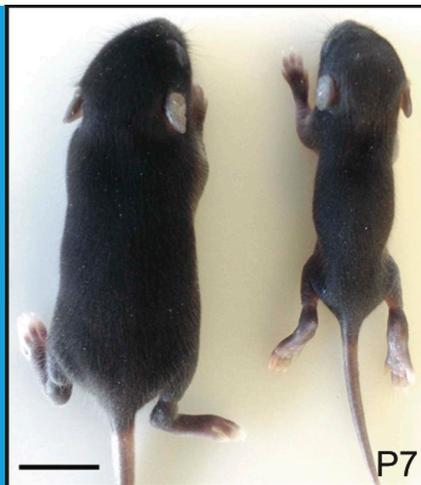


## GENETICS

# Probing the phenomics of noncoding RNA

Genetic knockout experiments on mice confirm that some long noncoding RNA molecules have developmental functions.

JOHN S MATTICK



**Related research article** Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, Liapis SC, Mallard W, Morse M, Swerdel MR, D'Ecclesiss MF, Moore JC, Lai V, Gong G, Yancopoulos GD, Frenthewey D, Kellis M, Hart RP, Valenzuela DM, Arlotta P, Rinn JL. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2:01749. doi: 10.7554/eLife.01749

**Image** A wild-type (left) and mutant (right) mouse seven days after birth; this particular mutation (*Mdgt*<sup>-/-</sup>) is partially lethal with growth defects in survivors

elements as body size goes up and population size goes down (Lynch, 2007).

High-throughput sequencing analyses over the past decade have shown that the majority of mammalian genome is transcribed, often from both strands, and have revealed an extraordinarily complex landscape of overlapping and interlacing sense and antisense, alternatively spliced, protein-coding and non-protein-coding RNAs, the latter generally referred to as long noncoding RNAs (lncRNAs). Moreover, the repertoire of these lncRNAs is different in different cells (Carninci et al., 2005; Cheng et al., 2005; Birney et al., 2007; Mercer et al., 2012). While some transcripts may encode previously unrecognized small proteins, the function or otherwise of the vast majority of lncRNAs remains to be determined.

Because many lncRNAs appear to be expressed at low levels, and many have lower sequence conservation than messenger RNAs, one interpretation has been that these RNAs represent transcriptional noise from complex genomes cluttered with evolutionary debris. However, assessments of sequence conservation rely on assumptions about the non-functionality and representative distribution of reference sequences, which are not verified and cannot be directly tested (Pheasant and Mattick, 2007). Nonetheless, many lncRNAs show patches of relative sequence conservation (Derrien et al., 2012), and even more do so at the secondary structural level (Smith et al., 2013).

Expression analyses have shown that lncRNAs originate from all over the genome and are expressed at different times during differentiation and development (Dinger et al., 2008), often exhibiting highly cell-specific patterns

It has been known since the late 1970s that many DNA sequences are transcribed but not translated. Moreover, most protein-coding genes in mammals are fragmented, with only a small fraction of the primary RNA transcript being spliced together to form messenger RNA. For many years it was assumed that untranslated RNA molecules served no useful purpose but, starting in the mid-1990s, a small body of researchers, including the present author (Mattick, 1994), have been arguing that these RNAs transmit regulatory information, possibly associated with the emergence of multicellular organisms. This is supported by the observation that the proportion of noncoding genomic sequences broadly correlates with developmental complexity, reaching over 98% in mammals (Liu et al., 2013), although others have argued that the increase in genome size is due to the inefficiency of selection against non-functional

(Mercer *et al.*, 2008). The precision of lncRNA expression is consistent with evidence suggesting that many are associated with chromatin-modifying complexes, thereby acting as regulators of the epigenetic control of differentiation and development (Mercer and Mattick, 2013).

A number of lncRNAs have also been linked to complex diseases like cancer (Mattick, 2009) and other complex physiological processes (see, for example, Rapicavoli *et al.*, 2013). However, these results seem at odds with the fact that few lncRNAs have been identified in traditional genetic screens. The reason for this is likely a combination of phenotypic, technical and expectational bias: mutations in protein-coding regions of the genome generally have phenotypes that are more severe, and are easier to identify, than those in non-coding regions. By contrast, in this context, it is worth noting that ~95% of all variants associated with complex (as opposed to monogenic) diseases in humans map to non-coding, presumably regulatory, sequences (Freedman *et al.*, 2011).

