

## Review

## The Genetic Signatures of Noncoding RNAs

John S. Mattick\*

Australian Research Council Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Australia

**Abstract:** The majority of the genome in animals and plants is transcribed in a developmentally regulated manner to produce large numbers of non-protein-coding RNAs (ncRNAs), whose incidence increases with developmental complexity. There is growing evidence that these transcripts are functional, particularly in the regulation of epigenetic processes, leading to the suggestion that they compose a hitherto hidden layer of genomic programming in humans and other complex organisms. However, to date, very few have been identified in genetic screens. Here I show that this is explicable by an historic emphasis, both phenotypically and technically, on mutations in protein-coding sequences, and by presumptions about the nature of regulatory mutations. Most variations in regulatory sequences produce relatively subtle phenotypic changes, in contrast to mutations in protein-coding sequences that frequently cause catastrophic component failure. Until recently, most mapping projects have focused on protein-coding sequences, and the limited number of identified regulatory mutations have been interpreted as affecting conventional *cis*-acting promoter and enhancer elements, although these regions are often themselves transcribed. Moreover, ncRNA-directed regulatory circuits underpin most, if not all, complex genetic phenomena in eukaryotes, including RNA interference-related processes such as transcriptional and post-transcriptional gene silencing, position effect variegation, hybrid dysgenesis, chromosome dosage compensation, parental imprinting and allelic exclusion, paramutation, and possibly transvection and transinduction. The next frontier is the identification and functional characterization of the myriad sequence variations that influence quantitative traits, disease susceptibility, and other complex characteristics, which are being shown by genome-wide association studies to lie mostly in non-coding, presumably regulatory, regions. There is every possibility that many of these variations will alter the interactions between regulatory RNAs and their targets, a prospect that should be borne in mind in future functional analyses.

## Introduction

Genome sequencing projects have shown that the numbers of protein-coding genes and the extent of protein-coding sequences do not change appreciably across the vertebrates nor indeed across the metazoa as a whole, despite large differences in developmental complexity [1]. On the other hand, the extent of non-protein-coding intronic and intergenic sequences in genomes does increase with developmental complexity, suggesting that these sequences may contain increasingly elaborate regulatory information [1].

In recent years it has also become evident that the vast majority of the mammalian genome, and that of other complex organisms, is transcribed, apparently in a developmentally regulated manner, to produce large numbers of ncRNAs that are antisense, intergenic, interleaved, or overlapping with protein-coding genes [2–5]. In

addition, there are increasing reports (Figure 1) of the functionality of individual ncRNAs in mammals (Table 1), other animals (see, e.g., [6–8]), plants (e.g., [9,10]), and fungi (e.g., [11]), particularly in relation to developmental processes [12]. These include the involvement of ncRNAs in the regulation of the expression of homeotic genes [7,13], oncogenes [14], and metabolic genes [15], as well as in the regulation of skeletal development [16], eye development [17], epithelial-to-mesenchymal transition [18], and subcellular structures [19–22], among many others (for reviews and additional examples, see [4,12,23–25]).

There is also compelling genome-wide evidence that the large numbers of identified but as yet unstudied intronic, intergenic, and antisense ncRNAs have intrinsic indices of functionality (Table 2), as indicated by the following: (i) the conservation of their promoters, splice junctions, exons, predicted structures, genomic position, and expression patterns [2,26–36]; (ii) their dynamic expression and alternative splicing during differentiation [13,31,32]; (iii) their altered expression or splicing patterns in cancer and other diseases [37–49]; (iv) their association with particular chromatin signatures that are indicative of actively transcribed genes [31,32]; (v) their regulation by key morphogens and transcription factors [31,32,49,50]; and (vi) their tissue- and cell-specific expression patterns and subcellular localization [16,17,19–22,49,51–56].

Indeed, a recent study of over 1,300 mouse ncRNAs showed that almost half of them exhibit precise expression patterns in different parts of the brain, including different subregions of the hippocampus, olfactory bulb, neocortex, and cerebellum [52]. While many ncRNAs, like other regulatory sequences (see below), appear to be evolving quickly, including those with well-validated functions [26], most retain conserved patches within them [26,32], and some may be under positive selection, especially in the brain [57]. At least some of these RNAs, including intronic RNAs, are precursors for small regulatory RNAs like microRNAs (miRNAs), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), and small nucleolar RNAs (snoRNAs) [58–68]. Many miRNAs are highly conserved from nematodes to humans, whereas others are primate-specific, and their full repertoire remains to be determined [4,69,70].

**Citation:** Mattick JS (2009) The Genetic Signatures of Noncoding RNAs. *PLoS Genet* 5(4): e1000459. doi:10.1371/journal.pgen.1000459

**Editor:** Jeannie T. Lee, Massachusetts General Hospital, Howard Hughes Medical Institute, United States of America

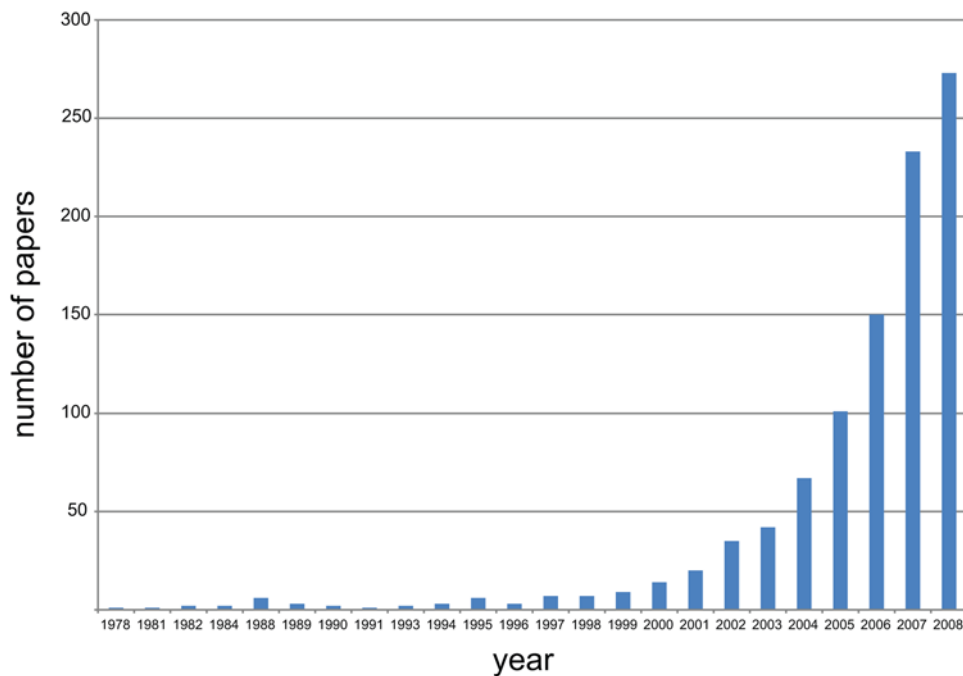
**Published:** April 24, 2009

**Copyright:** © 2009 John S. Mattick. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work was supported by the Australian Research Council (Fellowship Grant FF0561986), the University of Queensland, and the Queensland State Government. These agencies played no role in the preparation, review, or approval of the manuscript.

**Competing Interests:** The author has declared that no competing interests exist.

\* E-mail: j.mattick@imb.uq.edu.au



**Figure 1. The recent rise in papers on ncRNAs.** The number of indexed Medline entries with the words “non-coding RNA”, “noncoding RNA”, “non-protein-coding RNA” or “ncRNA” in the title, abstract or keywords is plotted per year. Data courtesy of Ryan J. Taft. doi:10.1371/journal.pgen.1000459.g001

It is widely accepted that animals have a relatively common set of protein-coding genes and that, notwithstanding lineage-specific innovations and splice variants, the primary basis of phenotypic, especially morphological, radiation and higher complexity has been the variation and expansion of the regulatory architecture that controls the deployment of these protein components and their isoforms during differentiation and development [71]. This regulatory architecture is generally more plastic than protein-coding sequences that are highly constrained by relatively strict structure-function relationships, which is reflected by the fact that regulatory sequences evolve at widely different rates [72–75]. These sequences range from promoter regions that have no recognizable sequence similarity yet direct orthologous patterns of gene expression between fishes and mammals [76] to highly conserved non-genic elements [77,78], and “ultraconserved” sequences that have remained essentially unchanged over hundreds of millions of years of vertebrate evolution and appear to act as tissue-specific enhancers that regulate gene expression during development [79–82].

Regulatory sequences are also generally assumed to operate through their interactions with sequence-specific transcription factors and other regulatory proteins, but this assumption has been made in ignorance, until recently, of the extent of developmentally regulated transcription of ncRNAs from the genome, including many regions spanning enhancers and promoters (see, e.g., [45,83–86]). The possibility is therefore that the genomes of mammals and other complex organisms encode a large repertoire of regulatory RNAs [4]. Indeed, the case has been made that a much higher degree of regulatory sophistication, aided by the co-option of the considerable powers of RNA to transmit sequence-specific information, was a prerequisite for the evolution of developmentally complex organisms [87,88], and that many of these RNAs may be involved in the regulation of developmental processes, including the epigenetic trajectories that underpin them, for which there is increasing evidence [12,89].

