FEBS Letters 567 (2004) 136–141 FEBS 28367

Minireview

Repetitive sequences that shape the human transcriptome

Anna Jasinska, Wlodzimierz J. Krzyzosiak*

Institute of Bioorganic Chemistry, Laboratory of Cancer Genetics, Polish Academy of Sciences, Noskowskiego 12/14 St., 61-704 Poznan, Poland

Received 29 February 2004; accepted 7 March 2004

Available online 22 April 2004

Edited by Horst Feldmann

Abstract Only a small portion of the total RNA transcribed in human cells becomes mature mRNA and constitutes the human transcriptome, which is context-dependent and varies with development, physiology and pathology. A small fraction of different repetitive sequences, which make up more than half of the human genome, is retained in mature transcripts and shapes their function. Among them are short interspersed elements (SINEs), of which Alu sequences are most frequent, and simple sequence repeats, which come in many varieties. In this review, we have focused on the structural and functional role of Alu elements and trinucleotide repeats in transcripts.

Keywords: Alu repeat; Trinucleotide repeat; RNA pathogenesis; Triplet repeat expansion disease

1. The human genome and transcriptome

A predominant part of the human genome consists of repetitive sequences of various types encompassing large segmental duplications, interspersed transposon-derived repeats and tandem repeats [1]. The latter include satellites, minisatellites and microsatellites known also as simple sequence repeats (SSRs). In contrast, the coding sequences of the nearly 25 000 human genes comprise about 1% of the genome, and about the same genome share can be assigned to their regulatory sequences (Fig. 1). This distinction between the repetitive and coding/regulatory sequences does not necessarily mean that these sequences are physically separated from each other in the genome. It often happens, that some of these repeats occur within genes and even within their coding sequences and perform regulatory functions. It also happens that the repeats increase the likelihood of deleterious mutations in their host genes, thus increasing the risk of disease.

The human genome is basically uniform and only about 0.3% of its sequence differs between individuals in a population. However, genetic information is the source of much higher diversity observed at the transcriptome level. The term transcriptome is young and its first appearance was in 1997 to describe the set of genes expressed from the yeast genome [2]. According to a more recent definition the term stands for the complete collection of transcribed elements of the genome [3]. In addition to mRNAs, it also includes various non-coding RNAs playing either structural or regulatory functions in cells. Hence, there are thousands of different transcriptomes in hundreds of different cell types and organs in their various physiological and pathological states. The variety of transcripts as compared to genes increases by about 50% due to alternative splicing [4,5], the numerous antisense transcripts are synthesized [6], and all non-protein-coding RNAs [7,8] including the recently recognized microRNA family [9] also make a significant contribution to the pool of human transcripts.

It has been known for some time that human mRNAs comprise about 2–3% of cellular RNA and they may be characterized as belonging to different abundance classes. A small number of mRNAs are synthesized in several thousands of copies, others occur in hundreds of copies, and the majority is present in less than ten copies per cell [10]. Altogether, as many as 500.000 mRNA molecules may exist in a single human cell. They are usually the products of about 25–50% human genes expressed in most tissues and cell types. Only in brain tissue is the number of expressed genes much higher. The above numbers give some impression of the complexity of the transcriptome, and show that its nearly complete characterization, as with the human genome sequence, may be an enormous task [3].

2. Alu sequences in human transcripts

Nearly half of the human genome derives from transposable elements (TEs) which are abundant in gene sequences and are also present in a significant portion of mature mRNAs, mostly in their untranslated regions [11]. Various TEs: the LTR, Alu, L1 and MIR sequences influence gene regulation at the level of transcription, e.g., by providing alternative promoters to many genes. Interestingly, the TEs are more prevalent within the mRNAs of the recently expanded gene families, which implies their role in genome evolution [12].

Among different TEs, the primate-specific Alu sequences are the most abundant and their 1.1 million copies account for more than 10% of the human genome [1]. They belong to five major subfamilies differing in the age of their appearance during hominoid evolution and have specific diagnostic differences in their sequences [13,14]. The oldest Alu sequences of the Alu-Jo and Alu-Jb elements are estimated to be 65–80 million years old, and the youngest Alu-Y class is about 15 million years old. The most abundant in the genome is the 30–50 million years old Alu S subfamily and in particular its Alu-Sx member. The

* Corrsponding author. Fax: +48-61-8520532.
E-mail address: wlodkrzy@ibch.poznan.pl (W.J. Krzyzosiak).

