

SURVEY AND SUMMARY

Experimental approaches to identify non-coding RNAs

Alexander Hüttenhofer* and Jörg Vogel¹

Innsbruck Biocenter, Division of Genomics and RNomics, Innsbruck Medical University, Fritz-Pregl-Str. 3, 6020 Innsbruck, Austria and ¹Max Planck Institute for Infection Biology, Schumannstrasse 21-22, 10117 Berlin, Germany

Received December 21, 2005; Revised and Accepted January 10, 2006

ABSTRACT

Cellular RNAs that do not function as messenger RNAs (mRNAs), transfer RNAs (tRNAs) or ribosomal RNAs (rRNAs) comprise a diverse class of molecules that are commonly referred to as non-protein-coding RNAs (ncRNAs). These molecules have been known for quite a while, but their importance was not fully appreciated until recent genome-wide searches discovered thousands of these molecules and their genes in a variety of model organisms. Some of these screens were based on biocomputational prediction of ncRNA candidates within entire genomes of model organisms. Alternatively, direct biochemical isolation of expressed ncRNAs from cells, tissues or entire organisms has been shown to be a powerful approach to identify ncRNAs both at the level of individual molecules and at a global scale. In this review, we will survey several such wet-lab strategies, i.e. direct sequencing of ncRNAs, shotgun cloning of small-sized ncRNAs (cDNA libraries), microarray analysis and genomic SELEX to identify novel ncRNAs, and discuss the advantages and limits of these approaches.

INTRODUCTION

Non-protein-coding RNAs (ncRNAs) do not encode proteins but function directly at the level of the RNA in the cell. Over the last few years, the importance of this surprisingly diverse class of molecules has been widely recognized (1–5). ncRNAs have been identified in unexpectedly large numbers, with present estimates—based on bioinformatical approaches—in the range of thousands per eukaryal and hundreds per bacterial genome (6–9). They play key roles in a variety of

fundamental processes in all three domains of life, i.e. Eukarya, Bacteria and Archaea. Their functions include DNA replication and chromosome maintenance, regulation of transcription, RNA processing (not only RNA cleavage and religation, but also RNA modification and editing), translation and stability of mRNAs, and even regulation of stability and translocation of proteins (4,5,10–13). Many of them have been discovered fortuitously, suggesting they merely represent the tip of the iceberg. Many known ncRNAs are small, i.e. typically <500 nt, and thus much shorter than the majority of mRNAs. However, eukaryotes also express a number of large ncRNAs, e.g. Xist or Air RNAs, which are several 1000 nt long (14–16). The highly specific roles of ncRNAs reflect in most cases their ability to selectively bind a small set of proteins as well as their potential to specifically recognize definite RNA targets via regions of sequence-complementarity.

In recent years, new bioinformatical and experimental strategies have been taken to identify a great number of novel ncRNA candidates in various model organisms from *Escherichia coli* to *Homo sapiens* (5–7,17–31). These findings demonstrated that the number of ncRNAs in genomes of model organisms is much higher than it had been anticipated.

In the following, we will review various experimental strategies that were employed to identify novel ncRNAs in genomes of model organisms. For these approaches, the term ‘Experimental RNomics’ has been coined (3). Four different methods will be presented and their advantages as well as their obstacles in the identification of novel ncRNA molecules will be discussed: (i) RNA sequencing (enzymatically or chemically) as the most traditional method to reveal novel ncRNA species; (ii) the parallel cloning of many ncRNA by generating specialized cDNA libraries; (iii) the use of microarrays to predict ncRNAs that are expressed under a given experimental condition; (iv) ‘genomic SELEX’ and its potential application to select ncRNA candidates from the sequence space represented by the genome of an organism of interest.

*To whom correspondence should be addressed. Tel: +43 (0) 512 507 3630; Fax: +43 (0) 512 507 9880; Email: alexander.huettnerhofer@i-med.ac.at

Alternatively to biochemical methods, genetic and bio-formatical tools may also be employed to identify ncRNAs in model organisms. In fact, some of the first chromosomally encoded regulatory ncRNAs, e.g. MicF, DsrA and RprA of *E.coli*, were discovered in the course of a genetic screen (32–34). Similarly, genetics also discovered the founding member, lin-4 RNA, of the ever-growing class of eukaryotic miRNAs (35). Due to space constraints, however, we would like to refer the reader to (6,36,37) for a more detailed review of genetic and bio-computational routes to ncRNA discovery.

Identification of ncRNAs by chemical or enzymatic sequencing

In the very early days of ncRNA research, e.g. some 35–40 years ago, single ncRNA species (at the time ribosomal RNAs, tRNAs or viral RNAs) were selected by size-separation of total RNA on denaturing gels, followed by visualization and excision of specific bands, ideally representing single ncRNA species. Thus, for its identification, the ncRNA of interest must be present in high amounts, e.g. visible as a

distinct band in an ethidium bromide-stained polyacrylamide gel, exposed to ultraviolet (UV) light (Figure 1A).

Subsequently (and prior to their identification by sequence analysis), ncRNAs are labelled either at their 5' end or at their 3' end: (i) for labelling of RNAs at their 5' end, the mono- or triphosphate group usually found at the 5' end of ncRNAs is removed first. This is achieved by the addition of calf intestinal alkaline phosphatase at an elevated temperature; inactivation of the enzyme is performed by repeated extraction with phenol/chloroform or by gel purification (38). Labelling of the RNA is then performed by the addition of polynucleotide kinase in the presence of [γ - 32 P]ATP (38). (ii) For labelling at their 3' end, ncRNAs can be labelled by the procedure described by Bruce and Uhlenbeck using 5'- 32 PpCp as a donor molecule in the presence of T4 RNA ligase (39). Subsequently, ncRNAs are gel-purified on denaturing polyacrylamide gels.

RNAs may also be labelled *in vivo* prior to extraction from an organism. In some early studies, *E.coli* total RNA was metabolically labelled with orthophosphate, 32 PO $_4^{3-}$ (40–43). Orthophosphate is readily taken up by

