Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Review

The plasticity of the mammalian transcriptome

Johan Lindberg *, Joakim Lundeberg

School of Biotechnology, Department of Gene Technology, AlbaNova University Center, Royal Institute of Technology, Stockholm, Sweden

A B S T R A C T

The dogmatic view of RNA as a mere necessity in the transfer of information between DNA and proteins has during recent years come into question. Novel approaches and new technology has revealed an unprecedented level of inherent complexity in the mammalian transcriptome. Here, the majority of nucleotides are expressed, in sharp contrast to the ~1.2% of the human genome harboring protein coding information. Also, >50% of genomic loci contain antisense and interleaved transcription, a conservative estimate since non-coding RNA is highly regulated between tissues and developmental stages, which has only been investigated to a limited extent. Subsequent focus on RNA with no coding potential has revealed numerous species with novel functions, and deep sequencing studies imply that many remain to be discovered. This review gives an overview of the plasticity and dynamics of the mammalian transcriptome and the prevailing interpretation of its effect on the complexity of species.

© 2009 Elsevier Inc. All rights reserved.

Contents

The transcriptome

Our view of genomes, corresponding transcriptomes and the flow of genetic information as described by the central dogma, have long been characterized by simple prokaryotic models. Here, the information is passed from DNA via messenger RNA (mRNA) to proteins, a process controlled by flanking regulatory 5′/3′ sequences working in concert with regulatory proteins. In addition, the genetic output is considered to be almost exclusively transacted by proteins, which lead to the oversimplified meaning of the term "gene" as a section of the DNA whose corresponding RNA encodes a protein. In the light of these assumptions the mammalian transcriptome was considered to consist mainly of ribosomal RNA (~80%, rRNA), transfer RNA (~15%, tRNA), mRNA (2–4%) and a small proportion of non-coding RNA (~1%, ncRNA) with regulatory functions. This model was extrapolated to higher multicellular organisms, despite conflicting reports on the composition of polyadenylated (poly A+) and non-polyadenylated (poly A−) RNA more than 30 years ago [1–3]. Moreover, the

indisputable correlation between proportions of non-coding DNA (ncDNA) and complexity (% ncDNA: prokaryotes <0.25%; simple eukaryotes 25–50%; more complex fungi and animals >50%; humans ~98.8%) [4] suggested that the common opinion of ncDNA as junk or evolutionary debris was fundamentally flawed.

It was not until 2002, when tiling arrays were introduced, allowing high-resolution, high-throughput scanning over large genomic regions that the extent of mammalian transcription was starting to be recognized [5]. Later, the same year, a collection of ~61,000 sequenced full-length cDNA clones of mRNA in mice combined with public sequence resources were analyzed and published by the FANTOM consortium [6]. Surprisingly, ~47% of ~30,000 identified transcriptionally active loci did not harbor protein coding potential and were identified to represent novel, potentially functional ncRNAs. These findings were verified by Bertone et al. [7], where triple selected poly A+ RNA from a pool of liver tissues was investigated with tiling arrays. Except for 64% (~11,000) of presently annotated RefSeq [8] genes, ~10,500 distinct regions located outside annotated exons were found to be transcriptionally active. In two subsequent studies by the FANTOM consortium, various cloning approaches were applied to retrieve, e.g., ~180,000 transcripts with paired initiation and termination boundaries in combination with public resources to

* Corresponding author. Fax: +41 468 314975.
E-mail addresses: johlin@kth.se (J. Lindberg), joakiml@biotech.kth.se (J. Lundeberg).

characterize the nature of the mammalian transcriptome [9,10]. Their analysis revealed a multifaceted, widespread, overlapping transcriptional landscape, which called for new definitions. When clustering transcriptional frameworks (overlapping transcripts sharing a splicing event, start or termination site) into transcriptional forests (non-interrupted transcriptionally active regions on either strand) it comprised ~63% of the genome with interspersed transcriptional deserts. ~65% of transcriptional units (overlapping mRNA which share direction and strand) were alternatively spliced which commonly affected domain content or organization and ~72% contained antisense transcription. Analysis of CAGE (cap analysis of gene expression, allows for 5′ end identification of transcripts) tags [11] revealed an overrepresentation of annotated transcripts initiating not only from the 5′ end, but also from the 3′ end. In addition, ~1/3 of all full-length cDNA did not appear to harbor any protein coding sequence, and many were only detected once, indicating low levels of expression. Although ncRNA were on average less conserved than 5/3′ UTRs (untranslated regions) or coding sequences, the relationship was surprisingly opposite for promoter regions, which may suggest functionality.

Another level of inherent complexity was added to prevailing model by Cheng and colleagues [12]. Here, tiling arrays, sequencing of reversely transcribed transcripts and real-time PCR was used to investigate poly A± RNA from the cytosol and nucleus separately. In total, ~44% of expressed nucleotides were poly A−, ~19% poly A+ and ~37% bimorphic (either poly A+ or poly A−). These patterns were further investigated by sequencing, since bimorphic transcripts could potentially arise from either strands, and the results showed that ~50% of investigated loci harbored antisense transcription. The nucleus was found to contain ca. five times the number of expressed nucleotides than the cytosol and of all transcripts, ~50% were unique for the nucleus, ~10% for the cytosol and ~40% were present in both. Subsequently in June 2007 the ENCODE pilot project was published [13].

