Hist. Phil. Life Sci., 28 (2006), 527-544

Molecular Epigenesis: Distributed Specificity as a Break in the Central Dogma

Karola Stotz

Cognitive Science Program
810 Eigenmann
Indiana University
Bloomington, IN 47408 USA

ABSTRACT – The paper argues against the central dogma and its interpretation by C. Kenneth Waters and Alex Rosenberg. I argue that certain phenomena in the regulation of gene expression provide a break with the central dogma, according to which sequence specificity for a gene product must be template derived. My thesis of 'molecular epigenesis' with its three classes of phenomena, sequence 'activation', 'selection', and 'creation', is exemplified by processes such as transcriptional activation, alternative *cis*- and *trans*-splicing, and RNA editing. It argues that other molecular resources share the causal role of genes; the sequence specificity for the linear sequence of any gene product is *distributed* between the coding sequence, *cis*-acting sequences, *trans*-acting factors, environmental signals, and the contingent history of the cell (thesis of distributed causal specificity). I conclude that the central dogma has unnecessarily restricted genetic research to the sequencing of protein-coding genes, unilinear pathway analyses, and the focus on exclusive specificity.

KEYWORDS – Waters, Rosenberg, causal specificity, distributed specificity, regulated recruitment, combinatorial control, regulation of gene expression,

1. Introduction

Francis Crick's restatement of his Central Dogma of Molecular Genetic, originally published 1958, clarified that the dogma comprises both the sequence or colinearity hypothesis and a statement about the direction of information flow between DNA and its RNA and protein products: 'The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that such information cannot be transferred from protein to either protein or nucleic acid' (Crick 1970, 561; Crick 1958; Sarabhai *et al.* 1964). I believe that the historical significance of the dogma lies not so much in hypothesizing about the direction of information flows, even though this came to commence an unfortunate 40 year long fixation of molecu-

lar genetics with genetic determinism and with linear pathway analyses at the expense of understanding cyclic processes. Rather, its significance lies in Crick's insight into the fundamental difference between nucleic acids' and amino acids' *specificity*.

The idea of specificity, first of macromolecular structure and than also of linear sequence, has been the touchstone for modern biology. It transformed our understanding of biological mechanism from a highly fluid and interactive process into an assembly of pieces each with its own specific and restricted part to play (Greenspan 2001). The first half of the last century was characterized by the concept of chemical or *conformational* specificity, namely the ability of an enzyme's binding site to recognize the chemical structure of its specific ligands. The fewer substrates a protein can bind, the greater its specificity. Quantum mechanics provided the necessary insight to explain the idea of structural complementarity, a key-and-lock system of recognition in terms of the stereospecificity of enzyme and substrate to form a certain number of weak hydrogen bonds. Crick's central dogma added to this concept of *analog* specificity based on the idea of 'form' the new concept of *informational* sequence, or *digital* specificity of nucleic acid based on the idea of 'information' encoded in the sequences of nucleotides.

In this essay I do not deal with the dogma's negative claim of information flow. I accept its definition of information as sequence specificity and take issue with the implicit or positive statement of the central dogma and its modern defenders, impersonated here by C. Kenneth Waters and Alex Rosenberg, that the gene is deterministic in gene expression and therefore all gene products are fully specified by the DNA code. Waters's thesis of causal specificity is basically restating Crick's central dogma of sequence specificity in causal language, with the slight modification of splicing agents as sharers of this specificity in certain cases. While he actually agrees with my main point that DNA has to share its sequence specificity, Waters's argument still appears as an attempt to 'rescue' DNA as the (more or less) sole bearer of causal specificity in order to a) justify 'why so much research attention in developmental biology is centered on DNA', and b) to 'reveal the fallacy of causal parity arguments' (Waters forthcoming b). Waters's newest position is not about the ontological status of genes, just their pragmatic values. But I do not believe that we can 'forcibly separate science's function as the facilitator of technology from its means of understanding things' (Laughlin 2005, xvi). Waters's position can be used by more metaphysically inclined theorists such as Rosenberg as another means of

*mis*understanding the role of genes development. Rosenberg, a stronger defender of the central dogma, agrees with Waters analysis of the primary status of genes and vehemently opposes the legitimacy of molecular epigenesis as laid down here and elsewhere (Stotz 2006, submitted) as a valid argument against the central dogma.