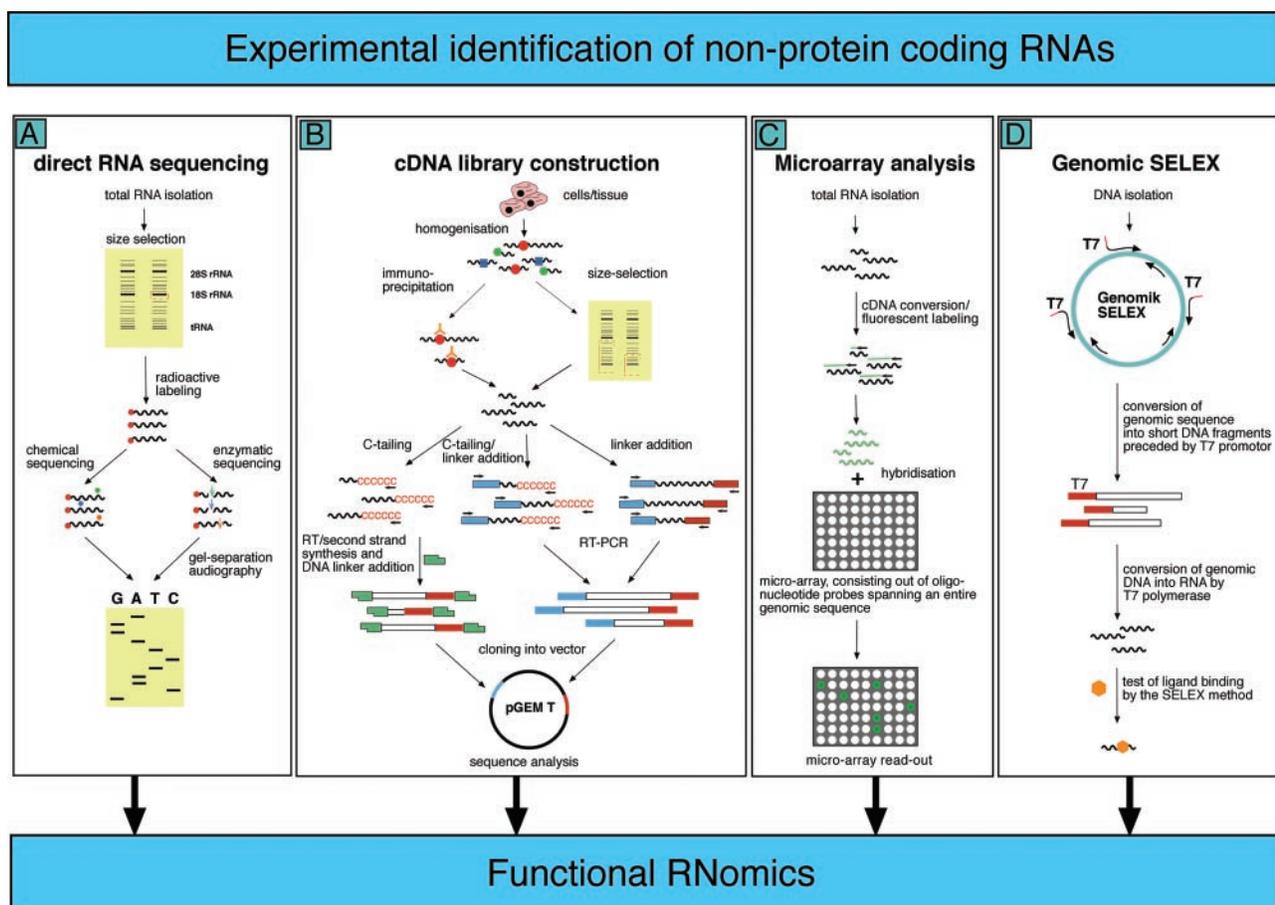


Figure 1. Four experimental approaches (A–D) to identify candidates for ncRNAs are shown. (A) Identification of ncRNAs by chemical or enzymatic sequencing of extracted abundant RNAs. (B) Identification of ncRNAs by cDNA cloning and sequencing; three different methods are indicated to reverse transcribe ncRNAs, usually lacking poly(A) tails, into cDNAs (e.g. by C-tailing, C-tailing and linker addition, or linker addition, only, followed by RT–RCR). (C) Identification of ncRNAs by micro-array analysis. DNA oligonucleotide covering the sequence space of an entire genome are spotted onto glass slides, to which fluorescently labelled samples, derived from cellular RNA, is hybridized. (D) Identification of ncRNAs by genomic SELEX. By random priming, the sequence of a genome is converted into short PCR fragments containing a T7 promoter at their 5' ends. Subsequently, *in vitro* transcription by means of T7 RNA polymerase converts this genomic sequence of an organism into RNA fragments, which can then be assayed for function, such as binding to a specific protein or small chemical ligand, by SELEX.

growing cells and incorporated into nucleic acids. Different from the aforementioned 5' or 3' labelling procedures, the extracted total RNA is randomly labelled at any nucleotide. Such uniformly labelled RNAs are mainly used for 'RNA fingerprinting' techniques (see below).

After extraction from a cell or organism, size separation by PAGE and elution from the gel, ncRNAs are identified by sequence analysis. This is either achieved by 2D RNA fingerprinting or by enzymatic or chemical sequencing of ncRNAs.

There are several versions of 2D RNA fingerprinting techniques to sequence small RNAs (or oligonucleotides) or prepare various RNase-digested oligonucleotide catalogs. The differences are use of uniformly or end-labelled RNAs, partial or complete digestion with various RNases, electrophoresis on cellulose acetate strips or in acrylamide gels for the first dimension, electrophoresis on DEAE-cellulose paper, or homochromatography on DEAE-cellulose plates, or gradient thin layer chromatography on DEAE-cellulose plates (44–47).

For enzymatic sequence analysis, labelled ncRNAs (at 5' or 3' ends) are subjected to partial digestion with base-specific ribonucleases at elevated temperatures (50–55°C) and in the presence of 7 M urea to avoid interference of the secondary/tertiary structure of the RNA with enzymatic hydrolysis steps. For base-specific cleavage, a plethora of RNases (RNase T1, T2, U2, PHY1, PHY M, CL3, A or M1) can be used which cleave preferentially 3' to either G, C, U or A bases (48–50). To resolve obtained RNA fragments by size, 1D gel electrophoresis is carried out on denaturing polyacrylamide gels (see below).

For chemical sequence analysis of ncRNAs, four different base-specific chemical reactions generate a means of directly sequencing RNA that was terminally labelled with ³²P (51). After a partial specific modification of each kind of RNA base, an amine-catalysed strand scission generates labelled fragments whose length determine the positions of each nucleotide in a sequence. Dimethyl sulfate modifies guanosine, diethyl pyrocarbonate attacks primarily adenosine, hydrazine attacks uridine and cytidine, but salt suppresses the reaction with uridine. In all cases, aniline induces a subsequent strand scission (51).

Subsequent to enzymatic or chemical sequencing, electrophoretic fractionation of the labelled fragments is achieved on denaturing polyacrylamide gels, followed by autoradiography, which allows determination of the RNA sequence of interest.

The earliest studies to identify RNA molecules by direct sequencing were performed on tRNAs as well as on ribosomal RNAs (48,52–54). In the case of 16S ribosomal RNA, exhibiting a size of ~1500 nt, smaller fragments were first generated by RNase T1 cleavage and subsequently analysed by RNase fingerprinting techniques (54). Direct RNA sequencing for identification of novel RNA species is far from being outdated, as was shown in more recent studies: by labelling and direct RNA sequencing, a novel class of ncRNAs, designated as small nucleolar RNAs, involved in rRNA modification (55) could be identified in eukaryotes. Lately, this technique was also used to visualize and subsequently sequence abundant RNAs of gram-positive bacteria (56,57).

Obstacles and advantages of the method

Identification of novel ncRNA species by RNA sequencing encounters four main obstacles. First, for identification, ncRNAs have to be highly abundant to be visible as single bands in ethidium-bromide stained gels; to circumvent this problem, labelling of total RNA, followed by size separation on a gel system (e.g. *vice versa* as described above), allows identification of less abundant ncRNA species.

Second, no other ncRNAs in the same size range should be present in the total RNA population, since it would hamper isolation of a single RNA species and thus would result in ambiguous sequencing data. If a band or spot is found to contain multiple RNA species, these can be resolved by 2D gel electrophoresis, which allows separation of RNA species with similar or identical sizes.

Third, chemical or enzymatic sequencing of ncRNAs sometimes results in sequencing data that are difficult to interpret. The reason being that, for enzymatic sequencing, RNases are not strictly specific for a distinct base but possess residual cleavage activity for other bases; similarly, chemical sequencing does not always result in unambiguous modification and cleavage of nucleotides, thus obscuring the readout of obtained sequence data.

Finally, due to the sequencing methods and resolution capacity of polyacrylamide gels, sequencing is limited to RNAs sized—at the most—a couple of hundred nucleotides. Thus, ncRNA species, which exceed this size range, cannot be directly analysed by this method, but have to be cleaved into smaller pieces (e.g. by T1 nuclease digestion) prior to further analysis.

The advantage of direct RNA sequencing, as compared with sequencing cDNA clones generated from ncRNAs (see below), is the fact that ncRNAs do not have to be reverse transcribed for analysis. Thus, RNA secondary/tertiary structures that might impede reverse transcription into cDNA do not interfere with RNA identification by using direct RNA sequencing.

