Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
w Poznaniu

mgr Karol Czubak

Analiza somatycznych zmian liczby kopii w genomie nowotworowym w celu identyfikacji nowych genów/biomarkerów odgrywających ważną rolę w raku płuca

Praca doktorska
wykonana pod kierunkiem
dr hab. Piotra Kozłowskiego, prof. IChB PAN
w Instytucie Chemii Bioorganicznej PAN w Poznaniu

Poznań, 2015

Niniejsza praca doktorska była finansowana z grantu
Narodowego Centrum Nauki OPUS1
„Identyfikacja nowych genów odgrywających ważną rolę w procesie
nowotworzenia w częstych nowotworach człowieka”
nr 2011/01/B/NZ5/02773

Praca została wykonana we współpracy z Centrum Onkologii
im. prof. Franciszka Łukaszczyka w Bydgoszczy
(dr hab. n med. Marzana Anna Lewandowska,
prof. dr hab. n med. Janusz Kowalewski,
dr n med. Wojciech Jóźwicki,
dr hab. n med. Krzysztof Roszkowski)

W trakcie realizacji pracy doktorskiej, autor otrzymywał specjalne
stypendium naukowe z dotacji podmiotowej w ramach Krajowego
Naukowego Ośrodka Wiodącego „Poznańskie Konsorcjum RNA”

Pragnę serdecznie podziękować mojemu Promotorowi,
Panu dr hab. Piotrowi Kozłowskiemu, prof. IChB PAN
za opiekę, cenne wskazówki, szereg inspirujących dyskusji,
a także za wykazaną cierpliwość i wyrozumiałość

Chciałbym podziękować również Pani dr hab. n. med.
Marzenie Lewandowskiej, za koordynowanie współpracy
z Centrum Onkologii w Bydgoszczy oraz za zaangażowanie
w realizowane wspólnie projekty

Dziękuję również wszystkim Pracownikom i Doktorantom
Europejskiego Centrum Bioinformatyki i Genomiki
za okazaną życzliwość i wszelką pomoc.
W szczególności chcę tutaj podziękować
Gosi Marcinkowskiej-Swojak, która pomagała mi w moich
pierwszych eksperymentach, służąc doświadczeniem,
dobrą radą i poświęcając swój czas,
a także Kasi Klonowskiej, za cenne dyskusje i wskazówki

Wyrazy podziękowania za okazane wsparcie składam
również mojej Rodzinie i Przyjaciołom

Niniejsza rozprawa doktorska składa się z następujących części:

STRESZCZENIE

ABSTRACT

OPIS WYNIKÓW PRACY DOKTORSKIEJ

LISTA PUBLIKACJI

PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

1. Lewandowska M. A.*, **Czubak K.***, Klonowska K., Jóźwicki W., Kowalewski J., Kozłowski P.
   The use of a two-tiered testing strategy for the simultaneous detection of small EGFR mutations and EGFR amplification in lung cancer
   PLoS One 2015, 10:e0117983 (IF 3.23)

2. **Czubak K.**, Lewandowska M. A., Klonowska K., Roszkowski K., Kowalewski J., Figlerowicz M., Kozłowski P.
   High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer
   Oncotarget 2015, 6(27):23399-416 (IF 6.35)

3. Klonowska K., **Czubak K.**, Wojciechowska M., Handschuh L., Żmieńko A., Figlerowicz M., Dams-Kozłowska H., Kozłowski P.
   Oncogenomic Portals for the Visualization and Analysis of Genome-wide Cancer Data
   Oncotarget 2015, doi: 10.18632/oncotarget.6128 (IF 6.35)

OŚWIADCZENIA WSPÓŁAUTORÓW

STRESZCZENIE

W genomie nowotworowym występuje szereg zmian somatycznych, których rezultatem może być utrata funkcji genów supresorowych lub aktywacja onkogenów, prowadząca do niekontrolowanego namnażania komórek. Jednym z obserwowanych rodzajów zmian genetycznych w nowotworach są mutacje typu zmiany liczby kopii (amplifikacje i delecje), które mogą obejmować znaczną część genomu. Większość takich zmian ma charakter neutralny, jednak część stanowią mutacje prowadzące do inicjacji i progresji nowotworu. Rozpoznanie genów ulegających częstym amplifikacjom lub delecjom może więc być ważnym narzędziem identyfikacji genów odgrywających kluczową rolę w procesie nowotworzenia, odpowiednio onkogenów i genów supresorowych.

Tematyka moich badań obejmowała analizę zmienności genetycznej w raku płuca, który stanowi najczęstszą przyczynę zgonów związanych z nowotworami. Materiałem wykorzystywanym w prowadzonych przeze mnie eksperymentach były próbki DNA pochodzące z pierwotnych niedrobnokomórkowych raków płuca (NSCLC), zdiagnozowanych w Centrum Onkologii w Bydgoszczy. Natomiast podstawową techniką, jaką stosowałem, była zależna od ligacji multipleksowa amplifikacja sond (MLPA).

W pierwszym etapie moich badań przeprowadziłem analizę zmian w jednym z kluczowych dla raka płuca onkogenów, genie *EGFR.* Analiza ta wskazała na zależność amplifikacji genu *EGFR* od występowania w nim mutacji punktowych. Ponadto, stwierdzenie bardzo wysokiej czułości i specyficzności testu MLPA było podstawą do zaproponowania strategii dwupoziomowej analizy mutacji genu *EGFR* w raku płuca.

Dalsze badania pozwoliły mi zaobserwować występowanie bardzo dużej częstości zmian liczby kopii genów mikroRNA. Wyniki wskazują, że zmienność liczby kopii może być istotnym mechanizmem odpowiedzialnym za nowotworowo-specyficzną ekspresję wielu mikroRNA. Ponadto, uzyskane wyniki wykazują zaskakująco częstą amplifikację dwóch kluczowych genów biogenezy mikroRNA, *DICER1* i *DROSHA*. Porównanie wyników analizy genetycznej z danymi klinicznymi wykazało związek podwyższonej lub obniżonej liczby kopii niektórych z badanych genów (*miR-30d*, *miR-200b* oraz *DROSHA*) ze zmianą przeżywalności pacjentów. Może to sugerować potencjalną możliwość wykorzystywania zmian liczby kopii tych genów jako biomarkerów nowotworowych. Otrzymane przeze mnie wyniki skonfrontowałem z całogenomowymi danymi generowanymi w dużych projektach badania nowotworów, co pozwoliło mi wykazać zarówno wpływ zmienności liczby kopii genów *DICER1* i *DROSHA* na ich ekspresję, jak również wpływ zmian ekspresji tych genów na okres przeżycia pacjentów.

Podsumowując, rezultaty moich badań ukazują nieznane dotąd zależności dotyczące ważnych dla raka płuca onkogenów, a także wskazują na potencjalnie istotne znaczenie kilku innych genów, w tym genów mikroRNA i genów biogenezy mikroRNA. Wyniki uzyskane w ramach mojej pracy doktorskiej zostały opublikowane w dwóch recenzowanych publikacjach eksperymentalnych oraz jednej przeglądowej.

ABSTRACT

In cancer genome, there is number of somatic changes, which may result in loss of function of suppressor genes or in the activation of oncogenes, leading to uncontrolled proliferation and growth of cancer cells. One of the most frequently observed type of genetic alterations in cancer is copy number alteration, which may often constitute a large part of the genome. Majority of such changes has a neutral character, but some of them lead to the initiation and/or progression of cancer. Therefore, recognition of genes undergoing frequent amplifications and/or deletions is important for the identification of those which play an important role in tumorigenesis, oncogenes and/or tumor suppressor genes, respectively.

The subject of my studies included analysis of a genetic variation in lung cancer, which is a type of cancer causing the highest number of cancer-related deaths. For my experiments, I used DNA samples extracted from non-small-cell lung cancers (NSCLCs), diagnosed in the Oncology Centre in Bydgoszcz. The basic experimental method employed in my studies was the multiplex ligation-dependent probe amplification (MLPA).

In the first stage of lung cancer studies, I performed the analysis of alterations in one of the key lung cancer related genes, the *EGFR* gene. This analysis revealed the association between the occurrence of *EGFR* mutations and *EGFR* amplification. Furthermore, the determination of very high sensitivity and specificity of the MPLA test was the basis for proposing a two-tiered *EGFR* mutation analysis strategy.

As a result of my further studies, I observed very frequent copy number alterations of microRNA genes. These results indicate, that copy number alterations may be an important mechanism responsible for cancer-specific expression of many microRNAs. This investigation also revealed a surprisingly frequent amplification of two key genes involved in microRNA biogenesis, *DICER1* and *DROSHA*. Comparing the outcomes of genetic analysis with clinical data, I showed the association of copy number alterations of some of the investigated genes (*miR-30d*, *miR-200b* and *DROSHA*) with patients' survival. This result suggests that copy number alterations of these genes may be considered biomarkers in cancer. In the next step, I confronted the results of my analyses with whole-genome data, generated in large-scale cancer genomic projects. It allowed me to show both the impact of copy number alterations of *DICER1* and *DROSHA* on their expression, and the influence of their expression changes on patients' survival.

To summarize, the outcomes of my studies have revealed unknown associations of key lung cancer oncogenes, and have indicated the potential significance of several other genes, including microRNA genes and microRNA biogenesis genes. The obtained results have been published in three peer-reviewed articles, including two experimental articles, and one review.

OPIS WYNIKÓW PRACY DOKTORSKIEJ

Wprowadzenie

Nowotworami nazywamy zmiany spowodowane niekontrolowanym rozrostem komórek. Według Światowej Organizacji Zdrowia, rocznie z powodu nowotworów umiera na świecie ponad 8 milionów ludzi, a przewidywany wzrost liczby zgonów w najbliższych dwóch dekadach wyniesie 70%. [1, 2]. Istnieje ponad 100 typów nowotworów, przy czym każdy z nich wymaga osobnej diagnostyki i leczenia. Spośród wszystkich typów, najczęstszą przyczyną zgonów związanych z nowotworami jest rak płuca, który jest jednocześnie jednym z najbardziej heterogenicznych genetycznie nowotworów. Wyróżniamy w nim kilka podtypów, z których najczęstszym jest niedrobnokomórkowy rak płuca (NSCLC, ang. *non-small-cell lung cancer*) [3].

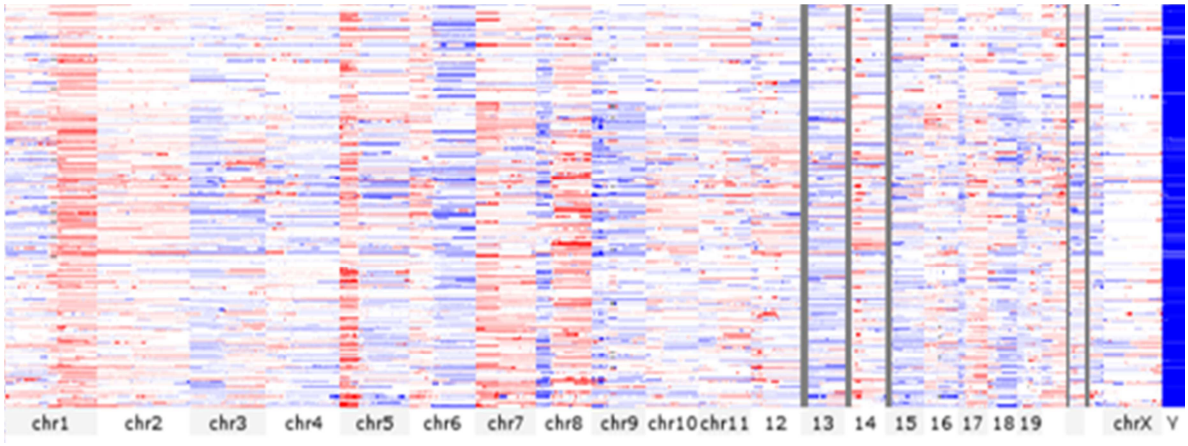
Rozwój nowotworów związany jest z występowaniem w genomie nowotworowym szeregu mutacji somatycznych. Pewna ich część może wpływać na funkcje genów, powodując utratę funkcji genów supresorowych lub aktywację (nabycie nowej funkcji i/lub zwiększenie ekspresji) onkogenów. Dodatkowo, ryzyko wystąpienia nowotworów może być modyfikowane przez tło genetyczne, w tym mutacje genetyczne o wysokiej penetracji, występujące w genach związanych z rodzinnymi formami nowotworów.

Obok mutacji punktowych oraz modyfikacji epigenetycznych, jednym z częstszych rodzajów zmian w genomie nowotworowym są mutacje typu zmiany liczby kopii [4]. Do zmian liczby kopii zaliczamy delecje, duplikacje oraz amplifikacje obejmujące regiony genomu o długości od około 1 kpz do kilku Mpz, czy nawet całe ramiona chromosomów. Analizy oparte na macierzach SNP wykazały, że w raku płuca regiony ulegające częstym zmianom liczby kopii obejmują 50% genomu [5]. W indywidualnych próbkach raka płuca, zmienności liczby kopii ulega średnio 40% genomu [6] (Rycina 1). Większość tych zmian stanowią zmiany neutralne, a tylko nieliczne odgrywają istotną rolę w inicjacji i progresji nowotworu.

Szacuje się, że liczba poznanych obecnie proto-onkogenów i genów supresorowych stanowi zaledwie około 15% wszystkich genów, które mogą mieć istotny udział w rozwoju nowotworów [4]. Skuteczną metodą poszukiwania nowych genów ważnych dla procesu nowotworzenia jest identyfikacja genów podlegających częstym mutacjom w genomie nowotworowym. Rozpoznanie takich genów może przyczynić się do lepszego zrozumienia

chorób nowotworowych, a tym samym do lepszej oceny ryzyka wystąpienia nowotworu i przebiegu choroby. Geny te również mogą stać się celami nowych terapii, w tym terapii personalizowanych.

W ostatnim czasie, poszukiwanie nowych biomarkerów nowotworowych w znacznym stopniu skupiło się na mikroRNA. mikroRNA są to niekodujące, jednoniciowe cząsteczki RNA o długości około 21 nukleotydów. Ich biogeneza z pierwotnych długich prekursorów jest procesem wieloetapowym, katalizowanym m. in. przez rybonukleazy DROSHA oraz DICER. Dojrzałe cząsteczki mikroRNA, poprzez oddziaływanie z regionem 3' UTR docelowych mRNA, przyczyniają się do wyciszania ekspresji genów [7, 8]. Biologiczna funkcja większości cząsteczek mikroRNA nie została dotąd poznana. Liczne badania wskazują jednak na potencjalną (zarówno onkogenną, jak i supresorową) rolę niektórych mikroRNA w procesie nowotworzenia [9].



Rycina 1 Wizualna reprezentacja (mapa cieplna) zmian liczby kopii w genomie nowotworowym na przykładzie NSCLC. Poziome rzędy reprezentują poszczególne próbki nowotworowe. Czerwonym i niebieskim kolorem zaznaczono regiony o odpowiednio podwyższonej i obniżonej liczbie kopii. Na osi X zaznaczone są pozycje kolejnych chromosomów (wygenerowano w UCSC Cancer Genomics Browser, na podstawie danych TCGA).

Cel pracy

Ogólnym celem mojej pracy doktorskiej było pogłębienie wiedzy na temat zmienności genetycznej w NSCLC oraz określenie jej roli w rozwoju nowotworu. Moje szczególne zainteresowanie skupiało się na somatycznych zmianach liczby kopii (delecjach i amplifikacjach), które poprzez efekt dawki mogą regulować aktywność wielu genów, a także mogą być wskaźnikiem roli poszczególnych genów w nowotworze (geny supresorowe, onkogeny).

W ramach powyżej określonego celu realizowałem dwa następujące cele szczegółowe:

- analiza amplifikacji genu *EGFR* oraz ich wzajemnych relacji z onkogennymi mutacjami punktowymi *EGFR,* oraz z podstawowymi danymi klinicznymi i epidemiologicznymi. Analiza genetyczna genu *EGFR* ma ogromne znaczenie ze względu na kluczową rolę tego genu w kontekście celowości stosowania terapii spersonalizowanej z wykorzystaniem inhibitorów kinaz tyrozynowych.
- analiza zmienności liczby kopii wybranych genów mikroRNA w kontekście mechanizmu ich regulacji (podniesienie lub obniżenie ekspresji) oraz potwierdzenia ich roli w raku płuca (NSCLC). W ramach tego celu prowadziłem również analizę dwóch kluczowych genów biogenezy mikroRNA, *DICER1* i *DROSHA.*

Materiały

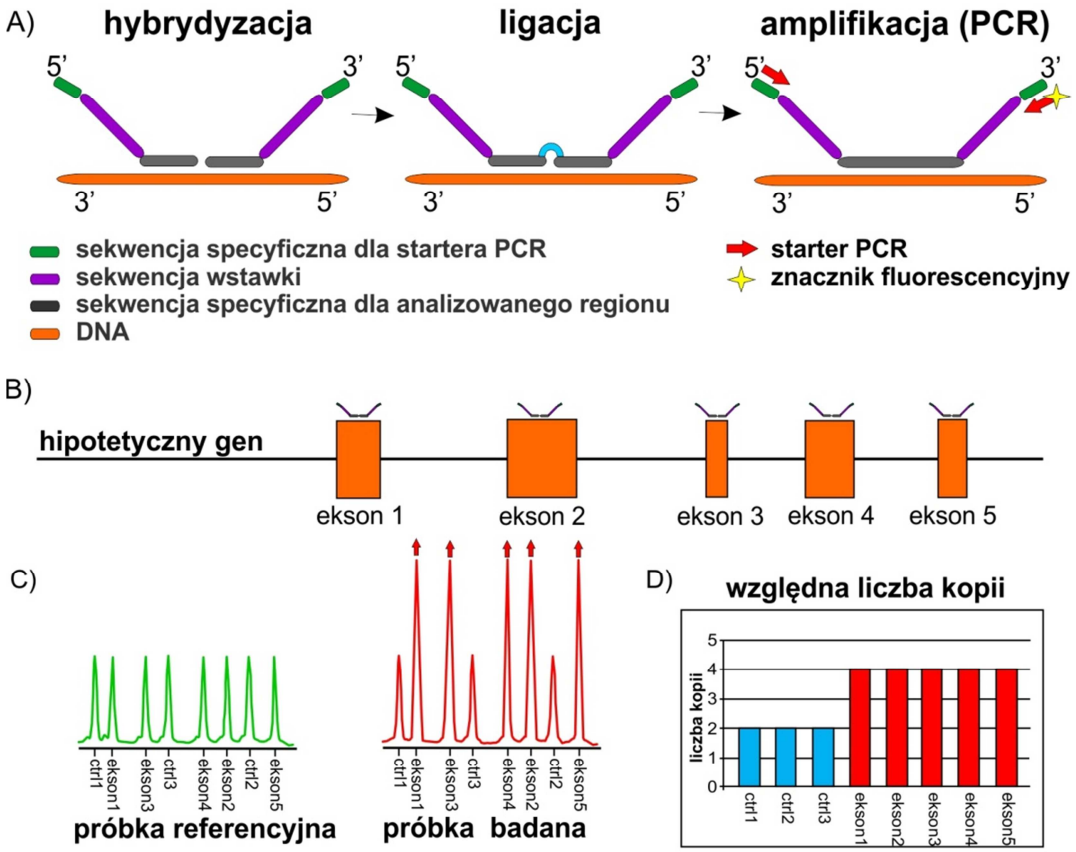
Materiałem wykorzystywanym w prowadzonych przeze mnie eksperymentach były próbki DNA pochodzące z pierwotnych NSCLC, zdiagnozowanych w Centrum Onkologii im. Franciszka Łukaszczyka w Bydgoszczy. Zostały one wyizolowane z bloczków parafinowych (FFPE; 60% próbek) oraz z rozmazów cytologicznych (40% próbek). Materiał poddany został wcześniej wstępnej ocenie histologicznej, mającej na celu ustalenie typu nowotworu oraz oszacowanie procentowej zawartości komórek nowotworowych w tkance. Część próbek scharakteryzowana została pod względem cech klinicznych pacjentów, takich jak wiek, płeć czy czas przeżycia. Próbki otrzymałem dzięki współpracy z dr. hab. n. med. Marzeną Lewandowską z Centrum Onkologii w Bydgoszczy.

Metody

Podstawową techniką wykorzystywaną w mojej pracy była zależna od ligacji multipleksowa amplifikacja sond (MLPA, ang. *multiplex ligation-dependent probe amplification*). W moich badaniach używałem testów MLPA zaprojektowanych według zmodyfikowanej strategii, która wykorzystuje krótkie, syntetyczne sondy oligonukleotydowe [10]. Zastosowanie tej strategii umożliwiło mi samodzielne projektowanie testów MLPA, odpowiednich dla celów analizy.

W skrócie, technika MLPA pozwala analizować jednocześnie wiele miejsc w genomie, z wykorzystaniem nawet 45 specyficznych sond [11]. Każda sonda MLPA składa się z dwóch

pół-sond zaprojektowanych tak, aby hybrydyzowały do bezpośrednio przylegających do siebie sekwencji docelowych. W dalszym etapie pary pół-sond rozpoznające prawidłowo sekwencje docelowe ulegają ligacji, a następnie są amplifikowane (PCR) przy użyciu pary uniwersalnych starterów, z których jeden jest wyznakowany fluorescencyjnie. Produkty amplifikacji rozdzielane są z wykorzystaniem elektroforezy kapilarnej. Wynikiem rozdziału elektroforetycznego jest charakterystyczny układ pików odpowiadających poszczególnym sondom. Intensywność pików jest proporcjonalna do dawki (liczby kopii) sekwencji docelowych. Dzięki temu, na etapie analizy wyników możliwe jest określenie względnej (oszacowanej względem próbek referencyjnych) liczby kopii danego regionu genomu. Ponadto, dzięki zastosowaniu specjalnie zaprojektowanych sond specyficznych dla mutacji, możliwe jest również wykrywanie za pomocą techniki MLPA małych mutacji [10] (Rycina 2)



Rycina 2 Schemat metody MLPA oraz analizy jej wyników. A) Etapy reakcji MLPA. Poszczególne sekwencje wchodzące w skład sondy MLPA zaznaczone zostały odpowiednimi kolorami. B) Mapa hipotetycznego genu, na której zaznaczone zostały poszczególne eksony oraz pozycje sond MLPA. C) Przykładowe elektroferogramy próbki referencyjnej i próbki badanej. Strzałkami zaznaczono podwyższone sygnały sond. D) Wykres słupkowy, przedstawiający stosunek intensywności sygnału poszczególnych sond w próbce badanej i referencyjnej. Przedstawiony przykład reprezentuje amplifikację hipotetycznego genu.

W moich badaniach wykorzystywałem również takie techniki laboratoryjne, jak sekwencjonowanie metodą Sangera, PCR w czasie rzeczywistym (RT PCR, ang. *Real-Time PCR*) oraz ilościowa analiza PCR techniką emulsyjną (ddPCR, ang. *droplet digital PCR*). W analizach otrzymywanych wyników wykorzystywałem odpowiednie testy statystyczne, w tym test Fishera, test Chi kwadrat dla trendu, analiza wariancji ANOVA czy korelacja Pearsona. Wykresy przeżywalności pacjentów tworzone były według metody Kaplana-Meiera, a w analizach statystycznych w tym przypadku wykorzystałem testy log-rank (Mantel-Cox) oraz log-rank dla trendu.

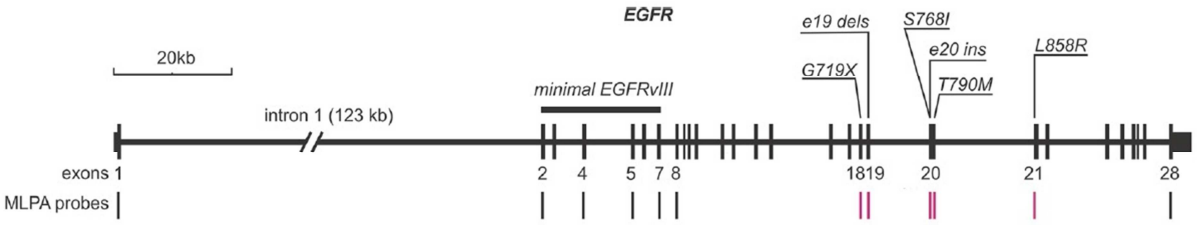
Skrótowe omówienie publikacji, będących wynikiem pracy doktorskiej

W pierwszej z omawianych prac, wchodzących w skład mojej rozprawy doktorskiej, to jest w pracy pt. *„The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small EGFR Mutations and EGFR Amplification in Lung Cancer”* [12], opisałem wynik analizy małych mutacji somatycznych oraz zmian liczby kopii genu receptora naskórkowego czynnika wzrostu (*EGFR* ang. *epidermal growth factor receptor*) w raku płuca. Gen *EGFR* jest jednym z ważniejszych znanych onkogenów odpowiedzialnych za rozwój raka płuca, a mutacje punktowe w tym genie są biomarkerami wskazującymi wrażliwość (np. mutacja p.L858R), lub oporność (np. mutacja p.T790M) nowotworu na inhibitory kinazy tyrozynowej, a tym samym na zasadność stosowania terapii celowanej [13]. Oprócz raka płuca, zmiany w genie *EGFR* są również biomarkerami często wykorzystywanymi w innych typach nowotworów [14, 15].

Analizę zmian w genie *EGFR* przeprowadziłem z wykorzystaniem zaprojektowanego wcześniej w naszym zespole testu MLPA, EGFRmut+ (Rycina 3) [10], który pozwala na jednoczesną i wzajemną analizę zmian liczby kopii (amplifikacji) genu *EGFR* oraz występujących w nim małych mutacji. Dodatkowo, test EGRFmut+ umożliwia wykrywanie amplifikacji dwóch innych onkogenów, biorących udział w rozwoju raka płuca, *MET* oraz *ERBB2.* Analizę *EGFR* przeprowadziłem dla 239 próbek NSCLC.

Równocześnie z prowadzoną przeze mnie analizą MLPA, dokonana została analiza z wykorzystaniem komercyjnego testu Real–Time PCR, EGFR-RT52 (Entrogen Inc.), który pozwala badać 29 najczęstszych mutacji występujących w domenie kinazy tyrozynowej genu *EGFR.* Wyniki uzyskane za pomocą dwóch wspomnianych technik wykazały zgodność na poziomie 98,7%. Specyficzność testu MLPA wyniosła 100%, co oznacza, że nie uzyskałem żadnych wyników fałszywie pozytywnych. Czułość testu wyniosła natomiast 90%. Obniżona

czułość wynika jednak z faktu, że dwie z trzech niewykrytych mutacji nie były objęte testem MLPA. Trzecia z nich występowała natomiast w próbce o bardzo niskiej zawartości komórek nowotworowych (15%).



Rycina 3 Schemat przedstawiający gen *EGFR* z zaznaczonymi (za pomocą pionowych linii umieszczonych pod mapą genu) pozycjami sond MLPA (test EGFRmut+). Różowym kolorem wyróżnione zostały sondy, dzięki którym możliwe jest wykrywanie małych mutacji (nazwy wykrywanych mutacji zaznaczone zostały powyżej miejsca ich występowania).

Przeprowadzona analiza wykazała obecność 30 mutacji w 29 (12%) spośród 239 próbek NSCLC. Najczęściej występującymi mutacjami były delecje w eksonie 19 oraz substytucja L858R w eksonie 21 genu *EGFR* (Tab. 1 w Lewandowska M.A., Czubak K. i wsp., 2015). Wynik ten jest zgodny z danymi literaturowymi, według których wspomniane mutacje stanowią łącznie ponad 80% wszystkich mutacji domeny kinazy tyrozynowej genu *EGFR* [16]. Analizy statystyczne uwzględniające charakterystykę pacjentów, pozwoliły mi zaobserwować kilka zależności. Jedną z nich była różna częstość występowania mutacji u kobiet (24,4%) i u mężczyzn (4,7%). Chociaż wyższa częstość mutacji u kobiet była już obserwowana wcześniej, to jednak w żadnym z poprzednich badań nie zauważono tak dużej dysproporcji [17-20]. Może to wynikać ze specyfiki warunków środowiskowych lub zachowań (np. palenie papierosów) w Polsce.

Jak już wspomniałem, zaletą testu EGFRmut+ w porównaniu do testów komercyjnych wykorzystywanych w badaniach mutacji w genie *EGFR,* jest możliwość jednoczesnej analizy mutacji punktowych oraz zmian liczby kopii genu *EGFR*, a także dwóch innych onkogenów*, MET* i *ERBB2.* Dzięki tej właściwości byłem w stanie zaobserwować zjawisko wzajemnego wykluczania się amplifikacji tych trzech onkogenów. Ponadto, wykazałem istnienie wyraźnego związku amplifikacji genu *EGFR* z wystąpieniem w nim mutacji. Zauważyłem bowiem, że mutacje występują znacznie częściej w próbkach z amplifikacją (90% próbek), niż w próbkach z normalną liczbą kopii genu *EGFR* (7%). Dokładna analiza wyników reakcji MLPA pozwoliła mi zauważyć, że w większości przypadków pierwszym etapem aktywacji *EGFR* jest

najprawdopodobniej mutacja punktowa, która pociąga za sobą amplifikację zmutowanego allelu. Dla potwierdzenia obserwowanych zmian liczby kopii zarówno genu *EGFR*, jak i dwóch pozostałych onkogenów (*MET* i *ERBB2*) przeprowadziłem analizę liczby kopii tych genów z wykorzystaniem dwóch innych technik, RT PCR oraz ddPCR. Dla wyników uzyskanych tymi metodami zauważyłem bardzo wysoką korelację z wynikami testu MLPA.

Podsumowując, głównym osiągnięciem uzyskanym przeze mnie w opisanej powyżej analizie było wykazanie związku pomiędzy występowaniem w genie *EGFR* mutacji punktowych a jego amplifikacją. Dodatkowo, przeprowadzona analiza potwierdziła przydatność testu EGFRmut+ w analizie mutacji *EGFR.* Pozwoliło to nam zaproponować strategię dwupoziomowej analizy mutacji genu *EGFR* w raku płuca. Proponowana strategia może być traktowana jako odpowiedź na najnowsze sugestie, które ze względu na możliwość generowania przez ultraczułe metody wyników fałszywie pozytywnych, rekomendują stosowanie jednocześnie dwóch niezależnych metod wykrywania mutacji. Zalecanym jest stosowanie jednej metody ultraczułej i jednej o umiarkowanej/standardowej czułości (wspólne rekomendacje towarzystw College of American Pathologists, International Association for the Study of Lung Cancer i Association for Molecular Pathology) [16, 21, 22].

W celu dalszego pogłębiania wiedzy dotyczącej genetyki raka płuca, podjąłem się zbadania somatycznej zmienności liczby kopii genów mikroRNA w tym nowotworze. Wyniki badań zostały opisane w pracy pt. *„High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer”* [23]. Opisywaną tutaj analizę przeprowadziłem dla genów kodujących te mikroRNA, które w raku płuca ulegają najczęstszej zmianie ekspresji (6 mikroRNA o podniesionej i 6 o obniżonej ekspresji). W celu selekcji takich genów wykorzystałem wyniki dwóch niezależnych meta-analiz, podsumowujących kilkadziesiąt prac eksperymentalnych, zawierających wyniki całogenomowych analiz ekspresji mikroRNA w raku płuca [24, 25]. Do analizy włączyłem także geny mikroRNA-17 i mikroRNA-155 (dla obu z nich opisano w literaturze związek zmiany ekspresji z przeżywalnością pacjentów), oraz dwa kluczowe geny biogenezy mikroRNA, *DICER1* i *DROSHA* (dla których w raku płuca również obserwuje się zmiany ekspresji).

