

AUTOREFERAT

DR ANNA PHILIPS

ZAŁĄCZNIK 4

I. IMIĘ I NAZWISKO

ANNA PHILIPS

II. POSIADANE DYPLOMY, STOPNIE NAUKOWE

- 2013 **Stopień doktora** nauk biologicznych (spec. bioinformatyka)
Uniwersytet im. Adama Mickiewicza w Poznaniu
Instytut Biologii Molekularnej i Biotechnologii
Promotor: prof. dr hab. Janusz M. Bujnicki
Tytuł: *„Nowe metody bioinformatyczne do przewidywania miejsc wiązania metali i ligandów w strukturach RNA”* (obrona z wyróżnieniem).
- 2008 **Stopień magistra** informatyki (spec. informatyka stosowana)
Uniwersytet im. Adama Mickiewicza w Poznaniu
Wydział Fizyki, Laboratorium Bioinformatyki
Promotor: prof. Andrzej Dobek, dr Leszek Rychlewski
Tytuł: *„Analiza strukturalna konformacji liganda w dokowaniu białko-ligand”*.

III. INFORMACJA O DOTYCHCZASOWYM ZATRUDNIENIU W JEDNOSTKACH NAUKOWYCH

- 2013– Adiunkt w Instytucie Chemii Bioorganicznej Polskiej Akademii Nauk
Od 01.2020 kierownik Pracowni Bioinformatyki
- 2008–2009 Bioinformatyk w Instytucie Biologii Molekularnej i Komórkowej w Warszawie

IV. OMÓWIENIE OSIĄGNIĘCIA

1. TYTUŁ

Opracowanie metod i narzędzi analizy ilościowej i jakościowej kolistych RNA oraz ich wykorzystanie w różnych układach biologicznych.

2. PUBLIKACJE WCHODZĄCE W SKŁAD OSIĄGNIĘCIA

- H1. **Philips A**, Nowis K, Stelmaszczuk M, Jackowiak P, Podkowiński J, Handschuh L, Figlerowicz M.
Expression Landscape of circRNAs in Arabidopsis thaliana Seedlings and Adult Tissues.
Front Plant Sci. 2020 Sep 10;11:576581. doi: 10.3389/fpls.2020.576581.
IF₂₀₂₀ = 5,753; 5IF₂₀₂₀ = 6,612; MNiSW: 100.
- H2. **Philips A**, Nowis K, Stelmaszczuk M, Podkowiński J, Handschuh L, Jackowiak P, Figlerowicz M.
Arabidopsis thaliana cbp80, c2h2, and flk Knockout Mutants Accumulate Increased Amounts of Circular RNAs.
Cells. 2020 Aug 21;9(9):1937. doi: 10.3390/cells9091937.
IF₂₀₂₀ = 6,6; 5IF₂₀₂₀ = 6,663; MNiSW: 140.
- H3. Nowis K, Jackowiak P, Figlerowicz M, **Philips A***
At-C-RNA database, a one-stop source for information on circRNAs in Arabidopsis thaliana in a unified format.
Database (Oxford), 2021, 00(0), 1–5, doi: 10.1093/database/baab07.
IF₂₀₂₀ = 3,451; 5IF₂₀₂₀ = 4,159; MNiSW: 100, *autor korespondencyjny.
- H4. Czubak K, Taylor K, Piasecka A, Sobczak K, Kozłowska K, **Philips A**, Sedehizadeh S, Brook JD, Wojciechowska M, Kozłowski P.
Global Increase in Circular RNA Levels in Myotonic Dystrophy.
Front Genet. 2019 Jul 18;10:649. doi: 10.3389/fgene.2019.00649.
IF₂₀₂₀ = 3,26; 5IF₂₀₂₀ = 4,007; MNiSW: 100.

3. OMÓWIENIE OSIĄGNIĘCIA

Omówione poniżej badania, składające się na moje osiągnięcie naukowe, zatytułowane „Opracowanie metod i narzędzi analizy ilościowej i jakościowej kolistych RNA oraz ich wykorzystanie w różnych układach biologicznych” realizowałam w latach 2014-2021 w Instytucie Chemii Bioorganicznej PAN (ICHB PAN) w Poznaniu. Prace te powstały w ramach grantu SONATA NCN, którego byłam kierownikiem. Moje osiągnięcie naukowe stanowi zbiór czterech wieloautorskich publikacji, z których wszystkie przedstawiają wyniki oryginalnych badań. Prace są opublikowane w modelu otwartego dostępu (ang. *open access*).

Jestem pierwszą autorką we dwóch z tych prac, w jednej autorką korespondencyjną. Pliki rzeczonych publikacji w formacie PDF zostały dołączone do elektronicznej wersji wniosku.

WPROWADZENIE

Pierwsze koliste RNA (circRNA) zostały opisane już ponad 20 lat temu (Nigro et al. 1991). Początkowo uważano je za rzadkie cząsteczki, najprawdopodobniej będące efektem błędów w procesie składania transkryptów (ang. *splicing*). Dlatego też przez długi czas ich występowanie łączono jedynie z charakterystyką komórek nowotworowych. Zastosowanie sekwencjonowania nowej generacji (NGS) do analizy transkryptomów (RNA-seq) przyniosło duży przełom w dziedzinie biologii circRNA. Okazało się, że circRNA są stabilne i powszechnie występują w zdrowych tkankach organizmów eukariotycznych (Salzman et al. 2012). Wykazano obecność i ewolucyjną konserwację circRNA u wielu gatunków w królestwach zwierząt i roślin (Wang et al. 2014). Jednak bardziej szczegółowe badania skupiały się dotychczas na ludziach i kilku zwierzętach modelowych (głównie badania prowadzone nad rakiem i innymi chorobami (Liu et al. 2019, Patop et al. 2019)).

Biologiczna funkcja większości z tych cząsteczek nie została jeszcze poznana. Istnieją dowody, że niektóre z circRNA odgrywają ważne role w regulacji ekspresji genów. Na przykład, koliste transkrypty CDR1as/CIRS-7 i *Sry* funkcjonują jako gąbki mikro RNA (miRNA) (Hansen et al. 2013, Memczak et al. 2013). Cząsteczki te posiadają wiele miejsc wiążących miRNA, dzięki czemu sekwestrują miRNA z ich miejsc docelowych w mRNA. Spekuluje się również, że niektóre circRNA mogą być związane z różnymi chorobami, głównie o podłożu neurologicznym, jak choroba Alzheimera (Song et al. 2021). Ponieważ wiadomo, że cząsteczki circRNA powstają w wyniku niekanonicznego splicingu liniowego pre-mRNA, a w wypadku wielu genów stosunek transkryptów kolistych do liniowych jest specyficzny dla typu komórki/tkanki (Salzman et al. 2013), spekuluje się również, że circRNA mogą działać w regulacji ekspresji genów poprzez konkutowanie z produkcją liniowych transkryptów (Ashwal-Fluss et al. 2014, Shao et al. 2021).

Koliste RNA u roślin pozostają niezbadane w jeszcze większym stopniu. Do tej pory wykazano jedynie, że circRNA pochodzący z eksonu 6 genu *SEPALLATA3* (*SEP3*) reguluje rozwój kwiatów u *Arabidopsis thaliana* (Conn et al. 2017). Niektóre doniesienia sugerują także, że circRNA w roślinach ulegają ekspresji w sposób specyficzny dla tkanek i stadium rozwoju (Sun et al. 2016, Chen et al. 2017, Liu et al. 2017). Pan i współpracownicy (Pan et al. 2018a) zidentyfikowali 1 583 circRNA specyficznych dla stresu cieplnego u *A. thaliana*, które nie występowały w próbkach kontrolnych. Nie wiadomo jednak jaką rolę odgrywają te cząsteczki.

Badania kolistych RNA utrudnia fakt, że do tej pory nie opracowano standardów ich analiz. Większość badań poświęconych circRNA była prowadzona na podstawie danych RNA-seq uzyskanych we wcześniejszych eksperymentach dedykowanych innym zaganieniom, na przykład ustalaniu poziomów mRNA. Niestety, ze względu na znaczne różnice w procedurach stosowanych w poszczególnych badaniach wygenerowane dane circRNA są nieporównywalne. Problem ten jest szczególnie widoczny na przykładzie roślin. Do tej pory circRNA zidentyfikowano w wielu gatunkach, m. in. w *Oryza sativa* (Lu et al. 2015), *Zea mays*

(Chen et al. 2018), *Solanum lycopersicum* (Zuo et al. 2016), *Glycine max*, *Gossypium hirsutum* (Zhao et al. 2017), *Triticum aestivum* (Wang et al. 2016), *Hordeum vulgare* (Darbani et al. 2016) oraz *A. thaliana* (Liu et al. 2017). Pomimo to, nadal niewiele wiadomo na temat ich funkcjonowania i biogenezy w tym królestwie. Obecnie, *A. thaliana* jest najlepiej scharakteryzowaną rośliną pod względem circRNA. Niestety, przez wspomniane powyżej problemy sama liczba circRNA znalezionych u *A. thaliana* waha się od 6 (Wang et al. 2014) do 6 012 (Ye et al. 2015), w zależności od podejścia eksperymentalnego. Według PlantcircBase (<http://ibi.zju.edu.cn/plantcircbase/index.php>; 17 września 2021 r.), całkowita liczba circRNA u *A. thaliana* zidentyfikowanych w różnych badaniach osiągnęła 52 393. Brak protokołu składającego się z izolacji, sekwencjonowania, identyfikacji i analizy ilościowej optymalizowanego pod kątem badań circRNA sprawia, że wyniki pochodzące z różnych źródeł są niespójne i praktycznie nieporównywalne. W związku z tym, na pytania dotyczące poziomów akumulacji circRNA w roślinach i ich tkankach, a co za tym idzie, o ich znaczenie funkcjonalne nie uzyskano do tej pory satysfakcjonującej odpowiedzi.