I do not aim to defy the dogma by claiming that the linear sequence from RNA to protein can be reversed or that proteins code for proteins. Instead I shall show that DNA shares its sequence specificity with other cellular actors. In other words, sequence specificity is not monopolized by DNA but is distributed among certain DNA sequences, plus regulatory RNAs, proteins, and environmental signals. If we focus on the regulation of gene expression instead of blindly taking the dogma for granted it becomes apparent that digital and analog structures work hand in hand as they are both recruited, supported by specific environmental signals into larger multi-molecular complexes comprising DNA, RNA and proteins to synthesize gene products and regulate cellular processes. The genome itself, besides containing digital specificity in form of its string of DNA base-pairs, is a complex three-dimensional structure that yields analog information which carries out important work. In addition, during the process of transcription and RNA processing the digital strings of single-stranded DNA and RNA have the tendency to form secondary and even tertiary structures that add a second layer of informational content to the one-dimensional DNA code. Comparing the human genome with its transcriptome reveals sequence information not encoded by the literal DNA code alone. Intra- and intercellular and even extra-organismal environmental signals impose instructional specificity on regulatory RNAs and proteins organized in expression mechanisms of mind-numbing complexity, which have an impact on the final sequence of the gene product.

Even if we restrict ourselves to the investigation of sequence specificity of gene products we see that the organism's molecular complexity is not specified by its limited number of protein coding genes but by what it can do with its genome. At another place I have proven this point with detailed examples of how nucleotide sequences are *activated, selected* and *created* by *causally specific* regulatory mechanisms of genome expression (Stotz submitted). There is no room here to introduce the reader to this exiting new research into the details of how gene expression is regulated. Many processes are only now beginning to become fully understood, while there are still many more where our knowledge is far from complete; but a new picture is slowly emerging.

Some conclusions drawn from it will be used here to argue against Waters and Rosenberg. My goal is not to understand the full complexity involved in the regulation of genome expression, much less the biological mechanisms beyond the production of the primary sequence of gene products. This paper has the limited agenda of using our new knowledge about sequence modifying processes to conclude which agents other than genes carry sequence specificity. I will conclude that *the distributed control of genome expression, the extent to which it amplifies the literal coding sequence of the 'reactive genome'*[1] *by providing additional sequence specificity to an underspecified DNA sequence, extends the range of 'constitutive epigenesis'*[2] *all the way down to the molecular level of sequence determination.*

2. Actual Difference Makers and Causal Specificity

C. Kenneth Waters has recently repeated, clarified, and justified a central thesis of his former analysis of the molecular gene concept. He identifies the privileged role of the molecular gene in many biological explanations as that of an 'actual difference maker' with 'causal specificity' (Waters forthcoming b). I argue that Waters's account clearly downplays some of the major theoretical insights into genome structure or function revealed by contemporary molecular genetics and genomics, including surprising ways in which DNA performs its traditional gene-like functions, new un-gene-like functions, and other cellular structures that may share some of DNA's cellular function. As I have argued elsewhere, his central claim is no longer suitable to capture our current knowledge of genome structure and function (Stotz 2006). Here I take issue with several of his most recent formulations of his genetic causation model phrased in terms of causal specificity:[3]

Thesis 1: 'Only the *activated* DNA segments (the genes) are actual difference makers of RNA sequences' (Waters forthcoming b, my emphasis).

Thesis 2a: 'The initial synthesis of RNA in prokaryotes and eukaryotes involves many causes, but only DNA is the *causally specific actual difference maker*' (Waters forthcoming b, my emphasis).

2b: '*Possible* exceptions involve cases of differential RNA splicing and editing. *If differential RNA splicing occurs within the same cell struc-*

[1] See Gilbert 2003.

[2] For a more detailed description of constituent epigenesis see Stotz 2006; Robert 2004.

[3] For argument's sake lets pretend that I accept his general model of causation.

ture at the same time, then differences in the linear sequences among *these* polypeptides … could be said to be caused by differences in splicing factors, rather than differences in DNA. It would still technically be true that different 'split genes' were involved'[4] (Waters forthcoming a, my emphasis).

Thesis 3: 'I will note that this qualifier does not need to be added for the case of genes for RNA or polypeptides in Prokaryotes or for the case of genes for unprocessed RNA in Eukaryotes' (Waters forthcoming a). 'DNA is *the causally specific actual difference maker* with respect to the population of RNA molecules first synthesized in eukaryotic cells' (Waters forthcoming b, emphasis in original).

The next section aims to show why these three theses give a wrong- or at least too weak-description of the underlying causation of gene expression.

