Genes without protein products: Is *H19* the norm or the exception?

Chandrasekhar Kanduri*, Susan Pfeifer*,†, Yiming Li* and Rolf Ohlsson*,§

*Department of Animal Development and Genetics, Uppsala University, Norbyvägen 18A, S-752 36 Uppsala, Sweden
†Department of Pediatrics, Uppsala University Hospital, S-751 85 Uppsala, Sweden

The increasing number of RNA polymease II transcripts without any apparent open reading frame has increased our awareness that gene functions can be selected for without involving a protein product. By using the *H19* gene as a point of reference, we highlight here several common features among non-coding genes, such as their antisense position in subchromosomal expression domains which are often genomically imprinted. We also discuss the need to critically examine the translatability of transcripts which are considered non-coding. Finally, we present a model to explain the origin of non-coding genes.

A central dogma in eukaryotic molecular biology holds that RNA polymerase II-driven transcription produces an mRNA, which is then translated in the cytoplasm to yield a primary protein product. This protein product represents the selection vehicle which determines whether or not the particular gene will survive evolutionary pressure. An increasing number of RNA polymerase II transcripts that have no obvious open reading frame (ORF) bear witness, however, to the evolution of alternative roles for 'mRNAs', other than for producing a primary protein product[1]. There is not, however, an obvious common denominator between these transcripts. For example, the *H19* (ref. 2) and *Ipw* (ref. 3) transcripts are predominantly localized in the cytoplasmic compartment, while the *Xist* (ref. 4), *rox1* (ref. 5) and *NTT* (ref. 6) transcripts are localized in the nucleus. In fact, current (although in most instances preliminary) data, suggests that noncoding transcripts are involved in many vital cellular processes, such as the maintenance of nuclear architectue[7] and the regulation of gene expression[8,9]. This data implies that despite the absence of an ORF, RNA polymerase II transcripts can be functionally recruited in a wide variety of biological processes to promote the survival of non-coding genes.

All of the above-mentioned genes produce polyadenylated transcripts, while some of these bear plausible ORFs. For example, both the mouse and human *H19* transcripts have rather long ORFs although these are not evolutionary conserved. It is currently unclear if the absence of a conserved ORF is a sufficient proof that a given gene does not produce a protein. Such a situation emphasizes the need to critically examine the requirements for classifying a transcript as non-coding in each individual case. In this review, we discuss the ORF issue, as well as structural and functional aspects of non-coding genes, using the imprinted *H19* gene as an example. We also present a simple model which explains the origin of non-coding, RNA polymerase II-derived transcripts.

Many mammalian non-coding genes locate at genomically imprinted subchromosomal domains

It is an unexpected observation that a significant proportion of the currently characterized non-coding transcripts in mammals are expressed in a parent of origin-specific manner, i.e. they belong to the class of imprinted genes. This rapidly increasing list of non-coding imprinted genes include nine members which have been identified in the following loci: *VIZ*, *H19*, *Igf2* (antisense)[10], *Igf2r* (antisense)[8], *Xist/XIST* (sense and antisense)[4,9], *Ipw/IPW*, *SNRPN*, *UBE3A* (antisense)[11] and *Znf127* (antisense)[12]. While the emphasis of this review is on the *H19* transcript (see below), we would first like to discuss the interesting fact that most of the imprinted, non-coding transcripts are derived from imprinted subchromosomal domains and consider the emergence of a common theme for a role for non-coding transcripts in the imprinting phenomenon.

One of the most well known cluster of imprinted genes contains the *Igf2* and *H19* genes, which are expressed monoallelically from the paternal and maternal alleles, respectively. A region upstream of *Igf2*, DMR (a differentially methylated region), produces three paternally expressed and non-coding transcripts; *Igf2*as-a, *Igf2*as-b and *Igf2*as-c, in antisense (AS) orientation. These AS transcripts have no likely ORFs and contain several tandem repeats[10]. The *IGF2/H19* cluster also includes the maternally expressed *KVLQT1* gene which has been shown to express several noncoding isoforms of *KVLQT1* transcripts[13].

