



## Analiza aktywności transkrypcyjnej genomu przy zastosowaniu mikromacierzy dachówkowych

Agnieszka Żmieńko<sup>1</sup>, Magdalena Guzowska-Nowowiejska<sup>2</sup>,  
Wojciech Płader<sup>2</sup>, Marek Figlerowicz<sup>1</sup>

<sup>1</sup> Centrum Doskonałości CENAT, Instytut Chemii Bioorganicznej,  
Polska Akademia Nauk, Poznań

<sup>2</sup> Katedra Genetyki Hodowli i Biotechnologii Roślin,  
Wydział Ogrodnictwa i Architektury Krajobrazu,  
Szkola Główna Gospodarstwa Wiejskiego, Warszawa

### Identification of transcriptionally active regions with genome tiling microarrays

#### Summary

Genome tiling microarrays, covering whole genomic sequence, have gained increasing popularity in transcript mapping studies. Functional analysis of model eucaryotic and procaryotic genomes proved their sensitivity and versatility in discovering actively transcribed regions of the genomes. Many novel proteins coding genes, miRNA coding genes, antisense transcripts and other non-protein coding regulatory RNAs, transcribed from introns, intergenic and centromeric regions have been identified this way. Their expression can often be linked to specific developmental stages, organs or stress response in plants and animals, giving further insight into processes which were considered to be already well characterized.

#### Adres do korespondencji

Agnieszka Żmieńko,  
Instytut Chemii  
Bioorganicznej,  
Polska Akademia Nauk,  
ul. Noskowskiego 12/14,  
61-704 Poznań;  
e-mail:  
akisiel@ibch.poznan.pl

#### Key words:

tiling arrays, antisense transcripts, genome annotation, transcript mapping.

### 1. Wprowadzenie

Transkryptomika to stosunkowo nowa dziedzina wiedzy zajmująca się kompleksową analizą funkcjonalną genomów. Jej zasadniczym celem jest określenie miejsca i czasu, w którym synte-

tyzowane są poszczególne cząsteczki RNA występujące w badanym organizmie. Można zatem powiedzieć, że głównym obiektem jej zainteresowań jest transkrypcjom, a zatem zbiór wszystkich RNA obecnych w danej komórce lub tkance. Niezwykle wysoki stopień złożoności tak zdefiniowanego układu eksperymentalnego sprawia, że podstawowym problemem, jaki napotykają badacze są ograniczenia technologiczne. Jednym z najciekawszych rozwiązań zmierzających do ich przezwyciężenia, jak się wydaje, są mikromacierze DNA (1-3). Stworzono je bowiem, jako narzędzie służące do jednoczesnego pomiaru poziomu ekspresji dużej ilości genów (nawet kilkudziesięciu tysięcy). W tym celu na podłożu o wymiarach szkiełka mikroskopowego nanosi się sondy oligonukleotydowe lub długie produkty PCR, specyficzne dla poszczególnych genów organizmu (komplementarne do wybranych fragmentów sekwencji/nici kodującej). W trakcie hybrydyzacji, do sond przyłączają się pochodne transkryptów z analizowanej próbki, uzyskane w procesie odwrotnej transkrypcji (cDNA) lub odwrotnej transkrypcji połączonej z transkrypcją *in vitro* (aRNA). Dzięki fluorescencyjnemu wyznakowaniu próbki związanej z sondą, można szacunkowo określić zawartość transkryptu danego genu w wyjściowej puli RNA. Uzyskana w ten sposób wartość uznawana jest za swoisty miernik aktywności genu (4-6). Ogromny potencjał mikromacierzy DNA wynika m.in. stąd, że dzięki zastosowaniu odpowiednich narzędzi bioinformatycznych i dostępnych baz danych, możliwa jest nie tylko analiza ekspresji pojedynczych genów, ale także poszukiwanie bardziej złożonych zależności m.in. powiązań metabolicznych, wspólnych mechanizmów regulacji itp. (7-13).

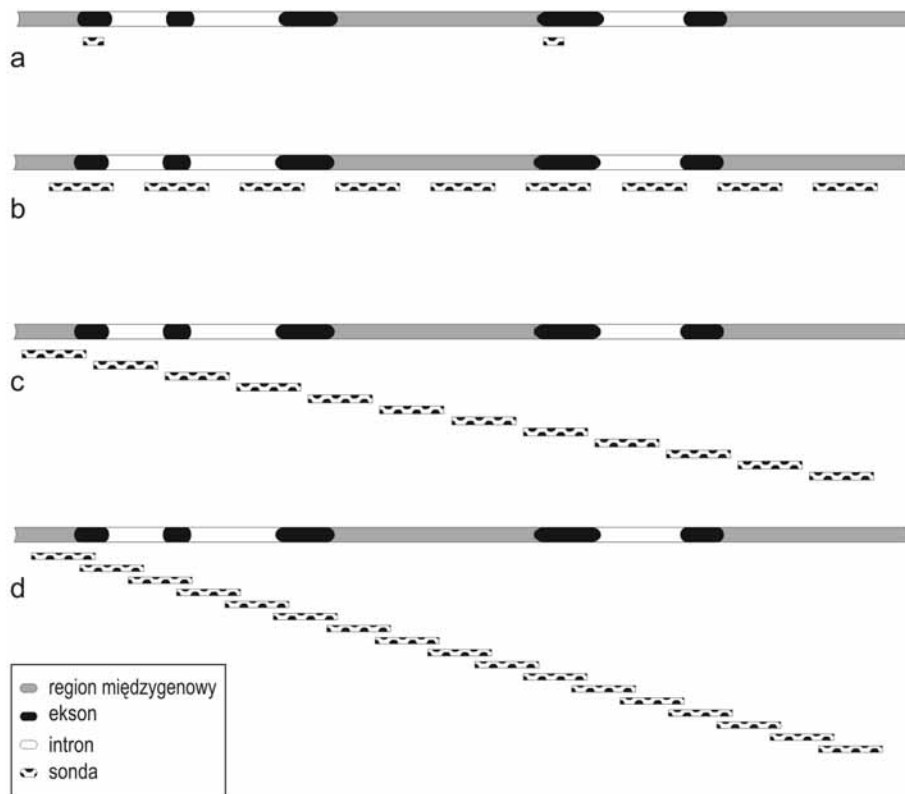
Głównym mankamentem stosowanych obecnie mikromacierzy ekspresyjnych jest brak pewności, że przy ich projektowaniu uwzględniono wszystkie transkrypcyjnie aktywne rejony genomu. Stwierdzenie to odnosi się także do najlepiej scharakteryzowanych modelowych genomów eukariotycznych (tj. genomu człowieka, muszki owocowej, nicienia *C. elegans*, rzodkiewnika, ryżu). Istnieje bowiem wiele dowodów świadczących, że aktywność transkrypcyjna nie jest ograniczona do genów kodujących białka (mRNA) oraz dobrze poznane niekodujące RNA (ncRNA), czyli rRNA, tRNA, małe jądrowe RNA (snRNA), małe jąderkowe RNA (snoRNA). Odkryto szereg nowych krótkich cząsteczek RNA, takich jak mikro RNA (miRNA), małe interferujące RNA (siRNA), antysensowe RNA, oraz dłuższych, dotąd nie scharakteryzowanych transkryptów, innych niż mRNA. Wielu z nich udało się przypisać istotne funkcje regulatorowe w tak różnorodnych procesach biologicznych jak: replikacja, transkrypcja, translacja, składanie, edycja i modyfikacja transkryptów RNA, czy translokacja białek i regulacja ich stabilności (14-20). Okazało się również, że ncRNA często powstają przez transkrypcję regionów genomu nie podejrzewanych dotąd o kodowanie i ekspresję jakiegokolwiek informacji. Dało to początek nowej dziedzinie badań genomów, przez niektórych zwanej RNomiką (21-23). Wraz z nią rozpoczęto wdrażanie metod pozwalających na masową identyfikację nie znanych dotąd produktów transkrypcji (24,25). W tym też celu opracowano specyficzny rodzaj mikromacierzy tzw. **mikromacierze równomiernie pokrywające genom** (ang. *genome*

