



POLITECHNIKA ŚLĄSKA

WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI

Autoreferat pracy doktorskiej

Selection of miRNA isoform markers differentiating between follicular thyroid cancer and follicular thyroid adenoma from high-throughput sequencing data

Autor: mgr inż. Tomasz Stokowy

Promotor: dr hab. inż. Krzysztof Fajarewicz, prof. nzw. Politechniki Śląskiej
Politechnika Śląska w Gliwicach, Instytut Automatyki

Promotor pomocniczy: dr n. med. Michał Jarząb

Centrum Onkologii – Instytut im. Marii Skłodowskiej - Curie Oddział w Gliwicach, III Klinika Radioterapii i Chemioterapii

Gliwice, kwiecień 2013



Spis treści

1. Wprowadzenie	3
2. Cele pracy	5
3. Metodologia badawcza	6
4. Wyniki	9
5. Wnioski	14
6. Źródła finansowania	15
7. Publikacje autora i zgłoszone do czasopism manuskrypty	16
8. Spis literatury (wybrane pozycje)	19

1. Wprowadzenie

Guzki tarczycy są często obserwowane u pacjentów, szczególnie w badaniach ultrasonograficznych. Częstość ich występowania u ludzi waha się w przedziale 4-76%, w zależności od badanej populacji (American Association of Clinical Endocrinologists/Associazione Medici Endocrinologi Task Force on Thyroid Nodules, Gharib et al. 2010). Wśród guzków tarczycy tylko 5% (Hegedüs 2004) należy do grupy złośliwych, to znaczy mogących wykazywać inwazyjny wzrost i predyspozycje do przerzutów. Najczęściej spotykanym złośliwym nowotworem tarczycy jest rak brodawkowaty będący stosunkowo łatwym w diagnostyce przedoperacyjnej. Drugim najczęściej diagnozowanym złośliwym guzem tarczycy jest rak pęcherzykowy (FTC), diagnozowany w 10-20% przypadków raka tarczycy, w zależności od badanej populacji (Handkiewicz-Junak et al. 2010). Nowotwór ten był przedmiotem wielu badań naukowych ze względu na jego znaczne cytologiczne, patomorfologiczne i biologiczne podobieństwo do łagodnego gruczolaka pęcherzykowego tarczycy (FA). W praktyce, niski stopień różnic pomiędzy tymi nowotworami zmusza klinicystów do operacji tarczycy u każdego pacjenta z guzem pęcherzykowym, niezależnie czy jest on łagodny czy złośliwy. Praktyka ta prowadzi do co najmniej 50% niepotrzebnych operacji tarczycy. Obecnie, problem rozróżnienia guzów pęcherzykowych tarczycy stał się pierwszym z trzech najważniejszych celów Europejskiego Stowarzyszenia Tarczycy (Zjazd European Thyroid Association w Pizie, Włochy 2012).

W związku ze stosunkowo rzadkim występowaniem omawianej choroby oraz ograniczonym dostępem do materiału klinicznego guzów pęcherzykowych tarczycy konieczna jest międzynarodowa współpraca mająca na celu rozwiązanie problemu. Jedynie wspólne badania wielu ośrodków pozwalają na uzyskanie statystycznie znamiennej grupy próbek do badań genetycznych.

Obecnie pojawiła się tendencja do poszukiwań różnic pomiędzy FTC i FA opartych o biopsję cienkoigłową tarczycy (FNAC), wykonywaną przedoperacyjnie (Ferraz et al. 2011). Alternatywne badania wykonywane są na podstawie materiału pochodzącego z operacji, zarówno w materiale mrożonym jak i umieszczanym w blokach parafinowych (FFPE) (Borup et al. 2010; Ferraz et al. 2012). Oprócz różnych metod pozyskiwania materiału klinicznego, znane są także różne metody klasyfikacji nowotworów, w zależności od regionu świata.

Na chwilę obecną społeczność międzynarodowa wiąże duże nadzieje z zastosowaniem badań genetycznych najnowszej generacji w celu poprawy dokładności klasyfikacji omawianych nowotworów (Rossing 2013). Ograniczona obecnie możliwość diagnostyki może zostać zastąpiona medycyną personalizowaną, ułatwiającą zastosowanie odpowiednich środków zaradczych dla podtypu choroby. Będzie to jednak możliwe jedynie dzięki pracy środowisk medycznych, biologicznych i informatycznych, tworzących konsorcjum badawcze działające nad wspólnym projektem.

Szczególne perspektywy dla poprawy diagnostyki dają badania nad krótkimi sekwencjami nukleotydowymi (miRNA). Badanie poziomu ekspresji tych stabilnych (nie ulegających degradacji niezależnie od metody pobrania materiału) sekwencji o średniej długości 22 zasad (Bartel 2004) daje możliwości rozróżnienia raków pęcherzykowych i gruczolaków pęcherzykowych tarczycy (Rossing 2013; Reddi et al. 2011; Lassalle et al. 2011; Nikiforova et al. 2008; Dettmer et al. 2013; Shen et al. 2012). Niestety, wyniki prac naukowych w tym temacie nie są zgodne (Stokowy et al., praca zgłoszona do czasopisma *Journal of the Clinical Endocrinology and Metabolism*). Może to wynikać z

faktu, że liczba próbek pęcherzykowych guzów tarczycy w tych badaniach była stosunkowo niska (maksimum 38 (Dettmer et al. 2013)) oraz z obserwacji, że liczba różnicujących miRNA jest stosunkowo niska (szacunkowo 10-20) w stosunku do liczby wszystkich znanych ludzkich miRNA (obecnie ponad 2000 (Kozomara and Griffiths-Jones 2011; Griffiths-Jones et al. 2008)). Podsumowując, konieczna jest wyższa moc statystyczna badań, oparta o większą liczbę próbek i czułą metodę detekcji ekspresji miRNA aby w racjonalny sposób wskazać a następnie potwierdzić na niezależnym zbiorze danych markery złośliwości nowotworów pęcherzykowych tarczycy.

Dzięki współpracy Politechniki Śląskiej w Gliwicach, Instytutu Onkologii Oddział w Gliwicach oraz Uniwersytetu w Lipsku (Niemcy) udało się zebrać znaczne ilości próbek do badań genetycznych ($N > 200$), pozyskać wcześniej opublikowane zbiory próbek danych z różnych guzów tarczycy ($N > 400$) oraz zaplanować, przeprowadzić i przeanalizować 50 próbek FTC i FA metodą głębokiego sekwencjonowania (sekwencjonowania nowej generacji - HTS). Szczegóły dotyczące materiału i platform badania ekspresji miRNA ujęte zostały w rozdziale „Metodologia badawcza”.

Przeprowadzone badania genetyczne dały w wyniku olbrzymie ilości informacji, których analiza jest przedmiotem rozprawy doktorskiej. Rolą badań informatycznych w projekcie jest powtórna analiza wcześniej opublikowanych danych a także nowatorska, dedykowana analiza danych uzyskanych z głębokiego sekwencjonowania miRNA. Tego typu praca badawcza jest jedną z pierwszych wykonanych w Polsce i pierwszą opartą o technologię sekwencjonowania Illumina/Solexa. Technologia głębokiego sekwencjonowania pozwala na precyzyjny odczyt sekwencji z dokładnością wyższą niż 99.9% dla pojedynczej zasady, nie osiągalny dla żadnej innej metody laboratoryjnej. Tak dokładny wynik pozwala nie tylko na analizę ekspresji samych miRNA, ale także na ocenę ilości poszczególnych izoform i wariantów miRNA (należących do tej samej rodziny).