ENCODE is a public research consortium with the aim to identify all functional elements in the human genome. The pilot project covered 1% of the human genome and although ~15% of interrogate nucleotides were found to be transcribed in one tissue sample ~93% of all bases were represented in a primary transcript according to at least two independent observations. ~40% of all identified transcripts were present in only one sample, and only ~2% were present in all samples. These findings confirm the extent of differential transcription in different tissues [12,14] and indicate a regulatory role. Amplification of 5′ ends was used, in combination with tiling arrays, to investigate the diversity of transcripts emanating from protein-coding loci. 90% of interrogated loci encoded transcripts with novel exons 50–100 kb upstream of the annotated TSSs (transcription start sites) and ~50% contained at least one exon from an upstream gene. On average, 5.4 transcripts were detected per locus. ~19% of pseudogenes were found to be transcriptionally active which concurs with previous work by the FANTOM consortium, where a high proportion (~10%) of all detected transcripts mapped to pseudogenes [15]. The correlation between queried protein factors associated with histones and DNaseI hypersensitive sites verified the pattern of unprecedented transcription. The proportions of novel TSSs associated with clusters of histone modifications and DNaseI hypersensitive sites were larger than for previously known and identified TSSs. Note, the known and novel TSSs contained the same histone modifications, indicating that they were regulated by the same processes. Observations in other species, such as yeast, worm and fruit fly confirm the pervasive (70–85%) transcriptional patterns [16–18]. Taken together, the studies from the FANTOM consortium, ENCODE and others have caused the gene definition to come under scrutiny as discussed by Gerstein and colleagues [19]. Here, focus was set on function and an updated definition is suggested to be "a union of genomic sequences encoding a coherent set of potentially overlapping functionally

products." This basically projects the functional products down to overlapping sequences and excludes regulatory sequences. Since regulatory sequences have a many-to-many relationship these are too complex to be included in the definition and are instead suggested to be defined in a class of "gene-associated sequences."

During 2008, new technology [20] has increased the throughput of sequencing to the extent where it is becoming feasible to sequence transcriptomes instead of using hybridization-based approaches [21]. Targeting the polyadenylated component of the transcriptome, ~35% mapped to intergenic regions [22,23]. Also, by sequencing RNA from transcriptionally active polymerases of primary human lung fibroblasts, 68% of RefSeq [8] genes were found to be active and in 59% antisense transcription was detected [24]. All-in-all, initial results using next-generation-sequencing technology support and bring novel knowledge on the current model of widespread interleaved, complex transcription in mammalian genomes. E.g., recently, Faulkner et al. [25] investigated the retrotransposon transcriptome, with focus on repetitive elements in the human genome. ~31% of detected TSS mapped to repetitive elements whose expression varied substantially between tissues. The majority retrotransposon transcripts initiated in novel promoters with a sharp dominant peak, bearing similarity to TATA-box controlled genes, which to a great extent are regulated between tissues and developmental stages [14]. Interestingly, a subset of detected retrotransposon were found to provide alternative promoters for Refseq genes [8].

Also, deep sequencing has allowed for unprecedented mapping of splice variants, which drastically increase the estimates of the proportion of alternatively spliced genes. 40–60% of all human genes [26] and >70% of all multi-exon genes have previously been suggested to be alternatively spliced [27] but Wang and colleagues demonstrated that ~93% of multi-exon genes are subject to alternative splicing and 52–80% of all alternative-splicing events are regulated between tissues depending on the type of splicing event [28]. In addition, genome-wide interrogation of RNA editing [29,30], an event leading to the conversion of, e.g., RNA adenosine to inosine, which is interpreted by the ribosome as guanosine, is now being made possible with next generation sequencing technology. In a recent study, Li and colleagues [31] used padlock probes [32] to sequence putative editing sites in DNA and RNA from several different tissues. This revealed hundreds of edited sites, many with an editing level as low as >2%, which is in stark contrast to the previously 13 known edited sites.

What is the role of all this non-coding RNA?

What is all this RNA used for? ncRNAs have been suggested to function as a digital information system with properties that allow organisms to surmount complexity limits imposed by using only regulatory proteins in conjunction with cis-regulatory elements. The key problems associated with complex protein networks lie in evolving systems capable of controlling them with the smallest possible amount of resources. The increasing sophistication of control architecture required as complexity increases is demonstrated by the connection between the numbers of regulatory proteins and genome sizes in prokaryotes, which scales quadratically. To overcome such limitations a parallel, overlying control system is needed, which could be provided by ncRNAs [33]. Furthermore, observations that organisms such as Caenorhabditis elegans and Strongylocentrotus purpuratus (sea urchin) host almost the same numbers of protein-coding genes (~19,000 and 23,000, respectively) as humans (~20,500), suggesting that the regulatory systems for higher complex organisms must lie beyond protein networks. Alternative splicing has been proposed to explain the surprisingly low number of genes in the human genome [34,35], since Saccharomyces cerevisiae (for instance) has virtually no introns [36]. However, in a comparison of seven species, the proportion of alternatively spliced genes was not found to be higher

in humans [37], although the outcome of the study has been questioned [38].

The complexity of mammalian transcriptomes is provocative and has catalyzed a discussion whether the pervasive transcription in mammalian transcriptomes is mainly noise [39] noise or not. Importantly, several studies have shown that long ncRNA (lRNA) does play a regulatory role by establishing that the pervasive ncRNA expression is differentially expressed between tissues, cell types and developmental stages [9,12–14,40–42]. In addition, lRNAs are subject to alternative splicing [43], localize to subcellular compartments [44], serve as structural components [45,46] and are strongly overrepresented in differentially methylated regions between tissues, which also argue for a role in regulation [47]. In a study worth attention, Guttman et al. used chromatin-state maps in four mouse cell types which identified ~1600 putatively functional lRNAs where ~95% displayed evolutionary conservation [48]. These lRNA appeared to be multi-exonic and to contain both poly A tail and 5′ cap. Subsequent analysis revealed these lRNA to be involved in various processes, e.g., cell proliferation, immune surveillance, morphogenesis, ESC pluripotency and muscle development, which was experimentally verified for a subset. The developmental role is in line with work by Efroni and coworkers [49] demonstrating that embryonic stem cells are transcriptionally hyperactive on a global scale, which is reduced after differentiation. Furthermore, the diversity of a species of small RNA (sRNA), microRNA (miRNA), has been shown to correlate to phenotypic complexity [50] and recent work, as exemplified in *S. cerevisiae*, demonstrates that chromatin-remodeling complexes actively regulate the chromatin state via mechanisms that minimize inappropriate transcription [51].