However, if these ncRNAs are functional and important in developmental and physiological processes, why have so few been identified in genetic screens to date? Here I outline the emerging evidence for ncRNA involvement in key molecular genetic phenomena and in specific functions and phenotypes. I also outline the expectational, perceptual, and practical factors that may collectively account for the low genetic visibility of individual ncRNAs. Awareness of these factors and the possibility that the structure of the genomic programming of complex organisms is different from our current understanding may lead to the increased recognition of ncRNAs in genetic analyses, assisted by the emerging fusion of genetics, genomics, and systems biology.

### Phenotypic Impact

The ability to detect a relevant mutation or variation is dependent on the sensitivity of the phenotypic screen. Mutations in protein-coding sequences usually give severely compromised (i.e., obvious) phenotypes, whereas those in regulatory sequences often do not. Proteins are the key structural and functional analogue components of cells, and the loss of their function is often disastrous, leading in many cases to obvious defects, and in some cases to embryonic lethality. Mutations in generic transcription factors and other “regulatory” proteins are included, and their loss causes pleiotropic effects on gene expression at many loci and plays an important role in the molecular etiology of cancer [71,90,91]. This is in contrast to regulatory sequences, which, when damaged, may only affect a small part of the network, with more restricted and subtle consequences, often referred to as quantitative trait variations. Indeed the use of the word “mutation”, as opposed to “variation”, reflects an inherent bias in the identification of genetic factors that influence phenotype in animals, with those exhibiting strong effects understandably having taken precedence over those that do not, both perceptually and practically. Consistent with this, until recently, there were

**Table 1.** Examples of functional mammalian noncoding RNAs.

Name	Characteristics	Function	Experimental Methodology <sup>a</sup>	References
Air	108 kb; transcribed from an antisense promoter located in intron 2 of <i>Igf2r</i>	Regulates genomic imprinting of a cluster of autosomal genes on mouse chromosome 17	Mutagenesis; FISH; ChIP; RNA-IP	[156,189,196]
BACE1AS	~2 kb; transcribed antisense to beta-secretase-1 (BACE1) gene, elevated in Alzheimer disease	Regulates BACE1 expression in vitro and in vivo, influences amyloid-beta 1–42 levels	shRNA and siRNA knockdown	[197]
BC1	152 nt; expressed by a specific subset of neurons in the central and peripheral nervous system; dendritic location	Affects exploratory behaviour and anxiety, represses translation by targeting initiation factor 4A helicase	In vivo knockout, biochemical analyses	[100,198]
Borg	~2.8 kb; induced by bone morphogenic proteins (BMPs) and osteogenic proteins	Regulates BMP-induced differentiation of C2C12 cells into osteoblastic cells	Antisense knockdown	[199]
CCND1 associated ncRNAs	Unspecified sizes; transcribed from the promoter region of the cyclin D (CCND1) gene; induced by DNA damage signals	Allosterically modifies the RNA-binding protein TLS (“translocated in liposarcoma”), leading to inhibition of CREB-binding protein and histone acetyltransferase activities to repress cyclin D1	ChIP; RNA-IP; siRNA knockdown	[200]
CUDR	~2.2 kb; up-regulated in a doxorubicin-resistant human squamous carcinoma	Regulates drug sensitivity, cellular transformation and apoptosis	Overexpression	[201]
EGO	~1.0 and 1.7 kb; highly expressed in human bone marrow and in eosinophil development	Regulates expression of myelin basic protein and eosinophil-derived neurotoxin mRNAs	siRNA knockdown	[202]
DHFR upstream	Unspecified size; transcribed upstream of the dihydrofolate reductase (DHFR) gene	Regulates DHFR expression by formation of triple helix in the DHFR promoter	siRNA knockdown; ChIP; RNA-IP, other	[15]
Evf-2	~3.8 kb; antisense to <i>Dlx6</i> ; developmentally regulated, expressed in the ventral forebrain	Cooperates with <i>Dlx-2</i> in vivo to increase the transcriptional activity of the <i>Dlx-5/6</i> enhancer in a target and homeodomain-specific manner	Overexpression; siRNA knockdown; mutagenesis; ChIP	[154]
Gadd7	754 nt; induced by lipotoxic-stress	Regulates lipid-induced oxidative and ER stress	Mutagenesis; shRNA knockdown	[203]
GAS5	~7 kb; growth arrest-specific transcript, multiple splice isoforms, encodes several snoRNAs in its introns, down-regulated in breast cancer	Controls apoptosis and the cell cycle in lymphocytes	siRNA knockdown; overexpression	[204]
H19	2.3 kb; imprinted (maternal allele active) at the <i>Igf2</i> locus; strongly expressed during embryogenesis; up-regulated in various tumours	Complex functions, influences growth by way of a cis control on <i>Igf2</i> expression implicated as both a tumour suppressor and an oncogene	siRNA knockdown; overexpression; in vivo knockout	[205–207]
HOTAIR	2.2 kb; transcribed from the <i>HOXC</i> locus from a position intergenic and antisense to the flanking <i>HOXC11</i> and <i>HOXC12</i> genes	Epigenetically silences gene expression at the <i>HOXD</i> locus	siRNA knockdown	[13]
HOTAIRM1	483 nt; specific to the myeloid lineage	Involved in RA-induced expression of <i>HOXA1</i> and <i>HOXA4</i> during myeloid differentiation, and induction of myeloid differentiation genes <i>CD11b</i> and <i>CD18</i>	shRNA knockdown	[49]
Hsr1	~600 nt; ubiquitously expressed	Required for heat shock response	siRNA and antisense knockdown	[208]
IGS RNAs	150–300 nt; originate from the intergenic spacers (IGS) that separates rRNA genes; bind to the chromatin remodelling complex NoRC	Required for the nucleolar localization of NoRC and epigenetic control of the rDNA locus	Mutagenesis; antisense knockdown	[209]
Kcnq1ot1	91 kb; paternally expressed from <i>Kcnq1</i> imprinting control region	Mediates organization of a lineage-specific nuclear domain involved in epigenetic silencing of the <i>Kcnq1</i> imprinting control region	Mutagenesis; ChIP; RNA/DNA FISH	[22,155, 157,210]
Khps1a	1,290 nt; originates from the CpG island and overlaps a tissue-dependent differentially methylated region of <i>Sphk1</i>	Regulates DNA methylation in the tissue-dependent differentially methylated region of <i>Sphk1</i>	Overexpression	[211]
lincENC1	Size unspecified; ~181 kb from <i>Enc1</i>	Regulates cell proliferation in embryonal stem cells	shRNA knockdown	[32]
MEG3	~1.6 kb; maternally expressed from the <i>Dlk1-Gtl2</i> imprinted locus	Regulates p53 expression, inhibits cell proliferation in the absence of p53	Overexpression, mutagenesis	[192]

**Table 1. cont.**

Name	Characteristics	Function	Experimental Methodology <sup>a</sup>	References
MEN ε/β (Neat1)	~3.5 and ~23 kb; up-regulated upon muscle differentiation; transcribed from the multiple endocrine neoplasia 1 (MEN1) locus	Required for the structural integrity of nuclear paraspeckles	Antisense, siRNA knockdown; overexpression; FISH	[19–21]
ncR-uPAR	~350 bp; upstream of human protease-activated receptor-1 (PAR-1) gene	Up-regulates Par-1 promoter	Overexpression	[212]
Nkx2.2AS	4.3 kb; antisense to Nkx2, preferentially expressed in the nervous system	Enhances oligodendrocytic differentiation	Overexpression	[213]
NRON	~2.7 kb; enriched in placenta, muscle, and lymphoid tissues	Modulates NFAT nuclear trafficking	siRNA knockdown	[214]
p15AS	38.4 kb; antisense to the tumour suppressor gene p15	Epigenetically silences p15 expression	Overexpression	[14]
PCGEM1	1,643 nt; prostate tissue-specific and prostate cancer-associated	Inhibits apoptosis induced by doxorubicin	Overexpression	[215]
PRINS	~3.6 kb; elevated expression in psoriatic epidermis; regulated by the proliferation and differentiation state of keratinocytes.	Required for cell viability after serum starvation	siRNA knockdown	[42]
PINC	~1 and 1.6 kb; developmentally regulated, expressed in mammary gland	Performs dual roles in cell survival and regulation of cell-cycle progression	siRNA knockdown; FISH	[216]
RepA	~1.6 kb; internal to <i>Xist</i>	Recruits the Polycomb complex, PRC2, to the inactive X chromosome, with Ezh2 serving as the RNA binding subunit	ChIP; FISH; overexpression; shRNA knockdown	[128]
SAF	1.5 kb; transcribed from the opposite strand of intron 1 of the human Fas gene	Regulates Fas-mediated but not TNF-alpha-mediated apoptosis	Overexpression	[217]
SatIII	Various sizes up to >1.4 kb; transcribed from satellite DNA associated with, and localized in, nuclear stress bodies	Mediates recruitment of RNA processing factors to, and formation of, nuclear stress bodies	Antisense and siRNA knockdown	[218]
SCA8	Predicted gene product underpinning the triplet repeat expansion-induced neurodegenerative disease Spinocerebellar Ataxia 8	Induces late-onset, progressive neurodegeneration in the <i>Drosophila</i> retina; associates with the RNA binding protein staufen	Ectopic expression in <i>Drosophila</i> ; genetic modifier screen	[39]
TERRA / TelRNAs	Various sizes; transcribed from and associated with telomeres; contain UUAGGG repeats	May form G-quartet structures with telomere DNA; inhibit telomerase activity	RNA FISH; oligo-nucleotide inhibition	[219,220]
Tsix	~40 kb; antisense to <i>Xist</i>	Epigenetically silences <i>Xist</i> expression by inhibiting RepA recruitment of polycomb complexes to maintain the active X chromosome in females	Mutagenesis	[128,221–223]
TUG1	6.7 kb; expressed in the developing retina and brain	Required for the proper formation of photoreceptors in the developing rodent retina	shRNA knockdown	[17]
UCA1	1,442 nt; expressed in embryonic development and bladder cancer	Enhances tumorigenic behaviour of bladder cancer cells in vitro and in vivo	Overexpression	[224]
<i>Xist</i>	~17 kb; mosaic expression in females	Epigenetically controls dosage compensation by silencing one of the two X chromosomes	Mutagenesis	for recent review see [99]
Y RNAs	83–112 nt; up-regulated in cancer; bound by Ro autoantigen	Regulate cell DNA replication and cell proliferation	siRNA knockdown	[47,225]
Zeb2NAT	>680 nt; antisense to Zeb2, a transcriptional repressor of E-cadherin	Regulates splicing of the Zeb2 5' UTR	Overexpression	[226]
Zfh-5AS	~10 kb; expressed in particular regions of the developing brain	Regulates expression of the transcription factor Zfh-5 mRNA	Mutagenesis	[227]