Analiza danych profilu miRNA wymaga zastosowania narzędzi statystycznych oraz algorytmów sztucznej inteligencji w celu możliwie precyzyjnej klasyfikacji badanych nowotworów. Dodatkowym aspektem w pracy jest fakt, że głównym celem jest precyzyjne wskazywanie nowotworów łagodnych (FA) w celu uniknięcia operacji, a nie jak w większości prac tego typu – wskazanie nowotworu złośliwego. Aby zrealizować tak przedstawione zadanie konieczna jest maksymalizacja specyficzności klasyfikatora i jego negatywnej wartości predykcyjnej (NPV) przy jednoczesnej minimalizacji liczby wskazanych markerów (aspekt praktyki klinicznej i możliwości wdrożenia). Ze względu na złożoność i brak dostępnego oprogramowania dedykowanego to podobnych analiz całość pracy opierana jest o oprogramowanie samodzielnie zaimplementowane przez autora pracy w językach R/Bioconductor, Perl i Unix Shell. Do wykonania analiz posłużył klaster obliczeniowy „Ziemowit” zakupiony w ramach projektu Śląska BIO-FARMA. Zastosowanie technik obliczeń równoległych pozwoliło na wykonanie projektu w terminie i wykonanie wielu symulacji, zarówno ujętych jak i nie opisanych w rozprawie doktorskiej.

2. Cele pracy

Głównym celem pracy jest selekcja markerów miRNA różnicujących pomiędzy rakiem pęcherzykowym tarczycy (FTC) i gruczolakiem pęcherzykowym tarczycy (FA). Dodatkowo, konieczna jest ocena wybranych markerów pod kątem możliwości klasyfikacji próbek pochodzących z biopsji cienkoigłowych guzów tarczycy.

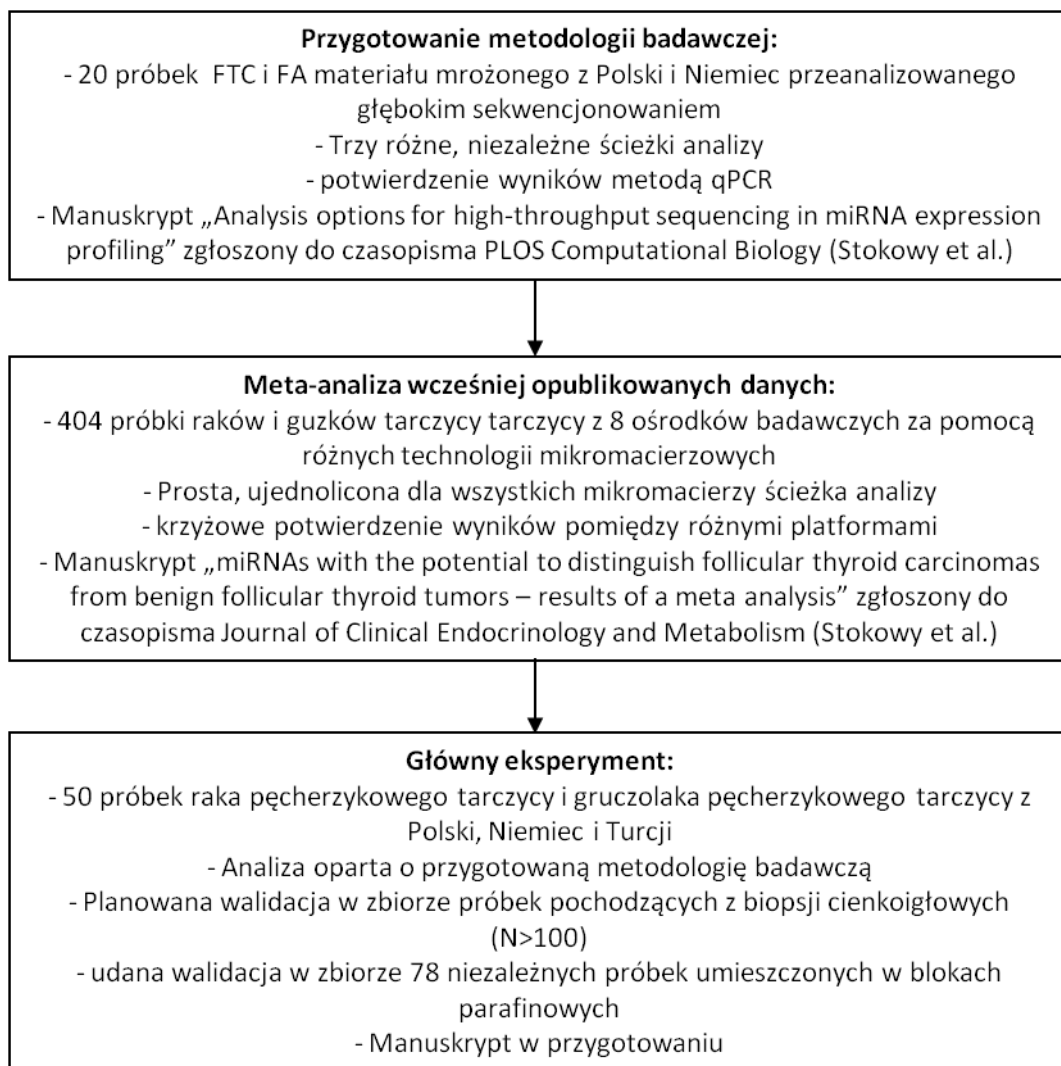
Cel pracy został osiągnięty w wyniku realizacji zadań cząstkowych:

1. Zweryfikować dotychczas opublikowane markery. Dane wielkoskalowe badające ekspresję miRNA dostępne w publicznych repozytoriach powinny być przedmiotem ponownej analizy oraz integracji z własnym zbiorem danych z głębokiego sekwencjonowania.
2. Wybrać markery złośliwości nowotworów tak, aby klasyfikowały próbki pochodzące z biopsji cienkoigłowych pęcherzykowych guzów tarczycy. Utworzony klasyfikator bądź lista markerów ekspresyjnych miRNA powinna być możliwie stabilnym i specyficznym narzędziem do rozróżniania złośliwego raka pęcherzykowego tarczycy od łagodnych gruczolaków pęcherzykowych tarczycy.
3. Zminimalizować liczbę markerów przy założeniu, żeby posiadała ona wystarczającą diagnostycznie specyficzną i negatywną wartość predykcyjną (NPV) pozwalającą na jej kliniczne zastosowanie.
4. Ocenąć wpływ izoform miRNA na działanie klasyfikatora na podstawie analizy danych pochodzących z głębokiego sekwencjonowania.

Zadania 2, 3 i 4 zostały zrealizowane poprzez zastosowanie samodzielnie zaprojektowanego i zaimplementowanego oprogramowania, ponieważ dotychczas opublikowane podejścia algorytmiczne nie były stosowne dla problemu rozróżnienia FTC i FA.

