sciendo ASIAN BIOMEDICINE DOI 10.2478/abm-2022-0032 — Asian Biomed (Res Rev News) 2022; 16(6):285–298

**Mini review** **Open access**

# Central resources of variant discovery and annotation and its role in precision medicine

Hashim Halim-Fikri[1], Sharifah-Nany Rahayu-Karmilla Syed-Hassan[1], Wan-Khairunnisa Wan-Juhari[1,2], Mat Ghani Siti Nor Assyuhada[1], Yetti Hernaningsih[3], Narazah Mohd Yusoff[3,4], Amir Feisal Merican[5,6], Bin Alwi Zilfalil[1,2,*]

**Abstract**

Rapid technological advancement in high-throughput genomics, microarray, and deep sequencing technologies has accelerated the possibility of more complex precision medicine research using large amounts of heterogeneous health-related data from patients, including genomic variants. Genomic variants can be identified and annotated based on the reference human genome either within the sequence as a whole or in a putative functional genomic element. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) mutually created standards and guidelines for the appraisal of proof to expand consistency and straightforwardness in clinical variation interpretations. Various efforts toward precision medicine have been facilitated by many national and international public databases that classify and annotate genomic variation. In the present study, several resources are highlighted with recognition and data spreading of clinically important genetic variations.

**Keywords:** annotation; discovery; precision medicine; resources; variant

Over the last 50 years, genetics has evolved in line with genomic technologies such as DNA sequencing [1], and currently possesses manifold applications, particularly in avenues of research where discoveries and continuously updated information pertaining to human diseases' genome variations is useful. This then led to the rapid and high-throughput generation of sequence data, which enabled the compilation of genetics information on a bigger scale, thereby resulting in massive consequences in terms of interpretation and correlation.

With the development of next-generation sequencing (NGS), researchers can easily identify the genetic sequences in the human genome [2]. This fundamental and landmark discovery provided valuable information for understanding various functions of genes or functional genetics as well as their role in human diseases.

Genome annotation is the process of identifying functional elements in the sequence of a genome. Based on the characteristics of genomics and genetics, the variants can be

***Correspondence to: Bin Alwi Zilfalil,** Malaysian Node of the Human Variome Project, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia; Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia, email: zilfalil@usm.my

[1]Malaysian Node of the Human Variome Project, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia

[2]Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia, Kelantan 16150, Malaysia

[3]Department of Clinical Pathology, Faculty of Medicine Universitas Airlangga, Dr. Soetomo Academic General Hospital, Surabaya, Indonesia

[4]Clinical Diagnostic Laboratory, Advanced Medical and Dental Institute, Universiti Sains Malaysia, Penang 13200, Malaysia

[5]Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia

[6]Center of Research for Computational Sciences and Informatics in Biology, Bio Industry, Environment, Agriculture and Healthcare (CRYSTAL), University of Malaya, Kuala Lumpur 50603, Malaysia

Open Access. © 2022 Halim-Fikri et al., published by Sciendo. This work is licensed under the Creative Commons Attribution 4.0 International License.

categorized or prioritized for further research. The genome that has been sequenced is not completed without an annotation process, which will help determine its clinical consequences. This information is fundamental for the practice of personalized medicine. Therefore, it is important to create a convergence between human genetics, molecular biology, biochemistry, and bioinformatics in such a way that the combination of their knowledge-pools enables arriving at an understanding over how variations of DNA exist in traits and diseases [3].

Human DNA consists of single nucleotide polymorphisms (SNP), mutations, and copy number variants (CNVs). These variants constitute a cause for the differentiation of individuals based on their unique traits, including disease susceptibility and pharmacogenomics. The understanding of our genome is not only useful for the geneticist but also relevant to those working on tools and health development [4].

The 2022 study of Samuels et al. [5] is a recent development in the literature that highlights empirically based annotation techniques, together with discussing their strengths and weaknesses. Further, the study of Hebbar and Sowmya [6] focuses on various tools and methods developed for variant annotation along with their datasets.

In this review, we discuss the features, strengths, and weaknesses of various resources as well as the establishment of a genomics variants database, i.e., the Malaysian Node of the Human Variome Project Database (MyHVPDb) and countries in Asia. We hope that this review will be useful for low and middle-income countries (LMIC) in establishing their own database for their population. This review also highlights how genomic variants are classified.

There are a few resources presently available that provide the basic information that is required in genetic annotation [4]. As of today, there are approximately 147,500 data entries pertaining to locus-specific databases curating genetic variants on specific genes or diseases. However, there are differences in the types of data and nomenclature (https://grenada.lumc.nl/LSDB_list/lsdbs).

The National Center for Biotechnology Information (NCBI) has developed a database, ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), to provide information about the relationships between human variants and phenotypes, and this information predominantly assumes the form of clinical interpretation and supporting evidence [7]. In addition to these international genomic databases, the Malaysian Node of the Human Variome Project (MyHVP) was established to collate the genomics variations of different ethnicities in Malaysia, including disease variants. Despite various resources and databases that have been established, there still remain changes and limitations pertaining to these databases.

## Reference genome

Locus Reference Genomic (LRG) (URL: http://www.lrg-sequence.org) was established to curate the variants with clinical implications and contains stable genomic DNA sequences, transcript, and protein reference sequences for reporting clinically relevant sequence variants [8]. The updatable section contains the mapping of the LRG to the most recent genome build and the most advanced and up-to-date biological knowledge for the LRG region from Ensembl and NCBI. To ensure that each LRG record is updated, the curators review all existing LRGs twice a year and when a new genome build is released.

The LRG is aligned to the reference genome to determine its mapping coordinates. Each area of contiguous alignment is described separately, along with any differences between the LRG and the reference. The existence of LRG is important for researchers to understand the function of a gene and how variations might affect its function and cause clinical implications [2]. The reference genome is not from a single person only. It is derived from the DNA of certain people, as data sequenced from thousands of genomes will show a drastic variation in DNA content. Among the advantages of having a reference is that the genome would be useful in analyses employing DNA testing for forensic purposes, as well as paternity determination, through the repetitive sequence of variant regions.

## Importance of variant annotation and interpretation

An individual's DNA varying from those of other individuals is caused by differences in ancestral pattern, unique traits, and susceptibility to certain diseases [9]. Therefore, a greater understanding of variant annotation and interpretation is essential and contributes significant roles to further determining how this part of the genome influences life and how knowledge of these variants can be applied toward optimal health management and care [3].

Variant annotation is an important step for the precise determination of sequence variants, as genetic alterations may result in various effects on genes and proteins, including a gain of function mutation that commonly affects the protein by introducing new protein functions such as missense mutation. Additionally, genetic alteration may also be caused by loss of function mutation, which could abolish the protein function.

The effect may result in nonsense mutation that terminates the proteins or splice site mutation that can disrupt RNA splicing, causing alteration to protein sequence. By annotating the variants according to their functional classification, it would be easier for researchers to prioritize their target variants and perform further interpretation.

