptom onset, was collected and recorded using Human Phenotype Ontology (HPO) terms to standardise clinical data input and facilitate genomic data interpretation.[17] Information on previous or planned genetic testing of the participants was considered important; prior genetic testing results available through the Hospital Authority's Genetic and Genomic Test Directory, covering conventional and NGS genetic tests offered across all public hospitals in Hospital Authority, the Department of Health, and at private diagnostic or research laboratories, were also collected.

Ethics approvals were granted by the Central Institutional Review Board (IRB) (HKGP-2021-001, HKGP-2022-001), and IRBs of Department of Health (L/M 257/2021), Joint Chinese University of Hong Kong-New Territories East Cluster (2021.423, 2023.120), and The University of Hong Kong/Hospital Authority Hong Kong West Cluster (UW 21–413, UW 23–289).

Short read genome sequencing

HKGP primarily collected whole blood samples, with occasional buccal mucosa or saliva samples when whole blood collection was not feasible. Following genomic DNA (gDNA) extraction, PCR-free srGS was performed with the KAPA HyperPlus Kit (Kapa Biosystems Inc.) on an Illumina NovaSeq 6000 sequencer (Illumina Inc.), which generated a mean depth of 39x (range: 29–76) and a depth greater than 10x for at least 95% of the reference human genome (Supplementary methods).[18] The generated srGS data then underwent sequencing quality control checks followed by secondary and tertiary analysis using an in-house bioinformatics pipeline. In brief, srGS reads were aligned to the reference human genome build 38 (GRCh38/hg38) using BWA (version 0.7.17) with duplicates removed using Picard (version 2.27.4).[19] Family-based variant calling for chromosomes 1 to 22, the X chromosome, and the mitochondrial genome was performed using Genome Analysis Toolkit (GATK, version 4.2.6.1),[20] CNVKit (version 0.9.9),[21] Manta (version 1.6.0),[22] INSurVeyor (version 1.1.0),[23] and ExpansionHunter (version 3.1.2)[24] to identify a wide range of variant types including SNVs, small insertion or deletions (indels), CNVs, SVs and STR expansions.

Phenotypic driven diagnostic workflow

An automated analytic pipeline was implemented to filter and prioritise rare and likely disease-causing variants that can potentially explain the patient and/or the family's primary clinical indication (Supplementary Method). Initial filtering based on population allele frequency, effect of variants and mode of inheritance aims to select for rare, segregating, predicted damaging candidate variants that affects the protein-coding and splice site regions.

A phenotype-driven variant prioritisation algorithm, Exomiser (version 12.1.0), was employed to match

patient phenotypes, expressed using HPO terms, with rare pathogenic SNV and indels based on expected disease inheritance models.[25] For cases without a causative variant identified through Exomiser, additional prioritisation was performed using virtual gene panels from Genomic England PanelApp and PanelApp Australia.[26,27]

For SVs and CNVs, an in-house algorithm was developed to prioritise phenotypically relevant variants by generating a similarity score based on HPO terms. This approach was used to systematically triage high-confidence SVs/CNVs and identify potential second variants in cases where a SNV/indel had been identified in recessive disease genes. SV/CNVs with a phenotype similar rate score >0.5, overlapping with the exonic and untranslated regions (UTRs) of genes and having a predicted loss-of-function (LoF) effect were evaluated. In parallel, ExpansionHunter was utilised to detect pathogenic STR expansions at known disease-associated loci.

Variant classification

Following variant prioritisation, candidate variants in genes with well-established gene–disease associations were investigated by HKGI genome curators, who are analysts trained with knowledge of genetic diseases and specialised in tertiary analysis, using an in-house pipeline and a commercial decision support system, Congenica. Mendelian disease genes curated as "definitive" and "strong" in ClinGen[28,29] and as "Green" in PanelApps are considered as high level of evidence of disease causation. Disease-causing variants in genes with an intermediate level of evidence for gene–disease association, scored as "moderate" in Clingen and/or "Amber" in PaneApp, were considered diagnostic after consensus was reached with referring clinician. Pathogenicity of the variants were determined by the American College of Medical Genetics (ACMG) guidelines[30] and up-to-date recommendations from ClinGen sequence variant interpretation (SVI) working Group.[31] STR calls were considered as pathogenic if the expansion is beyond the pathogenic reportable threshold.

Reporting of variants

HKGP values the process of variants identification and confirmation of disease-causing variants with the synergistic input of referring clinicians and other genetic specialists. Candidate variants were reviewed with clinical geneticists and referring clinicians for reporting of variants. Multi-disciplinary team (MDT) meeting (with a panel of members consisting of clinicians, clinical geneticists, laboratory scientists, genetic counsellors, genome curators, and bioinformaticians) were conducted for variants that require further assessment of the phenotypic fit for the proband or for VUS findings that have a posterior probability of pathogenicity (Post_P) ≥0.675 according to the Bayesian classification framework.[32] With the consensus of referring clinicians, VUS were returned to participants if the variant was deemed to be likely attributable to the proband's phenotype, and/or additional tests could be performed to confirm the pathogenicity of the variant (i.e. functional tests, biochemical and imaging testing, collection of parental samples to confirm the inheritance). Reportable likely pathogenic/pathogenic (LP/P) and VUS variants were confirmed using an appropriate orthogonal method (e.g. Sanger sequencing, digital PCR, lrGS) and a GS research report capturing the genetic finding and diagnosis, and when applicable, clinical management recommendations, was issued for each family. Participants can opt in to receive age-dependent additional findings within a pre-defined list of 13 genes,[8] but these findings are not reported in this study.

Long read genome sequencing

IrGS was used to (i) determine variant phasing without parental samples, (ii) resolve exact breakpoints for SV/CNVs, (iii) precise size detection of expanded STR, and (iv) provide orthogonal confirmation of findings from srGS that could not be resolved with Sanger Sequencing.

Following manufacturer's instructions, 1 μg of extracted gDNA was used to prepare libraries with the Oxford Nanopore Technology (ONT) Ligation Sequencing Kit (SQK-LSK114) which were sequenced on PromethION P-24 device using R10.4.1 flow cell (ONT, Oxford, UK) for up to 72 h to generate at least 30x coverage per sample. Additional sequencing through replenishing of additional libraries was performed when insufficient data was collected.

Raw data (qscore ≥10) were processed using the wf-human-variation workflow (version 1.7.2). The fast5 data was basecalled using Dorado in Super Accuracy (SUP) mode and aligned to GRCh38/hg38 human reference genome using Minimap2.[33] SNV and indels, CNV, SV and STR calling were performed using Clair3,[34] QDNAseq,[35] Sniffles2,[36] and Staglr,[37] respectively, and candidate variants were manually visualised in Integrative Genomics Viewer (IGV). Variant phasing was performed with WhatsHap.[38]

Length of diagnostic odyssey

The diagnostic odyssey, from symptom onset to reaching an accurate genetic diagnosis, is often long and difficult for many RD patients. Length of diagnostic odyssey was assessed for all diagnosed cases to illustrate the impact of GS. The diagnostic odyssey was estimated using the minimum number of years for cases that provided a range.

Clinical utility

Clinical utility is defined as the percentage of individuals experiencing potential changes to clinical management following a diagnosis, which helps to accelerate the

decision-making and consensus-formulation process for all relevant stakeholders during MDT meetings. The potential change in clinical management was classified into seven categories according to Riggs et al. and the UK 100,000 Genomes Project[10,39]: (i) referral to specialist(s); (ii) indication for further diagnostic tests to evaluate possible complications; (iii) initiation or contraindication of interventional or surgical procedures; (iv) surveillance for potential future complications; (v) initiation or contraindication of medications; (vi) lifestyle changes; and (vii) clinical trial eligibility. Management implications on genetic counselling for family screening and recurrence risk were not included because they were assumed to be applicable to all types of genetic test results.

