

Transkryptomika niekodujących RNA

Marek Żywicki

Division of Genomics and RNomics, Innsbruck Biocenter,
Innsbruck Medical University, Innsbruck, Austria

Noncoding RNA transcriptomics

Summary

The growing amount of knowledge about noncoding RNA molecules indicates their major role in regulation of cellular processes. For a long time there was no method which would allow for massive identification and analysis of such kind of transcripts. In recent years, the ongoing explosion of high throughput sequencing methods allowed a deep look into the transcriptome *via* the cDNA library analysis techniques, like RNA-seq. Since most of such protocols has been aimed for mRNA expression profiling, their adaptation for noncoding RNA research is necessary. It is mainly caused by differences in structure and properties of those two groups of transcripts. In this article, the main difficulties and challenges in investigation of noncoding RNAs with high throughput sequencing methods will be discussed.

Key words:

noncoding RNA, transcriptome, high-throughput sequencing, RNA-seq, RNomics.

Adres do korespondencji

Marek Żywicki,
Division of Genomics
and RNomics,
Innsbruck Biocenter,
Innsbruck Medical
University,
Austrian Academy of
Sciences,
Fritz-Pregl-Str. 3/II,
6020 Innsbruck;
e-mail:
Marek.Zywicki@i-med.ac.at

biotechnologia

3 (90) 191–201 2010

1. Wstęp

W ciągu ostatnich lat obserwować można znaczący wzrost zainteresowania cząsteczkami RNA nie posiadającymi otwartej ramki odczytu, nazywanymi niekodującymi RNA (ncRNA, ang. *noncoding* RNA). Fakt ten związany jest przede wszystkim z lawinowo rosnącą liczbą zidentyfikowanych cząsteczek tego typu. W 2001 r., gdy zsekwencjonowano genom człowieka, znane były pojedyncze cząsteczki niekodujących RNA, natomiast w 2004 r. liczba genów ncRNA u człowieka określana była na co najmniej równą liczbie genów kodujących białka (1). Dziś wydaje się, że

cząsteczki mRNA stanowiące matrycę dla biosyntezy białek stanowią tylko ułamek RNA obecnego w komórkach eukariotycznych, a transkrypcji ulegać może nawet 90% genomu człowieka (2).

Głównym wyzwaniem w badaniach nad niekodującymi RNA stało się określenie funkcji dla poszczególnych transkryptów. Obecnie znany jest mechanizm działania około 1% ncRNA. Dotyczy to głównie dobrze opisanych klas takich jak mikro RNA czy snoRNA (małe jąderkowe RNA, ang. *small nucleolar RNA*). W przypadku pozostałych setek tysięcy transkryptów znana jest jedynie sekwencja. Pilnie potrzebne są badania prowadzące do ustalenia warunków, w jakich ulegają one ekspresji oraz procesów, w które mogłyby być zaangażowane. Wybór pojedynczych cząsteczek do dokładnych badań funkcjonalnych jest utrudniony w świetle toczącej się dyskusji na temat funkcjonalności niekodujących RNA (3). Okazało się bowiem, że znaczna część ncRNA jest bardzo szybko degradowana (4-5). Obserwacja ta była podstawą hipotezy, według której sam fakt transkrypcji z nici DNA byłby istotny dla regulacji ekspresji położonych w pobliżu genów, natomiast powstający transkrypt byłby jedynie produktem ubocznym.

Istotnym wyzwaniem jest zatem identyfikacja cząsteczek istotnych dla procesów zachodzących w komórce. Skutecznymi narzędziami do tego typu analiz okazały się techniki opracowane pierwotnie do badań ekspresji mRNA – masowe sekwencjonowanie cDNA (RNA-seq) (6) oraz mikromacierze. Niestety specyfika cząsteczek ncRNA nie pozwala na analizę tego typu cząsteczek z zastosowaniem protokołów opracowanych z myślą o mRNA. Dopiero znaczące modyfikacje wspomnianych technik pozwalają na uzyskanie odpowiedzi na specyficzne pytania związane z funkcjonowaniem niekodujących RNA.

