Journal of Media and Information Warfare *Vol. 4, 93-104, 2011*

Securing DNA Information from Selective Attacks on Humans

Noor Elaiza A. Khalid[1,2], Siti Zaleha Zainal Abidin[1,3] & Izyan I. Kamsani[1]
[1]Faculty Computer and Mathematical Sciences
[2] Pharmacogenomics Centre (PROMISE)
[3] Centre for Media and Information Warfare Studies
Universiti Teknologi MARA

ABSTRACT

Research on human genome has evolved rapidly. In the past few decades, the understanding of deoxyribonucleic acid (DNA) has grown immensely towards wealth of information about human population diversity. This diversity trait can be grouped into ethnic, family and gender. Some of these DNA traits are susceptible to certain diseases, drugs or chemicals. Thus, it is critical to secure the DNA information especially the relationship between the DNA traits and its susceptibility. This paper discusses the importance of securing this susceptibility knowledge in order to provide awareness for people, agencies or organization that involve in managing genomic information. This awareness can alert the public on the possibilities of this kind of attack. Several guidelines are also proposed to safeguard DNA information from undesirable malicious attack.

Keywords: *Awareness, Deoxyribonucleic Acid, Genomic Traits, Human Attack, Information Security*

Introduction

Currently, the discovery of human genome sequence is exponentially being updated. Such valuable information contributes to the latest research in disease, mutation and human health related to genetic disease. Human genome provides the platform to analyze human traits especially susceptibility to disease, drugs or chemicals. In addition, knowledge can be harvested from deoxyribonucleic acid (DNA) helps to develop new

ISSN 1675-1302

ways to treat, cure, or even prevent thousands of diseases that afflict humankind. Since DNA shows the diversity of human traits, it can create various human populations that differ from one group to another. DNA contains sequence of genome information that provides an area of research where suitable drug can be matched to cure diseases associated to specific genomic type without any side effects (Buscemi and Tagliabracci, 2011). Furthermore, genome sequence contains instruction and blueprint of human. The instruction provides each cell with the role to play in human body. The blueprint map out the genetic factors associated with influence of drugs. This relationship leads to a new scientific discipline called pharmacogenomics.

Basically, genome can be defined as an organism's complete set of DNA (Genetic Home Reference, 2011). In humans, the entire genome is made up of more than 3 billion DNA base pairs in all cells that have a nucleus. Specifically a human genome is the combination of two genomics sources, a sperm cell (from the father) and an egg cell (from the mother). DNA is made up of four simple building blocks nucleotide bases namely Adenine (A), Cytosine (C), Thymine (T), Guanine (G); (A, C, T, G). The sizable combinations of these building blocks make it possible for each individual to be unique. Segments of DNA called genes contain specific instructions that produce other molecules called proteins. Proteins enable cell to perform various organ functions for human livelihood. Figure 1 shows a picture of DNA strand from many combination of A, C, T and G.

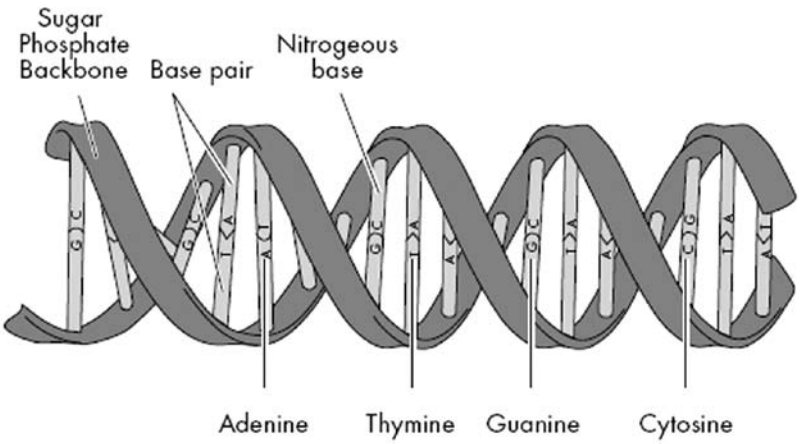

Figure 1: The DNA Helical Strand (Watson and Crick, 1953)

In this paper, we provide genomic information awareness by exploring the relationship between the DNA traits and how its susceptibility knowledge can affect our population. From the perspective of security, it is important to keep this information secure to avoid it from being misused. Susceptibility to a certain type of drugs is usually linked to a certain group of people. Using this knowledge, it is possible to selectively victimize a chosen group of people.

