

## REVIEW

# The indigenous populations as the model by nature to understand human genomic-phenomics interactions

Boon-Peng Hoh<sup>1,\*</sup>, Thuhairah Abdul Rahman<sup>2</sup>

<sup>1</sup> Faculty of Medicine and Health Sciences, UCSI University, Kuala Lumpur 56000, Malaysia

<sup>2</sup> Clinical Pathology Diagnostic Centre Research Laboratory, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh Campus, Sungai Buloh 47000, Malaysia

\* Correspondence: hoh.boonpeng@gmail.com

Received October 7, 2020; Revised December 23, 2020; Accepted January 26, 2021

**Background:** The advancement of genomics has progressed in lightning speed over the past two decades. Numerous large-scale genome sequencing initiatives were announced, along with the rise of the holistic precision medicine approach. However, the field of genomic medicine has now come to a bottleneck since genomic-phenomic interactions are not fully comprehended due to the complexity of the human systems biology and environmental influence, hence the emergence of human phenomics.

**Results:** This review attempts to provide an overview of the potential advantages of investigating the human phenomics of indigenous populations and the challenges ahead.

**Conclusion:** We believe that the indigenous populations serve as an ideal model to excavate our understanding of genomic-environmental-phenomics interactions.

**Keywords:** indigenous populations; Orang Asli; genomics; phenomics

**Author summary:** The advancement of genomic technology has progressed in a lightning speed. However, the understanding of genotype-phenotype interactions has come to a bottleneck owing to the complex interplays between the human biology and environment, hence posing hindrance to materialising precision medicine in a holistic manner. The newly emerging discipline on human phenomics may be the solution. We argue that the indigenous populations serve as an ideal model to excavate our understanding on genomic-environmental-phenomics interactions. This review provides an overview on the potential advantages of investigating the human phenomics in the indigenous populations, and the challenges ahead.

## INTRODUCTION

The ultimate goal of modern human biology is to completely understand how genetic make-up results in specific phenotypic properties, namely how human bodies function, and how to prevent deterioration. Although we often assume that the human genome is fixed, there are individual differences that drive phenotypic variations, such as skin pigmentation, weight, height, blood pressure. A large portion of these traits can be regulated by environmental changes, for instance diet, climate, sun-light exposure, microbial exposure. Such diverse interactive sources have led to an unprecedented collection of physical, environmental, biological, and medical threats

faced by human beings.

The past two decades have witnessed a revolutionary advancement in the field of human genomics. Following the announcement on the completion of the draft of human genome sequence in 2000 [1,2], genomic technologies, along with the analytical and digital revolution, have progressed at a tremendous pace – the inauguration of the International HapMap Project [3–5], followed by the launching of whole genome genotyping arrays, and subsequently, the genome-wide association studies (GWAS) on numerous Mendelian and complex diseases [6]. In addition, there is the blooming of the next-generation sequencing technologies [7–9] and the commencement of the 1000 Genomes Project [10,11]. Today,

we witness the completion of various ‘mega-scaled’ whole genome sequencing initiatives that involve thousands to tens of thousands of samples – the UK10K Project [12]; the SG10K Study [13]; the Genome Asia 100K [14]; the ChinaMap initiative [15]; Korean Genome Project [16]; the All of Us Research Programme; and the recently-launched Genome Sequencing of 1000 Indians, to name a few.

However, scientists realize that the power of genomics has come to a bottleneck, and that the understanding of genotype-phenotype interactions has encountered challenges due to the complexity of the biological systems that interplay with multiple environmental factors [17]. Although GWAS has shifted the paradigm for genetic investigations and the understanding of genetic architectures for many complex traits, the majority cannot be reliably replicated, and our understanding between the genotype-phenotype relationships remains scarce [18–20]. To overcome this challenge, comprehensive catalogue of genetic variations for both mainstream and marginalized populations [21] must couple with multiple dimensional phenotypes and considerations of harmonized environmental influence hence the emergence of human phenomics. The marginalized populations, such as the indigenous populations, referring to the populations residing within geographically distinct traditional habitats, ancestral territories, and relying on natural resources in these habitats and territories for survival, may be an ideal model by nature to comprehend human genomic-phenomics interactions. They maintain their unique cultural identities, as well as social, economic, cultural, and political practices separate from mainstream societies and norms. The indigenous populations usually have long population history and are genetically more homogenous compared to the mainstream populations. In addition, these communities are minimally exposed to modernization, genetically homogenous, and reside within a relatively ‘controllable’ habitat, which collectively ease the measurement of the phenotypic changes. This review attempts to provide an overview on the potential use of indigenous populations research in cataloguing the human phenome, and the challenges ahead.

## LESSONS LEARNT FROM GWAS

Since its first publication in 2002 [22], genome-wide association studies (GWAS) have revealed well-supported association and successfully mapped thousands of novel variants associated with complex traits, leading to discovery of new disease physiology (<https://www.ebi.ac.uk/gwas/>).

The underlying principle of GWAS is the ‘common disease-common variants’ hypothesis, which assumes that the major contributors to the genetic susceptibility of common diseases are the common genetic variants

amongst a population. These variants often have lower penetrance and small additive genetic effect, which explains why the vast majority of these loci generally explains only a small proportion of the phenotypic variance with modest effect size. In the case of human height, a total of 180 stringently validated loci collectively explains only ~10% of the genetic variation [23]. Thus, the problem of ‘missing heritability’ remains arguable [19]. Some speculate that ‘missing heritability’ could be due to rarer variants with high penetrance that contribute to disease susceptibility [19]. Future measures include applying stringent statistical tests to minimize false-positive findings [24], or simply increasing the sample size and density of genotypes [25–28].