MOTYWACJA I CEL BADAŃ

Podstawowym celem moich badań było opracowanie zoptymalizowanego protokołu służącego do identyfikacji, a także analizy ilościowej i jakościowej circRNA. Kolejnym moim celem była weryfikacja opracowanej metody w różnych układach biologicznych. Wykorzystałam układ roślinny, *A. thaliana* i przy pomocy opracowanej metody (i) scharakteryzowałam circRNA w siewce i w tkankach: korzeniu, liściu, kwiecie tej rośliny. Metodę użyłam także do (ii) identyfikacji czynników genetycznych potencjalnie uczestniczących w procesie biogenezy tych cząsteczek. Do badań wybrałam *A. thaliana*, jedną z najlepiej przebadanych roślin dwuliściennych, ponieważ roślina ta jest łatwa i szybka w hodowli, a przede wszystkim dlatego, że ma tysiące łatwo dostępnych wariantów, w tym dobrze scharakteryzowane nokauty pojedynczego genu. Pozwoliło to na selekcję roślin zawierających mutacje typu knock-out w genach kodujących białka biorące udział w splicingu na różnych jego etapach. Główna hipoteza badawcza zakładała, że analiza pul circRNA w wybranych mutantach *A. thaliana* pozwoli na zidentyfikowanie genów, których produkty potencjalnie wpływają na biogenezę tych cząsteczek. (iii) Opracowaną metodę wykorzystałam następnie, do ponownego przeanalizowania publicznie dostępnych danych NGS odpowiednich do identyfikacji circRNA u *A. thaliana*, a otrzymane wyniki umieściłam w nowo stworzonej bazie danych. (iv) Dodatkowo, protokół ten przetestowałam w systemie ludzkich tkanek. Zbadałam w jaki sposób zmieniają się pule generowanych circRNA i ich poziomy w mięśniach ludzi chorych na dystrofię miotoniczną typu 1 (DM1). Przeprowadzone badania udowodniły skuteczność opracowanego podejścia zarówno w układach roślinnych jak i ludzkich tkankach. Zaproponowany protokół postępowania służący do identyfikacji oraz analizy ilościowej i jakościowej circRNA ma zatem szerokie spektrum stosowalności, co pozwala na jego wykorzystanie niezależnie od testowanego układu biologicznego.

ZOPTYMALIZOWANY PROTOKÓŁ POZYSKANIA I ANALIZY DANYCH CIRC RNA

Pierwszym celem mojej pracy była wiarygodna identyfikacja i charakterystyka circRNA w siewkach i organach (korzeniach, liściach, kwiatach) *A. thaliana* typu dzikiego (ekotyp Columbia, w skrócie Col-0). W czasie, gdy zaczynałam swoje badania istniał już wymieniony powyżej szereg publikacji poświęcony badaniom circRNA u *A. thaliana* i w innych roślinach. Niestety, krytyczny przegląd tych prac ujawnił, że opublikowane wyniki są w dużej mierze niespójne ze względu na brak standaryzowanych metod. Przykładowo, wykorzystywano różne podejścia do przygotowania danych do sekwencjonowania, najczęściej z użyciem RNazy R trawiącej liniowe transkrypty lub bez. Dodatkowo, podczas analiz bioinformatycznych rutynowo uznawano circRNA za zidentyfikowany, jeśli jego obecność była poparta co najmniej dwoma odczytami NGS mapującymi do miejsca tzw. „kolistego” splicingu (tzw. odczyty typu back-spliced). Głębokość biblioteki RNA-seq zwykle nie była brana pod uwagę, a zatem liczba potencjalnych circRNA różniła się znacznie między badaniami. Zidentyfikowane na podstawie danych RNA-seq circRNA tylko w kilku przypadkach zostały poddane walidacji eksperymentalnej za pomocą ilościowego PCR (qPCR) (Wang et al. 2014, Liu et al. 2017, Cheng et al. 2018, Pan et al. 2018b).

Znając problemy napotymane przez innych badaczy, w pierwszej kolejności podjęłam próbę oceny i optymalizacji protokołów dedykowanych identyfikacji i analizie ilościowej circRNA, zarówno pod względem przygotowania bibliotek RNA-seq, jaki i późniejszej analizy bioinformatycznej danych z sekwencjonowania (Publikacja H1). Moje badania skupiły się na RNA-seq, ponieważ jest to obecnie jedyna metoda, której zastosowanie pozwala uzyskać kompletny obraz circRNA w organizmie, czy jego poszczególnych tkankach.

Przetestowałam dwa wiodące protokoły przygotowania bibliotek NGS do badań circRNA (z traktowaniem RNazą R i bez) oraz kilka podejść normalizacji danych RNA-seq. Obiektem moich badań były circRNA w siewkach i organach *A. thaliana* typu dzikiego, z których wyizolowano całkowity RNA a następnie przygotowano biblioteki RNA-seq przy użyciu wspomnianych dwóch protokołów. Zgodnie z pierwszym podejściem, użyto całkowitego RNA, po deplecji rRNA zatem próbki zawierały zarówno liniowe, jak i koliste transkrypty. Zgodnie z drugim podejściem, użyto całkowitego RNA zubożonego w rRNA, a następnie potraktowanego RNazą R, która trawi liniowe transkrypty. Próbkę zawierały zatem circRNA i śladowe ilości liniowego RNA. Wszystkie eksperymenty RNA-seq przeprowadzono w czterech powtórzeniach biologicznych.

Porównanie uzyskanych wyników wykazało, że traktowanie RNazą R w celu wzbogacenie próbek w circRNA, istotnie pozwala na identyfikację circRNA występujących mniej licznie, które to circRNA w standardowych bibliotekach są pomijane z powodu dużej liczby transkryptów liniowych. Biblioteki po traktowaniu RNazą R posiadały średnią liczbę odczytów w przybliżeniu o połowę mniejszą niż biblioteki bez traktowania RNazą R, mimo to, liczba zidentyfikowanych potencjalnych circRNA była w nich trzykrotnie wyższa. Tak więc, traktowanie RNazą R wyraźnie zwiększyło rozdzielczość analizy. Z drugiej strony stwierdziłam, że niektóre circRNA, które zostały zidentyfikowane w czterech powtórzeniach

biologicznych w bibliotekach nie traktowanych RNazą R (i zostały zwalidowane za pomocą RT-PCR), nie zostały zidentyfikowane w bibliotekach wzbogacanych w circRNA. Spowodowane to było prawdopodobnie nieenzymatyczną hydrolizą niektórych kolistych transkryptów, co uczyniło je liniowymi i narażonymi na trawienie RNazą R. Hydroliza nieenzymatyczna RNA jest dobrze znanym zjawiskiem, które zachodzi w buforze, w obecności wielu czynników, w tym jonów magnezu, polimerycznych związków organicznych, takich jak poliwinylpirolidon (PVP) oraz białek. Podsumowując, traktowanie próbek RNazą R wzbogaca circRNA, jednak ta procedura może również prowadzić do utraty informacji o niektórych z tych cząsteczek. W konsekwencji zdecydowałam, że podejście bez traktowania RNazą R jest bardziej odpowiednie do identyfikacji i oceny ilościowej całkowitej puli circRNA.

Aby porównać poziomy akumulacji circRNA w różnych organach/tkankach roślinnych, konieczne jest zastosowanie dedykowanej circRNA metody normalizacji danych RNA-seq. Postanowiłam wykorzystywać metodę ddPCR (ang. *droplet digital PCR*) do wyznaczenia wartości referencyjnych akumulacji wybranych circRNA, aby ustalić optymalny sposób normalizacji danych RNA-seq. Najpowszechniej stosowana normalizacja – przez wielkość biblioteki okazała się nieskuteczna w przypadku analiz circRNA u roślin. Uzyskane dane były słabo skorelowane z wynikami ddPCR (współczynnik korelacji Pearsona: 0,67). W następnym kroku przetestowałam cztery inne metody normalizacji: przez (i) tzw. liczbę odczytów back-spliced (odczyty mapujące do miejsca splicingowego typowego dla circRNA) (ii) współczynnik normalizacji oszacowany przez DESeq2 (iii) wielkość biblioteki bez odczytów rRNA i chloroplastowego RNA oraz (iv) poziom ekspresji genu *ACT2*. Zastosowanie tych podejść pozwoliło na osiągnięcie współczynnika korelacji Pearsona między wynikami ddPCR i RNA-seq rzędu 0,76-0,77 (Tabela 1 z publikacji H1). W przypadku jedno-organowych analiz ilościowych circRNA korelacja była jeszcze wyższa, sięgając 0,92 (dla liści). Wyniki przeprowadzonych badań potwierdziły pogląd, że skład transkryptomu obserwowany różnych tkankach rośliny ma istotny wpływ na analizę ilościową rzadkich gatunków RNA (w tym circRNA) w oparciu o RNA-seq. Ponadto, wyniki te wskazują na wysoką wiarygodność ilościowej charakterystyki circRNA w oparciu o RNA-seq w danym organie, pod warunkiem zastosowania odpowiedniej metody normalizacji.

