

**Mgr inż. Monika Pępek**

**Charakterystyka nowych zmian genetycznych w przewlekłej  
białaczce szpikowej u chorych w różnych stadiach choroby  
przy zastosowaniu sekwencjonowania następnej generacji**

**Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu  
w dyscyplinie nauki medyczne**

Promotor: prof. dr med. Tomasz Stokłosa

Zakład Immunologii / Zakład Biologii i Genetyki Nowotworów



Obrona rozprawy doktorskiej przed Radą Dyscypliny Nauk Medycznych  
Warszawskiego Uniwersytetu Medycznego

Warszawa 2023 r.

**Słowa kluczowe:** przewlekła białaczka szpikowa (PBSz), faza przewlekła, faza blastyczna, sekwencjonowanie następnej generacji, profil zmian genetycznych, przetworzony pseudogen, *SMAD4*

**Keywords:** chronic myeloid leukemia (CML), chronic phase (CP), blast phase (BP), next-generation sequencing (NGS), mutational landscape, processed pseudogene, *SMAD4*

**Nazwa i numer projektu badawczego dotyczącego rozprawy:**

Grant NCN Harmonia nr 2014/14/M/NZ5/00441 pt. „Zidentyfikowanie nowych, istotnych klinicznie zmian genetycznych i epigenetycznych przy zastosowaniu wysokoprzepustowego sekwencjonowania następnej generacji u chorych na przewlekłą białaczkę szpikową z niekorzystnym przebiegiem choroby”. Kierownik projektu: prof. dr med. Tomasz Stokłosa, główny wykonawca projektu - stypendystka NCN: mgr inż. Monika Pępek

*Pragnę serdecznie podziękować mojemu Promotorowi Prof. Tomaszowi Stokłosie za poświęcony czas, cenne wskazówki oraz konstruktywną krytykę na każdym etapie przygotowywania mojej pracy doktorskiej, a także za zaangażowanie w mój rozwój naukowy.*

*Serdecznie dziękuję wszystkim Pracownikom i Doktorantom Zakładu Immunologii oraz Zakładu Biologii i Genetyki Nowotworów WUM za współpracę, codzienne dyskusje i wspaniałą atmosferę pracy. W szczególności dziękuję Panu Prof. Jakubowi Gołębowskiemu za możliwość pracy w tak inspirującym środowisku jakim jest Zakład Immunologii. Dziękuję również dr Marcinowi Machnickiemu za nieustanną pomoc, rady i wprowadzenie w świat NGSu.*

*Dziękuję mojej Rodzinie oraz Przyjaciółom za cierpliwość, zrozumienie i wsparcie, które sprawiły, że każdy krok w kierunku ukończenia tego etapu był łatwiejszy.*

# SPIS TREŚCI

|   |           |
|---|-----------|
| <b>SPIS RYCIN.....</b>  | <b>7</b>  |
| <b>SPIS TABEL .....</b>   | <b>9</b>  |
| <b>WYKAZ STOSOWANYCH SKRÓTÓW .....</b>  | <b>11</b> |
| <b>STRESZCZENIE.....</b>  | <b>13</b> |
| <b>ABSTRACT.....</b>  | <b>15</b> |
| <b>1. WSTĘP .....</b>   | <b>17</b> |
| 1.1. Przewlekła białaczka szpikowa.....   | 17        |
| 1.1.1. Epidemiologia i etiopatogeneza .....   | 17        |
| 1.1.2. Obraz kliniczny i przebieg choroby .....   | 21        |
| 1.1.3. Diagnostyka .....  | 21        |
| 1.1.4. Leczenie.....  | 22        |
| 1.1.5. Dodatkowe mutacje .....  | 24        |
| 1.2. Pseudogeny .....   | 27        |
| 1.2.1. Klasyfikacja .....   | 27        |
| 1.2.2. Funkcje pseudogenów w procesie karcynogenezy .....   | 30        |
| 1.2.3. Pseudogeny w nowotworach układu krwiotwórczego .....   | 33        |
| 1.3. Gen i pseudogen <i>SMAD4</i> .....   | 34        |
| <b>2. ZAŁOŻENIA I CEL PRACY.....</b>  | <b>36</b> |
| <b>3. MATERIAŁY I METODY .....</b>  | <b>37</b> |
| 3.1. Materiały .....  | 37        |
| 3.1.1. Materiał kliniczny.....  | 37        |
| 3.1.2. Odczynniki .....   | 38        |
| 3.2. Metody .....   | 40        |
| 3.2.1. Izolacja ludzkiego materiału genetycznego.....   | 40        |
| 3.2.2. Sekwencjonowanie DNA za pomocą metody NGS .....  | 42        |
| 3.2.3. Reakcja łańcuchowej polimerazy PCR do oceny obecności $\Psi$ <i>SMAD4</i> .....                  | 48        |
| 3.2.4. Rozdział elektroforetyczny produktów reakcji PCR.....  | 50        |
| 3.2.5. Określenie miejsca insercji $\Psi$ <i>SMAD4</i> za pomocą sekwencjonowania metodą Sanger'a ..... | 50        |
| 3.2.6. Określenie sekwencji $\Psi$ <i>SMAD4</i> za pomocą sekwencjonowania NGS.....                     | 53        |
| 3.2.7. Sekwencjonowanie transkryptomu (RNAseq) .....  | 57        |
| 3.2.8. Analiza statystyczna .....   | 59        |
| <b>4. WYNIKI .....</b>  | <b>60</b> |
| 4.1. Analiza danych uzyskanych z celowanego sekwencjonowania próbek CML.....                            | 60        |

|  |            |
|--|------------|
| 4.2. Identyfikacja patogennych wariantów, zmian liczby kopii i aberracji chromosomowych w fazie BP .....                         | 62         |
| 4.3 Identyfikacja patogennych wariantów, zmian liczby kopii i aberracji chromosomowych w próbkach sprzed progresji choroby ..... | 66         |
| 4.4 Analiza porównawcza wybranych zmian punktowych, małych insercji-delekcji i zmian liczby kopii w 25 sparowanych próbkach..... | 67         |
| 4.5 Ocena częstości występowania $\Psi$ SMAD4 w różnych rodzajach białaczek.....   | 70         |
| 4.6 Potwierdzenie zarodkowego charakteru $\Psi$ SMAD4 .....  | 74         |
| 4.7 Określenie genomowej lokalizacji $\Psi$ SMAD4 .....  | 75         |
| 4.8 Określenie pełnej sekwencji $\Psi$ SMAD4 wraz z miejscem insercji.....   | 77         |
| 4.9 Ocena typu obserwowanego pseudogenu .....  | 80         |
| 4.10 Identyfikacja patogennych wariantów, zmian liczby kopii i aberracji chromosomowych u chorych z $\Psi$ SMAD4 .....           | 80         |
| 4.11 Analiza transkryptomu u chorych z $\Psi$ SMAD4 .....  | 82         |
| <b>5. DYSKUSJA.....</b>  | <b>86</b>  |
| <b>6. WNIOSKI .....</b>  | <b>101</b> |
| <b>PIŚMIENNICTWO.....</b>  | <b>102</b> |
| <b>OPINIE KOMISJI BIOETYCZNYCH .....</b>   | <b>112</b> |

## SPIS RYCIN

|   |    |
|---|----|
| Rycina 1. Schemat przedstawiający translokację t(9;22)(q34;q11) skutkującą powstaniem chromosomu Filadelfia z genem fuzyjnym <i>BCR::ABL1</i> . .....   | 18 |
| Rycina 2. Schematyczne przedstawienie struktury genów <i>ABL1</i> i <i>BCR</i> oraz transkryptów <i>BCR::ABL1</i> w białaczkach. ....   | 20 |
| Rycina 3. Klasyfikacja pseudogenów ze względu na sposób ich powstania. ....   | 29 |
| Rycina 4. Schemat badań przeprowadzonych w niniejszej pracy doktorskiej. ....   | 41 |
| Rycina 5. Przykładowy wykres analizy zmian liczby kopii wygenerowany za pomocą programu CNVkit. ....  | 47 |
| Rycina 6. Schematyczne przedstawienie zasady sekwencjonowania w trybie sparowanych końców oraz rodzajów odczytów, uzyskiwanych w sekwencjonowaniu NGS, wykorzystanych do identyfikacji $\Psi$ <i>SMAD4</i> . .... | 48 |
| Rycina 7. Schemat lokalizacji starterów w reakcjach PCR wykonanych w celu określenia miejsca insercji $\Psi$ <i>SMAD4</i> . ....  | 50 |
| Rycina 8. Schemat lokalizacji starterów oraz ampliconów powstałych w reakcjach PCR wykonanych w celu określenia sekwencji $\Psi$ <i>SMAD4</i> . ....  | 54 |
| Rycina 9. Statystyki dotyczące patogennych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 45 próbek CML od chorych w fazie BP. ....  | 61 |
| Rycina 10. Statystyki dotyczące patogennych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 25 sparowanych próbek CML sprzed progresji choroby. ....  | 62 |
| Rycina 11. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 45 próbek CML w fazie BP. ....   | 64 |
| Rycina 12. Profil zmian genetycznych w najczęściej zmutowanych genach w CML w fazie BP. ....  | 65 |
| Rycina 13. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 25 próbek CML sprzed progresji choroby. ....   | 67 |
| Rycina 14. Analiza porównawcza częstości zmian punktowych, małych insercji-delekcji i zmian liczby kopii u 25 chorych w sparowanych próbkach CML sprzed i po progresji CML. ....                                  | 68 |
| Rycina 15. Przykładowa analiza wizualna obecności $\Psi$ <i>SMAD4</i> w badanych próbkach za pomocą sekwencjonowania NGS. ....  | 71 |
| Rycina 16. Przykładowy obraz elektroforetyczny przedstawiający wynik reakcji PCR oceniającej obecność $\Psi$ <i>SMAD4</i> w badanych próbkach. ....   | 72 |

|   |    |
|---|----|
| Rycina 17. Obraz elektroforetyczny przedstawiający wyniki reakcji PCR potwierdzającej zarodkowy charakter $\Psi$ SMAD4.....                           | 74 |
| Rycina 18. Schemat genomowej lokalizacji $\Psi$ SMAD4.....  | 76 |
| Rycina 19. Schematyczne porównanie struktury $\Psi$ SMAD4 i mRNA genu SMAD4. ....   | 77 |
| Rycina 20. Porównanie całkowitej sekwencji $\Psi$ SMAD4 z sekwencją mRNA genu rodzicielskiego SMAD4. ....   | 79 |
| Rycina 21. Statystyki dotyczące patogennych lub istotnych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 9 próbek CML z $\Psi$ SMAD4. .... | 81 |
| Rycina 22. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 9 próbek CML z $\Psi$ SMAD4.....                             | 82 |
| Rycina 23. Sygnatura ekspresji genów u chorych na CML z $\Psi$ SMAD4.....   | 84 |
| Rycina 24. Sygnatura ekspresji genów u chorych na CML z $\Psi$ SMAD4 z podziałem na fazy choroby.....   | 85 |



## SPIS TABEL

|  |    |
|--|----|
| Tabela 1. Lista najistotniejszych badań naukowych opublikowanych w ostatnich latach, zawierających analizy porównawcze mutacji somatycznych w sparowanych próbkach z różnych faz CML. .... | 26 |
| Tabela 2. Dane kliniczne chorych na CML włączonych do badania NGS.....   | 38 |
| Tabela 3. Szczegóły celowanego sekwencjonowania wysokoprzepustowego dotyczące zastosowanych zestawów sond wzbogacających oraz statystyki sekwencjonowania. ....                            | 43 |
| Tabela 4. Lista 193 genów poddanych analizie. ....   | 45 |
| Tabela 5. Skład mieszaniny reakcyjnej PCR z użyciem zestawu HotStarTaq® Plus DNA Polymerase (Qiagen).....  | 49 |
| Tabela 6. Sekwencje starterów oraz długości ampliconów w przesiewowym PCR do oceny obecności $\Psi$ SMAD4. ....  | 49 |
| Tabela 7. Profil temperaturowo-czasowy reakcji PCR do oceny obecności $\Psi$ SMAD4..   | 49 |
| Tabela 8. Sekwencje starterów oraz długości ampliconów w PCR do określenia miejsca insercji, połączenie SCAI- $\Psi$ SMAD4. ....   | 51 |
| Tabela 9. Profil temperaturowo-czasowy reakcji PCR do określenia miejsca insercji, połączenie SCAI- $\Psi$ SMAD4. ....   | 51 |
| Tabela 10. Sekwencje starterów oraz długość ampliconu w PCR do określenia miejsca insercji, połączenie $\Psi$ SMAD4-SCAI. ....   | 51 |
| Tabela 11. Profil temperaturowo-czasowy reakcji PCR do określenia miejsca insercji, połączenie $\Psi$ SMAD4-SCAI. ....   | 52 |
| Tabela 12. Skład mieszaniny reakcji fluorescencyjnego znakowania. ....   | 53 |
| Tabela 13. Profil temperaturowo-czasowy reakcji fluorescencyjnego znakowania. ....   | 53 |
| Tabela 14. Skład mieszaniny reakcyjnej PCR z użyciem zestawu KAPA HiFi HotStart ReadyMix (Roche) do amplifikacji $\Psi$ SMAD4- reakcja 1 i 2. ....   | 54 |
| Tabela 15. Skład mieszaniny reakcyjnej PCR z użyciem zestawu Expand™ Long Template PCR System (Roche) do amplifikacji $\Psi$ SMAD4– reakcja 3. ....  | 55 |
| Tabela 16. Sekwencje starterów oraz długości ampliconów w reakcjach PCR do amplifikacji $\Psi$ SMAD4. ....   | 55 |
| Tabela 17. Profil temperaturowo-czasowy pierwszej reakcji PCR do amplifikacji $\Psi$ SMAD4. ....   | 55 |

|  |    |
|--|----|
| Tabela 18. Profil temperaturowo-czasowy drugiej reakcji PCR do amplifikacji<br>$\Psi$ SMAD4.....   | 56 |
| Tabela 19. Profil temperaturowo-czasowy trzeciej reakcji PCR do amplifikacji<br>$\Psi$ SMAD4.....  | 56 |
| Tabela 20. Aberracje chromosomowe w CML w fazie BP.....  | 66 |
| Tabela 21. Analiza porównawcza patogennych wariantów ukazująca dynamikę zmian<br>zachodzących w 25 sparowanych próbkach sprzed i po progresji CML..... | 69 |
| Tabela 22. Częstość występowania $\Psi$ SMAD4 w różnych rodzajach białaczek.....   | 73 |

## WYKAZ STOSOWANYCH SKRÓTÓW

| Skrót        | Rozwinięcie skrótu (ang.)                              | Wyjaśnienie skrótu (pol.)                              |
|--------------|--|--|
| 3'UTR        | <i>3' untranslated region</i>                          | rejon 3' niepodlegający translacji                     |
| 5'UTR        | <i>5' untranslated region</i>                          | rejon 5' niepodlegający translacji                     |
| ACA          | <i>additional chromosomal abnormalities</i>            | dodatkowe aberracje chromosomowe                       |
| ALL          | <i>acute lymphoblastic leukemia</i>                    | ostra białaczka limfoblastyczna                        |
| allo-HCT     | <i>allogeneic hematopoietic cell transplantation</i>   | alogeniczny przeszczep komórek krwiotwórczych          |
| AML          | <i>acute myeloid leukemia</i>                          | ostra białaczka szpikowa                               |
| AP           | <i>acceleration phase</i>                              | faza akceleracji                                       |
| BP           | <i>blast phase</i>                                     | faza blastyczna  |
| CAR-T        | <i>T cells modified with chimeric antigen receptor</i> | limfocyty T z chimerycznym receptorem antygenowym      |
| cDNA         | <i>complementary DNA</i>                               | komplementarny DNA                                     |
| CDS          | <i>coding DNA sequence</i>                             | region kodujący  |
| ceRNA        | <i>competitive endogenous RNA</i>                      | kompetycyjny endogenny RNA                             |
| CHIP         | <i>clonal hematopoiesis of indeterminate potential</i> | hematopoeza klonalna o nieokreślonym potencjale        |
| chromosom Ph | <i>Philadelphia chromosome</i>                         | chromosom Filadelfia                                   |
| CLL          | <i>chronic lymphocytic leukemia</i>                    | przewlekła białaczka limfocytowa                       |
| CML          | <i>chronic myeloid leukemia</i>                        | przewlekła białaczka szpikowa                          |
| CP           | <i>chronic phase</i>                                   | faza przewlekła  |
| DNA          | <i>deoxyribonucleic acid</i>                           | kwasy deoksyrybonukleinowe                             |
| EFS          | <i>event-free survival</i>                             | czas przeżycia wolny od zdarzeń                        |
| ELN          | <i>European LeukemiaNet</i>                            | Europejska sieć badań nad białaczką                    |
| endo-siRNA   | <i>endogenous small interfering RNA</i>                | endogenne małe interferujące RNA                       |
| ERK          | <i>extracellular signal-regulated kinases</i>          | rodzina kinaz regulowanych sygnałem zewnątrzkomórkowym |
| ESMO         | <i>European Society for Medical Oncology</i>           | Europejskie Towarzystwo Onkologii Medycznej            |

|        |  |   |
|--------|--|---|
| FDA    | <i>Food and Drug Administration</i>                    | Amerykańska Agencja ds. Żywności i Leków              |
| KRN    | -  | Krajowy Rejestr Nowotworów                            |
| LINE   | <i>Long Interspersed Nuclear Elements</i>              | długie, rozproszone elementy jądrowe                  |
| MDS    | <i>myelodysplastic syndromes</i>                       | zespoły mielodysplastyczne                            |
| miRNA  | <i>micro RNA</i>                                       | mikroRNA  |
| MPN    | <i>myeloproliferative neoplasms</i>                    | nowotwory mieloproliferacyjne                         |
| MRE    | <i>miRNA recognition elements</i>                      | elementy rozpoznające miRNA                           |
| mRNA   | <i>messenger RNA</i>                                   | informacyjny RNA                                      |
| NCCN   | <i>National Comprehensive Cancer Network</i>           | Narodowa kompleksowa sieć ds. nowotworów              |
| NGS    | <i>next-generation sequencing</i>                      | sekwencjonowanie następnej generacji                  |
| PCR    | <i>polymerase chain reaction</i>                       | reakcja łańcuchowej polimerazy                        |
| PFS    | <i>progression-free survival</i>                       | czas przeżycia wolny od progresji                     |
| PP     | <i>processed pseudogene</i>                            | przetworzony pseudogen                                |
| pz     | <i>base pair</i>                                       | para zasad  |
| RNA    | <i>ribonucleic acid</i>                                | kwasy rybonukleinowe                                  |
| RNAseq | <i>RNA-sequencing</i>                                  | sekwencjonowanie transkryptomu                        |
| RT-PCR | <i>reverse transcription polymerase chain reaction</i> | reakcja łańcuchowa polimerazy z odwrotną transkrypcją |
| SH1-3  | <i>Src-homology 1-3 domains</i>                        | domeny homologiczne 1-3 kinazy Src                    |
| SNV    | <i>single nucleotide variant</i>                       | wariant pojedynczego nukleotydu                       |
| TKI    | <i>tyrosine kinase inhibitors</i>                      | inhibitory kinazy tyrozynowej                         |
| TFR    | <i>treatment-free remission</i>                        | remisja wolna od leczenia                             |
| TSD    | <i>target site duplications</i>                        | duplikacja miejsca insercji/ miejsca docelowego       |
| VAF    | <i>variant allele frequency</i>                        | częstość allelu alternatywnego                        |
| VUS    | <i>variant of uncertain significance</i>               | wariant o nieokreślonej patogenności                  |
| WHO    | <i>World Health Organization</i>                       | Światowa Organizacja Zdrowia                          |
| ΨSMAD4 | <i>processed pseudogene SMAD4</i>                      | przetworzony pseudogen SMAD4                          |

## STRESZCZENIE

**Charakterystyka nowych zmian genetycznych w przewlekłej białaczce szpikowej u chorych w różnych stadiach choroby przy zastosowaniu sekwencjonowania następnej generacji.**

Przewlekła białaczka szpikowa (CML, ang. *chronic myeloid leukemia*) jest pierwszym nowotworem, którego rozwój powiązano ze zmianą genetyczną, translokacją chromosomową t(9;22)(q34;q11) prowadzącą do powstania chromosomu Filadelfia z onkogenem *BCR::ABL1*. Inhibitory kinaz tyrozynowych nakierowane na *BCR::ABL1* stanowią skuteczną metodę leczenia CML, jednak u części chorych obserwowana jest oporność na stosowaną terapię. Mechanizmy oporności dzieli się na zależne i niezależne od *BCR::ABL1*. Sekwencjonowanie następnej generacji (NGS, ang. *next-generation sequencing*) umożliwia identyfikację zmian genetycznych istotnych w patogenezie i progresji wielu chorób, w tym chorób nowotworowych. Przedmiotem niniejszej pracy doktorskiej była charakterystyka nowych zmian genetycznych u chorych na CML z niekorzystnym przebiegiem choroby. W tym celu sekwencjonowaniu NGS poddano 45 próbek pochodzących od chorych w fazie blastycznej (BP, ang. *blast phase*) oraz próbki sprzed progresji CML, które były dostępne dla 25 chorych. Analiza wariantów w 193 genach istotnych w patogenezie nowotworów układu krwiotwórczego oraz onkogenezie wykazała obecność zmian o potencjalnym znaczeniu prognostycznym i predykcyjnym. Zidentyfikowano zmiany punktowe i małe insercje-delecje w genach często zmutowanych w nowotworach układu krwiotwórczego takich jak: *ABL1*, *ASXL1*, *RUNX1*, *DNMT3A* i *TP53*. Zmiany liczby kopii (dotyczące genów *IKZF1*, *CDKN2A*, *CDKN2B*, *BCR::ABL1*) oraz dodatkowe aberracje chromosomowe (m.in. monosomia chromosomu 7 oraz złożony kariotyp) obserwowane były odpowiednio wyłącznie w próbkach z fazy BP oraz zarówno w fazie BP jak i przed progresją CML. Analiza porównawcza patogennych wariantów i zmian liczby kopii w 25 sparowanych próbkach sprzed i po progresji CML pozwoliła na identyfikację 7 wzorców dynamiki zmian genetycznych obserwowanych w wyniku nabywania, utrzymywania się i zanikania aberracji. Wariantem zarodkowym, dotychczas nieopisanym w nowotworach był przetworzony pseudogen *SMAD4* ( $\Psi$ *SMAD4*), którego obecność zidentyfikowano w 12% sparowanych próbek (3/25). Poszerzona analiza wykazała istotny wzrost częstości jego występowania w CML w porównaniu do częstości obserwowanej w populacji polskiej,

co więcej wzrost ten nie był obserwowany w innych typach białaczek tj. przewlekłej białaczce limfocytowej, ostrej białaczce szpikowej oraz ostrej białaczce limfoblastycznej. Analiza RNAseq wykazała, że  $\Psi$ SMAD4 nie ulega transkrypcji i tym samym nie reguluje ekspresji genu rodzicielskiego, ani innych, niepowiązanych genów. Poza obecnością  $\Psi$ SMAD4, profil wykrytych zmian u chorych z pseudogenem jest zbliżony do typowego profilu obserwowanego w CML. Na podstawie przeprowadzonych badań wydaje się więc, że  $\Psi$ SMAD4 jest wariantem zarodkowym niemającym bezpośredniego wpływu na progresję CML, a określenie jego roli w patogenezie CML wymaga dalszych badań.

## ABSTRACT

### **Characterization of new genetic aberrations by next-generation sequencing in patients with chronic myeloid leukemia at different stages of the disease.**

Chronic myeloid leukemia (CML) is a myeloproliferative neoplasm characterized by a reciprocal chromosomal translocation  $t(9;22)(9q34;q11)$ , resulting in the formation of the Philadelphia chromosome encoding the *BCR::ABL1* fusion oncogene. Despite successful therapy with tyrosine kinase inhibitors, some patients develop drug resistance that may be *BCR::ABL1*-dependent and independent. Next-generation sequencing (NGS) facilitates the search for genetic aberrations leading to the pathogenesis and progression of many diseases, including cancer. Hence, the goal of this dissertation was to characterize new genetic aberrations in CML patients with an unfavorable course of the disease. For this purpose, 45 samples from patients with CML progression to blast phase (CML-BP) and the corresponding samples, taken before the progression of CML available for 25 patients were sequenced using NGS. Mutation analysis of 193 genes related to oncogenesis and hematological malignancies revealed the presence of variants with potential prognostic and/or predictive significance. Point mutations and small insertions-deletions have been identified in genes frequently mutated in hematologic malignancies, such as *ABL1*, *ASXL1*, *RUNX1*, *DNMT3A* and *TP53*. Copy number variations (for *IKZF1*, *CDKN2A*, *CDKN2B*, *BCR::ABL1* genes) were observed only in CML-BP samples, while additional chromosomal abnormalities (such as monosomy of chromosome 7 and complex karyotype) were observed in CML-BP as well as in samples taken before progression of CML. Comparative analysis of point mutations, small insertions-deletions and copy number changes in 25 paired samples revealed 7 patterns of mutation dynamics resulting from mutations acquisition, persistence and clearance. In 12% of the paired samples (3/25), the germline processed pseudogene *SMAD4* ( $\Psi$ *SMAD4*) was identified, which has not yet been described in CML or other cancers. The screening study showed a significantly higher frequency of  $\Psi$ *SMAD4* in CML, but not in other types of leukemia, compared to the general Polish population. RNAseq analysis showed that  $\Psi$ *SMAD4* is not transcribed and therefore does not regulate expression of the parental gene or other unrelated genes. Except for the presence of  $\Psi$ *SMAD4*, the mutation profiles of patients with and without the pseudogene are similar.

Thus, it appears that  $\Psi$ *SMAD4* is a germline variant with no direct impact on CML progression, but its role in the pathogenesis of CML requires further research.



# 1. WSTĘP

## 1.1. Przewlekła białaczka szpikowa

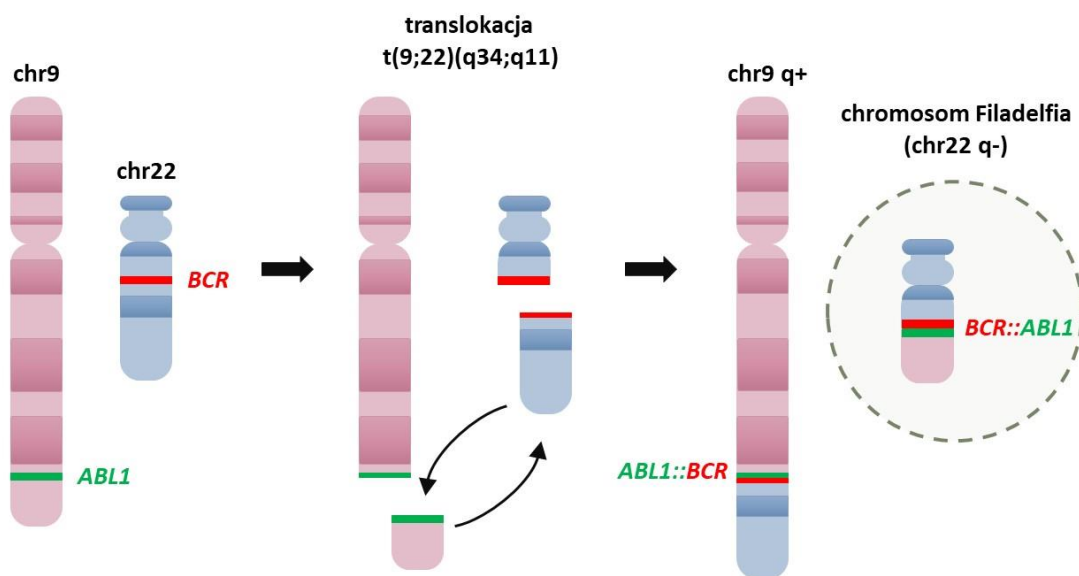
Według bazy danych GLOBOCAN, białaczki zajęły 15 miejsce na liście nowotworów pod względem liczby zachorowań na świecie w 2020 roku [1]. Przewlekła białaczka szpikowa (CML, ang. *chronic myeloid leukemia*) stanowi około 7% wszystkich białaczek diagnozowanych na świecie (2-13% w zależności od regionu) [2] a jej rozwój związany jest z ekspansją klonalną krwiotwórczej komórki macierzystej szpiku. Zwiększony potencjał proliferacyjny komórek CML jest efektem działania onkogenu *BCR::ABL1* powstałego w wyniku translokacji chromosomowej t(9;22)(q34;q11) [3]. CML jest chorobą modelową, w której po raz pierwszy rozwój nowotworu powiązано ze zmianą genetyczną [4], jak również po raz pierwszy z sukcesem zastosowano terapię celowaną z użyciem inhibitorów kinazy tyrozynowej *BCR::ABL1* [5].

### 1.1.1. Epidemiologia i etiopatogeneza

Aktualna klasyfikacja Światowej Organizacji Zdrowia (WHO, ang. *World Health Organization*) zalicza przewlekłą białaczkę szpikową do nowotworów mieloproliferacyjnych (MPN, ang. *myeloproliferative neoplasms*) [6]. CML jest chorobą nabytą, której etiologia jest nieznana, a jedynym potwierdzonym czynnikiem ryzyka powiązanim z rozwojem CML jest promieniowanie jonizujące [7]. Szacuje się, że zachorowalność na CML wynosi około 1 na 100 tys. osób, a choroba częściej dotyka mężczyzn niż kobiet (1,3 : 1). Mediana wieku w chwili rozpoznania wynosi 57 lat, chorzy powyżej 70 r.ż. stanowią więcej niż 20% wszystkich przypadków, a dzieci i młodzież poniżej 18 r.ż. mniej niż 5%. Chorobowość na CML w Europie oszacowano na 10-12 / 100 tys. osób, lecz liczba ta sukcesywnie wzrasta ze względu na poprawę jakości leczenia i przeżywalności pacjentów [8-10]. Określenie danych epidemiologicznych dla CML w populacji polskiej jest utrudnione ze względu na to, że Krajowy Rejestr Nowotworów (KRN) podaje ogólne dane dla „białaczek szpikowych”, do których oprócz CML wliczają się również ostra białaczka szpikowa (AML, ang. *acute myeloid leukemia*) i atypowe białaczki szpikowe bez chromosomu Filadelfia. Według KRN w 2018 roku zachorowalność na białaczki szpikowe wynosiła 3,1 / 100 tys. osób [11], dostępne są jednak dane pochodzące z Narodowego Funduszu Zdrowia, które wskazują,

że zachorowalność i chorobowość na CML w populacji polskiej wynosi odpowiednio 1 i 7,2 / 100 tys. osób [12].

Pierwsze opisy CML pochodzą z pierwszej połowy XIX wieku, kiedy to u chorych zaobserwowano powiększone śledziona i wątroby oraz podwyższoną liczbę leukocytów. W 1960 r. Peter Nowell i David Hungerford, pracujący w Filadelfii, opisali mały, akrocentryczny chromosom, występujący w komórkach krwi chorych na CML, który nazwano chromosomem Filadelfia (chromosom Ph, ang. *Philadelphia chromosome*). Jednak to Janet Rowley, 13 lat później, wykazała, że chromosom Ph jest wynikiem zrównoważonej translokacji chromosomów 9 i 22 [13, 14]. Obecnie wiadomo, że chromosom Ph powstaje w wyniku translokacji  $t(9;22)(q34;q11)$  poprzez przeniesienie fragmentu genu kodującego kinazę ABL1 z chromosomu 9 na chromosom 22 i jego fuzji z fragmentem genu BCR, prowadzącej do powstania onkogenu BCR::ABL1 (rycina 1).



**Rycina 1.** Schemat przedstawiający translokację  $t(9;22)(q34;q11)$  skutkującą powstaniem chromosomu Filadelfia z genem fuzyjnym BCR::ABL1. Chr9 — chromosom 9, chr22 — chromosom 22, chr9 q+ — wydłużony chromosom 9, chr22 q- — skrócony chromosom 22.

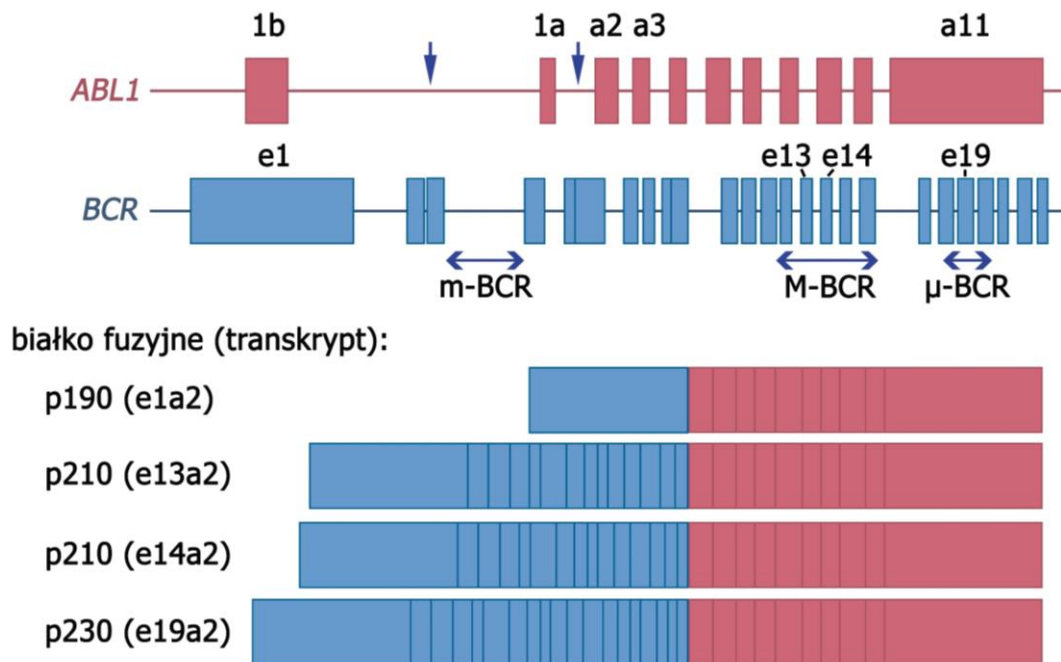
Gen ABL1 (*ABL proto-oncogene 1, non-receptor tyrosine kinase*) koduje niereceptorową kinazę tyrozynową, będącą białkiem o masie 145 kDa. ABL1 składa się z 11 egzonów i występuje w 2 izoformach powstałych w wyniku alternatywnego składania pierwszego egzonu (1a lub 1b). Aktywność kinazy tyrozynowej ABL1 jest zależna od trzech domen funkcyjnych, zlokalizowanych na N-końcu: SH1, SH2 i SH3 (ang. *Src-homology 1-3 domains*) występujących zarówno w prawidłowym białku, jak

również w białku fuzyjnym BCR::ABL1. Domena SH1 pełni funkcję kinazy tyrozynowej, która fosforyluje różne białka sygnałowe i adaptorowe, podczas gdy domena SH2 umożliwia interakcję z innymi białkami zawierającymi ufosforylowaną tyrozynę i odgrywa kluczową rolę w szlaku przekazywania sygnału. Domena SH3 bierze udział w mechanizmie autoinhibicji BCR::ABL1 poprzez negatywną regulację aktywności kinazowej domeny SH1. Białko ABL1 może przemieszczać się między jądrem a cytoplazmą, gdzie może wiązać, odpowiednio, DNA lub aktyne. Jądrowe białko ABL1 bierze udział w regulacji cyklu komórkowego i odpowiedzi komórkowej na stres genotoksyczny. Z kolei forma cytoplazmatyczna ABL1 jest zaangażowana w modyfikację struktury cytoszkieletu aktynowego oraz kontakt z innymi białkami wiążącymi się z cytoszkieletem [3, 15].

Gen *BCR* (*Breakpoint cluster region*) składa się z 23 egzonów i koduje białko o masie 160 kDa. BCR jest kinazą serynowo-treoninową biorącą udział w przekazywaniu sygnałów w komórkach (poprzez fosforylację i wiązanie guanozynotrifosforanu) oraz w naprawie uszkodzonego DNA. Podobnie jak w przypadku białka ABL1, na N-końcu znajdują się trzy domeny funkcyjne występujące zarówno w prawidłowym białku, jak i w białku fuzyjnym BCR::ABL1; domena kinazowa odpowiedzialna za autofosforylację oraz fosforylację innych białek, domena wiążąca SH2, łącząca się z domenami SH2 innych białek (w tym z białkiem ABL1), a także domena oligomeryzacyjna umożliwiająca wiązanie się białek ze sobą oraz pełniąca kluczową rolę w transformacji nowotworowej z udziałem onkogenu *BCR::ABL1* [3].

Gen fuzyjny *BCR::ABL1* koduje konstytutywnie aktywną kinazę tyrozynową, która wpływa na wiele ścieżek sygnalizacyjnych, powodując leukemogenezę. W trakcie tworzenia się *BCR::ABL1* pęknięcia w genie *ABL1* mogą wystąpić w obrębie intronów między egzonami 1b i 1a lub 1a i 2. Niezależnie od punktu pęknięcia *ABL1*, egzony 1a i 1b są zawsze wycinane, a z genem *BCR* łączą się egzony 2-11. W *BCR* scharakteryzowano trzy regiony częstych złamań: M-BCR (ang. *major cluster breakpoint region*), m-BCR (ang. *minor cluster breakpoint region*) i  $\mu$ -BCR (ang. *micro cluster breakpoint region*), odpowiadające trzem białkom fuzyjnym oznaczonym odpowiednio jako p210 BCR::ABL1, p190 BCR::ABL1 i p230 BCR::ABL1 (rycina 2). Transkrypty *BCR::ABL1* kodują białka fuzyjne o różnej masie co może prowadzić do różnic w fenotypie choroby. Transkrypty e13a2 (dawniej b2a2) i e14a2 (dawniej b3a2) kodują białko o masie 210 kDa występujące w ponad 95% wszystkich przypadków CML [16]. Z kolei transkrypt e1a2 koduje białko o masie 190 kDa charakterystyczne

(>80% przypadków) dla ostrej białaczki limfoblastycznej. Białko p190 BCR::ABL1 występuje u około 1% chorych na CML, u których częściej obserwuje się monocytozę w fazie przewlekłej, a rozpoznanie choroby następuje w podeszłym wieku, często w fazie zaawansowanej [17]. U niewielkiej części pacjentów z CML występuje rzadki transkrypt e19a2 kodujący białko p230 BCR::ABL1, którego obecność często wiąże się z wystąpieniem objawów charakterystycznych dla przewlekłej białaczki neutrofilowej [16].



**Rycina 2. Schematyczne przedstawienie struktury genów *ABL1* i *BCR* oraz transkryptów *BCR::ABL1* w białaczkach.** Niebieskimi strzałkami zaznaczono miejsca pęknięć w obrębie genów *ABL1* i *BCR* (na podstawie [2]).

Kinaza tyrozynowa BCR::ABL1 bierze udział w transformacji nowotworowej na drodze trzech mechanizmów: aktywacji przekazywania sygnałów w różnych szlakach sygnałowych, zahamowania apoptozy oraz upośledzenia adhezji komórek białaczkowych do komórek podścieliska szpiku [18]. Białko fuzyjne BCR::ABL1 aktywuje wiele szlaków sygnałowych takich jak: MAPK/ERK, JAK2/STAT, PI3K-AKT-mTOR, co prowadzi do zaburzenia podstawowych procesów zachodzących w komórce poprzez aktywację proliferacji, różnicowania i przeżycia komórek białaczkowych [19-21]. Ponadto, zmniejszona adhezja białaczkowych komórek progenitorowych do zrębu szpiku kostnego ułatwia ich uwalnianie do krwi [22].

### 1.1.2. Obraz kliniczny i przebieg choroby

Według najnowszej klasyfikacji WHO z 2022 roku wyróżnia się dwie fazy CML: początkową, najczęściej łagodną klinicznie fazę przewlekłą (CP, ang. *chronic phase*) i fazę ostrą nazywaną fazą blastyczną (dawniej kryzą blastyczną) (BP, ang. *blast phase*) [6]. Poprzednie edycje klasyfikacji WHO uwzględniały również trzecią fazę, którą czasami diagnozowano pomiędzy fazami CP i BP- fazę akceleracji (AP, ang. *acceleration phase*). Ponad 90% chorych diagnozowanych jest w fazie CP, a u prawie połowy z nich początek choroby ma charakter bezobjawowy- do rozpoznania CML dochodzi w sposób przypadkowy, najczęściej podczas rutynowych badań morfologii krwi. U pozostałych pacjentów mogą występować niespecyficzne objawy będące głównie wynikiem anemii i powiększenia śledziony takie jak: zmęczenie, utrata wagi, uczucie pełności lub dyskomfortu w jamie brzusznej, czy krwawienie związane z małopłytkowością [23, 24].

Nieleczona faza CP prowadzi do progresji choroby przeważnie w przeciągu 3-5 lat [25]. Mechanizm transformacji CML z fazy CP do BP nie jest w pełni poznany, najczęściej wiąże się go z niestabilnością genetyczną obserwowaną w komórkach białaczkowych [26]. Ekspresja i aktywność kinazy BCR::*ABL1* odgrywają rolę w regulacji długości telomerów, których skracanie ma bezpośredni wpływ na wzrost niestabilności genetycznej [27]. Obserwowana w komórkach białaczkowych niestabilność genetyczna prowadzi do ewolucji klonalnej, w trakcie której mogą pojawiać się nowe, wcześniej nieobecne, dodatkowe aberracje chromosomowe (ACA, ang. *additional chromosomal abnormalities*), będące negatywnym czynnikiem prognostycznym. Częstość występowania ACA wzrasta z około 10% w fazie CP na etapie diagnozy do ponad 80% w fazie BP [28].

### 1.1.3. Diagnostyka

W większości przypadków CML, diagnozę można postawić na podstawie charakterystycznej liczebności leukocytów i ich zróżnicowania (w rozmazie krwi obserwuje się wczesne postaci komórek układu granulocytarnego). Do potwierdzenia rozpoznania CML wymagany jest aspirat szpiku kostnego używany do badania kariotypu celem potwierdzenia obecności chromosomu Ph (konwencjonalne badanie cytogenetyczne lub fluorescencyjna hybrydyzacja *in situ*), jak również do oceny mielogramu pozwalającego określić fazę choroby [10]. Obecność i rodzaj transkryptu BCR::*ABL1* w komórkach krwi obwodowej lub szpiku kostnego jest oceniany metodą

RT-PCR (ang. *reverse transcription polymerase chain reaction*) [10, 29]. Rozpoznanie fazy BP możliwe jest na podstawie klinicznych i hematologicznych kryteriów, ustalanych przez organizacje takie jak WHO czy ELN (ang. *European LeukemiaNet*), opierających się m.in. o liczbę leukocytów, czy odsetek blastów i bazofili. Kryteria ELN są stosowane we wszystkich badaniach klinicznych dotyczących skuteczności inhibitorów kinazy tyrozynowej (TKI, ang. *tyrosine kinase inhibitors*) i są zalecane jako kryteria podstawowe przy podejmowaniu decyzji terapeutycznych [29].

Najważniejszym czynnikiem rokowniczym w CML jest poziom transkryptu *BCR::ABL1* i czas do uzyskania głębokiej odpowiedzi molekularnej. W chwili rozpoznania dodatkową pomocą są 4 skale prognostyczne: Sokala, EURO, EUTOS (ang. *European Treatment and Outcome Study*) i ELTS (ang. *EUTOS Long-Term Survival Score*) oparte na ocenie wielkości śledziony i wybranych parametrów w rozmazie krwi obwodowej, takich jak liczba płytek krwi i odsetek bazofili [10, 29].

#### **1.1.4. Leczenie**

Zatwierdzenie w 2001 r. przez Amerykańską Agencję ds. Żywności i Leków (FDA, ang. *Food and Drug Administration*) imatynibu, inhibitora kinazy tyrozynowej, do leczenia CML zrewolucjonizowało schemat jej leczenia na tyle, że z choroby śmiertelnej stała się chorobą przewlekłą [30, 31].

Rekomendacje dotyczące diagnostyki i leczenia CML w Polsce wydała Polska Grupa ds. Białaczek Dorosłych PALG [32] oraz Polskie Towarzystwo Onkologii Klinicznej PTOK [18]. Zawarte w nich schematy są zgodne z europejskimi (ESMO, ang. *European Society for Medical Oncology*) i amerykańskimi (NCCN, ang. *National Comprehensive Cancer Network*) wytycznymi [10, 25, 29].

Wytyczne zarówno ESMO jak i NCCN sugerują zastosowanie w pierwszej linii leczenia jednego z trzech TKI pierwszej lub drugiej generacji: imatynibu, dasatynibu lub nilotynibu. Wybór konkretnego inhibitora jest ustalany indywidualnie dla każdego pacjenta na podstawie oceny ryzyka określonego przez skale prognostyczne, profilu toksyczności TKI, wieku pacjenta oraz występowania chorób współistniejących [25, 29]. Mechanizm działania TKI polega na wiązaniu się z nieufosforylowaną, nieaktywną formą kinazy *BCR::ABL1* i kompetycyjnym blokowaniu miejsca wiązania ATP zapobiegającemu aktywacji kinazy. W konsekwencji następuje zahamowanie autofosforylacji kinazy *BCR::ABL1* oraz fosforylacji jej substratów, zaangażowanych

w kontrolę cyklu komórkowego, adhezję komórek i organizację cytoszkieletu, a to z kolei prowadzi do zahamowania proliferacji i śmierci apoptotycznej komórek białaczkowych [9, 15].

Wyróżnia się trzy rodzaje oceny odpowiedzi na terapię, tj. odpowiedź hematologiczną (ocena elementów morfotycznych krwi), cytogenetyczną (określenie odsetka komórek z chromosomem Ph w stadium metafazy w szpiku) i molekularną (ilościowa ocena transkryptu *BCR::ABL1* w szpiku lub krwi). Tempo uzyskiwania odpowiedzi cytogenetycznej i molekularnej stanowi istotny czynnik prognostyczny w kontekście ryzyka progresji choroby oraz przeżycia całkowitego [5, 29].

