

## Minireview

# Vertebrate pseudogenes

A.J. Mighell<sup>a,b,\*</sup>, N.R. Smith<sup>a</sup>, P.A. Robinson<sup>a,b</sup>, A.F. Markham<sup>a</sup>

<sup>a</sup>Molecular Medicine Unit, The University of Leeds, Clinical Sciences Building, St. James's University Hospital, Leeds LS9 7TF, UK

<sup>b</sup>Leeds Dental Institute, The University of Leeds, Clarendon Way, Leeds LS2 9LU, UK

Received 14 January 2000

Edited by Horst Feldmann

**Abstract** Pseudogenes are commonly encountered during investigation of the genomes of a wide range of life forms. This review concentrates on vertebrate, and in particular mammalian, pseudogenes and describes their origin and subsequent evolution. Consideration is also given to pseudogenes that are transcribed and to the unusual group of genes that exist at the interface between functional genes and non-functional pseudogenes. As the sequences of different genomes are characterised, the recognition and interpretation of pseudogene sequences will become more important and have a greater impact in the field of molecular genetics.

© 2000 Federation of European Biochemical Societies.

**Key words:** Pseudogene; Retrotransposon; Human genome; Transcription

### 1. What is a pseudogene and whence do they originate?

A pseudogene is a sequence which is present in the genome of a given population and typically is characterised by close similarities to one or more paralogous genes, yet is non-functional. This lack of function is a result of either failure of transcription or translation, or production of a protein that does not have the same functional repertoire as the protein encoded by the normal paralog gene. A fundamental feature of pseudogenes is that their nucleotide sequences differ from those of the paralogous functional genes at crucial points. Vanin [1] stressed that the term pseudogene is only applicable to sequences that are related to another sequence but are defective. Pseudogenes have been denoted in several ways including the prefixed Greek symbol  $\psi$ , for example  $\psi$ PGK-1, or by a capital 'P' suffix, for example CYP21P.

The term 'pseudogene' originally arose from investigation of the genome of *Xenopus laevis* [2]. Subsequently, pseudogenes have been identified in a diverse range of life forms including bacteria, plants, insects and vertebrates [3–5].

A schematic overview of the issues covered in this review is presented in Fig. 1. Pseudogenes are a consequence of gene duplication which can occur in two fundamentally different ways: firstly, by retrotransposition and, secondly, via the duplication of genomic DNA.

### 2. Pseudogenes arising by retrotransposition

Pseudogenes arising by retrotransposition are known as either processed pseudogenes or retro-pseudogenes. They are typically characterised by an absence of both 5'-promoter sequence and introns, the presence of flanking direct repeats and a 3'-polyadenylation tract [1,6]. Processed pseudogenes are retrotransposons, that in common with other retrotransposons, have been inserted into the genome as double-stranded sequence generated from single-stranded RNA [1,6,7]. Processed pseudogenes are probably generated by RNA polymerase II rather than the RNA polymerase III transcription responsible for other retrotransposons such as either *Alu* or *LINE* repeats [8].

Retrotransposition of any sequence is potentially threatening to normal genome function, yet the mechanisms that control this process remain poorly understood [9–11]. Processed pseudogenes have been generated in-vitro at a low frequency in human HeLa cells via mRNA from a reporter gene [8]. The resultant pseudogenes had features consistent with the processed pseudogenes observed in mammalian genomes. This study demonstrated that human cells possess an endogenous reverse transcription activity that is not restricted to transcripts of transposable elements [8].

The majority of retrotransposed genes are inactivated to processed pseudogenes, but in a few instances the retrotransposon is maintained as a functional, intronless gene. Two intronless retrotransposons,  $\psi$ PGK-1 and PGK-2, were derived from the functional, intron-containing human PGK-1 gene.  $\psi$ PGK-1 is a typical processed pseudogene, but PGK-2 generates transcripts that are translated into a protein with a high homology to that encoded by PGK-1 [12,13]. Similarly, there are three closely related murine *Amd* genes. *Amd-1* includes introns and codes for a 334 amino acid protein whereas *Amd-2* is an intronless retrotransposon that codes for a functional protein with only two amino acid differences from *Amd-1* [14]. *Amd-3* is also an intronless retrotransposon with a high nucleotide identity to *Amd-2*, but *Amd-3* is unable to encode a functional protein and is a processed pseudogene [14].

### 3. Pseudogenes arising by duplication of genomic DNA

Duplication of DNA segments has been essential in the development of complex genomes and explains the generation of gene families from a single ancestral gene [15,16]. The composition of any genome is not static and must inevitably change with time. The mechanisms by which duplicated genes either take on new adaptive functions, are maintained as they

\*Corresponding author. Fax: (44)-113-2444475.  
E-mail: a.j.mighell@leeds.ac.uk

were at the time of duplication or become neutralised to pseudogenes remain an area of intense debate [17–21]. In simple terms, pseudogenes represent genes that were not maintained as functional elements either because they conferred a selection disadvantage or because they did not give a selection advantage to the organism in evolutionary terms. This interpretation is made with the caveat that the duplicated sequence was functional immediately after duplication. If duplication was incomplete then the new sequence would be a pseudogene from the outset.