2. Sekwencjonowanie bibliotek cDNA

Jednym z najistotniejszych przełomów ostatniej dekady było opracowanie wysoko wydajnych technik sekwencjonowania DNA, które pozwalają na równoczesny odczyt sekwencji milionów cząsteczek. Technologia ta posiada kilka odmian, z których najważniejszymi i najczęściej stosowanymi są 454, Solexa oraz SOLID. Znaczące różnice metodologiczne pomiędzy tymi metodami mają swoje odbicie przede wszystkim w ilości sekwencji odczytywanych podczas jednego eksperymentu oraz ich długości (tab.). Zaraz po wprowadzeniu wymienionych technik główną ich wadą była ograniczona długość odczytywanych sekwencji, która wynosiła 30-50 nt. Obecnie można już otrzymać odczyty o długości ponad 500 nt, co jest wystarczające do sekwencjonowania *de novo* całych genomów eukariotycznych. Na podstawie tej technologii powstał projekt „1000 genomes” mający za zadanie poznanie sekwencji 1000 genomów człowieka do 2012 r.

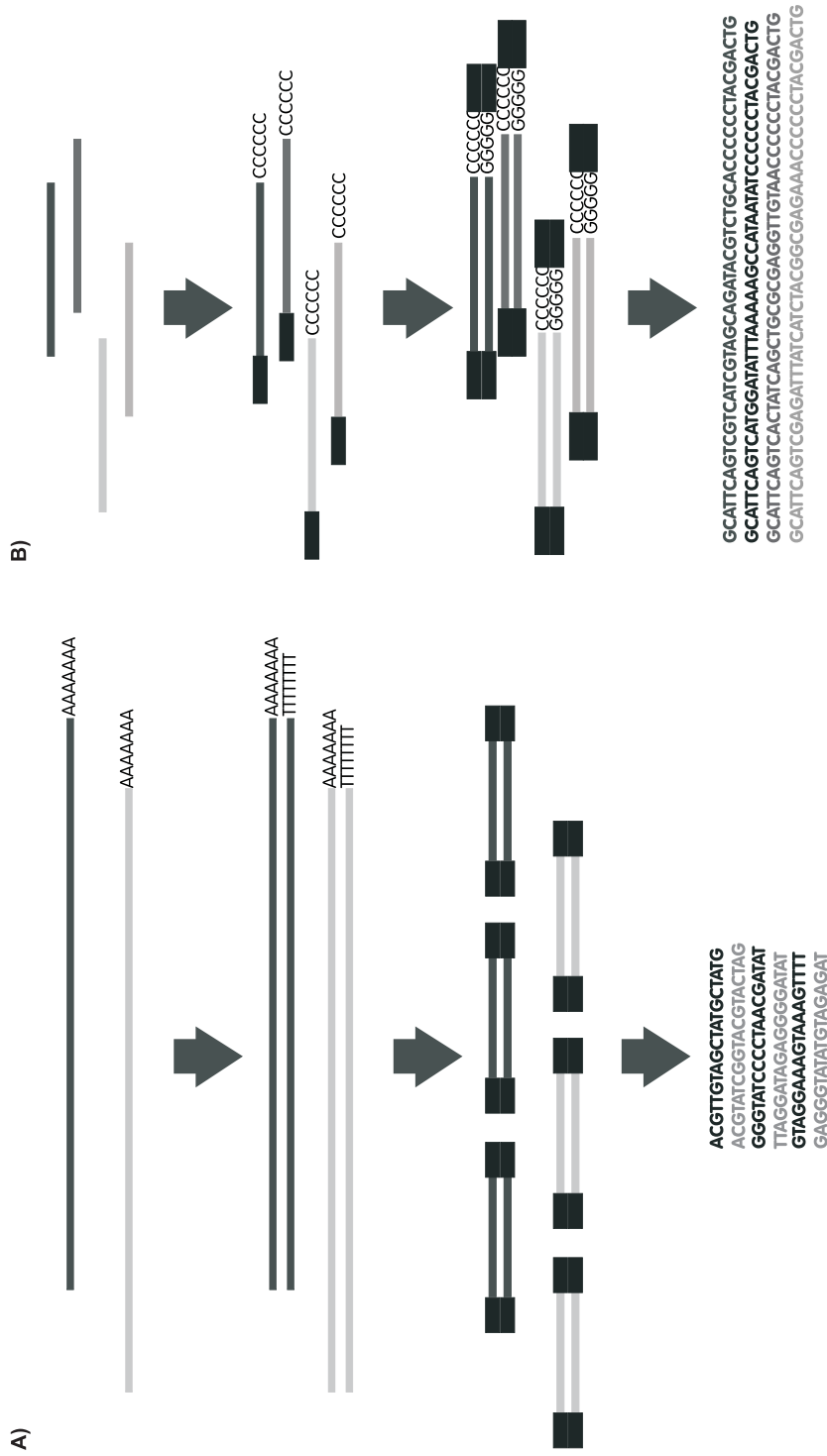
Tabela

Porównanie najpopularniejszych technik wysoko wydajnego sekwencjonowania DNA

	Roche/454	Illumina/Solexa	ABI/SOLID
maksymalna długość sekwencji (nt)	330-500	75-100	50
liczba zasad odczytywanych podczas jednego cyklu (mln nt)	0,45	18-35	30-50
koszt aparatu (USD)	500 000	540 000	595 000
koszt odczytu 1 mln zasad (USD)	60	2	2