In contrast, the estimation that only ~5% of the human genome is under purifying selection conflicts with the idea of a functional regulatory system mediated by ncRNA [13,52]. These numbers might seem paradoxical since sequences under no selective pressure should theoretically acquire functions or be lost. In accordance with this assumption is growing evidence that ancient and modern repeats act as factories for gene promoters, genes, regulatory elements, microRNA, etc., and newly detected classes of repeats with variable sequence conservation, some of which are among the slowest evolving entities in the human genome [53–59]. It is important to note that ncRNA regulatory signals are not bound by the same structure–function and multi-tasking relationships as proteins, and can therefore exhibit faster evolutionary rates, which has previously been discussed in great detail [60]. Collectively, these findings and the growing flood of functional ncRNAs [61–63] show that the enormous flows of information required during the development of complex organisms such as mammals are likely to be mediated by RNA, that proteins in conjunction with RNA execute the program and proteins alone mediate the interaction with the environment [62].

Regulatory RNA

The mammalian transcriptome consists of many types of non-coding regulatory RNA. Examples are small nuclear RNA (snRNA), transcripts ~200 nucleotides long, that guide the spliceosome complex and small nucleolar RNAs (snoRNAs), 60–300 nucleotides long RNA molecules, directing associated proteins to execute sequence-specific modifications (methylation, pseudouridylation) on tRNA, rRNA or snRNA molecules in the nucleolus. Interestingly, alternative functions have been suggested for snoRNA, e.g., regulation of alternative splicing by binding to silencer elements, [64] or as precursors for miRNA [65]. sRNAs are the most recent additions to the RNA family, Only <30 nucleotides long, which hindered an earlier discovery. piRNA (piwi-interacting RNA), discovered in *Drosophila* [66], are 24–30 nt poorly conserved sRNA interacting with Piwi proteins [67]. These have been associated with control of transpos-

able elements and normal spermatogenesis in mammals [68,69]. Exogeneous siRNA (small interfering RNA) have been widely used for gene silencing in laboratory settings by using the miRNA associated Risc complex but so far the only endogenous class of gene regulating sRNA that have been well characterized in humans is miRNA [70–72].

Mammalian miRNA, recently reviewed here [73], regulate gene expression as an integrated part of the RISC complex [74], which leads to inhibition of translation and destabilization of the target transcript, although the mechanistic details are still under debate and there are rare exceptions where slicing occurs [75,76]. Each broadly conserved mammalian miRNA targets hundreds of sites and >60% of all human protein coding genes are conserved targets of miRNA [77]. Interestingly, it was recently demonstrated that activation of T-cells causes a switch in expression in favor of transcript isoforms with shorter 3′ UTRs [78]. For some transcripts investigated in detail, the shortening of the 3′ UTR relieved negative miRNA regulation. Also, although the canonical miRNA target sites resides within the 3′UTR, simultaneous targeting of the 5′UTR increases the effect on the translational output [79]. Many miRNAs are differentially expressed between tissues and human cells transfected with tissue specific miRNA will change the expression profile to resemble the tissue from which the miRNA originated [80,81]. During 2008, the ability of individual miRNA to regulate hundreds of genes simultaneously was verified at the protein level by quantitative mass spectrometry, although levels of most proteins were only modestly changed [82,83]. This, however, does not mean unimportance for development and disease. Several different mutant mice carrying miRNA mutations display initial normal viability to subsequently reveal severe phenotypic abnormalities [84]. To date, >600 human miRNAs have been registered in the miRBase (Release 12.0) database [85].

lRNA have attracted a lot of attention after the extent of pervasive transcription was established for the mammalian genomes [5–7,9,10,12,13] and the number of known functional lRNA is rapidly increasing, as recently reported in two excellent reviews [86,87]. Briefly, lRNA has been linked to gene regulation as an integral component of chromatin [88,89] also affecting DNA methylation [90]. Subsequently, antisense lRNA was shown to regulate the methylation state of a tissue-dependent differentially regulated region [91] and to change the epigenetic states of promoter regions, leading to silencing of the affected genes [92,93]. lRNA is also involved in dosage compensation. The best-known example of this is X-chromosome inactivation, in which the transcriptional activity of large regions of one X-chromosome is silenced in human females. XIST, a 17-kb ncRNA, physically binds to large regions of the inactive X-chromosome, causing heterochromatin formation [94]. A subset of autosomal genes (~150) are only expressed on one allele due to regulatory lRNA, which controls the methylation status of a region denoted ICE (imprint control region) [95].

