

REGULACJA EKSPRESJI GENÓW PRZEZ DŁUGIE NATURALNIE WYSTĘPUJĄCE ANTYSENSOWNE TRANSKRYPTY

GENE EXPRESSION REGULATION BY LONG NATURALLY OCCURRING ANTISENSE TRANSCRIPTS

Adam MASTER, Alicja NAUMAN

Centrum Medyczne Kształcenia Podyplomowego, Zakład Biochemii i Biologii Molekularnej,
ul. Marymoncka 99/103, 01-813 Warszawa, Polska

Streszczenie: W sekwencji ludzkiego genomu zapisana jest informacja na temat budowy i funkcji całego organizmu. O charakterystycznych cechach poszczególnych komórek i tkanek decydują mechanizmy regulacji ekspresji genów, w których istotną rolę odgrywają niedoceniane dotąd długie, naturalnie występujące, antysensowne transkrypty NAT (ang. *Naturally occurring Antisense Transcripts, NATs*). NAT wykazują pełną lub częściową komplementarność względem funkcjonalnych genów, kontrolując ich ekspresję na poziomie ukierunkowanej metylacji DNA, transkrypcji pre-mRNA, różnicowego składania transkryptu pierwotnego jak również redagowania RNA, transportu i zlokalizowanej translacji białek. W przeciwieństwie do grupy krótkich, niekodujących RNA (sncRNA) takich jak mikroRNA, długie niekodujące RNA (lncRNA), będąc najliczniejszą grupą transkryptów NAT, wykazują zwykle większe dopasowanie sekwencji, a tym samym specyficzność działania względem docelowych genów. Wiązanie tych cząsteczek do DNA i RNA może przesłaniać elementy regulatorowe lub wpływać na stabilność funkcjonalnych struktur takich jak G-kwadrupleksy i wewnętrzne miejsca wejścia rybosomów (elementy IRES). Nowe technologie sekwencjonowania całych genomów i przeszukiwania bibliotek cDNA umożliwiły ustalenie, że ok. 20% ludzkich i mysich genów współdzieli sekwencje z innymi genami, a ich transkrypty tworzyć mogą komplementarne pary sens-antysens. Chociaż fizjologiczne znaczenie większości tych par pozostaje niejasne, zidentyfikowano i opisano antysensowne transkrypty, które regulują ekspresję genów związanych z podstawowymi procesami takimi jak wzrost, proliferacja, różnicowanie, apoptoza, metabolizm jak również cykl okołodobowy. Zaburzenie ekspresji NAT stwierdzono w przebiegu wielu chorób, w tym nowotworowych, co wskazuje na potrzebę ustalenia ich roli na poszczególnych poziomach przepływu informacji genetycznej. Niniejsza praca dokonuje przeglądu aktualnej wiedzy w zakresie regulacji ekspresji genów przez długie RNA wykazujące aktywność NAT. Zdolność do tworzenia par sens-antysens i aktywacji mechanizmów kontroli ekspresji docelowych genów, stanowiła podstawę zaproponowanej definicji i strukturalno-funkcjonalnej charakterystyki tej niejednorodnej grupy transkryptów.

Słowa kluczowe: długie naturalnie występujące antysensowne transkrypty - NAT, G-kwadrupleks, komórki nowotworowe, lncRNA, pseudogeny.

Summary: Human genome sequence contains information about structure and function of the whole organism. However, specific features of individual cells and tissues are determined by gene expression regulation mechanisms, in which long naturally occurring antisense transcripts (NATs) may play a substantial role, largely underestimated so far. These molecules show full or partial complementarity to functional genes and may control their expression on the level of RNA-directed DNA methylation, pre-mRNA transcription, alternative splicing as well as RNA editing, transport and localized protein translation. In contrast to a group of short non-coding RNAs (sncRNAs) such as microRNA, the NATs including long non-coding RNAs (lncRNAs) usually exhibit the higher complementarity, and hence the higher specificity towards target genes. Binding of these molecules to DNA or RNA may obstruct regulatory elements or influence the stability of some functional structures such as G-quadruplexes and internal ribosome entry sites (IRES elements). The new technologies of whole genome sequencing and screening of cDNA libraries revealed that about 20% of human and mouse genes overlap resulting in formation of potential sense-antisense pairs. Though, the physiological relevance of most of the pairs remains obscure, there are a number of identified and annotated antisense transcripts that are known to regulate expression of genes involved in cell growth, proliferation, differentiation, apoptosis, metabolism and circadian cycle. NATs expression have been found to be aberrant in the course of various diseases including cancers that indicates the need for understanding the roles of these transcripts at each step in the flow of genetic information. This paper provides a review of current knowledge on gene expression regulation by long

RNAs showing NATs activity. Moreover, a definition as well as structural and functional characteristics of the heterogeneous group of transcripts has been proposed on the basis of the NATs' ability to form sense-antisense pairs and the subsequent potential to activate mechanisms controlling target gene expression.

Keywords: cancer cells, G-quadruplex, long naturally occurring antisense transcripts - NATs, lncRNA, pseudogenes.

Wykaz zastosowanych skrótów: NAT, naturalnie występujące, antysensowne transkrypty (ang. *Naturally occurring Antisense Transcripts*, NATs); lncRNA, długie niekodujące RNA (ang. *long non-coding RNA*); sncRNA, krótkie niekodujące RNA (ang. *short non-coding RNA*); NR1D1, gen kodujący transkrypt Rev-Erb α (cis-NAT); GB.ac.no., NCBI GeneBank accession number; NCBI, Narodowe Centrum Informacji Biotechnologicznej w USA (ang. *National Center for Biotechnology Information*); GeneBank, Bank informacji o genach NCBI.

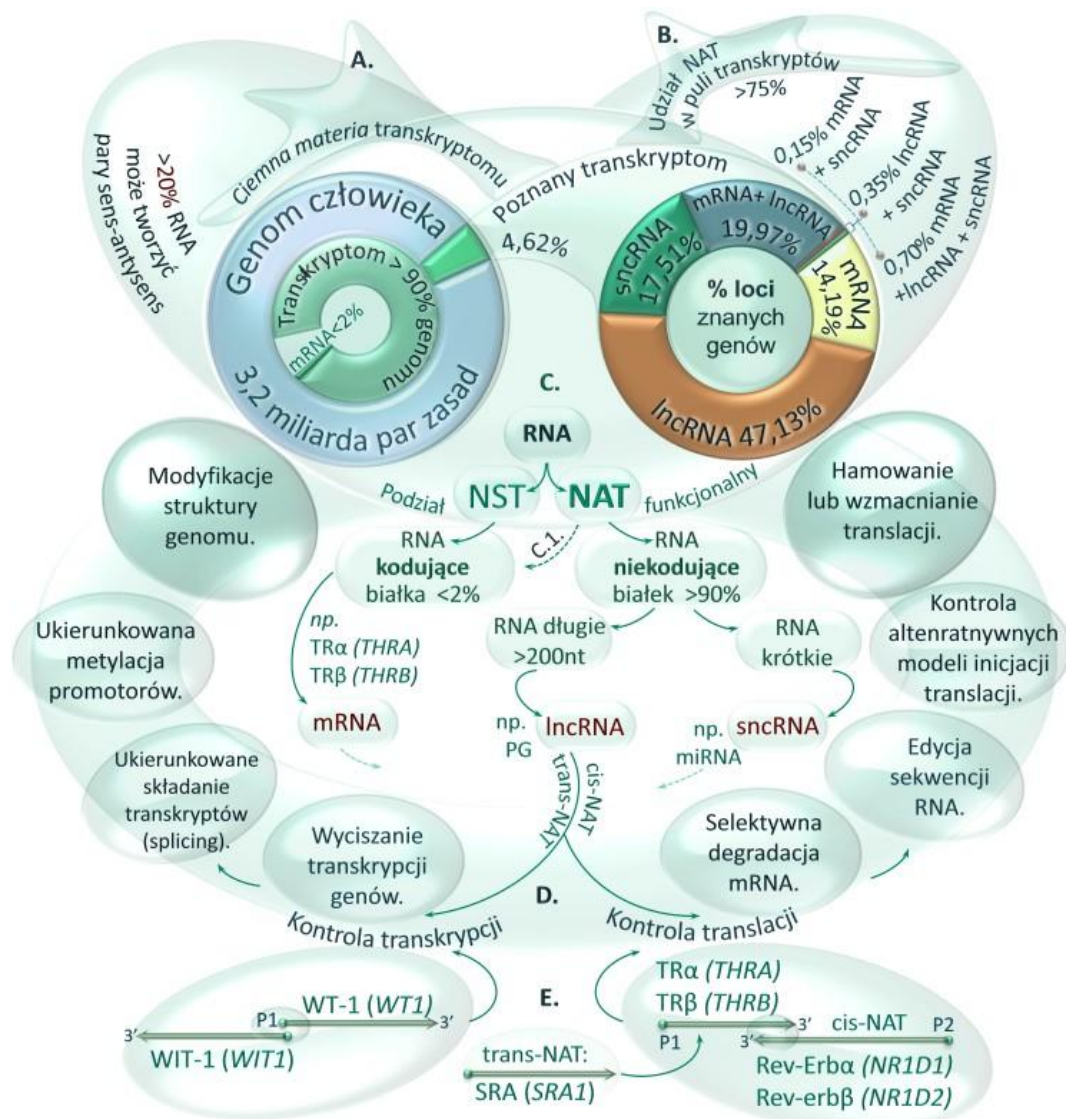
I. WSTĘP

Długie, naturalnie występujące antysensowne transkrypty **NAT** (ang. *Naturally occurring Antisense Transcripts*) stanowią niejednorodną klasę endogennych cząsteczek RNA (>200nt), które łączy funkcjonalna zdolność do komplementarnego i zwykle specyficznego wiązania się z docelowymi cząsteczkami sensownych kwasów nukleinowych (DNA, RNA). Takie pary (sens-antysens) uruchamiać mogą szereg mechanizmów prowadzących do zmiany ekspresji docelowych genów położonych w tym samym *locus* (**cis-NAT**) lub w oddalonych loci (**trans-NAT**) [51, 75]. Funkcję transkryptu NAT mogą pełnić zarówno długie niekodujące białek RNA (**lncRNA**), transkrypty pseudogenów, jak i cząsteczki RNA kodujące funkcjonalne białka, które rozpoznają inne cząsteczki RNA. Aktywność NAT wykazują także niektóre dojrzałe mRNA, tworzące kompleksy dwuniciowego RNA (dsRNA) z innymi mRNA [8, 45]. Odrębną klasę endogennych transkryptów NAT stanowią krótkie niekodujące RNA (sncRNA, zwykle < 90nt), tj. szeroko opisywane w literaturze mikroRNA, tRNA, rRNA, snRNA, snoRNA, piRNA oraz komórkowe siRNA i saRNA [17].

Komplementarność długich transkryptów cis-NAT względem docelowych RNA najczęściej nie jest przypadkowa i może wynikać z dwukierunkowej transkrypcji genów, prowadzonej jednocześnie na sensownej (kodującej) i antysensownej (matrycowej) nici DNA ($\rightarrow\leftarrow$). Transkrypcja dwukierunkowa prowadzić może do powstania dwóch częściowo pokrywających się cząsteczek transkryptu pierwotnego (pre-RNA), których komplementarne dopasowanie nie pozostaje bez wpływu na kolejne etapy przepływu informacji genetycznej. W przeciwieństwie do cis-NAT, długie cząsteczki trans-NAT wymagają dłuższych odcinków pełnej komplementarności, która umożliwia efektywne ich wiązanie do sekwencji DNA i RNA odległych loci [51, 75].

Każda komórka somatyczna dysponuje pełną informacją genetyczną na temat budowy i funkcji całego organizmu. Komórki różnych tkanek posiadają identyczną sekwencję DNA, chociaż odmienny program ekspresji genów, który prowadzi do ich zmienności fenotypowej [7]. Pierwszym poziomem kształtowania się tej zmienności jest transkryptom, który obejmuje zestaw wszystkich zsyntetyzowanych przez komórkę, kodujących i niekodujących cząsteczek RNA. Analiza genomu człowieka wykazała, że ponad 90% sekwencji DNA może ulegać mniej lub bardziej wydajnej transkrypcji, a niespełna 2% genomu koduje funkcjonalne białka [51], co oznacza, że zdecydowana większość transkryptów komórki nie koduje białek (Ryc.1). Ustalono także, że sekwencje długich niekodujących RNA zajmują samodzielnie 47% wszystkich poznanych dotąd loci chromosomów człowieka i dodatkowo 20% loci łącznie z sekwencjami mRNA (loci dwu-genowe: lncRNA-mRNA, Ryc.1) [51]. Co ciekawe, większość sekwencji NAT, zlokalizowanych jest w bliskim sąsiedztwie genów funkcjonalnych białek, zwłaszcza w pobliżu 3'-końcowej sekwencji kodującej, jak również w okolicy promotorowej tych genów [51]. Jednocześnie oszacowano, że ponad 20% ludzkich transkryptów może tworzyć pary sens-antysens i chociaż nie zbadano znaczenia większości z nich (skąd określenie *ciemnej materii transkryptomu*), wiadomo, że w skali całych sieci genetycznych ekspresja dziesiątek tysięcy (89 981 loci) lncRNA może odgrywać kluczową rolę w regulacji transkryptomu komórek [51]. Obserwacje te skłaniają zatem do rewizji dotychczasowych poglądów na temat roli długich transkryptów NAT w kontroli ekspresji genów, co z kolei wskazuje na konieczność analizy ich genomowego kontekstu.