Szybko zaczęto wykorzystywać nowe możliwości również do profilowania ekspresji mRNA za pomocą metody RNA-seq (6). Jest to możliwe dzięki przepisaniu mRNA na cDNA z użyciem starterów dla odwrotnej transkryptazy nakierowanych na sekwencje poli-A. Następnie przeprowadza się losową fragmentację, co pozwala na uzyskanie w końcowej puli sekwencji o kontrolowanej długości. Do tak przygotowanych fragmentów cDNA liguje się adaptory o znanej sekwencji potrzebne do przeprowadzenia amplifikacji cDNA za pomocą PCR. Adaptory te służą również do inicjacji reakcji wysoko wydajnego sekwencjonowania przez syntezę. Sam odczyt sekwencji może być przeprowadzony w dwojaki sposób (7). Jeśli fragmenty cDNA są odpowiednio krótkie, można odczytać ich całą sekwencję i użyć jej bezpośrednio do zmapowania rejonów genomowych ulegających ekspresji. Częściej jednak generuje się dłuższe fragmenty cDNA o podobnej długości, dla których odczytywane są jedynie końce 3' oraz 5'. Następnie można z dużą dokładnością zmapować na genomie oba odczyty z zachowaniem pomiędzy nimi odległości równej długości wygenerowanych fragmentów cDNA i przyjąć założenie, że rejon pomiędzy nimi również ulega ekspresji. Metoda ta pozwala na uzyskanie znacznie większego pokrycia genomu sekwencjami cDNA.

Opisana procedura posiada jednak kilka punktów, które znacząco wpływają na jakość uzyskiwanych wyników. Pierwszym jest przepisywanie cząsteczek mRNA na cDNA. Podczas tego procesu odwrotna transkryptaza posiada tendencję do przedwczesnej dysocjacji od matrycy mRNA powodując powstawanie skróconych produktów odpowiadających 3' części mRNA. Poddanie takiej puli cDNA fragmentacji prowadzi do podwyższenia liczby fragmentów pochodzących z 3' końca transkryptu. Dodatkowo równowaga ta zaburzona jest również poprzez zróżnicowaną długość mRNA. Z kolei przeprowadzana w następnym etapie ligacja adaptorów nie pozwala na zachowanie informacji o oryginalnej polaryzacji mRNA, co uniemożliwia późniejsze ustalenie nici DNA ulegającej transkrypcji. Ostatnim z etapów mających wpływ na jakość wyników jest stosowanie reakcji PCR w celu amplifikacji fragmentów cDNA. Skutkuje to zwiększeniem liczby krótkich fragmentów cDNA w stosunku do dłuższych, które są powielane z mniejszą wydajnością. Wszystkie te efekty mogą zostać uwzględnione podczas normalizacji i statystycznej analizy wyników (8), jed-



Rys. 1. Porównanie metod sekwencjonowania mRNA (RNA-seq) (A) oraz niekodujących RNA (RNomics) (B). A) Cząsteczki mRNA zostają przepisane na dwuniciowe cDNA, następnie poddawane są fragmentacji i przyłączone zostają adaptory. Tak przygotowany materiał poddawany jest sekwencjonowaniu, w wyniku którego uzyskuje się krótkie sekwencje odpowiadające fragmentom cDNA. B) Do niekodującego RNA nie zawierającego ogona poli-A dodaje się ogon poli-C co pozwala na ich selektywne przepisanie do dwuniciowych cDNA. Następnie przeprowadzana jest ligacja adaptorów i sekwencjonowanie.

nak ich obecność wskazuje jak pośrednią wartością jest poziom ekspresji odczytywany w ten sposób. Pomimo tych niedoskonałości walidacja wyników RNA-seq za pomocą reakcji PCR w czasie rzeczywistym (RT-PCR, Rang. *Real-Time* PCR) wykazała, że są one dokładniejsze niż dane uzyskane z mikromacierzy (9).