Identification of ncRNAs by specialized cDNA libraries

The second method for the identification of novel ncRNA species involves the generation of cDNA libraries, in analogy to expressed sequence tag libraries (EST libraries) for identification of mRNAs (58,59). The original mRNA cloning method is based on reverse transcription of mRNAs from an organism by an oligo(dT) primer and second strand synthesis, resulting in a cDNA library that ideally represents all protein-coding transcripts of a genome. Compared with these conventional EST libraries, the main difference for ncRNA library approaches is the source and treatment of the cloned RNA.

Since most mRNAs are >500 nt in length but many ncRNAs considerably smaller, first RNAs in the size range of ~20–500 nt are isolated. This fraction is usually depleted in EST libraries as it will not be present in poly(A)⁺ mRNA. The isolation of small-sized RNAs is achieved by size separation of total RNA (either from the entire organism at different developmental stages or from an individual organ) by denaturing PAGE (Figure 1B).

Alternatively, by employing an antibody against an RNA-binding protein of interest, entire groups of ncRNAs

can be isolated by immunoprecipitation. Thereby, RNAs are not selected by their size but rather based on their function since they bind to a common RNA binding protein, e.g. a library generated by immunoprecipitation with an antibody against a common small nucleolar RNA–protein will help identify ncRNAs from the class of snoRNAs (26).

In many cases, these size- or antibody-selected RNAs will lack poly-adenylated tails. In general, there are three different methods to reverse transcribe ncRNAs into cDNA as a prerequisite for cloning and sequencing (Figure 1B).

- (i) First, to generate cDNA from this ncRNA fraction, addition of an oligo(C) or oligo(A) tail to the RNA is performed in the presence of poly(A) polymerase, which uses ATP, but also—to a lesser extent CTP—as a substrate (60). Subsequently, tailed RNAs are reverse transcribed employing an oligo(dG) or oligo(dT) primer, respectively. Following second strand synthesis by employing DNA polymerase I and limited amounts of RNase H and subsequent ligation of double-stranded DNA linkers, the obtained double-stranded cDNAs are cloned into a standard vector system (e.g. pSPORT1/GibcoBRL), thus generating a cDNA library [for a detailed method description see (61)].
- (ii) As a second approach subsequent to C-tailing at the 3' end (see above), an oligonucleotide linker is ligated to 5' end of ncRNAs by T4 RNA ligase. The oligonucleotide can be made from RNA or almost entirely from DNA [for a more detailed method description see (62)]. To avoid multimerization of linker sequences, the 5'-oligonucleotide carries a 5'-hydroxyl group. Since T4 RNA ligase uses RNA as a template, the last 3 nt at the 3' end of the oligonucleotide should be ribonucleotides to increase efficiency of ligation. To add a linker to RNAs with modified 5' ends, such as a cap structure or a tri-phosphate group, the RNA is first treated with tobacco acid pyrophosphatase (TAP) which cleaves between the α and β phosphate group, thus leaving 5'-monophosphates (62). For RT-PCR of RNAs, an oligo(dC) or d(T) primer is used in combination with a 5'-primer that is complementary to the ligated 5' linker sequence.
- (iii) In a third method, RNA oligonucleotide linkers are sequentially ligated to both the 3' and the 5' end by T4 RNA ligase. To avoid multimerization of linker sequences, the oligonucleotide at the 5' end of the RNA lacks a phosphorylated 5' end, while the oligonucleotide ligated to the 3' end of the RNA contains a blocked 3' end. Typically, the entire RNA pool is subjected to another round of gel extraction after the first linker ligation step to remove excessive linker that would otherwise form dimers with the second adapter oligo. As described above, the terminal 3 nt of the 5'-oligonucleotide linker and the first three of the 3'-oligonucleotide linker might contain RNA bases to increase efficiency of ligation by T4 RNA ligase. RT-PCR of the ligated RNA fraction is achieved by DNA primers complementary to the respective 5'- or 3'-linker oligonucleotide [for a more detailed method description see (61)].

Subsequent to cDNA synthesis, cDNA fragments are cloned into standard vector systems and sequenced by cycle sequencing. Dependent on the expected complexity of the library, up

to 10 000 cDNA clones should be sequenced (for example in the case of large eukaryal genomes). Sequencing is usually followed by bioinformatical analyses, e.g. mapping of the ncRNA gene to a certain locus on the genome and identification of structure or sequence motifs, which might contribute to the identification of the function of the ncRNA species of interest.

In the recent past, numerous studies have been performed to identify ncRNAs in genomes of model organisms by constructing specialized cDNA libraries. The first study was initiated in the mouse *Mus musculus*, where by a cDNA library derived from size-selected RNAs (50–500 nt) 201 candidates for ncRNAs were identified from ~5000 cDNA clones analysed, about half of which belonging to the class of snoRNAs (19). This study was followed by using a similar approach for the plant *Arabidopsis thaliana* (20), the fruit fly, *Drosophila melanogaster* (23), the two archaeal species *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* (21,22) and the eubacteria *E.coli* (63,64) and *Aquifex aeolicus* (65).

Specialized cDNA library cloning was also applied to identify certain subclasses of ncRNAs, e.g. miRNAs, in different model organisms. Here, ncRNAs with a very narrow size range of about 18–25 nt, i.e. centering around the known sizes of miRNAs, were size-selected, cloned and sequenced (10,66–72).

Identification of small ncRNAs by generation of specialized cDNA libraries is now wide-spread and includes analysis of amoebozoia such as the slime mold *Dictyostelium discoideum* (62). For identification of specific classes of ncRNAs (such as snoRNAs) the method of cDNA library generation by immunoprecipitation with a snoRNA-binding protein like fibrillarlin followed by cloning of ncRNAs has been employed successfully for C/D and H/ACA snoRNAs (26,73).

Obstacles and advantages of the method

The above method for cloning of ncRNAs has its downfalls by the fact that it might not always be possible to reverse transcribe an ncRNA into cDNA because of its structure or modification (e.g. base or backbone modifications). Thus, a cDNA library is neither likely to reflect all ncRNAs in a cell, nor will it necessarily reflect—by number of individual cDNA clones—the abundance of the respective ncRNA. The rationale behind this is that less structured/modified ncRNAs are more easily reverse transcribed than others and will be over-represented within a cDNA library; similarly, smaller ncRNAs will be more abundant than longer ones, since they are more likely to be fully reverse transcribed.

For size-selected cDNA libraries, in general, it will not be possible to identify all ncRNAs of a cell type or organism, since the cut-off by size (e.g. 20–500 nt) will prohibit identification of longer ncRNAs (such as ncRNAs like Xist and Air RNA, which exhibit sizes in the range of many kb). In addition, by the very nature of a cDNA expression library, only those ncRNA species will be detected, which are transcribed from a genome. This might depend, however, on a specific developmental state of the organism or on expression in a certain tissue. Thus, to be able to clone all expressed RNA sequences from an organism, ideally, all developmental stages, in all tissues under all possible growth and nutrient conditions would have to be analysed and total RNA extracted

from these different states. This might not always be possible and hence some ncRNA species, which are expressed under certain conditions, only, will not be cloned.

As for cloning strategies, employing method I for conversion of ncRNAs into cDNAs, e.g. the method of reverse transcription, second strand synthesis and addition of DNA linkers (see above) will not result in full-length cDNA clones, according to our experience, but in truncated 5'-termini lacking about 10–15 nt of the full-length RNA.

Conversely, the disadvantage of methods II and III involving a linker oligonucleotide (see above) is the rather inefficient ligation step of the linkers to the potential ncRNA of interest, and the failure of linker attachment to modified termini. The advantage of this method is, however, that often full-length cDNA clones can be obtained, as compared with method I.

In general, cDNA cloning will result in identification of highly abundant known ncRNA species, such as tRNAs or small ribosomal RNAs (e.g. 5S or 5.8S rRNAs). To circumvent repeated sequencing of these already known ncRNA genes, one can try to excise these ncRNA species from the gel after PAGE. However, this might result in the loss of ncRNA species exhibiting the same or similar sizes as these known RNA species. Alternatively, one can spot cDNAs on filters (as a dot blot) and hybridize filters with radiolabelled oligonucleotides directed against the most abundant known ncRNAs species. Subsequently, only those cDNA clones are sequenced, which show no hybridization signal on autoradiograms of filters.

