

REVIEW ARTICLE

Roadmap to genomic inclusivity: collaboration to achieve efficient, affordable, and accessible genetic testing for Southeast Asia

Vincentius Simeon Weo Budhyanto¹, Stefanie Jürgens¹, Jonas Korlach², Cindy Oktavi Cicilia¹, Zuhrah Taufiq³

ABSTRACT

Background: Southeast Asia's ~700 million people contribute only a small fraction of the data underpinning global genomic reference resources. The clinical consequences are elevated rates of variants of uncertain significance, polygenic risk scores that transfer poorly across ancestries, and screening calibrated to non-representative populations. The equity case is established; the operational case is not.

Methods: We searched PubMed/MEDLINE, Scopus, and Google Scholar from January 1, 2010 to March 31, 2026, with a targeted documentation update on August 5, 2026. Sequencing requirements were modeled by applying phased clinical penetration rates to demographic and epidemiological denominators, with instrument counts derived at 70% utilization. Every input and formula is published as a supplementary workbook, and outputs are bounded by sensitivity analysis.

Results: Capacity is profoundly uneven, from more than 50,000 completed genomes in Singapore to none in Timor-Leste. The eleven member states record ~10.6 million annual births and 1.15 million incident cancers. At maturity, the base case yields approximately 46.0 petabases annually (23.0-56.8 across scenarios), requiring approximately 42 short-read instruments (23 throughput-derived, seven buffer, 12 mid-throughput) and a mixed long-read fleet of ~22. Modeled 15-year infrastructure investment is US\$733 million (US\$644 million to US\$1.0 billion), or US\$0.07 per capita annually, excluding recurrent testing at US\$5.3-US\$12.2 per capita annually.

Conclusion: Long-read sequencing is a required regional capability for structurally complex loci, alongside validated targeted assays. Specialist workforce, not hardware, is the most acute constraint identified where data exist. We propose a federated hub-and-spoke architecture aligned to the proposed Association of Southeast Asian Nations Genome Consortium.

Keywords: Genomic medicine, Southeast Asia, Pharmacogenomics, Long-read sequencing, Health systems.

Introduction

Southeast Asia accounts for ~700 million people and carries a disproportionate burden of thalassemia, actionable pharmacogenomic variation, and rising cancer incidence, yet remains among the most underrepresented regions in global genomic databases (1). Throughout this review, "Southeast Asia" denotes the 11 member states of the Association of Southeast Asian Nations (ASEAN), and "member states" denotes those 11 countries; "countries" is used only in generic comparative statements. The underrepresentation is not merely a

matter of scientific completeness. Reference panels dominated by individuals of European descent produce

Correspondence to: Cindy Oktavi Cicilia

*Satriabudi Dharma Setia Foundation (YSDS), Jakarta, Indonesia.

Email: cindy.o.cicilia@dharmas.or.id

Full list of author information is available at the end of the article.

Received: 05 August 2026 | **Revision (1):** 02 September 2026 | **Accepted:** 10 October 2026



elevated rates of variants of uncertain significance in non-European patients and have generated outright genetic misdiagnoses (2). Polygenic risk scores derived in European cohorts perform substantially worse in other ancestries and deploying them without recalibration risks widening rather than narrowing disparities (3).

The unsampled diversity is considerable. Whole-genome sequencing (WGS) of 227 West Javanese individuals identified 1.8 million novel single nucleotide variants, distinct from those of central Indonesian populations and poorly captured by existing Asian reference panels (4). Indonesia alone comprises more than 1,300 recognized ethnic groups, and founder effects in geographically isolated populations favor identification of rare disease-associated variants and population-specific markers (5).

Five member states have launched national genome initiatives since 2019 - Singapore, Thailand, Malaysia, Indonesia, and Vietnam - and the reagent cost of generating a 30× human genome is approaching US\$100 (cost tier 1, defined in Section “Cost conventions”). Sasongko and colleagues recently proposed an ASEAN Genome Consortium to convert this activity into regional capability, arguing for interoperable data standards, selective data sharing, co-developed polygenic risk scores, harmonized ethical frameworks, and joint workforce development (6). Their proposal establishes the political and ethical case. What has not been articulated is the quantitative one: how much sequencing capacity, of which technology classes, at what cost, and on what timeline.

Without that quantification, national programs risk continuing to procure independently, duplicating capital expenditure, generating datasets that cannot be jointly analyzed, and building infrastructure poorly matched to the region's specific clinical genetics, which is dominated by structurally complex alpha-globin rearrangements and highly polymorphic pharmacogenes that short-read platforms characterize incompletely.

This review documents current capacity across all 11 member states including Timor-Leste, admitted on October 26, 2025 (7); characterizes the clinical domains driving demand and the technology requirements each imposes; models the infrastructure and investment required under a phased 15-year scenario; and proposes an implementation architecture consistent with the proposed Consortium's governance principles. The scope is deliberately infrastructural: ethical, legal, and social dimensions are addressed only where they bear on implementation feasibility, and readers are referred to the ASEAN Genome Consortium proposal (6) for fuller treatment. The quantitative component is a scenario-based roadmap, not a forecast; measured and published facts are distinguished from model outputs throughout, and the model itself is published in full so that readers may substitute their own assumptions.

Evidence Acquisition

Literature and documentary sources

This is a narrative review incorporating a structured capacity-modeling component. Peer-reviewed literature

was identified through PubMed/MEDLINE, Scopus, and Google Scholar in a formal search covering January 1, 2010 to March 31, 2026, combining the terms genomic medicine, WGS, precision medicine, long-read sequencing, thalassemia, pharmacogenomics, carrier screening, and health system capacity with Southeast Asia, ASEAN, and the names of the 11 member states. Terms were combined with OR within the concept and geographic blocks and with AND between them. Reference lists of included articles were hand-searched. Targeted searches were additionally run for each national genome initiative by name, for the specific loci discussed here (*HLA-B*15:02*, *CYP2C19*, *CYP2D6*, alpha-globin deletions), and for comparator regional programs.

Because program, ministry, and manufacturer documentation changes quickly, a targeted update was run on August 5, 2026, restricted to the national programs in Table 1, the manufacturer documentation for the platform classes in Supplementary Table S4, and newly published primary studies on the loci and applications already covered. The update was not a repetition of the full search, and this asymmetry is noted in Section “Limitations”. All uniform resource locators were verified on the access dates shown in the reference list. Databases, search strings, dates, and yields for both stages are given in Supplementary Table S9. Search reproducibility is limited by the narrative design; Google Scholar ranking in particular is not fully reproducible.

Because much operationally relevant information on national infrastructure exists outside the peer-reviewed literature, the review also draws on four categories of grey literature: ministry and national agency publications; manufacturer technical specifications and launch announcements; intergovernmental policy instruments; and program documentation from non-governmental genomic initiatives. These are identified as such in the reference list and table footnotes and are assigned evidence categories under Section “Evidence categories”.

Denominators and modeling

Population and birth estimates are drawn from the United Nations World Population Prospects 2024 revision, medium-fertility variant, for 2025 (8). Cancer incidence is drawn from GLOBOCAN 2022 and the comparative Southeast Asian analysis (9). Alpha-thalassemia carrier frequencies are drawn from a regional meta-analysis pooling 83,674 subjects (10); that analysis did not cover Indonesia, the Philippines, Myanmar, Singapore, Brunei, or Timor-Leste, and estimates for those member states are flagged accordingly.

For each clinical application, projected annual test volumes were derived by applying a clinical penetration rate to the relevant denominator - namely annual births, incident cancers, or total population - with rates increasing across three implementation phases. The rationale, benchmark and source for every penetration rate are given in Supplementary Table S8; Phase 3 rates are scenario upper bounds achievable only under public reimbursement, not forecasts. Per-test raw data volumes were assigned from published platform specifications

Table 1. Genomic sequencing capacity across Southeast Asia, early 2026, by metric type. Verified completed output is cumulative output reported in a peer-reviewed publication or official program report, with the date of that report. Current operational annual capacity is output currently being produced per year. Announced target is a program target stated publicly by the responsible authority and not yet met. Theoretical or planned capacity is capacity that installed or planned instrumentation could in principle deliver and is not a commitment. “Not reported” indicates that the metric could not be verified for that member state. Evidence categories are defined in Section “Evidence categories”.

Member state	Principal initiative	Verified completed output (date)	Current operational annual capacity	Announced target	Theoretical or planned capacity	Source; category
Singapore	PRECISE-SG100K (National Precision Medicine Phase II)	>50,000 WGS (December 2023)	~36,000 WGS/yr (~3,000/month)	~500,000 participants (Phase III)	Not reported	(13); A
Thailand	Genomics Thailand; National Biobank of Thailand	48,652 WGS (Feb 2025)	~10,000 WGS/yr	50,000 genomes	Not reported	(14); B
Indonesia	National Biomedical and Genome Science Initiative; non-profit university network	12,378 hWGS datasets (10 Sept 2025)	~15,456 hWGS samples/yr	29 sequencing facilities	>100,000 WGS/yr at full deployment of all planned network centres ^a	(17)(18); B
Malaysia	MyGenom (Malaysia Genome and Vaccine Institute)	>2,400 genomes completed in Phase I	Not reported	2,400 genomes, expanding to 10,000	Not reported	(19); B
Vietnam	VN1K; Medical Genetics Institute	1,011 WGS at median 37× (2022)	Not reported	Not reported	~1,000-5,000 WGS/yr, research-grade	(20); B
Philippines	Philippine Genome Center; first hospital-based facility (April 2025)	Not reported	<1,000 WGS/yr, primarily WES and panels	Not reported	Not reported	(21); B
Brunei	National Molecular Laboratory	Not reported	Clinical panels only (prenatal, HPV, oncology)	Not reported	Not reported	(22); B
Cambodia	Institut Pasteur du Cambodge; national programs	Not reported	Pathogen surveillance only	Not reported	Not reported	(23); B
Laos	Institut Pasteur du Laos; national laboratories	Not reported	Pathogen surveillance only	Not reported	Not reported	(23); B
Myanmar	GenRe-Mekong; national public health laboratories	Not reported	Pathogen surveillance only	Not reported	Not reported	(23); B
Timor-Leste	None identified	Not reported	Zero domestic capacity identified	Not reported	Not reported	(24); B

^aTheoretical aggregate assuming full deployment of all planned centers; no formal production target has been announced and this should not be read as committed capacity (category D).

and standard coverage depths. Per-instrument annual outputs were taken as 1,650 Tb for conventional ultra-high-throughput (UHT) short-read platforms, 4,000 Tb for wafer-based platforms, 275 Tb of high-fidelity data for high-fidelity circular-consensus (HiFi-class) long-read platforms, and 650 Tb for high-capacity nanopore platforms, each at the assumed utilization rate; instrument capital was costed at class-midpoint list prices (Supplementary Table S4). Instrument counts were derived by dividing aggregate annual data volume by per-instrument annual output at an assumed 70% operational

utilization rate, allowing for maintenance, quality control, run failures, and scheduling inefficiency. Long-read demand is derived from the short-read volumes and demographic denominators rather than assumed, as set out in Section “Long-read infrastructure”.