The 5'-end of the *BRCA1* gene lies within a 30 kb region of chromosome 17 that has undergone tandem duplication. This has resulted in a  $\psi$ *BRCA1* gene paralogous to the 5'-end of *BRCA1*, but which only includes exons 1A, 1B and 2 [22]. It is extremely unlikely that this  $\psi$ *BRCA1* was ever functional at the protein level. The human  $\psi$ *PS*,  $\psi$ *SORD* and  $\psi$ *MTX* genes lack their respective exon 1 sequences which include the methionine start codons and so presumably these pseudogenes were never functional [23–25]. Alternatively, exon 1 deletions could have occurred some time after duplication. By contrast, some genes were almost certainly functional before being silenced into pseudogenes. For example, *Hox* genes have fundamental roles in patterning during development and are highly conserved between different species. However, during evolution of the puffer fish genome two *Hox* genes which are ubiquitously expressed in other species, have been silenced to pseudogenes [26,27].

#### 4. Numbers of pseudogenes

Pseudogenes are common, but it is not possible to quantify the number of pseudogenes within a particular genome until it has been completely sequenced. Even then, it may be difficult to differentiate between some pseudogenes and functional genes for reasons that will be discussed later. Initial interpretation of the sequence data from human chromosome 22 indicated that 19% of the coding sequences were pseudogenes [28]. Of these, 82% were processed pseudogenes.

Vertebrate genomes include a large number of genes with paralogous pseudogenes yet it is also probable that many genes do not have pseudogenes. The number of pseudogenes for any paralogous gene can vary tremendously from a single one to multiple non-functional copies. Ribosomal RNA genes can have hundreds of paralogous pseudogenes [29]. Highly polymorphic loci frequently include large numbers of pseudogenes and this is particularly well illustrated with immunoglobulin loci. For example, the murine immunoglobulin kappa light chain gene locus includes over 140 V $\kappa$  genes and paralogous pseudogenes of which at least 47 are pseudogenes [30,31]. Interpretation of such loci can be difficult and will be returned to in a later section.

Housekeeping genes may also have several pseudogenes. For example, in the rat there is one functional *Geranylgeranyltransferase I* gene, but at least 13 similar paralogous pseudogenes that have arisen by retrotransposition of a mis-spliced transcript [32].

In the human genome there are two loci, at chromosomes 15q11.2 and 16p11.2, that each includes a cassette of pseudogenes, where the cassette can be expanded to several copies at the same locus. These cassette expansions are believed to be a normal euchromatic variant and are of unknown significance [33–35].

#### 5. Location of pseudogenes

Potentially, pseudogenes can be located anywhere within a genome although they are more likely to persist over time in loci where they do not cause a deleterious effect. The mechanisms of retrotransposition mean that retrotransposons, including processed pseudogenes, are often clustered together at common loci, although a single retrotransposed element can occur in isolation [36,37]. Pseudogenes generated by duplication of genomic DNA are more likely than processed pseudogenes to be adjacent to their paralogous functional gene, but can be inserted into a different chromosome. For example, the majority of the human immunoglobulin V $\lambda$  light chain pseudogenes are clustered with their paralogous functional genes on chromosome 22q11.2. Two further 'orphan' human V $\lambda$  light chain pseudogenes are present at chromosome 8q11.2 and are believed to have arisen by a single duplication and translocation event that occurred before the divergence of humans and gorillas [38]. By contrast, human *olfactory receptor* pseudogenes have been distributed to most of the human chromosomes by duplication of genomic DNA [39,40].

Pseudogenes can also be identified both within genes or in their control elements. The promoter regions of the human *Type 1 angiotensin II receptor* gene, of each of the *amylase* genes and of the *prostate-specific transglutaminase* gene all include a pseudogene unrelated to the functional gene [15,41,42]. Intronic pseudogenes are present in the human *ATP7* and *p53* genes [43,44]. The insertion of pseudogenes within exons will almost inevitably alter gene function that will either be selected against or less likely, will generate a gene with a novel function.

Pseudogenes arising from human mitochondrial DNA can be inserted within nuclear DNA adding a further level of complexity to the analysis of mitochondrial disease [45,46].

#### 6. The fate of pseudogenes

The locus of insertion of a pseudogene is fundamental to its subsequent evolution. Insertion of some pseudogenes will have a deleterious effect on normal function of other genes and will be rapidly selected against and lost. Many other pseudogenes persist and evolve with time.

Established pseudogenes would be expected to undergo genetic drift as there are no apparent selection pressures to prevent either random mutations or deletions and insertions [47]. This is supported by the observation that processed pseudogenes mostly evolve more rapidly than their functional paralogs [48]. Analysis of 109 processed pseudogenes demonstrated that nucleotide deletions or insertions occurred once for every 40 or 100 nucleotide substitutions, respectively [49].

