

# A signal-responsive (cognitive) systems view of the genome

James A. Shapiro

University of Chicago

<http://shapiro.bsd.uchicago.edu>

# Exceptions to linear, unitary, atomistic coding in genomes

- Alternative promoters, splicing patterns, RNA and protein processing pathways
- Combining separate RNA and DNA segments to form novel coding domains
- Systemic rather than atomistic view of genotype and phenotype:
  - DNA, RNA and proteins composed of multiple interacting domains and signals
  - All phenotypes produced by multiple interacting molecules encoded in the genome & contingent structures not encoded in the genome

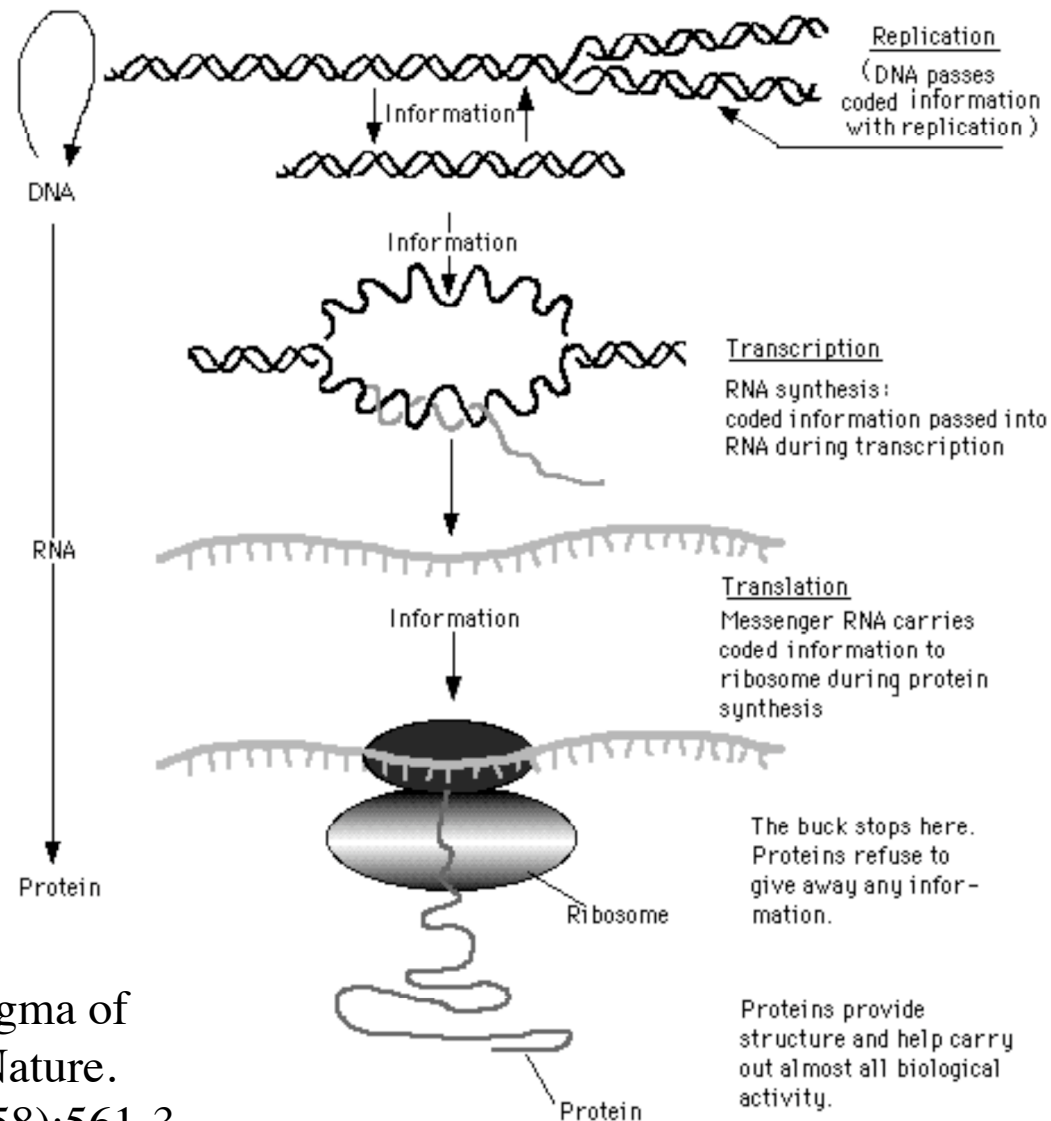
# Critical Issues in Genome Conceptualization

