Version 3 a0026653

Genetics as Explanation: Limits to the Human Genome Project

Irun R Cohen, *The Weizmann Institute of Science, Rehovot, Israel*

Henri Atlan, *Ecole des Hautes Etudes en Sciences Sociales, Paris, France and Hadassah University Hospital, Jerusalem, Israel*

AU:1

Sol Efroni, *Bar-Ilan University, Ramat Gan, Israel*

AU:2

Based in part on the previous versions of this eLS article 'Genetics as Explanation: Limits to the Human Genome Project' (2006, 2009).

Introductory article

Article Contents

Online posting date: 14th October 2016

Living organisms are composed of cells and all living cells contain a genome, the organism's stock of deoxyribonucleic acid (DNA). The role of the genome has been likened to a computer program that encodes the organism's development and its subsequent response to the environment. Thus, the organism and its fate can be explained by genetics, the plans written into the sequence of genomic DNA; the Human Genome Project was devised to decipher this program. However, it is now clear that the genome does not directly program the organism; the computer program metaphor has misled us. The genome is only one class of vital information that serves the organism. Indeed, we now know that the healthy individual human is an ecosystem that lives in symbiosis with hundreds of different species of bacteria – the microbiota. Metaphorically, the genome can be likened to a toolbox for accomplishing specific tasks.

Definitions

Genetics refers to the structure and function of genes in living organisms. Genes can be defined in various ways and at various scales of interest: genes are a concept used by scientists who study biological evolution, single organisms, populations of organisms, species, cells and molecules, heredity, embryonic development, health and disease and life management. These are quite diverse subjects, and the people who study them would seem to use the term gene in distinctly different ways. But genetics as a whole is organised by a single unifying principle, the deoxyribonucleic acid (DNA) code; all would agree that the information borne by a gene is linked to particular sequences of DNA. At the chemical level, we can define a gene as a sequence (or combination of sequences) of DNA that ultimately encodes a protein. The genome refers to the germline DNA that an organism has inherited from its progenitors. The genome includes DNA genes along with DNA sequences that do not appear to encode proteins. **See also**: **Gene Structure and Organization**; **Genome Organization of Vertebrates**; **Protein Coding**

a0005008
a0005001.pub2
a0005017.pub2

Now we can define the Human Genome Project: the genome project is a translation project. Its objective is to translate the chemical sequence information borne by the genome into the verbal information of human language and thought; the aim is to translate DNA sequences into words and ideas that can develop and spread among human minds. What we can manage to do with this information depends on how well we understand the functions of genomic DNA within the organism.

Metaphors and Programs

Most minds use metaphors to understand and explain; we grasp the essence of the unfamiliar (or the complex) by seeing its likeness to the familiar (or the simple). Metaphors are not merely literary devices; metaphors, which also include mathematical models, can aid precise thinking. Which metaphor is suitable for explaining the function of the genome?

Metaphorically, the genome is often likened to a computer program; just as the computer reads and executes the instructions of its program, the body is proposed to read and execute the

eLS subject area: Bioethics & Philosophy

How to cite:
Cohen, Irun R; Atlan, Henri; and Efroni, Sol (October 2016) Genetics as Explanation: Limits to the Human Genome Project. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0005881.pub3

instructions borne by the genome. The body, from this point of view, is mere hardware. The genome is the boss (**Figure 1**).

The computer program metaphor is often extended to explain evolution: evolution is thought to improve DNA programs. Diversification of genomes by random mutation combined with the selection of the most successful variants (survival of the fittest) leads, as it is claimed, to the continuous upgrading of existing DNA programs. The evolution of genomic DNA is automatic but costly – the death of the less fit drives the process. **See also**: **Evolutionary History of the Human Genome**

a0005070.pub2

Metaphors and Expectations

The computer program metaphor fosters high expectations of the Human Genome Project. Theoretically, if you know all the information borne by a computer program, you can expect to know how a computer using that program will operate; you can understand the computer's present behaviour and can predict its future behaviour with a high degree of accuracy. You would even be able to repair mistakes in the program, if that program were simple enough.

Metaphorically then, if the genome is really like a computer program, the genome project will empower us to understand the organism, predict its response to the changing environment and provide a key to the cure of its maladies. Or so many would have wished to believe.

Here we shall discuss what a program means to most people and then test whether the genome actually fits the bill. We shall see that the program metaphor is a misleading way to describe the genome, knowing the genome will not explain the organism. We will then go on to consider other metaphors for the genome.

Genome is Not a Simple Program

The *Oxford English Dictionary* (Draft Revision, 2009) defines a computer program as 'A series of coded instructions which when fed into a computer will automatically direct its operation in carrying out a specific task. Also in extended use: something conceived of as encoding and determining a process, esp. genetically'. A computer program is usually written intentionally by a computer programmer; the DNA program, by contrast, is written by evolution, without intention. But irrespective of who or what writes a program, at the very least, a program is a plan for a sequence of events. So most people would like a program to be unambiguous, coherent and definite. The program's task should be inherent in the program itself; the information in a program should be sufficient for the job. A program, like a blueprint, is a type of representation. But the genome, as every working biologist knows, is ambiguous, incoherent and indefinite. The most debilitating to the genetic program metaphor is that the genome is not autonomous or complete. Epigenetic processes, processes external to the DNA sequence, can markedly influence gene expression. Consider the following processes that modify the function of DNA sequences:

- Introns and exons: A DNA sequence that encodes a protein in a multicellular organism is usually discontinuous and is interrupted by chains of apparently meaningless DNA (introns); the discontinuous segments of the DNA that encode the sequence of the expressed protein are termed *exons*. The gene transcript [messenger ribonucleic acid (mRNA)] has to be spliced together by protein enzymes that cut out the introns and connect the exons. Thus, most DNA sequences are not intrinsically coherent; introns and exons have to be sorted out.
- Alternative splicing: Many DNA coding sequences, perhaps as many as a third, can undergo alternative splicing to produce different proteins – the identities of introns and exons in a single chain of DNA can vary; a given segment may be in intron (and so skipped) in one situation and an exon (and so expressed as part of a protein) in another situation. In other words, a single DNA sequence, by alternative splicing of introns and exons, can give rise to more than one species of protein. Moreover, the way the DNA actually gets spliced is not governed by the DNA sequence itself; enzymes and other proteins actually determine the gene – the spliced DNA sequence that is expressed in particular circumstances. Thus, the information encoded in many DNA sequences is intrinsically ambiguous until realised by the actions of proteins on the DNA.
- Conformational variations: The way a protein functions depends greatly on the three-dimensional shape assumed by the protein – its conformation. Different conformations of the same sequence of amino acids can expose different positively or negatively charged hydrophilic domains or hydrophobic domains, each with different arrays of interactions and functions. The conformation of the protein depends on how the chain of amino acids constituting the protein folds; different folds of the same sequence give rise to different functions. The sequence of the protein is encoded by a single DNA sequence, or gene, but the conformation of the protein is determined dynamically by its folding in response to interactions with other proteins in the environment and other factors, including the pH. Consequently, a single gene that gives rise to a single amino acid sequence of a protein may be said to function in more than one way.
- Posttranslational modifications: Moreover, the protein encoded by the gene can (and does) undergo chemical modifications (enzymatic cleavage, aggregation with other molecules, phosphorylation, glycosylation, methylation, binding with molecules such as ubiquitin and so forth) to carry out further functions independent of the gene that encoded the protein. The protein glyceraldehyde-3-phosphate dehydrogenase first discovered as an enzyme, for example, is now known to have a role in membrane fusion, microtubule bundling, RNA export, DNA replication and repair, apoptosis, cancer, viral infection and neural degeneration. The protein's gene obviously does not determine the functional program of the protein.
- Cellular environment: The sets of genes expressed at a particular time are determined by molecules external to the genome; the previous history of the DNA can be overridden. For example, the sheep Dolly was cloned by transplanting a

nucleus from an udder cell into an ovum. The molecular environment of the udder cell normally activated the milk genes of the nucleus; after the nucleus was transplanted to the ovum, the genes needed for making a new sheep became activated. The cellular environment 'reprograms' the genome epigenetically.

- Epigenetic modifications: Sequences of DNA that might otherwise serve to encode proteins can be inactivated epigenetically by enzymes that attach methyl groups (methylation) to the cytosine moieties of these segments of DNA. Strangely,<?xmltex the process of DNA methylation can be influenced by whether the particular DNA allele has been inherited from the mother or the father of the individual. This DNA imprinting is poorly understood but is essential to normal development.
- Regulatory or noncoding RNA: In the beginning, molecular biologists focused their attention on the question of how sequences of DNA – genes – determined the sequences of proteins; RNA molecules were viewed as messengers – mRNA – that mimicked a given DNA sequence and served as a template for the sequence of amino acids generated by ribosomes during protein synthesis. Molecules of mRNA served as functional copy of a gene for gene expression. Now however, it has become clear that many species of RNA, other than mRNA, are generated by sequences of DNA that do not encode proteins – noncoding RNA (ncRNA). These ncRNA molecules, also known as regulatory RNA (rRNA), include relatively short microRNA and relatively long ncRNA. Many different molecular and functional types of ncRNA are under study and much remains to be characterised and understood. Nevertheless, it clear that these ncRNA molecules play essential roles in the regulation of DNA gene expression and function – from DNA replication to expression, from suppression to enhancement of DNA expression and from DNA splicing to chromosome structure. The paradigm of one DNA gene–one mRNA molecule–one protein–one function is clearly outdated. How are we to define genes now? Should DNA sequences that encode ncRNA or rRNA be defined as genes? A new, more complex world of molecular biology has been discovered. The importance of ncRNA in differentiation, development, health, disease and evolution is in the offing. In any case, a genome project simply directed to sequencing DNA cannot alone explain the organism.
- Stage of development: A single protein can function in very different ways during prenatal development and later in life after development is completed. Thus, the gene encoding the protein can be seen to perform different functions at different times; the meaning of the gene varies with the stage of development.
- DNA deletions: Some DNA genes can be removed from the genomes of experimental animals (knockout genes) without producing an overt change in the form or behaviour of the animal – the phenotype. Knocking out other genes, in contrast, can lead to severe and unexpected effects on the phenotype. Scientists who knock out genes are not infrequently surprised by the resulting phenotype of the animal. In other words, the impact of a gene on an organism is not readily deducible from knowledge of its DNA sequence. The relationship between the genome – the genotype – and the phenotype is not direct.
- Immune somatic gene generation: The immune system exploits the genome to create novel genes after fertilisation. Each clone of lymphocytes in the immune system epigenetically constructs its unique antigen receptor by recombining otherwise unexpressed minigene elements inherited in the germline. The immune system thus manufactures millions of different genes that are not encoded as such in the genome. The immune system functions to heal the organism and protect it from foreign invaders and is also a key factor in causing autoimmune diseases. Yet, the ability of the immune system to recognise antigens, a major determinant of health or disease, is not inherent in genomic DNA. Indeed, monozygotic twins, born with identical genomic DNA, are often not concordant for autoimmune or inflammatory diseases such as type 1 diabetes, multiple sclerosis, lupus or rheumatoid arthritis; if one twin develops such a disease, the other twin has only about a 30% chance of developing the same disease; in other words, a twin set of inherited genes does not insure that each twin will develop in the same way. The immune system and its consequences on health or disease emerge from genetic and developmental events that take place subsequent to the formation of the individual inherited genome. Note that the development and function of the human brain, like those of the immune system, emerge post-genomially from somatic experience; identical twins do not develop identical brains. Clearly, one's individuality cannot be reduced to one's inherited genomic DNA.