3. Metodologia badawcza

W eksperymentach opisanych w pracy przeprowadzono analizę próbek pochodzących z 8 ośrodków badawczych z Polski, Niemiec, Turcji, Stanów Zjednoczonych, Francji i Danii. Próbki zostały w większości uzyskane na zasadach wymiany naukowej materiału lub wymiany wyników eksperymentalnych. Dokładny opis próbek wraz z technologiami analizy biologicznej i opisem histopatologicznym został zamieszczony w rozprawie doktorskiej w rozdziale „Material”. Opis ogólny z liczbami próbek i planem eksperymentu przedstawiono na schemacie (Rysunek 1).



Rysunek 1. Ogólne zestawienie metodologii badawczej i plan eksperymentu.

Główną częścią projektu była analiza próbek z głębokiego sekwencjonowania miRNA. Zastosowane sekwencjonowanie Illumina/Solexa, zwane też sekwencjonowanie drugiej generacji jest obecnie najczęściej stosowaną wielkoskalową metodą badania sekwencji nukleotydowych. Pozwala na pomiar ilościowy i jakościowy przy wysokiej precyzji odczytu, przekraczającej 99,9% dla pojedynczej zasady.

Głębokie sekwencjonowanie Illumina polega na odczycie milionów krótkich sekwencji (zazwyczaj o długościach 36, 50, 72 lub 101 zasad, w zależności od eksperymentu) i zdeponowaniu ich na serwerze akwizycyjnym. Eksperyment biologiczny przebiega w oparciu o przygotowanie biblioteki sekwencji dla próbek, namnożenie klastrów na płytce i ich odczyt w oparciu o system wizyjny z systemem laserów. Całość eksperymentów wykonywana była zgodnie z zaleceniami producenta. Szczegółowe informacje dotyczące metodologii sekwencjonowania można znaleźć w formie dokumentów w serwisie www.illumina.com, oraz w formie filmów instruktażowych w serwisie www.youtube.com pod hasłem „Illumina sequencing”.

W eksperymencie różnicowania FTC i FA analizowano sekwencje miRNA o średniej długości 22 nukleotydów za pomocą 50 cykli sekwencjonowania (odczytu 50 zasad). W celu uzupełnienia sekwencji miRNA biologicznie przyłączane były adaptory, które następnie w pierwszym etapie analizy były obcinane przy wykorzystaniu oprogramowania *cutadapt* (Martin 2011). Po obcięciu adapterów przeprowadzono kontrolę jakości odczytów za pomocą *FastQC* (Simon Andrews, Babraham Bioinformatics Institute, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, oprogramowanie nie opublikowane). Pozytywne wyniki kontroli jakości pozwoliły przejść do etapu normalizacji Reads Per Million (RPM) (Motameny et al. 2010), który podobnie jak w przypadku normalizacji danych mikromacierzowych ma na celu umożliwienie wiarygodnego porównywania ekspresji pomiędzy próbkami. W dalszym etapie dane transformowano funkcją logarytmu o podstawie 2 aby zredukować efekt skośności rozkładu ekspresji (podobnie jak w standardach analiz danych mikromacierzowych). Szczegółowy opis algorytmów i narzędzi przetwarzania wstępnego przedstawiono w rozprawie doktorskiej.

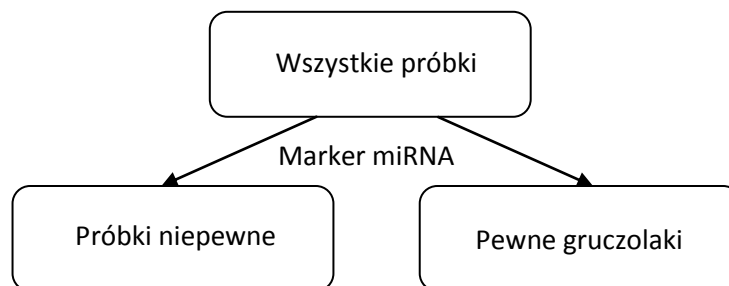
Dalszy etap analizy, mający na celu wskazanie markerów miRNA został oparty o samodzielnie zaimplementowane oprogramowanie. Wykorzystano w niej metody analizy nienadzorowanej, takie jak analiza głównych składowych (PCA), klasteryzacja hierarchiczna, czy metoda nienadzorowanej analizy Random Forests (Shi et al. 2005). W analizie nadzorowanej wykorzystane zostały metody statystycznego testowania – test t Studenta z poprawką na krotkość powtórzeń (False Discovery Rate – FDR) oraz ocena krotkości zmian oparta o mediany (w celu minimalizacji wpływu próbek odstających na otrzymany wynik).

W ramach analizy nadzorowanej wykorzystano także szereg algorytmów sztucznej inteligencji, takich jak drzewa decyzyjne i metoda nadzorowana Random Forests (Breiman 2001). Zaproponowano także własny klasyfikator, przygotowany w oparciu o algorytm drzew decyzyjnych. Charakter klasyfikatora jest dedykowany do separacji guzów łagodnych przy wykorzystaniu jak najmniejszej liczby cech (sekwencji miRNA). Koncepcja klasyfikatora oparta została o wiedzę ekspercką, wykorzystującą powszechnie znane w środowisku medycznym fakty:

- większość raków pęcherzykowych tarczycy ma ekspresję większości miRNA zdecydowanie niższą niż gruczolakach pęcherzykowych tarczycy

- w populacji ludzkiej gruczolak pęcherzykowy tarczycy jest diagnozowany częściej niż rak pęcherzykowy tarczycy
- metoda walidacyjna (qPCR) nie sprawdza się dla sekwencji miRNA o niskiej ekspresji (na przykład znacząco obniżonej na skutek procesu nowotworowego)

Koncepcję zaproponowanego klasyfikatora przedstawiono na schemacie (Rysunek 2). Należy podkreślić, że konstrukcja klasyfikatora nie została oparta lub dopasowana do zbioru danych. Na każdym kroku analizy separowano etapy uczenia i testowania aby uniknąć wyciekania informacji (tzw. information leakage). Podobnie w iteracyjnych procedurach testowania jakości klasyfikatora statystyczna selekcja markerów przebiegała w każdej iteracji uczenia niezależnie.



Rysunek 2. Schemat zaproponowanego klasyfikatora

Do realizacji celów pracy oraz wątków pobocznych opisanych w pracy zaimplementowano trzy pakiety oprogramowania. Pierwszy *Sequence_counter.pl* służy do obliczeniowo efektywnego zliczania poszczególnych sekwencji (izoform miRNA) w surowych danych z eksperymentów. Drugi pakiet – *Seed_sequencing.R* służy do normalizacji danych i uwzględnienia regionu wiązania miRNA (lub inaczej – jądra wiązania zwanego z języka angielskiego seed) będącego najważniejszą częścią sekwencji, wiążącą się do transkryptów genów i wpływającą na ich regulację. Trzecim samodzielnie zaimplementowanym narzędziem jest *Family_members_generator.R* wskazujący izoformy mogące potencjalnie wpływać na wyniki walidacji qPCR ze względu na podobieństwo do sekwencji docelowych. Wszystkie pakiety są szczegółowo opisane w rozprawie doktorskiej i dostępne do pobrania z internetu. Praca opisująca ich działanie została zgłoszona do czasopisma PLOS Computational Biology.

4. Wyniki

Celem pracy jest selekcja markerów różnicujących dla nowotworów pęcherzykowych tarczycy, dlatego sekcja wyników zawiera metody zaprojektowane przez autora rozprawy. Aspekty biologiczne i medyczne wyników są przedstawione równoległe z walidacją eksperymentalną na niezależnym zbiorze próbek.

