Genomics - Part 2: It's the 'junk DNA' that matters

'Next-generation sequencing is the microscope of the 21st century*,' so it is claimed. It's not within the framework of Public Health to work with the technology, but it should follow how genomics influences common diseases for risk assessment and prevention.

Health and diseases are influenced by the environment and genetics. The individual susceptibility to numerous diseases is influenced by genetic variation. Up to recently, the focus lay on 'genetic discovery efforts targeting variants with large effects' as a review paper about the 'history of human disease genetics' saw it (1). Less of interest were 'alleles that influence predisposition to common diseases.' Common diseases are significant public health problems, while those diseases commonly understood as 'genetic diseases,' or in memory of Mendel's work on heredity, named Mendelian diseases (page 14 ff) (2), are the domain of clinical medicine. Public health decision-makers and politicians' moderate interest in genetics favored curative therapy and supported research in genetics for personalized medicine. A telling example was the difference in attitude towards a proposal to look into the effects of environmental factors on child health versus an extended database on genomic health. At the same time, the latter was the more favored one (1, 3).

<u>Disappointment after publishing the 'finished' reference sequences of the Human Genome Project</u>

In 2003, it seemed that an overwhelming amount of human DNA, which means 97 percent of the 3.2 billion bases, had no function. Francis Crick, a hero among genetic scientists, was quoted at that time as saying that most of the DNA was 'little better than junk' (4). However, it was realized soon that the 'junk' is a complex part of our genome and might consist of operational elements (5). The expression 'junk DNA' was replaced by 'non-coding DNA' for protein. As explained in part one of this entry, the conventional understanding is that after the initiation of the primary RNA transcript, the mature mRNA is transferred from the nucleus, and translation occurs in the cytoplasm encoding protein. This is not the case for non-coding DNA but serves gene regulations (page 571) (2). It is now apparent that genetic variants affect diseases and have other significant roles in determining gene function.

Research in Mendelian disorders profited from the rapid advancement of genomic technology. Instead of Mendelian diseases, the term Monogenetic diseases is more clarifying. By 2000, from about 7,000 inherited diseases, about 1,000 could be traced to single genes (1). Celera Genomics (6) and scientists from countries with highly improved technologies and sufficient resources combined their efforts and worked on a massive database known as the 'total' human genome sequence and the Human Genome Project (HGP) (7). The abstract of the publication of Celera Genomics gives an impression of the undertaking. With the shotgun-sequencing method, 2.91 billion base pairs were generated from 27,271,853 high-quality sequence reads from the DNA of five individuals (6).

The immense efforts were undertaken in the hope of creating a complex map of the entire sequence of the human genome that will allow the investigation of the genetic basis of human diseases, in which multiple genes are involved (8). A genetic map describes the location of genes along chromosomes (page 118 ff) (2). However, in reviewing the Human Genome Project, it was concluded that in cases where multiple genes were involved, as in common human diseases, the 'genomic-based mapping techniques' were insufficient (8). For non-communicable diseases (NCDs) such as diabetes or schizophrenia, it was realized that the whole system of 'switches that govern where and how genes are expressed in the body' was overlooked (9).

The function of non-coding regions not fully understood

One of the main functions of non-coding regions is transcription, forming RNA particles. Eukaryotes have three types of polymerase enzymes for transcribing different kinds of genes. The well-known mRNA is described by RNA polymerase II, but also micro RNAs (miRNA), small nuclear RNA (snRNA), and small nucleolar RNA (snoRNA) genes. Polymerase I transcribe ribosomal RNA (rRNA) genes. RNA polymerase III transcribes transfer RNA (tRNA), and 5S rRNA is linked to other ribosomal genes (10).

The apparent, very complex configuration of the RNA system, which by no means is still fully understood, makes it an irresponsible undertaking to inject, on a global basis, mRNA molecules embedded in phospholipids into the human organism, as done with the mRNA vaccines during the COVID 19 affair. It is a very uneasy feeling not knowing what experts, with their goal to embark on 'excellent research' in mind, will come up with in the future while playing around with the generic structure on a population basis.