W celu analizy wybranych genów, zaprojektowałem testy MLPA w taki sposób, aby na każdy gen mikroRNA przypadały po 2 sondy, natomiast na geny *DICER1* i *DROSHA* po 3 sondy

MLPA. Dzięki takiemu podejściu mogłem wyłączyć z późniejszej analizy te przypadki, w których w danej próbce sondy dla danego genu nie dawały jednoznacznych rezultatów.

Z wykorzystaniem zaprojektowanych przeze mnie testów MLPA, przeprowadziłem analizę zmienności liczby kopii wybranych genów mikroRNA i genów biogenezy mikroRNA w 254 próbkach NSCLC. W wyniku dokonanej analizy zaobserwowałem zaskakująco dużą zmienność liczby kopii genów mikroRNA. Niektóre z nich, jak np. *miRNA-17* czy *miRNA-205*, ulegały amplifikacjom częstszym i większym niż dobrze znane onkogeny *EGFR* czy *MET*. Ponadto, dla niektórych analizowanych genów mikroRNA (przykładowo, *miRNA-17* czy *miRNA-205*), kierunek zmian liczby kopii pokrywał się z często obserwowanym kierunkiem zmian ich ekspresji (zwiększona liczba kopii odpowiadała podwyższonej ekspresji w nowotworze, a obniżona liczba kopii – obniżonej ekspresji). Obserwacja ta może sugerować istotny udział amplifikacji i delecji w regulacji ekspresji niektórych genów mikroRNA.

Oprócz genów mikroRNA, dużą częstość amplifikacji obserwowałem także dla obu badanych genów biogenezy mikroRNA, *DICER1* i *DROSHA*. Przeprowadzona przeze mnie analiza literatury wykazała, że w bezpośrednim sąsiedztwie genu *DROSHA* zlokalizowany jest gen *GOLPH3,* niedawno zidentyfikowany jako onkogen ważny dla rozwoju raka płuca [26]. Amplifikacja genu *GOLPH3* potencjalnie mogłaby pociągnąć za sobą obserwowaną zwiększoną liczbę kopii genu *DROSHA.* W celu zbadania takiej ewentualności, zaprojektowałem odpowiedni test MPLA, pozwalający badać jednocześnie liczbę kopii obu tych genów. Ponadto, dzięki sondom równomiernie rozmieszczonym wzdłuż krótkiego ramienia chromosomu 5, test ten pozwolił także sprawdzić, czy zmiany w regionie genu *DROSHA* mają charakter zlokalizowany, czy też obejmują większy region chromosomu. Analiza wyników testu wykazała, że obserwowana amplifikacja genu *DROSHA* jest niezależna od genu *GOLPH3* (w próbkach, w których obserwowałem amplifikacje genu *DROSHA,* gen *GOLPH3* nie ulegał amplifikacjom). Przeprowadzona analiza pokazała również, że zwiększenie liczby kopii genu *DROSHA* może być związane z amplifikacją znacznej części krótkiego ramienia chromosomu 5. Trzeba jednak zauważyć, że bez względu na to, jaka jest przyczyna amplifikacji genu *DROSHA*, a tym samym podniesienia poziomu jego ekspresji, może mieć ona poważne konsekwencje dla komórki. Gen *DROSHA* koduje bowiem jedną z kluczowych rybonukleaz biorących udział w procesie biogenezy mikroRNA. Zmiana ekspresji *DROSHA*

może więc wpływać na globalny poziom mikroRNA w komórce, a to z kolei może wpływać na ekspresję wielu innych genów.

W dalszych analizach, w celu zbadania wpływu zmian liczby kopii badanych genów na przeżywalność pacjentów, wykorzystałem dostępne dla części próbek dane kliniczne. Z moich obserwacji wynika, że zmiany liczby kopii trzech genów (*miR-30d*, *miR-200b* oraz *DROSHA*), miały wpływ na czas przeżycia pacjentów. Dla genu *DROSHA* stwierdziłem dodatkowo istnienie trendu, który pokazywał, że im bardziej zwiększona była liczba kopii tego genu, tym gorsze były rokowania pacjentów. Podobnego wyniku nie obserwowałem dla drugiego z badanych genów biogenezy mikroRNA, *DICER1.*

W celu skonfrontowania moich wyników, uzyskanych dla genów *DROSHA* i *DICER1*, z danymi pochodzącymi z dużych projektów badania genomów nowotworowych, wykorzystałem dane dostępne w portalach cBioPortal [27-29] i BioProfiling.de [30, 31]. Z ich pomocą wykazałem, że zwiększona liczba kopii zarówno genu *DROSHA*, jak i *DICER1,* koreluje z podwyższoną ekspresją tych genów. Podwyższona ekspresja tych genów skutkuje z kolei zmianą przeżywalności pacjentów (odpowiednio obniżoną i podwyższoną przeżywalnością). Zależności takie obserwowałem nie tylko dla raka płuca, ale również dla innych typów nowotworów (Figure 6 i Supplementary Figure S2 w Czubak K. i wsp.,).

Podsumowując, głównym wnioskiem, jaki wynika z opisanej powyżej pracy jest występowanie bardzo dużej częstości zmian liczby kopii genów mikroRNA oraz genów *DICER1* i *DROSHA* w raku płuca. Co więcej, zmienność ta może być istotnym mechanizmem regulującym ekspresję wspomnianych genów, a ta z kolei może wpływać na przeżywalność pacjentów.

Cennym doświadczeniem uzyskanym w trakcie moich badań było zapoznanie się z możliwościami, jakie niesie ze sobą powszechna dostępność do danych generowanych przez duże projekty badania genomów nowotworowych. Dane te udostępniane są za pomocą portali (przeglądarek) onkogenomicznych, takich jak np. wspomniane już cBioPortal czy BioProfiling.de. Doświadczenie zdobyte w pracy z tego typu portalami wykorzystałem do przygotowania, wraz z innymi współautorami, pracy przeglądowej pt. *„Oncogenomic Portals for the Visualization and Analysis of Genome-wide Cancer Data”* [32]. Praca ta w zwięzły sposób charakteryzuje strukturę oraz przedstawia możliwe zastosowania wybranych portali onkogenomicznych (Tumorscape, UCSC Cancer Genomics Browser, ICGC data portal,

COSMIC, cBioPortal, IntOGen oraz BioProfiling.de), które wyselekcjonowaliśmy na podstawie ich dostępności oraz przydatności w analizie danych uzyskiwanych w dużych projektach badania genomów nowotworowych, takich jak The Cancer Genome Atlas (TCGA) czy Cancer Genome Project (CGP). Moja rola w przygotowaniu niniejszego artykułu polegała na zebraniu informacji, przygotowaniu opisów oraz sporządzeniu ilustracji dla czterech z siedmiu omawianych szerzej portali, tj. Tumorscape, COSMIC, cBioPortal oraz BioProfiling.de. Wymienione portale opisane są w odpowiadających im podrozdziałach pracy.

Opisywane portale dostarczają możliwości analizy i wizualizacji występujących w genomie nowotworowym zmian takich jak małe mutacje, zmiany liczby kopii, zmiany epigenetyczne czy zmiany ekspresji genów. Dzięki zintegrowaniu ze sobą różnorodnych danych, wymienione cechy mogą być skorelowane z dostępnymi cechami klinicznymi, epidemiologicznymi czy molekularnymi (przykładowo, możliwe jest badanie związku ekspresji wybranych genów z przeżywalnością pacjentów, wpływu mutacji punktowej na funkcję białka i wiele innych).

Każdy z portali przedstawionych w omawianej pracy przeglądowej scharakteryzowaliśmy pod kątem szeregu cech takich jak struktura danego portalu, źródło danych w nim zawartych, typy nowotworów dostępnych do analiz oraz informacje, jakie można na podstawie tych analiz uzyskać. Ponadto, dla każdego z portali zamieściliśmy przykłady jego praktycznych zastosowań. Aby uczynić pracę bardziej przystępną dla czytelnika, dla prezentowanych portali sporządziliśmy ilustracje przedstawiające przykładowe analizy i wizualizacje danych. Należy nadmienić, że w pracy szerzej opisaliśmy te portale, które w trakcie jej powstawania wydawały się nam być najbardziej przydatnymi spośród wielu dostępnych portali onkogenomicznych. Wśród pozostałych, wartymi wspomnienia są portale takie, jak Oasis, Oncomine, Cancer Genetics Web czy CaSNP.

Podsumowując, prace wchodzące w skład mojej rozprawy doktorskiej opisują zagadnienia związane z genetyką raka płuca. Ujawniają one nowe, wcześniej nieznane zależności charakteryzujące poznane już, ważne w tym typie nowotworu onkogeny, jak również przedstawiają potencjalnie istotne znaczenie kilku innych genów, w tym genów mikroRNA i genów biogenezy mikroRNA. Wykorzystanie w analizach portali onkogenomicznych pozwoliło wpisać otrzymane wyniki w szerszy kontekst badań raka, jak

również przedstawić potencjalne możliwości wykorzystywania danych pochodzących z dużych projektów badania genomów nowotworowych.

Jestem również współautorem pracy eksperymentalnej pt. „Analysis of large mutations in *BARD1* in patients with breast and/or ovarian cancer: the Polish population as an example" [33]. W pracy tej przedstawione zostały wyniki analizy, mającej na celu określenie roli genu *BARD1* w predyspozycji do rodzinnego raka piersi i/lub jajnika. Rezultatem tej analizy było wykluczenie obecności dużych mutacji w *BARD1* oraz identyfikacja trzech nowych mutacji punktowych w tym genie. Testy funkcjonalne oraz kompleksowa analiza *in silico* potwierdziły duży potencjał funkcjonalny wszystkich 3 zidentyfikowanych mutacji. Chociaż projekt ten nie stanowił głównego wątku moich badań i nie wchodzi w skład mojej pracy doktorskiej, udział w nim pozwolił mi poszerzyć moją wiedzę w dziedzinie genetyki nowotworów o ważny aspekt, jakim jest dziedziczna predyspozycja do raka.

Bibliografia

Prace wchodzące w skład rozprawy doktorskiej wyróżnione zostały za pomocą podkreślenia.

1. http://www.who.int/mediacentre/factsheets/fs297/en/ref.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65:87-108.
3. http://www.who.int/selection_medicines/committees/expert/20/applications/NonSmallCell LungCancer.pdf?ua=1.
4. Stratton MR, Campbell PJ and Futreal PA. The cancer genome. Nature. 2009; 458:719-724.
5. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007; 450:893-898.
6. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899-905.
7. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116:281-297.
8. Boehm M and Slack FJ. MicroRNA control of lifespan and metabolism. Cell Cycle. 2006; 5: 837-840.
9. Esquela-Kerscher A and Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006; 6:259-269.
10. Marcinkowska M, Wong KK, Kwiatkowski DJ and Kozlowski P. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. ScientificWorldJournal. 2010; 10:2003-2018.
11. Jeuken J, Cornelissen S, Boots-Sprenger S, Gijsen S and Wesseling P. Multiplex ligation-dependent probe amplification: a diagnostic tool for simultaneous identification of different genetic markers in glial tumors. The Journal of molecular diagnostics : J Mol Diagn. 2006; 8:433-443.
12. Lewandowska MA, Czubak K, Klonowska K, Jozwicki W, Kowalewski J and Kozlowski P. The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small EGFR Mutations and EGFR Amplification in Lung Cancer. PloS One. 2015; 10:e0117983.
13. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497-1500.
14. Generali D, Leek R, Fox SB, Moore JW, Taylor C, Chambers P and Harris AL. EGFR mutations in exons 18-21 in sporadic breast cancer. Ann Oncol. 2007; 18:203-205.
15. Carlsson SK, Brothers SP and Wahlestedt C. Emerging treatment strategies for glioblastoma multiforme. EMBO Mol Med. 2014; 6:1359-1370.
16. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J, Thunnissen E and Ladanyi M. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol. 2013; 8:823-859.
17. Girard N, Sima CS, Jackman DM, Sequist LV, Chen H, Yang JC, Ji H, Waltman B, Rosell R, Taron M, Zakowski MF, Ladanyi M, Riely G, et al. Nomogram to predict the presence of EGFR activating mutation in lung adenocarcinoma. Eur Respir J. 2012; 39:366-372.
18. Liu Y, Xu ML, Zhong HH, Heng WJ and Wu BQ. EGFR mutations are more frequent in well-differentiated than in poor-differentiated lung adenocarcinomas. Pathol Oncol Res. 2008; 14:373-379.
19. Huang SF, Liu HP, Li LH, Ku YC, Fu YN, Tsai HY, Chen YT, Lin YF, Chang WC, Kuo HP, Wu YC, Chen YR and Tsai SF. High frequency of epidermal growth factor receptor mutations with

complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. Clin Cancer Res. 2004; 10:8195-8203.

20. Kondo M, Yokoyama T, Fukui T, Yoshioka H, Yokoi K, Nagasaka T, Imaizumi K, Kume H, Hasegawa Y, Shimokata K and Sekido Y. Mutations of epidermal growth factor receptor of non-small cell lung cancer were associated with sensitivity to gefitinib in recurrence after surgery. Lung cancer. 2005; 50:385-391.
21. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J, Thunnissen E and Ladanyi M. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. Arch Pathol Lab Med. 2013; 137:828-860.
22. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, Jenkins RB, Kwiatkowski DJ, Saldivar JS, Squire J, Thunnissen E, Ladanyi M, College of American Pathologists International Association for the Study of Lung C, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013; 15:415-453.
23. Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M and Kozlowski P. High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer. Oncotarget. 2015; 6:23399-23416.
24. Vosa U, Vooder T, Kolde R, Vilo J, Metspalu A and Annilo T. Meta-analysis of microRNA expression in lung cancer. Int J Cancer. 2013; 132:2884-2893.
25. Guan P, Yin Z, Li X, Wu W and Zhou B. Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. J Exp Clin Cancer Res. 2012; 31:54.
26. Scott KL, Kabbarah O, Liang MC, Ivanova E, Anagnostou V, Wu J, Dhakal S, Wu M, Chen S, Feinberg T, Huang J, Saci A, Widlund HR, et al. GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. Nature. 2009; 459:1085-1090.
27. http://www.cbioportal.org.
28. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401-404.
29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6: pl1.
30. http://www.bioprofiling.de.
31. Antonov AV. BioProfiling.de: analytical web portal for high-throughput cell biology. Nucleic Acids Res. 2011; 39:W323-327.
32. Klonowska K, Czubak K, Wojciechowska M, Handschuh L, Zmienko A, Figlerowicz M, Dams-Kozlowska H and Kozlowski P. Oncogenomic portals for the visualization and analysis of genome-wide cancer data. Oncotarget. 2015.
33. Klonowska K, Ratajska M, Czubak K, Kuzniacka A, Brozek I, Koczkowska M, Sniadecki M, Debniak J, Wydra D, Balut M, Stukan M, Zmienko A, Nowakowska B, et al. Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example. Sci Rep. 2015; 5:10424.

LISTA PUBLIKACJI

Publikacje wchodzące w skład rozprawy doktorskiej:

1. Lewandowska M. A.*, **Czubak K.***, Klonowska K., Jóźwicki W., Kowalewski J., Kozłowski P., 2015
   The use of a two-tiered testing strategy for the simultaneous detection of small EGFR mutations and EGFR amplification in lung cancer
   PLoS One 2015, 10:e0117983 (IF 3.23)

2. **Czubak K.**, Lewandowska M. A., Klonowska K., Roszkowski K., Kowalewski J., Figlerowicz M., Kozłowski P.
   High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer
   Oncotarget 2015, 6(27):23399-416 (IF 6.35)

3. Klonowska K., **Czubak K.**, Wojciechowska M., Handschuh L., Żmieńko A., Figlerowicz M., Dams-Kozłowska H., Kozłowski P.
   Oncogenomic Portals for the Visualization and Analysis of Genome-wide Cancer Data
   Oncotarget 2015, doi: 10.18632/oncotarget.6128 (IF 6.35)

Publikacje niewchodzące w skład rozprawy doktorskiej:

1. Klonowska K., Ratajska M., **Czubak K.**, Kuźniacka A., Brożek I., Koczkowska M., Śniadecki M., Dębniak J., Wydra D., Bałut M., Stukan M., Żmieńko A., Nowakowska B., Irminger-Finger I., Limon J. and Kozlowski P.
   Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example
   Scientific Reports 2015, 5:10424. (IF 5.58)

PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

1

Lewandowska M. A.*, Czubak K.*, Klonowska K., Jóźwicki W., Kowalewski J., Kozłowski P.

*Autorzy mieli taki sam wkład w przygotowanie publikacji

„The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer"

PLoS One 2015, 10:e0117983.

PLOS ONE

RESEARCH ARTICLE

The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer

Marzena Anna Lewandowska[1,2*‡], Karol Czubak[3,4‡], Katarzyna Klonowska[3,4], Wojciech Jozwicki[5], Janusz Kowalewski[6], Piotr Kozlowski[3,4*]

1 Molecular Oncology and Genetics Unit, Department of Tumour Pathology and Pathomorphology, Franciszek Lukaszczyk Oncology Centre, Bydgoszcz, Poland, **2** Department of Thoracic Surgery and Tumours, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland, **3** European Centre of Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, **4** Poznan University of Technology, Pl. Marii Sklodowskiej-Curie 5, 60-965, Poznan, Poland, **5** Department of Tumour Pathology and Pathomorphology, Franciszek Łukaszczyk Oncology Centre, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland, **6** Department of Thoracic Surgery and Tumours, Franciszek Łukaszczyk Oncology Centre, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland

‡ These authors contributed equally to this work.
* lewandowskam@co.bydgoszcz.pl (MAL); kozlowp@yahoo.com (PK)

Citation: Lewandowska MA, Czubak K, Klonowska K, Jozwicki W, Kowalewski J, Kozlowski P (2015) The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer. PLoS ONE 10(2): e0117983. doi:10.1371/journal.pone.0117983

Academic Editor: Rafael Rosell, Catalan Institute of Oncology, SPAIN

Received: May 16, 2014

Accepted: January 5, 2015

Published: February 26, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by research grant 2011/01/B/NZ5/02773 from National Science Centre (http://www.ncn.gov.pl/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Lung cancer is the leading cause of cancer-related death worldwide. Recent progress in lung cancer diagnosis and treatment has been achieved due to a better understanding the molecular mechanisms of the disease and the identification of biomarkers that allow more specific cancer treatments. One of the best known examples of personalized therapy is the use of tyrosine kinase inhibitors, such as gefitinib and erlotinib, for the successful treatment of non-small-cell lung cancer patients selected based on the specific *EGFR* mutations. Therefore, the reliable detection of mutations is critical for the application of appropriate therapy. In this study, we tested a two-tiered mutation detection strategy using real-time PCR assays as a well-validated high-sensitivity method and multiplex ligation-dependent probe amplification (MLPA)-based EGFRmut+ assay as a second-tier standard-sensitivity method. One additional advantage of the applied MLPA method is that it allows the simultaneous detection of EGFR mutations and copy-number alterations (i.e., amplifications) in *EGFR, MET* and *ERBB2*. Our analysis showed high concordance between these two methods. With the use of this two-tier strategy, we reliably determined the frequency of *EGFR* mutations and *EGFR, MET* and *ERBB2* amplifications in over 200 lung cancer samples. Additionally, taking advantage of simultaneous copy number and small mutation analyses, we showed a very strong correlation between EGFR mutations and EGFR amplifications and a mutual exclusiveness of EGFR mutations/amplifications with MET and ERBB2 amplifications. Our results proved the reliability and usefulness of the two-tiered EGFR testing strategy.

Introduction

Lung cancer is the leading cause of cancer-related death worldwide. The treatment of lung cancer is traditionally based on a histopathological evaluation that distinguishes two major types of lung cancer: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC), the latter of which can be subdivided into squamous cell carcinoma, large cell carcinoma, adenocarcinoma and cancers with mixed histology. Substantial recent progress in the treatment of lung cancer (especially adenocarcinomas) has been achieved by advances in our understanding of its pathology; the current treatment options include specialized agents based on the presence or absence of specific genetic biomarkers ("personalized therapy"), such as mutations in the epidermal growth factor receptor (*EGFR*) [1] or gain-of-function translocations or inversions involving the anaplastic lymphoma receptor tyrosine kinase (*ALK*) [2].

It was shown that certain somatic mutations within the kinase domain of *EGFR* sensitize cancers to treatment with *EGFR*-specific tyrosine kinase inhibitors (TKIs), such as erlotinib or gefitinib (reviewed in [3]). Among the most common sensitizing *EGFR* mutations are L858R in exon 21 and in-frame deletions in exon 19, which together account for over 80–85% of all *EGFR* mutations. However, the occurrence of the secondary T790M mutation in exon 20 causes acquired resistance to TKIs and causes the progression of cancers treated with TKIs [4,5]. Therefore, the reliable detection of *EGFR* mutations is an important factor that allows the personalized treatment of lung cancer patients.

In the last 10 years, numerous methods with different sensitivities and specificities have been used to detect *EGFR* mutations in cancer samples. These methods include Sanger sequencing, single strand conformation polymorphism (SSCP) [6], co-amplification at lower denaturation temperature-PCR (COLD PCR) [7], immunohistochemistry with *EGFR*-mutation specific antibodies [8], peptide nucleic acid-locked nucleic acid (PNA-LNA) PCR clamp assays, real-time PCR (RT-PCR) methods [9] and next generation sequencing [10]. However, none of the methods that have been used so far are ideal, and each of these methods has limitations that mostly relate to the following characteristics of cancer samples: (i) diversified types of tumor samples available for analysis (surgical, biopsied), (ii) contamination with normal cells, (iii) the genetic heterogeneity of tumors, (v) the often low frequency of the analyzed mutations, (iv) degradation of DNA, and (v) damage to or modification of DNA [the two latter factors also apply to formalin-fixed, paraffin-embedded (FFPE) samples, which are the most frequently available samples]. The most serious problems resulting from the limitations of tumor sample analysis are false negative and false positive errors that may lead to the misclassification and inadequate treatment of cancer. Therefore, to reduce the fraction of misclassified samples, it was recently proposed in the evidence-based guideline of three professional societies (College of American Pathologists, International Associations for the Study of Lung Cancer, and Association for Molecular Pathology) that, if possible, *EGFR* mutation testing should be carried out with two methods (a two-tiered testing strategy). This guideline represents state-of-the-art molecular lung cancer testing and was jointly published in three journals [11–13]. For simplicity we will subsequently refer only to [11]. This two-tier method should be based on different mutation detection principles and should cover different ranges of sensitivity, consisting of standard-sensitivity and high-sensitivity methods.

In this study, we tested over 200 NSCLC samples with the use of two complementary methods, a routinely used commercial RT-PCR assay (a high-sensitivity method) [9] and a new multiplex ligation-dependent probe amplification (MLPA)-based EGFRmut+ assay (a standard-sensitivity, second-tier method) [14]. Our analysis showed a high concordance between these two methods and thus proved the reliability and usefulness of the EGFRmut+ assay as a second-tier method for *EGFR* mutation testing. With the use of these methods, we characterized and

estimated the frequency of somatic *EGFR* mutations in a set of lung cancer samples from central Poland. One additional advantage of the EGFRmut+ assay is that it allows a mutation analysis and relative copy number determination (i.e., amplification detection) in parallel. We used this approach to find a very strong correlation between *EGFR* amplification and the occurrence of *EGFR* mutations and to determine the rough frequency of mutant alleles in our analyzed samples.

Materials and Methods

Selection and processing of NSCLC samples for molecular analysis

We retrospectively reviewed a cohort of 239 patients with histopathologically confirmed NSCLC diagnosed from 2011 to 2012 at the Franciszek Lukaszczyk Oncology Center in Bydgoszcz (central Poland). The age of the patients ranged from 35 to 81. A total of 239 specimens that passed the quality control steps (microscopic analysis and tumor content qualification as well as qualitative and quantitative DNA analysis) were obtained following 143 surgeries, 91 fine-needle aspiration (FNA) procedures, and 5 endobronchial ultrasound with guided transbronchial needle aspiration (EBUS-TBNA) procedures or pleural fluid samplings. The samples were stained with hematoxylin and eosin for the qualitative and quantitative analysis of tumor cells in the analyzed material (including macrodissection in marked out samples). The qualification of biological material for molecular analysis was based on the previously described qualitative and quantitative scales for cytological [9] and histological [15] material. Informed written consent for genetic testing, approved by the F. Lukaszczyk Oncology Center in Bydgoszcz was obtained from all of the patients and the study was approved by our local ethics committee, Bioethics Committee of Ludwig Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun. The data were analyzed anonymously.

DNA isolation

DNA isolation was performed after the macrodissection of a region indicated by the pathomorphologist. Genomic DNA was isolated from FFPE adenocarcinoma tissue or cytological smears using a QIAamp FFPE Mini Kit (QIAGEN) according to the manufacturer's instructions with the following modifications. We resuspended the pellet in 180 μl of Tissue Lysis Buffer with 20 μl of proteinase K, and then vortexed and continuously shook the cell pellet at short intervals during an overnight incubation at 56°C. The DNA quantity and quality were evaluated by NanoDrop absorbance analysis and agarose gel electrophoresis.

Mutation analysis by Real-Time PCR

The RT-PCR method uses mutation-specific probes to evaluate 29 point mutations, including the T790M mutation (EGFR-RT52; Entrogen, Inc., Tarzana, CA). A DNA quantity of 200–650 ng was adequate for the detection of 29 mutations in the samples of interest; the internal control VIC fluorescent probes and *EGFR* fluorescent probes were FAM dye-labeled. The amplification curves were evaluated according to the recommendations of the manufacturer.

Mutation and amplification analysis by MLPA

MLPA analysis was performed with the use of a custom-designed EGFRmut+ assay, which has previously been described in greater detail [14]. This assay simultaneously allows the detection of oncogenic *EGFR* mutations and an analysis of the copy number (amplification detection) of *EGFR*, *MET*, and *ERBB2*. All reagents except the EGFRmut+ probe mix were purchased form MRC-Holland Amsterdam, The Netherlands (www.mlpa.com). The MLPA reactions were run

according to the manufacturer's general recommendations (MRC-Holland), as described earlier in [16,17]. Briefly, 5 μl of genomic DNA (at a concentration of approximately 20 ng/nl) was incubated at 98°C for 5 min, cooled to room temperature and mixed with 1.5 μl of EGFRmut+ probes mix and 1.5 μl of SALSA hybridization buffer. The reaction was then denatured at 95°C for 2 min and hybridized at 60°C for 16 h. The hybridized probes were ligated at 54°C for 15 min by the addition of 32 μl of ligation mixture. Following heat inactivation, the ligation reaction was cooled to room temperature, mixed with 10 μl of PCR mixture (polymerase, dNTPs, and universal primers, one of which was labeled with fluorescein) and subjected to PCR amplification for 35 cycles. The MLPA products were subsequently diluted 20x in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated via capillary electrophoresis (POP7 polymer) in an ABI Prism 3130XL apparatus (Applied Biosystems). The obtained electropherograms were analyzed using GeneMarker software v1.91 (2.4.0). The signal intensities (peak heights) were retrieved and transferred to prepared Excel sheets (available upon request). For each individual sample, the signal intensity of each probe was divided by the average signal intensity of the control probes to normalize the obtained values and to equalize run-to-run variation, and the normalized value for each peak was then divided by a corresponding value in the reference samples and multiplied by 2. The final MLPA result of each sample is presented on bar-plot, in which the bars show the relative copy number value of the subsequent probes.

EGFR copy number analysis by quantitative PCR (qPCR) and droplet digital PCR (ddPCR)

The qPCR analysis was performed with the use of MESA GREEN qPCR MasterMix Plus for SYBR Assay (Eurogentec, Seraing, Belgium), according to the manufacturer's general recommendations. ddPCR was performed with the use of QX200 system and EvaGreen Supermix (BIO-RAD, CA, USA) according to manufacturer's general recommendations, as described before [18,19]. The ddPCR analysis was performed in a multiplex format with the co-amplification of control-amplicon and one of the test-amplicons in one reaction. To achieve proper separation of droplet types, the intensity of test and control signal was differentiated by amplicons' length and primers' concentrations. For both methods the same set of PCR primers was used: (i) test-amplicon for *EGFR* exon 2: forward primer GCAGTTGGGCACTTTTGAAG, reverse primer TTCCAAATTCCCAAGGACCA (concentration in qPCR—300 nM, concentration in ddPCR—100 nM, amplicon length 83 bp); (ii) test-amplicon for EGFR exon 18: forward primer TTGTGGAGCCTCTTACACCC, reverse primer CCTTCAAGATCCTCAAGAGAGC (concentration in qPCR—300 nM, concentration in ddPCR—100 nM, amplicon length 64 bp); (iii) control-amplicon: forward primer GCTGACCTGTTGGCTGAAAA, reverse primer GAATCGCTGTGGCCTTGATG (concentration in qPCR—300 nM, concentration in ddPCR—200 nM; amplicon length 113 bp). The amplicons either overlapped, or were closely located to the following MLPA probes: EGFR_e2, EGFR_e18, and ctrl_1, respectively. The optimized annealing temperature was set at 59°C and 58°C, in qPCR and ddPCR, respectively.

Results

All 239 samples were analyzed as blind samples by two methods: (i) a commercial RT-PCR assay (EGFR-RT52; Entrogen Inc.) that covers 29 of the most common oncogenic mutations in exons 18, 19, 20 and 21 of *EGFR* (for details, see the manufacturers webpage; http://www.entrogen.com/web2/egfr-mutation-analysis-kit) and (ii) a custom-designed MLPA assay (EGFRmut+) that simultaneously allows the detection of amplifications and small mutations in *EGFR* (for details see [14]) (Fig. 1). Briefly, in addition to the standard dosage-sensitive (DS) probes, the EGFRmut+ assay is also composed of two types of mutation-sensitive probes.

