Journal of King Saud University – Science 35 (2023) 102593

HOSTED BY

Contents lists available at ScienceDirect

Journal of King Saud University – Science

journal homepage: www.sciencedirect.com

Review

Population genetics and human health in the genomic era

Mat Ghani Siti Nor Assyuhada [a], Zefarina Zulkafli [b,*], Noor Haslina Mohd Noor [b], Siddhartha Pati [c], Zulhisyam Abdul Kari [d], Mahmoud A.O. Dawood [e], Geoffrey Keith Chambers [f,*], Hisham Atan Edinur [g,*]

[a] Biomedicine Programme, School of Health Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian 16150, Kelantan, Malaysia
[b] School of Medical Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian 16150, Kelantan, Malaysia
[c] Centre of Excellence, Khallikote University, Berhampur, Ganjam, Odisha 761008, India
[d] Faculty of Agro Based Industry, Universiti Malaysia Kelantan, Jeli Campus, 17600 Jeli, Kelantan, Malaysia
[e] Department of Animal Production, Faculty of Agriculture, Kafrelsheikh University, 33516 Kafrelsheikh, Egypt
[f] School of Biological Sciences, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand
[g] Forensic Programme, School of Health Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian 16150, Kelantan, Malaysia

ARTICLE INFO

Article history:
Received 6 December 2021
Revised 2 February 2023
Accepted 7 February 2023
Available online 13 February 2023

Keywords:
Genetic markers
Population genetic data
Molecular techniques
Human health

ABSTRACT

Population genetic data collected from variable regions in human genome have been extensively used for studying human origins and migration patterns, estimating the probative value of DNA evidence and in disease association studies. Here, we illustrate the value of population genetic data in the genomic era with an emphasis on their role in health science and practice. This commentary is intended to show the value of an ethnicity-based approach to medicine and the role of genetics via accumulated population genetic data as its rational support basis. For this specific reason, we feel that new genome-based knowledge and resources need to be disseminated urgently to health professionals, researchers, policy makers and the public, so that it may be fully integrated into health-related policy and decision making. Ideally, all population genetic databases should be freely accessible, but this creates several technical issues which need to be properly considered. Those highlighted here include such as factors sample size, marker type, population descriptions and genotyping coverage.

Contents

* Corresponding authors.
E-mail addresses: zefarina@usm.my (Z. Zulkafli), geoff.chambers@vuw.ac.nz (G.K. Chambers), edinur@usm.my (H.A. Edinur).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksus.2023.102593
1018-3647/

1. Introduction

Variable regions of the human genome are known as genetic markers. They are commonly used for studying human ancestry (Hatin et al., 2011; Reich et al., 2012; Schroeder et al., 2015), individual identification via DNA profiling in crime cases (Coble et al., 2004; Phillips et al., 2007), genetic mapping of disease loci and tissue matching in transfusion and transplant medicine (Morishima et al., 2002; Pratschke et al., 2016). A wide range of either serological or molecular techniques have been adopted for screening genetic variability between individuals or populations. This former suite of methodologies is now gradually being replaced by molecular techniques such as the polymerase chain reaction (PCR) with sequence specific primers (SSP), restriction fragment length polymorphism (RFLP) analysis and DNA sequencing, now including 'next generation' systems. These DNA-based technologies have high specificity and analysts enjoy easy access to testing reagents (Fukuen et al., 2002; Vilches et al., 2007; Ladouceur et al., 2012). In this regard, we do note that phenotyping with antibodies is still the gold standard methodology in tissue typing laboratories engaged in compatibility testing of donors and patients. However, even in this type of laboratory new molecular techniques have to be applied in complex cases, such as for genotyping chronically alloimmunized patients. The status of any such individual cannot be determined by serological techniques due to the presence of circulating alloantibodies (Kulkarni et al., 2018).

A significant landmark of human genomic study occurred in 2003 with the completion of Human Genome Project. Their database identified approximately 25,000 – 30,000 protein-coding genes in the genome (Southan, 2004). This initiative stimulated the invention of several new sequencing technologies starting from around 2010. Collectively these known as next-generation DNA sequencing (NGS). They have far higher genotyping capacity than the conventional Sanger DNA sequencing methodology and can routinely capture sequence diversity in genomes at an unprecedented scale (Lind et al., 2010; Liu et al., 2014; Duke et al., 2016; Steward et al., 2017). Their further development has led to the generation of genuinely abundant human genomic data (Steward et al., 2017). Today, both NGS and Sanger sequencing methodologies as well as other molecular techniques such as PCR-SSP and PCR-RFLP are widely used in various molecular biological research disciplines including population genetics, forensics and medical genetics. This commentary will focus on population genetic data and their value in the modern era of medicine and refer to Simpson et al. (2014), Karp et al. (2008), Ahmed and Shabani (2019) and Takashima et al. (2018) for ethical concerns related to the sharing and aggregation of genetic data.