Where only allele frequencies were published and carrier prevalence was required, or the converse, conversion assumed Hardy–Weinberg equilibrium, with carrier prevalence approximated as $1 - (1 - p)^2$ for allele frequency p . Frequencies are labeled explicitly as allele

frequency, carrier prevalence, or phenotype frequency wherever they appear

Every input, intermediate calculation and output is published as Supplementary File S1, a spreadsheet in which all intermediate values are live formulas and every figure quoted in this article is traceable to a named cell. The model was independently re-executed from the published inputs by C.O.C., and the outputs matched the reported figures to within rounding.

Cost conventions

All cost figures are expressed in 2026 United States dollars unless stated. Three cost tiers are distinguished throughout, and no cost of a genome is stated without naming its tier. Tier 1, reagent cost, is the manufacturer-stated consumable cost of generating the sequence, currently approximately US\$80-US\$200 per 30× genome. Tier 2, sequencing-production cost, adds library preparation, quality control, instrument depreciation, service contracts, and laboratory labor. Tier 3, fully loaded clinical test cost, further adds bioinformatics, storage, variant interpretation, reporting, genetic counseling, confirmatory testing where required, and accreditation overhead; published microcosting of clinical genome sequencing in a national health service laboratory processing 399 samples per year found £6,841 per cancer case and £7,050 per rare disease case, with consumables accounting for 68%-72% of the total (11), and concluded that the aspirational US\$1,000 genome is not achieved once complete costs are counted.

Every currency amount carries a price-basis label at first mention: list price (L), negotiated price (N), modeled cost (M), reported expenditure (R), or published market range (P). Amounts converted from other currencies state the rate and date used. Because regional procurement typically achieves discounts, and sequencing costs have declined historically at rates exceeding 30% annually, capital estimates at list price should be read as upper bounds; conversely, they exclude facilities, personnel, bioinformatics, and interpretation costs, addressed separately in Sections “Data volume, storage and recurrent cost” and “Implementation roadmap and investment”.

The model is expressed in platform classes rather than named commercial products. Five classes are used: conventional ultrahigh-throughput (UHT) short-read; wafer-based ultrahigh-throughput (UHT) short-read; mid-throughput short-read; high-fidelity circular-consensus long-read; and high-capacity nanopore long-read. The manufacturer documents from which each class parameter was taken are listed in Supplementary Table S4 with document title, date, and locator, so that every parameter remains auditable; named products appear only in that table and in the reference list, and their appearance there does not constitute endorsement. Section “Public-private partnership and safeguards” addresses vendor lock-in risk.

Evidence categories

Because this review combines peer-reviewed research, government announcements, manufacturer documents,

programme websites, and author-derived estimates, every substantive statement is assigned to one of four evidence categories:

A, independently peer-reviewed;

B, official government or program data;

C, manufacturer-reported;

D, modeled or author-derived.

Categories are shown in Tables 1, 3, 4, and 5, on the inputs sheet of Supplementary File S1, and in the reference-statement concordance audit in Supplementary Table S7; author-derived values in the text are identified as category D at first mention. Readers weighing an individual claim should consult its category, since the review's model outputs (category D) rest on inputs of categories A-C of differing strength.

The Genomic Landscape Across 11 Member States

Capacity spans an extraordinary range, from Singapore's population-scale program to Timor-Leste's complete absence of domestic sequencing. This gradient tracks economic disparity but also reflects policy choice: several middle-income member states have committed substantial public funds within the last 6 years while others of comparable income have not. Table 1 summarizes capacity across all 11 member states, separating verified completed output, current operational capacity, announced targets, and theoretical or planned capacity, which are not equivalent metrics and were previously reported in a single column.

Singapore operates the most mature ecosystem through its National Precision Medicine program, coordinated by Precision Health Research Singapore (PRECISE) (12,13). Phase I (SG10K) established a 10,000-genome Asian reference cohort; Phase II (PRECISE-SG100K) surpassed 50,000 genomes by December 2023 and sustains ~3,000 genomes monthly, combining short-read production with high-fidelity and nanopore long-read sub-projects. Phase III targets approximately 500,000 citizens with integration into electronic health records. The sustained monthly output is used in Section “Short-read infrastructure” as the only published real-world benchmark for national-scale production available in the region.

Thailand launched Genomics Thailand in 2019 across five clinical pillars - namely cancer, rare disease, non-communicable disease, infectious disease, and pharmacogenomics - anchored by the National Biobank of Thailand and targeting 50,000 genomes (14). Thailand is also the region's clearest case of genomic services reaching reimbursement: it is the only member state funding *BRCA1/BRCA2* testing through universal coverage, and *HLA-B*15:02* screening before carbamazepine prescription was piloted under government policy in 2013 and reinforced nationally by the National Health Security Office in 2018 (15,16).

Indonesia pursues a dual-track model. The government Biomedical and Genome Science Initiative, under the national biomedical agency, has equipped multiple

Table 2. Demographic and epidemiological estimates of clinical volume requirements.

Member state	Population (2025) ^a	Annual births ^a	Rare/congenital cases per year ^b	Cancer cases per year ^c	Actionable cancers per year ^d	Alpha-thal carrier prevalence ^e	Health expenditure per capita (US\$) ^f
Indonesia	285,721,000	4,629,000	138,870	408,661	122,598	5%-10% ^g	159
Philippines	116,787,000	1,876,000	56,280	145,000	43,500	5.0% ^g	208
Vietnam	101,599,000	1,450,000	43,500	183,000	54,900	51.5%	168
Thailand	71,620,000	600,000	18,000	175,000	52,500	20.1%	364
Myanmar	54,851,000	928,000	27,840	70,000	21,000	5%-15% ^g	71
Malaysia	35,978,000	446,000	13,380	59,000	17,700	17.3%	477
Cambodia	17,848,000	380,000	11,400	17,000	5,100	39.5%	120
Laos	7,873,000	171,000	5,130	8,400	2,520	26.8%	69
Singapore	5,871,000	46,000	1,380	20,000	6,000	5%-7% ^g	4,044
Timor-Leste	1,419,000	34,000	1,020	1,100	330	5.0% ^g	100
Brunei	466,000	6,500	195	630	189	5%-7% ^g	666
Total	700,033,000	10,566,500	316,995	1,087,791	326,337	-	160.8 bn total

^aUnited Nations World Population Prospects 2024 revision, medium-fertility variant, 2025 (8) (category B).

^bCalculated as 3% of annual births per World Health Organization estimates of 2%-5% (category D).

^cGLOBOCAN 2022 and Dee et al. (9); member-state figures may not sum precisely to the verified regional total of 1,146,810 owing to rounding across sources (category A).

^d30% of incident cancers, derived conservatively from the 36.7% reported by Zehir et al. (38); actual fractions vary by cancer type distribution, and member states with higher liver cancer proportions may be lower (category D).

^ePooled carrier prevalence from Hussin et al. (10), alpha-thalassemia only, excluding beta-thalassemia and HbE (category A).

^fCurrent health expenditure per capita, current US\$, 2021 (48) (category B); the Timor-Leste value was not in the retrieved series and is a placeholder to be replaced.

^gNot covered by Hussin et al. (10); derived from independent country-level studies, provisional pending meta-analytic verification (category D).

sequencing laboratories with plans to expand to 29 facilities (17). In parallel, a non-profit network supports instrument placement and consumables directly at university laboratories across more than 35 sites and 40 institutions (18). The approaches are complementary: centralized clinical throughput alongside distributed research access.

Malaysia entered population genomics in August 2024 with MyGenom, targeting 2,400 genomes with expansion to 10,000 contingent on the 13th Malaysia Plan (19). Vietnam completed 1,000 Vietnamese Genomes Project (VN1K), sequencing 1,011 individuals at median 37× depth and identifying ~39.7 million variants of which 8.5 million were novel, producing the first Vietnamese pangenome reference (20).

The Philippines continues to expand through the Philippine Genome Center and, from April 2025, its first hospital-based sequencing facility, though capacity remains oriented toward clinical exome and panel testing (21). Brunei launched its first molecular laboratory in August 2023, confined to prenatal, human papillomavirus, and oncology screening (22). Cambodia and Laos maintain pathogen-focused capacity built during the COVID-19 pandemic but no human clinical genomics service; Myanmar's activity is restricted by political instability to infectious disease surveillance (23). Timor-Leste has no domestic capacity identified and relies on Australian laboratories (24).

Three constraints recur irrespective of program maturity.

Workforce. Where quantified, the shortage is extreme. The Philippines operates with ~10 clinical geneticists for more than 117 million people, roughly 1 per 11.7 million, with all practising specialists trained as pediatricians and most concentrated in Metro Manila (25,26). Vietnam reports 13 medical geneticists for more than 90 million people, ~1 per 6.8 million, with no formally trained genetic counselors (27). No regional census exists, and aggregate region-wide figures should be treated as unverified.

Financing. Insurance coverage is largely absent. Meaningful public reimbursement exists only in Thailand and Singapore. Elsewhere testing is predominantly out-of-pocket, restricting access to urban private-sector patients and undermining any population-screening rationale.

Regulation. Initiatives operate independently within and across borders, with no harmonized regional framework governing testing standards. Absent mutual recognition of laboratory accreditation and variant classification standards, a test validated in one member state carries no standing in another - a direct obstacle to the shared-infrastructure model in Section “A federated hub-and-spoke architecture”.

Table 3. Projected annual testing volumes by clinical application and phase, with the unit of count and raw data yield per test. All volumes are category D scenario outputs; see Supplementary Table S8 for the rationale for each penetration rate.