This list is not exhaustive, and there are other examples of functional ncRNAs in mammals (see e.g. [228]) as well as of regulatory and structural ncRNAs in other animals, plants, fungi (see e.g. [6–11]) and bacteria [229].

<sup>a</sup>Abbreviations used are: siRNA, short interfering RNA; shRNA, short hairpin RNA; FISH, fluorescence in situ hybridization; CHIP, chromatin immunoprecipitation; RNA-IP, immunoprecipitation of RNA associated with particular proteins.

doi:10.1371/journal.pgen.1000459.t001

relatively few regulatory mutations identified among the catalogue of known human mutations that are associated with overt genetic disease.

There is, of course, a wide spectrum of effects of both coding and noncoding mutations (Figure 2), and there are exceptions to the rule in both directions. Loss-of-function mutations in some

protein-coding genes have mild effects [92], as exemplified by as knockouts of the mammalian genes encoding calbindin D9k [93] and C/EBPdelta [94,95], and the significant number of yeast genes that show no observable phenotype. Reciprocally, knockouts of some highly conserved miRNAs, many of which have multiple targets, give strong phenotypes [96–98], even though very few

**Table 2.** Indices of the functionality of ncRNAs.

Feature	References
Conservation of promoters	[2,27,32]
Conservation of splice junctions	[27]
Conservation of sequence	[26,27,32]
Conservation of genomic position	[31,33,34]
Conservation of secondary structure	[28–30]
Positive selection	[230]
Conservation of expression	[35,36]
Dynamic expression and alternative splicing	[13,31,32]
Altered expression or splicing in cancer and other diseases	[37–49]
Association with particular chromatin signatures	[31,32]
Regulation by morphogens and transcription factors	[31,32,49,50]
Tissue- and cell-specific expression patterns	[16,17,19–22,49,51–56]
Specific subcellular localization	[19–22,52,56]

doi:10.1371/journal.pgen.1000459.t002

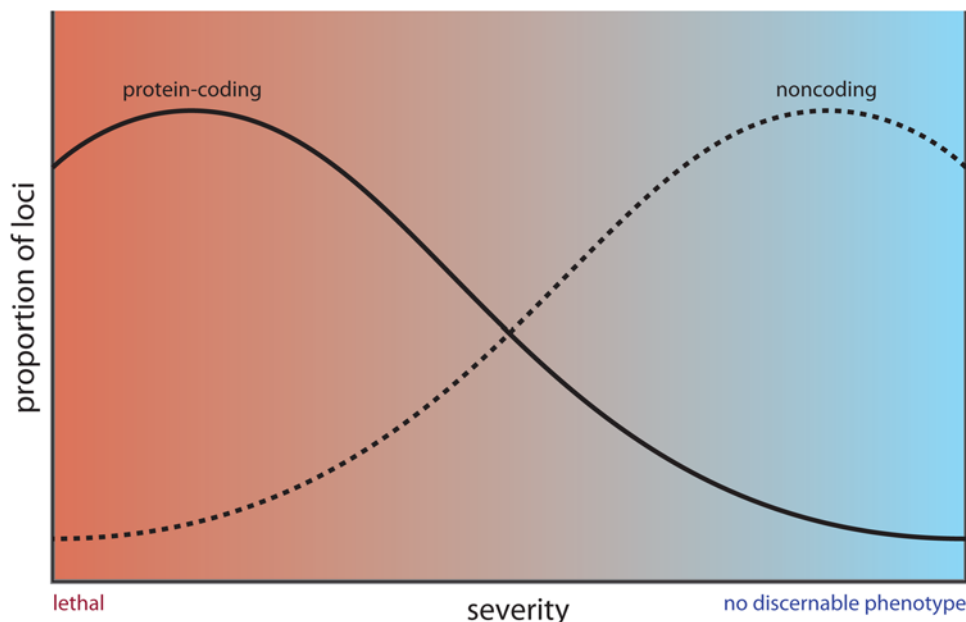
such genes have been identified in genetic screens in *Caenorhabditis elegans* and *Drosophila*, and none have been identified in mice, despite the intensity of such screens (see below). Moreover, to date, no naturally arising mutations have been discovered in the *Xist* gene in either humans or mice, despite the central role that this ncRNA plays in embryogenesis and in X-chromosome dosage control in females [99], possibly because such mutations are lethal.

On the other hand, the potential subtlety of mutations in ncRNAs, even when they are targeted by reverse genetics, is illustrated by the case of *BCI*, which is expressed in synaptodendritic domains of neurons in rodents. Knockout of this transcript produces no obvious physical or neurological abnormality, and the

mutant mice were initially indistinguishable from wild type, but were subsequently found to have reduced exploratory behaviour and consequently a higher mortality in field experiments [100]. Thus this ncRNA causes a subtle behavioural phenotype that is invisible in the cage, and would escape detection in superficial forward genetic screens, but is almost certainly strongly disadvantageous in the wild. Similarly, deletions or insertions in some ultraconserved enhancers yield no discernable abnormality [101,102], despite the fact that these sequences are clearly under intense selection [103], suggesting insensitivity of phenotypic screens in captivity or redundancy in the regulatory architecture (perhaps associated with developmental robustness) that we have yet to understand.

## Monogenic Diseases and High Penetrance Mutations

The generally stronger effects of protein-coding mutations leads to a sampling bias, in that more severe phenotypes are not only more easily discerned but have also been more likely to attract further study, both in medical contexts and in model organisms. In medicine, mutations involving catastrophic component damage have been traditionally referred to as “monogenic diseases” and were the primary targets of study in the pioneering days of human genomics, just a decade or two ago, when the protein-coding genes underpinning cystic fibrosis, Huntington disease, and Duchenne muscular dystrophy, among others, were identified by positional mapping and cloning approaches. Thus, in humans, mutation mapping is still both a young science and extraordinarily difficult due to the sheer complexity of the genome, which naturally led to an initial focus on severe loss-of-function diseases that exhibit simple inheritance patterns with high penetrance (“single gene–large effect”) making them amenable to identification. This is now changing with the increasing availability of human genome sequences, and the application of genome-wide association (GWA) studies to the mapping of the genetic components of



**Figure 2. The contrasting effects of mutations in protein-coding and regulatory sequences.** A conceptual diagram of the spectrum of phenotypic effects of mutations in sequences encoding proteins and other analogue components of cells (continuous line) versus variations in non-coding sequences that specify regulatory interactions (dashed line). doi:10.1371/journal.pgen.1000459.g002

complex traits, wherein lies the majority of the health burden and the majority of the interesting phenotypic differences between individuals and species.

## Complex Traits and GWA Studies

Indeed, functional classification of sequence variations that have been identified by GWA studies to be associated with complex traits, albeit still strongly focused on those of medical importance, shows that the vast majority of variations are located in noncoding regions [104,105]. In most cases, the causative mutations have yet to be defined, and their mechanistic basis is unknown—especially whether they affect *cis*-acting binding sites for regulatory proteins or the function or expression of regulatory RNAs. One good candidate for the latter is the ncRNA *ANRIL*, which lies antisense to *CDKN2A* and traverses a noncoding region centromeric to *CDKN2A*, a region implicated in a range of complex diseases including cancer, type 2 diabetes, periodontitis, and coronary heart disease [106–112]. Perplexingly, however, those variants mapped by GWA studies to date account for only a small proportion of genetic variation in disease or quantitative traits [113]. These traits are clearly multifactorial and may be affected by rare variants with strong effects that have yet to be recognized.