Microbes, Symbiosis and the Holobiont

The inherited genome, in recent years, has lost even more of its programmatic importance in determining the phenotype of the individual human: it has become clear that humans, like all multicellular organisms, live in symbiosis with myriads of bacteria. We each carry in our digestive and respiratory tracks and on and in our skin many hundreds if not thousands of diverse species of bacteria; indeed, our bodies bear tenfold more prokaryote DNA than the amount of inherited genomic DNA we were born with. Our guts, in particular, house astronomical numbers of bacteria, archaea and yeast cells. This situation is not merely benign parasitism; our health depends on our cooperative interactions with these 'foreign' resident bacteria. We each acquire our resident microbiota at birth, and their numbers and types vary dynamically as we grow and develop; the microbiota of individuals differ and reflect one's gender, age, diet, geography, style of life and state of health. Bearing the 'wrong' numbers and combinations of microbiota – termed dysbiosis – is associated with diseases such as inflammatory bowel disease, type 1 or type 2 diabetes, obesity, high blood pressure, atherosclerosis, autoimmune and chronic inflammation and even cancer. Consequently, our immune systems have evolved to select and maintain our symbiotic partnerships with many species of bacteria (and viruses), all the while rejecting potential pathogens. In other words, the

AU:3

AU:4

immune system functions to discriminate between different foreigners.

Thus, each of us is essentially an ecosystem formed by the combined interactions of a collective of cells bearing genomes of diverse origins – we are *holobionts*. The new appreciation of the individual is illustrated in **Figure 2**. It is clear that one's inherited genome is only part of one's being; the inherited genomic sequence is only a component in the holobiont that is the individual self.

It is ironic that the discovery of the holobiont, which detracts from the explanatory power of human genomic DNA alone, is itself a positive outcome of the Human Genome Project; our symbiotic microbiota are so well adapted to the living human that most of them cannot be cultured on artificial media outside the body and so could not be detected or studied by traditional microbiology techniques. The microbiota are now identified, classified and studied using DNA sequencing techniques developed as a consequence of the genome project – rather than trying to culture them, we simply sequence the microbial DNA present in faeces and tissues.

Stem Cells Express Multiple Genes

Activation of specific genes in the inherited genome cannot easily account for the development of the embryo: embryonic stem cells express no biological functions other than the capacity for self-renewal – they can replicate – and their potential to differentiate into all the types of specialised cells required to develop the organism such as the brain, blood cells, muscles, kidneys and so on. A simple concept of the genome as a program would lead us to expect that embryonic stem cells should express very few genes before they differentiate and that specialised differentiation would be marked by the progressive expression of specialised genes. This is not the case: embryonic stem cells express many more genes than do their differentiated daughter cells; and differentiation is marked by silencing the expression of the 'superfluous' genes (**Figure 3**). It remains to be seen why and how the stem cell state is associated with global gene expression. Nevertheless, the global state of gene expression in embryonic stem cells challenges the notion that specific DNA expression functions as the program that drives cell differentiation.

In their summation, these and other facts well known to biologists lead to the conclusion that the meaning of the information encoded in the genome is variable and conditional; the meaning of a DNA sequence cannot be derived from the sequence itself. Thus, the genome does not encode a coherent plan for a sequence of events. **See also**: **Alternative Processing: Neuronal Nitric Oxide Synthase**; **Alternative Splicing: Evolution**; **Epigenetic Factors and Chromosome Organization**; **Vertebrate Immune System: Evolution**

One may argue that the genome, despite its lack of intrinsic meaning, is still a set of instructions, albeit with many possible branching points. Even so, the extragenomic environment and the history of the organism and its attendant symbionts determine the path through which the genome is expressed. Since the given state of an actual person is not determined by the person's inherited genome, the genome is not a representation of the person. For this reason, the master-program metaphor does not clarify the role of the genome, but rather obscures it. The mere encoding of amino acid sequences and the sequences of ncRNA by DNA is not programming. On the contrary, the organism uses, manipulates, regulates and, in the case of the immune system, creates DNA genes. The genome acts as the organism's servant, not as its master. Why then have knowledgeable people likened the genome to a master program?

Meaning: Line or Loop?

The concept of the genome as program is associated with the idea that the connection between a gene and its meaning is linear: DNA–mRNA–protein–functional meaning.