3. Molecular Epigenesis and Distributed Specificity

For a much more extensive list of examples drawn from the most recent research into the complex mechanisms involved in the regulation of gene expression, especially some of the newest results about sequence-specifying actors of RNA splicing and editing, the two major sequence modifying processes, (see Stotz submitted). What follows is just the summary of my interpretation of these research results, applied to the two competing hypotheses of the central dogma with its (more or less) monopolized causal or sequence specificity on the one hand, and on the other my thesis of molecular epigenesis with its distributed sequence specificity.

Thesis 1:

To restate Waters's first thesis, he singles out 'activated' DNA as the causally specific agent responsible for the composition of a population of RNAs in a cell. The default position of eukaryotic DNA is inactivation and Waters deliberately neglects and downplays all the processes that are involved to *activate* DNA as causal agents. Second, he forgets to clarify between which two states DNA should function as the actual *difference* maker. It could be the difference in the linear sequence between any two gene products, or the difference between two populations of

[4] This move would depart from conventional molecular genetics, and it would mean that pre mRNA and final RNA are specified by two different genes; this would be a drastic step just to withhold causal specificity from splicing agents.

RNAs in two cells of an organism. The second problem is what is commonly called the foremost 'problem of development': the differentiation of cells from a single cell in multicellular organisms. The peculiarity of the differentiated cells is that despite their immense phenotypic differences they all share the *same* genotype (with some notable exception as immune cells). Hence the actual *difference* between two cells is not their DNA but activating agents such as specific transcription factors and inducing signals that co-differ between two cells. The latter orchestrate the tissue-dependent and time-specific *activation* and sequence *selection* of a subset of 'genes' that translates into different cellular phenotypes. The phenotypic difference between two daughter cells could result from the expression of *different* genes (with different causal specificity) or the time-, tissue-, and combination-dependent expression of *common* genes (with the same causal specificity). *Activation* of DNA is therefore a causally specifying mechanism by determining a particular RNA product to be there. In addition, since activation *selects* between different promoters and is likely to influence co-transcriptional activities such as splicing and editing, activation is causally specifying the particular sequence of a RNA product from the same DNA sequence through sequence *selection* and *creation*.

Thesis 2a:

Waters's main thesis states the exclusivity of DNA in providing causal sequence specificity. With some notable exceptions, only DNA provides the linear sequence specificity of any gene product. So while he agrees in principle that DNA alone is not the sole source of sequence specificity, I believe my argument presents a *radical* shift in focus from genetics (molecular) to distributed sequence specificity (systems biological). Against Waters's almost *exclusive* notion of causal specificity of DNA I set a picture of *distributed* causal specificity, where already *pre-selected and activated* DNA shares the stage with the RNA processing machineries of splicing, editing, modification, and translational recoding that further *select*, *modify,* and *newly create* DNA and RNA sequences.

It is the specific recruitment of transcription factors to varying complexes by *trans*-acting factors (proteins, RNA, and environmental factors) that imposes their specificity. Specificity is imposed by environmental induction of activators, differential recruitment and combinatorial control. Agents other than the original coding sequence have to provide sufficient splice-site specificity. In other words, the availability of certain *trans*-acting factors and the differential and combinatorial binding of spliceosomal binding RNAs and proteins to splice sites and regu-

latory sequences (the cellular splice code) seems to be the major contributor to splicing specificity. The selective use of nucleotide sequences through a range of transcriptional, co- and post-transcriptional mechanisms co-specifes the linear sequence of the final product.

Thesis 2b:

Under certain, restrictive conditions Waters is willing to extend causal specificity to splicing and editing agents, namely when different splice variants exist in the same cell at the same time; this is not credited when each cell produces its own splice variants, which would render the regulatory machinery as background condition. For the argument's sake, I interpret Waters to reason as follows: *From an observer's viewpoint*, in certain cellular conditions a gene is *always* specifying a particular splice variant, hence it holds the causal specificity. However, *from the viewpoint of the DNA sequence or the entire cell*, the relevant splicing and editing mechanisms are the providers of sufficient sequence specificity for the right product. In reality, however, most cells just differ in their *ratios* of a particular splice variant: '[F]or most alternatively spliced transcripts there is no 'default' or unregulated state; instead, the ratio of alternative splice forms observed for a given pre-mRNA results from a balance between positive and negative regulation' (Ladd and Cooper 2002, 3; Celottoa and Graveley 2001; Athanasiadis, Rich, and Maas 2004).