§For correspondence. (e-mail: rolf.ohlsson@devbiol.uu.se)

The human chromosome region 15q13-15, which is associated with genetic and epigenetic disturbances that generate both Prader-Willi syndrome and Angelmann syndrome, harbours non-coding genes such as *IPW/Ipw*, *PAR1*, *PAR5*, *PARSN* and *ASR1,2* (ref. 14). *IPW/Ipw* (imprinted gene in Prader-Willi syndrome) expresses a processed RNA with no significant ORFs, which is localized predominantly in the cytoplasm[15]. An antisense RNA is produced from the *ZNF127* locus, which encodes a protein with RING zinc-finger and multiple zinc-finger motifs. The *ZNF127* and *ZNF127* AS RNA vary in their expression patterns as well as in size of the transcripts[12]. The biological functions of IPW (sense) and *ZNF127* (antisense) transcripts, if any, are not clear at this moment. The *SNRPN* gene, which spans 360 kb of DNA produces both coding and non-coding imprinted transcripts[16]. The exons at the 3′ part of the gene generate a coding mRNA when spliced with exon 1 and non-coding RNAs when spliced with the 5′ BD exons[16]. BD transcripts, encoded by two alternative 5′ exons, BD1B and BD1A, have two alternative start sites and are subject to alternative splicing[16]. They are expressed from the paternal chromosome only, as the BD exons are heavily methylated on the maternal chromosome (16). It has recently been shown that paternally imprinted antisense RNA is produced at the 3′ UTR of Angelmann syndrome gene, UBE3A (11), which encodes ubiquitin protein ligase that functions in protein turnover[17].

Role of non-coding transcripts in the imprinting phenomenon

The observation that several non-coding transcripts originate in the neighbourhood of imprinted genes suggests that such transcripts may have a role in imparting monoallelic expression to target imprinted genes. This is implicated by the observation that the deletion of the *Igf2* upstream region, encompassing the *Igf2* AS transcripts, in a mouse fibroblast cell line results in a loss of imprinting[18]. Chromosomal rearrangements of the *KVLQT1* region are common in the Beckwith-Wiedemann syndrome which is linked with an overactive *IGF2* (ref. 13).

There is an emerging theme which suggests that non-coding AS transcripts actually regulate the imprinting status of genes expressing coding transcripts. This is exemplified by the *Igf2r* AS gene, which includes a stretch of differentially methylated repetitive sequences (termed region 2) and regulates the parent of origin-specific expression of the *Igf2r* gene[14]. Although a deletion of the region 2 resulted in biallelic expression of *Igf2r* when paternally inherited, it is not yet clear whether the *Igf2r* AS RNA plays a role in the imprinting process. By analogy to *Xist*, the observation that the *Igf2r* AS transcript coats the chromatin at the *Igf2r* locus even in metaphase chromosomes[8] (Barlow, personal information), could indicate a role in the formation of local heterochromatin. Alternatively, the epigenetic states of the promoters for the *Igf2r* and *Igf2r* AS RNA might set the stage for an enhancer competition situation[14]. This analogy may extend to the *UBE3A* gene, since an antisense transcript, which maps to the 3′UTR of *UBE3A* is derived from the paternal allele[11] whereas the brain-specific *UBE3A* transcript is derived from the maternal allele[17].

More compelling evidence for a role of non-coding antisense transcripts comes from studies on mammalian X-chromosome inactivation. To prevent overdosage, one of the two X chromosomes is randomly inactivated during early development in somatic cells in female mammals. Once the choice has been made, the inactive or active state is stably propagated during development. These events are controlled by the X-inactivation centre, which not only regulates the initiation of silencing, but also controls X chromosome counting and determines which X chromosome to (in)activate[19]. These properties have now been ascribed to two different non-coding transcripts which are derived from two overlapping genes, *Xist* and *Tsix*[9,20]. Whereas the silencing of the future inactive X chromosome involves an active *Xist* gene, the *Tsix* gene function seems to primarily involve repression of transcription of the *Xist* gene[9]. This conclusion is based on a combination of genetic evidence and the observation that at the onset of X-inactivation, *Tsix* expression becomes restricted to the active X chromosome and persists until *Xist* is turned off[9].

More complex scenarios can be depicted, however, for imprinted, non-coding transcripts. The background is that deletions in the BD exon portion of the *SNRPN* locus results in the Angelman syndrome phenotype[21–23]. It has been postulated, therefore, that the non-coding BD transcripts interact with a switch initiator site which has been mapped at or in the vicinity of exon 1 of the *SNRPN* gene[16]. It has been further hypothesized that the mutations in the imprint switch initiation site may block the maternal to paternal imprint switch, resulting in Prader-Willi syndrome phenotype. In contrast, the Angelman syndrome phenotype results from the block in the paternal to maternal imprint switch due to mutations in the BD exons.

