
Streszczenie

Zachorowalność i umieralność na choroby nowotworowe dotyczy rocznie milionów ludzi na całym świecie. Diagnostyka tych chorób odbywa się za pomocą różnych technik, takich jak obrazowanie, biopsje tkanek czy badanie krwi. Metody te są istotne dla wczesnego rozpoznania nowotworu, co jest kluczowe zarówno dla poprawy skuteczności leczenia pacjentów onkologicznych, jak i poprawy wyników tego leczenia [1].

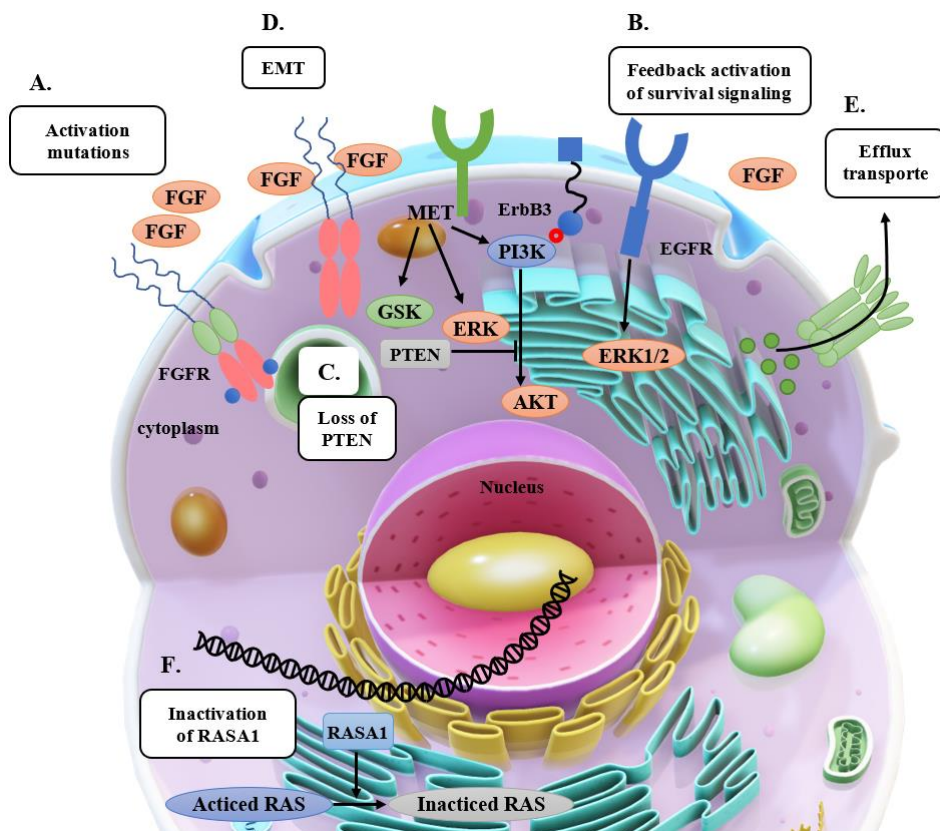
Jednym z elementów diagnozy, monitorowania, prognozy i leczenia spersonalizowanego nowotworu jest ocena różnych biomarkerów. Biomarkery nowotworowe to mierzalne substancje, które wskazują na obecność raka lub ryzyka rozwoju choroby nowotworowej i mogą być oznaczone w materiale tkankowym – w tym we krwi lub w innych płynach ustrojowych. Ponieważ biomarkery mogą także dostarczać informacji na temat biologicznego zachowania nowotworu, jego rozwoju i odpowiedzi na leczenie, wykorzystuje się je także do monitorowania przebiegu choroby oraz terapii [2].

Do tej pory mamy ponad 25 różnych markerów nowotworowych, które rutynowo są stosowane w oznaczeniach klinicznych, zarówno na potrzeby diagnostyki, jak i monitorowania leczenia [2, 3]. Jedne biomarkery są specyficzne dla określonego typu raka, inne są związane z dwoma lub więcej typami raka. Mimo, że potencjalnie każda cząsteczka biologiczna może zostać markerem, to jednak większość markerów to glikoproteiny lub białka [2].

W badaniach *in silico* nad poszukiwaniem biomarkerów, które opierają się o dane pochodzące z eksperymentów wysokoprzepustowych, wykorzystuje się różne metody do wytypowania potencjalnych biomarkerów. Mogą być stosowane tradycyjne techniki statystyczne, takie jak analiza wariancji (ANOVA) lub test t-Studenta, jak również bardziej zaawansowane techniki, takie jak UMAP (*ang. uniform manifold approximation and projection*) lub algorytmy uczenia maszynowego. Jednakże, ze względu na ograniczenia tych metod, często konieczne jest zastosowanie dodatkowych metod filtracji w celu zidentyfikowania biomarkerów, które spełnią wymagania kliniczne. Filtry te mogą uwzględniać kwestie takie jak: istotność biomarkera pod względem biologicznym, jego specyficzność i czułość oraz możliwość powtarzalności jego oznaczenia, aby zapewnić identyfikację najbardziej istotnych i wiarygodnych biomarkerów. Pomimo dekady intensywnych wysiłków oraz znacznych inwestycji i pracy, liczba biomarkerów, które zostały klinicznie zwalidowane i zatwierdzone przez odpowiednie agencje regulacyjne (np. *ang. Food and Drug Administration*, FDA), jest niezwykle mała [4].