Clinical application	Phase 1 (Yr 1-3)	Phase 2 (Yr 4-7)	Phase 3 (Yr 8-15)	Raw data per test (Gb)	Unit of count
Rare disease WES	30,000	100,000	200,000	6-8 (base 8)	Testing event (one per proband)
Rare disease WGS	5,000	20,000	100,000	90-120 (base 100)	Testing event (one per proband)
Cancer CGP (tissue)	50,000	200,000	500,000	1	Testing event (patients may also have liquid biopsy and germline panel)
Cancer liquid biopsy	10,000	50,000	200,000	0.5-5 (base 2.75)	Testing event (repeat monitoring counted per test)
Hereditary cancer panels	20,000	100,000	300,000	0.5	Testing event (one per individual)
NIPT	500,000	3,000,000	8,000,000	0.5-2 (base 1.5)	Testing event (one per screened pregnancy)
Carrier screening (NGS)	500,000	3,000,000	5,000,000	2	Testing event (one per individual)
Pharmacogenomics (NGS)	200,000	1,000,000	3,000,000	1	Testing event (one per individual; one-off)
Total short-read testing events	1,315,000	7,470,000	17,300,000	-	-
Long-read: Thalassaemia reflex testing from carrier screening ^a	5,000	50,000	250,000	10	Testing event
Long-read: Prenatal diagnosis of thalassaemia in at-risk pregnancies (primary indication) ^a	951	9,510	47,549	10	Testing event
Long-read: CYP2D6 star-allele phasing reflex ^a	3,000	30,000	150,000	10	Testing event
Long-read: Rare-disease reflex whole genomes (SV unresolved by short reads) ^a	720	7,200	36,000	100	Testing event
Long-read: CNS tumour methylation classification ^a	283	2,828	14,141	10	Testing event
Total long-read testing events	9,954	99,538	497,691	-	-
Bioinformatic queries against stored data ^b	200,000	3,000,000	15,000,000	0 (no new sequence)	Testing event (re-analysis)
Total testing events	1,524,954	10,569,538	32,797,691	-	-
Unique individuals tested (Phase 3)	-	-	17,100,000	-	40% cancer-test overlap assumed

^aLong-read volumes are derived, not assumed: see Section "Long-read infrastructure" and Supplementary File S1, sheet 3. Volumes.

^bBioinformatic queries are re-analyses of previously generated data and consume no sequencing capacity; they are included in the recurrent cost of Section "Data volume, storage and recurrent cost" and in the count of testing events, but not in the count of unique individuals.

Table 4. Consolidated sequencing infrastructure requirements by phase, in platform classes. Rows (a) and (b) are alternative single-platform fleets, each sized to the whole long-read demand; row (c) is the proposed mixed fleet, and is the configuration used in the capital estimate and in Table 5. The three long-read rows are alternatives and must not be added. All figures are category D outputs from category C inputs (Supplementary Table S4); price basis L at class-midpoint list prices; sensitivity ranges are in Section “Sensitivity analysis” and Supplementary Table S2.

Quantity	Phase 1 (Yr 1-3)	Phase 2 (Yr 4-7)	Phase 3 (Yr 8-15)
Short-read sequence generated (Pb)	2.8	14.7	37.8
Conventional-class UHT units, throughput-derived ^a	2	9 (9-10)	23 (21-26)
Wafer-class UHT units, alternative to the above ^b	1	4 (4-5)	10 (9-11)
Scheduling buffer units ^c	1	3 (2-3)	7 (6-8)
Mid-throughput urgent-care units ^d	4	8	12
Total short-read fleet (conventional-class configuration)	7	20	42
Long-read testing events	9,954	99,538	497,691
Long-read sequence generated (Pb)	0.2	1.6	8.2
(a) Single-platform fleet: HiFi-class only ^a	1	6 (6-7)	30 (28-33)
(b) Single-platform fleet: nanopore-class only ^a	1	3	13 (11-15)
(c) Proposed mixed fleet: HiFi-class units	1	3	15
(c) Proposed mixed fleet: nanopore-class units	1	2	7
(c) Proposed mixed fleet: total long-read units	2	5	22
Sequencing hardware capital of the proposed configuration (US\$ million, L) ^e	8.7	20.6	57.2
Annual storage added (PB)	0.3	1.8	6.6

^aAt 70% operational utilization and manufacturer-specified output, class-range values in parentheses use the low and high annual outputs of the class (Supplementary Table S4).

^bWafer-class units are an alternative to, not an addition to, conventional-class units.

^cBuffer set at 30% of the throughput-derived count (range 25-35%).

^dMid-throughput units are a turnaround provision, additional to the throughput-derived count: one per Tier 1 hub in Phase 1, two per Tier 1 hub in Phase 2, and two per Tier 1 and Tier 2 hub in Phase 3.

^eCapital covers the proposed configuration only - conventional-class units plus buffer, mid-throughput units, and the mixed long-read fleet - at class-midpoint list prices, and excludes reagents, bioinformatics, facilities, and personnel; it is the fleet value at each phase, not cumulative purchases, which are given in Table 5.

Clinical Domains and Technology Requirements

“Genomic testing” is not a single service. Five domains dominate demand, each imposing distinct requirements on depth, read length, turnaround, and analytical infrastructure; conflating them produces infrastructure that is simultaneously over- and under-specified.

Rare and inherited disorders

Whole-exome sequencing (WES) targets the ~1.5% protein-coding genome at 100-150× depth, generating 6-8 Gb per singleton. Diagnostic yields range from 25% to 40% in previously undiagnosed rare disease (28), rising with trio designs at approximately US\$1,000-US\$2,500 per trio (P, tier 3). WGS at 30-40× depth (90-120 Gb) adds an incremental 7%-15% yield by capturing structural variants, non-coding variants, and repeat expansions (29). As the cost differential has narrowed, an evidence-based clinical guideline of the American College of Medical Genetics and Genomics has made a strong recommendation for exome or genome sequencing as a first- or second-tier test in children with congenital anomalies, developmental delay, or intellectual disability (30); the analytical

validation requirements for clinical genome sequencing intended for germline diagnosis are set out separately (31).

Inherited hemoglobin disorders are the region's highest-volume inherited disease application and the clearest case in which read length determines diagnostic accuracy. Traditional gap polymerase chain reaction (gap-PCR) and reverse dot-blot hybridization interrogate only pre-specified variants. A 2025 comparison of short-read next-generation sequencing against a targeted long-read panel across five complex monogenic conditions found that short-read sequencing identified only 56.4% of carriers detected by long-read sequencing, and produced a 13.7% false-positive rate among alpha-thalassemia [Online Mendelian Inheritance in Man (OMIM) #604131]-positive samples by systematically miscalling the functionally normal HKαα variant as -α3.7 (32). In patients with suspected thalassemia and inconclusive conventional screening, circular consensus long-read sequencing identified clinically significant rare variants in a further 17.67% of cases (33). Regionally critical alleles - the -SEA deletion, the predominant alpha-zero-thalassemia deletion across Southeast Asia, and co-inherited alpha-globin triplications that exacerbate

Table 5. Modeled 15-year infrastructure and capital investment envelope by component and phase (US\$ million, 2026 prices, modeled cost basis M unless stated). Derived bottom-up in Supplementary File S1, sheet 9_ Investment. Double-counting rules: sequencing hardware is cumulative purchases converted from the fleet-at-phase counts in Table 4, including one Phase 3 replacement cycle for Phase 1 instruments; service contracts are charged on the installed base at 10% of installed list price per year; recurrent clinical testing is excluded and shown as a memorandum row. All component values are category D estimates built on the inputs listed in Supplementary File S1, sheet 1_ Inputs.

Component	Phase 1	Phase 2	Phase 3	Total	Basis of estimate
Sequencing hardware (cumulative purchases; Phase 3 includes replacement of Phase 1 units)	8.7	11.9	45.2	65.8	Fleet additions from Table 4 x class-midpoint list price (L); one Phase 3 replacement cycle for Phase 1 units
Facilities (hub and spoke laboratory fit-out)	26.5	36.5	0.0	63.0	4 Tier 1 hubs x US\$12M; 2 Tier 2 hubs x US\$5M; 5 spoke facilities x US\$1M; Tier 1 and spoke spend split across Phases 1-2
Maintenance and service contracts	2.6	8.2	45.8	56.6	Installed fleet value x 10% per year x years in phase
Data infrastructure (storage, compute, federated platform)	24.3	14.5	88.3	127.2	Storage PB-years x US\$50,000 per PB-year; analysis compute at US\$10 per whole-genome-equivalent and US\$0.20 per screening test; HPC per Tier 1 hub with Phase 3 refresh; federated platform
Workforce and training	19.0	32.0	34.0	85.0	50 fellows per year x US\$60,000; 100 clinical-geneticist training places x US\$200,000; e-learning platform and faculty development
Networking and sample logistics	6.5	4.0	12.0	22.5	Sample logistics and cold chain per year; one-off regional connectivity in Phase 1
Accreditation and quality assurance	5.2	9.3	13.2	27.8	11 laboratories: initial accreditation plus annual external quality assessment
Biobanking	0.0	32.0	4.0	36.0	Biobank facilities in Tier 1 and Tier 2 hubs; regional coordination from Phase 2
Reference-panel construction (50,000 genomes)	12.8	74.8	0.0	87.5	50,000 genomes x US\$1,500 fully loaded (M); 5,000-genome long-read subset; curation and release
Program management and governance	6.0	12.0	32.0	50.0	Consortium secretariat, governance bodies, ethical-legal-social and data-governance functions
Other capital (pilot programs; interpretation and variant-data-base infrastructure)	10.0	25.0	10.0	45.0	Thalassemia-prevention and cancer-pharmacogenomics pilots; regional variant database and interpretation infrastructure
Subtotal	121.6	260.3	284.5	666.4	
Contingency (10% of subtotal)	12.2	26.0	28.5	66.6	Modeled allowance for scope and price variation
Total infrastructure and capital investment	133.7	286.4	313.0	733.1	Scenario range US\$644-1,019 million^a
<i>Memorandum - excluded: recurrent clinical testing expenditure at Phase 3 (Section "Data volume, storage and recurrent cost")</i>			<i>3.7-8.5 bn per year</i>		<i>Operating cost borne by health systems, payers and patients; not part of the capital envelope</i>

^aScenario range from Section "Sensitivity analysis" and Supplementary Table S2, obtained by re-scaling the hardware, service-contract, storage, and reference-panel components under the envelope-low and envelope-high runs.

beta-thalassemia (OMIM #613985) severity - are resolved with substantially greater accuracy on long-read platforms (34), and single-molecule real-time sequencing identified causative variants in 83.2% of hematology-positive individuals in a carrier cohort against 73.3% by conventional methods (35).

Cancer genomics

Hereditary cancer panels covering 30 to more than 100 genes are sequenced at 200-500× depth at US\$250-US\$4,000 per test (P, tier 3). Guidance from the National Comprehensive Cancer Network and the American Society of Clinical Oncology has materially expanded indications, with universal germline testing now recommended for all epithelial ovarian and exocrine pancreatic cancers (36). The rationale is empirical: among patients with advanced cancer undergoing universal tumor-normal sequencing, 49.2% of those carrying actionable germline pathogenic variants would not have been identified by family-history-driven criteria (37).

Somatic comprehensive genomic profiling interrogates 300-plus genes at greater than 500× depth in tissue, or 80-plus genes from circulating tumor DNA, at US\$1,500-US\$5,000 (P, tier 3). Prospective sequencing of more than 10,000 patients through a single institutional program found that approximately 37% harbored at least one actionable somatic alteration (38); the model adopts a conservative 30% (category D). Liquid biopsy additionally enables minimal residual disease monitoring and provides an alternative where tissue is inaccessible, which carries particular weight where interventional radiology capacity is limited.