The *Igf2/H19* paradigm; is there a role for the *H19* gene?

The *H19* and *Igf2* genes, which are close physical neighbours separated by approximately 90 kb, are believed to share a common set of enhancers[24], as evidenced by a coordinated expression pattern during prenatal growth[25] and genetic dissection experiments[24,26]. To explain the parent of origin-specific si-

lencing of these genes, it was suggested that the ability of the *H19* and *Igf2* loci to compete for a common set of enhancers depended on their parental origin[27]. The paternal *H19* and maternal *Igf2* alleles are believed to be epigenetically modified to prevent the accessibility of the enhancer. A role for the structural *H19* gene in this process was indicated by the experiment in which its replacement with a reporter gene in a transgenic context resulted in loss of *H19 imprinting*[28]. This assumed role of the *H19* gene appeared to be analogous with an earlier observation that targeted replacement of a portion of the *Xist* gene on the inactive X-chromosome resulted in the upregulation of neigbouring genes[20]. It was later documented, however, that a targeted replacement of the *H19* structural gene did not result in any perturbed imprinting manifestation[29]. The absence of a marked phenotype in such *H19 -/-* mice and in mice experiencing an overexpressed *H19* transgene[30], seemed to rule out any obvious role for the structural *H19* gene. Rather, it was suggested that the only function of the *H19* locus is that its transcriptional activity is required to imprint *Igf2* and that the conservation of its product reflects the necessity of packaging and sequestering a very abundant RNA[29].

In parallel with the mouse genetic approaches, other reports have attempted to establish a function for the human *H19* transcript. The first indication appeared with the observation that an over-expressed *H19* gene rescued the normal phenotype when transfected into rhabdomyosarcoma cells[31]. Although this assay has been successfully applied to other cell types[31], its generality has been questioned[32]. The picture emerging from all of these results is confusing and deepens the engima. Is the case for a function of the *H19* transcripts a dead duck, or do we miss an important piece of information in the quest for an *H19* function?

The *H19* transcript is associated with polysomes and may regulate *IGF2* expressivity in *trans*

Initial studies reported that the *H19* transcript is excluded from the polysomal fraction and cannot be translated *in vitro*[2]. Although this could certainly be expected from a non-coding transcript, a recent report challenges this result. It was observed that the distribution of *H19* mouse RNA partially overlapped with the polysomal marker, β-actin mRNA[33]. The two transcripts appear to exist in both polysomal and non-polysomal pools. Moreover, it was shown that the polysome-association of the human *H19* transcript was sensitive to a translational inhibitor, pactamycin[33]. These experiments were performed in the absence of cycloheximide (in contrast to report by Brannan *et al.*[2]), which stabilizes the polysomes and yield larger polysomes than in untreated animals or cells. It is possible, therefore, that such polysomes would pellet quantitatively during centrifugation, perhaps explaining why *H19* RNA association to polysomes had gone unnoticed[2].

A potential clue to the function of the human *H19* transcript was offered by a unique Wilms' tumour specimen: it expressed *IGF2* and *H19* at high levels in almost identical patterns and displayed a postneoplastic loss of *IGF2* imprinting, in which only a subpopulation of the tumour cells express both parental alleles[33,34]. An *in situ* hybridization analysis revealed that the *H19* expression was absent in some cell subpopulations of the tumour[33,34]. A closer look showed that the *IGF2* hybridization signal was 2–3-fold higher in such *H19*-negative cells when compared to the neighbouring *H19*-positive cells[33]. This observation could be documented in all of the *H19*-negative areas, suggesting an inverse correlation between *H19* expression and cytoplasmic levels of *IGF2* mRNA. Since all the cells that expressed *IGF2* biallelically also lacked *H19* expression, there was an opportunity to address whether or not *H19* expression correlated with *IGF2* mRNA translatability. This strategy predicted that if *H19* expression modifies the translatability of *IGF2* mRNAs, the sedimentation properties of the transcripts would differ in a parent of origin-specific manner in a sucrose gradient analysis. Indeed, it turned out that the *IGF2* mRNA which was transcribed from the generally silent allele was much more efficiently translated than the *IGF2* transcripts derived from the more generally expressed allele[33]. These results are consistent with a role for *H19* in modifying the translatability and/or the stability of *IGF2* mRNAs.

