In this section of *Resonance*, we invite readers to pose questions likely to be raised in a classroom situation. We may suggest strategies for dealing with them, or invite responses, or both. “Classroom” is equally a forum for raising broader issues and sharing personal experiences and viewpoints on matters related to teaching and learning science.

**Genotype to Phenotype: Through ‘OMICS’: An Overview**

Wilhelm Johannsen (*Figure 1*) (1909) introduced two words to the genetics dictionary, namely ‘phenotype’ and ‘genotype’. ‘phenotype’ is the appearance of an organism while ‘genotype’ is the genetic constitution as revealed by breeding experiments. By taking cognizance of research findings since 1909, an overview of the evolution of our understanding of the relationship between genotype and phenotype is presented here.

**Foundation laid by Gregor Mendel**

Gregor Mendel, (*Figure 1*) through his systematic breeding experiments with pea plants, *Pisum sativum*, was able to deduce the fundamental rules of inheritance. He analysed the pattern of inheritance of seven pairs of contrasting characters or phenotypes. They were: round vs wrinkled seeds, yellow vs green endosperm, colored vs white flowers, inflated vs constricted seed pods, green

*Vol.27, No.11, DOI: https://doi.org/10.1007/s12045-022-1491-2

*On the occasion of the birth bicentenary of G. J. Mendel; Father of Genetics, this article is dedicated to the memory of my teachers Prof. M.R. Rajasekarasetty and Prof. L. Siddaveere Gowda, who taught me Genetics.*
vs yellow seed pods, long vs short plants, and axial vs terminal flower position. He published his results and postulates during 1865–1866. He assumed that for every character each parent in a cross between plants showing such contrasting characters would contribute a determinant called ‘factor’ and the progeny, which will carry two factors for every character, having received one from each parent. The character of the progeny resembled that of one of the parents, the ‘factor’ which was responsible for this was called ‘dominant’ over the other factor which did not manifest itself. The other factor which did not manifest in the hybrid progeny was called recessive. This led him to develop the ‘law of dominance’. At the time of reproduction, these dominant and recessive factors segregate, and are passed on to the next generation through gametes, which carry only one factor of the pair. Among the F₂ progeny Mendel could recover majority of the plants with character that was expressed in F₁ (dominant) and a few which resembled the character of the parent which was not expressed in the F₁ (recessive). In experiments, studying the inheritance of one character at a time (monohybrid), the F₂ plants with dominant and recessive characters appeared in a ratio of 3:1, respectively. This was the basis for the ‘law of segregation’. These results implied that, although not visible, both recessive and dominant factors coexist within the F₁ progeny. An important observation of Mendel was that during this phase of staying together in the hybrid progeny, factors coming from two different parents do not blend/mix/modify each another. Each of these factors would retain their original potential/information to determine a character, that is, ‘purity of the factors’ remains unchanged.

*The next question was whether the segregation of a factor responsible for a particular character (phenotype) has any influence on segregation of the factor determining another character.*

As an extension of these experiments, Mendel studied the pattern of inheritance of two different pairs of contrasting characters at a time. As expected on the basis of law of dominance, the F₁ progeny expressed only the dominant characters of each of the
two pairs of contrasting characters. although both the dominant and recessive factors of each pair were present in the progeny, which were hybrids for both pairs of characters (since two pairs of characters are involved in such crosses, they are called dihybrid crosses). When F₁ breeds, these factors would segregate during the formation of gametes, and again come together by the fusion of gametes coming from male and female parents (fertilization). In the zygote and in the individual that developed, the dominant and recessive factors for each character were brought together. In this experiment, ‘two different characters’ were involved, and so, Mendel was studying ‘two different pairs of factors’ simultaneously. Mendel was able to resolve this issue, by looking at the phenotypic ratios of the F₂ progeny. Among F₂ he observed a ratio of 9 (plants with both the dominant characters): 3 (plants with one dominant and are recessive character): 3 (plants with alternate combination of dominant and recessive character) and 1 (plants with both the recessive characters) (9:3:3:1). From this Mendel came to the conclusion, that during segregation, each pair of factors behaves independently since segregation of one pair has no influence on segregation the other pair of factors. Based on these results, he formulated the law of ‘independent assortment’ of factors. These observed ratios (3:1 and 9:3:3:1) were in agreement with the expectations of statistical theory of probability: Product Law of Probability: “The probability of two independent events occurring simultaneously is equal to the Product
of their separate probabilities”, and the Sum Law of Probability: “The probability of either one or the other of two mutually exclusive events is the Sum of their individual probabilities.”

