Histol Histopathol (2002) 17: 677-682

http://www.hh.um.es

Histology and Histopathology

Cellular and Molecular Biology

Review

Intergenic mRNAs. Minor gene products or tools of diversity?

C. Finta, S.C. Warner and P.G. Zaphiropoulos

Center for Nutrition and Toxicology, Department of Biosciences at NOVUM, Karolinska Institute, Huddinge, Sweden

Summary. The general model of gene expression implies that the units of genetic information, the genes, constitute the basis of the mRNA and protein products detected in living organisms. It is well known that in eukaryotic cells a single gene may give rise to various gene products due to alternative splicing and/or different positions of transcriptional start and termination. However, recent experimental results suggest that in addition to this variation in gene expression, another level of complexity may also exist as evidence for the presence of mRNAs that combine sequence information (exons) from distinct genes is accumulating. Moreover, the mechanisms that allow this production of intergenic mRNAs are starting to unravel and it appears that these follow two general pathways: (a) bypass of transcriptional termination, resulting in the generation of bicistronic pre-mRNAs; and (b) authentic *trans*-splicing events between pre-mRNAs of distinct genes.

Key words: Alternative splicing, *Trans*-splicing, Pre-mRNA, Evolutionary potential, Gene definition

Introduction

With the completion of the Human Genome Sequencing Project the number of human genes has been estimated to be in the range of 30000 to 40000 (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001). This has been achieved based on computer algorithms that scan genomic sequences and identify elements that are characteristic of functional genes. However, a recent comparison of the predicted genes between the two sequencing efforts revealed only a little overlap, casting therefore doubt on these low estimates of human gene numbers (Hogenesch et al., 2001). Moreover, predictions on the number of human genes based on methods that take advantage of assemblies of transcripts present in the Expressed Sequence Tag (EST) databases suggest a number closer to 70000 or even 120000 (Liang et al., 2000; Wright et al., 2001). This lack of consistency highlights the difficulty of providing a definite answer on the human gene number, reveals inherent limitations in the methods employed, and ultimately raises the question as to how a gene is defined and thus recognized (Aparicio, 2000).

Offprint requests to: Peter G. Zaphiropoulos, Center for Nutrition and Toxicology, Department of Biosciences at NOVUM, Karolinska Institute, SE-14157 Huddinge, Sweden. Fax: +46-8-6081501. e-mail: peter.zaphiropoulos@cnt.ki.se

Evaluation of the EST data as indicative of transcriptional events

It is generally thought that if a certain sequence is not present in the EST database or not linked to an EST clone then this sequence is not expressed. To test this hypothesis and evaluate the completeness of the EST database, an analysis using genes positioned in the recently characterized human cytochrome P450 (*CYP3A*) locus on chromosome 7q21-q22.1 (Finta and Zaphiropoulos, 2000b; Gellner et al., 2001) was performed. The three most abundantly expressed *CYP3A* genes in human liver are *CYP3A4*, *CYP3A5* and *CYP*3A7. Therefore, we asked ourselves how many EST clones represent these genes in the EST database. As a query sequence we used 260 bases from the 3´ terminal exon, exon 13, of these genes. The results revealed 105 hits that have significant similarity (score of 84 or above - search of August 10) implying that these represent expressed sequences from the three *CYP3A* genes.

Recently, a fourth *CYP3A* gene was identified: *CYP3A43* (Domanski et al., 2001; Gellner et al., 2001; Westlind et al., 2001). When a similar search with a *CYP3A43* query was performed only a single hit was identified. The results of the EST analysis therefore suggest that *CYP3A43* is expressed at about two orders of magnitude lower than the more abundant *CYP3A* genes, a value that is in agreement with the direct determination of the *CYP3A43* expression level (Gellner et al., 2001; Westlind et al., 2001).