## Next-generation genome annotation and the challenges

Curation of the human genome is a time-consuming process because it depends on unstructured sources such as published literature, collections of laboratory clinical data, disease-specific groups, and databases of functional experiments. While the genome sequencing revolution has led to the sequencing and assembly of many thousands of new genomes, genome annotation still uses very nearly the same technology that we have used for the past 2 decades. Variant identification and annotation are pivotal starting points for an accurate variant interpretation. The prevailing challenge in the variant annotation is the conversion of genomic variant coordinates to their corresponding localization of cDNA or amino acid [10]. Genome interpretation is highly dependent on the well-defined genomic variant coordinates, as the difference in the alignment of sequence variants may lead to incorrect variant nomenclature. An accurate variant interpretation is also determined by a consistent variant representation using a correct transcript accession [10].

Additionally, issues and arguments of data inconsistency may arise due to disparities between the laboratories and consequently cause difficulties for clinicians in fully utilizing genetic information for their patients' management. To successfully resolve these discrepancies, broad, up-to-date, and flexible standards that are best equipped and in line with the expanding genomics information should be deployed, so as to ensure the benefit of overcoming any argument in clinical classifications [11, 12]. Additionally, various variant detection software tools can strictly identify a specific alteration, including SNV, indels, structural variants, and CNV. Clinical laboratories must also understand and be aware of any limitations of the variant detection tools. The quality of the results from each laboratory should be maintained through an appropriate laboratory validation of the bioinformatics pipeline and the use of commercially purchased bioinformatics tools and software [11, 12].

Another critical point to be considered is the inclusion of significant clinical information and interpretation of the identified variants such as the gene symbol, the location of the variant, type of variant, correct HGVS nomenclature, and predicted protein sequence alterations. An additional source of information, such as cross-references to cancer-specific databases, multiple genomic databases, population frequency databases, and pre-computed in silico-based prediction tools, is also very useful in assisting in variant interpretation [10].

Variant annotation is greatly dependent on computational predictions to characterize the effect of the variant on a molecular level. While the use of this prediction tool was aided by the availability of published reports and an expert panel for validation, few limitations still exist in these tools. For annotation tools such as SIFT and PolyPhen that predict the functional effect of SNP on protein and phenotype, these tools have their own limitations. SIFT could not predict the protein structure while PolyPhen includes that information; however, the accuracy of this tool depends on the structural class of the protein [3, 13]. Various databases have now provided gene expression data in addition to genome annotation and allele frequencies for population-based information [14, 15]. However, these databases are still characterized by a few limitations, including an absence of data integration for different kinds of information, and an example of this particular limitation would be that, for a given DNA variant query at any one time, gene expression data are not available in a state of incorporation within the functional prediction and pathogenicity classification [16]. Most of these databases also only allow a limited number of gene and DNA variant queries [17]. Massive data for variant interpretation has been a major challenge for researchers and the integration of different information from the available resources to support the evaluation of evidence for data interpretation will be a better solution to resolve this limitation. At this point, variant annotation from multiple annotation tools and databases must be considered for a complete variant interpretation [18].

## Global initiatives on variant annotation

Selected resources, as described in **Table 1**, show the valuable information and analysis tools needed to portray genetic variations along the proof continuum [19]. These public resources consist of variation data from both normal and affected populations [20]. Advancements in innovation and investigation abilities, resulting in costs' and time consumption reduction, have alternately changed genomics, empowering one to pinpoint applicable genetic variants inside a large number of variations in a sequenced database [21]. However, recovering factual and useful annotation important at the single nucleotide level is still problematic [22].

**Table 1.** Selected resources, valuable information, and analysis tools available for variant annotation

| Resource | | Brief description | Features | References |
|---|---|---|---|---|
| **Database** | | | | |
| 1 | dbVar | dbVar is an archive of large-scale genomic variants (generally >50 bp) including insertions, deletions, duplications, inversions, mobile elements, translocations, and complex variants. The current number of submitted variants in the database is more than 6 million from over 190 studies including large international projects such as 1000 Genome Project, gnomAD, the CNV Global Population Survey, and clinical resources, which are ClinVar and ClinGen. | To date, the total in the dbVar database of 190 studies consists of 6 million regions and 36 million variants. dbVar continues to make it easier to find and use structural variation data by making selected datasets available on TrackHub for viewing in the NCBI GDV and other genome browsers, such as the UCSC browser. Tracks are available for the structural variants imported from ClinVar (https://www.ncbi.nlm.nih.gov/dbvar/content/clinvar_summary/#homepage) and for the NCBI Curated Common Structural Variants (nstd186). | [23–25] |
| 2 | HGMD | A thorough assortment of published germline mutations in nuclear genes that entail, or are firmly connected with, human acquired disease. | HGMD has been supported over the years by commercial partnerships with various industry leading biomedical research companies.<br>A stand-alone web application has been made available under license from BIOBASE GmbH.<br>The latest version of HGMD (2020.2) contains 289,346 different mutations in 11,076 genes (more than ClinVar & OMIM).<br>Articles identified as potential sources of mutation data are assessed by a team of experienced curators (with an average of more than 12 years' experience in curation).<br>Academic or non-profit users without a subscription may utilize the public version of HGMD (http://www.hgmd.org). However, this version is provided in a basic form that is searchable only by gene symbol or disease name, is only updated twice annually, is maintained permanently at least 3 years out of date, and does not contain any of the additional annotations or extra features present in HGMD Professional.<br>Another challenge is that an increasing number of journals do not appear to be systematically indexed by Medline, at least not immediately upon publication. | [26] http://www.hgmd.cf.ac.uk/ac/index.php |
| 3 | LOVD | The purpose of LOVD is to provide a flexible and freely available tool for genomic variant and phenotype collection, display, and curation. | LOVD allows both patient-centered and gene-centered views as it is open sourced and released under the GPL license. LOVD is actively being improved. Using Leiden server, LOVD offers free hosting and support of LOVD-powered gene variant databases. | [27, 28] https://www.lovd.nl/3.0/publiclist |
| 4 | gnomAD | An asset created by a global alliance of specialists, with the objective of amassing and blending both exome and genome sequencing information from a wide assortment of huge scope sequencing researches and making rundown information accessible for the more extensive academic local area. | The data released by gnomAD are available free of restrictions and proved to be fertile ground for testing new approaches to variant interpretation.<br>Major limitations on analysis is the quality and size of the available reference databases of normal genetic variation and the data is only focused entirely on small variants. | http://gnomad.broadinstitute.org |
| 5 | dbSNP | An information base of short genetic variants that list variations with population-genetic annotations, file germline variants for both rare mutation and polymorphism, and give alleles, genotypes, and their individual frequencies based on their population. | Data within dbSNP are available freely and in a variety of forms.<br>There is no requirement or assumption about minimum allele frequencies or functional neutrality for the polymorphisms in the database.<br>The validation code in the SNP record provides some limited information on how the SNP was identified and on the experimental evidence confirming its existence.<br>The quality of the data found on dbSNP has been questioned by many research groups, which suspect high false positive rates due to genotyping and base-calling errors. | [29] http://www.ncbi.nlm.nih.gov/snp |

