Turkish Journal of Molecular Biology & Biotechnology
(2016) 1: 67-71

REVIEW

Turkish JMBB

Genomic aspect of the human diversity

Peren H. KARAGİN, Zeynep BALABAN

Department of Medical Genetics, Başkent University, Faculty of Medicine, Maltepe, Ankara, Turkey

Keywords
Mutation, variation, polymorphism, human genome

ABSTRACT • *Homo sapiens* has been considered to be the most interesting species on earth for hundreds of years. Human genetic has progressed to accelerating period after the discovery of the DNA and development of recombinant DNA and molecular techniques during the last 50 years. *Homo sapiens* is a relatively young species. Therefore, *Homo sapiens* has not had much time to accumulate genetic variation comparing to the other species on earth. Nevertheless, there is considerable genetic variation between the individuals. The human genome comprises about $3x10^9$ base pairs of DNA. No two humans have been same DNA sequence in the world. Between any two humans, the ratio of genetic variation is about 0.1 percent. This means, nearly $6x10^6$ base pairs are different. The most common genetic difference in the human genome is single nucleotide differences. One person has 5-10 billion single nucleotide polymorphisms. The less frequent observed differences are deletion, insertion, inversion, duplication, copy number variation and genomic rearrangements (translocations). Most of the human genetic variations are insignificant. For example, neutral mutations and silent mutations do not change the phenotype of the organism significantly. Some of the genetic mutations can result positively for an individual. However, vast majority of the mutations have no known function. Benefits of understanding human genetic variations are its value for promoting human health and combating disease. It will also contribute to the rapidly growing field known a pharmacogenomics and help to develop personalized medicine. In the future, physicians will use genetic test before the prescription

INTRODUCTION

Why are we working human genetics? One reason is to know ourselves better. Another reason is practical value for human well-being. Studies showed that they have a large potential to improve the human life. Society has been willing to devote a significant amount of budget for research in this field. This perception and inventions of the past 20 years cause to increase in the number of genetic organizations and researchers.

Genetic variations in human genome

Genetic variants in humans are classified as polymorphisms, rare variants and mutations. Single nucleotide polymorphism (SNP) detected 1 in ev-

Correspondence: Peren H. KARAGİN, Department of Medical Genetics, Başkent University, Faculty of Medicine, Maltepe, Ankara, Turkey
Phone: +90 312 232 44 00/302 • E-mail: perenbaglan@yahoo.com

ery 1000 base pair, which means that there are 3.000.000 base pair differences between two individuals genome (1).

Translocation, deletion, insertion and inversion are also seen in human genome. When a chromosomal segment is moved from one position to another, either within the same chromosome or to another chromosome is called translocation. Deletions are the loss or absence of one or more nucleotides from a chromosome. Inversion is a chromosome rearrangement in which a segment of a chromosome is removed, turned upside down and reinserted back (2).

Recently, many projects have been done on human genomic variations to determine their functions and types. Most important and biggest projects were, Human Genome Project (HGP), HapMap Project, 1000 Genome Project and ENCODE Project (3-6,).

HUMAN GENOME PROJECT

United States Department of Energy (DOE) and the National Institutes of Health (NIH) conducted this project for 13 years (1990-2003) (7). Twenty institutes from six different countries (China, France, Germany, Japan, United Kingdom and USA) took part and totally three billion budgets were spent for the project.

The goals of the project were as follows:

1. To identify all the genes in the human DNA,
2. To know the function of genes,
3. To store this information in databases,
4. Developing tools for data analysis,
5. Related technology transfer to the private sector.

The euchromatic regions (90% of the whole genome) of the human genome sequenced at the end of the project (3). New information and technologies brought by the Human Genome Project shed light to the following international genome projects.

HapMap PROJECT

HapMap project began with a meeting on October 2002. It was planned to take 3 years. The project aimed to compare the difference in the sequence of genes between individuals and identify to chromosomal regions of the shared genetic variants (5). Therefore, this information would help people to discover the genes and response to therapeutic drugs.

The DNA samples for the HapMap come from a total of 270 people. Four populations were selected: Nigeria, Tokyo, Beijing and the northern and western European ancestry.

One of the many benefits of this project was to discover the link between our genes and common diseases such as depression, paralysis and asthma. Disease and the drug's effects can be customized according to the patient's genes and that provides resistance against to disease (5).

