








LITERATURE REVIEW

The Role of Genome-Wide Association Study in Pulmonary Disease Diagnostics: A Review in Medical Bioinformatics

Rossalyn Andromeda¹, Irandi Putra Pratomo^{2,3,4*}, Ahmad Fauzi⁵, Tutug Kinasih²,
Prasenohadi Prasenohadi^{1,4}, Arif Santoso^{4,6}, Kentaro Tokumo⁷

¹Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Universitas Indonesia/Persahabatan National Respiratory Referral Hospital, Jakarta, Indonesia.

²Department of Pulmonology and Respiratory Medicine, Faculty of Medicine/Pulmonology and Respiratory Medicine Staff Group, Universitas Indonesia Hospital, Universitas Indonesia, Depok, Indonesia.

³Bioinformatics Core Facilities, Indonesian Medical Education and Research Institute, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia.

⁴Indonesian Medical Association/Indonesian Society of Respiriology, Jakarta, Indonesia.

⁵Nursing Unit, Universitas Indonesia Hospital, Universitas Indonesia, Depok, Indonesia.

⁶Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Hasanuddin University, Makassar, Indonesia.

⁷Department of Clinical Oncology, Hiroshima University Hospital, Hiroshima, Japan.

ARTICLE INFO

Article history:

Received 4 August 2023

Received in revised form

21 December 2023

Accepted 28 January 2024

Available online 30 January 2024

Keywords:

Computational biology,
Genome-wide association study,
Human & health,
Respiratory system diagnostic
techniques,
Translational medical science.

Cite this as:

Andromeda R, Pratomo IP, Fauzi A, *et al.*
The Role of Genome-Wide Association
Study in Pulmonary Disease Diagnostics:
A Review in Medical Bioinformatics. *J
Respi* 2024; 10: 76–84.

ABSTRACT

Biomedical science, which initially required only conventional research in the laboratory, currently involves information technology and has created bioinformatics in its development. Bioinformatics, a branch of biology, quantitatively analyzes information within biological macromolecules using software. Contemporary applications of bioinformatics have advanced biotechnological, medical, and pharmaceutical practices. Among the established applications of bioinformatics is diagnosing lung diseases using the genome-wide association study (GWAS) technique. Owing to sequencing technology and rapid computational methods, this technique is applied to analyze the link between genes with essential traits in the population, thus mapping the target genes to diagnose and treat diseases. The lung diseases diagnosed using GWAS include the responsible locus in asthma, chronic obstructive pulmonary disease (COPD), and lung cancer. Moreover, it can identify the treatment for COPD and suggest a new locus in lung cancer. Advancing the current gene-mapping technology demands genotype and phenotype data to study disease-linked genomes. Currently, bioinformatics is barely known and receives little attention in Indonesia. However, it can grow rapidly through open-source basis data and cross-disciplinary collaboration.

INTRODUCTION

Bioinformatics synthesizes biology and information technology. It is defined as a technology to collect, store, analyze, interpret, disseminate, and apply biological molecular data. In its application, data considered key factors of a phenomenon are used, i.e., genetic macromolecule, including deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein. The most integral tool in bioinformatics is software supported by connections and databases.

Currently, large accessible databases include the GenBank DNA for nucleotide database, the Swiss-Prot protein database for amino acid sequence, and the Protein Data Bank for 3D protein structures. Through bioinformatics, all data from genome projects can be stored systematically and rapidly with high accuracy and analyzed for specific purposes. Among the established applications of bioinformatics is diagnosing lung diseases using the genome-wide association study (GWAS) technique.^{1,2}

*Corresponding author: irandiputra@ui.ac.id



GWAS technique involves rapid scans of complete sets of DNA or genomes from several individuals to identify genetic variation linked with a particular disease. Researchers may use the information to detect, treat, and prevent diseases. GWAS is valuable in diagnostics, especially in identifying genetic variation associated with general and complex diseases, including asthma, cancer, diabetes, cardiac disease, and psychiatric disorders.³

Genetics has identified the role of telomere and surfactant production genes in pulmonary fibrosis. The heterogeneity of lung cancer GWAS is found in each risk-associated locus. Another established role of GWAS in lung cancer is determining groups of patients who respond to therapies and the susceptibility of another group to the side effects of medications. In chronic obstructive pulmonary disease (COPD), GWAS is applied for responsible locus and its treatment, including diagnostic scoring called polygenic risk score (PRS), to predict disease risk. Meanwhile, most asthma GWAS reported single nucleotide polymorphisms (SNPs) at the 17q12-21 locus. In this study, the authors elaborate on the general framework of GWAS data analysis and the role of integrated bioinformatics, which elucidates disease pathology and its pharmacotherapy, especially in lung disease.⁴⁻⁷