Still, the gold standard in this field is the targeted *in vivo* silencing or deletion of specific genes, and since few of these have been conducted to date, some researchers have remained sceptical about the biological significance of lncRNAs. Now, in *eLife*, John Rinn, Paolo Arlotta and co-workers at Harvard, MIT, the Broad Institute, Rutgers and Regeneron Pharmaceuticals—including Martin Sauvageau, Loyal Goff and Simona Lodata as joint first authors—report the results of the first large-scale attack on the question (Sauvageau *et al.*, 2013). They selected 18 lncRNA genes in the mouse genome that had been stringently assessed for lack of protein-coding capacity and that did not overlap with known protein-coding genes or other known gene annotations—hence the name long intergenic noncoding RNAs (lincRNAs)—and generated knockout mouse mutants by replacing the lncRNA gene with a *lacZ* reporter cassette.

Sauvageau, Goff, Lodata *et al.* report discernable developmental problems in five of the 18 mutants, with three exhibiting embryonic or post-natal lethality, two of which exhibited growth defects in the survivors. The phenotypes of two of the mutants were analyzed in detail: one of the mutants that died showed defects in multiple organs (including the lung, heart and gastrointestinal tract), and one of the mutants that survived with growth defects also showed defects in the cerebral cortex. Other mutants that did not exhibit overt developmental defects showed brain-specific expression patterns and may be associated with

cognitive defects that are not grossly apparent at the developmental level.

Another group (Grote *et al.*, 2013) recently generated a different knockout allele for one of the 18 lincRNAs interrogated by Sauvageau *et al.*, and also reported an embryonic lethal phenotype, albeit with some differences. Importantly, the approach used by Grote *et al.* also provided strong evidence that the mutant defects were not caused by an indirect effect on an overlapping genomic element, such as an enhancer for a nearby gene.

The work of Sauvageau, Goff, Lodata *et al.* is a mini tour-de-force that shows that there are lncRNAs with important developmental functions *in vivo*, and it joins a small number of studies from other pioneering groups that show the same thing (Lewejohann *et al.*, 2004; Gutschner *et al.*, 2013; Li *et al.*, 2013), although not all of the targeted lncRNAs showed a phenotype. Similarly, other knockout experiments of widely expressed lncRNAs, as well as some of the most highly conserved elements in the mammalian genome, also did not yield discernable phenotypes (Ahituv *et al.*, 2007; Nakagawa *et al.*, 2011), which should sound a note of caution about the interpretation of negative results.

Indeed, since most lncRNAs are expressed in the brain (Mercer *et al.*, 2008) and many are primate-specific (Derrien *et al.*, 2012), it may be that much of the lncRNA-mediated genetic information in humans (and in mammals generally) is devoted to brain function, and therefore not easily detectable in developmental, as opposed to cognitive, screens. A good example is a noncoding RNA called BC1 that is widely expressed in the brain: knockout of BC1 causes no visible anatomical consequences, but it leads to a behavioural phenotype that would be lethal in the wild (Lewejohann *et al.*, 2004).

Although evidence for the hypothesis that lncRNAs have a role in mammalian development, brain function and physiology is growing, there is also a clear need for more sophisticated and comprehensive phenotypic screens, especially with respect to cognitive function.

**John S Mattick** is an *eLife* reviewing editor, and is at the Garvan Institute of Medical Research, Sydney, Australia and St Vincent's Clinical School and the School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, Australia  
j.mattick@garvan.org.au