Kapranov and colleagues mapped poly A+ lRNA in both the cytosol and nucleus to sRNAs from the same cells [96]. In the cited study, 2–10% of lRNA were identified as possible precursors for sRNAs based on overlapping transcription and conservation patterns. These sRNAs mapped mainly to 5′ and 3′ regions of known genes, 22–200 nt long and was denoted PASRs (promoter-associated RNAs) and TASRs (termini-associated RNAs). A large proportion of PASRs/TASRs mapped into syntenic regions between human and mouse and correlated with the expression of their corresponding genes. ~45% of all expressed genes harbored PASRs. Subsequently, PASRs were investigated in more detail in an attempt to map transcripts <200 nt in length in two human cell lines [97]. PASR expression patterns and positions correlated to a high degree with CAGE tags [98] of lRNA. sRNA and CAGE tags were ten times more likely to occur in exonic than intronic regions and although PASRs could be the output of independent transcription CAGE tag often bridged splice junctions with the 5′ end being <20 nt away. This makes it unlikely for such a

CAGE tag represents an independent transcript and suggests that they could be the result of a novel processing pathway of lRNA generating small, capped ncRNAs. Synthetic sense and antisense PASRs were demonstrated to reduce the expression of its corresponding mRNA, which implies a possible regulatory function of these sRNAs. Another novel short species of RNA was recently characterized by Taft and colleagues [99] denoted transcription initiation RNAs (tiRNA). tiRNA were predominantly 18 nucleotides in length, 190-fold strand-specifically enriched at TSS and found at >20% of expressed genes. Unlike PASRs the proportion of tiRNA did not increase when sequencing RNA enriched for 5′ end CAP structures. Although genes with tiRNA hosting promoters were on average more highly expressed, the tiRNA expression did not correlate to the abundance of its corresponding mRNA ($R < 0.05$).

Future perspectives — where do we go from here?

It is beyond reasonable doubt that the complexity of the mammalian transcriptome far exceeds what was expected only a decade ago. Different studies, e.g., deep sequencing [9,100] or mapping of secondary structure elements in the genome [101] suggest that many long and short non-coding RNAs still await discovery, especially taxon specific species [102]. It may very well be regulatory networks of RNA that convey the information required to control development and thereby the phenotypes of the ~100 trillion cells that make up the human body. The implications of its recently demonstrated complexity remain to be determined, especially in different disease contexts. It is important to note that even homogeneous population of cells will have different transcriptomes [103–105]. The stochastic variability of gene expression can lead to sub-phenotypes in genetically identical populations, which has been demonstrated to increase fitness in fluctuating environments [106] and to control lineage choices of mammalian progenitor cells [107]. Non-genetic variability was recently suggested by Brock and colleagues [108] to be a driving force in cancer development. By reason, the realization that the mammalian genomes can generate a large diversity of cell phenotypes, such as liver and brain cells with the same underlying genome, places the almost dogmatic somatic mutation theory of cancer under scrutiny.

The continuous increase in read-length and throughput by the next-generation of sequencers will aid in understanding the complex gene regulatory networks, especially due to the distribution of the mammalian transcriptome, where few transcripts are highly expressed [109]. But although it is becoming possible to retrieve a snapshot of the transcriptome, it is still not clear how to make best use of the massive amount of information retrieved by gene expression assays. Recently a promising approach was undertaken where an integrated analysis of genetic variation and gene expression assays revealed relevant disease signatures [110,111]. The parallel use of linkage- and gene expression data allowed for a direct connection between genotype and phenotype, a prerequisite to distinguish the variation that is causing a phenotype from the variation that is caused by the phenotype. This rationale has also been successfully applied on tumors from different forms of cancer [112,113]. Collectively, these and other studies demonstrate that variation in the transcriptome, the genome and the epigenome is best considered simultaneously in order to unlock their true potential to contribute to the understanding of disease pathogenesis and etiology.

Acknowledgments

References

[1] P.K. Katinakis, A. Slater, R.H. Burdon, Non-polyadenylated mRNAs from eukaryotes, FEBS Lett. 116 (1980) 1–7.

[2] C. Milcarek, R. Price, S. Penman, The metabolism of a poly(A) minus mRNA fraction in HeLa cells, Cell 3 (1974) 1–10.

[3] M. Salditt-Georgieff, M.M. Harpold, M.C. Wilson, J.E. Darnell, Jr., Large heterogeneous nuclear ribonucleic acid has three times as many 5′ caps as polyadenylic acid segments, and most caps do not enter polyribosomes, Mol. Cell. Biol. 1 (1981) 179–187.

[4] R.J. Taft, M. Pheasant, J.S. Mattick, The relationship between non-protein-coding DNA and eukaryotic complexity, Bioessays 29 (2007) 288–299.

[5] P. Kapranov, S.E. Cawley, J. Drenkow, S. Bekiranov, R.L. Strausberg, S.P. Fodor, et al., Large-scale transcriptional activity in chromosomes 21 and 22, Science 296 (2002) 916–919.

[6] Y. Okazaki, M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, N. Osato, R. Saito, H. Suzuki, I. Yamanaka, H. Kiyosawa, K. Yagi, Y. Tomaru, Y. Hasegawa, A. Nogami, C. Schonbach, T. Gojobori, R. Baldarelli, D.P. Hill, C. Bult, D.A. Hume, J. Quackenbush, L.M. Schriml, A. Kanapin, H. Matsuda, S. Batalov, K.W. Beisel, J.A. Blake, D. Bradt, V. Brusic, C. Chothia, L.E. Corbani, S. Cousins, E. Dalla, T.A. Dragani, C.F. Fletcher, A. Forrest, K.S. Frazer, T. Gaasterland, M. Gariboldi, C. Gissi, A. Godzik, J. Gough, S. Grimmond, S. Gustincich, N. Hirokawa, I.J. Jackson, E.D. Jarvis, A. Kanai, H. Kawaji, Y. Kawasawa, R.M. Kedzierski, B.L. King, A. Konagaya, I.V. Kurochkin, Y. Lee, B. Lenhard, P.A. Lyons, D.R. Maglott, L. Maltais, L. Marchionni, L. McKenzie, H. Miki, T. Nagashima, K. Numata, T. Okido, W.J. Pavan, G. Pertea, G. Pesole, N. Petrovsky, R. Pillai, J.U. Pontius, D. Qi, S. Ramachandran, T. Ravasi, J.C. Reed, D.J. Reed, J. Reid, B.Z. Ring, M. Ringwald, A. Sandelin, C. Schneider, C.A. Semple, M. Setou, K. Shimada, R. Sultana, Y. Takenaka, M.S. Taylor, R.D. Teasdale, M. Tomita, R. Verardo, L. Wagner, C. Wahlestedt, Y. Wang, Y. Watanabe, C. Wells, L.G. Wilming, A. Wynshaw-Boris, M. Yanagisawa, et al., Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs, Nature 420 (2002) 563–573.