APPLICATION OF CLINICAL GENOMICS

Bioinformatics application in medicine is widely known as translational bioinformatics (TBI). American Medical Informatics Association stated that TBI is a developed storage, analysis, and interpretation method

to improve biomedical and genomic data transformation for proactive, predictive, preventive, and participative health practices. The TBI developed a method to convert abundant data into health data. Russ Altman of Stanford University defined TBI as an informatic method to correlate biological entities, including genes, proteins, and small molecules, with clinical entities, such as diseases, symptoms, and medication.³

In [Figure 1](#), the Y-axis describes the central dogma of informatics, converting data into information and information into knowledge. The X-axis is the spectrum of translation from desk work to the patient. TBI incorporates data and knowledge spectrum, bridging the gap between laboratory research and its clinical application. TBI applications are classified into four, (1) Clinical big data or use of electronic medical records for genomic study or others, (2) Genomic and pharmacogenomic study and its application in clinical practice, (3) Omics study for medication development and reuse, and (4) Individual genomic test, including several ethical, law, and social issues associated with the service. Genomics is a discipline addressing the structure, function, evolution, and mapping of the genome to define and qualify the gene.³

Genomics study is the most popular of the four bioinformatics application categories and most implemented in clinical practice. In general, genomics study is categorized into structural genomics study, characterizing all physical traits of the genome, and functional genomics study, describing transcriptome with a total number of information containing molecules expressed by the gene of an organism and proteome. GWAS is an instance of functional genomics study.³

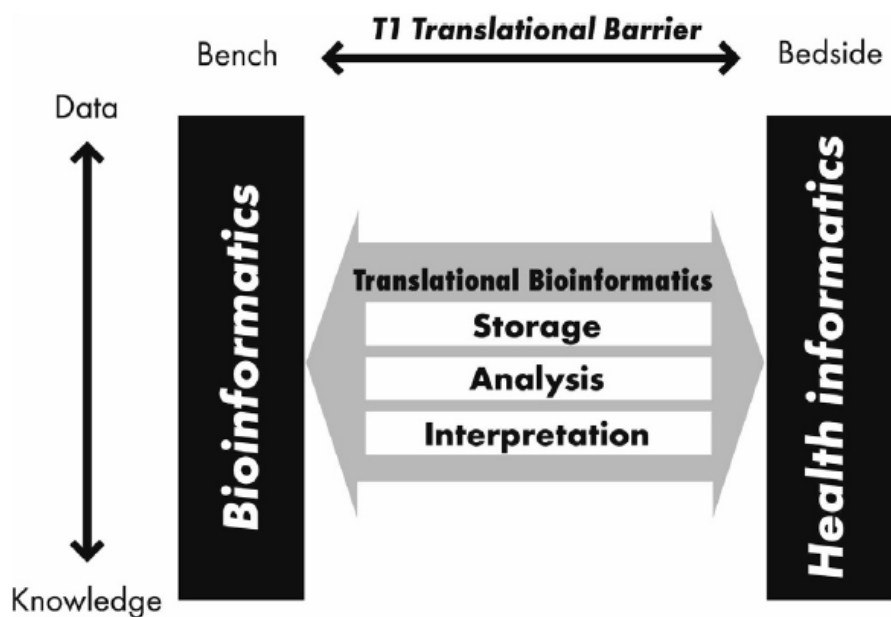


Figure 1. Translational bioinformatics in the health context⁷

GENOME-WIDE ASSOCIATION STUDY

GWAS identifies millions of SNPs in the human genome, where its subset compiles common genetic variations through a phenomenon called linkage disequilibrium (LD). LD occurs when a group of associated SNPs is likely more inherited than coincidentally inherited. GWAS has developed remarkably over the past 15 years in terms of sample size and number of traits involved, with 128,550 associations in 4,000 publications reported in the GWAS catalog.⁸

GWAS Definition

GWAS can identify SNP genetic variations linked with certain diseases and involves the whole genome of several individuals. The GWAS approach includes a complete set of DNA or genomes of several individuals affected or not by diseases. GWAS identifies correlations between genetic markers and phenotypic traits by analyzing thousands of individuals in case-control or quantitative studies.⁹

GWAS is a population-based study aiming to determine SNPs in different loci. SNPs are variations in a single nucleotide base structure, namely adenine, thymine, cytosine, or guanine, of a genome, resulting in genetic variations in a population.¹⁰

GWAS aims to identify correlations between genotype and phenotype through different tests of allele frequency in interindividual genetic variation. It can estimate the copy of variation number or sequence in the human genome, although most GWAS genetic study variation also investigates SNPs. GWAS reports that correlated SNP sequence are significantly associated with a trait called genomic risk locus.⁸