2. Population genetic data and health

Over recent years, population genetic data from many different human ethnic groups have been reported. These range from single nucleotide polymorphism (SNPs), simple bi-allelic gene loci such as human platelet antigen (HPA), human neutrophil antigen (HNA) through those with variable copy number (e.g., killer cell immunoglobulin-like receptor: KIR) to highly polymorphic regions (human leukocyte antigens: HLAs and short tandem repeats: STR). The now abundant data from these markers have been used for many different purposes including for ancestry analysis (Edinur et al., 2012, 2013a; Abidin et al., 2020; Hajar et al., 2020a; Hakim et al., 2020a), investigative leads (e.g., ancestry-informative panels of SNPs for ancestry inferences of suspects and missing persons), and estimating the probative value or weighting DNA identification evidence (Tvedebrink and Eriksen, 2019; Hakim et al., 2020b; Oldoni et al., 2020) and reference sets for disease association studies (Chasman et al., 2004; Baye et al., 2011; Jagannathan et al., 2011; Rohana et al., 2011). Indeed, many population genetic data originally acquired ancestry or forensic studies have turned out to have beneficial value for health as described below.

2.1. Searching for transfusion and transplant donor

The medically relevant HLA, blood group, HPA, HNA, KIR and cytokine population datasets are used for searching compatible blood and organ donors to avoid post-transplantation and transfusion complications (Morishima et al., 2002; Chambers et al., 2016; Orzinska et al., 2019). These experiences show the enduring value of attaching reliable ethnicity data to such collections of genetic information. For instance, individuals with Jk (a-b+) and Fy (a+b+) phenotypes are uncommon in African Americans, with frequencies around 8.1% and 1.0%, respectively (Barclay, 2001). Thus, compatible donor for a patient with Jk (a-b+) and Fy (a+b+) phenotypes might be difficult to find from within this population. However, these Kidd and Duffy blood group phenotypes are much more widely distributed among other groups; e.g., Malaysians (19.0–38.0% and 9.0–38.0%, respectively; see Hajar et al., 2020b; Hajar et al., 2021) and Koreans (28.3% and 11.4%, respectively; see Kim et al., 2003). This example amply demonstrates how completed population genetic data can be helpful when searching for donors with rare phenotypes and see sub-section 2.3 for notes on ethnicity/race vs genetic ancestry and health.

2.2. Estimating risk of disease susceptibility and the effectiveness of medical treatment

Certain population-specific genetic variants have been associated with a high risk of disease susceptibility including for autoimmune and infectious diseases and in predisposition to some cancers (Norhalifah et al., 2018; Sakuraba et al., 2020; Bukhari et al., 2021). They contribute to differences in disease profiles even when one controls for social factors like housing quality, diet, lifestyle, geographic location, poverty etc. For example, *APOL1*, a frequent genotype among populations with African ancestry is linked to kidney diseases (Borrell et al., 2021), while a SNP (rs370140172) in *EYS* gene on the long arm of chromosome 6 is associated with type-2 diabetes in Native Hawaiians (Sun et al., 2021). Equally, ancestry specific variants are highly significant in determining the effectiveness of medical treatments. One good example is the well-known antiplatelet medication, Plavix, that has been shown to be ineffective for treatment of heart disease and stroke in majority of Asians and Polynesians due to the absence of the *CYP2C19* allele in these population groups (Wu et al., 2015; Borrell et al., 2021). More recently, Zhang et al. (2021) reported SNPs in genes associated with coronavirus 2019 susceptibility in ethnically diverse subjects with African and European ancestry and their findings were validated against their health records. This study shows how different (e.g., frequency data for rs138390800, rs147311723, rs61735795, rs367866934 variants located in genes relevant to severe acute respiratory syndrome coronavirus 2 infection) might be between some groups due to local selection forces even when they are ancestrally related. These few examples demonstrate the value of population genetic data (if available) where associated disease prevalence is known. Medical decisions can be made and outcomes predicated based on ethnicity information (see later).

2.3. Searching for novel markers of disease susceptibility

As described in the earlier sections, population genetic data are regularly being reported for many ethnically related and unrelated population groups. Ethnicity classification tends to be defined by

Table 1
Here we show how social definition obscures the ancestral genetic signal using HLA-A and -B allele frequencies for Europeans, social and ancestry defined Maori. Data for Europeans and Maori were extracted from Edinur et al. (2012). Projected values for the wider set of socially self-defined Maori individuals can be calculated from the geometric mean across the two right hand columns.