*tiling arrays*), które proponujemy nazywać **mikromacierzami dachówkowymi**. Służą one m.in. do kompleksowej analizy aktywności transkrypcyjnej całych genomów. Uzyskane dzięki nim informacje zmuszają nas do radykalnej zmiany poglądów na temat roli, jaką odgrywać mogą tzw. niekodujące fragmenty genomu.

## 2. Mikromacierze równomiernie pokrywające genom

W odróżnieniu od mikromacierzy ekspresyjnych, zaprojektowanych wyłącznie do analizy znanych transkryptów, na mikromacierzach dachówkowych umieszcza się sondy komplementarne do położonych w mniej więcej regularnych odstępach fragmentów analizowanego genomu czy też chromosomu, bez względu na aktualny stan wiedzy o znaczeniu funkcjonalnym tego odcinka DNA (rys. 1) (26-28). Mikromacierze pokrywające genom mogą zawierać długie dwuniciowe DNA (produkty PCR) (29), jednak ich zastosowanie nie jest rozwiązaniem optymalnym przy poszukiwaniu nowych miejsc aktywnych transkrypcyjnie, nie pozwala bowiem na oddzielną analizę ekspresji nici sensowej i antysensowej. Obecnie mikromacierze pokrywające genom zawierają najczęściej sondy o długości kilkudziesięciu nukleotydów (27). Odległość między środkami sekwencji sąsiadujących sond mapujących genom, zwana rozdzielczością lub „krokiem” jest jednym z charakterystycznych parametrów danej mikromacierzy. Przykładowo, mikromacierz do analizy genomu ludzkiego GeneChip Human Tiling 2.0R, rozprowadzana przez firmę Affymetrix, składa się z siedmiu oddzielnych chipów, z których każdy zawiera około 6,5 miliona sond o długości 25 nt, pokrywających ludzki genom z rozdzielczością 35 pz. Jeżeli krok mikromacierzy jest równy lub mniejszy niż długość sondy, otrzymujemy mikromacierz, która pokrywa genom w sposób ciągły, a nawet gęściej, tworząc system „dachówkowy” (rys. 1c,d). Tak gęste rozłożenie sond w genomie zwiększa naturalnie ilość uzyskanych informacji i pozwala na udokładnienie wyników analiz (30).

Komercyjne mikromacierze dachówkowe są obecnie dostępne jedynie dla bardzo ograniczonej liczby organizmów modelowych, w przypadku chipów Affymetrixu dla: człowieka, małpy, muszki owocowej, *C. elegans* oraz rzodkiewnika. Przeprowadzenie takich analiz dla innych genomów wymaga samodzielnego zaprojektowania i przygotowania specyficznej mikromacierzy DNA. Jej gęstość zależy wówczas od wielkości i złożoności badanego genomu, celu eksperymentu oraz kosztów (syntezy odpowiedniej ilości sond), co może stanowić poważne ograniczenie całego przedsięwzięcia. Dlatego dość powszechnie produkuje się mikromacierze dachówkowe o dłuższych sondach oligonukleotydowych (np. 60-70 nt). Można również, uwzględniając aktualny stan wiedzy o rozkładzie genów w genomie, gęsto pokryć sondami obszary uznawane za niekodujące białek (introny, regiony międzygenowe, pseudogeny), a rzadziej – obszary kodujące (eksony) (31).



Rys. 1. Porównanie sposobu rozłożenia sond w a) mikromacierzach ekspresyjnych oraz b-d) mikromacierzach równomiernie pokrywających genom: b) o nie zachodzących sondach (małej rozdzielczości) c) o sąsiadujących sondach (rozdzielczości równej długości sond), co zapewnia całkowite pokrycie genomu oraz d) o sondach zachodzących (wysokiej rozdzielczości).

### 3. Nowe zastosowania – nowe problemy: kilka słów o analizie mikromacierzy dachówkowych

Podstawowe dane liczbowe uzyskiwane w wyniku hybrydyzacji wyznakowanego cDNA/aRNA dla mikromacierzy dachówkowej reprezentują intensywność fluorescencji poszczególnych sond, teoretycznie wskazując na poziom akumulacji komplementarnego transkryptu. Jednak wyłonienie z morza liczb takich danych, które rzeczywiście świadczą o ekspresji danego fragmentu genomu i uniknięcie tzw. wyników fałszywie pozytywnych, błędnie zakwalifikowanych jako znaczące, nie jest proste. Podstawowy problem w interpretacji uzyskanych danych polega na tym, że w organizmach eukariotycznych ekspresji ulega bardzo niewielka część genomu, a dodatkowo, poziom tej ekspresji może być niezwykle niski. Należy zatem być przygotowanym na to, że 1) przeważająca część sygnałów odczytanych z mikromacierzy będzie wykazywała intensywność na poziomie tła oraz 2) najbardziej intere-

sujące sondy (specyficzne dla dotąd nie zidentyfikowanych rejonów aktywnych transkrypcyjnie) nie będą wyróżniać się znacząco podwyższonym poziomem sygnału (26,32). Co za tym idzie, identyfikacja regionów aktywnych transkrypcyjnie jest niezwykle problematyczna i wymaga zastosowania wyrafinowanych sposobów analizy (33-38).