For over three decades, Mendel’s observations and interpretations remained unnoticed. Independent investigations of Carl Correns in Germany, Hugo de Vries in Holland and Erich von Tschermak in Austria, facilitated the rediscovery (1900) of Mendel’s work and validated his contributions.

*The next question was, where are these ‘Factors’ in the cell?*

Cytologists and Geneticists began to see parallels between the behaviour of chromosomes during cell division and the expected behaviour of Mendel’s Factors. Around 1902, Walter S. Sutton, Theodor Boveri and others independently noted these parallels and the chromosomal theory of inheritance began to take form, which implied that the Mendelian factors are located on chromosomes. Subsequent, elegant experiments by T.H. Morgan (1913) and his group using the *D. melanogaster* fruit flies led to the confirmation of the chromosome theory of inheritance. The abstract genetic factors postulated by Mendel were finally localised on the visible structures (chromosomes) within cells. With this, the principles of segregation and independent assortment can be explained in terms of behaviour of chromosomes during meiosis that produces the haploid gametes carrying only one pair of each chromosome present in the diploid individual. Even though the appearance of different phenotypes were logically deduced to due to different ‘factors’, information on the genotype as well as processes involved in the manifestation of the phenotype were not known. This is indicated by the ‘question marks’ in the equations. *Box 1* summarises the processes which resolved the relationship between genotype and phenotype.

To begin with, let us start with this equation.

![Genotype? ??? Phenotype](image-url)
Box 1.

Summary of the successive stages in the evolution of our understanding on the relationship between genotype and phenotype. The ‘Question marks’ indicate lack of information at that point of time. As and when the information was available, the question marks are deleted.

Post 1900 investigations revealed many more facets of inheritance than what Mendel visualized, such as incomplete dominance, codominance, interaction of genes, epistasis, hypostasis, polygenic-inheritance, sex-linkage, linkage, etc. During this period of ‘Classical Genetics’ (1900–1940), the pattern of inheritance was deduced from “Ratios” of ‘phenotypes’ – such as 3:1; 9:3:3:1; 1:1:1:1. Gene was considered as invisible, abstract and a “Statistical entity”.

The genotype of an individual for a character was deduced based on the outcome of breeding experiments. For example, in *Drosophila*, the genotype of the fly with vestigial wings, a recessive character, will be vg//vg and for the fly with normal wings, dominant character, will be vg+/vg+ (homozygote) or vg+/vg (heterozygote). Similarly, the genotype of a fly with red eye will be w+/w+ or w+/w and of white eye will be w/w.
these experiments revealed additional information, the equation is represented as:

\[ \text{Genotype?} \quad ??? \quad \text{Phenotype} \]

\[ w//w ? \quad \text{white eye} \]

Nature of Gene and its Functions:

*Even though the relationship between genotype and the phenotype was understood, no answers were available to questions like What are these Genes? How do they control different characters?*

Host of investigators working with bacteria and bacteriophages showed the chemical nature of the hereditary material, that is, genes. Experiments of Frederick Griffith (1928) with *Streptococcus pneumoniae*, the bacterium that causes pneumonia, whose colonies had two different phenotypes. Colonies with polysaccharide coat appeared ‘Smooth’ while colonies without polysaccharide coat showed ‘Rough’ surface. The ‘S’ strain is virulent and causes pneumonia; while the ‘R’ strain is non-virulent. Through his elegant experiments, he was able to demonstrate that the ‘R’ strain was transformed into an ‘virulent strain’ due to transfer of a substance from ‘S’ strain. Griffith called the material received from the ‘S’ strain as a ‘transforming principle’, which made the ‘R’ strain virulent. But these observations did not identify the biochemical nature of the transforming principle which was changing the phenotypes and properties of the colonies.

The bacterial transformation experiments of Oswald Avery, Colin MacLeod and Maclyn McCarty (Avery et al., 1944) showed that Deoxyribonucleic acid (DNA) of one bacterial strain when allowed to enter bacteria of a different strain, it could permanently transform the recipient strain with specific features of the donor strain. Alfred Hershey and Martha Chase (1952), working with *E. coli* bacteria, and the T2 bacteriophage, a virus which infects *E. coli*, convincingly established that the DNA of T2, but
not its protein coat, only enters the bacterium, carries the genetic information. Viral reproduction occurs within the host bacterium and progeny with DNA covered by a protein coat are released. On the other hand, in some viruses, such as tobacco mosaic virus (TMV), Ribonucleic acid serves as hereditary material. Following these elegant experiments, the gene which till now was a statistical abstract entity, was identified with as a distinct chemical substance, that is, Deoxyribonucleic acid (DNA), the hereditary material.