Sygnalizacja poprzez receptory czynników wzrostu fibroblastów (*ang. Fibroblast Growth Factor Receptor*, FGFR) stanowi ważny mechanizm regulujący procesy proliferacji i różnicowania komórki. W wielu nowotworach mechanizm ten jest zaburzony, czego główną przyczyną są różnego typu nieprawidłowości genomowe. Przykładowo w raku pęcherza moczowego, żołądka, piersi i płuc występuje amplifikacja FGFR1, podczas gdy w raku wątroby, macicy, płuc i żołądka może wystąpić zarówno amplifikacja, jak i mutacje oraz fuzje FGFR2 z innymi genami, a mutacje i fuzje FGFR3 często występują w raku pęcherza moczowego i płuc. Dlatego hamowanie sygnalizacji FGFR jest przedmiotem badań i prób klinicznych w ramach rozwoju nowych terapii celowanych.

Zastosowanie inhibitorów FGFR jako strategii przeciwnowotworowej przy użyciu małowczątkowych związków jest bardzo obiecujące, jednak poważnym wyzwaniem staje się rozwój oporności na te leki. W literaturze opisano kilka mechanizmów nabytej oporności (Rycina 1) [5].



Rycina 1. Mechanizmy oporności na inhibitory FGFR: (A) mutacje regionu kodującego bramkę (*ang. gatekeeper*) w domenie kinazy FGFR, (B) aktywacja alternatywnych szlaków sygnałowych, takich jak EGFR, ERBB3 lub MET, (C) utrata PTEN prowadząca do zwiększonej aktywacji PI3K-AKT, (D) przejście epitelia no-mezenchymalne (*ang. Epithelial-Mesenchymal Transition, EMT*) może prowadzić do oporności na inhibitory FGFR, (E) regulacja wypompowywania leku przez transportery ABCG2, (F) dezaktywacja RAS przez RASA1. Oporność na inhibitory FGFR może pojawić się w wyniku dezaktywacji RASA1 [6].

Projekt doktorski był realizowany częściowo w ramach projektu zatytułowanego „Opracowanie nowoczesnych biomarkerów oraz rozwój innowacyjnego inhibitora kinaz FGFR stosowanego w terapii nowotworów”, o akronimie CELONKO, sfinansowanego przez Narodowe Centrum Badań i Rozwoju (NCBR), w ramach programu STRATEGMED II. Inicjatorem i zarówno liderem projektu była polska firma farmaceutyczna Celon Pharma S.A., wynalazca nowego inhibitora FGFR (CPL304110) [7]. Projekt badawczy był realizowany we współpracy naukowo-biznesowej z Instytutem Gruźlicy i Chorób Płuc w Warszawie, Wojskowym Instytutem Medycznym w Warszawie, Narodowym Instytutem Onkologii im. Marii Skłodowskiej-Curie w Warszawie i Gliwicach oraz Gdańskim Uniwersytetem Medycznym.

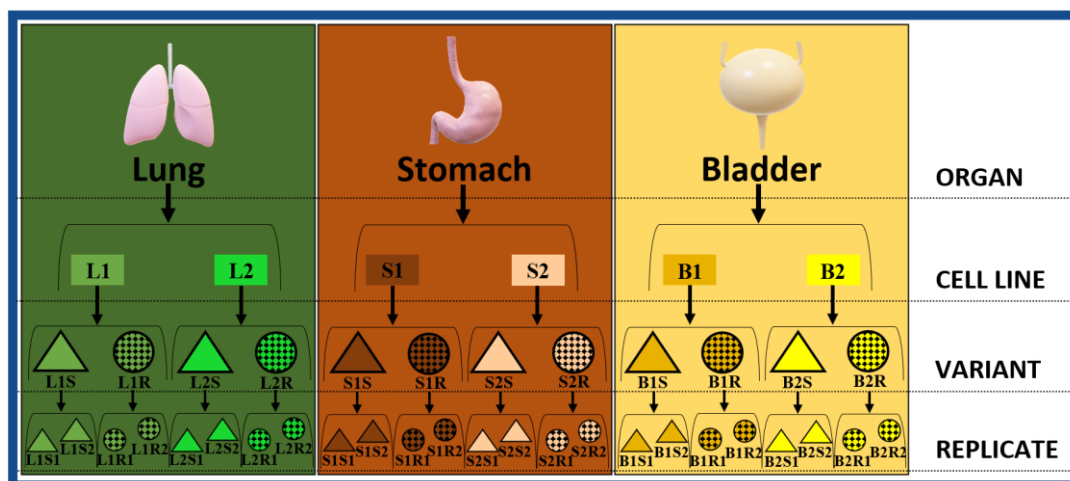
Lek CPL304110 przeznaczony jest do leczenia raka żołądka, pęcherza moczowego i płuc. W ramach projektu CELONKO opracowano test diagnostyczny do identyfikacji pacjentów ze znanymi aberracjami receptora FGFR, co umożliwi zastosowanie tego leku w sposób spersonalizowany u pacjentów, którzy mogą odnieść największą korzyść z proponowanej terapii.

Kolejnym celem projektu CELONKO było zidentyfikowanie potencjalnych nowych kandydatów na biomarkery, związanych z predykcją oporności na terapię opartą o inhibitory FGFR. Z powodu ograniczeń technicznych, było raczej niemożliwe poszukiwanie przyczyn wrażliwości guza na inhibitor FGFR. Dlatego skupiliśmy się na poszukiwaniu potencjalnych mechanizmów oporności i biomarkerów, badając linie komórkowe z nabytą opornością na inhibitor FGFR.