Podstawowym podejściem pozwalającym ominąć wymienione problemy jest zastosowanie tzw. ruchomego okna o określonej długości, znacznie przekraczającej długość pojedynczej sondy (38). Wielkość okna musi być odpowiednio dostosowana do rozmiaru sond i rozdzielczości mikromacierzy. Wówczas poszukuje się ciągu sąsiadujących sond o sygnałach przekraczających znacząco poziom fluorescencji tła i to jest wyznacznikiem transkrypcji w całym pokrytym sondami regionie. Na takim założeniu oparto szereg algorytmów, z powodzeniem stosowanych do analiz mikromacierzy dachówkowych, zarówno z krótkimi jak i długimi sondami. Przykładem tych pierwszych są ludzkie mikromacierze produkowane przez Nimblegen czy Affymetrix (gdzie obok sond o całkowitej komplementarności do analizowanego fragmentu sekwencji genomowej występują również ich odpowiedniki o niepełnej komplementarności, służące jako kontrole). Przykładem drugich są mikromacierze ludzkie Agilent o zachodzących sondach 60 nt (33,36,37). W niektórych z opracowanych metod analizy dodatkowo wykorzystuje się dane o aktualnej adnotacji funkcjonalnej genomu (37). Kilka przykładowych algorytmów przystosowanych do analiz mikromacierzy dachówkowych w przystępny sposób omówiono we wcześniejszej publikacji przeglądowej (32).

Jednym ze zjawisk obserwowanych podczas analiz mikromacierzy dachówkowych jest niejednorodny poziom sygnału sond uzyskiwany dla różnych eksonów tego samego genu. Co więcej, jak wykazali Royce i wsp. (32), analizując dane zebrane wcześniej z mikromacierzy dachówkowych człowieka, również intensywności sygnałów bezpośrednio sąsiadujących sond, leżących w obrębie tego samego eksonu różniły się nawet kilkukrotnie. Można wyobrazić sobie wiele przyczyn takiego stanu rzeczy: 1) istnienie alternatywnych odmian splicingowych tego samego transkryptu pierwotnego, 2) problemy niespecyficznego hybrydyzacji, 3) wpływ sekwencji sondy i jej temperatury topnienia na siłę wiązania (który można zmniejszyć na etapie projektowania, np. zezwalając na niewielkie przesunięcia sond w stosunku do zaplanowanych pozycji), czy wreszcie 4) obecność artefaktów, wynikająca z przyczyn technicznych. Niezależnie od źródeł zmienności, stanowi ona poważne utrudnienie w analizie statystycznej mikromacierzy dachówkowych. Jednak, jak wynika z analiz porównawczych eksperymentów przeprowadzonych dla różnych typów mikromacierzy ludzkiego chromosomu 22, to właśnie te najgęściej pokrywające genom dają najbardziej wiarygodne i zbieżne wyniki (30). Kolejnym istotnym elementem jest przeprowadzenie odpowiednio dużej liczby powtórzeń danego doświadczenia (30).

Jeśli badania wykraczają poza identyfikację regionów ulegających transkrypcji i są również punktem wyjścia do klasycznej analizy porównawczej ekspresji lub gdy należy uśrednić wyniki dla kilku powtórzeń technicznych eksperymentu (a tak jest

zazwyczaj), niezbędna jest wstępna normalizacja wyników uzyskanych dla poszczególnych kanałów/mikromacierzy (4,39,40). Jej celem jest zniwelowanie różnic zaistniałych podczas przygotowywania próby, hybrydyzacji mikromacierzy czy wynikających z ustawienia odmiennych parametrów skanowania. W tym miejscu pojawia się jednak kolejna trudność specyficzna dla mikromacierzy dachówkowych. W metodach normalizacji powszechnie stosowanych w analizie mikromacierzy ekspresyjnych, takich jak normalizacja totalna (podział przez medianę intensywności mikromacierzy) lub loess (ang. *locally weighted scatterplot smoothing*), uwzględniająca także wpływ intensywności fluorescencji sond na obserwowane różnice między kanałami (41) zakłada się, że co najmniej połowa analizowanych sond daje mierzalny sygnał. Tak jak opisano, w przypadku mikromacierzy równomiernie pokrywających genom nie ma to miejsca, co wymusza modyfikację lub opracowanie innych metod wstępnej normalizacji wyników (31,32,42).

Dodatkowym problemem w analizie tak złożonych mikromacierzy, gdzie znacząca ilość sond emituje sygnał na bardzo niskim poziomie, jest trudność w odróżnieniu istotnych informacji od artefaktów, które powstają z przyczyn technicznych bądź błędów w projektowaniu sond. Ciekawym przykładem takich problemów jest niespodziewanie wysoki poziom transkryptów antysensowych, obserwowany w wielu eksperymentach wykorzystujących mikromacierze dachówkowe (26). Wiadomo, że transkrypty antysensowe mogą odgrywać ważną rolę regulatorową w procesie transkrypcji odpowiadających im komplementarnych mRNA (14,15). Jednak w wielu eksperymentach z zastosowaniem mikromacierzy dachówkowych obserwowano charakterystyczną zbieżność poziomów akumulacji transkryptów antysensowych z poziomem komplementarnych mRNA oraz napotymano na problemy z uzyskaniem powtarzalnych wyników w technicznych powtórzeniach eksperymentów. Sugerowało to, że część transkryptów antysensowych może być generowana w sposób losowy, a ich źródło leży w błędach eksperymentalnych. W swoim niedawnym doniesieniu Perocchi i wsp. (43) wykazali, że istotnie, duża część takich transkryptów to artefakty, powstające na skutek syntezy drugiej nici cDNA podczas odwrotnej transkrypcji. Wiadomo bowiem, że odwrotne transkryptazy są polimerazami DNA zależnymi zarówno od RNA jak i DNA. Dodanie aktynomycyny D do mieszaniny transkrypcyjnej blokuje syntezę na matrycy DNA, a tym samym eliminuje artefakty, nie wpływając jednocześnie na poziom syntezy pierwszej nici (43). Obserwowane w tych warunkach sygnały wskazujące na obecność transkryptów antysensowych były o połowę mniej liczne, za to powtarzalne; co więcej, ich poziom nie był skorelowany z poziomem akumulacji nici sensowej. Na podstawie analiz RT PCR potwierdzono ogromną skuteczność tej prostej metody, wyższą niż stosowane w niektórych analizach wstępne filtrowanie danych (43,44).