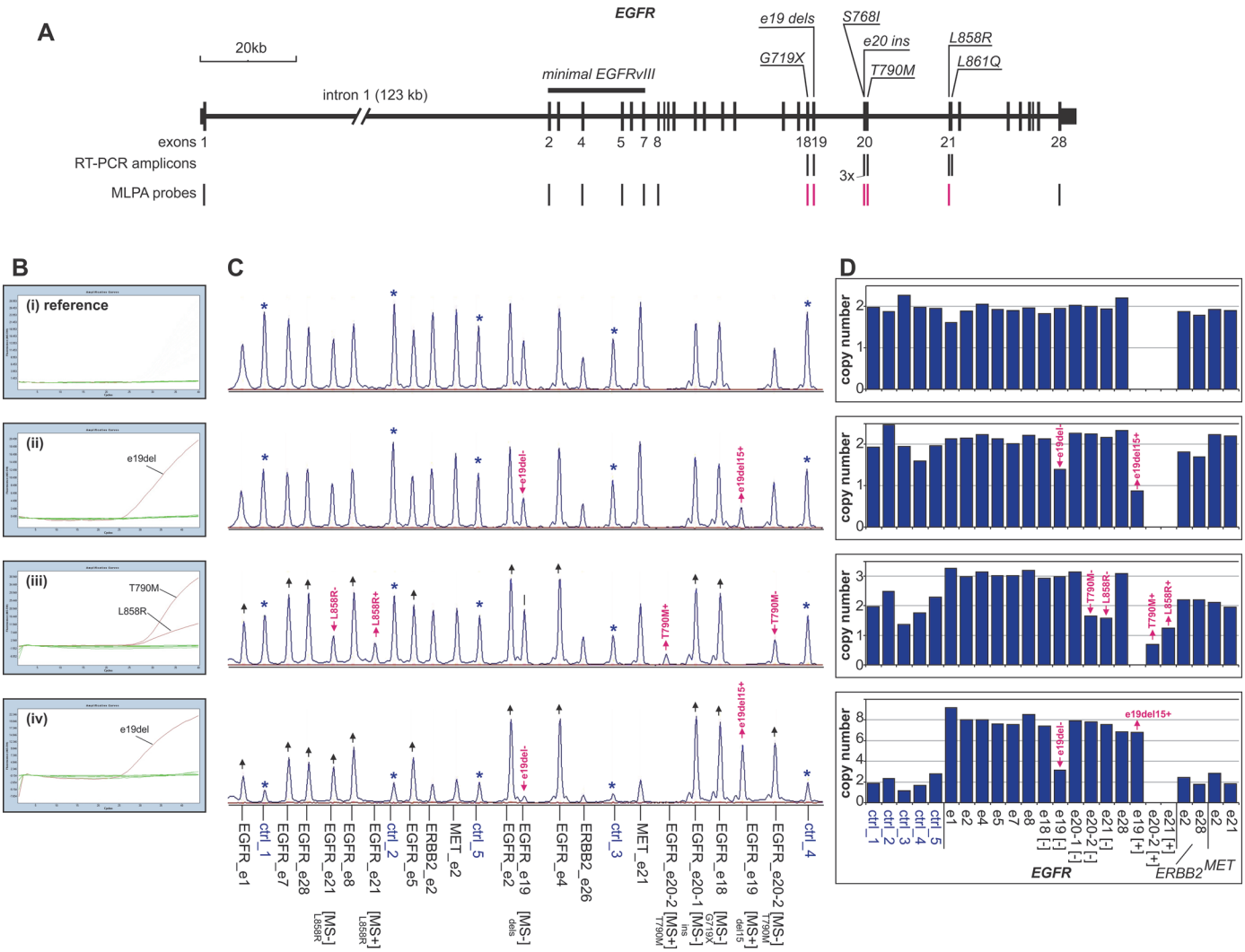


Fig 1. The strategy of *EGFR* mutation detection by combined RT-PCR and MLPA-based analyses. A) Map of the *EGFR* gene with the positions of the RT-PCR amplicons (EGFR-RT52) and MLPA probes (EGFRmut+ assay) indicated (vertical lines under the map). The mutation-sensitive EGFRmut+ probes are indicated in red. The positions of oncogenic *EGFR* mutations are indicated over the map. B) The RT-PCR results representing (from the top) (i) the reference sample (a sample with no mutations or amplification), (ii) a sample with the most common in-frame deletion in exon 19 (c.2235_2249del15), (iii) a sample with both L858R in exon 21 and T790M in exon 20, and (iv) a sample with an in-frame deletion in exon 19 (c.2235_2249del15) and *EGFR* amplification. In each graph, the overlapping results of the 8 RT-PCR reactions covering 29 *EGFR* mutations are shown. The red lines indicate the positive amplification-curves of specific *EGFR* mutations (pointed on graph), and the green base line represents the non-amplified signal due to the lack of the evaluated mutation in the analyzed sample. C) MLPA electropherograms of samples analyzed by RT-PCR (panel B). The probe IDs are indicated under the electropherograms. The asterisks indicate the control probes; the pink arrowheads indicate reduced signal of MS- probes and increased signal of MS+ probes, respectively; and the black arrowheads indicate amplified signals of *EGFR*-specific probes. D) Bar plots corresponding to the electropherograms shown in panel C and representing the normalized copy number value (y-axis) of each probe (x-axis). The pink arrowheads indicate a reduced copy number value of the MS- probes and an increased copy number value of the MS+ probes, respectively.

doi:10.1371/journal.pone.0117983.g001

Mutation-sensitive (MS-) probes are types of DS probes that are specific to the wild-type sequence but overlap with the sites of the following mutations: G719A/S/C (G719X) in exon 18 (probe e18), in-frame deletions in exon 19 (probe e19), S768I and in-frame insertions in exon 20 (probe e20-1), T790M in exon 20 (probe e20-2) and L858R in exon 21 (probe e21). The occurrence of one of these mutations in an analyzed sample causes a decrease in the signal of the corresponding MS- probe. The EGFRmut+ assay also contains three mutation-sensitive (MS+) probes that are specific for mutant sequences. The signals from these probes occur only if the

corresponding mutation is present. Two of these probes recognize the sequences of the most common *EGFR* mutations (the most common in-frame deletion in exon 19, c.2235_2249del15 (probe e19+) and L858R (probe e21+)), and the third recognizes the T790M mutation, which is associated with TKI resistance (probe e20-2+). Additionally, EGFRmut+ contains a few DS probes that are specific either for *MET* or *ERBB2* that allow amplifications of these genes to be detected.

Characteristics of the detected mutations

In total, we detected 30 mutations in 29 out of 239 samples (12.1%). Both L858R and T790M were found in one sample. Among the identified mutations were 16 in-frame deletions in exon 19 (53%), 9 substitutions L858R (30%), 2 substitutions L861Q (7%), one substitution G719X, one in-frame insertion in exon 20 and one substitution T790M (S1 Table; summarized in Table 1). The mutations were much more frequent in women 22/68 than in men 7/142 (24.4% versus 4.7%, respectively; $p<0.0001$). A stratification of the mutation frequency by age [<55 (13%), >55 (11.9%)] and sample type [FFPE (11.9%), cytological samples (12.5%)] did not show significant influences of these factors on mutation frequency (Table 1). We also did not observe a decrease in the mutation frequency in samples with a lower percentage of tumor cells (PTC). Note however, that only a small fraction (roughly 5%) of the analyzed samples had PTC≤30%.

Comparison of RT-PCR and MLPA mutation detection methods

There is no gold standard method for mutation detection in cancer samples, where a particular mutation may account for a very low fraction of the analyzed DNA. Therefore, to evaluate the quality of the MLPA-based EGFRmut+ assay, we compared it to the routinely used and well validated RT-PCR method. A direct comparison of the results of RT-PCR and MLPA showed a

Table 1. Sample characteristics and mutations detected by combined RT-PCR and MLPA-based analyses.

general statistics				types of *EGFR* mutations						gains + amplifications (%)		
		number of samples	samples with *EGFR* mut. (%)	G719X	in-frame dels in ex19 (*)	in-frame ins in ex20	L858R	L858R+ T790M	L861Q	EGFR	ERBB2	MET
	all patients	239	29 (12.1)	1†	16 (10)	1	8	1	2‡	19 (7.9)	5 (2.1)	28 (11.7)
sex	female	90	22 (24.4)	1	11 (7)	0	7	1	2	16 (17.8)	1 (1.1)	12 (13.3)
	male	149	7 (4.7)		5 (3)	1	1			3 (2.0)	4 (2.7)	16 (10.7)
age	35–55	54	7 (13.0)		4 (2)	1	1	1	2	6 (11.1)		6 (11.1)
	56–81	185	22 (11.9)	1	12 (8)		7			13 (7.0)	5 (2.7)	22 (11.9)
type of material	FFPE samples	143	17 (11.9)	1	9 (4)		5		2	8 (5.6)	2 (1.4)	20 (14)
	cytological samples §	96	12 (12.5)		7 (6)	1	3	1		11 (11.5)	3 (3.1)	8 (8.3)

* the most frequent in-frame deletion in exon 19 (c.2235_2249del15)

†, ‡ mutations that were not detected by MLPA-based assay due to low PTC and the lack of mutation-specific probe, respectively

§ includes FNA, EBUS-TBNA and pleural fluid sampling.

doi:10.1371/journal.pone.0117983.t001

very high concordance between these two methods (98.7% concordant results) as well as 100% specificity (no false positives) and 90% sensitivity (27 out of 30 mutations detected) of the EGFRmut+ test. Among the 3 mutations that were not detected by the EGFRmut+ assay were two cases with the L861Q substitution, which is not covered by EGFRmut+ probes (Fig. 1A), and one G719X mutation in a sample with a low PTC (15%). Additionally, among the samples analyzed by MLPA were 4 blind duplicates with 3 different mutations. In all cases, the results of the duplicate samples were concordant. Moreover, it must be noted that MS+ probes (signal occurrence) detect mutations more easily and with a higher confidence than MS- probes (signal decrease). The MS+ probes allowed the confident detection of mutations, even when those mutations account for a very small fraction of the analyzed DNA ($\sim$10%) in samples with PTC$\leq$30%. For example, the L858R mutation, which is covered by both MS+ and MS- probes, in three samples, could not be detected by a decrease in the signal of the MS- probe but was easily detected by the occurrence of a low but clear signal of the MS+ probe. The lower sensitivity of the MS- probes results mostly from the relatively high signal variation of the DS probes in cancer samples. The greater MLPA signal variation in cancer samples has been observed previously and results mostly from the genetic heterogeneity of cancer samples [20] and the lower quality and frequent degradation of DNA samples isolated from FFPE tissues [21–25].

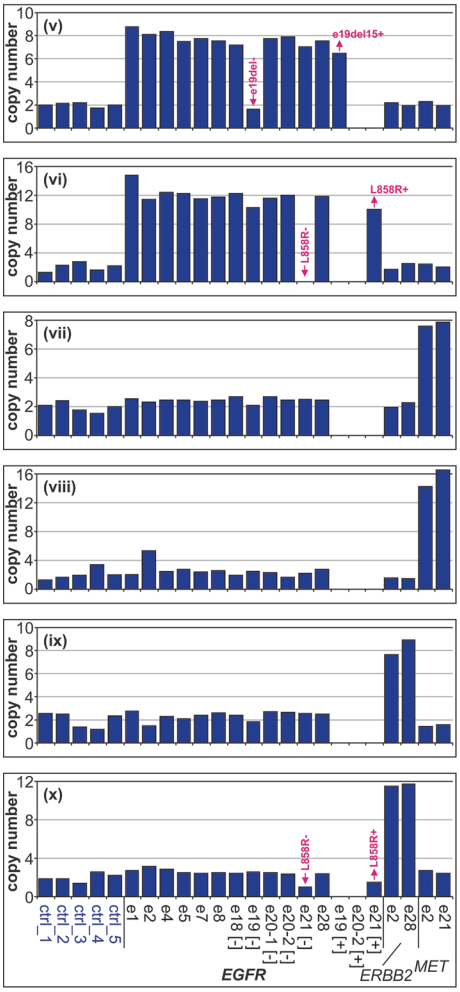
Amplification of EGFR, MET and ERBB2 and its relation to mutation frequency

The additional advantage of the MLPA assay is that in addition to detecting mutations, it also provides information about a relative copy number of all analyzed sequences/probes. It allows a rough estimation of the fraction of detected mutations in analyzed samples and allows detection of copy number changes (amplifications) in the analyzed genes: *EGFR*, *MET* and *ERBB2*. Defining relative copy number 3–4 as a "gain" and $\geq$4 as an "amplification," we identified 12 (5%), 17 (7%) and, 2 (1%) samples with gains and 7 (3%), 11 (5%) and, 3 (1%) samples with amplifications of *EGFR*, *MET*, and *ERBB2*, respectively (S1 Table and Fig. 2). The distribution of copy number amplitude was similar for all 3 genes, with the greatest amplifications exceeding 10 copies. As in case of the mutations, we observed a much higher frequency of *EGFR* gains/amplifications in women than in men (18% versus 2%, respectively; $p<0.0001$). A similar trend was not observed for *MET* (13% versus 11%, respectively), and a reversed but not significant trend was observed for *ERBB2* (1% versus 3%).

As shown in Fig. 2, the gains and amplifications of *EGFR*, *MET* and *ERBB2* were mutually exclusive. However, we observed very strong association between the occurrence of *EGFR* mutations and *EGFR* amplification (Chi square test for trend; $p<0.0001$). All but one cancer sample (90%) with *EGFR* amplification and 7 out of 12 samples (58%) with *EGFR* gain had an *EGFR* mutation.

A careful examination of the MLPA results with *EGFR* amplifications (see examples in Fig. 1 and Fig. 2) indicates that the signal decrease from the MS- probes and the signal increase of the MS+ probes (fraction of mutated copies) is approximately the same or even greater than the increase in *EGFR* copy number. This result suggests that all of the amplified copies of *EGFR* in the mutant samples contain the mutation (i.e., that only the mutant allele undergoes amplification) and that mutations occur as an early triggering event, making *EGFR* amplification beneficial for cancer. Note that all of the samples contain some amount of normal DNA, which may flatten the observed copy number changes and mask the effect in samples with a lower level of amplification.

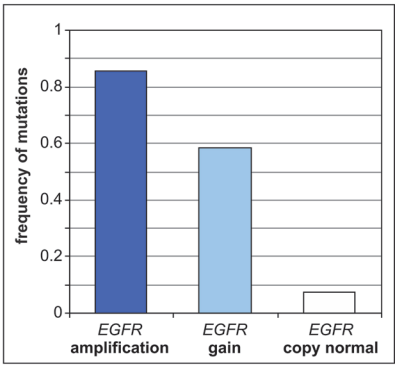
This observation prompted us to sequence exons 18–21 (encoding the tyrosine kinase domain) in samples with *EGFR* gains/amplifications in which no known *EGFR* mutation was



B

#	sex	mutations *EGFR*	gains/amplifications *EGFR*	*MET*	*ERBB2*
(vi)	F	L858R	12		
(v)	F	e19del	8		
(iv)	F	e19del	8		
	F		5.6		
	F	e19del	5.2		
	F	e19del	4.6		
	F	L858R	4		
	F	L858R	3.8		
	F		3.6		
	F	e19del	3.4		
	F	e19del	3.4		
	F	e19del	3.4		
	M	e19del	3.4		
	M		3.4		
	M		3.2		
(iii)	F	L858R*	3		
	F	L858R	3		
	F		3		
	F		3		
(viii)	M			15.4	
(vii)	M			7.6	
	F			5.6	
	F			5.6	
	M			4.8	
	M			4.8	
	F			4	
	F			4	
	M			4	
	M			4	
	M			4	
	F	L858R		3.6	
	F			3.6	
	M			3.4	
	M			3.4	
	F	e19del		3.2	
	F			3.2	
	F			3.2	
	M			3.2	
	M			3.2	
	M			3.2	
	M			3.2	
	F			3.2	
	F			3	
	F			3	
	M			3	
	M			3	
	M			3	
(x)	F	L858R			11.6
(ix)	M				8.6
	M				4
	M				3.2
	M				3.2

C



>=3 - gains
>=4 - amplifications
*sample with T790M

Fig 2. *EGFR*, *MET* and *ERBB2* amplifications in NSCLC samples. A) Examples of *EGFR* (samples v-vi), *MET* (samples vii-viii) and *ERBB2* (samples ix-x) amplification detected by the EGFRmut+ assay. B) Characteristics of samples with *EGFR*, *MET* and *ERBB2* gains/amplifications. The samples shown in Fig. 1D (ii-iii) and Fig. 2A (vi-x) are indicated in the first column. The sex and *EGFR* mutation status of each sample are indicated in the second and the third columns, respectively. Columns 4–6 show the copy number values of *EGFR*, *MET* and *ERBB2*, respectively. The dark blue cells indicate amplifications (copy number $\geq$4), and the light blue cells indicate gains (copy number 3–4). C) Frequency of *EGFR* mutations in samples with *EGFR* amplification, *EGFR* gain and normal *EGFR* copy number.

doi:10.1371/journal.pone.0117983.g002

found. The sequencing analysis did not reveal any new sequence variants that could be oncogenic mutations or trigger *EGFR* amplification.

Replication of EGFR copy number analysis (amplification detection) with the use of qPCR and ddPCR

To confirm the results of *EGFR* copy number analysis we reanalyzed a panel of cancer samples, including all samples with *EGFR* amplification, with the use of qPCR and ddPCR. Both methods are based on completely different principles than MLPA. The qPCR is a method utilizing real-time quantification of PCR product, most commonly used for verification of copy number variants, identified in both normal (non-cancer) and cancer samples (e.g. [26,27]). The ddPCR is a new quantification method, based on absolute counting of tested and reference DNA molecules, clonally amplified in thousands of water-in-oil droplet-reactions (emulsion PCR). The

usefulness of ddPCR with the EvaGreen dsDNA-binding dye for precise copy-number estimation was recently demonstrated in several studies [18,19,28]. With the use of both qPCR and ddPCR, the signals of two test-amplicons, located in exon 2 and exon 18 of *EGFR*, were analyzed against the signal of control-amplicon (corresponding to MLPA probe ctrl_1). The analysis performed showed that results of both qPCR and ddPCR correlate very well with the relative copy number values, determined with the use of MLPA (correlation coefficient R = 0.941 and R = 0.927, respectively), and that the signals of samples with *EGFR* amplification were well separated from signals of samples without amplification (S1 Fig.). It has to be noted however that it cannot be excluded that some samples with borderline copy number values may be misclassified. Some discrepancies in the absolute signal values, determined by MLPA, and the reference methods may result from different sets of control regions used for normalization and from the fact that MLPA is a multiplex method and it measures the copy number in multiple points in a gene of interest. Similar results were obtained for *MET* and *ERBB2* (data not shown).

Discussion

It is well known that the frequency of *EGFR* mutations differs substantially between human populations. The mutation frequency is highest in Asians (45–52%); lower in Europeans (24%), African Americans (20%) and Hispanics (17%); and lowest in White Australians (7%) ([11,29] and references within). Also among Europeans, different frequencies of *EGFR* mutations were reported: e.g. 17% in Spain [30] and 10% in southern Germany [31]. In this study, based on a comprehensive analysis of a substantial number of lung adenocarcinoma samples, we characterized and determined the frequency of *EGFR* mutations in Polish patients (12.1%). Our analysis showed a much higher (4.8x) frequency of *EGFR* mutations in women than in men (24% vs. 5%, respectively). We observed a similar trend for *EGFR* gains/amplifications, which showed even greater overrepresentation (9x) in women than in men (18% vs. 2%, respectively). Although the higher frequency of *EGFR* mutations in women than in men has been observed many times before, the overrepresentation of *EGFR* mutations in women observed in our study is, according to our knowledge, one of the highest reported so far (excluding the studies with a very small number of samples/mutations) ([11,29] and the references within). The information about population-specific characteristic and frequency of mutations are important epidemiological factors that may influence implementation of adequate diagnostic and treatment procedures optimal for particular population.

Molecular cancer diagnosticians address diversified heterogeneous tumor material on a daily basis. DNA is isolated from surgically resected tumors (fresh or FFPE) and from fine needle and endobronchial ultrasound transbronchial needle-aspirated cells. Each of these histological or cytological samples can have different and sometimes very low PTC, and they often show a substantial level of DNA degradation (especially the FFPE samples). Additionally, formalin fixation may damage DNA and lead to sequence modifications. All of these factors cause different types of artifacts that may lead to both false-positive and false-negative results.

In this study, we tested an EGFRmut+ MLPA assay (a standard-sensitivity method) as a second-tier method and compared our results with those of a well-validated commercial RT-PCR assay (a high-sensitivity method). These two methods are based on completely different principles. In RT-PCR, mutation detection is based on the hybridization of TaqMan mutation-specific probes, but in MLPA, mutation detection is based on the ligation and subsequent amplification of wild type or mutation-specific probes. Our results indicate that EGFRmut+ is a robust assay that exhibits high reproducibility, specificity and sensitivity. A small number of undetected mutations were either not covered by the MLPA probes (n = 2) or occurred in a

sample with a low PTC (n = 1). The EGFRmut+ assay allowed us to detect mutations in all types of samples (cytological and histological; fresh frozen and FFPA) and allowed us to detect mutations in samples with different PTCs (20–90%). However, it must be noted that MS+ probes allow more robust mutation detection than MS- probes do and that the quality of MLPA results is generally lower for cancer samples than for germ-line DNA samples. The lower quality of cancer MLPA results is due to the aforementioned characteristics of cancer samples and has been reported and discussed previously [23–25,32]. Additional factors that make the MLPA assay attractive as a second-tier *EGFR* mutation detection method are the low amount of DNA (50–100 ng) required for analysis (no additional sample extraction or preparation is required, and the same DNA sample can be used for both RT-PCR and MLPA analysis), a short turn-around time (<2 days), and low cost (roughly $5 plus the initial cost of probe synthesis, approximately $3,000). In the case of routinely used assays, the cost of probe synthesis may be ignored because the quantity of the synthesized probes is sufficient for hundreds of thousands of analyses.

In addition to *EGFR* mutation detection, the EGFRmut+ assay also allows the detection of *EGFR*, *MET* and *ERBB2* amplification. Although the data are not conclusive, it has been suggested that *EGFR* amplification may be an indicator or modifier of sensitivity to TKI treatment (reviewed in [11]). It also has been suggested that the amplification of either *MET* or *ERBB2* may be an alternative mechanism underlying acquired resistance to TKI treatment (reviewed in [11]). In our study, we found that *EGFR*, *MET* and *ERBB2* gains and amplifications are mutually exclusive and that *MET* and *ERBB2* gains/amplifications rarely co-occur with *EGFR* mutations. Only one *EGFR* mutation co-occurs with an *ERBB2* amplification, and two *EGFR* mutations co-occur with a gain of *MET*. These data suggest a rather independent occurrence of *EGFR* mutations and *MET*/*ERBB2* amplifications and may argue against the role of *MET*/*ERBB2* amplification as a mechanism of acquired resistance to TKIs. However, to investigate this observation further, analysis of lung cancer samples after treatment with TKI should be performed. On the other hand, we observed a very strong correlation between *EGFR* gains/amplifications and the occurrence of *EGFR* mutations. All but one sample with an *EGFR* amplification also had an *EGFR* mutation. This suggests that activating *EGFR* mutations are the triggers of *EGFR* amplifications; cases of *EGFR* amplification in which no mutations were found may therefore be triggered by mutations that are not covered by standard *EGFR* tests. Such mutations may occur outside the tyrosine kinase domain and in non-coding regulatory sequences. Although the co-occurrence of *EGFR* mutations and amplifications has been observed before, it is typically much less pronounced than was observed in our study (e.g., the recent whole-genome analysis of 183 lung adenocarcinomas with the use of massively parallel sequencing [33]). These discrepancies may result from the different methods used for *EGFR* copy number determination.

In our laboratory, we routinely analyze somatic *EGFR* mutations in NSCLC samples from a large area of Poland (mostly central Poland) with a commercial RT-PCR test (EGFR-RT52). As it was suggested in recommendation for *EGFR* testing [11], we attempted to set up a second-tier method for the detection of somatic *EGFR* mutations. For this purpose, we selected an MLPA-based assay (EGFRmut+), and the results of this assay were validated against those of RT-PCR. Our study showed a very high concordance between the results of these two methods and thus confirmed both the robustness of mutation detection by MLPA assays and its usefulness as second-tier method for *EGFR* testing. Two additional advantages of the MLPA assay are its ability to carry out copy number (amplification) analysis and the potential to estimate the relative proportion of the mutated allele in a sample. A similar strategy of analysis and a similar MLPA-based assay may be developed and used for the analysis of other important oncogenes in other types of cancer.

Supporting Information

S1 Fig. Replication of MLPA results of *EGFR* copy number analysis with the use of qPCR and ddPCR. Scatter plots showing correlation between relative copy number signals determined by MLPA (x-axis) and relative signals determined either by qPCR (A), or ddPCR (B) (y-axis). The values depicted on y-axis represent averaged signals, measured in exon 2 and 18 of *EGFR* (see Materials and Methods). The blue, light-blue and white dots indicate samples with *EGFR* amplification (N = 7), gain (N = 3) and normal copy number (N = 6), respectively. The trend line and correlation coefficient are indicated on each graph.
(PDF)

S1 Table. Mutations and Copy Number Alterations Detected in NSCLC Samples.
(XLS)

Author Contributions

Conceived and designed the experiments: MAL PK. Performed the experiments: MAL KC KK. Analyzed the data: MAL KC KK WJ JK PK. Contributed reagents/materials/analysis tools: WJ JK. Wrote the paper: MAL PK.

References

1. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, et al. (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304: 1497–1500. PMID: 15118125
2. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, et al. (2007) Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448: 561–566. PMID: 17625570
3. Gazdar AF (2009) Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. Oncogene 28 Suppl 1: S24–31. doi: 10.1038/onc.2009.198 PMID: 19680293
4. Ercan D, Zejnullahu K, Yonesaka K, Xiao Y, Capelletti M, et al. (2010) Amplification of EGFR T790M causes resistance to an irreversible EGFR inhibitor. Oncogene 29: 2346–2356. doi: 10.1038/onc.2009.526 PMID: 20118985
5. Kobayashi S, Boggon TJ, Dayaram T, Janne PA, Kocher O, et al. (2005) EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. N Engl J Med 352: 786–792. PMID: 15728811
6. Marchetti A, Martella C, Felicioni L, Barassi F, Salvatore S, et al. (2005) EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 23: 857–865. PMID: 15681531
7. Pennycuick A, Simpson T, Crawley D, Lal R, Santis G, et al. (2012) Routine EGFR and KRAS Mutation analysis using COLD-PCR in non-small cell lung cancer. Int J Clin Pract 66: 748–752. doi: 10.1111/j.1742-1241.2012.02961.x PMID: 22805266
8. Kawahara A, Taira T, Azuma K, Tominaga M, Hattori S, et al. (2012) A diagnostic algorithm using EGFR mutation-specific antibodies for rapid response EGFR-TKI treatment in patients with non-small cell lung cancer. Lung Cancer 78: 39–44. doi: 10.1016/j.lungcan.2012.07.002 PMID: 22858448
9. Lewandowska MA, Jozwicki W, Jochymski C, Kowalewski J (2013) Application of PCR methods to evaluate EGFR, KRAS and BRAF mutations in a small number of tumor cells in cytological material from lung cancer patients. Oncol Rep 30: 1045–1052. doi: 10.3892/or.2013.2579 PMID: 23817662
10. Tuononen K, Maki-Nevala S, Sarhadi VK, Wirtanen A, Ronty M, et al. (2013) Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-embedded tumor material of non-small cell lung carcinoma-superiority of NGS. Genes Chromosomes Cancer 52: 503–511. doi: 10.1002/gcc.22047 PMID: 23362162
11. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, et al. (2013) Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol 8: 823–859. doi: 10.1097/JTO.0b013e318290868f PMID: 23552377

12. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, et al. (2013) Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. Arch Pathol Lab Med 137: 828–860. doi: 10.5858/arpa.2012-0720-OA PMID: 23551194

13. Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, et al. (2013) Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn 15: 415–453. doi: 10.1016/j.jmoldx.2013.03.001 PMID: 23562183

14. Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P (2010) Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. ScientificWorldJournal 10: 2003–2018. doi: 10.1100/tsw.2010.195 PMID: 20953551

15. Lewandowska MA, Jozwicki W, Zurawski B (2013) KRAS and BRAF mutation analysis in colorectal adenocarcinoma specimens with a low percentage of tumor cells. Mol Diagn Ther 17: 193–203. doi: 10.1007/s40291-013-0025-8 PMID: 23606169

16. Kozlowski P, Roberts P, Dabora S, Franz D, Bissler J, et al. (2007) Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Hum Genet 121: 389–400. PMID: 17287951

17. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, et al. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30: e57. PMID: 12060695

18. McDermott GP, Do D, Litterst CM, Maar D, Hindson CM, et al. (2013) Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR. Anal Chem 85: 11619–11627. doi: 10.1021/ac403061n PMID: 24180464

19. Miotke L, Lau BT, Rumma RT, Ji HP (2014) High sensitivity detection and quantitation of DNA copy number and single nucleotide variants with single color droplet digital PCR. Anal Chem 86: 2618–2624. doi: 10.1021/ac403843j PMID: 24483992

20. Dopierala J, Damato BE, Lake SL, Taktak AF, Coupland SE (2010) Genetic heterogeneity in uveal melanoma assessed by multiplex ligation-dependent probe amplification. Invest Ophthalmol Vis Sci 51: 4898–4905. doi: 10.1167/iovs.09-5004 PMID: 20484589

21. Ji H, Ramsey MR, Hayes DN, Fan C, McNamara K, et al. (2007) LKB1 modulates lung cancer differentiation and metastasis. Nature 448: 807–810. PMID: 17676035

22. Liang MC, Ma J, Chen L, Kozlowski P, Qin W, et al. (2010) TSC1 loss synergizes with KRAS activation in lung cancer development in the mouse and confers rapamycin sensitivity. Oncogene 29: 1588–1597. doi: 10.1038/onc.2009.452 PMID: 19966866

23. Jeuken J, Cornelissen S, Boots-Sprenger S, Gijsen S, Wesseling P (2006) Multiplex ligation-dependent probe amplification: a diagnostic tool for simultaneous identification of different genetic markers in glial tumors. J Mol Diagn 8: 433–443. PMID: 16931583

24. Jeuken JW, Cornelissen SJ, Vriezen M, Dekkers MM, Errami A, et al. (2007) MS-MLPA: an attractive alternative laboratory assay for robust, reliable, and semiquantitative detection of MGMT promoter hypermethylation in gliomas. Lab Invest 87: 1055–1065. PMID: 17700563

25. van Dijk MC, Rombout PD, Boots-Sprenger SH, Straatman H, Bernsen MR, et al. (2005) Multiplex ligation-dependent probe amplification for the detection of chromosomal gains and losses in formalin-fixed tissue. Diagn Mol Pathol 14: 9–16. PMID: 15714058

26. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, et al. (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307: 1434–1440. PMID: 15637236

27. Scott KL, Kabbarah O, Liang MC, Ivanova E, Anagnostou V, et al. (2009) GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. Nature 459: 1085–1090. doi: 10.1038/nature08109 PMID: 19553991

28. Hwang VJ, Maar D, Regan J, Angkustsiri K, Simon TJ, et al. (2014) Mapping the deletion endpoints in individuals with 22q11.2 Deletion Syndrome by droplet digital PCR. BMC Med Genet 15: 106. doi: 10.1186/s12881-014-0106-5 PMID: 25312060

29. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, et al. (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97: 339–346. PMID: 15741570

30. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, et al. (2009) Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 361: 958–967. doi: 10.1056/NEJMoa0904554 PMID: 19692684

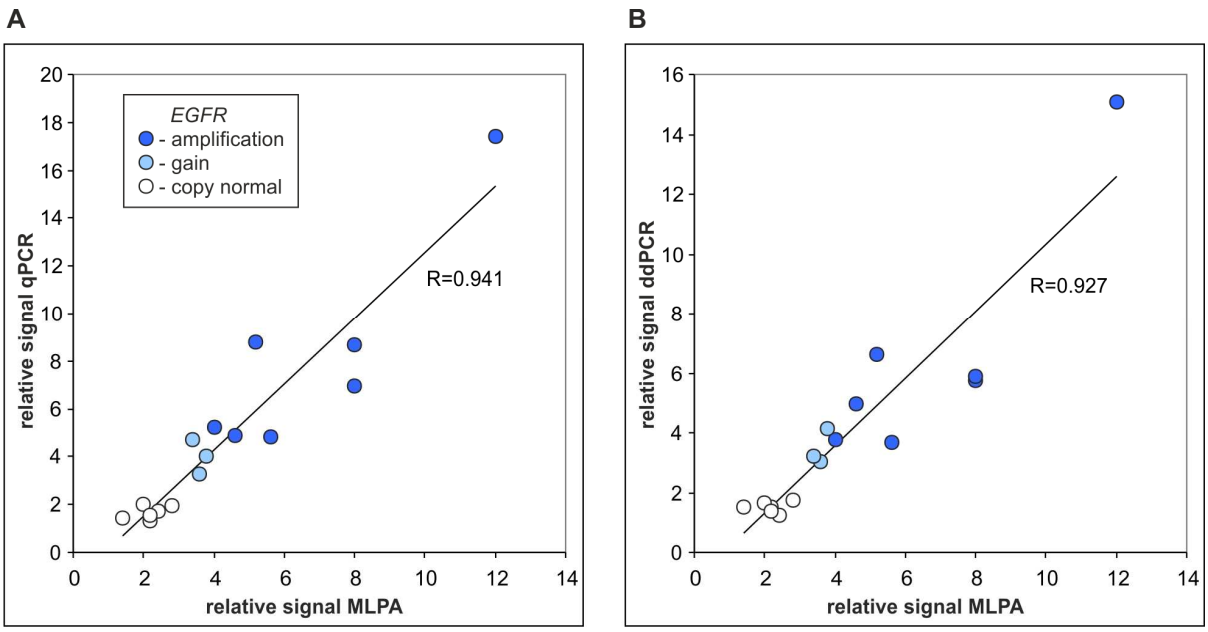
31. Gahr S, Stoehr R, Geissinger E, Ficker JH, Brueckl WM, et al. (2013) EGFR mutational status in a large series of Caucasian European NSCLC patients: data from daily practice. Br J Cancer 109: 1821–1828. doi: 10.1038/bjc.2013.511 PMID: 24002608