Do these results then imply that there are no additional expressed sequences from the *CYP3A* locus? Certainly not. By using highly sensitive techniques based on the Reverse Transcriptase/Polymerase Chain

Reaction (RT/PCR) technology a CYP3A7 variant mRNA was detected that uses an alternative 3' end sequence (Finta and Zaphiropoulos, 2000b). This variant is expressed in human liver at levels lower than CYP3A43 and accordingly is not represented in the EST database. Additional variant mRNAs from the *CYP3A* locus containing CYP3A43 and CYP3A4 sequences have also been detected and are expressed in human liver at about two orders of magnitudes lower than CYP3A43 (Finta and Zaphiropoulos, 2002). As expected, these mRNAs are also absent from the EST database.

Therefore, it appears that the EST database contains only the most abundantly expressed transcripts. However, RNA molecules of lower abundance are not identified by these "global" approaches. To detect these RNAs the employment of more gene-specific or locus-specific techniques is needed.

Intergenic mRNAs as products of bicistronic transcripts

Bypassing transcriptional termination

The general model of transcriptional termination by RNA polymerase II suggests that this is a rather random process occurring at sites between 200-2000 bp downstream of the poly(A) site (Dye and Proudfoot, 2001; Tran et al., 2001). Let us examine how this model applies in the case of transcription of the *CYP3A7* gene (Fig. 1). The major mRNAs identified employ two poly(A) signals positioned 343 and 440 bp downstream of the translation termination signal in exon 13. However, some mRNAs have been detected that contain an altered 3' end sequence. This is provided by sequences from exon 2 and exon 13 of the *CYP3AP1* gene, which are positioned about 10 kb and 20 kb downstream of the *CYP3A7* gene respectively. The resulting variant CYP3A7 mRNA molecule has substituted the last four codons of *CYP*3A exon 13 with 36 codons originating from exons 2 and 13 of the *CYP3AP1* gene. Interestingly, the *CYP3AP1* gene is composed of only 3 exons instead of the typical 13 exons of the *CYP*3A genes and formally represents a "pseudogene". No mRNA product composed solely of CYP3AP1 sequences has been detected, as the *CYP3AP1* promoter is apparently inactive. Therefore, the variant CYP3A7 mRNA containing exons 2 and 13 of *CYP3AP1* is the product of a CYP3A7 pre-mRNA that has bypassed canonical CYP3A7 transcriptional termination and extended into the *CYP3AP1* gene (Finta and Zaphiropoulos, 2000b).

These findings raise the question as to whether such a bypass of transcriptional termination is a unique phenomenon of the *CYP3A7* gene or may be a more widespread event that also occurs with other genes. In this context it is worth mentioning certain gene products generated by the cluster of the human *CYP2C* genes on chromosome10q24. The human *CYP2C18* gene is composed of nine exons and is positioned upstream of the *CYP2C19* gene in the same 5' to 3' orientation, with approximately 27 kb intergenic sequence (GenBank entry: AL583836, see also next section). In addition to the typical CYP2C18 and CYP2C19 mRNAs, a polyadenylated transcript containing exon 8 of

