

Original Article

Open Access



The Hong Kong genome project: building genome sequencing capacity and capability for advancing genomic science in Hong Kong

Annie T. W. Chu¹, Amy H. Y. Tong¹, Desiree M. S. Tse¹, Cario W. S. Lo¹, Carol C. F. Lau¹, Cecilia Y. F. Li¹, Nattily S. Y. Tai¹, Lap W. Wong¹, Gigi K. C. Choy¹, Belinda Y. Y. Tse¹, Su-vui Lo¹, Ken Sung¹, Mullin Yu¹, Hong Kong Genome Project, Brian H. Y. Chung^{1,2}

¹Hong Kong Genome Institute, Hong Kong, China.

²Department of Pediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

Correspondence to: Dr. Brian H. Y. Chung, Hong Kong Genome Institute, 2/F, Building 20E, Hong Kong Science Park, Hong Kong, China. E-mail: bhychung@genomics.org.hk

How to cite this article: Chu ATW, Tong AHY, Tse DMS, Lo CWS, Lau CCF, Li CYF, Tai NSY, Wong LW, Choy GKC, Tse BYY, Lo Sv, Sung K, Yu M, Hong Kong Genome Project, Chung BHY. The Hong Kong genome project: building genome sequencing capacity and capability for advancing genomic science in Hong Kong. *J Transl Genet Genom* 2023;7:196-212. <https://dx.doi.org/10.20517/jtgg.2023.22>

Received: 29 Jul 2023 **First Decision:** 9 Oct 2023 **Revised:** 12 Oct 2023 **Accepted:** 20 Oct 2023 **Published:** 27 Oct 2023

Academic Editor: Nathan A. Berger **Copy Editor:** Fangling Lan **Production Editor:** Fangling Lan

Abstract

Aim: The Hong Kong Genome Project (HKGP) is the first large-scale genome sequencing (GS) project in the Hong Kong Special Administrative Region. The Hong Kong Genome Institute (HKGI) is entrusted with the task of implementing the HKGP. With the aim to sequence 45,000-50,000 genomes in five years, it is the project's goal to provide participants with more precise diagnosis and personalised treatment, and to drive the application and integration of genomic medicine into routine clinical care.

Methods: The HKGI Laboratory's hardware and software components were customised to tailor to the needs of the project. Sample handling and storage protocol, DNA extraction, and PCR-free GS workflow were developed and optimised. Quality control indicators and metrics for assessing the quality of samples, sequencing libraries and sequencing data were established.

Results: The Laboratory is designed to facilitate a unidirectional GS workflow to minimise the risk of contamination. The Sample Manager system handles laboratory data generated from the HKGP samples and biobank. The



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Laboratory handles and analyses approximately 350-500 samples per week, the majority of which are whole blood. During the first 24 months since the launch of the HKGP, 12,937 participants and their family members (6,680 genomes) have been recruited and sequenced. The sequencing capacity of the Laboratory has been further enhanced to include the latest technologies, such as long-read sequencing and multi-omics in order to meet the target of the HKGP.

Conclusion: HKGI Laboratory established a robust GS workflow for the HKGP. The clinical utility of GS will bring precision medicine into routine clinical practice.

Keywords: Hong Kong Genome Project (HKGP), genome sequencing, genomic medicine, precision health, laboratory establishment

INTRODUCTION

Advances in DNA sequencing technologies have fueled the genome sequencing era and led to the initial draft of the human genome sequence and, more recently, the telomere-to-telomere genome assembly^[1]. Disease-risk and treatment response studies^[2,3] provided clinical interpretation of genomic variants and paved the way for the development of genomic diagnostics and therapies^[4]. Substantial improvements in next-generation sequencing and bioinformatic tools have significantly shortened sequencing times, yielded higher sensitivity, specificity, and accuracy, better coverage, and reduced cost^[5,6], democratising genome sequencing from the individual level to the population scale. This is reflected in the launch of national genome sequencing initiatives around the world, driving genomic medicine and improving healthcare through the collection, storage, and application of genomic data^[7]. In 2015, China launched its first Precision Medicine Initiative^[8], followed by neighboring Asian countries such as Japan^[9,10] and Thailand^[11,12]. Other national genome projects have also emerged in France^[13], the United Kingdom^[14-17], Denmark^[18], Australia^[19-21], Canada^[22], the United States^[23], Saudi Arabia^[24], and Turkey^[25,26]. The scale of these projects ranges from sequencing twenty-five thousand to one hundred million genomes using various genomic technologies, including DNA microarrays, RNA sequencing, targeted gene panel sequencing, exome sequencing (ES), and genome sequencing (GS)^[7,27,28], spanning four to fifteen years. As more genome projects are underway, we expect the catalog of human genomic variations and functional annotation to grow steadily^[29,30], marking a major step towards embedding genome sequencing into routine clinical care.

Hong Kong has been a Special Administrative Region of the People's Republic of China since 1997. Hong Kong continues to have its own economic, legal, social, healthcare and welfare infrastructures despite being a recognised financial hub with modern city standards and a high standard of living. The latest population forecasts released by the Census and Statistics Department predict an increase in Hong Kong's population from 7.70 million in 2023 to 8.10 million in 2039^[31]. Over 90% of the population of Hong Kong is of ethnic Chinese descent, according to the 2016 by-census^[32], with other ethnic groups making up the remaining 8%. Hong Kong is viewed as a prime location to carry out studies on Southern Chinese populations' health due to this relatively homogeneous population. Shouldering about 90% of the inpatient needs of the entire community, the Hospital Authority provides a strong public healthcare safety net through 43 hospitals and institutions, 49 Specialist Out-patient Clinics (SOPCs), and 74 General Out-patient Clinics (GOPCs). With a well-established dual-track healthcare system, the remaining population also has easy access to privately funded healthcare.

Hong Kong has embarked on her journey of developing genomic medicine. Following the release of the Policy Address in 2017, the Hong Kong Special Administrative Region Government established a Steering Committee on Genomic Medicine^[33]. Upon reviewing the local landscape, the Committee put forward the

recommendations for the strategic development of genomic medicine in Hong Kong, including the set-up of the Hong Kong Genome Institute (HKGI) in 2020 under the Health Bureau, to implement the Hong Kong Genome Project (HKGP) in 2021^[34]. HKGP is the first large-scale GS project in the city to catapult genetic and genomic services and research^[34], marking an important milestone in the revolution of the healthcare system in Hong Kong^[7,27]. Accelerating the advancement of genomic medicine, HKGP is conducted in two phases: the pilot phase and the main phase. In the pilot phase, from mid-2021 to late 2022, efforts were dedicated to developing and streamlining operational workflows for genomic diagnosis, with the aim to sequence and analyse approximately 5,000 genomes by GS to demonstrate the capability of HKGI and the feasibility of HKGP^[7,27]. This represents approximately 2,000 cases of patients with undiagnosed diseases or hereditary cancers and their family members, with the majority of the cases subjected to trio analysis to assist in genomic data interpretation. Valuable experience and insights gained from the pilot phase have laid a solid foundation for formulating directions for the main phase, adding “genomics and precision health” as a new theme, covering other disease and research cohorts, and aiming to sequence 45,000-50,000 genomes (approximately 18,000-20,000 cases) over three years from 2023 to 2025^[34].

The entire patient journey is the highlight of HKGP, with considerable efforts dedicated to designing and crafting the protocol and process according to international standards of medical ethics and participants' rights^[7]. According to the advice sought from the various professional bodies and personnel, this patient-centered process starts with clinician engagement, and the emphasis that all personnel are fully aware of the importance of a thorough informed consent process. Another primary and unique process is the element of genetic counselling. To facilitate the implementation of HKGP through addressing the needs and challenges of the profession, a representative group of experts and stakeholders in the fields of genetics and genomics was gathered to standardise genomic counselling practice. While pre-test genetic counselling focuses on the process of translating complex genetic information into colloquial and relevant information between the clinicians and the patient, post-test genetic counselling provides an opportunity for participants to understand the genetic diagnosis and discuss the implications of the findings with genetic counsellors or clinical geneticists. The project also has clear protocols for withdrawal procedures, in which withdrawal does not preclude joining the project again, for re-enrollment and consenting procedures were addressed in the project design.