[7] P. Bertone, V. Stolc, T.E. Royce, J.S. Rozowsky, A.E. Urban, X. Zhu, et al., Global identification of human transcribed sequences with genome tiling arrays, Science 306 (2004) 2242–2246.

[8] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 33 (2005) D501–D504.

[9] P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M.C. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, R. Kodzius, K. Shimokawa, V.B. Bajic, S.E. Brenner, S. Batalov, A.R. Forrest, M. Zavolan, M.J. Davis, L.G. Wilming, V. Aidinis, J.E. Allen, A. Ambesi-Impiombato, R. Apweiler, R.N. Aturaliya, T.L. Bailey, M. Bansal, L. Baxter, K.W. Beisel, T. Bersano, H. Bono, A.M. Chalk, K.P. Chiu, V. Choudhary, A. Christoffels, D.R. Clutterbuck, M.L. Crowe, E. Dalla, B.P. Dalrymple, B. de Bono, G. Della Gatta, D. di Bernardo, T. Down, P. Engstrom, M. Fagiolini, G. Faulkner, C.F. Fletcher, T. Fukushima, M. Furuno, S. Futaki, M. Gariboldi, P. Georgii-Hemming, T.R. Gingeras, T. Gojobori, R.E. Green, S. Gustincich, M. Harbers, Y. Hayashi, T.K. Hensch, N. Hirokawa, D. Hill, L. Huminiecki, M. Iacono, K. Ikeo, A. Iwama, T. Ishikawa, M. Jakt, A. Kanapin, M. Katoh, Y. Kawasawa, J. Kelso, H. Kitamura, H. Kitano, G. Kollias, S.P. Krishnan, A. Kruger, S.K. Kummerfeld, I.V. Kurochkin, L.F. Lareau, D. Lazarevic, L. Lipovich, J. Liu, S. Liuni, S. McWilliam, M. Madan Babu, M. Madera, L. Marchionni, H. Matsuda, S. Matsuzawa, H. Miki, F. Mignone, S. Miyake, K. Morris, S. Mottagui-Tabar, N. Mulder, N. Nakano, H. Nakauchi, P. Ng, R. Nilsson, S. Nishiguchi, S. Nishikawa, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.

[10] S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, et al., Antisense transcription in the mammalian transcriptome, Science 309 (2005) 1564–1566.

[11] R. Kodzius, M. Kojima, H. Nishiyori, M. Nakamura, S. Fukuda, M. Tagami, et al., CAGE: cap analysis of gene expression, Nat. Methods 3 (2006) 211–222.

[12] J. Cheng, P. Kapranov, J. Drenkow, S. Dike, S. Brubaker, S. Patel, et al., Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution, Science 308 (2005) 1149–1154.

[13] E. Birney, J.A. Stamatoyannopoulos, A. Dutta, R. Guigo, T.R. Gingeras, E.H. Margulies, Z. Weng, M. Snyder, E.T. Dermitzakis, R.E. Thurman, M.S. Kuehn, C.M. Taylor, S. Neph, C.M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J.A. Greenbaum, R.M. Andrews, P. Flicek, P.J. Boyle, H. Cao, N.P. Carter, G.K. Clelland, S. Davis, N. Day, P. Dhami, S.C. Dillon, M.O. Dorschner, H. Fiegler, P.G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K.D. James, B.E. Johnson, E.M. Johnson, T.T. Frum, E.R. Rosenzweig, N. Karnani, K. Lee, G.C. Lefebvre, P.A. Navas, F. Neri, S.C. Parker, P.J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F.S. Collins, J. Dekker, J.D. Lieb, T.D. Tullius, G.E. Crawford, S. Sunyaev, W.S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A. Sandelin, I.L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H.A. Hirsch, E.A. Sekinger, J. Lagarde, J.F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J.S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M.C. Dickson, D.J. Thomas, M.T. Weirauch, J. Gilbert, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (2007) 799–816.

[14] P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, et al., Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635.

[15] M.C. Frith, L.G. Wilming, A. Forrest, H. Kawaji, S.L. Tan, C. Wahlestedt, et al., Pseudo-messenger RNA: phantoms of the transcriptome, PLoS Genet. 2 (2006) e23.

[16] H. He, J. Wang, T. Liu, X.S. Liu, T. Li, Y. Wang, et al., Mapping the C. elegans noncoding transcriptome with a whole-genome tiling microarray, Genome Res. 17 (2007) 1471–1477.

[17] L. David, W. Huber, M. Granovskaia, J. Toedling, C.J. Palm, L. Bofkin, et al., A high-resolution map of transcription in the yeast genome, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5320–5325.

[18] J.R. Manak, S. Dike, V. Sementchenko, P. Kapranov, F. Biemar, J. Long, et al., Biological function of unannotated transcription during the early development of Drosophila melanogaster, Nat. Genet. 38 (2006) 1151–1158.