CHARAKTERYSTYKA CIRC RNA U *A. THALIANA* (COL-0)

Stosując opisaną powyżej zoptymalizowaną strategię, zidentyfikowałam i zbadałam akumulację kolistych RNA w siewkach i trzech organach (korzeniach, liściach i kwiatach) *A. thaliana* typu dzikiego, Col-0 (Publikacja H1). Dodatkowo przeprowadziłam porównanie pul circRNA z różnych tkanek i porównałam ich akumulację z akumulacją odpowiadających im liniowych transkryptów. Doprowadziło do identyfikacji circRNA występujących powszechnie w tej roślinie, jak i takich specyficznych tkankowo.

W wyniku przeprowadzonej analizy zidentyfikowałam łącznie 5 235 różnych circRNA popartych przez co najmniej dwa odczyty typu back-spliced. Już na początku analizy danych RNA-seq zauważyłam słabą powtarzalność circRNA w powtórzeniach biologicznych (w zakresie od 1,3 do 5,0%, Rysunek 2 A z publikacji H1). Było to zaskakujące odkrycie, gdyż

do tamtej pory żadna z publikacji poświęconych roślinnym circRNA nie informowała o podobnej charakterystyce danych. Błąd eksperymentalny został wyeliminowany (i) wprowadziliśmy dwa dodatkowe powtórzenia biologiczne; (ii) przeprowadziliśmy tożsame analizy circRNA w ludzkich liniach komórkowych Huh-7.5 i HeLa, które nie wykazały takiej zmienności; (iii) i analizę publicznie dostępnych danych RNA-seq dla circRNA u *A. thaliana*, która potwierdziła naszą obserwację. Otrzymane wyniki jasno sugerowały, że pula circRNA w *A. thaliana* zawiera znaczną część cząsteczek generowanych stochastycznie, których znaczenie fizjologiczne jest najpewniej zaniedbywalne. Większość wykrytych circRNA (4 297 circRNA = 82,1%) była poparta przez 2 do 5 odczytów back-spliced. Jednak zidentyfikowałam również 39 kolistych RNA wspieranych przez więcej niż 100 odczytów back-spliced. Pośród wszystkich tych potencjalnych circRNA 127 zostały znalezione przez dwa niezależne programy, CIRI2 (Gao et al. 2017) i find_circ (Memczak et al. 2013) oraz poparte przez co najmniej 10 znormalizowanych odczytów RNA-seq (back-splice), co odpowiadało ok. 50 kopiom circRNA na mikrogram całkowitego RNA. Zastosowanie tak restrykcyjnej selekcji pozwoliło na wyodrębnienie frakcji circRNA które występują w roślinie w sposób powtarzalny, w ilościach pozwalających na pełnienie przez nie potencjalnych funkcji w komórce. Zaproponowana przez mnie metoda umożliwia zatem nie tylko identyfikację i analizę ilościową, ale i oddzielenie circRNA produkowanych w sposób losowy od tych circRNA produkowanych w sposób powtarzalny, o potencjalnym znaczeniu biologicznym.

Liczba powtarzalnych circRNA wahała się między badanymi tkankami, od 87 w korzeniach, 79 w liściach, 99 w kwiatach po 109 circRNA w siewkach (Rysunek 3 A z publikacji H1). W sumie zidentyfikowałam 127 circRNA produkowanych w sposób powtarzalny. 40,9% (52) z nich było obecnych we wszystkich badanych tkankach rośliny. 88,2% (112) z nich zostało już zidentyfikowanych we wcześniejszych badaniach i zdeponowanych w bazie danych PlantcircBase (Chu et al. 2017), a 52,8% (67) w AtCircDB (Ye et al. 2017). Analiza ontologii genów (ang. *gene ontology*, GO) ujawniła, że geny dające początek powtarzalnym circRNA charakteryzują się znaczącym wzbogaceniem w te zaangażowane w fotosyntezę i odpowiedź na stres. Korzenie miały największą liczbę unikalnych circRNA (6) i najwięcej circRNA o dużej akumulacji (wartość $p \leq 0,05$) w porównaniu z innymi tkankami. Najmniejszą liczbę powtarzalnych circRNA o podwyższonym poziomie stwierdzono u siewek. Dalsze analizy koncentrowały się na circRNA obecnych w co najmniej dwóch tkankach. Porównanie par tkanek ujawniło specyficzny dla nich wzór akumulacji circRNA (Rysunek 4 A, B z publikacji H1). Największa liczba circRNA o podniesionym poziomie (wartość $p \leq 0,05$) w porównaniu z innymi tkankami, charakteryzowała korzenie (16 circRNA vs siewki, 25 vs liście i 9 vs kwiaty). Najmniejsza liczba circRNA o podniesionym poziomie została zidentyfikowana w siewkach (1 vs korzenie, 3 vs liście i 1 vs kwiaty).

Przeprowadzone analizy wykazały również obecność circRNA pochodzących z genomu chloroplastowego w siewkach (24), korzeniach (2), liściach (12) i kwiatach (14), Rysunek 3 C, D z publikacji H1. Chloroplasty zostały zidentyfikowane jako „hot spoty” produkcji circRNA w (Sun et al. 2016). Nasze wyniki ukazały, że w liściach rzeczywiście przeważały odczyty typu back-spliced wspierające circRNA pochodzące z genomu

chloroplastowego. Niestety do dziś nie wyjaśniono, w jaki sposób tak zasadniczo różne procesy, jak splicing jądrowy i organellarny, mogłyby prowadzić do tworzenia circRNA. Kwestię tę komplikuje jeszcze fakt, że niektóre organellarne mRNA są produktami *trans* splicingu, w którym eksony dwóch pre-mRNA zostają połączone (Aryamanesh et al. 2017). Niektóre odczyty pochodzące z takich mRNA mogą być błędnie zaklasyfikowane jako odczyty typu back-spliced. W związku z powyższym chloroplastowe circRNA zidentyfikowane w danych RNA-seq powinny być traktowane z dużą dozą ostrożności a ich obecność powinna zostać potwierdzona przy pomocy innych metod.

Analiza poziomów ekspresji mRNA wyraźnie wykazała, że nie ma bezpośredniego związku między akumulacją circRNA a akumulacją ich liniowych odpowiedników. Wybrane przykłady, poparte metodą ddPCR (Rysunek 7 z publikacji H1) pokazały, że poziom niektórych circRNA wzrasta wraz ze wzrostem ekspresji ich liniowych odpowiedników, jednak ogólnie nie ma korelacji między akumulacją tych dwóch frakcji transkryptomu (współczynnik korelacji Pearsona: 0,14). To odkrycie wskazało, na duże prawdopodobieństwo istnienia specyficznego mechanizmu regulującego powstawanie circRNA u *A. thaliana*. Co więcej, poziom niektórych circRNA był zwiększony w poszczególnych tkankach, podczas gdy poziomy ich liniowych odpowiedników pozostawały względnie stabilne w całej roślinie. Można zatem spekulować, że biogenezę tych circRNA reguluje specyficzny tkankowo czynnik molekularny, który zmienia „wzór” splicingu.

CZYNNIKI UCZESTNICZĄCE W BIOGENEZIE CIRC RNA U *A. THALIANA*

Kolejnym etapem moich badań było poszukiwanie czynników potencjalnie zaangażowanych w biogenezę circRNA. Wcześniej pokazałam, że circRNA u roślin są generowane w sposób specyficzny dla tkanek/komórki, a ich poziomy akumulacji nie są skorelowane z poziomami odpowiadających im mRNA. Wymienione fakty pozwoliły na postawienie hipotezy, że oprócz czynników białkowych/RNA zaangażowanych w kanoniczny splicing, istnieją inne nieznanne czynniki, które promują wytwarzanie circRNA.

Aby zidentyfikować te czynniki, wykorzystałam ponownie *A. thaliana*. Roślina ta ma tysiące łatwo dostępnych mutantów, w tym dobrze scharakteryzowane nokauty pojedynczego genu. Jako roślinę referencyjną wybrałam typ dziki Col-0. Następnie dokonałam selekcji jej wariantów, tak by każdy nosił mutację typu knockout w genie kodującym białko biorące udział w splicingu na różnych jego etapach. Według Arabidopsis Splicing Related Data Base Genes (ASRG), 395 genów koduje białka związane ze splicingiem u *A. thaliana* (Wang et al. 2004). Zestaw mutantów został zawężony do tych z tłem Col-0, aby uzyskać wspólne odniesienie dla wszystkich analizowanych rośliny. Wykluczyłam również heterozygoty i mutacje letalne. Wstępna selekcja wariantów doprowadziła mnie do listy 86 ze zmianami w genach związanych ze splicingiem, 23 z nich miało dokładnie opisany fenotyp. W rezultacie, uzyskałam listę 18 mutantów *A. thaliana* wymienionych w Tabeli S1 z publikacji H2. Skład transkryptomów wszystkich tych roślin i Col-0 określono za pomocą RNA-seq, a identyfikację circRNA przeprowadzono przy pomocy opracowanego wcześniej protokołu.