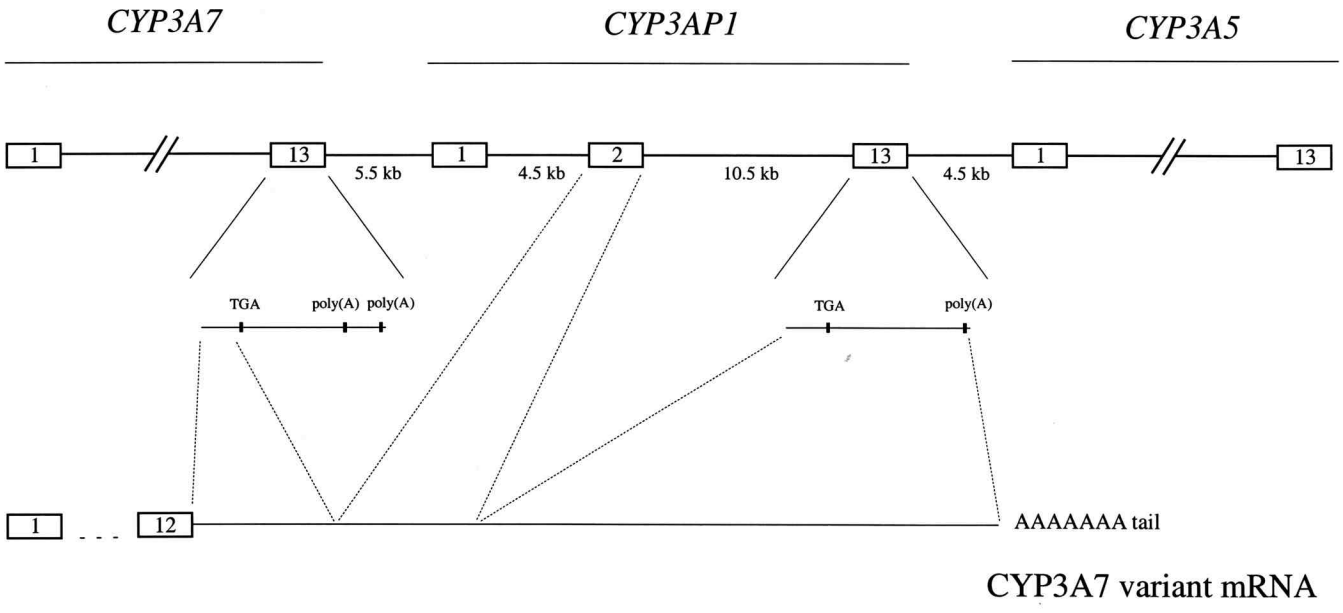

Fig. 1. Schematic diagram of the *CYP3A* locus encompassing the *CYP3A7*, *CYP3AP1* and *CYP3A5* genes. Exons are indicated by boxes and are numbered. The *CYP3A7* and *CYP3AP1* exons 13 are shown in detail with the translation termination signals TGA and the poly(A) signals indicated. The segment of the *CYP3A* locus that results in the variant 3' end of the *CYP3A7* mRNAs is highlighted.

CYP2C18 spliced to exon 2 of CYP2C19 has been detected in human epidermis and liver (Zaphiropoulos, 1999). The most plausible explanation for this finding is that certain 2C18 transcripts have extended beyond the borders of the *CYP2C18* gene into the downstream *CYP2C19* gene.

Thus, it may be that in addition to the predominant transcriptional termination events that generally occur within 1 to 2 kb of the major poly(A) sites, some transcripts may extend further downstream and encompass sequences from neighboring genes. Consequently, poly(A) sites provided by the neighboring genes are used in these bicistronic transcripts. The presence in single pre-mRNA molecules of sequences originating from more than one gene may give rise to splicing events that bring in juxtaposition exons from neighboring genes (Fig. 2, panel A). This type of intergenic splicing is not only restricted to CYP3A7/CYP3AP1 or CYP2C18/CYP2C19 but has also been reported in several additional cases (Benson et al., 1995; Fears et al., 1996; Magrangeas et al., 1998; Kowalski et al., 1999; Moore et al., 1999; Krauss and Reuter, 2000; Millar et al., 2000; Thomson et al., 2000; Communi et al., 2001; Courseaux and Nahon, 2001). Moreover, intergenic transcription has been observed in some gene clusters (Ashe et al., 1997; Rogan et al., 1999; Gribnau et al., 2000; Plant et al., 2001) raising the interesting possibility as to whether it may be followed by intergenic splicing.

