RNA driving the epigenetic bus

John S Mattick

Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland, Australia. Correspondence to: j.mattick@imb.uq.edu.au

The EMBO Journal (2012) 31, 515–516. doi:10.1038/emboj.2011.479

For many years, it has been assumed that mammalian development is primarily controlled by transcription factors, but recent studies, including that reported by Ng *et al* (2012) in this issue, indicate that there is an additional massive hidden layer of regulatory RNAs that control the site specificity of chromatin-modifying complexes and other regulatory proteins involved in the control of embryological and neural differentiation.

Dynamic changes to chromatin structure appear to constitute the senior level of gene regulation during differentiation and development by determining which loci may be open or closed for expression, although it has been assumed that this is itself controlled by transcription factors. The term ‘transcription factors’ is, in fact, used rather loosely as DNA- or chromatin-binding proteins are many and varied, and act in different ways, presumably at different levels in decisional hierarchies. It is also increasingly clear that there are other levels of regulation of gene expression that involve target-specific RNAs, such microRNAs that control mRNA translation and stability.

It appears that molecular biology may have misunderstood the nature and extraordinary complexity of the regulatory information that controls the fine details of human differentiation and development, on the superficially reasonable but most likely mistaken assumption that most genetic information is transacted by proteins, and the derived assumption that combinatoric interactions between regulatory proteins can provide all of the necessary computational power to put 100 trillion specialized cells in the right places. Indeed, while some transcription factors are powerful controllers of cell type, the far bigger challenge is to specify the precise architecture of different organs, muscles and bones, as well as functional neural networks in the brain.

Less than 1.5% of the human genome encodes proteins, most of which are orthologous to those in other animals, but genome-wide analyses have shown that virtually the entire genome is differentially transcribed in highly complex cell-specific patterns, to produce tens if not hundreds of thousands of long non-coding RNAs (lncRNAs; Mercer *et al*, 2012). Over the past few years, a number of studies have shown that these lncRNAs are specifically expressed, especially in the brain (Mercer *et al*, 2008), and are dynamically regulated during cell differentiation, including during embryonal and neural stem cell differentiation (Dinger *et al*, 2008; Mercer *et al*, 2010; Guttman *et al*, 2011).

These studies have now been extended by Ng *et al* (2012), who analysed a set of lncRNAs that are dynamically regulated

during the differentiation of human stem cells into neurons and, using siRNA knockdown, identified subsets that are involved either in the maintenance of pluripotency or in neuronal differentiation. This mirrors similar findings by Guttman *et al* (2011) that lncRNAs are also involved in embryonal stem cell maintenance and lineage specification, and that the expression of lncRNAs is itself regulated by various transcription factors, that is, behave similarly to protein-coding genes, presumably in a largely predetermined feed-forward program that unfolds during normal embryogenesis. Ng *et al* (2012) also show that lncRNAs themselves associate with ‘transcription factors’ such as SOX2 and REST, whose mode of action is not well understood.

These papers also showed that lncRNAs are associated with chromatin-modifying complexes, such as polycomb components and histone methyltransferases, extending earlier studies showing association of lncRNAs with both activating and repressive chromatin-modifying enzymes and states (Dinger *et al*, 2008; Khalil *et al*, 2009; Koziol and Rinn, 2010). The inescapable conclusion is that these RNAs are likely acting as adaptors to assemble different suites of generic effector proteins that are recognized and bound by secondary structural features embedded within the RNA, and to direct these to specific genomic positions by virtue of RNA–DNA interactions (Koziol and Rinn, 2010; Figure 1).

This previously hidden world of RNA-directed epigenetic control of gene structure and expression may be extremely sophisticated, not simply operating at the regional level, but extending to individual exons and other features such as

Figure 1 Long non-coding RNAs as adaptors that assemble chromatin-modifying complexes and direct them to target sites by RNA–DNA interactions.

promoters and enhancers. It has recently been discovered that nucleosomes are preferentially positioned over exons in both somatic and germ cells (Nahkuri *et al*, 2009), raising the prospect of the preprogrammed epigenetic control of splicing, which has since obtained experimental support (Luco *et al*, 2011). When one considers the extraordinary amount of information required to impose such fine control of gene structure and expression, and the myriad of histone modifications involved, it is not surprising that most of the genome may be differentially transcribed to supply this information. Many if not most lncRNAs are themselves alternatively spliced (Mercer *et al*, 2012), adding further complexity to this scenario, and painting an entirely new picture of the genomic programming of human development and brain

function, whereby relatively generic, albeit often state-specific, effector proteins (chromatin-modifying complexes and ‘transcription factors’) are guided to their sites of action in the genome by an army of adaptor RNAs, which themselves may be subject to allosteric interactions. Clearly long-held ideas of gene regulation in development and cognition will have to be reassessed, although there is a long way to go in understanding the details and mechanics of RNA interactions with DNA and with complexes involved in writing and erasing chromatin marks.

Conflict of interest

The author declares that he has no conflict of interest.

References

- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Perkins AC, Mattick JS (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res* **18**: 1433–1445
- Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L, Yang X, Amit I, Meissner A, Regev A, Rinn JL, Root DE, Lander ES (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**: 295–300
- Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, Regev A, Lander ES, Rinn JL (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA* **106**: 11667–11672
- Koziol MJ, Rinn JL (2010) RNA traffic control of chromatin complexes. *Curr Opin Genet Dev* **20**: 142–148
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* **144**: 16–26
- Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* **105**: 716–721
- Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh JA, Mattick JS, Rinn JL (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol* **30**: 99–104
- Mercer TR, Qureshi IA, Gokhan S, Dinger ME, Li G, Mattick JS, Mehler MF (2010) Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci* **11**: 14
- Nahkuri S, Taft RJ, Mattick JS (2009) Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**: 3420–3424
- Ng S-Y, Johnson R, Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* **31**: 522–533