

Praca doktorska

**Mutacje somatyczne i zmiany liczby kopii  
w raku podstawnokomórkowym skóry  
- ze szczególnym uwzględnieniem mutacji w niekodujących  
częściach genów i miRNomie**

Paulina Maria Nawrocka-Muszyńska

Promotor: prof. dr hab. Piotr Kozłowski  
Promotor pomocniczy: dr Paulina Gałka-Marciniak

Praca doktorska została wykonana w  
Zakładzie Genetyki Molekularnej Instytutu Chemii Bioorganicznej  
Polskiej Akademii Nauk

Poznań, 2023

*Spis treści*

<b>I. STRESZCZENIE</b> .....	<b>4</b>
<b>II. ABSTRACT</b> .....	<b>6</b>
<b>III. WYKAZ PUBLIKACJI</b> .....	<b>8</b>
<b>IV. WSTĘP</b> .....	<b>9</b>
<b>1. Charakterystyka chorób nowotworowych</b> .....	<b>9</b>
<b>1.1 Cechy komórek nowotworowych</b> .....	<b>9</b>
<b>1.2 Wpływ mutacji napędzających oraz towarzyszących na rozwój nowotworów</b> .....	<b>11</b>
<b>1.3 Mutacje w genach supresorowych oraz onkogenach</b> .....	<b>11</b>
<b>1.4 Identyfikacja nowotworowych mutacji somatycznych w regionach niekodujących</b> .....	<b>13</b>
1.4.1 Promotory i potencjalny efekt mutacji na ich funkcje .....	16
1.4.2 Regiony 5'UTR i potencjalny efekt mutacji na ich funkcje.....	18
1.4.3 Introny i potencjalny efekt mutacji na ich funkcje.....	19
1.4.4 Regiony 3'UTR i potencjalny efekt mutacji na ich funkcje.....	20
1.4.5 Geny miRNA i potencjalny efekt mutacji w genach miRNA.....	22
<b>1.5 Metody wykrywania mutacji w nowotworach</b> .....	<b>24</b>
<b>2. Rak podstawnokomórkowy skóry (BCC)</b> .....	<b>25</b>
<b>2.1 Czynniki genetyczne i środowiskowe wpływające na rozwój BCC</b> .....	<b>26</b>
2.1.1 Najczęściej mutowane geny napędzające, kluczowe dla rozwoju BCC.....	27
<b>V. CEL BADAŃ</b> .....	<b>32</b>
<b>VI. WYNIKI</b> .....	<b>34</b>
<b>1. Sekwencjonowanie całego eksomu (WES)</b> .....	<b>35</b>
<b>1.1. Analiza podstawowych parametrów sekwencjonowania oraz lokalizacji i typów wykrytych mutacji</b> .....	<b>36</b>
<b>1.2 Walidacja mutacji metodą sekwencjonowania Sangera - estymacja częstości mutacji fałszywie pozytywnych</b> .....	<b>39</b>
<b>1.3 Obciążenie mutacyjne (TMB) w BCC</b> .....	<b>42</b>
<b>1.4 Sygnatury mutacyjne w BCC</b> .....	<b>44</b>
<b>1.5 Analiza gorących punktów mutacji w BCC</b> .....	<b>46</b>
1.5.1 Gorące punkty mutacji w regionach kodujących białko .....	50
1.5.2 Gorące punkty mutacji w regionach niekodujących.....	54
<b>1.6 Geny często mutowane w BCC</b> .....	<b>60</b>
1.6.1 Geny często mutowane w regionach kodujących białko .....	61
1.6.3 Geny często mutowane w regionie 5'UTR.....	67
1.6.4 Geny często mutowane w regionie 3'UTR.....	69
1.6.5 Geny często mutowane w intronach .....	70
<b>1.7 Identyfikacja genów napędzających rozwój nowotworu</b> .....	<b>71</b>
1.7.1. Geny napędzające wyselekcjonowane na podstawie mutacji w regionach kodujących białko.....	71
1.7.2 Geny napędzające wyselekcjonowane na podstawie mutacji w regionie 5'UTR.....	73
1.7.3 Geny napędzające wyselekcjonowane na podstawie mutacji w regionie 3'UTR.....	74
<b>1.8 Identyfikacja regionów istotnie wzbogaconych w zmiany liczby kopii</b> .....	<b>75</b>

## SPIS TREŚCI

1.8.1 Zmiany liczby kopii na poziomie ramienia chromosomu napędzające proces nowotworzenia.....	75
1.8.2 Ogniskowe zmiany liczby kopii napędzające proces nowotworzenia.....	76
1.8.3 Walidacja regionów ulegających zmianie liczby kopii napędzających rozwój BCC .....	77
<b>2. Sekwencjonowanie całego miRNomu (WMS) .....</b>	<b>79</b>
2.1 Panel do sekwencjonowania genów miRNA .....	79
2.2 Lokalizacja i typy mutacji w genach miRNA .....	80
2.3 Dodatkowa walidacja mutacji metodą sekwencjonowania Sangera.....	85
2.4 Często mutowane geny miRNA, gorące punkty mutacji i geny MIRNA napędzające rozwój BCC.....	87
2.5 Analiza mutacji w regionie źródłowym .....	89
<b>3. Mutacje w promotorach genów TERT i DPH3 .....</b>	<b>91</b>
<b>VII. DYSKUSJA.....</b>	<b>94</b>
<b>VIII. GŁÓWNE WYNIKI I WNIOSKI .....</b>	<b>104</b>
<b>IX. MATERIAŁY I METODY .....</b>	<b>106</b>
<b>1.1 Materiały.....</b>	<b>106</b>
1.1.1 Enzymy.....	106
1.1.2 Zestawy komercyjne.....	106
1.1.3 Bufory i roztwory.....	106
1.1.4 Inne odczynniki chemiczne .....	107
1.1.5 Wzorce wielkości DNA .....	108
1.1.6 Żele agarozowe.....	108
1.1.7 Materiał biologiczny.....	108
1.1.8 Oligonukleotydy.....	109
1.1.9 Sprzęt laboratoryjny.....	114
1.1.10 Programy komputerowe i bazy danych.....	114
<b>1.2 Metody.....</b>	<b>117</b>
1.2.1 Homogenizacja tkanek oraz izolacja materiału genetycznego .....	117
1.2.2 Ocena ilościowa i jakościowa kwasów nukleinowych .....	117
1.2.3 Strącenie kwasów nukleinowych .....	118
1.2.4 Przygotowanie DNA do utworzenia bibliotek.....	118
1.2.5 Przygotowanie, wzbogacenie oraz sekwencjonowanie bibliotek DNA.....	119
1.2.5.1 Biblioteka DNA do WES.....	119
1.2.5.2 Biblioteka DNA do WMS.....	119
1.2.6 Analiza danych otrzymanych z sekwencjonowania NGS.....	120
1.2.7 Analiza sygnatur nowotworowych .....	122
1.2.8 Wykrywanie gorących miejsc mutacji.....	122
1.2.9 Analiza mutacji dostępnych w bazie danych TCGA.....	122

## SPIS TREŚCI

1.2.10 Wizualizacja mutacji w genach kodujących białko oraz genach miRNA.....	123
1.2.11 Wpływ mutacji na wybór kodonu start i efektywność inicjacji translacji .....	123
1.2.12 Analiza i wizualizacja struktury drugorzędowej 2D i 3D RNA.....	123
1.2.13 Przewidywanie miRNA regulujących ekspresję genów .....	124
1.2.14 Przewidywanie genów, których ekspresję mogą regulować miRNA .....	124
1.2.15 Wykrywanie mutacji napędzających.....	124
1.2.16 Analiza regionów istotnie wzbogaconych w zmiany liczby kopii.....	124
1.2.17 Reakcja odwrotnej transkrypcji (RT).....	125
1.2.18 Łańcuchowa reakcja polimerazy (PCR).....	125
1.2.19 Sekwencjonowanie produktów PCR metodą sekwencjonowania Sangera.....	126
1.2.20 Reakcja MLPA .....	126
1.2.21 Dodatkowe analizy statystyczne.....	128
<b>X. SPIS RYCIN I TABEL.....</b>	<b>129</b>
1.1 Ryciny.....	129
1.2 Tabele.....	130
<b>XI. WYKAZ SKRÓTÓW .....</b>	<b>131</b>
<b>XII. ŹRÓDŁA FINANSOWANIA BADAŃ .....</b>	<b>134</b>
<b>XIII. BIBLIOGRAFIA.....</b>	<b>135</b>
<b>XIV. ZAŁĄCZNIKI .....</b>	<b>149</b>

## I. STRESZCZENIE

Rak podstawnokomórkowy skóry (ang. *basal cell carcinoma*; BCC) jest najbardziej powszechnym nowotworem człowieka, występującym głównie u przedstawicieli populacji kaukaskiej. Rozwój BCC jest napędzany głównie przez mutacje w genach zaangażowanych w szlak sygnałowy Sonic Hedgehog (Shh), niemniej jednak patogeneza BCC nie jest do końca poznana. Podobnie jak w innych nowotworach, do tej pory analizy genetyczne BCC skupiały się niemal wyłącznie na regionach kodujących białka, stanowiących zaledwie ~2% genomu. W związku z powyższym, zasadniczym celem niniejszej pracy była wstępna charakterystyka mutacji somatycznych w niekodujących elementach genów kodujących białka (5'UTR, fragmenty 3'UTR oraz fragmenty intronów) oraz w genach miRNA. W pracy poddano analizie panel 27 par próbek DNA wyizolowanych z BCC oraz z odpowiadających im tkanek normalnych. Próbkę poddano dwóm procedurom sekwencjonowania nowej generacji, rutynowemu sekwencjonowaniu całego eksomu (ang. *Whole Exome Sequencing*; WES) oraz unikalnej, zaprojektowanej w Zakładzie Genetyki Molekularnej, Instytutu Chemii Bioorganicznej PAN, procedurze sekwencjonowania całego miRNomu (ang. *Whole miRNome Sequencing*; WMS).

W wyniku przeprowadzonego sekwencjonowania łącznie zostało wykrytych ponad 80 tysięcy mutacji somatycznych z czego ponad połowa występowała w regionach niekodujących, w których częstość mutacji była zbliżona do tej w regionach kodujących. Średnie obciążenie mutacyjne analizowanych próbek wynosiło >50 mutacji/Mpz co wskazuje, że częstość mutacji w BCC jest najwyższa wśród nowotworów człowieka. Najczęstszą sygnaturą mutacyjną, występującą we wszystkich próbkach, była sygnatura 7, co potwierdza, że głównym czynnikiem mutagennym w BCC jest promieniowanie słoneczne.

Analiza wykrytych mutacji wykazała szereg regionów niekodujących o podwyższonej częstości mutacji, w tym regionów zawierających gorące punkty mutacji. Niektóre mutacje w regionach niekodujących wskazują na duży potencjał mutowanych genów jako elementów napędzających nowotworzenie w BCC m.in. 3'UTR w genie *BAD*, sekwencja Kozak w 5'UTR w genie *DHODH* oraz 5'UTR w genie *CHCHD2*. Wszystkie te geny są funkcjonalnie zaangażowane w procesy związane z nowotworzeniem (apoptoza, metabolizm mitochondrialny i synteza *de novo* pirymidyn). Dodatkowo analizy przeprowadzone z

## STRESZCZENIE

wykorzystaniem zewnętrznych zasobów danych, w tym The Cancer Genome Atlas (TCGA) wykazały, że mutacje w genach *BAD* i *CHCHD2* często występują także w czerniaku, natomiast mutacje wykryte w genie *DHODH* są specyficzne i występują wyłącznie w BCC.

Wśród mutacji niekodujących 171 zostało wykrytych w genach miRNA, w tym w różnych funkcjonalnych elementach tych genów, kluczowych dla biogenezy i funkcjonowania cząsteczek miRNA. Najczęściej mutowanym genem miRNA był *MIR3928* o dobrze udokumentowanej funkcji w nowotworzeniu, w którym zostały zidentyfikowane 4 mutacje, w tym 3 w gorącym punkcie mutacji zlokalizowanym w sekwencji flankującej 5'.

Wygenerowane dane pozwoliły również na przeprowadzenie pierwszej kompleksowej analizy zmian liczby kopii w BCC, w wyniku której, między innymi, zostały zidentyfikowane częste delecje długiego ramienia chromosomu 9 (chr9q), obejmującego kluczowy dla BCC gen supresorowy *PTCH1* oraz częste duplikacje krótkiego ramienia chromosomu 9 (chr9p) zawierającego gen *JAK2* oraz geny ligandów PD-L1 oraz PD-L2 będących kluczowymi elementami punktów kontrolnych odpowiedzi immunologicznej.

Wśród genów najczęściej mutowanych w sekwencjach kodujących zidentyfikowano takie geny jak *PTCH1*, *TP53*, *MYCN* czy geny *NOTCH*. Duża zgodność wyników analizy sekwencji kodujących z ich wcześniejszymi analizami w BCC potwierdziła rzetelność przeprowadzonych analiz i wiarygodność uzyskanych wyników.

Podsumowując w ramach niniejszej pracy przeprowadzono pierwszą analizę i charakterystykę mutacji w regionach niekodujących w BCC. W ramach analizy zidentyfikowano szereg mutacji niekodujących o potencjalnym znaczeniu dla rozwoju BCC. Niniejsze wyniki stanowią podstawę do dalszych analiz niekodujących wariantów w BCC i innych typach raka.

## II. ABSTRACT

### **Somatic mutations and copy number alterations in basal cell carcinoma- focus on gene-associated noncoding variants and miRNome**

Basal cell carcinoma (BCC) of the skin is the most common human cancer, especially frequent in the Caucasian population. It is known that the development of BCC is driven by mutations in genes involved in the Sonic Hedgehog (Shh) signaling pathway, however, the pathogenesis of BCC is not fully understood. As in other cancer types, genetic analyses in BCC have been focused almost exclusively on the coding regions, constituting only ~2% of the genome.

The main aim of this study was the preliminary characteristics of somatic mutations in noncoding elements of protein-coding genes (5'UTR, fragments of 3'UTR, and introns) and in miRNA genes. DNA samples isolated from 27 pairs of BCC and corresponding healthy tissues were sequenced with the use of two next generation sequencing procedures, the routine Whole Exome Sequencing (WES) and the unique procedure of Whole miRNome Sequencing (WMS), developed at the Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences.

In total, over 80,000 somatic mutations were detected of which more than half in noncoding regions. The average tumor mutational burden (TMB) was >50 mutations/Mb, which indicates that BCC is the most frequently mutated human cancer. The most common mutational signature found in all samples was signature 7, confirming that the most important mutagenic factor in BCC is UV radiation.

The analysis of mutations in noncoding regions showed that some of the mutations were clustered in specific regions, including hotspots. Moreover, some mutations in noncoding regions strongly suggest the cancer-driving potential of the mutated genes, e.g., mutations in 3'UTR in *BAD*, mutations in the Kozak sequence in the *DHODH* 5'UTR, and mutations in 5'UTR in *CHCHD2*. All these genes are functionally involved in processes related to carcinogenesis (apoptosis, mitochondrial metabolism, and de novo pyrimidines synthesis). The analysis of external resources of cancer genomic data such as The Cancer Genome Atlas

## ABSTRACT

(TCGA) demonstrated that mutations in *BAD* and *CHCHD2* occur also in melanoma, while mutations in *DHODH* are specific for BCC.

Among the noncoding mutations, 171 were detected in miRNA genes, including mutations in various functional elements of these genes, crucial for miRNA functioning and biogenesis. The most frequently mutated miRNA gene was *MIR3928* whose functional role in cancer is well documented. All the mutations in *MIR3928* were localized in the 5' flanking region of the miRNA precursor of which 3 were in a single hotspot position.

Taking advantage of the generated sequencing data allowed for the first comprehensive analysis of copy number variation in BCC. Among the identified most frequent copy number alterations are deletions of the long arm of chromosome 9 (chr9q), encompassing *PTCH1*, a key BCC tumor suppressor, and duplications of the short arm of chromosome 9 (chr9p), encompassing the oncogene *JAK2* and the genes encoding the key immune checkpoint ligands PD-L1 and PD-L2.

Among the genes identified as most frequently mutated in coding sequences, are genes such as *PTCH1*, *TP53*, *MYCN*, and *NOTCH* well known as key drivers in BCC. This high consistency of the results of the analysis of coding sequences with previous BCC genetic analyzes confirms the reliability of the performed analyzes and the credibility of the obtained results.

In summary, the study is the first analysis of noncoding mutations in BCC. The analysis enabled the identification of noncoding mutations with the potential of cancer drivers in BCC. The results provide a strong basis for further analyses of noncoding variants in BCC and cancer in general.



### III. WYKAZ PUBLIKACJI

Publikacja, której wyniki stanowią znaczną część rozprawy doktorskiej:

1. Nawrocka, P.M., Galka-Marciniak, P., Urbanek-Trzeciak, M.O., Ilamathi, M., Szostak, N., Philips, A., Susok, L., Sand, M., & Kozłowski, P. Profile of basal cell carcinoma mutations and copy number alterations - focus on gene-associated noncoding variants. *Front Oncol.* 2021 Nov 25;11:752579. doi: 10.3389/fonc.2021.752579. PMID: 34900699; PMCID: PMC8656283.

Pozostałe publikacje autorki:

1. Lewandowski, K., Kandała, Z., Gniot, M., Pczkowska E., Nawrocka P., Wojtaszewska M., Janowski M., Mariak M., Handschuh L., Kozłowski P. Essential thrombocythaemia progression to the fibrotic phase is associated with a decrease in JAK2 and PDL1 levels. *Ann Hematol.* 2022 Dec;101(12):2665-2677. doi: 10.1007/s00277-022-05001-8. Epub 2022 Oct 21. PMID: 36266510.
2. Bernardi, A.\* , Gobelli, D.\* , Serna, J.\* , Nawrocka, P.\*, March-Rosselló, G., Orduña, A., Kozłowski, P., Simarro, M., & de la Fuente, M. A. Novel fluorescent-based reporter cell line engineered for monitoring homologous recombination events. *PLoS One.* 2021 Apr 30;16(4):e0237413. doi: 10.1371/journal.pone.0237413. PMID: 33930025; PMCID: PMC8087102.  
\* równorzędni pierwsi autorzy pracy
3. Galka-Marciniak, P., Urbanek-Trzeciak, M. O., Nawrocka, P. M., & Kozłowski, P. A pan-cancer atlas of somatic mutations in miRNA biogenesis genes. *Nucleic Acids Res.* 2021 Jan 25;49(2):601-620. doi: 10.1093/nar/gkaa1223. PMID: 33406242; PMCID: PMC7826265.
4. Urbanek-Trzeciak, M. O., Galka-Marciniak, P., Nawrocka, P. M., Kowal, E., Szwec, S., Giefing, M., & Kozłowski, P. (2020). Pan-cancer analysis of somatic mutations in miRNA genes. *EBioMedicine.* 2020 Nov;61:103051. doi: 10.1016/j.ebiom.2020.103051. Epub 2020 Oct 7. PMID: 33038763; PMCID: PMC7648123.
5. Galka-Marciniak, P., Urbanek-Trzeciak, M. O., Nawrocka, P. M., Dutkiewicz, A., Giefing, M., Lewandowska, M. A., & Kozłowski, P. (2019). Somatic Mutations in miRNA Genes in Lung Cancer—Potential Functional Consequences of Non-Coding Sequence Variants. *Cancers (Basel).* 2019 Jun 8;11(6):793. doi: 10.3390/cancers11060793. PMID: 31181801; PMCID: PMC6627760.

## IV. WSTĘP

### 1. Charakterystyka chorób nowotworowych

#### 1.1 Cechy komórek nowotworowych

Nowotwór to nieprawidłowa tkanka, rozwijająca się z pojedynczej komórki, która w wyniku zmian genetycznych nabyła zdolność do niekontrolowanych podziałów komórkowych. Komórka może zapobiegać rozwojowi nowotworu przez indukcję procesu programowanej śmierci (apoptozy) lub naprawę błędów przez aktywację mechanizmów naprawy DNA. Niestety geny zaangażowane w te i inne ważne procesy biologiczne mogą ulegać mutacjom, co w konsekwencji może prowadzić do niekontrolowanych podziałów i powstania nowotworu. Wyróżniamy 14 cech charakteryzujących tkankę nowotworową. W 2000 r. D. Hanahan i R. A. Weinberg zaproponowali pierwsze 6 takich cech (Hanahan & Weinberg, 2000): (i) nieograniczony potencjał replikacyjny, (ii) spadek aktywności genów supresorowych, (iii) unikanie programowanej śmierci komórki, (iv) indukcja angiogenezy, (v) inwazyjność i zdolność do przerzutów oraz (vi) wytwarzanie własnych czynników wzrostowych. W roku 2011 do powyżej listy dodano: (vii) niestabilność genetyczną wynikającą z nagromadzenia mutacji, (viii) indukowanie stanu zapalnego organizmu, (ix) unikanie odpowiedzi układu immunologicznego, a także (x) przeprogramowanie metabolizmu energetycznego pod wpływem hipoksji (Hanahan & Weinberg, 2011). W 2022 r. D. Hanahan opisał kolejne 4 cechy komórek nowotworowych: (xi) plastyczność fenotypową, (xii) zakłócone różnicowanie komórek, (xiii) niemutacyjne przeprogramowanie epigenetyczne oraz (xiv) polimorfizm mikrobiomu (Hanahan, 2022).

Warto zaznaczyć, że nowotwór to choroba o podłożu genetycznym, charakteryzująca się nagromadzeniem mutacji w obrębie materiału genetycznego komórki, dlatego wszystkie wyżej wymienione cechy są warunkowane zmianami genetycznymi, wśród których najpowszechniejsze to somatyczne mutacje, somatyczne zaburzenia liczby kopii (amplifikacje, delecje) jak również zmiany epigenetyczne. Należy również zaznaczyć, że nowotwór rozwija się na tle konstytutywnego tła genetycznego, którego warianty, w tym dziedziczne mutacje mogą prowadzić do podwyższonej predyspozycji niektórych nowotworów i ich rodzinnych agregacji oraz wpływać na przebieg choroby.

## WSTĘP

W przypadku zmian somatycznych bardzo często ilość mutacji określa się przy użyciu parametru wskazującego obciążenie mutacyjne nowotworu (ang. *tumor mutational burden*; TMB). Parametr ten określa ilość mutacji w przeliczeniu na milion par zasad (Mpz), a jego wartość różni się w zależności od typu nowotworu (Lawrence et al., 2013). Ekstremalnie wysoka wartość parametru TMB określana jako zjawisko hipermutacji może być efektem działania endogennych procesów mutagennych m.in. i) uszkodzeń genów biorących udział w mechanizmie naprawy błędnie sparowanych zasad azotowych w DNA (ang. *DNA mismatch repair*; MMR) co prowadzi do zjawiska niestabilności regionów mikrosatelitarnych (ang. *microsatellite instability*; MSI), ii) zwiększeniem aktywności deaminaz APOBEC, a także może być spowodowana działaniem silnych mutagenów egzogennych np. światła ultrafioletowego (UV), czy dymu papierosowego (Schlesner & Eils, 2015). Nowotwory charakteryzujące się najwyższym obciążeniem mutacyjnym to głównie nowotwory skóry takie jak czerniak, rak podstawnokomórkowy skóry (ang. *basal cell carcinoma*; BCC) i kolczystokomórkowy skóry (ang. *squamous cell carcinoma*; SCC), których rozwój jest spowodowany m.in. przez nadmierną ekspozycję na promieniowanie UV, nowotwory płuc wywołane działaniem substancji zawartych w dymie tytoniowym, nowotwory jelita grubego charakteryzujące się częstymi uszkodzeniami mechanizmu MMR, a także nowotwory pęcherza moczowego i szyjki macicy wykazujące zwiększoną aktywność APOBEC.

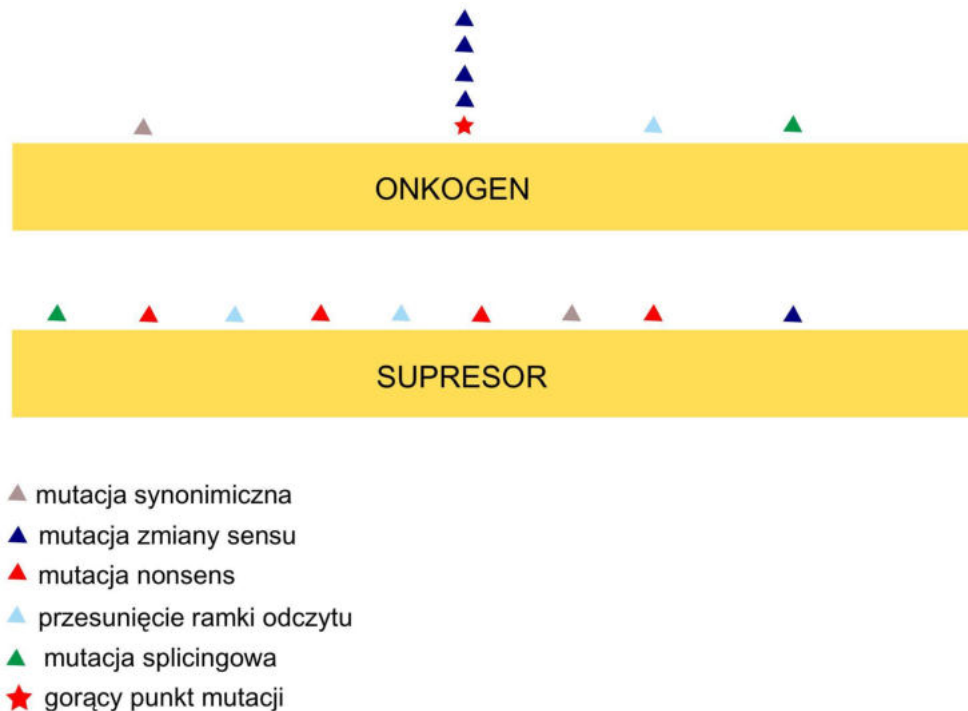
Na podstawie analizy przeprowadzonej przez L.B. Alexandrova i wsp. (Alexandrov et al., 2013) na podstawie danych całogenomowych dla różnych typów nowotworów wykazano, że mutacje mogą układać się w tzw. sygnatury mutacyjne - wzór częstości występowania różnych typów mutacji w określonym kontekście nukleotydowym (biorąc pod uwagę bezpośredni kontekst nukleotydowy w pozycji -1 i +1), charakterystyczny dla konkretnego nowotworu lub typu nowotworu, procesu mutagennego oraz rodzaju uszkodzeń mechanizmu naprawy DNA. Początkowo analizując mutacje typu substytucja pojedynczego nukleotydu (ang. *single base substitution*; SBS) w 39 typach nowotworów rozpoznano 30 sygnatur mutacyjnych. Aktualnie wyselekcjonowano 60 sygnatur SBS, 11 sygnatur DBS (ang. *double base substitution*), analizujących substytucje dwóch nukleotydów, 18 sygnatur ID (ang. *small insertions and deletions*; ID), analizujących indele, oraz 21 sygnatur CN (ang. *copy number*), biorących pod uwagę różnej wielkości zmiany liczby kopii (Steele et al., 2022). Szczegółowa charakterystyka poszczególnych sygnatur jest dostępna w bazie danych Catalogue of Somatic Mutations in Cancer (COSMIC; <https://cancer.sanger.ac.uk/cosmic>).

### **1.2 Wpływ mutacji napędzających oraz towarzyszących na rozwój nowotworów**

Warto zaznaczyć, że nie wszystkie mutacje bezpośrednio wpływają na proces kancerogenezy, dlatego niezwykle istotnym wyzwaniem jest rozróżnienie, które zmiany genetyczne są czynnikiem napędzającym rozwój nowotworu, są zdolne do inicjacji procesu nowotworzenia, a które są tylko zmianami towarzyszącymi. Co ważne, tylko bardzo niewielka część mutacji w nowotworze to mutacje napędzające (ang. *driver mutation*) dające komórkom nowotworowym selektywną przewagę, zwiększając ich przeżywalność i proliferację. Geny, które charakteryzują się występowaniem mutacji napędzających nazywane są genami napędzającymi (ang. *driver gene*). Identyfikacja genów napędzających jest istotna nie tylko dla zrozumienia przyczyny rozwoju nowotworu, ale może być także wykorzystywana w diagnostyce czy terapiach celowanych. Zdecydowana większość mutacji należy do grupy mutacji towarzyszących (ang. *passenger mutations*), które są funkcjonalnie obojętne (neutralne) (Bozic et al., 2010), często selekcionowane są razem z mutacjami napędzającymi, w tych samych klonach lub poprzez bliską lokalizację chromosomową.

### **1.3 Mutacje w genach supresorowych oraz onkogenach**

Mutacje napędzające bardzo często lokalizują się w dwóch typach genów, które możemy klasyfikować według ich potencjału onkogenego jako protoonkogeny/onkogeny lub geny supresorowe. Z uwagi na fakt, że geny te pełnią różną funkcję w rozwoju nowotworów, występujące w nich zmiany charakteryzują się różnym rodzajem mutacji oraz często różnym wzorem ich rozmieszczenia wzdłuż sekwencji genu (Ryc. 1) (Liu, Xing, Yang, & Tian, 2011).



**Rycina 1. Schematyczny rozkład mutacji charakterystycznych dla onkogenów i genów supresorowych.**

Wzdłuż sekwencji genu, oznaczonej żółtym prostokątem, zaznaczono rodzaj i dystrybucję mutacji charakterystycznych dla onkogenów oraz supresorów.

Protoonkogeny to geny potencjalnie zdolne do wyzwolenia procesu nowotworzenia. W wyniku nabycia mutacji aktywujących funkcję onkogenną (ang. *gain-of-function*) protoonkogen może stać się onkogenem, genem którego nadmierna aktywność, stymuluje proces nowotworzenia przez m.in. aktywację podziałów, wzrostu, dojrzewania i różnicowania komórek, jak również unikania apoptozy, unikania rozpoznania i aktywacji procesów immunologicznych czy aktywacji angiogenezy (Anderson, Reynolds, You, & Maronpot, 1992). Zazwyczaj mutacja występująca na jednym z alleli protoonkogenu jest wystarczająca do aktywacji jego onkogennej funkcji. Często mutacje występujące w onkogenach są mutacjami zmiany sensu (ang. *missense*), prowadzącymi do zmiany jednego aminokwasu na inny oraz są zlokalizowane w ściśle określonej pozycji (Bookstein & Allred, 1993) lub występują w bliskiej lokalizacji (np. w obrębie domeny białkowej lub centrum aktywnym enzymu) tworząc tzw. gorące miejsca mutacji (ang. *hotspot*). Na wzrost aktywności onkogenu może mieć również wpływ wzrost liczby kopii genu. Przykładem protoonkogenu jest *EGFR*, kodujący receptor naskórkowego czynnika wzrostu (ang. *epidermal growth factor receptor*), który stymuluje

## WSTĘP

wzrost komórek po uprzedniej stymulacji odpowiednim ligandem. W wyniku mutacji EGFR ulega aktywacji pomimo braku ligandu, prowadząc do aktywacji szlaku sygnałowego prowadzącego do nieograniczonych podziałów komórkowych i nadmiernego wzrostu komórek nowotworowych (Sigismund, Avanzato, & Lanzetti, 2018).

Kolejną grupę genów uczestniczących w onkogenezie stanowią geny supresorowe, kodujące białka hamujące proces kancerogenezy, m.in. przez inhibicję proliferacji komórek czy blokadę przejścia uszkodzonej komórki na ścieżkę apoptozy, dlatego onkogenne mutacje w genach supresorowych zwykle mają charakter mutacji utraty funkcji (ang. *loss-of-function*). Według hipotezy Knudsona inaktywacja genu supresorowego wymaga wyłączenia obu alleli (ang. *double hit, biallelic*) dlatego mutacje w genach supresorowych często występują w obu allelach (Bookstein & Allred, 1993). Supresory charakteryzują się większą niż onkogeny częstością mutacji definitywnych (ang. *deleterious*) obejmujących mutacje typu: i) nonsens (przedwczesne wprowadzenie kodonu STOP), co skutkuje występowaniem skróconej formy białka, oraz ii) przesunięcia ramki odczytu (ang. *frameshift*), co ma wpływ na zmianę sekwencji aminokwasowej białka, a także iii) zwiększoną częstością mutacji splicingowych (mutacje występujące w kanonicznych miejscach splicingowych +/-2 nukleotydy (nt) od granicy ekson/intron). Mutacje w genach supresorowych są zazwyczaj rozproszone na całej długości genu (Liu et al., 2011) i zwykle nie tworzą ewidentnych (jak w onkogenach) gorących miejsc mutacji. Przykładem supresora jest gen *TP53*, kodujący białko p53 nazywane "strażnikiem genomu". W przypadku uszkodzenia DNA białko to bierze udział w zahamowaniu cyklu komórkowego i skierowaniu komórki na drogę apoptozy (Aubrey, Kelly, Janic, Herold, & Strasser, 2018). W nowotworze, mutacje lokalizujące się w genie *TP53* skutkują zahamowaniem funkcji supresorowej białka p53.

### **1.4 Identyfikacja nowotworowych mutacji somatycznych w regionach niekodujących**

Do niedawna analiza mutacji w nowotworach ograniczała się głównie do analizy kodującej części genów, stanowiącej zaledwie ~2% całego genomu. Jednoznaczny sposób przewidzenia konsekwencji mutacji w sekwencji kodującej usprawnia funkcjonalne analizy regionów kodujących. Kod genetyczny oraz znajomość regionów kodujących funkcjonalne domeny białek umożliwiają ustalenie konsekwencji mutacji dla zmiany sekwencji oraz funkcji

## WSTĘP

białka. Analizę dodatkowo usprawnia spory wybór algorytmów i narzędzi bioinformatycznych dostosowanych do oceny konsekwencji mutacji w kodującej części genów, ułatwiających analizę dużych zbiorów danych, a także umożliwiających wyselekcjonowanie genów o potencjale napędzającym.

Warto zaznaczyć jednak, że większość mutacji somatycznych występuje w regionach, które nie kodują białek, w skrócie nazywanych regionami niekodującymi, które łącznie stanowią >98% genomu (UniProt Consortium, 2019). Dotychczasowe niewielkie zainteresowanie niekodującą częścią genomu ma przełożenie na ograniczoną dostępność narzędzi bioinformatycznych pomagających w rzetelnej analizie wpływu mutacji na funkcję regionów niekodujących i proces nowotworzenia. Wyzwaniem pozostaje także identyfikacja niekodujących elementów/genów napędzających. Analizowanie mutacji w celu wyselekcjonowania genów napędzających jest trudnym zadaniem statystycznym i obliczeniowym zarówno w przypadku oceny kodującej jak i niekodującej części genomu. Napędzający wpływ genów można zaobserwować na podstawie pozytywnej presji selekcyjnej mutacji w danym genie. Sygnały takiej presji, które wykorzystuje się do identyfikacji genów/regionów napędzających to m.in. i) podwyższona częstość mutacji, oraz lokalizowanie mutacji w gorących punktach, ii) lokalizacja mutacji w miejscach o wysokiej konserwatywności i/lub w motywach funkcjonalnych, czy też przede wszystkim iii) nadwyżka w danym regionie mutacji funkcjonalnych nad mutacjami neutralnymi, np. w sekwencjach kodujących takie rozróżnienie dokonuje się na podstawie porównania częstości mutacji zmiany sensu czy mutacji definitywnych z częstością mutacji synonimicznych, które jest możliwe dzięki znajomości kodu genetycznego. Niestety podobna analiza funkcjonalnych mutacji nie jest obecnie możliwa w regionach niekodujących. Wynika to głównie z różnorodności typów regionów niekodujących oraz z niewystarczającego rozpoznania elementów funkcjonalnych w regionach niekodujących co uniemożliwia łatwe przewidzenie konsekwencji mutacji w tych regionach. Pewnym wskaźnikiem funkcjonalności mutacji w regionach niekodujących mogą być: i) wpływ mutacji na strukturę RNA, ii) wpływ mutacji na motywy sekwencyjne/strukturalne odpowiedzialne za oddziaływanie i wiązanie różnych białek, w tym czynników transkrypcyjnych i innych białek wiążących DNA oraz białek wiążących RNA, iii) wpływ mutacji na oddziaływanie z innymi RNA, np. oddziaływanie miRNA z ich cząsteczkami docelowymi. Jednak powyższe czynniki mają charakter bardzo generalny i mogą być uwzględniane tylko w kontekście odpowiednich elementów niekodujących. Powyższe

## WSTĘP

ograniczenia prowadzą do braku jakichkolwiek narzędzi informatycznych/statystycznych pozwalających na rozpoznawanie sygnału mutacji napędzających dedykowanych dla specyficznych typów regionów niekodujących (np. miRNA, lncRNA czy 3'UTR). Dostępne obecnie narzędzia, takie jak np. OncodriveFML uwzględniają jedynie bardzo ogólne kryteria, takie jak wpływ na strukturę RNA czy konserwatywność bez możliwości rozróżnienia typu analizowanej sekwencji.

Pomimo trudności w analizie mutacji w regionach niekodujących w ostatnich latach pojawiają się nieliczne badania, które uwzględniają analizy tych regionów oraz wskazują na występowanie mutacji w regionach niekodujących w różnych typach nowotworów (Elliott & Larsson, 2021; Rheinbay et al., 2020). Najbardziej rozpoznawalnym przykładem mutacji w regionie niekodującym są mutacje w promotorze genu *TERT* (ang. *telomerase reverse transcriptase*), kodującego odwrotną transkryptazę, katalityczną podjednostkę telomerazy. Telomeraza jest enzymem rybonukleoproteinowym odpowiedzialnym za utrzymywanie stałej długości telomerów, elementów zlokalizowanych na końcach chromosomów i zapewniających ich stabilność. W normalnych warunkach podziały komórkowe skutkują skracaniem się telomerów, ponieważ dojrzałe komórki somatyczne nie mają aktywnej telomerazy (wyjątek stanowią komórki tkanek mających zdolność do odnawiania się, np. komórki macierzyste). Podwyższony poziom telomerazy, obserwowany w wielu komórkach nowotworowych, wynikający m.in. z mutacji w promotorze genu *TERT*, zapobiega skracaniu się telomerów podczas wielokrotnych podziały komórek nowotworowych, a tym samym prowadzi do nieśmiertelności tych komórek i rozwoju nowotworu. Mutacje występujące w promotorze *TERT* tworzą miejsca wiązania dla czynników transkrypcyjnych, prowadząc do zwiększonej transkrypcji i aktywności *TERT* (Horn et al., 2013; Huang et al., 2013). Najczęściej występujące mutacje w promotorze *TERT* są to substytucje zlokalizowane w pozycjach -124 oraz -146 przed kodonem start, skutkujące utworzeniem miejsca wiązania czynnika transkrypcyjnego GABP (Bell et al., 2015).

Kolejnym przykładem często występujących mutacji niekodujących są mutacje w genie *MIR142*, który jest najczęściej mutowanym genem miRNA w nowotworach (Urbanek-Trzeciak et al., 2020). Mutacje w tym genie występują niemal wyłącznie w nowotworach krwi, w tym chłoniaku rozlanym z dużych komórek B (ang. *diffuse large B-cell lymphoma*; DLBCL), chłoniaku grudkowym, ostrej białaczce szpikowej oraz przewlekłej białaczce limfatycznej (Morin et al., 2010; Kwanhian et al., 2012; Thol et al., 2015; Puente et al., 2015; Hezaveh et



## WSTĘP

al., 2016; Bouska et al., 2017; Urbanek-Trzeciak et al., 2020). W przypadku chorób mieloproliferycyjnych oraz guzów litych mutacje w *MIR142* występują bardzo rzadko (Galka-Marciniak et al., 2022; Urbanek-Trzeciak et al., 2020). Mutacje wykryte w *MIR142* są rozmieszczone wzdłuż całej sekwencji genu miRNA, ale skupiają się głównie w obrębie sekwencji regionu źródłowego (ang. *seed*) miRNA generowanego z ramienia 3p, wpływając na zmianę rozpoznawania docelowych transkryptów, co sugeruje ich definitywny wpływ na funkcję dojrzałego miRNA (Urbanek-Trzeciak et al., 2020). Mutacje w regionie źródłowym miR-142-3p skutkują brakiem regulacji ekspresji *ASH1L*, co ma przełożenie na wzrost poziomu *HOXA9/A10* i zaburzenie różnicowania krwiotwórczych komórek progenitorowych prowadząc do rozwoju białaczki (Trissal et al., 2018). Inne geny regulowane przez miR-142-3p, których ekspresja może być zaburzona przez mutacje w *MIR142* to *CD274* (kodujący białko PD-L1), *RAC1*, *PROM1*, *PTEN*, *IL6* oraz *SOCS1* (Kwanhian et al., 2012; Su et al., 2015; Sun et al., 2011; Tolba, Foda, kamal, & Elshabrawy, 2013; Wan, Ling, Peng, & Ding, 2018).

Powyższe przykłady mutacji zlokalizowanych w niekodujących częściach genów, ze względu na swój znaczący wpływ na rozwój nowotworów wskazują na potrzebę analizy mutacji w niekodujących częściach genomu i pokazują, że w celu poznania wpływu mutacji zlokalizowanych w tych regionach, konieczne jest dokładne zrozumienie funkcji pełnionej przez poszczególne elementy niekodujące.

Regiony niekodujące stanowią część genomu, która nie koduje białek. Są to m.in. i) geny kodujące cząsteczki RNA takie jak miRNA, rRNA, tRNA, snRNA, lncRNA, ii) elementy regulatorowe genów, takie jak promotory, regiony 5' oraz 3'UTR, introny, a także iii) pseudogeny, iv) transpozony, oraz v) telomery. W niniejszej pracy skupiono się na regionach niekodujących genów kodujących białka oraz genach miRNA.

### **1.4.1 Promotory i potencjalny efekt mutacji na ich funkcje**

Promotor to fragment DNA o długości od kilkudziesięciu do kilkuset nukleotydów, zwykle położony powyżej sekwencji kodującej genu, biorący udział w regulacji transkrypcji genów. Promotor to region o niezwykle różnorodnej budowie, złożony z podstawowej części rdzeniowej/minimalnej (ang. *core*) oraz dodatkowych sekwencji regulatorowych położonych w większej odległości od miejsca startu transkrypcji (ang. *transcription start site*; TSS). Regulacja transkrypcji odbywa się głównie za pośrednictwem czynników transkrypcyjnych

## WSTĘP

(ang. *transcription factors*; TFs), które wiążą się ze specyficznymi krótkimi sekwencjami DNA występującymi w promotorach genów docelowych oraz inicjują związaną z polimerazą RNA niezbędną do przeprowadzenia procesu transkrypcji. Część rdzeniowa promotora to region 40-50 par zasad (pz) w kierunku 3' oraz 5' od miejsca TSS posiadający motywy sekwencyjne konieczne do zainicjowania procesu transkrypcji. Obecność wszystkich motywów nie jest warunkiem koniecznym do inicjacji transkrypcji. Większość promotorów eukariotycznych zawiera sekwencję TATAWAW (inaczej kasetę TATA; gdzie W to A lub T) zlokalizowaną zazwyczaj około 24-31 pz przed TSS (Hahn, Buratowski, Sharp, & Guarente, 1989). Kasetę TATA stanowi miejsce wiązania kompleksu inicjującego transkrypcję złożonego z podstawowych czynników transkrypcyjnych oraz polimerazy RNA. Sekwencja otaczająca kasetę TATA jest również bardzo ważna, wpływa ona na wydajność inicjacji transkrypcji. W kierunku 5' oraz 3' w stosunku do kasety TATA mogą występować tzw. elementy rozpoznające (ang. *TFIIIB recognition element*; BRE), wiążące podstawowe czynniki transkrypcyjne TFIIIB. Sekwencja konsensusowa BRE powyżej kasety TATA to SSRGCC (gdzie S to C lub G, R to A lub G) (Lagrange, Kapanidis, Tang, Reinberg, & Ebright, 1998) a poniżej RTDKKKK (gdzie D to A/G/T, K to G/T) (Deng & Roberts, 2005). Oba motywy mogą mieć różny wpływ na wydajność transkrypcji powodując zarówno spadek jak i wzrost efektywności transkrypcji. Niektóre promotory posiadają tzw. element inicjatorowy (ang. *initiator element*; Inr), zawierający sekwencję YR (gdzie Y to C lub T, R to TSS oraz A lub G) obejmującą region TSS (Carninci et al., 2006). Inr może samodzielnie inicjować transkrypcję, jednak najczęściej występuje wspólnie z kasetą TATA. Zdarzają się także promotory, które nie mają ani kasety, ani elementu inicjatorowego, poziom transkrypcji tych genów jest zazwyczaj bardzo niski, a miejsce inicjacji nie jest precyzyjnie zdefiniowane. Kolejny motyw kluczowy dla funkcjonowania promotora stanowi DPE (ang. *downstream promoter element*) zlokalizowany w pozycji od +28 do +33 w stosunku do TSS, i jest to element wiążący podstawowy czynnik transkrypcyjny TFIID (Louder et al., 2016). W promotorze mogą występować także elementy kontrolujące transkrypcję położone zazwyczaj nie dalej niż kilkaset pz powyżej kasety TATA i stanowią one kolejne miejsca wiązania dodatkowych TFs poza regionem rdzeniowym promotora. W większości przypadków eukariotyczne TFs to aktywatory transkrypcji, jednak istnieją także TFs będące represorami procesu transkrypcji.

Promotory to odcinki DNA bogate w nukleotydy CG tworzące tzw. wyspy CpG, charakterystyczne dla regulacji transkrypcji genów metabolizmu podstawowego (ang.

## WSTĘP

*housekeeping genes*). Region bogaty w pary GC może skutkować tworzeniem stabilnej struktury w formie kwadrupleksu, która sprzyja wiązaniu TFs i ma przełożenie na wysoką wydajność transkrypcji (Spiegel et al., 2021). Wyspy CpG mogą także prowadzić do spadku poziomu ekspresji genów na skutek metylacji cytozyny w ich obrębie.

Mutacje występujące w promotorach mogą przede wszystkim uszkadzać istniejące sekwencje wiązania czynników transkrypcyjnych lub tworzyć nowe miejsca ich wiązania, a tym samym wpływać na regulację ekspresji genu. Mogą także wpływać na tworzenie struktury kwadrupleksu zwiększając lub obniżając efektywność inicjacji transkrypcji.

### 1.4.2 Regiony 5'UTR i potencjalny efekt mutacji na ich funkcje

Region 5'UTR to nieulegająca translacji część mRNA o długości od kilku do nawet kilku tysięcy nukleotydów (średnia długość u człowieka to 210 nt; (Leppek, Das, & Barna, 2018; Pesole et al., 2001)) położona w kierunku 5' od sekwencji kodującej białko. Region ten charakteryzuje się występowaniem szeregu miejsc mających wpływ zarówno na stabilność transkryptu jak i proces translacji. Region 5'UTR zawiera na swoim 5' końcu tzw. strukturę czapeczki (ang. *cap*; 7-metyloguanozyna) zapewniającą stabilność transkryptów. Dodatkowo, czapeczka stanowi miejsce wiązania małej podjednostki rybosomowej w procesie inicjacji translacji zależnej od czapeczki. Region 5'UTR to region o wysokiej zawartości par GC, które są związane z zapewnieniem stabilności transkryptu i ułatwiają tworzenie struktur m.in. spinek, kwadrupleksów oraz pseudowęzłów. Stabilność lokalnych struktur tworzonych przez region 5'UTR ma znaczący wpływ na regulację procesu translacji, ponieważ zbyt wysoka stabilność ( $dG < -50$  kcal/mol; (Mignone, Gissi, Liuni, & Pesole, 2002)) zaburza migrację kompleksu inicjacyjnego podczas skanowania mRNA w poszukiwaniu kodonu start i inhibuje lub osłabia efektywność inicjacji translacji zależnej od czapeczki. 5'UTR to także miejsce występowania wewnętrznych miejsc wiązania rybosomu (ang. *internal ribosome entry site*; IRES) odpowiedzialnej za proces inicjacji translacji niezależnej od czapeczki, który odbywa się głównie w warunkach stresowych, w trakcie apoptozy, podczas fazy G2/M cyklu komórkowego oraz przy blokadzie lub osłabieniu inicjacji translacji zależnej od czapeczki. Co więcej, region 5'UTR jest miejscem występowania części sekwencji Kozak, 5'- GCCRCCAAUGG, 10 nt sekwencji obejmującej kodon start AUG, mającej wpływ na rozpoznanie przez rybosom kodonu start jako miejsca początku syntezy białka.

## WSTĘP

W nowotworze mutacje w 5'UTR mogą przyczyniać się do zmiany stabilności mRNA, co może skutkować zwiększeniem poziomu ekspresji onkogenów i/lub spadkiem poziomu transkryptów genów supresorowych. Co więcej mutacje w 5'UTR mogą wpływać na wydajność translacji powodując wzrost ilości białek o funkcji onkogennej lub spadek poziomu białek supresorowych na skutek: i) mutacji w sekwencji Kozak, które mogą powodować osłabienie inicjacji translacji co może skutkować wyborem kolejnego kodonu start jako miejsca rozpoczęcia translacji i krótszym białkiem, ii) wzmocnienia sekwencji Kozak może mieć przełożenie na wysoką efektywność procesu inicjacji syntezy białka, iii) zwiększenia stabilności lokalnych struktur drugorzędowych utworzonych w regionie 5'UTR co osłabia lub uniemożliwia proces inicjacji translacji, lub wymusza konieczność inicjacji translacji niezależnej od czapeczki iv) spadku stabilności struktur co ułatwia inicjację translacji, v) wprowadzenia na skutek mutacji nowego kodonu start i utworzenia nowej otwartej ramki odczytu (ang. *open reading frame*; ORF) co może skutkować utworzeniem dodatkowej izoformy białka i blokadą/spadkiem efektywności translacji głównej izoformy.

### 1.4.3 Introny i potencjalny efekt mutacji na ich funkcje

Introny to niekodujące części genów, które oddzielają od siebie eksony. Sekwencje intronów stanowią ~80% ludzkich genów i zawierają elementy funkcjonalne, które kontrolują dojrzewanie transkryptu w procesie składania genów (ang. *splicing*) zwanym także splicingiem (Jung, Lee, & Choi, 2021). W trakcie splicingu z cząsteczki prekursorowego transkryptu (pre-mRNA) zostają wycięte introny a powstały dojrzały transkrypt mRNA staje się matrycą do syntezy białek. Za przebieg składania genów odpowiedzialny jest wieloskładnikowy, rybonukleoproteinowy kompleks składania pre-mRNA zwany spliceosomem. Kluczowym etapem jest rozpoznanie fragmentu, który powinien być wycięty. W intronach znajdują się konserwatywne sekwencje: i) miejsce donorowe, nazywane również miejscem splicingowym 5' (ang. *5' splice site, donor site; 5'ss*), które stanowi sekwencja o długości dziewięciu nukleotydów 5'YAG/GURAGU3' (gdzie "/" oznacza granicę ekson-intron), ii) miejsce akceptorowe, nazywane także miejscem akceptorowym splicingu (ang. *3' splice site, acceptor site; 3'ss*), które stanowi sekwencja o długości pięciu nukleotydów 5'NYAG/G3', iii) miejsce rozgałęzienia (ang. *branch point; BP*), sekwencja o długości siedmiu nukleotydów 5'YNYURAY3' (gdzie N oznacza dowolny nukleotyd) znajdująca się około 40 nt od miejsca 3'ss, zawierająca adenozyne biorącą udział w reakcji transestryfikacji w trakcie splicingu, oraz iv)

## WSTĘP

trakt polipirymidynowy (ang. *polypirimidine tract*; PPT) o różnej długości, znajdujący się pomiędzy 3' ss a BP (Zhang, 1998). Rozpoznanie przez spliceosom sekwencji konserwatywnych w intronie jest niezbędne do przeprowadzenia konstytutywnego składania genów. Możliwy jest także proces alternatywnego splicingu, na który wpływ mają występujące w intronach wzmacniacze (ang. *intrinsic splicing enhancers*; ISEs) lub wyciszacze (ang. *intrinsic splicing silencers*; ISSs) splicingu (Diederichs et al., 2016). ISEs stanowią sekwencje zlokalizowane w intronie, do których mogą wiązać się białka wiążące RNA z rodziny SR (ang. *serine/arginine-rich protein family*) promujące wiązanie spliceosomu i aktywujące proces splicingu. ISSs to sekwencje wiążące białka hnRNP (ang. *heterogeneous nuclear ribonucleoproteins*), które blokują wiązanie spliceosomu i osłabiają splicing (Wang, Ma, Xiao, & Wang, 2012).

Najczęściej mutacje w intronach w miejscach kluczowych dla splicingu wykrywane są w miejscu sekwencji donorowej lub akceptorowej, a mutacje w BP oraz PPT są bardzo rzadkie (Lewandowska, 2013). Mutacje w tych regionach mogą skutkować zmianą sekwencji kodującej białko poprzez zaburzenie procesu składania genów w wyniku tworzenia nowych granic ekson/intron, osłabienia lub wzmocnienia miejsc splicingowych a co za tym idzie utratę eksonów lub włączenie intronów do dojrzałego transkryptu. Mutacje w intronach mogą skutkować także osłabieniem/wzmocnieniem wiązania białek SR oraz hnRNP do miejsc ISEs i ISSs lub skutkować utworzeniem nowych miejsc wiązania wspomnianych białek.

### 1.4.4 Regiony 3'UTR i potencjalny efekt mutacji na ich funkcje

Region 3'UTR to nieulegająca translacji część mRNA położona poniżej sekwencji kodującej białko. Średnia długość regionu 3'UTR u człowieka wynosi ~1030 nt (Pesole et al., 2001).

3'UTR to miejsce występowania sekwencji i wiązania białek odpowiedzialnych za stabilność i lokalizację komórkową transkryptu oraz inicjację translacji. 3'UTR to miejsce występowania sygnału poliadenylacji (ang. *polyadenylation signal*; PAS), który stanowi sekwencja AAUAAA (Beaudoing, Freier, Wyatt, Claverie, & Gautheret, 2000). Sekwencja ta jest sygnałem do przecięcia transkryptu blisko jego 3' końca oraz dołączenia do transkryptu ogona poliA (szeregu nukleotydów adeninowych o długości 50-300 nt). PAS to miejsce wiązania czynnika specyficzności cięcia i poliadenylacji (ang. *cleavage and polyadenylation specific factor*; CPSF) stanowiącego jedno z wielu białek kompleksu biorącego udział w poliadenylacji. Pozostałe białka kompleksu łączą się do sekwencji UGUA występującego

## WSTĘP

powyżej PAS oraz do regionu bogatego w U lub G/U położonego poniżej PAS. Dodatkowo, w 3'UTR mogą występować także miejsca alternatywnej poliadenylacji (ang. *alternative polyadenylation site*; APAS) prowadzące do powstawania izoform transkryptu o różnej długości 3'UTR. Ogon poliA stanowi miejsce wiązania białek PABP (ang. *poly(A)-binding protein*). Wiązanie białka PABP zapobiega deadenylacji ogona poliA, a tym samym degradacji mRNA. Ogon poliA nie tylko stabilizuje transkrypt, ale bierze też udział w inicjacji translacji. Białka PABP oddziałują z czynnikiem inicjacyjnym eIF4G, wskutek czego dochodzi do zbliżenia końca 5' i 3' oraz cyrkularyzacji mRNA. Oddziaływanie to stymuluje wiązanie podjednostki 40S oraz zwiększa wydajność translacji dzięki ponownemu wykorzystaniu podjednostek 40S, które po zakończeniu translacji znajdują się w pobliżu mRNA i jego końca 5'. 3'UTR to region bogaty w nukleotydy AU mogące tworzyć, poza sygnałem poliadenylacji, elementy AREs (ang. *AU-rich elements*) stanowiące sekwencję o długości 40-15 nt posiadającą co najmniej jeden motyw AUUUA otoczony sekwencją bogatą w nukleotydy AU, wiążące białka kierujące transkrypt do degradacji (Schwerk & Savan, 2015). Elementy AREs są charakterystyczne dla około 8-10% ludzkiego transkryptomu.

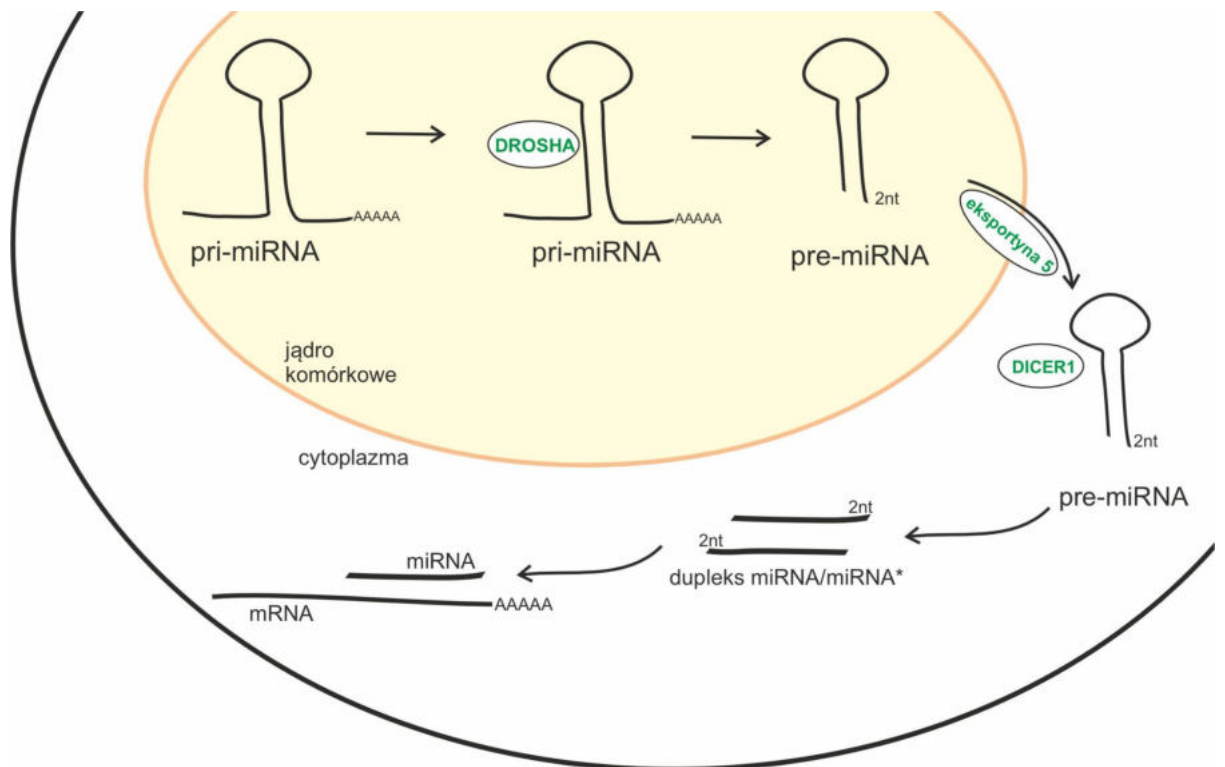
3'UTR stanowi także miejsca wiązania miRNA, cząsteczek regulujących ekspresję genów. Dokładny mechanizm regulacji ekspresji genów przez miRNA opisano w kolejnym rozdziale pracy.

Mutacje w regionie 3'UTR mogą skutkować zmianą miejsca lub brakiem sygnału poliadenylacji co może prowadzić do zmiany długości transkryptu i w konsekwencji wpływać na stabilność i transport mRNA oraz wydajność inicjacji translacji. Dodatkowo mogą powodować brak/osłabienie lub tworzyć nowe miejsca wiązania miRNA (Schuster & Hsieh, 2019; Wu et al., 2018).