Pharmacogenomics

Southeast Asia is a global hotspot for actionable pharmacogenomic variation, with allele frequencies diverging sharply from European reference populations (39,40). This is where the case for regionally specific reference data is strongest, because the frequencies determining screening cost-effectiveness are precisely the parameters that differ. Country- and ancestry-specific frequencies, with the type of frequency stated for each entry, are given in Supplementary Table S5; regional aggregates are not reported, because the available data do not support them.

*HLA-B*15:02*, conferring markedly elevated risk of Stevens–Johnson syndrome and toxic epidermal necrolysis on carbamazepine exposure, is rare in European and Japanese populations. Regulatory label language reports a carrier prevalence above 15% in Hong Kong, Thailand, Malaysia, and parts of the Philippines (41); that source does not cover Indonesia, and the previous version of this review extended it to Indonesia in error. Indonesian primary data instead report the allele in 22.9% of 236 healthy Javanese and Sundanese controls (carrier prevalence) (42), consistent with reported allele frequencies of ~11.6% in Javanese-Sundanese and 15.5% in Indonesian Malay participants (43) under Hardy–Weinberg conversion. *CYP2C19* poor metabolizers,

at increased risk of adverse cardiovascular events on standard clopidogrel dosing, occur at a phenotype frequency of ~14%-15% in East Asian populations against ~2% in Europeans (39,44). *CYP2D6*10*, affecting tamoxifen, codeine, antidepressants, and beta-blockers, reaches a median allele frequency of ~41% across East and Southeast Asian populations against ~2% in Europeans (40), with direct implications for breast cancer management through impaired endoxifen conversion.

These frequencies also reshape platform economics. As the fully loaded clinical cost of WGS (tier 3) falls toward that of array-based panels, the cost advantage of arrays erodes, and genome sequencing becomes attractive both because it resolves structural complexity in genes such as *CYP2D6* and because it produces a durable resource capturing pharmacogenomic, disease-risk, and population reference data simultaneously. In our judgment, the constraint then shifts to interpretation infrastructure: regional reference databases, bioinformatics capacity, and prescriber education.

Reproductive and carrier screening

Non-invasive prenatal testing (NIPT) uses ultralow-pass sequencing of cell-free fetal DNA (0.06-0.5× depth, 0.5-2 Gb), achieving >99% sensitivity for trisomy 21 at US\$200-US\$500 (P, tier 3) (45). Expanded carrier screening across 100-300 genes costs US\$250-US\$600 per individual (P, tier 3). Where carrier couples are identified, preimplantation genetic testing for monogenic disorders within an *in vitro* fertilization cycle (US\$7,000-US\$12,000, P) remains the most definitive preventive intervention. Alpha-thalassemia carrier prevalence reaches 40%-50% in specific subpopulations, with Vietnam highest at a pooled 51.5% (10), so the population arithmetic for couple-based screening is unusually favorable.

Pathogen genomics

Endemic tuberculosis, dengue, and malaria alongside documented zoonotic spillover vulnerability make pathogen genomics a health-security requirement. Clinical metagenomic sequencing offers culture-independent detection for cryptogenic meningitis, encephalitis, and sepsis; validated clinical implementations carry total costs of US\$1,000- US\$3,000 per sample (P, tier 3) once bioinformatics and expert interpretation are included (46). WGS of *Mycobacterium tuberculosis* isolates provides resistance profiling of direct relevance given rising drug-resistant disease, and rapid long-read sequencing supports field outbreak response and resolution of mobile resistance plasmids.

Notably, this is the one domain where several member states already possess functioning capacity built during the COVID-19 response. Cambodia, Laos, and Myanmar maintain pathogen sequencing infrastructure without any human clinical genomics service - instruments, laboratory space, trained operators, and sample logistics that represent the most realistic foundation on which human genomics could be layered.

Regional Demand

Following Timor-Leste's admission, the combined ASEAN population is ~700 million, the third-largest population bloc globally, with ~10.6 million annual births (8). The scale is large enough to justify dedicated regional infrastructure and large enough that no single member state can serve it alone.

Demand for rare and congenital conditions is estimated at ~3% of annual births, consistent with World Health Organization data indicating that 2%-5% of live births present with anomalies requiring genetic evaluation. Oncology demand is conservatively projected at 30% of incident cancers, the proportion expected to harbor actionable biomarkers [category D, derived from the 36.7% reported prospectively (38)].

GLOBOCAN 2022 recorded 1,146,810 new cancer cases across the 11 member states, with breast cancer most common in women in every member state and lung cancer leading in men in the Philippines, Malaysia, Myanmar, and Indonesia (9). Liver cancer dominates the Mekong subregion, driven by chronic hepatitis B prevalence. Absent effective intervention, ~2.03 million annual cases are anticipated by 2050, an increase of 89.2% in men and 65.6% in women relative to 2022 (9).

Running parallel is a hemoglobinopathy burden with few global parallels. Pooled alpha-thalassemia carrier prevalence reaches 51.5% in Vietnam, 39.5% in Cambodia, and 20.1% in Thailand, though studies concentrated in ethnic minority populations may inflate pooled estimates (10). Incorporating beta-thalassemia and hemoglobin E (HbE) trait, the number of individuals carrying at least one hemoglobinopathy trait is an author-derived estimate (category D) that varies substantially with the assumptions adopted and is therefore presented as three assumption sets rather than a single figure (Supplementary Table S10): applying the pooled meta-analytic rates unadjusted with a combined-trait uplift gives ~148-175 million, which implies national carrier prevalences we consider implausible, applying an explicit downward representativeness adjustment to those member states whose constituent studies concentrated on ethnic minority populations gives ~116-143 million; and a conservative set using the lower bound of country-specific rates with no combined-trait uplift gives ~83 million. The adjustment and uplift factors are authors' judgments rather than published corrections and are the largest single source of uncertainty in the estimate.

This aggregate is an indicator of disease burden and is not an input to the capacity model. Long-read thalassemia demand in Section "Long-read infrastructure" is derived instead from projected carrier-screening volumes multiplied by a carrier-positive rate and a reflex fraction, plus prenatal diagnostic volumes in at-risk pregnancies, as set out there and in Supplementary File S1. In Indonesia, with an alpha-thalassemia carrier prevalence estimated at 5%-10%, thalassemia major is among the most costly catastrophic diseases for the national insurance scheme, with lifelong transfusion and chelation reported at approximately IDR 800 billion = \$45,088,640 USD (1 USD = ~17,742 IDR per September 2026) annually (R).

Regionally, Southeast Asia carries among the highest thalassemia-related morbidity and mortality worldwide (47).

Actionable pharmacogenomic variation is similarly widespread, but the previous version of this review reported regional aggregates for *CYP2C19* metabolizer phenotypes and *HLA-B*15:02* carriers that were extrapolated in part from East Asian rather than Southeast Asian data. Those aggregates have been withdrawn. Country- and ancestry-specific frequencies, with the frequency type named and the member states for which no Southeast Asian data could be located identified as such, are given in Supplementary Table S5. Glucose-6-phosphate dehydrogenase deficiency, following historical malaria endemicity, affects a substantial further proportion. Collectively these describe a population in which a large fraction of routine prescribing - antiplatelet therapy, antiepileptics, antidepressants, tamoxifen, and primaquine - proceeds without genotype information known to be predictive and, in several cases, already subject to regulatory guidance.

Modeled Capacity Requirements

The model uses conservative penetration estimates and assumes that comprehensive coverage realistically spans a 10-15-year horizon. Rare disease targets reflect ~3%-6% of annual births plus backlog clearance; cancer profiling assumes that 20%-45% of incident cases eventually receive testing; prenatal testing scales from 5% to 76% of annual births. Phase 3 rates are scenario upper bounds attainable only where public reimbursement exists, not forecasts; the rationale, benchmark and source for each is given in Supplementary Table S8, together with an affordability check. Prenatal testing at the Phase 3 rate would cost ~US\$1.6-US\$3.2 billion per year (Section "Data volume, storage and recurrent cost"), equivalent to ~1.0%-2.0% of current health expenditure across the eleven member states (48), which implies reimbursement decisions of a scale that only some member states could presently contemplate; the constrained scenario in Section "Sensitivity analysis" therefore models prenatal testing at half the Phase 3 rate. Table 3 gives projected volumes.

Short-read infrastructure

Current ultrahigh-throughput (UHT) short-read platforms deliver ~1,500-4,000 Tb of sequence annually per instrument at 70% operational utilization, at reagent costs (tier 1) of ~US\$80-US\$200 per 30× genome. Conventional-class platforms deliver up to ~16 Tb per run from dual high-capacity flow cells, or up to ~7 Tb daily from four independent flow cells, at ~US\$150-US\$200 per genome (tier 1); wafer-based platforms are specified above 4,000 Tb annually at ~US\$80 per genome (tier 1). All are manufacturer-stated figures (category C), documented in Supplementary Table S4. The most recent chemistry, using mostly natural sequencing-by-synthesis on open wafer fluidics (49), is manufacturer-specified below US\$100 per genome at tier 1 but has not been independently validated at population scale. The only sustained real-world output published for a national

program in the region is ~3,000 genomes per month (13); the sensitivity analysis in Section “Sensitivity analysis” accordingly varies realized throughput down to 50% of specification.

Raw data volume alone understates hardware requirements, because turnaround times and batching govern operational efficiency. Under a fixed schedule of, for example, twice weekly per application, a dynamic queue classifying samples as urgent or non-urgent allows clinical samples to displace population screening to the next run, preserving diagnostic turnaround without partially filled runs that inflate per-test cost. Because routine samples are displaced unpredictably, a hardware buffer is required; the model sets it at 30% of the throughput-derived fleet (range 25%-35%), which at Phase 3 corresponds to 6-8 instruments. This buffer is a first-order determinant of fleet size and is frequently omitted from estimates built on throughput alone.

Where the fully loaded clinical cost of a 30× genome (tier 3) approaches the cost of a targeted panel, the economics favor centralized sequencing with decentralized bioinformatics under a “sequence once, query often” model. The regional short-read portfolio is therefore specified in three parts, stated here by phase and by platform class so that it agrees with Table 4. At Phase 3, throughput requires ~23 conventional-class ultrahigh-throughput (UHT) units (class-range 21-26) or, equivalently, ~10 wafer-class units, for population sequencing, large-batch exomes and prenatal testing; a scheduling buffer of seven further units absorbs queueing variability; and 12 mid-throughput units distributed across hubs serve hyper-acute needs for which even a twice-weekly schedule is insufficient, such as neonatal rapid sequencing and urgent oncology profiling. At Phase 2, the corresponding figures are nine conventional-class or four wafer-class units, three buffer units, and eight mid-throughput units; the previous version of this review quoted the Phase 2 wafer-class figure alongside the Phase 3 conventional-class figure without stating either phase or class, which is the source of the apparent inconsistency between Section “Short-read infrastructure” and Table 4. Cloud infrastructure replaces distributed genotyping arrays for applications servable computationally at minimal marginal cost.