The identification of the sequence changes that directly underpin quantitative trait variations has thus far been possible, or at least achieved, only in well-structured pedigrees in plants and animals. In the few cases where such quantitative trait loci (QTLs) have been mapped to completion, most have been found to be located in noncoding sequences, specifically: (i) regulatory sequences in promoters and distal enhancers (e.g., the “*teosinte branched1*” mutation affecting branching and inflorescence in maize [114]); (ii) 3′ untranslated regions (UTRs) (e.g., those underlying Tourette’s syndrome [115], and muscular hypertrophy in sheep [116]; see also below); (iii) introns (e.g., a QTL affecting muscle growth in domestic pigs [117]); or (iv) intergenic sequences of unknown transcriptional status (e.g., the “*calipyge*” mutation causing posterior muscular hypertrophy in sheep [118]). The latter occurs in an imprinted locus and affects the expression of a number of protein-coding and ncRNA genes [119] associated with an unusual genetic phenomena termed “polar overdominance” [120], which may also occur in humans [121]. While these mutations are reasonably assumed to be regulatory in nature, their mechanistic basis has not been determined, although in the latter case, there is some evidence for the involvement of *trans*-acting miRNAs [119,122]. In addition, linkage studies in a large family have recently identified the ncRNA *AK023948* as a candidate susceptibility gene for papillary thyroid carcinoma [123].

## Type of Mutation and Sensitivity of the Model System

The nature of the organism under study and the type of mutations also affect the outcome of genetic screens—relevant single-base mutations are not only harder to identify than insertions/deletions, especially in mammalian genomes and even in inbred mice, but also have milder effects on regulatory sequences.

Most mutations induced by ENU mutagenesis involve single-base changes, which are also, along with small indels and copy number variations, the most common type of natural variation in humans and other mammals, where few regulatory mutations have yet been identified. Whereas a nonsynonymous mutation in a protein-coding sequence can have severe effects on the structure and function of the protein, many regulatory sequences have loose

consensus sequences, and variations in them, as noted already, may have subtle effects and go unnoticed, especially in superficial phenotypic screens.

On the other hand, insertions and deletions dominate mutational screens in *Drosophila*, and a large number map to noncoding intergenic and intronic regions. This is exemplified by the intensively studied *bithorax* complex, where there are not only mutations known in the coding sequences for the homeotic protein Ultrabithorax (Ubx), but also in upstream (*bithoraxoid* or *bxd*) and intronic (*Contrabithorax* or *Cbx* and *anterobithorax* or *abx*) sequences [124], which contain conserved blocks within them [125–127]. Such noncoding mutations are interpreted as affecting orthodox *cis*-acting enhancer sequences (i.e., those that bind *cis*-acting regulatory proteins), despite the fact that (for example) *bxd* mutations fall within a region that is transcribed during early embryogenesis into a complex set of short polyadenylated RNAs with no coding potential. These RNAs arise by alternative splicing of at least 11 exons derived from a 26-kb primary transcript [126]. Moreover, these regulatory regions involve interaction with Polycomb-group and Trithorax-group proteins, which are increasingly implicated as being directed to their sites of action by ncRNAs [7,13,128] (see below).

Similarly the *iab* regulatory elements of the bithorax complex that control the expression of *abdominal-A* and *Abdominal-B*, and consequently the identities of the 2nd–9th abdominal segments, are transcribed into ncRNAs in a spatially ordered pattern [129]. It has also recently been shown that 231 ncRNAs are expressed from the four human *HOX* loci in a spatially and temporally ordered progression along developmental axes, one of which (termed *HOTAIR*) from the *HOXC* locus controls expression of the *HOXD* locus *in trans*, as shown by siRNA-mediated knockdown experiments [13]. None of these ncRNAs has yet been specifically associated with a genetic variant in mammals, although in *Drosophila* there are many homeotic mutations that lie in regions encompassed by ncRNAs [12,124], and *Drosophila* geneticists were clearly intrigued by them [130]. Moreover, as pointed out by Rinn et al. [13], the existence of such transacting regulatory ncRNAs may explain the observation that the deletion of the entire *HOXC* locus exhibits a milder phenotype than the deletion of individual *HOXC* genes [131].

Despite the now known importance of miRNAs in the control of gene expression [67,132], only four miRNA loci have been identified in intense genetic screens in *C. elegans* and *Drosophila*, and none in mammals. This may be partly due to the fact that *C. elegans* is hermaphroditic and naturally driven to homozygosity in individual isolates, without the need for laborious back-crossing, and flies are routinely bred to homozygosity in mutational screens, making screening of recessive mutations far more efficient than in mice. There was also an element of serendipity, in that the first identified miRNA locus, *lin-4*, expressed a small RNA whose complementary sequence was present in multiple copies in the 3′ UTR of its target gene, *lin-14*, and thus was relatively easy to pinpoint [133,134], which also applied in the subsequent case of *let-7* [135].

Following the discovery of miRNAs, *Drosophila* mutants of uncertain provenance that mapped in gain-of-function screens to noncoding regions were re-analysed, and one that regulates growth [136], termed *bantam*, was identified to encode a miRNA [137]. The miRNAs *let-7* [135] and *ly-6* [138] were also identified genetically in *C. elegans*, not in other organisms, despite the former being not only highly conserved in sequence and expression pattern throughout metazoan evolution [139], but also fundamental to normal and abnormal developmental processes [140–142], as are many other miRNAs [143]. On the other hand, *ly-6*



is expressed in only a few cells [144], and has only rarely turned up in deep sequencing libraries. Subsequently most known miRNAs, of which there are hundreds in mammals, and later piRNAs, have been identified by biochemical not genetic means. In view of the *ly-6* example and the clearly incomplete sampling of the small RNA transcriptome, even using deep sequencing [70], there are likely to be many more.

## Expectations and Interpretations

There has also been a strong expectation that mutations that have phenotypic effects will map to protein-coding genes or *cis*-regulatory elements that interact with regulatory proteins. The former has influenced the practical strategies for mutation searching, in terms of a focus on exon scanning of candidate genes (see below), and the latter has influenced the interpretation of regulatory variations, although in only a few cases has the mechanistic basis been determined [71,145]. Some mutations map to gene “deserts” (see, e.g., [146]), and while it is conceivable that they affect distal enhancers (see, e.g., [147]), it is interesting and relevant to note that there is good evidence that enhancers and other regulatory sequences are transcribed into ncRNAs in the cells in which they are active [45,83–86], and hence may act in part via ncRNAs.

## Transvection and Locally Acting ncRNAs

Many loci, such as the *bithorax* complex referred to earlier, exhibit a genetic phenomenon called “transvection,” whereby a wild-type regulatory region upstream of a defective protein-coding sequence on one chromosome can rescue a relatively normal phenotype, when it is combined with a mutant regulatory region linked to a wild-type protein-coding sequence on the homologous chromosome (both of which give mutant phenotypes when homozygous) [130]. This phenomenon is well documented in *Drosophila* but appears to occur in most animals and has been interpreted as a physical cross-talk between functional *cis*-acting promoters or enhancers on one chromosome to engender transcription of adjacent protein-coding genes on the other, since the effect is usually pairing-dependent and lost when the regulatory and protein-coding sequences are separated to nonsyntenic positions in the genome [130,148]. However, this is not always the case—at some loci, transvection between regulatory elements and protein-coding sequences can operate over large distances (even between different chromosomes) [149–152], suggesting the involvement of a trans-acting signal. Moreover, many promoter elements that exhibit transvection are transcribed into ncRNAs, and transvection is altered in *Polycomb* and *zeste* mutants [125–127,130], indicating that epigenetic processes (which may be regulated by ncRNAs, see below) are involved. Taken together, these observations raise the possibility that transvection is mediated by *trans*-acting RNAs [153], in which case the observed cross-complementation may occur simply as a result of a compound heterozygosity between a regulatory ncRNA locus and a nearby protein-coding locus whose expression is controlled by it.

In support of this proposition, there is now rapidly emerging evidence that many ncRNAs derived from either same or opposite strands act locally to regulate the epigenetic status and expression of nearby protein-coding genes, often involving the recruitment of chromatin-activator or repressor complexes [7,13–16,89,154–157], with sense-antisense pairs in some cases being the substrate for the generation of siRNAs [63–65,68]. Moreover the many deletion studies of gene promoter regions to define regulatory sequences have almost always assumed, physically and mechanis-

tically, that resultant changes to expression patterns are due to the loss of *cis*-acting protein binding sites rather than deletions in the same or opposite strand ncRNAs that frequently traverse and are expressed from the same region. The complexity of these relationships is illustrated by the examples of the ncRNA *DLeu2* (deleted in lymphocytic leukemia 2), which has multiple splice variants and lies antisense to genes in a region deleted in various malignancies [158], and the ncRNA *ANRIL* referred to earlier. Thus the interpretation of the mechanisms by which such mutations operate remains not only an open question but a difficult problem to disentangle, given the complex interlacing and overlapping coding and noncoding transcripts, and splice variants thereof, that are expressed from many loci in different cells and tissues [2,3].

## Transinduction, Ectopic Expression and Gene Knockouts

RNA is also implicated in a curious genetic phenomenon called “transinduction,” whereby transient transfection of a  $\beta$ -globin gene induces transcription of the “locus control region” and intergenic regions at the chromosomal  $\beta$ -globin locus in non-erythroid cell lines. This effect is dependent on transcription of the globin gene from the transfected plasmid and its association with the endogenous  $\beta$ -globin locus, but not on protein expression, and therefore is RNA-mediated [83], although the responsible sequences have not been mapped. Indeed, the general assumption that mRNA is simply an intermediate between gene and protein, albeit with *cis*-acting regulatory elements, may be incorrect, and there may be a false dichotomy between coding and noncoding RNA [159]. This is indicated by the complexity of overlapping sense and antisense coding and noncoding transcripts from most genomic loci [2–5] and evidence that protein-coding sequences are under constraint not only at the amino acid sequence level, but also within their RNA sequence [160]. Moreover, given that many gene knockout studies concomitantly delete potential sources of regulatory ncRNAs such as introns (given that many miRNAs and all snoRNAs are sourced from introns) and antisense transcribed sequences, all aspects of the observed phenotypes cannot be unequivocally or solely ascribed to the loss of the protein without complementation studies or more precise deletions that are rarely done for reasons of technical difficulty.

