With the advent of new DNA sequencing and array-based technologies we will be able to uncover aspects of the genome and the transcriptome previously unimagined. One of these aspects is the identification and characterization of non-coding RNAs. These new players are becoming a common feature of eukaryotic transcriptomes. They are associated with various diseases opening new possibilities in drug design and biomarker identification by the pharmaceutical and biotechnology industry.

Non-coding RNAs and new opportunities for the private sector

Fabrício F. Costa

Cancer Biology and Epigenomics Program, Children's Memorial Research Center and Northwestern University's Feinberg School of Medicine, 2300 Children's Plaza, Box 220, Chicago, IL, 60614, USA

Dr. F.F. Costa
Dr. F.F. Costa has obtained his PhD in cancer genetics in the year 2004 from the Ludwig Institute for Cancer Research in Sao Paulo, Brazil. He has completed his training for two years at the Massachusetts General Hospital, affiliated with Harvard University in Boston, USA. Dr. Costa has a special interest in the emerging field of non-coding RNAs and epigenomics. He has worked as a consultant for newspapers in the biotechnology sector and has also given interviews to the scientific journal "Nature Biotechnology" and the British magazine "The Economist" about his work. He has recently founded a web-based company named Genomic Enterprise with the objective of bridging the fields of genetics, genomics and epigenomics.

Non-coding RNAs (ncRNAs) have been recently implicated in several molecular mechanisms in eukaryotes. They are a group of transcripts with no protein-coding potential that may have multiple functions and in many cases they have been associated with diseases. Some companies have already started to launch platforms such as arrays and products on the basis of new DNA sequencing technologies aimed at identifying and studying different types of ncRNAs but this represents just a small step toward the understanding of this new area of research. The private sector should start paying more attention to ncRNAs in order to improve the pipeline for drug discovery, drug development and facilitate the identification of new diagnostic and prognostic markers.

Introduction

One of the main objectives of scientific research is to identify and develop new drugs and/or therapies that could ameliorate the impact of diseases in individuals as well as in society in general. Most efforts in drug development and discovery in the private sector have been mainly focused on targeting molecules such as proteins. The drugs targeting proteins include various monoclonal antibodies that can block receptors on cells, vaccines against proteins from pathogens and antagonist proteins. Drugs produced by the pharmaceutical industry also include several small molecules that can be natural or chemically designed targeting defective proteins in diseases such as cancer. Non-coding RNAs (ncRNAs) such as ribosomal RNAs, which will not be the focus of this article, serve for example as targets for different classes of antibiotics and have been used for decades with success in various diseases. One of these drugs – aminoglycoside antibiotics – can target human ribosomal RNAs and have been used as therapeutics for different genetic disorders. Other types of ncRNAs, however, such as the new classes that have been described recently might also be good drug targets as exemplified by several reports implicating them in a range of different pathologies. ncRNAs are transcriptional units lacking protein-coding potential produced by eukaryotic cells. There is still skepticism on whether the majority of the ncRNAs produced by eukaryotic cells have specific functions but growing evidence suggests that this might be the case. For example, in-depth transcriptome analyses using second generation DNA sequencers combined with other technologies have shown that our genomes work as an

E-mail address: fcosta@childrensmemorial.org.

1359-6446/06/$ - see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.drudis.2009.01.008

BOX 1

MicroRNAs and companies developing drugs and/or therapies on the basis of their mode of action.

MicroRNAs (miRNAs) are a large group of ncRNAs that act by blocking mRNA translation. It is estimated that the human genome has thousands of miRNA genes, but just ~800 were described so far (http://mirnamap.mbc.nctu.edu.tw/). This class of noncoding genes is able to regulate at least 30% of all human protein-coding genes by targeting their 3′-UTR sequences [46]. There is also evidence that miRNAs can regulate the expression of other types of ncRNAs (such as long ncRNAs), indicating that these small genes might have a big impact in transcriptome networks [39,45]. Several studies have already shown that miRNAs are deregulated in diseases [47]. The private sector has increased their interest in targeting these regulators of gene expression as a means of countering diseases. Companies that were focusing their research and development on different types of RNA molecules are now expanding their interest in miRNAs. By measuring the activity of genes encoding miRNAs, signatures that enable molecular classification of diseases can be determined. Companies such as Rosetta Genomics, Exiqon, Crogen Pharmaceuticals and Asuragen currently focus on using miRNAs as cancer biomarkers. Regulus Therapeutics and Santaris Pharma work on the development of miRNA inhibitors as a novel class of RNA-based drugs, which have shown promising therapeutic effects in preclinical models. Another company, Miragen Therapeutics, has special interest in developing miRNA-based drugs to cardiovascular and muscular diseases.