32. Kozlowski P, Jasinska AJ, Kwiatkowski DJ (2008) New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis 29: 4627–4636. doi: 10.1002/elps.200800126 PMID: 19053154

33. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150: 1107–1120. doi: 10.1016/j.cell.2012.08.029 PMID: 22980975

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Lewandowska M.A., Czubak K. i wsp., PLoS One 2015



S1 Figure. Replication of MLPA results of *EGFR* copy number analysis with the use of qPCR and ddPCR. Scatter plots showing correlation between relative copy number signals determined by MLPA (x-axis) and relative signals determined either by qPCR (A), or ddPCR (B) (y-axis). The values depicted on y-axis represent averaged signals, measured in exon 2 and 18 of *EGFR* (see Materials and Methods). The blue, light-blue and white dots indicate samples with *EGFR* amplification (N=7), gain (N=3) and normal copy number (N=6), respectively. The trend line and correlation coefficient are indicated on each graph. (PDF)

Supplementary Table S1. Mutations and Copy Number Alterations Detected in NSCLC Samples

SAMPLE NUMBER	SAMPLES shown in Figures	SEX	AGE at diagnosis	sampleTYPE	PTC	EGFR mutations; RT-PCR	EGFR mutations; MLPA	EGFR gain/amp.	MET gain/amp.	ERBB2 gain/amp.
1		M	68	FFPE						
2		M	58	FFPE						
3		W	69	FFPE						
4		M	70	FFPE						
5		M	63	FFPE						
6		M	66	FFPE						
7		M	67	FFPE						
8		M	60	FFPE					4.8	
9		W	74	FFPE		L858R	L858R	3.0		
10		M	58	FFPE						
11		M	40	FFPE					3.0	
12		W	76	FFPE					3.1	
13		M	65	FFPE						
14		M	53	FFPE						
15		M	78	FFPE						
16		M	76	FFPE						
17		M	58	FFPE						
18		M	56	FFPE						
19		M	76	FFPE		L858R	L858R			
20		M	54	FFPE						
21		W	52	FFPE	15%				4.0	
22		M	81	FFPE	50%					
23		M	59	cyto						
24		W	59	FFPE	30%				3.2	
25		M	61	FFPE	40%					
26		M	60	FFPE						
27		M	64	FFPE	10%					
28		W	62	FFPE						
29		W	55	FFPE						
30		W	60	cyto						
31		W	63	FFPE	60%			3.1		
32		M	53	FFPE	50%					
33		W	56	cyto						
34		W	54	FFPE						
35		M	70	FFPE	10%					
36		M	59	FFPE					4.1	
37		W	67	FFPE	30%					
38		W	48	FFPE	70%					
39		M	60	cyto						
40	(iii)	W	53	cyto		T790M;L858R	T790M;L858R	3.2		
41		M	63	FFPE	40%					
42		M	57	FFPE	30%					
43		W	49	FFPE	60%				3.2	
44		W	67	FFPE		L858R	L858R		3.6	
45		M	68	FFPE	40%	e19del	e19del			
46		W	53	FFPE						
47		W	55	FFPE	30%					
48		W	50	FFPE	80%					
49		W	61	FFPE		L861Q				
50		W	55	FFPE	20%	e19del	e19del15	3.4		
51		W	52	FFPE		e19del	e19del			
52		M	64	FFPE						
53		W	69	FFPE	10%				3.6	
54	(viii)	M	64	cyto	80%				15.4	
55		M	47	FFPE	20%				3.3	
56		W	52	FFPE	20%			5.6		
57		M	61	FFPE	35%				4.8	
58		M	53	cyto						
59		M	76	FFPE	30%					
60		M	72	FFPE	20%					
61		M	69	FFPE					4.2	
62		W	60	FFPE	70%					
63		M	74	FFPE	65%					
64		M	55	FFPE	40%					
65		W	35	cyto		e19del	e19del			
66	(x)	W	66	cyto	95%	L858R	L858R			11.6
67		M	67	FFPE	50%				4.3	
68		M	65	cyto	50%					
69		M	56	FFPE	30%					
70		M	65	FFPE	50%					
71		M	65	cyto	50%					
72		W	58	FFPE	20%	e19del	e19del15		3.2	
73		W	49	FFPE	70%	e19del	e19del15	4.6		
74		M	56	FFPE	70%					
75		M	73	FFPE	40%					
76		W	61	FFPE	35%					
77		W	56	cyto	40%					
78		W	75	FFPE	30%					
79		M	68	FFPE	85%					
80		M	66	FFPE	30%				3.2	
81		W	73	cyto	60%					

82		W	49	FFPE	55%					
83		W	57	cyto						
84		W	71	FFPE	50%					
85		M	63	cyto	80%					
86	(iv)	W	66	cyto	80%	e19del	e19del15	8.0		
87		W	59	FFPE	30%	e19del	e19del	3.4		
88		W	65	FFPE	50%					
89		W	58	FFPE	10%	L861Q				
90		W	36	FFPE	50%				3.2	
91		M	55	cyto	50%					
92		W	57	cyto	30%					
93		M	66	FFPE	60%					
94		W	78	FFPE	55%	e19del	e19del	3.4		
95		M	51	cyto	50%					
96		M	58	FFPE	20%					
97		W	54	FFPE	40%					
98		M	56	FFPE	60%					
99		W	63	FFPE						
100		M	48	FFPE	70%					
101		M	64	FFPE	15%					
102		M	42	FFPE	70%					
103		W	64	cyto	90%					
104		M	58	FFPE	10%					
105		M	68	cyto	70%					
106		M	59	cyto	60%					
107		M	54	FFPE	70%					
108		W	64	FFPE	60%					
109		M	62	FFPE	40%					
110		W	67	FFPE	15%	G719X				
111		M	62	cyto	80%					
112		M	44	cyto	80%				3.4	
113	(ix)	M	59	FFPE	50%					8.6
114		M	66	FFPE	50%					
115		W	38	cyto	55%			3.3		
116		W	54	FFPE	25%	L858R	L858R	4.0		
117		M	37	cyto	80%					
118		M	60	FFPE	50%					
119		M	73	FFPE	30%					
120		M	66	FFPE	50%					
121		M	71	cyto	40%					
122		W	65	cyto	80%	e19del	e19del15	5.2		
123		M	55	cyto	65%					
124		M	73	FFPE	65%				3.2	
125		M	68	FFPE	10%				3.4	
126		M	61	FFPE	15%					
127		W	48	FFPE	15%					
128		M	49	FFPE	15%					
129		M	70	cyto	70%				3.2	
130		M	52	cyto	60%					
131		M	65	cyto	90%					
132		M	64	cyto	80%					3.2
133		M	57	cyto	60%					
134		M	58	cyto	70%					
135		M	57	cyto	80%					
136		W	42	cyto	80%					
137	(v)	W	62	cyto	50%	e19del	e19del15	8.0		
138		M	66	cyto	90%					
139		W	63	cyto	60%					
140		M	79	cyto	60%					
141		M	78	cyto	90%					
142		W	57	cyto	80%					
143		W	52	cyto	60%					
144		W	80	FFPE	15%					
145		M	59	cyto	70%					
146		M	64	cyto	60%				3.2	
147		M	59	cyto	60%					
148		W	77	cyto	60%					
149		M		FFPE	40%					
150		M	72	FFPE	65%					
151		M	60	cyto	90%					
152		M	59	FFPE	2%					
153		M	70	FFPE	0%					
154		M	64	cyto	70%					
155		M	69	cyto	80%					
156		M	59	cyto	70%					
157		M	56	FFPE	35%					
158		M	59	FFPE	40%					
159		M	58	FFPE	60%					
160		M	54	FFPE	30%					
161		M	74	cyto	80%					
162		M	76	cyto	60%					
163		M	52	cyto	90%	e20ins	e20ins			
164		M	55	cyto	90%					
165		M	66	FFPE	30%					
166		M	56	FFPE	25%					
167		W	74	FFPE	50%					

168		M	61	FFPE	25%					
169		W	64	FFPE	75%				3.2	
170		W	77	FFPE	25%					
171		W	65	cyto	85%					
172		M	54	FFPE	70%					
173		M	67	cyto	40%					3.2
174		M	59	FFPE	50%					
175		W	36	FFPE	20%					
176		W	62	FFPE	10%					
177		W	47	cyto	80%					
178		W	59	FFPE	80%					
179		M	56	FFPE	85%					
180		M	73	cyto	80%				3.4	
181		M	60	FFPE	25%					
182		M	56	cyto	90%					
183		M	75	cyto	80%					
184		M	61	FFPE	25%					
185		W	69	FFPE	20%					
186		W	76	cyto	60%					
187	(vi)	W	62	cyto	90%	L858R	L858R	12.0		
188		M	51	FFPE	25%					
189		M	60	cyto	80%					
190		M	65	FFPE	40%					
191		M	66	cyto	80%					
192	(ii)	M	77	cyto	80%	e19del	e19del15			
193		M	59	FFPE	30%					
194		M	66	cyto	80%					
195		W	59	cyto	80%					
196		W	58	FFPE	60%	L858R	L858R			
197		M	66	cyto	80%					
198		M	65	cyto	70%			3.4		
199		M	53	FFPE	70%					
200		W	51	FFPE	20%					
201		M	66	cyto	80%					
202		M	71	FFPE	30%	e19del	e19del15			
203		W	77	cyto	80%				5.6	
204		M	58	FFPE	60%					
205		W	61	cyto	10%			3.6		
206		W	56	FFPE	45%					
207		M	60	cyto	5%			3.2		
208		M	72	cyto	90%					
209		W	71	cyto	90%	e19del	e19del15			
210		M	58	cyto	80%					
211		M	68	cyto	80%					
212		M	76	FFPE	70%					
213		W	78	cyto	90%				4.4	
214		M	62	FFPE	20%	e19del	e19del			
215		M	64	FFPE	25%					
216		W	59	cyto	50%					
217		M	48	cyto	80%					
218		M	68	cyto	80%					
219	(vii)	M	79	FFPE	30%				7.6	
220		W	55	FFPE	20%					
221		M	66	FFPE	35%					
222		M	65	FFPE	20%					4.0
223		M	70	cyto	70%	e19del	e19del15	3.4		
224		M	60	cyto	60%					
225		W	76	FFPE	20%					
226		W	59	cyto	80%					
227		W	64	cyto	80%	L858R	L858R	3.8		
228		M	72	cyto	70%					
229		W	65	cyto	80%					
230		M	44	FFPE	55%					
231		W	53	cyto	60%					
232		W	70	cyto	50%					
233		M	47	cyto	60%					
234		M	60	cyto	60%					
235		W	72	cyto	60%					
236		W	66	FFPE	30%					
237		M	56	FFPE	30%					
238		W	59	cyto	80%				5.6	
239		W	55	cyto	80%					

LEGEND
PTC - percentage of cancer cells
e19del - one of in-frame deletions in exon 19
e19del15 - the most common in-frame deletion in exon 19, c.2235_2249del15
e20ins - one of in-frame insertions in exon 20
M - man
W - woman

2

Czubak K., Lewandowska M. A., Klonowska K., Roszkowski K., Kowalewski J., Figlerowicz M., Kozłowski P.

„High copy number variation of cancer-related microRNA genes and frequent amplification of *DICER1* and *DROSHA* in lung cancer"

Oncotarget 2015, 6:23399-416

High copy number variation of cancer-related microRNA genes and frequent amplification of *DICER1* and *DROSHA* in lung cancer

Karol Czubak[1], Marzena Anna Lewandowska[2,3], Katarzyna Klonowska[1], Krzysztof Roszkowski[4,5], Janusz Kowalewski[3,6], Marek Figlerowicz[1], Piotr Kozlowski[1]

[1]European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

[2]Molecular Oncology and Genetics Department, Innovative Medical Forum, The Franciszek Lukaszczyk Oncology Center, Bydgoszcz, Poland

[3]Department of Thoracic Surgery and Tumors, Nicolaus Copernicus University, Torun, Collegium Medicum, Bydgoszcz, Poland

[4]Department of Radiotherapy, The Franciszek Lukaszczyk Oncology Center, Bydgoszcz, Poland

[5]Department of Oncology, Radiotherapy and Gynecologic Oncology, Nicolaus Copernicus University, Torun, Collegium Medicum, Bydgoszcz, Poland

[6]Department of Thoracic Surgery and Tumors, The Franciszek Lukaszczyk Oncology Center, Bydgoszcz, Poland

Correspondence to:
Piotr Kozlowski, **e-mail:** kozlowp@yahoo.com; kozlowp@ibch.poznan.pl

Keywords: *microRNA, DROSHA, DICER1, non-small cell lung cancer NSCLC, MLPA*

Received: *May 04, 2015* ***Accepted:*** *June 08, 2015* ***Published:*** *June 22, 2015*

ABSTRACT

A growing body of evidence indicates that miRNAs may be a class of genetic elements that can either drive or suppress oncogenesis. In this study we analyzed the somatic copy number variation of 14 miRNA genes frequently found to be either over- or underexpressed in lung cancer, as well as two miRNA biogenesis genes, *DICER1* and *DROSHA*, in non-small-cell lung cancer (NSCLC). Our analysis showed that most analyzed miRNA genes undergo substantial copy number alteration in lung cancer. The most frequently amplified miRNA genes include the following: *miR-30d*, *miR-21*, *miR-17* and *miR-155*. We also showed that both *DICER1* and *DROSHA* are frequently amplified in NSCLC. The copy number variation of *DICER1* and *DROSHA* correlates well with their expression and survival of NSCLC and other cancer patients. The increased expression of *DROSHA* and *DICER1* decreases and increases the survival, respectively. In conclusion, our results show that copy number variation may be an important mechanism of upregulation/downregulation of miRNAs in cancer and suggest an oncogenic role for *DROSHA*.

INTRODUCTION

Cancer initiation and development are associated with the accumulation of numerous genetic alterations in the cancer genome. These alterations include both small-size mutations and large-scale genomic alterations consisting of copy number variants (CNVs - deletions, duplications or amplifications), as well as copy-number-neutral genomic rearrangements (inversions or translocations). Interactions between these alterations (in certain situations, in addition to germline mutations) allow cancer to clonally evolve due to deactivation of tumor suppressor genes (loss-of-function mutations) and activation of oncogenes (gain-of-function mutations).

Lung cancer is the leading cause of cancer-related death (http://www.who.int/mediacentre/factsheets/fs297/en/; [1]). There are several subtypes of lung cancer, the most common of which is non-small-cell lung cancer (NSCLC). NSCLC can be further divided into adenocarcinoma, squamous-cell carcinoma, and large-cell carcinoma. Lung cancer occurs predominantly in smokers (>60%). Regardless of histological and

risk-factor divisions, lung cancer is one of the most genomically heterogeneous type of cancer. Recently, several whole-genome sequencing projects utilizing next-generation sequencing technologies revealed the presence of thousands of small-size mutations in the individual lung cancer genome [2–5], with an almost 10 times higher frequency of mutations in smoker than in non-smoker samples [6]. An even higher level of variation seems to be attributed to copy number alterations. It was shown with the use of SNP-array-based analysis that approximately 50% of the lung cancer genome undergoes recurrent copy number alterations [7]. On average, over 40% of the genome undergoes copy number alteration in individual lung cancers [8]. However, only a small fraction of alterations occurring in cancer genomes are functional ("driver") mutations; others are "passenger" mutations that occur as a consequence of the general cancer genome destabilization. Although "passenger" mutations are not critical for cancer genome evolution, they are often selected in parallel with closely located or commonly regulated targets of "driver" mutations. The role of "passenger" mutations for particular cancers is mostly unknown (it is not necessarily neutral).

A substantial progress in lung cancer treatment (especially adenocarcinomas) has been made recently due to personalized therapy based on genomic biomarkers. The distinctive biomarkers in lung cancer are mutations in the *epidermal growth factor receptor* (*EGFR*) [9] or gain-of-function translocations and inversions involving the *anaplastic lymphoma receptor tyrosine kinase* (*ALK*) [10]. However, the general prognosis of lung cancer is still poor and its 5-year survival is one of the lowest among cancer patients at approximately 10%. Therefore, many lung cancer studies are currently focused on understanding the impact of genetic alterations on cancer biology and development and on the identification of new prognostic biomarkers.

Among the most intensively studied candidate biomarkers are microRNAs (miRNAs), a class of short (~21 nt long), single-stranded, noncoding RNAs. MiRNAs are primarily involved in the post-transcriptional regulation of gene expression, either by mRNA degradation or inhibition of translation efficiency [11, 12]. Mature miRNAs are generated in two subsequent steps from long primary precursors (pri-miRNAs). Pri-miRNAs are encoded either by independent transcriptional units or by protein-coding genes. In the first step of miRNA biogenesis that takes place in the nucleus, the secondary precursor (~60 nt long pre-miRNA), which adopts a hairpin structure, is cleaved out from pri-miRNA by the nuclease DROSHA. Upon export to the cytoplasm, the pre-miRNA is further processed into a miRNA-duplex by the nuclease DICER. One of the miRNA-duplex strands is released, and the other becomes the mature miRNA that, as a key element of the miRNA-induced silencing complex (miRISC) recognizes complementary target sequences usually located within the 3' untranslated regions of mRNAs.

The biological functions of most miRNAs identified so far (miRBase; http://www.mirbase.org; [13, 14] remain unknown. However, it has been well documented that miRNAs downregulate numerous genes and either stimulate or inhibit many important biological processes and diseases, including cell proliferation and differentiation, apoptosis, development and cancer [15–18].

The role of miRNAs in the development of cancer was first identified in chronic lymphocytic leukemia in 2002 [19]. Since then, it has been shown that overexpression or downregulation of certain miRNAs contributes to the development, progression and metastasis of many types of cancer. Such miRNAs can therefore be classified as either oncogenes (oncomirs) or tumor suppressors [20]. It has also been shown that some miRNAs, such as *miR-21*, *miR-205* or *miR-155*, seem to be universal for different cancers [12].

There have been numerous studies of miRNA expression in lung cancer, and many miRNAs that are specifically over- or underexpressed in lung cancer or in particular lung cancer subtypes were identified. For example, it was shown that 6 miRNAs constituting the polycistronic miRNA cluster, *miR-17/92*, are overexpressed in lung cancer and enhance cell proliferation [21]. It was later shown that an elevated level of these miRNAs may be detected in the plasma of lung cancer patients [22, 23] and is associated with poor disease prognosis [24]. Other miRNAs consistently found to be either overexpressed or underexpressed in lung cancer are *miR-21*, *miR-210* and *miR-126*. However, it should be noted that substantial discordances between miRNA profiling results also exist.

Although the functional relevance of some of the miRNAs that are differentially expressed in lung cancer has been demonstrated (e.g., [25–27]), the roles of most of these miRNAs in cancer are unknown or poorly recognized. One factor that may shed more light on the role of particular miRNAs in cancer is the mechanism underlying their aberrant expression. Among the most pronounced mechanisms underlying aberrant expression in cancer are point mutations, epigenetic modifications and copy number alterations. However, it has been suggested that point mutations and epigenetic modifications are not important factors in the global miRNA regulation in lung cancer [24, 28]. It has also been shown that miRNA genes are overrepresented and cluster in genomically fragile sites and other regions that undergo frequent copy number changes in cancer genomes. Thus, it has been suggested that somatic copy number variation may lead to the activation/deactivation of miRNAs in cancer [29, 30]. For example, systematic analysis of three cancer types (ovarian, breast, and melanoma) with the use of comparative genome hybridization microarrays showed

that 37% (ovarian) to 89% (melanoma) of analyzed miRNA genes undergo copy number changes [30]. There are known examples of both miRNA- and protein-coding genes whose expression in cancer is either increased or decreased by their copy number variation. These high-copy-number amplifications and recurrent deletions (loss of heterozygosity) are often used as a confirmation of oncogenic and tumor-suppressive function of the analyzed gene, respectively. The role of copy number variations in the regulation of miRNAs in cancer and the potential cancer-related implications have been reviewed before [31–33]. The most recent review provides an excellent summary and discusses this notion using ovarian cancer as an example [33].

In this study, with the use of homemade multiplex ligation-dependent probe amplification (MLPA) assays, we analyzed the somatic copy number variation of 14 miRNA genes consistently found to be either over or underexpressed in lung cancer. Additionally, we analyzed the copy number variation of *DICER1* and *DROSHA*, two main miRNA biogenesis genes. We analyzed these genes in 254 NSCLC samples and observed high copy number variation in most of the analyzed genes. Among the frequently amplified miRNA genes were *miR-21*, *miR-17/92* and *miR-155*, which are commonly recognized as oncomirs, as well as *miR-30a* and *miR-30d* which were downregulated in lung cancer. Surprisingly, a high average copy number value and frequent amplifications were present in both miRNA biogenesis genes. We also showed that amplification of *DROSHA* is not driven by other closely located oncogenes. The most frequently deleted miRNA gene turned out to be *miR-126*, which is commonly found to be downregulated in lung cancer. Our analysis showed that a substantial fraction of differentially expressed miRNAs may be explained by and are consistent with the copy number variation of their genes.

RESULTS

Selection of miRNA genes for copy number analysis in lung cancer

To select miRNA genes for our analysis, we took advantage of two recently published meta-analysis studies [34, 35] summarizing the results of dozens of whole-genome miRNA expression studies in lung cancer (references within [34, 35]). Although these two studies utilized completely different strategies of meta-analyses, the top significantly up- and downregulated miRNAs identified in both studies overlap perfectly (with minor differences in the order of identified miRNAs). Based on these meta-analyses, we selected 6 genes/genomic regions (*miR-21*, *miR-210*, *miR-182*, *mir-31*, *mir-200b*, *mir-205*) encoding miRNAs most consistently identified as upregulated, and 6 genes (*miR-126*, *miR-30a*, *miR-30d*, *miR-486*, *miR-451a*, *miR-143*) encoding miRNAs most consistently identified as downregulated in lung cancer. Additionally, for our analysis we selected the genomic regions of *miR-155* and *miR-17* (identified in one meta-analysis), which were consistently associated with poor prognosis of lung cancer, as well as two genes (*DICER1* and *DROSHA*) encoding miRNA processing enzymes. The genomic positions of all selected genes are indicated in Figure 1, and the criteria for their selection are summarized in Table 1. Note that some of the selected miRNA genes encompass miRNA clusters (e.g., *miR-17/92* and *miR-143/145*).

MLPA assays design

To analyze the somatic copy number variation of selected genomic regions, we designed two MLPA assays, each covering 7 miRNA genes and 1 miRNA biogenesis gene. Each miRNA or miRNA cluster region was covered by two MLPA probes located in close proximity (mostly within 1 kb) to an annotated pre-miRNA sequence, preferentially on both sides of the pre-miRNA sequence. Each of the miRNA biogenesis genes (*DICER1* and *DROSHA*) was covered by 3 MLPA probes located in exons distributed about equally across the genes. Additionally, each MLPA probe-set contained 4 control probes specific for different chromosomes. The exact genomic location and sequence of each probe is indicated in Supplementary Table S1. MLPA assays were designed and generated according to a strategy developed and have been described in detail previously [36, 37]. We validated the performance of the assays with the panel of reference non-cancer DNA samples and showed that all covered genomic regions are genetically stable and always occur in 2 copies.

Analysis of the somatic copy number variation of selected miRNA genes

With the use of the developed MLPA assay, we analyzed 254 NSCLC samples and determined the relative copy number value of all analyzed regions in these samples. As shown in Figure 2, the signals of probes representing particular regions in most cases are strongly synchronized. If one probe in a particular region indicates a copy number increase, the other probe or probes in these regions also show similar levels of copy number increase. As each MLPA probe recognizes different target sequence, such a correlation provides independent validation of the obtained results. The copy number value of a particular region was calculated as the average of the copy number values of the respective probes. The regions for which inter-probe variation was too high were considered uninterpretable and were excluded from further analysis. The relative copy number values of all analyzed regions are shown in Supplementary Table S2 and graphically summarized in Figure 3. As analyzed NSCLC samples are contaminated with different amounts of normal DNA (in most samples percentage of tumor cells (PTC) is >50%,

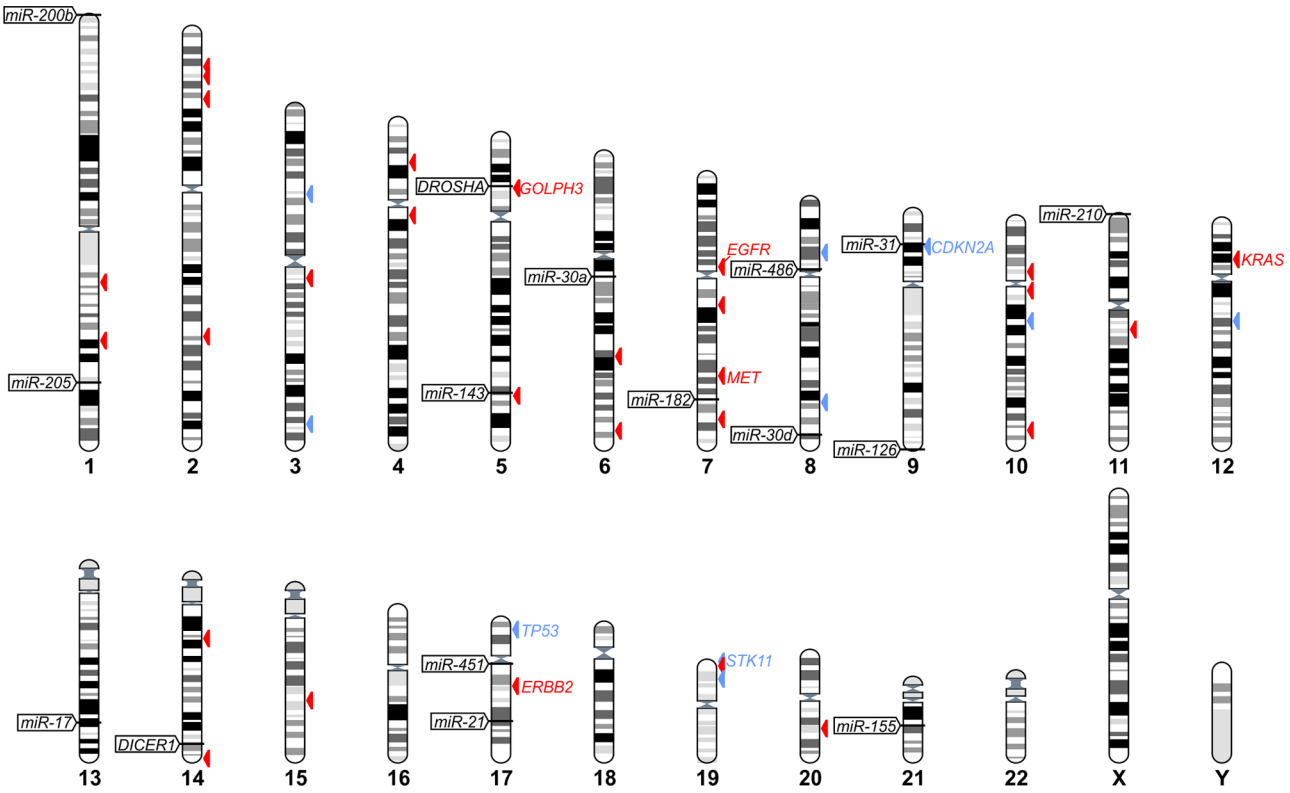


Figure 1: The positions of selected miRNA and miRNA biogenesis genes in human genome. The positions of miRNA and miRNA biogenesis genes are indicated on chromosome ideograms (left-hand side). Arrowheads on the right-hand side of the ideograms indicate lung cancer relevant genes catalogued in COSMIC: the Cancer Gene Census (associated with one of the following terms: "lung", "NSCLC" or "multiple tumor types"), the most reliable list of cancer-related genes annotated and curated by the Wellcome Trust Sanger Institute (originally published in [78]). Additionally, the position of *GOLPH3*, which is discussed in this study, is indicated. Red and blue arrowheads indicate oncogenes and tumor suppressor genes, respectively. IDs of the most relevant genes are indicated next to the arrowheads. The figure was prepared with the use of the "Ensembl karyotypes" tool available on the Ensembl portal.

and an average PTC is approximately 70%) the estimated copy number changes are generally diluted and lower than in actual cancer cells. For comparison, copy number values corrected for PTC (dilution) factor are shown in Supplementary Figure S1. As shown in Figure 3 and Supplementary Figure S1, the average copy number of analyzed regions differs substantially and is highest for *DROSHA*, *miR-30d*, *miR-30a*, *miR-21*, *DICER1*, *miR-205*, *miR-17*, and *miR-155* and lowest for the *miR-126* region (Table 2).

Pronounced copy number changes may be more indicative of the role of a particular region in cancer. Therefore, based on criteria similar to those applied before [38, 39], we classified the identified copy number changes to the following categories (from highest to lowest copy number): amplifications (≥4 copies; ≥2x increase), gains (≥3 copies; ≥1.5x increase), losses (≤1.33 copies; ≤1.5x decrease) and homozygous deletions (≤1 copy; ≤2x decrease). The categorized copy number changes of all analyzed samples are visualized in a heatmap graph (Figure 3B) and are summarized in Table 2. The results indicate that the number of amplifications detected in particular genes generally correlates with an increase in the average copy number value of these genes. The highest frequency of amplifications was observed in *miR-30a*, *miR-30d*, *miR-21*, *miR-17*, *DROSHA*, *DICER1*, and *miR-155*. In some of these genes both amplifications and isolated cases of deletions were detected. The genes for *miR-182*, *miR-200b* and *miR-210* turned out to be relatively stable, showing neither amplifications nor homozygous deletions. Only a few homozygous deletions but no amplifications were detected in *miR-126* and *miR-451a*. For comparison, the results of copy number changes of genes analyzed in this study are presented along with corresponding results of two oncogenes, *EGFR* and *MET*, obtained previously with the use of a similar methodology [40].

