

# Genomic medicine

Saied Jaradat, Ph.D. and Wajih Owais, Ph.D.  
 Biotechnology Center/ Jordan University of Science & Technology

Fifty years after the discovery of double helical structure of DNA by Watson and Crick, the Human genome Project consortium published its working draft of the human genome sequence map (1).



Dr. wajeeh owees



Dr. saeed jaradat

The order of the 3.2 billion chemical nucleotide bases represented by the letters;

- ◆ A (Adenine)
- ◆ T (thymine)
- ◆ G (guanine)
- ◆ C (cytosine)

Which makes up our genome and determined what we are, how we look, the diseases may be predisposed and the drug we might not respond to, is in the public domain and available for downloading (<http://www.ncbi.nlm.nih.gov/genbank>). The order of almost 99.9% of the 3.2 billion bases is exactly the same in all people. Considering there are 3,2 billion letters in the DNA code in every cell and there are 100 trillion (100,000,000,000,000) cells in the body, this suggests if all the DNA in the human body put end to end it would be enough to reach the sun and back over 600 times.

In contrast to the previous estimation of the genes number in the human genome (120,000); The human genome sequence unveiled that the gene number is 4 times lower (~30,000 genes) and just two times higher than the genes number in lower complexity organism such as worms and flies and is equal to the genes number in a mouse (1,2). (Table. 1; Fig.1).

The 30,000 human genes split into 24 chromosomes and distributed unevenly across the human genome. The nucleus of most human cells contains two sets of chromosomes, one set is given by each parent. Each set has 22 autosomes and an X or Y sex chromosome. Gene makes up less than 2% of the human genome. Majority of the 3.2 Billion nucleotide bases (97%) have no function and historically named "Junk DNA". Most of this "Junk DNA" is repetitive sequences and thought to have no

Organism	Estimated genome size/bases	Estimated gene Number	Chromosome number
Human	3,200,000,000	30,000	46
Chimp	3,100,000,000	30,000	46
Mouse	2,500,000,000	30,000	44
Fruit Fly	180,000,000	13,000	8
Nematode	100,000,000	19,000	6
Yeast	12,000,000	6,000	16
<i>E. Coli</i>	5,000,000	3,200	1
<i>H.influenzae</i>	1,800,000	1,700	1
<i>A. thaliana</i>	100,000,000	25,000	5

Table 1 - comparative genome sizes

direct functions. Chromosome 1 has the highest number of genes (2968 genes) while chromosome Y has the fewest number of genes (231 genes); <http://www.ncbi.nlm.nih.gov/genome/guide/human>). Human chromosomes range in lengths from 46,976,537 base pairs (chromosome 21) to 245,203,898 base pairs (chromosome 1). There is no correlation between the size of the human chromosome and the number of genes located in these chromosomes, for example, chromosome 22 contains 4 time less DNA size (49,476,972 base pairs) than chromosome 4 (191,610,523 base pairs), however chromosome 22 is a gene dense chromosome compare to chromosome 4. Additionally, the gene distribution varies within the same chromosome; a dense GC nucleotide region on the chromosome has more genes than an AT nucleotide rich region. The average gene size consists of 3000 bases, but the sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases and the smallest for tRNA genes at 500 bases.

**Comparative Genomics and medicine**

Numerous genomes have been sequenced including the mouse *Mus musculus*, the fruit fly *Drosophila melanogaster*, the worm *Caenorhabditis*

*elegans*, over 65 microbial genome, include the bacterium *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, the plant *Arabidopsis thaliana* and recently our closest relative the chimpanzee *Pan troglodytes* (<http://www.ncbi.nlm.nih.gov/genbank>). The tremendous amount of DNA sequences from different species lead to the development of new genomic science discipline namely “comparative genomics”. Comparative genomics is the analysis and comparison of genomes in different species (3). We gain a better understanding about the functions of so many human genes in health and disease by examining their counterparts in simpler model organism such as mouse. Mice and humans have the same number of genes, 90% of genes associated with disease are identical in human and mouse. Additionally, the similarities between mouse and human genes at the DNA level reached on average to 85%. Knock-out mouse (altered a given gene function in the germline of mice by an insertion of foreign genetics material