Niepowodzenie w osiągnięciu odpowiedzi hematologicznej, cytogenetycznej i molekularnej po rozpoczęciu leczenia TKI nazywane jest opornością pierwotną. Natomiast mianem oporności wtórnej określa się utratę odpowiedzi na terapię TKI. Szacuje się, że około 25% pacjentów rozwinię oporność wtórną lub nietolerancję na TKI stosowany w pierwszej linii leczenia i zmieni TKI przynajmniej raz w ciągu życia [33]. Oporność na TKI może rozwinąć się w wyniku mechanizmów zależnych i niezależnych od *BCR::ABL1*. Mechanizmy niezależne od *BCR::ABL1* występują częściej w przypadku oporności pierwotnej i są związane m.in. z aktywacją innych kinaz tyrozynowych, biodostępnością leków i ich zmiennością farmakokinetyczną. Natomiast mechanizmy zależne od *BCR::ABL1*, obejmujące mutacje w domenie kinazowej *BCR::ABL1* i nadekspresję *BCR::ABL1*, częściej obserwowane są w przypadku oporności wtórnej [16, 34].

U pacjentów z objawami nietolerancji lub oporności na stosowane leczenie, w drugiej linii leczenia włączane są TKI drugiej generacji: dasatynib, nilotynib lub bosutynib. Wybór odpowiedniego inhibitora uzależniony jest od obecności i typu mutacji w domenie kinazowej *BCR::ABL1* [29]. Ponatynib jest inhibitorem trzeciej generacji i ze względu na zwiększone ryzyko wystąpienia tętniczej lub żyłnej choroby zakrzepowozatorowej jego stosowanie jest ograniczone wyłącznie do pacjentów z mutacją p.Thr315Ile oraz w przypadku oporności na inne inhibitory [35]. Od końca 2021 roku, nową opcją terapeutyczną dla pacjentów z mutacją p.Thr315Ile stał się również asciminib, nakierowany na inną część kinazy *BCR::ABL1* niż pozostałe TKI [36]. Niefarmakologiczną metodą pozwalającą na całkowite wyleczenie jest alogeniczny przeszczep komórek krwiotwórczych (allo-HCT, ang. *allogeneic hematopoietic cell transplantation*), jednak ze względu na ryzyko zgonu i liczne powikłania, opcja ta jest obecnie zalecana u niewielkiej grupy chorych z opornością na dostępne TKI [18, 29, 37].

Od kilkunastu lat na całym świecie u wybranych chorych, którzy uzyskali pełną remisję na poziomie molekularnym trwającą minimum 2 lata, praktykuje się całkowite odstawienie leczenia. Przeprowadzone dotychczas badania wskazują na możliwość uzyskania remisji wolnej od leczenia (TFR, ang. *treatment-free remission*) odpowiadającej stabilnej, głębokiej odpowiedzi molekularnej bez potrzeby leczenia TKI u ok. 40-60% chorych, a efektywność tego podejścia zależy od czasu trwania leczenia przed odstawieniem, czasu trwania głębokiej odpowiedzi molekularnej przed odstawieniem oraz oceny ryzyka według wskaźnika Sokala na etapie diagnozy CML. Co istotne, zdecydowana większość osób, u których stwierdzono nawrót choroby po odstawieniu leczenia, odzyskała odpowiedź molekularną po ponownym wdrożeniu leczenia TKI [38-40].

### 1.1.5. Dodatkowe mutacje

Obserwowane w ostatnich latach postępy w technologii wysokoprzepustowego sekwencjonowania następnej generacji (NGS, ang. *next-generation sequencing*) umożliwiły badanie nieprawidłowości genetycznych w poszczególnych subpopulacjach komórek i zapoczątkowały erę medycyny personalizowanej w leczeniu nowotworów [41, 42]. Profile mutacji somatycznych są obecnie wykorzystywane do diagnostyki, oceny ryzyka i rokowania oraz klasyfikacji do terapii celowanej praktycznie we wszystkich nowotworach, przykładowo w: nowotworach ośrodkowego układu nerwowego [43], raku jajnika [44], raku jelita grubego [45], jak również w nowotworach układu krwiotwórczego takich jak: AML [46], ostrej białaczki limfoblastycznej (ALL, ang. *acute lymphoblastic leukemia*) [47], MPN [48] i w zespołach mielodysplastycznych (MDS, ang. *myelodysplastic syndromes*) [49].

W 1981 roku Fialkow i in. [50] zasugerował, że nabycie chromosomu Ph może być niewystarczające do powstania CML, a w patogenezie tej białaczki biorą udział dodatkowe zdarzenia genetyczne. Teorię o wieloetapowym procesie transformacji nowotworowej prowadzącej do powstania CML wspierają doniesienia o obecności transkryptu *BCR::ABL1* u osób zdrowych [51].

Niestabilność genetyczna spowodowana przez *BCR::ABL1* prowadzi do nabycia dodatkowych zaburzeń genetycznych, od mutacji punktowych po ACA, bezpośrednio przyczyniających się do progresji choroby [26]. Do najczęstszych ACA zalicza się trisomię chromosomów 8 i 19, izochromosom 17q oraz dodatkowy chromosom Ph, rzadziej obserwowane są również trisomia chromosomów 21 i 17, monosomia



chromosomu 7 oraz inne zmiany. ACA występują częściej w przypadku kryzy mieloidalnej niż limfoidalnej [20, 52, 53].

Pojawiają się również doniesienia sugerujące wpływ mutacji powiązanych ze zjawiskiem hematopoezy klonalnej o nieokreślonym potencjale (CHIP, ang. *clonal hematopoiesis of indeterminate potential*) na proces rozwoju CML. Cechą charakterystyczną CHIP jest obecność mutacji somatycznych w komórkach krwiotwórczych, przy braku objawów nowotworów układu krwiotwórczego i braku cytopenii, a mutacje te dotyczą głównie genów *DNMT3A*, *TET2* i *ASXL1* oraz rzadziej *TP53*, *SF3B1*, czy *SRSF2*. Jak dotąd, rola CHIP w patogenezie CML nie została do końca wyjaśniona, jednak przypuszcza się, że może ona nieznacznie zwiększać ryzyko wystąpienia nowotworu hematologicznego [54, 55].

Mutacje w domenie kinazowej *BCR::ABL1* są przyczyną oporności na TKI, która z kolei prowadzi do progresji choroby [56]. Jednak u około połowy chorych, którzy nie odpowiadają na leczenie TKI nie wykrywa się tychże mutacji, co sugeruje istnienie innych ścieżek rozwoju oporności na TKI, niezależnych od mutacji w *BCR::ABL1* [57]. Do niedawna, wiele badań naukowych, poświęconych poszukiwaniu genetycznych przyczyn i mechanizmów progresji CML, ograniczonych było do analizy próbek z tylko jednej fazy choroby (CP lub BP) lub do analizy jedynie kilku-kilkunastu genów [52, 58-60]. Pojawiło się jednak kilka kompleksowych prac poświęconych profilowaniu mutacji somatycznych w różnych stadiach CML, opartych o wysokoprzepustowe sekwencjonowanie całego egzomu lub nawet genomu. Tabela 1 przedstawia podsumowanie największych takich badań opublikowanych w przeciągu ostatnich 6 lat. W badaniach tych wykonano analizy porównawcze zmian genetycznych obserwowanych w fazach CP, AP i BP, z których wynika, że genami najczęściej zmutowanymi w fazie BP są m.in. *ABL1* oraz geny zaangażowane w rearanżację chromatyny (*ASXL1*, *IKZF1*) i regulację transkrypcji (*RUNX1*), co może sugerować ich udział w progresji choroby. W czterech z sześciu tych prac, oprócz badania genomu, zawarto również wyniki z analizy transkryptomu lub metylomu [61-64].

**Tabela 1. Lista najistotniejszych badań naukowych opublikowanych w ostatnich latach, zawierających analizy porównawcze mutacji somatycznych w sparowanych próbkach z różnych faz CML. CP- faza przewlekła, AP- faza akceleracji, BP-faza blastyczna, WES-sekwencjonowanie całego egzomu, WGS- sekwencjonowanie całego genomu, RNAseq-sekwencjonowanie transkryptomu, CHIPseq- analiza interakcji białek z DNA wykorzystująca immunoprecypitację chromatyny oraz wysokoprzepustowe sekwencjonowanie**

| <b>Publikacja</b>            | <b>Metoda sekwencjonowania, liczba przebadanych próbek</b>                                  | <b>Najczęściej zmutowane geny</b>   | <b>Dodatkowe analizy</b>  |
|------------------------------|---|---|---|
| Branford i in., 2018 [61]    | WES, 38 (18 CP, 4 BP i 16 sparowanych próbek CP+BP)   | w fazie CP: <i>ASXL1</i> , <i>RUNX1</i> , <i>TP53</i> , <i>IKZF1</i> , <i>SETD1B</i><br>w fazie BP: <i>ABL1</i> , <i>RUNX1</i> , <i>ASXL1</i> , <i>IKZF1</i> , <i>BCORL1</i> , <i>IDH1</i>  | analiza zmian liczby kopii, fuzji genowych, RNAseq                          |
| Magistrioni i in., 2019 [65] | WES, 10 sparowanych próbek CP+BP  | tylko w fazie BP, ale nie w CP: <i>RUNX1</i> , <i>IKZF1</i> , <i>NRAS</i> , <i>ASXL1</i> , <i>ABL1</i> , <i>UBE2A</i>   | analiza dodatkowych próbek (41 CP i 24 BP) pod kątem mutacji w <i>UBE2A</i> |
| Adnan Awad i in., 2020 [62]  | WES lub celowane sekwencjonowanie 578 genów, 59 (40 CP, 16 AP/BP, 3 sparowane próbki CP+BP) | w fazie CP: <i>ASXL1</i> , <i>PRKDC</i> , <i>KIAA1549</i><br>w fazie BP: <i>ABL1</i> , <i>ASXL1</i> , <i>RUNX1</i> , <i>BCOR</i> , <i>JAK3</i>  | RNAseq, analiza fuzji genowych  |
| Ko i in., 2020 [63]          | WGS, 13 sparowanych próbek CP+BP<br>WES, 39 (2 AP, 37 BP)                                   | WGS, tylko w fazie BP: <i>ABL1</i> , <i>RUNX1</i> , <i>IKZF1</i><br>WES: <i>CDKN2A/B</i> , <i>PAX5</i> , <i>EZH2</i>  | analiza zmian liczby kopii, metylomu, RNAseq, CHIPseq                       |
| Adnan Awad i in., 2021 [64]  | WES, 8 (4 BP, 4 sparowane próbki CP+BP)   | <i>RUNX1</i> dodatkowo z: <i>PHF6</i> , <i>BCORL1</i> , <i>ABL1</i> , <i>IKZF1</i>  | RNAseq  |
| Ochi i in., 2021 [66]        | WES, 52 sparowane próbki CP+BP, celowane sekwencjonowanie 104 genów: 19 CP, 60 BP           | WES w fazie CP: <i>ASXL1</i> , <i>TET2</i><br>WES w fazie BP: <i>RUNX1</i> , <i>ABL1</i> , <i>ASXL1</i> , <i>BCOR/BCORL1</i> , <i>TP53</i> , <i>WT1</i> , <i>UBE2A</i> , <i>SETD1B</i> , <i>KLC2</i> , <i>NBEAL2</i> , <i>PTPN11</i> , <i>IKZF1</i> , <i>CREBBP</i> | analiza zmian liczby kopii  |

|  |  |  |  |
|--|--|--|--|
|  |  | celowane<br>sekwencjonowanie<br>w fazie BP: <i>RUNX1</i> ,<br><i>ASXL1</i> , <i>ABL1</i> , <i>IKZF1</i> ,<br><i>BCORL1</i> , <i>TP53</i> |  |
|--|--|--|--|

## 1.2. Pseudogeny

Termin „pseudogen” został po raz pierwszy użyty w 1977 r., kiedy to Jacq i in. [67] opisali niefunkcjonalną kopię genu kodującego cząsteczkę 5S rRNA u płatanu szponiastej (łac. *Xenopus laevis*), różniącą się od funkcjonalnego homologu skróconym końcem 5'. Pseudogeny opisywane są jako *loci* genomowe o sekwencji zbliżonej do funkcjonalnych genów, jednak uznawane za biologicznie niefunkcjonalne ze względu na występujące mutacje (głównie nonsensowne i zmiany ramki odczytu), zaburzające procesy transkrypcji lub translacji białka. Funkcjonalne homologi pseudogenów są określane mianem genów rodzicielskich. Z uwagi na zaburzoną zdolność ekspresji, pseudogeny początkowo nazywane były „śmieciowymi genami” (ang. *junk genes*). Obecnie, naukowcy przyznają jednak, że terminologia „śmieciowego DNA” jest daleka od prawdy z uwagi na pojawiające się badania naukowe wskazujące na funkcjonalną rolę niektórych pseudogenów [68-71]. Ich znaczenie wydaje się być szczególnie istotne w ewolucji wielu gatunków, gdzie ich rola została powiązana m.in. z: chondroplazją odpowiadającą za fenotypowo krótkie łapy u kilkunastu ras psów [69], czy inaktywacją genu *GULO*, niezbędnego w procesie syntezy endogennej witaminy C, wskutek kompensacji dietetycznej tej witaminy u niektórych gatunków, w tym ludzi, świnek morskich i nietoperzy [72].

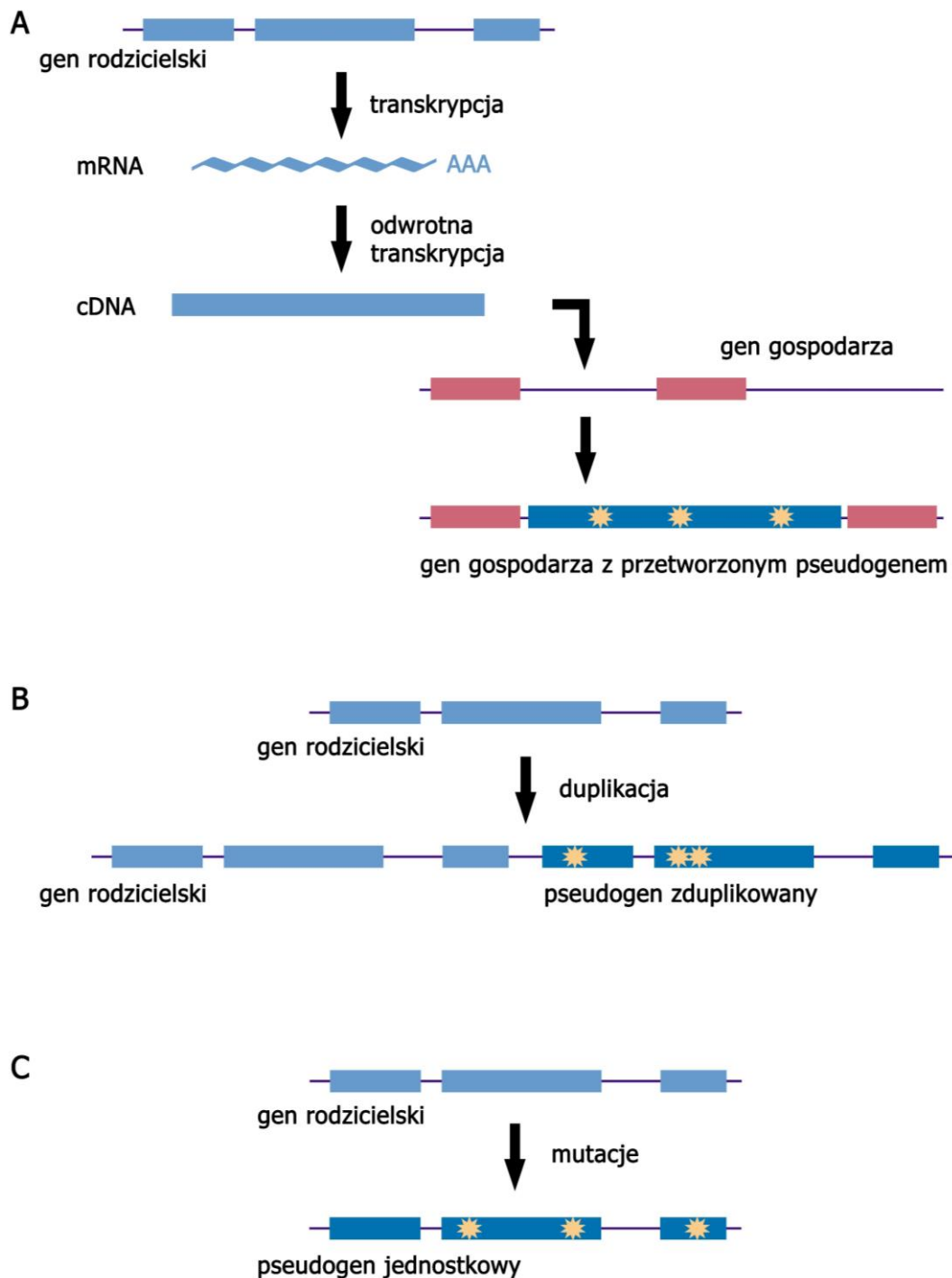
### 1.2.1. Klasyfikacja

Ze względu na sposób powstania wyróżnia się pseudogeny przetworzone (PP, ang. *processed pseudogenes*) i nieprzetworzone; te drugie z kolei dzieli się na zduplikowane i jednostkowe (rycina 3).

W ramach projektu GENCODE (v38, 05.2021), zajmującego się opisaniem funkcjonalnych elementów w ludzkim genomie, zidentyfikowano prawie 15 tys. pseudogenów, z czego ponad 72% (10 667 przypadków) stanowiły pseudogeny przetworzone [73]. Zdecydowana większość ludzkich genów (>23tys.) nie ma żadnego PP, ponad 2 tysiące genów posiada od 1 do 5 PP, ale istnieje 20 genów, kodujących m.in. białka rybosomalne, enzymy metaboliczne (np. *GAPDH*) i białka strukturalne

(np.  $\beta$ -aktynę), z których każdy ma >50 kopii PP [74]. Przetworzone pseudogeny powstają w wyniku retrotranspozycji z udziałem elementów LINE (ang. *Long Interspersed Nuclear Elements*). W procesie tym, część transkryptu informacyjnego RNA (mRNA, ang. *messenger RNA*) ulega spontanicznej odwrotnej transkrypcji do komplementarnego DNA (cDNA, ang. *complementary DNA*) a następnie integracji do genomowego DNA w nowym, przypadkowym miejscu, często zlokalizowanym na innym chromosomie niż gen rodzicielski (rycina 3A). Cechami charakterystycznymi PP wynikającymi z mechanizmu ich powstawania są: brak wszystkich lub większości intronów oraz obecność sekwencji poli-A na końcu 3'. Ponadto, PP są oflankowane na końcach 5' i 3' identycznymi wstawkami o długości ok. 5-20 nukleotydów, będącymi duplikacją miejsca insercji (TSD, ang. *target site duplications*). Mimo iż sekwencja PP jest zbliżona do sekwencji genu rodzicielskiego, pseudogeny te z reguły stają się niefunkcjonalne już w momencie integracji do genomu z powodu utraty niepodlegających transkrypcji sekwencji 5' promotora i innych elementów regulacyjnych (np. wzmacniaczy) oraz nabycia mutacji unieczyniających [68, 75, 76]. Opublikowano badania naukowe pokazujące, że PP zlokalizowane w intronach innych genów mogą „podczepiać się” do maszynerii transkrypcyjnej genu gospodarza i ulegają ekspresji dwa razy częściej niż PP znalezione w regionach międzygenowych [77]. Przetworzone pseudogeny mogą powstawać nie tylko w komórkach linii zarodkowej, ale również w komórkach somatycznych i tym samym stanowić nowy rodzaj mutacji przyczyniających się do rozwoju nowotworów [78].

Drugą co do częstości grupą są pseudogeny zduplikowane (ang. *duplicated pseudogenes*) powstające wskutek duplikacji genu rodzicielskiego (rycina 3B). W odróżnieniu od PP, pseudogeny zduplikowane są zwykle zlokalizowane w pobliżu genów rodzicielskich oraz posiadają sekwencje intronowe i regulatorowe. Utrata funkcjonalności tego typu pseudogenów wynika z nabycia mutacji unieczyniających takich jak mutacje nonsensowne, miejsc splicingowych i zmiany ramki odczytu [79].



**Rycina 3. Klasyfikacja pseudogenów ze względu na sposób ich powstania.** (A) Pseudogeny przetworzone powstają z informacyjnego RNA (mRNA), które podlega odwrotnej transkrypcji do komplementarnego DNA (cDNA), a następnie jest integrowane, ze zlokalizowanym często na innym chromosomie, innym genem - genem gospodarza (jak zaznaczono na rycinie) lub regionem międzygenowym, (B) Zdublikowane pseudogeny pochodzą z niedokładnych (nieukończonych lub wprowadzających mutacje) duplikacji genów, przy czym oryginalna kopia pozostaje w pełni funkcjonalna (gen rodzicielski) a zdublikowana kopia staje się pseudogenem, (C) Jednostkowe pseudogeny powstają wskutek różnych mutacji w genie, które znoszą jego transkrypcję lub translację zamieniając gen w pseudogen jednostkowy. Żółte gwiazdki na schemacie symbolizują mutacje zachodzące w pseudogenach (na podstawie [79]).

Pseudogeny jednostkowe (ang. *unitary pseudogenes*) są niewielką grupą pseudogenów, powstającą w wyniku spontanicznych mutacji skutkujących zahamowaniem procesu transkrypcji lub translacji. Mutacji nie poprzedza duplikacja genu, co sprawia, że pseudogeny jednostkowe nie mają swoich funkcjonalnych odpowiedników (genów rodzicielskich), charakterystycznych dla PP i pseudogenów zduplikowanych (rycina 3C). Pseudogeny jednostkowe są więc wynikiem wyłączenia funkcjonalnego genu, który w wyniku ewolucji przestał być niezbędny do przeżycia [79].

### 1.2.2. Funkcje pseudogenów w procesie karcynogenezy

Początkowo pseudogeny opisywane były jako niekodujące kopie genów rodzicielskich, przez co z góry określano je jako niefunkcjonalne i nieistotne. Ostatnie badania naukowe wskazują jednak, że część pseudogenów może pełnić wiele różnorodnych funkcji na poziomie DNA, RNA i białka, w procesach fizjologicznych jak i patologicznych, zarówno w stosunku do genów rodzicielskich jak również innych, niezwiązanych genów. Przykładowe funkcje pseudogenów w procesie nowotworzenia w oparciu o różne mechanizmy ich działania opisano poniżej.

#### 1.2.2.1. Funkcje na poziomie DNA

Na poziomie DNA pseudogeny mogą pełnić swoje funkcje poprzez konwersję genów i rekombinację homologiczną (głównie w odniesieniu do genów rodzicielskich) oraz egzonizację i mutacje insercyjne (dotyczące zwykle wpływu pseudogenu na gen gospodarza).

W mechanizmie konwersji genów, dochodzi do jednokierunkowego przeniesienia DNA pseudogenu do genu rodzicielskiego skutkującego zmianą sekwencji genu rodzicielskiego. Przykładowo, gen *CYP2A6* kodujący enzym wątrobowy metabolizujący m.in. nikotynę, posiada pseudogen *CYP2A7*, który nie koduje funkcjonalnego białka. Konwersja pseudogenu *CYP2A7* do genu *CYP2A6* skutkuje powstaniem allelu *CYP2A6\*1B* z miejscem polimorficznym na końcu 3'UTR stabilizującym mRNA, dzięki czemu *CYP2A6\*1B* wykazuje zwiększoną aktywność metabolizmu nikotyny *in vivo*, tym samym wpływając na zwiększoną konsumpcję papierosów oraz związane z nią ryzyko rozwoju raka płuc [70].

W procesie nazywanym egzonizacją dochodzi do nabycia przez pseudogen egzonu *de novo*. Kalikreina 3 (KLK3), znana również jako antygen specyficzny dla

prostata (ang. *prostate specific antigen*), jest powszechnie stosowanym markerem raka gruczołu krokowego. Wykryto pseudogen genu *KRK3- KRK31P*, w którym trzy z pięciu egzonów powstały wskutek egzonizacji *de novo* powtórzeń rozproszonych (ang. *interspersed repeats*). Transkrypcja *KRK3* i *KRK31P* jest regulowana przez androgeny, jednak w przeciwieństwie do *KRK3*, którego podwyższony poziom obserwuje się w surowicy pacjentów z rakiem prostaty, ekspresja *KRK31P* maleje w komórkach nowotworowych, co może sugerować jego niezależną rolę w patogenezie raka gruczołu krokowego [80].

W mechanizmie rekombinacji genetycznej może dochodzić do wzajemnej wymiany części sekwencji pseudogenu i genu rodzicielskiego. U dwóch rodzin, w których występowały przypadki raka piersi i jajnika, zaobserwowano rekombinację homologiczną intronu 2, pomiędzy genem *BRCA1* i jego pseudogenem  $\Psi$ *BRCA1*, prowadzącą do delecji 37 tys. par zasad obejmujących sekwencję promotorową i kodon startowy genu *BRCA1*, wskutek czego dochodziło do inaktywacji genu supresorowego [81].

#### 1.2.2.2. Funkcje na poziomie RNA

Chociaż tylko niewielka część pseudogenów podlega transkrypcji, transkrypty pseudogenów odgrywają różne role w potranskrypcyjnej regulacji ekspresji genów, gdzie przykładowo mogą funkcjonować na zasadzie endogennego małego interferującego RNA (endo-siRNA, ang. *endogenous small interfering RNA*) lub stanowić kompetycyjny endogenny RNA (ceRNA, ang. *competitive endogenous RNA*) konkurujący o wiązanie z mikroRNA (miRNA, ang. *micro RNA*).

Niektóre transkrypty pseudogenów mogą zostać przetworzone w endo-siRNA za pomocą rybonukleazy DICER. Przyłączenie endo-siRNA do komplementarnego mRNA aktywuje kompleks RISC (ang. *RNA-induced silencing complex*), który degradowe mRNA prowadząc do obniżenia jego ekspresji w mechanizmie interferencji RNA. Przykładem pseudogenu działającego na zasadzie endo-siRNA jest  $\Psi$ *PPMIK* regulujący ekspresję genu rodzicielskiego *PPMIK* i genu *NEK8* w raku wątroby [82].

Cząsteczki miRNA wiążą się z częściowo komplementarnymi elementami rozpoznającymi miRNA (MRE, ang. *miRNA recognition elements*), zlokalizowanymi głównie na końcach 3'UTR docelowych genów, co prowadzi do aktywacji kompleksu RISC i obniżenia ekspresji tychże genów. W ostatnich latach wykryto nowy mechanizm regulacji ekspresji wykorzystujący ceRNA. ceRNA posiadają wspólne MRE z genami

docelowymi, przez co regulują wzajemnie swoją ekspresję, rywalizując o tę samą pulę miRNA [83]. Do grupy ceRNA zalicza się m.in. pseudogeny ze względu na wysokie podobieństwo do genów rodzicielskich, a co za tym idzie - posiadanie takich samych MRE. Badania nad rakiem prostaty pokazują, że mRNA pseudogenu *PTENP1* konkuruje z mRNA genu *PTEN* o wiązanie z miR-17, miR-19, miR-21, miR-26 i miR-214, co z kolei prowadzi do zmniejszenia supresji indukowanej miRNA i w efekcie do wzrostu poziomu mRNA *PTEN* w komórce [71].

### 1.2.2.3. Funkcje na poziomie białka

Z reguły, pseudogeny posiadają mutacje genetyczne zaburzające procesy transkrypcji lub translacji białka, jednak niewielka garstka przetworzonych pseudogenów zachowuje lub odzyskuje zdolność kodowania białka.

Niektóre badania sugerują, że pseudogeny mogą kodować w pełni funkcjonalne białka o identycznej funkcji jak ich rodzicielskie odpowiedniki, jednak o innym wzorze ekspresji. W raku piersi koneksyna Cx43 spowalnia wzrost komórek nowotworowych i uwrażliwia je na chemioterapię, przez co często dochodzi do zahamowania ekspresji tego białka. Wykryto pseudogen *ΨCx43* ulegający translacji do białka wysoce homologicznego do białka Cx43. Białko *ΨCx43* jest transkrybowane wyłącznie w liniach komórkowych raka piersi i tak jak Cx43 prowadzi do zahamowania wzrostu komórek [84].

Nie wszystkie pseudogeny ulegające translacji posiadają w pełni funkcjonalne białko. Pseudogen *BRAF* obserwowany u chorych z rakiem tarczycy posiada wiele kodonów stop uniemożliwiających powstanie w pełni funkcjonalnego białka. Pseudogen ten koduje jednak polipeptyd o wysokiej homologii do domeny CR1 dzikiego białka BRAF i oddziałuje z nim aktywując w ten sposób szlak MAPK, co w efekcie prowadzi do rozwoju nowotworu. Co ciekawe, pseudogen *BRAF* był częściej obserwowany w próbkach bez mutacji w genie *BRAF*, niż w tych z mutacją, co sugeruje, że oba te zdarzenia mogą mieć podobne znaczenie w rozwoju raka tarczycy [85].

Krótkie peptydy pochodzące z pseudogenów mogą stanowić źródło peptydów antygenowych. Peptydy antygenowe są prezentowane przez cząsteczki głównego układu zgodności tkankowej MHC (ang. *major histocompatibility complex*) klasy I, indukując odpowiedź cytotoksycznych limfocytów T. Dlatego też, krótkie oligopeptydy pochodzące z pseudogenów mogą przyczyniać się do stymulacji układu odpornościowego przeciwko komórkom nowotworowym, co zostało zaobserwowane



w przypadku mięsaka i peptydu pochodzącego z pseudogenu dehydrogenazy Hsd17b12 [86].

### 1.2.3. Pseudogeny w nowotworach układu krwiotwórczego

Z uwagi na niewielką liczbę badań naukowych, rola pseudogenów w patogenezie nowotworów układu krwiotwórczego nie jest do końca poznana. W badaniach *in vivo* w modelu mysim opisano pseudogeny Braf-rs1 [87] i HMGA1P7 [88], działające na zasadzie ceRNA konkurującego o wiązanie z miRNA, których ekspresja prowadziła do wzrostu ekspresji, odpowiednio, genu rodzicielskiego (Braf) i innych genów (regulujących procesy zapalne, genów szlaku NF-κB, czy ścieżek sygnałowych JAK/STAT3 i mTOR). W obu przypadkach, u myszy z ekspresją pseudogenów zaobserwowano powstanie nowotworu fenotypowo przypominającego ludzkiego chłoniaka rozlanego z dużych komórek B (ang. *diffuse large B cell lymphoma*). Z kolei w chłoniaku Hodgkina, wysoka ekspresja pseudogenu *DUSP5P1*, obserwowana jedynie w komórkach nowotworowych, jest związana z jednoczesnym obniżeniem ekspresji genu rodzicielskiego *DUSP5*, będącego negatywnym regulatorem rodziny kinaz regulowanych sygnałem zewnątrzkomórkowym (ERK, ang. *extracellular signal-regulated kinases*). Spadek ekspresji *DUSP5* zależny od pseudogenu prowadzi więc do aktywacji szlaku ERK, a następnie do obniżenia ekspresji czynnika proapoptotycznego *BCL2L11* i w konsekwencji zahamowania apoptozy w komórkach nowotworowych [89]. Właściwości pronowotworowe wykazuje również pseudogen *NANOGP8* [90]. Zahamowanie ekspresji tego pseudogenu za pomocą krótkiego RNA o strukturze spinki do włosów (ang. *short hairpin RNA*) w liniach komórkowych ostrej białaczki limfoblastycznej T-komórkowej prowadziło do zahamowania proliferacji, zahamowania cyklu komórkowego w fazie G0/G1 oraz indukcji apoptozy. Z kolei w AML, poziom ekspresji pseudogenów *BMI1P1* [91], *DUSP5P1* [92] i *POU5F1B* [93] może stanowić niezależny czynnik prognostyczny w zakresie czasu przeżycia całkowitego (ang. *overall survival*). Wiedza dotycząca roli pseudogenów w CML ogranicza się do jednej publikacji, w której zbadano wpływ pseudogenu *OCT4-PG1* na komórki FEPS, będące modelem oporności wielolekowej, wyprowadzone z linii komórkowej K562. Wyciszenie pseudogenu *OCT4-PG1* w komórkach FEPS skutkowało jednoczesnym spadkiem ekspresji genu *OCT-4*, będącego czynnikiem transkrypcyjnym transporterów ABC powiązanych z opornością wielolekową [94].

### 1.3. Gen i pseudogen *SMAD4*

Gen *SMAD4* (*SMAD family member 4*) zlokalizowany jest na chromosomie 18 (18q21.2) i składa się z 12 egzonów kodujących białko o masie 60 kDa. Białko *SMAD4* składa się z trzech głównych części obejmujących: N-kończącą domenę MH1, C-kończącą domenę MH2 oraz łącznika pomiędzy nimi zawierającego domenę SAD. Domena MH1 odpowiada za wiązanie DNA poprzez rozpoznawanie elementu wiążącego (sekwencji GTCTAGAC). Domena MH2 bierze udział w homo- i heterodimeryzacji oraz lokalizacji jądrowej białka *SMAD4*, a także, podobnie jak domena SAD, ma znaczenie w aktywacji transkrypcji. *SMAD4* odgrywa kluczową rolę w szlaku transformującego czynnika wzrostu nowotworów beta (TGF- $\beta$ , ang. *transforming growth factor beta*), regulującego wiele procesów biologicznych, takich jak wzrost komórek, różnicowanie, apoptoza, migracja oraz rozwój i progresja nowotworu. Aktywacja szlaku podstawowego TGF- $\beta$ , zależna od białek rodziny *SMAD*, zaczyna się od połączenia ligandu z receptorem typu II (T- $\beta$ R II), prowadzącego do fosforylacji i aktywacji białek poniżej w szlaku, w tym białek *SMAD2* i *SMAD3*. Białka te, po połączeniu ze *SMAD4*, tworzą heterooligomer, który przedostaje się do jądra komórkowego i tam bierze udział w regulacji transkrypcji genów docelowych. *SMAD4* jest genem supresorowym, którego homo- i heterozygotyczne delecje zostały po raz pierwszy opisane w gruczolakoraku trzustki, stąd alternatywna nazwa tego genu, *DPC4* (*deleted in pancreatic carcinoma, locus 4*), tłumaczona na język polski jako „usunięty w raku trzustki”. Utrata *SMAD4* rzadko prowadzi do bezpośredniego rozwoju nowotworu, częściej przyspiesza jego rozwój lub progresję zapoczątkowaną przez mutacje w innych genach. Somatyczne mutacje *SMAD4* obserwowane są w wielu nowotworach, najczęściej we wspomnianym wyżej raku trzustki oraz w raku jelita grubego, dróg żółciowych, żołądka czy prostaty [95, 96].

Przetworzony pseudogen *SMAD4* ( $\Psi$ *SMAD4*) został po raz pierwszy opisany w 2015 roku, kiedy podczas badania u osób z podejrzeniem zespołu polipowatości młodzieńczej odkryto duplikację genu *SMAD4*, będącą w rzeczywistości pseudogenem [97].  $\Psi$ *SMAD4* ulega integracji na chromosomie 9 w ostatnim intronie genu *SCAI*, a jego częstość występowania określa się na poziomie ok. 0,24%. W porównaniu do mRNA genu rodzicielskiego, sekwencja  $\Psi$ *SMAD4* pozostaje skrócona na końcach rejonów 5' i 3' niepodlegających translacji (5'UTR, 3'UTR, ang. 5', 3' *untranslated region*), ale jego sekwencja, odpowiadająca regionowi kodującemu (CDS, ang. *coding*

*DNA sequence*), pozostaje niezmienną, co może sugerować funkcjonalną rolę pseudogenu [98].

## 2. ZAŁOŻENIA I CEL PRACY

U około 10–20% chorych na przewlekłą białaczkę szpikową dochodzi do oporności na standardowe leczenie z użyciem inhibitorów kinazy tyrozynowej. Cennej wiedzy na temat mechanizmów oporności, często niezależnych od *BCR::ABL1*, może dostarczyć analiza mutacji w sparowanych próbkach pobranych od chorych z niekorzystnym przebiegiem choroby, jednak liczba takich badań pozostaje wciąż ograniczona.

Głównym celem niniejszej pracy była kompleksowa analiza profilu genetycznego komórek CML od chorych z niekorzystnym przebiegiem choroby i poszukiwanie nowych zmian genetycznych, mogących mieć wpływ na progresję przewlekłej białaczki szpikowej.

Cele szczegółowe pracy obejmowały:

1. Zbadanie profilu mutacji genetycznych, zmian liczby kopii i dodatkowych aberracji chromosomowych w grupie 25 sparowanych próbek sprzed i po progresji CML oraz 20 próbek wyłącznie z fazy blastycznej, a następnie identyfikację znanych oraz wytypowanie nowych zmian genetycznych o potencjalnym znaczeniu prognostycznym lub predykcyjnym.
2. Porównanie zmian genetycznych w 25 sparowanych próbkach oraz ocenę dynamiki zachodzących zmian, mogących mieć wpływ na progresję CML.
3. Scharakteryzowanie na poziomie molekularnym przetworzonego pseudogenu *SMAD4* oraz ocenę jego roli w patogenezie CML.

## 3. MATERIAŁY I METODY

### 3.1. Materiały

#### 3.1.1. Materiał kliniczny

Materiał pierwotny do badań stanowiły krew, szpik, pelety leukocytów oraz cebulki włosowe. Dodatkowo, do oceny częstości występowania  $\Psi$ SMAD4 uzyskano materiał w postaci wyizolowanego wcześniej DNA. W celu anonimizacji chorych wszystkie próbki włączone do badań zostały odpowiednio zakodowane.

Materiał do badań pozyskiwany był w ramach współpracy z następującymi ośrodkami:

- Klinika Chorób Wewnętrznych Hematologii i Onkologii, Samodzielny Publiczny Centralny Szpital Kliniczny, Uniwersyteckie Centrum Kliniczne Warszawskiego Uniwersytetu Medycznego, Warszawa;
- Poradnia Hematologiczna oraz Zakład Diagnostyki Hematologicznej, Instytut Hematologii i Transfuzjologii, Warszawa;
- Oddział Kliniczny Hematologii oraz Zakład Diagnostyki Hematologicznej i Genetyki, Szpital Uniwersytecki w Krakowie, Kraków;
- Klinika Chorób Wewnętrznych i Hematologii, Wojskowy Instytut Medyczny, Centralny Szpital Kliniczny MON, Warszawa;
- Klinika Hematologii z Pododdziałem Chorób Naczyń, Uniwersytecki Szpital Kliniczny w Białymstoku, Białystok;
- Zakład Hematoonkologii Doświadczalnej, Uniwersytet Medyczny w Lublinie, Lublin;
- Katedra i Klinika Pediatrii, Hematologii i Onkologii Dziecięcej w Zabrze, Śląski Uniwersytet Medyczny, Katowice;
- Zakład Genetyki Medycznej, Warszawski Uniwersytet Medyczny, Warszawa;
- Department of Molecular Genetics, Institute of Hematology and Blood Transfusion, Praga, Czechy.

Na przeprowadzenie badań zostały wydane odpowiednie zgody Komisji Bioetycznych przy Warszawskim Uniwersytecie Medycznym (KB/227/2016), Instytucie Hematologii i Transfuzjologii w Warszawie (Opinie Nr. 31/2008 i 27/2016) oraz Wojskowym Instytucie Medycznym w Warszawie (Uchwała Nr 65/WIM/2016). Materiał od chorych uzyskiwano po wyrażeniu pisemnej, świadomej zgody na udział w badaniu, które było prowadzone zgodnie z zasadami etycznymi określonymi w Deklaracji Helsińskiej.

### 3.1.1.1. Charakterystyka grupy chorych

Podstawowe dane kliniczne chorych, dla których wykonano sekwencjonowanie wybranych genów metodą NGS przedstawia tabela 2.

**Tabela 2. Dane kliniczne chorych na CML włączonych do badania NGS.**

| <b>Parametr</b>  | <b>Wartość</b>   |
|--|------------------|
| liczba chorych   | 51               |
| wiek [mediana (zakres)] [lata]                                   | 52,5 (18 – 77)   |
| płeć [n, (%)]  |                  |
| kobiety  | 31 (61)          |
| mężczyźni  | 20 (39)          |
| faza CML w czasie diagnozy [n, (%)]                              |                  |
| CP   | 28 (55)          |
| AP   | 5 (10)           |
| BP   | 14 (27)          |
| brak danych  | 4 (8)            |
| typ kryzy blastycznej [n, (%)]                                   |                  |
| mieloidalny  | 23 (45)          |
| limfoidalny  | 7 (14)           |
| nie dotyczy (wyłącznie faza CP)                                  | 5 (10)           |
| brak danych  | 16 (31)          |
| liczba białych krwinek [mediana (zakres)] [tys/mm <sup>3</sup> ] | 140 (9,4 – 640)  |
| poziom hemoglobiny [mediana (zakres)] [g%]                       | 9,8 (6,2 – 15,5) |
| liczba płytek krwi [mediana (zakres)] [tys/mm <sup>3</sup> ]     | 275 (32 – 1291)  |
| splenomegalia [n, (%)]   |                  |
| tak  | 27 (53)          |
| nie  | 11 (22)          |
| brak danych  | 13 (25)          |
| wskaźnik Sokala [n, (%)]   |                  |
| niski  | 7 (14)           |
| wysoki   | 12 (24)          |
| nie określono (chorzy zdiagnozowani w fazie AP lub BP)           | 19 (37)          |
| brak danych  | 13 (25)          |

### 3.1.2. Odczynniki

- a) agaroz- Agarose LE, analytical grade (Promega), nr kat.: V3125;
- b) agaroz low melting- QA-Agarose TM low melting (MP Biomedicals), nr kat.: AGAL0050;
- c) alkohol etylowy bezwodny 99,8% CZDA ODCZ.FP (POCH), nr kat.: 396480111;
- d) alkohol izopropylowy cz.d.a. (Chempur), nr kat.: 117515002;
- e) bromek etydyny- Ethidium bromide solution (Sigma), nr kat.: 46067;
- f) bufor AE (10mMTris-Cl, 0.5mMEDTA, pH9.0), (Qiagen), nr kat.: 19077;

- g) bufor TBE, 5-krotnie stężony (445 mM Tris, 445 mM kwas borowy i 10 mM EDTA, pH=8.3);
- h) formamid- HiDi™ Formamide (Applied Biosystems), nr kat.: 4311320;
- i) kulki paramagnetyczne VAHTS DNA Clean Beads (Vazyme), nr kat.: N411-01;
- j) marker wielkości GeneRuler™ 50 bp DNA Ladder (Thermo Fisher Scientific), nr kat.: SM0372;
- k) marker wielkości GeneRuler™ 1kb Plus DNA Ladder (Thermo Fisher Scientific), nr kat.: SM1343;
- l) mieszanina deoksynukleotydów, dNTP- 100mM dNTP Set (Invitrogen), nr kat.: 10297-018;
- m) proteinaza K- Puregene Proteinase K (Qiagen), nr kat.: 158920;
- n) RNaza- RNase A Solution (Qiagen), nr kat.: 158924;
- o) woda- AccuGENE™ Molecular Biology Water (Lonza), nr kat.: BE51223;
- p) zestaw Agilent High Sensitivity DNA Kit (Agilent Technologies), nr kat.: 5067-4626;
- q) zestaw Agilent RNA 6000 Nano Kit (Agilent Technologies), nr kat.: 5067-1511;
- r) zestaw BigDye™ Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), nr kat.: 4336917;
- s) zestaw dsDNA High Sensitivity Assay Kit (Invitrogen), nr kat.: Q32851;
- t) zestaw Expand™ Long Template PCR System (Roche), nr kat.: 11681834001;
- u) zestaw Gentra Puregene Blood Kit (Qiagen), nr kat.: 158389;
- v) zestaw HotStarTaq® Plus DNA Polymerase (Qiagen), nr kat.:203605;
- w) zestaw KAPA HiFi HotStart ReadyMix (Roche), nr kat.: 7958935001;
- x) zestaw KAPA HyperPlus Library Preparation Kit (Roche), nr kat.: 07962428001;
- y) zestaw KAPA Lib Prep kit (Kapa Biosystems, Illumina), nr kat.: KK8201;
- z) zestaw Maxwell® FSC DNA IQ™ Casework Kit (Promega), nr kat.: AS1550;
- aa) zestaw Monarch® DNA Gel Extraction Kit (New England BioLabs), nr kat.: T1020S;
- bb) zestaw Nextera XT DNA Library Preparation Kit (Illumina), nr kat.: FC-131-1024;
- cc) zestaw Nextera XT Index Kit v2 Set D (Illumina), nr kat.: FC-131-2004;
- dd) zestaw QIAamp® DNA Blood Mini Kit (Qiagen), nr kat.: 51104;
- ee) zestaw QIAamp® RNA Blood Mini Kit (Qiagen), nr kat.: 52304;
- ff) zestaw SepCap Adapter Kit (Roche), nr kat.: 07141530001, 07141548001;

- gg) zestaw SeqCap HE-Oligo Kit (Roche), nr kat.: 06777287001, 06777317001;
- hh) zestaw SeqCap Hybridization and Wash Kit (Roche), nr kat.: 05634261001;
- ii) zestaw TruSeq RNA Library Preparation Kit v2, Set A, nr kat: RS-122-2001;
- jj) zestawy sond wzbogacających SeqCap EZ Choice (Roche NimbleGen), na zamówienie;

## **3.2. Metody**

Schemat przeprowadzonych badań prezentuje rycina 4.

### **3.2.1. Izolacja ludzkiego materiału genetycznego**

Materiał do badań stanowiły próbki krwi, szpiku, pelety leukocytów oraz cebulki włosowe, pobierane od chorych na CML, z których za pomocą komercyjnych zestawów izolowano DNA genomowe, DNA linii zarodkowej oraz całkowite RNA.

Izolację genomowego DNA z krwi lub szpiku wykonywano za pomocą komercyjnego zestawu QIAamp® DNA Blood Mini Kit (Qiagen) zgodnie z zaleceniami producenta. W skrócie, procedura polegała na lizie komórek, związaniu DNA do złoza na kolumnie, płukaniu i elucji DNA w buforze AE (Qiagen).