HLA allele	Europeans (n= 545)	Admixed Maori (n = 65)	Full ancestry Maori (n = 49)
A*01	0.22	0.08	0.05
A*02	0.27	0.31	0.36
A*03	0.13	0.07	0.00
A*11	0.05	0.18	0.16
A*23	0.02	0.00	0.00
A*24	0.08	0.19	0.38
A*25	0.02	0.02	0.00
A*26	0.02	0.02	0.00
A*29	0.04	0.00	0.00
A*30	0.02	0.01	0.00
A*31	0.04	0.01	0.00
A*32	0.04	0.01	0.00
A*33	0.01	0.01	0.00
A*34	0.00	0.08	0.05
A*68	0.04	0.02	0.00
B*07	0.14	0.07	0.02
B*08	0.13	0.04	0.02
B*13	0.02	0.00	0.04
B*14	0.04	0.04	0.00
B*15	0.00	0.05	0.01
B*18	0.04	0.02	0.00
B*27	0.04	0.01	0.01
B*35	0.06	0.01	0.00
B*37	0.01	0.01	0.00
B*38	0.01	0.01	0.00
B*39	0.01	0.12	0.09
B*40	0.08	0.16	0.20
B*41	0.01	0.00	0.00
B*44	0.17	0.08	0.02
B*45	0.01	0.01	0.00
B*47	0.01	0.01	0.00
B*48	0.00	0.10	0.14
B*49	0.01	0.00	0.00
B*50	0.02	0.02	0.00
B*51	0.07	0.01	0.00
B*53	0.03	0.00	0.00
B*54	0.00	0.01	0.00
B*55	0.03	0.16	0.37
B*56	0.01	0.05	0.07
B*57	0.04	0.01	0.00
B*58	0.01	0.02	0.00

the original purpose of the study. Here, forensics and population genetics provide the perfect contrast. Forensics is primarily interested in a social definition of ethnicity (in New Zealand this is self-declared ethnic affiliation), but population genetics demands ethnicity information to be based on ancestry. Hence, any forensic database in New Zealand will include 'Maori' who are both full blood Maori (as medicine and population genetics require) and admixed Maori – e.g., refer (Edinur et al., 2012, 2013a, 2013b, 2013c). When DNA samples obtained during the development of a New Zealand forensic DNA database were examined (with all due Ethics permits), we found differences between members of two sub-sets of data (i.e., those from full ancestry vs admixed heritage Maori); refer Table 1. These two type are about 50:50 among our volunteers and allele frequency data for the latter were more or less exactly intermediate between those for the former and for matching reference data for Europeans (interpreted by us to mean that the admixture fraction was around 0.50 in this sub-set). Therefore, it is imperative that medical analysts properly understand where their data come from and how participants are defined. Here, population stratification can lead to low statistical power (increasing the genomic variability of each cohort) or leading false positive results (if case and controls belong to different ancestral groups) in medical research. In admixed populations, the individual ancestral mosaic is different from one person to another and may possibly create problems when searching for genetic variants associated with disease pathogenesis. Possible solutions would be to create different case-control cohorts by ancestral origin, or by considering ancestral variables and not ethnicity in etiological formulations; e.g., refer Peterson et al. (2019) for details. In our experience, population stratification continues to occur especially in multi-ethnic countries, as previously observed for Polynesians and Europeans in New Zealand. This is associated with extensive inter-marriage in the community (Chambers et al., 2016). For this specific reason, multiple cohorts, ancestral variables and tabulation of admixed and full ancestry population datasets (refer Table 1) can be used to estimate frequency distribution of a particular markers for medical purposes. In general, similarities and differences between population datasets can be powerful tools because they may reveal variants that can be used as candidates in disease association studies. A good marker would then be one with large differences in frequency and disease prevalence between populations, such as the ones described earlier for *APOL1* and *CYP2C19* in sub-section 2.2 and see later on 1000 Genomes Project in section 3.

2.4. Reference controls for other diseases

We would also like to note that any genetic markers that are being screened as control data for a specific disease may also have other unforeseen extended medical benefits. This is because, many diseases map to a common region with otherwise anonymous markers in the human genome. One good example is HLA where many infectious diseases and cancers are linked to this extended gene region (Shukla et al., 2015; Shi et al., 2020). In this context, a single set of HLA population data can be developed and used for studying multiple diseases.