2.1. Specyfika bibliotek cDNA opartych na niekodujących RNA

Biblioteki cDNA zawierające niekodujące RNA znacząco różnią się od bibliotek opartych na mRNA. O ile długie, poliadenylowane ncRNA można wykryć i analizować razem z mRNA, o tyle ogromna pula małych ncRNA nie może być przepisana do cDNA poprzez odwrotną transkrypcję z użyciem startera poli-T. Ponieważ cząsteczki te są w większości produktami polimerazy RNA III, bądź też podlegają wycinaniu z dłuższych transkryptów polimerazy RNA II, brak jest w tych cząsteczkach ogona poli-A. Dlatego niezbędne jest dołączenie dodatkowych sekwencji na końcu 3' ncRNA, które służyć będą jako miejsca przyłączenia starterów dla odwrotnej transkryptazy. W tym celu stosuje się jedną z dwóch metod – ligację krótkich adaptorów RNA, bądź znacznie bardziej efektywne enzymatyczne wydłużanie 3' końca łańcuchem polinukleotydowym.

Kolejną różnicą jest znacznie większa różnorodność rozmiarów cząsteczek ncRNA. Może ona się wahać od 18-19 do kilku tysięcy nukleotydów. W związku z tym utrudnione staje się przeprowadzenie fragmentacji cDNA tak, żeby otrzymać homogeną pulę cDNA. W warunkach reakcji zoptymalizowanych dla dłuższych ncRNA, krótkie cząsteczki, które również uległyby przecięciu, stałyby się za małe dla celów późniejszej analizy bioinformatycznej. Alternatywnie, złagodzenie warunków doprowadziłoby do powstania zbyt dużych fragmentów długich ncRNA, które nie mogłyby zostać w pełni zsekwencjonowane.

Najpoważniejszym problemem wydaje się jednak fakt, że niepoliadenylowane cząsteczki ncRNA nie posiadają cech charakterystycznych, które pozwoliłyby na ich separację od rybosomalnych RNA. Ponieważ rRNA stanowi około 90-95% totalnego RNA, jego obecność w bibliotece cDNA skutecznie obniża możliwości identyfikacji innych cząsteczek o niższym stopniu ekspresji. Dlatego właśnie zamiast sekwencjonowania cDNA uzyskanego z totalnego RNA, stosuje się rozmaite techniki pozwalające na wyłowienie potencjalnie interesujących cząsteczek bądź zminimalizowanie udziału rRNA w bibliotece.

2.2. Metody usuwania rybosomalnych RNA

Najprostszą metodą pozwalającą na manipulację składem puli RNA wyizolowanego z komórek jest selekcja cząsteczek o wybranej długości. Najczęściej dokonuje się jej na podstawie rozdziału elektroforetycznego RNA w denaturującym żelu polia-

krylamidowym (10). Na podstawie obrazu uzyskanego po wybarwieniu, elucji poddaje się wyłącznie fragmenty żelu nie zawierające pełnej długości rRNA. Niewątpliwą zaletą jest możliwość równoczesnego wzbogacenia puli RNA w potencjalnie interesujące cząsteczki, np. mikro RNA czy małe niekodujące RNA, poprzez elucję RNA o określonej długości (11). Metodę tę charakteryzuje kilka poważnych ograniczeń. Najistotniejszym z nich jest to, że wraz z rRNA usuwane są również inne cząsteczki o zbliżonej wielkości. Dodatkowo, podczas elucji z żelu łatwiej uwalniane są krótsze RNA, co znacząco wpływa na skład późniejszej biblioteki cDNA. Co więcej, podczas licznych kroków eksperymentalnych, związanych z przygotowaniem próbki do rozdziału (denaturacja) oraz oczyszczeniem i zagęszczeniem po elucji, wzrasta ryzyko degradacji RNA, co w przypadku niektórych bibliotek cDNA stanowi poważną przeszkodę.

Ze względu na te niedogodności w przypadku konstrukcji bibliotek cDNA nastawionych na ustalenie poziomu ekspresji ncRNA stosuje się metody oparte na specyficznym „wylawianiu” rRNA. Metoda na to pozwalająca opiera się na zastosowaniu oligonukleotydów DNA lub LNA komplementarnych do rRNA. Sondy takie znakuje się biotyną i poddaje się hybrydyzacji do ekstraktu RNA. Następnie za pomocą ziaren pokrytych streptawidyną, specyficznie wiążącą biotynę, usuwa się powstałe heterodupleksy z roztworu. Metoda ta pozwala na usunięcie ponad 90% rybosomalnych RNA, jest wysoce specyficzna i nie zaburza równowagi ilościowej pomiędzy pozostałymi transkryptami. Jednak rezultaty jej zastosowania są silnie zależne od sekwencji sond użytych do hybrydyzacji. Muszą być one wysoce specyficzne i nakierowane na rejony rRNA, które są dostępne strukturalnie.