For instance, microRNAs seem involved in development issues, and small RNAs (sRNAs) are found in 'silencing' genes. While only the function of several different non-coding RNAs (ncRNA) could be identified, while others are still under investigation. Non-coding DNA is required for the 'proper' expression of the genes and is found close to transcription factors, referred to as cis-acting elements. Among cis-acting regulators are promoters, enhancers, and silencers. Close to the transcriptional sites are response elements. Several non-coding regions might be causing diseases in humans, but how these regions function is unclear. Trans-regulatory factors bind to cis-acting sequences and modify gene expression. Introns, as mentioned in part one of this entry, splice the exon, the part which is finally involved in translating polypeptides. Shanmugam et al. (11) give a more detailed overview of the non-coding RNA and DNA.

Experts in genetics have their terms

After the HGP was published more than 20 years ago, genomics gave biomedical research an enormous push. To get an impression of how fast the scientific field of genetics developed, one might compare the latest edition of the standard textbook about genetics published in 2021 (2) with the seventh edition from 2011. Numerous new aspects and findings were included in the latest edition. However, it became increasingly difficult to read and understand genomic publications because of the detection of unique genomic characteristics described by terms used by experts but unknown to the laymen.

What has not changed since we finished high school is the understanding of what is a ‘gene.’ The gene, in the words of an expert, ‘is a segment of DNA on one of the chromosomes, which function as a unit encoding (‘directing’ so to say) to create a certain RNA or polypeptide’ and finally a protein. Based on Mendel findings with peas, there are two copies of genes inherited from the mother and the father, such as resulting in a child's eye color. A specific eye color for one child is part of the phenotype of the child, or one might say the particular trait of that person. The traits can be very complex and might be controlled by many genes. The alternative forms of a single gene are called alleles. Many traits are measured in quantities, so in one of the same population, the height of people is different. It can be calculated as the variance of size in the population (so the industry, knowing the variation of height, can produce fashionable dresses for the youngsters, being either quite tall, have a ‘normal’ height, or are a little bit smaller compared to their friends).

The specific location of genes in a particular chromosome is named loci. A genetic map relates to the actual distance in base pairs of DNA. The mapping of quantitative trait loci (QTLs) helps to explain genetic variant effects (11). If the linkage of the variant is quite close to the trait loci, the probability increases that the markers are inherited. Trait loci along DNA rows add to a phenotype. As mentioned above, inherited genes shape the trait, such as the height, the color of the skin or the hair, or the blood group a particular person belongs to. Variation in the expression of the genes determines the different phenotypes. However, if one gene is expressed in favor of the other one, it influences the effect on the trait and might increase the risk for diseases such as hemoglobinopathies along the different traits of blood groups.

Annotation means that genome analysis is customarily performed, deposited in a gene bank, and published, or simply it means ‘which sequence of DNA does which task’ (page 327ff) (2). The results from the ‘annotation’ of the genome were accumulated on big-data bases, such as Encyclopedia of DNA Elements (ENCODE), GTex (Genotype-Tissue Expression) eQTL Browser, Roadmap Epigenomics Mapping Consortium (12). One such public archive of genetic variations is ClinVar (13). From about one million entries into ClinVar from patients with severe genetic diseases, 47% of variants have uncertain effects or conflicting analytical results (14). Generally, variants clearly define monogenic diseases happen to be rare. Common variants seldom have large impact on diseases. Variants with large effects are close to protein-coding regions, contrary to variants falling outside protein-coding regions. That is why QTLs are a meaningful research variable for interpreting the function of genes.

Genetics of common diseases

The relationship of the genetic background of human disease could be interpreted, given our evolution. Genetic variants with a large effect on disease risk might be partly removed from the population and nowadays impress as rare genetic diseases. In contrast, common variants are less likely to have very large effects on diseases. Nevertheless, the effect, especially in later life, could be life-threatening, with environmental and additional traits. Often, non-coding regions are related to human diseases, which could be large effect variants such as breast, ovarian, and prostate cancer, or fortunately with small individual effects, such as Autism spectrum disorder (ASD) or developmental delay, but are more frequent within the population (14, 15).