(*Continued*)

**Table 1.** Continued

| Resource | | Brief description | Features | References |
|---|---|---|---|---|
| 6 | GenBank | A public database that contains 9.9 trillion base pairs from over 2.1 billion nucleotide sequences for 478,000 formally described species. | A public database in which nucleotide sequences are available for 400,000 formally described species, and which handles large sequence records.<br>Daily data exchange with the Europe Nucleotide Archive and the DNA Data Bank of Japan ensures worldwide coverage. | [30, 31] |
| 7 | ClinGen | Makes an assembly of organization assets to reform our perception of genomic variants and perk up its use in clinical thought with related data set in ClinVar. | The ClinGen variant curation process combines clinical, genetic, population, and functional evidence with expert review to classify variants into 1 of 5 categories, which are pathogenic, likely pathogenic, uncertain, likely benign, and benign according to the ACMG guidelines. | [32] https://clinicalgenome.org/ |
| 8 | ClinVar | Totals data on variation of sequence and its association to human wellbeing and giving affirmations of variation that are of clinical significance. | In December 2019, ClinVar reached the milestone of 1 million submitted records, representing more than half a million variants. ClinVar allows a user to "follow" a particular variant and be notified if the overall clinical interpretation in ClinVar changes, for example from a pathogenic category to a non-pathogenic one. This feature makes it easier for a laboratory to become aware of variants that may need to be re-evaluated, and for clinicians to know when they should contact their clinical testing laboratory, patient, or both with new information. | [7, 23] http://www.ncbi.nlm.nih.gov/clinvar |
| **Tools** | | | | |
| 1 | Phenotype Based Gene Analyzer (Phenolyzer) | An instrument zeroing in on finding genes dependent on client explicit disease or phenotype terms. | Phenolyzer includes 5 components and works based on an intuitive approach.<br>Phenolyzer exhibits superior performance over competing methods for prioritizing Mendelian and complex disease genes, based on disease or phenotype terms entered as free text.<br>Compared to wANNOVAR, Phenolyzer is better at prioritizing candidate genes for Mendelian diseases and complex diseases. | [33] http://phenolyzer.wglab.org/ |
| 2 | VarAFT software | Provides annotation and determines human disease-causing mutations by accessing various levels of data. | It combines information from the dbNSFP and ANNOVAR, OMIM, HPO, Gene Ontology, pathways (Reactome), KEGG, PID, predictions from UMD-Predictor, and HSF.<br>The filtration features the prioritization of candidate mutations and it is a freely available software. | [34] http://varaft.eu |
| 3 | Ensembl | A product framework of genome data sets for vertebrates and other eukaryotic species, which delivers and keeps up programed annotation on chosen eukaryotic genomes | The framework indicates the role played by variant effect prediction (VEP) "plugins" in expanding variant functionality, and the plugins considered include Missense Tolerance Ratio (MTR) and Rare Exome Variant Ensemble Learner (REVEL); additionally, the variant data integration from NCBI dbSNP and European Variation Archive (EVA) is considered, as well as the population frequency data from Genome Aggregation Database (gnomAD).<br>It displays variant locations on experimentally derived 3D protein structures. | [35–37] http://www.ensembl.org |
| 4 | SNPnexus | An online variant annotation instrument intended to improve and aid the determination and prioritization of known and novel genomic alterations | It applies DeepSEA scoring algorithm for non-coding variants and CGI for identification of somatic alterations.<br>Genes are linked to associated pathways using Reactome. | [38, 39] https://www.snp-nexus.org/v4/ |
| 5 | UCSC genome browser | An online instrument for rapidly showing a mentioned segment of a genome at any magnitude, joined by a progression of aligned annotation "tracks" | Researchers are able to customize their annotation track based on the track hubs available in the genome browser.<br>The framework allows a broad selection of annotation track; for example, GTEx Gene track, which is used to display gene expression levels across various tissue types; NCBI RefSeq tracks; and gnomAD, which is used for whole genome and whole exome datasets. | [40–44] http://www.genome.ucsc.edu/ |

(*Continued*)

**Table 1.** Continued

| Resource | | Brief description | Features | References |
|---|---|---|---|---|
| 6 | COSMIC | A worldwide asset infiltrating the world literature on somatic mutations in human cancer | It provides functional descriptions of cancer genes and consists of 2 tiers of CGC to classify genes based on their strength of evidence associated to their role in cancer.<br>A new platform of COSMIC-3D, for understanding of cancer mutations in 3D protein structure | [45–47]<br>http://www.sanger.ac.uk/perl/genetics/CGP/cosmic |
| 7 | ITHANET | A global site devoted to thalassemia and different hemoglobinopathies, which gives free admittance to data sets and devices applicable to various parts of hemoglobinopathies, for example, sequence variations, epidemiology, and phenotype | The website provides the most recent content update with its friendly user interface and is easy to navigate. It provides multiple external links to other databases. | [48, 49]<br>https://www.ithanet.eu/ |
| 8 | Genenames.org | The HGNC based at EMBL-EBI assigns unique symbols and names to human genes. | It contains almost 42,000 approved gene symbols, over 19,000 of which represent protein-coding genes, and displays all approved nomenclature within Symbol Reports that contain data curated by HGNC editors and links to related genomic, phenotypic, and proteomic information.<br>It is accessible via https://www.genenames.org | [50–52] |
| **Consortium** | | | | |
| 1 | ICGC | The ICGC is a global initiative to build a comprehensive catalog of mutational abnormalities in the major tumor types. | ICGC's Data Portal is a user-friendly platform for efficient visualization, analysis, and interpretation of large, diverse cancer datasets. The portal currently consists of data from 84 worldwide cancer projects, collectively representing about 77 million somatic mutations and molecular data from over 20,000 contributors.<br>The ICGC dataset comprises several high-level data types, including donor, gene, mutation, and cancer drug. To support fast, simultaneous querying of these entities and their relationships, the ICGC developed a sophisticated search framework based on the document-oriented indexing technology Elasticsearch (https://www.elastic.co/), a distributed search and analytics engine that features fast query speed, massive scalability, and flexible JSON schemas. It takes ~100 ms in a search across the 77 million ICGC mutations to retrieve the 10 most frequent mutations found in brain cancer donors and those that occur within genes cataloged in the COSMIC CGC (https://cancer.sanger.ac.uk/census). | [53] |

(*Continued*)