Niniejsza praca zawiera przegląd literatury prezentującej aktualny stan wiedzy na temat strukturalno-funkcjonalnych właściwości długich, naturalnie występujących antysensownych transkryptów (NAT) oraz ich roli w wielopoziomowej regulacji ekspresji genów (Ryc.1).



RYCINA 1. Funkcjonalno-strukturalna klasyfikacja transkryptów (RNA).

A. Szacowana wielkość transkryptomu w kontekście sekwencji genomu człowieka. Ocenia się, że ponad 90% sekwencji genomu (ok. 3,2 miliarda par nukleotydów, nt) ulega transkrypcji, podczas gdy niespełna 2% koduje transkrypty funkcjonalnych białek (mRNA). Biologiczne znaczenie większości transkryptów nie jest znane (skąd określenie *ciemnej materii transkryptomu*), choć wiadomo, że ponad 20% komórkowych RNA może hybrydyzować ze sobą tworząc pary sens-antysens. **B. Udział poszczególnych klas RNA w poznanym dotąd fragmencie 4,62% transkryptomu człowieka.** Na wykresie kołowym przedstawiono udział procentowy liczby *loci* (miejsz) zajmowanych przez poszczególne klasy RNA tj: matrycowych RNA (mRNA), długich niekodujących RNA (lncRNA), krótkich niekodujących RNA (sncRNA), jak również *loci* zajmowanych jednocześnie przez: mRNA i lncRNA, mRNA i sncRNA, lncRNA i sncRNA oraz mRNA i lncRNA i sncRNA. Chociaż większość z tych cząsteczek może pełnić funkcję naturalnie występującego, antysensownego transkryptu (NAT), najliczniejszą klasę tych transkryptów stanowią lncRNA (>200nt), które łącznie występują w ponad 67% *loci*. Grupa niekodujących RNA obejmuje podklasę transkryptów pseudogenów (PG), które zajmują 19,39% *loci* przypisanych lncRNA i sncRNA. Wartości procentowe odpowiadają liczbie zajmowanych *loci* (nie nukleotydów w sekwencji genomu). Opracowanie przygotowano na podstawie danych literaturowych omówionych w treści artykułu. **C. Funkcjonalna klasyfikacja cząsteczek RNA.** Transkrypty podzielono na dwie funkcjonalne klasy cząsteczek: naturalnie występujących, sensownych RNA (NST) i antysensownych (NAT), takich jak lncRNA i pseudogeny (PG), które w większości nie kodują białek, jakkolwiek funkcję długiego NAT może pełnić również cząsteczka kodującego mRNA (C.1) tj. Rev-Erba, która jest zdolna do rozpoznania i związania innej sensownej nici RNA tj. TRα. Na schemacie zaznaczono również małe niekodujące RNA (sncRNA), obejmujące klasy: mikroRNA (miRNA), tRNA, rRNA, snRNA, snoRNA, siRNA, saRNA, piRNA, które mogą pełnić funkcję krótkich NAT. **D. Kontrola przepływu informacji genetycznej przez cząsteczki wykazujące aktywność transkryptów NAT.** Hybrydyzacja transkryptu NAT z docelową cząsteczką RNA, która ulega transkrypcji w tym samym *locus* (cis-NAT) lub odległych *loci* (trans-NAT) uruchamia szereg specyficznych dla NAT mechanizmów związanych z regulacją ekspresji docelowych genów, zwłaszcza na poziomie kontroli ich transkrypcji i translacji, co zostało omówione w treści artykułu. **E. Wiązanie transkryptów NAT przez mRNA.** Cis-NAT

Rev-Erb α kodowany przez gen *NR1D1* ulega transkrypcji z promotora (P2) w kierunku zbieżnym do transkrypcji położonego w tym samym *locus* (17q11.2) genu *THRA* (P1, transkrypcja dwukierunkowa), którego mRNA TR α 2 wykazuje pełną komplementarność w regionie 3' z antysensowną cząsteczką Rev-Erb α , co umożliwia hybrydyzację obydwu RNA i aktywuje specyficzną odpowiedź NAT. Na schemacie zaznaczono również parę komplementarnych transkryptów *cis*-NAT Rev-Erb β (*NR1D2*) i mRNA TR β (*THRB*), oraz transkrypty: mRNA WT-1 (WT1) i antysensownej cząsteczki WIT-1 (*WIT1*), które ulegają jednak dwukierunkowej transkrypcji rozbieżnej z jednego promotora (P1) i prezentują typowy układ genów związanych z genomowym piętnowaniem rodzicielskim. Po środku zaznaczono transkrypt *trans*-NAT SRA kodowany przez gen *SRA1*, który wiążąc się z transkryptami odległych *loci* może zmieniać ekspresję ich genów, jak również pełnić funkcję niebiałkowego koaktywatora wielu czynników transkrypcyjnych w tym TR α 1 i TR β 1.

FIGURE 1. Functional and structural classification of transcripts (RNAs).

A. Transcriptome size in the context of the human genome sequence. Although it is estimated that over 90% of the genome sequence can be transcribed, the cellular proteome is encoded by less than 2% of the genome. The biological functions of the most transcripts are unknown (*dark matter* of the transcriptome), however, it was calculated that at least 20% of cellular RNAs have the potential to form sense-antisense pairs. **B. Participation of various RNA classes in an annotated part (4.62 %) of the human transcriptome.** Percentage values of gene *loci* number occupied by 4 categories of transcripts are shown in right circle graph. **C. Functional classification of RNA molecules.** Transcripts were divided into two groups of RNAs: naturally occurring sense (NST) and antisense (NAT) transcripts including lncRNAs and pseudogenes (PGs) – a class of non-protein coding or non-functional protein coding RNAs. Nevertheless, the function of NATs may be also fulfilled by protein coding mRNAs transcribed from overlapping genes such as *THRA* and *NR1D1* encoding TR α 2 and Rev-Erb α proteins. **D. The flow of genetic information controlled by transcripts showing NAT's activity.** Complementarity of transcripts may be observed when their genes share the common sequence in the same *locus* (*cis*-NATs) or distant *loci* (*trans*-NATs) that may result from gene duplication during genome evolution. Furthermore, the long double-stranded RNAs (dsRNAs) can trigger specific responses at the level of transcription or translation. **E. NATs' binding to mRNAs.** *Cis*- and *trans*-NATs can interact with various mRNA regions forming long RNA pairs such as WT-1 / WIT-1 transcribed from the same bidirectional promoter, TR α / Rev-Erb α with overlapping 3'UTR regions or SRA affecting the expression and function of the thyroid hormone receptors (TRs).

II.1. PSEUDOGENY STANOWIĄ NATURALNE ŹRÓDŁO CZĄSTECZEK NAT

Naturalnym źródłem długich, antysensownych RNA są pseudogeny, czyli zmienione kopie genu powstałe podczas ekspansji rodziny genowej, wykazujące ok. 75%-80% homologii z funkcjonalnym genem. Pseudogeny powstają przez duplikację genu lub jego fragmentu, jak również przez retrotranspozycję obejmującą proces odwrotnej transkrypcji genu, po której następuje integracja z genomem w odwrotnym kierunku, czemu towarzyszą zwykle niewielkie zmiany w ich sekwencji [46]. Długie antysensowne transkrypty mogą znajdować się zarówno na nici plus (+) jak i minus (-) DNA pseudogenów, ale również w sekwencji intronów, eksonów, promotorów, enhancerów, sekwencji między genowych i regionów nieulegających translacji [38, 46]. Obecnie, liczbę pseudogenów szacuje się na ok. 20 tysięcy. Nie jest to jednak ich ostateczna ilość, która może przekroczyć liczbę genów kodujących białka [46]. W najnowszych badaniach wskazuje się na proces duplikacji genów, jako kluczowy mechanizm powstawania funkcjonalnych *trans*-NAT, co przynajmniej w części tłumaczyć może dlaczego w toku ewolucji te *bezużyteczne* geny (jak wcześniej określano pseudogeny) nie zostały wyeliminowane [51]. Funkcjonalne znaczenie NAT w regulacji ekspresji genów kodujących białka obrazować może doświadczenie, w którym wykazano, że mysz transkrypt pseudogenu *Mkrn1-ps1* (ang. *Makorin1-p1*) odpowiada za stabilizację mRNA homologicznego genu kodującego białko *Mkrn1* (ang. *Makorin1*) [27]. Zablockowanie transkrypcji tego pseudogenu powoduje destabilizację mRNA *Mkrn1* i zahamowanie syntezy białka, którego niedobór objawia się wielotorbielowością nerek i deformacją kości myszy [27].

II.2. DŁUGIE NIEKODUJĄCE RNA TWORZĄ NAJLICZNIEJSZĄ KLASĘ FUNKCJONALNEJ GRUPY TRANSKRYPTÓW NAT

Transkrypty pseudogenów stanowią naturalną podklasę cząsteczek długich niekodujących RNA (**lncRNA**, ang. *long non-coding RNA*), które obok długich kodujących antysensownych RNA (np. Rev-Erb α , Rev-Erb β , SRA) i krótkich niekodujących RNA (**sncRNA**, ang. *short non-coding RNA*) takich jak mikroRNA (miRNA), należą do funkcjonalnej grupy naturalnie występujących antysensownych transkryptów **NAT** (Ryc.1). Badane ostatnio mechanizmy kontroli transkrypcji wskazują na niedocenianą dotąd rolę długich niekodujących transkryptów, które oprócz wpływu na proces metylacji [56] i modyfikacji chromatyny [56], mogą również być wiązane przez czynniki transkrypcyjne i pełnić funkcję korepresorów lub koaktywatorów transkrypcji [59], blokować transkrypcję przez wiązanie się do komplementarnych sekwencji kodujących cząsteczek mRNA [46], jak również kierować procesem alternatywnego składowania eksonów [70]. Chociaż molekularny mechanizm ich działania nie został dotąd dobrze poznany [64], coraz więcej prac naukowych łączy zmiany ekspresji lncRNA z występowaniem chorób związanych z zaburzeniem kontroli transkrypcji i translacji białek [55, 30].

II.3. GENY CIS-NAT WSPÓŁDZIELĄ SEKWENCJE DNA Z KONTROLOWANYMI PRZEZ NIE GENAMI TEGO SAMEGO LOCUS

Funkcjonalny podział NAT związany jest z docelowym miejscem ich działania. Długie antysensowne transkrypty syntetyzowane w obrębie jednego *locus*, które wykazują przy tym komplementarność względem sensownej nici kodującego mRNA tego samego genu lub genów w obrębie jednego *locus*, nazywane są naturalnie występującymi antysensownymi transkryptami *cis* (*cis*-NAT) [75]. Przykładem tego typu interakcji są geny *WT1* i *WIT1*, ulegające transkrypcji w przeciwnych, rozbieżnych kierunkach z jednego promotora [19], jak również gen *NR1D1*, położony w jednym *locus*, tuż za genem *THRA*, który koduje dwie funkcjonalnie antagonistyczne izoformy receptora jądrowego trijodotyroniny - właściwy receptor TR α 1 oraz receptor TR α 2, którego wydłużony C-końcowy łańcuch aminokwasów uniemożliwia zależną od T3 aktywację transkrypcji docelowych genów [10, 52]. Chociaż *NR1D1* i *THRA* ulegają transkrypcji z dwóch różnych promotorów znajdujących się w jednym *locus*, transkrypcja odbywa się w przeciwnych, zbieżnych kierunkach, co powoduje, że końce 3' transkryptów Rev-Erb α (*NR1D1*) i TR α 2 (*THRA*), nakładają się częściowo i są względem siebie komplementarne [45], hamując tym samym powstawanie izoformy białkowej TR α 2 na korzyść TR α 1 [22, 54].