A model for an *H19* function in *trans*

These observations provide yet another glimpse into the complex regulation of the function of *IGF2*. This involves multiple steps of controls from gene dosage, differential promoter usage, splicing patterns, translational control(s) to post-secretory attenuation of *IGF2* function by IGF-binding proteins[35]. Importantly, the cytoplasmic and extra-cellular levels of control appear to be generally uncoordinated with *IGF2* activity, i.e. that the expression of *IGF2* and IGF-binding protein genes, for example, are differentially regulated. Conversely, the expression of the *IGF2* and *H19* genes and hence the antagonistic function of *H19* in *trans* are coordinated, since they share common regulatory elements[36]. This allows us to formulate a model in which *H19* serves to prevent overshoot of *IGF2* expression in *trans*. In this model (Figure 1), medium increments of *IGF2* expression would lead to only a moderate increase in the levels of IGF II ligand due to the uncoordinated types of negative controls, such as IGF-binding proteins. High levels of *IGF2* expression, however, would saturate the uncoordinated types of negative cytoplasmic and/or extracellular controls. On the other hand, high levels of *IGF2*

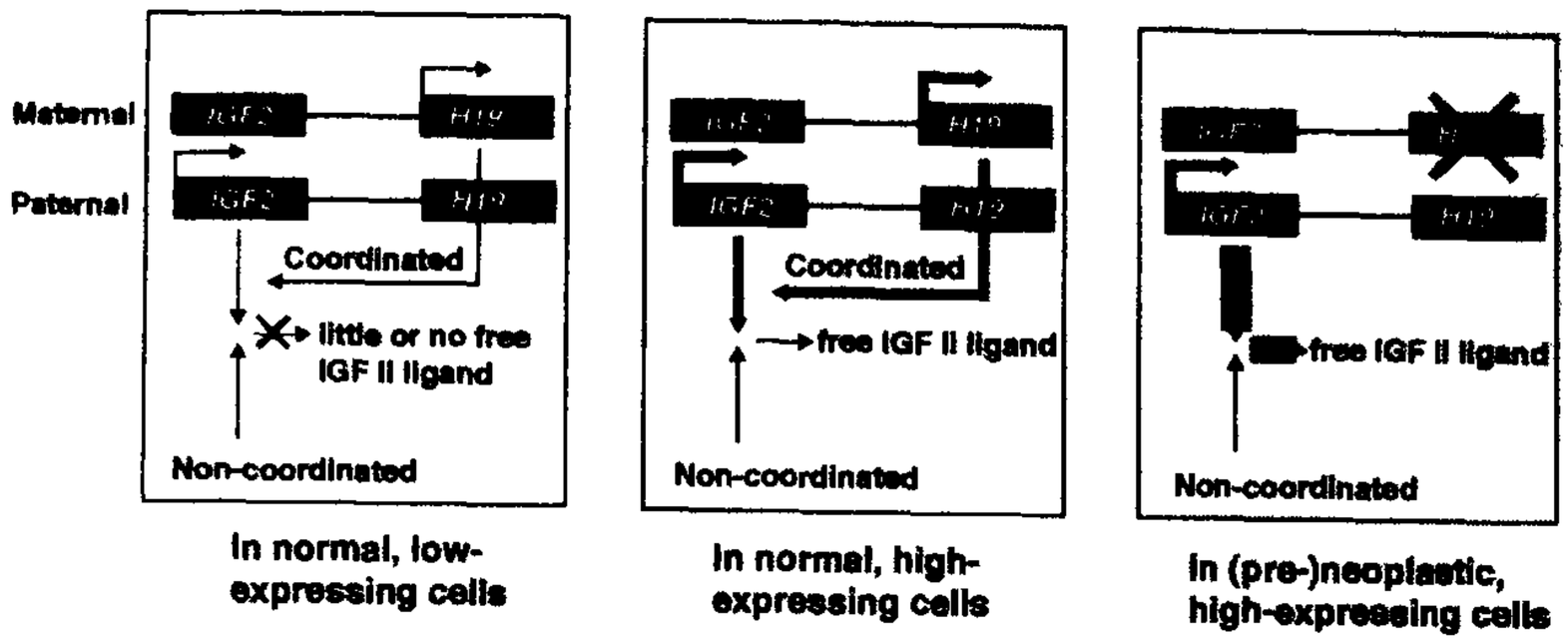

Figure 1. A model for the postulated *H19*-specific control loop. Whereas the *H19*-specific control would be coordinated with *IGF2* activity, other levels of controls, represented by the IGF-binding proteins, for example, may be uncoordinated. In the normal context with cells displaying low levels of *IGF2* mRNAs, the noncoordinated controls would be expected to successfully prevent production of free IGF II ligand. The coordinated control, represented by *H19* transcripts, will be of minor importance because the levels of and pattern of *H19* expression closely follows those of *IGF2* (ref. 33). By the same token, increased activity of the *IGF2* gene will be followed by increased activity of the *H19* gene, albeit from opposite parental alleles. In this scenario, the coordinated *H19* control would gain importance in direct proportion to the inability of the uncoordinated controls to deal with abnormal levels of *IGF2* activity. When the *H19* function is abnormally silenced in a situation with high levels of *IGF2* transcripts, as has been documented in numerous contexts, the model suggests that production of free IGF II ligand will increase abnormally to contribute to overgrowth syndromes, such as the Beckwith-Wiedemann syndrome, and neoplasia (outlined in the right-most panel).