RNA machine and that the great majority of these RNAs do not code for proteins. In fact, it is estimated that ~90% of the human genome is transcribed into RNAs with unknown functions [1].

ncRNAs can be divided into groups depending on their size, ranging from 25 to 30-nt long for microRNAs (miRNAs), 30–300-nt long for small RNAs and from sizes ranging from 300 nt to several kilobases for long ncRNAs (for review see [2,3]). miRNAs are a group of ncRNAs that have been extensively studied by different groups and there are companies specifically using them as a new therapeutic strategy and as diagnostic or prognostic markers (see Box 1 for more details). Other groups, however, such as small and long ncRNAs, have been neglected so far. This is probably because small and long ncRNAs have only recently emerged as potential players in different aspects of eukaryotic cells. It is speculated that the majority of these RNAs play an important role in epigenetic mechanisms and gene regulation in eukaryotic cells (for review see [4]); however, the exact function of a great proportion of these transcripts is not yet understood. Several recent publications reported an association of their deregulated expression with diseases (some examples are shown in Table 1 and further discussed in this article).

The new discoveries showing that the human genome works as an RNA machine with the majority of the DNA being transcribed into RNAs with unknown functions [5] opens a new area of research and new possibilities for drug discovery, improvement of therapeutics and identification of new diagnostic and prognostic markers. The private sector should start paying more attention to these new discoveries, especially the identification of small and long ncRNAs that could be used to develop new drugs. Several companies have already started changing their focus developing new platforms such as arrays and products based on new DNA sequencing technologies that could be used to study known long ncRNAs and also to identify new examples with important functions (see Box 2 and Table 2). This article will discuss the potential use of small and long ncRNAs in drug development, drug discovery and for the identification of better diagnostic and prognostic molecular markers in diseases.

The rise of non-coding RNAs

The new classes of small and long ncRNAs were first described and identified by chance in the early 1990s [6]. The increase in the number of ncRNA examples coincided with the use of high-throughput sequencing technologies and functional screening attempting to clone new genes associated with different aspects of eukaryotic cells and disease states. Some databases have already been launched to collect and store all new examples of small and long ncRNAs (see Table 3). With the advent of new DNA sequencing technologies for in-depth transcriptome analyses (see Box 2 for more details) it is becoming clear that thousands of protein-coding genes have antisense transcripts that, in some cases, lead to

TABLE 1

Examples of long ncRNAs that are deregulated in diseases

ncRNA	Function	Disease	Refs
MIAT	Polymorphism in the gene increases the risk of the disease. Its function is still unknown	Cardiovascular disease	[32]
FMR4	Antiapoptotic function	X-fragile syndrome	[31]
SCA8	Downregulates KLHL1 expression through an antisense mechanism which can lead to SCA8 neuropathogenesis	Spinocerebellar ataxia	[33]
BACE1-AS	Regulation of the sense transcript (BACE1) by a feed-forward mechanism is implicated in driving Alzheimer's disease pathology	Alzheimer's disease	[30]
aHIF	Regulation of the sense transcript HIF1alfa	Renal and breast cancers	[26,27]
DD3	Large ncRNA overexpressed in prostate cancers. Its function is still unknown	Prostate cancer	[19]
MALAT-1	mRNA metabolism and splicing?	Cancer	[21,22]
H19	H19 gene behaves as an oncogene and may serve as a potential new target for antitumor therapy	Cancer	[23]
Y RNAs	Essential factors for chromosomal DNA replication and cell proliferation	Cancer	[25]
p15AS	p15 tumor suppressor gene silencing by antisense and epigenetic mechanisms	Cancer?	[28]
NCTs	Group of newly identified ncRNAs with changes in expression and mutations in tumors. Their function is still unknown	Cancer	[29]

BOX 2

High-throughput technologies and small/long ncRNA identification.

