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 ncRNAs, are precursors for small regulatory RNAs like microRNAs (miRNAs), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), and small nucleolar RNAs ( snoRNAs) [38–68]. 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
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., [43,83–86]). 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

![Figure 1. The recent rise in papers on ncRNAs.](https://example.com/figure1.png)

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
| Name          | Characteristics                                                                 | Function                                                                                      | Experimental Methodology | References |
|--------------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|--------------------------|------------|
| Air          | 108 kb; transcribed from an antisense promoter located in intron 2 of Igf2r     | Regulates genomic imprinting of a cluster of autosomal genes on mouse chromosome 17         | Mutagenesis; FISH; ChIP; RNA-IP | [156,189,190] |
| BACE1AS      | ~2 kb; transcribed antisense to betasecretase-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                 | Anti-sense 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 esoinophil development | Regulates expression of myelin basic protein and esoinophil-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 Dlx6; developmentally regulated, expressed in the ventral forebrain | Cooperates with Dlx-2 in vivo to increase the transcriptional activity of the Dlx-5/6 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]      |
| GASS         | ~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 Igf2 locus; strongly expressed during embryogenesis; up-regulated in various tumours | Complex functions, influences growth by way of a cis control on Igf2 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 HOXC locus from a position intergenic and antisense to the flanking HOXC11 and HOXC12 genes | Epigenetically silences gene expression at the HOXD locus                                     | siRNA knockdown           | [13]       |
| HOTAIRM1     | 483 nt; specific to the myeloid lineage                                         | Involved in RA-induced expression of HOXA1 and HOXDA4 during myeloid differentiation, and induction of myeloid differentiation genes CD11b and CD18 | shRNA knockdown           | [49]       |
| Hsr1         | ~600 nt; ubiquitously expressed                                                 | Required for heat shock response                                                            | siRNA and anti-sense 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 Kcnq1 imprinting control region                | Mediates organization of a lineage-specific nuclear domain involved in epigenetic silencing of the Kcnq1 imprinting control region | Mutagenesis; ChIP; RNA/DNA FISH | [22,155, 157,210] |
| Khsps1a      | 1.290 nt; originates from the CpG island and overlaps a tissue-dependent differentially methylated region of Sphk1 | Regulates DNA methylation in the tissue-dependent differentially methylated region of Sphk1 | Overexpression            | [211]      |
| lincENC1     | Size unspecified; ~181 kb from Enc1                                            | Regulates cell proliferation in embryonal stem cells                                        | shRNA knockdown           | [32]       |
| MEG3         | ~1.6 kb; maternally expressed from the Dlk1-Gtl2 imprinted locus               | Regulates p53 expression, inhibits cell proliferation in the absence of p53                | Overexpression, mutagenesis | [192]      |
Table 1. cont.

| Name         | Characteristics                      | Function                                                                 | Experimental Methodology* | References               |
|--------------|--------------------------------------|--------------------------------------------------------------------------|---------------------------|-------------------------|
| MEN c/βi     | AT-rich, 40 kb; essential for muscle differentiation; expressed from the MEN1 locus | Required for the structural integrity of muscle fiber differentiation    | 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]                   |
| p15AS        | ~2.7 kb; enriched in placenta, muscle, and lymphoid tissues                    | Modulates NFAT nuclear trafficking                                       | siRNA knockdown           | [214]                   |
| 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 keratinocyes. | 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 Xist                                                      | 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 | Regulates recruitment of RNA processing factors to, and formation of, nuclear stress bodies | Antisense and siRNA knockdown | [218]                   |
| TERRA / TelRNAs | Various sizes; transcribed from and associated with telomeres; contain UAAAGG repeats | May form G-quartet structures with telomere DNA; inhibit telomerase activity | RNA FISH; oligo-nucleotide modification | [219,220]             |
| Tix          | ~40 kb; antisense to Xist                                                      | Epigenetically silences Xist 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]                   |
| Xist         | ~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 Ilo 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].