3. Current developments, technical considerations & future directions

Medicine today is an increasingly dependent on genetic knowledge and we are moving towards precision medicine where even individual treatment and disease prevention will be based on information about personal genetics, lifestyle, environmental factors etc. But until then, there are already plenty of known genetic differences between people of different ethnicities, ancestry or simply between those from different geographic regions that are highly valuable for estimating disease susceptibility and response to medical treatment. However, as the new age of personalized genomic medicine dawns we remain in total agreement with Borrell et al. (2021):

> *"Indeed, we contend that the epidemiologic importance of race/ethnicity will never disappear."*

For this reason, several public databases are now dedicated to storing these kinds of datasets. They include the Allele Frequency Net Database (AFND), Immuno Polymorphism Database (IPD), gnomAD and Ensembl - refer https://www.allelefrequencies.net/, https://www.ebi.ac.uk/ipd/, https://gnomad.broadinstitute.org/ and https://www.ensembl.org/index.html, respectively for details.

AFND is the largest population genetic database storing frequency data (alleles, genes, haplotypes and genotypes) of HLA, KIR, major histocompatibility complex class I chain related genes (MIC) and a number of cytokine gene polymorphisms found in worldwide populations. Information from this resource has been extensively used in a variety of settings such as histocompatibility, immunology, epidemiology, pharmacogenetics, and population

genetics (Gonzalez-Galarza et al., 2020). However, data stored in AFND represent only one relatively small (but also one of the most polymorphic) region of the human genome. Any population genetic data deposited in these collections should only ever be collected from well-characterized and well-described populations (i.e., indicating full or admixed ancestry) and be of sufficient sample size to properly reflect the particular population as a whole (Borrell et al., 2021; Edinur et al., 2022). Population data collected from limited numbers of individuals may not adequately represent the actual genetic diversity in that particular population and subsequently may affect the accurate interpretation of any statistical results if they are used as a reference standard.

Efforts should also be made to study and include as many as population groups as possible in any particular population database. This will not only ensure equity in medical research, but will also help to give better insights into genetic variability across the planet and its wider influences on health (Banerjee and Chaudhury, 2010; Landry et al., 2018). Currently, AFND holds data from more than 1600 populations including records from over ten million healthy individuals belonging to 141 countries. However, there are many more population datasets for Europeans in AFND than there are people from Africa and Central Asia (Tshabalala et al., 2015; Santos et al., 2016; Gonzalez-Galarza et al., 2020).

Currently, NGS is becoming a widely affordable platform and has the ability to capture entire whole genomes. One good example is the 1000 Genome Project which compiles genomic information generated using NGS and aims to create public databases of human genetic variants that can be used for a wide range of biological research (Larson et al., 2015; McKenna et al., 2010). The power of this type of data is shown where as many as 53% of rare variants detected are only found in individual populations, and 17% of low-frequency variants are detected exclusively in single ancestry groups. It is thus crucial for this project to be expanded further as there are many population groups that are either not included at all and/or underrepresented (e.g. Persians and Austronesians) in this global genome-wide survey (Wong et al., 2013; Thareja et al., 2015). Our own genetic research programme showed that there are differences even between ethnically related populations. For example, products of HLA genes (class I and II) that determine tissue specificity are diverse and different between Malays and Polynesians, even though they share a partial Austronesian origin in common (Chambers and Edinur, 2013, 2021; Edinur et al., 2012, 2013c).

4. Conclusions

Genetic screening using various molecular techniques have produced large amounts of population data that can be used for improving human health, as described in this short commentary. Here we endorse the value of an ethnicity/genetic ancestry-based approach to medicine and the role of genetics via the generated population genetic data as its rational support basis. This view has becoming clearer and is even significant when we are moving towards the era of precision medicine. However, several key elements such as sample size, marker type, population description/stratification and genotyping coverage should be properly evaluated before any population genetic data can be used as valid resources to health professionals, researchers or policy makers.

Funding

This research was funded by Ministry of Higher Education Malaysia (Fundamental Research Grant Scheme: FRGS/1/2020/STG03/USM/03/5).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to acknowledge financial support given by Ministry of Higher Education Malaysia (Fundamental Research Grant Scheme: FRGS/1/2020/STG03/USM/03/5). GKC thanks VUW for alumnus scholar support.

References

Abidin, N.W.H.Z., Nor, N.M., Sundararajulu, P., Zafarina, Z., 2020. Understanding the genetic history of Malay populations in Peninsular Malaysia via KIR genes diversity. Am. J. Hum. Bio. e23545. https://doi.org/10.1002/ajhb.23545.