Inserted retrotransposons, such as either *Alu* or *LINE* sequences, may persist in pseudogenes whereas they would be expected to have a significantly deleterious effect on function when inserted into the paralogous functional gene. *Alu* sequences are present in some functional genes and their mature mRNAs, but *Alu* sequences are rarely identified in protein-coding regions [50,51]. Accordingly, the recognition of an *Alu* sequence within an apparently protein coding region can allow rapid identification of a pseudogene [42,52,53]. The presence of intronic *Alu* sequences in one but not another paral-

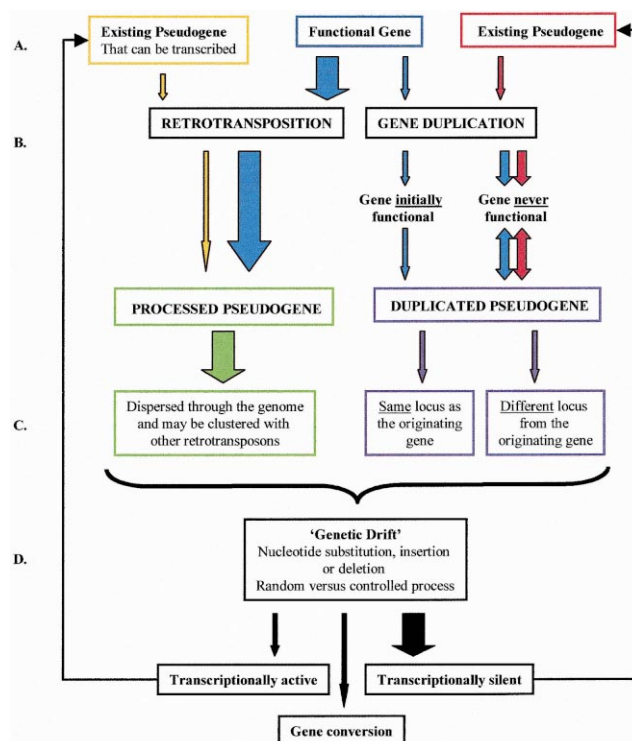


Fig. 1. A schematic summary, to be interpreted in conjunction with the text, of the principal issues covered in this review of vertebrate pseudogenes. The width of the arrows gives a rough guide to the relative importance of the different pathways at each level. A: Origin of pseudogenes. The majority of vertebrate pseudogenes are probably derived from functional genes. B: Mechanism of pseudogene formation. The majority of vertebrate pseudogenes are a result of retrotransposition of transcripts derived from genes that encode functional proteins. C: Pseudogene location. Pseudogenes persist in parts of the genome where they do not have a deleterious effect on fitness of the organism. D: Pseudogene fate. Most pseudogenes undergo genetic drift and are never transcribed. By contrast, in some instances there appears to be selectional pressures that prevent major changes to the pseudogene sequence. A few pseudogenes are involved in gene conversion and a few can be transcribed. Accordingly, not all pseudogenes are unequivocally functionless.

ogous gene can also be a convenient pointer to which gene is a pseudogene [25].

Established pseudogenes can potentially be copied to generate further pseudogenes. A processed  $\psi$ PCNA has later been partially duplicated to give a second pseudogene at the same locus [54]. The human *amylase* gene family (*AMY*) originated from a single ancestral gene that had a  $\gamma$ -actin processed pseudogene inserted into its promoter. Subsequent gene duplication and evolution resulted in an uninterrupted  $\gamma$ -actin pseudogene in *AMY2B*, a partially deleted  $\gamma$ -actin pseudogene that includes a long terminal repeat (LTR) in *AMY2A*, and partially deleted  $\gamma$ -actin pseudogenes that include two LTRs in *AMY1A*, *AMY1B* and *AMY1C* [12].

The evolution of many pseudogenes follows the patterns expected for non-functional sequences. However, it is apparent that factors other than evolutionary time can have a significant influence on pseudogene evolution [49]. Nucleotide changes are not always random and, for example, several disease-associated nucleotide substitutions in human *PMM2* are also present in the paralogous pseudogene [48,55]. Some pseudogenes exhibit accelerated genetic drift for reasons which are not readily apparent [56]. Other pseudogenes are remarkably similar to their paralogous functional gene. The murine processed  $\psi$ Tsg101 has 98% nucleotide identity to *Tsg101* cDNA, but the pseudogene includes a frameshift mutation and a premature stop codon which would generate a very different protein lacking some of the key functional do-

main of *Tsg101* [19]. The transcribed intronless human  $\psi$ PTEN has 98.6% homology to functional *PTEN* cDNA, but 18 of its 19 nucleotide differences are located in the protein coding region and  $\psi$ PTEN lacks a translation initiation methionine codon [57]. The high nucleotide similarity may simply represent recent pseudogene formation. However, examples have been identified where pseudogene sequences are maintained by functional demands.