Long-read infrastructure

Long-read sequencing is a required regional capability for structurally complex loci, principally in thalassemia diagnostics, complex pharmacogenomic genotyping, and methylation-informed tumor classification. It is not proposed as the primary platform for every relevant application. gap-PCR, reverse dot-blot hybridization, multiplex ligation-dependent probe amplification, targeted genotyping and orthogonal confirmation remain appropriate, validated first-line assays for known variants, and a large share of thalassemia testing is servable by targeted panels of approximately 10 Gb rather than by whole genomes. The role modeled here is threefold: reflex or second-line testing where conventional or short-read results are inconclusive, discordant with phenotype, or suggestive of a complex rearrangement; primary

testing for the specific indications in which head-to-head comparative evidence favors it, namely prenatal diagnosis of thalassemia (50) and *CYP2D6* star-allele phasing; and construction of the long-read subset of the regional reference panel.

High-fidelity circular consensus platforms currently deliver approximately 250-300 Tb annually per instrument at approximately US\$1,000-US\$1,300 per 30× genome (tier 1, C), with announced chemistry targeting approximately US\$250; high-capacity nanopore platforms deliver comparable annual genome counts at ~US\$500-US\$800 per genome (tier 1, C). High-fidelity systems generate ~400-480 Gb per four-cell run, and high-capacity nanopore systems up to ~14 Tb per run across 48 flow cells, a theoretical maximum that real-world operation does not reach. These are manufacturer specifications and have not been independently validated at population scale.

Long-read demand at Phase 3 is derived from the same volume model rather than assumed, and totals approximately 497,691 tests per year (category D). Its components are thalassemia reflex testing arising from carrier screening, calculated as carrier-screening volume multiplied by a 20% carrier-positive rate and a 25% reflex fraction, ~250,000 tests; prenatal diagnosis in at-risk pregnancies, calculated as annual births multiplied by a 1.5% at-risk-couple rate and 30% uptake, ~47,549 tests; *CYP2D6* structural phasing at 5% of pharmacogenomic tests, ~150,000 tests; rare-disease reflex genomes at 12% of exome and genome tests, ~36,000 tests; and central nervous system tumor classification by direct methylation detection at 1.3% of incident cancers, ~14,141 tests. The reflex and uptake fractions are authors’ assumptions informed by the comparative studies cited above because a larger share of cases may be served by targeted assays, Section “Sensitivity analysis” reports a scenario in which long-read demand is halved.

A mixed fleet is indicated, with high-fidelity platforms for clinical-grade accuracy in rare disease and prenatal diagnosis and nanopore platforms for scalable targeted screening. Table 4 reports the two single-platform alternatives and the proposed mixed configuration separately because the previous version’s notation implied incorrectly that the two platform counts were additive.

Data volume, storage and recurrent cost

Two distinct quantities were previously reported with the same symbol. This review now uses Pb (petabase) exclusively for sequence generated and PB (petabyte) exclusively for storage retained. At Phase 3, maturity the region would generate ~37.8 Pb of short-read and 8.2 Pb of long-read sequence annually, ~46.0 Pb in total.

Sequence generated is not storage retained. The retention policy modeled here, set out in full in Supplementary Table S3, is as follows: raw instrument output - base-call and signal files and FASTQ - is deleted after alignment and quality-control sign-off; whole-genome and exome data are retained as compressed reference-oriented alignment map with quality-score binning, at ~9 GB

per 30× genome and 1.2 GB per exome; long-read data are retained as aligned binary alignment map with base-modification tags but without kinetics, at approximately 30 GB per rare-disease genome and 0.5 GB per targeted panel; high-volume screening assays - prenatal, carrier and pharmacogenomic panels - retain only variant-call files and reports after a 90-day hold, because their raw reads have no re-analysis value proportionate to their storage cost; and every retained object is held in two copies, active and archival. Applied to the Phase 3 volumes in Table 3, these assumptions yield approximately 6.6 PB added per year (3.3-8.2 PB across the scenarios in Section “Sensitivity analysis”) and an accumulated 61 PB by Year 15. Retention policy is a governance decision with large cost consequences and is among the parameters varied in Section “Sensitivity analysis”.

Fully loaded recurrent testing expenditure (tier 3) at Phase 3 is modeled at ~US\$3.7-US\$8.5 billion annually across ~32.8 million testing events, of which ~17.8 million are new sequencing events and ~15 million are bioinformatic queries against previously generated whole-genome data at ~US\$5-US\$10 each (M). Testing events are not unique individuals: allowing for an assumed 40% overlap between tissue profiling, liquid biopsy and hereditary panel testing in the same patients, and noting that long-read reflex tests arise from carrier-screening and prenatal pathways already counted as events, the corresponding number of unique individuals tested is ~17.1 million per year (category D). Capacity and recurrent cost are driven by events, so the fleet and expenditure figures are unaffected by this distinction; only statements about the population served depend on it. The largest single components are prenatal testing at ~US\$1.6-US\$3.2 billion and cancer profiling at ~US\$0.8-US\$1.9 billion, with per-test costs of US\$200-US\$400 and US\$500-US\$2,500, respectively (P, tier 3).

For a population of 700 million this corresponds to ~US\$5.3-US\$12.2 per capita annually in recurrent expenditure, which is distinct from, and roughly two orders of magnitude larger than, the infrastructure and capital investment of ~US\$0.07 per capita annually reported in Section Implementation roadmap and investment: the former is the cost of delivering the modeled testing volumes and is borne by health systems, payers, and patients; the latter is the cost of building the shared infrastructure that makes the delivery possible. Against current health expenditure ranging from ~US\$69 per capita (Laos) to more than US\$4,000 (Singapore) (48), recurrent genomic services at maturity would represent ~2.3%-5.3% of combined regional health expenditure - material, and prohibitive for the lowest-spending member states without external financing, as the member-state breakdown in Supplementary Table S8 shows.

These costs should continue falling as sequencing prices decline, historically exceeding 30% annually with substantial year-to-year variability, and as regional procurement compresses per-test pricing. With tier 1 reagent costs approaching US\$100, the dominant constraint shifts away from sequencing hardware toward clinical interpretation, genetic counseling, and the cloud

infrastructure sustaining a “sequence once, query often” model - a shift the cost structure reported in published microcosting supports, with consumables accounting for 68%-72% of total cost today (11).

Sensitivity analysis

Because the model's outputs depend on assumptions that cannot be verified in advance, five drivers were varied one at a time around the base case, with all other inputs held constant: operational utilization, 50%-85% (base 70%); realized per-instrument throughput, 50%-100% of manufacturer specification; Phase 3 clinical penetration, -50% to +25% for every application in Table 3; raw data yield per test, ±25%, with prenatal testing additionally run at both ends of its 0.5-2 Gb range; and unit cost, comprising instrument list prices ±20%, service contracts at 8%-12% of installed list price per year, and reference-panel production at US\$800-US\$3,000 per genome. Three combined scenarios were then run: *Base*; *Constrained* (penetration -50%, utilization 55%, realized throughput 60%, prices +20%); and *Accelerated* (penetration +25% with prenatal testing capped at 90% of births, utilization 80%, realized throughput 100%, prices -20%). A further scenario halves long-read demand. Full results are in Supplementary Table S2, a tornado plot of the Phase 3 outputs is Supplementary Figure S1, and the calculations are live formulas on sheet 10_Sensitivity of Supplementary File S1, so that readers may vary the inputs further.

The Phase 3 outputs behave as follows. Sequence generated ranges from 23.0 Pb (Constrained) to 56.8 Pb (Accelerated) against a base case of 46.0 Pb and is driven almost entirely by penetration and data yield. Instrument counts are driven principally by realized throughput and utilization: across the one-way runs the throughput-derived conventional-class fleet spans 12-46 units and the mixed long-read fleet 12-43 units, against base cases of 23 and 22. Because penetration and realized throughput move fleet size in opposite directions, the combined Constrained and Accelerated scenarios both give a total short-read fleet of ~45 units, close to the base case of 42 - an instructive result, since it means that the fleet estimate is more robust than any single driver suggests, while remaining sensitive to realized throughput alone. Sequencing hardware capital at Phase 3 spans US\$33-110 million (base US\$57 million), and annual storage added spans 3.3-8.2 PB (base 6.6 PB). The fifteen-year investment envelope spans US\$644-US\$1,019 million (base US\$733 million), and mature recurrent expenditure spans US\$2.7-US\$14.9 per capita per year (base US\$5.3-US\$12.2). Halving long-read demand reduces the mixed long-read fleet to 12 units and the Phase 3 hardware capital to US\$49 million and leaves recurrent expenditure and the investment envelope substantially unchanged.

Two limitations of this analysis should be stated. The drivers are varied deterministically rather than probabilistically, so the reported ranges are scenario bounds and not confidence intervals; and correlations between drivers - for example between procurement price and realized throughput - are not modeled except in the three combined scenarios.

A Federated Hub-and-Spoke Architecture

No single member state can build comprehensive services for 700 million people. A workable framework must reconcile national sovereignty over sensitive genetic data with the efficiency of shared infrastructure, common reference resources, and integrated clinical networks. Three precedents inform the design: H3Africa, which invested ~US\$176-US\$180 million (R) across 51 projects in 30 countries over a decade (51); the European 1+Million Genomes Initiative, in which 27 signatory countries adopted federated rather than centralized data architecture (52); and the GenomeAsia 100K Project, whose pilot phase demonstrated both the feasibility and the analytical value of a pan-Asian reference dataset spanning 1,739 individuals from 219 population groups, while also illustrating how thinly Southeast Asian populations are sampled in such efforts (53).

Three capability tiers

Tier 1, regional hubs (Singapore, Thailand, Malaysia, Indonesia). These possess the regulatory frameworks, clinical workforces, and established programs to serve as reference centers, each hosting 5-8 ultrahigh-throughput (UHT) and long-read instruments and providing economies of scale for population sequencing, complex variant interpretation, training, and quality assurance.

Tier 2, developing hubs (Vietnam, Philippines). Positioned to reach primary hub status within 3-7 years, operating 2-4 mid-throughput instruments for time-sensitive panels with population screening batch-routed to Tier 1.

Tier 3, spoke countries (Myanmar, Cambodia, Laos, Brunei, Timor-Leste). Focused on sample collection, deep phenotyping, telegenetics consultation, and point-of-care testing, with samples requiring comprehensive profiling transported to hubs and secure cloud portals providing access to centrally generated data.

Tier assignment reflects current capability, not permanent classification, and the architecture is designed to allow promotion. Its principal risk is entrenching an extractive relationship in which spoke countries supply samples while hubs accumulate data, expertise, and publications; this is addressed in Section “Public-private partnership and safeguards”.