A striking example is hypertension. During its earlier stages, GWAS failed to identify the genes responsible for the genetic susceptibility of hypertension [6,29]. Subsequently, a large-scale meta-GWAS was performed, which identified several hundreds of associated loci. Yet, they only explain ~3%–5% of hypertension heritability [30], and many of the association signals are not replicable. This is plausibly due to the fact that the etiology of hypertension is primarily classified into different intermediate phenotypes based on its physiological mechanisms, which potentially dilutes the statistical effects. Therefore, we believe that one solution to address genotype-phenotype interactions is via deep phenotyping.

## WHAT IS THE HUMAN PHENOME?

Phenotype refers to a set of properties or characteristics that arise from the interaction between an individual’s genetic make-up and the environment, while phenome refers to the complete set of all human characteristics, determined by complex interactions between genes and the environment. The human phenome includes a comprehensive collection of all phenotypes, ranging from macro- to micro-scales, from external appearance to internal mechanisms, from biochemical characteristics to microbiota and psychological behavior, from population to individual levels, and from system to tissue and cellular characteristics [31].

Human phenomics integrates and models multiple ‘-omics’ parameters together with other biological metadata and their impact on disease risk at individual and population levels. It is the key driver towards elucidating the ultimate gene-environment interactions that underpin the differential risks, prevalence, and emergence of disease phenotypes.

## HUMAN PHENOME AND PRECISION MEDICINE

Precision medicine is a holistic idea that customizes the healthcare approach of an individual patient or a

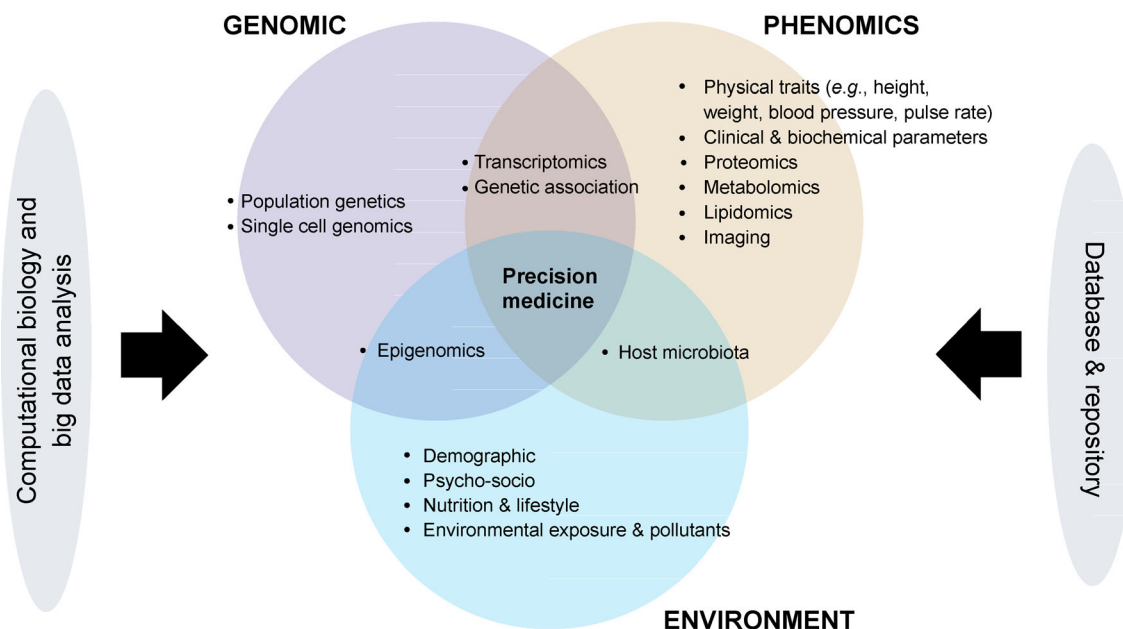
particular group of patients [32]. Although genetic make-up is no doubt the key element to the implementation of precision medicine, individual history, environment, lifestyle, as well as heterogeneous disease phenotypes and manifestations must be taken into consideration. Therefore, in addition to physical and biochemical parameters, numerous ‘-omics’ technologies including transcriptomics, proteomics, metabolomics, lipidomics, metagenomics, epigenomics, and microbiomes, have been proposed as an alternative solution to overcome challenges for precision medicine [33]. In other words, the prerequisite to warrant the success of precision medicine is comprehensive cataloguing the cell-, individual- and population-based phenomes (Fig. 1). Precision medicine in hypertension is a typical model that reflects the importance of phenomics [34].

Hypertension is a classic model to exemplify the importance of phenomics in precision medicine. Hypertension, defined as the persistent elevation in blood pressure (BP) greater than 140/90 mmHg, affects 1.2 billion people worldwide [35]. It has been increasingly recognized as a syndrome, rather than a disease [36]. In addition, only ~30% of hypertensive individuals have their blood pressure well-controlled, despite being prescribed various anti-hypertensive medications [37].