2.3. Konstrukcja specjalistycznych bibliotek cDNA opartych na niekodujących RNA

Obok konstruowania bibliotek cDNA opartych na totalnym RNA dużą popularnością cieszą się techniki pozwalające na wyselekcjonowanie i sekwencjonowanie wyłącznie wybranych grup cząsteczek ncRNA. Pomimo że podczas przygotowywania tego typu bibliotek cDNA dochodzi do zaburzenia równowagi pomiędzy cząsteczkami i nie można na ich podstawie wnioskować na temat poziomu ekspresji genów, są one niezwykle pomocne dla identyfikacji cząsteczek ncRNA biorących udział w konkretnych procesach komórkowych.

2.3.1. Zawężanie puli ncRNA na podstawie oddziaływań RNA-białko

Istnieje kilka metod pozwalających na identyfikację oddziaływań cząsteczek RNA z białkami. Podstawową stosowaną techniką jest koimmunoprecypitacja. Polega ona na izolacji kompleksów RNA-białko za pomocą immobilizowanych przeciwciał spe-

cyficznie rozpoznających białko. Podstawową trudnością związaną z tego typu eksperymentem jest odpowiednie dobranie warunków, tak aby jednocześnie uzyskać wysokie powinowactwo białka do przeciwciała i nie spowodować dysocjacji RNA. Głównie z tego względu można w ten sposób izolować wyłącznie bardzo stabilne kompleksy RNA-białko. W celu wzmocnienia interakcji stosuje się fotozszywanie kompleksów za pomocą promieniowania UV w obecności formaldehydu (12). Wytworzone w ten sposób wiązania mogą zostać później zredukowane za pomocą ciepła co prowadzi do uwolnienia RNA. Ogromną zaletą tej procedury jest fakt, że można ją stosować zarówno *in vitro* jak i *in vivo*. Pozwala ona również na stosunkowo dokładną identyfikację miejsca wiązania białka do RNA, jeśli w czasie izolacji kompleksów przeprowadzona zostanie hydroliza RNA z użyciem RNazy I. W takim przypadku w bibliotece uzyskamy wyłącznie fragmenty cDNA otaczające rejony RNA chronione przed degradacją przez białko.

Inną ciekawą metodą pozwalającą na izolację cząsteczek ncRNA oddziałujących z białkami jest izolacja kompleksów RNP (ang. *ribonucleoprotein*) (13). W odróżnieniu od koimmunoprecypitacji, gdzie identyfikowane są RNA oddziałujące z konkretnym białkiem, w tym przypadku uzyskana pula RNA zawiera zróżnicowaną populację cząsteczek RNA będących składnikami bardzo różnych kompleksów białkowych, takich jak RISC (ang. *RNA-Induced Silencing Complex*), spliceosom i inne. Selekcję RNP prowadzi się poprzez wirowanie ekstraktów komórkowych w gradiencie glicerolowym, a następnie izolację RNA migrującego z kompleksami o zadanej stałej sedymentacji. Można również pozyskać całkowite RNA wchodzące w skład różnych RNP i uzyskać w ten sposób globalny obraz kompozycji tego typu cząsteczek. Jednocześnie w prosty sposób można wyeliminować frakcje gradientu zawierające rybosomalne RNA.