## Technical Limitations

The focus on protein-coding sequences has been reinforced by a practical problem, especially in mammals. Mutation mapping studies using whole-genome scanning techniques usually have not pinpointed the causative mutation or variation, and the affected region may encompass one megabase or more. Until recently, comparative sequencing of such regions was not feasible, and in any case it can be very difficult to sort the relevant polymorphism from the considerable background variation in huge intergenic and intronic regions, especially in outbred human populations. In most circumstances, understandably, investigators have resorted to analysing the most plausible candidate genes (recently including those expressing noncoding transcripts or ESTs [158,161]) usually involving scanning of known exons (and sometimes the immediate 5′ flanking promoter sequences and UTRs) in the region by PCR amplification, in the hope that they can identify the causative mutation in these locations, which in turn are the ones that are then reported in the literature. However, there are many informal reports of mapping studies that have not identified such exonic mutations, and which consequently lie in abeyance, including (as noted already) the large number of GWA studies that have mapped disease-associated variations to noncoding, presumably

regulatory, regions [104]. A reasonable strategy for searching for the relevant variations in these regions may be to focus on sequences that exhibit evolutionary conservation and/or whose expression is altered [162]. Conversely, reverse genetic screens looking for phenotypes associated with ncRNAs might target conserved blocks within them and focus on tissues where they are known to be expressed.

### Mutations Affecting *trans*-Acting Functions of mRNAs

As noted above, regulatory mutations have been identified in the 3' UTRs of mRNAs, such as those underlying Tourette's syndrome [115] and muscular hypertrophy in sheep [116], which appear to involve gain- or loss-of-function of miRNA binding sites. Interestingly, however, a number of other reported 3' UTR mutations do not appear to act in *cis* to regulate the expression of the associated mRNA, as is normally assumed, but rather in *trans* as ncRNAs. For example the 3' UTR of prohibitin (in the absence of the associated protein-coding sequences) can inhibit cell cycle progression in one complementation group of breast cancer-derived cells that is characterized by naturally occurring mutations in the 3' UTR, indicating that these sequences are in fact functioning, in part, as *trans*-acting ncRNAs [163,164]. Similarly, the oogenesis defect observed in *Drosophila oskar* null mutants is rescued by the *oskar* 3' UTR alone [165].

A *trans*-acting function for mRNA sequences, both coding and noncoding, may be more general than expected. For example, the introduction of cancer-associated silent point-mutations in *p53* mRNA alters its binding to the protein Mdm2, which in turn alters *p53* expression and function [166]. The 3' UTRs of troponin I, tropomyosin, and  $\alpha$ -cardiac actin have been shown to reactivate muscle-specific promoters in a differentiation-defective myoblast mutant, enhance the differentiation of wild-type muscle cells, and suppress the proliferation of fibroblasts independently of their normally associated protein-coding sequences [167]. Similarly, the 3' UTRs of tropomyosin [168] and ribonucleotide reductase [169] can suppress tumour formation, and the 3' UTR of the DM protein kinase gene, which is involved in myotonic dystrophy, inhibits the differentiation of C2C12 myoblasts [170]. Moreover, many 3' UTRs in mouse appear to be expressed separately from their mRNAs in a developmentally regulated manner [171].

There are other examples of mutations in sequences encoding 3' UTRs that do not act via the UTR. A single nucleotide polymorphism that determines susceptibility to an autoimmune thyroid disease occurs both within the 3' UTR of the *ZFAT* gene (zinc-finger gene in AITD susceptibility region) and also within the promoter of an antisense transcript (*SAS-ZFAT*), and increases the expression of *ZFAT* not through increasing mRNA stability, but by repressing the expression of the antisense transcript [172].

### Regulation of Complex Genetic Processes by ncRNAs

Apart from the general presumption that most ncRNAs will be involved in regulation, variations in which will often have, individually, subtle effects on phenotype, there is, in fact, general evidence of their positive genetic signatures, as ncRNAs underpin most, if not all, complex genetic processes in the higher organisms. These include RNA interference-related phenomena such as co-suppression and transcriptional gene silencing [132,173–176], as well as position effect variegation [177,178], hybrid dysgenesis [179], parental imprinting, X-chromosome dosage compensation

and allelic exclusion [180], germ cell reprogramming [181], and paramutation [182,183], all of which involve epigenetic processes. Indeed, as noted already, there is increasing evidence that a major function of ncRNAs, both small and large, is the regulation of epigenetic memory through modifications to DNA and chromatin structure, involving the recruitment of DNA methyltransferases, histone-modifying enzymes, and chromatin remodelling complexes to their appropriate sites of action (including ncRNA genes themselves) in particular cells at particular stages of differentiation (for reviews, see [12,89,184]; also Table 1).

### Examples of Mutations in ncRNAs

There are some known examples of mutations in ncRNAs, aside from those mentioned already, that give recognizable phenotypes or that are strongly implicated in altered phenotypic states. These include a triplet repeat expansion in the ncRNA SCA8, which causes the human neurodegenerative disease Spinocerebellar Ataxia 8 (which as a transgene can induce progressive retinal neurodegeneration in *Drosophila*) [39] and other examples of deleterious gain-of-function mutations in noncoding RNAs associated with diseases such as myotonic dystrophy [185], deletions encompassing ncRNA loci and alterations to ncRNA splicing patterns in various cancers [47,106,158,161,186,187], and a SNP variant in an ncRNA MIAT that confers risk of myocardial infarction [188]. They also include many ncRNAs, including small nucleolar RNAs, that appear to be important in the mechanism of imprinting [189] and the molecular etiology of associated pathologies such as Prader-Willi syndrome [190,191], some that are implicated as tumour suppressors [168,192], or that are located at chromosomal translocation breakpoints associated with B-cell lymphoma [193] and schizophrenia [194]. It has also been shown that the translocation and induced expression of an antisense, long ncRNA can cause the epigenetic silencing of the adjacent  $\alpha$ -globin gene, resulting in  $\alpha$ -thalassemia [195].

### Conclusion

There is not (yet) a huge catalogue of mutations in ncRNAs that have been shown to affect phenotype, compared to those in protein-coding sequences. However, on the assumption that most ncRNAs are regulatory and that most regulatory regions have yet to be assigned genetic signatures, it is no surprise that this may be the case. On the other hand, as screening for variations affecting complex traits becomes more sophisticated, it is reasonable to anticipate that many will map to, and affect the function of, ncRNAs. Certainly this possibility should be borne in mind in the interpretation of such variations and the consequent studies to define their mechanism of action. The functional analysis of ncRNAs is in its infancy, but in situ hybridization, genomic, and structural characteristics, and the perturbation of their expression by overexpression and siRNA-mediated knockdown are emerging as major tools (Tables 1 and 2). There seems little doubt that there is a hidden world of regulatory architecture underpinning the development of complex organisms that we have yet to explore, both genetically and functionally.

### Acknowledgments

I thank Paulo Amaral, Grant Montgomery, and Manolis Dermitzakis for helpful comments on the manuscript; Paulo Amaral for assistance in the construction of Table 1; Ryan Taft for providing the analysis shown in Figure 1; and Tim Mercer for drawing the figures.



