

Jarosław Jendrzewski

**Rola polimorfizmu rs944289 w predyspozycji do raka
brodawkowego tarczycy**

Praca na stopień doktora nauk medycznych
Promotor: prof. dr hab. med. Krzysztof Sworczak

Gdański Uniwersytet Medyczny

Gdańsk 2013

SPIS TREŚCI

1. WYKAZ STOSOWANYCH SKRÓTÓW	5
1.1. Najczęściej stosowane skróty	5
1.2. Skróty nazw genów omawianych w pracy	7
2. WSTĘP	9
2.1. Epidemiologia raka tarczycy	9
2.2. Klasyfikacja i charakterystyka kliniczna raka tarczycy	10
2.3. Środowiskowe czynniki ryzyka rozwoju raka tarczycy	11
2.4. Podłoże molekularne raka tarczycy	12
2.5. Metody wykorzystywane w identyfikacji genów biorących udział w patogenezie chorób nowotworowych	16
2.5.1. Analiza sprzężeń	16
2.5.2. Badania asocjacyjne	16
2.6. Warianty polimorficzne predysponujące do wysokozróżnicowanego raka tarczycy	22
2.6.1. Wariant rs2910164	24
2.6.2. Wariant rs1867277	25
2.6.3. Wariant rs6983267	25
2.6.4. Wariant rs944289	26
3. CELE PRACY	31
4. MATERIAŁ i METODY	32
4.1. Materiał.	32
4.1.1. Linie komórkowe	32
4.1.2. Badana populacja	33
4.1.3. Dawcy narządów	34
4.2. Metody	34
4.2.1. Odczynniki i aparatura	34
4.2.2. Roztwory	38
4.2.3. Izolacja DNA i RNA	39
4.2.4. Analiza bioinformatyczna	40
4.2.5. Oznaczanie genotypu polimorfizmu rs944289	41

4.2.6. Reakcja PCR z odwrotną transkryptazą	43
4.2.7. Szybka amplifikacja końców cDNA z użyciem ligazy RNA	45
4.2.8. Klonowanie wytypowanych transkryptów do wektora pcDNA3	49
4.2.9. Hybrydyzacja northern	50
4.2.10. Reakcja PCR z analizą ilości produktu w czasie rzeczywistym	51
4.2.11. Podwójny test lucyferazy	53
4.2.12. Test opóźnienia w żelu	55
4.2.13. Sekwencjonowanie wytypowanych transkryptów	57
4.2.14. Test żywotności (wzrostu) komórek	60
4.2.15. Analiza wpływu wytypowanych transkryptów na ekspresję genów przy użyciu mikromacierzy	62
4.2.16. Analiza statystyczna	62
5. WYNIKI	63
5.1. Analiza <i>in silico</i> regionu 14q13.3	63
5.2. Tkankowa ekspresja CB987890	65
5.3. Struktura CB987890	66
5.3.1. Szybka amplifikacja końców 5' i 3' CB987890 z użyciem ligazy RNA	66
5.3.2. Klonowanie <i>PTCSC3</i> do wektora pcDNA3	70
5.3.3. Analiza struktury <i>PTCSC3</i> metodą northern blot	70
5.3.4. Analiza <i>in silico</i> struktury <i>PTCSC3</i>	71
5.4. Ekspresja <i>PTCSC3</i> w liniach komórkowych raka tarczycy	73
5.5. Analiza genotypu rs944289 w badanej populacji	73
5.6. Wpływ polimorfizmu rs944289 na ekspresję <i>PTCSC3</i>	74
5.6.1. Ekspresja <i>PTCSC3</i> w tkance tarczycowej	75
5.6.2. Ekspresja <i>PTCSC3</i> w tkance guza	75
5.6.3. Analiza ekspresji <i>PTCSC3</i> wyrażonej jako stosunek transkrypcji w tkance guza do tkanki zdrowej	76
5.7. Regulacja transkrypcji <i>PTCSC3</i>	79

5.7.1. Analiza <i>in silico</i> regionu promotora <i>PTCSC3</i>	79
5.7.2. Test opóźnienia w żelu dla C/EBP α i C/EBP β	80
5.7.3. Regulacja genu reporterowego <i>luc2P</i> przez C/EBP α i C/EBP β	82
5.8. Sekwencjonowanie <i>PTCSC3</i>	83
5.9. Funkcja <i>PTCSC3</i>	84
5.9.1. Wpływ <i>PTCSC3</i> na żywotność (wzrost) komórek	84
5.9.2. Wpływ <i>PTCSC3</i> na ekspresję genów	85
6. DYSKUSJA	91
6.1. Znaczenie badań GWAS w identyfikacji wariantów związanych z chorobami nowotworowymi	91
6.2. Struktura <i>PTCSC3</i>	94
6.3. <i>PTCSC3</i> jako duży niekodujący RNA gen	95
6.4. Funkcja <i>PTCSC3</i>	98
6.5. Wpływ rs944289 na ekspresję <i>PTCSC3</i>	103
7. WNIOSKI	108
8. STRESZCZENIE	109
9. SUMMARY	111
10. PIŚMIENNICTWO	113

1. WYKAZ STOSOWANYCH SKRÓTÓW

1.1. Najczęściej stosowane skróty.

ATC - rak anaplastyczny tarczycy (ang. *anaplastic thyroid cancer*)

BCPAP - ludzka linia komórkowa raka brodawkowego tarczycy

C/EBP - białko wiążące się z sekwencją CCAAT (ang. *CCAAT/enhancer binding protein*)

CD-CV - częsta choroba-częsty wariant (ang. *common disease-common variant*)

CD-RV - częsta choroba-rzadki wariant (ang. *common disease-rare variant*)

cDNA - komplementarne DNA (ang. *complementary DNA*)

CNVs - zmienność liczby kopii (ang. *copy-number variations*)

COS-7 - linia komórkowa wywodząca się z nerki koczkodana zielonego (łac. *Cercopithecus aethiops*)

DNA - kwas deoksyrybonukleinowy (ang. *deoxyribonucleic acid*)

dNTP - deoksyrybonukleotyd (ang. *deoxyribonucleotide*)

DTC - rak wysokorozóżnicowany tarczycy (ang. *differentiated thyroid cancer*)

EMSA - test opóźnienia w żelu (ang. *electrophoretic mobility shift assay*)

EST - znaczniki ekspresji (ang. *expressed sequence tags*)

FRR - współczynnik rodzinnego ryzyka (ang. *family risk ratio*)

FTC - rak pęcherzykowy tarczycy (ang. *follicular thyroid cancer*)

GWAS - badanie asocjacyjne całego genomu (ang. *genome wide association study*)

HEK293 - linia komórkowa wywodząca się z nerki ludzkiego zarodka

IPA - oprogramowanie *Ingenuity Pathway Analysis*

LD - nierównowaga sprzężeń (ang. *linkage disequilibrium*)

lincRNA - duży międzygenowy niekodujący RNA (ang. *large intergenic non-coding RNA*)

LINE - elementy LINE (ang. *long interspersed nuclear elements*)

lncRNA - duże niekodujące RNA (ang. *large non coding RNA*)

LTR-ERV1 - elementy LTR-ERV1 (ang. *long terminal repeat endogenous retrovirus-like elements*)

MAF - częstość rzadszego allelu (ang. *minor allele frequency*)

mRNA - przekaźnikowy RNA (ang. *messenger RNA*)

MTC - rak rdzeniasty tarczycy (ang. *medullary thyroid cancer*)

ncRNA - niekodujące RNA (ang. *non-coding RNA*)

NMTC - rak nierdzeniasty tarczycy (ang. *non medullary thyroid cancer*)

nt- nukleotydy (ang. *nucleotides*)

ORF - ramka odczytu (ang. *open reading frame*)

PCR - reakcja łańcuchowa polimerazy (ang. *polymerase chain reaction*)

PRC - kompleks PRC (ang. *polycomb repressive complex*)

PTC - rak brodawkowy tarczycy (ang. *papillary thyroid carcinoma*)

PWM - wpływ pozycji w macierzy (ang. *position weight matrix*)

RLM-RACE - szybka amplifikacja końców cDNA z użyciem ligazy RNA (ang. *RNA ligase mediated rapid amplification of cDNA ends*)

RNA - kwas rybonukleinowy (ang. *ribonucleic acid*)

RT-PCR - reakcja PCR z odwrotną transkryptazą (ang. *reverse transcriptase-polymerase chain reaction*)

SINE - elementy SINE (ang. *short interspersed nuclear elements*)

SIR - współczynnik częstości występowania (ang. *standard incidence ratio*)

SNP - polimorfizm pojedynczego nukleotydu (ang. *single nucleotide polymorphism*)

STATU1 - białko *double-stranded RNA-binding protein* *Staufen homolog 1*

TC - rak tarczycy (ang. *thyroid cancer*)

TF - czynniki transkrypcyjne (ang. *transcription factors*)

TFBS - miejsce wiążące czynnik transkrypcyjny (ang. *transcription factor binding site*)

TPC-1 - ludzka linia komórkowa raka brodawkowego tarczycy

3'UTR - region 3' nie ulegający translacji (ang. *3' untranslated region*)

5'UTR - region 5' nie ulegający translacji (ang. *5' untranslated region*)

1.2. Skróty nazw genów omawianych w pracy.

Wszystkie nazwy genów zostały podane w języku angielskim.

AKT - *V-Akt Murine Thymoma Viral Oncogene Homolog*

ANRIL (CDKN2BAS) - *Antisense RNA in the INK4 Locus (CDKN2B antisense RNA 1)*

BRMS1L - *Breast Cancer Metastasis-Suppressor 1-Like*

CDK2/CDK4 - *Cyclin-Dependent Kinase 2/Cyclin-Dependent Kinase 4*

CDKN2A - *Cyclin-Dependent Kinase Inhibitor 2A*

CDKN2B - *Cyclin-Dependent Kinase Inhibitor 2B*

CEBPA - *CCAAT/enhancer Binding Protein (C/EBP), alpha*

CEBPB - *CCAAT/enhancer Binding Protein (C/EBP), beta*

CFH - *Complement Factor H*

FOXE1 - *Forkhead Box E1*

GAPDH - *Glyceraldehyde-3-Phosphate Dehydrogenase*

GAS5 - *Growth Arrest-Specific 5*

GDF15 - *Growth Differentiation Factor 15*

HNF3 β /FoxA2 - *Hepatocyte Nuclear Factor 3/Forkhead Box A2*

HOTAIR – *HOX Transcript Antisense RNA*

HSP90 - *Heat Shock Protein 90 kDa*

MALAT-1 - *Metastasis Associated Lung Adenocarcinoma Transcript 1*

MBIP - *MAP3K12 Binding Inhibitory Protein 1*

MEG3 - *Maternally Expressed 3*

MOAPI - *Modulator of Apoptosis 1*

NIS - *Sodium/Iodide Symporter*

NKX2-1 - *NK2 Homeobox 1*

NRG1 - *Neuregulin 1*

PI3K - *Phosphoinositide-3-Kinase*

PLK1 – *Polo-Like Kinase 1*

POU5F1B - *POU Class 5 Homeobox 1B*

PPAR γ - *Peroxisome Proliferator-Activated Receptor γ*

PTMA - *Prothymosin, alpha*

RASEF - RAS and EF-hand Domain Containing

RHOB - Ras Homolog Gene Family, Member B

S100A4 - S100 Calcium Binding Protein A4

SRA - Steroid Receptor RNA Activator

TG - Thyroglobulin

TPD52L1 - Tumor Protein D52-Like 1

TPO - Thyroid Peroxidase

TSHR - Thyroid Stimulating Hormone Receptor

TSIX - TSIX Transcript, XIST Antisense RNA

USF1 - Upstream Transcription Factor 1

USF2 - Upstream Transcription Factor 2

XIST - X (Inactive)-Specific Transcript

2. WSTĘP.

2.1. Epidemiologia raka tarczycy.

Szacuje się, że na całym świecie rozpoznaje się rocznie około 212 000 nowych przypadków raka tarczycy (ang. *thyroid cancer*, TC), co stanowi nieco poniżej 2% wszystkich pacjentów z rozpoznaną *de novo* chorobą nowotworową [1, 2]. O ile w ogólnej populacji pacjentów z chorobą rozrostową przypadki TC ciągle stanowią niewielki procent, to liczba rozpoznań tego raka w krajach uprzemysłowionych liczona rok do roku wykazuje (w przeciwieństwie do wielu innych guzów) silny trend wzrostowy [3, 4]. W USA od 2004 roku zapadalność na TC zwiększała się dla mężczyzn i kobiet odpowiednio o 5.5% i 6.6% na rok [5, 6]. W Polsce notuje się >1700 nowych zachorowań rocznie przy zapadalności dla kobiet i mężczyzn wynoszącej odpowiednio ~7.3/100000/rok i ~1.7/100000/rok [7]. Wydaje się, że wzrost częstości rozpoznań TC można wiązać (przynajmniej częściowo) z postępowaniem w technikach diagnostycznych, głównie ultrasonografii (ang. *ultrasonography*, USG) i biopsji cienkoigłowej (ang. *fine needle aspiration biopsy*, FNA) jaki dokonał się na przestrzeni ostatnich trzech dekad. Szerokie wprowadzenie USG do użytku klinicznego w latach '80 zrewolucjonizowało diagnostykę narządów miękkich dzięki jego nieinwazyjności, prostocie i dostępności, co w połączeniu z FNA pozwoliło efektywniej diagnozować podejrzaną o podłoże nowotworowe zmiany chorobowe w obrębie tarczycy. Analiza danych z USA z lat 1973-2002 pokazała, że w tym okresie zapadalność na TC wzrosła dwukrotnie, z czego około 87% wzrostu dotyczyło rozpoznań małych guzów (<2 cm) postaci brodawkowatej raka tarczycy (ang. *papillary thyroid cancer*, PTC). W tym samym okresie czasu całkowita śmiertelność z powodu TC nie uległa zmianie, co może sugerować, że wzrost zapadalności był efektem skuteczniejszej diagnostyki małych guzów niż realnego wzrostu częstości TC [8]. Za powyższą hipotezę dodatkowo przemawiają wyniki badań autopsyjnych, w których częstość nierozpoznanych za życia PTC była stosunkowo wysoka i sięgała powyżej 1/3 chorych [9-12]. Dane z amerykańskiego National Cancer Institute (Surveillance Epidemiology and End Results, SEER; <http://seer.cancer.gov/>) przedstawiające trendy zapadalności na najczęstsze nowotwory w USA wskazują, że w latach 1988-2005 dokonał się wzrost rozpoznań głównie małych guzów PTC (<1 cm), ale także guzów o średnicy >4 cm oraz postaci z regionalnymi i odległymi przerzutami [13]. Wydaje się więc, że dostępność

badania diagnostycznych może być głównym choć zapewne nie jedynym wytłumaczeniem wzrostu częstości PTC. Część badaczy sugeruje, że krzywa zachorowalności na PTC może odzwierciedlać wpływ innych czynników (środowiskowych i dietetycznych), które poprzez działania pro-onkogenne zwiększają liczbę rozpoznań PTC w krajach wysokorozwiniętych [14-16].

2.2. Klasyfikacja i charakterystyka kliniczna raka tarczycy.

Rak tarczycy jest najczęstszym nowotworem gruczołów endokrynnych. Z uwagi na rodzaj komórek, z których pochodzi dzielimy go na dwie zasadnicze grupy - raki rdzeniaste tarczycy (ang. *medullary thyroid cancer*, MTC) wywodzące się z komórek C produkujących kalcytoninę i nierdzeniaste TC (ang. *non medullary thyroid cancer*, NMTC) wywodzące się z komórek pęcherzykowych tarczycy. Histologicznie NMTC dzielimy na trzy typy: raki brodawkowate, raki pęcherzykowe (ang. *follicular thyroid cancer*, FTC) i raki anaplastyczne (niezróżnicowane) (ang. *anaplastic thyroid cancer*, ATC). PTC i FTC czasami określa się wspólnym mianem raków wysokozróżnicowanych tarczycy (ang. *differentiated thyroid cancer*, DTC), które w przeciwieństwie do ATC i MTC charakteryzują się lepszym rokowaniem oraz podobnym (pomimo pewnych biologicznych różnic) sposobem leczenia.

Rak pęcherzykowy stanowi około 10% wszystkich TC [17]. Obraz histologiczny FTC jest różnorodny, od stosunkowo dobrze zróżnicowanych komórek raka formujących pęcherzyki wypełnione koloidem, do niskozróżnicowanych komórek z nasiloną jądrową atypią, naciekaniami torebki guza i naczyń krwionośnych. W rozpoznaniu FTC kluczowe jest stwierdzenie infiltracji torebki guza i/lub naczyń krwionośnych, co odróżnia go od gruczolaka tarczycy. FTC charakteryzuje się stosunkowo wolnym wzrostem, rozsiewem drogą krwiopochodną głównie do kości i płuc, rzadziej do wątroby, ośrodkowego układu nerwowego, pęcherza moczowego i skóry [18].

Rak brodawkowaty jest najczęstszą postacią histologiczną TC stanowiącą około 80% wszystkich przypadków. W obrazie mikroskopowym komórki zazwyczaj układają się w struktury brodawkowate, otaczające rdzeń z naczyniem jedną lub kilkoma warstwami. Przyjmują kształt sześcienny. Istnieje ponadto odmiana pęcherzykowa, a także wysokokomórkowa. Podścielisko guza jest dobrze unaczynione, zawierające w połowie przypadków koncentryczne mikrozwapnienia („ciałka piaszczakowate”) [19]. Podobnie jak

w przypadku FTC jego wzrost jest powolny, aczkolwiek często wielogniskowy. Guz daje przerzuty drogą limfatyczną do okolicznych węzłów chłonnych, których powiększenie jest nierzadko pierwszym objawem choroby [7, 20, 21].

Rak anaplastyczny jest jednym z najbardziej złośliwych nowotworów. W obrazie histologicznym widoczne są lite pola składające się z małych, niezróżnicowanych komórek wymagające różnicowania z chłoniakami. Guz ten nie formuje struktur pęcherzykowych ani brodawkowatych [22]. Jego rozwój jest bardzo szybki, przerzuty odległe są zwykle obecne w momencie rozpoznania. Nowotwór ten słabo reaguje na chemioterapię i radioterapię [7].

Rak rdzeniasty tarczycy jest zaliczany do guzów neuroendokrynnych. W obrazie histologicznym widoczne są atypowe komórki zawierające w cytoplazmie neurosekrecyjne ziarnistości. W podścielisku obecne są zwykle złogi amyloidu, których ilość wzrasta w miarę wzrostu guza i przejścia z fazy bogatokomórkowej do ubogokomórkowej [19, 22]. MTC stanowi około 5% wszystkich przypadków TC [23]. Rokowanie u chorych z MTC jest złe. W około połowie przypadków dochodzi do zajęcia śródpiersiowych węzłów chłonnych i lokalnego rozprzestrzenienia się choroby do otaczających tarczycę narządów. U jednej piątej pojawiają się ogniska odległych przerzutów, głównie do kości, płuc i wątroby [7, 24].

2.3. Środowiskowe czynniki ryzyka rozwoju raka tarczycy.

Ekspozycja na promieniowanie radioaktywne (szczególnie w dzieciństwie) stanowi silny i utrzymujący się przez całe życie czynnik ryzyka zachorowania na PTC [25]. Dane zebrane po zrzućeniu bomb atomowych na Hiroszimę i Nagasaki oraz po katastrofie elektrowni atomowej w Czarnobylu wskazały na wyraźny wzrost zachorowań na PTC wśród napromienionych osób [26]. Wykazano, że postacie indukowane ekspozycją na promieniowanie (ang. *radiation induced PTC*) rozwijają się u młodszych chorych, charakteryzują się większą agresywnością oraz są mniej związane z płcią żeńską [27]. Wpływ promieniowania jest ważnym czynnikiem ryzyka rozwoju PTC ale wydaje się, że do indukcji nowotworu muszą zaistnieć dodatkowe warunki takie jak odpowiednie podłoże genetyczne, które może modulować osobniczą wrażliwość na promieniowanie. Przemawiają za tym dane wskazujące, że PTC rozwija się u stosunkowo niewielkiego odsetka osób narażonych na podobną dawkę promieniowania [28-30].

Uznany środowiskowym czynnikiem ryzyka rozwoju FTC jest niedobór jodu [31]. W regionach świata, w których występuje deficyt tego pierwiastka w diecie FTC jest dominującą postacią histologiczną [32, 33]. W krajach, w których wprowadzono obowiązek jodowania soli kuchennej zaobserwowano zmniejszenie częstości występowania FTC [34-37].

2.4. Podłoże molekularne raka tarczycy.

Czynniki predysponujące do nowotworów można podzielić na środowiskowe (ang. *environmental*) i genetyczne (ang. *genetic*). Te ostatnie można ocenić na podstawie retrospektywnej analizy porównawczej chorych z danym fenotypem i odpowiednio dobranej grupy kontrolnej (ang. *case-control studies*). W badaniach tych identyfikuje się probanta (z badanym nowotworem) oraz kontrolę (bez danego nowotworu), a następnie określa się występowanie lub brak występowania danego fenotypu u członków rodziny. Wyniki przedstawia się jako iloraz liczby obserwowanych przypadków choroby u krewnych probantów do oczekiwanej liczby przypadków choroby wyliczonej na podstawie występowania danego fenotypu u krewnych kontroli. Opisany współczynnik zwany jest ilorazem szans (ang. *odds ratio*, OR) lub współczynnikiem rodzinnego ryzyka (ang. *family risk ratio*, FRR) i jest miarą rodzinności (ang. *familiality*) danego fenotypu. W przypadku, kiedy grupę kontrolną stanowi cała populacja wskaźnik ten zwany jest standardowym współczynnikiem częstości występowania (ang. *standard incidence ratio*, SIR). Badania retrospektywne oceniające rodzinność danych fenotypów wymagają analizy bardzo dużych i dobrze opisanych populacji. Dwa najbardziej znane doniesienia oceniające rodzinność nowotworów pochodzą z analizy populacji ze stanu Utah w USA i mieszkańców Szwecji [38, 39]. W badaniu analizującym populację z Utah dopasowano dane z Utah Genealogic Database zawierającej opis genealogii mieszkańców tego stanu do danych z Utah Cancer Registry identyfikujących chorych z rozpoznanym nowotworem. Autorzy tego opracowania przedstawili wyniki dotyczące 28 nowotworów (w tym TC) oparte na analizie 35 228 probantów i ich rodzin oraz 399 786 krewnych I-go stopnia z odpowiednio dobranej grupy kontrolnej [38]. W drugim badaniu przeprowadzono analizę danych 9.6 miliona mieszkańców Szwecji zawartych w Swedish Family Cancer Database i Swedish Cancer Registry zawierającej opis ponad 700 000 chorych z rozpoznaną chorobą nowotworową [39].

W obydwu badaniach genetyczna predyspozycja do TC była jedną z najwyższych wśród wszystkich analizowanych nowotworów (odpowiednio FRR=8.48 i SIR=292.0). Z uwagi na nieco inny sposób przedstawienia końcowych wyników Neil Risch dokonał metaanalizy tych badań. Zunifikowana wartość współczynnika FRR dla TC wynosiła 8.48 (całkowite FRR) dla populacji Utah i 12.42 (FRR dla krewnych I-go stopnia) dla mieszkańców Szwecji. Dla porównania FRR dla raka płuc i piersi, czyli jednych z najczęstszych nowotworów w Polsce wynosił dla Utah i Szwecji, odpowiednio 2.55 i 1.83 oraz 3.16 i 2.01 (Tabela 1) [40].

Tabela 1. Rodzinność wybranych nowotworów wyrażona jako współczynnik rodzinnego ryzyka (*family risk ratio*, FRR) w populacji Utah i Szwecji [40]. Rak tarczycy charakteryzuje się jednym z najwyższych współczynników FRR w obydwu populacjach.

Nowotwór	Utah		Szwecja	
	Całkowity FRR	FRR dla wczesnego rozpoznania	FRR dla dziecka	FRR dla krewnego I-go stopnia
Prostaty	2.21	4.08	2.82	9.41
Piersi	1.83	3.70	1.86	2.01
Jelita grubego i odbytnicy	2.54	4.53	1.86	4.41
Płuc	2.55	2.50	1.68	3.16
Czerniak	2.10	6.43	2.50	3.41
Pęcherza moczowego	1.53	5.00	1.53	3.30
Chłoniak nieziarniczy	1.68	2.40	1.68	2.37
Mózgu/OUN	1.97	8.95	1.72	2.37
Jajników	2.04	-	2.94	2.52
Żołądka	2.46	-	1.72	8.82
Nerek	2.46	-	1.60	5.26
Tarczycy	8.48	-	9.52	12.42
Szpiczak mnogi	4.29	-	4.25	5.62
Białaczka granulocytowa	2.94	-	1.69	3.53
Narządów rodnych	2.22	-	2.85	3.97
Jąder	8.57	-	4.31	8.50
Wszystkie guzy	2.15	4.08	1.86	3.53

Zasadniczym pytaniem wynikającym z opisanych powyżej obserwacji było określenie na ile oceniana rodzinność wynikała z predyspozycji genetycznej, a na ile miała ona związek z wpływem czynników środowiskowych, które z natury rzeczy oddziałują na wszystkich członków nuklearnej rodziny. Analiza oceniająca proporcję wpływu genów i środowiska dla piętnastu najczęściej rozpoznawanych w Szwecji raków wykazała, że połowa rodzinności (53%) dla TC miała podłoże genetyczne - najwięcej ze wszystkich badanych nowotworów. Następnymi po TC nowotworami o najsilniejszej komponente genetycznej były guzy innych gruczołów endokrynych (28%), jąder (25%), piersi (25%) i szyjki macicy (22%) [41].

Ważnym ograniczeniem opisywanych badań było nierozróżnianie rodzinności w zależności od typu histologicznego TC co jest istotne w ocenie genetycznego podłoża tego nowotworu, ponieważ postać rdzeniasta w około 25% przypadków wykazuje autosomalnie dominujący sposób dziedziczenia mutacji w genie *RET*, a więc nadreprezentacja MTC w badanej populacji mogła zawyżyć wartości FRR [40]. Jednakże badania uwzględniające postać histologiczną wskazują, że rodzinność DTC pozostaje stosunkowo silna, aczkolwiek słabsza niż łącznie dla wszystkich typów TC. Dla przykładu w pracy Hemminki i wsp. oceniano genetyczne podłoże TC analizując 2435 chorych opisanych w latach 1958-1996 w Swedish Cancer Registry. Wartość SIR dla syna lub córki probanta z DTC wynosiła odpowiednio 7.8 (95% CI: 3.9-13.2) lub 2.8 (95% CI: 1.5-4.5) [42]. W badaniu Frich i wsp. analizującym 1025 przypadków TC zarejestrowanych w latach 1960-1995 w Norwegian Cancer Registry wykazano, że SIR dla postaci nierdzeniastej wynosił 5.2 (95% CI: 2.1-10.7) i 4.9 (95% CI: 3.0-7.7) odpowiednio dla krewnych I-go stopnia płci męskiej i żeńskiej. W tym samym badaniu, SIR dla krewnych I-go stopnia chorych z PTC wynosił 5.8 (95% CI: 2.1-12.6) dla mężczyzn i 4.1 (95% CI: 2.1-7.1) dla kobiet [43]. Podobne dane przedstawiono w badaniu Hrafinkelsson i wsp. analizującym 712 przypadków NMTC. Względne ryzyko (ang. *relative risk*, RR) dla krewnych I-go stopnia płci męskiej i żeńskiej wynosiło odpowiednio 4.1 (95% CI: 2.7-5.9) i 1.9 (95% CI: 1.3-2.7) (Tabela 2) [44].

W badaniach poświęconych rodzinności TC nie stwierdzono częstszego występowania tego nowotworu u współmałżonków, a więc osób, które dzielą to samo środowisko ale nie są spokrewnione genetycznie. Tym niemniej wpływ czynników środowiskowych oddziałujących na członków rodziny w dzieciństwie i okresie dojrzewania oraz styl życia, który zazwyczaj (w jakimś stopniu) jest powielany z pokolenia na pokolenie mogą mieć duże znaczenie w predyspozycji do NMTC. Dlatego w badaniach

dziedziczności używa się pojęcia „rodzinności”, które szerzej definiuje predysponujące czynniki. Badaniami, które mogłyby bardziej precyzyjnie odpowiedzieć na pytanie na ile geny, a na ile czynniki środowiskowe determinują rozwój NMTC są badania z udziałem bliźniąt jedno- i dwujajowych. W badaniach tych zakłada się, że wpływ środowiska i/lub stylu życia na oboje z bliźniąt jest bardzo zbliżony. W związku z tym, że bliźnięta monozygotyczne mają identyczny zestaw genów, podczas gdy bliźnięta dwuzygotyczne dzielą je w 50%, różnica współwystępowania TC u bliźniąt jedno- i dwujajowych obrazuje rolę czynnika genetycznego niezwiązanego z oddziaływaniem środowiska. Dla przykładu, jeśli zgodność występowania TC u bliźniąt jednojajowych byłaby większa niż u bliźniąt dwujajowych, świadczyłoby to o większym wpływie czynnika genetycznego na jego rozwój. Ograniczeniem tego typu badań jest konieczność analizy dużej liczby par bliźniąt, z których co najmniej jedno choruje na raka. W największej znanej tego typu analizie badającej 44 788 par zaobserwowano wzrost ryzyka dla raka żołądka, odbytnicy i jelita grubego, płuc, piersi i prostaty. Istotny statystycznie komponent dziedziczności uzyskano tylko dla trzech nowotworów (prostaty, piersi, odbytnicy i jelita grubego - ten ostatni analizowany łącznie jako „*colorectal cancer*”). Dla TC odnotowano 39 monozygotycznych i 63 dwuzygotycznych par. W żadnej nie stwierdzono współwystępowania TC. Dane te nie pozwalały na wyciągnięcie wniosków odnośnie stopnia dziedziczności NMTC, aczkolwiek mogą pośrednio sugerować, że wpływ genów nie jest decydujący o rozwoju choroby [45].

Tabela 2. Sumaryczne przedstawienie trzech retrospektywnych badań oceniających genetyczną predyspozycję do nierzemiastego raka tarczycy (NMTC).

Populacja	Typ histologiczny	Typ ryzyka	Ryzyko dla krewnych I-go stopnia		Referencja
			Mężczyźni	Kobiety	
Szwedzka	NMTC	SIR	7.8 (3.9-13.2)	2.8 (1.5-4.5)	[42]
Norweska	NMTC	SIR	5.2 (2.1-10.7)	4.9 (3.0-7.7)	[43]
	PTC	SIR	5.8 (2.1-12.6)	4.1 (2.1-7.1)	
Islandzka	NMTC	RR	4.1 (2.7-5.9)	1.9 (1.3-2.7)	[44]

2.5. Metody wykorzystywane w identyfikacji genów biorących udział w patogenezie chorób nowotworowych.

2.5.1. Analiza sprzężeń.

Analiza sprzężeń wykorzystuje znajomość zjawiska rekombinacji (ang. *cross-over*) czyli wymiany materiału genetycznego pomiędzy homologicznymi chromosomami zachodzącym w czasie podziału mejotycznego do lokalizacji loci związanych z daną chorobą. W analizie sprzężeń określa się współwystępowanie markerów genetycznych i badanego fenotypu u poszczególnych członków rodziny. Im odległość pomiędzy dwoma loci (markerem genetycznym i predysponującym do choroby) na chromosomie jest mniejsza, tym prawdopodobieństwo, że będą one dziedziczone razem (sprzężone) jest większe, ponieważ zmniejsza się prawdopodobieństwo wystąpienia rekombinacji i segregacji loci do różnych gamet. Miarą oddalenia loci jest częstość rekombinacji (θ), która przyjmuje wartości $0 \leq \theta \leq 0.5$ (0 =pełne sprzężenie; 0.5 =brak sprzężenia). Do oceny sprzężenia oblicza się współczynnik LOD (ang. *logarithm of odds*), który jest logarytmem dziesiętnym stosunku prawdopodobieństwa sprzężenia (θ) do prawdopodobieństwa niesprzężenia dwóch loci. Wynik $LOD \geq 3$ wskazuje na sprzężenie (szanse na sprzężenie 1000:1); wynik $LOD \leq -2$ świadczy przeciwko sprzężeniu (szanse na sprzężenie 1:100) [46]. Analiza sprzężeń jest przydatna w przypadkach, w których badany fenotyp występuje rodzinnie a geny inicjujące proces chorobowy wykazują dużą penetrację. Zaletą analizy sprzężeń jest brak konieczności znajomości patogenezy choroby. Wadą tej metody jest trudność wykrycia genów o średniej lub niskiej penetracji oraz konieczność poszukiwania do analizy dużych wielopokoleniowych rodzin.

2.5.2. Badania asocjacyjne.

W badaniach asocjacyjnych szuka się związku pomiędzy markerem genetycznym a daną chorobą porównując występowanie wybranych wariantów w populacji z badanym fenotypem i w odpowiednio dobranej grupie kontrolnej. Badania asocjacyjne z reguły wymagają znajomości patofizjologii choroby, ponieważ markery genetyczne selekcjonuje się w oparciu o lokalizację genów biorących udział w istotnych z punktu widzenia danej tkanki

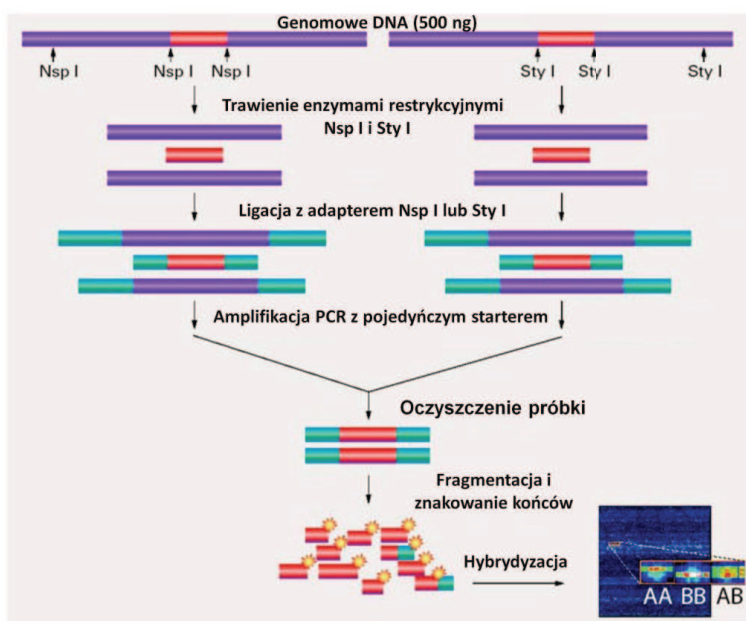
funkcjach takich jak różnicowanie tkanek, proliferacja czy apoptoza. Wybrane na podstawie postawionej hipotezy badawczej geny nazywa się genami kandydatami (ang. *candidate genes*). Zaletą badań asocjacyjnych w porównaniu z analizą sprzężeń jest większa łatwość znalezienia do badania odpowiednich populacji oraz możliwość skuteczniejszej identyfikacji genów o średniej lub małej penetracji. Wadą jest zwykle dokonywanie selekcji markerów w oparciu o regiony zawierające geny biorące udział w znanych szlakach komórkowych, co powoduje, że rzadko odkrywa się nowe ścieżki metaboliczne i nowe transkrypty zamieszcane w proces nowotworowy. Pozbawionymi tego ograniczenia są badania asocjacyjne całego genomu (ang. *genome wide association study, GWAS*), których szybki rozwój na przestrzeni ostatnich lat był jednym z ważniejszych osiągnięć współczesnej genetyki. Badania GWAS są tradycyjnymi badaniami asocjacyjnymi z tą różnicą, że nie dokonuje się wyboru genów kandydatów, tylko przy użyciu platform mikromacierzowych genotypuje się na pojedynczej płytce od 300 000 do 1 000 000 polimorfizmów rozmieszczonych na całej długości genomowego DNA. Dzięki tej technice można testować jednocześnie wiele tysięcy genów kandydatów. Kluczowa w projektowaniu badań GWAS jest znajomość regionów wykazujących nierównowagę sprzężeń, która pozwala przy pomocy stosunkowo niewielkiej liczby polimorfizmów i ich sprzężenia przeanalizować cały genom pod kątem zmienności genetycznej regionów będących w nierównowadze sprzężeń (ang. *linkage disequilibrium, LD*) i tym samym wskazać loci związane z ryzykiem wystąpienia danego fenotypu. Dane opisujące częstość występowania wariantów genetycznych i ich sprzężenia są zawarte w bazie HapMap (International HapMap Project, <http://hapmap.ncbi.nlm.nih.gov/>). Baza ta jest międzynarodowym projektem, który ma na celu poznanie i skatalogowanie wariantów polimorficznych ludzkiego genomu występujących w różnych rejonach świata. Zawiera ona dane o kilku milionach SNP'ów w populacji pochodzenia europejskiego, afrykańskiego i wschodnioazjatyckiego. Dzięki bazie HapMap można nakreślić mapę sprzężeń dla ludzkiego genomu oraz oszacować jaki jego procent będzie „pokryty” przez konkretny zestaw SNP'ów znajdujący się na danej płytce. Jako mikromacierzy używa się zestawów firm Affymetrix (USA) lub Illumina (USA). Te pierwsze (Affymetrix's Genome-Wide Human SNP Array 6.0; 1 852 600 markerów na płytce) posługują się techniką hybrydyzacji wyznakowanego DNA (<http://www.affymetrix.com/estore/>). Zestawy Illumina (HumanOmni2.5-Quad BeadChip; 2 379 855 wariantów na płytce) używają techniki hybrydyzacji sond z następczym przyłączaniem znakowanych nukleotydów (<http://www.illumina.com/>) (Rycina 1). W

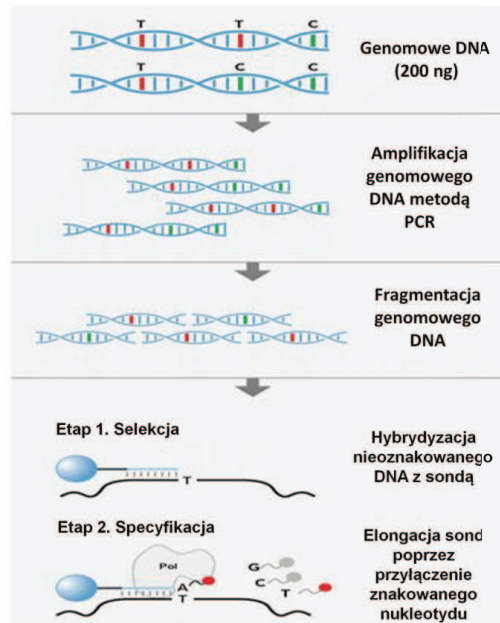
badaniach GWAS ważny jest dobór możliwie licznej grupy z danym fenotypem i bardzo staranne wyselekcjonowanie grupy kontrolnej pod kątem wieku, pochodzenia etnicznego i płci tak, by uniknąć rozwarstwienia (stratyfikacji) badanych populacji. Wyniki genotypowania przy użyciu mikromacierzy są poddawane szczegółowej kontroli jakości poprzez analizę odsetka osób z odczytanym genotypem, ocenę zgodności częstości alleli z prawem Hardy'ego-Weinberga, analizę niezgodności dziedziczenia wariantów sprzężonych z płcią oraz dziedziczenia alleli w relacji rodzic-dziecko. Przetworzone w ten sposób wyniki bada się za pomocą testu chi-kwadrat dla każdego wariantu z osobna. Z uwagi na stosunkowo duże prawdopodobieństwo otrzymania wyników fałszywie dodatnich zaleca się przyjmowanie wartości $p < 5 \times 10^{-8}$ jako istotnych statystycznie. Wszystkie warianty wykazujące związek z chorobą testuje się w dodatkowej populacji chorych i kontroli przy użyciu tradycyjnych metod genotypowania. Ostatnim etapem badania GWAS jest potwierdzenie związku zreplikowanych SNP'ów w niezależnych (walidacyjnych) populacjach chorych (Rycina 2).