0014-5793/$22.00
doi:10.1016/j.febslet.2004.03.109

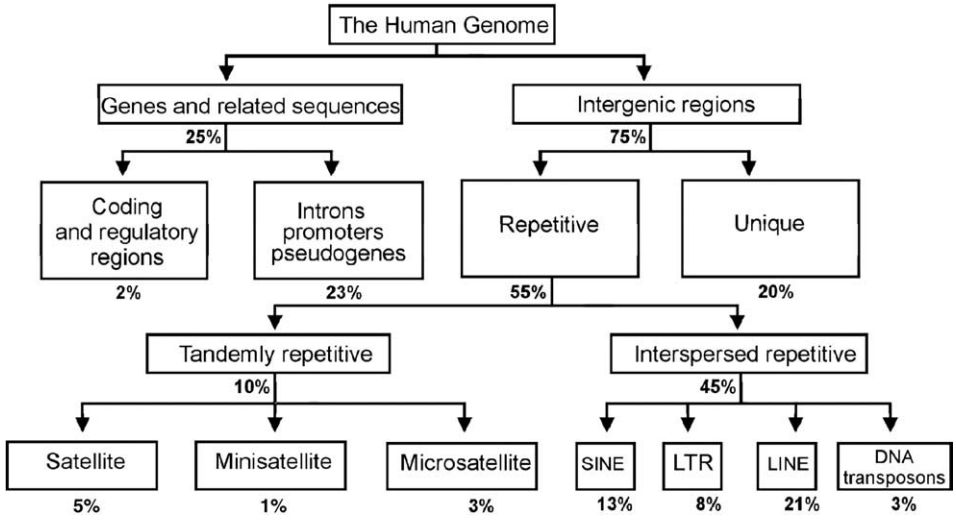

Fig. 1. Composition of the human genome. The percentage shares of various functional and non-functional sequences are shown.

common feature of all these Alu sequences is their length of about 280 nucleotides and their structure composed of two unequal monomer units [13,14].

The Alu sequences may change the genome in several different ways including insertion mutagenesis, gene conversion and recombination between the repeats [15]. They are also known as the genome-wide modifiers of gene expression. This can be achieved either by disrupting promoters, changing their methylation status or by inserting new regulatory signals, e.g., binding sites for steroid hormone receptors [16]. They may also shape human transcriptome by interfering with the splicing process. Harboring the motifs in their nucleotide sequences that resemble the splice sites, they can mimic exons and become a part of mature mRNAs [17]. As most of the Alu insertions to protein coding regions are deleterious to cells, the internal exons containing Alu sequences are often alternatively spliced, and selected against [18]. The Alu sequences may also create genetic novelties being a rich source of raw material for experiments carried out by the evolution [19].

The extreme example of the Alu-rich gene is *BRCA1* involved in the hereditary predisposition to breast and ovarian cancer. As much as 40% of its 80 kb genomic sequence is composed of Alu sequences [20]. The Alu elements occurring in various *BRCA1* introns were reported to cause numerous complex changes in the genomic, sequence which were associated with the increased risk of these cancers [21]. A number of other hereditary human diseases, e.g., hemophilia, have also been attributed to Alu insertions in the implicated genes, and more disorders including hypercholesterolemia, α-thalassemia and thrombofilia were caused by the Alu/Alu recombination [22]. These examples show that Alu sequences may be involved in diseases not only through their high content but also due to their specific locations within gene sequences.

Importantly, the Alu elements that occur in mRNAs may also influence their translation [23]. When Alu occurs in a leader sequence of mRNA, it may regulate the initiation step of translation. This is the case of the BRCA1 mRNA which has two forms differing in their leader sequences and patterns of expression [24]. The mRNA containing a shorter leader is expressed in normal mammary tissue, whereas the one with a longer leader containing the Alu sequence is expressed in breast cancer tissue. We have shown that mRNA with the longer 5′-UTR is translated with 10 times lower efficiency and that the stable structure formed by the Alu element is the major factor responsible for this effect [25]. Thus, the Alu-mediated *BRCA1* downregulation in sporadic breast cancer may be considered the effect equivalent to the gene mutation in hereditary cancer, which contributes to the decrease of BRCA1 protein in the tumor tissue. It is worth noting, that the Alu insert present in the BRCA1 leader is 60 nt shorter than the full length Alu, but its right monomer that forms the stable structure hampering the ribosome scanning is basically intact. Our transcriptome-wide survey performed with bioinformatic tools revealed that about 4% of human mRNAs harbor Alu elements and those containing intact right monomers in their leaders may exploit the same regulatory mechanism [25].