Microarray analysis

Microarrays have become the preferred method to monitor the levels of many transcripts in parallel and often at the whole-genome level (Figure 1C). Microarrays, also known as DNA chips or expression arrays, are glass (or silicon) slides onto whose surface DNA probes have been printed in a grid-like arrangement. To date, single-stranded DNA oligonucleotides of 25–70 in length are the predominant type of DNA probe on commercial microarrays, though double-stranded PCR products may also serve as probes.

To analyse the entire level of cellular transcripts, samples are prepared from total RNA of an organism. The samples used for microarray hybridization can be the extracted RNA, the converted cDNA or cRNA; in any case, these probes will generally be labelled with fluorescent dyes, such as Cy3 or Cy5. For more details on the various labelling protocols that are currently being used, see references in (74) and the work cited below. The prepared sample is then mixed with hybridization buffer and applied to the glass slide so that they will hybridize to a spot on the microarray.

The fluorescence of the spots to which the sample hybridized is read by a scanner and the results are displayed as a pattern of coloured, e.g. red or green dots, with the colour intensity reflecting the amount of transcripts that was present in the cell. If two samples labelled with different dyes were hybridized in parallel onto the same microarray, additional colours such as yellow or orange would indicate relative amounts of the individual transcripts in the two RNA pools.

Microarrays are mostly used for mRNA expression profiling but they could also be a means for studying

ncRNA expression or even for ncRNA discovery (Figure 2). The main caveat for their use with ncRNAs, however, was—and in many cases still is—the design of the commercially available microarrays. Since tailored for mRNA profiling, most of these arrays carry probes only for coding regions, thus transcripts from non-coding genome regions will not be detected. Nonetheless, the last few years have seen considerable improvement of this situation.

In bacteria, most of the functional ncRNAs are encoded in intergenic regions (IGRs). The first microarray to include IGRs in addition to coding regions was introduced for the model bacterium *E.coli* by (75). Their high-density array (tiling array) carries ~300 000 strand-specific 25mer oligonucleotide probes for all mRNA, tRNA and rRNA regions at a 30 bp resolution as well as for all IGRs of >40 bp with 6 bp resolution.

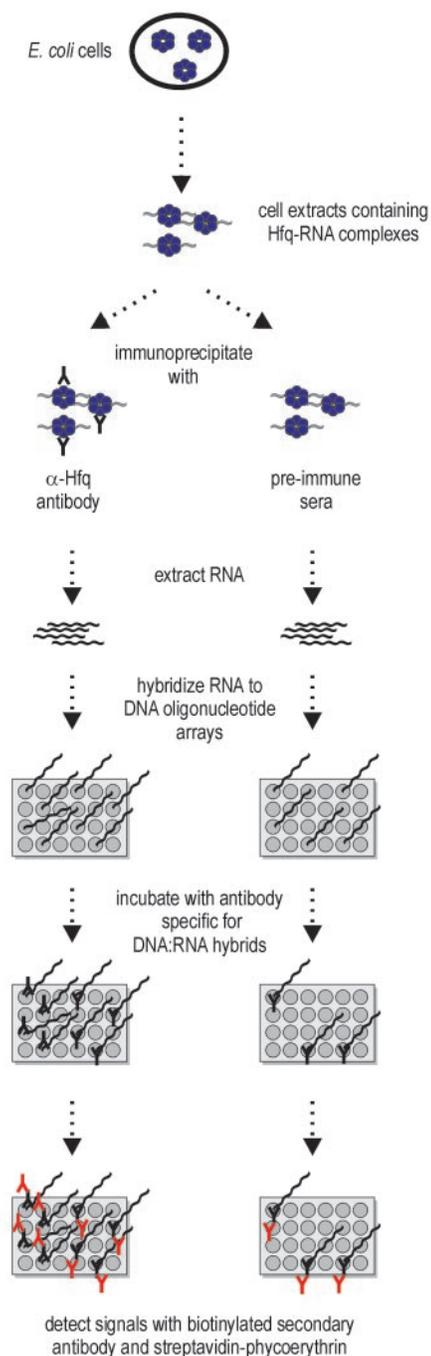
While this initial study primarily focused on technical issues of mRNA level profiling, Wassarman *et al.* (2001) subsequently used this microarray type to specifically analyse the transcriptional output from IGRs. They found that array hybridization with RNA extracted from three different growth conditions yielded signals for at least a third of the ncRNAs that were detected by parallel probing on northern blots.

These global analyses of the *E.coli* transcriptome were subsequently extended by (76). By including a much broader set of growth conditions, additional transcripts from IGRs that may be novel ncRNA candidates were detected. Notably, the extraordinarily high probe density here facilitated detection of 3'- or 5'- UTR RNA fragments that accumulate independently after the processing of mRNA transcripts.

In a third study with this microarray type, cellular RNAs that associate with *E.coli* Hfq protein were analysed (77). This bacterial Sm-like protein, over the previous years, has emerged as a key player in regulation by small regulatory ncRNAs (78) and was known to bind a number of bacterial ncRNAs (i.e. in addition to mRNAs). By the time of the study of Zhang *et al.* (2003), 46 ncRNAs were known in *E.coli*, of which ~30% were detected by array hybridizations of RNA that co-immunoprecipitated with Hfq. For bacteria other than *E.coli*, microarrays have been applied to support biocomputational prediction of ncRNAs of *Staphylococcus aureus* (56). In contrast to the aforementioned *E.coli* oligonucleotide tiling array, selected *S.aureus* IGRs were PCR amplified to yield double-stranded DNA probes that were then spotted on glass slides.

Similar to bacteria, microarrays have been increasingly used to confirm global predictions of certain classes of eukaryotic ncRNAs as well as to study their expression profile in different tissues. One such class, the ~22 nt microRNAs, is matured from 60 to 110 nt pre-miRNA hairpin transcripts thought to derive from longer pri-miRNA products. Microarrays with 40 or 60mer oligonucleotides to detect known microRNAs or their hairpin precursors were introduced recently (79,80). Barad *et al.* (80) evaluated several aspects of the methodology in order to standardize it and define the parameters needed to achieve efficient hybridization and reliable results, including mismatch analysis to determine the specificity of microRNA probes. It was observed that signal intensity correlates with the location of the microRNA sequence within the 60mer probes, showing that location at the 5' region yields the highest signals, whereas the 3' end