The remainder of this paper is organized as follows: a brief explanation on the evolution of genomic research is given in Section 2; Section 3 explains the annotation and drugs, while security factors are discussed in Section 4. Finally, Section 5 gives the concluding remarks.

Evolution of Human Genomic Research

Human genomic research begins with the discovery of DNA helical structure in 1953 by Kristine B. Stewart (2007). Between the year 1866-1883, Greggor Mandel (1865) proposed basic law of heredity based on pea plants. In the same period, Walter Fleming and Francis Galton (1882) discovered human race improvement by sub-dividing the tiny biological thread into chromosome. In the 20th century, other discoveries such as X-rays, DNA, genetic engineering, DNA fingerprinting, disease founding, cloning and genome project have flourished. In the year 2000, Genome project has advanced into sequencing human chromosome. Year 2003 sees the mapping of genes to deferent types of body functions. Recently, investigation on impacts of gene changes within the human genome has emerged exponentially.

Human evolutionary benefits scientists in determining the environmental and other factors (e.g. population structure, genetic drift, migration and changes in population size) that shape human genomic diversity (Tennesen et al., 2011). In addition, human genome contains all information about genetic variant including inherited genetic disease. In evolutionary theory, a majority of population evolution is caused by selective pressures acting as the catalyst for the earlier evolution of australopithecines to Homo sapiens and finally, to modern humans (Platter, 2009).

The evolutionary process of natural selection is divided into two parts which are positive and negative natural selection. The positive natural selection occurs when changes to the genes has positive impact to the human life and vice versa to the negative natural selection. Examples

of natural selection factors include diseases, body immune system and physical fitness. Changes to the genes play the main role in the evolution of new species (divergence of gene) as a new tactical alternative for survival (genetic variation). Genetic variants can be associated to disease susceptibility, disease progression and variability in drug response.

Annotation and Drugs Responses

The advancement of DNA sequencing machine has introduced vast amount of human genomic data. These genomic data varies from one human to another. The data variation can be caused by mutation, crossover, replication and others. In addition, the causes could be due to various hereditary diseases, surrounding contracted diseases or drugs ingested responses during their life time. The relationship between human genome, disease and drugs is depicted in Figure 2.

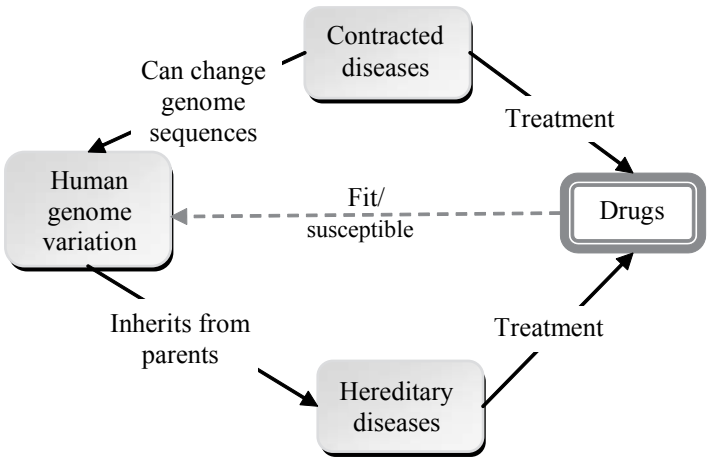

Figure 2: Conceptual Relation Diagram between Human Genome, Disease and Drugs

Large numbers of researchers focus on identifying specific location of DNA variant that causes different diseases or responses to different drugs (Francisco et al., 2002). The DNA variation can influence the way human is affected by specific drugs where the treatment can be successful or harmful to individuals. Therefore, different people may

have different illness and they need different drugs even which they have the same illness.

These issues have been investigated by Sir William Osler (*The Principles and Practice of Medicine, 1892*) where he discovered that patients will not give the same response to the same drugs. Additionally, patients may have adverse side effects due to their unique genetic make-up. Since DNA traits are susceptible to certain diseases, drugs or chemicals, this information will be used by other researchers to explore new avenue in preventing, analyzing and treating disease. Thus, the mapping of body system and drugs acceptance can be identified.