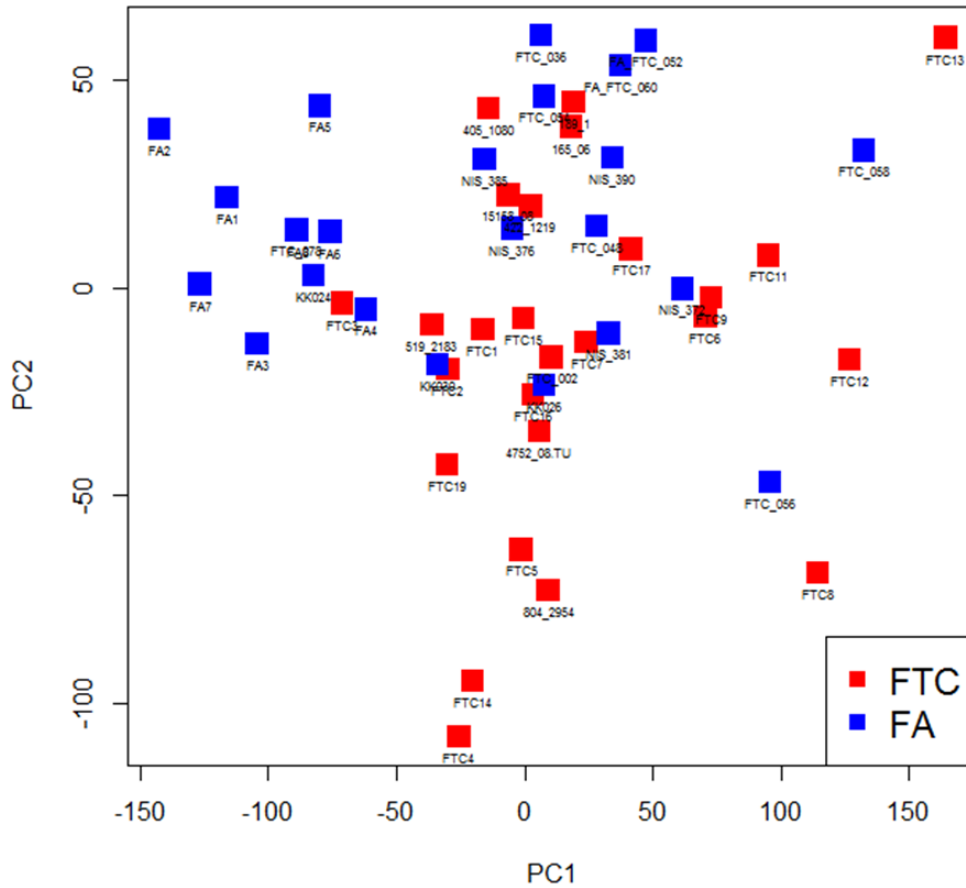
Wyniki analizy zostały oparte o $4,7 \pm 1,5$ miliona surowych sekwencji na próbkę (podane w formie średnia \pm odchylenie standardowe) pozyskanych z sekwencjonowania Illumina. W sumie, wykryto 48939 różnych sekwencji (głównie izoform miRNA) obecnych w co najmniej połowie z 50 badanych próbek. Sekwencje odpowiadały 450 ± 70 wykrytym (na próbkę) w tkance tarczycy miRNA, wskazanym wśród wszystkich 2042 zdeponowanych w bazie miRBase v19 (Kozomara and Griffiths-Jones 2011; Griffiths-Jones et al. 2008).

Nienadzorowana analiza sekwencji potwierdziła istnienie problemu histopatologicznego w danych profilu miRNA (brak separowalności liniowej próbek, Rysunek 3 i Rysunek 4). W analizie PCA zauważono klaster gruczolaków pęcherzykowych, widoczny po lewej stronie wykresu (Rysunek 3). Wynik potwierdziła analiza metodą klasteryzacji hierarchicznej (Rysunek 4) oraz metoda nienadzorowanych lasów losowych (wynik przedstawiony w rozprawie doktorskiej).

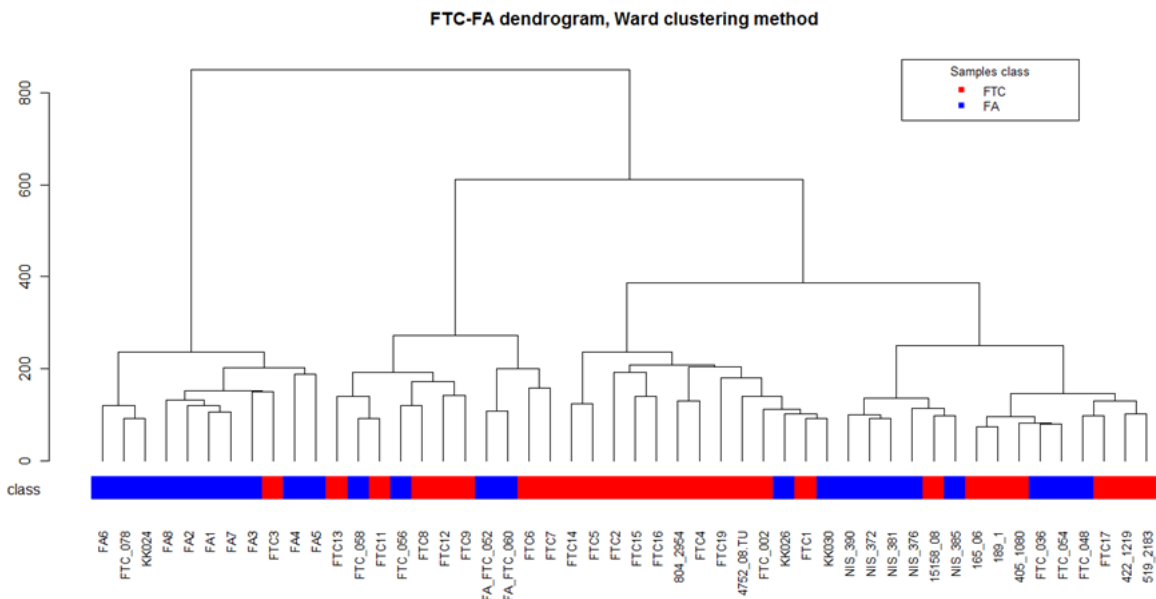
Wyniki powtórnej analizy dotychczas opublikowanych prac naukowych i danych w nich zawartych wspierają rezultaty głębokiego sekwencjonowania. Można z nich wnioskować że liczba miRNA różnicujących pomiędzy FTC i FA jest stosunkowo niska (10-20), a globalne różnice pomiędzy badanymi klasami próbek nie są wyraźne (Stokowy et al., manuskrypt zgłoszony do czasopisma Journal of the Clinical Endocrinology and Metabolism).

W oparciu o zbiór 50 próbek wybrano 12 sekwencji miRNA różnicujących pomiędzy FTC i FA (kryteria wyboru szczegółowo przedstawione w rozprawie doktorskiej). Stopień zmian ekspresji pomiędzy klasami próbek przedstawiony został na wykresach pudełkowych, wskazujących medianę (linia), pierwszy i trzeci kwartył (pudełko), wartość średnią w klasie (czarny kwadrat), 1.5 przedziału pomiędzy kwartylami (wąsy) oraz ekspresję miRNA w próbkach (punkty)(Rysunek 5).