In radical cases the *linear sequence* of the final product is not mirrored by the DNA sequence but is extensively *scrambled*, *modified,* or literally *created* through a variety of co- and post-transcriptional processes, which often are interdependent with mechanisms of sequence activation and selection. Cases of sequence creation are even stronger counterarguments to Waters's main thesis of exclusive DNA sequence specificity than any of the 'conservative' cases provided above. In many cases, for instance in the human brain, the editing-derived coding information is essential for the normal functioning of the organism. *This phenomenon provides a potential break in the central dogma according to which coding information must be template derived.*

Thesis 3: The Cotranscriptional Machinery

Waters names prokaryotic gene expression and the specification of pre-mRNA as the clearest case for an exclusive DNA causal specificity. But it turns out that even in prokaryotes and in the production of *preliminary* mRNAs in eukaryotic cells, the DNA sequence does not exclu-

sively specify its products. There exist RNA modifying mechanisms in bacteria and transcription in eukaryotes is carried out by what has come to be known as the cotranscriptional machinery or mRNA assembly line. This means that there is indeed no time at which a fully sequenced pre-mRNA exists in the cell.

Although all mechanisms of DNA expression and regulation have their biochemical identity, all of them feature in an 'extensive network of coupling among gene expression machines'. It is now clear that alternative splicing does not represent a distinct and decoupled step but is tightly coupled to transcription, polyadenylation, RNA editing, RNA surveillance and transport.

> Recent studies suggest that this task is facilitated by a combination of protein – RNA and protein – protein interactions within a 'mRNA factory' that comprises the elongating RNA polymerase and associated processing factors. This 'factory' undergoes dynamic changes in composition as it traverses a gene and provides the setting for regulatory interactions that couple processing to transcriptional elongation and termination. (Bentley 2005, 251)

Polymerase II and many other transcriptional proteins cooperate with the cotranscriptional processing factors. For instance, some SR proteins involved in the spliceosome have been known to react with transcription factors, while other proteins even exhibit a dual function as transcription and splicing regulator (Maniatis and Reed 2002; Bentley 2002). The cotranscriptional assembly of the spliceosome in this 'mRNA assembly line' suggests profound implications for the regulation of splice site choice. Splicing has also been implicated in downstream processes such as RNA transport, stability, translation, and location (Black 2003, 323). In addition, important links between RNA editing and other co- and posttranscriptional events that regulate gene expression have been suggested (Davidson 2002). *These co-transcriptional agents in combinatorial interplay with each other share causal specificity with genomic coding sequences protein synthesis through their involvement in sequence selection (e.g. splice-site specificity) and sequence creation (e.g. editing-site specificity).*

In summary, details into the processes of transcriptional activation, alternative splicing, *trans*-splicing, RNA editing, and translational recoding, among others, can show that *the specifying relationship between DNA and gene product is indirect, mediated and specifically intervened by other sequence specifying agents.*

4. Sequence is not Destiny

The view of specificity as selective and exclusive has recently given way to a highly distributed, modular, and combinatorial picture. Through a multitude of combinatorial associations and interactions macromolecules can expand on their intrinsic *molecular* functions to achieve more sophisticated and varied *cellular* functions. Especially in eukaryotes, the regulation of gene expression works by means of the *regulated recruitment* of *trans*-acting factors (proteins, RNAs, metabolites, and other environmental signals) into larger complexes and to *cis*-acting sequence modules, so that the specificity of an enzyme, a sequence, transcription or splicing factor comes to depend on its proper recruitment and combinatorial interaction (Ptashne and Gann 2002; Buchler, Gerland and Hwa 2003). A gene product is specified as much by the genomic template as by the differential recruitment of agents of genome expression mechanisms that *activate*, *select*, and *modify* the transcript specifically. The function of a transcription factor may not depend on a particular promoter sequence on which it can bind but its interacting protein partners only a few of which need a particular binding specificity. This versatility of the genome by means of the combinatorial complexity of its regulation resolves the 'N-value' paradox (Claverie 2001; Harrison *et al.* 2002). The proportion of protein-coding sequences seems to decline as a function of complexity, but the ratio of non-coding DNA rises, and so does the number of functional, regulatory roles played by non-coding RNAs and other cellular factors that help to translate sequential information encoded in the genome into developmental complexity (Mattick 2004).