II.4. TRANS-NAT WYMAGAJĄ DŁUŻSZYCH ODCINKÓW DOPASOWANIA WZGLĘDEM DOCELOWYCH TRANSKRYPTÓW ODLEGŁYCH LOCI

Jeżeli długie NAT wykazują komplementarność i regulują aktywność odległych genów, położonych poza własnymi *loci*, takie cząsteczki klasyfikowane są w grupie antysensownych, naturalnie występujących transkryptów *trans* (*trans*-NAT) [46]. Skuteczna regulacja „na odległość” wymaga od cząsteczek *trans*-NAT dłuższych odcinków komplementarności względem regulowanych transkryptów w porównaniu do *cis*-NAT, choć najczęściej takie dopasowanie nie jest pełne (<100%) [46]. W przeciwieństwie do cząsteczek mikroRNA, zaliczanych do grupy krótkich *trans*-NAT i regulujących aktywność mRNA głównie na poziomie translacji, *trans*-NAT kontrolują przede wszystkim etap transkrypcji na różnych jego poziomach (Ryc.1). Jednym z takich długich niekodujących RNA jest omówiony powyżej aktywator RNA receptorów steroidowych SRA, który działając „na odległość” pełni funkcję niebiałkowego koaktywatora wielu czynników transkrypcyjnych [35, 48], w tym TR α i TR β [70].

II.5. NAT KONTROLUJĄ EKSPRESJĘ GENÓW NA WIELU POZIOMACH PRZEPŁYWU INFORMACJI GENETYCZNEJ

Aktywność długich, naturalnie występujących transkryptów NAT streścić można do funkcji komplementarnego oddziaływania (interferencji) z różnymi elementami regulatorowymi sekwencji DNA jak i RNA, które decydują o kierunku, wydajności i szybkości procesów replikacji, transkrypcji i translacji białek [38]. Funkcjonalny potencjał tych cząsteczek związany jest także z ukierunkowaną metylacją wysp CpG i epigenetyczną regulacją struktury chromatyny [16].

II.5.1. NAT WYCISZAJĄ TRANSKRYPCJĘ GENÓW

Mechanizm ten opisano szczegółowo dla procesu inaktywacji jednego z chromosomów X ssaków płci żeńskiej [63]. Zasadniczą rolę w tym procesie odgrywa układ długich niekodujących transkryptów typu *cis*-NAT: antysensownego Tsix i sensownego Xist, który odpowiada za modyfikację struktury chromatyny przez naprowadzanie i rekrutację kompleksów wyciszających transkrypcję, metylację histonu H3 oraz metylację wysp CpG promotorów genów chromosomu X_i przeznaczonego do wyciszenia. Zwiększona ekspresja transkryptu Tsix na aktywnym chromosomie X_a neutralizuje sensowny transkrypt Xist zarówno przez komplementarne wiązanie antysensownego Tsix jak również zależną od niego metylację promotora Xist, którego mechanizm nie został jednak dobrze wyjaśniony, jakkolwiek odpowiada on również za proces piętnowania genomowego [63]. W podobny sposób wyciszana jest transkrypcja pojedynczych genów, których ekspresja regulowana jest przez selektywne rozpoznawanie i wiązanie się cząsteczek NAT, najczęściej w regionach promotorów tych genów [17, 13].

II.5.2. NAT BIORĄ UDZIAŁ W METYLACJI DNA I REGULACJI AKTYWNOŚCI PROMOTORÓW GENÓW

W wielu przypadkach cząsteczki RNA są czynnikiem inicjującym ukierunkowane dziedziczenie epigenetyczne, prowadzące do wzmocnienia lub wyciszenia określonych sekwencji DNA [63]. Proces wyciszania transkrypcji odbywać się może przez zależną od długich niekodujących RNA (lncRNA) rekrutację białkowych kompleksów wyciszających transkrypcję. Transkrypty NAT takie jak lncRNA mogą kierować te kompleksy do odpowiednich regionów genomu lub sekwencji DNA tak w obrębie własnego *locus* (regulacja *cis*) jak i oddalonych *loci* (regulacja *trans*) [16]. Mechanizmy metylacji DNA kontrolują istotne procesy związane z rozwojem organizmów, w tym inaktywacją jednego z chromosomów X embrionów żeńskich, jak również z

ekspresją genów istotnych z punktu widzenia embriogenezy (np. WT1, IGF2) [56]. Zaburzenie tych procesów może prowadzić do chorób nowotworowych [66].

Klasycznym przykładem wyciszania ekspresji genów przez metylację DNA jest inaktywacja jednego z chromosomów X ssaków przez transkrypt Xist [63]. Zasadniczą rolę w tym procesie odgrywa rejon zwany XIC (ang. *X-inactivation center*), obejmujący gen *XIST*, który koduje długi, podlegający splicingowi i poliadenylacji RNA Xist (ang. *X inactive-specific transcript*) oraz położony na 3' jego końcu gen *TSIX*. RNA Tsix pełni funkcję długiego antysensownego transkryptu NAT, częściowo komplementarnego do RNA Xist, który w nieobecności Tsix hybrydyzuje z sekwencjami DNA przeznaczonymi do wyciszenia. Zależna od Xist modyfikacja struktury chromatyny polega m.in. na rekrutacji enzymatycznych kompleksów wyciszających transkrypcję, które przeprowadzają metylację lizyny w pozycji 9 i 27 histonu H3 a następnie metylację wysp CpG promotorów chromosomu X przeznaczonego do wyciszenia. Wyższa ekspresja transkryptu Tsix na aktywnym chromosomie X neutralizuje sensowny transkrypt Xist zarówno przez komplementarne wiązanie antysensownego Tsix jak również przez zależną od Tsix metylację promotora Xist, którego jednak mechanizm nie został dotąd wystarczająco opisany [63].

II.5.3. NAT KONTROLUJĄ SKŁADANIE ALTERNATYWNYCH WARIANTÓW mRNA

W najnowszych badaniach, podkreśla się niedocenianą dotąd rolę cząsteczek NAT zarówno w ukierunkowanym składaniu specyficznych wariantów mRNA [39], jak i regulacji biosyntezy białek [38]. Cząsteczki długich antysensownych RNA mogą oddziaływać z intronowymi i eksonowymi sekwencjami pre-mRNA [46], wpływając na obserwowaną często korelację między ekspresją poszczególnych wariantów mRNA a poziomem długich antysensownych transkryptów typu NAT [53, 50]. Zaburzenie tej regulacji może towarzyszyć rozwojowi wielu chorób, w tym neurodegeneracyjnych [55] i nowotworowych [30, 66]. Szacuje się, że ok. 15% ludzkich chorób genetycznych związanych jest z zaburzeniem różnicowego składania transkryptu pierwotnego [50], w którym istotną rolę odgrywać mogą długie antysensowne transkrypty NAT [30, 76]. Długie transkrypty NAT mogą oddziaływać z pre-mRNA i ukierunkowywać proces wycinania intronów oraz alternatywnego łączenia eksonów, który często rozpoczyna się w fazie elongacji transkrypcji [2]. Eksony tworzą ciąg sekwencji kodującej (CDS) oraz regulatorowe regiony nieulegające translacji (UTR, ang. *UnTranslated Regions*), położone terminalnie, na 5' końcu (5'UTR) oraz 3' końcu (3'UTR) cząsteczki mRNA [3, 18]. W wyniku różnicowego łączenia eksonów powstają alternatywne warianty mRNA, które mogą różnić się między sobą w zakresie sekwencji kodującej izoforny białkowe, lub też w regionach nieulegających translacji 5'UTR [42] i 3'UTR [3]. Zarówno alternatywne składanie eksonów jak i proces potranskrypcyjnej modyfikacji RNA mogą być regulowane przez NAT [45, 67], prowadząc do powstania odmiennych białek, modyfikowanych dodatkowo w procesie potranslacyjnej obróbki, co zwiększa bioróżnorodność makrocząsteczek różnych typów komórek [49].

II.5.4. NAT UCZESTNICZĄ W PROCESIE REDAGOWANIA RNA

Redagowanie RNA, nazywane także edycją RNA (ang. *RNA editing*) polega na modyfikacji informacji genetycznej przez enzymatyczne wprowadzanie drobnych zmian w sekwencji RNA, które prowadzić mogą do syntezy odmiennego białka lub degradacji zmodyfikowanego RNA [57]. Modyfikacje te obejmować mogą deaminację adenozyiny do inozyiny (A→I), deaminację cytozyny do urydyny (C→U), jak również małe insercje lub delecje, które zmieniają zwykle ramkę odczytu (ORF) [32]. Długie Antysensowne transkrypty, które w jądrze komórkowym rozpoznają komplementarne fragmenty sensownych RNA, pełnią kluczową rolę w konwersji adenozyiny do inozyiny, rozpoznawanej następnie przez aparat translacyjny w cytoplaźmie jako guanozyna [16]. Powstające w jądrze komórkowym dupleksy antysensownego RNA i sensownego mRNA aktywują deaminazę adenozyiny dwuniciowego RNA - ADAR lub ADAT (w przypadku edycji tRNA), która przeprowadza zmianę (A→I), znakując w ten sposób dupleksy RNA. Modyfikacje te mogą zmieniać sekwencję białka lub uruchamiać proces degradacji znakowanych inozyną dupleksów RNA. Zmodyfikowane w wysokim stopniu dwuniciowe RNA, akumulowane są w jądrze i degradowane przez nukleazy specyficznie rozpoznające inozynę [16]. Mechanizm ten wskazuje na kolejną funkcję NAT, związaną z regulacją transportu jądrowego i kontrolą poziomu specyficznych mRNA [16]. Akumulacja w jądrze komórkowym antysensownego RNA, a w konsekwencji również jego sensownego partnera, jest częstym zjawiskiem obserwowanym w warunkach stresu komórkowego, hipoksji i zwiększonej ilości nadtlenu wodoru [16].

Modelowym przykładem zależnej od NAT edycji RNA jest przeprowadzana przez deaminazę ADAR konwersja A→I w mRNA 4f-rnp, inicjowana przez związanie antysensownego transkryptu sas-10 muszki owocowej (*Drosophila melanogaster*) [16]. ADAR przeprowadza podobną edycję w mRNA podjednostki GluR-B receptora kwasu glutaminowego typu AMPA. Zmiana pojedynczego aminokwasu Gln (CAG) na Arg (CGG) w eksonie 11, obniża przepuszczalność wapnia przez kanały jonowe zawierające tę podjednostkę GluR-B [57]. Uważa się, że redagowanie przez ADAR odbywa się przed etapem alternatywnego składania eksonów, o czym

świadczyć może wykazany udział intronów pre-mRNA w wyznaczaniu specyficznych miejsc edycji tego RNA. Oprócz edycji eksonu 11 (zmiana Gln/Arg), redagowanie GluR-B obejmuje także intron 11, jak również ekson 13 (zmiana Arg/Gly) oraz intron 13. Eksony wraz z intronami tworzą wewnątrzcząsteczkowe pary dwuniciowego RNA ekson-intron (sens-antysens), stanowiącego substrat deaminazy ADAR [57]. Wykazano także, że redagowanie pre-mRNA GluR-B determinuje proces różnicowego składania tego transkryptu, jak również, że edycja regionów UTR wpływa na stabilność mRNA i wydajność jego translacji [57].

Inny typ modyfikacji C→U przeprowadza rodzina deaminaz APOBEC/AID. APOBEC-1 redaguje mRNA apolipoproteiny B czego wynikiem jest wprowadzenie dodatkowego kodonu stop i skrócenie białka 100kDa → 48kDa [32]. Deaminaza cytydyny indukowana aktywacją limfocytów B (AID) odpowiedzialna jest za tworzenie zmienności genów przeciwciał przez wprowadzanie mutacji do jednoliciowego DNA [32]. W linii komórek wątrobowych zakażonych wirusem HBV wykazano, że AID bierze udział w modyfikacji RNA, jak również modyfikacji DNA tego wirusa, gdzie stwierdzono zmiany C→U oraz G→A [40]. Chociaż mechanizm edycji G→A nie jest znany, sugeruje się że zmiana ta może być wynikiem modyfikacji prowadzonej przez białko z rodziny APOBEC/AID na antysensownej nici dwuniciowego RNA [20, 21]. Zmianę G→A zaobserwowano także w mRNA HnRNP-K. Wynikiem edycji RNA tej rybonukleoproteiny jest specyficzna dla raka jelita grubego modyfikacja G274A, prowadząca do zmiany Ala92Thr na końcu domeny KH1 białka hnRNP-K [31]. Innym przykładem antysensownego transkryptu, który w mitochondriach kinetoplastów bierze udział w edycji RNA jest gRNA (ang. *guide RNA*). Pełni on funkcję wzorca blokowych insercji i delecji urydyn, wprowadzanych do edytowanego RNA [32].