Analysis of Hfq-associated RNAs



Analysis of Lhp1p-associated RNAs

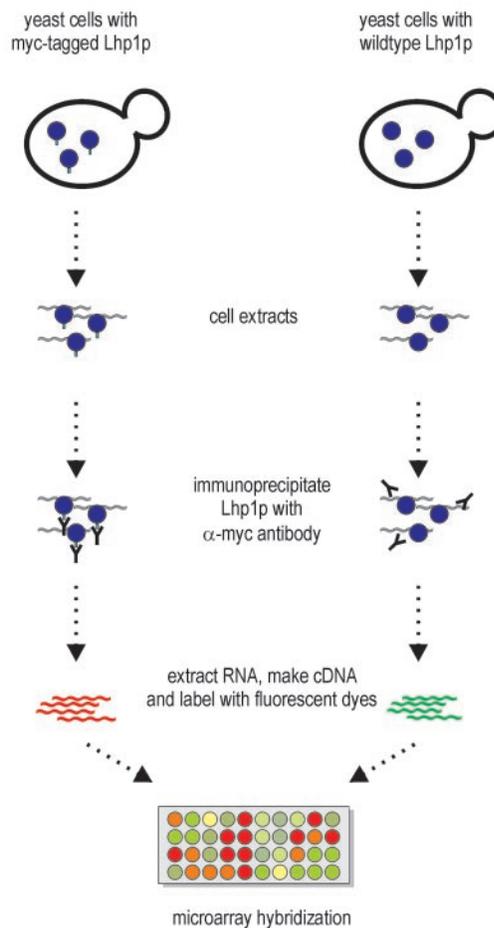


Figure 2. Microarray detection of cellular RNAs, including ncRNAs, that associate with the bacterial Sm-like protein, Hfq, and the La homologous protein (Lhp1p) of yeast (*S.cerevisiae*), respectively (77,87). Left panel: RNAs are co-immunoprecipitated from *E.coli* cell extracts with anti-Hfq antibodies, purified and hybridized to high-density microarrays that carry DNA oligonucleotides covering the entire *E.coli* genome. RNAs that hybridized to probes on the arrays are detected with an antibody that specifically sees DNA:RNA hybrids. Subsequently, such signals are detected as indicated and subtracted from those obtained in a control experiment in which cell extracts were incubated with pre-immune sera, that is no immunoprecipitation of Hfq. Right panel: cell extracts from either wild-type yeast or yeast cells that express epitope (myc)-tagged Lhp1p are incubated with an anti-myc antibody, RNA is extracted from immunoprecipitates and reverse transcribed. The two obtained cDNAs are labelled with different fluorescent dyes (Cy3, red; Cy5, green), mixed and hybridized to yeast whole-genome microarrays. Spots that yield red signals indicate that the corresponding RNA was enriched in Lhp1p-myc immunoprecipitates.

location results in poor signals. These results were subsequently used to develop an integrative approach to the discovery of new microRNAs, in which potential microRNA precursor regions were predicted in the human genome and

5300 of these candidates tested in a high-throughput manner on the aforementioned microarrays (81).

Several groups have recently used microarrays to study ncRNAs of the yeast *Saccharomyces cerevisiae*. Following

earlier work with microarrays that carry individual probes for a representative set of certain ncRNAs, e.g. snoRNAs (82), the Hughes laboratory designed a tiling microarray to cover all known and several predicted yeast ncRNAs (83). Here, each ncRNA transcript is covered by oligonucleotide probes at ~5 nt intervals including 100 nt of flanking sequence on both the 5' and 3' ends. Thus far, however, these arrays have mainly been used to monitor the synthesis, processing and modification of known ncRNAs (84,85).

New yeast ncRNAs were identified by means of a truly whole-genome microarray that contains 6700 PCR fragments to cover all yeast open reading frames, annotated small RNAs and all intergenic regions (86,87). Here, Inada and Guthrie (87) sought to identify the RNA binding partners of the yeast La protein (Lhp1) at a global scale. La is a ubiquitous, nuclear RNA-binding protein that is conserved among eukaryotes. Aside from binding mRNAs, it is known to associate with the primary transcripts of RNA polymerase III, including all tRNAs and other small RNAs. To selectively identify La binding RNAs in yeast, a Myc-tagged Lhp1 protein was immunoprecipitated with its associated RNAs and an untagged strain was used as the reference sample in subsequent microarray hybridizations (Figure 2). The La targets identified in this work included 20 annotated snoRNAs. Furthermore, at least three novel H/ACA snoRNAs that were not before annotated as such were newly discovered in intergenic regions. Additional highly enriched signals from other intergenic regions suggest that these also represent novel unannotated transcripts which may be unknown ncRNAs.

Customized tiling arrays have now also been applied to systematically search for functional ncRNAs in higher eukaryotes. For example, a biocomputational approach was taken to extract 3478 intergenic and intronic sequences that are conserved between the human, mouse and rat genomes, and that showed characteristics of ncRNAs by a number of other criteria (88). This information was then used to design tiling arrays that contained probes for this candidate set, and these arrays were probed with RNA isolated from 16 wild-type mouse tissues. Subsequently, 55 candidates for highly expressed novel ncRNAs were tested on northern blots, thus confirming eight of these as small, highly and ubiquitously expressed RNAs in mouse. Interestingly, only five of these ncRNAs could also be detected in rat tissues, but none in human tissues or cultured cells. The conserved expression of these five ncRNAs in mouse and rat may indicate these molecules to be functional in these two organisms albeit not in human.

Obstacles and advantages of the method

The aforementioned studies provided valuable clues as to the potential of this technique for ncRNA discovery as well as the problems associated with microarrays when assaying small and highly structured ncRNAs. Analyzing the hybridization signals from *E.coli* tiling arrays it was noted that often only a subset of the oligonucleotide probes within the range of a given ncRNA transcript region yielded a signal peak, even though the same sRNA locus gave a strong and distinct band on northern blots [cf. Figures 2 and 3 in (18)]. Tjaden *et al.* (76) occasionally observed transcripts on the strand opposite an experimentally validated ncRNA, which may account

either for unknown ncRNA antisense transcripts or simply for experimental noise.

Although so far techniques for the reliable microarray detection of bacterial small RNAs, which are usually highly structured, have not been thoroughly evaluated, sample preparation would seem a major issue. To date, most microarray approaches involve fluorescent labelling of the RNA to be used as sample. Frequently, the RNA is converted into cDNA in the presence of modified nucleotides that carry fluorescent dyes. Most bacterial ncRNAs cluster in a size range of 100–150 nt (8,9), and thus reverse transcription may not be efficient and could further be hampered by tight secondary structure.

Whether direct labelling approaches, e.g. chemical labelling of fragmented RNA as alternatively used in (18), would fully solve these problems is currently unknown. However, Zhang *et al.* (2003) drastically improved detection sensitivity by directly hybridizing RNA to oligonucleotide arrays without labelling or cDNA synthesis (Figure 2). Instead, hybridization was assayed using an antibody that sees RNA:DNA hybrids. The highly improved sensitivity of this method is demonstrated by the detection of the oxidative stress-induced OxyS RNA, which is present in very low concentrations under the growth conditions used in this study.

Hence, microarrays bear a great potential to not only detect many RNAs in parallel but also to point to transcripts that are present at low levels. As a note of caution, the fact that the vast majority of the mouse ncRNA candidates suggested by microarray analysis failed in downstream northern analysis (88) clearly emphasizes the need for validation of microarray hybridization results by independent methods.

These authors also point out that hybridizing covalently labelled total RNA as applied in their study, as opposed to reverse transcribed RNA derived from poly-adenylated RNA, would be important in tiling array analyses, since any amplification or enrichment steps are likely to skew the representation of the large noncoding regions of eukaryotic genomes and may thus make it difficult to distinguish such signals from global 'transcriptional noise'. The application of stringent criteria when using microarrays for ncRNA discovery seems to be imperative as more data from whole genome tiling microarrays are becoming available (89–91) and this data will increasingly serve as input for biocomputational ncRNA predictions by others [e.g. (7)].

Genomic SELEX

Many ncRNAs form ribonucleo-protein particles (RNPs) at various time points in their life cycle. Such RNA-binding proteins may help an ncRNA fold into its active conformation, shield it from nucleases prior to exerting its function or promote its annealing with target RNAs up to guiding a protein to its proper target. Other ncRNAs interact with proteins to directly regulate their activity.

The techniques discussed so far allow to identify ncRNAs from the pool of expressed cellular RNAs after co-purification with proteins, i.e. by cloning, direct sequencing or microarray analysis. Given that many such proteins bind their RNA ligands in a nanomolar range, it should also be possible to select RNA ligands from the pool of ncRNAs that an organism can possibly express even without isolating their *in vivo* transcripts.

This approach, termed genomic SELEX (92), is based on the *in vitro* generation of RNA species that are derived from a library of an organism's entire genomic DNA (Figure 1D). The generated RNA pool will undergo successive rounds of association with a given RNA-binding protein, partitioning and re-amplification. As a result, RNA sequences that are stringently bound by the protein partner will be enriched. Once the sequence of the bound RNAs is determined, this information can be used to search for matches in the genome, and so predicted genomic regions could then be tested for the expression of unknown ncRNAs. Genomic SELEX has been successfully applied to select mRNA binding partners of proteins [e.g. (93,94)], but to the best of our knowledge, studies that focused on ncRNAs have not been published for any organism, yet.

