

Systematic discovery of the grammar of translational inhibition by RNA hairpins

Clark Jeffries^{1*}, Diana O. Perkins², Michael Jarstfer³

¹Renaissance Computing Institute and School of Pharmacy, Campus Box 7360, University of North Carolina at Chapel Hill, Chapel Hill NC 27599-7360 USA

Email: clark_jeffries@med.unc.edu , FAX 919-843-7624, TEL 919-260-6576

²Department of Psychiatry, University of North Carolina at Chapel Hill, USA

Email: diana_perkins@med.unc.edu

³School of Pharmacy, University of North Carolina at Chapel Hill, USA

Email: jarstfer@email.unc.edu

* Corresponding author

Research supported by: (Jeffries) 5-P20-RR020751-01-02 from the National Institutes of Health/National Center for Research Resources; (Perkins) USP#SMH01905-01 North Carolina Foundation of Hope, Anonymous donor administered through the San Francisco Foundation; (Jarstfer) Elsa U. Pardee Foundation and NSF MCB 0446019. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health/National Center for Research Resources or the other supporting Foundations.

Introduction

This paper addresses the problem of discovering from databases the organizational principles of regulation of gene expression. Text-based searches for gene relationships inherently must be artificially terminated. By contrast, we present a hybrid algorithm of text and biochemical alignment searches that inherently terminates. The product of the algorithm is a gene organization hypothesis.

The hybrid search works, we assert, because some genes are not just "selfish" (seeking above all else their own survival and proliferation in future generations); they are also "devious" in the sense that their transcription byproducts automatically inhibit translation of other genes with counteracting functions (so they quietly undermine opposing genes). In terms of control theory in cell biology, this would be an example of a positive feedforward mechanism, resulting in a "fervid" response as described in (Sinclair NR, 1993). The mechanisms include microRNA (miRNA) hairpins and also the main emphasis of this paper: larger regulatory hairpins from introns. Three well-formed hairpins from one intron of pro-apoptosis gene PAWR (GenelD 5074) are shown in Fig. 3. The first two have stems that resemble Alu repeat consensus sequence on one side and a sequence that is approximately the reverse complement (revcom) of an Alu, that is, an Alurc, on the other. The third hairpin, however, comes from an intronic portion of PAWR that is first Alu-like, then about 100 other nucleotides (nts), then about 240 nt of an Alurc-like pattern (a full Alu or Alurc is about 285 nt). Thus the stem of the third hairpin in Fig. 3 is not fully Alu/Alurc, but is restricted to only a part of an Alu/Alurc double-stranded RNA (dsRNA).

The predicted hairpins provide feedstock dsRNA for an miRNA pathway that eventually causes inhibition of other genes, as considered in (Hannon, 2002). In particular, the anti-apoptosis gene BIRC4 (331) is expressed at the same time as PAWR in certain cancer cell lines, as discussed below. According to the output of the hybrid algorithm, PAWR can be hypothesized to inhibit BIRC4 by means of an interference mechanism (and possibly other mechanisms). Thus once apoptosis begins, this double mechanism including PAWR expression and automatic BIRC4 repression might drive the cell efficiently and irreversibly to disintegration (hence fervid, positive feedforward control). Such deviousness by some genes might suggest a logical module that helps researchers to organize concepts of gene expression.

That there is a need to find organizational principles for gene expression including various types of control agents is self-evident. The purpose of this paper is progress in that direction with emphasis on the miRNA pathway applied primarily to larger, non-miRNA hairpins (miRNA hairpins contain ~70 to ~120 nt). Again, a fundamental assumption is that portions of dsRNA of various sources might be exported to the cytosol through an approximately common pathway and participate in translational inhibition of target mRNAs (Hannon, 2002). The dsRNA sources can include, intronic single-stranded RNA (ssRNA) that folds to make dsRNAs, or long dsRNAs generated by bidirectional transcription of repetitive sequences from adjacent promoters (Grewal and Rice, 2004). The subsequent miRNA pathway might lead to binding by chemical affinity to target mRNA and then repression of target gene expression by: inhibition of ribosomal processing; hybridization with and accelerated degradation of mRNA in a processing body (PB); or heterochromatin alteration (Murchison and Hannon, 2004), (Pillai RS, 2005), (Grewal and Rice, 2004). In this paper the target sequence is assumed to lie in the 3'UTR, but also known are related mechanisms including instances of a functional target in the 5'UTR (Jin et al., 2004) and transcription factor binding site in an intron (Xu Y, 1999) (St Clair DK, 2002). The full range of types of sequence specific pre-mRNA and mRNA targeting is no doubt very large. In addition is the recent finding that half or even more of the human genome might be transcribed, antisense as well as sense; this is a fraction that dwarfs the one or two percent of the genome that contains conventional genes. For a survey of ideas and current research, see (Mattick, 2005) and other articles on noncoding RNA (ncRNA) in the same special issue of *Science*. Since there are many small and some large sequences in the genome with nearby revcom sequences, there is potentially a vast supply of dsRNAs that could participate in regulatory systems. At any rate, the bipartite graph of such relationships (regulated genes versus blocking agents) would ideally be

part of an organized theory or grammar of combinatorial control, the description of which could proceed in terms of its syntax and semantics. By “syntax” we mean sets of rules starting with the list of permitted combinations of nucleotides (nts) that make the allowed sequences (words) engaged as blocking agents in translational inhibition control. In the case of miRNAs, this syntactic description might, for example, include the set of miRNAs and their mature regions with “seeds” (short nt sequences that hybridize with target sequences in an mRNA (Lewis et al., 2003)). The next level would include the rules for permitted tables of genes versus control agents (sentences in a relational grammar). Higher levels that make gradual or abrupt control of gene expression possible can be imagined. By “semantics” we define the union of all levels that somehow endows meaning (life) to syntactically correct sets of symbols.

In gene expression theory just as in computer science (Aho, 1986), finding formal yet useful definitions of semantics and meaning is difficult, necessitating human intuition and understanding. There are similarities but also extreme differences between gene expression systems and computer systems; for a given set of tasks, humans usually make computers to be as simple and efficient as possible, while nature often allows redundancy, tolerates errors (even fatal), and seems indifferent to inefficiency (e.g. retention of long-disused components and steady production of certain molecules most of which are immediately degraded). The most important human design goals include simplicity and near-optimality, while the goals of nature emphasize robustness and the capacity for self-organization. Still, synthetic and natural control systems are again comparable in that humans believe that they can both be profitably described using two concepts enabled by syntactical structure: *module* and *hierarchy*. Modules are logically interchangeable units, for example, all genes (including isoforms) that produce proteins that are biochemically interchangeable. A hierarchy (of relational syntax) is the organizational pyramid of intra- and intercellular communication that makes multicellular life possible. If gene expression organization is modular and hierarchical, then signaling among genes should occur mainly between genes in the same organizational unit; module-to-module signals likely are relatively infrequent and possibly more important (silencing them is especially problematic to life).