Thus, genome technology used to look into the function of key variants differs between rare diseases, such as 'genetic diseases,' and common diseases, such as diabetes mellitus, asthma, and depression. Large data sets of genes and variants combined with advanced laboratory technologies and clinical observations will be the future main approaches to search for causal variants for rare diseases. At the same time, genetic analysis of human populations with 'direct relevance to human physiology and health should dominate research for common diseases' (14).

Common diseases are polygenetic

Common diseases are polygenetic, with many loci related to the phenotype and often joining risk patterns with environmental factors. Gene expression for common diseases depends predominantly on non-coding genome regulation. Those with common diseases often have no family history, but relatives of those with a genetic disposition might have a higher risk of suffering from the disease (15). Single nucleotide polymorphism (SNP) is the genetic basis of common diseases. However, the 'effect size' of loci significantly associated with common diseases is difficult to identify. Developments in computerization and the SNP array technology made it possible to associate thousands of genomic loci to an illness or trait. One technique compared cases (affected) with unaffected (controls). The program related to these efforts is the genome-wide association studies (GWAS) (16).

Besides the case-control endeavor, population studies such as the volunteer cohort of five hundred thousand people between 50 and 70 years old from the UK Biobank (UKB) contributed to the challenge of insight into the polygenicity of common diseases. Individuals with a polygenetic risk for a disease do not necessarily have the disease. This phenomenon might be due to the process called negative selection. Here, alleles with a highly dangerous effect on health tend to be removed from the population, as mentioned above. For instance, from 117 individuals with T2DM SNPs, two individuals within the lowest polygenetic risk have 98 and 92 risk alleles, while two persons are T2DM patients show 131 and 130 T2DM-associated SNPs and are in the highest centile of polygenetic risk. Similarly, 306 women with breast cancer-associated SNPs, two not having the disease, were found to have 273 and 266 risk SNPs, while two other patients with the disease were found with 326 and 322 risk-associated SNPs (15).

Risk assessment in genetics

Risk assessment in common diseases related to omics leaves students, lecturers, and public health experts, including epidemiologists and statisticians, quite bewildered. Common diseases are one of the domains of public health. Our 'daily bread' are relative risk, odds ratio, and the Mantel-Haenszel statistics (page 634 ff) (17). The Dictionary of Epidemiology covers almost two pages with numerous keywords about the risk subject (pages 250 ff) (18). Generally, risk assessment is defined as 'the qualitative estimation of the likelihood of adverse effects that may result from exposure to specified health hazards or the absence of beneficial influence.'

Those in the field of genomics admit that there are several 'traditional' risk factors, and several risk prediction models have existed for quite some time already. Diseases occur independently of genetic conditions, mainly due to environmental factors. However, it is argued that, in addition, genetic variables play a role in the occurrence of common diseases and might even be the 'most

informative risk factor in pre-symptomatic individuals.' Coronary Heart Disease predominantly occurs in older age, but genetic risk scoring could be used as a prevention tool for people in the middle age groups (19). Similarly, risk prediction could be helpful in breast- and prostate cancer, obesity, type 1- and type 2 diabetes mellitus, and Alzheimer's disease (20). Not only 'relative risk' and 'odds ratio' might be a 'household name' for public health, but in the future, 'polygenetic indices' or 'polygenetic risk scoring' should be a common entity observed by public health.

Polygenetic scoring for common diseases

Polygenetic scoring for common diseases was made possible not only through the rapid development from 'Sanger sequencing' to the 'whole genome shotgun technique' up to 'next-generation sequencing' as mentioned in part one of this entry. Likewise, rapid advancements in data processing and specialized statistical methods applied to genetic investigation resulted in thousands of research results now available in the genome-wide association study (GWAS).

A user guide explains how polygenetic scoring works (21). For 47 phenotypes in 11 databases, 'DNA-based predictors' are given. Besides GWAS, other data sources were used, including the UK Biobank, as mentioned above. All available summary statistics were used for each phenotype, and the highest SNP heritability was kept. Through complex statistics, an additive SNP factor was identified as a true 'regressor.' The polygenetic score, also named polygenetic index, was understood as a proxy variable, which could be 'biased' by the additional variables necessary for the phenotype. Phenotypes included behavioral variables such as cognitive ability and education, alcohol consumption and smoking, and anthropometry as BMI and height. With this technique, polygenic risk scores (PRS) for inflammatory bowel disease, atrial fibrillation, and glaucoma are revealed (15). An attempt is made to generate a PRS catalog as a database in a standardized format for systematic evaluation (20).