Several reports have been showing extensive and pervasive transcription of RNAs in a big fraction of the human genome. Estimates reported so far, utilizing technologies such as tilling arrays with a high resolution, indicated that ~90% of eukaryotic genomes can be transcribed [48]. These studies have also been showing that our transcriptome is more complex than previously anticipated. It is important to note that the fraction of the genome that is analyzed by these technologies do not take into account repetitive elements and some groups have already reported that repeats might also be transcribed by eukaryotic cells, although their function is still unknown. In addition, tiling arrays have shown that the human genome contains roughly comparable numbers of protein-coding and noncoding genes that are bound by common transcription factors and regulated by common environmental signals [49]. To reinforce this concept, the development of second and third generation DNA sequencing technologies started to have a big impact in genomics and transcriptomics. These new DNA sequencers are able to generate a better resolution of the transcriptome (even better than tilling arrays) facilitating the identification of new and promising noncoding transcripts that might be disease-associated (for more details on these new technologies see http://www.genomicenterprise.com). In this regard, two studies using these new DNA sequencing technologies were able to conduct the most comprehensive analysis of the mammalian transcriptome so far generating a vast collection of RNAs transcribed from mice embryonic stem cells [50,51]. These new studies have used the SOLiD™ system from Applied Biosystems (ABI) to identify new small noncoding RNAs [50,51]. ABI is also planning to launch a new technology based on the SOLiD™ system to identify long ncRNAs. The new technology is named SOLiD™ Whole Transcriptome Expression Kit and it is expected to provide researchers a greater insight into biological pathways and molecular mechanisms that regulate cell fate decision, development and disease progression. The whole transcriptome kit by ABI is based on Ambion® technology and is expected to provide researchers with an innovative workflow that greatly reduces the time, cost and experimental variability associated with RNA library preparation. It also applies multiplexing that further simplifies the downstream workflow for RNA analysis. For example, using the multiplexing capability of the kit, up to ten RNA libraries can be sequenced simultaneously, reducing the cost of analysis per sample. In addition, the technology will maintain the strandedness of cDNAs which will allow researchers to discern between overlapping RNAs transcribed from the sense or antisense strand. This feature will be important to identify new antisense ncRNA transcripts. The other technologies for ncRNA identification are mainly based on microarray hybridization or array-based technologies (see also Table 2). For example, a high-throughput study that was recently published has used an array-based normalization of RACE libraries to uncover new transcripts produced by the genome [40]. These new high-throughput technologies will enable a better coverage of the hidden layer of RNAs transcribed in normal cells and increase our molecular understanding of common diseases.

the repression of the sense transcript (discordant regulation) or can enhance the stability of the sense transcript (concordant regulation). The great majority of the antisense transcripts are long ncRNAs. In this regard, natural antisense transcripts have been identified as important players in different mechanisms in

TABLE 2

Companies/institutions developing products for small and long ncRNA gene discovery and analyses

Company/institution	ncRNA class	Platform	Drug Discovery potential
Allen Institute for Brain Science	Long	*In situ* hybridization	Expression analyses of long ncRNAs in different areas of the brain could speed the development of new drugs for diseases of the central nervous system
CuRNA	Long	New technologies being developed based on the therapeutic potential of ncRNAs	Patented technologies that use ncRNAs that are up or downregulated in diseases could improve the identification of new drugs and biomarkers
Invitrogen[a]	Long	Microarray hybridization	Expression analyses of long ncRNAs in different samples might help in the identification of functional examples
Applied Biosystems[a]	Small and long	Whole transcriptome analyses by high-throughput sequencing with the SOLiD™ system	Identification of new small and/or long ncRNAs that might be implicated in diseases and a better coverage of mammalian transcriptomes
LC Sciences	Small	Microarray hybridization	Identification of new small ncRNAs that are deregulated in diseases

[a] These companies have recently merged.

TABLE 3

Public databases of small and long ncRNAs

Database	Website	Refs
Rfam	http://rfam.sanger.ac.uk/	[41]
RNAdb	http://jsm-research.imb.uq.edu.au/rnadb/default.aspx	[42]
fRNAdb	http://www.ncrna.org/frnadb/	[43]
ncRNA Portal	http://www.ncrna.org/	[43]
NONCODE	http://www.noncode.org/	[44]
NRED	http://jsm-research.imb.uq.edu.au/NRED	[52]

eukaryotic genomes [7]. For example, Nakaya *et al*. have reported that antisense, intronic, ncRNAs have tissue-specific expression and are enriched in genes that regulate transcription [8]. There is also growing evidence that ncRNAs could be responsible for organism complexity in evolution as exemplified by studies with miRNAs [9]. Evidence that long ncRNAs could be associated with organism complexity was also proposed but still needs further studies [4]. ncRNAs have also been associated with epigenetic inheritance mechanisms that are still not completely understood [10] indicating that there are several unknown mechanisms involving ncRNAs.