[19] M.B. Gerstein, C. Bruce, J.S. Rozowsky, D. Zheng, J. Du, J.O. Korbel, et al., What is a gene, post-ENCODE? History and updated definition, Genome Res. 17 (2007) 669–681.

[20] E.R. Mardis, Next-generationDNA , sequencing methods, Annu. Rev. Genomics Hum. Genet. 9 (2008) 387–402.

[21] J. Shendure, The beginning of the end for microarrays? Nat. Methods 5 (2008) 585–587.

[22] N. Cloonan, A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, et al., Stem cell transcriptome profiling via massive-scale mRNA sequencing, Nat. Methods 5 (2008) 613–619.

[23] M. Sultan, M.H. Schulz, H. Richard, A. Magen, A. Klingenhoff, M. Scherf, et al., A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome, Science 321 (2008) 956–960.

[24] L.J. Core, J.J. Waterfall, J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, Science 322 (2008) 1845–1848.

[25] G.J. Faulkner, Y. Kimura, C.O. Daub, S. Wani, C. Plessy, K.M. Irvine, et al., The regulated retrotransposon transcriptome of mammalian cells, Nat. Genet. 41 (2009) 563–571.

[26] B. Modrek, C. Lee, A genomic view of alternative splicing, Nat. Genet. 30 (2002) 13–19.

[27] J.M. Johnson, J. Castle, P. Garrett-Engele, Z. Kan, P.M. Loerch, C.D. Armour, et al., Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays, Science 302 (2003) 2141–2144.

[28] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, et al., Alternative isoform regulation in human tissue transcriptomes, Nature 456 (2008) 470–476.

[29] K. Nishikura, C. Yoo, U. Kim, J.M. Murray, P.A. Estes, F.E. Cash, et al., Substrate specificity of the dsRNA unwinding/modifying activity, EMBO J. 10 (1991) 3523–3532.

[30] A.G. Polson, B.L. Bass, Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase, EMBO J. 13 (1994) 5701–5711.

[31] J.B. Li, E.Y. Levanon, J.K. Yoon, J. Aach, B. Xie, E. Leproust, et al., Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing, Science 324 (2009) 1210–1213.

[32] G.J. Porreca, K. Zhang, J.B. Li, B. Xie, D. Austin, S.L. Vassallo, et al., Multiplex amplification of large sets of human exons, Nat. Methods 4 (2007) 931–936.

[33] J.S. Mattick, RNA regulation: a new genetics? Nat. Rev. Genet. 5 (2004) 316–323.

[34] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., The sequence of the human genome, Science 291 (2001) 1304–1351.

[35] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921.

[36] C.A. Davis, L. Grate, M. Spingola, M. Ares, Jr., Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast, Nucleic Acids Res. 28 (2000) 1700–1706.

[37] D. Brett, H. Pospisil, J. Valcarcel, J. Reich, P. Bork, Alternative splicing and genome complexity, Nat. Genet. 30 (2002) 29–30.

[38] H. Kim, R. Klein, J. Majewski, J. Ott, Estimating rates of alternative splicing in mammals and invertebrates, Nat. Genet. 36 (2004) 915–916 author reply 916-7.

[39] K. Struhl, Transcriptional noise and the fidelity of initiation by RNA polymerase II, Nat. Struct. Mol. Biol. 14 (2007) 103–105.

[40] T. Ravasi, H. Suzuki, K.C. Pang, S. Katayama, M. Furuno, R. Okunishi, et al., Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome, Genome Res. 16 (2006) 11–19.

[41] T.R. Mercer, M.E. Dinger, S.M. Sunkin, M.F. Mehler, J.S. Mattick, Specific expression of long noncoding RNAs in the mouse brain, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 716–721.

[42] J.L. Rinn, M. Kertesz, J.K. Wang, S.L. Squazzo, X. Xu, S.A. Brugmann, et al., Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, Cell 129 (2007) 1311–1323.

[43] K.C. Pang, S. Stephen, P.G. Engstrom, K. Tajul-Arifin, W. Chen, C. Wahlestedt, et al., RNAdb—a comprehensive mammalian noncoding RNA database, Nucleic Acids Res. 33 (2005) D125–D130.

[44] J.E. Wilusz, S.M. Freier, D.L. Spector, 3′ end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA, Cell 135 (2008) 919–932.

[45] Y.T. Sasaki, T. Ideue, M. Sano, T. Mituyama, T. Hirose, MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 2525–2530.

[46] C.M. Clemson, J.N. Hutchinson, S.A. Sara, A.W. Ensminger, A.H. Fox, A. Chess, et al., An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles, Mol. Cell 33 (2009) 717–726.

[47] F. Eckhardt, J. Lewin, R. Cortese, V.K. Rakyan, J. Attwood, M. Burger, et al., DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 38 (2006) 1378–1385.

[48] M. Guttman, I. Amit, M. Garber, C. French, M.F. Lin, D. Feldser, et al., Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, Nature (2009).

[49] S. Efroni, R. Duttagupta, J. Cheng, H. Dehghani, D.J. Hoeppner, C. Dash, et al., Global transcription in pluripotent embryonic stem cells, Cell. Stem Cell 2 (2008) 437–447.

[50] A. Grimson, M. Srivastava, B. Fahey, B.J. Woodcroft, H.R. Chiang, N. King, et al., Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals, Nature 455 (2008) 1193–1197.

[51] I. Whitehouse, O.J. Rando, J. Delrow, T. Tsukiyama, Chromatin remodelling at promoters suppresses antisense transcription, Nature 450 (2007) 1031–1035.