How do family history and polygenetic risk score correspond?

In the clinical setting, asking the patients about disease occurrence within the family is a routine procedure. The limitations of family history (FH) assessment are well known, such as recall bias, misunderstanding between the patient and the medical doctor, and not knowing much about the diseases for relatives while the size of the families declines. Comparing FH with genetic susceptibility to PRS will set the suitability of one method against the other. Genetic loci for most common diseases are available through GWAS. A population database (FinnGen) of family relationships up to 50 years based on nationwide registries was used to examine how FH interrelates with PRS for twenty-four common diseases in a systematic way (22). Expecting a clear-cut interrelationship between both approaches was somehow disappointing. On average, PRS was met to 10% by first-degree family history. The other way around, first-degree family history was correlated to only 3% of PRS. According to the authors' judgment, both variables reacted more or less independently, while for coronary arteria diseases, glaucoma, and type 2 diabetes, FH was met with a 'considerably elevated PRS risk.' Besides cardiometabolic diseases, breast, and ovarian cancers with the high-risk variants BRCA1 and BRCA2, as well as depression FH and PRS, are moderately related. Generally, PRS is not recommended for prevention measures, given that FH is available. As an exemption, PRS could be used to assess

breast, prostate, and colorectal cancer risk assessment. In cases FH indicates the cancer risk as well, the determination of PRS might add valid information in assuring the severity of the risk. It is also argued that a high PRS for cardiovascular diseases and type 2 diabetes mellitus could benefit preventive treatment and motivate patients to change their lifestyles.

Randomized control trials versus Mendelian randomization

Risk assessment through PRS also could strengthen risk estimation for public health. Another genetic tool may interest epidemiology. Randomized control trials comparing exposed groups under intervention with suitable control groups without intervention. Such trials are hampered by several biases, particularly by confounding. Mendelian randomization (MR) uses exposed alleles with control alleles, while for the two groups, the confounders are equal. This is easily said as done!

Claiming that smoking causes bladder cancer will not be readily accepted by the tobacco industry. Arguments against that hypothesis will come up, pointing towards numerous confounders and biases. A fictional situation pointing towards a confounder could arise in Egypt, where bladder cancer is very common due to chronic infection with Schistosoma haematobium. A smoking Egypt farmer, with his feet frequently in the water, will point towards the parasite and strictly argue against the idea that his smoking will cause cancer in his bladder. The MR example here, however, is taken from an investigation in Spain.

Carcinogens in tobacco smoke include aromatic and heterocyclic amines, which could cause bladder cancer. The amines are detoxified by N-acetyl transferase 2 (NAT2). NAT2 enzyme variation could either slower or faster react in detoxification. In this case, those with the gene variance with lower detoxification will be at a higher risk for bladder cancer when smoking. It was found that the NAT2 slow acetylation genotype has an increased risk for bladder cancer (OR, 95% CI): Overall 1,4 (1.2-1.7), never smokers 0.9 (0.6-1.3) and ever smokers 1.6 (1.3-1.9) (23).

Mendel randomization needs nucleotide polymorphism traits (NPT) for a potential causal relationship (24). These are available in GWAS, but withdrawing the variants needed to reject or allow the hypothesis from the database is a complex task. An appropriate MR-Base platform is available for 2-sample Mendelian randomization (2SMR). From the original base, out of 1673 GWAS, eleven billion NPT are regularly updated. The technique seems suitable for the academic segment of public health experts to adopt the methodology for epidemiological studies (25).