A specific DNA sequence was seen as the plan for making, through the agency of mRNA molecules, a particular protein. The protein (e.g. an enzyme that builds or degrades molecules or a transcription factor that activates genes) is the agent that carries out a defined activity. As the DNA encodes the protein, the meaning of the information borne by the DNA is transformed ultimately into the precise action performed by the protein as an enzyme, transcription factor or other agent. A one-to-one relationship was envisioned: one gene for each protein and one protein for each function. Thus, the activities of the protein – the meaning of the gene – were held to be inherent in the gene: the information.

But, in reality, the living system is not a linear progression from DNA information to protein function; the system is a recursive loop. Proteins, as we have discussed earlier, are required to make sense out of the DNA sequence; the proteins are required to activate and even to manufacture the very genes that encode the proteins. This way of drawing the connection is closer to reality (**Figure 4**).

A circle has no beginning and no end: the information actually expressed by DNA is formatted by proteins recursively generated in the process. There is no linear transformation of information (DNA) into meaning (protein action). Genetic information itself is one of the products of protein action; the activities of proteins generate legible DNA, in a reiterating loop. We must add the environment to the loop; the influence of proteins on genes is modulated by intracellular and extracellular factors. ncRNA regulates much of gene expression and cellular function and so affects the loop. The symbiotic components of the holobiont individual are also critical to the process. There is no fixed hierarchy, no one-to-one relationship. The living system is not simply a transformation of DNA information into protein information. The living system is an ongoing process – a reactive ecosystem. The meaning of the process that connects DNA and protein is not an outcome of the process; the meaning of the process is the process itself.

Self-Organisation and Program

Scientists had hoped that the genome might function as a simple program because people, especially scientists, think programmatically. Planning is a characteristic of the human mind. We have

intentions and goals; we scheme and we plot. We implement programs, so we take programs for granted; every building has to have an architect; a watch implies a watchmaker.

But we also know of many complex natural phenomena that organise themselves without recourse to a master plan; the world is filled with them. A colony of ants or a hive of bees seem wonderfully organised, yet no single ant or bee, not even the queen, has an idea in mind of what a colony or hive should look like. (Queens are just egg-laying machines.) Each ant and each bee only responds mindlessly to what it senses. What seems to us to be a master plan actualised by each insect colony or hive emerges from the combined actions of the insects themselves, each insect autonomous and entirely ignorant of a world beyond its own sensations.

Similarly, an organism is built and operates with the help of its genome; but the genome is only one element in a recursive process. The iterating cycle of genes that form RNA and proteins that define and regulate genes is the self-organising process from which the organism emerges. If there be a genetic program, then such a program writes itself collectively. The action, as it were, precedes the plan. But how can that be?

Complexity, Reduction and Emergence

Physics is the paradigm of sciences; the other sciences try to emulate physics. Physicists explain the behaviour of matter by reducing material phenomena to the basic laws of matter and energy. Underlying the physical world are fundamental laws that account for what we see; the material world – the phenotype of reality, as it were – is explainable by these laws and so is reducible to these laws. Reduction is done by analysing the data of sense and experiment to uncover the underlying elements (mathematical laws or component parts) that give rise to or 'cause' the data.

Biologists, noting the success of reduction in physics, have attempted to reduce the phenomena of living organisms – the living phenotype – to the DNA code. Unfortunately, it does not work; life is far too complex to be explained entirely by genomes, either inherited or acquired through symbiosis. **See also**: **Systems Biology: Genomics Aspects**

a0005928

We do not mean to say that reduction should not be done in biology. On the contrary, scientific reduction has been the key to the identification and characterisation of the elements – the cells and molecules – that constitute living organisms. The power of modern biology must be credited to reductive analysis. Our point is that reduction to component parts is only the beginning of wisdom. The essence of biology, like that of other complex systems, is the emergence of high-level complexity created by the interactions of component parts.

Emergence is not a mystical concept. A physical basis for the emergence of self-organisation has been established in studies of nonequilibrium thermodynamics: open systems that exchange matter and energy with their surroundings can maintain themselves in steady states far from equilibrium. The decrease in internal entropy in such systems can be offset by increased entropy in the surroundings; this makes it possible for macroscopic organisation to emerge from the coupling of multiple microscopic reactions. Certain coupled chemical reactions exemplify such processes experimentally. Computer simulations of networks of automata have also provided examples of the emergence of high-level non-programmed functions created by the interactions of component parts. But these simple examples only illustrate the bare principle; present models of emergence will need upgrading to deal with the complexity of actual biological systems.

Emergence in biology is difficult to study because we have not yet devised a mathematical language suitable for modelling and simulating the generation of high-level complexity out of simple parts. Fortunately, the Human Genome Project, with its need for advanced bioinformatic technology, has invigorated collaborations among biologists, mathematicians, physicists and computer scientists. New ways to model and study the emergence of complexity are already emerging from these activities. As we mentioned earlier, the microbiome – a new way of seeing the human – has emerged from the Human Genome Project. But until biology and the informatic sciences develop a common language, we shall have to make do with examples; fortunately examples of emergence abound. Think of your mind. The mind emerges not from neurons but from the interactions of functioning neurons; all the neurons may be intact and alive, but there will be no local or global brain functions unless individual neurons interact. The functions of the brain are not reducible to neurons in isolation; brain functions emerge from the ongoing interactions of neurons. The interactions create brain functions we call the mind. Emergent functions, like your mind, are not mere abstractions; they work. **See also**: **Information Theories in Molecular Biology and Genomics**

a0005927

Evolving Genomes

Evolution, as we have noted, does write genomes, but does not improve genomes, even metaphorically. Genomes, at the level of the species, develop from the processes that adapt an organism to its world. Now a bacterium is no less adapted to its environment than is a human being to its environment. A bacterium as a form of life might, in fact, enjoy a more robust future than the fragile and pugnacious human species. The life and survival of a bacterium would not be improved by making the bacterium more like a human. Self-consciousness would not help a bacterium. Improvement is relative to one's point of view; people like to see themselves as superior to bacteria.