expression would normally be accompanied by high levels of *H19* expression, since their expression patterns are coordinated[24,36]. By extrapolation, the higher the levels of *IGF2* expression the more important the *H19* function becomes. According to this model, a loss of the *H19* function in *trans* would be expected to be a key event in cells expressing high levels of *IGF2* mRNAs, subsequently increasing production of free IGF II ligands to a significant degree. Our model would highlight the consequences of losing the *H19* function in *trans* when *IGF2* is overexpressed. In a parallel study, we have been able to show that this loss of *H19* expression is an early event that potentially predisposes for Wilms' tumours[34]. Another interesting case is the previous documentation that *IGF2* could not be genetically linked with a familial form of BWS[37], in spite of the close link between BWS and *IGF2*[38,39]. The possibility that *H19* can be the direct and *IGF2* the indirect culprit in this cancer-predisposing disease, at least in some instances, would appear to be compatible with the proposed model.

Has the coding potential of the *H19* transcript been rigorously tested?

The documentation that the *H19* transcript is associated with polysomes is compatible with either a non-coding or a coding *H19* transcript. It could be envisaged, for example, that the *H19* transcript interacts with mRNP complexes or with ribosomes to modulate translatability or stability of target transcripts. This possibility is suggested by the observation that the major IGF2 tra and *H19* transcript co-sediment in sucrose gradi Alternatively, the *H19* transcript could be transl produce a regulatory protein product. Althou murine and human genes have a similar structur exons and 4 small introns and a significant homol the primary sequence level, it has been claime there is no consensus ORF[2]. A closer inspection sequence, however, reveals that one of the smaller encodes a hypothetical protein which displays so quence similarity between murine and human *H1* scripts (Figure 2 *c*, *d*). It is also striking that the *H19* transcript has an ORF for a hypothetical pro 26 kd (Figure 2 *a*). This product would be high usual if it exists, with an isoelectric point of 11 sequence similarities with RNA-binding proteins. are other potential ORFs, however, and a splice has been observed which lacks exon 4. This v which is genotype-specific and expressed in only humans, produces a new ORF (Figure 2 *b*). It is known if a corresponding splice variant with a ORF is expressed in the mouse. The possibility least some of the *H19* transcripts display an OR produces a protein cannot, therefore, be ruled out.

The origin of non-coding transcripts: The gene hypothesis

When considering the functional aspects of non-pol II transcripts, it might be fruitful to consider general understanding of their origin. One pos

A

MGPFQAERARGQGGSTGVASVGSSTWGGTPGLGQTGTWQGTQDRGVPSCHLTHRNSFSSRHRGSS
GTAFSGLCRSLEGWRAGRQAVLGELQQDVTRRAKRPREGGPGTLRSKEAAGARPAGAGPTRHCGP
GSGAERAHSGLGDAVCPHRSAPGAGLADSTASRGVKGMQRDQTRRGGRGGVNELSGGRMVQAGVR
SAAGGEREALASRARGQGGPRGRRRSRDRCLSVRAGDEARSPAGVDVPTSCRRPRRQVRWT

B

MAALWAPRTHNMKGEARGTRFHQPNIKDTIGTAAPAAPTPHRRGLHLHGHPLRWTVDHQPFHHPRA
ELLQRDDAVPTTSLFFFFILLSLCF

C

MGWRLVVEAVCSFTFPKELTLLCCSLDPPPPTLNPQDERNGATQLMSGPLNPGTSLSPSRSESRR
CCNQNHYTTCLRNLLQGEAERTDGVNILKEQTHSTHPPLRIHLHGQLCLTRETTTHIILEPSLYP
GMTSSSPSCLFLLFPFL