1st and 2nd Principal Component: FTC-FA analysis

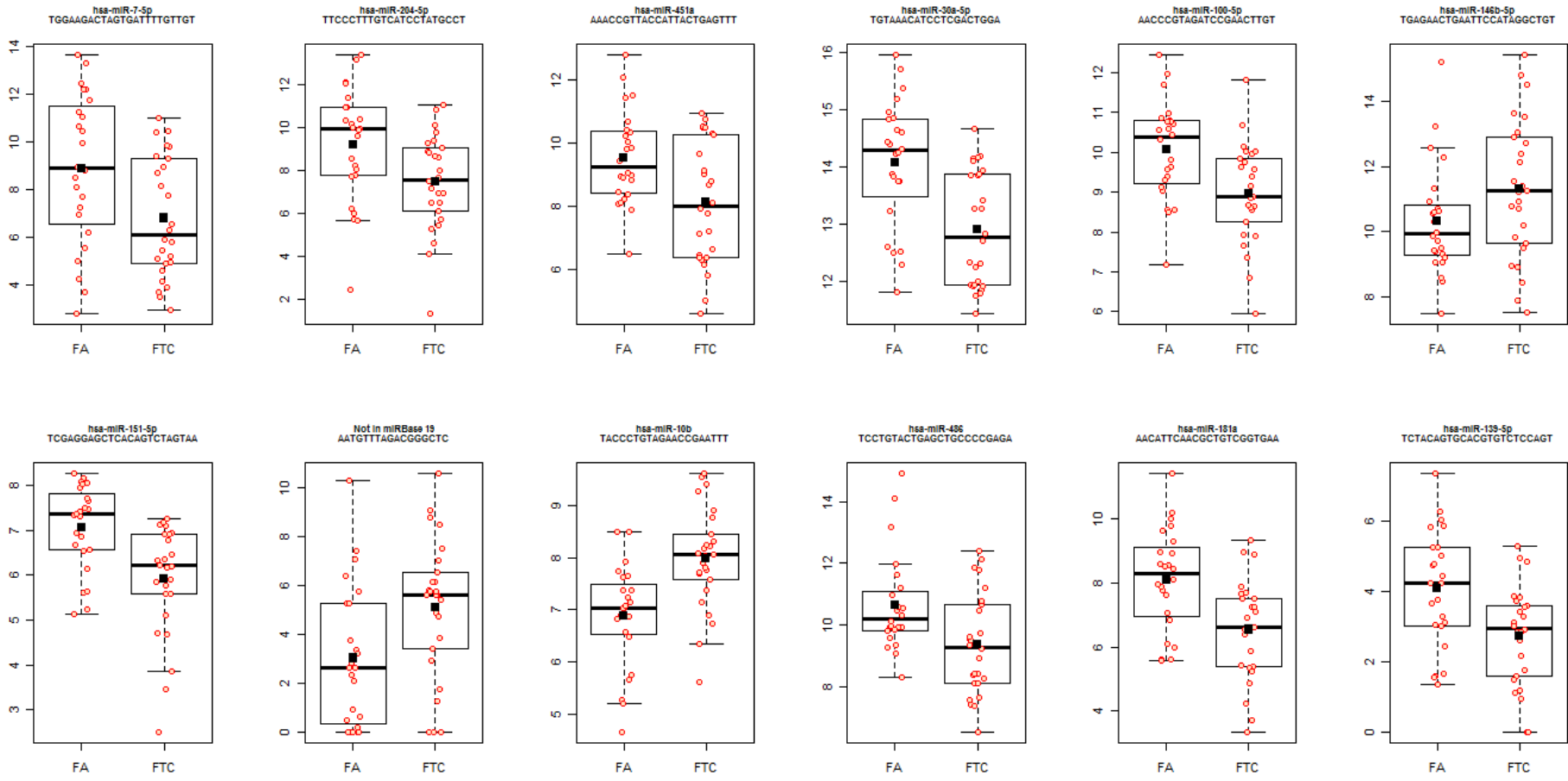


Rysunek 3. Wynik nienadzorowanej analizy głównych składowych wykonanej na całym zbiorze miRNA.

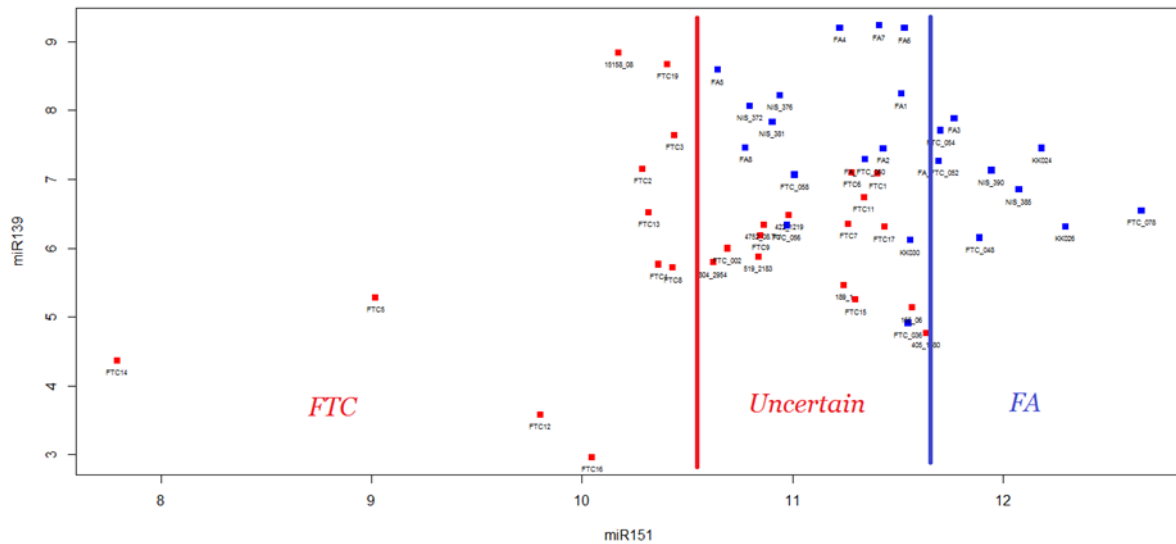


Rysunek 4. Wynik klasteryzacji hierarchicznej na całym zbiorze danych pochodzących z głębokiego sekwencjonowania.