Początkowo, za pomocą techniki Western blot próbowano wytypować potencjalnych kandydatów poprzez analizę szlaków sygnałowych związanych z receptorami FGFR. Wyniki nie były wystarczające do

wyboru biomarkera, więc zdecydowano się przeprowadzić eksperyment sekwencjonowania RNA (RNA-seq) i dalszą analizę danych w celu zidentyfikowania *in silico* kandydatów na predykcyjny biomarker związany z potencjalnym mechanizmem oporności na inhibitory FGFR.

Mając na uwadze wyzwania związane z heterogenicznością genetyczną nowotworów, konieczne jest uwzględnienie tej bioróżnorodności zarówno w badaniach onkologicznych, jak i w praktyce klinicznej [8]. Na potrzeby eksperymentu RNA-seq, opisanego w mojej rozprawie doktorskiej wybrano kilka typów nowotworów, a konkretnie raka płuc, żołądka i pęcherza moczowego (Rycina 2). Te trzy typy nowotworów zostały wybrane, ponieważ w nich aberracje FGFR są obserwowane najczęściej [9, 10]. Wybór rodzaju raka był również uzasadniony związaną z nimi wysoką zachorowalnością i śmiertelnością, co wymaga nowych rozwiązań terapeutycznych.



Rycina 2. Do eksperymentu RNA-seq wykorzystano 6 linii komórkowych (L1: NCI-H1581, L2: NCI-H1703, S1: SNU-16, S2: KATO III, B1: RT-112, and B2: UM-UC-14), z każdej po dwa warianty (wariant dziki i jednocześnie wrażliwy (S: L1S, L2S, S1S, S2S, B1S, B2S) na inhibitor FGFR i wariant oporny (R: L1R, L2R, S1R, S2R, B1R, B2R)). Przygotowano dwa powtórzenia biologiczne na każdy wariant, co daje łącznie 24 próbki eksperymentalne.

W celu uwzględnienia bioróżnorodności związanej z różnicą między różnymi guzami jednego typu nowotworu (*ang. inter-tumor diversity*), na potrzeby eksperymentu RNA-seq, wybrano po dwie różne linie komórkowe dla każdego typu raka (Tabela 1, Rycina 2). Wybór był oparty o obecność profilu molekularnego sprzyjającego wrażliwości na inhibitory FGFR, w szczególności związanego z obecnością amplifikacji jednego z genów FGFR1-4 (Tabela 1). Dodatkowo, wybrane zostały linie komórkowe o największej wrażliwości na testowany inhibitor, jak również te, dla których było możliwe otrzymanie wariantu linii odpornej na testowany inhibitor (Rycina 2). Dla uwzględnienia bioróżnorodności związanej z różnicą występującą w obrębie jednego pacjenta (*ang. intra-patient diversity*), przygotowano dwa powtórzenia biologiczne dla każdego wariantu linii komórkowej (Rycina 2).

Tabela 1. Linie komórkowe wykorzystane w eksperymencie RNA-seq.

The organ origin of the cell line	Cell line symbol	Disease	Symbol of cell line variant		Amplification	Studied inhibitor 304-110-01 (IC50 [μM])	Max concentra- tion 304-110-01 tolerated by de- rived cell lines [μM]
			sensitive	resistant			
Lung	NCI-H1581	Non-small cell lung cancer. Cell type: large cell	L1	L1R	FGFR 1	0.074	2.500
	NCI-H1703	Non-small cell lung cancer. Cell type: squamous cell	L2	L2R	FGFR 1	1.300	5
Stomach	SNU 16	Gastric adenocarcinoma	S1	S1R	FGFR 2	0.005	0.700

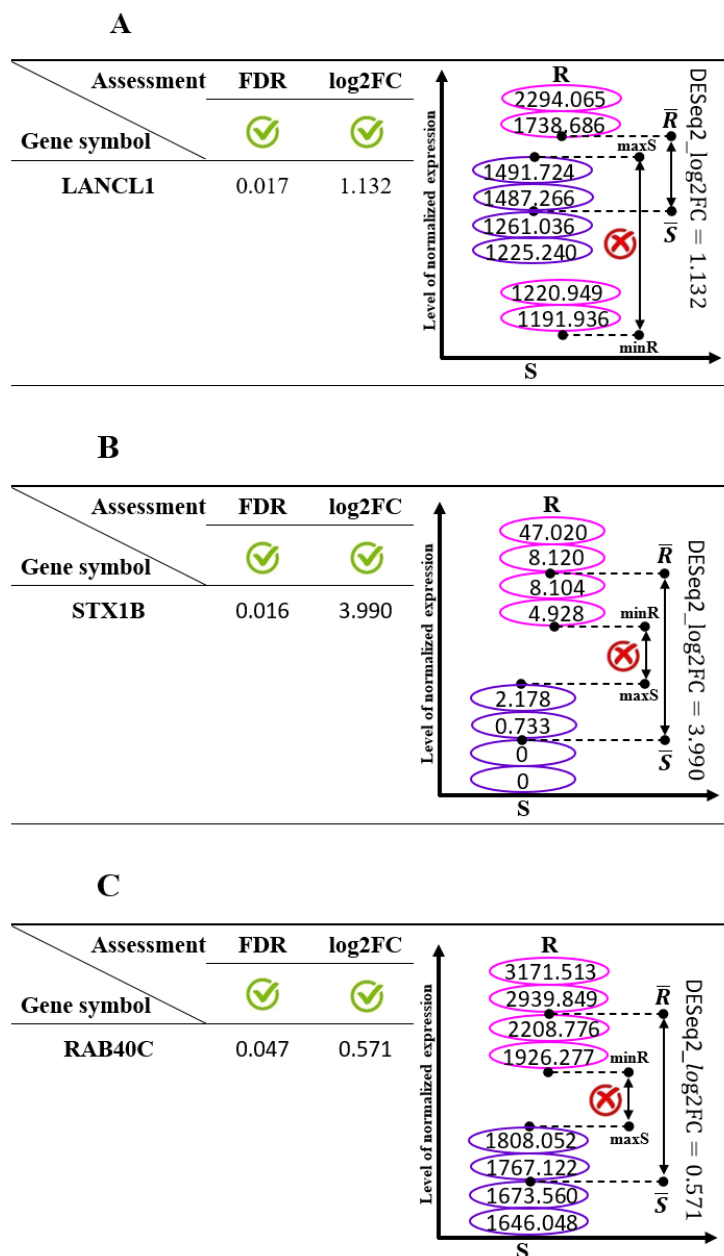
		Derived from metastatic site: ascites.					
	KATOIII	Gastric signet ring cell adenocarcinoma. Derived from metastatic site: pleural effusion	S2	S2R	FGFR 2	0.040	0.350
Bladder	RT112/84	Bladder carcinoma	B1	B1R	FGFR 3	0.239	1
	UM-UC 14	Renal pelvis carcinoma	B2	B2R	FGFR 3	0.031	0.100