## References

- Taft RJ, Pheasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29: 288–299.
- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005) The transcriptional landscape of the mammalian genome. *Science* 309: 1559–1563.
- Kapranov P, Drenkow J, Cheng J, Long J, Helt G, et al. (2005) Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res* 15: 987–997.
- Mattick JS, Makunin IV (2006) Non-coding RNA. *Hum Mol Genet* 15: R17–29.
- Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* 8: 413–423.
- Rajendra TK, Prasanth KV, Lakhotia SC (2001) Male sterility associated with overexpression of the noncoding hromosome gene in cyst cells of testis of *Drosophila melanogaster*. *J Genet* 80: 97–110.
- Sanchez-Elsner T, Gou D, Kremmer E, Sauer F (2006) Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to *Ultrabithorax*. *Science* 311: 1118–1123.
- Hellwig S, Bass BL (2008) A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. *Proc Natl Acad Sci U S A* 105: 12897–12902.
- Ma J, Yan B, Qu Y, Qin F, Yang Y, et al. (2008) Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J Cell Biochem* 105: 136–146.
- Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, et al. (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Res* 19: 57–69.
- Camblong J, Iglesias N, Fickentscher C, Dieppois G, Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131: 706–717.
- Amaral PP, Mattick JS (2008) Noncoding RNA in development. *Mamm Genome* 19: 454–492.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* 129: 1311–1323.
- Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, et al. (2008) Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* 451: 202–206.
- Martianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445: 666–670.
- Blin-Wakkach C, Lezot F, Ghoul-Mazgar S, Hotton D, Monteiro S, et al. (2001) Endogenous *Mxl1* antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. *Proc Natl Acad Sci U S A* 98: 7336–7341.
- Young TL, Matsuda T, Cepko CL (2005) The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr Biol* 15: 501–512.
- Cano A, Nieto MA (2008) Non-coding RNAs take centre stage in epithelial-to-mesenchymal transition. *Trends Cell Biol* 18: 357–359.
- Sunwoo H, Dinger ME, Wilusz JE, Amaral PP, Mattick JS, et al. (2009) *MEV*  $\epsilon/\beta$  nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res* 19: 347–359.
- Sasaki YT, Ideue T, Sano M, Mituyama T, Hirose T (2009) MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc Natl Acad Sci U S A* 106: 2525–2530.
- Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, et al. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*, epub ahead of print; doi:10.1016/j.molcel.2009.01.026.
- Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, et al. (2009) The long noncoding RNA *Kcnq1ot1* organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development* 136: 525–530.
- Prasanth KV, Spector DL (2007) Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev* 21: 11–42.
- Louro R, Smirnova AS, Verjovski-Almeida S (2009) Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93: 291–298.
- Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10: 155–159.
- Pang KC, Frith MC, Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 22: 1–5.
- Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* 17: 556–565.
- Washietl S, Hofacker IL, Lukasser M, Huttenhofer A, Stadler PF (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat Biotechnol* 23: 1383–1390.
- Torarinsson E, Sawera M, Havgaard JH, Fredholm M, Gorodkin J (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res* 16: 885–889.
- Torarinsson E, Yao Z, Wiklund ED, Bramsen JB, Hansen C, et al. (2008) Comparative genomics beyond sequence-based alignments: RNA structures in the ENCODE regions. *Genome Res* 18: 242–251.
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, et al. (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* 18: 1433–1445.
- Guttman M, Amit I, Garber M, French C, Lin MF, et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458: 223–227.
- Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otilar RP, et al. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res* 14: 62–66.
- Engstrom PG, Suzuki H, Ninomiya N, Akalin A, Sessa L, et al. (2006) Complex loci in human and mouse genomes. *PLoS Genet* 2: e47. doi:10.1371/journal.pgen.0020047.
- Tupy JL, Bailey AM, Dailey G, Evans-Holm M, Siebel CW, et al. (2005) Identification of putative noncoding polyadenylated transcripts in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 102: 5495–5500.
- Louro R, El-Jundi T, Nakaya HI, Reis EM, Verjovski-Almeida S (2008) Conserved tissue expression signatures of intronic noncoding RNAs transcribed from human and mouse loci. *Genomics* 92: 18–25.
- Thrash-Bingham CA, Tartof KD (1999) aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia. *J Natl Cancer Inst* 91: 143–151.
- Ji P, Diederichs S, Wang W, Boing S, Metzger R, 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: 6087–6097.
- Mutsuddi M, Marshall CM, Benzow KA, Koob MD, Rebay I (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in *Drosophila*. *Curr Biol* 14: 302–308.
- Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschart AV, et al. (2004) Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene* 23: 6684–6692.
- Reis EM, Ojopi EP, Alberto FL, Rahal P, Tsukumo F, et al. (2005) Large-scale transcriptome analyses reveal new genetic marker candidates of head, neck, and thyroid cancer. *Cancer Res* 65: 1693–1699.
- Sonkoly E, Bata-Csorgo Z, Pivarcsi A, Polyanka H, Kenderessy-Szabo A, et al. (2005) Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS. *J Biol Chem* 280: 24159–24167.
- Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J (2005) A new frontier for molecular medicine: noncoding RNAs. *Biochim Biophys Acta* 1756: 65–75.
- Angeloni D, ter Elst A, Wei MH, van der Veen AY, Braga EA, et al. (2006) Analysis of a new homozygous deletion in the tumor suppressor region at 3p12.3 reveals two novel intronic noncoding RNA genes. *Genes Chromosomes Cancer* 45: 676–691.
- Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, et al. (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* 12: 215–229.
- Pang KC, Stephen S, Dinger ME, Engstrom PG, Lenhard B, et al. (2007) RNAdb 2.0—an expanded database of mammalian non-coding RNAs. *Nucleic Acids Res* 35: D178–182.
- Christov CP, Trivier E, Krude T (2008) Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. *Br J Cancer* 98: 981–988.
- Perez DS, Hoage TR, Pritchett JR, Ducharme-Smith AL, Halling ML, et al. (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum Mol Genet* 17: 642–655.
- Zhang X, Lian Z, Padden C, Gerstein MB, Rozowsky J, et al. (2009) A myelopoiesis-associated regulatory intergenic non-coding RNA transcript within the human *HOXA* cluster. *Blood* 113: 2526–2534.
- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, 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.
- Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, et al. (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res* 16: 11–19.
- Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci U S A* 105: 716–721.
- Kohtz JD, Fishell G (2004) Developmental regulation of EVF-1, a novel non-coding RNA transcribed upstream of the mouse *Dlx6* gene. *Gene Expr Patterns* 4: 407–412.
- Inagaki S, Numata K, Kondo T, Tomita M, Yasuda K, et al. (2005) Identification and expression analysis of putative mRNA-like non-coding RNA in *Drosophila*. *Genes Cells* 10: 1163–1173.

55. Brena C, Chipman AD, Minelli A, Akam M (2006) Expression of trunk Hox genes in the centipede *Strigamia maritima*: sense and anti-sense transcripts. *Evol Dev* 8: 252–265.
56. Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, et al. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci* 120: 2498–2506.
57. Pollard KS, Salama SR, King B, Kern AD, Dreszer T, et al. (2006) Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2: e168. doi:10.1371/journal.pgen.0020168.
58. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A (2004) Identification of mammalian microRNA host genes and transcription units. *Genome Res* 14: 1902–1910.
59. Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318: 761–764.
60. Ruby JG, Jan CH, Bartel DP (2007) Intronic microRNA precursors that bypass Droscha processing. *Nature* 448: 83–86.
61. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC (2007) The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130: 89–100.
62. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC (2007) Mammalian mirtron genes. *Mol Cell* 28: 328–336.
63. Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, et al. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453: 534–538.
64. Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, et al. (2008) Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453: 539–543.
65. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, et al. (2008) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453: 803–806.
66. Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of piRNAs. *Development* 135: 3–9.
67. Stefani G, Slack FJ (2008) Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* 9: 219–230.
68. Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA interference and X-inactivation pathways. *Science* 320: 1336–1341.
69. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37: 766–770.
70. Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, et al. (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res* 16: 1289–1298.
71. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134: 25–36.
72. Smith NG, Brandstrom M, Ellegren H (2004) Evidence for turnover of functional noncoding DNA in mammalian genome evolution. *Genomics* 84: 806–813.
73. Frith MC, Ponjavic J, Fredman D, Kai C, Kawai J, et al. (2006) Evolutionary turnover of mammalian transcription start sites. *Genome Res* 16: 713–722.
74. Taylor MS, Kai C, Kawai J, Carninci P, Hayashizaki Y, et al. (2006) Heterotachy in mammalian promoter evolution. *PLoS Genet* 2: e30. doi:10.1371/journal.pgen.0020030.
75. Pheasant M, Mattick JS (2007) Raising the estimate of functional human sequences. *Genome Res* 17: 1245–1253.
76. Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 312: 276–279.
77. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, et al. (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420: 578–582.
78. Dermitzakis ET, Reymond A, Scamuffa N, Ucla C, Kirkness E, et al. (2003) Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* 302: 1033–1035.
79. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, et al. (2004) Ultraconserved elements in the human genome. *Science* 304: 1321–1325.
80. Stephen S, Pheasant M, Makunin IV, Mattick JS (2008) Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol Biol Evol* 25: 402–408.
81. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499–502.
82. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, et al. (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet* 40: 158–160.
83. Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ (1997) Intergenic transcription and transduction of the human beta-globin locus. *Genes Dev* 11: 2494–2509.
84. Ling J, Baibakov B, Pi W, Emerson BM, Tuan D (2005) The HS2 enhancer of the beta-globin locus control region initiates synthesis of non-coding, polyadenylated RNAs independent of a cis-linked globin promoter. *J Mol Biol* 350: 883–896.
85. Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C (2008) The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol Cell* 32: 129–139.
86. Lempradl A, Ringrose L (2008) How does noncoding transcription regulate *Hox* genes? *Bioessays* 30: 110–121.
87. Mattick JS (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 25: 930–939.
88. Mattick JS (2007) A new paradigm for developmental biology. *J Exp Biol* 210: 1526–1547.
89. Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF (2009) RNA regulation of epigenetic processes. *Bioessays* 31: 51–59.
90. Liu ET (2008) Functional genomics of cancer. *Curr Opin Genet Dev* 18: 251–256.
91. Tomasini R, Mak TW, Melino G (2008) The impact of p53 and p73 on aneuploidy and cancer. *Trends Cell Biol* 18: 244–252.
92. Barbaric I, Miller G, Dear TN (2007) Appearances can be deceiving: phenotypes of knockout mice. *Brief Funct Genomic Proteomic* 6: 91–103.
93. Kutuzova GD, Akhter S, Christakos S, Vanhooke J, Kimmel-Jehan C, et al. (2006) Calbindin D(9k) knockout mice are indistinguishable from wild-type mice in phenotype and serum calcium level. *Proc Natl Acad Sci U S A* 103: 12377–12381.
94. Crawley JN (1999) Behavioral phenotyping of transgenic and knockout mice: experimental design and evaluation of general health, sensory functions, motor abilities, and specific behavioral tests. *Brain Res* 835: 18–26.
95. Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, et al. (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129: 303–317.
96. Thai TH, Calado DP, Casola S, Ansel KM, Xiao C, et al. (2007) Regulation of the germinal center response by microRNA-155. *Science* 316: 604–608.
97. Rodriguez A, Vigorito E, Clare S, Warren MV, Couttet P, et al. (2007) Requirement of bic/microRNA-155 for normal immune function. *Science* 316: 608–611.
98. van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, et al. (2007) Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 316: 575–579.
99. Ng K, Pullirsch D, Leeb M, Wutz A (2007) Xist and the order of silencing. *EMBO Rep* 8: 34–39.
100. Lewejohann L, Skryabin BV, Sachser N, Prehn C, Heiduschka P, et al. (2004) Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. *Behav Brain Res* 154: 273–289.
101. Poulin F, Nobrega MA, Plajzer-Frick I, Holt A, Afzal V, et al. (2005) In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics* 85: 774–781.
102. Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, et al. (2007) Deletion of ultraconserved elements yields viable mice. *PLoS Biol* 5: e234. doi:10.1371/journal.pbio.0050234.
103. Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, et al. (2007) Human genome ultraconserved elements are ultraselected. *Science* 317: 915.
104. Hindorf LA, Junkins HA, Manolio TA (2008) A catalog of published genome-wide association studies. Available at: [www.genome.gov/26525384/](http://www.genome.gov/26525384/).
105. Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 118: 1590–1605.
106. Pasmant E, Laurendeau I, Heron D, Vidaud M, Vidaud D, et al. (2007) Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. *Cancer Res* 67: 3963–3969.
107. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316: 1341–1345.
108. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316: 1336–1341.
109. McPherson R, Pertsemidis A, Kavaslar N, Stewart A, Roberts R, et al. (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316: 1488–1491.
110. Helgadóttir A, Thorleifsson G, Manolescu A, Gretarsdóttir S, Blondal T, et al. (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 316: 1491–1493.
111. Broadbent HM, Peden JF, Lorkowski S, Goel A, Ongen H, et al. (2008) Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet* 17: 806–814.
112. Schaefer AS, Richter GM, Groessner-Schreiber B, Noack B, Nothnagel M, et al. (2009) Identification of a shared genetic susceptibility locus for coronary heart disease and periodontitis. *PLoS Genet* 5: e1000378. doi:10.1371/journal.pgen.1000378.
113. Wray NR, Goddard ME, Visscher PM (2008) Prediction of individual genetic risk of complex disease. *Curr Opin Genet Dev* 18: 257–263.
114. Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 38: 594–597.