Na inne źródło artefaktów w mikromacierzach dachówkowych wskazują Nelson i wsp. (45). Wykazali oni, że zastosowanie mikromacierzy o krótkich sondach (poniżej 50 nt) łącznie z metodami znakowania próby obejmującymi transkrypcję *in vitro* z promotora T7 oraz fragmentację aRNA powoduje, że w grupie najsilniejszych za-

notowanych sygnałów znacząco wzrasta udział niektórych sond o charakterystycznym składzie nukleotydowym. Autorzy powiązali te obserwacje z występowaniem pewnego motywu w sekwencji startera używanego do zapoczątkowania reakcji transkrypcji *in vitro* i udowodnili, że komplementarność sond do tego obszaru, nawet niepełna i na bardzo krótkim odcinku (6-9 nukleotydów) jest wystarczająca do hybrydyzacji z wyznakowanymi fragmentami aRNA pochodzącymi od startera, co wpływa tak drastycznie na intensywność uzyskanego sygnału.

Chociaż tego typu artefakty można w znacznym stopniu wyeliminować *post factum*, stosując odpowiednie filtry (44,45), to jednak bioinformatyczne metody również nie są doskonałe i dodatkowo mogą prowadzić do omyłkowego wykluczenia niektórych sygnałów (43).

Na podstawie tych przykładów pokazano jak wiele dodatkowych czynników może wpływać na końcowy wynik eksperymentu z zastosowaniem mikromacierzy dachówkowych. Tymczasem możliwość weryfikacji wszystkich uzyskanych rezultatów mniej czułymi i specyficznymi metodami może być ograniczona, nie tylko ze względu na koszty, lecz także z uwagi na bardzo niski poziom ekspresji badanych fragmentów genomu. Kolejnym problemem komplikującym właściwą identyfikację transkryptów, któremu nie są w stanie zaradzić nawet mikromacierze dachówkowe o gęsto rozłożonych sondach, jest analiza sekwencji wielokrotnie powtarzających się w genomie. Dokładna analiza aktywności transkrypcyjnej takich obszarów jest niezwykle trudna, z uwagi na niemożność odróżnienia loci, będących właściwym źródłem obserwowanego sygnału. Poza tym sumaryczny sygnał uzyskany dla powtarzających się regionów, może w znaczący sposób przewyższać sygnał pochodzący od sekwencji występujących w genomie tylko jeden raz. Przesunięcie skali detekcji w górę, niezbędne dla prawidłowego odczytania takich sygnałów przez skaner, spowoduje odcięcie wyników o najsłabszej fluorescencji. Dlatego regiony powtarzające się są często nieuwzględniane na etapie projektowania mikromacierzy dachówkowych (27).

## **4. Wykorzystanie mikromacierzy dachówkowych w analizie transkryptomów**

### **4.1. Mapowanie transkryptomu**

Dzięki swoim właściwościom, mikromacierze dachówkowe są najintensywniej wykorzystywane w badaniach transkryptomów dobrze poznanych organizmów modelowych. W 2000 r. wykorzystano je do określenia aktywności transkrypcyjnej całego genomu *E. coli* (31), skupiając się wszakże głównie na obszarach kodujących dotychczas scharakteryzowane geny. Wykryto w ten sposób ekspresję prawie wszystkich znanych genów, a także obecność licznych transkryptów z nici antysensowej. Dwa lata później Tjaden i wsp., używając tych samych mikromacierzy, prze-

badali transkryptom *E. coli* w trzynastu różnych warunkach hodowlanych, co zaowocowało zwiększeniem liczby znalezionych transkryptów o 25%: wśród 1102 nowych transkryptów zidentyfikowano elementy znanych operonów, nowe otwarte ramki odczytu, regiony 5'-UTR i 3'-UTR (ang. *untranslated region*, obszar nie ulegający translacji), a także małe RNA i inne, nie zidentyfikowane wcześniej transkrypty (46).

Do tej pory do organizmów dość dobrze scharakteryzowanych za pomocą mikromacierzy dachówkowych można zaliczyć, oprócz *E. coli*, także drożdże (44), *C. elegans* (47) muszkę owocową (48-50), człowieka (33,35,36,51,52), a także takie rośliny jak rzodkiewnik (53-55) i ryż (56-58). Cechą wspólną wszystkich uzyskanych tą drogą wyników, było stwierdzenie znacznie wyższej aktywności transkrypcyjnej genomów, niż wskazywały na to dotychczasowa wiedza i przewidywania bioinformatyczne. Między innymi zidentyfikowano szereg nowych: genów kodujących białka (33,49,59), eksonów będących częścią specyficznych i nie wykrytych dotąd form splicingowych (33,49,56-58) czy też regionów 3' i 5' UTR o niskim poziomie ekspresji (33). Scharakteryzowano też miejsca aktywne transkrypcyjnie, których ekspresja prowadzi do powstania nie znanych wcześniej RNA niekodujących białek (ncRNA). Najczęściej miejsca takie ulokowane są w obszarach międzygenowych i w intronach. Ekspresję obserwowano także w regionach centromerowych chromosomów, w pseudogenach oraz w sekwencjach transpozonowych, powszechnie występujących w genomach eukariotycznych (33,54-56,58).