W ramach tego eksperymentu RNA-seq nie udało mi się uwzględnić zmienności związanej z bioróżnorodnością występującą w obrębie guza (*ang. intra-tumor diversity*). Aby to zrobić, należałoby użyć próbek pobranych z różnych obszarów guza pochodzącego od jednego pacjenta. Ponieważ praca badawcza nad moją rozprawą doktorską poprzedzała fazy badań klinicznych projektu CELONKO, nie miałem bezpośredniego dostępu do materiału pochodzącego od pacjenta. Jednakże w trwającym badaniu klinicznym koordynowanym przez firmę Celon Pharma S.A. zbierany jest różny materiał biologiczny, co pozwoli nam kontynuować badania i uwzględnić ten aspekt.

Wraz z coraz większą dostępnością danych transkryptomicznych, zwłaszcza z eksperymentów o małej próbce, coraz ważniejsze staje się opracowanie solidnych i niezawodnych metod selekcji potencjalnych kandydatów na biomarkery do fazy ich walidacji. Celem moich badań było opracowanie nowej metody selekcji potencjalnych kandydatów na biomarkery, specjalnie dostosowanej do danych sekwencjonowania RNA (RNA-seq) uzyskanych z eksperymentów o małej próbce (Rycina 2).

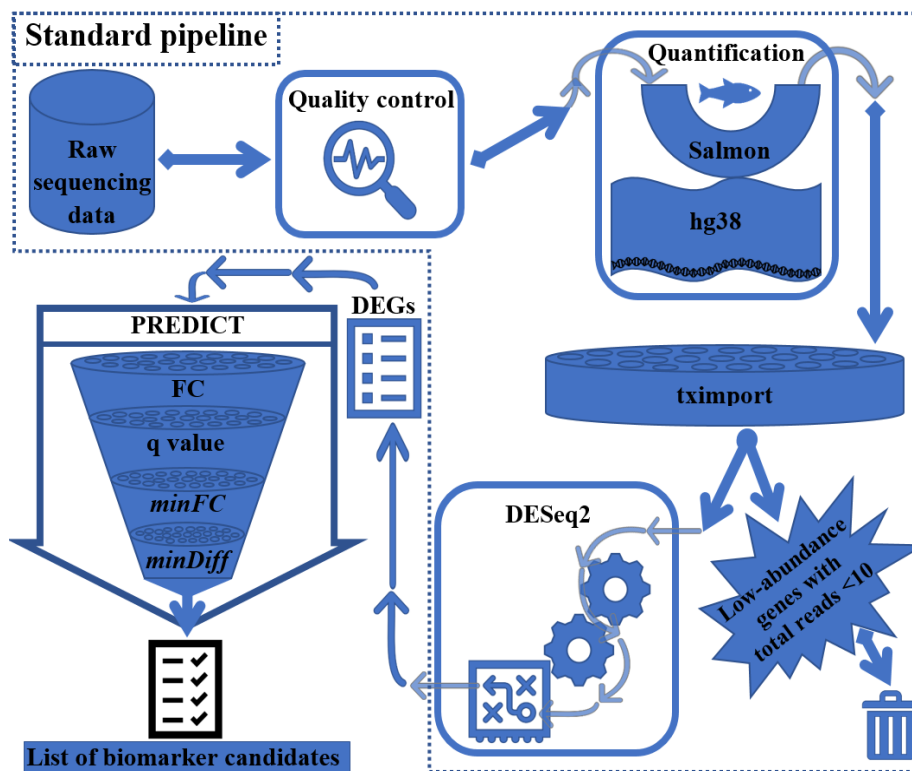
Korzystając z danych z eksperymentu RNA-seq, przeprowadzono kompleksową analizę profilu ekspresji genów w liniach komórkowych z trzech różnych typów nowotworów (płuca, żołądka i pęcherza moczowego), aby zidentyfikować potencjalne biomarkery predykcyjne związane z mechanizmami oporności na inhibitor FGFR.