115. Abelson JF, Kwan KY, O'Roak BJ, Back DY, Stillman AA, et al. (2005) Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science* 310: 317–320.
116. Clop A, Marq F, Takeda H, Pirrottin D, Tordoir X, et al. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* 38: 813–818.
117. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, et al. (2003) A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* 425: 832–836.
118. Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, et al. (2003) Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* 163: 453–456.
119. Hagan JP, O'Neill BL, Stewart CL, Kozlov SV, Croce CM (2009) At least ten genes define the imprinted Dlk1-Dio3 cluster on mouse chromosome 12qF1. *PLoS ONE* 4: e4352. doi:10.1371/journal.pone.0004352.
120. Georges M, Charlier C, Cockett N (2003) The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet* 19: 248–252.
121. Wernter AK, Scherag A, Meyre D, Reichwald K, Durand E, et al. (2008) Preferential reciprocal transfer of paternal/maternal *DLK1* alleles to obese children: first evidence of polar overdominance in humans. *Eur J Hum Genet* 16: 1126–1134.
122. Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, et al. (2005) RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus. *Curr Biol* 15: 743–749.
123. He H, Nagy R, Liyanarachchi S, Jiao H, Li W, et al. (2009) A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res* 69: 625–631.
124. Duncan I (1987) The bithorax complex. *Annu Rev Genet* 21: 285–319.
125. Akam ME, Martinez-Arias A (1985) The distribution of Ultrabithorax transcripts in *Drosophila* embryos. *Embo J* 4: 1689–1700.
126. Lipshitz HD, Peattie DA, Hogness DS (1987) Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev* 1: 307–322.
127. Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, et al. (2006) Transcription of bxd noncoding RNAs promoted by trithorax represses *Ubx* in cis by transcriptional interference. *Cell* 127: 1209–1221.
128. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322: 750–756.
129. Sanchez-Herrero E, Akam M (1989) Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* 107: 321–329.
130. Duncan IW (2002) Transvection effects in *Drosophila*. *Annu Rev Genet* 36: 521–556.
131. Suemori H, Noguchi S (2000) *HoxC* cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev Biol* 220: 333–342.
132. Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 20: 759–771.
133. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843–854.
134. Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855–862.
135. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901–906.
136. Hipfner DR, Weigmann K, Cohen SM (2002) The bantam gene regulates *Drosophila* growth. *Genetics* 161: 1527–1537.
137. Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM (2003) bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113: 25–36.
138. Sarin S, O'Meara MM, Flowers EB, Antonio C, Poole RJ, et al. (2007) Genetic screens for *Caenorhabditis elegans* mutants defective in left/right asymmetric neuronal fate specification. *Genetics* 176: 2109–2130.
139. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, et al. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408: 86–89.
140. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, et al. (2005) RAS is regulated by the let-7 microRNA family. *Cell* 120: 635–647.
141. Yu F, Yao H, Zhu P, Zhang X, Pan Q, et al. (2007) let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 131: 1109–1123.
142. Roush S, Slack FJ (2008) The let-7 family of microRNAs. *Trends Cell Biol*.
143. Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* 6: 259–269.
144. Johnston RJ, Hobert O (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* 426: 845–849.
145. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, et al. (2006) A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* 312: 1215–1217.
146. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, et al. (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of *PTGER4*. *PLoS Genet* 3: e58. doi:10.1371/journal.pgen.0030058.
147. Marcellini S, Simpson P (2006) Two or four bristles: functional evolution of an enhancer of scute in *Drosophilidae*. *PLoS Biol* 4: e386. doi:10.1371/journal.pbio.0040386.
148. Sipos L, Mihalj J, Karch F, Schedl P, Gausz J, et al. (1998) Transvection in the *Drosophila Abd-B* domain: extensive upstream sequences are involved in anchoring distant cis-regulatory regions to the promoter. *Genetics* 149: 1031–1050.
149. Hendrickson JE, Sakonju S (1995) Cis and trans interactions between the *iab* regulatory regions and *abdominal-A* and *abdominal-B* in *Drosophila melanogaster*. *Genetics* 139: 835–848.
150. Hopmann R, Duncan D, Duncan I (1995) Transvection in the *iab-5,6,7* region of the bithorax complex of *Drosophila*: homology independent interactions in trans. *Genetics* 139: 815–833.
151. Muller M, Hagstrom K, Gyurkovics H, Pirrotta V, Schedl P (1999) The *mcp* element from the *Drosophila melanogaster bithorax* complex mediates long-distance regulatory interactions. *Genetics* 153: 1333–1356.
152. Vazquez J, Muller M, Pirrotta V, Sedat JW (2006) The *Mcp* element mediates stable long-range chromosome-chromosome interactions in *Drosophila*. *Mol Biol Cell* 17: 2158–2165.
153. Mattick JS, Gagen MJ (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol* 18: 1611–1630.
154. Feng J, Bi C, Clark BS, Mady R, Shah P, et al. (2006) The *Evf-2* noncoding RNA is transcribed from the *Dlx-5/6* ultraconserved region and functions as a *Dlx-2* transcriptional coactivator. *Genes Dev* 20: 1470–1484.
155. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32: 232–246.
156. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, et al. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322: 1717–1720.
157. Terranova R, Yokobayashi S, Stadler MB, Otte AP, van Lohuizen M, et al. (2008) Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev Cell* 15: 1–12.
158. Corcoran MM, Hammarsund M, Zhu C, Lerner M, Kapanadze B, et al. (2004) DLEU2 encodes an antisense RNA for the putative bicistronic *RFP2/LEU5* gene in humans and mouse. *Genes Chromosomes Cancer* 40: 285–297.
159. Dinger ME, Pang KC, Mercer TR, Mattick JS (2008) Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput Biol* 4: e1000176. doi:10.1371/journal.pcbi.1000176.
160. Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet* 7: 98–108.
161. Davison EJ, Tarpey PS, Fiegler H, Tomlinson IP, Carter NP (2005) Deletion at chromosome band 20p12.1 in colorectal cancer revealed by high resolution array comparative genomic hybridization. *Genes Chromosomes Cancer* 44: 384–391.
162. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. *Nat Genet* 39: 1202–1207.
163. Jupe ER, Liu XT, Kiehlbauch JL, McClung JK, Dell'Orco RT (1996) Prohibitin in breast cancer cell lines: loss of antiproliferative activity is linked to 3' untranslated region mutations. *Cell Growth Differ* 7: 871–878.
164. Jupe ER, Liu XT, Kiehlbauch JL, McClung JK, Dell'Orco RT (1996) The 3' untranslated region of prohibitin and cellular immortalization. *Exp Cell Res* 224: 128–135.
165. Jenny A, Hachet O, Zavorszky P, Cyrklaff A, Weston MD, et al. (2006) A translation-independent role of oskar RNA in early *Drosophila* oogenesis. *Development* 133: 2827–2833.
166. Candéas MM, Malbert-Colas L, Powell DJ, Daskalogianni C, Maslon MM, et al. (2008) p53 mRNA controls p53 activity by managing Mdm2 functions. *Nat Cell Biol*.
167. Rastinejad F, Blau HM (1993) Genetic complementation reveals a novel regulatory role for 3' untranslated regions in growth and differentiation. *Cell* 72: 903–917.
168. Rastinejad F, Conboy MJ, Rando TA, Blau HM (1993) Tumor suppression by RNA from the 3' untranslated region of alpha-tropomyosin. *Cell* 75: 1107–1117.
169. Fan H, Villegas C, Huang A, Wright JA (1996) Suppression of malignancy by the 3' untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs. *Cancer Res* 56: 4366–4369.
170. Amack JD, Paguio AP, Mahadevan MS (1999) *Cis* and *trans* effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum Mol Genet* 8: 1975–1984.
171. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, et al. (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38: 626–635.
172. Shirasawa S, Harada H, Furugaki K, Akamizu T, Ishikawa N, et al. (2004) SNPs in the promoter of a B cell-specific antisense transcript, SAS-ZFAT, determine susceptibility to autoimmune thyroid disease. *Hum Mol Genet* 13: 2221–2231.
173. Mette MF, Aufsatz W, van Der Winden J, Matzke MA, Matzke AJ (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* 19: 5194–5201.