*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
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 *BC1*, 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

| 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 |

**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 AK023948, 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 branched2” 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 “callipyge” 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 box) 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) *bad* 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 Polycob-mutant 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 hermaphrodite 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 by-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 non-syntenic 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 Polycand 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 mechanistically, 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 Dm5 (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 β-globin gene induces transcription of the “locus control region” and intergenic regions at the chromosomal β-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 β-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 mRNA 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 that 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 ß-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 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 ADTD susceptibility region) and also within the promoter of an antisense transcript (AS-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 [47,106,158,186,187], 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 ß-globin gene, resulting in ß-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

1. Taft RJ, Pfeasant M, Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29: 288–299.

2. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005) The transcriptional landscape of the mammalian genome. Science 309: 1559–1563.

3. 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: 907–997.

4. Mattick JS, Makkin J (2006) Non-coding RNA. Hum Mol Genet 15: R17–29.

5. Kapranov P, Willingham AT, Gingeras TR (2007) Genome-wide transcription as a function of genome organization. Nat Rev Genet 8: 413–423.

6. Rajendra TK, Prasanth KV, Lakhotia SC (2001) Male sterility associated with transcriptional silencing of 3p12.3 reveals two novel intronic noncoding RNA genes. Genes Chromosomes Evolut 3: 407–412.

7. Sanchez-Elinier T, Gou D, Kremmer E, Sauer F (2006) Noncoding RNAs of S. cerevisiae are essential for structural integrity of nuclear paraspeckles. Proc Natl Acad Sci U S A 103: 155–160.

8. Hellwig S, Bass BL (2008) A starvation-induced noncoding RNA modulates the 5′-end processing of Dicer-regulated genes. Proc Natl Acad Sci U S A 105: 12907–12912.

9. Ma J, Yan B, Qu Y, Qin F, Yang Y, et al. (2008) Zrn401, a short-open reading frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the flower formation in maize. J Cell Biol 180: 105–116.

10. Ben Amor B, Wirth S, Merchan F, Laporte P, A’hernon-Carafa Y, et al. (2009) Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. Genome Res 19: 57–69.

11. Cambilong I, Neggeis F, Kiepferch G, Diepgo G, Stutz F (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell 131: 706–717.

12. Amrall JP, Mattick JS (2008) Noncoding RNA in development. Mamman Geno Sci 19: 454–492.

13. Rina JL, Kertesz M, Wang JK, Squazzo SL, Xu X, et al. (2007) Functional demarcation of active and silent chromatin domains in human H1X loci by noncoding RNAs. Cell 129: 1311–1323.

14. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, et al. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense transcript. Nature 451: 202–206.

15. Mattianov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A (2007) Evidence for selection within long noncoding RNAs. Genome Res 17: 1383–1390.

16. Prasanth KV, Spector DL (2007) Eukaryotic regulatory RNAs: an answer to the functional noncoding RNAs in the human genome. Nat Biotechnol 23: 837–847.

17. Young TL, Matsuda T, Cepko CL (2005) The noncoding RNA taurine-iota promoter regulates RNA transcription in vivo. J Biol Chem 280: 24159–24167.

18. Martinez J, 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.

19. Blin-Wiakack G, Lefot F, Ghoul-Mazzar S, Houton D, Montrieu S, et al. (2000) Endogenous Maxi antisense transcript in vivo and in vitro evidences structural and potential involvement in sleep development in mammals. Proc Natl Acad Sci U S A 107: 7336–7341.

20. Young TL, Matsuda T, Cepko CL (2005) The noncoding RNA taurine-iota promoter regulates RNA transcription in vivo. J Biol Chem 280: 24159–24167.

21. Mutsudhi M, Marshall CM, Benzies KA, Koob MD, Reday R (2004) The spinocerebellar ataxia 8 noncoding RNA causes neurodegeneration and associates with staufen in Drosophila. Curr Biol 14: 302–308.

22. Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschert AV, et al. (2004) Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. Oncogene 23: 6608–6609.

23. Inagaki S, Numata K, Kondo T, Tomita M, Yasuda K, et al. (2005) Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc Natl Acad Sci U S A 102: 5495–5500.

24. 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.

25. Torarinsson E, Hartman SJ, Schroeder DI, 0ueller RP, et al. (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14: 62–66.

26. 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.

27. 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: 9452–9457.

28. Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, 0ueller RP, et al. (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14: 62–66.