Produkcja circRNA u trzech mutantów była znacząco różna od tej obserwowanej w typie dzikim i pozostałych mutantach. Pierwszy ze wspomnianych wariantów nie posiadał funkcjonalnego genu *cbp80*, którego produktem jest składnik kompleksu wiążącego kap na końcu 5' mRNA. Drugi mutant, *c2h2*, pozbawiony był genu kodującego białko U4/U6.U5 tri-snRNP – elementu wstępnie złożonego kompleksu spliceosomowego, który odgrywa kluczową rolę w tworzeniu katalitycznie aktywnego spliceosomu. Trzeci mutant, *flk*, nie posiadał genu produkującego białko z grupy jądrowych rybonukleoprotein (hnRNP), które, jeśli są związane z cząsteczkami pre-mRNA, sygnalizują, że pre-mRNA nie jest jeszcze w pełni dojrzała, a zatem nie jest gotowy do eksportu do cytoplazmy. Wspólną cechą tych trzech mutantów była zwiększona akumulacja kolistych transkryptów (w porównaniu z typem dzikim i innymi mutantami (Rysunek 2 z publikacji H2)), której nie towarzyszyły adekwatne zmiany w schemacie ekspresji genów. Zaobserwowana większa akumulacja circRNA mogła być spowodowana częstszym back-splicingiem lub zwiększoną stabilnością circRNA u tych mutantów. Dodatkowo, zaobserwowałam zwiększoną liczbę wystąpień alternatywnego splicingu w rzeczonych trzech mutantach, co sugeruje, że nokaut genu miał wpływ na splicing w ogóle.

Warto również nadmienić, że mutant *cbp80* wytwarza bardzo dużą liczbę (129) unikatowych circRNA (Rysunek 1 z publikacji H2). Spośród nich, tylko 11 było wspólnych z Col-0, w której to roślinie zidentyfikowałam w sumie 35 circRNA. Białko CBP80 wraz z CBP20 tworzy kompleks zaangażowany w splicing oraz biogenezę miRNA. Mutanty *A. thaliana* pozbawione tego kompleksu lub jego składników kumulują częściowo złożone transkrypty (Laubinger et al. 2008) i pri-miRNA (Kim et al. 2008). Ponadto braki w kompleksie wiążącym kap, a w szczególności brak CBP80, wpływają na powstawanie zjawiska alternatywnego splicing, szczególnie na końcu 5' pierwszego intronu (Raczynska et al. 2010). Moje badania po raz pierwszy dostarczyły dowodów na to, że brak CBP80 nie tylko zaburza produkcję liniowych izoform transkryptów, ale także dramatycznie zwiększa wytwarzanie różnych circRNA. Co ciekawe, większość circRNA specyficznych dla mutantu *cbp80* zawierała pierwszy ekson, podczas gdy circRNA powstałe w innych mutantach i Col-0 zostały wyprodukowane głównie z drugiego lub kolejnych eksonów (Rysunek 4 z publikacji H2). Ta obserwacja jest zgodna z wcześniejszymi odkryciami, że CBP80 preferencyjnie wywiera swój wpływ na splicing w mRNA w okolicy końca 5'.

Białko CBP80 bierze również udział w transdukcji sygnału kwasu abscysynowego (ABA) (Hugouvieux et al. 2002) i szlaku kwitnienia (Kuhn et al. 2007). Mutanty pozbawione CBP80 wykazują wywołane przez ABA podwyższenie cytozolowego poziomu wapnia w komórkach, prowadzące do lepszego zamykania aparatów szparkowych zapewniając przez to większą tolerancję na suszę (Hugouvieux et al. 2002). Te rośliny wykazują również fenotyp wczesnego kwitnienia, wynikający z wadliwego splicingu intronów FLOWING LOCUS C (FLC). Niestety nie wiadomo czy circRNA generowane w mutantach *cbp80* również przyczyniają się do zaobserwowanych fenotypów. Repertuar circRNA specyficznych dla nokautów *cbp80* musi zostać zbadany bardziej szczegółowo w celu przedstawienia ostatecznych wniosków dotyczących tego aspektu. CircRNA mogą być generowane w sposób kontrolowany lub stanowić produkty nieprawidłowego splicingu.

Na podstawie moich badań można postawić hipotezę, że ogólnie niezakłócone funkcjonowanie czynników białkowych zaangażowany w splicing sprzyja tworzeniu liniowych transkryptów. Niemniej jednak otrzymane wyniki sugerują, że zmiana kompleksów białkowych związanych ze splicingiem może być łatwym mechanizmem uruchamiającym/wzmacniającym produkcję circRNA. Dynamiczny skład kompleksu białek wiążących kap w *A. thaliana* jest dobrze udokumentowany i zmienia się w trakcie wzrostu rośliny (Bush et al. 2009). CBP80 jest jednym z białek, które pozostają związane z kapem mRNA zarówno w komórkach proliferujących, jak i w spoczynku, ale ulegają zmianie lokalizacja subkomórkowej. Na wczesnych etapach cyklu wzrostu białko to lokalizuje się preferencyjnie w cytoplazmie, a później gromadzi się w jądrze. Z kolei jego partner, CBP20, najpierw lokalizuje się zarówno w jądrze, jak i w cytoplazmie, a następnie koncentruje się w jądrze (Bush et al. 2009). Można zatem postawić hipotezę, że niektóre wystąpienia niekanonicznego splicingu na początku cyklu wzrostu rośliny mogą być spowodowane niedoborem CBP80. Takie zjawisko potencjalnie może aktywować tworzenie circRNA. Idąc dalej, kontrola występowania różnorodnych czynników związanych ze splicingiem może stanowić potencjalny mechanizm regulacji ekspresji genów w roślinach. Co ciekawe, pokazano już, że ekspresja circRNA wzrasta, gdy działanie maszyny odpowiedzialnej za splicing pre-mRNA u *Drosophila* jest upośledzone (Liang et al. 2017). Sugeruje to konserwację takiego mechanizmu w różnych królestwach.

BAZA DANYCH AT-C-RNA

Opracowaną w pracy H1 metodę analizy danych RNA-seq do identyfikacji i oceny ilościowej circRNA w roślinach, wykorzystałam do ponownego przeanalizowania wszystkich spełniających kryteria opracowanego protokołu publicznie dostępnych zbiorów danych RNA-seq dla *A. thaliana*. Chociaż circRNA u *A. thaliana* scharakteryzowano w wielu badaniach, porównanie ich wyników wykazuje wyraźne rozbieżności, a ich analiza porównawcza nie była możliwa. Główną przyczyną takiej sytuacji był brak standaryzacji metod stosowanych do analiz circRNA. Aby zmienić ten stan rzeczy zdecydowałam, że ujednolicone wyniki otrzymane z analiz publicznych danych zostaną udostępnione online szerszemu gronu odbiorców. W efekcie powstała baza danych At-C-RNA (<http://plantcircrna.ibch.poznan.pl/>; publikacja H3), zaimplementowana przez moją doktorantkę, panią Katarzynę Nowis, która pozwala na analizę i porównanie danych o circRNA z różnych źródeł, co wcześniej nie było możliwe. At-C-RNA to obecnie jedyne źródło, w którym dane z różnych badań zostały ponownie przeanalizowane przy pomocy tej samej metody, ustandaryzowane i ujednolicone. Co więcej, At-C-RNA jest jedyną bazą danych, która dostarcza informacji o występowaniu circRNA zarówno w dzikich, jak i zmutowanych roślinach *A. thaliana* i ich tkankach. Baza została tak zaprojektowana, aby jej interfejs był jak najbardziej przyjazny użytkownikowi. Informacje o circRNA zaprezentowane są w interaktywnej tabeli, która umożliwia użytkownikowi zdefiniowanie kryteriów przeglądania. Możliwe jest sortowanie i filtrowanie danych podług każdej z kolumn. Istnieje również możliwość ręcznego usunięcia wybranych circRNA z tabeli. Co więcej, użytkownicy mogą filtrować tabelę, klikając interaktywne wykresy znajdujące się pod nią.

Obecnie w At-C-RNA zdeponowanych jest 113 327 circRNA. Warto zauważyć, że tylko 16,2%, 18,9% i 19,7% z nich odnotowano wcześniej w bazach poświęconych

roślinnym circRNA, kolejno w AtCircDB (Ye et al. 2017), PlantcircNet (Zhang et al. 2017) i PlantcircBase (Chu et al. 2017) (dane z 02.09.2021). Każdemu circRNA znajdującemu się w bazie przypisany został unikalny identyfikator zgodnie z ogólnie przyjętym wzorcem: AT_numer_chromosomu:circRNA_start-circRNA_stop, co umożliwia wygodne porównywanie wyników z innych baz danych/badań. Na temat każdego circRNA dostępne są następujące informacje: (i) publikacja, w ramach której wygenerowano zestawy danych RNA-seq, (ii) linia rośliny i tkanka, w której zidentyfikowano circRNA, (iii) średnia punktacja At-C-RNA obliczona dla wszystkich circRNA zgodnie z tą samą procedurą, która wyraża poziom akumulacji circRNA (iv) punktacja At-C-RNA poszczególnych circRNA wchodzących w skład punktacji średniej, (v) informacja, czy circRNA został potwierdzony eksperymentalnie, np. przy użyciu RNazy R, ddPCR (vi) informacja na temat „powtarzalności” circRNA. Ostatnie kryterium jest szczególnie istotne, gdyż we wcześniejszych badaniach wykazałam, że większość circRNA produkowana jest w sposób spontaniczny i niepowtarzalny w *A. thaliana* (i prawdopodobnie w innych roślinach również) a zatem frakcja tych cząsteczek najpewniej nie pełni funkcji biologicznej. Tylko circRNA produkowane w sposób powtarzalny mogą mieć potencjał funkcjonalny, a co za tym idzie są szczególnie ważnymi obiektami dla szerokiego zakresu badań.