The architecture of gene expression is today studied by use of powerful online search engines, of which more than 100 exist. Some of the engines used in the preparation of this paper are: general resources like NCBI (<http://www.ncbi.nlm.nih.gov/>), GeneCards (<http://www.genecards.org/>) (Rebhan, 1997), and Ensembl (<http://www.ensembl.org/index.html>) (EMBL-EBI/Sanger Institute); relationship engines based upon text searches such as Information Hyperlinked over Proteins (iHOP) (<http://www.ihop-net.org/UniPub/iHOP/>) (Hoffmann R, 2004) (Hoffmann R, 2005); resource portals like UBiC (<http://bioinformatics.ubc.ca/resources/>) (Fox JA, 2005) and IMB Jena (<http://www.imb-jena.de/RNA.html>) (Sühnel, 1997); and specialized tools for such tasks as optimizing alignments including ClustalW (<http://www.ebi.ac.uk/clustalw/>) (EMBL-EBI), RNA folding including Sfold (<http://sfold.wadsworth.org/index.pl>) (Ding et al., 2004), and studying miRNA/mRNA relationships including miRNA Viewer at the Memorial Sloan-Kettering Cancer Center site (<http://www.wetdry.org/cgi-bin/mirnaviewer/mirnaviewer.pl>) (John et al., 2004).

One basic problem of using genomic search engines is limiting the proliferation of searches. Starting with a gene, finding genes related by the literature to it, and then finding genes related to those genes, and so on, clearly creates the problem of artificial search termination. Thus a goal of this paper is to present a different type of search that hybridizes two types of gene-to-gene relationships and is inherently limited. The first search type is a literature or table search and the second is a biochemical alignment search. For the text search part, we next discuss the concept of counteracting genes.

Counteracting genes

Given a selected gene, we may define the set of its *contemporary* genes to be all other genes that could be expressed at roughly the same time and in the same locus of the same cell. Next we define *counteracting genes* to be a subset of a set of contemporary genes with the property that their action opposes in some sense the action of the first gene. The primary example in this

paper is a pro-apoptosis gene versus the set of contemporary anti-apoptosis genes. We will also need the concept of *coacting* genes, contemporary genes that in some way have the same effect as the first gene.

The counteraction between two genes might be understood in detail by discovery of opposite effects on the rate of expression of a third gene. For example, one gene might be expressed to generate a protein that in certain reactions has the effect of activating a transcription factor, while a counteracting gene inhibits activation of the same transcription factor. Thus the level of detail of the definition of the concept of counteracting gene is variable. Also, defining counteracting genes might proceed by first describing any action of the first gene, then adding all other genes with any type of related actions. The next step would be to choose a subset of all the added genes that can be reasonably described as being counteractive to one function of the first gene. For example, using the apoptosis web page of N. Inohara (<http://www-personal.umich.edu/~ino/List/alphabet.html>), we selected the pro-apoptosis gene PAWR (GeneID 5074; alias PAR4, PAR-4), leading to a certain row in an interaction matrix. In the row that includes PAWR, we find under an "inhibit" hyperlink reference to research showing that PAWR suppresses expression of BCL2 (596) and other genes. Automating such searches is an active research area, and early versions of some products such as GeneWays (<http://geneways.cu-genome.org/>) (Rzhetsky, 2004) are becoming available.

The central theme of this paper that enables a hybrid, inherently terminating search is that in some cases: transcription of one gene might automatically create a byproduct from excised introns that causes translational inhibition of a counteracting gene.

A scenario that fits the mechanism of this paper might be described as follows (see Fig. 1). Suppose a system switches on and off production of mRNA X used to make protein X; normally production is on. Production is switched off by an action gene as follows. The default state is maintained by steady expression of a counteraction gene that yields a protein that activates a transcription factor (TF) needed for production of mRNA X. The action gene produces a protein that represses activation of the same TF. Furthermore, suppose the action gene also yields a transcription byproduct that inhibits translation of the counteracting gene. Thus expression of the action gene not only represses activation of the TF, but it also represses translation of the counteracting gene that is trying to activate the same TF. The action gene might be not only selfish (Dawkins, 1990), but also devious (emphasizing its importance whenever expressed by automatically disabling counteracting genes).