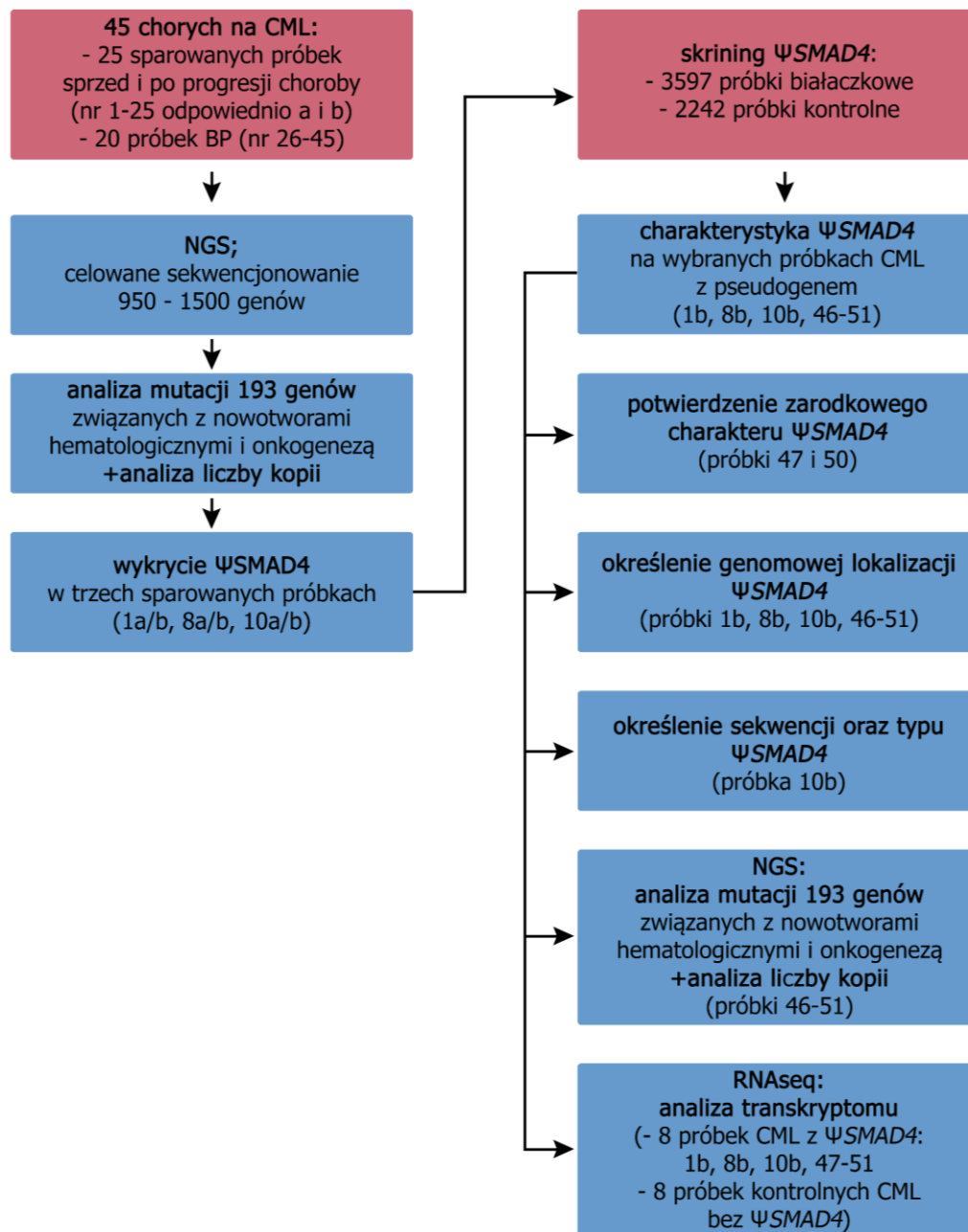
Genomowe DNA z peletu leukocytów izolowano w wykorzystaniem komercyjnego zestawu Gentra Puregene Blood Kit (Qiagen) według dołączonego protokołu z własnymi modyfikacjami. Pelety leukocytowe lizowano, a następnie trawiono białka (inkubacja całonocna) i RNA (inkubacja 1-godzinna), z użyciem odpowiednio proteiny K i RNazy (Qiagen). DNA wytrącano z roztworu za pomocą izopropanolu, płukano w 70% etanolu, suszono i zawieszano w buforze AE.

Izolację DNA linii zarodkowej z cebulek włosowych wykonano we współpracy z Zakładem Medycyny Sądowej Warszawskiego Uniwersytetu Medycznego z wykorzystaniem systemu Maxwell RSC (Promega). DNA wyizolowano za pomocą zestawu Maxwell® FSC DNA IQ™ Casework Kit (Promega) opartego na wiązaniu DNA do kulek paramagnetycznych. Procedurę automatycznej izolacji wykonano zgodnie z zaleceniami producenta, DNA zawieszono w buforze do elucji dołączonym do zestawu.

RNA z krwi lub szpiku izolowano za pomocą komercyjnego zestawu QIAamp® RNA Blood Mini Kit (Qiagen), według protokołu producenta. Procedura ta polegała na lizie erytrocytów, następnie lizie leukocytów, homogenizacji lizatów komórkowych, wiązaniu RNA do złoza na kolumnie, płukaniu i elucji RNA w wodzie (Lonza).



Do czasu dalszych analiz DNA i RNA przechowywano odpowiednio w +4°C i -80°C.



**Rycina 4. Schemat badań przeprowadzonych w niniejszej pracy doktorskiej.** Kolorem różowym oznaczono materiały a niebieskim - badania jakie wykonano. W badaniach dotyczących charakterystyki  $\Psi$ SMAD4, z uwagi na ograniczoną dostępność materiału do badań w nawiasach podano zakodowane nazwy próbek dla których wykonano poszczególne analizy.

### 3.2.2. Sekwencjonowanie DNA za pomocą metody NGS

DNA wyizolowane od chorych na CML poddawano celowanemu sekwencjonowaniu genów związanych z nowotworami układu krwiotwórczego i onkogenezą. Badanie wykonano dla 76 próbek pochodzących od 51 chorych ze zdiagnozowaną przewlekłą białaczką szpikową (25 sparowanych próbek pochodzących sprzed i po progresji choroby, 20 próbek BP i dodatkowe 6 próbek z  $\Psi$ SMAD4; w tym 5 z fazy CP i jedna z fazy AP).

#### 3.2.2.1. Przygotowanie bibliotek genomowych, wzbogacanie celowane i sekwencjonowanie

Biblioteki genomowe przygotowano z wysokiej jakości genomowego DNA (100-1000 ng) chorych na CML za pomocą zestawów KAPA HyperPlus Library Preparation Kit lub KAPA Lib Prep kit (Kapa Biosystems). Procedury wykonano według zaleceń producenta. Główne etapy konstrukcji biblioteki obejmowały: fragmentację genomowego DNA, naprawę końców, adenylację, ligację adaptorów w systemie 24-indeksowym (SepCap Adapter Kit, Roche), dwustronną selekcję wielkości fragmentów biblioteki za pomocą kulek paramagnetycznych VAHTS DNA Clean Beads (Vazyme) oraz amplifikację bibliotek za pomocą reakcji PCR z użyciem zestawu KAPA HiFi HotStart ReadyMix (Kapa Biosystems). W zależności od użytego zestawu do konstrukcji bibliotek, DNA poddawano fragmentacji mechanicznej z wykorzystaniem aparatu Covaris M220 (Covaris) lub fragmentacji enzymatycznej a jej czas wynosił odpowiednio 225 sekund lub 20 minut.

Celowane sekwencjonowanie wykonano używając 3 różne zestawy sond wzbogacających (SeqCap EZ Choice, Roche NimbleGen) wychwytyjących sekwencje kodujące około 950 (~7Mb), 1200 (~10Mb) lub 1500 (~8,5Mb) genów. Szczegóły dotyczące sekwencjonowania poszczególnych próbek przedstawiono w tabeli 3.

Po ocenie ilościowej i jakościowej biblioteki genomowe łączono w pule, zwykle o wielkości 24 bibliotek, a następnie poddawano 24-72 godzinnej hybrydyzacji z sondami. Po tym czasie, za pomocą magnetycznych kulek sprzężonych ze streptawidyną wychwytywano biotynylowane sondy zhybrydyzowane z DNA, a następnie poddawano je serii płukań (SeqCap Hybridization and Wash Kit, Roche) i amplifikacji za pomocą reakcji PCR (KAPA HiFi HotStart ReadyMix, Kapa Biosystems).

W trakcie procedury wykonywania bibliotek trzykrotnie dokonywano pomiaru stężenia DNA: przed przystąpieniem do procedury, po skonstruowaniu indywidualnych bibliotek genomowych oraz biblioteki wzbogaconej. Pomiary stężeń wykonywano za pomocą fluorymetru Qubit 2.0 (Life Technologies) z wykorzystaniem zestawu odczynników dsDNA High Sensitivity Assay Kit (Invitrogen). Ponadto, rozkład wielkości bibliotek genomowych oceniano za pomocą elektroforezy w 2% żelu agarozowym a rozkład wielkości biblioteki wzbogaconej - z wykorzystaniem aparatu Bioanalyser 2100 (Agilent Technologies) i zestawu Agilent DNA 1000 kit (Agilent Technologies).

Wzbogacone biblioteki były sekwencjonowane w trybie sparowanych końców (2×100 pz) na platformach Illumina HiSeq 1500 lub NovaSeq 6000.

**Tabela 3. Szczegóły celowanego sekwencjonowania wysokoprzepustowego dotyczące zastosowanych zestawów sond wzbogacających oraz statystyki sekwencjonowania.**

| <b>Kod próbki</b>   | <b>Liczba zsekwencjonowanych genów</b> | <b>Średnie pokrycie – zakres ; mediana</b> | <b>Odsetek zasad pokrytych min. 20 razy – zakres ; mediana</b> |
|---|--|--|--|
| 1b, 2a, 2b, 3a, 3b, 4a, 4b, 6a, 6b, 7a, 7b, 8b, 9a, 9b, 10a, 11a, 11b, 12a, 12b, 22a, 22b, 23a, 23b, 24a, 24b, 25a, 25b, 26, 28, 29, 30, 31, 32, 33, 34, 35, 43, 44 | 950 (~7Mb)                             | 51,3 – 276,8×<br>; 170,4×                  | 79,3 – 96,7% ;<br>95,3%  |
| 5a, 5b, 13a, 13b, 14a, 14b, 15a, 15b, 16a, 16b, 17a, 17b, 20a, 20b, 36, 37, 38, 45, 47, 48  | 1200 (~10Mb)                           | 117,4 –<br>177,7× ;<br>151,5×              | 94,2 – 97,8% ;<br>96,2%  |
| 1a, 8a, 10b, 18a, 18b, 19a, 19b, 21a, 21b, 27, 39, 40, 41, 42, 46, 49, 50, 51   | 1500 (~ 8,5Mb)                         | 96,1 – 254,7×<br>; 113,4×                  | 90,3 – 92,6% ;<br>91,4%  |

### 3.2.2.2. Obróbka surowych danych i adnotacja wariantów

Podstawową obróbkę surowych danych oraz identyfikację i adnotację wariantów genetycznych i zmian liczby kopii wykonali pracownicy Zakładu Genetyki Medycznej oraz Zakładu Biologii i Genetyki Nowotworów Warszawskiego Uniwersytetu Medycznego.

Obróbkę danych przeprowadzono w oparciu o wytyczne GATK (ang. *Genome Analysis Toolkit*) [99]. W skrócie, procedura ta obejmowała konwersję surowych danych i kontrolę ich jakości. Następnie za pomocą programu Trimmomatic [100] usuwano sekwencje adaptorowe oraz odczyty niskiej jakości a program BWA-MEM [101] wykorzystano do mapowania odczytów do genomu referencyjnego hg19. Usuwanie duplikatów i rekaliibrację jakości wyników wykonano za pomocą pakietów GATK i Picard, a identyfikację wariantów z użyciem m.in. HaplotypeCaller. Wykryte warianty zostały przekonwertowane do formatu MS Access do dalszej analizy manualnej.

Identyfikację zmian liczby kopii w próbkach poddanych sekwencjonowaniu NGS wykonano za pomocą programu CNVkit v0.9.5 [102]. Ustawienia programu pozostawiono domyślne, maksymalna wielkość analizowanych regionów (ang. *bins*) wynosiła do 400 nukleotydów. Identyfikacji zmian liczby kopii dokonywano na podstawie różnicy w pokryciu analizowanych regionów w próbkach badanych w stosunku do próbek referencyjnych, które stanowiły próbki DNA (białaczkowe i niebiałaczkowe) o znanej, pozbawionej znaczących zmian liczbie kopii. Regiony o podobnym odchyleniu od oczekiwanego pokrycia łączono w grupy w ramach tzw. segmentacji. W próbce o numerze 28 dokonano ręcznej korekty z uwagi na widoczną w wynikach, niepoprawną identyfikację regionów wykazujących normalną liczbę kopii. Wyniki przedstawiono w formie wykresów oraz tabel z bezwzględną liczbą kopii genów do dalszej analizy manualnej.

### 3.2.2.3. Analiza patogennych wariantów

Analizie poddano 193 geny istotne w patogenezie nowotworów układu krwiotwórczego oraz najczęściej zmutowane w nowotworach ludzkich protoonkogeny i geny supresorowe wspólne dla wszystkich trzech użytych paneli genowych (tabela 4). Do wizualizacji zmapowanych odczytów oraz weryfikacji wykrytych wariantów wykorzystano oprogramowanie IGV (*Integrative Genome Viewer*).

**Tabela 4. Lista 193 genów poddanych analizie.**

|               |               |               |               |                |                |                 |
|---------------|---------------|---------------|---------------|----------------|----------------|-----------------|
| <i>ABL1</i>   | <i>CCND1</i>  | <i>ETV6</i>   | <i>IKZF1</i>  | <i>MLLT10</i>  | <i>PHF6</i>    | <i>SMC1A</i>    |
| <i>AKAP9</i>  | <i>CCND2</i>  | <i>EZH2</i>   | <i>IKZF3</i>  | <i>MN1</i>     | <i>PIK3C2A</i> | <i>SMC3</i>     |
| <i>AKT1</i>   | <i>CCND3</i>  | <i>FAM46C</i> | <i>IRF4</i>   | <i>MPL</i>     | <i>PIK3CA</i>  | <i>SPEN</i>     |
| <i>APC</i>    | <i>CCNE1</i>  | <i>FANCA</i>  | <i>JAK1</i>   | <i>MSH2</i>    | <i>PIK3CB</i>  | <i>SRSF2</i>    |
| <i>ARID1A</i> | <i>CD274</i>  | <i>FANCB</i>  | <i>JAK2</i>   | <i>MSH6</i>    | <i>PIK3CD</i>  | <i>STAG2</i>    |
| <i>ARID2</i>  | <i>CD79B</i>  | <i>FANCC</i>  | <i>JAK3</i>   | <i>MTOR</i>    | <i>PIK3CG</i>  | <i>STAT2</i>    |
| <i>ARID4B</i> | <i>CDK4</i>   | <i>FANCD2</i> | <i>KDM5A</i>  | <i>MUTYH</i>   | <i>PIK3R1</i>  | <i>STAT3</i>    |
| <i>ASXL1</i>  | <i>CDK6</i>   | <i>FANCE</i>  | <i>KDM6A</i>  | <i>MYC</i>     | <i>PIK3R2</i>  | <i>STAT5B</i>   |
| <i>ATM</i>    | <i>CDKN1B</i> | <i>FANCF</i>  | <i>KDR</i>    | <i>MYD88</i>   | <i>PIM1</i>    | <i>STK11</i>    |
| <i>ATR</i>    | <i>CDKN2A</i> | <i>FANCG</i>  | <i>KIT</i>    | <i>MYH11</i>   | <i>PMS1</i>    | <i>SYNE2</i>    |
| <i>ATRX</i>   | <i>CDKN2B</i> | <i>FANCI</i>  | <i>KLHL6</i>  | <i>NF1</i>     | <i>PMS2</i>    | <i>TERT</i>     |
| <i>BAP1</i>   | <i>CDKN2C</i> | <i>FANCL</i>  | <i>KMT2A</i>  | <i>NF2</i>     | <i>POT1</i>    | <i>TET2</i>     |
| <i>BCL10</i>  | <i>CEBPA</i>  | <i>FANCM</i>  | <i>KMT2C</i>  | <i>NFE2L2</i>  | <i>PRDM1</i>   | <i>TGFBR2</i>   |
| <i>BCL2</i>   | <i>CHD2</i>   | <i>FAT1</i>   | <i>KMT2D</i>  | <i>NOTCH1</i>  | <i>PRDM9</i>   | <i>TNFAIP3</i>  |
| <i>BCL6</i>   | <i>CHEK2</i>  | <i>FAT3</i>   | <i>KRAS</i>   | <i>NOTCH2</i>  | <i>PTEN</i>    | <i>TNFRSF14</i> |
| <i>BCOR</i>   | <i>CREBBP</i> | <i>FAT4</i>   | <i>LRP1B</i>  | <i>NPM1</i>    | <i>PTPN11</i>  | <i>TP53</i>     |
| <i>BIRC2</i>  | <i>CSF3R</i>  | <i>FBXW7</i>  | <i>MALT1</i>  | <i>NRAS</i>    | <i>PTPRD</i>   | <i>TSC1</i>     |
| <i>BIRC3</i>  | <i>CTNNB1</i> | <i>FGFR1</i>  | <i>MAP2K1</i> | <i>NSD1</i>    | <i>PTPRJ</i>   | <i>TYK2</i>     |
| <i>BIRC5</i>  | <i>CYLD</i>   | <i>FGFR2</i>  | <i>MAP2K2</i> | <i>NTRK1</i>   | <i>RAD21</i>   | <i>U2AF1</i>    |
| <i>BRAF</i>   | <i>DAXX</i>   | <i>FGFR3</i>  | <i>MAP3K1</i> | <i>NTRK2</i>   | <i>RAD51B</i>  | <i>VHL</i>      |
| <i>BRCA1</i>  | <i>DDX3X</i>  | <i>FLT1</i>   | <i>MAPK1</i>  | <i>NUP214</i>  | <i>RAD51C</i>  | <i>WHSC1</i>    |
| <i>BRCA2</i>  | <i>DNMT3A</i> | <i>FLT3</i>   | <i>MAX</i>    | <i>PALB2</i>   | <i>RB1</i>     | <i>WT1</i>      |
| <i>BRD3</i>   | <i>EGFR</i>   | <i>FOXP1</i>  | <i>MCL1</i>   | <i>PAX5</i>    | <i>RUNX1</i>   | <i>XPO1</i>     |
| <i>BRIP1</i>  | <i>EGLN1</i>  | <i>GATA2</i>  | <i>MDM4</i>   | <i>PBRM1</i>   | <i>SETD2</i>   | <i>ZMYM3</i>    |
| <i>CARD11</i> | <i>EP300</i>  | <i>GNAS</i>   | <i>MECOM</i>  | <i>PCLO</i>    | <i>SF3B1</i>   | <i>ZRSR2</i>    |
| <i>CASP8</i>  | <i>EPCAM</i>  | <i>HRAS</i>   | <i>MED12</i>  | <i>PDE4DIP</i> | <i>SLX4</i>    |                 |
| <i>CBL</i>    | <i>ERBB2</i>  | <i>IDH1</i>   | <i>MET</i>    | <i>PDGFRA</i>  | <i>SMAD4</i>   |                 |
| <i>CBLB</i>   | <i>ERBB3</i>  | <i>IDH2</i>   | <i>MLH1</i>   | <i>PDGFRB</i>  | <i>SMARCA4</i> |                 |

W ramach analizy danych uzyskanych z sekwencjonowania NGS wykluczono warianty:

- częste w populacji ogólnej tj. takie, których częstość jest większa od 0,1% w bazie danych gnomAD [103],
- częste (>5%) w wewnętrznej bazie danych Zakładu Genetyki Medycznej WUM,
- wykryte przez inne programy do wykrywania wariantów, ale nie przez HaploTypeCaller,
- zakwalifikowane jako artefakty sekwencjonowania na podstawie: obecności w wielu próbkach (warianty powtarzające się każdorazowo w sekwencjonowanych pulach próbek), występowania zmienionych odczytów tylko na jednej nici oraz

jedno- lub kilkunukleotydowych insercji-delecji w sekwencjach tandemowych i homopolimerycznych.

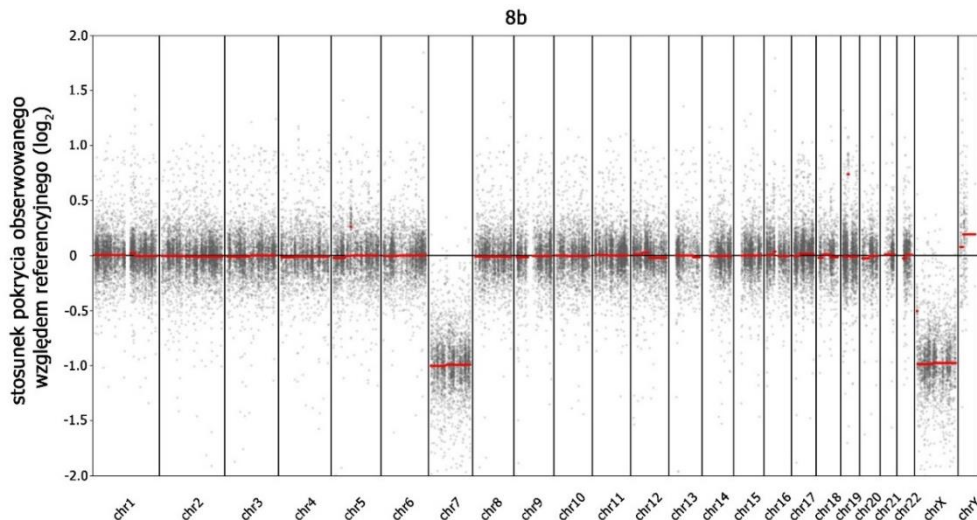
Dalszej analizie poddano warianty o potencjalnie funkcjonalnym wpływie na kodowane białko. Patogenność wykrytych wariantów oceniano z wykorzystaniem bioinformatycznych predyktorów: CADD [104], PolyPhen2 [105], SIFT [106], FATHMM [107] i Mutation Taster [108] oraz informacji dostępnych w publikacjach naukowych i bazach danych: Varsome [109], COSMIC [110] i ClinVar [111]. W przypadku zbyt małej ilości informacji wskazujących na patogenny charakter wariantu lub sprzecznych danych, warianty określano mianem wariantów o nieokreślonej patogenności (VUS, ang. *variant of uncertain significance*). Pozostałe warianty, tj. te, dla których nie znaleziono informacji o ich potencjalnie patogennej roli, zaklasyfikowano jako warianty łagodne / prawdopodobnie łagodne.

#### 3.2.2.4. Analiza zmian liczby kopii

Prawdziwość wariantów oceniano liczbowo i wizualnie (na podstawie obecności w tabeli i na rycinie oraz jakości odczytów w programie IGV), jak również uwzględniając płeć pacjenta. W opisie wyników przyjęto następujące nazewnictwo: amplifikacja ( $\geq 4$  kopii), nabycie liczby kopii (3 kopie), płytka delecja (1 kopia), głęboka delecja (0 kopii). W przypadku gdy zmianie liczby kopii ulegały wszystkie analizowane regiony na danym chromosomie, zmiany te opisywano jako monosomie lub trisomie. Przykładowy wykres analizy zmiany liczby kopii ukazujący monosomię chromosomu 7 przedstawiono na rycinie 5.

#### 3.2.2.5. Analiza umożliwiająca identyfikację $\Psi$ SMAD4

Wykrycie pseudogenu było możliwe dzięki zastosowaniu sekwencjonowania w trybie sparowanych końców (ang. *paired-end sequencing*) polegającego na sekwencjonowaniu każdego fragmentu DNA z obu końców pozostawiając niezsekwencjonowaną wstawkę pomiędzy nimi (rycina 6A). Analiza obecności  $\Psi$ SMAD4 polegała na manualnej inspekcji genów SMAD4 i SCAI (miejsca insercji pseudogenu) w programie IGV w celu poszukiwania odczytów na złączach (ang. *junction reads*) oraz, co ważniejsze, odczytów sprzecznych (ang. *discordant reads*).

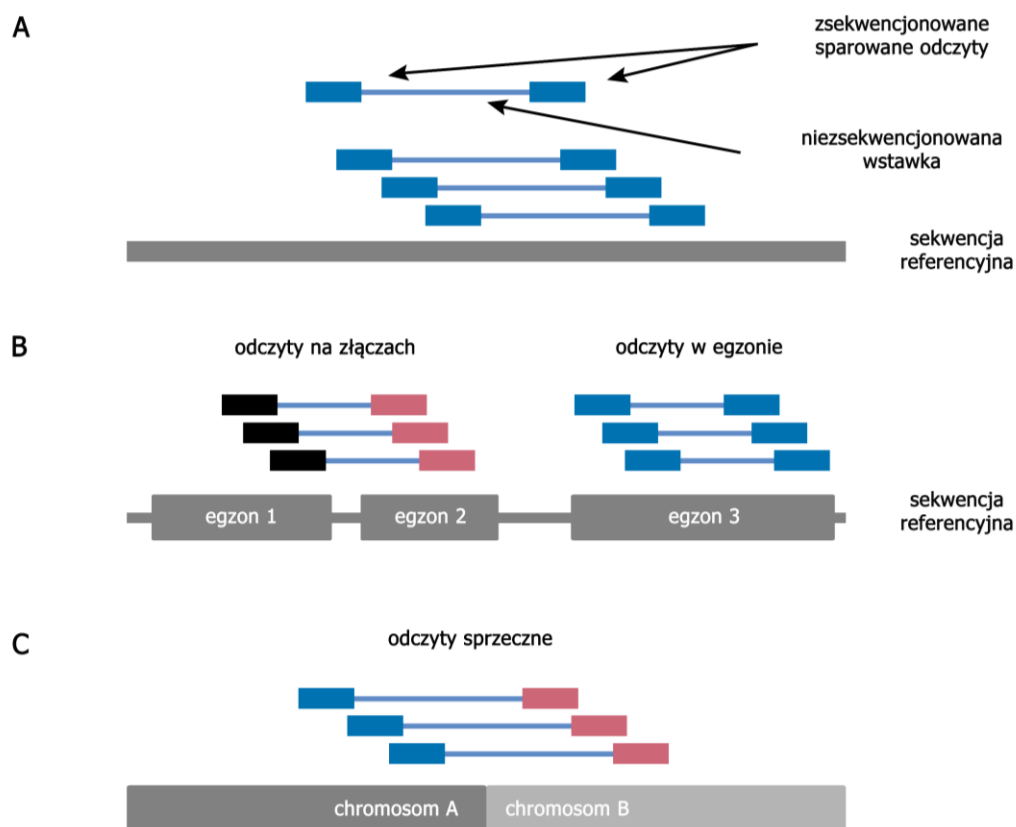


**Rycina 5. Przykładowy wykres analizy zmian liczby kopii wygenerowany za pomocą programu CNVkit.** Wykres dotyczy próbki 8b. Pojedyncze szare punkty określają zlogarytmowany stosunek pokrycia analizowanych regionów (bins) względem próbek referencyjnych. Regiony o podobnym odchyleniu od oczekiwanego pokrycia, połączone w wyniku segmentacji, oznaczono czerwonymi liniami. Numery poszczególnych chromosomów znajdują się poniżej wykresu.

W wyniku sekwencjonowania w trybie sparowanych końców, oprócz standardowych, sparowanych odczytów w egzonach (ang. *exon reads*) można uzyskać również sparowane odczyty na złączach, w których pary odczytów mapują się do sąsiadujących egzonów (rycina 6B). Obecność odczytów na złączach w sytuacji, gdy długość intronu znacznie przekracza długość sekwencjonowanego fragmentu DNA świadczy o braku sekwencji intronowej w analizowanym fragmencie a tym samym potwierdza obecność przetworzonego pseudogenu. Z kolei obecność odczytów sprzecznych tj. takich, w których sparowane odczyty mapowały się do różnych chromosomów (rycina 6C), może wskazywać m.in. translokację lub tak jak w tym przypadku miejsce insercji pseudogenu.

### 3.2.2.6. Graficzna prezentacja wyników

Wyniki analizy danych uzyskanych z celowanego sekwencjonowania sparowanych próbek CML oraz próbek z *PSMAD4* przedstawiono w postaci wykresów wykonanych z użyciem funkcji *plomaSummary*, *oncplot*, *mafCompare* i *coBarplot* pakietu *mafTools* (2.12.0) dla języka R (wersja 4.2.1).



**Rycina 6. Schematyczne przedstawienie zasady sekwencjonowania w trybie sparowanych końców oraz rodzajów odczytów, uzyskiwanych w sekwencjonowaniu NGS, wykorzystanych do identyfikacji  $\Psi$ SMAD4.** (A) Sekwencjonowanie w trybie sparowanych końców, (B) Typy odczytów w obrębie jednego genu, (C) Odczyty sprzeczne mapujące się do różnych genów/chromosomów.

### 3.2.3. Reakcja łańcuchowej polimerazy PCR do oceny obecności $\Psi$ SMAD4

Reakcję łańcuchowej polimerazy (PCR, ang. *polymerase chain reaction*) wykorzystano do szybkiej oceny częstości występowania  $\Psi$ SMAD4 (matryca: DNA genomowe z krwi/szpiku) oraz potwierdzenia jego zarodkowego charakteru (matryca: DNA zarodkowe z cebulek włosów). Wiedząc, że przetworzone pseudogeny, w porównaniu do genów rodzicielskich, cechują się brakiem sekwencji intronowych, w reakcji PCR wykorzystano startery zlokalizowane w dwóch różnych egzonach genu SMAD4 (ang. *intron-spanning primers*) tj. w egzonie 3 i 4. W wyniku takiej reakcji PCR, w próbkach bez pseudogenu obserwowano wyłącznie produkt PCR o długości 623 pz (zamplifikowany fragment genu SMAD4 zawierający sekwencję intronu 3), podczas gdy próbki z  $\Psi$ SMAD4 zawierały zarówno pochodzący z genu amplikon o długości 623 pz, jak również produkt PCR o długości 189 pz tj. zamplifikowany fragment  $\Psi$ SMAD4 pozbawiony sekwencji intronowej.



Reakcję PCR przeprowadzono z użyciem zestawu HotStarTaq® Plus DNA Polymerase (Qiagen) a objętość mieszaniny reakcyjnej w przeliczeniu na jedną próbkę wynosiła 25 µl. Skład mieszaniny reakcyjnej, sekwencje starterów wraz z długością produktów oraz warunki reakcji PCR przedstawiono odpowiednio w tabelach 5, 6, 7. Reakcje amplifikacji PCR przeprowadzono w termocyklerach Eppendorf Mastercycler ep gradient lub Biorad T100 Thermal Cycler. Długość produktów reakcji PCR sprawdzono za pomocą elektroforezy w żelu agarozowym opisanej w podrozdziale 3.2.4. *Rozdział elektroforetyczny produktów reakcji PCR.*

**Tabela 5. Skład mieszaniny reakcyjnej PCR z użyciem zestawu HotStarTaq® Plus DNA Polymerase (Qiagen).**

| SKŁAD  | OBJĘTOŚĆ [µL] | STĘŻENIE KOŃCOWE |
|--|---------------|------------------|
| 10x bufor CoralLoad (15 mM MgCl <sub>2</sub> ) | 2,5           | 1x               |
| mieszanina dNTP (10 mM)                        | 0,4           | 0,16 mM          |
| starter prawy (10 µM)                          | 0,5           | 0,2 µM           |
| starter lewy (10 µM)                           | 0,5           | 0,2 µM           |
| polimeraza HotStarTaq Plus 5U/µl               | 0,25          | 1,25 U           |
| woda   | 19,85         | -                |
| matryca DNA (50-100ng)                         | 1             | -                |

**Tabela 6. Sekwencje starterów oraz długości ampliconów w przesiewowym PCR do oceny obecności ΨSMAD4.**

| STARTER | SEKWENCJA 5'- 3'       | DŁUGOŚĆ AMPLIKONU [pz] |
|---------|------------------------|------------------------|
| SMAD4pF | CGGAAAGGATTCCTCATGT    | 623                    |
| SMAD4pR | CTGCAGTGTTAATCCTGAGAGA | lub<br>623 i 189       |

**Tabela 7. Profil temperaturowo-czasowy reakcji PCR do oceny obecności ΨSMAD4.**

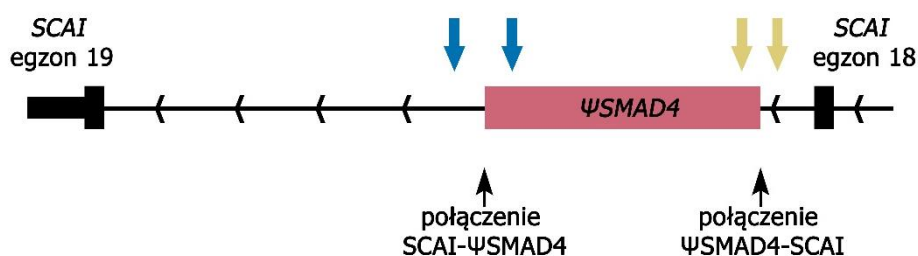
| ETAP                   | TEMPERATURA [°C] | CZAS   | LICZBA CYKLI |
|------------------------|------------------|--------|--------------|
| denaturacja wstępna    | 95               | 5 min  | 1 x          |
| denaturacja            | 94               | 40 sek | 35 x         |
| przyłączanie starterów | 55               | 40 sek |              |
| wydłużanie             | 72               | 40 sek |              |
| wydłużanie końcowe     | 72               | 5 min  | 1 x          |
| inkubacja końcowa      | 4                | O/N    | 1 x          |

### 3.2.4. Rozdział elektroforetyczny produktów reakcji PCR

Długości produktów reakcji PCR sprawdzano za pomocą elektroforezy w żelu agarozowym. Rozdział elektroforetyczny produktów PCR prowadzono w 1,5% żelu agarozowym (Promega) z dodatkiem bromku etydyny (Sigma, 1  $\mu$ l/ 50 ml żelu) przy napięciu 100V w buforze TBE. Czas elektroforezy uzależniony był od długości amplikonów i wynosił około 40-100 minut (elektroforezę wykonywano do momentu dokładnego rozdziału amplikonów w żelu, umożliwiającącego odczyt ich długości). Wizualizacji wyników dokonano z użyciem aparatu ChemiDoc™ Touch Imaging System (BioRad), do oceny długości produktów PCR wykorzystano markery wielkości GeneRuler™ 50 bp DNA Ladder lub GeneRuler™ DNA Ladder Mix (Thermo Fisher Scientific).

### 3.2.5. Określenie miejsca insercji $\Psi$ SMAD4 za pomocą sekwencjonowania metodą Sangera

Dokładne miejsce złamania genu *SCAI*, będące jednocześnie miejscem insercji  $\Psi$ SMAD4, określono z wykorzystaniem reakcji PCR i sekwencjonowania metodą Sangera. Sekwencje starterów zaprojektowano na podstawie danych uzyskanych z mapowania odczytów genu *SMAD4* z sekwencjonowania NGS. Wykonano dwie reakcje PCR, w których jeden ze starterów był komplementarny do sekwencji genu *SCAI* a drugi do sekwencji  $\Psi$ SMAD4 (rycina 7) a następnie zsekwencono powstałe produkty PCR za pomocą metody Sangera.



**Rycina 7. Schemat lokalizacji starterów w reakcjach PCR wykonanych w celu określenia miejsca insercji  $\Psi$ SMAD4.** Pary starterów do reakcji PCR oznaczono niebieskimi i beżowymi strzałkami.

#### 3.2.5.1. Reakcja PCR do określenia miejsca insercji, połączenie *SCAI*- $\Psi$ SMAD4

Reakcję PCR przeprowadzono z użyciem zestawu HotStarTaq® Plus DNA Polymerase (Qiagen). Skład mieszaniny reakcyjnej, sekwencje starterów wraz

z długością produktu oraz warunki reakcji PCR przedstawiono odpowiednio w tabelach 5, 8, 9.

**Tabela 8. Sekwencje starterów oraz długości amplikonów w PCR do określenia miejsca insercji, połączenie *SCAI-ΨSMAD4*.**

| STARTER  | SEKWENCJA 5' - 3'   | DŁUGOŚĆ AMPLIKONU [pz] |
|----------|---------------------|------------------------|
| SM_SCpsF | CACTGCCTAGTGTTCACAG | 300                    |
| SM_SCpsR | CTAGGGGAGAGCAGGAAGG |                        |

**Tabela 9. Profil temperaturowo-czasowy reakcji PCR do określenia miejsca insercji, połączenie *SCAI-ΨSMAD4*.**

| ETAP                   | TEMPERATURA [°C] | CZAS   | LICZBA CYKLI |
|------------------------|------------------|--------|--------------|
| denaturacja wstępna    | 95               | 5 min  | 1 x          |
| denaturacja            | 94               | 40 sek | 6 x          |
| przyłączanie starterów | 65, -1°C/cykl    | 40 sek |              |
| wydłużanie             | 72               | 30 sek |              |
| denaturacja            | 94               | 40 sek | 32 x         |
| przyłączanie starterów | 60               | 40 sek |              |
| wydłużanie             | 72               | 30 sek |              |
| wydłużanie końcowe     | 72               | 5 min  | 1 x          |
| inkubacja końcowa      | 4                | O/N    | 1 x          |

### 3.2.5.2. Reakcja PCR do określenia miejsca insercji, połączenie *ΨSMAD4-SCAI*

Reakcję PCR przeprowadzono z użyciem zestawu HotStarTaq® Plus DNA Polymerase (Qiagen). Skład mieszaniny reakcyjnej, sekwencje starterów wraz z długością produktu oraz warunki reakcji PCR przedstawiono odpowiednio w tabelach 5, 10, 11.

**Tabela 10. Sekwencje starterów oraz długość amplikonu w PCR do określenia miejsca insercji, połączenie *ΨSMAD4-SCAI*.**

| STARTER  | SEKWENCJA 5' - 3'         | DŁUGOŚĆ AMPLIKONU [pz]                                       |
|----------|---------------------------|--|
| SC_SM_3F | TCCCTGATCCTATCTCCCTCT     | około 580-610<br>(w zależności od długości sekwencji poli-A) |
| psSMAD4R | TCTAAGCCCTTTGCCATCAATGATC |  |

**Tabela 11. Profil temperaturowo-czasowy reakcji PCR do określenia miejsca insercji, połączenie  $\Psi$ SMAD4-SCAI.**

| ETAP                   | TEMPERATURA [°C] | CZAS   | LICZBA CYKLI |
|------------------------|------------------|--------|--------------|
| denaturacja wstępna    | 95               | 5 min  | 1 x          |
| denaturacja            | 94               | 40 sek | 38 x         |
| przyłączanie starterów | 60               | 40 sek |              |
| wydłużanie             | 72               | 40 sek |              |
| wydłużanie końcowe     | 72               | 4 min  | 1 x          |
| inkubacja końcowa      | 4                | O/N    | 1 x          |

### 3.2.5.3. Rozdział elektroforetyczny produktów PCR

Długość produktów reakcji PCR sprawdzano za pomocą elektroforezy w żelu agarozowym opisanej w podrozdziale 3.2.4. *Rozdział elektroforetyczny produktów reakcji PCR.*

### 3.2.5.4. Oczyszczenie amplikonów za pomocą kulek paramagnetycznych

Amplikony PCR zmieszano z kulkami paramagnetycznymi VAHTS DNA Clean Beads (Vazyme) w stosunku 1:1,8 (10  $\mu$ l amplikonu i 18  $\mu$ l kulek) i inkubowano 10 min w temp. pokojowej w celu związania DNA na powierzchni kulek. Następnie próbki wstawiono na separator magnetyczny w celu wychwycenia kulek, usunięto supernatant a DNA związane z kulkami dwukrotnie przepłukano w 200  $\mu$ l świeżo przygotowanego 80% etanolu. Próbki wysuszono, po czym zdjęto je z separatora magnetycznego i zawieszono w 10  $\mu$ l wody. Po 10 min inkubacji, próbki ponownie wstawiono na separator magnetyczny, a wyluowane DNA przeniesiono do nowych próbek.

### 3.2.5.5. Reakcja fluorescencyjnego znakowania, oczyszczanie produktów i przygotowanie do sekwencjonowania metodą Sanger

Oczyszczony amplikon poddano reakcji fluorescencyjnego znakowania z użyciem zestawu BigDye<sup>TM</sup> Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Do reakcji wykorzystano startery z poprzednich reakcji PCR, których sekwencje przedstawiono w tabelach 8 i 10, z jednym wyjątkiem, gdzie zamiast startera SC\_SM\_3F użyto starter SC\_SM\_aw1R (5'-GATTATAGGCGCCCACCAC-3'). Skład mieszaniny reakcyjnej oraz warunki reakcji PCR przedstawiono odpowiednio w tabelach 12 i 13.

**Tabela 12. Skład mieszaniny reakcji fluorescencyjnego znakowania.**

| SKŁAD                         | OBJĘTOŚĆ [μL] |
|-------------------------------|---------------|
| 5x Sequencing buffer          | 3,6           |
| Ready reaction mix            | 0,4           |
| starter prawy lub lewy (1 μM) | 2             |
| woda                          | 2             |
| matryca                       | 2             |

**Tabela 13. Profil temperaturowo-czasowy reakcji fluorescencyjnego znakowania.**

| ETAP                   | TEMPERATURA [°C] | CZAS        | LICZBA CYKLI |
|------------------------|------------------|-------------|--------------|
| denaturacja wstępna    | 95               | 5 min       | 1 x          |
| denaturacja            | 95               | 10 sek      | 55 x         |
| przyłączanie starterów | 50               | 15 sek      |              |
| wydłużanie             | 60               | 1min 30 sek |              |
| wydłużanie końcowe     | 60               | 5 min       | 1 x          |
| inkubacja końcowa      | 4                | O/N         | 1 x          |

Po skończonej reakcji produkt ponownie oczyszczono za pomocą kulek paramagnetycznych VAHTS DNA Clean Beads (Vazyme). W tym celu do 10 μl amplikonu dodano 10 μl kulek, wymieszano i dodano 42 μl świeżo przygotowanego 90% etanolu, całość dokładnie wymieszano i inkubowano 5 min. Próbkę wstawiono na separator magnetyczny w celu wychwycenia kulek ze związanym DNA, usunięto supernatant a osad kulek z DNA przepłukano dwukrotnie w 100 μl 90% etanolu. Próbkę wysuszono pozostawiając je na separatorze. Następnie zdjęto je z magnesu, zawieszono w 10 μl wody i inkubowano przez 5 min. Próbkę ponownie wstawiono na separator magnetyczny a 8 μl oczyszczonego amplikonu przeniesiono do nowych próbek zawierających 8 μl formamidu (Applied Biosystems).

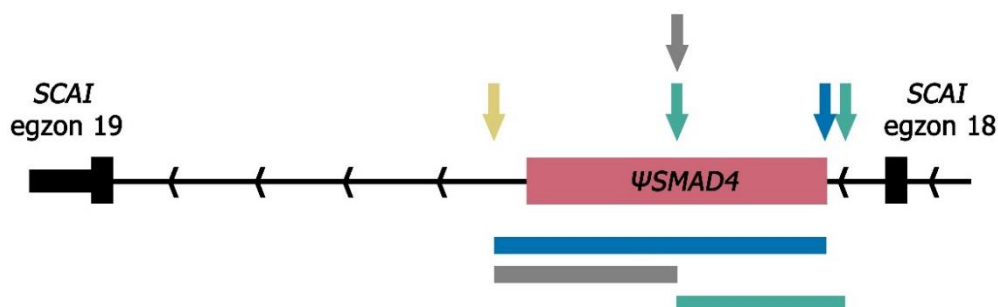
Oczyszczone produkty zsekwencjonowano w Pracowni Sekwencjonowania DNA i Syntezy Oligonukleotydów Instytutu Biochemii i Biofizyki PAN. Chromatogramy analizowano za pomocą oprogramowania FinchTV.

### 3.2.6. Określenie sekwencji ΨSMAD4 za pomocą sekwencjonowania NGS

W celu określenia dokładnej, całkowitej sekwencji ΨSMAD4, zamplifikowano go za pomocą reakcji PCR, otrzymane amplikony zsekwencjonowano metodą NGS a następnie złożyło sekwencję *de novo* bez wykorzystania genomu referencyjnego.

### 3.2.6.1. PCR do amplifikacji $\Psi$ SMAD4

Opracowano 3 reakcje PCR umożliwiające całkowitą bądź częściową amplifikację  $\Psi$ SMAD4 wraz z miejscem insercji (rycina 8).



**Rycina 8. Schemat lokalizacji starterów oraz amplikonów powstałych w reakcjach PCR wykonanych w celu określenia sekwencji  $\Psi$ SMAD4.** Pary starterów do reakcji PCR oznaczono strzałkami nad schematem  $\Psi$ SMAD4, zaś amplikony powstałe w wyniku reakcji PCR oznaczono prostokątami pod schematem  $\Psi$ SMAD4. Kolory strzałek oraz prostokątów odpowiadają kolejnym reakcjom PCR: 1 (niebieski), 2 (szary) oraz 3 (zielony). Starter oznaczony kolorem beżowym został wykorzystany w reakcjach PCR 1 i 2.

Reakcje PCR przeprowadzono z użyciem zestawów KAPA HiFi HotStart ReadyMix (Roche) lub Expand™ Long Template PCR System (Roche). Składy mieszanin reakcyjnych, sekwencje starterów wraz z długościami produktów oraz warunki reakcji PCR przedstawiono w tabelach 14-19.