Statistical analysis

Differences of diagnostic yield between subgroups with respect to clinical and demographic characteristics were compared using chi-square test. Spearman correlation was used to test the association of number of HPO term with Exomiser's rank. Binary logistic regression was used to examine association of number of HPO terms with result of WGS. Multiple logistic regression was used to evaluate association in diagnostic rate and presence of prior genetic test. The significant level was set at $P < 0.05$ for 2 tails for all analyses. All statistical analyses were performed using STATA version 17.26.

Role of the funding source

The HKGP is a publicly funded genome sequencing initiative commissioned by the Health Bureau of the HKSAR Government. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Results

Between July 2021 and December 2022, 1264 participants (520 probands and 744 family members) with suspected genetic disease by attending clinician across a broad spectrum were enrolled in the HKGP. The characteristics of the recruited patients were summarised in Table 1. Demographic characteristics were reasonably well-balanced with no significant difference between the positive and negative cases. Among the 520 prospectively recruited probands, the proportions of male and female probands (53% vs 47%), and adult (> 18 yo at recruitment) and paediatric probands (≤18yo at recruitment) (52% vs 48%) were similar. The median (IQR range) age at time of enrolment for adult and paediatric probands was 43 years (29 years–57.5 years) and 9 years (4 years–13 years), respectively. The self-reported ethnicity and race of the probands was in line with what was expected of the general local population (Hong Kong 2021 Population Census), in which majority (94%) of the probands were Chinese. Clinical presentations were highly diverse, with the most common indication for testing being neurodevelopmental disorders (28%), neurological disorders (27%), cancers (10%) and multiple congenital anomalies (8.1%) which match the common primary medical specialty of patients with rare diseases.[40] Consanguinity was uncommon (n = 9, 2%).

Diagnostic performance analysis

GS achieved a molecular diagnosis in 146/520 probands, corresponding to a diagnostic yield of 28%. In total, 81% of the diagnoses were SNV/small indels, 12% were SV/CNV, 4% were non-coding SNV/small indels, 3% were STRs, and <1% were mitochondrial DNA (mtDNA) variants (Fig. 1). One-hundred-and-forty-one patients received a single molecular diagnosis with 79% (n = 111) having autosomal dominant (AD) (34% *de novo*), 15% (n = 21) autosomal recessive (AR) (52% compound heterozygous, 48% homozygous) and 6% (n = 8) X-linked disorders (50% *de novo*). One proband was diagnosed with a mitochondrial disorder, carrying an mtDNA variant with 85% heteroplasmy. In three of the compound heterozygous cases, the variants were confirmed to be *in trans* using lrGS. Additionally, five patients received dual diagnoses. Diagnostic variants were found in 151 genes (116 distinct), with 15 genes accounting for multiple findings: *ACVR1* (n = 2), *ATXN2* (n = 2), *BRCA2* (n = 2), *CAPN3* (n = 2), *GJB2* (n = 3), *LDLR* (n = 5), *NF1* (n = 5), *PAX2* (n = 2), *PKD1* (n = 11), *PKD2* (n = 4), *PRRT2* (n = 2), *PTPN11* (n = 3), *TARDBP* (n = 2), *TSC2* (n = 3), *ZNF462* (n = 2).

An additional 4% (22/520) of the cases with strong candidate variants that could potentially explain the proband's phenotype were classified as inconclusive and were returned to referring clinicians at the discretion of MDT meeting. This included 15 VUS with a Post_P ≥ 0.675 according to the Bayesian classification framework, two probands with a single heterozygous variant in recessive genes, and three cases with unknown phasing of two causative variants due to lack of parental samples. In addition, two diagnoses with LP/P variants that did not match the clinical indication or family history were also returned to facilitate further clinical correlation. srGS was performed for 144 singletons (28%), 70 duos (13%), 277 trios (53%), 12 larger family structures (2%), and 17 others (3%). No significant difference in diagnostic yield was found among different family structures while cases that were sequenced as larger family structure showed a trend to be higher than those that were sequenced as singletons (33.7% vs 25.7; p = 0.602) (Fig. 2a).

Diagnostic yield was found to be significantly higher among adult probands compared to paediatric probands (32% vs 24%; p = 0.025). Additionally, the diagnostic rate varies vastly across disease categories, ranging from

Characteristic	Overall (n = 520)[a]	Positive (n = 146)[b]	Inconclusive (n = 22)[b]	Negative (n = 352)[b]
Sex				
Male	276 (53%)	74 (27%)	12 (4%)	190 (69%)
Female	244 (47%)	72 (30%)	10 (4%)	162 (66%)
Age at recruitment				
≤18 y	251 (48%)	59 (24%)	11 (4%)	181 (72%)
>18 y	269 (52%)	87 (32%)	11 (4%)	171 (64%)
Median (IQR)	20 y (9 y–43 y)	23 y (11.75 y–43.5 y)	19 y (7.75 y–46.25 y)	17 y (8 y–43 y)
Self-reported ethnicity and race				
Chinese	490 (94%)	139 (28%)	19 (4%)	332 (68%)
Chinese with mixed ancestry	7 (1%)	4 (57%)	0 (0%)	3 (43%)
Filipino	2 (<1%)	1 (50%)	0 (0%)	1 (50%)
Indian	2 (<1%)	0 (0%)	0 (0%)	2 (100%)
Nepalese	1 (<1%)	0 (0%)	0 (0%)	1 (100%)
Pakistani	4 (<1%)	1 (25%)	0 (0%)	3 (75%)
Thai	1 (<1%)	0 (0%)	0 (0%)	1 (100%)
White	4 (<1%)	1 (25%)	0 (0%)	3 (75%)
Did not specify	8 (2%)	0 (0%)	2 (25%)	6 (75%)
Unknown	1 (<1%)	0 (0%)	1 (100%)	0 (0%)
Known parental consanguinity	9 (2%)	5 (56%)	2 (25%)	2 (25%)
Disease category				
Cancer	54 (10%)	21 (39%)	0 (0%)	33 (61%)
Cardiovascular	20 (4%)	7 (35%)	0 (0%)	13 (65%)
Dermatological	9 (2%)	6 (67%)	0 (0%)	3 (33%)
Endocrine	7 (1%)	2 (29%)	1 (14%)	4 (57%)
Gastrointestinal	9 (2%)	1 (11%)	1 (11%)	7 (78%)
Growth disorder	9 (2%)	2 (22%)	0 (0%)	7 (78%)
Hearing or ear disorder	4 (1%)	2 (50%)	0 (0%)	2 (50%)
Immune	10 (2%)	0 (0%)	0 (0%)	10 (100%)
Multiple congenital abnormalities	42 (8%)	9 (21%)	2 (5%)	31 (74%)
Metabolic	8 (2%)	3 (38%)	2 (25%)	3 (38%)
Musculoskeletal	18 (3%)	3 (17%)	1 (6%)	14 (78%)
Neurodevelopmental	143 (28%)	39 (27%)	5 (3%)	99 (69%)
Neurological	140 (27%)	26 (19%)	10 (7%)	104 (74%)
Ophthalmological	11 (2%)	7 (64%)	0 (0%)	4 (36%)
Renal	31 (6%)	16 (52%)	0 (0%)	15 (48%)
Respiratory	1 (<1%)	1 (100%)	0 (0%)	0 (0%)
Urogenital	4 (1%)	1 (25%)	0 (0%)	3 (75%)

% may not sum to 100 due to rounding. [a]% is expressed as the number of cases/total 520 cases. [b]% is expressed as the number of positive/inconclusive/negative cases over the overall number of cases in the same row.

Table 1: **Demographic characteristics of the 520 probands.**

0% to 100% (Fig. 2b). Diagnostic yield among probands with dermatological disorder (67%), ophthalmological disorder (64%), and renal disorder (52%) was significantly higher than those with other disorders (p = 0.009, p = 0.008, p = 0.003, respectively).