> There is increasing awareness that multiple, often overlapping mechanisms exist for amplifying the repertoire of protein products specified through the mammalian genome. An expanding array of processing and targeting mechanisms is now emerging, each representing a potentially important restriction point in the regulation of eukaryotic gene expression, and each expanding the possibilities specified by the literal code of the genome. These co- and posttranscriptional regulatory events include capping, alternative splicing, differential polyadenylation, RNA editing, nuclear export, alternative decay and degradation pathways, as well as alterations in ribosomal loading or translation. (Davidson 2002)

These template-modifying mechanisms are the rule rather than exceptions in normal gene product synthesis. Latest estimates place the number of alternatively spliced human genes to over 70%, with around 100 genes with over 5,000 splice variants. The *Drosophila* cell adhesion molecule gene (*DSCAM*) can produce more than 38,000 temporally and

spatially regulated splice variants (Kapranov *et al.* 2005; Celottoa and Graveley 2001; Leipzig, Pevzner and Heber 2004). In some mitochondria of higher plants, a total of more than 1000 C-to-U changes are known to alter the total coding text of the entire RNA population, mostly within the first two positions of codons, hence changing the amino acid. The RNA editing of cellular RNAs of many eukaryotic organisms can result in up to 50% modified adenosine residues in a transcript (Gott and Emeson 2000). This form of editing is absolutely critical for normal brain function in humans and very prevalent in mammalian cells with a suspected 85% of all mRNAs targeted (Athanasiadis, Rich and Maas 2004). A recent study shows a highly overlapping, complex, and dynamic nature of the human transcriptome, where one base pair can be part of many transcripts emanating from both strands of the genome. The data further suggest that base pairs normally thought to contribute to transcripts from different genes can be joined together in a single RNA molecule (Kapranov *et al.* 2005; Kampa *et al.* 2004; Cheng *et al.* 2005).

Suspecting this, the National Human Genome Research Institute launched the ENCyclopedia Of DNA Elements Project (ENCODE) in 2003, which aims to identify comprehensively all the functional elements in the human genome, and has basically supported the picture described above (Gerstein *et al.* 2007).

The factors that interactively regulate genomic expression are on a par with coding information since they *co-specify* the linear sequence of the gene product together with the target DNA sequence. From this follows the radical thesis of 'molecular epigenesis': *Networks of genome regulation made up of cis-regulatory sequences, trans-acting factors and environmental signals causally specify the physical structure of a gene and the range of its products through the activation, the selective use, and, more radically, the creation of nucleotide sequence information* (Stotz 2006).

5. Epigenesis Does not Reduce to Genetics

Rosenberg (this issue) interprets the central dogma as an 'important version of genocentrism,' which he translates into the claim that 'nucleotide sequences carry semantic information.' Surprisingly, he does not understand this overused concept as selected effect (Maynard Smith 2000; Griffiths 2001) but as 'derived intentionality,' from which he mys-

tically infers 'the unique role of DNA in programming the embryo, and it is this unique programming role that gives the central dogma its significance.' While it remains unclear how 'specifying proteins' can become 'programming the embryo', Rosenberg expects opponents of genocentrism and defenders of the idea of epigenetic inheritance to defy the literal wording and to 'reject the central dogma's core claim, the negative thesis of once in the protein molecule, information cannot get out again.' I maintain instead that the epigenetic skeptic may legitimately oppose the dogma's implicit positive claim that, quote Rosenberg, 'only DNA and RNA can transfer information' and 'that it is always the genes that program the organism's development by directing protein synthesis.'

Rosenberg argues that epigenetic inheritance is a 'DNA directed mechanism . . . which does not convey information.' He uses as an example parental imprinting of target genes. Histone deacetylation, DNA methylation, and other chromatin modification processes, however, are much more diverse than Rosenberg gives them credit for. So is epigenetic cellular inheritance an important mechanism of multicellular organisms by which populations of cells maintain their pattern of differentiation. It is also known that DNA modification is one of the ways by which environmental agents, such as diet, temperature, tactile stimulation, and other factors, effect normal physiological reactions during development (Gilbert 2005; Jaenisch and Bird 2003; Fish *et al.* 2004). Lastly, DNA methylation is a crucial step in memory formation, such as conditioning, and is therefore dynamically regulated in the adult nervous system (Miller and Sweatt 2007). While a whole range of gene products, such as microRNAs and diverse proteins, are involved in chromatin and DNA modification, environmental signals are also present, and the concrete targeting processes are rarely fully known. We can not infer direct genetic control in any process in which some gene products are involved as part of a variety of stimuli! In any case, DNA modification is an important agent in protein synthesis and therefore conveys information, and most importantly, it helps us understand one of the mechanisms by which the genome is learning from experience.