2.3.2. Zawężanie puli ncRNA na podstawie oddziaływań z kwasami nukleinowymi

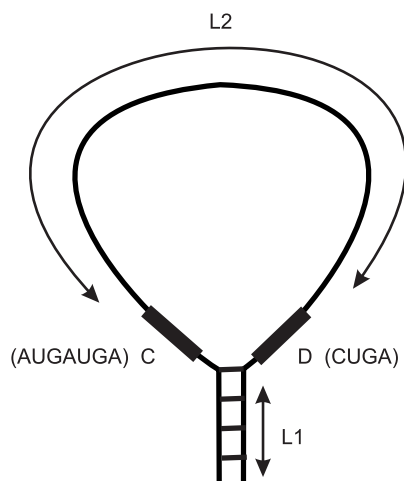
Drugą grupę technik konstrukcji specjalistycznych bibliotek cDNA stanowią metody oparte na specyficznych, zależnych od sekwencji interakcjach pomiędzy kwasami nukleinowymi. Metodą o największym potencjale, jak się wydaje, jest SHORT (ang. *Subtractive Hybridization of RNA Transcripts*) (14). Pozwala ona na znaczne wzbogacenie puli RNA w cząsteczki ulegające różnicowej ekspresji. Pierwszym krokiem jest immobilizacja cDNA uzyskanego w wyniku odwrotnej transkrypcji RNA z próbki A na złożu umieszczonym w kolumnie. Następnie ma miejsce hybrydyzacja RNA pochodzącego z próbki B. Cząsteczki, które znajdują się w obu próbkach zostają związane, natomiast te, które występują tylko w próbce B wypływają z kolumny. Co istotne, uzyskana pula RNA zawiera znikomą ilość rybosomalnych oraz transferowych RNA, co znacznie upraszcza kolejne etapy przygotowania biblioteki cDNA.

2.4. Analiza bioinformatyczna bibliotek cDNA opartych na ncRNA

Ogromne ilości danych uzyskiwane z użyciem wysoko wydajnego sekwencjonowania wymusiły opracowanie nowych metod pozwalających na ich analizę. Podstawowym wyzwaniem okazało się złożenie milionów krótkich fragmentów w sekwencje pełnej długości transkryptów. Podstawową metodą stosowaną do tej pory było łączenie sekwencji w dłuższe fragmenty na podstawie wzajemnego podobieństwa (15). W przypadku najnowszych technik sekwencjonowania metoda ta jest jednak nieefektywna zarówno z uwagi na liczbę, jak i ograniczoną długość otrzymywanych fragmentów. Z tego względu najpopularniejszą techniką jest porównanie otrzymanych odczytów do sekwencji referencyjnej, najczęściej genomowej (16). Pozwala to na stosunkowo szybkie zgrupowanie sekwencji pochodzących z tego samego transkryptu oraz opracowanie mapy aktywności transkrypcyjnej poszczególnych rejonów genomowych.

W przypadku badania poziomu ekspresji genów białkowych, zarówno sposób postępowania jak i większość metod statystycznych może być bezpośrednio zaczerpnięta z metodologii analizy danych pochodzących z mikromacierzy. Do analizy statystycznej używa się wyłącznie transkryptów zmapowanych w rejonach egzonów genów kodujących białka. Wszystkie sekwencje posiadające więcej niż jedno odpowiadające im miejsce w genomie są odrzucane. W takim przypadku nie ma pewności, z którego loci pochodzi transkrypt, co uniemożliwia jednoznaczne określenie poziomu ekspresji. Tylko nieliczne geny kodujące białka występują w genomach eukariotycznych w wielu kopiach, natomiast niekodujące RNA często kodowane są w wielu powtórzeniach. Doskonałymi przykładami są klastery ułożonych liniowo genów snoRNA oraz rozproszone powtórzenia genów tRNA i rRNA. Pomimo że nie jest możliwe określenie poziomu ekspresji osobno dla każdego z powtórzonych genów, to sama ich identyfikacja (szczególnie w specjalistycznych bibliotekach) może być wysoce pożądana. Z tego względu w przypadku analizy bibliotek cDNA opartych na ncRNA należy brać pod uwagę wszystkie transkrypty i wszystkie loci, do których mogą one zostać przypisane.

Kolejnym wyzwaniem jest funkcjonalna adnotacja zaobserwowanych transkryptów. Sekwencje genomowe organizmów modelowych posiadają dokładnie opisane pozycje genów kodujących białka. W większości przypadków zostały one opracowane na podstawie mapowania pełnej długości sekwencji mRNA znajdujących się w bazie danych sekwencji referencyjnych RefSeq (ang. *The Reference Sequence Database*) (17). Jednakże w przypadku genów ncRNA zazwyczaj adnotowane są wyłącznie geny znajdujące się w bazie Rfam (ang. *RNA Family Database*) (18), która nie zawiera pewnych grup ncRNA, takich jak długie mRNA-podobne ncRNA, czy słabo jeszcze poznane rodziny małych RNA (piRNA, rasiRNA itp.). Z tego względu dla pozostałej części ncRNA zlokalizowanych w rejonach międzygenowych w celu adnotacji przeprowadza się analizę porównawczą względem innych baz sekwencji, w szczególności sekwencji nukleotydowej bazy GeneBank (19).