W teorii chorób złożonych (wieloczynnikowych), a więc takich, w których wzajemna interakcja genów i czynników środowiskowych determinuje rozwój danego fenotypu istnieją dwie przeciwstawne hipotezy dotyczące ich genetycznego podłoża. Pierwsza zakłada, że w genezie tych chorób biorą udział liczne i często występujące w populacji warianty, z których każdy z osobna niesie stosunkowo małe ryzyko ale zsumowane razem i we współdziałaniu z czynnikami środowiskowymi mogą one wyzwolić proces chorobowy („częsta choroba-częsty wariant”; ang. „*common disease-common variant*”, CD-CV) [47]. W przypadku prawdziwości hipotezy CD-CV badania GWAS analizujące stosunkowo częste warianty (częstość rzadszego allelu $> 5\%$; ang. *minor allele frequency*, MAF) byłyby bardzo przydatne w wykrywaniu genów związanych z chorobami wieloczynnikowymi. Badania asocjacyjne całego genomu wykazały swoją skuteczność w identyfikacji transkryptów biorących udział w patogenezie chorób złożonych takich jak cukrzyca typu 2, udar niedokrwienny mózgu, choroba wieńcowa, stwardnienie rozsiane, rak piersi, reumatoidalne zapalenia stawów czy choroba Alzheimer'a [48-54]. Dla większości z powyższych chorób ryzyko jakie niósł predysponujący allel nie było wysokie ($1.2 < OR < 2.0$) i potwierdzało hipotezę CD-CV. Z drugiej strony niektórzy badacze sugerują, że w etiologii chorób złożonych biorą udział rzadkie (MAF $< 1\%$) ale niosące duże ryzyko rozwoju choroby warianty („częsta choroba-rzadki wariant”; ang. „*common disease-rare variant*”, CD-RV). Innymi słowy choroby

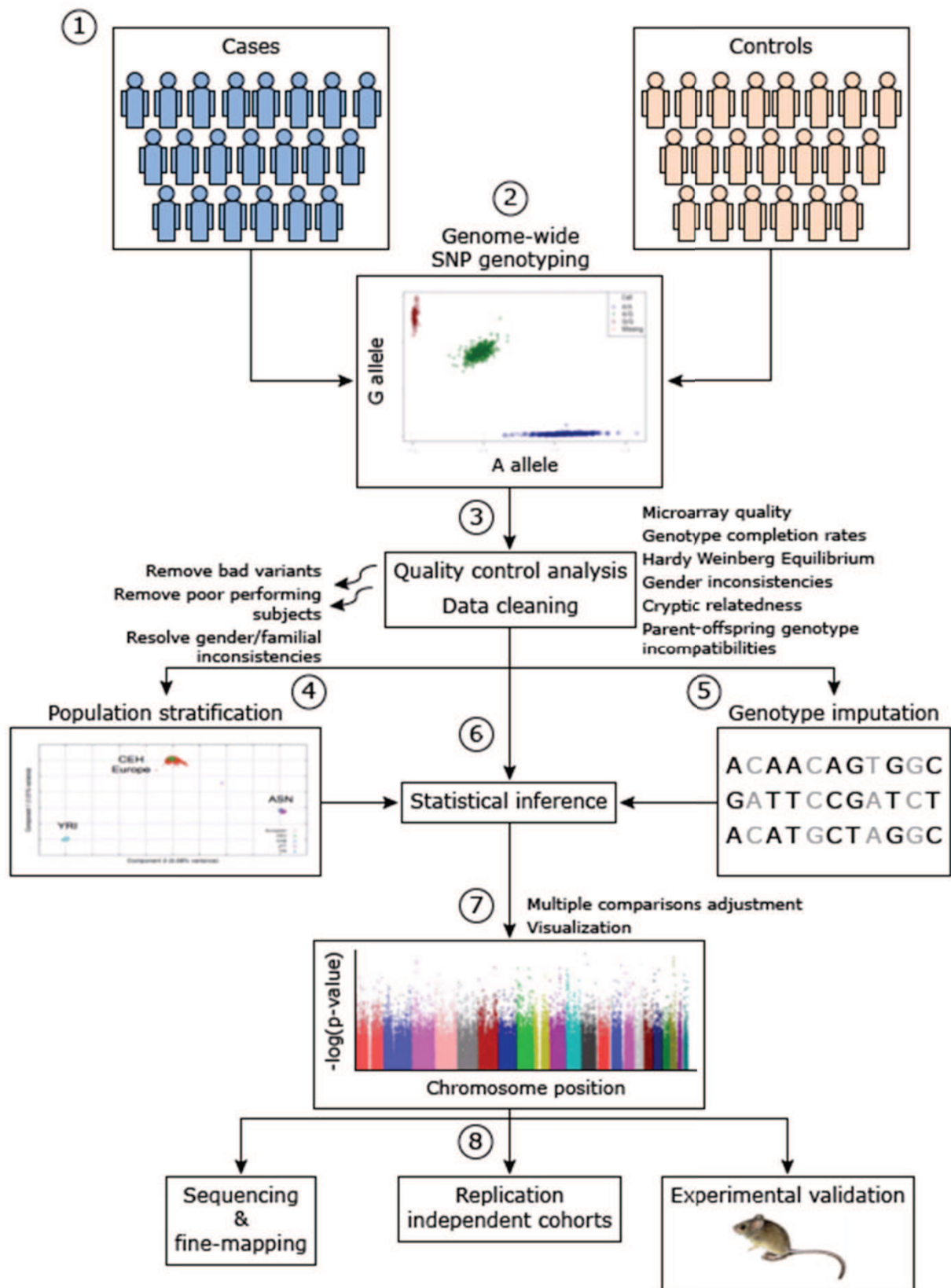
złożone są efektem występowania rzadkich ale licznych wariantów, z których każdy z osobna może odpowiadać za obserwowany fenotyp [55]. Zakładając prawdziwość tej teorii badania GWAS byłyby mniej przydatne w identyfikacji tych polimorfizmów i jedynie poprzez sekwencjonowanie całego genomu (ang. *whole-genome sequencing*) byłibyśmy w stanie wykryć predysponujące mutacje. Jednakże dane z dotychczas opublikowanych badań wskazują, że w chorobach wieloczynnikowych można wskazać zarówno częste polimorfizmy z małym ryzykiem jak i bardzo rzadkie, niosące duże ryzyko warianty. Sugeruje to, że nie ma dychotomii pomiędzy tymi dwiema teoriami i rola badań GWAS w poszukiwaniu częstych wariantów jest bardzo istotna [56, 57].

A



B

Rycina 1. Schemat genotypowania z wykorzystaniem mikromacierzy. (A) Mikromacierze firmy Affymetrix (USA) (<http://www.affymetrix.com/estore/>). (B) Mikromacierze firmy Illumina (USA) (<http://www.illumina.com/>).



Rycina 2. Schemat badania GWAS (Stages of a genome-wide association study, <http://www.uptodate.com/contents/search>). (1, 2) Genotypowanie chorych i kontroli przy użyciu mikromacierzy DNA. Na wykresie przedstawiono przykładowy wynik genotypowania dla pojedynczego SNP'u. Homozygoty z częstszym allelem [A] zostały przedstawione kolorem niebieskim, kolorem czerwonym zaznaczono homozygoty z rzadszym allelem [G], heterozygoty reprezentuje kolor zielony. (3) Kontrola jakości genotypowania przeprowadzona na poziomie wyników dla pojedynczego markera. (4) Analiza głównych składowych (ang. *principal component analysis*, PCA) celem wykrycia ewentualnego rozwarstwienia badanych populacji (stratyfikacja populacji). (5) Genotypowanie *in silico*. Typowanie genotypu (kolor szary) wariantów sąsiadujących z markerami o znanym dzięki mikromacierzom genotypie (kolor czarny). Wytypowane warianty włącza się do analizy asocjacji. (6) Wnioskowanie statystyczne (7) Prezentacja wyników za pomocą „wykresu Manhattan”, gdzie pojedyncze wyniki są przedstawiane jako funkcja $-\log_{10}p$ (oś Y) nałożonych na lokalizację danego markera na chromosomie (oś X). (8) Replikacja wyników w niezależnych populacjach chorych z ewentualną dalszą walidacją biologiczną i/lub sekwencjonowaniem.

2.6. Warianty polimorficzne predysponujące do wysokozróżnicowanego raka tarczycy.

Dane z licznych badań retrospektywnych wskazały na stosunkowo istotną rolę genów w patogenezie PTC [42-44]. Wydawało by się więc, że predysponujące mutacje powinny wykazywać dużą penetrację z mendlowskim typem dziedziczenia w obrębie dotkniętych tą chorobą rodzin. W istocie, odsetek chorych z wywiadem rodzinnym w kierunku PTC sięga około 10% i jak na tak rzadki nowotwór jest stosunkowo wysoki [58, 59]. Jednakże drzewa genealogiczne pokazujące mendlowski typ dziedziczenia są bardzo nieliczne, a rodziny z więcej niż pięcioma członkami z rozpoznanym PTC są wyjątkowo rzadkie. Co więcej, w rodzinach, w których mendlowski typ dziedziczenia wydaje się mieć zastosowanie zdarzają się generacje wolne od choroby. Powyższe obserwacje sugerują, że geny predysponujące do PTC wykazują niską penetrację oraz prawdopodobnie wzajemne współdziałanie, a ich interakcja z czynnikami środowiskowymi może odgrywać rolę w inicjacji nowotworu [60]. Dotychczas opublikowane badania analizujące sprzężenie genetyczne w rodzinach z PTC

wskazały na regiony, gdzie mogą znajdować się potencjalne mutacje (1q21 [61], 2q21 [62], 6q22 [63], 8p23 [64], 8q24 [65], 14q31 [66], 19p13.2 [67]) ale w żadnym z tych badań nie zidentyfikowano i opisano konkretnych genów. Większość przekonująco udokumentowanych polimorfizmów związanych z ryzykiem dla PTC wykryto przy pomocy badań asocjacyjnych (Tabela 3).

Tabela 3. Najważniejsze warianty genetyczne związane z predyspozycją do NMTC wskazane w dużych badaniach asocjacyjnych (>200 chorych, >300 kontroli).

Typ raka tarczycy	Locus	Wariant	OR	p	Badanie	Zreplikowano
NMTC	2q35	rs966423	1.34	1.3×10^{-9}	[68]	-
PTC	5q34	rs2910164	1.62	7.0×10^{-6}	[69]	Nie [70]
NMTC	8p12	rs2439302	1.36	2.0×10^{-9}	[68]	-
PTC	8q24.21	rs6983267	1.37	0.04	[71]	Tak [70] Nie [72, 73]
NMTC	8q24	rs1133076	1.60	0.015	[74]	-
PTC+FTC	9q22.33	rs965513	1.75	1.7×10^{-27}	[75]	Tak [70, 76, 77]
PTC+FTC	9q22	rs1867277	1.49	5.9×10^{-9}	[78]	Tak [70, 76]
PTC	11p11.2	rs4752904	1.61	0.0053	[79]	-
PTC+FTC	14q13.3	rs944289	1.37	2.0×10^{-9}	[75]	Tak [70, 77]
NMTC	14q13.3	rs116909374	2.09	4.6×10^{-11}	[68]	-
PTC+FTC	19q13.2	rs25487	0.70	0.03	[80]	Tak [81]

Najważniejsze z predysponujących do NMTC SNP'ów, które wymagają krótkiego omówienia to badania Jażdżewskiego i wsp. dotyczące polimorfizmu rs2910164 i jego wpływu na funkcje miR-146a [69, 82], badanie Landa i wsp. pokazujące wpływ wariantu rs1867277 zlokalizowanego w genie *FOXE1* na wiązanie czynników transkrypcyjnych USF1/USF2 [78], badanie Wokołorczyk i wsp. wykazujące związek rs6983267 z różnymi

nowotworami w tym z TC [71] oraz dwa asocjacyjne badania całego genomu, które wskazały pięć polimorfizmów (rs965513, rs944289, rs966423, rs2439302 i rs116909374) predysponujących do NMTC [68, 75].

2.6.1. Wariant rs2910164.

Predyspozycja polimorfizmu rs2910164 do PTC została po raz pierwszy opisana przez Jażdżewskiego i wsp. w roku 2008. W analizie 601 chorych z PTC i 901 kontroli badacze ci wykazali, że SNP rs2910164 znajdujący się 60 nukleotydów od początku sekwencji dla pre-miR-146a (w tak zwanej „nici pasażera”, ang. *passenger strand*) niesie zwiększone ryzyko dla heterozygot rs2910164[GC] (OR=1.62 (95% CI:1.3-2.0), p=0.000007) oraz zmniejszone dla homozygot rs2910164[CC] (OR=0.42 (95% CI:0.24-0.73), p=0.0027) i homozygot rs2910164[GG] (OR=0.69 (95% CI:0.57-0.85), p=0.0006) [69]. Analiza genotypu rs2910164 w tkance guza wykazała, że u 4.7% chorych doszło do spontanicznej mutacji somatycznej ze zmianą genotypu z formy homozygotycznej (CC lub GG) do postaci heterozygotycznej, a więc zwiększającej ryzyko dla PTC. W kolejnej pracy ci sami badacze wykazali, że nić pasażera u heterozygot ulega procesowaniu w podobny sposób jak nić przodująca (ang. *leading strand*) i ostatecznie u tych chorych można wykryć 3 formy dojrzałego mikroRNA (jedną z nici przodującej (miR-146a) i dwie z nici pasażera (miR-146a*G i miR-146a*C)), z których każda ma swój specyficzny zestaw docelowych genów. Ocena ekspresji transkryptów w zdrowej i nowotworowej tkance tarczycowej pokazała, że chorzy z genotypem rs2910164[GC] w porównaniu z pacjentami z genotypem rs2910164[GG] różnią się ekspresją genów regulujących procesy apoptozy (tkanka zdrowa i guza odpowiednio 27 genów, p=0.011 i 44 genów, p=0.0001), co prowadzi do zmienionej odpowiedzi komórkowej na uszkodzenie DNA i może wyjaśniać mechanizm predyspozycji heterozygotyczności [82]. Prace te były jednymi z pierwszych opartymi na analizie dużej populacji chorych i kontroli, wskazującymi na konkretny wariant genetyczny predysponujący do PTC oraz wyjaśniającymi w nowatorski sposób mechanizm predyspozycji. W dużym badaniu (781 chorych z NMTC i 6122 kontroli) analizującym pięć najważniejszych SNP'ów predysponujących do PTC potwierdzono związek dla czterech z nich. Jedynym wariantem, który nie wykazał asocjacji z NMTC był rs2910164 (genotyp CG: OR=1.032 (95% CI:0.876-1.214), p=0.728; genotyp CC: OR=0.987 (95% CI:0.682-1.384), p=0.985; allel

rs2910164[C]: OR=1.013 (95% CI:0.893-1.148), p=0.845) [70]. Subanaliza dla chorych z PTC dla różnych modeli ryzyka także pokazała brak związku rs2910164 z tą postacią histologiczną (p>0.784). Autorzy badania wskazali, że w badaniu Jażdżewskiego i wsp. widoczne było odchylenie od równania Hardy'ego-Weinberg'a dla genotypów chorych z PTC. Nie do końca jest jasne, czy był to rezultat czy przyczyna związku heterozygotyczności z PTC. Innym wyjaśnieniem mogą być osobnicze różnice pomiędzy badanymi populacjami aczkolwiek z uwagi na to, że w obydwu projektach analizowano chorych rasy kaukaskiej wydaje się to mało prawdopodobne.

2.6.2. Wariant rs1867277.

Gen *FOXE1* znajdujący się na chromosomie 9q22 koduje czynnik transkrypcyjny biorący udział w organogenezie i różnicowaniu tarczycy, utrzymywaniu postembrionalnej dojrzałości gruczołu tarczowego oraz regulacji ekspresji tarczycowo-specyficznych genów takich jak tyreoglobulina (TG) i tyreoperoksydaza (TPO) [83-86]. Mutacje w tym genie prowadzą do agenezji tarczycy oraz rozszczepu podniebienia [87, 88]. Wariant rs1867277 znajdujący się w regionie 5' nie ulegającym translacji (ang. *5' untranslated region*) *FOXE1* wykazał silny związek z TC w populacji 1097 chorych z NMTC i 1057 kontroli (OR=1.49 (95% CI:1.30-1.70), p=5.9x10⁻⁹) [78]. Allel ryzyka rs1867277[A] zmienił powinowactwo czynników transkrypcyjnych USF1/USF2 do wiązania w miejscu wariantu, co wpływało na aktywność promotora *FOXE1*. Predyspozycja rs1867277 do TC została potwierdzona w niezależnej populacji chorych z NMTC i chorych z PTC indukowanym promieniowaniem radioaktywnym [70, 76].

2.6.3. Wariant rs6983267.

SNP rs6983267 został po raz pierwszy opisany jako wariant predysponujący do raka prostaty i jelita grubego [89-91]. Marker ten znajduje się na chromosomie 8q24 w 19 kb bloku wykazującym silną nierównowagę sprzężeń. W badaniu Wokołorczyk i wsp. analizującym 7665 chorych i 1910 kontroli potwierdzono związek rs6983267 z rakiem prostaty i jelita grubego oraz wykazano, że wariant ten predysponuje także do raka nerek, tarczycy i krtani [71]. Dane z analizy 485 chorych z TC i 1910 kontroli wskazały na zwiększone ryzyko dla

osób z genotypem rs6983267[GG] (OR=1.37 (95% CI: 1.02-1.82), p=0.04). Autorzy publikacji zwracali uwagę, że istotność statystyczna tego związku była graniczna (prawdopodobnie z uwagi na małą liczebność badanej populacji) oraz dopuszczali możliwość, że opisywana asocjacja była fałszywie dodatnia. W dwóch kolejnych badaniach z Hiszpanii i USA (odpowiednio 398 chorych z NMTC i 479 kontroli oraz 344 chorych z PTC i 452 kontroli) nie znaleziono dowodu na związek rs6983267 z TC [72, 73]. Kolejna analiza przeprowadzona przez Jones i wsp. pokazała, że allel rs6983267[G] istotnie statystycznie zwiększał ryzyko dla PTC (OR=1.140 (95% CI: 1.025-1.268), p=0.016). Siła związku była największa dla modelu recesywnego predyspozycji ($p_{GG \text{ vs } GT+TT}=0.004$), czyli inaczej niż dla ryzyka jakie niesie ten polimorfizm dla innych nowotworów. Metaanaliza badania polskiego i brytyjskiego wskazała, że model recesywny ulega wzmocnieniu (rs6983267[GG] vs [GT], OR=1.215 (95% CI: 1.051-1.404), p=0.004) natomiast nie ma istotnej różnicy pomiędzy heterozygotami a homozygotami [TT] (OR=1.087 (95% CI:0.933-1.266), p=0.142) [70].

2.6.4. Wariant rs944289.

Pierwsze badanie asocjacyjne całego genomu opublikowane przez Gudmundsson i wsp. w roku 2009 analizowało związek pomiędzy DTC a 304 083 wariantami genetycznymi u 378 chorych i 37 196 kontroli [75]. Skan całego genomu wskazał na dziewięć potencjalnych polimorfizmów wykazujących asocjację z DTC, z których siedem było zlokalizowanych w jednym bloku LD na chromosomie 9q22.33 (rs965514[A], rs10759944[A], rs907580[A], rs10984103[A], rs925487[G], rs7024345[A], rs1443434[G]), a pozostałe dwa znajdowały się na chromosomach 14q13.3 (rs944289[T]) i 1p36.13 (rs622450[T]). Polimorfizmy te zostały pojedynczo genotypowane (ang. *single track assay*) w dodatkowej populacji 241 chorych z DTC przy użyciu zestawu Centaurus. Zbiorcza analiza wyników dla obu badanych grup wskazała na trzy warianty najsilniej związane z ryzykiem dla TC (rs965513[A]: OR=1.77 (95% CI:1.57-2.00), p=6.8x10⁻²⁰; rs10759944[A]: OR=1.77 (95% CI: 1.57-2.01), p=1.7x10⁻¹⁹; rs944289[T]: OR=1.44 (95% CI:1.26-1.63), p=2.5x10⁻⁸). Polimorfizmy rs965513[A] i rs10759944[A] zlokalizowane w bloku LD na chromosomie 9q22.33 były silnie ze sobą skorelowane ($r^2=1$ dla Utah CEPH HapMap i $r^2=0.998$ dla populacji islandzkiej) i do dalszej analizy został wybrany wariant rs965513. W analizie wieloczynnikowej regresji żaden z pozostałych pięciu wariantów będących w

nierównowadze sprzężeń z rs965513 (9q22.33) nie wykazał istotności statystycznej. Wyniki dla polimorfizmów rs965513 i rs944289 zostały zreplikowane w dwóch niezależnych populacjach (USA: PTC=342, kontrole=384; Hiszpania: DTC=90, kontrole=1343). Skumulowane ryzyko dla TC we wszystkich grupach (islandzkiej, amerykańskiej i hiszpańskiej) wynosiło OR=1.75 (95% CI:1.59-1.94), $p=1.7 \times 10^{-27}$ i OR=1.37 (95% CI: 1.24-1.52), $p=2.0 \times 10^{-9}$ dla odpowiednio rs965513[A] i rs944289[T] (Tabela 4).

Tabela 4. Wyniki badania asocjacyjnego całego genomu (GWAS) dla DTC opublikowane przez Gudmundsson i wsp. (2009) [75]. CTR=kontrola.

Wariant (locus) Populacja	Liczba		Allel ryzyka %		OR (95% CI)	p
	DTC	CTR	DTC	CTR		
rs965513[A] (9q22.33)						
Islandia	579	37196	49.0%	35.2%	1.77 (1.57-2.00)	6.8×10^{-20}
Hiszpania	89	1343	44.4%	34.2%	1.51 (1.13-2.09)	6.5×10^{-3}
USA	294	384	47.1%	32.9%	1.81 (1.45-2.26)	1.2×10^{-7}
Razem	962	38923	-	34.1%	1.75 (1.59-1.94)	1.7×10^{-27}
rs944289[T] (14q13.3)						
Islandia	574	37083	64.4%	55.8%	1.44 (1.26-1.63)	2.5×10^{-8}
Hiszpania	90	881	60.0%	56.9%	1.14 (0.83-1.55)	4.3×10^{-1}
USA	342	381	65.4%	59.1%	1.32 (1.06-1.63)	1.2×10^{-2}
Razem	1006	38345	-	57.3%	1.37 (1.24-1.52)	2.0×10^{-9}

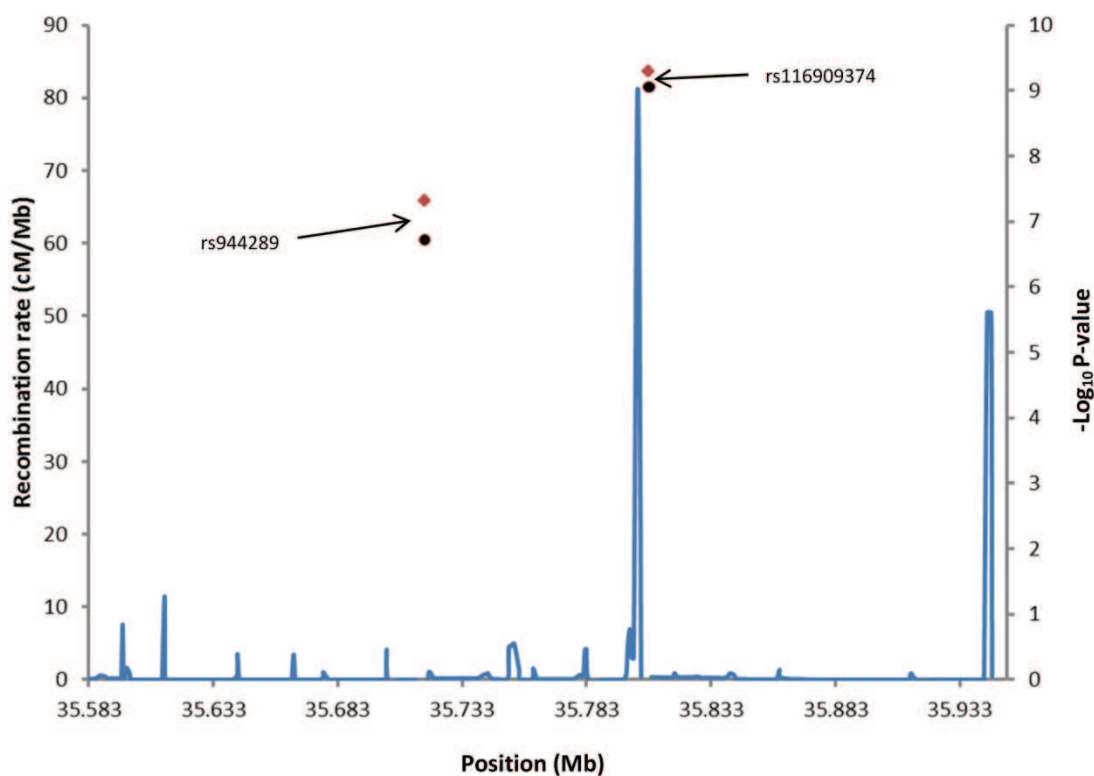
W badanej populacji 11% i 32% osób było homozygotycznymi nosicielami dla odpowiednio rs965513[A] i rs944289[T]. Homozygoty rs965513[A] i rs944289[T] miały zwiększone ryzyko dla TC o odpowiednio 3.1 i 1.9 raza. Osoby będące homozygotycznymi nosicielami obydwu wariantów (3.7% populacji) miały oszacowany OR na poziomie 5.7. Subanaliza dla podtypów histologicznych wykazała, że dla postaci brodawkowatych OR dla wariantu zlokalizowanego na chromosomie 9 (PTC=738, kontrole=38537) było nieznacznie wyższe (OR=1.80 (95% CI: 1.60-2.02), $p=4.7 \times 10^{-27}$), a dla polimorfizmu umiejscowionego

na chromosomie 14 (PTC=779, kontrole=38345) nieznacznie niższe (OR=1.32 (95% CI: 1.18-1.48), $p=2.0 \times 10^{-6}$). Dla FTC ryzyko jaki niósł wariant rs965513[A] (FTC=65, kontrole=38537) wynosiło OR=1.55 (95% CI: 1.09-2.20) ($p=0.016$), natomiast dla wariantu rs944289[T] (FTC=66, kontrole=37964) było wyższe i wynosiło OR=1.63 (95% CI: 1.14-2.33) ($p=0.0071$). Wyniki tej subanalizy wskazują, że obydwa SNP'y zwiększają ryzyko dla najczęstszych typów histologicznych TC.

Wariant rs944289 jest zlokalizowany w 249 kb bloku wykazującym nierównowagę sprzężeń, w którym nie opisano w bazie RefSeq żadnych genów. Najbliższe znane geny w sąsiedztwie rs944289 znajdują się 338 kb centromerycznie (*BRMS1L*) i 108 kb telomerycznie (*MBIP*) względem tego wariantu. Najbliższy tarczycowo-specyficzny gen (*NKX2-1*) znajduje się 338 kb od rs944289. Dla SNP rs965513 najbliższym genem jest *FOXO1*, który jest zlokalizowany 57 kb od tego polimorfizmu. Geny *NKX2-1* i *FOXO1* są ważnymi czynnikami transkrypcyjnymi regulującymi ekspresję innych tarczycowo-specyficznych genów [92]. Gudmundsson i wsp. analizowali asocjację obydwu wariantów z poziomami hormonu tyreotropowego (ang. *thyroid-stimulating hormone*, TSH), wolnej trijodotyroniny (ang. *free triiodothyronine*, fT_3) i wolnej tyroksyny (ang. *free thyroxine*, fT_4). Zarówno rs944289[T] jak i rs965513[A] wykazały związek z niższymi poziomami TSH o odpowiednio -1.7% ($p=0.030$) i -5.9% ($p=2.9 \times 10^{-14}$) dla pojedynczej kopii allelu ryzyka. Allel rs965513[A] wykazał związek z wyższymi poziomami fT_3 (+1.2%, $p=3.0 \times 10^{-3}$) i niższymi poziomami fT_4 (-1.2%, $p=6.1 \times 10^{-5}$). W drugim badaniu asocjacyjnym całego genomu opublikowanym w roku 2012 przez ten sam zespół badaczy wskazano na trzy kolejne warianty genetyczne, które wykazały związek z niskim poziomem TSH i ryzykiem dla TC [68]. W badaniu tym analizowano 27 758 osób bez rozpoznanego TC identyfikując 22 polimorfizmy wykazujące związek z obniżonym poziomem TSH. Wśród tych SNP'ów zidentyfikowano odkryty wcześniej wariant rs965513 zlokalizowany na chromosomie 9q22.33, który wyłączono z dalszej analizy. Na podstawie danych otrzymanych z genotypowania 21 wariantów wykazujących asocjację z niskim poziomem TSH u 561 chorych z DTC i 40 013 kontroli zidentyfikowano wstępnie pięć polimorfizmów wykazujących związek z TC, z których po włączeniu danych otrzymanych z niezależnych populacji walidacyjnych (DTC=595, kontrole=2604) zawężono ostatecznie do trzech najsilniej predysponujących do DTC (2q35, rs966423: OR=1.34 (95% CI: 1.22-1.47), $p=1.3 \times 10^{-9}$; 8p12, rs2439302: OR=1.36 (95% CI: 1.23-1.50), $p=2.0 \times 10^{-9}$; 14q13.3, rs116909374: OR=2.09 (95% CI: 1.68-2.60), $p=4.6 \times 10^{-9}$).

¹¹). Wariant znajdujący się na chromosomie 14 jest zlokalizowany 89 kb od polimorfizmu rs944289 opisanego w pierwszym badaniu GWAS. Oba SNP'y leżą w dwóch sąsiadujących lecz różnych regionach wykazujących nierównowagę sprzężeń. Analiza korelacji wskazała na bardzo słaby związek tych wariantów ($n=3693$, $r^2=0.005$). Zarówno rs944289 jak i rs116909374 wykazują niezależną od siebie istotność statystyczną (model bez wzajemnego dopasowania odpowiednio: $OR=1.36$, $p=4.9 \times 10^{-8}$ i $OR=2.07$, $p=5.0 \times 10^{-10}$; model z dopasowaniem odpowiednio: $OR=1.32$, $p=1.9 \times 10^{-7}$ i $OR=1.99$, $p=8.7 \times 10^{-10}$), co sugeruje, że region 14q13.3 zawiera dwa predysponujące niezależnie od siebie polimorfizmy o prawdopodobnie innym mechanizmie działania (Rycina 3).

Związek polimorfizmu rs944289[T] z NMTC został zreplikowany w dwóch niezależnych populacjach chorych. W pierwszym badaniu analizującym 507 pacjentów z PTC i 2766 kontroli z populacji japońskiej iloraz szans dla rs944289[T] wynosił $OR=1.21$ (95% CI: 1.04-1.39), $p=0.0121$ [70]. W drugim badając 781 chorych z NMTC i 6122 kontroli pochodzących z Wielkiej Brytanii także potwierdzono istotność statystyczną ryzyka jakie niesie rs944289[T] ($OR=1.33$ (95% CI: 1.19-1.49), $p=6.95 \times 10^{-7}$) [77]. Zbieżność wyników dla rs944289 w różnych populacjach w połączeniu z danymi dla wariantu rs116909374 wskazują na locus 14q13.3 jako jeden z ważniejszych regionów ludzkiego genomu związanych z predyspozycją do DTC.



Rycina 3. Wartości $-\log_{10}p$ dla związku rs944289 i rs116909374 z TC (czerwone diamnety: bez wzajemnego dopasowania; czarne punkty: z wzajemnym dopasowaniem) oraz wskaźnik rekombinacji (cM/Mb) w 375 kb regionie 14q13.3. Wskaźnik rekombinacji został obliczony w oparciu o dane z CEU HapMap II.22.

3. CELE PRACY

1. Identyfikacja potencjalnych genów zlokalizowanych w bliskim sąsiedztwie wariantu polimorficznego rs944289.
2. Strukturalna charakterystyka wytypowanych transkryptów zlokalizowanych w locus 14q13.3.
3. Określenie zależności pomiędzy wytypowanymi genami a wariantem rs944289.
4. Wstępne scharakteryzowanie biologicznej funkcji genów zidentyfikowanych w locus 14q13.3.

4. MATERIAŁ i METODY

4.1 MATERIAŁ.

4.1.1. Linie komórkowe.

Linia TPC-1.

Ludzką linię komórkową TPC-1 PTC otrzymano dzięki uprzejmości dr Matthew D. Ringel'a i dr Motoyasu Saji z Ohio State University (USA). Komórki hodowano w Dulbecco's Modified Eagle Medium (DMEM) firmy Gibco (USA) zawierającym 10% płodowej surowicy wołowej (ang. *fetal bovine serum*, FBS; Gibco, USA) oraz 1x endogennych aminokwasów (Gibco, USA).

Linia BCPAP.

Ludzką linię komórkową BCPAP PTC otrzymano dzięki uprzejmości dr Matthew D. Ringel'a i dr Motoyasu Saji z Ohio State University (USA). Komórki hodowano w Roswell Park Memorial Institute 1640 medium (RPMI 1640) (Gibco, USA) zawierającym 10% FBS (Gibco, USA).

Linia COS-7.

Linię komórkową COS-7 wywodzącą się z nerki koczodana zielonego (łac. *Cercopithecus aethiops*) zakupiono z firmy ATCC (nr katalogowy CRL-1651). Komórki hodowano w pożywce DMEM (Gibco, USA) suplementowaną 10% FBS (Gibco, USA).

Linia HEK293T.

Ludzką linię komórkową HEK293T wywodzącą się z komórek embrionalnych nerki zakupiono z firmy ATCC (nr katalogowy CRL-11268). Komórki HEK293T hodowano w DMEM (Gibco, USA) uzupełnionym 10% FBS (Gibco, USA).

Wszystkie linie komórkowe były hodowane w temperaturze 37°C, w atmosferze zawierającej 95% powietrza i 5% CO₂.

4.1.2. Badana populacja.

W pracy własnej analizowano RNA wyizolowane z pobranej w czasie zabiegu operacyjnego (a następnie zamrożonej i przechowywanej w ciekłym azocie) tkanki nowotworowej i zdrowej tkanki tarczycowej (pobranej z przeciwległego do lokalizacji guza płata tarczycy). W pracy analizowano także DNA wyizolowane z limfocytów krwi obwodowej. Materiał do badań pobrano od 46 chorych z rozpoznaniem histologicznie PTC. Wszyscy pacjenci po zapoznaniu się z protokołem badania wyrazili pisemną zgodę na udział w projekcie. Badanie uzyskało zgodę Komisji Bioetycznej działającej przy The Ohio State University (The Institutional Review Board at the Office of Responsible Research Practices Ohio State University, numer zgody 2006C0047). Charakterystykę badanej populacji przedstawiono w Tabeli 5.

W reakcjach sekwencjonowania (opisanych w rozdziale 4.2.13) użyto 33 próbek RNA (par zdrowa tkanka tarczycowa-tkanka PTC) dostępnych w laboratorium w momencie wykonywania eksperymentu (Tabela 6).

Tabela 5. Charakterystyka badanej populacji.

Charakterystyka	Pacjenci z PTC (n=46)
Wiek przy rozpoznaniu choroby (lata)	43.2 (SD±13.7)
Mężczyźni	8 (17.4%)
Typ histologiczny:	
• klasyczny	39 (84.8%)
• wariant pęcherzykowy	7 (15.2%)
Rasa:	
• kaukaska	45 (97.8%)
• azjatycka	1 (2.2%)

Tabela 6. Charakterystyka chorych z PTC, u których wykonano sekwencjonowanie wytypowanych transkryptów w zdrowej tkance tarczycowej i tkance guza.

Charakterystyka	Pacjenci z PTC (n=33)
Wiek przy rozpoznaniu choroby (lata)	43.1 (SD±14.8)
Mężczyźni	6 (18.2%)
Typ histologiczny: <ul style="list-style-type: none"> • klasyczny • wariant pęcherzykowy 	29 (87.9%) 4 (12.1%)
Rasa: <ul style="list-style-type: none"> • kaukaska • azjatycka 	32 (97.0%) 1 (3.0%)

4.1.3. Dawcy narządów.

W eksperymentach RLM-RACE (rozdział 4.2.7), reakcjach klonowania (rozdział 4.2.8) i northern blot (rozdział 4.2.9) i wykorzystano RNA wyizolowane z tkanki tarczycowej pobranej od dawców organów (Lifeline of Ohio; <http://www.lifelineofohio.org>). Ekstrakcja RNA została wykonana przy użyciu roztworu TRIzol (Ambion, USA) według protokołu zalecanego przez producenta.

4.2. METODY

4.2.1. Odczynniki i aparatura.

Tabela 7. Odczynniki użyte w badaniu.