Extended analysis of the *DROSHA* locus on chromosome 5

One of the genes with the highest average copy number and the highest frequency of amplifications

Table 1. List of miRNA and miRNA biogenesis genes selected for analysis

analyzed loci IDs		miRs in cluster	expression change top-ranked in meta-analysis		other lung cancer relevant features		
			Vosa et al.[35] [corrected *p*-value]	Guan et al.[34] [mean fold change]	association with poor prognosis	potential biomarkers[B]	frequently deregulated in other solid tumors
miRNA and miRNA-cluster genes	*miR-21*		↑ 2E-14	↑ 4.4	+ [24, 80, 81]	+ [23, 82, 83]	+ [12, 84]
	miR-210		↑ 6E-11	↑ 2.7		+ [23]	
	miR-182	*182*[E], *183*[E], *96*	↑ 3E-8, 4E-2	↑ 6.3, 5.9		+ [23, 83, 85, 86]	+ [84]
	miR-31		↑ 1E-4	↑ 2.89			+ [84]
	miR-200b	*200b*[E], *200a*, *429*	↑ 1E-3	↑ 3.7		+ [82, 83]	+ [84]
	miR-205		↑ 7E-3	↑ 23.2		+ [83]	
	miR-126		↓ 7E-12	↓ .33		+ [23, 86]	+ [56]
	miR-30a	*30a*[E], *30b*	↓ 1E-9	↓ .36			
	miR-30d		↓ 2E-8	↓ .34			
	miR-486		↓ 4E-7	↓ .39		+ [23, 82]	
	miR-451a	*451a*[E], *4732*, *144*	↓ 7E-5	↓ .37		+ [83]	
	miR-143	*143*[E], *145*[E]	↓ 7E-4, 1E-3	↓ .33, .23	+ [24]	+ [83]	
	miR-155		↑* [12, 24, 80, 85]		+ [24, 81]	+ [85]	+ [12, 84]
	miR-17	*17*, *18a*, *19a*, *20a*, *19b-1*, *92a-1*	↑* [12, 51, 80]		+ [24]		+ [12, 84]
miRNA biogenesis genes	*DICER1*		↑/↓* [87–89]		+ [87, 88]		+ [90–93]
	DROSHA		↑* [87]		+ [87]		+ [73, 91–93]

[E]miRNAs reported as top-ranked in both meta-analyses;
*expression changes non top-ranked or not analyzed in meta-analyses;
[B]altered in plasma/serum/blood/sputum of lung cancer patients and/or associated with early stage NSCLC [23, 82, 85, 86]

was *DROSHA*. To investigate whether copy number increases in *DROSHA* result from amplification of other nearby genes or regions, we designed an additional MLPA assay (LC-5p) covering the short arm of chromosome 5 (5p-arm). Except for the 4 control probes that were used before, the assay was composed of (i) 6 probes more or less evenly distributed along the entire chromosome arm, (ii) 5 probes covering the *DROSHA* gene (3 probes used before and 2 new probes), and 3 probes covering the *GOLPH3* gene, recently identified as oncogene [41] located in close proximity (~0.5 Mb upstream) to *DROSHA*. The locations of the probes are indicated in Figure 4A and Supplementary Table S1. With the use of the developed assay, we analyzed 20 samples selected based on the increased signal of *DROSHA* observed in the first experiment (18 amplifications and 2 gains). The copy number values of *DROSHA* determined by two independent experiments (with the use of LC-miR_1 and LC-5p assays) showed a very strong correlation ($R = 0.92$, $p < 0.0001$, data not shown). As shown in Figure 4, increased copy number is observed along almost the

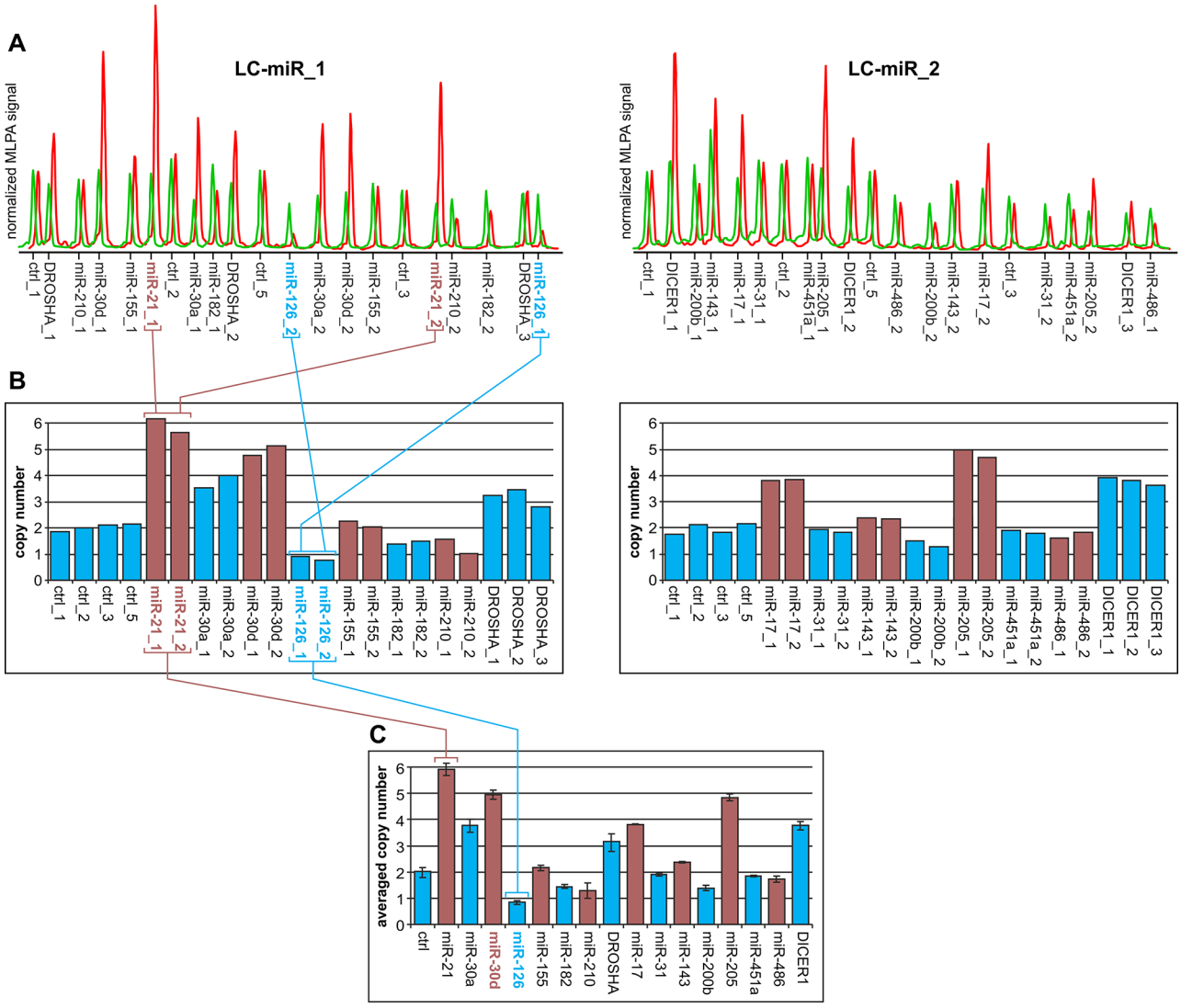


Figure 2: Copy number analysis of the selected genomic regions in a representative lung cancer sample. A. Electropherograms of MLPA results obtained with the use of LC-miR_1 (left-hand) and LC-miR_2 (right-hand) MLPA assays. The electropherograms of the cancer sample (red) are presented along the electropherograms from a reference non-cancer sample (green) and normalized against the signal of control probes. Probe IDs are indicated below the electropherograms. The probe signals (peak heights) correspond to the copy number of targeted regions. **B.** Bar plots (corresponding to the electropherograms of the cancer sample shown above (A)) represent the copy number value (y-axis) of each probe (x-axis) normalized by comparison of its signal in cancer samples to the corresponding signal in reference sample. The colors were used purely for sake of visualization purposes to better distinguish probes of subsequent genomic regions. Note that the signals of probes specific to the same genomic region are synchronized (e.g., probes miR-21_1 and miR-21_1 or miR-126_1 and miR-126_2; indicated in panels A and B). **C.** Bar plot representing the average copy number values of investigated regions in analyzed samples. Whiskers indicate maximum and minimum copy number values detected in particular regions, as shown in panel B. Note that genomic regions in which the difference between the maximum and minimum signal was higher than one-third of an average copy number value were excluded from further analysis (*miR-210*).

entire 5p-arm and no specific region shows sign of focal amplification. The region of amplifications observed in particular samples extends from the probe 5p_10, 2M to the probes covering *DROSHA*, and usually does not encompass *GOLPH3* (Figure 4). The above experiment clearly demonstrates that amplification of *DROSHA* is part of a chromosome-level amplification of the 5p-arm and is not a "passenger" effect of focal amplification of some other oncogene.

Survival analysis of patients stratified by copy number categories of miRNA and miRNA biogenesis genes

The overall survival data were available for 120 of the analyzed patient samples. Median overall survival of these patients was 416 days (14 months). Kaplan-Meier survival analysis of patients grouped based on copy number categories showed significant decreases in the

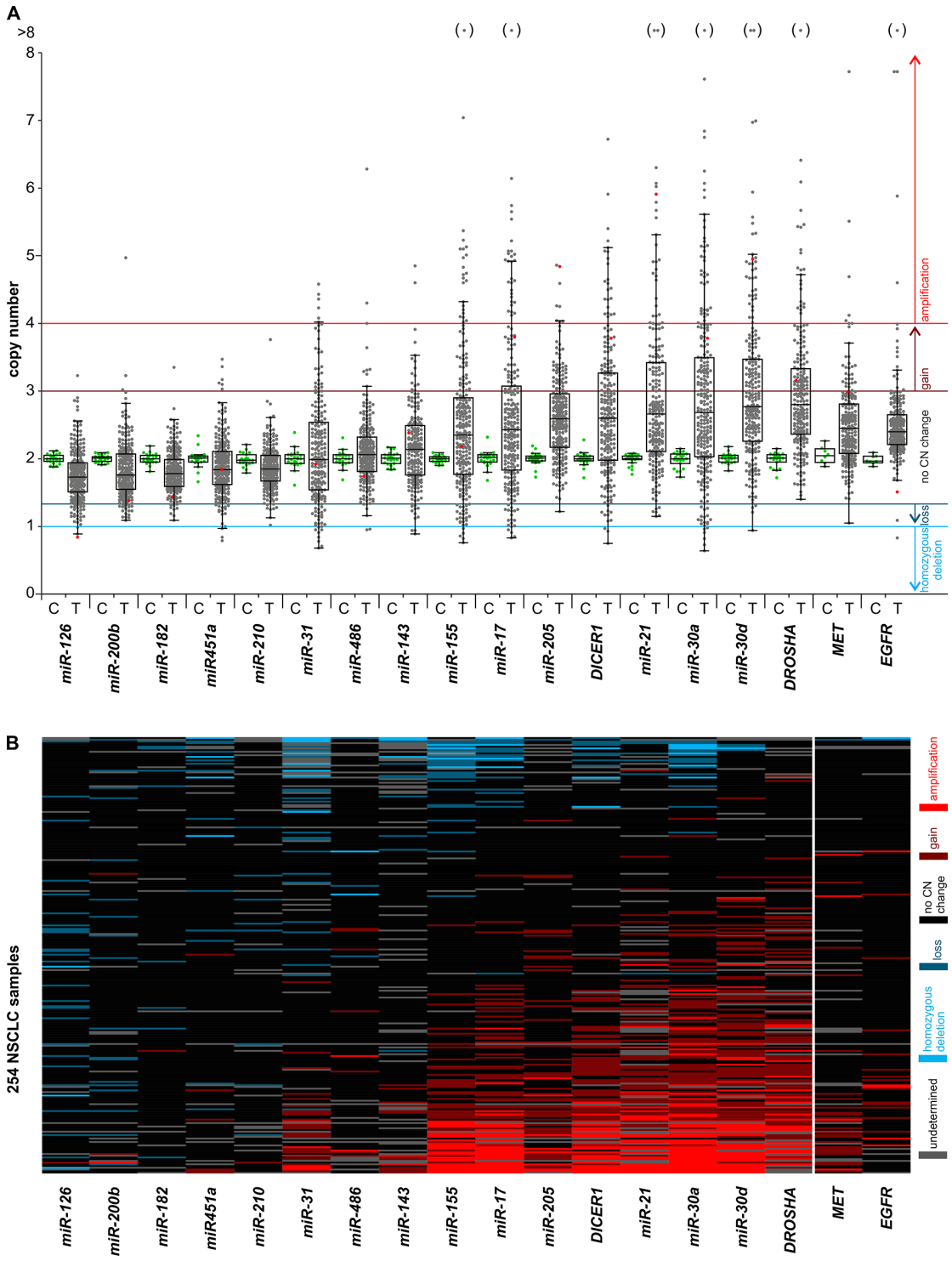


Figure 3: Graphical summary of the copy number variation of the analyzed genes in NSCLC samples. The graph shows the results of copy number analysis of the selected miRNA and miRNA biogenesis genes as well as two lung cancer related oncogenes, *MET* and *EGFR*. **A.** The graph shows the relative copy number values (y-axis) of selected genes (x-axis) of all studied samples. The genes were ordered from the lowest to highest median copy number value. Each dot represents the copy number value of individual control (C – green dot) or lung cancer (T – grey dot) samples. Red dots indicate copy number values of the representative lung cancer sample, analysis of which is shown in Figure 2. Dots in brackets (above) indicate samples with a copy number value >8. Color lines represent threshold values of homozygous deletions, losses, gains and amplifications. The outlined Tukey box-and-whisker plots indicate 1st quartile, median and 3rd quartile,and summarize the distribution of the presented copy number values. **B.** The heatmap graph showing the distribution of copy number categories of analyzed genes (columns) in 254 lung cancer samples (rows). The genes (from the left) and samples (from the top) were ordered from the lowest to highest average copy number value. Copy number categories are indicated by colors as shown in the legend on the right.

Table 2. Summary of copy number changes observed in analyzed miRNA and miRNA biogenesis genes in NSCLC samples

	expression	**copy number**: median (average)	**gains**: number (%)	**amplifications**: number (%)	**losses**:number (%)	**hom. deletions**: number (%)	informative samples #
miR-126	↓	1.73 (1.76)	1 (0.4)	0 (0)	26 (10.8)	3 (1.2)	241
miR-200b	↑	1.76 (1.84)	2 (0.9)	1 (0.4)	18 (7.7)	0 (0)	235
miR-182	↑	1.78 (1.81)	1 (0.4)	0 (0)	9 (3.8)	0 (0)	240
miR-451a	↓	1.84 (1.89)	6 (2.5)	0 (0)	12 (5.0)	4 (1.7)	239
miR-210	↑	1.85 (1.87)	1 (0.4)	0 (0)	9 (4.0)	0 (0)	227
miR-31	↑	1.99 (2.11)	22 (10.0)	6 (2.7)	20 (9.1)	8 (3.7)	219
miR-486	↓	2.06 (2.11)	7 (3.0)	2 (0.9)	3 (1.3)	2 (0.9)	233
miR-143	↓	2.14 (2.16)	12 (5.9)	2 (1.0)	8 (3.9)	3 (1.5)	205
miR-155	↑	2.33 (2.50)	33 (13.5)	23 (9.4)	21 (8.6)	5 (2.0)	245
miR-17	↑	2.42 (2.62)	39 (16.0)	28 (11.5)	15 (6.1)	5 (2.0)	244
miR-205	↑	2.59 (2.61)	45 (19.0)	8 (3.4)	1 (0.4)	0 (0)	237
DICER1	↓/↑	2.60 (2.69)	51 (21.8)	23 (9.8)	8 (3.4)	3 (1.3)	234
miR-21	↑	2.63 (2.90)	48 (22.7)	25 (11.8)	6 (2.8)	0 (0)	211
miR-30a	↓	2.67 (2.90)	59 (23.8)	40 (16.1)	14 (5.6)	6 (2.4)	248
miR-30d	↓	2.77 (3.02)	63 (26.6)	35 (14.8)	3 (1.3)	1 (0.4)	237
DROSHA	↑	2.79 (3.00)	67 (30.7)	23 (10.6)	0 (0)	0 (0)	218
MET	↑	2.45 (2.50)	23 (9.9)	5 (2.2)	1 (0.4)	0 (0)	232
EGFR	↑	2.41 (2.55)	13 (5.3)	5 (2.0)	1 (0.4)	1 (0.4)	246

survival of patients with the *miR-200b* deletion (log-rank test, $p = 0.022$) and patients with gain or amplification of *miR-30d* ($p = 0.013$) (Figure 5). This corresponds to a lower 5-year survival rate (0%) of patients with the above mentioned copy number aberrations compared to patients without the aberrations in *miR-200b* (6%) and *miR-30d* (10%).

Similar analyses performed for *DICER1* and *DROSHA* showed that samples with an increased copy number of *DROSHA* have significantly decreased survival and that the survival rate corresponds to the degree of copy number increase (log-rank test for trend, $p = 0.032$) (Figure 5).

Association of clinical data with copy number categories of miRNA and miRNA biogenesis genes

The copy number categories of any of the analyzed regions showed substantial association with the sex or age of the analyzed patients (Supplementary Table S3). Somewhat higher average age of diagnosis showed samples with *miR-126* deletion (with del/without del; 64.9/61.2 years; $p = 0.046$), *miR-451a* deletion (with del/without del; 66.8/61.2 years; $p = 0.041$), and with *miR-31* deletion (with del/without del; 65.7/61.0 years; $p = 0.017$). It has to be noted, however, that these associations are only marginally significant on the nominal level but not after adjustment for multiple comparisons. We also did not find any significant association of copy number categories with clinical data, such as stage of lung cancer at time of sample collection and metastasis/progression/remission status during the last examination (Supplementary Table S3). It has to be noted, however, that clinical data were available only for part of the analyzed samples ($N = 120$) and therefore, the lack of association may result from relatively low statistical power of our analysis.

Computational analysis of the association of *DICER1* and *DROSHA* copy number categories with their expression and cancer patient survival

Because we do not have access to mRNA/cDNA material or the expression data for our samples to

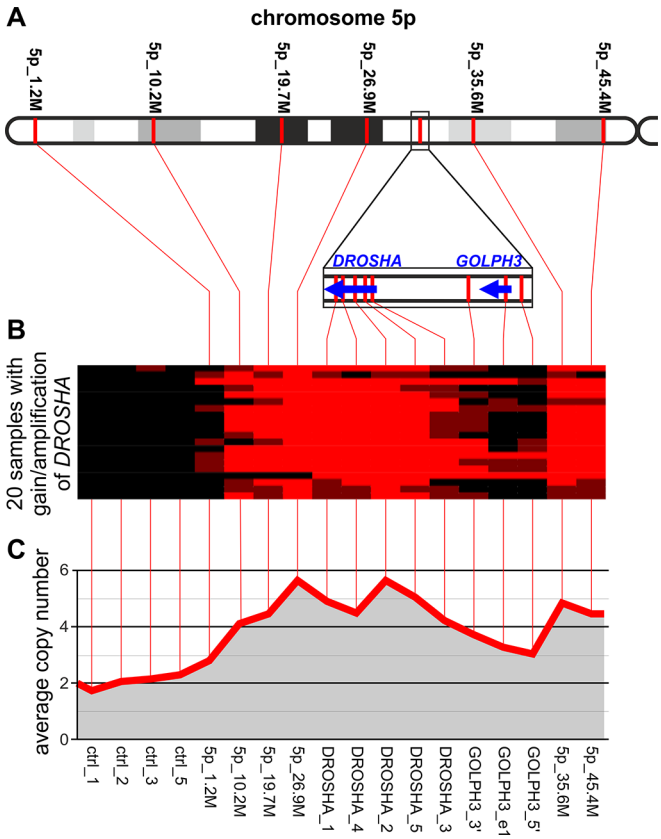


Figure 4: Analysis of copy number changes in the 5p-arm in NSCLC samples with gain/amplification of *DROSHA*. **A.** The schematic map of the 5p-arm with indicated positions of LC-5p MLPA probes (spaced by approximately 10 Mbp). The *DROSHA*/*GOLPH3* region, more densely covered by MLPA probes, is zoomed in on below. **B.** A heatmap graph showing copy number categories of all analyzed samples (18 with *DROSHA* amplification and 2 with *DROSHA* gain; rows) in control regions and in regions along the 5p-arm (columns). Red, brown and black colors indicate amplification, gain and no copy number change, respectively. **C.** A line-graph indicating the average copy number values (y-axis) of analyzed samples in control regions and in regions along the 5p-arm (x-axis). Note that in B and C, the spacing of consecutive probe signals depicted on the graphs does not correspond to their exact genomic distance.

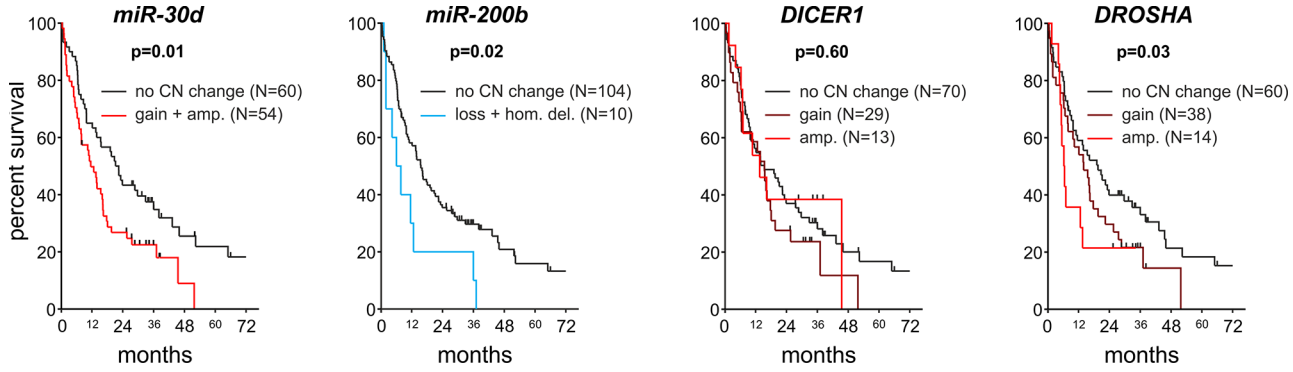


Figure 5: Survival analysis of NSCLC patients. Kaplan-Meier graphs presents the survival of patients stratified based on copy number categories of (from the left) *miR-30d*, *miR-200b*, *DICER1* and *DROSHA*.

determine whether copy number changes in *DICER1* and *DROSHA* correlate with their expression, we used data deposited in the cBioPortal for Cancer Genomics [42, 43]. As shown in Figure 6, there is a dose-dependent correlation between the copy number categories and the expression of *DICER1* and *DROSHA* in lung cancer (based on TCGA Cancer Genome ATLAS data [44]). A similar correlation can be observed in other cancers analyzed in different studies (Supplementary Figure S2). Further analysis with the use of another oncogenomic tool,

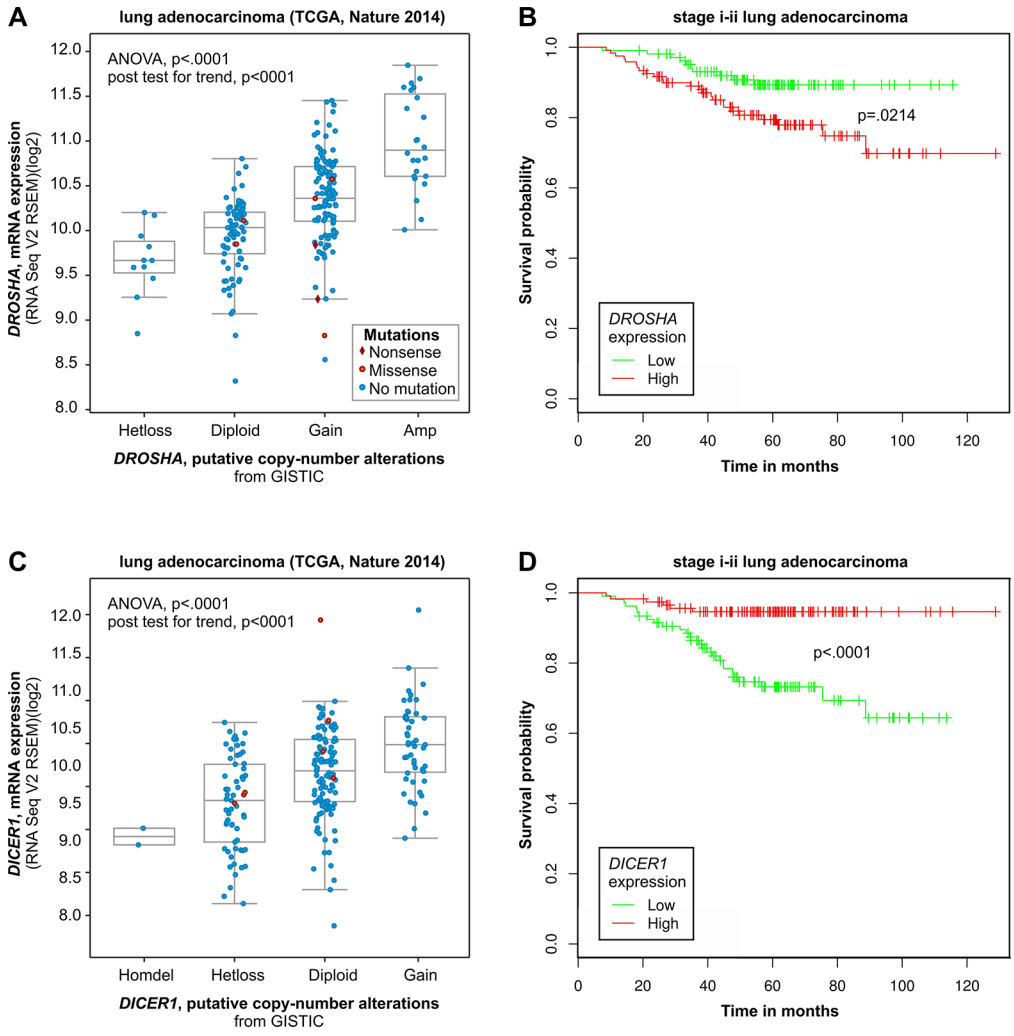


Figure 6: Computational analysis of clinical (survival) and oncogenomic data of *DROSHA* and *DICER1*. Mutual relation between copy number and expression (oncogenomic data) of *DROSHA* **A and B.** and *DICER1* **C and D.** and the relation of their expression to survival of cancer patients. A) and C) Correlation analysis of copy number categories and expression level performed with the use of a dataset (lung adenocarcinoma TCGA [79]) deposited and tools available in cBioPortal for Cancer Genomics. B) and D) Survival analysis performed with the use of a dataset (stage i-ii lung adenocarcinoma; GEO: GSE31210) deposited in and tools available from the PPISURV web portal.

PPISURV [45], showed that the increased expression of *DROSHA* generally (across cancers/datasets) correlates with decreased survival (Figure 6B and Supplementary Figure S2). In most deposited datasets/cancer types, including lung cancer, correlations show the same negative direction (in 6 of 36 datasets association show significance at $p \leq 0.05$). Similar analysis performed for *DICER1* shows the opposite effect of increased expression. In most deposited datasets, increased expression of *DICER1* shows the association (positive correlation) with increased survival (14 of 42 datasets show association at p-≤ 0.05; Figure 6B and Supplementary Figure S2).

DISCUSSION

With the use of two homemade MLPA assays, we analyzed the copy number variation of 14 miRNA genes reported as either over- or underexpressed in lung cancer. Additionally, we analyzed two critical miRNA biogenesis genes, *DROSHA* and *DICER1*. Each analyzed gene was tested by at least two independent MLPA probes, providing additional internal validation for the obtained results. To avoid any potential false results, the substantially discordant signals of matched probes were excluded from analysis. A similar strategy of somatic copy number variation analysis may be applied to almost any genomic region of interest in cancer samples. It should be noted, however, that the obtained copy number values are relative and to some extent may depend on the copy number variation of selected control regions (probes).

The analysis showed a substantial somatic copy number variation (both gains and losses) of all selected regions in cancer samples (compared variation in cancer vs. control, non-cancer samples; Figure 3). However,

the observed copy number alterations are not random, and some regions show a substantial increase (frequent amplifications), while the others show decrease in the average copy number value. The genes showing the highest average level of copy number include *miR-30d*, *miR-30a*, *miR-21*, *miR-205*, *miR-17*, *miR-155* as well as *DROSHA* and *DICER1*. Surprisingly, the average copy number and the frequency of amplifications of some of these genes (e.g., *DROSHA, miR-30d, miR-30a* and *miR-21)* are substantially higher than the corresponding values of well-known lung cancer-related oncogenes, *EGFR* and *MET*, analyzed in the same set of samples. In contrast, *miR-126* showed the lowest average copy number and a relatively high frequency of deletions and homozygous deletions. It should be noted, however, that due to the contamination of cancer samples with normal DNA and the inherent lower amplitude of copy number losses than copy number gains, the power of our analysis to detect deletions was substantially lower than the power to detect copy number gains/amplifications. Some genes, such as *miR-31*, show a relatively high frequency of both gains/amplifications and deletions.

As expected, the copy number variation of analyzed miRNAs does not correlate perfectly with the global expression changes of these miRNAs observed in lung cancer. However, our results indicate that copy number gains/amplifications may contribute substantially and may be an important mechanism underlying overexpression of miRNAs such as *miR-21*, *miR-17*, *miR-205* or *miR-155*. In short, these miRNAs are the best known oncomirs implicated not only in lung cancer but also in many other types of cancer (reviewed in [20, 26, 46, 47]). *MiR-21* was originally recognized as an antiapoptotic miRNA [48] that was strongly overexpressed in most types of cancer. Later, it was shown that *miR-21* promotes growth, metastasis and invasiveness, as well as chemo- and radioresistance of NSCLC, most likely by targeting tumor suppressor *PTEN* [49, 50]. In our experiment, *miR-17* represents 6 miRNAs coded in the *miR-17/92* cluster located within intron 3 of the *C13orf25* on chromosome 13. It was shown that the *miR-17/92* cluster may be upregulated by gene amplification, which is consistent with our results, or by *MYC* overexpression. It was also shown that upregulation of the *miR-17/92* cluster promotes cell proliferation and inhibits lung cell differentiation (the role of *miR-17/92* cluster was reviewed in [51]). *MiR-205* acts either as a tumor suppressor or as an oncogene. As an oncogene, it promotes tumor initiation, progression, resistance to therapies and inhibits apoptosis. It was shown that the oncogenic role of *miR-205* is expressed mostly by downregulation of tumor suppressors such as *PTEN* and *SHIP2* (references within [46]). *MiR-155* is encoded by the non-protein-coding gene *BIC*, originally identified as B-cell integration cluster for the avian leukosis virus, inducing lymphomas [52]. It was shown that *miR-155* targets several tumor suppressors such as *SOCS1*, *FOXO3*, and *VHL* and is involved in the regulation of cell survival, growth, chemosensitivity and tumor angiogenesis [53–55].

On the other hand, *miR-126* showed the lowest average copy number and frequent deletions in our study and is also recurrently found as downregulated in lung cancer. *MiR-126* was recognized as a tumor suppressor in most of the cancers studied. It was shown that *miR-126* may negatively control and inhibit cell proliferation, migration, invasion, and cancer cell survival. Among the validated targets of *miR-126* are such oncogenes as *ADAM9*, *CRK*, *EGFL7*, *HOXA9*, *IRS1*, *KRAS*, *PI3K*, *SLC7A5*, *SOX2*, and *VEGF* (reviewed and references within [56]).