3. SSRs in the human genome

The SSRs known also as short tandem repeats or microsatellites are defined as tandemly repeated tracts of DNA composed of 1–6 base pair long motifs. The number of repeats in a tract is usually less than ten and if it is higher the repeat length polymorphism is often observed [26]. In rare cases, the repeat number may reach hundreds and even thousands. All SSRs taken together occupy about 3% of the human genome in which they are widely dispersed and associated with many genes [27]. The genomic distribution of the SSRs of different motif lengths and sequences is strongly biased. With the exception of the most prevalent SSRs composed of monomeric motifs, the classes of repeated dimers, tetramers and hexamers, which show nearly equal density in the genome, are 3–4 times more abundant than those of trimers and pentamers [28]. When the occurrence of SSRs in different functional genome regions is considered, it turns out that most of them show much higher density in non-coding regions. Exceptions to the rule are trimers and hexamers which are nearly two times more prevalent in exons compared to introns and intergenic regions [28,29]. Their high frequency in coding regions may be explained by the fact that they do not change the reading frames and gene coding properties and, thus, are much better tolerated than other

SSRs. Their positive selection in exons suggests also some function for the repeats. This, however, remains barely known despite the fact that the properties of trinucleotide repeats have been extensively studied over the past decade. The high mutation rate of these repeats and their frequent length polymorphism have raised speculations that they may be involved in the regulation or "fine tuning" of gene expression and function, and have quantitative effects on phenotype [30].

4. Trinucleotide repeats in the human transcriptome

Focusing now on triplet repeats in the human transcriptome, we have recently asked the questions: how many different mRNAs harbor such repeats in their sequences? At what frequencies do certain types of repeats occur? And, what are the preferred repeat locations in mRNAs? To answer these questions, the GenBank database was trawled for all 20 different triplet repeat tracts composed of at least six repeats, and the 718 repeat tracts were found in 619 mRNAs [31]. Most of the identified mRNAs (87%) carried short repeat tracts 6–10, those harboring 11–20 repeats contributed 11%, and those containing more than 20 repeats – only 2% to the total. The most frequently occurring repeated motifs were CAG, CGG, CCG, CUG, AGG and ACC, whereas the ACG, AUC, CUU, AGU, CGU and ACU were very poorly represented. Considering different mRNA regions, most of the repeats were located in the ORF (67%), followed by 5′-UTR (24%) and 3′-UTR (9%). The GC-rich repeats were more prevalent in the 5′-UTR, whereas the AU-rich were more abundant in the 3′-UTR. Taking into account length differences between the mRNA untranslated regions, the repeats were strongly over-represented in mRNA leaders implying their role in translation regulation [31].

5. CNG repeats form hairpin structures in transcripts

As the first step in searching for any biological functions of trinucleotide repeats in transcripts, we performed a systematic structure analysis using a battery of chemical and enzymatic probes. These experiments, besides answering important questions concerning the structure formation abilities of different repeats ([32], unpublished data), provided unique information regarding the specificity of different structure probes acting on highly regular RNA sequences. Out of the investigated 20 model transcripts composed of all possible triplet repeat motifs reiterated 17 times, 6 transcripts were shown to form hairpin structures. Among them were all the CNG repeats shown by our bioinformatic analysis to be the most abundant in transcripts [31]. The quasistable hairpin stem structure is composed of the periodically occurring C–G, G–C pairs and N–N mismatches, and the repeats show a tendency to assume several variantive alignments if this process is not suppressed by the presence of a suitable GC clamp [32]. The characteristic feature of hairpins formed by the CNG repeats is that their structure rigidity increases with repeat length. For example, hairpin formed by the (CUG)49 has a stem structure which shows a melting temperature much higher than that of (CUG)21 and is completely resistant to the single strand specific probes [33]. Moreover, the long CUG and CAG hairpins behave as regular double-stranded RNA structures with typical RNA-A geometry [34].