Governance and the reference panel

The proposed ASEAN Genome Consortium provides the governance architecture this network requires (6), with terms of reference covering interoperable data standards, selective sharing, co-developed polygenic risk scores, harmonized ethical frameworks, and workforce development. Anchoring it within ASEAN Health Cluster 3 would allow it to draw on the ASEAN Health Ministers' Meeting and the ASEAN Centre for Public Health Emergencies and Emerging Diseases.

The Consortium should adopt federated data architecture modeled on the European 1 + MG initiative (52), under which genomic data remain physically and legally resident within each member state but are queryable

through distributed analysis frameworks. This addresses the sovereignty concerns that have historically impeded cross-border sharing while enabling the joint analysis that rare variant interpretation requires.

Given more than 1,000 ethnic groups and correspondingly high allelic diversity, we propose an ASEAN Reference Panel of at least 50,000 diverse genomes. Three considerations set that target, with the arithmetic in Supplementary Table S11. First, allele-frequency estimation: 50,000 genomes provide 100,000 chromosomes, so a variant at 0.1% frequency is expected ~100 times, giving a relative standard error of the frequency estimate of ~10%, and a variant at 0.01% is expected approximately 10 times, with a probability above 0.9999 of at least one observation. This is the frequency range in which variants of uncertain significance are most often reclassified, and it is the range in which European-derived panels are least informative for Southeast Asian patients (1,2). Second, population structure: the panel is stratified as approximately 35,000 population-weighted genomes plus ~15,000 genomes oversampling the smaller ethnic groups, averaging 75 genomes per group across more than 200 groups with a floor of 100-250 per group, the range in which population-specific reference data have been shown to improve imputation accuracy for previously unrepresented groups (4,53). Third, precedent: the target is of the same order as Singapore's completed Phase II cohort (13) and the national contributions to the European initiative (52). The limits of the target should also be stated: 50,000 genomes support frequency estimation, variant classification and imputation for the major populations, but not rare-variant association discovery, for which the clinical data accumulating under the sequence-once, query-often model are the intended resource.

Public-private partnership and safeguards

Every national program in the region depends on private-sector participation, and the terms of that participation will largely determine whether regional capability is built or merely rented. Four arrangements are already visible: corporate philanthropy funding instrument placement through non-profit intermediaries; vendor-government instrument partnerships; outsourced service provision under contract; and commercial clinical laboratory networks operating across multiple jurisdictions ahead of any public program.

These have been overwhelmingly transactional: instrument placement paired with reagent supply transfers capital equipment without transferring capability. Three consequences follow. Vendor lock-in arises where proprietary consumables, closed data formats, and platform-specific pipelines make later procurement effectively irreversible, eliminating the price competition on which the cost projections above depend. Skills asymmetry persists where operational expertise remains with vendor field scientists rather than accumulating locally, so capability leaves when the contract ends. And value asymmetry emerges where samples and phenotype data flow toward commercial partners while intellectual property and publications flow away from contributing

countries, the pattern H3Africa was designed to interrupt (51).

Partnership agreements under a regional framework should therefore carry five standing provisions: training mandates with certified, contractually specified outputs rather than discretionary vendor goodwill; progressive technology transfer and local content milestones tied to procurement volume; data portability guarantees including open output formats, third-party reanalysis rights, and prohibition of exclusive consumable-supply clauses; explicit benefit-sharing terms allocating intellectual property, publication rights, and reciprocal access for spoke-country institutions; and pricing transparency across the Consortium.

The largest available financial lever is demand aggregation. Member states negotiating separately for small fleets cannot access the volume pricing underpinning a sub-US\$100 reagent cost; a consortium negotiating for 700 million people can. Pooled procurement is well established for other health commodities, though we are not aware of a published evaluation specific to sequencing consumables.

Possible financing sources (not evaluated). No single source is likely to fund an investment of the magnitude modeled in Section “Implementation roadmap and investment”. Multilateral development bank lending, bilateral assistance, national budgets, philanthropy, and private investment in service delivery are each plausible components, and concessional capital is conventionally better matched to long-payback infrastructure than commercial capital is. We evaluated none of these options in this review: no financing model was tested, the sequencing of concessional and commercial capital described here is illustrative, and the composition of any actual capital stack would depend on member-state fiscal space, debt sustainability, and lender appetite. These are proposals for consideration, not findings.

Workforce and regulation

In our judgment, the most likely binding constraint is human rather than technological, although this is a judgement and not a demonstrated result: no regional workforce census exists and quantified data are available for only two member states. Three observations support it. First, the two quantified specialist ratios - approximately one clinical geneticist per 11.7 million people in the Philippines and one per 6.8 million in Vietnam (25-27) - fall an order of magnitude short of any published service norm, whereas the sensitivity analysis in Section “Sensitivity analysis” varies hardware requirements only within a factor of approximately four across all one-way runs. Second, lead times are asymmetric: specialist training requires roughly 4-10 years, against months for instrument procurement, installation, and validation. Third, hardware can be redeployed between applications and member states; trained specialists cannot be redeployed at the same speed and are subject to outward migration.

The scale of the requirement is derived in Supplementary Table S6, and the derivation is reproduced on sheet

11_Workforce of Supplementary File S1. Bioinformatics staffing is derived from the annual volume of whole-genome-equivalents analyzed - Phase 3 whole genomes and long-read genomes at full weight, exomes at 0.25, cancer tests at 0.05 and targeted long-read tests at 0.02, giving ~245,234 whole-genome-equivalents per year - divided by an assumed 500 whole-genome-equivalents per full-time equivalent per year, which yields ~490 bioinformaticians. Genetic counseling staffing is derived from the annual number of results requiring pre- or post-test counseling: rare-disease diagnoses at a 35% yield, hereditary-cancer positives at 10%, at-risk carrier couples at 2% of screened couples, and high-risk prenatal results at 0.5%, totaling ~225,000 episodes per year; at an assumed caseload of 250 episodes per counselor per year this implies ~900 counselors across all member states, of which the Consortium-level increment for hub-based and reference functions is ~180. Clinical geneticist staffing is derived from complex cases requiring specialist review at hubs, ~66,000 per year at an assumed caseload of 650, giving ~102 posts.

A regional strategy proportionate to the capacity modeled above would therefore need to train on the order of 500 bioinformaticians, 200 genetic counselors, and 100 clinical geneticists within a decade. These are Consortium-level increments needed to operate the modeled regional network - hubs, reference functions, and interpretation - and not the total national workforce that a population norm would imply. For comparison, Thailand's national target of 30 clinical geneticists, 100 genetic counselors, 100 molecular geneticists, and 500 bioinformaticians over five years indicates the order of investment a single middle-income member state considers necessary for a 50,000-genome program (14). Delivery would run through an ASEAN Genomics Fellowship with rotations across Tier 1 hubs, a regional e-learning platform modeled on H3ABioNet, which has trained more than 2,000 researchers across Africa since 2012 (54), and structured hub-spoke university partnerships.

Distribution matters as much as aggregate. Training concentrated in Tier 1 hubs will not by itself produce capacity in Tier 3 countries and may accelerate outward migration of the few specialists they have. Return-of-service commitments, dual appointments, and remote clinical supervision are more likely to build distributed capability than open-competition scholarships.

Regulatory harmonization should build on the ASEAN Medical Device Directive, which classifies genetic syndrome test kits as Class C (55). An ASEAN-specific annex covering next-generation sequencing diagnostics, laboratory-developed tests, and pharmacogenomic prescribing guidance would supply the certainty required for both clinical implementation and sustained investment. Two further elements are prerequisites for shared infrastructure: mutual recognition of laboratory accreditation, so that a sample collected in one member state can be sequenced in another without duplicate validation; and common variant classification standards applied to a shared reference panel, preventing the same variant being reported as pathogenic in one member state and uncertain in another.

Implementation Roadmap and Investment

The fifteen-year investment envelope is built from the bottom up in Table 5, with one row for each capital component, three phase columns, the basis of each estimate, and its evidence category. The supporting arithmetic is on sheet 9_Investment of Supplementary File S1. Two rules prevent double counting. Sequencing hardware is taken from Table 4 and converted from fleet-at-phase to cumulative purchases, with one replacement cycle in Phase 3 for instruments installed in Phase 1. Recurrent clinical testing - reagents and labor for routine service delivery, modeled separately in Section “Data volume, storage and recurrent cost” - is excluded entirely, and appears in Table 5 only as an explicit exclusion row, because it is an operating cost borne by health systems rather than a capital investment.

The bottom-up total is ~US\$733 million over 15 years, with a scenario range of US\$644 million to US\$1.0 billion (Section “Sensitivity analysis”). This replaces the US\$750 million to US\$1.4 billion range asserted in the previous version of this review, which was not derived; the revised figure is lower at the upper bound principally because the earlier range was not constrained by a component-level build-up. Sequencing hardware is only ~9.0% of the envelope. The three largest components are data infrastructure (US\$127 million), reference-panel construction (US\$88 million), and workforce and training (US\$85 million) - which is to say that the roadmap is predominantly an investment in storage, computation, reference resources, and people rather than in sequencers.

Phase 1 (years 1-3), ~US\$134 million. The Consortium is established under ASEAN Health Cluster 3 with Timor-Leste included from inception. Pipelines are standardized across Tier 1 hubs through deployment of initial ultrahigh-throughput (UHT) and clinical long-read systems, and a federated data architecture pilot links three to five member states. An initial reference panel of 5,000 diverse genomes is constructed, pilot programs launch in thalassemia prevention and cancer pharmacogenomics, and fellowship training begins at 50 fellows annually.

Phase 2 (years 4-7), ~US\$286 million. Vietnam and the Philippines reach Tier 2 status and sovereign national biobanks are established. The reference panel is completed at 50,000 genomes across more than 200 ethnic groups, a sub-US\$100 reagent-cost standard (tier 1) is consolidated through volume procurement, pharmacogenomics enters national formularies in hub member states, and prenatal testing scales to ~30% of births under harmonized *in vitro* diagnostic regulation.

Phase 3 (years 8-15), ~US\$313 million. Genomic testing is integrated into routine care across all 11 member states, with population screening driven by centralized sequencing and local cloud querying. Regional precision oncology becomes operational, pharmacogenomics-guided prescribing becomes standard practice in hub member states, and a localized diagnostics industry reduces reliance on imported reagents.

Spread across 15 years and 700 million people, the envelope corresponds to approximately US\$0.07

per capita annually (range US\$0.06-US\$0.10) in infrastructure and capital investment only, excluding the recurrent clinical testing expenditure modeled in Section “Data volume, storage and recurrent cost”. More consequentially, this investment should be assessed net of the expenditure it displaces - avoidable adverse drug reactions, prolonged diagnostic odysseys, and lifelong management of severe thalassemia. The magnitude of that offset is the single most important open question in the economic case, and it is not one this review can settle: no cost-effectiveness analysis was performed here. Formal cost-effectiveness analyses conducted in Southeast Asian health system contexts, using local unit costs and carrier frequencies, are largely absent and represent the highest-value near-term research contribution available in this field.