1000 GENOME PROJECT

The aim of the 1000 genome project was to determine the DNA polymorphism in different populations. This project consisted DNA sequence analysis from 26 different populations and 2.504 individuals. Identification of geographical and functional spectrum of human genetic variation will help us understand the effects on the patient's genetic (6).

As a result, 38 million SNPs, 1.4 million short indels and more than 14.000 verified for large deletions are revealed in haplotype map. Individuals from different populations have different genetic profiles which includes rare or frequent variants.

Studies have been show that, evolutionary conservation and exon regions are key elements to determine the strength of selection. In this context it showed rare variant load changes significantly between the biological pathways involved in the non-coding region of each individual transcription factors such as changes disrupt conserved binding

motifs were identified as being in hundreds of variants (6).

ENCODE (Encyclopedia of DNA Elements) PROJECT

The euchromatic regions (90% of the whole genome) of the human genome sequenced in 2003 by the human genome project. However studies have been conducted to show that only 1% of the human genes encoded a protein. Remaining 99% parts of the DNA (base sequences) is called as 'Junk DNA'. Scientist have been thought to cause of this ambiguity is closely connected with junk DNA which masked in 99% parts of the DNA. Therefore, ENCODE project is designed for discovered to these mysteries. The aim of the project was to detect the functional DNA sequences and to investigate the effect of the genome packaging on expression and regulation of the genes.

ENCODE project revealed that 80% of the genome have biochemical function. It was noticed that, non-coding regions of human genome have important functions. However, many embodiments appear to arise largely in this manner. To term "Junk DNA" has been disappeared in this way (4).

WHAT IS THE IMPORTANCE OF GENETIC DIVERSITY IN PEOPLE?

Only a small percentage of DNA in the human genome is coding sequence (sequence which eventually converts to protein) or regulatory sequence (sequence which can affect the level, timing and tissue specificity of gene expression).

In some mutations (for example neutral mutation) change the amino acid sequence of the protein but it does not account detectable change in its function. Other variations (for example silent mutations) do not change the amino acid sequences, gene products and also gene product's function.

Some genetic variations may have positive effect and provide advantages to the changing environmental conditions. For example: heterozygous form of the sickle-cell mutation causes resistance to malaria in areas where the disease is common.

Another example to positive effected mutation is mutation in the *CCR5 gene*. This *gene* encodes the protein on the surface of immune cells in humans. Human indeficiency virus (HIV) infects the immune cells by binding to *CCR5 gene* product protein and other proteins in the cell surface. Mutation in the *CCR5 gene* changes the level of expression or structure of the gene product protein may inhibit the HIV infection.

Some genetic variations are associated with diseases. Classic single-gene disorders are sickle cell disease, cystic fibrosis, Huntington's disease o and Duchenne muscular dystrophy (8-10). Whereas, some genetic variations are associated with common diseases. These diseases cause to morbidity and mortality in developed countries (for example heart disease, cancer, diabetes, schizophrenia and manic depression). However, more common diseases are caused by the interaction of many genes and environmental variables. Such diseases are called polygenic and multifactorial. In fact, the vast majority of human traits and diseases are multifactorial such as Parkinson's disease, Alzheimer's disease or stroke.

HOW GENETIC DIFFERENCES SPREAD WITHIN and BETWEEN THE POPULATIONS

Human population could be identified as geographic, political, linguistic, religious or ethnic boundaries. Numerous studies have demonstrated that, genetic differences could be observed in 90% of the population and only 10% of these differences separated the populations from the others. People have also been adapted to share of their own DNA, for this reason genetic boundaries are typically undetermined. DNA sequence of two people from different continent may resemble as genetically. In general, humans are resembled as biologically at the DNA level (11).