II.5.5. NAT MOGĄ TAKŻE WZMACNIAĆ EKSPRESJĘ GENÓW

Aktywator RNA receptorów steroidowych (SRA), kodowany jest przez gen *SRA1* (ang. *Steroid Receptor RNA Activator*) [35], jest klasycznym przykładem *trans*-NAT, który pełni funkcję koaktywatora wielu czynników transkrypcyjnych [48]. Pierwotnie SRA został zidentyfikowany jako poliadenylowany, ulegający alternatywnemu składaniu, niekodujący transkrypt, będący koaktywatorem hormonów steroidowych (progesteronu, estrogenów, androgenów i glukokortykoidów), co tłumaczy jego nazwę [35]. W późniejszym jednak czasie wykazano, że na matrycy tego samego genu powstaje alternatywny, kodujący wariant mRNA, którego białko (SRAP) [29] może działać antagonistycznie względem SRA, jako korepresor hormonów steroidowych, co obrazuje złożoność regulacji dwu-funkcyjnych genów takich jak *SRA1* [48]. Obecnie, w bazie *NCBI Gene* opublikowane są dwa kodujące (GB.ac.no.: NM_001035235.3, NM_001253764.1) i dwa niekodujące warianty transkrypcyjne (NR_045586.1, NR_045587.1), które różnią się alternatywnym miejscem składania eksonów w regionie 5' genu *SRA1*. Zaburzenie właściwego stosunku tych wariantów, wykazujących zwykle aktywność korepresorów (białka SRAP) i koaktywatorów (lncRNA SRA) transkrypcji, może zmieniać fenotyp komórek nowotworowych, w tym raka piersi z ekspresją jądrowych receptorów progesteronu i estrogenów [17]. Jak wykazano w 2004r, niekodujący transkrypt SRA może być wiązany przez receptory jądrowe trijodotyroniny (TRα1 i TRβ1) za pośrednictwem domeny wiążącej RNA zlokalizowanej w sąsiedztwie domeny wiążącej DNA (C) i mocno wzmacniać aktywność transkrypcyjną tych receptorów [70].

Innym przykładem pozytywnej roli NAT w ekspresji genów może być układ dwukierunkowej transkrypcji w obrębie genu *HIF1A*, kodującego podjednostkę α czynnika indukowanego w hipoksji (HIF-1α). Układ ten prowadzi do syntezy dwóch cząsteczek RNA: sensownego mRNA i długiego antysensownego transkryptu (NAT), komplementarnego względem regionu 3' HIF-1α [68]. Związanie mRNA kodującego białko HIF-1α przez transkrypt NAT powoduje zwiększenie stabilności mRNA HIF-1α, co tłumaczy się zablokowaniem destabilizującego motywu tego mRNA [68]. Ze względu na aktywację transkrypcji genów związanych procesami angiogenezy przez HIF-1, regulacja ekspresji tego czynnika transkrypcyjnego przez NAT może mieć szczególne znaczenie w badaniach nowotworowych. Podobny efekt wzmocnienia ekspresji przez zwiększenie stabilizacji mRNA obserwuje się w przypadku dwukierunkowej transkrypcji genu *BACE1*, kodującego β-sekretazę, chociaż w tym przypadku stabilizacja mRNA wynika z zablokowania przez NAT miejsca wiązania destabilizującego mikroRNA [68]. W badaniach regeneracji, z wykorzystaniem modelu psich płuc poddanych pneumonektomii wykazano, że jednoczesna ko-transkrypcja sensownego RNA sEPO-R i antysensownego asEPO-R prowadzi do zwiększenia ilości białka receptora erytropoetyny (EPO-R) i przyspieszonej regeneracji, obejmującej powiększenie, wzrost i rearanżację zachowanego płuca. Wyniki tych badań mogą stanowić przykład synergizmu działania pary transkryptów sens-antysens w podnoszeniu wydajności syntezy białka [74].

II.5.6. TRANSKRYPTY NAT REGULUJĄ SYNTĘZĘ BIAŁEK

Kontrola translacji jest jednym z końcowych mechanizmów regulacji ekspresji genów, który na poziomie proteomu decyduje o charakterystycznych cechach fenotypowych komórki [25]. Szybkość oraz wydajność syntezy niektórych białek może być bezpośrednio regulowana przez komplementarne transkrypty

NAT, których związanie zmienia strukturę przestrzenną mRNA wpływając na dostępność miejsc wiązania czynników inicjacji, wydłużania łańcucha polipeptydowego i terminacji translacji [38]. Ponieważ inicjacja translacji jest etapem limitującym szybkość syntezy białek [47], szczególnie istotną dla tego procesu może być komplementarne wiązanie cząsteczek NAT przez sekwencje 5'UTR mRNA. Alternatywne warianty 5'UTR nie mają wpływu na sekwencję białka (nie kodują odmiennych izoform białkowych), chociaż mogą decydować o trwałości mRNA, lokalizacji subkomórkowej jak również o końcowej wydajności syntezy białek [47]. W regionach UTR znajdują się regulatorowe elementy *cis* (ang. *cis-acting elements*), rozpoznawane przez czynniki *trans* (ang. *trans-acting elements*) takie jak transkrypty NAT, wchodzące w skład rybonukleoprotein ITAF (ang. *IRES trans-acting factor*) [37], które rozpoznają wewnętrzne miejsca wejścia rybosomów IRES (ang. *Internal Ribosome Entry Site*) [33], czy też krótkie, niekodujące mikroRNA (sncRNA), rozpoznające swoiste miejsca wiązania w sekwencji 3'UTR [28]. NAT mogą pośredniczyć w interakcji mRNA z rybosomem i czynnikami inicjacji translacji. Wchodząc w skład kompleksów ITAF, obejmujących m.in. heterogenną grupę jądrowych rybonukleoprotein (hnRNP), zwykle stymulują translację niezależną od czapeczki 5' (zależną od elementu IRES) [37]. Oddziaływanie mRNA-NAT (sens-antysens) prowadzi do rearanżacji przestrzennej struktury 5'UTR, co ułatwia wiązanie małej podjednostki rybosomu 40S z sekwencją IRES i dopasowanie jej do kodonu start, bez potrzeby skanowania całego 5'UTR [33].

II.5.7. NAT MOGĄ KONTROLOWAĆ TWORZENIE FUNKCJONALNYCH STRUKTUR DNA I RNA

Kontrola ekspresji niektórych genów realizowana jest przez wpływ na proces formowania struktur przestrzennych regionów regulatorowych DNA (promotorów, intronów) [2] i RNA (UTR) [4]. Najnowsze badania wskazują, że szczególną rolę w tym procesie mogą odgrywać czteroniciowe struktury kwasów nukleinowych nazywane G-kwadrupleksami (ang. *G-quadruplex*), które charakteryzuje zwykle obecność motywu czterech tetrad guanidynowych (GGG) rozdzielonych przez trzy pętłe 1-7 dowolnych nukleotydów $d(G_{3+N_{1-7}G_{3+N_{1-7}+G_{3+N_{1-7}+G_{3+N_{1-7}}}})$ [2, 5]. Tworzą one wielowarstwowe, równoległe lub antyrównoległe struktury stabilizowane jonami metali i wiązaniami wodorowymi. G-kwadrupleksy składają się z jednej lub z wielu cząsteczek kwasów nukleinowych, które formują dynamiczne struktury przestrzenne, zmieniające się w trakcie progresji cyklu komórkowego [2]. Obecność G-kwadrupleksów wykazano w sekwencji promotorowej wielu genów odgrywających istotną rolę w cyklu komórkowym i procesach nowotworowych, w tym: *MYC* [14], *HIF1A* [60], oraz innych [2].

Jak wykazała analiza bioinformatyczna genów *THRA* i *THRB* (Materiały uzupełniające), promotory tych genów zawierają sekwencje zgodne z sekwencją konsensusową G-kwadrupleksu: **GGGCCGGGGCGGGCCGGG** (*TRHB*) oraz **GGGGCCAGGGGGGTGCTG GGGTAATTAGGG** (*THRA*), co wskazuje że mogą one formować funkcjonalne struktury przestrzenne, jak zaprezentowano na schematycznym modelu G-kwadrupleksu (Ryc.2). W przypadku *THRB*, na tworzenie tej potencjalnej struktury może wpływać komplementarny względem sekwencji promotorowej tego genu, antysensowny transkrypt LOC644990, co jednak wymaga weryfikacji w badaniach bezpośrednich. Podobny model regulacji zaproponowano dla tworzącego struktury G-kwadrupleksu promotora ludzkiego genu *RFP2*, kontrolowanego przez niekodujący antysensowny transkrypt *RFP2OS* [61].

Obecność G-kwadrupleksu stwierdzono także w sekwencji 5'UTR transkryptów genu *THRA*, gdzie zidentyfikowano ciąg tetrad guanidynowych (**GGGCCTGGGTGGCAGGGGGT GGG**) odpowiedzialnych za 35% zahamowanie syntezy jądrowego receptora trijodotyroniny TR α [1]. Najnowsze badania wskazują, że obecne w regionie 5'UTR G-kwadrupleksy mogą pełnić zasadniczą rolę w przełączaniu alternatywnych mechanizmów inicjacji translacji [5]. Wykazano, że charakterystyczne dla kwadrupleksów wielowarstwowe struktury mogą z jednej strony hamować translację zależną od czapeczki 5', a z drugiej ułatwiać inicjację translacji niezależnej od czapeczki 5' (zależnej od elementów IRES) [5]. Kwadrupleksy RNA mogą również powstawać w ciągu powtórzeń trinukleotydowych STR (ang. *Short Tandem Repeats*) takich jak: CGG, AGG i UGG, które zależą jednak silnie od sąsiadujących sekwencji RNA jak i od warunków środowiska [4]. W zależności od położenia i konformacji, struktury kwadrupleksów mRNA mogą hamować inicjację translacji zależnej od czapeczki 5' lub wzmacniać syntezę białek inicjowaną w wewnętrznych miejscach wejścia rybosomów (IRES) [5]. Przestrzenną konformację 5'UTR modyfikują dodatkowo wymienione wcześniej rybonukleoproteiny ITAF [33, 37]. NAT mogą ułatwiać lub utrudniać proces formowania się bogatych w guaninę, funkcjonalnych struktur G-kwadrupleksów przez komplementarne oddziaływanie z sekwencją konsensusową samego kwadrupleksu lub jego otoczenia, zarówno w RNA [5, 51] jak i w nici kodującej lub matrycowej DNA [23, 60].

W komórkach eukariotycznych mogą tworzyć się także trójniciowe kompleksy RNA lub RNA-DNA [6]. Tripleksy mogą być rozpoznawane i regulowane przez białka takie jak rybonukleoproteiny HnRNP, białka IF, białka HMG, różne białka cyklu komórkowego, czynniki transkrypcyjne i białka naprawy DNA [6]. Typowym przykładem regulacji przez trójniciową strukturę DNA-DNA-RNA jest kontrola ekspresji genu *DHFR* przez długi niekodujący transkrypt, którego transkrypcja inicjowana jest z alternatywnego promotora, położonego powyżej

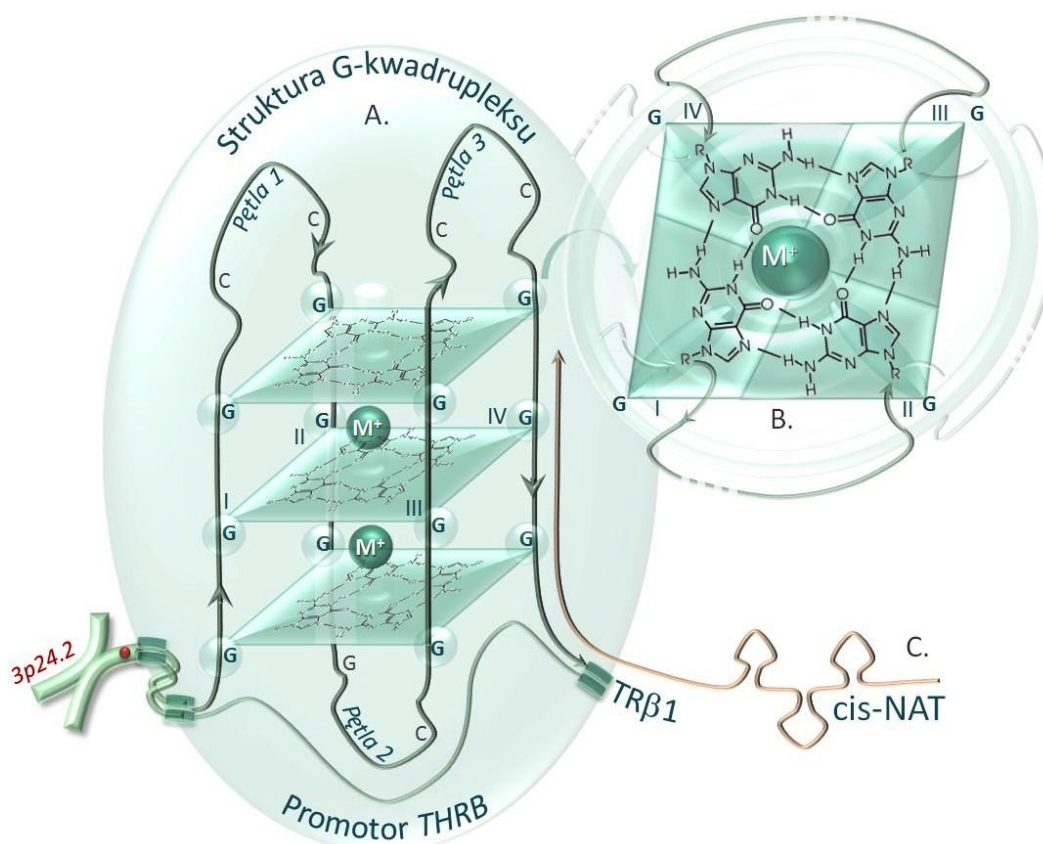
głównego promotora *DHFR* [6]. Niekodzący transkrypt oddziałuje z właściwym promotorem, z którym tworzy trójniciowy kompleks. Chociaż główny promotor odpowiada za 99% transkrypcji *DHFR*, utworzenie trójniciowego kompleksu powoduje dysocjację kompleksu pre-inicjacyjnego i hamuje transkrypcję *DHFR* [6].