It is known that the long double-stranded RNAs are toxic to human cells possibly due to their non-specific effects on gene expression [35]. They are likely to induce the interferon (IFN-α/β)-related pathways stimulating apoptosis and activate the 2′,5′-oligoA synthetase/RNaseL enzymes resulting in RNA degradation. The long dsRNAs activate also the RNA-dependent protein kinase (PKR)-mediated antiviral responses including the general inhibition of translation and induction of cell death [35]. In this context it is interesting to note that the CAG and CUG repeat hairpins were reported to activate PKR [36,37], whereas the CGG hairpin was shown to be inactive [38]. The double-stranded RNAs can also induce RNA interference (RNAi), which is a sequence specific RNA degradation mechanism developed to combat viral RNA, operating in many organisms including humans [39]. The ribonuclease Dicer, which cuts the dsDNA into about 22nt long RNA duplexes, was recently shown to cut the long CGG repeat hairpins also [38]. According to our experience, all types of the CNG repeat hairpins are Dicer substrates if they are long enough (unpublished data).

6. Mechanism of pathogenesis in triplet repeat expansion diseases

The pathways described above in which the long dsRNAs execute their toxic effects in cells are not the only ones. There is also another mechanism of RNA pathogenesis mediated by SSRs. Before discussing its various aspects, we provide some background in brief. In 20 human genes the polymorphic repeats, in most cases triplet repeats of the CNG type, undergo pathogenic expansions that cause hereditary human neurological diseases, the so called triplet repeat expansion diseases (TREDs) [40]. The best known examples of these diseases are: myotonic dystrophy, fragile X syndrome, Huntington disease and a number of spinocerebellar ataxias. The expandable repeats are present in all parts of the implicated genes, mostly in their ORFs but also in 5′-UTRs, 3′-UTRs and introns as shown schematically in Fig. 2. In the majority of the disease-related genes containing the CAG repeats in their coding sequences, the pathological repeat number begins at about 40 and may reach 100. On the contrary, the expansions of repeats located in non-coding regions are usually larger and more variable. The repeats present in the translated regions are thought to exert their pathogenic functions on the level of proteins which in most cases contain expanded polyglutamine tracts [41]. Transcripts with the expanded repeats present in non-coding regions, as they do not give rise to aberrant proteins, may be transcriptionally silenced or translationally inhibited as in the fragile-X-syndrome or tremor ataxia syndrome, respectively [42]. In both disorders the *FMR1* gene is implicated. They may also activate some other RNA-mediated mechanisms like that of the specific protein sequestration postulated for myotonic dystrophies type 1 and type 2 (involving the *DMPK* and *ZNF1* genes, respectively) [43].

The originally proposed mechanism of RNA pathogenesis in myotonic dystrophy implied that long CUG repeats confiscate specific repeat binding proteins from their normal binding sites in other transcripts and compromise their function [44]. We have shown that such repeats form stable hairpins, thus the sequestered proteins must be the dsCUG repeat binding proteins [33]. Such proteins were then isolated [45] and shown to

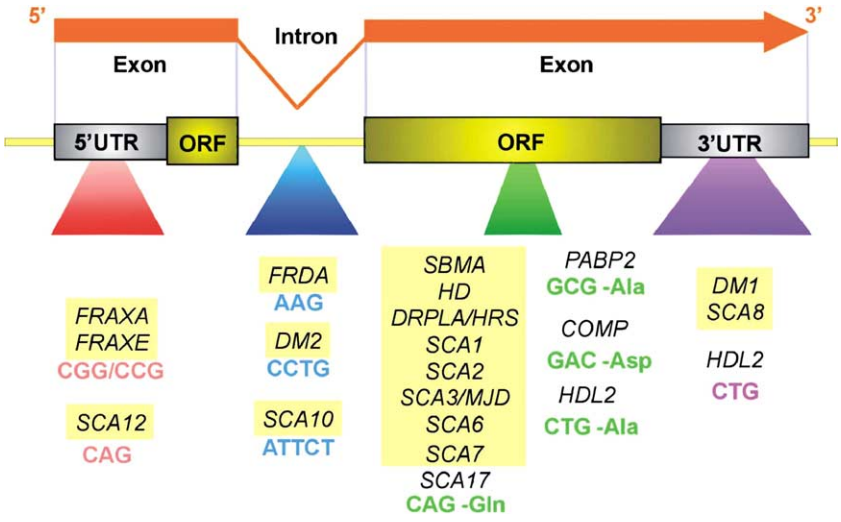

Fig. 2. Location of expandable repeats in the genes involved in repeat expansion diseases. The disorders for which the involved genes and transcripts were subjected to detailed studies in our laboratory are highlighted with a yellow background.