Stwierdzono, że ncRNA są istotnym składnikiem transkryptomów u większości przebadanych dotąd organizmów (17-20). Za wyjątek można uznać nicienie, reprezentowane przez *C. elegans*, u którego ncRNA stanowi jedynie kilkanaście procent transkryptomu (47). Postulowany w przypadku nicieni niski poziom regulacji epigenetycznej może być powodem, dla którego ten prosty organizm potrzebuje do prawidłowego funkcjonowania aż 21 000 genów kodujących białka, co jest liczbą zbliżoną do liczby genów u znacznie bardziej skomplikowanych organizmów zwierzęcych (47,60). Uważa się, że u roślin i zwierząt wyższych wysoki stosunek zawartości sekwencji niekodujących do kodujących białka jest wyznacznikiem udziału ncRNA w regulacji ekspresji genów (17). Niektóre ncRNA nie są silnie zakonserwowane ewolucyjnie, co jednak nie świadczy o braku ich funkcjonalności, czego przykładem może być *Xist*, zaangażowany w inaktywację chromosomu X (61,62). Może okazać się, że pewne ncRNA są specyficzne dla pojedynczych gatunków albo dla gatunków blisko spokrewnionych. Na przykład, osiem nowo opisanych mysich ncRNA nie ma swoich odpowiedników w genomie człowieka, co pozwala sądzić, że są gatunkowo specyficzne (ewentualnie mogą one być specyficzne dla wszystkich gryzoni, ponieważ homologiczne sekwencje wykryto w genomie szczura) (63).

Do RNA niekodujących białek zalicza się także transkrypty nici antysensowej obszarów kodujących geny (33,53,56,58,60). Dopiero wyniki eksperymentów mikromacierzowych pokazały na jak szeroką skalę zjawisko to występuje w komórkach roślinnych i zwierzęcych (33,56,58), nawet uwzględniając fakt, że niektóre z dotychczasowych analiz mogły nie brać pod uwagę udziału artefaktów w tej puli transkryptów



(por. rozdz. 3). Uważa się, że transkrypty antysensowe mogą blokować polimerazę RNA lub pełnić rolę substratów w biogenezie siRNA. W konsekwencji mogą uczestniczyć w tworzeniu heterochromatyny i wyciszaniu odpowiedniej części chromosomu lub poszczególnych genów (28,56,64-66). Borsani i wsp. wykazali zależność pomiędzy zwiększeniem odporności rzodkiewnika na zasolenie, a obniżeniem aktywności genu *P5CDH*, którego transkrypty po połączeniu ze swoim antysensowym odpowiednikiem *SR05* stawały się substratem do produkcji specyficznych siRNA (66). W badaniach przeprowadzonych przez Yamadę i wsp. wykazano, że aż 30% genów rzodkiewnika kodowanych na nici sensowej ma swoje odpowiedniki na nici antysensowej (53). Natomiast ilość genów posiadających swój antysensowy odpowiednik w genomie myszy szacuje się na 70%, co, jak się zdaje, potwierdza powszechny udział transkryptów antysensowych w epigenetycznej regulacji genów (64).

#### 4.2. Badania funkcji nowych regionów aktywnych transkrypcyjnie

Odkrycie nowych jednostek aktywnych transkrypcyjnie w genomach skłania do dalszych poszukiwań, które pozwolą na przynajmniej częściowe rozpoznanie ich funkcji i profilu ekspresji. Dlatego też mikromacierze dachówkowe często stosuje się w porównawczej analizie funkcjonalnej genomów, w sposób analogiczny do klasycznych mikromacierzy ekspresyjnych (49,53,59,67). Doświadczenia takie są źródłem informacji o potencjalnej roli odkrytych transkryptów, dają jednocześnie szansę na wgląd w regulację wielu, dotąd słabo poznanych pod kątem genetycznym procesów, takich jak embriogeneza, różnicowanie tkanek, czy też odpowiedź na warunki stresowe.

Stosując mikromacierze dachówkowe reprezentujące cały genom rzodkiewnika, scharakteryzowano i porównano regiony aktywne transkrypcyjnie w czterech rodzajach tkanek (z kwiatów, liści, zawieszin komórkowych oraz korzeni) (53). Stwierdzono znaczące różnice w ekspresji genów (w kwiatach 16% wszystkich transkryptów stanowiły produkty ekspresji genów związanych z metabolizmem, w zawieszinach – jedynie 6,5%). Dodatkowo zbadano wpływ warunków wzrostu siewek (wzrost na świetle, w ciemności i chłodzie) na ekspresję genów i wykazano, że tylko 88% genów ulegało transkrypcji we wszystkich przeanalizowanych przypadkach. Z kolei na podstawie porównania z typem dzikim (ekotyp Columbia) dwóch mutantów rzodkiewnika *dif1* i *myb98*, nie mających zdolności do wytwarzania funkcjonalnych makrospor wykazano, że wiele specyficznych dla tego procesu genów kodowało małe, najprawdopodobniej wydzielane białka (przypuszczalnie cząsteczki sygnałowe związane z ostatnimi etapami rozwoju woreczka zalążkowego) (59). Znane jest, że mniej niż połowa z nich jest reprezentowana na popularnej mikromacierzy ekspresyjnej ATH1 dla rzodkiewnika, co wyraźnie obrazuje przewagę mikromacierzy równomiernie pokrywających genom w badaniach funkcjonalnych. Interesującym odkryciem było również scharakteryzowanie 22 genów o wysokiej homologii do genu *S1*, od-

powiadającego za samoniezdgodność u maku, mimo że występujący u roślin z rodziny krzyżowych typ samoniezdgodności ma zupełnie inne podłoże genetyczne. Co więcej, większość genów związanych z rozwojem woreczka zalążkowego ulegała wysokiej ekspresji jedynie w zalążni, wskazując na ich istotną i specyficzną rolę w rozwoju gametofitu żeńskiego (59).