Biology of non-coding RNAs

Long ncRNAs have been implicated in several aspects of eukaryotic biology. There are many examples of long ncRNAs that can give rise to small RNAs and miRNAs after processing indicating that some might function as primary transcripts 'coding' for other classes of ncRNAs [11]. There is also growing evidence that a large fraction of intergenic genomic regions can be transcribed as ncRNAs [12]. Furthermore, the expression of ncRNAs can be regulated by transcriptional factors, epigenetic mechanisms and also by other ncRNAs suggesting that they might be functional.

Signal-induced ncRNAs localized to regulatory regions of transcription units were recently reported [13]. They are able to act in cooperation with selective ligands, modulating the activity of RNA-binding proteins in response to specific signals, providing an ncRNA-protein-based strategy to integrate transcriptional programs [13]. A recent study has identified a new ncRNA in yeast, named Csnk1a1, that can be induced after a transcriptional ripple effect [14]. This study also showed that ncRNAs could be activators of the transcription of neighboring genes [14]. Some of these events could be because of pervasive transcription but the authors suggested that it is very improbable that all ncRNAs are not functional [14]. There are also several examples implicated in important epigenetic mechanisms, especially in epigenetic inheritance, as was already discussed [4]. Some studies have also shown that long polyadenylated ncRNAs might help to position recombination protein complexes at dsDNA break (DSB) hotspots within chromosomes in meiotic recombination indicating that ncRNAs might be implicated in DNA repair mechanisms [15]. ncRNAs might even function as important players in the mechanisms of gene splicing as has recently described [16]. Moreover, a study has shown that transcription of ncRNAs by RNA polymerase II is required for chromatin remodeling in yeast during transcription activation [17]. Thus, there is strong evidence that long ncRNAs are implicated in various molecular mechanisms.

Non-coding RNAs and diseases – more harm than good?

Deregulated expression of long ncRNAs has been associated with a variety of multigenic diseases reinforcing their importance in normal mechanisms (see Fig. 1 and Table 1 for more details). For example, DD3 was cloned almost a decade ago as one of the most prostate cancer-specific gene indicating that it could be used as a promising marker for the early diagnosis of prostate cancer [18,19]. In fact, a group has proved that quantitative assays for DD3 could potentially be a valuable tool for the detection of malignant cells in plasma, urine or other clinical specimens, and it could have important implications for the earlier diagnosis and molecular staging of prostate cancer [20]. It was also suggested that the identification of transcription factors and other molecules that interact with the DD3 promoter region could represent a promising area for investigation. Other example is the MALAT-1 ncRNA gene that was first cloned as a metastasis-associated gene in lung cancers and lately its overexpression has been associated with different types of tumors [21,22]. In the same way, the H19 ncRNA has been associated with human cancers and it is able to enhance tumorigenic potential in different tumor types [23]. An antisense to the H19 ncRNA was also identified and it is overexpressed in breast tumors [24]. Other examples include ncRNAs transcribed from human chromosome Y that are overexpressed in human cancers and were associated with cell proliferation [25].

New mechanisms of gene regulation that might be mediated by long ncRNAs in tumors have also been reported in the past five years. Widespread sense–antisense transcripts have been identified in mammalian cells by different groups and global transcriptome analysis have been showing that approximately 60–70% of transcripts have antisense pairs and that perturbation of antisense RNA can alter the expression of the sense gene. For example, a natural antisense transcript of the hypoxia-inducible factor 1α (aHIF) was identified after cell exposure to hypoxia [26]. aHIF is overexpressed in renal tumors and was also described as a strong prognostic marker in breast cancer [27]. In addition, a recent study has identified an antisense transcript to the tumor suppressor gene p15 that is able to trigger heterochromatin formation and DNA methylation in the genomic region of the p15 [28]. p15 antisense (p15AS) is able to regulate negatively the expression of p15 indicating that mechanisms of gene silencing by antisense ncRNAs could be implicated in the tumorigenesis process [28].

A study has also identified new noncoding transcripts (named NCTs) in normal and tumor samples and some examples had altered expression in cancers [29]. Mutations in several NCTs were identified when sequences of normal samples were compared with a panel of cancer-derived cell lines [29]. The authors have suggested that these NCTs might participate in important mechanisms in normal and cancer cells [29]. Interestingly, one of the NCTs identified by this group is the MALAT-1 gene that was previously associated with lung cancer [21,22].