**Competing interests:** The author declares that no competing interests exist.

**Published** 31 December 2013

## References

- Ahituv N**, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biology* **5**:e234. doi: [10.1371/journal.pbio.0050234](https://doi.org/10.1371/journal.pbio.0050234).
- Birney E**, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**:799–816. doi: [10.1038/nature05874](https://doi.org/10.1038/nature05874).
- Carninci P**, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Carninci P, Kasukawa T, Katayama S, Gough J, et al. 2005. The transcriptional landscape of the mammalian genome. *Science* **309**:1559–1563. doi: [10.1126/science.1112014](https://doi.org/10.1126/science.1112014).
- Cheng J**, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**:1149–1154. doi: [10.1126/science.1108625](https://doi.org/10.1126/science.1108625).
- Derrien T**, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research* **22**:1775–1789. doi: [10.1101/gr.132159.111](https://doi.org/10.1101/gr.132159.111).
- Dinger ME**, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Soldà G, Simons C, et al. 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research* **18**:1433–1445. doi: [10.1101/gr.078378.108](https://doi.org/10.1101/gr.078378.108).
- Freedman ML**, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, et al. 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nature Genetics* **43**:513–518. doi: [10.1038/ng.840](https://doi.org/10.1038/ng.840).
- Grote P**, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, Macura K, Blass G, Kellis M, Werber M, et al. 2013. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Developmental Cell* **24**:206–214. doi: [10.1016/j.devcel.2012.12.012](https://doi.org/10.1016/j.devcel.2012.12.012).
- Gutschner T**, Hammerle M, Eissmann M, Hsu J, Kim Y, Hung G, Revenko A, Arun G, Stentrup M, Gross M, et al. 2013. The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Research* **73**:1180–1189. doi: [10.1158/0008-5472.CAN-12-2850](https://doi.org/10.1158/0008-5472.CAN-12-2850).
- Lewejohann L**, Skryabin BV, Sachser N, Prehn C, Heiduschka P, Thanos S, Jordan U, Dell’Omo G, Vysotski AL, Pleskacheva MG, et al. 2004. Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. *Behavioural Brain Research* **154**:273–289. doi: [10.1016/j.bbr.2004.02.015](https://doi.org/10.1016/j.bbr.2004.02.015).
- Li L**, Liu B, Wapinski OL, Tsai MC, Qu K, Zhang J, Carlson JC, Lin M, Fang F, Gupta RA, et al. 2013. Targeted disruption of Hota1r leads to homeotic transformation and gene derepression. *Cell Reports* **5**:3–12. doi: [10.1016/j.celrep.2013.09.003](https://doi.org/10.1016/j.celrep.2013.09.003).
- Liu G**, Mattick JS, Taft RJ. 2013. A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle* **12**:2061–2072. doi: [10.4161/cc.25134](https://doi.org/10.4161/cc.25134).
- Lynch M**. 2007. *The origins of genome architecture*. Sunderland, MA: Sinauer Associates.
- Mattick JS**. 1994. Introns: evolution and function. *Current Opinion in Genetics & Development* **4**:823–831. doi: [10.1016/0959-437X\(94\)90066-3](https://doi.org/10.1016/0959-437X(94)90066-3).
- Mattick JS**. 2009. The genetic signatures of noncoding RNAs. *PLoS Genetics* **5**:e1000459. doi: [10.1371/journal.pgen.1000459](https://doi.org/10.1371/journal.pgen.1000459).
- Mercer TR**, Dinger ME, Sunken SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **105**:716–721. doi: [10.1073/pnas.0706729105](https://doi.org/10.1073/pnas.0706729105).
- Mercer TR**, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddelloh JA, Mattick JS, Rinn JL. 2012. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnology* **30**:99–104. doi: [10.1038/nbt.2024](https://doi.org/10.1038/nbt.2024).
- Mercer TR**, Mattick JS. 2013. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural and Molecular Biology* **20**:300–307. doi: [10.1038/nsmb.2480](https://doi.org/10.1038/nsmb.2480).
- Nakagawa S**, Naganuma T, Shioi G, Hirose T. 2011. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. *Journal of Cell Biology* **193**:31–39. doi: [10.1083/jcb.201011110](https://doi.org/10.1083/jcb.201011110).
- Pheasant M**, Mattick JS. 2007. Raising the estimate of functional human sequences. *Genome Research* **17**:1245–1253. doi: [10.1101/gr.6406307](https://doi.org/10.1101/gr.6406307).
- Rapicavoli NA**, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY. 2013. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *eLife* **2**:e00762. doi: [10.7554/eLife.00762](https://doi.org/10.7554/eLife.00762).
- Sauvageau M**, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**:01749. doi: [10.7554/eLife.01749](https://doi.org/10.7554/eLife.01749).
- Smith MA**, Gesell T, Stadler PF, Mattick JS. 2013. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Research* **41**:8220–8236. doi: [10.1093/nar/gkt596](https://doi.org/10.1093/nar/gkt596).