Assay/Reagent	Zestaw/Odczynnik	nr katalogowy	Producent
Agilent Low Input One-Color Labeling	zestaw Agilent Low Input One-Color Labeling	5190-2305	Agilent Technologies
alamarBlue®	odczynnik alamarBlue®	DAL1100	Invitrogen
Amersham Hyperfilm™ MP	film Amersham Hyperfilm™ MP	28-9068-42	GE Healthcare Life Sciences

Assay/Reagent	Zestaw/Odczynnik	nr katalogowy	Producent
AmpliTaQ Gold® DNA Polymerase	polimeraza AmpliTaQ Gold® DNA	N8080258	Applied Biosystems
anti-C/EBP α	przeciwciało anty-C/EBP α	sc-61x	Santa Cruz Biotechnology Inc.
anti-C/EBP β	przeciwciało anty-C/EBP β	sc-150x	Santa Cruz Biotechnology Inc.
Bovine Serum Albumin (BSA)	surowicza albumina wołowa	B9001S	New England BioLabs® Inc
cOmplete Mini, EDTA-free Protease Inhibitor Cocktail Tablets	mieszanina inhibitorów proteaz	4693159001	Roche
Costar® 6 Well Clear TC-Treated Multiple Well Plates	płytki 6 dołkowa do hodowli komórkowej	3516	Corning Incorporated
Costar® 12 Well Clear TC-Treated Multiple Well Plates	płytki 12 dołkowa do hodowli komórkowej	3513	Corning Incorporated
Costar® 24 Well Clear TC-Treated Multiple Well Plates	płytki 24 dołkowa do hodowli komórkowej	3524	Corning Incorporated
Corning® 96 Well Flat Clear Bottom Black Polystyrene TC-Treated Microplates	płytki 96 dołkowa do hodowli komórkowej	3603	Corning Incorporated
DMEM	pożywka DMEM	11995-065	Gibco
DNA SurePrint G3 Human Gene Expression 8x60000 Arrays	zestaw mikromacierzowy DNA SurePrint G3 Human Gene Expression 8x60000 Arrays	G4851A (design ID 028004)	Agilent Technologies
DNA-free™ Kit	zestaw do reakcji oczyszczania próbek RNA z genomowego DNA	AM1906	Ambion
dNTP Mix (dATP, dCTP, dGTP, dTTP)	mieszanina dNTP (dATP, dCTP, dGTP, dTTP)	11814362001	Roche
DTT	ditiotreitol	V3151	Promega
Dual-Luciferase® Reporter Assay System	zestaw do reakcji podwójnej lucyferazy	E1910	Promega
E-Gel® CloneWell 0.8% SYBR® Safe	żel E-Gel® CloneWell 0.8% SYBR® Safe	G6618-08	Invitrogen
EDTA (0.5 M, pH 8.0)	EDTA (0.5 M, pH 8.0)	15694	USB (Affymetrix)
Ethidium Bromide Solution	bromek etydyny	161-0433	BIO-RAD
ExoSAP-IT® For PCR Product Clean-Up	Zestaw ExoSAP-IT® do oczyszczania reakcji PCR	78201	Affymetrix
Expand Long Template PCR System	zestaw PCR do powielania długich fragmentów DNA	11759060001	Roche
Fetal Bovine Serum	plodowa surowica wołowa	26140-079	Gibco
FirstChoice® Human Total RNA Survey Panel	zestaw RNA wyizolowanego z ludzkich tkanek	AM6000	Ambion
FirstChoice® RLM-RACE Kit	zestaw RLM-RACE	AM1700	Ambion
High Capacity cDNA Reverse Transcription Kit	zestaw do reakcji odwrotnej transkrypcji	4368813	Applied Biosystems
Hybond N	membrana Hybond N	RPN119N	GE Healthcare Life

Assay/Reagent	Zestaw/Odczynnik	nr katalogowy	Producent
			Sciences
illustra MicroSpin G-50 Columns	kolumny do oczyszczania radioaktywnej sondy	27-5330-01	GE Healthcare Life Sciences
Isoamyl alcohol	alkohol isoamylowy	BP1150	Fischer Scientific
KpnI HF	enzym restrykcyjny KpnI HF	R3142L	New England BioLabs® Inc
Lipofectamine® 2000 Transfection Reagent	Lipofektamina® 2000	52887	Invitrogen
MEM Non-Essential Amino Acids Solution	mieszanina endogennych aminokwasów	11140-050	Gibco
MiniElute PCR Purification Kit	zestaw do oczyszczania reakcji PCR	28006	QIAGEN
MOPS	MOPS	BP308	Fischer Scientific
NEB2 buffer	bufor NEB2	B7002S	New England BioLabs® Inc
NEB3 buffer	bufor NEB3	B7003S	New England BioLabs® Inc
NEB4 buffer	bufor NEB4	B7004S	New England BioLabs® Inc
Opti-MEM® Reduced Serum Medium	pożywka Opti-MEM®	31985-070	Gibco
pGL4.24	wektor pGL4.24	E8421	Promega
Phosphate-Buffered Saline (PBS)	roztwór soli fizjologicznej buforowanej fosforanem	10010-023	Gibco
phRL-TK	wektor phRL-TK	GenBank® Acc.#: AF362547	Promega
Poly(deoxyinosinic-deoxycytidylic) acid sodium salt	poli(dI-dC)	P4929	Sigma Aldrich
QIAGEN Plasmid Maxi Kit	zestaw do izolacji plazmidowego DNA z dużych (500 ml) kultur bakteryjnych	12163	QIAGEN
QIAprep Spin Miniprep Kit	zestaw do izolacji plazmidowego DNA z małych (5 ml) kultur bakteryjnych	27106	QIAGEN
RPMI 1640	pożywka RPMI	11875-093	Gibco
Salmon Sperm DNA	DNA z nasienia łososia	AM9680	Ambion
SAP	enzym SAP	78390	Affymetrix
SNaPshot® Multiplex Kit	zestaw SNaPshot® Multiplex Kit	4323161	Applied Biosystems
SYBR® Green PCR Master Mix	mieszanina SYBR® Green do reakcji real time PCR	4364346	Applied Biosystems
T4 DNA Ligase	ligaza T4 DNA	15224-025	Invitrogen
T4 DNA Ligase Buffer	bufor do ligazy T4 DNA	46300-018	Invitrogen
TEMED	TEMED	161-0801	Bio-Rad
TOPO® TA Cloning® Kit for Sequencing with One Shot® TOP10 Chemically Competent E. Coli	zestaw do klonowania TOPO® TA zawierający One Shot® TOP10 E.Coli	K4575	Invitrogen
TRIzol® Reagent	roztwór TRIzol®	15596018	Ambion
ULTRAhyb® Ultrasensitive	bufor ULTRAhyb® Ultrasensitive Hybridization	AM8670	Ambion

Assay/Reagent	Zestaw/Odczynnik	nr katalogowy	Producent
Hybridization Buffer	Buffer		
UltraPure™ 20X SSC	roztwór UltraPure™ 20X SSC	15557-044	Invitrogen
XhoI	enzym restrykcyjny XhoI	R0146L	New England BioLabs® Inc
[α - ³² P]-dCTP	[α - ³² P]-dCTP	BLU513H	Perkin Elmer
40% Acrylamide/Bis Solution	40% roztwór akrylamidu	161-0144	BIO-RAD
20% SDS	20% SDS	AM9820	Ambion
5x Passive Lysis Buffer	5x bufor lizujący do reakcji lucyferazy	E194	Promega

Tabela 8. Aparatura użyta w badaniu.

Appliance	Aparat	Producent
ABI Prism 7900HT Sequence Detection System	cykler ABI Prism 7900HT	Applied Biosystems
Agilent BioAnalyzer 2100	aparat Agilent BioAnalyzer 2100	Agilent Technologies
Agilent G2505C Microarray Scanner	skaner do mikromacierzy Agilent G2505C	Agilent Technologies
AlphaImager 3400 Gel Imaging System	system do obrazowania żeli AlphaImager 3400	Alpha Innotech
Applied Biosystems 3730 DNA Analyzers	sekwencjoner Applied Biosystems 3730	Applied Biosystems
E-Gel® iBase Power System	system E-Gel® iBase	Invitrogen
eppendorf centrifuge 5417R	wirówka eppendorf 5417R	Eppendorf
eppendorf centrifuge 5804	wirówka eppendorf 5804	Eppendorf
eppendorf centrifuge 5810R	wirówka eppendorf 5810R	Eppendorf
GeneAmp® PCR System 9700	cykler GeneAmp® PCR System 9700	Applied Biosystems
Incubator Isotemp 525D	inkubator Isotemp 525D	Fischer Scientific
Isotemp 210 waterbath	łaźnia wodna Isotemp 210	Fischer Scientific
Kodak X-Omat 2000A	aparat Kodak X-Omat 2000A	Commonwealth X-Ray Inc.
Sorvall RC6 centrifuge	wirówka Sorvall RC6	Thermo Scientific
SpectraMax M2	czytnik absorbancji i fluorescencji SpectraMax M2	Molecular Devices
Stratalinker 1800 UV	aparat Stratalinker 1800 UV	Stratagene, USA
Thermo Centra CL2 centrifuge	wirówka Thermo Centra CL2	Thermo Scientific
Thermo FORMA Class IIA/B3 Biological Safety Cabinet	wyciąg Thermo FORMA Class IIA/B3 Biological Safety Cabinet	Thermo Scientific
Thermo FORMA Series II Jacketed CO ₂ Incubator	inkubator Thermo FORMA Series II Jacketed CO ₂	Thermo Scientific
ThermoScientific MAXQ 500	inkubator ThermoScientific MAXQ 500	Thermo Scientific
Veritas Microplate Luminometer	luminometer Veritas Microplate	Turner Biosystems

4.2.2. Roztwory.

RCL bufor

8.4 g NH₄Cl
1.0 g KHCO₃
2.0 ml 0.5 M EDTA pH 8.0
rozpuść w 1000 ml dH₂O
dostosuj do pH 7.3-7.4 używając HCl lub NaOH
wysterylizuj przez przefiltrowanie (0.22 µm)

3M NaAc pH 5.5

24.6 g NaAc rozpuść w 80 ml dH₂O
dostosuj do pH 5.5 używając CH₃COOH
uzupełnij do 100 ml dodając dH₂O
wysterylizuj w autoklawie

Bufor A do EMSA

10 mmol Hepes-KOH (pH 7.9)
1.5 mmol MgCl₂
10 mmol KCl
0.5 mmol DTT

Bufor C do EMSA

20 mmol Hepes-KOH (pH 7.9)
25% glicerol
420 mmol NaCl
1.5 mmol MgCl₂
0.2 mmol EDTA
0.5 mmol DTT

SE buffer

4.4 g NaCl
9.3 g Na₂EDTA x 2H₂O
rozpuść w 950 ml dH₂O
dostosuj do pH 8.0 (HCl lub NaOH)
uzupełnij dH₂O do 1000 ml
wysterylizuj w autoklawie

CIA

1000 ml chloroformu (C)
42 ml alkoholu isoamyłowego (IA)

LoTE

3 ml 1 M Tris-HCl (pH 7.5)
400 µl 0.5 M EDTA (pH 8.0)
997 ml dH₂O
wysterylizuj w autoklawie

5x EMSA bufor

50 mmol HEPES-KOH (pH 7.9)
250 mmol MgCl₂
5 mmol DTT
10% glicerol

Bufor AB (annealing buffer)

20 µl 1 M Tris (pH 7.5)

10 µl 1 M MgCl₂

10 µl 5 M NaCl

10 µl 100 mM DTT

950 µl dH₂O

4% poliakrylamidowy żel do EMSA

12 ml 40 % akrylamidu (19:1)

0.75 ml 10% nadsiarczanu amonu

120 µl TEMED'u

83.7 ml dH₂O

1% formaldehydowy żel do northern blot

4.0 g agarozy

288.0 ml dH₂O

40.0 ml 10x MOPS'u

72 ml 37% formaldehydu

4.0 µl bromku etydyny

4.2.3. Izolacja DNA i RNA.

Izolację DNA przeprowadzono wykorzystując standardową (opisaną poniżej) procedurę z użyciem chloroformu i fenolu. Ekstrakcję RNA wykonano w oparciu o metodę opisaną przez Chomczyńskiego i wsp. wykorzystując roztwór TRIzol i protokół zalecany przez producenta (Ambion, USA) [93].

Izolacja DNA z limfocytów krwi obwodowej:

1. Zmieszać krew obwodową z 4 objętościami buforu RCL w probówce typu Falcon.
2. Inkubować próbkę przez 30 min. na lodzie mieszając delikatnie co 15 min.
3. Wirować (650 x g) przez 10 min. w temperaturze 4°C.
4. Usunąć supernatant.
5. Rozpuścić osad z leukocytów w 10 ml buforu SE przy pomocy sterylnej pipety.
6. Dodać 50 µl proteinazy K (10 mg/ml) i 500 µl 20% SDS.
7. Inkubować w temperaturze 37°C przez około 18 h.
8. Dodaj równoważną objętość fenolu. Mieszać przez 15 min.
9. Wirować (2600 x g) przez 10 min.

10. Przenieść supernatant do sterylnej probówki typu Falcon. Dodać równoważną objętość fenolu i CIA. Mieszać przez 15 min i odwirować jak powyżej.
11. Przenieść supernatant do sterylnej probówki typu Falcon. Dodać równoważną objętość CIA. Mieszać przez 15 min i odwirować jak powyżej.
12. Przenieść supernatant do nowej probówki typu Falcon. Dodać 1/30 objętości 3M NaAc (pH 5.5) i wymieszać.
13. Dodać 2.5 objętości 100% etanolu o temperaturze 4°C. Delikatnie wymieszać w celu wytrącenia DNA.
14. DNA powinno być widoczne w postaci „kłaczką”. Jeśli nie jest widoczne - inkubować próbkę w temperaturze –20°C przez kilka godzin, a następnie wirować (2600 x g) przez 20 min.
15. Przenieść DNA do nowej probówki typu eppendorf i przepłukać 80% etanolem.
16. Wirować (2600 x g) przez 10 min. w temperaturze 4°C.
17. Usunąć etanol.
18. Osuszyć osad pod wyciągiem.
19. Rozpuścić DNA w 250 µl buforu LoTE.

4.2.4. Analiza bioinformatyczna.

Do poszukiwania transkryptów zlokalizowanych w bliskim sąsiedztwie (± 40 kb) wariantu polimorficznego rs944289 wykorzystano bazę dbESTs (<http://www.ncbi.nlm.nih.gov/dbEST/>) zawierającą charakterystykę znaczników ekspresji (ang. *expressed sequence tags*, EST) oraz dane zawarte w UCSC Genome Browser, Feb.2009 (GRCh 37/hg19) (<http://genome.ucsc.edu/cgi-bin/hgGateway>) [94]. Program MATCH (<http://www.gene-regulation.com/pub/programs.html>) został użyty do identyfikacji czynników transkrypcyjnych (ang. *transcription factors*, TF) wiążących się *in silico* z sekwencją DNA zawierającą rs944289 [95]. W badaniu analizowano dwie 51 nukleotydowe sekwencje zawierające w centralnej pozycji rs944289[C] lub rs944289[T]. Dla każdej sekwencji szukano miejsc wiążących dla TF i obliczano ewentualne zamiany wyniku wpływu pozycji w macierzy (ang. *position weight matrix*, PWM) dla rs944289[C] i rs944289[T].

4.2.5. Oznaczanie genotypu polimorfizmu rs944289.

U każdego chorego włączonego do badania (n=46) oznaczono genotyp polimorfizmu rs944289 wykorzystując DNA wyizolowane z limfocytów krwi obwodowej. Reakcję genotypowania przeprowadzono przy pomocy zestawu SNaPshot firmy Applied Biosystems (USA) według protokołu zalecanego przez producenta (<http://www.appliedbiosystems.com>). Specyficzne startery oraz warunki reakcji amplifikacji, SNaPshot i oczyszczania produktów przedstawiono w Tabelach 9-15. Genotypowanie oczyszczonego produktu reakcji SNaPshot wykonano przy użyciu aparatu 3730 DNA Analyzer firmy Applied Biosystems. Wyniki analizowano przy pomocy oprogramowania GeneMapper[®]v3.7 (Applied Biosystems, USA).

Tabela 9. Specyficzne startery użyte w reakcji genotypowania metodą SNaPshot (F: starter sensowny; R: starter antysensowny; Ext: starter *extension*).

Starter	Sekwencja 5'→3'
rs944289_F	CTCTCTGGAGAACCAACAGGA
rs944289_R	GCCTGTGAATGGACATTAGATT
rs944289_Ext_38F	GTAACCTGCAATTTAATTTGGTTGAAAGATAGTCATTG

Tabela 10. Składniki reakcji PCR dla amplifikacji DNA zawierającego rs944289.

Składnik	Objętość (μl)
10x PCR bufor	1.0
dNTP (10 mM)	0.2
MgCl ₂ (25 mM)	0.4
Starter rs944289_F (10 μM)	0.6
Starter rs944289_R (10 μM)	0.6
Woda wolna od nukleaz	6.1
AmpliGold Taq polimeraza (5 U/μl)	0.1
DNA (20 ng/μl)	1.0
Suma	10.0

Tabela 11. Warunki PCR dla amplifikacji DNA zawierającego rs944289.

Temperatura	95°C	95°C	58°C	72°C	72°C
Czas	10 min.	30 sek.	30 sek.	1 min.	10 min.
Liczba cykli	1	30			1

Tabela 12. Skład mieszaniny dla reakcji oczyszczania produktu PCR przy użyciu enzymu ExoSAP-IT. Reakcję prowadzono w temperaturze 37°C przez 1 h z następczą dezaktywacją enzymu poprzez inkubację próbki w temperaturze 80°C przez 15 minut.

Składnik	Ilość (µl)
Produkt PCR	10.0
ExoSAP-IT	2.0
Suma	12.0

Tabela 13. Składniki reakcji SNaPshot.

Składnik	Objętość (µl)
SNaPshot Mix	1.0
10x PCR bufor	1.0
Starter rs94289_ext_38F (10 µM)	0.1
Woda wolna od nukleaz	4.9
Oczyszczony produkt PCR	3.0
Suma	10.0

Tabela 14. Warunki reakcji SNaPshot.

Temperatura	96°C	50°C	60°C
Czas	10 sek.	5 sek.	30 sek.
Liczba cykli	25		

Tabela 15. Skład mieszaniny dla reakcji oczyszczania produktu SNaPshot przy użyciu enzymu SAP. Reakcję prowadzono w temperaturze 37°C przez 1 h z następczą dezaktywacją enzymu poprzez inkubację próbki w temperaturze 80°C przez 15 minut.

Składnik	Ilość (μl)
Produkt SNaPshot	10.0
SAP	1.0
Suma	11.0

4.2.6. Reakcja PCR z odwrotną transkryptazą (ang. *reverse transcriptase polymerase chain reaction*, RT-PCR).

Ocenę tkankowej ekspresji wytypowanych transkryptów przeprowadzono przy użyciu reakcji łańcuchowej polimerazy z odwrotną transkryptazą (ang. *reverse transcriptase polymerase chain reaction*, RT-PCR). W reakcji tej polimeraza DNA zależna od RNA transkrybuje komplementarną do RNA nić DNA (ang. *complementary DNA*, cDNA). Następnie wykorzystując odpowiednio zaprojektowane startery amplifikuje się wybrane fragmenty cDNA. W reakcjach odwrotnej transkryptazy użyto zestawu High Capacity cDNA Reverse Transcription Kit firmy Applied Biosystem (USA). Skład mieszaniny i warunki reakcji przedstawiono w Tabelach 16 i 17. Materiałem biologicznym użytym w RT-PCR było dostępne komercyjnie RNA wyizolowane z ludzkich tkanek (FirstChoice[®] Human Total RNA Survey Panel, Ambion, USA). Jako genu referencyjnego użyto dehydrogenazy aldehydu-3-fosfoglicerynowego (*GAPDH*) (starter 5': 5'-CACCAACTGCTTAGCACCCC-3'; starter 3': 5'-TCTTCTGGGTGGCAGTGATG-3'). Reakcje amplifikacji przeprowadzono w termocyklerze GeneAmp[®] PCR System 9700 firmy Applied Biosystems (USA) używając mieszaniny opisanej w Tabeli 18 i warunków reakcji zawartych w Tabeli 19.

Tabela 16. Skład mieszaniny dla reakcji odwrotnej transkryptazy.

Składnik	Objętość (μl)
10x RT bufor	2.0
dNTP Mix (100 mM)	0.8
10x RT losowe startery	2.0
MultiScribe odwrotna transkryptaza (50 U/μl)	1.0
RNase inhibitor (40 U/μl)	1.0
Woda wolna od nukleaz	3.2
RNA (100 ng/μl)	10.0
Suma	20.0

Tabela 17. Warunki reakcji odwrotnej transkryptazy.

Temperatura	25°C	37°C	85°C	4°C
Czas	10 minut	120 minut	5 minut	∞
Liczba cykli	1			

Tabela 18. Składniki mieszaniny reakcji RT-PCR.

Składnik	Objętość (μl)
10x PCR bufor	2.5
dNTP (10 mM)	0.5
MgCl ₂ (25 mM)	1.0
Starter 5' (10 μM)	1.5
Starter 3' (10 μM)	1.5
Woda wolna od nukleaz	15.7
AmpliGold Taq polimeraza (5 U/μl)	0.3
cDNA	2.0
Suma	25.0

Tabela 19. Warunki reakcji RT-PCR.

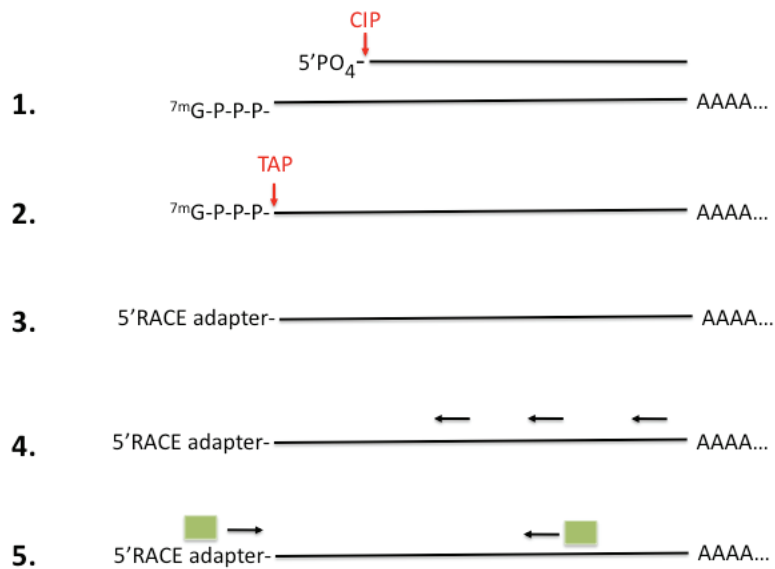
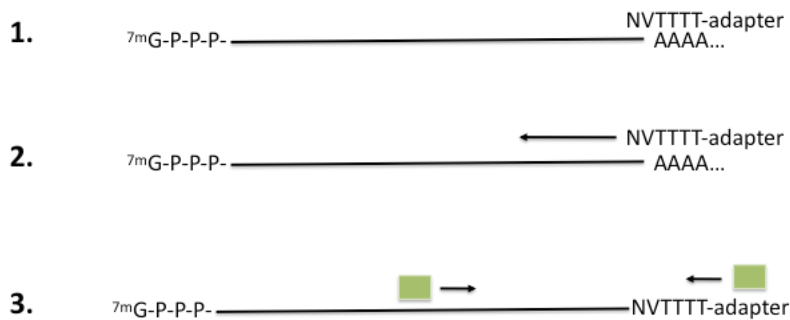
Temperatura	95°C	95°C	60°C	72°C	72°C
Czas	10 min.	30 sek.	30 sek.	1 min.	10 min.
Liczba cykli	1	30			1

4.2.7. Szybka amplifikacja końców cDNA z użyciem ligazy RNA (*RNA ligase mediated rapid amplification of cDNA ends*, RLM-RACE).

W celu określenia pełnej sekwencji wytypowanych transkryptów zastosowano szybką amplifikację końców 5' i 3' cDNA z użyciem ligazy RNA (*RNA ligase mediated rapid amplification of cDNA ends*, RLM-RACE). RACE jest reakcją typu PCR, która pozwala na powielenie pełnej nici cDNA w przypadku, gdy znany jest tylko fragment sekwencji danego transkryptu. Klasyczny 5'RACE opiera się na reakcji odwrotnej transkrypcji RNA na komplementarną nić DNA z następczą ligacją do końca 3' adaptera o znanej sekwencji lub dodania przy pomocy terminalnej transferazy dideoksynukleotydów. Następnie w reakcji PCR łączy się specyficzny dla danego genu starter ze starterem hybrydującym do adaptera/dideoksynukleotydów, co prowadzi do amplifikacji cDNA pomiędzy adapterem/dideoksynukleotydami (koniec 3') a specyficznym starterem (koniec 5'). Otrzymane produkty zawierają ostatecznie komplementarną sekwencję DNA od końca 5'RNA (3'adapter) do miejsca hybrydyzacji specyficznego startera. Największym ograniczeniem tradycyjnej techniki RACE jest udział w reakcji amplifikacji nici cDNA powstałych ze zdegradowanych mRNA, które zawierają niepełną sekwencję badanego transkryptu. Powielanie krótszych fragmentów jest zwykle bardziej wydajne, co powoduje, że klasyczny produkt RACE zawiera heterogenne amplikony o różnej długości z przewagą niepełnych sekwencji dla końca 5'. Szybka amplifikacja końców cDNA z użyciem ligazy RNA (RLM-RACE) zastosowana w niniejszej pracy pozwala uniknąć powielania niekompletnych fragmentów mRNA poprzez wcześniejsze usunięcie przy użyciu zasadowej fosfatazy (ang. *calf intestine alkaline phosphatase*, CIP) 5'-fosforanów ze zdefragmentowanych mRNA, tRNA, rybosomalnego RNA lub DNA. Następnie poprzez zastosowanie kwaśnej pirofosfatazy (ang. *tobacco acid pyrophosphatase*, TAP) usuwa się czapkę (ang. *cap*) z kompletnych nici mRNA pozostawiając na ich końcach 5'-

monofosforany. W kolejnym etapie dokonuje się przyłączenia za pomocą ligazy (T4 RNA ligaza) 45 nukleotydowego adaptera RNA do końca 5'. W reakcji tej konieczna jest obecność 5'-fosforanu, a więc do niekompletnych, uprzednio zdefosforylowanych nici mRNA ligacja adaptera nie zachodzi. W końcowym etapie przeprowadza się typową reakcję „zakotwiczonego” PCR (ang. *nested PCR*), w której używa się dwóch starterów zlokalizowanych na adapterze (zewnątrzny i wewnętrzny, ang. *outer* i *inner*) oraz dwóch specyficznych (zewnątrzny i wewnętrzny) dla sekwencji danego genu. Otrzymany produkt PCR (zwykle po wcześniejszym sklonowaniu) poddaje się sekwencjonowaniu.

Reakcja 3'RLM-RACE opiera się na hybrydyzacji do łańcucha poliA adaptera przy pomocy którego syntetyzuje się komplementarną nić cDNA. Adapter zawiera sekwencję do pary starterów (zewnątrzny i wewnętrzny), które łączy się w reakcji zakotwiczonego PCR z odpowiednimi starterami zlokalizowanymi na danym genie (zewnątrzny i wewnętrzny). Schemat procedury 5'RLM-RACE i 3'RLM-RACE przedstawiono na Rycinie 4. W reakcji amplifikacji końców 3' i 5' cDNA z użyciem ligazy RNA wykorzystano całkowite RNA (odpowiednio 1 µg i 10 µg) wyizolowane z tkanki tarczycowej pobranej od dawców narządów (Lifeline of Ohio; <http://www.lifelineofohio.org>). RNA procesowano według protokołu opisanego przez producenta zestawu (<https://products.invitrogen.com/ivgn/product/AM1700>). Reakcję zakotwiczonego PCR przeprowadzono używając zestawu Expand Long Template PCR System (Roche, USA) i odpowiednich par starterów. Skład mieszaniny i warunki PCR zostały przedstawione w Tabelach 20 i 21. Produkty reakcji PCR wklonowano do wektora pCR4-TOPO po uprzedniej adenylacji końców 3' (Skład mieszaniny: 10x PCR bufor, 50 mM dNTP, 1 U AmpliGold Taq polimeraza; inkubacja w 72°C przez 10 minut). Wklonowane do wektora pCR4-TOPO amplikony sekwencjonowano metodą Sanger'a używając starterów M13 sensownego (5'-GTAAAACGACGGCCAG-3') i M13 antysensownego (5'-CAGGAAACAGCTATGAC-3').

A**5'RLM-RACE****B****3'RLM-RACE**

Rycina 4. Schemat poszczególnych etapów procedury RLM-RACE. (A) 5'RLM-RACE. (1) Usunięcie 5'-fosforanów ze zdegradowanych mRNA, rRNA, tRNA i DNA poprzez zastosowanie zasadowej fosfatazy (CIP). (2) Usunięcie czapeczki (ang. *cap*) z pełnych nici mRNA przy użyciu kwaśnej pirofosfatazy (TAP). (3) Ligacja 5'-adaptera RNA. (4)

Odwrotna transkrypcja. (5) Reakcja PCR przy użyciu starterów hybrydujących do adaptera i specyficznych dla danego genu. (B) 3'RLM-RACE. (1) Hybrydyzacja adaptera do łańcucha poliA. (2) Odwrotna transkrypcja. (3) Reakcja PCR przy użyciu starterów łączących się do adaptera i specyficznych dla danego genu.

Tabela 20. Skład mieszaniny powielania długich fragmentów DNA (Expand Long Template PCR System, Roche, USA).

Składnik	Objętość (μl)
10x bufor 3	5.0
dNTP Mix (10 mM)	2.5
Starter 5' (10 μM)	1.5
Starter 3' (10 μM)	1.5
Polimerazy Taq i Tgo (5 U/μl)	0.5
Woda wolna od nukleaz	36.0
cDNA	3.0
Suma	50.0

Tabela 21. Warunki reakcji powielania długich fragmentów DNA (Expand Long Template PCR System, Roche, USA).

Temperatura	95°C	95°C	60°C	68°C	68°C
Czas	2 min.	10 sek.	30 sek.	3 minuty+20 sek. dla każdego kolejnego cyklu >10	5 min.
Liczba cykli	1	10+20			1

4.2.8. Klonowanie wytypowanych transkryptów do wektora pcDNA3.

Całkowite RNA wyizolowane z tkanki tarczycowej otrzymanej od dawców organów (Lifeline of Ohio; <http://www.lifelineofohio.org>) zostało przepisane na cDNA przy użyciu High Capacity Reverse Transcription Kit firmy Applied Biosystems (USA). Skład mieszaniny i warunki reakcji zostały opisane w rozdziale 4.2.6. Wytypowany transkrypt został powielony przy użyciu Expand Long Template PCR System (Roche, USA) ze starterami zawierającymi na końcach 5' miejsca trawienia dla enzymów restrykcyjnych (KpnI i XhoI dla odpowiednio startera 5' i 3') (Tabela 20 i 21). Produkt reakcji PCR został oczyszczony przy pomocy zestawu MiniElute PCR Purification Kit (Qiagen, USA), a następnie poddany trawieniu w temperaturze 37°C przez 1 h przy użyciu restryktaz KpnI HF i XhoI (New England BioLabs[®] Inc, USA). Skład mieszaniny dla reakcji trawienia został przedstawiony w Tabeli 22. Identyczną reakcję przeprowadzono dla wektora ekspresyjnego pcDNA3 (Invitrogen, USA). Strawiony produkt PCR i wektor zostały oczyszczone przy użyciu E-Gel[®] iBase Power System (Invitrogen, USA) na żelu E-Gel[®] CloneWell 0.8% SYBR[®] Safe. Następnie w reakcji ligacji (Tabela 23) wytypowany transkrypt wprowadzono do pcDNA3. Produkt ligacji transformowano do One Shot[®] TOP10E.coli (Invitrogen, USA) zgodnie z protokołem zalecanym przez producenta. Obecność wklonowanego amplikonu potwierdzono poprzez sekwencjonowanie DNA wyizolowanego z wytypowanych kolonii bakteryjnych (starter sensowny: 5'-TAATACGACTCACTATAGGG-3', starter antysensowny: 5'-AGAAGGCACAGTCGAGG-3').

Tabela 22. Skład mieszaniny dla reakcji trawienia. Mieszaninę inkubowano przez 1 h w temperaturze 37°C.

Składnik	Objętość (μl)
DNA (25 ng/μl)	40.0
KpnI HF (20 U/μl)	2.0
XhoI (20 U/μl)	2.0
10x NEB4 bufor	5.0
BSA (1 μg/μl)	1.0
Suma	50.0

Tabela 23. Skład mieszaniny dla reakcji ligacji. Mieszaninę inkubowano w temperaturze pokojowej przez 1 h.

Składnik	Objętość (μl)
Wektor (10 ng/μl)	1.0
Insert (20 ng/μl)	1.0
5x bufor dla T4 DNA ligazy	3.0
T4 DNA ligaza (1 U/μl)	0.5
Woda wolna od nukleaz	9.5
Suma	15.0

4.2.9. Hybrydyzacja northern.

W celu niezależnego od RLM-RACE określenia długości wytypowanych transkryptów zastosowano hybrydyzację northern (ang. *northern blot*, NB). Do reakcji NB wykorzystano 20 μg całkowitego RNA wyizolowanego z tkanki tarczycowej pobranej od dawców organów (Lifeline of Ohio; <http://www.lifelineofohio.org>). Jako negatywnej i pozytywnej kontroli użyto 20 μg RNA wyizolowanego z komórek HEK293T transfekowanych 24 h przed ekstrakcją odpowiednio pustym wektorem pcDNA3 lub wektorem pcDNA3 zawierającym badany gen. RNA rozdzielono na 1% formaldehydowym

zelu, a następnie transferowano na membranę Hybond N (GE Healthcare Life Sciences, USA) i związano przy użyciu Stratalinkera 1800 UV (Stratagene, USA). Jako znacznika użyto [α -P³²]-dCTP, który został wbudowany do sondy DNA za pomocą zestawu Random Primed Labeling Kit (Roche, USA) przy użyciu polimerazy Klenow. Skład mieszanki dla tej reakcji przedstawiono w Tabeli 24. Mieszaninę inkubowano w temperaturze 37°C przez 30 minut, a następnie po dodaniu 180 μ l dH₂O oczyszczono w kolumnie Quick Spin Column G-50 (GE Healthcare, USA) zgodnie z zaleceniami producenta. Sondę hybrydyzowano do RNA przez 18 godzin w temperaturze 42°C w roztworze UltraHyb Ultrasensitive Buffer (Ambion, USA). Następnie membrana została przemyta roztworem 2x SSC/0.2% SDS (2x20 minut w temperaturze 42°C) i poddana 24 godzinnej ekspozycji w temperaturze -70°C z filmem Amersham Hyperfilm™ MP (GE Healthcare, USA). Film został wywołany przy użyciu aparatu Kodak X-Omat 2000A firmy Commonwealth X-Ray Inc. (USA).

Tabela 24. Skład mieszaniny dla reakcji znakowania sondy DNA izotopem [α -P³²]-dCTP. dNTP=dATP, dGTP, dTTP.

Składnik	Objętość (μ l)
dNTP (2 mM)	3.0
10x NEB2 bufor	2.0
[α -P ³²]-dCTP (10 mCi/ml)	5.0
Klenow (2 U/ μ l)	1.0
DNA (25 ng)	9.0
Suma	20.0

4.2.10. Reakcja PCR z analizą ilości produktu w czasie rzeczywistym (ang. *real-time* PCR).

Ocenę ekspresji wytypowanych transkryptów przeprowadzono w oparciu o reakcję łańcuchową polimerazy z analizą ilości produktu w czasie rzeczywistym (ang. *real-time* PCR) używając SYBR Green jako fluorescencyjnego znacznika. Związek ten silnie wiąże się do dwuniciowego DNA absorbując promieniowanie o kolorze niebieskim (długość fali λ_{\max} =

497 nm) i emitując promieniowanie o kolorze zielonym (długość fali $\lambda_{\max}=520$ nm). Właściwości SYBR Green są wykorzystywane do pomiaru ilości zamplifikowanych cząsteczek w reakcji łańcuchowej polimerazy, gdzie w miarę wzrostu liczby kopii DNA w kolejnych cyklach PCR wzrasta ilość emitowanego promieniowania, które jest proporcjonalne do ilości powstających *de novo* amplikonów. Im większa liczba wprowadzonych kopii DNA do reakcji *real-time PCR*, tym natężenie promieniowania osiąga szybciej tak zwany cykl odcięcia (ang. *cycle threshold*, Ct), który jest arbitralnie wyznaczany na początku fazy logarytmicznej amplifikacji DNA. Ekspresję badanego transkrypty wyrażoną jako Ct_{target} normalizuje się do wyznaczonej w tej samej próbce ekspresji genu referencyjnego (Ct_{ref}), który w teorii wykazuje taki sam poziom ekspresji w różnych tkankach. Standaryzacja pozwala uniknąć błędów wynikających z niewielkich różnic w ilości wprowadzonego RNA do reakcji odwrotnej transkryptazy związanego z pipetowaniem. Metodą normalizacji zastosowaną w projekcie była relatywna ocena ekspresji określona wzorem $2^{-\Delta Ct}$, gdzie $\Delta=Ct_{\text{target}}-Ct_{\text{ref}}$ [96]. Jako genu referencyjnego użyto dehydrogenazy aldehydu-3-fosfoglicerynowego (*GAPDH*) (Tabela 36). Po zakończeniu każdej reakcji analizowano krzywe topnienia celem monitorowania obecności niespecyficzných produktów takich jak na przykład dimery starterów. Składniki mieszaniny i warunki reakcji *real-time PCR* przedstawiono w Tabeli 25 i 26.

Związek pomiędzy ekspresją wytypowanych transkryptów a genotypem rs944289 badano na próbie 46 chorych z rozpoznaniem PTC. Do oceny ekspresji badanych genów zostało użyte całkowite RNA wyizolowane z tkanki guza i tkanki tarczycowej nieobjętej procesem nowotworowym. RNA po oczyszczeniu z ewentualnej obecności genomowego DNA (DNA-free™ Kit, Ambion, USA) zostało przepisane na cDNA według protokołu opisanego w rozdziale 4.2.6. Specyficzne startery użyte w reakcji *real-time PCR* zostały opisane w rozdziale 5.6.

Tabela 25. Składniki mieszaniny dla reakcji *real-time* PCR.

Składnik	Objętość (μl)
2x SYBR [®] Green PCR Mix	5.0
Starter 5' (10 μM)	0.2
Starter 3' (10 μM)	0.2
Woda wolna od nukleaz	3.6
cDNA	1.0
Suma	10.0

Tabela 26. Warunki reakcji *real-time* PCR.

Temperatura	95°C	95°C	60°C
Czas	10 min.	15 sek.	1 min..
Liczba cykli	1	40	

4.2.11. Podwójny test lucyferazy (ang. *dual luciferase assay*).

1. Klonowanie fragmentów DNA zawierających rs944289[C] lub rs944289[T] do wektora reporterowego pGL4.24[luc2P/minP].

522 nukleotydowy insert został powielony z matrycy genomowego DNA wyizolowanego od homozygot [C] lub [T] przy pomocy starterów (sensowny: 5'-atcaggtaccGGCAATTGAAGTTCCTCCAAAA-3'; antysensowny : 5'-atcactcgagGCCTCCAGACTTGGACTGAG-3') zawierających na końcach 5' miejsca trawienia dla enzymów restyrykcyjnych KpnI i XhoI. Skład i warunki reakcji amplifikacji przedstawiono odpowiednio w Tabeli 27 i 28. Produkt reakcji PCR został oczyszczony przy pomocy zestawu MiniElute PCR Purification Kit (Qiagen, USA), a następnie strawiony przy użyciu enzymów restyrykcyjnych KpnI HF i XhoI (New England BioLabs[®] Inc, USA) w reakcji opisanej w Tabeli 22. Wektor reporterowy pGL4.24[luc2P/minP] (Promega, USA) zawierający gen dla lucyferazy świetlika (łac. *Photinus pyralis*) został strawiony przy użyciu

enzymów restrykcyjnych KpnI HF i XhoI (New England BioLabs[®] Inc, USA) (Tabela 22). Strawione produkt PCR i DNA wektora zostały oczyszczone na żelu E-Gel[®] CloneWell 0.8% SYBR[®] Safe przy użyciu E-Gel[®] iBase Power System (Invitrogen, USA). Reakcję ligacji przeprowadzono inkubując pGL4.24 i insert w mieszaninie opisanej w Tabeli 23. Produkt ligacji został transformowany do One Shot[®] TOP10E.coli (Invitrogen, USA) zgodnie z protokołem zalecanym przez producenta. Obecność wklonowanych do pGL4.24 amplikonów zawierających allel rs944289[C] lub rs944289[T] potwierdzono sekwencjonując DNA wyizolowane z wytypowanych kolonii bakteryjnych.

Tabela 27. Składniki mieszaniny dla reakcji powielania 522 nt insertu zawierającego rs944289.

Składnik	Objętość (μl)
10x PCR bufor	2.0
dNTP Mix (10 mM)	0.4
MgCl ₂ (25 mM)	0.8
Starter 5' (10 μM)	0.8
Starter 3' (10 μM)	0.8
Woda wolna od nukleaz	14.0
AmpliGold Taq polimeraza (5 U/μl)	0.2
DNA (20 ng/μl)	1.0
Suma	20.0

Tabela 28. Warunki reakcji dla powielania 522 nt insertu zawierającego rs944289.

Temperatura	95°C	95°C	60°C	72°C	72°C
Czas	10 min.	30 sek.	30 sek.	1 min.	10 min.
Liczba cykli	1	33			1

2. Regulacja genu reporterowego *luc2P* przez czynniki transkrypcyjne wiążące się *in silico* z sekwencją DNA zawierającą polimorfizm rs944289.

Komórki COS-7 zostały posiane na 24 dołkową płytkę w ilości 4×10^4 komórek/dołek. Po 24 h komórki transfekowano wektorem ekspresyjnym (100 ng) zawierającym wytypowany czynnik transkrypcyjny razem z wektorem pGL4.24 (50 ng), w którym gen reporterowy pozostawał pod kontrolą 522 nukleotydowego fragmentu DNA zawierającego allel rs944289[C] lub rs944289[T]. Do kontroli wydajności transfekcji użyto wektora phRL-TK (20 ng) kodującego gen lucyferazy żebroplawa (łac. *Renilla reniformis*). Jako nośnika DNA użyto 2 μ l Lipofektaminy 2000 (Invitrogen, USA) zgodnie z protokołem zalecanym przez producenta. Każdą transfekcję wykonano w 4 powtórzeniach. Po 24 godzinach komórki zostały poddane lizie przy użyciu 100 μ l Passive Lysis Buffer (Promega, USA). Pomiar luminescencji został wykonany w 20 μ l lizatu za pomocą aparatu Veritas Microplate Luminometer (Turner Biosystems, USA) przy wykorzystaniu Dual-Luciferase[®] Reporter Assay System (Promega, USA). Wyniki przedstawiono jako średnia \pm SEM z 6 niezależnych eksperymentów.

