nature genetics

The four Rs of RNA-directed evolution

Alan Herbert

The way we quantify the human genome has changed markedly. The estimated percentage of the genome derived from retrotransposition has increased (now 45%; refs. 1,2), as have the estimates for alternative splicing (now 41–60% of multiexon genes)[3,4], antisense transcription (now 10–20% of genes)[5,6] and non–protein coding RNA (now ~7% of full-length cDNAs)[7]. Concomitantly, the estimated number of protein-coding genes (now ~24,500) has decreased[8]. These numbers support an RNA-centric view of evolution in which phenotypic diversity arises through extensive RNA processing and widespread RNA-directed rewriting of DNA enables dissemination of 'selfish' RNAs associated with successful outcomes[9]. The numbers also indicate important roles for sense-antisense transcription units (SATs) and coregulatory RNAs (coRNAs) in directing the read-out of genetic information, in reconciling different regulatory inputs and in transmitting epigenetic information to progeny. Together, the actions of reading, 'riting, 'rithmetic and replication constitute the four Rs of RNA-directed evolution.

The read-out of genetic information in eukaryotes is not only more complex than previously realized, but also more dynamic. Much of the information is 'soft-wired', with extensive RNA processing necessary to generate phenotypes[10]. A subset of the RNA, referred to here as coRNA (**Fig. 1**), coordinates the read-out of genetic information and reconciles regulatory inputs. For each cell type, the sum of these interactions defines a distinct RNA profile or ribotype[10]. RNA-based processes also facilitate replication of ribotypes in subsequent generations. The flexibility of these programs allows soft-wired organisms to evolve in a more rapid and dynamic way than 'hard-wired' organisms constrained by DNA mutation[10]. Together, the actions of reading, 'riting, 'rithmetic and replication constitute the four Rs of RNA-directed evolution. Here, we describe roles for coRNAs and SATs (**Fig. 2**) in these events.

RNA-directed RNA readout

RNA coregulates the sequence-specific read-out of genetic information from RNA by targeting the evolutionarily conserved machinery of RNA interference (RNAi)[11] and translational repression. RNAi leading to post-transcriptional gene silencing (PTGS) is induced by double-stranded RNA (dsRNA) arising from infection by dsRNA viruses, sense-antisense transcription pairs, transcription of inverted repeats and delivery of dsRNAs manufactured *in vitro*[12]. Both endogenous and exogenous dsRNAs are processed cytoplasmically by RNase III dicer paralogs into small interfering RNAs (siRNAs) that are 21–25 bases long[12,13]. These siRNAs are incorporated into an RNA-induced silencing complex (RISC) that targets any RNA with a sequence complementary to the siRNA[12,14]. The target RNA is cleaved by RISC, which has endonucleolytic and presumed exonucleolytic activity[12,15]. Each cycle results in the destruction of the target RNA and the regeneration of an active RISC[16,17].

Alan Herbert is in the Department of Genetics and Genomics, Boston University School of Medicine, 715 Albany Street, Boston, Massachusetts 02118, USA. e-mail: aherbert@bu.edu

In *Neurospora crassa*[18,19], plants[20], *Schizosaccharomyces pombe*[21] and *Caenorhabditis elegans*[22], but not in flies, humans or mouse oocytes[18,19,23,24], PTGS can also be initiated by an RNA-dependent RNA polymerase (RdRP) that synthesizes complementary RNA from highly expressed transcripts to generate dsRNA. This process can spread from the 3′ to the 5′ end of the mRNA, leading to progressive silencing of a gene and other mRNAs with exons in common (transitive RNAi)[18,22]. From the evolutionary viewpoint, transitive RNAi favors the generation of new phenotypes, as related genes in species with polyploid genomes will escape cosuppression by sequence divergence, resulting in altered function or expression.

The number of dicer paralogs also differs between species. *Arabidopsis*, which has at least four dicer family members, seems to use different enzymes to defend against viruses and to process endogenously produced dsRNA[13,25]. In humans, mice and worms, there is only one dicer ortholog, suggesting that any pathway-specific functions are guided by ancillary regulatory proteins associated with RISC. These ancillary proteins include members of the argonaute family, some of which are involved in cell fate determination[12,26]. Argonaute proteins may therefore target RISC to specific pathways during development, suggesting an important role for RNA-dependent processes in specifying phenotypes, and thus in the evolution of these phenotypes.