Przy bliższej ocenie uzyskanych wyników analizy różnicowej przy użyciu narzędzia DESeq2 zauważyłem, że standardowa metoda typuje wiele DEGs (*ang. differentially expressed genes*), które nie spełniają wymagań stawianych dla biomarkerów. Na przykład, wiele wyników nie miało spójnego kierunku zmiany, ale były losowe, jak przedstawiono na przykładzie ujętym w Rycina 3.A, co wskazywałoby na potencjalnie niską czułość takiego kandydata i niską powtarzalność wyniku testowania takiego markera. Biomarker predykcyjny w momencie testowania u pacjenta z rakiem, który nie jest wrażliwy na terapię opartą o inhibitory FGFR, nie może jednocześnie wskazywać na obecność określonego mechanizmu oporności, a za chwilę przy powtórzeniu badania na jego brak. Inne wyniki, pomimo dużej wartości współczynnika krotności zmiany (*ang. fold change*, FC), miały zbyt małą różnicę pomiędzy wartościami znormalizowanych odczytów, by osiągnąć próg wykrywalności dla technik stosowanych w rutynowej praktyce diagnostycznej (Rycina 3.B). Ponadto, niektóre wyniki, pomimo wystarczająco dużej różnicy między odczytami, miały niewielki współczynnik miary minimalnej krotności zmiany (*ang. minimal fold change*, *minFC*), co mogłoby skutkować niewykryciem biologicznego efektu dla tak wytypowanego biomarkera [11, 12] (Rycina 3.C). Tacy potencjalni kandydaci na biomarker cechowałiby się bardzo niską czułością i swoistością, jeśli nie byłoby możliwe wykrycie różnicy w ich poziomie między zdrową tkanką, a tkanką nowotworową.



Rycina 3. Przykłady kandydatów na biomarker zidentyfikowanych za pomocą standardowej analizy danych RNA-seq, charakteryzujących się istotnym q wartościami ($FDR < 0,050$ i odpowiednim $\log_2FC > 0,500$), ale nie posiadających cech, jakie powinien mieć biomarker: (A) LANCL1 to przykład braku spójnego kierunku zmiany, (B) STX1B to przykład, który pomimo dużego współczynnika krotności zmiany (FC), posiada bardzo małą różnicę między skrajnymi wartościami wewnętrznymi zbiorów R i S ($\min R$ i $\max S$), (C) RAB40C to przykład, który pomimo, że charakteryzuje się odpowiednio dużą wartością miary minimalnej krotności zmiany ($\min FC$), to różnica między skrajnymi wartościami wewnętrznymi zbiorów R i S jest bardzo mała, co sprawia, że jest mało prawdopodobne, aby wykryć taką różnicę za pomocą metod typowo stosowanych w praktyce klinicznej. Oznaczenie ✓ wskazuje na pożądaną wartość, natomiast znak ✗ wskazuje na niepożądaną wartość.

Zestaw danych z sekwencjonowania RNA charakteryzował się małą liczebnością próbek, co jest typowym ograniczeniem wielu podobnych eksperymentów. Standardowe metody analizy nie radzą sobie dobrze z tym typem danych. Ponadto, brak odpowiednich filtrów sprawia, że wyniki mogą nie spełniać wymagań stawianych biomarkerom do zastosowań klinicznych. Dlatego, w ramach niniejszej pracy opracowano schemat obliczeniowy nazwany „Pipeline for Rapid Evaluation, and Discovery of Important biomarker Candidates” (PREDICT) (Rycina 4), który opiera się między innymi na sekwencyjnym stosowaniu progów $\log_2FC > 0,500$, wartość $q < 0,050$, $\log_2\min FC > 0,100$ i $\min Diff > 100$ do wyników uzyskanych standardową metodą analizy różnicowej.



Rycina 4. Schemat standardowego potoku analizy danych RNA-seq oraz schemat „Pipeline for Rapid Evaluation and Discovery of Important biomarker Candidates” (PREDICT).

Schemat obliczeniowy PREDICT obejmuje dwie miary, a mianowicie minimalną krotność zmiany (*ang. minimal fold change, minFC*) i minimalną różnicę (*ang. minimal Difference, minDiff*) (Rycina 5. A i B, odpowiednio).

minFC (Rycina 5.A) – niech $X = \{x_1, x_2, \dots, x_n\}$ będzie zbiorem pomiarów poziomu ekspresji danego genu dla próbek należących do jednej grupy, a $Y = \{y_1, y_2, \dots, y_m\}$ będzie zbiorem pomiarów poziomu ekspresji danego genu dla próbek należących do drugiej grupy. Wówczas definiujemy *minFC* jako:

$$\min FC = \begin{cases} \frac{\min X}{\max Y} & \text{if } \bar{X} > \bar{Y} \\ 1 & \text{if } \bar{X} = \bar{Y} \\ \frac{\min Y}{\max X} & \text{if } \bar{X} < \bar{Y} \end{cases}$$

gdzie *minX* i *minY* oznaczają najmniejszą wartość w zbiorze X, Y , odpowiednio, a *maxX* i *maxY* oznaczają największą wartość w zbiorze X, Y , odpowiednio. \bar{X} , i \bar{Y} oznaczają wartość średnią dla zbiorów X i Y , odpowiednio. $\min FC > 1$ (wartość $\log_2 \min FC > 0$) pokazuje, że przedziały wartości ekspresji dla porównywanych grup nie pokrywają się, a $\min FC \leq 1$ (wartość $\log_2 \min FC \leq 0$) pokazuje, że przedziały wartości ekspresji dla porównywanych grup pokrywają się.