The example of miRNAs which show discordant directions of expression and copy number changes are *miR-30a* and *miR-30d*, both belonging to *miR-30* family. *MiR-30a* and *miR-30d* belong to the miRNAs most frequently reported to be downregulated in lung cancer. On the other hand, these two miRNAs exhibit average copy number values and amplification frequencies that are among the highest of the genes analyzed in our study. It should be noted, however, that the copy number increases in *miR-30d* observed in our study correspond well to the results obtained previously by Li et al.. They showed that *miR-30d* is frequently amplified in different types of cancer (~30%) including lung cancer (27%), and that amplification of *miR-30d* correlates with its overexpression [57]. It was also shown that *miR-30d* downregulates many cancer-related genes, including apoptotic caspase *CASP3*, and is involved in the upregulation of such processes as cell proliferation, apoptosis, and migration [57]. The above facts strongly suggest the oncogenic character of *miR-30d*. Additionally, our results suggest that increased copy number of *miR-30d* (gains or amplifications vs. others) correlate with significantly reduced survival (Figure 5). On the other hand, *miR-30a* has been frequently implicated as a tumor suppressor. It was shown that *miR-30a* targets and downregulates the transcription factor Snai1 and consequently inhibits the epithelial-to-mesenchymal transition (EMT), invasion, mobility and metastasis of NSCLC cells [25]. The opposite characteristics of these two miRNAs may be reflected by the different frequency of deletions of these two miRNAs observed in our study. Although *miR-30a* showed a substantially increased average copy number, it was also one of the most frequently deleted in our analysis. Of our analyzed samples, 20 (8%) showed deletion of *miR-30a*, including 6 samples (2.4%) with homozygous deletions. For comparison, only 5 samples showed deletion of *miR-30d*.

Another example of miRNA with opposite trends in global expression and copy number changes is *miR-200b*. Although upregulation of *miR-200b* was recurrently identified in lung cancer, its character suggests it is likely a tumor suppressor. *MiR-200b* belongs to the *miR-200* family that maintains the general characteristics of the epithelia and inhibits EMT, tumor cell motility,

and invasiveness ([58] and references within). Among the experimentally identified and validated targets of *miR-200b* are numerous genes involved in the regulation of cytoskeletal organization and cell morphology in addition to *EGFR* [58]. Additionally, our analysis showed significantly decreased survival of patients with either deletion or homozygous deletion of *miR-200b*.

In addition to miRNA genes, we analyzed also two key miRNA biogenesis genes, *DICER1* and *DROSHA*. Both of these genes, but especially *DROSHA*, show substantial copy number increases and frequent high-copy number amplifications in analyzed samples. Review of the Cancer Gene Census (COSMIC database) reveals no proto-oncogene in close proximity of either *DROSHA* or *DICER1* that might drive their amplification. However, meticulous review of the literature allowed us to identify *GOLPH3* located in direct proximity (~600 kb upstream) of *DROSHA*. *GOLPH3* encodes a Golgi-localizing protein that was recently identified as a candidate oncogene driving the amplification of the 5p13 region. This amplification has frequently been observed in multiple solid tumors, including lung cancer [41]. It was shown that Golph3 enlarges cell size, enhances growth-factor-induced mTOR signaling in human cancer cells, and increases the sensitivity to an mTOR inhibitor [41]. The detailed analysis showed that the region of amplification comprising *GOLPH3* is very narrow and does not extend to *DROSHA*. However, the frequency of *GOLPH3* amplification in lung cancer observed previously (56%) corresponded well to the frequency of gains/amplifications of *DROSHA* observed in our study (42%). To verify whether the *DROSHA* amplifications observed in our study might be driven by the closely located *GOLPH3*, we reanalyzed this region with the use of the new 5p-arm-specific MLPA assay. This experiment confirmed *DROSHA* amplifications in analyzed samples and showed that amplification of *DROSHA* results mostly from the chromosome-level amplification of almost the entire 5p-arm. This experiment clearly demonstrated that amplification of *DROSHA* does not depend on the focal amplification of closely located *GOLPH3* or any other specific oncogene on the 5p-arm. Regardless of whether *DROSHA* and *DICER1* are drivers of their amplifications, the amplifications of these two key miRNA biogenesis genes may increase their expression and, as a consequence, may contribute to the global destabilization of miRNA expression observed in many types of cancer.

The computational analysis of publically available oncogenomic data showed that the copy number variation of *DROSHA* correlates well with its expression and that increased expression of *DROSHA* is associated with worse survival. The above analyses of oncogenomic data are in line with our experimental results suggesting decreased survival of patients with gain or amplification of *DROSHA* (Figure 5). A similar computational analysis of *DICER1* also showed a good correlation between its copy number categories and expression. However, in contrast to *DROSHA*, increased expression of *DICER1* was associated with longer survival in various cancers including lung cancer. Although such results must be interpreted with caution, the opposite effects of increased expression of *DROSHA* and *DICER1* on survival (positive and negative, respectively) may suggest the oncogenic role of intermediate products of these two enzymes, that is, pre-miRNAs (either specific or as a class). It should be noted that the advantage of the computational results discussed above is that they are based on independent (objectified) whole genome datasets generated in projects not focused specifically on *DICER1*, *DROSHA* or any other miRNA biogenesis gene.

Our results add to the complex picture of the role of *DICER1* and *DROSHA* in cancer. The miRNA biogenesis genes were primarily considered as haploinsufficient tumor suppressors [59]. This notion results mostly from the observation that the overall level of miRNAs is often reduced in cancer [60–62] and from the fact that germline loss-of-function mutations in *DICER1* are causative variants in the so called DICER1 syndrome, which is associated with increased risk of numerous, mostly early, childhood malignancies and benign tumors [63]. The representative (most common) malignancy for this syndrome is pleuropulmonary blastoma, which occurs in the lungs. More recently, analysis of cancers associated with DICER1 syndrome as well as other early childhood cancers (e.g., Wilms tumor) led to the identification of a peculiar pattern of somatic second-hit mutations in *DICER1* and *DROSHA*. These mostly missense mutations are not randomly distributed over the genes but form clear hotspots, mostly affecting few amino acid residues located in or adjacent to metal-ion-binding residues in the RNaseIIIb domain of either *DICER1* (D1709, E1813) or *DROSHA* (E1147, D1151) [63–68]. Functional analyses suggest that these mutations are not deleterious (as expected for typical second-hit mutations) but rather modify the function of DICER1 or DROSHA, making it favorable for cancer (oncogenic) (recently discussed in [69, 70]). It was shown that modified enzymes selectively reduce the processing of miRNAs generated from the 5′ arm of pre-miRNA hairpins and as a consequence modify the miRNA expression profile in cancer [65, 66, 71, 72].

Our results and the notion about the oncogenic role of *DROSHA* are very much in line with previous results suggesting that *DROSHA* is a key gene driving frequent gains of the 5p-arm in cervical squamous cell carcinoma (SCC) [73, 74]. Analysis of primary cervical SCC samples and cell lines showed that the frequent copy number gains and overexpression of *DROSHA* led to an altered profile of miRNA expression, including the expression of many cancer-related miRNAs. Among the miRNAs showing the highest overexpression was *miR-31*. Functional *in vitro* analyses (including wound healing test) showed that upregulation of *DROSHA* increases motility and

invasiveness of squamous SCC cell lines [73, 74]. It was also shown that overexpression of *DROSHA* is associated with metastasis and decreased survival in esophageal cancer patients [75].

It should be noted that other genes of miRNA biogenesis enzymes may contribute to the regulation of global or individual miRNA expression in cancer. Therefore, to better understand and evaluate the impact of somatic copy number variation of miRNA biogenesis genes on miRNA expression in cancer, a more complex analysis is needed.

In conclusion, our results show a substantial somatic copy number variation in genomic regions comprising miRNA genes. Among these regions were those showing a substantial increase in the average copy number (frequently amplified), and regions with decreased average copy number. Concordance of copy number and expression changes of some miRNAs suggest that copy number variation may be an important mechanism responsible for regulation of these miRNAs in lung cancer. Therefore our observations support the proposed earlier notion, implying the high genomic instability of miRNA gene regions and contribution of copy number variation in the regulation of miRNA expression in cancer [29, 30]. It should be emphasized however that the amplitude and recurrence of copy number changes cannot be simply interpreted as the oncogenic role of a variable region/gene in cancer.

Our results also indicate the important role of miRNA biogenesis genes, especially *DROSHA*, in lung cancer. Even if these genes are not drivers of their copy number changes, they may affect global regulation of miRNA expression in cancer.

Finally, somatic copy number changes of some of the analyzed genes including *DROSHA* correlate with survival of cancer patients. Although the results of our survival analyses are only marginally significant (relatively low number of samples) and must be replicated in an independent group of samples, the copy number changes would be attractive biomarkers due to (i) the relatively high stability of genomic DNA, even extracted from formalin-fixed paraffin embedded (FFPE) samples; (ii) the small amount of DNA necessary for analysis; (iii) relatively low cost; (iv) simplicity; and (v) the reliability of copy number analysis. The drawback of such analysis is, however, contamination of the cancer samples with a difficult to estimate amount of normal DNA.

MATERIALS AND METHODS

Selection and processing of NSCLC samples for molecular analysis

We retrospectively reviewed a cohort of 254 patients with histopathologically confirmed NSCLC diagnosed at the Franciszek Lukaszczyk Oncology Center in Bydgoszcz (central Poland). The age of the patients ranged from 35 to 81. A total of 254 specimens that passed the quality control steps (microscopic analysis and tumor content qualification as well as qualitative and quantitative DNA analysis) were obtained following surgeries, fine-needle aspirations (FNAs), endobronchial ultrasound with guided transbronchial needle aspiration (EBUS-TBNA) procedures or pleural fluid sampling. The samples were stained with hematoxylin and eosin for the qualitative and quantitative analysis of tumor cells in the analyzed material (including macrodissection in marked out samples) as described previously [76]. The study was approved by the Committee of Ethics of Scientific Research of Collegium Medicum of Nicolaus Copernicus University, Poland (KB 265/2012). The data were analyzed anonymously.

DNA extraction was performed after the microdissection of a region indicated by the pathomorphologist, and the quality and quantity of DNA samples were evaluated as described previously [40].

Copy number analysis by MLPA

MLPA analysis was performed with the use of three in-house designed and generated assays, LC-miR_1, LC-miR_2 and LC-5p. Both LC-miR_1 and LC-miR_2 assays contained 14 probes specific for 7 miRNA genes (two probes for each miRNA or miRNA-cluster gene), 3 probes specific for one of miRNA biogenesis gene, and 4 control probes (located on different chromosomes outside of chromosome 5 and regions of known cancer-related genes). The LC-5p assay contained 6 probes more or less evenly covering the short arm of chromosome 5 (5p-arm), 5 probes specific for *DROSHA*, 3 probes specific for *GOLPH3*, and 4 control probes. The detailed characteristics, genomic positions and sequences of all probes used in this study are presented in Supplementary Table S1.

The MLPA probes and the general layout of the probe sets were designed according to a previously proposed strategy [36, 37]. This strategy utilizes only short oligonucleotide probes that can easily be generated via standard chemical synthesis. Briefly, each probe was composed of two half-probes of equal size, and the total probe length ranged from 93 to 164 nt. The target sequences for the probes were selected to avoid SNPs, repeat elements and sequences of extremely high or low GC content. The MLPA probes were synthesized by IDT (Skokie, IL, USA).

The MLPA reactions were run according to the manufacturer's general recommendations (MRC-Holland, Amsterdam, the Netherlands), as described earlier in [37, 77]. All reagents except the probe mixes were purchased form MRC-Holland (http://www.mlpa.com). The products of the MLPA reaction were subsequently diluted 20x in HiDi formamide containing GS Liz600,

which was used as a DNA sizing standard, and separated via capillary electrophoresis (POP7 polymer) in an ABI Prism 3130XL apparatus (Applied Biosystems, Carlsbad, CA, USA).

The obtained electropherograms were analyzed using GeneMarker software v2.4.0 (SoftGenetics, State College, PA, USA). The signal intensities (peak heights) were retrieved and transferred to prepared Excel sheets (available upon request). For each individual sample, the signal intensity of each probe was divided by the average signal intensity of the control probes to normalize the obtained values and to equalize run-to-run variation. Due to high signal variation, the control probe 3 (ctrl_3) was excluded from analysis. To calculate relative copy number value of particular probe, the normalized signal of this probe was divided by a corresponding value of this probe in the reference (non-cancer) sample and multiplied by 2. The relative copy number of a particular gene was calculated as an average of the normalized copy number value of 2 or 3 probes specific to this gene. If the difference between the maximum and minimum signal of the averaged probes was higher than one-third of an average copy number value or if the coefficient of variation of the averaged probes was higher than 0.3, the result was excluded from further analyses.

Databases and statistical analysis

All statistical analyses were performed using Statistica (StatSoft, Tulsa, OK) or Prism v. 4.0 (GraphPad, San Diego, CA). All *p*-values were provided for two-sided tests. All human genome positions indicated in this report refer to the February 2009 (GRCh37/hg19) human reference sequence. The datasets for analysis and visualization of the relationship between copy number category and expression level of *DROSHA* and *DICER1* were obtained from cBioPortal for Cancer Genomics (MemorialSloan-Kettering Cancer Center, New York, NY, USA; http://www.cbioportal.org/) [42, 43] and were analyzed with the use of the cBioPortal Plots tool. The survival analyses of the cancer patients with high and low levels of either *DICER1* or *DROSHA* expression were performed with the use of datasets and tools available in the PPISURV portal (http://www.bioprofiling.de/GEO/PPISURV/ppisurv. html) [45].

ACKNOWLEDGMENTS

This work was supported by National Science Centre 2011/01/B/NZ5/02773, and KNOW program of the Polish Ministry of Science and Higher Education.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65:87–108.
2. Hagemann IS, Devarakonda S, Lockwood CM, Spencer DH, Guebert K, Bredemeyer AJ, Al-Kateb H, Nguyen TT, Duncavage EJ, Cottrell CE, Kulkarni S, Nagarajan R, Seibert K, et al. Clinical next-generation sequencing in patients with non-small cell lung cancer. Cancer. 2015; 121:631–639.
3. Liu P, Morrison C, Wang L, Xiong D, Vedell P, Cui P, Hua X, Ding F, Lu Y, James M, Ebben JD, Xu H, Adjei AA, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. Carcinogenesis. 2012; 33:1270–1276.
4. Zhao Y, Yang J, Chen Z, Gao Z, Zhou F, Li X, Li J, Shi S, Feng X, Sun N, Yao R, Zhou C, Chang S, et al. Identification of somatic alterations in stage I lung adenocarcinomas by next-generation sequencing. Genes Chromosomes Cancer. 2014; 53:289–298.
5. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature. 2008; 455:1069–1075.
6. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150:1107–1120.
7. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007; 450:893–898.
8. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463:899–905.
9. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004; 304:1497–1500.
10. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007; 448:561–566.
11. Du L, Pertsemlidis A. microRNAs and lung cancer: tumors and 22-mers. Cancer Metastasis Rev. 2010; 29:109–122.

12. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A. 2006; 103:2257–2261.
13. Griffiths-Jones S. The microRNA Registry. Nucleic Acids Res. 2004; 32:D109–111.
14. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic acids research. 2014; 42:D68–73.
15. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116:281–297.
16. Harfe BD. MicroRNAs in vertebrate development. Curr Opin Genet Dev. 2005; 15:410–415.
17. Boehm M, Slack FJ. MicroRNA control of lifespan and metabolism. Cell Cycle. 2006; 5:837–840.
18. Carleton M, Cleary MA, Linsley PS. MicroRNAs and cell cycle regulation. Cell Cycle. 2007; 6:2127–2132.
19. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, Rassenti L, Kipps T, Negrini M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002; 99:15524–15529.
20. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006; 6:259–269.
21. Hayashita Y, Osada H, Tatematsu Y, Yamada H, Yanagisawa K, Tomida S, Yatabe Y, Kawahara K, Sekido Y, Takahashi T. A polycistronic microRNA cluster, miR-17–92, is overexpressed in human lung cancers and enhances cell proliferation. Cancer Res. 2005; 65:9628–9632.
22. Zandberga E, Kozirovskis V, Abols A, Andrejeva D, Purkalne G, Line A. Cell-free microRNAs as diagnostic, prognostic, and predictive biomarkers for lung cancer. Genes Chromosomes Cancer. 2013; 52:356–369.
23. Shen J, Todd NW, Zhang H, Yu L, Lingxiao X, Mei Y, Guarnera M, Liao J, Chou A, Lu CL, Jiang Z, Fang H, Katz RL, et al. Plasma microRNAs as potential biomarkers for non-small-cell lung cancer. Lab Invest. 2011; 91:579–587.
24. Yanaihara N, Caplen N, Bowman E, Seike M, Kumamoto K, Yi M, Stephens RM, Okamoto A, Yokota J, Tanaka T, Calin GA, Liu CG, Croce CM, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell. 2006; 9:189–198.
25. Kumarswamy R, Mudduluru G, Ceppi P, Muppala S, Kozlowski M, Niklinski J, Papotti M, Allgayer H. MicroRNA-30a inhibits epithelial-to-mesenchymal transition by targeting Snai1 and is downregulated in non-small cell lung cancer. Int J Cancer. 2012; 130:2044–2053.
26. Kumarswamy R, Volkmann I, Thum T. Regulation and function of miRNA-21 in health and disease. RNA Biol. 2011; 8:706–713.
27. Zernecke A, Bidzhekov K, Noels H, Shagdarsuren E, Gan L, Denecke B, Hristov M, Koppel T, Jahantigh MN, Lutgens E, Wang S, Olson EN, Schober A, et al. Delivery of microRNA-126 by apoptotic bodies induces CXCL12-dependent vascular protection. Sci Signal. 2009; 2. ra81.
28. Diederichs S, Haber DA. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. Cancer Res. 2006; 66:6097–6104.
29. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M, Croce CM. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004; 101:2999–3004.
30. Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, Yao G, Medina A, O'Brien-Jenkins A, et al. microRNAs exhibit high frequency genomic alterations in human cancer. Proc Natl Acad Sci U S A. 2006; 103:9136–9141.
31. Huppi K, Volfovsky N, Mackiewicz M, Runfola T, Jones TL, Martin SE, Stephens R, Caplen NJ. MicroRNAs and genomic instability. Semin Cancer Biol. 2007; 17:65–73.
32. Zhang L, Coukos G. MicroRNAs: a new insight into cancer genome. Cell Cycle. 2006; 5:2216–2219.
33. Kan CW, Howell VM, Hahn MA, Marsh DJ. Genomic alterations as mediators of miRNA dysregulation in ovarian cancer. Genes Chromosomes Cancer. 2015; 54:1–19.
34. Guan P, Yin Z, Li X, Wu W, Zhou B. Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. J Exp Clin Cancer Res. 2012; 31:54.
35. Vosa U, Vooder T, Kolde R, Vilo J, Metspalu A, Annilo T. Meta-analysis of microRNA expression in lung cancer. Int J Cancer. 2013; 132:2884–2893.
36. Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. ScientificWorldJournal. 2010; 10:2003–2018.
37. Kozlowski P, Roberts P, Dabora S, Franz D, Bissler J, Northrup H, Au KS, Lazarus R, Domanska-Pakiela D, Kotulska K, Jozwiak S, Kwiatkowski DJ. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Hum Genet. 2007; 121:389–400.
38. Lacle MM, Kornegoor R, Moelans CB, Maes-Verschuur AH, van der Pol C, Witkamp AJ, van der Wall E, Rueschoff J, Buerger H, van Diest PJ. Analysis of copy number changes on chromosome 16q in male breast cancer by multiplex ligation-dependent probe amplification. Mod Pathol. 2013; 26:1461–1467.
39. Rygiel AM, Milano F, Ten Kate FJ, Schaap A, Wang KK, Peppelenbosch MP, Bergman JJ, Krishnadath KK. Gains and amplifications of c-myc, EGFR, and 20.q13 loci in

the no dysplasia-dysplasia-adenocarcinoma sequence of Barrett's esophagus. Cancer Epidemiol Biomarkers Prev. 2008; 17:1380–1385.

40. Lewandowska MA, Czubak K, Klonowska K, Jozwicki W, Kowalewski J, Kozlowski P. The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small EGFR Mutations and EGFR Amplification in Lung Cancer. PloS One. 2015; 10:e0117983.
41. Scott KL, Kabbarah O, Liang MC, Ivanova E, Anagnostou V, Wu J, Dhakal S, Wu M, Chen S, Feinberg T, Huang J, Saci A, Widlund HR, et al. GOLPH3 modulates mTOR signalling and rapamycin sensitivity in cancer. Nature. 2009; 459:1085–1090.
42. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2:401–404.
43. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6:pl1.
44. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068.
45. Antonov AV, Krestyaninova M, Knight RA, Rodchenkov I, Melino G, Barlev NA. PPISURV: a novel bioinformatics tool for uncovering the hidden role of specific genes in cancer survival outcome. Oncogene. 2014; 33:1621–1628.
46. Qin AY, Zhang XW, Liu L, Yu JP, Li H, Wang SZ, Ren XB, Cao S. MiR-205 in cancer: an angel or a devil? Eur J Cell Biol. 2013; 92:54–60.
47. Olive V, Li Q, He L. mir-17–92: a polycistronic oncomir with pleiotropic functions. Immunol Rev. 2013; 253:158–166.
48. Chan JA, Krichevsky AM, Kosik KS. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. Cancer Res. 2005; 65:6029–6033.
49. Liu ZL, Wang H, Liu J, Wang ZX. MicroRNA-21 (miR-21) expression promotes growth, metastasis, and chemo- or radioresistance in non-small cell lung cancer cells by targeting PTEN. Mol Cell Biochem. 2013; 372:35–45.
50. Zhang JG, Wang JJ, Zhao F, Liu Q, Jiang K, Yang GH. MicroRNA-21 (miR-21) represses tumor suppressor PTEN and promotes growth and invasion in non-small cell lung cancer (NSCLC). Clin Chim Acta. 2010; 411:846–852.
51. Osada H, Takahashi T. let-7 and miR-17–92: small-sized major players in lung cancer development. Cancer Sci. 2011; 102:9–17.
52. Tam W, Ben-Yehuda D, Hayward WS. bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol Cell Biol. 1997; 17:1490–1502.
53. Jiang S, Zhang HW, Lu MH, He XH, Li Y, Gu H, Liu MF, Wang ED. MicroRNA-155 functions as an OncomiR in breast cancer by targeting the suppressor of cytokine signaling 1 gene. Cancer Res. 2010; 70:3119–3127.
54. Kong W, He L, Coppola M, Guo J, Esposito NN, Coppola D, Cheng JQ. MicroRNA-155 regulates cell survival, growth, and chemosensitivity by targeting FOXO3a in breast cancer. J Biol Chem. 2010; 285:17869–17879.
55. Kong W, He L, Richards EJ, Challa S, Xu CX, Permuth-Wey J, Lancaster JM, Coppola D, Sellers TA, Djeu JY, Cheng JQ. Upregulation of miRNA-155 promotes tumour angiogenesis by targeting VHL and is associated with poor prognosis and triple-negative breast cancer. Oncogene. 2014; 33:679–689.
56. Ebrahimi F, Gopalan V, Smith RA, Lam AK. miR-126 in human cancers: clinical roles and current perspectives. Exp Mol Pathol. 2014; 96:98–107.
57. Li N, Kaur S, Greshock J, Lassus H, Zhong X, Wang Y, Leminen A, Shao Z, Hu X, Liang S, Katsaros D, Huang Q, Butzow R, et al. A combined array-based comparative genomic hybridization and functional library screening approach identifies mir-30d as an oncomir in cancer. Cancer Res. 2012; 72:154–164.
58. Bracken CP, Li X, Wright JA, Lawrence DM, Pillman KA, Salmanidis M, Anderson MA, Dredge BK, Gregory PA, Tsykin A, Neilsen C, Thomson DW, Bert AG, et al. Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. EMBO J. 2014; 33:2040–2056.
59. Kumar MS, Pester RE, Chen CY, Lane K, Chin C, Lu J, Kirsch DG, Golub TR, Jacks T. Dicer1 functions as a haploinsufficient tumor suppressor. Genes Dev. 2009; 23:2700–2704.
60. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, et al. MicroRNA expression profiles classify human cancers. Nature. 2005; 435:834–838.
61. Kumar MS, Lu J, Mercer KL, Golub TR, Jacks T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nat Genet. 2007; 39:673–677.
62. Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, Ambros VR, Israel MA. Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. Cancer Res. 2007; 67:2456–2468.
63. Hill DA, Ivanovich J, Priest JR, Gurnett CA, Dehner LP, Desruisseau D, Jarzembowski JA, Wikenheiser-Brokamp KA, Suarez BK, Whelan AJ, Williams G, Bracamontes D, Messinger Y, et al. DICER1 mutations in familial pleuropulmonary blastoma. Science. 2009; 325:965.
64. Pugh TJ, Yu W, Yang J, Field AL, Ambrogio L, Carter SL, Cibulskis K, Giannikopoulos P, Kiezun A, Kim J,

McKenna A, Nickerson E, Getz G, et al. Exome sequencing of pleuropulmonary blastoma reveals frequent biallelic loss of TP53 and two hits in DICER1 resulting in retention of 5p-derived miRNA hairpin loop sequences. Oncogene. 2014; 33:5295–5302.

65. Heravi-Moussavi A, Anglesio MS, Cheng SW, Senz J, Yang W, Prentice L, Fejes AP, Chow C, Tone A, Kalloger SE, Hamel N, Roth A, Ha G, et al. Recurrent somatic DICER1 mutations in nonepithelial ovarian cancers. N Engl J Med. 2012; 366:234–242.

66. Torrezan GT, Ferreira EN, Nakahata AM, Barros BD, Castro MT, Correa BR, Krepischi AC, Olivieri EH, Cunha IW, Tabori U, Grundy PE, Costa CM, de Camargo B, et al. Recurrent somatic mutation in DROSHA induces microRNA profile changes in Wilms tumour. Nat Commun. 2014; 5:4039.

67. Seki M, Yoshida K, Shiraishi Y, Shimamura T, Sato Y, Nishimura R, Okuno Y, Chiba K, Tanaka H, Kato K, Kato M, Hanada R, Nomura Y, et al. Biallelic DICER1 mutations in sporadic pleuropulmonary blastoma. Cancer Res. 2014; 74:2742–2749.

68. Rakheja D, Chen KS, Liu Y, Shukla AA, Schmid V, Chang TC, Khokhar S, Wickiser JE, Karandikar NJ, Malter JS, Mendell JT, Amatruda JF. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. Nat Commun. 2014; 2:4802.

69. Foulkes WD, Priest JR, Duchaine TF. DICER1: mutations, microRNAs and mechanisms. Nat Rev Cancer. 2014; 14:662–672.

70. Kurzynska-Kokorniak A, Koralewska N, Pokornowska M, Urbanowicz A, Tworak A, Mickiewicz A, Figlerowicz M. The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. Nucleic Acids Res. 2015; 1.

71. Anglesio MS, Wang Y, Yang W, Senz J, Wan A, Heravi-Moussavi A, Salamanca C, Maines-Bandiera S, Huntsman DG, Morin GB. Cancer-associated somatic DICER1 hotspot mutations cause defective miRNA processing and reverse-strand expression bias to predominantly mature 3p strands through loss of 5p strand cleavage. J Pathol. 2013; 229:400–409.

72. Gurtan AM, Lu V, Bhutkar A, Sharp PA. *In vivo* structure-function analysis of human Dicer reveals directional processing of precursor miRNAs. RNA. 2012; 18:1116–1122.

73. Muralidhar B, Goldstein LD, Ng G, Winder DM, Palmer RD, Gooding EL, Barbosa-Morais NL, Mukherjee G, Thorne NP, Roberts I, Pett MR, Coleman N. Global microRNA profiles in cervical squamous cell carcinoma depend on Drosha expression levels. J Pathol. 2007; 212:368–377.

74. Muralidhar B, Winder D, Murray M, Palmer R, Barbosa-Morais N, Saini H, Roberts I, Pett M, Coleman N. Functional evidence that Drosha overexpression in cervical squamous cell carcinoma affects cell phenotype and microRNA profiles. J Pathol. 2011; 224:496–507.

75. Sugito N, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Kurehara H, Ando T, Mori R, Takashima N, Ogawa R, Fujii Y. RNASEN regulates cell proliferation and affects survival in esophageal cancer patients. Clin Cancer Res. 2006; 12:7322–7328.

76. Lewandowska MA, Jozwicki W, Zurawski B. KRAS and BRAF mutation analysis in colorectal adenocarcinoma specimens with a low percentage of tumor cells. Mol Diagn Ther. 2013; 17:193–203.

77. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002; 30:e57.

78. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer. 2004; 4:177–183.

79. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511:543–550.

80. Saito M, Schetter AJ, Mollerup S, Kohno T, Skaug V, Bowman ED, Mathe EA, Takenoshita S, Yokota J, Haugen A, Harris CC. The association of microRNA expression with prognosis and progression in early-stage, non-small cell lung adenocarcinoma: a retrospective analysis of three cohorts. Clin Cancer Res. 2011; 17:1875–1882.

81. Yang M, Shen H, Qiu C, Ni Y, Wang L, Dong W, Liao Y, Du J. High expression of miR-21 and miR-155 predicts recurrence and unfavourable survival in non-small cell lung cancer. Eur J Cancer. 2013; 49:604–615.

82. Yu L, Todd NW, Xing L, Xie Y, Zhang H, Liu Z, Fang H, Zhang J, Katz RL, Jiang F. Early detection of lung adenocarcinoma in sputum by a panel of microRNA markers. Int J Cancer. 2010; 127:2870–2878.

83. Wang Q, Wang S, Wang H, Li P, Ma Z. MicroRNAs: novel biomarkers for lung cancer diagnosis, prediction and treatment. Exp Biol Med (Maywood). 2012; 237:227–235.

84. Barbarotto E, Schmittgen TD, Calin GA. MicroRNAs and cancer: profile, profile, profile. Int J Cancer. 2008; 122:969–977.

85. Zheng D, Haddadin S, Wang Y, Gu LQ, Perry MC, Freter CE, Wang MX. Plasma microRNAs as novel biomarkers for early detection of lung cancer. Int J Clin Exp Pathol. 2011; 4:575–586.

86. Lin Q, Mao W, Shu Y, Lin F, Liu S, Shen H, Gao W, Li S, Shen D. A cluster of specified microRNAs in peripheral blood as biomarkers for metastatic non-small-cell lung cancer by stem-loop RT-PCR. J Cancer Res Clin Oncol. 2012; 138:85–93.

87. Diaz-Garcia CV, Agudo-Lopez A, Perez C, Lopez-Martin JA, Rodriguez-Peralto JL, de Castro J, Cortijo A, Martinez-Villanueva M, Iglesias L,

Garcia-Carbonero R, Fresno Vara JA, Gamez-Pozo A, Palacios J, et al. DICER1, DROSHA, and miRNAs in patients with non-small cell lung cancer: implications for outcomes and histologic classification. Carcinogenesis. 2013; 34:1031–1038.

88. Karube Y, Tanaka H, Osada H, Tomida S, Tatematsu Y, Yanagisawa K, Yatabe Y, Takamizawa J, Miyoshi S, Mitsudomi T, Takahashi T. Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci. 2005; 96:111–115.

89. Chiosea S, Jelezcova E, Chandran U, Luo J, Mantha G, Sobol RW, Dacic S. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. Cancer Res. 2007; 67:2345–2350.

90. Khoshnaw SM, Rakha EA, Abdel-Fatah TM, Nolan CC, Hodi Z, Macmillan DR, Ellis IO, Green AR. Loss of Dicer expression is associated with breast cancer progression and recurrence. Breast Cancer Res Treat. 2012; 135:403–413.

91. Han Y, Liu Y, Gui Y, Cai Z. Inducing cell proliferation inhibition and apoptosis via silencing Dicer, Drosha, and Exportin 5 in urothelial carcinoma of the bladder. J Surg Oncol. 2013; 107:201–205.