Ahmed, E., Shabani, M., 2019. DNA data marketplace: an analysis of the ethical concerns regarding the participation of the individuals. Front. Genet. 10, 1107. https://doi.org/10.3389/fgene.2019.01107.

Banerjee, A., Chaudhury, S., 2010. Statistics without tears: Populations and samples. Ind. Psychiatry J. 19, 60. https://doi.org/10.4103/0972-6748.77642.

Barclay, S., 2001. Red blood cell antigens and human blood groups. In: Hillyer, C.D., Strobl, F.J., Silberstein, L.E., Hillyer, K.L., Jefferies, L.C. (Eds.), Handbook of Transfusion Medicine. Elsevier, The Netherland, pp. 91–106.

Baye, T.M., Kovacic, M.B., Myers, J.M.B., Martin, L.J., Lindsey, M., Patterson, T.L., He, H., Ericksen, M.B., Gupta, J., Tsoras, A.M., et al., 2011. Differences in candidate gene association between European ancestry and African American asthmatic children. PLoS One 6, e16522.

Borrell, L.N., Elhawary, J.R., Fuentes-Afflick, E., Witonsky, J., Bhakta, N., Wu, A.H.B., Bib-bins-Domingo, K., Rodríguez-Santana, J.R., Lenoir, M.A., Gavin 3rd, J.R., et al., 2021. Race and Genetic Ancestry in Medicine - A Time for Reckoning with Racism. N. Engl. J. Med. 384, 474–480. https://doi.org/10.1056/NEJMms2029562.

Bukhari, W., Khalilidehkordi, E., Mason, D.F., Barnett, M.H., Taylor, B.V., Fabis-Pedrini, M., Kermode, A.G., Subramanian, S., Waters, P., Broadley, S.A., et al., 2021. NMOSD and MS prevalence in the Indigenous populations of Australia and New Zealand. J. Neurol. 1–10. https://doi.org/10.1007/s00415-021-10665-9.

Chambers, G.K., Edinur H.A., 2013. Genetic relationships between Malays and Maori. In: The Changing Values of Malays, Maori and Pacific Islanders, Rahman W.R.A., Higgins T.R. (Eds.), Chair of Malay Studies & Victoria University Foundation., Wellington, New Zealand, pp. 6–37.

Chambers, G.K., Edinur, H.A., 2021. Reconstruction of the Austronesian Diaspora in the Era of Genomics. Hum Biol., 182

Chambers, G.K., Edinur, H.A., Dunn, P.P.J., 2016. New insights into ancestry and health of Polynesians and New Zealand Mäori. New Zealand Sci. Rev. 73, 3–10.

Chasman, D.I., Posada, D., Subrahmanyan, L., Cook, N.R., Stanton Jr, V.P., Ridker, P.M., 2004. Pharmacogenetic study of statin therapy and cholesterol reduction. JAMA 291, 2821–2827. https://doi.org/10.1001/jama.291.23.2821.

Coble, M.D., Just, R.S., O'Callaghan, J.E., Letmanyi, I.H., Peterson, C.T., Irwin, J.A., Parsons, T.J., 2004. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int. J. Legal Med. 118, 137–146. https://doi.org/10.1007/s00414-004-0427-6.

Duke, J.L., Lind, C., Mackiewicz, K., Ferriola, D., Papazoglou, A., Gasiewski, A., Heron, S., Huynh, A., McLaughlin, L., Rogers, M., et al., 2016. Determining performance characteristics of an NGS-based HLA typing method for clinical applications. HLA 87, 141–152. https://doi.org/10.1111/tan.12736.

Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Velickovic, Z.M., Lea, R.A., Chambers, G.K., 2012. Using HLA loci to inform ancestry and health in Polynesian and Maori populations. Tissue Antigens 80, 509–522. https://doi.org/10.1111/tan.12026.

Edinur, H.A., Dunn, P.P.J., Lea, R.A., Chambers, G.K., 2013a. Human platelet antigens frequencies in Maori and Polynesian populations. Transfus. Med. 23, 330–337. https://doi.org/10.1111/tme.12061.

Edinur, H.A., Dunn, P.P.J., Lea, R.A., Chambers, G.K., 2013b. Molecular approaches to transfusion medicine in Polynesians and Maori in New Zealand. Int. J. Immunogenet. 40, 460–470. https://doi.org/10.1111/iji.12073.

Edinur, H.A., Dunn, P.P.J., Hammond, L., Selwyn, C., Brescia, P., Askar, M., Reville, P., Velickovic, Z.M., Lea, R.A., Chambers, G.K., 2013c. HLA and MICA polymorphism in Polynesians and New Zealand Maori: Implications for ancestry and health. Hum. Immunol. 74, 1119–1129. https://doi.org/10.1016/j.humimm.2013.06.011.