Many studies with higher organisms also contributed indirectly to relating genes with chromosomes and DNA: Physical mapping of genes on polytene chromosomes, beginning in 1933; Constancy of DNA, measured by cytophotometry of Feulgen stained cells and 2C-1C relationship between somatic and gametic cells (1940s) and In situ hybridization of specific DNA sequences on chromosomes and their correlation with genetic maps.

*With this breakthrough, that is, identification of the material basis of gene, a string of questions like: What is the mechanism by which genetic information is passed on from one generation to the next? How exactly DNA determines different phenotype(s)/character(s) of individuals? Basis for variations in the phenotypes – What are Mutations? Heritable changes resulting in new phenotypes. What is the structure of the gene etc. were resolved.*

To resolve these, information was required about the composition of DNA. Hydrolysis of DNA revealed its chemistry. DNA is a polymer with repeating units of nucleotides. Each nucleotide is made up of a nitrogenous base, a deoxyribose sugar and a phosphate. Nitrogenous bases are of two types, purines and pyrimidines. Adenine and Guanine bases are purines, while cytosine and Thymine are pyrimidines. Biochemist Erwin Chargaff and his colleagues (1945 and 1950) showed that in DNA, the number of guanine units is equal to the number of cytosine units, and the number of adenine units is equal to the number of thymine units. Therefore, the sum of (A+G:Purines) equals the sum of (C+T :
Pyrimidines). This observation gave a strong clue for pairing between purines and pyrimidines. This is known as Chargaff’s rule. Chargaff’s team also established that the ratio between A+G and C+T differs between different organisms/species.

During this period, Rosalind Franklin and Maurice Wilkins (1950–1953) were examining the X-ray diffraction patterns of DNA. Their data suggested that DNA exists in the form of a helix. The X-ray diffraction pictures revealed the periodicity of bases, which are 3.4 Å apart. James Watson and Francis Crick (1953), taking cognizance of Chargaff’s data as well as of X-ray diffraction data of Rosalind Franklin and Maurice Wilkins, combined with their expertise in model building, proposed the three dimensional structure of DNA. They suggested that DNA exists as a double helical structure. Importantly, they also expressed “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material”. Research findings during the next two decades revealed the mechanism of replication of DNA (Gene), the most fundamental event for inheritance; how the genetic information present in the form the sequence of bases (nucleotides) in DNA is involved in defining/deciding the products of the gene, that is protein, which in turn takes care of the phenotype. Similarly processes of genetic recombination, mutation as well as DNA repair were also being understood. To study the pattern of inheritance of a character, the basic requirement is the availability of heritable alternate phenotypes. Mutations, the heritable changes in gene structure, i.e., changes in the nucleotide sequence, are the ones which bring about heritable variations. Today, radiation and chemical mutagenesis, more specific experimental protocols like gene knockouts, transposon-mediated mutagenesis etc are also available to generate mutations and thus altered heritable phenotypes. This information helps us to rewrite the equation as:

\[
\text{Genotype DNA} \quad ??? \quad \text{Phenotype PROTEIN}
\]
The genotype is represented by DNA (Gene) but the underlying mechanism(s) by which the DNA/Gene directs the manifestation of a phenotype (protein) was unclear. Let me take up this later.

Epigenetics:

Let us briefly look into a few of the investigations/concepts prevailing prior to the advent of Genomics, which had vital importance for further growth of genetics. C.H. Waddington (1942) introduced the concept of Epigenetics (literal meaning “above genetics”). Historically, the word “epigenetics” was used to describe a few genetic events that could not be explained by the then known genetic principles of Mendel as could be seen in the following examples:

1. Position effect variegation in Drosophila – (PEV) (Figure 2). This phenomenon was defined in the 1940s. The $w^+$ gene in Drosophila melanogaster is located on the X-chromosome, away from the centromeric heterochromatin and regulates the development of red eyes, with uniform pigmentation in all the ommatidia. But if the same gene is brought closer to the centromere by the process of inversion, its expression is altered resulting in a variegated eye, with uneven distribution of pigment in different ommatidia. When the same gene is relocated to its original position by reinversion, the normal red eye phenotype reappears. The inference at that point of time was that the local chromatin environment of a gene determines its expression (Figure 2).

![Figure 2](image.png)

Figure 2. Expression of the white gene in Drosophila melanogaster varies with its location on the chromosome.

Figure 3.  a. Mealy bugs have $2n=10$. In Males, 5 chromosomes inherited from the father (Paternal; P) is heterochromatinised while 5 chromosomes received from the mother (Maternal : M) remain euchromatic. In females, both paternal (P) and maternal (M) chromosomes remain euchromatic. b Interphase nucleus of a human female, with one of the X-chromosomes is heterochromatinised and is referred to as Barr Body.