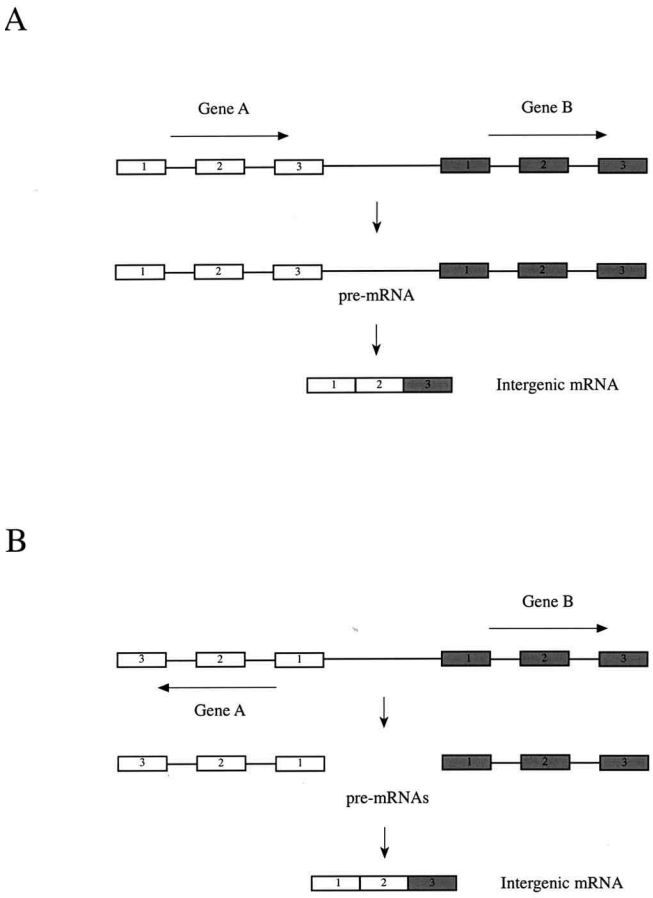

Fig. 2. Schematic representation of how relative gene orientation provides clues to the mechanisms that result in the generation of intergenic mRNAs. Horizontal arrows indicate the orientation of the genes, and boxes the exons. **Panel A.** Intergenic mRNAs resulting from genes with the same 5' to 3' orientation. A bypass of transcriptional termination of Gene A results in a bicistronic pre-mRNA that is followed by splicing of exon 2 of Gene A to exon 3 of Gene B. **Panel B.** Intergenic mRNAs resulting from genes with opposite 5' to 3' orientation (coded by different DNA strands). The cotranscriptional interpretation of panel A is not applicable in this case. Thus, splicing between distinct pre-mRNAs from Gene A and Gene B, i.e. *trans*-splicing, has to be postulated. Note that, in principle, *trans*-splicing may also interpret the case of panel A.

Alternative promoters

When the results described above are viewed from the perspective of the downstream instead of the upstream gene, a question that is likely to be raised may be formulated as follows: does transcriptional initiation solely occur about 20-25 bases downstream of canonical TATA boxes or can some transcripts significantly extend 5' of the major TATA box? For this analysis we will concentrate on the *CYP2C* genes. About a decade ago transcription initiation of the major rat *CYP2C* genes was determined by primer extension analysis and was found to be about 25 bases downstream of the TATA box, as expected (Morishima et al., 1987; Zaphiropoulos et al., 1990; Eguchi et al., 1991). It was therefore a surprise when an additional rat *CYP2C* gene, *CYP2C24*, was analyzed in its 5' region by the more sensitive RACE technique (Rapid Amplification of cDNA Ends) that resulted in the identification of transcripts that extended beyond the expected transcriptional start into the TATA box and further upstream (Zaphiropoulos, 1993). More recently, analysis of a solitary *CYP2C* exon 1 in the human *CYP2C* locus, which is positioned 33 kb upstream of the *CYP2C9* gene in the same orientation, revealed mRNAs that contain combinations of that exon with CYP2C9 exons (Warner et al., 2001). This implies that some transcripts may not start at about 25 bases downstream of the canonical TATA box of *CYP2C9* gene, but that by the fact that these include the solitary exon 1 some 33 kb upstream are likely to initiate at that exon 1 promoter.

Intergenic mRNAs as products of *trans*-splicing

Clearly bypassing transcriptional termination/alternative promoter usage provides the means that allows sequences from neighboring genes to be present in the same pre-mRNA molecule. However, a pertinent question is whether the presence of exons from different genes in a single pre-mRNA molecule is a prerequisite for the juxtaposition event, which is the joining of these exons. The clearest example that this may not necessarily be true is provided by the phenomena of

exon repetition in three mammalian genes (Caudevilla et al., 1998; Akopian et al., 1999; Frantz et al., 1999). In these cases, mRNA molecules were identified where exons were present more than once in the same pre-mRNA molecule. No partial duplication has been detected at the genomic DNA level. Assuming that the RNA polymerase complex faithfully transcribes the DNA without any flip-back; that is that every DNA segment is represented only once in the pre-mRNA, then the only interpretation for these findings has to be a splicing process between two pre-mRNAs, which is *trans*-splicing.