GWAS Workflow

GWAS experimental workflow incorporates several steps described in [Figure 2](#).⁹ GWAS steps: (a) Determining study population. GWAS requires a large data to produce GWAS-appropriate data. The data may be determined by utilizing special software, such as CaTs, a simple platform for estimating power in extensive genetic association studies such as the Genetics Personality Consortium. Data may be collected from cohort studies or available genetic and phenotypic information from Biobanks or repositories by disease-focused recruitment. Randomization should be thoroughly considered, and recruitment strategy should not lead to collider bias.^{8,11,12} (b) Genotype data may be gathered through microarrays to identify general

variants or next-generation sequencing methods, such as whole genome sequencing (WGS) or whole exome sequencing, to identify rare variants. Microarray is the most commonly applied method for GWAS due to its more reasonable cost than that of next-generation sequencing.⁸ (c) Data processing. Selected data includes anonymous identity numbers, coded familial relationships, gender, phenotype information, and all genotype variations and information. Data input requires precise quality control to produce more significant results. Quality control involves steps on many levels, including determining DNA and genotypes, deleting incomplete SNPs, detecting population strata, and estimating principal components.⁸ (d) Genotype data is gradually examined, wherein the expected genotype is tested based on the information of population reference or on a repository, such as in the case of the 1,000 genomes project, a project aiming to describe human genetic variation in general by applying WGS on various individuals from populations or the transomics for precision medicine, a program describing the genetic and biological architecture of heart, lung, blood, and sleep disorder, with its ultimate goal of improving diagnosis capacity, medication, and disease prevention. In this case, SNP1 and SNP3 genotypes are estimated based on other SNP genotype tests.^{8,13,14} (e) Genetic association test is conducted on every genetic variant using an appropriate model, including additive, nonadditive, linear, or logistic regression. Confounding factors are controlled, including population strata. The outcome is examined for irregular patterns, and summary statistics should be established.⁸ (f) Several small cohorts are combined using standard statistical analysis. Meta-analysis is presented in summary statistics, which may be outlined using the fixed or random effect models to assess the heterogeneity of results.⁸ (g) The result may be replicated for independent cohorts. External replication for independent cohorts should match the lineage, and the identified cohort should not share with any individual or family member.⁸ (h) In silico GWAS analysis uses external source information. It may include in silico mapping, SNPs to genetic mapping, and gene to functional mapping, including pathway analysis, genetic correlation analysis, Mendel randomization, and polygenic risk prediction. The functional hypothesis may be examined following GWAS using experimental techniques, such as clustered regularly interspaced short palindromic repeats (CRISPR), or the result can be confirmed through human trait or disease models.⁸

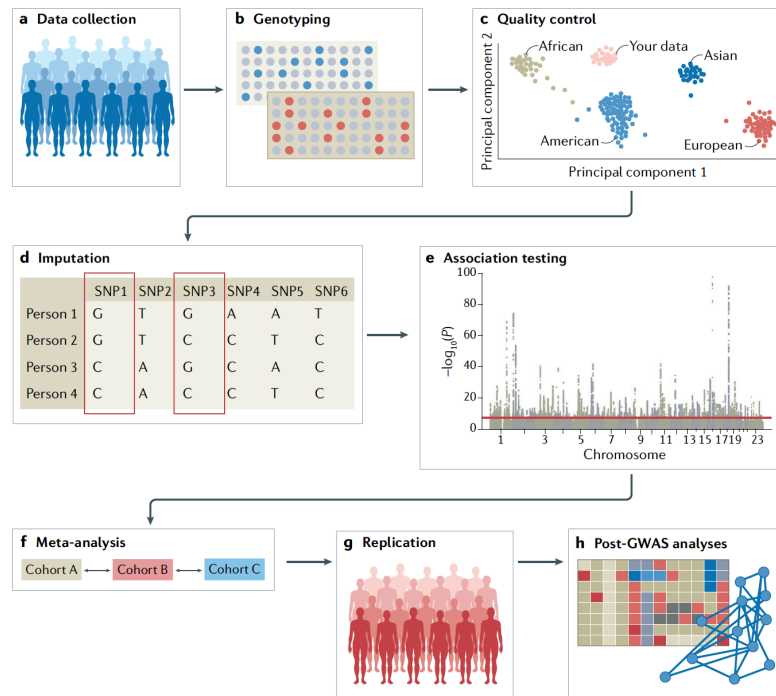


Figure 2. GWAS steps¹⁰

APPLICATION OF GWAS IN PULMONARY DISEASE

Role of GWAS in pulmonary fibrosis

GWAS shows that pulmonary fibrosis is a sporadic disease with a locus linked with another locus containing genes associated with lung defense, telomere maintenance, and cell adhesion. The genetic alteration has a 30% chance of idiopathic pulmonary fibrosis (IPF), and the molecular mechanism of the disease is mostly unclear. A study conducted by Allen, *et al.* (2020) is currently the largest collaborative GWAS IPF study.⁴ The study population involved all patients and control variables of European descent, with 2,668 patients with IPF and 8,951 controls. Two sets of cases and independent controls were included in the cohort, with a total of 1,467 patients with IPF and 11,874 controls.^{4,15,16}