For a seamless GS operation at the beginning of the HKGP, a sequencing service provider was engaged to provide the wet-lab process while the HKGI laboratory was designed and built to provide a bespoke GS workflow with a laboratory information management system. To handle the increasing workload and tight timeline, a NovaSeq 6000 was installed in the HKGI laboratory in mid-2022 to boost the sequencing capacity for the project. A bioinformatics pipeline and variant curation workflow were also developed to assist in the identification of causal variants. Multidisciplinary team (MDT) meeting is an integral part of the HKGP, allowing exchanges from relevant specialists to attain consensus on a molecular diagnosis and clinical management plan. By leveraging the collective knowledge and skills of different specialties, such as clinicians, genetic counsellors, laboratory scientists, genome curators, bioinformaticians, allied health professionals, and trainees, HKGP has been tackling complex genomic challenges and delivering cutting-edge advancements in healthcare. This collaborative environment allows for prime integration of diverse expertise and perspectives, fostering innovative approaches and ensuring comprehensive patient care.

The present paper focuses on sharing the experience in setting up the HKGI laboratory and genome sequencing workflow, and scaling up the sequencing capability and capacity in the laboratory to handle the increasing workload for the HKGP. The HKGI laboratory (Laboratory) comprises four major components: (i) multidisciplinary talents; (ii) laboratory infrastructure tailored for clinical genomics; (iii) scalable semi-

automation sequencing workflow; and (iv) quality assurance/quality control measures, privacy protection, and electronic records. By documenting the experiences and challenges in designing the laboratory and establishing the GS workflow in a tight timeline, lessons learnt might assist international counterparts in steering their own course in the genomic medicine era.

METHODS

Participant recruitment

The HKGP starts its operational workflow with the participants' journey, from engaging referring clinicians in recruitment to ending the diagnostic odyssey with personalised treatment by offering the first end-to-end GS service in Hong Kong [Figure 1]. Eligible patients and families are recruited by a HKGP team set up in each hospital (known as the Partnering Centre; PC). The University of Hong Kong/Queen Mary Hospital (HKU/QMH; HKWC), the Chinese University of Hong Kong/Prince of Wales Hospital (CUHK/PWH; NTEC), and the Hong Kong Children's Hospital (HKCH; KCC) are PCs in the Pilot Phase. In the main phase, recruitment has been extended to other cluster institutions of Hong Kong West Cluster (HKWC, namely The Duchess of Kent Children's Hospital at Sandy Bay (DKCH) and Grantham Hospital (GH) and New Territories East Cluster (NTEC, namely North District Hospital (NDH) and Alice Ho Miu Ling Nethersole Hospital (AHNH). Eligible participants are referred to PCs by clinicians after screening and informed consent is conducted through face-to-face interviews^[7].

Sample collection and transfer

For each participant, 6 mL of blood is obtained and stored in two 3-mL EDTA-containing anticoagulation tubes. RapidDri Pouch kit (Isohelix) is used to collect cells inside the cheek for buccal swabs, and saliva samples are collected using the GeneFix Saliva DNA Collection and Stabilization Kit (Isohelix). Specimen collection is performed at the PCs. The specimens are packed in zip bags and stored at 4 °C until being transferred to the HKGI Laboratory, or for a maximum of 72 h. The specimens are packed with cooling packs and temperature-logging devices in the isothermal transfer boxes with combination locks.

Genomic DNA extraction and QC

Following sample registration, blood samples were aliquoted either manually or using the liquid handling system Freedom EVO100 (Tecan). Genomic DNA (gDNA) was extracted from 400 µL of whole blood using the QIASymphony SP system and QIASymphony SP DNA Midi Kit (Qiagen). For saliva and buccal swab samples, gDNA was extracted using the EZ2 Connect system and EZ1&2 DNA Tissue Kit (Qiagen). gDNA concentration was determined using the Qubit dsDNA BR assay kit and was measured with the Qubit 4 Fluorometer (Thermo Fisher Scientific). gDNA purity was determined using a NanoDrop One Spectrophotometer (Thermo Fisher Scientific). gDNA integrity was assessed for degradation using a 1% E-Gel precast gel electrophoresis system (Thermo Fisher Scientific). Approximately 100 ng of each gDNA sample was loaded into the precast agarose gel.

Illumina GS sequencing and QC

PCR-free GS libraries were constructed using the KAPA HyperPlus kit for PCR-free workflow and KAPA Unique Dual-Indexed adapter kit (Roche) following the instructions provided by the manufacturer. 1 µg of gDNA is fragmented enzymatically at 37 °C for 15 min, end-repaired, 3'dA-tailed, ligated to dual-index adapters, and size-selected. Reaction cleanup and double-sided size selection steps were performed using KAPA HyperPure Beads (Roche). For the double-sided size selection, 50 µL of beads were added to the adapter-ligated library to remove large-sized DNA fragments in the first cut. To remove small-sized DNA fragments, 8 L of beads were added to the supernatant from the first size cut, resulting in the final library size range of 400-700 bp.

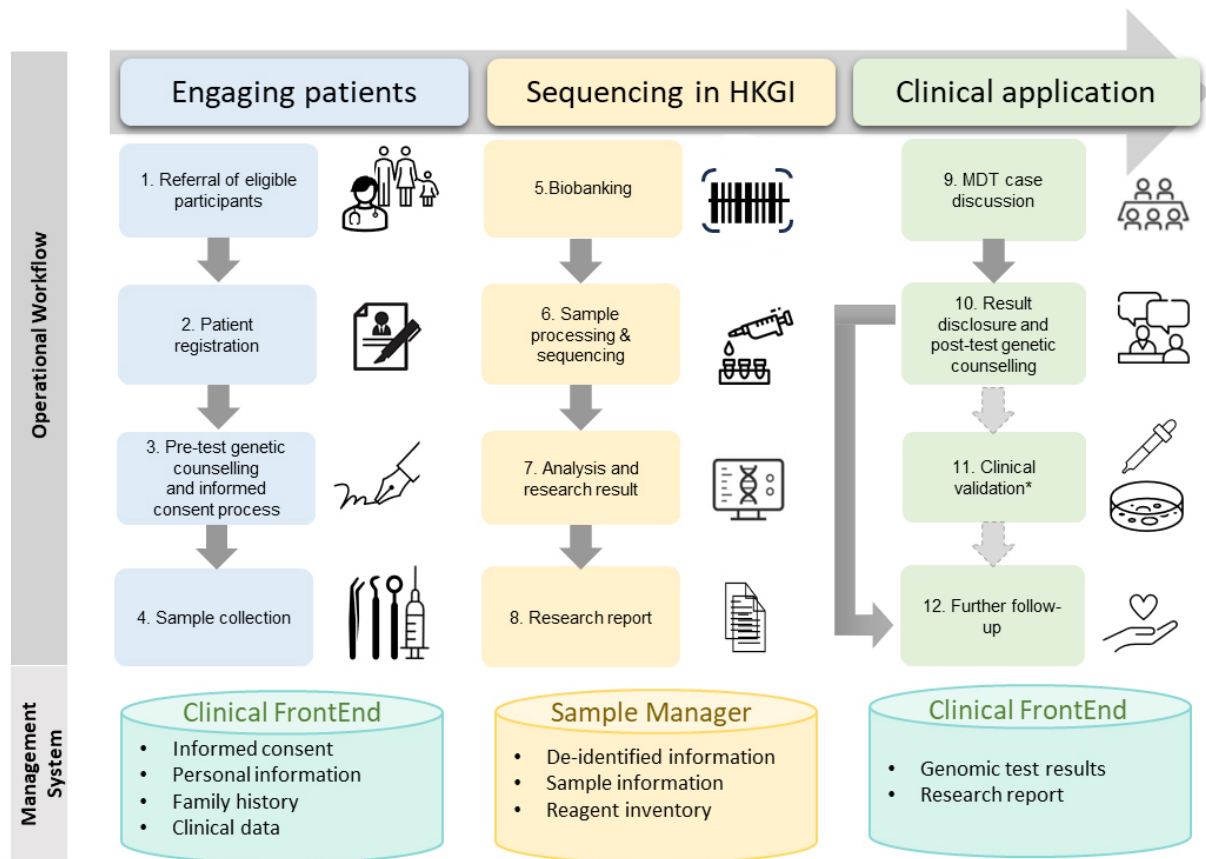


Figure 1. The operational workflow of HKGP using the main data managers: clinical FrontEnd stores all clinical-related data and documents, connected with Sample Manager using de-identified sample IDs. Sample Manager manages the biobank, records the GS journey of the sample, and works as a reagent inventory.