Therefore, if pre-mRNA molecules from the same gene may interact in *trans*-splicing processes resulting in highly complex mRNAs, then it is logical to ask whether pre-mRNAs from neighboring but distinct genes are also involved in similar *trans*-splicing events. However, the mere detection of hybrid mRNAs from neighboring genes does not necessarily imply *trans*-splicing because this could be the result of cotranscriptional splicing, which is a bypass of transcriptional termination / alternative promoter usage as discussed above. An unambiguous evidence that intergenic mRNAs are the result of splicing of distinct pre-mRNAs is provided when the genes in questions are coded by different DNA strands (Fig. 2, panel B). This is exactly the case of the CYP3A43/CYP3A4 and CYP3A43/CYP3A5 hybrid mRNAs detected. The *CYP3A43* gene is coded by the opposite strand to the one that codes the *CYP3A4* and *CYP3A5* genes (Finta and Zaphiropoulos, 2002).

An analysis of the expression pattern of the *CYP*2C class of cytochrome P450 genes has also been performed. This provided evidence for the presence of hybrid mRNA molecules containing exons from more than one *CYP2C* gene, including certain exonic combinations originating from the genes which are at the two ends of the cluster; *CYP2C8* and *CYP2C18* (Finta and Zaphiropoulos, 2000a). The relative orientation of these two outer genes was deduced by the analysis of two Yeast Artificial Chromosomes (YACs) suggesting the same orientation. However, comparison of sequences recently available from the Human Genome Project (GenBank entries AL133513, AL157834, AL157835, AL359672 and AL583836) indicate that the orientation of *CYP2C8* is opposite to the other three genes, *CYP2C9*, *CYP2C19* and *CYP2C18*. This inconsistency between the two approaches may be the result of the unreliability of the YACs, as these are prone to rearrangements and internal deletions. Moreover, the sequencing data indicate that the *CYP2C8* gene is in a tail-to-tail orientation with *CYP2C9* with approximately 48 kb of intergenic sequence (GeneBank entry AL359672). Therefore, it appears that the hybrid CYP2C8/CYP2C18 mRNAs identified are also likely to be the result of *trans*-splicing events.

An additional case of *trans*-splicing has recently been reported in the *Drosophila mod(mdg4)* locus where exons from independent transcripts are combined to generate mature mRNAs (Dorn et al., 2001; Labrador et al., 2001).

Intergenic mRNAs as a source of complexity in gene expression

The apparently low number of human genes relative to model organisms such as *Drosophila melanogaster* (Adams et al., 2000) and *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998) has raised the question as to how to ascribe the complexity of the human species relative to these simpler organisms. In addition, is *C. elegans* with a total gene number of 19000 more complex than *Drosophila* which has only 14000 genes? One way out of this may be that a single gene can give rise to many more mRNAs and consequently proteins in higher eukaryotes than in simpler organisms (Graveley, 2001). Moreover, the discussion in the above sections implies that in addition to the variations in splicing within single genes, another level of complexity may also occur that encompasses splicing combinations between distinct, often neighboring genes. Even sequences that do not have the characteristics of functional genes, such as the *CYP3AP1* or the solitary *CYP2C* exon 1, may also be used in these highly complex transcripts. Thus, the number of unique mRNAs that can be produced in *H. sapiens* may be much higher than *Drosophila*, which in turn could produce more mRNAs than *C. elegans*. At this point it is worth mentioning that although significant progress in unraveling the mechanisms of pre-mRNA splicing has been made, the regulation of this process, that is, how a certain pre-mRNA sequence is recognized as an exon and spliced to another exon in one cellular context but not in another, has not been fully unraveled (Elliott, 2000; Smith and Valcárcel, 2000).