1. Re-evaluating long-held DNA-centric views on the role of the genome in the cellular information economy
2. Incorporating the ability of cells and organisms to deal with predictable and unpredictable contingencies by responding to inputs containing information about the environment and the status of the biological system

# Implications (and Irony) of Molecular Biology

- Revisiting the Central Dogma shows that information transfer is interactive and involves all molecular classes and compartments of the cell (no Cartesian dualism).
- External and internal signals modulate all kinds of cellular information transfer (signal transduction).
- Genome operates as RW memory system at various time scales based on obligatory interactivity of DNA with other molecules.
- Combinatorial nature of biological specificity ( $\neq$  from lock-and-key) enhances the information content and processing power of molecular interactions.

# The Central Dogma of Molecular Biology



Crick, F. Central dogma of molecular biology. *Nature*. 1970 Aug 8;227(5258):561-3.

# Revisiting the Central Dogma

## Conventional expression of the Central Dogma of Molecule Biology:

(DNA  $\implies$  2X DNA)  $\implies$  RNA  $\implies$  Protein  $\implies$  Phenotype

## Contemporary statements of molecular information transfer in cell:

1. DNA + 0  $\implies$  0
2. DNA + Protein + ncRNA  $\implies$  Chromatin
3. Chromatin + Protein + ncRNA  $\implies$  DNA replication, chromatin maintenance/reconstitution
4. Protein + RNA + lipids + small molecules  $\implies$  Signal transduction
5. Chromatin + Protein + signals  $\implies$  RNA (primary transcript)
6. RNA + Protein + ncRNA  $\implies$  RNA (processed transcript)
7. RNA + Protein + ncRNA  $\implies$  Protein (primary translation product)
8. Protein + nucleotides + Ac-CoA + SAM + sugars + lipids  $\implies$  Processed and decorated protein
9. DNA + Protein  $\implies$  new DNA sequence (mutator polymerases)
10. Signals + Chromatin + Protein  $\implies$  new DNA structure (stimulated DNA rearrangements)
11. RNA + Protein + chromatin  $\implies$  new DNA structure (retrotransposition, retroduction, retrohoming)
12. Signals + chromatin + proteins + ncRNA + lipids  $\implies$  nuclear/nucleoid localization of genome domains

## SUMMARY:

DNA + Protein + ncRNA + signals + other molecules  $\iff$  Genome Structure & Phenotype