The future in genetics with single-cell sequencing and more

PRS and MR expanded the opportunity to include genetics in applied epidemiology and statistics; there is more to come. ‘Single-cell sequencing (scRNAThe-seq) opens the door to a new area of functional genetics’ (26). This element of the non-coding protein genetics is highly involved in regulatory gene expression on the cell level. eQTL, as mentioned above, measures the distance between RNA variants and gene expression. This requires a high number of cell populations. For example, through the single-cell eQTL mapping, the specific cell type for genetic control of autoimmune diseases is investigated (27), and the regulation of gene

expression of the human tissue (28). For those who believe that the difference between men and females is a social construct, a recent publication in Science might be of interest (29). From forty-four human tissues of 838 adults, 16,245 RNA sequencing samples were taken.

All in all, 13,294 genes are related either to men or females. The function of the genes involves drug and hormone response, fat metabolism, and immune response. Predictably, embryonic development and tissue morphogenesis, fertilization, sexual reproduction, and spermatogenesis had been identified. There are 369 eQTLs across specific tissues 'biased' by sex. Gene regulations in single sex drive '58 gene traits.' Epigenetic marks are highly expressed in females. For instance, eQTL for the hexokinase HKDC1, which is connected to glucose metabolism in pregnancy, influences the child's birth weight.

The European dominance in GWAS

Up to 2020, about 4500 GWAS reported from 4300 papers verified 55,000 loci for almost 5,000 diseases and traits (12). Despite all this enhancement, it is criticized that loci detected in GWAS, related to disease risk increase, are moderate and 'explain only a fraction of the heritability.' On the other hand, there is evidence that 'the strength of association of a GWAS locus is not proportional to its biological importance.

A more serious shortcoming is the overwhelming focus on European ancestry over non-European ancestry, amounting to only 7.4% of Asians and 1% of Africans of all participants (30). This is going to change. Research in genome sequence analysis in many different ethnic cohorts has been identified, and published results by comparing genetic variation between populations so far would justify a separate entry into this blog.

Examples of some drawbacks justify large multi-ethnic biobanks, which will be beneficial for those population groups with functional variants and causal genes not yet included in GWAS. 'Genetic drift and adaptation to the environment over thousands of years are important' (31). A publication about human genetic variation in Europe, the Middle East, Central and South Asia, East Asia, Oceania, and the Americas, the magnitude of variation on continents was compared to Africa, setting it to a hundred percent, and for the archaic admixture to the Oceana, where it is the highest variation in the world. Total genetic variation is relatively low in the Americans compared to the other six continents, and archaic admixture is the lowest worldwide (32). In the past, the Asian population was not included in genetic studies of obesity since the prevalence was low at that time, which has drastically changed (33). In Singapore, a specific SNP connected to T2DM was detected in citizens with an Indian background but not in Chinese and those from the Malay (34). The most recently published results of a Chinese survey for thirty-one complex phenotypes found that, on average, 54.5% of European associates SNPs were significant in East Asia tool, and heterogeneity and allele frequency pattern often did not match the European phenotypes (35).

Viewpoint

How the insufficient concern of population groups in Asia and here in Thailand within the existing databases is relevant for medical genetic services is unclear. In a 2017 publication,

genetic services in clinical medicine, like in other middle-income countries, at that time seemed to be neglectable (36). More than five years later, this might have changed, at least for major university hospitals. A well-known private hospital in Bangkok, caring not only for Thai patients but is well attended by patients from the Middle East and other countries worldwide, started to advertise genetic testing, among other risks for cancer, allergies, cardiovascular diseases, and hyperlipidemia.

So far, as it is known for public health, research and application of omics is neglectable. Throughout the COVID-19 calamity, decision-making was based not on public health expertise, with devastating consequences, as appears by now. As a sensitive issue, this is not discussed in Thailand but is a major controversy in Western countries. In the future, besides immunology, genetic issues for common diseases should be an essential issue of interest and not be overruled by advice from experts in biological science only (37).

