International Haplotype Mapping Project

By

Clement A. Adebamowo
BM ChB Hons., FWACS, FACS, Dsc

Key words: Genomics, HapMap, Low resource environment

Clement A. Adebamowo is a Professor of Surgery and Director, West African Bioethics Training Program Division of Oncology, Department of Surgery, College of Medicine, University of Ibadan, University College Hospital, Ibadan, Oyo State, Nigeria

Introduction:

After the re-discovery of Mendel's work[1], the foundation of modern molecular biology was laid by Avery, McLeod and McCarty's classic identification of DNA as the carrier of genetic information[2]; a finding that induced Erwin Chargaff to change his line of research and he soon found that though the base composition of DNA varied among species, the amount of thymine always equaled that of adenine, while that of cytosine always equaled that of guanine; thus establishing the Chargaff rules. These rules paved the way for the work on the structure of the human genome by Watson, Crick and Franklin[3-5]. Shortly after, the biological mechanism for reading the information encoded in genes and the necessary recombinant technology was developed. This was followed by sequencing of the genome of a variety of plasmid, viruses, bacteria, some animals and plants. It was therefore just a matter of time before the sequence of the human genome would be tackled. That time came, after discussions at scientific meetings organized by the U.S. Department of Energy from 1984 to 1986[6] lead to the formation of the International Human Genome Sequencing Consortium (IHGSC) in 1990, an open consortium of twenty centers in six countries, charged with the task of producing an accurate sequence of the euchromatic part of the human genome.

The result of the work of the IHGSC was the publication of a draft of the Human Genome sequence in February 2001[6] followed by an update in 2004 that was suggested as a "firm foundation for biomedical research in the years ahead"[7]. Despite these huge strides, there has been a crisis of expectation among scientists about the human genome project. In the period leading up to and immediately following the sequencing of the human genome, several wildly, sometimes self-serving predictions were made about the potential of the Human Genome Sequence for understanding disease, natural variation, population structure and drug response. It was not unusual for articles to be published linking one gene or the other to diseases or other outcomes, only for follow-up studies to fail to confirm the initial findings or when they do, to find more conservative effects than the initial studies suggest[8]. A lot of these problems were due to poor study design, lack of power, small sample sizes, population heterogeneity and stratification, the multitude of end-points, sampling bias, publication bias, injudicious sub-set analysis and the unprecedented scale of the data generated. Regardless of these problems however, the potential that knowledge about the human genome has to fundamentally influence gene-disease association studies cannot be overstated. Indeed what we need is a handle for the pump[9].

There have been many responses to the problems. Authors and journals are more cautious about making claims. Bigger studies that take longer to mature are now being published for several gene-disease associations, more negative results are being published and the problem of heterogeneity and population stratification is being addressed both at the design and analytical stage of studies[9]. Furthermore, it is obvious now that to find genes that are associated with diseases or response to drug treatment, more powerful tools that permit efficient

Address for correspondence:

Division of Oncology, Department of Surgery, College of Medicine, University of Ibadan, University College Hospital, Ibadan, Oyo State, Nigeria.
E-mail: cadebamo@yahoo.com
Fax/Phone Number: 234-2-2410995

utilization of the human genome sequence data are needed.

The Science of the HapMap

One such tool is the Haplotype Map. Observation from population genetics suggest that ~90% of the genome variation in humans is due to common variants and that most of these variations arose from single historical mutation and demographic events, hence they remain in close association with other nearby mutations. The HapMap is predicated on the existence of combinations of these single nucleotide polymorphism (SNP) markers occurring in linkage disequilibrium (LD) thus permitting creation of blocks of genome, irrespective of functional status, which can be used for association studies[10]. The strong associations between these SNPs mean that knowledge of a few of them (tag SNPs) is sufficient to predict information about the rest. Given that the strength of the association between SNPs is highly variable, empirical methods must be used to determine the best tag SNPs. In this way, it is possible to use a much smaller subset of SNPs (approximately 200,000 to 1 million) to efficiently describe the approximately 10 million common SNPs in the population (Figure 1)[10].