RNA can also induce PTGS by repressing mRNA translation. This mechanism was first demonstrated in *C. elegans*, in which the short non–protein coding RNAs *lin-4* and *let-7* are used to time development programs by regulating the stage at which *lin-14*, *lin-28* and *lin-41* messages are translated[27]. These RNAs are microRNAs (miRNAs)[27–30], which are encoded genomically as short inverted repeats that have a dsRNA stem of about 70 bp. They are found in introns as well as in intergenic clusters[29,31]. The precursors are processed by dicer and other RNase III family members to produce effectors of 21–25 nucleotides derived usually from one strand of the stem[32]. miRNAs are defined by conservation between species; there are an estimated 200–255 miRNAs in humans[30].
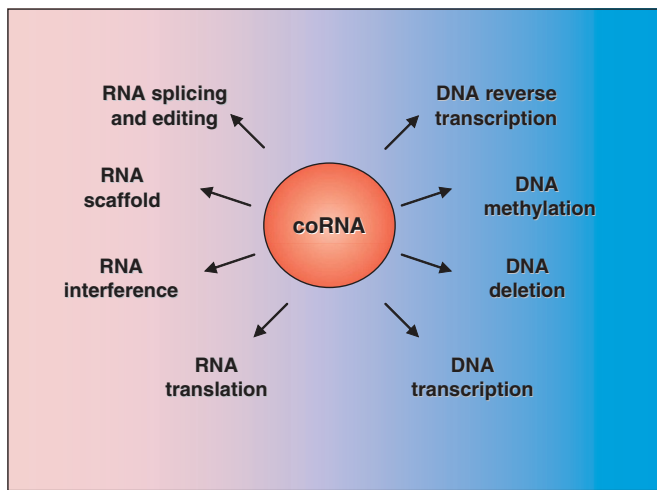
Figure 1 coRNAs (shown in red) have a central role in RNA and DNA read-out. For example, miRNAs can act through the RNA interference pathway or by inhibiting translation of RNA. Similarly, coRNAs affect DNA read-out through transcriptional silencing pathways or by modulating transcription. They are also essential for the rewriting of RNA during editing and splicing, generating new nucleotide sequences not found in the gene. Together, these processes allow information from RNA-processing events to be reconciled with other regulatory inputs, including those from the environment, and to affect both the short-term (through epigenetic mechanisms) and long-term storage (through reverse transcription) of outcomes in DNA.

Some miRNAs are capable of targeting RNAs to the RNAi pathway[16]. Other miRNAs are predicted to bind conserved elements in mRNA important for translational control, such as the K box, Brd box and GY box motifs[33]. A role for *trans*-acting RNAs in these pathways had not been previously recognized. A wide variety of miRNA targets have been identified, mostly in plants, where perfect sequence matches between miRNAs and their targets are more common than in other species[27]. The tissue-specific expression of miRNAs in other species suggests that they have roles in development analogous to those found in *C. elegans*[34,35].

RNA-directed DNA read-out

The read-out of genetic information from DNA can be silenced by processes that generate dsRNA, including SATs and transcription of inverted repeats. The activity of transcriptional complexes can also be modulated by coRNAs either serving as scaffolds or acting as allosteric effectors.

Transcriptional gene silencing (TGS) can be induced in *trans* in plants by dsRNA homologous to a promoter and is associated with DNA methylation in the region of sequence overlap[36]. The involvement of RISC in this pathway is suggested by the requirement for argonaute family members. For example, in *Arabidopsis*, mutant argonaute 4 protein diminishes methylation of the gene *SUPERMAN*, which defines boundaries between carpels and stamens[13,37].

More commonly, RNA-directed DNA silencing uses elements encoded in *cis*. Direct evidence has been obtained in *S. pombe* for the involvement of the RNAi pathway in establishing the silenced state[38,39]. Silencing is associated with bidirectional transcription of a repeat element in heterochromatin that generates a dsRNA substrate and is dependent on the *S. pombe* orthologs of *dicer*, *argonaute* and RdRP[38,39]. The pathway initiates a signaling cascade that produces methylation of Lys9 on histone H3. Spread of silencing to adjacent

chromosomal regions depends on the heterochromatin protein 1 ortholog *swi6* (refs. 21,38), a member of the chromodomain protein family that recognizes both histone H3 Lys9 and RNA[40]. A related mechanism seems to silence polycomb response elements (PREs) in *Drosophila melanogaster*; studies using chromatin crosslinking and immunoprecipitation showed that promoters in these elements are engaged with the transcriptional machinery. These promoters have a low level of transcriptional activity, give rise to short dsRNA products[41,42] and produce methylation of histone H3 at Lys27 (ref. 43). Involvement of RISC in the silencing of PREs is supported by its dependence on *piwi*, an argonaute family member[42].

Similar mechanisms may control transcription from retroelements[44]. For example, the methylation status of the agouti A^{vy} allele, which contains an antisense retroviral IAP insertion[45,46], determines its effect on phenotype[44]. Methylation occurs stochastically, producing phenotypic variation at the level of the individual, the tissue or the cell. For example, methylation of only a subset of cells produces mottled mice with both yellow and agouti coloration. Variable methylation of other retroelements may have a similar impact on many phenotypes[44]. Such alleles, whose phenotypic expression depends on epigenetic modification, have been referred to as epialleles[47].

TGS can potentially be initiated by dsRNA formed from pairs of transcriptional units arranged in a tail-to-tail orientation (SATs). In humans, SATs account for most overlapping transcriptional units (~70%; refs. 5,48). They are far more frequent than previously anticipated[5,48]. A recent computational survey estimated that there are ~1,600 human SATs (~3,200 transcriptional units)[5]. Similar results were obtained from an analysis of the mouse FANTOM2 clone set. A total of 2,481 sense-antisense pairs (4,962 transcriptional units) in which exons from both transcriptional units overlapped were identified. In 351 of these pairs, neither transcriptional unit seems to encode a protein product, whereas in 519 pairs, both are coding[6]. Another 899 pairs (1,798 transcriptional units) in which the overlap was between an exon and an intron were identified. The number of SATs is probably underestimated in these analyses due to our limited knowledge of the degree to which the 5′ and 3′ untranslated regions of adjacent genes overlap. Furthermore, the role of RNA polymerase III transcripts (~10^6 in the human genome[49]) in SAT formation was not evaluated in these reports. Given the current estimate of 24,500 human protein-coding genes[8], 10–20% of these seem to belong to SATs.