### 1.4.5 Geny miRNA i potencjalny efekt mutacji w genach miRNA

MiRNA to krótkie (o długości 21-23 nt), jednoniciowe, niekodujące cząsteczki RNA, biorące udział w regulacji ekspresji genów. Biogeneza miRNA jest procesem wieloetapowym, w który zaangażowanych jest wiele białek. W pierwszym etapie biogenezy transkrypcji ulega pierwotny prekursor pri-miRNA o długości od kilkuset do kilku tysięcy nukleotydów (Saini, Griffiths-Jones, & Enright, 2007). Najistotniejszą częścią pri-miRNA jest fragment tworzący charakterystyczną strukturę typu spinki do włosów, zakończoną pętlą terminalną.

Ta charakterystyczna struktura rozpoznawana jest przez kompleks zwany Mikroprocesorem, którego kluczowym elementem jest rybonukleaza DROSHA (RNaza III) (Han et al., 2006), która przecina dwuniciową helisę RNA, uwalniając w ten sposób około 60 nt strukturę typu spinki do włosów, stanowiącą wtórny prekursor pre-miRNA, pozostawiając na jego końcu 3' 2-nt wolny koniec (ang. *overhang*). Pierwsze etapy biogenezy miRNA zachodzą w jądrze komórkowym. Następnie pre-miRNA zostaje wyeksportowany do cytoplazmy, w czym główną rolę odgrywa białko Exportyna-5. W cytoplazmie pre-miRNA jest cięty przez rybonukleazę DICER1 (RNaza III) działającą w kompleksie z białkami TRBP oraz AGO, co w konsekwencji prowadzi do utworzenia dwuniciowego dupleksu miRNA-miRNA\* (Chendrimada et al., 2005). Jedna nić w dupleksie jest nazywana wiodącą (miRNA), druga zaś nicią pasażerską (miRNA\*). Aktywna postać miRNA to pojedynczo niciowa wiodąca cząsteczka miRNA, generowana w wyniku rozplecenia dupleksu w kompleksie białkowym miRISC (ang. *microRNA induced silencing complex*), w którym ważną rolę odgrywają białka Argonaute (AGO) (Ryc. 2).



**Rycina 2. Schematyczny proces biogenezy miRNA.**

Powstający na terenie jądra komórkowego pri-miRNA ulega cięciu do krótszej cząsteczki pre-miRNA, która jest transportowana do cytoplazmy. Pre-miRNA ulega kolejnemu cięciu, w wyniku którego powstaje dupleks miRNA/miRNA\*. Następnie nie wiodąca bierze udział w potranskrypcyjnym wyciszaniu ekspresji genów.

Dzięki komplementarności miRNA do regionów 3'UTR w mRNA biorą one udział w regulacji ekspresji genów na poziomie potranskrypcyjnym. Niemal wszystkie geny kodujące białka w komórkach ludzkich podlegają regulacji przez miRNA (Berezikov et al., 2005). W przypadku niepełnej komplementarności dojrzałej cząsteczki miRNA do mRNA wyciszenie odbywa się na zasadzie blokowania translacji i jest to przeważający mechanizm regulacji ekspresji genów przez miRNA u ludzi (Zeng, Wagner, & Cullen, 2002). Dużo rzadziej regulacja ekspresji odbywa się przez cięcie i degradację mRNA w przypadku pełnej komplementarności dojrzałej cząsteczki miRNA do docelowego transkryptu. Minimalna sekwencja miRNA, której pełna komplementarność jest wymagana przy rozpoznaniu docelowego mRNA, to region źródłowy (ang. *seed region*). Region źródłowy tworzą nukleotydy w pozycjach 2-8 od końca 5' dojrzałego miRNA.

Mutacje w regionie źródłowym mogą wpływać na interakcje miRNA:sekwencja komplementarna, drastycznie zmieniając pulę potencjalnie regulowanych genów. Mutacje



## WSTĘP

lokalizujące się przy końcach 5' dupleksu miRNA mogą wpływać na wybór nici wiodącej, która jest wybierana na podstawie niższej stabilności końca 5' dupleksu (względem dojrzałego miRNA). Mutacje mogą także lokalizować się wzdłuż sekwencji prekursora, poza regionem źródłowym miRNA i wpływać na biogenezę dojrzałej cząsteczki poprzez dwa kluczowe parametry, które stanowi sekwencja i struktura prekursorów miRNA. Zmiana sekwencji może wpływać na motywy strukturalne biorące udział w interakcji prekursora z licznymi białkami ważnymi dla biogenezy miRNA. Na wiązanie białek do prekursora ma także wpływ budowa i stabilność struktury prekursorów (Ha & Kim, 2014). Zaburzenie struktury może mieć wpływ na wiązanie nukleaz takich jak DROSHA i DICER1 co w konsekwencji może prowadzić do zmiany miejsc (precyzji) oraz efektywności cięcia prekursorów przez te nukleazy (Galka-Marciniak et al., 2019; Urbanek-Trzeciak et al., 2020).

### 1.5 Metody wykrywania mutacji w nowotworach

Mutacje w nowotworach można wykryć z użyciem metod sekwencjonowania DNA. Tradycyjne metody sekwencjonowania opierały się głównie na technologii sekwencjonowania pierwszej generacji - metodzie Sangera (Sanger, Nicklen, & Coulson, 1977). Metoda ta jest szybka i skuteczna przy sekwencjonowaniu krótkich fragmentów DNA o długości do około 1000 pz. W trakcie sekwencjonowania pierwszej generacji możliwe jest jednoczesne sekwencjonowanie wyłącznie produktu pojedynczej reakcji PCR. Analiza nawet pojedynczego genu często wymaga wielu oddzielnych reakcji, co powoduje, że sekwencjonowanie Sangera jest metodą bardzo czasochłonną i nieefektywną przy analizach większej części genomu. W ostatnich latach postęp w sekwencjonowaniu genomu doprowadził do opracowania technologii sekwencjonowania nowej generacji (ang. *next generation sequencing*; NGS) (Morey et al., 2013; Reuter, Spacek, & Snyder, 2015). NGS odnosi się do zbioru technik, które wykorzystują masowe równoległe sekwencjonowanie milionów zwykle krótkich (około 100-500 pz) sekwencji DNA jednocześnie (Kanzi et al., 2020). Sekwencjonowanie nowej generacji pozwala na analizę całego genomu (ang. *Whole Genome Sequencing*; WGS), eksomu (ang. *Whole Exome Sequencing*; WES) lub sekwencjonowanie wybranych regionów/paneli genów (ang. *target sequencing*) w trakcie pojedynczej analizy. W ostatnich latach, dzięki prowadzonym na szeroką skalę międzynarodowym, wielośrodkowym projektom badawczym, zastosowanie technologii sekwencjonowania NGS umożliwiło stworzenie

## WSTĘP

zbiorów danych zawierających mutacje charakterystyczne dla różnych typów nowotworów. W 2006 r. uruchomiono The Cancer Genome Atlas (TCGA; Atlas Genomu Raka) a w 2008 r. International Cancer Genome Consortium (ICGC) jako dwa główne projekty przyspieszające kompleksowe zrozumienie genetyki nowotworów. Celem projektu TCGA było stworzenie kompleksowego „atlasu” profili mutacyjnych nowotworów. TCGA zawiera dane genomiczne, transkryptomiczne, proteomiczne i epigenomiczne oraz dane kliniczne ponad 11 tysięcy próbek nowotworowych, 33 różnych typów nowotworów.

## 2. Rak podstawnokomórkowy skóry (BCC)

BCC stanowi najczęstszy nowotwór złośliwy człowieka. Częstość zachorowania na BCC może wynosić nawet 30% i jest najwyższa w populacji kaukaskiej (białej). BCC wywodzi się z komórek warstwy podstawnej naskórka i należy do grupy nieczerniakowych/niebarwnikowych nowotworów skóry (ang. *nonmelanoma skin cancers*; NMSC) (Pellegrini et al., 2017; Samarasinghe, Madan, & Lear, 2011). BCC jest miejscowo złośliwy, lecz charakteryzuje się niską śmiertelnością i powolnym wzrostem. Ryzyko przerzutów jest bardzo małe i wynosi < 0,1% (Piva de Freitas, Senna, Tabai, Chone, & Altemani, 2017).

Ze względu na cechy morfologiczne guza wyróżnia się kilka podtypów BCC. Najczęstsza postać - guzkowa, charakteryzuje się występowaniem małych, twardych guzków, często z długo niegojącym się owrzodzeniem, które łatwo ulegają uszkodzeniom mechanicznym i lokalizują się głównie na twarzy i szyi. Drugi, najczęściej występujący podtyp stanowi postać powierzchniową, mającą charakter rumieniowy, często pokryta strupami, lokalizująca się głównie na tułowie (Sexton, Jones, & Maloney, 1990). Zarówno podtyp guzkowy jak i powierzchniowy, charakteryzują się powolnym wzrostem i niską inwazyjnością (Crowson, 2006). Do rzadszych podtypów BCC należą podtyp drobnoguzkowy oraz lejkowo-torbielowaty, charakteryzujące się powolnym wzrostem, a także bardziej agresywne podtypy takie jak naciekający i twardzinopodobny.

### 2.1 Czynniki genetyczne i środowiskowe wpływające na rozwój BCC

W ostatnich latach obserwuje się ogólnosiwiatowy wzrost liczby zachorowań na BCC. Mimo licznych badań naukowych, etiologia rozwoju BCC wciąż pozostaje niejasna. Rozwój BCC jest zależny zarówno od uwarunkowania genetycznego jak i narażenia na działanie czynników środowiskowych. BCC może rozwijać się w obrębie skóry całego ciała, jednak najczęściej lokalizuje się na tych częściach, które są najbardziej narażone na działanie promieniowania słonecznego, będącego głównym czynnikiem sprzyjającym powstawaniu BCC. BCC może rozwijać się jako konsekwencja stanów przedrakowych, takich jak rogowacenie starcze/słoneczne (ang. *actinic keratosis*, AK) powstałe na podłożu skóry przewlekłe uszkodzonej promieniowaniem słonecznym (Balcere et al., 2019; Siegal et al., 2016) dlatego głównym środowiskowym czynnikiem ryzyka BCC jest ekspozycja na promieniowanie słoneczne i UV. Może rozwinąć się także jako konsekwencja porażenia promieniowania jonizującego, lub w wyniku rogowacenia chemicznego będącego skutkiem działania arsenu (Hogan, To, Gran, Wong, & Lane, 1989; Yu, Liao, & Chai, 2006). Częstość występowania BCC związana jest ze stopniem pigmentacji skóry. Nowotworzeniu sprzyjają przede wszystkim jasna karnacja, piegi i liczne znamiona barwnikowe. W związku z tym BCC częściej obserwowany jest u przedstawicieli populacji kaukaskiej, rzadziej u ludzi o ciemniejszym zabarwieniu skóry. Wiek pacjenta jest kolejnym czynnikiem wpływającym na narażenie na BCC. BCC najczęściej pojawia się u pacjentów powyżej 65 roku życia, częściej u mężczyzn niż u kobiet (Scrivener, Grosshans, & Cribier, 2002).

BCC występuje zwykle w formie sporadycznej, dużo rzadziej w formie rodzinnej/dziedzicznej jako konsekwencja zespołu Gorlina (ang. *Gorlin syndrome*). W populacji ogólnej zespół Gorlina występuje z częstością około 1:30 000 i charakteryzuje się zmianami nowotworowymi w postaci licznych BCC na powierzchni skóry pacjenta, wraz z nieprawidłowościami układu nerwowego, narządu wzroku i kości, które rozwijają się w młodym wieku (Evans et al., 2010). Zespół Gorlina to choroba dziedziczona w sposób autosomalny dominujący, powodowana przez mutacje w genach szlaku sygnałowego Sonic Hedgehog (Shh), głównie w genie *PTCH1*, rzadziej *SUFU*, *SMO* czy *PTCH2* (Peris et al., 2019). Szlak sygnałowy Shh pełni kluczową rolę również w rozwoju sporadycznej formy BCC (patrz informacje poniżej). W zdrowych komórkach szlak Shh pełni kluczową rolę w procesie embriogenezy. Szlak ten jest mniej aktywny u osób dorosłych, u których odpowiada głównie

## WSTĘP

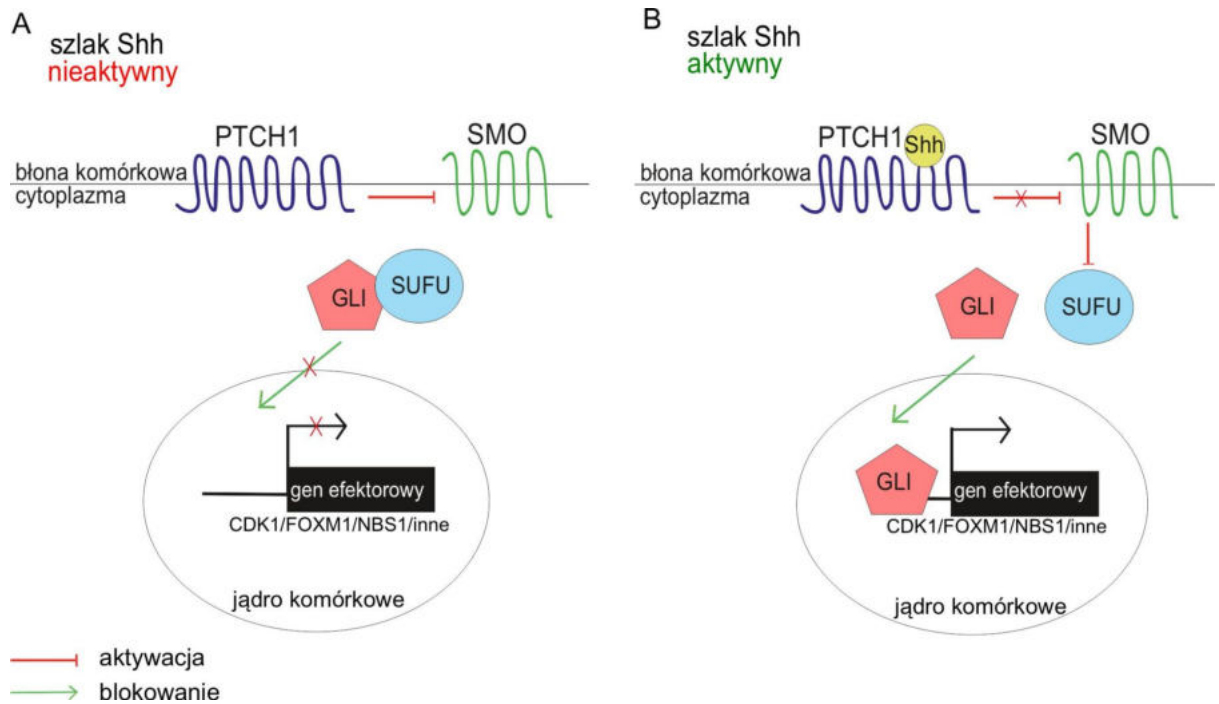
za utrzymanie homeostazy (Mak et al., 2008; Ohba et al., 2008). Zwiększenie aktywności szlaku Shh u dorosłych zachodzi podczas zjawiska regeneracji i naprawy tkanek, a zaburzenia w przebiegu tych procesów mogą prowadzić do procesu nowotworzenia (Petrova & Joyner, 2014).

### **2.1.1 Najczęściej mutowane geny napędzające, kluczowe dla rozwoju BCC**

#### **2.1.1.1 Mutacje somatyczne w genach zaangażowanych w szlak sygnałowy Shh**

Dotychczasowe badania naukowe przeprowadzone z udziałem pacjentów z BCC, wykazały częste mutacje w genach zaangażowanych w szlak sygnałowy Shh, wskazując na jego fundamentalny wpływ na rozwój BCC (Gailani et al., 1996; Pan, Dong, Sun, & Li, 2010). Gen *PTCH1*, kluczowy dla funkcjonowania szlaku Shh jest najczęściej mutowanym genem w BCC (mutacje występują w ponad 70% próbek). Aktywatorem szlaku Shh jest ligand Sonic Hedgehog (SHH), oddziałujący z komórkami docelowymi za pośrednictwem receptora błonowego zbudowanego z dwóch transbłonowych białek: receptora *PTCH1* (ang. *Protein patched homolog 1*), kodowanego przez gen supresorowy *PTCH1*, i związanego z nim białka *SMO* (ang. *Smoothed*) kodowanego przez protoonkogen *SMO*. *PTCH1* ma zdolność odbierania sygnału spoza komórki przez wiązanie ligandu SHH, natomiast przeniesienie sygnału do wnętrza komórki odbywa się z udziałem białka *SMO*. W przypadku braku pobudzenia receptora ligandem SHH, podjednostka *SMO* jest związana i hamowana przez białko *PTCH1* (Ryc. 3A). Natomiast wiązanie SHH z *PTCH1* skutkuje zniesieniem inhibicji białka *SMO* (Ryc. 3B). *SMO* aktywuje wówczas kaskadę sygnałową prowadzącą do aktywacji i translokacji czynników transkrypcyjnych z rodziny GLI (ang. *Glioma-associated oncogene*) do jądra komórkowego (Otsuka i in., 2015). Aktywacja GLI następuje w wyniku zniesienia jego inhibicji przez białko supresorowe *SUFU* (ang. *Suppressor of fused*) (Huang, Wang, Tang, & Luo, 2018). Następnie znajdujące się w jądrze czynniki z rodziny GLI aktywują transkrypcję genów zależnych od SHH związanych m.in. z proliferacją komórki i regulacją cyklu komórkowego (np. *CDK1*, *FOXM1*) oraz procesem naprawy uszkodzeń w DNA (np. *NBS1*) (Avery, Zhang, & Boohaker, 2021; Carballo, Honorato, de Lopes, & Spohr, 2018).

## WSTĘP



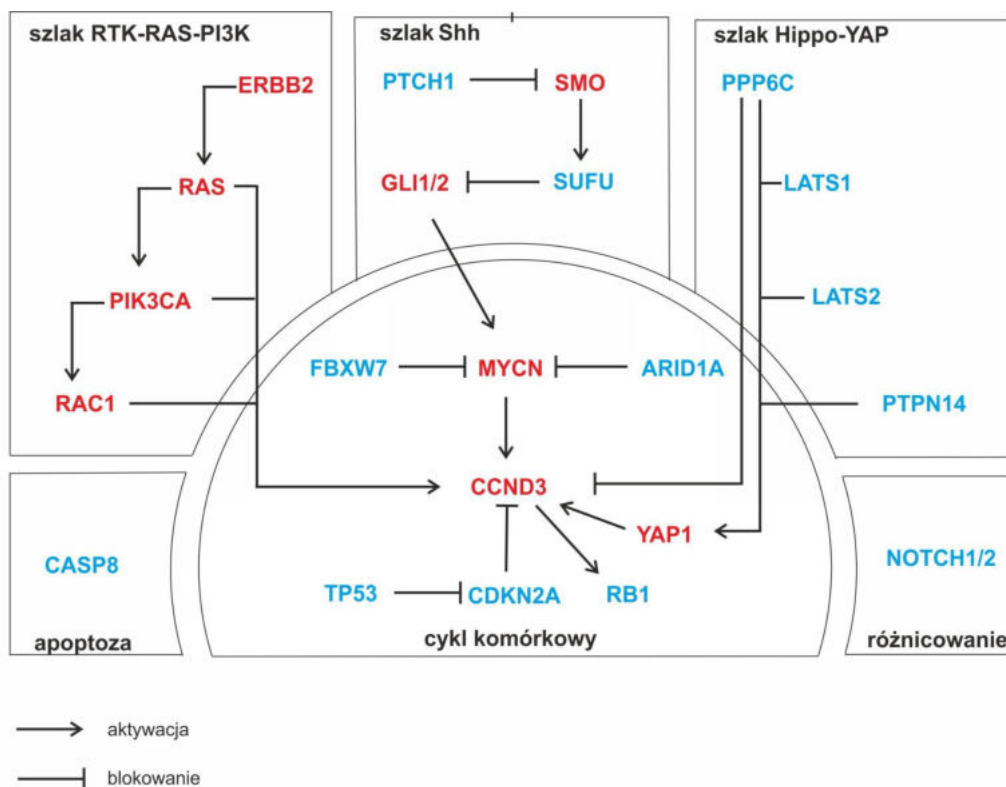
### Rycina 3. Schemat szlaku sygnałowego Shh.

A) nieaktywny szlak Shh - brak ligandu Shh skutkuje zablokowaniem szlaku, białko receptorowe PTCH1 blokuje wówczas białko SMO, co skutkuje zablokowaniem czynników transkrypcyjnych GLI i zablokowaniem transkrypcji genów efektorowych szlaku Shh B) aktywny szlak Shh - związanie ligandu Shh aktywującego kaskadę sygnałową prowadzącą do aktywacji ekspresji genów efektorowych szlaku Shh.

Sekwencjonowanie całego eksomu próbek BCC wykazało, że spośród genów zaangażowanych w szlak Shh najczęściej mutowanym genem jest *PTCH1* (ponad 70%), następnie *SMO* (20%) oraz *SUFU* (8%) (Bonilla et al., 2016; Jayaraman, Rayhan, Hazany, & Kolodney, 2014). Mutacje w genach *PTCH1* oraz *SUFU* mają charakter mutacji utraty funkcji i powodują zniesienie funkcji supresorowej białek kodowanych przez te geny, natomiast mutacje w *SMO* skutkują aktywacją białka SMO mimo braku pobudzenia szlaku przez ligand SHH. Większość mutacji w genie *PTCH1* to mutacje zmiany sensu, mutacje typu nonsens, przesunięcia ramki odczytu, a także mutacje splicingowe. Wykryto także pojedyncze przypadki delecji całego chromosomu 9, na którym zlokalizowany jest gen *PTCH1*, a także sporadyczną utratę heterozygotyczności w genie *PTCH1* (Bonilla et al., 2016; Jayaraman et al., 2014).

### 2.1.1.2 Geny poza szlakiem Shh, często mutowane w BCC

Drugim, zaraz po *PTCH1* najczęściej mutowanym genem, jest *TP53* (mutowany w 61% próbek) (Bonilla et al., 2016), kodujący białko p53, które w zdrowych komórkach odpowiada za aktywację mechanizmów naprawy DNA i stymuluje proces apoptozy (Ryc. 4). Mutacje w tym genie powodują zahamowanie funkcji supresorowej p53, co skutkuje inhibicją procesu apoptozy. Poza *TP53* także inne geny zaangażowane w regulację cyklu komórkowego takie jak: *MYCN* (mutowany w 30%), *RB1* (mutowany w 8%) oraz *ARID1A* (mutacje w 26%) zostały uznane za często mutowane w BCC (Ryc. 4) (Bonilla et al., 2016). Innymi potencjalnymi genami napędzającymi, często mutowanymi w BCC są: *NOTCH1* (40%) oraz *NOTCH2* (26%) zaangażowane w proces różnicowania komórek; *CSMD1* (63%) kodujący transbłonowe białko supresorowe (Di Nardo et al., 2021); geny zaangażowane w szlaki sygnałowe RTK-RAS-PI3K oraz Hippo-YAP, *LATS1* (16%) i *LATS2* (12%) (Bednarski et al., 2021; Bonilla et al., 2016). Natomiast aktywacja szlaku sygnałowego RTK-RAS-PI3K jest charakterystyczna dla złośliwego podtypu BCC (Piérard-Franchimont et al., 2015).



**Rycina 4. Geny często mutowane w BCC.**

Geny posiadające mutacje skutkujące zwiększeniem aktywności genu zaznaczono kolorem czerwonym. W przypadku obniżenia aktywności geny oznaczono kolorem niebieskim.

## WSTĘP

Prawie wszystkie dotychczasowe badania genetyczne BCC obejmowały wyłącznie analizę sekwencji kodujących. Dlatego niemal wszystkie mutacje wykryte dotychczas w BCC to mutacje kodujące (mutacje w częściach genów kodujących białko). Jedyne mutacje niekodujące, które analizowano w BCC to mutacje we wspomnianym wcześniej promotorze genu *TERT* oraz mutacje w promotorze genu *DPH3*. Mutacje w promotorze genu *TERT* zidentyfikowano w 56% do 59% próbek BCC (Di Nardo et al., 2021; Griewank et al., 2013; Maturo et al., 2020) a mutacje w promotorze *DPH3* w 38% do 49% analizowanych próbek BCC (Denisova et al., 2015; Di Nardo et al., 2021; Maturo et al., 2020). Mutacje w promotorze *TERT*, jak wspomniano wyżej, skutkują utworzeniem miejsca wiązania czynników transkrypcyjnych GABP. W przypadku mutacji w *DPH3*, w związku z brakiem jasnej funkcjonalności *DPH3* w procesie nowotworzenia, efekt mutacji zlokalizowanych w promotorze tego genu nie został jak dotąd określony (Denisova et al., 2015; Di Nardo et al., 2021; Maturo et al., 2020).

### 2.1.1.3 miRNA w BCC

Do tej pory badania nad miRNA w BCC skupiały się wyłącznie na analizie zmian ich poziomów. Analiza wykonana przez Heffelfinger i wsp. wykazała różny profil ekspresji miRNA w podtypie guzkowym oraz naciekającym. Dodatkowo, wskazano na supresorową rolę miR-183 (Heffelfinger et al., 2012). Sonkoly i wsp. opisali miR-203 jako supresor nowotworu w mysim modelu BCC. miR-203 występuje na wysokim poziomie w zdrowych komórkach skóry, natomiast w BCC, w wyniku aktywacji szlaku Shh jego poziom jest znacząco obniżony (Sonkoly et al., 2012). Sun i wsp. wykazali obniżenie poziomu *miR-451a* zarówno w ludzkich komórkach BCC jak i w mysim modelu (Sun & Zhang, 2018), co skutkowało zwiększeniem poziomu proliferacji komórek. Natomiast, zwiększenie poziomu *miR-451a* skutkowało zatrzymaniem cyklu komórkowego w fazie G1.

Całogenomowa analiza profilu miRNA w BCC w porównaniu do zdrowej skóry wykonana przez Sand i wsp. pozwoliła na identyfikację 16 miRNA (miR-17, miR-18a, miR-18b, miR-19b, hsa-miR-19b-1-5p, miR-93, miR-106b, miR-125a-5p, miR-130a, miR-181c-5p, miR-181c-3p, miR-181d, miR-182, miR-455-3p, miR-455-5p i miR-542-5p), których poziom był podwyższony w BCC oraz 10 miRNA (miR-29c-3p, miR-29c-5p, miR-139-5p, miR-140-3p, miR-145, miR-378, miR-572, miR-638, miR-2861 i miR-3196), których poziom był obniżony (Sand et al., 2012). miR-130a wycisza ekspresję genu *BCL2* wykazującego działanie

## WSTĘP

antyapoptotyczne, zatem zwiększona poziom miR-130a w BCC może prowadzić do unikania przez nowotwór śmierci w wyniku apoptozy. Kolejny miRNA mogący pełnić ważną funkcję w BCC to miR-29c, którego poziom jest obniżony w BCC. miR-29c obniża ekspresję genów *DNMT3A* i *DNMT3B* kodujących metylotransferazy biorące udział w wyciszeniu ekspresji genów supresorowych w wyniku metylacji. Obniżenie poziomu miR-29c może prowadzić do hipermetylacji genów supresorowych i skutkować obniżeniem ich aktywności. Kolejne analizy przeprowadzone przez Sand i wsp., z udziałem pacjentów z twardzinopodobnym podtypem BCC wykazały 8 miRNA, których poziom był znacząco podwyższony: miR-21, miR-99a, miR-26a, let-7f, let-7g, let-7i, miR-100, and miR-205 (Sand et al., 2019). Większość docelowych cząsteczek, do których wiąże się miR-21 to transkrypty genów supresorowych takich jak *BCL2* (opisany powyżej) i *PTEN* (Meng et al., 2007). *PTEN* to gen kodujący białko supresorowe blokujące aktywność szlaku RAS/PI3K/AKT. Szlak ten może mieć wpływ na aktywność ścieżki sygnałowej Shh przez aktywację czynników z rodziny Gli. Inaktywacja *PTEN* przez wysoki poziom miR-21 może skutkować zatem zwiększoną aktywnością dwóch szlaków sygnałowych kluczowych dla rozwoju BCC (Ming & He, 2009; Riobo, Lu, & Emerson, 2006).

Wei i wsp. wykazali 181 miRNA o zmienionym poziomie w BCC w porównaniu do zdrowej skóry (53 podwyższonych, 128 obniżonych) (Wei, Li, Wang, Song, & Dong, 2021). MiRNA podwyższone w BCC, regulujące ekspresję genów supresorowych ważnych z punktu widzenia rozwoju BCC to miR-320b i miR-30c-1-3p regulujące ekspresję *PTCH1*, oraz miR-3944-5p regulujący *SUFU*. Natomiast miRNA o obniżonym poziomie w BCC hamujące ekspresję genów zwiększających aktywność szlaku Shh to miR-4701-5p oraz miR-3194-3 regulujące ekspresję *SMO* oraz miR-3622a-5p regulujący *GLI1*.



## V. CEL BADAŃ

BCC to najczęstszy nowotwór złośliwy człowieka. Ryzyko zachorowania wśród populacji kaukaskiej wynosi aż 30%. W ostatnich latach obserwuje się ogólnoswiatowy wzrost liczby zachorowań na BCC. Mimo licznych badań, etiologia rozwoju BCC wciąż nie jest do końca wyjaśniona.

Pomimo, że regiony niekodujące stanowią przeważającą część genomu (~98%) oraz są bogate w elementy funkcjonalne, to dotychczas ich analiza była niemal całkowicie pomijana, przez co, z wyjątkiem pojedynczych przypadków, brak informacji na temat nowotworowych mutacji somatycznych w tych regionach. Dotychczas nie przeprowadzono żadnej kompleksowej analizy mutacji niekodujących w BCC.

- **Głównym celem niniejszej pracy była wstępna charakterystyka mutacji somatycznych w BCC w wybranych niekodujących częściach genomu obejmujących niekodujące elementy genów kodujących białka (5'UTR, fragmenty 3'UTR i fragmenty intronów graniczące z eksonami) oraz geny miRNA.**
- Dodatkowym celem pracy była charakterystyka mutacji w regionach kodujących białka, która miała posłużyć przede wszystkim do walidacji uzyskanych wyników poprzez porównanie do wyników wcześniejszych analiz mutacji w tych regionach w BCC.

Realizację powyższych ogólnych celów podjęto w formie następujących zadań badawczych (celów szczegółowych):

- 1) Przeprowadzenie sekwencjonowania typu WES pozwalającego na analizę mutacji zarówno w regionach kodujących białka jak i w regionach niekodujących obejmujących regiony 5'UTR, a także fragmenty 3'UTR oraz intronów;
- 2) Przeprowadzenie sekwencjonowania genów miRNA z zastosowaniem opracowanego w Zakładzie Genetyki Molekularnej ICHB PAN unikatowego systemu do sekwencjonowania całego miRNomu (ang. *Whole miRNome Sequencing*; WMS);
- 3) Ogólna charakterystyka mutacji we wcześniej wymienionych regionach niekodujących oraz eksonach kodujących białka obejmująca analizę typów mutacji, lokalizacji mutacji

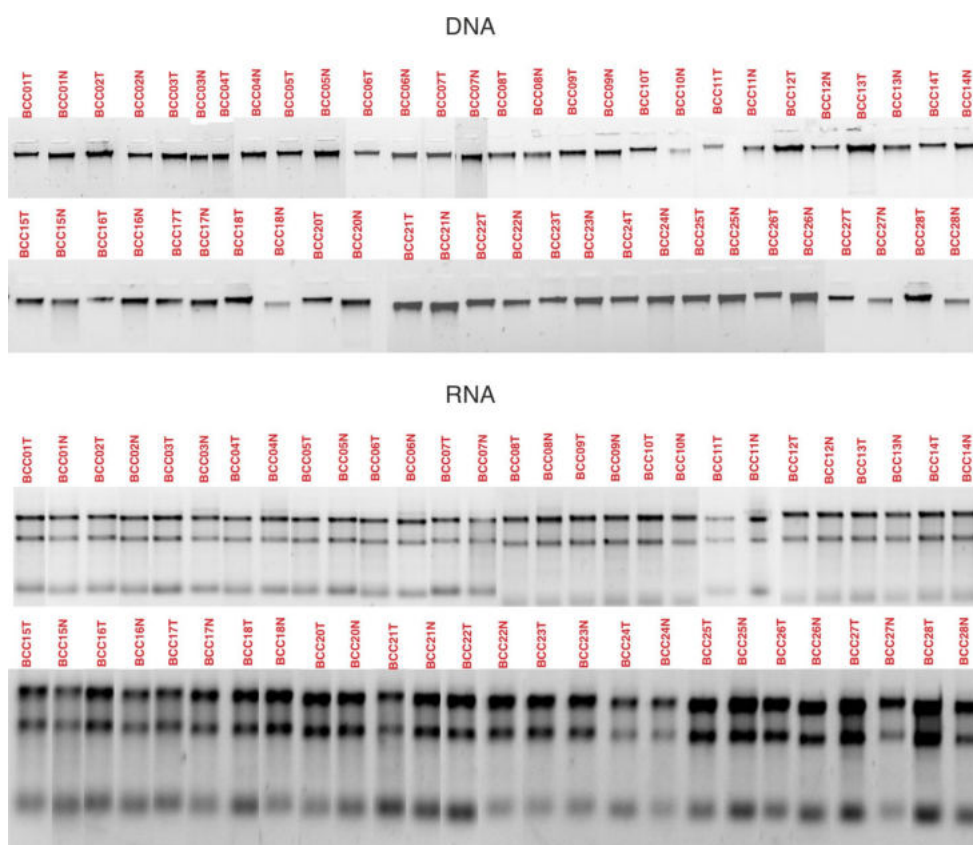
## CEL BADAŃ

w poszczególnych regionach, kalkulację obciążenia nowotworu mutacjami, a także analizę sygnatur nowotworowych;

- 4) Identyfikacja genów często (powtarzalnie) mutowanych w BCC, gorących punktów mutacji oraz regionów niekodujących, w których mutacje są czynnikiem potencjalnie napędzających rozwój BCC (OncodriveFML);
- 5) Wykonanie pierwszej całogenomowej, kompleksowej analizy zmian liczby kopii w BCC i identyfikacja regionów/genów ulegającym częstym delecjom i amplifikacjom (analiza danych WES z użyciem GISTIC2.0; MLPA);
- 6) Wstępna analiza funkcjonalna wybranych mutacji, z użyciem narzędzi bioinformatycznych oraz zewnętrznych zasobów danych (np. baza danych TCGA) gromadzących informacje na temat mutacji somatycznych w innych typach nowotworów (poza BCC).

## VI. WYNIKI

Wyniki przedstawione w niniejszej pracy otrzymano dzięki analizie materiału genetycznego, który stanowił DNA wyizolowany z 27 par próbek (nowotwór oraz zdrowa tkanka) od pacjentów z BCC, w tym 22 z podtypem guzkowym i 5 z podtypem powierzchniowym. Fragment tkanki nieprzekraczający 20 mg poddano homogenizacji, a następnie izolowano DNA oraz całkowity RNA. RNA wyizolowano do celu ewentualnej walidacji lub analizy funkcjonalnych konsekwencji mutacji na poziomie RNA, np. do analizy skutków mutacji splicingowej opisaney w punkcie 1.6.1.1. Jakość i ilość materiału genetycznego określono za pomocą pomiarów z użyciem spektrofotometru NanoDrop One (ThermoScientific) oraz fluorymetru Qubit 3.0 (Invitrogen), a także za pomocą rozdziału elektroforetycznego prób (Ryc. 5).



**Rycina 5. Rozdział elektroforetyczny DNA oraz RNA próbek BCC.**

Rycina zawiera fragmenty kilku żeli agarozowych po rozdziale elektroforetycznym.

## WYNIKI

W celu wykrycia mutacji somatycznych DNA poddano dwóm procedurom sekwencjonowania nowej generacji, WES oraz nowatorskiemu sekwencjonowaniu całego miRNomu (ang. *Whole miRNome Sequencing*; WMS).

Jako że nie istnieją komercyjne zestawy WMS i według aktualnej wiedzy takie sekwencjonowanie nigdy wcześniej nie było wykonywane, panel do WMS został zaprojektowany w Zakładzie Genetyki Molekularnej Instytutu Chemii Bioorganicznej PAN. Panel ten obejmuje niemal 1900 genów miRNA, 28 genów kodujących białka zaangażowane w biogenezę miRNA oraz 10 genów kontrolnych często mutowanych w nowotworach.

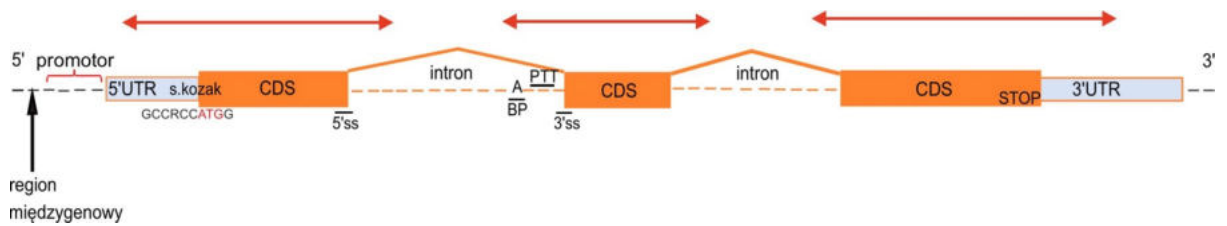
Mutacje wykryte zarówno za pomocą analiz WES jak i WMS przypisano do genomu referencyjnego człowieka Hg38 (GRCh38). Dodatkowo, zgodnie z nomenklaturą Human Genome Variation Society (HGVS) zmiany wykryte w genach kodujących białko oznaczono względem pozycji w sekwencji kodującej (oznaczone literą „c.”), a mutacje wykryte w genach miRNA oznaczono względem odpowiednich transkryptów prekursorów miRNA wg bazy miRBase (oznaczone literą „n.”). W niniejszej pracy sekwencje kodujące białko nazywane są skrótowo sekwencjami kodującymi.

Dalsza analiza otrzymanych wyników obejmowała walidację wybranych mutacji metodą sekwencjonowania Sangera, analizę typów mutacji oraz określenie ich lokalizacji w genach, określenie gorących punktów mutacji i często mutowanych genów, a także wykrycie genów o potencjale napędzającym na podstawie mutacji typu SBS i indeli oraz zmian liczby kopii. Dla wybranych mutacji/genów przeprowadzono także analizę porównawczą zmian wykrytych w BCC oraz w innych typach nowotworów dostępnych w bazie TCGA.

### **1. Sekwencjonowanie całego eksomu (WES)**

W przypadku WES sekwencjonowanie obejmowało zbiór eksonów genów kodujących białka (wraz z eksonami obejmującymi regiony 5'UTR oraz fragmenty regionów 3'UTR (~100 pz)), a także fragmenty intronów sąsiadujące z eksonami (~100 pz) (Ryc. 6).

## WYNIKI



**Rycina 6. Schemat regionów poddanych sekwencjonowaniu WES.**

Kolorem pomarańczowym zaznaczono gen, w skład którego wchodzi: kodujące części eksonów (CDS), niekodujące regiony 5'UTR i 3'UTR, oraz introny. Na początku i końcu regionu CDS zaznaczono kodon start oraz stop. Przestrzeń międzygenowa została oznaczona czarnymi przerywanymi liniami. Symbolem 5' i 3' oznaczono orientację nici kodującej. Regiony poddane sekwencjonowaniu zaznaczono czerwonymi strzałkami.

### 1.1. Analiza podstawowych parametrów sekwencjonowania oraz lokalizacji i typów wykrytych mutacji

W celu otrzymania listy mutacji somatycznych konieczny był wieloetapowy proces analizy danych NGS (szczegóły w sekcji Materiały i Metody). Analiza głębokości sekwencjonowania wykazała średnie pokrycie celowanych regionów na poziomie 183x (185x w próbkach normalnych i 180x w próbkach guza), wahając się w różnych próbkach od 134x do 232x. Średnie pokrycie zmutowanych pozycji wynosi 169x i jest nieco wyższe w sekwencjach kodujących (195x) niż w regionach niekodujących (142x). Finalnie z wykorzystaniem algorytmu MuTect2 wykryto 84 571 mutacji somatycznych. Średni udział zmutowanego allelu (ang. alternative *allele fraction*; AAF) wynosi 0,35 (0,33 w regionach kodujących i 0,40 w regionach niekodujących) (Tabela 1).

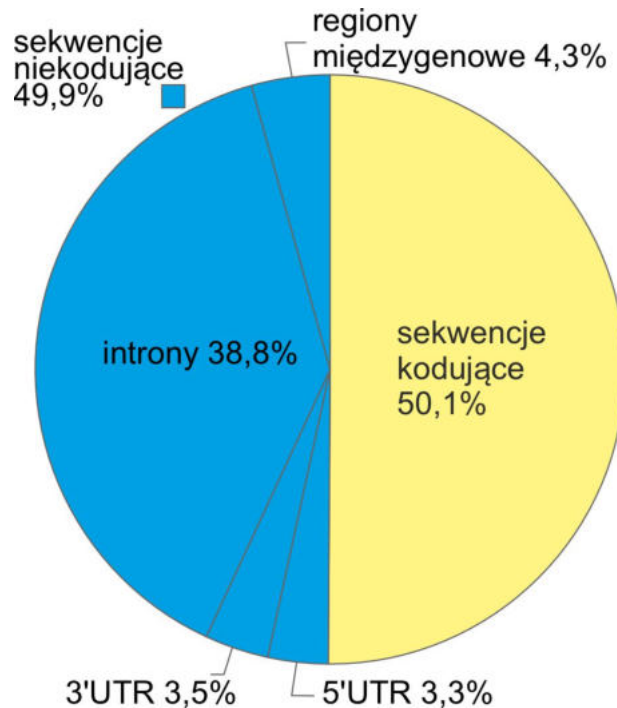
## WYNIKI

**Tabela 1. Rozkład i typy mutacji w BCC.**

Region	Liczba (%) mutacji	Średnie pokrycie	Średnie AAF	Liczba (%) mutacji typu SBS	Liczba (%) mutacji typu DBS	Liczba (%) krótkich indelii
Wszystkie regiony	84 571 (100,0)	169	0,35	79 960 (94,5)	3 128 (3,7)	1 483 (1,8)
Sekwencje kodujące	42 380 (50,1)	195	0,33	40 503 (95,6)	1 697 (4,0)	180 (0,4)
Region niekodujący	42 400 (49,9)	142	0,37	39 457 (93,5)	1 431 (3,4)	1 303 (3,1)
Introny	32 805 (38,8)	136	0,38	30 655 (93,4)	1 076 (3,3)	1 074 (3,3)
3'UTR	2 926 (3,5)	154	0,39	2 721 (93,0)	100 (3,4)	105 (3,6)
5'UTR	2 832 (3,3)	163	0,37	2 650 (93,6)	125 (4,4)	57 (2,0)
Regiony międzygenowe	3 628 (4,3)	167	0,35	3 431 (94,6)	130 (3,6)	67 (1,8)

42 380 (50,1%) mutacji jest zlokalizowanych w regionach kodujących białka, a pozostałe 42 191 (49,9%) w regionach niekodujących. Wśród mutacji w regionach niekodujących większość stanowią warianty intronowe (38,8%). Mutacje w regionie 5'UTR oraz 3'UTR stanowią odpowiednio 3,5% oraz 3,3% wszystkich mutacji. Wykryto także 4,3% mutacji w sekwencjach innych niż sklasyfikowane powyżej, obejmujące głównie regiony międzygenowe (Ryc. 7).

## WYNIKI



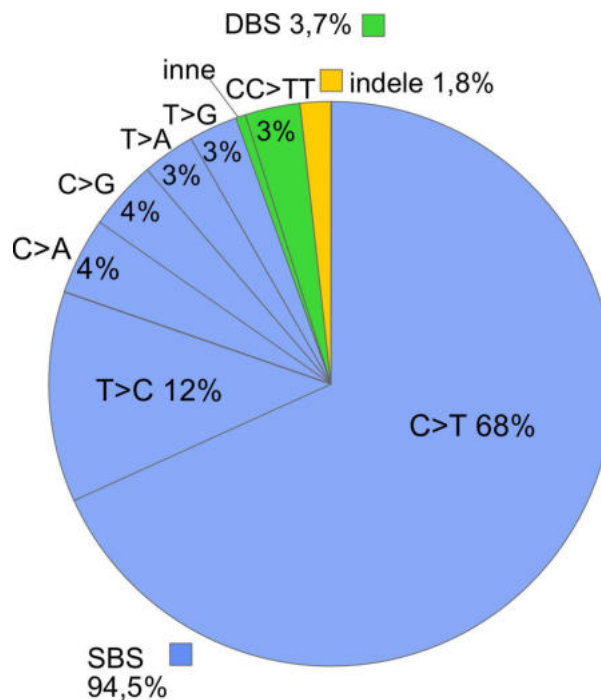
**Rycina 7. Lokalizacja mutacji w BCC.**

Żółtym kolorem zaznaczono mutacje zlokalizowane w części kodującej genów. Na niebiesko zaznaczono mutacje lokalizujące się w części niekodującej genomu obejmującej introny, 5'UTR, 3' UTR oraz regiony międzygenowe.

Zdecydowaną większość wszystkich wykrytych mutacji stanowią mutacje typu SBS (79 960 (94,5%)), następnie DBS (3 128 (3,7%)), i krótkie (<4 nt) indel (1 483 (1,8%)) (Tabela 1; Ryc. 8). Liczba mutacji typu SBS w sekwencjach kodujących (95,6%) i niekodujących (93,5%) jest zbliżona. Natomiast wyższa częstotliwość indeli w regionach niekodujących (3,1%) w stosunku do kodujących (0,4%) najprawdopodobniej wynika z wysokiej ilości w tych regionach sekwencji o niskiej złożoności (ang. *low-complexity regions*) i sekwencji repetytywnych, które powodują błędy polimerazy w trakcie replikacji w wyniku tzw. poślizgu replikacji.

Szczegółowa analiza mutacji SBS wykazała największy udział tranzycji C>T (68,0%) wśród wszystkich wykrytych mutacji. Najliczniejszą grupę spośród DBS stanowią zmiany CC>TT (3,0%) (Ryc. 8). Oba typy mutacji są charakterystyczne dla nowotworów, których rozwój jest związany z działaniem promieniowania słonecznego.

## WYNIKI



**Rycina 8. Typy mutacji występujące w BCC.**

Kolorem niebieskim oznaczono mutacje typu: SBS, zielonym DBS, a żółtym indele.

### 1.2 Walidacja mutacji metodą sekwencjonowania Sangera - estymacja częstości mutacji fałszywie pozytywnych

W celu ustalenia ilości mutacji fałszywie pozytywnych wybrane mutacje wykryte z użyciem NGS zostały poddane walidacji przez powtórne sekwencjonowanie z zastosowaniem metody Sangera. Analizie poddano 52 mutacje reprezentujące różne typy mutacji: 39 substytucji (SBS oraz DBS) oraz 13 indeli. W wyniku przeprowadzonej analizy potwierdzonych zostało 51 mutacji. Załedwie 2% mutacji (1 mutacja) to zmiany fałszywie pozytywne (Tabela 2). Niepotwierdzona mutacja (CC>TT w *MYCN*) charakteryzuje się bardzo niskim AAF=0,07 (poniżej zwykle przyjmowanego progu wykrywalności metodą sekwencjonowania Sangera) w związku z czym faktyczny poziom mutacji fałszywie pozytywnych może być jeszcze niższy niż stwierdzone 2%.



## WYNIKI

**Tabela 2. Lista mutacji potwierdzanych z użyciem sekwencjonowania metodą Sangera.**

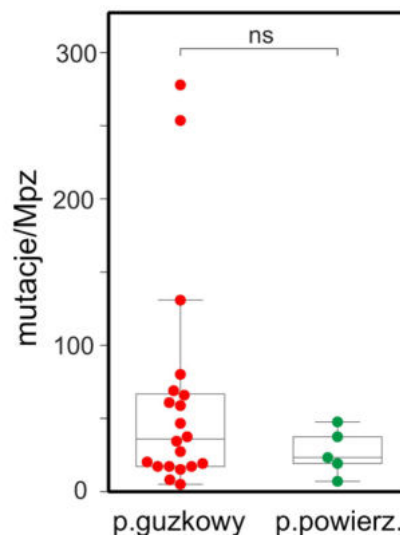
Rodzaj mutacji	Nazwa genu	Lokalizacja mutacji	Rodzaj mutacji	AAF	Liczba analizowanych próbek	Walidacja
Substytucje	<i>SCNM1</i>	chr1:151,168,176	C>T	0,09	1	Tak
	<i>FLG2</i>	chr1:152,350,960	G>A	0,29	1	Tak
	<i>FLG2</i>	chr1:152,353,014	C>T	0,26	1	Tak
	<i>FLG2</i>	chr1:152,354,686	G>A	0,32; 0,35	2	Tak
	<i>RPS27</i>	chr1:153,990,763	C>T	0,17; 0,38; 0,46	3	Tak
	<i>DYRK3</i>	chr1:206,648,309	C>T	0,07; 0,09	2	Tak
	<i>MYCN</i>	chr2:15,942,194	CC>TT	0,18; 0,37	2	Tak
	<i>MYCN</i>	chr2:15,942,195	C>T	0,19; 0,23	2	Tak
	<i>MYCN</i>	chr2:15,942,240	C>T	0,38	1	Tak
	<i>MYCN</i>	chr2:15,945,676	C>T	0,38	1	Tak
	<i>MYCN</i>	chr2:15,945,997	CC>TT	0,07	1	Nie
	<i>PTCH1</i>	chr9:95,459,622	G>T	0,61; 0,77	2	Tak
	<i>PTCH1</i>	chr9:95,469,014	G>A	0,11; 0,13	2	Tak
	<i>PTCH1</i>	chr9:95,480,502	C>T	0,82	1	Tak
	<i>AMBRA1</i>	chr11:46,433,601	G>A	0,05	1	Tak
	<i>AMBRA1</i>	chr11:46,435,033	GG>AA	0,36	1	Tak
	<i>BAD</i>	chr11:64,270,066	G>T	0,2	1	Tak
	<i>BAD</i>	chr11:64,270,066	G>A	0,05; 0,07; 0,23; 0,32	4	Tak
	<i>BAD</i>	chr11:64,270,066	GG>AA	0,06; 0,4	2	Tak
	<i>BAD</i>	chr11:64,270,067	G>A	0,08	1	Tak
<i>DHODH</i>	chr16:72,008,760	GG>AA	0,11; 0,19; 0,25; 0,46	4	Tak	

WYNIKI

	<i>DHODH</i>	chr16:72,008,760	G>A	0,08	1	Tak
	<i>TP53</i>	chr17:7,676,071	G>A	0,29	1	Tak
	<i>MISP</i>	chr19:757,463	C>T	0,11	1	Tak
Indele	<i>CASZ1</i>	chr1:10,659,766	A>AT	0,22	1	Tak
	<i>TCHH</i>	chr1:152,110,315	GC>G	0,28	1	Tak
	<i>BIRC6</i>	chr2:32,464,744	AGC>A	0,25	1	Tak
	<i>TTN</i>	chr2:178,536,516	CA>C	0,32	1	Tak
	<i>SCAP</i>	chr3:47,413,919	C>CA	0,24	1	Tak
	<i>HERC5</i>	chr4:88,463,980	TG>T	0,28	1	Tak
	<i>KCNN2</i>	chr5:114,473,131	C>CA	0,41	1	Tak
	<i>IBTK</i>	chr6:82,227,236	CA>C	0,18	1	Tak
	<i>ADGB</i>	chr6:146,672,360	GA>G	0,32	1	Tak
	<i>SSPO</i>	chr7:149,779,582	C>CT	0,38	1	Tak
	<i>PTCH1</i>	chr9:95,477,628	GAC>G	0,33	1	Tak
	<i>ALG10B</i>	chr12:38,320,258	CTA>C	0,3	1	Tak
	<i>FRMPD3</i>	chrX:107,598,126	AC>A	0,34	1	Tak

### 1.3 Obciążenie mutacyjne (TMB) w BCC

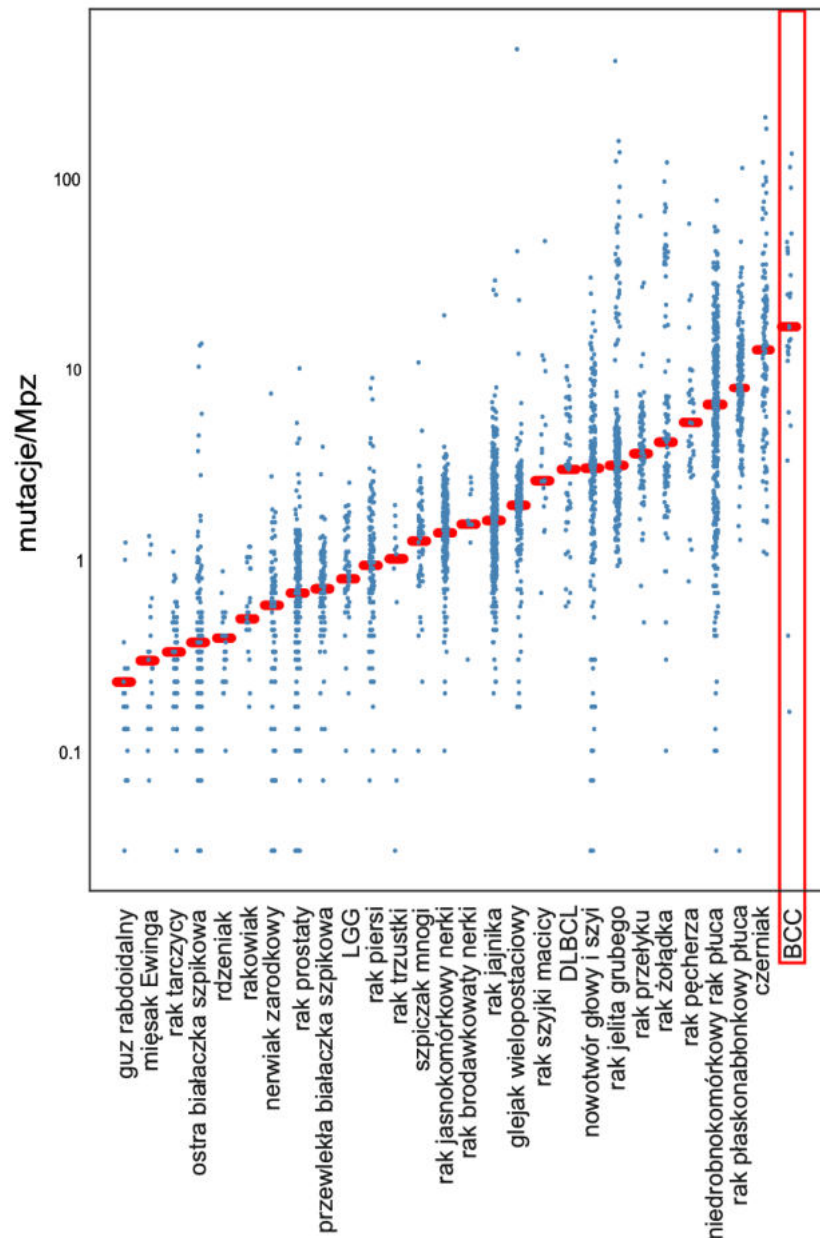
TMB to parametr określający ilość mutacji w przeliczeniu na wielkość analizowanego regionu (podaną w Mpz). W niniejszej pracy średnia wielkość parametru TMB obliczona na podstawie mutacji lokalizujących się w regionach kodujących genów wynosi 52,8 mutacje/Mpz (w przedziale od 0,1 do 287,5). Podtyp guzkowy i powierzchniowy nie różnią się pod względem obciążenia mutacjami (Ryc. 9;  $p = 0,74$ , test Manna-Whitneya-Wilcoxona; dla podtypu guzkowego - mediana 31,5, dla podtypu powierzchniowego - mediana 25,9).



**Rycina 9. Porównanie obciążenia mutacyjnego pomiędzy podtypami BCC.**

Na osi y przedstawiono wartość parametru TMB (oznaczoną kropką dla każdej analizowanej próbki nowotworowej) obliczonego na podstawie mutacji niesynonimicznych lokalizujących się w części kodującej genów.

W celu porównania TMB w BCC z wartościami tego parametru dla innych typów nowotworów wykorzystano wyniki pracy Lawrence et al., 2013 zawierającej dane dla 27 typów nowotworów (bez BCC). Jako że w pracy Lawrence et al. 2013 obliczano parametr TMB z pominięciem mutacji synonimicznych, przed porównaniem przeliczono wartości TMB dla próbek BCC również według tej zasady. Wówczas średnia wartość TMB wynosiła 34,1 mutacje/Mpz (w przedziale od 0,1 do 166,3). Bezpośrednie porównanie wartości TMB w BCC i innych nowotworach wykazało, że TMB w BCC znacząco przewyższa wysokość tego parametru we wszystkich innych typach nowotworów (Ryc. 10).

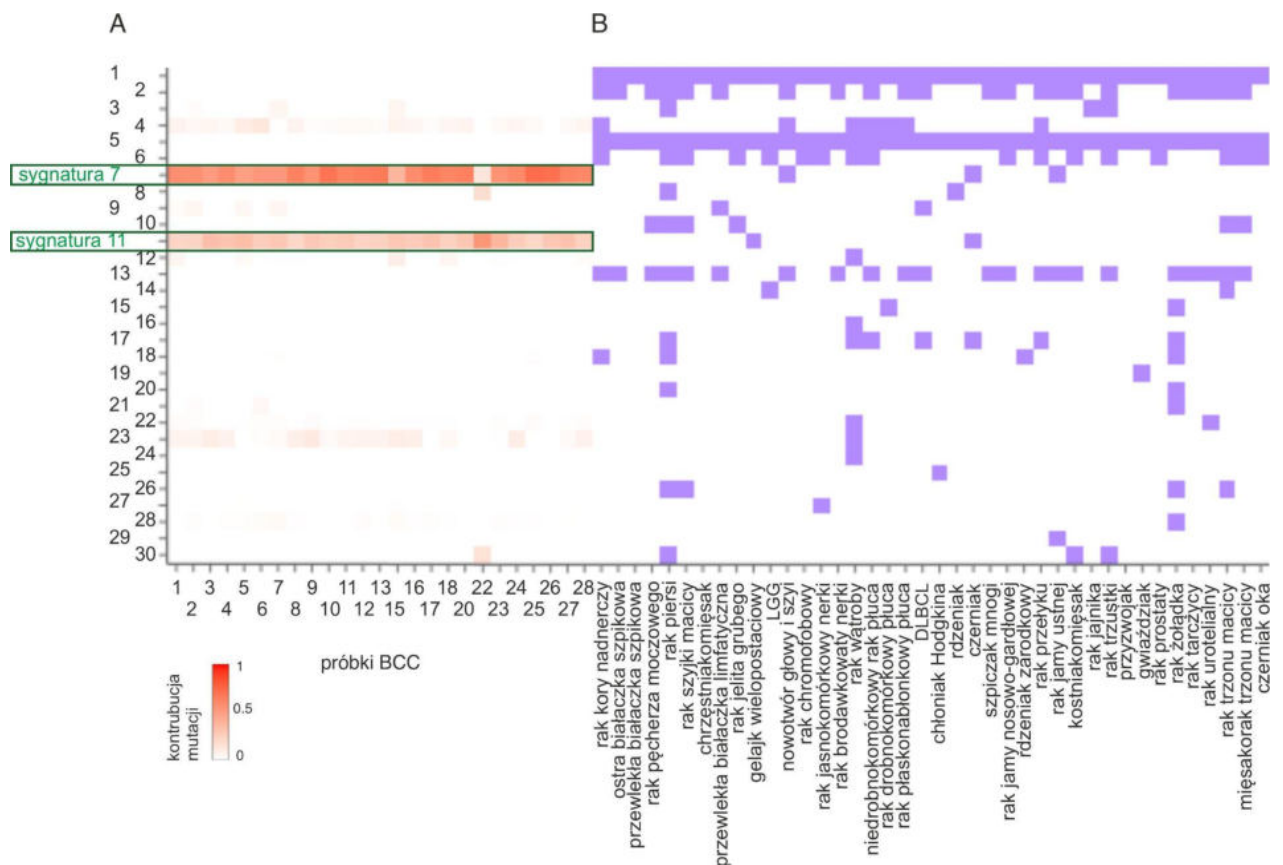


**Rycina 10. Obciążenie mutacyjne w nowotworach.**

Na osi y przedstawiono (w skali logarytmicznej) wartość parametru TMB (oznaczoną kropką dla każdej analizowanej próbki) w 28 typach nowotworów zaznaczonych na osi x i uszeregowanych rosnąco według wartości mediany (oznaczonej czerwoną linią). Rozwinięcia wprowadzonych skrótów nowotworów: LGG (ang. low grade glioma) - nowotwory pochodzenia glejowego o niskim stopniu złośliwości; DLBCL (ang. diffuse large B-cell lymphoma) - chłoniaki rozlane z dużych komórek B.