Rosenberg then turns to another example of epigenetic inheritance, host imprinting of the brood-parasitic widowbird that learns the song from its finch foster parent. The real explanation, so Rosenberg claims, is provided by 'the genetically-encoded program for the neurology of singing.' Following this logic, genes must ultimately explain all learning! If we apply Waters's account of causation, however the *actual* differ-

ence-making cause between two potential bird songs learned by an offspring is the template song sung by the parent, not any neurological mechanism, which may be the same in many species of birds. Rosenberg goes so far as to describe the song as part of the bird's extended phenotype of the bird's genes. Even if many geneticists treat the phenotype as if it were a direct 'readout' of the genome, it remains a truism that the genotype interacts with the environment to construct the phenotype. Hardly a single protein is produced without necessary cellular signals. For most organisms, but the delicately breed model organism in the lab, the environment plays an expected and evolutionarily selected role in development (Gilbert 2003). Rosenberg points out the longevity of genetic inheritance but does not seem to grasp how ubiquitous non-genetic parental effects, niche construction and other extra-genetic inheritance systems are (Mousseau and Fox 2003; Jablonka and Lamb 2005; Odling-Smee, Laland and Feldman 2003). While Rosenberg misunderstands these systems as an 'alternative,' they are in fact complementary to genetic heredity. The whole discussion of host imprinting is full of biological errors and logical non-sequiturs.

Rosenberg misunderstands my thesis of molecular epigenesis as just adding *quantitative* information to the production of RNAs and proteins via the process of sequence *activation*. The main epigenetic challenge for the central dogma, however, comes from the *qualitative* processes of sequence *selection* and *creation*, which add or modify sequential *information* to the message.

6. Conclusion

Both Waters and Rosenberg misunderstand the principle of 'causal parity', which derives its name from Oyama's earlier call for 'parity of reasoning', when thinking about the roles of DNA elements and other developmental resources. She argued that if one of the above distinctions applies to some but not all DNA elements and also applies to some non-DNA influences in development, we should treat both the DNA and the non-DNA factors alike in the area of theory where the distinction is useful. In order to be able to follow this principle of parity it is essential not to build grand, metaphysical distinctions, like that between form and matter or information and matter, on top of the many empirical differences between the roles of DNA elements and the roles of other causal factors in development. DNA does play a distinctive set of roles in development, but it does not play just one role (partly because

DNA elements are themselves so diverse) and the important roles of those various DNA elements are sometimes played by non-DNA factors in development (Griffiths and Gray 2003, 421).

When distinguishing different causal processes (italized below) in an organism, there are always more than one agent fitting this causal role. Sequence specificity is held by DNA, but also by splicing and editing agents, as well as other regulatory mechanisms that are involved in modifying the primary sequence of RNA. Important for transgenerational reliability, many of the necessary factors involved in this expression are reliably reproduced in each generation next to DNA. The causal role of *inheritance* is carried out by DNA, histone and methylation patterns, structural cellular components, maternal RNA and transcription factors, provisioning of resources, preference induction (oviposition, imprinting on food, habitat, and mates), social learning, plus whatever else is provided by the parental generation in form of an 'ontogenetic niche' as a providing environment. *Enzymatic activity* has for the longest time been attributed to proteins alone but is, as we now know, regularly achieved by tertiary RNA structures (ribozymes). Protein transcription factors now have to share their fame with regulatory non-coding RNAs and inducing environmental factors such as lactose in the *regulation of genome expression*.

Waters's focus on the specificity of single genes in RNA and protein synthesis overstates the importance of single nodes, individual events, and isolated pathways over highly interactive genes, signaling and regulatory networks. Important are not so much the gene sequences but their differentiated expression (Meaney 2004). The central dogma has misdirected research into dogmatic pathway analyses involving a few protein-coding genes and led away from systems thinking (Werner 2005). Living systems are characterized through cyclic feedback networks and emergent organization (Bechtel in press).

> Feedback loops and back-up *pathways* have been invoked to account for these properties. […] A more flexible and fluid view of the relationships among these signaling and regulatory systems allows for the same net result *without invoking a predetermined mechanism for it*. The malleability and versatility of gene *networks* and their ability to find new solutions when constituents are changed, help to account for the properties of robustness, buffering and emergence. (Greenspan 2001, 386).