Limitations

This is a narrative rather than systematic review: it was not registered, no formal risk-of-bias assessment was applied, and study selection involved author judgment. The formal search closed on March 31, 2026, and the subsequent update was targeted rather than a repetition of the full search, so peer-reviewed literature published between April and August 2026 outside the specific topics updated may not be represented.

A substantial proportion of the capacity data derives from ministry announcements and manufacturer documentation rather than peer-reviewed sources. These are the only sources in which such information exists, but they are not independently verified, may be promotional, and become stale quickly; several figures, notably completed output against announced national targets, could not be confirmed against published program reporting, and are reported in Table 1 as “not reported” rather than inferred.

Platform throughput and cost figures are manufacturer specifications (category C) without independent validation at population scale, and real-world throughput routinely falls below specification once sample quality, run failures, and quality control are accounted for. The 70% utilization assumption absorbs part of this but is itself an assumption; Section “Sensitivity analysis” accordingly varies realized throughput down to 50% of specification, and the resulting fleet requirements are substantially larger. Clinical penetration rates, test mix, and phase timings are scenario assumptions rather than forecasts. The sensitivity analysis varies drivers deterministically and one at a time, apart from three combined scenarios, so its ranges are scenario bounds and not confidence intervals.

The aggregate hemoglobinopathy estimate is an author-derived calculation extrapolated from published frequency data and incorporating subjective adjustment factors; it is presented as three assumption sets, is not an input to the capacity model, and should not be cited as a measured prevalence. Alpha-thalassemia carrier prevalence for six member states remains provisional. Pharmacogenomic frequencies are reported by country and ancestry where published data exist and are marked as absent where they do not; regional aggregates

extrapolated from East Asian data have been removed. No regional clinical genetics workforce census exists, so the workforce conclusion rests on data from two member states and on the derivation in Supplementary Table S6.

One author is an employee and shareholder of a manufacturer of long-read sequencing instruments. That interest could in principle have influenced the platform parameters adopted, the derivation of long-read demand, and the prominence given to long-read reflex testing in the conclusions. Several safeguards reduce that possibility: the complete model is published with every input and formula; the model was independently re-executed by authors without manufacturer affiliation; product names have been removed from the main text and tables, with class parameters traceable to the public documents listed in Supplementary Table S4; the long-read conclusion is stated as a capability requirement with validated alternatives named; and Section “Sensitivity analysis” reports a scenario in which long-read demand is halved. These safeguards reduce but cannot eliminate the possibility of unconscious influence, and readers should weigh the long-read findings with that in mind.

Ethical, legal, and social implications, consent frameworks, genetic discrimination protections, and public acceptability are addressed only insofar as they affect implementation feasibility, though each could independently determine political viability. Environmental costs of large-scale sequencing and data storage are not addressed.

Conclusions

What the evidence shows. Capacity spans from more than 50,000 completed genomes in Singapore to none in Timor-Leste. The region recorded ~1.15 million incident cancers in 2022, projected to approach 2 million by 2050 on demographic growth alone (9). The two member states that have enumerated their specialist workforce report approximately one clinical geneticist per 6.8 and 11.7 million people (25-27). For structurally complex loci, long reads outperform conventional methods: in 278 at-risk fetuses they detected 7.9% more clinically significant variants than PCR-based methods and corrected the predicted phenotype in eight cases (50).

What the scenario model indicates. Regional coverage at maturity would generate ~46.0 Pb annually (23.0-56.8 Pb across scenarios), requiring ~42 short-read and 22 long-read instruments, at ~US\$733 million over 15 years (US\$644 million to US\$1.0 billion), or US\$0.07 per capita annually excluding recurrent testing. Three implications follow, each a scenario result rather than a measured finding: long-read sequencing is a required regional capability for complex loci, deployed alongside validated targeted assays; specialist workforce is, in our judgment, the most acute constraint, since hardware requirements vary within a factor of four across scenarios while the specialist gap exceeds any published norm by an order of magnitude; and most of the envelope belongs to interpretation, storage, reference resources and people rather than to sequencers. Whether the expenditure is justified by the health gain is a question for the cost-effectiveness analyses this region still lacks.

List of Abbreviations

ASEAN	Association of Southeast Asian Nations
CGP	Comprehensive genomic profiling
gap-PCR	Gap polymerase chain reaction
Gb	gigabase
GB	gigabyte
HbE	hemoglobin E
HiFi-class	High-fidelity circular-consensus long-read platform class
NGS	next-generation sequencing
NIPT	non-invasive prenatal testing
OMIM	Online Mendelian Inheritance in Man
Pb	petabase
PB	petabyte
PRECISE	Precision Health Research Singapore
Tb	terabase;
UHT	ultra-high-throughput
VN1K	1,000 Vietnamese Genomes Project
WES	whole-exome sequencing
WGS	whole-genome sequencing.

Conflict of interests

J.K. is an employee and shareholder of a commercial manufacturer of long-read sequencing instruments whose platform class is discussed in this review. V.S.W.B., S.J., and C.O.C. are affiliated with a non-profit foundation that operates one of the national genomic networks described in Section “The genomic landscape across eleven member states”. Z.T. declares no competing interests. Named, individual disclosure statements corresponding exactly to these five positions are provided on the separate title page. No manufacturer had any role in the conception, design, analysis, drafting or decision to submit this work, and no manufacturer reviewed the manuscript before submission. Six measures limit the influence of the declared interest on the content. The main text and all tables refer to platform classes rather than named commercial products; named products appear only in Supplementary Table S4 and in the reference list, as the documentary sources of class-level parameters. Every throughput and cost parameter is identified as manufacturer-stated (category C) and, in Section “Limitations”, as not independently validated at population scale. The capacity model is technology-neutral, expressing requirements as instrument-class counts rather than named-product procurement recommendations, and the review makes no comparative product recommendation. The complete model is published as Supplementary File S1 and was by C.O.C., whose outputs matched the reported figures to within rounding. Section “Sensitivity analysis” reports a scenario in which long-read demand is halved. Section “Limitations” states that these safeguards reduce but cannot eliminate the possibility of unconscious influence. Author 3 contributed to design, definition of intellectual content and manuscript editing, and did not select the numerical input values for any platform class, which were taken from the public documents listed in Supplementary Table S4 by authors without manufacturer affiliation.

Funding

No external grant funding was received for this specific manuscript.

Consent for publication

Not applicable.

Consent for participation

Not applicable.

Ethical approval

Not applicable. This review involved no human participants, human material, or human data collected by the authors.

Author contributions

Concepts: VSWB. Design: VSWB, JK. Definition of intellectual content: SJ, JK. Literature search: VSWB, SJ. Data acquisition: VSWB. Data analysis: SJ. Independent re-execution of the model: COC. Manuscript preparation: VSWB. Manuscript editing: SJ, JK, COC, ZT. Manuscript review: COC. Guarantor: VSWB.

Data availability

All data supporting this review derive from published literature, publicly available demographic and epidemiological databases, and publicly available manufacturer documentation, as cited. No new primary data were generated. The complete calculation framework underlying every quantitative statement in Sections “Regional demand,” “Modeled capacity requirements,” “A federated hub-and-spoke architecture,” and “Implementation roadmap and investment” is published as Supplementary File S1, in which all intermediate values are live formulas and every figure quoted in the text is traceable to a named cell; the supplementary tables listed in the supplementary material section give the derivations, sources, and audits on which it rests.

Author details

Vincentius Simeon Weo Budhyanto¹, Stefanie Jürgens¹, Jonas Korlach², Cindy Oktavi Cicilia¹, Zuhrah Taufiq³

1. Satriabudi Dharma Setia Foundation (YSDS), Jakarta, Indonesia
2. Pacific Biosciences of California, Inc., Menlo Park, California, USA
3. Faculty of Medicine, Negeri Padang University, Padang, Indonesia

References

1. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161–4. <https://doi.org/10.1038/538161a>
2. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. 2016;375:655–65. <https://doi.org/10.1056/NEJMs1507092>
3. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584–91. <https://doi.org/10.1038/s41588-019-0379-x>
4. Ardiansyah E, Riza AL, Dian S, Ganiem AR, Alisjahbana B, Setiabudiawan TP, et al. Sequencing whole genomes of the West Javanese population in Indonesia reveals novel variants and improves imputation accuracy. *Front Genet*. 2024;15:1492602. <https://doi.org/10.3389/fgene.2024.1492602>
5. Tumonggor MK, Karafet TM, Hallmark B, Lansing JS, Sudoyo H, Hammer MF, et al. The Indonesian archipelago: an ancient genetic highway linking Asia and the Pacific. *J Hum Genet*. 2013;58:165–73. <https://doi.org/10.1038/jhg.2012.154>
6. Sasongko TH, Gunadi, Hoh BP, Minari J, Seow SW, Shotelersuk V. The ASEAN genome consortium: advancing equitable precision medicine through regional solidarity. *Lancet Reg Health West Pac*. 2025;60:101611. <https://doi.org/10.1016/j.lanwpc.2025.101611>
7. ASEAN Secretariat. Forging a new era: Timor-Leste admitted into ASEAN. Jakarta: ASEAN Secretariat; 2025 [cited 2026 Aug 5]. Available from: <https://asean.org/forging-a-new-era-timor-leste-admitted-into-asean/>
8. United Nations, Department of Economic and Social Affairs, Population Division. World population prospects 2024: summary of results. New York: United Nations; 2024 [cited 2026 Aug 5]. Available from: <https://population.un.org/wpp/>
9. Dee EC, Ang CDU, Ting FIL, Yu AV, Jacomina LE, Alcasabas AP, et al. Cancer incidence and mortality estimates in 2022 in Southeast Asia: a comparative analysis. *Lancet Oncol*. 2025;26:516–28. [https://doi.org/10.1016/S1470-2045\(25\)00017-8](https://doi.org/10.1016/S1470-2045(25)00017-8)
10. Hussin NH, Saidin B, Johan MF, Hussein AR, Mohd Nor NS, Zilfalil BA, et al. Prevalence of alpha-thalassemia in Southeast Asia (2010–2020): a meta-analysis involving 83,674 subjects. *Int J Environ Res Public Health*. 2020;17:7354. <https://doi.org/10.3390/ijerph17207354>
11. Schwarze K, Buchanan J, Fermont JM, Dreau H, Tilley MW, Taylor JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med*. 2020;22:85–94. <https://doi.org/10.1038/s41436-019-0618-7>
12. Bylstra Y, Davila S, Lim WK, Wu R, Teo JX, Kam S, et al. Implementation of genomics in medical practice to deliver precision medicine for an Asian population. *NPJ Genom Med*. 2019;4:12. <https://doi.org/10.1038/s41525-019-0085-8>
13. Wong E, Bertin N, Hebrard M, Tirado-Magallanes R, Bellis C, Lim WK, et al. The Singapore National Precision Medicine Strategy. *Nat Genet*. 2023;55:178–86. <https://doi.org/10.1038/s41588-022-01274-x>
14. Shotelersuk V, Tongsimas S, Pithukpakorn M, Mahasirimongkol S, Eu-ahsunthornwattana J, Palittapongarnpim P, et al. Advancing precision public health at Genomics Thailand. *Nature Health*. 2026;1:461–4.
15. Tassaneeyakul W, Tiamkao S, Jantararoungtong T, Chen P, Lin SY, Chen WH, et al. Association between HLA-B*1502 and carbamazepine-induced severe cutaneous adverse drug reactions in a Thai population. *Epilepsia*. 2010;51:926–30. <https://doi.org/10.1111/j.1528-1167.2010.02533.x>
16. Tiwattanont K, John S, Koomdee N, Jinda P, Rachanakul J, Jantararoungtong T, et al. Implementation of HLA-B*15:02 genotyping as standard-of-care for reducing carbamazepine/oxcarbazepine induced cutaneous adverse drug reactions in Thailand. *Front Pharmacol*. 2022;13:867490. <https://doi.org/10.3389/fphar.2022.867490>
17. Ministry of Health of the Republic of Indonesia. Menteri kesehatan tinjau fasilitas BGSi dan uji klinik di Gedung Eijkman RSCM. Kementerian Kesehatan RI; 2025. Available from: <https://www.kemkes.go.id/id/menteri-kesehatan-tinjau-fasilitas-bgsi-dan-uji-klinik-di-gedung-eijkman-rscm>