[52] R.H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J.F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S.E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M.R. Brent, D.G. Brown, S.D. Brown, C. Bult, J. Burton, J. Butler, R.D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A.T. Chinwalla, D.M. Church, M. Clamp, C. Clee, F.S. Collins, L.L. Cook, R.R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K.D. Delehaunty, J. Deri, E.T. Dermitzakis, C. Dewey, N.J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D.M. Dunn, S.R. Eddy, L. Elnitski, R.D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G.A. Fewell, P. Flicek, K. Foley, W.N. Frankel, L.A. Fulton, R.S. Fulton, T.S. Furey, D. Gage, R.A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T.A. Graves, E.D. Green, S. Gregory, R. Guigo, M. Guyer, R.C. Hardison, D. Haussler, Y. Hayashizaki, L.W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D.B. Jaffe, L.S. Johnson, M. Jones, T.A. Jones, A. Joy, M. Kamal, E.K. Karlsson, et al., Initial sequencing and comparative analysis of the mouse genome, Nature 420 (2002) 520–562.

[53] M. Kamal, X. Xie, E.S. Lander, A large family of ancient repeat elements in the human genome is under strong selection, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 2740–2745.

[54] H. Nishihara, A.F. Smit, N. Okada, Functional noncoding sequences derived from SINEs in the mammalian genome, Genome Res. 16 (2006) 864–874.

[55] N.R. Smalheiser, V.I. Torvik, Mammalian microRNAs derived from genomic repeats, Trends Genet. 21 (2005) 322–326.

[56] J. Brosius, RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements, Gene 238 (1999) 115–134.

[57] J. Hasler, K. Strub, Alu elements as regulators of gene expression, Nucleic Acids Res. 34 (2006) 5491–5497.

[58] R. Britten, Transposable elements have contributed to thousands of human proteins, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 1798–1803.

[59] G. Bejerano, C.B. Lowe, N. Ahituv, B. King, A. Siepel, S.R. Salama, et al., A distal enhancer and an ultraconserved exon are derived from a novel retroposon, Nature 441 (2006) 87–90.

[60] M. Pheasant, J.S. Mattick, Raising the estimate of functional human sequences, Genome Res. 17 (2007) 1245–1253.

[61] F.F. Costa, Non-coding RNAs, epigenetics and complexity, Gene 410 (2008) 9–17.

[62] J.S. Mattick, A new paradigm for developmental biology, J. Exp. Biol. 210 (2007) 1526–1547.

[63] F.M. Pauler, M.V. Koerner, D.P. Barlow, Silencing by imprinted noncoding RNAs: is transcription the answer? Trends Genet. 23 (2007) 284–292.

[64] S. Kishore, S. Stamm, The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C, Science 311 (2006) 230–232.

[65] C. Ender, A. Krek, M.R. Friedlander, M. Beitzinger, L. Weinmann, W. Chen, et al., A human snoRNA with microRNA-like functions, Mol. Cell 32 (2008) 519–528.

[66] A.A. Aravin, N.M. Naumova, A.V. Tulin, V.V. Vagin, Y.M. Rozovsky, V.A. Gvozdev, Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline, Curr. Biol. 11 (2001) 1017–1027.

[67] A.A. Aravin, G.J. Hannon, J. Brennecke, The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race, Science 318 (2007) 761–764.