A drug can be defined as any material used in diagnosis to cure and treat diseases (Hussar, 2007). Predicting and detecting specific diseases relating to individual will encourage custom made designer-drugs. Thus, the relationship between drug prediction and disease detection can be constructed. This relationship can be classified into three; (a) drug-drug interaction (person taking more than one drug), (b) drug-nutrient interaction (drug consumes with supplement) and (c) drug-disease interaction (person having diseases taking drug). Unsuccessful treatment causes side effects and may be harmful to the person. People are unique and different from each other due to differential in genetic makeup. The divergence gene factor makes us react to different kind of drugs.

Human Genome Variation

Recently, studies in human genome variation have lead to the identification variability of species. Genes contain coded information that allows instructions to be carried out. The variation in genes occurs when there are changes in this information called mutation (Kristine B. Stewart, 2007). The role of these variations may cause or contribute to human development, growth and health problems. Figure 3 shows the type of variations that occurs in human genome.

Four common genetic variations in human genomes are (a) Copy Number Variants (CNVs), (b) Single Nucleotide Polymorphism (SNP), (c) Insertion and Deletion (InDel) and (d) Inversion. These variations may result in different human characteristics representing the individual phenotypes, genotypes, disease and blood type. Genotype is the whole set of genes inherited from the parents while phenotype is the combination of morphology and physiological traits. It represents the differentiation of individual from their own relatives.

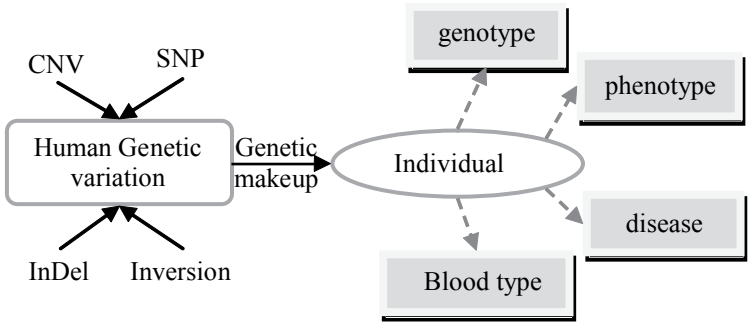

Figure 3: Annotation of Genetic Variation

Contracted and Hereditary Diseases

Contracted disease occurs due to environmental factor (also known as non-genetic factors). This factor contributes to disease risks and development disorders (Digitale, 2011). The factors include maternal infections, maternal diseases, drug, chemical and irradiation. Examples of contracted diseases are autism and Alzheimer that affects thought, memory and language.

Hereditary diseases are defective genes inherited from the parents. These are genetic diseases caused by abnormality in an individual's genome. Examples of hereditary diseases are diabetes, sickle cell anaemia and cancer.

Drugs Responses and Disease Mapping

The mapping of drug responses to diseases has advanced tremendously in recent years (Weinshilboum and Wang, 2004). Information on the mapping is kept in various distributed global databases and individual/organization's research publications. Furthermore, software development has evolved rapidly to allow fast information retrieval to perform the mapping. This mapping will group individuals with the same drugs responses or disease susceptibility. These groups of individuals usually belong to the same family, ethnic groups or gender. This information should be kept as confidential data. However, data confidentiality is less realistic. This is due to the fact that genomic information is difficult to obscure, therefore, threatening one's privacy (Buscemi and Tagliabracci, 2011), especially disease markers in drug response predictions. Thus, it is important to restrain any people with bad intentions to misuse the information that could adversely affect other people.

Security Factor

It is important to be aware of how the human genomic informations are kept and used. Genomic information consists of immune body system such as drug susceptibility, afflicted diseases and physical status of a person. Currently, there is a lack of protection and security measures to control the access of this information. In order to enrich the knowledge in genomic research, researchers are encouraged to have public data sharing. Vast increase in the number of data may lead to uncontrolled flow of information. The information is readily exposed at anyone's disposal.