# Signals at all steps of cell information transfer

<b>INFORMATION TRANSFER</b>	<b>SIGNAL (example)</b>	<b>REFERENCE</b>
<b>DNA + Protein + ncRNA ==&gt; chromatin</b>	<b>Bacterial infection</b>	Hamon MA, Cossart P. Histone modifications and chromatin remodeling during bacterial infections. <i>Cell Host Microbe</i> . 2008
<b>Chromatin + Protein + ncRNA ==&gt; DNA replication, chromatin maintenance/reconstitution</b>	<b>Ecdysone</b>	Sun, J., Smith, L., Armento, A., Deng, W.-M. 2008. Regulation of the endocycle/gene amplification switch by Notch and ecdysone signaling. <i>Journal of Cell Biology</i> 182 (5), pp. 885-896.
<b>Signals + Protein + RNA + lipids + small molecules ==&gt; Signal transduction</b>	<b>Too many to count (by definition)</b>	
<b>Chromatin + Protein + signals ==&gt; RNA (primary transcript)</b>	<b>Extracellular matrix</b>	Xu R, Spencer VA, Bissell MJ. Extracellular matrix-regulated gene expression requires cooperation of SWI/SNF and transcription factors. <i>J Biol Chem</i> . 2007 May 18;282(20):14992-9
<b>RNA + Protein + ncRNA ==&gt; RNA (processed transcript)</b>	<b>Insulin</b>	Patel NA, Apostolatos HS, Mebert K, Chalfant CE, Watson JE, Pillay TS, Sparks J, Cooper DR. Insulin regulates protein kinase CbetaII alternative splicing in multiple target tissues: development of a hormonally responsive heterologous minigene.
<b>RNA + Protein + ncRNA ==&gt; Protein (primary translation product)</b>	<b>LPS</b>	Chen YL, Huang YL, Lin NY, Chen HC, Chiu WC, Chang CJ. Differential regulation of ARE-mediated TNFalpha and IL-1beta mRNA stability by lipopolysaccharide in RAW264.7 cells. <i>Biochem Biophys Res Commun</i> . 2006 Jul 21;346(1):160-8.
<b>Protein + nucleotides + Ac-CoA + SAM + sugars + lipids ==&gt; Processed and decorated protein</b>	<b>Osmotic, hypoxic, heat, oxidative and genotoxic stresses</b>	Tempé D, Piechaczyk M, Bossis G. SUMO under stress. <i>Biochem Soc Trans</i> . 2008 Oct;36(Pt 5):874-8.
<b>DNA + Protein ==&gt; new DNA sequence (mutator polymerases and repair processes)</b>	<b>Antigen binding</b>	Odegard VH, Schatz DG (2006) Targeting of somatic hypermutation. <i>Nat Rev Immunol</i> 6: 573-583.
<b>Signals + Chromatin + Protein ==&gt; new DNA structure</b>	<b>Aerobic starvation</b>	Maenhaut-Michel, G. & Shapiro, J.A. 1994c. The roles of starvation and selective substrates in the emergence of araB-lacZ fusion clones. <i>EMBO J</i> . 13, 5229-5239.
<b>Signals + RNA + Protein + chromatin ==&gt; new DNA structure (retrotransposition, retroduplication, retrohoming)</b>	<b>Mating pheromone</b>	Kinsey PT, Sandmeyer SB. Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3. <i>Genetics</i> . 1995 Jan;139(1):81-94.
<b>Signals + chromatin + proteins + ncRNA + lipids ==&gt; nuclear/nucleoid localization</b>	<b>TNF alpha</b>	Gao C, Cheng X, Lam M, Liu Y, Liu Q, Chang KS, Kao HY. Signal-dependent regulation of transcription by histone deacetylase 7 involves recruitment to promyelocytic leukemia protein nuclear bodies. <i>Mol Biol Cell</i> . 2008 Jul;19(7):3020-7.

# Genome as RW Memory System: Functional Content

- Data files (Protein and RNA coding sequences)
- Signals formatting data files for transcriptional and post-transcriptional access
- Signals formatting genome for organization as chromatin
- Signals formatting genome for localization in the nucleus or nucleoid
- Signals formatting genome for replication and transmission
- Signals formatting genome for proofreading and repair
- Signals formatting genome for restructuring