The importance of strict definitions

As discussed in the introduction, the number of human genes based on computer algorithms that scan genomic sequences appears to be surprisingly low. However, the use of different algorithms results in different sets of predicted genes, implying an inherent uncertainty in this methodology (Hogenesch et al., 2001). Moreover, the number of genes based on the transcriptional information present in the EST databases gives generally higher values. Furthermore, we believe it is evident to the reader of this article that the human transcriptome contains a lot of additional sequences that are not represented in the EST database. Therefore, the relation between a gene, as defined by the computer algorithms, and the detected unique transcripts containing sequences from that gene, is not one to one. It may be at least an order of magnitude if not more in favor of the unique transcripts. For example, the *CYP3AP1* gene discussed above is formally a pseudogene. It is missed by the computer algorithms but it has a role in diversifying the expression of the *CYP3A7* gene. The same is true for the *CYP2C* solitary

exon 1. It is not expressed on its own but allows variation in the *CYP2C9* expression. Thus, sequences that may not be considered as "genes" have a role in the expression potential of the human genome. Moreover, combinations of exons from distinct functional genes, such as *CYP3A43/CYP3A4*, *CYP2C8/CYP2C18*, or *CYP2C18/CYP2C19*, may result in additional variations in gene expression. In fact, this process is not limited to the production of bicistronic pre-mRNAs, but can also employ *trans*-splicing events. Strictly speaking, based on the observation of these transcriptional complexities, the term “gene” may not accurately represent the unit of expressed genomic information as combinations between parts of distinct genes do occur. Moreover, sequences outside genes (as defined by the computer algorithms) may also be expressed (Finta and Zaphiropoulos, 2000c). Consequently, it has been proposed that a more appropriate term for these complicated patterns of transcriptional activities may be “genome transcription” (Finta and Zaphiropoulos, 2001).

Certainly, it could be argued that as these transcripts are generally of low frequency, no significant biological function is associated with them. Apart from the lack of a direct correlation between level of expression and biological significance (e.g. knockout mouse models of some low frequency mRNAs result in more dramatic phenotypes than others of higher frequency) the presence of a large collection of rare mRNAs may safeguard an optimal response to a constantly changing environment. As the future is unpredictable and some genotypes that are currently considered unfavorable they may provide advantages under different conditions (Brosius and Kreitman, 2000), in the same way as that rare mRNAs that are not clearly linked with a specific function at present may prove to be indispensable in the future. It could even be argued that this plasticity in gene expression of higher eukaryotes versus the apparent rigidity of prokaryotes (lack of introns and alternative splicing) may have provided these organisms with significant evolutionary advantages.

Acknowledgements. Funding for the research performed in our laboratory was provided by the foundations Åke Wiberg, Lars Hiertas Minne, Tullberg, Otto Swärd, and the Karolinska Institute. Csaba Finta is a recipient of a postdoctoral fellowship from the Swedish Cancerfonden.

References

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F. et al. (195 authors) (2000). The genome sequence of *Drosophila melanogaster*. Science 287, 2185-2195.

Akopian A.N., Okuse K., Souslova V., England S., Ogata N. and Wood J.N. (1999). Trans-splicing of a voltage-gated sodium channel is regulated by nerve growth factor. FEBS Lett. 445, 177-182.

Aparicio S.A. (2000). How to count ... human genes. Nature Genet. 25, 129-130.

Ashe H.L., Monks J., Wijgerde M., Fraser P. and Proudfoot N.J. (1997). Intergenic transcription and transinduction of the human ß-globin locus. Genes Dev. 11, 2494-2509.

Benson G.V., Nguyen T.-H.E. and Maas R.L. (1995). The expression pattern of the murine Hoxa-10 gene and the sequence recognition of its homeodomain reveal specific properties of *Abdominal B*-like genes. Mol. Cell. Biol. 15, 1591-1601.

Brosius J. and Kreitman M. (2000). Eugenics-evolutionary nonsense? Nature Genet. 25, 253.

Caudevilla C., Serra D., Miliar A., Codony C., Asins G., Bach M. and Hegardt F.G. (1998). Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. Proc. Natl. Acad. Sci. USA 95, 12185-12190.

Communi D., Suarez-Huerta N., Dussossoy D., Savi P. and Boeynaems J.M. (2001). Cotranscription and intergenic splicing of human *P2Y11* and *SSF1* genes. J. Biol. Chem. 276, 16561-16566.