W At-C-DB zdefiniowano, że circRNA jest powtarzalny, jeśli został zidentyfikowany w co najmniej 4 biologicznych powtórzeniach w określonej linii rośliny i tkance w ramach badania. W sumie 655 circRNA sklasyfikowano jako powtarzalne. Spośród nich 226 zidentyfikowano we wszystkich analizowanych tkankach roślinnych (kwiecie, liściu, korzeniu i siewce) i w całej roślinie. Największą liczbę unikatowych, powtarzalnych circRNA (35) znaleziono w liściu. Natomiast dwa inne organy, korzeń i kwiat, charakteryzowały się odpowiednio tylko dwoma i jednym circRNA typowym dla danej tkanki. W siewce nie zidentyfikowano żadnego unikatowego circRNA.

Większość genów (362) dających początek powtarzalnym circRNA wytwarza ponad pięć izoform circRNA, a tylko 11 genów wytwarza jedną izoformę. Punktacja At-C-DB większości powtarzalnych circRNA (91,6%) mieści się w zakresie 1-15, co odpowiada raczej niskiej liczebności. 55 circRNA (8,4%) przekracza średnią ocenę powyżej 15. Rozkład powtarzalnych circRNA na chromosomach wskazuje, że większość z nich pochodzi z genów zlokalizowanych na chromosomie 1, co jest w zgodzie z obserwacjami dokonanymi przeze mnie we wcześniejszej pracy (Rysunek 3 D z publikacji H1). Szeroko zakrojona analiza danych pochodzących z różnych źródeł pozwoliła na ujawnienie pełnego obrazu rozkładu circRNA w *A. thaliana* i jej tkankach, co wcześniej było ograniczone do pojedynczego eksperymentu.

CHARAKTERYSTYKA CIRC RNA W MODELU LUDZKIM, DM1.

Dodatkowym aspektem moich badań, wskazującym na generalne znaczenie wypracowanych rozwiązań analitycznych, było zaadaptowanie wcześniej opracowanej metody analizy circRNA u roślin do badania circRNA u ludzi chorych na dystrofię miotoniczną typu 1 (DM1), publikacja H4. Mutacją powodującą DM1 jest ekspansja powtórzeń CTG w regionie 3' UTR genu kinazy DMPK (Brook et al. 1992). Osoby zdrowe mają od 5 do ~34 powtórzeń, podczas gdy u pacjentów z DM1 powtórzenia sięgają często do setek, a nawet tysięcy kopii

(Brook et al. 1992). Obecność tych zmutowanych transkryptów powoduje sekwestracja białek typu muscleblind (MBNL), które regulują alternatywny splicing (Konieczny et al. 2014). Co ważne, Ashwal Fluss i wsp. (Ashwal-Fluss et al. 2014) pokazali, że białka MBNL mogą odgrywać też ważną rolę w biogenezie circRNA. Biorąc pod uwagę powyższe fakty, że białka MBNL biorą udział w biogenezie circRNA, a ich funkcjonalność jest znacząco ograniczona w DM1 (Miller et al. 2000) postawiono hipotezę, że poziom przynajmniej niektórych circRNA będzie obniżony w DM1.

Aby przetestować wysuniętą hipotezę, wykorzystałam publicznie dostępne zestawy danych RNA-Seq zdeponowane w bazie danych DMseq (<http://www.dmseq.org/>) (Wang et al. 2019). Do analizy wybrałam tylko dane, które zostały wygenerowane w ramach tej samej procedury, gdzie biblioteki przygotowano z całkowitego RNA pozbawionego rRNA, i zastosowano sekwencjonowanie Illumina typu pair-end. Ostatecznie przebrałam dwa zestawy danych najliczniej reprezentowane w bazie. Próbkę były uzyskane z dwóch różnych mięśni szkieletowych, czworogłowego uda (ang. *quadriceps femoris*, QF; 23 próbki, 11 próbek kontrolnych i 12 próbek DM1) i piszczelowego przedniego (ang. *tibialis anterior*, TA; 27 próbek, 6 próbek kontrolnych i 21 próbek DM1). Zgodnie z wynikami uzyskanymi wcześniej, poziomy akumulacji circRNA znormalizowałam względem liczby odczytów mapujących do genów *ACTB* i *GAPDH*.

W sumie w próbkach QF wykryłam 22 816 różnych circRNA, z czego 4 168 (18%) sklasyfikowałam jako powtarzalne, potwierdzone przez co najmniej pięć odczytów typu back-splice w co najmniej dwóch próbkach, a 152 (0,7%) sklasyfikowałam jako powszechne, obecne we wszystkich lub wszystkich oprócz jednej próbkach kontrolnych/DM1. W przypadku próbek TA łącznie zidentyfikowałam 38 403 circRNA, z czego 7 537 (20%) powtarzalnych i 403 (1%) powszechnych. Porównanie globalnego poziomu akumulacji wszystkich circRNA w kontroli i w próbkach DM1, ujawniło istotnie zwiększony poziom circRNA w DM1 ($p = 0,004$ w QF i $p < 0,0001$ w TA), Rysunek 3 z publikacji H4. Co ważne, podobne różnice nie zostały zaobserwowane dla ich liniowych odpowiedników ($p = 0,6$ i $p = 0,1$ odpowiednio w QF i TA). Zwiększony poziom circRNA w próbkach DM1 był również widoczny w grupach „powszechnych” circRNA. Aby wykluczyć, że obserwowane zmiany poziomu circRNA są wynikiem wzrostu lub spadku transkrypcji liniowych prekursorów, znormalizowałam również poziom circRNA w stosunku do odpowiadających im liniowych transkryptów. Ponownie, wyniki wykazały, wyższe poziomy circRNA w próbkach DM1 niż w próbkach kontrolnych (prawe wykresy na Rysunku 3 z publikacji H4). Dodatkowo, przy użyciu wygenerowanych danych, przeanalizowałam różnicową akumulację circRNA w DM1. Analiza tą ograniczałam tylko do zbiorów circRNA występujących powszechnie we wszystkich lub wszystkich próbkach kontrolnych/DM1 ($n = 152$ w QF i $n = 403$ w TA). Zarówno w QF, jak i TA, zmiany w poziomach circRNA obliczone na podstawie danych normalizowanych obydwiema metodami były wysoce skorelowane, wskazując, że zmiany poziomu circRNA nie zależą od poziomu ekspresji genów, z których są generowane. Jak pokazano na Rysunku 4 z publikacji H4, wartości zmian akumulacji są znacznie przesunięte w kierunku wartości dodatnich, wskazujący na więcej circRNA o podwyższonym poziomie

w próbkach DM1. Ten efekt jest zgodny z zaobserwowanym wcześniej globalnym wzrostem poziomu circRNA w DM1 (zarówno w QF, jak i TA).

Przeprowadzona analiza wykazała, że globalny poziom circRNA w DM1 jest podwyższony. Dzięki zastosowanemu protokołowi analizy danych, a w szczególności, dzięki zastosowaniu różnych metod normalizacji (względem liczby odczytów mapujących do genów *ACTB* i *GAPDH*; względem globalnej liczby mapujących odczytów; oraz względem liczby odczytów mapujących do liniowych odpowiedników circRNA) udało mi się wykazać, że wzrost poziomu circRNA w DM1 nie jest wynikiem zmian globalnego poziomu transkrypcji czy podwyższenia ekspresji ich liniowych odpowiedników. Otrzymane w niniejszym badaniu wyniki tym samym poddają w wątpliwość rolę MBNL jako czynnika odgrywającego istotną rolę w biogenezie circRNA u człowieka. Warto jednak zauważyć, że podwyższony globalny poziom circRNA jak i podwyższony poziom indywidualnych circRNA w DM1 okazał się dobrze korelować z ostrością przebiegu choroby co implikuje potencjalne wykorzystanie circRNA jako biomarkerów w DM1.

POTENCJAŁ

Opracowana metoda analizy danych RNA-seq do badań circRNA u *A. thaliana* może zostać w łatwy sposób rozszerzona na badania circRNA w innych układach biologicznych.

Zidentyfikowane powtarzalne circRNA które są specyficzne tkankowo jawią się jako atrakcyjni kandydaci do badań funkcjonalnych. Moje wyniki mogą prowadzić do dalszych badań nad circRNA specyficznymi dla narządów czy tkanek, a w szczególności nad ich funkcją. Ponadto, mogą być punktem wyjściowym do prac nad odkryciem mechanizmów, które kierują przełączaniem między tworzeniem się kolistych i liniowych transkryptów.

Użycie mutantów *A. thaliana* o zmienionych fenotypach, może w przyszłości ułatwić kolejne badania funkcjonalne circRNA. Na przykład w przypadku niedoboru circRNA u mutantu typu knockout możliwe byłoby wprowadzenie go sztucznie i natychmiastowe obserwowanie jego efektów fenotypowych.

Identyfikacja nieznanymi czynnikami zaangażowanymi w splicing i wspierających produkcję circRNA ma ogromne znaczenie dla zrozumienia mechanizmów leżących u podstaw niekanonicznego łączenia eksonów. Otrzymane wyniki mogą przyczynić się do lepszego zrozumienia i identyfikacji nowych mechanizmów regulacji ekspresji genów.

Baza danych która pozwala na analizę i porównanie danych o circRNA z różnych źródeł, pozwoli naukowcom z całego świata na wykorzystanie wyników badań wygenerowanych przez inne grupy, co wcześniej nie było możliwe. At-C-RNA to obecnie jedyne źródło, w którym dane z różnych eksperymentów zostały ponownie przeanalizowane, ustandaryzowane i ujednolicone.