A total of 335 (64%) probands had previously undergone genetic testing, with a median of two prior tests done per proband (range 1–7). A total of 41 (8%) probands had previously undergone karyotyping, 105 (20%) chromosomal microarray (CMA) (array-based comparative genomic hybridisation (aCGH)/single nucleotide polymorphism (SNP) array), 134 (26%) targeted single-gene testing using polymerase chain reaction (PCR) or Sanger sequencing, 53 (10%) multiplex ligation-dependent probe amplification (MLPA), four (0.8%) fluorescence *in situ* hybridisation (FISH), three (0.6%) methylation studies, 20 (4%) pharmacogenetic testing, 100 (19%) NGS single gene/gene panel testing, 25 (5%) mitochondrial DNA/RNA, 169 (33%) ES (clinical/medical/whole), and 16 (3%) GS.

The overall diagnostic yield among probands who had no previous genetic testing was 37%, significantly higher than those who had been tested at least once (23%) (p = 0.001). The diagnostic yield provided by GS

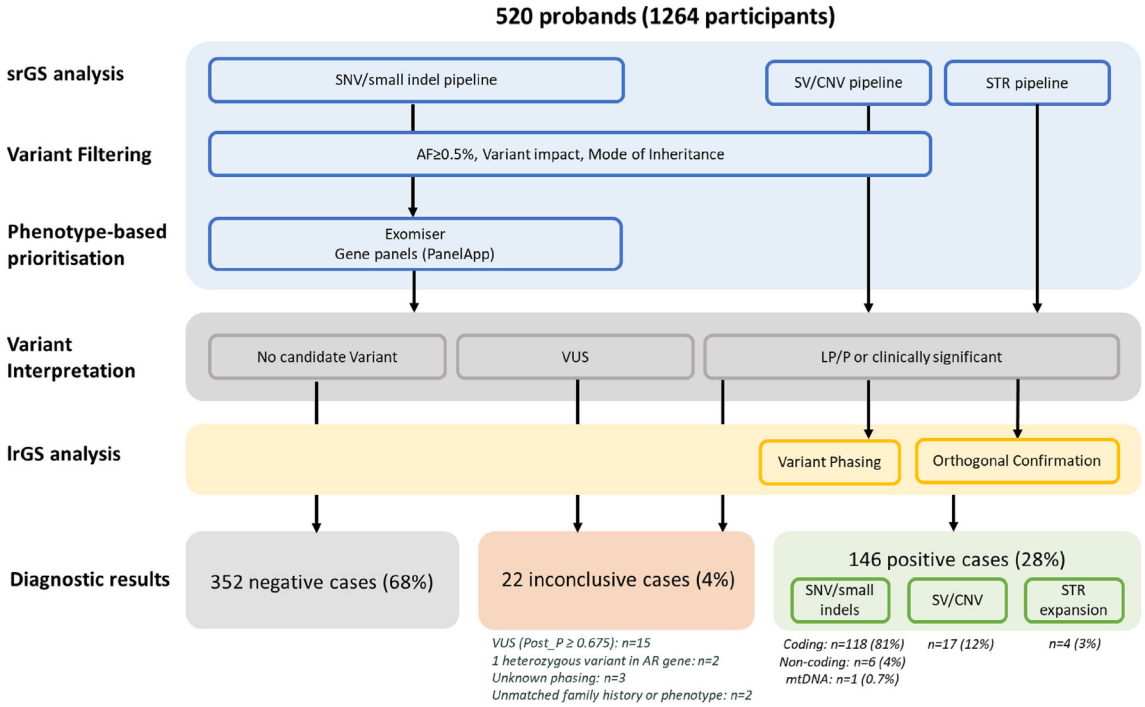

***Fig. 1:* GS findings based on automated diagnostic pipelines**. Our diagnostic pipeline, which integrated srGS and lrGS, along with robust variant filtering and prioritisation workflows, achieved a positive diagnosis in 28% of the probands. Additionally, 4% of the cases harboured strong candidate variants that were classified as inconclusive and returned to the clinicians for further clinical correlation and review. The remaining 68% of the probands received negative results. CNV copy number variant; LP likely pathogenic; lrGS long read genome sequencing; P pathogenic; SNV single nucleotide variant; srGS short read genome sequencing; STR short tandem repeat; SV structural variant; VUS variant of unknown significance.

varied between 18% and 32%, depending on the type of previous genetic test being performed (Fig. 2c). For the entire cohort, the diagnostic rate among adult probands (32%) is higher than paediatric probands (24%). After adjusting for previous genetic testing in a multivariate logistic regression, the diagnostic rates between paediatric and adult probands were statistically similar (OR = 1.271, 95% CI = 0.836–1.937, p = 0.262).

Among probands without prior ES/GS previously (n = 345), the diagnostic rate was 33%. For probands who had previously undergone clinical/medical/whole ES or GS but were undiagnosed (n = 175), GS achieved a molecular diagnosis in 31 probands, corresponding to a diagnostic rate of 18%. The variants were missed by clinical/medical/whole ES or GS previously due to a variety of reasons: methodological limitation in detecting SVs, STRs, and non-coding variants (39%; 12/31), high GC region/low genome coverage (16%; 5/31), recently discovered gene–disease associations (16%; 5/31), updated variant annotation (3%; 1/31), analysis with additional family members (3%; 1/31), and missed by previous tertiary analysis pipeline (3%; 1/31). The remaining seven cases had unknown reasons.

Variants identified and reported

Among 520 probands, 193 variants were reported, including 173 LP/P variants and 20 VUS. The majority were SNVs (64%, 123/193) and indels (25%, 49/193), while deletions (8%, 15/193) and duplications, complex SVs and STR expansions were rare (3% 6/193). Despite the lower coverage of srGS, one mosaic variant with a 24% variant allelic frequency and one mtDNA variant with 85% heteroplasmy level were called, which were further confirmed by visual inspection of IGV and Sanger sequencing (Supplementary Table S1).

When comparing the pathogenicity of the different variant types, deletions, duplications, complex SVs, STR expansion were all classified as LP/P as these molecular findings either disrupt coding regions of the genes or are known disease-causing loci. In addition, 62 unique variants (54 LP/P and 8 VUS) were novel when compared against public variant databases such as ClinVar,[41] LoVD[42] and literature.

PCR-free srGS provides uniform genome coverage, allowing the detection of SVs, STRs, rare intronic and spice region variants missed by prior conventional and NGS methods. A total of 17 SVs were reported, including 15 deletions, one duplication and one