Courseaux A. and Nahon J.L. (2001). Birth of two chimeric genes in the Hominidae lineage. Science 291, 1293-1297.

Domanski T.L., Finta C., Halpert J.R. and Zaphiropoulos P.G. (2001). cDNA cloning and initial characterization of CYP3A43, a novel human cytochrome P450. Mol. Pharmacol. 59, 386-392.

Dorn R., Reuter G. and Loewendorf A. (2001). Transgene analysis proves mRNA trans-splicing at the complex *mod(mdg4)* locus in *Drosophila*. Proc. Natl. Acad. Sci. USA 98, 9724-9729.

Dye M.J. and Proudfoot N.J. (2001). Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase. Cell 105, 669-681.

Eguchi H., Westin S., Ström A., Gustafsson J.-Å. and Zaphiropoulos, P.G. (1991). Gene structure and expression of the rat cytochrome P450IIC13, a polymorphic, male-specific cytochrome in the P450IIC subfamily. Biochemistry 30, 10844-10849.

Elliott D.J. (2000). Splicing and the single cell. Histol. Histopathol. 15, 239-249.

Fears S., Mathieu C., Zeleznik-Le N., Huang S., Rowley J.D. and Nucifora G. (1996). Intergenic splicing of *MDS1* and *EVI1* occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family Proc. Natl. Acad. Sci. USA 93, 1642-1647.

Finta C. and Zaphiropoulos P.G. (2000a). The human *CYP2C* locus: A prototype for intergenic and exon repetition splicing events. Genomics 63, 433-438.

Finta C. and Zaphiropoulos P.G. (2000b). The human cytochrome P450 3A locus. Gene evolution by capture of downstream exons. Gene 260, 13-23.

Finta C. and Zaphiropoulos P.G. (2000c). Complexity of gene expression – Evolutionary implications. Recent Res. Dev. Mol. Cell. Biol. 1, 109-116.

Finta C. and Zaphiropoulos P.G. (2001). A statistical view of genome transcription? J. Mol. Evol. 53, 160-162.

Finta C. and Zaphiropoulos P.G. (2002). Intergenic mRNA molecules resulting from *trans*-splicing. J. Biol. Chem. 277, 5882-5890.

Frantz S.A., Thiara A.S., Lodwick D., Ng L.L., Eperon I.C. and Samani N.J. (1999). Exon repetition in mRNA. Proc. Natl. Acad. Sci. USA 96, 5400-5405.

Gellner K., Eiselt R., Hustert E., Arnold H., Koch I., Haberl M., Deglmann C.J., Burk O., Buntefuss D., Escher S., Bishop C., Koebe H.G., Brinkmann U., Klenk H.P., Kleine K., Meyer U.A. and Wojnowski L. (2001). Genomic organization of the human *CYP3A* locus: identification of a new, inducible *CYP3A* gene. Pharmacogenetics 11, 111-121.

Graveley B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. Trends Genet. 17, 100-107.

Gribnau J., Diderich K., Pruzina S., Calzolari R. and Fraser P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human ß-globin locus. Mol. Cell 5, 377-386.

Hogenesch J.B., Ching K.A., Batalov S., Su A.I., Walker J.R., Zhou Y., Kay S.A., Schultz P.G. and Cooke M.P. (2001) A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. Cell 106, 413-415.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Kowalski P.E., Freeman J.D. and Mager D.L. (1999). Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. Genomics 57, 371-379.

Krauss V. and Reuter G. (2000). Two genes become one: the genes encoding heterochromatin protein Su(var)3-9 and translation initiation factor subunit eIF-2γ are joined to a dicistronic unit in holometabolic insects. Genetics 156, 1157-1167.

Labrador M., Mongelard F., Plata-Rengifo P., Baxter E.M., Corces V.G. and Gerasimova T.I. (2001). Protein encoding by both DNA strands. Nature 409, 1000.

Liang F., Holt I., Pertea G., Karamycheva S., Salzberg S.L. and Quackenbush J. (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. Nature Genet. 25, 239-240.