co-localize with the expanded transcripts in nuclear foci [46–48]. More recently, the muscleblind knockout mouse model for myotonic dystrophy revealed that the lack of this protein leads indeed to abnormalities characteristic for the disease in humans [49]. In this way the protein sequestration mechanism was further supported. Interestingly, the CUG repeat expansion also causes the up-regulation of the single-stranded CUG repeat binding proteins (CELF), which results in the altered expression of several target genes [50].

7. Unified model of RNA-mediated pathogenesis in TREDs

The question whether a similar RNA-mediated mechanism of pathogenesis may operate also in other TREDs was recently addressed by several authors. Not only the diseases caused by the repeat expansions in non-coding sequences [42,51] but also the so-called polyglutamine diseases were considered in this context [31,32,52]. We have shown that at least the RNA hairpin structures, which are behind the RNA pathogenesis in myotonic dystrophy have the same architecture also in other TREDs-related transcripts [32,33]. Thus, they have the potential to interact with the specific dsRNA binding proteins. In line with that, the dsCAG repeat binding proteins were described [53] and identification of the dsCGG repeat binding protein was reported [54].

It has been argued by the proponents of the more prominent role of RNA in pathogenesis of TREDs that the toxic effect of the expanded polyglutamine containing proteins may not be the only mechanism of toxicity and may not be the one responsible for neuronal dysfunction [52]. Indeed, the polyglutamine protein aggregation was shown to be neither necessary nor sufficient for neurodegeneration [41]. Among other arguments were the facts that most TREDs show dominant inheritance independent of the localization of mutation, and that SCA8 is characterized by similar clinical phenotype as other spinocerebellar ataxias despite the fact that SCA8 RNA is a non-coding antisense transcript containing expanded CUG repeat [52]. Thus, the common molecular feature of these diseases independent of the position of the repeat expansion could be the presence of expanded transcripts in the nucleus, in some diseases also in cytoplasm, in most cases in the form of long stable hairpin structures.

According to the hypothesis which was put forward recently [31], not only the expanded repeat hairpins but also the overexpressed transcripts containing long repeats (still within normal length range) could either trigger pathogenesis or at least modulate the pathogenic effect. This hypothesis was rooted in the observation that not only the expanded CUG repeat hairpin but also the much shorter one (CUG)20 showed a considerable binding of the muscleblind protein [45]. From the perspective of the agent triggering pathogenesis, two factors are important for the protein sequestration mechanism: the suitable structure and sufficient length of the repeat hairpin stem, and the copy number of its harboring transcript per cell. This is so because the number of effective protein binding sites is the factor which matters.

In support of this hypothesis and the unified model for RNA pathogenesis in TREDs, there are recent results showing RNA-mediated neurodegeneration caused by the *FMR1* permutation [55]. The hairpin structure formed by the CGG repeats in the FMR1 transcript [38,56] together with its 5–10-fold increased level [57] likely to activate RNA pathogenesis in permutation carriers [55]. The resulting syndromes, tremor-ataxia and premature ovarian failure, differ significantly in their clinical features from the fragile X syndrome developed in full mutation carriers [57]. The *FMR1* example shows that two different mechanisms of pathogenesis may be triggered depending on the different number of repeats within the same gene.

8. Other factors relevant to protein sequestration mechanism

The mechanism of RNA pathogenesis has to be considered also from the perspective of transcripts depleted from the proteins they normally bind and having their normal functions impaired. A group of such transcripts has been preliminarily selected using bioinformatic methods [31]. The difficulty with their selection is that the status of their repeat length polymorphism is known for about 10% of the identified repeat

tracts only. Thus, systematic genotyping is required in order to determine their normal repeat length ranges and stable hairpin structure formation abilities.