**Tabela 14. Skład mieszaniny reakcyjnej PCR z użyciem zestawu KAPA HiFi HotStart ReadyMix (Roche) do amplifikacji  $\Psi$ SMAD4- reakcja 1 i 2.**

| SKŁAD                                 | OBJĘTOŚĆ<br>[ $\mu$ L] | STĘŻENIE<br>KOŃCOWE |
|---------------------------------------|------------------------|---------------------|
| 2x ReadyMix (5 mM MgCl <sub>2</sub> ) | 12,5                   | 1x                  |
| starter prawy (10 $\mu$ M)            | 1,5                    | 0,6 $\mu$ M         |
| starter lewy (10 $\mu$ M)             | 1,5                    | 0,6 $\mu$ M         |
| woda                                  | 8,5                    | -                   |
| matryca DNA (50-100ng)                | 1                      | -                   |

**Tabela 15. Skład mieszaniny reakcyjnej PCR z użyciem zestawu Expand™ Long Template PCR System (Roche) do amplifikacji  $\Psi$ SMAD4– reakcja 3.**

| SKŁAD                                    | OBJĘTOŚĆ<br>[ $\mu$ L] | STĘŻENIE<br>KOŃCOWE |
|--|------------------------|---------------------|
| 10x bufor 2 (27,5 mM MgCl <sub>2</sub> ) | 2,5                    | 1x                  |
| mieszanina dNTP (10 mM)                  | 1,25                   | 0,5 mM              |
| starter prawy (10 $\mu$ M)               | 0,75                   | 0,3 $\mu$ M         |
| starter lewy (10 $\mu$ M)                | 0,75                   | 0,3 $\mu$ M         |
| polimeraza 5U/ $\mu$ l                   | 0,38                   | 1,9 U               |
| woda                                     | 18,37                  | -                   |
| matryca DNA (50-100ng)                   | 1                      | -                   |

**Tabela 16. Sekwencje starterów oraz długości amplikonów w reakcjach PCR do amplifikacji  $\Psi$ SMAD4.**

| REAKCJA | STARTER  | SEKWENCJA 5'- 3'      | DŁUGOŚĆ<br>AMPLIKONU [pz]  |
|---------|----------|-----------------------|--|
| 1       | SC_SM_3F | TCCCTGATCCTATCTCCCTCT | ok. 3900-3930<br>(w zależności od<br>długości sekwencji<br>poli-A) |
|         | SM_SCzIR | AAAAATTTCCCCTGGTCGTC  |  |
| 2       | SM_SCpsF | CACTGCCTAGTGTTGCACAG  | 1926   |
|         | SMps1F   | CTCTGTACAAAGACCGCGTG  |  |
| 3       | SC_SM_3F | TCCCTGATCCTATCTCCCTCT | ok. 2190-2220<br>(w zależności od<br>długości sekwencji<br>poli-A) |
|         | SC_SM_3R | CACGCGGTCTTTGTACAGAG  |  |

**Tabela 17. Profil temperaturowo-czasowy pierwszej reakcji PCR do amplifikacji  $\Psi$ SMAD4.**

| ETAP                   | TEMPERATURA<br>[°C] | CZAS   | LICZBA<br>CYKLI |
|------------------------|---------------------|--------|-----------------|
| denaturacja wstępna    | 95                  | 3 min  | 1 x             |
| denaturacja            | 98                  | 20 sek | 38 x            |
| przyłączenie starterów | 65                  | 15 sek |                 |
| wydłużanie             | 72                  | 1 min  |                 |
| wydłużanie końcowe     | 72                  | 4 min  | 1 x             |
| inkubacja końcowa      | 4                   | O/N    | 1 x             |

**Tabela 18. Profil temperaturowo-czasowy drugiej reakcji PCR do amplifikacji  $\Psi$ SMAD4.**

| ETAP                   | TEMPERATURA [°C] | CZAS   | LICZBA CYKLI |
|------------------------|------------------|--------|--------------|
| denaturacja wstępna    | 95               | 3 min  | 1 x          |
| denaturacja            | 98               | 20 sek | 38 x         |
| przyłączanie starterów | 65               | 15 sek |              |
| wydłużanie             | 72               | 45 sek |              |
| wydłużanie końcowe     | 72               | 3 min  | 1 x          |
| inkubacja końcowa      | 4                | O/N    | 1 x          |

**Tabela 19. Profil temperaturowo-czasowy trzeciej reakcji PCR do amplifikacji  $\Psi$ SMAD4.**

| ETAP                   | TEMPERATURA [°C] | CZAS          | LICZBA CYKLI |
|------------------------|------------------|---------------|--------------|
| denaturacja wstępna    | 94               | 2 min         | 1 x          |
| denaturacja            | 94, -0,2°C/cykl  | 10 sek        | 10 x         |
| przyłączanie starterów | 60               | 30 sek        |              |
| wydłużanie             | 68               | 40 sek        |              |
| denaturacja            | 94, -0,1°C/cykl  | 15 sek        | 25 x         |
| przyłączanie starterów | 60               | 30 sek        |              |
| wydłużanie             | 68               | 4 min, +20sek |              |
| wydłużanie końcowe     | 72               | 7 min         | 1 x          |
| inkubacja końcowa      | 4                | O/N           | 1 x          |

3.2.6.2. Rozdział elektroforetyczny produktów PCR i oczyszczanie amplikonów z żelu agarozowego

Długość produktów reakcji PCR sprawdzono za pomocą elektroforezy w żelu agarozowym opisanej w podrozdziale 3.2.4. *Rozdział elektroforetyczny produktów reakcji PCR*, z tą różnicą, że rozdział elektroforetyczny prowadzono w 1% żelu wykonanym z agarozy o niskiej temperaturze topnienia (MP Biomedicals). Z powodu obecności niespecyficznych prążków, specyficzne produkty PCR wycięto z żelu i oczyszczono za pomocą zestawu Monarch® DNA Gel Extraction Kit (New England BioLabs) zgodnie z zaleceniami producenta. Oczyszczony produkt PCR zawieszono w 10  $\mu$ l wody.



### 3.2.6.3. Sekwencjonowanie amplikonów metodą NGS

Oczyszczone produkty PCR przygotowano do sekwencjonowania NGS poprzez wykonanie bibliotek za pomocą komercyjnego zestawu Nextera XT DNA Library Preparation Kit (Illumina), zgodnie z zaleceniami producenta. Biblioteki przygotowano z 1 ng amplikonów PCR a ogólne etapy procedury obejmowały: tagmentację (proces jednoczesnej fragmentacji DNA i dołączenia sekwencji adaptorowych), amplifikację bibliotek w reakcji PCR, oczyszczenie bibliotek za pomocą kulek paramagnetycznych oraz ocenę jakościową bibliotek z użyciem aparatu Bioanalyser 2100 (Agilent Technologies) i zestawu Agilent High Sensitivity DNA Kit (Agilent Technologies). Gotowe biblioteki zostały zsekwencjonowane na platformie Illumina HiSeq 1500 w trybie sparowanych końców (2×100 pz).

### 3.2.6.4. Złożenie sekwencji *de novo* bez wykorzystania genomu referencyjnego

Wyniki otrzymane z sekwencjonowania amplikonów metodą NGS analizowano we współpracy z pracownikami Zakładu Sztucznej Inteligencji Politechniki Warszawskiej. W skrócie, przeprowadzona analiza polegała na selekcji wysokiej jakości odczytów z sekwencjonowania amplikonów w trybie sparowanych końców, a następnie złożeniu sekwencji *de novo* bez wykorzystania genomu referencyjnego z użyciem programu ABySS 2.0 [112] oraz autorskiej aplikacji *dnaasm* [113].

## 3.2.7. Sekwencjonowanie transkryptomu (RNAseq)

Sekwencjonowanie transkryptomu (RNAseq, ang. *RNA-sequencing*) wykonano dla 16 chorych na CML. Próbki od chorych podzielono na dwie grupy: badaną (próbki CML z  $\Psi$ SMAD4) i kontrolną (próbki CML bez pseudogenu), a w obu grupach znajdowało się po 5 próbek z fazy CP i 3 próbki z fazy BP.

### 3.2.7.1. Przygotowanie bibliotek cDNA i sekwencjonowanie NGS

Biblioteki do sekwencjonowania transkryptomu przygotowano za pomocą komercyjnego zestawu TruSeq RNA Library Prep Kit (Illumina), umożliwiającego konwersję mRNA do biblioteki cDNA poddawanej sekwencjonowaniu. Procedurę wykonano zgodnie z zaleceniami producenta w ścieżce dla mniejszej ilości próbek („*Low sample protocol*”).

Przed przystąpieniem do konstrukcji bibliotek, jakość RNA została oceniona z wykorzystaniem aparatu do elektroforezy mikroprzepływowej Bioanalyser 2100 (Agilent Technologies) i zestawu Agilent RNA 6000 Nano Kit (Agilent Technologies). Do dalszej procedury zostały dopuszczone jedynie takie próbki RNA, dla których RIN (ang. *RNA Integrity Number*) określający integralność RNA wynosił >8.

Główne etapy konstrukcji biblioteki obejmowały: wychwycenie mRNA z całkowitego RNA za pomocą kulek magnetycznych skoniugowanych z oligo-dT, fragmentację mRNA, syntezę pierwszej a następnie drugiej nici cDNA, naprawę końców, adenylację, ligację adaptorów w systemie 24-indeksowym, amplifikację biblioteki cDNA za pomocą reakcji PCR.

Stężenie gotowych bibliotek było mierzone z użyciem fluorymetru Qubit 2.0 (Life Technologies) z wykorzystaniem zestawu odczynników dsDNA High Sensitivity Assay Kit (Invitrogen). Rozkład wielkości bibliotek oceniano za pomocą aparatu Bioanalyser 2100 (Agilent Technologies) i zestawu Agilent High Sensitivity DNA Kit (Agilent Technologies).

Biblioteki były sekwencjonowane w trybie sparowanych końców na platformach Illumina HiSeq 1500 lub NovaSeq 6000.

#### 3.2.7.2. Obróbka i analiza danych uzyskanych z sekwencjonowania

Wyniki RNAseq analizowano we współpracy z firmą Xenstats. W skrócie obróbka danych polegała na ocenie jakości surowych odczytów za pomocą programów fastqc (v0.11.9) i multiqc (v1.0.dev0), zmapowaniu surowych odczytów do sekwencji genomu ludzkiego (Ensembl GRCh38) z wykorzystaniem STAR (v2.7.10b) i zliczeniu odczytów zmapowanych do sekwencji egzonów z użyciem *featureCounts* (v2.0.1). Ekspresja różnicowa genów została obliczona za pomocą pakietu *edgeR* (v3.36.0) a jej wyniki zobrazowano za pomocą heatmap wygenerowanych z wykorzystaniem pakietu *mixOmics* (v6.18.1) w języku R (v4.1.2).

Dodatkowym sposobem oceny ekspresji  $\Psi$ *SMAD4* było zmapowanie surowych odczytów do ręcznie przygotowanej sekwencji referencyjnej, tj. sekwencji pseudogenu wraz z miejscem jego insercji w genie *SCAI* (egzon 18 i intron 18) a następnie poszukiwanie pojedynczych odczytów oraz par odczytów mapujących się zarówno do pseudogenu jak i miejsca insercji.

### 3.2.8. Analiza statystyczna

Analiza statystyczna została wykonana dla określenia częstości występowania  $\Psi$ SMAD4. Różnice pomiędzy kontrolą (osoby bez białaczki) a próbami badanymi (osoby z różnymi rodzajami białaczki) oceniono za pomocą testu dokładnego Fishera (Fisher's exact test). Wartość  $p < 0,05$  uznano za istotną statystycznie. Analizy statystyczne zostały wykonane z użyciem programu PQStat 1.8.2.

W przypadku analizy porównawczej częstości zmian punktowych i małych insercji-delecji w 25 sparowanych próbkach CML analiza statystyczna została wykonana za pomocą funkcji *mafCompare* pakietu *maftools* (test Fishera).

## 4. WYNIKI

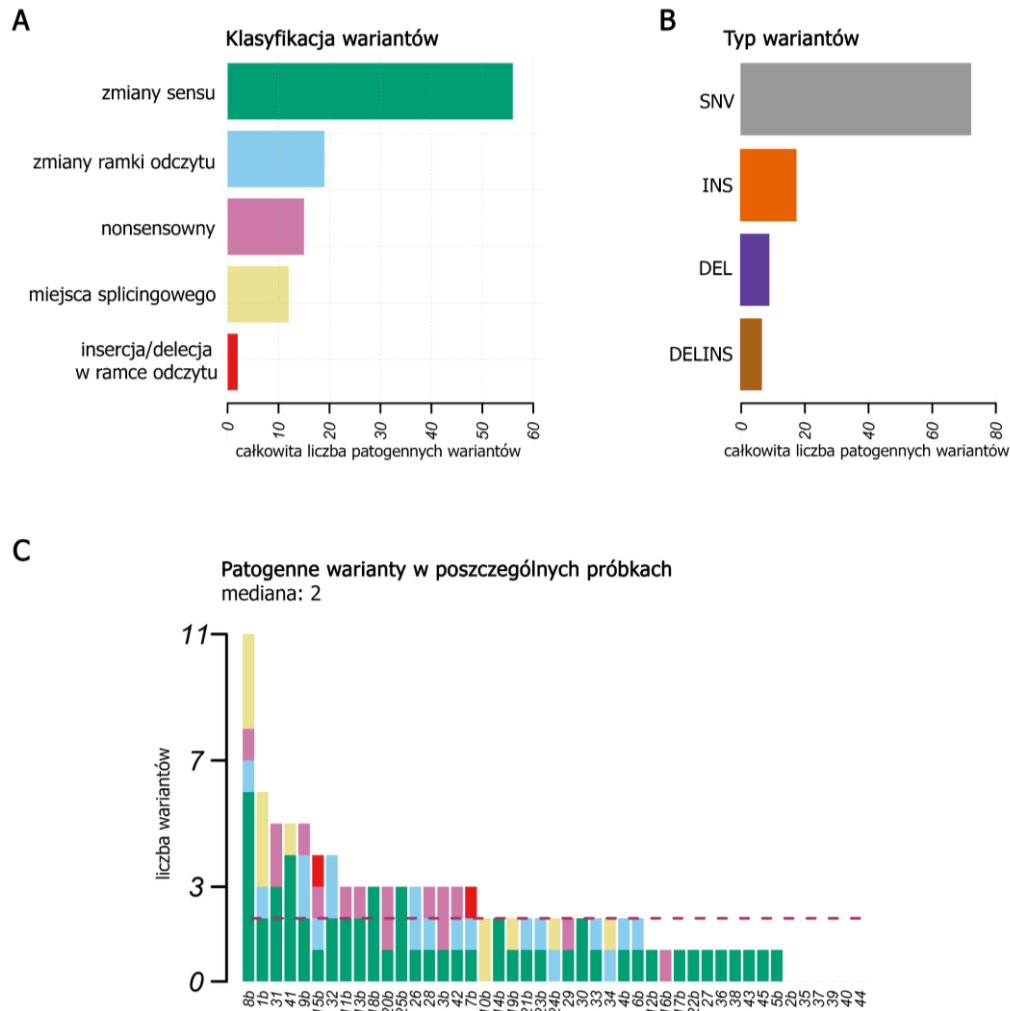
### 4.1. Analiza danych uzyskanych z celowanego sekwencjonowania próbek CML

W celu identyfikacji nowych zmian genetycznych w CML wykonano celowane sekwencjonowanie 950–1500 genów powiązanych z onkogenezą i nowotworami układu krwiotwórczego u 45 chorych w fazie blastycznej a dla 25 z nich zsekwenconowano również dostępne próbki sprzed progresji choroby, głównie z fazy CP, jak również AP lub pierwszej fazy BP (w sumie 70 próbek CML). Analiza uzyskanych wyników obejmowała identyfikację mutacji małej skali tj. zmian punktowych i małych insercji-delekcji w 193 genach (tabela 4), zmian liczby kopii genów (analiza wykonana dla wszystkich zsekwenconowanych genów, ale prezentacja wyników ograniczona do istotnych zmian w genach: *IKZF1*, *CDKN2A*, *CDKN2B*, *TP53* i *BCR::ABL1*) oraz dodatkowych aberracji chromosomowych- ACA.

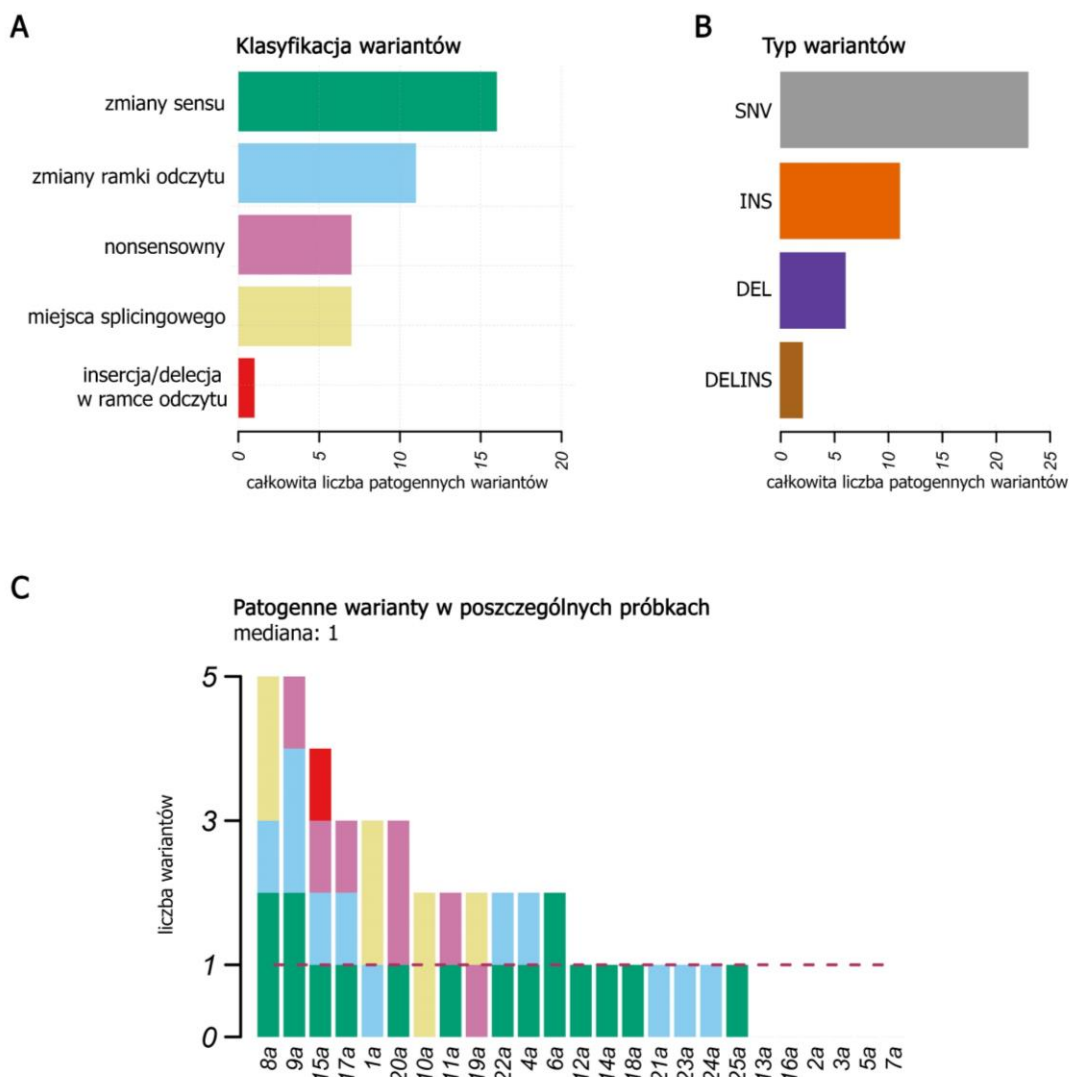
Patogenne zmiany wykryte w ramach niniejszej pracy doktorskiej określane są mianem patogennych wariantów, a nie mutacji, gdyż brak materiału zarodkowego uniemożliwiało jednoznaczne potwierdzenie somatycznego charakteru wykrytych zmian. W toku analizy mutacji małej skali zidentyfikowano w sumie 146 wariantów patogennych / prawdopodobnie patogennych (dla uproszczenia w dalszej części pracy określane jako warianty patogene); 104 i 42 odpowiednio w próbkach BP i próbkach sprzed progresji choroby. Szczegółowe informacje dotyczące wykrytych patogennych wariantów przedstawiono na rycinach 9 (próbki BP) i 10 (próbki sprzed progresji choroby). Patogenne zmiany punktowe i małe insercje-delekcje wykryto w 39/45 (87%) próbkach BP oraz odpowiadającym im 19/25 (76%) próbkach sprzed progresji CML (ryciny 9C i 10C), mediana liczby patogennych wariantów na próbkę wynosiła 2 i 1 odpowiednio w fazie BP i próbkach sprzed progresji choroby. W obu grupach największy odsetek patogennych wariantów stanowiły warianty pojedynczych nukleotydów (SNV, ang. *single nucleotide variant*) w postaci wariantów zmiany sensu (ryciny 9A-B i 10A-B).

Według przyjętych kryteriów (opisanych w podrozdziale 3.2.2.3. *Analiza patogennych wariantów*), przeprowadzona analiza pozwoliła na wykrycie 126 wariantów VUS o nieokreślonej patogenności, w tym 89 w próbkach BP i 37 w próbkach sprzed progresji. Do genów, w których warianty VUS wykryto u minimum 3 chorych zaliczono:

*MYH11* (5/45; 11%), *PDE4DIP* (4/45; 9%), *LRP1B* (4/45; 9%), *NTRK1* (3/45; 7%) oraz *PCLO* (3/45; 7%).



**Rycina 9. Statystyki dotyczące patogennych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 45 próbek CML od chorych w fazie BP.** (A) Klasyfikacja wykrytych wariantów ze względu na wywoływany przez nie efekt, (B) Klasyfikacja wykrytych wariantów ze względu na ich typ; SNV- wariant pojedynczego nukleotydu, INS- insercja, DEL- delecja, DELINS- delecja-insercja, (C) Liczba patogennych wariantów wykryta w poszczególnych próbkach; kody próbek znajdują się pod odpowiadającymi im słupkami. Przerywaną linią oznaczono medianę patogennych wariantów. Wykresy wygenerowano za pomocą funkcji *plomafSummary* pakietu *mafTools* dla patogennych zmian punktowych i małych insercji-delekcji, z pominięciem zmian liczby kopii genów oraz aberracji chromosomowych.



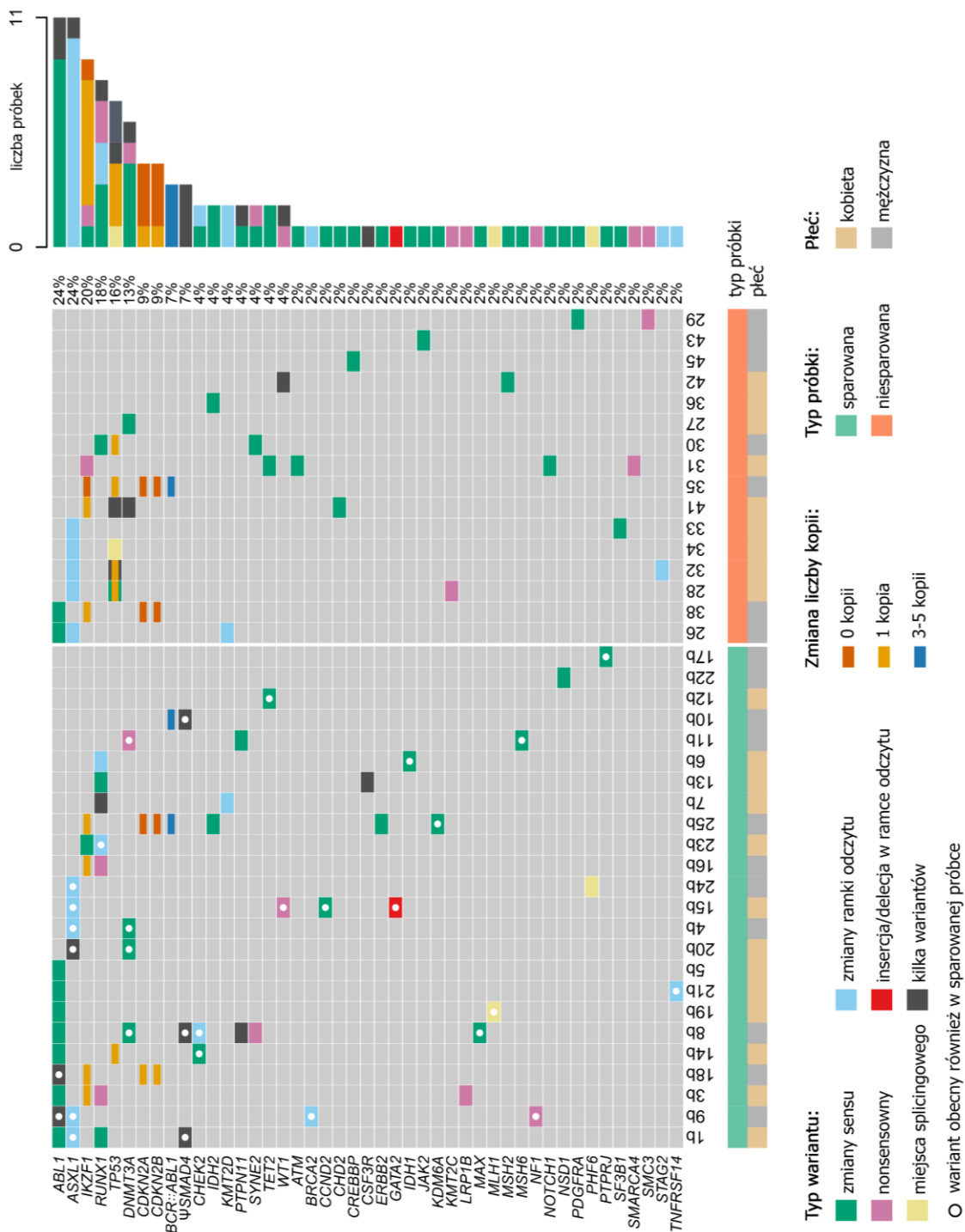
**Rycina 10. Statystyki dotyczące patogenicznych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 25 sparowanych próbek CML sprzed progresji choroby.** (A) Klasyfikacja wykrytych wariantów ze względu na wywoływany przez nie efekt, (B) Klasyfikacja wykrytych wariantów ze względu na ich typ; SNV- wariant pojedynczego nukleotydu, INS- insercja, DEL- delecja, DELINS- delecja-insercja, (C) Liczba patogenicznych wariantów wykryta w poszczególnych próbkach; kody próbek znajdują się pod odpowiadającymi im słupkami. Przerwaną linią oznaczono medianę patogenicznych wariantów. Wykresy wygenerowano za pomocą funkcji *plomafSummary* pakietu *maftools* dla patogenicznych zmian punktowych i małych insercji-delekcji, z pominięciem zmian liczby kopii genów oraz aberracji chromosomowych.

## 4.2. Identyfikacja patogenicznych wariantów, zmian liczby kopii i aberracji chromosomowych w fazie BP

Przeprowadzona analiza wykazała patogenne zmiany łącznie u 89% (40/45) chorych w fazie BP, z czego warianty punktowe i małe insercje-delekcje wykryto u 87% (39/45) chorych a zmiany liczby kopii wybranych genów u 22% (10/45).

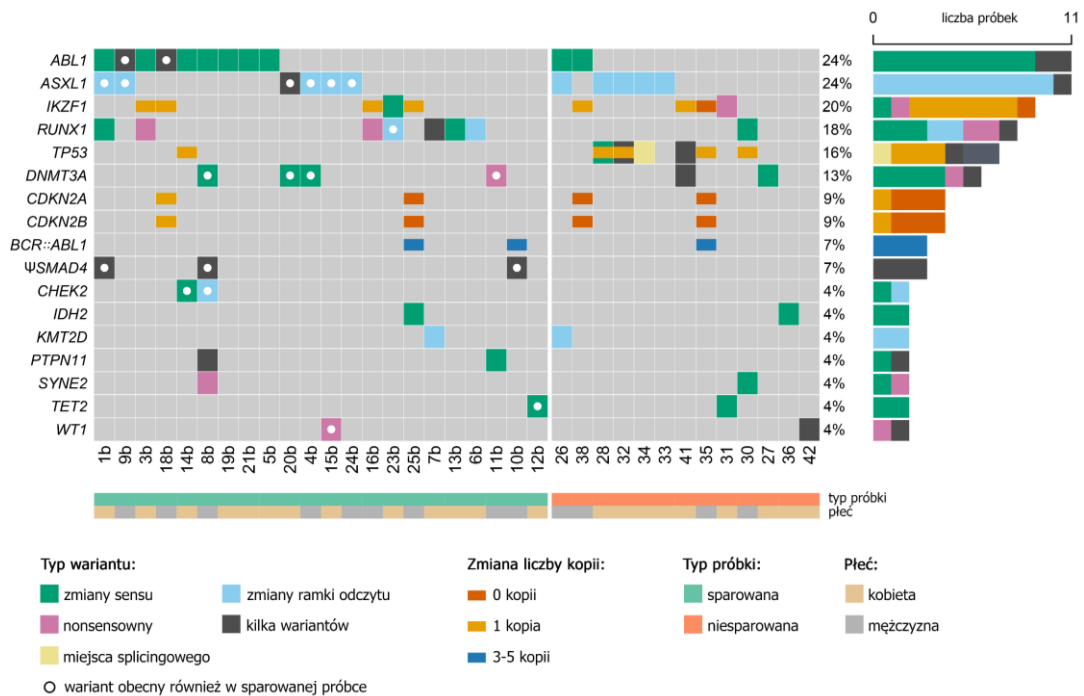
Rycina 11 przedstawia wszystkie istotne, patogenne zmiany jakie wykryto w trakcie analizy, natomiast dla ułatwienia analizy tych najczęstszych, rycina 12 ograniczona została do genów, w których patogenne warianty lub zmiany liczby kopii zidentyfikowano u minimum dwojga chorych.

Zmiany punktowe, małe insercje-delecje lub zmiany liczby kopii wykryto w 45 ze 193 przeanalizowanych genów (rycina 11). Do genów w których patogenne warianty/zmiany liczby kopii zidentyfikowano u minimum 3 chorych należały: *ABL1* (SNV) (11/45; 24%), *ASXLI* (11/45; 24%), *IKZF1* (9/45; 20%), *RUNXI* (8/45; 18%), *TP53* (7/45; 16%), *DNMT3A* (6/45; 13%), *CDKN2A* i *CDKN2B* (4/45; 9%), *BCR::ABL1* (zmiana liczby kopii) (3/45; 7%) i  $\Psi$ *SMAD4* (3/45; 7%) (rycina 11-12). Obserwowane patogenne warianty miały charakter zmian charakterystycznych dla poszczególnych genów. Przykładowo, w *ABL1* obserwowano typowe warianty zmiany sensu w domenie kinazowej, w *ASXLI* były to warianty o charakterze zmian ramki odczytu oraz warianty nonsensowne w egzonie 12, natomiast w *RUNXI* zmiany występowały w obrębie funkcjonalnych domen, zaburzających naturalną funkcję kodowanego białka. W genach *IKZF1* i *TP53*, oprócz wariantów SNV, zaobserwowano również zmiany liczby kopii: w przypadku genu *TP53* - płytkie delecje, a w *IKZF1* – płytkie i głębokie delecje. Natomiast w genach *CDKN2A* i *CDKN2B* zidentyfikowano płytkie i głębokie delecje bez dodatkowych wariantów SNV (rycina 11-12).



**Rycina 11. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 45 próbek CML w fazie BP.** Wykres przedstawia wszystkie istotne, patogenne zmiany jakie zidentyfikowano w ramach analizy zmian punktowych, małych insercji-delekcji i zmian liczby kopii genów. Lewa część wykresu dotyczy próbek sparowanych a prawa niesparowanych. Na wykresie pominięto próbki, w których nie wykryto żadnych istotnych zmian. Kody poszczególnych próbek umieszczono pod wykresem. Po prawej stronie podano odsetek próbek, w których wykryto istotne zmiany w poszczególnych genach. Wykres wygenerowano za pomocą funkcji *oncoplot* pakietu *maftools*.





**Rycina 12. Profil zmian genetycznych w najczęściej zmutowanych genach w CML w fazie BP.** W porównaniu do ryciny 11, wykres został ograniczony jedynie do genów, w których patogenne warianty i zmiany liczby kopii wykryto w próbkach BP u minimum 2 z 45 chorych. Lewa część wykresu dotyczy próbek sparowanych a prawa niesparowanych. Kody poszczególnych próbek umieszczono pod wykresem. Po prawej stronie podano odsetek próbek, w których wykryto istotne zmiany w poszczególnych genach. Wykres wygenerowano za pomocą funkcji *oncplot* pakietu *maftools*.

Aberracje chromosomowe, w tym ACA, wykryte za pomocą badania cytogenetycznego (wykonanego przez ośrodek, z którego pochodzi dany materiał) lub analizy zmian liczby kopii (przeprowadzonej na podstawie badania NGS) zostały przedstawione w tabeli 20. U 5/45 (11%) chorych wykryto po 3 lub więcej aberracje chromosomowe, tym samym określając ich kariotyp jako złożony. W żadnej z próbek BP nie wykryto dodatkowego chromosomu Ph, jednak analiza zmian liczby kopii wykazała amplifikację (4-5 kopii) lub nabycie liczby kopii (3 kopie) genu fuzyjnego *BCR::ABL1* u trzech chorych (7%, rycina 12; próbki 10b, 25b i 35).

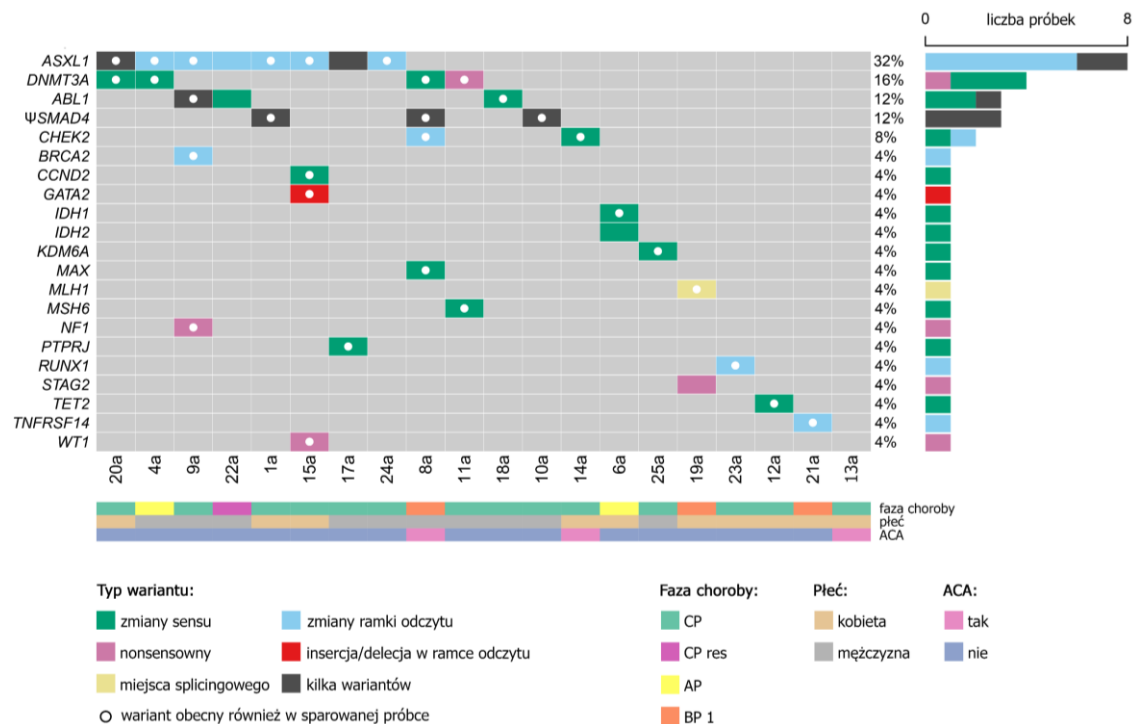
**Tabela 20. Aberracje chromosomowe w CML w fazie BP.** Aberracje wykryte w ramach analizy zmian liczby kopii z danych NGS lub badania cytogenetycznego (gdy dostępne). <sup>1</sup>zmiana również w sparowanej próbce sprzed progresji, <sup>2</sup>brak pełnego kariotypu, zmiany wykryte w badaniu NGS, <sup>3</sup>w próbce CP złożony kariotyp obserwowany w jednym klonie komórek.

| Wykryta aberracja chromosomowa                           | Numer próbki                       |
|--|------------------------------------|
| <b>ACA</b>   |                                    |
| trisomia chromosomu 8                                    | 32                                 |
| monosomia chromosomu 7                                   | 8b <sup>1</sup> , 13b <sup>1</sup> |
| izochromosom 17q   | 40                                 |
| rearanżacja MECOM (3q26.2)                               | 42, 43                             |
| kariotyp złożony   | 6b, 10b, 23b, 28, 41               |
| <b>inne zmiany</b>                                       |                                    |
| der(7;9)(q10;q10)  | 21b                                |
| del(7p), del(9p), subklonalna tetrasomia 8q <sup>2</sup> | 18b                                |
| del(7p), del(9p) <sup>2</sup>                            | 38                                 |
| del(17p), del(9p21.3-p13.2) <sup>2</sup>                 | 35                                 |
| del(9p), amp(9q) <sup>2</sup>                            | 25b                                |
| del(17p), amp(17q) <sup>2</sup>                          | 14b <sup>3</sup> , 32              |
| del(17p13.3-p13.1), amp(17q21.33-q25.3) <sup>2</sup>     | 30                                 |

### 4.3 Identyfikacja patogennych wariantów, zmian liczby kopii i aberracji chromosomowych w próbkach sprzed progresji choroby

Za pomocą metody NGS zsekwencjonowano również pochodzące sprzed progresji choroby, sparowane próbki dostępne dla 25 chorych, w tym 19 próbek z fazy CP, 1 próbkę oporności wtórnej w fazie CP (CP res, próbka 22a), dwie próbki z fazy AP (próbka 4a, 6a) i trzy próbki z pierwszej fazy blastycznej (BP 1, próbki 8a, 19a i 21a). W toku analizy zidentyfikowano patogenne zmiany w 80% (20/25) próbek sprzed progresji, w tym: zmiany punktowe i małe insercje-delecje w 19 próbkach (76%) oraz ACA w 3 próbkach (12%). Zdecydowana większość wykrytych patogennych wariantów była obecna również w sparowanych próbkach z fazy BP, a do najczęściej zmutowanych genów zaliczono: *ASXL1* (8/25; 32%), *DNMT3A* (4/25; 16%), *ABL1* (3/25; 12%), *PSMAD4* (3/25; 12%) i *CHEK2* (2/25; 8%) (rycina 13). Podobnie jak w przypadku próbek z fazy BP, zidentyfikowane warianty miały charakter typowych, opisanych w literaturze zmian. W 3 próbkach sprzed progresji CML wykryto monosomię

chromosomu 7, obserwowaną również w fazie BP w przypadku dwóch z nich (próbki 8 i 13; tabela 20, rycina 13). W próbkach sprzed progresji nie stwierdzono istotnych zmian liczby kopii genów *IKZF1*, *CDKN2A*, *CDKN2B*, *TP53* i *BCR::ABL1* obecnych w próbkach z fazy BP.

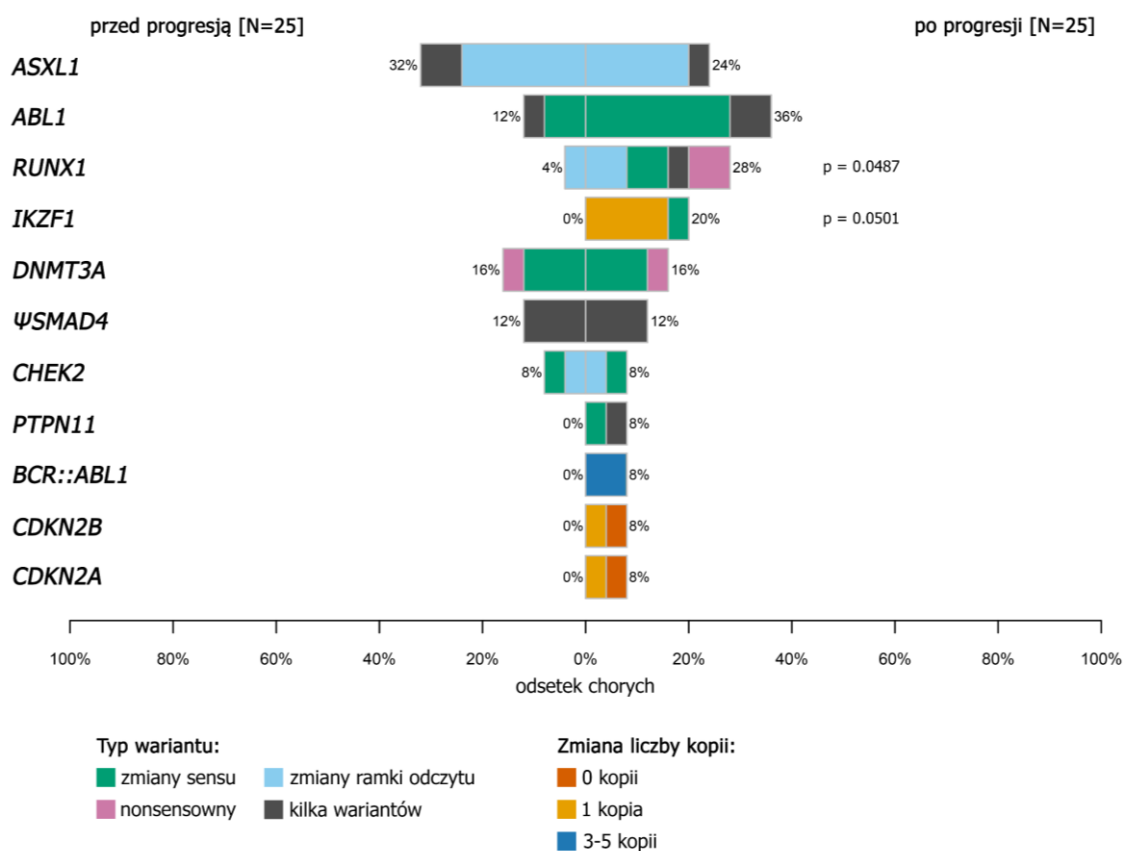


**Rycina 13. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 25 próbek CML przed progresją choroby.** Wykres przedstawia wszystkie patogene zmiany jakie zidentyfikowano w ramach analizy zmian punktowych i małych insercji-delekcji (górny panel) oraz ACA, typowych dla CML (dolny panel). Na wykresie pominięto próbki, w których nie wykryto żadnych istotnych zmian genetycznych. Kody poszczególnych próbek umieszczono pomiędzy górnym a dolnym panelem. Po prawej stronie górnego panelu podano odsetek próbek, w których wykryto patogene warianty w poszczególnych genach. Wykres wygenerowano za pomocą funkcji *oncoplot* pakietu *maftools*. Skróty; ACA- dodatkowe aberracje chromosomowe, CP- faza przewlekła, CP res- oporność wtórna w fazie przewlekłej, AP- faza akceleracji, BP 1- pierwsza faza blastyczna

#### 4.4 Analiza porównawcza wybranych zmian punktowych, małych insercji-delekcji i zmian liczby kopii w 25 sparowanych próbkach

W celu wytypowania genów, których status mutacji mógłby mieć potencjalny wpływ na niekorzystny przebieg choroby, porównano profile zmian punktowych, małych insercji-delekcji i zmian liczby kopii genów u 25 chorych, dla których dostępne były próbki sprzed i po progresji CML. Analizę wykonano dla genów zmutowanych u min. dwojga chorych a wyniki przedstawiono na rycinie 14. Wykazano wzrost ilości

wariantów patogennych w genach *ABL1* (SNV) i *RUNX1* w próbkach po progresji do fazy BP a w przypadku genu *RUNX1* wzrost ten był statystycznie istotny ( $P=0,0487$ , OR 0,11, 95% CI 0,0022 – 0,99, test Fishera). W przypadku genu *ASXL1* patogene warianty obserwowano częściej w próbkach przed progresją choroby, z kolei w genach *IKZF1*, *PTPN11*, *BCR::ABL1* (zmiany liczby kopii), *CDKN2A* i *CDKN2B* zmiany występowały jedynie w próbkach po progresji do fazy BP. Dla genu *IKZF1* zaobserwowano zmiany na granicy istotności statystycznej;  $P=0,0501$ , OR 0,00, 95% CI 0,0000 – 1,0012 (test Fishera). Przeprowadzona analiza nie wykazała różnic w ilości patogennych wariantów/próbkę pomiędzy obiema grupami w genach *DNMT3A*, *SMAD4* i *CHEK2*.

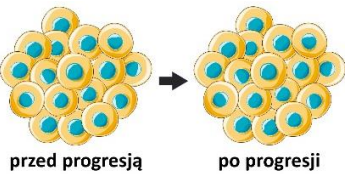
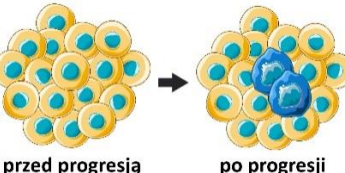
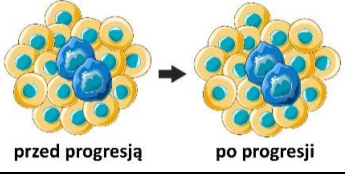


**Rycina 14. Analiza porównawcza częstości zmian punktowych, małych insercji-delekcji i zmian liczby kopii u 25 chorych w sparowanych próbkach CML sprzed i po progresji CML.** Funkcje *coBarplot* i *mafCompare* pakietu *mafTools* wykorzystano odpowiednio do wygenerowania wykresu oraz przeprowadzenia analizy statystycznej (test Fishera).

Dodatkowo, porównanie występowania patogennych wariantów pomiędzy próbkami sprzed i po progresji CML pozwoliło na wyodrębnienie 7 wzorców dynamiki patogennych zmian zachodzących w poszczególnych sparowanych próbkach (tabela 21). U jednego chorego nie zaobserwowano patogennych zmian zarówno przed jak i po

progresji CML (wzorzec 1). U 5 chorych patogenne warianty zidentyfikowano dopiero po progresji (wzorzec 2). Utrzymywanie się tych samych patogennych wariantów w próbkach sprzed, jak i po progresji CML zaobserwowano u 6 chorych (wzorzec 3), a u kolejnych 9 dochodziło do nabycia dodatkowych patogennych zmian po progresji (wzorzec 4). U jednego chorego doszło do utraty patogennych wariantów w próbce po progresji CML (wzorzec 5), a u następnych 2 chorych, w próbkach z fazy BP obserwowano utratę części patogennych wariantów oraz jednoczesne nabycie innych (wzorzec 6). Wreszcie, u 1 chorego w próbce z fazy BP obserwowano zanik patogennych wariantów obecnych przed progresją choroby i pojawienie się nowego patogennego wariantu w innym genie (wzorzec 7) co jednak jest najprawdopodobniej związane z tym, że chory został poddany procedurze przeszczepu allo-HCT.