### 1.4 Sygnatury mutacyjne w BCC

W celu charakterystyki procesów odgrywających rolę w mutageniezie BCC przeprowadzono analizę przypisania rozkładu mutacji do sygnatur, w tym celu zastosowano narzędzie MuSiCa (Díaz-Gay et al., 2018). Podobieństwo między otrzymanym a referencyjnym (wg COSMIC) rozkładem określono za pomocą parametru ASC (ang. *average signature contribution*) określającego średnią kontrybucję rozkładu mutacji do poszczególnych sygnatur we wszystkich analizowanych próbkach BCC. Analiza wykazała, że większość próbek BCC posiada wzór mutacji związany głównie z sygnaturą 7 (ASC = 0,7) oraz w mniejszym stopniu z sygnaturą 11 (ASC = 0,2) (Ryc. 11A).

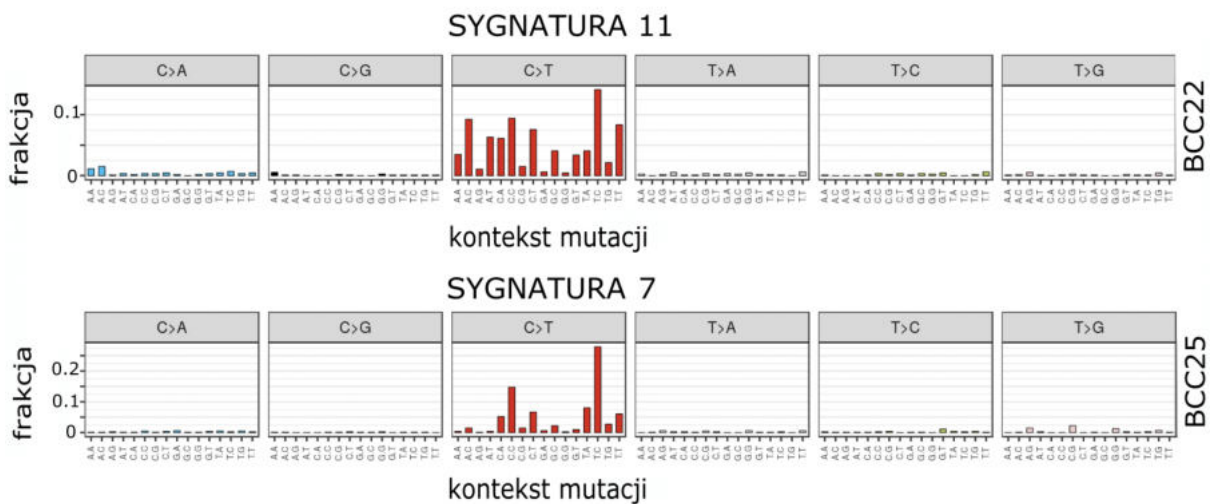


**Rycina 11. Sygnatury mutacyjne w BCC oraz innych typach nowotworów.**

Na osi y zaznaczono sygnatury mutacyjne. Zielonym kolorem wyróżniono sygnatury najczęstsze w badanych próbkach BCC. A) Mapa ciepła kontrybucji mutacji wykrytych w próbkach BCC (zaznaczonych na osi x) do poszczególnych sygnatur. Wyższa intensywność koloru pomarańczowego oznacza wyższą kontrybucję mutacji. B) Sygnatury charakterystyczne dla nowotworów innych niż BCC (oznaczone kolorem fioletowym). Rozwinięcia zastosowanych skrótów: LGG (ang. low grade glioma) - nowotwory pochodzenia glejowego o niskim stopniu złośliwości; DLBCL (ang. diffuse large B-cell lymphoma) - chłoniaki rozlane z dużych komórek B.

## WYNIKI

Sygnatura 7 jest charakterystyczna dla czerniaka oraz raka głowy i szyi. Sygnatura 11 często występuje w czerniaku i glejaku wielopostaciowym (Ryc. 11B), głównie u pacjentów leczonych środkiem alkilującym - temozolomidem, który jest również stosowany w terapii BCC. Obie sygnatury charakteryzują się występowaniem głównie substytucji C>T, ale różnią się kontekstem sekwencji, w obrębie której występuje mutacja (Ryc. 12). Sygnatura 7 charakteryzuje się występowaniem mutacji C>T głównie w kontekście nukleotydowym TCC i związana jest z ekspozycją na promieniowanie UV. W przypadku sygnatury 11 mutacje C>T występują głównie w kontekście nukleotydowym TCC oraz CCC i ACC. Tylko jedna próbka (BCC22) wykazała silniejszy związek z sygnaturą 11 (ASC = 0,6) niż z sygnaturą 7 (ASC = 0,3) (Ryc. 12). Rozkład mutacji dla wszystkich próbek BCC przedstawiono w Załączniku 1.



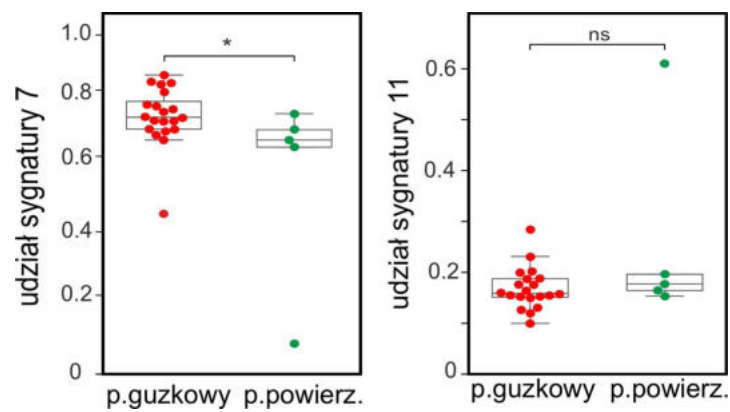
**Rycina 12. Sygnatury mutacyjne charakterystyczne dla BCC.**

Na osi y przedstawiono częstość występowania mutacji w kontekście mutacyjnym oznaczonym na osi x. Przedstawiono rozkład mutacji dwóch próbek BCC, których wzór mutacyjny jest najbardziej zbliżony do rozkładu charakterystycznego dla sygnatury: 11 (na przykładzie próbki BCC22) oraz 7 (na przykładzie próbki BCC25).

Żadna z analizowanych próbek BCC nie wykazała związku z sygnaturami 1, 2, 5 i 13, które występują bardzo często w większości typów nowotworów (Ryc. 11B). Może to sugerować, że deaminacja 5-metylocytozyny (5mC) indukowana głównie przez deaminazy cytydynowe AID/APOBEC (przypisywane powyższym sygnaturom) nie odgrywa roli w patogenezie BCC.

## WYNIKI

Porównanie sygnatur dla próbek BCC z podtypu guzkowego i powierzchniowego wykazało istotną różnicę pod względem ASC do sygnatury 7. Kontrybucja ta była wyższa dla próbek podtypu guzkowego w porównaniu z podtypem powierzchniowym ( $p = 0,03$ ; test Manna-Whitneya-Wilcoxona; Ryc. 13). Przyczyną może być zwiększona ekspozycja na promieniowanie słoneczne części ciała, na których rozwija się podtyp guzkowy (twarz, głowa, szyja). W przypadku sygnatury 11 analiza nie wykazała różnicy pod względem kontrybucji mutacji w podtypie guzkowym i powierzchniowym ( $p = 0,2$ ; test Manna-Whitneya-Wilcoxona; Ryc. 13).



**Rycina 13. Kontrybucja mutacji do sygnatury 7 oraz 11 w podtypach BCC.**

Na osi y przedstawiono kontrybucję mutacji do sygnatury 7 oraz 11 (kropką oznaczono wartość kontrybucji dla każdej analizowanej próbki) dla dwóch podtypów BCC zaznaczonych na osi x. Gwiazdką oznaczono istotną statystycznie różnicę między analizowanymi grupami, ns- różnica nieistotna statystycznie.

### 1.5 Analiza gorących punktów mutacji w BCC

Powtarzalne występowanie mutacji w określonych pozycjach w wielu próbkach może być nieprzypadkowe oraz sugerować funkcjonalny wpływ mutacji i być wskaźnikiem kancerogennej roli zmutowanych genów. W niniejszej pracy doktorskiej za gorące punkty mutacji uznano pozycje genomowe zmutowane w co najmniej 3 próbkach BCC (>10% analizowanych próbek). W sumie na podstawie analiz WES zidentyfikowano 43 gorące punkty mutacji, w tym 23 zlokalizowane w sekwencjach kodujących genów a 20 w regionach niekodujących (8 w 5'UTR, 1 w 3'UTR oraz 11 w intronach). W przypadku mutacji występujących w regionie kodującym sprawdzono wpływ mutacji na zmianę sekwencji białka (Tabela 3).

## WYNIKI

**Tabela 3. Gorące punkty mutacji wykryte w BCC.**

Pozycja genomowa gorącego punktu	Rodzaj mutacji	Lokalizacja/Nazwa genu	Liczba (%) próbek z mutacją w gorącym punkcie	Wpływ na sekwencję białka
chr11:64,270,066-64,270,067	G>A; GG>A; G>T	3'UTR/ <i>BAD</i>	8 (29,6)	-
chr14:103,131,144-103,131,145	C>T; CC>TT	CDS/ <i>TNFAIP2</i>	6 (22,2)	mutacja zmiany sensu, Ser431Phe, c.1292C>T, c.1292_1293delinsTT; domena Sec6; mutacja synonimiczna, Ser431Ser, c.1293C>T
chr15:57,706,717	C>T	5'UTR; intron/ <i>POLR2M</i> ; <i>GCOM1</i>	5 (18,5)	-
chr16:72,008,760-72,008,761	G>A; GG>AA	5'UTR/ <i>DHODH</i>	5 (18,5)	-
chr1:19,596,899	G>A	Intron/ <i>MICOS10</i>	4 (14,8)	-
chr4:47,463,639	C>T	CDS/ <i>COMMD8</i>	4 (14,8)	mutacja zmiany sensu, Glu5Lys, c.14C>T
chr7:56,106,490	G>A	5'UTR/ <i>CHCHD2</i>	4 (14,8)	-
chr9:32,552,281	G>A	Intron; 5'UTR/ <i>TOPORS</i> ; <i>SMIM27</i>	4 (14,8)	-
chr12:110,450,362-110,450,363	G>A; GG>AA	5'UTR/ <i>ARPC3</i>	4 (14,8)	-
chr16:2,460,094-2,460,095	G>A; GG>AA	5'UTR/ <i>TEDC2</i>	4 (14,8)	-
chr18:23,586,506-23,586,507	G>A; GG>AA	5'UTR/ <i>NPC1</i>	4 (14,8)	-



## WYNIKI

chr19:3,788,7566	C>T	CDS/ <i>WDR87</i>	4 (14,8)	mutacja synonimiczna, Arg2035Arg, c.6105C>T
chr1:20,090,617	C>T	CDS/ <i>PLA2G5</i>	3 (11,1)	mutacja synonimiczna, Leu114Leu, c.342C>T
chr1:99,306,852	G>A	CDS/ <i>PLPPR4</i>	3 (11,1)	mutacja zmiany sensu, Glu606Lys, c.1816G>A
chr1:103,432,947	C>T	CDS/ <i>LOC101928436</i>	3 (11,1)	mutacja zmiany sensu, Gly39Glu, c.116C>T
chr1:153,990,763	C>T	5'UTR/ <i>RPS27</i>	3 (11,1)	-
chr1:155,965,725	G>A	CDS/ <i>ARHGEF2</i>	3 (11,1)	mutacja zmiany sensu, Pro99Ser, c.295G>A
chr2:28,628,556	C>T	Intron/ <i>PLB1</i>	3 (11,1)	-
chr2:33,397,321	C>T	Intron/ <i>LTBP1</i>	3 (11,1)	-
chr3:26,710,432	G>A	CDS/ <i>LRR3B</i>	3 (11,1)	mutacja zmiany sensu, Asp254Asn, c.760G>A
chr3:121,928,459	C>T	CDS/ <i>SLC15A2</i>	3 (11,1)	mutacja synonimiczna, Phe415Phe, c.1245C>T
chr4:68,069,918	C>T	Intron/ <i>UBA6-AS1</i>	3 (11,1)	-
chr5:140,863,551	G>A	Intron/ <i>PCDHA3</i>	3 (11,1)	-
chr6:36,961,899	G>A	CDS/ <i>PI16</i>	3 (11,1)	mutacja zmiany sensu, Gly173Arg, c.517G>A
chr6:69,339,554	C>T	Intron/ <i>BAI3</i>	3 (11,1)	-
chr6:118,565,500- 118,565,501	G>A; GG>AA	Intron/ <i>CEP85L</i>	3 (11,1)	-
chr6:165,301,652	C>T	CDS/ <i>Q8TC11</i> ; <i>C6orf118</i>	3 (11,1)	mutacja zmiany sensu, Glu224Lys, c.670C>T

## WYNIKI

chr7:143,935,997	G>A	CDS/ <i>OR2F2</i>	3 (11,1)	mutacja synonimiczna, Thr255Thr, c.765G>A
chr7:148,827,237	G>A	CDS/ <i>EZH2</i>	3 (11,1)	mutacja zmiany sensu, Pro219Ser, c.655G>A
chr9:8,451,932	C>T	Intron/ <i>PTPRD</i>	3 (11,1)	-
chr9:136,979,271	C>T	CDS/ <i>PTGDS</i>	3 (11,1)	mutacja synonimiczna, Ser101Ser, c.303C>T
chr10:67,097,691	G>A	CDS/ <i>LRRTM3</i>	3 (11,1)	mutacja synonimiczna, Thr547Thr, c.1641G>A
chr12:10,160,032	G>A	Intron/ <i>OLR1</i>	3 (11,1)	-
chr13:36,695,128	G>A	CDS/ <i>SERTM1</i>	3 (11,1)	mutacja zmiany sensu, Gly17Glu, c.50G>A
chr15:24,676,062	C>T	CDS/ <i>NPAP1</i>	3 (11,1)	mutacja synonimiczna, Phe65Phe, c.195C>T
chr15:40,382,906- 40,382,907	C>T; CC>TT	CDS/ <i>KNSTRN</i>	3 (11,1)	mutacja zmiany sensu, Ser24Phe, c.71C>T, c.71_72delinsTT
chr15:50,482,094	G>C/A	Intron/ <i>USP8</i>	3 (11,1)	-
chr15:74,615,846	G>A	CDS/ <i>CLK3</i>	3 (11,1)	mutacja zmiany sensu, Arg131Lys, c.392G>A
chr16:4,813,725	G>A	Intron/ <i>GLYR1</i>	3 (11,1)	-
chr17:35,441,269	G>A	CDS/ <i>SLFN13</i>	3 (11,1)	mutacja zmiany sensu, Arg674Cys, c.2020G>A
chr17:44,355,031	G>A	CDS/ <i>FAM171A2</i>	3 (11,1)	-
chr19:4,513,208	C>T	CDS/ <i>PLIN4</i>	3 (11,1)	mutacja zmiany sensu, Gly252Val, c.755C>T
chrX:135,868,028	G>A	CDS/ <i>CT45A9</i>	3 (11,1)	mutacja zmiany sensu, Pro12Ser, c.34G>A

## WYNIKI

Niektóre gorące miejsca mutacji znajdują się w genach zdefiniowanych w bazie COSMIC Cancer Gene Census (CGC) jako związanych z procesem nowotworzenia. Jednym z nich jest *EZH2* posiadający gorący punkt w regionie kodującym (szerzej opisany w punkcie 1.5.1.2 niniejszej pracy doktorskiej). Kolejny gen obecny w bazie CGC to *USP8* kodujący enzym biorący udział w deubikwitynacji onkogenów (Islam, Chen, & Chen, 2021) i posiadający gorący punkt w sekwencji intronowej.

Ze względu na sporą liczbę genów posiadających gorące punkty mutacji (43 geny), w kolejnych akapitach pracy doktorskiej skupiono się tylko na tych genach, których funkcja w BCC lub innych nowotworach została najlepiej poznana bazując na przeglądzie literaturowym.

### **1.5.1 Gorące punkty mutacji w regionach kodujących białko**

Spośród 23 gorących punktów mutacji zlokalizowanych w regionach kodujących genów 15 stanowiły gorące miejsca posiadające mutacje zmiany sensu, a 7 mutacje synonimiczne.

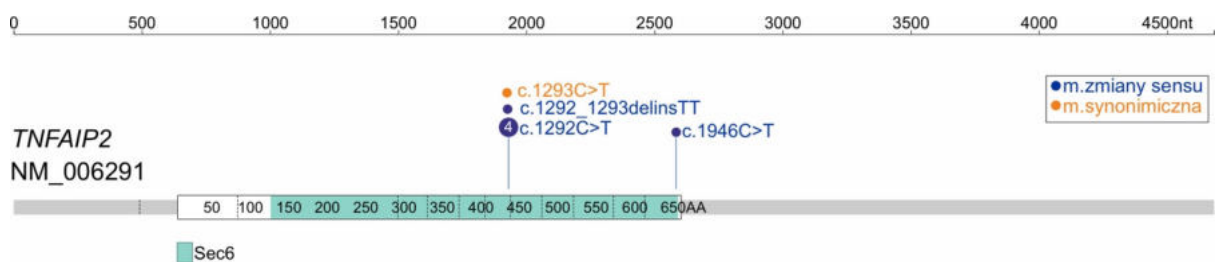
Ponieważ większość mutacji synonimicznych wynika z losowo występujących neutralnych funkcjonalnie zmian, oraz wykryte mutacje synonimiczne nie były zlokalizowane w miejscach ważnych z punktu widzenia splicingu (eksonowe sekwencje wzmacniające i wyciszające splicing, miejsca donorowe i akceptorowe splicingu; analiza z użyciem ESEfinder (Cartegni, Wang, Zhu, Zhang, & Krainer, 2003) w dalszym etapie pracy skupiono się na analizie gorących punktów mutacji posiadających mutacje zmiany sensu.

W celu zobrazowania lokalizacji gorących punktów mutacji, zmiany naniesiono na kanoniczne transkrypty poszczególnych genów. Na rycinach przedstawionych w poniższych akapitach oprócz gorących punktów mutacji, zobrazowano także pozostałe mutacje zlokalizowane w sekwencji kodującej, w regionach 5'UTR oraz 3'UTR a także w intronach (w przypadku, gdy mutacje występowały do 40 nt od granicy ekson/intron).

## WYNIKI

### 1.5.1.1 Gorący punkt mutacji w genie *TNFAIP2*

Spośród gorących punktów mutacji zlokalizowanych w sekwencji kodującej genów najczęściej występujący gorący punkt mutacji jest zlokalizowany w chr14:103,131,144-103,131,145 w genie *TNFAIP2* (w 6 próbkach (22,2 %)). Zidentyfikowane w gorącym punkcie mutacje to substytucja c.1292C>T lub podwójna substytucja c.1292delinsTT (podwójne podstawienia są opisane jako warianty delekcji/insercji (delins) zgodnie z nomenklaturą HGVS) skutkujące zmianą aminokwasu p.Ser431Phe oraz substytucja c.1293C>T (p.Ser431Ser) (Ryc. 14).



#### Rycina 14. Rozkład mutacji w genie *TNFAIP2*.

Schemat transkryptu (NM\_006291) genu *TNFAIP2* wraz z zaznaczonymi pozycjami mutacji wykrytych w BCC. Region kodujący (z podziałem na eksony i zaznaczonymi fragmentami kodującymi poszczególne domeny białkowe) został zaznaczony w formie szerszego prostokąta, regiony 5'UTR i 3'UTR zostały zaznaczone w formie węższych szarych prostokątów. Pozycje w transkrypcie (nt) i sekwencji kodującej (AA) zaznaczono odpowiednio powyżej schematu i na sekwencji kodującej. Rozwinięcie nazwy domeny: Sec6- domena posiadająca sekwencję aminokwasową podobną do białka SEC6.

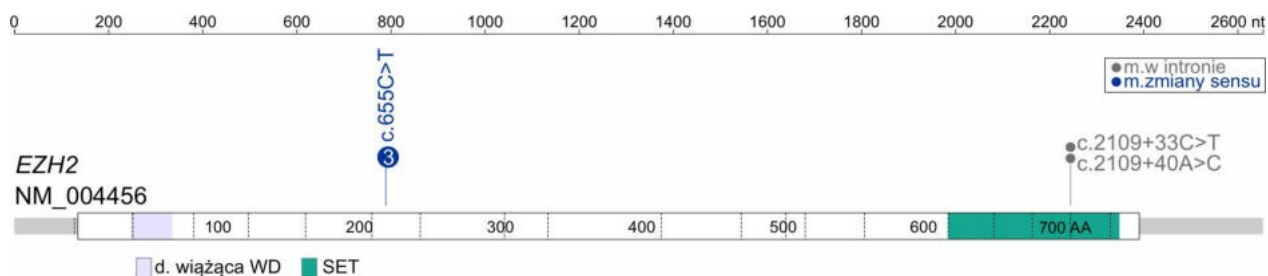
*TNFAIP2* to onkogen, kodujący wielofunkcyjne białko odgrywające rolę w angiogenezie, wywoływaniu stanu zapalnego, migracji i inwazji komórek oraz przebudowie cytoszkieletu i tworzeniu się błony komórkowej (Hase et al., 2009; Jia et al., 2016, 2018; Xie & Wang, 2017). Zidentyfikowany gorący punkt mutacji nie był wcześniej zaobserwowany w żadnym nowotworze, także w BCC w związku z tym wpływ wykrytej mutacji na funkcję *TNFAIP2* wymaga dalszych badań.

## WYNIKI

### 1.5.1.2. Gorący punkt mutacji w genie *EZH2*

Kolejny gorący punkt mutacji (mutacje w 3 próbkach) zlokalizowany w sekwencji kodującej stanowi zlokalizowana w pozycji chr7:148,827,237 w genie *EZH2* substytucja c.655C>T (p.Pro219Ser).

*EZH2* to gen odnotowany w bazie CGC, kodujący podstawową podjednostkę (metylotransferazę) kompleksu represyjnego polycomb 2 (PRC2), która odgrywa rolę w metylacji histonów i wyciszaniu ekspresji genów (Kodach et al., 2010). *EZH2* jest dobrze znanym onkogenem związanym z bardziej agresywną postacią i gorszym rokowaniem wielu nowotworów, w tym BCC oraz innych nowotworów złośliwych skóry, czerniaka i SCC (Mahmoud et al., 2016). Zwiększona ekspresja *EZH2* została wykazana w SCC (Xie et al., 2014) i agresywnej formie BCC (Rao, Chan, Andrews, & Kahana, 2016). W przeciwieństwie do wcześniej wykrytych mutacji skupiających się głównie w katalitycznej domenie SET (Bödör et al., 2013; Donaldson-Collier et al., 2019) wykryty w niniejszej pracy gorący punkt mutacji jest zlokalizowany w N-końcowej części białka (Ryc. 15), która odpowiada m.in. za interakcję z histonami (Weaver et al., 2019).



#### **Rycina 15. Rozkład mutacji w genie *EZH2*.**

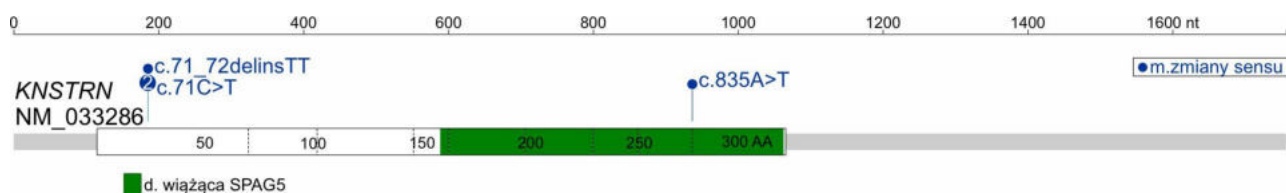
Schemat transkryptu (NM\_004456) genu *EZH2* wraz z zaznaczonymi pozycjami mutacji wykrytych w BCC. Schemat ryciny analogiczny do Ryc.14. Rozwinięcie nazw domen: d.wiążąca WD- domena wiążąca region bogaty w aminokwasy W (Trp) oraz D (Asp), SET- domena o aktywności metylotransferazy.

Zidentyfikowany gorący punkt mutacji nie był wcześniej zaobserwowany w żadnym nowotworze, także w BCC w związku z tym wpływ wykrytej mutacji na funkcję *EZH2* wymaga dalszych badań.

## WYNIKI

### 1.5.1.3 Gorący punkt mutacji w genie *KNSTRN*

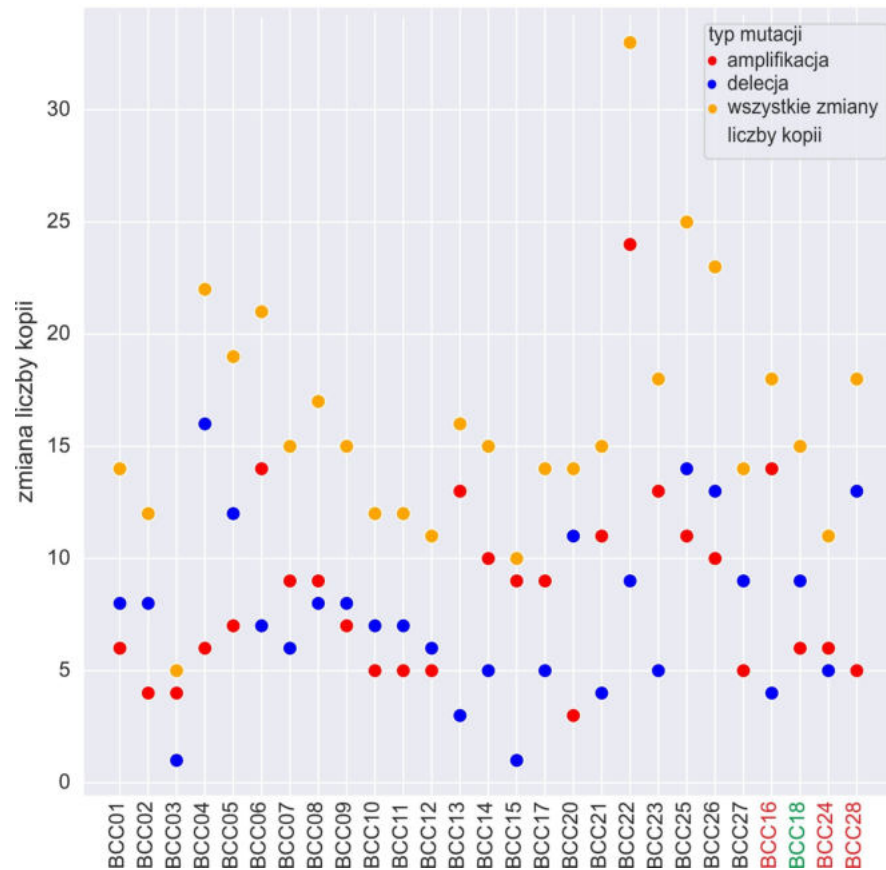
Kolejny gorący punkt mutacji (mutacja w 3 próbkach) stanowi zlokalizowana w chr15:40,382,906-40,382,907 substytucja c.71C>T lub podwójna substytucja c.71\_72delinsTT. Obie substytucje skutkują zmianą aminokwasu p.Ser24Phe, zlokalizowane w N-końcowej części białka *KNSTRN* (Ryc. 16), które odgrywa rolę w utrzymaniu kohezji chromatyd i prawidłowej separacji chromatyd podczas anafazy (Fang, Seki, & Fang, 2009).



#### **Rycina 16. Rozkład mutacji w genie *KNSTRN*.**

Schemat transkryptu (NM\_033286) genu *KNSTRN* wraz z zaznaczonymi pozycjami mutacji wykrytych w BCC. Schemat ryciny analogicznie do Ryc.14.

Mutacje w *KNSTRN* (głównie mutacja Ser24Phe) zostały wcześniej wykryte u pacjentów z BCC, SCC, AK oraz czerniakiem (Jaju et al., 2015; Lee et al., 2014). Mutacje w genie *KNSTRN* w SCC wpływają na separację chromosomów podczas podziałów komórkowych i są związane ze zwiększoną niestabilnością chromosomów, wyrażoną jako frakcja genomu ze zmianami liczby kopii (Lee i in., 2014). W celu zweryfikowania niniejszego mechanizmu w BCC, przeprowadzono analizę ilości zmian liczby kopii (z użyciem GISTIC2.0) w próbkach z mutacją oraz bez mutacji w genie *KNSTRN*. Związek mutacji *KNSTRN* z ilością zmian liczby kopii nie został potwierdzony ani w niniejszej pracy (Ryc. 17;  $p = 0,7392$ , test T-studenta) ani na podstawie analiz próbek BCC przedstawionych w pracy Jaju i in. (Jaju i in., 2015). Warto zauważyć, że *KNSTRN* odgrywa też rolę w apoptozie wywołanej promieniowaniem UV. Jednak wpływ mutacji w *KNSTRN* na unikanie apoptozy przez komórki BCC lub jakiegokolwiek inne komórki nowotworowe nie został jeszcze zbadany.



**Rycina 17. Wpływ mutacji w KNSTRN na zmiany liczby kopii w BCC.**

Kolorem czerwonym zaznaczono liczbę amplifikacji, niebieskim liczbę delecji, żółtym liczbę wszystkich zmian liczby kopii (amplifikacje oraz delecje) w poszczególnych próbkach BCC zaznaczonych na osi x. Próbki BCC oznaczone kolorem czerwonym to próbki z mutacją w gorącym punkcie mutacji w KNSTRN, natomiast próbka oznaczona kolorem zielonym posiada mutację w sekwencji kodującej KNSTRN poza gorącym punktem.

### 1.5.2 Gorące punkty mutacji w regionach niekodujących

Analiza 20 gorących punktów mutacji zlokalizowanych w regionach niekodujących wykazała 11 gorących punktów w intronach, 8 w regionie 5'UTR oraz 1 w regionie 3'UTR. W przypadku intronów gorące punkty lokalizowały się poza dobrze poznanymi, funkcjonalnymi częściami intronów (eksonowe sekwencje wzmacniające i wyciszające splicing, miejsca donorowe i akceptorowe splicingu - analiza z zastosowaniem ESEfinder, oraz miejsce rozgałęzienia- analiza z zastosowaniem LaBranchoR (Paggi & Bejerano, 2018) w związku z tym geny posiadające gorące punkty w intronach nie zostały dalej analizowane. Dla wybranych gorących punktów zlokalizowanych w regionach 5'UTR oraz 3'UTR przeprowadzono analizę porównawczą mutacji wykrytych w BCC z innymi typami nowotworów. W tym celu

## WYNIKI

wykorzystano dane dostępne w bazie danych TCGA, która zawiera informacje o mutacjach wykrytych w ponad 10 000 próbek z 33 typów nowotworów (nie obejmuje BCC).

### 1.5.2.1 Gorący punkt mutacji w regionie 3'UTR genu *BAD*

Najczęstszy ze wszystkich wykrytych gorących punktów mutacji w BCC to punkt zmutowany w 8 próbkach (30%) w pozycji chr11:64,270,066-64,270,067 w regionie 3'UTR genu *BAD*. Obejmuje on 4 różne substytucje (c.\*142C>A, c.\*142C>T, c.\*142\_\*143delinsTT i c.\*143C>T), zlokalizowane 142 i/lub 143 nukleotydy za kodonem stop. Gorący punkt w *BAD* jest zlokalizowany w ramieniu 5' motywu stabilnej struktury RNA o długości ~40 pz ( $dG=-39,6$  kcal/mol), który jest destabilizowany (spadek stabilności o ~2 kcal/mol) w wyniku mutacji (Ryc. 18). Wykryty gorący punkt nie był wcześniej zidentyfikowany w żadnym nowotworze. Białko kodowane przez *BAD* należy do rodziny białek BCL-2 i odgrywa kluczową rolę w pozytywnej regulacji apoptozy (Stickles et al., 2015).

Analiza danych z TCGA pozwoliła na zidentyfikowanie 28 mutacji w regionie 3'UTR genu *BAD* (Ryc. 18). Mutacje znaleziono głównie w próbkach czerniaka (15 mutacji w 12 próbkach (2,6% analizowanych próbek czerniaka)), w tym 4 mutacje pokrywające się z gorącym punktem (c.\*142C i c.\*143C) i 6 mutacji (c.\*77C>T) stanowiących dodatkowy gorący punkt nie występujący w BCC. Poza czerniakiem, mutacje regionu 3'UTR były bardzo rzadkie (Ryc. 18). Wyłącznie mutacji w 3'UTR w czerniaku i BCC wyklucza przypadkowe wystąpienie mutacji i przemawia za napędzającą funkcją mutacji w BCC i czerniaku.







## WYNIKI

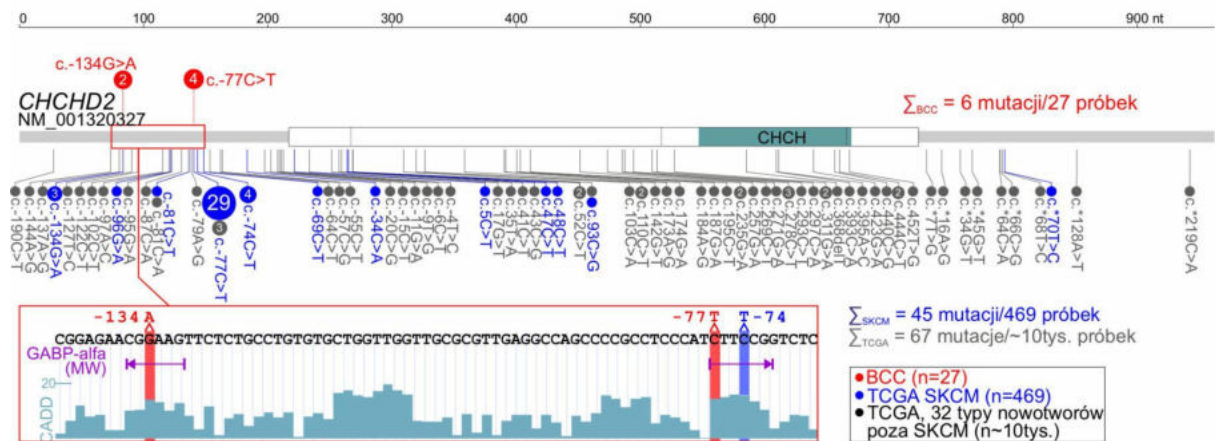
Mutacje tworzące gorący punkt w genie *DHODH* w BCC lokalizują się w sekwencji Kozak, która jest kluczową sekwencją wpływającą na proces inicjacji translacji. Analiza z użyciem narzędzia ATGpr (Salamov, Nishikawa, & Swindells, 1998) oraz NetStart 1.0 (Peitsch, 1997) nie wykazała jednak znacznego wpływu mutacji w pozycjach -4 oraz -5 na zmianę wyboru kodonu start i efektywność inicjacji translacji. Analiza struktury drugorzędowej RNA wykazała, że mutacje w gorącym punkcie *DHODH* nieznacznie obniżyły stabilność przewidywanego motywu spinki do włosów, tworzonego przez RNA w miejscu mutacji (Ryc. 19). Dalsza analiza sekwencji 5'UTR wykazała, że podwójna substytucja GG>AA w gorącym miejscu tworzy konsensusowe miejsce wiązania dla czynnika transkrypcyjnego NFAT1 (Ryc. 19), który występuje na wysokim poziomie w wielu tkankach, w tym zarówno w narażonej jak i chronionej przed promieniowaniem UV skórze (wg GTExPortal; GTEx Consortium Science 2020) i ma związek z wieloma nowotworami, w tym z czerniakiem (Shen et al., 2019; Shoshan et al., 2016).

### 1.5.2.3 Gorący punkt mutacji w regionie 5'UTR genu *CHCHD2*

Kolejny gorący punkt mutacji zidentyfikowany w niekodującej części genu to punkt zmutowany w 4 próbkach BCC (15%), w pozycji genomowej chr7:56,106,490 w regionie 5'UTR genu *CHCHD2* (znanego również jako *MNRR1*). Gorący punkt obejmuje substytucje c.-77C>T, natomiast warto zaznaczyć, że dalsza analiza wykazała jeszcze jedną powtarzającą się (występującą w 2 próbkach) substytucję c.-134G>A, zlokalizowaną w chr7:56,106,547, co przekłada się na łączną ilość 6 mutacji w 6 próbkach BCC.

Analiza porównawcza mutacji w BCC z mutacjami dostępnymi w bazie danych TCGA wykazała, że obie wspomniane pozycje genomowe są także często mutowane w czerniaku, co potwierdzono także w pracy Hayward i in. (Hayward et al., 2017).

## WYNIKI



### Rycina 20. Rozkład mutacji w CHCHD2.

Schemat transkryptu (NM\_001320327) genu CHCHD2 wraz z zaznaczonymi pozycjami mutacji. Schemat ryciny analogiczny do Ryc.14. Kolorem czerwonym zaznaczono mutacje występujące w BCC, niebieskim mutacje w czerniaku (SKCM) a czarnym mutacje w pozostałych typach nowotworów dostępnych w bazie TCGA. Na dole ryciny zaznaczono miejsca wiązania czynników transkrypcyjnych. Wartość wskaźnika CADD, który może być wskaźnikiem funkcjonalnego potencjału dowolnej mutacji we wskazanych pozycjach genomowych zaznaczono na dole wykresu. Rozwinięcie nazwy domeny CHCH- domena tworząca strukturę typu coiled-coil-helix-coiled-coil-helix.

W sumie na podstawie danych z TCGA zidentyfikowano 63 mutacje w regionie 5'UTR genu CHCHD2 (Ryc. 20). Mutacje stwierdzono głównie w próbkach czerniaka (40 mutacji w 39 (8,5%) próbkach), w tym 29 mutacji c.-77C>T i 3 mutacje c.-134G>A oraz dodatkowo zidentyfikowano 4 próbki z mutacją c.-74C>T, stanowiące dodatkowy gorący punkt w regionie 5'UTR. Tylko 5 mutacji w czerniaku było zlokalizowanych poza regionem 5'UTR, 4 mutacje znajdowały się w regionie kodującym (2 mutacje zmiany sensu i 2 synonimiczne) i 1 w regionie 3'UTR.

W innych nowotworach mutacje w rejonie 5'UTR występowały bardzo rzadko- 4 mutacje w nowotworach głowy i szyi oraz raku trzonu macicy, 3 mutacje w raku piersi i 12 mutacji w innych typach nowotworów na łącznie blisko 10 000 próbek.

Gorące punkty występujące zarówno w BCC jak i czerniaku znajdują się w dwóch różnych miejscach wiązania czynnika transkrypcyjnego GABP-alfa. Przeprowadzona w niniejszej pracy doktorskiej analiza z użyciem MotifMap (Daily, Patel, Rigor, Xie, & Baldi, 2011) wykazała potencjalny wpływ mutacji na zaburzenie wiązania czynników GABP (Ryc. 20).

## 1.6 Geny często mutowane w BCC

Za często mutowane geny uznano geny mające (i) mutacje niesynonimiczne w regionie kodującym w co najmniej 5 próbkach, (ii) mutacje w regionie 5'UTR w co najmniej 4 próbkach, (iii) mutacje w regionie 3'UTR w co najmniej 4 próbkach, oraz (iv) mutacje w intronach (do 40 nt od granicy ekson/intron) w co najmniej 5 próbkach (Tabela 4). Analizę częstości występowania mutacji przeprowadzono osobno dla każdego analizowanego regionu. W przypadku intronów mutacje w regionach kluczowych dla przebiegu splicingu, zlokalizowane w pozycjach +/- 2 nt od eksonów dołączono do analiz regionu kodującego. Z listy często mutowanych genów zostały usunięte geny sklasyfikowane w pracy Lawrence i in. (Lawrence i in., 2013) jako posiadające liczne mutacje pasażerskie, których wysoka częstość wynika z dużej wielkości genu (geny kodujące białka o długości >4000 aminokwasów oraz geny posiadające bardzo długie introny, obejmujące region genomowy >1Mb) a także z dużej zmienności genetycznej analizowanego regionu (m.in. geny kodujące receptory węchowe).

W związku z wykryciem dużej liczby często mutowanych genów (915 genów) w dalszych etapach analizy skupiono się na genach obecnych w bazie CGC i/lub genach wybranych na podstawie przeglądu literaturowego wykazujących wpływ na proces nowotworzenia i/lub funkcję skóry. Na rycinach przedstawionych w poniższych akapitach zaznaczono wyłącznie mutacje występujące w regionie, na podstawie którego gen został sklasyfikowany jako często mutowany.

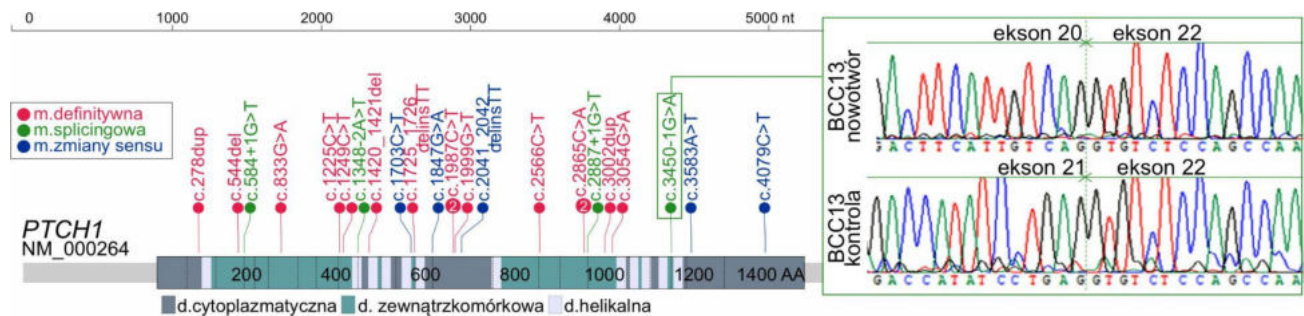
**Tabela 4. Liczba genów posiadających częste mutacje w poszczególnych częściach genów.**

Region	Minimalna liczba mutacji	Liczba często mutowanych genów
CDS	5	606
5'UTR	4	10
3'UTR	4	10
Intron	5	289

## 1.6.1 Geny często mutowane w regionach kodujących białko

### 1.6.1.1 Częste mutacje w genach zaangażowanych w szlak sygnałowy Hedgehog

Mutacje w genach szlaku Shh są charakterystyczne dla rozwoju BCC. Spośród 606 genów często mutowanych w regionie kodującym (Załącznik 2), *PTCH1* charakteryzuje się największą liczbą mutacji posiadając łącznie 24 mutacje w 20 próbkach BCC, w tym 5 mutacji zmiany sensu, 4 mutacje splicingowe oraz 15 mutacji definitywnych (Ryc. 21).



**Rycina 21. Rozkład mutacji niesynonimicznych w genie *PTCH1*.**

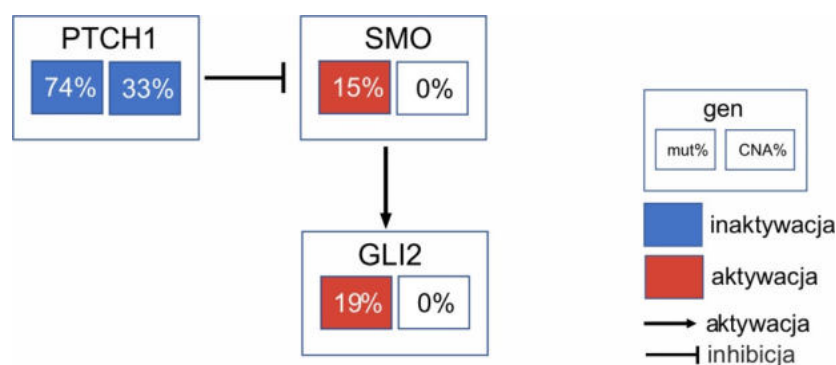
Schemat transkryptu (NM\_000264) genu *PTCH1* wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14. Załączony elektroferogram przedstawia funkcjonalny wpływ mutacji splicingowej wykrytej w pozycji c.3450-1G>A, skutkującej pominięciem eksonu 21 w dojrzałym transkrypcie.

Funkcjonalny wpływ mutacji splicingowej c.3450-1G>A zlokalizowanej powyżej eksonu 21 (w intronie 20) został potwierdzony z użyciem metody sekwencjonowania Sangera. Analiza sekwencjonowania wykazała pominięcie eksonu 21 w procesie splicingu (Ryc. 21).

Mutacje w regionie kodującym zaobserwowano także w innych genach zaangażowanych w funkcjonowanie szlaku sygnałowego Shh. *GLI2* posiadał mutacje w 5 próbkach BCC (19%) i został podobnie jak *PTCH1* sklasyfikowany jako gen często mutowany w regionie kodującym. Dodatkowo, wykryto 4 mutacje w 4 próbkach (15%) w genie *SMO* (Ryc. 22). Łączna częstość mutacji w *SMO* i *GLI2* jest niższa w próbkach z mutacją w genie *PTCH1* (4/20; 20%) niż w próbkach bez (4/7; 57%) mutacji (wynik nieistotny statystycznie; wartość  $p=0,1445$  - test Fishera), co sugeruje tendencję do wzajemnego wykluczania się mutacji w *PTCH1* z innymi genami szlaku Shh. Do zwiększenia aktywności szlaku Shh konieczna jest mutacja w jednym z kilku genów biorących udział w szlaku, stąd rzadkie mutacje w więcej niż

## WYNIKI

jednym genie ze szlaku. W genach szlaku Shh wykryto także zmiany liczby kopii, szerzej opisane w dalszej części pracy (analiza z zastosowaniem GISTIC2.0).



**Rycina 22. Mutacje w genach szlaku Shh.**

Kolorem niebieskim zaznaczono mutacje skutkujące inaktywacją genu, natomiast kolorem czerwonym mutacje wywołujące aktywację genu. Częstość substytucji i indeli (mut) oraz zmian liczby kopii (CNA; ang. copy number alteration) w analizowanych próbkach wyrażono w procentach.

### 1.6.1.2 Geny poza szlakiem Hedgehog, często mutowane w BCC

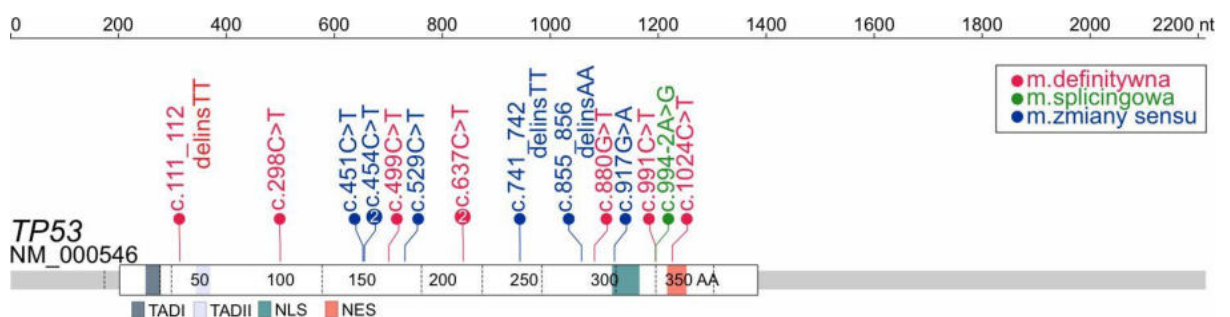
Poza mutacjami w genach zaangażowanych w ścieżkę sygnałową Shh częste mutacje występują także w genach zaangażowanych w inne szlaki komórkowe (Tabela 5).

**Tabela 5. Geny poza szlakiem Shh często mutowane w regionie kodującym.**

Gen	Mutacje zmiany	Mutacje	Mutacje	Łączna liczba (%)
TP53	7	8	1	13
PTPRD	18	1	-	13
FLG	12	3	-	10
FLG2	5	2	-	9
MYCN	8	-	-	8
NOTCH1	8	2	-	8
NOTCH2	3	3	2	7
NOTCH3	6	-	-	5

### 1.1.6.2.1 Mutacje w genie *TP53*

*TP53* to jeden z najlepiej poznanych genów supresorowych, często mutowany w wielu typach nowotworów. W analizowanych próbkach BCC mutacje w *TP53* wykryto w 13 próbkach (48%) (Ryc.23). Zaobserwowano, że aż 50% (8 mutacji) wykrytych mutacji stanowią zmiany definitywne. Dodatkowo, wykryto 7 mutacji zmiany sensu oraz jedną mutację splicingową. W niniejszej pracy wykazano, że *TP53* jest drugim najczęściej mutowanym w regionie kodującym genem w BCC, co jest zgodne z wcześniejszymi doniesieniami literaturowymi (Bonilla et al., 2016; Jayaraman et al., 2014).



**Rycina 23. Rozkład mutacji niesynonimicznych w genie *TP53*.**

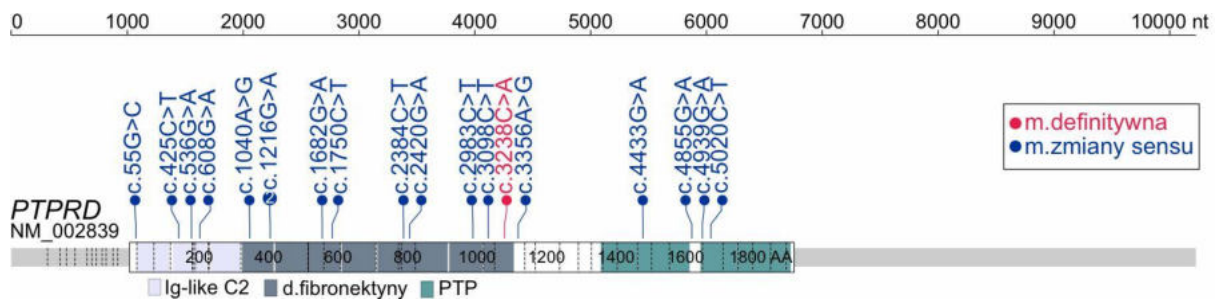
Schemat transkryptu (NM\_000546) genu *TP53* wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazw domen: TAD/TADII- domena transktywacyjna, NLS- domena lokalizacji jądrowej, NES- domena zawierając sygnał eksportu jądrowego.

### 1.1.6.2.2 Mutacje w genie *PTPRD*

Częste mutacje zidentyfikowano także w *PTPRD* (Ryc. 24), genie supresorowym często mutowanym w wielu nowotworach, w tym nowotworze głowy i szyi, glejaku, czerniaku i SCC (Peyser et al., 2015; Purdie et al., 2007; Veeriah et al., 2009; Walia et al., 2014). *PTPRD* katalizuje defosforylację wybranych białek prowadząc do zwiększenia aktywności onkogenów m.in. STAT3. Gen ten nie był wcześniej uznany za często mutowany w BCC. W niniejszej pracy wykryto 18 mutacji w regionie CDS w tym 18 mutacji zmiany sensu oraz 1 mutację definitywną. Mutacje posiadało 13 próbek BCC (48%).



## WYNIKI

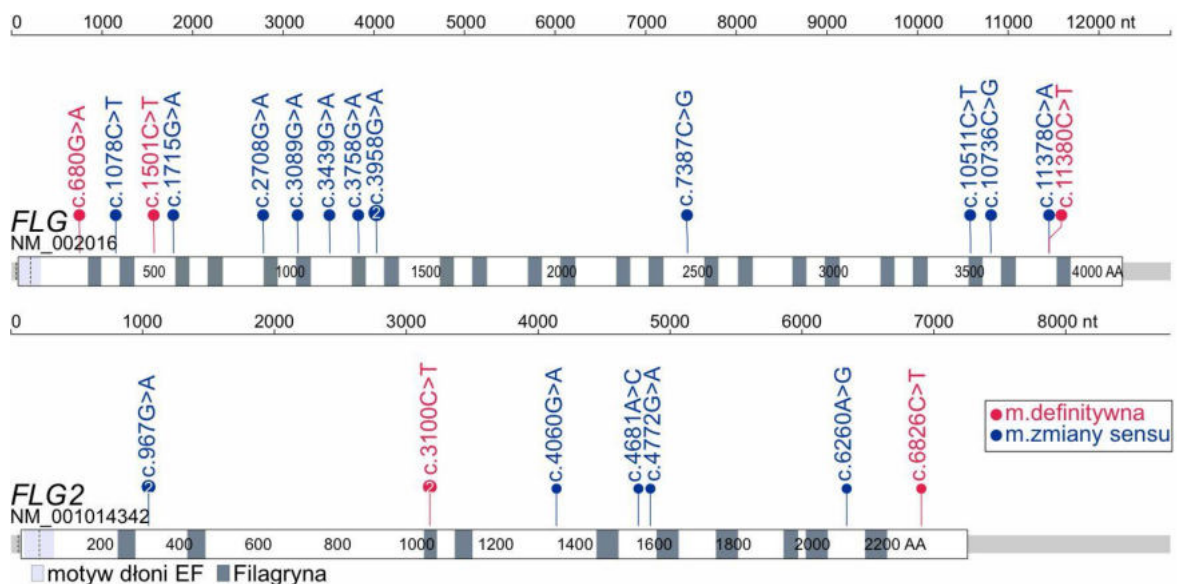


### Rycina 24. Rozkład mutacji niesynonimicznych w genie PTPRD.

Schemat transkryptu (NM\_002839) genu PTPRD wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazw domen: Ig-like C2- domena immunoglobulino-podobna, PTP- domena o aktywności fosfatazy.

### 1.1.6.2.3 Mutacje w genie FLG oraz FLG2

W analizowanych próbkach BCC wykryto wysoką częstość mutacji w genach FLG (15 mutacji w 10 próbkach) i FLG2 (9 mutacji w 9 próbkach) (Ryc. 25), kodujących profilagrynę i białko podobne do filagryny (ang. *filaggrin-like protein*), będące prekursorami dojrzałej filagryny, która jest ważnym składnikiem warstwy rogowej naskórka. W niniejszej pracy w genach FLG oraz FLG2 zaobserwowano wysoki odsetek definitywnych nonsensownych mutacji, występujących łącznie w 5 próbkach BCC (Ryc.25).

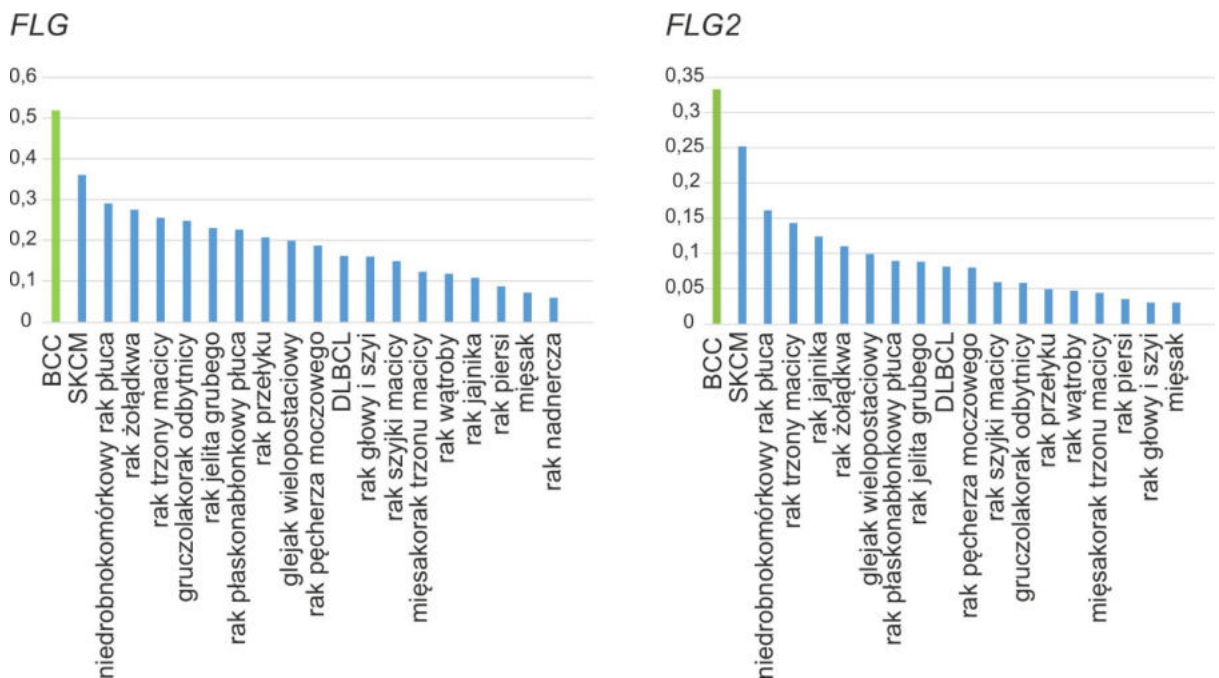


### Rycina 25. Rozkład mutacji niesynonimicznych w genie FLG i FLG2.

Schemat transkryptów (NM\_002016) genu FLG oraz (NM\_001014342) FLG2 wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14.

## WYNIKI

Mutacje w *FLG* i *FLG2* wykryto wcześniej także w innych nowotworach takich jak czerniak, rak szyjki macicy, czy rak prostaty (Wu et al., 2022; Xu, Luo, Hu, & Zhu, 2021; Zhang, An, Sun, Xu, & Zhang, 2022). W celu porównania częstości mutacji w BCC z innymi typami nowotworów przeprowadzono analizę porównawczą BCC i próbek dostępnych w bazie TCGA. Wykazano, że częstość występowania mutacji w genach *FLG* i *FLG2* obserwowanych w BCC znacznie przewyższa częstość mutacji w innych nowotworach, w tym czerniaku (kolejny najczęściej mutowany w genach *FLG/FLG2* nowotwór) (Ryc. 26).



**Rycina 26. Częstość mutacji w genach *FLG* i *FLG2*.**

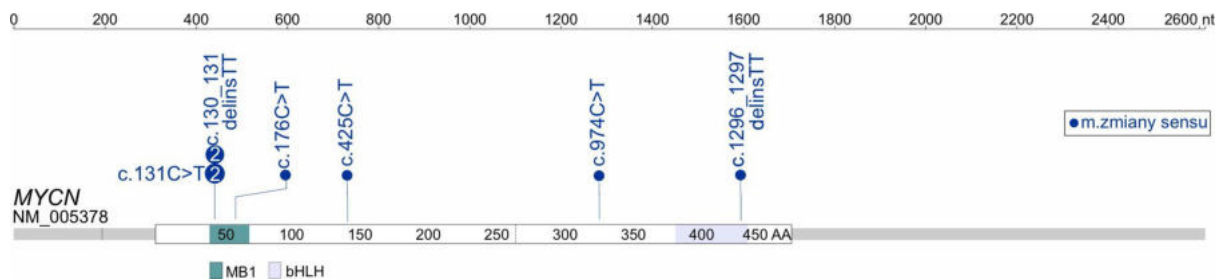
Na osi y zaznaczono frakcję próbek posiadającą mutację w genach *FLG* oraz *FLG2* w różnych typach nowotworów oznaczonych na osi x.

