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*

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

Based in part on the previous version of this Encyclopedia of Life Sciences (ELS) article, Genetics as Explanation: Limits to the Human Genome Project by Irun R Cohen and Henri Atlan.

Introductory article

Article Contents

Online posting date: 15th December 2009

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 program that encodes the organism's development and its subsequent response to the environment. This computer metaphor teaches that the organism – the hardware – is controlled by the software – the genome. Thus, the organism and its fate can be explained by genetics, the plans written into the sequence of genomic DNA; the 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. Metaphorically, the genome can be likened to a vocabulary of words that can be deployed to make meaningful sentences, or 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 organized 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

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.

ELS subject area: Science and Society

How to cite:
Cohen, Irun R; Atlan, Henri; and Efroni, Sol (December 2009) Genetics as Explanation: Limits to the Human Genome Project. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0005881.pub2

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.

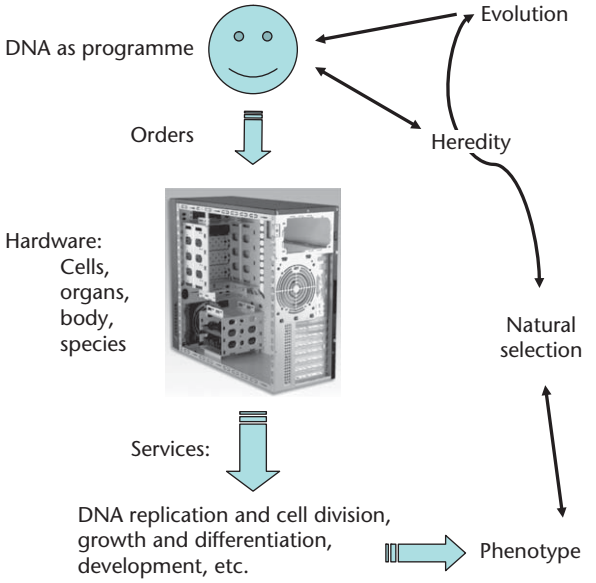

Figure 1 Genomic DNA has been viewed as a master program. According to this metaphor, the cell's DNA is considered to function like the brain of the cell; the DNA is likened to a computer program that sends *orders* to the *hardware* – the cells, organs, body and species that bear the genome and its 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).

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 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, 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

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. Most debilitating to the genetic program metaphor, 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:

- 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 gene transcript (messenger ribonucleic acid (mRNA)) has to be spliced together by protein enzymes that cut out the introns. Thus most DNA sequences are not intrinsically coherent.
- Many DNA-coding sequences, perhaps as many as a third, can undergo alternative splicing to produce different proteins. In other words, a single DNA sequence 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; 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 realized by the action of proteins.
- The protein products encoded by a gene may also vary: a single protein may assume several functionally different

conformations, and so the gene that gives rise to the protein may be said to function in more than one way. Moreover, the protein encoded by the gene can (and does) undergo chemical modifications (enzymatic cleavage, aggregation with other molecules, phosphorylation, glycosylation, methylation 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 is obviously not the program of the protein.

- 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.
- 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, 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.
- MicroRNA molecules (miRNA) are short segments of RNA that regulate gene expression. Such miRNA molecules are encoded in DNA sequences that are not translated into proteins. Thus, DNA that does not encode protein (nongenetic DNA) can regulate the function of DNA sequences that do encode proteins.
- 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.
- Some genes can be removed from the genomes of experimental animals (knocked-out 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.
- The immune system exploits the genome to create novel genes. 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 recognize antigens, a major determinant of health or disease, is not inherent in genomic DNA.

Stem Cells Express Multiple Genes

Embryonic stem cells express no biologic functions other than the capacity for self-renewal – they can replicate – and the potential to differentiate into all the types of specialized cells required to develop the organism – brain, blood cells, muscles, kidneys, etc. 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 specialized differentiation would be marked by the progressive expression of specialized 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 2**). 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 contradicts 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 determine the path through which the genome is expressed. Since the given state of an actual person is not determined by the person's 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 within DNA nucleotide sequences is not programming. On the contrary, the organism uses, manipulates, regulates and, in the case of the immune system, creates 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?