**Tabela 21. Analiza porównawcza patogennych wariantów ukazująca dynamikę zmian zachodzących w 25 sparowanych próbkach sprzed i po progresji CML.** W kolumnie pierwszej w sposób schematyczny przedstawiono zmiany zachodzące w sparowanych próbkach sprzed i po progresji CML. Kolorem żółtym oznaczono komórki nowotworowe bez patogennych wariantów, natomiast kolorem niebieskim, czerwonym i zielonym komórki nowotworowe ze zidentyfikowanymi patogennymi zmianami.

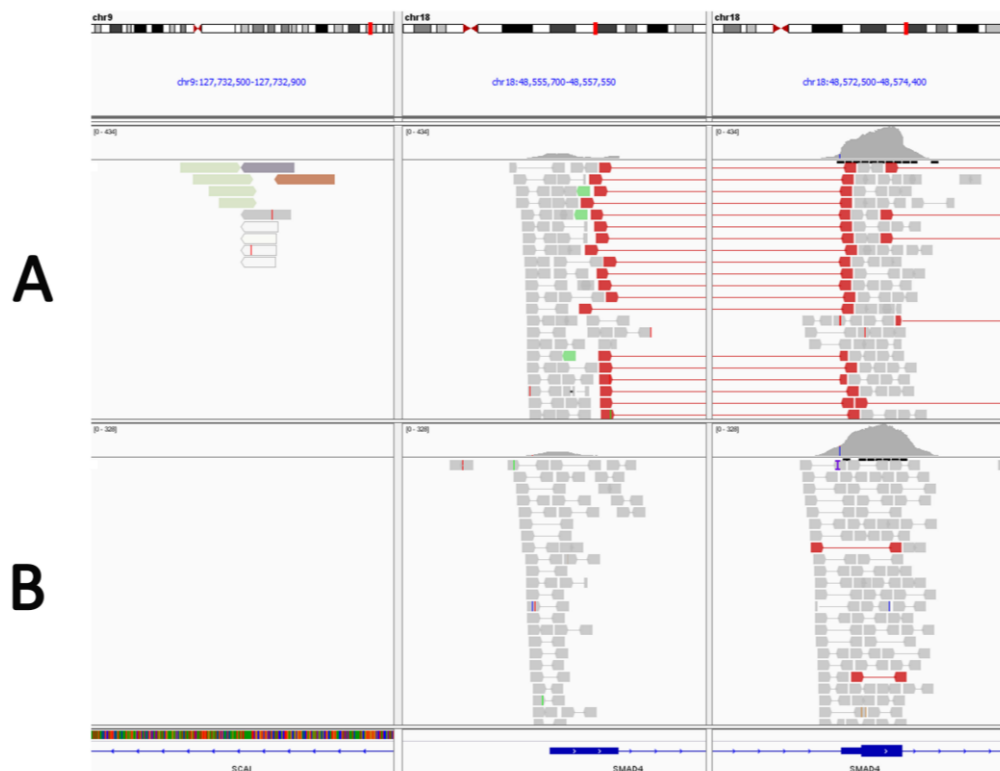
| Wzorzec   | Kod chorego | Geny z patogennymi wariantami przed progresją CML | Geny z patogennymi wariantami po progresji CML |
|---|-------------|---|--|
| 1<br><br>przed progresją      po progresji | 2           | brak  | brak   |
| 2<br><br>przed progresją      po progresji | 3           | brak  | <i>ABL1, RUNX1, LRP1B</i>                      |
|   | 5           | brak  | <i>ABL1</i>                                    |
|   | 7           | brak  | <i>RUNX1, KMT2D</i>                            |
|   | 13          | brak  | <i>CSF3R, RUNX1</i>                            |
|   | 16          | brak  | <i>RUNX1</i>                                   |
| 3<br><br>przed progresją      po progresji | 4           | <i>ASXL1, DNMT3A</i>                              | <i>ASXL1, DNMT3A</i>                           |
|   | 9           | <i>ABL1, ASXL1, NF1, BRCA2</i>                    | <i>ABL1, ASXL1, NF1, BRCA2</i>                 |
|   | 10          | <i>ΨSMAD4</i>                                     | <i>ΨSMAD4</i>                                  |
|   | 12          | <i>TET2</i>                                       | <i>TET2</i>                                    |

|   |   |   |  |
|---|---|---|--|
|   | 15  | <i>ASXL1, GATA2, CCND2, WT1</i>         | <i>ASXL1, GATA2, CCND2, WT1</i>                              |
|   | 20  | <i>ASXL1, DNMT3A</i>                    | <i>ASXL1, DNMT3A</i>   |
| <p>4</p> <p>przed progresją      po progresji</p> | 1   | $\Psi$ <i>SMAD4, ASXL1</i>              | $\Psi$ <i>SMAD4, ASXL1, ABL1, RUNX1</i>                      |
|   | 8   | $\Psi$ <i>SMAD4, DNMT3A, CHEK2, MAX</i> | $\Psi$ <i>SMAD4, DNMT3A, CHEK2, MAX, ABL1, PTPN11, SYNE2</i> |
|   | 11  | <i>DNMT3A, MSH6</i>                     | <i>DNMT3A, MSH6, PTPN11</i>                                  |
|   | 14  | <i>CHEK2</i>                            | <i>CHEK2, ABL1</i>   |
|   | 18  | <i>ABL1</i>                             | <i>ABL1</i> (kilka wariantów)                                |
|   | 21  | <i>TNFRSF14</i>                         | <i>TNFRSF14, ABL1</i>  |
|   | 23  | <i>RUNX1</i>                            | <i>RUNX1, IKZF1</i>  |
|   | 24  | <i>ASXL1</i>                            | <i>ASXL1, PHF6</i>   |
|   | 25  | <i>KDM6A</i>                            | <i>KDM6A, ERBB2, IDH2</i>                                    |
|   | <p>5</p> <p>przed progresją      po progresji</p> | 17                                      | <i>PTPRJ, ASXL1</i> (kilka wariantów)                        |
| <p>6</p> <p>przed progresją      po progresji</p> | 6   | <i>IDH1, IDH2</i>                       | <i>IDH1, RUNX1</i>   |
|   | 19  | <i>MLH1, STAG2</i>                      | <i>MLH1, ABL1</i>  |
| <p>7</p> <p>przed progresją      po progresji</p> | 22  | <i>ABL1, ASXL1</i>                      | <i>NSD1</i>  |

#### 4.5 Ocena częstości występowania $\Psi$ *SMAD4* w różnych rodzajach białaczek

Oprócz dobrze scharakteryzowanych mutacji w genach *ABL1*, *ASXL1*, *IKZF1*, *RUNX1*, *TP53* czy *DNMT3A*, analiza danych uzyskanych z sekwencjonowania NGS ujawniła obecność licznych wariantów miejsc splicingowych w genie *SMAD4* w 12%

(3/25) sparowanych próbek (ryciny 11-13). Dokładna analiza tych wariantów wykazała, że w rzeczywistości reprezentowały one obecność  $\Psi$ SMAD4, który jak dotąd nie został opisany w nowotworach złośliwych. Warianty zadnotowane jako warianty w miejscach splicingowych SMAD4 odpowiadały sparowanym odczytom mapującym się na końcach sąsiadujących egzonów SMAD4 (tzw. odczytom na złączach), wskazującym na brak sekwencji intronowych i tym samym sugerującym obecność przetworzonego pseudogenu (rycina 15A, część centralna i prawa strona). Dodatkowo, wyłącznie w próbkach z  $\Psi$ SMAD4 zidentyfikowano odczyty sprzeczne, w których jeden odczyt z pary mapował się do genu SMAD4 położonego na chromosomie 18, a drugi odczyt z pary mapował się do genu SCAI zlokalizowanego na chromosomie 9, tj. w miejscu insercji pseudogenu (rycina 15A, lewa strona i część centralna).

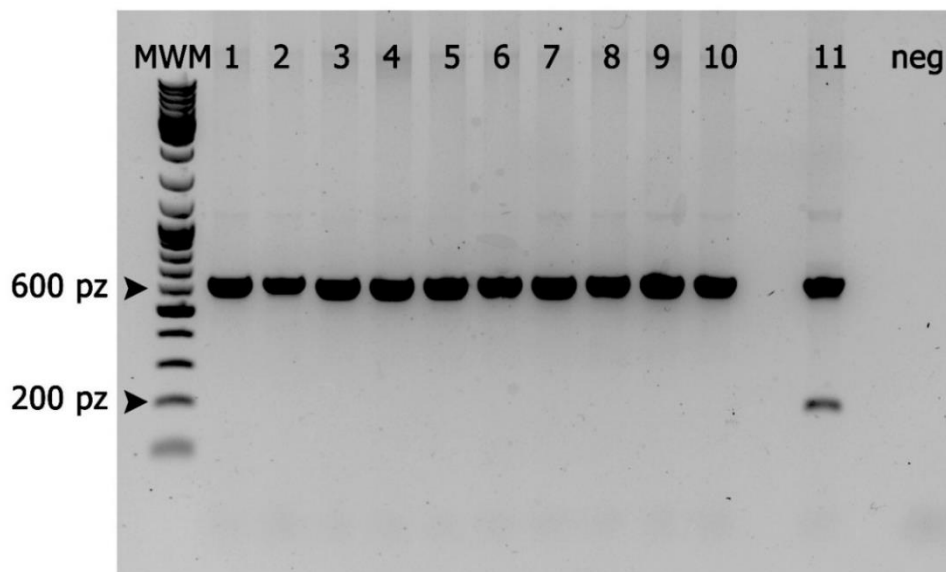


**Rycina 15. Przykładowa analiza wizualna obecności  $\Psi$ SMAD4 w badanych próbkach za pomocą sekwencjonowania NGS. Zrzut ekranu z programu IGV: (A) próbka z  $\Psi$ SMAD4, (B) próbka bez  $\Psi$ SMAD4. Kolorem zielonym zaznaczono odczyty sprzeczne wskazujące na obecność  $\Psi$ SMAD4, z których jeden mapuje się w miejscu insercji pseudogenu, tj. w intronie 18 genu SCAI (część lewa) a drugi w egzonie 1 genu SMAD4 (część centralna). Kolorem czerwonym oznaczone są pary odczytów zlokalizowanych na końcach sąsiadujących egzonów, tzw. odczyty na złączach, wskazujące na brak sekwencji intronowej pomiędzy nimi (część centralna i prawa).**

Dane z sekwencjonowania NGS zgromadzone w wewnętrznej bazie Zakładu Genetyki Medycznej WUM pozwoliły na retrospektywną analizę częstości

występowania  $\Psi$ SMAD4 zarówno w próbkach nowotworowych jak i nienowotworowych - łącznie ok. 2 tys. próbek, z czego zdecydowaną większość stanowiły próbki nienowotworowe.  $\Psi$ SMAD4 został wykryty jedynie w 3 próbkach CML oraz w 2 próbkach nienowotworowych pochodzących od krewnych pierwszego stopnia (pseudogenu nie wykryto w żadnym innym typie nowotworu). Ponieważ w przypadku CML, wszystkie trzy próbki z  $\Psi$ SMAD4 pochodziły od pacjentów mających progresję do fazy blastycznej, postanowiono sprawdzić, czy obserwowany pseudogen jest nowym markerem w CML lub ma związek z progresją tej choroby.

Ostatecznie, częstość występowania  $\Psi$ SMAD4 określano głównie za pomocą reakcji PCR (opisanej w podrozdziale 3.2.3 *Reakcja łańcuchowej polimerazy PCR do oceny obecności  $\Psi$ SMAD4*), rzadziej wykorzystując wyłącznie metodę sekwencjonowania NGS. Część próbek została przebadana obiema metodami w celu ich walidacji. Przykładowy wynik reakcji PCR umożliwiającej identyfikację  $\Psi$ SMAD4 przedstawiono na rycinie 16.



**Rycina 16. Przykładowy obraz elektroforetyczny przedstawiający wynik reakcji PCR oceniającej obecność  $\Psi$ SMAD4 w badanych próbkach.** MWM- marker wielkości mas (GeneRuler™ DNA Ladder Mix, Thermo Fisher Scientific), strzałkami oznaczono prążki o wielkości 200 i 600 pz; ścieżki: 1-10- próbki bez  $\Psi$ SMAD4, amplikon o długości 623 pz; 11- próbka z  $\Psi$ SMAD4, amplikony o długości 623 pz i 189 pz; neg- kontrola negatywna reakcji PCR. Kolory odwrócone.

Początkowo, częstość występowania  $\Psi$ SMAD4 oceniono w próbkach CML pochodzących od polskich pacjentów i porównano ją z częstością populacyjną, którą określono na podstawie analizy próbek zgromadzonych w Zakładzie Genetyki



Medycznej WUM w ramach uczestnictwa w Wieloośrodkowym Ogólnopolskim Badaniu Stanu Zdrowia Ludności (WOBASZ)<sup>1</sup>. Częstość występowania  $\Psi$ SMAD4 w CML wyniosła 1,35% (9 z 669 próbek) i była 2,5-krotnie wyższa od częstości badanego pseudogenu w populacji polskiej, która wyniosła 0,54% (12 z 2242 próbek). Do osób z grupy kontrolnej, u których wykryto  $\Psi$ SMAD4 wysłano ankiety z pytaniami pozwalającymi wykluczyć obecność choroby nowotworowej, w szczególności białaczki, u badanych osób lub najbliższych członków ich rodzin. Przeprowadzony wywiad wykluczył obecność choroby nowotworowej w przypadku wszystkich osób, które odpowiedziały na pytania zawarte w ankiecie (n= 9).

W związku ze zwiększoną częstością  $\Psi$ SMAD4 w CML, badanie postanowiono poszerzyć o próbki z innych, najczęstszych nowotworów układu krwiotwórczego, tj. AML oraz ALL i przewlekłej białaczki limfocytowej (CLL, ang. *chronic lymphocytic leukemia*). Dodatkowo, dla dokładniejszej oceny częstości występowania  $\Psi$ SMAD4, nieograniczającej się jedynie do polskiej populacji, badanie wykonano również dla 273 czeskich próbek CML. Uzyskane wyniki oceny częstości występowania  $\Psi$ SMAD4 w różnych rodzajach białaczek przedstawiono w tabeli 22.

**Tabela 22. Częstość występowania  $\Psi$ SMAD4 w różnych rodzajach białaczek.** CML- przewlekła białaczka szpikowa, CLL- przewlekła białaczka limfocytowa, AML- ostra białaczka szpikowa, ALL- ostra białaczka limfoblastyczna. Wyniki dla CML przedstawiono osobno dla próbek polskich i czeskich, oraz po połączeniu tych danych. Znamienność statystyczna względem próbek kontrolnych została oznaczona za pomocą testu dokładnego Fishera (*Fisher's exact test*).

| Grupa                       |                           | Liczba zbadanych próbek | Liczba próbek z $\Psi$ SMAD4 | Wartość p względem grupy kontrolnej |
|-----------------------------|---------------------------|-------------------------|------------------------------|-------------------------------------|
| CML                         | populacja polska          | 669                     | 9 (1,35%)                    | <b>0,034</b>                        |
|                             | populacja czeska          | 273                     | 3 (1,1%)                     | 0,217                               |
|                             | populacja polska + czeska | 942                     | 12 (1,27%)                   | <b>0,028</b>                        |
| CLL                         |                           | 458                     | 2 (0,44%)                    | 0,567                               |
| AML                         |                           | 614                     | 4 (0,65%)                    | 0,462                               |
| ALL                         |                           | 641                     | 5 (0,78%)                    | 0,321                               |
| kontrola (populacja polska) |                           | 2242                    | 12 (0,54%)                   |                                     |

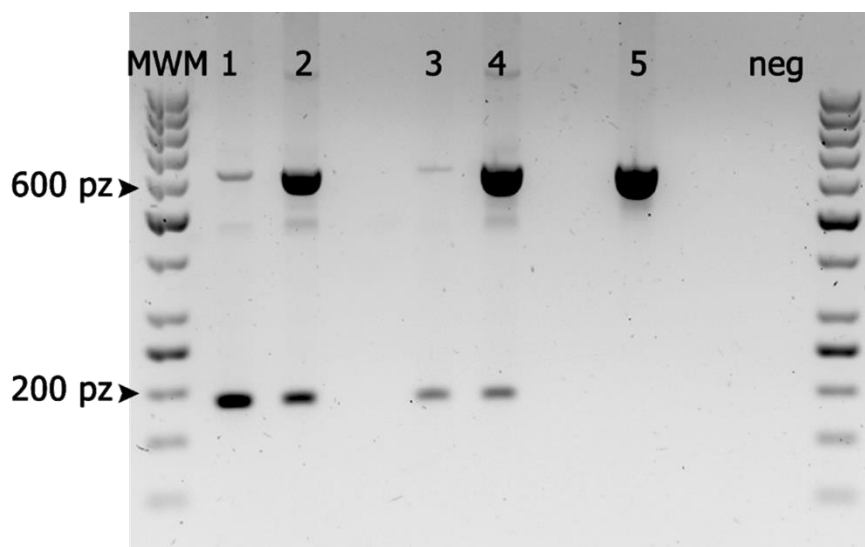
<sup>1</sup> Celem badania WOBASZ była ocena czynników ryzyka chorób sercowo-naczyniowych a do badania włączono reprezentatywną liczbę osób z całej Polski w wieku powyżej 19 roku życia.

Ostatecznie, wyniki istotne statystycznie otrzymano jedynie dla próbek CML (populacja polska i polska połączona z czeską). Dlatego też, w dalszych badaniach dotyczących  $\Psi$ SMAD4 analizowano już wyłącznie 9 polskich chorych na CML z badanym pseudogenem, dla których została zabezpieczona odpowiednia ilość materiału genetycznego do kolejnych badań.

#### 4.6 Potwierdzenie zarodkowego charakteru $\Psi$ SMAD4

W następnym etapie prowadzonych badań postanowiono potwierdzić zarodkowy (obecny nie tylko w komórkach nowotworowych, ale we wszystkich komórkach ciała) charakter  $\Psi$ SMAD4. W tym celu ponownie wykonano reakcję PCR do określania częstości występowania  $\Psi$ SMAD4 (opisaną w podrozdziale 3.2.3 *Reakcja łańcuchowej polimerazy PCR do oceny obecności  $\Psi$ SMAD4*), jednak tym razem do reakcji użyto DNA pochodzenia zarodkowego, które wyizolowano z cebulek włosowych (materiał możliwy był do uzyskania jedynie od 2 chorych na CML z pseudogenem).

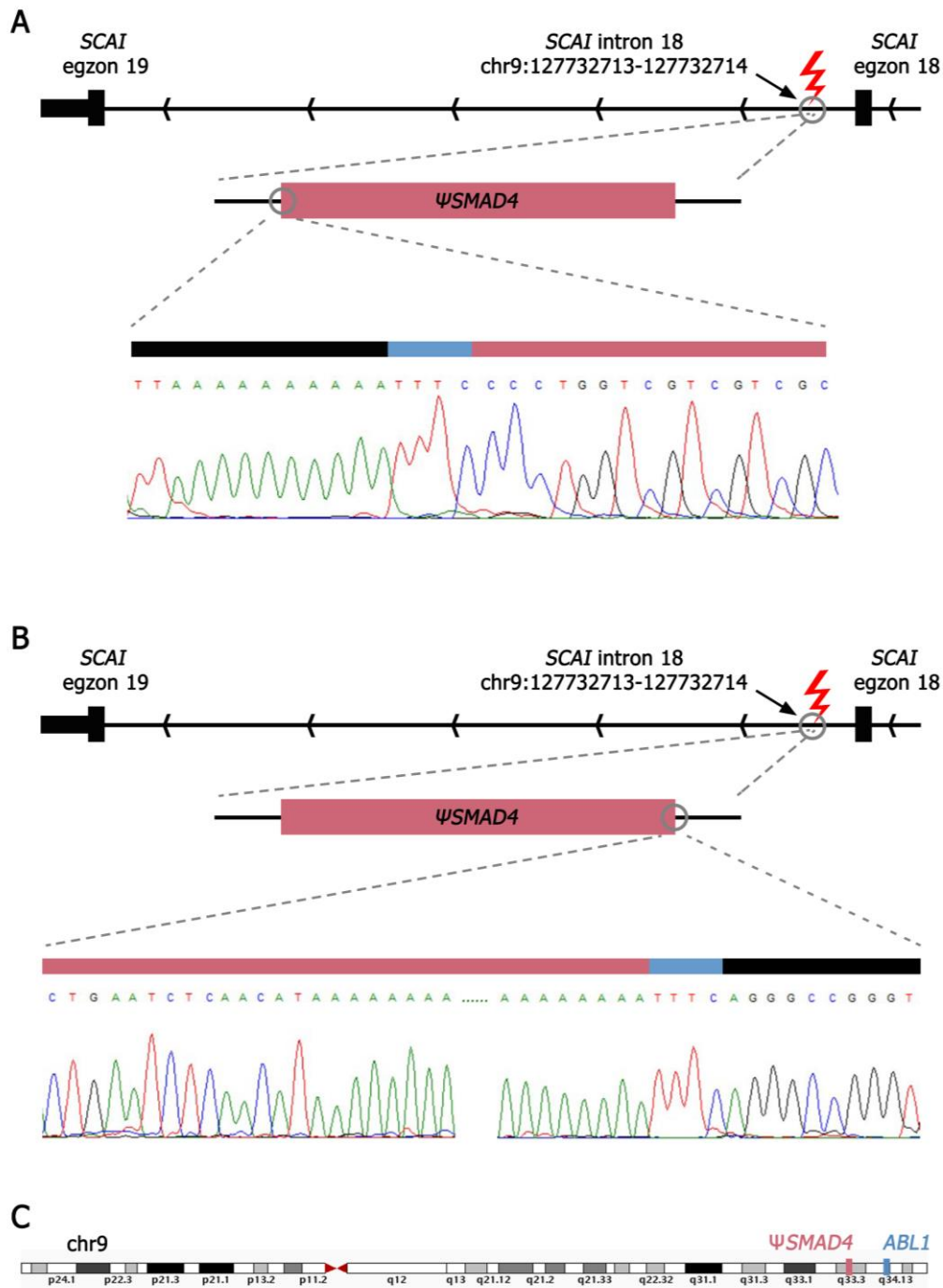
Przeprowadzona analiza wykazała obecność prążka pochodzącego z pseudogenu zarówno w próbkach z DNA zarodkowym jak i nowotworowym, co potwierdziło zarodkowy charakter  $\Psi$ SMAD4 u obojga chorych. Zdjęcie obrazu elektroforetycznego prezentującego uzyskane wyniki przedstawia rycina 17.



**Rycina 17. Obraz elektroforetyczny przedstawiający wyniki reakcji PCR potwierdzającej zarodkowy charakter  $\Psi$ SMAD4.** MWM- marker wielkości mas (GeneRuler™ 50 bp DNA Ladder, Thermo Fisher Scientific), strzałkami oznaczono prążki o wielkości 200 i 600 pz; ścieżki: 1 i 3- DNA zarodkowe wyizolowane z cebulek włosowych chorych o numerach odpowiednio 47 i 50 ; 2 i 4- DNA nowotworowe wyizolowane z krwi chorych o numerach odpowiednio 47 i 50; 5- DNA nowotworowe wyizolowane z krwi chorego na CML bez  $\Psi$ SMAD4; neg- kontrola negatywna reakcji PCR. Kolory odwrócone.

## 4.7 Określenie genomowej lokalizacji $\Psi$ SMAD4

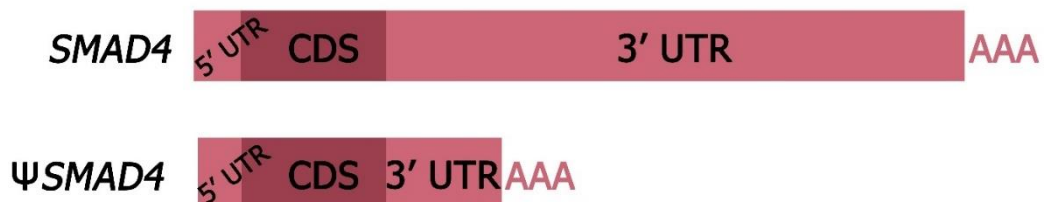
Przybliżoną lokalizację przetworzonego  $\Psi$ SMAD4 określono na podstawie danych uzyskanych z sekwencjonowania NGS, natomiast dokładne miejsce insercji określono poprzez zsekwencjonowanie metodą Sangera produktów PCR powstałych w wyniku amplifikacji regionów na złączach (rycina 7). Gen SMAD4 zlokalizowany jest na długim ramieniu chromosomu 18. Analiza wyników sekwencjonowania NGS wykazała obecność odczytów sprzecznych w genie SMAD4, dla których pary mapowały się na chromosomie 9 w obrębie genu SCAI (*Suppressor of cancer cell invasion*), a dokładniej - w jego 18 intronie (NM\_173690.5). Na tej podstawie zaprojektowano startery do reakcji PCR, które umożliwiły precyzyjne określenie miejsca insercji pseudogenu oraz sekwencji na złączach (tj. fragmentów sekwencji  $\Psi$ SMAD4 i SCAI, które uległy połączeniu). Przeprowadzona analiza wykazała, że miejsce pęknięcia genu SCAI, będące jednocześnie miejscem insercji  $\Psi$ SMAD4, zlokalizowane jest w ostatnim intronie, w pozycji chromosomowej chr9:127732713-127732714 (rycina 18). Wykazano, że w miejscu pęknięcia intronu SCAI, dochodzi do jego połączenia z sekwencją  $\Psi$ SMAD4, odpowiadającą fragmentowi mRNA genu SMAD4 (NM\_005359.6), pochodzącemu z egzonu 1 (połączenie SCAI- $\Psi$ SMAD4, rycina 18A) i egzonu 12 (połączenie  $\Psi$ SMAD4-SCAI, rycina 18B). W miejscu połączenia  $\Psi$ SMAD4-SCAI pomiędzy sekwencjami  $\Psi$ SMAD4 i SCAI zidentyfikowano sekwencję poli-A. Potwierdzono, że miejsce insercji  $\Psi$ SMAD4 oraz sekwencje na złączach są identyczne u wszystkich 9 pacjentów CML z badanym pseudogenem. Ponieważ  $\Psi$ SMAD4 ulega insercji na chromosomie 9, porównano jego lokalizację z położeniem genu ABLI zaangażowanego w powstawanie chromosomu Ph. Ustalono, że  $\Psi$ SMAD4 zlokalizowany jest proksymalnie od ABLI o niecałe 6 Mpz, co schematycznie przedstawiono na rycinie 18C.



**Rycina 18. Schemat genomowej lokalizacji  $\Psi$ SMAD4.** (A) połączenie *SCAI*- $\Psi$ SMAD4, (B) połączenie  $\Psi$ SMAD4-*SCAI*, (C) położenie względem *ABL1*. Na rycinach 18A-B, czerwona błyskawica oznacza miejsce insercji  $\Psi$ SMAD4 w 18 intronie genu *SCAI*. Czarny i różowy kolor prostokątów nad chromatogramami określa pochodzenie sekwencji odpowiednio z genu *SCAI* i  $\Psi$ SMAD4. Prostokąty w kolorze błękitnym odpowiadają wstawce będącej duplikacją miejsca insercji (TSD, ang. *target site duplications*). Na rycinie 18B zielonymi kropkami w sekwencji oznaczono sekwencję poli-A o różnej długości w każdej badanej próbce. Rycina 18C przedstawia schemat chromosomu 9, na którym zaznaczono lokalizację  $\Psi$ SMAD4 względem genu *ABL1*.

## 4.8 Określenie pełnej sekwencji $\Psi$ SMAD4 wraz z miejscem insercji

Powszechnie wiadomo, że pseudogeny różnią się od genów rodzicielskich obecnością mutacji zakłócających powstawanie funkcjonalnego białka. Dlatego też, kolejnym etapem badań było określenie całkowitej sekwencji  $\Psi$ SMAD4. W tym celu trzy amplikony, obejmujące cały  $\Psi$ SMAD4 lub jego część, zsekwencjonowano metodą NGS a uzyskane sekwencje złożono *de novo*, bez wykorzystania genomu referencyjnego, jak opisano w podrozdziale 3.2.6 *Określenie sekwencji  $\Psi$ SMAD4 za pomocą sekwencjonowania NGS*. Przeprowadzona analiza wykazała, że sekwencja  $\Psi$ SMAD4 ma długość 3466 pz i w porównaniu do sekwencji mRNA genu rodzicielskiego SMAD4 jest krótsza o początkowe 41 pz na końcu 5'UTR oraz 5265 ostatnich pz na końcu 3'UTR, co zostało schematycznie przedstawione na rycinie 19. Na końcu 3'  $\Psi$ SMAD4 występuje sekwencja poli-A o długości ok. 60-90 pz.



**Rycina 19. Schematyczne porównanie struktury  $\Psi$ SMAD4 i mRNA genu SMAD4.** Końce 5'UTR i 3'UTR pseudogenu są krótsze o odpowiednio 41 i 5265 pz w porównaniu do mRNA genu SMAD4 (NM\_005359.6). Kolorem jasnoróżowym oznaczono fragmenty odpowiadające rejonom 5' i 3' niepodlegającym translacji (5'UTR, 3'UTR, ang. 5', 3' *untranslated region*), natomiast kolorem ciemnoróżowym regionowi kodującemu (CDS, ang. *coding DNA sequence*).

Cała sekwencja kodująca białko została zachowana i jest identyczna w przypadku genu i pseudogenu – nie wykryto żadnych wariantów SNV, występujących powszechnie w sekwencjach pseudogenów. Rycina 20 przedstawia porównanie całkowitej sekwencji  $\Psi$ SMAD4 z odpowiadającym fragmentem sekwencji mRNA genu SMAD4.

|                |      |  |      |
|----------------|------|--|------|
| SMAD4          | 1    | ATGCTCAAGTGGCTTCTCGACAAGTTGGCAGCAACACACGGCCCTGGTGGTGGTGGTGGTGG   | 80   |
| <i>ψ</i> SMAD4 | 1    | .....CCCTGGTGGTGGTGGTGGTGGTGGTGGTGG                              | 19   |
| SMAD4          | 81   | TGCGGTAACGGAGCGGTTTGGGTGGCGGAGCCTGCCGTTCCCGCTCTCCCTCGG           | 120  |
| <i>ψ</i> SMAD4 | 20   | TGCGGTAACGGAGCGGTTTGGGTGGCGGAGCCTGCCGTTCCCGCTCTCCCTCGG           | 79   |
| SMAD4          | 121  | GAGGCCCTTCTGCTCTCCCTAGGCTCCGCGGCGCCAGGGGGTGGGAGCGGGTGAGG         | 180  |
| <i>ψ</i> SMAD4 | 80   | GAGGCCCTTCTGCTCTCCCTAGGCTCCGCGGCGCCAGGGGGTGGGAGCGGGTGAGG         | 139  |
| SMAD4          | 181  | GGAGCCAGGCGCCACGAGAGAGGCCCCCGCCGCGCAGGGCGGCCGGGGAGCTCGAGGGC      | 240  |
| <i>ψ</i> SMAD4 | 140  | GGAGCCAGGCGCCACGAGAGAGGCCCCCGCCGCGCAGGGCGGCCGGGGAGCTCGAGGGC      | 199  |
| SMAD4          | 241  | GTCCGGCCCGCGCGGGCAGCGCGCGGGCGCTGAGGAGGGGCGGCCTGGCCGGGAGCCCTC     | 300  |
| <i>ψ</i> SMAD4 | 200  | GTCCGGCCCGCGCGGGCAGCGCGCGGGCGCTGAGGAGGGGCGGCCTGGCCGGGAGCCCTC     | 259  |
| SMAD4          | 301  | GGGGCGGGGCGGAGGAGCTCTCCGGCGCGCGGGGAAAGCTACGGGCGCGGTGCGTCCG       | 360  |
| <i>ψ</i> SMAD4 | 280  | GGGGCGGGGCGGAGGAGCTCTCCGGCGCGCGGGGAAAGCTACGGGCGCGGTGCGTCCG       | 319  |
| SMAD4          | 381  | GAGACGAGAGCGCGGGAGAGCGGACTCCCTCGCCACCGCCGAGCCAGGTTATCCTG         | 420  |
| <i>ψ</i> SMAD4 | 320  | GAGACGAGAGCGCGGGAGAGCGGACTCCCTCGCCACCGCCGAGCCAGGTTATCCTG         | 379  |
| SMAD4          | 421  | AATACATGCTAACAAATTTCCCTGCAACGTTAGCTGTTGTTTTCACTGTTTCCAAAGG       | 480  |
| <i>ψ</i> SMAD4 | 380  | AATACATGCTAACAAATTTCCCTGCAACGTTAGCTGTTGTTTTCACTGTTTCCAAAGG       | 439  |
| SMAD4          | 481  | ATCAAATTTGCTTCAGAATTTGGAGACATATTTGATTTAAAGGAAAGAACTTGAACAAAT     | 540  |
| <i>ψ</i> SMAD4 | 440  | ATCAAATTTGCTTCAGAATTTGGAGACATATTTGATTTAAAGGAAAGAACTTGAACAAAT     | 499  |
| SMAD4          | 541  | GGACAATATGCTATTACGAATACACCAACAGTAATGATGCCGTGCTGAGCATTGTGCA       | 600  |
| <i>ψ</i> SMAD4 | 500  | GGACAATATGCTATTACGAATACACCAACAGTAATGATGCCGTGCTGAGCATTGTGCA       | 559  |
| SMAD4          | 601  | TAGTTTGATGTGCCATAGACAAGGTGGAGAGAGTAAACATTTGCCAAAAAGAGCAATTTGA    | 660  |
| <i>ψ</i> SMAD4 | 560  | TAGTTTGATGTGCCATAGACAAGGTGGAGAGAGTAAACATTTGCCAAAAAGAGCAATTTGA    | 619  |
| SMAD4          | 661  | AAGTTTGGTAAAGAAGCTGAAGGAGAAAAAGATGAATTGGATTCTTTAATAACAGCTAT      | 720  |
| <i>ψ</i> SMAD4 | 620  | AAGTTTGGTAAAGAAGCTGAAGGAGAAAAAGATGAATTGGATTCTTTAATAACAGCTAT      | 679  |
| SMAD4          | 721  | AACACAAATGGAGCTCATCCTAGTAAATGTGTTACCATACAGAGAACAATTGGATGGGAG     | 780  |
| <i>ψ</i> SMAD4 | 680  | AACACAAATGGAGCTCATCCTAGTAAATGTGTTACCATACAGAGAACAATTGGATGGGAG     | 739  |
| SMAD4          | 781  | GCTTCAAGGTGGCTGGTCGGAAAGGATTTCCCTCATGTGATCTATGCCCGTCTCTGGAGGTG   | 840  |
| <i>ψ</i> SMAD4 | 740  | GCTTCAAGGTGGCTGGTCGGAAAGGATTTCCCTCATGTGATCTATGCCCGTCTCTGGAGGTG   | 799  |
| SMAD4          | 841  | GCCTGATCTTCACAAAAATGAACATAAACATGTTAAATATGTCAGTATGCCGTTTGAATT     | 900  |
| <i>ψ</i> SMAD4 | 800  | GCCTGATCTTCACAAAAATGAACATAAACATGTTAAATATGTCAGTATGCCGTTTGAATT     | 859  |
| SMAD4          | 901  | AAAAATGATAGTGTCTGTGTGAATCCATATCCTACGAAACGAGTTGATCACCTGGAAAT      | 960  |
| <i>ψ</i> SMAD4 | 860  | AAAAATGATAGTGTCTGTGTGAATCCATATCCTACGAAACGAGTTGATCACCTGGAAAT      | 919  |
| SMAD4          | 961  | TGATCTCTCAGGATTAACACTGCAGAGTAATGCTCCATCAAGTATGATGGTGAAAGGATGA    | 1020 |
| <i>ψ</i> SMAD4 | 920  | TGATCTCTCAGGATTAACACTGCAGAGTAATGCTCCATCAAGTATGATGGTGAAAGGATGA    | 979  |
| SMAD4          | 1021 | ATATGTGCATGACTTTGAAGGACAGCCATCGTTGTCCACTGAAGGACATTCAATTCAAAC     | 1080 |
| <i>ψ</i> SMAD4 | 980  | ATATGTGCATGACTTTGAAGGACAGCCATCGTTGTCCACTGAAGGACATTCAATTCAAAC     | 1039 |
| SMAD4          | 1081 | CATCCAGCATCCACCAAGTAATCGTGCATCGACAGAGACATACAGCACCCGCTCTGTT       | 1140 |
| <i>ψ</i> SMAD4 | 1040 | CATCCAGCATCCACCAAGTAATCGTGCATCGACAGAGACATACAGCACCCGCTCTGTT       | 1099 |
| SMAD4          | 1141 | AGCCCCATCTGAGTCTAATGTCTACCAGCACTGCCAATTTCCCAACATTCTCTGTTGTTT     | 1200 |
| <i>ψ</i> SMAD4 | 1100 | AGCCCCATCTGAGTCTAATGTCTACCAGCACTGCCAATTTCCCAACATTCTCTGTTGTTT     | 1159 |
| SMAD4          | 1201 | CACAAGTCAGCCTGCCAGTATACTGGGGGGCAGCCATAGTGAAGGACTGTTGCAGATAGC     | 1260 |
| <i>ψ</i> SMAD4 | 1160 | CACAAGTCAGCCTGCCAGTATACTGGGGGGCAGCCATAGTGAAGGACTGTTGCAGATAGC     | 1219 |
| SMAD4          | 1261 | ATCAGGGCCCTCAGCCAGGACAGCAGCAGAATGGATTTACTGGTCAAGCCAGCTACTTACCA   | 1320 |
| <i>ψ</i> SMAD4 | 1220 | ATCAGGGCCCTCAGCCAGGACAGCAGCAGAATGGATTTACTGGTCAAGCCAGCTACTTACCA   | 1279 |
| SMAD4          | 1321 | TCATAACAGCACTACCACCTGGACTGGAAGTGGAGCTGCACCATACACACCTAATTTGCC     | 1380 |
| <i>ψ</i> SMAD4 | 1280 | TCATAACAGCACTACCACCTGGACTGGAAGTGGAGCTGCACCATACACACCTAATTTGCC     | 1339 |
| SMAD4          | 1381 | TCACCACCAAAAACGGCCATCTTCCAGCACCACCAGCCATATGCCGCCCCCATCCCCGACATTA | 1440 |
| <i>ψ</i> SMAD4 | 1340 | TCACCACCAAAAACGGCCATCTTCCAGCACCACCAGCCATATGCCGCCCCCATCCCCGACATTA | 1399 |
| SMAD4          | 1441 | CTGGCCCTTTCACAATGAGCTTGCATTTCCAGCCTCCATTTCCAATCATCCTGCTCTGA      | 1500 |
| <i>ψ</i> SMAD4 | 1400 | CTGGCCCTTTCACAATGAGCTTGCATTTCCAGCCTCCATTTCCAATCATCCTGCTCTGA      | 1459 |
| SMAD4          | 1501 | GATTTGGTGTTCCATTGCTTACTTTGAAATGGATGTTTCAAGGTAGGAGAGACATTTAAGGT   | 1560 |
| <i>ψ</i> SMAD4 | 1460 | GATTTGGTGTTCCATTGCTTACTTTGAAATGGATGTTTCAAGGTAGGAGAGACATTTAAGGT   | 1519 |
| SMAD4          | 1561 | TCCTTCAAGCTGCCCTATTGTTACTGTTGATGGATACGTGGACCTCTCTGGAGGAGATCG     | 1620 |
| <i>ψ</i> SMAD4 | 1520 | TCCTTCAAGCTGCCCTATTGTTACTGTTGATGGATACGTGGACCTCTCTGGAGGAGATCG     | 1579 |
| SMAD4          | 1621 | CTTTTGTGTTGGTCAACTCTCCAATGTCCACAGGACAGAAAGCCATTGAGAGAGCAAGGTT    | 1680 |
| <i>ψ</i> SMAD4 | 1580 | CTTTTGTGTTGGTCAACTCTCCAATGTCCACAGGACAGAAAGCCATTGAGAGAGCAAGGTT    | 1639 |
| SMAD4          | 1681 | GCACATAGGCAAAGGTGTCAGTTGGAAATGTAAAGGTGAAGGTGATGTTTGGTTCAGGTG     | 1740 |
| <i>ψ</i> SMAD4 | 1640 | GCACATAGGCAAAGGTGTCAGTTGGAAATGTAAAGGTGAAGGTGATGTTTGGTTCAGGTG     | 1699 |
| SMAD4          | 1741 | CCTTAGTGACCACGGCTTTTGTACAGAGTTACTACTTAGACAGAGAAAGCTGGGGCTGCG     | 1800 |
| <i>ψ</i> SMAD4 | 1700 | CCTTAGTGACCACGGCTTTTGTACAGAGTTACTACTTAGACAGAGAAAGCTGGGGCTGCG     | 1759 |
| SMAD4          | 1801 | ACCTGGAGATGCTGTTTCAAAAGATCTACCCAAGTGCATATATAAAGGTCTTTGATTTGCG    | 1860 |
| <i>ψ</i> SMAD4 | 1760 | ACCTGGAGATGCTGTTTCAAAAGATCTACCCAAGTGCATATATAAAGGTCTTTGATTTGCG    | 1819 |

|                     |      |   |      |
|---------------------|------|---|------|
| <i>SMAD4</i>        | 1881 | TCAGTGTTCATCGACAGATGCAGCAGCAGGCGGGCTACTGCACAAAGCTGCAGCAGCTGCCCA | 1920 |
| $\Psi$ <i>SMAD4</i> | 1820 | TCAGTGTTCATCGACAGATGCAGCAGCAGGCGGGCTACTGCACAAAGCTGCAGCAGCTGCCCA | 1879 |
| <i>SMAD4</i>        | 1921 | GGCAGCAGCCGTGGCAGGAAACATCCCTGGCCDAGGATCAGTAGGTGGAATAGCTCCAGC    | 1980 |
| $\Psi$ <i>SMAD4</i> | 1880 | GGCAGCAGCCGTGGCAGGAAACATCCCTGGCCDAGGATCAGTAGGTGGAATAGCTCCAGC    | 1939 |
| <i>SMAD4</i>        | 1981 | TATCAGTCTGTCTCAGCTGCTGCTGGAAATTGGTGTGGATGACCTTCGTCGCTTATGCATACT | 2040 |
| $\Psi$ <i>SMAD4</i> | 1940 | TATCAGTCTGTCTCAGCTGCTGCTGGAAATTGGTGTGGATGACCTTCGTCGCTTATGCATACT | 1999 |
| <i>SMAD4</i>        | 2041 | CAGGATGAGTTTTGTGAAAGGCTGGGGACCGGATTACCCAAGACAGAGCATCAAAGAAAC    | 2100 |
| $\Psi$ <i>SMAD4</i> | 2000 | CAGGATGAGTTTTGTGAAAGGCTGGGGACCGGATTACCCAAGACAGAGCATCAAAGAAAC    | 2059 |
| <i>SMAD4</i>        | 2101 | ACCTTGCTGGATTGAAATTCACCTTACACCGGGCCCTCCAGCTCCTAGACGAAAGTACTTCA  | 2160 |
| $\Psi$ <i>SMAD4</i> | 2080 | ACCTTGCTGGATTGAAATTCACCTTACACCGGGCCCTCCAGCTCCTAGACGAAAGTACTTCA  | 2119 |
| <i>SMAD4</i>        | 2161 | TACCATGCCGATTGCAGACCCACAACCTTTAGACTGAGGCTTTTTACCGTTGGGGCCCTT    | 2220 |
| $\Psi$ <i>SMAD4</i> | 2120 | TACCATGCCGATTGCAGACCCACAACCTTTAGACTGAGGCTTTTTACCGTTGGGGCCCTT    | 2179 |
| <i>SMAD4</i>        | 2221 | AACCTTATCAGGATGGTGGACTACAAAATACAATCCTGTTTATAATCTGAAAGATATATT    | 2280 |
| $\Psi$ <i>SMAD4</i> | 2180 | AACCTTATCAGGATGGTGGACTACAAAATACAATCCTGTTTATAATCTGAAAGATATATT    | 2239 |
| <i>SMAD4</i>        | 2281 | CACCTTTTGTCTGCTTTATCTTTTCATAAAGGGTTGAAAATGTGTTTGCTGCCTTGCTCC    | 2340 |
| $\Psi$ <i>SMAD4</i> | 2240 | CACCTTTTGTCTGCTTTATCTTTTCATAAAGGGTTGAAAATGTGTTTGCTGCCTTGCTCC    | 2299 |
| <i>SMAD4</i>        | 2341 | TAGCAGACAGAACTGGATTAACAATTTTTTTTTCTCTTCAGAACTTGTCAGGCA          | 2400 |
| $\Psi$ <i>SMAD4</i> | 2300 | TAGCAGACAGAACTGGATTAACAATTTTTTTTTCTCTTCAGAACTTGTCAGGCA          | 2359 |
| <i>SMAD4</i>        | 2401 | GGCTCAGAGCTTGAAGATTAGGAGAAACACATCTTATTAATCTTCACCTGTTATGTAT      | 2460 |
| $\Psi$ <i>SMAD4</i> | 2380 | GGCTCAGAGCTTGAAGATTAGGAGAAACACATCTTATTAATCTTCACCTGTTATGTAT      | 2419 |
| <i>SMAD4</i>        | 2461 | GAAAGGAATCATTCCAGTGCTAGAAAATTTAGCCCTTTAAAACGCTTTAGAGCCTTTTATC   | 2520 |
| $\Psi$ <i>SMAD4</i> | 2420 | GAAAGGAATCATTCCAGTGCTAGAAAATTTAGCCCTTTAAAACGCTTTAGAGCCTTTTATC   | 2479 |
| <i>SMAD4</i>        | 2521 | TGCAGAACATCGATATGTATATCATTCTACAGAATAATCCAGTATTGCTGATTTTAAAGG    | 2580 |
| $\Psi$ <i>SMAD4</i> | 2480 | TGCAGAACATCGATATGTATATCATTCTACAGAATAATCCAGTATTGCTGATTTTAAAGG    | 2539 |
| <i>SMAD4</i>        | 2581 | CAGAGAAGTTCTCAAAGTTAATTCACCTATGTTATTTTGTGTACAAGTTGTTATTGTTGA    | 2640 |
| $\Psi$ <i>SMAD4</i> | 2540 | CAGAGAAGTTCTCAAAGTTAATTCACCTATGTTATTTTGTGTACAAGTTGTTATTGTTGA    | 2599 |
| <i>SMAD4</i>        | 2641 | ACATACTTCAAAAATAATGTGCCATGTGGGTGAGTTAATTTTACCAAGAGTAACTTTACT    | 2700 |
| $\Psi$ <i>SMAD4</i> | 2600 | ACATACTTCAAAAATAATGTGCCATGTGGGTGAGTTAATTTTACCAAGAGTAACTTTACT    | 2659 |
| <i>SMAD4</i>        | 2701 | CTGTGTTTAAAAAGTAAGTTAATAATGTATTGTAATCTTTTCAATCAAAAATATTTTGTCA   | 2760 |
| $\Psi$ <i>SMAD4</i> | 2680 | CTGTGTTTAAAAAGTAAGTTAATAATGTATTGTAATCTTTTCAATCAAAAATATTTTGTCA   | 2719 |
| <i>SMAD4</i>        | 2761 | AGTTATATTAGTGAAGATGGTTTCAATTCAGATTGTCTTGCAACTTCAGTTTTATTTTGT    | 2820 |
| $\Psi$ <i>SMAD4</i> | 2720 | AGTTATATTAGTGAAGATGGTTTCAATTCAGATTGTCTTGCAACTTCAGTTTTATTTTGT    | 2779 |
| <i>SMAD4</i>        | 2821 | CCAAGGCAAAAACTCTTAATCTGTGTGTATATTGAGAATCCCTTAAAATACAGACAA       | 2880 |
| $\Psi$ <i>SMAD4</i> | 2780 | CCAAGGCAAAAACTCTTAATCTGTGTGTATATTGAGAATCCCTTAAAATACAGACAA       | 2839 |
| <i>SMAD4</i>        | 2881 | AAAAATTTAAAATACGTTTGTATTTCCTAGTGGATGACTGTTGATGAAGTACTTTTC       | 2940 |
| $\Psi$ <i>SMAD4</i> | 2840 | AAAAATTTAAAATACGTTTGTATTTCCTAGTGGATGACTGTTGATGAAGTACTTTTC       | 2899 |
| <i>SMAD4</i>        | 2941 | CCCTGTTAAACAGTAGTTGTATTCTTCTGTATTCTAGGCCACAAGGTTGGTTGCTAAGAA    | 3000 |
| $\Psi$ <i>SMAD4</i> | 2900 | CCCTGTTAAACAGTAGTTGTATTCTTCTGTATTCTAGGCCACAAGGTTGGTTGCTAAGAA    | 2959 |
| <i>SMAD4</i>        | 3001 | GCCATAAGAGGAATTTCTTTCCCTTCATTTCATAGGGAAGGTTTGTATTTTTTAAAC       | 3060 |
| $\Psi$ <i>SMAD4</i> | 2960 | GCCATAAGAGGAATTTCTTTCCCTTCATTTCATAGGGAAGGTTTGTATTTTTTAAAC       | 3019 |
| <i>SMAD4</i>        | 3061 | ACTAAAAGCAGCGTCACTCACTAATGTCTCACTGTTCTGCAAAGGTGGCAATGCTTAA      | 3120 |
| $\Psi$ <i>SMAD4</i> | 3020 | ACTAAAAGCAGCGTCACTCACTAATGTCTCACTGTTCTGCAAAGGTGGCAATGCTTAA      | 3079 |
| <i>SMAD4</i>        | 3121 | ACTAAATAATGAATAAAGTGAATATTTTGGAAACTGCTAAATCTATGTTAAATACTGTG     | 3180 |
| $\Psi$ <i>SMAD4</i> | 3080 | ACTAAATAATGAATAAAGTGAATATTTTGGAAACTGCTAAATCTATGTTAAATACTGTG     | 3139 |
| <i>SMAD4</i>        | 3181 | CAGAATAATGGAACATTACAGTTTCATAATAGGTAGTTTGGATATTTTGTACTTGATTT     | 3240 |
| $\Psi$ <i>SMAD4</i> | 3140 | CAGAATAATGGAACATTACAGTTTCATAATAGGTAGTTTGGATATTTTGTACTTGATTT     | 3199 |
| <i>SMAD4</i>        | 3241 | GATGTGACTTTTTTGGTATAATGTTTAAATCATGTATGTTATGATATTGTTTAAATTC      | 3300 |
| $\Psi$ <i>SMAD4</i> | 3200 | GATGTGACTTTTTTGGTATAATGTTTAAATCATGTATGTTATGATATTGTTTAAATTC      | 3259 |
| <i>SMAD4</i>        | 3301 | AGTTTTGTATCTTGGGGCAAGACTGCAAACCTTTTATATCTTTTGGTTATTCTAAGCC      | 3360 |
| $\Psi$ <i>SMAD4</i> | 3260 | AGTTTTGTATCTTGGGGCAAGACTGCAAACCTTTTATATCTTTTGGTTATTCTAAGCC      | 3319 |
| <i>SMAD4</i>        | 3361 | CTTTGCCATCAATGATCATATCAATTGGCAGTGACTTTGTATAGAGAATTTAAGTAGAAA    | 3420 |
| $\Psi$ <i>SMAD4</i> | 3320 | CTTTGCCATCAATGATCATATCAATTGGCAGTGACTTTGTATAGAGAATTTAAGTAGAAA    | 3379 |
| <i>SMAD4</i>        | 3421 | AGTTGCAGATGATTGACTGTACCACAGACACAATATGTATGCTTTTACCTAGCTGGTA      | 3480 |
| $\Psi$ <i>SMAD4</i> | 3380 | AGTTGCAGATGATTGACTGTACCACAGACACAATATGTATGCTTTTACCTAGCTGGTA      | 3439 |
| <i>SMAD4</i>        | 3481 | GCATAAATAAACTGAATCTCAACATACAAAGTTGAATCTAGGTTTGTATTTTTAAGATT     | 3540 |
| $\Psi$ <i>SMAD4</i> | 3440 | GCATAAATAAACTGAATCTCAACATA.....                                 | 3466 |