### 1.1.6.2.4 Mutacje w genie *MYCN*

*MYCN* to dobrze znany onkogen, który ulega częstym mutacjom w różnych typach nowotworów (Liu, Shi, Wang, Yuan, & Cui, 2021). *MYCN* stanowi jeden z wielu genów efektorowych, którego ekspresja jest regulowana przez aktywność szlaku Shh (Freier et al., 2006). Dotychczas mutacje w *MYCN* w BCC zostały wykryte głównie w regionie MB1 (ang. *Myc box 1*) odpowiedzialnym za wiązanie FBXW7 – supresora prowadzącego do degradacji *MYCN* przez aktywację ścieżki ubikwitynizacji (Welcker et al., 2004). Wykazano, że mutacje występujące w MB1 zaburzają wiązanie FBXW7 co prowadzi do wzrostu ekspresji *MYCN*

## WYNIKI

(Bonilla et al., 2016). W niniejszej pracy w *MYCN* zidentyfikowano 8 mutacji zmiany sensu w 8 próbkach BCC, z czego 5 mutacji lokalizuje się w regionie domeny MB1 (Ryc.27).



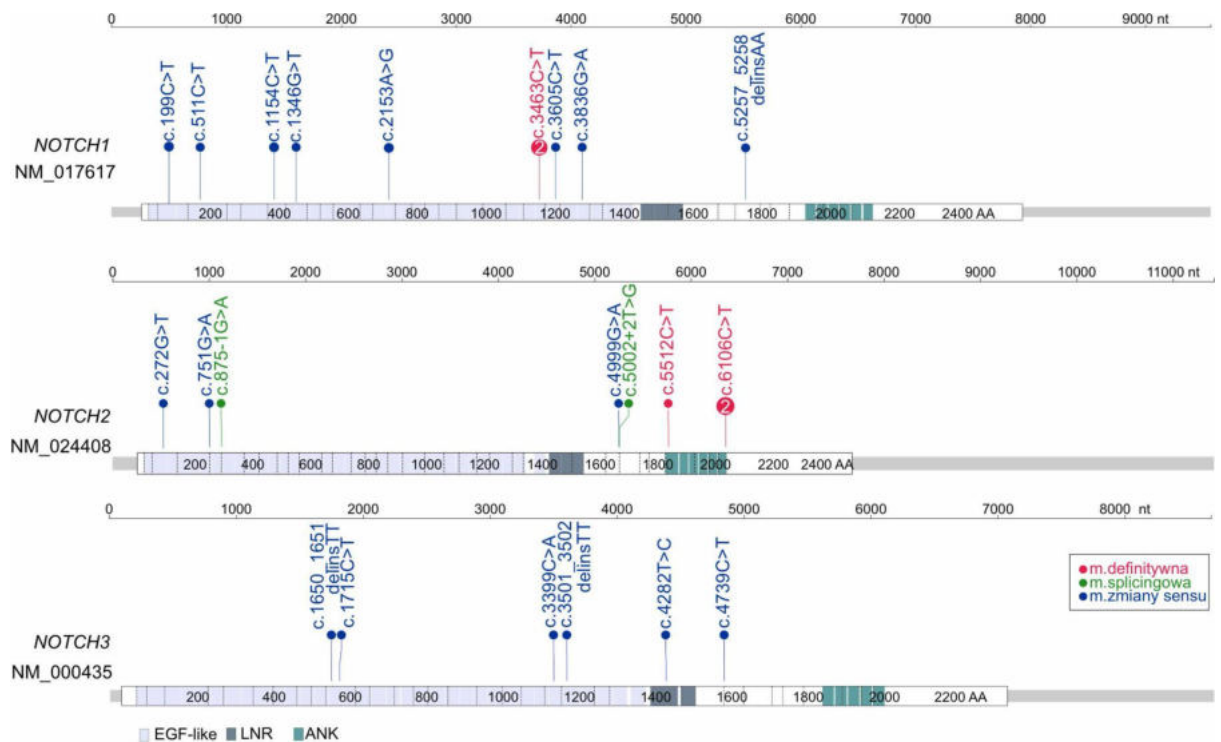
### **Rycina 27. Rozkład mutacji niesynonimicznych w genie MYCN.**

Schemat transkryptu (NM\_005378) genu *MYCN* wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazwy domeny bHLH- domena helisa-pętla-helisa.

#### **1.1.6.2.5 Mutacje w genach NOTCH**

Geny należące do rodziny *NOTCH*, w tym: *NOTCH1*, *NOTCH2* oraz *NOTCH3* to geny supresorowe posiadające częste mutacje w BCC (Ryc.28). *NOTCH* pełnią funkcję genów supresorowych w rozwoju BCC oraz SCC (Lefort et al., 2007; Thélu, Rossio, & Favier, 2002) a spadek ich ekspresji sprzyja przetrwaniu komórek nowotworowych (Eberl i in., 2018). W niniejszej pracy doktorskiej mutacje w *NOTCH1* zidentyfikowano w 8 próbkach, w tym 8 mutacji zmiany sensu a także 2 mutacje nonsensowne. *NOTCH2* posiada 8 mutacji w 7 próbkach BCC, w tym 3 mutacje zmiany sensu, 3 warianty definitywne oraz 2 mutacje splicingowe. W *NOTCH3* zaobserwowano 6 mutacji zmiany sensu w 5 próbkach BCC.

## WYNIKI



**Rycina 28. Rozkład mutacji niesynonimicznych w genach NOTCH.**

Schemat transkryptów (NM\_017616) genu NOTCH1, (NM\_024408) NOTCH2 oraz (NM\_000435) NOTCH3 wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie kodującym w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazw domen: EGF-like-domena EGF-podobna, ANK- domena odpowiedzialna za oddziaływanie białko-białko, LNR-domena bogata w cysteinę, zawierająca powtórzenia Lin12/Notch charakterystyczne dla rodziny NOTCH (LIN-12/Notch repeats).

### 1.6.3 Geny często mutowane w regionie 5'UTR

Za geny często mutowane w 5'UTR uznano geny posiadające mutacje w co najmniej 4 próbkach BCC. Analiza mutacji pozwoliła na wyselekcjonowanie 10 genów często mutowanych w regionie 5'UTR (Tabela 6).

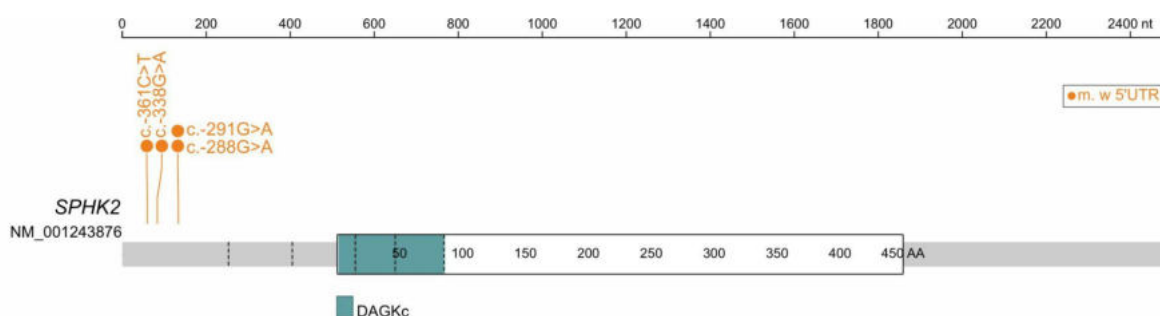
## WYNIKI

**Tabela 6. Geny często mutowane w regionie 5'UTR.**

Nazwa genu	Liczba mutacji	Liczba zmutowanych próbek
<i>CHCHD2</i>	6	6
<i>DHODH</i>	5	5
<i>POLR2M</i>	5	5
<i>PES1</i>	4	4
<i>RPL18</i>	4	4
<i>NPC1</i>	4	4
<i>PPME1</i>	4	4
<i>SPHK2</i>	4	4
<i>AIM1</i>	4	4
<i>NELL2</i>	4	4

Wśród często mutowanych genów znajdują się *DHODH* i *CHCHD2* opisane w poprzednim podrozdziale o gorących punktach mutacji. Oba geny charakteryzują się występowaniem mutacji wyłącznie w regionie 5'UTR w analizowanych próbkach BCC.

Kolejny gen często mutowany w BCC w regionie 5'UTR to posiadający 4 mutacje gen *SPHK2* (Ryc.29).



**Rycina 29. Rozkład mutacji w genie *SPHK2*.**

Schemat transkryptu (NM\_001243876) genu *SPHK2* wraz z zaznaczonymi pozycjami mutacji wykrytych w regionie 5'UTR w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazwy domen DAGKc- domena katalityczna o aktywności kinazy.

## WYNIKI

Rola *SPHK2* w rozwoju BCC nie została opisana, natomiast wiadomo, że ulega on wysokiej ekspresji w SCC (inny nowotwór skóry należący do nieczerniakowych nowotworów skóry) a jego inhibicja skutkuje zahamowaniem wzrostu komórek, co sugeruje onkogenną funkcję *SPHK2* (Zhou, Chen, & Yu, 2018).

Funkcja pozostałych genów często mutowanych w regionie 5'UTR nie jest dobrze poznana lub nie jest związana z procesem nowotworzenia, dlatego geny te nie były dalej analizowane w pracy doktorskiej.

### 1.6.4 Geny często mutowane w regionie 3'UTR

Za geny często mutowane w 3'UTR uznano geny posiadające mutacje w co najmniej 4 próbkach BCC. Analiza mutacji pozwoliła na wyselekcjonowanie 10 genów często mutowanych w regionie 3'UTR (Tabela 7).

**Tabela 7. Geny często mutowane w regionie 3'UTR.**

Nazwa genu	Liczba mutacji	Liczba zmutowanych próbek
<i>BAD</i>	8	8
<i>SMIM27</i>	8	7
<i>OPRM1</i>	5	5
<i>RBM48</i>	5	4
<i>C4orf17</i>	5	5
<i>PHKG2</i>	4	4
<i>FCRL5</i>	5	4
<i>TEX35</i>	5	4
<i>ZNF780B</i>	4	4
<i>C19orf54</i>	4	4

Spośród genów często mutowanych w regionie 3'UTR jedynym genem biorącym udział w rozwoju nowotworów skóry jest gen *BAD*, który został opisany we wcześniejszym podrozdziale pracy dotyczącym gorących miejsc mutacji. Funkcja pozostałych genów nie jest

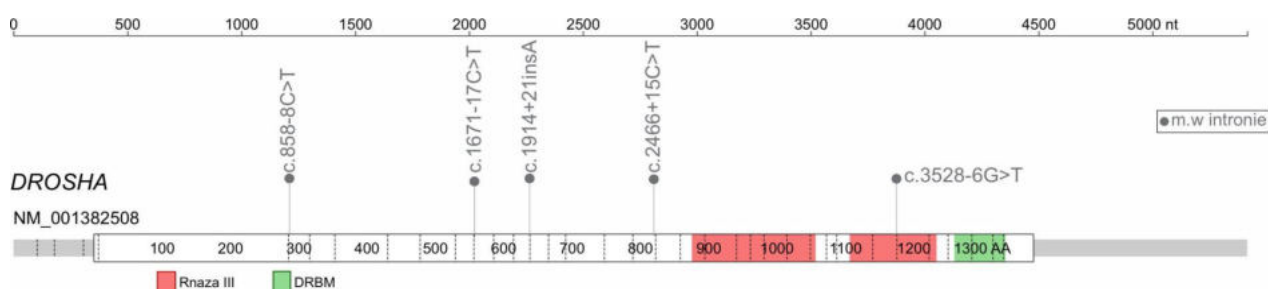
## WYNIKI

związana z procesem nowotworzenia ani rozwojem skóry, dlatego geny te nie były dalej analizowane w pracy doktorskiej.

### 1.6.5 Geny często mutowane w intronach

Analizę często mutowanych intronów przeprowadzono dla regionów obejmujących 40 nukleotydów od granicy z eksonami. Analiza pozwoliła na wyselekcjonowanie 289 często mutowanych genów (Załącznik 3). Wśród nich znalazł się m.in. *PTCH1*, który oprócz 4 mutacji w miejscach splicingowych (wspomnianych we wcześniejszym podrozdziale) posiadał również 4 inne mutacje intronowe (c.1504-8T>C, c.2560+9G>C, c.2887+21A>G oraz c.\*-21C>T; łącznie 8 mutacji intronowych w 5 próbkach).

Kolejny gen często mutowany w intronach to *DROSHA* (posiadający 5 mutacji w 5 próbkach BCC), kodujący jeden z głównych enzymów (rybonukleazę) szlaku biogenezy miRNA (Ryc. 30). Żadna z mutacji wykrytych w intronach nie znajduje się w miejscach kluczowych dla splicingu ani w miejscu donorowym, akceptorowym, ani w miejscu rozgałęzienia (analiza z zastosowaniem ESEfinder oraz LaBranchor).



#### **Rycina 30. Rozkład mutacji w genie *DROSHA*.**

Schemat transkryptu (NM\_001382508) genu *DROSHA* wraz z zaznaczonymi pozycjami mutacji wykrytych w intronach w BCC. Schemat ryciny analogicznie do Ryc.14. Rozwinięcie nazw domen: RNaza III- domena o aktywności katalitycznej RNazy III, DRBM- domena wiążąca RNA.

### 1.7 Identyfikacja genów napędzających rozwój nowotworu

Identyfikacja genów napędzających jest kluczowa dla zrozumienia mechanizmu prowadzącego do rozwoju danego typu nowotworu. Analizę genów napędzających przeprowadzono z użyciem OncodriveFML (Mularoni, Sabarinathan, Deu-Pons, Gonzalez-Perez, & López-Bigas, 2016). OncodriveFML identyfikuje geny napędzające analizując substytucje oraz indele zlokalizowane nie tylko w regionach kodujących, ale też niekodujących genów. Potencjał napędzający genów jest oceniany na podstawie szkodliwości (ang. *deleteriousness*) wykrytych mutacji szacowanej za pomocą współczynnika CADD (ang. *Combined Annotation Dependent Depletion*; (Rentzsch, Witten, Cooper, Shendure, & Kircher, 2019)) oceniającego szkodliwość mutacji poprzez zintegrowanie w jedną metrykę wielu adnotacji (>60), w tym m.in. konserwatywności badanych regionów, analizy miejsc wiązania czynników transkrypcyjnych i miRNA oraz wpływu mutacji na sekwencje aminokwasów i funkcje białek. Według założenia zastosowanego w OncodriveFML geny napędzające charakteryzują się tendencją do akumulacji mutacji o wysokiej szkodliwości. W przypadku, gdy obserwowany funkcjonalny wpływ mutacji na analizowany region jest wyższy niż wartość oczekiwana przez przypadek losowy dla danego regionu, oznacza to, że gen posiadający analizowane mutacje może pełnić funkcję napędzającą.

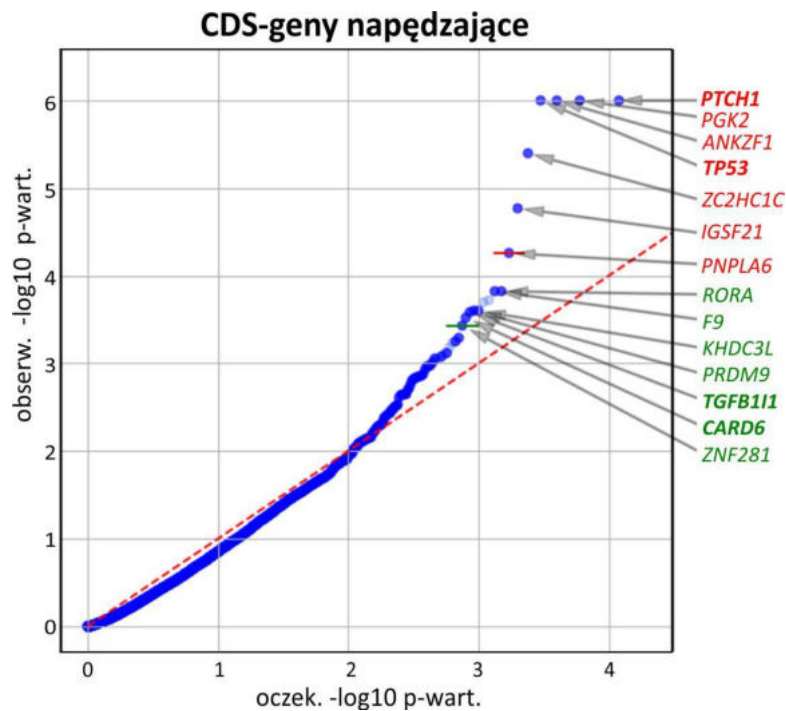
Analizę mutacji przeprowadzono osobno dla sekwencji kodujących oraz dla rozpatrywanych regionów niekodujących: 5'UTR i 3'UTR oraz intronów. Na podstawie przeprowadzonej analizy zidentyfikowano 14 potencjalnych genów napędzających w oparciu o mutacje w regionach kodujących (Ryc.31), 36 genów w oparciu o mutacje w 5'UTR (Ryc.32) oraz 7 genów na podstawie mutacji w 3'UTR (Ryc.33). Analiza mutacji w intronach nie wykazała genów o potencjale napędzającym.

W dalszej analizie skupiono się na wybranych genach o potencjale napędzającym pełniących funkcje związane z nowotworzeniem i/lub funkcją skóry.

#### 1.7.1. Geny napędzające wyselekcjonowane na podstawie mutacji w regionach kodujących białko

Analiza mutacji w regionach kodujących genów pozwoliła na wyselekcjonowanie 14 genów o potencjale napędzającym rozwój BCC (Ryc. 31).





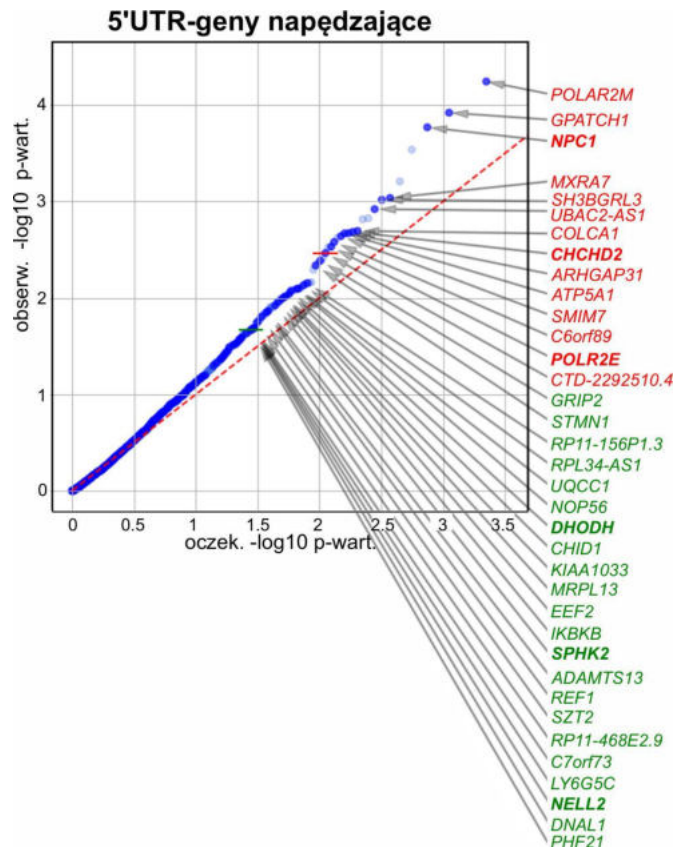
**Rycina 31. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji w regionie kodującym.**

Kolorem czerwonym zaznaczono geny, dla który  $q < 0.1$ , a zielonym  $< 0.25$ . Pogrubioną czcionką zaznaczono geny, które zostały sklasyfikowane jako geny często mutowane w regionie CDS. Objaśnienie skrótów: obserw.- obserwowana oczek.- oczekiwana, p-wart. – p-wartość.

Wśród wykrytych genów znajdują się geny często mutowane w regionie kodującym takie jak *PTCH1*, *TP53*, *TGFB11I* oraz *CARD6*. Dodatkowo, warto zwrócić uwagę na gen *RORA* pełniący ważną rolę w hamowaniu rozwoju alergicznego zapalenia skóry (Malhotra et al., 2018). Inne interesujące geny to *PRDM9* i *ZNF281*, odgrywające rolę w naprawie i generowaniu dwuniciowych pęknięć DNA (ang. *double strand breaks*; DSBs; (Houle et al., 2018; Nicolai et al., 2020). *ZNF281* bierze udział w naprawie DSBs przez niehomologiczne łączenie końców. Natomiast *PRDM9* koduje metylotransferazę pośrednio generującą DSBs. *PRDM9* wiąże DNA w miejscu, gdzie białko SPO11 (nakierunkowane do miejsca cięcia przez *PRDM9*) generuje DSBs konieczne do rekombinacji homologicznej między chromosomami w trakcie mejozy (Paiano et al., 2020). Mutacje zarówno w *PRDM9* jak i *ZNF281* mogą prowadzić do zwiększonej niestabilności genetycznej genomów nowotworowych.

### 1.7.2 Geny napędzające wyselekcjonowane na podstawie mutacji w regionie 5'UTR

Wśród 36 genów o potencjale napędzającym wyselekcjonowanych na podstawie mutacji w regionie 5'UTR (Ryc. 32) znajdują się geny *SPHK2*, *POLR2M*, *NPC1* i *NELL2* zidentyfikowane wcześniej jako często mutowane w BCC oraz *DHODH*, *CHCHD2* nie tylko często mutowane, ale także posiadające gorące miejsca mutacji.



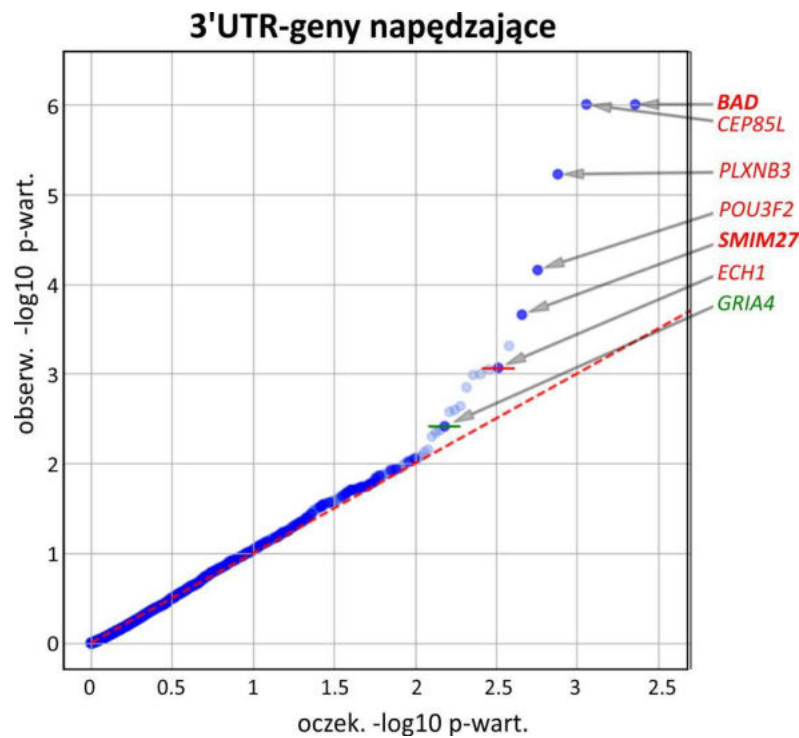
**Rycina 32. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji w regionie 5'UTR.**

Schemat ryciny analogiczny do Ryc.31.

Poza genami *DHODH*, *SPHK2* oraz *CHCHD2*, które zostały opisane we wcześniejszych rozdziałach, przeprowadzona analiza wykazała potencjał napędzający genu *IKBKB* (gen zmutowany w 3 próbkach), który wykazuje działanie supresorowe w nieczerniakowych nowotworach skóry i nienowotworowych zmianach skórnych a jego delecje prowadzą do zapalenia skóry, uszkodzenia mieszków włosowych, hiperplazji i rozwoju SCC (Cornish, Tung, Marshall, Ley, & Seddon, 2012; Kirkley, Walton, Duncan, & Tjalkens, 2017; Pasparakis et al., 2002; Stratis et al., 2006).

### 1.7.3 Geny napędzające wyselekcjonowane na podstawie mutacji w regionie 3'UTR

Wśród 7 genów o potencjale napędzającym rozwój BCC na podstawie mutacji w regionie 3'UTR (Ryc. 33) znajdują się dwa geny: *BAD* oraz *SMIM27* zidentyfikowane we wcześniejszej części pracy jako geny często mutowane w BCC.



**Rycina 33. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji regionu 3'UTR.**

Schemat ryciny analogiczny do Ryc.31.

Funkcja genu *BAD* została już wspomniana we wcześniejszej części pracy. Natomiast *SMIM27* (znany także jako lncRNA *TOPORS-AS1*) to gen, którego ekspresja jest regulowana przez receptor witaminy D (ang. *vitamin D receptor*) a nadekspresja jest związana z dobrym rokowaniem u pacjentów z nowotworem piersi i jajnika, w związku z blokowaniem aktywności ścieżki sygnałowej Wnt/B-keniny (Fu et al., 2021; Su et al., 2014).

Warto wspomnieć o genie *POU3F2* (zmutowany w 3 próbkach) kodującym czynnik transkrypcyjny, który odgrywa rolę w inwazyjności i przerzutowaniu czerniaka a jego ekspresja jest kontrolowana przez miR-211 (Boyle et al., 2011; Simmons, Pierce, Al-Ejeh, & Boyle, 2017) i miR-107 (Zhao, Wei, & Guo, 2020). Chociaż mutacje wykryte w niniejszej pracy w 3'UTR *POUF32* nie są zlokalizowane w przewidywanych miejscach wiązania miR-107 i miR-

## WYNIKI

211, mogą wpływać na strukturę 3'UTR, a tym samym w sposób pośredni zmieniać dostępność miejsc wiązania miRNA.

### **1.8 Identyfikacja regionów istotnie wzbogaconych w zmiany liczby kopii**

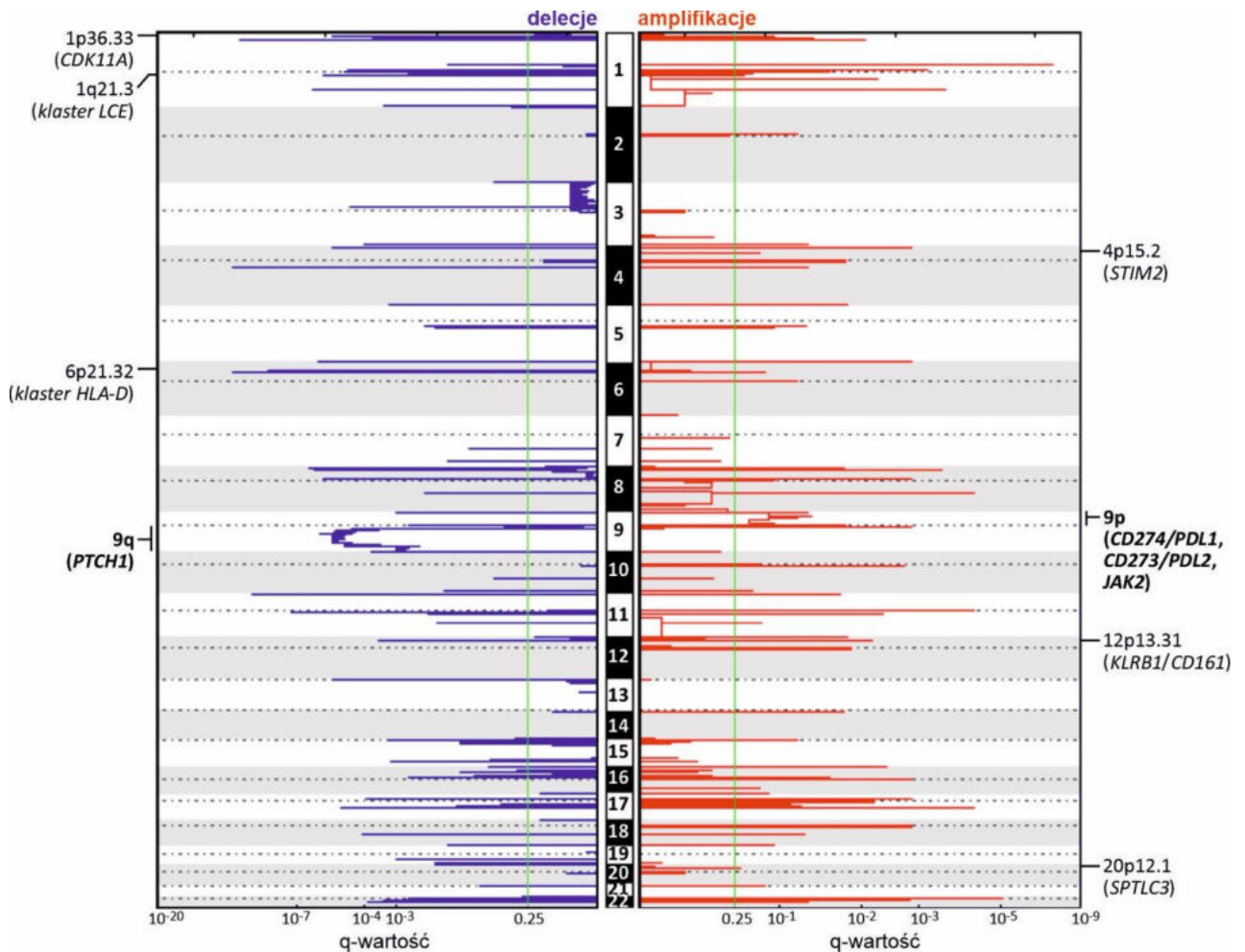
W celu wyselekcjonowania regionów ulegających w sposób istotny powtarzalnym zmianom liczby kopii przeprowadzono analizę z użyciem GISTIC 2.0 (ang. *Genomic Identification of Significant Targets in Cancer*) wykorzystując model bazujący na częstości występowania zdarzeń o określonej długości i amplitudzie w analizowanym zbiorze prób (Mermel i in., 2011). Takie regiony mogą być wskaźnikiem na występowanie w nich ważnych onkogenów (amplifikacje) lub genów supresorowych (delecje). Wśród wykrytych regionów można wyróżnić zmiany na poziomie całego ramienia chromosomu (obejmujące >50% ramienia) oraz zmiany lokalne, ogniskowe (ang. *focal*) obejmujące mniejszą część chromosomu. Analiza pozwoliła wyselekcjonować regiony ulegające zarówno amplifikacjom jak i delecjom.

#### **1.8.1 Zmiany liczby kopii na poziomie ramienia chromosomu napędzające proces nowotworzenia**

Na poziomie ramienia chromosomu wykryto powtarzającą się delecję krótkiego ramienia chromosomu 9 (chr9q ( $q=1,4 \times 10^{-6}$ ; występującą w 9 próbkach BCC)), obejmującą *PTCH1* (Ryc. 34), co jest zgodne z wcześniejszymi doniesieniami literaturowymi sugerującymi utratę heterozygotyczności *PTCH1* (Hahn et al., 1989; Mizuno et al., 2006) jako mechanizm napędzający rozwój BCC.

W przypadku amplifikacji zaobserwowano częstą duplikację długiego ramienia chromosomu 9 (chr9p ( $q=0,05$ ; występującą w 5 próbkach)), obejmującą geny *CD274* i *CD273* (geny znane również jako *PDL1* oraz *PDL2*, kodujące białka PD-L1 i PD-L2 stanowiące punkt kontroli układu immunologicznego), oraz *JAK2* (gen kodujący kinazę tyrozynową JAK2 zaangażowaną w funkcjonowanie szlaku sygnałowego JAK-STAT) (Ryc. 34). Dotychczas amplifikacja tego regionu została odnotowana tylko w jednym przypadku rzadkiego przerzutującego BCC (Ikeda et al., 2016).

## WYNIKI



**Rycina 34. Regiony napędzające rozwój BCC.**

Kolorem czerwonym zaznaczono regiony o potencjale napędzającym ulegające amplifikacjom, a kolorem niebieskim delecjom. Wykres przedstawia zarówno zmiany na poziomie ramienia chromosomu jak i zmiany ogniskowe. Na środku ryciny oznaczono 22 chromosomy. Próg wartości  $q=0.25$  oznaczono zieloną linią.

### 1.8.2 Ogniskowe zmiany liczby kopii napędzające proces nowotworzenia

Analiza ogniskowych zmian liczby kopii wykazała 54 regiony ulegające delecjom oraz 56 regionów ulegających amplifikacjom (Załącznik 4; Ryc. 34).

Geny znajdujące się w regionach najbardziej znaczących delecji ogniskowych to i) *CDK11A* (chr1p36.33;  $q=2,4 \times 10^{-5}$ ; delecja w 6 próbkach), spadek aktywności tego genu indukuje kancerogenezę w skórze (Chandramouli et al., 2007); ii) klaster *LCE* (chr1q21.3;  $q=2,4 \times 10^{-6}$ ; delecja w 4 próbkach), obejmujący geny takie jak *LCE2* i *LCE3*, które odgrywają rolę w utrzymaniu funkcji bariery ochronnej skóry, delecja tych genów ma wpływ na rozwój łuszczycy (de Cid et al., 2009); oraz iii) klaster *HLA-D* (*HLA-DP*, *-DQ* i *-DR*, chr6p21.32;  $q=2 \times 10^{-6}$ ).

## WYNIKI

<sup>4</sup>; delecja w 3 próbkach), kodujący składniki głównego układu zgodności tkankowej (MHC) klasy II, których zwiększona ekspresja wiąże się ze zwiększoną immunogennością raka i lepszym rokowaniem w BCC, SCC i czerniaku (Chen, Chang, Lin, Chen, & Kuo, 2019; García-Plata, Mozos, Carrasco, & Solana, 1993; Hua, Kagen, Carpenter, & Goltz, 1985; Johnson et al., 2016; Markey, Churchill, & MacDonald, 1990; Myskowski, Pollack, Schorr, Dupont, & Safai, 1985; Streilein, 1991).

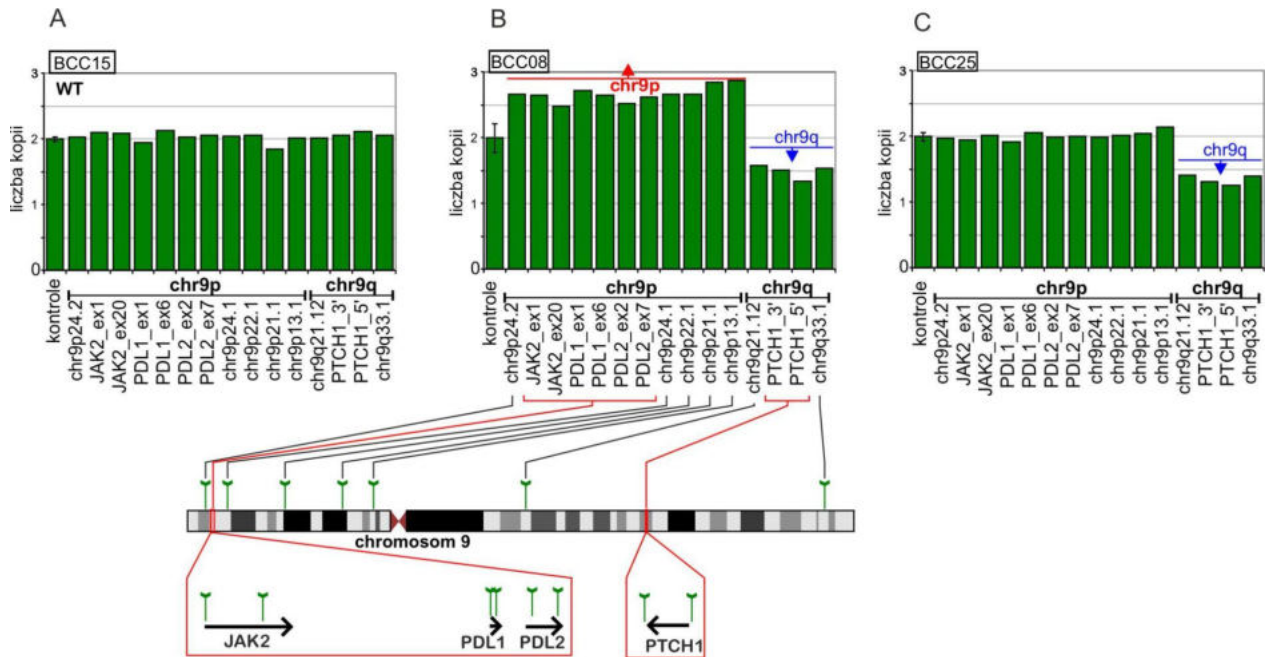
Geny w regionach ulegającym amplifikacjom to m.in. i) *STIM2* (chr4p15.2; q=0,16; amplifikacja w 2 próbkach) ulegający nadekspresji w czerniaku (Stanisz et al., 2014); ii) *KLRB1/CD161* (chr12p13.31; q=0,007; amplifikacja w 2 próbki) kodujący receptor ekspresjonowany na powierzchni limfocyty T regulujący odpowiedź układu immunologicznego przeciw nowotworom (Isvoranu et al., 2019; Sanchez-Canteli et al., 2020); oraz iii) *SPTLC3* (chr20p12.1 q=0,23; amplifikacja w 2 próbkach) szczególnie wysoce ekspresjonowany w skórze (Lone et al., 2020).

### 1.8.3 Walidacja regionów ulegających zmianie liczby kopii napędzających rozwój BCC

Aby potwierdzić obecność zmian liczby kopii chromosomu 9, samodzielnie opracowano test wykorzystujący metodę zależnej od ligacji multipleksowej amplifikacji sond (ang. *multiplex ligation probe amplification*; MLPA). Zaprojektowano sondy rozłożone wzdłuż chromosomu 9, ale w szczególności skupiające się na regionie zawierającym gen *PTCH1* (chr9q22.32) i regionie obejmującym *PDL1*, *PDL2* i *JAK2* (chr9p24.1). Test składa się z 20 sond, w tym (i) 7 sond rozmieszczonych wzdłuż ramienia chr9p (5 sond) i chr9q (2 sondy), ii) po 2 sondy zlokalizowane w samym genie lub w pobliżu genów *JAK2*, *PDL2*, *PDL1* i *PTCH1* (łącznie 8 sond) oraz iii) 5 sond kontrolnych (zlokalizowanych na różnych chromosomach poza chromosomem 9 i regionami genów związanych z rakiem) (Ryc. 35., Tab.11 w sekcji Materiały i Metody).

Analiza MLPA potwierdziła zmianę liczby kopii (wykrytą wcześniej przez GISTIC 2.0) we wszystkich analizowanych próbkach (przykładowe analizy pokazano na Ryc. 35A-C, wszystkie wyniki umieszczono w Załączniku 5). Całościowo zmiany liczby kopii analizowanych regionów wykryto w 9 próbkach BCC. Spośród nich 5 próbek wykazało zarówno amplifikację regionu obejmującego geny *PDL1*, *PDL2* oraz *JAK2*, jak i delecję *PTCH1*. Pozostałe 4 próbki posiadały wyłącznie delecję *PTCH1*.

## WYNIKI



**Rycina 35. Zmiany liczby kopii w BCC na podstawie analiz MLPA.**

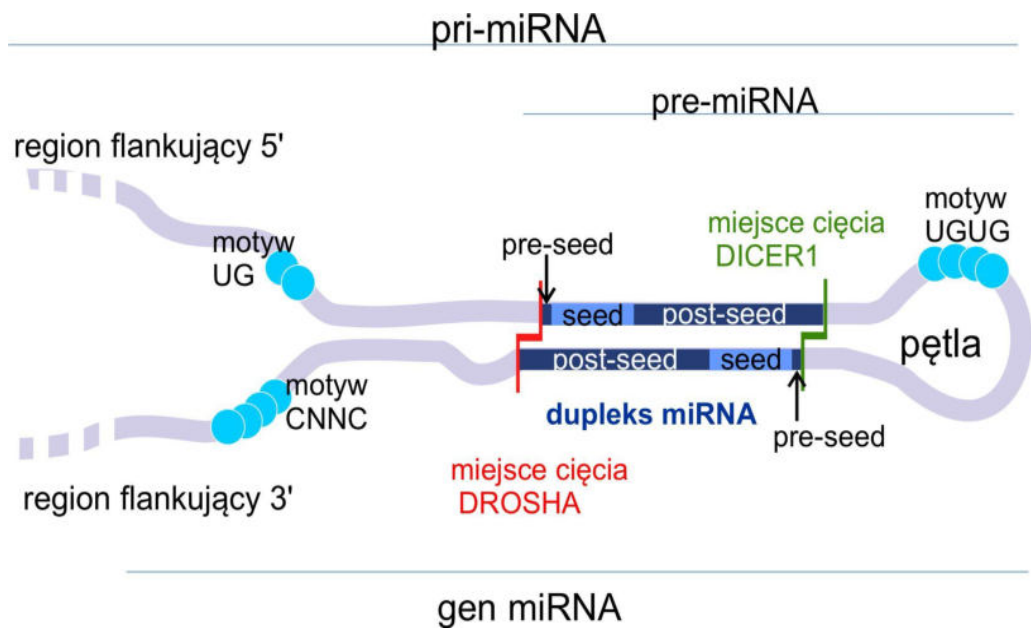
Rycina przedstawia liczbę kopii genów i regionów zlokalizowanych na chromosomie 9. A) Brak zmian liczby kopii w analizowanej próbce, B) Amplifikacja ramienia 9p oraz delecja ramienia 9q, C) Delecja ramienia 9q. Lokalizację sond przedstawiono na schemacie na dole wykresu.

## 2. Sekwencjonowanie całego miRNomu (WMS)

### 2.1 Panel do sekwencjonowania genów miRNA

W celu wykrycia mutacji somatycznych w genach miRNA, DNA poddano autorskiej procedurze wzbogacania oraz sekwencjonowania WMS. Panel WMS zaprojektowano w Zakładzie Genetyki Molekularnej ICHB PAN jednak jego opracowanie nie jest częścią niniejszej pracy doktorskiej. Panel umożliwia sekwencjonowanie wszystkich genów miRNA oraz dodatkowo szeregu genów kodujących białka biorące udział w biogenezie miRNA. Listę genów miRNA wchodzącą w skład panelu miRNom wyselekcjonowano na podstawie bazy danych miRBase v21 (obejmującej 1881 ludzkich miRNA), sondy do wzbogacenia zaprojektowano dla 1849 genów miRNA (wyselekcjonowanych przez oprogramowanie SureCall, które odfiltrowało geny miRNA pokrywające się z sekwencjami repetytywnymi). W tej pracy, geny miRNA zdefiniowano jako sekwencje kodujące najbardziej kluczowe i dobrze rozpoznane sekwencje prekursorów miRNA obejmujące sekwencje pre-miRNA (~60nt) oraz bezpośrednio do nich przylegające 25 nukleotydowe sekwencje flankujące powyżej i poniżej pre-miRNA (region flankujący 5' oraz 3'). Podobną definicję genów miRNA, ze względów praktycznych, coraz częściej stosuje się w literaturze naukowej (Galka-Marciniak et al., 2019, 2022; Machowska, Galka-Marciniak, & Kozłowski, 2022; Urbanek-Trzeciak et al., 2020). Na strukturze przedstawiającej gen miRNA wyróżniono również regiony zawierające sekwencje dojrzałych cząsteczek miRNA w tym regiony źródłowe (ang. *seed*) oraz duplekisy przed i za regionem źródłowym (ang. *pre-seed*; *post-seed*) (Ryc.36).





**Rycina 36. Schemat prekursora miRNA.**

Pri-miRNA składa się z pre-miRNA oraz sekwencji flankujących 5' oraz 3'. Gen miRNA obejmuje sekwencję kodującą cząsteczkę pre-miRNA oraz regiony flankujące 5' oraz 3' o długości 25 nt. Dojrzały dupleks miRNA zaznaczono kolorem granatowym między miejscem cięcia enzymami DROSHA (kolor czerwony) i DICER1 (kolor zielony). Sekwencję źródłową zaznaczono kolorem jasnoniebieskim. Motywy sekwencyjne charakterystyczne dla miRNA oznaczono niebieskimi kropkami.

Dane z sekwencjonowania NGS poddane zostały wieloetapowemu procesowi analizy (szczegóły w sekcji Materiały i Metody) w celu uzyskania listy mutacji somatycznych.

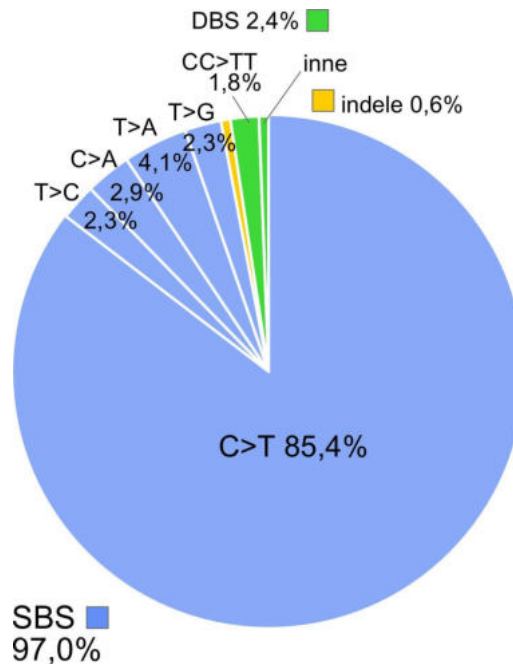
W wyniku przeprowadzonej analizy, łącznie wykryto 171 somatycznych mutacji w 155 genach miRNA. Średnie pokrycie regionów celowanych wynosiło 699x (700x dla próbek nowotworowych oraz 698,5x dla próbek kontrolnych), wahając się od 394x do 1177x. Średnie pokrycie mutowanych pozycji wynosi 1084,3x (15x - 3130x), a średnia wartość AAF wynosi 0,19.

## 2.2 Lokalizacja i typy mutacji w genach miRNA

WMS umożliwiło zidentyfikowanie 171 mutacji somatycznych w 155 genach miRNA w 22 próbkach (81,5%) BCC. Średnia ilość mutacji na próbkę wynosi 6,3 mutacje. Zdecydowaną większość wszystkich mutacji stanowią SBS (166 (97,0%)), następnie DBS (4 (2,4%)), i indele (1 (0,6%)). Analiza pojedynczych substytucji wykazała największy udział (85,4%) tranzycji C>T wśród wszystkich wykrytych mutacji (Ryc. 37). Spośród podwójnych substytucji najliczniejszą

## WYNIKI

grupę stanowią substytucje CC>TT (1,8% wszystkich mutacji). Oba typy mutacji są charakterystyczne dla nowotworów, których rozwój jest związany z działaniem promieniowania słonecznego.



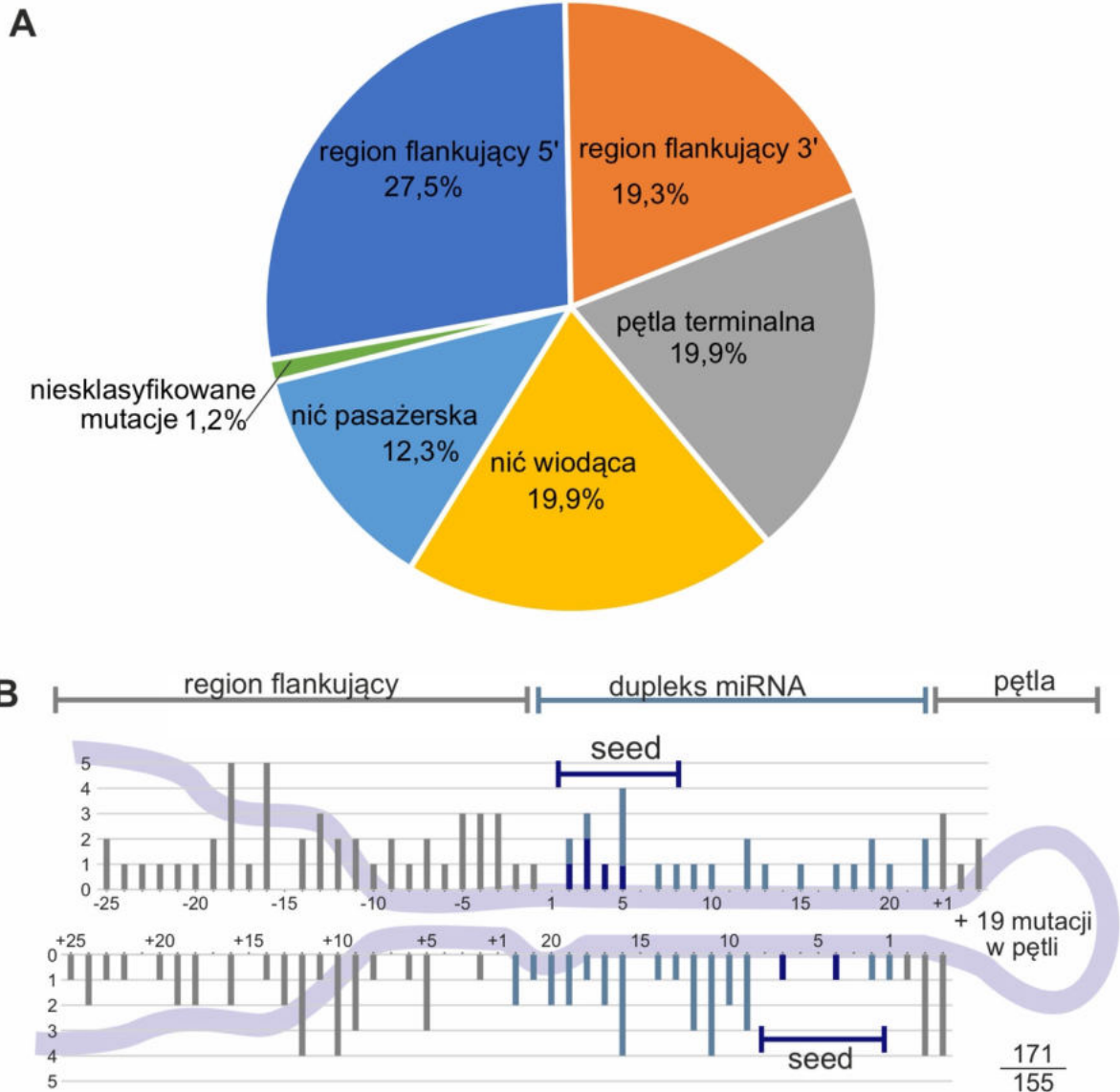
**Rycina 37. Typy mutacji występujące w genach miRNA w BCC.**

Niebieskim kolorem oznaczono substytucje pojedynczych nukleotydów, zielonym podwójne substytucje, a żółtym indele.

Analiza lokalizacji mutacji w poszczególnych subregionach prekursorów miRNA wykazała, że 47 mutacji (27,5%) jest zlokalizowanych w regionie flankującym 5', 34 (19,9%) w nici wiodącej z czego 7 w (4,1%) w regionie źródłowym (seed), 21 (12,3%) w nici pasażerskiej, 34 (19,9%) w pętli terminalnej oraz 33 (19,3%) w regionie flankującym 3' (Ryc. 38A). W przypadku 2 mutacji niemożliwe było określenie lokalizacji (mutacje niesklasyfikowane, Ryc. 38A) ze względu na ich lokalizację w regionie dupleksu miRNA w genach (prekursorach), dla których nie została zdefiniowana nić wiodąca (miRBase).

Mutacje lokalizują się we wszystkich subregionach genów miRNA. W celu dokładniejszej analizy lokalizacji mutacji w subregionach prekursorów miRNA mutacje zwizualizowano na ogólnej strukturze prekursora (analiza z zastosowaniem miRMut; Ryc. 38B).

## WYNIKI

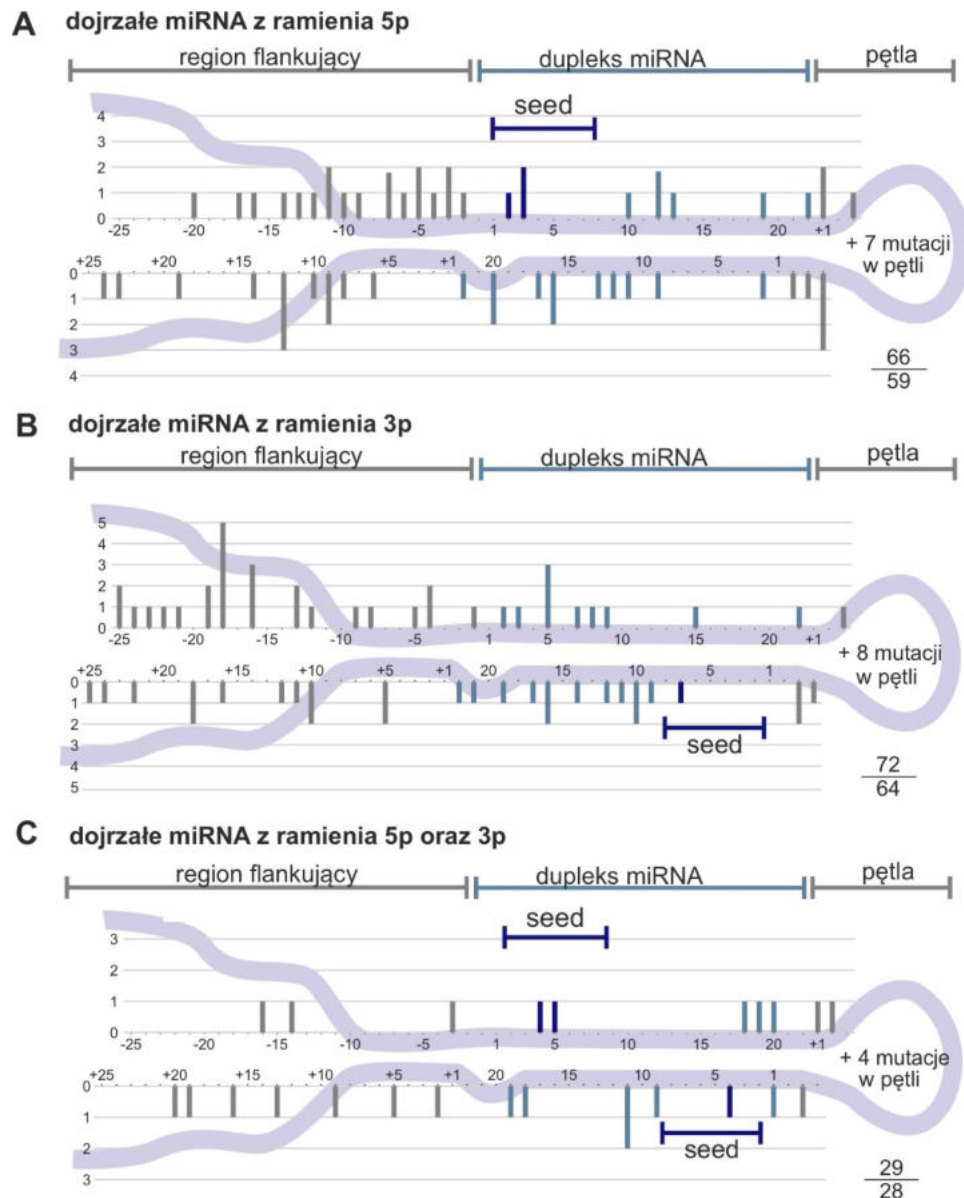


### Rycina 38. Lokalizacja mutacji w subregionach prekursorów miRNA.

A) Rozkład mutacji (wyrażony w %) wykrytych w BCC w poszczególnych częściach prekursorów miRNA; B) Rozkład mutacji wzdłuż wszystkich prekursorów miRNA. Mutacje w duplekcie miRNA oznaczono kolorem niebieskim, w regionie źródłowym zaznaczono kolorem granatowym, a w regionach flankujących oraz początkowe i końcowe pozycje pętli terminalnej oznaczono kolorem szarym. Liczby w prawym dolnym rogu reprezentują łączną liczbę mutacji (wartość na górze) oraz liczbę zmutowanych genów miRNA (wartość na dole). W przypadku, gdy dojrzała cząsteczka miRNA jest dłuższa niż zaznaczone na wykresie 22 nt mutacje w pozycjach powyżej 22 zaznaczono kumulatywnie w pozycji 22. Ponieważ wielkość i struktura pętli różnią się istotnie pomiędzy genami miRNA, na rycinie nie przedstawiono mutacji zlokalizowanych wewnątrz pętli. Wykres pokazuje mutacje w obrębie 6nt pętli (pierwsze i ostatnie 3 nt pętli). Liczbę pozostałych mutacji wpisano wewnątrz pętli.

## WYNIKI

W celu weryfikacji czy rozkład mutacji w miRNA może różnić się ze względu na ramię (zawierające nić wiodącą lub pasażerską) z którego powstaje dojrzała cząsteczka, prekursorzy miRNA sklasyfikowano do jednej z 3 kategorii na podstawie informacji z bazy danych miRBase v.21 na temat poziomu miRNA generowanych z poszczególnych ramion: (i) prekursorzy generujące dojrzałe miRNA głównie z ramienia 5p ( $\geq 90\%$  odczytów z ramienia 5'); (ii) prekursorzy generujące dojrzałe miRNA głównie z ramienia 3p ( $\geq 90\%$  odczytów z ramienia 3'); (iii) zrównoważone prekursorzy generujące dojrzałe miRNA zarówno z ramienia 3p jak i 5p ( $>10\%$  odczytów z obu ramion). Następnie za mutacje w nici wiodącej uznano mutacje lokalizujące się w sekwencji źródłowej, pre-seed oraz post-seed tylko z tego ramienia, z którego uwalniane są dojrzałe miRNA. W przypadku 2 mutacji niemożliwe było określenie lokalizacji na nici wiodącej lub pasażerskiej ze względu na brak informacji w bazie miRBase o ramieniu, z którego powstaje dojrzała cząsteczka miRNA. Analizy wykazały podobny rozkład częstości mutacji wzdłuż prekursora zarówno dla tych prekursorów, z których dojrzałe miRNA powstają z ramienia 5p (Ryc. 39A; 66 mutacji w 59 prekursorach, sekwencja źródłowa obecna wyłącznie w ramieniu 5p), 3p (Ryc. 39B; 72 mutacje w 64 prekursorach, sekwencja źródłowa zlokalizowana w ramieniu 3p), oraz z obu ramion (Ryc. 39C; 29 mutacji w 28 prekursorach, sekwencja źródłowa w ramieniu 3p oraz 5p).



**Rycina 39. Rozkład mutacji w subregionach prekursorów miRNA z zastosowaniem podziału ze względu na ramię, z którego powstaje dojrzała cząsteczka.**

Schemat ryciny analogicznie do Ryc.38B A) Rozkład mutacji wzdłuż prekursorów miRNA generujących dojrzałą cząsteczkę miRNA z ramienia 5p; B) Rozkład mutacji wzdłuż prekursorów miRNA generujących dojrzałą cząsteczkę miRNA z ramienia 3p; C) Rozkład mutacji wzdłuż prekursorów miRNA generujących dojrzałe cząsteczki miRNA z obu ramion.

Analiza częstości mutacji w poszczególnych subregionach prekursorów miRNA nie wykazała żadnych istotnych różnic (Tabela 8). Zaobserwowano jedynie niewielką tendencją do niższej częstości mutacji w regionie źródłowym (FC=0,58; p=0,16530), oraz wyższą częstość mutacji w regionie pętli terminalnej (FC=1,22; p=0,22445) oraz regionie flankującym 5' (FC=1,22; p=0,11746), jednak żadna z tych zmian nie była istotna statystycznie, najprawdopodobniej z powodu zbyt małej liczby mutacji wykrytych w eksperymencie.

**Tabela 8. Rozkład mutacji w prekursorach miRNA.**

Region	Liczba mutacji	mutacje/Mpz	FC	wartość p (rozkład dwumianowy)
cały gen miRNA	171	29,77	1,00	-
region flankujący 5'	47	36,47	1,22	0,11746
region flankujący 3'	33	25,61	0,86	0,42155
pętla	34	36,22	1,22	0,22445
nić wiodąca	34	27,07	0,91	0,65951
nić pasażerska	21	22,79	0,77	0,23558
region źródłowy	7	17,19	0,58	0,16630

### 2.3 Dodatkowa walidacja mutacji metodą sekwencjonowania Sangera

W celu ustalenia liczby mutacji fałszywie pozytywnych wybrane mutacje wykryte z użyciem NGS, zostały dodatkowo potwierdzone sekwencjonowaniem metodą Sangera. Analizie poddano 15 mutacji SBS. Potwierdzono 14 z 15 mutacji. Nie potwierdzono jednej mutacji co może wskazywać na poziom 6,6% mutacji fałszywie pozytywnych (Tabela 9), jednak wartość ta może być niższa, ponieważ niepotwierdzona mutacja (substytucja C>T w *MIR3928*) charakteryzuje się niskim AAF=0,12.