4.2.12. Test opóźnienia w żelu (ang. *electrophoretic mobility shift assay*, EMSA).

Celem oceny wiązania wytypowanych (na podstawie analizy *in silico*) TF do DNA zawierającego polimorfizm rs944289 użyto testu opóźnienia w żelu (EMSA). W technice tej wykorzystuje się fakt, że kompleks kwas nukleinowy-białko migruje w żelu wolniej niż niezwiązany z białkiem DNA (ang. *shift*). Specyficzność kompleksu DNA-białko można określić dodając do mieszaniny reakcyjnej odpowiednie dla badanego czynnika transkrypcyjnego przeciwciało. Kompleks DNA-białko-przeciwciało migruje wolniej niż kompleks DNA-białko (ang. *super-shift*).

W teście opóźnienia w żelu wykonanym w niniejszej pracy komórki HEK293T posiano na 100 mm płytkę i po upływie 24 h transfekowano (przy konfluencji 80%) wektorami zawierającymi analizowane czynniki transkrypcyjne (3 μ g) przy użyciu Lipofektaminy 2000 (6 μ l) zgodnie z protokołem zalecanym przez producenta (Invitrogen, USA). Po 24 godzinach komórki zostały przemyte lodowatym roztworem PBS (Gibco, USA), zawieszono w hypotonicznym buforze A uzupełnionym mieszaniną inhibitorów

proteaz (Complete Mini Protease Inhibitor Mixture, Roche, USA) i inkubowane na lodzie przez 15 minut. Następnie dokonano lizy komórek poprzez 10-krotny pasaż przez igłę [26G × 1/2" (0.45 × 12mm)]. Jądra komórkowe wyizolowano poprzez wirowanie mieszaniny z szybkością 18 000 x g przez 20 sekund w temperaturze 4°C. Po usunięciu supernatantu osad (składający się z jąder komórkowych) zawieszono w buforze C w ilości odpowiadającej 2/3 objętości pelletu i inkubowano na lodzie przez 30 minut. Uwolnione z jąder komórkowych białka wyizolowano poprzez wirowanie z szybkością 18 000 x g przez 5 minut w temperaturze 4°C. Supernatant zawierający proteinowy ekstrakt przeniesiono do nowych probówek i zamrożono w temperaturze -80°C do czasu wykonania testu opóźnienia w żelu. Jako sondy użyto oligonukleotydów (5'- gGAAAGATAGTCATTG**C**AGATTTGTAATA-3' i 5'- ggTATTACAAATCTGCAATGACTATCTTT-3'), które zostały zhybrydyzowane w reakcji opisanej w Tabeli 29 tworząc dwuniciową sondę. Końce sondy oznakowano [α -P³²]-dCTP przy pomocy polimerazy Klenow (Tabela 30). Charakterystyka tej reakcji powoduje, że tylko dwuniciowe (zhybrydyzowane) oligonukleotydy ulegają oznakowaniu radioaktywnym dCTP. Reakcje wiązania sondy i czynników transkrypcyjnych przeprowadzono inkubując mieszaninę opisaną w Tabeli 31 przez 20 minut na lodzie z następczym rozdzieleniem jej produktów na 4% poliakrylamidowym żelu (150 V, 4°C, 2 h).

Tabela 29. Skład mieszaniny dla reakcji hybrydyzacji komplementarnych oligonukleotydów (Oligo 1 i Oligo 2). Mieszaninę inkubowano przez 5 minut w temperaturze 95°C, a następnie powoli schłodzono do temperatury pokojowej.

Składnik	Objętość (μl)
Oligo 1 (1 μg/μl)	10.0
Oligo 2 (1 μg/μl)	10.0
10x bufor AB	10.0
Woda wolna od nukleaz	70.0
Suma	100.0

Tabela 30. Skład mieszaniny dla reakcji znakowania dwuniciowych oligonukleotydów (dsOligo) [α -P³²]-dCTP. Mieszaninę inkubowano przez 30 minut w 37°C. dNTP=dATP, dGTP, dTTP.

Składnik	Ilość (μ l)
dsOligo (50 ng/ μ l)	1.0
10x NEB2 bufor	5.0
dNTP (2 mM)	1.0
[α -P ³²]-dCTP (10 mCi/ml)	5.0
Klenow (2 U/ μ l)	1.0
Woda wolna od nukleaz	37.0
Suma	50.0

Tabela 31. Skład mieszaniny dla reakcji EMSA. Ekstrakt z jąder komórkowych i odpowiednie przeciwciała dodano do wybranych reakcji (patrz Rycina 19). W reakcjach pozbawionych tych składników dodano równoważne objętości wody.

Składnik	Objętość (μ l)
Sonda DNA znakowana [α -P ³²]-dCTP	1.0 (50.000 cpm)
Ekstrakt z jąder komórkowych (10 μ g)	1.0
5x EMSA bufor	5.0
Acetylowany BSA (1 μ g/ μ l)	1.0
poli(dI-dC) (0.5 μ g/ μ l)	3.0
Przeciwciało (2 μ g/ μ l)	1.0
Woda wolna od nukleaz	14.0
Suma	25.0

4.2.13. Sekwencjonowanie wytypowanych transkryptów.

Wytypowane i scharakteryzowane strukturalnie dzięki RLM-RACE i northern blot transkrypty poddano sekwencjonowaniu metodą dideoksy opisaną przez Fredericka Sangera

[97, 98]. Metoda dideoksy opiera się na amplifikacji analizowanego fragmentu DNA w mieszaninie dNTP i dideoksynukleotydów (ddNTP), w której losowe włączenie do syntetyzowanego łańcucha ddNTP kończy reakcję polimerazy. Dzięki elektroforetycznemu rozdzielaniu powstałych fragmentów DNA z następczą analizą długości emitowanej fali świetlnej przez wbudowany do ddNTP fluorescencyjny znacznik porządkuje się je pod względem wielkości i rodzaju terminalnego dideoksynukleotydu (kolor czarny: guanina; kolor czerwony: tymina; kolor zielony: adenina; kolor niebieski: cytozyna).

Do sekwencjonowania wykorzystano całkowite RNA wyizolowane z tkanki tarczycowej i tkanki guza od chorych z PTC (n=33). RNA oczyszczono z genomowego DNA (DNA-free™ Kit, Ambion, USA), a następnie przepisano przy pomocy High Capacity cDNA Reverse Transcription Kit (Roche, USA) na cDNA zgodnie z protokołem opisanym w rozdziale 4.2.6. W reakcji sekwencjonowania użyto oczyszczonego przy użyciu ExoSap-IT (Affymetrix, USA) produktu PCR powstałego w reakcji amplifikacji cDNA z odpowiednimi dla danego transkryptu starterami (Tabela 12 oraz rozdział 5.8). Reakcje powielania wytypowanych transkryptów przeprowadzono przy użyciu Expand Long Template PCR System firmy Roche (USA). Skład i warunki reakcji PCR zawarto w Tabeli 32 i 33. Sekwencjonowanie przeprowadzono wykorzystując ABI Prism BigDye Terminator Cycle Sequencing Kit version 3.1 (Applied Biosystems, nr kat 4337455) w 48 kapilarnym sekwencjonerze Applied Biosystems 3730 DNA Analyzers (Applied Biosystems, USA) zgodnie z protokołem zaleconym przez producenta. Skład mieszaniny dla reakcji sekwencjonowania przedstawiono w Tabeli 34. Dane analizowano przy pomocy ABI PRISM® DNA Sequencing Analysis Software v.5.0 (Applied Biosystem, nr kat. 4331940).

Tabela 32. Skład mieszaniny dla reakcji Expand Long Template PCR, w której powielano wytypowane do sekwencjonowania transkrypty.

Składnik	Ilość (μl)
10x bufor 3	2.5
dNTP Mix (10 mM)	1.3
Starter 5' (10 μM)	0.7
Starter 3' (10 μM)	0.7
Polimerazy Taq i Tgo (5 U/μl)	0.3
Woda wolna od nukleaz	17.5
cDNA	2.0
Suma	25.0

Tabela 33. Warunki reakcji Expand Long Template PCR, w której powielano wytypowane do sekwencjonowania transkrypty.

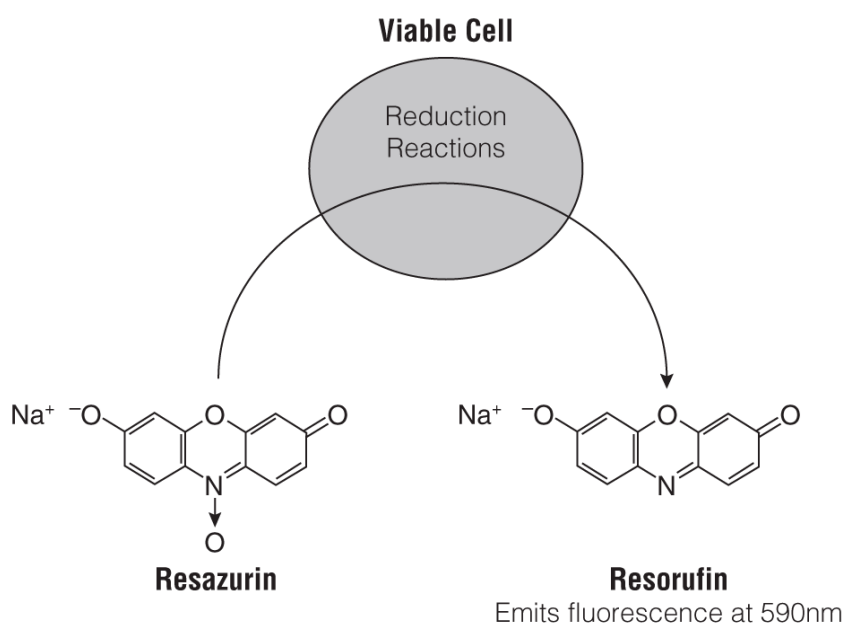
Temperatura	95°C	95°C	60°C	68°C	68°C
Czas	2 min.	15 sek.	30 sek.	3 minuty+20 sek. dla każdego kolejnego cyklu >10	7 min.
Liczba cykli	1	10+20			1

Tabela 34. Skład mieszaniny dla reakcji sekwencjonowania wytypowanych transkryptów.

Składnik	Ilość (μl)
Oczyszczony product PCR	5.4
Starter (10 μM)	0.6
Woda wolna od nukleaz	6.0
Suma	12.0

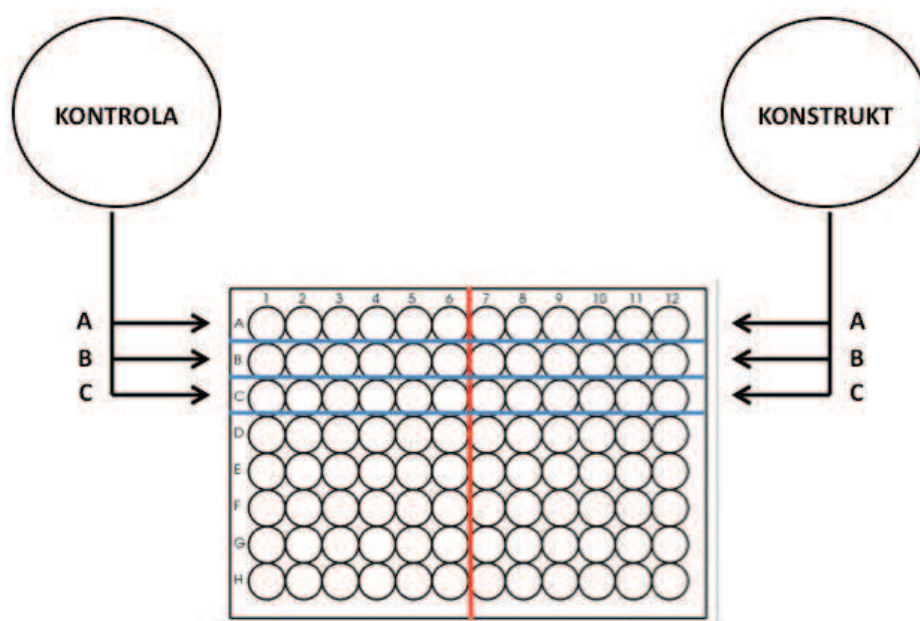
4.2.14. Test żywotności (wzrostu) komórek.

Celem oceny wpływu badanych transkryptów na żywotność (ang. *viability*) komórek linii PTC zastosowano zestaw alamarBlue[®] firmy Invitrogen (USA). Test ten opiera się na reakcji redukcji przez komórki rezazuryny (ang. *resazurin*) do rezorufiny (ang. *resorufin*) połączonej ze zmianą koloru z niebieskiego (rezazuryna) do czerwonego (rezorufina) i przejściem od związku praktycznie niefluorescencyjnego (rezazuryna) do fluorescencyjnego (rezorufina) (Rycina 5). Intensywność reakcji redukcji a tym samym wielkość zmierzonej fluorescencji (lub absorbancji) jest tym większa, im komórki są liczniejsze, szybciej proliferujące, bardziej żywotne, o wyższym poziomie metabolizmu. Komórki uszkodzone (apoptyczne) wykazują mniejszą aktywność metaboliczną i w związku z tym generują proporcjonalnie słabszy sygnał niż komórki zdrowe i proliferujące. W związku z tym, że wielkość fluorescencji jest wprost proporcjonalna do liczby komórek test alamarBlue jest dość powszechnie wykorzystywany do oceny różnic w tempie proliferacji [99, 100].



Rycina 5. Reakcja redukcji nie wykazującej właściwości fluorescencyjnych rezazuryny do fluorescencyjnej rezorufiny przez aktywne metabolicznie komórki. Intensywność redukcji a tym samym poziomu fluorescencji jest proporcjonalny do liczby żywych komórek.

W protokole przyjętym w niniejszej rozprawie komórki linii PTC posiano na dwie 100 mm płytki (5×10^5 komórek/płytkę) i następnego dnia (przy konfluencji 90%) transfekowano przy użyciu 5 μg konstrukt (zawierającego badany transkrypt, rozdział 4.2.8) lub 5 μg pcDNA3. Jako nośnika dla DNA użyto 12 μl Lipofektaminy 2000 (Invitrogen, USA) zgodnie z protokołem zalecanym przez producenta. Po 12 h komórki zostały przeniesione na 96-dołkową płytkę (10^3 komórek/dołek) po 6 dołków (powtórzeń) dla kontroli i konstrukt dla każdego analizowanego przedziału czasowego (Rycina 6). Po 24, 48 i 72 godzinach od transfekcji dodawano do odpowiednich dołków 10 μl odczynnika alamarBlue®, a następnie po 2 godzinach inkubacji w cieplarni (37°C , 95% powietrza, 5% CO_2) dokonywano pomiaru fluorescencji (długość fali ekscytacji: 560 nm; długość fali emisji: 590 nm) przy pomocy aparatu SpectraMax M2 firmy Molecular Devices (USA). Wyniki przedstawiono jako średnia \pm SEM z 6 niezależnych eksperymentów.



Rycina 6. Schemat przeprowadzonego eksperymentu dla oceny wpływu badanych transkryptów na żywotność (wzrost) komórek linii PTC. Komórki transfekowano wektorem pcDNA3 (KONTROLA) i konstrukt (pcDNA3 z wklonowanym transkryptem: KONSTRUKT), a następnie po 12 godzinach od transfekcji przeniesiono na 96 dołkową płytkę. Dla każdego analizowanego przedziału czasowego (A-24 h, B-48 h, C-72 h) przygotowano po 6 dołków dla kontroli i konstrukt. Po 24, 48 i 72 godzinach od transfekcji

do odpowiednio dołków A, B i C dodawano 10 μ l odczynnika alamarBlue® i po 2 h inkubacji w cieplarni mierzone wielkość fluorescencji. Całość eksperymentu powtórzono 6 razy.

4.2.15. Analiza wpływu wytypowanych transkryptów na ekspresję genów przy użyciu mikromacierzy.

Komórki linii PTC posiano na 12 dołkową płytkę (5×10^4 komórek/dołek), a następnie po 24 godzinach transfekowano w czterech powtórzeniach wektorem ekspresyjnym pcDNA3 (50 ng) zawierającym analizowany transkrypt. Jako kontroli użyto komórek transfekowanych w czterech powtórzeniach pustym wektorem pcDNA3 (50 ng). DNA wprowadzono do komórek za pomocą Lipofektaminy 2000 (2 μ l) zgodnie z protokołem zalecanym przez producenta (Invitrogen, USA). Po 24 godzinach od transfekcji z komórek wyizolowano całkowite RNA przy użyciu roztworu TRIzol (Invitrogen, USA). Próbkę RNA oczyszczono z ewentualnej obecności genomowego DNA dzięki zestawowi DNA-free™ firmy Ambion (USA), a następnie oceniono jego integralność przy pomocy Agilent BioAnalyzer 2100 (Agilent, USA). Oznaczenie poziomu ekspresji genów przeprowadzono przy użyciu zestawu mikromacierzy DNA SurePrint G3 Human Gene Expression 8x60000 Arrays firmy Agilent (AMADID 028004). Zestaw ten zawiera sondy dla 37 357 genów, w tym dla 7419 niekodujących RNA. Analizowane próbki zostały oznakowane barwnikiem cyjaninowym Cy3 przy użyciu zestawu Agilent Low Input One-Color Labeling (Agilent, USA), a następnie hybrydyzowane do mikromacierzy w temperaturze 65°C przez 17 godzin. Sygnał z płytek mikromacierzowych odczytano za pomocą skanera Agilent G2505C Microarray Scanner (Agilent, USA). Otrzymane obrazy analizowano wykorzystując oprogramowanie Feature Extraction 10.9 (Agilent, USA).

4.2.16. Analiza statystyczna.

Dane z reakcji PCR z oceną ilości produktu w czasie rzeczywistym analizowano przy pomocy testu Kruskal-Wallis, testu Wilcozona dla par obserwacji i testu U Manna-Whitneya. Wyniki eksperymentów pomiaru aktywności lucyferazy i wzrostu komórek były

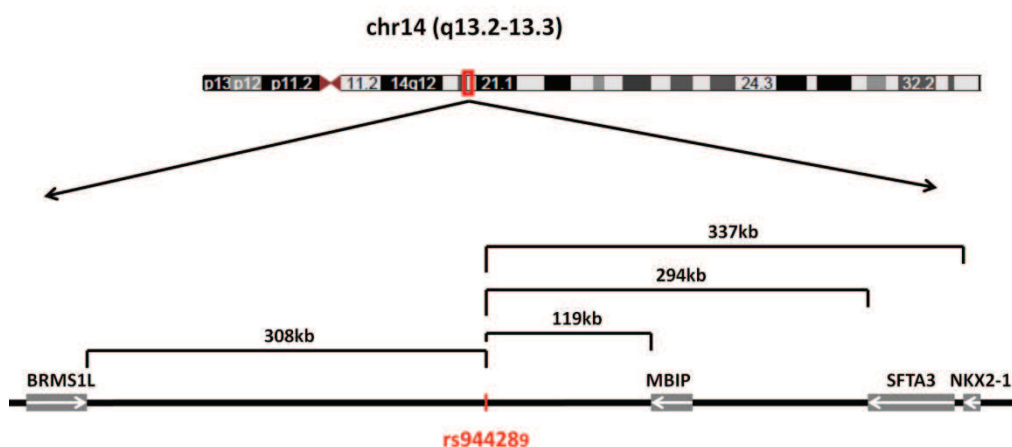
analizowane przy użyciu testu t studenta. Dane z mikromacierzy zostały transformowane logarytmicznie (\log_2), a następnie przy pomocy t testu dla losowego modelu wariancji dokonano porównania wpływu analizowanego transkryptu (konstrukt) i kontroli (pusty wektor) na ekspresję genów. Obliczenia wykonano wykorzystując oprogramowanie R (<http://www.r-project.org>) z pakietem LIMMA i BRB-Array Tools [101, 102].

Wszystkie podane w pracy wartości p są dwustronne.

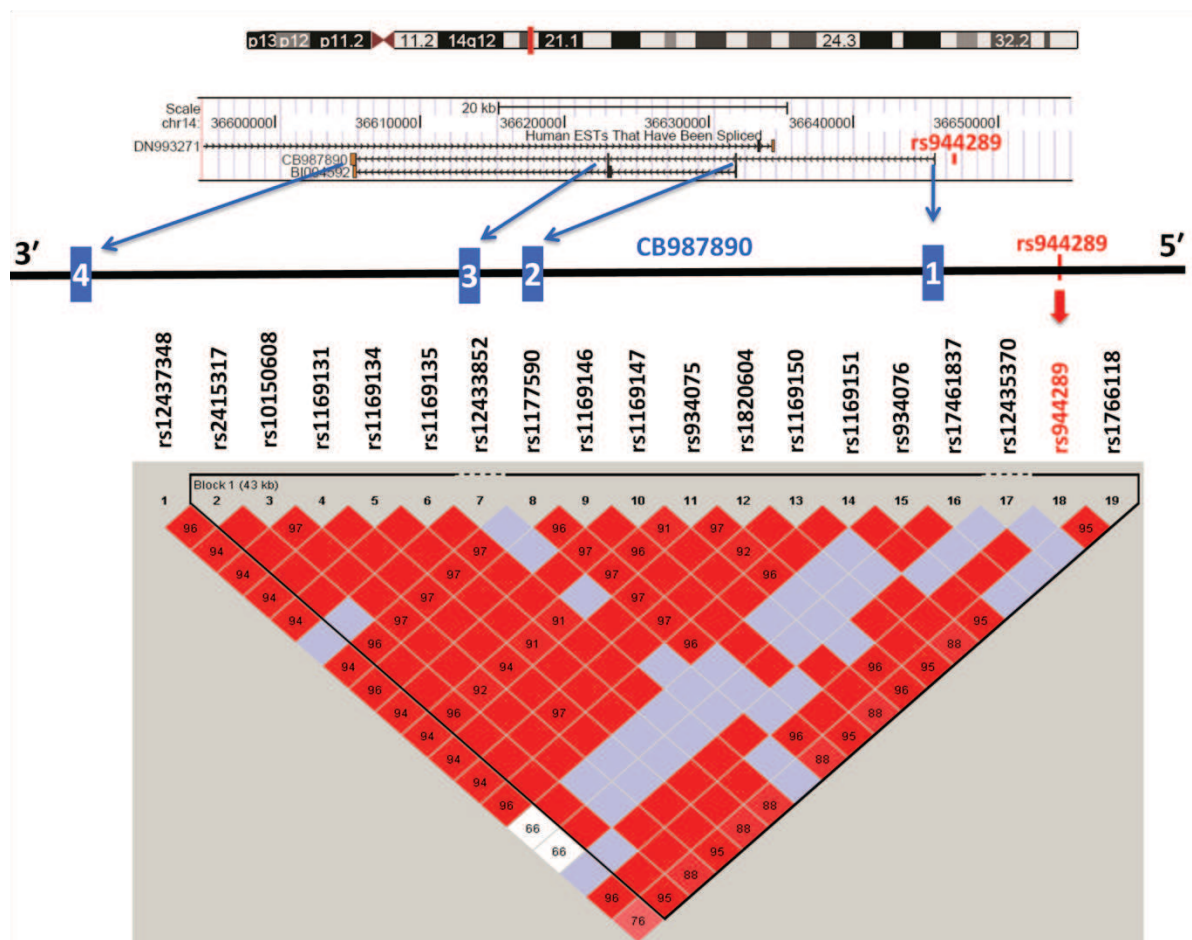
5. WYNIKI

5.1. Analiza *in silico* regionu 14q13.3.

Region 14q13.3 nie zawiera w bliskim (± 100 kb) sąsiedztwie rs944289 opisanych genów. Najbliższe odnotowane w bazie danych UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>, assembly February 2009, GRCh37/hg19; dane z listopada 2009) geny to *Breast Cancer Metastasis-Suppressor 1-Like (BRMS1L)* i *MAP3K12 Binding Inhibitory Protein 1 (MBIP)* znajdujące się odpowiednio 308 kb centrometrycznie i 119 kb telometrycznie od rs944289. Najbliższym tarczycowo-specyficznym genem jest *NK2 homeobox 2 (NKX2-1)* zlokalizowany 336 kb telometrycznie od rs944289 (Rycina 7). W celu poszukiwania znaczników ekspresji znajdujących się w sąsiedztwie rs944289 analizowano dane zawarte w UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>, assembly February 2009, GRCh37/hg19; dane z listopada 2009) i bazie Expressed Sequence Tags Database (<http://www.ncbi.nlm.nih.gov/dbEST/>; dane z listopada 2009). Według UCSC Genome Browser w bliskim (± 40 kb) sąsiedztwie rs944289 znajdują się następujące EST: CB987890, BI004592, DN993271, AA921750, AA632637, AW341561, CV391331, AW899174 i DB325123. Dane zawarte w bazie dbEST wskazały, że CB987890 został sklonowany z mieszaniny RNA wyizolowanego z tkanki tarczycowej, przytarczyczej, nadnerczowej i szyszynki. Ten potencjalny gen ulega posttranskrypcyjnej obróbce, co sugeruje stabilność transkryptu w cytoplazmie komórkowej i możliwość translacji. CB987890 składa się z 4 eksonów o łącznej długości 809 nt. Ekson 1 znajduje się 3.2 kb w kierunku końca 3' nici DNA (ang. *downstream*) od polimorfizmu rs944289. CB987890 i rs944289 znajdują się w 43 kb bloku genomowego DNA charakteryzującym się nierównowagą sprzężeń (Rycina 8).



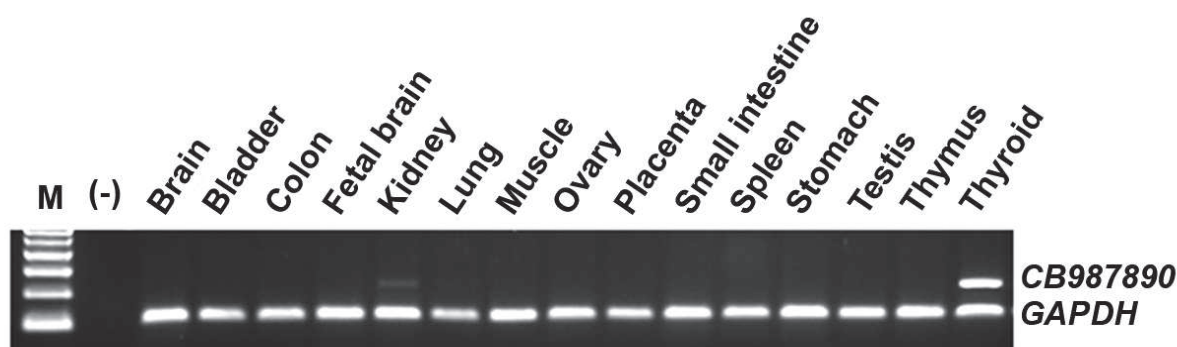
Rycina 7. Schematyczne przedstawienie regionu 14q13.3. Górny panel przedstawia chromosom 14 z zaznaczonym (czerwony prostokąt) locus 14q13.3. Dolny panel przedstawia region 14q13.3 z pozycją rs944289 (kolor czerwony) i odległością względem niego genów odnotowanych w bazie UCSC Genome Browser (assembly February 2009, GRCh37/hg19; dane z listopada 2009).



Rycina 8. Lokalizacja EST CB987890 względem wariantu rs944289. Górny panel przedstawia chromosom 14 z zaznaczonym kolorem czerwonym regionem 14q13.3. Poniżej przedstawiono schematycznie rs944289 (kolor czerwony) i CB987890 znajdujący się w bliskim (3.2 kb) sąsiedztwie badanego polimorfizmu. Dolny panel przedstawia rs944289 (kolor czerwony) i CB987890 (kolor niebieski) znajdujące się w 43 kb bloku charakteryzującym się nierównowagą sprzężeń.

5.2. Tkankowa ekspresja CB987890.

Analizę tkankowej ekspresji CB987890 przeprowadzono przy użyciu RT-PCR (starter 5': 5'-TCAAACCTCCAGGGCTTGAAC-3'; starter 3': 5'-ATTACGGCTGGGTCTACCT-3') zgodnie z protokołem opisanym w rozdziale 4.2.6. Produkt RT-PCR został rozdzielony elektroforetycznie na 2% agarozowym żelu. Analiza prążków PCR wykazała, że CB987890 ulega silnej transkrypcji w tkance tarczycowej (Thyroid) i niewielkiej w tkance nerkowej (Kidney). W tkance mózgowej (Brain, Fetal brain), pęcherza moczowego (Bladder), jelita grubego (Colon) i cienkiego (Small intestine), płuc (Lung), mięśni (Muscle), jajnika (Ovary), łożyska (Placenta), śledziony (Spleen), żołądka (Stomach), jąder (Testis) i grasicy (Thymus) nie stwierdzono transkrypcji CB987890 (Rycina 9). Z uwagi na silną, tarczycowo-specyficzną ekspresję CB987890 wytypowano ten transkrypt do dalszej analizy.



Rycina 9. Analiza tkankowej ekspresji CB987890 oznaczona metodą RT-PCR. CB987890 ulega silnej transkrypcji w tkance tarczycowej (Thyroid) i bardzo słabej w tkance nerkowej (Kidney). Ekspresja *GAPDH* posłużyła jako kontrola.

5.3. Struktura CB987890.

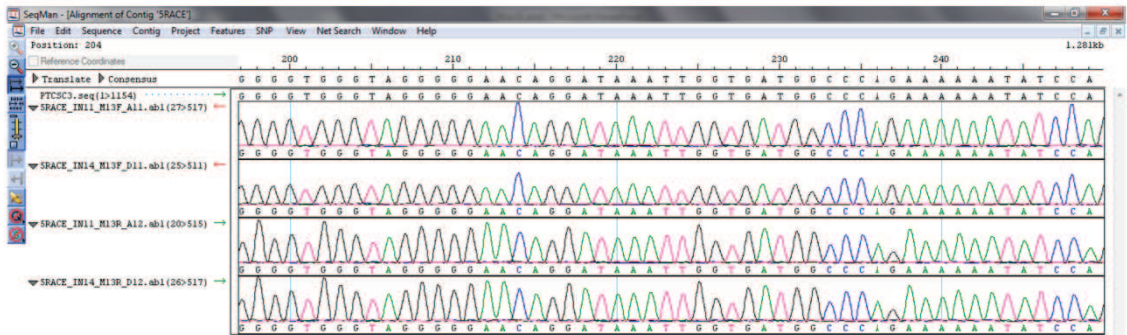
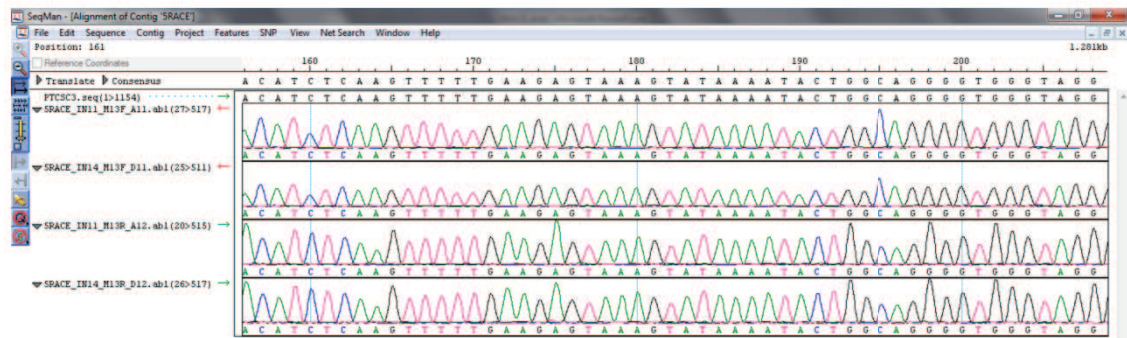
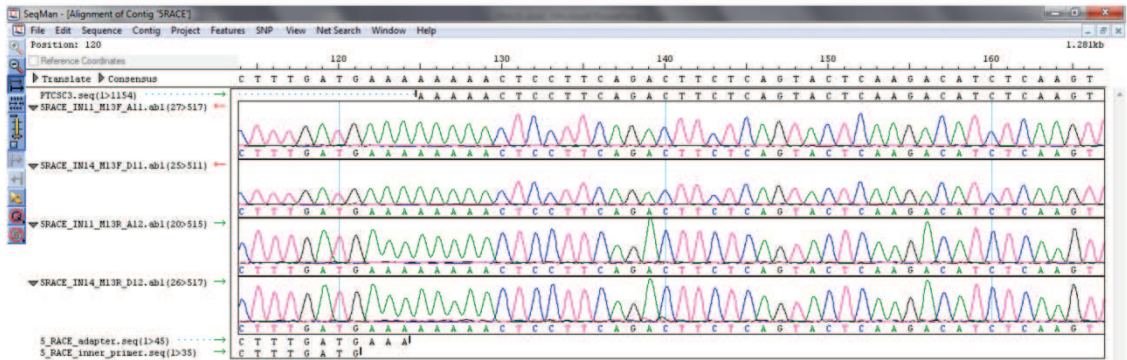
5.3.1 Szybka amplifikacja końców 5' i 3' CB987890 z użyciem ligazy RNA (RLM-RACE).

W celu określenia pełnej sekwencji CB987890 zastosowano metodę szybkiej amplifikacji końców cDNA z użyciem ligazy RNA (RLM-RACE). Specyficzne startery użyte do amplifikacji końców 5' i 3' zostały opisane w Tabeli 35.

Tabela 35. Sekwencja specyficznych starterów użytych w reakcji RLM-RACE.

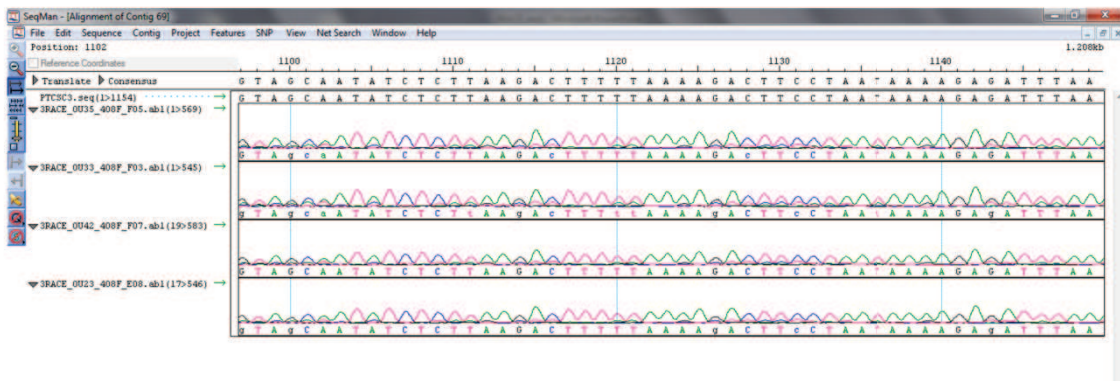
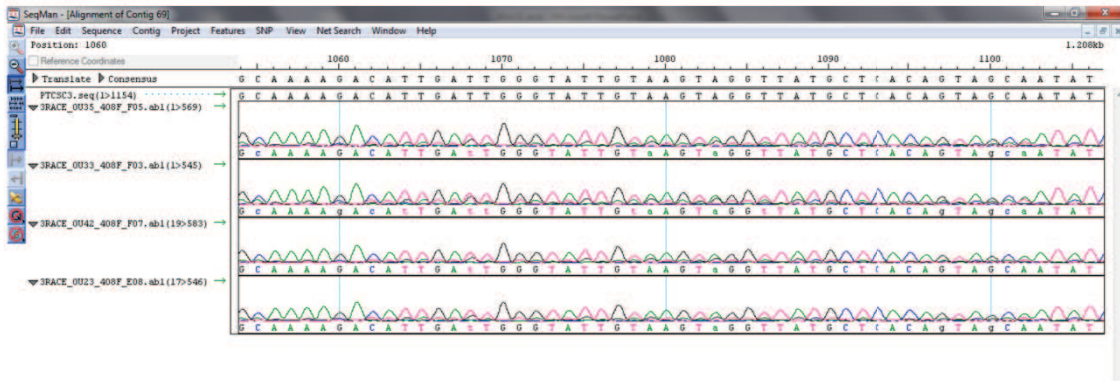
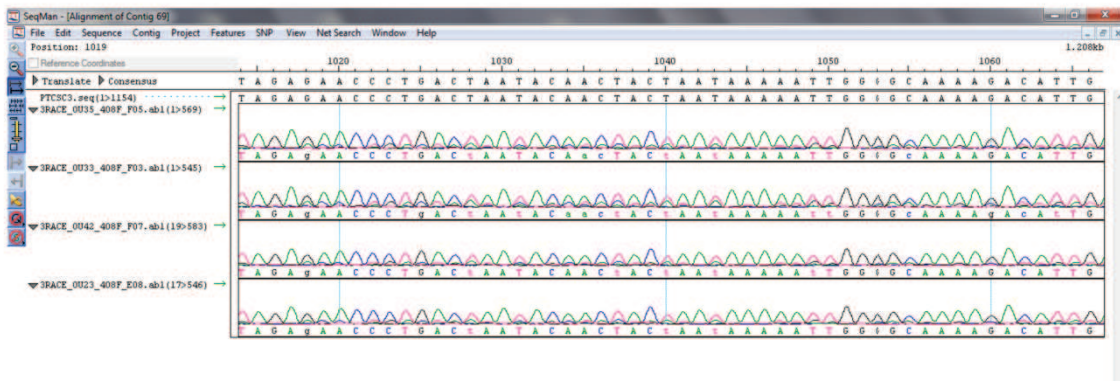
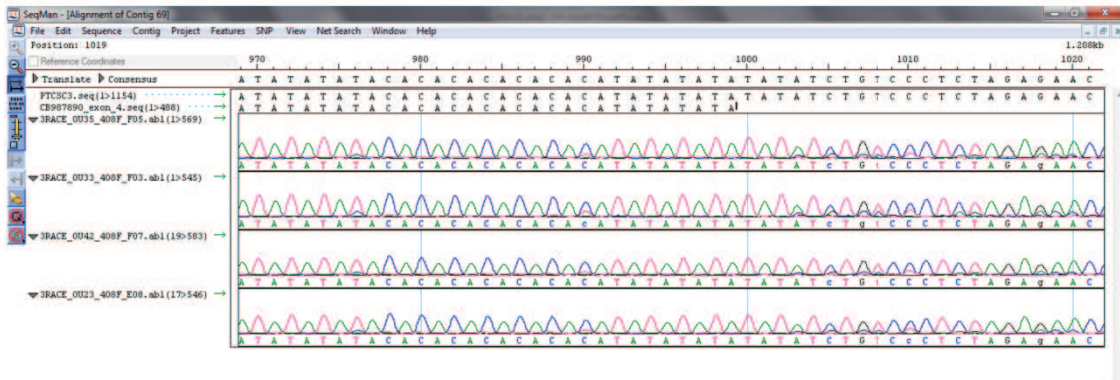
Starter	Sekwencja (5'→3')
5' RLM-RACE: zewnątrzny wewnętrzny	CATTTTGCCACCATGGTCTA TCTTCTGCTTGGCCTTTGAC
3' RLM-RACE: zewnątrzny wewnętrzny	TGGGACCTGTTGTTTTTCTTG TCAAACCTCCAGGGCTTGAAC

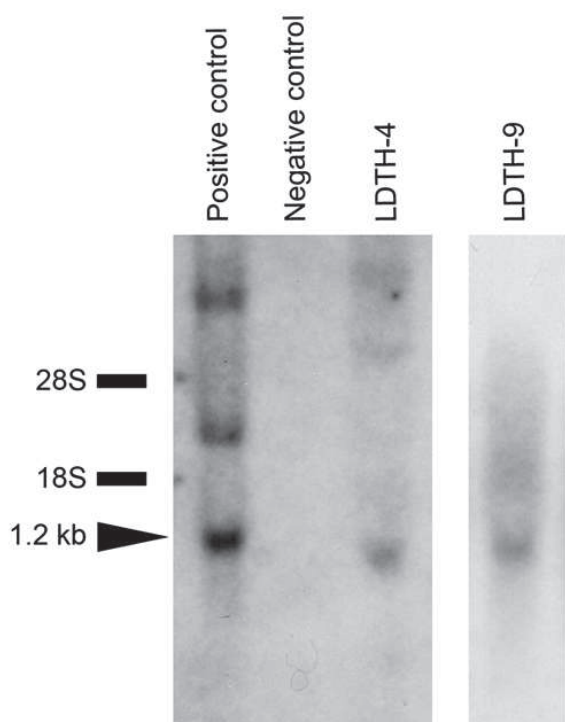
Dane z sekwencjonowania wklonowanych do wektora pCR4-TOPO produktów reakcji 5' i 3'RLM-RACE przedstawiono na Rycinie 10 i 11. Wykazano, że koniec 5' eksonu 1 składa się z dodatkowych 183 nukleotydów (nt). Koniec 3' eksonu 4 składa się z dodatkowych 192 nt. Całkowita długość CB987890 wynosi 1154 nt (Rycina 13). Sekwencja tego nowego genu nazwanego *Papillary Thyroid Carcinoma Susceptibility Candidate 3* (PTCSC3) została zdeponowana w GenBank (<http://www.ncbi.nlm.nih.gov/nucleotide/>) pod numerem dostępowym [JN689234](#).