2 Another example is from Mealy bugs (coccids). Both males and females have $2n=10$ as diploid number of chromosomes. Of these, 5 are inherited from father and 5 from the mother. In females, both paternal and maternal sets are euchromatic. In males, only the maternal set of chromosomes is Euchromatic (functional) while the paternal set is Heterochromatic (condensed and inactive or non-functional) (Figure 3a). Therefore, the phenotypes of males resemble those of their mothers only! In males, during spermatogenesis, only the euchromatic set (inherited from the mother) is included in the haploid sperm.

3 Human males have one X-chromosome and one Y-chromosome while females have two X-chromosomes: $X^pX^m$, inherited one from each parent. In human females, one of the two X-chromosomes becomes randomly inactivated in different somatic cells of the body (Barr body, Figure 3b, source Wikipedia). It means that the phenotype of different cells in a female may vary depending upon which of the two X-chromosomes remains active. If the female is heterozygous for a sex-linked trait, the phenotype will be a combination of both dominant and recessive traits. This is related to the phenomenon of Dosage Compensation, which will equalize the X-chromosome linked gene products between males and females. On the other hand, in Marsupial females, the paternal X- is always inactivated resulting in the appearance of mother’s phenotypes only for the X-linked genes.

Although unclear at the time of their discoveries regarding the mechanisms underlying these manifestations of genes, these phenomena have now been shown to be related and to share a
common feature – that is, these heritable changes did not require alternations in DNA sequence, and were said to be due to the epigenetic events, yet to be understood at that time.

Even long before the concept of epigenetics was conceived, the phenomena known as ‘Epigenesis’ was discussed (Wolff 1759, 1764). Belief at that time was that the Germ cell (Egg) consists of a homogeneous substance. The parts of an organism with phenotypes, that is, varied structures and functions, arise progressively. Development was thus viewed as a process of “new formation” for which the term ‘epigenesis’ was used. Development of an organism from a zygote involves a sequence of steps in which cells differentiate and organs, organ systems form. Different systems in the body of an organism, e.g., skeletal system, nervous system, hematopoietic organ systems etc., are structurally and functionally different. Even though all of them are derived from the same single cell, the zygote, they have varied and differentiated PHENOTYPES. Mechanism(s) underlying this was not clear when the term epigenesis was conceived. Waddington’s classical epigenetic landscape (Figure 4) is a metaphor to illustrate depending on the path of differentiation, different lineages of cells develop into structurally functionally different phenotypes.

Figure 4. Adapted from the Epigenetic Landscape proposed by Waddington. Source: adapted from Wikipedia and modified.
Even when Waddington proposed the idea of ‘epigenetic landscape’ in 1942, the genetic events which facilitate differential expression and regulation of genes during development were yet to be resolved. More than 65 years ago, Huxley said ‘I shall use ‘epigenetics’ as meaning the science of developmental process in general’ (Huxley, *Nature*, Vol.177, pp.807–809, 1956). Most of the genetic regulatory events that are involved in influencing the development and differentiation are indeed of “epigenetic” in nature.

What is the Meaning of ‘Epigenetic Events’?

Epigenetics describes heritable but reversible changes in gene function in the absence of changes in primary DNA sequence. It means, inheritance (from one cell to another during cell division) of altered gene expression which results in different phenotypes, without altering the nucleotide sequence of the gene! The consequence of altered gene expression is associated with, altered phenotype. Therefore, in Epigenetics, inheritance is involved, but not as we know it in terms of Mendelian inheritance. Hence, there is a profound shift in our understanding of inheritance. The new phenotype is not due to ‘mutation’ because there is no change in the information of the gene, that is its nucleotide sequence.

*Now the question is who are the other players, which are capable of bringing about heritable phenotypic changes?*