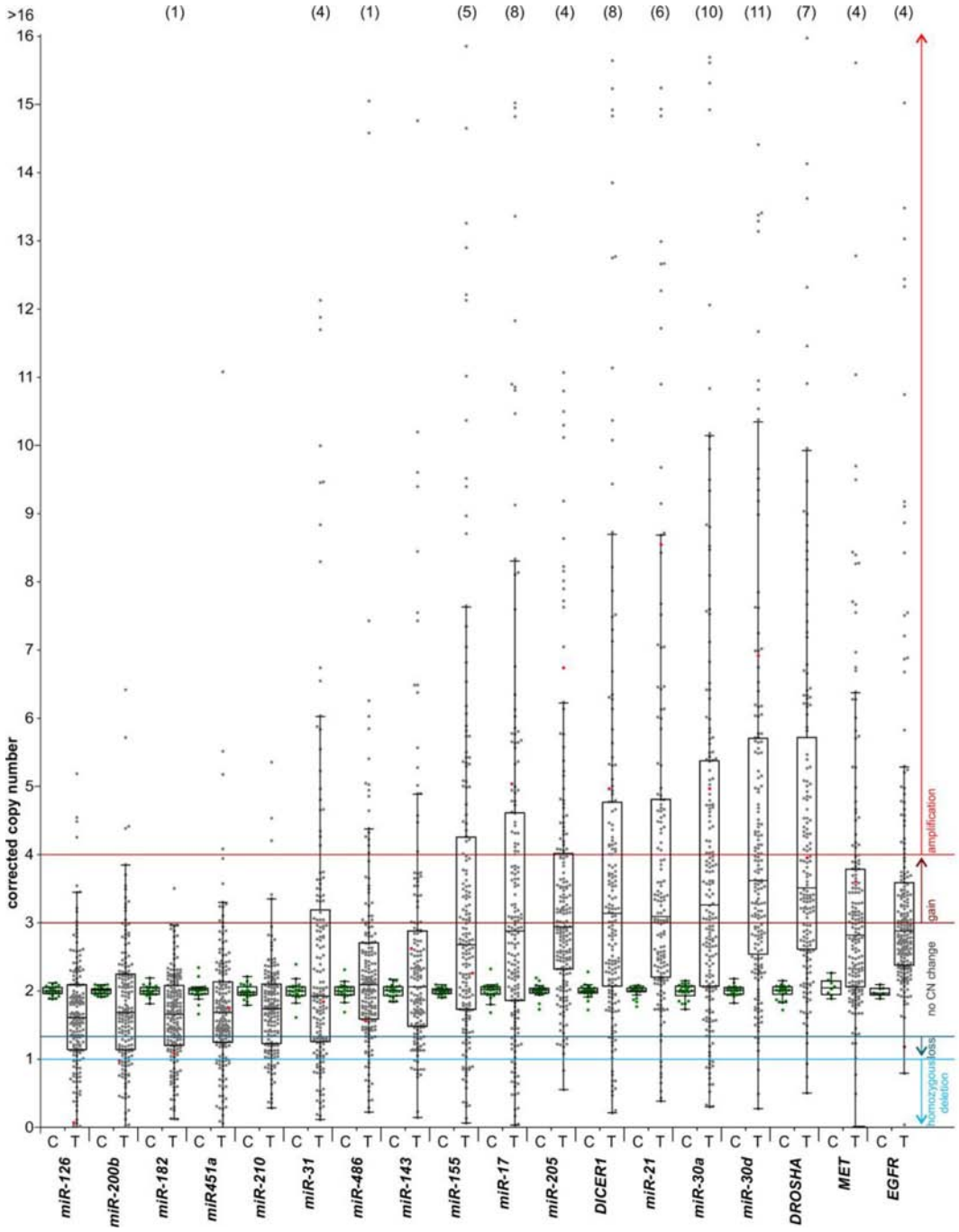
92. Sand M, Gambichler T, Skrygan M, Sand D, Scola N, Altmeyer P, Bechara FG. Expression levels of the microRNA processing enzymes Drosha and dicer in epithelial skin cancer. Cancer Invest. 2010; 28:649–653.

93. Avery-Kiejda KA, Braye SG, Forbes JF, Scott RJ. The expression of Dicer and Drosha in matched normal tissues, tumours and lymph node metastases in triple negative breast cancer. BMC Cancer. 2014; 14:253.

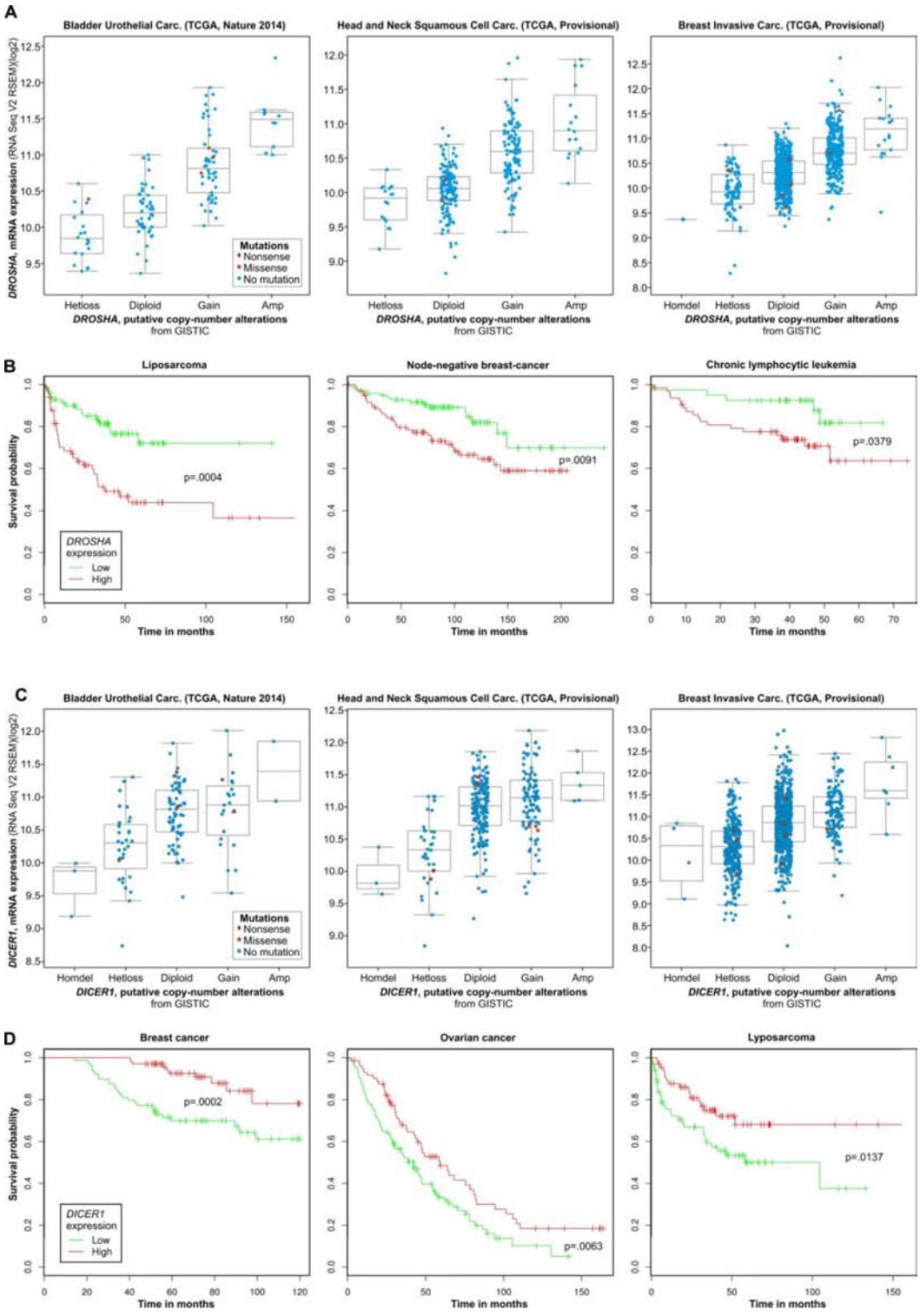
MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Czubak K. i wsp., Oncotarget 2015

SUPPLEMENTARY FIGURES AND TABLES



Supplementary Figure S1: Graphical summary of the copy number variation of the analyzed genes in NSCLC samples. The graph shows the results of copy number analysis of the selected miRNA and miRNA biogenesis genes as well as two lung cancer related oncogenes, *MET* and *EGFR*. The y-axis shows the copy number value corrected for PTC (dilution of cancer cells). Numbers in brackets (above graph) indicate samples with a copy number value >16. All other Figure legend details are the same as Figure 3A.



Supplementary Figure S2: Computational analysis of clinical (survival) and oncogenomic data of *DROSHA* and *DICER1* performed with the use of datasets representing different types of cancer. Mutual relation of copy number and expression (oncogenomic data) of *DROSHA* **A** and **B.** and *DICER1* **C** and **D.**, and the relation of their expression to survival of cancer patients (clinical data). A) and C) Correlation analysis of copy number categories and expression level, performed with the use of datasets (indicated above the graphs) deposited and tools available in cBioPortal for Cancer Genomics. B) and D) Survival analysis performed with the use of datasets (indicated above the graphs; GEO: GSE30929, GSE11121, GSE22762, GSE24450, GSE26712, GSE30929) deposited in and tools available from the PPISURV web-portal.

Supplementary Table S1. MLPA assays - detailed characteristics.

Supplementary Table S2. Relative copy number values of the analyzed regions.

Supplementary Table S3. Comparison of clinical data with copy number categories of analyzed miRNA and miRNA biogenesis genes.

Supplementary Table S1. MLPA assays - detailed characteristics

LC-miR_1 assay

probe ID	ligation position	5'PPS	length	5'SS	length	5'TSS	length	Tm (°C)	3'TSS	length	Tm (°C)	3'SS	length	3'PSS	length	5'HPS	5'HPL	3'HPS	3'HPL	TPS	TPL
ctrl_1	chr22:30,069,316	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGGA	21	75.6	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGAGGCAAAACTTCTGGCCCAGAACCGAGGAGGA	46	GGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	47	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGAGGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	93
DROSHA_1	chr5:31,410,887	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	GCCAAGGTCTTGGTGCGAAGC	21	73.2	GCCACAGGCCTCTTGGTCTTG	21	72.3	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactaGCCAAGGTCTTGGTGCGAAGC	48	GCCACAGGCCTCTTGGTCTTGGTCTTGGTGCGAAGCGCCCACAGGCCTCTTGGTCTTGctacTCTAGATTGGATCTTGCTGGCGC	48	GGGTTCCCTAAGGGTTGGAcgctactaGCCAAGGTCTTGGTGCGAAGCGCCCACAGGCCTCTTGGTCTTGctacTCTAGATTGGATCTTGCTGGCGC	96
miR-210_1	chr11:567,488	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	TCCAAAGGCCAGTCCGAGTCCAT	23	73.0	GCCAGGCGTCCACTAAATCTGC	22	72.0	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactaTCCAAAGGCCAGTCCGAGTCCATGCCAGGCGTCCACTAAATCTGCctacTCTAGATTGGATCTTGCTGGCGC	50	GCCAGGCGTCCACTAAATCTGCctacTCTAGATTGGATCTTGCTGGCGC	49	GGGTTCCCTAAGGGTTGGAcgctactaTCCAAAGGCCAGTCCGAGTCCATGCCAGGCGTCCACTAAATCTGCctacTCTAGATTGGATCTTGCTGGCGC	99
miR-30d_1	chr8:135,815,796	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	GTTGGAAAGCCACCTGAGTCTGAAG	25	71.8	CAACTGACCCCCATCCCAAATAA	24	72.3	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactGTTGGAAAGCCACCTGAGTCTGAAG	51	CAACTGACCCCCATCCCAAATAActacTCTAGATTGGATCTTGCTGGCGC	51	GGGTTCCCTAAGGGTTGGAcgctactGTTGGAAAGCCACCTGAGTCTGAAGCAACTGACCCCCATCCCAAATAActacTCTAGATTGGATCTTGCTGGCGC	102
miR-155_1	chr21:26,944,245	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	CTCTGCCCAGTGCCTTGGCTAATA	24	71.9	GACTGGCCTTTGTTGTCTTGACCAC	26	71.3	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactCTCTGCCCAGTGCCTTGGCTAATA	53	GACTGGCCTTTGTTGTCTATGACCACtacTCTAGATTGGATCTTGCTGGCGC	52	GGGTTCCCTAAGGGTTGGAcgctactactCTCTGCCCAGTGCCTTGGCTAATAGACTGGCCTTTGTTGTCTATGACCACtacTCTAGATTGGATCTTGCTGGCGC	105
miR-21_1	chr17:57,918,134	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	GGCTGCTGCATACTGCTAAATGGC	24	71.7	ACCTCTGGGATTGGCCTACCTGG	23	73.0	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactaGGCTGCATCACTGCTAAATGGC	54	ACCTCTGGGATTGGCCTACCTGGaaatctacTCTAGATTGGATCTTGCTGGCGC	54	GGGTTCCCTAAGGGTTGGAcgctactactaGGCTGCTGCATACTGCTAAATGGCACCTCTGGGATTGGCCTACCTGGaaatctacTCTAGATTGGATCTTGCTGGCGC	108
ctrl_2	chr1:156,105,841	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	CTGGACATCCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGACGAGTACCAGGAGCTT	55	CTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGC	56	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTTCTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	111
miR-30a_1	chr6:72,113,379	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	CTTTAGCCTTCTGTTGGGTTAACCTGAAGAA	31	72.2	GTAATCCCAGCAAGTGTTTCCAAGATGTG	29	72.6	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactCTTTAGCCTTCTGTTGGGTTAACCTGAAGAAGTAATCCCAGCAAGTGTTTCCAAGATGTG	57	GTAATCCCAGCAAGTGTTTCCAAGATGTGtctacTCTAGATTGGATCTTGCTG	57	GGGTTCCCTAAGGGTTGGAcgctactCTTTAGCCTTCTGTTGGGTTAACCTGAAGAAGTAATCCCAGCAAGTGTTTCCAAGATGTGtctacTCTAGATTGGATCTTGCTGGCGC	114
miR-182_1	chr7:129,409,814	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	GCCCAGGAAGGAACCACTCTTAGT	24	71.0	GACCTGTGGAAGGTTAGGAGACCC	24	72.3	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtGCCCAGGAAGGAACCACTCTTAGTGACCTGTGGAAGGTTAGGAGACCCactaaatctacTCTAGATTGGATCTTGCTGGCGC	59	GACCTGTGGAAGGTTAGGAGACCCactaaatctacTCTAGATTGGATCTTGCTGGCGC	58	GGGTTCCCTAAGGGTTGGAcgctactactattagtGCCCAGGAAGGAACCACTCTTAGTGACCTGTGGAAGGTTAGGAGACCCactaaatctacTCTAGATTGGATCTTGCTGGCGC	117
DROSHA_2	chr5:31,472,248	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	CCAGCTCCTCCCCACTGAAGCATA	23	71.9	TTGGCAATCTCCTCCTCAGGCA	22	71.0	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagCCAGCTCCTCCCCACTGAAGCATA	60	TTGGCAATCTCCTCCTCAGGCAtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	60	GGGTTCCCTAAGGGTTGGAcgctactactattagtagCCAGCTCCTCCCCACTGAAGCATATTGGCAATCTCCTCCTCAGGCAtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	120
ctrl_5	chr2:109,545,837	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72.8	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaatAGTCCTGTGGCTACGGCACCAAAGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	62	AGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	62	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaatAGTCCTGTGGCTACGGCACCAAAGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	124
miR-126_2	chr9:139,565,840	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	CCCGGGTACAGGCTCCGTAGT	21	73.1	CTCCTGGGACATGGTCCCGAC	21	74.7	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgaCCCGGGTACAGGCTCCGTAGTCTCCTGGGACATGGTCCCGACaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	64	CTCCTGGGACATGGTCCCGACaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	64	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgaCCCGGGTACAGGCTCCGTAGTCTCCTGGGACATGGTCCCGACaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	128
miR-30a_2	chr6:72,114,604	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CTTAGTACTTGGCATTCATATACCACTGCTGAC	33	71.2	TAGAGCTTGAGTCAATCCACCAAACTCAGCTATct	33	72.3	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattaCTTAGTACTTGGCATTCATATACCACTGCTGACTAGAGCTTGAGTCAATCCACCAAACTCAGCTATctaaatctacTCTAGATTGGATCTTGCTGGCGC	66	TAGAGCTTGAGTCAATCCACCAAACTCAGCTATctaaatctacTCTAGATTGGATCTTGCTGGCGC	66	GGGTTCCCTAAGGGTTGGAcgctactactattaCTTAGTACTTGGCATTCATATACCACTGCTGACTAGAGCTTGAGTCAATCCACCAAACTCAGCTATctaaatctacTCTAGATTGGATCTTGCTGGCGC	132
miR-30d_2	chr8:135,819,855	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	GCAGCTGGGAAGGTAAACACAAAGA	25	70.8	GTAGTAGCAGGGTGGTGACAGGGA	24	72.2	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgaCAGCTGGGAAGGTAAACACAAAGAGTAGTAGCAGGGTGGTGACAGGGAtaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	68	GTAGTAGCAGGGTGGTGACAGGGAtaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	68	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgaCAGCTGGGAAGGTAAACACAAAGAGTAGTAGCAGGGTGGTGACAGGGAtaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	136
miR-155_2	chr21:26,946,400	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgg	26	GCATTCACATGGAACAAATTGCTGC	25	70.9	CGTGGGAGGATGACAAGAAGCATctaatggtcaaactaaatctac	24	71.4	tctaatggtcaaactaaatctac	23	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgGCATTCACATGGAACAAATTGCTGCtaatggtcaaactaaatctac	70	CGTGGGAGGATGACAAGAAGCATctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	70	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgGCATTCACATGGAACAAATTGCTGCtaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	140
ctrl_3	chr17:3,397,683	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	TATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctac	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgTCCCTGCGCCATTGAGGTCTATAAAATTATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctac	72	TATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	72	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgTCCCTGCGCCATTGAGGTCTATAAAATTATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	144
miR-21_2	chr17:57,920,168	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgat	25	GGTTTTTAGTTTCATTTGGATAGTGGCCAC	30	71.0	CTCTCCATTCCCTACGAATGCTAGAATGA	29	71.6	ctaatggtcaaactaaatctac	22	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatGGTTTTTAGTTTCATTTGGATAGTGGCCACCAATGCTAGAATGActaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	74	CTCTCCATTCCCTACGAATGCTAGAATGActaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	74	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatGGTTTTTAGTTTCATTTGGATAGTGGCCACCAATGCTAGAATGActaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	148
miR-210_2	chr11:569,809	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttc	36	GGGAAGGTGCCCAGAGGATCA	31	73.4	CGGAGCCACTCGGTTCTATTGG	22	72.0	ttgcgaaatgtatctaattggtcaaactaaat	31	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcGGGAAGGTGCCCAGAGGATCA	76	CGGAGCCACTCGGTTCTATTGGttgcgaaatgtatctaattggtcaaactaaat	76	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcGGGAAGGTGCCCAGAGGATCACGGAGCCACTCGGTTCTATTGGttgcgaaatgtatctaattggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	152
miR-182_2	chr7:129,415,089	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttt	35	CTTCTGGCCTTGCAGAGGAGAAGA	24	72.2	CCACACCAAGGTGTCTCCTATCTCC	25	72.5	aaatgtatctaattggtcaaactaa	30	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcCTTCTGGCCTTGCAGAGGAGAAGA	78	CCACACCAAGGTGTCTCCTATCTCCaaatgtatctaattggtcaaactaaatctac	78	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcCTTCTGGCCTTGCAGAGGAGAAGAAACACACCAAGGTGTCTCCTATCTCCaaatgtatctaattggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	156
DROSHA_3	chr5:31,526,333	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagctc	39	CTCTCACCTCGCCCATGACTGT	22	71.8	GATCTCGGTGCCTGTGGCATCATC	22	71.9	ttgcgaaatgtatctaattggtcaaactaaatctac	35	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcagcCTCTCACCTCGCCCATGACTGTGATCTCGGTGCCTGTGGCATCATCttgcgaaatgtatctaattggtcaaactaaatctac	80	GATCTCGGTGCCTGTGGCATCATCttgcgaaatgtatctaattggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	80	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcagcCTCTCACCTCGCCCATGACTGTGATCTCGGTGCCTGTGGCATCATCttgcgaaatgtatctaattggtcaaactaaatctac	160
miR-126_1	chr9:139,564,281	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagctcg	42	CCGAGCAAAGCACCCACTCCTAG	21	73.7	CGCAGAGCACCCACTCCTAG	21	73.0	catttgcgaaatgtatctaattggtcaaactaaatctac	38	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcagctcgCCGAGCAAAGCACCACCTTGC	82	CGCAGAGCACCCACTCCCTAGcatttgcgaaatgtatctaattggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	82	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcagctcgCCGAGCAAAGCACCACCTTGCCGCAGAGCACCCACTCCCTAGcatttgcgaaatgtatctaattggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	164

LC-miR_2 assay

probe ID	ligation position	5'PPS	length	5'SS	length	5'TSS	length	Tm (°C)	3'TSS	length	Tm (°C)	3'SS	length	3'PSS	length	5'HPS	5'HPL	3'HPS	3'HPL	TPS	TPL
ctrl_1	chr22:30,069,316	GGGTTCCCTAAGGGTTGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGGA	21	75.6	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGA	46	GGCAAAACTTCTGGCCCAGAACTCTAGATTGGATCTTGCTGGCGC	47	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	93
DICER1_1	chr14:95,557,143	GGGTTCCCTAAGGGTTGGA	19	cgct	4	GGAAACACGCGTTACGACTTACTGC	25	71.4	AATTGCATACCCCCGACGACAGG	24	72.0	c	1	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctGGAAACACGCTACGACTTACTGC	48	AATTGCATACCCCCGACAGACAGGcTCTAGATTGGATCTTGCTGGCGC	48	GGGTTCCCTAAGGGTTGGAcgctGGAAACACGCGTTACGACTTACTGCAATTGCATACCCCCGACAGACAGGcTCTAGATTGGATCTTGCTGGCGC	96
miR-200b_1	chr1:1,102,025	GGGTTCCCTAAGGGTTGGA	19	cgctactac	9	CGTTCTGTCTCGAGAGCCTCGC	22	73.4	AGACACCGGGCCTTTGAGAAGAG	23	71.2	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactacCGTTCTGTCTCGAGAGCCTCGC	50	AGACACCGGGCCTTTGAGAAGAGtacTCTAGATTGGATCTTGCTGGCGC	49	GGGTTCCCTAAGGGTTGGAcgctactacCGTTCTGTCTCGAGAGCCTCGCAGACACCGGGCCTTTGAGAAGAGtacTCTAGATTGGATCTTGCTGGCGC	99
miR-143_1	chr5:148,808,365	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	CAGGCCACAGACAGGAAACACAGTT	25	72.6	GTGAGGAATTACAACAGCCTCCCG	24	71.9	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactCAGGCCACAGACAGGAAACACAGTT	51	GTGAGGAATTACAACAGCCTCCCGctacTCTAGATTGGATCTTGCTGGCGC	51	GGGTTCCCTAAGGGTTGGAcgctactCAGGCCACAGACAGGAAACACAGTTGTGAGGAATTACAACAGCCTCCCGctacTCTAGATTGGATCTTGCTGGCGC	102
miR-17_1	chr13:92,002,604	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	GGAGAGGCCAGCCATTGGAAGA	22	73.5	GCCACCACTTCCAGTGCAGTTG	23	71.1	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactatGGAGGCCAGCCATTGGAAGA	53	GCCACCACTTCCAGTGCTAGTTGatctacTCTAGATTGGATCTTGCTGGCGC	52	GGGTTCCCTAAGGGTTGGAcgctactactatGGAGAGGCCAGCCATTGGAAGACCACTTCCAGTGCTAGTTGatctacTCTAGATTGGATCTTGCTGGCGC	105
miR-31_1	chr9:21,510,188	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	AGGATGGCGCTTGGCAAGCAATG	23	74.2	TTCATTTCCACACATGGACCCAGC	25	71.9	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagAGGATGGCGCTTGGCAAGCAATG	57	TTCATTTCCACACATGGAACCCAGCatctacTCTAGATTGGATCTTGCTGGCGC	54	GGGTTCCCTAAGGGTTGGAcgctactactattagAGGATGGCGCTTGGCAAGCAATGTTCATTTCCACACATGGAACCCAGCatctacTCTAGATTGGATCTTGCTGGCGC	108
ctrl_2	chr1:156,105,841	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTT	55	CTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	56	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTTCTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	111
miR-451a_1	chr17:27,188,349	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	GAGCTGGAGTCTCCCTTCCAGTAGT	25	71.4	CTGCTTCAGGGTCTCAGAGTTCTCTT	25	71.1	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattGAGCTGGAGTCTCCCTTCCAGTAGT	57	CTGCTTCAGGGTCCTGAGTTCTCTTtaaatctacTCTAGATTGGATCTTGCTGGCGC	57	GGGTTCCCTAAGGGTTGGAcgctactactattGAGCTGGAGTCTCCCTTCCAGTAGTCTGCTTCAGGGTCCTGAGTTCTCTTtaaatctacTCTAGATTGGATCTTGCTGGCGC	114
miR-205_1	chr1:209,605,273	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	CAGACCCACCTCCCCAACTCAG	22	74.1	CACTGTCTGTCTGTGCAGCAGGT	23	71.1	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagC AGACCCACCTCCCCAACTCAG	59	CACTGTCTGTCTGTGCAGCAGGTaactaaatctacTCTAGATTGGATCTTGCTGGCGC	58	GGGTTCCCTAAGGGTTGGAcgctactactattagtagCAGACCCACCTCCCCAACTCAGCACTGTCTGTCTGTGCAGCAGGTaactaaatctacTCTAGATTGGATCTTGCTGGCGC	117
DICER1_2	chr14:95,577,684	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	GGTAGCACTGCCTTCGTTTCGTG	23	71.5	GAACCTGGTCTTCCTGGAACACTG	24	71.5	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagGGTAGCACTGCCTTCGTTTCGTG	60	GAACCTGGTCTTCCTGGAACACTGaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	60	GGGTTCCCTAAGGGTTGGAcgctactactattagtagGGTAGCACTGCCTTCGTTTCGTGGAACCTGGTCTTCCTGGAACACTGaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	120
ctrl_5	chr2:109,545,837	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72.8	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaTAGTCCTGTGGCTACGGCACCAA	62	AGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	62	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaTAGTCCTGTGGCTACGGCACCAAGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	124
miR-486_2	chr8:41,517,622	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaaa	20	CAGGACCCAACAGCAGCTACTCAGA	25	73.3	GTCTTCTAAACCAGCAGCTAGTGGG	26	72.0	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaCAGGACCCAACAGCAGCTACTCAGA	64	GTCTTCTAAACCAGCAGCCCTAGTGGGtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	64	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaCAGGACCCAACAGCAGCTACTCAGAGTCTTCTAAACCAGCAGCCTAGTGGGtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	128
miR-200b_2	chr1:1,104,685	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgg	26	CTGTAGTCCCCGGGCAGATGC	21	74.0	TGTGGACCCCGTGCTATCAGC	21	71.5	ctaatggtcaaactaaatctac	22	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgCTGTAGTCCCCGGGCAGATGC	66	TGTGGACCCCGTGCTATCAGCctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	66	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgCTGTAGTCCCCGGGCAGATGCTGTGGACCCCGTGCTATCAGCctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	132
miR-143_2	chr5:148,810,997	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	CTCTCCATTGTTGCACCACCAGC	23	72.9	AATGCCCCCATCACTTCCATCACTTCCTCATT	25	71.0	ctaatggtcaaactaaatct	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgCTCTCCATTGTTGCACCACCAGC	68	AATGCCCCCATCACTTCCTCTCATTctaatggtcaaactaaatctTCTAGATTGGATCTTGCTGGCGC	68	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgCTCTCCATTGTTGCACCACCAGCAATGCCCCCATCACTTCCTCTCATTctaatggtcaaactaaatctTCTAGATTGGATCTTGCTGGCGC	136
miR-17_2	chr13:92,003,414	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgc	27	AGCTGTAGAACTCCAGCTCGGCC	24	72.1	TGTCGCCCAATCAAACTCCTCATT	25	71.8	ctaatggtcaaactaaatct	22	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgcAGCTGTAGAACTCCAGCTTCGGCC	70	TGTCGCCCAATCAAACTGTCCTGTTctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	70	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgcAGCTGTAGAACTCCAGCTTCGGCCTGTCGCCCAATCAAACTGTCCTGTTctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	140
ctrl_3	chr17:3,397,683	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	TATAGAGAAAGTTGATTACCCCCGGGATGAATGGTCAAACTAAATCTACTCTAGATTGGATCTTGCTGGCGC	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgTCCCTGCGCCATTGAGGTCTATAAAAT	72	TATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	72	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgTCCCTGCGCCATTGAGGTCTATAAAATTATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	144
miR-31_2	chr9:21,512,347	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	CCTTAGAGTGTCCGCTTTTGAAGATGTAG	29	70.6	TTAGGCACAGGAGGTTTGTGGGTAGTTC	28	72.6	tctaatggtcaaactaaatctac	23	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgCCTTAGAGTGTCCGCTTTTGAAGATGTAG	74	TTAGGCACAGGAGGTTTGTGGGTAGTTCtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	74	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgCCTTAGAGTGTCCGCTTTTGAAGATGTAGTAGTTCCCGCTTTTGAAGATGTAGAGCTTCTTGGGCTTGCaatgtatcctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	148
miR-451a_2	chr17:27,188,489	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacct	32	GGCAGTCAGTAGGTTGTGACAGGCT	25	72.8	GAGCAGAGAGCTTCTTGGGCTTGC	24	73.7	aatgtatctaatggtcaaactaa atctac	29	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgccacctGGCAGTCAGTAGGTTGTGACAG	76	GAGCAGAGAGCTTCTTGGGCTTGCaatgtatcctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	76	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctGGCAGTCAGTAGGTTGTGACAGGCTGAGCAGAGAGCTTCTTGGGCTTGCaatgtatcctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	152
miR-205_2	chr1:209,606,077	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttca	37	GCAAGGAACCCTGCTGCCCTTA	22	73.2	GACTCCCCACAGAGCAAGCAGA	22	72.0	gcgaaatgtatctaatggtcaaactaaatctac	33	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgccacctttttcaGCAAGGAACCCTGCTGCCCTTA	78	GACTCCCCACAGAGCAAGCAGAGcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	78	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcaGCAAGGAACCCTGCTGCCCTTAGACTCCCCACAGAGCAAGCAGAGcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	156
DICER1_3	chr14:95,607,817	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttc	36	GACATCAGACTTGGTTTGCGCTCAG	25	72.5	GTCTGAGCCTTCTGCAATGCCTC	23	71.8	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgccacctttttcGACATCAGACTTGGTTTGCGCTCAG	80	GTCTGAGCCTTCTGCAATGCCTCtgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	80	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcGACATCAGACTTGGTTTGCGCTCAGtgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	160
miR-486_1	chr8:41,518,110	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagct	40	GAGATGGTCGCACAGCGTCTTCA	23	73.2	TGTCCCTTCCTCATGGCAGCTTCA	21	73.0	catttgcgaaatgtatctaatggtcaaactaaatctac	38	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaa ttgatgccacctttttcagctGATGGTCGCACAGCGTCTTCA	82	TGTCCCTTCCTCATGGCACGGcatttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	82	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttttcagctGAGATGGTCGCACAGCGTCTTCATGTCCCTTCCTCATGGCACGGcatttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	164