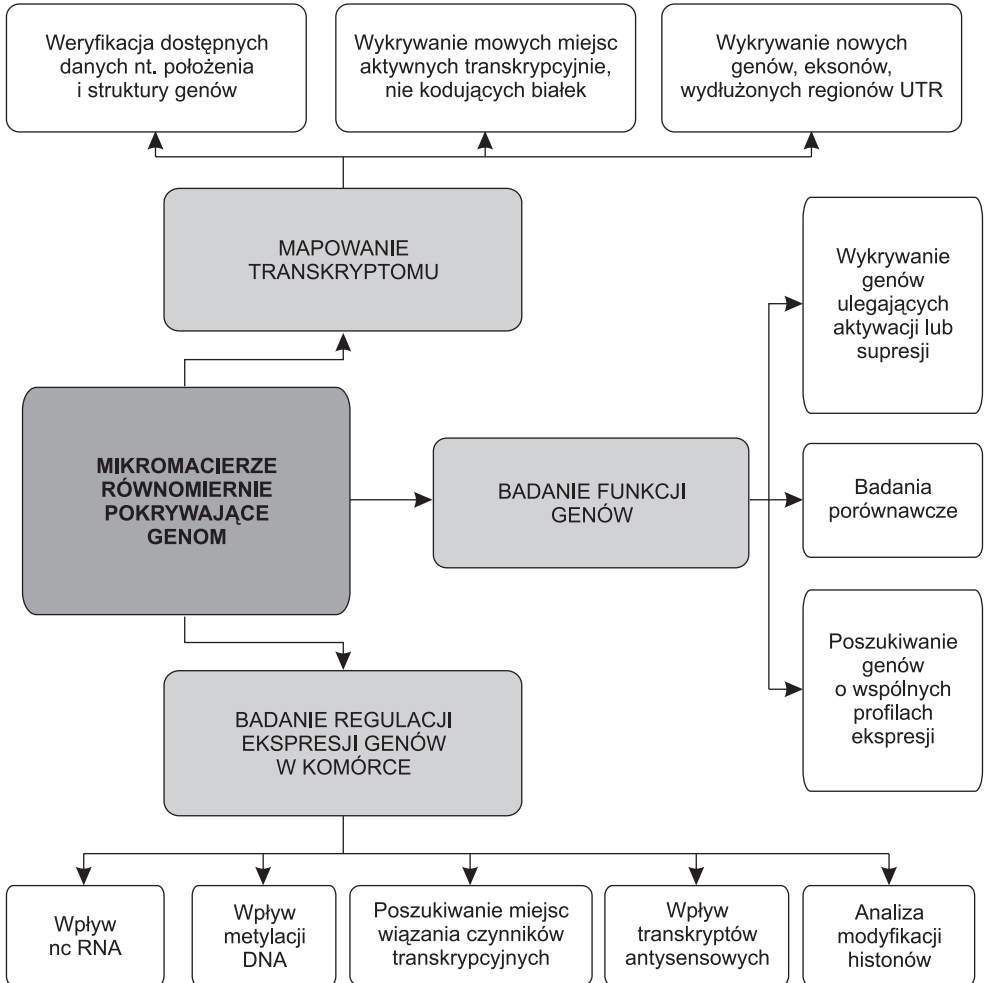
W badaniach dotyczących rozwoju osi grzbietowo-brzuszej u muszki owocowej charakteryzowano odpowiedź genów aktywowanych przez gradient białek Dor-sal (49). Dzięki zastosowaniu mikromacierzy dachówkowych wykryto 29 nowych genów biorących udział w regulacji prawidłowego przebiegu tego etapu rozwoju. Zidentyfikowano także dodatkowe eksony, wchodzące w skład znanych wcześniej białek, nie wykryte wcześniej transkrypty oraz wykazano ekspresję szeregu regionów międzygenowych, kodujących m.in. dwie cząsteczki miRNA. Z kolei Stölc i wsp., badając różne stadia rozwojowe muszki owocowej określili, że ekspresja około 15% genów jest zależna od stadium rozwojowego muszki, przy czym wiele z nich nie zostało dotąd scharakteryzowanych (67).

### 4.3. Badania porównawcze genomów

Mikromacierze dachówkowe zostały zastosowane również do porównywania pokrewnych genotypów (58,67). Przykładowo, w wyniku analizy porównawczej aktywności transkrypcyjnej chromosomu 10 ryżu, przeprowadzonej dla dwóch gatunków: *O. indica* i *O. japonica* wykazano, że dwie trzecie aktywnych transkrypcyjnie miejsc występuje u obu tych gatunków (58). Zauważono również, że poziom zachowawczości sekwencji nowo odkrytych transkryptów między oboma gatunkami był podobny do poziomu zachowawczości znanych genów. Zjawisko dość silnej zachowawczości nowych miejsc aktywnych transkrypcyjnie pomiędzy blisko spokrewnionymi organizmami zauważono także w genomach muszki owocowej *D. melanogaster*, oraz *D. pseudoobscura* (67).

### 4.4. Inne zastosowania

Na podstawie omówionych wyników analiz – z wykorzystaniem mikromacierzy dachówkowych – pokazano, że wszystkie przebadane dotąd genomy cechują się: 1) wysoką zawartością genów posiadających alternatywne formy splicingowe, 2) obecnością licznych transkryptów o bardzo niskim poziomie ekspresji i dotąd nieznanej funkcji, często nie kodujących białek, 3) niezwykle powszechną w genomach ekspresją z nici antysensowej oraz, 4) skomplikowanym mechanizmem regulacji ekspresji. Zauważono również, że nie opisane wcześniej miejsca aktywne transkrypcyjnie najliczniej występują w regionach o niskiej zawartości zidentyfikowanych dotąd genów (56). Jest to zgodne z obserwacjami, że zdecydowana większość no-



Rys. 2. Zastosowanie mikromacierzy równomiernie pokrywających genom w badaniach funkcjonalnych genomu.

wych regionów podlegających ekspresji odpowiada transkryptom nie kodującym białek i wynika z braku wcześniejszych dokładnych analiz regionów międzygenowych i intronów pod kątem potencjalnej aktywności transkrypcyjnej oraz tego, że dotychczas stosowane metody nie były dość czułe.

Mikromacierze dachówkowe przeżywają dynamiczny rozwój i znalazły już szeregi nowych zastosowań, takich jak badanie poziomu metylacji DNA, poszukiwanie nowych miejsc wiążących czynniki transkrypcyjne (ChIP-chip) czy detekcja zaburzeń w strukturze genomu (array-CGH) (26,28,68,69) (rys. 2), jednak nawet związane ich omówienie zdecydowanie wykracza poza ramy niniejszej pracy. W najbliższych la-

tach można z pewnością spodziewać się rozwoju kolejnych zastosowań dla mikromacierzy równomiernie pokrywających genom oraz dalszego udoskonalenia metod do ich analizy.