Long ncRNAs have also been associated with other multigenic diseases such as Alzheimer's [30], X-fragile syndrome [31], cardiovascular disease [32] and spinocerebellar ataxia [33]. A recent study has identified a ncRNA that is able to regulate a protein-coding transcript BACE1 and is associated with Alzheimer's disease pathophysiology [30]. The BACE1 protein-coding gene is responsible for cleaving amyloid precursor protein (APP) and the product

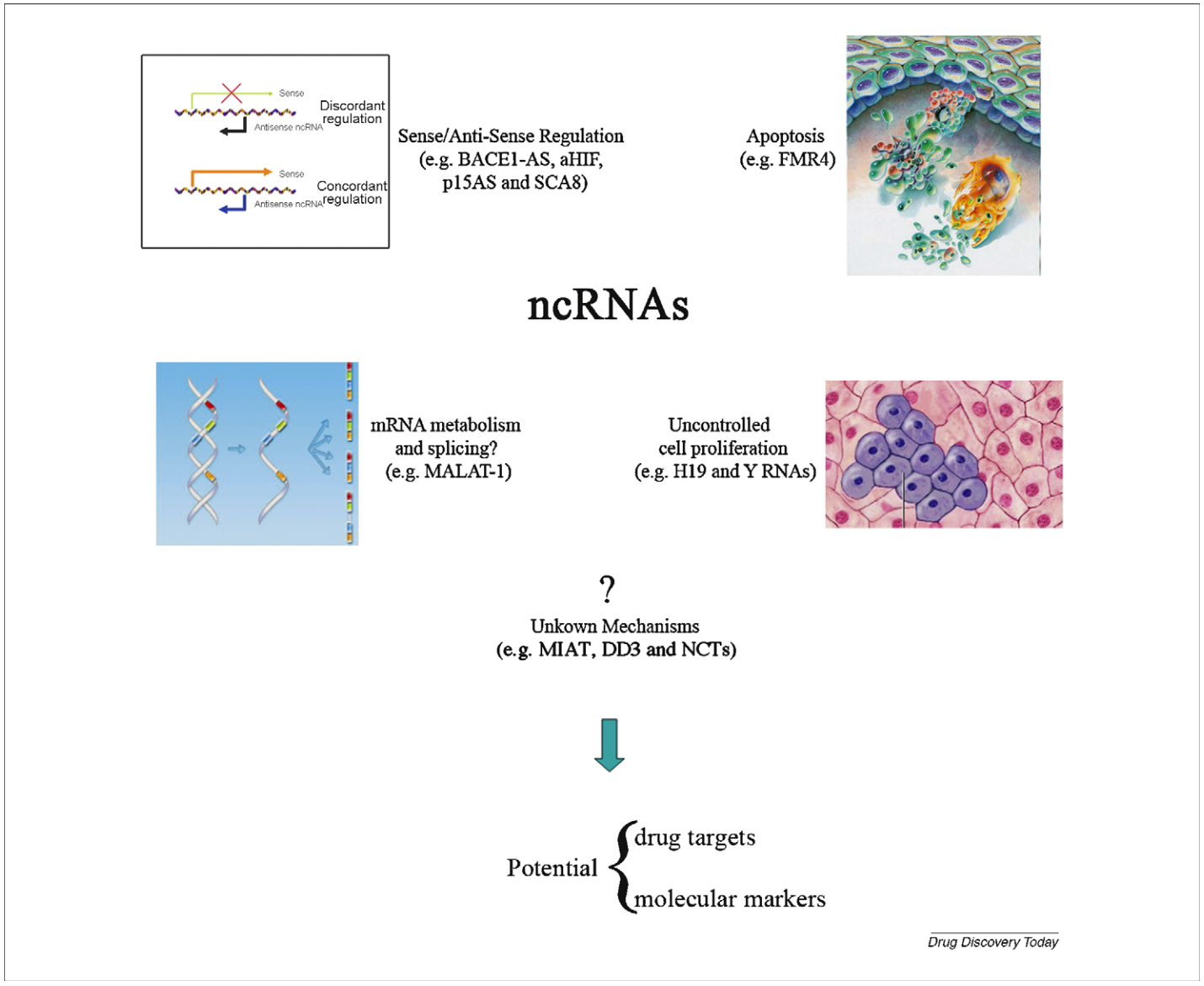

FIGURE 1

ncRNAs as potential drug targets and molecular markers. The figure depicts some of the mechanisms in which small and long ncRNAs have been recently implicated such as sense/antisense regulation, apoptosis, mRNA metabolism and splicing and cell proliferation. There is also a group of ncRNAs with unknown function opening a new area of research.