## WYNIKI

**Tabela 9. Lista mutacji walidowanych metodą sekwencjonowania Sangera.**

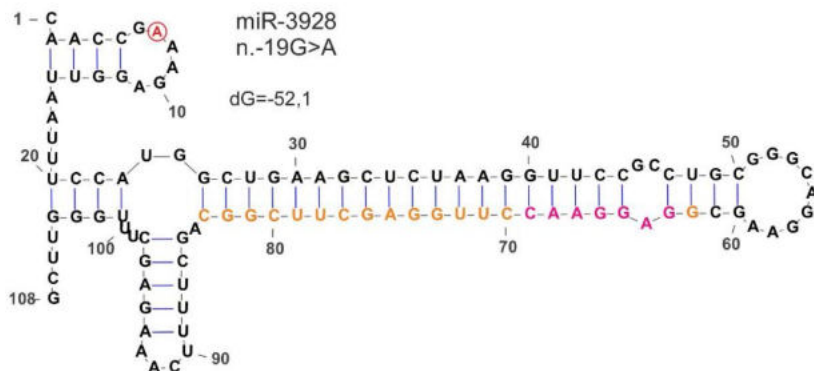
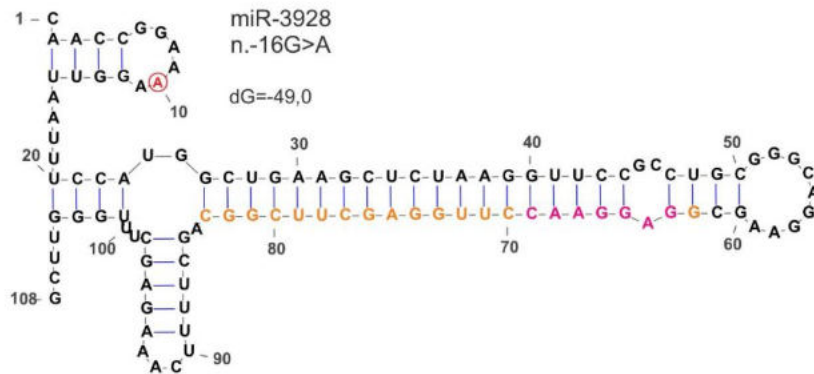
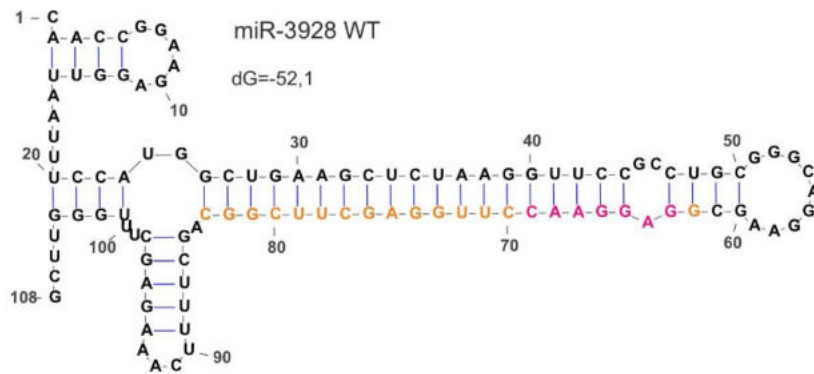
Rodzaj mutacji	Nazwa genu [orientacja genu]	Lokalizacja mutacji w genomie	Lokalizacja mutacji w prekursorze miRNA	AAF	Liczba analizowanych próbek	Walidacja
C>T	<i>MIR3928</i> [-]	chr22:31,160,135	n.-16G>A	0,12; 0,28; 0,68	3	Tak (2)/ Nie (1)
C>T	<i>MIR3928</i> [-]	chr22:31,160,138	n.-19G>A	0,2	1	Tak
T>C	<i>MIR3612</i> [+]	chr12:128,294,142	n.51T>C	0,37; 0,32	2	Tak
C>T	<i>MIR448</i> [+]	chrX:11,4823,502	n.49C>T	0,34; 0,3	2	Tak
G>A	<i>MIR7108</i> [-]	chr19:2,434,934	n.67C>T	0,33	1	Tak
G>A	<i>MIR7108</i> [-]	chr19:2,434,947	n.54C>T	0,29	1	Tak
G>A	<i>MIR6796</i> [+]	chr19:40,369,887	n.41G>A	0,41	1	Tak
G>A	<i>MIR409</i> [+]	chr14:10,1065,311	n.12G>A	0,17	1	Tak
C>T	<i>MIR409</i> [+]	chr14:101,065,383	n.+8C>T	0,16	1	Tak
C>T	<i>MIR6507</i> [-]	chr10:98,924,575	n.-7G>A	0,14	1	Tak

## 2.4 Często mutowane geny miRNA, gorące punkty mutacji i geny MIRNA napędzające rozwój BCC

Analiza często mutowanych genów miRNA (mutacje w  $\geq 3$  próbek - 10%) wykazała częste mutacje tylko w 1 genie miRNA: *MIR3928*. W *MIR3928* zidentyfikowano 4 mutacje w 4 próbkach BCC. Wszystkie mutacje są zlokalizowane w sekwencji flankującej 5'. 3 próbki posiadają mutację zlokalizowaną w pozycji chr22:31,160,135C>T[-] (n.-16G>A) tworząc gorące miejsce mutacji, a 1 próbka posiada mutację w pozycji chr22:31,160,138C>T[-] (n.-19G>A). Gorący punkt mutacji wykryty w prekursorze miR-3928 jest jedynym gorącym punktem wykrytym w genach miRNA. MiR-3928 jest określony w bazie miRBase jako miRNA o wysokiej wiarygodności, natomiast nie został uwzględniony w bazie MiRGeneDB. Przeprowadzona analiza stabilności struktury typu spinki do włosów tworzonej przez *MIR3928* wykazała nieznaczny spadek stabilności struktury 2D w wyniku mutacji n.-16G>A (spadek stabilności o  $\sim 3$  kcal/mol) oraz brak zmian stabilności w wyniku mutacji n.-19G>A. Modelowanie struktury przestrzennej (3D), przy użyciu programu RNAComposer nie wykazało znaczącego wpływu mutacji na strukturę prekursora (Ryc. 40).



## WYNIKI



### Rycina 40. Wpływ mutacji na strukturę prekursora miR-3928.

Efekt mutacji pokazany jest na poziomie struktur drugorzędowych 2D oraz 3D tworzonych przez prekursor miRNA. Pomarańczowym kolorem zaznaczono sekwencję dojrzałej, wiodącej cząsteczki miRNA a różowym sekwencję źródłową. Czerwonym kółkiem zaznaczono miejsce występowania mutacji. Zmutowane struktury 3D (struktura zmutowanego prekursora - kolor zielony, mutacja - kolor czerwony) są nałożone na odpowiadające im struktury typu dzikiego (kolor czarny).

## WYNIKI

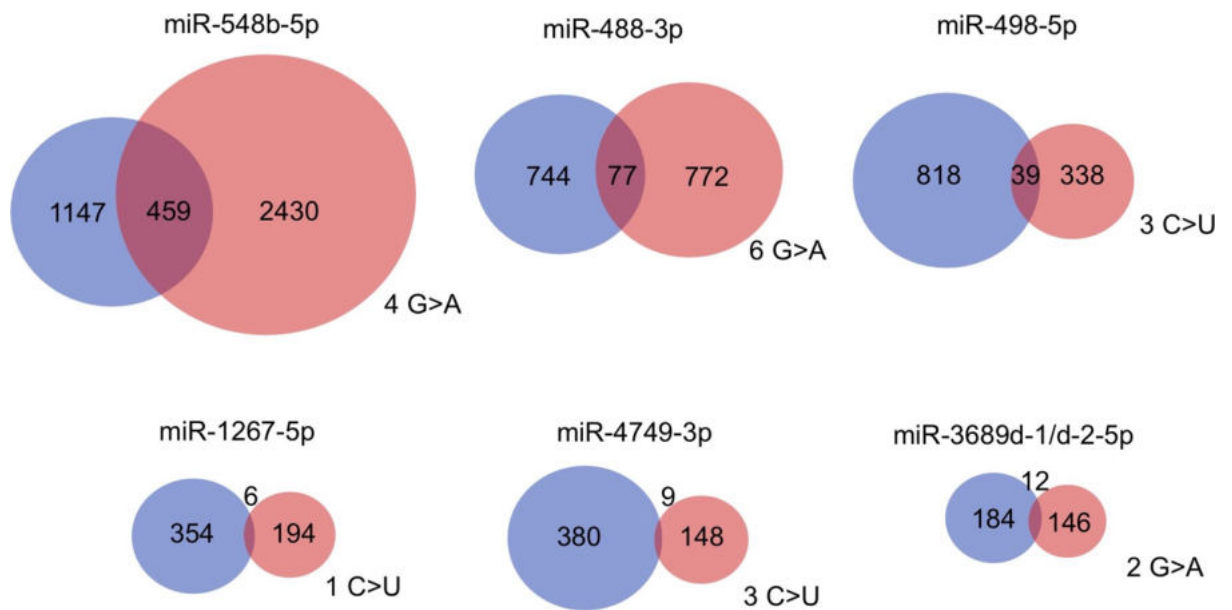
Aktualne doniesienia literaturowe wskazują zarówno na onkogenny jak i supresorowy potencjał miR-3928 w raku płuca, glejaku, czy nowotworach głowy i szyi (Fadhil, Wei, Nikolarakos, Good, & Nair, 2020; Mulcahy et al., 2022; Xia et al., 2018).

Analiza mutacji wykrytych w miRNomie obejmowała także analizę genów napędzających (z zastosowaniem OncodriveFML) na podstawie mutacji w genach miRNA. Analiza nie wykazała jednak żadnego genu miRNA o potencjale napędzającym.

### 2.5 Analiza mutacji w regionie źródłowym

W celu weryfikacji wpływu mutacji w regionie źródłowym na pulę regulowanych genów docelowych przeprowadzono analizę z wykorzystaniem narzędzia TargetScan. Całościowo w analizowanych próbkach BCC wykryto 7 mutacji zlokalizowanych w regionie źródłowym. Generalnie, mutacje w regionie źródłowym skutkują znaczącą zmianą puli potencjalnie regulowanych genów docelowych. W wyniku mutacji w regionie źródłowym dochodzi zarówno do spadku jak i wzrostu ilości regulowanych genów (Ryc. 41). Przykładowo mutacja w czwartej pozycji sekwencji źródłowej w miR-548b-5p skutkująca substytucją G>A powoduje 1,8-krotne zwiększenie puli regulowanych genów. W przypadku mutacji C>U w miR-4749-3p w trzeciej pozycji sekwencji źródłowej dochodzi do 2,5-krotnego zmniejszenia puli regulowanych genów. Największą liczbę nakładających się genów docelowych między dziką i zmutowaną formą miRNA wykryto dla miR-548b-5p (459 genów, 10,2%), który posiada mutację G>A w czwartej pozycji regionu źródłowego, a najmniejszą liczbę nakładających się genów (6 genów, 1,1%) wykryto dla miR-1267-5p posiadającego mutację w pierwszej pozycji regionu źródłowego.

## WYNIKI



**Rycina 41. Wpływ mutacji w regionie źródłowym na liczbę przewidzianych genów docelowych.**

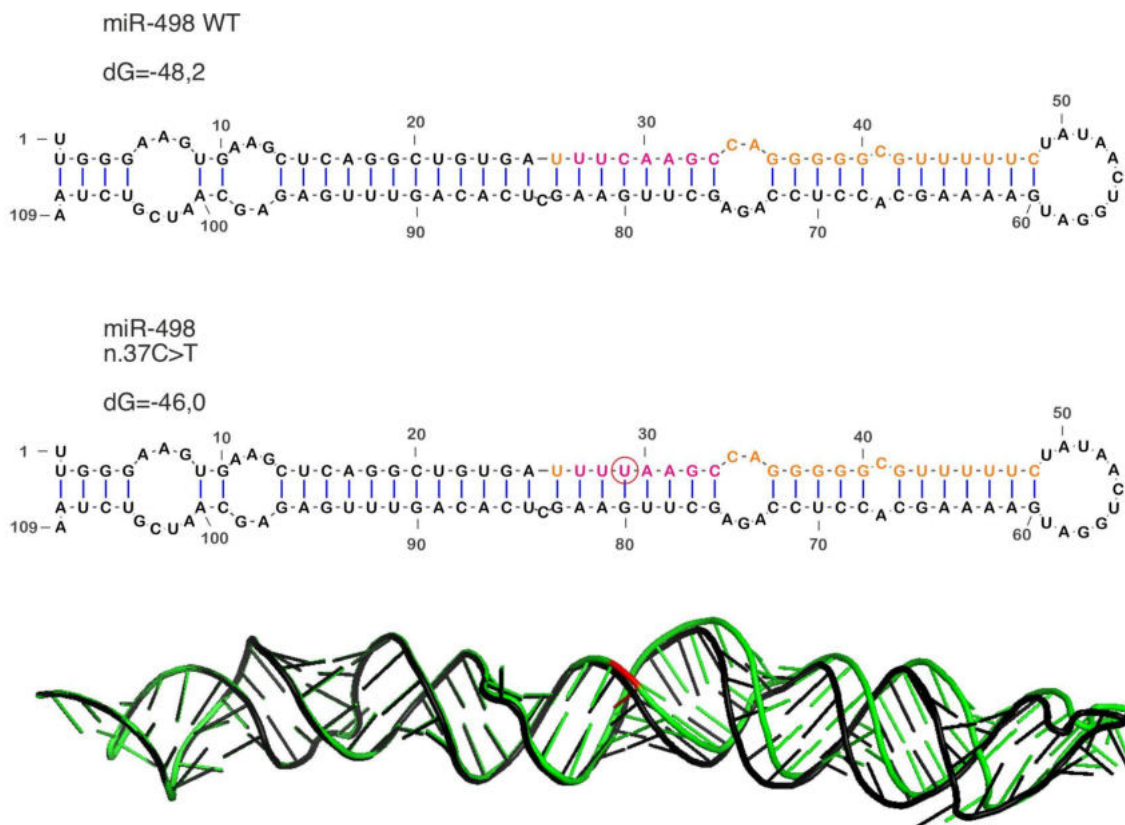
Diagramy przedstawiają liczbę przewidywanych transkryptów wiążących dziką sekwencję miRNA (kolor niebieski), liczba transkryptów, do których wiąże się zmutowana cząsteczka miRNA (kolor czerwony). Mutowana pozycja w regionie źródłowym oraz rodzaj mutacji są wskazane po prawej stronie każdego diagramu. Wpływ mutacji 2 G>A występującej w miR-3689d-1 oraz miR-3689d-2 został zaznaczony na jednym diagramie w związku z taką samą sekwencją źródłową tych miRNA.

Część mutacji w regionie źródłowym wykryto w genach miRNA o wysokiej wiarygodności oraz dobrze poznanej funkcji w rozwoju nowotworów skóry. Mutację wykryto m.in. w trzecim nukleotydzie regionu źródłowego w pozycji chr19:53,674,233C>T[+] (n.37C>T) w *MIR498*.

Analiza z użyciem TargetScan wykazała 857 genów posiadających w regionie 3'UTR miejsce wiązania miR-498-5p. Mutacja drastycznie zmienia pulę przewidzianych genów potencjalnie regulowanych przez miR-498-5p i skutkuje zmniejszeniem ich do 377, a liczba genów pokrywających się z genami potencjalnie regulowanymi przez miRNA bez mutacji to 39 (Ryc. 42). Dodatkowo, występowanie mutacji skutkuje nieznacznym spadkiem stabilności struktury 2D tworzonej przez prekursor miRNA (spadek stabilności o ~2 kcal/mol). Wykryta mutacja nie wpływa jednak na zmianę struktury 3D prekursora. Przegląd literaturowy wykazał, zarówno onkogenny jak i supresorowy potencjał miR-498. W raku prostaty oraz raku piersi miR-498 wykazuje działanie onkogenne i obniża ekspresję genu supresorowego *PTEN* (Chai, Wu, Wang, Eisenstat, & Leng, 2018; Duan et al., 2019). W czerniaku miR-498 działa jako

## WYNIKI

supresor nowotworzenia, prowadząc do obniżenia ekspresji genu *UBE2T*, co spowalnia rozwój nowotworu (Cao et al., 2022). MiR-498 pełni także funkcję supresora rozwoju raka wątroby regulując ekspresję *CREB1* (Zhu, Gao, & Ren, 2022) oraz spowalnia rozwój raka wątroby obniżając ekspresję genu docelowego *ZEB2* (Zhang et al., 2019). Mutacje n.37C>T skutkuje utratą regulacji wyżej wymienionych genów docelowych.



**Rycina 42. Wpływ mutacji na strukturę prekursora miR-498 oraz analiza regionów wiążących miRNA.**

Schemat ryciny analogicznie do Ryc.40. Mutację oznaczono względem transkryptu prekursora miR-498 wg bazy miRBase (n.37C>T). W związku z krótszą sekwencją uznaną w niniejszej pracy za gen miRNA mutacja ta znajduje się w pozycji 29 na pokazanej strukturze.

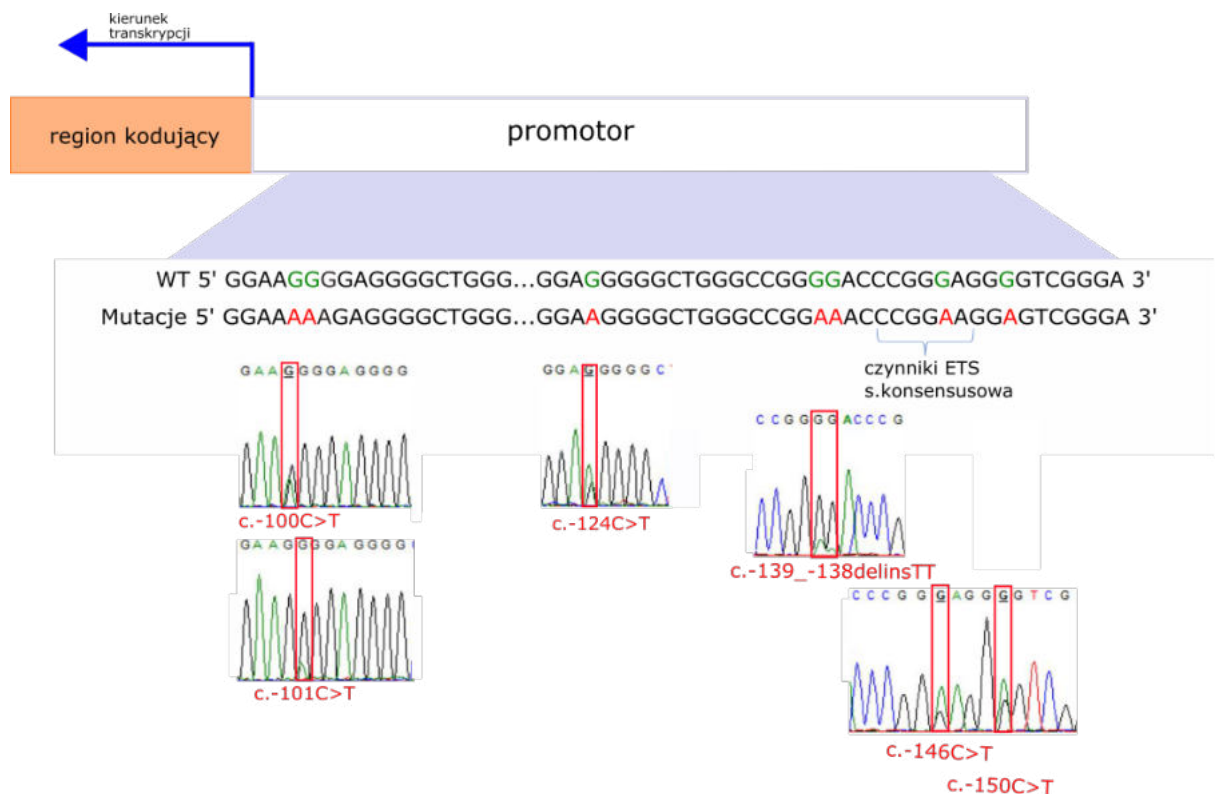
### 3. Mutacje w promotorach genów *TERT* i *DPH3*

Jedynie niekodujące mutacje badane wcześniej w BCC to mutacje w promotorach genów *TERT* i *DPH3* (Griewank et al., 2013; Maturro et al., 2020). Ponieważ regiony te nie były pokryte przez panel Exome V5 ani miRNom, w celu wykrycia mutacji w tych regionach w próbkach analizowanych w niniejszej pracy, przeprowadzono sekwencjonowanie Sangera. W

## WYNIKI

rezultacie wykryto 16 mutacji u 11 (41%) pacjentów w promotorze *TERT* i 6 mutacji u 5 (19%) pacjentów w promotorze *DPH3*.

Wszystkie mutacje *TERT* wykryto w pozycjach genomowych wcześniej opisanych w literaturze, obejmujących mutacje typu SBS lokalizujące się w gorących punktach. Całościowo zidentyfikowano 2 podwójne substytucje c.-139\_-138delinsTT, 9 substytucji c.-146C>T, 2 substytucje c.-101C>T, 3 inne podstawienia (c.-150C>T, c.-124C>T oraz c.-100C>T) (Ryc. 43, Załącznik 6). Najczęściej występująca mutacja c.-146 skutkuje utworzeniem motywu CCGGA/T wiążącego czynniki transkrypcyjne z rodziny ETS m.in. GABP-alfa, ELK1 oraz ELK4 i prowadzi do zwiększenia ekspresji *TERT* (Min & Shay, 2016).



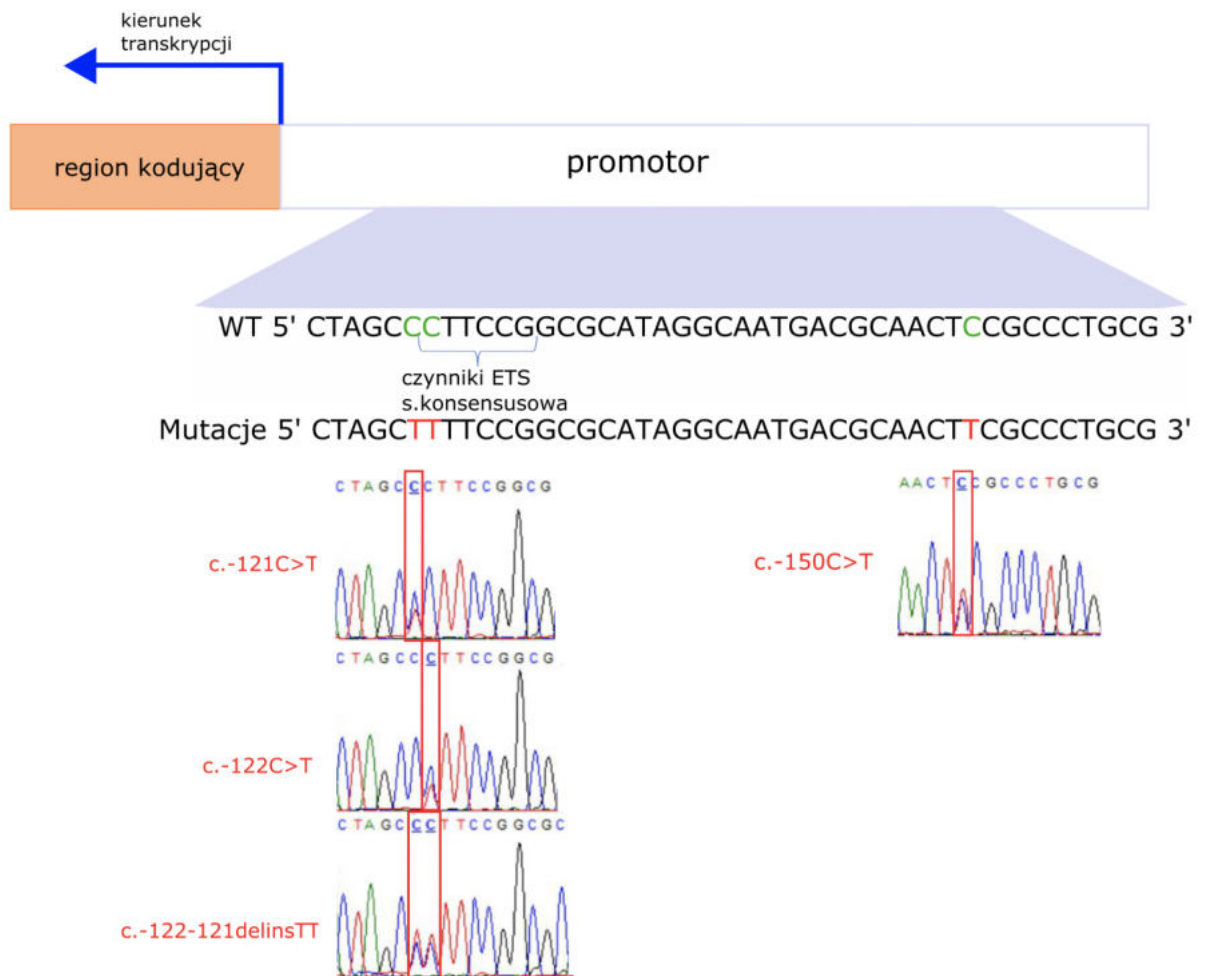
### **Rycina 43. Mutacje w promotorze genu *TERT* zidentyfikowane w niniejszej pracy w próbkach BCC.**

Kolorem zielonym w sekwencji dzikiej promotora zaznaczono miejsca ulegające mutacjom, a czerwonym substytucje, które zachodzą w mutowanych pozycjach. Sekwencję konsensusową wiążącą czynniki transkrypcyjne oznaczono klamrą. Elektroferogramy pokazujące wykryte mutacje rozmieszczono zgodnie z sekwencją promotora.

W przypadku *DPH3* mutacje znajdowały się w pozycjach wcześniej opisanych w literaturze, obejmując 3 podwójne podstawienia c.-122\_-121delinsTT i 3 pojedyncze

## WYNIKI

substytucje (c.-150C>T, c.-122C>T i c.-121C>T) (Ryc. 44, Załącznik 7). Mutacja w pozycji c.-122 zaburza sekwencję kolejnego motywu CTTCCG wiążącego czynniki transkrypcyjne z rodziny ETS co skutkuje najprawdopodobniej zmniejszoną ekspresją *DPH3*. Wcześniejsze badania dotyczące funkcjonalnego wpływu mutacji c.-121/122 nie wykazały jednak jednoznacznego wpływu mutacji na poziom ekspresji genu *DPH3* (Denisova et al., 2015).



**Rycina 44. Mutacje w promotorze genu *DPH3* zidentyfikowane w niniejszej pracy w próbkach BCC.**

Schemat ryciny analogicznie do Ryc.43.

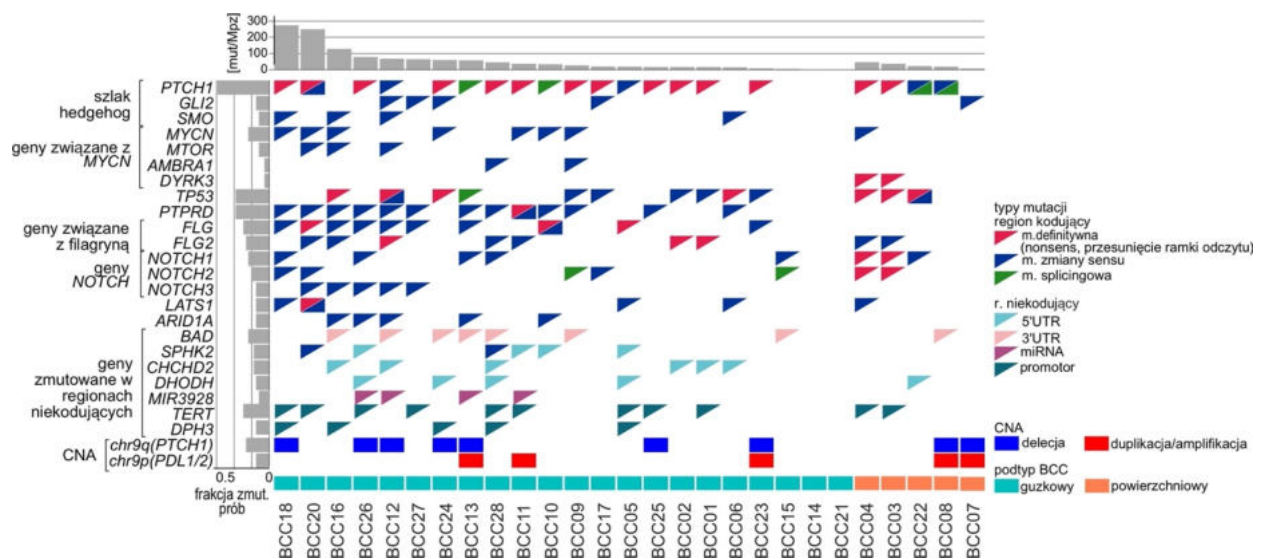
## VII. DYSKUSJA

Częstość zachorowań na nowotwory skóry nieustannie wzrasta, co stanowi znaczący problem epidemiologiczny, dlatego badania mające na celu lepsze zrozumienie mechanizmów związanych z rozwojem BCC, który jest najczęściej występującym nowotworem skóry, są niezwykle istotne z punktu widzenia współczesnej medycyny. Pomimo licznych badań patogeneza BCC wciąż nie jest do końca poznana. Dotychczasowe analizy BCC jak i większości nowotworów skupiały się na analizie regionów kodujących genomu. Wiadomo jednak, że niekodujące części genów tj. promotory, introny, 5'UTR, 3'UTR, a także inne regiony niekodujące np. geny miRNA obejmują liczne elementy funkcjonalne/regulatorowe ważne dla prawidłowego funkcjonowania genów. W ostatnich latach pojawiły się badania, które uwzględniają analizy tych regionów w niektórych typach nowotworów oraz wskazują na napędzającą rolę pojedynczych mutacji niekodujących.

W niniejszej pracy doktorskiej, aby kompleksowo scharakteryzować mutacje somatyczne zlokalizowane w różnych częściach genomu, zastosowano zarówno standardowe sekwencjonowanie typu WES jak i nowatorską strategię sekwencjonowania WMS. Sekwencjonowanie typu WES jest powszechną/standardową metodą służącą do analizy głównie kodującej części genomu. Pomimo, że niektóre jego wersje obejmują także część regionów niekodujących związanych z genami kodującymi białka (znaczną część 3'UTR i 5'UTR oraz intronów) mutacje w tych regionach są zwykle pomijane w trakcie analiz ze względu na brak jasnych reguł, według których możliwe jest określenie funkcjonalnych konsekwencji mutacji w tych regionach. Badania regionów niekodujących przeprowadzone w niniejszej pracy wykonano z zastosowaniem szeregu narzędzi bioinformatycznych, wyselekcjonowanych pod kątem możliwie jak najlepszego ich dostosowania do analiz sekwencji niekodujących.

Analiza wyników dla 27 próbek BCC pozwoliła na wykrycie dziesiątek tysięcy mutacji somatycznych występujących z podobną częstością zarówno w kodującej jak i niekodującej części genomu. Podsumowanie najważniejszych wyników, tj. występowanie mutacji w najczęściej mutowanych genach kodujących i niekodujących oraz zmian liczby kopii przedstawione jest na Ryc. 45.

## DYSKUSJA



**Rycina 45. Geny, promotory i/lub szlaki często mutowane w BCC.**

Typy mutacji występujące w genach często mutowanych (oś y) w podtypach BCC (oś x) oznaczono według kolorów podanych w legendzie po prawej stronie wykresu. Obciążenie mutacyjne próbek umieszczono na górze wykresu (mut/Mpz).

Wiele zidentyfikowanych niekodujących gorących punktów mutacji było zlokalizowanych w sekwencjach genów funkcjonalnie związanych z rakiem. W niniejszej pracy zidentyfikowano trzy geny pełniące funkcje związane z aktywnością mitochondriów, posiadające gorące punkty mutacji w regionach niekodujących oraz zidentyfikowane jako potencjalne geny napędzające: *BAD*, *DHODH* i *CHCHD2*. W związku z faktem, że geny te posiadają mutacje wyłącznie w regionie niekodującym, nie zostały zidentyfikowane jako często mutowane w BCC ani innych nowotworach na podstawie wcześniejszych analiz (Jayaraman et al., 2014; Bonilla et al., 2016).

Spośród wszystkich gorących punktów mutacji wykrytych w niniejszej pracy najczęściej mutowany był punkt zlokalizowany w 3'UTR genu *BAD* (sklasyfikowany także jako gen często mutowany oraz napędzający w BCC) posiadający substytucje występujące w 2 sąsiadujących pozycjach genomowych (142 i 143 nt poniżej kodonu stop). *BAD* koduje białko *BAD* należące do rodziny BCL-2, składającej się zarówno z białek proapoptotycznych, jak i antyapoptotycznych. *BAD* promuje śmierć komórki poprzez indukcję permeabilizacji mitochondrialnej błony zewnętrznej (ang. *mitochondrial outer membrane permeabilization*; MOMP), umożliwiając uwalnianie cytochromu c oraz antagonizowanie antyapoptotycznych białek BCL-2 (Datta et al., 1997; Yang et al., 1995). Z drugiej strony ufosforylowane *BAD* mogą



## DYSKUSJA

mieć również właściwości antyapoptotyczne i sprzyjać m.in. przeżywalności melanocytów (Sastry, Ibrahim, & Chouchane, 2020; Sastry et al., 2014). Inne funkcje BAD obejmują regulację metabolizmu mitochondriów (regulacja zależnych od napięcia kanałów anionowych i przejścia metabolitów przez zewnętrzną błonę mitochondrialną) oraz dynamikę mitochondriów (regulacja zmiany kształtu) (Aouacheria et al., 2017; Berman et al., 2009; Danial et al., 2003, 2008; Roy et al., 2009; Seo et al., 2004; Vander Heiden et al., 2001). Chociaż funkcja BAD nie była wcześniej opisana jako związana z nowotworami skóry, wykazano, że utrata lub zmniejszenie innych proapoptotycznych białek rodziny BCL-2, tj. BAX i PUMA, sprzyja rozwojowi BCC, SCC i czerniaka (Karst, Dai, Martinka, & Li, 2005; Tomková, Fujimoto, & Arata, 1998), dlatego podobny wpływ na rozwój nowotworów skóry mogą wywołać mutacje powodujące spadek aktywności BAD. Przeprowadzona w niniejszej pracy analiza porównawcza z TCGA wykazała gorący punkt w BCC jako często mutowany także w czerniaku, ale nie w innych nowotworach. Wyłączność mutacji w 3'UTR w czerniaku i BCC zmniejsza prawdopodobieństwo przypadkowego wystąpienia mutacji i przemawia za napędzającą funkcją mutacji w BCC i czerniaku (i prawdopodobnie również w innych nowotworach związanych z promieniowaniem UV).

Wykazano, że gen *CHCHD2* jest często mutowany w BCC w regionie 5'UTR, posiada gorące punkty mutacji w pozycjach c.-77C>T oraz c.-134G>A i został sklasyfikowany jako potencjalny gen napędzający. Ponadto wykazano, że *CHCHD2* jest często mutowany nie tylko w BCC, ale także w czerniaku (analiza TCGA). Mutacje w czerniaku występują nie tylko w gorącym punkcie c.-77C>T, ale także w dodatkowym punkcie c.-74C>T charakterystycznym wyłącznie dla czerniaka. Funkcja genu *CHCHD2* w nowotworzeniu nie była intensywnie badana, ale wykazano, że w warunkach hipoksji *CHCHD2* przemieszcza się z mitochondrialnej przestrzeni międzybłonowej do jądra, gdzie wiąże się z elementem reagującym na tlen (ang. *oxygen-responsive element*) zlokalizowanym w promotorze oksydazy cytochromowej 4I2 (*COX4I2*). *COX4I2* koduje podjednostkę kompleksu IV łańcucha transportu elektronów. Wiązanie *CHCHD2* w rejonie promotora *COX4I2* skutkuje zwiększeniem jego ekspresji. Spadek ilości *CHCHD2* powoduje spadek poziomu *COX4I2* i zmniejsza zużycie tlenu w komórce (Aras et al., 2015). Wykazano również, że *CHCHD2* jest negatywnym regulatorem apoptozy za pośrednictwem mitochondriów (ang. *mitochondria-mediated apoptosis*) (Liu et al., 2015). Liu i in. wykazali, że *CHCHD2* wchodzi w interakcję z antyapoptotycznym białkiem BCL-XL (z rodziny BCL-2), co prowadzi do hamowania proapoptotycznego BAX, a w konsekwencji

## DYSKUSJA

zmniejsza MOMP i inhibuje apoptozę. Ponadto wykazano, że *CHCHD2* reguluje ekspresję wielu genów, które odgrywają rolę w migracji komórek i przerzutach w raku, oraz że poziom *CHCHD2* jest wyższy w liniach komórkowych pochodzących z agresywnych guzów piersi (Aras, 2019). W niniejszej pracy doktorskiej wykazano, że wszystkie mutacje wykryte w gorących punktach w BCC oraz czerniaku pokrywają się z dwoma odrębnymi miejscami wiązania czynników transkrypcyjnych GABP-alfa i osłabiają ich wiązanie. Ponieważ wiadomo, że GABP-alfa jest czynnikiem transkrypcyjnym zaangażowanym w regulację metabolizmu energii komórkowej i regulację cyklu komórkowego (Rosmarin, Resendes, Yang, McMillan, & Fleming, 2004), odkrycie to może wskazywać na funkcjonalną rolę mutacji *CHCHD2* w BCC.

*DHODH* to gen, który wykazuje częste mutacje w sekwencji Kozak w 5'UTR, i charakteryzuje się występowaniem substytucji w gorącym punkcie obejmującym 2 pozycje genomowe c.-5G>A i c.-5\_-4delinsAA. Dodatkowo *DHODH* został sklasyfikowany jako potencjalny gen napędzający rozwój BCC. Chociaż analiza sekwencji Kozak z zastosowaniem ATGpr oraz NetStart 1.0 nie wykazała znaczącego wpływu mutacji w pozycjach -4 oraz -5 na zmianę efektywności translacji jednak wynik ten nie wyklucza wpływu mutacji na proces translacji w specyficznych warunkach takich jak hipoksja, narażenie na promieniowanie UV, czy nowotwór. Analiza porównawcza BCC z danymi z TCGA wykazała, że poza BCC żaden inny nowotwór nie miał mutacji ani w gorącym punkcie ani w ogóle w regionie 5'UTR, co wskazuje, że mutacje te są specyficzne wyłącznie dla BCC. Brak tych mutacji w innych nowotworach związanych z promieniowaniem UV sprawia, że prawdopodobieństwo, aby częste występowanie mutacji w BCC było spowodowane wyłącznie przypadkowym efektem wywołanym promieniowaniem UV jest niewielkie. Chociaż mutacje w 5'UTR genu *DHODH* nigdy wcześniej nie były obserwowane i badane w żadnym nowotworze, wykazano, że *DHODH* odgrywa kluczową rolę w kancerogenezie SCC i innych nowotworów skóry wywołanych promieniowaniem UV oraz ułatwia rozwój przedrakowych zmian skórnych (Hosseini et al., 2018, 2019). Hosseini i in. wykazali, że poziom białka *DHODH* i jego aktywność enzymatyczna są znacznie podwyższone w skórze poddanej działaniu promieniowania słonecznego i że zwiększony poziom *DHODH* uwrażliwia skórę na uszkodzenia wywołane promieniowaniem UV. Wykazano również, że poziom białka *DHODH* jest podwyższony w czerniaku, w którym hamowanie aktywności *DHODH* prowadzi do wyraźnego zmniejszenia wzrostu guza zarówno *in vitro*, jak i w badaniach ksenograftów myszy (White et al., 2011). Inaktywacja *DHODH* hamuje proliferację komórek i indukuje zatrzymanie cyklu komórkowego

## DYSKUSJA

w fazie S w komórkach czerniaka z niedoborem BCL-2 (proapoptotyczne białko) (Garcia-Bermudez et al., 2018). DHODH jest białkiem osadzonym w wewnętrznej błonie mitochondrialnej, a jego kanoniczną rolą jest utlenianie dihydroorotanu do orotanu, co stanowi ważny etap w syntezie pirymidyny de novo (kluczowy etap w replikacji i naprawie DNA). Produkt uboczny szlaku syntezy pirymidyn, ubichinol (QH<sub>2</sub>), jest źródłem elektronów w łańcuchu transportu elektronów. Dodatkowo DHODH odgrywa rolę w alternatywnym (niezależnym od glukozy) oddychaniu komórkowym (wykorzystując aminokwasy jako źródło energii) (Garcia-Bermudez et al., 2018; Hosseini et al., 2018, 2019), ułatwiając rozwój raka w warunkach niedotlenienia. Ponadto stwierdzono, że w raku kolczystokomórkowym przełyku podwyższone poziomy DHODH sprzyjają proliferacji komórek poprzez stabilizację  $\beta$ -kateniny (Qian et al., 2020).

Funkcjonalne efekty mutacji wykrytych w niniejszej pracy w *DHODH* mogą wynikać z samej zmiany sekwencji Kozak, ale także z utworzenia miejsca wiązania czynnika transkrypcyjnego NFAT1, które nie występuje w sekwencji typu dzikiego. NFAT1 należy do rodziny czynników transkrypcyjnych NFAT i występuje na podwyższonym poziomie w wielu komórkach nowotworowych (Mancini & Toker, 2009). Aktywacja i nadekspresja NFAT1 w wielu typach nowotworów sprzyja transkrypcji genów, które są kluczowe dla rozwoju i progresji nowotworu, w tym *COX2*, *MMP7*, *MMP9* i *MDM2* (Shoshan et al., 2016; Zhang et al., 2012). W związku z ważną funkcją białka DHODH w rozwoju zmian skórnych oraz wykazaniem w niniejszej pracy wpływem mutacji na funkcjonalność *DHODH*, gen ten może wykazywać potencjał napędzający dla rozwoju BCC.

Analizowane w niniejszej pracy dane z sekwencjonowania typu WES posłużyły także do przeprowadzenia pierwszej kompleksowej (całogenomowej) oceny zmian liczby kopii w BCC. Analiza z użyciem GISTIC2.0 umożliwiła wykrycie zmian o potencjalnie napędzającym wskazując 54 regiony ulegające ogniskowym delecjom, 56 regionów ulegających ogniskowym amplifikacjom oraz przede wszystkim dwie bardzo istotne zmiany na poziomie ramienia chromosomów. Oprócz częstej delecji chr9q obejmującej gen *PTCH1*, wykryto również częstą duplikację chr9p, obejmującą geny *PDL1* i *PDL2*, których nadekspresja umożliwia komórkom nowotworowym unikanie odpowiedzi układu odpornościowego gospodarza. W związku z potrzebą potwierdzenia zmian liczby kopii wykrytych przez analizy całogenomowe wykorzystujące wyniki sekwencjonowania NGS zaprojektowano test MLPA dedykowany analizom liczby kopii regionów chromosomu 9 obejmujących geny *PTCH1*, *PDL1* oraz *PDL2*.

## DYSKUSJA

Test pozwolił na potwierdzenie wszystkich zmian liczby kopii w genach *PTCH1*, *PDL1* oraz *PDL2* wykrytych z zastosowaniem GISTIC 2.0.

Wzrost liczby kopii regionu 9p24.3-9p22.2 obejmującego *PDL1* oraz *PDL2* zaobserwowano wcześniej tylko w jednym przypadku przerzutującego BCC (Ikeda et al., 2016). Pacjent, który był oporny na wismodegib i sonidegib,- inhibitory szlaku Shh, wykazał odpowiedź na niwolumab (przeciwciało anti-PD-1 (receptor, do którego wiąże się PD-L1) blokujące interakcję PD-1/PD-L1), co sugeruje, że wzrost liczby kopii *PDL1* może być biomarkerem wrażliwości na leczenie z zastosowaniem inhibitorów punktów kontrolnych (Ikeda et al., 2016). W niezależnym badaniu wykazano również, że niektórzy pacjenci (do ~40%) z zaawansowanym BCC (nie testowanym pod kątem amplifikacji *PDL1*) odpowiadają na terapię pembrolizumabem (kolejne przeciwciało anti-PD-1) (Chang et al., 2019; Ikeda et al., 2016). Identyfikacja zmian liczby kopii genów kodujących białka stanowiące punkty kontrolne układu immunologicznego może pomóc w ocenie zasadności leczenia z udziałem inhibitorów punktów kontrolnych, jednak wymagane są dalsze badania na większej grupie próbek.

Analiza mutacji w regionach niekodujących poza częściami niekodującymi genów kodujących białka obejmowała także geny miRNA i była możliwa dzięki zaprojektowaniu panelu miRNom dedykowanemu analizom genów miRNA. Na konieczność przeprowadzania analizy mutacji w genach miRNA w nowotworach może wskazywać przykład genu *MIR142*, którego mutacje mają napędzający wpływ na rozwój nowotworów (głównie nowotworów krwi) (Morin et al., 2010; Kwanhian et al., 2012; Puente et al., 2015; Thol et al., 2015). Przeprowadzona w niniejszej pracy analiza mutacji w genach miRNA jest pierwszą taką analizą w BCC. WMS pozwoliło na wykrycie 171 mutacji zlokalizowanych w 155 genach miRNA w 22 próbkach (81,5%) BCC. 33% mutowanych genów miRNA to miRNA o wysokiej wiarygodności (wg miRBase i/lub miRGeneDB). Mutacje zidentyfikowano we wszystkich subregionach genów miRNA, a częstość mutacji jest podobna we wszystkich subregionach. Częstość rozkładu mutacji w poszczególnych subregionach jest zbliżona do częstości wykrytej w innych pracach analizujących te regiony w innych nowotworach na podstawie danych TCGA (Galka-Marciniak et al., 2019; Urbanek-Trzeciak et al., 2020). 7 mutacji wykrytych w regionach źródłowych miRNA skutkuje znaczącą zmianą puli potencjalnie regulowanych genów docelowych prowadząc zarówno do spadku jak i wzrostu ilości regulowanych genów.

Analiza często mutowanych genów miRNA pozwoliła na zidentyfikowanie tylko jednego genu miRNA często mutowanego w BCC. Jest to *MIR3928* posiadający 4 substytucje

## DYSKUSJA

w 4 próbkach. Przegląd literaturowy wykazał, że miR-3928-3p pełni funkcję zarówno supresorową jak i onkogenną. Wykazano, że zwiększony poziom miR-3928-3p jest związany z gorszą przeżywalnością pacjentów z płaskonabłonkowym rakiem płuca (Xia et al., 2018). Z kolei wysoki poziom miR-3928-3p w glejaku spowalnia rozwój nowotworu przez obniżenie ekspresję wielu onkogenów m.in. *MDM2* (Mulcahy et al., 2022). Obniżony poziom miR-3928-3p może być stosowany jako potencjalny biomarker w diagnostyce nowotworów głowy i szyi (Fadhil et al., 2020). Niski poziom miR-3928-3p może także przyspieszać rozwój kostniakomięsaka (Xu, Liu, & Zhao, 2014).

W związku z małą liczebnością prób i stosunkowo niską liczbą wykrytych mutacji ustalenie, czy zidentyfikowane często mutowane, ale także pozostałe geny miRNA charakteryzują się wzorem mutacyjnym charakterystycznym dla supresorów lub onkogenów wymagałoby dalszej walidacji statystycznej na większej liczbie prób.

Aby lepiej zrozumieć potencjalną funkcję zidentyfikowanych częstych mutacji, scharakteryzowano je pod kątem lokalizacji w różnych subregionach prekursorów miRNA oraz wpływu na strukturę 2D i 3D prekursorów miRNA. Wszystkie mutacje w miR-3928 były zlokalizowane w regionie flankującym 5'. W prekursorze miR-3928 zlokalizowany jest jedyny wykryty w BCC w genach miRNA gorący punkt mutacji, który wpływa na zmniejszenie stabilności struktury prekursora (analiza z zastosowaniem MFold). W związku z jedynie niewielkim wpływem mutacji na strukturę prekursorów trudno jest przewidzieć jaki funkcjonalny wpływ na geny miRNA mogą mieć wykryte mutacje.

Analiza z zastosowaniem OncodriveFML nie wykryła żadnego miRNA o potencjale napędzającym rozwój BCC. Niska liczebność próbek i liczba wykrytych mutacji utrudniają określenie czy mutacje mają potencjał napędzający. Warto jednak zaznaczyć, że nawet jeśli wykryte zmiany somatyczne nie są głównymi czynnikami napędzającymi, mogą nadal odgrywać rolę w raku. Wykazano, że mutacje pasażerskie mogą zarówno zapobiegać dalszej progresji nowotworu (McFarland et al., 2017), jak i brać udział w stymulowaniu rozwoju nowotworu i lekooporności (Castro-Giner, Ratcliffe, & Tomlinson, 2015; Nussinov & Tsai, 2015). Zaprezentowane w pracy nowatorskie podejście do celowanego sekwencjonowania genów miRNA może doprowadzić do poznania roli tych niekodujących cząsteczek w wielu jednostkach chorobowych.

Jedynie niekodujące zmiany badane dotychczas w BCC to mutacje w promotorach *TERT* oraz *DPH3*. Przeprowadzona w niniejszej pracy analiza tych regionów wykazała, że wszystkie

## DYSKUSJA

mutacje wykryte zarówno w *TERT* jak i *DPH3* lokalizują się w pozycjach genomowych już wcześniej opisanych w literaturze (Denisova et al., 2015; Griewank et al., 2013; Maturo et al., 2020). W związku z faktem, że najczęściej występująca mutacja w promotorze *TERT* (w pozycji c.-146) jest uznana za zmianę napędzającą rozwój m.in. czerniaka, mutacje w *TERT* mogą mieć charakter napędzający także w przypadku BCC. Z kolei najczęściej występująca mutacja w *DPH3* (w pozycji c.-122) skutkuje najprawdopodobniej zmniejszoną ekspresją *DPH3* sugerując utratę funkcji supresorowej tego genu, jednak w związku z brakiem jasnej funkcjonalności *DPH3* w procesie nowotworzenia, efekt mutacji zlokalizowanych w promotorze tego genu nie został jak dotąd określony (Denisova i in., 2015).

Poza analizą regionów niekodujących wyniki otrzymane w pracy doktorskiej dostarczają także cennych informacji związanych z ogólną charakterystyką mutacji, takich jak np. określenie TMB, czy sygnatur mutacyjnych. Analiza TMB wykazała, że dla BCC wartość tego parametru jest najwyższa i przewyższa obciążenie mutacyjne w innych typach nowotworów (analiza porównawcza z danymi z pracy Lawrence i in., 2013). Wartość parametru TMB dla analizowanego w niniejszej pracy zbioru próbek jest, podobnie jak w poprzednich pracach o BCC (Bonilla et al., 2016; Jayaraman et al., 2014) bardzo wysoka. W przypadku analizy sygnatur wykazano, że sygnaturą najsilniej związaną z analizowanymi próbkami BCC jest sygnatura 7 co potwierdza, że głównym czynnikiem mutagennym w BCC jest promieniowanie UV. Analiza różnic molekularnych pomiędzy podtypem guzkowym i powierzchniowym BCC nie wskazuje na istotne różnice, z wyjątkiem nieco wyższego poziomu przypisania sygnatury mutacyjnej do sygnatury 7 w próbkach podtypu guzkowego.

W przypadku analiz części kodującej wykazano, że najczęściej mutacje lokalizują się w genach zaangażowanych w szlak Shh, głównie *PTCH1*, który jest genem napędzającym rozwój BCC. Kolejny najczęściej mutowany w BCC gen to *TP53*, a także *MYCN* oraz geny rodziny *NOTCH*. Wyniki te są zgodne z rezultatami przedstawionymi we wcześniejszych pracach o BCC (Bonilla et al., 2016; Jayaraman et al., 2014), co potwierdza wiarygodność procedur wykrywania mutacji przeprowadzonych w niniejszej pracy doktorskiej.

Wykryto także częste mutacje w sekwencji kodującej genów *FLG* i *FLG2* kodujących profilagrynę i białko podobne do filagryny, będące prekursorami dojrzałej filagryny. Filagryna jest ważnym składnikiem warstwy rogowej naskórka, odgrywa rolę w utrzymaniu homeostazy nabłonka (Wu, Hansmann, Meyer-Hoffert, Gläser, & Schröder, 2009) oraz jest substratem w reakcji prowadzącej do powstania kwasu transurokanowego (UCA) i kwasu

## DYSKUSJA

pirolidonokarboksylowego (PCA), które wchodzi w skład naturalnego czynnika nawilżającego (ang. *Natural Moisturizing Factor*) i wykazują właściwości hydrofilowe zapobiegając wysychaniu skóry. Dodatkowo UCA działa jak naturalna bariera chroniąca skórę przed promieniowaniem UVB (Leman, Moosbrugger-Martinez, Blunder, Pavel, & Dubrac, 2019). Chociaż mutacje w *FLG* i *FLG2* były wcześniej sporadycznie obserwowane w BCC, były uważane za przypadkowe (pasażerskie) ze względu na wysoki polimorfizm i powtórzenia sekwencji w obrębie tych genów. Przeprowadzona przez J. Kaae analiza porównawcza częstości występowania mutacji germinalnych w *FLG* u pacjentów z BCC i u kontrolnej grupy nie wykazała różnicy częstości mutacji wśród analizowanych grup (Kaae et al., 2013). Mutacje germinalne w *FLG/FLG2* skutkujące utratą funkcjonalnego produktu genów są częste w populacji kaukaskiej i wiąże się je głównie z występowaniem chorób skóry takich jak atopowe zapalenie skóry (ang. *atopic dermatitis*), rogowacenie starcze, rybia łuska (ang. *ichthyosis vulgaris*) oraz z alergiami pokarmowymi a także astmą (Kaae et al., 2013; Kalb et al., 2022; McAleer & Irvine, 2013). W niniejszej pracy spośród 23 mutacji somatycznych wykrytych w *FLG* 15 stanowią mutacje niesynonimiczne (3 mutacje definitywne (13%)) a 8 synonimiczne (34,8%). W przypadku *FLG2* wykryto 10 mutacji z czego 9 stanowią mutacje niesynonimiczne (3 mutacje definitywne (30%)) a 1 to mutacja synonimiczna (10%). Duży odsetek mutacji niesynonimicznych oraz występowanie mutacji definitywnych mogą jednak sugerować funkcjonalny charakter mutacji w *FLG* oraz *FLG2* w BCC.

Należy jednak zwrócić uwagę na ograniczenia analizy przeprowadzonych w niniejszej pracy. Przede wszystkim, jako że niniejsza praca była pierwszą i wstępną analizą sekwencji niekodujących w BCC przeprowadzona ona została na relatywnie niewielkiej liczbie próbek. Skutkuje to ograniczoną siłą statystyczną eksperymentu i w konsekwencji brakiem istotności statystycznej wielu analiz. Ponadto w pracy ograniczono się do charakterystyki zidentyfikowanych wariantów wyłącznie pod kątem genetycznych analiz opisowych. Mutacje o potencjale napędzającym wymagają zatem eksperymentalnego potwierdzenia ich wpływu na rozwój BCC. Należy również zauważyć, że analiza objęła tylko część (~1%) genomu niekodującego, który łącznie stanowi ~98% genomu i zawiera wiele różnych elementów funkcjonalnych nieobjętych niniejszą analizą, w tym promotory, wzmacniacze, oraz geny innych klas niekodujących RNA. Warto podkreślić, że w związku z tym, że mutacje w regionach niekodujących są rzadko badane, dostępność narzędzi bioinformatycznych pomagających w analizie tych regionów jest bardzo ograniczona, w szczególności narzędzi ułatwiających

## DYSKUSJA

analizę dużych zbiorów danych. Mniejsza dostępność narzędzi dedykowanych analizom regionów niekodujących wynika także z różnorodności typów regionów niekodujących oraz z niewystarczającego rozpoznania elementów funkcjonalnych w tych regionach co uniemożliwia łatwe przewidzenie konsekwencji występujących mutacji. Rozszerzenie analiz o dodatkowe regiony niekodujące mogłoby zatem spowodować trudności/niemożność wielkoskalowej oceny wpływu mutacji na funkcję mutowanych regionów. W niniejszej pracy OncodriveFML pozwolił na identyfikację genów o potencjale napędzających na podstawie mutacji zlokalizowanych zarówno w regionach niekodujących jak i kodujących genów, zatem nie jest algorytmem dedykowanym analizom specyficznych regionów niekodujących. Algorytm korzysta z narzędzia CADD, które generuje wartość określającą szkodliwość wykrytych mutacji na podstawie ponad 60 cech genomowych, jednak znaczna ich część odnosi się głównie dla regionów kodujących a tylko niektóre dla regionów niekodujących.

Podsumowując w niniejszej pracy wykorzystując dane z sekwencjonowania WES oraz WMS wykryto nie tylko mutacje w regionach kodujących genów związanych z rozwojem BCC, ale przede wszystkim zidentyfikowano częste mutacje w niekodujących białka częściach genów związanych z rakiem, które mogą wykazywać potencjał funkcjonalny i odgrywać rolę w rozwoju BCC. Analiza przeprowadzona w niniejszej pracy jest jak dotąd jedyną analizą obejmującą zarówno kodującą jak i niekodującą część genomu BCC. Przeprowadzona pierwszy raz w BCC kompleksowa analiza zmian liczby kopii, a także analiza często mutowanych genów, gorących punktów mutacji oraz genów napędzających przyczyniła się do wszechstronnej analizy mutacji w BCC. Aby jednoznacznie określić rolę poszczególnych zidentyfikowanych genów/mutacji w BCC konieczne są dalsze badania funkcjonalne. Jednak uzyskane wyniki znacząco przyczyniają się do poszerzenia wiedzy dotyczącej molekularnych podstaw tej choroby, co może być pomocne w diagnostyce BCC lub nawet w dalszej perspektywie może znaleźć zastosowanie w terapii celowanej. Przedstawione w pracy wyniki stanowią mocną podstawę do dalszych, badań niekodujących wariantów nie tylko w BCC, ale także w innych nowotworach jak i chorobach nienowotworowych.