18. Ministry of Health of the Republic of Indonesia. Laporan kinerja Kementerian Kesehatan tahun 2024. Kementerian Kesehatan RI — Laporan Kinerja 2024; 2025. Available from: <https://repository.badankebijakan.kemkes.go.id/id/eprint/5852/1/LKj%20Kemenkes%202024.pdf>
19. National Institutes of Biotechnology Malaysia. Majlis Apresiasi MyGenom. NIBM — MyGenom programme page; 2026. Available from: https://nibm.my/v6/?page_id=15180
20. Tran T.T.H, Hoang TH, Tran MH, Pham TM, Nguyen NN, Vu GM, et al. VN1K is a pangenome-informed multi-omics and phenomics resource for the Vietnamese population. *Nat Commun*. 2026;17(1):6679. <https://doi.org/10.1038/s41467-026-75375-0>
21. Philippine Genome Center. DNA sequencing. University of the Philippines; n.d. Available from: <https://pgc.up.edu.ph/services/dna-sequencing/>
22. Ministry of Finance and Economy Brunei Darussalam. Ministry of Health Brunei receives ancillary laboratory items and equipments for COVID-19 genome sequencing machine. Ministry of Finance and Economy Brunei Darussalam; 2021. Available from: <https://www.mofe.gov.bn/Press%20Room/Press%20Release%2017.11.2021%20%5BEnglish%5D%20-%20Equipment%20for%20Genome%20Sequencing%20Machine%20from%20BIA.pdf>
23. Institut Pasteur du Cambodge. Annual report 2022. Institut Pasteur du Cambodge; 2022. Available from: <https://www.pasteur-kh.org/wp-content/uploads/2023/05/20230511-Rapport-Annuel-2022-FINAL.pdf>
24. World Health Organization. Timor-Leste: does the country have genomic sequencing capacity? No. WHO South-East Asia Region. Timor Leste. Available from: <https://iris.who.int/server/api/core/bitstreams/f1f3754e-c245-450b-aa26-b650c7290d96/content>
25. Padilla CD, Cutiongco-de la Paz EM, Ngelangel CA. Genetics and genomic medicine in the Philippines. *Mol Genet Genomic Med*. 2016;4:354–72. <https://doi.org/10.1002/mgg3.247>
26. Alcausin MMLB, Abacan MAR, Cutiongco-de la Paz EM, Silao CLT, Chiong MAD, Padilla CD, et al. Integrating genetic services in the Philippine public health delivery system: the value of networks. *Genes (Basel)*. 2024;15:780. <https://doi.org/10.3390/genes15060780>
27. Ormiston S, Hanson B. World of genomics: Vietnam. *Front Line Genomics*; 2023 [2026 Aug 5]. Available from: <https://frontlinegenomics.com/world-of-genomics-vietnam/>
28. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312:1870–9. <https://doi.org/10.1001/jama.2014.14601>
29. Lionel AC, Costain G, Monfared N, Walker S, Reuter MS, Hosseini SM, et al. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med*. 2018;20:435–43. <https://doi.org/10.1038/gim.2017.119>
30. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23:2029–37. <https://doi.org/10.1038/s41436-021-01242-6>
31. Marshall CR, Chowdhury S, Taft RJ, Lebo MS, Buchan JG, Harrison SM, 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. <https://doi.org/10.1038/s41525-020-00154-9>
32. Zhu Y, Wang Y, Wang J, Zhang X, Li M, Chen H, et al. Targeted long-read sequencing facilitates effective carrier screening for complex monogenic diseases including spinal muscular atrophy, alpha-/beta-thalassemia, 21-hydroxylase deficiency, and fragile-X syndrome. *J Transl Med*. 2025;23:301. <https://doi.org/10.1186/s12967-025-06345-1>
33. Zhou X, Chen Y, Chen S, Li J, Wang H, Zhang L, et al. Detection of rare thalassemia variants using accurate circular consensus long-read sequencing. *Hum Mutat*. 2026;2026:8897128. <https://doi.org/10.1155/humu/8897128>
34. Liang Q, Gu W, Chen P, Li Y, Liu Y, Tian M, et al. A more universal approach to comprehensive analysis of thalassemia alleles (CATSA). *J Mol Diagn*. 2021;23:1195–204. <https://doi.org/10.1016/j.jmoldx.2021.06.007>
35. Luo S, Chen X, Zhong Q, Wang Q, Xu Z, Qin L, et al. The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype-genotype correlation. *J Hum Genet*. 2022;67:197–205. <https://doi.org/10.1038/s10038-021-00983-1>
36. Tung N, Ricker C, Messersmith H, Balmana J, Domchek S, Stoffel EM, et al. Selection of germline genetic testing panels in patients with cancer: ASCO guideline. *J Clin Oncol*. 2024;42:2599–615. <https://doi.org/10.1200/JCO.24.00662>
37. Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, et al. Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. *JAMA*. 2017;318:825–35. <https://doi.org/10.1001/jama.2017.11137>
38. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23:703–13. <https://doi.org/10.1038/nm.4333>
39. Fricke-Galindo I, Llerena A, Lopez-Lopez M. An update on the pharmacogenomics of CYP2C19 in relation to clinical outcomes. *Pharmacogenomics*. 2016;17:1269–91. <https://doi.org/10.2217/pgs-2016-0074>
40. Gaedigk A, Sangkuhl K, Whirl-Carrillo M, Klein T, Leeder JS. Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med*. 2017;19:69–76. <https://doi.org/10.1038/gim.2016.80>
41. Dean L. Carbamazepine therapy and HLA genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kattman BL, Malheiro AJ, editors. *Medical genetics summaries*. Bethesda: National Center for Biotechnology Information; 2018 [2026 Aug 5]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK321445/>
42. Yuliwulandari R, Kristin E, Prayuni K, Sachrowardi Q, Suyatna FD, Menaldi SL, et al. Association of the HLA-B alleles with carbamazepine-induced Stevens-Johnson syndrome/toxic epidermal necrolysis in the Javanese and Sundanese population of Indonesia: the important role of the HLA-B75 serotype. *Pharmacogenomics*.

- 2017;18:1643–8. <https://doi.org/10.2217/pgs-2017-0103>
43. Yuliwulandari R, Prayuni K, Viyati K, Mahasirimongkol S, Wichukchinda N. Frequencies of HLA-B alleles in Indonesian Malay Ethnic. *Heliyon* 2024;10(5):e26713. <https://doi.org/10.1016/j.heliyon.2024.e26713>
44. Dean L. Clopidogrel therapy and CYP2C19 genotype. In: Pratt VM, Scott SA, Pirmohamed M, Esquivel B, Kattman BL, Malheiro AJ, editors. *Medical genetics summaries*. Bethesda: National Center for Biotechnology Information; 2022 [2026 Aug 5]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK84114/>
45. Gil MM, Accurti V, Santacruz B, Plana MN, Nicolaides KH. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: updated meta-analysis. *Ultrasound Obstet Gynecol*. 2017;50:302–14. <https://doi.org/10.1002/uog.17484>
46. Benoit P, Brazer N, de Lorenzi-Tognon M, Kelly E, Servellita V, Oseguera M, et al. Seven-year performance of a clinical metagenomic next-generation sequencing test for diagnosis of central nervous system infections. *Nat Med*. 2024;30:3522–33. <https://doi.org/10.1038/s41591-024-03275-1>
47. Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. *Blood*. 2010;115:4331–6. <https://doi.org/10.1182/blood-2010-01-251348>
48. World Bank. Current health expenditure per capita (current US\$), indicator SH.XPD.CHEX.PC.CD, derived from the WHO Global Health Expenditure Database. Washington (DC): World Bank; 2024 [2026 Aug 5]. Available from: <https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD>
49. Simmons SK, Lithwick-Yanai G, Adiconis X, Oberstrass F, Iremadze N, Geiger-Schuller K, et al. Mostly natural sequencing-by-synthesis for scRNA-seq using Ultima sequencing. *Nat Biotechnol*. 2023;41:204–11. <https://doi.org/10.1038/s41587-022-01452-6>
50. Liang Q, He J, Li Q, Zhou Y, Liu Y, Li Y, et al. Evaluating the clinical utility of a long-read sequencing-based approach in prenatal diagnosis of thalassemia. *Clin Chem*. 2023;69:239–50. <https://doi.org/10.1093/clinchem/hvac200>
51. Sengupta D, Choudhury A, Ramsay M. H3Africa: a model for implementing biobank-based genomic research in resource-constrained settings. *Hum Mol Genet*. 2025;ddaf113. <https://doi.org/10.1093/hmg/ddaf113>
52. European Commission. European '1+ million genomes' initiative. Brussels: European Commission; 2024 [cited 2026 Aug 5]. Available from: <https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes>
53. GenomeAsia100K Consortium. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576:106–11. <https://doi.org/10.1038/s41586-019-1793-z>
54. Mulder NJ, Adebiyi E, Alami R, Benkahla A, Brandful J, Doumbia S, et al. H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. *Genome Res*. 2016;26:271–7. <https://doi.org/10.1101/gr.196295.115>
55. ASEAN Secretariat. ASEAN medical device directive. Jakarta: ASEAN Secretariat; 2015 [2026 Aug 5]. Available from: <https://asean.org>