Four populations were selected based on pilot data showing a range of haplotype frequencies[11], scientific recommendation to include at least 3 "Old World" populations, exclusion of small isolated populations, established relationship between researchers and the communities and interest of the funding agencies. It is believed that samples from these four populations would include most of the genetic variation found in all populations throughout the world. A total of 270 samples were obtained; 90 samples from a US population of people with Northern and Western European ancestry which had been collected in 1980 by the Centre d'Etude du Polymorphisme Humain (CEPH), 30 trios of two parents and a child older than 18 years of age from Yoruba living in Ibadan, 45 unrelated Han Chinese and 45 unrelated Japanese in Tokyo, Japan. The samples were transported to Coriell Laboratory in the United States where the cells were immortalized and shipped to the genotyping centers. The data generated by the HapMap is being released rapidly into the public database.

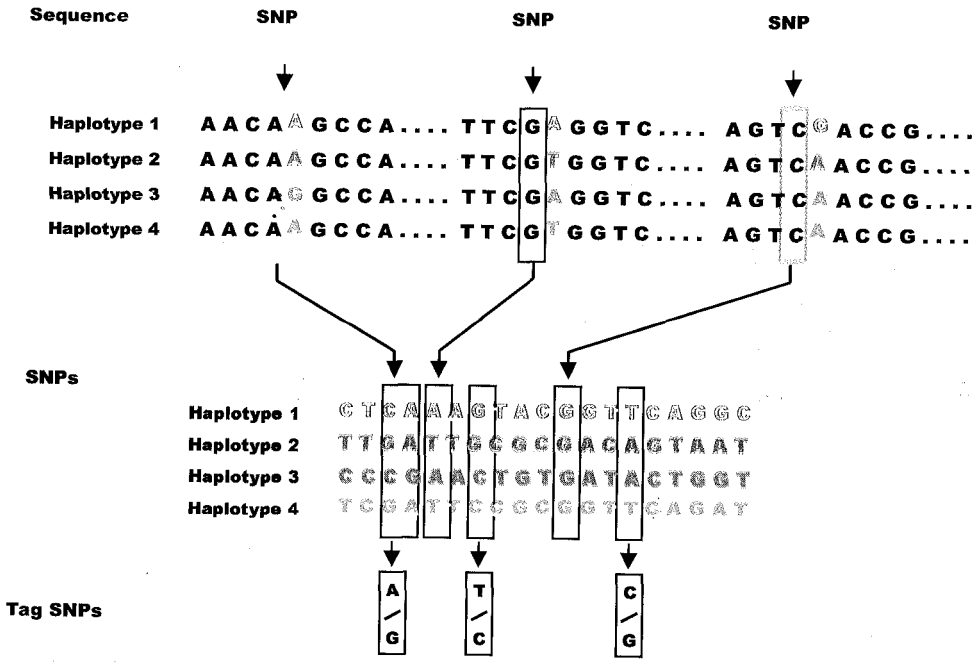

Figure 1: Tag SNPs can be used to create haplotype blocks that allow efficient use of the gnome sequence for association studies.

In order to realize the objective of creating a functional haplotype map that would be useful for most, if not all of the world's population, the International HapMap Consortium was formed in July 2001 and the project initiated in October 2002[10].

The Ethics of the HapMap

Uniquely among big science projects of this nature was the decision to identify and incorporate ethical concerns into the study right from its inception[12]. The inclusion of a population from a low resource country like Nigeria in a study of this nature seems hardly justifiably given the socio-economic and disease profile of the country. However several international guidelines have argued strongly for the

inclusion of the population from developing countries in genetics research[13]. In addition the profile of incident diseases and the causes of mortality indicate that Nigeria like many developing countries is undergoing a period of epidemiologic transition characterized by high incidences of both communicable and complex diseases[14]. It was also considered important to ensure that the concerns of the communities where sampling was done was given due consideration during the study as well as obtain their input into how their samples should be described. To accomplish this, an extensive community engagement process was put in place[12]. This involved several town meetings, key informant interviews, focus group discussions and community survey. A working group and a community advisory group was set up in an effort to ensure that all relevant ethical concerns are addressed and there was unimpeded multilevel interaction between the community members and the research team[12].

Other ethical issues are related to the decision to identify the population from where samples where obtained given the risk of stigmatization and over-generalization of the result of limited studies to whole populations. Failure to identify the populations would only create a false sense of security as these can always be discovered with a little effort[12]. Giving the communities the opportunity to name their samples gave members the opportunity to evaluate the implications of naming the population and deciding how they wished their community to be named[12]. It is also argued that citizen of developed countries are likely to access the benefits of these kinds of research faster than those of developing countries. However, disregarding the assumptions inherent in this statement, genomic research has already shown promise to deliver improved health care to people in low resource environment[15].