Rysunek 5. Różnice w profilu ekspresji dla 12 wybranych miRNA różnicujących FTC i FA



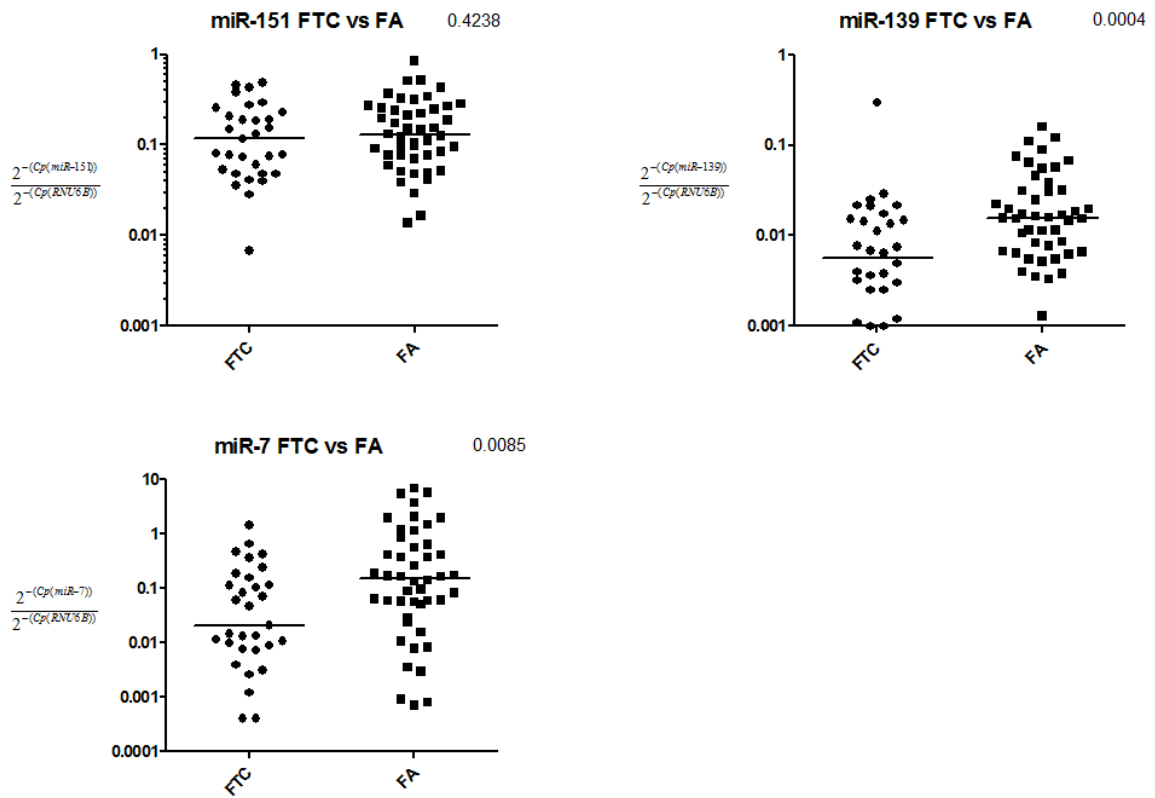
Wśród 12 wybranych markerów szczególną uwagę zwrócić należy na hsa-miR-151a-5p oraz hsa-miR-139-5p. Te dwa miRNA pozwoliły na stosunkowo precyzyjną klasyfikację guzów pęcherzykowych tarczycy na podstawie danych z głębokiego sekwencjonowania (Rysunek 6, szczegóły w rozprawie doktorskiej).



Rysunek 6. Rozróżnienie FTC i FA za pomocą ekspresji hsa-miR-151a-5p i hsa-miR-139-5p.

Należy dodatkowo podkreślić fakt, że zaprojektowany klasyfikator (rozdział „Metodologia badawcza”) pomimo niezależnego wyboru cech w pętłach walidacji krzyżowej wykorzystywał najczęściej właśnie te dwie cechy. Liczba cech nie była narzucana z góry, jednak w każdej iteracji 2 miRNA wystarczyły aby w procesie uczenia rozdzielić próbki na klasy FTC i FA. Ocena klasyfikatora w walidacji krzyżowej leave-one-out dała wynik dokładności 0.82, specyficzności 0.94 i negatywnej wartości predykcyjnej równej 0.67.

Walidacja trzech miRNA: hsa-miR-151a-5p, hsa-miR-139-5p oraz hsa-miR-7-5p pozwoliła na potwierdzenie dwóch markerów (hsa-miR-139-5p i hsa-miR-7-5p) w niezależnym zbiorze 78 próbek (Rysunek 7). hsa-miR-151a-5p nie został potwierdzony ze względu na wpływ izoform, które potencjalnie mogły zaburzyć proces walidacji qPCR. W związku z laboratoryjnym potwierdzeniem istnienia izoform negatywnie wpływających na badanie profilu hsa-miR-151a-5p metodą qPCR dalsze badania nad tym zjawiskiem będą prowadzone w przyszłości.



Rysunek 7. Wyniki walidacji qPCR dla trzech markerów miRNA w zbiorze niezależnych 78 próbek

5. Wnioski

Cele pracy zostały wykonane poprzez zbiór zintegrowanych bioinformatycznych, biologicznych i medycznych eksperymentów. Następująca lista wniosków jest wynikiem pracy wykonanej przez autora rozprawy:

1. Zbiór 12 markerów został zaproponowany do walidacji na niezależnym zbiorze próbek w oparciu o analizę danych z głębokiego sekwencjonowania miRNA (HTS). Liczba markerów jest wystarczająco niska aby umożliwić walidację na próbkach biopsji cienkoigłowej tarczycy i wprowadzić wyniki do praktyki klinicznej.
2. hsa-miR-151a-5p i hsa-miR-139-5p są najistotniejszymi markerami złośliwości guzów pęcherzykowych tarczycy na podstawie danych pochodzących z głębokiego sekwencjonowania. Wspólnie z hsa-miR-7-5p i hsa-miR-486-5p (pozytywnie walidowanymi we wcześniej opublikowanych danych) tworzą listę miRNA znamienne obniżonych w utkaniu raka pęcherzykowego tarczycy.
3. Wskazane markery są stabilnymi sekwencjami nukleotydów, które można potwierdzić w innych zbiorach, niezależnie od typu materiału biologicznego.
4. hsa-miR-151a-5p jest celem dalszych badań i ponownej oceny techniką reakcji łańcuchowej polimerazy w czasie rzeczywistym (qPCR), ponieważ profil tego miRNA wymaga walidacji specyficznej dla poszczególnych izoform.
5. Głębokie sekwencjonowanie wprowadza wyższą precyzję do wszystkich eksperymentów RNA. Łączy w sobie zalety klasycznego sekwencjonowania Sangera, metody qPCR i mikromacierzy ekspresyjnych w jednym eksperymencie. W niektórych przypadkach technologia HTS jest bardziej specyficzna dla poszczególnych izoform niż metoda qPCR.
6. FTC i FA nie są jednorodnymi grupami nowotworów. Dalsze badania nad próbkami o niepewnej klasyfikacji guza, podtypem oksyfilnym raka tarczycy lub łagodnymi guzkami tarczycy są konieczne dla pełnego zrozumienia molekularnego profilu i zróżnicowania guzów tarczycy.

Wyniki i wnioski pracy prowadzą do optymistycznych oczekiwań, że zaproponowane markery miRNA poprawią możliwości diagnostyczne i zostaną wprowadzone w praktyce klinicznej. Praca doktorska przygotowana została w języku angielskim aby przedstawić wyniki wykonanej pracy społeczności współpracujących naukowców z różnych krajów.

6. Źródła finansowania

Projekt doktorski był realizowany w ramach programu MPD Fundacji na rzecz Nauki Polskiej, współfinansowany przez Unię Europejską z Europejskiego Funduszu Rozwoju Regionalnego.

Praca była wspierana przez NCN – Narodowe Centrum Nauki granty [N N519 647840] i [N 401 072 637] oraz grant dla młodych naukowców Politechniki Śląskiej w Gliwicach – BK Young [BKM-233/RAu1/2012/t.8].