W ostatnim czasie zwraca się uwagę na potencjalną rolę NAT w promowaniu zjawiska zmiany translacyjnej ramki odczytu, nazywanej również ślizganiem się rybosomów (ang. *ribosomal frameshifting*), które natrafiając na różnego rodzaju przeszkody natury strukturalnej, obecne w sekwencji poślizgowej (ang. *slippery sequence*) np. UUUAAAC, mogą przesuwać się z pominięciem tej przeszkody w inne miejsce mRNA, w tej samej lub innej ramce odczytu [15]. Mechanizm ten zmienia pierwotne znaczenie kodu genetycznego i prowadzi do zwiększenia bioróżnorodności białek [15].

Wszystkie wymienione powyżej etapy ekspresji genów mogą być kontrolowane przez długie, naturalnie występujące antysensowne transkrypty NAT [38], których komplementarność względem docelowych cząsteczek DNA i RNA może prowadzić do selektywnego wzmocnienia lub osłabienia ekspresji poszczególnych genów na różnych poziomach przepływu informacji genetycznej.

II.6. GENY lncRNA ZAWIERAJĄ WIĘCEJ SEKWENCJI STABILNYCH G-KWADRUPLEKSÓW

Badania *in silico* opublikowanych dotąd sekwencji długich niekodujących RNA (lncRNA), które stanowią najliczniejszą podklasę NAT, wykazała, że ok. 60% lncRNA może tworzyć G-kwadrupleksy, których pętle (rozdzielające tetrazy GGG) zawierają zwykle jeden lub dwa nukleotydy (najczęściej A lub T), co wskazuje na charakterystyczną dla nich cechę formowania bardziej stabilnych kwadrupleksów i wyróżnia spośród innych klas RNA [8, 39]. Chociaż obecna wiedza na temat znaczenia zwiększonej częstości występowania kwadrupleksów w sekwencjach lncRNA jest ograniczona, wiadomo że mogą one wiązać białkowe czynniki *trans*, które kontrolują proces transkrypcji i translacji mRNA [26]. Wykazano, że G-kwadrupleksy obecne w sekwencji aptamerów DNA i RNA umożliwiają swoiste rozpoznanie specyficznych białek i stabilizację ich wiązania, podobnie jak ma to miejsce w przypadku stosowanych powszechnie przeciwciał monoklonalnych [26].



RYCINA 2. Schemat oddziaływania NAT z modelową strukturą G-kwadrupleksu.

A. Model wewnętrznej struktury G-kwadrupleksu w konformacji antyrównoległej, opracowany na podstawie wytypowanej *in silico* sekwencji promotora genu *THRB* (GGGCCGGGGCGGGCCGGG), zgodnej z sekwencją konsensusową stabilnego G-kwadrupleksu. Na matrycowej nici DNA (w kolorze ciemno-zielonym) zaznaczono pozycje kolejnych nukleotydów guanylanowych w czterech powtórzeniach GGG, rozdzielonych przez trzy pętle dowolnych nukleotydów (tu:

CC, GC, CC), które łącznie tworzą trzy płaszczyzny oddziaływania czterech fragmentów tej samej nici. **B.** Powiększenie pojedynczej płaszczyzny stabilizowanej jonem metalu (M^+ , np. K^+) i wiązaniami wodorowymi guaniny. Rzymskimi cyframi (I, II, III, IV) oznaczono kolejne pozycje G w pojedynczej płaszczyźnie, zgodnie z kierunkiem przebiegu nici DNA. **C.** Schematycznie zaznaczony transkrypt cis-NAT (LOC644990), który mógłby zaburzać tworzenie struktury G-kwadrupleksu przez komplementarne oddziaływanie z nicią matrycową DNA sekwencji promotorowej *THRB*.

FIGURE 2. Scheme of interaction between NATs and G-quadruplex structures.

A. Intermolecular G-quadruplex structure in an antiparallel conformation. The model was designed on the basis of a G-quadruplex consensus sequence found *in silico* in *THRB* gene promoter (GGGCCGGGGCGGGCCGGG). **B. Magnification of a single G-quadruplex plain,** cooperatively stabilized by a metal ion (M^+) and hydrogen bonds of guanine (G). The G-quadruplex structure could be disturbed by a cis-NAT (e.g. LOC644990) that is complementary to this promoter region.

II.7. WYCISZANIE HAMUJĄCYCH NAT MOŻE WZMACNIAĆ EKSPRESJĘ GENÓW

W badaniach *in vivo* wykazano, że wyciszenie długich antysensownych transkryptów może prowadzić do zwiększenia wydajności transkrypcji i translacji genów [44]. Zastosowanie cząsteczek siRNA skierowanych przeciwko konserwatywnemu, antysensownemu transkryptowi BDNF-AS, który zwykle hamuje ekspresję czynnika BDNF (ang. Brain-Derived Neurotrophic Factor), powoduje wielokrotne zwiększenie wydajności transkrypcji i translacji tego czynnika, co podkreśla znaczenie długich, antysensownych RNA w pozytywnej regulacji ekspresji genów [44].

Znane są również cząsteczki dwuniciowego RNA (dsRNA), które wykazując homologię do sekwencji promotorowych, uruchamiają transkrypcję docelowych genów [24]. Cząsteczki te aktywują zjawisko RNAa (ang. RNA activation) i nazywane są **saRNA** (ang. *small activating RNA*) [24], przez analogię do przeciwnie działających siRNA (ang. *small interfering RNA*) i zjawiska RNAi (ang. *RNA interferencje*). Cząsteczki saRNA działają przeciwnie do epigenetycznego wyciszania transkrypcji, jak wykazano na przykładzie aktywacji genu *NANOG*, związanego z ewolucją komórek nowotworowych, którego aktywacja za pomocą saRNA, specyficznych dla sekwencji promotorowej tego genu, zwiększyła wydajność transkrypcji i hamowała indukowane przez kwas retinowy różnicowanie ludzkich embrionalnych komórek nowotworowych linii NCCIT [65]. Chociaż mechanizm działania saRNA jest słabo poznany, jedna z hipotez zakłada oddziaływanie tych cząsteczek z długimi antysensownymi transkryptami, wykrywanymi w rejonie promotorowym [71], co może wskazywać, że saRNA to siRNA, które wyciszają hamujące NATs.

Hipotezę tą potwierdzają wyniki badań, wskazujące, że długi antysensowny transkrypt wiążący się do sekwencji promotorowej receptora progesteronu (PR) pełni zasadniczą rolę w rozpoznaniu i wiązaniu zarówno hamujących jak i aktywujących agRNA (ang. *promoter-directed antigene RNAs*) [58]. Cząsteczki agRNA, jak zostały nazwane przez autorów pracy [58], to krótkie (19nt) dwuniciowe RNA (dupleksy RNA), komplementarne do sekwencji promotorów i wykazujące aktywność saRNA lub siRNA. Jak wykazano na linii T47D i MCF7 komórek raka piersi, rzeczywistym miejscem wiązania agRNA okazał się nie sam promotor, ale wiążące się do niego długie (do 70 000 nt) antysensowne transkrypty (NAT), komplementarne do promotora PR [58]. Poziom transkrypcji tych NAT (AT1, AT2-T47D, AT2-MCF7) był 10-1000 krotnie niższy od ekspresji RNA samego receptora PR, co wskazuje, że niewielki poziom NAT może kontrolować ekspresję docelowego genu [58]. Zaproponowany przez autorów mechanizm tej regulacji obejmuje kolejno: związanie agRNA z NAT, zależną od agRNA rekrutację białka AGO, przesunięcie lokalizacji rybonukleoproteiny hnRNP-K z DNA na RNA, włączenie polimerazy RNA II oraz białka heterochromatyny HP1y [58]. Wyniki tych badań rzucają światło na nieznaną dotąd mechanizm, w którym aktywność promotorów może być regulowana przez małe, dwuniciowe RNA takie jak saRNA/siRNA/agRNA, za pośrednictwem wiążących je długich NAT, komplementarnych do tych promotorów.

Z drugiej strony, proces transkrypcji może być wzmocniony bezpośrednio przez naturalnie występujące antysensowne transkrypty [43, 48]. Wykazano również, że zastosowanie antysensów może destabilizować hamujące elementy takie jak G-kwadrupleksy w promotorach genów *MYC* lub *HIF1A*, a w konsekwencji istotnie podnosić poziom transkrypcji [14, 60].

II.8. DŁUGIE TRANSKRYPTY NAT MOGĄ BRAĆ UDZIAŁ W ODPOWIEDZI NA STRES KOMÓRKOWY

Wysoki poziom białek C/EBP, regulujących m.in. ekspresję genów na szlaku Rev-Erba/TRa, stwierdza się także w odpowiedzi na stres komórkowy i nieprawidłowo sfałdowane białka (ang. *Unfolded Protein Response*, UPR), w wyniku aktywacji szlaków: IRE1/XBP1 [69], PERK/ATF4 [70] jak również kinazy PKR (ang. *double-stranded RNA-activated protein kinase*) [19]. PKR rozpoznaje dwuniciowe RNA (dsRNA) o długości >30 nukleotydów, które powstaje np. przez komplementarne wiązanie cząsteczek wirusowych transkryptów [19].

Zależna od dsRNA aktywacja PKR umożliwia dimeryzację i autofosforylację tego białka [19], co w dalszej kolejności prowadzi do fosforylacji podjednostki α czynnika inicjacji translacji eIF2, zahamowania inicjacji translacji zależnej od czapeczki 5' i uruchomienia interferonu (IFN) na szlaku I κ B/NF κ B/IFN [19]. Procesy te mogą zatrzymać cykl komórkowy, a przedłużający się stres siateczki śródplazmatycznej może aktywować również zależną od PKR apoptozę [19].

Podobny, 269-nukleotydowy odcinek całkowicie komplementarnego, dsRNA tworzą w regionie 3'UTR również dojrzałe transkrypty Rev-Erb α i TR α 2 [22]. Chociaż brak jest bezpośrednich badań potwierdzających lub wykluczających udział tych długich endogennych cząsteczek NAT w stresie siateczki śródplazmatycznej (ER) i odpowiedzi UPR, wiadomo, że podobne efekty biologiczne (zwykle niepożądane) wywołuje wprowadzenie do komórek dwuniciowych cząsteczek siRNA lub antysensów RNA, których długość przekracza 30 nukleotydów [19]. Wiadomo również, że specyficzny dla mysich neuronów dopaminergicznych, antysensowny lncRNA Uchl1, wzmacnia translację białka UCHL1 jedynie w pewnych warunkach stresu komórkowego [9]. Antysensowny transkrypt Uchl1 (*cis*-NAT) i mRNA genu Uchl1 (ang. *Ubiquitin carboxy-terminal hydrolase L1*) związanego z chorobami neurodegeneracyjnymi, kodowane są w tym samym *locus* przez geny, które ulegają dwukierunkowej transkrypcji, tak że ich dojrzałe transkrypty pokrywają się w regionie 5' na długości 72 nukleotydów. Antysensowny Uchl1, podobnie jak znaczna część niekodujących ncRNA akumulowana jest w obszarze jądra komórkowego. Niemniej, w warunkach stresu komórkowego wywołanego przez rapamycynę (inhibitora kompleksu mTORC1 i translacji zależnej od czapeczki 5') dochodzi do translokacji antysensownego Uchl1 z jądra do cytoplazmy, gdzie po utworzeniu pary sens-antysens z mRNA Uchl1 następuje aktywacja polisomów zwiększających wydajność translacji UCHL1 [9]. Obserwowany w stresie komórkowym efekt wzmocnienia translacji specyficznego białka wymagał obecności w regionie 5' *cis*-NAT odwróconego elementu SINEB2 i komplementarnej sekwencji, umożliwiającej utworzenie fragmentu dsRNA [9]. Podobne struktury długich dsRNA mogą powstawać przez hybrydyzację cząsteczek mRNA oraz ich antysensownych transkryptów [9, 46], biorąc udział w procesach fizjologicznej apoptozy towarzyszącej morfogenezie narządów i tkanek [13, 62].