The study by Allen, *et al.* (2020) confirmed that risk in locus 11p15 was promoted by a promotor variant of MUC5B.⁴ Three new signals of association adjacent to KIF15 in 3p21.31, MAD1L1 in 7p22.3, and DEP domain-containing mTOR-interacting protein (DEPTOR) in 8q24.12 were also identified and remained significant following adjustment for diversity. The authors concluded that the variant adjacent to KIF15 was associated with low gene expression in the brain, and the MAD1L1 variant was associated with low gene expression in the heart. The variant adjacent to DEPTOR was associated with low gene expression in the lungs, colon, and integument. Furthermore, DEPTOR encodes a protein interacting with mTOR and inhibits its kinase activities. The authors concluded that

the variant contributed to IPF pathogenesis by reducing mTOR signal inhibition, thus promoting collagen synthesis and fibrogenesis induction by transforming growth factor β 1 (TGF- β 1).^{4,15,16}

The study by Allen, *et al.* (2020) is a huge milestone in the TBI for pulmonary fibrosis.⁴ However, the study only involved European descent, preventing generalization for a large population. Around 60–70% of non-Hispanic whites carry at least a copy of MUC5B rs35705950 T, the most powerful and replicated variant of IPF risk. However, this variant is rarely identified in the Korean population with IPF. Thus, enhancing our understanding of genetic risk for IPF in different racial groups is crucial.^{4,15,16}

Further study is recommended to identify the genetic factors in disease risk of all phenotype spectrums of interstitial lung disease (ILD). GWAS has exclusively focused on IPF until now. However, a study addressing rheumatoid arthritis and chronic hypersensitivity pneumonitis associated ILD reported a similar prevalence of MUC5B promotor polymorphism and rare variant in telomere-associated gene as reported in the IPF study.¹⁵ In the future, ILD genetic study is expected to identify relationships between genetic risk and certain ILD phenotypes.^{4,15,16}

Role of GWAS in lung cancer

The first GWAS in lung cancer was reported in 2008.⁵ Three independent studies identified cancer-susceptible locus in chromosome 15q.⁵ A study has identified two SNPs linked with lung cancer, chromosome 15q25.⁵ Other studies suggest new loci with cancer risk.⁵ Two SNPs residing in separate genes,

BAG6 and MSH5, are 600 kilobases apart in chromosome 6p21 and are significantly correlated with lung cancer. The strongest association is also identified in chromosome 5p15 in CLPTMIL, apart from 15q25 and 6p21. Locus 5p15 is reported as cancer-susceptible and supported by a GWAS study with a larger population size than that of prior studies.⁵

Three lung cancer-susceptible loci were identified in 2008, including 15q25, 6p21, and 5p15. 15q25 locus is linked with smoking habits, the risk allele is correlated with high tobacco consumption. Conversely, the 5p15 and 6p21 loci are not associated with smoking behavior. It is also reported that a DNA variant in 5p15 is linked with a subtype of histological lung cancer and a high frequency of risk allele on a type of adenocarcinoma. GWAS meta-analysis in 2010 confirmed that lung cancer locus at 6p21 is strongly associated with squamous cell carcinoma.⁵

Another GWAS meta-analysis in 2014 involved the 1,000 genomes project and tested rare SNPs, which were ignored in previous studies.⁵ A rare genetic variant was linked with squamous cell carcinoma at the BRCA2 gene of 13q13.1 and CHEK2 of 22q12.1. 3q28 locus was associated with lung adenocarcinoma found in the Asian population. A recent GWAS on lung cancer of European descent population involved 29,266 cases and 56,450 controls.⁵ The GWAS emphasized genetic heterogeneity in all subtypes of histological lung cancer and reported new loci of lung cancer, including 1p31.1, 6q27, 8p21.1, and 15q21.1, and adenocarcinoma, including 8p12, 10q24.3, 11q23.3, and 20q13.33. The lung cancer locus previously reported is highly specific to squamous cell carcinoma in this study, including 6p21.33, 12p13.33, and 22q12.1.⁵

Genetic heterogeneity in lung cancer susceptibility is observed among European and Asian populations. The variant with the highest susceptibility to lung cancer at the 15q25 locus has the lowest allele frequency among the Asian population. Furthermore, the 6p21 variant found in Europe is not polymorphic in Asia. Therefore, conducting GWAS on the Asian population is necessary. GWAS of non-small cell carcinoma lung cancer (NSCLC) in the Korean population reported a new locus at the 3q29 chromosome.⁵ The study confirmed the susceptibility of the 5p15 locus of Koreans. Two susceptible loci were identified in the Japanese and Korean populations, confirming a new locus of 3q28. 5p15 and 3q28 locus are also confirmed by GWAS involving the Han in China.⁵ Furthermore, two new loci at 13q12 and 12q12 were also identified. The same GWAS involving a large sample identified five new lung cancer loci: 10p14, 5q32, 20q13, 5q31.1, and 1p36.32. A recent GWAS addressing squamous cell carcinoma in the Han

population reported a new locus at 12q23.1.⁵ Rare variations at 6p21.33, 20q11.21, and 6p22.2 were also identified by applying an exome genotype chip. 5p15, 3q28, and 6p21 loci association with adenocarcinoma were also confirmed but not for 13q12.12 and 22q12.2 loci.⁵