Badanie wpływu niedoboru czynnika uważanego za biorącego udział w biogenezie niektórych circRNA niewątpliwie przyczyni się do dalszych badań i weryfikacji jego wpływu na produkcję circRNA, a także na weryfikację hipotezy powstawania tych cząsteczek przy udziale białek z rodziny MBNL.

4. OMÓWIENIE POZOSTAŁYCH OSIĄGNIĘĆ NAUKOWO-BADAWCZYCH

Na przestrzeni mojej pracy naukowej, moje badania dotyczyły zastosowania i rozwoju narzędzi bioinformatycznych w celu poznania i zrozumienia zjawisk i procesów biologicznych.

W przeszłości pracowałam nad stworzeniem metod komputerowych, które pozwalałyby na przewidzenie miejsca wiązania jonów metali i ligandów w strukturach białek i RNA. Podczas studiów magisterskich na wydziale Fizyki UAM, pod kierunkiem pana dr. Leszka Rychlewskiego, stworzyłam meta-metodę oceniającą różne konformacje przestrzenne kompleksu białko-ligand na podstawie cząstkowych ocen uzyskanych z sześciu znanych programów do dokowania. Ocena opracowanej meta-metody, nazwanej HarmonyDock dawała lepsze wyniki (mniejsze RMSD do struktury natywnej) aniżeli każdy z programów dokujących zastosowany osobno. Wyniki moich badań zostały zawarte w publikacji N14. Badania nad strukturami biomolekuł, kontynuowałam podczas studiów doktoranckich na wydziale Biologii UAM, pod kierunkiem pana prof. dr. hab. Janusza Bujnickiego, między innymi w ramach grantu PRELUDIUM NCN, którego byłam kierownikiem. Opracowałam wtedy anizotropowy potencjał statystyczny, który wyróżniał się wysoką wiarygodnością dzięki zastosowaniu nie tylko parametru odległości między atomami, ale także parametru orientacji przestrzennej (kąta). Potencjał ten następnie zaimplementowałam w dwóch programach. MetalionRNA jest metodą służącą do przewidywania miejsc wiązania jonów metali w strukturach RNA, natomiast LigandRNA pozwala na ocenę kompleksów RNA-ligand. Opracowanie tych narzędzi, pozwoliło na badanie zależności między RNA a oddziałującymi z nim cząsteczkami. Narzędzia mogą zostać wykorzystane do identyfikacji potencjalnych leków np. antybiotyków. Badania prowadzone przeze mnie w trakcie doktoratu zaowocowały czterema publikacjami w renomowanych czasopismach naukowych (N10, N11, N12, N13) oraz pracą doktorską, którą obroniłam z wyróżnieniem i została ona uznana przez Polskie Towarzystwo Bioinformatyczne jako najlepszy doktorat z bioinformatyki w 2013 roku.

Następnie, podczas pracy w ICHB PAN głównym tematem moich badań były koliste cząsteczki RNA w roślinach. Na te prace przyznano mi grant SONATA NCN, a jej wyniki, w postaci cyklu publikacji, stanowią moje osiągnięcie naukowe prezentowane w niniejszym wniosku. Pokróctce, początkowo zajęłam się identyfikacją circRNA u *A. thaliana*, gdyż w czasie, gdy zaczynałam swoje badania cząsteczki te były bardzo słabo poznane i opisane w roślinach. Opracowałam wtedy metodę do identyfikacji i standaryzacji danych circRNA, która później posłużyła do ponownego przeanalizowania publicznie dostępnych zbiorów danych. W wyniku tego powstała baza danych (<http://plantcircrna.ibch.poznan.pl/>), która umożliwia analizę i porównanie danych o circRNA z różnych źródeł, co wcześniej nie było możliwe. Wykazałam też, że większość circRNA u *A. thaliana* jest produkowana w sposób losowy i zaproponowałam metodę umożliwiającą oddzielenie tych circRNA od takich produkowanych w sposób powtarzalny, gdyż tylko takie cząsteczki mogą pełnić biologiczne funkcje. Następnie zajęłam się identyfikacją czynników genetycznych uczestniczących w biogenezie circRNA. Wykazałam, że mutanty rośliny pozbawione genów *cbp80*, *c2h2*, i *flk* akumulują więcej circRNA, co może wskazywać na udział produktów tych genów w regulacji ekspresji circRNA. Pokazałam też, że niedobór czynnika uważanego za promotora powstawania circRNA, białka z rodziny muscleblind, nie doprowadził do zmniejszenia produkcji circRNA w warunkach

in vivo, tj. w tkankach mięśni ludzi chorych na dystrofię miotoniczną typu 1. Wiarygodna identyfikacja i oznaczenie ilościowe circRNA u roślin oraz identyfikacja czynników zaangażowanych w ich biogenezę przyczyni się do właściwego zrozumienia uniwersalnych zasad, które rządzą powstawaniem tych RNA w różnych królestwach, oraz znaczenia circRNA w szerokim kontekście ewolucyjnym.

Równolegle z badaniami nad circRNA rozwijałam swoje zainteresowania tematyką mikrobiomów. Początkowo zajmowałam się badaniem składu bakterii w kopalnych szczątkach ludzkich i pokazałam, że poza występującymi powszechnie bakteriami środowiskowymi, możliwa jest identyfikacja kopalnych szczepów, typowych dla jamy ustnej człowieka. Opracowałam metodę identyfikacji tych bakterii, a także walidacji ich wieku za pomocą analiz modyfikacji i zniszczeń DNA. Wykazałam, że czas przechowywania szczątków w magazynie, wiek prób czy stanowisko archeologiczne na jakim zostały pobrane nie mają wpływu na ilość kopalnych bakterii zachowanych w próbie. Badania porównawcze starożytnych i współczesnych genomów bakterii dostarczają ważnych informacji na temat pochodzenia i ewolucji bakterii, zwłaszcza w kontekście czynników wirulencji i adaptacji do organizmu żywiciela w wyniku zmiany diety i nawyków higienicznych. Wyniki tych badań były przedmiotem trzech publikacji (N5, N7, N9) w renomowanych czasopismach naukowych oraz zaowocowały współautorstwem w dwóch monografiach (N1, N2). W tamtym czasie nawiązałam też współpracę z grupą z Uniwersytetu w Louisville, która specjalizuje się w badaniu patogenów jamy ustnej. W wyniku tej współpracy zrekonstruowałam cztery kopalne genomy bakterii *Tannerella forsythia* i porównałam je ze genomami współczesnych *T. forsythia*. Ujawniło to istotne różnice w sekwencjach niektórych genów związanych z wirulencją tej bakterii (publikacja N5).

Obecnie, w mojej samodzielnej pracy badawczej jako kierownik Pracowni Bioinformatyki w ICHB PAN dalej rozwijam badania nad mikrobiomami. Wspólnie z partnerem biznesowym, firmą Ardigen SA, otrzymałam grant NCBR „Mapa mikrobiomu Polski” na stworzenie referencyjnego zbioru mikrobiomów przewodu pokarmowego Polaków. Zbiór ten będzie punktem wyjścia do wielu badań nad relacją między składem bakteryjnym jelit a stylem życia, dietą, chorobami, a także odpowiedzią pacjentów na immunoterapię, gdyż projekt zakłada również przebadanie grupy pacjentów onkologicznych. Powstanie tego zbioru danych i stworzenie sieci korelacji pomiędzy mikrobiomem a szeregiem cech fenotypowych, stylem życia jaki dana osoba prowadzi, czy też chorobami, a w szczególności skutecznością terapii immunologicznej otwiera nowy rozdział w badaniach nad zrozumieniem mechanizmów wpływu mikrobiomu jelit na nasz organizm. Dotychczas badania zaowocowały zgłoszeniem patentowym P1 oraz dwoma publikacjami, z których jedna znajduje się w recenzji czasopisma *Scientific Reports*, a druga jest w przygotowaniu.

W ramach współpracy z fizykami, chemikami, biologami i lekarzami w kraju i zagranicą, zaangażowana jestem dodatkowo w kilka projektów badawczych o interdyscyplinarnym charakterze, w których służę wiedzą i doświadczeniem w wyborze, projektowaniu i interpretacji wyników analiz bioinformatycznych. Przykładowo, w wyniku współpracy z grupą z Wydziału Biologii Uniwersytetu im. A. Mickiewicza w Poznaniu powstały dwie publikacje (N4, N6), w których przeanalizowałam skład bakteryjny ścieków

przed, w trakcie i po oczyszczeniu. Dzięki temu badaniu udało się wykazać istotne różnice w analizowanych grupach. W szczególności prace dowiodły, że oczyszczanie ścieków zmniejsza znacząco liczbę bakterii, ale równocześnie w ściekach po oczyszczeniu znajduje się więcej szczepów antybiotykoopornych. Oczyszczalnie sprzyjają więc transmisji genów związanych z antybiotykoopornością, co jest zjawiskiem bardzo niekorzystnym, niesie bowiem ze sobą duże zagrożenie dla zdrowia i życia ludzi.

V. INFORMACJA O WYKAZYWANIU SIĘ ISTOTNĄ AKTYWNOŚCIĄ NAUKOWĄ ALBO ARTYSTYCZNĄ REALIZOWANĄ W WIĘCEJ NIŻ JEDNEJ UCZELNI, INSTYTUCJI NAUKOWEJ LUB INSTYTUCJI KULTURY, W SZCZEGÓLNOŚCI ZAGRANICZNEJ.

