

Editorial

BEYOND THE CENTRAL DOGMA

The central dogma, DNA makes RNA makes protein, has long been a staple of biology textbooks. More recently, this paradigm has been extended from individual genes to whole genomes by advances in genomic technologies. For example, probing of DNA microarrays accomplishes on a large scale what was previously achieved for single genes using filter hybridization. High-throughput technology, not breakthrough biology, is becoming synonymous with genomics.

The longevity of the central dogma has also meant that basic computational tools for analyzing sequence data reached maturity before whole genomic sequences became available. For example, gene finding systems were introduced over a decade ago (Fields and Soderlund, 1990), and the importance of the problem attracted many computational biologists. As a result of numerous incremental advances made already, diminishing returns may be expected for this problem. A similar situation appears to hold for database searching (Schaffer *et al.*, 2001). Thus, the computational tools that are most widely used now may be difficult to improve upon in the post-genomic era.

Technologies based on textbook biology will continue to generate opportunities in bioinformatics. However, more exciting prospects may come from new discoveries that extend or even violate the central dogma. Consider developmental biology. The central dogma says nothing about the differences between the cells in a human body, as each one has the same DNA. However, recent findings have begun to shed light on how these differences arise and are maintained, and the biochemical rules that govern these differences are only being worked out now. The emerging understanding of developmental inheritance follows a series of fundamental discoveries that have led to a realization that there is more to life than the central dogma.

The central dogma was first challenged by the discovery of reverse transcription (Baltimore *et al.*, 1970; Temin and Mizutani, 1970). Thought at the time to be peculiar to retroviruses, we now know from large-scale sequencing that our genome contains an order of magnitude more copies of sequences encoding reverse transcriptase than sequences encoding all other proteins combined (Lander *et al.*, 2001)! Half of our genome is devoted to retroelements and their remnants, compared to only a few percent devoted to gene coding regions. Humans are not alone in having genomes dominated by retroelements. The genomes of many plants are even more infested: for example, retrotransposons occupy about 80% of the maize genome (SanMiguel *et al.*, 1996).

With so much genomic territory taken over by selfish elements, they are prime candidates for involvement in important genetic processes. One example is the propagation of silencing along the inactivated X chromosome of mammalian females: abundant LINE-1 retrotransposons were proposed to act as 'way-stations' (Lyon, 1998). Evidence in support of this idea was obtained by analyzing genomic sequence data (Bailey *et al.*, 2000), an illustration of how thinking about genetic mechanisms creates opportunities in bioinformatics. Selfish elements also reveal evolutionary processes that continue to shape genomes: arguably the major scientific story of the draft human genome sequence was the history of retrotransposon evolution (Lander *et al.*, 2001), a story missed by others who may have been so focused on the genes that they overlooked the junk (Venter *et al.*, 2001).

The success of selfish DNA elements does not mean that our genomes are entirely at their mercy. A widespread view is that genomes are protected by an immunity system (Yoder *et al.*, 1997). Among the weapons that are thought to help protect genomes, especially in plants, are DNA methyltransferases, enzymes that mark sequences for silencing by covalent modification. Silencing of retrotransposon transcription, which must precede reverse transcription and integration, should be an effective defense against their mobilization. A major unsolved problem has been the basis for recognition of transposons and their ilk by the DNA methylation machinery. Without obvious sequence cues, it has been difficult to understand how a genome defense system protects against invaders. This question goes beyond DNA methylation: organisms such as the fruit fly, which has an almost unmethylated genome, may effectively prevent transposition by packaging retroelements in silent chromatin (van Steensel *et al.*, 2001).

Recently, a surprising solution to the problem of retroelement recognition has been proposed: RNA interference (RNAi). First elucidated in the nematode, where genes could be shut down by introduction of double-stranded RNA, this powerful gene silencing technique is now known to utilize enzymatic machinery that is common to animals and plants (Carthew, 2001). Small interfering RNAs (siRNAs) of only 22–25 bp can traverse intracellular spaces to enter cells and trigger rapid degradation of homologous RNAs. The same mechanism underlies post-transcriptional gene silencing (PTGS) in plants, where unintended post-transcriptional silencing of transgenes has been the bane of genetic engineers for over a decade. Thus, PTGS appears to be a natural mechanism for defending against RNA-based invaders. In addition, PTGS may be involved in the recognition and targeting of genomic DNA sequences: siRNA made in the cytoplasm would be targeted to the nucleus where it guides a DNA methyltransferase to covalently modify

its homologous DNA (Matzke *et al.*, 2001). Whereas the central dogma begins with DNA, in this hypothesized RNAi-based process, DNA participates only at the end.

Perhaps the simplest exceptions to the central dogma are prions, proteins that undergo heritable conformational changes and seed polymeric forms of themselves (Lindquist, 1997). Prions were discovered in the search for the infectious agents of scrapie and mad cow disease, and their existence was confirmed by demonstrating protein inheritance in yeast. Although no nucleic acid component has been reported for scrapie, these searches were carried out before the discovery of RNAi, raising the possibility that siRNA guides that have eluded detection are causally involved in prion diseases.