Our initial sampling of the 10 TREDs-related genes (20 alleles) showed that the number of long alleles (containing more than 20 repeats) ranged from 1 to 8 in genomes of the investigated individuals [Jasinska et al. submitted]. This result shows that the genetic background in which the expanded repeat exerts its pathogenic effect may vary significantly between individuals in a population. However, for the mechanism of RNA pathogenesis in TREDs the transcriptome background is more relevant than that of the genome. Therefore, the expression patterns and levels of the repeat containing transcripts need to be determined in cells and tissues which are affected by the diseases and in those unaffected as well. This knowledge, which is being gathered now, is required because not only the presence or absence of the specific transcripts but also their abundance in a given cell type may be important.

As the structures formed by the repeats alone and by the repeats in their host transcripts may differ, it was necessary to find out what is the architecture of the repeat regions within their natural sequence context. This problem was also important from the perspective of the cellular function of normal length repeats as well as from that of the putative role of long normal repeats in pathogenesis. According to structure prediction, in a number of TREDs-related transcripts: the FMR2, AR, SCA6, SCA7 and SCA12 repeat flanking sequences contribute significantly to the stability of the repeat hairpins [31]. For most of these transcripts, these predictions were experimentally confirmed ([56], Michlewski and Krzyzosiak, submitted). In several other studied TREDs-related transcripts, the repeats have more freedom of alignment and form a number of slipped hairpin variants ([33], unpublished data). If the unified RNA-mediated mechanism of pathogenesis indeed operates in cells of TREDs patients, the various contributions from flanking sequences may play the role of co-factors involved in determining different pathogenic repeat length thresholds in different diseases.

9. Structural role of repeat interruptions in transcripts

In three out of the 20 TREDs-related genes: the *SCA1*, *SCA2* and *FMR1*, their normal variants contain specific interruptions in the repeat tracts, and these interruptions are absent in the expanded mutant alleles. The repeat expansions are thought to occur predominantly as the result of DNA slippage during replication and the loss of interruptions was postulated either to precede or to follow the repeat expansion [58]. As a result of a comprehensive genotyping study, the patterns of repeat interruptions in the *SCA1* and *SCA2* were determined for the Polish population and their relevance to the repeat expansion mechanism and disease incidence was discussed [58].

The presence of interruptions in the repeated sequence observed in some genes is the cellular strategy which prevents repeat expansions in DNA. To characterize the normal functions of the repeats in RNA, the roles played by the repeat interruptions have to be established. This has been done for the transcripts of all three implicated genes using chemical and biochemical methods [Napierala et al. submitted, Sobczak and Krzyzosiak submitted]. In the case of the SCA1 RNA, the structures of representative alleles containing either one, two or three CAU interruptions present at different localizations within the repeat tract were analyzed. It turned out that, that the interruptions either enlarge the terminal loop of the hairpin formed by the CAG repeats, nucleate internal loops in the hairpin stem, or force the repeats to form two smaller hairpin structures. Thus, their role in RNA is to destabilize the repeat hairpin structure, most likely to decrease its ability to interact with dsCAG repeat binding proteins and participate in the putative RNA pathogenesis mechanism [Sobczak and Krzyzosiak, submitted]. A similar role can be assigned to repeat interruptions in the SCA2 and FMR1 transcripts, although several details of structure destabilization strategy are different.

10. Concluding remarks

In this article, we have shown the variety of ways in which just two different classes of human repetitive sequences may influence gene expression and function at the level of transcripts. In the case of Alu sequences, it turns out that the insertion site matters as well as the size and nature of the inserted fragment. All these factors may be important for the effects that Alu elements generate and these effects may range from modulatory to deleterious. In the case of triplet repeats, it is their length variation and localization within a transcript which determine their variable effects on phenotype, and may decide which mechanism of the pathogenesis is activated. As the field develops the views regarding the contributions of different pathomechanisms evolved, and the concept of RNA-mediated pathogenesis, unifying most if not all TREDs, is becoming more sound. The specific protein sequestration by dsRNA repeat, which triggers the mechanism of RNA pathogenesis, was discussed here in more details. This is a sort of protein "gain and loss" game in which numerous RNA and protein players participate. Our ongoing research is focused on the identification and detailed characterization of all RNA players in this game. We use RNA structure analysis and the tools of genetics to answer the question of how the repeat polymorphism in genes translates to RNA structure varieties and shapes transcriptomes in humans.