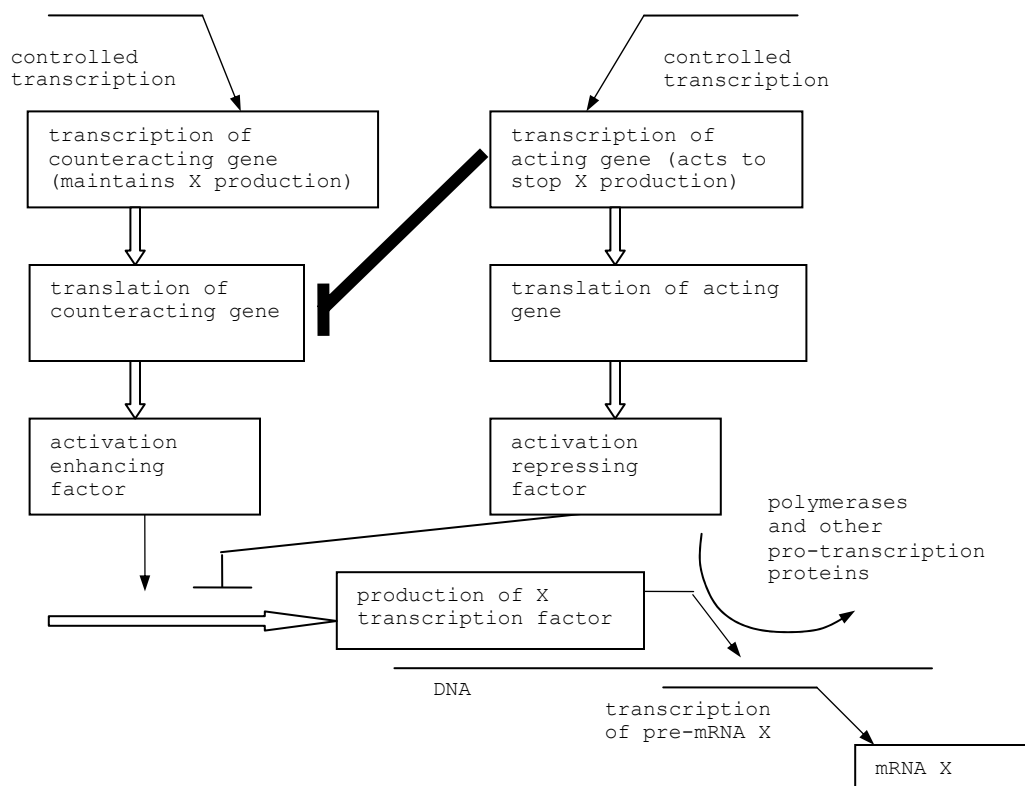


Figure 1. Simplified schematic of one arrangement in which an automatic inhibition process simplifies control of switching off mRNA X production. In this case, the novel mechanism proposed in the paper is shown as heavy line with T inhibition symbol. Lines with arrowheads denote promotion and open flowchart arrows denote processes. Precision of the controlled transcription of the counteracting gene (left) becomes less essential with the automatic inhibition process. The overall effect is simple, efficient, and decisive switching on production of mRNA X and protein X. Such an automatic blocking relationship might be a syntactically valid pattern and might be useful in describing modular and hierarchical organization of gene expression control.

Variations on this central theme will be suggested in the next section.

Algorithm for construction of gene control modules

The flowchart of an algorithm we call Hybrid Search is shown in Fig. 2. It includes an ordering of bioinformatic research steps that could lead to discover of modules of genes in which automatic inhibition of counteracting genes occurs.

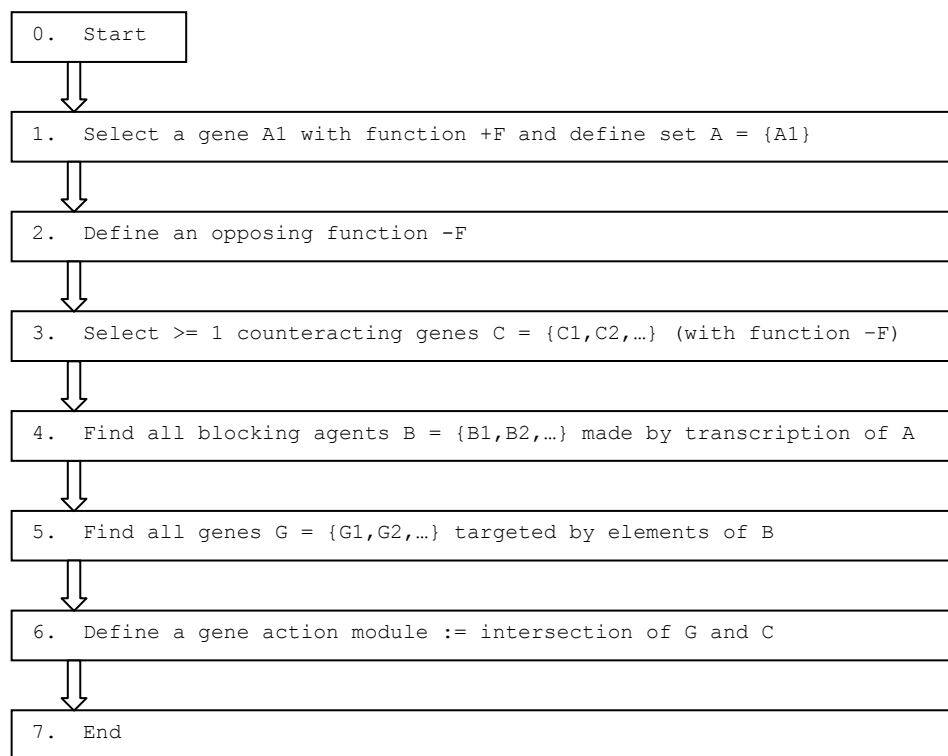


Figure 2. The Hybrid Search Algorithm. From an input consisting of starting gene A_1 and a function $+F$ of A_1 , a counteracting function $-F$ is defined. Next a text search produces a set of counteracting genes with function $-F$. Then a sequence search based on all the possible blocking agents among transcription byproducts of the starting gene A_1 produces the union of genes that could be blocked by any of the blocking agents. The algorithm output is the intersection of the set of counteracting genes with the set of potentially blocked genes. Such an intersection, if nonempty, is an organizational module in the syntax of gene expression.

An extension of the Hybrid Search Algorithm is to use as input a set A with two or more genes $A = \{A_1, A_2, \dots\}$ consistently having action function $+F$. The output set of genes is the union of the outputs over all elements of A .

The Hybrid Search Algorithm can be extended as follows (suggested by a reviewer). A first gene might encode a transcription factor that upregulates expression of a second gene. At the same time, the same transcription factor might target and upregulate a cluster of noncoding genes (miRNAs) that in turn inhibit translation of the second gene. Thus expression of the first gene both directly upregulates and indirectly downregulates expression of the second gene. A human instance might be the proto-oncogene MYC (4609, alias c-MYC), that expresses a transcription factor that upregulates another transcription factor E2F1 (1869) associated with proliferation, differentiation, and inhibition (Levens, 2002). At the same time, the product of MYC also binds to a cluster of six miRNAs, at least two of which downregulate expression of E2F1 (O'Donnell et al., 2005); listed target sites in the 3' UTR of E2F1 include the same hsa-mir-17-5p and hsa-mir-20a, according to the miRNA TargetViewer web engine cited above. Thus MYC would seem to directly upregulate E2F1 and indirectly downregulate E2F1. This organizational module could be

found by a modified algorithm.

The Hybrid Search Algorithm might be extended in another logical direction. Suppose a gene has no direct regulatory effect on a chemical signal, but upregulation of the gene is coacting with upregulation of the signal. Suppose also a transcription byproduct of the first gene downregulates translation of a second gene that inhibits levels of the signal. A putative case would be pro-apoptotic PAWR regulation of manganese superoxide dismutase SOD2 (6648, alias MnSOD), since SOD2 is a gene that limits cytosolic levels of reactive oxygen species (ROS) that can signal initiation of apoptosis (Forman HJ, 2002). The possibility of a PAWR/SOD2 system is further discussed below.