So what does evolution accomplish if improvement is spurious? Evolution leads to accumulating complexity; humans are objectively more complex than isolated bacteria; in effect, humans are symbionts composed of diverse bacteria (and viruses) interacting in a complex ecosystem. Evolution is a process that, rather than generating improvement, generates new information.

The Environment and the Genome

We can summarise the limitations of the genome most easily by repeating what has already been said many, many times: one's

genes are only an incomplete explanation for one's being; the present environment and its history, at the scales of the cell, the person, the group and the biosphere, interact with the genome to determine its expressions and effects. Note that the interactions between a genome and the environment are exceedingly complex; the contribution of a gene to a phenotype cannot always be separated from the contribution of the environment, despite sophisticated calculations, because the interactions between genome and environment are not linear and not additive. The effect of a given gene, as we saw regarding the gene for haemoglobin S, can depend greatly on the environment.

Language Metaphor

What metaphor might be generally useful for appreciating the function of the genome? Genomic DNA is a reservoir of raw information that, suitably processed, can be translated into the amino acid sequence of functional proteins or into RNA molecules that perform various regulatory functions in the cell. The genome, then, is a store of information suitable for transmission into meaningful effects. Systems useful for storing and transmitting information are the defining features of humans; language is surely the transmission system we all are most at home with. Perhaps we could think of the genome as akin to a list of words, a vocabulary, which can be used to build and express a language. Vocabularies and genomes, despite the very different ways each bears information, turn out to share many features:

- Both reservoirs of information mutate and undergo complex evolution over time.
- Both reservoirs of information are transmitted from generation to generation (the mode of transmission differs markedly – fertilisation compared to learning).
- Both reservoirs of information replicate and are shared by interacting groups (species compared to language communities).
- Both reservoirs of information manifest individual differences between persons (each individual bears a unique genotype and a unique mind expressing an individual pattern of vocabulary and word usage).
- Both reservoirs of information can be processed to generate meaning by influencing concrete behaviours and structures (proteins and cells compared to ideas, social interactions and communities). But neither reservoir of information alone programs meaning: you use your vocabulary to express your thoughts, but your vocabulary alone does not tell you what to think or what to say. Similarly, you use your DNA to express proteins, but your DNA alone does not tell your cells what proteins have to be expressed or what the proteins mean (their function) in a particular situation. The meaning of a string of DNA, like the meaning of a string of words, emerges from the complexities of history, dynamic interactions, context and circumstance.

Fragments of genomic DNA, like words, acquire different meanings in different contexts. They can be used artfully to tell different stories. The genome, like a vocabulary, is information, transmissible from generation to generation, that is available for processing into meaning. The process itself, as we have discussed, is the storey.

Tool and Toolbox Metaphors

The genome, in summary, is a collection of information inherent in DNA sequences that in time of need can be used by the cell to construct proteins or rRNA sequences for specific structural and functional applications. The cell may materialise the same segment of the genome in different ways; cell enzymes, for example, may use a particular stretch of DNA to combine various coding regions and splice sequence information to construct different proteins. The cell thus packages and expresses a given stretch of DNA in different ways depending on the state of the cell. The stretch of DNA is not a *program* for making these different proteins; the stretch of DNA is a *tool* used by the cell for making proteins (and/or regulatory DNA).

From this point of view, a stretch of DNA comprises a tool or a number of tools useful for further construction and function. The one-dimensional sequence of DNA serves ultimately as a template for synthesising one-dimensional sequences of proteins, which, depending on the ionic and molecular environments of the protein, fold into variable three-dimensional functional structures – enzymes, growth and differentiation factors, transcription factors, hormones, replication agents, death molecules and so forth; these molecules interact dynamically over time (the fourth dimension) in organised pathways to generate a functioning organism. If a stretch of DNA is a tool, then we may define the genome as a toolbox (**Figure 5**).

Tools and toolboxes are metaphors used in computer discourse. The *Oxford English Dictionary* (Draft Editions 2007) defines a computer tool as 'any item of software … used as the means of accomplishing a specific task', and a toolbox as 'a set of software tools designed to facilitate the construction of more advanced tools … in specific application areas'. In the case of the living organism, the more advanced tools are the proteins and all the rest of the services provided by the toolbox of DNA sequences needed for constructing the phenotype of the organism. The genome toolbox is inherited from generation to generation, and the interaction between the phenotype of the organism and the process of natural selection leads, via heredity, to evolution of the toolbox. According to the toolbox metaphor, there is no master program; the organism emerges from an ongoing loop of interactions (**Figure 5**).