RNAi is only the most recent RNA-based phenomenon to grab the attention of biologists. It is widely believed that an ancestral ribozyme-based 'RNA world' has been mostly supplanted by protein enzymes (Woese, 2001). In addition to protein synthesis, several RNA-based processes are known: a large structural RNA, Xist, coats the inactive X chromosome (Mlynarczyk and Panning, 2000), and numerous small RNAs are involved in diverse processing reactions (Eddy, 2001). RNA-based regulatory mechanisms have recently been documented, including transcription from the opposite strand which represses the Xist locus (Mlynarczyk and Panning, 2000), and siRNAs which repress production of a sperm-specific protein in flies (Aravin *et al.*, 2001). Clearly, the dogmatic view of RNA as playing merely intermediary roles in the synthesis of protein is becoming increasingly outmoded. Yet the analysis of genomic sequences to identify non-coding RNAs is still in its infancy (Eddy, 2001).

Genetic mechanisms that challenge the central dogma contribute to the complexity of eukaryotic organisms. However, when we get down to the nitty-gritty of a biological mechanism, what can emerge is elegant in its simplicity. Indeed, it now appears that all of the examples of gene silencing that I have mentioned rely on a simple on/off code. Recall that DNA comprises less than half of the chromosome, because it wraps tightly around a core octamer of four histones to form a nucleosome. Each histone has an N-terminal tail that exits from the nucleosome core, and the tail of histone H3 is covalently and permanently modified by methylation of two lysine residues, K4 and K9 (Jenuwein and Allis, 2001). Methylation of K4 but not K9 is found on nucleosomes that are transcriptionally active and methylation of K9 but not K4 is found on nucleosomes that are present in silent chromatin. Over the past few months, evidence has emerged that this simple difference distinguishes active versus silent chromatin, whether it is constitutively silent chromatin found around centromeres (Jenuwein and Allis, 2001) or chromatin found on the inactive but not the active mammalian X chromosome (Heard *et al.*, 2001).

By heritably affecting DNA accessibility, nucleosomes appear to be ultimately responsible for maintaining differences in gene expression that occur during development. Nucleosomes are ubiquitous components of eukaryotic chromosomes, and so there is reason to expect that the histone code is a generally applicable. Centromeres, the points on chromosomes that are responsible for mitotic movements, are exceptional, being inhabited by nucleosomes containing an H3-like histone that replaces H3 (Henikoff *et al.*, 2001). Centromeric sequences are notoriously repetitive and diverse, and yet all centromeres and only centromeres contain these specialized nucleosomes. Centromeres appear to be inherited by the continued presence of centromeric nucleosomes, with DNA sequence playing at most a secondary role. Thus, mitosis, a defining feature of eukaryotes, may rely upon a protein-based inheritance mechanism.

I have little doubt that these exciting developments will continue to open up new areas for computational biologists willing to look beyond the central dogma.

REFERENCES

- Aravin,A.A., Naumova,N.M., Tulin,A.V., Vagin,V.V., Rozovsky,Y.M. and Gvozdev,V.A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* genome. *Curr. Biol.*, **11**, 1017–1027.
- Bailey,J.A., Carrel,L., Chakravarti,A. and Eichler,E.E. (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. USA*, **97**, 6634–6639.
- Baltimore,D. (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature*, **226**, 1209–1211.
- Carthew,R.W. (2001) Gene silencing by double-stranded RNA. *Curr. Op. Cell Biol.*, **13**, 244–248.
- Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
- Fields,C.A. and Soderlund,C.A. (1990) gm: A practical tool for automating DNA sequence analysis. *Comput. Appl. Biosci.*, **6**, 263–270.
- Heard,E., Rougeulle,C., Amaid,D., Avner,P., Allis,C.D. and Spector,D.L. (2001) Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell*, **107**, 727–738.
- Henikoff,S., Ahmad,K. and Malik,H.S. (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098–1102.
- Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
- Lander,E.S., Linton,L.M., Birren,B. and Nusbaum,C. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Lindquist,S. (1997) Mad cows meet psi-chotic yeast: the expansion of the prion hypothesis. *Cell*, **89**, 495–498.
- Lyon,M.F. (1998) X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet.*, **80**, 133–137.

- Matzke,M., Matzke,A.J.M, and Kooter,J.M. (2001) RNA: Guiding gene silencing. *Science*, **293**, 1080–1083.
- Mlynarczyk,S.K. and Panning,B. (2000) X inactivation: Tsix and Xist as yin and yang. *Curr. Biol.*, **10**, R899–R903.
- SanMiguel,P., Tikhonov,A., Jin,Y.K., Motshoulskaia,N., Zakharov,D., Melake-Berhan,A., Springer,P.S., Edwards,K.J., Lee,M., Avramova,Z. and Bennetzen,J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.
- Schaffer,A.A., Aravind,L., Madden,T.L., Shavirin,S., Spouge,J.L., Wolf,Y.I., Koonin,E.V. and Altschul,S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.*, **29**, 2994–3005.
- Temin,H.M. and Mizutani,S. (2001) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature*, **226**, 1211–1213.
- van Steensel,B., Delrow,J. and Henikoff,S. (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nature Gen.*, **27**, 304–308.
- Venter,J.C., Adams,M.D., Myers,E.W. and Li,P.W. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
- Woese,C.R. (2001) Translation: in retrospect and prospect. *RNA*, **7**, 1055–1067.
- Yoder,J.A., Walsh,C.P. and Bestor,T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.

Steven Henikoff
Howard Hughes Medical Institute
Fred Hutchinson Cancer Research Center
Seattle WA 98109-1024, USA
E-mail: steveh@fhcrc.org