29. Martini M, Aimat I, Garber M, French G, Lin MF, et al. (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458: 223–227.

30. Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, 0ueller RP, et al. (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14: 62–66.

31. Reis EM, Ojopi EP, Alberto FL, Rhad 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.

32. Szymanski M, Barciszewska ZM, Erdmann VA, Barciszewski J (2005) A new frontier for molecular medicine: noncoding RNAs. Biochem Biophys Acta 1736: 65–75.

33. Angellou D, ter Elst A, Wei MH, van der Ven AY, Reega EA, et al. (2006) Analysis of a new homozygous deletion in the tumor suppressor region at 3p12.5 reveals two novel intronic noncoding RNA genes. Genes Chromosomes Cancer 43: 676–691.

34. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, et al. (2004) Genome-wide mapping of conserved RNA secondary structures predicts thousands of conserved genomic regions unalignable to human and mouse genomes. Proc Natl Acad Sci U S A 101: 7336–7341.

35. torarinsson E, Hartman SJ, Schroeder DI, 0ueller RP, et al. (2004) An abundance of bidirectional promoters in the human genome. Genome Res 14: 62–66.
84. Ling J, Baibakov B, Pi W, Emerson BM, Tuan D (2005) The HS2 enhancer of
82. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, et al. (2008)
79. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, et al. (2004)
77. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, et al. (2002)
75. Pheasant M, Mattick JS (2007) Raising the estimate of functional human
72. Smith NG, Brandstrom M, Ellegren H (2004) Evidence for turnover of
69. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, et al. (2005)
68. Ogawa Y, Sun BK, Lee JT (2008) Intersection of the RNA interference and X-
67. Stefani G, Slack FJ (2008) Small non-coding RNAs in animal development. Nat
66. Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of
62. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC (2007) Mammalian
59. Aravin AA, Hannon GJ, Brennecke J (2007) The Piwi-piRNA pathway
57. Pollard KS, Salama SR, Kern AD, Kern AD, Dreszer T, et al. (2006) Forces
56. Sone M, Hayashi T, Takeichi M, et al. (2007) The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of
55. Brena C, Chipman AD, Minelli A, Akam M (2006) Expression of trunk Hox
core promoters in the beta-globin locus control region initiates synthesis of non-coding, trans-translation of the human beta-globin locus. Genes Dev 444: 499–502.
54. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312: 276–279.
53. 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: 1299–1298.
52. Carroll SB (2005) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134: 25–36.
51. Smith NG, Brandstrom M, Ellegren H (2004) Evidence for turnover of noncoding DNA in mammalian genome evolution. Genomics 84: 309–317.
50. Frith MC, Pontjavich J, Fredman D, Kai C, Kawai J, et al. (2006) Evolutionary turnover of mammalian transcription start sites. Genome Res 16: 713–722.
49. Taylor MS, Kai C, Kawai J, Carminpi, Hayashizaki Y, et al. (2006) Heterochromatin in mammalian promoter evolution. PLoS Genet 2: e30 doi:10.1371/journal.pgen.0020030.
48. Pheasant M, Mattick JS (2007) Raising the estimate of functional human sequences. Genome Res 17: 1245–1253.
47. Fisher S, Greve EA, Vinton RM, Besling SL, McCallion AS (2006) Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312: 276–279.
46. dermitzakis ET, Reymond A, Ellegren H, et al. (2002) Expression of conserved non-coding sequences on human chromosome 21. Nature 420: 578–582.
45. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
44. CNGs). Science 302: 1033–1035.
43. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
42. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
41. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
40. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
39. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
38. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
37. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
36. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
35. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
34. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
33. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
32. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
31. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
30. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
29. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
28. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
27. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
26. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
25. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
24. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
23. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
22. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
21. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
20. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
19. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
18. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
17. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
16. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
15. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
14. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
13. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
12. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
11. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
10. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
9. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
8. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
7. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
6. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
5. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
4. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
3. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
2. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
1. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420: 578–582.
115. Abeles JF, Kwan KY, O’Reak BJ, Back DW, Stillman AA, et al. (2005) Sequence variants in SRRT1 are associated with Tourette’s syndrome. Science 310: 317–320.