The bias of the last 50 years of genetic research in its focus on (protein) coding genes has neglected our growing understanding that the complexity of higher organisms lies not in its number of genes but within the flexibility, versatility, and reactivity of its whole genome.

Complexity is not encoded in the literal sequence of coding genes but in the processes that can amplify this information. These regulatory mechanisms involve, among other agents, a large number of different non-coding RNAs and non-coding DNA sequences with important binding or structural domains and even transcriptional capacity, for the longest time dismissed as 'junk' (Levine and Tjian 2003; Buchler, Gerland and Hwa 2003). 'We continue to learn new ways in which nature has exploited the specificity of interactions between RNA and nucleotide sequences. We now know that RNA, after being transcribed from DNA, can feed back to direct modifications of the genome. These modifications can be inherited through cell divisions and influence development' (Kawasaki and Taira 2004).

To conclude, the central dogma may not be literally wrong in its negative formulation, but its implicit positive formulation has unnecessarily restricted genetic research for too long. We are long overdue to change course.

> Isaac Newton might have liked the neat view of biological systems made up of dedicated components, with causal roles that can be studied in isolation, and in which particular starting conditions give rise to uniquely predictable responses. Charles Darwin, by contrast, might have felt more at home with the idea of a complex, emergent system made up of many non-identical components, with non-exclusive roles, non-exclusive relationships, several ways of producing any given output, and a great deal of slop along the way. We have been Newtonians for the past several decades in our thinking about gene action. It is time to become Darwinians. (Greenspan 2001, 386).

References

Athanasiadis A., Rich A., Maas S., 2004, 'Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome', *PSoS Biology*, 2(12):e391.

Bechtel W., (in press), 'Biological Mechanisms: Organized to Maintain Autonomy'. In: Boogerd F., *et. al.* (eds), *Systems Biology: Philosophical Foundations*, New York: Elsevier.

Bentley D., 2002, 'The mRNA Assembly Line: Transcription and Processing Machines in the Same Factory', *Current Opinion in Cell Biology*, 14(3):336-342.

Bentley D.L., 2005, 'Rules of Engagement: Co-transcriptional Recruitment of Pre-mRNA Processing Factors', *Current Opinion in Cell Biology*, 17(3): 251-256.

Black D.L., 2003, 'Mechanisms of Alternative pre-messenger RNA splicing', *Annual Reviews of Biochemistry*, 72:291-336.

Buchler N.E., Gerland U., Hwa T., 2003, 'On Schemes of Combinatorial Transcription Logic', *Proceedings of the National Academy of Sciences of the United States of America*, 100:5136-5141.
Celottoa A.M., Graveley B.R., 2001, 'Alternative Splicing of the Drosophila Dscam Pre-mRNA Is Both Temporally and Spatially Regulated', *Genetics*, 159:599-608.
Cheng J. *et. al.*, 2005, 'Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution', *Science*, 308:1149-1154.
Claverie J. M., 2001, 'Gene Number: What if There are Only 30,000 Human Genes?', *Science*, 291:1255-1257.
Crick F.H.C., 1958, 'The Biological Replication of Macromolecules', S*ymposia of the Society for Experimental Biology*, XII:138-163.
Crick F.H.C., 1970, 'Central Dogma of Molecular Biology', *Nature*, 227:561-563.
Davidson N.O., 2002, 'The Challenge of Target Sequence Specificity in C?U RNA Editing', *The Journal of Clinical Investigation*, 109(3):291-294.
Fish E.W. *et. al.*, 2004, 'Epigenetic Programming of Stress Responses through Variations in Maternal Care', *Annals of the New York Academy of Sciences*, 1036:167-180.
Gerstein M.B. *et. al.*, 2007, 'What is a Gene, post-ENCODE? History and Updated Definition', *Genome Research*, 17:669-681.
Gilbert S.F., 2003, 'The Reactive Genome'. In: Müller G.B., Newman S.A. (eds), *Origination of Organismal Form: Beyond the Gene in Developmental and Evolutionary Biology*, Cambridge, MA: The MIT Press, 87-101.
Gilbert S.F., 2005, 'Mechanisms for the Environmental Regulation of Gene Expression: Ecological Aspects of Animal Development', *Journal of Biosciences*, 30(1):65-74.
Gott J.M., Emeson R.B., 2000, 'Functions and Mechanisms of RNA Editing', *Annual Review of Genetics*, 34:499-531.
Greenspan R.J., 2001, 'The Flexible Genome', *Nature Reviews Genetic*, 2:383-387.
Griffiths P.E., 2001, 'Genetic Information: A Metaphor in Search of a Theory', *Philosophy of Science*, 68(3):394-412.
Griffiths P.E., Gray R.D., 2003, 'The Developmental Systems Perspective: Organism-Environment Systems as Units of Evolution'. In: Preston K., Pigliucci M. (eds), *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, Oxford / New York: Oxford University Press, 409-431.
Harrison P.M., *et. al.*, 2002, 'A Question of Size: the Eukaryotic Proteome and the Problems in Defining it', *Nucleic Acids Research,* 30(5):1083-1090.
Jablonka E., Lamb M.J., 2005, *Evolution in Four Dimenesions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, Cambridge, MA: The MIT Press.
Jaenisch R., Bird A., 2003, 'Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals', *Nature Genetics*, 33 (Suppl):245-254.