The miRNA pathway

ncRNA is a product of transcription that is noncoding, that is, does not subsequently appear in mRNA that is used as a template for translation. Many recent publications have greatly expanded the proposed and sometimes observed roles of ncRNA in regulation of gene expression. In particular, much recent research scrutiny has been focused on dsRNAs processed from ncRNAs that function as translation-inhibiting or, in plants, mRNA-cleaving (Lai, 2003), agents supported in RNA Induced Silencing Complexes (RISCs). The miRNA pathway includes small dsRNA hairpins (~50 nt on each side) and leads to translational inhibition or accelerated degradation of targeted mRNA by *mature* miRNAs (~20 nt). *In vitro* the miRNA-associated enzyme Dicer will in fact produce many species of likewise small dsRNA molecules from much larger dsRNA. The small product dsRNA molecules are described by Ambion TechNotes as a "cocktail" (<http://www.ambion.com/techlib/tn/114/5.html>). Different small dsRNAs in the cocktail typically have a range of potencies as antisense inhibitors of mRNAs. *In vivo* dsRNA from other sources including folded introns with repeats appearing in stems might likewise serve as feedstock for the same apparatus (Meister and Tuschl, 2004). We note that processing of repeats into repeat-associated short interfering RNAs (siRNAs) was reported in worms by (Ambros et al., 2003) and, in zebrafish, by (Chen et al., 2005); instances in plants have been documented in (Hannon, 2002).

In multicellular eukaryotes, miRNAs are now widely recognized to perform regulation of gene expression essential to life (Giraldez et al., 2005) (Leaman, 2005). The site <http://microrna.sanger.ac.uk/sequences/index.shtml> (Griffiths-Jones, 2004) (Griffiths-Jones, 2004) presently lists ~321 human miRNAs. Combinations of the miRNAs might target 5300 human genes at 13,000 target sites on mRNAs (Lewis et al., 2005). Large maps between genes and miRNAs are available at in the miRNA Viewer cited above.

Again, miRNA control of gene expression (reviewed in (Eccles, 2004) and companion papers in the same issue of *Nature*) involves a transcribed pre-miRNA sequence of ~70 nts to ~120 nts, the ends of which contain a sequence and an approximation of the revcom of the sequence. The ssRNA miRNA strand is excised in processing of the ambient ncRNA. Of the human miRNA transcriptional units, some are in introns of ambient mRNAs, so they presumably become available for gene control along with transcription of the host gene; experimental evidence of this is reported in (Baskerville S, 2005). Using UCSC Golden Path and GeneCards (<http://bioinfo1.weizmann.ac.il/genecards/index.shtml>) we observe ~40 genes with introns that contain miRNAs. Such miRNAs tend to appear in clusters of up to five, totaling 57 miRNAs in introns. The pre-miRNA folds into a stable hairpin molecule with a dsRNA stem that can be exported from the nucleus. In the cytosol, the dsRNA stem can be recognized and processed, and from one side, the short mature subsequence (~20 nt) can be selected for binding to a complementary portion of a target mRNA. The proposed miRNA mechanisms for repressing protein production include impeding processing in the ribosome, enhancing mRNA degradation, and epigenetic mechanisms (Hammond et al., 2000) (Reinhart et al., 2000) (Meister and Tuschl, 2004) (Bartel, 2004) (Scott, 2004) (Pillai RS, 2005). Though details of miRNA translational inhibition remain elusive, the technology of gene silencing has quickly become mature; a summary appears in the text (Sohail, 2004).

Generalizing miRNAs, we note that if an ncRNA contains a sequence followed closely by its revcom, then that part of the ssRNA could also fold into a hairpin or other RNA secondary structure, parts of which might be exported from the nucleus and then acted upon by the enzyme Dicer to produce numerous short dsRNA molecules. Possibly the progress of Dicer in cleaving a long dsRNA is not dependent upon the specific nts at the cut sites; rather, cuts might be made repeatedly depending upon the distance from the end of the molecule. The dsRNA side that enters the RISC is the one with the less tightly paired 5' end (Khvorova et al., 2003), (Schwarz et al., 2003).

Alu-type hairpins as blocking agents and counteracting genes

Alus are ~285 nt in length and are characteristic of primates; they are identified relative to a consensus pattern or 31 closely related patterns (Rebase browser, Genetic Information Research Institute at <http://www.girinst.org/index.html>, (Jurka, 1998) (Jurka, 2000). Alus as a family are the most common type of repeat, accounting for perhaps 10% of the genome. A recent article (Athanasiadis et al., 2004) on adenosine-to-inosine (A-to-I) editing of mRNAs considers intramolecular pairs of sequences, namely, an Alu-type sequence and its revcom, all within a span of several hundred nts. A prerequisite for A-to-I editing is that such RNA fold into a secondary structure containing suitable dsRNA stems (Bass, 2002). As stated above regarding pre-miRNAs, the general questions of the quality of revcom alignment both for hairpin formation and for eventual target binding are not completely resolved. However, Athanasiadis et al. suggest that for stability, the dsRNA should have at most ten mismatches per 100 bp and that the size of the loop be limited to a few tens or hundreds of nts. Furthermore, "Apart from Alu repeats, many more low- and high-frequency repeats exist in the human genome and might give rise to RNA foldback structures that result in exonic A-to-I editing." Or, we assert, parts of the stems of the same foldback structures might be exported to the cytosol and processed by Dicer, allowing the targeting of mRNAs by RISCs.

Recently we described a bioinformatics search providing evidence that pro-apoptosis gene PAWR might automatically inhibit anti-apoptosis gene BIRC4 (Jeffries, 2005). In three positions in intron 2-3 of PAWR, an Alu-like sequence is almost consecutive to a sequence much like Alurc. Thus excision of the 2-3 intron from the pre-mRNA of PAWR might produce three large, stable hairpins. The hypothetical hairpins are shown in Fig. 3, drawn using the fold prediction tool *Srna* (in *Sfold* cited above).