W komórkach eukariotycznych mogą tworzyć się trójniciowe kompleksy RNA (tripleksy RNA), które biorą udział w po-transkrypcyjnej obróbce RNA [6]. Sugeruje się również, że tworzenie tripleksów może zabezpieczać komórki przed hamowaniem wzrostu oraz stresem komórkowym, jaki indukuje PKR w odpowiedzi na zakażenie dwuniciowym RNA wirusów [6]. Jak wykazano, utworzenie potrójnej helisy między dwuniciowym RNA i oligonukleotydem formującym tripleksy - TFO (ang. *Triplex-Forming Oligonucleotide*) istotnie zahamowało wiązanie PKR i jej aktywację, co może częściowo tłumaczyć dlaczego w warunkach fizjologicznych, pomimo komplementarności wielu komórkowych transkryptów, zwykle nie dochodzi do aktywacji PKR [6]. TFO wiążą się do nici bogatej w nukleotydy purynowe dwuniciowego RNA za pośrednictwem wiązań wodorowych Hoogsteen'a [6].

II.9. NAT REGULUJĄ CYKL OKOŁODOBOWY

Naturalne transkrypty antysensowne stanowią integralną część zegara biologicznego, sterującego rytmem okołodobowym, co wykazano po raz pierwszy na modelowym organizmie *Neurospora crassa* [34]. W dzikim szczepie tego grzyba, zegar biologiczny dostrajany jest do światła aktywującego transkrypcję antysensownego RNA frq (*frequency*), którego ilość w ciemności spada na korzyść sensownego transkryptu frq. W zmutowanym szczepie, w którym światło nie może indukować antysensownego frq, czas wewnętrznego zegara jest opóźniony, czemu towarzyszy zaburzenie ekspresji genów związanych z cyklem okołodobowym [34]. NAT frq uczestniczy w regulacji pętli sprzężenia zwrotnego między transkrypcją RNA i translacją białek bezpośrednio zaangażowanych w zależną od czasu regulację ekspresji genów [34].

W regulacji cyklu okołodobowego człowieka bierze udział opisany wcześniej gen *NR1D1*, którego ekspresja w cyklu okołodobowym zależy ściśle od innych, sprzężonych czynników transkrypcyjnych tj. BMAL1/CLOCK [41]. Gen *NR1D1* koduje antysensowny transkrypt Rev-Erb α (NAT), który reguluje stosunek izoform białkowych receptora jądrowego trijodotyroniny TR α 1/TR α 2 genu *THRA* [36]. Dobowe wahania transkrypcji Rev-Erb α mogą wpływać na cykliczność zmian poziomu ekspresji TR α 1/TR α 2 [72], a w konsekwencji komórkowej odpowiedzi na T3.

II. PODSUMOWANIE

W latach ubiegłych, uwaga naukowców skupiała się głównie na podstawowych badaniach liniowego przepływu informacji genetycznej DNA $\rightarrow$ RNA $\rightarrow$ białko. W erze post-genomowej pojawiła się możliwość sekwencjonowania całych genomów, analizy transkryptomów, proteomów i fenomów pojedynczych komórek,

co zmieniło rozumienie znaczenia regulacji ekspresji genów, które wymaga systemowego ujęcia wzajemnie powiązanych ze sobą układów biologicznych funkcjonalno-strukturalnej organizacji komórki. Poznanie zatem słabo opisanych dotąd mechanizmów regulacji przepływu informacji genetycznej przez długie, naturalnie występujące antysensowne transkrypty NAT wydaje się koniecznością. Ludzki genom wciąż kryje tajemnice związane z nieokreślonym nadal znaczeniem transkrypcji nieulegających translacji cząsteczek RNA, w tym niedocenianych dotąd pseudogenów, które w całości lub w części wykazują komplementarność względem genów w pełni funkcjonalnych. Jak wskazują ostatnie badania, co najmniej część z tych RNA może oddziaływać z docelowymi sekwencjami regulowanych genów na poziomie replikacji DNA, modyfikacji struktury chromatyny, transkrypcji i translacji białek. Zaburzenie ekspresji tych cząsteczek wykazano w patogenezie wielu chorób, w tym nowotworowych, co wskazuje na potrzebę ustalenia roli NAT w regulacji ekspresji poszczególnych genów.

LITERATURA

- [1] Beaudoin JD, Perreault JP. 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* 2010 Nov;38(20):7022-36. doi: 10.1093/nar/gkq557. Epub 2010 Jun 22. PubMed PMID: 20571090; PubMed Central PMCID: PMC2978341.
- [2] Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem.* 2013 Mar;5(3):182-6. doi: 10.1038/nchem.1548. Epub 2013 Jan 20. PubMed PMID: 23422559.
- [3] Boue S, Letunic I, Bork P. Alternative splicing and evolution. *Bioessays.* 2003 Nov;25(11):1031-4. Review. PubMed PMID: 14579243.
- [4] Broda M, Kierzek E, Gdaniec Z, Kulinski T, Kierzek R. Thermodynamic stability of RNA structures formed by CNG trinucleotide repeats. Implication for prediction of RNA structure. *Biochemistry.* 2005 Aug 16;44(32):10873-82. PubMed PMID: 16086590.
- [5] Bugaut A, Balasubramanian S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* 2012 Jun;40(11):4727-41. doi: 10.1093/nar/gks068. Epub 2012 Feb 20. Review. PubMed PMID: 22351747; PubMed Central PMCID: PMC3367173.
- [6] Buske FA, Mattick JS, Bailey TL. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol.* 2011 May-Jun;8(3):427-39. Epub 2011 May 1. Review. PubMed PMID: 21525785; PubMed Central PMCID: PMC3218511.
- [7] Butte AJ, Kohane IS. Creation and implications of a phenome-genome network. *Nat Biotechnol.* 2006 Jan;24(1):55-62. PubMed PMID: 16404398; PubMed Central PMCID: PMC2716377.
- [8] Caley DP, Pink RC, Trujillano D, Carter DR. Long noncoding RNAs, chromatin, and development. *ScientificWorldJournal.* 2010 Jan 8;10:90-102. Review. PubMed PMID: 20062956.
- [9] Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C, Forrest AR, Carninci P, Biffo S, Stupka E, Gustincich S. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature.* 2012 Nov 15;491(7424):454-7. doi: 10.1038/nature11508. Epub 2012 Oct 14. PubMed PMID: 23064229.
- [10] Cheng SY, Leonard JL, Davis PJ. Molecular aspects of thyroid hormone actions. *Endocr Rev.* 2010 Apr;31(2):139-70. Epub 2010 Jan 5. Review. PubMed PMID: 20051527; PubMed Central PMCID: PMC2852208.
- [11] Conley AB, Jordan IK. Epigenetic regulation of human cis-natural antisense transcripts. *Nucleic Acids Res.* 2012 Feb;40(4):1438-45. PubMed PMID: 22371288; PubMed Central PMCID: PMC3287164.
- [12] Cooper C, Guo J, Yan Y, Chooniedass-Kothari S, Hube F, Hamedani MK, Murphy LC, Myal Y, Leygue E. Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. *Nucleic Acids Res.* 2009 Jul;37(13):4518-31. Epub 2009 May 29. PubMed PMID: 19483093; PubMed Central PMCID: PMC2715257.
- [13] Dallosso AR, Hancock AL, Malik S, Salpekar A, King-Underwood L, Pritchard-Jones K, Peters J, Moorwood K, Ward A, Malik KT, Brown KW. Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer. *RNA.* 2007 Dec;13(12):2287-99. Epub 2007 Oct 16. PubMed PMID: 17940140; PubMed Central PMCID: PMC2080606.
- [14] De Armond R, Wood S, Sun D, Hurley LH, Ebbinghaus SW. Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry.* 2005 Dec 13;44(49):16341-50. PubMed PMID: 16331995.
- [15] Dinman JD. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA.* 2012 Jun 19. doi: 10.1002/wrna.1126. [Epub ahead of print] PubMed PMID: 22715123.

- [16] Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol.* 2009 Sep;10(9):637-43. doi: 10.1038/nrm2738. Epub 2009 Jul 29. Review. PubMed PMID: 19638999; PubMed Central PMCID: PMC2850559.
- [17] Faghihi MA, Wahlestedt C. Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol.* 2009 Sep;10(9):637-43. doi: 10.1038/nrm2738. Epub 2009 Jul 29. Review. PubMed PMID: 19638999; PubMed Central PMCID: PMC2850559.
- [18] Frankton S, Harvey CB, Gleason LM, Fadel A, Williams GR. Multiple messenger ribonucleic acid variants regulate cell-specific expression of human thyroid hormone receptor beta1. *Mol Endocrinol.* 2004 Jul;18(7):1631-42. Epub 2004 Apr 22. PubMed PMID: 15105435.
- [19] García MA, Gil J, Ventoso I, Guerra S, Domingo E, Rivas C, Esteban M. Impact of protein kinase PKR in cell biology: from antiviral to antiproliferative action. *Microbiol Mol Biol Rev.* 2006 Dec;70(4):1032-60. Review. PubMed PMID: 17158706; PubMed Central PMCID: PMC1698511.
- [20] Gu T, Buaas FW, Simons AK, Ackert-Bicknell CL, Braun RE, Hibbs MA. Canonical A-to-I and C-to-U RNA editing is enriched at 3'UTRs and microRNA target sites in multiple mouse tissues. *PLoS One.* 2012;7(3):e33720. doi: 10.1371/journal.pone.0033720. Epub 2012 Mar 20. PubMed PMID: 22448268; PubMed Central PMCID: PMC3308996.
- [21] Hamilton CE, Papavasiliou FN, Rosenberg BR. Diverse functions for DNA and RNA editing in the immune system. *RNA Biol.* 2010 Mar-Apr;7(2):220-8. Epub 2010 Mar 29. Review. PubMed PMID: 20220309.
- [22] Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH. Expression of the thyroid hormone receptor gene, *erbA* alpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. *Nucleic Acids Res.* 1997 Nov 1;25(21):4296-300. PubMed PMID: 9336460; PubMed Central PMCID: PMC147039.
- [23] Hastings ML, Wilson CM, Munroe SH. A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA.* 2001 Jun;7(6):859-74. PubMed PMID: 11421362; PubMed Central PMCID: PMC1370135.
- [24] Huang V, Qin Y, Wang J, Wang X, Place RF, Lin G, Lue TF, Li LC. RNAa is conserved in mammalian cells. *PLoS One.* 2010 Jan 22;5(1):e8848. PubMed PMID: 20107511; PubMed Central PMCID: PMC2809750.
- [25] Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* 2010; 11:113-27.
- [26] Jayaraj GG, Pandey S, Scaria V, Maiti S. Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol.* 2012 Jan;9(1):81-6. doi: 10.4161/rna.9.1.18047. Epub 2012 Jan 1. PubMed PMID: 22258148.
- [27] Kaneko S, Aki I, Tsuda K, Mekada K, Moriwaki K, Takahata N, Satta Y. Origin and evolution of processed pseudogenes that stabilize functional *Makorin1* mRNAs in mice, primates and other mammals. *Genetics.* 2006 Apr;172(4):2421-9. Epub 2006 Jan 16. PubMed PMID: 16415359; PubMed Central PMCID: PMC1456392.
- [28] Karaa ZS, Iacovoni JS, Bastide A, Lacazette E, Touriol C, Prats H. The VEGF IRESes are differentially susceptible to translation inhibition by miR-16. *RNA* 2009; 15:249-54.
- [29] Kawashima H, Takano H, Sugita S, Takahara Y, Sugimura K, Nakatani T. A novel steroid receptor co-activator protein (SRAP) as an alternative form of steroid receptor RNA-activator gene: expression in prostate cancer cells and enhancement of androgen receptor activity. *Biochem J.* 2003 Jan 1;369(Pt 1):163-71. PubMed PMID: 12350225; PubMed Central PMCID: PMC1223065.
- [30] Kitagawa M, Kotake Y, Ohhata T. Long Non-coding RNAs Involved in Cancer Development and Cell Fate Determination. *Curr Drug Targets.* 2012 Sep 7. [Epub ahead of print] PubMed PMID: 22974399.
- [31] Klimek-Tomczak K, Mikula M, Dzwonek A, Paziewska A, Karczmarski J, Hennig E, Bujnicki JM, Bragoszewski P, Denisenko O, Bomszyk K, Ostrowski J. Editing of hnRNP K protein mRNA in colorectal adenocarcinoma and surrounding mucosa. *Br J Cancer.* 2006 Feb 27;94(4):586-92. PubMed PMID: 16404425; PubMed Central PMCID: PMC2361188.
- [32] Knoop V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci.* 2011 Feb;68(4):567-86. doi: 10.1007/s00018-010-0538-9. Epub 2010 Oct 12. Review. PubMed PMID: 20938709.
- [33] Komar AA, Hatzoglou M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle.* 2011 Jan 15;10(2):229-40. Epub 2011 Jan 15. Review. PubMed PMID: 21220943; PubMed Central PMCID: PMC3048795.
- [34] Kramer C, Loros JJ, Dunlap JC, Crosthwaite SK. Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature.* 2003 Feb 27;421(6926):948-52. PubMed PMID: 12607002.