Badania wspierane były przez grant Centro Nacional de Pesquisa e Tecnologia do Brasil (CNPq), Brazil przyznany dla Carolina Ferraz [290023/2009-2], granty DFG [ES162/4-1] i Krebshilfe [109994] przyznane dla Markus Eszlinger oraz nieograniczony grant Novo Nordisk Foundation przyznany dla Laszlo Hagedüs.

Analiza głębokiego sekwencjonowania przeprowadzona została w IZKF Leipzig, Interdyscyplinarnym Centrum Badań Klinicznych na Uniwersytecie w Lipsku (Niemcy) w ramach projektu Z03.

Obliczenia zostały wykonane na klastrze obliczeniowym Ziemowit (<http://ziemowit.hpc.polsl.pl>) zakupionym ze środków projektu Śląska BIO-FARMA [POIG.02.01.00-00-166/08] w Laboratorium Biologii Obliczeniowej i Bioinformatyki, Centrum Biotechnologii na Politechnice Śląskiej w Gliwicach.

7. Publikacje autora i zgłoszone do czasopism manuskrypty

Manuskrypty wysłane do czasopism naukowych z listy filadelfijskiej:

Stokowy T., Wojtaś B, Fujarewicz K, Jarzab B, Eszlinger M, Paschke R. **Analysis options for high-throughput sequencing in miRNA expression.**

Submitted to PLOS Computational Biology. **(IF 5.22)**

Stokowy T., Wojtaś B, Fujarewicz K, Jarzab B, Eszlinger M, Paschke R. **miRNAs with the potential to distinguish follicular thyroid carcinomas from benign follicular thyroid tumors - results of a meta-analysis.**

Submitted to Journal of the Clinical Endocrinology and Metabolism. **(IF 5.96)**

Wojtaś B, Ferraz C, Stokowy T, Jarzab B, Paschke R, Eszlinger M. **Differential miRNA expression defines migration and reduced apoptosis in Follicular Thyroid Carcinomas.**

Submitted to Journal of the Clinical Endocrinology and Metabolism. **(IF 5.96)**

Pfeifer A, Wojtaś B, Oczko-Wojciechowska M, Kukulska A, Czarniecka A, Eszlinger M, Musholt T, Stokowy T, Stobiecka E, Rusinek D, Tyszkiewicz T, Kowal M, Jarzab M, Hauptmann S, Lange D, Paschke R, Jarzab B. **Molecular discrimination between follicular thyroid carcinoma and adenoma based on gene expression profiling in formalin-fixed paraffin-embedded samples.**

Submitted to BMC Medical Genomics. **(IF 3.69)**

Malcher A, Rozwadowska N, Stokowy T, Jedrzejczak P, Zietkowiak W, Kurpisz M. **The gene expression analysis of paracrine/autocrine factors in patients with spermatogenetic failure compared to normal spermatogenesis.**

Submitted to American Journal of Reproductive Immunology. **(IF 3.05)**

2013:

Waclawska A, Rozwadowska N, Stokowy T, Jedrzejczak P, Zietkowiak W, Kurpisz M. **Identification of potential molecular indicators in non-obstructive azoospermia.** The 10th International Congress of Andrology, 23-26 February 2013, Melbourne, Australia. Abstract published in Andrology (2013), 1(Suppl. 1):76. **(IF 2.97) (MNISW 32)**

2012:

Waclawska A, Rozwadowska N, Stokowy T, Jedrzejczak P, Zietkowiak W, Kurpisz M. **Potential biomarkers of non-obstructive azoospermia identified in gene expression analysis** – oral presentation at 7th European Congress of Andrology (ECA), November 2012, Berlin, Germany. Abstract published in Journal of Reproductive Medicine and Endocrinology 9(5):352-353. **(IF 2.05) (MNISW 27)**

Labun K, Stokowy T, Swierniak M, Wojtas B, Krohn K, Jarzab B, Fujarewicz K. **Detection of novel miRNA in deep sequencing data from thyroid cancer** – poster presentation at 16th Gliwice Scientific Meetings, Conference Materials page 54, November 2012, Gliwice, Poland

Stokowy T, Swierniak M, Wojtas B, Danch M, Krohn K, Fujarewicz K, Jarzab B, Paschke R, Eszlinger M. **Zastosowanie techniki sekwencjonowania nowej generacji, mikromacierzy i qPCR do badania miRNA dla różnicowania molekularnego guzów pęcherzykowych tarczycy**

– poster presentation at 20th Meeting of Polish Endocrinology Society, Conference Materials page 50, September 2012, Poznan, Poland.

Wojtas B, Ferraz C, Stokowy T, Jarzab B, Paschke R, Eszlinger M. **Identification of microRNAs regulating mRNA expression in follicular thyroid cancer** – poster presentation at Polish-German Biochemical Societies Joint Meeting, September 2012, Poznan, Poland

Stokowy T, Swierniak M, Wojtas B, Danch M, Krohn K, Fujarewicz K, Jarzab B, Paschke R, Eszlinger M. **RNA-Seq provides isoform specific miRNA expression data which require isoform qPCR verification** – oral presentation at European Thyroid Association conference, September 2012, Pisa, Italy. Abstract published in Eur Thyroid J 1(suppl 1) p.88-89. Presentation awarded with ETA Travel Grant.

Wojtas B, Pfeifer A, Stokowy T, Oczko-Wojciechowska M, Eszlinger M, Kukulska A, Musholt T, Jarzab B, Czarniecka A, Stobiecka E, Hauptmann S, Lange D, Paschke R, Jarzab B. **Follicular thyroid cancer molecular markers validation in FFPE material** – poster presentation at the 22nd European Association for Cancer Research meeting, 7-10 July 2012, Barcelona, Spain. Abstract published in European Journal of Cancer 48 suppl. (5):142. **(IF 4.78) (MNISW 32)**

Stokowy T, Swierniak M, Wojtas B, Danch M, Krohn K, Fujarewicz K, Jarzab B, Paschke R, Eszlinger M. **Correlation of NGS, microarray and qPCR data in a small RNA study** – poster presentation at 20th Annual International Conference Intelligent Systems for Molecular Biology, Conference Materials page 67, July 2012, Long Beach, CA, USA.

Stokowy T, Swierniak M, Wojtas B, Danch M, Krohn K, Fujarewicz K, Jarzab B, Paschke R, Eszlinger M. **RNA-Seq reveals usefulness of small RNA isoforms in thyroid tumors diagnosis** – poster presentation at 8th ISMB Student Council Symposium, Conference Materials page 25, July 2012, Long Beach, CA, USA. Presentation scored ISMB SCS Travel Grant award: <http://symposium.iscbsc.org/content/travel-fellowships>.

Wirth H, von Bergen M, Murugaiyan J, Rösler U, Stokowy T, Binder H. – **MALDI-typing of infectious algae of the genus Prototheca using SOM portraits**. J Microbiol Methods. 2012 Jan;88(1):83-97. Epub 2011 Oct 26. **(IF 2.018) (MNISW 27)**

2011:

Stokowy T, Wojtas B, Pfeifer A, Swierniak M, Oczko-Wojciechowska M, Fujarewicz K, Jarzab B, Eszlinger M, Krohn K, Paschke R, Jarzab B, Swierniak A. – **Selection of markers differentiating between follicular thyroid carcinoma (FTC) and follicular thyroid adenoma (FA)** – poster presentation at 10th Leipzig Research Festival for Life Sciences page 346, December 2011, Leipzig, Germany.