With the realization of the importance and critical roles of epigenetic changes during normal development and cell functions as well in diseases like cancer etc., and studies by molecular and developmental biologists, substantial information about the mechanisms underlying of epigenetics is now available. I will try to summarise the candidates and the processes involved in bringing about epigenetic changes resulting in different phenotypes. The candidates are DNA and Histones of the chromatin. In addition, some non-histone proteins, and the non-coding RNAs have also been implicated in regulating the epigenetic changes. Some of the major epigenetic changes are noted below.
DNA Methylation: For a gene to express, that is to, the promoter region of the gene should be readily accessible to transcription factors and other regulatory units (e.g. enhancers etc.). DNA methylation refers to the addition of a methyl group (–CH₃) co-valently to a base. Addition of a methyl groups to certain bases in the DNA, results in changes in structure of chromatin which prevents its interaction with the transcription machinery and leads to repression of the given gene’s activity. Of the four types of bases, the Cytosine (C) in the dinucleotide 5’-CpG-3’ is more frequently methylated. The methylated CpGs attract methyl-CpG-binding domain proteins that recruit ‘repressor complexes’, resulting in the modification of nucleosomes of the local chromatin region. The modification promotes the recruitment of repressor complexes leading to a more condensed chromatin structure as opposed to an open and active chromatin structure required for transcription. DNA methylation is catalysed by the DNA methyltransferases, including those which establish methylation and those that maintain methylation. In mammals, DNA methylation patterns are established and maintained by three DNA methyltransferases: Dnmt3a and Dnmt3b for establishment, and Dnmt1 for maintenance of methylation. DNA methylation is associated with stable gene silencing (for example on the inactive X-chromosome) either through interference with transcription – factor binding or through recruitment of repressors that specifically bind to sites containing methylated CG. As and when, demethylation occurs, chromatin will be accessible for transcription.

HISTONES, components of nucleosomes are globular small basic proteins. The first 20 or so amino acids of histones, known as the histone tail, are highly conserved and are subjected to post-translational modifications. In the histone tail, serine, threonine, and tyrosine can undergo phosphorylation; lysine and arginine can be methylated, acetylated, ubiquitinated, and sumoylated. Furthermore, lysine residues have potential to be mono-, di- or trimethylated. These Histone modifications are known as “HISTONE MARKS” and provide docking sites for many chromatin-associated proteins. The Histone code hypothesis predicts that the
type, location, and combination of Histone MARKS determine the recruitment of specific chromatin associated proteins and/or transcription factors which determine whether the gene would be expressed or silenced. For example, Histone Acetylation is associated with transcriptionally active genes while Deacetylation is associated with inactive genes (gene silencing). Involvement of other factors like nucleosome remodeling machines and histone variants have also been reported to introduce epigenetic modifications (Seo et al., 2013). Thus, SWI/SNF a nucleosome remodeling multi-protein complex remodels the way DNA is packaged in chromatin. Similarly, Polycomb and Trithorax group of proteins act to remodel chromatin and alter the accessibility of DNA to factors required for gene transcription. The coding in the histones may be heritable from parent to daughter cell during cell division. These histone modifications are reversible.

**NON-CODING RNA:** Non-coding RNAs, as the name indicates, are not carrying message for proteins. Instead, they participate in many activities concerned with the gene expression. Involvement of long non-coding RNAs in controlling multiple epigenetic phenomena are well known. For example, dosage compensation mechanisms in *Drosophila* and mammals involve epigenetic actions of long non-coding RNAs. In human females, one of the X-chromosomes in somatic cells is heterochromatinised (inactivated) as Barr Body. Long non-coding RNAs such as XIST-RNA and TSIX-RNA, along with other chromatin modifiers and repressor proteins are involved in this process. In *Drosophila* males, the single X-chromosome, under the influence of a ribonucleoprotein complex (RNP) called ‘Compensasome’, becomes ‘hyperactive’. The compensasome includes long non-coding RNAs called *rox1* and *rox2* as its major component.

Finally the crosstalk among modifications of DNA, Histones and non-coding RNAs decide the functional status of the chromatin.

The expression and repression of gene expression are reversible epigenetic changes that are caused/influenced by chromatin remodelling factors. This results in getting multiple types of phenotypes from the same genotype. Therefore, an additional step
needs to be introduced in the equation:

Genotype → Epigenotype → Phenotype

**Genome, Epigenome and Epigenomics:**

What is a genome? The entire genetic complement of a cell, that is, entire DNA content of the cell is the genome. What is genomics? It is a discipline that unravels the genetic information present in a the genome, commonly through DNA sequencing, that is, determining the exact nucleotide sequence (A,T,G,C) of the genome. The human genome is made up of 23 chromosome pairs with a total of about 3 billion DNA base pairs and it has about 20,000–22,000 protein-coding genes. The genome of *D. melanogaster*, commonly known as ‘Cinderella of Genetics’ contains four pairs of chromosomes with about 139 million base pairs and about 15,000 protein coding genes.

The vital question in Developmental biology is to understand processes guiding the appearance of new structures such as tissues, organs and organ systems with diverse phenotypes, both in normal healthy and disease systems. It was realised that the key phenomenon is the differential expression of genes in space and time, during development (*Figure 5*).