Currently, the Schroeder laboratory has taken this approach to identify new Hfq-binding RNAs from *E. coli* (C. Lorenz and R. Schroeder, personal communication). A representative library of the *E. coli* genome was constructed from random 50–500 bp genomic DNA fragments to which defined linkers, one of these containing a T7 RNA polymerase promoter, were attached in the course of the initial library generation step (92). These fragments were *in vitro* transcribed with T7 RNA polymerase, incubated with Hfq and selected for Hfq binding on filters. Taking the standard SELEX route (95), the retained RNA was converted to cDNA and subjected to additional (eight) re-amplification and selection rounds, which finally resulted in a pool of RNAs that bound Hfq with K_d values of 5–50 nM. Subsequently, specific Hfq interaction of the thus enriched RNAs was determined *in vivo* using a yeast three-hybrid screen (96). Preliminary results suggest that these experiments identified a number of novel Hfq-binding RNAs, including antisense RNAs and candidate ncRNAs from intergenic regions.

Obstacles and advantages of the method

Genomic SELEX would clearly have its strength in finding ncRNAs that are overlooked by methods that require an ncRNA gene to be expressed at a certain level. With their small genome sizes, prokaryotes should be particularly amenable to this type of approach. Since in bacteria functional ncRNAs are mostly encoded by intergenic regions, the original pool of DNA fragments could be loaded by specifically amplifying this portion of the genome, which in bacteria typically constitute <10% of the entire genome. As a further advantage of genomic SELEX, the tight association of an ncRNA with a given protein that is a prerequisite for its successful selection could also point to a biological role of this ncRNA, e.g. its function as an antagonist or cofactor of the protein's activity.

At present, very few general RNA-binding proteins are known that specifically form complexes with ncRNAs. Two of the proteins discussed above, Hfq and La (Lhp1p), also associate with mRNAs [references in (78,87)]. Thus, similar to cDNA cloning and microarray analysis, a genomic SELEX approach with such general RNA-binding proteins is expected to yield many additional RNA candidates that one would not readily consider as ncRNAs. What is more, this method only indicates that a certain genomic locus could have a function when transcribed into RNA. However, the exact condition

under which such an RNA is expressed—if it is at all—will still have to be determined.

A major advantage of the genomic SELEX method, compared with the cDNA cloning strategy (see above), is, however, that the latter requires isolation of ncRNAs from an organism or cell under all possible developmental and growth conditions, which might not be always feasible. In contrast, genomic SELEX generates RNA species from all regions of a genome and thus is not dependent on isolating RNAs from all these different states.

Functional RNomics approaches: techniques following RNA identification

Identification of ncRNAs can be only regarded as a first step towards the elucidation of their functions. The term 'candidate' should be used as a suffix to the ncRNA, as long as the function of an ncRNA has not been elucidated. Only then, the RNA species should be designated as a *bona fide* ncRNA.

To obtain hints towards the function of an ncRNA candidate, several approaches can be performed:

- (i) Since most functional ncRNAs are part of a ribonucleoprotein particle (RNP), the protein components of ncRNAs can be searched for. This is achieved, for example, by using the RNA as a 'bait' to fish for these RNA binding proteins in cell extracts. RNAs can be synthesized with an 'affinity-tag' such as biotin by T7 RNA polymerase *in vitro* transcription in the presence of biotin-UTP. The biotinylated RNAs are then coupled to a streptavidin column. Alternatively, an RNA sequence binding to a known protein can be cloned 5'- or 3'- to the ncRNA gene. By attaching the known RNA binding protein to a solid support, the ncRNP can be isolated by using the known RNA tag as a bait (97). Elucidation of the protein-components of an RNP can hint towards its functions, since the proteins might exhibit domains with known catalytic activity. For *in vivo* analysis, the yeast two-hybrid system has been expanded to a three-hybrid system, where the ncRNA is used as a bait *in vivo* to fish for proteins which bind to it (98).
- (ii) Many of the ncRNAs hitherto found exhibit specific RNA targets, which they recognize by an antisense mechanism, e.g. Watson–Crick base pairing (99). Target RNAs include mRNAs or other ncRNAs such as ribosomal RNAs, snRNAs or tRNAs. For elucidation of ncRNA targets, either bioinformatical or experimental methods can be employed. For bioinformatical methods, search for complementarity can be performed. This was successfully achieved, for example, in the case of miRNAs targets (100,101). Experimental methods could include that by fishing the ncRNA of interest through an RNA-binding protein (see above), the target RNA, complementary to the ncRNA, could be co-isolated as well. This might require a cross-link prior to isolation of the RNA heteroduplex, depending on the stability of the RNA–RNA interaction. Alternatively, by expression/overexpression of an ncRNA of interest and subsequent microarray analysis, potential mRNA targets can be identified, if the ncRNAs influences the abundance of its respective mRNA target(s) in the cell (102).

- (iii) Analysis of expression patterns from an ncRNA of interest: for example, the cellular/subcellular localization of the RNA/RNP particle might shed some additional light on its function, e.g. localization in the nucleolus, nucleus or cytoplasm might hint towards an ncRNA involvement in functions exerted in these cellular compartments. To this end, fluorescent *in situ* hybridization techniques can be used to localize the RNA of interest (103). In addition to sub-cellular localization of ncRNAs, the tissue-specific or developmental expression of ncRNAs can be analysed by northern blotting, using total RNA from different tissues or developmental states. Thus, if an ncRNA is only expressed in the brain, at a certain developmental stage, for example, the function of the ncRNA can be searched for within this temporal and spatial expression window of the respective organism.
- (iv) Ultimately, to address the function of ncRNAs, their genes have to be eliminated in the genomes of respective organisms. In other cases, overexpression of ncRNA genes has been useful to obtain a more prominent phenotype [see discussion of multicopy plasmid approaches in (36)].

For certain model bacteria such as *E.coli*, gene deletions are usually accomplished in a few days (104,105). For most other organisms, only the conventional time-consuming knock-out technology is available for this purpose. Very recently, the more elegant knock-down strategies by RNA interference, so far applied only to protein-coding mRNAs, have been shown - in some cases - to be also suitable for rapid ncRNA depletion (106,107); however, the mechanism by which RNAi targets ncRNAs is completely unknown. In addition, a very elegant study has also very recently demonstrated the potential of chemically modified antisense miRNAs (so called 'antagomirs') for the knock-down of certain miRNA species (108).

CONCLUSION

The methods presented above offer a rich tool-box to search for and identify ncRNAs at both large and small scale in virtually any genome. In humans, thousands of apparently non-coding RNA transcripts were observed by microarray analyses and proposed to be involved in regulating gene expression (109). That some of these transcripts could also be verified by northern blot analysis or quantitative PCR makes it difficult to write them all off as experimental artefacts, mRNA degradation fragments or other transcriptional 'noise' (30). Furthermore, several of the ncRNAs that were discovered in genome-wide searches in *E.coli* only a couple of years ago (17,18) have meanwhile been assigned regulatory functions (110–116).

Nonetheless, without a clue as to their biological functions, newly identified ncRNAs molecules should rather be considered as 'candidates for ncRNAs' (see above). So, the burning question is: what are the functions of all of these RNA transcripts? Or, if they are not functional, why does the cell devote its resources to producing them? Thus, next to novel methods to identify them in model organisms, also novel methods—preferentially as high-throughput approaches as used in (106,108)—are needed to tackle the biological roles of ncRNAs.