Rycina 10. Przykładowe wyniki sekwencjonowania wklonowanych do wektora pCR4-TOPO produktów reakcji 5'RLM-RACE. Na rycinie pokazano sekwencję dodatkowych 183 nt składających się na ekson 1 (koniec 5') CB987890 (*PTCSC3*).



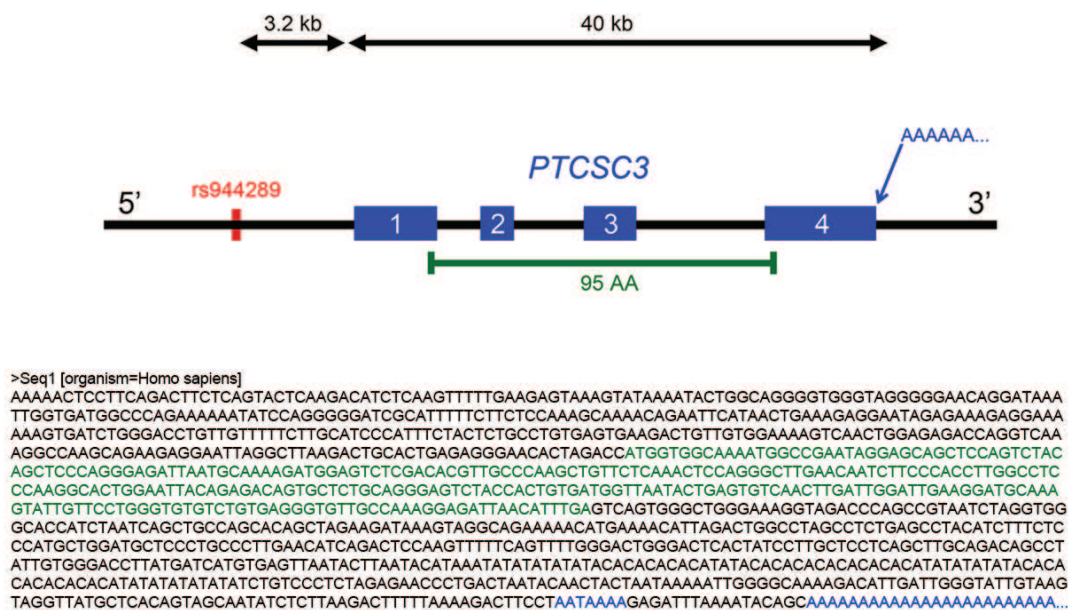


Rycina 12. Analiza długości mRNA *PTCSC3* metodą northern blot. Jako sondy użyto DNA hybrydującego do nukleotydów 125-307 *PTCSC3* mRNA. Pozytywna kontrola (Positive control): RNA wyizolowane z komórek 293T (nie wykazujących endogennej ekspresji *PTCSC3*) transfekowanych wektorem pcDNA3 z wklonowanym *PTCSC3*. Negatywna kontrola (Negative control): RNA wyizolowane z komórek 293T transfekowanych pustym wektorem pcDNA3. Specyficzne prążki dla *PTCSC3* (około 1.2 kb) zostały uwidocznione dla tarczycowego RNA (LDTH-4, LDTH-9) i dla pozytywnej kontroli (Positive control) potwierdzając wyniki uzyskane dzięki RLM-RACE.

5.3.4. Analiza *in silico* struktury *PTCSC3*.

Analiza bioinformatyczna oparta o dane dostępne w bazie UCSC Genome Browser wykazała, że z 4 eksonów składających się na *PTCSC3* jedynie sekwencja eksonu 1 jest unikalna dla ludzkiego genomu. Sekwencje eksonów 2 i 3 pokrywają się całkowicie z odpowiednio elementami LINE (ang. *long interspersed nuclear elements*) z rodziny L1 (ang. *family L1*) i SINE (ang. *short interspersed nuclear elements*) z rodziny Alu. Sekwencja

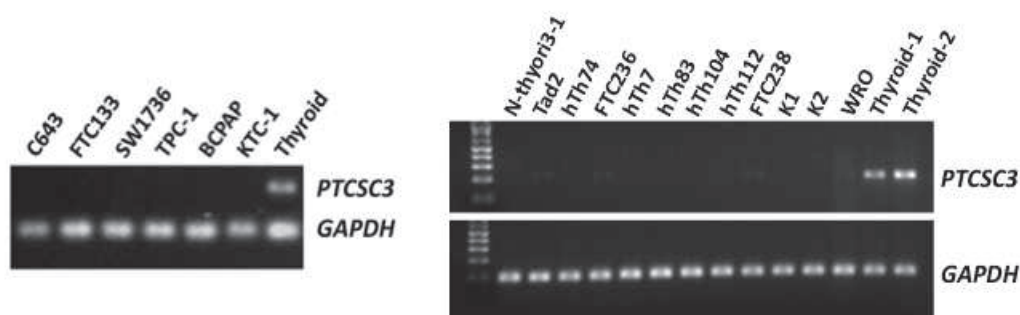
eksonu 4 w około 75% pokrywa się z elementami LTR ERVL (ang. *long terminal repeat endogenous retrovirus-like elements*). Elementy LINE, SINE i LTR są charakterystyczne dla mobilnych elementów genetycznych (ang. *mobile genetic elements*) takich jak transpozony klasy I (retrotranspozony). *PTCSC3* posiada sekwencję dla sygnału poliadenylacji (AATAAA) zlokalizowaną 18 nt od końca 3'. Analiza *in silico* wskazała na obecność potencjalnej 95 aminokwasowej otwartej ramki odczytu (ang. *open reading frame, ORF*) zaczynającej się typową sekwencją Kozak (AccAUGg) (Rycina 13). Jednakże struktura *PTCSC3* z unikalną sekwencją eksonu 1 i eksonami 2, 3 i 4 zawierającymi przeplatane powtórzenia DNA (ang. *interspersed repeat DNA*) sugeruje, że gen ten prawdopodobnie nie koduje białka i może należeć do grupy dużych międzygenowych niekodujących RNA (ang. *large intergenic noncoding RNA, lincRNA*).



Rycina 13. Schemat struktury *PTCSC3* (górny panel) wraz z pełną sekwencją tego genu (dolny panel). *PTCSC3* składa się z 4 eksonów znajdujących się w 40 kb bloku genomowego DNA. SNP rs944289 (kolor czerwony) znajduje się 3.2 kb od eksonu 1 w regionie dalszego promotora. Całkowita długość *PTCSC3* wynosi 1154 nukleotydów. Transkrypt ten posiada potencjalną ORF o długości 95 aminokwasów (kolor zielony) oraz sekwencję sygnałową dla poliadenylacji (AATAAA) i łańcuch poliA (kolor niebieski).

5.4. Ekspresja *PTCSC3* w liniach komórkowych raka tarczycy.

Analizę ekspresji *PTCSC3* w 18 liniach komórkowych NMTC przeprowadzono przy użyciu RT-PCR (starter 5': 5'-TCAAACCTCCAGGGCTTGAAC-3'; starter 3': 5'-ATTACGGCTGGGTCTACCT-3') zgodnie z protokołem opisanym w rozdziale 4.2.6. Jako pozytywnej kontroli użyto RNA wyizolowanego ze zdrowej tkanki tarczycowej (Thyroid). Produkt reakcji RT-PCR rozdzielono na 2% agarozowym żelu (85 V, 45 minut). W 14 liniach komórkowych stwierdzono brak transkrypcji *PTCSC3*. Gen ten wykazywał bardzo słaby poziom ekspresji jedynie w 4 liniach (Tad2, FTC236, FTC238 i WRO) (Rycina 14).

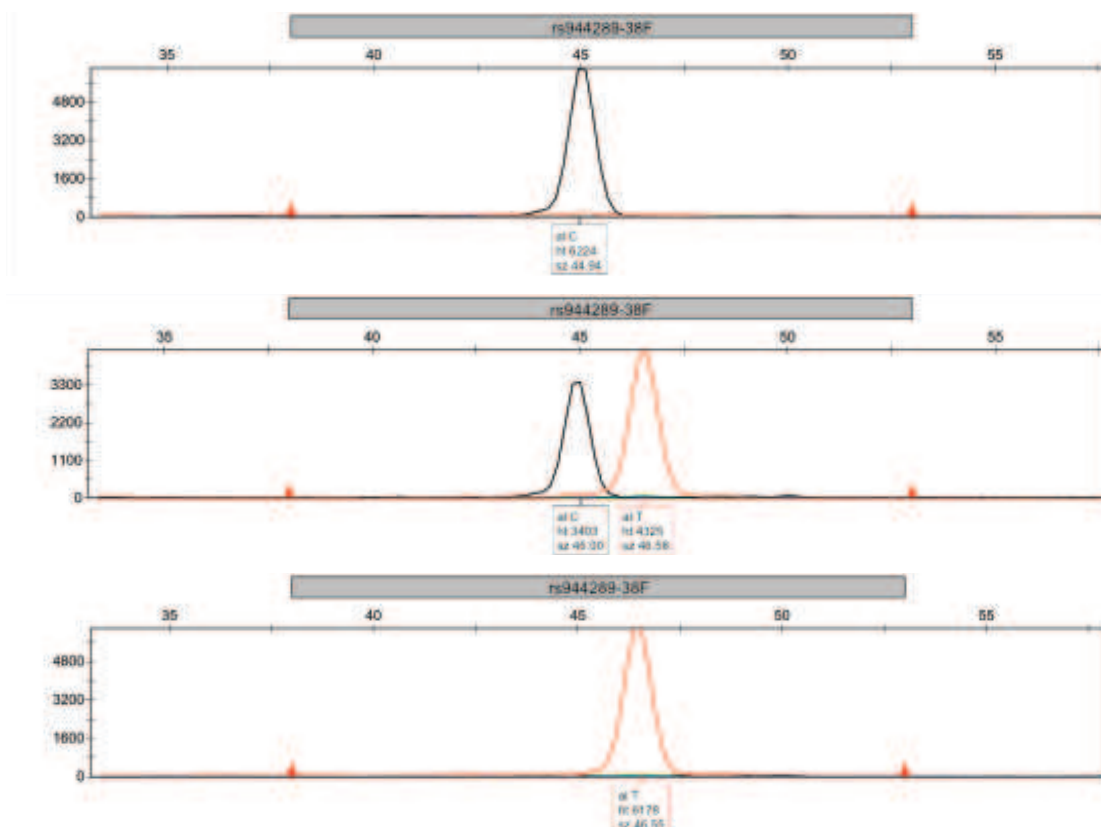


Rycina 14. Ekspresja *PTCSC3* w ludzkich liniach komórkowych NMTC oceniona metodą RT-PCR. W większości linii nowotworowych nie uwidoczniono ekspresji *PTCSC3*. Słaby poziom transkrypcji udokumentowano w liniach Tad2, FTC236, FTC238 i WRO. Tkanka tarczycowa (Thyroid) została wykorzystana jako pozytywna kontrola. Ekspresja *GAPDH* posłużyła jako kontrola.

5.5. Analiza genotypu rs944289 w badanej populacji.

U każdego chorego oznaczono genotyp polimorfizmu rs944289 używając jako materiału DNA wyizolowanego z limfocytów krwi obwodowej. Genotypowanie przeprowadzono wykorzystując zestaw SNaPshot® Multiplex Kit firmy Applied Biosystems (USA) zgodnie z protokołem opisanym w rozdziale 4.2.5. Dane z genotypowania wykazały, że w badanej populacji n=6 pacjentów było homozygotami rs944289[C], n=19

heterozygotami rs944289[CT] i n=21 homozygotami rs944289[T]. Przykładowe wyniki genotypowania przedstawiono na Rycinie 15.



Rycina 15. Przykładowe wyniki genotypowania wariantu rs944289 otrzymane metodą SNaPshot. Górny, środkowy i dolny panel przedstawiają wyniki dla odpowiednio homozygot rs944289[C], heterozygot rs944289[CT] i homozygot rs944289[T].

5.6. Wpływ polimorfizmu rs944289 na ekspresję *PTCSC3*.

Związek pomiędzy ekspresją *PTCSC3* a genotypem rs944289 badano na próbie 46 chorych z rozpoznaniem PTC. Do pomiaru ekspresji *PTCSC3* wykorzystano reakcję PCR z analizą ilości produktu w czasie rzeczywistym. Startery użyte w reakcji *real-time* PCR zostały przedstawione w Tabeli 36.

Tabela 36. Startery użyte w reakcji *real-time* PCR

Starter	Sekwencja (5'→3')
PTCSC3_F (200nM)	TCAAACCTCCAGGGCTTGAAC
PTCSC3_R (200nM)	ATTACGGCTGGGTCTACCT
GAPDH_F (200nM)	CACCAACTGCTTAGCACCCC
GAPDH_R (200nM)	TCTTCTGGGTGGCAGTGATG

5.6.1. Ekspresja *PTCSC3* w tkance tarczycowej.

W tkance tarczycowej nieobjętej procesem nowotworowym ekspresja *PTCSC3* u chorych z genotypem rs944289[CT] (n=19; mediana=0.0114 (Q1-Q3:0.0060-0.0153)) w porównaniu z homozygotami [TT] (n=21; mediana=0.0141 (Q1-Q3:0.0110-0.0189)) była istotnie statystycznie wyższa (p=0.034). Nie stwierdzono różnic w ekspresji analizowanego genu pomiędzy chorymi z genotypem [CC] (n=6; mediana=0.0135 (Q1-Q3:0.0121-0.0152)) i [CT] (n=19; mediana=0.0114 (Q1-Q3:0.0060-0.0153)) (p=0.303) oraz pomiędzy homozygotami [CC] (n=6; mediana=0.0135 (Q1-Q3:0.0121-0.0152)) i homozygotami [TT] (n=21; mediana=0.0141 (Q1-Q3:0.0110-0.0189)) (p=0.932) (Tabela 37, Rycina 16).

5.6.2. Ekspresja *PTCSC3* w tkance guza.

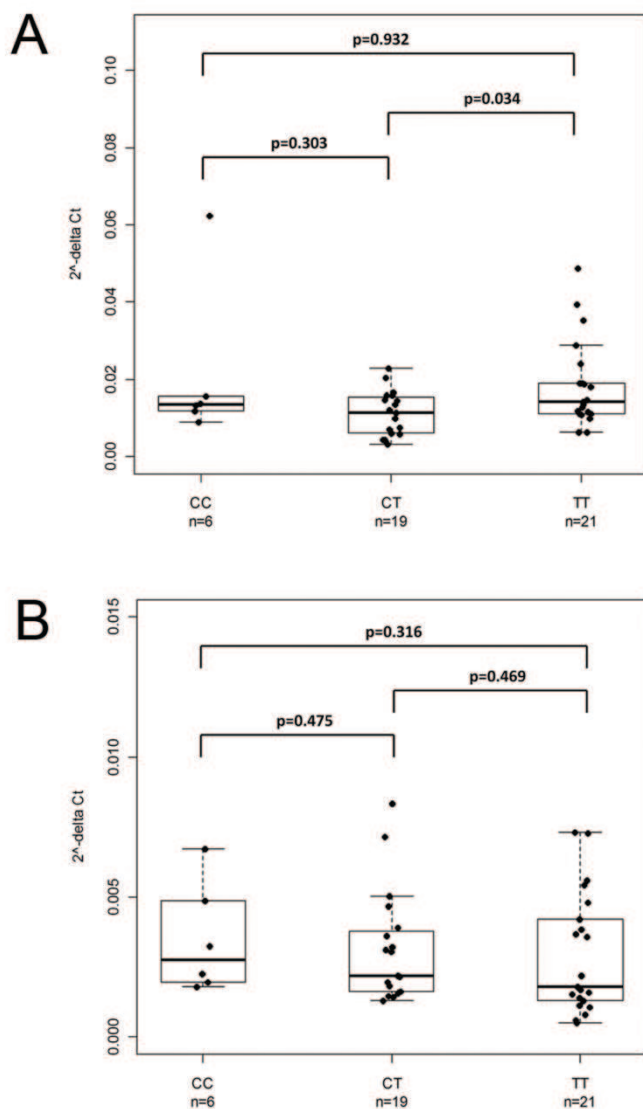
Mediana ekspresji *PTCSC3* w tkance guza układała się według grup chorych: genotyp [CC] (n=6; mediana=0.0027 (Q1-Q3:0.0020-0.0045)) > genotyp [CT] (n=19; mediana=0.0022 (Q1-Q3:0.0016-0.0038)) > genotyp [TT] (n=21; mediana=0.0018 (Q1-Q3:0.0013-0.0042)). Różnice pomiędzy grupami nie wykazywały jednakże istotności statystycznej (Tabela 37, Rycina 16).

5.6.3. Analiza ekspresji *PTCSC3* wyrażonej jako stosunek transkrypcji w tkance guza do tkanki zdrowej.

W celu oceny stopnia supresji lub aktywacji *PTCSC3* w tkance nowotworowej analizowano stosunek ekspresji tego genu w tkance guza (ang. *tumor*, T) do tkanki zdrowej (ang. *normal*, N). Współczynnik T/N wyraża siłę względnego hamowania lub aktywacji transkrypcji *PTCSC3* w tkance guza w porównaniu ze zdrową tkanką tarczycową. Analiza wartości T/N wskazała na silną supresję *PTCSC3* w tkance PTC u wszystkich chorych ($p=2.84 \times 10^{-14}$) (Rycina 17). Ekspresja *PTCSC3* u homozygot [TT] (n=21; mediana=0.1325 (Q1-Q3:0.0860-0.2599)) w porównaniu z heterozygotami [CT] (n=19; mediana=0.2789 (Q1-Q3:0.1940-0.4902)) ulegała silniejszemu hamowaniu ($p=0.004$). Nie stwierdzono istotnych statystycznie różnic pomiędzy pacjentami z genotypem [TT] (n=21; mediana=0.1325 (Q1-Q3:0.0860-0.2599)) i [CC] (n=6; mediana=0.2395 (Q1-Q3:0.1402-0.4335)) ($p=0.110$) oraz pomiędzy homozygotami [CC] (n=6; mediana=0.2395 (Q1-Q3:0.1402-0.4335)) i heterozygotami [CT] (n=19; mediana=0.2789 (Q1-Q3:0.1940-0.4902)) ($p=0.687$) (Tabela 37, Rycina 17).

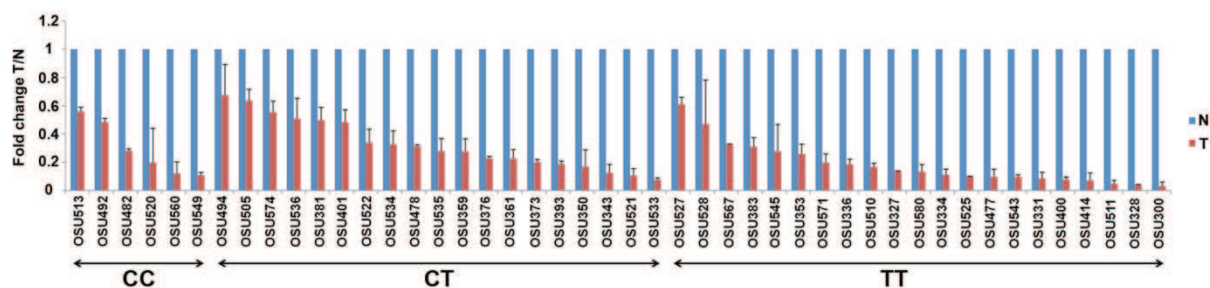
Tabela 37. Wyniki analizy ekspresji *PTCSC3* metodą *real-time PCR* u chorych z PTC (n=46) w zależności od genotypu rs944289. U każdego pacjenta oceniano ekspresję *PTCSC3* w tkance tarczycowej (*normal*, N) i tkance guza (*tumor*, T). Współczynnik T/N wyraża siłę względnej supresji lub aktywacji *PTCSC3* w tkance guza.

Genotyp	Tkanka zdrowa (<i>normal</i> , N)	Tkanka guza (<i>tumor</i> , T)	Współczynnik T/N
CC (n=6)	0.0135 (Q1-Q3:0.0121-0.0152)	0.0027 (Q1-Q3:0.0020-0.0045)	0.2395 (Q1-Q3:0.1402-0.4335)
CT (n=19)	0.0114 (Q1-Q3:0.0060-0.0153)	0.0022 (Q1-Q3:0.0016-0.0038)	0.2789 (Q1-Q3:0.1940-0.4902)
TT (n=21)	0.0141 (Q1-Q3:0.0110-0.0189)	0.0018 (Q1-Q3:0.0013-0.0042)	0.1325 (Q1-Q3:0.0860-0.2599)

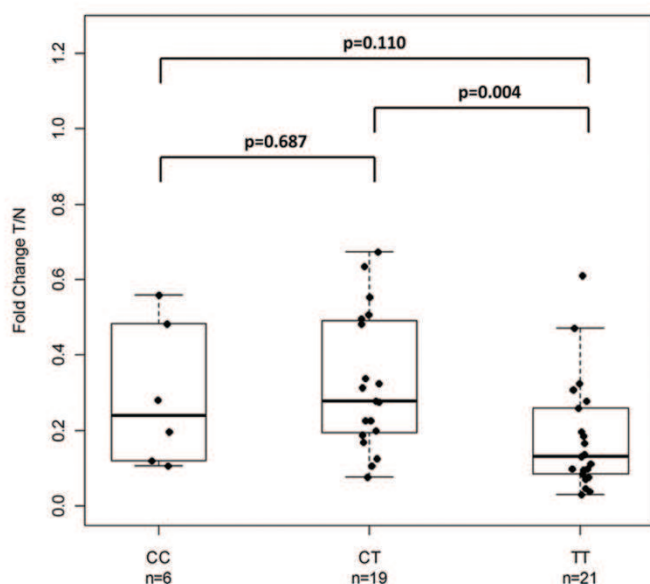


Rycina 16. Ekspresja *PTCSC3* u chorych z PTC (n=46) w tkance guza i tkance tarczycowej nieobjętej procesem nowotworowym oznaczona metodą *real-time PCR*. **(A)** Ekspresja *PTCSC3* w tkance tarczycowej nieobjętej procesem nowotworowym. Poziom transkrypcji u chorych z genotypem [TT] (n=21) w porównaniu z pacjentami z genotypem [CT] (n=19) był istotnie statystycznie wyższy (p=0.034). Nie stwierdzono istotnych statystycznie różnic w transkrypcji *PTCSC3* pomiędzy homozygotami [CC] (n=6) a homozygotami [TT] (n=21) (p=0.932) i homozygotami [CC] (n=6) a heterozygotami [CT] (n=19) (p=0.303). **(B)** Analiza ekspresji *PTCSC3* w tkance guza. Nie stwierdzono istotnych statystycznie różnic w poziomie mRNA pomiędzy analizowanymi grupami.

A



B



Rycina 17. Ilościowa ocena ekspresji *PTCSC3* u chorych z PTC (n=46) w tkance guza (T) i tkance tarczycowej nieobjętej procesem nowotworowym (N). Wyniki przedstawiono jako iloraz ekspresji w tkance guza do ekspresji w zdrowej tkance tarczycowej (T/N). **(A)** Transkrypcja *PTCSC3* w tkance PTC (T, kolor czerwony) była istotnie statystycznie słabsza w porównaniu z tkanką tarczycową nieobjętą procesem nowotworowym (N, kolor niebieski) ($p=2.84 \times 10^{-14}$). Na osi X przedstawiono chorych uszeregowanych według genotypu rs944289 (CC, n=6; CT, n=19; TT, n=21) i stopnia supresji *PTCSC3* w tkance guza. **(B)** Supresja *PTCSC3* w tkance guza była istotnie statystycznie silniejsza u chorych z genotypem rs944289[TT] (n=21) w porównaniu z heterozygotami rs944289[CT] (n=19) ($p=0.004$). Nie wykazano istotności statystycznej w poziomie supresji *PTCSC3* pomiędzy homozygotami

rs944289[TT] (n=21) i rs944289[CC] (n=6) (p=0.110) oraz pomiędzy homozygotami rs944289[CC] (n=6) i heterozygotami rs944289[CT] (n=19) (p=0.687).

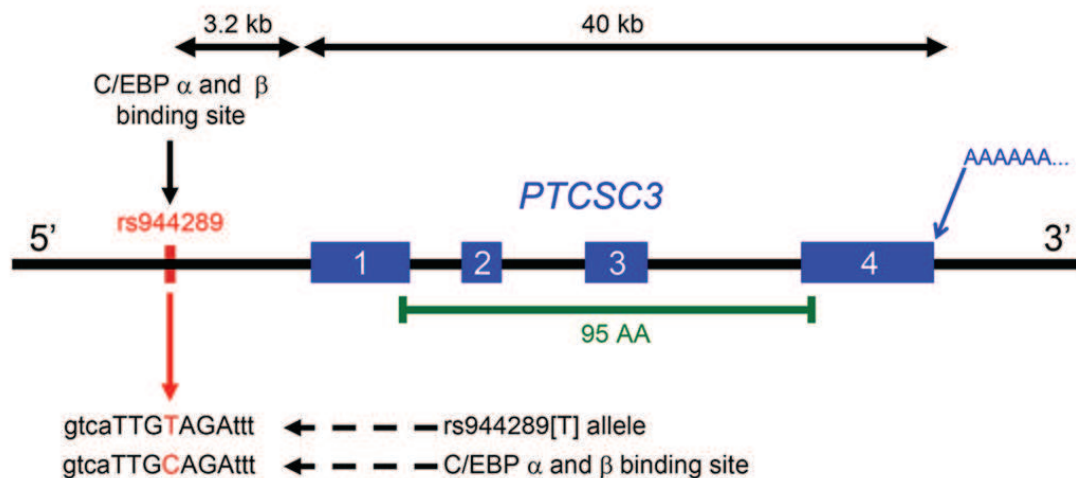
5.7. Regulacja transkrypcji *PTCSC3*.

5.7.1. Analiza *in silico* regionu promotora dla *PTCSC3*.

SNP rs944289 jest zlokalizowany w regionie dalszego promotora dla *PTCSC3* (3.2 kb od eksonu 1). Przy użyciu programu MATCH zidentyfikowano sekwencję gtcaTTG**C**AGAtt, do której powinowactwo wykazuje białko wiążące się z sekwencją CCAAT α i β (ang. *CCAAT/enhancer binding protein*, C/EBP α i C/EBP β). Miejsce wiązania zawiera w części konserwatywnej (ang. *conserved part*) polimorfizm rs944289 (zaznaczony kolorem czerwonym). Allel rs944289[T] osłabia *in silico* powinowactwo C/EBP α i C/EBP β do wiązania się z DNA obliczone jako wpływ pozycji w macierzy (ang. *position weight matrix*, PWM) (Tabela 38, Rycina 18).

Tabela 38. Przewidywane miejsce wiążące dla czynników transkrypcyjnych (TFBS) zawierające polimorfizm rs944289 (zaznaczony czerwoną czcionką). Analizę przeprowadzono dla obydwu alleli przy użyciu programu MATCH osobno dla rdzenia wiązania (części konserwatywnej) i całej sekwencji TFBS. Allel rs944289[T] osłabia *in silico* powinowactwo C/EBP α (V\$CEBP_Q2) i C/EBP β (V\$CEBPB_01) do wiązania z DNA wyrażone jako wpływ pozycji w macierzy (PWM).

Matryca	Przewidywane miejsce wiązania dla allelu C				Przewidywane miejsce wiązania dla allelu T			
	Ni \acute{e} DNA	Wynik dla rdzenia	Wynik całkowity	TFBS	Ni \acute{e} DNA	Wynik dla rdzenia	Wynik całkowity	TFBS
V\$CEBP_Q2	(+)	1.000	0.908	gtcaTTG C AGAttg	(+)	0.949	0.864	gtcaTTG T AGAttg
V\$CEBPB_01	(+)	0.897	0.896	gtcaTTG C AGAttg	(+)	0.889	0.890	gtcaTTG T AGAttg

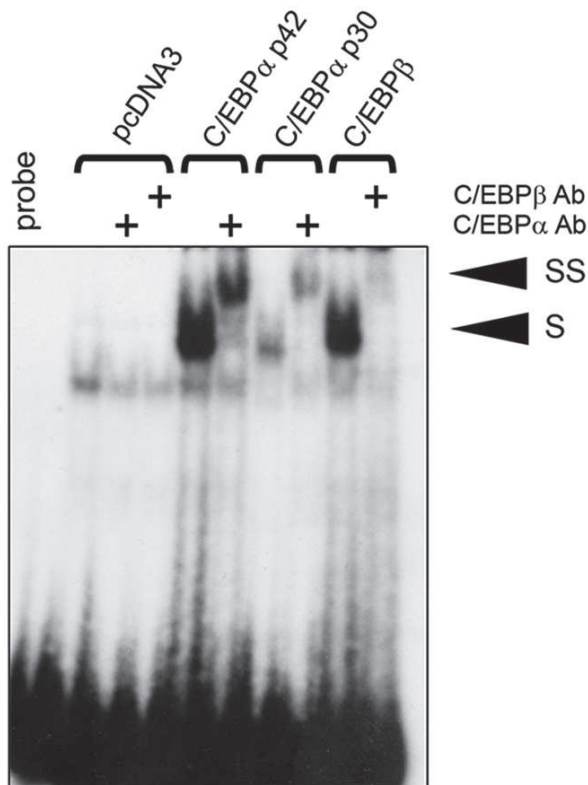


Rycina 18. Schemat *PTCSC3* z regionem promotora i pozycją rs944289. SNP jest zlokalizowany w regionie dalszego promotora dla *PTCSC3* wewnątrz części konserwatywnej miejsca wiążącego czynniki transkrypcyjne C/EBP α i C/EBP β . Allel ryzyka [T] zmniejsza *in silico* powinowactwo wiązania dla obydwu TF wyrażone jako PWM.

5.7.2. Test opóźnienia w żelu dla C/EBP α i C/EBP β .

W celu potwierdzenia wiązania *in vitro* C/EBP α i C/EBP β do DNA w miejscu lokalizacji rs944289 zastosowano test opóźnienia w żelu (EMSA) wykorzystując sondę zawierającą allel rs944289[C], który nie redukuje wielkości PWM i stanowi lepsze miejsce wiązania dla C/EBP. Jako materiału użyto wyciągu białkowego z jąder komórek HEK293T transfekowanych 24 h przed ekstrakcją protein pustym wektorem pcDNA3, pcDNA3 zawierającym izoformę C/EBP α p30, izoformę C/EBP α p42 lub C/EBP β (izoformę LAP*). Wektory ekspresyjne zawierające gen dla C/EBP α p30, C/EBP α p42 lub C/EBP β otrzymano od dr Daniela G. Tenen'a (Harvard Medical School, USA). Wyniki eksperymentu przedstawiono na Rycinie 19. C/EBP α p42 i C/EBP β wiązały się stosunkowo silnie z analizowaną sekwencją, podczas gdy izoforma C/EBP α p30 wykazywała słabsze powinowactwo. Specyfika wiązania C/EBP α p30, C/EBP α p42 i C/EBP β została potwierdzona poprzez brak prążków w obecności ekstraktu z komórek transfekowanych

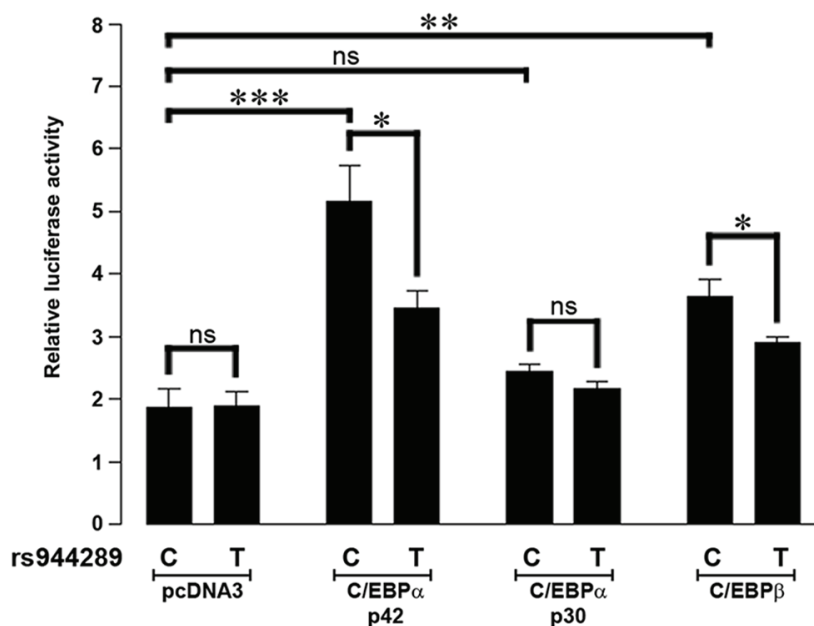
pustym wektorem (pcDNA3) oraz poprzez widoczny prążek „*super-shift*” (SS) (lub abrogację prążka S) w reakcjach ze specyficznymi dla badanych TF przeciwciałami.



Rycina 19. Test opóźnienia w żelu dla wiązania C/EBP α p30, C/EBP α p42 i C/EBP β (izoforma LAP*) z sekwencją zawierającą rs944289[C]. Prążek przesunięcia (*shift*, S) dla C/EBP α p30, C/EBP α p42 i C/EBP β wskazuje na wiązanie się tych TF z DNA zawierającym rs944289[C]. Specyficzność kompleksów C/EBP-DNA potwierdza brak prążka S dla ekstraktu z komórek transfekowanych pustym wektorem (pcDNA3) i widoczny prążek „*super-shift*” (SS) (lub abrogacja prążka S) w obecności specyficznych dla C/EBP α i C/EBP β przeciwciał. Probe: negatywna kontrola (mieszanka reakcyjna bez proteinowego ekstraktu).

5.7.3. Regulacja genu reporterowego *luc2P* przez C/EBP α i C/EBP β .

W celu oceny rodzaju oddziaływania C/EBP α i C/EBP β na promotor *PTCSC3* (aktywacja lub supresja transkrypcji) oraz wpływu badanego polimorfizmu na efekt działania tych czynników wklonowano 522 nt fragment promotora zawierający rs944289[C] lub rs944289[T] do wektora reporterowego pGL4.24 (Promega, USA). Komórki COS-7 (nie wykazujące endogennej ekspresji C/EBP α i C/EBP β) transfekowano wektorem reporterowym razem z wektorem ekspresyjnym zawierającym C/EBP α p30, C/EBP α p42 lub C/EBP β (izoforma LAP*). Jako kontroli użyto komórek COS-7 transfekowanych pustym wektorem pGL4.24. Wektor pcDNA3 służył jako nośnik DNA do utrzymywania stałej ilości materiału genetycznego w dołkach, do których nie dodawano wektorów zawierających C/EBP. Aktywność promotora wyrażono jako współczynnik *firefly/renila* względem współczynnika uzyskanego dla komórek transfekowanych pustym wektorem pGL4.24 (względna aktywność lucyferazy, ang. *relative luciferase activity*). Zarówno C/EBP α p42 jak i C/EBP β istotnie statystycznie aktywowały promotor dla *PTCSC3* (rs944289[C]: odpowiednio p=0.0009 i p=0.0014; rs944289[T]: odpowiednio p=0.0019 i p=0.0048). Nie zaobserwowano takiego efektu dla C/EBP p30 (rs944289[C]: p=0.107; rs944289[T]: p=0.327). Allel ryzyka [T] zmniejszał aktywność promotora generowaną przez C/EBP α p42 i C/EBP β (odpowiednio p=0.026 i p=0.048) (Rycina 20).



Rycina 20. Test podwójnej lucyferazy w komórkach COS-7 transfekowanych wektorem reporterowym pGL4.24 z wklonowanym powyżej genu *luc2P* fragmentem promotora *PTCSC3* (zawierającym allel [C] lub [T]) i wektorem ekspresyjnym zawierającym C/EBP α p30, C/EBP α p42 lub C/EBP β (izoforma LAP*). Czynniki C/EBP α p42 i C/EBP β aktywowały transkrypcję dla obydwu alleli (rs944289[C]: odpowiednio p=0.0009 i p=0.0014; rs944289[T]: odpowiednio p=0.0019 i p=0.0048). Obecność allelu [T] zmniejszała aktywność promotora wywołaną przez C/EBP α p42 i C/EBP β (odpowiednio p=0.026 i p=0.048). Izoforma C/EBP α p30 nie wpływała na aktywność promotora (rs944289[C]: p=0.107; rs944289[T]: p=0.327). Wyniki przedstawiono jako średnia \pm SEM z 6 niezależnych eksperymentów. Oś Y: względna aktywność lucyferazy (*relative luciferase activity*). *p<0.05; **p=0.01-0.001; ***p<0.001; ns, nieistotne statystycznie.

5.8. Sekwencjonowanie *PTCSC3*.

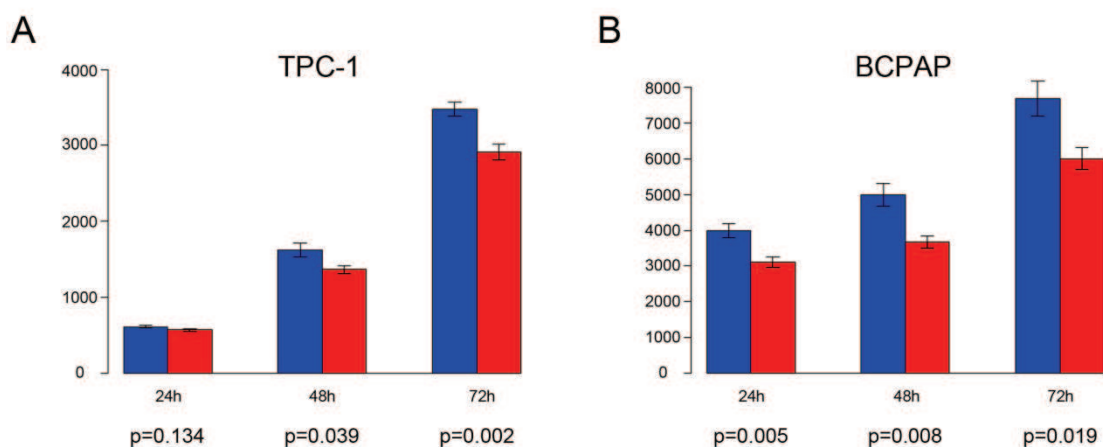
W celu wykrycia wrodzonych i/lub somatycznych mutacji w genie *PTCSC3* wykonano sekwencjonowanie cDNA otrzymanego ze zdrowej tkanki tarczycowej i tkanki guza od 33 chorych z PTC. *PTCSC3* cDNA zamplifikowano przy pomocy PCR (starter 5':

5'-AAAAACTCCTTCAGACTTCTCAGT-3'; starter 3': 5'-
GTCCCAGTCCCAAAGTAA-3') zgodnie z protokołem opisanym w rozdziale 4.2.13. Wyniki sekwencjonowania nie wskazały na obecność w badanej populacji wrodzonych i/lub somatycznych mutacji zlokalizowanych w *PTCSC3*.