of this cleavage can influence key aspects of Alzheimer's disease [30]. The identification of an antisense ncRNA that regulates in a concordant manner the BACE1 protein-coding gene opens new possibilities for therapeutic intervention and changes some paradigms in Alzheimer's disease. It shows that the cascade of plaque formation can have an ncRNA that may be a major player and this antisense transcript could be potentially used as a drug target.

Using mainly genomic approaches, a group has recently identified a primate-specific ncRNA transcript, named FMR4, that is implicated in X-fragile syndrome disease [31]. This ncRNA resides upstream of the protein-coding gene FMR1 and both share a bidirectional promoter [31]. The authors verified that the CGG expansion in the 5′-untranslated region (UTR) of FMR1, which is a common characteristic of this disease, appears to affect transcription in both directions and FMR4 is silenced in fragile X patients [31]. The authors also showed that FMR4 ncRNA has antiapoptotic functions and its expression might contribute to several aspects of the clinical presentation of fragile X syndrome and/or related disorders [31]. Another example of antisense regulation by a ncRNA is the SCA8 gene in Spinocerebellar Ataxia 8 [33]. This group has identified that SCA8 ncRNA transcription is colocalized with the protein-coding gene KLHL1 in many brain regions whose functions are correlated to the clinical symptoms of the patients [33]. These findings suggest that SCA8 transcript downregulates KLHL1 expression through an antisense mechanism, which then leads to SCA8 neuropathogenesis [33].

A large-scale analyses using SNP (Single nucleotide polymorphisms) was used recently to identify a susceptibility locus for myocardial infarction and as a result the MIAT gene was cloned [32]. The MIAT gene has five exons and after *in vitro* translation assays the authors showed that MIAT did not encode any protein product, indicating that this is likely to be a functional ncRNA [32]. It

is suggested by this group that MIAT might play an important role in the pathophysiology of cardiovascular diseases [32].

The examples of ncRNAs presented here represent just a small fraction of noncoding transcripts in which deregulation can lead to disease. There is growing evidence that most, if not all, ncRNA genes may have a function and their deregulation might cause diseases. It is not improbable that several new functional examples implicated in diseases will be identified and characterized by new technologies that have been launched (see Table 2 and Box 2 for more details) opening new opportunities in drug development.

Riboswitches and a new avenue

Riboswitches are part of mRNA molecules that can directly bind to a small target and whose binding to the target affects gene activity [34,35]. An mRNA that contains a riboswitch can directly regulate its own activity, depending on the presence or absence of its target molecule. Although the metabolic pathways in which riboswitches are involved have been studied for several years, the first experimental validations of riboswitches have been described only recently [36]. Riboswitches are recognized as a powerful mechanism of genetic control but they were mainly described in prokaryotes. Recently, functional riboswitches have been discovered in eukaryotes, opening a new avenue in gene regulation [16]. It was demonstrated that eukaryotic cells might use metabolite-binding RNAs to regulate splicing events that are important for the control of key biochemical processes [16]. It is tempting to speculate that higher organisms might use the same type of gene regulation and ncRNAs could be key molecules in this process. Some studies have already highlighted how riboswitches, which are RNA sequences that undergo a ligand-induced conformational change to alter gene expression, can be used to reprogram how bacteria respond to small molecules [37]. Since there is growing evidence that eukaryotes express a large number of ncRNAs, there is a possibility that riboswitches could control the expression and action of these RNAs. Riboswitches could also pave new ways for drug design and to develop alternative therapeutic interventions to treat diseases.

New therapeutic targets?

A large fraction of the human genome is still unexplored in drug discovery and research is mainly focused in finding protein targets leading to just a few successful drug candidates so far. Estimates indicate that ~2–3% of our genomes code for proteins. In-depth transcriptome analyses have been showing that the majority of the other 97–98% that do not code for proteins can be transcribed into RNAs. This fact will have profound implications for the understanding of human diseases and also for the design of new therapies.