Hypertension is categorized into primary (or essential) and secondary hypertension. Approximately 90% of hypertension appears with no identified cause, hence

defined as primary hypertension, while the rest are due to endocrine or renal pathologies, hence termed as secondary hypertension. Primary hypertension is further categorized into intermediate phenotypes, including salt-resistant (~40% of the primary hypertension) and salt-sensitive, which can be further sub-classified into low-renin (~55% of the salt-sensitive) and normal-renin levels (non-modulation) [36]. Other hypertension sub-intermediate phenotypes being suggested including obesity-related hypertension [38] and deoxycorticosterone acetate (DOCA) salt hypertension, which exhibits salt-dependent excess mineralocorticoid [39,40]. What aggravates this complication is the fact that numerous factors, including age, sex, circadian rhythm, individual lifestyle, and physical activity. These factors have been proposed to further explain the increased risk of hypertension and individual differential response to anti-hypertensive medication. In addition, variation in average blood pressure levels have been shown in different populations [41], and that hypertension susceptibility is correlated with geographical latitude [42–44].

Collectively, these findings suggest that the realization of precision medicine for hypertension relies on comprehensive phenotyping and the understanding of the complex, dynamic interactions between “genotype-environment-phenotypes” [34,36]. Essentially, the holistic approach of human phenomics to the implications of precision medicine should be reflected by Fig. 1.



**Figure 1. Holistic approach to precision medicine.** An ideal model for precision medicine should consider not only individual genomic information, but also the totality of phenotypical and environmental influences. Essentially, an established analysis pipeline must be developed to handle the generated metadata. Ultimately, a database or a data repository would be able to systematically manage the outcome of the investigations and serve as a future reference.

## INDIGENOUS POPULATIONS AS THE MODEL POPULATION FOR PHENOMIC STUDY

The idea of phenomics was first proposed by Nelson Freimer and Chiara Sabatti in 2003 [31]. However, owing to the limitation of technology, the idea was initially unsuccessful. Subsequently, a handful of phenome-wide association studies (PheWAS) were published, which principally examined phenotypic data derived from electronic health records across large population cohorts and associated them with genome-wide genotyping data [45–48], while others selected a list of chosen phenotypes [49,50]. A handful of single-nucleotide polymorphisms (SNPs) with pleiotropic effects were identified from these studies, further advancing our understanding of genotype-phenotype interactions.

Phenome research has progressed into utilizing animal models with several recently-initiated projects [51,52]. Although utilizing animal models simplifies the analytical framework, the practical limitation of performing bench studies in a laboratory setting has often restricted findings and does not reflect the larger picture of systems biology of human traits [53]. For instance, non-human primates can sustain viral replication in relevant cell types and develop a robust immune response without manifesting the disease itself. A similar phenomenon can be observed in the murine model, where the dengue virus generally will not cause pathology in wild type mice.

As Sydney Brenner rightfully quoted, *“We don’t have to search for a model organism anymore. Because we are the model organisms”* – the indigenous populations serve as a natural model to deepen our understanding of genomic-environmental-phenomics interactions.

Take the indigenous populations of Peninsular Malaysia (locally known as Orang Asli) as an example. These populations are classified into three major tribes, namely the Negrito, Senoi, and Proto Malay, each of which is further categorized into six sub-tribes. Archaeological and population genetic studies suggest the possibility of these populations inhabiting the Southeast Asia region more than 50,000 years ago [54,55]. Although genome sequencing of these populations has been reported, such research has been rather modest [54,56] compared to many genome sequencing initiatives worldwide.

Traditionally, the Orang Asli inhabit remote areas neighboring the tropical rainforest. Because of this, the majority of these tribes were nomadic hunter-gatherers or swiddening agriculturalists that practiced egalitarianism. Specifically, the Proto Malays were more settled down and demonstrated advanced farming practices. Due to the nature of the tropical rainforest, the Orang Asli were exposed to numerous environmental stress stimuli,

including poor personal hygiene, parasites and pathogens, as well as a lack of dietary nutrition. However, in the last several decades, certain Orang Asli communities have been relocated to semi-urban areas due to unavoidable modernization and the government’s initiative to alleviate poverty [57]. Consequently, there has been an increase in the prevalence of metabolic syndromes within the Orang Asli, especially among those being resettled near urban areas. Yet, the trajectory of the reported incident rates is still slower than that of the mainstream populations [58–60].

Given such ‘natural’ exposure to the environmental stressors of the tropical rainforest and a fairly ‘controllable’ lifestyle and diet, attribution to any observed phenotypic variation could be narrowed down to the variability of intrinsic factors, thus comprehending the resiliency of these populations would be biological meaningful.

## PHENOTYPIC STUDIES OF INDIGENOUS POPULATIONS

Genomic and phenomic studies of indigenous populations worldwide have been largely underrepresented. Until 2019, indigenous populations only accounted ~0.02% of all GWAS conducted [61,62]. Indeed, most global large-scale genomic research studies lack representation of indigenous populations (to name a few, the HapMap Project; 1,000 Genomes Project; Genome Aggregation Database (gnomAD); Simons Genome Diversity Project). Unfortunately, only a handful of studies reported their respective genomic structures and population histories [56,63–70]. Studies that correlate their genomic structure and phenotypes have been primarily focusing on identifying signatures of positive selection [71–74]. Research related to phenome-wide association studies (PheWAS) or phenomics of indigenous populations have been scarce, as evidenced by a PubMed search (on 18th Dec 2020) that revealed no publications related to this topic.