Another established role of GWAS in lung cancer is determining groups of patients who respond to therapies and the susceptibility to side effects of medications. In the case of determining groups responding to treatment, Chang (2017) reported that GWAS identified SNPs at 4q12, which was linked with progression-free survival in patients with lung adenocarcinoma and undergoing first-line therapy with epidermal growth factor receptor tyrosine kinase inhibitor (EGFR-TKI).¹⁷ Another GWAS project by Li, *et al.* (2021) from Harbin Medical University in China, which applied a drug-gene interaction database, suggested that the SNPs variant of rs191188930 was correlated with the cardiotoxic effect of TKI types, such as sunitinib, pazopanib, sorafenib, dasatinib, and nilotinib.¹⁸

A study by Xu, *et al.* (2020) revealed that deletion of exon 19 EGFR in NSCLC is linked with TKI therapy sensitivity.¹⁹ The study also revealed that 50% of EGFR exon 19 deletion occurred at the S752 locus, followed by 21% at the A750 and T751 loci, and 2.6% at the L747, E749, and p753 locus.¹⁹ In conclusion, the study approved TKI therapy as first-line therapy for the advanced NSCLC stage.¹⁹ Proper detection methods for medication sensitivity are integral to treating the advanced stage, suggesting the importance of mutation tests on the locus linked with EGFR exon 19 deletions.¹⁹

Role of GWAS in asthma

Since the first GWAS of asthma, which identified a variant at the 17q21 locus and its correlation with expression of the orosomucoid-like3 (ORMDL3) gene, the locus has become the primary focus of studies. The region possesses haploblock, a genomic sequence with high LD and SNP-dense with overlapping of at least four genes, including IKAROS family zinc finger 3 (IKZF3), zona pellucida binding protein 2 (ZPBP2), gasdermins B (GSDMB), and ORMDL3. The locus has been extended to incorporate the core region involving post-GPI attachment to proteins 3 (PGAP3) and erb-b2 receptor tyrosine kinase 2 (ERBB2) at the proximal end and gasdermins A (GSDMA) at distal ends, which may represent independent loci for asthma. Moffatt, *et al.* from the GABRIEL Consortium Project funded by the European Commission, as cited by Kim, *et al.* (2019), analyzed a subgroup of childhood-onset asthma and reported a specific association between the region, although the study also includes a group of individuals

with late-onset asthma to be analyzed in a meta-analysis based on asthma GWAS consortium.⁷

Meta-analysis of the transnational asthma genetic consortium GWAS also revealed that the 17q12-21 locus centered in ORMDL3/GSDMB is specifically linked with early-onset asthma.⁷ The meta-analysis concluded that asthma-related signals adjacent to PGAP3/ERBB2 and ORMDL3/GSDMB might affect the risk for asthma through different gene expressions. Notably, the genotype effect of the locus on risk and protection of asthma has been reported to be altered by early exposure to tobacco, smoking environment, and rhinovirus within the first three years of life. SNPs at the locus are associated with childhood-onset asthma in European, Asian, and Latin American populations.⁷

A study conducted by Stein, as cited by Kim, *et al.* (2019), also revealed that SNPs at 17q12-21 were correlated with expression-quantitative trait loci (eQTL) asthma in GSDMA, ORMDL3, GSDMB, and PGAP3 within immune and lung cells.⁷ However, the role of the 17q12-21 gene in asthma pathogenesis remains underexplored. The limitation of GWAS is identifying SNPs without providing information about affected genes or causal SNPs. Consequentially, almost every GWAS reported proximate gene as a potential asthma gene, yet not every SNP affects the function or closest gene expression despite SNPs residing within the gene. A study by Luo, *et al.*, as cited by Kim, *et al.* (2019), combined results of asthma GWAS and available eQTL data of epithelial cells in small and large human airways.⁷ The study revealed that the GWAS of asthma attacks was enriched due to eQTL of airway epithelial cells, and the gene regulated by asthma GWAS locus in the epithelium was enriched in the immune response.⁷

Role of GWAS in COPD

A recent study addressing COPD conducted by Moll, *et al.* (2020) involving 257,811 individuals from various databases, including the International Cancer Genome Consortium and United Kingdom (UK) Biobank, succeeded in identifying 82 loci.⁶ Thirty-five new loci were identified. Considering the significant correlation between lung function test and COPD, the authors examined the association between all variants in each locus and forced expiratory volume in the first second (FEV1) or VEPI/forced vital capacity (FVC) among 79,055 individuals and reported 13 loci, including C1orf87, DENND2D, DDX1, SLMAP, BTC, FGF18, CITED2, ITGB8, STN1, ARNTL, SERP2, DTWD1, and ADAMTSL3, with significant correlation ($p < 0.05$).⁶