Epigenetic events are behind the variable expression of genes. Because of its importance, the epigenome project was initiated. The epigenome is the set of chemical modifications to the DNA and DNA-associated proteins in the cell, which alter gene expression, and are heritable (through meiosis and mitosis). The modifications occur as a regulated process during development and tissue differentiation and can be altered in response to environmental exposures or disease (NIH: National Institutes of Health). Epigenomics is the study of the complete set of epigenetic modifications on the genetic material of a cell (Russell, 2010). To uncover the relationship between genotype and phenotype, an intermediary transient state, that is, epigenetic influence on the genome needs to be understood. Because of its importance, two
global collaborative studies are in progress. They are, (1) Human Epigenome project (HEP) which aims to identify, catalogue and interpret genome-wide DNA methylation patterns and profiles of all human genes in all major tissues and their disease variants; and, (2) International Human Epigenome Consortium (IHEC) which focuses on all areas of epigenetics. The roadmap of Epigenomics Project includes study of the epigenomic signatures of a broad spectrum of human tissues and cells undergoing crucial developmental transitions. Epigenomics Mapping Consortium has published non-genetic modifications to the genome – epigenetic modifications – that crucially determine which genes are expressed by which cell type, and when.

As of now, reference epigenomes of 127 human tissues and cell types, drawn from Embryonic stem cells, Induced pluripotent stem cells and Adult tissues, from Healthy individuals and those with cancer, alzheimer’s and autoimmune disease are already available. It is anticipated that initiatives to define the normal human epigenome will enhance progress towards better understanding of the role of epigenetics in human disease phenotypes. Recent molecular techniques such as ‘DNA barcoding of individual cells during development’ have been applied to trace and demonstrate in real-time the transition of different lineages which are passing
Box 2. DNA Barcoding for Cell Development

Deciphering the Genomes-to-Phenomes (G2P) Relationship is considered as one of the grand challenges in biology. DNA barcoding for cell development tracing has advanced to include single cells and single nucleic acid mutations. Techniques are available to trace how and when 26 billion cells present in human body arise from one zygote (Reza Kalhor et al., *Science*, 2018: DOI: 10.1126/science.aat9804; Wang et al., Mol Med Rep. 2021 Dec; 24(6): 849). They were able to record every cell’s history in real-time. This New technique can reconstruct history of every mature cell’s development. With this, biologists are within the reach of resolving one of the grand challenges of developmental biology, that is, how single cell grows to form to an adult animal and with diverse phenotypes.

through varied paths of development (*Box 2*).

Realization of the phenotype related to a particular gene depends on the epigenetic modifications is aptly captured by Kellis (2015) with the statement that “The epigenome brings the genome to life”. Epigenetic changes are not only HERITABLE but also REVERSIBLE. If a Human disease is diagnosed as due to faulty epigenetic modification, it is possible to reverse it. This is referred as Epigenetic Therapy. Also popularly called “Epi-Drugs: from chemistry via biology to medicine and back” (Altucci, L., Rots, M.G. Epigenetic drugs: from chemistry via biology to medicine and back. Clin Epigenet 8, 56, 2016). More than 100 epigenetic agents are currently under investigation. Some of the approved drugs are: DNA methyltransferase inhibitors: 5-azacitidine and Decitabine; Histone deacetylase inhibitors: Vorinostat and Romidepsin.

With the realization of the importance of epigenetic modifications underlying normal as well as pathological phenotypes, many questions have arisen. How Many Epigenomes? How to Analyze Epigenomic Data? Can we keep our epigenome healthy? How can we get it back in shape? Etc. The Epigenetic markings are done by different proteins/enzymes and noncoding RNAs. Therefore, genome of an individual has to code for these different
players and also has to regulate their expression. Therefore, the big question is – Genetics of Epigenetics!

**Holy polymers of Molecular Biology:**

The theme of this article is to understand how genotype determines the phenotype. As mentioned earlier, the genetic information for a phenotype is carried in the nucleotide sequence of the gene. The question is how this informational content is translated in the form of a phenotype. Molecular biologists have dissected the process and showed the participation as well as inter-relationship of three major polymers. They are DNA, RNA and Protein. The first two are polynucleotides with a string of nucleotides while the third one is polypeptide with chain of amino acids. Since the entire molecular biology revolves around these, they may be called as ‘Holy polymers’ of biology. The genetic information present in the form nucleotide sequence in the Gene, that is, DNA, determines the amino acid sequence of the protein coded by it. DNA achieves this through RNA. The flow of information from one polymer to the other, forms the central dogma of molecular biology which was conceived by Crick (1957, 1970) to reflect the directional flow of information from nucleic acids to protein. Since the mechanisms involved in the transfer of information were not known, question marks are placed in the equation. These were resolved subsequently.