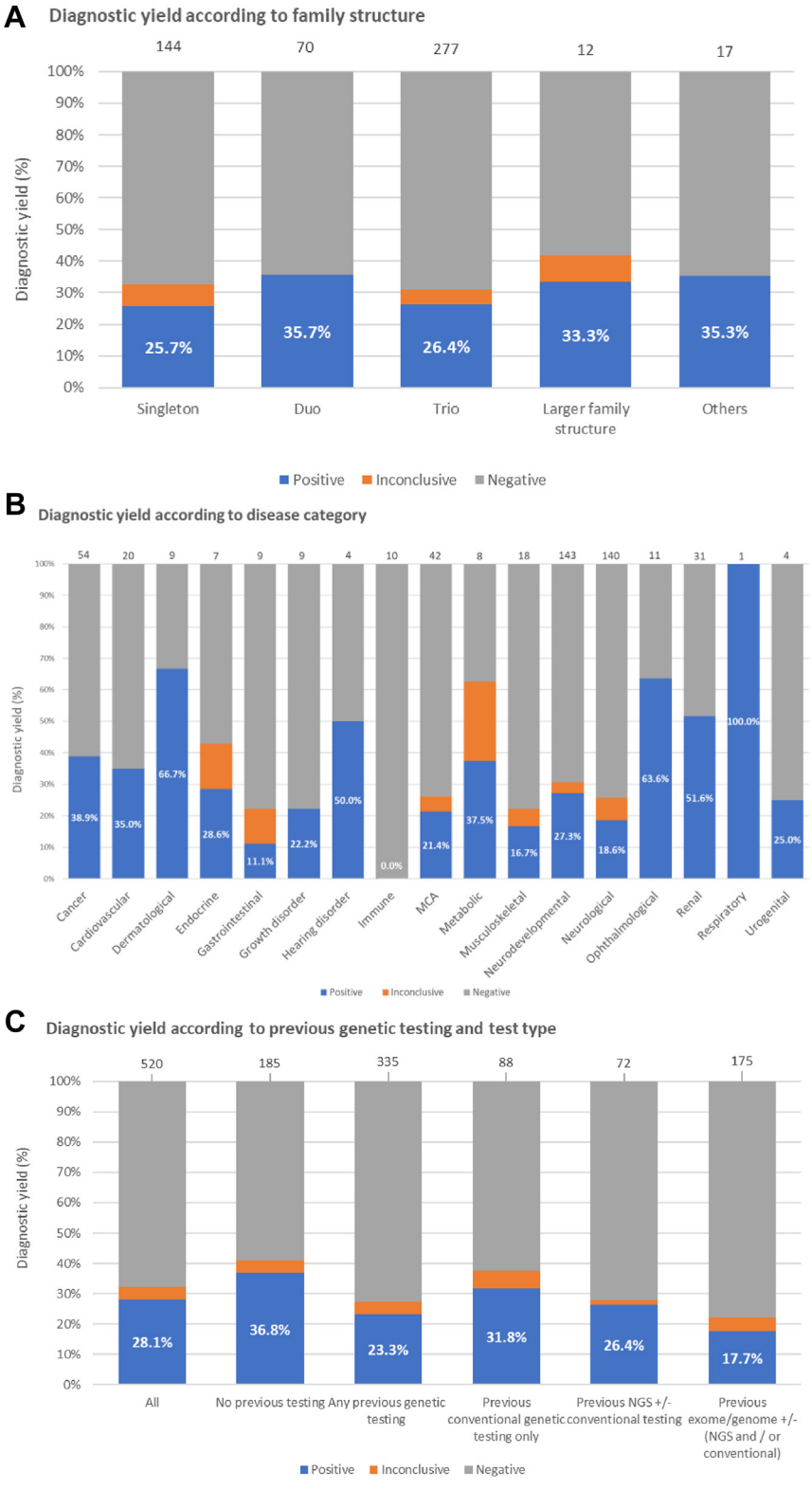

***Fig. 2*: Diagnostic yield by family structure, disease category, and previous genetic testing and test type**. Fig. 2A shows the diagnostic yield according to family structure. Different family structures were recruited based on the indication for testing. A "singleton" refers to a proband, or

translocation. Of which, 11 SVs were smaller than the standard CMA resolution (200 kb with Affymetrix Chromosome Analysis Suite) and one translocation was undetected by CMA. Most SVs (94%; 16/17) were novel, except for a duplication overlapping the well-reported 17p12 recurrent region. The breakpoints of the identified SVs were located in the intronic regions. Over half (53%; 9/17) of the SVs breakpoints could not be precisely determined by srGS due to overlap with known repetitive elements, including Alu repeat regions (32%, 11/34 breakpoints) and long terminal repeat retrotransposons (3%, 1/34 breakpoints). Six coding exon-deleting deletions were found in five AR cases (22%) with a prior LP/P SNV or indel detected from NGS gene panel or ES testing (Supplementary Results).

Four STRs were identified in *ATXN2* (n = 2), *CACNA1A* (n = 1) and *PPP2R2B* (n = 1), which are associated with spinocerebellar ataxia (SCA) 2 (OMIM: 183090), SCA 6 (OMIM: 183086) and SCA 12 (OMIM: 604326), respectively. Six intronic variants were identified including four indels (range 4–19 bp) that span across the splicing regions and disrupt the canonical donor or acceptor splice sites. Negative NGS gene panel or ES results were previously returned for three out of these six cases (50%).

The efficiency of phenotypic driven diagnostic workflow

Efficient diagnostic pipelines are critical for analysing the millions of variants typically returned from GS. These pipelines need to filter out unlikely variants and prioritise likely disease-causing variants to reduce the number to a few plausible candidates, accelerating interpretation and reporting.

In this study, 90% of diagnoses (132/146) with disease-causing SNVs and indels were made using a combination of Exomiser for phenotype-based variant prioritisation (107/132; 81%) and custom virtual gene panels from PanelApp (25/132; 19%). Exomiser proves to be an effective filtering and prioritisation tool, with 67%, 77%, 82%, and 83% of the 132 diagnoses being ranked as the top 1, top 3, top 5, and top 10 candidates, respectively (Supplementary Figure S1).

HPO terms were collected for each proband to standardise phenotypic data and facilitate Exomiser prioritisation. A median of nine HPO terms (range 1–40), were recorded for each proband. Although having more HPO terms per proband slightly increased the chances of a diagnostic variant being ranked top by Exomiser, the improvement was not statistically significant (p = 0.823) and annotating with more than five HPO terms only led to very minor performance gains. Interestingly, there is a negative association between the number of HPO terms and the diagnostic rate (binary logistic regression, OR = 0.964, 95% CI = 0.932–0.996, p = 0.029), suggesting that more HPO terms does not necessarily lead to higher diagnostic success. Further, sequencing additional family members did not substantially improve diagnostic yield compared to singleton cases. These findings underscore the importance of comprehensive yet accurate phenotyping for effective variant prioritisation (Supplementary Figure S2).

A minority of diagnostic variants were not prioritised by Exomiser due to factors like recently reported gene–disease associations (1%), partially explained phenotype (2%), high minor allele frequencies from gnomAD (5%), or incomplete penetrance (7%). However, these variants were identified alternatively via gene-panel analysis, suggesting the use of two different phenotype-based prioritisation tools were able to complement each other and make 100% of the 132 diagnoses.

Complementing srGS with lrGS

Many studies have demonstrated the benefits of combining long-read sequencing approaches and srGS. For the HKGP, srGS identified reportable SV/CNV and STRs, and lrGS was used to confirm the expected findings in 21 individuals. lrGS was able to provide phasing information to determine the compound heterozygous cases (n = 3), resolve inconclusive finding in 2of the 3 cases, accurately deduce the repeat size within STR expansion regions (n = 4) and confirm the exact size of SVs (n = 2) (Fig. 3).

For adult cases where parental samples were unavailable, lrGS provided the long-range information and resolved the phasing of alleles for AR disease diagnosis. In a 70-year-old female presented with limb girdle dystrophy (HKGP000276-1), two heterozygous pathogenic variants (c.2120A > G p.(Asp707Gly) and c.2305C > T p.(Arg769Trp)), located 493 bp apart in the *CAPN3* gene, were identified by srGS, but the phasing could not be determined. lrGS recovered multiple reads spanning across both variant positions, confirming the variants were *in trans* and established the diagnosis. Similarly, in

affected individual, for whom no other family members were recruited for the study. A "duo" consists of a parent–proband pair, while a "trio" includes both parents and the proband. Probands recruited with parents and one or more siblings were grouped into "larger family structures". Any cases that could not be classified into these defined family group categories were considered "others". Fig. 2B shows the diagnostic yield according to disease category, which were categorised according to the most prominent phenotypes. Fig. 2C shows the diagnostic yield according to previous genetic testing: conventional (karyotyping, array-based comparative genomic hybridisation, single nucleotide polymorphism array, targeted single-gene testing, multiplex ligation-dependent probe amplification, fluorescence *in situ* hybridisation, methylation, pharmacogenetic testing), NGS (single gene or targeted panels, mitochondrial DNA/RNA), exome (clinical, medical, whole) and genome. MCA multiple congenital anomalies; NGS next generation sequencing. *The values above the bars are the total number of probands included in that category.