Conclusion

The Human Genome Project, like putting a man on the moon, has been a costly undertaking of great technical virtuosity. It is good that the project has been done for the daring of it and because it has already provided much important information about genetics and the organism. No less important, the genome project has spawned powerful technologies and has opened biology to the age of informatics. Biology has learned that it is an informatic science. Finally, the very success of the genome project has

dispelled the simplistic illusion of the genetic program; biology is now aware of its true complexity. The genome project, wittingly or not, has built the foundation for deeper probes into the complexity of life. The limitations of the project are only the limitations of the genome itself. **See also**: **Biological Complexity: Beyond the Genome**; **Complex Genetic Systems and Diseases**; **Genetics, Reductionism and Autopoiesis**; **Systems Biology: Genomics Aspects**

AU:6
a0005886.pub2
a0005887.pub2
a0005895.pub2
a0005928

Reference

Efroni S, Duttagupta R, Cheng J, *et al.* (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**: 437–447.

AU:7

Further Reading

Atlan H (1987) Self-creation of meaning. *Physica Scripta* **36**: 563–576.

Atlan H and Koppel M (1990) The cellular computer DNA: program or data? *Bulletin of Mathematical Biology* **52**: 3335–3348.

Atlan H (2001) Epigenesis and self-organization: new perspectives in biology and medicine. In: Compendium A, Solomon B, Taraboulos A and Katchalski-Katsir E (eds) *Conformational Diseases*, pp. 291–297. Basel: The Center for the Study of Emerging Diseases, Bialik Institute Jerusalem and Karger.

Belizário JE and Napolitano M (2015) Human microbiomes and their roles in dysbiosis, common diseases, and novel therapeutic approaches. *Frontiers in Microbiology* **6**: 1050.

Cohen IR (2000) Tending Adam's Garden*:* Evolving the Cognitive Immune Self. San Diego, CA: Academic Press.

Cohen IR (2006) Informational landscapes in art, science, and evolution. *Bulletin of Mathematical Biology* **68**: 1213–1229.

Cohen IR and Harel D (2007) Explaining a complex living system: dynamics, multi-scaling and emergence. *Journal of the Royal Society, Interface/the Royal Society* **4**: 175–182.

Efroni S, Harel D and Cohen IR (2005) A theory for complex systems: reactive animation. In: Paton R and McNamara L (eds) *Multidisciplinary Approaches to Theory in Medicine: Studies in Multidisciplinarity*, vol. **3**, pp. 309–324. Amsterdam: Elsevier.

Efroni S, Duttagupta R, Cheng J, *et al.* (2008) Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**: 437–447.

Zilber-Rosenberg I and Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* **32** (5): 723–735.

Article Title: Genetics as Explanation: Limits to the Human Genome Project

Article ID: a0026653

Article DOI: 10.1002/9780470015902.a0005881.pub3

Version: 3

Previous version(s):
10.1002/9780470015902.a0005881.pub2

Article Type: Standard

Readership Level: Introductory article

Top level subject categories: Bioethics & Philosophy

Keywords: genetics # computer program # evolution # information # complexity symbiosis

Key Concepts

- The genome does not function as a master plan or computer program for controlling the organism; the genome is the organism's servant, not its master.
- The genome is a reservoir of DNA sequence information and a vehicle for transmitting this information; the meaning of DNA emerges from the cellular processing of this raw information into proteins and other functional molecules.
- Complex organisms like humans live in symbiosis with a microbiota of many hundreds of micro-organisms. The microbiota are essential to human health and the phenotype of the individual is greatly influenced by interactions with the microbiota. The body houses manyfold more DNA of microbial origin than it does the genomic DNA inherited from the parents. The individual is thus an ecosystem of diverse cellular origins.
- DNA is only one class of vital information that serves the organism; the organism epigenetically uses, manipulates, regulates and, in the case of the immune system, creates genes.
- The effects of a gene vary with the organism's environment; the interactions between genes and environment are not linear and, in most cases, not additive. Therefore, one cannot compute with certainty the relative contributions of genes and environment to an organism's observed features – its phenotype.
- Metaphorically, we can think of the genome as akin to a list of words, a vocabulary, which can be used to build and express a meaningful language; like a vocabulary, a genome by itself has no functional meaning.
- The genome is thus akin to a toolbox of DNA sequences that provide molecular tools as requested by the internal state of the organism and the state of the environment.
- One's genes cannot explain one's being: an organism is the expression of a dynamic and ongoing interaction between the state of its environment and its internal state, which includes its past history and its toolbox of DNA sequences.

Glossary

***Biologic evolution*#** Evolution is the hereditary change over time in species of organisms resulting from the ability of individuals within the species to thrive in the given environment. Evolution involves genetics, epigenetics, the states of the environment and the states of individuals within species.

***Complex system*#** Complex systems are formed by the interactions of many different kinds of components that can be arranged in different alternative ways; thus, different properties can emerge from various alternative arrangements of such systems (a complex system such as a cell or an organism can function in different ways).

***Ecosystem*#** A network of interacting species of living organisms together with the environments that house them. The healthy functioning of the ecosystem is essential to the well-being of the participating species and of their environments.

***Epigenetics*#** Changes in the expressions of genes or in the appearance of the organism that are not caused by changes in the DNA sequence.

***Heredity*#** The passing of traits from parent to offspring, which involves both genetic and epigenetic mechanisms.

***Information*#** A 'just-so' arrangement of elements as distinct from random arrangements of the same elements (e.g. words, numbers, DNA sequences, amino acids and proteins). Information per se has no meaning unless something or somebody responds to it; meaning emerges from the effects of information, and so the same information (DNA sequence, words, etc.) can have different meanings in different situations.