Edinur, H.A., Mat-Ghani, S.N.A., Chambers, G.K., 2022. Ethnicity-based classifications and medical genetics: One Health approaches from a Western Pacific perspective. Front. Genet. 13,. https://doi.org/10.3389/fgene.2022.970549 970549.

Fukuen, S., Fukuda, T., Maune, H., Ikenaga, Y., Yamamoto, I., Inaba, T., Azuma, J., 2002. Novel detection assay by PCR–RFLP and frequency of the CYP3A5 SNPs,

CYP3A5*3 and *6, in a Japanese population. Pharmacogenet. Genomics 2002 (12), 331–334. https://doi.org/10.1097/00008571-200206000-00009.
Gonzalez-Galarza, F.F., McCabe, A., Santos, E.J.M.d., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al., 2020. Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 48, D783–D788. https://doi.org/10.1093/nar/gkz1029.
Hajar, C.G.N., Zefarina, Z., Riffin, N.S.M., Mohammad, T.H.T., Hassan, M.N., Poonachi, P., Safuan, S., ElGhazali, G., Chambers, G.K., Edinur, H.A., 2020a. Extended blood group profiles for Malays, Chinese, and Indians in peninsular Malaysia. Egypt. J. Med. Hum. Genet. 21, 1–10. https://doi.org/10.1186/s43042-020-00096-y.
Hajar, C.G.N., Zulkafli, Z., Riffin, N.S.M., Mohammad, T.H.T., Safuan, S., Nelson, B.R., Abdullah, M.T., Chambers, G.K., Edinur, H.A., 2020b. Human neutrophil antigen frequency data for Malays, Chinese and Indians. Transfus. Apher. Sci. 59,. https://doi.org/10.1016/j.transci.2019.09.004 102651.
Hajar, C.G.N., Zefarina, Z., Riffin, N.S.M., Mohammad, T.H.T., Hassan, M.N., Poonachi, P., Safuan, S., ElGhazali, G., Chambers, G.K., Edinur, H.A., 2021. Correction to: Extended blood group profiles for Malays, Chinese, and Indians in peninsular Malaysia. Egypt. J. Med. Hum. Genet. 22, 8.
Hakim, H.M., Khan, H.O., Ismail, S.A., Lalung, J., Kofi, A.E., Abdullah, M.T., Chambers, G.K., Edinur, H.A., 2020a. Forensic parameters and ancestral fractions in the Kedayan population inferred using 21 autosomal STR loci. Meta Gene 25,. https://doi.org/10.1016/j.mgene.2020.100741 100741.
Hakim, H.M., Khan, H.O., Lalung, J., Nelson, B.R., Chambers, G.K., Edinur, H.A., 2020b. Autosomal STR Profiling and Databanking in Malaysia: Current Status and Future Prospects. Genes 11, 1112.
Hatin, W.I., Nur-Shafawati, A.R., Zahri, M.-K., Xu, S., Jin, L., Tan, S.-G., Rizman-Idid, M., Zilfalil, B.A., Consortium, H.P.-A.S., 2011. Population genetic structure of peninsular Malaysia Malay sub-ethnic groups. PLoS One 6, e18312. https://doi.org/10.1371/journal.pone.0018312.
Jagannathan, L., Chaturvedi, M., Satish, B., Satish, K.S., Desai, A., Subbakrishna, D.K., Satishchandra, P., Pitchappan, R., Balakrishnan, K., Kondaiah, P., et al., 2011. HLA-B57 and gender influence the occurrence of tuberculosis in HIV infected people of south India. J. Immunol. Res. 2011,. https://doi.org/10.1155/2011/549023 549023.
Karp, D.R., Carlin, S., Cook-Deegan, R., Ford, D.E., Geller, G., Glass, D.N., Greely, H., Guthridge, J., Kahn, J., Kaslow, R., et al., 2008. Ethical and practical issues associated with aggregating databases. PLoS Med. 5, e190.
Kim, I.T., Suh, I.B., Ma, K.R., Lim, C.S., Kim, Y.K., Lee, K.N., 2003. The genotyping of Kell, Duffy, and Kidd system in Korean. Korean J. Blood Transfus. 14, 9–19.
Kulkarni, S., Choudhary, B., Gogri, H., Patil, S., Manglani, M., Sharma, R., Madkaikar, M., 2018. Molecular genotyping of clinically important blood group antigens in patients with thalas-saemia. Indian J. Med. Res. 148, 713. https://doi.org/10.4103/ijmr.IJMR_455_17.
Ladouceur, M., Dastani, Z., Aulchenko, Y.S., Greenwood, C.M.T., Richards, J.B., 2012. The empirical power of rare variant association methods: Results from sanger sequencing in 1,998 individu-als. PLoS Genet. 8, e1002496. https://doi.org/10.1371/journal.pgen.1002496.