The authors also recognized that none of the 35 new loci were significantly associated with tobacco consumption and asthma history.⁶ The study also

described up to 7.0% of phenotypic variants of COPD out of 82 significantly associated genome variants.⁶ Using 10% data on COPD prevalence, the authors revealed a 48% increase in COPD phenotypic variant described by genetic locus.⁶ In contrast, there is only a 4.7% increase in 22 loci reported in previous COPD GWAS.⁶

Through GWAS, Moll, *et al.* (2020) also attempted to identify medication targets for genetic and genome-wide levels.⁶ Of 472 potential genes, 59 were the target for at least one established or in-development COPD medication type. New loci that became the target of COPD medication and lung function included ABHD6, CDKL2, GSTO2, KCNC4, PDHB, SLK, and TRPM7. The authors also managed to identify reusable medications for COPD through transcriptome-wide associations and drug-induced gene expression.⁶

Aside from determining the responsible locus for COPD and its treatment, GWAS may also be applied for diagnostic scoring, such as PRS. PRS could predict disease risk by using summary statistics of GWAS to determine high-risk individuals requiring immediate intervention. Figure 3 describes the steps for performing PRS.⁴

The first step is to consider the summary statistic of GWAS and recognize each effect of SNPs on the predicted phenotype. The second step is to compare individual genotypes. The previous figure shows the genotype data of 4 SNPs in four individuals. The third step is calculating the PRS for every individual and combining the effect size of all risk alleles. The fourth step is to perform linear regression analysis on calculated PRS to estimate its effect.⁸

Concerning COPD GWAS, Moll, *et al.* (2020) estimated the PRS of COPD by using GWAS of lung function (FEV1 or FEV1/FVC) of UK Biobank and SpiroMeta data.⁶ The authors examined the PRS in nine multiethnic cohorts to identify the association between the score and moderate to severe COPD, described as FEV1/FVC < 0.7 and FEV1 $< 80\%$ of the prediction score.⁶ The association was examined by a logistic regression model, which was adjusted to age, gender, height, Brinkmann index, and the main component of the hereditary gene. The study also assessed computed tomography (CT) scan phenotype, which reflected pathological parenchyma and airway, including patterns of reduced lung size.⁶

Moll, *et al.* (2020) revealed the formula of $PRS_{\text{Combination}} = 0.43847 \times PRS_{\text{VEPI}} + 0.58833 \times PRS_{\text{VEPI/KVP}}$.⁶ This risk scoring has an odds ratio (OR) of 1.81 [95% CI 1.74–1.88] for the European population and OR of 1.42 [95%CI 1.34–1.51] for the non-European population. This PRS is superior to the previous genetic scoring risk and, when combined with

clinical data, including age, gender, and Brinkmann index, can predict COPD better than the old model, with an area under the curve of 0.80 [0.79–0.81] vs 0.76 [0.75–0.76], respectively. The authors also confirmed

that PRS was linked with CT scan phenotype, qualitative and quantitative emphysema size, and reduced lung size patterns.⁶