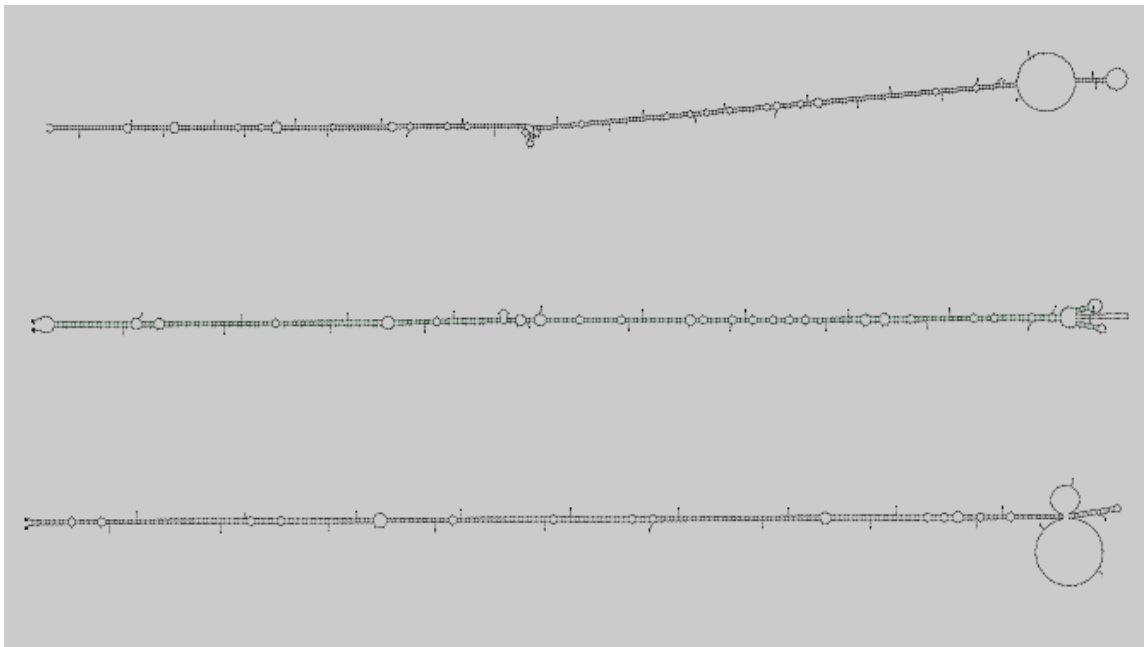


Figure 3. Hairpins predicted by the tool Srna from intronic RNA of intron 2-3 of pro-apoptosis gene PAWR. The first two have Alu/Alurc-like stems; the third has a stem like a portion of an Alu/Alurc dsRNA. The sequences that fold into these hairpins are available in a supplemental file.

Due to the presence of the Alu-like and Alurc-like sequences, parts of the stem, if exported, might hybridize in a RISC complex to various mRNA sequences with Alu or Alurc in their 3'UTRs. In particular, it turns out that Alurc and then Alu subsequences appear in the last exon (including 3' UTR) of the mRNA for BIRC4. (Since this Alu/Alurc pair is separated by 1868 nts, a hairpin would presumably not be formed.) Thus transcription of PAWR would primarily promote apoptosis and secondarily inhibit translation of BIRC4. Hypothetically, this would seem to be part of an efficient, irreversible system for driving initiated apoptosis to its conclusion. If proven, then this mechanism would join (nonexclusively!) the other mechanisms that characterize the potent apoptotic activity of PAWR (El-Guendy, 2003).

The PAWR/BIRC4 example can readily be couched in terms of the Hybrid Search Algorithm in Fig. 2. In the prostate cancer cell line LNCaP, both PAWR and BIRC4 are expressed (El-Guendy, 2003) (Berezovskaya, 2005). Since BIRC4 has an Alurc-like and Alu-like sequences in its 3'UTR, it appears to be a gene in the intersection of counteracting genes and blocked genes (step 6).

Once PAWR and BIRC4 are selected, the definition of action and counteraction can be refined somewhat by the observation that overexpression of full-length PAWR can result in repression of activation of nuclear factor of kappa light polypeptide gene enhancer in B-cells (NF- κ B) following trophic factor withdrawal. (The current description of NF- κ B would include the transcription factor derived from genes NFKB1 4790, NFKB2 4791, NFKBIA 4792, NFKBIB 4793.) Decreased NF- κ B activity is correlated with enhanced apoptosis (Camandola, 2000). In contrast, an anti-apoptotic pathway leads to activation of NF- κ B that regulates the expression of genes including members of the IAP gene family. In particular, BIRC4 is an NF- κ B-dependent member of the IAP gene family and is a strong stimulator of NF- κ B activation (Hofer-Warbinek, 2000). Thus the action of PAWR is to inhibit activation of NF- κ B while BIRC4 has a positive-positive feedback loop with the same transcription factor (see also (Levkau, 2001)). This scenario seems to fit the overall plan of Fig. 1 in which protein X is the transcription factor product of gene NF- κ B.

energy considerations, homology and heuristic goals should figure into an RNA hairpin computer modeling project, analogous to protein modeling approaches (Chou, 2004).

Housekeeping genes

Suppose hybridization between a RISC built with a ~22 nt sequence from an Alu hairpin suffices somehow to silence expression of any mRNA with Alu or Alurc anywhere in its pre-mRNA. In that case, it would seem that transcription of, for example, PAWR would silence many or even most other contemporary genes in some cells. It could be that silencing most genes is desirable in some circumstances such as execution of apoptosis or passage of a cell cycle checkpoint. Our guess, however, is that such wide-scale intronic hybridization and gene repression is unlikely because: mRNA splicing might happen before RISC inhibition or accelerated degradation; RISC binding might be mostly part of control that is graduated control, not absolute, so that many such RISC interactions are required to completely silence a gene; and, nature does not see much of a design penalty in wasting large numbers of newly-assembled pre-mRNAs.

However, one would expect that some genes would be unlikely to be silenced by any such master switch, namely, the housekeeping genes. By examining small (< ~5kb) housekeeping genes in particular, we observe a marked absence of Alu and Alurc, even though nearby intergenic regions are, as usual, littered with the same. On this point Fig. 5 shows relevant tracks for the five, small housekeeping genes: ACTB 60, GAPDH 2597, HTF9C 27037, TUBA1 7277, and UBB 7314. The track graphs were made by the UCSC Genome Browser (<http://genome.ucsc.edu/index.html?org=Human>) (also accessible through GeneCards as UCSC Golden Path) (Karolchik D, 2003).