**Rycina 20. Porównanie całkowitej sekwencji  $\Psi$ SMAD4 z sekwencją mRNA genu rodzicielskiego SMAD4.** Sekwencja  $\Psi$ SMAD4 o długości 3466 pz została porównana z sekwencją mRNA SMAD4 (NM\_005359.6). Dla zachowania przejrzystości sekwencję SMAD4 ograniczono do pierwszych 3540 pz, w obrębie których przyrównano sekwencję pseudogenu. Kolor ciemnoniebieski odpowiada sekwencji identycznej a jasnoniebieski wskazuje różnice w porównywanych sekwencjach. Rycinę przygotowano z użyciem oprogramowania Jalview w wersji 2.11.2.5.

## 4.9 Ocena typu obserwowanego pseudogenu

Określenie pełnej sekwencji  $\Psi$ *SMAD4* pozwoliło na zidentyfikowanie i potwierdzenie cech charakterystycznych dla pseudogenów przetworzonych takich jak:

- brak sekwencji intronowych,
- obecność sekwencji poli-A na końcu 3',
- obecność wstawki TSD (w tym przypadku „TTTC”) będącej duplikacją miejsca insercji.

Powyższe cechy, pomimo braku dodatkowych wariantów SNV, pozwoliły na jednoznaczne zakwalifikowanie obserwowanego  $\Psi$ *SMAD4* do grupy pseudogenów przetworzonych.

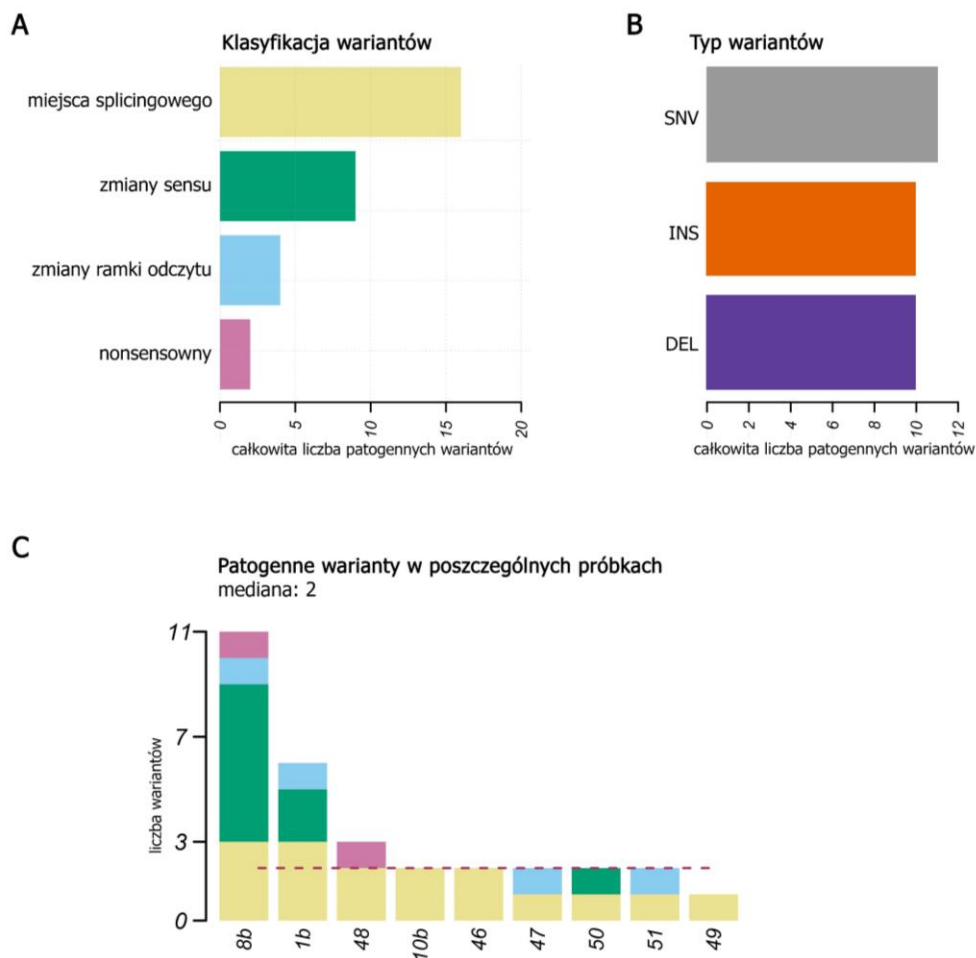
## 4.10 Identyfikacja patogennych wariantów, zmian liczby kopii i aberracji chromosomowych u chorych z $\Psi$ *SMAD4*

Następnie postanowiono sprawdzić, czy próbki CML z  $\Psi$ *SMAD4* cechują się unikalnymi zmianami genetycznymi, nieobecnymi w próbkach CML bez pseudogenu. W tym celu za pomocą metody NGS zsekwencjonowano pozostałe próbki CML z  $\Psi$ *SMAD4* wykryte podczas skriningu, tj. próbki 46-51. Ostatecznie, analizę danych pochodzących z sekwencjonowania NGS wykonano dla 5 próbek CP (próbki 47-51), jednej AP (próbka 46) i trzech, zsekwencjonowanych wcześniej, próbek BP z  $\Psi$ *SMAD4* (próbki 1b, 8b i 10b). Analizę przeprowadzono jak dla sparowanych próbek CML. Szczegółowe informacje dotyczące wykrytych patogennych wariantów przedstawiono na rycinach 21 i 22.

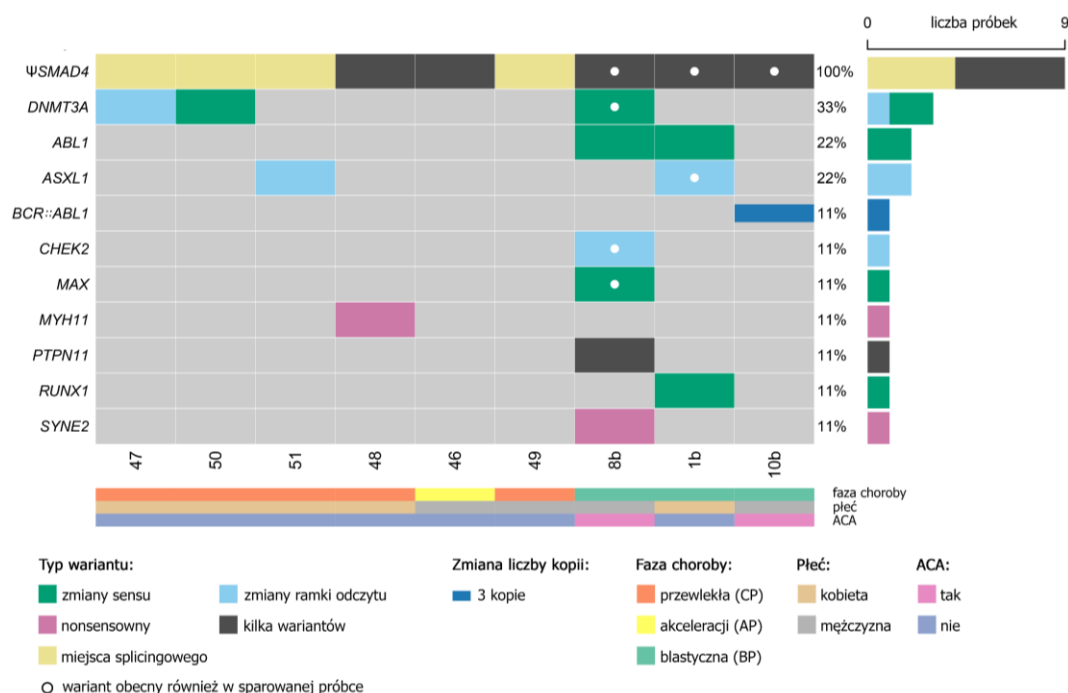
Analiza mutacji małej skali umożliwiła identyfikację 31 patogennych lub istotnych wariantów, z czego 16 potwierdzało obecność pseudogenu we wszystkich badanych próbkach (9/9; 100%). Patogenne warianty u więcej niż jednego chorego zidentyfikowano w genach: *DNMT3A* (3/9; 33%), *ABL1* (2/9; 22%) i *ASXL1* (2/9; 22%) (rycina 22). Liczba wariantów patogennych lub istotnych wahała się od 1 do 11 na próbkę; więcej zmian obserwowano w próbkach z fazy BP (rycina 21C). Z uwagi na fakt, że przetworzone pseudogeny nie posiadają sekwencji intronowych, odczyty potwierdzające obecność  $\Psi$ *SMAD4* interpretowane były przez programy do wykrywania wariantów jako insercje lub delecje na końcach egzonów genu *SMAD4*, w obrębie miejsc splicingowych i stanowiły największy odsetek istotnych wariantów wykrytych w analizowanych próbkach (rycina 21A-B). W dwóch próbkach z fazy AP lub CP (odpowiednio próbki 46 i 49), poza wariantami w genie *SMAD4* wspierającymi obecność pseudogenu nie wykryto patogennych zmian genetycznych. Z kolei w dwóch próbkach



BP, oprócz zmian punktowych i małych insercji-delekcji, dodatkowo zidentyfikowano zwiększoną liczbę kopii *BCR::ABL1* oraz trisomie chromosomów 8 i 21 (próbka 10b) a także monosomię chromosomu 7 (próbka 8b). Powyższe wyniki przedstawiono na rycinie 22. Ponadto, w próbce 49 wykryto delecję chromosomu Y a w próbce 50- translokację  $t(1;9;22)(q21;q34;q11)$  - dane pominięto na rycinie 22 z uwagi na to, że aberracje te nie są typowymi ACA obserwowanymi w CML.



**Rycina 21. Statystyki dotyczące patogennych lub istotnych wariantów zidentyfikowanych w wyniku sekwencjonowania NGS 9 próbek CML z  $\Psi$ SMAD4.** (A) Klasyfikacja wykrytych wariantów ze względu na wywoływany przez nie efekt, (B) Klasyfikacja wykrytych wariantów ze względu na ich typ; SNV- wariant pojedynczego nukleotydu, INS- insercja, DEL- delecja, (C) Liczba patogennych lub istotnych wariantów wykryta w poszczególnych próbkach; kody próbek znajdują się pod odpowiadającymi im słupkami. Przerywaną linią oznaczono medianę patogennych wariantów. Wykresy wygenerowano za pomocą funkcji *plomafSummary* pakietu *mafTools* dla zmian punktowych i małych insercji-delekcji, z pominięciem zmian liczby kopii genów oraz aberracji chromosomowych.



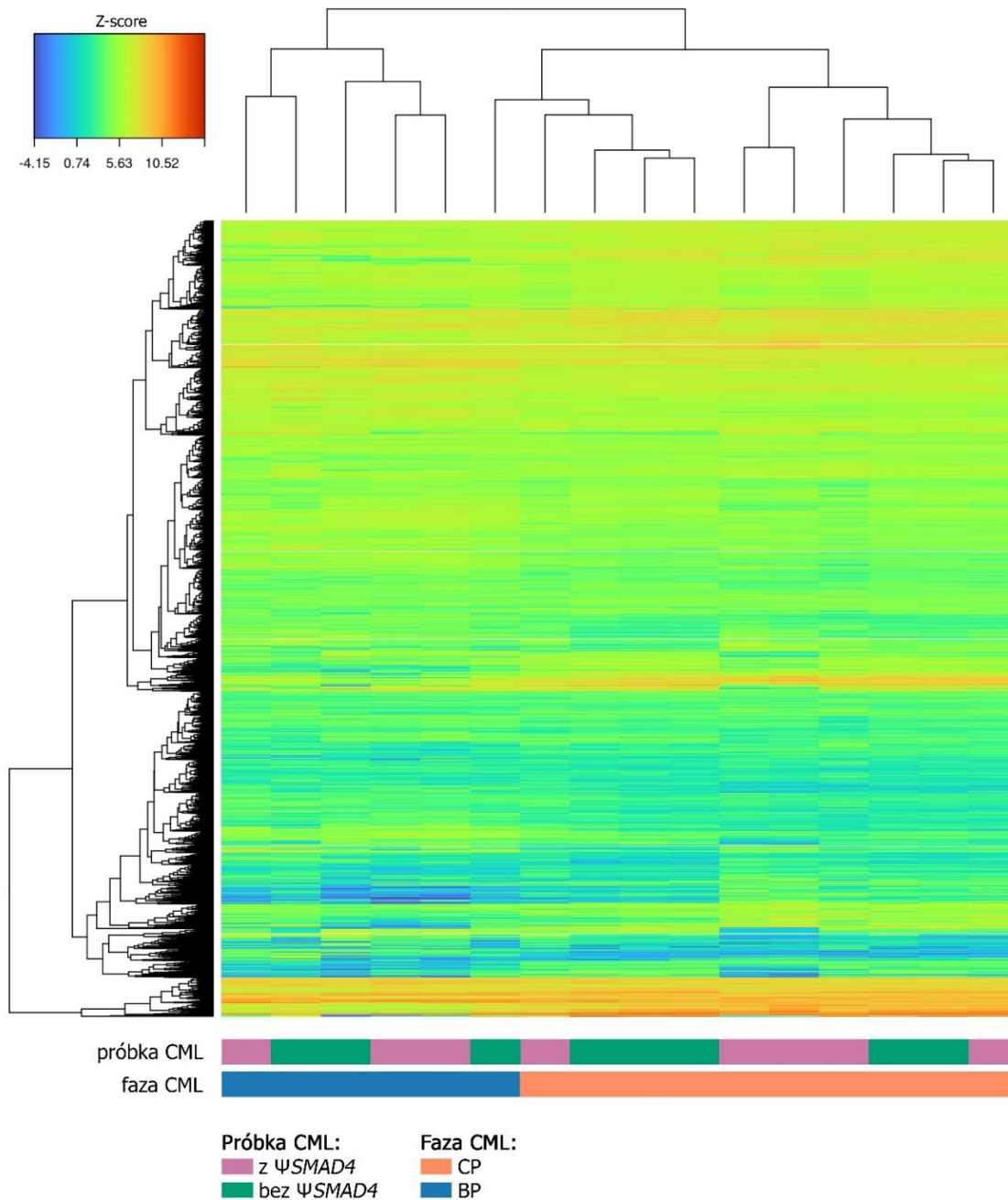
**Rycina 22. Profil zmian genetycznych zidentyfikowanych na podstawie sekwencjonowania NGS 9 próbek CML z  $\Psi$ SMAD4.** Wykres przedstawia wszystkie patogenne lub istotne zmiany jakie zidentyfikowano w ramach analizy zmian punktowych i małych insercji-delecji, zmian liczby kopii genów (górny panel) oraz ACA (dolny panel). Kody poszczególnych próbek umieszczono pomiędzy górnym a dolnym panelem. Po prawej stronie górnego panelu podano odsetek próbek, w których wykryto patogenne lub istotne zmiany w poszczególnych genach. Wykres wygenerowano za pomocą funkcji *oncoplot* pakietu *maftools*. Skrót; ACA- dodatkowe aberracje chromosomowe.

#### 4.11 Analiza transkryptomu u chorych z $\Psi$ SMAD4

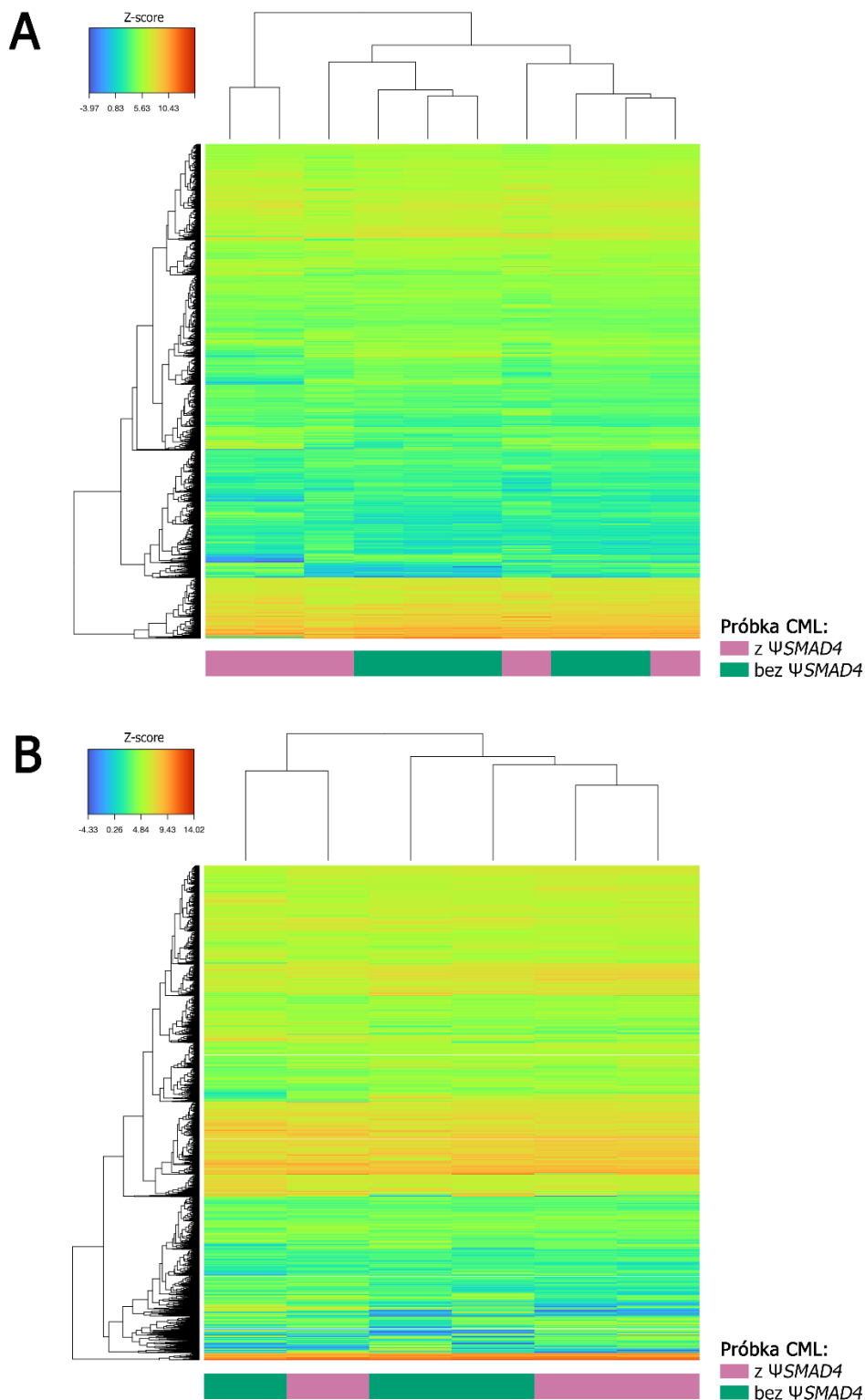
Niewielka część pseudogenów może podlegać transkrypcji i w konsekwencji regulować ekspresję różnych genów. W celu sprawdzenia czy badany  $\Psi$ SMAD4 ulega ekspresji wykonano analizę transkryptomu za pomocą metody RNAseq dla 8 próbek CML z  $\Psi$ SMAD4. Kontrolę w eksperymencie stanowiło 8 próbek CML bez badanego pseudogenu. Układ próbek w grupie badanej i kontrolnej był taki sam i obejmował 5 próbek CP i 3 BP. Przy założeniach minimum dwukrotnej różnicy ekspresji (*fold change*;  $\log_2FC >1$  lub  $<-1$ ) i odrzuceniu wyników fałszywie pozytywnych (*false discovery rate*; współczynnik FDR  $<0,05$ ), przeprowadzona analiza nie wykazała istotnych różnic w ekspresji genów *SMAD4* (wartość  $p= 0,35$  ;  $\log_2FC= -0,15$  ; FDR= 0,84) i *SCAI* (wartość  $p= 0,43$  ;  $\log_2FC= -0,39$  ; FDR= 0,87) oraz pozostałych genów ulegających ekspresji w próbkach z  $\Psi$ SMAD4 względem próbek kontrolnych. Całościowe dane uzyskane z sekwencjonowania RNAseq zostaną zdeponowane

w repozytorium Gene Expression Omnibus na etapie składania manuskryptu z uzyskanych wyników w recenzowanym czasopiśmie naukowym. Zmiany ekspresji genów nie były obserwowane ani w analizie wszystkich próbek jednocześnie (rycina 23), ani z podziałem na fazy choroby (rycina 24). Przedstawiona na rycinie 23 zbiorcza analiza próbek z fazy CP i BP wykazała klastrowanie hierarchiczne próbek względem fazy choroby, ale nie względem obecności  $\Psi$ *SMAD4*. Analizy porównawcze ekspresji genów w fazie CP (rycina 24A) i BP (rycina 24B) również nie wykazały klastrowania hierarchicznego próbek względem obecności  $\Psi$ *SMAD4*.

Z powodu braku różnic w sekwencji  $\Psi$ *SMAD4* i mRNA genu rodzicielskiego porównanie ekspresji *SMAD4* w analizowanych grupach jedynie pośrednio dowodzi braku ekspresji pseudogenu. Z tego względu, drugim sposobem oceny potencjalnej ekspresji  $\Psi$ *SMAD4* było przyrównanie surowych odczytów z RNAseq do sekwencji referencyjnej stanowiącej sekwencję  $\Psi$ *SMAD4* wraz z miejscem jego insercji w 18 intronie *SCAI*. Analiza nie wykazała obecności odczytów (lub par odczytów) mapujących się jednocześnie do *SCAI* i  $\Psi$ *SMAD4*, co dodatkowo potwierdza brak jego ekspresji.



**Rycina 23. Sygnatura ekspresji genów u chorych na CML z  $\Psi$ SMAD4.** Na wykresie przedstawiono klastrowanie hierarchiczne danych z RNAseq dla próbek CML z  $\Psi$ SMAD4 i próbek kontrolnych CML bez pseudogenu. Heatmapa została wygenerowana za pomocą pakietu *mixOmics* dla 10,7 tys. genów, po wcześniejszym odfiltrowaniu genów o niskim i średnim poziomie ekspresji ( $\log_2\text{CPM}>2$ ); rzędy przedstawiają poziom ekspresji genów zobrazowanych za pomocą Z-score, kolumny odpowiadają chorym poddany analizie. CPM- counts per milion.



**Rycina 24. Sygnatura ekspresji genów u chorych na CML z  $\Psi$ SMAD4 z podziałem na fazy choroby.** Na wykresach przedstawiono klastrowanie hierarchiczne danych z RNAseq dla próbek CML z  $\Psi$ SMAD4 i próbek kontrolnych CML bez pseudogenu w fazie CP (A) i BP (B). Heatmapy zostały wygenerowane za pomocą pakietu *mixOmics* odpowiednio dla 10,2 tys. i 11,1 tys. genów, po wcześniejszym odfiltrowaniu genów o niskim i średnim poziomie ekspresji ( $\log_2\text{CPM}>2$ ); rzędy przedstawiają poziom ekspresji genów zobrazowanych za pomocą Z-score, kolumny odpowiadają chorym poddanych analizie. CPM- *counts per milion*.

## 5. DYSKUSJA

Chromosom Filadelfia, będący wynikiem translokacji t(9;22)(q34;q11), stanowi modelowy przykład tego, jak obecność pojedynczej aberracji genetycznej inicjuje proces transformacji nowotworowej, a jednocześnie pozwala na zastosowanie skutecznej terapii celowanej. Wprowadzenie terapii z użyciem TKI sprawiło, że CML ma jedno z najlepszych rokowań wśród nowotworów hematologicznych. Mimo to, oporność na imatynib lub TKI drugiej generacji zastosowany w pierwszej linii leczenia występuje u, odpowiednio, 10-15% i <10% chorych [10]. Najczęstszą przyczyną oporności są mutacje w domenie kinazowej *BCR::ABL1*, jednak zróżnicowana odpowiedź na leczenie z użyciem TKI wskazuje udział niezależnych od *BCR::ABL1* mechanizmów w progresji CML. Mechanizmy nabywania oporności oraz przyczyny progresji z fazy CP do BP nie zostały do końca poznane, jednak mogą być one powiązane m.in. z niestabilnością genetyczną, indukowaną przez onkogen *BCR::ABL1*, prowadzącą do nabycia dodatkowych aberracji genetycznych, począwszy od mutacji małej skali, aż po aberracje chromosomowe.

Ciągły rozwój wysokoprzepustowych technologii, wzrost ich dostępności oraz obniżenie kosztów sekwencjonowania NGS umożliwia nie tylko poszukiwanie nowych czynników prognostycznych i predykcyjnych, ale również odkrywanie molekularnych mechanizmów patogenezy, oporności i progresji wielu chorób, w tym chorób nowotworowych. Prowadzone w ostatnich latach badania wielkoskalowe z użyciem sekwencjonowania NGS wskazują, że słaba odpowiedź na leczenie z zastosowaniem TKI oraz progresja z fazy CP do BP w CML może po części wynikać z nabycia dodatkowych mutacji genetycznych, dotyczących w szczególności genów zaangażowanych w mechanizmy epigenetyczne tj. modyfikację struktury chromatyny (np. *ASXL1*, *BCOR*, *EZH2*, *KDM6A*) i metylację DNA (np. *DNMT3A*, *IDH1*, *IDH2*, *TET2*) [52, 61, 62, 66, 114]. Podobne profile mutacji obserwuje się również w innych nowotworach układu krwiotwórczego takich jak: AML [115, 116], MPN [117] i MDS [118, 119].

W poznaniu molekularnych przyczyn oporności na TKI oraz progresji CML uzasadnioną strategią badań jest więc identyfikacja i porównanie profili mutacji genetycznych w próbkach z fazy CP (pobranej jeszcze przed wdrożeniem leczenia) ze sparowanymi próbkami zebranymi w trakcie progresji choroby. Analiza takich danych pozwala na wytypowanie zmian genetycznych, zarówno tych już istniejących na etapie diagnozy jak i nabytych w trakcie leczenia, o istotnym znaczeniu prognostycznym

i predykcyjnym, co może mieć konkretne przełożenie na wybór najkorzystniejszej dla chorego terapii. O ile dostępnych jest wiele wyników badań naukowych wskazujących geny zmutowane w sposób istotnie częsty w CML nawet na dużych zbiorach próbek z fazy CP lub BP, o tyle liczba publikacji uwzględniająca dane dla sparowanych próbek jest wciąż ograniczona i zwykle dotyczy kilku-kilkunastu par (tabela 1). Dlatego celem niniejszej pracy doktorskiej było znalezienie nowych zmian genetycznych w sparowanych próbkach z różnych stadiów CML, mogących mieć związek z opornością na TKI i progresją choroby. Za pomocą metody NGS zsekwencjonowano 25 sparowanych próbek CML oraz dodatkowe 20 próbek wyłącznie z fazy BP, a następnie przeprowadzono analizę mutacji 193 genów (tabela 4), powiązanych z patogenezą nowotworów układu krwiotwórczego oraz onkogenezą (najczęściej zmutowane protoonkogeny i geny supresorowe).

Przedstawiona w rozprawie analiza wykazała patogenne zmiany w genach zaangażowanych w regulację metylacji DNA (*DNMT3A*, *IDH2*, *TET2*), modyfikację struktury chromatyny (*ASXL1*, *KMT2D*) i mechanizmy naprawy DNA (*TP53*, *CHEK2*), a także w genach kodujących czynniki transkrypcyjne (*IKZF1*, *RUNX1*, *WT1*), genach ścieżki sygnałowej RTK-RAS (*ABL1*, *PTPN11*) oraz genach uczestniczących w regulacji cyklu komórkowego (*CDKN2A*, *CDKN2B*) (ryciny 11-13). Najczęstszą i zarazem najlepiej poznaną przyczyną oporności na TKI w CML są mutacje w domenie kinazowej *BCR::ABL1*, co potwierdzają również wyniki przedstawione w ramach tej rozprawy. Analiza wykazała obecność patogennych wariantów w *ABL1* w 11 z 45 (24%) próbek z fazy BP, w tym w 9 sparowanych próbkach i 2 niesparowanych (rycina 12). W przypadku sparowanych próbek BP, te same patogenne warianty obserwowane były jedynie w dwóch próbkach sprzed progresji choroby, co oznacza, że u pozostałych 7 chorych wykryte warianty powstały pod presją terapii TKI. Z drugiej strony, w 10 z 11 (91%) próbek z fazy BP ze zmutowanym *ABL1* wykryto dodatkowe, patogenne warianty w innych genach, co może wskazywać na udział w progresji choroby mutacji w genach innych niż w *BCR::ABL1*. Wszystkie patogenne warianty wykryte w *ABL1* miały charakter mutacji zmiany sensu, a większość z nich została wcześniej opisana w literaturze jako mutacje typu „*hot spot*”, tj. jako najczęściej występujące. Częstość patogennych wariantów w genie *ABL1* jest podobna lub nieco niższa niż częstość obserwowana w innych, opublikowanych wynikach badań naukowych. W pracy Curik i in. mutacje w genie *BCR::ABL1* wykryto u 28% (24/87) chorych z nieoptymalną odpowiedzią na TKI lub jej brakiem, z czego u 8 z nich (33%) zidentyfikowano również

współistniejące mutacje w innych genach, w większości występujące również w próbkach z momentu diagnozy CML [120]. Z kolei analiza 33 próbek z fazy BP po leczeniu TKI wykazała mutacje w *ABL1* w 19 próbkach (58%) i występowały one częściej u chorych z kryzą limfoidalną, niż mieloidalną. W 89% próbek ze zmutowanym *ABL1* zidentyfikowano istotne mutacje w innych genach, które w ponad połowie przypadków występowały jeszcze przed pojawieniem się mutacji w *ABL1* [61]. Ilość wykrywanych wariantów w dużej mierze zależy od czułości stosowanej metody. Złotym standardem w ocenie mutacji w domenie kinazowej *BCR::ABL1* od lat pozostaje sekwencjonowanie metodą Sangera, którego czułość ocenia się na poziomie około 15-20%. W powyższej pracy częstość allelu alternatywnego (VAF, ang. *variant allele frequency*) patogennych wariantów wykrytych w *ABL1* wynosiła od 1,5 do 56%; połowa tych wariantów (9/18) miała VAF <20%, a więc poniżej czułości sekwencjonowania metodą Sangera. Co istotne, wśród wariantów o niskim VAF występowały zmiany zaliczane do najbardziej opornych mutacji w domenie kinazowej tj. p.Gly250Glu, p.Glu255Val, p.Glu255Lys, p.Thr315Ile oraz p.Phe317Leu, w przypadku których stosuje się szczególne protokoły leczenia [121]. Wariant p.Thr315Ile został wykryty u dwóch chorych: w próbce CP 22a (VAF 19%) oraz BP 19b (VAF 4,6%) i obu chorych skierowano do terapii komórkowej, tj. allo-HCT lub immunoterapii z użyciem limfocytów T z chimerycznym receptorem antygenowym (CAR-T, ang. *T cells modified with chimeric antigen receptor*). Istotę czułości metody używanej do oceny mutacji w domenie kinazowej *BCR::ABL1* podkreśla również praca Soverini i in., w której porównano wyniki sekwencjonowania *BCR::ABL1* uzyskane metodą Sangera i NGS [122]. Równoległa analiza obiema metodami 236 próbek CML wykazała obecność mutacji wykrywanych jedynie metodą NGS u 51 chorych. Dodatkowo, u 29 z 60 chorych, u których wykryto mutacje metodą Sangera, analiza wyników z sekwencjonowania NGS wykazała obecność dodatkowych mutacji z niższym VAF. 18% wszystkich mutacji niewykrywalnych metodą Sangera miało istotne znaczenie kliniczne. Co ciekawe, sekwencjonowanie metodą NGS umożliwia identyfikację mutacji w domenie kinazowej *ABL1* odpowiedzialnych za progresję choroby już po 3 miesiącach od rozpoczęcia terapii z TKI [123].

Gen *ASXL1* (*Additional sex comb like 1*) koduje białko biorące udział w regulacji procesów epigenetycznych i transkrypcji. Prawidłowe białko ASXL1, poprzez interakcję z kompleksem białek PRC2 (*Polycomb repressive complex 2*) wpływa na trimetylację lizyny 27 histonu H3, prowadząc do zahamowania transkrypcji genów docelowych [124].



Mutacje genu *ASXLI* są zazwyczaj monoalleliczne, występują w obrębie egzonu 12 i mają charakter zmian ramki odczytu lub mutacji nonsensownych, skutkujących przedwczesnym skróceniem C-końcowej części białka i w konsekwencji utratą jego funkcji [125]. Liczne doniesienia literaturowe potwierdzają wysoką częstość mutacji w *ASXLI* w nowotworach hematologicznych linii mieloidalnej m.in. w MDS, MPN, CML i AML wtórnej do innych chorób hematologicznych [114, 126-128]. W niniejszej pracy patogenne warianty w genie *ASXLI* zostały wykryte, podobnie jak w przypadku *ABL1*, w 24% (11/45) próbek z fazy BP (rycina 11-12). Dodatkowo, był to również najczęściej zmutowany gen w próbkach sprzed progresji CML (8/25 próbek, 32%) (rycina 13). Przeprowadzona analiza wykazała obecność typowych dla *ASXLI* wariantów o charakterze zmian ramki odczytu oraz wariantów nonsensownych w egzonie 12, do najczęstszych zmian należały warianty: p.Gly646Trpfs\*12 (5/45, 11%) oraz p.Glu635Argfs\*15 (3/45, 7%). Mutacje w *ASXLI* uważane są za niekorzystny czynnik prognostyczny w różnych nowotworach. Retrospektywna analiza 81 genów powiązanych z nowotworzeniem u 115 chorych na CML w różnych stadiach choroby wykazała, że najczęstszym genem zmutowanym w fazie CP i AP był *ASXLI*, odpowiednio 14 i 40%. Co więcej, w analizie wieloczynnikowej pokazano, że mutacje w *ASXLI* były jedynym niezależnym czynnikiem ryzyka związanym z gorszym czasem przeżycia wolnym od zdarzeń (EFS, ang. *event-free survival*) w fazie przewlekłej CML [129]. Ostatnie badania wskazują, że mutacje w *ASXLI*, są nie tylko czynnikiem prognostycznym, ale i predykcyjnym w CML. Schönfeld i in. przeprowadzili analizę, w której wykazali, że chorzy na CML ze zmutowanym *ASXLI* w fazie CP, odpowiadali gorzej na terapię z nilotynibem w stosunku do pozostałych badanych chorych, mimo zaobserwowanej eradykacji większości mutacji w trakcie leczenia [130]. Do niedawna mutacje w *ASXLI* uważano za zdarzenia powiązane z podeszłym wiekiem, jednak badanie przeprowadzone przez Ernsta i in. na grupie młodych dorosłych i dzieci chorych na CML wykazało, że *ASXLI* był jedynym z analizowanych 30 genów który ulegał mutacji (6/21 chorych) [131]. W niniejszej pracy, u dwóch chorych zaobserwowano zanikanie patogennych wariantów w *ASXLI* w trakcie progresji choroby. U chorego o numerze 22 zanik wariantów jest wynikiem przeprowadzonego przeszczepu allo-HCT, natomiast w przypadku chorego o numerze 17 doszło do utraty dwóch wariantów: zamiany ramki odczytu (VAF= 29%) i nonsensownego (VAF= 16%), co może świadczyć o wznowie nowotworu z innego klonu komórek białaczkowych. Zanik mutacji w genie *ASXLI* wraz z progresją choroby został już wcześniej opisany

w AML [132] i CML [133], co wskazuje, że status mutacji *ASXL1* może ulegać zmianie w trakcie ewolucji choroby, a mutacje te nie stanowią dobrego markera do monitorowania choroby resztkowej w AML.

Analiza przeprowadzona w ramach niniejszej pracy pozwoliła na identyfikację patogennych wariantów w genie *IKZF1* (*IKAROS family zinc finger 1*), kodującym czynnik transkrypcyjny odgrywający kluczową rolę w hematopoezie, szczególnie w linii limfoidalnej, jako główny regulator wczesnego rozwoju limfocytów, ich różnicowania i homeostazy [134]. W wyniku alternatywnego splicingu mogą powstawać skrócone izoformy *IKZF1*, które utraciły zdolność wiązania DNA, ale mogące hamować funkcjonalne białko o pełnej długości (ang. *dominant-negative isoforms*) [135]. Heterozygotyczne mutacje *IKZF1* prowadzą zazwyczaj do szybkiego rozwoju ostrych białaczek i chłoniaków [136]. Delecje *IKZF1* są najczęstszym typem mutacji występującej w ALL z chromosomem Ph i stanowią negatywny czynnik prognostyczny, zarówno u dorosłych, jak i dzieci [137, 138]. Delecji może ulegać cały gen lub jedynie część egzonów (najczęściej egzony 2-7 i 4-7), a zmiany te obserwowane są zarówno w ALL, jak i CML w fazie BP o fenotypie limfoidalnym; brak jest ich w kryzie mieloidalnej i fazie CP [139, 140]. Zgodnie z powyższym, w przeprowadzonej analizie patogene zmiany w genie *IKZF1* występowały jedynie w fazie BP (rycina 11-12), głównie w postaci delecji 1 lub 2 kopii obejmujących cały gen lub egzony 2–7 albo 4–7. Podobne zmiany w próbkach CML zaobserwowali wcześniej Branford i in.; co ciekawe, delecje wykryto nie tylko w fazie BP, ale również w kilku próbkach z fazy CP o niekorzystnym przebiegu [61]. Opisano również nietypowy przypadek, w którym u 1,5-roczonej dziewczynki z CML zidentyfikowano delecje w genach *IKZF1*, *CDKN2A*, *PAX5* i *ETV6*. Zmiany te przyczyniły się do progresji do fazy BP, a następnie transformacji do ALL w przeciągu 6 miesięcy od diagnozy [141].

*RUNX1* (*RUNX family transcription factor 1*) jest kolejnym czynnikiem transkrypcyjnym pełniącym kluczową rolę w regulacji hematopoezy. Mutacje somatyczne w *RUNX1* występują często w nowotworach i stanowią niezależny negatywny czynnik prognostyczny, co zaobserwowano m.in. w: AML [142], MDS [143] i ALL [144]. Ostatnie badania wskazują jednak, że efekt ten może zależeć od zastosowanej terapii [145]. Z kolei zarodkowe mutacje *RUNX1* powiązano z występowaniem rodzinnego defektu płytek z predyspozycją do AML (ang. *familial platelet disorder with predisposition to AML*) [146]. W przeprowadzonej analizie patogene warianty w *RUNX1* zidentyfikowano u 18% chorych w fazie BP, z czego 2/3

wykryto w domenie RUNT, odpowiedzialnej za wiązanie DNA, i były to głównie warianty zmiany sensu. Pozostałe warianty były zlokalizowane w domenie transaktywacyjnej TAD (lub w jej pobliżu), pośredniczącej w oddziaływaniach białko-białko, i miały charakter wariantów zmiany ramki odczytu lub wariantów nonsensownych (rycina 11-12). Profil zmian w *RUNX1* jest zgodny z opublikowanymi dotychczas danymi [64, 142]; przypuszcza się że obserwowane zmiany zaburzają fizjologiczną funkcję *RUNX1* poprzez upośledzenie zdolności wiązania DNA lub przedwczesną terminację białka [147]. Mutacje w *RUNX1* są często identyfikowane również w CML [63, 64]. Wykazano także ich potencjalny udział w transformacji do AML [148]. Oprócz mutacji mono i biallelicznych, *RUNX1* jest również partnerem fuzyjnym w ponad 50 translokacjach chromosomowych, obserwowanych w nowotworach układu krwiotwórczego [149].