Acknowledgements: We thank all present and former laboratory members and students for their contributions to the research presented here. This work was supported by the State Committee for Scientific Research Grants 6P04B-03118 and PBZ/KBN/040/P04/2001 and the Foundation for Polish Science Grants 117/96 8/2000.

References

[1] International Human Genome Sequencing Consortium (2001) Nature 409, 860–921.
[2] Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B. and Kinzler, K.W. (1997) Cell 88, 243–251.
[3] Ruan, Y., Le Ber, P., Ng, H.H. and Liu, E.T. (2004) Trends Biotechnol. 22, 23–30.
[4] Black, D.L. (2003) Annu. Rev. Biochem. 72, 291–336.
[5] Zavolan, M., Kondo, S., Schonbach, C., Adachi, J., Hume, D.A., Hayashizaki, Y. and Gaasterland, T. (2003) RIKEN GER Group; GSL Members. Genome Res. 13, 1290–1300.
[6] Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K. and Rotman, G. (2003) Nat. Biotechnol. 21, 379–386.

[7] Numata, K., Kanai, A., Saito, R., Kondo, S., Adachi, J., Wilming, L.G., Hume, D.A., Hayashizaki, Y. and Tomita, M. (2003) RIKEN GER Group; GSL Members. Genome Res. 13, 1301–1306.
[8] Herbert, A. (2004) Nat. Genet. 36, 19–25.
[9] Ambros, V. (2001) Cell 107, 823–826.
[10] Jackson, D.A., Pombo, A. and Iborra, F. (2000) FASEB J. 14, 242–254.
[11] Yulug, I.G., Yulug, A. and Fisher, E.M. (1995) Genomics 27, 544–548.
[12] van de Lagemaat, L.N., Landry, J.R., Mager, D.L. and Medstrand, P. (2003) Trends Genet. 19, 530–536.
[13] Kapitonov, V. and Jurka, J. (1996) J. Mol. Evol. 42, 59–65.
[14] Dagan, T., Sorek, R., Sharon, E., Ast, G. and Graur, D. (2004) Nucleic Acids Res. 32, D489–D492.
[15] Batzer, M.A. and Deininger, P.L. (2002) Nat. Rev. Genet. 3, 370–379.
[16] Vansant, G. and Reynolds, W.F. (1995) Proc. Natl. Acad. Sci. USA 92, 8229–8233.
[17] Makalowski, W., Mitchell, G.A. and Labuda, D. (1994) Trends Genet. 10, 188–193.
[18] Sorek, R., Ast, G. and Graur, D. (2002) Genome Res. 12, 1060–1067.
[19] Makalowski, W. (2003) Science 300, 1246–1247.
[20] Welcsh, P.L. and King, M.C. (2001) Hum. Mol. Genet. 10, 705–713.
[21] Puget, N., Stoppa-Lyonnet, D., Sinilnikova, O.M., Pages, S., Lynch, H.T., Lenoir, G.M. and Mazoyer, S. (1999) Cancer Res. 59, 455–461.
[22] Deininger, P.L. and Batzer, M.A. (1999) Mol. Genet. Metab. 67, 183–193.
[23] Rubin, C.M., Kimura, R.H. and Schmid, C.W. (2002) Nucleic Acids Res. 30, 3253–3261.
[24] Xu, C.F., Brown, M.A., Chambers, J.A., Griffiths, B., Nicolai, H. and Solomon, E. (1995) Hum. Mol. Genet. 4, 2259–2264.
[25] Sobczak, K. and Krzyzosiak, W.J. (2002) J. Biol. Chem. 277, 17349–17358.
[26] Wren, J.D., Forgacs, E., Fondon III, J.W., Pertsemlidis, A., Cheng, S.Y., Gallardo, T., Williams, R.S., Shohet, R.V., Minna, J.D. and Garner, H.R. (2000) Am. J. Hum. Genet. 67, 345–356.
[27] Subramanian, S., Madgula, V.M., George, R., Mishra, R.K., Pandit, M.W., Kumar, C.S. and Singh, L. (2003) Bioinformatics 19, 549–552.
[28] Subramanian, S., Mishra, R.K. and Singh, L. (2003) Genome Biol. 4, R13.