The sequence similarity between a functional gene and its paralogous pseudogene(s) can provide a good matrix for genomic rearrangement and gene conversion, particularly where the two genes are situated close to each other. This process can have a key functional role and is best understood through study of the chicken immunoglobulin genes. Both chickens and mammals need to generate diversity in the variable domains of the immunoglobulin heavy ( $V_H$ ) and light ( $V_L$ ) chains during B-cell development, but achieve this diversity in very different ways. Mammals rely on a number of different functional immunoglobulin  $V_H$  and  $V_L$  genes whereas chickens have only a single functional copy each of a  $V_H$  and  $V_L$  gene. Instead, chickens generate diversity by somatic cell gene conversion of the functional  $V_H$  and  $V_L$  genes by a number of different paralogous  $V_H$  and  $V_L$  pseudogenes [58].

In mammals, gene conversion events involving pseudogenes are rarer and as yet have not been identified to play such fundamental roles as in chickens. Nucleotide point mutations with significant effects on protein function have been gener-

ated by gene conversion of the murine *lactate dehydrogenase* gene by its pseudogene [59]. In an unusual example, a human pseudogene can potentially become functional again. The human  $\alpha$ -globin cluster of genes on chromosome 16 has arisen by gene duplication and divergence. This cluster includes  $\zeta 2$ , which is briefly expressed in the embryonic yolk sac, and its paralogous pseudogene  $\psi\zeta 1$  [60].  $\zeta 2$  and  $\psi\zeta 1$  are almost identical at the nucleotide level, but  $\psi\zeta 1$  has a non-functional promoter. In some individuals, gene conversion of  $\psi\zeta 1$  by  $\zeta 2$  has resulted in restoration of a functional promoter and the generation of  $\zeta 1$  from  $\psi\zeta 1$  [61].  $\zeta 2$ - $\zeta 1$  alleles are common in a number of populations, but despite being potentially functional, it is not clear if  $\zeta 1$  is then transcribed [61]. The evolution of *seminal ribonuclease* genes and their paralogous pseudogenes has been investigated in mammals and there is evidence which supports conversion of a *seminal ribonuclease* pseudogene to a functional gene in recent evolutionary history [62].

Gene conversion involving pseudogenes can have a deleterious effect on the fitness of the organism and is the fundamental event in the pathogenesis of some diseases where part of a gene is replaced by its pseudogene. These include mucopolysaccharidosis type II, congenital adrenal hyperplasia and Gaucher disease with the respective genes involved being *IDS*, *CYP21* and *GBA* [63–68].

Other gene conversion events involving pseudogenes are of less apparent significance. For example, the long-standing bovine  $\beta$ -*lactoglobulin* pseudogene appears to have had its 3'-end recently converted by the functional  $\beta$ -*lactoglobulin* gene [69]. The 3'-end of the murine *H2a* pseudogene is the target of frequent gene conversion by functional *H2a* genes [70].

The contrasts between the evolution of genes and their paralogous pseudogenes can be quite dramatic, as illustrated above by the comparison of mammalian and chicken immunoglobulin genes. Many other examples are recognised. Nucleotide analysis of gene families including their pseudogene paralogs can thus provide a powerful tool for phylogenetic studies that investigate genome evolution between different species and between different populations within individual species [62,71–76].

## 7. Some pseudogenes can be transcribed

Consideration of how pseudogenes are formed suggests that most are unlikely to be transcribed, but pseudogene transcripts can nevertheless be identified [13,77,78]. The functional relevance of pseudogene transcripts remains unclear.

Processed pseudogenes cannot include all the transcriptional control elements present in their paralogous functional gene. Accordingly, they must use other transcriptional elements in the same way that functional retrotransposed genes do. The intronless, functional *PGK-2* originated as a retrotransposon of *PGK-1*, yet it is still transcribed despite the GC-rich promoter of *PGK-1* being replaced by an AT repeat in *PGK-2* [13]. A similar situation occurs with the functional murine retrotransposon *Amd-2* that has to employ a very different and much less efficient promoter than the one used by the intron-containing *Amd-1* gene from which it originated [14]. Alternatively, pseudogene transcription can be driven by a nearby promoter present in unrelated sequence. For example, the promoter of human *CYP21A1* drives transcription of an adjacent, unrelated pseudogene [79]. A human *topoisomer-*

*ase I* pseudogene and a murine  $\psi$ *Fgfr-3* are transcribed in an antisense direction reflecting the insertion of these pseudogenes adjacent to a functional, but unrelated promoter [80,81].