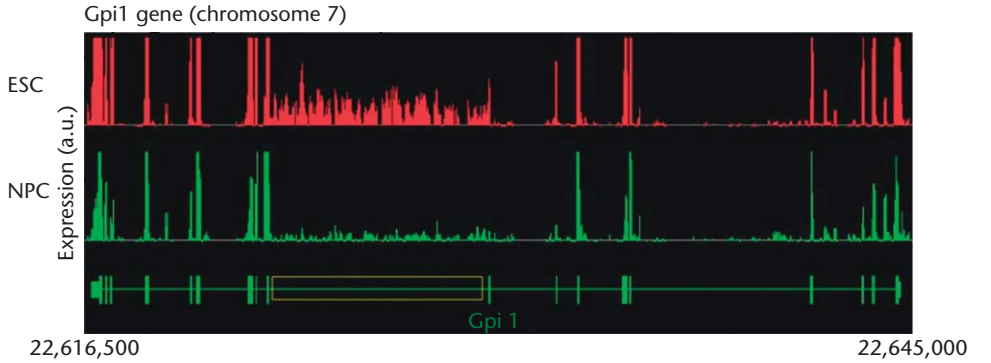

Figure 2 Embryonic stem cells express more segments of genomic DNA than do their differentiated offspring cells. The figure depicts the relative degree of DNA expression as the height of the vertical lines along the strand of DNA encoding the Gpi1 gene on chromosome 7 of the mouse in embryonic stem cells (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) with permission from Elsevier.

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 → messenger RNA → 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. Since 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 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 3**).

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 loop. We might even add the environment to the loop; the influence of proteins on genes is modulated by intracellular and extracellular factors. There is no fixed hierarchy, no one-to-one relationship. The living system is not transformational. The living

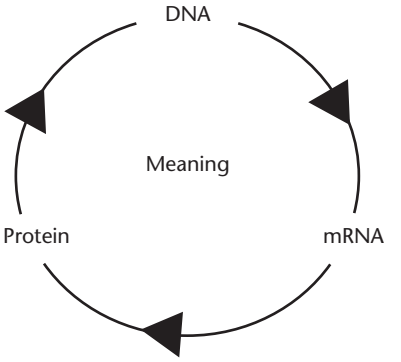

Figure 3 The relationship between DNA, RNA and proteins, the expressed products of genomic DNA, is circular. It takes DNA and RNA to produce proteins, but it takes RNA and proteins to make genes. The circular relationship is an ongoing process. The structures and behaviours generated by the process is the meaning of the process (see the text).

system is an ongoing process. 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-organization 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 organize themselves without recourse to a master plan; the world is filled with them. A colony of ants or a hive of bees seem wonderfully organized, 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 actualized 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 proteins that form genes is the self-organizing 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 (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. **See also**: Systems Biology: Genomics Aspects

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 characterization 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-organization 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 organization 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 nonprogrammed 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 between biologists, mathematicians, physicists and computer scientists. New ways to model and study the emergence of complexity are already emerging from these activities. 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

Genetic Causality: The Case of Sickle Cell Disease

Detailed knowledge of the DNA sequence, the outcome of the genome project, will not suffice to explain health or disease. We shall have to look to the activation of genes and the dynamic functions of proteins and other molecules involved in the processes of life. Biological and cultural evolution, and the environment too, have their place in the action. Take, for example, the case of sickle cell disease. **See also**: Sickle Cell Disease as a Multifactorial Condition

Sickle cell disease is a deficiency of red blood cells (anemia) characterized by an abnormality in the haemoglobin molecule, such that the affected red blood cells assume an elongated shape (like a sickle) at low oxygen tension. The sickle-shaped red blood cells stick together and are destroyed, producing the anaemia. Small blood vessels get clogged by the clumps of sickled red cells and tissues suffer from the lack of blood flow. The disease, untreated, results in an early death. What is the cause of sickle cell disease?

The answer is deceptively simple. Sickle cell disease is a genetic disease; the sickle haemoglobin molecule is abnormal because of a mutation in the gene encoding the β chain of the molecule: a single glutamic acid in the protein is replaced by the amino acid valine – this abnormal haemoglobin is called haemoglobin S. Persons who have inherited the gene for haemoglobin S from both parents can

make only haemoglobin S and so manifest the disease. Persons who have inherited one haemoglobin S gene and one normal gene (heterozygotes) are essentially free of the disease. Thus we could define the disease as caused by having inherited two copies of the haemoglobin S gene. But this is not the whole story. **See also**: Genetic Variation: Polymorphisms and Mutations

The haemoglobin S gene is present mostly in populations of people who have originated in equatorial Africa, and the incidence of the gene is much higher than would be expected from the spontaneous mutation rate of standard haemoglobin to haemoglobin S. The relatively high frequency of a potentially lethal mutation suggests that the mutated gene must have some selective advantage, must contribute to fitness. Well, it turns out that children infected with a certain type of malaria (and who are untreated) will die of the infection if their genome contains only the standard haemoglobin gene. The children who carry one allele for haemoglobin S (heterozyotes), however, are relatively resistant to malaria, and so do not die of that disease. The heterozygous children also do not die of sickle cell disease. Of course, children whose genomes include two of the haemoglobin S alleles (homozygotes) die of sickle cell disease. Thus, we might say that haemoglobin S is an advantageous adaptation to malaria, and the gene is maintained in the population, despite the loss of homozyzgous children, at a rate that reflects the selective pressure exerted by the rate and severity of malaria infection. The death of homozygous individuals is the price paid by the population for heterozygous resistance to malaria. So we could say that the high frequency of haemoglobin S (and sickle cell disease) is caused by malaria. By this reasoning, we might say that sickle cell disease has value as a trade-off in exchange for malaria. In environments free of malaria, however, haemoglobin S provides no advantage.