5.9. Funkcja *PTCSC3*.

5.9.1. Wpływ *PTCSC3* na żywotność (wzrost) komórek.

Komórki ludzkich linii PTC (TPC-1 i BCPAP) nie wykazujące endogennej ekspresji *PTCSC3* transfekowano pustym wektorem ekspresyjnym pcDNA3 (kontrola) i wektorem pcDNA3 zawierającym *PTCSC3*. Po 24 h, 48 h i 72 h oceniano żywotność komórek wykorzystując zestaw alamarBlue® (Invitrogen, USA) zgodnie z protokołem opisanym w rozdziale 4.2.14. W związku z tym, że wielkość zmierzonej fluorescencji jest pośrednio proporcjonalna do liczby komórek wyniki testu interpretowano jako różnice w tempie wzrostu pomiędzy badanymi populacjami komórkowymi [99, 100]. Przywrócenie ekspresji *PTCSC3* hamowało wzrost obu linii (TPC-1: $p=0.039$ i $p=0.002$ po odpowiednio 48 h i 72 h; BCPAP: $p=0.005$, $p=0.008$ i $p=0.019$ po odpowiednio 24 h, 48 h i 72 h) (Rycina 21).



Rycina 21. Wpływ ekspresji *PTCSC3* na wzrost komórek linii PTC. (A) Komórki TPC-1 (nie wykazujące endogennej transkrypcji *PTCSC3*) transfekowane wektorem ekspresyjnym zawierającym *PTCSC3* (kolor czerwony) wykazywały mniejsze tempo wzrostu w

porównaniu z komórkami transfekowanymi pustym wektorem (kolor niebieski) ($p=0.039$ i $p=0.002$ po odpowiednio 48 h i 72 h). **(B)** Komórki BCPAP (nie wykazujące endogennej transkrypcji *PTCSC3*) transfekowane wektorem ekspresyjnym zawierającym *PTCSC3* (kolor czerwony) wykazywały mniejsze tempo wzrostu w porównaniu z komórkami transfekowanymi pustym wektorem (kolor niebieski) ($p=0.005$, $p=0.008$ i $p=0.019$ po odpowiednio 24 h, 48 h i 72 h). Oś X: punkty czasowe pomiaru fluorescencji; oś Y: intensywność fluorescencji. Wyniki przedstawiono jako średnia \pm SEM z 6 niezależnych eksperymentów.

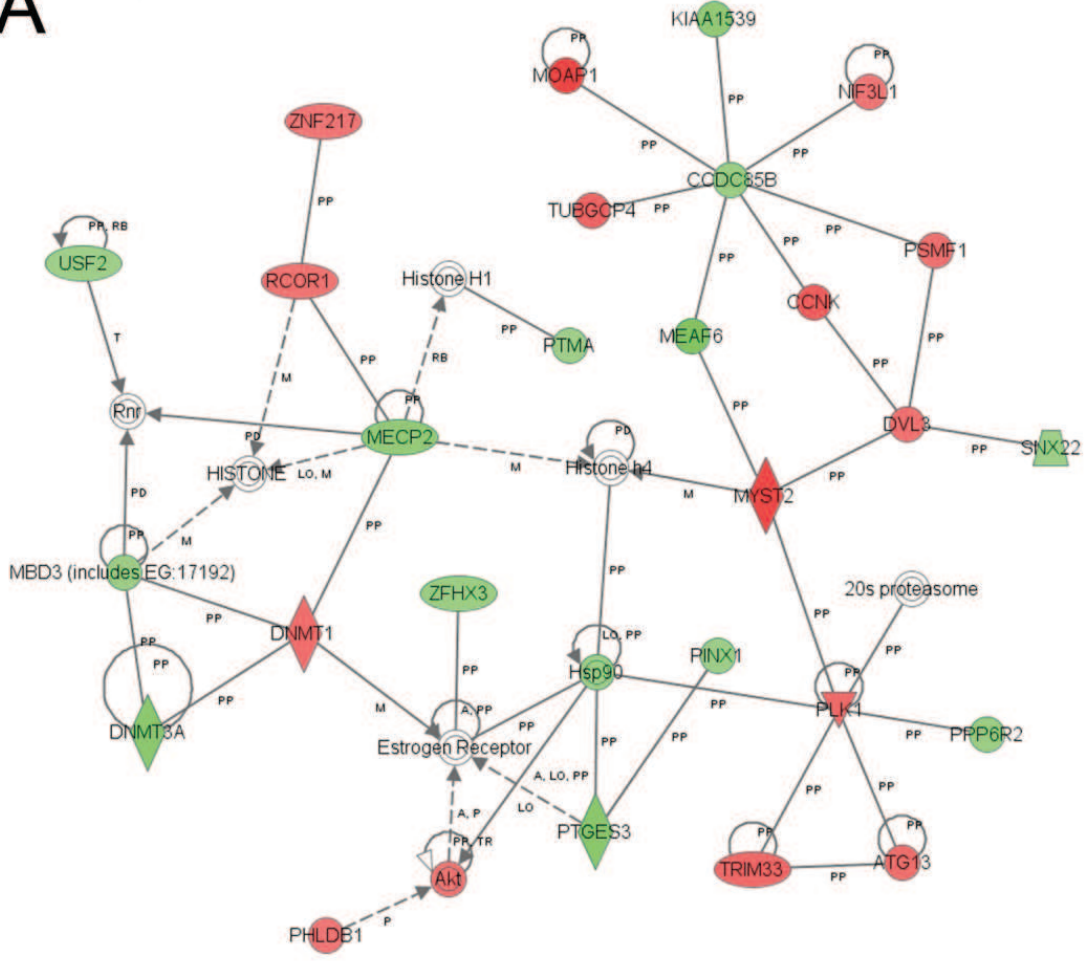
5.9.2. Wpływ *PTCSC3* na ekspresję genów.

Celem oceny wpływu *PTCSC3* na ekspresję innych genów komórki TPC-1 transfekowano pustym wektorem pcDNA3 (kontrola) i wektorem pcDNA3 zawierającym *PTCSC3*. Po 24 h od transfekcji analizowano poziom transkrypcji dla 37 357 genów przy użyciu zestawu mikromacierzy DNA SurePrint G3 Human Gene Expression 8x60000 Arrays firmy Agilent (USA). Otrzymane dane zostały przetworzone przy użyciu „volcano plot” ($p<0.05$, współczynnik zmiany ekspresji $>1.5x$) i przedstawione jako lista transkryptów ulegających dysregulacji w odpowiedzi na przemijającą indukcję ekspresji *PTCSC3*. Wyniki z eksperymentu zostały zdeponowane w bazie danych Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) pod numerem dostępowym [GSE32195](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32195). Przywrócenie transkrypcji *PTCSC3* skutkowało zwiększeniem lub zmniejszeniem ekspresji odpowiednio 229 (49%) i 238 (51%) genów. Celem identyfikacji metabolicznych szlaków i interakcji pomiędzy genami o zmienionym przez *PTCSC3* poziomie transkrypcji użyto oprogramowania Ingenuity Pathway Analysis (IPA) (<http://www.ingenuity.com/>). Najistotniejsze kanoniczne szlaki metaboliczne określone w analizie IPA przedstawiono w Tabeli 39. Wzajemne interakcje wzbudzanych lub hamowanych przez *PTCSC3* genów przedstawiono na Rycinie 22.

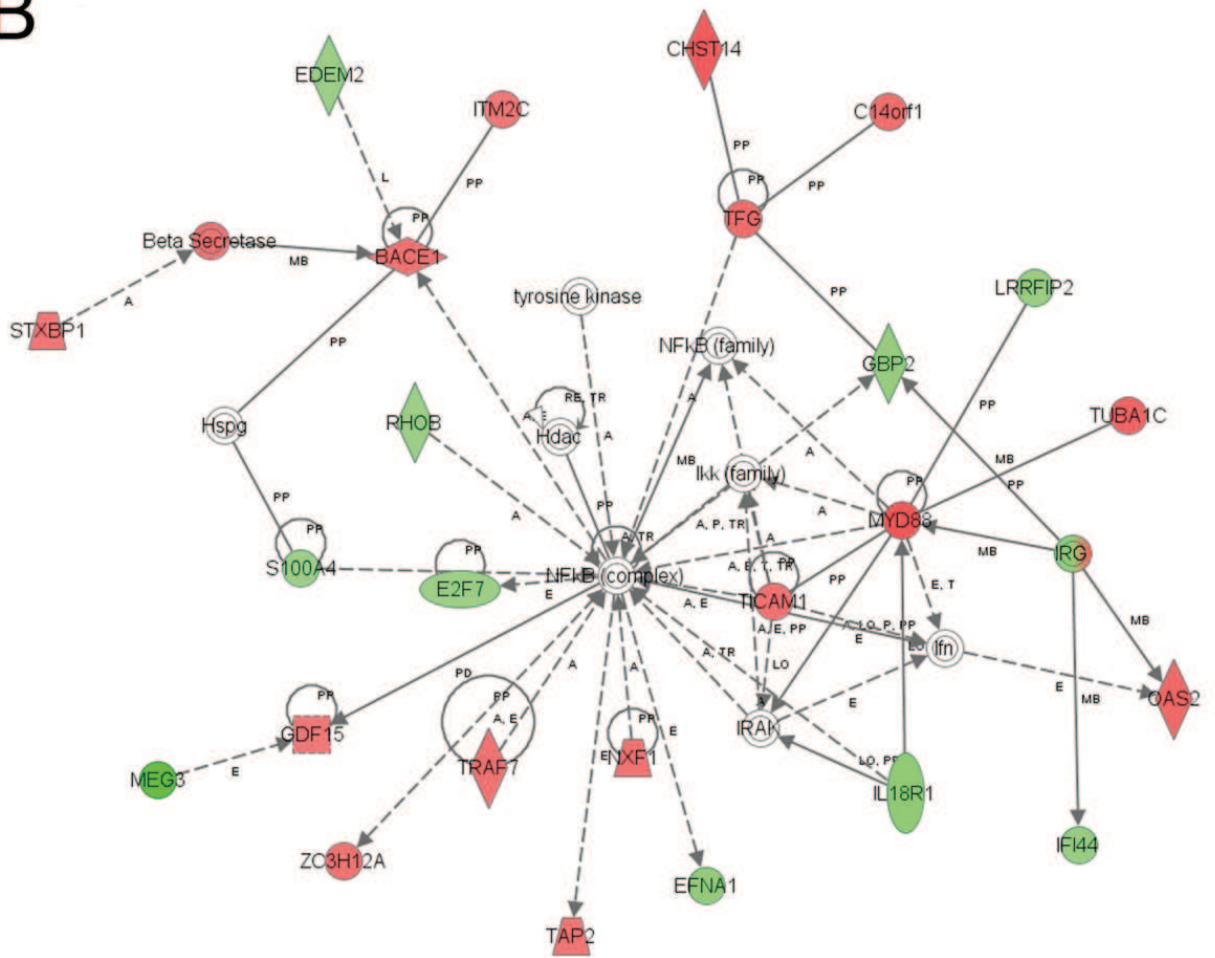
Tabela 39. Analiza IPA. Pięć najistotniejszych szlaków metabolicznych (ang. *canonical pathways*) przypisanych genom, których ekspresja uległa zmianie pod wpływem indukcji ekspresji *PTCSC3* w komórkach TPC-1. Szlaki uszeregowano według malejącej istotności statystycznej (p) obliczonej przy pomocy dokładnego testu Fischer'a. Współczynnik R wyraża stosunek liczby genów w danym szlaku, których ekspresja została zmieniona przez *PTCSC3* do całkowitej liczby genów tworzących dany szlak.

Szlaki metaboliczne	p	Współczynnik R
Procesowanie amyloidu	3.8×10^{-4}	1.07×10^{-1}
Metylacja i transkrypcyjna represja DNA	5.62×10^{-4}	1.74×10^{-1}
Przebieżnictwo neuregulinowe	9.54×10^{-4}	6.86×10^{-2}
Metabolizm metioniny	1.38×10^{-3}	5.13×10^{-2}
Przebieżnictwo zależne od kompleksu beta-gamma	5.01×10^{-3}	5.13×10^{-2}

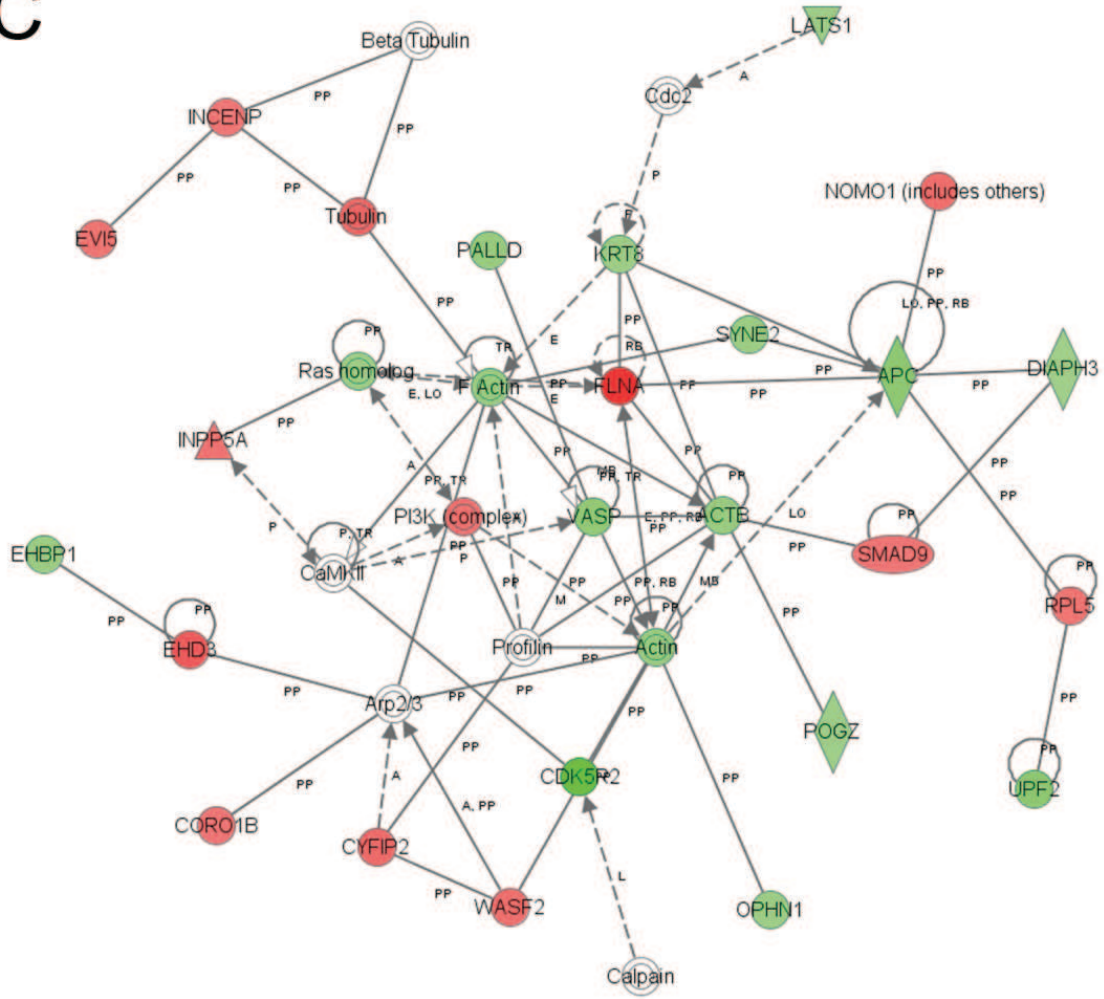
A

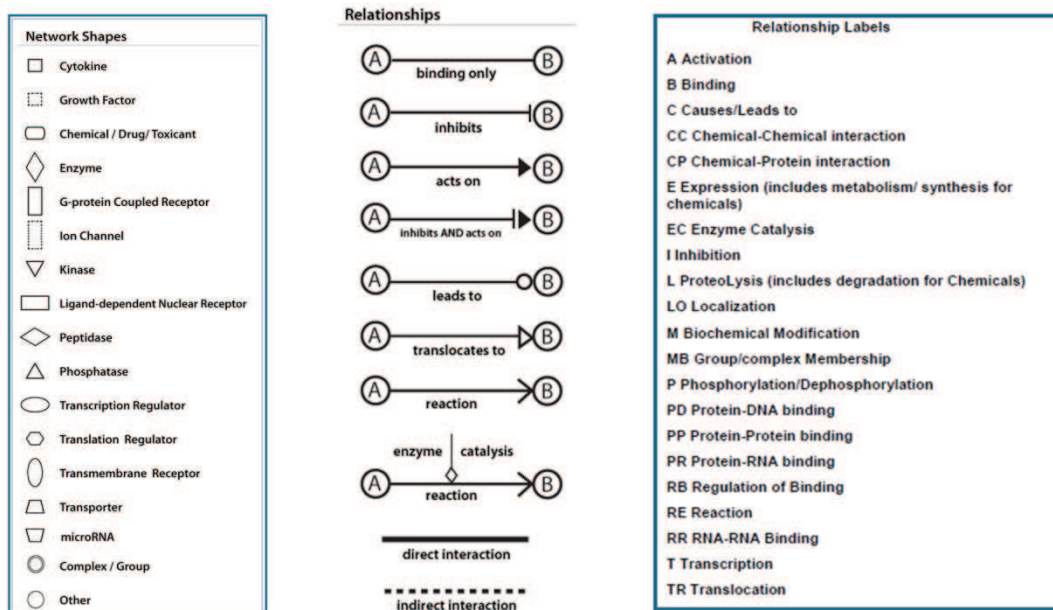


B



C





Rycina 22. Sieć wzajemnych interakcji genów, których transkrypcja uległa zmianie w odpowiedzi na indukcję ekspresji *PTCSC3* w komórkach linii PTC (TPC-1) zidentyfikowana przez program IPA. Na rycinie przedstawiono trzy najważniejsze sieci zawierające największą liczbę genów. (A) Sieć-1: replikacja, rekombinacja i naprawa DNA, ekspresja genów, metabolizm aminokwasów. (B) Sieć-2: mobilność komórek, morfologia guza, śmierć komórek. (C) Sieć-3: Powstawanie i organizacja komórek, funkcje komórkowe, morfologia tkanek.

Geny o zmniejszonej lub zwiększonej ekspresji oznaczono odpowiednio kolorem zielonym lub czerwonym.

6. DYSKUSJA

6.1. Znaczenie badań GWAS w identyfikacji wariantów związanych z chorobami nowotworowymi.

Liczne badania wskazują na stosunkowo ważną rolę genetycznej predyspozycji w rozwoju chorób wieloczynnikowych takich jak cukrzyca typu 2, schizofrenia, czy choroby nowotworowe [103-105]. Jednakże pomimo wysiłków wielu grup badawczych postęp dokonany na przestrzeni ostatnich kilkadziesiąt lat w identyfikacji wariantów odpowiedzialnych za rozwój tych chorób był relatywnie niewielki. Przełomowe wyniki pierwszego projektu GWAS opublikowane w roku 2005 zapoczątkowały nowy rozdział w badaniach poświęconych genetycznemu podłożu chorób o złożonej etiologii. [106]. Od tamtej pory dzięki analizom asocjacyjnym całego genomu zidentyfikowano wiele wariantów związanych z ryzykiem dla różnych jednostek chorobowych. Baza National Human Genome Research Institute (<http://www.genome.gov/gwastudies>) katalogująca wyniki GWAS zawiera dane z 1356 publikacji opisujących związek 7235 SNP'ów z kilkuset fenotypami (<http://www.genome.gov/gwastudies>; dane z września 2012). Dzięki badaniom całogenomowym udało się zidentyfikować co najmniej kilka predysponujących polimorfizmów dla każdej z analizowanych chorób. Przykładowo, tylko w jednej z ostatnio opublikowanych prac badającej związek wariantów genetycznych z wielkością czasu aktywowanej częściowej tromboplastyny (aPTT) i czasu protrombinowego (PT) wskazano osiem wariantów zlokalizowanych na sześciu różnych chromosomach [107]. W raku tarczycy będącym przedmiotem tej pracy w dwóch publikacjach opisano łącznie pięć wariantów umiejscowionych w czterech loci (2q35, 8p12, 9q22.33 i 14q13.3) [68, 75]. Efekt biologiczny tych polimorfizmów nie był duży (dla wariantu rs944289[T] wynosił OR=1.37 (95% CI:1.24-1.54)) [75]. Niewielkie ryzyko jakie niesie allel rs944289[T] jest zgodne z obserwacjami z innych badań GWAS, gdzie efekt zidentyfikowanych polimorfizmów mierzony ilorazem szans wystąpienia danego fenotypu był zwykle niewielki (OR<2.0) [108]. Ma to prawdopodobnie związek z dwoma czynnikami. Po pierwsze, istnieje odwrotna korelacja pomiędzy liczebnością analizowanej populacji a siłą statystyczną badania, które na większej liczbie osobników jest w stanie wykryć bardziej subtelne związki pomiędzy wariantem a chorobą [109]. W badaniu Gudmundsson i wsp. analizowano >1000 chorych z

NMTC i >38 000 kontroli, a więc populacje przekraczające liczebnością większość tradycyjnych badań asocjacyjnych. Po drugie, badania GWAS do niedawna wykorzystywały głównie platformy zawierające stosunkowo częste warianty genetyczne, a więc zgodnie z teorią CD-CV biologiczny efekt niesiony przez te polimorfizmy powinien być niewielki [110]. Co ciekawe, w pierwszym badaniu GWAS na stosunkowo małej próbie 96 chorych ze zwyrodnieniem plamki żółtej zależnej od wieku (ang. *age-related macular degeneration*, AMD) i 50 kontroli zidentyfikowano wariant rs380390 znajdujący się w intronie genu *Complement Factor H (CFH)* w locus 1q31, którego efekt do dzisiaj pozostaje jednym z najwyższych wśród wszystkich polimorfizmów zidentyfikowanych dzięki badaniom całogenomowym (OR=7.4 (95% CI: 2.9-19.0), $p < 10^{-7}$) [106]. Co więcej, wskazanie na gen *CFH* było zaskoczeniem, ponieważ w tamtym czasie nie wiązano go z makulopatią degeneracyjną. Kolejne doniesienia potwierdziły wyniki tego badania i wyjaśniły rolę *CFH* w patogenezie AMD [111-113]. Rezultaty pierwszego GWAS wskazały na potencjalne możliwości badań asocjacyjnych całego genomu w identyfikacji nowych mechanizmów patogenetycznych chorób wieloczynnikowych.

Pomimo niewątpliwych sukcesów badania asocjacyjne całego genomu posiadają kilka ograniczeń, z których najistotniejszy wydaje się problem tak zwanej „brakującej dziedziczności” (ang. *missing heritability*). Równanie dziedziczności związanej z danym zestawem wariantów genetycznych wyraża się stosunkiem dziedziczności ocenionej na podstawie zaobserwowanego ich wspólnego efektu (nominator równania) do całkowitej dziedziczności wywnioskowanej na podstawie obserwacji dla całej populacji (denominator równania) [114]. Innymi słowy w badaniach poświęconych dziedziczności ocenia się na ile zmienność badanego fenotypu można wyjaśnić przez dany zestaw wariantów genetycznych, a na ile poprzez wpływ innych czynników genetycznych i/lub środowiskowych. Dla wspomnianej powyżej pracy poświęconej czasom aPTT i PT osiem wskazanych polimorfizmów wyjaśniało odpowiednio tylko ~29% i ~14% zmienności tych parametrów, co oznacza, że odpowiednio ~71% i ~86% zmienności aPTT i PT ma inne niż wykryte w tym badaniu determinaty [107]. Dla większości badań GWAS frakcja „objaśnionej” dziedziczności nie przekracza 30% [115]. Istnieje wiele hipotez próbujących wyjaśnić ten problem. Najważniejsze z nich sugerują, że część „brakującej dziedziczności” niosą bardzo rzadkie warianty, które mogą być wykryte jedynie poprzez sekwencjonowanie całego genomu oraz, że pewna frakcja dziedziczności może być związana ze zmiennością liczby

kopii (ang. *copy-number variations*, CNVs), których do niedawna nie można było analizować w oparciu o platformy mikromacierzowe [116-118]. Innym wyjaśnieniem jest możliwość przeszacowywania wielkości denominatora w równaniu dziedziczności, który wywodzi się na podstawie danych z obserwacji dla całej populacji [114]. Ostatnim i najprostszym rozwiązaniem problemu „brakującej dziedziczności” jest założenie istnienia wielu częstych wariantów genetycznych związanych z danym fenotypem, które nie zostały jeszcze wykryte (nominator równania dziedziczności) [56, 119].

Większość wariantów genetycznych wykrytych dzięki badaniom GWAS lokalizuje się w regionach, które nie zawierają genów lub zawierają transkrypty, których funkcja nie była dotychczas związana z analizowaną chorobą [120]. To spostrzeżenie odnosi się także do badania Gudmundsson i wsp., w którym predysponujący do NMTC polimorfizm znajdował się w 249 kb bloku LD nie zawierającym opisanych w bazie RefSeq genów. Lokalizacja wariantów GWAS w tak zwanych „pustyniach genowych” (ang. „*gene desert region*”) powoduje, że dla dużego odsetka tych badań nie udało się do tej pory wyjaśnić mechanizmu predyspozycji zidentyfikowanych SNP’ów. Może to wynikać z faktu, że markery znajdujące się na platformach mikromacierzowych są wybierane głównie w oparciu o lokalizację i częstość występowania w regionach wykazujących LD, a nie w oparciu o ich potencjalną funkcję związaną na przykład ze zmianą aktywności regionów promotorowych. Projektując zestawy mikromacierzowe zakłada się, że polimorfizmy umieszczone na platformie mają wskazać (poprzez nierównowagę sprzężeń) locus związany z danym fenotypem, w którym znajdują się funkcjonalne warianty sprzężone z predysponującymi genami. Oczywiście tryb selekcji markerów nie wyklucza możliwości, że wytypowane w badaniach GWAS warianty są bezpośrednio zamieszane w proces chorobowy [119].

W niniejszej pracy wykazano, że polimorfizm rs944289 znajduje się w bloku nierównowagi sprzężeń z niezidentyfikowanym dotychczas, dużym niekodującym międzygenowym genem RNA (*PTCSC3*). Udowodniono, że ekspresja tego transkryptu jest związana z genotypem polimorfizmu, który poprzez allel ryzyka (T) osłabia aktywność promotora dla *PTCSC3* stymulowaną przez białka wiążące się z sekwencją CCAAT (C/EBP). Wstępna charakterystyka *PTCSC3* wskazała, że gen ten ulega selektywnej transkrypcji w tarczycy, jego ekspresja jest silnie hamowana w tkance PTC, a jej odtworzenie w komórkach raka tarczycy hamuje ich wzrost. Przedstawione wyjaśnienie mechanizmu działania rs944289 jest jednym z nielicznych doniesień, które wskazały i scharakteryzowały

strukturalnie oraz funkcjonalnie nieznan wcześniej gen znajdujący się w bliskim sąsiedztwie wariantu wykrytego w badaniu asocjacyjnym całego genomu oraz zaproponowały mechanizm w jakim może on być odpowiedzialny za predyspozycję do choroby.

6.2. Struktura *PTCSC3*.

Sekwencja eksonów drugiego, trzeciego i czwartego *PTCSC3* pokrywa się prawie całkowicie z transpozycyjnymi elementami (ang. *transposable elements*, TE). TE dzieli się na retrotranspozony (klasa I) i transpozony DNA (klasa II). Te pierwsze dodatkowo dzieli się na autonomiczne zawierające sekwencje LTR i nieautonomiczne zawierające elementy SINE i LINE [121].

Elementy LINE stanowiące około 21% całego ludzkiego genomu składają się z trzech rodzin- LINE1, LINE2 i LINE3, z których tylko pierwsza wykazuje aktywność transkrypcyjną. Długość LINE1 zwanych także elementami L1 wynosi ~6 kb. Elementy te posiadają promotor dla polimerazy II oraz dwie ORF (ORF1 i ORF2) kodujące białka niezbędne do retrotranspozycji (ORF1p i ORF2p). Białko ORF1p służy jako „białko opiekuńcze” (ang. *chaperone*) dla LINE1 RNA. ORF2p posiada dwie domeny (endonukleazową i odwrotnej transkryptazy), przy pomocy których dokonuje przecięcia nici DNA i wbudowania komplementarnego do LINE1 RNA fragmentu DNA [122, 123]. Okresowo reakcja odwrotnej transkrypcji (biegnąca od końca 3' do końca 5' LINE1 RNA) ulega przedwczesnej terminacji skutkując powielaniem niekompletnych i niefunkcyjnych elementów. W efekcie długość większości LINE1 w ludzkim genomie nie przekracza 900 nt [124].

Elementy SINE składają się z krótkich, niekodujących sekwencji (100-400 nt) zawierających promotor dla polimerazy III. Większość elementów SINE w ludzkim genomie należy do rodziny Alu (elementy Alu), nazwanej tak z powodu posiadania miejsca dla restryktazy AluI [125]. SINE z rodziny Alu wywodzą się z RNA wchodzącego w skład kompleksu rybosomalnego 7SL i wykorzystują do retrotranspozycji elementy L1 [126, 127].

Retrotranspozony LTR są autonomicznymi mobilnymi elementami składającymi się z genów *gag*, *pol* i *env*, które kodują proteazę, odwrotną transkryptazę, RNazę H i integrazę. Transpozycja zachodzi w mechanizmie podobnym do powielania się retrowirusów (przy użyciu tRNA jako startera w reakcji odwrotnej transkrypcji zachodzącej w cytoplazmie), co

odróżnia LTR od LINE1 [128]. Obecnie większość elementów LTR znajdujących się w ludzkim genomie jest nieaktywna [124]. Transpozony DNA (klasa II mobilnych elementów) w przeciwieństwie do retrotranspozonów nie zwiększają całkowitej ilości DNA, ponieważ ulegają transpozycji w mechanizmie „wytnij-wklej” (ang. *cut-paste*) a nie „kopiuj-wklej” (ang. *copy-paste*). Elementy te kodują enzym zwany transpozyną, który wycina transpozon, a następnie wkleja go do innego miejsca w genomie. Ostatecznie liczba kopii transpozonu pozostaje taka sama, zmienia się tylko jego lokalizacja [129].

Eksony drugi (60 nt) i trzeci (91 nt) *PTCSC3* pokrywają się całkowicie z sekwencjami charakterystycznymi dla odpowiednio elementów LINE (rodzina L1) i SINE (rodzina Alu), podczas gdy ekson czwarty (647 nt) w około 75% wykazuje homologię z powtarzającymi się sekwencjami LTR (rodzina ERVL). Całkowicie unikalnym fragmentem *PTCSC3* jest ekson pierwszy (356 nt), którego sekwencja nie wykazuje podobieństwa do żadnego innego regionu ludzkiego genomu. Wydaje się więc, że *PTCSC3* wyewoluował częściowo z mobilnych elementów (ME) tworząc specyficzną mozaikę powtórzeń i unikalnej sekwencji. ME składające się na około 45% ludzkiego genomu były uważane w przeszłości za „śmieciowe DNA” (ang. „*junk DNA*”), których „samolubne” powielanie się zwiększało objętość genomu bez pożytku dla komórki gospodarza. Obecnie wiemy, że ME były istotnymi elementami biorącymi udział w ewolucji genomu różnych organizmów, służącymi jako rezerwuar sekwencji, które poprzez powielanie, insercję, delecję i/lub mutacje dawały początek genetycznym innowacjom [130, 131]. Dla przykładu, pewne dane wskazują, że mobilne elementy brały udział w tworzeniu genów kontrolujących powstawanie łożyska, a więc były kluczowe w wyłonieniu się ssaków jako odrębnej gromady zwierząt [132]. Do tej pory zidentyfikowano co najmniej kilkadziesiąt transkryptów wywodzących się z mobilnych elementów [124, 133, 134]. Biorąc pod uwagę fakt, że wiele kodujących genów poprzez nabycie fragmentów regulatorowych jakie niosą ME moduluje swoją ekspresję, jest prawdopodobne, że część lncRNA (których główna funkcja w komórce polega na kontroli ekspresji innych genów) mogła wyewoluować z mobilnych elementów [135-138].

6.3 *PTCSC3* jako duży niekodujący RNA gen.

Wprowadzenie i rozwój automatycznego sekwencjonowania metodą Sangera w latach '90 ubiegłego wieku umożliwił efektywniejsze odczytywanie informacji niesionej przez

DNA. Jednym z pierwszych zastosowań tej techniki było mapowanie EST do odpowiednich regionów w genomie celem identyfikacji loci ulegających transkrypcji, a więc zawierających potencjalne geny [139]. Wstępne wyniki wskazywały, że ludzki genom może zawierać nawet >100 000 genów. Jednakże zakończenie w kwietniu 2003 roku projektu Human Genome Project (HGP) (<http://www.genome.gov/10001772>) oraz dalsze badania dostarczyły zaskakujących danych sugerujących, że całkowita liczba kodujących genów nie przekracza ~25 000, a więc większość z ujawnionych transkryptów nie koduje białek [124]. Kolejne prace wskazały, że transkrypty nie ulegające translacji zwane niekodującymi RNA (ang. *non-coding RNA*, ncRNA) stanowią znaczącą frakcję wszystkich genów [140, 141]. W zależności od długości dzieli się je na duże niekodujące RNA (ang. *large non coding RNA*, lncRNA; >200 nt) i małe niekodujące RNA (ang. *small non coding RNA*; ≤200 nt). Nazwa duże niekodujące RNA odnosi się do różnych typów transkryptów takich jak duże międzygenowe niekodujące RNA (ang. *large intergenic non-coding RNA*, lincRNA), pseudogeny (ang. *pseudogenes*), enhancerowe RNA (ang. *enhancer RNA*), długie intronowe niekodujące RNA (ang. *long intronic non-coding RNA*), antysensowne RNA (ang. *antisense RNA*), długie RNA związane z promotorem (ang. *promoter-associated long RNA*), transkrybowane ultrakonserwatywne regiony (ang. *transcribed ultraconserved regions*), niekodujące RNA związane z powtórzeniami (ang. *repeat-associated non-coding RNA*) i długie niekodujące transkrypty indukowane stresem (ang. *long stress-induced non-coding transcripts*). Małe ncRNA także składają się z kilku typów transkryptów, z których najlepiej poznane są mikroRNA [142]. Liczne badania na przestrzeni ostatnich lat dostarczyły wielu dowodów na udział mikroRNA w patogenezie różnych nowotworów takich jak rak wątrobowokomórkowy, rak jelita grubego, rak piersi, rak jajnika, rak płuc, białaczki i szeregu innych [143]. Zidentyfikowanie polimorfizmu rs2910164 w łańcuchu pre-miR-146a zwiększającego ryzyko dla PTC (genotyp [GC]: OR=1.62 (95% CI:1.3-2.0), p=0.000007) było jednym z ważniejszych odkryć wskazujących na potencjalną rolę ncRNA w patogenezie NMTC [69]. O ile dzisiejsza wiedza na temat mikroRNA jest stosunkowo duża, o tyle mechanizmy działania i funkcje lncRNA pozostają relatywnie mało poznane. Prace próbujące zdefiniować duże niekodujące geny wskazują na kilka istotnych cech odróżniających te transkrypty od tak zwanego „szumu” transkrypcyjnego, czyli niespecyficznego przepisywania informacji z DNA na RNA. Duże niekodujące RNA charakteryzują się występowaniem epigenetycznych znaczników (wspólnych z genami

kodującymi), takich jak trimetylacja lizyny 36 w histonie H3 w genie (H3K36me3) i trimetylacja lizyny 4 w histonie H3 (H3K4me3) w promotorze. Geny te zazwyczaj są transkrybowane przez polimerazę RNA II oraz stosunkowo często ulegają poliadenylacji, splicing'owi i tkankowo-specyficznej ekspresji, która jest regulowana poprzez tradycyjne czynniki transkrypcyjne [144]. Wstępna charakterystyka *PTCSC3* pokazała, że gen ten ulega tarczycowo-specyficznej ekspresji, aktywność jego promotora jest regulowana poprzez powszechnie znane czynniki transkrypcyjne (C/EBP α i C/EBP β), a końcowy produkt genu ulega poliadenylacji i składaniu (splicing'owi). Kluczowe dla ostatecznego zaklasyfikowania danego transkryptu do lincRNA jest ustalenie jego kodującego potencjału. Duży odsetek lincRNA posiada różnej długości ORF. Jednym z pierwszych i najlepiej poznanych lincRNA jest gen *XIST* znajdujący się w locus Xq13.2 [145]. *XIST* posiada podwójny motyw „spinki do włosów” (ang. *hairpin*) w domenie RepA łańcucha RNA, poprzez którą wiąże Polycomb Repressive Complex 2 (PRC2), co prowadzi do epigenetycznego wyciszenia chromosomu X. Aktywny chromosom X hamuje ekspresję *XIST* poprzez transkrypcję genu *TSIX* zlokalizowanego na przeciwległej do *XIST* nici DNA (antysensowny RNA) [146]. Z uwagi na obecność ramki odczytu *XIST* został początkowo zakwalifikowany jako kodujący. Dopiero późniejsze prace wskazały na jego niekodujący charakter [147]. *PTCSC3* posiada krótką 95 aminokwasową ORF. Analiza *in silico* jego sekwencji wskazała, że trzy z czterech eksonów zawierają motywy SINE, LINE i LTR. Przemawia to za niekodującym charakterem tego genu, ponieważ elementy te, jakkolwiek mogą składać się na retrotranspozony i kodować białka, to w konstelacji jaka występuje w *PTCSC3* wydaje się to mało prawdopodobne [148]. W niezależnej od niniejszego badania publikacji zespół badaczy (Cabili i wsp.) z Massachusetts Institute of Technology (USA) analizując (poprzez głębokie sekwencjonowanie RNA) transkryptomy różnych tkanek (w tym tarczycy) zidentyfikował ponad 8000 nowych lincRNA [149]. Celem rozróżnienia genów niekodujących od kodujących autorzy tego badania używali metody PhyloCSF opracowanej przez Linn i wsp. polegającej na projektowaniu filogenetycznych modeli reprezentujących ewolucję kodonów w genach kodujących i ewolucję nukleotydowych tripletów w regionach niekodujących [150]. Dodatkowo dla każdego z transkryptu analizowano wszystkie potencjalne ramki odczytu pod kątem ewentualnego kodowania jakiegokolwiek z 31 912 białkowych domen skatalogowanych w bazie Pfam [151]. Badacze ci zidentyfikowali w tkance tarczycowej transkrypt XLOC_010996 (chr14:36598757-36645961, UCSC Genome Browser,

GRCh37/h19) odpowiadający *PTCSC3*, który został zakwalifikowany do klasy lncRNA [149]. Analiza struktury *PTCSC3* przeprowadzona w niniejszej rozprawie oraz wyniki pracy Cabili i wsp. wskazują, że potencjał kodujący *PTCSC3* jest niewielki, aczkolwiek nie można go całkowicie wykluczyć wyłącznie na podstawie analizy bioinformatycznej. Kolejnym etapem charakterystyki tego genu mogłoby być wykonanie testu western blot z użyciem przeciwciał przeciwko wybranym epitopom potencjalnego peptydu *PTCSC3*. Empiryczne udowodnienie translacji dla zidentyfikowanej 95 aminokwasowej ORF *PTCSC3* byłoby ambitnym zadaniem wykraczającym jednakże poza cele tej rozprawy.