**Table 1.** Continued

| Resource | | Brief description | Features | References |
|---|---|---|---|---|
| **Project** | | | | |
| 1 | 100,000 Genome Project | The biggest worldwide coordinated effort, which has sequenced 100,000 genomes from around 85,000 NHS patients influenced by an uncommon disease, or malignant growth | The 100,000 Genomes Project aims to sequence 100,000 genomes from NHS patients with cancer and rare diseases. Data collected from the 100,000 Genomes Project can inform research on rare diseases, or benefit patient care potentially by streamlining the diagnostic process and tailoring care to the individual.<br>The project was established to sequence 100,000 genomes from around 85,000 NHS patients affected by a rare disease, or cancer.<br>The Project would also create a new genomic medicine service for the NHS – transforming the way people are cared for and bringing advanced diagnosis and personalized treatments to all those who need them.<br>The 100,000 Genomes Project improves the field of molecular medicine through the improvement of diagnosis of disease, earlier detection of genetic predispositions to disease, rational drug design, gene therapy, and control systems for drugs and pharmacogenomics "custom drugs."<br>The Human Genome Project could help with the diagnosis and prevention of human disease, which allow to modify medication for more effective treatment cycles, improve criminal justice proceedings, helped to boost the economy and can help more than just humans.<br>The Human Genome Project may cause a loss in human diversity, by developing a trend in "designer" humans. Its information could be used to form new biological weapons and could become the foundation of genetic racism as it would be most accessible to wealthy cultures. | [54]<br>https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/ |
| 2 | GENCODE | A project to identify and classify all gene features in the human and mouse genomes with high accuracy based on biological evidence, and to release these annotations for the benefit of biomedical research and genome interpretation | The project annotates human and mouse genes and transcripts supported by experimental data with high accuracy, providing a foundational resource that supports genome biology and clinical genomics.<br>GenCode annotation is accessible via Ensembl, the UCSC Genome Browser, and https://www.gencodegenes.org | [55–57] |
| **Guidelines** | | | | |
| 1 | ACMG Guidelines | An established guidance for the interpretation of sequence variants. | It provides the guidelines of variant classification, namely "pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign," based on criteria using typical types of variant evidence (population data, computational data, functional data, segregation data, etc.).<br>These 5 features are applied to the ClinVar database. | [58]<br>https://www.acmg.net/docs/standards_guidelines_for_the_interpretation_of_sequence_variants.pdf |

ACMG, American College of Medical Genetics and Genomics; CGC, cancer gene census; CGI, cancer genome interpreter; COSMIC, catalogue of somatic mutations in cancer; ClinGen, Clinical Genome Resource; CNV, copy number variant; dbSNP, database for single nucleotide polymorphism; EMBL-EBI, EMBL's European Bioinformatics Institute; EVA, European Variation Archive; gnomAD, Genome Aggregation Database; GTEx, genotype-tissue expression; HGMD, Human Gene Mutation Database; HSF, human splicing finder; HGNC, HUGO Gene Nomenclature Committee; ICGC, International Cancer Genome Consortium; JSON, JavaScript Object Notation; LOVD, Leiden Open (source) Variation Database; MTR, Missense Tolerance Ratio; NCBI; National Center for Biotechnology Information; OMIM, Online Mendelian Inheritance in Man; REVEL, Rare Exome Variant Ensemble Learner; SNP, single nucleotide polymorphism; VEP, variant effect prediction.

### *Initiatives by LMICs in establishing genomics variants database*

Several LMICs have established their own ethnic-specific genome databases in preparation for future implementation of personalized medicine. As an example, Thailand has developed the Thai Reference Exome (T-REx) variant database (https://trex.nbt.or.th/) that can be used to filter and prioritize pathogenic candidate mutations among the Thai population [59]. T-REx is part of the Genomic Thailand Initiative, which the Thai government supports, with the aim of sequencing the genomes of 50,000 diverse individuals among the Thai population so as to establish an infrastructure for harnessing genomic information. The T-REx variant database contains information regarding short variants and CNV genotyped from 1092 unrelated populations of Thailand [59].

In the same context, Iran has genotyped 800 individuals from 8 major ethnic groups, which has contributed to establishing the Iranome database (www.iranome.com). This database contains 1,575,702 variants, of which 308,311 were novel variants (19.6%). The clinical information in this database can be a useful tool for categorizing the pathogenicity of uncommon mutations [60].

Malaysia is a middle-income country located in Southeast Asia and is a member of the Human Variome Project. Its services in medical genetics started in 1994 and are improving within 2 decades. The availability of medical genetic services includes genetic testing, diagnosis, and counseling for both adult-onset inherited and pediatric conditions [61]. Up to now, variant information of the Malaysian genome is under-represented in many databases. Therefore, with the objective of ascertaining and documenting molecular genomic variations across the Malay population, Malaysia has initiated the first database for this population, namely the Malaysian Node of the Human Variome Project (MyHVP), which includes the Chinese, Indian, and Orang Asli ethnicities in Peninsular Malaysia, as well as various minority ethnicities in Sabah and Sarawak. The MyHVPDatabase (MyHVPDb) was developed as a database containing genetic variants of both healthy and diseased Malaysians based on their ethnic groups [62].

MyHVPDb serves as a repository of information about mutations, and can have applications in future undertakings by researchers worldwide, especially those focusing on precision medicine. To date, a total of 189 genes consisting of 2571 variants are recorded in MyHVPDb, and a majority of the sources of genetic information originate from individuals belonging to the Malay, Orang Asli, Indian, and Chinese ethnic groups in Malaysia. These variants are associated with 112 diseases including beta-thalassemia, multiple endocrine neoplasia, breast cancer, and retinoblastoma. About 291,718 SNPs were genotyped from 101 healthy Malay individuals and deposited in MyHVPDb.

MyHVPDb was developed using the Leiden Open Variation Database (LOVD) version 3. LOVD is widely used for the standard variant databases system for the collection, display, and curation of DNA variants including phenotypes and diseases [63]. The current version of LOVD 3 introduced a more sophisticated patient phenotype model as an important feature for this version. This allowed phenotype information to be recorded over time as a disease is diagnosed, treated, or progresses in a patient [63]. Having the phenotype feature will fulfill our future plan to add the clinical data in the MyHVPDb as an important resource to the clinicians in the field of genetic medicine.

The Singapore Genome Variation Project (SGVP) provided a publicly available resource of 1.6 million SNPs in the past 12 years that were genotyped in 268 individuals from the Chinese, Malay, and Indian population groups in Southeast Asia [64]. This project aims to characterize genomic variation and positive natural selection in 3 major Asian population groups (Chinese, Malay, and Indian). With the genomic information that is already available, Singapore is currently working on a national strategy for developing personalized medicine, which is called Singapore's National Precision Medicine (NPM). The goal is to improve and protect the health of every Singaporean by strengthening public health, enhancing disease prevention, and discovering the most appropriate therapies for the most appropriate individuals and groups [65]. Until now, Singapore has been the leading country in developing the field of personalized medicine when compared with other countries in the Asian region. The NPM was launched in 2017 and is a 10-year project, which has already entered the second phase in April 2021. The goal is to enhance and accelerate Singapore's biomedical research, health outcomes, and economic growth [65].

