

# **Podejścia bioinformatyczne w badaniach kolistych RNA u roślin**

mgr Katarzyna Nowis

## **STRESZCZENIE**

Koliste cząsteczki RNA (circRNA, ang. *circular RNA*) zostały odkryte już w latach 70. ubiegłego wieku. W przypadku roślin był to wiroidowy RNA (Sanger i in., 1976) a w przypadku człowieka transkrypt występujący w cytoplazmie linii komórkowej HeLa (Hsu i in., 1979). Jednakże, przez długi czas circRNA nie cieszyły się większym zainteresowaniem badaczy, gdyż powszechnie sądzono, że ich powstanie jest wynikiem błędów aparatu splicingowego. Sytuacja ta uległa zmianie dopiero na początku XXI wieku wraz z powstaniem technik wysokoprzepustowego sekwencjonowania nowej generacji (NGS, ang. *next-generation sequencing*), które dały pełniejszy wgląd w strukturę transkryptomów. Stwierdzono między innymi, że circRNA są zachowawcze ewolucyjnie i występują powszechnie w organizmach żywych. Obserwacje te pozwalają przypuszczać, że koliste transkrypty pełnią istotne funkcje biologiczne. W rezultacie liczba publikacji dotyczących circRNA zaczęła gwałtownie wzrastać, wciąż jednak niewiele wiadomo na temat biogenezy i funkcji tych cząsteczek szczególnie w odniesieniu do roślin.

Przeprowadzony przegląd literaturowy ujawnił szereg problemów z jakimi zmagają się badacze analizując circRNA u roślin. Pierwszym z nich okazał się brak standardowych protokołów laboratoryjnych i bioinformatycznych dedykowanych analizie circRNA. Powodowało to, że autorzy większości prac jedynie identyfikowali circRNA, nie przeprowadzając porównawczych analiz ich akumulacji w organach czy ekotypach. Kolejnym problemem był brak jednolitego sposobu deponowania informacji na temat roślinnych circRNA w publicznie dostępnych bazach danych. Nie istniała zatem możliwość bezpośredniego porównywania już zidentyfikowanych circRNA oraz poziomów ich akumulacji w różnych organach czy liniach. Jak dotąd

dostępne były trzy bazy danych zawierające informacje na temat circRNA u *Arabidopsis thaliana*. Repozytoria te udostępniają informacje na temat circRNA, całkowicie ignorując zagadnienia dotyczące wiarygodności przeprowadzonych analiz jakościowych i ilościowych.

W celu rozwiązania tych problemów opracowałam protokoły bioinformatyczne umożliwiające jakościową i ilościową analizę circRNA występujących w różnych tkankach czy organach roślin. Stworzone protokoły zastosowałam do scharakteryzowania circRNA w kwiatach, liściach, korzeniach i siewkach *A. thaliana* typu dzikiego (ekotyp Columbia, w skrócie Col-0) oraz do zidentyfikowania białek wpływających na biogenezę circRNA. Uzyskane wyniki wskazują, że u *A. thaliana* większość circRNA powstaje w procesach stochastycznych regulowanych niezależnie od transkrypcji liniowych odpowiedników. Ponieważ circRNA są najprawdopodobniej produktami składania transkryptów, stąd, aby zidentyfikować białka uczestniczące w ich biogenezie scharakteryzowaliśmy circRNA powstające w mutantach *A. thaliana* pozbawionych pojedynczych genów kodujących białka splicingowe. Analizy circRNA w typie dzikim oraz 18 mutantach typu knock-out ujawniły, że wyłączenie trzech genów związanych ze splicingiem wywiera istotny wpływ na produkcję i akumulację circRNA (mutanty: *cbp80*, *flk*, *c2h2*). Brak białek kodowanych przez te geny zaburzał proces składania transkryptów i przyczyniał się do wzrostu produkcji circRNA. Co więcej, w przypadku dwóch wariantów *cbp80*, *c2h2* stwierdziliśmy obecność dużej liczby unikatowych circRNA nie występujących w typie dzikim *A. thaliana*. Dodatkowo, opracowane przeze mnie protokoły dedykowane badaniom circRNA wykorzystałam do analizy publicznie dostępnych danych RNA-seq dla *A. thaliana*. Wszystkie uzyskane wyniki zdeponowałam w stworzonej specjalnie w tym celu publicznej bazie danych (<http://plantcircrna.ibch.poznan.pl/>). Tym samym powstała możliwość jakościowej i ilościowej analizy porównawczej wszystkich circRNA zidentyfikowanych u *A. thaliana*.