Analiza danych z sekwencjonowania NGS pozwoliła na wykrycie w 13% próbek (6/45) z fazy BP patogennych wariantów w genie *DNMT3A* (*DNA methyltransferase 3 alpha*), kodującym metylotransferazę DNA, zaangażowaną w proces metylacji *de novo* (rycina 11-12). U 4 chorych warianty te były obecne również w próbkach sprzed progresji choroby, przez co *DNMT3A* był drugim najczęściej zmutowanym genem w tej grupie (16%) (rycina 13). Mutacje w *DNMT3A* występują z różną częstotliwością w wielu nowotworach hematologicznych, w tym głównie w AML i MDS [150] i na ogół mają charakter mutacji punktowych, zlokalizowanych w obrębie funkcjonalnych domen, prowadzących do utraty funkcjonalnego białka [151]. Mutacje typu „*hot spot*” w *DNMT3A* najczęściej dotyczą kodonu p.Arg882 [152], jednak nie zidentyfikowano ich w przeprowadzonej, w ramach niniejszej rozprawy, analizie. Ciekawe wydają się jednak dwie inne, dotychczas nieopisane patogenne zmiany (p.Leu798Phe i p.Pro799Thr) znajdujące się w tej samej domenie katalitycznej *DNMT3A*, co najczęstsze mutacje p.Arg882. Badanie przeprowadzone przez Kim i in. wykazało, że somatyczne mutacje w *DNMT3A* i *ASXL1* mogą pełnić istotną rolę w progresji CML i rozwoju oporności na terapię TKI niezależnej od mutacji w *BCR::ABL1* [57]. Istnieją również dane wskazujące, że mutacje w genach zaangażowanych w procesy epigenetyczne (w tym *DNMT3A*) występują częściej u chorych, u których po odstawieniu TKI doszło do nawrotu CML, w porównaniu z chorymi w remisji [153]. Mimo, iż mutacje w *DNMT3A* są często łączone ze zjawiskiem CHIP i podeszłym wiekiem chorych [54], wykrywane są również w pediatrycznych próbkach nowotworowych [154].

*TP53* (*Tumor protein p53*) jest genem supresorowym zaangażowanym w wiele kluczowych procesów na poziomie komórkowym, w tym regulację cyklu komórkowego i mechanizmy naprawy uszkodzeń DNA. *TP53* jest często zmutowany w guzach litych, natomiast w przypadku CML mutacje obserwuje się u około 20% chorych w fazie BP [155]. W badaniu prezentowanym w rozprawie, patogenne warianty w genie *TP53* zostały zidentyfikowane jedynie w fazie BP u 16% chorych (7/45, rycina 11-12). Obserwowane zmiany miały charakter typowych dla tego genu patogennych wariantów SNV lub płytkich delecji wynikających z delecji całego krótkiego ramienia chromosomu 17 lub jego części. Połączenie wariantu SNV i delecji, skutkujące biallelicznym wyłączeniem *TP53* zaobserwowano u 2 chorych. Mutacje w *TP53*, chociaż rzadko, są obserwowane również w CML w fazie CP, a ich obecność może mieć związek z nieoptymalną odpowiedzią na leczenie lub progresją choroby [156]. Istnieją badania wskazujące, że *TP53* jest selektywnie aktywowany przez imatynib w komórkach CML, a mutacje w tym genie mogą przyczyniać się do indukowania oporności na terapię z tym inhibitorem [157].

W 9% próbek BP (4/45) zaobserwowano płytkie i głębokie delecje genów *CDKN2A* i *CDKN2B* (*Cyclin Dependent Kinase Inhibitor 2A/B*), których produkty białkowe zaangażowane są w regulację cyklu komórkowego i apoptozy. Co ciekawe, we wszystkich próbkach z delecją *CDKN2A/B* stwierdzono również kodelecję *IKZF1*, a zmiany te nie występowały w próbkach sprzed progresji choroby (rycina 11-12). Delecje *CDKN2A/B* zostały wcześniej opisane w CML jako zmiany charakterystyczne dla limfoidalnej fazy BP [63, 66], jednak najczęściej identyfikowane są w ALL u dorosłych i dzieci [158]. Wykazano, że bialleliczna utrata *CDKN2A* często współwystępuje z delecją *IKZF1* lub *PAX5*, a także jest związana ze słabą odpowiedzią na leczenie sterydami oraz zwiększonym ryzykiem nawrotu ALL z prekursorów limfocytów B u dzieci [159]. Natomiast kodelecja *CDKN2A* z *IKZF1* u dorosłych chorych na ALL koreluje z istotnie gorszym czasem przeżycia całkowitego w porównaniu do delecji tylko jednego lub żadnego z tych genów [160].

U dwóch pacjentów wykryto również patogenne warianty w *PTPN11* (*Protein tyrosine phosphatase non-receptor type 11*) w próbkach BP (brak w sparowanych próbkach sprzed progresji CML) (ryciny 11-12). W prawidłowej hematopoezie *PTPN11* jest regulatorem szlaku sygnałowego RAS/MAPK, a mutacje nabycia funkcji (ang. *gain-of-function*) prowadzą do hiperaktywacji tego szlaku i w konsekwencji do transformacji nowotworowej [161]. Mutacje w *PTPN11* są często obserwowane w różnych

nowotworach układu krwiotwórczego tj. AML, CML, MDS, ALL a także w guzach litych np. w raku płuc, piersi, jelita grubego. Przeprowadzona analiza wykazała obecność patogennych wariantów w najczęściej zmutowanych kodonach Ala72, Gly60 i Glu76 m.in. w AML, rasopatiach i zespole Noonan [162].

Znanym czynnikiem prognostycznym w CML są ACA [163]. Analiza danych uzyskanych z sekwencjonowania NGS pozwoliła na identyfikację aberracji chromosomowych, typowych dla CML, takich jak monosomia chromosomu 7, trisomia chromosomu 8, złożony kariotyp, jak również innych, mniej typowych zmian (tabela 20). Do innych aberracji ACA najczęściej pojawiających się w publikacjach naukowych [61, 66], ale niezaobserwowanych w wynikach niniejszej pracy doktorskiej zalicza się trisomię chromosomów 17 i 19 oraz dodatkowy chromosom Ph (w prezentowanej rozprawie zaobserwowano jedynie zwiększoną liczbę kopii *BCR::ABL1*, ale nie pozostałych genów zlokalizowanych na chromosomie Ph). Kompleksowej analizy ACA dokonali Hehlmann i in., którzy na podstawie danych dla ponad 1500 chorych na CML leczonych imatynibem wytypowali aberracje, które mimo niskiego odsetka blastów zwiększały ryzyko progresji choroby oraz śmierci [163].

Znajomość patogennych zmian może mieć również przełożenie na dokładniejszą klasyfikację i określenie podtypu choroby. Jak wynika z przeprowadzonych dotychczas badań naukowych, mimo wielu nakładających się zmian genetycznych, fazy blastyczne limfo- i mieloidalne mają odmienne profile molekularne. Do zmian typowych dla limfoidalnej fazy BP zalicza się: delecje *IKZF1*, *CDKN2A/B*, monosomie chromosomu 7 i 9. Z kolei w mieloidalnej fazy BP częściej obserwuje się mutacje w *ASXL1* i *TP53* oraz trisomie chromosomów 8, 19 i 21. W obu rodzajach kryzy z podobną częstością obserwuje się mutacje w *RUNX1* i dodatkowy chromosom Ph [61, 64, 66, 164].

Analiza danych klinicznych w połączeniu z wynikami sekwencjonowania NGS umożliwia identyfikację nie tylko czynników prognostycznych, ale i predykcyjnych. W 2019 roku Nteliopoulos i in. opublikowali wyniki badań przeprowadzone na grupie 124 chorych na CML w fazie CP, od których zebrano materiał do badań przed i po rozpoczęciu leczenia z użyciem imatynibu lub TKI drugiej generacji, a następnie próbki te poddano sekwencjonowaniu 71 genów metodą NGS. Analiza umożliwiła identyfikację somatycznych mutacji w 21 genach u 30% chorych, z których większość nie odpowiadała na leczenie. Do genów z największą liczbą mutacji należały: *ASXL1*, *IKZF1*, *DNMT3A*, *CREBBP*, *KMT2D*, *KMT2E* i *EP300*. Co ciekawe, obecność patogennych wariantów negatywnie wpływała na szybkość osiągnięcia większej

odpowiedzi molekularnej (ang. *major molecular response*), czas przeżycia wolny od progresji (PFS, ang. *progression-free survival*) i EFS w kohorcie chorych leczonych imatynibem, ale nie u chorych leczonych TKI drugiej generacji [165]. Związek pomiędzy mutacjami w genach innych niż *BCR::ABL1* a odpowiedzią na leczenie wykazały również inne badania, w których analiza *in silico* mutacji w genach często zmutowanych w nowotworach pozwoliła na identyfikację pacjentów z wysokim ryzykiem niepowodzenia terapii z zastosowaniem imatynibu [166].

Ewolucja klonalna komórek nowotworowych jest efektem gromadzenia się w nich różnych mutacji somatycznych, mających wpływ na fenotyp komórek, czy wrażliwość na leki. Modelowym przykładem rozwoju nowotworu w wyniku ewolucji klonalnej jest AML [167], pojawiają się jednak doniesienia o jej roli również w patogenezie CML [59, 114]. Przedstawiona w rozprawie analiza porównawcza profili mutacyjnych 25 sparowanych próbek pozwoliła na określenie dynamiki patogennych zmian i ich różny charakter. Progresji choroby towarzyszył zarówno zanik patogennych wariantów/próbek (*ASXLI*), jak również pojawianie się lub wzrost ich ilości (*ABL1*, *RUNX1*, *IKZF1*, *PTPN11*, *CDKN2A/B*) albo brak zmian (*DNMT3A*, *SMAD4*, *CHEK2*) (tabela 21). Różnice obserwowane w obrębie próbek z odmiennych punktów czasowych u poszczególnych pacjentów sugerują obecność różnych klonów białaczkowych ulegających ekspansji lub regresji wraz z progresją choroby. W przeprowadzonej analizie, brak dostępu do DNA linii zarodkowej uniemożliwił jednoznaczne określenie charakteru wykrytych zmian, jednak VAF większości z nich sugeruje, że są to warianty somatyczne. Niewątpliwą wartością miałyby skorelowanie powyższej analizy z dokładnymi danymi klinicznymi dotyczącymi zastosowanej terapii i poziomu transkrypty *BCR::ABL1*, które umożliwiłyby precyzyjne określenie zależności pomiędzy obserwowanymi aberracjami a odpowiedzią na leczenie i przebiegiem choroby. Bez tych danych można jedynie spekulować, że utrata i pojawianie się patogennych zmian w trakcie choroby wiążą się odpowiednio z eradykacją białaczkowych klonów wskutek właściwie dobranej terapii lub progresją choroby / opornością na zastosowanie leczenia. Schmidt i in. wykonali głębokie sekwencjonowanie NGS 25 genów w komórkach z i bez chromosomu Ph, pochodzących od chorych na CML. Analiza wykazała obecność mutacji w genach innych niż *BCR::ABL1* (np. w *ASXLI*, *DNMT3A*, *EZH2*, *RUNX1*, *TP53* i *TET2*) zarówno w klonach komórkowych z chromosomem Ph, wyizolowanych z próbek CP na etapie diagnozy, jak i klonach bez tego chromosomu w trakcie leczenia TKI. W niektórych przypadkach obserwowane zmiany były obecne jednocześnie w klonach

z chromosomem Ph, jak też w sparowanych próbkach bez chromosomu Ph, przez co autorzy sugerują, że mutacje niezależne od *BCR::ABL1* mogą stanowić ważny czynnik ewolucji klonalnej w CML. Autorzy powyższego badania proponują 3 modele ewolucji klonalnej CML, w których mutacje dodatkowe występują w obrębie klonu z translokacją *BCR::ABL1*, poprzedzają ją lub występują w całkowicie odrębnych klonach [59]. Inne badanie, w ramach którego zsekwencjonowano seryjne próbki od 100 chorych na CML, pochodzące sprzed rozpoczęcia i po zakończeniu terapii z TKI, wyodrębniły 5 wzorców ukazujących dynamikę mutacji wynikającą z zastosowanego leczenia. Autorzy wykazali korelację pomiędzy nabywaniem, utrzymywaniem lub zanikiem mutacji w trakcie terapii a przebiegiem choroby. Najlepszą odpowiedź molekularną zaobserwowano u chorych, u których mutacje wykryte na etapie diagnozy, wykrywane były również w próbkach po zastosowaniu TKI pomimo skutecznej eradykacji klonów z chromosomem Ph, ale nie były obserwowane w limfocytach T. Z kolei najgorszy przebieg leczenia zaobserwowano u chorych, w przypadku których mutacje pojawiały się dopiero po terapii z TKI, a które nie były obecne na etapie diagnozy oraz w limfocytach T [114].

Analiza przeprowadzona w ramach niniejszej pracy pozwoliła na identyfikację patogennych wariantów w genach stanowiących potencjalne cele terapeutyczne, takich jak: *RUNX1*, *IKZF1*, *ASXL1*, *TP53* czy *IDH1/2*. Wysokoprzepustowe testy wrażliwości i oporności na leki (ang. *high-throughput drug sensitivity and resistance testing*) przeprowadzone na komórkach pochodzących od chorych na CML ze zmutowanym *RUNX1* wykazały ich wrażliwość na inhibitory mTOR, VEGFR i BCL2 oraz na glikokortykoidy [64]. Co więcej, połączenie tych leków z imatynibem dawało efekt synergistyczny, obserwowany szczególnie w przypadku kombinacji z deksametazonem. Ponadto, ze względu na obserwowany wzrost ekspresji antygenu CD19 u chorych z aberracją w *RUNX1*, zastosowanie immunoterapii z użyciem CAR-T skierowanych przeciwko cząsteczce CD19 wykazało *ex vivo* silną aktywność cytotoksyczną komórek CAR-T w stosunku do komórek białaczkowych. Opisano przykład chorego na AML z t(8;21)(*RUNX1*–*RUNX1T1*), u którego po wdrożeniu immunoterapii CD19 CAR-T zaobserwowano remisję choroby [168]. Inną, równie obiecującą strategią leczenia chorych ze zmutowanym *RUNX1* wydaje się być zastosowanie inhibitorów BET. BRD4 (*Bromodomain containing 4*) jest jednym z białek rodziny BET, które wiążąc się do acetylowanych histonów reguluje transkrypcję wielu genów, w tym *RUNX1* [169]. Zastosowanie inhibitorów BET prowadzi do zahamowania funkcji *RUNX1*, prowadząc tym samym do nasilonej apoptozy komórek białaczkowych

ze zmutowanym *RUNX1* w porównaniu do komórek z niezmutowanym genem. Inne badania pokazują, że inhibitory BET mogą przynieść korzyść również w leczeniu pacjentów z mutacjami *ASXL1*, skracającymi część C-końcową białka, gdyż skrócona forma białka ASXL1 ulega interakcji z BRD4, co w konsekwencji powoduje uwrażliwienie komórek szpiku na inhibitory BET [170]. Kolejnym obiecującym celem terapeutycznym jest *IKZF1*. Mutacje w *Ikzf1* w mysich komórkach AML z chromosomem Ph osłabiają odpowiedź na leczenie dasatynibem. Wykazano, że dodanie do terapii retinoidów nasila działanie dasatynibu, zarówno w komórkach mysich jak i ludzkich, oraz odwraca efekt wywołany mutacjami w *IKZF1*, powodując obniżenie adhezji i proliferacji komórek, zatrzymanie cyklu komórkowego i w konsekwencji śmierć komórek białaczkowych [171]. Potencjalną korzyść kliniczną mogliby również odnieść chorzy, u których wykryto mutacje w *IDH1/2*, gdyż inhibitory tych genów tj. enasidenib (inhibitor *IDH2*) [172] i ivosidenib (inhibitor *IDH1*) [173] zostały zatwierdzone przez FDA w leczeniu AML z mutacjami *IDH1/2*. Pojawia się coraz więcej badań naukowych i klinicznych, w których celami terapeutycznymi są także inne geny często zmutowane w różnych nowotworach hematologicznych, w tym w CML tj.: *TP53* [174], *DNMT3A* [175], *EZH2* [176] i *NOTCH1* [177]. Analiza przeprowadzona w ramach niniejszej pracy doktorskiej wykazała delecję *CDKN2A* w 9% (4/45) analizowanych próbek z fazy BP mającą potencjalnie istotne znaczenie w kontekście odpowiedzi na inhibitory kinaz zależnych od cyklin (CDK4/6) [178]. Ponadto, u pojedynczych chorych wykryto również patogenne warianty w genach uczestniczących w naprawie DNA: *CHEK2*, *ATM*, *BRCA2*, *MLH1*, *MSH2* i *MSH6* stanowiących obiecujący cel terapeutyczny z zastosowaniem inhibitorów PARP w terapii opartej o zjawisko tzw. syntetycznej letalności (ang. *synthetic lethality*), stosowanej m.in. w raku jajnika [179]. Co więcej, istnieją badania sugerujące, że z terapii z użyciem inhibitorów PARP mogliby skorzystać również chorzy ze zmutowanym *TET2*, ale nie *DNMT3A*, gdyż mutacje w tych genach powodują odpowiednio uwrażliwienie i oporność na te inhibitory [151]. Niemniej, w przypadku CML podejście takie wydaje się dość ryzykowne wobec pojawiających się doniesień o wzroście ryzyka rozwoju AML lub MDS po terapii z inhibitorami PARP [180].

Oprócz patogennych zmian genetycznych w genach omówionych powyżej, których rola w CML została już opisana, w toku przeprowadzonej analizy zidentyfikowano również 126 wariantów typu VUS. Zaliczały się do nich głównie warianty zmiany sensu, na których patogenność wskazywały bioinformatyczne predyktory (minimum 3 z zastosowanych 5) i bazy danych, jednak brakowało



dotatkowych danych jednoznacznie potwierdzających ich patogenny charakter. Warianty VUS, wykryte u minimum 3 chorych dotyczyły genów: *MYH11*, *PDE4DIP*, *LRP1B*, *NTRK1* i *PCLO*. Co ciekawe, żaden z tych genów nie został jak dotąd opisany jako często zmutowany w CML za wyjątkiem *PDE4DIP*, w którym mutacje w różnych typach białaczek opisano po raz pierwszy u chińskich chorych w 2020 roku [181]. Według danych TCGA zgromadzonych na portalu cBioPortal (dostępne zbiory Firehose Legacy) mutacje w tych genach występują w wielu typach nowotworów, najczęściej w czerniaku, raku płaskonabłonkowym przetyku i płuc oraz chłoniaku rozlanym z dużych komórek B [182]. *MYH11* jest dobrze opisanym partnerem fuzyjnym w AML, gdzie fuzja *CBFB::MYH11* wynikająca z *inv(16)(p13.1q22)* jest powiązana z korzystnym przebiegiem choroby [183]. Innym genem, uczestniczącym w fuzjach genowych, obserwowanych w AML jest *NTRK1*: wykazano, że niektóre mutacje występujące w tym genie mogą indukować oporność na leczenie za pomocą inhibitorów kinaz tyrozynowych z rodziny TRK [184]. *LRP1B* jest genem supresorowym zmutowanym w wielu nowotworach. Chorzy ze zmutowanym *LRP1B* mogą odnieść korzyść z terapii z użyciem inhibitorów punktów kontrolnych (ang. *checkpoint inhibitors*) [185]. Podsumowując, istnieje potrzeba dalszych badań, zwłaszcza funkcjonalnych, które pozwolą na określenie roli powyższych genów zarówno w patogenezie, jak i przebiegu CML.

Interesującą zmianą wykrytą w trakcie przeprowadzonej analizy a jednocześnie nieopisaną jak dotąd w nowotworach, w tym w CML, był przetworzony pseudogen *SMAD4*. W niniejszej pracy doktorskiej podjęto próbę oceny jego znaczenia w przebiegu CML.  $\Psi$ *SMAD4* został początkowo wykryty u 3 chorych w fazie BP (7%, 3/45). Przeprowadzono poszerzony skrining w różnych typach białaczek, który wykazał istotny statystycznie wzrost częstości występowania  $\Psi$ *SMAD4* jedynie w przypadku CML, ale nie AML, ALL i CLL, w porównaniu do obserwowanej częstości w polskiej populacji (tabela 22). Co więcej, wydaje się, że częstość występowania pseudogenu może różnić się w zależności od badanej populacji. Wykazano, że częstość występowania  $\Psi$ *SMAD4* w badanej grupie kontrolnej była dwukrotnie wyższa niż obserwowana w amerykańskich (0,26%) i brytyjskich (0,24%) badaniach naukowych, w których materiał stanowiły próbki od osób skierowanych do badań genetycznych oceniających m.in. predyspozycje do nowotworów dziedzicznych [97, 98]. W następnym etapie potwierdzono zarodkowy charakter  $\Psi$ *SMAD4*, określono jego dokładną pozycję genomową oraz sekwencję (ryciny 17-20). Co ciekawe, jedyną zmianą odróżniającą pseudogen od genu

rodzicielskiego były skrócone końce 5' i 3' UTR, nie zaobserwowano natomiast żadnych różnic w sekwencji CDS (rycina 19-20). Otrzymane wyniki dotyczące lokalizacji i sekwencji  $\Psi$ *SMAD4* są zgodne z opublikowanymi dotychczas danymi [98]. Kolejno postanowiono sprawdzić, czy próbki z  $\Psi$ *SMAD4* posiadają inne warianty, odróżniające je od próbek CML bez pseudogenu. Analiza danych z sekwencjonowania 193 genów wykazała, że profil zmian genetycznych próbek z  $\Psi$ *SMAD4* był zbliżony do profilu próbek CML bez pseudogenu, przeanalizowanych w ramach niniejszej pracy doktorskiej (rycina 22). Poza odczytami w genie *SMAD4* potwierdzającymi istnienie pseudogenu, wykryto, obserwowane u minimum 2 chorych, patogenne zmiany w genach *DNMT3A*, *ABL1* i *ASXL1*, których znaczenie omówiono powyżej. U jednego chorego w fazie CP wykryto wariant nonsensowny w genie *MYH11* (*Myosin heavy chain 11*). *MYH11* jest dobrze opisanym partnerem fuzyjnym genu *CBFB* w AML [183], jednak mutacje punktowe w tym genie w nowotworach układu krwiotwórczego pojawiają się rzadko, a ich znaczenie nie jest do końca poznane [186]. Istnieją badania naukowe wskazujące, że obniżona ekspresja *MYH11* lub warianty SNV mogą mieć związek z rozwojem raka jelita grubego [187] i wywołują fenotyp zbliżony do fenotypu zespołu polipowatości młodzieńczej [188]. Co ciekawe, w patogenezie tych chorób istotną rolę odgrywa również *SMAD4*.

W celu sprawdzenia czy  $\Psi$ *SMAD4* ulega ekspresji wykonano analizę transkryptomu techniką RNAseq. Z powodu braku różnic w sekwencji  $\Psi$ *SMAD4* i mRNA *SMAD4* niemożliwe było wykrycie odczytów specyficznych jedynie dla pseudogenu, które bezsprzecznie dowodziłyby jego ekspresji. Oceny potencjalnej ekspresji  $\Psi$ *SMAD4* dokonano więc w sposób pośredni, poprzez porównanie poziomu ekspresji *SMAD4* w próbkach z i bez pseudogenu. Analiza nie wykazała istotnych różnic w ekspresji *SMAD4*, co sugeruje prawdopodobny brak ekspresji i wpływu  $\Psi$ *SMAD4* na funkcjonowanie genu rodzicielskiego na poziomie RNA. Drugim sposobem na potwierdzenie braku ekspresji  $\Psi$ *SMAD4* była manualna inspekcja miejsca insercji pseudogenu w programie IGV, podczas której nie zaobserwowano odczytów sprzecznych (ang. *discordant reads*), z których jeden odczyt z pary mapowałby się do pseudogenu a drugi do miejsca jego insercji, ani odczytów rozszczepionych (ang. *split reads*), gdzie część jednego odczytu mapowałaby się do *SMAD4*, a druga część tego samego odczytu - do intronu *SCAI*. Wreszcie, trzecim sposobem była bioinformatyczna analiza danych, w której surowe odczyty z RNAseq przyrównano do sztucznej referencji stanowiącej dokładną sekwencję  $\Psi$ *SMAD4* wraz z okalającą ją sekwencją ostatniego

intronu *SCAI*. Ponownie nie wykryto odczytów mapujących się jednocześnie do  $\Psi$ *SMAD4* i miejsca jego insercji. Podobne podejście do detekcji ekspresji pseudogenów w swojej pracy naukowej zastosowali Cooke i in., którzy za pomocą odczytów sprzecznych potwierdzili ekspresję somatycznego przetworzonego pseudogenu *KTNI* zlokalizowanego w intronie genu *PSD3* w raku płuca [78].

Pseudogeny mogą mieć wpływ na funkcjonowanie nie tylko genów rodzicielskich, ale również innych, niezwiązanych genów, w tym tych będących miejscem ich insercji. Z tego względu porównano poziom ekspresji *SCAI* w próbkach z  $\Psi$ *SMAD4* i próbkach kontrolnych, jednak analiza nie wykazała istotnych zmian. Porównując obie grupy nie zaobserwowano także istotnych różnic w poziomie ekspresji wszystkich pozostałych genów poddanych analizie. Klastrowanie hierarchiczne wykazało grupowanie się próbek jedynie względem fazy choroby, a różnice ekspresji genów pomiędzy fazą CP i BP CML zostały już wcześniej opisane [63, 189, 190]. Uzyskane wyniki z analizy RNAseq sugerują, że  $\Psi$ *SMAD4* nie ulega ekspresji, a tym samym nie pełni istotnej roli na poziomie RNA w stosunku do genu rodzicielskiego *SMAD4*, względem *SCAI*, stanowiącego miejsce jego insercji ani też względem innych genów, ulegających ekspresji w analizowanych próbkach. Mając na uwadze umiejscowienie pseudogenu można przypuszczać, że  $\Psi$ *SMAD4* ulega wycięciu razem intronem *SCAI* (stanowiącym miejsce insercji) w trakcie obróbki potranskrypcyjnej podczas splicingu pre-mRNA.

Ciekawą kwestią wydaje się położenie  $\Psi$ *SMAD4*, zwłaszcza w kontekście istotnie zwiększonej częstości jego występowania w CML, ponieważ jest on zlokalizowany proksymalnie od genu *ABLI*, w odległości niecałych 6 Mbp (rycina 18C). Otwartym pozostaje więc pytanie, czy możliwe jest aby insercja DNA o długości niecałych 3,5 kbp wpływała na osłabienie struktury długiego ramienia chromosomu 9, przyczyniając się w konsekwencji do zwiększonej częstości translokacji t(9;22)(q34;q11)? Istnieje potrzeba przeprowadzenia dodatkowych badań, w których ocenionoby potencjalny wpływ  $\Psi$ *SMAD4* na powstawanie translokacji t(9;22). Dokładna przyczyna powstawania chromosomu Ph na poziomie molekularnym nie została jak dotąd odkryta, a narażenie na promieniowanie jonizujące jest jedynym znanym zewnętrznym czynnikiem zwiększającym ryzyko rozwoju CML [7]. Opublikowano również badania wskazujące na przestrzennie bliskie położenie regionów 9q34 i 22q11 względem siebie w jądrze, ułatwiające ich wzajemną translokację [191]. Innym czynnikiem potencjalnie ułatwiającym powstawanie chromosomu Ph może być również duplikon (specyficzne dla

chromosomu powtórzenie o niskiej liczbie kopii) o długości 76 kpb mapujący się w pobliżu genu *BCR* oraz *ABL1* [192]. Istnieją także sprzeczne doniesienia dotyczące udziału sekwencji Alu w powstawaniu translokacji t(9;22) [193, 194].

Podsumowując, analiza danych uzyskanych z sekwencjonowania NGS 25 sparowanych próbek CML o niekorzystnym przebiegu i kolejnych 20 próbek z fazy BP umożliwiła wytypowanie zmian o charakterze prognostycznym (m.in. ACA oraz patogene warianty w genach *ASXL1*, *IKZF1*, *RUNX1*) oraz zmian o potencjalnym znaczeniu terapeutycznym (patogene warianty w *RUNX1*, *IKZF1*, *ASXL1*, *TP53*, *IDH1/2*). Z uwagi na długi okres czasu (kilka lat), w którym zbierano materiał do badań od chorych na CML w fazie BP i wykonywano sekwencjonowanie z użyciem różnych, dostępnych w danym momencie paneli genowych, analiza mutacji została przeprowadzona dla niemal 200 wspólnych genów, zsekwencjonowanych we wszystkich próbkach. Ograniczenie liczby analizowanych genów miało przełożenie na uzyskane wyniki, które częściowo stanowiły potwierdzenie opublikowanych dotychczas danych, ale zidentyfikowano również nowe, dotychczas nieopisane patogene warianty (np. w *DNMT3A*), jak również warianty VUS w genach (m.in. *MYH11*, *PDE4DIP*, *LRP1B*, *NTRK1* i *PCLO*), których rola w patogenezie CML nie została jasno określona i wymaga dalszych badań. Przedstawione w rozprawie wyniki z analizy porównawczej profili mutacyjnych sparowanych próbek pozwoliły na określenie dynamiki patogennych zmian w trakcie progresji CML, mogących wskazywać na aktywny proces ewolucji klonalnej w komórkach białaczkowych. Wartość analizy porównawczej z pewnością zwiększyłoby skorelowanie otrzymanych wyników z danymi klinicznymi, co być może pozwoliłoby na identyfikację dodatkowych, istotnych zmian o znaczeniu prognostycznym lub predykcyjnym. Do przeprowadzenia takiej analizy potrzebne byłyby dane dotyczące m.in. poziomu transkryptu *BCR::ABL1*, linii leczenia oraz przeżycia, które nie były dostępne dla wszystkich chorych włączonych do analizy. Nieoczekiwaną zmianą zarodkową o zwiększonej częstości występowania w CML okazał się  $\Psi$ *SMAD4*, który jak dotąd nie został opisany w kontekście chorób nowotworowych. Przeprowadzona analiza wykazała brak jego ekspresji a tym samym brak funkcjonalnej roli na poziomie RNA. Jednakże jego bliskie położenie względem genu *ABL1* będącego partnerem fuzyjnym w translokacji prowadzącej do powstania chromosomu Ph sugeruje, że badany pseudogen potencjalnie mógłby mieć funkcjonalne znaczenie na poziomie DNA, zwiększając podatność chromosomu 9 na złamania prowadzące do rozwoju CML. Hipoteza ta wymaga jednak potwierdzenia w dalszych badaniach.

## 6. WNIOSKI

Dzięki zastosowaniu sekwencjonowania następnej generacji w próbkach od chorych z CML o niekorzystnym przebiegu choroby zidentyfikowano zarówno nowe, jak i wcześniej opisane patogenne warianty, w tym warianty o charakterze zmian punktowych oraz zmiany liczby kopii i dodatkowe aberracje chromosomowe.

1. Zidentyfikowane zmiany mają charakter prognostyczny (m.in. ACA oraz patogenne warianty w genach *ASXL1*, *IKZF1*, *RUNX1*) jak również potencjalne znaczenie predykcyjne (patogenne warianty w *RUNX1*, *IKZF1*, *ASXL1*, *TP53*, *IDH1/2*). Część tych zmian była możliwa do wykrycia w próbkach od chorych w diagnozie lub przed progresją choroby, przez co wskazane może być w przyszłości włączenie w diagnostykę genetyczną genów innych niż *BCR::ABL1*.

2. Analiza porównawcza 25 sparowanych próbek CML umożliwiła określenie różnic w występowaniu patogennych zmian w poszczególnych genach przed i po progresji CML. Zidentyfikowano 7 wzorców dynamiki zmian genetycznych wynikających z nabywania, utrzymywania się lub zanikania aberracji genetycznych.

3. Przetworzony pseudogen *SMAD4* jest wariantem zarodkowym, dotychczas nieopisanym w przewlekłej białaczce szpikowej, a jego wzrost częstości występowania w tej chorobie jest znamieny statystycznie. Jednocześnie wykazano, że  $\Psi$ *SMAD4* nie ulega transkrypcji i nie wpływa na ekspresję innych genów. Przeprowadzone badania sugerują, że obserwowany  $\Psi$ *SMAD4* jest najprawdopodobniej aberracją genetyczną, nie mającą bezpośredniego wpływu na progresję CML, jednak mogącą mieć rolę w inicjacji i patogenezie tej choroby, co wymaga dalszych badań.

## PIŚMIENNICTWO

- [1] Sung H., Ferlay J., Siegel R.L., et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 71. 2021/02/05 ed.; 2021:209-49.
- [2] Dong Y., Shi O., Zeng Q., et al. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. *Exp Hematol Oncol* 2020;9:14.
- [3] Żołnierowicz J., Kawiak J., Hoser G. Patogeneza przewlekłej białaczki szpikowej - od genu do terapii celowanej. *Hematologia* 2010;1(3):195-218.
- [4] Nowell P.C., Hungerford D.A. Chromosome studies on normal and leukemic human leukocytes. *J Natl Cancer Inst* 1960;25:85-109.
- [5] Druker B.J., Guilhot F., O'Brien S.G., et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N Engl J Med* 2006;355(23):2408-17.
- [6] Khoury J.D., Solary E., Abla O., et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* 2022;36(7):1703-19.
- [7] Preston D.L., Kusumi S., Tomonaga M., et al. Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987. *Radiat Res* 1994;137(2 Suppl):S68-97.
- [8] Hoglund M., Sandin F., Simonsson B. Epidemiology of chronic myeloid leukaemia: an update. *Ann Hematol* 2015;94 Suppl 2:S241-7.
- [9] An X., Tiwari A.K., Sun Y., et al. BCR-ABL tyrosine kinase inhibitors in the treatment of Philadelphia chromosome positive chronic myeloid leukemia: a review. *Leuk Res* 2010;34(10):1255-68.
- [10] Hochhaus A., Baccarani M., Silver R.T., et al. European LeukemiaNet 2020 recommendations for treating chronic myeloid leukemia. *Leukemia* 2020;34(4):966-84.
- [11] Wojciechowska U., Didkowska J., Michałek I., Olasek P., Ciuba A. Nowotwory złośliwe w Polsce w 2018 roku, [http://onkologia.org.pl/wp-content/uploads/Nowotwory\\_2018.pdf](http://onkologia.org.pl/wp-content/uploads/Nowotwory_2018.pdf) 2018.
- [12] Budziszewska B.K., Więckowska B., Lech-Marańda E., et al. Zachorowalność i chorobowość na nowotwory układu krwiotwórczego w Polsce (2009–2015) określone na podstawie analizy danych Narodowego Funduszu Zdrowia wykorzystanych w projekcie „Mapy potrzeb zdrowotnych - baza analiz systemowych i wdrożeniowych”. *Hematologia* 2017;8(2):89-104.
- [13] Geary C.G. The story of chronic myeloid leukaemia. *Br J Haematol* 2000;110(1):2-11.
- [14] Goldman J.M. Chronic myeloid leukemia: a historical perspective. *Semin Hematol* 2010;47(4):302-11.
- [15] Ali M.A. Chronic Myeloid Leukemia in the Era of Tyrosine Kinase Inhibitors: An Evolving Paradigm of Molecularly Targeted Therapy. *Mol Diagn Ther* 2016;20(4):315-33.
- [16] Zhou T., Medeiros L.J., Hu S. Chronic Myeloid Leukemia: Beyond BCR-ABL1. *Curr Hematol Malig Rep* 2018;13(6):435-45.
- [17] Gong Z., Medeiros L.J., Cortes J.E., et al. Clinical and prognostic significance of e1a2 BCR-ABL1 transcript subtype in chronic myeloid leukemia. *Blood Cancer J* 2017;7(7):e583.
- [18] Sacha T. Zalecenia postępowania diagnostyczno-terapeutycznego w nowotworach złośliwych 2019 rok: Przewlekła Białaczka Szpikowa. *Polskie Towarzystwo Onkologii Klinicznej*; 2019.
- [19] Kang Z.J., Liu Y.F., Xu L.Z., et al. The Philadelphia chromosome in leukemogenesis. *Chin J Cancer* 2016;35:48.
- [20] Quintas-Cardama A., Cortes J. Molecular biology of bcr-abl1-positive chronic myeloid leukemia. *Blood* 2009;113(8):1619-30.

- [21] Sattler M., Griffin J.D. Molecular mechanisms of transformation by the BCR-ABL oncogene. *Semin Hematol* 2003;40(2 Suppl 2):4-10.
- [22] Cheng K., Kurzrock R., Qiu X., et al. Reduced focal adhesion kinase and paxillin phosphorylation in BCR-ABL-transfected cells. *Cancer* 2002;95(2):440-50.
- [23] Thompson P.A., Kantarjian H.M., Cortes J.E. Diagnosis and Treatment of Chronic Myeloid Leukemia in 2015. *Mayo Clin Proc* 2015;90(10):1440-54.
- [24] Granatowicz A., Piatek C.I., Moschiano E., et al. An Overview and Update of Chronic Myeloid Leukemia for Primary Care Physicians. *Korean J Fam Med* 2015;36(5):197-202.
- [25] Radich J.P., Deininger M., Abboud C.N., et al. Chronic Myeloid Leukemia, Version 1.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* 2018;16(9):1108-35.
- [26] Melo J.V., Barnes D.J. Chronic myeloid leukaemia as a model of disease evolution in human cancer. *Nat Rev Cancer* 2007;7(6):441-53.
- [27] Deregowska A., Pepek M., Pruszczyk K., et al. Differential Regulation of Telomeric Complex by BCR-ABL1 Kinase in Human Cellular Models of Chronic Myeloid Leukemia-From Single Cell Analysis to Next-Generation Sequencing. *Genes (Basel)* 2020;11(10).
- [28] Hehlmann R., Lauseker M., Voskanyan A., et al. Impact of emerging ACA on survival in chronic myeloid leukemia (CML). *Leukemia* 2022;36(10):2544-7.
- [29] Hochhaus A., Saussele S., Rosti G., et al. Chronic myeloid leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017;28(suppl\_4):iv41-iv51.
- [30] Ning L., Hu C., Lu P., et al. Trends in disease burden of chronic myeloid leukemia at the global, regional, and national levels: a population-based epidemiologic study. *Exp Hematol Oncol* 2020;9(1):29.
- [31] Cohen M.H., Williams G., Johnson J.R., et al. Approval summary for imatinib mesylate capsules in the treatment of chronic myelogenous leukemia. *Clin Cancer Res* 2002;8(5):935-42.
- [32] Sacha T., Lewandowski K., Hellmann A., et al. Rekomendacje PALG dotyczące diagnostyki i leczenia przewlekłej białaczki szpikowej w 2013r. *Acta Haematologica Polonica* 2013;44:345-62.
- [33] Rinke J., Hochhaus A., Ernst T. CML - Not only BCR-ABL1 matters. *Best Pract Res Clin Haematol* 2020;33(3):101194.
- [34] Goldman J.M., Melo J.V. BCR-ABL in chronic myelogenous leukemia--how does it work? *Acta Haematol* 2008;119(4):212-7.
- [35] Cortes J.E., Kim D.W., Pinilla-Ibarz J., et al. A phase 2 trial of ponatinib in Philadelphia chromosome-positive leukemias. *N Engl J Med* 2013;369(19):1783-96.
- [36] Yeung D.T., Shanmuganathan N., Hughes T.P. Asciminib: a new therapeutic option in chronic-phase CML with treatment failure. *Blood* 2022;139(24):3474-9.
- [37] Jabbour E., Kantarjian H. Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. *Am J Hematol* 2020;95(6):691-709.
- [38] Saussele S., Richter J., Hochhaus A., Mahon F.X. The concept of treatment-free remission in chronic myeloid leukemia. *Leukemia* 2016;30(8):1638-47.
- [39] Saussele S., Richter J., Guilhot J., et al. Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. *Lancet Oncol* 2018;19(6):747-57.
- [40] Richter J., Lubking A., Soderlund S., et al. Molecular status 36 months after TKI discontinuation in CML is highly predictive for subsequent loss of MMR-final report from AFTER-SKI. *Leukemia* 2021;35(8):2416-8.
- [41] Ashley E.A. Towards precision medicine. *Nat Rev Genet* 2016;17(9):507-22.
- [42] Morganti S., Tarantino P., Ferraro E., et al. Next Generation Sequencing (NGS): A Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer. *Adv Exp Med Biol* 2019;1168:9-30.

- [43] Zacher A., Kaulich K., Stepanow S., et al. Molecular Diagnostics of Gliomas Using Next Generation Sequencing of a Glioma-Tailored Gene Panel. *Brain Pathol* 2017;27(2):146-59.
- [44] Endris V., Stenzinger A., Pfarr N., et al. NGS-based BRCA1/2 mutation testing of high-grade serous ovarian cancer tissue: results and conclusions of the first international round robin trial. *Virchows Arch* 2016;468(6):697-705.
- [45] D'Haene N., Fontanges Q., De Neve N., et al. Clinical application of targeted next-generation sequencing for colorectal cancer patients: a multicentric Belgian experience. *Oncotarget* 2018;9(29):20761-8.
- [46] Dohner H., Estey E., Grimwade D., et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;129(4):424-47.
- [47] Coccaro N., Anelli L., Zagaria A., Specchia G., Albano F. Next-Generation Sequencing in Acute Lymphoblastic Leukemia. *Int J Mol Sci* 2019;20(12).
- [48] Vainchenker W., Kralovics R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* 2017;129(6):667-79.
- [49] Maurya N., Mohanty P., Dhangar S., et al. Comprehensive analysis of genetic factors predicting overall survival in Myelodysplastic syndromes. *Sci Rep* 2022;12(1):5925.
- [50] Fialkow P.J., Martin P.J., Najfeld V., et al. Evidence for a multistep pathogenesis of chronic myelogenous leukemia. *Blood* 1981;58(1):158-63.
- [51] Boquett J.A., Alves J.R., de Oliveira C.E. Analysis of BCR/ABL transcripts in healthy individuals. *Genet Mol Res* 2013;12(4):4967-71.
- [52] Grossmann V., Kohlmann A., Zenger M., et al. A deep-sequencing study of chronic myeloid leukemia patients in blast crisis (BC-CML) detects mutations in 76.9% of cases. *Leukemia* 2011;25(3):557-60.
- [53] Fabarius A., Leitner A., Hochhaus A., et al. Impact of additional cytogenetic aberrations at diagnosis on prognosis of CML: long-term observation of 1151 patients from the randomized CML Study IV. *Blood* 2011;118(26):6760-8.
- [54] Genovese G., Kahler A.K., Handsaker R.E., et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014;371(26):2477-87.
- [55] Steensma D.P. Clinical consequences of clonal hematopoiesis of indeterminate potential. *Blood Adv* 2018;2(22):3404-10.
- [56] Khorashad J.S., Kelley T.W., Szankasi P., et al. BCR-ABL1 compound mutations in tyrosine kinase inhibitor-resistant CML: frequency and clonal relationships. *Blood* 2013;121(3):489-98.
- [57] Kim T., Tyndel M.S., Zhang Z., et al. Exome sequencing reveals DNMT3A and ASXL1 variants associate with progression of chronic myeloid leukemia after tyrosine kinase inhibitor therapy. *Leuk Res* 2017;59:142-8.
- [58] Makishima H., Jankowska A.M., McDevitt M.A., et al. CBL, CBLB, TET2, ASXL1, and IDH1/2 mutations and additional chromosomal aberrations constitute molecular events in chronic myelogenous leukemia. *Blood* 2011;117(21):e198-206.
- [59] Schmidt M., Rinke J., Schafer V., et al. Molecular-defined clonal evolution in patients with chronic myeloid leukemia independent of the BCR-ABL status. *Leukemia* 2014;28(12):2292-9.
- [60] Togasaki E., Takeda J., Yoshida K., et al. Frequent somatic mutations in epigenetic regulators in newly diagnosed chronic myeloid leukemia. *Blood Cancer J* 2017;7(4):e559.
- [61] Branford S., Wang P., Yeung D.T., et al. Integrative genomic analysis reveals cancer-associated mutations at diagnosis of CML in patients with high-risk disease. *Blood* 2018;132(9):948-61.



- [62] Adnan Awad S., Kankainen M., Ojala T., et al. Mutation accumulation in cancer genes relates to nonoptimal outcome in chronic myeloid leukemia. *Blood Adv* 2020;4(3):546-59.
- [63] Ko T.K., Javed A., Lee K.L., et al. An integrative model of pathway convergence in genetically heterogeneous blast crisis chronic myeloid leukemia. *Blood* 2020;135(26):2337-53.
- [64] Adnan Awad S., Dufva O., Ianevski A., et al. RUNX1 mutations in blast-phase chronic myeloid leukemia associate with distinct phenotypes, transcriptional profiles, and drug responses. *Leukemia* 2021;35(4):1087-99.
- [65] Magistrini V., Mauri M., D'Aliberti D., et al. De novo UBE2A mutations are recurrently acquired during chronic myeloid leukemia progression and interfere with myeloid differentiation pathways. *Haematologica* 2019;104(9):1789-97.
- [66] Ochi Y., Yoshida K., Huang Y.J., et al. Clonal evolution and clinical implications of genetic abnormalities in blastic transformation of chronic myeloid leukaemia. *Nat Commun* 2021;12(1):2833.
- [67] Jacq C., Miller J.R., Brownlee G.G. A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* 1977;12(1):109-20.
- [68] Pei B., Sisu C., Frankish A., et al. The GENCODE pseudogene resource. *Genome Biol* 2012;13(9):R51.
- [69] Parker H.G., VonHoldt B.M., Quignon P., et al. An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 2009;325(5943):995-8.
- [70] Wang J., Pitarque M., Ingelman-Sundberg M. 3'-UTR polymorphism in the human CYP2A6 gene affects mRNA stability and enzyme expression. *Biochem Biophys Res Commun* 2006;340(2):491-7.
- [71] Poliseno L., Salmena L., Zhang J., et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 2010;465(7301):1033-8.
- [72] Yang H. Conserved or lost: molecular evolution of the key gene *GULO* in vertebrate vitamin C biosynthesis. *Biochem Genet* 2013;51(5-6):413-25.
- [73] GENCODE, <https://www.genencodegenes.org/human/stats.html>.
- [74] Zhang Z., Harrison P.M., Liu Y., Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003;13(12):2541-58.
- [75] Esnault C., Maestre J., Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000;24(4):363-7.
- [76] Ding W., Lin L., Chen B., Dai J. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* 2006;58(12):677-85.
- [77] Vinckenbosch N., Dupanloup I., Kaessmann H. Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A* 2006;103(9):3220-5.
- [78] Cooke S.L., Shlien A., Marshall J., et al. Processed pseudogenes acquired somatically during cancer development. *Nat Commun* 2014;5:3644.
- [79] Poliseno L. Pseudogenes: newly discovered players in human cancer. *Sci Signal* 2012;5(242):re5.
- [80] Lu W., Zhou D., Glusman G., et al. *KLK31P* is a novel androgen regulated and transcribed pseudogene of kallikreins that is expressed at lower levels in prostate cancer cells than in normal prostate cells. *Prostate* 2006;66(9):936-44.
- [81] Puget N., Gad S., Perrin-Vidoz L., et al. Distinct *BRCA1* rearrangements involving the *BRCA1* pseudogene suggest the existence of a recombination hot spot. *Am J Hum Genet* 2002;70(4):858-65.
- [82] Chan W.L., Yuo C.Y., Yang W.K., et al. Transcribed pseudogene *psiPPM1K* generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. *Nucleic Acids Res* 2013;41(6):3734-47.