Rys. 2. Deskryptor sekwencyjno-strukturalny opisujący charakterystyczne cechy cząsteczek snoRNA typu C/D. Są to terminalny trzon o długości L1, konserwatywne sekwencje C i D oraz dowolny znajdujący się pomiędzy tymi elementami fragment snoRNA o długości L2.

W następnym kroku wobec nieadnotowanych transkryptów przeprowadza się przewidywanie funkcji. W tym celu wykorzystuje się przede wszystkim metody oparte na wyszukiwaniu znanych rodzin ncRNA. Polegają one na zastosowaniu deskryptorów sekwencyjno-strukturalnych zawierających silnie konserwatywne motywy charakterystyczne dla danej klasy ncRNA. W ten sposób skutecznie można zidentyfikować nowe cząsteczki snoRNA typu H/ACA oraz C/D, SRP RNA (ang. *Signal Recognition Particle*) oraz tmRNA (ang. *transfer-messenger RNA*). Znacznie bardziej skomplikowana jest identyfikacja nowych mikro RNA. Wykorzystuje się tutaj fakt, że te 20-24-nukleotydowe cząsteczki wycinane są z dłuższych prekursorów, które posiadają strukturę spinki do włosów. Podobne motywy są szeroko rozprzestrzenione w genomach eukariotycznych. Dlatego w celu weryfikacji wykorzystuje się fakt, że metody wysoko wydajnego sekwencjonowania pozwalają na uzyskanie wielu kopii danej sekwencji, co umożliwia wnioskowanie o biogenezie cząsteczki z obserwowanego rozkładu odczytów (20). Ponieważ dojrzałe mikro RNA wycinane są ze ściśle zdefiniowanego fragmentu trzonu struktury spinki do włosów, należy spodziewać się, że obserwowane sekwencje będą w zdecydowanej większości pochodzić z jednego rejonu prekursora. W większości przypadków tylko jedna z nici trzonu ulega włączeniu do kompleksu RISC, co skutkuje szybszą degradacją cząsteczek pochodzących z przeciwnej nici. Proces ten również ma odzwierciedlenie w obserwowanym rozkładzie sekwencji.

3. Weryfikacja poziomu ekspresji ncRNA

Ze względu na liczne manipulacje towarzyszące procedurom związanym z tworzeniem bibliotek cDNA, w szczególności nastawionych na identyfikację konkretnych klas ncRNA, w większości przypadków nie jest możliwe bezpośrednie wnioskowanie o poziomie ekspresji ncRNA na podstawie wyników wysoko wydajnego sekwencjonowania. Często stosuje się inne techniki pozwalające na dokładne określenie poziomu zidentyfikowanych ncRNA w komórkach. W przypadku niewielkiej liczby cząsteczek najdokładniejszą metodą jest ilościowa reakcja PCR w czasie rzeczywistym. Pozwala ona na bardzo dokładne określenie ilości transkryptu poprzez monitorowanie przyrostu produktu reakcji w czasie.

Ponieważ technika ta jest pracochłonna i stosunkowo kosztowna w przypadku dużych ilości ncRNA, do analizy poziomu ekspresji konstruuje się mikromacierze (21). Analiza ncRNA z użyciem tej metody wymaga jednak uwzględnienia dwóch głównych problemów związanych ze specyfiką ncRNA. Pierwszy związany jest z projektowaniem sond oligonukleotydowych, które mają zostać umieszczone na powierzchni mikromacierzy. W celu uzyskania miarodajnych wyników wszystkie sondy muszą posiadać zbliżoną energię wiązania analizowanych transkryptów oraz długość zapewniającą specyficzną interakcję. Wiele małych ncRNA (np. mikro RNA) jest zbyt krótkich by umożliwić zaprojektowanie efektywnych sond. Rozwiązaniem jest zastosowanie modyfikacji LNA (ang. *Locked Nucleic Acids*), które zwiększają trwałość interakcji. Tego typu sondy są bardzo kosztowne, co często stanowi barierę dla większych projektów identyfikacji małych RNA w oparciu na technologii LNA.