## 5. Podsumowanie

Jeden z najbardziej intrygujących wniosków wypływających z zakończonych już projektów sekwencjonowania genomów kilku modelowych organizmów dotyczył fragmentów kodujących białka. Stwierdzono, że wbrew wcześniejszym oczekiwaniom stanowią one zwykle jedynie niewielką część genomu (u człowieka około 3%). W rezultacie szereg pytań dotyczących tak fundamentalnych zagadnień jak: kodowanie i uwalnianie informacji genetycznej, nadal pozostaje bez odpowiedzi. Co więcej, okazało się, że stosując dostępne dotychczas metody badania procesu ekspresji genów nie jesteśmy w stanie dowiedzieć się, jak funkcjonuje zasadnicza część genomu. Ciekawą próbą zmierzenia się z tym problemem jest opracowanie nowego typu mikromacierzy zwanych dachówkowymi. Pozwalają one na jednoczesną analizę aktywności transkrypcyjnej całego genomu, a dzięki zastosowaniu odpowiednich metod statystycznych dają możliwość bezpośredniego porównania wyników uzyskanych dla wielu różnych tkanek, czy stadiów rozwojowych. Są również wystarczająco czułe, aby umożliwić detekcję transkryptów obecnych w bardzo niewielkich ilościach. Dodatkowo ich projektowanie wymaga „jedynie” znajomości pełnej sekwencji nukleotydowej genomu. Szersze zastosowanie mikromacierzy dachówkowych już w niedługim czasie może, jak się wydaje, zrewolucjonizować nasze poglądy na temat wielu procesów związanych z ekspresją informacji genetycznej (60).

Opracowanie powstało w ramach realizacji zamawianego projektu badawczego finansowanego przez MNiSW: nr PBZ-MNiSW-2/3/2006/32.

## Literatura

1. Schena M., Shalon D., Davis R. W., Brown P. O., (1995), *Science*, 270, 467-470.
2. Shalon D., Smith S. J., Brown P.O., (1996), *Genome Res.*, 6, 639-645.
3. Schena M., Shalon D., Heller R., Chai A., Brown P. O., Davis R. W., (1996), *Proc. Natl. Acad. Sci. USA*, 93, 10614-10619.
4. Hegde P., Qi R., Abernathy K., Gay C., Dharap S., Gaspard R., Hughes J. E., Snesrud E., Lee N., Quackenbush J., (2000), *BioTechniques*, 29, 548-562.
5. Kisiel A., Skąpska A., Markiewicz W. T., Figlerowicz M., (2004), *Kosmos*, 53, 295-303.
6. Venkatasubbarao S., (2004), *Trends Biotechnol.*, 22, 630-637.
7. Koehler J., Rawlings C., Verrier P., Mitchell R., Skusa A., Ruegg A., Philippi S., (2005), *In Silico Biol.*, 5, 33-44.
8. Hackl H., Burkard T. R., Sturn A., Rubio R., Schleiffer A., Tian S., Quackenbush J., Eisenhaber F., Trajanoski Z., (2005), *Genome Biol.*, 6, R108.

9. Ghassemian M., Lutes J., Tepperman J. M., Chang H. S., Zhu T., Wang X., Quail P. H., Lange B. M., (2006), *Arch. Biochem. Biophys.*, 448, 45-59.
10. Ryan P. A., Kirk B. W., Euler C. W., Schuch R., Fischetti V. A., (2007), *PLoS Comput. Biol.*, 3, e132.
11. Rotter A., Usadel B., Baebler S., Stitt M., Gruden K., (2007), *Plant Methods*, 3, 10.
12. Adler P., Reimand J., Jänes J., Kolde R., Peterson H., Vilo J., (2008), *Bioinformatics*, 24, 588-590.
13. Saito K., Hirai M. Y., Yonekura-Sakakibara K., (2008), *Trends Plant Sci.*, 13, 36-43.
14. Lapidot M., Pilpel Y., (2006), *EMBO Rep.*, 7, 1216-1222.
15. Werner A., (2005), *RNA Biol.*, 2, 53-62.
16. Yazgan O., Krebs J. E., (2007), *Biochem. Cell Biol.*, 85, 484-496.
17. Mattick J. S., Makunin I. V., (2006), *Hum. Mol. Genet.*, 15 Spec No 1, R17-R29.
18. Zhang B., Wang Q., Pan X., (2007), *J. Cell Physiol.*, 210, 279-289.
19. Chu C. Y., Rana T. M., (2007), *J. Cell Physiol.*, 213, 412-419.
20. Filipowicz W., Bhattacharyya S. N., Sonenberg N., (2008), *Nat. Rev. Genet.*, 9, 102-114.
21. Filipowicz W., (2000), *Proc. Natl. Acad. Sci. USA*, 97, 14035-14037.
22. Hüttenhofer A., Brosius J., Bachelier J. P., (2002), *Curr. Opin. Chem. Biol.*, 6, 835-843.
23. Beggs J. D., Tollervey D., (2005), *Nat. Rev. Mol. Cell Biol.*, 6, 423-429.
24. Hüttenhofer A., Schattner P., Polacek N., (2005), *Trends Genet.*, 21, 289-297.
25. Hüttenhofer A., Vogel J., (2006), *Nucleic Acids Res.*, 34, 635-646.
26. Johnson J. M., Edwards S., Shoemaker D., Schadt E. E., (2005), *Trends Genet.*, 21, 93-102.
27. Bertone P., Trifonov V., Rozowsky J. S., Schubert F., Emanuelsson O., Karro J., Kao M. Y., Snyder M., Gerstein M., (2006), *Genome Res.*, 16, 271-281.
28. Yazaki J., Gregory B. D., Ecker J. R., (2007), *Curr. Opin. Plant Biol.*, 10, 534-542.
29. Gräf S., Nielsen F. G., Kurtz S., Huynen M. A., Birney E., Stunnenberg H., Flicke P., (2007), *Bioinformatics*, 23, 195-204.
30. Emanuelsson O., Nagalakshmi U., Zheng D., Rozowsky J. S., Urban A. E., Du J., Lian Z., Stolc V., Weissman S., Snyder M., Gerstein M. B., (2007), *Genome Res.*, 17, 886-897.
31. Selinger D. W., Cheung K. J., Mei R., Johansson E. M., Richmond C. S., Blattner F. R., Lockhart D. J., Church G. M., (2000), *Nat. Biotechnol.*, 18, 1262-1268.
32. Royce T. E., Rozowsky J. S., Bertone P., Samanta M., Stolc V., Weissman S., Snyder M., Gerstein M., (2005), *Trends Genet.*, 21, 466-475.
33. Bertone P., Stolc V., Royce T. E., Rozowsky J. S., Urban A. E., Zhu X., Rinn J. L., Tongprasit W., Samanta M., Weissman S., Gerstein M., Snyder M., (2004), *Science*, 306, 2242-2246.
34. Ghosh S., Hirsch H. A., Sekinger E. A., Kapranov P., Struhl K., Gingeras T. R., (2007), *BMC Bioinformatics*, 8, 359.
35. Kapranov P., Cawley S. E., Drenkow J., Bekiranov S., Strausberg R. L., Fodor S. P., Gingeras T. R., (2002), *Science*, 296, 916-919.
36. Kampa D., Cheng J., Kapranov P., Yamanaka M., Brubaker S., Cawley S., Drenkow J., Piccolboni A., Bekiranov S., Helt G., Tammana H., Gingeras T. R., (2004), *Genome Res.*, 14, 331-342.
37. Schadt E. E., Edwards S. W., GuhaThakurta D., Holder D., Ying L., et al., (2004), *Genome Biol.*, 5, R73.
38. Royce T. E., Carriero N. J., Gerstein M. B., (2007), *BMC Bioinformatics*, 8, 186.
39. Do J. H., Choi D. K., (2006), *Mol. Cells*, 22, 254-261.
40. Leung Y. F., Cavalieri D., (2003), *Trends Genet.*, 19, 649-659.
41. Quackenbush J., (2002), *Nat. Genet.*, 32 Suppl, 496-501.
42. Zeller G., Henz S. R., Laubinger S., Weigel D., Rättsch G., (2008), *Pac Symp Biocomput.*, 12, 527-538.
43. Perocchi F., Xu Z., Clauder-Münster S., Steinmetz L. M., (2007), *Nucleic Acids Res.*, 35, e128.
44. David L., Huber W., Granovskaia M., Toedling J., Palm C. J., Bofkin L., Jones T., Davis R. W., Steinmetz L. M., (2006), *Proc. Natl. Acad. Sci. USA*, 103, 5320-5325.
45. Nelson D. C., Wohlbach D. J., Rodesch M. J., Stolc V., Sussman M. R., Samanta M. P., (2007), *FEBS Lett.*, 581, 3363-3370.
46. Tjaden B., Saxena R. M., Stolyar S., Haynor D. R., Kolker E., Rosenow C., (2002), *Nucleic Acids Res.*, 30, 3732-3738.