6.4. Funkcja *PTCSC3*.

Do niedawna na patogenezę nowotworów patrzono głównie poprzez pryzmat mutacji w genach białkowych regulujących kluczowe dla inicjacji i progresji choroby nowotworowej procesy takie jak proliferacja, apoptoza, różnicowanie komórek czy naprawa DNA [152]. Liczne badania ostatnich lat skupiające się na mechanizmach kontrolujących transkrypcję zwróciły uwagę na fakt, że zmiany w epigenetycznej represji genów mediowane przez lncRNA mogą zaburzać delikatną homeostazę komórkową i także inicjować nowotworową transformację [153]. Jednym z lepiej opisanych lncRNA wpływających na ekspresję innych transkryptów jest *HOTAIR* zlokalizowany w locus 12q13.13. Gen ten posiada dwie funkcjonalne domeny wiążące PRC2 (koniec 5'RNA) i LSD1/CoREST1 (koniec 3'RNA), poprzez które wpływa na genomową lokalizację PRC2, co prowadzi do zmiany w konformacji chromatyny i ostatecznie represji transkrypcji [154, 155]. Nadekspresja *HOTAIR* w raku piersi, wątroby i jelita grubego jest związana ze zwiększonym ryzykiem pojawienia się przerzutów i mniejszą przeżywalnością, co wskazuje na pro-onkogenne działanie tego genu w różnych tkankach [156-158]. *HOTAIR* jest transkryptem zlokalizowanym międzygenowo (ang. *intergenic*) wpływającym na ekspresję genów zlokalizowanych w odległych od 12q13.13 loci. Jego działanie jest określane mianem *trans* (ang. *trans-acting*). Innym przykładem lncRNA są transkrypty regulujące lokalną ekspresję genów, co określa się mianem działania *cis* (ang. *cis-acting*). Przykładem takiego lncRNA jest *ANRIL* zlokalizowany w 9p21.3. Gen ten jest transkrybowany z przeciwległej do locus *INK4A/ARF* nici DNA (antysensowny RNA). Locus *INK4A/ARF* zawiera geny supresorowe *INK4n/ARF/INK4a*, *p16/CDKN2A* i *p15/CDKN2B*, które kontrolują cykl komórkowy i

proces starzenia się komórki. W raku prostaty zwiększona ekspresja *ANRIL* poprzez lokalne zmiany w strukturze chromatyny wywoływane przez PRC1 i PRC2 hamuje ekspresję genów supresorowych *INK4n/ARF/INK4a* i *p15/CDKN2B* [159, 160].

HOTAIR i *ANRIL* poprzez zwiększoną ekspresję w tkance guza i mechanizm działania promujący chorobę nowotworową wykazują cechy charakterystyczne dla onkogenów. Duże niekodujące RNA potrafią także wywoływać działania charakterystyczne dla genów supresorowych. Przykładem może być *GAS5*, który poprzez antagonizowanie receptora glukokortykosteroidowego (ang. *glucocorticoid receptor*, GR) moduluje procesy wpływające na przeżycie i metabolizm komórek. Koniec 3'*GAS5* RNA zawiera motyw podobny do sekwencji DNA wiążącej receptor GR. Związanie GR z *GAS5* RNA uniemożliwia temu pierwszemu oddziaływanie na promotory genów regulowanych przez glukokortykosteroidy [161]. Wykazano, że ekspresja *GAS5* ulega supresji w guzach piersi, a stymulacja jego transkrypcji w liniach komórkowych tego raka indukuje apoptozę i hamuje proliferację [162]. Innym przykładem dużego niekodującego RNA o cechach supresora jest gen *MEG3* znajdujący się na chromosomie 14q32 [163]. *MEG3* stymuluje ekspresję TP53 wzmacniając jednocześnie wiązanie p53 z promotorami docelowych genów [164]. Utrata ekspresji *MEG3* została udokumentowana w neuroblastomie, guzach wątroby, glejakach, guzach przysadki mózgowej i białaczkach [165-169]. Restytucja transkrypcji *MEG3* w komórkach raka szyjki macicy, raka wątrobowokomórkowego i glejakach hamowała ich wzrost poprzez indukcję apoptozy [166-168].

Przytoczone powyżej przykłady lncRNA obrazują różnorodność mechanizmów, poprzez które transkrypty te wpływają na ekspresję docelowych genów. Jest to cecha istotnie odróżniająca lncRNA od mikroRNA, które zawiadują transkrypcją poprzez wiązanie się z odpowiednią sekwencją docelowego genu zlokalizowaną w regionie 3' nie ulegającym translacji (ang. *3' untranslated region*, *3'UTR*) [170]. Jak pokazano powyżej, duże niekodujące RNA wpływają na inne transkrypty poprzez lokalną (działanie *cis*: *XIST*, *ANRIL*) lub odległą (działanie *trans*: *HOTAIR*) zmianę w konformacji chromatyny, bezpośredni wpływ na transkrypcję (*MEG3*) lub działania kompetycyjne dla receptorowego białka (*GAS5*). Mechanizmy te nie wyczerpują jednakże całego spektrum działań tych genów. Kilka z ostatnio opublikowanych prac wskazało na inne, potencjalnie ważne mechanizmy kontroli ekspresji regulowane przez lncRNA. Wykazano na przykład, że niekodujący transkrypt *MALAT-1* znajdujący się na chromosomie 11q13.1 potrafi zmieniać

aktywność genów poprzez wpływ na procesy składania mRNA [171]. Wzmoczona ekspresja *MALAT-1* w raku niedrobnokomórkowym płuca była związana ze zwiększonym ryzykiem wystąpienia przerzutów i zmniejszoną przeżywalnością [172]. Inne badanie wskazało na komplementarne wiązanie dużych niekodujących RNA poprzez sekwencje Alu z przekaźnikowymi RNA, co prowadzi do rozpoznania kompleksu lncRNA-mRNA przez białko STAU1 i w konsekwencji do degradacji kwasu rybonukleinowego [173]. Ciekawym mechanizmem kontrolującym poziom mRNA w komórce jest zaproponowana przez Polisenio i wsp. hipoteza sugerująca, że lncRNA mogą sekwestrować mikroRNA, co prowadzi do ich zmniejszonego wpływu na 3'UTR genów docelowych [174]. Złożoność mechanizmów działania lncRNA sprawia, że określenie funkcji *PTCSC3* stanowi duże wyzwanie. Wstępna charakterystyka tego lncRNA sugeruje supresorowy typ działania. Przynależność danego genu do supresorów określa się na podstawie trzech głównych kryteriów. Po pierwsze dany transkrypt powinien ulegać funkcjonalnej deaktywacji w tkance guza (supresja transkrypcji, mutacje, delecje i inne). Po drugie, restytucja jego ekspresji w tkance nowotworowej powinna hamować wzrost komórek w modelach *in vitro* i *in vivo*. Po trzecie, deaktywacja genu w modelu *in vivo* powinna prowadzić do transformacji nowotworowej, formowania guza i/lub defektów rozwojowych [175]. W analizie ekspresji *PTCSC3* w tkance guza i zdrowej tkance tarczycowej u 46 chorych z PTC wykazano, że transkrypcja tego lncRNA ulegała silnej supresji w tkance raka w każdej z testowanych par ($p=2.84 \times 10^{-14}$) (Rycina 17). Ta obserwacja jest zgodna z wynikami oceny ekspresji *PTCSC3* w liniach komórkowych raka tarczycy metodą RT-PCR, w której udokumentowano utratę ekspresji *PTCSC3* w 14 z 18 testowanych linii i znaczącą supresję w pozostałych 4 (Rycina 14). Wyniki te wskazują, że ekspresja *PTCSC3* jest silnie hamowana w komórkach tarczycy, które uległy nowotworowej transformacji. Indukcja transkrypcji *PTCSC3* w liniach komórkowych PTC (TPC-1 i BCPAP) hamowała ich wzrost (odpowiednio $p=0.002$ i $p=0.019$) (Rycina 21). Powyższe dane sugerują, że *PTCSC3* posiada właściwości genu supresorowego. Mechanizm w jaki *PTCSC3* hamuje wzrost komórek pozostaje nieznany. Wyniki analizy porównawczej ekspresji genów w komórkach TPC-1 (z ektopową transkrypcją *PTCSC3* i komórkach niewykazujących transkrypcji *PTCSC3*) wskazały, że gen ten wpływał na ekspresję szeregu ważnych transkryptów biorących udział w patogenezie PTC (AKT [176], PI3K [177], USF2 [178], RHOB [179], MOAP1 [180], PLK1 [181], HSP90 [182], PTMA [183], S100A4 [184, 185], GDF15 [186] i MYD88 [187]) (Rycina 22, Gene Expression Omnibus: [GSE32195](#)).

Geny te są ważnymi czynnikami wpływającymi bezpośrednio lub pośrednio na procesy proliferacji i apoptozy. Hamowanie wzrostu komórek raka tarczycy mogło być efektem działania *PTCSC3* na jeden lub więcej z tych transkryptów. Siła z jaką *PTCSC3* oddziaływał na proliferację nie była znaczna. Można to wiązać z efektywnością transfekcji, która dla użytych w eksperymencie linii komórkowych nie przekraczała 50%. Wydaje się, że dalsze badania z wykorzystaniem linii, w których gen ten będzie ulegał stabilnej (system pTet-On/Off) ekspresji może lepiej zobrazować wpływ *PTCSC3* na proliferację komórkową. Należy także zaznaczyć, że jakkolwiek test alamarBlue[®] jest wykorzystywany do oceny wzrostu komórek to w rzeczywistości mierzy on ich aktywność metaboliczną (żywołność), która jest tylko pośrednio zależna od tempa proliferacji. Nie można zatem wykluczyć możliwości, że *PTCSC3* hamował metabolizm komórek raka tarczycy nie wpływając na ich wzrost. Ten mało prawdopodobny, aczkolwiek możliwy aspekt działania *PTCSC3* także należy wziąć pod uwagę.

Lista genów ulegających dysregulacji w odpowiedzi na przywrócenie ekspresji *PTCSC3* w komórkach raka tarczycy obejmowała 467 transkrypty. *PTCSC3* wpływał bezpośrednio na niewielki odsetek tych transkryptów, pozostałe uległy dysregulacji na zasadzie „efektu domina” w przebiegu wzajemnych międzygenowych interakcji. Dlatego w analizie wpływu danego genu na ekspresję innych transkryptów ważne jest wskazanie szlaków metabolicznych do jakich należą te ostatnie, co pozwala pośrednio określić rolę badanego transkryptu w komórce. Trzy najważniejsze szlaki metaboliczne i ścieżki sygnałowe jakie modulował *PTCSC3* to powstawanie i procesowanie amyloidu ($p=3.80 \times 10^{-4}$), metylacja DNA i represja transkrypcji ($p=5.62 \times 10^{-4}$) oraz przekaźnictwo neuregulinowe ($p=9.54 \times 10^{-4}$) (Tabela 39). Dane sugerujące związek amyloidu i przekaźnictwa neuregulinowego z DTC wydają się zaskakujące. O ile złogi amyloidu w podścielisku tkanki nowotworowej u chorych z MTC są stosunkowo częstym zjawiskiem, to do tej pory opisano zaledwie kilka przypadków współwystępowania skrobiawicy tarczycy i NMTC [188-194]. U pacjentów z MTC depozycja amyloidu ogranicza się zwykle do guza i jest efektem niekontrolowanej produkcji kalcytoniny przez transformowane nowotworowo komórki C [188, 194]. Złogi amyloidu są skutkiem, a nie przyczyną rozwoju MTC. Dla postaci nierdzeniastych wole amyloidowe jest zwykle jedną z manifestacji układowej amyloidozy [190, 193]. Nie ma dowodów, że skrobiawica inicjuje proces nowotworowy, a więc współwystępowanie amyloidu w tkance tarczycowej z NMTC traktuje się jako koincydencję

dwóch rzadkich i nie wykazujących związku przyczynowo-skutkowego chorób. Rola przekąźnictwa neuregulinowego w patogenezie raka tarczycy wydaje się być intrygująca w świetle danych wskazujących na bardzo silny związek ekspresji genu *NRG1* kodującego białko neuregulinę 1 z genotypem wariantu rs2439302 ($p=9.1 \times 10^{-91}$), który został zidentyfikowany w badaniu GWAS jako predysponujący do NMTC (OR=1.36 (95% CI: 1.23-1.50), $p=2.0 \times 10^{-9}$) [68]. Neureguliny należą do rodziny naskórkowych czynników wzrostu (ang. *epidermal growth factor family*, EGF) odgrywających ważną rolę w procesie dojrzewania układu nerwowego, serca i gruczołów piersiowych. Wykazano, że metylacja *NGR1* i w efekcie wyciszenie jego ekspresji jest częstym zjawiskiem w raku piersi. Supresja transkrypcji *NGR1* w linii komórkowej tego raka stymulowała proliferację, co wskazuje na supresorowe właściwości tego genu [195]. Jakkolwiek związek *PTCSC3* z przekąźnictwem neuregulinowym wydaje się interesujący to najwyższą wartość współczynnika liczby genów ulegających dysregulacji do całkowitej liczby genów tworzących dany szlak odnotowano dla ścieżek związanych z metylacją DNA i represją transkrypcji, co może wskazywać, że są to jedne z najważniejszych mechanizmów działania *PTCSC3* (Tabela 39). Niezależna od analizy kanonicznych szlaków ocena wzajemnych interakcji genów ulegających dysregulacji w odpowiedzi na przywrócenie ekspresji *PTCSC3* w komórkach raka tarczycy wskazała, że gen ten wpływa na transkrypty regulujące replikację, rekombinację i naprawę DNA (Sieć-1), mobilność komórek, morfologię guza i śmierć komórki (Sieć-2) oraz powstawanie i organizację komórek, funkcje komórkowe i morfologię tkanek (Sieć-3) (Rycina 22). Wyniki analizy wzajemnych interakcji transkryptów oraz kanonicznych ścieżek metabolicznych są zgodne z obserwacjami wiążącymi funkcje lincRNA z regulacją ekspresji genów. Niektóre z doniesień wskazują, że lincRNA w przeciwieństwie do genów białkowych wykazują w dużym odsetku tkankową specyficzność, co może sugerować ich istotną rolę w różnicowaniu i dojrzewaniu tkanek [142]. Dane z analizy IPA wskazujące na prawdopodobny związek *PTCSC3* z różnicowaniem, dojrzewaniem i morfologią tkanki tarczycowej mogą wyjaśniać jego tarczycowo-specyficzną ekspresję. Tkankowa specyficzność *PTCSC3* może także wskazywać na udział w swoistych dla tarczycy funkcjach. Wszystkie poznane do tej pory tarczycowo-specyficzne geny (*TG*, *TPO*, *NIS*, *TSHR*) biorą udział w produkcji hormonów tarczycowych. Wyniki eksperymentu mikromacierzowego nie wykazały jednakże wpływu *PTCSC3* na poziom transkrypcji żadnego z powyższych genów. Jakkolwiek lincRNA regulują ekspresję głównie na poziomie transkrypcji (*XIST*, *HOTAIR*, *ARNIL*) to mogą one także

modelować procesy posttranslacyjne (*GAS5*). Nie można całkowicie wykluczyć wpływu *PTCSC3* na tarczycowo-specyficzne geny w mechanizmie innym niż kontrola transkrypcji. Jednakże z uwagi na obserwowane zmiany w poziomach RNA dla 467 transkryptów, wyniki analizy kanonicznych szlaków metabolicznych oraz wzajemnych interakcji wzbudzanych lub hamowanych genów wydaje się, że transkrypcyjna regulacja ekspresji jest zapewne głównym mechanizmem działania tego lincRNA. Jest więc mało prawdopodobne, że tkankowa specyficzność *PTCSC3* ma związek z produkcją hormonów tarczycowych, aczkolwiek aby całkowicie wykluczyć ten aspekt potencjalnego działania *PTCSC3* należałoby przeprowadzić dalsze, ukierunkowane na tą hipotetyczną funkcję badania.

6.5. Wpływ rs944289 na ekspresję *PTCSC3*.

Polimorfizm rs944289 znajduje się 3.2 kb od początku pierwszego eksonu *PTCSC3* w regionie dalszego promotora (Rycina 18). Analiza *in silico* wskazała na czynniki transkrypcyjne C/EBP α i C/EBP β jako potencjalnie wiążące się z DNA w miejscu polimorfizmu, który poprzez allel ryzyka zmniejszał siłę wiązania wyrażoną jako wpływ pozycji w macierzy (PWM). Wyniki testu opóźnienia w żelu potwierdziły *in vitro* wiązanie obydwu TF z fragmentem DNA zawierającym allel C (Rycina 19). W podwójnym teście lucyferazy C/EBP α i C/EBP β aktywowały promotor dla *PTCSC3* zarówno dla allelu C (odpowiednio p=0.0009 i p=0.0014) jak i allelu T (odpowiednio p=0.0019 i p=0.0048). Allel ryzyka (T) osłabiał aktywność promotora generowaną przez C/EBP α i C/EBP β (odpowiednio p=0.026 i p=0.048) (Rycina 20).

Rodzina czynników wiążących się z sekwencją CCAAT składa się z sześciu protein: C/EBP α , C/EBP β , C/EBP γ , C/EBP δ , C/EBP ϵ i C/EBP θ [196]. Nazwa grupy pochodzi od pierwszego zidentyfikowanego członka rodziny (C/EBP α), który wykazywał zdolność wiązania się z motywem CCAAT występującym w promotorach różnych genów [197]. C/EBP α i C/EBP β są kodowane przez geny składające się z pojedynczych eksonów zlokalizowane odpowiednio w loci 19q13.3 i 20q13.3. Obydwa białka posiadają szereg izoform na skutek inicjacji translacji przez kilka alternatywnych kodonów startowych. Każda z postaci posiada C-kończową domenę podstawowego zamka leucynowego (bZIP) odpowiedzialnego za wiązanie z DNA oraz w zależności od długości łańcucha białkowego funkcjonalną domenę znajdującą się w N-kończowym fragmencie danej izoformy. Dla

C/EBP α rozróżniamy izoformy p42 i p30, których nazwy odnoszą się do ich molekularnej masy (odpowiednio 42 kDa i 30 kDa). Uważa się, że forma dłuższa (p42) posiada pełną aktywność czynnika transkrypcyjnego. C/EBP β posiada izoformy LAP* (38 kDa), LAP (34 kDa) i LIP (20 kDa), z których dwie pierwsze uznaje się za aktywne czynniki transkrypcyjne [198, 199]. Klasyczne miejsce wiążące dla C/EBP zawiera palindromową sekwencję RTTGC*GYAAAY, gdzie R=A lub G i Y=C lub T [200]. Nukleotydy w pozycji ± 3 i ± 4 odgrywają kluczową rolę w identyfikacji przez C/EBP miejsca wiążącego. Następna decydująca o sile wiązania jest pozycja ± 2 . Największą zmienność odnotowano dla nukleotydów ± 1 i ± 5 [201]. Czynniki należące do rodziny bZIP takie jak C/EBP posiadają zdolność rozpoznawania miejsc zawierających tylko połowę sekwencji palindromowej (ang. *half site*) [200, 202, 203]. C/EBP oddziałują na DNA w formie homo- lub heterodimerów. Te ostatnie powstają poprzez wiązanie cząsteczek C/EBP w praktycznie dowolnych kombinacjach [204]. Dodatkowo C/EBP posiadają zdolność bezpośredniej interakcji z innymi czynnikami transkrypcyjnymi, w tym bardzo istotnymi z punktu widzenia fizjologii komórki tarczycowej takich jak NF- κ B czy NKX2-1 [205, 206].

Podstawowa funkcja C/EBP α jest związana z regulacją procesów kierujących dojrzewaniem tkanek. Wyniki badań na myszach pozbawionych ekspresji C/EBP α (ang. *Cebp null mice*) wskazują na kluczową rolę tego czynnika transkrypcyjnego w różnicowaniu adipocytów, granulocytów i komórek nabłonkowych [207-209]. Mechanizmy stojące za zdolnością C/EBP α do wpływania na dojrzewanie różnych tkanek pozostają nie do końca jasne. Wydaje się, że C/EBP α współdziała z innymi czynnikami transkrypcyjnymi modulując ekspresję docelowych genów w zależności od kolaborującego z nim czynnika transkrypcyjnego. Dla przykładu, w adipocytach C/EBP α wpływa na różnicowanie i zahamowanie wzrostu poprzez współdziałanie z PPAR γ , podczas gdy interakcje NKX2-1 i HNF3 β z C/EBP α zawiadują dojrzewaniem komórek nabłonkowych dróg oddechowych [210, 211]. Drugą funkcją C/EBP α jest zdolność hamowania wzrostu dojrzałych komórek. Kierowanie różnicowaniem komórek z następczym hamowaniem ich wzrostu po osiągnięciu przez nich pełnej dojrzałości powoduje, że C/EBP α jest uważane za „przełącznik dojrzewania” kontrolujący transformację proliferujących i nie w pełni zróżnicowanych tkanek w formy dojrzałe i nieproliferujące [212]. Regulacja wzrostu odbywa się poprzez bezpośredni represyjny wpływ C/EBP α na kompleks E2F, który kontroluje przejście komórek z fazy G1 do S w cyklu komórkowym poprzez hamowanie *CDK2/CDK4*, aktywację

p21^{waf1/cip1} i remodeling kompleksu chromatyny [213-216]. Z uwagi na funkcje związane z regulowaniem procesów dojrzewania i proliferacji C/EBP α jest istotnym czynnikiem transkrypcyjnym biorącym udział w patogenezie szeregu nowotworów. Rola jaką odgrywa C/EBP α w różnicowaniu granulocytów i monocytów sprawia, że jest jednym z najważniejszych czynników biorących udział w transformacji nowotworowej tych linii komórkowych. Mutacje somatyczne w genie *CEBPA* są jednymi z najczęstszych zmian genetycznych stwierdzanych w ostrej białaczce szpikowej (ang. *acute myeloid leukemia*, AML) [217]. Dodatkowo opisano metylację promotora, supresję translacji i posttranslacyjne modyfikacje wpływające na ekspresję C/EBP α u chorych z AML [218]. Oprócz białaczek C/EBP α odgrywają także istotną rolę w onkogenezie guzów litych. Wykazano, że wysoka ekspresja C/EBP α w dojrzałych hepatocytach ulega prawie całkowitemu wyhamowaniu w tkance wątrobowej ulegającej regeneracji. Podobną supresję odnotowano w komórkach raka wątrobowokomórkowego [219]. Przywrócenie ekspresji C/EBP α w linii komórkowej tego raka hamowało proliferację wskazując na supresorowe działanie C/EBP α w tym guzie [220]. Podobny wpływ C/EBP α zaobserwowano w raku płuc i skóry [221, 222].

C/EBP β zwany także jądrowym czynnikiem dla ekspresji interleukiny 6 (ang. *Nuclear Factor for Interleukin-6 Expression*, NF-IL6) jest jednym z ważniejszych mediatorów wrodzonej odpowiedzi immunologicznej wpływającym na ekspresję szeregu cytokin takich jak TNF, IL-1 β , IL-6, IL-12, p35 i IL-10 [223]. C/EBP β jest także istotnym czynnikiem regulującym procesy proliferacji i różnicowania monocytów [224]. Deficyt C/EBP β w komórkach niskozróżnicowanych szpiku będących we wczesnych fazach dojrzewania (przed podziałem na linię monocytową i granulocytową) hamuje ich proliferację i formowanie kolonii, podczas gdy w komórkach wykazujących większy stopień zróżnicowania supresja C/EBP β stymuluje podziały komórkowe [225, 226]. W porównaniu z C/EBP α rola C/EBP β w patogenezie nowotworów jest mniej poznana. Wykazano, że ekspresja izoformy LIP jest podwyższona w białaczkach mielomonocytowych oraz w raku piersi [227, 228]. Z kolei stymulacja transkrypcji izoform LAP*/LAP hamuje proliferację i inicjuje dojrzewanie leukocytów [229]. C/EBP β zwiększa przeżycie komórek poprzez redukcję ekspresji i aktywności p53 w odpowiedzi na uszkodzenie DNA [230]. Z drugiej strony indukcja C/EBP β w komórkach raka wątrobowokomórkowego hamuje ich wzrost poprzez zatrzymanie cyklu komórkowego na granicy faz G1/S [231]. Wydaje się więc, że w zależności od rodzaju tkanki lub stopnia jej dojrzałości C/EBP β indukuje lub hamuje procesy

pro-proliferacyjne. Dlatego niektórzy autorzy stawiają tezę o roli C/EBP β jako ważnego czynnika transkrypcyjnego kontrolującego proces przejścia niezłośliwej proliferacji komórkowej (na przykład proliferacji leukocytów w odpowiedzi na czynnik immunologiczny) w niekontrolowany rozrost nowotworowy [223].

C/EBP α i C/EBP β ulegają transkrypcji w wielu różnych komórkach takich jak hepatocyty, adipocyty, komórki szpiku kostnego i komórki nabłonkowe dróg oddechowych [232, 233]. Według danych zawartych w bazie Cancer Genome Anatomy Project czynniki te są także transkrybowane w zdrowej tkance tarczycowej i hamowane w tkance NMTC (<http://cgap.nci.nih.gov/cgap.html>). Baza Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed>; dane z sierpnia 2012) zawiera jedynie dwie publikacje (Pomerance i wsp. oraz Akagi i wsp.) poświęcone związkowi C/EBP z TC [234, 235]. W badaniu Pomerance i wsp. analizowano regulację ekspresji i lokalizację C/EBP β w tyreocytach. Wykazano, że TSH stymulował transkrypcję C/EBP β w komórkach tarczycowych, który z kolei hamował ekspresję symportera sodowo-jodowego (NIS) stymulowaną przez cykliczny monofosforan adenozyliny (cAMP). Analiza immunohistochemiczna wykazała, że ekspresja C/EBP β w tyreocytach była ograniczona ściśle do jądra komórkowego, podczas gdy w komórkach PTC lokalizowała się głównie w cytoplazmie [234]. Ta ostatnia obserwacja została potwierdzona w pracy Akagi i wsp. analizującej rolę HNF3 β /FoxA2, NKX2-1, C/EBP α i C/EBP β w TC. Dodatkowo, w badaniu tym przy użyciu ilościowego PCR wykazano, że transkrypcja C/EBP α i C/EBP β ulega supresji w komórkach TC, a indukcja ekspresji C/EBP β w komórkach BHP hamuje ich wzrost [235].

Wyniki niniejszej rozprawy są kolejnymi, które wskazują na związek C/EBP α i C/EBP β z PTC. Test opóźnienia w żelu i podwójnej lucyferazy wskazały na wiązanie się i aktywację promotora *PTCSC3* przez C/EBP α i C/EBP β *in vitro*. Potwierdzeniem wpływu tych czynników na transkrypcję *PTCSC3* *in vivo* powinno być wykazanie zależności poziomu ekspresji tego genu od genotypu rs944289. Wyniki analizy transkrypcji *PTCSC3* u 46 chorych z PTC wykazały silniejszą supresję tego lincRNA w tkance guza u homozygot [TT] (n=21) w porównaniu z heterozygotami (n=19) (p=0.004). Porównując dostępne w laboratorium nieliczne homozygoty [CC] (n=6) z homozygotami [TT] (n=21) nie wykazano istotnych różnic w supresji *PTCSC3* (p=0.110). Liczebność pacjentów bez allelu ryzyka była jednakże dużym ograniczeniem tego porównania (Tabela 37, Rycina 17). Analiza ekspresji *PTCSC3* w zdrowej tkance tarczycowej wskazała, że gen ten jest silniej transkrybowany u

homozygot [TT] (n=21) w porównaniu z heterozygotami (n=19) (p=0.034) (Tabela 37, Rycina 16). O ile słabszą ekspresję *PTCSC3* w tkance guza u homozygot [TT] można wiązać z wpływem C/EBP α i/lub C/EBP β , o tyle zwiększoną transkrypcję tego lincRNA w zdrowej tkance tarczycowej nie można wytłumaczyć oddziaływaniem C/EBP. Wiadomo jednakże, że regulacja ekspresji zachodzi na poziomie zwykle kilku TF współdziałających w odpowiedniej dla danej tkanki konstelacji. Jest więc możliwe, że działanie innego czynnika transkrypcyjnego podlega zmianie poprzez niezidentyfikowany jeszcze wariant będący w LD z rs944289. Identyfikacja takich polimorfizmów wymagałaby pełnego sekwencjonowania *PTCSC3* (eksony+introny) wraz z regionem promotora w odpowiednio licznej populacji chorych i kontroli. W niniejszej pracy sekwencjonowano jedynie eksony *PTCSC3* w zdrowej tkance tarczycowej i tkance guza u 33 pacjentów z PTC. Wyniki tego przesiewowego badania nie wykazały obecności mutacji lub zmian MAF dla opisanych w bazie dbSNP wariantów.

Zaobserwowana nadekspresja *PTCSC3* w tkance niedotkniętej procesem nowotworowym oraz silniejsza supresja w tkance guza u chorych noszących allel ryzyka w porównaniu z heterozygotami może sugerować, że gen ten oprócz właściwości supresora posiada także potencjał onkogenny. Możliwe, że stymulacja ekspresji *PTCSC3* w zdrowej tkance tarczycowej wpływa na ekspresję genów ważnych w procesie nowotworowej transformacji, podczas gdy jego silniejsza supresja w tkance TC wpływa na inny zestaw genów istotnych dla wzrostu guza. Do tej pory scharakteryzowano co najmniej kilka genów wykazujących działanie zarówno anty- jak i pro-onkogenne, aczkolwiek udowodnienie takiej roli dla *PTCSC3* wymagałoby dalszych badań ukierunkowanych na funkcję tego genu w tyreocycie [236-238].

7. WNIOSKI

1. W bliskim sąsiedztwie (3.2 kb) polimorfizmu rs944289 zidentyfikowano nieznany wcześniej lincRNA (*PTCSC3*) charakteryzujący się wysoką, tarczycowo-specyficzną ekspresją. *PTCSC3* i rs944289 znajdują się w 43 kb bloku genomowego DNA wykazującym nierównowagę sprzężeń.
2. Silna supresja transkrypcji *PTCSC3* w tkance PTC i komórkach linii komórkowych PTC oraz hamowanie wzrostu tych ostatnich w odpowiedzi na przywrócenie ekspresji *PTCSC3* sugeruje supresorowe działanie tego genu w PTC.
3. Poziom ekspresji *PTCSC3* w tkance tarczycowej i tkance guza u chorych z PTC wykazywał zależność od genotypu rs944289. Transkrypcja *PTCSC3* w tkance guza była silniej hamowana u chorych z allelem ryzyka (genotyp [TT]) w porównaniu z heterozygotami. W tkance tarczycowej genotyp [TT] w porównaniu z genotypem [CT] był związany z nieznacznie silniejszą ekspresją *PTCSC3*.
4. SNP rs944289 znajduje się w miejscu wiążącym czynniki transkrypcyjne C/EBP α i C/EBP β . Allel ryzyka osłabia aktywację promotora *PTCSC3* wyzwalaną przez C/EBP α i C/EBP β , co może tłumaczyć mechanizm wpływu rs944289 na ekspresję *PTCSC3*.

8. STRESZCZENIE

Pomimo względnie silnej predyspozycji genetycznej do raka brodawkowatego tarczycy (PTC) do tej pory odkryto zaledwie kilka wariantów związanych z tym nowotworem. Wydaje się, że może to mieć związek z niską penetracją genów odpowiedzialnych za rozwój PTC. Tego typu transkrypty są z natury trudne do wykrycia w analizie sprzężeń ale można je wskazać w badaniach asocjacyjnych. Badanie asocjacyjne całego genomu dla DTC zidentyfikowało dwa warianty polimorficzne (rs944289 i rs965513) zlokalizowane w regionach nie zawierających genów (odpowiednio 14q13.3 i 9q22.33). Celem tej pracy było wyjaśnienie mechanizmu, poprzez który polimorfizm rs94489 predysponuje do PTC.

Aby zidentyfikować potencjalne transkrypty zlokalizowane w locus 14q13.3 przeprowadzono analizę danych zawartych w bazie znaczników ekspresji (dbEST). W bliskim sąsiedztwie rs944289 opisano kilka EST, z których tylko jeden (CB987890) został wyizolowany z mieszaniny RNA z różnych narządów, w tym tarczycy. Analiza tkankowej transkrypcji metodą RT-PCR wykazała, że CB987890 charakteryzuje się silną, tarczycowo-specyficzną ekspresją. W celu określenia pełnej struktury CB987890 zastosowano technikę RLM-RACE i northern blot. Wyniki tych eksperymentów wykazały, że CB987890 składa się z 4 eksonów o całkowitej długości 1154 nukleotydów. Ten nowy gen nazwany *Papillary Thyroid Carcinoma Susceptibility Candidate 3 (PTCSC3)* znajduje się razem z rs944289 w 43 kb bloku genomowego DNA wykazującym nierównowagę sprzężeń. SNP rs944289 jest zlokalizowany w regionie promotora dla *PTCSC3* (3.2 kb od eksonu 1). *PTCSC3* posiada 95 aminokwasową ramkę odczytu oraz sekwencję sygnałową dla poliadenylacji znajdującą się 18 nukleotydów przed łańcuchem poliA. Struktura eksonów 2, 3 i 4 pokrywa się z mobilnymi elementami LINE, SINE i LTR wskazując na prawdopodobnie niekodujący charakter *PTCSC3* (duży międzygenowy niekodujący RNA).

Ekspresja *PTCSC3* w tkance PTC w porównaniu z tkanką tarczycową nieobjętą procesem nowotworowym była silnie hamowana u wszystkich chorych włączonych do badania ($n=46$; $p=2.84 \times 10^{-14}$). Aby ocenić wpływ allelu ryzyka na supresję *PTCSC3* analizowano współczynnik ekspresji w guzie do ekspresji w zdrowej tkance tarczycowej. Allel rs944289[T] był związany z silniejszym hamowaniem transkrypcji *PTCSC3* u homozygot [TT] ($n=21$) w porównaniu z heterozygotami [CT] ($n=19$) ($p=0.004$). W zdrowej

tkance tarczycowej ekspresja *PTCSC3* u homozygot [TT] (n=21) w porównaniu z heterozygotami [CT] (n=19) była nieznacznie wyższa (p=0.034). Mała liczebność homozygot [CC] (n=6) ograniczała wartość analizy statystycznej dla tej grupy chorych.

Analiza bioinformatyczna wskazała, że rs944289 jest zlokalizowany wewnątrz konserwatywnej części miejsca wiążącego czynniki transkrypcyjne C/EBP α i C/EBP β . Allel ryzyka zmniejszał *in silico* wielkość wpływu pozycji w macierzy (PWM). Obecność kompleksów C/EBP α i C/EBP β z DNA zawierającym rs944289[C] została potwierdzona *in vitro* w teście opóźnienia w żelu. Aby wykazać efekt wywierany przez C/EBP α i C/EBP β na aktywność promotora *PTCSC3* zastosowano test podwójnej lucyferazy w komórkach COS-7 (nie wykazujących endogennej ekspresji C/EBP α i C/EBP β). Zarówno C/EBP α (izofорма p42) jak i C/EBP β (izofорма LAP*) stymulowały promotor (odpowiednio p=0.0009 i p=0.0014). Allel ryzyka [T] w porównaniu z allelem [C] redukował aktywność generowaną przez C/EBP α p42 i C/EBP β (odpowiednio p=0.026 i p=0.048).

Analiza ekspresji *PTCSC3* wykonana przy użyciu RT-PCR w 18 liniach komórkowych TC wskazała na utratę transkrypcji tego genu w 14 liniach i bardzo silną supresję w pozostałych 4. Przywrócenie transkrypcji *PTCSC3* w liniach TPC-1 i BCPAP hamowało wzrost komórek (odpowiednio p=0.002 i p=0.019) oraz wpływało na ekspresję genów mających istotne znaczenie w patogenezie PTC. Analiza przeprowadzona przy użyciu oprogramowania Ingenuity Pathway Analysis wskazała, że geny, których ekspresja uległa zmianie w odpowiedzi na indukcję transkrypcji *PTCSC3* są odpowiedzialne za replikację, rekombinację i naprawę DNA, mobilność komórek, morfologię guza i śmierć komórki.

Reasumując, dzięki danym uzyskanym w niniejszym badaniu zidentyfikowano tarczycowo-specyficzny duży niekodujący międzygenowy RNA gen (*PTCSC3*) znajdujący się razem z predysponującym do DTC polimorfizmem rs944289 w 43 kb bloku wykazującym nierównowagę sprzężeń. Allel ryzyka zmniejszał aktywność promotora dla *PTCSC3* generowaną przez C/EBP α p42 i C/EBP β . Ekspresja *PTCSC3* była silnie hamowana w tkance guza w porównaniu ze zdrową tkanką tarczycową, a allel ryzyka był związany z najsilniejszą supresją *PTCSC3*. Przywrócenie ekspresji *PTCSC3* w liniach komórkowych PTC wpływało na ekspresję genów związanych z onkogenezą TC oraz hamowało wzrost komórek.

9. SUMMARY

Despite a strong genetic component in the etiology of papillary thyroid carcinoma (PTC) only few predisposing variants have been uncovered so far. This might be due to low penetrance of the genes involved in PTC pathogenesis. Such genes are typically difficult to detect by linkage analysis but can be revealed by association studies. A genome wide association study (GWAS) of PTC pinpointed two independent SNPs predisposing to PTC (rs944289 and rs965513) located in regions containing no annotated genes (14q13.3 and 9q22.33, respectively). The aim of this study was to elucidate the mechanism by which polymorphism rs944289 predisposes to PTC.

To identify potential candidate genes in the 14q13.3 region *in silico* analysis of the Expressed Sequence Tags database (dbEST) was performed. Several expressed sequence tags were described in the close vicinity of rs944289. Only one (CB987890) was derived from RNA from human organs including thyroid. The analysis of its transcription in multiple human tissues indicated that CB987890 displays strong thyroid-specific expression. In order to establish its full structure RNA Ligase Mediated Rapid Amplification of cDNA Ends technique and northern blot analysis were applied. The results showed that CB987890 consists of 4 exons with a total length of 1154 bp. This novel gene that we named *Papillary Thyroid Carcinoma Susceptibility Candidate 3 (PTCSC3)* is located in a 43 kb linkage disequilibrium block also containing rs944289. The SNP rs944289 is located in the promoter region of *PTCSC3* (3.2 kb upstream of exon 1). *PTCSC3* has a potential 95 amino acid open reading frame and a polyadenylation signal sequence located 18 bp before a polyA tail. The sequence of exons 2, 3 and 4 overlaps with LINE, SINE and LTR mobile elements suggesting that *PTCSC3* is probably a large non-coding intergenic RNA gene (lincRNA).

PTCSC3 transcription was strongly downregulated in tumor tissue compared with unaffected thyroid in all PTC patients included in the study (n=46; $p=2.84 \times 10^{-14}$). To assess the impact of the risk allele on suppression of *PTCSC3* fold changes in expression between tumor and unaffected thyroid tissue were analyzed. The rs944289[T] allele was associated with stronger downregulation of *PTCSC3* when [TT] homozygotes (n=21) were compared with heterozygous [CT] patients (n=19) ($p=0.004$), but in the comparison with the few available homozygous [CC] subjects (n=6) the difference did not reach statistical significance ($p=0.110$). The small sample size was a major limitation of this comparison. In adjacent

unaffected thyroid tissue *PTCSC3* expression was slightly higher in [TT] homozygotes compared with [CT] heterozygotes ($p=0.034$).

Computational analysis showed that rs944289 is located within the conserved part of the binding site for the CCAAT/enhancer binding proteins (C/EBP) α and β . The risk allele reduces a position weight matrix score (PWM) *in silico*. The presence of complexes of C/EBP α and C/EBP β proteins with DNA containing rs944289[C] was confirmed *in vitro* by an electrophoretic mobility shift assay. To investigate the effect of C/EBP α and C/EBP β on the promoter activity, a dual luciferase assay in COS-7 cells (with no endogenous expression of C/EBP α and C/EBP β) was employed. Both C/EBP α (isoform p42) and C/EBP β (isoform LAP*) activated the *PTCSC3* promoter in reporter assays ($p=0.0009$ and $p=0.0014$, respectively) and the risk allele [T] reduced the activation compared to the non-risk allele [C] ($p=0.026$ and $p=0.048$, respectively).

The analysis of *PTCSC3* expression performed by RT-PCR in 18 thyroid cancer cell lines showed loss of expression in 14 cell lines and very strong suppression in the remaining 4. Restoration of *PTCSC3* transcription in TPC-1 and BCPAP cell lines inhibited cell growth ($p=0.002$ and $p=0.019$, respectively) and affected the expression of genes of major interest to PTC. The Ingenuity Pathway Analysis showed that the transcripts dysregulated by *PTCSC3* are involved in DNA replication, recombination and repair, cellular movement, tumor morphology and cell death.