Genome System Architecture - taxonomically specific

# Genome as RW Memory System: Time Scales

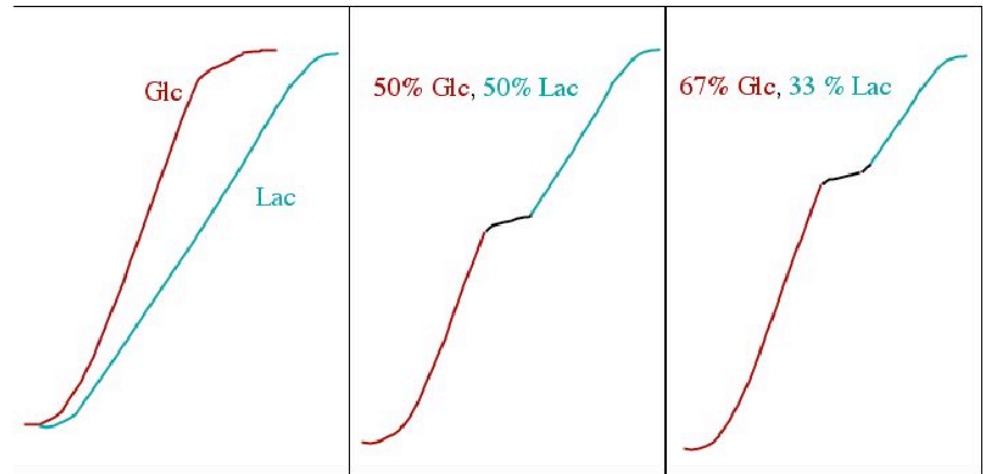
- Within cell cycle by adjustment of DNA binding protein complexes (e.g. replication factors, transcription factors, cell cycle and checkpoint monitors, cohesins)
- Over multiple cell cycles by chromatin reformatting (DNA methylation, histone modification, chromatin binding proteins)
- Within organismal life cycle and over evolutionary time by natural genetic engineering (immune system, antigenic variation, macronuclear development; duplications, transpositions, chromosome rearrangements)

# Genome as RW Memory System: Cellular Responsiveness of Genome Restructuring

transposon and retrotransposon activation	<i>Drosophila melanogaster</i>	Georgiev, P.G. et al. 1990. Mitomycin C induces genomic rearrangements involving transposable elements in <i>Drosophila melanogaster</i> . <i>Molec. Gen. Genet.</i> 220: 229-233.
Alu retransposition	<i>Homo sapiens</i>	Hagan, C.R., R.F. Sheffield, & C.M. Rudin. 2003. Human Alu element retrotransposition induced by genotoxic stress. <i>Nat. Genet.</i> 35: 219-220.
mouse mammary tumor virus (MMTV) activation	<i>Mus musculus</i>	Truss, M., G. Chalepakis, M. Beato. 1992. Interplay of steroid hormone receptors and transcription factors on the mouse mammary tumor virus promoter. <i>J. Steroid Biochem. Mol. Biol.</i> 43: 365-378.
retrotransposon activation	<i>Nicotiana tabacum</i>	Beguiristain, T. et al. 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. <i>Plant Physiol.</i> 127: 212-221.
MITE DNA transposons	rice	Lin, X. et al. 2006. In planta mobilization of mPing and its putative autonomous element Pong in rice by hydrostatic pressurization. <i>J Exp Bot.</i> 57: 2313-2323.
Retrotransposon activation	<i>N. tabacum</i>	Sugimoto, K., S. Takeda, & H. Hirochika. 2000. MYB-related transcription factor NtMYB2 induced by wounding and elicitors is a regulator of the tobacco retrotransposon Tto1 and defense-related genes. <i>Plant Cell</i> 12: 2511-2527.
Transposon and retrotransposon activation	various plants	Grandbastien, M.-A. 1998. Activation of plant retrotransposons under stress conditions. <i>Trends Plant Sci.</i> 3: 181-187. Hirochika, H. 1993. Activation of tobacco retrotransposons during tissue culture. <i>EMBO J.</i> 12: 2521-2528. Kikuchi, K. et al. 2003. The plant MITE mPing is mobilized in anther culture. <i>Nature</i> , 421: 167-170.