116. Gloz A, Marq F, Takeda H, Piroon 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, Nezár C, Collette C, et al. (2003) A regulatory mutation in RGF2 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) Mosacism of Solid Gold supports the causality of a noncoding A-to-G transition in the determination of the calliphore phenotype. Genetics 163: 435–456.

119. Hagan JP, O'Neill BL, Stewart CL, Kozlov SV, Croce CM (2009) At least ten genes define the imprinted DKL-Dio cluster on mouse chromosome 12FL. PLoS ONE 4: e4352. doi:10.1371/journal.pone.0004352.

120. Georges M, Charlier C, Cocklet N (2003) The calliphore locus: evidence for the trans interaction of reciprocally imprinted genes. Trends Genet 19: 248–252.

121. Wiesener AK, Scherga A, Meyer D, Reichwald S, Durand E, et al. (2008) Preferential reciprocal transfer of paternal/maternal DKL alleles to obes 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) RNA-mediated allelic trans-interaction at the imprinted Rpl10/Pgfl locus. Curr Biol 15: 743–749.

123. He H, Nagy R, Lyranarachchi S, Jiao H, Li W, et al. (2009) A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. Cancer Res 69: 622–631.

124. Duncan I (1987) The bithorax complex. Annu Rev Genet 21: 205–319.

125. Akam ME, Martinez-Arias A (1985) The distribution of Ultrabithorax transcripts in Drosophila embryos. Embryo J 4: 1689–1700.

126. Lipshtiz HD, Portie DA, Hogness DS (1987) Novel transcripts from the Drosophila Willy gene. Science 232: 1717–1720.

127. Zhao J, Sun BK, Erwin JA, Song J, Lee JT (2008) Polychrom protein targets by a short-repetitive repeat to the RNA molecule Xenopus. Science 322: 750–756.

128. Sanchez-Herrero E, Akam M (1989) Spatially ordered trascription of regulatory DNA in the bithorax complex of Drosophila. Development 107: 309–325.

129. Duncan IW (2002) Transvection effects in Caenorhabditis elegans. Annu Rev Genet 36: 1031–1050.

130. 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: 304–391.

131. Jupe ER, Liu XT, Kiehlbauch JL, McClung JK, Dell’Orco RT (1996) The 3 untranslated region of prohibitin and cellular immortalization. Exp Cell Res 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 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Cell 75: 843–854.

134. Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 855–862.

135. Reinhart BJ, Slack PJ, Basson M, Pasquonelle AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature 403: 901–906.

136. Hiepfl DR, Weigmann K, Cohen SM (2003) The bantam gene regulates Roush S, Slack FJ, Pedersen A, Dufour J, Jørgensen N (2008) The let-7 family of microRNAs. Trends Cell Biol 18: 53–60.

137. Marcellini S, Simpson P (2006) Two or four bristles: functional evolution of an enhancer of scute in Drosophila. PLoS Biol 4: e586. doi:10.1371/journal.pbio.0040586.

138. Sipos L, Mihaly J, Karch F, Schell P, Gauja J, et al. (1998) Translocation in the Drosophila Igl-D domain: extensive upstream sequences are involved in anchoring distal cis-regulatory regions to the promoter. Genetics 149: 1031–1050.

139. Hendrickson JE, Sakonju S (1995) cis and trans interactions between the Iab regulatory regions and Iab-A and Iab-B in Drosophila melanogaster. Genetics 139: 833–848.

140. 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.

141. Moulal M, Hagstrom K, Grynkiewicz H, Pirrotta V, Schell P (1999) The seq element from the Drosophila melanogaster bithorax complex mediates long-distance regulatory interactions. Genetics 135: 1333–1356.

142. van der Winden J, Matzke MA, Matzke AJ (2009) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J 19: 5194–5201.

143. Mattick JS, Gagen MJ (2001) The evolution of controlled multiskilled gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol Biol Evol 18: 150–160.

144. Feng J, Bi C, Clark BS, Marky R, Shah P, et al. (2006) The Es-f2 noncoding RNA is transcribed from the Dvl-5/7e ultraconserved region and functions as a Dvl-2 transcriptional coactivator. Genes Dev 20: 1470–1481.