47. He H., Wang J., Liu T., Liu X. S., Li T., Wang Y., Qian Z., Zheng H., Zhu X., Wu T., Shi B., Deng W., Zhou W., Skogerbø G., Chen R., (2007), *Genome Res.*, 17, 1471-1477.
48. Stolc V., Gauhar Z., Mason C., Halasz G., van Batenburg M. F., Rifkin S. A., Hua S., Herreman T., Tongprasit W., Barbano P. E., Bussemaker H. J., White K. P., (2004), *Science*, 306, 655-660.
49. Biemar F., Nix D. A., Piel J., Peterson B., Ronshaugen M., Sementchenko V., Bell I., Manak J. R., Levine M. S., (2006), *Proc. Natl. Acad. Sci. USA*, 103, 12763-12768.
50. Manak J. R., Dike S., Sementchenko V., Kapranov P., Biemar F., Long J., Cheng J., Bell I., Ghosh S., Piccolboni A., Gingeras T. R., (2006), *Nat. Genet.*, 38, 1151-1158.
51. Cheng J., Kapranov P., Drenkow J., Dike S., Brubaker S., Patel S., Long J., Stern D., Tammanna H., Helt G., Sementchenko V., Piccolboni A., Bekiranov S., Bailey D. K., Ganesh M., Ghosh S., Bell I., Gerhard D. S., Gingeras T. R., (2005), *Science*, 308, 1149-1154.
52. Rinn J. L., Euskirchen G., Bertone P., Martone R., Luscombe N. M., Hartman S., Harrison P. M., Nelson F. K., Miller P., Gerstein M., Weissman S., Snyder M., (2003), *Genes Dev.*, 17, 529-540.
53. Yamada K., Lim J., Dale J. M., Chen H., Shinn P., Palm C. J., et al., (2003), *Science*, 302, 842-846.
54. Stolc V., Samanta M. P., Tongprasit W., Sethi H., Liang S., Nelson D. C., et al., (2005), *Proc. Natl. Acad. Sci. USA*, 102, 4453-4458.
55. Hanada K., Zhang X., Borevitz J. O., Li W. H., Shiu S. H., (2007), *Genome Res.*, 17, 632-640.
56. Li L., Wang X., Sasidharan R., Stolc V., Deng W., He H., Korbel J., Chen X., Tongprasit W., Ronald P., Chen R., Gerstein M., Wang Deng X., (2007), *PLoS ONE* 2, e294.
57. Li L., Wang X., Stolc V., Li X., Zhang D., Su N., Tongprasit W., Li S., Cheng Z., Wang J., Deng X. W., (2006), *Nat. Genet.*, 38, 124-129.
58. Li L., Wang X., Xia M., Stolc V., Su N., Peng Z.; Waraporn T., Li S., Wang J., Wang X., Deng X. W., (2005), *Genome Biol.*, 6, R52.
59. Jones-Rhoades M. W., Borevitz J. O., Preuss D., (2007), *PLoS Genet.*, 3, 1848-1861.
60. Willingham A. T., Gingeras T. R., (2006), *Cell*, 125, 1215-122.
61. Pang K. C., Frith M. C., Mattick J. S., (2006), *Trends Genet.*, 22, 1-5.
62. Khaitovich P., Kelso J., Franz H., Visagie J., Giger T., Joerchel S., Petzold E., Green R. E., Lachmann M., Pääbo S., (2006), *PLoS Genet.*, 2, e171.
63. Babak T., Blencowe B. J., Hughes T. R., (2005), *BMC Genomics*, 6, 104.
64. Katayama S., Tomaru Y., Kasukawa T., Waki K., Nakanishi M., Nakamura M., et al., (2005), *Science*, 309, 1564-1566.
65. Swiezewski S., Crevillen P., Liu F., Ecker J. R., Jerzmanowski A., Dean C., (2007), *Proc. Natl. Acad. Sci. USA*, 104, 3633-3638.
66. Borsani O., Zhu J., Verslues P. E., Sunkar R., Zhu J. K., (2005), *Cell*, 123, 1279-1291.
67. Stolc V., Gauhar Z., Mason C., Halasz G., van Batenburg M. F., Rifkin S. A., Hua S., Herreman T., Tongprasit W., Barbano P. E., Bussemaker H. J., White K. P., (2004), *Science*, 306, 655-660.
68. Liu X. S., (2007), *PLoS Comput. Biol.*, 3, 1842-1844.
69. Hanlon S. E., Lieb J. D., (2004), *Curr. Opin. Genet. Dev.*, 14, 697-705.