

# Chapter 17

## Genetic Data Analysis

**M. Shamila**

 <https://orcid.org/0000-0001-9105-9531>

*Malla Reddy Engineering College, India*

**Amit Kumar Tyagi**

 <https://orcid.org/0000-0003-2657-8700>

*Vellore Institute of Technology, Chennai, India*

### ABSTRACT

*Genome-wide association studies (GWAS) or genetic data analysis is used to discover common genetic factors which influence the health of human beings and become a part of a disease. The concept of using genomics has increased in recent years, especially in e-healthcare. Today there is huge improvement required in this field or genomics. Note that the terms genomics and genetics are not similar terms here. Basically, the human genome is made up of DNA, which consists of four different chemical building blocks (called bases and abbreviated A, T, C, and G). Based on this, we differentiate each and every human being living on earth. The term ‘genetics’ originated from the Greek word ‘genetikos’. It means ‘origin’. In simple terms, genetics can be defined as a branch of biology, which deals with the study of the functionalities and composition of a single gene in an organism. There are mainly three branches of genetics, which include classical genetics, molecular genetics, and population genetics.*

### INTRODUCTION

Genetics (WHO, 2002) is the investigation of genes, genetic variety and heredity in living organism. Genetics is the basic of heredity. Heredity is the transmission of genetically based characteristics from ancestor to descendant. Any form of heritable feature is known as a character. Some examples for this type of characters are eye color, hair color, height etc. DNA (deoxyribo nucleic acid) is the main genetic substance of all living organism. DNA of Human being consists of 23 pairs of chromosomes. For example, chimpanzees consist of 24 pair of chromosomes. This means only 2% difference between human and chimpanzee in DNA structure. That is the main reason because humans are said to genetically evolved from ape like creature especially chimpanzee. Genetic studies were carried out since the

DOI: 10.4018/978-1-7998-2742-9.ch017

classical era where the scientist like Aristotle and Hypocrites put forward some theories like how the parental characteristics are passed to the offspring. However, breakthrough was achieved in 19<sup>th</sup> century when Gregor Mendal experiments with pea plants and discovered the fundamental laws of inheritance.

In cells gene is part of DNA. Genes can be defined as instruction manual to create a living being. Every gene has two alternative forms. And this alternative form is known as alleles. The four nitrogen bases present in the DNA are adenine (A), thymine (T), guanine (G) and cytosine (C). A pair with T and C with G. Thus, in the double stranded form, each strand contains all the necessary information, redundant with its partner strand. Genes are arranged linearly along the DNA sequence. Human DNA consists of 3billion bases. According to the U.S. National Library of Medicine (NLM) more than 99 percent of these bases are identical in all people.

Genetic analysis is the general procedure of contemplating and looking into in fields of science that include hereditary qualities and molecular biology. There are various applications that are created from this exploration, and these are likewise viewed as parts of the procedure. Genetic analysis can be utilized for the most part to depict strategies both utilized in and coming about because of the sciences of hereditary qualities and molecular biology, or to applications coming about because of this research. The process of genetic analysis began in the primitive days itself. Early people found that they could use selective breeding to improve yields and creatures. They additionally distinguished acquired qualities in people that were dispensed with throughout the years. The numerous hereditary investigations step by step advanced after some time.

As we mentioned above genetics is the study of hereditary, which emphasis on the study of limited number of genes with specific function whereas genomics (WHO, 2002; WHA, 2004)is a new term which focus on study of whole organism gene set called as genome. Genomics become popular in the last couple of decades due to the advancement of technologies. But most of the time these two terms are often used interchangeably. We can consider genetics as a subset of genomics. The table 1 shown below give insight to distinction between these two terms.

## Algorithm

Genetic algorithm: Genetic algorithm works based on the principle of genetics and natural science. It is commonly used to solve optimization problems, in research, and in machine learning.

*Table 1. Difference between Genetics and Genomics*

Genetics	Genomics
It focuses on study of functionalities of a single gene	It includes study of whole gene and the inter relationships among them
Gene refers to particular DNA sequence of a single chromosome	Genome is complete hereditary information of an organism
Father of genetics is Gregor Mendal	Introduced by Tom Roderick
Biochemistry and biology help to explore this field	Can be explored with the help of bioinformatics and molecular biology
It emphasis on single gene behaviour	It emphasis entire genome of an organism

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/genetic-data-analysis/263327](http://www.igi-global.com/chapter/genetic-data-analysis/263327)

## Related Content

---

### Autoencoder Based Anomaly Detection for SCADA Networks

Sajid Nazir, Shushma Patel and Dilip Patel (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 83-99).

[www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436](http://www.irma-international.org/article/autoencoder-based-anomaly-detection-for-scada-networks/277436)

### Churn Prediction in a Pay-TV Company via Data Classification

Ilayda Ulku, Fadime Uney Yuksektepe, Ozgur Yilmaz, Merve Ulku Aktas and Nergiz Akbalik (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 39-53).

[www.irma-international.org/article/churn-prediction-in-a-pay-tv-company-via-data-classification/266495](http://www.irma-international.org/article/churn-prediction-in-a-pay-tv-company-via-data-classification/266495)

### Binary Search Approach for Largest Cascade Capacity of Complex Networks

Natarajan Meghanathan (2023). *Encyclopedia of Data Science and Machine Learning* (pp. 2500-2513).

[www.irma-international.org/chapter/binary-search-approach-for-largest-cascade-capacity-of-complex-networks/317690](http://www.irma-international.org/chapter/binary-search-approach-for-largest-cascade-capacity-of-complex-networks/317690)

### Secure Chaotic Image Encryption Based on Multi-Point Row-Column-Crossover Operation

K. Abhimanyu Kumar Patro, Mukesh Drolia, Akash Deep Yadav and Bibhudendra Acharya (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science* (pp. 286-328).

[www.irma-international.org/chapter/secure-chaotic-image-encryption-based-on-multi-point-row-column-crossover-operation/268762](http://www.irma-international.org/chapter/secure-chaotic-image-encryption-based-on-multi-point-row-column-crossover-operation/268762)

### An Integrated Process for Verifying Deep Learning Classifiers Using Dataset Dissimilarity Measures

Darryl Hond, Hamid Asgari, Daniel Jeffery and Mike Newman (2021). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-21).

[www.irma-international.org/article/an-integrated-process-for-verifying-deep-learning-classifiers-using-dataset-dissimilarity-measures/289536](http://www.irma-international.org/article/an-integrated-process-for-verifying-deep-learning-classifiers-using-dataset-dissimilarity-measures/289536)