***Microbiota*#** The collective of micro-organisms (bacteria, archaea and yeasts) that live in symbiosis with the healthy human body or with the bodies of other multicellular organisms. The microbiota typically includes many hundreds of different species. An abnormal composition of the microbiota – termed dysbiosis – may be associated with various diseases. The term microbiome is often used to refer to the microbiota.

***Symbiosis*#** The living together in close interaction of individuals of different species for the mutual benefit of the participants.

***System state*#** The state of a system is a particular arrangement of all of its component parts and their interactions.

***System*#** Formed by elements that are held together by their mutual interactions and separated (relatively) from their environment by a definable boundary. Systems are often arranged in nested hierarchies: systems of interacting atoms form molecules, systems of interacting molecules form cells, systems of interacting cells form organisms and so forth.

Author(s) and Affiliation(s):

Irun R Cohen, *The Weizmann Institute of Science, Rehovot, Israel*

Henri Atlan, *Ecole des Hautes Etudes en Sciences Sociales, Paris, France and Hadassah University Hospital, Jerusalem, Israel*

Sol Efroni, *Bar-Ilan University, Ramat Gan, Israel*

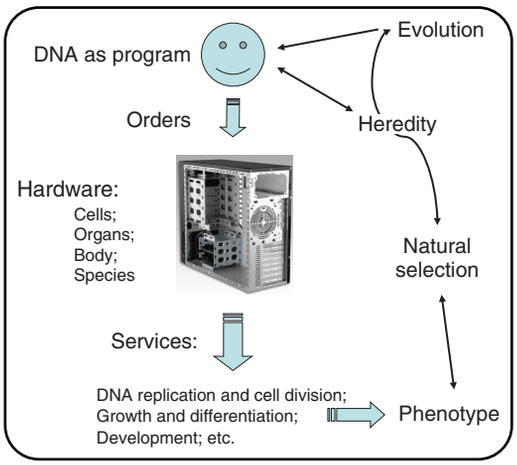

Figure 1 Genomic DNA has been viewed as a master program. According to this metaphor, the cell's DNA is considered to function like the brain of the cell; the DNA is likened to a computer program that sends *orders* to the *hardware* – the cells, organs, body and species that bear the genome and its variants. The cellular hardware performs *services* as ordered by the program to effect DNA replication, cell division, growth, differentiation, development, hereditary transmission and other vital functions; these in total give rise to the *phenotype* of the organism. The *phenotype* through *natural selection* and *heredity* leads to *evolution* of the DNA program (arrows).

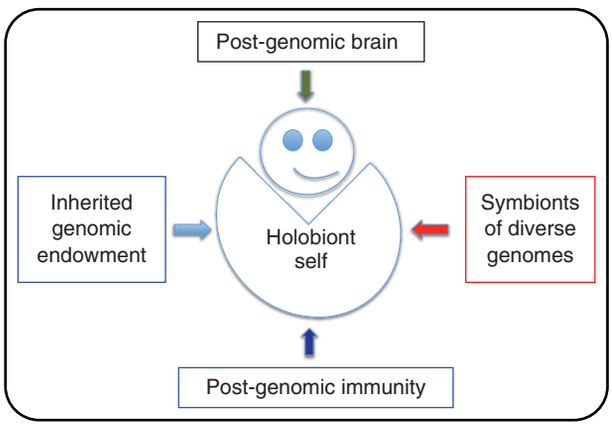

Figure 2 The individual is a holobiont that includes eukaryote cells bearing genomic DNA inherited from its parents along with myriads of symbiotic microbial cells. The species of microbial cells in the microbiota bear diverse genomes; the individual phenotype is thus the expression of multiple genotypes. In addition, the individual contains a brain and an immune system that develop beyond their initial genomic information.

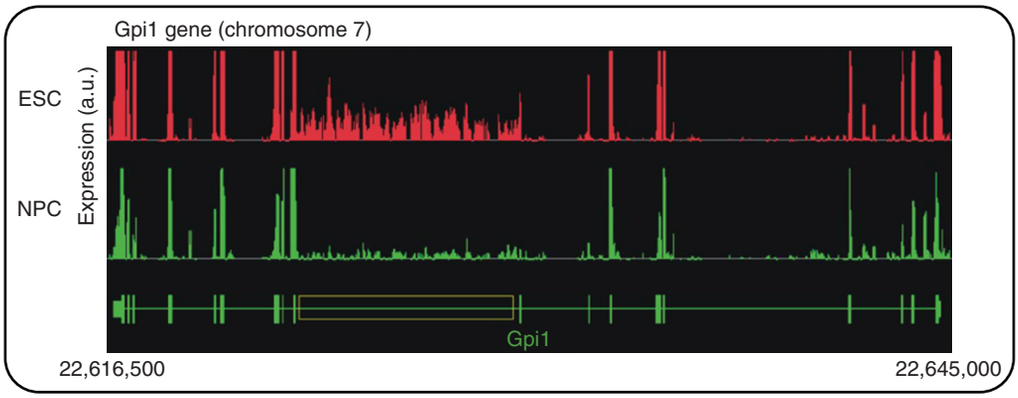

Figure 3 Embryonic stem cells express more segments of genomic DNA than their differentiated offspring cells do. The figure depicts the relative degree of DNA expression as the height of the vertical lines along the strand of DNA encoding the Gpi1 gene on chromosome 7 of the mouse in embryonic stem cells (ESC; upper red) and in significantly more differentiated neuronal precursor cells (NPC; lower green). It is clear that the Gpi1 locus (delineated at the bottom of the figure) is expressed to a greater degree in the ESC than in the NPC. The product of the Gpi1 gene is expressed in many blood cells and is involved in the synthesis of a glycolipid that serves to anchor proteins to the cell surface. The specific panels in the figure are part of a genome-wide assay of gene expression using microarrays that tile the entire mouse genome. Reproduced from Efroni *et al.* (2008) © Elsevier.