## VIII. GŁÓWNE WYNIKI I WNIOSKI

Przeprowadzone badania pozwoliły na szczegółową charakterystykę molekularnych podstaw BCC oraz pozwoliły na wyciągnięcie następujących wniosków:

1. Wykorzystanie standardowego systemu WES, poza identyfikacją mutacji w częściach kodujących genów, pozwala na wykrycie znacznej ilości nowotworowych mutacji somatycznych w niekodujących częściach genów. Globalna częstość mutacji w regionach niekodujących zbliżona jest do częstości mutacji w regionach kodujących.
2. BCC charakteryzuje się najwyższym obciążeniem mutacyjnym wśród nowotworów, które średnio wynosi  $\sim 34,1$  mutacje/Mpz (na podstawie analizy TMB z wyłączeniem mutacji synonimicznych) i 52,8 mutacji/Mpz (na podstawie analizy wszystkich mutacji w części kodującej).
3. Analiza typów mutacji wskazuje, że najczęstszą sygnaturą mutacyjną (najsilniej związaną z analizowanymi próbkami BCC) jest sygnatura 7. Wynik ten potwierdza, że głównym czynnikiem mutagennym w BCC jest promieniowanie UV. Może być to dodatkowy argument stosowania ochrony przed światłem słonecznym jako zachowania profilaktycznego chroniącego przed BCC.
4. Analiza mutacji w regionach niekodujących wykazała istnienie regionów o podwyższonej częstości mutacji, regionów zawierające gorące miejsca mutacji oraz regionów, w których mutacje wskazują ich duży potencjał jako elementów napędzających dla rozwoju BCC. Do najciekawszych takich regionów należą 3'UTR genu *BAD*, 5'UTR (sekwencja Kozak) genu *DHODH* oraz 5'UTR genu *CHCHD2*.
5. Analiza mutacji w sekwencjach kodujących potwierdziła dużą zgodność z wcześniejszymi wynikami sekwencjonowania tych regionów w BCC, zidentyfikowano między innymi częste mutacje i/lub napędzający charakter takich genów jak *PTCH1*, *TP53*, *MYCN* czy geny z rodziny *NOTCH*. Ta duża zgodność z wcześniejszymi wynikami potwierdza poprawność przeprowadzonej analizy a tym samym wiarygodność uzyskanych wyników.
6. Przeprowadzenie sekwencjonowania genów miRNA pozwoliło na wykrycie znacznej liczby mutacji somatycznych w genach miRNA w BCC. Wcześniej tego typu

## GLÓWNE WYNIKI I WNIOSKI

sekwencjonowanie nie było prowadzone w żadnym nowotworze. Średnia liczba mutacji w genach miRNA wynosiła 6,3 mutacje na próbkę BCC, a wykryte mutacje występowały we wszystkich subregionach genów miRNA, w tym w elementach funkcjonalnych takich jak region źródłowy, co wskazuje, że mutacje somatyczne w genach miRNA mogą zakłócać/uszkadzać funkcjonowanie tych genów. Należy jednak zaznaczyć, że ze względu na małą liczbę analizowanych próbek i w konsekwencji małą liczbę wykrytych mutacji nie możliwe jest na tym etapie wyciąganie wniosków o konsekwencjach czy to molekularnych czy klinicznych tych mutacji. Sugeruje to konieczność kontynuowania podobnych badań w większych grupach próbek.

7. Z wyjątkiem nieco wyższego poziomu przypisania sygnatury mutacyjnej do sygnatury 7 w próbkach typu guzkowego, przeprowadzona analiza nie wskazuje na żadne istotne molekularne różnice między podtypem guzkowym i powierzchniowym BCC.
8. Przeprowadzona pierwszy raz w BCC całogenomowa, kompleksowa analiza zmian liczby kopii wskazała na istnienie licznych regionów o podwyższonej częstości zmian liczby kopii, zarówno lokalnych jak i zmian obejmujących całe ramiona chromosomów. Do najważniejszych wykrytych zmian należą częste delecje długiego ramienia chromosomu 9 i częste duplikacje krótkiego ramienia chromosomu 9 zawierające odpowiednio gen *PTCH1* oraz geny *JAK2*, *PDL1* i *PDL2*, które mogą mieć duże znaczenie dla rozwoju BCC (analiza z użyciem GISTIC2.0).
9. Analizy przeprowadzone w zakresie mutacji niekodujących miały charakter wstępny (pionierski) i zostały wykonane na relatywnie niewielkiej grupie próbek, dlatego uzyskane wyniki, na tym etapie, nie pozwalają na wyciągnięcie głębszych konkluzji, czy to o znaczeniu molekularnym czy klinicznym wykrytych mutacji. Uzyskane wyniki wymagają weryfikacji w większych grupach próbek i w odpowiednich testach funkcjonalnych. Przeprowadzona analiza, jest jednak pierwszym krokiem w analizie sekwencji niekodujących w BCC, silnie wskazującym potencjał funkcjonalny zmienności somatycznej w tych regionach i konieczność dalszej bardziej dogłębnej analizy mutacji niekodujących w BCC i w innych nowotworach.

## IX. MATERIAŁY I METODY

### 1.1 Materiały

#### 1.1.1 Enzymy

- Polimeraza DNA GoTaq 5 U/ $\mu$ l (Promega)
- Odwrotna transkryptaza Superscript IV 200 U/ $\mu$ l (Invitrogen)
- Ligaza-65 (MRC Holland)
- Polimeraza SALSA (MRC Holland)
- DNaza I 4 U/ $\mu$ l (Qiagen)
- Inhibitor RNaz RNaseOUT 40 U/ $\mu$ l (Invitrogen)
- Proteinaza K 20 mg/ml (Qiagen)

#### 1.1.2 Zestawy komercyjne

- Zestaw do izolacji AllPrep DNA/RNA/miRNA Universal Kit (Qiagen)
- Zestaw do odwrotnej transkrypcji - SuperScript IV First-Strand Synthesis System (Invitrogen)
- Zestaw SALSA MLPA Reagent Kit (MRC Holland)
- Zestaw do pomiaru stężenia DNA - dsDNA Qubit HS Assay Kit (ThermoFisher)
- Zestaw do wzbogacania SureSelect Custom - samodzielnie zaprojektowany zestaw sond dla miRNomu (Agilent)

#### 1.1.3 Bufory i roztwory

- Octan sodu 3M, pH 5,2
- TBE 10x stężony, pH 8,3
  - Tris-HCl 890 mM
  - Kwas borowy 890 mM
  - EDTA 20 mM

## MATERIAŁY I METODY

- **TE 1x, pH 8,0**
  - Tris-HCl 10 mM
  - EDTA 1 mM
- **Roztwór obciążający do nakładania próbek na żele agarozowe, 6x stężony**
  - Glicerol 30%
  - EDTA 10 mM
  - Błękit bromofenolowy (BB) 0,04%
  - Cyjanol ksylenowy (FF) 0,04%
- **Bufory i roztwory dostępne komercyjnie**
  - Bufor dla polimerazy GoTaq - Green GoTaq Buffer 5x (Promega)
  - Bufor dla odwrotnej transkryptazy z zestawu SuperScript IV First-Strand Synthesis System (ThermoFisher)
  - Bufor dla ligazy do reakcji MLPA (MRC Holland)
  - Bufor EB (Qiagen)
  - Bufor RPE (Qiagen)
  - Bufor FRN (Qiagen)
  - Bufor AW1 (Qiagen)
  - Bufor AW2 (Qiagen)
  - Bufor RLT Plus (Qiagen)
  - Bufor A do MLPA (MHC Holland)
  - Bufor B do MLPA (MHC Holland)

### 8.1.4 Inne odczynniki chemiczne

- Agaroza BioStandard (Prona)
- Etanol 96% CZDA (POCH)
- Izopropanol 99,7% CZDA (POCH)
- SimplySafe (EURx)
- Błękit bromofenolowy (BB; Sigma Aldrich)
- Cyjanol ksylenowy (FF; Sigma Aldrich)
- Błękit trypanu 0,4% (Gibco)

## MATERIAŁY I METODY

- Glicerol 99% (Sigma Aldrich)
- EDTA (BioShop)
- Tris-HCl (BioShop)
- Kwas borowy (BioShop)
- dNTP 10 mM (Invitrogen)
- DTT 0,1 M (Invitrogen)
- RNAlater (Qiagen)
- Reagent DX (Qiagen)

### 1.1.5 Wzorce wielkości DNA

- Quick-Load 1 Kb DNA Ladder (NEB)
- 1 Kb Plus DNA Ladder (Invitrogen)

### 1.1.6 Żele agarozowe

- **Żel agarozowy do analizy produktów PCR**  
1,3% agarosa w buforze 0,5x TBE z SimplySafe (4  $\mu$ l na 100 ml TBE)
- **Żel agarozowy do oceny jakości gDNA i RNA**  
1% agarosa w buforze 0,5x TBE z SimplySafe (4  $\mu$ l na 100 ml TBE)

### 1.1.7 Materiał biologiczny

- **Tkanki**

54 tkanki (27 par) umieszczonych w roztworze RNAlater: tkanka skórna obejmująca zmianę nowotworową (rak podstawnokomórkowy skóry, podtyp guzkowy lub powierzchniowy) oraz zdrowa skóra otaczająca nowotwór. Tkanekę skórną pozyskano w szpitalu św. Józefa (Essen, Niemcy) w Zakładzie Chirurgii Plastycznej, kierowanym przez Profesora Michaela Sanda. Zabieg wycięcia tkanki przeprowadzono w znieczuleniu miejscowym. Pobrano 4 mm biopsje ze środka zmiany nowotworowej oraz z nienaruszonego, zdrowego nabłonka. Tkanekę umieszczono w roztworze RNAlater (Qiagen) i przechowywano w temperaturze  $-80^{\circ}\text{C}$ . Następnie materiał umieszczono na suchym lodzie i wysłano do Zakładu Genetyki Molekularnej, Instytutu Chemii Bioorganicznej, Polskiej Akademii Nauk w Poznaniu.

### 1.1.8 Oligonukleotydy

- Startery syntetyzowane były przez firmę Sigma-Aldrich

**Tabela 10. Zestawienie starterów użytych do reakcji PCR.**

Nazwa starteru	Sekwencja 5` - 3`
SCNM1_F	TCTGAAATCCTTTGTATGCAGAAC
SCNM1_R	CTTCTTGTTGGGCTCATGG
FLG2_ex3_F1	TCCAGATGTCTGTCCTGAAT
FLG2_ex3_R1	GGTCAGGCTCAGGTCAATC
FLG2_ex3_F2	GGCTAGATCTCTGTCTTCCAGT
FLG2_ex3_R2	ACTCACGGCCAAAGTGGAT
FLG2_ex3_F3	CATGCCTTGCCTTTCCTACTAC
FLG2_ex3_R3	CTGGAAGACAAAGATCTAGTCACA
RPS27_F	TGATCAAACGCTGACCTCAA
RPS27_R	GAAAATAGCCCCACTGTCA
DYRK3_F	CAAGTTTGCCAGTCCATCT
DYRK3_R	GAAGTTTTGGTGGTGGCAGT
MYCN_F1	GTGTCTGTCGGTTGCAGTGT
MYCN_R1	CACAGCTCGTTCTCAAGCAG
MYCN_F2	GCTGTCACCACATTACCAT
MYCN_R2	GTGAGAAAGCTGGACCGAAG
MYCN_F3	CTTCGGTCCAGCTTCTCAC
MYCN_R3	AACCGTCACCAACGTTTAGC
MYCN_F4	CCCTACGTGGAGAGTGAGGA
MYCN_R4	AAATGTGCAAAGTGGCAGTG
PTCH1_ex17_F	ACACATCCTCGTCTCCAGA
PTCH1_ex17_R	TGTTCTGATGGGATTTTCGAC
PTCH11_ex14_F	TGGCTTTTGGTTTCAAGAGG
PTCH11_ex14_R	CAGCAATGTGATGTTTCTCTCC
PTCH1_ex12_F	TAAAGCAGCATCCACCCAGT
PTCH1_ex12_R	TGCTGGTGTAGGGGGTAAG
PTCH1_ex6_F	TGTTTTGCTCTCCACCCTTC
PTCH1_ex6_R	GCATTTGTTTTGATGCCAGA
AMBRA1_F1	AGTGGACTIONACTCGGTCTT
AMBRA1_R1	TGTAGTTGCAGGAGCACAGG
AMBRA1_F2	CTAGGGGAGCTGCCATACAG
AMBRA1_R2	TGAGCAAGTGAATATGAAGATTAGAGA
BAD_3UTR_F	CAGACGCGGGCTTTATTAAC
BAD_3UTR_R	AGTGACCTTCGCTCCACATC
DHODH_kozak_F	GTGCACCAACCAGAACCT
DHODH_kozak_R	GGCGGTCATGTGTTCACTAA
TP53_ex4_F	GCCAGGCATTGAAGTCTCAT

## MATERIAŁY I METODY

TP53_ex4_R	CGTTCTGGTAAGGACAAGGGTT
MISP_F	AAGACCTACCGCTGGATG
MISP_R	AGCACGCACCTTGTTCTCTT
CASZ1_F	CGAGGGACTCACTTCTCTG
CASZ1_R	TCCACACCTGTCTTCACTGC
TCHH_F	GGTGCCTGTTTTTGTGGAT
TCHH_R	CCATGCCTGTGCCAGTATAA
BIRC6_F	CACAACCTCTGTGCCTTCCAA
BIRC6_R	TTGGTATTAATAATTTTGCCTTCT
TTN_F	TTCACTGAAAAGACAGTTTCTCG
TTN_R	TCCTAACCCATTGTAATAATAGAA
SCAP_F	CCCCAAGTCCAGGTTCAAGT
SCAP_R	GTAGAGGGGATGGGGATCAT
HERC5_F	CTGGAAAACATGGGCAACTT
HERC5_R	GGAGACAGCAGAAAAGTGAATTG
KCNN2_F	CTTGTTTTTCAAGGTGCTGGT
KCNN2_R	GACTCGTCTTTATCCAATACCAA
IBTK_F	TCTTTCAGCTTATCAGCATTTTTC
IBTK_R	TTTTTCGTGTCAATGTTTTTAGAC
ADGB_F	TCTTAGCCCGGGATATATGG
ADGB_R	GGCATTATATGCAATGTCCA
SSPO_F	GTCCTGGTTAACGTCCCTCA
SSPO_R	ACATATATGAGCGGGCCATT
PTCH1_del_F1	TGGCTTTTGGTTTCAAGAGG
PTCH1_del_R1	GCGTCAGCAGAGTGATTGAG
ALG10B_F	AGGTAAGCAGAAATAGCATTATCA
ALG10B_R	GGATGTAGGGCCAAGTCAAA
FRMPD3_F	TTCCGAGACAACAGCTCTGA
FRMPD3_R	CCAAAGAGATGCAATGACGA
TERTpromoter_F	CCCACGTGCGCAGCAGGAC
TERTpromoter_R	CTCCCAGTGGATTGCGGGGC
DPH3promoter_F2	CCTCGTCATGAAACTGCCC
DPH3promoter_R2	GCAGACGATAAGAAGCAGGG
miR604_F	CAGGGAGCACAGGTGTTTG
miR604_R	AACAACCTTGACACGGTGCAG
miR409_F	ACGTCCTTGGAGAGTGTGAG
miR409_R	GCTAGTGGACCAGGTGAAGT
miR448_F	CCAGAAGAGGCTTTCATGACAG
miR448_R	ACTCACATACCCACTGCCTC
miR-3928_F	GTCTCGAACTCCTGACCTCA
miR-3928_R	CCAGCCCAGTCAATCAGTTT
miR-3612_F	AAGGGGAATCAGGTGGTTGT
miR-3612_R	GAGAGGGGAGAGGAAGTCAGC
miR-6507_F	TCTTGTAGGCTCCAGGTGTG
miR-6507_R	CTAGTGCTAGGCTGGGCTTC
miR-6796_F	GGGGATCACAGGGCAACTAA
miR-6796_R	CAGAGATCAGGGGCCAGG

## MATERIAŁY I METODY

miR-7108_F	GTCATCACCTGCTTCTGGA
miR-7108_R	TCACGAAGCCTTGGTCCG
DHODH_CDS_F	TCTCTTCGCCTCCTACCTGA
DHODH_CDS_R	TGGGTCTAGGGTTTCCTCC
GAPDH_F2	GAAGGTGAAGGTCGGAGTC
GAPDH_R2	GAAGATGGTGATGGGATTTC



## MATERIAŁY I METODY

- Sondy do MLPA syntetyzowane były przez firmę IDT

**Tabela 11. Sondy do reakcji MLPA.**

Nazwa sondy	koordynaty miejsca wiązania sondy	5'PSS	Długość	5'SS	długość	5'TSS	Długość	Tm	3'TSS	Długość	Tm	3'SS	długość	3'PSS	długość	5'HPL	3'HPL	TPL
Ctrl_1	chr22:30,069,296-30,069,338	GGGTTCCCTA AGGGTTGGA	19	cgctac	6	GGCCCAGATCACC GAGGAGGA	21	75.6	GGCAAACTTCTG GCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGC TGGCGC	23	46	47	93
JAK2_ex1	chr9:4,985,343-4,985,382	GGGTTCCCTA AGGGTTGGA	19	cgctactac	9	CTTGGGTGTTGCG GTCGCCA	20	75.4	CTTCGGCTTCTCG GCCGTC	20	75.5	tctac	5	TCTAGATTGGATCTTGC TGGCGC	23	48	48	96
PDL1_ex6	chr9:5,467,928-5,467,969	GGGTTCCCTA AGGGTTGGA	19	cgctactac	9	GCTGAGCGTGACA AGAGGAAGG	22	71.4	AATGGGCCGTG GGATGCAG	20	74.0	tactac	6	TCTAGATTGGATCTTGC TGGCGC	23	50	49	99
chr9p24.2	chr9:2,811,377-2,811,420	GGGTTCCCTA AGGGTTGGA	19	cgctactac t	10	CCAGATGCAGTC CTGTTGCTG	22	72.5	CCAAGTGGCGAT GGCATTAT	22	72.9	tactac	6	TCTAGATTGGATCTTGC TGGCGC	23	51	51	102
chr9q21.12	chr9:70,419,662-70,419,715	GGGTTCCCTA AGGGTTGGA	19	cgcta	5	GCAGAGTAGGGT GTCGAGATGACAT CAG	28	74.0	CTTCTAAACATAG CAAGCACGCACC	26	71.7	ctac	4	TCTAGATTGGATCTTGC TGGCGC	23	52	53	105
ctrl_2	chr1:156,105,818-156,105,862	GGGTTCCCTA AGGGTTGGA	19	cgctactac tat	12	CAGCTGGACGAGT ACCAGGAGCTT	24	72.8	CTGGACATCAAGC TGGCCCTG	21	72.7	aactaatctac	12	TCTAGATTGGATCTTGC TGGCGC	23	55	56	111
PTCH1_3'	chr9:95,438,463-95,438,519	GGGTTCCCTA AGGGTTGGA	19	cgctactac ta	11	TATGCATGTGGAT GTGTTGGGCAAC C	27	74.1	GTGTGCATGCATG AAGTATGTTTTAC AGCC	30	72.5	ctac	4	TCTAGATTGGATCTTGC TGGCGC	23	57	57	114
PDL2_ex2	chr9:5,522,543-5,522,598	GGGTTCCCTA AGGGTTGGA	19	cgctactac t	10	GAACATGATCTTC CTCCTGCTAATGTT GAG	30	71.5	CCTGGAATTGCAG CTTCACCAGATAG	26	72.0	taaatctac	9	TCTAGATTGGATCTTGC TGGCGC	23	59	58	117
PTCH1_5'	chr9:95,521,997-95,522,048	GGGTTCCCTA AGGGTTGGA	19	cgctactac tatt	13	TGTGGGAGAAGC TGAGCCTCCAACA GTC	28	77.1	CTTTCATTAGGCT GTGCCCTGGGG	24	75.0	aaactaatctac	13	TCTAGATTGGATCTTGC TGGCGC	23	60	60	120
ctrl_5	chr2:109,545,794-109,545,837	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aat	21	AGTCTGTGGCTA CGGCACCAA	22	72.8	AGACGAGGACTAC GGCTGCGTC	22	71.8	ggtcaactaat ctac	17	TCTAGATTGGATCTTGC TGGCGC	23	62	62	124
JAK2_ex20	chr9:5,089,805-5,089,863	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag	18	CCTGAAATCCCTA CAGCATGACAACA TTG	29	72.9	TAAAGTACAAGGG AGTGTGCTACAGT GCTG	30	71.1	aaactaatctac	13	TCTAGATTGGATCTTGC TGGCGC	23	66	66	132
PDL2_ex7	chr9:5,570,132-5,570,183	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattg	23	GAAACTTGCCCT TCACTGATCTGGA	26	73.4	CTCACCTCTGGAG CCTATGGCTTTAA	26	71.2	atggtcaactaa atctac	19	TCTAGATTGGATCTTGC TGGCGC	23	68	68	136

## MATERIAŁY I METODY

chr9q33.1	chr9:132,886,392-132,886,438	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatg	23	GCTGGTTGAACC CAACAGAGCTGCA TG	28	77.4	GTCTCCAGTAAT CACATGGAATGT AGACAC	32	72.4	tcaactaaatct ac	15	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	70	70	140
ctrl_3	chr17:3,397,657-3,397,712	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatg	26	TCCCTGCGCCATT GAGGCTATAAAA T	27	70.6	TATAGAGAAAGTT ATTACCCCGGG ATG	29	70.9	aatggtcaacta aatctac	20	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	72	72	144
chr9p24.1	chr9:8,517,968-8,518,021	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc c	28	GGGTGTCAGTAC ATTGTGTTTCATCC	27	71.9	AGTTGTTGACATG TTGAGTGGGATCC A	27	71.1	atctaagtgtcaa actaaatctac	24	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	74	74	148
chr9p22.1	chr9:19,515,344-19,515,395	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc caccttt	34	TTGGTCAGGGTAC TCTGGTGGAGGT	25	72.8	GGGCAGATAGTCT CTCTGTTTGGCTT G	27	72.9	atgtatctaattg tcaactaaatct ac	28	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	78	78	156
PDL1_ex1	chr9:5,450,545-5,450,586	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc caccttttc agc	39	TCTGCGCAGCTTC CCGAGGCTC	22	76.8	CGCACCAGCCCGG CTTCTGT	20	76.8	atttgcgaaatgt atctaagtgtcaa actaaatctac	37	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	80	80	160
chr9p21.1	chr9:32,987,566-32,987,612	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc caccttttc ag	38	CTTCTTTAGGGGC ACAGAGCATTGG	25	72.2	CCAGAGTTGCTCC CAGTTCCA	22	73.2	atttgcgaaatgt atctaagtgtcaa actaaatctac	37	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	82	82	164
chr9p13.1	chr9:38,752,528-38,752,568	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc caccttttc agctcgcgc	45	TCTGGGTCTGGAA GGGCTGC	20	71.6	GCCAGCCTGGCCT GCATCTGA	21	76.4	accatttgcgaaa tgtatctaattggtc aaactaaatctac	40	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	84	84	168
ctrl_4	chr11:14,515,205-14,515,256	GGGTTCCCTA AGGGTTGGA	19	cgctactac tattagtag aattgatgc caccttttc agctcgcgc	44	TGCATGTTGGAG CATCGACACA	23	70.4	GCTATGTTAGAAG AAATGCTGTTTTG GCC	29	70.5	tgcgaaatgtatc taagtgtcaact aaatctac	34	TCTAGATTGGATCTTGC TGGC <u>GC</u>	23	86	86	172

PSS- sekwencja komplementarna do startera (ang. primer-specific sequence); SS- dodatkowa sekwencja wypełniająca (ang. stuffer sequence); TSS- sekwencja komplementarna do regionu docelowego (ang. target-specific sequence);

HPL- długość półsondy (ang. half-probe length); TPL- całkowita długość sondy (ang. total-probe length)

### 1.1.9 Sprzęt laboratoryjny

- pH-metr pHenomenal 1100L (VWR)
- Spektrofotometr NanoDrop One (ThermoScientific)
- Fluorymetr Qubit 3.0 (Invitrogen)
- System dokumentacji żeli Q9 Alliance (Uvitec Cambridge)
- Termocykler Masterecycler nexus (Eppendorf)
- Termocykler Thermal Cycler T100 (BioRad)
- Wirówka 5424 R (Eppendorf)
- Wirówko-worteks Multi-spin MSC-6000 (BioSan)
- Zasilacz do elektroforezy PowerPac Basic (BioRad)
- Aparat do elektroforezy Wide Mini-Sub cell GT (BioRad)
- Waga laboratoryjna precyzyjna model LPC-423p (VWR)
- Mieszadło magnetyczne ES 21H (Wigo)
- Termoblok TS-100C (bioSan)
- Worteks V-1 plus (bioSan)
- Homogenizator TissueLyser LT (Qiagen)
- Stalowe kulki do homogenizacji, średnica 5mm (Qiagen)

### 1.1.10 Programy komputerowe i bazy danych

- Primer3 – projektowanie starterów do PCR  
<https://primer3.ut.ee/>
- UCSC-Genome Browser – baza danych sekwencji genomowych  
<https://genome.ucsc.edu/>
- GDC – baza Genomic Data Commons zawierająca dane mutacyjne z projektu TCGA  
<https://portal.gdc.cancer.gov/>
- MuSiCa – analiza sygnatur nowotworowych  
<https://toolsr.fcrb.es/sample-apps/MuSiCa/>
- SnpEff – adnotacja rodzaju mutacji  
[https://pcingola.github.io/SnpEff/se\\_introduction/](https://pcingola.github.io/SnpEff/se_introduction/)
- OncodriveFML – identyfikacja genów napędzających  
<https://oncodrivefml.readthedocs.io/en/latest/oncodriveFML.html>

## MATERIAŁY I METODY

- GISTIC 2.0 – identyfikacja regionów driverowych  
[https://www.genepattern.org/modules/docs/GISTIC\\_2.0](https://www.genepattern.org/modules/docs/GISTIC_2.0)
- ATGpr oraz NetStart 1.0 - określanie wyboru kodonu start w procesie inicjacji translacji
- ESEfinder - identyfikacja mutacji w miejscu akceptorowym/donorowym splicingu lub w wyciszaczach/wzmacniaczach splicingu
- LaBranchoR - identyfikacja mutacji w miejscu rozgałęzienia w intronach
- VEP tool – identyfikacja wpływu mutacji na sekwencję genu, transkryptu i białka  
[https://www.ensembl.org/Homo\\_sapiens/Tools/VEP](https://www.ensembl.org/Homo_sapiens/Tools/VEP)
- ProteinPaint – wizualizacja mutacji  
<https://proteinpaint.stjude.org/>
- miRMut - oprogramowanie stworzone w Zakładzie Genetyki Molekularnej ICHB PAN, służące do identyfikacji mutacji w genach miRNA oraz wizualizacji mutacji na strukturze genu (Urbanek-Trzeciak, Kozłowski, & Galka-Marciniak, 2022)
- TargetScanHuman wersja 8.0 – przewidywanie miRNA wiążących się do wskazanych genów  
[http://www.targetscan.org/vert\\_80/](http://www.targetscan.org/vert_80/)
- TargetScanHuman Custom 5.2 - przewidywanie genów, do których wiążą się wskazane miRNA  
[http://www.targetscan.org/vert\\_50/seedmatch.html](http://www.targetscan.org/vert_50/seedmatch.html)
- RegRNA 2.0, MotifMap – analiza występowania motywów regulatorowych oraz wiązania czynników transkrypcyjnych  
<http://regrna.mbc.nctu.edu.tw/>  
<http://motifmap.ics.uci.edu/>
- miRBase – baza danych sekwencji miRNA
- miRGeneDB – baza danych miRNA
- UniProt – baza danych sekwencji i domen funkcjonalnych białek  
<https://www.uniprot.org/>
- MFold – przewidywanie struktur RNA  
<http://www.unafold.org/>
- VARNA - wizualizacja struktury drugorzędowej RNA
- RNAComposer – przewidywanie struktur przestrzennych RNA  
<https://rnacomposer.cs.put.poznan.pl/>

## MATERIAŁY I METODY

- PyMol 2.5.3 – trójwymiarowa analiza struktur przestrzennych RNA (Shrodinger, LLC)
- Chromas 2.5 – analiza sekwencji DNA
- GeneMarker v2.4.0 (SoftGenetics) – analiza intensywności sygnału produktów PCR po reakcji MLPA
- Klaster obliczeniowy Eagle dostępny w ramach grantu obliczeniowego w PCSS (Poznańskie Centrum Superkomputerowo-Sieciowe)

### 1.2 Metody

#### 1.2.1 Homogenizacja tkanek oraz izolacja materiału genetycznego

Probówkę o pojemności 2 ml z umieszczoną w środku stalową kulką o średnicy 5mm (Qiagen) chłodzono na suchym lodzie przez 15 minut. Tkankę skórną wyciągnięto z roztworu RNeasy Lysis Buffer (Qiagen), pozbyto się nadmiaru odczynnika i wycięto skalpelem fragment tkanki nieprzekraczający wagi 20 mg. Następnie zważony fragment przełożono do uprzednio schłodzonej probówki, do której dodano także 600  $\mu$ l buforu do lizy RLT Plus (Qiagen) z dodatkiem reagentu DX (Qiagen) zapobiegającemu pienieniu się buforu. Ograniczenie do 20 mg wynikało z zaleceń dołączonych do instrukcji użytkowania homogenizatora. Ograniczenie wielkości homogenizowanej tkanki skórnej wpływa na wzrost efektywności homogenizacji. Homogenizację prowadzono przez 2 minuty w 50 Hz.

Izolację genomowego DNA oraz całkowitego RNA przeprowadzono zgodnie z wytycznymi zawartymi w protokole załączonym do zestawu AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). Warto wspomnieć, że zgodnie z protokołem, RNA poddano działaniu DNAzy I w celu pozbycia się zanieczyszczenia genomowym DNA. RNA rozpuszczono w wodzie i przechowywano w  $-80^{\circ}\text{C}$ . DNA rozpuszczono w buforze do elucji EB i przechowywano w  $-20^{\circ}\text{C}$ .

#### 1.2.2 Ocena ilościowa i jakościowa kwasów nukleinowych

Stężenie DNA i RNA oznaczono na podstawie pomiarów absorbancji przy długości fali 260 nm ( $A_{260}$ ) z użyciem spektrofotometru NanoDrop One (ThermoScientific). Czystość materiału genetycznego określono na podstawie stosunku absorbancji  $A_{260}/A_{280}$  oraz  $A_{260}/A_{230}$ . W przypadku DNA, które ma być przeznaczone do przygotowania bibliotek do sekwencjonowania nowej generacji, stężenie materiału genetycznego musi być oszacowane z najwyższą dokładnością. W związku z tym, ilościowy pomiar otrzymany z NanoDropa potwierdzano z użyciem alternatywnej metody – pomiaru z użyciem fluorymetru Qubit 3.0 (Invitrogen) z zastosowaniem zestawu dsDNA Qubit HS Assay Kit.

Dodatkowo, w celu potwierdzenia ilości i jakości (braku degradacji) materiału genetycznego przeprowadzono rozdział elektroforetyczny prób. Elektroforezę prowadzono w 1% żelu agarozowym w buforze 0,5x TBE. Do wystudzonego roztworu agarozy, bezpośrednio przed wylaniem do aparatu do elektroforezy, dodawano SimplySafe (5  $\mu$ l na 100 ml buforu TBE), który pełni funkcję interkalatora. Na żel nakładano 6  $\mu$ l próby (1  $\mu$ l DNA/RNA + 5  $\mu$ l obciążacz

## MATERIAŁY I METODY

stężony 6x). Rozdział prowadzono przez 10-15 min w 100 V w przypadku DNA, oraz 30-40 min w 100 V w przypadku RNA.

### **1.2.3 Strącenie kwasów nukleinowych**

W celu otrzymania DNA i RNA o możliwie najwyższej czystości część prób, dla których parametry A260/280 oraz A260/230 wykazywały odstępstwa od pożądanych wartości (A260/280 około 1,8 dla DNA oraz 2,0 dla RNA, A260/230 około 1,8-2,0), poddano precipitacji z użyciem etanolu. W celu wytrącenia kwasów nukleinowych z roztworu wodnego dodawano mieszaninę składającą się z 3 M octanu sodu stanowiącego 1/10 objętości roztworu wodnego kwasu nukleinowego oraz 3 objętości 100% alkoholu etylowego. Całość inkubowano przez noc w -20°C, a następnie wirowano w 4°C przez 30 minut w 13 000 rpm. W kolejnym kroku usuwano supernatant i wirowano przez 10 s przy maksymalnej prędkości wirówki. Powtórnie usuwano resztki supernatantu, a osad suszono. Osad zawierający RNA rozpuszczono w wodzie, a DNA w buforze do elucji EB. Następnie przeprowadzono rozdział elektroforetyczny według zasad przedstawionych w punkcie wyżej.

### **1.2.4 Przygotowanie DNA do utworzenia bibliotek**

300 ng DNA (200 ng do stworzenia biblioteki z nadmiarem) odpipetowano do probówek o pojemności 1,5 ml. W przypadku wysokiego stężenia DNA próby rozcieńczano buforem TE. Końcowa objętość prób wynosiła 15-30 µl. W celu zabezpieczenia przed otwarciem, każda probówka została dokładnie owinięta parafilmem. Następnie probówki umieszczone w suchym lodzie wraz z wkładami chłodzącymi wysłano do firm CeGaT (Niemcy) oraz Macrogen (Korea), którym zlecono przygotowanie, wzbogacenie i sekwencjonowanie bibliotek z użyciem sekwenatora nowej generacji.

### **1.2.5 Przygotowanie, wzbogacenie oraz sekwencjonowanie bibliotek DNA**

#### **1.2.5.1 Biblioteka DNA do WES**

Przygotowanie biblioteki zlecono firmie CeGaT i wykonano z użyciem zestawu SureSelectXT Library Prep Kit (Agilent). Bibliotekę przygotowano z 200 ng DNA zgodnie z instrukcjami producenta. Wzbogacenie biblioteki polegające na zastosowaniu sond komplementarnych do docelowego regionu o wielkości 60 Mbp, wykonano z użyciem zestawu SureSelectXT Human All Exon V6 (Agilent), umożliwiającego pokrycie regionu, obejmującego nie tylko części kodujące genów, ale także 5'UTR oraz fragmenty 3'UTR. Wzbogaconą bibliotekę poddano sekwencjonowaniu wysokoprzepustowemu w trybie sparowanych końców (ang. *paired-end*) 2 x 100 bp z użyciem platformy firmy Illumina na aparacie NovaSeq 6000.

#### **1.2.5.2 Biblioteka DNA do WMS**

Wykonanie biblioteki zlecono firmie MacroGen. Do przygotowania biblioteki zastosowano zestaw SureSelectXT Library Prep Kit (Agilent). Bibliotekę przygotowano z 200 ng DNA zgodnie z instrukcjami producenta. Wzbogacenie biblioteki wykonano z użyciem miRNomu- panelu SureSelect Custom Target Tier 1 (1-499 Kbp) (Agilent) zaprojektowanego w Zakładzie Genetyki Molekularnej i nazwanego miRNom.

Zaprojektowane sondy pokrywały regiony o łącznej wielkości 0,50 Mbp, i umożliwiały pokrycie 1849 genów miRNA (miRNA obecne w bazie danych miRBase wersja v21; koordynaty genomowe obejmują sekwencje pre-miRNA oraz sekwencje flankujące prekursor obejmujące 25 nukleotydów występujących przed i po sekwencji pre-miRNA), 461 eksonów 28 genów biorących udział w biogenezie miRNA oraz 92 eksonów 10 genów kontrolnych, często mutowanych w wielu typach nowotworów. Wzbogaconą bibliotekę poddano sekwencjonowaniu wysokoprzepustowemu 2 x 100 bp z użyciem platformy firmy Illumina i aparatu HiSeq 4000.



### 1.2.6 Analiza danych otrzymanych z sekwencjonowania NGS

Wstępne etapy analizy danych przeprowadzono w firmie CeGaT (w przypadku WES) oraz Macrogen (w przypadku WMS). Dla każdej sekwencjonowanej próbki otrzymano po dwa pliki w formacie FASTQ wynikające z przeprowadzania sekwencjonowania w dwóch kierunkach 2 x 100 pz. Pliki zdeponowano w zewnętrznym repozytorium danych BioProject/BioSample, PRJNA747148. Plik FASTQ to tekstowy format do przechowywania zarówno sekwencji nukleotydów, jak i odpowiadających jej wyników jakości. Jakość sekwencji została zakodowana w skali phred +33 a rozkład jakości sprawdzono z użyciem FastQC. Demultipleksowanie odczytów sekwencjonowania przeprowadzono za pomocą Illumina bcl2fastq (wersja 2.19). Adaptory zostały przycięte z użyciem Skewer (wersja 0.2.2) (Jiang, Lei, Ding, & Zhu, 2014). Kolejne etapy wykonano na superkomputerze Eagle w Poznańskim Centrum Superkomputerowo-Sieciowym (PCSS). Przeniesienie plików na maszynę obliczeniową wykonano za pomocą WinSCP. Skrypty umożliwiające analizę danych uruchamiano w powłoce BASH. Łączenie z superkomputerem oraz uruchamianie skryptów wykonano za pomocą oprogramowania PuTTY. Przy użyciu Burrows-Wheeler Alignment Tool (BWA) wykonano mapowanie odczytów (pliki fastq) do sekwencji referencyjnej genomu człowieka w wersji Hg38, co umożliwiło otrzymanie pliku w formacie SAM. Duplikaty PCR zostały usunięte z użyciem Picard. Rekalibrację oceny jakości detekcji nukleotydów wykonano z użyciem GATK (wersja 4.1.2.0). Konwersja SAM do BAM (binarna wersja pliku SAM) została wykonana przy użyciu SAMtools. Następnie w celu przypisania (ang. *calling*) somatycznych mutacji punktowych i indeli zastosowano MuTect2 (wersja 4.1.0.0. z wykorzystaniem trybu nowotwór – próbka normalna). Z pomocą MuTect2 warianty dodatkowo oznaczono za pomocą panelu prób normalnych (ang. *panel of normals*; PoN), obejmującego warianty reprezentujące powszechnie występujące błędy sekwencjonowania, które mogą naśladować warianty somatyczne o niskiej frakcji alleli. Aby uniknąć fałszywie pozytywnych mutacji somatycznych, przeprowadzono filtrowanie pod kątem wariantów germinalnych obecnych w bazie danych gnomAD (wersja 2.1.1) oraz częstych wariantów  $\geq 1\%$  (ang. *common SNP*) obecnych w bazie dbSNP (wersja 147 dla WES oraz wersja 151 dla WMS). Z listy otrzymanych mutacji somatycznych dodatkowo usunięto te, które nie spełniały następujących kryteriów: (i) co najmniej pięć alternatywnych odczytów występujących w próbce nowotworowej w przypadku WES oraz 2 odczyty w przypadku WMS; (ii) AAF co najmniej 0,05 w przypadku WES oraz 0,01 dla WMS; (iii) ilość alternatywnych odczytów w próbce nowotworowej co najmniej

## MATERIAŁY I METODY

5 razy wyższa niż w odpowiadającej próbce normalnej. Dalsza analiza wariantów wykrytych dzięki WES oraz WMS nieznacznie się od siebie różniła. W przypadku WES do końcowej listy mutacji dodano informacje o lokalizacji mutacji w subregionach genów kodujących białka (CDS, 5'UTR, 3'UTR lub introny). Adnotacje dodano za pomocą skryptów przygotowanych w języku programowania python przez dr Martynę Urbanek-Trzeciak. Aby przypisać rodzaj mutacji do podanych wariantów zastosowano SnpEff, dostępny na platformie Subio-oprogramowanie do analizy danych muti-omicznych (Subio, Inc., Kagoshima, Japonia, <http://www.subio.jp>). Dodatkowo, zgodnie z nomenklaturą HGVS wybrane zmiany wykryte w genach kodujących białka oznaczono względem pozycji w sekwencji kodującej (oznaczone literą „c.”). Nazwy genów określono zgodnie z nazewnictwem według HUGO Gene Nomenclature Committee (HGNC; Komitet Nomenklatury Genów).

W przypadku WMS w celu przypisania mutacji do poszczególnych genów miRNA użyto narzędzia miRMut (Urbanek-Trzeciak et al., 2022). Mutacje wykryte w genach miRNA oznaczono względem odpowiednich transkryptów prekursorów miRNA wg bazy miRBase (zgodnie z nomenklaturą HGVS oznaczone literą „n.”). Określono także wiarygodność miRNA według dwóch baz danych miRBase oraz MiRGeneDB. Wiarygodność miRNA określana w bazie danych miRBase zależy od poziomu ekspresji wiodącej nici miRNA bazując na danych z sekwencjonowania małych cząsteczek RNA (ang. *small RNA sequencing*). Im wyższa ilość odczytów mapujących się w regionie miRNA tym wyższe prawdopodobieństwo prawidłowego zakwalifikowania cząsteczki do grupy miRNA. miRBase zawiera miRNA zarówno o wysokiej jak i niskiej wiarygodności. MiRGeneDB zawiera natomiast listę wyłącznie miRNA o wysokiej wiarygodności co wiąże się z mniejszą ilością miRNA zdeponowanych w tej bazie. Wiarygodność jest określana na podstawie konserwatywności nici wiodącej miRNA (Fromm et al., 2015).

### 1.2.7 Analiza sygnatur nowotworowych

W celu wykrycia sygnatur mutacyjnych wykorzystano program MuSiCa (ang. *Mutational Signatures in Cancer*; (Díaz-Gay et al., 2018) (<http://bioinfo.ciberehd.org:3838/MuSiCa/>). Jako pliki wejściowe posłużyły pliki z rozszerzeniem VCF zawierające listę mutacji somatycznych zebranych osobno dla każdej próbki nowotworowej. Dwie próbki (BCC14 oraz BCC21) zostały wykluczone z analizy ze względu na bardzo niską liczbę mutacji somatycznych zidentyfikowanych w tych próbkach.

### 1.2.8 Wykrywanie gorących miejsc mutacji

Jako gorące punkty mutacji sklasyfikowano pozycje genomowe zmutowane w co najmniej 3 próbkach BCC (>10% analizowanych próbek BCC). Mutacje występujące w bezpośrednio sąsiadujących pozycjach genomowych łączono w jeden gorący punkt. Do wykrycia gorących punktów zastosowano skrypt napisany w języku python, przygotowany przez dr Martynę Urbanek-Trzeciak.

### 1.2.9 Analiza mutacji dostępnych w bazie danych TCGA

W celu porównania częstych mutacji wykrytych w BCC z innymi typami nowotworów, przeanalizowano dane z WES dostępne dla 10 369 próbek reprezentujących 33 typy nowotworów. Dane te wygenerowano i zdeponowano w repozytorium TCGA (<http://cancergenome.nih.gov>). Mutacje somatyczne zidentyfikowano za pomocą standardowego potoku analitycznego (ang. *pipeline*) TCGA analizującego pary próbek (nowotwór i zdrowa tkanka) z zastosowaniem algorytmów takich jak Mutect2, Muse, VarScan i SomaticSnipper. Za mutacje somatyczne uznano tylko zmiany z adnotacją "PASS". Ekstrakcję mutacji na podstawie danych z TCGA przeprowadzono dla mutacji zlokalizowanych tylko w wybranych genach, często mutowanych w BCC (*BAD*, *DHODH*, *CHCHD2*, *FLG* oraz *FLG2*). Analizę przeprowadzono z użyciem skryptu napisanego w języku python przez dr Martynę Urbanek-Trzeciak (Galka-Marciniak i in., 2021).

### **1.2.10 Wizualizacja mutacji w genach kodujących białko oraz genach miRNA**

Skutki mutacji określono za pomocą narzędzia Ensembl Variant Effect Predictor (VEP). Do wizualizacji mutacji na schematach przedstawiających budowę poszczególnych genów zastosowano narzędzie ProteinPaint (Zhou et al., 2016). Domeny białkowe przedstawione na mapach zostały umieszczone zgodnie z danymi dostępnymi w bazie UniProt (Apweiler et al., 2004).

Wykres komutacyjny pokazujący często mutowane geny został stworzony przy użyciu biblioteki Python CoMut (Crowdis, He, Reardon, & Van Allen, 2020).

W celu wizualizacji mutacji wzdłuż genów miRNA zastosowano algorytm miRMut (Urbanek-Trzeciak et al., 2022).

### **1.2.11 Wpływ mutacji na wybór kodonu start i efektywność inicjacji translacji**

Analiza wyboru kodonu start w procesie inicjacji translacji została przeprowadzona z użyciem narzędzi ATGpr (Salamov et al., 1998) oraz NetStart 1.0 (Peitsch, 1997). Analizie poddano sekwencję cDNA (o minimalnej długości 50 nt) typu dzikiego oraz z mutacją w sekwencji Kozak.

### **1.2.12 Analiza i wizualizacja struktury drugorzędowej 2D i 3D RNA**

Przewidywanie struktury drugorzędowej dla podanej sekwencji RNA wykonano z użyciem narzędzia Mfold (Zuker, 2003). Predykcję struktury zapisano w formacie Vienna (format kropkowo-nawiasowy) do wykorzystania w dalszych analizach. Parametry programu zostały ustalone w taki sposób, aby przewidywać optymalną strukturę RNA w temperaturze 37°C oraz 1 M NaCl. Mfold umożliwia także ocenę stabilności termodynamicznej struktury tworzonej przez RNA. Wizualizację struktury przewidywanej przez Mfold przeprowadzono z użyciem VARNA (Darty, Denise, & Ponty, 2009).

Struktury określone za pomocą Mfold zastosowano do przewidywania struktury przestrzennej (3D) z zastosowaniem RNAComposer (Popenda et al., 2012) a wizualizację struktury 3D wykonano z użyciem PyMOL (wersja 2.5.3). PyMOL umożliwił określenie rozbieżności struktur 3D tworzonych przez geny miRNA bez oraz z mutacjami. Podobieństwo struktur analizowano za pomocą komendy align.

### **1.2.13 Przewidywanie miRNA regulujących ekspresję genów**

Przewidywanie miRNA wiążących się w regionie 3'UTR wybranych genów przeprowadzono za pomocą narzędzia TargetScanHuman wersja 8.0. Po wpisaniu nazwy genu oraz wyborze odpowiedniego transkryptu otrzymano listę miRNA mogących wiązać się do wybranego wcześniej genu wraz ze wskazaniem sekwencji, do której wiąże się miRNA.

### **1.2.14 Przewidywanie genów, których ekspresję mogą regulować miRNA**

Przewidywanie transkryptów genów, do których może wiązać się miRNA przeprowadzono za pomocą narzędzia TargetScanHuman Custom wersja 5.2. W wyszukiwarce wpisywano sekwencję źródłową miRNA bez mutacji lub z mutacją i otrzymano listę genów, do których może wiązać się miRNA. Wizualizację porównania listy genów, do których wiąże się miRNA wykonano za pomocą diagramu Venna (<https://bioinformatics.psb.ugent.be/webtools/Venn/>), który przedstawia jaka część genów jest osobna a jaka wspólna dla dzikiego i mutowanego typu miRNA.

### **1.2.15 Wykrywanie mutacji napędzających**

Wykrycie mutacji napędzających wykonano z użyciem OncodriveFML (Mularoni et al., 2016). Zmiany napędzające analizowano nie tylko pod kątem kontekstu nukleotydowego (sygnatury) i częstości występowania mutacji, ale także szkodliwości (ang. *deleteriousness*) mutacji, którą oceniano z użyciem narzędzia CADD v.1.6 ((Rentsch et al., 2019); Hg38, wersja 1.6). Analizę wykonano osobno dla każdego z analizowanych regionów: region kodujący, 5'UTR, 3'UTR, introny, miRNA. Zastosowano następujące parametry OncodriveFML: sygnatura "complement" analizująca mutacje przy użyciu 96 różnych kontekstów nukleotydowych; metoda do porównywania wartości obserwowanych i symulowanych "amean"- średnia arytmetyczna; indele włączone do analiz; wartość  $q < 0.1$ , oraz  $< 0.25$ .

### **1.2.16 Analiza regionów istotnie wzbogaconych w zmiany liczby kopii**

W celu wykrycia regionów ulegających znaczącej amplifikacji lub delecji wykorzystano GISTIC2.0 (Mermel et al., 2011). GISTIC2.0 umożliwia wykrywanie zmian o potencjale napędzającym, występujących na poziomie ramienia chromosomu lub na poziomie zmian ogniskowych (fokalnych). Plik segmentowy konieczny do uruchomienia GISTIC2.0 wygenerowano za pomocą narzędzia CNVkit (Talevich, Shain, Botton, & Bastian, 2016).

## MATERIAŁY I METODY

Analizę z użyciem GISCITC2.0 wykonano z zastosowaniem następujących parametrów: próg dla amplifikacji/delecji wynosi +/- 0,2; poziom ufności 0,9; wartość  $q < 0,25$ ; zastosowanie metody 'arm peel-off'; analiza 'broad-level' biorąca pod uwagę zmiany powyżej połowy długości ramienia chromosomu. Analizę wykonano z użyciem superkomputera Eagle. Skrypt został przygotowany przez dr Natalię Szóstak z Pracowni Bioinformatyki ICHB PAN.

### 1.2.17 Reakcja odwrotnej transkrypcji (RT)

Reakcję odwrotnej transkrypcji przeprowadzono z wykorzystaniem zestawu SuperScript IV First-Strand Synthesis System (Invitrogen) zgodnie z protokołem producenta. Do reakcji użyto 500 ng RNA, a jako startery posłużyły heksamery o losowej sekwencji (ang. *random hexamers*). Mieszaninę inkubowano w 23°C przez 10 min, 55°C przez 10 min oraz 80°C przez 10 min. Produkty reakcji przechowywano w -20°C. Uzyskany w wyniku reakcji odwrotnej transkrypcji komplementarny DNA (cDNA) wykorzystywano jako matryce w reakcji PCR.

### 1.2.18 Łańcuchowa reakcja polimerazy (PCR)

Reakcję PCR przeprowadzano w celu amplifikacji regionów posiadających mutacje somatyczne zidentyfikowane przez sekwencjonowanie NGS. Składniki mieszaniny reakcyjnej oraz warunki reakcji przedstawiono w poniższych tabelach (Tabela 12 oraz 13). Jako matrycę stosowano genomowy DNA lub cDNA.

**Tabela 12. Skład mieszaniny do reakcji PCR.**

Składnik	Ilość
bufor Green GoTaq 5x	2 $\mu$ l
starter <i>forward</i> F (10 $\mu$ M)	1 $\mu$ l
starter <i>reverse</i> R (10 $\mu$ M)	1 $\mu$ l
dNTP (10 mM)	0,2 $\mu$ l
polimeraza GoTaq (5U/ $\mu$ l)	0,05 $\mu$ l
DNA	10-50 ng
Woda	do 10 $\mu$ l

**Tabela 13. Warunki reakcji PCR.**

Etap	Temperatura	Czas	Liczba cykli
Denaturacja wstępna	95°C	5 min	1
Denaturacja	95°C	30 s	35-40
Przyłączanie starterów	57-63 °C	30 – 60 s	
Wydłużanie	72 °C	45 s	
Synteza końcowa	72 °C	5 min	1

### 1.2.19 Sekwencjonowanie produktów PCR metodą sekwencjonowania Sangera

Sekwencjonowanie produktów PCR zlecono Pracowni Techniki Biologii Molekularnej UAM. Do sekwencjonowania wykorzystano startery o stężeniu 10 µM. Standardowo do sekwencjonowania przygotowywano matrycę o objętości 9 µl i dołączano zdjęcie po rozdiale elektroforetycznym 1 µl matrycy. Wszystkie fragmenty poddano sekwencjonowaniu w dwóch kierunkach za pomocą zestawu BigDye v3.1 (Applied Biosystems, Foster City, CA, USA). Produkty sekwencjonowania rozdzielono metodą elektroforezy kapilarnej (polimer POP7; aparat ABI Prism 3130xl; Applied Biosystems, Foster City, CA, USA) zgodnie z zaleceniami producenta. Pliki w formacie AB1 zawierające wyniki sekwencjonowania były wizualizowane z użyciem programu Chromas (wersja 2.5).

### 1.2.20 Reakcja MLPA

Test MLPA zaprojektowano w celu weryfikacji zmian liczby kopii wykrytych w części chromosomu 9 przy udziale algorytmu GISTIC 2.0. Test składa się z 20 sond, w tym (i) 7 sond rozmieszczonych wzdłuż ramienia chr9p (5 sond) i chr9q (2 sondy), (ii) po 2 sondy zlokalizowane w samym genie lub w pobliżu genów *JAK2*, *PDL2*, *PDL1* i *PTCH1* (łącznie 8 sond) oraz (iii) 5 sond kontrolnych (zlokalizowanych na różnych chromosomach poza chromosomem 9 i regionami genów związanych z rakiem). Sekwencje i szczegółowa charakterystyka sond zostały podane w tabeli w punkcie 8.1.8 (Materiały). Każda sonda składa się z dwóch półsond o jednakowej lub różniącej się o 1 nt długości. Całkowita długość sondy waha się od 93 do 172 nt. Sekwencje docelowe dla sond wybrano tak, aby uniknąć miejsc występowania częstych

## MATERIAŁY I METODY

SNP, regionów powtórzeniowych oraz sekwencji o wyjątkowo wysokiej lub niskiej zawartości par GC.

Przed przystąpieniem do reakcji przygotowano mieszaninę sond (mix). W tym celu przeniesiono po 2  $\mu$ l każdej półsondy o stężeniu 0,2 mM do próbki o pojemności 1,5 ml i uzupełniono wodą do końcowej objętości 400  $\mu$ l. Tak przygotowaną mieszaninę przechowywano w temperaturze 4°C. Reakcję MLPA przeprowadzono zgodnie z ogólnymi zaleceniami producenta zestawu do MLPA (MRC-Holland). W pierwszym etapie 5  $\mu$ l DNA (50-250 ng) poddano denaturacji w 98°C przez 5 minut. Następnie próbki schłodzono do temperatury 25°C oraz dodano 1,5  $\mu$ l buforu do MLPA oraz 1,5  $\mu$ l uprzednio przygotowanej mieszaniny sond. Całość dokładnie zmieszano przez przepipetowanie zawartości próbki. Następnie przeprowadzono denaturację w 95°C przez 1 min oraz hybrydyzację sond w 60°C przez 16-20 h. W kolejnym etapie przygotowano 32  $\mu$ l mieszaniny, w której skład wchodziło: 3  $\mu$ l buforu A oraz 3  $\mu$ l buforu B (bufory do ligazy), 25  $\mu$ l wody oraz 1  $\mu$ l ligazy. Temperaturę w termocyklerze obniżono do 54°C i w momencie, gdy próbki znajdowały się w termocyklerze dodano uprzednio przygotowaną mieszaninę w celu przeprowadzenia ligacji sond, które ulegają hybrydyzacji do DNA. Ligację przeprowadzono w następujących warunkach: 15 min w 54°C, 5 min w 95°C, schłodzenie do 20°C. Następnie przeprowadzono amplifikację zligowanych sond. Do amplifikacji użyto uniwersalnych starterów komplementarnych do regionów 5' oraz 3' PSS zaprojektowanych sond. Skład mieszaniny reakcyjnej obejmował: 7,5  $\mu$ l wody, 2  $\mu$ l mieszaniny starterów, 0,5  $\mu$ l polimerazy. Mieszaninę dodano do reakcji MLPA i amplifikację przeprowadzono w następujących warunkach: 35 cykli denaturacji w 95°C przez 30 s, przyłączania starterów 60°C przez 30 s, oraz elongacji 72°C przez 1 min. Reakcję zakończono inkubacją w 72°C przez 20 min i schłodzeniem próbek do temperatury 20°C.

Następnie produkty reakcji MLPA zostały dostarczone do Pracowni Techniki Biologii Molekularnej UAM, gdzie zostały rozcieńczone 20x w formamidzie HiDi zawierającym GS Liz600, który zastosowano jako wzorzec wielkości DNA, i rozdzielone metodą elektroforezy kapilarnej (polimer POP7) w aparacie ABI Prism 3130XL (Applied Biosystems, Carlsbad, CA, USA).

Otrzymane elektroferogramy analizowano za pomocą oprogramowania GeneMarker v2.4.0 (SoftGenetics, State College, PA, USA). Dla każdej pojedynczej próbki, intensywność sygnału każdej sondy została podzielona przez średnią geometryczną intensywności sygnału sond kontrolnych w celu znormalizowania zmienności sygnału. Następnie znormalizowany sygnał



## MATERIAŁY I METODY

każdej sondy w próbkach nowotworu podzielono przez sygnał w normalnych/zdrowych próbkach i pomnożono przez 2. Ostateczny wynik MLPA dla każdej próbki przedstawiano na wykresie słupkowym, na którym wysokość słupka obrazuje względną wartość liczby kopii regionów oznaczonych poszczególnymi sondami.

### **1.2.21 Dodatkowe analizy statystyczne**

Obliczenia statystyczne wykonano w środowisku R. Dla każdego testu statystycznego obliczano wartość  $p$ , za istotne uznano wartości  $p < 0,05$ . Stosowane testy statystyczne to: i) test Manna-Whitneya-Wilcoxon (stosowany przy analizach różnic w wartości TMB oraz przy porównaniu kontrybucji mutacji do sygnatur pomiędzy podtypami BCC; ii) test Fishera (stosowany przy analizie wzajemnie wykluczających się mutacji); iii) test T-studenta dla prób niezależnych (stosowany przy określaniu wpływu mutacji w *KNSTRN* na ilość zmian liczby kopii); iv) rozkład dwumianowy (stosowany przy analizach rozkładu mutacji wzdłuż prekursorów miRNA). W przypadku testowania wielokrotnego wprowadzono korektę wartości  $p$  za pomocą procedury Benjaminiego-Hochberga.