174. Pal-Bhadra M, Bhadra U, Birchler JA (2002) RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. *Mol Cell* 9: 315–327.
175. Buhler M, Moazed D (2007) Transcription and RNAi in heterochromatic gene silencing. *Nat Struct Mol Biol* 14: 1041–1048.
176. Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJ (2007) Targets of RNA-directed DNA methylation. *Curr Opin Plant Biol* 10: 512–519.
177. Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, et al. (2004) Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* 303: 669–672.
178. Singh J, Freeling M, Lisch D (2008) A position effect on the heritability of epigenetic silencing. *PLoS Genet* 4: e1000216. doi:10.1371/journal.pgen.1000216.
179. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322: 1387–1392.
180. Yang PK, Kuroda MI (2007) Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* 128: 777–786.
181. Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 9: 129–140.
182. Chandler VL (2007) Paramutation: from maize to mice. *Cell* 128: 641–645.
183. Cuzin F, Grandjean V, Rassoulzadegan M (2008) Inherited variation at the epigenetic level: paramutation from the plant to the mouse. *Curr Opin Genet Dev* 18: 193–196.
184. Costa FF (2008) Non-coding RNAs, epigenetics and complexity. *Gene* 410: 9–17.
185. Osborne RJ, Thornton CA (2006) RNA-dominant diseases. *Hum Mol Genet* 15: R162–169.
186. Angeloni D, Ter Elst A, Wei MH, van der Veen AY, Braga EA, et al. (2006) Analysis of a new homozygous deletion in the tumor suppressor region at 3p12.3 reveals two novel intronic noncoding RNA genes. *Genes Chromosomes Cancer* 45: 676–691.
187. Dallosso AR, Hancock AL, Malik S, Salpekar A, King-Underwood L, et al. (2007) Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer. *Rna* 13: 2287–2299.
188. Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, et al. (2006) Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J Hum Genet* 51: 1087–1099.
189. Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415: 810–813.
190. Ding F, Prints Y, Dhar MS, Johnson DK, Garnacho-Montero C, et al. (2005) Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader-Willi syndrome mouse models. *Mamm Genome* 16: 424–431.
191. Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, et al. (2008) Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 40: 719–721.
192. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, et al. (2007) Activation of p53 by MEG3 non-coding RNA. *J Biol Chem* 282: 24731–24742.
193. Tanaka R, Satoh H, Moriyama M, Satoh K, Morishita Y, et al. (2000) Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. *Genes Cells* 5: 277–287.
194. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, et al. (2000) Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Human Mol Genet* 9: 1415–1423.
195. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, et al. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34: 157–165.
196. Seidl CI, Stricker SH, Barlow DP (2006) The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *Embo J* 25: 3565–3575.
197. Faghilhi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, 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.
198. Lin D, Pestova TV, Hellen CU, Tiedge H (2008) Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol Cell Biol* 28: 3008–3019.
199. Takeda K, Ichijo H, Fujii M, Mochida Y, Saitoh M, et al. (1998) Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J Biol Chem* 273: 17079–17085.
200. Wang X, Arai S, Song X, Reichart D, Du K, et al. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* 454: 126–130.
201. Tsang WP, Wong TW, Cheung AH, Co CN, Kwok TT (2007) Induction of drug resistance and transformation in human cancer cells by the noncoding RNA CUDR. *Rna* 13: 890–898.
202. Wagner LA, Christensen CJ, Dunn DM, Spangrude GJ, Georgelas A, et al. (2007) EGO, a novel, non-coding RNA gene, regulates eosinophil granule protein transcript expression. *Blood* 109: 5191–5198.
203. Brookheart RT, Michel CI, Listenberger LL, Ory DS, Schaffer JE (2009) The non-coding RNA *gadd7* is a regulator of lipid-induced oxidative and ER stress. *J Biol Chem* 284: 7446–7454.
204. Mourada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT (2008) Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J Cell Sci* 121: 939–946.
205. Matouk IJ, DeGroot N, Mezan S, Ayesh S, Abu-lail R, et al. (2007) The H19 non-coding RNA is essential for human tumor growth. *PLoS ONE* 2: e845. doi:10.1371/journal.pone.0000845.
206. Gabory A, Ripoché MA, Yoshimizu T, Dandolo L (2006) The H19 gene: regulation and function of a non-coding RNA. *Cytogenet Genome Res* 113: 188–193.
207. Yoshimizu T, Miroglio A, Ripoché MA, Gabory A, Vernucci M, et al. (2008) The H19 locus acts in vivo as a tumor suppressor. *Proc Natl Acad Sci U S A* 105: 12417–12422.
208. Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E (2006) RNA-mediated response to heat shock in mammalian cells. *Nature* 440: 556–560.
209. Mayer C, Schmitz KM, Li J, Grummt I, Santoro R (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* 22: 351–361.
210. Thakur N, Tiwari VK, Thomassin H, Pandey RR, Kanduri M, et al. (2004) An antisense RNA regulates the bidirectional silencing property of the *Kcnq1* imprinting control region. *Mol Cell Biol* 24: 7855–7862.
211. Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, et al. (2004) Non-coding RNA directed DNA demethylation of *Sphk1* CpG island. *Biochem Biophys Res Commun* 322: 593–600.
212. Madamanchi NR, Hu ZY, Li F, Horaist C, Moon SK, et al. (2002) A noncoding RNA regulates human protease-activated receptor-1 gene during embryogenesis. *Biochim Biophys Acta* 1576: 237–245.
213. Tochitani S, Hayashizaki Y (2008) Nks2.2 antisense RNA overexpression enhanced oligodendrocytic differentiation. *Biochem Biophys Res Commun* 372: 691–696.
214. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, et al. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309: 1570–1573.
215. Fu X, Ravindranath L, Tran N, Petrovics G, Srivastava S (2006) Regulation of apoptosis by a prostate-specific and prostate cancer-associated noncoding gene, PCGEM1. *DNA Cell Biol* 25: 135–141.
216. Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, et al. (2006) A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A* 103: 5781–5786.
217. Yan MD, Hong CC, Lai GM, Cheng AL, Lin YW, et al. (2005) Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. *Hum Mol Genet* 14: 1465–1474.
218. Valgardsdottir R, Chiodi I, Giordano M, Cobianchi F, Riva S, et al. (2005) Structural and functional characterization of noncoding repetitive RNAs transcribed in stressed human cells. *Mol Biol Cell* 16: 2597–2604.
219. Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318: 798–801.
220. Schoeffner S, Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* 10: 228–236.
221. Lee JT, Davidow LS, Warshawsky D (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genet* 21: 400–404.
222. Luikenhuis S, Wutz A, Jaenisch R (2001) Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Mol Cell Biol* 21: 8512–8520.
223. Sado T, Hoki Y, Sasaki H (2005) Tsix silences Xist through modification of chromatin structure. *Dev Cell* 9: 159–165.
224. Wang F, Li X, Xie X, Zhao L, Chen W (2008) UCA1, a non-protein-coding RNA up-regulated in bladder carcinoma and embryo, influencing cell growth and promoting invasion. *FEBS Lett* 582: 1919–1927.
225. Sim S, Weinberg DE, Fuchs G, Choi K, Chung J, et al. (2009) The subcellular distribution of an RNA quality control protein, the ro autoantigen, is regulated by noncoding y RNA binding. *Mol Biol Cell* 20: 1555–1564.
226. Beltran M, Puig I, Pena C, Garcia JM, Alvarez AB, et al. (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 22: 756–769.
227. Komine Y, Nakamura K, Katsuki M, Yamamori T (2006) Novel transcription factor zfh-5 is negatively regulated by its own antisense RNA in mouse brain. *Mol Cell Neurosci* 31: 273–283.
228. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, et al. (2005) Antisense transcription in the mammalian transcriptome. *Science* 309: 1564–1566.
229. Toledo-Arana A, Repoila F, Cossart P (2007) Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* 10: 182–188.
230. Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, et al. (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443: 167–172.