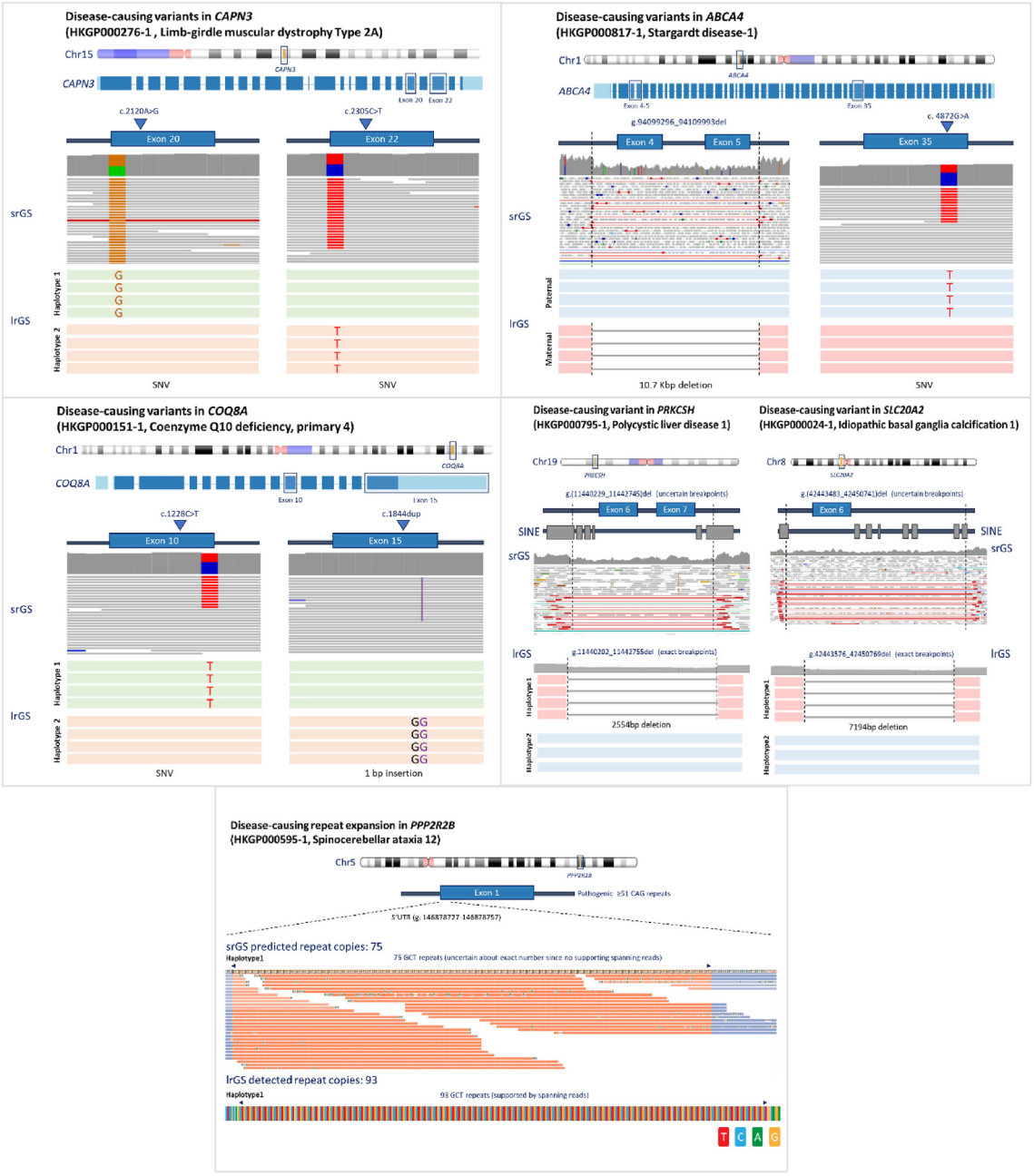

***Fig. 3:* Utility of lrGS: case illustrations on variant phasing**. The first three cases demonstrated the value of lrGS in determining the phasing of variants within the same gene, which could not be deciphered using srGS alone due to the lack of parental samples (top panel). lrGS was able to provide long-range information necessary to confirm the variants to be in trans, leading to the molecular diagnosis (bottom panel). The utility of lrGS in resolving the breakpoints located in repetitive regions (SINEs; top panel) helped to confirm the correct size of the SV (bottom panel) in the fourth and fifth case. In the final case, srGS was unable to correctly predict the repeat size due to its limitation in resolving large repeat expansions. This finding was then accurately confirmed through the use of lrGS. lrGS: long read genome sequencing; SNV: single nucleotide variant.

a 33-year-old individual clinically diagnosed with Stargardt's disease (HKGP000817-1), lrGS phased a SV (g.(94099300_94109997)del) and SNV (c.4872G > A p.(Trp1624Ter) that were 80 kb apart in the *ABCA4* gene, determining they were in trans. These cases demonstrated the ability of lrGS in resolving the phasing of both SNVs and SVs.

Moreover, lrGS helped refine clinical management for patients. In a 22-year-old male presented with cerebellar atrophy (HKGP000151-1), lrGS confirmed

biallelic LP/P *COQ8A* variants, c.1228C > T p.(Arg410Ter) and c.1844dup p.(Ser616LeufsTer114), which supported the diagnosis of primary coenzyme Q10 (CoQ10) deficiency and provided potential treatment options such as CoQ10 supplementation.

In four cases, lrGS also provided more precise sizing and phasing of the pathogenic repeat expansions detected by srGS. In particular, it determined the exact expanded allele size in the *PPP2R2B* gene in patient HKGP000595-1, which was underestimated by srGS. Due to the limited read length of srGS, ExpansionHunter predicted an expansion containing 75 repeats (95% CI: 66–105), but was confirmed to be 93 repeats using lrGS.

Lastly, lrGS clarified the boundaries and confirmed the exact size of the SVs in the *PRKCSH* and *SLC20A2* genes in HKGP000795-1 and HKGP000024-1 respectively, which were ambiguous by short-read data due to repetitive sequences at the breakpoints.

The impact of MDT meeting

A total of 80 cases were discussed at 17 MDT meetings, which facilitated the evaluation of the genetic findings with the context of clinical phenotype and variant pathogenicity (Supplementary Results). With further phenotypes supplemented by the referring clinician, two cases were agreed by MDT members to be compatible with patient's phenotype and specific to the disease, thereby enabling the reclassification of variant to LP and substantiating a new molecular diagnosis. Extensive discussion on variant-and-phenotype matching with the support of extra phenotypic information at MDT meetings, such as immunohistochemistry of muscle biopsy, plasma/cerebrospinal fluid amino acid profile and plasma cholesterol profile, also led to an additional four clinically diagnosed cases. Additional investigations were suggested to potentially upgrade the variant to LP or reportable VUS, and for further clinical correlation. With the additional four clinically diagnosed cases, the diagnostic rate can be uplifted by 1% (150/520; 29%). On the other hand, 12 cases were not reported due to incompatible phenotype or family history (n = 7) or were compatible with phenotype but were considered to have minimal clinical actionability (n = 5). These findings demonstrate the value and effectiveness of pooling diverse expertise through MDT meetings to facilitate consensus-based decision-making.

Diagnosis-predicated implications on diagnostic odyssey

The application of srGS, complemented by lrGS in some cases, has ended prolonged diagnostic trajectories for the patients and families (Fig. 4, Panel 1, Supplementary Results). Among the 146 probands receiving a positive molecular diagnosis, GS ended diagnostic odysseys with an average length of 15 years (0.5–59).

Potential changes in clinical management

Among the 146 positive molecular diagnoses, GS could potentially impact clinical management in 113 (77%) probands (Panel 2). GS would most likely aid management in the areas of surveillance (n = 84; 58%), followed by indication or contraindication of medications (n = 79; 54%), lifestyle changes (n = 46; 32%), indication or contraindication of procedures (n = 41; 28%), clinical trial eligibility (n = 21; 14%), specialist referral (n = 17; 12%), and further diagnostic testing (n = 1; 1%) (Supplementary Figure S3). Implications on clinical management could be increased to 100% if genetic counselling for family screening and recurrence risk were also included. Assessment in the potential changes in diagnosis-predicated clinical management helped to accelerate discussions and the decision-making process for the referring clinicians during MDT meetings.