Phenotypic research carried out on the indigenous populations were largely related to disease traits, noncommunicable diseases have been of interest in recent years [58–60,72,75–77]. Numerous studies reported high prevalence rates of infectious diseases (*e.g.*, plasmodium, soil transmitted helminths, viral infections) among indigenous populations [78–83], yet many did not demonstrate symptoms of severe manifestations. Intriguingly, some indigenous populations exhibit unique phenotypic characteristics that cannot be explained merely by environmental exposure. A striking example is the Negrito populations of Peninsular Malaysia who predominantly inhabit the remote areas of Northern

Peninsular Malaysia and demonstrated an average blood pressure higher than other populations, despite minimally attributed to hypertension risk factors [84]. In addition, studies conducted on inflammatory and endothelial activation biomarker levels suggest that genetic factors may be, in part, a plausible explanation of lower cardiovascular disease risk among the Negrito populations compared to other populations [58]. In contrast, studies conducted on hypertension and cardiovascular diseases in mainstream populations would be challenging since these populations are exposed to various risk factors (*e.g.*, sedentary lifestyle, stress stimuli, excess salt intake, poor diet, pollution, poor health awareness) and eliminating these factors would pose a great challenge.

Another example is the Greenlandic Inuit population who live in extreme arctic climates. By consuming a low-carbohydrate diet, they rely primarily on fatty acids and ketone bodies as their main source of energy. The Inuits have lower levels of cardiometabolic risk factors for the same level of body mass index (BMI) or waist circumference (WC) compared to Europeans [85], but have shown a significant increase in type 2 diabetes mellitus over the last 30 years [86].

## CHALLENGES AHEAD

While investigating the phenomics of indigenous populations sounds promising, several challenges and trade-offs in contrast to mainstream populations are acknowledged as followed:

(i) Large-scale genome sequencing initiatives on diverse indigenous populations, in particular those within Southeast Asia, are essential. This is because Southeast Asia houses approximately 70% of the global human genetic diversity. Indeed, the fact that the 1000 Genomes Project data does not sufficiently cover the global human genetic diversity [87] highlights the need for global collaborative initiatives to complete the catalogue of human genetic variations.

(ii) Comprehensive and accurate phenotyping is crucial to the improvement and advancement of modern precision medicine. However, it is practically difficult to collect complete sets of phenotypic data and biological samples (*e.g.*, blood, urine, saliva, tissue, imaging data) from the same individual. Such data fragmentation poses challenges to downstream analysis [88]. Substantial efforts are required to engage with indigenous communities, establish trust and mutual respect, in order to ease the process of sample and metadata collection.

(iii) The majority of indigenous populations have a small population size, which consequently means that genetic variations would be too small to (a) detect

mutations with large effects; and (b) explain variants commonly observed in other populations. Of particular importance, the statistical power of attributing the phenotypic variability to genetics or environmental exposures may be restricted [53]. This is even more challenging when the phenotypic data collection is prospective, whereby follow-ups are required to measure changes or trends. Alternative analysis pipelines are required to overcome these restraints, without compromising the biological effects.

(iv) Extending from (iii), the issue of replicability within indigenous populations may arise due to a lower statistical power resulting from a small sample size. In addition, the power of study may be confounded by a high inbreeding rate in the indigenous populations. However, considering the relatively homogenous genotype and environmental exposure, a proper study design may be able to mitigate the potential error, for example, selecting the top percentile versus the bottom percentile of a quantitative trait from an indigenous cohort.

(v) The revolutionary change of the field of life sciences and materialization of precision medicine (Fig. 1) will eventually face a ‘data tsunami’. Scientists will be overwhelmed by huge, diversified, and complicated metadata. Therefore, harmonizing protocols and analysis pipelines that integrate such diversified data are crucial. However, this approach will undoubtedly lead to drastically higher costs for experimental procedures, infrastructure establishment, and analysis pipeline. Consequently, this may not be affordable to many developing countries, especially those in Southeast Asia. To this end, establishing an integrated genomic-phenomic database for indigenous populations would be meaningful [89] to systematically organize the resources for future reference.

In contrast, despite heterogeneous genetic make-up and environmental exposure, there are several pros to studying the mainstream populations. Due to their large population sizes:

(i) Collection of genotypic and phenotypic data is relatively less burdensome. In addition, the availability of reference panels (*e.g.*, 1000 Genomes Project and the GnomAD database) allows for more accurate genotype imputation compared to indigenous populations that lack a representative reference panel. In addition, there is a lower possibility of inbreeding within mainstream populations as opposed to indigenous populations.

(ii) The statistical power to detect mutations with greater genetic effects and genetic variation responsible for phenotypic variability would be higher. In addition, the issue of replicability could be mitigated.

For the ease of comparison, the pros and cons of researching indigenous and mainstream populations are listed in Table 1.