The GS library insert size was determined using the 4200 TapeStation and D1000 ScreenTape assay (Agilent). The library concentration was determined using the dsDNA HS assay kit and measured with the Qubit 4 Fluorometer (Thermo Fisher Scientific). The libraries were quantified by quantitative PCR using KAPA Library Quantification kit (Roche) and QuantStudio™ 5 Real-Time PCR system, 384-well or StepOnePlus™ Real-Time PCR system (Thermo Fisher Scientific). An equimolar library pool containing 24 dual-indexed GS libraries was combined prior to sequencing on the Illumina NovaSeq 6000 sequencer using NovaSeq 6000 S4 Reagent kit v1.5 (300 cycles), with 1% spike-in PhiX control (Illumina).

Sequence data analysis and validation

Base-calling was done using DRAGEN version 4.1.5. The secondary analysis workflow followed the best practice guidelines provided by the Genome Analysis Toolkit (GATK)^[35]. Reads were aligned to the GATK-provided reference genome Homo_sapiens_assembly38.fasta using BWA version 0.7.17^[36] and duplicates were removed using Picard version 2.27.4^[37]. Base quality score recalibration, variant calling, and variant filtering were performed using GATK version 4.2.6.1 and in-house tools. Annotation was performed using Variant Effect Predictor version 104, BCFtools version 1.13, and in-house tools^[38,39].

Following sequence data quality control steps, the bioinformatic pipelines identify and filter a list of variants for each GS sample. Candidate variants are prioritised based on the phenotype-based Exomiser^[40], and the expert crowdsourced reviewed PanelApp software^[41]. Sequence variants are classified according to the

standards and guidelines of the American College of Medical Genetics (ACMG)^[42]. Post-analysis multidisciplinary team (MDT) meetings facilitate the exchange of views on the GS findings with respect to patient clinical indications, refining diagnoses and clinical management plans for individual patients. GS findings are validated by appropriate orthogonal methods such as Sanger sequencing, RNA sequencing, long-read sequencing, and digital PCR.

RESULTS

Hong Kong genome institute laboratory

Taking references and guidelines from various professional bodies into consideration, the development of the laboratory data information management system and GS workflow is fine-tuned to meet the recommendations set out by the Medical Genome Initiative for clinical GS^[43-46]. One of the critical features of the Laboratory design is the capability of conducting a unidirectional workflow, and in conjunction with proper laboratory practices and operation, can minimise the risk of sample contamination. It requires physical separation of different stages of the procedure, dedicated equipment and supplies for each stage, and a workflow that prevents samples or laboratory personnel from moving “backwards” to potentially contaminate upstream workspaces [Figure 2]. In addition, differential air pressure is maintained to prevent contamination between the rooms. Pass-through chambers with interlocking doors and UV sterilisation capability enable the transfer of reagents or samples between physically separated rooms without compromising isolation and minimising cross-contamination. It ensures that reagents and samples pass through the Laboratory according to the designated route as demarcated in the GS workflow. Briefly, the GS workflow is divided into six wet-bench processes, in the order of: (i) reagent preparation; (ii) biological sample reception and registration; (iii) sample processing and biobanking; (iv) nucleic acid extraction; (v) GS library preparation and quality assessment; and (vi) sequencing. The main steps are performed in six key rooms, including the Reagent Preparation Room, Sample Reception Room, Sample Processing Room, Freezer/Biobank Room, Library Preparation Room, and Sequencing Room^[47] [Figure 2].

The Hong Kong genome project biobank

HKGI developed and employed the data manager, Clinical FrontEnd, to facilitate and standardise the patient recruitment process and clinical data collection across different recruitment sites, by properly handling and housing different types of data from the participants [Figure 3]. The interface of this portal includes an e-consent form and a clinical information collection form for supporting patient recruitment, while the specimen collection form and GS report provide efficient information exchange between the HKGI and PCs. The participants' samples are collected at the PCs, and then delivered to the HKGI Laboratory, which acts as a central processing hub for registration, processing, and biobanking. After verification of the participant's identification data in the Clinical FrontEnd, the samples are de-identified to protect participants' personal data and confidentiality, and a HKGI Laboratory ID, a unique alphanumeric identifier, is assigned for downstream processing [Figure 3]. The de-identified HKGI Laboratory IDs are handed over to the LabKey Sample Manager, which is an independent environment for handling laboratory data generated from the HKGP samples and biobank. Its logical data structure, fine-grained security management of data access, and intuitive user-friendly interface facilitate the tracking of samples and reagents in the Laboratory [Figure 3].

The Laboratory routinely handles and analyses approximately 350-500 samples per week of various natures ranging from whole blood, buccal swab, saliva, and tissues. Sample processing uses a hybrid of manual and automated approaches; details of the operations such as date and time, operator identifiers, and location are logged in the Sample Manager system. To minimise repeated freeze-thaw cycles and potential sample degradation and contamination, samples are first divided into aliquots, barcoded, and transferred to the ultra-low temperature archives of the HKGP biobank [Figure 3]. The system maintains audit logs

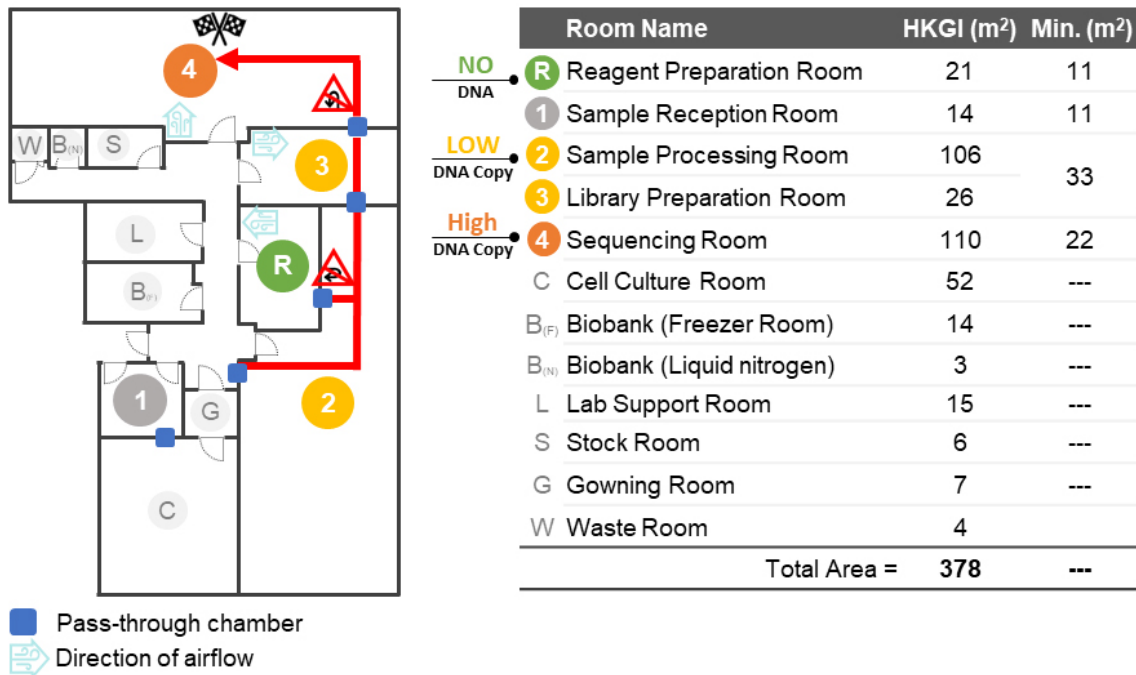


Figure 2. The HKGI genomic laboratory is designed for a unidirectional workflow, with differential air pressure to prevent contamination between the rooms. Pass-through chambers with interlocking doors allow the transfer of reagents or samples between physically separated rooms without compromising isolation and minimising cross-contamination. The movement of biological samples, reagents, and laboratory personnel strictly follows the unidirectional workflow, from the “No DNA” room to “Low DNA copy” rooms, and finally to the “High DNA copy” room. The Reagent Preparation room holds positive pressure, keeping environmental contaminants from entering, while the Library Preparation and Sequencing rooms hold a negative pressure to contain potential contaminants within the rooms and reduce the risk of contaminating corridors and other rooms.

containing detailed linked records of the sample type, number and volume of aliquots, storage location, sample status, derivatives of each aliquot, and associated assay data. As HKGP has a relatively tight timeline with reference to the number of genomes to be sequenced, optimising and enhancing the throughput of high-quality GS is a priority for the laboratory. Automation is integrated into the labor-intensive workflow to maximise productivity, reduce human errors, and increase consistency and reproducibility. Laboratory personnel organise and monitor the automation systems, performing quality assessments of extracted genomic DNA and sequencing libraries, while standard operating procedures (SOPs) and pre-formatted data worksheets guide routine operations. All controlled laboratory documents, including quality manuals, laboratory safety manuals, equipment maintenance records, SOPs, and staff training records, are managed in the Laboratory Document Management System for up-to-date distribution and access by authorised personnel.