Pseudogenes derived from genomic DNA duplication potentially include the transcriptional control elements. These usually represent a relatively small proportion of the whole gene compared to its protein coding regions. Accordingly, it can be envisaged that a pseudogene might be non-functional due to nucleotide mutations in the protein-coding parts of its sequence, whereas the control elements are maintained and pseudogene transcripts are produced [17]. Transcriptional control element nucleotide changes can be identified in pseudogenes of this class and the loss of methylated CpG residues may contribute to gene silencing [82,83]. Even where transcriptional control elements appear potentially functional, nuclear architecture and intranuclear trafficking of transcription factors may exert a powerful influence on transcription [84].

Pseudogene transcripts can be more prevalent than the functional paralogous transcripts. *PTEN* is a tumour suppressor gene implicated in cancer development as a result of either somatic or germline mutations. In the analysis of *PTEN* and  $\psi$ *PTEN* transcripts in a variety of tissues,  $\psi$ *PTEN* was the predominant transcript in liver and glioblastoma cell lines, but represented a minor species in spleen and kidney [77]. The spatial expression of pseudogene transcripts can differ from that of the paralogous gene transcripts. Human *5-HT7 receptor* pseudogene transcripts can be identified in tissues such as kidney and liver whereas the functional *5-HT7 receptor* transcripts cannot be detected [78]. Pseudogene transcripts can even be alternatively spliced [53,78,85].

Recognition of the extensive sequence identity between some pseudogenes and their paralogous functional genes is essential to meaningful mutational analysis by RT-PCR. For example, RT-PCR assay design may have to distinguish between a functional transcript, an almost identical pseudogene transcript and the genomic processed pseudogene present in DNA contaminating the RNA sample [86–90]. Detection of human *cytokeratin 19* (*CK19*) transcripts has been proposed as a sensitive marker of metastatic epithelial cancer cells in tissues that normally do not express *CK19*, such as blood, lymph nodes and bone marrow. However, the oligonucleotide primers initially used for this assay allowed detection of a novel  $\psi$ *CK19* which casts significant doubts on the validity of previously published results [91].

## 8. The pseudogene twilight zone

A small subset of apparent pseudogenes appear, on current evidence, to be non-functional although this may have to be re-evaluated in the future. The human *laminin receptor* (*LAMR*) gene family includes a number of processed pseudogenes. One of these, *LAMRL5*, has 97.9% nucleotide identity with the functional *LAMR1*, includes upstream TATA and CAAT boxes and codes for a protein with 97% amino acid homology to *LAMR1*. Currently there are no data to suggest that *LAMRL5* is ever transcribed [91]. Similarly, a human processed  $\psi$ *CK2\alpha* includes transcriptional control elements and has a complete open reading frame, but evidence of transcription has not been established [92].

Interpretation of highly polymorphic loci can be extremely difficult. The murine immunoglobulin V $\kappa$  gene locus includes over 140 genes and paralogous pseudogenes [30,31]. Eighteen

of these sequences are unequivocally pseudogenes and have been designated 'relics' due to multiple nucleotide substitutions, deletions and insertions that would prevent encoding a functional kappa light chain. Equally, there are other sequences that fulfill the strict criteria for a functional V $\kappa$  gene. Between these two extremes there are a group of sequences that are less easy to interpret. These include sequences that have single base pair differences in their sequence compared to the consensus V $\kappa$  sequences of key functional domains. Analysis of V $\kappa$  cDNA populations must be made with caution. As illustrated above, the presence or absence of a specific cDNA does not necessarily confirm or refute whether a gene is functional or not. Furthermore, the V $\kappa$  cDNA mutations that are a feature of this highly polymorphic group of genes can make it difficult to identify the originating gene.

In these and other examples it cannot be stated with certainty that a gene is unequivocally either a pseudogene or a gene. It is possible that analysis has not been performed in the appropriate temporo-spatial conditions to detect expression. Equally, a greater understanding about the biology of certain genes and their paralogous pseudogenes may be required before interpretation can be made with more certainty.

## 9. Conclusions

Pseudogenes are common and are encountered in a diverse range of life forms, but particularly vertebrates. Genome complexity has evolved by the generation of gene families via gene duplication, but for reasons that remain contentious, some of these duplicated genes have become non-functional pseudogenes. Retrotransposed pseudogenes arise by a completely different mechanism and reflect a different aspect of genome evolution. Once established within a genome, pseudogenes evolve with time, although the mechanisms that control these changes remain very poorly understood.

Most pseudogenes have multiple features that confirm their non-functional status. However, there are genes that have many features of pseudogenes, but which are functional, and a separate group of genes that are currently considered as pseudogenes, but with the recognition that these genes are potentially functional. Accordingly, experimental design and interpretation across the whole field of molecular genetics must take pseudogenes into careful consideration.

**Acknowledgements:** A.J.M. thanks the Medical Research Council for specific support. Research in the authors' laboratory is also funded by The Wellcome Trust and Yorkshire Cancer Research.