Dorobek naukowy jaki posiadam wiąże się z moją pracą w trzech jednostkach badawczych.

1. Uniwersytet im. Adama Mickiewicza w Poznaniu, na którym ukończyłam studia doktoranckie i uzyskałam tytuł doktora w 2013 roku. Podczas swojej pracy przyznano mi grant PRELUDIUM NCN.

2. Międzynarodowy Instytut Biologii Molekularnej i Komórkowej w Warszawie, gdzie w latach 2010-2012 byłam zatrudniona w granie TEAM prof. dr. hab. J. Bujnickiego. Zajmowałam się modelowaniem oddziaływać RNA-ligand.

3. Po uzyskaniu stopnia doktora, od 2014 roku pracuję w ICHB PAN, gdzie realizuję własne projekty oraz projekty w ramach krajowych i międzynarodowych współprac. W szczególności:

- Od 2018 roku współpracuję z grupą pana prof. Jana Potempy z Department of Oral Immunity and Infectious Diseases, University of Louisville School of Dentistry. Jest to stała współpraca eksperymentalna i koncepcyjna. Dotyczy głównie badań mikrobiomu jamy ustnej, ze szczególnym uwzględnieniem ewolucji czynników patogennych. Współpraca zaowocowała już publikacją naukową w renomowanym czasopiśmie międzynarodowym N5. Planowany był także mój staż w grupie pana prof. Potempy w celu kontynuacji badań, jednak ze względu na pandemię COVID-19 jest stale przekładany na okres późniejszy.
- Jako wykonawca interdyscyplinarnego grantu SYMFONIA NCN (kierownik prof. dr. hab. M. Figlerowicz, ICHB PAN – lider) współpracowałam z pozostałymi jednostkami wchodzącymi w skład konsorcjum realizującego grant, tj. Instytutem Antropologii UAM oraz grupą prof. Jasińskiego z Wydziału Historii UAM. Prowadzone badania zaowocowały szeregiem wspólnych publikacji (N1, N2, N5, N7, N9), wystąpień konferencyjnych oraz wspólną organizacją konferencji naukowych. We wszystkich wymienionych aktywnościach brałam czynny udział. Mój wkład w realizację grantu polegał na zastosowaniu nowoczesnych metod obliczeniowych do uzyskania odpowiedzi na zagadnienia historyczne. Np. badania kopalnego DNA pozwoliły na odpowiedź na pytanie o pochodzenie genetyczne naszej populacji. Zastosowane metody

wniosły nową jakość do badań antropologicznych i historycznych na polskich uczelniach.

- Od 2019 roku współpracuję stale z grupą pani prof. dr hab. Joanny Mokrackiej (Wydział Biologii UAM). W wyniku tej współpracy powstały dotychczas dwie publikacje (N4 i N6), w których przeanalizowałam skład bakteryjny ścieków przed, w trakcie i po oczyszczaniu. Warto nadmienić, że pani dr Nicoletta Makowska, pierwsza autorka obu prac po obronie swojej pracy doktorskiej na UAM uzyskała grant SONATINA NCN, który realizuje w moim zespole.
- Ponadto pełniłam funkcję promotora dwóch prac magisterskich na UAM i PP oraz recenzenta pracy licencjackiej na PP.

VI. INFORMACJA O OSIĄGNIĘCIACH DYDAKTYCZNYCH, ORGANIZACYJNYCH ORAZ POPULARYZUJĄCYCH NAUKĘ LUB SZTUKĘ.

1. OSIĄGNIĘCIA DYDAKTYCZNE

Po uzyskaniu stopnia doktora:

- Do tej pory pod moją opieką znajdowało się dwoje doktorantów – pani Katarzyna Nowis i pan Michał Stelmaszczuk. Osoby te były zatrudnione w kierowanym przeze mnie grantie SONATA NCN mającym na celu badanie kolistych RNA u *A. thaliana*. Po uzyskaniu stopnia doktora habilitowanego chciałabym zostać ich promotorem.
- Byłam promotorem dwóch prac magisterskich:
 - Bogny Kuczkowskiej, temat pracy: „*Identyfikacja mikroorganizmów towarzyszących kopalnym materiałom kostnym*”, UAM, kierunek Bioinformatyka, 2016. Ocena: bardzo dobra. Współpraca ta zaowocowała wspólną publikacją w *GigaScience*, N9.
 - Kai Chmielewskiej, temat pracy: „*Porównywanie dostępnych metod do identyfikacji kolistego RNA na podstawie danych z RNA-Seq*” PP, kierunek Bioinformatyka, 2016. Ocena: celująca.

Za szczególne osiągnięcie dydaktyczne uważam fakt, że obie magistrantki, po obronie prac dyplomowych, zdecydowały się kontynuować naukę na studiach doktoranckich.

- Recenzowałam także licencjat: Marek Justa, PP kierunek Bioinformatyka, 2016.
- Opiekowałam się dwoma stażystami.

Podczas doktoratu:

- Opiekowałam się dwoma projektami licencjackimi:
 - Krzysztofa Formanowicza, UAM kierunek Bioinformatyka, 2011.
 - Fryderyka Kozłowskiego, UAM kierunek Bioinformatyka, 2011.

2. OSIĄGNIĘCIA ORGANIZACYJNE

- Od 2020 roku jestem kierownikiem Pracowni Bioinformatyki w ICHB PAN.
- Kierowałam projektami finansowanymi ze źródeł zewnętrznych.
 - Po uzyskaniu stopnia doktora grantem NCN SONATA oraz grantem strukturalnym NCBR (grant w konsorcjum z Ardigen SA).
 - Przed uzyskaniem stopnia doktora grantem NCN PRELUDIUM.
- W 2021 roku byłam członkiem zespołów przygotowujących materiały do oceny instytutów PAN i ewaluacji jednostek naukowych, prowadzonej przez MNiSW w celu kategoryzacji.
- W 2018 roku byłam członkiem komisji oceniającej pracowników ICHB PAN.
- Współorganizowałam cztery konferencje:
 - „*Fascinating World of Bioorganic Chemistry*”, konferencja międzynarodowa dedykowana prof. A. Legockiemu z okazji 80 lecia, organizowana przez ICHB PAN, Poznań, 12-13.11.2019.
 - „*Piast & Přemyslovci Meeting*”, Poznań, 29-30.05.2018, konferencja międzynarodowa organizowana przez ICHB PAN, UAM (Wydział Historyczny) oraz Instytut Archeologii Czeskiej Akademii Nauk w Pradze.
 - „*DNA – język życia*”, Poznań, 27–28.10.2016, konferencja krajowa organizowana przez ICHB PAN i PP.
 - „*Tradycje i nowoczesność - początki państwa polskiego na tle środkowoeuropejskim w badaniach interdyscyplinarnych*”, Poznań, 10-12.06.2015 - konferencja międzynarodowa, organizowana przez Instytut Historii i Instytut Prahistorii UAM oraz ICHB PAN.

VII. OPRÓCZ KWESTII WYMIENIONYCH W PKT. 1-6, WNIOSKODAWCA MOŻE PODAĆ INNE INFORMACJE, WAŻNE Z JEGO PUNKTU WIDZENIA, DOTYCZĄCE JEGO KARIERY ZAWODOWEJ.

1. NAGRODY, WYRÓŻNIENIA I STYPENDIA

- Stypendium Ministra Edukacji i Nauki dla wybitnych młodych naukowców, 2021.
- Nagroda ICHB PAN za najlepszą publikację eksperymentalną w 2019 (N9).
- Nagroda Polskiego Towarzystwa Bioinformatycznego za najlepszą pracę doktorską z bioinformatyki, 2014.
- Obrona pracy doktorskiej z wyróżnieniem, 2013.
- Stypendium dla doktorantów Wojewódzkiego Urzędu Pracy w Poznaniu 2013/2014.

- Zwiększenie stypendium doktoranckiego z dotacji projakościowej UAM, lata 2011-2013.
- Stypendium naukowe Wydziału Biologii UAM (studia doktoranckie), lata 2010-2013.
- Nagroda Dziekana Wydziału Biologii UAM za publikacje, 2012.
- Nagroda absolutoryjna Dziekana Wydziału Fizyki UAM za bardzo dobre wyniki w nauce, 2008.
- Stypendium naukowe Wydziału Fizyki UAM (studia magisterskie), lata 2004-2008.

VIII. REFERENCJE

Aryamanesh, N., H. Ruwe, L. V. Sanglard, L. Eshraghi, J. D. Bussell, K. A. Howell, I. Small and C. C. des Francs-Small (2017). "The Pentatricopeptide Repeat Protein EMB2654 Is Essential for Trans-Splicing of a Chloroplast Small Ribosomal Subunit Transcript." *Plant Physiol* 173(2): 1164-1176.

Ashwal-Fluss, R., M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evtantal, S. Memczak, N. Rajewsky and S. Kadener (2014). "circRNA biogenesis competes with pre-mRNA splicing." *Mol Cell* 56(1): 55-66.

Brook, J. D., M. E. McCurrach, H. G. Harley, A. J. Buckler, D. Church, H. Aburatani, K. Hunter, V. P. Stanton, J. P. Thirion, T. Hudson and et al. (1992). "Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member." *Cell* 69(2): 385.

Bush, M. S., A. P. Hutchins, A. M. Jones, M. J. Naldrett, A. Jarmolowski, C. W. Lloyd and J. H. Doonan (2009). "Selective recruitment of proteins to 5' cap complexes during the growth cycle in Arabidopsis." *Plant J* 59(3): 400-412.