The genome sequencing workflow and quality metrics

Taking reference from the Medical Genome Initiative, the Laboratory established a list of stringent quality control indicators and metrics for assessing the quality of the samples and sequencing libraries [Figure 4 and Table 1]. The Laboratory developed and optimised protocols to process different sample types and prepare a PCR-free GS library for the HKGP. To promote genomics research in Hong Kong, the Laboratory further co-developed the GS protocol and established related quality metrics with the DNA sequencing core facility at the University of Hong Kong, Centre of PanorOmics Sciences (CPOS), for the sequencing of HKGP samples. While the role of the sequencing service provider remains critical for the project, the sequencing capacity of the Laboratory has been further enhanced and will take on the responsibility of the

Table 1. Summary of quality metrics for genome sequencing (GS)

Metric	Threshold or expected value ^[45,48]	Mean of 240 GS data performed by HKGI
Yield of data \geq Q30 ^a	\geq 80 Gb	162 Gb
Mean coverage ^b	\geq 30X	41.0X
Base \geq Q30 % ^c	\geq 85%	90.0%
Clusters passing filter % ^d	\geq 70%	80.9%
Sample identity ^e	Match/not match	All match
Contamination % ^f	\leq 2%	0.0055%
Mapping rate % ^g	$>$ 95%	99.9%
10X percentage (%) ^h	\geq 95%	95.7%
Gene passed 15X % ⁱ	\geq 90%	99.3%
Adapter-dimer % ^j	$<$ 0.2%	0.0014%
Duplication % ^k	$<$ 15%	14.4%
Mean insert size ^l	$>$ 300 bp	496bp

^aTotal yield of data with \geq Q30 scores; ^bMean coverage across the human reference genome, after all filters are applied; ^cPercentage of bases that meet Q30 scores; ^dA cluster is considered to pass the filter when its chastity value is below 0.6 in the first 25 cycles. Cluster passing filter % is the percentage of clusters with passing filter; ^eConcordance with genotype when family structure is available; ^fEstimated level of sample cross-individual contamination based on a genotype-free estimation; ^gPercentage of unique reads that mapped to the human reference genome; ^hPercentage of bases in human genome with sequencing depth of \geq 10X; ⁱA measure of completeness. Percentage of genes with sequencing depth of \geq 15X; ^jFraction of pass-filtered reads that are unaligned and match to a known adapter sequence; ^kPercentage of mapped sequence that is marked as duplicate; ^lMedian insert size for all paired-end reads where both ends map to the same chromosome.

sequencing workload in the next few years. During the first 24 months since the launch of the HKGP, 12,937 participants and their family members (6,680 genomes) have been recruited and sequenced by the end of July 2023. As expected, the majority of the cases are from the undiagnosed disease category and a smaller cohort (~12.1%) from hereditary cancer. In order to illustrate the performance of the GS workflow, ten sequencing runs carried out at the Laboratory consisting of 240 samples will be presented in the following sections.

After a series of pilot studies on the method of genomic DNA (gDNA) extraction, two automated magnetic bead-based protocols were established for the extraction of DNA: a high-throughput system for whole blood samples, and a medium-throughput system for saliva, buccal swab, and tissue samples. The extracted DNA is eluted in a slightly alkaline buffer, 10 mM Tris-HCl, pH 8.0, and EDTA is omitted as it interferes with enzymatic reactions in the downstream sequencing library preparation. The extracted gDNA is assessed for degradation using agarose gel electrophoresis. Each DNA sample migrates as a high-molecular weight band without any smearing or signs of degradation, indicative of intact gDNA of high quality and integrity [Supplementary Figure 1]. The purity of extracted DNA is evaluated using the NanoDrop spectrophotometer. A typical pure gDNA has an A260/A280 absorbance ratio of 1.7-2.0 and an A260/A230 ratio of 1.8-2.5. All extracted gDNA samples showed an absorbance ratio of A260/A280 and A260/A230 within the acceptable range, denoting the absence of protein, carbohydrate, salts, and other contaminants. In addition to using UV absorbance, gDNA is quantified using the Qubit fluorometer. Figure 4B shows the concentration and total yield for the 240 gDNA samples. On average, 400 μ L of whole blood yields ~6 μ g of gDNA, at a concentration ~122 ng/ μ L. As indicated, all 240 blood samples resulted in high-quality gDNA, sufficient for GS coverage of 30X.

The GS library preparation protocol has been optimised for both manual and automated operations. Figure 4C shows the quality metrics for the 240 GS libraries. Using 1 μ g of gDNA as input, the average final library concentration is about 15.5 nM (in 25 μ L volume), which is more than sufficient to reach 30-100X genome coverage. The insert size of the GS libraries is analysed on an automated electrophoresis system