D

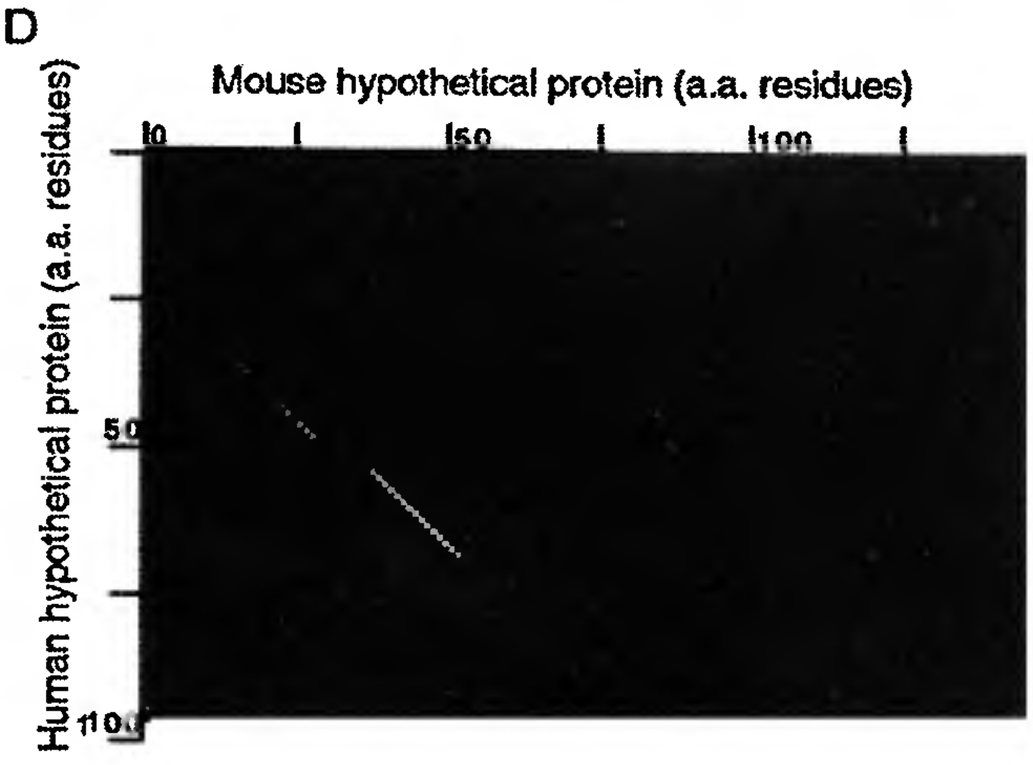

Figure 2. Hypothetical H19 proteins and their sequence similarity. Panel A depicts the amino acid sequence of the human hypothetical H19 protein derived from the long ORF. Panel B shows the amino acid sequence of the new ORF found only in a splice variant of the human *H19* transcript. Panel C shows the amino acid sequence of the hypothetical H19 protein derived from the long ORF in the mouse *H19* transcript. Panel D shows a 2D dot matrix analysis of sequence similarities of the hypothetical mouse H19 protein with one of the minor ORFs in the major human *H19* transcript.

that comes to mind is that the requirements for any sequence to make up a 'gene' that is transcribed at any given time are not very stringent. It includes a promoter, a cap site and a transcription termination signal that is generally associated with a polyA addition signal. Each of these requirements involves only a limited amount of sequence information. Gene-searching program using these algorithms are currently being applied to decipher the primary sequences of whole genomes and have been instrumental in identifying numerous, previously unknown genes.

It is not difficult to perceive that the majority of non-coding genes will be located in gene clusters. The reason for this postulate reflects the ability of locus control regions to regulate transcription of multiple genes in a long-range manner within subchromosomal domains[40]. By extrapolation, any sequence within a gene cluster could be transcriptionally activated, depending on the stochastic juxtaposition of simple sequence elements which make up a promoter and a transcriptional start site, particularly since such elements exhibit fairly loose sequence stringency. This model of the fortitious generation of 'transcriptional noise' genes is in keeping with the increased number of non-coding transcripts which emanate from gene clusters, as accounted for above. Although such 'noise' genes have been postulated to become silenced by CpG methylation[41], some may escape the effects of the epigenetic modification. It is in this context interesting that many of the non-coding transcripts, such as *Igf2* AS, *Igf2r* AS and *UBE3A* transcripts, originate from methylated alleles.