In conclusion, the data obtained in this study identified a thyroid-specific lincRNA gene (*PTCSC3*) located in a linkage disequilibrium block with the polymorphism rs944289 that predisposes to PTC. The risk allele decreased *PTCSC3* promoter activation by weakening the binding affinity of the p42 C/EBP α and C/EBP β transcription factors. *PTCSC3* expression was significantly downregulated in tumor compared with unaffected thyroid tissue, and the risk allele was associated with the strongest suppression of *PTCSC3*. The restoration of *PTCSC3* expression in PTC cell lines affected genes involved in thyroid tumorigenesis and inhibited cell growth.

10. PIŚMIENNICTWO

1. Ferlay, J., et al., *Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008*. Int J Cancer, 2010. **127**(12): p. 2893-917.
2. Jemal, A., et al., *Global cancer statistics*. CA Cancer J Clin, 2011. **61**(2): p. 69-90.
3. Kohler, B.A., et al., *Annual report to the nation on the status of cancer, 1975-2007, featuring tumors of the brain and other nervous system*. J Natl Cancer Inst, 2011. **103**(9): p. 714-36.
4. Liu, S., et al., *Increasing thyroid cancer incidence in Canada, 1970-1996: time trends and age-period-cohort effects*. Br J Cancer, 2001. **85**(9): p. 1335-9.
5. ACS, American Cancer Society. *Cancer Facts & Figures 2012*. Atlanta: American Cancer Society; 2012. <http://www.cancer.org/acs/>. 2012.
6. Ehemann, C., et al., *Annual Report to the Nation on the status of cancer, 1975-2008, featuring cancers associated with excess weight and lack of sufficient physical activity*. Cancer, 2012. **118**(9): p. 2338-66.
7. Szczeklik, A., *Choroby wewnętrzne - stan wiedzy na rok 2011/2011*, Kraków: Medycyna Praktyczna.
8. Davies, L. and H.G. Welch, *Increasing incidence of thyroid cancer in the United States, 1973-2002*. Jama, 2006. **295**(18): p. 2164-7.
9. Bondeson, L. and O. Ljungberg, *Occult thyroid carcinoma at autopsy in Malmo, Sweden*. Cancer, 1981. **47**(2): p. 319-23.
10. Heitz, P., H. Moser, and J.J. Staub, *Thyroid cancer: a study of 573 thyroid tumors and 161 autopsy cases observed over a thirty-year period*. Cancer, 1976. **37**(5): p. 2329-37.
11. Sobrinho-Simoes, M.A., M.C. Sambade, and V. Goncalves, *Latent thyroid carcinoma at autopsy: a study from Oporto, Portugal*. Cancer, 1979. **43**(5): p. 1702-6.
12. Harach, H.R., K.O. Franssila, and V.M. Wasenius, *Occult papillary carcinoma of the thyroid. A "normal" finding in Finland. A systematic autopsy study*. Cancer, 1985. **56**(3): p. 531-8.
13. Chen, A.Y., A. Jemal, and E.M. Ward, *Increasing incidence of differentiated thyroid cancer in the United States, 1988-2005*. Cancer, 2009. **115**(16): p. 3801-7.
14. Mack, W.J., et al., *Lifestyle and other risk factors for thyroid cancer in Los Angeles County females*. Ann Epidemiol, 2002. **12**(6): p. 395-401.
15. Markaki, I., D. Linos, and A. Linos, *The influence of dietary patterns on the development of thyroid cancer*. Eur J Cancer, 2003. **39**(13): p. 1912-9.
16. Dijkstra, B., et al., *Changing patterns of thyroid carcinoma*. Ir J Med Sci, 2007. **176**(2): p. 87-90.
17. Kondo, T., S. Ezzat, and S.L. Asa, *Pathogenetic mechanisms in thyroid follicular-cell neoplasia*. Nat Rev Cancer, 2006. **6**(4): p. 292-306.
18. Grebe, S.K. and I.D. Hay, *Follicular thyroid cancer*. Endocrinol Metab Clin North Am, 1995. **24**(4): p. 761-801.
19. Kruś, S., *Patomorfologia kliniczna 2007*: Wydawnictwo Lekarskie PZWL.
20. Shattuck, T.M., et al., *Independent clonal origins of distinct tumor foci in multifocal papillary thyroid carcinoma*. N Engl J Med, 2005. **352**(23): p. 2406-12.
21. McCarthy, R.P., et al., *Molecular evidence for the same clonal origin of multifocal papillary thyroid carcinomas*. Clin Cancer Res, 2006. **12**(8): p. 2414-8.
22. Stevens, A., *Patologia 2004*, Lublin: Wydawnictwo CZELEJ.
23. Randolph, G.W. and D. Maniar, *Medullary carcinoma of the thyroid*. Cancer Control, 2000. **7**(3): p. 253-61.
24. Saad, M.F., et al., *Medullary carcinoma of the thyroid. A study of the clinical features and prognostic factors in 161 patients*. Medicine (Baltimore), 1984. **63**(6): p. 319-42.
25. Schneider, A.B. and D.H. Sarne, *Long-term risks for thyroid cancer and other neoplasms after exposure to radiation*. Nat Clin Pract Endocrinol Metab, 2005. **1**(2): p. 82-91.

26. Boice, J.D., Jr., *Thyroid disease 60 years after Hiroshima and 20 years after Chernobyl*. *Jama*, 2006. **295**(9): p. 1060-2.
27. Pacini, F., et al., *Post-Chernobyl thyroid carcinoma in Belarus children and adolescents: comparison with naturally occurring thyroid carcinoma in Italy and France*. *J Clin Endocrinol Metab*, 1997. **82**(11): p. 3563-9.
28. Cardis, E., et al., *Risk of thyroid cancer after exposure to 131I in childhood*. *J Natl Cancer Inst*, 2005. **97**(10): p. 724-32.
29. Jacob, P., et al., *Thyroid cancer among Ukrainians and Belarusians who were children or adolescents at the time of the Chernobyl accident*. *J Radiol Prot*, 2006. **26**(1): p. 51-67.
30. Likhtarev, I., et al., *Questionnaire- and measurement-based individual thyroid doses in Ukraine resulting from the Chornobyl nuclear reactor accident*. *Radiat Res*, 2006. **166**(1 Pt 2): p. 271-86.
31. Feldt-Rasmussen, U., *Iodine and cancer*. *Thyroid*, 2001. **11**(5): p. 483-6.
32. Kalk, W.J., F. Sitas, and A.C. Patterson, *Thyroid cancer in South Africa--an indicator of regional iodine deficiency*. *S Afr Med J*, 1997. **87**(6): p. 735-8.
33. Woodruff, S.L., et al., *Global variation in the pattern of differentiated thyroid cancer*. *Am J Surg*, 2010. **200**(4): p. 462-6.
34. Harach, H.R., et al., *Thyroid carcinoma and thyroiditis in an endemic goitre region before and after iodine prophylaxis*. *Acta Endocrinol (Copenh)*, 1985. **108**(1): p. 55-60.
35. Harach, H.R. and E.D. Williams, *Thyroid cancer and thyroiditis in the goitrous region of Salta, Argentina, before and after iodine prophylaxis*. *Clin Endocrinol (Oxf)*, 1995. **43**(6): p. 701-6.
36. Burgess, J.R., et al., *The changing incidence and spectrum of thyroid carcinoma in Tasmania (1978-1998) during a transition from iodine sufficiency to iodine deficiency*. *J Clin Endocrinol Metab*, 2000. **85**(4): p. 1513-7.
37. Pettersson, B., et al., *Trends in thyroid cancer incidence in Sweden, 1958-1981, by histopathologic type*. *Int J Cancer*, 1991. **48**(1): p. 28-33.
38. Goldgar, D.E., et al., *Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands*. *J Natl Cancer Inst*, 1994. **86**(21): p. 1600-8.
39. Dong, C. and K. Hemminki, *Modification of cancer risks in offspring by sibling and parental cancers from 2,112,616 nuclear families*. *Int J Cancer*, 2001. **92**(1): p. 144-50.
40. Risch, N., *The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches*. *Cancer Epidemiol Biomarkers Prev*, 2001. **10**(7): p. 733-41.
41. Czene, K., P. Lichtenstein, and K. Hemminki, *Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database*. *Int J Cancer*, 2002. **99**(2): p. 260-6.
42. Hemminki, K. and C. Dong, *Familial relationships in thyroid cancer by histo-pathological type*. *Int J Cancer*, 2000. **85**(2): p. 201-5.
43. Frich, L., E. Glatte, and L.A. Akslen, *Familial occurrence of nonmedullary thyroid cancer: a population-based study of 5673 first-degree relatives of thyroid cancer patients from Norway*. *Cancer Epidemiol Biomarkers Prev*, 2001. **10**(2): p. 113-7.
44. Hrafinkelsson, J., et al., *Familial non-medullary thyroid cancer in Iceland*. *J Med Genet*, 2001. **38**(3): p. 189-91.
45. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland*. *N Engl J Med*, 2000. **343**(2): p. 78-85.
46. Węgleński, P., *Genetyka molekularna* 2006, Warszawa: PWN.
47. Iyengar, S.K. and R.C. Elston, *The genetic basis of complex traits: rare variants or "common gene, common disease"?* *Methods Mol Biol*, 2007. **376**: p. 71-84.
48. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes*. *Science*, 2007. **316**(5829): p. 1336-41.

49. Matarin, M., et al., *A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release*. *Lancet Neurol*, 2007. **6**(5): p. 414-20.
50. Samani, N.J., et al., *Genomewide association analysis of coronary artery disease*. *N Engl J Med*, 2007. **357**(5): p. 443-53.
51. Hafler, D.A., et al., *Risk alleles for multiple sclerosis identified by a genomewide study*. *N Engl J Med*, 2007. **357**(9): p. 851-62.
52. Easton, D.F., et al., *Genome-wide association study identifies novel breast cancer susceptibility loci*. *Nature*, 2007. **447**(7148): p. 1087-93.
53. Plenge, R.M., et al., *TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study*. *N Engl J Med*, 2007. **357**(12): p. 1199-209.
54. Coon, K.D., et al., *A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease*. *J Clin Psychiatry*, 2007. **68**(4): p. 613-8.
55. Pritchard, J.K., *Are rare variants responsible for susceptibility to complex diseases?* *Am J Hum Genet*, 2001. **69**(1): p. 124-37.
56. Gibson, G., *Rare and common variants: twenty arguments*. *Nat Rev Genet*, 2011. **13**(2): p. 135-45.
57. Magi, R., et al., *Genome-Wide Association Analysis of Imputed Rare Variants: Application to Seven Common Complex Diseases*. *Genet Epidemiol*, 2012.
58. Uchino, S., et al., *Familial nonmedullary thyroid carcinoma characterized by multifocality and a high recurrence rate in a large study population*. *World J Surg*, 2002. **26**(8): p. 897-902.
59. Ito, Y., et al., *Biological behavior and prognosis of familial papillary thyroid carcinoma*. *Surgery*, 2009. **145**(1): p. 100-5.
60. de la Chapelle, A. and K. Jazdzewski, *MicroRNAs in thyroid cancer*. *J Clin Endocrinol Metab*, 2011. **96**(11): p. 3326-36.
61. Malchoff, C.D., et al., *Papillary thyroid carcinoma associated with papillary renal neoplasia: genetic linkage analysis of a distinct heritable tumor syndrome*. *J Clin Endocrinol Metab*, 2000. **85**(5): p. 1758-64.
62. McKay, J.D., et al., *Localization of a susceptibility gene for familial nonmedullary thyroid carcinoma to chromosome 2q21*. *Am J Hum Genet*, 2001. **69**(2): p. 440-6.
63. Suh, I., et al., *Distinct loci on chromosome 1q21 and 6q22 predispose to familial nonmedullary thyroid cancer: a SNP array-based linkage analysis of 38 families*. *Surgery*, 2009. **146**(6): p. 1073-80.
64. Cavaco, B.M., et al., *Mapping a new familial thyroid epithelial neoplasia susceptibility locus to chromosome 8p23.1-p22 by high-density single-nucleotide polymorphism genome-wide linkage analysis*. *J Clin Endocrinol Metab*, 2008. **93**(11): p. 4426-30.
65. He, H., et al., *A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24*. *Cancer Res*, 2009. **69**(2): p. 625-31.
66. Bignell, G.R., et al., *Familial nontoxic multinodular thyroid goiter locus maps to chromosome 14q but does not account for familial nonmedullary thyroid cancer*. *Am J Hum Genet*, 1997. **61**(5): p. 1123-30.
67. Canzian, F., et al., *A gene predisposing to familial thyroid tumors with cell oxyphilia maps to chromosome 19p13.2*. *Am J Hum Genet*, 1998. **63**(6): p. 1743-8.
68. Gudmundsson, J., et al., *Discovery of common variants associated with low TSH levels and thyroid cancer risk*. *Nat Genet*, 2012. **44**(3): p. 319-22.
69. Jazdzewski, K., et al., *Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma*. *Proc Natl Acad Sci U S A*, 2008. **105**(20): p. 7269-74.
70. Jones, A.M., et al., *Thyroid cancer susceptibility polymorphisms: confirmation of loci on chromosomes 9q22 and 14q13, validation of a recessive 8q24 locus and failure to replicate a locus on 5q24*. *J Med Genet*, 2012. **49**(3): p. 158-63.

71. Wokolorczyk, D., et al., *A range of cancers is associated with the rs6983267 marker on chromosome 8*. *Cancer Res*, 2008. **68**(23): p. 9982-6.
72. Neta, G., et al., *Common genetic variants in the 8q24 region and risk of papillary thyroid cancer*. *Laryngoscope*, 2012. **122**(5): p. 1040-2.
73. Akdi, A., et al., *Common variants of the thyroglobulin gene are associated with differentiated thyroid cancer risk*. *Thyroid*, 2011. **21**(5): p. 519-25.
74. Matakidou, A., et al., *Risk of non-medullary thyroid cancer influenced by polymorphic variation in the thyroglobulin gene*. *Carcinogenesis*, 2004. **25**(3): p. 369-73.
75. Gudmundsson, J., et al., *Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations*. *Nat Genet*, 2009. **41**(4): p. 460-4.
76. Takahashi, M., et al., *The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl*. *Hum Mol Genet*, 2010. **19**(12): p. 2516-23.
77. Matsuse, M., et al., *The FOXE1 and NKX2-1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population*. *J Med Genet*, 2011. **48**(9): p. 645-8.
78. Landa, I., et al., *The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors*. *PLoS Genet*, 2009. **5**(9): p. e1000637.
79. Iuliano, R., et al., *Role of PTPRJ genotype in papillary thyroid carcinoma risk*. *Endocr Relat Cancer*, 2010. **17**(4): p. 1001-6.
80. Akulevich, N.M., et al., *Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma*. *Endocr Relat Cancer*, 2009. **16**(2): p. 491-503.
81. Ho, T., et al., *Association of XRCC1 polymorphisms and risk of differentiated thyroid carcinoma: a case-control analysis*. *Thyroid*, 2009. **19**(2): p. 129-35.
82. Jazdzewski, K., et al., *Polymorphic mature microRNAs from passenger strand of pre-miR-146a contribute to thyroid cancer*. *Proc Natl Acad Sci U S A*, 2009. **106**(5): p. 1502-5.
83. Zannini, M., et al., *TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation*. *EMBO J*, 1997. **16**(11): p. 3185-97.
84. Cuesta, I., K.S. Zaret, and P. Santisteban, *The forkhead factor FoxE1 binds to the thyroperoxidase promoter during thyroid cell differentiation and modifies compacted chromatin structure*. *Mol Cell Biol*, 2007. **27**(20): p. 7302-14.
85. De Felice, M. and R. Di Lauro, *Thyroid development and its disorders: genetics and molecular mechanisms*. *Endocr Rev*, 2004. **25**(5): p. 722-46.
86. Parlato, R., et al., *An integrated regulatory network controlling survival and migration in thyroid organogenesis*. *Dev Biol*, 2004. **276**(2): p. 464-75.
87. De Felice, M., et al., *A mouse model for hereditary thyroid dysgenesis and cleft palate*. *Nat Genet*, 1998. **19**(4): p. 395-8.
88. Clifton-Bligh, R.J., et al., *Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia*. *Nat Genet*, 1998. **19**(4): p. 399-401.
89. Tomlinson, I., et al., *A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21*. *Nat Genet*, 2007. **39**(8): p. 984-8.
90. Yeager, M., et al., *Genome-wide association study of prostate cancer identifies a second risk locus at 8q24*. *Nat Genet*, 2007. **39**(5): p. 645-9.
91. Haiman, C.A., et al., *A common genetic risk factor for colorectal and prostate cancer*. *Nat Genet*, 2007. **39**(8): p. 954-6.
92. Ambroziak, M. and A. Nauman, *[Thyroid transcription factors involved in the thyroid hormones biosynthesis]*. *Postepy Biochem*, 2005. **51**(3): p. 288-96.
93. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction*. *Anal Biochem*, 1987. **162**(1): p. 156-9.
94. Fujita, P.A., et al., *The UCSC Genome Browser database: update 2011*. *Nucleic Acids Res*, 2011. **39**(Database issue): p. D876-82.

95. Kel, A.E., et al., *MATCH: A tool for searching transcription factor binding sites in DNA sequences*. Nucleic Acids Res, 2003. **31**(13): p. 3576-9.
96. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method*. Methods, 2001. **25**(4): p. 402-8.
97. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase*. J Mol Biol, 1975. **94**(3): p. 441-8.
98. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
99. Nakayama, G.R., et al., *Assessment of the Alamar Blue assay for cellular growth and viability in vitro*. J Immunol Methods, 1997. **204**(2): p. 205-8.
100. Gloeckner, H., T. Jonuleit, and H.D. Lemke, *Monitoring of cell viability and cell growth in a hollow-fiber bioreactor by use of the dye Alamar Blue*. J Immunol Methods, 2001. **252**(1-2): p. 131-8.
101. Wright, G.W. and R.M. Simon, *A random variance model for detection of differential gene expression in small microarray experiments*. Bioinformatics, 2003. **19**(18): p. 2448-55.
102. Smyth, G.K., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments*. Stat Appl Genet Mol Biol, 2004. **3**: p. Article3.
103. Visscher, P.M. and G.W. Montgomery, *Genome-wide association studies and human disease: from trickle to flood*. JAMA, 2009. **302**(18): p. 2028-9.
104. Thomson, G. and M.S. Esposito, *The genetics of complex diseases*. Trends Cell Biol, 1999. **9**(12): p. M17-20.
105. Willard, H.F. and K.E. Davies, *Genetics of disease Complex genetics, complex diseases*. Curr Opin Genet Dev, 1998. **8**(3): p. 271-3.
106. Klein, R.J., et al., *Complement factor H polymorphism in age-related macular degeneration*. Science, 2005. **308**(5720): p. 385-9.
107. Tang, W., et al., *Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease*. Am J Hum Genet, 2012. **91**(1): p. 152-62.
108. Witte, J.S., *Genome-wide association studies and beyond*. Annu Rev Public Health, 2010. **31**: p. 9-20 4 p following 20.
109. Kraft, P., *Curses--winner's and otherwise--in genetic epidemiology*. Epidemiology, 2008. **19**(5): p. 649-51; discussion 657-8.
110. Lander, E.S., *The new genomics: global views of biology*. Science, 1996. **274**(5287): p. 536-9.
111. Thakkinian, A., et al., *Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration*. Hum Mol Genet, 2006. **15**(18): p. 2784-90.
112. Despriet, D.D., et al., *Complement factor H polymorphism, complement activators, and risk of age-related macular degeneration*. JAMA, 2006. **296**(3): p. 301-9.
113. Shaw, P.X., et al., *Complement factor H genotypes impact risk of age-related macular degeneration by interaction with oxidized phospholipids*. Proc Natl Acad Sci U S A, 2012. **109**(34): p. 13757-62.
114. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability*. Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
115. Lander, E.S., *Initial impact of the sequencing of the human genome*. Nature, 2011. **470**(7333): p. 187-97.
116. Wain, L.V., J.A. Armour, and M.D. Tobin, *Genomic copy number variation, human health, and disease*. Lancet, 2009. **374**(9686): p. 340-50.
117. Maher, B., *Personal genomes: The case of the missing heritability*. Nature, 2008. **456**(7218): p. 18-21.
118. Cirulli, E.T. and D.B. Goldstein, *Uncovering the roles of rare variants in common disease through whole-genome sequencing*. Nat Rev Genet, 2010. **11**(6): p. 415-25.

119. Freedman, M.L., et al., *Principles for the post-GWAS functional characterization of cancer risk loci*. Nat Genet, 2011. **43**(6): p. 513-8.
120. Donnelly, P., *Progress and challenges in genome-wide association studies in humans*. Nature, 2008. **456**(7223): p. 728-31.
121. Wicker, T., et al., *A unified classification system for eukaryotic transposable elements*. Nat Rev Genet, 2007. **8**(12): p. 973-82.
122. Martin, S.L., *Nucleic acid chaperone properties of ORF1p from the non-LTR retrotransposon, LINE-1*. RNA Biol, 2010. **7**(6): p. 706-11.
123. Mathias, S.L., et al., *Reverse transcriptase encoded by a human transposable element*. Science, 1991. **254**(5039): p. 1808-10.
124. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. Nature, 2001. **409**(6822): p. 860-921.
125. Houck, C.M., F.P. Rinehart, and C.W. Schmid, *A ubiquitous family of repeated DNA sequences in the human genome*. J Mol Biol, 1979. **132**(3): p. 289-306.
126. Ullu, E. and C. Tschudi, *Alu sequences are processed 7SL RNA genes*. Nature, 1984. **312**(5990): p. 171-2.
127. Dewannieux, M., C. Esnault, and T. Heidmann, *LINE-mediated retrotransposition of marked Alu sequences*. Nat Genet, 2003. **35**(1): p. 41-8.
128. Cohen, M. and E. Larsson, *Human endogenous retroviruses*. Bioessays, 1988. **9**(6): p. 191-6.
129. Makalowski, W., et al., *Transposable elements and their identification*. Methods Mol Biol, 2012. **855**: p. 337-59.
130. Brosius, J., *The contribution of RNAs and retroposition to evolutionary novelties*. Genetica, 2003. **118**(2-3): p. 99-116.
131. Kazazian, H.H., Jr., *Mobile elements: drivers of genome evolution*. Science, 2004. **303**(5664): p. 1626-32.
132. Ono, R., et al., *Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality*. Nat Genet, 2006. **38**(1): p. 101-6.
133. Volff, J.N., *Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes*. Bioessays, 2006. **28**(9): p. 913-22.
134. Cordaux, R., et al., *Birth of a chimeric primate gene by capture of the transposase gene from a mobile element*. Proc Natl Acad Sci U S A, 2006. **103**(21): p. 8101-6.
135. Nekrutenko, A. and W.H. Li, *Transposable elements are found in a large number of human protein-coding genes*. Trends Genet, 2001. **17**(11): p. 619-21.
136. Marino-Ramirez, L., et al., *Transposable elements donate lineage-specific regulatory sequences to host genomes*. Cytogenet Genome Res, 2005. **110**(1-4): p. 333-41.
137. Medstrand, P., et al., *Impact of transposable elements on the evolution of mammalian gene regulation*. Cytogenet Genome Res, 2005. **110**(1-4): p. 342-52.
138. Hung, T. and H.Y. Chang, *Long noncoding RNA in genome regulation: prospects and mechanisms*. RNA Biol, 2010. **7**(5): p. 582-5.
139. Schuler, G.D., et al., *A gene map of the human genome*. Science, 1996. **274**(5287): p. 540-6.
140. Carninci, P., et al., *The transcriptional landscape of the mammalian genome*. Science, 2005. **309**(5740): p. 1559-63.
141. Bertone, P., et al., *Global identification of human transcribed sequences with genome tiling arrays*. Science, 2004. **306**(5705): p. 2242-6.
142. Prensner, J.R. and A.M. Chinnaiyan, *The emergence of lncRNAs in cancer biology*. Cancer Discov, 2011. **1**(5): p. 391-407.
143. Visone, R. and C.M. Croce, *MiRNAs and cancer*. Am J Pathol, 2009. **174**(4): p. 1131-8.
144. Guttman, M., et al., *Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals*. Nature, 2009. **458**(7235): p. 223-7.
145. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome*. Nature, 1991. **349**(6304): p. 38-44.

146. Lee, J.T., *Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome*. Genes Dev, 2009. **23**(16): p. 1831-42.
147. Borsani, G., et al., *Characterization of a murine gene expressed from the inactive X chromosome*. Nature, 1991. **351**(6324): p. 325-9.
148. Deininger, P.L. and M.A. Batzer, *Mammalian retroelements*. Genome Res, 2002. **12**(10): p. 1455-65.
149. Cabili, M.N., et al., *Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses*. Genes Dev, 2011. **25**(18): p. 1915-27.
150. Lin, M.F., I. Jungreis, and M. Kellis, *PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions*. Bioinformatics, 2011. **27**(13): p. i275-82.
151. Punta, M., et al., *The Pfam protein families database*. Nucleic Acids Res, 2012. **40**(Database issue): p. D290-301.
152. Cowin, P.A., et al., *Profiling the cancer genome*. Annu Rev Genomics Hum Genet, 2010. **11**: p. 133-59.
153. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer*. Cell, 2007. **128**(4): p. 683-92.
154. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs*. Cell, 2007. **129**(7): p. 1311-23.
155. Tsai, M.C., et al., *Long noncoding RNA as modular scaffold of histone modification complexes*. Science, 2010. **329**(5992): p. 689-93.
156. Gupta, R.A., et al., *Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis*. Nature, 2010. **464**(7291): p. 1071-6.
157. Kogo, R., et al., *Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers*. Cancer Res, 2011. **71**(20): p. 6320-6.
158. Yang, Z., et al., *Overexpression of long non-coding RNA HOTAIR predicts tumor recurrence in hepatocellular carcinoma patients following liver transplantation*. Ann Surg Oncol, 2011. **18**(5): p. 1243-50.
159. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA*. Nature, 2008. **451**(7175): p. 202-6.
160. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a*. Mol Cell, 2010. **38**(5): p. 662-74.
161. Kino, T., et al., *Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor*. Sci Signal, 2010. **3**(107): p. ra8.
162. Mourtada-Maarabouni, M., et al., *GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer*. Oncogene, 2009. **28**(2): p. 195-208.
163. Miyoshi, N., et al., *Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q*. Genes Cells, 2000. **5**(3): p. 211-20.
164. Zhou, Y., et al., *Activation of p53 by MEG3 non-coding RNA*. J Biol Chem, 2007. **282**(34): p. 24731-42.
165. Astuti, D., et al., *Epigenetic alteration at the DLK1-GTL2 imprinted domain in human neoplasia: analysis of neuroblastoma, pheochromocytoma and Wilms' tumour*. Br J Cancer, 2005. **92**(8): p. 1574-80.
166. Braconi, C., et al., *microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer*. Oncogene, 2011. **30**(47): p. 4750-6.
167. Wang, P., Z. Ren, and P. Sun, *Overexpression of the long non-coding RNA MEG3 impairs in vitro glioma cell proliferation*. J Cell Biochem, 2012. **113**(6): p. 1868-74.
168. Zhang, X., et al., *A pituitary-derived MEG3 isoform functions as a growth suppressor in tumor cells*. J Clin Endocrinol Metab, 2003. **88**(11): p. 5119-26.
169. Benetatos, L., et al., *CpG methylation analysis of the MEG3 and SNRPN imprinted genes in acute myeloid leukemia and myelodysplastic syndromes*. Leuk Res, 2010. **34**(2): p. 148-53.

170. Nielsen, C.B., et al., *Determinants of targeting by endogenous and exogenous microRNAs and siRNAs*. RNA, 2007. **13**(11): p. 1894-910.
171. Tripathi, V., et al., *The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation*. Mol Cell, 2010. **39**(6): p. 925-38.
172. Ji, P., et al., *MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer*. Oncogene, 2003. **22**(39): p. 8031-41.
173. Gong, C. and L.E. Maquat, *lncRNAs transactivate STAUI-mediated mRNA decay by duplexing with 3' UTRs via Alu elements*. Nature, 2011. **470**(7333): p. 284-8.
174. Poliseno, L., et al., *A coding-independent function of gene and pseudogene mRNAs regulates tumour biology*. Nature, 2010. **465**(7301): p. 1033-8.
175. Weinberg, R.A., *Tumor suppressor genes*. Science, 1991. **254**(5035): p. 1138-46.
176. Shinohara, M., et al., *AKT in thyroid tumorigenesis and progression*. Endocrinology, 2007. **148**(3): p. 942-7.
177. Ringel, M.D., et al., *Overexpression and overactivation of Akt in thyroid carcinoma*. Cancer Res, 2001. **61**(16): p. 6105-11.
178. Jung, H.S., et al., *USF inhibits cell proliferation through delay in G2/M phase in FRTL-5 cells*. Endocr J, 2007. **54**(2): p. 275-85.
179. Marlow, L.A., et al., *Detailed molecular fingerprinting of four new anaplastic thyroid carcinoma cell lines and their use for verification of RhoB as a molecular therapeutic target*. J Clin Endocrinol Metab, 2010. **95**(12): p. 5338-47.
180. Nakamura, N., et al., *RASSF1A and NORE1A methylation and BRAFV600E mutations in thyroid tumors*. Lab Invest, 2005. **85**(9): p. 1065-75.
181. Nappi, T.C., et al., *Identification of Polo-like kinase 1 as a potential therapeutic target in anaplastic thyroid carcinoma*. Cancer Res, 2009. **69**(5): p. 1916-23.
182. Samadi, A., et al., *A novel HSP90 modulator with selective activity against thyroid cancers in vitro*. Surgery, 2009. **146**(6): p. 1196-207.
183. Kashat, L., et al., *Secretome-based identification and characterization of potential biomarkers in thyroid cancer*. J Proteome Res, 2010. **9**(11): p. 5757-69.
184. Zou, M., et al., *S100A4 (Mts1) gene overexpression is associated with invasion and metastasis of papillary thyroid carcinoma*. Br J Cancer, 2005. **93**(11): p. 1277-84.
185. Min, H.S., et al., *S100A4 expression is associated with lymph node metastasis in papillary microcarcinoma of the thyroid*. Mod Pathol, 2008. **21**(6): p. 748-55.
186. Weber, F., et al., *Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination*. J Clin Endocrinol Metab, 2005. **90**(5): p. 2512-21.
187. Coste, I., et al., *Dual function of MyD88 in RAS signaling and inflammation, leading to mouse and human cell transformation*. J Clin Invest, 2010. **120**(10): p. 3663-7.
188. Khurana, R., et al., *Unraveling the amyloid associated with human medullary thyroid carcinoma*. Endocrinology, 2004. **145**(12): p. 5465-70.
189. Liftin, A.J., *Nonmedullary carcinoma of the thyroid gland in association with amyloid goiter: case report*. Mt Sinai J Med, 1985. **52**(3): p. 225-7.
190. Nessim, S. and M. Tamilia, *Papillary thyroid carcinoma associated with amyloid goiter*. Thyroid, 2005. **15**(4): p. 382-5.
191. Polliack, A. and U. Freund, *Mixed papillary and follicular carcinoma of the thyroid gland with stromal amyloid*. Am J Clin Pathol, 1970. **53**(5): p. 592-5.
192. Valenta, L.J., et al., *Microfollicular thyroid carcinoma with amyloid rich stroma, resembling the medullary carcinoma of the thyroid (MCT)*. Cancer, 1977. **39**(4): p. 1573-86.
193. Weng, M.T., et al., *Papillary thyroid cancer associated with light-chain amyloidosis initially presenting with small-bowel bleeding and protein-losing enteropathy*. J Clin Oncol, 2010. **28**(29): p. e557-9.
194. Reches, M., Y. Porat, and E. Gazit, *Amyloid fibril formation by pentapeptide and tetrapeptide fragments of human calcitonin*. J Biol Chem, 2002. **277**(38): p. 35475-80.

195. Chua, Y.L., et al., *The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene.* *Oncogene*, 2009. **28**(46): p. 4041-52.
196. Cao, Z., R.M. Umek, and S.L. McKnight, *Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells.* *Genes Dev*, 1991. **5**(9): p. 1538-52.
197. Johnson, P.F. and S.L. McKnight, *Eukaryotic transcriptional regulatory proteins.* *Annu Rev Biochem*, 1989. **58**: p. 799-839.
198. Calkhoven, C.F., C. Muller, and A. Leutz, *Translational control of C/EBPalpha and C/EBPbeta isoform expression.* *Genes Dev*, 2000. **14**(15): p. 1920-32.
199. D'Alo, F., et al., *The amino terminal and E2F interaction domains are critical for C/EBP alpha-mediated induction of granulopoietic development of hematopoietic cells.* *Blood*, 2003. **102**(9): p. 3163-71.
200. Tsukada, J., et al., *The CCAAT/enhancer (C/EBP) family of basic-leucine zipper (bZIP) transcription factors is a multifaceted highly-regulated system for gene regulation.* *Cytokine*, 2011. **54**(1): p. 6-19.
201. Osada, S., et al., *DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family.* *J Biol Chem*, 1996. **271**(7): p. 3891-6.
202. Metallo, S.J. and A. Schepartz, *Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA.* *Nat Struct Biol*, 1997. **4**(2): p. 115-7.
203. Hollenbeck, J.J. and M.G. Oakley, *GCN4 binds with high affinity to DNA sequences containing a single consensus half-site.* *Biochemistry*, 2000. **39**(21): p. 6380-9.
204. Ramji, D.P. and P. Foka, *CCAAT/enhancer-binding proteins: structure, function and regulation.* *Biochem J*, 2002. **365**(Pt 3): p. 561-75.
205. Cassel, T.N., et al., *Synergistic transactivation of the differentiation-dependent lung gene Clara cell secretory protein (secretoglobin 1a1) by the basic region leucine zipper factor CCAAT/enhancer-binding protein alpha and the homeodomain factor Nkx2.1/thyroid transcription factor-1.* *J Biol Chem*, 2002. **277**(40): p. 36970-7.
206. Hsu, W., et al., *Fos and Jun repress transcription activation by NF-IL6 through association at the basic zipper region.* *Mol Cell Biol*, 1994. **14**(1): p. 268-76.
207. Wang, N.D., et al., *Impaired energy homeostasis in C/EBP alpha knockout mice.* *Science*, 1995. **269**(5227): p. 1108-12.
208. Zhang, D.E., et al., *Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice.* *Proc Natl Acad Sci U S A*, 1997. **94**(2): p. 569-74.
209. Basseres, D.S., et al., *Respiratory failure due to differentiation arrest and expansion of alveolar cells following lung-specific loss of the transcription factor C/EBPalpha in mice.* *Mol Cell Biol*, 2006. **26**(3): p. 1109-23.
210. Tontonoz, P., E. Hu, and B.M. Spiegelman, *Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor.* *Cell*, 1994. **79**(7): p. 1147-56.
211. Cassel, T.N., G. Suske, and M. Nord, *C/EBP alpha and TTF-1 synergistically transactivate the Clara cell secretory protein gene.* *Ann N Y Acad Sci*, 2000. **923**: p. 300-2.
212. Umek, R.M., A.D. Friedman, and S.L. McKnight, *CCAAT-enhancer binding protein: a component of a differentiation switch.* *Science*, 1991. **251**(4991): p. 288-92.
213. Slomiany, B.A., et al., *C/EBPalpha inhibits cell growth via direct repression of E2F-DP-mediated transcription.* *Mol Cell Biol*, 2000. **20**(16): p. 5986-97.
214. Timchenko, N.A., et al., *CCAAT/enhancer-binding protein alpha (C/EBP alpha) inhibits cell proliferation through the p21 (WAF-1/CIP-1/SDI-1) protein.* *Genes Dev*, 1996. **10**(7): p. 804-15.
215. Wang, H., et al., *C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4.* *Mol Cell*, 2001. **8**(4): p. 817-28.
216. Pedersen, T.A., et al., *Cooperation between C/EBPalpha TBP/TFIIB and SWI/SNF recruiting domains is required for adipocyte differentiation.* *Genes Dev*, 2001. **15**(23): p. 3208-16.

217. Pabst, T., et al., *Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia.* Nat Genet, 2001. **27**(3): p. 263-70.
218. Paz-Priel, I. and A. Friedman, *C/EBPalpha dysregulation in AML and ALL.* Crit Rev Oncog, 2011. **16**(1-2): p. 93-102.
219. Birkenmeier, E.H., et al., *Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein.* Genes Dev, 1989. **3**(8): p. 1146-56.
220. Watkins, P.J., et al., *Impaired proliferation and tumorigenicity induced by CCAAT/enhancer-binding protein.* Cancer Res, 1996. **56**(5): p. 1063-7.
221. Halmos, B., et al., *Down-regulation and antiproliferative role of C/EBPalpha in lung cancer.* Cancer Res, 2002. **62**(2): p. 528-34.
222. Loomis, K.D., et al., *Genetic ablation of CCAAT/enhancer binding protein alpha in epidermis reveals its role in suppression of epithelial tumorigenesis.* Cancer Res, 2007. **67**(14): p. 6768-76.
223. Huber, R., et al., *Regulation of C/EBPbeta and resulting functions in cells of the monocytic lineage.* Cell Signal, 2012. **24**(6): p. 1287-96.
224. Friedman, A.D., *Transcriptional control of granulocyte and monocyte development.* Oncogene, 2007. **26**(47): p. 6816-28.
225. Hirai, H., et al., *C/EBPbeta is required for 'emergency' granulopoiesis.* Nat Immunol, 2006. **7**(7): p. 732-9.
226. Screpanti, I., et al., *Lymphoproliferative disorder and imbalanced T-helper response in C/EBP beta-deficient mice.* EMBO J, 1995. **14**(9): p. 1932-41.
227. Duprez, E., *A new role for C/EBPbeta in acute promyelocytic leukemia.* Cell Cycle, 2004. **3**(4): p. 389-90.
228. Zahnw, C.A., et al., *Overexpression of C/EBPbeta-LIP, a naturally occurring, dominant-negative transcription factor, in human breast cancer.* J Natl Cancer Inst, 1997. **89**(24): p. 1887-91.
229. Guerzoni, C., et al., *Effects of C/EBPalpha and C/EBPbeta in BCR/ABL-expressing cells: differences and similarities.* Cell Cycle, 2006. **5**(12): p. 1254-7.
230. Ewing, S.J., et al., *C/EBPbeta represses p53 to promote cell survival downstream of DNA damage independent of oncogenic Ras and p19(Arf).* Cell Death Differ, 2008. **15**(11): p. 1734-44.
231. Buck, M., H. Turler, and M. Chojkier, *LAP (NF-IL-6), a tissue-specific transcriptional activator, is an inhibitor of hepatoma cell proliferation.* EMBO J, 1994. **13**(4): p. 851-60.
232. Zahnw, C.A., *CCAAT/enhancer-binding protein beta: its role in breast cancer and associations with receptor tyrosine kinases.* Expert Rev Mol Med, 2009. **11**: p. e12.
233. Koschmieder, S., et al., *Dysregulation of the C/EBPalpha differentiation pathway in human cancer.* J Clin Oncol, 2009. **27**(4): p. 619-28.
234. Pomerance, M., et al., *Expression, hormonal regulation, and subcellular localization of CCAAT/enhancer-binding protein-beta in rat and human thyrocytes.* Thyroid, 2005. **15**(3): p. 197-204.
235. Akagi, T., et al., *Induction of sodium iodide symporter gene and molecular characterisation of HNF3 beta/FoxA2, TTF-1 and C/EBP beta in thyroid carcinoma cells.* Br J Cancer, 2008. **99**(5): p. 781-8.
236. Yang, L., et al., *A tumor suppressor and oncogene: the WT1 story.* Leukemia, 2007. **21**(5): p. 868-76.
237. Kossatz, U. and N.P. Malek, *p27: tumor suppressor and oncogene ...?* Cell Res, 2007. **17**(10): p. 832-3.
238. Pelosi, G., et al., *Dual role of RASSF1 as a tumor suppressor and an oncogene in neuroendocrine tumors of the lung.* Anticancer Res, 2010. **30**(10): p. 4269-81.