Discussion

In the past decade, significant global and concerted efforts have been made to accelerate clinical implementation of genomics in healthcare. Few have reported outcomes to date, and to the best of knowledge, this

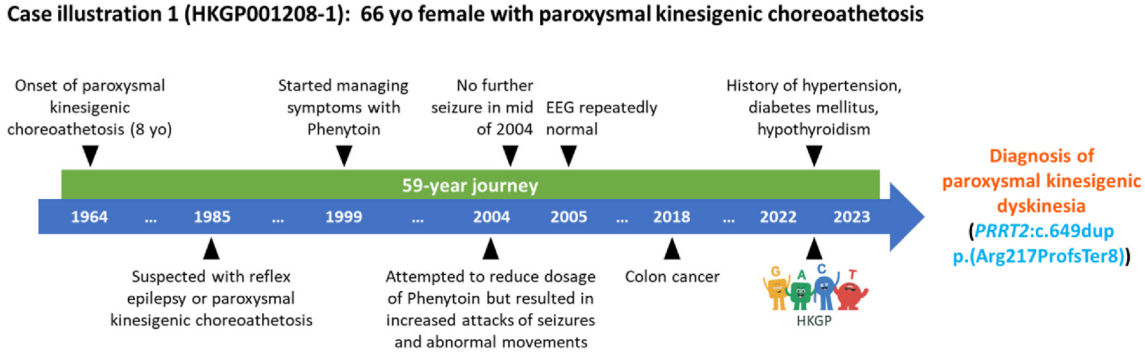

Fig. 4: **Ending prolonged diagnostic odysseys: case illustration.**

study serves to be the first in Asia to report preliminary results of a RD cohort in a population-wide GS project.

This study demonstrated the impact of GS in 1264 predominantly Chinese paediatric and adult participants across a broad spectrum of RDs. The use of srGS complemented from lrGS identified causal variants in 28% of the probands, with approximately one-third being novel. Lack of diversity in genomic datasets has been recognised, with the majority of variants catalogued in international genomic databases biased towards White individuals.[1,4,43,44] The identification of unique and prevalent variants from Southeast Asia in this study contributes to genomic diversity by increasing the inclusiveness of Chinese genomic data. This helps to expand the mutation spectrum and the clinical characterisation of phenotypes that are distinctively or more frequently observed in Chinese disease cohorts, potentially delineating genotype-phenotype correlations in Chinese patients. This study identified two known Chinese founder variants in the *TARDBP* and *GJB2* genes. In particular, the c.892G > A variant identified in *TARDBP* gene is a founder mutation in East Asia that was most found in Guangdong and GuangXi, China.[45,46] It is associated with a rapid disease progression in amyotrophic lateral sclerosis (ALS) Chinese patients. The identification of this founder mutation is critical in informing disease prognosis and family reproduction planning for ALS. Additionally, two hotspot and three recurrent variants exclusively reported in Chinese patients were identified, potentially representing founder pathogenic variants. Forty percent of the reportable VUS in this study were novel, which may have the potential to be reclassified when further cases could be confirmed in a wider population of the same ethnicity. This highlights the indispensable role of the HKGP in identifying novel and unique disease variants relevant to the Chinese population.

The diagnostic rate of 28% in this study is comparable to international studies, including preliminary findings from the UK 100,000 Genomes Project (25%).[4,10] Interestingly, as contrasted with other cohort studies, this study demonstrated a significantly higher diagnostic rate among adult probands (32%) than paediatric probands (24%),[4,10] potentially due to the presence or absence of prior genetic testing, which was found to be a major factor affecting the diagnostic yield (23% vs 37%; $p = 0.001$). This may also explain why the diagnostic yield of trios in this study showed no significant difference with that of singleton (26.4% vs 25.7%; $p = 0.884$) since a larger portion of trios received prior NGS testing compared to that of singletons (43.0% vs 17.4%). Traditionally, children are more heavily investigated as 50–70% of the RDs have a paediatric onset.[47,48] Technological advancement and the growing knowledge of the field have also allowed higher adoption of genetic testing in recent years. In fact, 83% of the paediatric probands in this cohort had undergone at least one genetic test previously, and over half had undergone ES/GS. In contrast, 53% of the adult probands had not undergone any previous genetic test, potentially due to the limited availability of genetic tests under the public healthcare system in HK in earlier years. Importantly, after adjusting for previous genetic testing, the diagnostic rates between paediatric and adult probands were statistically similar, illustrating the similar diagnostic capability of GS across age groups, and

Panel 1. **Ending a prolonged diagnostic odyssey for a patient with paroxysmal kinesigenic dyskinesia**

A 66-year-old woman (HKGP001208-1) had the longest diagnostic odyssey in this cohort. She experienced paroxysmal kinesigenic choreoathetosis (PKC) since eight-year-old, with seizures and dystonia attacks triggered by noise, including being called by name. Her younger brother, daughter and son were also affected by PKC. She was initially suspected with reflex epilepsy or PKC, but repeated electroencephalogram (EEG) showed normal results. She was managed by the anti-epileptic drug Phenytoin all along. However, at 49-year-old, an attempt to reduce Phenytoin dosage resulted in abnormal movements and increased attacks. She was then referred to the HKGP, where srGS revealed a heterozygous pathogenic variant in the PRRT2 gene (c.649dup p.(Arg217ProfsTer8)), leading to a diagnosis of AD paroxysmal kinesigenic dyskinesia (OMIM: 128200). The molecular diagnosis ended her 59-year diagnostic odyssey and enabled cascade testing to her children and at-risk relatives, which is important for disease management and reproductive planning.

Panel 2. **Personalised treatment for a patient with GLUT1 deficiency syndrome**

A 24-year-old female patient (HKGP000046-1) with onset of epilepsy at six-month-old and paroxysmal dystonia and dyskinesia at seven-year-old underwent extensive diagnostic tests over the years, including nerve conduction studies, electromyogram, EEG, and brain magnetic resonance imaging, but failed to reach a conclusive diagnosis. She had been managed with various anti-epileptic drugs, including valproate sodium, carbamazepine, and lamotrigine, without success. She was referred to the HKGP, and srGS revealed a heterozygous pathogenic variant in the SLC2A1 gene, confirming a diagnosis of AD glucose transporter 1 (GLUT1) deficiency syndrome 2 with childhood onset (OMIM: 612126). This ended the patient's 18-year diagnostic odyssey, and enabled the clinician to refine clinical management for the patient. In particular, she was recommended a ketogenic diet, which is highly effective in controlling seizures and improving gait disturbance.[56–58] Anti-epileptic drugs including barbiturates and valproic acid were reported to be generally ineffective in treating GLUT1 deficiency syndrome and should be contraindicated.[56] The diagnosis allowed the contraindication of lamotrigine, saving a minimum annual cost of HK$5505 (US$706) based on the patient's previous prescription history, and potentially saving a lifetime cost of HK$319,476 (US$40,958) given patients' normal lifespan and the life expectancy of 87.2 years in HK in 2022.[57,59] Moreover, if srGS was offered earlier, it could have avoided the prescription of valproate sodium for nine years and carbamazepine for one year, saving an additional cost of over HK$4282 (US$549). With the confirmed diagnosis, the patient's younger brother and four-year-old son, also suspected of having dystonia, may also benefit from cascade genetic testing and proactive disease management.

suggesting the equitable use of GS in both paediatric and adult patients.

GS examines all exons and 90% of the genome, which provides a more uniform genome coverage across the nuclear and mitochondrial genomes, enabling comprehensive detection of variants that are often intractable to ES and conventional genetic tests.[4] The diagnostic capability of GS was illustrated in 23% of cases that were previously undetected by a range of conventional and NGS genetic tests. Notably, srGS identified diagnoses in 18% (31/175) of cases that remained undiagnosed by previous ES/GS. Consistent with prior evidence, these included disease-causing variants that were missed by ES such as deep intronic, non-coding, and SVs.[49,50] In the present study, srGS identified two indels spanning across splice junctions and eight SVs with breakpoints residing in the intronic regions that have evaded ES detection. srGS also uncovered six deletions and one intronic variant as the "second hit" in AR cases where only a single heterozygous pathogenic SNV has been detected. For instance, a second LP intronic variant in *CNGA3* gene was identified in a 34-year-old female affected with Achromatopsia 2 (OMIM: 216900), where prior NGS panel testing only detected a pathogenic indel in the coding region. The value of srGS was further demonstrated by uncovering an additional genetic diagnosis for a one-year-old patient with multiple phenotypes. Despite ES identifying a pathogenic *GLI2* variant associated with Culler-Jones syndrome (OMIM: 615849), the more severe phenotypes could not be explained. Application of srGS through HKGP revealed another homozygous 1.5 kb deletion in the *FANCA* gene, which is associated with Fanconi anaemia complementation group A (OMIM: 227650). This highlights the diagnostic capability of srGS in resolving complicated cases, such as patients with "one-man-two-diseases". The diagnosis necessitated frequent monitoring of complete blood count for bone marrow failure and early surveillance of solid tumours. More importantly, it could improve the patient's prognosis through early initiation of pre-bone marrow transplant workup by assessing the high-resolution human leukocyte antigen (HLA)-typing of the patient and immediate family members for hematopoietic stem cell transplantation. These findings underscore the importance of the improved genome coverage and uniformity of srGS, which allows for the interrogation of variants often overlooked by other genetic analyses.