## References

- [1] Vanin, E.F. (1985) *Annu. Rev. Genet.* 19, 253–272.
- [2] Jacq, C., Miller, J.R. and Brownlee, G.G. (1977) *Cell* 12, 109–120.
- [3] Fsihi, H., De Rossi, E., Salazar, L., Cantoni, R., Labo, M., Riccardi, G., Takiff, H.E., Eiglmeier, K., Bergh, S. and Cole, S.T. (1996) *Microbiology* 142, 3147–3161.
- [4] Loguercio, L.L. and Wilkins, T.A. (1998) *Curr. Genet.* 34, 241–249.
- [5] Ramos-Onsins, S. and Aguade, M. (1998) *Genetics* 150, 157–171.
- [6] Rogers, J.H. (1985) *Int. Rev. Cytol.* 93, 187–279.
- [7] Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) *Annu. Rev. Biochem.* 55, 631–661.
- [8] Maestre, J., Tchénio, T., Dhellin, O. and Heidmann, T. (1995) *EMBO J.* 14, 6333–6338.
- [9] Curcio, M.J. and Garfinkel, D.J. (1999) *Trends Genet.* 15, 43–45.
- [10] Werle-Schneider, G., von Brevern, M.C., Sylla, B.S. and Hollstein, M.C. (1999) *Genes Chromosomes Cancer* 26, 84–91.
- [11] Moran, J.V., DeBerardinis, R.J. and Kazazian Jr., H.H. (1999) *Science* 283, 1530–1534.
- [12] McCarrey, J.R. (1990) *Nucleic Acids Res.* 18, 949–955.
- [13] McCarrey, J.R., Kumari, M., Aivaliotis, M.J., Wang, Z., Zhang, P., Marshall, F. and Vandeberg, J.L. (1996) *Dev. Genet.* 19, 321–332.
- [14] Nishimura, K., Liisanantti, M., Muta, Y., Kashiwagi, K., Shirahata, A., Janne, M., Kankare, K., Janne, O.A. and Igarashi, K. (1998) *Biochem. J.* 332, 651–659.
- [15] Samuelson, L.C., Wiebauer, K., Snow, C.M. and Meisler, M.H. (1990) *Mol. Cell. Biol.* 10, 2513–2520.
- [16] Hardison, R. (1998) *J. Exp. Biol.* 201, 1099–1117.
- [17] Cooke, J., Nowak, M.A., Boerlijst, M. and Maynard-Smith, J. (1997) *Trends Genet.* 13, 360–364.
- [18] Tachida, H. and Kuboyama, T. (1998) *Genetics* 149, 2147–2158.
- [19] Wagner, K.U., Dierisseau, P., Rucker, E.B., Robinson, G.W. and Hennighausen, L. (1998) *Oncogene* 17, 2761–2770.
- [20] Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y. and Postlethwait, J. (1999) *Genetics* 151, 1531–1545.
- [21] Ganfornina, M.D. and Sanchez, D. (1999) *BioEssays* 21, 432–439.
- [22] Brown, M.A., Xu, C.F., Nicolai, H., Griffiths, B., Chambers, J.A., Black, D. and Solomon, E. (1996) *Oncogene* 12, 2507–2513.
- [23] Ploos van Amstel, H.K., Reitsma, P.H., van der Logt, C.P. and Bertina, R.M. (1990) *Biochemistry* 29, 7853–7861.
- [24] Long, G.L., Winfield, S., Adolph, K.W., Ginns, E.I. and Bornstein, P. (1996) *Genomics* 33, 177–184.
- [25] Carr, I.M., Markham, A.F. and Coletta, P.L. (1997) *Eur. J. Biochem.* 245, 760–767.
- [26] Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R. and Brenner, S. (1997) *Nature Genet.* 16, 79–83.
- [27] Meyer, A. (1998) *Nature* 391, 225–228.
- [28] Dunham, I., Hunt, A.R. and Collins, J.E. et al. (1999) *Nature* 402, 489–495.
- [29] Frederiksen, S., Cao, H., Lomholt, B., Levan, G. and Hallenberg, C. (1997) *Cytogenet. Cell Genet.* 76, 101–106.
- [30] Thiebe, R., Schäble, K.F., Bensch, A., Brensing-Küppers, J., Heim, V., Kirschbaum, T., Mitlöhner, H., Ohnrich, M., Pourrajabi, S., Rösenthaller, F., Schwendinger, J., Wichelhaus, D., Zocher, I. and Zachau, H.G. (1999) *Eur. J. Immunol.* 29, 2072–2081.
- [31] Schäble, K.F., Thiebe, R., Bensch, A., Brensing-Küppers, J., Heim, V., Kirschbaum, T., Lamm, R., Ohnrich, M., Pourrajabi, S., Rösenthaller, F., Schwendinger, J., Wichelhaus, D., Zocher, I. and Zachau, H.G. (1999) *Eur. J. Immunol.* 29, 2082–2086.
- [32] Dhawan, P., Yang, E., Kumar, A. and Mehta, K.D. (1998) *Gene* 210, 9–15.
- [33] Barber, J.C., Cross, I.E., Douglas, F., Nicholson, J.C., Moore, K.J. and Browne, C.E. (1998) *Hum. Genet.* 103, 600–607.
- [34] Ritchie, J.R., Mattei, M.-G. and Lalande, M. (1998) *Hum. Mol. Genet.* 8, 1253–1260.
- [35] Barber, J.C., Reed, C.J., Dahoun, S.P. and Joyce, C.A. (1999) *Hum. Genet.* 104, 211–218.
- [36] Liu, Q.R. and Chan, P.K. (1990) *J. Mol. Biol.* 212, 453–459.
- [37] Sadhu, A., Shen, M.L., Hackbarth, M., Hume, E. and McKeithan, T.W. (1997) *Immunogenetics* 45, 365–370.
- [38] Frippiat, J.P., Dard, P., Marsh, S., Winter, G. and Lefranc, M.P. (1997) *Eur. J. Immunol.* 27, 1260–1265.
- [39] Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J. and Giorgi, D. (1998) *Nature Genet.* 18, 243–250 [published erratum: (1998) *Nature Genet.* 19, 102].
- [40] Brand-Arpon, V., Rouquier, S., Massa, H., de Jong, P.J., Ferraz, C., Ioannou, P.A., Demaille, J.G., Trask, B.J. and Giorgi, D. (1999) *Genomics* 56, 98–110.
- [41] Espinosa, L., Martin, M., Nicolas, A., Fabre, M. and Navarro, E. (1997) *Biochim. Biophys. Acta* 1354, 58–64.
- [42] Dubbink, H.J., de Waal, L., van Haperen, R., Verkaik, N.S., Trapman, J. and Romijn, J.C. (1998) *Genomics* 51, 434–444.
- [43] Reisman, D., Balint, E., Loging, W.T., Rotter, V. and Almon, E. (1996) *Genomics* 38, 364–370.