**Table 1 Comparison of pros and cons in conducting phenomic research between indigenous populations and mainstream populations**

	Indigenous	Mainstream
Environment exposure	Homogenous; minimally influenced by urbanization	Heterogenous; strongly influenced by urbanization
Genotypic and phenotypic heterogeneity	Isolated; genetically and phenotypically homogenous	Genetically and phenotypically heterogeneous; may be influenced by admixture
Population size	Small Higher inbreeding rate, potentially affect statistical analysis for genotype-phenotype correlation	Large Lower possibility of inbreeding
Statistical power	May be affected by a small population size	Higher statistical power
Replicability	May be affected by a small sample size	Higher likelihood with sufficient power of study

## CONCLUDING THOUGHTS

The completion of the human genome sequence in 2000 had brought a world-shattering shift of paradigm to genome medicine research. Since then, significant technological advancements have drastically reduced the cost of genome sequencing within a relatively short period of time, increasing both its output quality and quantity. The establishment of the human genome-phenome reference is anticipated to provide a more complete understanding of the origins and diversity of various human traits and diseases, ultimately, to create a new translational medicine paradigm for optimizing and improving healthcare and patient management. Provided that the aforementioned challenges can be overcome, the indigenous populations serve as an ideal model population to excavate our understanding on genomic-environmental-phenomics interactions.

## ACKNOWLEDGEMENTS

The author acknowledges the Department of Orang Asli Development (JKOA) of Malaysia and the Orang Asli communities for their participation and engagement in earlier studies.

## COMPLIANCE WITH ETHICS GUIDELINES

The authors Boon-Peng Hoh and Thuhairah Abdul Rahman declare that they have no conflict of interests.

This article is a review article and does not contain any studies with human or animal subjects performed by any of the authors.

## OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's

Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., *et al.* (2001) The sequence of the human genome. *Science*, 291, 1304–1351
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921
- International HapMap Consortium. (2003) The International HapMap Project. *Nature*, 426, 789–796
- Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449, 851–861
- The International HapMap 3 Consortium. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, 467, 52–58
- Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447, 661–678
- Fujimoto, A., Nakagawa, H., Hosono, N., Nakano, K., Abe, T., Boroevich, K. A., Nagasaki, M., Yamaguchi, R., Shibuya, T., Kubo, M., *et al.* (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat. Genet.*, 42, 931–936
- Ahn, S. M., Kim, T. H., Lee, S., Kim, D., Ghang, H., Kim, D. S., Kim, B. C., Kim, S. Y., Kim, W. Y., Kim, C., *et al.* (2009) The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.*, 19, 1622–1629
- Kim, J. I., Ju, Y. S., Park, H., Kim, S., Lee, S., Yi, J. H., Mudge, J., Miller, N. A., Hong, D., Bell, C. J., *et al.* (2009) A highly annotated whole-genome sequence of a Korean individual. *Nature*, 460,

- 1011–1015
10. The 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature*, 467, 1061–1073
  11. The 1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. *Nature*, 526, 68–74
  12. The UK10K Consortium. (2015) The UK10K project identifies rare variants in health and disease. *Nature*, 526, 82–89
  13. Wu, D., Dou, J., Chai, X., Bellis, C., Wilm, A., Shih, C. C., Soon, W. W. J., Bertin, N., Lin, C. B., Khor, C. C., *et al.* (2019) Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. *Cell*, 179, 736–749.e15
  14. The GenomeAsia 100K Consortium. (2019) The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, 576, 106–111
  15. Cao, Y., Li, L., Xu, M., Feng, Z., Sun, X., Lu, J., Xu, Y., Du, P., Wang, T., Hu, R., *et al.* (2020) The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals. *Cell Res.*, 30, 717–731