In recent years, lrGS has improved the diagnosis and clinical utility by resolving variants that were unclear from srGS. In this study, the clinical benefits of lrGS were particularly evident when phasing information from srGS was unavailable. For instance, lrGS confirmed two pathogenic variants in *ABCA4* on different haplotypes, ending a 21-year diagnostic odyssey and allowing the patient to become eligible for clinical trials for Stargardt disease 1 (OMIM: 248200).[51] Similarly, lrGS could refine treatment choices, as illustrated by a patient carrying two LP/P variants in *COQ8A*, enabling exploration of CoQ10 supplementation. Furthermore, the importance of lrGS was illustrated in patients with genetic findings that could not be precisely resolved by srGS, such as confirming the presence and exact breakpoints of CNVs, which is crucial for predicting the potential frameshift consequences. These findings reinforced the diagnostic value of lrGS complementing srGS, supporting previous studies.[52,53] Further development and widespread application of lrGS in the clinical setting are expected to provide significant advantages for patients with unsolved RDs.

An accurate molecular diagnosis is crucial for optimal clinical management and outcomes.[1] Potential changes in diagnosis-predicated clinical management were observed in 77% of probands in this study, facilitating clinical decision-making among managing clinicians. Importantly, genetic diagnoses could potentially save costs by avoiding unnecessary procedures and treatment, such as in the case of *SLC2A1*. Notably, the significance of early genetic diagnosis has been highlighted in previous studies, illustrating its socio-economic impact in reducing RD costs and risk of financial hardship.[54,55] These results emphasize the need of a formal health economics study to precisely calculate the economic benefits of HKGP on its participants. Beyond improved diagnostic and clinical outcomes, shortening of diagnostic odysseys is increasingly valued. The application of srGS and lrGS has ended diagnostic trajectories lasting up to 60 years in this cohort. Over time, the diagnostic pathway will focus on simplifying the diagnostic evaluation process, with a single test like srGS potentially superseding a succession of diagnostic tests to reduce time-to-diagnosis and healthcare inequities, with lrGS complementing or replacing this as costs reduce over time.

The findings from the HKGP pilot study support the use of srGS with complementation from lrGS in the diagnosis of RDs in HK. Limitations were acknowledged. Firstly, lrGS was used to complement srGS in specific cases only. When to utilise lrGS in a clinical diagnostic pathway remains unclear. The frequent change of pipelines and the identification of unchartered regions not covered by srGS and ES also pose challenges for interpretation. Further benchmarking efforts are needed to standardise the tools before more widespread clinical application. Secondly, potential changes in clinical management based on the genetic diagnosis were assessed to facilitate discussions during MDT meetings, but actual changes were not evaluated for the entire cohort. Long-term monitoring on patient's healthcare and utilisation records is needed to fully evaluate the clinical utility of srGS and lrGS.

Conclusion

This study demonstrates the pioneering effort of the HKGP pilot phase in providing empirical evidence of the capabilities of GS, and in building a solid foundation for the main phase of the project. As the first study in Asia to report findings from a population-based genome project for RDs, this study contributes significant and representative Chinese genomic data to improve global genomic diversity and equity. The use of srGS and complementary lrGS achieved a diagnostic rate of 28%, comparable to other international population-wide GS projects. These findings shed light on the consideration of integrating srGS and lrGS in the clinical workflows for patients with RDs in other health systems.

Contributors

HKGP, ATWC, BHYC and SVL contributed to the conception and design of the study. All authors carried out acquisition, analysis, or interpretation of data. CCYC and HMLo conducted the statistical analyses and were involved in data organisation and presentation. ATWC, CCYC, and HMLo drafted the manuscript. All authors critically reviewed the manuscript with suggestions for improvement and revision. HKGP and SVL obtained funding. ATWC, BHYC and SVL oversaw and supervised the project. All authors contributed to the overall data interpretation, reviewed, and approved the final draft for submission.

Data sharing statement

Upon reasonable request, the study protocol, informed consent forms, and individual participant data reported in this article, after de-identification, will be made available to investigators whose proposed use of the data has been approved by an independent review committee. Data will be available from the corresponding authors up to five years following publication.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests

Dr Wai Kei Jacky Lam is a Director of DRA. Prof Herbert Ho Fung Loong reported receiving research funding outside the submitted work from MSD, Mundipharma, and Novartis; travel support for Bayer, Boehringer-Ingelheim, MSD, Novartis, Pfizer; serving as a speaker's bureau for AbbVie, Bayer, Eisai, Eli-Lilly, Guardant Health, Novartis; providing advice to Boehringer-Ingelheim, Celgene, Eli-Lilly, Illumina, Novartis, Merck Sereno, Takeda, George Clinical; and serving as a member for Pharmacy and Poisons (Registration of Pharmaceutical Products and Substances: Certification of Clinical Trial/Medicinal Test) Committee, Pharmacy & Poisons Board of Hong Kong. Dr Desmond Yat Hin Yap reported lectures honorarium from GlaxosmithKline, Boerhinger Ingelheim, and AstraZeneca; and serving as Honorary Secretary for the Hong Kong Society of Nephrology and Honorary Treasurer for the Asian Pacific Society of Nephrology. No other disclosures were reported.

Acknowledgements

The authors would like to thank all members of the HKGI for their efforts in preparing the launch of the HKGP. We thank all patients and family members for participating in the Project. We extend our heartfelt appreciation to the healthcare and recruitment teams at the Partnering Centres of HKGI – The University of Hong Kong/Queen Mary Hospital, The Chinese University of Hong Kong/Prince of Wales Hospital, and Hong Kong Children's Hospital – for their efforts in recruiting and managing the patients and families. The authors also wish to acknowledge members of the HKGI Scientific Team and Bioinformatics Team for their role in supporting sample sequencing, development of bioinformatics pipelines, and data curation, analysis, and reporting process. Finally, the authors would like to express their gratitude to the HKGI Board of Directors and Advisory Committees for their continued support and guidance, as well as to the key HKGP stakeholders – The Health Bureau, Hospital Authority, and Department of Health – for their overall coordination and financial support of the Project. The HKGP is a publicly funded genome sequencing initiative commissioned by the Health Bureau of the HKSAR Government. The sponsor had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2025.101473.