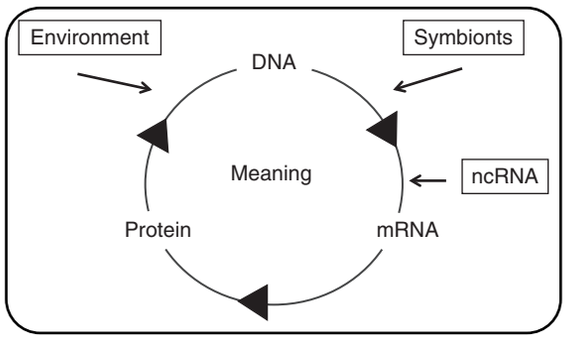

Figure 4 The relationship between DNA, RNA and proteins, the expressed products of genomic DNA, is circular. It takes DNA and RNA to produce proteins, but it takes RNA and proteins to make genes. The circular relationship is an ongoing process strongly influence by ncRNA, symbiotic microbiota and environment. The structures and behaviours generated by the process are the meaning of the process (see the text).

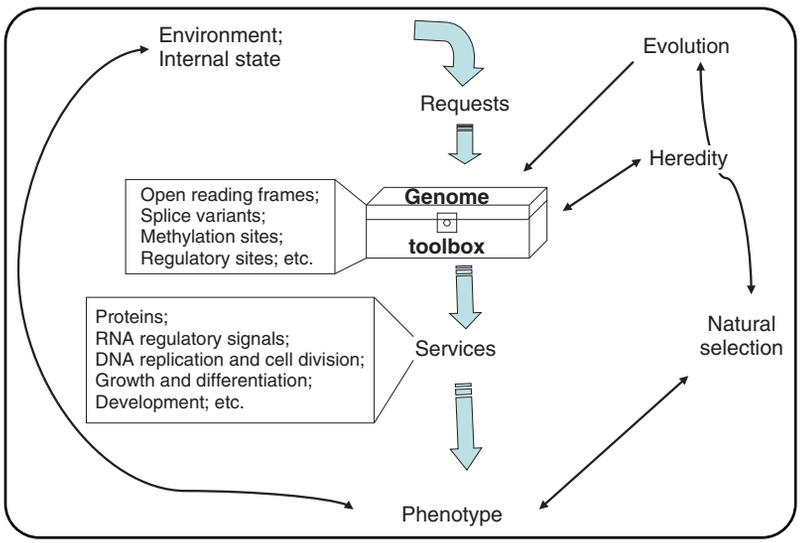

Figure 5 The genome is a toolbox. The genome is envisioned as a box of *DNA sequence tools* that are materialised into *proteins* and *RNA regulatory signals* by expressing open reading frames, splice variants and so forth. These various expressions of DNA tools are *services* made in response to *requests* generated by the *internal state* of the organism and by the state of the *environment*. The proteins and other service molecules made using the DNA tools affect cell and body structure, cell division, growth, differentiation and other organismal functions that together generate the organism's *phenotype* – its observed characteristics. In contrast to the genome viewed as master program (**Figure 1**), the output of the genome generates the input that submits *requests* to the genome toolbox: the *internal state* of the organism's *phenotype* together with the state of the *environment* feeds back to generate *requests* to the *genome toolbox*. In other words, there is no master program a priori; the genome is an element in an ongoing circular loop (**Figure 4**). The *phenotype*, through the processes of *natural selection* and *heredity*, generates the *evolution* of the toolbox.

Version 3 a0026653

Author Query Form

Title: eLS

Entry Title/Article No: Genetics as Explanation: Limits to the Human Genome Project/a0026653

Dear Author,

During the preparation of your manuscript for typesetting some questions have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

Queries and/or remarks

AU:1	Two affiliations have been provided for the author 'Henri Atlan'. To keep in line with the encyclopedia's copyediting style, please indicate only the principal affiliation to be retained in the article.
AU:2	In order to ensure that the work of authors who have contributed to the previous version of this article, but who are not actively contributing to this new version, is properly recognised, we have added the following byline to your article: Based in part on the previous versions of this eLS article, Genetics as Explanation: Limits to the Human Genome Project (2006, 2009) (amend if necessary). If you did NOT use material from the previous version of the article please let us know, and we will ensure that this additional byline is not included.
AU:3	Please check and confirm if the head level is fine here.
AU:4	Generally, the term bacteria should be used instead of prokaryotes. When used in the classic sense, authors should substitute the term prokaryotes with "bacteria and the archaeans". Please check your usage of the word prokaryotes in this article and indicate whether this should be changed to "bacteria" or "bacteria and the archaeans".
AU:5	Please provide the high resolution artwork for Figure 3.
AU:6	As per style the section 'Reference(s)' is mandatory for this article. so please provide maximum of 50 cited references.
AU:7	Reference Efroni et al. (2008) is cited in Figure 3, so it has been moved to 'Reference' section from 'Further Reading' section. Please check.