Stokowy T, Wojtas B, Pfeifer A, Swierniak M, Oczko-Wojciechowska M, Fujarewicz K, Jarzab B, Eszlinger M, Paschke R, Jarzab B, Swierniak A. – **Selection of markers differentiating between follicular thyroid carcinoma (FTC) and follicular thyroid adenoma (FA)** – poster presentation at Genome Informatics Meetings 2011, November 2011, Cold Spring Harbor, NY, USA

Stokowy T – **Selection of differentiating features in RNA-Sequencing of thyroid cancer samples.**, OWD International PhD workshop materials p. 269-274, Wisła 2011, ISBN 83-922242-8-0. Presentation scored two workshop awards.

Stokowy T, Wojtaś B, Fajarewicz K, Jarzab M, Pfeifer A, Jarzab B, Krogdahl A, Hauptmann S, Paschke R, Eszlinger M – **miRNA expression patterns differentiate between follicular thyroid carcinoma (FTC) and follicular thyroid adenoma (FA)** – oral presentation at European Thyroid Association conference, September 2011, Kraków

Wojtas B, Pfeifer A, Stokowy T, Oczko-Wojciechowska M, Eszlinger M, Kukulska A, Musholt T, Jarzab M, Czarniecka A, Stobiecka E, Hauptmann S, Lange D, Paschke R, Jarzab B. – **Follicular thyroid cancer molecular markers Discovery** – poster presentation at European Thyroid Association conference, September 2011, Kraków

Stokowy T, Wojtaś B, Pfeifer A, Oczko-Wojciechowska M, Świerniak M, Danch M, Jarzab M, Fajarewicz K, Jarzab B - **Bioinformatics analysis of follicular (FTC) and papillary thyroid cancer (PTC) RNA-Sequencing data** – oral presentation at The 5th polish-french conference on cancer biology and therapy May 2011, Paris

2010:

Pfeifer A, Oczko-Wojciechowska M, Wojtaś B, Kowalska M, Stokowy T, Jarzab M, Jarzab B – **Detecting genes expressed differentially in thyroid follicular adenomas and carcinomas: a meta-analysis.** 14th Gliwice Scientific Meetings 2010 conference materials - <http://gsn.io.gliwice.pl/streszczeniaplakat2010.html>

Stokowy T - **Classification of DNA microarray data with PCA based Support Vector Machines.**, OWD International PhD workshop materials, p. 31-36, Wisła 2010, ISBN 83-922242-7-2

Stokowy T, Jarzab M, Fajarewicz K, Oczko-Wojciechowska M, Pfeifer A, Paschke R, Eszlinger M - **Próba zastosowania techniki lasów losowych (RandomForests) do dyskryminacji typów histologicznych nowotworów tarczycy.** EndokrynologiaPolska, t. 61 nr 5, str. 632 / 2010 Conference Materials of IV „Thyroid Cancer” conference, Zakopane May 2010

Stokowy T - **Classification of DNA Microarray Data with Random Forests.** Advances in Soft Computing, Volume 69/2010, Pages 305-308, Springer Berlin / Heidelberg, ISBN 978-3-642-13104-2 Part of book with a separate title **(MNI SW 7)**

8. Spis literatury (wybrane pozycje)

- Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**: 281–297.
- Borup R, Rossing M, Henao R, Yamamoto Y, Krogdahl A, Godballe C, Winther O, Kiss K, Christensen L, Høgdall E, et al. 2010. Molecular signatures of thyroid follicular neoplasia. *Endocr Relat Cancer* **17**: 691–708.
- Breiman L. 2001. Random Forests. *Mach Learn* **45**: 5–32.
- Dettmer M, Vogetseder A, Durso MB, Moch H, Komminoth P, Perren A, Nikiforov YE, Nikiforova MN. 2013. MicroRNA expression array identifies novel diagnostic markers for conventional and oncocytic follicular thyroid carcinomas. *J Clin Endocrinol Metab* **98**: E1–7.
- Ferraz C, Eszlinger M, Paschke R. 2011. Current state and future perspective of molecular diagnosis of fine-needle aspiration biopsy of thyroid nodules. *J Clin Endocrinol Metab* **96**: 2016–2026.
- Ferraz C, Rehfeld C, Krogdahl A, Precht Jensen EM, Bösenberg E, Narz F, Hegedüs L, Paschke R, Eszlinger M. 2012. Detection of PAX8/PPARG and RET/PTC rearrangements is feasible in routine air-dried fine needle aspiration smears. *Thyroid Off J Am Thyroid Assoc* **22**: 1025–1030.
- Gharib H, Papini E, Paschke R, Duick DS, Valcavi R, Hegedüs L, Vitti P. 2010. American Association of Clinical Endocrinologists, Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules: executive summary of recommendations. *J Endocrinol Invest* **33**: 51–56.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154–158.
- Handkiewicz-Junak D, Czarniecka A, Jarzab B. 2010. Molecular prognostic markers in papillary and follicular thyroid cancer: Current status and future directions. *Mol Cell Endocrinol* **322**: 8–28.
- Hegedüs L. 2004. Clinical practice. The thyroid nodule. *N Engl J Med* **351**: 1764–1771.
- Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**: D152–157.
- Lassalle S, Hofman V, Ilie M, Bonnetaud C, Puisségur M-P, Brest P, Loubatier C, Guevara N, Bordone O, Cardinaud B, et al. 2011. Can the microRNA signature distinguish between thyroid tumors of uncertain malignant potential and other well-differentiated tumors of the thyroid gland? *Endocr Relat Cancer* **18**: 579–594.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: pp. 10–12.
- Motameny S, Wolters S, Nürnberg P, Schumacher B. 2010. Next Generation Sequencing of miRNAs – Strategies, Resources and Methods. *Genes* **1**: 70–84.
- Nikiforova MN, Tseng GC, Steward D, Diorio D, Nikiforov YE. 2008. MicroRNA expression profiling of thyroid tumors: biological significance and diagnostic utility. *J Clin Endocrinol Metab* **93**: 1600–1608.

- Reddi HV, Madde P, Milosevic D, Hackbarth JS, Algeciras-Schimmich A, McIver B, Grebe SKG, Eberhardt NL. 2011. The Putative PAX8/PPAR γ Fusion Oncoprotein Exhibits Partial Tumor Suppressor Activity through Up-Regulation of Micro-RNA-122 and Dominant-Negative PPAR γ Activity. *Genes Cancer* **2**: 46–55.
- Rossing M. 2013. Classification of Follicular Cell-derived Thyroid Cancer by Global RNA Profiling. *J Mol Endocrinol*. <http://jme.endocrinology-journals.org/content/early/2013/01/25/JME-12-0170> (Accessed March 8, 2013).
- Shen R, Liyanarachchi S, Li W, Wakely PE Jr, Saji M, Huang J, Nagy R, Farrell T, Ringel MD, de la Chapelle A, et al. 2012. MicroRNA signature in thyroid fine needle aspiration cytology applied to “atypia of undetermined significance” cases. *Thyroid Off J Am Thyroid Assoc* **22**: 9–16.
- Shi T, Seligson D, Belldegrun AS, Palotie A, Horvath S. 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Mod Pathol Off J United States Can Acad Pathol Inc* **18**: 547–557.