Landry, L.G., Ali, N., Williams, D.R., Rehm, H.L., Bonham, V.L., 2018. Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. Health Aff. 37, 780–785. https://doi.org/10.1377/hlthaff.2017.1595.
Larson, J.L., Silver, A.J., Chan, D., Borroto, C., Spurrier, B., Silver, M.S., 2015. Validation of a high resolution NGS method for detecting spinal muscular atrophy carriers among phase 3 participants in the 1000 Genomes Project. BMC Med. Genet. 16, 100. https://doi.org/10.1186/s12881-015-0246-2.
Lind, C., Ferriola, D., Mackiewicz, K., Heron, S., Rogers, M., Slavich, L., Walker, R., Hsiao, T., McLaughlin, L., D'Arcy, M.D., et al., 2010. Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing. Hum. Immunol. 71, 1033–1042. https://doi.org/10.1016/j.humimm.2010.06.016.
Liu, Z., Liu, M., Mercado, T., Illoh, O., Davey, R., 2014. Extended blood group molecular typing and next-generation sequencing. Transfus. Med. Rev. 28, 177–186. https://doi.org/10.1016/j.tmrv.2014.08.003.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
Morishima, Y., Sasazuki, T., Inoko, H., Juji, T., Akaza, T., Yamamoto, K., Ishikawa, Y., Kato, S., Sao, H., Sakamaki, H., et al., 2002. The clinical significance of human leukocyte antigen (HLA) allele compatibility in patients receiving a marrow transplant from serologically HLA-A, HLA-B, and HLA-DR matched unrelated donors. Blood 99, 4200–4206. https://doi.org/10.1182/blood.v99.11.4200.
Norhalifah, H.K., Mat, N.F.C., Edinur, H.A., 2018. Cytokine Gene Polymorphisms in Cancer and Inflammatory Disorders. Curr. Immunol. Rev. 14, 81–93. https://doi.org/10.2174/1573395514666180724121419.
Oldoni, F., Yoon, L., Wootton, S.C., Lagacé, R., Kidd, K.K., Podini, D., 2020. Population genetic data of 74 microhaplotypes in four major U.S. population groups. Forensic Sci. Int. 49,. https://doi.org/10.1016/j.fsigen.2020.102398 102398.
Orzinska, A., Guz, K., Brojer, E., 2019. Potential of next-generation sequencing to match blood group antigens for transfusion. Int. J. Clin. Transfus. Med. 7, 11. https://doi.org/10.2147/IJCTM.S175142.
Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., et al., 2019. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell 179, 589–603. https://doi.org/10.1016/j.cell.2019.08.051.
Phillips, C., Salas, A., Sanchez, J.J., Fondevila, M., Gomez-Tato, A., Alvarez-Dios, J., Calaza, M., de Cal, M.C., Ballard, D., Lareu, M.V., et al., 2007. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci. Int. Genet. 1 (3–4), 273–280. https://doi.org/10.1016/j.fsigen.2007.06.008.
Pratschke, J., Dragun, D., Hauser, I.A., Horn, S., Mueller, T.F., Schemmer, P., Thaiss, F., 2016. Immunological risk assessment: The key to individualized immunosuppression after kidney transplantation. Transplant. Rev. 30, 77–84. https://doi.org/10.1016/j.trre.2016.02.002.
Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al., 2012. Reconstructing native American population history. Nature 488, 370–374. https://doi.org/10.1038/nature11258.
Rohana, A.G., Loh, K.C., Tin, S.K., Soh, C.H., Nazaimoon, W.M.W., Fong, K.Y., Azmi, K. N., Khalid, B.A.K., 2011. HLA-DQ A1,-DQB1 and -DRB1 gene polymorphism–in Malay type 1 diabetes mellitus patients and their use for risk prediction. Med. J. Malaysia 66, 133–137.
Sakuraba, A., Haider, H., Sato, T., 2020. Population Difference in Allele Frequency of HLA-C*05 and Its Correlation with COVID-19 Mortality. Viruses 12, 1333. https://doi.org/10.3390/v12111333.
Santos, E.J.M.D., McCabe, A., Gonzalez-Galarza, F.F., Jones, A.R., Middleton, D., 2016. Allele Frequencies Net Database: Improvements for storage of individual genotypes and analysis of existing data. Hum. Immunol. 77, 238–248. https://doi.org/10.1016/j.humimm.2015.11.013.