- [83] Thomson D.W., Dinger M.E. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* 2016;17(5):272-83.
- [84] Bier A., Oviedo-Landaverde I., Zhao J., et al. Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. *Mol Cancer Ther* 2009;8(4):786-93.
- [85] Zou M., Baitei E.Y., Alzahrani A.S., et al. Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia* 2009;11(1):57-65.
- [86] Hendrickson R.C., Cicinnati V.R., Albers A., et al. Identification of a 17beta-hydroxysteroid dehydrogenase type 12 pseudogene as the source of a highly restricted BALB/c Meth A tumor rejection peptide. *Cancer Immunol Immunother* 2010;59(1):113-24.
- [87] Karreth F.A., Reschke M., Ruocco A., et al. The BRAF pseudogene functions as a competitive endogenous RNA and induces lymphoma in vivo. *Cell* 2015;161(2):319-32.
- [88] De Martino M., De Biase D., Forzati F., et al. HMGA1-pseudogene7 transgenic mice develop B cell lymphomas. *Sci Rep* 2020;10(1):7057.
- [89] Staeger M.S., Muller K., Kewitz S., et al. Expression of dual-specificity phosphatase 5 pseudogene 1 (DUSP5P1) in tumor cells. *PLoS One* 2014;9(2):e89577.
- [90] Cao J., Li L., Chen C., et al. RNA interference-mediated silencing of NANOG leads to reduced proliferation and self-renewal, cell cycle arrest and apoptosis in T-cell acute lymphoblastic leukemia cells via the p53 signaling pathway. *Leuk Res* 2013;37(9):1170-7.
- [91] Zhou L.Y., Zhai L.L., Yin J.Y., et al. Pseudogene BMI1P1 expression as a novel predictor for acute myeloid leukemia development and prognosis. *Oncotarget* 2016;7(30):47376-86.
- [92] Zhou L.Y., Yin J.Y., Tang Q., et al. High expression of dual-specificity phosphatase 5 pseudogene 1 (DUSP5P1) is associated with poor prognosis in acute myeloid leukemia. *Int J Clin Exp Pathol* 2015;8(12):16073-80.
- [93] Yi J., Zhou L.Y., Yi Y.Y., et al. Low Expression of Pseudogene POU5F1B Affects Diagnosis and Prognosis in Acute Myeloid Leukemia (AML). *Med Sci Monit* 2019;25:4952-9.
- [94] Lettnin A.P., Wagner E.F., Carrett-Dias M., et al. Silencing the OCT4-PG1 pseudogene reduces OCT-4 protein levels and changes characteristics of the multidrug resistance phenotype in chronic myeloid leukemia. *Mol Biol Rep* 2019;46(2):1873-84.
- [95] McCarthy A.J., Chetty R. Smad4/DPC4. *J Clin Pathol* 2018;71(8):661-4.
- [96] Zhao M., Mishra L., Deng C.X. The role of TGF-beta/SMAD4 signaling in cancer. *Int J Biol Sci* 2018;14(2):111-23.
- [97] Millson A., Lewis T., Pesaran T., et al. Processed Pseudogene Confounding Deletion/Duplication Assays for SMAD4. *J Mol Diagn* 2015;17(5):576-82.
- [98] Watson C.M., Camm N., Crinnion L.A., et al. Characterization and Genomic Localization of a SMAD4 Processed Pseudogene. *J Mol Diagn* 2017;19(6):933-40.
- [99] Wytyczne GATK, <https://gatk.broadinstitute.org/hc/en-us/sections/360007226651-Best-Practices-Workflows>.
- [100] Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114-20.
- [101] Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754-60.
- [102] Talevich E., Shain A.H., Botton T., Bastian B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* 2016;12(4):e1004873.
- [103] Karczewski K.J., Francioli L.C., Tiao G., et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581(7809):434-43.
- [104] Kircher M., Witten D.M., Jain P., et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46(3):310-5.

- [105] Adzhubei I., Jordan D.M., Sunyaev S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* 2013;Chapter 7:Unit7 20.
- [106] Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4(7):1073-81.
- [107] Shihab H.A., Gough J., Cooper D.N., et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 2013;34(1):57-65.
- [108] Schwarz J.M., Cooper D.N., Schuelke M., Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11(4):361-2.
- [109] Kopanos C., Tsiolkas V., Kouris A., et al. VarSome: the human genomic variant search engine. *Bioinformatics* 2019;35(11):1978-80.
- [110] Tate J.G., Bamford S., Jubb H.C., et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019;47(D1):D941-D7.
- [111] Landrum M.J., Lee J.M., Benson M., et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46(D1):D1062-D7.
- [112] Jackman S.D., Vandervalk B.P., Mohamadi H., et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* 2017;27(5):768-77.
- [113] Kusmirek W., Nowak R. De novo assembly of bacterial genomes with repetitive DNA regions by dnaasm application. *BMC Bioinformatics* 2018;19(1):273.
- [114] Kim T., Tyndel M.S., Kim H.J., et al. Spectrum of somatic mutation dynamics in chronic myeloid leukemia following tyrosine kinase inhibitor therapy. *Blood* 2017;129(1):38-47.
- [115] He J., Abdel-Wahab O., Nahas M.K., et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood* 2016;127(24):3004-14.
- [116] Cancer Genome Atlas Research N., Ley T.J., Miller C., et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013;368(22):2059-74.
- [117] Delic S., Rose D., Kern W., et al. Application of an NGS-based 28-gene panel in myeloproliferative neoplasms reveals distinct mutation patterns in essential thrombocythaemia, primary myelofibrosis and polycythaemia vera. *Br J Haematol* 2016;175(3):419-26.
- [118] Nagata Y., Maciejewski J.P. The functional mechanisms of mutations in myelodysplastic syndrome. *Leukemia* 2019;33(12):2779-94.
- [119] Yu J., Li Y., Li T., et al. Gene mutational analysis by NGS and its clinical significance in patients with myelodysplastic syndrome and acute myeloid leukemia. *Exp Hematol Oncol* 2020;9:2.
- [120] Curik N., Polivkova V., Burda P., et al. Somatic Mutations in Oncogenes Are in Chronic Myeloid Leukemia Acquired De Novo via Deregulated Base-Excision Repair and Alternative Non-Homologous End Joining. *Front Oncol* 2021;11:744373.
- [121] Patel A.B., O'Hare T., Deininger M.W. Mechanisms of Resistance to ABL Kinase Inhibition in Chronic Myeloid Leukemia and the Development of Next Generation ABL Kinase Inhibitors. *Hematol Oncol Clin North Am* 2017;31(4):589-612.
- [122] Soverini S., Bavaro L., De Benedittis C., et al. Prospective assessment of NGS-detectable mutations in CML patients with nonoptimal response: the NEXT-in-CML study. *Blood* 2020;135(8):534-41.
- [123] Kizilers A., Crisa E., Lea N., et al. Effect of low-level BCR-ABL1 kinase domain mutations identified by next-generation sequencing in patients with chronic myeloid leukaemia: a population-based study. *Lancet Haematol* 2019;6(5):e276-e84.
- [124] Abdel-Wahab O., Adli M., LaFave L.M., et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell* 2012;22(2):180-93.

- [125] Alvarez Argote J., Dasanu C.A. ASXL1 mutations in myeloid neoplasms: pathogenetic considerations, impact on clinical outcomes and survival. *Curr Med Res Opin* 2018;34(5):757-63.
- [126] Kakosaiou K., Panitsas F., Daraki A., et al. ASXL1 mutations in AML are associated with specific clinical and cytogenetic characteristics. *Leuk Lymphoma* 2018;59(10):2439-46.
- [127] Stein B.L., Williams D.M., O'Keefe C., et al. Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: analysis of molecular genetics and clinical phenotypes. *Haematologica* 2011;96(10):1462-9.
- [128] Boulwood J., Perry J., Pellagatti A., et al. Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia* 2010;24(5):1062-5.
- [129] Bidikian A., Kantarjian H., Jabbour E., et al. Prognostic impact of ASXL1 mutations in chronic phase chronic myeloid leukemia. *Blood Cancer J* 2022;12(10):144.
- [130] Schonfeld L., Rinke J., Hinze A., et al. ASXL1 mutations predict inferior molecular response to nilotinib treatment in chronic myeloid leukemia. *Leukemia* 2022;36(9):2242-9.
- [131] Ernst T., Busch M., Rinke J., et al. Frequent ASXL1 mutations in children and young adults with chronic myeloid leukemia. *Leukemia* 2018;32(9):2046-9.
- [132] Chou W.C., Huang H.H., Hou H.A., et al. Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations. *Blood* 2010;116(20):4086-94.
- [133] Xue M. Z.Z., Wang Q., Wen L., Xu Y., Xie J., Wang Q., Ruan C., Wu D., Chen S. Mutational Profiles during the Progression of Chronic Myeloid Leukemia. *Blood* 2021;138:3596.
- [134] Sellars M., Kastner P., Chan S. Ikaros in B cell development and function. *World J Biol Chem* 2011;2(6):132-9.
- [135] Marke R., van Leeuwen F.N., Scheijen B. The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia. *Haematologica* 2018;103(4):565-74.
- [136] Winandy S., Wu P., Georgopoulos K. A dominant mutation in the Ikaros gene leads to rapid development of leukemia and lymphoma. *Cell* 1995;83(2):289-99.
- [137] Mullighan C.G., Su X., Zhang J., et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* 2009;360(5):470-80.
- [138] Yao Q.M., Liu K.Y., Gale R.P., et al. Prognostic impact of IKZF1 deletion in adults with common B-cell acute lymphoblastic leukemia. *BMC Cancer* 2016;16:269.
- [139] Mullighan C.G., Miller C.B., Radtke I., et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* 2008;453(7191):110-4.
- [140] Iacobucci I., Storlazzi C.T., Cilloni D., et al. Identification and molecular characterization of recurrent genomic deletions on 7p12 in the IKZF1 gene in a large cohort of BCR-ABL1-positive acute lymphoblastic leukemia patients: on behalf of Gruppo Italiano Malattie Ematologiche dell'Adulto Acute Leukemia Working Party (GIMEMA AL WP). *Blood* 2009;114(10):2159-67.
- [141] Klumb C.E., Barbosa T.D.C., Nestal de Moraes G., et al. IKZF1 deletion and co-occurrence with other aberrations in a child with chronic myeloid leukemia progressing to acute lymphoblastic leukemia. *Pediatr Blood Cancer* 2019;66(4):e27570.
- [142] Tang J.L., Hou H.A., Chen C.Y., et al. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood* 2009;114(26):5352-61.
- [143] Chen C.Y., Lin L.I., Tang J.L., et al. RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. *Br J Haematol* 2007;139(3):405-14.

- [144] Grossmann V., Kern W., Harbich S., et al. Prognostic relevance of RUNX1 mutations in T-cell acute lymphoblastic leukemia. *Haematologica* 2011;96(12):1874-7.
- [145] Venugopal S., DiNardo C.D., Loghavi S., et al. Differential prognostic impact of RUNX1 mutations according to frontline therapy in patients with acute myeloid leukemia. *Am J Hematol* 2022;97(12):1560-7.
- [146] Fournier E., Debord C., Soenen V., et al. Baseline dysmegakaryopoiesis in inherited thrombocytopenia/platelet disorder with predisposition to haematological malignancies. *Br J Haematol* 2020;189(4):e119-e22.
- [147] Mangan J.K., Speck N.A. RUNX1 mutations in clonal myeloid disorders: from conventional cytogenetics to next generation sequencing, a story 40 years in the making. *Crit Rev Oncog* 2011;16(1-2):77-91.
- [148] Zhao L.J., Wang Y.Y., Li G., et al. Functional features of RUNX1 mutants in acute transformation of chronic myeloid leukemia and their contribution to inducing murine full-blown leukemia. *Blood* 2012;119(12):2873-82.
- [149] De Braekeleer E., Douet-Guilbert N., Morel F., et al. RUNX1 translocations and fusion genes in malignant hemopathies. *Future Oncol* 2011;7(1):77-91.
- [150] Hou H.A., Kuo Y.Y., Liu C.Y., et al. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood* 2012;119(2):559-68.
- [151] Maifrede S., Le B.V., Nieborowska-Skorska M., et al. TET2 and DNMT3A Mutations Exert Divergent Effects on DNA Repair and Sensitivity of Leukemia Cells to PARP Inhibitors. *Cancer Res* 2021;81(19):5089-101.
- [152] Scheller M., Ludwig A.K., Gollner S., et al. Hotspot DNMT3A mutations in clonal hematopoiesis and acute myeloid leukemia sensitize cells to azacytidine via viral mimicry response. *Nat Cancer* 2021;2(5):527-44.
- [153] Adnan Awad S., Bruck O., Shanmuganathan N., et al. Epigenetic modifier gene mutations in chronic myeloid leukemia (CML) at diagnosis are associated with risk of relapse upon treatment discontinuation. *Blood Cancer J* 2022;12(4):69.
- [154] Liang D.C., Liu H.C., Yang C.P., et al. Cooperating gene mutations in childhood acute myeloid leukemia with special reference on mutations of ASXL1, TET2, IDH1, IDH2, and DNMT3A. *Blood* 2013;121(15):2988-95.
- [155] Peterson L.F., Mitrikeska E., Giannola D., et al. p53 stabilization induces apoptosis in chronic myeloid leukemia blast crisis cells. *Leukemia* 2011;25(5):761-9.
- [156] Menezes J., Salgado R.N., Acquadro F., et al. ASXL1, TP53 and IKZF3 mutations are present in the chronic phase and blast crisis of chronic myeloid leukemia. *Blood Cancer J* 2013;3(11):e157.
- [157] Wendel H.G., de Stanchina E., Cepero E., et al. Loss of p53 impedes the antileukemic response to BCR-ABL inhibition. *Proc Natl Acad Sci U S A* 2006;103(19):7444-9.
- [158] DeBoer R., Koval G., Mulkey F., et al. Clinical impact of ABL1 kinase domain mutations and IKZF1 deletion in adults under age 60 with Philadelphia chromosome-positive (Ph+) acute lymphoblastic leukemia (ALL): molecular analysis of CALGB (Alliance) 10001 and 9665. *Leuk Lymphoma* 2016;57(10):2298-306.
- [159] Braun M., Pastorczak A., Fendler W., et al. Biallelic loss of CDKN2A is associated with poor response to treatment in pediatric acute lymphoblastic leukemia. *Leuk Lymphoma* 2017;58(5):1162-71.
- [160] Patel S., Mason C.C., Glenn M.J., et al. Genomic analysis of adult B-ALL identifies potential markers of shorter survival. *Leuk Res* 2017;56:44-51.
- [161] Chan R.J., Feng G.S. PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* 2007;109(3):862-7.
- [162] Stasik S., Eckardt J.N., Kramer M., et al. Impact of PTPN11 mutations on clinical outcome analyzed in 1529 patients with acute myeloid leukemia. *Blood Adv* 2021;5(17):3279-89.

- [163] Hehlmann R., Voskanyan A., Lauseker M., et al. High-risk additional chromosomal abnormalities at low blast counts herald death by CML. *Leukemia* 2020;34(8):2074-86.
- [164] Adnan-Awad S., Kankainen M., Mustjoki S. Mutational landscape of chronic myeloid leukemia: more than a single oncogene leukemia. *Leuk Lymphoma* 2021;62(9):2064-78.
- [165] Nteliopoulos G., Bazeos A., Claudiani S., et al. Somatic variants in epigenetic modifiers can predict failure of response to imatinib but not to second-generation tyrosine kinase inhibitors. *Haematologica* 2019;104(12):2400-9.
- [166] Mologni L., Piazza R., Khandelwal P., Pirola A., Gambacorti-Passerini C. Somatic mutations identified at diagnosis by exome sequencing can predict response to imatinib in chronic phase chronic myeloid leukemia (CML) patients. *Am J Hematol* 2017;92(10):E623-E5.
- [167] Morita K., Wang F., Jahn K., et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* 2020;11(1):5327.
- [168] Danylesko I., Jacoby E., Yerushalmi R., et al. Remission of acute myeloid leukemia with t(8;21) following CD19 CAR T-cells. *Leukemia* 2020;34(7):1939-42.
- [169] Mill C.P., Fiskus W., DiNardo C.D., et al. RUNX1-targeted therapy for AML expressing somatic or germline mutation in RUNX1. *Blood* 2019;134(1):59-73.
- [170] Yang H., Kurtenbach S., Guo Y., et al. Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood* 2018;131(3):328-41.
- [171] Churchman M.L., Low J., Qu C., et al. Efficacy of Retinoids in IKZF1-Mutated BCR-ABL1 Acute Lymphoblastic Leukemia. *Cancer Cell* 2015;28(3):343-56.
- [172] Dogra R., Bhatia R., Shankar R., Bansal P., Rawal R.K. Enasidenib: First Mutant IDH2 Inhibitor for the Treatment of Refractory and Relapsed Acute Myeloid Leukemia. *Anticancer Agents Med Chem* 2018;18(14):1936-51.
- [173] DiNardo C.D., Stein E.M., de Botton S., et al. Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. *N Engl J Med* 2018;378(25):2386-98.
- [174] Granowicz E.M., Jonas B.A. Targeting TP53-Mutated Acute Myeloid Leukemia: Research and Clinical Developments. *Onco Targets Ther* 2022;15:423-36.
- [175] Wong K.K., Lawrie C.H., Green T.M. Oncogenic Roles and Inhibitors of DNMT1, DNMT3A, and DNMT3B in Acute Myeloid Leukaemia. *Biomark Insights* 2019;14:1177271919846454.
- [176] Rinke J., Chase A., Cross N.C.P., Hochhaus A., Ernst T. EZH2 in Myeloid Malignancies. *Cells* 2020;9(7).
- [177] Sorrentino C., Cuneo A., Roti G. Therapeutic Targeting of Notch Signaling Pathway in Hematological Malignancies. *Mediterr J Hematol Infect Dis* 2019;11(1):e2019037.
- [178] Kreuger I.Z.M., Sliker R.C., van Groningen T., van Doorn R. Therapeutic Strategies for Targeting CDKN2A Loss in Melanoma. *J Invest Dermatol* 2023;143(1):18-25 e1.
- [179] Miller R.E., El-Shakankery K.H., Lee J.Y. PARP inhibitors in ovarian cancer: overcoming resistance with combination strategies. *J Gynecol Oncol* 2022;33(3):e44.
- [180] Oliveira J.L., Greipp P.T., Rangan A., Jatoi A., Nguyen P.L. Myeloid malignancies in cancer patients treated with poly(ADP-ribose) polymerase (PARP) inhibitors: a case series. *Blood Cancer J* 2022;12(1):11.
- [181] Yao H., Wu C., Chen Y., et al. Spectrum of gene mutations identified by targeted next-generation sequencing in Chinese leukemia patients. *Mol Genet Genomic Med* 2020;8(9):e1369.
- [182] Cerami E., Gao J., Dogrusoz U., et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2(5):401-4.
- [183] Salem A., Loghavi S., Tang G., et al. Myeloid neoplasms with concurrent BCR-ABL1 and CBFB rearrangements: A series of 10 cases of a clinically aggressive neoplasm. *Am J Hematol* 2017;92(6):520-8.

- [184] Somwar R., Hofmann N.E., Smith B., et al. NTRK kinase domain mutations in cancer variably impact sensitivity to type I and type II inhibitors. *Commun Biol* 2020;3(1):776.
- [185] Brown L.C., Tucker M.D., Sedhom R., et al. LRP1B mutations are associated with favorable outcomes to immune checkpoint inhibitors across multiple cancer types. *J Immunother Cancer* 2021;9(3).
- [186] Sammalkorpi H., Alhopuro P., Niittymaki I., et al. Mutation analysis of MYH11 in acute myeloid leukemia. *Leuk Lymphoma* 2008;49(9):1829-31.
- [187] Wang R.J., Wu P., Cai G.X., et al. Down-regulated MYH11 expression correlates with poor prognosis in stage II and III colorectal cancer. *Asian Pac J Cancer Prev* 2014;15(17):7223-8.
- [188] Wallace K.N., Dolan A.C., Seiler C., et al. Mutation of smooth muscle myosin causes epithelial invasion and cystic expansion of the zebrafish intestine. *Dev Cell* 2005;8(5):717-26.
- [189] Zhang W.W., Cortes J.E., Yao H., et al. Predictors of primary imatinib resistance in chronic myelogenous leukemia are distinct from those in secondary imatinib resistance. *J Clin Oncol* 2009;27(22):3642-9.
- [190] Radich J.P., Dai H., Mao M., et al. Gene expression changes associated with progression and response in chronic myeloid leukemia. *Proc Natl Acad Sci U S A* 2006;103(8):2794-9.
- [191] Neves H., Ramos C., da Silva M.G., Parreira A., Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. *Blood* 1999;93(4):1197-207.
- [192] Saglio G., Storlazzi C.T., Giugliano E., et al. A 76-kb duplison maps close to the BCR gene on chromosome 22 and the ABL gene on chromosome 9: possible involvement in the genesis of the Philadelphia chromosome translocation. *Proc Natl Acad Sci U S A* 2002;99(15):9882-7.
- [193] Ross D.M., O'Hely M., Bartley P.A., et al. Distribution of genomic breakpoints in chronic myeloid leukemia: analysis of 308 patients. *Leukemia* 2013;27(10):2105-7.
- [194] Albano F., Anelli L., Zagaria A., et al. Genomic segmental duplications on the basis of the t(9;22) rearrangement in chronic myeloid leukemia. *Oncogene* 2010;29(17):2509-16.

# OPINIE KOMISJI BIOETYCZNYCH



## Komisja Bioetyczna przy Warszawskim Uniwersytecie Medycznym

Tel.: 022/ 57 - 20 -303  
Fax: 022/ 57 - 20 -165

ul. Żwirki i Wigury nr 61  
02-091 Warszawa

e-mail: komisja.bioetyczna@wum.edu.pl  
www.komisja-bioetyczna.wum.edu.pl

KB/...../2016

Komisja Bioetyczna przy Warszawskim Uniwersytecie Medycznym  
w dniu 08 listopada 2016 r. po zapoznaniu się z wnioskiem:

**Dr Tomasz Stokłosa**  
Zakład Immunologii WUM,  
ul. Banacha 1a, 02-097 Warszawa

**dotyczący:** wyrażenia opinii w sprawie badania pt. „Zbadanie podłoża genetycznego oraz mechanizmów progresji i niestabilności genomowej przewlekłej białaczki szpikowej BCR/ABL- dodatniej oraz wybranych innych chorób mieloproliferacyjnych BCR-ABL1-ujemnych.”

wyraża następującą  
opinię

- stwierdza, że jest ono dopuszczalne i zgodne z zasadami naukowo-etycznymi\*.
- stwierdza, że jest ono niedopuszczalne i niezgodne z zasadami naukowo-etycznymi.\*

Uwagi Komisji – *verte*

Komisja działa na podstawie art.29 ustawy z dnia 5.12.1996r. o zawodzie lekarza /Dz.U.nr 28/97 poz.152 wraz z późn.zm./, zarządzenia MZiOS z dn.11.05.1999r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych /Dz.U.nr 47 poz.480/, Ustawy prawo farmaceutyczne z dnia 6 września 2001r. (Dz.U.Nr 126, poz. 1381 z późn. zm.) oraz Zarządzenie nr 56/2007 z dnia 15 października 2007r. w sprawie działania Komisji Bioetycznej przy Warszawskim Uniwersytecie Medycznym /Regulamin Komisji Bioetycznej przy Warszawskim Uniwersytecie Medycznym/.

Komisja działa zgodnie z zasadami GCP .

W załączeniu: skład komisji oraz lista obecności

Przewodniczący Komisji Bioetycznej

Prof. dr hab. n. med. Zbigniew Wierzbicki

\*niepotrzebne skreślić



Komisja wyraża pozytywną opinię w sprawie przeprowadzenia wnioskowanych badań- na warunkach określonych we wniosku oraz dodatkowo zastrzegając:

1/ obowiązek przedstawienia Komisji:

- wszystkich zmian w protokole mających wpływ na przebieg oraz ocenę badania,
- wszystkich przypadków zdarzeń niepożądanych,
- zawiadomienia o przyczynach przedwczesnego zakończenia badania,
- sprawozdania w toku przeprowadzonych badań-za sześć miesięcy,
- raportu końcowego.

strona podpisowa do uchwały Komisji Bioetycznej przy Warszawskim  
Uniwersytecie Medycznym nr KB/.....<sup>227</sup>..... z dnia 08 listopada 2016r.

- |   |   |
|---|---|
| 1. Prof. dr hab. med. Zbigniew WIERZBICKI         | .....<br><i>Zbigniew Wierzbicki</i>         |
| 2. Prof. dr hab. Barbara GAJKOWSKA                | .....<br><i>Barbara Gajkowska</i>           |
| 3. Prof. dr hab. med. Jadwiga KOMENDER            | .....<br><i>Jadwiga Komender</i>            |
| 4. Prof. dr hab. med. Bożenna WOCIAL              | .....<br><i>Bożenna Wocial</i>              |
| 5. Prof. nadzw. dr hab. med. Andrzej KAŃSKI       | .....<br><i>Andrzej Kański</i>              |
| 6. Prof. dr hab. med. Jadwiga DWILEWICZ-TROJACZEK | .....<br><i>Jadwiga Dwilewicz-Trojaczek</i> |
| 7. Prof. dr hab. med. Krzysztof J. FILIPIAK       | .....<br><i>Krzysztof J. Filipiak</i>       |
| 8. Dr n. med. Zygmunt JAMROZIK                    | .....<br><i>Zygmunt Jamrozik</i>            |
| 9. Dr hab. n. med. Mariusz JASIK                  | .....<br><i>Mariusz Jasiak</i>              |
| 10. Dr hab. n. med. Andrea HORVATH-STOLARCZYK     | .....<br><i>Andrea Horvath-Stolarczyk</i>   |
| 11. Dr Agnieszka PIECHAL                          | .....<br><i>Agnieszka Piechal</i>           |
| 12. Mec. Ryszard PŁACZKOWSKI                      | .....<br><i>Ryszard Placzkowski</i>         |
| 13. Prof. dr hab. Joanna GÓRNICKA-KALINOWSKA      | .....<br><i>Joanna Górnicka-Kalinowska</i>  |
| 14. Alicja JAWORSKA                               | .....<br><i>Alicja Jaworska</i>             |
| 15. Ksiądz Władysław GRĘDOWSKI                    | .....<br><i>Władysław Grędowski</i>         |

**KOMISJA BIOETYCZNA**

przy  
Instytucie Hematologii i Transfuzjologii  
02-776 Warszawa, ul. I. Gandhi 14  
tel. 022 34 96 304, fax 022 34 96 305

**Komisja Bioetyczna****przy Instytucie Hematologii i Transfuzjologii**

ul. I. Gandhi 14, 02 - 776 Warszawa

tel: (022) 349 61 00 wew. 239

**Opinia Nr. 31/2008r**

z dnia 12.11.2008r.

Na podstawie art. 29 ustawy z dnia 5 grudnia o zawodzie lekarza (Dz.U. z 1997r. Nr 28 poz. 152 z późn. zm.) oraz rozporządzenia Ministra Zdrowia i Opieki Społecznej z dnia 11 maja 1999r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych (Dz.U. Nr 47, poz. 480 oraz Dz.U. z 2007 r. Nr 6, poz.46) Komisja Bioetyczna przy Instytucie Hematologii i Transfuzjologii w Warszawie na posiedzeniu w dniu 12 listopada 2008 r. przeanalizowała wniosek złożony przez Panią dr n. med. Ilonę Seferyską główną wykonawczynią projektu p.t.: „Zbadanie molekularnych mechanizmów lekooporności oraz próby jej przełamania w przewlekłej białaczce szpikowej BCR/ABL-dodatniej.”, który realizowany będzie w Zakładzie Immunologii i Zakładzie Histologii Centrum Biostruktury, Warszawskiego Uniwersytetu Medycznego w Warszawie, kierownikiem projektu jest dr n. med. Tomasz Stokłosa. W wyniku przeprowadzonej dyskusji oraz głosowania podjęła uchwałę o pozytywnym zaopiniowaniu badania.

Komisja zapoznała się z następującymi dokumentami, które były dołączone do wniosku:

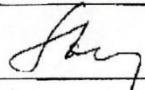

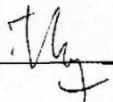


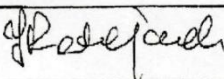
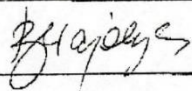
1. Opis projektu badania.
2. Informacja dla pacjenta o programie badawczym.
3. Karta zgody na pobranie krwi lub szpiku do badania.

Komisja działa zgodnie z zasadami Prawidłowego Prowadzenia Badań Klinicznych (GCP) opracowanymi w oparciu o Deklarację Helsińską.

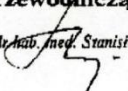
Od niniejszej opinii przysługuje odwołanie do Odwoławczej Komisji Bioetycznej przy Ministrze Zdrowia za pośrednictwem Komisji Bioetycznej przy Instytucie Hematologii i Transfuzjologii w Warszawie.

**Przewodniczący**  
Prof. dr hab. med. Stanisław Maj

**Lista obecności członków Komisji Bioetycznej przy Instytucie Hematologii  
i Transfuzjologii na posiedzeniu w dn. 12.11.2008 r.**

| Lp. | Imię i nazwisko                      | podpis  |
|-----|--------------------------------------|---|
| 1.  | prof. dr hab. med. Stanisław Maj     |   |
| 2   | prof. dr hab. med. Bożena Mariańska  |    |
| 3   | prof. dr hab. med. Jerzy Meissner    |   |
| 4   | prof. dr hab. Wincenty Tryzno        |   |
| 5   | doc. dr hab. med. Magdalena Łętowska |    |
| 6   | dr med. Bernadeta Ceglarek           |   |
| 7   | dr med. Jerzy Ratajczak              |   |
| 8   | mgr Andrzej Kaliński                 |   |
| 9   | mgr farm. Bożena Wajdyk              |  |
| 10  | Mgr Małgorzata Piątek                |   |
| 11  | Ks. dr Artur Filipowicz              |   |
| 12  | Marzena Wójcik                       |   |

**KOMISJA BIOETYCZNA**  
przy  
Instytucie Hematologii i Transfuzjologii  
02-776 Warszawa, ul. Indrzej Gandhi 14  
tel. 022 34 96 334, fax 022 34 96 335

**Przewodniczący**  
  
Prof. dr hab. med. Stanisław Maj

### **Informacja dla chorego o protokole projektu badawczego**

#### **Zbadanie molekularnych mechanizmów lekooporności oraz próby jej przelamywania w przewlekłej białaczce szpikowej BCR/ABL-dodatniej.**

Szanowni Państwo,

Proponujemy Pani/Panu udział w badaniu naukowym, którego celem jest zbadanie molekularnych mechanizmów lekooporności oraz próby jej przelamywania w przewlekłej białaczce szpikowej z obecnością chromosomu Filadelfia. Badania mogą być przydatne w przyszłości w stworzeniu nowych, skuteczniejszych metod terapii. Badania przeprowadzane będą w Zakładzie Immunologii Centrum Biostruktury, Akademii Medycznej w Warszawie, ul. Żwirki i Wigury 61.

Udział w badaniu polegać będzie na oddaniu krwi lub szpiku w ilości od 3 do 10 ml. Pobranie krwi do badań odbywać się będzie przy okazji pobierania krwi lub szpiku do rutynowych badań diagnostycznych. Nadzór nad pobieraniem próbek zgodnym z zasadami GCP (Good Clinical Practice) oraz gromadzenie danych klinicznych o pacjentach będzie sprawowała dr med. Ilona Seferyńska - jedna z głównych wykonawczyń projektu.

Rezygnacja Pani/Pana z udziału w badaniu nie będzie miała żadnego wpływu na jakość oferowanej opieki medycznej w Instytucie Hematologii i Transfuzjologii.

### Karta zgody chorego na udział w badaniu

**Zbadanie molekularnych mechanizmów lekooporności oraz próby jej przelamywania w przewlekłej białaczce szpikowej BCR/ABL-dodatniej.**

Ja .....urodzony.....

Wyrażam zgodę na wzięcie udziału w wyżej wymienionym badaniu, tj. oddaniu krwi lub szpiku w ilości 3-10 ml dla celów badania.

Oświadczam, że przeczytałem/am, zrozumiałem i jestem w posiadaniu kopii ulotki informacyjnej **Informacja dla chorego o protokole projektu badawczego**, dotyczącej wymienionego badania.

Wyrażam zgodę na anonimowy (posługiwanie się inicjałami) dostęp do danych medycznych dotyczących mojej osoby dla osób bezpośrednio zaangażowanych w prowadzenie badania.

Podpis chorego ..... Data .....

Podpis lekarza ..... Data .....

Imię i nazwisko lekarza.....

**Komisja Bioetyczna**  
**przy Instytucie Hematologii i Transfuzjologii**  
ul. I. Gandhi 14, 02 – 776 Warszawa  
tel.: (022) 34 96 472

**Opinia Nr 27/2016**  
z dnia 12.09.2016 r.

Na podstawie art. 29 ustawy z dnia 5 grudnia 1996 r. o zawodzie lekarza (Dz.U. z 1997r. Nr 28 poz. 152 z późn. zm.) oraz rozporządzenia Ministra Zdrowia i Opieki Społecznej z dnia 11 maja 1999r. w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych (Dz.U. Nr 47, poz. 480 oraz Dz.U. z 2007 r. Nr 6, poz.46) Komisja Bioetyczna przy Instytucie Hematologii i Transfuzjologii w Warszawie na posiedzeniu w dn. 12.09.2016 r. przeanalizowała wniosek: **„Zbadanie podłoża genetycznego oraz mechanizmów progresji i niestabilności genomowej przewlekłej białaczki szpikowej BCR/ABL-dodatniej oraz innych chorób mieloproliferacyjnych”**.

Kierownik samodzielnej jednostki organizacyjnej: Prof. dr hab. med. Krzysztof Warzocha  
Wykonawcy: Dr n. med. Tomasz Stokłosa wraz z zespołem badawczym – Zakład Immunologii Centrum Biostruktury, Warszawski Uniwersytet Medyczny; Dr med. Ilona Seferyńska, Prof. Joanna Góra-Tybor – Klinika Hematologii Instytutu Hematologii i Transfuzjologii w Warszawie.

Finansowanie:

1. Grant Harmonia z Narodowego Centrum Nauki (NCN) Nr rej. 2014/14/M/NZ5/00441.
2. Grant Opus z Narodowego Centrum Nauki (NCN) Nr rej. 2015/19/B/NZ5/03501.

**W wyniku przeprowadzonej dyskusji oraz głosowania komisja podjęła uchwałę o pozytywnym zaopiniowaniu w/w badania naukowego.**

Komisja zapoznała się z następującymi dokumentami:

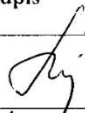

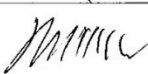
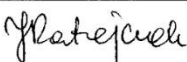



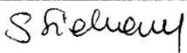
1. Wniosek do Komisji Bioetycznej z prośbą o wyrażenie zgody na prowadzenie badania.
2. Streszczenie projektu.
3. Załącznik Nr 1 „Deklaracja świadomej zgody na wykonanie badań genetycznych”.
4. Załącznik Nr 2 „Informacja dla chorego o protokole projektu badawczego”.

Komisja działa zgodnie z zasadami Prawidłowego Prowadzenia Badań Klinicznych (GCP) opracowanymi w oparciu o Deklarację Helsińską. Od niniejszej opinii przysługuje odwołanie do Odwoławczej Komisji Bioetycznej przy Ministrze Zdrowia za pośrednictwem Komisji Bioetycznej przy Instytucie Hematologii i Transfuzjologii w Warszawie.

**Przewodniczący**  
*Prof. dr hab. med. Stanisław Maj*



**Lista obecności członków Komisji Bioetycznej przy Instytucie Hematologii  
i Transfuzjologii na posiedzeniu w dniu 12.09.2016 r.**

| Lp. | Imię i nazwisko                          | podpis   |
|-----|--|--|
| 1.  | Prof. dr hab. med. Stanisław Maj         |   |
| 2   | Prof. dr hab. med. Bożena Mariańska      |    |
| 3   | Prof. dr hab. med. Alfred Jerzy Meissner |    |
| 4   | Prof. dr hab. med. Magdalena Łętowska    |  |
| 5   | Dr med. Bernadetta Ceglarek              |  |
| 6   | Dr med. Jerzy Ratajczak                  |    |
| 7   | Mgr Andrzej Kaliński                     |  |
| 8   | Mgr farm. Bożena Wajdyk                  |   |
| 9   | Mgr Małgorzata Piątek                    |  |
| 10  | Ks. Eugeniusz Klimiński                  |  |
| 11  | Piel. Sylwia Siarkowska                  |  |

**KOMISJA BIOETYCZNA**  
przy  
Instytucie Hematologii i Transfuzjologii  
02-776 Warszawa, ul. Indrzej Gandhi 14  
tel. 022 34 96 334, fax 022 34 96 335

**UCHWAŁA NR 65/WIM/2016**

**Komisji Bioetycznej Wojskowego Instytutu Medycznego**  
**z dnia 19.10.2016 r.**

w sprawie wyrażenia opinii o projekcie eksperymentu medycznego zgłoszonego przez:

**dr. n. med. Waldemara SAWICKIEGO**

zatrudnionego w Klinice Chorób Wewnętrznych i Hematologii Centralnego Szpitala Klinicznego Ministerstwa Obrony Narodowej Wojskowy Instytut Medyczny, ul. Szaserów 128, 04-141 Warszawa.

Na podstawie art. 29 ust. 2 ustawy z dn. 5 grudnia 1996 roku o zawodzie lekarza (Dz. U. z 1997 r. Nr 28, poz. 152 z późn. zm.) oraz § 6 ust. 5 rozporządzenia Ministra Zdrowia i Opieki Społecznej z 11 maja 1999 roku w sprawie szczegółowych zasad powoływania i finansowania oraz trybu działania komisji bioetycznych (Dz. U. Nr 47, poz. 480) uchwała się co, następuje:

§ 1

Komisja Bioetyczna Wojskowego Instytutu Medycznego – zwanej dalej „Komisją” w składzie:

|                            |  |
|----------------------------|--|
| <b>Przewodniczący:</b>     | <b>plk prof. dr hab. med. Dariusz Jurkiewicz</b> |
| <b>Wiceprzewodniczący:</b> | <b>mec. Andrzej Niewiadomski</b>                 |
| <b>Członkowie:</b>         | <b>prof. dr hab. med. Włodzimierz Baranowski</b> |
|                            | <b>prof. dr hab. med. Wojciech Kozłowski</b>     |
|                            | <b>prof. dr hab. med. Marek Maruszyński</b>      |
|                            | <b>prof. dr hab. med. Anna Jung</b>              |
|                            | <b>prof. dr hab. med. Jerzy Kruszewski</b>       |
|                            | <b>prof. dr hab. med. Adam Stępień</b>           |
|                            | <b>prof. dr hab. med. Piotr Rzepecki</b>         |
|                            | <b>dr hab. med. Andrzej Chciałowski</b>          |
|                            | <b>dr hab. med. Andrzej Cwetsch</b>              |
|                            | <b>prof. dr hab. med. Marek Rękas</b>            |
|                            | <b>plk mgr farm. Tadeusz Nierebiński</b>         |
| <b>Kapelan:</b>            | <b>ks. dr filoz. i teol. Wiesław Bożejewicz</b>  |
| <b>Pielęgniarka:</b>       | <b>dr n. o zdr. Olga Dembicka</b>                |

zapoznała się z projektem eksperymentu medycznego pt.: „Zbadanie podłoża genetycznego oraz mechanizmów progresji i niestabilności genomowej przewlekłej białaczki szpikowej BCR/ABL-dodatniej oraz innych nowotworów mieloproliferacyjnych”,

do którego dołączone zostały następujące dokumenty:

- Informacja o badaniu
- Informacja dla pacjenta,
- Deklaracja Świadomej zgody na wykonanie badań genetycznych.

#### § 2

Badanie finansowane jest ze środków własnych.

#### § 3

Komisja przeanalizowała projekt, o którym mowa § 1 w oparciu o Deklarację Helsińską z roku 1996 i zasady opracowane przez Unię Europejską pt.: „Good clinical practice for on medical products in the European Community. The rules governing medical products in the European Community”, Vol. III, 1990, zatwierdzone przez Komisję Etyczną w Polsce od 1 stycznia 1993 roku.

#### § 4

Każde poważniejsze powikłanie nie wchodzące w zakres znanych i mogących wystąpić skutków leczenia badanym lekiem, powinno być natychmiast zgłoszone Komisji Bioetycznej. Komisja zastrzega sobie zmianę badań lub ich wstrzymanie.

#### § 5

Każde pojedyncze badanie musi być poprzedzone wyjaśnieniem przez lekarza pacjentowi o zakresie badań klinicznych i wymaga podpisu przez pacjenta i lekarza protokołu świadomej zgody. Po zakończeniu badań Zespół prowadzący projekt zobowiązany jest dołączyć badania opublikowane kończące opracowanie.

§ 6

Komisja:

1) oceniając pod względem etycznym, celowości oraz wykonalność projektu eksperymentu medycznego dotyczącego: „Zbadanie podłoża genetycznego oraz mechanizmów progresji i niestabilności genomowej przewlekłej białaczki szpikowej BCR/ABL-dodatniej oraz innych nowotworów mieloproliferacyjnych”,

2) na podstawie opinii Recenzenta o projekcie i członków Komisji,  
postanawia wyrazić pozytywną opinię o wymienionym w § 1 projekcie badania.

§ 7

Uchwała wchodzi w życie z dniem podjęcia

Uchwałę podjęto w głosowaniu tajnym.

Za przyjęciem opinii głosowało <sup>12</sup>..... członków Komisji.

Za odrzuceniem opinii głosowało ..... członków Komisji.

W głosowaniu uczestniczyła ponad połowa członków Komisji pod przewodnictwem przewodniczącego z udziałem, co najmniej dwóch członków Komisji nie będących lekarzami.

Podpisy członków Komisji:

1. Przewodniczący

2. Wiceprzewodniczący

3. Członkowie

4.

5.

6.

7.

8.

9.

10.

11.

12.

13.

14.

15.

.....  
.....  
.....  
.....  
.....  
.....  
.....nie głosował.....  
.....  
.....  
.....  
.....  
.....  
.....

SKŁAD KOMISJI BIOETYCZNEJ PODEJMUJĄCEJ DECYZJĘ

WARSZAWA, dnia 19.10.2016r.

| Lp. | Imię i nazwisko                           | Zawód                           | Miejsce pracy                                     |
|-----|---|---------------------------------|---|
| 1.  | plk prof. dr hab. med. Dariusz Jurkiewicz | Lekarz/Otolaryngolog            | Wojewódzki Instytut Medyczny                      |
| 2.  | mec. Andrzej Niewiadomski                 | Radca Prawny                    | Departament Kadr MON                              |
| 3.  | prof. dr hab. med. Włodzimierz Baranowski | Lekarz/Ginekolog                | Wojewódzki Instytut Medyczny                      |
| 4.  | prof. dr hab. med. Wojciech Kozłowski     | Lekarz/Patomorfolog             | Wojewódzki Instytut Medyczny                      |
| 5.  | prof. dr hab. med. Marek Matuszyński      | Lekarz/Chirurg                  | Wojewódzki Instytut Medyczny                      |
| 6.  | prof. dr hab. med. Anna Jung              | Lekarz/Pediatra                 | Wojewódzki Instytut Medyczny                      |
| 7.  | prof. dr hab. med. Jerzy Kruszewski       | Lekarz/Algolog                  | Wojewódzki Instytut Medyczny                      |
| 8.  | prof. dr hab. med. Adam Szepeń            | Lekarz/Neurolog                 | Wojewódzki Instytut Medyczny                      |
| 9.  | prof. dr hab. med. Piotr Rzepecki         | Lekarz/Hematolog                | Wojewódzki Instytut Medyczny                      |
| 10. | plk dr hab. med. Andrzej Chatałowski      | Lekarz/Pulmonolog-<br>Alergolog | Wojewódzki Instytut Medyczny                      |
| 11. | dr hab. med. Andrzej Cwetsch              | Lekarz/Kardiolog                | Wojewódzki Instytut Medyczny                      |
| 12. | prof. dr hab. med. Marek Rekas            | Lekarz/Okulista                 | Wojewódzki Instytut Medyczny                      |
| 13. | mgr Tadeusz Nierebiński                   | Farmaceuta                      | Inspektorat Wojewódzkiej Szpitali Zdrowia         |
| 14. | ks. dr filozji teol. Wiesław Bożejewicz   | Kapelan                         | Ordynariat Polowy WP                              |
| 15. | dr n. o zdr. Olga Dembicka                | Pielęgniarka                    | Wojewódzki Szpital Chirurgii Urazowej w Warszawie |