SATs offer many potential advantages for gene regulation (**Fig. 2**). When both transcriptional units are active, formation of dsRNA occurs by default, leading to histone modification and TGS: SATs effectively cancel each other, relaxing the need for stringent regulation of promoter engagement by transcription factors. An example is the imprinting of the gene *Igf2r* in mice[50]. Expression of *Igfr2* depends on whether the locus is inherited from either the mother or the father. Expression from the paternal chromosome is prevented by an antisense transcript produced by the non–protein coding gene *Air*. Net expression of *Igfr2* occurs only from the maternal chromosome, as methylation of the maternal *Air* gene promoter prevents production of an antisense transcript[50]. This mechanism might be quite general. A search for imprinted genes identified 159 differentially expressed loci with SATs[51], and an examination of 58 known mouse imprinted genes found an antisense transcript for 22 of them[6]. Besides acting on themselves, SATs may cause differential expression of other loci by silencing genes in adjacent chromosomal regions.

SATs also have the potential to regulate silencing of adjacent chromosomal regions by switching one transcriptional unit off, allowing

net transcription of the other. This process could be stochastic. Such an event may operate in the random silencing of paternal and maternal X chromosomes in mouse (but not human[52]) somatic tissue[53]. In mice, the expression of the X-inactivating *Xist* transcript and its antisense transcript *Tsix* are mutually exclusive[53].

Switching to a promoter outside the region of silencing provides a different mechanism for producing net transcription from only one transcriptional unit in a SAT (**Fig. 2a**). Such a process may regulate expression of bithorax complex genes in flies[54,55]. Here, transcription through a silenced PRE from an intergenic promoter outside the region of silencing relieves suppression of the PRE and results in an active chromatin conformation. Other genes silenced by spreading of heterochromatin from the PRE also become transcriptionally active[54,55]. This outcome represents an epigenetic switch as the active chromosomal state is transmitted to daughter cells[54,55]. This mechanism also allows the coordinated switching of one gene on and another off (**Fig. 2b**). For example, a bidirectional promoter in the cardiac myosin locus that activates read-out of *MYHC6* also generates an antisense transcript that switches off the upstream gene *MYHC7* (ref. 56). Expression of *MYHC6* is thus accompanied by silencing of *MYHC7*. Similarly, polyoma virus uses late transcripts to downregulate early ones[57].

coRNAs act in other ways to regulate read-out of information from DNA. Many pathways involve *trans*-acting coRNAs that affect transcriptional initiation and elongation. Examples include the steroid receptor RNA activator[58], dsRNA motifs bound by the transcriptional regulator NF110 (ref. 59), the U1 small nuclear RNA that associates with TFIIH[60] and the 7SK RNA that inhibits Cdk9 (the RNA polymerase II C-terminal domain kinase)[61,62]. In these cases, coRNA seems to act as an allosteric effector. In other pathways, coRNAs may serve as scaffolds for the assembly of activation complexes. For example, the roX1 and roX2 transcripts in flies nucleate protein complexes that double the rate of X-chromosome transcription in XY males to match the quantity of X-chromosome transcripts present in XX females, achieving dosage compensation[63]. In addition to binding proteins, RNA scaffolds could bind other coRNAs in a sequence-specific manner. Furthermore, both scaffolds and allosteric coRNAs provide a means to switch SATs in *trans*[64].

Candidate coRNAs have been identified by searching for non–protein coding RNAs other than transfer RNA and ribosomal RNA in the RIKEN mouse cDNA collection[7]. These transcripts constitute 7% of the library. Only 7.5% are antisense to another transcript. Some have orthologs in other species[65]. Such findings indicate that coRNAs may be quite numerous in soft-wired organisms.