In Indonesia, a developing country, approximately 2500–5000 babies are born each year with a clinical pattern of beta-thalassemia major and beta-thalassemia intermedia [66, 67]. The prevalence rate for beta-thalassemia carriers varies in the range of 3.0%–10.0% throughout regions of Indonesia [68].

In general, DNA diagnosis is not routinely performed in the health service, and diagnosis of suspected thalassemia individuals usually comprises complete blood count and hemoglobin electrophoresis [68]. Several studies have mapped several local mutations in large, multi-ethnic cities in Indonesia [66, 67]. In Indonesia, the task of collection of thalassemia patients' information and compiling these into databases has not been implemented except on a very limited scale, since many of these patients have no allelic mutation detected. Since this patient database has still not been managed effectively, a national effort in this direction is needed, which may be initiated by a government agency dedicated especially to this task.

### *Guidelines for the interpretation of variants*

The American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines were predominantly developed as an educational guideline for clinical laboratory geneticists, to assist them in providing a quality clinical laboratory service [58].

### *ACMG Guidelines*

In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) released joint recommendations for interpreting variants in genes associated with Mendelian disorders. These recommendations are part of an effort to establish a common framework for variant classification based on a standardized and transparent assessment of different lines of evidence. ACMG Guidelines provide several levels of evidence to ascertain the clinical significance of each genomic variant. However, there is still a challenging task to determine the pathogenicity of variants, especially in cancer. Several factors and information need to be taken into account as each genomic alteration can have an extent of much clinical efficacy, such as the diagnosis, prognosis, therapy selection, and monitoring, though only a few cancers have strong evidence of pathogenicity [69]. Until now, peer-reviewed literature, guidelines for clinical practice, and any relevant mutation database are being utilized as the main sources in equipping these variants with useful clinically significant evidence for further assessment [70].

The ACMG–AMP Guidelines propose the use of specific terminology, namely "pathogenic," "likely pathogenic," "uncertain significance," "likely benign," and "benign," to characterize variants in Mendelian disorders (**Figure 1**). The variants are classified into these 5 categories according to different types of evidence including population data, computational data, functional data, and segregation data. This framework defined 28 evidence criteria organized by type and strength, which are used to classify the variants into the 5 categories. The framework proposed by the guidelines, encompassing various genes, diseases, and inheritance patterns, is intended for general use; however, to effectively execute the process of variant interpretation, expert judgment in evaluating and weighing evidence is needed." By using the guidelines, a mutation is characterized by a change in the DNA base sequence, whereas a polymorphism is characterized by a frequency of greater than 1%. Both mutation and polymorphism were later changed to "variant" to avoid confusion arising from inaccurate assumptions of pathogenic or benign effects [58].

The Clinical Genome Resource (ClinGen) convenes all Variant Curation Expert Panels (VCEPs) to perform gene- and disease-specific modifications to the ACMG–AMP framework for variants related to a particular disease and inheritance pattern [32]. Additionally, members of ClinGen have formed a working group, called the Sequence Variant Interpretation Working Group (SVI WG). This working group aims to provide refinements of ACMG–AMP guidelines across multiple domains and to harmonize guidelines made by individual VCEPs [71]. SVI WG has published these recommendations (https://www.clinicalgenome.org/svi/).

The current effort of ClinGen is the development of the ClinGen Variant Curation Interface (VCI). The ClinGen VCI is a global open-source variant classification platform that helps with the use of evidence criteria and the classification

| Benign | Likely benign | Variants of unknown significance | Likely pathogenic | Pathogenic |
|---|---|---|---|---|
| Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
| Common polymorphism or not causative for clinical phenotype | Probably not causative for clinical phenotype | Unknown effect on clinical phenotype | Probably causative for clinical phenotype | Causative for clinical phenotype |