- [35] Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. A steroid receptor coactivator, SRA, functions as an RNA and is present in anSRC-1 complex. *Cell*. 1999 Apr 2;97(1):17-27. PubMed PMID: 10199399.
- [36] Lazar MA, Hodin RA, Darling DS, Chin WW. A novel member of the thyroid/steroidhormone receptor family is encoded by the opposite strand of the rat c-erbA alphanatranscriptional unit. *Mol Cell Biol*. 1989 Mar;9(3):1128-36. PubMed PMID: 2542765;PubMed Central PMCID: PMC362703.
- [37] Lewis SM, Holcik M. For IRES trans-acting factors, it is all about location. *Oncogene* 2008; 27:1033-5.
- [38] Li K, Ramchandran R. Natural antisense transcript: a concomitant engagement with protein-coding transcript. *Oncotarget*. 2010 Oct;1(6):447-52. Review. PubMed PMID: 21311100; PubMed Central PMCID: PMC3248111.
- [39] Li YY, Qin L, Guo ZM, Liu L, Xu H, Hao P, Su J, Shi Y, He WZ, Li YX. In silico discovery of human natural antisense transcripts. *BMC Bioinformatics*. 2006 Jan13;7:18. PubMed PMID: 16409644; PubMed Central PMCID: PMC1369008.
- [40] Liang G, Kitamura K, Wang Z, Liu G, Chowdhury S, Fu W, Koura M, Wakae K, Honjo T, Muramatsu M. RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. *Proc Natl Acad Sci U S A*. 2013 Feb 5;110(6):2246-51. doi: 10.1073/pnas.1221921110. Epub 2013 Jan 22. PubMed PMID: 23341589; PubMed Central PMCID: PMC3568302.
- [41] Liu C, Li S, Liu T, Borjigin J, Lin JD. Transcriptional coactivator PGC-1alpha integrates the mammalian clock and energy metabolism. *Nature*. 2007 May 24;447(7143):477-81. Epub 2007 May 2. PubMed PMID: 17476214.
- [42] Master A, Wójcicka A, Piekiełko-Witkowska A, Bogusławska J, Popławski P, Tański Z, Darras VM, Williams GR, Nauman A. Untranslated regions of Thyroid hormone receptor beta 1 mRNA are impaired in human clear cell renal cell carcinoma. *Biochim Biophys Acta*. 2010 Nov;1802(11):995-1005. Epub 2010 Aug 3. PubMed PMID: 20691260.
- [43] Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 2010 Jul8;466(7303):253-7. PubMed PMID: 20613842.
- [44] Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, van der Brug MP, Wahlestedt C. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat Biotechnol*. 2012 Mar 25;30(5):453-9. doi: 10.1038/nbt.2158. PubMed PMID: 22446693.
- [45] Munroe SH, Lazar MA. Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. *J Biol Chem*. 1991 Nov 25;266(33):22083-6. PubMed PMID: 1657988.
- [46] Muro EM, Andrade-Navarro MA. Pseudogenes as an alternative source of natural antisense transcripts. *BMC Evol Biol*. 2010 Nov 3;10:338. PubMed PMID: 21047404; PubMed Central PMCID: PMC2984423.
- [47] Myasnikov AG, Simonetti A, Marzi S, Klaholz BP. Structure-function insights into prokaryotic and eukaryotic translation initiation. *Curr Opin Struct Biol*. 2009 Jun;19(3):300-9. Epub 2009 Jun 1. Review. PubMed PMID: 19493673.
- [48] Novikova IV, Hennelly SP, Sanbonmatsu KY. Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res*. 2012 Feb 22. [Epub ahead of print] PubMed PMID: 22362738.
- [49] Ostrowski J.; Integrative genomics - a basic and essential tool for the development of molecular medicine. *Acta Pol Pharm*. 2008 Nov-Dec;65(6):621-4. PubMed PMID: 19172842.
- [50] Peacey E, Rodriguez L, Liu Y, Wolfe MS. Targeting a pre-mRNA structure with bipartite antisense molecules modulates tau alternative splicing. *Nucleic Acids Res*. 2012 Jul 25. [Epub ahead of print] PubMed PMID: 22844088.
- [51] Pertea M. The Human Transcriptome: An Unfinished Story. *Genes (Basel)*. 2012 Sep;3(3):344-360. PubMed PMID: 22916334; PubMed Central PMCID: PMC3422666..
- [52] Puzianowska-Kuznicka M, Pietrzak M, Turowska O, Nauman A. Thyroid hormones and their receptors in the regulation of cell proliferation. *Acta Biochim Pol*. 2006;53(4):641-50. Epub 2006 Nov 19. Review. PubMed PMID: 17115080.
- [53] Rinaldi C, Haddad F, Bodell PW, Qin AX, Jiang W, Baldwin KM. Intergenic bidirectional promoter and cooperative regulation of the Ilx and Ilb MHC genes in fast skeletal muscle. *Am J Physiol Regul Integr*

- Comp Physiol. 2008Jul;295(1):R208-18. Epub 2008 Apr 23. PubMed PMID: 18434443; PubMed CentralPMCID: PMC2494810.
- [54] Salato VK, Rediske NW, Zhang C, Hastings ML, Munroe SH. An exonic splicing enhancer within a bidirectional coding sequence regulates alternative splicing of an antisense mRNA. *RNA Biol.* 2010 Mar-Apr;7(2):179-90. Epub 2010 Mar 8. PubMed PMID: 20200494; PubMed Central PMCID: PMC2893273.
- [55] Salta E, De Strooper B. Non-coding RNAs with essential roles in neurodegenerative disorders. *Lancet Neurol.* 2012 Feb;11(2):189-200. doi: 10.1016/S1474-4422(11)70286-1. Review. PubMed PMID: 22265214.
- [56] Saxena A, Carninci P. Long non-coding RNA modifies chromatin: epigeneticsilencing by long non-coding RNAs. *Bioessays.* 2011 Nov;33(11):830-9. doi:10.1002/bies.201100084. Epub 2011 Sep 14. Review. PubMed PMID: 21915889; PubMedCentral PMCID: PMC3258546.
- [57] Schoft VK, Schopoff S, Jantsch MF. Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. *Nucleic Acids Res.* 2007;35(11):3723-32. Epub 2007 May 21. PubMed PMID: 17517775; PubMed Central PMCID: PMC1920255.
- [58] Schwartz JC, Younger ST, Nguyen NB, Hardy DB, Monia BP, Corey DR, Janowski BA. Antisense transcripts are targets for activating small RNAs. *Nat Struct Mol Biol.* 2008 Aug;15(8):842-8. doi: 10.1038/nsmb.1444. Epub 2008 Jul 6. PubMed PMID: 18604220; PubMed Central PMCID: PMC2574822.
- [59] Shamovsky I, Nudler E. Gene control by large noncoding RNAs. *Sci STKE.* 2006Oct 3;2006(355):pe40. Review. PubMed PMID: 17018852.
- [60] Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A.* 2002 Sep 3;99(18):11593-8. Epub 2002 Aug 23. PubMed PMID: 12195017; PubMed Central PMCID: PMC129314.
- [61] Skoblov M, Shakhbazov K, Oshchepkov D, Ivanov D, Guskova A, Ivanov D, Rubtsov P, Prasolov V, Yankovsky N, Baranova A. Human RFP2 gene promoter: unique structure and unusual strength. *Biochem Biophys Res Commun.* 2006 Apr 14;342(3):859-66. Epub 2006 Feb 17. PubMed PMID: 16499869; PubMed Central PMCID: PMC1994241.
- [62] Tata JR. Metamorphosis: an exquisite model for hormonal regulation of post-embryonic development. *Biochem Soc Symp.* 1996;62:123-36. Review. PubMed PMID: 8971345.
- [63] Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell.* 2010 Oct 29;143(3):390-403. doi: 10.1016/j.cell.2010.09.049. PubMed PMID: 21029862; PubMed Central PMCID: PMC2994261.
- [64] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell.* 2011Sep 16;43(6):904-14. Review. PubMed PMID: 21925379; PubMed Central PMCID: PMC3199020.
- [65] Wang X, Wang J, Huang V, Place RF, Li LC. Induction of NANOG expression by targeting promoter sequence with small activating RNA antagonizes retinoic acid-induced differentiation. *Biochem J.* 2012 May 1;443(3):821-8. PubMed PMID: 22339500; PubMed Central PMCID: PMC3327998.
- [66] Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol.* 2011 Jun;21(6):354-61. Epub 2011 May 6. Review. Erratum in: *Trends Cell Biol.* 2011 Oct;21(10):561. PubMed PMID: 21550244.
- [67] Werner A, Sayer JA. Naturally occurring antisense RNA: function and mechanisms of action. *Curr Opin Nephrol Hypertens.* 2009 Jul;18(4):343-9. doi: 10.1097/MNH.0b013e3283283cb982. Review. PubMed PMID: 19491676.
- [68] Werner A. Biological functions of natural antisense transcripts. *BMC Biol.* 2013 Apr 12;11:31. doi: 10.1186/1741-7007-11-31. PubMed PMID: 23577602; PubMed Central PMCID: PMC3626547.
- [69] Wouters BG, Koritzinsky M. Hypoxia signalling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer.* 2008 Nov;8(11):851-64. Epub 2008 Oct10. Review. PubMed PMID: 18846101.
- [70] Xu B, Koenig RJ. An RNA-binding domain in the thyroid hormone receptor enhances transcriptional activation. *J Biol Chem.* 2004 Aug 6;279(32):33051-6. Epub 2004 Jun 4. PubMed PMID: 15180993.
- [71] Yang K, Shen J, Xie YQ, Lin YW, Qin J, Mao QQ, Zheng XY, Xie LP. Promoter-targeted double-stranded small RNAs activate PAWR gene expression in human cancer cells. *Int J Biochem Cell Biol.* 2013 Jul;45(7):1338-46. doi: 10.1016/j.biocel.2013.03.022. Epub 2013 Apr 10. PubMed PMID: 23583662.
- [72] Yang X, Downes M, Yu RT, Bookout AL, He W, Straume M, Mangelsdorf DJ, Evans RM. Nuclear receptor expression links the circadian clock to metabolism. *Cell.* 2006 Aug 25;126(4):801-10. PubMed PMID: 16923398.
- [73] Zaporozhan V, Ponomarenko A. Mechanisms of geomagnetic field influence on gene expression using influenza as a model system: basics of physical epidemiology. *Int J Environ Res Public Health.* 2010

- Mar;7(3):938-65. doi: 10.3390/ijerph7030938. Epub 2010 Mar 10. PubMed PMID: 20617011; PubMed Central PMCID: PMC2872305.
- [74] Zhang Q, Zhang J, Moe OW, Hsia CC. Synergistic upregulation of erythropoietin receptor (EPO-R) expression by sense and antisense EPO-R transcripts in the canine lung. *Proc Natl Acad Sci U S A*. 2008 May 27;105(21):7612-7. doi: 10.1073/pnas.0802467105. Epub 2008 May 21. PubMed PMID: 18495932; PubMed Central PMCID: PMC2396693.
- [75] Zhang Y, Liu XS, Liu QR, Wei L. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res*. 2006 Jul 18;34(12):3465-75. Print 2006. PubMed PMID: 16849434; PubMed Central PMCID: PMC1524920.
- [76] Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. *J Mol Neurosci*. 2012 Sep 5. [Epub ahead of print] PubMed PMID: 22949041.

Redaktor prowadzący - Jerzy Kawiak

Otrzymano: 2013 r.