Magrangeas F., Pitiot G., Dubois S., Bragado-Nilsson E., Cherel M., Jobert S., Lebeau B., Boisteau O., Lethe B., Mallet J., Jacques Y. and Minvielle S. (1998). Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor α-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution. J. Biol. Chem. 273, 16005-16010.

Millar J.K., Christie S., Semple C.A.M. and Porteous D.J. (2000). Chromosomal location and genomic structure of the human translin-associated factor X gene (*TRAX;TSNAX*) revealed by intergenic splicing to *DISC1*, a gene disrupted by a translocation segregating with schizophrenia. Genomics 67, 69-77.

Moore R.C., Lee I.Y., Silverman G.L., Harrison P.M., Strome R., Heinrich C., Karunaratne A., Pasternak S.H., Chishti M.A., Liang Y., Mastrangelo P., Wang K., Smit A.F., Katamine S., Carlson G.A., Cohen F.E., Prusiner S.B., Melton D.W., Tremblay P., Hood L.E. and Westaway D. (1999). Ataxia in prion protein (PrP)-deficient mice is associated with upregulation of the novel PrP-like protein doppel. J. Mol. Biol. 292, 797-817.

Morishima N., Yoshioka H., Higashi Y., Sogawa K. and Fujii-Kuriyama Y. (1987). Gene structure of cytochrome P-450(M-1) specifically expressed in male rat liver. Biochemistry 26, 8279-8285.

Plant K. E., Routledge S.J.E. and Proudfoot N.J. (2001). Intergenic transcription in the human ß-globin gene cluster. Mol. Cell. Biol. 21, 6507-6514.

Rogan D.F., Cousins D.J. and Staynov D.Z. (1999). Intergenic transcription occurs throughout the human IL-4/IL-13 gene cluster. Biochem. Biophys. Res. Commun. 255, 556-561.

Smith C.W. and Valcárcel J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. 25, 381-388.

The *C. elegans* Sequencing Consortium (1998). Genome sequencing of the nematode *C. elegans*: a platform for investigative biology. Science 282, 2012-2018.

Thomson T.M., Lozano J.J., Loukili N., Carrio R., Serras F., Cormand B., Valeri M., Diaz V.M., Abril J., Burset M., Merino J., Macaya A., Corominas M. and Guigo R. (2000). Fusion of the human gene for the polyubiquitination coeffector UEV1 with *kua*, a newly identified gene. Genome Res. 10, 1743-1756.

Tran D.P., Kim S.J., Park N.J., Jew T.M. and Martinson H.G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. Mol Cell Biol. 21, 7495-7508.

Venter J.C., Adams M.D., Myers E.W., Li P.W., Mural R.J., Sutton G.G., Smith H.O., Yandell M., Evans C.A., Holt R.A. et al. (276 authors) (2001). The sequence of the human genome. Science 291, 1304-1351.

Warner S.C., Finta C. and Zaphiropoulos P.G. (2001). Intergenic transcripts containing a novel human cytochrome P450 *2C* exon 1 spliced to sequences from the *CYP*2C9 gene. Mol. Biol. Evol. 18, 1841-1848.

Westlind A., Malmebo S., Johansson I., Otter C., Andersson T.B., Ingelman-Sundberg M. and Oscarson M. (2001). Cloning and tissue distribution of a novel human cytochrome P450 of the *CYP3A* subfamily, *CYP3A43*. Biochem. Biophys. Res. Commun. 281, 1349-1355.

Wright F.A., Lemon W.J., Zhao W.D., Sears R., Zhuo D., Wang J.-P, Yang H.-Y., Baer T., Stredney D., Spitzner J., Stutz A., Krahe R. and Yuan B. (2001). A draft annotation and overview of the human genome. Genome Biology 2, research0025.

Zaphiropoulos P.G. (1993). Differential expression of cytochrome P450 2C24 transcripts in rat kidney and prostate. Evidence indicative of alternative and possibly *trans* splicing events. Biochem. Biophys. Res. Commun. 192, 778-786.

Zaphiropoulos P.G. (1999). RNA molecules containing exons originating from different members of the cytochrome P450 2C gene subfamily (*CYP2C*), in human epidermis and liver. Nucleic Acids Res. 27, 2585-2590.

Accepted January 21, 2002