The danger of this exposure has raised alarm among researchers. For example, safeguarding genetic identity (profile information) from public access (Weiser, 2002) and addressing the opinion of privacy issue on individual DNA data (Boyle, 2003) as follows:

- **Randy, Port Angeles, Wash**: "I would never hand over my DNA to any government or private data bank. I am very suspicious of any group that wants to track people. In the USA we are not required to carry ID of any kind, and I believe that any national ID or DNA databank is dangerous."
- **Robert**: "I have many privacy concerns with requiring a DNA profile for everyone. We have long been protected against unreasonable search and seizure. Requiring an individual to supply a DNA sample to police or other governmental authority without probable cause or reasonable suspicion is very troublesome."
- **Barrie, Los Angeles**: "I can't even imagine the privacy and security issues this would generate. DNA is filled with so many indicators of potential illness, etc., that I can foresee a great deal of misuse and abuse of this database. There would have to be stringent controls and absolute privacy, and even then, just because your DNA is present at a crime site doesn't mean you yourself were present at the time of the crime."

In addition, there are several possible threats listed as follows:

a. *Manipulation of DNA Information (Weiser, 2002)*
   With DNA sequence and its disease prediction knowledge, it is possible to modify information regarding potential diseases that a person might have.

b. *The relationship between DNA traits and its susceptibility to disease (Franciscoet al, 2002)*
The advancement of technology has speed up the mapping process of DNA traits and diseases. This may provide information about a person's susceptibility which can be dangerous if it falls into the hands of people who have bad intention.

c. *Use as weapon to attack victim (legal crime)(Robertson J.A., 2003)*
The advent of custom-made designers' drug may also lead to the development of harmful designers' drug. The information about harmful designers' drug can be used as a weapon to target specific victims.

d. *Misuse of DNA information*
Knowledge of DNA information can be misused by third parties for their own profit. For example, an insurance company may use the DNA information to decide on providing insurance coverage that is profitable to the company.

Robertson J.A. (2003) assures consumers and public that creating ethical and legal guidelines will minimize and prevent DNA's information being stolen and accessed by others. The suggested guidelines are divided into two; individuals and organizations.

Guidelines for individuals

a. Individuals are advice to provide their DNA information strictly to their doctors for health benefit.
b. Ensure only the authorized person can collect the DNA information.
c. Individuals should have awareness that their DNA sample could expose information about their health.

Guidelines for Organisations / Agencies:

1. DNA information database should limit access to authorized personal only.
2. Every organization dealing with DNA information should formulate policies and rules for managing DNA information.
3. The software developers who are involved in DNA information system development should include the mechanisms for protection and security.

4. Any DNA information transfer must be accompanied by a consent agreement between the organizations and individuals.

Discussion and Conclusion

DNA discovery has invoked issues on privacy and human rights. Protecting DNA information has become necessary due to the following factors:

a. Since DNA information is stored in distributed databases where data are easily accessible, DNA data could readily be modified or misused.
b. DNA data could be used to impersonate an individual to gain benefit.
c. Drugs and DNA information are closely mapped; thus, contain many health and genetic relationships between human.

These issues threaten the privacy protection rights. People should be aware of selective attack that could occur to human if this information is not well protected and secured. In the near future, genes to drug mapping will be extensively being researched in the area of pharmacogenomics. Present researchers are moving towards revolutionizing designer drug administration. The success of these discoveries may have its pros and cons. The proposed guidelines could help secure the DNA information.

References

Alex L. (2011). *The Future of Human Evolution*. Retrieved on 15/10/11 from http://alexlod.com/2011/01/14/the-future-of-human-evolution/

Anthony J. Brookes. (1999). The Essence of SNPs. *The International Journal on Genes and Genomes* p. 177-186.

Boyle, Alan (2003). "*The Great DNA Debate: Do you think everyone's DNA fingerprints should be on file somewhere*?", Cosmic Log Archive, MSNBC http://groups.msn.com/AlanBoylesCosmicLog/42003archive.msnw?action=get_message&mview=0&ID_Message=692&LastModified=4675420716692968809

Branden E. Platter. (2009). Evidence of Cotemporary Modern Human Evolution Contained within the Human Genome. *Lethbridge Undergraduate Research Journal*, 4(1).

Buscemi, L. & Tagliabracci, A. (2011). *Pharmacogenetics Role in Forensic Sciences*. Intech.

Collins, F. S., E. D. Green, A. E. Guttmacher, and M. S. Guyer. (2003). A vision for the future of genomics research, *Nature*, 422: 835-847.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C. & Campbell, P. (2009). Origins and functional impact of copy number variation in the human genome. *Nature*, 464: 704-712.