Given genetic background differences, the optimal arrangement of crucial *cis* elements may be found in only a subpopulation of the species. The *IGF2R* gene, which is expressed in a parent of origin-dependent manner in only some individuals[42], may provide a particularly interesting case. Could it be that the *IGF2R* antisense transcript, which is believed to prevent the paternal *IGF2r* allele from being expressed, is absent in many humans? Alternatively, the population-dependent sequence composition could generate different versions of non-coding transcripts. For example, a splice variant of the *H19* gene which generates a new potential open reading frame, can only be found in a subpopulation of humans[43]. Moreover, the genotype and epigenotype synergize to regulate the spatial distribution of the alternatively spliced *H19* transcript in the developing human placenta[43]. This synergy between genotype and epigenotype may increase the amplitude of phenotype variation which could underlie the selection of particular functions for non-coding as well as coding transcripts. Whether or not the *H19* gene belongs to this category of genes is open to question. We would like to emphasize, however, that it cannot currently be ruled out that the *H19* gene has a coding potential in some species or even only in some subpopulations within a species. A more careful assessment of the open reading potential of the *H19* gene and other non-coding genes in a number of individual mammals with different genetic backgrounds is warranted.

Conclusion

A consensus view is emerging that non-coding transcripts are abundant in gene clusters which belong to imprinted domains. These transcripts may arise fortuitously due to long-range acting locus control regions and/or play a fundamental role in the imprinting process. This is underscored by the fact that these transcripts are often transcribed from the opposite strand of protein-coding genes and that the non-coding and coding genes are active from either of the parental alleles. It is uncertain if the *H19* gene belongs to this class of non-coding genes, since its transcript is not known to be antisense to any protein-coding gene and it cannot be excluded that in some instances may be translated to yield a protein product.