As discussed above, ncRNAs could be used as molecular markers for the early detection of diseases (e.g. DD3 in prostate cancer) and as prognostic factors (e.g. MALAT-1 in lung cancer). In addition, a group has reported that fragments of ncRNAs (small and long) can be detected in samples from human blood plasma indicating that these molecules could be used as molecular markers in diseases [38]. Furthermore, long ncRNAs that are up or downregulated in diseases could be targeted by oligos or small molecules in an attempt to restore their expression. Unique shapes of various target RNAs can create potential binding sites for small molecules. One example is the long ncRNA BACE1-AS that was identified as upregulated in Alzheimer's disease [30]. In this case, the authors proposed that this long ncRNA could potentially constitute a drug target by systemically administrating siRNA oligos against it or even small molecules to inhibit its expression [30]. In this regard, a new company – CuRNA – was founded with the objective of developing new drugs that will be specific for long ncRNAs (for more details see Table 2). This opens a completely new area of research full of potential for drug discovery and development.

Recent studies have also shown that a large fraction of genomic ultraconserved regions (UCRs) encode a particular subset of long ncRNAs whose expression is altered in human cancers [39]. Moreover, both miRNAs and UCRs are frequently located at fragile sites and genomic regions affected in various cancers [39]. It was also reported that long ncRNAs could be regulated by specific miRNAs making a biological connection between two different groups of ncRNAs [39]. ncRNAs have also been extensively associated with epigenetic regulation in synergy with chromatin modifications such as histone marks and DNA methylation, suggesting that they could be important as new molecular diagnostic markers and molecular targets for epigenetic therapy. Thus, small and long ncRNAs are emerging as new players in various diseases and they could be a gold mine for the identification of new molecular markers and the development of better therapies.

ncRNAs represent attractive therapeutic targets because they might work as riboswitches, control gene expression of several other ''conventional'' genes that give rise to different proteins and their secondary structure can be targeted by an array of strategies such as designing small molecules to block their action.

Conclusions and future prospects

The new discoveries involving ncRNAs will have profound implications for the understanding of human diseases. Because we have a limited number of protein-coding genes in the human genome and there is growing evidence that the vast majority of the genome is transcribed into RNA with a low protein-coding potential there is a completely new area of research emerging for drug development and for molecular marker identification (for more details on public databases for new small and long ncRNAs see Table 3 and Refs. [41–44,52]). The increasing number of reports demonstrating alterations in ncRNA expression patterns in cancer and other diseases discussed here indicate that the private sector may start increasing their efforts in the exploration of this new field. The identification and discovery of new functional ncRNAs in this hidden layer of the human genome will not just help us better understand the cellular circuitries governing eukaryotic cells but will also help us improve disease treatment. Biotechnology and pharmaceutical companies are urged to catch up in this new promising field of research.

Conflict of interest

The author declares that he is the founder and a consultant for Genomic Enterprise.

Acknowledgements

I thank Dr Thomas Wurdinger, Dr Elio Vanin and Dr George S. Mack for critically reading this manuscript. This work is supported by the Maeve MacNicholas Memorial Foundation and Children's Memorial Research Center.