  16. Jeon, S., Bhak, Y., Choi, Y., Jeon, Y., Kim, S., Jang, J., Jang, J., Blazyte, A., Kim, C., Kim, Y., *et al.* (2020) Korean Genome Project: 1094 Korean personal genomes with clinical information. *Sci. Adv.*, 6, eaaz7835
  17. Houle, D., Govindaraju, D. R. and Omholt, S. (2010) Phenomics: the next challenge. *Nat. Rev. Genet.*, 11, 855–866
  18. Cano-Gamez, E. and Trynka, G. (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.*, 11, 424
  19. Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747–753
  20. Altshuler, D., Daly, M. J. and Lander, E. S. (2008) Genetic mapping in human disease. *Science*, 322, 881–888
  21. Bustamante, C. D., De La Vega, F. M. and Burchard, E. G. (2011) Genomics for the world. *Nature*, 475, 163–165
  22. Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, H., Sato, H., Hori, M., Nakamura, Y., *et al.* (2002) Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.*, 32, 650–654
  23. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., Jackson, A. U., Vedantam, S., Raychaudhuri, S., *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467, 832–838
  24. Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, 42, 565–569
  25. Nikpay, M., Goel, A., Won, H.-H., Hall, L. M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C. P., Hopewell, J. C., *et al.* (2015) A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.*, 47, 1121–1130
  26. Doherty, A., Smith-Byrne, K., Ferreira, T., Holmes, M. V., Holmes, C., Pulit, S. L. and Lindgren, C. M. (2018) GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.*, 9, 5257
  27. Band, G., Le, Q. S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., *et al.* (2013) Imputation-based meta-analysis of severe malaria in three African populations. *PLoS Genet.*, 9, e1003509
  28. Klarin, D., Lynch, J., Aragam, K., Chaffin, M., Assimes, T. L., Huang, J., Lee, K. M., Shao, Q., Huffman, J. E., Natarajan, P., *et al.* (2019) Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat. Med.*, 25, 1274–1279
  29. Zhang, K., Weder, A. B., Eskin, E. and O'Connor, D. T. (2010) Genome-wide case/control studies in hypertension: only the 'tip of the iceberg'. *J. Hypertens.*, 28, 1115–1123
  30. Wang, Y. and Wang, J.-G. (2019) Genome-wide association studies of hypertension and several other cardiovascular diseases. *Pulse (Basel)*, 6, 169–186
  31. Freimer, N. and Sabatti, C. (2003) The human phenome project. *Nat. Genet.*, 34, 15–21
  32. Dominiczak, A., Delles, C. and Padmanabhan, S. (2017) Genomics and precision medicine for clinicians and scientists in hypertension. *Hypertension*, 69, e10–e13
  33. Peck, R. W. (2018) Precision medicine is not just genomics: the right dose for every patient. *Annu. Rev. Pharmacol. Toxicol.*, 58, 105–122
  34. Hoh, B. P., Abdul Rahman, T. and Yusoff, K. (2019) Natural selection and local adaptation of blood pressure regulation and their perspectives on precision medicine in hypertension. *Hereditas*, 156, 1
  35. Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., Chen, J. and He, J. (2016) Global disparities of hypertension prevalence and control: A systematic analysis of population-based studies from 90 countries. *Circulation*, 134, 441–450
  36. Manosroi, W. and Williams, G. H. (2019) Genetics of Human Primary Hypertension: Focus on Hormonal Mechanisms. *Endocr. Rev.*, 40, 825–856
  37. Tocci, G., Ferrucci, A., Pontremoli, R., Ferri, C., Rosei, E. A., Morganti, A., Trimarco, B., Mancina, G., Borghi, C. and Volpe, M. (2015) Blood pressure levels and control in Italy: comprehensive analysis of clinical data from 2000–2005 and 2005–2011 hypertension surveys. *J. Hum. Hypertens.*, 29, 696–701
  38. Korner, P. I. (2010) The phenotypic patterns of essential hypertension are the key to identifying “high blood pressure” genes. *Physiol. Res.*, 59, 841–857
  39. Basting, T. and Lazartigues, E. (2017) DOCA-salt hypertension: an update. *Curr Hypertens Rep.*, 19, 1–8
  40. Coffman, T. M. (2011) Under pressure: the search for the essential mechanisms of hypertension. *Nat. Med.*, 17, 1402–1409
  41. Kato, N. (2012) Ethnic differences in genetic predisposition to hypertension. *Hypertens. Res.*, 35, 574–581