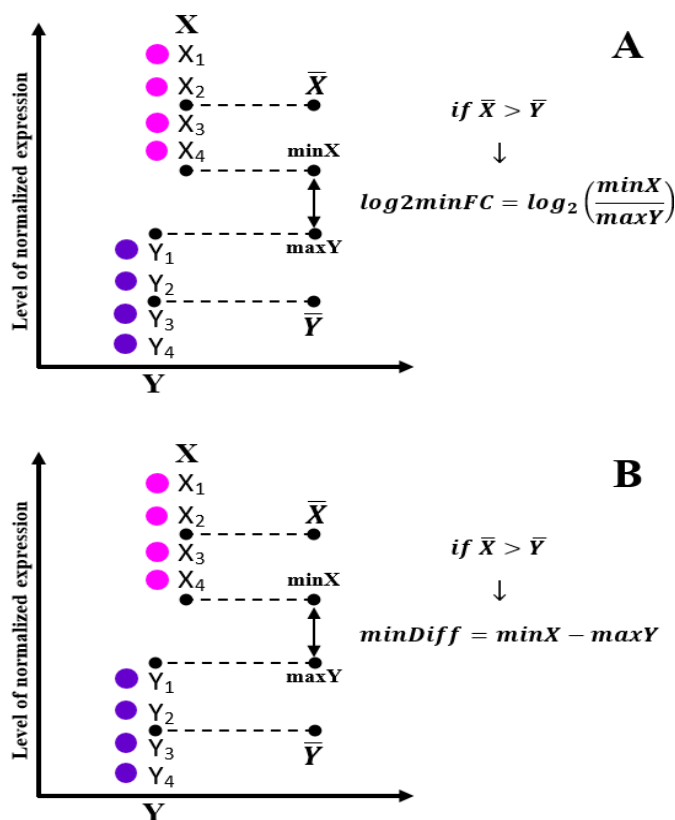
Przyjęto wartość progową $\log_2 \min FC = 0,100$. Ustalenie poziomu tego progu opierało się o moją wiedzę ekspercką i doświadczenie oraz o doniesienia literaturowe dotyczące potencjalnego poziomu powyżej którego uważa się, że występują wyniki mające znaczenie biologiczne [11, 12]. Geny z wartością $\log_2 \min FC$ poniżej progu są usuwane.

minDiff (Rycina 5.B) – niech $X = \{x_1, x_2, \dots, x_n\}$ będzie zbiorem pomiarów poziomów ekspresji danego genu dla próbek należących do jednej grupy, a $Y = \{y_1, y_2, \dots, y_m\}$ będzie zbiorem pomiarów poziomów ekspresji danego genu dla próbek należących do drugiej grupy. Wówczas definiujemy *minDiff* jako:

$$\text{minDiff} = \begin{cases} \text{min}X - \text{max}Y & \text{if } \bar{X} \geq \bar{Y} \\ \text{min}Y - \text{max}X & \text{if } \bar{X} < \bar{Y} \end{cases}$$

gdzie *minX* i *minY* oznaczają najmniejszą wartość w zbiorze *X* i *Y*, odpowiednio, a *maxX* i *maxY* oznaczają największą wartość w zbiorze *X* i *Y*, odpowiednio. \bar{X} , i \bar{Y} oznaczają wartość średnią dla zbiorów *X* i *Y*, odpowiednio. Wartość *minDiff* > 0 pokazuje, że przedziały wartości ekspresji dla porównywanych grup nie pokrywają się, a wartość *minDiff* ≤ 0 pokazuje, że przedziały wartości ekspresji dla porównywanych grup pokrywają się.