RNA-directed rewriting of RNA

Rewriting of information in nascent RNA transcripts is a well characterized process: some nucleotides are deleted during splicing and others are changed by editing. Around 41–60% of mouse multiexon genes generate alternatively spliced transcripts[4]; the frequency of edited transcripts is unknown. These processes generate new sequences not found in the gene. Trypanosomes show the importance of RNA rewriting. Their survival depends on editing defective mitochondrial transcripts using *trans*-encoded RNA sequences to guide insertion and deletion of uridines[66]. The rewriting of RNA restores the correct reading frame, allowing the production of functional gene products. RNA guides are also used to direct rewriting of RNA during editing and splicing of pre-mRNA[3,67]. In some cases, editing targets splice sites and in others, splicing prevents editing[67]. Whether *trans*-acting coRNAs
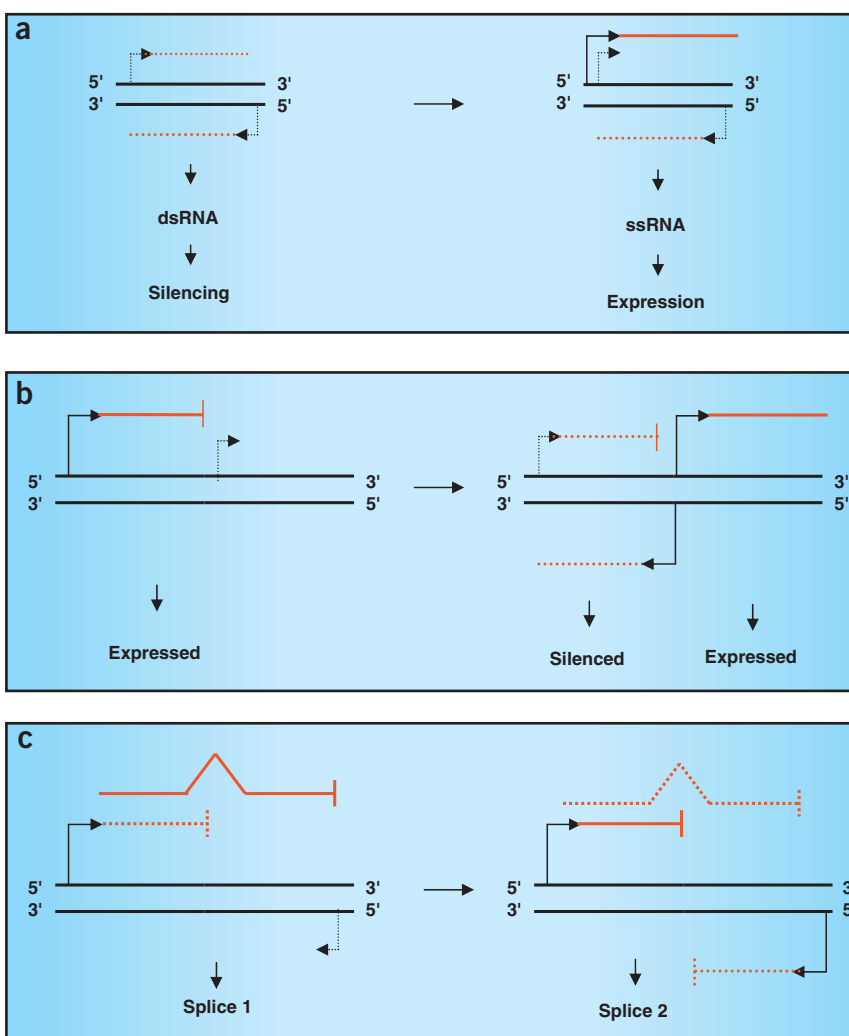
Figure 2 Use of SATs to regulate RNA expression and processing. (**a**) Transcription from both transcriptional units initiates a signaling cascade that results in gene silencing (indicated by dotted lines for transcripts and promoters): the transcriptional units effectively cancel each other out. In this example, net transcription of one gene is achieved by using a promoter outside the affected region (both the transcript and the promoter are shown with solid lines). ssRNA, single-stranded RNA. (**b**) Switching one gene off and another gene on by using a bidirectional promoter. Silencing of the upstream gene results from the antisense transcript generated from the promoter. At the same time, transcription of the downstream gene begins. (**c**) A switch to a shorter RNA splice isoform by using an antisense transcript to interfere with production of the longer isoform. Transcripts are shown in red and the DNA strands are shown as parallel, solid horizontal black lines, with direction of transcription being 5′ to 3′.

Table 1 Genomic mechanisms for the generation of new coRNAs

DNA-mediated[a]	RNA-mediated[b]
Recombination	Template switching
	trans-splicing
	Insertion into a transcript
	Insertion into a transcriptional unit
Insertion of DNA transposon promoter	Insertion of retroviral promoter
	Insertion RNA polymerase promoter
Duplication	Reverse transcription into multiple sites
Deletion	Splicing
Point mutation	Editing
Gene conversion	Repair using a paralogous RNA template

[a]Some events may be RNA-dependent, such as DNA deletion in *Tetrahymena*[76–78] and the segmental duplications associated with insertion of *Alu* retroelements[83]. [b]In these events, RNA intermediates are written to DNA by reverse transcription or are transmitted epigenetically[1,2].

The high frequency of retrotransposition in the human genome[1,2] underscores the importance of RNA-mediated events in the generation of new coRNAs. Events mediated by both DNA and RNA generate new coRNAs by altering the sequence of old ones, by producing new combinations of old sequence elements, by altering processing of old transcripts directly through changes to regulatory elements or indirectly through other RNA-directed or protein-based events, by creating new promoter elements that lead to expression of new transcripts and by assembling new transcriptional complexes that recognize new promoter sites.

generate phenotypic diversity by regulating alternative splicing is currently unknown. But the potential to treat mendelian diseases by altering splicing patterns with transfected antisense molecules has been shown[68].

Rewriting of RNA is also impacted by SATs. For example, an overlapping antisense transcript inhibits formation of the longer splice isoform from the thyroid receptor α gene (*THRA*; **Fig. 2c**)[69]. dsRNAs produced by SATs are also substrates for the ADAR family of editing enzymes[70]. These enzymes rewrite adenosine with inosine, destabilizing dsRNA[67]. This process downregulates RNAi induced by transgenes in *C. elegans*, probably by disrupting the formation of siRNA[70]. Similarly, dsRNA-activated TGS mechanisms may be inhibited by ADARs, resulting in the net expression of both transcriptional units in a SAT. This process may have other sequelae: RNAs with a large number of inosines are stable and selectively retained in the nucleus[57]. These modified transcripts, perhaps produced by a particular class of SATs, are proposed to participate in the organization of the nuclear matrix[71].
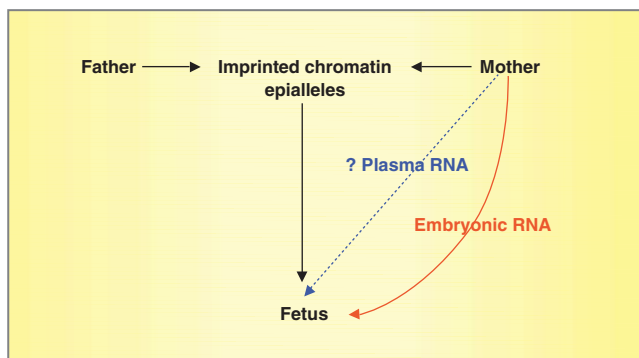

Figure 3 Replication of RNA-based programs in the next generation. In soft-wired organisms, imprinting of parental chromosomes and the stochastic methylation of epialleles probably depend on RNA-based mechanisms. The development of the embryo also relies on maternal RNAs transferred to the ovum. These RNAs initiate the establishment of appropriate ribotypes during embryogenesis. Maternal RNAs present in plasma could also influence these events.