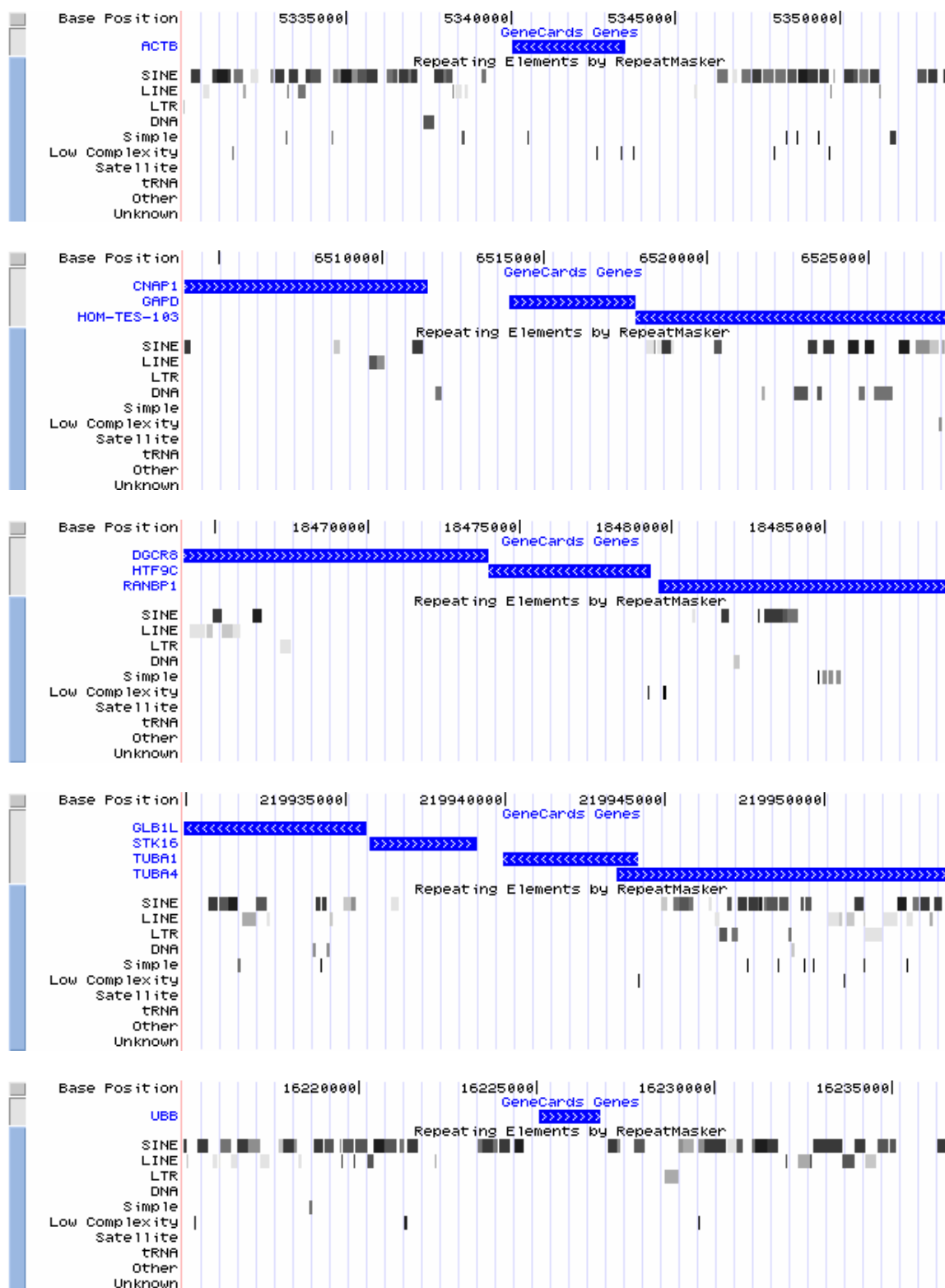


Figure 5. Five small housekeeping genes (ACTB 60, GAPDH 2597, HTF9C 27037, TUBA1 7277, UBB 7314) occur in regions devoid of repeats. Alu repeats (SINEs) and Alurc repeats (also SINEs, not shown) occur in both 3' and 5' directions from the genes, however.

Observations and testable predictions

Among genes involved in various apoptosis pathways, we have found that BAD (572), BIRC4 (331), BIRC6 (57448), DIABLO (56616), and PAWR (5074) are potential sources of Alu-like dsRNA and APAF1 (317), BIRC4 (331), CASP2 (835), SOD2 (6648), and TP53 (7157) are possible targets having Alu or Alurc sequences in their last exons. (Preliminary examination found no such hairpins or targets in several related apoptosis genes: APC, the other six BIRCs, BCL2, CASP3, CASP9, FADD, and FAS.) If any such source and target are contemporaries, then the source might be predicted to inhibit translation of the target by means of the hairpin and Alu or Alurc target mechanism. Of course the same source and target could react in other ways directly or through additional cellular entities. Therefore, isolation of the effect would be difficult, and multiple types of experiments that should have outcomes influenced by the interaction must be devised.

Prediction 1. If the hairpin inhibition hypothesis is correct, then we would expect that silencing of PAWR intron 2-3 would result in decreased downregulation of BIRC4 during the onset of apoptosis.

Prediction 2. As discussed above, Stuart et al. (Stuart, 2000) studied SOD2 in a human cancer cell line. Evidently, naturally occurring ~300 nucleotide (nt) Alu-like subsequences bind by chemical affinity to an approximate Alurc sequence in the 3' UTR of the mRNA of SOD2. PAWR might be a source of such interference. Thus PAWR upregulation, by downregulating SOD2, would lead to higher reactive oxygen species (ROS) concentrations. In turn, upregulation of ROS should act in pathways to accelerate apoptosis (Forman HJ, 2002). If the hairpin inhibition hypothesis is correct, then silencing of PAWR intron 2-3 might be expected to decrease downregulation of SOD2 and indirectly to decrease upregulation of ROS.

Prediction 3. It is known that the pro-apoptosis gene DIABLO (56616, alias SMAC) prevents some BIRCs including BIRC4 from inhibiting some caspases. According to (Fesik, 2001), it is likely that the N-terminus of the protein product of DIABLO simply displaces the N-terminus of a subunit of protein caspase-9, releasing the activated enzyme. A second possible (nonexclusive) mechanism by which DIABLO might inhibit BIRC4 is that the Alu-like hairpin in intron 4-5 of DIABLO targets the Alu and Alurc patterns in the last exon of BIRC4. Another point regarding BIRC4 is that it both seems capable of producing Alu/Alurc hairpins and should be targeted by them. This implies a prediction that BIRC4 levels can only rise when other mRNAs titrate the Alu/Alurc hairpin RISCs.

Prediction 4. Alu-like hairpins conceivably could be employed by anti-apoptosis genes against pro-apoptosis genes. We have observed in BIRC6 (a large gene with 74 introns) that introns 45-46 and 66-67 include Alu-like hairpins with loops of sizes 93 nt and 33 nt, shown in Fig. 6. Thus we may predict, for example, that BIRC6 targets TP53, if they are contemporaries. Again, experiments to silence selected hairpin structures in BIRC6 pre-mRNA might be devised to test whether or not the hairpins have a role in TP53 control. For example, suitable insertion of a gene trap vector in an upstream intron should silence production of the hairpins, causing upregulation of TP53.