Chen, G., J. Cui, L. Wang, Y. Zhu, Z. Lu and B. Jin (2017). "Genome-Wide Identification of Circular RNAs in Arabidopsis thaliana." *Front Plant Sci* 8: 1678.

Chen, L., P. Zhang, Y. Fan, Q. Lu, Q. Li, J. Yan, G. J. Muehlbauer, P. S. Schnable, M. Dai and L. Li (2018). "Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize." *New Phytol* 217(3): 1292-1306.

Cheng, J., Y. Zhang, Z. Li, T. Wang, X. Zhang and B. Zheng (2018). "A lariat-derived circular RNA is required for plant development in Arabidopsis." *Sci China Life Sci* 61(2): 204-213.

Chu, Q., X. Zhang, X. Zhu, C. Liu, L. Mao, C. Ye, Q. H. Zhu and L. Fan (2017). "PlantcircBase: A Database for Plant Circular RNAs." *Mol Plant* 10(8): 1126-1128.

Conn, V. M., V. Hugouvieux, A. Nayak, S. A. Conos, G. Capovilla, G. Cildir, A. Jourdain, V. Tergaonkar, M. Schmid, C. Zubieta and S. J. Conn (2017). "A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation." *Nat Plants* 3: 17053.

- Darbani, B., S. Noeparvar and S. Borg (2016). "Identification of Circular RNAs from the Parental Genes Involved in Multiple Aspects of Cellular Metabolism in Barley." *Front Plant Sci* 7: 776.
- Gao, Y., J. Zhang and F. Zhao (2017). "Circular RNA identification based on multiple seed matching." *Brief Bioinform.*
- Hansen, T. B., T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen, C. K. Damgaard and J. Kjems (2013). "Natural RNA circles function as efficient microRNA sponges." *Nature* 495(7441): 384-388.
- Hugouvieux, V., Y. Murata, J. J. Young, J. M. Kwak, D. Z. Mackesy and J. I. Schroeder (2002). "Localization, ion channel regulation, and genetic interactions during abscisic acid signaling of the nuclear mRNA cap-binding protein, ABH1." *Plant Physiol* 130(3): 1276-1287.
- Kim, S., J. Y. Yang, J. Xu, I. C. Jang, M. J. Prigge and N. H. Chua (2008). "Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs." *Plant Cell Physiol* 49(11): 1634-1644.
- Konieczny, P., E. Stepniak-Konieczna and K. Sobczak (2014). "MBNL proteins and their target RNAs, interaction and splicing regulation." *Nucleic Acids Res* 42(17): 10873-10887.
- Kuhn, J. M., G. Breton and J. I. Schroeder (2007). "mRNA metabolism of flowering-time regulators in wild-type Arabidopsis revealed by a nuclear cap binding protein mutant, abh1." *Plant J* 50(6): 1049-1062.
- Laubinger, S., T. Sachsenberg, G. Zeller, W. Busch, J. U. Lohmann, G. Ratsch and D. Weigel (2008). "Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana." *Proc Natl Acad Sci U S A* 105(25): 8795-8800.
- Liang, D., D. C. Tatomer, Z. Luo, H. Wu, L. Yang, L. L. Chen, S. Cherry and J. E. Wilusz (2017). "The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting." *Mol Cell* 68(5): 940-954 e943.
- Liu, J., D. Li, H. Luo and X. Zhu (2019). "Circular RNAs: The star molecules in cancer." *Mol Aspects Med* 70: 141-152.
- Liu, T., L. Zhang, G. Chen and T. Shi (2017). "Identifying and Characterizing the Circular RNAs during the Lifespan of Arabidopsis Leaves." *Front Plant Sci* 8: 1278.
- Lu, T., L. Cui, Y. Zhou, C. Zhu, D. Fan, H. Gong, Q. Zhao, C. Zhou, Y. Zhao, D. Lu, J. Luo, Y. Wang, Q. Tian, Q. Feng, T. Huang and B. Han (2015). "Transcriptome-wide investigation of circular RNAs in rice." *RNA* 21(12): 2076-2087.
- Memczak, S., M. Jens, A. Elefsinioti, F. Torti, J. Krueger, A. Rybak, L. Maier, S. D. Mackowiak, L. H. Gregersen, M. Munschauer, A. Loewer, U. Ziebold, M. Landthaler, C.

- Kocks, F. le Noble and N. Rajewsky (2013). "Circular RNAs are a large class of animal RNAs with regulatory potency." *Nature* 495(7441): 333-338.
- Miller, J. W., C. R. Urbinati, P. Teng-Umnuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton and M. S. Swanson (2000). "Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy." *EMBO J* 19(17): 4439-4448.
- Nigro, J. M., K. R. Cho, E. R. Fearon, S. E. Kern, J. M. Ruppert, J. D. Oliner, K. W. Kinzler and B. Vogelstein (1991). "Scrambled exons." *Cell* 64(3): 607-613.
- Pan, T., X. Sun, Y. Liu, H. Li, G. Deng, H. Lin and S. Wang (2018a). "Correction to: Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis." *Plant Mol Biol* 96(3): 231.
- Pan, T., X. Sun, Y. Liu, H. Li, G. Deng, H. Lin and S. Wang (2018b). "Heat stress alters genome-wide profiles of circular RNAs in Arabidopsis." *Plant Mol Biol* 96(3): 217-229.
- Patop, I. L., S. Wust and S. Kadener (2019). "Past, present, and future of circRNAs." *EMBO J* 38(16): e100836.
- Raczynska, K. D., C. G. Simpson, A. Ciesiolka, L. Szewc, D. Lewandowska, J. McNicol, Z. Szweykowska-Kulinska, J. W. Brown and A. Jarmolowski (2010). "Involvement of the nuclear cap-binding protein complex in alternative splicing in Arabidopsis thaliana." *Nucleic Acids Res* 38(1): 265-278.
- Salzman, J., R. E. Chen, M. N. Olsen, P. L. Wang and P. O. Brown (2013). "Cell-type specific features of circular RNA expression." *PLoS Genet* 9(9): e1003777.
- Salzman, J., C. Gawad, P. L. Wang, N. Lacayo and P. O. Brown (2012). "Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types." *PLoS One* 7(2): e30733.
- Shao, T., Y. H. Pan and X. D. Xiong (2021). "Circular RNA: an important player with multiple facets to regulate its parental gene expression." *Mol Ther Nucleic Acids* 23: 369-376.
- Song, C., Y. Zhang, W. Huang, J. Shi, Q. Huang, M. Jiang, Y. Qiu, T. Wang, H. Chen and H. Wang (2021). "Circular RNA Cwc27 contributes to Alzheimer's disease pathogenesis by repressing Pur-alpha activity." *Cell Death Differ*.
- Sun, X., L. Wang, J. Ding, Y. Wang, J. Wang, X. Zhang, Y. Che, Z. Liu, X. Zhang, J. Ye, J. Wang, G. Sablok, Z. Deng and H. Zhao (2016). "Integrative analysis of Arabidopsis thaliana transcriptomics reveals intuitive splicing mechanism for circular RNA." *FEBS Lett* 590(20): 3510-3516.
- Wang, B. B. and V. Brendel (2004). "The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing." *Genome Biol* 5(12): R102.
- Wang, E. T., D. Treacy, K. Eichinger, A. Struck, J. Estabrook, H. Olafson, T. T. Wang, K. Bhatt, T. Westbrook, S. Sedehizadeh, A. Ward, J. Day, D. Brook, J. A. Berglund, T. Cooper,

D. Housman, C. Thornton and C. Burge (2019). "Transcriptome alterations in myotonic dystrophy skeletal muscle and heart." *Hum Mol Genet* 28(8): 1312-1321.

Wang, P. L., Y. Bao, M. C. Yee, S. P. Barrett, G. J. Hogan, M. N. Olsen, J. R. Dinneny, P. O. Brown and J. Salzman (2014). "Circular RNA is expressed across the eukaryotic tree of life." *PLoS One* 9(6): e90859.

Wang, Y., M. Yang, S. Wei, F. Qin, H. Zhao and B. Suo (2016). "Identification of Circular RNAs and Their Targets in Leaves of *Triticum aestivum* L. under Dehydration Stress." *Front Plant Sci* 7: 2024.

Ye, C. Y., L. Chen, C. Liu, Q. H. Zhu and L. Fan (2015). "Widespread noncoding circular RNAs in plants." *New Phytol* 208(1): 88-95.

Ye, J., L. Wang, S. Li, Q. Zhang, Q. Zhang, W. Tang, K. Wang, K. Song, G. Sablok, X. Sun and H. Zhao (2017). "AtCircDB: a tissue-specific database for Arabidopsis circular RNAs." *Brief Bioinform*.

Zhang, P., X. Meng, H. Chen, Y. Liu, J. Xue, Y. Zhou and M. Chen (2017). "PlantCircNet: a database for plant circRNA-miRNA-mRNA regulatory networks." *Database (Oxford)* 2017.

Zhao, T., L. Wang, S. Li, M. Xu, X. Guan and B. Zhou (2017). "Characterization of conserved circular RNA in polyploid *Gossypium* species and their ancestors." *FEBS Lett* 591(21): 3660-3669.

Zuo, J., Q. Wang, B. Zhu, Y. Luo and L. Gao (2016). "Deciphering the roles of circRNAs on chilling injury in tomato." *Biochem Biophys Res Commun* 479(2): 132-138.