**Figure 1.** The 5 variant classifications of ACMG Guidelines [58].

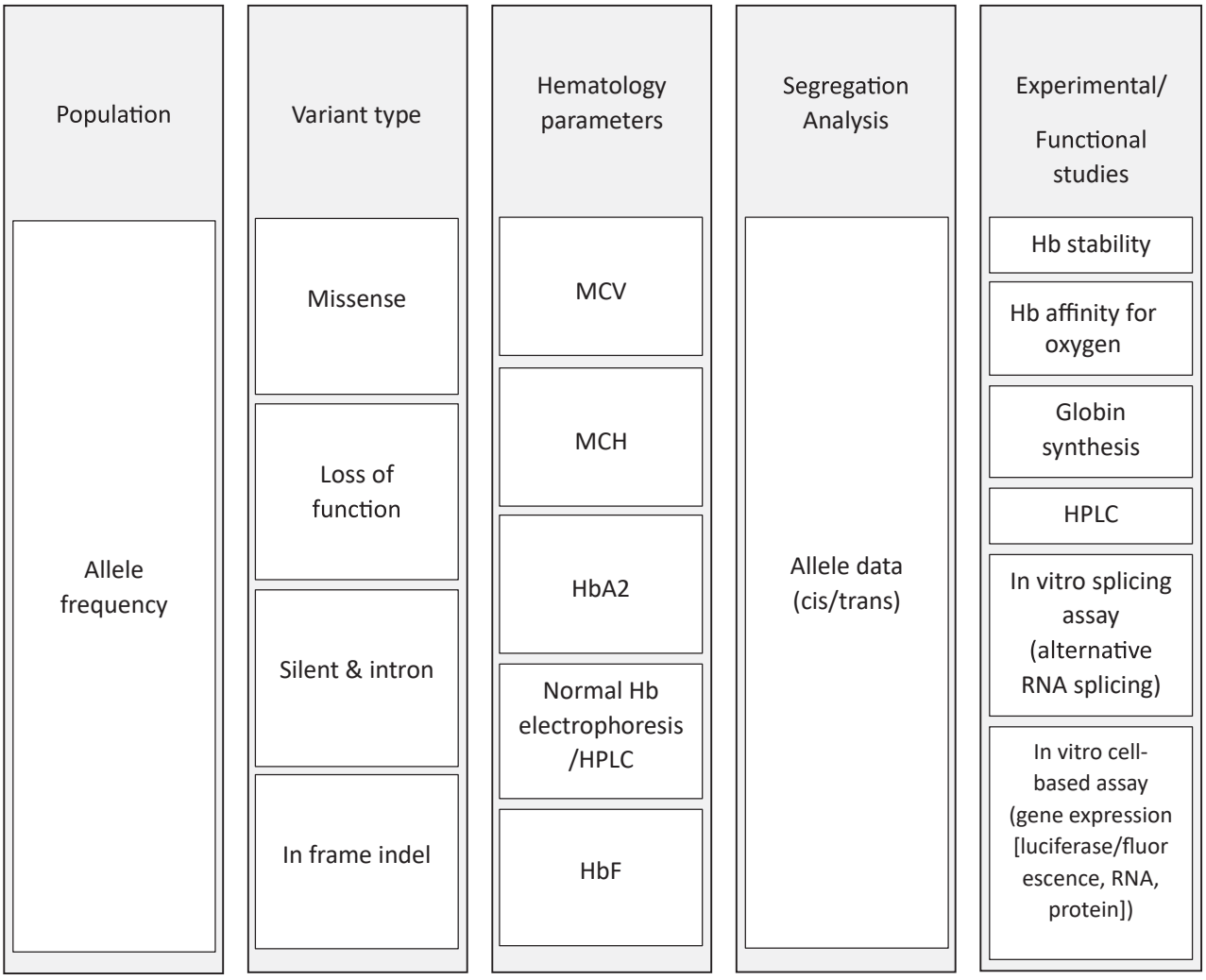



**Figure 2.** Parameters required for variant annotation in hemoglobinopathies.

of variants based on the ACMG–AMP variant classification guidelines. There are 43 VCEPs and the majority are in the process of the establishment (https://clinicalgenome.org/affiliation/vcep/#ep_table_heading). **Figure 2** indicates the parameters used by the ClinGen Hemoglobinopathy VCEP for annotation variants in disorders of hemoglobinopathies.

Although these modifiers may not fully represent human phenotypes, they comprise a 5-tier system of classification for variants that are significant to Mendelian disease, which should be supported with a clinical condition and inheritance pattern.

## Variant interpretation in precision medicine and its challenges

Precision Medicine Initiative (PMI) was initiated in 2015 to accelerate and enhance the understanding of precision medicine and to take into consideration factors such as genetics, physiology, and environmental parameters for ascertaining the best approach to disease treatment for any individual patient [72]. Previous and current medical practices usually do not consider this biomedical information due to inaccessible and insufficient sources. Additionally, genome sequencing and extensive biomedical data collection are usually still not an option due to the costs involved and hence not available for all [73, 74]. However, recently, high-throughput DNA sequencing technologies are starting to have an impact on various medical fields and it has become feasible to gather a substantial amount of health information data from patients to be applied for individualized medical treatment [75].

Precision medicine will aid physicians to choose the best treatment and prevention approach for each patient according to their personal health information and their genetic makeup [76]. Certain diseases are associated with and influenced by genetic variations as well as patients' responses to

certain medications. Therefore, the most suitable treatment can be considered to improve the patients' outcomes. For instance, the finding of hundreds of genetic susceptibility loci in inflammatory bowel disease (IBD) patients contributes to better management of IBD [77]. The development of precision medicine improves our healthcare surroundings by providing precise knowledge of current trends of morbidity and mortality in populations or groups with specific diseases [74].

Various efforts toward precision medicine have been facilitated by many national and international public databases that classify and annotate genomic variation [78], including database for single nucleotide polymorphism (dbSNP) [29], ClinVar [7], and Online Mendelian Inheritance in Man (OMIM) [79]. The basis of precision medicine lies in cataloging genomic variation data of local genomes. Strong local reference databases can make precision medicine meaningful and the inadequacy of local reference data is the biggest challenge in promoting precision medicine in Asia.

Additionally, a decision from expert panels for variant annotation might be one of the challenges in precision medicine. Experts disclose that the disadvantages of personalized medicine are classifiable into 3 recognized issues, which are validity uncertainty, equity issues, and implementation [80]. Therefore, the availability of artificial intelligence (AI) is critical for the implementation of precision medicine. When AI is introduced into precision medicine, organizations can implement precision medicine using their own innovative approach. For instance, AI controls deep learning approaches to defeat the difficulties involved in massive data sets and unstructured data. To illustrate, Gubatan et al. [81] explained the most recent advanced technologies of AI in disease diagnosis, risk prediction, and disease severity evaluation, as well as prediction of clinical decision-making in IBD patients. Again, Jiang et al. [82] developed predictive models learned from historical data for predicting COVID-19 severity. Therefore, we deduce that variant annotation can be implemented into AI to assist in precision medicine in the future.

In clinical settings, AI facilitates clinicians to work more efficiently to make diagnoses more accurate. Additionally, AI helps companies in drug development to implement time and cost-effective measures as well as prevent errors in the system.

## Global initiative toward precision medicine

Globally, there are many initiatives to predict, with a high level of accuracy, which treatment and prevention strategies would be an appropriate choice for any particular disease that may present in a population. A primary challenge for clinical laboratories is to determine, with adequate evidence, the genes that can be associated with clinical findings and patients' phenotypes. Therefore, ClinGen developed a method to help the community in this effort to assess the strength of evidence for each gene in specific diseases. ClinGen (clinicalgenome.org) is a multi-institutional initiative to standardize gene and variant curation as well as to improve the utility of genomics in medicine. ClinGen developed a method to evaluate the strength of evidence for providing supporting claims for every gene that causes particular diseases. Additionally, this effort was purposely initiated to establish an authoritative central resource for the characterization of clinically relevant genes and variants for precision medicine and future research [32].

ClinGen gene and variant curation is currently organized into 8 Clinical Domain Working Groups (CDWG) to emphasize on curation and standardization efforts in clarification. These CDWG include Cardiovascular, Hearing Loss, Hemostasis or Thrombosis, Hereditary Cancer, Inborn Errors of Metabolism, Neurodevelopmental, Neuromuscular, and RASopathies (mutations in genes of the Ras-MAPK pathway). Each CDWG manages numerous Gene or Variant Curation Expert Panels (GCEPs or VCEPs) performing curation activities on gene sets and diseases within the domain group [83].

As the ACMG–AMP Variant Interpretation Guidelines clarify and specify, ClinGen VCI curates the variants and expert review, and the VCEPs can provide interpreted variant sets to ClinVar and the ClinGen evidence repository (eRepo; erepo.clinicalgenome.org). This step encourages the sharing of data from diagnostic laboratories, creation of consensus interpretations, and development of a collection of gene-specified ACMG–AMP evidence codes, thereby paving the way for eventually ensuring more accurate interpretations for clinical use [75].

However, the critically important factor would also depend on ancestry and population reference data that are supported by previous data, developed through a detailed clinical investigation of affected patients and at-risk family members [84]. As the composition of pathogenic variants differs greatly among human populations, most data focused mainly on populations of European ancestry. In contrast, little is known about pathogenic DNA variants among populations from other parts of the world. This also includes those non-annotated variants that could be pathogenic in the minority of populations [85].

## Conclusion

In the era of precision medicine, variant discovery and annotation are required for diagnosis, prevention, and therapy improvement, based on an individually tailored approach; however, progress is still slow and the application of precision medicine

in patients' treatment still needs enhancement. Despite the abundance of raw data, much genetic variation in the human genome still remains uncharacterized. For precision medicine initiatives to progress, it is crucial that the discovery and annotation pipeline for novel variants also be incorporated during the analysis and interrogation of genomic data. Providing a uniform set of annotation resources for certain genetic diseases will ease comparisons and the interpretation of results. It is hoped that the discovery of novel genomic variants and their annotations would accelerate the development of precision medicine, thus benefitting mankind.