- [44] Dierick, H.A., Mercer, J.F. and Glover, T.W. (1997) *Gene* 198, 37–41.
- [45] Taylor, R.W., Taylor, G.A., Morris, C.M., Edwardson, J.M. and Turnbull, D.M. (1998) *Biochem. Biophys. Res. Commun.* 251, 883–887.
- [46] Herrnstadt, C., Clevenger, W., Ghosh, S.S., Anderson, C., Fahy, E., Miller, S., Howell, N. and Davis, R.E. (1999) *Genomics* 60, 67–77.
- [47] Blake, R.D., Hess, S.T. and Nicholson-Tuell, J. (1992) *J. Mol. Evol.* 34, 189–200.
- [48] Ophir, R., Itoh, T., Graur, D. and Gojobori, T. (1999) *Mol. Biol. Evol.* 16, 49–53.
- [49] Ophir, R. and Graur, D. (1997) *Gene* 205, 191–202.
- [50] Makalowski, W., Mitchell, G.A. and Labuda, D. (1994) *Trends Genet.* 10, 188–193.
- [51] Mighell, A.J., Markham, A.F. and Robinson, P.A. (1997) *FEBS Lett.* 47, 1–5.
- [52] Foord, S.M., Marks, B., Stolz, M., Bufflier, E., Fraser, N.J. and Lee, M.G. (1996) *Genomics* 35, 182–188.
- [53] Cronin, C.N., Zhang, X., Thompson, D.A. and McIntire, W.S. (1998) *Gene* 220, 71–76.
- [54] Taniguchi, Y., Katsumata, Y., Koido, S., Suemizu, H., Yoshimura, S. and Moriuchi, T. (1996) *Mamm. Genome* 7, 906–908.
- [55] Schollen, E., Pardon, E., Heykants, L., Renard, J., Doggett, N.A. and Callen, D.F. (1998) *Hum. Mol. Genet.* 7, 157–164.
- [56] Grimsley, C., Mather, K.A. and Ober, C. (1998) *Mol. Biol. Evol.* 15, 1581–1588.
- [57] Dahia, P.L.M., FitzGerald, M.G., Zhang, X., Marsh, D.J., Zheng, Z., Pietsch, T., von Deimling, A., Haluska, F.G., Haber, D.A. and Eng, C. (1998) *Oncogene* 16, 2403–2406.
- [58] McCormack, W.T., Hurley, E.A. and Thompson, C.B. (1993) *Mol. Cell. Biol.* 13, 821–830.
- [59] Pretsch, W., Chatterjee, B., Favor, J., Merkle, S. and Sandulache, R. (1998) *Mamm. Genome* 9, 144–149.
- [60] Liebhaber, S.A. and Russell, J.E. (1998) *Ann. NY Acad. Sci.* 850, 54–63.
- [61] Hill, A.V.S., Nicholls, R.D., Thein, S.L. and Higgs, D.R. (1985) *Cell* 42, 809–819.
- [62] Trabesinger-Ruef, N., Jermann, T., Zankel, T., Durrant, B., Frank, G. and Benner, S.A. (1996) *FEBS Lett.* 382, 319–322.
- [63] Zimran, A., Sorge, J., Gross, E., Kubitz, M., West, C. and Beutler, E. (1990) *J. Clin. Invest.* 85, 219–222.
- [64] Strasberg, P.M., Triggs-Raine, B.L., Warren, I.B., Skomorowski, M.A., McInnes, B., Becker, L.E., Callahan, J.W. and Clarke, J.T. (1994) *J. Clin. Lab. Anal.* 8, 228–236.
- [65] Birot, A.M., Bouton, O., Froissart, R., Maire, I. and Bozon, D. (1996) *Hum. Mutat.* 8, 44–50.
- [66] Bunge, S., Rathmann, M., Steglich, C., Bondeson, M.L., Tylki-Szymanska, A., Popowska, E. and Gal, A. (1998) *Eur. J. Hum. Genet.* 6, 492–500.