42. Thompson, E. E., Kuttub-Boulos, H., Witonsky, D., Yang, L., Roe, B. A. and Di Rienzo, A. (2004) CYP3A variation and the evolution of salt-sensitivity variants. *Am. J. Hum. Genet.*, 75, 1059–1069
43. Nakajima, T., Wooding, S., Sakagami, T., Emi, M., Tokunaga, K., Tamiya, G., Ishigami, T., Umemura, S., Munkhbat, B., Jin, F., *et al.* (2004) Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *Am. J. Hum. Genet.*, 74, 898–916
44. Young, J.H., Chang, Y.P.C., Kim, J.D.O., Chretien, J.P., Klag, M.J., Levine, M.A., Ruff, C.B., Wang, N.Y., Chakravarti, A. (2005) Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet.*, 1, e82
45. Verma, A., Lucas, A., Verma, S. S., Zhang, Y., Josyula, N., Khan, A., Hartzel, D. N., Lavage, D. R., Leader, J., Ritchie, M. D., *et al.* (2018) PheWAS and beyond: the landscape of associations with medical diagnoses and clinical measures across 38,662 individuals from geisinger. *Am. J. Hum. Genet.*, 102, 592–608
46. Safarova, M. S., Satterfield, B. A., Fan, X., Austin, E. E., Ye, Z., Bastarache, L., Zheng, N., Ritchie, M. D., Borthwick, K. M., Williams, M. S., *et al.* (2019) A phenome-wide association study to discover pleiotropic effects of *PCSK9*, *APOB*, and *LDLR*. *NPJ Genom. Med.*, 4, 3
47. Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., Field, J. R., Pulley, J. M., Ramirez, A. H., Bowton, E., *et al.* (2013) Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.*, 31, 1102–1111
48. Diogo, D., Tian, C., Franklin, C. S., Alanne-Kinnunen, M., March, M., Spencer, C. C. A., Vangjeli, C., Weale, M. E., Mattsson, H., Kilpeläinen, E., *et al.* (2018) Phenome-wide association studies across large population cohorts support drug target validation. *Nat. Commun.*, 9, 4285
49. Zhao, X., Geng, X., Srinivasasainagendra, V., Chaudhary, N., Judd, S., Wadley, V., Gutiérrez, O. M., Wang, H., Lange, E. M., Lange, L. A., *et al.* (2019) A PheWAS study of a large observational epidemiological cohort of African Americans from the REGARDS study. *BMC Med. Genomics*, 12, 26
50. Pendergrass, S. A., Buyske, S., Jeff, J. M., Frase, A., Dudek, S., Bradford, Y., Ambite, J. L., Avery, C. L., Buzkova, P., Deelman, E., *et al.* (2019) A phenome-wide association study (PheWAS) in the Population Architecture using Genomics and Epidemiology (PAGE) study reveals potential pleiotropy in African Americans. *PLoS One*, 14, e0226771
51. Mashimo, T., Voigt, B., Kuramoto, T. and Serikawa, T. (2005) Rat Phenome Project: the untapped potential of existing rat strains. *J. Appl. Physiol.*, 98, 371–379
52. Bogue, M. A., Grubb, S. C., Walton, D. O., Philip, V. M., Kolishovski, G., Stearns, T., Dunn, M. H., Skelly, D. A., Kadakuzha, B., TeHennepe, G., *et al.* (2018) Mouse Phenome Database: an integrative database and analysis suite for curated empirical phenotype data from laboratory mice. *Nucleic Acids Res.*, 46, D843–D850
53. FitzGerald, G., Botstein, D., Califf, R., Collins, R., Peters, K., Van Bruggen, N. and Rader, D. (2018) The future of humans as model organisms. *Science*, 361, 552–553
54. Yew, C. W., Lu, D., Deng, L., Wong, L. P., Ong, R. T., Lu, Y., Wang, X., Yunus, Y., Aghakhanian, F., Mokhtar, S. S., *et al.* (2018) Genomic structure of the native inhabitants of Peninsular Malaysia and North Borneo suggests complex human population history in Southeast Asia. *Hum. Genet.*, 137, 161–173
55. Hill, C., Soares, P., Mormina, M., Macaulay, V., Meehan, W., Blackburn, J., Clarke, D., Raja, J. M., Ismail, P., Bulbeck, D., *et al.* (2006) Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol. Biol. Evol.*, 23, 2480–2491
56. Deng, L., Lou, H., Zhang, X., Thiruvahindrapuram, B., Lu, D., Marshall, C. R., Liu, C., Xie, B., Xu, W., Wong, L. P., *et al.* (2019) Analysis of five deep-sequenced trio-genomes of the Peninsular Malaysia Orang Asli and North Borneo populations. *BMC Genomics*, 20, 842
57. Endicott, K. (2016) Malaysia's Original People: Past, Present and Future of the Orang Asli. Singapore: NUS Press
58. Mokhsin, A., Mokhtar, S. S., Mohd Ismail, A., M Nor, F., Shaari, S. A., Nawawi, H., Yusoff, K., Abdul Rahman, T. and Hoh, B. P. (2018) Observational study of the status of coronary risk biomarkers among Negritos with metabolic syndrome in the east coast of Malaysia. *BMJ Open*, 8, e021580
59. Aghakhanian, F., Wong, C., Tan, J. S. Y., Yeo, L. F., Ramadas, A., Edo, J., Hoh, B. P., Khalid, B. A. K. and Phipps, M. E. (2019) Metabolic syndrome and cardiometabolic risk factors among indigenous Malaysians. *Public Health*, 176, 106–113
60. Tuan Abdul Aziz, T. A., Teh, L. K., Md Idris, M. H., Bannur, Z., Ashari, L. S., Ismail, A. I., Ahmad, A., Isa, K. M., Nor, F. M., Rahman, T. H., *et al.* (2016) Increased risks of cardiovascular diseases and insulin resistance among the Orang Asli in Peninsular Malaysia. *BMC Public Health*, 16, 284
61. Mills, M. C. and Rahal, C. (2019) A scientometric review of genome-wide association studies. *Comm Biol.*, doi:10.1038/s42003-018-0261-x
62. Garrison, N. A., Hudson, M., Ballantyne, L. L., Garba, I., Martinez, A., Tauaii, M., Arbour, L., Caron, N. R. and Rainie, S. C. (2019) Genomic research through an indigenous lens: understanding the expectations. *Annu. Rev. Genomics Hum. Genet.*, 20, 495–517
63. Jinam, T. A., Phipps, M. E., Aghakhanian, F., Majumder, P. P., Datar, F., Stoneking, M., Sawai, H., Nishida, N., Tokunaga, K., Kawamura, S., *et al.* (2017) Discerning the origins of the negritos, first sundaland people: deep divergence and archaic admixture. *Genome Biol. Evol.*, 9, 2013–2022
64. Scholes, C., Siddle, K., Ducourneau, A., Crivellaro, F., Järve, M., Rootsi, S., Bellatti, M., Tabbada, K., Mormina, M., Reidla, M., *et al.* (2011) Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. *Am. J. Phys. Anthropol.*, 146, 62–72
65. Kusuma, P., Brucato, N., Cox, M. P., Letellier, T., Manan, A., Nuraini, C., Grangé, P., Sudoyo, H. and Ricaut, F. X. (2017) The last sea nomads of the Indonesian archipelago: genomic origins and dispersal. *Eur. J. Hum. Genet.*, 25, 1004–1010