![Diagram showing DNA, RNA, and Proteins]

**Transcription, Transcriptome and Transcriptomics:**

After epigenetic and other regulatory decisions, the gene is ready to express. The first step in this is called ‘Transcription’. The information content of an organism is recorded in the DNA (GENES) of its genome and expressed through transcription. The sequence of bases present in the gene/DNA acts as a template to synthesise a complementary RNA. In almost all instances, the primary transcript is subjected to varieties of post-transcriptional modifications to generate the mature functional RNAs. If this
carries a message for a polypeptide, it is called as messenger RNA (mRNA) or coding RNA. The mRNA serves as a transient intermediary molecule in the information network that leads to the development of a character/phenotype. On the other hand, the transcribed RNA, if it is not coding for a protein but is involved in other diverse activities, it is called non-coding RNA, e.g., transfer RNAs, ribosomal RNAs, regulatory RNAs, catalytic RNAs, microRNAs, snRNAs, etc. The non-coding RNAs are assisting the functions of coding sequences at different levels. Organism’s transcriptome is the sum of all of its RNA transcripts. A transcriptome captures a snapshot in time of the total transcripts present in a CELL. It will be a heterogeneous pool OF RNAs. Transcriptome represents an important Biochemical Phenotype of the gene expression. Transcriptomics is the study of Transcriptome.

The key aims of transcriptomics are: to catalogue all species of transcript, including mRNAs, non-coding RNAs and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5’ and 3′ ends, splicing patterns and other post-transcriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions (Zhong Wang, Mark Gerstein, and Michael Snyder, 2009). Whole-transcriptome analysis is of growing importance in understanding how altered expression of genetic variants contributes to complex diseases (disease phenotypes) such as cancer, diabetes, and heart disease. Analysis of genome-wide differential RNA expression (transcription profiling) provides researchers with greater insights into biological pathways and molecular mechanisms that regulate cell fate, development, and disease progression, that is, leading to different phenotypes (Thermo Fischer Scientific).

*The next question is, how to get the phenotype determined by a gene which has conveyed its message to its messenger RNA?*
Translation and Proteins: Proteome and Proteomics:

The Translation process follows transcription, subjected to post-transcriptional regulations. The translation machinery is a complex of coordinated participation of many players such as ribosomes, tRNAs, enzymes, protein factors etc. The genetic code carried by mRNA is decoded and translated to produce the specific sequence of amino acids in a polypeptide chain. The nascent polypeptide chain, the first product of gene is subjected to post-translational modifications, passing through primary, secondary and tertiary changes before assuming the functional configuration.

Even before heredity, gene and its relationship to protein were thought of, Sir Archibald Garrod’s work (initial observations in 1902; book in 1923) on inherited metabolic diseases like Alkaptonuria, Phenylketonuria etc., showed the importance of proteins in ‘disease phenotypes’. During the next three decades, the concepts like one gene – one protein, one gene – one polypeptide, co-linearity between gene and its product, that is protein, etc., established the relationship between two polymers namely DNA and Proteins.

![DNA Transcription RNA Translation PROTEINS](image)

Proteome and Proteomics

A proteome is the collection of proteins (Products of genes, representing a protein phenotype) that are present in a biological entity – a cell, an organ, the blood or an individual. Proteomics is a science that focuses on the study of proteins. It embraces three levels of study. They are, (1) Structural proteomics – in-depth analysis of protein structure; (2) Expression proteomics – analysis of expression and differential expression of proteins; (3) Interaction proteomics – analysis of interactions between proteins to characterize complexes and determine function (Handbook of Systems Biology, 2013). This is one of the examples to illustrate how challenging it would be to establish genotype-phenotype correlations.
The Human Proteome Project (HPP) is designed to map the entire human protein set. The width (the number of various protein species, proteoforms) and the depth (the number of copies of each proteoform in a tissue) have been the focus of proteomics (Ponomarenko et al., 2016; 2017; Lauren A. Blake and Bin Wu, 2019). Human proteome consists of 92,179 proteins out of which 71,173 are splicing variants. It is hoped that this project will enhance understanding of human biology at the cellular level and lay a foundation for development of diagnostic, prognostic, therapeutic, and preventive medical applications.

The gene/genotype has conveyed the information through transcription and translation, and a functional protein is available. But this output alone cannot contribute to a phenotype.