*Quotation from (8)

References

1. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, et al. A brief history of human disease genetics. Nature. 2020;577(7789):179-89.
2. Goldberg ML. Genetics. From Genes to Genomes. Seventh Edition ed. New York: Mc Graw Hill; 2021.
3. Greely HT. Public health versus personalized medicineTyranny of the Gene James Tabery Knopf, 2023. 336 pp. Science. 2023;381(6659):738.
4. Hal SS, Birney, E. Journey yo the Genetic Interier. Scientefic American 2012;307(4):10.
5. Doolittle WF. Is junk DNA bunk? A critique of ENCODE. Proc Natl Acad Sci U S A. 2013;110(14):5294-300.
6. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
8. Rood JE, Regev A. The legacy of the Human Genome Project. Science. 2021;373(6562):1442-3.
9. Pennisi E. Massive project reveals complexity of gene regulation. Science. 2020;369(6509):1286-7.
10. Carter R, Drouin G. Structural differentiation of the three eukaryotic RNA polymerases. Genomics. 2009;94(6):388-96.
11. Flynn ED, Lappalainen T. Functional Characterization of Genetic Variant Effects on Expression. Annu Rev Biomed Data Sci. 2022;5:119-39.
12. Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. Nat Commun. 2020;11(1):5900.
13. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-D7.
14. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. Science. 2021;373(6562):1464-8.

15. Visscher PM, Yengo L, Cox NJ, Wray NR. Discovery and implications of polygenicity of common diseases. Science. 2021;373(6562):1468-73.
16. Wray NR, Kemper KE, Hayes BJ, Goddard ME, Visscher PM. Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans: Genomic Prediction. Genetics. 2019;211(4):1131-41.
17. Daniel WW. Biostatistics. Basic concepts and methodology for health science: John Wiley & Sons, Inc.; 2010. 781 p.
18. Porta M. A Dictonary of Epidemiology Sixth Edition ed: Oxford University Press; 2014.
19. Natarajan P. Polygenic Risk Scoring for Coronary Heart Disease: The First Risk Factor. J Am Coll Cardiol. 2018;72(16):1894-7.
20. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet. 2019;28(R2):R133-R42.
21. Becker J, Burik CAP, Goldman G, Wang N, Jayashankar H, Bennett M, et al. Resource profile and user guide of the Polygenic Index Repository. Nat Hum Behav. 2021;5(12):1744-58.
22. Mars N, Lindbohm JV, Della Briotta Parolo P, Widen E, Kaprio J, Palotie A, et al. Systematic comparison of family history and polygenic risk across 24 common diseases. Am J Hum Genet. 2022;109(12):2152-62.
23. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. Lancet. 2005;366(9486):649-59.
24. Smith GD, Ebrahim, S. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Interference in Observational Studies. In: Weinstein M, Vaupel, J.M., Wachter, K.W., editor. Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveyes. Devision of Behaviorral and Social Sciences and Education. Washington, D.C.: The National Academies Press; 2008. p. 336.
25. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.
26. Sumida TS, Hafler DA. Population genetics meets single-cell sequencing. Science. 2022;376(6589):134-5.
27. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science. 2022;376(6589):eabf3041.
28. Kim-Hellmuth S, Aguet F, Oliva M, Munoz-Aguirre M, Kasela S, Wucher V, et al. Cell type-specific genetic regulation of gene expression across human tissues. Science. 2020;369(6509).
29. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, et al. The impact of sex on gene expression across human tissues. Science. 2020;369(6509).
30. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. Nat Genet. 2020;52(3):242-3.
31. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. Nat Rev Genet. 2019;20(9):520-35.
32. Bergstrom A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science. 2020;367(6484).
33. Young KL, Graff M, Fernandez-Rhodes L, North KE. Genetics of Obesity in Diverse Populations. Curr Diab Rep. 2018;18(12):145.

34. Sim X, Ong RT, Suo C, Tay WT, Liu J, Ng DP, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. PLoS Genet. 2011;7(4):e1001363.
35. Qiao J, Wu Y, Zhang S, Xu Y, Zhang J, Zeng P, et al. Evaluating significance of European-associated index SNPs in the East Asian population for 31 complex phenotypes. BMC Genomics. 2023;24(1):324.
36. Limwongse C. Medical genetic services in a developing country: lesson from Thailand. Curr Opin Pediatr. 2017;29(6):6.
37. Gostin LO, Wetter S. Fix the backlash against public health. Science. 2023;379(6639):1277.

Frank P. Schelp is responsible for the manuscript's content, and the points of view expressed might not reflect the stance and policy of the Faculty of Public Health, Khon Kaen University, Thailand.

For comments and questions, please contact <awuso11@gmail.com>