References

1 Chung CCY, Hong Kong Genome Project, Chu ATW, Chung BHY. Rare disease emerging as a global public health priority. *Front Public Health*. 2022;10:1028545.
2 Boycott KM, Rath A, Chong JX, et al. International cooperation to enable the diagnosis of all rare genetic diseases. *Am J Hum Genet*. 2017;100(5):695–705.
3 Splinter K, Adams DR, Bacino CA, et al. Effect of genetic diagnosis on patients with previously undiagnosed disease. *N Engl J Med*. 2018;379(22):2131–2139.
4 Chung CCY, Hue SPY, Ng NYT, Doong PHL, Chu ATW, Chung BHY. Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. *Genet Med*. 2023;25(9):100896.
5 Wojcik MH, Reuter CM, Marwaha S, et al. Beyond the exome: what's next in diagnostic testing for Mendelian conditions. *Am J Hum Genet*. 2023;110(8):1229–1248.
6 Sanchis-Juan A, Megy K, Stephens J, et al. Genome sequencing and comprehensive rare-variant analysis of 465 families with neurodevelopmental disorders. *Am J Hum Genet*. 2023;110(8):1343–1355.
7 Cohen ASA, Farrow EG, Abdelmoity AT, et al. Genomic answers for children: dynamic analyses of >1000 pediatric rare disease genomes. *Genet Med*. 2022;24(6):1336–1348.
8 Chu ATW, Fung JLF, Tong AHY, et al. Potentials and challenges of launching the pilot phase of Hong Kong Genome Project. *J Transl Genet Genomics*. 2022;6(2):290–303.
9 Chung BHY, Chau JFT, Wong GK. Rare versus common diseases: a false dichotomy in precision medicine. *NPJ Genom Med*. 2021;6(1):19.
10 Genomes Project Pilot Investigators, Smedley D, Smith KR, et al. 100,000 genomes pilot on rare-disease diagnosis in health care - preliminary report. *N Engl J Med*. 2021;385(20):1868–1880.
11 The Government of the United Kingdom. Genome UK: the futre of healthcare 2020. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920378/Genome_UK_-_the_future_of_healthcare.pdf.
12 Lewis C. Genomic testing for rare disease diagnosis—where are we now, and where should we be heading? The reflections of a behavioural scientist. *Eur J Hum Genet*. 2023;31(11):1204–1206.
13 Census and Statistics Department HKSAR. 2021 Population Census. Thematic report: ethnic minorities. 2022. Available from: https://www.census2021.gov.hk/doc/pub/21c-ethnic-minorities.pdf.
14 *Race Relations Unit, Home Affairs Department, The Government of the Hong Kong Special Administrative region*. The Demographics: Ethnic Groups; 2023. Available from: https://www.had.gov.hk/rru/english/info/demographics.htm.
15 Chiu ATG, Chung CCY, Wong WHS, Lee SL, Chung BHY. Healthcare burden of rare diseases in Hong Kong - adopting ORPHAcodes in ICD-10 based healthcare administrative datasets. *Orphanet J Rare Dis*. 2018;13(1):147.
16 Chu ATW, Chung CCY, Hong Kong Genome Project, Lo SV, Chung BHY. Marketing and publicity strategies for launching the pilot phase of the Hong Kong Genome Project. *J Transl Genet Genomics*. 2023;7(1):66–78.
17 Gargano MA, Matentzoglu N, Coleman B, et al. The Human Phenotype Ontology in 2024: phenotypes around the world. *Nucleic Acids Res*. 2024;52(D1):D1333–D1346.

18 Chu ATW, Tong AHY, Tse DMS, et al. The Hong Kong Genome Project: building genome sequencing capacity and capability for advancing personalised genomic treatments in Hong Kong. *J Transl Genet Genomics*. 2023;7:196–212.
19 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
20 McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
21 Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.
22 Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.
23 Rajaby R, Liu DX, Au CH, et al. INSurVeyor: improving insertion calling from short read sequencing data. *Nat Commun*. 2023;14(1):3243.
24 Dolzhenko E, Deshpande V, Schlesinger F, et al. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics*. 2019;35(22):4754–4756.
25 Robinson PN, Köhler S, Oellrich A, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res*. 2014;24(2):340–348.
26 Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019;51(11):1560–1565.
27 Robertson AJ, Tran K, Patel C, Sullivan C, Stark Z, Waddell N. Evolution of virtual gene panels over time and implications for genomic data re-analysis. *Genet Med Open*. 2023;1(1):100820.
28 Thaxton C, Good ME, DiStefano MT, et al. Utilizing ClinGen gene-disease validity and dosage sensitivity curations to inform variant classification. *Hum Mutat*. 2022;43(8):1031–1040.
29 Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the clinical genome resource. *Am J Hum Genet*. 2017;100(6):895–906.
30 Richards S, Aziz N, Bale S, et al. 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*. 2015;17(5):405–424.
31 ClinGen. Sequence variant interpretation. Available from: https://clinicalgenome.org/working-groups/sequence-variant-interpretation/.
32 Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med*. 2018;20(9):1054–1060.
33 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–3100.
34 Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci*. 2022;2(12):797–803.
35 Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res*. 2014;24(12):2022–2032.
36 Smolka M, Paulin LF, Grochowski CM, et al. Detection of mosaic and population-level structural variants with Sniffles 2. *Nat Biotechnol*. 2024;42(10):1571–1580.
37 Chiu R, Rajan-Babu I-S, Friedman JM, Birol I. Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biol*. 2021;22(1):224.
38 Patterson M, Marschall T, Pisanti N, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol*. 2015;22(6):498–509.
39 Riggs ER, Wain KE, Riethmaier D, et al. Chromosomal microarray impacts clinical management. *Clin Genet*. 2014;85(2):147–153.
40 Al-Attar M, Butterworth S, McKay L. A quantitative and qualitative analysis of patient group narratives suggests common bio-psychosocial red flags of undiagnosed rare disease. *Orphanet J Rare Dis*. 2024;19(1):172.
41 Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062–D1067.
42 Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557–563.
43 Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161–164.
44 Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell*. 2019;179(3):589–603.
45 Pang SY, Hsu JS, Teo KC, et al. Burden of rare variants in ALS genes influences survival in familial and sporadic ALS. *Neurobiol Aging*. 2017;58:238.e9–238.e15.
46 Xu F, Huang S, Li XY, et al. Identification of TARDBP Gly298Ser as a founder mutation for amyotrophic lateral sclerosis in Southern China. *BMC Med Genom*. 2022;15(1):173.
47 Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*. 2020;28(2):165–173.
48 The Lancet Diabetes E. Spotlight on rare diseases. *Lancet Diabetes Endocrinol*. 2019;7(2):75.
49 Wu B, Kang W, Wang Y, et al. Application of full-spectrum rapid clinical genome sequencing improves diagnostic rate and clinical outcomes in critically ill infants in the China neonatal genomes project. *Crit Care Med*. 2021;49(10):1674–1683.
50 Bertoli-Avella AM, Beetz C, Ameziane N, et al. Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet*. 2021;29(1):141–153.
51 Khan M, Cornelis SS, Pozo-Valero MD, et al. Resolving the dark matter of ABCA4 for 1054 Stargardt disease probands through integrated genomics and transcriptomics. *Genet Med*. 2020;22(7):1235–1246.
52 Miller DE, Sulovari A, Wang T, et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet*. 2021;108(8):1436–1449.
53 Kaplun L, Krautz-Peterson G, Neerman N, et al. ONT long-read WGS for variant discovery and orthogonal confirmation of short read WGS derived genetic variants in clinical genetic testing. *Front Genet*. 2023;14:1145285.
54 Chung CCY, Ng NYT, Ng YNC, et al. Socio-economic costs of rare diseases and the risk of financial hardship: a cross-sectional study. *Lancet Reg Health West Pac*. 2023;34:100711.
55 Willmen T, Völkel L, Ronicke S, et al. Health economic benefits through the use of diagnostic support systems and expert knowledge. *BMC Health Serv Res*. 2021;21(1):947.
56 Wang D, Pascual JM, De Vivo D. Glucose transporter type 1 deficiency syndrome. In: Adam MP, Mirzaa GM, Pagon RA, et al., eds. *GeneReviews((R))*. Seattle WA; 1993.
57 Tang M, Park SH, De Vivo DC, Monani UR. Therapeutic strategies for glucose transporter 1 deficiency syndrome. *Ann Clin Transl Neurol*. 2019;6(9):1923–1932.
58 Ragona F, Matricardi S, Castellotti B, et al. Refractory absence epilepsy and glut 1 deficiency syndrome: a new case report and literature review. *Neuropediatrics*. 2014;45(5):328–332.
59 Centre for Health Protection. Department of health, the government of the Hong Kong special administrative region. Life Expectancy at Birth (Male and Female), 1971 - 2022. Available from: https://www.chp.gov.hk/en/statistics/data/10/27/111.html.