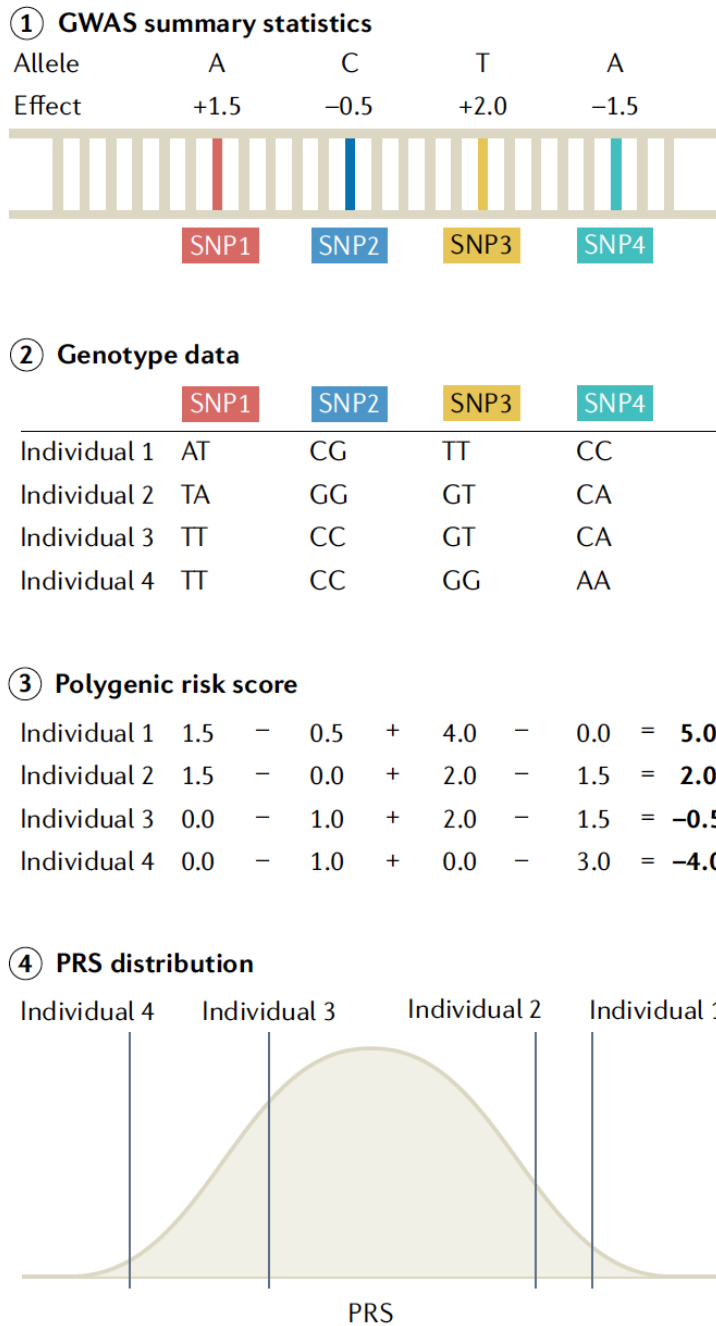


Figure 3. Steps for calculating PRS⁸

LIMITATION OF GWAS

The limitation of GWAS is its weak statistical power due to the simplicity of effect size. Several studies are combined in a meta-analysis to overcome this shortcoming. Meta-analysis improves the power to detect association signals and reduce false positives by increasing sample size and surveying the number of variants in every genome. A major challenge in GWAS interpretation lies in several variants not replicated in

studies, reflecting the complexity and recognizable heterogeneity in the observed diseases. Furthermore, deciding the functional relevance of GWAS results remains challenging since most risk variants are introns. Therefore, integrating GWAS data with additional relevant biological information is pivotal to describing the phenomenon and biological intervention that connects genetic variation and phenotype traits.^{8,20}

ROLE OF GWAS IN THE FUTURE

GWAS addresses genetic differences among individuals' genomes to identify the relationship between genetic code/genotype and its expression in individual/phenotype. The method requires several genetic variants from the genomes of many individuals, these are examined to identify their association. The

method demands an understanding of gene inheritance and statistics since it involves seeking patterns in a large amount of data that requires proper interpretation by researchers to obtain beneficial results, especially in improving the diagnostic accuracy of various diseases and their treatment tailored for each individual through genetic information.^{20,21}

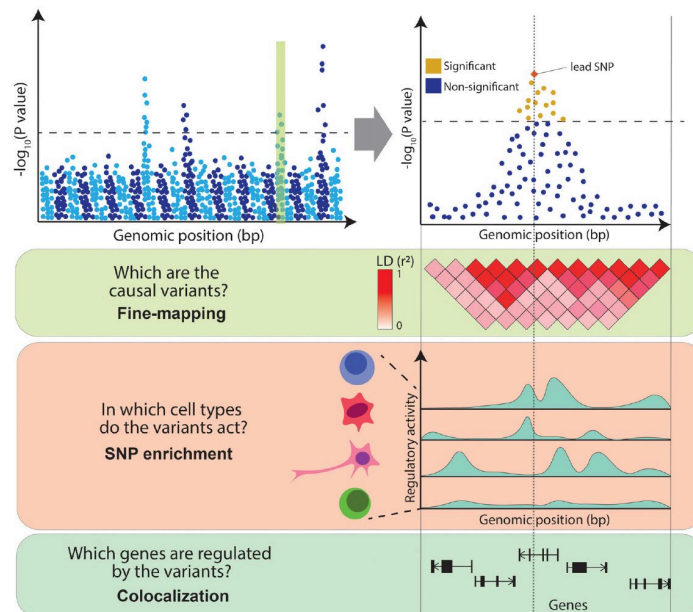


Figure 4. Challenges in interpretations of GWAS association. From top: Manhattan plot describes the link between genetic variants and certain traits, disease at the genomic level (left panel), and locus (right panel). The variants above the dotted line represent all significant associations across the genome. The panel below describes the major challenges in interpreting GWAS association: High LD across variants (shaded in red), level of regulatory activity variable in all types of cells (different peaks represent various levels of chromatin activity), and several genes in related locus.²⁰

SUMMARY

A major challenge in bioinformatics, as described in Figure 4, is obtaining valuable or relevant knowledge from millions of data. GWAS method facilitates finding new molecular biomarkers, which assists in early disease detection and diagnosis, leading to highly selective and effective prevention and treatment. In the context of treatment, GWAS can predict medication effects in terms of efficacy and safety (potential side effects, especially the harmful ones) that are linked to the genetic variability of each individual.