Przyjęto 2013 r.

prof. dr hab. n. med. Alicja Nauman, Centrum Medyczne Kształcenia Podyplomowego,
Zakład Biochemii i Biologii Molekularnej, ul. Marymoncka 99/103, 01-813 Warszawa,
e-mail: a.nauman@cmkp.edu.pl

TOM 41 2014 NR 1

MATERIAŁY UZUPEŁNIAJĄCE

I. ANALIZA SEKWENCJI *THRB* I *THRA*

W analizie sekwencji wykorzystano ogólnie dostępne oprogramowanie i bazy danych NCBI (ang. National Center for Biotechnology Information, U.S.): Map viewer, Gene, Genome, Nucleotide, BLAST, Epigenomics, EST, GEO Profiles, SRA, Pubmed, ACEDB, ACEWiew, jak również: Neural Network Promoter Prediction, RNAhybrid 2.2.

II. IDENTYFIKACJA ELEMENTÓW REGULATOROWYCH W REGIONIE 5' GENU *THRB*, KTÓRYCH FUNKCJE MOGĄ BYĆ ZABURZONE PRZEZ ANTYSSENSOWNE CZĄSTECZKI TYPU NAT

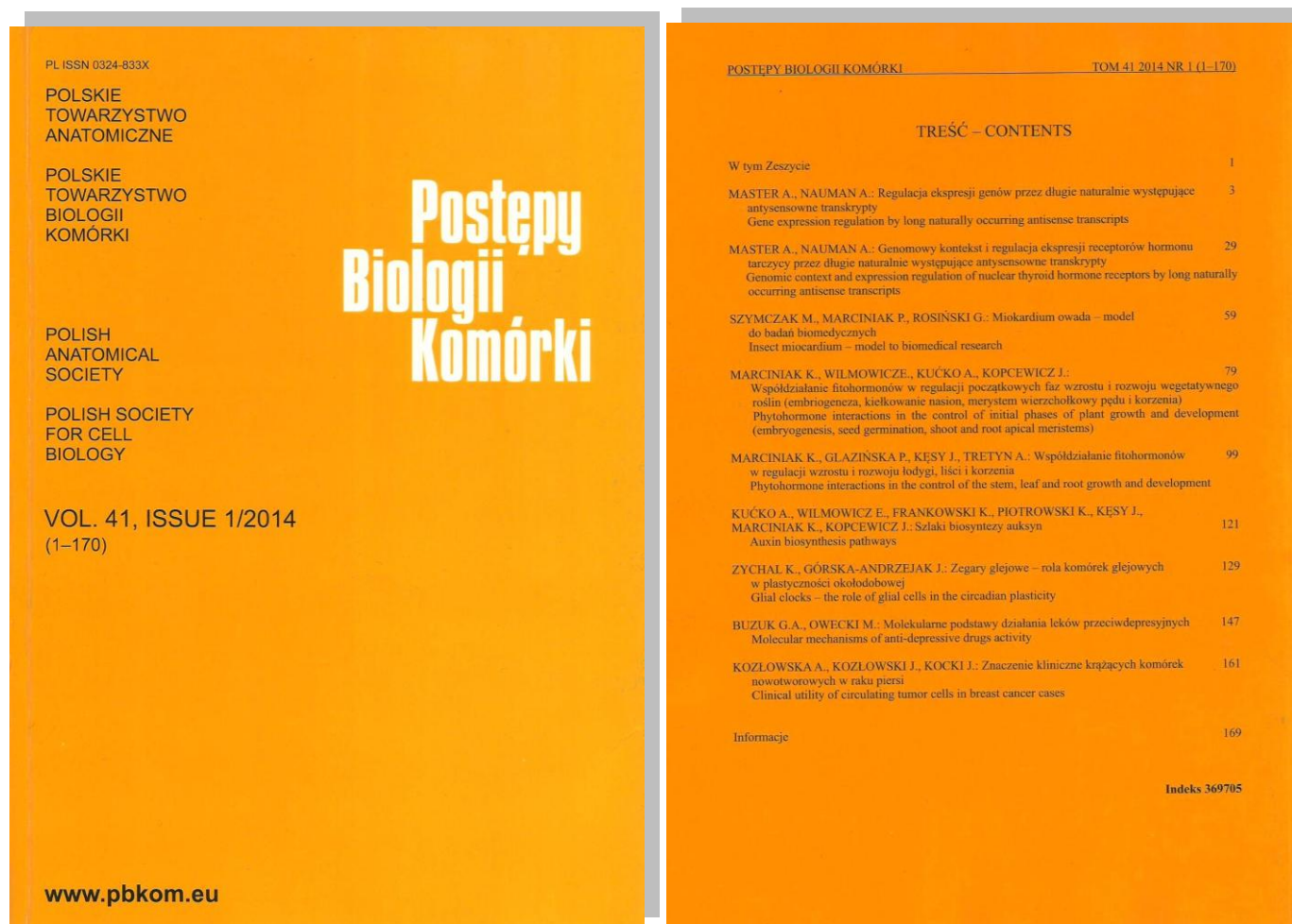
Zakres analizy: region analizy obejmował znaną sekwencję promotorową +/- 1000nt. Pozycję elementu określono względem miejsca inicjacji transkrypcji najdłuższego znanego eksonu 1a, zdefiniowanego w wariantie F TRβ1 (NCBI Nucleotides, Acc.no. : AY286470.1).

- **II.1.** Analiza sekwencji regionu 5' genu *THRB* wykazała także czteroniciowe struktury DNA nazywane **G-kwadrupleksami** (ang. *G-quadruplex*), zgodne z motywem czterech tetrad guanidynowych (GGG) rozdzielonych przez trzy pętle 1-7 dowolnych nukleotydów d(G_{3+N₁₋₇G_{3+N₁₋₇G_{3+N₁₋₇G₃}}):}
- W sekwencji bogatej w guaninę GGAGGGAGGAGCGCTGCGGCACGGGTCGGGCCGGTCCGGCCGGGCAA zidentyfikowano element spełniający kryteria G-kwadrupleksu **GGGTCGGGCCGGTCCGGCCGGG**, położony w sekwencji promotorowej w pozycji -881 od miejsca inicjacji wariantu F TRβ1 (NCBI acc.no. : AY286470.1)
- W sekwencji bogatej w guaninę: CCCGGGAAGCGGGCCGGGGCGGGCCGGGCACGCGGGGGACCCGGAGAGGCGGGGAC, znaleziono element **GGGCCGGGGCGGGCCGGG** spełniający kryteria stabilnego G-kwadrupleksu (Ryc. poniżej), położony w sekwencji promotorowej w pozycji -132 od miejsca inicjacji wariantu F TRβ1 (NCBI acc.no. : AY286470.1). Sekwencja ta tworzyć może również inne struktury G-kwadrupleksów tj: **GGGAAGCGGGCCGGGGCGGG** oraz **GGGGCGGGCCGGGCACGCGGG**.
- W sekwencji bogatej w guaninę: ACAGGGTGGGCGGGATGGTGACGGAGCGTCGAAGAACCCGGAGGGGTGCGGGCGGCTAAGCCGAGCGCGC GCGGGCGGGCAGGCGGGTGAGCGTGGGTGGTGGGGT, znaleziono element **GGGCGGGCAGGCGGGTGAGCGTGGG** spełniający kryteria G-kwadrupleksu, położony w sekwencji pierwszego intronu w pozycji +109 od ostatniego nukleotydu eksonu 1c wariantu A TRβ1 (NCBI acc.no. : AY286465.1). Sekwencja ta tworzyć może również inne struktury G-kwadrupleksów tj: **GGGCAGGCGGGTGAGCGTGGGTGGTGGG** oraz mniej stabilne **GGGTGGGCGGGATGGT**.

II.2. Analiza sekwencji regionu 5' genu *THRA* wykazała także czteroniciowe struktury DNA nazywane **G-kwadrupeksami** (ang. *G-quadruplex*), zgodne z motywem czterech tetrad guanidynowych (GGG) rozdzielonych przez trzy pętle 1-7 dowolnych nukleotydów $d(G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+})$:

- W sekwencji bogatej w guaninę **GGGTGGGCAGGGTGAGTGGGGTGGGAGGACAGATGAGGCAACACAGGG** zidentyfikowano element spełniający kryteria G-kwadrupeksu **GGGTGGGCAGGGTGAGTGGG** (oraz w drugim wariancie sekwencyjnym **GGGCAAGGGTGAGTGGGGTGGG**) położony w sekwencji promotorowej w pozycji -1975 (-1970 w drugim wariancie) od miejsca inicjacji transkrypcji TRα (zgodnie z sekwencją wzorcową NCBI acc.no. : NG_023345.1)
- W sekwencji bogatej w guaninę: W sekwencji bogatej w guaninę **GGGGAACCGGCGGGGCCAGGGGGGTGCTGGGGTAATTAGGG** zidentyfikowano element spełniający kryteria G-kwadrupeksu **GGGAACCGGCGGGGCCAGGGGGTGCTGGG** (oraz w drugim wariancie sekwencyjnym **GGGGCCAGGGGGTGCTGGGTAATTAGGG**, gdzie można wyróżnić inne nukleotydy G które mogą aktywnie brać udział w budowaniu płaszczyzny G-kwadrupeksu), położony w sekwencji promotorowej w pozycji -318 (-308 w drugim wariancie) od miejsca inicjacji transkrypcji TRα (zgodnie z sekwencją wzorcową NCBI acc.no. : NG_023345.1)

This version of the paper contains text and figures of the article that has been sent (2013) to the Journal (Postepy Biologii Komórki).



This article should be referenced as such:

Master A, Nauman A. Regulacja ekspresji genów przez długie naturalnie występujące antysensowne transkrypty [Gene expression regulation by long naturally occurring antisense transcripts].
Post. Biol. Kom. 2014;41(1):3-28. Review.

Full-text, original article: <http://www.pbkom.pl/> <http://pbkom.eu/en/>

REGULACJA EKSPRESJI GENÓW PRZEZ DŁUGIE NATURALNIE WYSTĘPUJĄCE ANTYSENSOWNE TRANSKRYPTY

GENE EXPRESSION REGULATION
BY LONG NATURALLY OCCURRING ANTISENSE TRANSCRIPTS

Adam MASTER, Alicja NAUMAN

Centrum Medyczne Kształcenia Podyplomowego,
Zakład Biochemii i Biologii Molekularnej, Warszawa

Streszczenie: W sekwencji ludzkiego genomu zapisana jest informacja na temat budowy i funkcji całego organizmu. O charakterystycznych cechach poszczególnych komórek i tkanek decydują mechanizmy regulacji ekspresji genów, w których istotną rolę odgrywają niedoceniane dotąd długie, naturalnie występujące, antysensowne transkrypty NAT (ang. *Naturally occurring Antisense Transcripts, NATs*). NAT wykazują pełną lub częściową komplementarność względem funkcjonalnych genów, kontrolując ich ekspresję na poziomie ukierunkowanej metylacji DNA, transkrypcji pre-mRNA, różnicowego składania transkryptu pierwotnego jak również redagowania RNA, transportu i zlokalizowanej translacji białek. W przeciwieństwie do grupy krótkich, niekodujących RNA (sncRNA) takich jak mikroRNA, długie niekodujące RNA (lncRNA), będąc najliczniejszą grupą transkryptów NAT, wykazują zwykle większe dopasowanie sekwencji, a tym samym specyficzność działania względem docelowych genów. Wiązanie tych cząsteczek do DNA i RNA może przesłaniać elementy regulatorowe lub wpływać na stabilność funkcjonalnych struktur takich jak G-kwadrupleksy i wewnętrzne miejsca wejścia rybosomów (elementy IRES). Nowe technologie sekwencjonowania całych genomów i przeszukiwania bibliotek cDNA umożliwiły ustalenie, że ok. 20% ludzkich i mysich genów współdzieli sekwencje z innymi genami, a ich transkrypty tworzyć mogą komplementarne pary sens-antysens. Chociaż fizjologiczne znaczenie większości tych par pozostaje niejasne, zidentyfikowano i opisano antysensowne transkrypty, które regulują ekspresję genów związanych z podstawowymi procesami takimi jak wzrost, proliferacja, różnicowanie, apoptoza, metabolizm jak również cykl okołodobowy. Zaburzenie ekspresji NAT stwierdzono w przebiegu wielu chorób, w tym nowotworowych, co wskazuje na potrzebę ustalenia ich roli na poszczególnych poziomach przepływu informacji genetycznej. Niniejsza praca dokonuje przeglądu aktualnej wiedzy w zakresie regulacji ekspresji genów przez długie RNA wykazujące aktywność NAT. Zdolność do tworzenia par sens-antysens i aktywacji mechanizmów kontroli ekspresji docelowych genów, stanowiła podstawę zaproponowanej definicji i strukturalno-funkcjonalnej charakterystyki tej niejednorodnej grupy transkryptów.