**Author contributions.** All authors contributed to the conception and design of the study. HHF, SNRKSH, WKWJ, MGSNA, YH, and BAZ contributed to the drafting of the manuscript. NMY, AFM, and BAZ contributed to the revising of the manuscript. All authors approved the final version submitted for publication and take responsibility for statements made in the published article.

**Acknowledgments.** This work was supported by the Universiti Sains Malaysia (USM), Grant No. 1001/PPSP/8012301.

**Conflict of interest statement.** Each author has completed and submitted an International Committee of Medical Journal Editors Disclosure Form. Apart from the funding source listed in the acknowledgments, none of the authors have any competing relationship, activity, or interest to disclose.

**Data sharing statement.** The present review is based on the references cited. Further details, opinions, and interpretations are available from the corresponding author on reasonable request.

## References

[1] Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of genetic techniques: past, present, and beyond. Biomed Res Int. 2015; 2015:461524. doi: 10.1155/2015/461524

[2] Steward CA, Parker APJ, Minassian BA, Sisodiya SM, Frankish A, Harrow J. Genome annotation for clinical genomic diagnostics: strengths and weaknesses. Genome Med. 2017; 9:49. doi: 10.1186/s13073-017-0441-1

[3] Miller JE, Veturi Y, Ritchie MD. Innovative strategies for annotating the "relationSNP" between variants and molecular phenotypes. Bio Data Min. 2019; 12:10. doi: 10.1186/s13040-019-0197-9

[4] Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, et al. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur J Hum Genet. 2014; 22:144–7.

[5] Samuels DC, Yu H, Guo Y. Is it time to reassess variant annotation? Trends Genet. 2022; 38:521–3.

[6] Hebbar P, Sowmya SK, editors. Genomic variant annotation: a comprehensive review of tools and techniques. Cham: Springer International Publishing; 2022.

[7] Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42:D980–5.

[8] MacArthur JA, Morales J, Tully RE, Astashyn A, Gil L, Bruford EA, et al. Locus reference genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic Acids Res. 2014; 42:D873–8.

[9] Sosnay PR, Cutting GR. Interpretation of genetic variants. Thorax. 2014; 69:295–7.

[10] Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the association for molecular pathology, American Society of Clinical Oncology, and College of American Pathologists. J Mol Diagn. 2017; 19:4–23.

[11] Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020; 22:245–57.

[12] Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Correction: technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2021; 23:2230. doi: 10.1038/s41436-021-01150-9

[13] Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat. 2011; 32:358–68.

[14] Chaitankar V, Karakülah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: perspectives for retinal research. Prog Retin Eye Res. 2016; 55:1–31.

[15] Serin EA, Nijveen H, Hilhorst HW, Ligterink W. Learning from co-expression networks: possibilities and challenges. Front Plant Sci. 2016; 7:444. doi: 10.3389/fpls.2016.00444

[16] Asplund A, Edqvist PH, Schwenk JM, Ponten F. Antibodies for profiling the human proteome-the human protein atlas as a resource for cancer research. Proteomics. 2012; 12:2067–77.

[17] Lee CY, Chattopadhyay A, Chiang LM, Juang JJ, Lai LC, Tsai MH, et al. VariED: the first integrated database of gene annotation and expression profiles for variants related to human diseases. Database (Oxford). 2019; 2019:baz075. doi: 10.1093/database/baz075

[18] Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. Genome Med. 2020; 12:103. doi: 10.1186/s13073-020-00803-9

[19] Ramos EM, Din-Lovinescu C, Berg JS, Brooks LD, Duncanson A, Dunn M, et al. Characterizing genetic variants for clinical action. Am J Med Genet C Semin Med Genet. 2014; 166C:93–104.

[20] Riggs ER, Ledbetter DH, Martin CL. Genomic variation: lessons learned from whole-genome CNV analysis. Curr Genet Med Rep. 2014; 2:146–50.

[21] Dahary D, Golan Y, Mazor Y, Zelig O, Barshir R, Twik M, et al. Genome analysis and knowledge-driven variant interpretation with TGex. BMC Med Genomics. 2019; 12:200. doi: 10.1186/s12920-019-0647-8
[22] Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-centered annotation browser. Bioinformatics. 2015; 31:1334–6.
[23] Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020; 48(D1):D9–D16.
[24] Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2021; 49(D1):D10–D7.
[25] Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022; 50(D1):D20–D6.
[26] Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet. 2017; 136:665–77.
[27] Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. Hum Mutat. 2011; 32:557–63.
[28] Fokkema IFAC, Kroon M, López Hernández JA, Asscheman D, Lugtenburg I, Hoogenboom J, den Dunnen JT. The LOVD3 platform: efficient genome-wide sharing of genetic variants. Eur J Hum Genet. 2021; 29:1796–803.
[29] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29:308–11.
[30] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. GenBank. Nucleic Acids Res. 2018; 46(D1):D41–D7.
[31] Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, et al. GenBank. Nucleic Acids Res. 2021; 49(D1):D92–D6.
[32] Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen – the clinical genome resource. N Engl J Med. 2015; 372:2235–42.
[33] Yang H, Robinson PN, Wang K. Phenolyzer: phenotype-based prioritization of candidate genes for human diseases. Nat Methods. 2015; 12:841–3.
[34] Desvignes JP, Bartoli M, Delague V, Krahn M, Miltgen M, Béroud C, Salgado D. VarAFT: a variant annotation and filtration system for human next generation sequencing data. Nucleic Acids Res. 2018; 46(W1):W545–W53.
[35] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 2019:531210. doi: 10.1101/531210
[36] Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020; 48(D1):D682–D8.
[37] [37] Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, et al. An overview of Ensembl. Genome Res. 2004; 14:925–8.
[38] Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 2015; 12:931–4.
[39] Oscanoa J, Sivapalan L, Gadaleta E, Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). Nucleic Acids Res. 2020; 48:W185–W92.
[40] Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 2019; 47(D1):D853–D8.
[41] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–60.
[42] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016; 44(D1):D733–45.
[43] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–91.
[44] Karolchik D, Hinrichs AS, Kent WJ. The UCSC genome browser. Curr Protoc Bioinformatics. 2009; Chapter 1:Unit1. 4. doi: 10.1002/0471250953.bi0104s28
[45] Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue of somatic mutations in cancer. Nucleic Acids Res. 2019; 47(D1):D941–D7.
[46] Jubb HC, Saini HK, Verdonk ML, Forbes SA. COSMIC-3D provides structural perspectives on cancer genetics for drug discovery. Nat Genet. 2018; 50:1200–2.
[47] Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, et al. The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet. 2008;Chapter 10:Unit 10. 11. doi: 10.1002/0471142905.hg1011s57
[48] Stephanou C, Kountouris P, Tamana S, Minaidou A, Lederer CW, Kleanthous M. ITHANET: an information and database community portal for hemoglobinopathies. Hemoglobin. 2019; 43:363.
[49] Kountouris P, Lederer CW, Fanis P, Feleki X, Old J, Kleanthous M. IthaGenes: an interactive database for haemoglobin variations and epidemiology. PLoS One. 2014; 9:e103020. doi: 10.1371/journal.pone.0103020
[50] Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, et al. Genenames.org: the HGNC and VGNC resources in 2019. Nucleic Acids Res. 2019; 47(D1):D786–D92.
[51] Seal RL, Gordon SM, Lush MJ, Wright MW, Bruford EA. genenames.org: the HGNC resources in 2011. Nucleic Acids Res. 2011; 39(Database issue):D514–9.
[52] Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Res. 2021; 49(D1):D939–D46.
[53] Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The international cancer genome consortium data portal. Nat Biotechnol. 2019; 37:367–9.
[54] Turnbull C, Scott RH, Thomas E, Jones L, Murugaesu N, Pretty FB, et al. The 100 000 genomes project: bringing whole genome sequencing to the NHS. BMJ. 2018; 361:k1687. doi: 10.1136/bmj.k1687
[55] Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47(D1):D766–D73.
[56] Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. Nucleic Acids Res. 2021; 49(D1):D916–D23.
[57] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome

annotation for The ENCODE Project. Genome Res. 2012; 22:1760–74.

[58] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17:405–24.

[59] Shotelersuk V, Wichadakul D, Ngamphiw C, Srichomthong C, Phokaew C, Wilantho A, et al. The Thai reference exome (T-REx) variant database. Clin Genet. 2021; 100:703–12.

[60] Fattahi Z, Beheshtian M, Mohseni M, Poustchi H, Sellars E, Nezhadi SH, et al. Iranome: a catalog of genomic variations in the Iranian population. Hum Mutat. 2019; 40:1968–84.

[61] Lee JM, Thong MK. Genetic counseling services and development of training programs in Malaysia. J Genet Couns. 2013; 22:911–6.

[62] Halim-Fikri H, Etemad A, Abdul Latif AZ, Merican AF, Baig AA, Annuar AA, et al. The first Malay database toward the ethnic-specific target molecular variation. BMC Res Notes. 2015; 8:176.

[63] Plazzer JP, Macrae F. DNA variant databases: current state and future directions. Methods Mol Biol. 2014; 1168:263–73.

[64] Teo YY, Sim X, Ong RT, Tan AK, Chen J, Tantoso E, et al. Singapore genome variation project: a haplotype map of three Southeast Asian populations. Genome Res. 2009; 19:2154–62.

[65] 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.

[66] Indonesia MoHotRo. Pedoman Nasional Pelayanan Kedokteran Tata laksana Thalasemia. Indonesia: MHRI; 2018.

[67] Rujito L, Basalamah M, Mulatsih S, Sofro AS. Molecular scanning of beta-thalassemia in the southern region of Central Java, Indonesia; a step towards a local prevention program. Hemoglobin. 2015; 39:330–3.

[68] Rujito L. Talasemia: genetik dasar dan pengelolaan terkini. Indonesia Universitas Jenderal Soedirman; 2019.

[69] Joseph L, Cankovic M, Caughron S, Chandra P, Emmadi R, Hagenkord J, et al. The spectrum of clinical utilities in molecular pathology testing procedures for inherited conditions and cancer: a report of the association for molecular pathology. J Mol Diagn. 2016; 18:605–19.

[70] Schrijver I, Farkas DH, Gibson JS, Lyon E, Committee AMPE. The evolving role of the laboratory professional in the age of genome sequencing: a vision of the association for molecular pathology. J Mol Diagn. 2015; 17:335–8.

[71] Harrison SM, Biesecker LG, Rehm HL. Overview of specifications to the ACMG/AMP variant interpretation guidelines. Curr Protoc Hum Genet. 2019; 103:e93. doi: 10.1002/cphg.93

[72] Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015; 372:793–5.

[73] Lin JZ, Long JY, Wang AQ, Zheng Y, Zhao HT. Precision medicine: in need of guidance and surveillance. World J Gastroenterol. 2017; 23:5045–50.

[74] Gavan SP, Thompson AJ, Payne K. The economic case for precision medicine. Expert Rev Precis Med Drug Dev. 2018; 3:1–9.

[75] Alzu'bi AA, Zhou L, Watzlaf VJM. Genetic variations and precision medicine. Perspect Health Inf Manag. 2019; 16(Spring):1a.

[76] Wynn RM, Adams KT, Kowalski RL, Shivega WG, Ratwani RM, Miller KE. The patient in precision medicine: a systematic review examining evaluations of patient-facing materials. J Healthc Eng. 2018; 2018:9541621. doi: 10.1155/2018/9541621

[77] Boyapati RK, Kalla R, Satsangi J, Ho GT. Biomarkers in search of precision medicine in IBD. Am J Gastroenterol. 2016; 111:1682–90.

[78] Hoffman JM, Dunnenberger HM, Kevin Hicks J, Caudle KE, Whirl Carrillo M, Freimuth RR, et al. Developing knowledge resources to support precision medicine: principles from the Clinical Pharmacogenetics Implementation Consortium (CPIC). J Am Med Inform Assoc. 2016; 23:796–801.

[79] Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM ®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 2015; 43(Database issue):D789–98.

[80] Najafzadeh M, Davis JC, Joshi P, Marra C. Barriers for integrating personalized medicine into clinical practice: a qualitative analysis. Am J Med Genet A. 2013; 161A:758–63.

[81] Gubatan J, Levitte S, Patel A, Balabanis T, Wei MT, Sinha SR. Artificial intelligence applications in inflammatory bowel disease: emerging technologies and future directions. World J Gastroenterol. 2021; 27:1920–35.

[82] Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. Comput Mater Contin. 2020; 63:537–51.

[83] Rivera-Muñoz EA, Milko LV, Harrison SM, Azzariti DR, Kurtz CL, Lee K, et al. Clingen variant curation expert panel experiences and standardized processes for disease and gene-level specification of the ACMG/AMP guidelines for sequence variant interpretation. Hum Mutat. 2018; 39:1614–22.

[84] Easteal S, Arkell RM, Balboa RF, Bellingham SA, Brown AD, Calma T, et al. Equitable expanded carrier screening needs indigenous clinical and population genomic data. Am J Hum Genet. 2020; 107:175–82.

[85] Kessler MD, Yerges-Armstrong L, Taub MA, Shetty AC, Maloney K, Jeng LJB, et al. Challenges and disparities in the application of personalized genomic medicine to populations with African ancestry. Nat Commun. 2016; 7:12521. doi: 10.1038/ncomms12521