1. Erdmann, A. V., Szymanski, M., Hochberg, A., De Groot, N. and Barciszewski, J., *Nucleic Acids Res.*, 1999, **27**, 192–195.
2. Brannan, C., Dees, E., Ingram, R. and Tilghman, S., *Mol. Cell. Biol.*, 1990, **10**, 28–36.
3. Wevrick, R., Kerns, J. A. and Francke, U., *Hum. Mol. Genet.*, 1994, **3**, 1877–1882.
4. Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. and Rastan, S., *Cell*, 1992, **71**, 515–526.
5. Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. and Davis, R. K., *Cell*, 1997, **88**, 445–457.
6. Liu, A. Y., Torchia, B. S., Migeon, B. R. and Siliciano, R. F., *Genomics*, 1997, **39**, 171–184.
7. Nickerson, J. A., Blencowe, B. J. and Penman, S., *Int. Rev. Cytol.*, 1995, **A162**, 67–123.
8. Wutz, A., Smrzka, O., Schweifer, N., Schellander, K., Wagner, E. and Barlow, D., *Nature*, 1997, **389**, 745–749.
9. Lee, J. T., Davidow, L. S. and Warshawsky, D., *Nature Genet.*, 1999, **21**, 400–404.
10. Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H. and Reik, W., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 12509–12514.
11. Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, K. and Lalande, M., *Nature Genet.*, 1998, **19**, 15–16.
12. Jong, M. T., Gray, T. A., Ji, Y., Glenn, C. C., Saitoh, S., Driscoll, D. J. and Nicholls, R. D., *Hum. Mol. Genet.*, 1999, **8**, 783–793.
13. Lee, M. P., Hu, R. J., Johnson, L. A. and Feinberg, A. P., *Nature Genet.*, 1997, **15**, 181–185.
14. Barlow, D. P., *EMBO J.*, 1997, **16**, 6899–6905.
15. Wevrick, R., Kerns, J. and Francke, U., *Acta Genet. Med. Gemellol. (Roma)*, 1996, **45**, 191–197.
16. Dittrich, B., Buiting, K., Korn, B., Rickard, S., Buxton, J., Saitoh, S., Nicholls, R. D., Poustka, A., Winterpacht, A., Zabel, B. and Horsthemke, B., *Nature Genet.*, 1996, **14**, 163–170.
17. Rougeulle, C., Glatt, H. and Lalande, M., *Nature Genet.*, 1997, **17**, 14–15.
18. Hu, J. F., Vu, T. H. and Hoffman, A. R., *J. Biol. Chem.*, 1997, **272**, 20715–20720.
19. Brockdorff, N., *Curr. Opin. Genet. Dev.*, 1998, **8**, 328–333.
20. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. and Brockdorff, N., *Nature*, 1996, **379**, 131–137.
21. Sutcliffe, J. S., Nakao, M., Christian, S., Örstavik, K. H., Tommerup, N., Ledbetter, D. H. and Beaudet, A. L., *Nature Genet.*, 1994, **8**, 52–58.
22. Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D. and Horsthemke, B., *Nature Genet.*, 1995, **9**, 395–400.
23. Saitoh, S., Buiting, K., Rogan, P., Buxton, J., Driscoll, D., Arnemann, J., König, R., Malcolm, S., Horsthemke, B. and Nicholls, R., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 7811–7815.
24. Leighton, P., Saam, J., Ingram, R., Stewart, C. and Tilghman, S., *Genes Dev.*, 1995, **9**, 2079–2089.
25. Poirier, F., Chan, C.-T., Timmons, P., Robvertson, E., Evans, M. and Rigby, P., *Development*, 1991, **113**, 1105–1114.
26. Ripoche, M., Kress, C. F. P. and Dandolo, L., *Genes Dev.*, 1997, **11**, 1596–1604.
27. Bartolomei, M. E. and Tilghman, S. M., *Sem. Dev.. Biol.*, 1992, **3**, 107–117.
28. Pfeifer, K., Leighton, P. and Tilghman, S. M., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 13876–13883.
29. Jones, B. K., Levorse, J. M. and Tilghman, S. M., *Genes Dev.*, 1998, **12**, 2200–2207.
30. Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. and M, T. S., *Nature*, 1995, **375**, 34–39.
31. Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E. and Tycko, B., *Nature*, 1993, **365**, 764–767.
32. Rachmilewitz, J., Elkin, M., Rosensaft, J., Gelman-Kohan, Z., Ariel, I., Lustig, O., Schneider, T., Goshen, R., Biran, H., De Groot, N. and Hochberg, A., *Oncogene*, 1995, **11**, 863–870.
33. Li, Y.-M., Franklin, G., Cui, H. M., Svensson, K., He, X. B., Adam, G., Ohlsson, R. and Pfeifer, S., *J. Biol. Chem.*, 1998, **273**, 28247–28252.
34. Cui, J., Hedborg, F., He, L., Nordenskjöld, A., Sandstedt, P.-O. S. and Ohlsson, R., *Cancer Res.*, 1997, **57**, 4469–4473.
35. Stewart, C. E. and Rotwein, P., *Physiol. Rev.*, 1996, **76**, 1005–1026.
36. Ohlsson, R., Hedborg, F., Holmgren, L., Walsh, C. and Ekström, T. J., *Development*, 1994, **120**, 361–368.
37. Nyström, A., Hedborg, F. and Ohlsson, R., *Eur. J. Pediatr.*, 1994, **153**, 574–580.
38. Reeve, A. E., *Med. Pediatr. Oncol.*, 1996, **27**, 470–475.
39. Hedborg, F., Holmgren, L., Sandstedt, B. and Ohlsson, R., *Am. J. Pathol.*, 1994, **145**, 802–817.
40. Evans, T., Felsenfeld, G. and Reitman, M., *Annu. Rev. Cell Biol.*, 1990, **6**, 95–124.
41. Bird, A. and Tweedie, S., *Phil. Trans. R. Soc. London*, 1995, **B349**, 249–253.
42. He, L., Cui, H., Walsh, C., Mattsson, R., Lin, W., Anneren, G., Pfeifer-Ohlsson, S. and Ohlsson, R., *Oncogene*, 1998, **16**, 113–119.
43. Lin, W. L., He, X. B., Svensson, K., Adam, G., Li, Y. M., Tang, T. W., Paldi, A., Susan, P. and Ohlsson, R., *Mech. Dev.*, 1999, **82**, 195–197.

ACKNOWLEDGEMENTS. We thank Dr Gary Franklin for valuable advice. This work was supported by the Swedish Cancer Research Foundation (RmC/CF), The Swedish Pediatric Cancer Foundation (BCF) and the Swedish Natural Science Research Council (NFR).