145. Pandey RR, Mondal T, Mohammad F, Eroukh S, Redrup L, et al. (2006) Kcrip177 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32: 232–246.

146. Van Laere AS, Nguyen M, Braunschweig M, Nezár C, Collette C, et al. (2003) The Air noncoding RNA epigenetically silences transcription by targeting G9a. Development 130: 1219–1229.

147. Maruccini P, Simpson P (2006) Two or four bristles: functional evolution of an enhancer of scute in Drosophila. PLoS Biol 4: e586. doi:10.1371/journal.pbio.0040586.
174. Pal-Bhadra M, Bhadra U, Birchler JA (2002) RNAi related mechanisms affect both transcriptional and posttranscriptional gene silencing in Drosophila. Mol Cell 9: 313–327.

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, Lebovits 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. Bremecue J, Malon 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 intranucleosomal positioning in monolayer cell gene expression. Cell 128: 777–786.

181. Sasaki H, Matsui Y (2008) Epigenetic events in mammalian germ-cell development: programming and reprogramming. Nat Rev Genet 9: 129–140.

182. Chandler VL (2007) Paramutation: from maize to mice. Cell 128: 641–645.

183. Cuzin F, Grandjean V, Russoladzean M (2008) Inherited variation at the epigenetic level: paramutation from the plant to the mouse. Curr Opin Genet Dev 18: 193–196.

184. Costa NF (2008) Non-coding RNAs, epigenetics and complexity. Gene 410: 9–17.

185. Osborne RJ, Thornton CA (2007) Paramutation: from maize to mice. Cell 128: 641–645.

186. Ishii N, Ozaki K, Sato H, Mizuno H, Saito S, et al. (2006) Identification of a novel bone morphogenetic protein-responsive gene that may function as a novel anti-malarial drug target. J Biol Chem 281: 28323–28332.

187. Dallosso AR, Hancock AL, Malik S, Salpekar A, King-Underwood L, et al. (2005) A novel, non-coding RNA gene, regulates eosinophil granule RNA CUDR. RNA 13: 890–898.

188. Thakur N, Tiwari VK, Thomassin H, Pandey RR, Kanduri M, et al. (2004) An antisense RNA regulates the bidirectional silencing property of the King/ imprinting control region. Mol Cell Biol 24: 7855–7862.

189. Fujita N, Matsui Y, Ishii N, Ozaki K, Sato H, et al. (2006) Identification of a novel non-coding RNA directed DNA demethylation of \(K\)-CpG island. Biochem Biophys Res Commun 343: 593–599.

190. Matouk IJ, DeGroot N, Mezan S, Abu-lail R, et al. (2007) The H19 antisense RNA regulates muscle development and promotes cell proliferation in bladder carcinoma. J Biol Chem 282: 17079–17085.

191. Osbourne RJ, Thornton CA (2006) RNA-dominant diseases. Hum Mol Genet 15: R162–R169.

192. Angeloni D, Ter Elst A, Wei MH, van den Veen AJ, Braga EA, et al. (2005) Analysis of a new homoygous deletion in the tumor suppressor region at 5p12.3 reveals two novel intronic noncoding RNA genes. Genes Chromosomes Cancer 45: 676–679.

193. 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: 2297–2299.

194. 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.

195. Sado T, del Gaudio D, German JR, Shinawi M, Prens U, et al. (2008) Prader-Willi syndrome phenotype caused by maternal deficiency for the H19-5UTR box small nuclear RNA cluster. Nat Genet 40: 719–721.

196. Zhou Y, Zhong Y, Wang Y, Zhang X, Barista DL, et al. (2007) Activation of p53 by MEG3 non-coding RNA. J Biol Chem 282: 24731–24742.

197. Tanaka R, Sato H, Moriyama M, Saito K, Morishita Y, et al. (2000) Identification of a novel bone morphogenetic protein-responsive gene that may function as a novel anti-malarial drug target. J Biol Chem 281: 28323–28332.

198. Fuji T, Ravidrannath L, Tran P, 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.

199. Talevi T, Morita K, Shigemori M, Tashiro Y, et al. (2004) Anti-malarial drug resistance and transformation in human cancer cells by the noncoding RNA CUDR. Rna 10: 228–236.