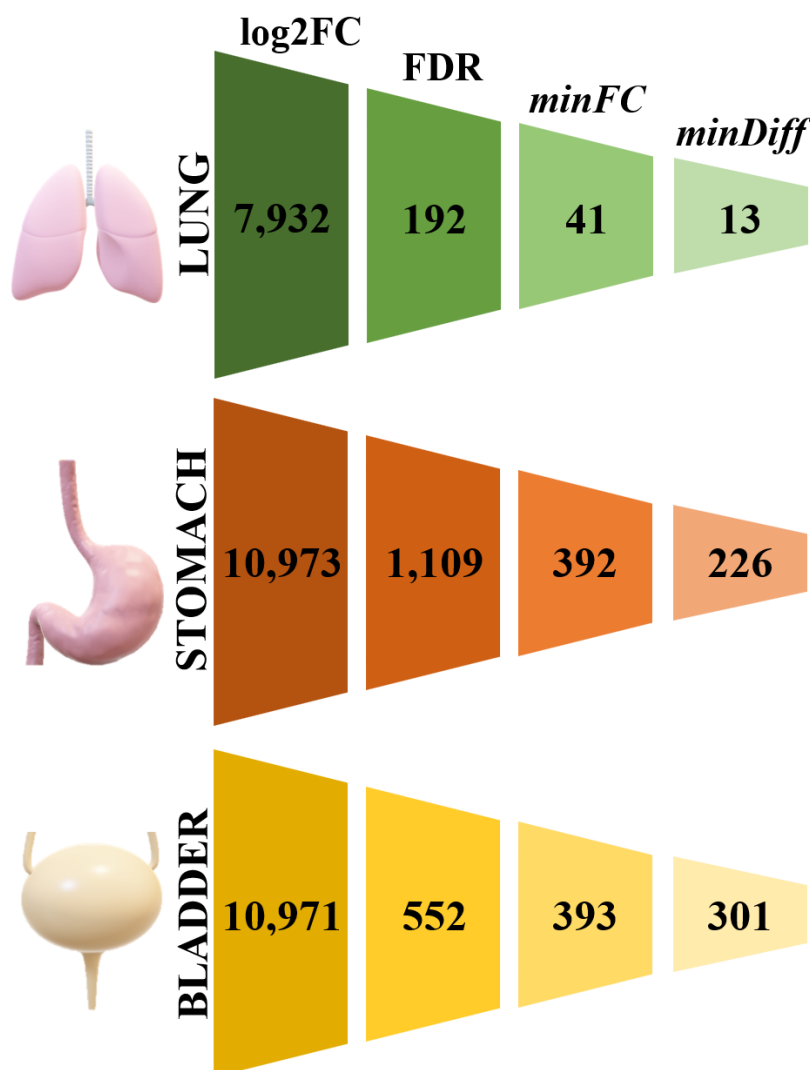
Przyjęto wartość progową *minDiff* = 100. Przy ustalaniu tego progu oceniono możliwość wykrycia takiej różnicy za pomocą metod stosowanych w oznaczeniach biomarkerów w praktyce klinicznej, takich jak IHC, ELISA i qPCR, a także to, czy taka różnica miałaby znaczenie biologiczne. W tym celu oceniano poziom znormalizowanych odczytów (średnie odczyty dla wariantów dzikich linii komórkowych: LS, SS i BS (Rycina 2)) dla białek, które są już literaturowo scharakteryzowane jako powiązane z danym narządem i/lub nowotworem w naszym przypadku płuc, żołądka i pęcherza moczowego. Informacje o takich białkach pozyskano z repozytorium Human Protein Atlas [13-15] oraz w większości przypadków dodatkowo zweryfikowano w materiałach dostępnych w repozytorium GeneCards®: The Human Gene Database [16]. Ponieważ są to białka o znanym funkcjonalnym znaczeniu w danej tkance, ich poziom ekspresji musi być wykrywalny przez standardowe techniki wymienione powyżej. Dlatego obecność takiej minimalnej różnicy między próbkami badanymi, a kontrolnymi będzie możliwa do wykrycia za pomocą wyżej wymienionych technik. Ponieważ 75% tak ocenianych białek mieściło się w zakresie do 86 (Q3 = 86,107), wartość 100 przyjęto jako minimalny próg różnicy. Geny o wartości *minDiff* poniżej progu są usuwane.



Rycina 5. Schemat (A) miary *minFC* i (B) miary *minDiff*.

Zastosowanie własności statystycznych zaimplementowanych w schemacie PREDICT pozwoliło wyselekcjonować mniejsze liczby potencjalnych biomarkerów, ale o bardziej obiecujących cechach. Eliminacja niepewnych kandydatów na etapie obliczeniowym, dzięki zastosowaniu schematu PREDICT, pozwoli na redukcję kosztów i wysiłku na etapie walidacji, który jest kolejną fazą rozwoju potencjalnego biomarkera.

Wykorzystanie schematu PREDICT doprowadziło do odfiltrowania niechcianych wyników. Liczby DEGs uzyskane tą metodą wynosiły odpowiednio 13, 226 i 301 dla danych z płuc, żołądka i pęcherza (Rysunek 6).



Rycina 6. Liczba zidentyfikowanych DEGs uzyskanych przez sekwencyjne zastosowanie progów dla miar: log₂ krotności zmiany (log₂FC) < 0,500, q wartość (FDR) < 0,050, log₂minFC > 0,100 i minDiff > 100.

Wyselekcjonowani kandydaci na biomarker charakteryzowali się cechami odpowiednimi dla markera, który może być stosowany w oznaczeniach klinicznych.

- Na podstawie analizy różnicowej przeprowadzonej za pomocą narzędzia DESeq2, zidentyfikowano istotnie statystycznie DEGs. Przyjęto próg istotności dla q wartości = 0,050. Geny o q wartościach poniżej progu zostały odfiltrowane.
- Potencjalne biomarkery charakteryzowały się wartością log₂FC > 0,500.
- Markery cechowały się odpowiednio dużą wartością miary log₂minFC > 0,100. Geny o wartości log₂minFC poniżej progu zostały odfiltrowane.

-
- Markery cechowały się także odpowiednio dużą wartością miary $minDiff > 100$. Geny o wartości $minDiff$ poniżej progu zostały odfiltrowane.
 - Implementacja miar $minFC$ i $minDiff$ zapewniała, że przedziały wartości ekspresji dla porównywanych grup się nie pokrywają. Dzięki temu uzyskano potencjalnych kandydatów o spójnym kierunku zmiany, co pozwala na powtarzalność testowania takiego markera.
 - Przyjęcie progów dla miar \log_2FC , q wartości, $minFC$ i $minDiff$ pozwala na selekcje kandydatów o pożądanym cechach czułości i swoistości.