Should we now conclude that one of the causes of sickle cell disease as a disease (rather than as a trade-off) is the absence of malaria? Sickle cell disease is a serious health problem in the African Americans whose ancestors were taken to America as slaves from West Africa. Should we include the slave trade among the causes of sickle cell disease in an African-American child? Or has the disease been caused by two heterozygotes falling in love?

Persons who have inherited two haemoglobin S genes do much better clinically if they continue to produce fetal haemoglobin after birth; the fetal haemoglobin makes the affected red cell more resistant to the deleterious effects of haemoglobin S. Indeed, homozygous persons are now treated with a drug that induces the production of fetal haemoglobin after birth. Are we to conclude that the normal termination of fetal haemoglobin production is a causal factor in sickle cell disease?

In short, a 'simple' genetic disease like sickle cell disease, which is associated with a defined mutation resulting in a defined molecular abnormality, presents us with a complex causal chain of events. How much more complex are the possible genetic explanations for diseases such as type 1 diabetes or multiple sclerosis, diseases that have been associated with many different susceptibility genes. Indeed, inheriting susceptibility genes does not make the disease inevitable. Take for example identical twins, who bear identical genomic DNA; if one twin develops type 1 diabetes or multiple sclerosis, the other twin will develop the disease in only about a third of the pairs. Having susceptible DNA does not suffice to explain the disease. Indeed, prevalent genes that are associated with susceptibility to complex diseases are probably advantageous trade-offs. **See also**: Complex Genetic Systems and Diseases

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 are bacteria. Evolution is a process that, rather than generating improvement, generates new information. But that issue is beyond the scope of this discussion.

The Environment and the Genome

We can summarize 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 – fertilization 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 story.

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 regulatory RNA sequences for specific structural and functional applications. The cell may materialize 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).

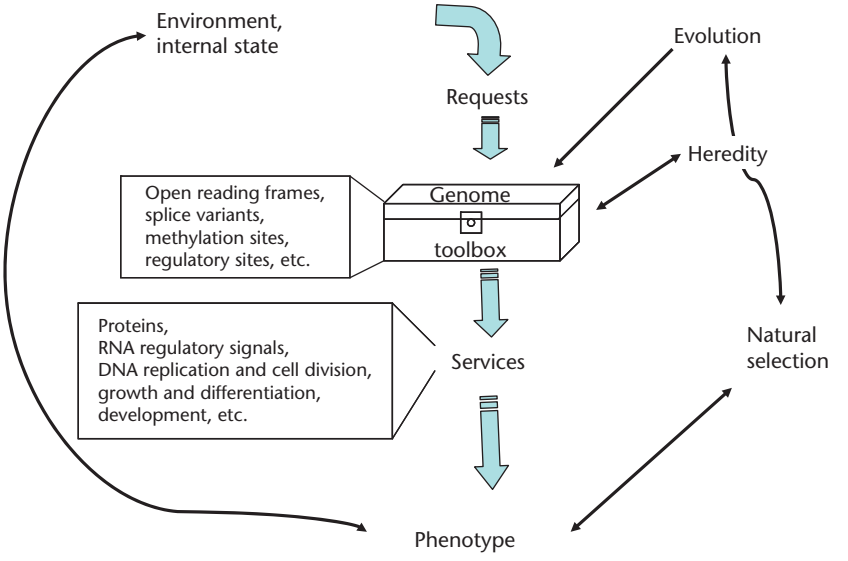

Figure 4 The genome is a toolbox. The genome is envisioned as a box of *DNA sequence tools* that are materialized into *proteins* and *RNA regulatory signals* by expressing open reading frames, splice variants and so forth. These various expressions of DNA tools are *services* made in response to *requests* 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 (see **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 (see **Figure 3**). The *phenotype*, through the processes of *natural selection* and *heredity*, generate the *evolution* of the toolbox.

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 creating three-dimensional protein structures – enzymes, growth and differentiation factors, transcription factors, hormones, replication agents, death molecules and so forth – that interact dynamically over time (the fourth dimension) in organized pathways to generate a functioning organism. If a stretch of DNA is a tool, then we may define the genome as a toolbox (**Figure 4**).

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 lead, 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 (see **Figure 4**).

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

Further Reading

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

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.

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

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 Royal Society, Interface* **4**: 175–182.

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

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.

Fox Keller E (1995) Refiguring Life. In: *Metaphors of Twentieth-century Biology*. New York: Columbia University Press.

Neuman Y (2008) *Reviving the living: meaning making in living systems. Studies in Multidisciplinarity*, vol. 6. Amsterdam: Elsevier.