**On average, there is at least one base pair (one SNP) different between every two unrelated individuals**

into their DNA), contributed to better understanding of 100’s of human genes function at both the molecular and biochemical levels. For example two novel tumor suppressor loci has been identified in bone tumor (Osteosarcoma) patients by comparative study between human and mouse (4). Recently and after depositing the genomic sequence draft of the chimp genome in the gene bank, comparative genetics instead of looking for similarities between the human sequence and the chimp, they start looking at the dif-

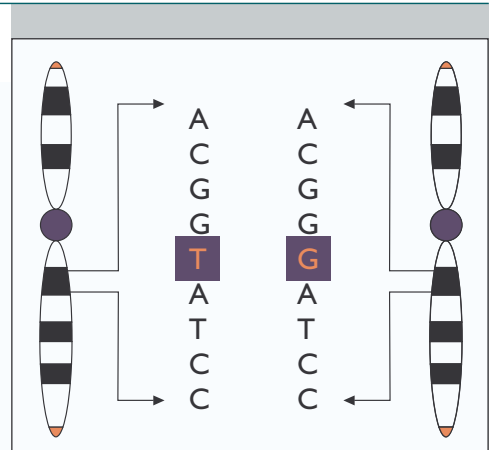


Figure 2 - SNPs Human Genetic Variation

ferences. By pinpointing the genomic regions where the chimpanzee DNA sequence differs from that of humans, we will be able to discover which part of the genetics code give chimps their increase resistance to some diseases such as AIDS, Malaria and Alzheimer. The absent of the 2 genes which encodes for the enzyme that adds the sialic acid (5) (type of acidic sugar) to protein and the gene encode for the cell-surface receptor that recognizes the sialic acid (6) from human genome and their present in chimp genome might influence the ability of certain viruses to infect human, as opposed to chimp cells. Chimpanzees are so similar to humans that veterinarians often refer to human medical textbooks when treating them. Chimp genome is 99% identical to the human yet no one would mistake a human for a chimp. Soon comparative genomics between chimp and human will provide us with a list of humanness genes (what make humans human). Recently, identification for a mutation in FOXP2 gene (7) in a family with a speech problem, suggested that FOXP2 might be important in speech development in human. Interestingly, the same gene exists in chimps, however, it is significantly different. Finally, *Drosophila* is another model organism, which provided an insight into our standing of human genes pathways that are mutated in human diseases. In a set of 289 genes implicated in human diseases, 177 are closely similar to fruit fly genes (8), including genes that play a role in cancer such as menin which

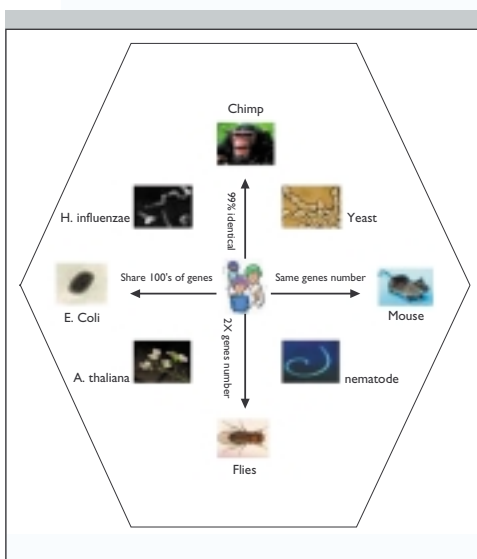


Figure 1



Figure 3 - A SNP profile can be used to stratify patients and differentiated between different ethnicity

causes multiple endocrine neoplasia type 1, neurological genes such as Notch which cause CADASIL syndrome and other genes implicated in human renal diseases (8).