Drugi problem związany jest z brakiem możliwości przepisania cząsteczek ncRNA na cDNA bez zaburzania równowagi pomiędzy cząsteczkami, gdyż większość z nich nie posiada ogona poli-A, który w przypadku mRNA stanowi miejsce przyłączenia startera dla odwrotnej transkryptazy. Hybrydyzacja materiału RNA do mikromacierzy pociąga za sobą pewne trudności. Jedną z nich jest fakt, że cząsteczki ncRNA posiadają często bardzo stabilne struktury drugorzędowe, co obniża efektywność reakcji. Problem ten można częściowo rozwiązać poprzez zaprojektowanie sond, które hybrydują w stosunkowo wysokich temperaturach. Pociąga to za sobą wydłużenie sond, co znacznie wpływa na wspomniane możliwości identyfikacji krótkich cząsteczek. Jednym z możliwych rozwiązań jest konstrukcja dwóch mikromacierzy – jednej zawierających długie sondy pozwalające na hybrydyzacje silnie ustrukturyzowanych ncRNA w wysokich temperaturach oraz drugiej, z krótszymi sondami, skierowanymi na małe ncRNA, które zazwyczaj nie posiadają tak silnych struktur drugorzędowych.

4. Perspektywy

Dalszy rozwój technik wysoko wydajnego sekwencjonowania nakierowany jest na wyeliminowanie większości z opisanych problemów utrudniających wnioskowanie na temat poziomu ekspresji genów. Niedawno opublikowane zostały pierwsze protokoły pozwalające na sekwencjonowanie pierwszej nici cDNA otrzymanej podczas odwrotnej transkrypcji RNA (22), oraz na bezpośredni odczyt sekwencji z RNA podczas przyłączania nukleotydów przez odwrotną transkryptazę (23). Techniki te pozwalają nie tylko na dokładne zliczenie liczby transkryptów, ale również pozwalają na zachowanie informacji na temat nici DNA, z której uległ on ekspresji. Na jednoznaczną ocenę ich skuteczności należy jednak poczekać do ukazania się pierwszych niezależnych publikacji wykorzystujących te technologie. Obserwowany w ostatnich latach rozwój dotychczasowych metod, nakierowanych na wydłużenie odczytywanych sekwencji oraz obniżenie kosztów pozwala przypuszczać, że w najbliższym czasie analiza transkryptomu z zastosowaniem opisanych technik stanie się dominującym sposobem badania różnorodności niekodujących RNA.

Literatura

1. Cawley S., et al., (2004), *Cell*, 116, 499-509.
2. Birney E., et al., (2007), *Nature*, 447, 799-816.
3. Louro R., Smirnova A.S., Verjovski-Almeida S., (2009), *Genomics*, 93, 291-298.
4. Neil H., et al., (2009), *Nature*, 457, 1038-1042.
5. Arigo J. T., Eyler D. E., Carroll K. L., Corden J. L., (2006), *Mol. Cell.*, 23, 841-851.
6. Mortazavi A., et al., (2008), *Nat. Methods*, 5, 621-628.
7. Wang Z., Gerstein M., Snyder M., (2009), *Nat. Rev. Genet.*, 10, 57-63.
8. Pepke S., Wold B., Mortazavi A., (2009), *Nat. Methods*, 6, S22-S32.
9. Marioni J. C., et al., (2008), *Genome Res.*, 18, 1509-1517.
10. Huttenhofer A., et al., (2001), *EMBO J.*, 20, 2943-2953.
11. Lu C., et al., (2005), *Science*, 309, 1567-1569.
12. Wang Z., et al., (2009), *Methods*, 48, 287-293.
13. Rederstorff M., et al., (2010), *Nucleic Acids Res.*, 38, e113.
14. Mrazek J., et al., (2007), *Nucleic Acids Res.*, 35, e73.
15. Huang X., Madan A., (1999), *Genome Res.*, 9, 868-877.
16. Flicek P., Birney E., (2009), *Nat. Methods*, 6, 8 S6-S12.
17. Pruitt K. D., Tatusova T., Maglott D. R., (2005), *Nucleic Acids Res.*, 33, D501-D504.
18. Gardner P. P., et al., (2009), *Nucleic Acids Res.*, 37, D136-D140.
19. Benson D. A., et al., (2010), *Nucleic Acids Res.*, 38, D46-D51.
20. Friedlander M. R., et al., (2008), *Nat. Biotechnol.*, 26, 407-415.
21. Aftab M. N., He H., Skogerbo G., Chen R., (2008), *BMC Genomics*, 9, 278.
22. Ozsolak F., et al., (2009), *Nature*, 461, 814-818.
23. Mamanova L., et al., (2010), *Nat. Methods*, 7, 130-132.