Rewriting of RNA is associated with a high turnover of transcripts. Of all the RNA transcribed in human nuclei, only about 5% enters the cytoplasm[72]. Quality control mechanisms, such as nonsense-mediated decay, dispose of incompletely or improperly processed messages encoding flawed proteins[73]. They ensure that rewriting of information in RNA in soft-wired organisms yields functional ribotypes.

RNA-directed rewriting of DNA

Genomes can be rewritten, using reverse transcription to record elements of successful ribotypes and create an evolutionary scratchpad for new ones. Around 45% of the human genome is derived from retrotransposition[1]. These processes generate genetic variability (**Table 1**)[1,2]. Silencing of genes may result when insertion produces new SATs. Alternative splicing may develop when reverse transcription of an antisense RNA copies into a gene those sense-strand sequences that encode introns[74]. Species-specific promoters may evolve through spread of RNA scaffolds by reverse transcription.

RNA-directed rewriting of DNA also has an essential role in maintaining genome stability. Telomerase is a reverse transcriptase that uses an RNA guide to rewrite chromosomal ends and prevent their loss through fusion[75]. Other roles for RNA-directed rewriting of DNA have been discovered in tetrahymena. This organism has two nuclei: a germline micronucleus that is transmitted to progeny and a somatic macronucleus that is actively transcribed[76–78]. The macronucleus is created from the micronucleus. During this process, the deletion of DNA sequences is guided by homologous RNA transcripts from the micronucleus[76–78]. New sequences appearing in the macronucleus during one reproductive cycle are targeted for deletion in the next generation, ensuring, for example, that the genome is not destabilized by the spread of DNA transposons[78]. Similar types of processes may impact genome stability in other organisms. For example, centromeric repeats are deleted from *S. pombe* when the RNAi machinery is inactivated by mutation, suggesting that unrepressed bidirectional transcription, due to a lack of silencing, promotes loss of these elements[38]. One consequence of repeat deletion is failure to assemble and correctly segregate chromosomes during meiosis and mitosis[79]. Microdeletion in areas of SAT overlap, such as those observed in 70% of human Angelman and Prater-Willi syndromes[80], may have a similar basis, as may chromosomal loss during oncogenesis. Deletion of retroelements from inverted repeats, such as those found in the human Y chromosome[81], may be favored by the enhanced transcription and chromatin remodeling occurring during spermatogenesis[82]. Recombination between actively transcribed retroelements may also produce the small segmental DNA duplications that account for ~5% of the human genome[83]. RNA-directed processes thus have important roles in rewriting and maintaining information stored in DNA. Besides influencing DNA read-out, SATs also shape the genomic architecture of soft-wired organisms.

Reconciliation of information

In each ribotype, there is a rule set to ensure that specific transcripts are produced and particular mRNAs translated (**Fig. 1**). These outcomes are achieved by using coRNAs to coordinate the action of highly conserved pathways[84,85]. An RNA product from one processing event may regulate a downstream event, making the second outcome contingent on the first[10]. For example, a miRNA encoded in an intron would only be expressed when the host gene was transcribed[31,84,85]. Alternatively, coRNAs may facilitate coordination of pathways by interacting with sequence motifs shared by a number of targets[64].

Evolution of rule sets requires creation of new coRNAs (**Table 1**). New coRNAs could arise by duplication and mutation of older versions to

generate new ones with different expression patterns or altered sequence specificity. Alternatively, coRNAs with different permutations of regulatory elements could be created *de novo* by template-switching during reverse transcription. This mechanism is similar to that proposed for retroviral recombination[86]. New coRNAs could also evolve from intron sequences or from genes that have lost protein-coding function. Many of these coRNAs, unlike miRNAs, will be species-specific.

In this scheme, coRNAs and mRNAs interact to modify the linear flow of information from DNA to protein[84,85]. They may evolve in parallel[85] or in concert[84]. New phenotypes arise when newly generated coRNAs extract different subsets of information from the genome. This process exploits the inherent flexibility of the underlying RNA-based programs. It is not dependent on mutation of protein-coding genes in the way that hard-wired plans are. Rather, it depends on a coRNA-dependent change in the way RNAs are produced and processed[10]. In this view, the phenotypic diversity underlying the evolution of soft-wired organisms is generated through the creation and selection of new coRNAs.

These processes favor the rapid evolution of soft-wired organisms. Unlike the creation of new protein domains by exon shuffling, this mode of evolution is not constrained by the need to maintain reading frame[74]. Nor is it constrained by the selective pressures that have conserved promoter and enhancer sequences over the millions of years that separate fish and humans[87]. Instead, new coRNAs would lead to the assembly of new regulatory complexes on conserved DNA elements, new patterns of TGS and PTGS, altered processing or translation of transcripts and changes to patterns of tissue-specific gene expression during development. Exploration of new evolutionary spaces can occur without completely abandoning adaptations that were advantageous in the past: potentially, only the subset of information extracted from DNA need change during this process. The genome thus serves as a 'junkyard' of possibilities through which RNA-directed processes sift to create a selective advantage for their host[10].