**Pharmacogenetics**

The greatest immediate benefit of the human genome sequence for medicine will derive from sequence differences rather than consensus. Single nucleotide polymorphism (SNP; pronounced "snip") is a single-base difference in the DNA sequence that can be observed between individuals in the population. SNP variation occurs when a single nucleotide, such as a T, replaces one of the other three nucleotide letters-A, G, or C (Fig. 2). SNPs occur in human population at a frequency = or > than 1% of the time for the least common allele, in contrast to mutation which occurs in less than 1% of the population. On average, there is at least one base pair (one SNP) different between every two unrelated individuals. Given the human genome size is more than 3 billion base pair and two unrelated people share 99.9% of their DNA sequence, this suggests that two unrelated humans vary in

their genomes. Additionally these SNPs are varied between different ethnicity (Fig. 3; <http://www.ncbi.nlm.nih.gov/SNP/>; <http://snp.cshl.org/>). Effort under way to characterize more SNPs and as of 3/3/2003 there is over 6,107,661 SNPs, which have been mapped and deposited

in the database. The biggest challenge is to catalogue these SNPs in relation to disease phenotype and their relevance to drug response.

Pharmacogenetics is the study of how genetic polymorphisms mainly SNPs influence the variability in patients' responses to drugs (9). Patient responses to drugs are variable and often unpredictable. A given drug might be effective in one patient but may not be effective in others and not only that some patients developed adverse reactions. Over 2 million hospitalizations and 100,000 deaths every year are due to adverse reaction to drugs. It is evident an individual-to-individual variability in drug response is a major clinical problem, and again because we vary in our DNA sequence, we vary in our response to drugs. Rapid advances in SNPs genotyping, Biocomputing processing, and advances in statistical methods can be applied now to characterize the individuals patients who suffer from an adverse event to drugs, or those for whom a drug shows efficacy and classify the individuals at risk (Fig. 3). By applying whole genome SNP linkage disequilibrium (non-random association between SNPs

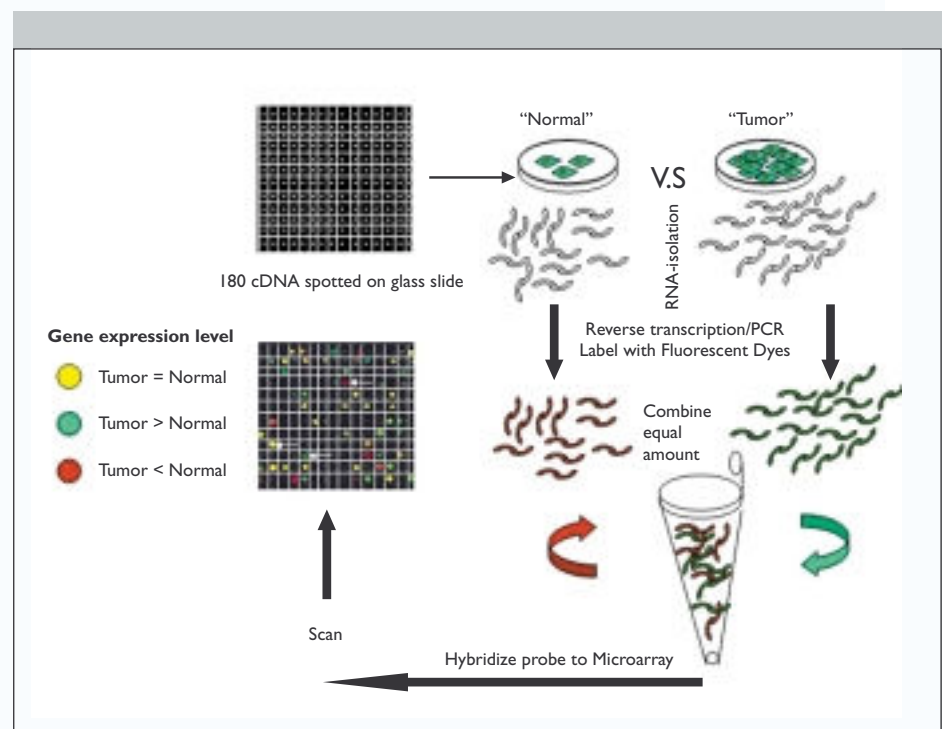


Figure 4 - Prepare cDNA microarray and probe

in proximity to each other) mapping the patients during phase II clinical trails of a given drug, may be possible to identify group of SNPs and their association with efficacy and common adverse event phenotypes (Fig. 3). The future potential of SNPs profiling is illustrated in Fig. 3, which depicts the hypothetical effect of 17 polymorphic SNPs in relation to hypothetical drug trail in phase II. For example, SNPs T, G and A at positions 5, 6, and 17, respectively, are associated with drug efficacy, while SNPs T and G at positions 10 and 12 are associated with adverse reactions to drugs. Identifications of such SNPs profile will enable more rapid and inexpensive screening of patients, who are likely to experience efficacy and common adverse event phenotype.