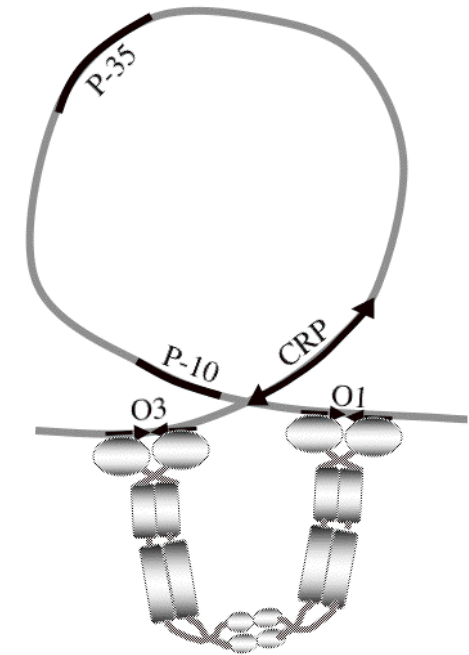
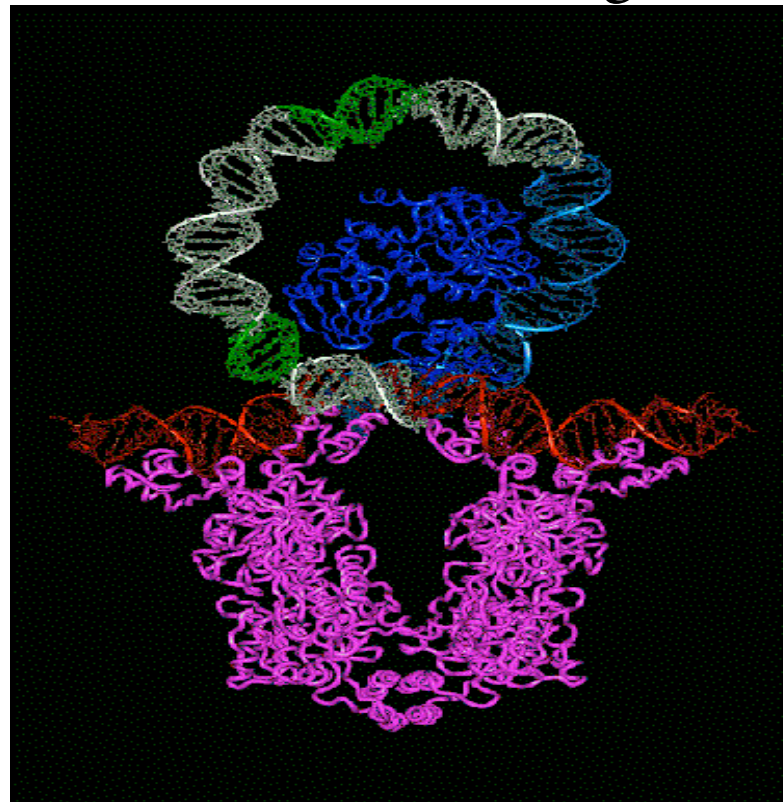
Excerpt of Table 1, Shapiro, J.A. 1999. Revisiting the Central Dogma in the 21st Century. *Ann. NY Acad Sci*, *in press*.

# Combinatorial nature of biological specificity put to cognitive use



Monod, 1942: Diauxic growth. *E. coli* distinguishes glucose and lactose

Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. Mitchell Lewis, et al. *Science*, Vol. 271(Mar. 1, 1996), pp. 1247-1254.



# Significance of the combinatorial nature of biological specificity

- Limitless specificities can be built from a finite number of components.
- Domain repeats in biological molecules (DNA, RNA, proteins) are essential elements of system architecture.
- Each molecule is a system of domains, not a unitary structure.
- Combinatorial interactions of domains and molecules can each assume various logical forms (arithmetic, geometric, Boolean, Bayesian, fuzzy, etc.)

# Significance of cognitive systems view of genome operation

- “Turtles all the way down” = no informatic units, only systems (proteins, RNAs, genetic loci, traits); system > collection of many parts, has functional architecture and can respond logically to inputs.
- Biological functions have a contingent, historical nature rather than hardwired outcomes ==> molecular structure alone insufficient to explain maintenance of information content and reliable operation of system.
- Need to find additional sources of hereditary and developmental reliability in response to unlimited range of possible inputs (features of system architecture).
- Evolution by natural genetic engineering responses to challenge (McClintock, 1984; Shapiro, 2005)