VARIATIONS DETECTED IN A TYPICAL HUMAN GENOME

Individuals vary greatly in a wide range of biological functions because of the variation among their genomes. Any individual genome will contain the following (12):

- ≈5-10 million SNPs (varies by population)
- 25.000-50.000 rare variants (private mutations or seen previously in < 0.5% of individuals tested)
- ≈75 new base pair mutations not detected in parental genomes
- 3-7 new CNVs involving ≈500 kb of DNA
- 200.000-500.000 indels (150 bp) (varies by population)
- 500-1.000 deletions 145 kb, overlapping ≈200 genes
- ≈150 inframe indels
- ≈200-250 shifts in reading frame
- 10.000-12.000 synonymous SNPs
- 8.000-11.000 nonsynonymous SNPs in 4.000-5.000 genes
- 175-500 rare nonsynonymous variants
- 1 new nonsynonymous mutation
- ≈100 premature stop codons
- 40-50 splice site disrupting variants
- 250-300 genes with likely loss of function variants
- ≈25 genes predicted to be completely inactivated

ETHIC

About the benefits and harm of understanding human genetic variation at the molecular level is still unknown. This information has a great potential to improve the human health. The human genome project architectures realized that the growing ethical, legal, they social problems. So in 1990 they set up the Ethical, Legal and Social Implications (ELS) program. This program works on critical and basic analysis of the sciences with the philosophical disciplines of pure and applied ethics. This program is focused on people, especially on educators. People have increased their understanding of the ongoing debate about the genetic and contribute.

As a result, science plays a significant role in making the right choice about people's personal health and public health as ethics. Ethics provides clarify basic structure for our values and our understanding of us (1).

CONCLUSION

Benefits of understanding human genetic variations are its value for promoting human health and combating disease. It will also contribute to the rapidly growing field known a pharmacogenomics and help to develop personalized medicine. In the future, physicians will use genetic test before the prescription so that the safest and most effective drugs and dosages can be prescribed.

REFERENCES

1. American Society of Human Genetics, American College of Medical Genetics. Points to consider: Ethical, legal, and psychological implications of genetic testing in children and adolescents. Journal of Human Genetics. 1995; 57:1233–1241.
2. Griffiths AJF, Gelbart WM, Miller JH, et al. Modern Genetic Analysis. New York: W. H. Freeman; 1999. Chromosomal Rearrangements. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21367/
3. The Human Genome Project FAQ". Genoscope. Centre National de Séquençage. http://www.genoscope.cns.fr/spip/The-Human-Genome-Project.html?artsuite=1
4. Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., & Barnes, I. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome research, 22(9), 1760-1774.

5. Gibbs, R. A., Belmont, J. W., Hardenbol, P., Willis, T. D., Yu, F., Yang, H., & Tam, P. K. H. (2003). The international HapMap project. Nature, 426(6968), 789-796.

6. 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature, 491(7422), 56-65.

7. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. Human Genome Project: 1998–2003. Science, 282(5389), 682-689.

8. Peccate C, Mollard A, Le Hir M, Julien L, McClorey G, Jarmin S, Le Heron A,Dickson G, Benkhelifa-Ziyyat S, Piétri-Rouxel F, Wood MJ, Voit T, Lorain S.Antisense pre-treatment increases gene therapy efficacy in dystrophic muscles.Hum Mol Genet. 2016 Jul 4. pii: ddw201. [Epub ahead of print].

9. Lettre G, Bauer DE. Fetal hemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. Lancet. 2016 Jun 18;387(10037):2554-64.

10. Edmondson C, Davies JC. Current and future treatment options for cystic fibrosis lung disease: latest evidence and clinical implications. Ther Adv Chronic Dis. 2016 May;7(3):170-83.

11. Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., & Hunt, S. E. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409(6822), 928-933.

12. Robert L. Nussbaum MD, FACP, FACMG, Roderick R. McInnes CM, MD, PhD, FRS(C), FCAHS, FCCMG and Huntington F. Willard PhD Thompson & Thompson Genetics in Medicine, Chapter 4, 4356

13. Saltus R. Tailor-made drugs. The Boston Globe. 1998 April 20

14. Bartlett, J. M. S.; Stirling, D. (2003). "A Short History of the Polymerase Chain Reaction".PCR Protocols. Methods in Molecular Biology 226 (2nd ed.). pp. 3–6.

15. Sillon, G., Joly, Y., Feldman, S., & Avard, D. (2008). Ethical and Legal Overview of Pharmacogenomics: Perspectives and Issues, An. Med. & L., 27,

16. Maher, B., "The Human Encyclopedia", Nature, 489, 7414, s.46-48, 2012.

17. Steinberg MH, Forget BG, Higgs DR, Nagel RL. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. Cambridge University Press; Cambridge: 2001.