66. Deng, L., Hoh, B. P., Lu, D., Fu, R., Phipps, M. E., Li, S., Nur-Shafawati, A. R., Hatin, W. I., Ismail, E., Mokhtar, S. S., *et al.* (2014) The population genomic landscape of human genetic structure, admixture history and local adaptation in Peninsular Malaysia. *Hum. Genet.*, 133, 1169–1185
67. Aghakhanian, F., Yunus, Y., Naidu, R., Jinam, T., Manica, A., Hoh, B. P. and Phipps, M. E. (2015) Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol. Evol.*, 7, 1206–1215
68. McEvoy, B. P., Lind, J. M., Wang, E. T., Moyzis, R. K., Visscher, P. M., van Holst Pellekaan, S. M. and Wilton, A. N. (2010) Whole-genome genetic diversity in a sample of Australians with deep Aboriginal ancestry. *Am. J. Hum. Genet.*, 87, 297–305
69. Tokunaga, K., Ohashi, J., Bannai, M. and Juji, T. (2001) Genetic link between Asians and native Americans: evidence from HLA genes and haplotypes. *Hum. Immunol.*, 62, 1001–1008
70. Zhou, S., Xie, P., Quobion, A., Ambalavanan, A., Dionne-Laporte, A., Spiegelman, D., Bourassa, C. V., Xiong, L., Dion, P. A. and Rouleau, G. A. (2019) Genetic architecture and adaptations of Nunavik Inuit. *Proc. Natl. Acad. Sci. USA*, 116, 16012–16017
71. Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jørgensen, M. E., Korneliussen, T. S., Gerbault, P., Skotte, L., Linneberg, A., *et al.* (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, 349, 1343–1347
72. Liu, X., Yunus, Y., Lu, D., Aghakhanian, F., Saw, W. Y., Deng, L., Ali, M., Wang, X., Nor, F. M., Rahman, T. A., *et al.* (2015) Differential positive selection of malaria resistance genes in three indigenous populations of Peninsular Malaysia. *Hum. Genet.*, 134, 375–392
73. Ilardo, M. A., Moltke, I., Korneliussen, T. S., Cheng, J., Stern, A. J., Racimo, F., de Barros Damgaard, P., Sikora, M., Seguin-Orlando, A., Rasmussen, S., *et al.* (2018) Physiological and genetic adaptations to diving in sea nomads. *Cell*, 173, 569–580.e15
74. Kapoor, A. K. and Kaur, J. (2012) Natural selection in a population group of Andaman and Nicobar Islands. *J. Nat. Sci. Biol. Med.*, 3, 71–77
75. Roberts-Thomson, R. A. and Roberts-Thomson, P. J. (1999) Rheumatic disease and the Australian aborigine. *Ann. Rheum. Dis.*, 58, 266–270
76. Leung, L. (2016) Diabetes mellitus and the Aboriginal diabetic initiative in Canada: An update review. *J. Family Med. Prim. Care*, 5, 259–265
77. Wong, E. C. L. and Kapoor, A. (2017) Epidemiology of prostate and kidney cancer in the Aboriginal population of Canada: A systematic review. *Can. Urol. Assoc. J.*, 11, E222–E232
78. Thomas, V., Sng, K. H. and Yap, P. L. (1980) Seroepidemiology of Malaria in Orang Asli (Aborigines) Children in Kelantan, Malaysia. Geneva, available from: <http://apps.who.int/iris/handle/10665/65793>
79. Baer, A. S. (1999) Health, disease, and survival : a biomedical and genetic analysis of the Orang Asli of Malaysia. 1st ed. Subang Jaya: Center for Orang Asli Concerns
80. Kaur, G. (2009) Prevalence of clinical malaria among an Orang Asli community in Malaysia. *Southeast Asian J Trop Med Public Health*, 40, 665–673
81. Muslim, A., Mohd Sofian, S., Shaari, S. A., Hoh, B. P. and Lim, Y. A. L. (2019) Prevalence, intensity and associated risk factors of soil transmitted helminth infections: A comparison between Negritos (indigenous) in inland jungle and those in resettlement at town peripheries. *PLoS Negl. Trop. Dis.*, 13, e0007331
82. Sahlan, N., Fadzilah, M. N., Muslim, A., Shaari, S. A., Abdul Rahman, T. and Hoh, B. P. (2019) Hepatitis B virus infection: Epidemiology and seroprevalence rate amongst Negrito tribe in Malaysia. *Med. J. Malaysia*, 74, 320–325
83. Ishida, T., Yamamoto, K. and Omoto, K. (1988) A seroepidemiological survey of HTLV-I in the Philippines. *Int. J. Epidemiol.*, 17, 625–628
84. Phipps, M. E., Chan, K. K. L., Naidu, R., Mohamad, N. W., Hoh, B.-P., Quek, K.-F., Ahmad, B., Harnida, S. M., Zain, A. Z. and Kadir, K. A. (2015) Cardio-metabolic health risks in indigenous populations of Southeast Asia and the influence of urbanization. *BMC Public Health*, 15, 47
85. Rønn, P. F., Lucas, M., Laouan Sidi, E. A., Tvermosegaard, M., Andersen, G. S., Lauritzen, T., Toft, U., Carstensen, B., Christensen, D. L. and Jørgensen, M. E. (2017) The obesity-associated risk of cardiovascular disease and all-cause mortality is not lower in Inuit compared to Europeans: A cohort study of Greenlandic Inuit, Nunavik Inuit and Danes. *Atherosclerosis*, 265, 207–214
86. Appel, E. V. R., Moltke, I., Jørgensen, M. E., Bjerregaard, P., Linneberg, A., Pedersen, O., Albrechtsen, A., Hansen, T. and Grarup, N. (2018) Genetic determinants of glycated hemoglobin levels in the Greenlandic Inuit population. *Eur. J. Hum. Genet.*, 26, 868–875
87. Lu, D. and Xu, S. (2013) Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. *Front. Genet.*, 4, 127
88. Wang, L., Olson, J. E., Bielinski, S. J., St Sauver, J. L., Fu, S., He, H., Cicek, M. S., Hathcock, M. A., Cerhan, J. R. and Liu, H. (2020) Impact of diverse data sources on computational phenotyping. *Front. Genet.*, 11, 556
89. Caron, N. R., Chongo, M., Hudson, M., Arbour, L., Wasserman, W. W., Robertson, S., Correard, S. and Wilcox, P. (2020) Indigenous Genomic Databases: pragmatic considerations and cultural contexts. *Front. Public Health*, 8, 111