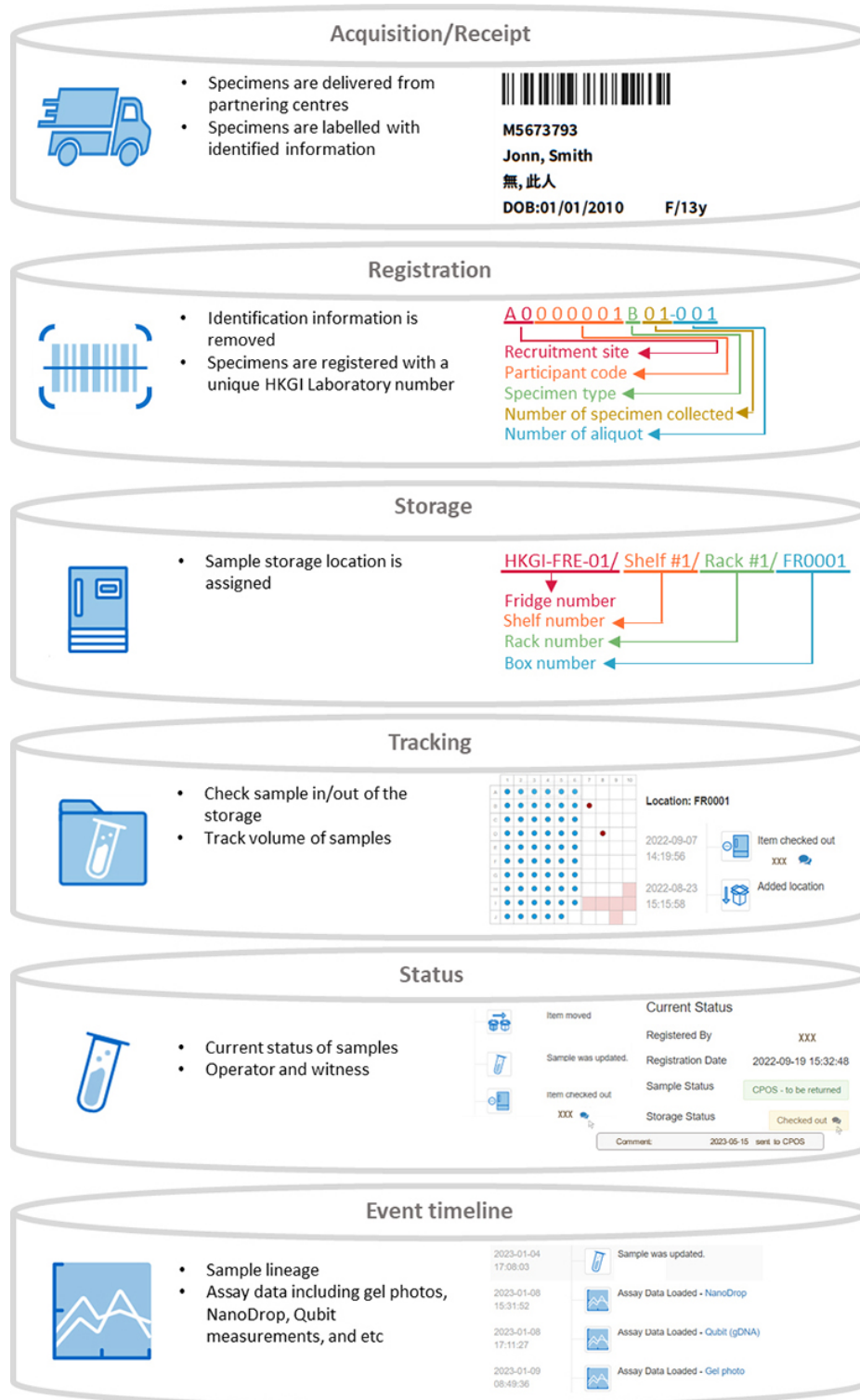


Figure 3. HKGI biobank and data management in Sample Manager for tracing sample lineages, storage, laboratory data, and relevant information during the genome sequencing process.

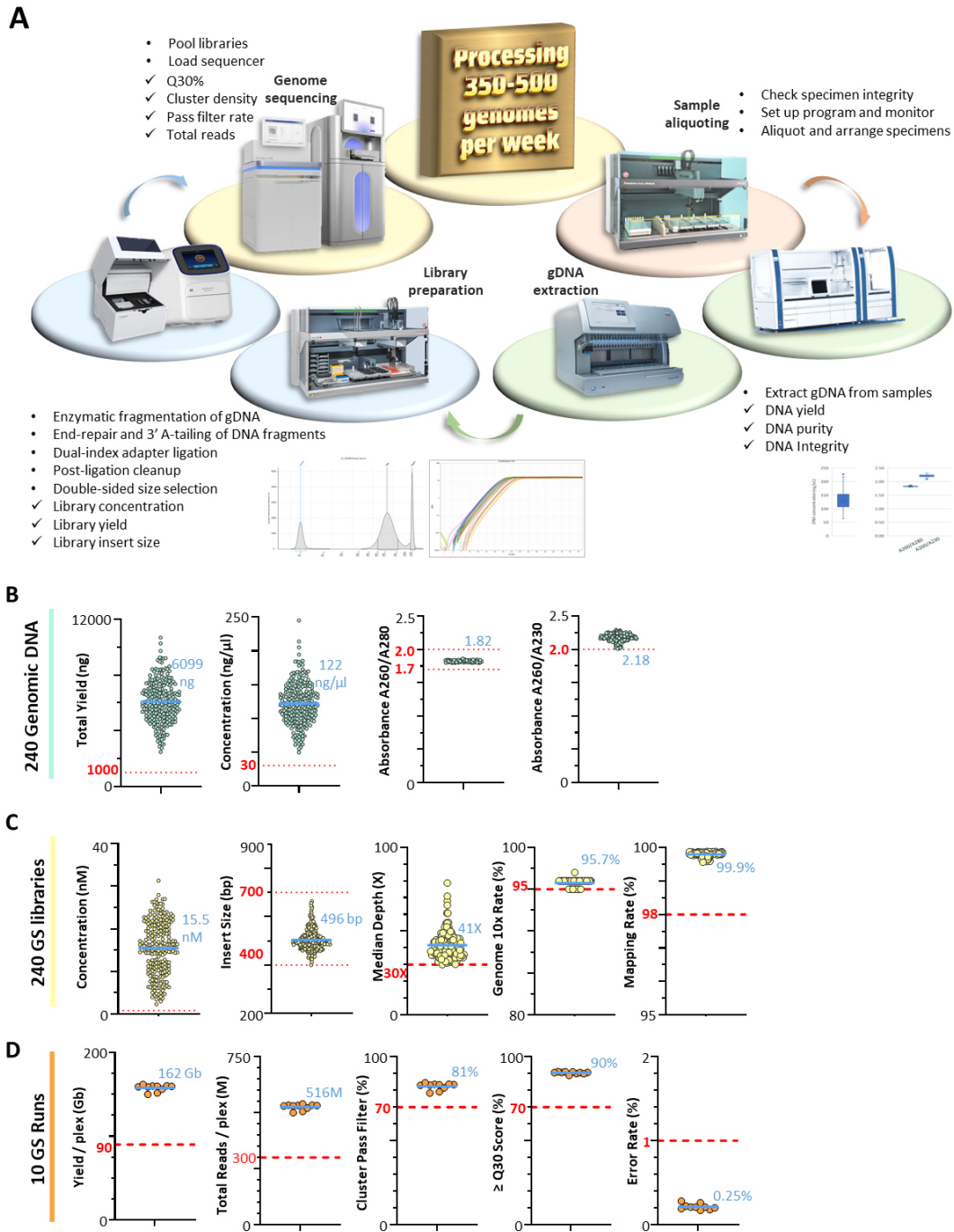


Figure 4. Performance statistics of 240 genome sequencing (GS) conducted by HKGI. (A) Quality indicators and thresholds used in monitoring the quality of different stages in the GS workflow. (B) Plots showing quality and quantity statistics of the 240 extracted gDNA. (C) Plots showing quality and quantity statistics of the 240 GS libraries. (D) Performance statistics of 240 GS libraries in ten NovaSeq 6000 runs. The 240 samples all meet the stringent quality indicators and metrics for assessing the quality of the samples and sequencing libraries.

(TapeStation, Agilent). Amongst the 240 GS libraries shown, the library insert size ranges from 400 to 700 bp, with a median insert size of 496 bp, which is within the range of efficient cluster formation on the Illumina patterned flow cells. GS libraries that pass the above quality controls then proceed to qPCR