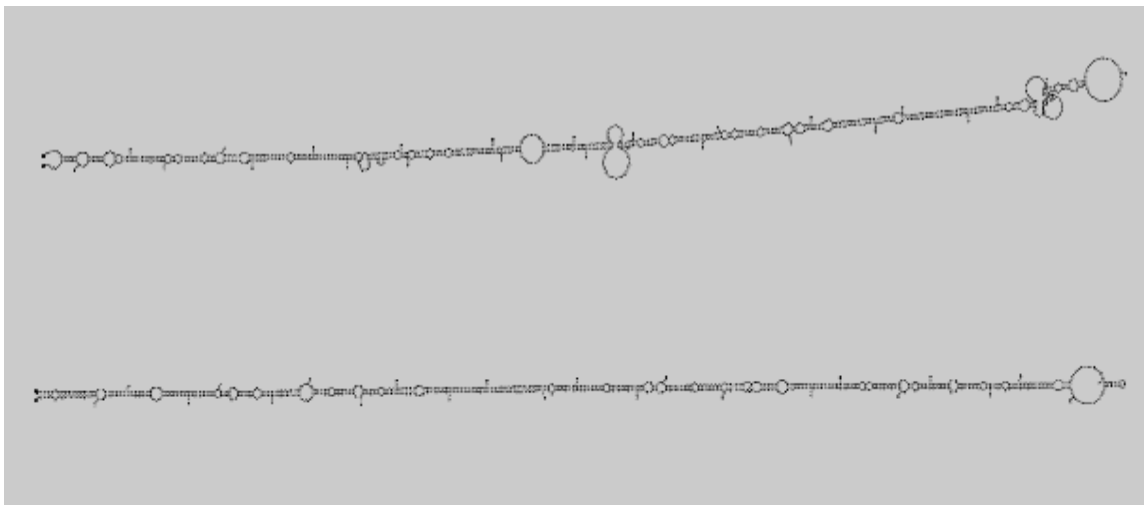


Figure 6. Two hairpins from introns of BIRC6.

Rational drug design from new drug platforms

Many other gene expression systems might involve dsRNA hairpins and translational inhibition. In particular, the transmembrane aspartyl protease β -secretase is the product of gene BACE1 (23621); this zymogen is described as the rate-limiting step in production of brain plaques *in vivo* (Chou KC, 2002). Hence BACE1 expression is a potential target for drugs to treat Alzheimer's disease. We have found that intron 3-4 of the pre-mRNA of BACE1 contains a well-formed Alu-like hairpin, comprising over half of this small intron. The hairpin is shown in Fig. 7. One might hypothesize that the balance of β -secretase is partly controlled by some mechanism in which the hairpin has a role, such as inhibition of translation of a counteracting gene. Disruption of the balance might be etiologic, and restoration of it might be a pharmaceutical goal worth investigation. Truly, there is no shortage of such possibilities.

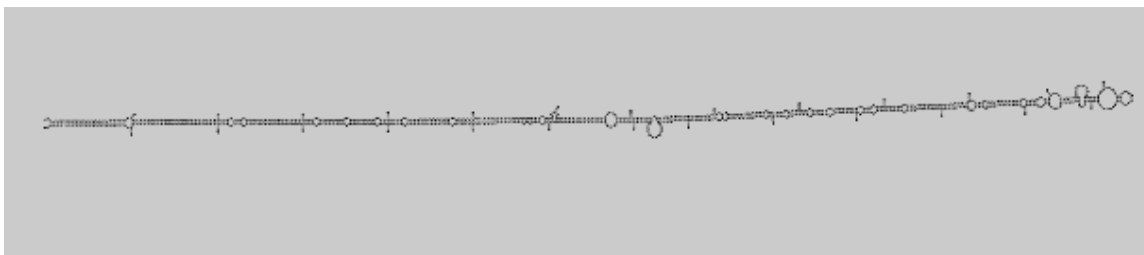


Figure 7. A hairpin from an intron of BACE1.

As another example, let us consider characteristics of prostate cancer. In the normal adult prostate and the primary prostate cancer, the majority of cells are androgen-dependent (Gurumurthy S, 2004). Consequently, upon withdrawal of androgen, the rate of apoptosis exceeds the rate of cell proliferation, resulting in involution of the normal prostate as well as regression of the tumor. Thus androgen-ablation treatment of prostate cancer targets some cells and is often successful, at least initially. However, androgen-ablation might subsequently cause development of a more aggressive tumor formed from surviving androgen-independent cells. For several reasons given in (Gurumurthy S, 2004), drug discovery enabling control of PAWR levels in the prostate might result in the general destruction of tumor cells, regardless of

androgen dependency. Part of the normal potency of PAWR, according to this paper, might be due to the production of Alu-like hairpins. Therefore, we can speculate that investigation of that property of PAWR or supplementary sources of hairpins might be worthwhile.

Conclusions

Some sort of system for combinatorial control with organized syntax and semantics of transcription, editing, translation, and signaling must be employed by nature, but attaining an understanding of it remains a major challenge to molecular geneticists. Thus it is reasonable to seek in nature instances of something humans can understand, namely, modular and hierarchical organization.

In addition to miRNAs, intronic sections of certain transcriptional units might contain sequences that are like Alus (or subsequences of Alus) with nearby revcoms that fold into dsRNA hairpins. Reactions involving dsRNA from the hairpins might lead eventually through a miRNA-like pathway to repression of translation of genes with opposing functions. While cogent and potentially an important theme in gene regulation, such a mechanism obviously requires case-by-case experimental verification. If preliminary experiments were to indicate the mechanism between counteracting genes occurs at least *in vitro*, then refined algorithms for searching for such patterns might be devised.

Acknowledgements

A reviewer suggested the extension of the Hybrid Search Algorithm to the above MYC example and the possibility of other extensions. A second reviewer pointed to BACE1 and suggested investigation into several other potential types of examples. The authors are sincerely grateful for these ideas and also for several suggested and adopted organizational improvements.