Replication of ribotypes

Transmission of successful ribotypes to subsequent generations is essential to any evolutionary program (**Fig. 3**). In soft-wired organisms, elements written from RNA to DNA during gametogenesis by reverse transcription are fixed in the nucleotide sequence of the genome, ensuring their communication to offspring. But epigenetic modifications, such as DNA and histone methylation, can also be relayed to descendents. As discussed above, many of these modifications probably arise through RNA-dependent processes and thus represent a means by which ribotypic images can be written to chromatin and carried from one generation to the next. For example, imprinting is determined by the parent of origin of a chromosome, indicating that at some point maternal and paternal chromosomes are marked so that they can be distinguished during embryogenesis. In other cases, a particular modification can be transmitted from either the father or the mother. The apparent nonmendelian transmission of the mouse axin fused epiallele (*Axin*[Fu]) is one example of this process[88]. The *Axin*[Fu] allele contains an intronic IAP retroelement insertion that produces a truncated transcript, leading to a kinked tail. Silencing of the IAP promoter by methylation restores a wild-type phenotype. The silent methylated state is transmitted to offspring in a strain-specific manner. For the 129P4Rr/Rk strain there is a 70% chance of transmitting the modification when present in the mother and a 40% chance when coming from the father[88]. There is no paternal transmission in C57BL/6J crosses[88]. Similarly, methylation of other retroviral regulatory elements may undergo variable erasure during primordial germ

cell development[89] producing 'compound epigenetic mosaic' individuals[44]. The persistence of such epigenetic marks is of relevance to the origin of complex diseases, which are characterized by the absence of mendelian inheritance[90]. Here, the susceptibility of offspring to disease can depend on whether there is maternal or paternal history of disease as well as ethnicity[91].

Transmission of ribotypes also occurs more directly. The embryo receives RNA from the mother that is important in specifying cell fate. Disruption of this process in flies by mutation of the maternal-effect genes enhancer of zeste and extra sex combs leads to the homeotic substitution of one segmental body plan for another by altering PRE methylation patterns[92,93]. By establishing chromosomal states in the embryo, maternal RNAs initiate read-out of appropriate ribotypes. Variations in this process may alter the temporal and spatial expression of developmental programs, leading to phenotypic variation between individuals and eventually to the heterochronic evolution of new body plans[94].

The fetus is also exposed to the maternal environment, which can influence the fetal phenotype. For example, pregnant female mice fed a diet rich in methyl donors rather than a standard laboratory diet have litters with fewer yellow-colored agouti A^{vy} offspring, reflecting enhanced silencing of the retroviral promoter in this allele[46,95]. In other cases, integration of signals received from maternal hormones may trigger epigenetic modifications that alter long-term phenotypic development by modulating RNA coregulatory networks[96]. Low birth weight, for example, has been shown to correlate with lifetime risk of cardiovascular disease and diabetes mellitus, presumably by resetting developmental programs[97]. Furthermore, the recent demonstration that plasma RNA is quite stable[98] raises the question of whether coRNAs secreted by various somatic tissues are used to transmit information from mother to fetus. Epigenetic changes programmed before the germ line is segregated from the soma during gastrulation[99] not only will affect fetal phenotype but also may undergo transmission to the next generation. Collectively, the above mechanisms allow the communication of successful maternal ribotypes to progeny.

Origins

The four Rs have always been part of the evolutionary process, their genetic coding changing with time and circumstance. Mechanisms in which information was copied from one nonrandom surface to another, allowing for their replication, emerged early. Depending on the fidelity of this process and the nature of the interaction, surfaces were rewritten and read-out in different ways, generating diversity and new functionality. Among the different surface combinations possible, those that catalyzed each other's formation were favored. The surfaces that best reconciled the prevailing physical and chemical constraints persisted, while those capable of adaptation prevailed. The surfaces may at times have been composed of clay, lipids, carbohydrates, polypeptides or polynucleotides of various descriptions and in different combinations, depending on such processes as noncovalent assembly, oligomerization and polymerization, but the emergence of life as we know it was marked by the arrival of 'RNA world' (ref. 100). RNA was informational, acted as its own template and was capable of catalysis; it could promote its own perpetuation, either directly or by co-opting other partners along the way. The preferential replication of selfish RNA has defined the evolutionary space for soft-wired organisms ever since[2,9]. In humans, we find that much of the DNA genome has arisen from retrotransposition and that RNA-directed processes impact the read-out of RNA from genes and also rewrite their RNA transcripts. We find that non–protein coding RNAs are central to the translation of coding

RNAs and to the coregulation of other cellular events. Through combinatorial processes that lead to new coRNAs, these events can vary over time. New RNA spaces can be explored in search of those that create a selective advantage for their host. In this view, the genome of soft-wired organisms is co-opted as a canvas for the creation, storage and dissemination of successful, but ultimately selfish, RNAs.