Kampa D. *et. al.*, 2004, 'Novel RNAs Identified from an in-Depth Analysis of the Transcriptome of Human Chromosomes 21 and 22', *Genome Research*, 14:331-342.
Kapranov P. *et. al.*, 2005, 'Examples of the Complex Architecture of the Human Transcriptome Revealed by RACE and High-density Tiling Arrays', *Genome Research*, 15:987-997.
Kawasaki H., Taira K., 2004, 'Induction of DNA Methylation and Gene Silencing by Short Interfering RNAs in Human Cells', *Nature*, 431:211-217.
Ladd A.N., Cooper T.A., 2002, 'Finding Signals that Regulate Alternative Splicing in the Post-Genomic Era', *Genome Biology*, 3(11):1-16.
Laughlin R.B., 2005, *A Different Universe (Reinventing Physics from the Bottom down)*, New York: Basic Books.
Leipzig J., Pevzner P., Heber S., 2004, 'The Alternative Splicing Gallery (ASG): Bridging the Gap between Genome and Transcriptome', *Nucleic Acids Research*, 32(13): 3977-3983.
Levine M., Tjian R., 2003, 'Transcription Regulation and Animal Diversity', *Nature*, 424:147-151.
Maniatis T., Reed R., 2002, 'An Extensive Network of Coupling among Gene Expression machines', *Nature*, 416:499-506.
Mattick J.S., 2004, ' RNA Regulation: a New Genetics?', *Nature Reviews Genetics*, 5(4):316-323.
Maynard Smith J., 2000, 'The Concept of Information in Biology', *Philosophy of Science*, 67(2):177-194.
Meaney M.J., 2004, 'The Nature of Nurture: Maternal Effect and Chromatin Modelling'. In: Cacioppo J.T., Berntson G.G. (eds), *Essays in Social Neuroscience*, Cambridge, MA: The MIT Press, 1-14.
Miller C.A., Sweatt J.D., 2007, 'Covalent Modification of DNA Regulates Memory Formation', *Neuron*, 53:857-869.
Mousseau T.A., Fox C.W. (eds), 2003, *Maternal Effects as Adaptations*, Oxford: Oxford University Press.
Odling-Smee F.J., Laland K.N., Feldman M.W., 2003, *Niche Construction: The Neglected Process in Evolution*, Princeton, NJ: Princeton University Press.
Ptashne M., Gann A., 2002, *Genes and Signals*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Robert J.S., 2004, *Embryology, Epigenesis and Evolution: Taking Development Seriously*, Cambridge: Cambridge University Press.
Sarabhai A.S., Stretton A.O.W., Brenner S., Bolleh A., 1964, 'The Colinearity Hypothesis', *Nature*, 201:13-17.
Stotz K., 2006, 'With Genes Like That, Who Needs an Environment? Postgenomics' Argument for the Ontogeny of Information', *Philosophy of Science*, 73(5), proceedings PSA 04, December 2006. Preprint: http://philsci-archive.pitt.edu/archive/00002456/.
Stotz K., (submitted), '2001 and All That: a Tale of a Third Science', *Biology & Philosophy*.
Waters C.K. (forthcoming a), 'A Pluralist Interpretation of Gene-centered

Biology'. In: Kellert S., Longino H.E., Waters C.K. (eds), *Scientific Pluralism*, Minneapolis: University of Minnesota Press.
Waters C.K., (forthcoming b), 'Causes That Make a Difference', *Journal of Philosophy*.
Werner E., 2005, 'Genome Semantics, in Silico Multicellular Systems and the Central Dogma', *FEBS Letters*, 579(8):1779-1782.