ACKNOWLEDGEMENTS

The authors thank R. Gupta for helpful suggestions and critical reading of the manuscript. A.H. was supported by an Austrian FWF grant FWF-171370 and a German DFG grant 467/5-1. Support to J.V. was given by the Max Planck Society and by a DFG grant VO 875/1-1. The Open Access publication charges for this article were waived by Oxford University Press.

Conflict of interest statement. None declared.

REFERENCES

- Mattick,J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.
- Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nature Rev. Genet.*, **2**, 919–929.
- Huttenhofer,A., Brosius,J. and Bachelierie,J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.
- Huttenhofer,A., Schattner,P. and Polacek,N. (2005) Non-coding RNAs: hope or hype? *Trends Genet.*, **21**, 289–297.
- Storz,G. (2002) An expanding universe of noncoding RNAs. *Science*, **296**, 1260–1263.
- Washietl,S., Hofacker,I.L., Lukasser,M., Huttenhofer,A. and Stadler,P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.
- Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.
- Zhang,Y., Zhang,Z., Ling,L., Shi,B. and Chen,R. (2004) Conservation analysis of small RNA genes in *Escherichia coli*. *Bioinformatics*, **20**, 599–603.
- Hershberg,R., Altuvia,S. and Margalit,H. (2003) A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res.*, **31**, 1813–1820.
- Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Tuschl,T. (2003) Functional genomics: RNA sets the standard. *Nature*, **421**, 220–221.
- Ambros,V. (2001) microRNAs: tiny regulators with great potential. *Cell*, **107**, 823–826.
- Volpe,T.A., Kidner,C., Hall,I.M., Teng,G., Grewal,S.I. and Martienssen,R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
- Avner,P. and Heard,E. (2001) X-chromosome inactivation: counting, choice and initiation. *Nature Rev. Genet.*, **2**, 59–67.
- Plath,K., Mlynarczyk-Evans,S., Nusinow,D.A. and Panning,B. (2002) Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.*, **36**, 233–278.
- Rougeulle,C. and Heard,E. (2002) Antisense RNA in imprinting: spreading silence through Air. *Trends Genet.*, **18**, 434–437.
- Argaman,L., Hershberg,R., Vogel,J., Bejerano,G., Wagner,E.G., Margalit,H. and Altuvia,S. (2001) Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.*, **11**, 941–950.
- Wassarman,K.M., Repoila,F., Rosenow,C., Storz,G. and Gottesman,S. (2001) Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.*, **15**, 1637–1651.
- Huttenhofer,A., Kiefmann,M., Meier-Ewert,S., O'Brien,J., Lehrach,H., Bachelierie,J.P. and Brosius,J. (2001) RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, **20**, 2943–2953.
- Marker,C., Zemann,A., Terhorst,T., Kiefmann,M., Kastenmayer,J.P., Green,P., Bachelierie,J.P., Brosius,J. and Huttenhofer,A. (2002) Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant *Arabidopsis thaliana*. *Curr. Biol.*, **12**, 2002–2013.
- Tang,T.H., Bachelierie,J.P., Rozhdstvensky,T., Bortolin,M.L., Huber,H., Drungowski,M., Elge,T., Brosius,J. and Huttenhofer,A.

- (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl Acad. Sci. USA*, **99**, 7536–7541.
22. Tang, T.-H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachelier, J.P. and Huttenhofer, A. (2004) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.*, in press.
 23. Yuan, G., Klambt, C., Bachelier, J.P., Brosius, J. and Huttenhofer, A. (2003) RNomics in *Drosophila melanogaster*: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.*, **31**, 2495–2507.
 24. Klein, R.J., Misulovin, Z. and Eddy, S.R. (2002) Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl Acad. Sci. USA*, **99**, 7542–7547.
 25. Rivas, E., Klein, R.J., Jones, T.A. and Eddy, S.R. (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr. Biol.*, **11**, 1369–1373.
 26. Vitali, P., Royo, H., Seitz, H., Bachelier, J.P., Huttenhofer, A. and Cavaille, J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, **31**, 6543–6551.
 27. Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 563–573.
 28. Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K. *et al.* (2004) Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genet.*, **36**, 40–45.
 29. Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J. *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, **116**, 499–509.
 30. Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G. *et al.* (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, **14**, 331–342.
 31. Storz, G., Opdyke, J.A. and Zhang, A. (2004) Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol.*, **7**, 140–144.
 32. Mizuno, T., Chou, M.Y. and Inouye, M. (1983) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the Japan Academy Series B-Physical and Biological Sciences*, **59**, 335–338.
 33. Sledjeski, D. and Gottesman, S. (1995) A small RNA acts as an antisilencer of the H-NS-silenced *rcaA* gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **92**, 2003–2007.
 34. Majdalani, N., Chen, S., Murrow, J., St John, K. and Gottesman, S. (2001) Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol. Microbiol.*, **39**, 1382–1394.
 35. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
 36. Vogel, J. and Sharma, C.S. (2005) How to find small non-coding RNAs in bacteria. *Biol. Chem.*, **386**, 1219–1238.
 37. Eddy, S.R. (2002) Computational genomics of noncoding RNA genes. *Cell*, **109**, 137–140.
 38. Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning 3rd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.
 39. Bruce, A.G. and Uhlenbeck, O.C. (1978) Reactions at the termini of tRNA with T4 RNA ligase. *Nucleic Acids Res.*, **5**, 3665–3677.
 40. Ikemura, T. and Dahlberg, J.E. (1973) Small ribonucleic acids of *Escherichia coli*. II. Noncoordinate accumulation during stringent control. *J. Biol. Chem.*, **248**, 5033–5041.
 41. Ikemura, T. and Dahlberg, J.E. (1973) Small ribonucleic acids of *Escherichia coli*. I. Characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *J. Biol. Chem.*, **248**, 5024–5032.
 42. Griffin, B.E. (1971) Separation of ³²P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5 S' and 5 S from *E. coli*. *FEBS Lett.*, **15**, 165–168.
 43. Hindley, J. (1967) Fractionation of ³²P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *J. Mol. Biol.*, **30**, 125–136.
 44. Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965) A two-dimensional fractionation procedure for radioactive nucleotides. *J. Mol. Biol.*, **13**, 373–398.
 45. Brownlee, G.G. and Sanger, F. (1969) Chromatography of ³²P-labelled oligonucleotides on thin layers of DEAE-cellulose. *Eur. J. Biochem.*, **11**, 395–399.
 46. Silberklang, M., Gillum, A.M. and RajBhandary, U.L. (1979) Use of *in vitro* ³²P labeling in the sequence analysis of nonradioactive tRNAs. *Methods Enzymol.*, **59**, 58–109.
 47. Branch, A.D., Benenfeld, B.J. and Robertson, H.D. (1989) RNA fingerprinting. *Methods Enzymol.*, **180**, 130–154.
 48. Donis-Keller, H., Maxam, A.M. and Gilbert, W. (1977) Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.*, **4**, 2527–2538.
 49. Gupta, R.C. and Randerath, K. (1977) Use of specific endonuclease cleavage in RNA sequencing. *Nucleic Acids Res.*, **4**, 1957–1978.
 50. Gupta, R.C. and Randerath, K. (1977) Use of specific endonuclease cleavage in RNA sequencing—an enzymic method for distinguishing between cytidine and uridine residues. *Nucleic Acids Res.*, **4**, 3441–3454.
 51. Peattie, D.A. (1979) Direct chemical method for sequencing RNA. *Proc. Natl Acad. Sci. USA*, **76**, 1760–1764.
 52. Yarus, M. and Barrell, B.G. (1971) The sequence of nucleotides in tRNA Ile from *E. coli* B. *Biochem. Biophys. Res. Commun.*, **43**, 729–734.
 53. Brownlee, G.G., Cartwright, E., McShane, T. and Williamson, R. (1972) The nucleotide sequence of somatic 5 S RNA from *Xenopus laevis*. *FEBS Lett.*, **25**, 8–12.
 54. Ehresmann, C., Stiegler, P., Carbon, P. and Ebel, J.P. (1977) Recent progress in the determination of the primary sequence of the 16 S RNA of *Escherichia coli*. *FEBS Lett.*, **84**, 337–341.
 55. Balakin, A.G., Smith, L. and Fournier, M.J. (1996) The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell*, **86**, 823–834.
 56. Pichon, C. and Felden, B. (2005) Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc. Natl Acad. Sci. USA*, **102**, 14249–14254.
 57. Trotochaud, A.E. and Wassarman, K.M. (2005) A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Struct. Mol. Biol.*, **12**, 313–319.
 58. Gerhold, D. and Caskey, C.T. (1996) It's the genes! EST access to human genome content. *Bioessays*, **18**, 973–981.
 59. Ohlrogge, J. and Benning, C. (2000) Unraveling plant metabolism by EST analysis. *Curr. Opin. Plant Biol.*, **3**, 224–228.
 60. Martin, G. and Keller, W. (1998) Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA*, **4**, 226–230.
 61. Huttenhofer, A., Cavaille, J. and Bachelier, J.P. (2004) Experimental RNomics: a global approach to identifying small nuclear RNAs and their targets in different model organisms. *Methods Mol. Biol.*, **265**, 409–428.
 62. Aspegren, A., Hinas, A., Larsson, P., Larsson, A. and Soderbom, F. (2004) Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res.*, **32**, 4646–4656.
 63. Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jager, J.G., Huttenhofer, A. and Wagner, E.G. (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res.*, **31**, 6435–6443.
 64. Kawano, M., Reynolds, A.A., Miranda-Rios, J. and Storz, G. (2005) Detection of 5'- and 3'-UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.*, **33**, 1040–1050.
 65. Willkomm, D.K., Minnerup, J., Huttenhofer, A. and Hartmann, R.K. (2005) Experimental RNomics in *Aquifex aeolicus*: identification of small non-coding RNAs and the putative 6S RNA homolog. *Nucleic Acids Res.*, **33**, 1949–1960.
 66. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.
 67. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Identification of tissue-specific microRNAs from mouse. *Curr. Biol.*, **12**, 735–739.
 68. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, **294**, 862–864.