1. Jordan, I.K., Rogozin, I.B., Glazko, G.V. & Koonin, E.V. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19**, 68–72 (2003).
2. Brosius, J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118**, 99–116 (2003).
3. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
4. Zavolan, M. *et al.* Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.* **13**, 1290–1300 (2003).
5. Yelin, R. *et al.* Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21**, 379–386 (2003).
6. Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S. & Hayashizaki, Y. Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.* **13**, 1324–1334 (2003).
7. Numata, K. *et al.* Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res.* **13**, 1301–1306 (2003).
8. Pennisi, E. Human genome. A low number wins the GeneSweep Pool. *Science* **300**, 1484 (2003).
9. Orgel, L.E. & Crick, F.H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
10. Herbert, A. & Rich, A. RNA processing and the evolution of eukaryotes. *Nat. Genet.* **21**, 265–269 (1999).
11. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
12. Hannon, G.J. RNA interference. *Nature* **418**, 244–251 (2002).
13. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
14. Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, 25–33 (2000).
15. Caudy, A.A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411–414 (2003).
16. Hutvagner, G. & Zamore, P.D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
17. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
18. Roignant, J.Y. *et al.* Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in *Drosophila*. *RNA* **9**, 299–308 (2003).
19. Chi, J.T. *et al.* Genomewide view of gene silencing by small interfering RNAs. *Proc. Natl. Acad. Sci. USA* **100**, 6343–6346 (2003).
20. Mourrain, P. *et al.* Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101**, 533–542 (2000).
21. Schramke, V. & Allshire, R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301**, 1069–1074 (2003).
22. Smardon, A. *et al.* EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10**, 169–178 (2000).
23. Schwarz, D.S., Hutvagner, G., Haley, B. & Zamore, P.D. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol. Cell* **10**, 537–548 (2002).
24. Stein, P., Svoboda, P., Anger, M. & Schultz, R.M. RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA* **9**, 187–192 (2003).
25. Finnegan, E.J., Margis, R. & Waterhouse, P.M. Posttranscriptional gene silencing is not compromised in the *Arabidopsis* CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from *Drosophila*. *Curr. Biol.* **13**, 236–240 (2003).
26. Carmell, M.A., Xuan, Z., Zhang, M.Q. & Hannon, G.J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
27. Carrington, J.C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
28. Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. & Tuschl, T. New microRNAs from mouse and human. *RNA* **9**, 175–179 (2003).
29. Dostie, J., Mourelatos, Z., Yang, M., Sharma, A. & Dreyfuss, G. Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* **9**, 180–186 (2003).
30. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. & Bartel, D.P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
31. Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* **34**, 261–262 (2003).
32. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
33. Lai, E.C. Micro RNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* **30**, 363–364 (2002).
34. Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M. & Ambros, V. Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-Complex gene activity. *Dev. Biol.* **259**, 9–18 (2003).
35. Bashirullah, A. *et al.* Coordinate regulation of small temporal RNAs at the onset of *Drosophila* metamorphosis. *Dev. Biol.* **259**, 1–8 (2003).
36. Wassenegger, M. RNA-directed DNA methylation. *Plant Mol. Biol.* **43**, 203–220 (2000).
37. Zilberman, D., Cao, X. & Jacobsen, S.E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**, 716–719 (2003).
38. Volpe, T.A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
39. Hall, I.M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
40. Akhtar, A., Zink, D. & Becker, P.B. Chromodomains are protein-RNA interaction modules. *Nature* **407**, 405–409 (2000).
41. Breiling, A., Turner, B.M., Bianchi, M.E. & Orlando, V. General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* **412**, 651–655 (2001).
42. Pal-Bhadra, M., Bhadra, U. & Birchler, J.A. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol. Cell* **9**, 315–327 (2002).
43. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
44. Whitelaw, E. & Martin, D.I. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**, 361–365 (2001).
45. Argeson, A.C., Nelson, K.K. & Siracusa, L.D. Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics* **142**, 557–567 (1996).
46. Waterland, R.A. & Jirtle, R.L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).
47. Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
48. Lehner, B., Williams, G., Campbell, R.D. & Sanderson, C.M. Antisense transcripts in the human genome. *Trends Genet.* **18**, 63–65 (2002).
49. Paule, M.R. & White, R.J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.* **28**, 1283–1298 (2000).
50. Sleutels, F., Zwart, R. & Barlow, D.P. The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
51. Nikaido, I. *et al.* Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res.* **13**, 1402–1409 (2003).
52. Migeon, B.R., Lee, C.H., Chowdhury, A.K. & Carpenter, H. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am. J. Hum. Genet.* **71**, 286–293 (2002).
53. Lee, J.T. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell* **103**, 17–27 (2000).
54. Drewell, R.A., Bae, E., Burr, J. & Lewis, E.B. Transcription defines the embryonic domains of *cis*-regulatory activity at the *Drosophila* bithorax complex. *Proc. Natl. Acad. Sci. USA* **99**, 16853–16858 (2002).
55. Rank, G., Prestel, M. & Paro, R. Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol. Cell. Biol.* **22**, 8026–8034 (2002).
56. Haddad, F., Bodell, P.W., Qin, A.X., Giger, J.M. & Baldwin, K.M. Role of antisense RNA in coordinating cardiac myosin heavy chain gene switching. *J. Biol. Chem.* **278**, 37132–37138 (2003).
57. Kumar, M. & Carmichael, G.G. Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc. Natl. Acad. Sci. USA* **94**, 3542–3547 (1997).
58. Lanz, R.B., Razani, B., Goldberg, A.D. & O'Malley, B.W. Distinct RNA motifs are important for coactivation of steroid hormone receptors by steroid receptor RNA activator (SRA). *Proc. Natl. Acad. Sci. USA* **99**, 16081–16086 (2002).
59. Reichman, T.W. & Mathews, M.B. RNA binding and intramolecular interactions modulate the regulation of gene expression by nuclear factor 110. *RNA* **9**, 543–554 (2003).
60. Kwek, K.Y. *et al.* U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.* **9**, 800–805 (2002).
61. Nguyen, V.T., Kiss, T., Michels, A.A. & Bensaude, O. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322–325 (2001).
62. Yang, Z., Zhu, Q., Luo, K. & Zhou, Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317–322 (2001).
63. Franke, A. & Baker, B.S. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol. Cell* **4**,

117–122 (1999).