Daniel A. Hussar (2007). *Drugs Interaction*. Retrieved on 15/10/11 from http://www.merckmanuals.com/home/drugs/factors_affecting_response_to_drugs/drug_interactions.html

Dennis K., Kinney, Daniel H. Barch, Bogdan Chayka, Siena Napoleon & Kerim M. Munir (2010). Environmental risk factors for autism: Do they help cause de novo genetic mutations that contribute to the disorder. *Medical Hypotheses*, 74: 102-10.

Digitale E. (2011). *Non-genetic factors play surprisingly large role in determining autism, says study by group*. Retrieved on 30/10/11 from http://med.stanford.edu/ism/2011/july/autism.html.

Francisco, M.D.L.V., Issac S.K., Julie, A. S. & J. Claiborne S. (2002). Human Genome Variation: Drug Response and Clinical Phenotype. *Pacific Symposium*, 7: 3-5.

Freedman, Scott & Herron, Jon C. (2006). *Evolutionary Analysis: Fourth Edition*. New Jersey; Prentice Hall.

Guttmacher, A. E. & Collins, F. S. (2002). Genomic medicine–a primer. *N Eng J Med*, 347: 1512-1520.

Jacob A. Tennessen, Timothy, D., O'Connor, Michael, J. Bamshad & Joshua M. Akey (2011). *The Promise and Limitation of Population Exomics for Human Evolution Studies*. Genome Biology.

Kristine B. Stewart (2007). *The Human Genetic Code - The Human Genome Project and Beyond*, The Australian Genetics Resource Book.

Lagerstedt, K. et al. (1997). Double-strand breaks may initiate the inversion mutation causing the Hunter syndrome, *Human Molecular Genetics*, 6(4).

Lee, C. & Scherer, S.W. (2010). The clinical context of copy number variation in the human genome. *Expert Rev Mol Med*, 12: e8.

Maiti, S., Kumar, K. H. B. G., Castellani, C. A., O'Reilly R. & Singh, S. M. (2011). Ontogenetic De Novo Copy Number Variations (CNVs) as a Source of Genetic Individuality: Studies on Two Families with MZD Twins for Schizophrenia, *Plos One*, vol. 6, Issue.

Nielsan, R., Hellman, I., Hubisz, M., Bustamante, C. & Clark, A.G. (2007). Recent and Ongoing Selection in the Human Genome, *Nature Reviews Genetics*, vol. 8.

Robertson, J.A. (2003). The $1000 Genome: Ethical and Legal Issues in Whole Genome Sequencing of Individuals, *The American Journal of Bioethics*, 3(3): InFocus.

Ryan E. Mills, Stephen Pittard, W., Julienne M. Mullaney, et al. (2011). *Natural genetic variation caused by small insertions and deletions in the human genome*, Cold Spring Harbor Laboratory Press.

Sir William Osler. (1892). *The Principles and Practice of Medicine*.

Stankiewicz, P., & Lupski, J.R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med*, 61: 437-455.

The Revolution in Medicine retrieved on 8/10/11 from http://www.answers.com/topic/genetics#ixzz1afgOV4VN

Toland, A. E. (2011). *DNA Mutations*. Retrieved on 30/10/11 from http://www.geneticheath.com/G101_Changes_in_DNA.shtml

Wain, L.V., Armour, J.A.L. & Tobin, M.D. (2009). Genomic copy number variation, human health, and disease. *The Lancet*, 374: 340-350.

Watson, J.D. & Crick, F.H.C. (1953). Molecular structure of nucleic acids: A structure for deoxyribonuceliec acid. *Nature*, 171, 737.

Weinshilboum, R. & Wang, L. (2004). Pharmacogenomics: bench to bedside. *Nature Reviews Drug Discovery*, 3: 739-748.

Weiser, B. (2002). *Identity ring said to victimize 30,000*. New York Times Nov. 26, 2002, p. A1.

William C. Shiel Jr. (n.d.). *Sickle Cell Disease (Sickle Cell Anemia)*. Retrieved on 30/10/11 from http://www.medicinenet.com/sickle_cell/article.htm#1whatis

Your Genome (2011). *What is Genome*. Retrieved on 8/10/11 from http://www.yourgenome.org/dgg/general/genomes/genomes_1.shtml