69. Chen, P.Y., Manning, 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.
70. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
71. Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C. and Green, P.J. (2005) Elucidation of the small RNA component of the transcriptome. *Science*, **309**, 1567–1569.
72. Pfeffer, S., Lagos-Quintana, M. and Tuschl, T. (2003) Cloning of small RNA molecules. *Current Protocols in Molecular Biology*, 26.24.21–26.24.18.
73. Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E. and Kiss, T. (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O- methylation and pseudouridylation guide RNAs. *EMBO J.*, **21**, 2746–2756.
74. Stoughton, R.B. (2005) Applications of DNA microarrays in biology. *Annu. Rev. Biochem.*, **74**, 53–82.
75. Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J. and Church, G.M. (2000) RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.*, **18**, 1262–1268.
76. Tjaden, B., Saxena, R.M., Stolyar, S., Haynor, D.R., Kolker, E. and Rosenow, C. (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.*, **30**, 3732–3738.
77. Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G. and Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.*, **50**, 1111–1124.
78. Valentin-Hansen, P., Eriksen, M. and Udesen, C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.*, **51**, 1525–1533.
79. Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA*, **101**, 9740–9744.
80. Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y. *et al.* (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.*, **14**, 2486–2494.
81. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nature Genet.*, **37**, 766–770.
82. Peng, W.T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W. *et al.* (2003) A panoramic view of yeast noncoding RNA processing. *Cell*, **113**, 919–933.
83. Hiley, S.L., Babak, T. and Hughes, T.R. (2005) Global analysis of yeast RNA processing identifies new targets of RNase III and uncovers a link between tRNA 5' end processing and tRNA splicing. *Nucleic Acids Res.*, **33**, 3048–3056.
84. Hiley, S.L., Jackman, J., Babak, T., Trocheset, M., Morris, Q.D., Phizicky, E. and Hughes, T.R. (2005) Detection and discovery of RNA modifications using microarrays. *Nucleic Acids Res.*, **33**, e2.
85. Xing, F., Hiley, S.L., Hughes, T.R. and Phizicky, E.M. (2004) The specificities of four yeast dihydrouridine synthases for cytoplasmic tRNAs. *J. Biol. Chem.*, **279**, 17850–17860.
86. Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.
87. Inada, M. and Guthrie, C. (2004) Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **101**, 434–439.
88. Babak, T., Blencowe, B.J. and Hughes, T.R. (2005) A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription. *BMC Genomics*, **6**, 104.
89. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
90. Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S. *et al.* (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science*, **306**, 2242–2246.
91. Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. and Gingeras, T.R. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res.*, **15**, 987–997.
92. Singer, B.S., Shtatland, T., Brown, D. and Gold, L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.*, **25**, 781–786.
93. Shtatland, T., Gill, S.C., Javornik, B.E., Johansson, H.E., Singer, B.S., Uhlenbeck, O.C., Zichi, D.A. and Gold, L. (2000) Interactions of *Escherichia coli* RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acids Res.*, **28**, E93.
94. Kim, S., Shi, H., Lee, D.K. and Lis, J.T. (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res.*, **31**, 1955–1961.
95. Gold, L., Polisky, B., Uhlenbeck, O. and Yarus, M. (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.*, **64**, 763–797.
96. Bernstein, D.S., Buter, N., Stumpf, C. and Wickens, M. (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods*, **26**, 123–141.
97. Bardwell, V.J. and Wickens, M. (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res.*, **18**, 6587–6594.
98. Zhang, B., Kraemer, B., SenGupta, D., Fields, S. and Wickens, M. (1999) Yeast three-hybrid system to detect and analyze interactions between RNA and protein. *Methods Enzymol.*, **306**, 93–113.
99. Wagner, E.G., Altuvia, S. and Romby, P. (2002) Antisense RNAs in bacteria and their genetic elements. *Adv. Genet.*, **46**, 361–398.
100. 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.
101. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. *et al.* (2005) Combinatorial microRNA target predictions. *Nature Genet.*, **37**, 495–500.
102. Lim, L.P., Lau, N.C., Garrett-Engle, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. and Johnson, J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
103. Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavaille, J. and Huttenhofer, A. (2005) ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.*, **169**, 745–753.
104. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA*, **97**, 6640–6645.
105. Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **97**, 5978–5983.
106. Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B. and Schultz, P.G. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*, **309**, 1570–1573.
107. Nakamoto, M., Jin, P., O'Donnell, W.T. and Warren, S.T. (2005) Physiological identification of human transcripts translationally regulated by a specific microRNA. *Hum. Mol. Genet.*, **14**, 3813–3821.
108. Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M. and Stoffel, M. (2005) Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature*, **438**, 685–689.
109. Mattick, J.S. (2005) The functional genomics of noncoding RNA. *Science*, **309**, 1527–1528.
110. Vogel, J., Argaman, L., Wagner, E.G. and Altuvia, S. (2004) The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.*, **14**, 2271–2276.

111. Udekwi, K.I., Darfeuille, F., Vogel, J., Reimegard, J., Holmqvist, E. and Wagner, E.G. (2005) Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. *Genes Dev.*, **19**, 2355–2366.
112. Rasmussen, A.A., Eriksen, M., Gilany, K., Udesen, C., Franch, T., Petersen, C. and Valentin-Hansen, P. (2005) Regulation of ompA mRNA stability: the role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.*, **58**, 1421–1429.
113. Opydyke, J.A., Kang, J.G. and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J. Bacteriol.*, **186**, 6698–6705.
114. Chen, S., Zhang, A., Blyn, L.B. and Storz, G. (2004) MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J. Bacteriol.*, **186**, 6689–6697.
115. Massé, E. and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **99**, 4620–4625.
116. Vanderpool, C.K. and Gottesman, S. (2004) Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, **54**, 1076–1089.