64. Britten, R.J. & Davidson, E.H. Gene regulation for higher cells: a theory. *Science* **165**, 349–357 (1969).

65. Erdmann, V.A. *et al*. The non-coding RNAs as riboregulators. *Nucleic Acids Res.* **29**, 189–193 (2001).

66. Horton, T.L. & Landweber, L.F. Rewriting the information in DNA: RNA editing in kinetoplastids and myxomycetes. *Curr. Opin. Microbiol.* **5**, 620–626 (2002).

67. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).

68. Vacek, M., Sazani, P. & Kole, R. Antisense-mediated redirection of mRNA splicing. *Cell. Mol. Life Sci.* **60**, 825–833 (2003).

69. Hastings, M.L., Ingle, H.A., Lazar, M.A. & Munroe, S.H. Post-transcriptional regulation of thyroid hormone receptor expression by *cis*-acting sequences and a naturally occurring antisense RNA. *J. Biol. Chem.* **275**, 11507–11513 (2000).

70. Knight, S.W. & Bass, B.L. The role of RNA editing by ADARs in RNAi. *Mol. Cell* **10**, 809–817 (2002).

71. Zhang, Z. & Carmichael, G.G. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* **106**, 465–475 (2001).

72. Jackson, D.A., Pombo, A. & Iborra, F. The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *FASEB J.* **14**, 242–254 (2000).

73. Maquat, L.E. & Carmichael, G.G. Quality control of mRNA function. *Cell* **104**, 173–176 (2001).

74. Borsu, L., Presse, F. & Nahon, J.L. The AROM gene, spliced mRNAs encoding new DNA/RNA-binding proteins are transcribed from the opposite strand of the melanin-concentrating hormone gene in mammals. *J. Biol. Chem.* **275**, 40576–40587 (2000).

75. Feldser, D.M., Hackett, J.A. & Greider, C.W. Opinion: Telomere dysfunction and the initiation of genome instability. *Nat. Rev. Cancer* **3**, 623–627 (2003).

76. Madireddi, M.T. *et al*. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell* **87**, 75–84 (1996).

77. Mochizuki, K., Fine, N.A., Fujisawa, T. & Gorovsky, M.A. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* **110**, 689–699 (2002).

78. Yao, M.C., Fuller, P. & Xi, X. Programmed DNA deletion as an RNA-guided system of genome defense. *Science* **300**, 1581–1584 (2003).

79. Hall, I.M., Noma, K. & Grewal, S.I. RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. *Proc. Natl. Acad. Sci. USA* **100**, 193–198 (2003).

80. Cassidy, S.B., Dykens, E. & Williams, C.A. Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am. J. Med. Genet.* **97**, 136–146 (2000).

81. Skaletsky, H. *et al*. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).

82. Eddy, E.M. Male germ cell gene expression. *Recent Prog. Horm. Res.* **57**, 103–128 (2002).

83. Bailey, J.A., Liu, G. & Eichler, E.E. An alu transposition model for the origin and expansion of human segmental duplications. *Am. J. Hum. Genet.* **73**, 823–834 (2003).

84. Herbert, A. RNA editing, introns and evolution. *Trends Genet.* **12**, 6–9 (1996).

85. Mattick, J.S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2**, 986–991 (2001).

86. Negroni, M. & Buc, H. Mechanisms of retroviral recombination. *Annu. Rev. Genet.* **35**, 275–302 (2001).

87. Aparicio, S. *et al*. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).

88. Rakyan, V.K. *et al*. Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc. Natl. Acad. Sci. USA* **100**, 2538–2543 (2003).

89. Hajkova, P. *et al*. Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).

90. Risch, N.J. Searching for genetic determinants in the new millennium. *Nature* **405**, 847–856 (2000).

91. Kardia, S.L., Modell, S.M. & Peyser, P.A. Family-centered approaches to understanding and preventing coronary heart disease. *Am. J. Prev. Med.* **24**, 143–151 (2003).

92. Czermin, B. *et al*. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).

93. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).

94. Gould, S.J. *Ontogeny and Phylogeny* (Belknap, Cambridge, 1977).

95. Wolff, G.L., Kodell, R.L., Moore, S.R. & Cooney, C.A. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J.* **12**, 949–957 (1998).

96. Godfrey, K.M. The role of the placenta in fetal programming-a review. *Placenta* **23** Suppl A, S20–S27 (2002).

97. Barker, D.J., Eriksson, J.G., Forsen, T. & Osmond, C. Fetal origins of adult disease: strength of effects and biological basis. *Int. J. Epidemiol.* **31**, 1235–1239 (2002).

98. Ng, E.K. *et al*. mRNA of placental origin is readily detectable in maternal plasma. *Proc. Natl. Acad. Sci. USA* **100**, 4748–4753 (2003).

99. Saffman, E.E. & Lasko, P. Germline development in vertebrates and invertebrates. *Cell. Mol. Life Sci.* **55**, 1141–1163 (1999).

100. Joyce, G.F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).