References

- Aho, A., Sethi, R., Ullman, J. (1986). *Compilers: Principles, Techniques, and Tools*. 796.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* 13, 807-818.
- Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* 2, e391.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.
- Baskerville S, B. D. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11, 241-247.
- Bass, B. (2002). RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 71, 817-846.
- Berezovskaya, O., Schimmer, A., Glinskii, A., Pinilla, C., Hoffman, R., Reed, J., Glinsky, G. (2005). Increased expression of apoptosis inhibitor protein XIAP contributes to anoikis resistance of circulating human prostate cancer metastasis precursor cells. *Cancer Res* 65, 2378-2386.
- Camandola, S., Mattson M. (2000). Pro-apoptotic action of PAR-4 involves inhibition of NF-kappaB activity and suppression of BCL-2 expression. *Journal of Neuroscience Research* 61, 134-139.
- Chen, P. Y., Manninga, H., Slanchev, K., Chien, M., Russo, J. J., Ju, J., Sheridan, R., John, B., Marks, D. S., Gaidatzis, D., *et al.* (2005). The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev* 19, 1288-1293.
- Chou, K. (2004). Structural bioinformatics and its impact to biomedical science. *Curr Med Chem* 11, 2105-2134.
- Chou KC, H. W. (2002). Prediction of the tertiary structure of the beta-secretase zymogen. *Biochem Biophys Res Commun* 292, 702-708.
- Dawkins, R. (1990). *The Selfish Gene*. Oxford University Press.
- Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32, W135-141.
- Eccles, M. R. (2004). How to interfere with RNA. *Nature* 431, 229.
- El-Guendy, N., Zhao Y., Gurumurthy, S., Burikhanov, R., Rangnekar, V. (2003). Identification of a unique core domain of Par-4 sufficient for selective apoptosis induction in cancer cells. *Molecular and Cellular Biology* 23, 5516-5525.

- Fesik, S. W., Shi, Yigong (2001). Controlling the Caspases. *Science* 294, 1477-1478.
- Forman HJ, T. M., Fukuto J (2002). Redox signaling. *Molecular and Cellular Biochemistry* 234, 49-62.
- Fox JA, B. S., McMillan S, Campbell G, Ouellette BF (2005). The Bioinformatics Links Directory: a compilation of molecular biology web servers. *Nucleic Acids Res* 33, W3-W24.
- Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P., and Schier, A. F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308, 833-838.
- Grewal, S. I., and Rice, J. C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. *Curr Opin Cell Biol* 16, 230-238.
- Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Res* 32, D109-111.
- Gurumurthy S, R. V. (2004). Par-4 inducible apoptosis in prostate cancer cells. *J Cell Biochem* 91, 504-512.
- Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293-296.
- Hannon, G. J. (2002). RNA interference. *Nature* 418, 244-251.
- Hofer-Warbinek, R., Schmid, J., Stehlik, C., Binder, B., Lipp, J., de Martin, R. (2000). Activation of NF-kappa B by XIAP, the X chromosome-linked inhibitor of apoptosis, in endothelial cells involves TAK1. *Journal of Biological Chemistry* 275, 22064-22068.
- Hoffmann R, V. A. (2004). A gene network for navigating the literature. *Nat Genet* 36, 664.
- Hoffmann R, V. A. (2005). Implementing the iHOP concept for navigation of biomedical literature. *Bioinformatics* 21, 252-258.
- Jeffries, C., Jarstfer, M., Perkins, D. (2005). Folded RNA from an intron of one gene might inhibit expression of a counteracting gene. *in silico Biology* 5, 0037.
- Jin, P., Alisch, R. S., and Warren, S. T. (2004). RNA and microRNAs in fragile X mental retardation. *Nat Cell Biol* 6, 1048-1053.
- John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biol* 2, e363.
- Jurka, J. (1998). Repeats in genomic DNA: mining and meaning. *Curr Opin Struct Biol* 8, 333-337.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* 16, 418-420.

Karolchik D, B. R., Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ (2003). The UCSC Genome Browser Database. *Nucleic Acids Res* 31, 51-54.

Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209-216.

Lai, E. C. (2003). microRNAs: runts of the genome assert themselves. *Curr Biol* 13, R925-936.

Leaman, D., Chen, P., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., Marks, D., Sander, C., Tuschl, T., and Gaul, U. (2005). Antisense-Mediated Depletion Reveals Essential and Specific Functions of MicroRNAs in *Drosophila* Development. *Cell* 121, 1097-1108.

Levens, D. (2002). Disentangling the MYC web. *PNAS* 99, 5757-5759.

Levkau, B., Garton, K., Ferri, N., Kloke, K., Nofer, J., Baba, H., Raines, E., Breithardt, G. (2001). XIAP induces cell-cycle arrest and activates nuclear factor-kappaB : new survival pathways disabled by caspase-mediated cleavage during apoptosis of human endothelial cells. *Circulation Research* 88, 282-290.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15-20.

Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787-798.

Mattick, J. (2005). The functional genomics of noncoding RNAs. *Science* 309, 1527-1528.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 343-349.

Murchison, E. P., and Hannon, G. J. (2004). miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Curr Opin Cell Biol* 16, 223-229.

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., and Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839-843.

Pillai RS, B. S., Artus C, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W (2005). Inhibition of translational initiation by let-7 microRNA in human cells. *Science* 309, 1573-1576.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D. (1997). GeneCards: encyclopedia for genes, proteins and diseases. Weizmann Institute of Science, Bioinformatics Unit and Genome Center (Rehovot, Israel).

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901-906.

Rzhetsky, A., Iossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., Yu, H., Duboue, P., Weng, W., Wilbur, W., Hatzivassiloglou, V., Friedman C. (2004). GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *Journal of Biomedical Information* 37, 43-53.

Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199-208.

Scott, R., Spielman, M. (2004). Epigenetics: imprinting in plants and mammals-the same but different? *Curr Biol* 14, 201-203.

Sinclair NR, C. J. (1993). Tentativeness and fervor in cell biology require negative and positive feedforward control. *Life Sci* 52, 1985-1993.

Sohail, M. (2004). *Gene Silencing by RNA Interference: Technology and Application*. University of Oxford Press, 416.

St Clair DK, P. S., Xu Y, Kiningham K (2002). Transcription regulation of human manganese superoxide dismutase gene. *Methods in Enzymology* 349, 306-312.

Stuart, J., Egry, L., Wong, G., Kaspar, R. (2000). The 3' UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. *Biochem Biophys Res Commun* 274, 641-648.

Sühnel, J. (1997). Views of RNA on the World Wide Web. *Trends in Genetics* 13, 206-207.

Xu, P., Guo, M., and Hay, B. A. (2004). MicroRNAs and the regulation of cell death. *Trends Genet* 20, 617-624.

Xu Y, K. K., Devalaraja MN, Yeh CC, Majima H, Kasarskis EJ, St Clair DK (1999). An intronic NF-kappaB element is essential for induction of the human manganese superoxide dismutase gene by tumor necrosis factor-alpha and interleukin-1beta. *DNA and Cell Biology* 18, 709-722.

Yildirim I, T. D. (2005). RNA Challenges for Computational Chemists. *Biochemistry* 44, 13225-13234.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.