## X. SPIS RYCIN I TABEL

### 1. Ryciny

<i>Rycina 1. Schematyczny rozkład mutacji charakterystycznych dla onkogenów i genów supresorowych.</i>	12
<i>Rycina 2. Schematyczny proces biogenezy miRNA.</i>	23
<i>Rycina 3. Schemat szlaku sygnałowego Shh.</i>	28
<i>Rycina 4. Geny często mutowane w BCC.</i>	29
<i>Rycina 5. Rozdział elektroforetyczny DNA oraz RNA próbek BCC.</i>	34
<i>Rycina 6. Schemat regionów poddanych sekwencjonowaniu WES.</i>	36
<i>Rycina 7. Lokalizacja mutacji w BCC.</i>	38
<i>Rycina 8. Typy mutacji występujące w BCC.</i>	39
<i>Rycina 9. Porównanie obciążenia mutacyjnego pomiędzy podtypami BCC.</i>	42
<i>Rycina 10. Obciążenie mutacyjne w nowotworach.</i>	43
<i>Rycina 11. Sygnatury mutacyjne w BCC oraz innych typach nowotworów.</i>	44
<i>Rycina 12. Sygnatury mutacyjne charakterystyczne dla BCC.</i>	45
<i>Rycina 13. Kontrybucja mutacji do sygnatury 7 oraz 11 w podtypach BCC.</i>	46
<i>Rycina 14. Rozkład mutacji w genie TNFAIP2.</i>	51
<i>Rycina 15. Rozkład mutacji w genie EZH2.</i>	52
<i>Rycina 16. Rozkład mutacji w genie KNSTRN.</i>	53
<i>Rycina 17. Wpływ mutacji w KNSTRN na zmiany liczby kopii w BCC.</i>	54
<i>Rycina 18. Rozkład mutacji w BAD.</i>	56
<i>Rycina 19. Rozkład mutacji w DHODH.</i>	57
<i>Rycina 20. Rozkład mutacji w CHCHD2.</i>	59
<i>Rycina 21. Rozkład mutacji niesynonimicznych w genie PTCH1.</i>	61
<i>Rycina 22. Mutacje w genach szlaku Shh.</i>	62
<i>Rycina 23. Rozkład mutacji niesynonimicznych w genie TP53.</i>	63
<i>Rycina 24. Rozkład mutacji niesynonimicznych w genie PTPRD.</i>	64
<i>Rycina 25. Rozkład mutacji niesynonimicznych w genie FLG i FLG2.</i>	64
<i>Rycina 26. Częstość mutacji w genach FLG i FLG2.</i>	65
<i>Rycina 27. Rozkład mutacji niesynonimicznych w genie MYCN.</i>	66
<i>Rycina 28. Rozkład mutacji niesynonimicznych w genach NOTCH.</i>	67
<i>Rycina 29. Rozkład mutacji w genie SPHK2.</i>	68
<i>Rycina 30. Rozkład mutacji w genie DROSHA.</i>	70
<i>Rycina 31. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji w regionie kodującym.</i>	72
<i>Rycina 32. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji w regionie 5'UTR.</i>	73
<i>Rycina 33. Geny napędzające rozwój BCC wyselekcjonowane na podstawie mutacji regionu 3'UTR.</i>	74
<i>Rycina 34. Regiony napędzające rozwój BCC.</i>	76
<i>Rycina 35. Zmiany liczby kopii w BCC na podstawie analiz MLPA.</i>	78
<i>Rycina 36. Schemat prekursora miRNA.</i>	80
<i>Rycina 37. Typy mutacji występujące w genach miRNA w BCC.</i>	81
<i>Rycina 38. Lokalizacja mutacji w subregionach prekursorów miRNA.</i>	82

<b>Rycina 39. Rozkład mutacji w subregionach prekursorów miRNA z zastosowaniem podziału ze względu na ramię, z którego powstaje dojrzała cząsteczka.</b>	84
<b>Rycina 40. Wpływ mutacji na strukturę prekursora miR-3928.</b>	88
<b>Rycina 41. Wpływ mutacji w regionie źródłowym na liczbę przewidzianych genów docelowych.</b>	90
<b>Rycina 42. Wpływ mutacji na strukturę prekursora miR-498 oraz analiza regionów wiążących miRNA.</b>	91
<b>Rycina 43. Mutacje w promotorze genu TERT zidentyfikowane w niniejszej pracy w próbkach BCC.</b>	92
<b>Rycina 44. Mutacje w promotorze genu DPH3 zidentyfikowane w niniejszej pracy w próbkach BCC.</b>	93
<b>Rycina 45. Geny, promotory i/lub szlaki często mutowane w BCC.</b>	95

## 2. Tabele

<b>Tabela 1. Rozkład i typy mutacji w BCC.</b>	37
<b>Tabela 2. Lista mutacji potwierdzanych z użyciem sekwencjonowania metodą Sangera.</b>	40
<b>Tabela 3. Gorące punkty mutacji wykryte w BCC.</b>	47
<b>Tabela 4. Liczba genów posiadających częste mutacje w poszczególnych częściach genów.</b>	60
<b>Tabela 5. Geny poza szlakiem Shh często mutowane w regionie kodującym.</b>	62
<b>Tabela 6. Geny często mutowane w regionie 5'UTR.</b>	68
<b>Tabela 7. Geny często mutowane w regionie 3'UTR.</b>	69
<b>Tabela 8. Rozkład mutacji w genach miRNA.</b>	85
<b>Tabela 9. Lista mutacji walidowanych metodą sekwencjonowania Sangera.</b>	86
<b>Tabela 10. Zestawienie starterów użytych do reakcji PCR.</b>	109
<b>Tabela 11. Sondy do reakcji MLPA.</b>	112
<b>Tabela 12. Skład mieszaniny do reakcji PCR.</b>	125
<b>Tabela 13. Warunki reakcji PCR.</b>	126

## XI. WYKAZ SKRÓTÓW

- 3'ss- miejsce akceptorowe splicingu (ang. *3' splice site, acceptor site*)  
3'UTR- region nieulegający translacji zlokalizowany po stronie 3' sekwencji kodującej białko (ang. *3' untranslated region*)  
5'ss- miejsce donorowe splicingu (ang. *5' splice site, donor site*)  
5'UTR- region nieulegający translacji zlokalizowany po stronie 5' sekwencji kodującej białko (ang. *5' untranslated region*)  
5mC- 5-metylocytozyna  
A-adenina  
aa- aminokwas  
AAF- frakcja zmutowanego allelu (ang. *alternative allele fraction*)  
AK- rogowacenie starcze/słoneczne (ang. *actinic keratosis*)  
ASC- średnia kontrybucja mutacji do sygnatur mutacyjnych (ang. *average signature contribution*)  
Asp- kwas asparaginowy  
BCC- rak podstawnokomórkowy skóry (ang. *basal cell carcinoma*)  
BP- miejsce rozgałęzienia (ang. *branch point*)  
C-cytozyna  
CADD- Combined Annotation Dependent Depletion  
cDNA- komplementarny DNA (ang. *complementary DNA*)  
CDS- region kodujący (ang. *coding sequence*)  
CGC- Cancer Gene Census  
COSMIC- Catalogue of Somatic Mutations in Cancer  
CPSF- czynnik specyficzności cięcia i poliadenylacji (ang. *cleavage and polyadenylation specific factor*)  
D- adenina lub guanina lub tymina  
DBS- substytucja dwóch zasad azotowych (ang. *double base substitution*)  
DLBCL- chłoniaki rozlane z dużych komórek B (ang. *diffuse large B-cell lymphoma*)  
DNA- kwas deoksyrybonukleinowy  
dsDNA- dwuniciowe DNA (ang. *double stranded DNA*)  
F- starter forward  
G-guanina  
HPL- długość półsondy (ang. *half-probe length*)  
ICGC- International Cancer Genome Consortium  
ID- indele  
IRES- wewnętrzne miejsce wiązania rybosomu (ang. *internal ribosome entry site*)  
ISE- wzmacniacz splicing (ang. *intronic splicing enhancer*)  
ISS- wyciszacz splicing (ang. *intronic splicing silencer*)  
K- guanina lub tymina  
Kpz- tysiąc par zasad  
LGG- nowotwory pochodzenia glejowego o niskim stopniu złośliwości (ang. *low grade glioma*)  
lncRNA- długie niekodujące RNA (ang. *long noncoding RNA*)  
miRISC- ang. *microRNA induced silencing complex*  
miRNA- mikroRNA (ang. *microRNA*)  
MLPA- metoda zależnej od ligacji multipleksowej amplifikacji sond (ang. *multiplex*)

## WYKAZ SKRÓTÓW

*ligation probe amplification*)

MMR- naprawa błędnie sparowanych zasad azotowych (ang. *mismatch repair*)

MOMP- permeabilizacja mitochondrialnej błony zewnętrznej (ang. *mitochondrial outer membrane permeabilization*)

Mpz – milion par zasad

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

MSI- niestabilność regionów mikrosatelitarnych (ang. *microsatellite instability*)

NGS- sekwencjonowanie nowej generacji (ang. *Next Generation Sequencing*)

nt – nukleotydy

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

PAS- sygnał poliadenylacji (ang. *polyadenylation signal*)

PCA- kwas pirolidonokarboksylowy

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

Phe- fenyloalanina

post-seed- nukleotydy w duplesie miRNA znajdujące się za regionem źródłowym

PPT- trakt polipirymidynowy (ang. *polypirimidine tract*)

pre-miRNA- (wtórny) prekursor miRNA

pre-seed- nukleotydy w duplesie miRNA znajdujący się przed regionem źródłowym

pri-miRNA- pierwotny prekursor miRNA, pierwotny transkrypt miRNA

Pro- prolina

PSS- sekwencja komplementarna do startera (ang. *primer-specific sequence*)

pz- pary zasad

R- adenina lub guanina

R- starter reverse

RNA- kwas rybonukleinowy

rpm- obroty na minutę (ang. *rotations per minute*)

rRNA- rybosomalny RNA (ang. *ribosomal RNA*)

S- cytozyna lub guanina

SBS – substytucja pojedynczej zasady azotowej (ang. *single base substitution*)

SCC- rak kolczystokomórkowy skóry (ang. *squamous cell carcinoma*)

seed- region źródłowy

Ser- seryna

snRNA- mały jądrowy RNA (ang. *small nuclear RNA*)

SS- dodatkowa sekwencja (ang. *stuffer sequence*)

Szlak Shh- szlak sygnałowy Sonic Hedgehog

T-tymina

TCGA- The Cancer Genome Atlas

TF- czynnik transkrypcyjny (ang. *transcription factor*)

Tm- temperatura topnienia (ang. *melting temperature*)

TMB- obciążenie mutacyjne nowotworu (ang. *tumor mutational burden*)

TPL- długość sondy (ang. *total-probe length*)

tRNA- transportujący RNA (ang. *transfer RNA*)

Trp- tryptofan

TSS- miejsce startu transkrypcji (ang. *transcription start site*)

TSS- sekwencja komplementarna do regionu docelowego (ang. *target-specific sequence*)

U-uracyl

UCA- kwas transurokanowy

## WYKAZ SKRÓTÓW

UV- światło ultrafioletowe

W- adenina lub tymina

WES- sekwencjonowanie całego eksomu (ang. *Whole Exome Sequencing*)

WMS- sekwencjonowanie całego miRNomu (ang. *Whole miRNome Sequencing*)

Y- cytozyna lub tymina

## XII. ŹRÓDŁA FINANSOWANIA BADAŃ

Instytut Chemii Bioorganicznej PAN w Poznaniu



Narodowe Centrum Nauki



GRANT NCN MAESTRO, kierownik Piotr Kozłowski

“Analiza somatycznych mutacji w genach miRNA i w genach biogenezy miRNA w celu identyfikacji nowych celów terapii i biomarkerów chorób nowotworowych”

(2016/22/A/NZ2/00184)

### XIII. BIBLIOGRAFIA

- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., et al. 2013. Signatures of mutational processes in human cancer. *Nature*, 500(7463): 415–421.
- Anderson, M. W., Reynolds, S. H., You, M., & Maronpot, R. M. 1992. Role of proto-oncogene activation in carcinogenesis. *Environmental Health Perspectives*, 98: 13–24.
- Aouacheria, A., Baghdiguan, S., Lamb, H. M., Huska, J. D., Pineda, F. J., et al. 2017. Connecting mitochondrial dynamics and life-or-death events via Bcl-2 family proteins. *Neurochemistry International*, 109: 141–161.
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., et al. 2004. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Research*, 32(Database issue): D115–D119.
- Aras, S. 2019. *Mitochondrial autoimmunity and MNRR1 in breast carcinogenesis*, 12.
- Aras, S., Bai, M., Lee, I., Springett, R., Hüttemann, M., et al. 2015. MNRR1 (formerly CHCHD2) is a bi-organellar regulator of mitochondrial metabolism. *Mitochondrion*, 20: 43–51.
- Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J., & Strasser, A. 2018. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death & Differentiation*, 25(1): 104–113.
- Avery, J. T., Zhang, R., & Boohaker, R. J. 2021. GLI1: A Therapeutic Target for Cancer. *Frontiers in Oncology*, 11: 673154.
- Balcere, A., Karls, R., Čēma, I., Rone Kupfere, M., Vīksna, L., et al. 2019. Treatment of Actinic Keratoses Facilitates Dermatoscopic Diagnosis of Early Basal Cell Carcinoma: A Case Report and Review. *Case Reports in Dermatology*, 11(1): 16–22.
- Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J.-M., & Gautheret, D. 2000. Patterns of Variant Polyadenylation Signal Usage in Human Genes. *Genome Research*, 10(7): 1001–1010.
- Bednarski, I. A., Ciężyńska, M., Wódz, K., Drózdź, I., Skibińska, M., et al. 2021. Hippo Signaling Pathway as a New Potential Target in Non-Melanoma Skin Cancers: A Narrative Review. *Life (Basel, Switzerland)*, 11(7): 680.
- Bell, R. J. A., Rube, H. T., Kreig, A., Mancini, A., Fouse, S. D., et al. 2015. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science (New York, N.Y.)*, 348(6238): 1036–1039.
- Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R. H. A., et al. 2005. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, 120(1): 21–24.
- Berman, S. B., Chen, Y., Qi, B., McCaffery, J. M., Rucker, E. B., et al. 2009. Bcl-xL increases mitochondrial fission, fusion, and biomass in neurons. *The Journal of Cell Biology*, 184(5): 707–719.
- Bödör, C., Grossmann, V., Popov, N., Okosun, J., O’Riain, C., et al. 2013. EZH2 mutations are frequent and represent an early event in follicular lymphoma. *Blood*, 122(18): 3165–3168.
- Bonilla, X., Parmentier, L., King, B., Bezrukov, F., Kaya, G., et al. 2016. Genomic analysis identifies new drivers and progression pathways in skin basal cell carcinoma. *Nature Genetics*, 48(4): 398–406.
- Bookstein, R., & Allred, D. C. 1993. Recessive oncogenes. *Cancer*, 71(3 Suppl): 1179–1186.



## BIBLIOGRAFIA

- Bouska, A., Zhang, W., Gong, Q., Iqbal, J., Scuto, A., et al. 2017. Combined copy number and mutation analysis identifies oncogenic pathways associated with transformation of follicular lymphoma. *Leukemia*, 31(1): 83–91.
- Boyle, G. M., Woods, S. L., Bonazzi, V. F., Stark, M. S., Hacker, E., et al. 2011. Melanoma cell invasiveness is regulated by miR-211 suppression of the BRN2 transcription factor. *Pigment Cell & Melanoma Research*, 24(3): 525–537.
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., et al. 2010. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences of the United States of America*, 107(43): 18545–18550.
- Cao, W., Ni, L., Li, P., Wang, Q.-X., Li, M.-M., et al. 2022. MiR-498 Targets UBE2T to Inhibit the Proliferation of Malignant Melanoma Cells. *Technology in Cancer Research & Treatment*, 21: 15330338221082432.
- Carballo, G. B., Honorato, J. R., de Lopes, G. P. F., & Spohr, T. C. L. de S. E. 2018. A highlight on Sonic hedgehog pathway. *Cell Communication and Signaling: CCS*, 16(1): 11.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., et al. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, 38(6): 626–635.
- Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., & Krainer, A. R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Research*, 31(13): 3568–3571.
- Castro-Giner, F., Ratcliffe, P., & Tomlinson, I. 2015. The mini-driver model of polygenic cancer evolution. *Nature Reviews. Cancer*, 15(11): 680–685.
- Chai, C., Wu, H., Wang, B., Eisenstat, D. D., & Leng, R. P. 2018. MicroRNA-498 promotes proliferation and migration by targeting the tumor suppressor PTEN in breast cancer cells. *Carcinogenesis*, 39(9): 1185–1196.
- Chandramouli, A., Shi, J., Feng, Y., Holubec, H., Shanas, R. M., et al. 2007. Haploinsufficiency of the *cdc21* gene contributes to skin cancer development in mice. *Carcinogenesis*, 28(9): 2028–2035.
- Chang, A. L. S., Tran, D. C., Cannon, J. G. D., Li, S., Jeng, M., et al. 2019. Pembrolizumab for advanced basal cell carcinoma: An investigator-initiated, proof-of-concept study. *Journal of the American Academy of Dermatology*, 80(2): 564–566.
- Chen, Y.-Y., Chang, W.-A., Lin, E.-S., Chen, Y.-J., & Kuo, P.-L. 2019. Expressions of HLA Class II Genes in Cutaneous Melanoma Were Associated with Clinical Outcome: Bioinformatics Approaches and Systematic Analysis of Public Microarray and RNA-Seq Datasets. *Diagnostics*, 9(2): 59.
- Chendrimada, T. P., Gregory, R. I., Kumaraswamy, E., Norman, J., Cooch, N., et al. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 436(7051): 740–744.
- Cornish, G. H., Tung, S. L., Marshall, D., Ley, S., & Seddon, B. P. 2012. Tissue specific deletion of inhibitor of kappa B kinase 2 with OX40-Cre reveals the unanticipated expression from the OX40 locus in skin epidermis. *PloS One*, 7(2): e32193.
- Crowdis, J., He, M. X., Reardon, B., & Van Allen, E. M. 2020. CoMut: Visualizing integrated molecular information with comutation plots. *Bioinformatics (Oxford, England)*, 36(15): 4348–4349.
- Crowson, A. N. 2006. Basal cell carcinoma: Biology, morphology and clinical implications. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc*, 19 Suppl 2: S127-147.

## BIBLIOGRAFIA

- Daily, K., Patel, V. R., Rigor, P., Xie, X., & Baldi, P. 2011. MotifMap: Integrative genome-wide maps of regulatory motif sites for model species. *BMC Bioinformatics*, 12(1): 495.
- Danial, N. N., Gramm, C. F., Scorrano, L., Zhang, C.-Y., Krauss, S., et al. 2003. BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature*, 424(6951): 952–956.
- Danial, N. N., Walensky, L. D., Zhang, C.-Y., Choi, C. S., Fisher, J. K., et al. 2008. Dual role of proapoptotic BAD in insulin secretion and beta cell survival. *Nature Medicine*, 14(2): 144–153.
- Darty, K., Denise, A., & Ponty, Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15): 1974–1975.
- Datta, S. R., Dudek, H., Tao, X., Masters, S., Fu, H., et al. 1997. Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell*, 91(2): 231–241.
- de Cid, R., Riveira-Munoz, E., Zeeuwen, P. L. J. M., Robarge, J., Liao, W., et al. 2009. Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nature Genetics*, 41(2): 211–215.
- Deng, W., & Roberts, S. G. E. 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & Development*, 19(20): 2418–2423.
- Denisova, E., Heidenreich, B., Nagore, E., Rachakonda, P. S., Hosen, I., et al. 2015. Frequent DPH3 promoter mutations in skin cancers. *Oncotarget*, 6(34): 35922–35930.
- Di Nardo, L., Pellegrini, C., Di Stefani, A., Ricci, F., Fossati, B., et al. 2021. Molecular alterations in basal cell carcinoma subtypes. *Scientific Reports*, 11(1): 13206.
- Díaz-Gay, M., Vila-Casadesús, M., Franch-Expósito, S., Hernández-Illán, E., Lozano, J. J., et al. 2018. Mutational Signatures in Cancer (MuSiCa): A web application to implement mutational signatures analysis in cancer samples. *BMC Bioinformatics*, 19(1): 224.
- Diederichs, S., Bartsch, L., Berkmann, J. C., Fröse, K., Heitmann, J., et al. 2016. The dark matter of the cancer genome: Aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Molecular Medicine*, 8(5): 442–457.
- Donaldson-Collier, M. C., Sungalee, S., Zufferey, M., Tavernari, D., Katanayeva, N., et al. 2019. EZH2 oncogenic mutations drive epigenetic, transcriptional, and structural changes within chromatin domains. *Nature Genetics*, 51(3): 517–528.
- Duan, X.-M., Liu, X.-N., Li, Y.-X., Cao, Y.-Q., Silayiding, A., et al. 2019. MicroRNA-498 promotes proliferation, migration, and invasion of prostate cancer cells and decreases radiation sensitivity by targeting PTEN. *The Kaohsiung Journal of Medical Sciences*, 35(11): 659–671.
- Elliott, K., & Larsson, E. 2021. Non-coding driver mutations in human cancer. *Nature Reviews Cancer*, 21(8): 500–509.
- Evans, D. G., Howard, E., Giblin, C., Clancy, T., Spencer, H., et al. 2010. Birth incidence and prevalence of tumor-prone syndromes: Estimates from a UK family genetic register service. *American Journal of Medical Genetics. Part A*, 152A(2): 327–332.
- Fadhil, R. S., Wei, M. Q., Nikolarakos, D., Good, D., & Nair, R. G. 2020. Salivary microRNA miR-let-7a-5p and miR-3928 could be used as potential diagnostic biomarkers for head and neck squamous cell carcinoma. *PloS One*, 15(3): e0221779.
- Fang, L., Seki, A., & Fang, G. 2009. SKAP associates with kinetochores and promotes the metaphase-to-anaphase transition. *Cell Cycle (Georgetown, Tex.)*, 8(17): 2819–2827.
- Freier, K., Flechtenmacher, C., Devens, F., Hartschuh, W., Hofele, C., et al. 2006. Recurrent NMYC copy number gain and high protein expression in basal cell carcinoma. *Oncology Reports*, 15(5): 1141–1145.

## BIBLIOGRAFIA

- Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., et al. 2015. A Uniform System For The Annotation Of Human microRNA Genes And The Evolution Of The Human microRNAome. *Annual Review of Genetics*, 49: 213–242.
- Fu, Y., Katsaros, D., Biglia, N., Wang, Z., Pagano, I., et al. 2021. Vitamin D receptor upregulates lncRNA TOPORS-AS1 which inhibits the Wnt/ $\beta$ -catenin pathway and associates with favorable prognosis of ovarian cancer. *Scientific Reports*, 11(1): 7484.
- Gailani, M. R., Stähle-Bäckdahl, M., Leffell, D. J., Glynn, M., Zaphiropoulos, P. G., et al. 1996. The role of the human homologue of *Drosophila* patched in sporadic basal cell carcinomas. *Nature Genetics*, 14(1): 78–81.
- Galka-Marciniak, P., Kanduła, Z., Tire, A., Wegorek, W., Gwozdz-Bak, K., et al. 2022. Mutations in the miR-142 gene are not common in myeloproliferative neoplasms. *Scientific Reports*, 12(1): 10924.
- Galka-Marciniak, P., Urbanek-Trzeciak, M. O., Nawrocka, P. M., Dutkiewicz, A., Giefing, M., et al. 2019. Somatic Mutations in miRNA Genes in Lung Cancer—Potential Functional Consequences of Non-Coding Sequence Variants. *Cancers*, 11(6): 793.
- Garcia-Bermudez, J., Baudrier, L., La, K., Zhu, X. G., Fidelin, J., et al. 2018. Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumors. *Nature Cell Biology*, 20(7): 775–781.
- García-Plata, D., Mozos, E., Carrasco, L., & Solana, R. 1993. HLA molecule expression in cutaneous squamous cell carcinomas: An immunopathological study and clinical-immunohistopathological correlations. *Histology and Histopathology*, 8(2): 219–226.
- Griewank, K. G., Murali, R., Schilling, B., Schimming, T., Möller, I., et al. 2013. TERT promoter mutations are frequent in cutaneous basal cell carcinoma and squamous cell carcinoma. *PloS One*, 8(11): e80354.
- Ha, M., & Kim, V. N. 2014. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, 15(8): 509–524.
- Hahn, S., Buratowski, S., Sharp, P. A., & Guarente, L. 1989. Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and nonconsensus DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 86(15): 5718–5722.
- Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., et al. 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5): 887–901.
- Hanahan, D. 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, 12(1): 31–46.
- Hanahan, D., & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell*, 100(1): 57–70.
- Hanahan, D., & Weinberg, R. A. 2011. Hallmarks of cancer: The next generation. *Cell*, 144(5): 646–674.
- Hase, K., Kimura, S., Takatsu, H., Ohmae, M., Kawano, S., et al. 2009. M-Sec promotes membrane nanotube formation by interacting with Ral and the exocyst complex. *Nature Cell Biology*, 11(12): 1427–1432.
- Hayward, N. K., Wilmott, J. S., Waddell, Nicola, Johansson, P. A., Field, M. A., et al. 2017. Whole-genome landscapes of major melanoma subtypes. *Nature*, 545(7653): 175–180.
- Heffelfinger, C., Ouyang, Z., Engberg, A., Leffell, D. J., Hanlon, A. M., et al. 2012. Correlation of Global MicroRNA Expression With Basal Cell Carcinoma Subtype. *G3 (Bethesda, Md.)*, 2(2): 279–286.
- Hezaveh, K., Kloetgen, A., Bernhart, S. H., Mahapatra, K. D., Lenze, D., et al. 2016. Alterations of microRNA and microRNA-regulated messenger RNA expression in

## BIBLIOGRAFIA

- germinal center B-cell lymphomas determined by integrative sequencing analysis. *Haematologica*, 101(11): 1380–1389.
- Hogan, D. J., To, T., Gran, L., Wong, D., & Lane, P. R. 1989. Risk factors for basal cell carcinoma. *International Journal of Dermatology*, 28(9): 591–594.
- Horn, S., Figl, A., Rachakonda, P. S., Fischer, C., Sucker, A., et al. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)*, 339(6122): 959–961.
- Hosseini, M., Dousset, L., Mahfouf, W., Serrano-Sanchez, M., Redonnet-Vernhet, I., et al. 2018. Energy Metabolism Rewiring Precedes UVB-Induced Primary Skin Tumor Formation. *Cell Reports*, 23(12): 3621–3634.
- Hosseini, M., Dousset, L., Michon, P., Mahfouf, W., Muzotte, E., et al. 2019. UVB-induced DHODH upregulation, which is driven by STAT3, is a promising target for chemoprevention and combination therapy of photocarcinogenesis. *Oncogenesis*, 8(10): 52.
- Houle, A. A., Gibling, H., Lamaze, F. C., Edgington, H. A., Soave, D., et al. 2018. Aberrant PRDM9 expression impacts the pan-cancer genomic landscape. *Genome Research*, 28(11): 1611–1620.
- Hua, L. A., Kagen, C. N., Carpenter, R. J., & Goltz, R. W. 1985. HLA and beta 2-microglobulin expression in basal and squamous cell carcinomas of the skin. *International Journal of Dermatology*, 24(10): 660–663.
- Huang, D., Wang, Y., Tang, J., & Luo, S. 2018. Molecular mechanisms of suppressor of fused in regulating the hedgehog signalling pathway. *Oncology Letters*, 15(5): 6077–6086.
- Huang, F. W., Hodis, E., Xu, M. J., Kryukov, G. V., Chin, L., et al. 2013. Highly recurrent TERT promoter mutations in human melanoma. *Science (New York, N.Y.)*, 339(6122): 957–959.
- Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., et al. 2020. miRTarBase 2020: Updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Research*, 48(D1): D148–D154.
- Ikeda, S., Goodman, A. M., Cohen, P. R., Jensen, T. J., Ellison, C. K., et al. 2016. Metastatic basal cell carcinoma with amplification of PD-L1: Exceptional response to anti-PD1 therapy. *Npj Genomic Medicine*, 1(1): 1–5.
- Islam, Md. T., Chen, F., & Chen, H. 2021. The oncogenic role of ubiquitin specific peptidase (USP8) and its signaling pathways targeting for cancer therapeutics. *Archives of Biochemistry and Biophysics*, 701: 108811.
- Isvoranu, G., Surcel, M., Huică, R.-I., Munteanu, A. N., Pîrvu, I. R., et al. 2019. Natural killer cell monitoring in cutaneous melanoma—New dynamic biomarker. *Oncology Letters*, 17(5): 4197–4206.
- Jaju, P. D., Nguyen, C. B., Mah, A. M., Atwood, S. X., Li, J., et al. 2015. Mutations in the Kinetochores Gene KNSTRN in Basal Cell Carcinoma. *The Journal of Investigative Dermatology*, 135(12): 3197–3200.
- Jayaraman, S. S., Rayhan, D. J., Hazany, S., & Kolodney, M. S. 2014. Mutational landscape of basal cell carcinomas by whole-exome sequencing. *The Journal of Investigative Dermatology*, 134(1): 213–220.
- Jia, L., Shi, Y., Wen, Y., Li, W., Feng, J., et al. 2018. The roles of TNFAIP2 in cancers and infectious diseases. *Journal of Cellular and Molecular Medicine*, 22(11): 5188–5195.
- Jia, L., Zhou, Z., Liang, H., Wu, J., Shi, P., et al. 2016. KLF5 promotes breast cancer proliferation, migration and invasion in part by upregulating the transcription of TNFAIP2. *Oncogene*, 35(16): 2040–2051.

## BIBLIOGRAFIA

- Jiang, H., Lei, R., Ding, S.-W., & Zhu, S. 2014. Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, 15: 182.
- Johnson, D. B., Estrada, M. V., Salgado, R., Sanchez, V., Doxie, D. B., et al. 2016. Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. *Nature Communications*, 7: 10582.
- Jung, H., Lee, K. S., & Choi, J. K. 2021. Comprehensive characterisation of intronic missplicing mutations in human cancers. *Oncogene*, 40(7): 1347–1361.
- Kaae, J., Thyssen, J. P., Johansen, J. D., Meldgaard, M., Linneberg, A., et al. 2013. Filaggrin gene mutations and risk of basal cell carcinoma. *British Journal of Dermatology*, 169(5): 1162–1164.
- Kalb, B., Marenholz, I., Jeanrenaud, A. C. S. N., Meixner, L., Arnau-Soler, A., et al. 2022. Filaggrin loss-of-function mutations are associated with persistence of egg and milk allergy. *The Journal of Allergy and Clinical Immunology*, 150(5): 1125–1134.
- Kanzi, A. M., San, J. E., Chimukangara, B., Wilkinson, E., Fish, M., et al. 2020. Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance. *Frontiers in Genetics*, 11: 544162.
- Karst, A. M., Dai, D. L., Martinka, M., & Li, G. 2005. PUMA expression is significantly reduced in human cutaneous melanomas. *Oncogene*, 24(6): 1111–1116.
- Kirkley, K. S., Walton, K. D., Duncan, C., & Tjalkens, R. B. 2017. Spontaneous Development of Cutaneous Squamous Cell Carcinoma in Mice with Cell-specific Deletion of Inhibitor of  $\kappa$ B Kinase 2. *Comparative Medicine*, 67(5): 407–415.
- Kodach, L. L., Jacobs, R. J., Heijmans, J., van Noesel, C. J. M., Langers, A. M. J., et al. 2010. The role of EZH2 and DNA methylation in the silencing of the tumour suppressor RUNX3 in colorectal cancer. *Carcinogenesis*, 31(9): 1567–1575.
- Kwanhian, W., Lenze, D., Alles, J., Motsch, N., Barth, S., et al. 2012. MicroRNA-142 is mutated in about 20% of diffuse large B-cell lymphoma. *Cancer Medicine*, 1(2): 141–155.
- Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., & Ebright, R. H. 1998. New core promoter element in RNA polymerase II-dependent transcription: Sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12(1): 34–44.
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457): 214–218.
- Lee, C. S., Bhaduri, A., Mah, A., Johnson, W. L., Ungewickell, A., et al. 2014. Recurrent point mutations in the kinetochore gene KNSTRN in cutaneous squamous cell carcinoma. *Nature Genetics*, 46(10): 1060–1062.
- Lefort, K., Mandinova, A., Ostano, P., Kolev, V., Calpini, V., et al. 2007. Notch1 is a p53 target gene involved in human keratinocyte tumor suppression through negative regulation of ROCK1/2 and MRCK $\alpha$  kinases. *Genes & Development*, 21(5): 562–577.
- Leman, G., Moosbrugger-Martinz, V., Blunder, S., Pavel, P., & Dubrac, S. 2019. 3D-Organotypic Cultures to Unravel Molecular and Cellular Abnormalities in Atopic Dermatitis and Ichthyosis Vulgaris. *Cells*, 8(5): 489.
- Leppek, K., Das, R., & Barna, M. 2018. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews. Molecular Cell Biology*, 19(3): 158–174.
- Lewandowska, M. A. 2013. The missing puzzle piece: Splicing mutations. *International Journal of Clinical and Experimental Pathology*, 6(12): 2675–2682.

## BIBLIOGRAFIA

- Liu, H., Xing, Y., Yang, S., & Tian, D. 2011. Remarkable difference of somatic mutation patterns between oncogenes and tumor suppressor genes. *Oncology Reports*, 26(6): 1539–1546.
- Liu, R., Shi, P., Wang, Z., Yuan, C., & Cui, H. 2021. Molecular Mechanisms of MYCN Dysregulation in Cancers. *Frontiers in Oncology*, 10. <https://www.frontiersin.org/articles/10.3389/fonc.2020.625332>.
- Liu, Y., Clegg, H. V., Leslie, P. L., Di, J., Tollini, L. A., et al. 2015. CHCHD2 inhibits apoptosis by interacting with Bcl-x L to regulate Bax activation. *Cell Death and Differentiation*, 22(6): 1035–1046.
- Lone, M. A., Hülsmeier, A. J., Saied, E. M., Karsai, G., Arenz, C., et al. 2020. Subunit composition of the mammalian serine-palmitoyltransferase defines the spectrum of straight and methyl-branched long-chain bases. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27): 15591–15598.
- Louder, R. K., He, Y., López-Blanco, J. R., Fang, J., Chacón, P., et al. 2016. Structure of promoter-bound TFIID and insight into human PIC assembly. *Nature*, 531(7596): 604–609.
- Machowska, M., Galka-Marciniak, P., & Kozłowski, P. 2022. Consequences of genetic variants in miRNA genes. *Computational and Structural Biotechnology Journal*, 20: 6443–6457.
- Mahmoud, F., Shields, B., Makhoul, I., Hutchins, L. F., Shalin, S. C., et al. 2016. Role of EZH2 histone methyltransferase in melanoma progression and metastasis. *Cancer Biology & Therapy*, 17(6): 579–591.
- Mak, K. K., Bi, Y., Wan, C., Chuang, P.-T., Clemens, T., et al. 2008. Hedgehog signaling in mature osteoblasts regulates bone formation and resorption by controlling PTHrP and RANKL expression. *Developmental Cell*, 14(5): 674–688.
- Malhotra, N., Leyva-Castillo, J. M., Jadhav, U., Barreiro, O., Kam, C., et al. 2018. ROR $\alpha$ -expressing T regulatory cells restrain allergic skin inflammation. *Science Immunology*, 3(21): eaao6923.
- Mancini, M., & Toker, A. 2009. NFAT Proteins: Emerging Roles in Cancer Progression. *Nature Reviews. Cancer*, 9(11): 810–820.
- Markey, A. C., Churchill, L. J., & MacDonald, D. M. 1990. Altered expression of major histocompatibility complex (MHC) antigens by epidermal tumours. *Journal of Cutaneous Pathology*, 17(2): 65–71.
- Maturo, M. G., Rachakonda, S., Heidenreich, B., Pellegrini, C., Srinivas, N., et al. 2020. Coding and noncoding somatic mutations in candidate genes in basal cell carcinoma. *Scientific Reports*, 10(1): 8005.
- McAleer, M. A., & Irvine, A. D. 2013. The multifunctional role of filaggrin in allergic skin disease. *The Journal of Allergy and Clinical Immunology*, 131(2): 280–291.
- McFarland, C. D., Yaglom, J. A., Wojtkowiak, J. W., Scott, J. G., Morse, D. L., et al. 2017. The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Research*, 77(18): 4763–4772.
- Meng, F., Henson, R., Wehbe-Janek, H., Ghoshal, K., Jacob, S. T., et al. 2007. MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer. *Gastroenterology*, 133(2): 647–658.
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., et al. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology*, 12(4): R41.
- Mignone, F., Gissi, C., Liuni, S., & Pesole, G. 2002. Untranslated regions of mRNAs. *Genome Biology*, 3(3): REVIEWS0004.

## BIBLIOGRAFIA

- Min, J., & Shay, J. W. 2016. TERT Promoter Mutations Enhance Telomerase Activation by Long-range Chromatin Interactions. *Cancer Discovery*, 6(11): 1212–1214.
- Ming, M., & He, Y.-Y. 2009. PTEN: New Insights into Its Regulation and Function in Skin Cancer. *The Journal of Investigative Dermatology*, 129(9): 2109–2112.
- Mizuno, T., Tokuoka, S., Kishikawa, M., Nakashima, E., Mabuchi, K., et al. 2006. Molecular basis of basal cell carcinogenesis in the atomic-bomb survivor population: P53 and PTCH gene alterations. *Carcinogenesis*, 27(11): 2286–2294.
- Morey, M., Fernández-Marmiesse, A., Castiñeiras, D., Fraga, J. M., Couce, M. L., et al. 2013. A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*, 110(1–2): 3–24.
- Morin, R. D., Johnson, N. A., Severson, T. M., Mungall, A. J., An, J., et al. 2010. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genetics*, 42(2): 181–185.
- Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A., & López-Bigas, N. 2016. OncodriveFML: A general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biology*, 17(1): 128.
- Mulcahy, E. Q. X., Zhang, Y., Colón, R. R., Cain, S. R., Gibert, M. K., et al. 2022. MicroRNA 3928 Suppresses Glioblastoma through Downregulation of Several Oncogenes and Upregulation of p53. *International Journal of Molecular Sciences*, 23(7): 3930.
- Myskowski, P. L., Pollack, M. S., Schorr, E., Dupont, B., & Safai, B. 1985. Human leukocyte antigen associations in basal cell carcinoma. *Journal of the American Academy of Dermatology*, 12(6): 997–1000.
- Nicolai, S., Mahen, R., Raschella, G., Marini, A., Pieraccioli, M., et al. 2020. ZNF281 is recruited on DNA breaks to facilitate DNA repair by non-homologous end joining. *Oncogene*, 39(4): 754–766.
- Nussinov, R., & Tsai, C.-J. 2015. “Latent drivers” expand the cancer mutational landscape. *Current Opinion in Structural Biology*, 32: 25–32.
- Ohba, S., Kawaguchi, H., Kugimiya, F., Ogasawara, T., Kawamura, N., et al. 2008. Patched1 haploinsufficiency increases adult bone mass and modulates Gli3 repressor activity. *Developmental Cell*, 14(5): 689–699.
- Otsuka, A., Dreier, J., Cheng, P. F., Nägeli, M., Lehmann, H., et al. 2015. Hedgehog pathway inhibitors promote adaptive immune responses in basal cell carcinoma. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 21(6): 1289–1297.
- Paggi, J. M., & Bejerano, G. 2018. A sequence-based, deep learning model accurately predicts RNA splicing branchpoints. *RNA*, 24(12): 1647–1658.
- Paiano, J., Wu, W., Yamada, S., Sciascia, N., Callen, E., et al. 2020. ATM and PRDM9 regulate SPO11-bound recombination intermediates during meiosis. *Nature Communications*, 11(1): 857.
- Pan, S., Dong, Q., Sun, L.-S., & Li, T.-J. 2010. Mechanisms of inactivation of PTCH1 gene in nevoid basal cell carcinoma syndrome: Modification of the two-hit hypothesis. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 16(2): 442–450.
- Pasparakis, M., Courtois, G., Hafner, M., Schmidt-Supprian, M., Nenci, A., et al. 2002. TNF-mediated inflammatory skin disease in mice with epidermis-specific deletion of IKK2. *Nature*, 417(6891): 861–866.
- Peitsch, M. C. 1997. Large scale protein modelling and model repository. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 5: 234–236.

## BIBLIOGRAFIA

- Pellegrini, C., Maturo, M. G., Di Nardo, L., Ciciarelli, V., Gutiérrez García-Rodrigo, C., et al. 2017. Understanding the Molecular Genetics of Basal Cell Carcinoma. *International Journal of Molecular Sciences*, 18(11): E2485.
- Peris, K., Fargnoli, M. C., Garbe, C., Kaufmann, R., Bastholt, L., et al. 2019. Diagnosis and treatment of basal cell carcinoma: European consensus-based interdisciplinary guidelines. *European Journal of Cancer (Oxford, England: 1990)*, 118: 10–34.
- Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., et al. 2001. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 276(1–2): 73–81.
- Petrova, R., & Joyner, A. L. 2014. Roles for Hedgehog signaling in adult organ homeostasis and repair. *Development (Cambridge, England)*, 141(18): 3445–3457.
- Peysers, N. D., Du, Y., Li, H., Lui, V., Xiao, X., et al. 2015. Loss-of-Function PTPRD Mutations Lead to Increased STAT3 Activation and Sensitivity to STAT3 Inhibition in Head and Neck Cancer. *PLOS ONE*, 10(8): e0135750.
- Piérard-Franchimont, C., Hermanns-Lê, T., Paquet, P., Herfs, M., Delvenne, P., et al. 2015. Hedgehog- and mTOR-targeted therapies for advanced basal cell carcinomas. *Future Oncology*, 11(22): 2997–3002.
- Piva de Freitas, P., Senna, C. G., Tabai, M., Chone, C. T., & Altemani, A. 2017. Metastatic Basal Cell Carcinoma: A Rare Manifestation of a Common Disease. *Case Reports in Medicine*, 2017: 8929745.
- Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K. J., Lukasiak, P., et al. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Research*, 40(14): e112.
- Puente, X. S., Beà, S., Valdés-Mas, R., Villamor, N., Gutiérrez-Abril, J., et al. 2015. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*, 526(7574): 519–524.
- Purdie, K. J., Lambert, S. R., Teh, M.-T., Chaplin, T., Molloy, G., et al. 2007. Allelic imbalances and microdeletions affecting the PTPRD gene in cutaneous squamous cell carcinomas detected using single nucleotide polymorphism microarray analysis. *Genes, Chromosomes & Cancer*, 46(7): 661–669.
- Qian, Y., Liang, X., Kong, P., Cheng, Y., Cui, H., et al. 2020. Elevated DHODH expression promotes cell proliferation via stabilizing  $\beta$ -catenin in esophageal squamous cell carcinoma. *Cell Death & Disease*, 11(10): 862.
- Rao, R. C., Chan, M. P., Andrews, C. A., & Kahana, A. 2016. EZH2, Proliferation Rate, and Aggressive Tumor Subtypes in Cutaneous Basal Cell Carcinoma. *JAMA Oncology*, 2(7): 962–963.
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. 2019. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1): D886–D894.
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. 2015. High-throughput sequencing technologies. *Molecular Cell*, 58(4): 586–597.
- Rheinbay, E., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., et al. 2020. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*, 578(7793): 102–111.
- Riobo, N. A., Lu, K., & Emerson, C. P. 2006. Hedgehog signal transduction: Signal integration and cross talk in development and cancer. *Cell Cycle (Georgetown, Tex.)*, 5(15): 1612–1615.
- Rosmarin, A. G., Resendes, K. K., Yang, Z., McMillan, J. N., & Fleming, S. L. 2004. GA-binding protein transcription factor: A review of GABP as an integrator of intracellular signaling and protein-protein interactions. *Blood Cells, Molecules & Diseases*, 32(1): 143–154.



## BIBLIOGRAFIA

- Roy, S. S., Madesh, M., Davies, E., Antonsson, B., Danial, N., et al. 2009. Bad targets the permeability transition pore independent of Bax or Bak to switch between Ca<sup>2+</sup>-dependent cell survival and death. *Molecular Cell*, 33(3): 377–388.
- Saini, H. K., Griffiths-Jones, S., & Enright, A. J. 2007. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(45): 17719–17724.
- Salamov, A. A., Nishikawa, T., & Swindells, M. B. 1998. Assessing protein coding region integrity in cDNA sequencing projects. *Bioinformatics*, 14(5): 384–390.
- Samarasinghe, V., Madan, V., & Lear, J. T. 2011. Focus on Basal cell carcinoma. *Journal of Skin Cancer*, 2011: 328615.
- Sanchez-Canteli, M., Hermida-Prado, F., Sordo-Bahamonde, C., Montoro-Jiménez, I., Pozo-Agundo, E., et al. 2020. Lectin-Like Transcript 1 (LLT1) Checkpoint: A Novel Independent Prognostic Factor in HPV-Negative Oropharyngeal Squamous Cell Carcinoma. *Biomedicines*, 8(12): 535.
- Sand, M., Bromba, A., Sand, D., Gambichler, T., Hessem, S., et al. 2019. Dicer Sequencing, Whole Genome Methylation Profiling, mRNA and smallRNA Sequencing Analysis in Basal Cell Carcinoma. *Cellular Physiology and Biochemistry: International Journal of Experimental Cellular Physiology, Biochemistry, and Pharmacology*, 53(5): 760–773.
- Sand, M., Skrygan, M., Sand, D., Georgas, D., Hahn, S. A., et al. 2012. Expression of microRNAs in basal cell carcinoma. *British Journal of Dermatology*, 167(4): 847–855.
- Sanger, F., Nicklen, S., & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12): 5463–5467.
- Sastry, K. S., Ibrahim, W. N., & Chouchane, A. I. 2020. Multiple signaling pathways converge on proapoptotic protein BAD to promote survival of melanocytes. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 34(11): 14602–14614.
- Sastry, K. S. R., Al-Muftah, M. A., Li, P., Al-Kowari, M. K., Wang, E., et al. 2014. Targeting proapoptotic protein BAD inhibits survival and self-renewal of cancer stem cells. *Cell Death and Differentiation*, 21(12): 1936.
- Schlesner, M., & Eils, R. 2015. Hypermutation takes the driver's seat. *Genome Medicine*, 7(1): 31.
- Schuster, S. L., & Hsieh, A. C. 2019. The untranslated regions of mRNAs in cancer. *Trends in Cancer*, 5(4): 245–262.
- Schwerk, J., & Savan, R. 2015. Translating the Untranslated Region. *The Journal of Immunology*, 195(7): 2963–2971.
- Scrivener, Y., Grosshans, E., & Cribier, B. 2002. Variations of basal cell carcinomas according to gender, age, location and histopathological subtype. *The British Journal of Dermatology*, 147(1): 41–47.
- Seo, S. Y., Chen, Y., Ivanovska, I., Ranger, A. M., Hong, S. J., et al. 2004. BAD Is a Pro-survival Factor Prior to Activation of Its Pro-apoptotic Function. *The Journal of Biological Chemistry*, 279(40): 42240–42249.
- Sexton, M., Jones, D. B., & Maloney, M. E. 1990. Histologic pattern analysis of basal cell carcinoma. Study of a series of 1039 consecutive neoplasms. *Journal of the American Academy of Dermatology*, 23(6 Pt 1): 1118–1126.
- Shen, Y., Song, Z., Lu, X., Ma, Z., Lu, C., et al. 2019. Fas signaling-mediated TH9 cell differentiation favors bowel inflammation and antitumor functions. *Nature Communications*, 10(1): 2924.

## BIBLIOGRAFIA

- Shoshan, E., Braeuer, R. R., Kamiya, T., Mobley, A. K., Huang, L., et al. 2016. NFAT1 Directly Regulates IL8 and MMP3 to Promote Melanoma Tumor Growth and Metastasis. *Cancer Research*, 76(11): 3145–3155.
- Siegal, J. A., Walker, J. L., Sachar, M., Lew, R. A., Chen, S. C., et al. 2016. Actinic keratoses progress more quickly to squamous cell carcinoma than basal cell carcinoma. *Dermatology Online Journal*, 22(9). <https://doi.org/10.5070/D3229032516>.
- Sigismund, S., Avanzato, D., & Lanzetti, L. 2018. Emerging functions of the EGFR in cancer. *Molecular Oncology*, 12(1): 3–20.
- Simmons, J. L., Pierce, C. J., Al-Ejeh, F., & Boyle, G. M. 2017. MITF and BRN2 contribute to metastatic growth after dissemination of melanoma. *Scientific Reports*, 7(1): 10909.
- Sonkoly, E., Lovén, J., Xu, N., Meisgen, F., Wei, T., et al. 2012. MicroRNA-203 functions as a tumor suppressor in basal cell carcinoma. *Oncogenesis*, 1: e3.
- Spiegel, J., Cuesta, S. M., Adhikari, S., Hänsel-Hertsch, R., Tannahill, D., et al. 2021. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biology*, 22(1): 117.
- Stanisz, H., Saul, S., Müller, C. S. L., Kappl, R., Niemeyer, B. A., et al. 2014. Inverse regulation of melanoma growth and migration by Orai1/STIM2-dependent calcium entry. *Pigment Cell & Melanoma Research*, 27(3): 442–453.
- Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., et al. 2022. Signatures of copy number alterations in human cancer. *Nature*, 606(7916): 984–991.
- Stickles, X. B., Marchion, D. C., Bicaku, E., Al Sawah, E., Abbasi, F., et al. 2015. BAD-mediated apoptotic pathway is associated with human cancer development. *International Journal of Molecular Medicine*, 35(4): 1081–1087.
- Stratis, A., Pasparakis, M., Markur, D., Knaup, R., Pofahl, R., et al. 2006. Localized inflammatory skin disease following inducible ablation of I kappa B kinase 2 in murine epidermis. *The Journal of Investigative Dermatology*, 126(3): 614–620.
- Streilein, J. W. 1991. Immunogenetic factors in skin cancer. *The New England Journal of Medicine*, 325(12): 884–887.
- Su, S., Zhao, Q., He, C., Huang, D., Liu, J., et al. 2015. MiR-142-5p and miR-130a-3p are regulated by IL-4 and IL-13 and control profibrogenic macrophage program. *Nature Communications*, 6: 8523.
- Su, X., Malouf, G. G., Chen, Y., Zhang, J., Yao, H., et al. 2014. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget*, 5(20): 9864–9876.
- Sun, X., & Zhang, H. 2018. MiR-451 elevation relieves inflammatory pain by suppressing microglial activation-evoked inflammatory response via targeting TLR4. *Cell and Tissue Research*, 374(3): 487–495.
- Sun, Y., Varambally, S., Maher, C. A., Cao, Q., Chockley, P., et al. 2011. Targeting of microRNA-142-3p in dendritic cells regulates endotoxin-induced mortality. *Blood*, 117(23): 6172–6183.
- Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. 2016. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Computational Biology*, 12(4): e1004873.
- Thélu, J., Rossio, P., & Favier, B. 2002. Notch signalling is linked to epidermal cell differentiation level in basal cell carcinoma, psoriasis and wound healing. *BMC Dermatology*, 2: 7.
- Thol, F., Scherr, M., Kirchner, A., Shahswar, R., Battmer, K., et al. 2015. Clinical and functional implications of microRNA mutations in a cohort of 935 patients with

## BIBLIOGRAFIA

- myelodysplastic syndromes and acute myeloid leukemia. *Haematologica*, 100(4): e122-124.
- Tolba, F. M., Foda, M. E., kamal, H. M., & Elshabrawy, D. A. 2013. Expression of CD133 in acute leukemia. *Medical Oncology*, 30(2): 527.
- Tomková, H., Fujimoto, W., & Arata, J. 1998. Expression of the bcl-2 homologue bax in normal human skin, psoriasis vulgaris and non-melanoma skin cancers. *European Journal of Dermatology: EJD*, 8(4): 256–260.
- Trissal, M. C., Wong, T. N., Yao, J.-C., Ramaswamy, R., Kuo, I., et al. 2018. MIR142 Loss-of-Function Mutations Derepress ASH1L to Increase HOXA Gene Expression and Promote Leukemogenesis. *Cancer Research*, 78(13): 3510–3521.
- UniProt Consortium. 2019. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1): D506–D515.
- Urbanek-Trzeciak, M. O., Galka-Marciniak, P., Nawrocka, P. M., Kowal, E., Szewc, S., et al. 2020. Pan-cancer analysis of somatic mutations in miRNA genes. *EBioMedicine*, 61: 103051.
- Urbanek-Trzeciak, M. O., Kozłowski, P., & Galka-Marciniak, P. 2022. miRMut: Annotation of mutations in miRNA genes from human whole-exome or whole-genome sequencing. *STAR Protocols*, 3(1): 101023.
- Vander Heiden, M. G., Li, X. X., Gottleib, E., Hill, R. B., Thompson, C. B., et al. 2001. Bcl-xL promotes the open configuration of the voltage-dependent anion channel and metabolite passage through the outer mitochondrial membrane. *The Journal of Biological Chemistry*, 276(22): 19414–19419.
- Veeriah, S., Brennan, C., Meng, S., Singh, B., Fagin, J. A., et al. 2009. The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 106(23): 9435–9440.
- Walia, V., Prickett, T. D., Kim, J.-S., Gartner, J. J., Lin, J. C., et al. 2014. Mutational and functional analysis of the tumor-suppressor PTPRD in human melanoma. *Human Mutation*, 35(11): 1301–1310.
- Wan, J., Ling, X., Peng, B., & Ding, G. 2018. MiR-142-5p regulates CD4+ T cells in human non-small cell lung cancer through PD-L1 expression via the PTEN pathway. *Oncology Reports*, 40(1): 272–282.
- Wang, Y., Ma, M., Xiao, X., & Wang, Z. 2012. Intronic Splicing Enhancers, Cognate Splicing Factors and Context Dependent Regulation Rules. *Nature Structural & Molecular Biology*, 19(10): 1044–1052.
- Weaver, T. M., Liu, J., Connelly, K. E., Coble, C., Varzavand, K., et al. 2019. The EZH2 SANT1 domain is a histone reader providing sensitivity to the modification state of the H4 tail. *Scientific Reports*, 9(1): 987.
- Wei, L., Li, X., Wang, L., Song, Y., & Dong, H. 2021. Comprehensive Analysis of RNA Expression Profile Identifies Hub miRNA-circRNA Interaction Networks in the Hypoxic Ischemic Encephalopathy. *Computational and Mathematical Methods in Medicine*, 2021: 6015473.
- Welcker, M., Orian, A., Jin, J., Grim, J. A., Harper, J. W., et al. 2004. The Fbw7 tumor suppressor regulates glycogen synthase kinase 3 phosphorylation-dependent c-Myc protein degradation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(24): 9085–9090.
- White, R. M., Cech, J., Ratanasirintrao, S., Lin, C. Y., Rahl, P. B., et al. 2011. DHODH modulates transcriptional elongation in the neural crest and melanoma. *Nature*, 471(7339): 518–522.

## BIBLIOGRAFIA

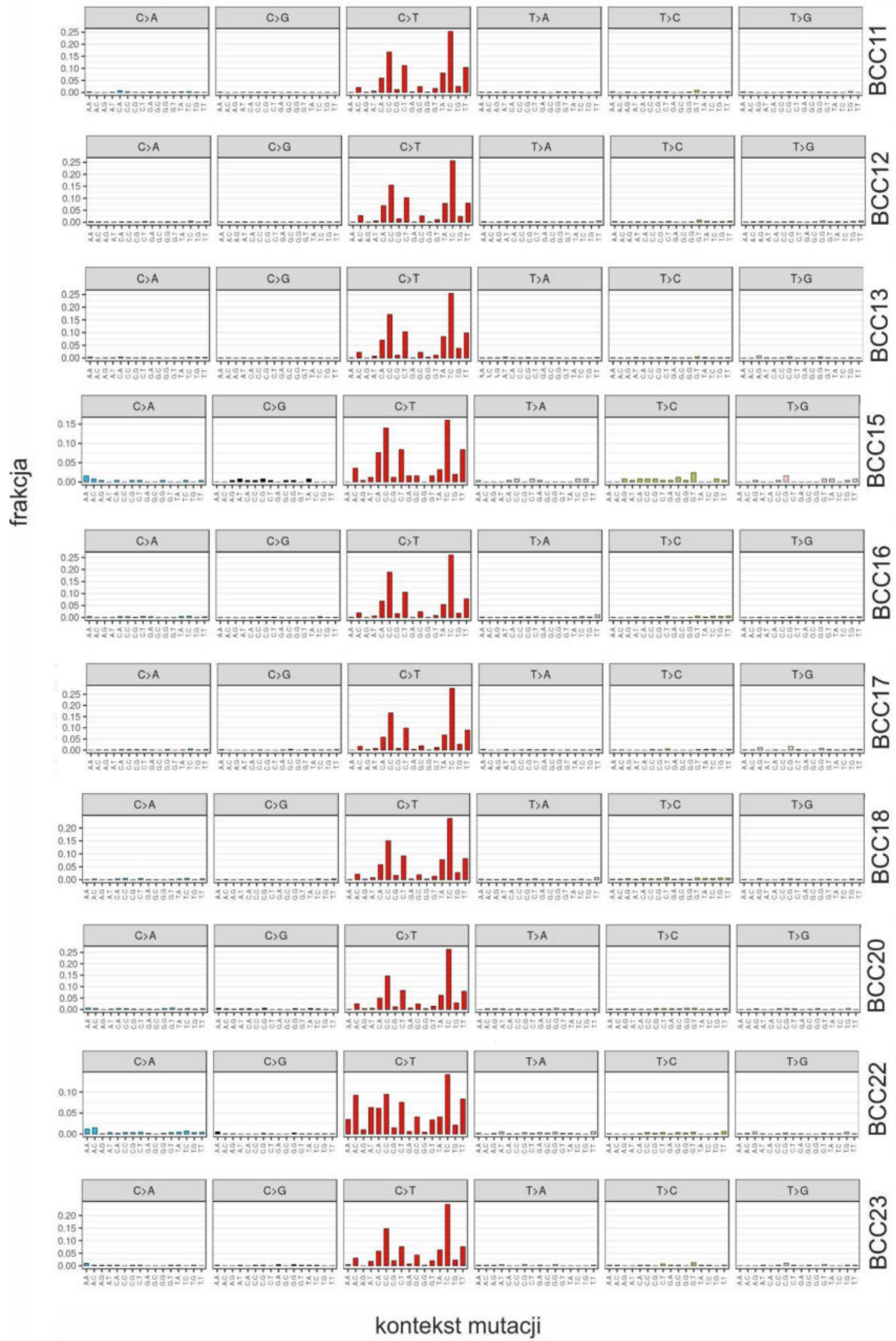
- Wu, S., Liang, Y., Zang, Q., Xing, Z., Yin, P., et al. 2022. FLG Is a Potential Biomarker of Prognosis and Immunotherapy in Skin Cutaneous Melanoma. *Applied Bionics and Biomechanics*, 2022: e5160748.
- Wu, W., Wu, L., Zhu, M., Wang, Z., Wu, M., et al. 2018. MiRNA Mediated Noise Making of 3'UTR Mutations in Cancer. *Genes*, 9(11): 545.
- Wu, Z., Hansmann, B., Meyer-Hoffert, U., Gläser, R., & Schröder, J.-M. 2009. Molecular identification and expression analysis of filaggrin-2, a member of the S100 fused-type protein family. *PLoS One*, 4(4): e5227.
- Xia, E., Kanematsu, S., Suenaga, Y., Elzawahry, A., Kondo, H., et al. 2018. MicroRNA induction by copy number gain is associated with poor outcome in squamous cell carcinoma of the lung. *Scientific Reports*, 8(1): 15363.
- Xie, Q., Wang, H., Heilman, E. R., Walsh, M. G., Haseeb, M. A., et al. 2014. Increased expression of enhancer of Zeste Homolog 2 (EZH2) differentiates squamous cell carcinoma from normal skin and actinic keratosis. *European Journal of Dermatology*, 24(1): 41–45.
- Xie, Y., & Wang, B. 2017. Downregulation of TNFAIP2 suppresses proliferation and metastasis in esophageal squamous cell carcinoma through activation of the Wnt/ $\beta$ -catenin signaling pathway. *Oncology Reports*, 37(5): 2920–2928.
- Xu, H., Liu, X., & Zhao, J. 2014. Down-Regulation of miR-3928 Promoted Osteosarcoma Growth. *Cellular Physiology and Biochemistry*, 33(5): 1547–1556.
- Xu, Y., Luo, H., Hu, Q., & Zhu, H. 2021. Identification of Potential Driver Genes Based on Multi-Genomic Data in Cervical Cancer. *Frontiers in Genetics*, 12: 598304.
- Yang, E., Zha, J., Jockel, J., Boise, L. H., Thompson, C. B., et al. 1995. Bad, a heterodimeric partner for Bcl-XL and Bcl-2, displaces Bax and promotes cell death. *Cell*, 80(2): 285–291.
- Yu, H.-S., Liao, W.-T., & Chai, C.-Y. 2006. Arsenic carcinogenesis in the skin. *Journal of Biomedical Science*, 13(5): 657–666.
- Zeng, Y., Wagner, E. J., & Cullen, B. R. 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Molecular Cell*, 9(6): 1327–1333.
- Zhang, M. Q. 1998. Statistical features of human exons and their flanking regions. *Human Molecular Genetics*, 7(5): 919–932.
- Zhang, P., An, Z., Sun, C., Xu, Y., & Zhang, Z. 2022. FLG Gene Mutation Up-regulates the Abnormal Tumor Immune Response and Promotes the Progression of Prostate Cancer. *Current Pharmaceutical Biotechnology*, 23(14): 1658–1670.
- Zhang, X., Xu, X., Ge, G., Zang, X., Shao, M., et al. 2019. MiR-498 inhibits the growth and metastasis of liver cancer by targeting ZEB2. *Oncology Reports*, 41(3): 1638–1648.
- Zhang, X., Zhang, Z., Cheng, J., Li, M., Wang, W., et al. 2012. Transcription Factor NFAT1 Activates the mdm2 Oncogene Independent of p53. *The Journal of Biological Chemistry*, 287(36): 30468–30476.
- Zhao, G., Wei, Z., & Guo, Y. 2020. MicroRNA-107 is a novel tumor suppressor targeting POU3F2 in melanoma. *Biological Research*, 53(1): 11.
- Zhou, J., Chen, J., & Yu, H. 2018. Targeting sphingosine kinase 2 by ABC294640 inhibits human skin squamous cell carcinoma cell growth. *Biochemical and Biophysical Research Communications*, 497(2): 535–542.
- Zhou, X., Edmonson, M. N., Wilkinson, M. R., Patel, A., Wu, G., et al. 2016. Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nature Genetics*, 48(1): 4–6.