[29] Toth, G., Gaspari, Z. and Jurka, J. (2000) Genome Res. 10, 967–981.
[30] Kashi, Y., King, D. and Soller, M. (1997) Trends Genet. 13, 74–78.
[31] Jasinska, A., Michlewski, G., de Mezer, M., Sobczak, K., Kozlowski, P., Napierala, M. and Krzyzosiak, W.J. (2003) Nucleic Acids Res. 31, 5463–5468.
[32] Sobczak, K., de Mezer, M., Michlewski, G., Krol, J. and Krzyzosiak, W.J. (2003) Nucleic Acids Res. 31, 5469–5482.
[33] Napierala, M. and Krzyzosiak, W.J. (1997) J. Biol. Chem. 272, 31079–31085.
[34] Michalowski, S., Miller, J.W., Urbinati, C.R., Paliouras, M., Swanson, M.S. and Griffith, J. (1999) Nucleic Acids Res. 27, 3534–3542.
[35] McManus, M.T. and Sharp, P.A. (2002) Nat. Rev. Genet. 3, 737–747.
[36] Tian, B., White, R.J., Xia, T., Welle, S., Turner, D.H., Mathews, M.B. and Thornton, C.A. (2000) RNA 6, 79–87.
[37] Peel, A.L., Rao, R.V., Cottrell, B.A., Hayden, M.R., Ellerby, L.M. and Bredesen, D.E. (2001) Hum. Mol. Genet. 10, 1531–1538.
[38] Handa, V., Saha, T. and Usdin, K. (2003) Nucleic Acids Res. 31, 6243–6248.
[39] Hannon, G.J. (2002) Nature 418, 244–251.
[40] Cummings, C.J. and Zoghbi, H.Y. (2000) Hum. Mol. Genet. 9, 909–916.
[41] Orr, H.T. (2001) Genes Dev. 15, 925–932.
[42] Bardoni, B. and Mandel, J.L. (2002) Curr. Opin. Genet. Dev. 12, 284–293.
[43] Tapscott, S.J. and Thornton, C.A. (2001) Science 293, 816–817.
[44] Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R.C., Angelini, C. and Hoffman, E.P. (1995) Hum. Mol. Genet. 4, 599–606.
[45] Miller, J.W., Urbinati, C.R., Teng-Umnuay, P., Stenberg, M.G., Byrne, B.J., Thornton, C.A. and Swanson, M.S. (2000) EMBO J. 19, 4439–4448.
[46] Fardaei, M., Larkin, K., Brook, J.D. and Hamshere, M.G. (2001) Nucleic Acids Res. 29, 2766–2771.
[47] Mankodi, A., Urbinati, C.R., Yuan, Q.P., Moxley, R.T., Sansone, V., Krym, M., Henderson, D., Schalling, M., Swanson, M.S. and Thornton, C.A. (2001) Hum. Mol. Genet. 10, 2165–2170.
[48] Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Hum. Mol. Genet. 11, 805–814.
[49] Kanadia, R.N., Johnstone, K.A., Mankodi, A., Lungu, C., Thornton, C.A., Esson, D., Timmers, A.M., Hauswirth, W.W. and Swanson, M.S. (2003) Science 302, 1978–1980.
[50] Faustino, N.A. and Cooper, T.A. (2003) Genes Dev. 17, 419–437.
[51] Ranum, L.P. and Day, J.W. (2002) Curr. Opin. Genet. Dev. 12, 266–271.
[52] Broude, N.E. and Cantor, C.R. (2003) Expert Rev. Mol. Diagn. 3, 269–274.
[53] McLaughlin, B.A., Spencer, C. and Eberwine, J. (1996) Am. J. Hum. Genet. 59, 561–569.
[54] Rosser, T.C., Johnson, T.R. and Warren, S.T. (2002) Am. J. Hum. Genet. 71, 507.
[55] Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K. and Warren, S.T. (2003) Neuron 39, 739–747.
[56] Krzyzosiak, W.J., Napierala, M. and Drozdz, M. (1999) in: RNA Biochemistry and Biotechnology (Barciszewski, J. and Clark, F.C., Eds.), Vol. 70, pp. 303–314, Kluwer Academic Publishers, Netherlands.
[57] Hagerman, R.J. and Hagerman, P.J. (2002) Curr. Opin. Genet. Dev. 12, 278–283.
[58] Sobczak, K. and Krzyzosiak, W. (2004) Hum. Mutat. (in press).