- [67] Krone, N., Roscher, A.A., Schwarz, H.P. and Braun, A. (1998) *Clin. Chem.* 44, 2075–2082.
- [68] Lee, H.H., Chao, H.T., Lee, Y.J., Shu, S.G., Chao, M.C., Kuo, J.M. and Chung, B.C. (1998) *Hum. Genet.* 103, 304–310.
- [69] Passey, R.J. and Mackinlay, A.G. (1995) *Eur. J. Biochem.* 233, 736–743.
- [70] DeBry, R.W. (1998) *J. Mol. Evol.* 46, 355–360.
- [71] Casane, D., Boissinot, S., Chang, B.H., Shimmin, L.C. and Li, W. (1997) *J. Mol. Evol.* 45, 216–226.
- [72] Devor, E.J., Dill-Devor, R.M., Magee, H.J. and Waziri, R. (1998) *J. Exp. Zool.* 282, 150–156.
- [73] Pompei, F., Ciminelli, B.M. and Modiano, G. (1998) *Hum. Biol.* 70, 659–666.
- [74] Persson, K., Heby, O. and Berger, F.G. (1999) *Mamm. Genome* 10, 784–788.
- [75] Jiddou, R.R., Wei, W.L., Sane, K.S. and Killeen, A.A. (1999) *Clin. Chem.* 45, 625–629.
- [76] Boyson, J.E., Iwanaga, K.K., Urvater, J.A., Hughes, A.L., Golos, T.G. and Watkins, D.I. (1999) *Immunogenetics* 49, 86–98.
- [77] Fujii, G.H., Morimoto, A.M., Berson, A.E. and Bolen, J.B. (1999) *Oncogene* 18, 1765–1769.
- [78] Olsen, M.A. and Schechter, L.E. (1999) *Gene* 227, 63–69.
- [79] Bristow, J., Gitelman, S.E., Tee, M.K., Staels, B. and Miller, W.L. (1993) *J. Biol. Chem.* 268, 12919–12924.
- [80] Zhou, B.S., Beidler, D.R. and Cheng, Y.C. (1992) *Cancer Res.* 52, 4280–4285.
- [81] Weil, D., Power, M.A., Webb, G.C. and Li, C.L. (1997) *Gene* 187, 115–122.
- [82] Sim, B.C., Grosveld, F.G. and Hui, K.M. (1990) *J. Immunogenet.* 17, 133–150.
- [83] Sved, J. and Bird, A. (1990) *Proc. Natl. Acad. Sci. USA* 87, 4692–4696.
- [84] Stein, G.S., van Wijnen, A.J., Stein, J.L., Lian, J.B., Pockwinse, S.H. and McNeil, S. (1999) *FASEB J.* 13 (Suppl), S157–166.
- [85] Huang, Z., Fasco, M.J. and Kaminsky, L.S. (1997) *Arch. Biochem. Biophys.* 343, 101–108.
- [86] Tayebi, N., Stern, H., Dymarskaia, I., Herman, J. and Sidransky, E. (1996) *Am. J. Med. Genet.* 66, 316–319.
- [87] Whang, Y.E., Wu, X. and Sawyers, C.L. (1998) *J. Natl. Cancer Inst.* 90, 859–861.
- [88] Liu, J. and Kagan, J. (1999) *BioTechniques* 26, 19–22.
- [89] Tschentscher, P., Wagener, C. and Neumaier, M. (1997) *Clin. Chem.* 43, 2244–2250.
- [90] Kreuzer, K.A., Lass, U., Landt, O., Nitsche, A., Laser, J., Ellerbok, H., Pauli, G., Huhn, D. and Schmidt, C.A. (1999) *Clin. Chem.* 45, 297–300.
- [91] Ruud, P., Fodstad, O. and Hovig, E. (1999) *Int. J. Cancer* 80, 119–125.
- [92] Richardson, M.P., Braybrook, C., Tham, M., Moore, G.E. and Stanier, P. (1998) *Gene* 206, 145–150.