One of the most extensively studied genes in regard to Pharmacogenetics is TPMT which encode thiopurine methyltransferase. thiopurine methyltransferase is an important route in metabolism of thiopurine drugs. The three major thiopurine drugs are 6-mercaptopurine, 6-thioguanine and azathioprine which are used in treatment for lymphoblastic leukemia, acute myeloblastic leukemia and immunosuppressant patients. It is now well established that polymorphism in the TPMT gene is responsible for patients' variations in responding to the thiopurine drug (10). For example patients'

with a genotype wt/wt for the TPMT gene (produce 100% of the thiopurine methyltransferase protein) are at high risk of leukemia relapse, while a genotype mut/mut (produce 0% of the thiopurine methyltransferase protein) are at high risk of developing Myelosuppression and secondary cancer in addition to leukemia, and finally, patients with mut/wt genotype have a better chance to be cured from leukemia. An additional application of Pharmacogenetics in anti-cancer drug therapy is the metabolism of 5-Fluorouracil, an anti-solid tumors drugs. Breast and colorectal cancer patients with 2 tandem repeats (28 base pair of the DNA repeated twice) in the enhancer region of thymidylate synthase gene (11) has a better chance for recovery than patients with three tandem repeats do.

### Genomics Technology and medicine

As a result of the Human genome project, there has been a tremendous amount of information available about DNA sequence of the human genome. However, we are still lagging behind on how much we know about the functions of the 30,000 genes, what are the role of these genes in health and disease? How they interact with each other at the cellular level and when and where this interaction occurs? DNA Microarrays

(Fig. 4) is powerful new technologies hold the promise to answer some of these questions (12). The most attractive application of microarrays is the study of differential gene expression in disease. In other words, we can know by using this technology which genes is silent or active in any given type of cancer in comparison to normal as illustrated in fig. 4. For example we know the genes, which are spotted in column number 9, and row number 4 is down regulated in cancer cells compare to normal cells. In contrast to the gene spotted in row number 10 and column number 5, which showed an expression level higher in cancer cells. Additionally genes showed no expression differences between the cancer and the normal cell lines such as the gene spotted in row 9 and column 2. It is clear that genes showed an up-regulation or down-regulation in cancer cell can be tagged as genes of interest. These genes can make a diagnostics profile for such type of cancer.

### REFERENCES

1. International Human Genome Sequencing Consortium (2001). *Nature* 409: 860-921
2. Mouse Genome Sequencing Consortium (2002). *Nature* 420: 520-562
3. Furlong RF, Yang Z (2003). *Heredity* 91:533-4.
4. Nathrath MH, Kuosaitte V, Rosemann M, Kremer M, Poremba C, Wakana S, Yanagi M, Nathrath WB, Hofler H, Imai K, Atkinson MJ (2002). *Oncogene* 21: 5975-80.
5. Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A (1998). *Proc Natl Acad Sci U S A* 95: 11751-6.
6. Angata T, Varki NM, Varki A (2001). *J Biol Chem*. 276: 40282-7
7. Liegeois F, Baldeweg T, Connolly A, Gadian DG, Mishkin M, Vargha-Khadem F (2003). *Nat Neurosci*. 6: 1230-7.
8. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, et al (2000). *Science* 287: 2204-15.
9. Roses AD (2002). *Nat Rev Drug Discov* 1: 541-9.
10. Murphy LA and Atherton DJ. (2003). *Pediatr Dermatol*. 20: 531-4
11. Ma T, Zhu ZG, Ji YB, Zhang Y, Yu YY, Liu BY, Yin HR, Lin YZ (2004). *World J Gastroenterol* 10: 172-6
12. Churchill GA (2002). *Nat Genet* 32 Suppl: 490-5