Zidentyfikowane DEGs za pomocą narzędzia DESeq2 i PREDICT wykorzystano w dalszej analizie dotyczącej oceny kontekstu biologicznego. Kontekst ten został oceniony przez przeprowadzenie analizy ścieżek sygnałowych, w której zastosowano dwie metody: analizę nadreprezentacji (*ang. over-representation analysis*, ORA) oraz analizę wzbogacenia zbiorów genów (*ang. gene set enrichment analysis*, GSEA). Uzyskane zestawy ścieżek sygnałowych na podstawie DEGs żołądka i pęcherza poddano klasteryzacji celem wyodrębnienia grup ścieżek sygnałowych, które mogą być potencjalnie związane z mechanizmem oporności na inhibitory FGFR. Następnie z zestawu DEGs związanych z wybranymi grupami ścieżek wyselekcjonowano te, które spełniały statystyczne własności zaimplementowane w schemacie PREDICT. Dla żołądka i pęcherza zidentyfikowano odpowiednio 57 i 54 geny (DEGs), które następnie poddano ocenie na podstawie literatury naukowej. Ponieważ zidentyfikowano tylko 13 DEGs za pomocą schematu PREDICT dla danych z płuca, wykonano ocenę literaturową dla wszystkich tych genów.

Na podstawie analizy szlaków sygnałowych, połączonej z użyciem schematu PREDICT oraz przeglądem literatury, odkryto związek z potencjalnymi mechanizmami oporności na inhibitor FGFR dla większości wyselekcjonowanych genów. Otrzymane wyniki wskazują, że guzy uodpornione na działanie inhibitora FGFR wykształciły kompensacyjną aktywację szlaków regulujących proliferację komórek, tempo migracji, przeżycie, inwazyjność i hamowanie apoptozy.

Porównując zestawy genów wyselekcjonowane w trzech różnych typach raka, zidentyfikowano kilka potencjalnie uniwersalnych biomarkerów oporności na inhibitory FGFR, a mianowicie *SSRP1* (*ang. Structure Specific Recognition Protein 1*), *CCNB2* (*ang. Cyclin B2*), *CDT1* (*ang. Chromatin Licensing And DNA Replication Factor 1*) i *CENPO* (*ang. Centromere Protein O*). Te geny miały zmienioną ekspresję zarówno w raku żołądka, jak i pęcherza moczowego i wykazywały ten sam kierunek zmiany ekspresji w obydwu typach raka. Dlatego mogą one służyć jako uniwersalne biomarkery do predykcji oporności na inhibitory FGFR u pacjentów ze zdiagnozowanym rakiem żołądka lub pęcherza moczowego.

Podsumowując, użycie schematu PREDICT skutkuje odfiltrowaniem niepożądanych wyników, a wyselekcjonowane geny kandydackie posiadają cechy odpowiednie dla biomarkera, który może znaleźć praktyczne zastosowanie kliniczne. Przegląd literatury pozwolił na określenie związku większości wyselekcjonowanych genów z potencjalnymi mechanizmami oporności na inhibitory FGFR. Następnym krokiem w rozwoju biomarkera byłaby faza walidacji, aby potwierdzić, że zaobserwowane różnice dla wyselekcjonowanych kandydatów w fazie odkrywczej są także widoczne za pomocą innych metod i na różnym materiale biologicznym.

Podziękowania

Alexander Cortez był współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Społecznego (POWR.03.02.00-00-I029).

BIBLIOGRAFIA

1. World Health Organization, *Cancer*. 2023.
2. Dasgupta, A. and A. Wahed, *Chapter 13- Tumor markers*, in *Clinical Chemistry, Immunology and Laboratory Quality Control (Second Edition)*, A. Dasgupta and A. Wahed, Editors. 2021, Elsevier. p. 269-293.
3. Füzéry, A.K., et al., *Translation of proteomic biomarkers into FDA approved cancer diagnostics: issues and challenges*. *Clinical Proteomics*, 2013. **10**(1): p. 13.
4. McDermott, J.E., et al., *Challenges in Biomarker Discovery: Combining Expert Insights with Statistical Analysis of Complex Omics Data*. *Expert Opin Med Diagn*, 2013. **7**(1): p. 37-51.
5. Zarczynska, I., et al., *p38 Mediates Resistance to FGFR Inhibition in Non-Small Cell Lung Cancer*. *Cells*, 2021. **10**(12).
6. Zhou, Y., et al., *FGF/FGFR signaling pathway involved resistance in various cancer types*. *J Cancer*, 2020. **11**(8): p. 2000-2007.
7. Yamani, A., et al., *Discovery and optimization of novel pyrazole-benzimidazole CPL304110, as a potent and selective inhibitor of fibroblast growth factor receptors FGFR (1-3)*. *Eur J Med Chem*, 2021. **210**: p. 112990.
8. Mroz, E.A. and J.W. Rocco, *The challenges of tumor genetic diversity*. *Cancer*, 2017. **123**(6): p. 917-927.
9. Helsten, T., et al., *The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing*. *Clin Cancer Res*, 2016. **22**(1): p. 259-67.
10. Wu, Y.M., et al., *Identification of targetable FGFR gene fusions in diverse cancers*. *Cancer Discov*, 2013. **3**(6): p. 636-47.
11. McCarthy, D.J. and G.K. Smyth, *Testing significance relative to a fold-change threshold is a TREAT*. *Bioinformatics*, 2009. **25**(6): p. 765-71.
12. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biol*, 2014. **15**(12): p. 550.
13. The Human Protein Atlas, *The lung-specific proteome*. 2023.
14. The Human Protein Atlas, *The urinary bladder-specific proteome*. 2023.
15. The Human Protein Atlas, *The stomach-specific proteome*. 2023.
16. GeneCards®: The Human Gene Database, 2023.