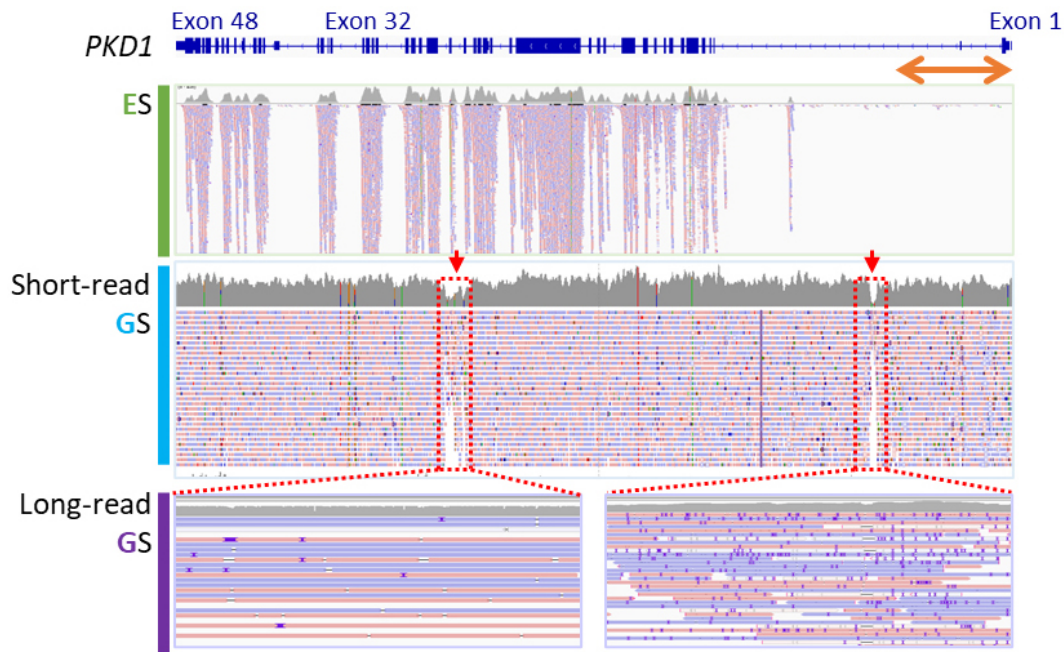


Figure 5. Comparison of exome sequencing (ES), short-read and long-read genome sequencing (GS) in resolving complex regions (“dark regions”) of the human genome. An example of such regions is the *PKD1* (polycystic kidney disease 1) gene, where the first 32 exons are located in a segmental duplicated region on chromosome 16p13, with six pseudogenes located 13 Mb proximal to the *PKD1* locus. In addition to high GC content, the sequences of these six pseudogenes are highly homologous to *PKD1* and share 97% sequence similarity, making amplification- and capture-based approaches challenging. The *PKD1* region is visualised with Integrative Genomics Viewer (IGV) using different sequencing approaches. Despite improvements in the capture probe design, ES of exons 1 to 14 of *PKD1* showed lower coverage, while GS achieved a more uniform coverage for the entire locus, including the duplicated region. Long-read GS enables unambiguous alignment of reads, complementing short-read GS, and enhances disease diagnosis. The orange double arrow indicates the “dark region”. The red dotted box and arrow indicate regions where short-read GS covers poorly.

quantification, and only libraries that pass all quality indicators are sequenced. An equimolar library pool containing 24 dual-indexed GS libraries is combined prior to sequencing on the Illumina NovaSeq 6000 sequencer using S4 flow cells with 300 cycles (2×150 bases).

As the number of nanowells is fixed in the patterned flow cells, the optimal loading concentration is determined by comparing the nanowell occupied rate (%Occupied) and pass filter rate (%PF). The optimal cluster density was attained after several rounds of optimisation on the S4 flow cell. The overall performance of 10 sequencing runs is consistent and of high quality, as shown in Figure 4D. Over 81% of clusters passed the chastity filter; more than 90% of bases had Q30, while the error rate using spike-in PhiX control was less than 0.25%. On average, each GS library yielded 162 Gb of data with 41X depth and over 516 million reads [Figure 4C and D]. Nearly all reads (99.9%) can be mapped to the human reference genome (GRCh38), while cross-individual contamination and adapter-dimer were merely detected. In summary, the overall statistics indicate that the established GS workflow is robust, and the data generated from the HKGI laboratory and CPOS are comparable, and on par with international standards.

The performance of GS in detecting variants in complex regions (“dark regions”) of the human genome is illustrated in the example of the *PKD1* (polycystic kidney disease 1) gene. Mutations in the *PKD1* gene contribute to 80%-85% of autosomal dominant polycystic kidney disease (ADPKD) cases. ADPKD is an inherited renal disease characterised by many fluid-filled cysts in the kidneys that progressively impair kidney functions and eventually result in end-stage renal disease. *PKD1* lies in a segmental duplication

region, with six pseudogenes located at 13 Mb proximal to the original *PKD1*. The sequence of these six pseudogenes is highly homologous to the *PKD1* gene, with a sequence similarity as high as 97.7%^[47], making the genetic diagnosis of ADPKD challenging when using exome sequencing (ES) or targeted enrichment approaches. Compared to ES, GS showed a more uniform coverage of the entire *PKD1*, including the duplicated region [Figure 5]. Our preliminary data also showed that long-read GS uniformly covered more of the “dark regions” in the genome, including the duplicated regions of *PKD1* that are challenging for short-read GS [Figure 5]. Our findings have demonstrated the capability of GS in clinical applications. The genetic diagnosis informed the patient’s clinical management and treatment choices, such as the use of Tolvaptan to slow down the progression of kidney failure.

DISCUSSION

As of today, HKGI is the only organisation in Hong Kong to provide free end-to-end WGS with a goal to cater to participants’ diagnostics needs and answer clinical management questions. With the opportunity to customise the Laboratory’s hardware and software components to tailor its own needs, the design, and layout of the Laboratory, the monitoring and data management systems were all carefully planned and crafted to serve the specific needs of the HKGP. While local accreditation programs for medical tests involving next-generation sequencing are still under development, the design, construction, and outfitting of the HKGI Laboratory adopt international standards for clinical genomics laboratory for DNA sequencing. The HKGP biobank has the capacity to house more than 200,000 tubes of sample aliquots currently, and can be scaled up as the project extends to various focused disease areas. Together with the genomic database, the HKGP biobank opens new research opportunities in a collaborative environment.

Successful genomics research requires broad public participation and informed collaboration between researchers and society, which relies on trustworthy sharing, effective management, and appropriate privacy and data security protections^[47]. Informed consent guidelines, data collection and storage protocols, and responsible sharing policies will be reviewed, improved, and augmented to provide the best practices in genomic research. A research environment is developed to facilitate efficient and effective genomic data sharing and analysis with clinicians and researchers. A well-developed infrastructure that facilitates active genomic research collaborations enables the integration of scientific discoveries and genomic findings into clinical practice. Collaboration between researchers and clinicians could promote the development of standardised protocols for data collection, analysis, and interpretation in genomic medicine. This infrastructure enhances the accuracy and reliability of genomic testing and enables more effective treatment decision-making based on individual genetic profiles. Building a biobank with the inclusion of a population deviated from that of European descent could also drive research and innovation in genetic medicine, leading to the discovery of new gene therapies and interventions to further improve patient outcomes.

The Laboratory adopted a PCR-free GS library preparation for the HKGP. The advantages of an amplification-free GS have been demonstrated by many studies^[49,50], including significantly reduced biases introduced by the DNA polymerase, reduced duplication rates and false positives, and greater sensitivity in calling indel and copy number variants. In addition, it provides better mapping and more uniform genome coverage^[50,51], allowing comprehensive detection of a wide range of variants, from single nucleotide variants to structural variants. Following the evaluation of different commercial kits for GS library preparation, an enzymatic fragmentation-based approach was selected. It does not only offer greater flexibility in the input gDNA amount, but also adaptability to liquid handling systems for routine library preparation. The Laboratory has devoted considerable investment to fine-tuning the library insert size range and final library yield, specifically the fragmentation and double-sided size selection steps, to achieve longer insert lengths for better coverage and indel variant detection.

The optimal loading concentration of the sequencing libraries is essential for a successful sequencing run. A loading concentration that is too high results in over-clustering and run failure. On the other hand, a loading concentration that is too low results in under-clustering and reduced output and accuracy. The optimal loading concentration is dependent on the library type, sequencing system, and reagent kit, and it requires to be adjusted empirically for each sequencer.

In order to achieve the ambitious target of sequencing 45,000-50,000 genomes by 2025, the Laboratory has recently installed additional automation systems and the latest sequencer, NovaSeq X Plus, which promises an increase of 2.5X throughput with the 25B flow cell that will be released later this year (2023). The higher throughput and lower per unit sequencing cost will benefit population-scale projects like the HKGP. Through standardisation of GS workflow and genomic data, it paves the way for data sharing and collaboration, ultimately advancing the field of genomics and improving patient care^[52].

Stepping into the future: the important role of WGS laboratory workflow in enhancing the development of precision medicine in the genomic era

Due to lower assay cost and faster turn-around time, ES and gene panels have been the routine tests employed in clinical genomic diagnosis for the past decade^[53,54]. However, this targeted approach requires PCR enrichment of the targeted regions, limiting the detection of small-sized variants found in exonic regions and the overall efficacy. In contrast, GS enables comprehensive interrogation of the entire genome with the option of PCR-free library preparation, allowing the unbiased identification of different types of genetic variants, including protein-coding, regulatory and noncoding regions, as well as regions affecting RNA splicing. Given the superior performance of GS and higher clinical utility^[55] compared to other sequencing technologies, it is not only adopted by the HKGP and other international genome projects^[8-10,14,15,18,26], but is also replacing ES as a first-tier test in the clinic^[53,56].