Schroeder, H., Ávila-Arcos, M.C., Malaspinas, A.-S., Poznik, G.D., Sandoval-Velasco, M., Carpenter, M.L., Moreno-Mayar, J.V., Sikora, M., Johnson, P.L.F., Allentoft, M. E., et al., 2015. Genome-wide ancestry of 17th-century enslaved Africans from the Caribbean. PNAS 112, 3669–3673. https://doi.org/10.1073/pnas.1421784112.
Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., Bucci, E., Piacentini, M., Ippolito, G., Melino, G., 2020. COVID-19 infection: the perspectives on immune responses. Cell Death Differ. 27, 1451–1454. https://doi.org/10.1038/s41418-020-0530-3.
Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al., 2015. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 33, 1152–1158. https://doi.org/10.1038/nbt.3344.
Simpson, C.L., Goldenberg, A.J., Culverhouse, R., Daley, D., Igo, R.P., Jarvik, G.P., Mandal, D.M., Mascalzoni, D., Montgomery, C.G., Pierce, B.L., et al., 2014. Practical barriers and ethical challenges in genetic data sharing. Int. J. Environ. Res. Public Health 11, 8383–8398. https://doi.org/10.3390/ijerph110808383.
Southan, C., 2004. Has the yo-yo stopped? An assessment of human protein-coding gene number. Proteomics 4, 1712–1726. https://doi.org/10.1002/pmic.200300700.
Steward, C.A., Parker, A.P.J., Minassian, B.A., Sisodiya, S.M., Frankish, A., Harrow, J., 2017. Genome annotation for clinical genomic diagnostics: Strengths and weaknesses. Genome Med. 9, 1–19. https://doi.org/10.1186/s13073-017-0441-1.
Sun, H., Lin, M., Russell, E.M., Minster, R.L., Chan, T.F., Dinh, B.L., Naseri, T., Reupena, M.S., Lum-Jones, A., Cheng, I., et al., 2021. The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. PLoS Genet. 17, e1009273.
Takashima, K., Maru, Y., Mori, S., Mano, H., Noda, T., Muto, K., 2018. Ethical concerns on sharing genomic data including patients' family members. BMC Med. Ethics 19, 1–6. https://doi.org/10.1186/s12910-018-0310-5.
Thareja, G., John, S.E., Hebbar, P., Behbehani, K., Thanaraj, T.A., Alsmadi, O., 2015. Sequence and analysis of a whole genome from Kuwaiti population subgroup of Persian ancestry. BMC Genomics 16, 1–14. https://doi.org/10.1186/s12864-015-1233-x.
Tshabalala, M., Mellet, J., Pepper, M.S., 2015. Human Leukocyte Antigen Diversity: A Southern African Perspective. J. Immunol. Res. 2015, 746151. https://doi.org/10.1155/2015/746151.
Tvedebrink, T., Eriksen, P.S., 2019. Inference of admixed ancestry with Ancestry Informative Markers. Forensic Sci. Int. Genet. 42, 147–153. https://doi.org/10.1016/j.fsigen.2019.06.013.
Vilches, C., Castano, J., Gomez-Lozano, N., Estefania, E., 2007. Facilitation of KIR genotyping by a PCR-SSP method that amplifies short DNA fragments. Tissue Antigens 70, 415–422. https://doi.org/10.1111/j.1399-0039.2007.00923.x.
Wong, L.-P., Ong, R.-H., Poh, W.-T., Liu, X., Chen, P., Li, R., Lam, K.-Y., Pillai, N.E., Sim, K.-S., Xu, H., et al., 2013. Deep Whole-Genome Sequencing of 100 Southeast Asian Malays. Am. J. Hum. Genet. 92, 52–66. https://doi.org/10.1016/j.ajhg.2012.12.005.
Wu, A.H., White, M.J., Oh, S., Burchard, E., 2015. The Hawaii clopidogrel lawsuit: the possible effect on clinical laboratory testing. Per. Med. 12, 179–181. https://doi.org/10.2217/pme.15.4.
Zhang, C., Verma, A., Feng, Y., Melo, M.C.R., McQuillan, M., Hansen, M. Lucas, A., Park, J., Ranciaro, A., Thompson, S., et al., 2021. Impact of natural selection on global patterns of genetic variation, and association with clinical phenotypes, at genes involved in SARS-CoV-2 infection. medRxiv https://doi.org/10.1101/2021.06.28.21259529.