[68] W. Deng, H. Lin, miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis, Dev. Cell 2 (2002) 819–830.
[69] M.A. Carmell, A. Girard, H.J. van de Kant, D. Bourc'his, T.H. Bestor, D.G. de Rooij, et al., MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline, Dev. Cell 12 (2007) 503–514.
[70] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, Identification of novel genes coding for small expressed RNAs, Science 294 (2001) 853–858.
[71] N.C. Lau, L.P. Lim, E.G. Weinstein, D.P. Bartel, An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans, Science 294 (2001) 858–862.
[72] R.C. Lee, V. Ambros, An extensive class of small RNAs in Caenorhabditis elegans, Science 294 (2001) 862–864.
[73] V.N. Kim, J. Han, M.C. Siomi, Biogenesis of small RNAs in animals, Nat. Rev., Mol. Cell Biol. 10 (2009) 126–139.
[74] S.M. Hammond, E. Bernstein, D. Beach, G.J. Hannon, An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells, Nature 404 (2000) 293–296.
[75] T. Du, Zamore, microPrimer: the biogenesis and function of microRNA, Development 132 (2005) 4645–4652.
[76] W. Filipowicz, S.N. Bhattacharyya, N. Sonenberg, Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 9 (2008) 102–114.
[77] R.C. Friedman, K.K. Farh, C.B. Burge, D.P. Bartel, Most mammalian mRNAs are conserved targets of microRNAs, Genome Res. (2008).
[78] R. Sandberg, J.R. Neilson, A. Sarma, P.A. Sharp, C.B. Burge, Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites, Science 320 (2008) 1643–1647.
[79] I. Lee, S.S. Ajay, J.I. Yook, H.S. Kim, S.H. Hong, N.H. Kim, et al., New class of microRNA targets containing simultaneous 5′-UTR and 3′-UTR interaction sites, Genome Res. (2009).
[80] A.J. Giraldez, Y. Mishima, J. Rihel, R.J. Grocock, S. Van Dongen, K. Inoue, et al., Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs, Science 312 (2006) 75–79.
[81] L.P. Lim, N.C. Lau, P. Garrett-Engele, A. Grimson, J.M. Schelter, J. Castle, et al., Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs, Nature 433 (2005) 769–773.
[82] M. Selbach, B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin, N. Rajewsky, Widespread changes in protein synthesis induced by microRNAs, Nature (2008).
[83] D. Baek, J. Villen, C. Shin, F.D. Camargo, S.P. Gygi, D.P. Bartel, The impact of microRNAs on protein output, Nature (2008).
[84] A.S. Flynt, E.C. Lai, Biological principles of microRNA-mediated regulation: shared themes amid diversity, Nat. Rev. Genet. 9 (2008) 831–842.
[85] S. Griffiths-Jones, R.J. Grocock, S. van Dongen, A. Bateman, Enright, miRBase: microRNA sequences, targets and gene nomenclature, Nucleic Acids Res. 34 (2006) D140–D144.
[86] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long non-coding RNAs: insights into functions, Nat. Rev. Genet. 10 (2009) 155–159.
[87] J.E. Wilusz, H. Sunwoo, D.L. Spector, Long noncoding RNAs: functional surprises from the RNA world, Genes Dev. 23 (2009) 1494–1504.
[88] J. Paul, J.D. Duerksen, Chromatin-associated RNA content of heterochromatin and euchromatin, Mol. Cell. Biochem. 9 (1975) 9–16.
[89] A. Rodriguez-Campos, F. Azorin, RNA is an integral component of chromatin that contributes to its structural organization, PLoS ONE 2 (2007) e1182.
[90] A. Weiss, I. Keshet, A. Razin, H. Cedar, DNA demethylation in vitro: involvement of RNA, Cell 86 (1996) 709–718.
[91] T. Imamura, S. Yamamoto, J. Ohgane, N. Hattori, S. Tanaka, K. Shiota, Non-coding RNA directed DNA demethylation of Sphk1 CpG island, Biochem. Biophys. Res. Commun. 322 (2004) 593–600.
[92] C. Tufarelli, J.A. Stanley, D. Garrick, J.A. Sharpe, H. Ayyub, W.G. Wood, et al., Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease, Nat. Genet. 34 (2003) 157–165.
[93] K.V. Morris, S. Santoso, A.M. Turner, C. Pastori, P.G. Hawkins, Bidirectional transcription directs both transcriptional gene activation and suppression in human cells, PLoS Genet. 4 (2008) e1000258.
[94] P. Navarro, D.R. Page, P. Avner, C. Rougeulle, Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program, Genes Dev. 20 (2006) 2787–2792.
[95] P.P. Luedi, F.S. Dietrich, J.R. Weidman, J.M. Bosko, R.L. Jirtle, A.J. Hartemink, Computational and experimental identification of novel human imprinted genes, Genome Res. 17 (2007) 1723–1730.
[96] P. Kapranov, J. Cheng, S. Dike, D.A. Nix, R. Duttagupta, A.T. Willingham, et al., RNA maps reveal new RNA classes and a possible function for pervasive transcription, Science 316 (2007) 1484–1488.
[97] Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs, Nature 457 (2009) 1028–1032.
[98] T. Shiraki, S. Kondo, S. Katayama, K. Waki, T. Kasukawa, H. Kawaji, et al., Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 15776–15781.
[99] R.J. Taft, E.A. Glazov, N. Cloonan, C. Simons, S. Stephen, G.J. Faulkner, et al., Tiny RNAs associated with transcription start sites in animals, Nat. Genet. 41 (2009) 572–578.
[100] J.M. Cummins, Y. He, R.J. Leary, R. Pagliarini, L.A. Diaz Jr., T. Sjoblom, et al., The colorectal microRNAome, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 3687–3692.
[101] E. Torarinsson, M. Sawera, J.H. Havgaard, M. Fredholm, J. Gorodkin, Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure, Genome Res. 16 (2006) 885–889.
[102] E. Berezikov, G. van Tetering, M. Verheul, J. van de Belt, L. van Laake, J. Vos, et al., Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis, Genome Res. 16 (2006) 1289–1298.
[103] I.L. Ross, C.M. Browne, D.A. Hume, Transcription of individual genes in eukaryotic cells occurs randomly and infrequently, Immunol. Cell Biol. 72 (1994) 177–185.
[104] J.M. Levsky, S.M. Shenoy, R.C. Pezo, R.H. Singer, Single-cell gene expression profiling, Science 297 (2002) 836–840.
[105] A. Raj, C.S. Peskin, D. Tranchina, D.Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS Biol. 4 (2006) e309.
[106] M. Acar, J.T. Mettetal, A. van Oudenaarden, Stochastic switching as a survival strategy in fluctuating environments, Nat. Genet. 40 (2008) 471–475.
[107] H.H. Chang, M. Hemberg, M. Barahona, D.E. Ingber, S. Huang, Transcriptome-wide noise controls lineage choice in mammalian progenitor cells, Nature 453 (2008) 544–547.
[108] A. Brock, H. Chang, S. Huang, Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours, Nat. Rev. Genet. 10 (2009) 336–342.
[109] V.A. Kuznetsov, G.D. Knott, R.F. Bonner, General statistics of stochastic process of gene expression in eukaryotic cells, Genetics 161 (2002) 1321–1332.
[110] Y. Chen, J. Zhu, P.Y. Lum, X. Yang, S. Pinto, D.J. MacNeil, et al., Variations in DNA elucidate molecular networks that cause disease, Nature 452 (2008) 429–435.
[111] V. Emilsson, G. Thorleifsson, B. Zhang, A.S. Leonardson, F. Zink, J. Zhu, et al., Genetics of gene expression and its effect on disease, Nature 452 (2008) 423–428.
[112] D.W. Parsons, S. Jones, X. Zhang, J.C. Lin, R.J. Leary, P. Angenendt, et al., An integrated genomic analysis of human glioblastoma multiforme, Science 321 (2008) 1807–1812.
[113] S. Jones, X. Zhang, D.W. Parsons, J.C. Lin, R.J. Leary, P. Angenendt, et al., Core signaling pathways in human pancreatic cancers revealed by global genomic analyses, Science 321 (2008) 1801–1806.