Advancements in long-read sequencing technologies and broad application have earned it the “Method of the Year 2022”^[57]. Long-read sequencing has been shown to complement the shortcomings of short-read technologies, such as directly identifying structural variants and methylation patterns, resolving complex rearrangements, sequencing homologous and repetitive regions, phasing of alleles, and so on^[58-60], making previously computationally challenging and inferential approaches more straightforward. Other applications of long-read sequencing, like the characterisation of full-length RNA isoforms while preserving native RNA modifications^[61], and the detection of aberrant splicing and gene fusion events^[58], have the potential to facilitate functional interpretation of genomic variation^[62]. Considering the wide range of applications in clinical genomics^[63-65], the Laboratory introduced the Oxford Nanopore Technology (ONT) PromethION system to handle challenging cases such as certain neurological and neuromuscular diseases that are known to be caused by short tandem repeat expansions, and others that are unresolved by short-read GS. The longer reads also allow haplotype phasing of compound recessive variants and the detection of structural variants with precise breakpoint details, as shown in our analysis of *PKD1*. Our preliminary data showed that long-read GS uniformly covered many of the “dark regions” in the genome, including the duplicated regions of *PKD1* that are challenging for short-read GS.

To improve our understanding of the underlying biology of different diseases at the cellular and tissue levels, single-cell genomics, spatial multi-omics, and proteomics technologies have been applied to investigate thousands of individual cells in an unbiased approach^[66-71]. In recent years, single-cell sequencing has been widely adopted in cancer research to interrogate tumour microenvironment and heterogeneity, and tumour clonal lineage^[72,73]. The Laboratory has also established a single-cell sequencing platform to integrate multi-omics information to enhance the characterisation of disease at the cellular and tissue levels and enable the discovery of biomarkers for therapeutic targets. Integration of clinical information, GS,

single-cell sequencing and other omics data holds the promise of transforming healthcare as it facilitates diagnosis, targeted treatment and prognostic prediction, and may allow better clinical management and inform targeted therapeutic development.

Since commencing its operation, HKGI has overcome many challenges while embracing opportunities to improve the workflow along the way. Tremendous efforts have been invested in establishing its Laboratory from scratch, from recruiting talents and training new bloods, to setting GS quality and QC benchmarks, developing operational workflow and completing the sequencing of over 6,680 genomes, continuous fine-tuning and enhancing laboratory workflow. All within the first two years. HKGI sees the GS Laboratory as an important entry point into the scientific workflow. Our preliminary findings have demonstrated the capability of GS in advancing personalised genomic treatment, as illustrated in the *PKD1* example. While shouldering the important responsibility of offering the first end-to-end GS in Hong Kong, HKGI also values the privilege of paving the way for the career training and development of GS laboratory personnel from medical technologists, laboratory scientists and researchers, laboratory assistants to interns and trainees aspired to embark on the journey. As we move into the main phase, the HKGI GS platform will enable the implementation of precision medicine in clinical research and practice, going beyond genomics.

DECLARATIONS

Acknowledgments

We would like to thank the participants of the HKGP. We also thank all members of the HKGI for preparing the launch of the HKGP, especially Jimmy Jiang, Janice Wong, Raven Lee, Nathan Lau, and Sau-Dan Lee. We thank The University of Hong Kong, Centre of PanorOmics Sciences (CPOS), for the sequencing of HKGP samples. We also wish to acknowledge the support of the HKGP stakeholders: The Health Bureau, Hospital Authority, Department of Health, and Partnering Centres at The University of Hong Kong/Queen Mary Hospital, The Chinese University of Hong Kong/Prince of Wales Hospital, and Hong Kong Children's Hospital.

Authors' contributions

Made substantial contributions to the conception and design of the study: Chung BHY, Chu ATW, Tong AHY

Drafted the article and made critical revisions: Chung BHY, Chu ATW, Tong AHY, Tse DMS, Lo CWS

Performed data analysis and interpretation: Tong AHY, Lo CWS

Performed data acquisition: Lau CCF, Li CYF, Tai NSY, Wong LW, Choy GKC, Tse BYY

Provided administrative, technical, and material support: Lo SV, Tse DMS, Sung K, Yu M

Resources: Hong Kong Genome Project

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author, [BHYC]. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Ethical approval for this study was obtained from Central Institutional Board, Hospital Authority (HKGP-2021-001 & HKGP-2022-001), The Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (2021.423 & 2023.120) and Institution Review Board of the University of Hong Kong/ Hospital Authority Hong Kong West Cluster (UW 21-413 & UW23-289). Written informed consent has been obtained from all participants.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2023.

REFERENCES

1. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science* 2022;376:44-53. DOI
2. Beck T, Rowlands T, Shorter T, Brookes AJ. GWAS central: an expanding resource for finding and visualising genotype and phenotype data from genome-wide association studies. *Nucleic Acids Res* 2023;51:D986-93. DOI PubMed PMC
3. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062-7. DOI PubMed PMC
4. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature* 2011;470:204-13. DOI
5. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: an overview. *Hum Immunol* 2021;82:801-11. DOI PubMed
6. The cost of sequencing a human genome. 2020. Available from: <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost> [Last accessed on 23 Oct 2023].
7. Chu ATW, Fung JLF, Tong AHY, et al. Potentials and challenges of launching the pilot phase of Hong Kong genome project. *J Transl Genet Genom* 2022;6:290-303. DOI
8. Liu J, Hui RT, Song L. Precision cardiovascular medicine in China. *J Geriatr Cardiol* 2020;17:638-41. DOI
9. Japan Agency for Medical Research and Development. GA4HG. 2022. Available from: https://www.amed.go.jp/en/aboutus/collaboration/ga4gh_gem_japan.html [Last accessed on 23 Oct 2023].
10. Global Alliance for Genomics & Health. GEM Japan releases largest-ever open-access Japanese variant frequency panel. 2022. Available from: https://www.ga4gh.org/news_item/gem-japan-releases-largest-ever-open-access-japanese-variant-frequency-panel/ [Last accessed on 23 Oct 2023].
11. Genomics Thailand. Available from: <https://genomicsthailand.com/Genomic/home> [Last accessed on 23 Oct 2023].
12. A twisted ladder to prosperity. Available from: <https://www.nature.com/articles/d42473-020-00211-y> [Last accessed on 23 Oct 2023].
13. Aviesan. Genomic medicine France 2025. 2022. Available from: https://solidarites-sante.gouv.fr/IMG/pdf/genomic_medicine_france_2025.pdf [Last accessed on 23 Oct 2023].
14. Genomics England. The 100,000 genomes project by numbers. Available from: <https://www.genomicsengland.co.uk/news/the-100000-genomes-project-by-numbers> [Last accessed on 23 Oct 2023].
15. Smedley D, Smith KR, Martin A, et al. 100,000 Genomes pilot on rare-disease diagnosis in health care - preliminary report. *N Engl J Med* 2021;385:1868-80. DOI PubMed PMC
16. Cookies on our future health. Available from: <https://ourfuturehealth.org.uk/research-programme/> [Last accessed on 23 Oct 2023].
17. Genome UK: 2021 to 2022 implementation plan. Available from: <https://www.gov.uk/government/publications/genome-uk-2021-to-2022-implementation-plan/genome-uk-2021-to-2022-implementation-plan> [Last accessed on 23 Oct 2023].
18. Danish ministry of health. Personalised medicine for the benefit of the patients. 2022. Available from: <https://eng.ngc.dk/Media/637614364621421665/Danish%20Strategy%20for%20personalised%20medicine%202021%202022.pdf> [Last accessed on 23 Oct 2023].
19. Australian genomics. Available from: <https://www.australiangenomics.org.au/our-history/> [Last accessed on 23 Oct 2023].
20. Stark Z, Boughtwood T, Phillips P, et al. Australian genomics: a federated model for integrating genomics into healthcare. *Am J Hum Genet* 2019;105:7-14. DOI PubMed PMC
21. Department of health and aged care, australian government. Genomics health futures mission. Available from: <https://www.health.gov.au/initiatives-and-programs/genomics-health-futures-mission> [Last accessed on 23 Oct 2023].
22. Genome Canada. Available from: <https://genomecanada.ca/challenge-areas/all-for-one/> [Last accessed on 23 Oct 2023].
23. National human genome research institute. The undiagnosed diseases program. Available from: <https://www.genome.gov/Current-NHGRI-Clinical-Studies/Undiagnosed-Diseases-Program-UDN> [Last accessed on 23 Oct 2023].
24. Saudi genome program. Available from: <https://www.vision2030.gov.sa/en/projects/saudi-genome-program/> [Last accessed on 23 Oct 2023].