Clinical applications of HapMap

The HapMap was conceived and designed as a tool to aid the identification of the genetic basis of complex diseases. For this purpose, it is assumed that common diseases are due to common genetic variants (CD/CV hypothesis)[16]. Haplotypes may help to identify genetic factors associated with disease by serving as proxies for ungenotyped SNPs[17]. The HapMap may also be particularly useful in pharmacogenomics[17]. Pharmacogenomics holds the promise of individualized medicine whereby patients' predisposition to response or to side effects can be predicted from genomic tests and potentially useful therapies will not be discarded on the basis of crude adverse reaction reports[18,19]. These potential impacts of the HapMap and molecular biology in general on clinical practice must be complemented by readiness of clinicians to routinely obtain detailed phenotypic information about their patients and preparedness for the new type of medical practice that would result. Instead of the current practice of 'average dose' of treatment for the 'average patient', more individualized response to patients' illness that may include genomic test and interpretation of these tests are likely in the future. We need large scale pharmacogenomic studies that will confirm or refute many of these hypotheses.

Like any tool, the HapMap has many real and potential limitations. It has been argued that the sampling strategy chosen for the project is inadequate because linkage disequilibrium varies substantially among populations and the CD/CV hypothesis is not yet proven [20]. The International HapMap Consortium is committed to further studies among additional populations if studies on the current samples prove inadequate. However regardless of the potential limitations of the project and the risk that hitherto unanticipated ethical risks may occur, the HapMap is already leading to significant progress in genomics and clinical medicine.

References

1. Correns C. Untersuchungen uber die Xenien bei Zea mays. Berichte der Deutsche Botanische Gesellschaft 1899; 17:410 - 418.

2. Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med 1944; 79:137 - 158.

3.Franklin RE, Gosling RG. The structure of sodium thymonucleate fibres: II. The cylindrically symmetrical Patterson function. Acta Crystallogr 1953; 6:678 - 685.

4. Franklin RE, Gosling RG. The structure of sodium thymonucleate fibres: I. Influence of water content. Acta Crystallogr 1953; 6:673 - 677.

5. Watson JD, Crick FHC. Molecular structure of nucleic acids. Nature 1953; 171:737 - 738.

6.International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860 - 921.

7.International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431:931 - 945.

8.Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nature 2001; 29:306 - 309.

9. Little J, Khoury MJ, Bradley L, et al. The Human Genome Project is complete. How do we develop a

handle for this pump? Am J Epidemiol 2003; 157:667 - 673.
10.International HapMap Consortium. The International HapMap Project. Nature 2003; 426:789 - 796.
11.Gabriel SB, Schaffner SF, Nguyen H, et al. The Structure of Haplotype Blocks in Human Genome. Science 2002; 296:2225 - 2229.
12.International HapMap Consortium. Integrating ethics and science in the International HapMap Project. Nature Review Genetics 2004; 5:467 - 475.
13. Council for International Organizations of Medical Sciences (CIOMS). International Ethical Guidelines for Biomedical Research Involving Human Subjects. 2002.
14. Rodgers A, Ezzati M, Hoorn SV, et al. Distribution of Major Health Risks: Findings from the Global Burden of Disease Study. PLoS Med 2004; 1:e27.
15.Pang T, Weatherall D. Genomics and Global Health. BMJ 2002; 324:1051 - 1052.
16. Reich D, Lander ES. On the allelic spectrum of human disease. Trend Genet 2001; 17:199 - 204.
17. Austin CP. The impact of the completed human genome sequence on the development of novel therapeutics for human disease. Annu Rev Med 2004; 55:1 - 13.
18.Evans WE, Relling MV. Moving towards individualized medicine with pharmacogenomics. Nature 2004; 429:464 - 468.
19. Pao W, Miller VA, Politi KA, Riely GJ, Somwar R, Zakowski MF. Acquired resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib is associated with a second mutation in the EGFR Kinase Domain. PLoS Med 2005; 2:e73.
20. Sawyer SL, Mukherjee N, Pakstis AJ, et al. Linkage disequilibrium patterns vary substantially among populations. Vol. 2005: Eur J Hum Genet, 2005:doi: 10. 1038/sj.ejhg.5201368.