Acknowledgments

The authors would like to thank Dr. Triya Damayanti, Puspita Widyaningrum, Dr. Aulia Pranandrari, Defri Dwi Yana Putra, and M. Ferdiansyah for their technical support.

Conflict of Interest

The authors declared there is no conflict of interest.

Funding

This study was self-funded by the authors.

Authors' Contributions

Principal manuscript writer and data collector: RA. Concept writers: RA and IPP. Manuscript reviewers: IPP, AF, P, AS, and KT. Manuscript writers and proofreaders: TK, AF, and KT.

REFERENCES

1. Aprilyanto V, Sembiring L. *Bioinformatika*. 1st ed. Yogyakarta: Innosain, <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1155075> (2017).
2. Ethica SN. *Pengantar Bioinformatika untuk Mahasiswa Laboratorium Medis*. Sleman: Deepublish, <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1279557> (2019).
3. Tenenbaum JD. Translational Bioinformatics: Past, Present, and Future. *Genomics Proteomics Bioinformatics* 2016; 14: 31–41. [PubMed]
4. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019; 201: 564–574. [PubMed]
5. Bossé Y, Amos CI. A Decade of GWAS Results in Lung Cancer. *Cancer Epidemiol Biomarkers Prev* 2018; 27: 363–379. [PubMed]

6. Moll M, Sakornsakolpat P, Shrine N, *et al.* Chronic Obstructive Pulmonary Disease and Related Phenotypes: Polygenic Risk Scores in Population-Based and Case-Control Cohorts. *Lancet Respir Med* 2020; 8: 696–708. [PubMed]
7. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* 2019; 11: 170–187. [PubMed]
8. Uffelmann E, Huang QQ, Munung NS, *et al.* Genome-Wide Association Studies. *Nat Rev Methods Prim* 2021; 1: 59. [Journal]
9. Health NI of. National Human Genome Research Institute (NHGRI). *The NIH Almanac*, <https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-human-genome-research-institute-nhgri> (2023).
10. Chen X, Sullivan PF. Single Nucleotide Polymorphism Genotyping: Biochemistry, Protocol, Cost and Throughput. *Pharmacogenomics J* 2003; 3: 77–96. [PubMed]
11. Skol AD, Scott LJ, Abecasis GR, *et al.* Joint Analysis is More Efficient than Replication-Based Analysis for Two-Stage Genome-Wide Association Studies. *Nat Genet* 2006; 38: 209–213. [PubMed]
12. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: Design of Linkage and Association Genetic Mapping Studies of Complex Traits. *Bioinformatics* 2003; 19: 149–150. [PubMed]
13. Auton A, Abecasis GR, Altshuler DM, *et al.* A Global Reference for Human Genetic Variation. *Nature* 2015; 526: 68–74. [PubMed]
14. Taliun D, Harris DN, Kessler MD, *et al.* Sequencing of 53,831 Diverse Genomes from the NHLBI TOPMed Program. *Nature* 2021; 590: 290–299. [Journal]
15. Spagnolo P, Kropski JA. Genome-Wide Association Studies in Idiopathic Pulmonary Fibrosis: Bridging the Gap between Sequence and Consequence. *Am J Respir Crit Care Med* 2019; 201: 508–509. [PubMed]
16. Hall R, Hall IP, Sayers I. Genetic Risk Factors for the Development of Pulmonary Disease Identified by Genome-Wide Association. *Respirology* 2019; 24: 204–214. [PubMed]
17. Chang I-S, Jiang SS, Yang JC-H, *et al.* Genetic Modifiers of Progression-Free Survival in Never-Smoking Lung Adenocarcinoma Patients Treated with First-Line Tyrosine Kinase Inhibitors. *Am J Respir Crit Care Med* 2017; 195: 663–673. [PubMed]
18. Li Y, Wang W, Gao R, *et al.* Genome-Wide Prioritization Reveals Novel Gene Signatures associated with Cardiotoxic Effects of Tyrosine Kinase Inhibitors. *Oncol Lett* 2021; 21: 94. [PubMed]
19. Xu C, Lei L, Wang W, *et al.* Molecular Characteristics and Clinical Outcomes of EGFR Exon 19 C-Helix Deletion in Non-Small Cell Lung Cancer and Response to EGFR TKIs. *Transl Oncol* 2020; 13: 100791. [PubMed]
20. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*; 11, <https://www.frontiersin.org/articles/10.3389/fgene.2020.00424> (2020).
21. Zhu Z, Li J, Si J, *et al.* A Large-Scale Genome-Wide Association Analysis of Lung Function in the Chinese Population Identifies Novel Loci and Highlights Shared Genetic Aetiology with Obesity. *Eur Respir J*; 58. Epub ahead of print October 2021. [PubMed]