25. BBMRI-ERIC. Turkish genome project launched. Available from: <https://www.bbmri-eric.eu/news-events/turkish-genome-project-launched/> [Last accessed on 23 Oct 2023].
26. Özçelik T. Medical genetics and genomic medicine in Turkey: a bright future at a new era in life sciences. *Mol Genet Genomic Med* 2017;5:466-72. DOI PubMed PMC
27. 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. DOI PubMed PMC
28. Chung BHY, Chau JFT, Wong GK. Rare versus common diseases: a false dichotomy in precision medicine. *NPJ Genom Med* 2021;6:19. DOI PubMed PMC
29. Posey JE, O'Donnell-Luria AH, Chong JX, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med* 2019;21:798-812. DOI
30. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434-43. DOI
31. Census and statistics department, Hong Kong special administrative region. Table 150-11001: projected mid-year population by sex and age group. Available from: https://www.censtatd.gov.hk/en/web_table.html?id=150-11001 [Last accessed on 23 Oct 2023].
32. Census and statistics department, Hong Kong special administrative region. 2016 population by-census: summary results. Available from: <https://www.byccensus2016.gov.hk/data/16bc-summary-results.pdf> [Last accessed on 23 Oct 2023].
33. The steering committee of genomic medicine, Hong Kong special administrative region. Strategic development of genomic medicine in Hong Kong. 2022. Available from: https://www.healthbureau.gov.hk/download/press_and_publications/otherinfo/200300_genomic/SCGM_report_en.pdf [Last accessed on 23 Oct 2023].
34. Hong Kong genome institute. strategic plan 2022-25. Available from: <https://hkgp.org/ebook/HKGI-Strategic-Plan/> [Last accessed on 23 Oct 2023].
35. Van der Auwera G, O'Connor B. Genomics in the cloud: using docker, GATK, and WDL in Terra, 1st edition. Sebastopol, CA: O'Reilly Media 2020.
36. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. Available from: <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L/abstract> [Last accessed on 23 Oct 2023].
37. Picard tools. Version 2.17.8. Broad institute, GitHub repository. Available from: <http://broadinstitute.github.io/picard/> [Last accessed on 23 Oct 2023].
38. McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol* 2016;17:122. DOI PubMed PMC
39. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008. DOI PubMed PMC
40. Smedley D, Jacobsen JO, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 2015;10:2004-15. DOI PubMed PMC
41. Martin AR, Williams E, Foulger RE, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet* 2019;51:1560-5. DOI
42. 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;17:405-24. DOI PubMed PMC
43. Hall L, Wilson JA, Bernard K, et al. Establishing molecular testing in clinical laboratory environments; Proposed guideline. 2011. Available from: https://clsi.org/media/1474/mm19a_sample.pdf [Last accessed on 23 Oct 2023].
44. Aysal A, Pehlivanoglu B, Ekmekci S, Gundogdu B. How to set up a molecular pathology lab: a guide for pathologists. *Turk Patoloji Derg* 2020;36:179-87. DOI PubMed PMC
45. Marshall CR, Chowdhury S, Taft RJ, et al. Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *NPJ Genom Med* 2020;5:47. DOI PubMed PMC
46. Aziz N, Zhao Q, Bry L, et al. College of American pathologists' laboratory standards for next-generation sequencing clinical tests. *Arch Pathol Amp Lab Med* 2015;139:481-93. PubMed
47. Bogdanova N, Markoff A, Gerke V, McCluskey M, Horst J, Dworniczak B. Homologues to the first gene for autosomal dominant polycystic kidney disease are pseudogenes. *Genomics* 2001;74:333-41. DOI
48. Hübschmann D, Schlesner M. Evaluation of whole genome sequencing data. *Methods Mol Biol Clifton N J* 2019;1956:321-36. DOI
49. Ribarska T, Bjørnstad PM, Sundaram AYM, Gilfillan GD. Optimization of enzymatic fragmentation is crucial to maximize genome coverage: a comparison of library preparation methods for Illumina sequencing. *BMC Genom* 2022;23:92. DOI PubMed PMC
50. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* 2009;6:291-5. DOI PubMed PMC
51. Quail MA, Swerdlow H, Turner DJ. Improved protocols for the illumina genome analyzer sequencing system. *Curr Protoc Hum Genet* 2009;62:18.2.1-27. DOI PubMed
52. Rehm HL, Page AJH, Smith L, et al. GA4GH: international policies and standards for data sharing across genomic research and healthcare. *Cell Genom* 2021;1:100029. DOI
53. Retterer K, Jussola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med* 2016;18:696-704. DOI
54. Lee H, Deignan JL, Dorrani N, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014;312:1880-7. DOI PubMed PMC

55. 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;25:100896. DOI PubMed
56. Scocchia A, Wigby KM, Masser-Frye D, et al. Clinical whole genome sequencing as a first-tier test at a resource-limited dysmorphismology clinic in Mexico. *NPJ Genom Med* 2019;4:5. DOI PubMed PMC
57. Marx V. Method of the year: long-read sequencing. *Nat Methods* 2023;20:6-11. DOI PubMed
58. Xing L, Shen Y, Wei X, et al. Long-read Oxford nanopore sequencing reveals a de novo case of complex chromosomal rearrangement involving chromosomes 2, 7, and 13. *Mol Genet Genom Med* 2022;10:e2011. DOI
59. Roberts HE, Lopopolo M, Pagnamenta AT, et al. Short and long-read genome sequencing methodologies for somatic variant detection; genomic analysis of a patient with diffuse large B-cell lymphoma. *Sci Rep* 2021;11:6408. DOI PubMed PMC
60. Chung BHY, Kan ASY, Chan KYK, et al. Analytical validity and clinical utility of whole-genome sequencing for cytogenetically balanced chromosomal abnormalities in prenatal diagnosis: abridged secondary publication. *Hong Kong Med J* 2022;28:4-7. PubMed
61. Garalde DR, Snell EA, Jachimowicz D, et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 2018;15:201-6. DOI
62. Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. *J Hum Genet* 2020;65:3-10. DOI PubMed PMC
63. Turner H. Clinical long-read sequencing. WYNG foundation briefing. 2022. Available from: <https://www.phgfoundation.org/briefing/clinical-long-read-sequencing> [Last accessed on 23 Oct 2023].
64. Turner H. Long-read sequencing: clinical applications and implementation. WYNG foundation briefing. 2022. Available from: <https://www.phgfoundation.org/briefing/lrs-clinical-applications-and-implementation> [Last accessed on 23 Oct 2023].
65. Sanford Kobayashi E, Batalov S, Wenger AM, et al. Approaches to long-read sequencing in a clinical setting to improve diagnostic rate. *Sci Rep* 2022;12:16945. DOI PubMed PMC
66. Rood JE, Maartens A, Hupalowska A, Teichmann SA, Regev A. Impact of the human cell atlas on medicine. *Nat Med* 2022;28:2486-96. DOI PubMed
67. Alfaro JA, Bohländer P, Dai M, et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat Methods* 2021;18:604-17. DOI PubMed PMC
68. Tang X, Huang Y, Lei J, Luo H, Zhu X. The single-cell sequencing: new developments and medical applications. *Cell Biosci* 2019;9:53. DOI PubMed PMC
69. Nomura S. Single-cell genomics to understand disease pathogenesis. *J Hum Genet* 2021;66:75-84. DOI PubMed PMC
70. Cuomo ASE, Nathan A, Raychaudhuri S, MacArthur DG, Powell JE. Single-cell genomics meets human genetics. *Nat Rev Genet* 2023;24:535-49. DOI PubMed
71. Sreenivasan VKA, Balachandran S, Spielmann M. The role of single-cell genomics in human genetics. *J Med Genet* 2022;59:827-39. DOI PubMed PMC
72. Lei Y, Tang R, Xu J, et al. Applications of single-cell sequencing in cancer research: progress and perspectives. *J Hematol Oncol* 2021;14:91. DOI PubMed PMC
73. Liu J, Xu T, Jin Y, Huang B, Zhang Y. Progress and clinical application of single-cell transcriptional sequencing technology in cancer research. *Front Oncol* 2020;10:593085. DOI PubMed PMC