References

1 ENCODE project Consortium. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816
2 Costa, F.F. (2005) Non-coding RNAs: new players in eukaryotic biology. *Gene* 357, 83–94
3 Costa, F.F. (2007) Non-coding RNAs: lost in translation? *Gene* 386, 1–10
4 Costa, F.F. (2008) Non-coding RNAs, epigenetics and complexity. *Gene* 410, 9–17
5 Amaral, P.P. *et al.* (2008) The eukaryotic genome as an RNA machine. *Science* 319, 1787–1789
6 Brannan, C.I. *et al.* (1990) The product of the H19 gene may function as an RNA. *Mol. Cell Biol.* 10, 28–36
7 Røsok, Ø. and Sioud, M. (2004) Systematic identification of sense–antisense transcripts in mammalian cells. *Nat. Biotechnol.* 22, 104–108
8 Nakaya, H.I. *et al.* (2007) Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* 8, R43
9 Heimberg, A.M. *et al.* (2008) MicroRNAs and the advent of vertebrate morphological complexity. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2946–2950
10 Wagner, K.D. *et al.* (2008) RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev. Cell.* 14, 962–969
11 He, L. *et al.* (2005) A microRNA polycistron as a potential human oncogene. *Nature* 435, 828–833
12 Mercer, T.R. *et al.* (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* 105, 716–721
13 Wang, X. *et al.* (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454, 126–130
14 Ebisuya, M. *et al.* (2008) Ripples from neighbouring transcription. *Nat. Cell Biol.* [Epub ahead of print]
15 Wahls, W.P. *et al.* (2008) Meiotic recombination hotspots of fission yeast are directed to loci that express non-coding RNA. *PLoS ONE* 3, e2887
16 Cheah, M.T. *et al.* (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447, 497–500
17 Hirota, K. *et al.* (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* [Epub ahead of print]
18 Bussemakers, M.J. *et al.* (1999) DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res.* 59, 5975–5979
19 de Kok, J.B. *et al.* (2002) DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res.* 62, 2695–2698
20 Hessels, D. *et al.* (2003) DD3 (PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur. Urol.* 44, 8–15
21 Ji, P. *et al.* (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* 22, 8031–8041
22 Lin, R. *et al.* (2007) A large noncoding RNA is a marker for murine hepatocellular carcinomas and a spectrum of human carcinomas. *Oncogene* 26, 851–858
23 Matouk, I.J. *et al.* (2007) The H19 non-coding RNA is essential for human tumor growth. *PLoS ONE* 2, e845
24 Berteaux, N. *et al.* (2008) A novel h19 antisense rna overexpressed in breast cancer contributes to paternal igf2 expression. *Mol. Cell Biol.* [Epub ahead of print]
25 Christov, C.P. *et al.* (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br. J. Cancer* 98, 981–988
26 Thrash-Bingham, C.A. and Tartof, K.D. (1999) aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J. Natl. Cancer Inst.* 91, 143–151
27 Cayre, A. *et al.* (2003) aHIF but not HIF-1alpha transcript is a poor prognostic marker in human breast cancer. *Breast Cancer Res.* 5, R223–R230
28 Yu, W. *et al.* (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* 451, 202–206
29 Perez, D.S. *et al.* (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum. Mol. Genet.* 17, 642–655
30 Faghihi, M.A. *et al.* (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat. Med.* 14, 723–730
31 Khalil, A.M. *et al.* (2008) A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS ONE* 3, e1486
32 Ishii, N. *et al.* (2006) Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet.* 51, 1087–1099
33 Chen, W.L. *et al.* (2008) SCA8 mRNA expression suggests an antisense regulation of KLHL1 and correlates to SCA8 pathology. *Brain Res.* 1233, 176–184
34 Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Opin. Struct. Biol.* 15, 342–348
35 Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8, 776–790
36 Mandal, M. *et al.* (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* 113, 577–586
37 Gallivan, J.P. (2007) Toward reprogramming bacteria with small molecules and RNA. *Curr. Opin. Chem. Biol.* 11, 612–619
38 Semenov, D.V. *et al.* (2008) Fragments of noncoding RNA in plasma of human blood. *Ann. N. Y. Acad. Sci.* 1137, 130–134
39 Calin, G.A. *et al.* (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell.* 12, 215–229
40 Djebali, S. *et al.* (2008) Efficient targeted transcript discovery via array-based normalization of RACE libraries. *Nat. Methods* 5, 629–635
41 Griffiths-Jones, S. *et al.* (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33 (Database issue), D121–D124
42 Pang, K.C. *et al.* (2007) RNAdb 2.0 – an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res.* 35 (Database issue), D178–182
43 Kin, T. *et al.* (2007) fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Res.* 35 (Database issue), D145–D148
44 He, S. *et al.* (2008) NONCODE v2.0: decoding the non-coding. *Nucleic Acids Res.* 36 (Database issue), D170–D172
45 Rossi, S. *et al.* (2008) Cancer-associated genomic regions (CAGRs) and noncoding RNAs: bioinformatics and therapeutic implications. *Mamm. Genome* [Epub ahead of print]
46 Filipowicz, W. *et al.* (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.* 9, 102–114
47 Wurdinger, T. and Costa, F.F. (2007) Molecular therapy in the microRNA era. *Pharmacogenom. J.* 7, 297–304
48 Wilhelm, B.T. *et al.* (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239–1243
49 Cawley, S. *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
50 Cloonan, N. *et al.* (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* 5, 613–619
51 Mortazavi, A. *et al.* (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628
52 Dinger, M.E. *et al.* (2008) NRED: a database of long noncoding RNA expression. *Nucleic Acids Res.* [Epub ahead of print]