## BIBLIOGRAFIA

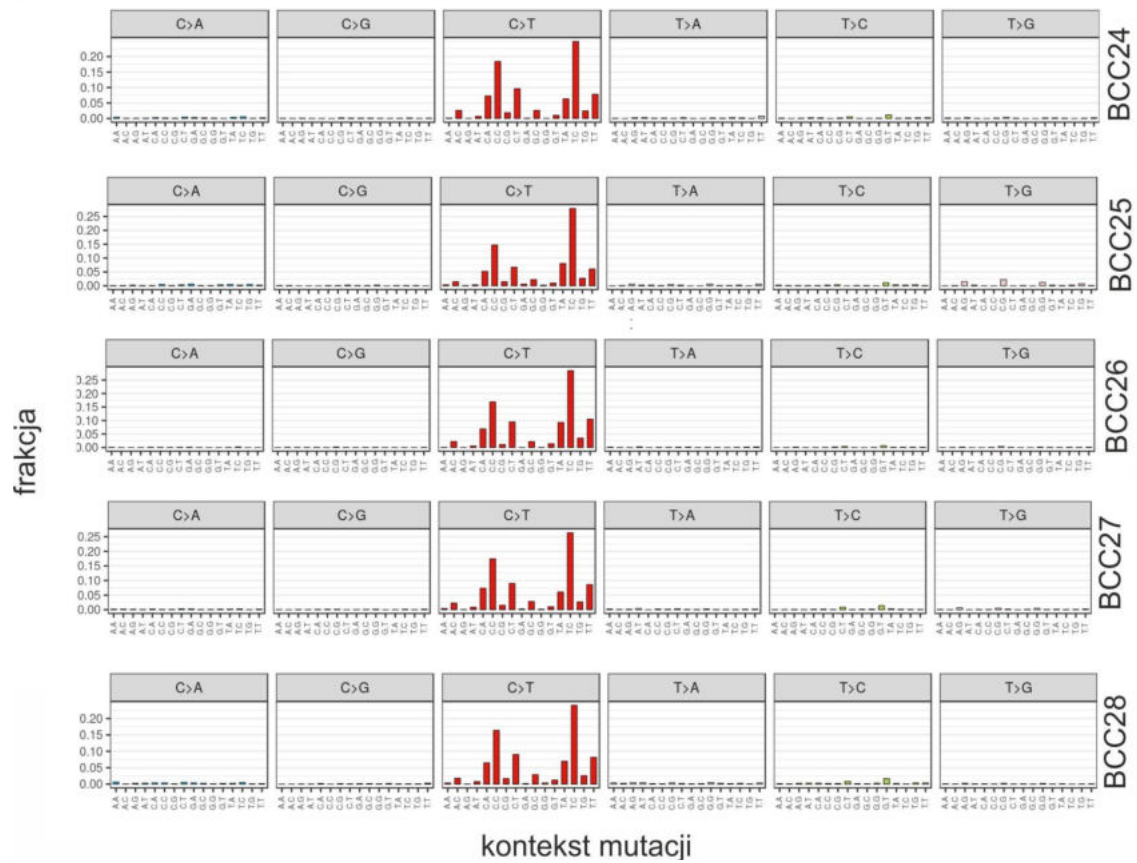
- Zhu, M., Gao, S., & Ren, X. 2022. Role of miR-498 Combined with CREB1 in Apoptosis and Invasion of Hepatoma Cell Line. *Computational and Mathematical Methods in Medicine*, 2022: e9621764.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13): 3406–3415.



# ZAŁĄCZNIKI



# ZAŁĄCZNIKI





## ZAŁĄCZNIKI

Załącznik 2. Geny często mutowane w regionie CDS.

Nazwa genu	Liczba mutacji	Liczba zmutowanych próbek	Obecność w bazie CGC
PTCH1	24	20	Tak
PCDHA1	78	17	-
USH2A	27	16	-
GPR98	34	16	-
SYNE1	29	16	-
PKHD1L1	22	15	-
FSIP2	30	15	-
MUC5B	22	15	-
APOB	28	14	-
XIRP2	19	14	-
PCDHGA1	36	14	-
ZNF469	29	14	-
PTPRD	19	13	Tak
PKHD1	16	13	-
TP53	16	13	Tak
ANK3	18	12	-
IGFN1	19	12	-
RP1	19	12	-
PCDH15	19	12	-
EYS	25	12	-
GRIN2A	17	11	Tak
FAT4	16	11	Tak
DIP2A	13	11	-
PKD1	11	11	-
DCHS2	17	11	-
MUC17	12	11	-
DNAH7	17	11	-
RIMS1	13	11	-
MYO15A	16	11	-
ABCA13	17	11	-
RYR1	15	10	-

MLL2	19	10	-
CSMD2	14	10	-
BAI3	12	10	-
ACAN	17	10	-
PREX1	10	10	-
DSCAM	16	10	-
DNAH9	17	10	-
FBN1	12	10	-
NRG3	12	10	-
FLG	15	10	-
NEFM	11	10	-
MAGEC1	20	10	-
RELN	16	10	-
SI	13	10	-
CACNA1H	15	10	-
PRUNE2	12	10	-
FAM135B	15	10	Tak
ZFHX4	13	9	-
SIPA1L2	10	9	-
LRRC7	12	9	-
MYOCD	13	9	-
HDAC9	14	9	-
FLG2	9	9	-
AHNAK2	14	9	-
HEATR7B	14	9	-
FREM2	13	9	-
BNC2	10	9	-
TAF1L	13	9	-
KDR	10	9	Tak
MYH13	10	9	-
SLIT2	12	9	-
THSD7A	13	9	-
VPS13B	12	9	-
SCN9A	12	9	-
STAB2	11	9	-
FREM1	14	9	-
WDFY4	9	9	-
KANK4	9	9	-
PXDNL	12	9	-

## ZAŁĄCZNIKI

LPA	13	9	-
POM121L	9	9	-
KIAA1211	12	9	-
SVEP1	13	9	-
CMYA5	37	9	-
NOTCH1	10	8	Tak
CKAP5	9	8	-
SLC8A3	8	8	-
MYCN	8	8	Tak
AHNAK	12	8	-
TDRD6	11	8	-
FMN2	11	8	-
ZFYVE26	9	8	-
ARHGAP3	9	8	-
NEB	14	8	-
RREB1	14	8	-
GRM8	13	8	-
BIRC6	9	8	Tak
KIAA1109	13	8	-
GREB1	11	8	-
MED12L	8	8	-
DLC1	9	8	-
VWF	9	8	-
WDR87	11	8	-
ZNF804B	19	8	-
NLRP12	9	8	-
MYO7B	12	8	-
TEP1	11	8	-
KLHL4	8	8	-
CNGB1	9	8	-
PLEC	8	8	-
NAV3	9	8	-
GPR112	18	8	-
MYEF2	8	8	-
TTBK2	9	8	-
CASS4	9	8	-
RP1L1	12	8	-
ZFH3	10	8	Tak
NES	13	8	-

KIAA1210	9	8	-
ALPK2	10	8	-
LRRK2	9	8	-
SRRM2	9	8	-
DIDO1	8	8	-
SPHKAP	10	8	-
KNDC1	11	8	-
ZNF831	11	8	-
SPTA1	16	8	-
NACA	10	8	Tak
NOTCH2	8	7	Tak
SLC4A9	8	7	-
FAT2	7	7	-
TMEM13	9	7	-
CRB1	9	7	-
CILP2	10	7	-
UNC80	8	7	-
PLCL1	7	7	-
HCN1	8	7	-
DOCK10	9	7	-
MBD5	9	7	-
SHANK1	10	7	-
ATRNL1	8	7	-
FSTL5	9	7	-
RNU6-28	7	7	-
FBN2	12	7	-
NPAP1	9	7	-
ODZ1	10	7	-
CELSR3	10	7	-
CACNA1A	10	7	-
SPAG17	8	7	-
MYO18B	11	7	-
SEZ6L	8	7	-
EPHA5	7	7	-
PDE10A	7	7	-
LRP1	8	7	-
LAMA2	7	7	-
ABCA6	7	7	-
MYO16	8	7	-

## ZAŁĄCZNIKI

ZNF521	7	7	Tak
CELSR2	9	7	-
PSME4	7	7	-
UBR4	7	7	-
CUBN	8	7	-
PCDHB6	9	7	-
LPPR4	7	7	-
ST18	7	7	-
CXorf22	10	7	-
NALCN	9	7	-
ANK2	8	7	-
TRIOBP	12	7	-
MYOM3	11	7	-
ABCA7	8	7	-
SPTBN2	8	7	-
ASXL3	7	7	-
CUX2	7	7	-
STARD9	8	7	-
MDN1	11	7	-
ZNF541	8	7	-
ACACB	7	7	-
NWD1	7	7	-
FRY	8	7	-
NLRC5	10	7	-
MYO1H	8	7	-
MACF1	7	7	-
PCDHB16	9	7	-
C7	10	7	-
MRC1	7	7	-
EPG5	7	7	-
VWDE	7	7	-
CFH	8	7	-
HIVEP3	7	7	-
SH2D7	7	7	-
MEFV	9	7	-
PRRC2C	9	7	-
FNDC1	8	7	-
PLCE1	10	7	-
FCGBP	11	7	-

WDR33	8	7	-
PPP1R3A	8	7	-
ALMS1	17	7	-
AK5	8	6	-
ADCY8	7	6	-
ZNF536	7	6	-
ZIM2	9	6	-
GCN1L1	6	6	-
PTPRF	6	6	-
FAM59A	6	6	-
TG	8	6	-
CHRNA9	6	6	-
SACS	8	6	-
IL16	8	6	-
ADAMTS	11	6	-
SLC4A4	8	6	-
CDKL5	7	6	-
SORL1	6	6	-
CYP17A1	6	6	-
HMCN1	7	6	-
ADCY1	6	6	-
RNU6-71	13	6	-
ZMAT4	6	6	-
NELL2	6	6	-
HUWE1	6	6	-
SPTB	7	6	-
MDGA2	7	6	-
AATK	7	6	-
ZEB1	6	6	Tak
BRD4	6	6	Tak
ODZ4	6	6	-
MLL	7	6	-
ROBO3	7	6	-
URGCP-	6	6	-
NOS1	6	6	-
ANKRD30	8	6	-
ZNF208	6	6	-
NPR1	7	6	-
TPTE	6	6	-

## ZAŁĄCZNIKI

ROS1	9	6	Tak
CPS1	6	6	-
LRRC3B	6	6	-
PRDM16	7	6	Tak
DCC	8	6	Tak
UNC13C	8	6	-
OPCML	6	6	-
SPEG	6	6	-
EPHA10	11	6	-
CNTNAP2	6	6	Tak
SLC8A1	8	6	-
MYO9A	6	6	-
SPEN	7	6	Tak
EPHA3	8	6	Tak
FLNC	7	6	-
MAP2	6	6	-
EPB41L3	6	6	-
NYAP2	7	6	-
LRBA	6	6	-
ZNF609	6	6	-
SYNE2	9	6	-
ABCA4	7	6	-
DMXL2	6	6	-
ULK1	6	6	-
EPHA6	8	6	-
FAM75D	6	6	-
TEX11	6	6	-
SLCO1C1	8	6	-
GRIN3A	6	6	-
PTPRZ1	8	6	-
TNFAIP2	6	6	-
HCFC1	7	6	-
HECTD1	7	6	-
PCDHB8	9	6	-
TEX15	9	6	-
FRMPD4	9	6	-
TTC21A	10	6	-
PLIN4	8	6	-
NRAP	13	6	-

ZMIZ1	6	6	-
IGSF10	8	6	-
SIGLEC1	7	6	-
HEATR8-	7	6	-
PCDH9	6	6	-
SLFN13	6	6	-
PKD1L1	10	6	-
GRIN2B	6	6	-
CHD4	6	6	Tak
PRIC285	8	6	-
CHD5	9	6	-
TTLL6	6	6	-
CD163L1	9	6	-
TIE1	6	6	-
AKD1	7	6	-
LEPR	8	6	-
FAM186A	10	6	-
RASGRP3	8	6	-
ZNF804A	7	6	-
FREM3	6	6	-
TRANK1	7	6	-
KIF6	7	6	-
CFHR5	6	6	-
ATP10B	8	6	-
ITGAM	7	6	-
CC2D1A	7	6	-
CEP112	6	6	-
TRIO	6	6	-
KCNH8	6	6	-
TMPRSS1	7	6	-
FAT1	12	6	Tak
F5	9	6	-
REV3L	7	6	-
AXDND1	8	6	-
WDR7	6	6	-
INADL	6	6	-
CPAMD8	12	6	-
FCRL5	7	6	-
KLC1	7	6	-

## ZAŁĄCZNIKI

ZEB2	6	6	-
CDYL2	6	6	-
CDH9	7	6	-
AKAP13	7	6	-
VGLL3	6	6	-
DOCK5	6	6	-
MYBPC3	6	6	-
CYP4A11	6	6	-
PPP1R9A	9	6	-
PKDREJ	7	6	-
POLQ	9	6	Tak
FCRL4	6	6	Tak
GPRIN1	6	6	-
MLL4	7	6	-
SLC22A2	7	6	-
AHRR	6	6	-
GNAS	9	6	Tak
ANKRD24	6	6	-
PIGQ	8	6	-
KDM4C	8	6	-
BANK1	9	6	-
C14orf39	7	6	-
FOXR2	6	6	-
GPR158	7	6	-
TGFB111	5	5	-
CARD6	5	5	-
CTNND2	6	5	Tak
PCDH7	5	5	-
FLNA	7	5	Tak
NOS3	5	5	-
FSHR	6	5	-
CNGA3	6	5	-
A1CF	5	5	Tak
FLNB	8	5	-
DNAH3	8	5	-
ADCY4	5	5	-
PYGB	6	5	-
NELL1	5	5	-
SYNPO	10	5	-

SYT3	5	5	-
SLC36A2	5	5	-
CREBBP	7	5	Tak
CTNNA3	6	5	-
SETD1A	5	5	-
DDX60L	8	5	-
CNBD1	5	5	Tak
MYH7	7	5	-
NCAN	7	5	-
CREB3L4	7	5	-
DNMT3B	5	5	-
CELSR1	6	5	-
MAEL	5	5	-
TMPRSS1	5	5	-
NCKAP5L	5	5	-
LATS1	6	5	Tak
ABCB5	5	5	-
MYH3	6	5	-
C16orf96	7	5	-
OTOL1	5	5	-
ADAMTS	5	5	-
TIMELESS	5	5	-
PDE6C	5	5	-
PLXNC1	7	5	-
FZD9	5	5	-
ZNF318	6	5	-
SHROOM	5	5	-
MATN4	5	5	-
KANK1	9	5	-
MYO3B	5	5	-
BRWD3	8	5	-
XKR4	5	5	-
SDK2	6	5	-
TNIK	5	5	-
TLN1	6	5	-
FAM5B	6	5	-
TANC1	5	5	-
TECTA	5	5	-
PBRM1	5	5	Tak

## ZAŁĄCZNIKI

CHD7	8	5	-
RBP3	6	5	-
NEBL	8	5	-
BSN	6	5	-
FAM194A	6	5	-
PCDH1	10	5	-
TRPM3	7	5	-
GRIA2	5	5	-
DYNC1H1	5	5	-
FAM171A	6	5	-
DNAJC13	5	5	-
BCL9L	6	5	Tak
KIF1A	5	5	-
DLG2	5	5	-
DCAF4	5	5	-
KIAA0240	5	5	-
BAI2	5	5	-
DCAF4L2	5	5	-
BRD1	5	5	-
RHOT2	8	5	-
GRIK2	5	5	-
AR	5	5	Tak
CADPS	6	5	-
CACNA1F	6	5	-
PRDM14	5	5	-
FAM135A	7	5	-
GRIK4	5	5	-
ABCC11	8	5	-
ATP13A1	5	5	-
PLXNA3	6	5	-
CADPS2	6	5	-
FAM155A	5	5	-
RBMXL2	5	5	-
AKAP6	6	5	-
ADAMTS	6	5	-
IRS2	6	5	-
C6	7	5	-
ARFGEF2	5	5	-
SAMD9L	5	5	-

EPHA8	7	5	-
GIN1	5	5	-
UBASH3A	5	5	-
IVL	5	5	-
ANKRD11	5	5	-
TACC2	6	5	-
DIRC2	5	5	-
CNKSR2	6	5	-
CEACAM	5	5	-
UGT2B11	8	5	-
RGS11	6	5	-
SPATA20	5	5	-
CDH23	5	5	-
ITGAX	5	5	-
UMODL1	10	5	-
USP26	6	5	-
DAB2	5	5	-
PCDHB4	6	5	-
ZSCAN10	5	5	-
ARID1A	5	5	Tak
RBM19	5	5	-
FBXO40	5	5	-
PRKCB	5	5	Tak
PDE3A	5	5	-
ARHGEF2	6	5	-
NLRP5	9	5	-
DOT1L	6	5	-
MYH8	5	5	-
SYT16	5	5	-
CCDC155	5	5	-
NBAS	5	5	-
ALOX15B	6	5	-
SIM1	5	5	-
F8	6	5	-
MMRN2	7	5	-
RIMS2	6	5	-
L1CAM	8	5	-
DNAH6	6	5	-
ROBO1	6	5	-

## ZAŁĄCZNIKI

ADAMTSL	6	5	-
SLFN5	6	5	-
ARID1B	6	5	Tak
ROR1	6	5	-
INTS5	6	5	-
MYO3A	5	5	-
EIF2AK3	5	5	-
MYLK3	5	5	-
PTPN21	6	5	-
TLN2	6	5	-
GPR123	6	5	-
NBEA	5	5	Tak
ATP2C2	6	5	-
KALRN	6	5	-
FLT4	6	5	Tak
NCOA3	5	5	-
SMCHD1	5	5	-
MEIS2	6	5	-
BCORL1	6	5	Tak
GREB1L	5	5	-
TGM6	5	5	-
C6orf118	6	5	-
GPR128	5	5	-
SLC4A10	6	5	-
FAM5C	5	5	-
MSH4	5	5	-
STARD13	5	5	-
TRIP11	6	5	Tak
EYA2	5	5	-
MEGF8	7	5	-
RTTN	5	5	-
KCNT2	6	5	-
IHH	5	5	-
ARHGEF1	6	5	Tak
NOTCH3	6	5	-
RADIL	7	5	-
ADAMTS	6	5	-
ATP13A5	6	5	-
FAM71B	5	5	-

CDON	8	5	-
PIK3CG	6	5	-
ADGB	5	5	-
BCHE	6	5	-
VPS13A	5	5	-
SHROOM	7	5	-
ATP13A4	5	5	-
ZC3HC1	6	5	-
FSCB	7	5	-
TRPM5	5	5	-
DCAF8L1	5	5	-
MANBA	5	5	-
WWC2	5	5	-
VWA3B	8	5	-
ITGAD	5	5	-
C1orf173	5	5	-
CACNA1E	5	5	-
LCP2	5	5	-
SNED1	5	5	-
EML5	5	5	-
IBSP	6	5	-
PTPRN	7	5	-
MYH4	10	5	-
CNTN3	6	5	-
NYNRIN	6	5	-
METTL11	6	5	-
FAM47A	6	5	-
PCDHB5	6	5	-
TCHH	5	5	-
ABCA9	5	5	-
TMEM23	5	5	-
ABCC8	5	5	-
DPP10	6	5	-
GAA	5	5	-
ATRX	5	5	Tak
DHX34	5	5	-
PRSS38	5	5	-
GLI2	5	5	-
CCDC88C	7	5	-

## ZAŁĄCZNIKI

CCDC129	5	5	-
PTGFR	5	5	-
RTL1	5	5	-
PEAR1	5	5	-
MAMLD1	9	5	-
FRYL	5	5	-
PRRC2B	7	5	-
NHS	5	5	-
HHIPL2	5	5	-
OTOF	8	5	-
PCYOX1L	6	5	-
PTPRB	6	5	Tak
GABRE	7	5	-
CCDC88B	6	5	-
TSPEAR	6	5	-
KLHDC7A	12	5	-
MAGEC3	14	5	-
ANKRD30	5	5	-
KIAA2026	5	5	-
SGIP1	7	5	-
SPEF2	8	5	-
TNRC18	7	5	-
PTPN13	7	5	Tak
STOX1	7	5	-
NLRP4	6	5	-
HGF	9	5	-
BLM	6	5	Tak
OAS3	5	5	-
SPOCD1	5	5	-
ATP8B4	6	5	-
FN1	6	5	-
CCDC79	5	5	-
BPIFB6	5	5	-
TCP11L1	5	5	-
LAD1	5	5	-
BTN1A1	6	5	-
ATP1A1	5	5	Tak
PARP14	7	5	-
MKI67	23	5	-

TCP11	5	5	-
SLC6A16	5	5	-
HELB	6	5	-
EP400	5	5	-
DYNC2H1	7	5	-
TNN	6	5	-
STAB1	5	5	-
LAMC3	5	5	-
WNK1	5	5	-
FSTL4	5	5	-
TTBK1	6	5	-
ADAM7	5	5	-
USP29	5	5	-
REXO1	7	5	-
CD93	5	5	-
ADAMTS	5	5	-
PSG6	5	5	-
GLIS3	6	5	-
NCOR2	5	5	Tak
C10orf71	15	5	-
VCAN	7	5	-
WDR17	6	5	-
KIAA1107	6	5	-
C12orf35	9	5	-
NACAD	8	5	-
STK11IP	5	5	-
MS4A14	6	5	-
JMJD1C	7	5	-
MUC12	9	5	-



## ZAŁĄCZNIKI

Załącznik 3. Geny często mutowane w intronach.

Nazwa genu	Liczba mutacji	Liczba zmutowanych próbek	Obecność w bazie CGC
PTCH1	24	20	Tak
PCDHA1	78	17	-
USH2A	27	16	-
GPR98	34	16	-
SYNE1	29	16	-
PKHD1L1	22	15	-
FSIP2	30	15	-
MUC5B	22	15	-
APOB	28	14	-
XIRP2	19	14	-
PCDHGA1	36	14	-
ZNF469	29	14	-
PTPRD	19	13	Tak
PKHD1	16	13	-
TP53	16	13	Tak
ANK3	18	12	-
IGFN1	19	12	-
RP1	19	12	-
PCDH15	19	12	-
EYS	25	12	-
GRIN2A	17	11	Tak
FAT4	16	11	Tak
DIP2A	13	11	-
PKD1	11	11	-
DCHS2	17	11	-
MUC17	12	11	-

DNAH7	17	11	-
RIMS1	13	11	-
MYO15A	16	11	-
ABCA13	17	11	-
RYR1	15	10	-
MLL2	19	10	-
CSMD2	14	10	-
BAI3	12	10	-
ACAN	17	10	-
PREX1	10	10	-
DSCAM	16	10	-
DNAH9	17	10	-
FBN1	12	10	-
NRG3	12	10	-
FLG	15	10	-
NEFM	11	10	-
MAGEC1	20	10	-
RELN	16	10	-
SI	13	10	-
CACNA1H	15	10	-
PRUNE2	12	10	-
FAM135B	15	10	Tak
ZFHX4	13	9	-
SIPA1L2	10	9	-
LRRC7	12	9	-
MYOCD	13	9	-
HDAC9	14	9	-
FLG2	9	9	-
AHNAK2	14	9	-
HEATR7B2	14	9	-
FREM2	13	9	-
BNC2	10	9	-

## ZAŁĄCZNIKI

TAF1L	13	9	-
KDR	10	9	Tak
MYH13	10	9	-
SLIT2	12	9	-
THSD7A	13	9	-
VPS13B	12	9	-
SCN9A	12	9	-
STAB2	11	9	-
FREM1	14	9	-
WDFY4	9	9	-
KANK4	9	9	-
PXDNL	12	9	-
LPA	13	9	-
POM121L12	9	9	-
KIAA1211	12	9	-
SVEP1	13	9	-
CMYA5	37	9	-
NOTCH1	10	8	Tak
CKAP5	9	8	-
SLC8A3	8	8	-
MYCN	8	8	Tak
AHNAK	12	8	-
TDRD6	11	8	-
FMN2	11	8	-
ZFYVE26	9	8	-
ARHGAP35	9	8	-
NEB	14	8	-
RREB1	14	8	-
GRM8	13	8	-
BIRC6	9	8	Tak
KIAA1109	13	8	-
GREB1	11	8	-

MED12L	8	8	-
DLC1	9	8	-
VWF	9	8	-
WDR87	11	8	-
ZNF804B	19	8	-
NLRP12	9	8	-
MYO7B	12	8	-
TEP1	11	8	-
KLHL4	8	8	-
CNGB1	9	8	-
PLEC	8	8	-
NAV3	9	8	-
GPR112	18	8	-
MYEF2	8	8	-
TTBK2	9	8	-
CASS4	9	8	-
RP1L1	12	8	-
ZFHX3	10	8	Tak
NES	13	8	-
KIAA1210	9	8	-
ALPK2	10	8	-
LRRK2	9	8	-
SRRM2	9	8	-
DIDO1	8	8	-
SPHKAP	10	8	-
KNDC1	11	8	-
ZNF831	11	8	-
SPTA1	16	8	-
NACA	10	8	Tak
NOTCH2	8	7	Tak
SLC4A9	8	7	-
FAT2	7	7	-

## ZAŁĄCZNIKI

TMEM132D	9	7	-
CRB1	9	7	-
CILP2	10	7	-
UNC80	8	7	-
PLCL1	7	7	-
HCN1	8	7	-
DOCK10	9	7	-
MBD5	9	7	-
SHANK1	10	7	-
ATRNL1	8	7	-
FSTL5	9	7	-
RNU6-28	7	7	-
FBN2	12	7	-
NPAP1	9	7	-
ODZ1	10	7	-
CELSR3	10	7	-
CACNA1A	10	7	-
SPAG17	8	7	-
MYO18B	11	7	-
SEZ6L	8	7	-
EPHA5	7	7	-
PDE10A	7	7	-
LRP1	8	7	-
LAMA2	7	7	-
ABCA6	7	7	-
MYO16	8	7	-
ZNF521	7	7	Tak
CELSR2	9	7	-
PSME4	7	7	-
UBR4	7	7	-
CUBN	8	7	-
PCDHB6	9	7	-

LPPR4	7	7	-
ST18	7	7	-
CXorf22	10	7	-
NALCN	9	7	-
ANK2	8	7	-
TRIOBP	12	7	-
MYOM3	11	7	-
ABCA7	8	7	-
SPTBN2	8	7	-
ASXL3	7	7	-
CUX2	7	7	-
STARD9	8	7	-
MDN1	11	7	-
ZNF541	8	7	-
ACACB	7	7	-
NWD1	7	7	-
FRY	8	7	-
NLRC5	10	7	-
MYO1H	8	7	-
MACF1	7	7	-
PCDHB16	9	7	-
C7	10	7	-
MRC1	7	7	-
EPG5	7	7	-
VWDE	7	7	-
CFH	8	7	-
HIVEP3	7	7	-
SH2D7	7	7	-
MEFV	9	7	-
PRRC2C	9	7	-
FNDC1	8	7	-
PLCE1	10	7	-

## ZAŁĄCZNIKI

FCGBP	11	7	-
WDR33	8	7	-
PPP1R3A	8	7	-
ALMS1	17	7	-
AK5	8	6	-
ADCY8	7	6	-
ZNF536	7	6	-
ZIM2	9	6	-
GCN1L1	6	6	-
PTPRF	6	6	-
FAM59A	6	6	-
TG	8	6	-
CHRNA9	6	6	-
SACS	8	6	-
IL16	8	6	-
ADAMTS20	11	6	-
SLC4A4	8	6	-
CDKL5	7	6	-
SORL1	6	6	-
CYP17A1	6	6	-
HMCN1	7	6	-
ADCY1	6	6	-
RNU6-71	13	6	-
ZMAT4	6	6	-
NELL2	6	6	-
HUWE1	6	6	-
SPTB	7	6	-
MDGA2	7	6	-
AATK	7	6	-
ZEB1	6	6	Tak
BRD4	6	6	Tak
ODZ4	6	6	-

MLL	7	6	-
ROBO3	7	6	-
URGCP-MRPS24	6	6	-
NOS1	6	6	-
ANKRD30A	8	6	-
ZNF208	6	6	-
NPR1	7	6	-
TPTE	6	6	-
ROS1	9	6	Tak
CPS1	6	6	-
LRRC3B	6	6	-
PRDM16	7	6	Tak
DCC	8	6	Tak
UNC13C	8	6	-
OPCML	6	6	-
SPEG	6	6	-
EPHA10	11	6	-
CNTNAP2	6	6	Tak
SLC8A1	8	6	-
MYO9A	6	6	-
SPEN	7	6	Tak
EPHA3	8	6	Tak
FLNC	7	6	-
MAP2	6	6	-
EPB41L3	6	6	-
NYAP2	7	6	-
LRBA	6	6	-
ZNF609	6	6	-
SYNE2	9	6	-
ABCA4	7	6	-
DMXL2	6	6	-
ULK1	6	6	-

## ZAŁĄCZNIKI

EPHA6	8	6	-
FAM75D1	6	6	-
TEX11	6	6	-
SLCO1C1	8	6	-
GRIN3A	6	6	-
PTPRZ1	8	6	-
TNFAIP2	6	6	-
HCFC1	7	6	-
HECTD1	7	6	-
PCDHB8	9	6	-
TEX15	9	6	-
FRMPD4	9	6	-
TTC21A	10	6	-
PLIN4	8	6	-
NRAP	13	6	-
ZMIZ1	6	6	-
IGSF10	8	6	-
SIGLEC1	7	6	-
HEATR8-TTC4	7	6	-
PCDH9	6	6	-
SLFN13	6	6	-
PKD1L1	10	6	-
GRIN2B	6	6	-
CHD4	6	6	Tak
PRIC285	8	6	-
CHD5	9	6	-
TLL6	6	6	-
CD163L1	9	6	-
TIE1	6	6	-
AKD1	7	6	-
LEPR	8	6	-
FAM186A	10	6	-

RASGRP3	8	6	-
ZNF804A	7	6	-
FREM3	6	6	-
TRANK1	7	6	-
KIF6	7	6	-
CFHR5	6	6	-
ATP10B	8	6	-
ITGAM	7	6	-
CC2D1A	7	6	-
CEP112	6	6	-
TRIO	6	6	-
KCNH8	6	6	-
TMPRSS11F	7	6	-
FAT1	12	6	Tak
F5	9	6	-
REV3L	7	6	-
AXDND1	8	6	-
WDR7	6	6	-
INADL	6	6	-
CPAMD8	12	6	-
FCRL5	7	6	-
KLC1	7	6	-
ZEB2	6	6	-
CDYL2	6	6	-
CDH9	7	6	-
AKAP13	7	6	-
VGLL3	6	6	-
DOCK5	6	6	-
MYBPC3	6	6	-
CYP4A11	6	6	-
PPP1R9A	9	6	-
PKDREJ	7	6	-

## ZAŁĄCZNIKI

POLQ	9	6	Tak
FCRL4	6	6	Tak
GPRIN1	6	6	-
MLL4	7	6	-
SLC22A25	7	6	-
AHRR	6	6	-
GNAS	9	6	Tak
ANKRD24	6	6	-
PIGQ	8	6	-
KDM4C	8	6	-
BANK1	9	6	-
C14orf39	7	6	-
FOXR2	6	6	-
GPR158	7	6	-
TGFB111	5	5	-
CARD6	5	5	-
CTNND2	6	5	Tak
PCDH7	5	5	-
FLNA	7	5	Tak
NOS3	5	5	-
FSHR	6	5	-
CNGA3	6	5	-
A1CF	5	5	Tak
FLNB	8	5	-
DNAH3	8	5	-
ADCY4	5	5	-
PYGB	6	5	-
NELL1	5	5	-
SYNPO	10	5	-
SYT3	5	5	-
SLC36A2	5	5	-
CREBBP	7	5	Tak

CTNNA3	6	5	-
SETD1A	5	5	-
DDX60L	8	5	-
CNBD1	5	5	Tak
MYH7	7	5	-
NCAN	7	5	-
CREB3L4	7	5	-
DNMT3B	5	5	-
CELSR1	6	5	-
MAEL	5	5	-
TMPRSS11B	5	5	-
NCKAP5L	5	5	-
LATS1	6	5	Tak
ABCB5	5	5	-
MYH3	6	5	-
C16orf96	7	5	-
OTOL1	5	5	-
ADAMTS7	5	5	-
TIMELESS	5	5	-
PDE6C	5	5	-
PLXNC1	7	5	-
FZD9	5	5	-
ZNF318	6	5	-
SHROOM2	5	5	-
MATN4	5	5	-
KANK1	9	5	-
MYO3B	5	5	-
BRWD3	8	5	-
XKR4	5	5	-
SDK2	6	5	-
TNIK	5	5	-
TLN1	6	5	-

## ZAŁĄCZNIKI

FAM5B	6	5	-
TANC1	5	5	-
TECTA	5	5	-
PBRM1	5	5	Tak
CHD7	8	5	-
RBP3	6	5	-
NEBL	8	5	-
BSN	6	5	-
FAM194A	6	5	-
PCDH1	10	5	-
TRPM3	7	5	-
GRIA2	5	5	-
DYNC1H1	5	5	-
FAM171A1	6	5	-
DNAJC13	5	5	-
BCL9L	6	5	Tak
KIF1A	5	5	-
DLG2	5	5	-
DCAF4	5	5	-
KIAA0240	5	5	-
BAI2	5	5	-
DCAF4L2	5	5	-
BRD1	5	5	-
RHOT2	8	5	-
GRIK2	5	5	-
AR	5	5	Tak
CADPS	6	5	-
CACNA1F	6	5	-
PRDM14	5	5	-
FAM135A	7	5	-
GRIK4	5	5	-
ABCC11	8	5	-

ATP13A1	5	5	-
PLXNA3	6	5	-
CADPS2	6	5	-
FAM155A	5	5	-
RBMXL2	5	5	-
AKAP6	6	5	-
ADAMTS19	6	5	-
IRS2	6	5	-
C6	7	5	-
ARFGEF2	5	5	-
SAMD9L	5	5	-
EPHA8	7	5	-
GIN1	5	5	-
UBASH3A	5	5	-
IVL	5	5	-
ANKRD11	5	5	-
TACC2	6	5	-
DIRC2	5	5	-
CNKSR2	6	5	-
CEACAM8	5	5	-
UGT2B11	8	5	-
RGS11	6	5	-
SPATA20	5	5	-
CDH23	5	5	-
ITGAX	5	5	-
UMODL1	10	5	-
USP26	6	5	-
DAB2	5	5	-
PCDHB4	6	5	-
ZSCAN10	5	5	-
ARID1A	5	5	Tak
RBM19	5	5	-

## ZAŁĄCZNIKI

FBXO40	5	5	-
PRKCB	5	5	Tak
PDE3A	5	5	-
ARHGEF2	6	5	-
NLRP5	9	5	-
DOT1L	6	5	-
MYH8	5	5	-
SYT16	5	5	-
CCDC155	5	5	-
NBAS	5	5	-
ALOX15B	6	5	-
SIM1	5	5	-
F8	6	5	-
MMRN2	7	5	-
RIMS2	6	5	-
L1CAM	8	5	-
DNAH6	6	5	-
ROBO1	6	5	-
ADAMTSL3	6	5	-
SLFN5	6	5	-
ARID1B	6	5	Tak
ROR1	6	5	-
INTS5	6	5	-
MYO3A	5	5	-
EIF2AK3	5	5	-
MYLK3	5	5	-
PTPN21	6	5	-
TLN2	6	5	-
GPR123	6	5	-
NBEA	5	5	Tak
ATP2C2	6	5	-
KALRN	6	5	-

FLT4	6	5	Tak
NCOA3	5	5	-
SMCHD1	5	5	-
MEIS2	6	5	-
BCORL1	6	5	Tak
GREB1L	5	5	-
TGM6	5	5	-
C6orf118	6	5	-
GPR128	5	5	-
SLC4A10	6	5	-
FAM5C	5	5	-
MSH4	5	5	-
STARD13	5	5	-
TRIP11	6	5	Tak
EYA2	5	5	-
MEGF8	7	5	-
RTTN	5	5	-
KCNT2	6	5	-
IHH	5	5	-
ARHGEF10L	6	5	Tak
NOTCH3	6	5	-
RADIL	7	5	-
ADAMTS12	6	5	-
ATP13A5	6	5	-
FAM71B	5	5	-
CDON	8	5	-
PIK3CG	6	5	-
ADGB	5	5	-
BCHE	6	5	-
VPS13A	5	5	-
SHROOM3	7	5	-
ATP13A4	5	5	-



## ZAŁĄCZNIKI

ZC3HC1	6	5	-
FSCB	7	5	-
TRPM5	5	5	-
DCAF8L1	5	5	-
MANBA	5	5	-
WWC2	5	5	-
VWA3B	8	5	-
ITGAD	5	5	-
C1orf173	5	5	-
CACNA1E	5	5	-
LCP2	5	5	-
SNED1	5	5	-
EML5	5	5	-
IBSP	6	5	-
PTPRN	7	5	-
MYH4	10	5	-
CNTN3	6	5	-
NYNRIN	6	5	-
METTL11B	6	5	-
FAM47A	6	5	-
PCDHB5	6	5	-
TCHH	5	5	-
ABCA9	5	5	-
TMEM232	5	5	-
ABCC8	5	5	-
DPP10	6	5	-
GAA	5	5	-
ATRX	5	5	Tak
DHX34	5	5	-
PRSS38	5	5	-
GLI2	5	5	-
CCDC88C	7	5	-

CCDC129	5	5	-
PTGFR	5	5	-
RTL1	5	5	-
PEAR1	5	5	-
MAMLD1	9	5	-
FRYL	5	5	-
PRRC2B	7	5	-
NHS	5	5	-
HHIPL2	5	5	-
OTOF	8	5	-
PCYOX1L	6	5	-
PTPRB	6	5	Tak
GABRE	7	5	-
CCDC88B	6	5	-
TSPEAR	6	5	-
KLHDC7A	12	5	-
MAGEC3	14	5	-
ANKRD30B	5	5	-
KIAA2026	5	5	-
SGIP1	7	5	-
SPEF2	8	5	-
TNRC18	7	5	-
PTPN13	7	5	Tak
STOX1	7	5	-
NLRP4	6	5	-
HGF	9	5	-
BLM	6	5	Tak
OAS3	5	5	-
SPOCD1	5	5	-
ATP8B4	6	5	-
FN1	6	5	-
CCDC79	5	5	-

## ZAŁĄCZNIKI

BPIFB6	5	5	-
TCP11L1	5	5	-
LAD1	5	5	-
BTN1A1	6	5	-
ATP1A1	5	5	Tak
PARP14	7	5	-
MKI67	23	5	-
TCP11	5	5	-
SLC6A16	5	5	-
HELB	6	5	-
EP400	5	5	-
DYNC2H1	7	5	-
TNN	6	5	-
STAB1	5	5	-
LAMC3	5	5	-
WNK1	5	5	-
FSTL4	5	5	-
TTBK1	6	5	-
ADAM7	5	5	-
USP29	5	5	-
REXO1	7	5	-
CD93	5	5	-
ADAMTS2	5	5	-
PSG6	5	5	-
GLIS3	6	5	-
NCOR2	5	5	Tak
C10orf71	15	5	-
VCAN	7	5	-
WDR17	6	5	-
KIAA1107	6	5	-
C12orf35	9	5	-
NACAD	8	5	-

STK11IP	5	5	-
MS4A14	6	5	-
JMJD1C	7	5	-
MUC12	9	5	-

## ZAŁĄCZNIKI

Załącznik 4. Regiony ulegające amplifikacjom oraz delecjom na podstawie analiz z użyciem GISTIC 2.0.

Prążek chromosomowy	q-wartość	Nazwa genu
<b>amplifikacje</b>		
1p36.11	1,57E-11	RSRP1
1p36.21	0,0002	PRAMEF6
1p36.33	2,44E-05	MMP23A;SLC35E2;CDK11A
1q21.1	1,75E-05	RNU1-4;RNU1-3;RNU1-2;RNU1-1;LOC388692;FAM72C;FAM231D;LOC645166;FCGR1CP;RNVU1-17
1q21.3	2,38E-06	LCE2D;LCE3A;LCE3B;LCE3C
1q31.3	6,97E-07	CFHR1
1q44	0,0004	OR2T10;OR2T11
2p25.3	0,1622	TPO
2q37.3	0,0938	LOC285095;LOC102723927
3p12.3	3,13E-05	FAM86DP;MIR1324;MIR4444-1
3q29	0,0001	MUC4
3q29	0,0001	MUC20;LINC00969;MIR570;SDHAP2;LOC105374297
4p16.1	5,80E-06	USP17L9P;USP17L24;USP17L25;USP17L26;USP17L27;USP17L28;USP17L29;USP17L30;USP17L1;USP17L12;USP17L17;USP17L19;USP17L20
4q13.2	3,30E-12	UGT2B28
4q13.2	1,19E-06	TMPRSS11E
5p15.33	0,0497	ZDHHC11
5q13.2	0,0052	LOC102724392
5q35.3	1,19E-06	RACK1;BTNL3;OR2V1;TRIM7;TRIM52;TRIM41;BTNL9;OR2V2;SNORD95;SNORD96A;TRIM52-AS1;MIR4638;CTC-338M12.4;MIR8089;LOC102577426
5q35.3	4,54E-05	OR4F16;OR4F29
6p21.32	3,30E-12	HLA-DQA1;HLA-DRA;HLA-DRB1;HLA-DRB5;HLA-DRB6

## ZAŁĄCZNIKI

6p21.32	0,0002	HLA-DPA1;HLA-DPB1;HLA-DPB2;HCG24
6p22.1	2,01E-09	HLA-A;HCG4B
7q22.1	0,0431	RASA4;POLR2J2;SPDYE2;POLR2J3;UPK3BL;RASA4B
7q34	0,0154	TAS2R38;MGAM
8p11.22	3,10E-06	ADAM3A;ADAM5
8p23.1	4,31E-07	FAM90A10P
8p23.1	4,15E-06	DEFB109P1;LOC649352;FAM66A
8p23.1	8,98E-07	USP17L2;LOC392196;USP17L7;FAM66D
8q21.2	0,0052	REXO1L2P
9p11.2	0,0022	SPATA31A6
9p11.2	0,029	ANKRD20A3;ANKRD20A2;AQP7P3;GXYLT1P3;FAM95B1;MIR1299;LOC101929583;LOC102724580
9p24.3	0,001	FOXD4;FAM138A;FAM138C;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11;PGM5P3-AS1
10q26.3	1,27E-10	DUX4
11q11	6,04E-08	OR4P4;OR4C16;OR4C11;OR4S2;OR4C6
11q12.2	0,0071	PGA5;PGA3;PGA4
11q14.3	0,0098	TRIM49
12p13.2	0,0003	TAS2R43;TAS2R46
13q11	5,86E-06	FAM230C;LINC00417
14q32.33	0,1795	LINC00226;LINC00221;MIR5195;MIR7641-2
15q11.2	0,0006	OR4N4;IGHV1OR15-1;OR4M2;OR4N3P
15q11.2	0,0071	GOLGA6L1;GOLGA6L22;GOLGA8DP;MIR4509-1;MIR4509-2
15q26.3	0,0813	OR4F4;WASH3P;FAM138E;DDX11L9;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11;MIR6859-2;MIR6859-3;MIR6859-1
16p11.2	0,002	TP53TG3;LOC390705;TP53TG3C;TP53TG3B;TP53TG3E;TP53TG3F
16p13.11	0,0297	PDXDC1;MIR1972-1;MIR1972-2
17p11.2	0,0001	FAM106A;USP32P2;LGALS9C
17q21.31	1,38E-05	LRRC37A

## ZAŁĄCZNIKI

18q21.1	0,0001	TCEB3CL;TCEB3CL2
19p13.3	0,0158	OR4F17;WASH5P;FAM138A;FAM138C;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11
19q13.31	0,0009	PSG1;PSG2;PSG6;PSG7;PSG11;PSG10P;LOC100289650
20p13	0,009	CSNK2A1;FKBP1A;SOX12;TCF15;PSMF1;SNPH;SIRPB1;RBCK1;SDCBP2;ANGPT4;TMEM74B;NSFL1C;TRIB3;NRSN2;DEFB126;FAM110A;ZCCHC3;SCRT2;SLC52A3;TBC1D20;SIRPD;C20orf96;SRXN1;DEFB127;DEFB129;DEFB125;DEFB128;SIRPB2;RSPO4;DEFB132;C20orf202;RAD21L1;NRSN2-AS1;SDCBP2-AS1;FKBP1A-SDCBP2;MIR6869;LOC105372493
21p11.2	0,0564	TPTE
22q11.1	0,0571	POTEH;OR11H1;CCT8L2;DUXAP8;POTEH-AS1;BMS1P17;LINC01297;BMS1P22
22q11.21	0,0021	FAM230A;TMEM191B;LOC729461
22q11.23	0,0001	DDT;GSTT2;GSTTP1;LOC391322;GSTT2B;DDTL
<b>delecje</b>		
1p11.2	0,0007	RNVU1-19
1p21.1	1.5273e-07	AMY1A;AMY1B;AMY1C;AMY2A
1p36.11	0,0085	RHD;RSRP1
1q23.3	0,0056	HSPA7;FCGR2C
1q31.3	0,0003	CFHR1;CFHR3
2p11.2	0,067	LOC654342
2p11.2	0,1305	LINC00152;MIR4435-2;MIR4771-1;MIR4435-1
3q29	0,0496	MUC20;LINC00969;MIR570;SDHAP2
4p15.2	0,155	STIM2
4p16.1	0,0014	USP17L6P;USP17L9P;USP17L24;USP17L25;USP17L26;USP17L27;USP17L28;USP17L29;USP17L30;USP17L11;USP17L12;USP17L13;USP17L17;USP17L19;USP17L20;USP17L15
4q11	0,0167	DCUN1D4
4q13.2	0,0496	UGT2B15;UGT2B17;TMPRSS11E
4q35.2	0,016	DUX4;DBET
5p15.33	0,155	ZDHHC11
5q13.2	0,0533	SMA5;LOC441081;GUSBP3;GUSBP9;LOC102724392

## ZAŁĄCZNIKI

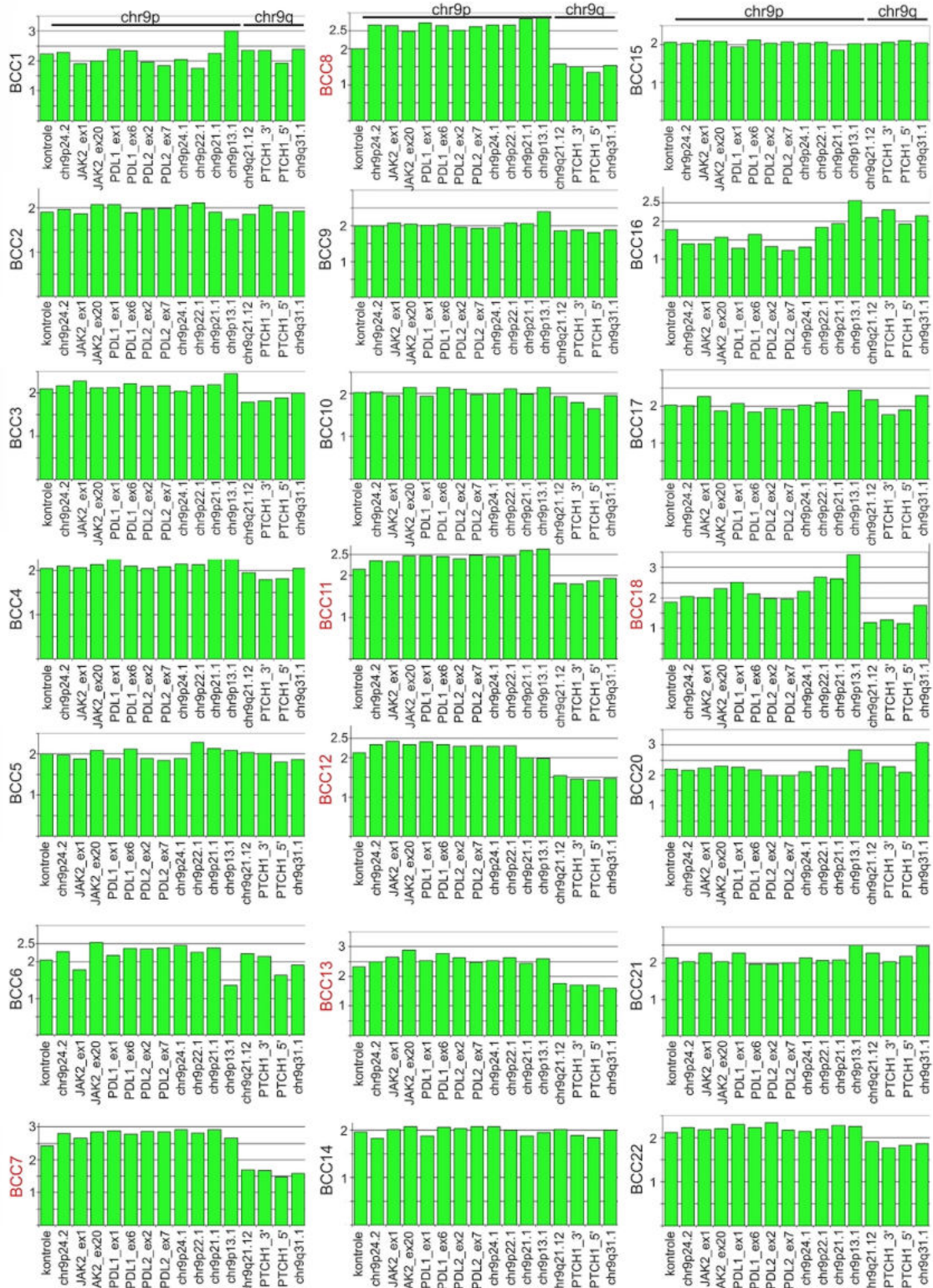
5q35.3	0,0014	BTNL3;BTNL8
6p11.1	0,067	LINC00680-GUSBP4
6p21.33	0,1404	CDSN;HLA-B;HLA-C;MICB;POU5F1;TCF19;DDX39B;HCP5;C6orf15;CCHCR1;PSORS1C1;PSORS1C2;HCG27;HCG26;MCCD1;SNORD117;PSORS1C3;MICA;ATP6V1G2-DDX39B;MIR6891
6p25.3	0,1787	IRF4;DUSP22
8p11.1	0,1128	POTEA
8p11.22	0,0014	ADAM3A;ADAM18;ADAM32;ADAM5;LOC100130964
8p23.1	0,0003	DEFB109P1;LOC649352;FAM66A
8p23.1	0,0179	FAM90A7P
8p23.1	0,1268	LINC00965
8q21.2	0,0001	REXO1L2P
9p24.3	0,0496	FOXD4;FAM138A;FAM138C;WASH1;DDX11L5;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11;PGM5P3-AS1
9q11	0,0014	SPATA31A7
9q12	0,0556	SPATA31A7;FAM74A1;SPATA31A5
10q11.22	0,0019	ANXA8L1;FAM25BP
10q26.3	0,0195	DUX4
11p11.11	0,0001	OR4C46
11q12.2	0,0045	PGA5;PGA3;PGA4
11q14.3	0,1492	TRIM49
12p13.31	0,0068	KLRB1
12p13.33	0,016	IQSEC3;FAM138D;LOC100288778
12q11	0,0143	ALG10
14q11.2	0,0179	OR4K5;OR11H2;OR4K1;OR4N2;OR4K2;OR4Q3;OR4M1
15q11.2	0,1268	NBEAP1;HERC2P3;GOLGA6L6;GOLGA8CP
15q26.3	0,0039	LOC100128108

## ZAŁĄCZNIKI

15q26.3	0,067	OR4F4;WASH3P;FAM138E;DDX11L9;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11;MIR6859-2;MIR6859-3;MIR6859-1
16p11.1	0,027	LINC01566
16q11.1	0,0014	ANKRD26P1
16q24.3	0,1268	FAM157C
17p11.1	0,0062	LOC105371703
17p11.2	0,0014	FAM106A;USP32P2;CCDC144B;LGALS9C;KRT16P1
17p11.2	0,1404	KCNJ12;C17orf51;KCNJ18
17q12	0,0602	TBC1D3F;TBC1D3C;LOC440434;TBC1D3;TBC1D3H;TBC1D3G;TBC1D3K;TBC1D3L;LOC101060389;TBC1D3E;TBC1D3I
17q21.31	0,0001	NSF;LRRC37A;ARL17A;LRRC37A2;NSFP1;ARL17B
18p11.1	0,0014	ROCK1
18q21.1	0,0556	TCEB3B;KATNAL2;TCEB3C;TCEB3CL;TCEB3CL2
19p13.3	0,1128	OR4F17;WASH5P;FAM138A;FAM138C;MIR1302-2;MIR1302-9;MIR1302-10;MIR1302-11
20p12.1	0,2314	SPTLC3;TASP1;ISM1;LOC100505515;ISM1-AS1;LOC101929486;LOC102606466
21p11.2	0,14	TPTE
22q11.21	8.541e-06	RIMBP3;TMEM191B;PI4KAP1
22q11.23	0,0496	GRK3;CRYBB2;CRYBB2P1;IGLL3P;LRP5L;MIR6817
22q11.23	0,1492	DDT;GSTT2;GSTTP1;LOC391322;GSTTP2;GSTT2B;DDTL

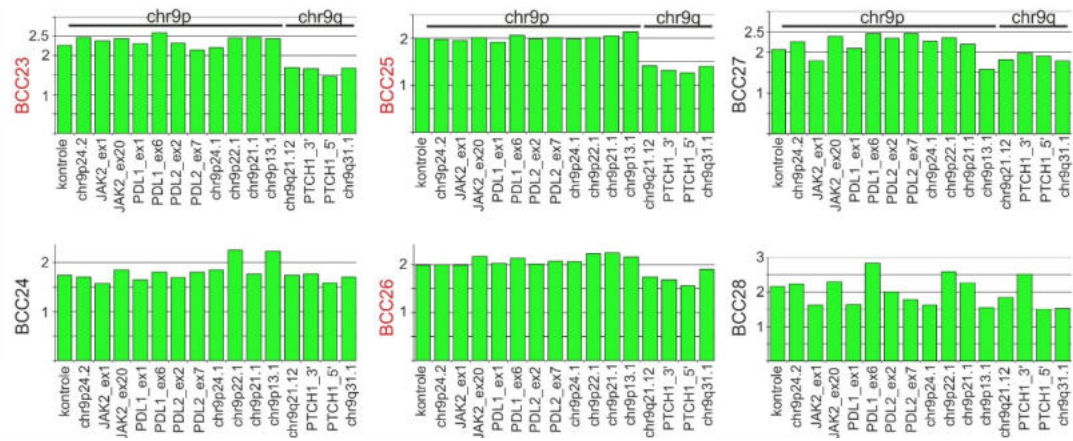
# ZAŁĄCZNIKI

Załącznik 5. Zmiany liczby kopii zlokalizowane na chromosomie 9, wykryte przez reakcję MLPA.





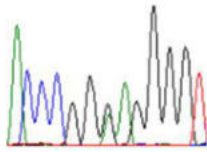
# ZAŁĄCZNIKI



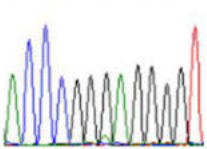
## ZAŁĄCZNIKI

Załącznik 6. Mutacje w promotorze genu *TERT* wykryte przez sekwencjonowanie metodą Sangera.

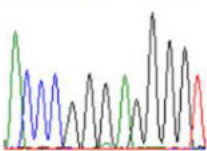
BCC1 c.-146 G>A  
A C C C G G G A G G G G T



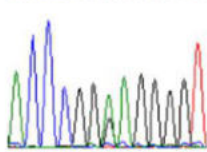
BCC3 c.-146 G>A  
A C C C G G G A G G G G T



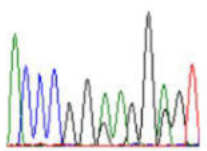
BCC4 c.-146 G>A  
A C C C G G G C G G G G T



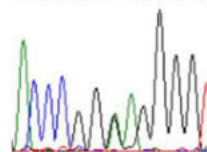
BCC18 c.-146 G>A  
A C C C G G G A G G G G T



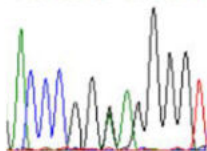
BCC20 c.-146 G>A; c.-150 G>A  
A C C C G G G A G G G G T



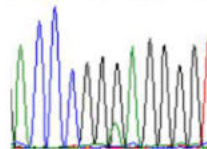
BCC25 c.-146 G>A  
A C C C G G G A G G G G T



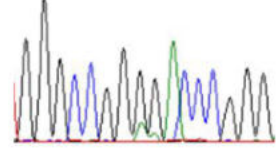
BCC26 c.-146 G>A  
A C C C G G G A G G G G T



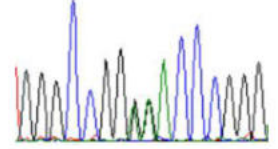
BCC27 c.-146 G>A  
A C C C G G G A G G G G T



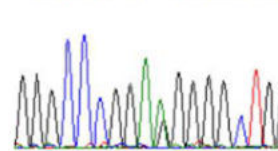
BCC5 c.-139\_138delinsAA  
G G G C C G G G G A C C C G G G



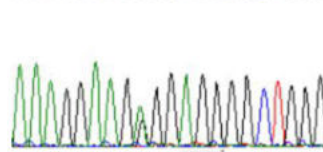
BCC11 c.-139\_138delinsAA  
G G G C C G G G G A C C C G G G



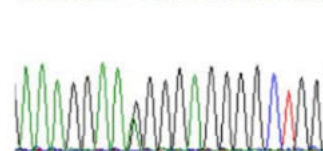
BCC28 c.-124 G>A  
G G G C C C G G A G G G G G C T G G G



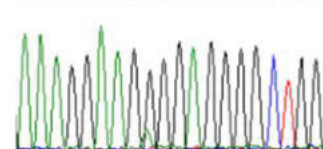
BCC18 c.-101 G>A  
A A A G G A A G G G G A G G G G C T G G G



BCC26 c.-100 G>A  
A A A G G A A G G G G A G G G G C T G G G



BCC27 c.-101 G>A  
A A A G G A A G G G G A G G G G C T G G G

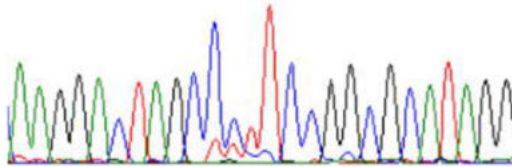


## ZAŁĄCZNIKI

Załącznik 7. Mutacje w promotorze genu *DPH3* wykryte sekwencjonowaniem metodą Sangera.

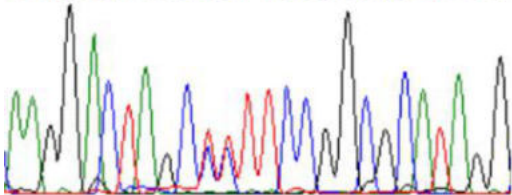
BCC5 c.-122\_121delinsTT

AAGGAC TAGC CCTT CCGGC GCATAGG



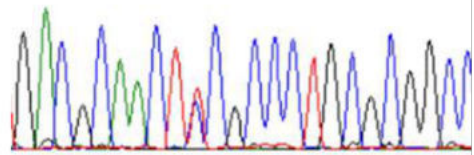
BCC16 c.-122\_121delinsTT

AAGG AC TAGC CCTT CCGGC GCATAGG



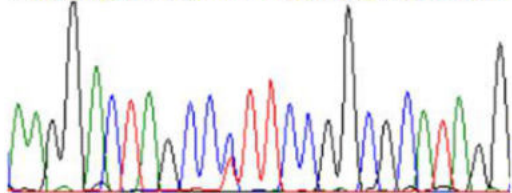
BCC16 c.-150 C>T

GACGCAACTCCGCCCTGCGCGGCC



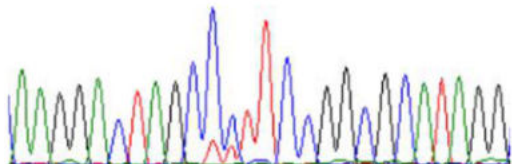
BCC18 c.-122 C>T

AAGG AC TAGC CCTT CCGGC GCATAGG



BCC24 c.-122\_121delinsTT

AAGGAC TAGC CCTT CCGGC GCATAGG



BCC28 c.-121 C>T

AAGG AC TAGC CCTT CCGGC GCATAGG

