The difference between genetics and genomics is that the former is the study of a single gene while the latter focuses on all the genes (the genome) as well as their interrelationships (1). Furthermore, genomics examines *the interactions of those genes with the person's environment* thus contributing to the efficient research-study of complex genetic diseases (CGD's) such as heart disease, asthma, diabetes, and cancer (2)

Almost 20 years ago in 2003, The Genome Project(commenced in 1990) was completed. The outcome wasthe'decoding' of the*complete human genome sequence*. It covered approximately 99% of the human genome's gene-containing regions and itsscientific significance has beencompared to splitting the atom or going to the moon (2).

Humans share 99.9% DNA instructions – the genome - which are made up of 3 billion 'bases' or smaller molecules adenine (A), thymine (T), cytosine (C), and guanine (G). The remainder is what makes us unique, and each person's version of this difference – the single nucleotide polymorphism (SNP) is called their genotype. (3)

The ability to sequence genomes and compare populations called 'genome-wide association studies' (GWAS) have been made possible by the collection of samples in Biobanks. However, 80% of this data has been collected from individuals with European ancestry which produces multiple impacts on genomic findings.

Firstly, robustness and representation for world health conclusions and recommendations. The population of Europe makes up only 5.9% of the world's population (4). Even if we were to allow for population mixing caused by migration of Europeans across the world e.g. European-Americans, the overwhelming majority of the world's population is non-European.

Secondly, an assessment of multiple GWAS concluded that associations between certain disease traits and variants found in European populations cannot be replicated in other populations, which can produce 'false positives' and opportunities to discover new associations with disease traits in other populationsare therefore missed**.**(5)

The value of findings and discoveries based on the current data sets for *all* populations - albeit from 80% European ancestry - is however enormous. The findings have contributed to a better understanding of CGDssuch asCoronary cardiovascular diseases, diabetes, cancers, and chronic obstructive pulmonary disease which are the most prevalent diseases worldwide and the major cause for socioeconomic burden *in all* the World Health Organization (WHO) classified geographical locations.(6)

The inclusion of more diverse populations may however hold the key to completing the gaps in genomic sequencing.For example,a study on Indians suggests that the unique genetic predisposition toward complex diseases of Indians could be due to their unique genetic constitution (6). The significance of this on world health is twofold. Firstly, at one-sixth, a sizable proportion of the world's population is Indian. Secondly, India ranks second in the world for CGDs.

The understanding of a worldwide major chronic disorder – nicotine dependence – has also been enhanced by multi-ethnic and diverse studies. For example, by comparing European-American with African American samples, 29 genes associated significantly with nicotine dependence were revealed (7).The use of Japanese samples has contributed to the identification of seven new loci – a very specific place where a gene is located - including three loci associated with the number of cigarettes per day with its authors claiming to have "characterised the genetic architecture of smoking behaviors" (8)

Finally, the value of using more diverse population samples for the understanding of SARS-CoV-2is evident in global efforts such as the *COVID-19 Host Genetics Initiative.* This will have a multi-ethnic human leukocyte antigen (HLA) reference panel constructed using whole-genome sequencing data from 21,546 individuals of five different populations: European, African, Latino, Asian, and South Asian. The advantage of such a diverse sample base is that this will allow the testing of each HLA allele and also each of the amino acid site position within HLA genes, in order to assess if they explain COVID risk. (9)

It has been almost 20 years since the geneticist James Franklin Crow, raised the importance of the mere (0.1%) of human DNA which contributes to different characteristics among populations.He highlighted its importance from both a genetic but also social science perspective. (10) Ten years ago, the European bias for DNA samples was as high as 96% (5) suggesting that today's 80% European representation is a small but important improvement. The need for a more diverserepresentation of DNA samples does not just serve to statistically'correct' the issue of representation. Rather, it indisputably serves to enhance genomic understanding and ultimately provide a fairer and all-inclusiveworld health improvement approach.

References

1) WHO definitions of Genetics and Genomics World Health Organisation(2021) https://www.who.int/genomics/geneticsVSgenomics/en

2) Human Genome Project Results, National Human Genome Institute(22 December 2020)

https://www.genome.gov/human-genome-project/results

3)Fine R. (27 February 2019)Lessons from the Genome Project Blog SITN 20 Celebrating 20 years.*Science In the News*, Harvard University

https://sitn.hms.harvard.edu/flash/2019/lessons-from-the-human-genome-project/

4) The EU in The World Population, Eurostat Statics Explained 2020

https://ec.europa.eu/eurostat/statistics-explained/index.php/The_EU_in_the_world_-_population#Population_size_and_population_density

5)PopejoyA.B. & Fullerton S.M.(2016) Genomics is failing on diversity, *Nature International Weekly Journal Of Science*, **538**, (7624)

https://www.nature.com/news/genomics-is-failing-on-diversity-1.20759

6)PranavchandR.& Reddy BM(2016) Genomics era and complex disorders: Implications of GWAS with special reference to coronary artery disease, type 2 diabetes mellitus, and cancers. *Journal of Postgraduate Medicine* Jul-Sep; **62**(3): 188–198.oi: 10.4103/0022-3859.186390

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970347/

7) Fan R, Cui. J, Ma Y, Yang Z, Payne TJ, Ma JZ, Li MD. Gene-based association analysis reveals involvement of LAMA5 and cell adhesion pathways in nicotine dependence in African- and European-American samples. *Addiction Biology*. 2021 Mar;**26**(2) e12898. doi:10.1111/adb.12898. PMID: 32281736.

http://europepmc.org/article/MED/32281736

8)Matoba N., Akiyama M., Ishigaki K., Kanai M., Takahashi A.,&YoichiroKamatani. (25 March 2019)GWAS of smoking behaviour in 165,436 Japanese people reveals seven new loci and shared genetic architecture.

https://www.nature.com/articles/s41562-019-0557-y

9)Eur J Hum Genet.(June 2020) The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *European Journal of Human Genetics*. 2020 Jun; 28(6): 715–718. Published online 2020 May 13. doi: 10.1038/s41431-020-0636-6 https://pubmed.ncbi.nlm.nih.gov/32404885/

10) Crow J.F *Daedalus*Winter 2002 Report Unequal by nature: a geneticist's perspective on human differences

https://www.amacad.org/publication/unequal-nature-geneticists-perspective-human-differences

Unequal by nature: a geneticist's perspective on human differences - American Academy of Arts & Sciences