**Metabolome and Metabolomics:**

The Protein, product of ‘Translation’ interacts with other components within the cell, that is, the biochemical composition of the cell/tissue, called ‘metabolites’. This assembly is called ‘Metabolome’ (*Box 3*). The study of the contents of metabolome is referred to as ‘Metabolomics’. The picture that emerges out of this is known as “metabolic fingerprint”, representative of the state of the organism at that time. This leads to the realization of the phenotype of the cell/individual/disease. The Human Metabolome Database (HMDB) is a comprehensive, high-quality, freely accessible, online database of small molecule metabolites found in the human body. 50 different human disease phenotypes can be easily detected using “metabolic fingerprint” and cancer cells are compared to the normal cells for metabolite comparison. The metabolome biomarkers for many lysosomal diseases, types of diabetes etc., are available (*Box 3*).

**Genotype-Phenotype correlations in Complex traits:**

Some traits are called complex because they are controlled by many genes and by environmental factors. A Complex problem. Complex or quantitative traits are important in medicine, agriculture and evolution. Many classical text books on Genetics used to
have chapters on ‘Nature and Nurture’, or ‘Heredity and Environment’. Of late, these are missing. The phenotype of a quantitative trait depends on the interaction between Genotype – multiple genes and the Environmental factors. Hence it is represented as

**Genotype (G) + Environment (E) + Genotype and Environment interactions (GE) → Phenotype (P)**

The animal and plant breeders have further expanded this to study by looking at genotype variance, phenotype variance and environmental variance while evolving different strains adapted to different environmental situations. Statistical approach is required to understand this relationship. Also, the phenomena like phenotypic robustness, phenotypic plasticity, variable penetrance and variable expressivity also to be noted, (Box 4).

**Phenome and Phenomics:**

The ‘Phenome’ refers to all the phenotypes exhibited by the individual. The study of such an assembly of phenotypes referred to as phenomics. In this regard, an interdisciplinary procedure, namely ‘Phenotyping’ is being adopted. According to Guo and Zhu (2014) “the purpose of phenotyping is to produce a description of the plant’s anatomical, ontological, physiological, and bio-
At every stage of transmission of genetic information, there will be the impact of environmental factors.

chemical properties”. Similarly, phenotyping of animal models like mouse has yielded a comprehensive baseline invaluable characterization to investigators with normal anatomy, histology, physiology, metabolism, age- or strain-related states etc. For instance, the Michigan Mouse Metabolic Phenotyping Center (Animal Phenotyping Core) provides access to a broad range of comprehensive, state of the art, specialized phenotyping services for mouse models of diabetes, diabetic complications, obesity, and related metabolic diseases and conditions. Phenomic concepts are used in functional genomics, pharmaceutical research, metabolic engineering, agricultural research etc.

The Extended Phenotype:

This is another concept related to Genotype and Phenotype. Richard Dawkins, the author of the book, *Selfish Gene* has written another book called *The Extended Phenotype*. He has proposed that the expression of a gene is not limited simply to the biological processes and the organism’s physical appearance or phenotype. He has proposed that the phenotype also includes its impact on the or-
ganism’s environment. This is known as ‘Extended Phenotype’ and this concept has profound influence on concepts in ‘Evolutionary Biology’

**How Do All the ‘Omics’ Fit Together?**

To understand the genotype – phenotype correlations, one has to look at the big-data of many ‘omics’ platforms’, genomics to phenomics (Lauwen et al., 2017). One need to have advanced integrated bioinformatic methodology. Babelomics is an upcoming bioinformatic science. Named after the tale ‘The Babel library’, a masterpiece by the famous Argentinean writer Jorge Luís Borges that describes an infinite library containing all the possible books. Alonso et al., (2015) have discussed the functional interpretation of new generations of genomic data. Babelomics is an integrative web-based platform for the functional analysis of transcriptomics, proteomics and genomic data with advanced functional profiling.

**Message:**

Johanssen proposed a seemingly straightforward relationship between genotype and phenotype. Over the years, knowledge on both genetics and phenotype has expanded and continue to be enriched by interdisciplinary investigations. The study of genotype-phenotype relationship fits into the philosophy of Systems biol-
ogy, that is, instead of focusing on individual parts, the focus has to be on a complete system made up of different parts interacting with each other – integrated and interacting network of genes and their products, proteins and biochemical reactions, and environmental factors (Figure 6).

This story is also a good example of how scientific understanding builds up with curiosity driven research coupled with raising questions one after another.

Acknowledgements

I profusely thank Prof. S.C. Lakhotia for Critical reading of the draft of the article; and Smt. Ashwini Ranganath and Smt. Thanmayi Ranganath for assistance during the preparation of the manuscript.

Suggested Reading