LC-5p assay

probe ID	ligation position	5'PPS	length	5'SS	length	5'TSS	length	Tm (°C)	3'TSS	length	Tm (°C)	3'SS	length	3'PSS	length	5'HPS	5'HPL	3'HPS	3'HPL	TPS	TPL
ctrl_1	chr22:30,069,316	GGGTTCCCTAAGGGTTGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGA	46	GGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	47	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGAGGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	93
DROSHA_1	chr5:31,410,887	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	GCCAAGGTCTTGGTGCGAAGC	21	73.2	GCCACAGGCCTCTTGGTCTTG	21	72.3	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactaGCCAAGGTCTTGGTGCGAAGC	48	GCCACAGGCCTCTTGGTCTTGctacTCTAGATTGGATCTTGCTGGCGC	48	GGGTTCCCTAAGGGTTGGAcgctactaGCCAAGGTCTTGGTGCGAAGCGCCACAGGCCTCTTGGTCTTGctacTCTAGATTGGATCTTGCTGGCGC	96
5p_1.2M	chr5:1,244,805	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	CTGCTGACCATCTTTGTGGCTTACA	25	70.6	TCATCCTCCTGTTCTGGAAGCCAC	24	71.6	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctacCTGCTGACCATCTTTGTGGCTTACA	50	TCATCCTCCTGTTCTGGAAGCCACacTCTAGATTGGATCTTGCTGGCGC	49	GGGTTCCCTAAGGGTTGGAcgctacCTGCTGACCATCTTTGTGGCTTACATCATCCTCCTGTTCTGGAAGCCACacTCTAGATTGGATCTTGCTGGCGC	99
DROSHA_4	chr5:31,431,731	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	CATTGCATGTCGAAGGTCCGATTC	24	71.9	TCTACAAAGGTCAGGCCGTGAG	23	70.9	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactaCATTGCATGTCGAAGGTCCGATTC	51	TCTACAAAGGTCAGGCCCGTGAGtctacTCTAGATTGGATCTTGCTGGCGC	51	GGGTTCCCTAAGGGTTGGAcgctactaCATTGCATGTCGAAGGTCCGATTCGGGTTCCCTAAGGGTTGGAcgctactaCATTGCATGTCGAAGGTCCGATTCGTGAGtctacTCTAGATTGGATCTTGCTGGCGC	102
5p_10.2M	chr5:10,236,639	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	TGGCAGATCCATGCACACCTTCTC	24	72.5	GCCAAGCGCGGTATCTGGAATAC	23	71.9	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactTGGCAGATCCATGCACACCTTCTC	53	GCCAAGCGCGGTATCTGGAATACatctacTCTAGATTGGATCTTGCTGGCGC	52	GGGTTCCCTAAGGGTTGGAcgctactactTGGCAGATCCATGCACACCTTCTCGCCAAGCGCGGTATCTGGAATACatctacTCTAGATTGGATCTTGCTGGCGC	105
5p_45.4M	chr5:45,396,650	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	GACTGGATTAAAGCGGTGGCATG	23	70.8	GCCGACAAACATGGCATGCAG	22	70.8	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactatGACTGGATTAAAGCGGTGGCATG	54	GCCGACAAACATGGCATAGCAGtaaatctacTCTAGATTGGATCTTGCTGGCGC	54	GGGTTCCCTAAGGGTTGGAcgctactactatGACTGGATTAAAGCGGTGGCATGGCCGACAAACATGGCATAGCAGtaaatctacTCTAGATTGGATCTTGCTGGCGC	108
ctrl_2	chr1:156,105,841	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTT	55	CTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	56	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTTCTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	111
5p_19.7M	chr5:19,721,515	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	GGAGAATGCTGTAAACCACCCGAG	24	71.9	CGCTGTTTCCATAGGTAGGGTCATC	25	70.9	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattaGGAGAATGCTGTAAACCACCCGAGGGTCATC	57	CGCTGTTTCCATAGGTAGGGTCATCtaaatctacTCTAGATTGGATCTTGCTGGCGC	57	GGGTTCCCTAAGGGTTGGAcgctactactattaGGAGAATGCTGTAAACCACCCGAGCGCTGTTTCCATAGGTAGGGTCATCtaaatctacTCTAGATTGGATCTTGCTGGCGC	114
GOLPH3_5'	chr5:32,010,496	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	GAGCTTGCCAGATCTGACCAGCT	23	71.6	CGGTAGAAGATGTGTCCTCCTGG	23	70.4	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagGAGCTTGCCAGATCTGACCAGCT	59	CGGTAGAAGATGTGTCCTCCTGGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	58	GGGTTCCCTAAGGGTTGGAcgctactactattagtaGAGCTTGCCAGATCTGACCAGCTCGGTAGAAGATGTGTCCTCCTGGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	117
DROSHA_2	chr5:31,472,248	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	CCAGCTCCTCCCCACTGAAGCATA	23	71.9	TTGGCAATCTCCTCCTCAGGCA	22	71.0	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagCCAGCTCCTCCCCACTGAAGCATA	60	TTGGCAATCTCCTCCTCAGGCAtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	60	GGGTTCCCTAAGGGTTGGAcgctactactattagtagCCAGCTCCTCCCCACTGAAGCATATTGGCAATCTCCTCCTCAGGCAtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	120
ctrl_5	chr2:109,545,837	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72.8	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaGTCCTGTGGCTACGGCACCAA	62	AGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	62	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaatAGTCCTGTGGCTACGGCACCAAAGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	124
5p_26.9M	chr5:26,903,788	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	CCCTTCCTGTGTATCCTTGTCAGTG	25	70.9	ATGACATCGAACATGTCTGCACCATC	26	70.5	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaCCCTTCCTGTGTATCCTTGTCAGTG	64	ATGACATCGAACATGTCTGCACCATCtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	64	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaaCCCTTCCTGTGTATCCTTGTCAGTGATGACATCGAACATGTCTGCACCATCtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	128
DROSHA_5	chr5:31,504,712	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	ATTCCTGTGCGTCTTTGCC	23	70.9	GCTGCATTTGCAGAGTGGTCCAT	23	71.5	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagattgaATTCCTGTGCGTCTTGCCTTTGCC	66	GCTGCATTTGCAGAGTGGTCCATaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	66	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgaATTCCTGTGCGTCTTGCCTTTGCCATTGCAGAGTGGTCCATaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	132
GOLPH3_e1	chr5:32,173,966	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	TTCCTCCATCAGGGTCAGCCG	21	72.4	CGTTTCCTTGGAGTCGCCTT	21	71.4	atctaatggtcaaactaaatctac	24	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagattgatgccAGGCTTATCGGGAGGAACAGCTGAT	68	CGTTTCCTTGGAGTCGCCCTTatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	68	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccTTCCTCCATCAGGGTCAGCCGCGTTTCCTTGGAGTCGCCCTTatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	136
5p_35.6M	chr5:35,659,144	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	AGGCTTATCGGGAGGAACAGCTG	23	70.6	ATTAACCGGCTGATGCGCTGAGTCGCCC	23	72.4	atctaatggtcaaactaaatctac	24	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccAGGCTTATCGGGAGGAACAGCTGATTAACCGGCTGATGCGGCAGTCatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	70	ATTAACCGGCTGATGCGGCAGTCatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	70	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccAGGCTTATCGGGAGGAACAGCTGATTAACCGGCTGATGCGGCAGTCatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	140
ctrl_3	chr17:3,397,683	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagattgatgTCCCTGCGCCATTGAGGTCTATAAAAT	72	TATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	72	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgTCCCTGCGCCATTGAGGTCTATAAAATTATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	144
GOLPH3_3'	chr5:32,239,139	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccctttt	34	GTTCTGGCGGAGATGGTTGCA	21	71.9	GCGATCCAAGAAACAGTGCC	21	70.9	aaatgtatctaatggtcaaactaaatctac	30	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagattgccaccttttGTTCTGGCGGAGATGGTTGCA	74	GCGATCCAAGAAACAGTGGCCaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	74	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccaccttttcagcCTCTCACCTCGCCCATGACTGTGATCTCGGTGCCTGTGGTCATCttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	148
DROSHA_3	chr5:31,526,333	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagc	39	CTCTCACCTCGCCCATGACTGT	22	71.8	GATCTCGGTGCCTGTGGTCATC	22	71.9	ttgcgaaatgtatctaatggtcaaactaaatctac	35	TCTAGATTGGATCTTGCTGGCGC	23	GGGTTCCCTAAGGGTTGGAcgctactactattagtagattgccaccttttcagcCTCTCACCTCGCCCATGACTGT	80	GATCTCGGTGCCTGTGGTCATCttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	80	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccaccttttcagcCTCTCACCTCGCCCATGACTGTGATCTCGGTGCCTGTGGTCATCttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	160

Legend:

5'PSS, 3'PSS - 5' and 3' primer-specific sequence, respectively
5'SS, 3'SS - 5' and 3' stuffer sequence, respectively
5'TSS, 3'TSS - 5' and 3' target-specific sequence, respectively
Tm - melting temperature
5'HPL, 3'HPL - 5' and 3' half-probe sequence
5'HPL, 3'HPL - 5' and 3' half-probe length
TPS, TPL - total probe sequence and length

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA
SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

Sequence used for generation of all 5' and 3' stuffer sequences: AC# V00604, Phage M13 genome, position 3-99
5'-cgctactactattagtagaattgatgccaccttttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccatttgcgaaatgtatctaatggtcaaactaaatctac-3'

Supplementary Table S2. Relative copy number values of the analyzed regions

SAMPLE NUMBER	tumor stage	remission (R) progression (P) metastasis (M)	sex	age at diagno-sis	sample type	PTC	relative copy number values of all regions																		EGFR mutation
							miR-126	miR-200b	miR-182	miR-451a	miR-210	miR-31	miR-486	miR-143	miR-155	miR-17	miR-205	DICER1	miR-21	miR-30a	miR-30d	DROSHA	MET	EGFR	
1	IIB	P	M	68	FFPE		1.16	1.55	1.60	2.15	1.64	2.59	2.22	2.54	2.93	2.73	3.62	3.17	2.76	3.27	2.70	2.59	2.41	2.20	
2	IV	R	M	58	FFPE		0.89	excluded	1.60	2.00	1.70	2.00	excluded	2.40	2.38	2.27	3.03	2.82	2.72	3.30	2.52	3.95	2.57	1.68	
3	IA	R	W	69	FFPE		1.06	1.74	1.52	2.25	1.64	2.10	2.45	2.77	2.48	2.32	3.04	2.86	2.73	2.52	2.97	2.40	2.12	2.43	
4	IIA	P	M	70	FFPE		1.42	1.49	1.66	2.54	1.69	2.44	2.66	2.61	2.71	2.33	3.34	3.06	3.76	3.50	3.32	3.43	2.92	2.56	
5	IV	P	M	63	FFPE		1.53	1.98	1.56	2.30	1.55	2.08	2.35	2.51	2.36	2.25	2.57	2.49	2.67	2.95	2.60	2.45	2.53	2.50	
6	IIIA	P	M	66	FFPE		1.18	1.98	1.69	2.10	1.99	1.47	3.32	2.13	1.55	1.81	2.71	2.38	2.08	2.11	3.65	3.08	2.61	2.51	
7	IIB	M	M	67	FFPE		1.24	1.88	1.70	2.67	1.71	2.51	2.22	2.54	2.49	2.48	3.14	2.80	2.69	3.10	2.71	2.96	2.63	2.33	
9	IB	R	W	74	FFPE		1.63	2.36	1.37	2.83	1.80	excluded	2.16	3.34	2.70	2.52	3.54	2.51	3.50	3.74	2.77	3.71	2.26	3.22	+
10	IA	R	M	58	FFPE		1.79	1.72	2.39	2.46	2.02	2.31	2.94	2.91	2.53	2.81	2.98	3.01	2.81	2.92	2.92	2.73	2.79	2.83	
11	IV	P	M	40	FFPE		1.56	1.62	1.78	2.80	2.11	3.63	2.60	3.08	3.42	3.80	3.30	3.89	3.99	3.78	3.08	3.13	3.06	2.47	
12		R	W	76	FFPE		1.41	1.44	1.78	2.47	1.95	3.11	2.40	3.10	3.47	3.92	3.69	4.14	3.55	4.13	3.25	3.16	2.99	2.40	
13	IIIB	P	M	65	FFPE		1.51	1.76	excluded	1.94	1.87	1.88	1.87	2.16	2.32	2.55	2.26	2.30	2.73	2.89	1.80	2.43	2.20	2.14	
14			M	53	FFPE		1.66	2.04	1.85	2.07	1.97	2.37	2.23	2.31	2.18	2.53	2.51	2.48	2.53	2.72	2.09	2.28	2.28	2.15	
15			M	78	FFPE		excluded	1.66	1.62	2.23	2.05	2.93	2.16	2.78	3.67	3.29	2.89	3.36	4.70	3.98	2.93	2.80	2.70	2.12	
16	IIB	R	M	76	FFPE		1.73	excluded	1.37	2.45	2.08	4.08	3.29	4.6	3.21	5.22	3.28	4.59	3.26	4.99	5.94	4.63	2.99	2.38	
17	IIB	P	M	58	FFPE		1.20	1.56	1.55	2.31	1.50	2.15	2.26	2.49	2.86	2.44	2.74	2.98	3.29	4.23	2.70	2.74	2.29	2.59	
18	IV	P	M	56	FFPE		1.07	1.60	1.49	2.11	1.76	2.66	1.96	2.55	2.87	3.04	2.62	2.96	3.35	3.45	2.35	2.51	2.49	2.50	
19	IV	P	M	76	FFPE		1.19	1.69	1.31	2.37	2.07	2.84	excluded	3.01	5.13	4.94	3.73	4.52	5.56	5.97	6.99	3.87	2.55	2.93	+
20	IV	M	M	54	FFPE		1.27	1.78	1.27	2.06	2.02	2.70	1.94	2.21	2.86	3.23	2.92	3.24	3.34	3.94	2.46	2.86	2.19	2.10	
21	IV	R	W	52	FFPE	15%	excluded	1.57	1.55	3.36	1.95	4.36	excluded	3.91	5.23	6.14	4.59	5.91	5.79	5.21	5.48	excluded	4.04	2.78	
22	IIIA	R	M	81	FFPE	50%	1.37	1.49	2.15	1.85	1.32	2.52	2.82	2.25	2.06	2.43	2.70	2.60	1.95	1.96	2.64	2.60	3.00	2.86	
23			M	59	cyto		1.71	1.30	1.80	1.72	1.51	2.16	1.81	1.65	1.64	2.34	3.35	1.58	2.92	3.31	4.67	2.72	2.62	2.27	
24	IIIB	P	W	59	FFPE	30%	1.70	1.68	1.93	1.89	1.77	4.23	2.13	2.28	5.37	4.63	2.84	4.50	4.91	6.07	5.33	4.23	3.12	2.44	
25	IIIB	P	M	61	FFPE	40%	1.64	1.79	1.73	1.82	1.65	2.41	2.95	1.99	3.92	4.44	3.12	5.22	3.60	4.72	5.57	4.72	2.92	2.36	
26	IV	M	M	60	FFPE		1.25	2.08	1.59	2.21	1.54	1.99	6.28	2.29	2.49	1.80	2.64	2.71	3.25	3.37	3.10	3.10	2.40	2.69	
27	IIA	P	M	64	FFPE	10%	1.50	1.55	1.74	2.13	1.83	2.23	2.42	2.33	2.35	2.98	2.62	2.59	3.02	2.61	2.88	2.64	2.75	2.47	
28	IV	M	W	62	FFPE		1.29	1.74	1.25	1.78	2.02	2.56	2.21	1.61	2.21	2.80	2.07	1.98	3.44	3.13	2.49	4.80	1.90	2.10	
29	IV	M	W	55	FFPE		1.67	1.65	1.56	1.97	1.87	2.21	1.92	2.20	2.11	2.36	2.77	2.45	3.24	2.41	2.73	3.20	2.54	2.53	
30			W	60	cyto		1.94	2.01	1.57	1.67	1.69	0.84	1.51	1.23	1.18	1.35	1.92	1.52	1.92	1.88	2.11	3.03	2.27	2.71	
31	IV	M	W	63	FFPE	60%	1.71	1.97	1.75	2.27	2.17	1.97	2.18	3.48	2.81	2.63	2.78	5.07	3.61	3.15	3.47	4.79	2.24	3.31	
32	IV	M	M	53	FFPE	50%	1.74	1.56	1.86	1.52	1.89	2.06	1.78	1.59	1.44	1.48	1.86	1.91	1.77	1.66	2.44	2.30	2.17	2.31	
33			W	56	cyto		excluded	1.33	excluded	1.53	excluded	0.71	2.44	0.94	excluded	1.09	1.22	1.59	excluded	excluded	excluded	excluded	1.05	1.09	
34	IV	M	W	54	FFPE		1.61	1.72	1.54	2.07	1.72	4.43	2.22	2.51	7.04	5.74	3.17	5.06	5.16	6.84	6.97	excluded	2.89	2.90	
35			M	70	FFPE	10%	1.42	1.35	1.76	2.01	1.59	1.79	2.54	1.79	2.53	2.63	2.72	3.32	4.26	3.86	3.14	4.24	2.64	1.78	
36	IV	M	M	59	FFPE		1.40	1.94	1.78	3.47	1.63	3.77	2.68	3.26	5.01	4.99	4.03	4.62	8.95	7.61	11.02	excluded	4.00	2.84	
37	IA	P	W	67	FFPE	30%	1.89	1.26	1.87	2.02	1.71	2.49	1.84	2.30	2.90	3.14	2.72	3.24	3.52	3.11	4.09	3.43	2.89	2.40	
38	IV	P	W	48	FFPE	70%	1.65	1.14	1.61	1.74	2.13	2.36	2.72	2.43	3.01	2.98	3.39	3.08	3.29	2.67	4.19	3.98	2.65	2.23	
39	IV	P	M	60	cyto		2.28	1.77	2.31	1.62	1.94	1.61	2.17	1.44	1.87	2.12	1.95	2.00	2.08	2.09	1.82	2.39	2.55	2.29	
40			W	53	cyto		1.11	2.51	1.84	1.95	1.77	2.16	0.96	2.30	2.82	2.45	2.79	1.52	2.21	2.31	2.66	2.59	2.03	3.05	+
41			M	63	FFPE	40%	1.59	1.66	1.66	1.84	1.57	1.94	2.25	excluded	2.30	2.61	2.19	2.97	3.15	2.57	4.25	2.96	2.60	2.28	
42			M	57	FFPE	30%	1.90	1.64	1.77	1.84	1.70	2.22	2.25	excluded	2.67	2.73	2.70	3.09	2.62	3.11	3.21	2.92	2.22	2.42	
43			W	49	FFPE	60%	1.55	1.43	1.42	1.64	1.70	2.79	2.42	excluded	3.18	3.39	3.61	3.80	4.0	3.31	4.09	4.82	3.26	3.25	
44	IV	P	W	67	FFPE		1.86	1.53	1.89	2.72	2.37	1.91	excluded	2.96	3.42	3.62	2.90	3.97	4.94	3.67	3.07	3.10	3.51	3.02	+
45	IV	M	M	68	FFPE	40%	1.76	2.40	1.91	1.76	2.30	1.87	2.32	2.40	2.08	1.86	2.48	1.77	2.21	2.26	2.44	3.06	2.57	2.56	+
46			W	53	FFPE		1.89	1.84	1.86	1.99	1.79	excluded	2.02	2.21	1.45	1.80	2.49	1.92	2.04	1.72	1.89	2.49	2.10	2.34	
47			W	55	FFPE	30%	1.86	1.82	1.90	2.13	1.76	3.06	2.24	2.27	2.27	2.87	2.31	2.60	2.76	2.71	2.40	2.68	2.80	2.52	
48	IIIA	M	W	50	FFPE	80%	1.92	1.65	1.77	2.70	2.15	2.25	excluded	2.99	2.87	3.72	3.24	3.71	3.98	3.69	2.87	3.33	2.99	2.51	
49	IV	M	W	61	FFPE		1.73	1.97	1.73	2.09	2.27	excluded	2.10	2.19	1.95	2.33	2.45	2.41	2.39	2.08	2.11	2.51	2.45	2.43	+
50			W	55	FFPE	20%	1.97	2.06	1.89	2.06	2.05	excluded	2.00	2.07	2.66	2.95	2.67	3.10	2.80	3.12	3.12	3.39	2.94	3.74	+
51	IV	M	W	52	FFPE		2.54	1.49	1.56	2.25	2.61	1.99	2.28	2.11	2.33	3.19	4.86	2.02	3.67	3.33	2.97	2.85	2.79	2.57	+
52			M	64	FFPE		1.93	1.77	1.80	1.78	2.03	2.36	2.45	2.18	2.25	2.61	2.89	2.92	2.35	2.35	3.27	2.61	2.61	2.36	
53			W	69	FFPE	10%	1.68	1.64	1.99	2.03	1.79	3.89	3.64	1.98	5.10	4.39	4.25	4.10	5.31	5.13	4.88	4.29	3.51	2.68	
55	IV	M	M	47	FFPE	20%	1.76	1.62	1.68	2.07	1.67	3.36	2.22	2.17	3.67	4.60	3.32	4.15	3.42	4.01	3.70	4.32	3.13	2.45	
56	IIIA	R	W	52	FFPE	20%	1.49	1.40	excluded	1.89	excluded	2.77	2.60	1.94	4.04	3.69	2.85	3.34	3.77	3.49	5.32	5.67	2.99	5.88	
57			M	61	FFPE	35%	1.67	1.79	1.66	1.78	1.91	1.37	1.93	1.95	1.50	1.81	2.02	1.76	1.91	1.86	2.33	2.89	4.69	2.20	
58			M	53	cyto		1.75	1.77	1.74	1.99	1.82	1.36	2.28	1.67	1.71	1.65	1.94	1.56	2.04	1.81	1.85	2.06	1.59	2.08	
59			M	76	FFPE	30%	1.43	1.23	1.54	2.00	1.28	excluded	2.18	1.96	2.21	2.19	2.60	2.87	2.44	2.32	3.25	2.42	2.98	2.89	
60	IIIA	M	M	72	FFPE	20%	1.37	1.46	1.76	2.05	1.82	2.64	2.30	2.51	3.13	3.77	3.18	3.48	3.10	3.29	3.17	3.40	2.95	2.66	
61	IV	M	M	69	FFPE		1.85	1.73	2.52	3.12	1.50	3.00	2.22	excluded	2.77	2.36	3.41	3.42	4.64	3.35	3.75	5.04	4.12	2.12	
62			W	60	FFPE	70%	1.39	2.04	1.52	2.02	1.68	1.81	2.38	excluded	1.53	2.02	2.63	2.12	1.90	1.81	1.94	1.99	2.04	2.12	
63			M	74	FFPE	65%	1.14	2.00	1.47	2.19	1.13	1.05	3.31	excluded	2.39	2.67	2.50	2.83	2.37	2.71	3.13	3.01	2.68	2.51	
65	IV	M	W	35	cyto		1.58	2.12	2.12	1.81	2.08	2.08	1.26	2.08	2.94	2.05	3.43	2.75	1.73	2.43	2.56	3.71	2.56	2.58	+
66	IV		W	66	cyto	95%	1.90	1.78	1.71	1.22	1.95	1.10	2.26	excluded	2.46	2.67	2.89	3.29	3.00	3.02	2.77	2.68	2.56	2.60	+
67	IIIB	M	M	67	FFPE	50%	1.37	1.39	2.36	2.44	excluded	3.33	2.77	excluded	3.09	3.39	3.12	4.14	excluded	3.99	4.37	4.20	4.00	2.82	
68	IIIA	M	M	65	cyto	50%	2.17	2.25	2.31	2.51	1.40	3.33	2.83	1.40	2.32	2.81	2.75	2.68	excluded	3.54	3.07	2.95	2.64	2.27	
69			M	56	FFPE	30%	1.62	1.62	1.74	2.17	1.71	2.84	2.39	2.90	2.89	3.68	2.89	3.03	2.77	2.61	2.60	3.22	2.64	2.39	
70			M	65	FFPE	50%	excluded	1.59	1.80	1.93	excluded	2.48	2.15	2.16	2.79	3.39	3.37	3.08	2.69	3.81	2.64	2.66	2.82	2.27	
72			W	58	FFPE	20%	1.61	1.55	1.56	2.25	excluded	3.93	2.08	2.58	4.02	4.27	3.14	4.36	4.13	5.26	3.93	3.89	3.25	2.60	+
74			M	56	FFPE	70%	1.15	1.63	1.28	1.19	excluded	0.99	1.43	1.27	1.23	1.32	1.87	1.17	1.15	1.29	1.67	1.78	2.01	2.40	
75			M	73	FFPE	40%	1.33	1.49	1.36	1.35	1.63	1.95	1.41	1.66	3.49	2.16	2.98	3.72	2.53	3.21	3.46	3.23	2.09	2.14	
76			W	61	FFPE	35%	1.86	1.86	1.81	1.94	2.10	2.32	1.92	2.60	2.79	3.18	2.24	2.75	2.84	3.01	3.13	2.98	2.45	2.59	
77			W	56	cyto	40%	2.09	2.03	2.00	1.79	1.87	2.32	2.14	2.14	2.65	2.69	2.46	2.66	2.33	2.99	2.76	2.57	2.45	2.28	
78			W	75	FFPE	30%	1.73	1.76	1.83	1.76	1.96	1.52	1.90	2.07	1.66	1.54	1.90	1.91	2.16	1.87	2.35	2.01	1.81	2.03	
79	IIA	R	M	68	FFPE	85%	1.72	1.50	excluded	1.57	excluded	3.27	2.01	2.17	4.32	4.31	2.78	3.91	3.35	4.98	4.68	4.15	2.94	2.52	
80	IIA	R	M	66	FFPE	30%	1.41	1.53	1.64	2.18	1.76	3.16	2.17	2.50	3.51	3.82	3.07	3.65	3.92	4.19	3.32	3.25	3.14	2.52	
81	IV	P	W	73	cyto	60%	1.61	1.77	1.97	2.16	1.64	1.97	1.68	1.69	2.08	2.67	2.91	2.61	2.63	2.96	3.78	3.03	2.90	2.89	
82	IIIA	M	W	49	FFPE	55%	1.52	1.37	1.88	1.92	1.47	2.00	2.02	1.74	1.66	2.51	2.75	2.28	2.27	2.27	2.40	2.33	2.81	3.00	
83			W	57	cyto		2.13	2.22	2.04	1.77	1.82	2.41	1.98	1.53	2.71	2.73	2.92	2.53	excluded	2.69	3.47	2.92	2.93	2.51	
84			W	71	FFPE	50%	1.90	1.87	1.82	1.78	1.87	2.57	2.08	2.20	2.77	3.08	2.51	2.57	2.50	3.10	2.78	2.73	2.64	2.37	
86			W	66	cyto	80%	1.92	2.29	1.88	1.61	1.91	1.03	0.95	1.52	1.25	1.38	1.84	1.60	1.84	1.90	2.31	1.93	excluded	7.72	+
87	IIIA	M	W	59	FFPE	30%	1.69	1.53	1.88	2.09	1.76	2.89	1.63	2.61	3.36	2.79	2.56	3.30	3.51	3.32	3.36	3.33	2.98	3.92	+
88			W	65	FFPE	50%	excluded	1.32	1.40	1.69	1.67	1.93	1.52	2.45	2.00	2.47	2.74	2.34	2.29	2.42	2.23	2.37	2.33	2.41	
89	IV	M	W	58	FFPE	10%	1.54	1.46	1.75	1.94	1.72	3.01	2.09	2.54	3.38	3.29	2.90	3.58	3.06	3.49	3.52	3.03	2.90	2.55	+
90	IIIA	R	W	36	FFPE	50%	1.60	1.41	1.57	1.88	1.79	4.02	2.10	2.70	4.09	4.17	2.99	4.65	3.92	4.78	4.50	3.73	3.38	2.51	
91			M	55	cyto	50%	1.93	1.84	2.02	1.62	1.62	2.02	1.84	1.66	2.17	2.21	2.66	2.17	2.81	2.57	2.49	3.43	2.56	2.56	
92			W	57	cyto	30%	2.17	1.89	1.99	1.95	1.90	2.01	2.04	2.10	2.23	1.98	2.24	2.14	2.42	2.31	2.39	2.41	2.04	2.21	
93	IIIA	P	M	66	FFPE	60%	0.84	1.39	1.45	1.85	excluded	1.89	1.73	2.37	2.16	3.82	4.84	3.78	5.91	3.78	4.95	3.17	2.96	1.51	
94	IIIA	P	W	78	FFPE	55%	1.89	1.32	1.51	2.19	1.67	1.98	2.02	2.14	2.76	1.92	3.07	3.45	2.77	2.49	2.50	2.41	1.98	3.58	+
95			M	51	cyto	50%	2.43	2.12	2.48	1.64	2.12	2.06	2.27	1.62	2.88	3.00	2.30	2.61	excluded	3.28	2.73	3.03	2.79	2.57	
96	IIIB	P	M	58	FFPE	20%	1.96	1.86	2.10	1.68	1.98	2.01	1.95	2.15	2.18	2.04	2.16	2.35	2.42	2.53	2.54	2.46	2.21	2.21	
97			W	54	FFPE	40%	1.40	1.38	1.66	1.74	1.55	1.96	1.98	2.13	2.11	2.06	2.38	2.34	2.19	2.32	2.64	2.52	2.30	2.36	
98			M	56	FFPE	60%	1.91	1.50	1.89	1.97	1.47	2.79	2.01	1.96	2.92	2.99	2.93	3.27	3.15	3.17	3.26	3.35	2.87	2.68	
99	IIA	R	W	63	FFPE		1.76	1.60	1.83	1.71	1.74	1.72	1.69	2.30	2.49	2.05	2.61	2.31	2.37	2.22	4.69	2.50	1.96	2.19	
100			M	48	FFPE	70%	1.73	1.23	1.50	3.10	1.84	2.61	2.44	2.42	2.49	2.76	2.80	3.12	2.72	2.53	4.91	6.09	2.20	2.67	
101	IIIB	P	M	64	FFPE	15%	1.44	1.43	1.81	2.09	1.54	2.07	2.08	2.24	2.57	2.35	2.48	2.58	2.66	2.57	2.79	2.63	2.22	2.41	
102	IIIB	R	M	42	FFPE	70%	2.13	1.81	excluded	1.85	excluded	1.76	1.83	2.02	1.76	1.80	2.02	2.06	2.25	2.31	2.53	2.23	2.08	2.22	
103	IV	M	W	64	cyto	90%	2.09	2.40	1.94	1.86	1.82	1.40	2.02	1.91	2.05	2.25	2.55	2.17	2.16	2.36	2.74	2.24	2.07	2.17	
104	IIIA	P	M	58	FFPE	10%	1.93	1.92	1.88	1.79	1.89	1.77	2.23	2.14	1.48	1.66	2.10	2.02	1.71	1.69	2.50	2.03	2.00	2.14	
105			M	68	cyto	70%	1.54	excluded	1.77	excluded	1.26	excluded	excluded	excluded	1.93	excluded	excluded	excluded	1.82	1.69	2.44	2.47	2.81	1.86	
106			M	59	cyto	60%	1.67	1.82	1.54	1.71	1.97	1.74	1.58	excluded	1.98	2.12	1.78	excluded	excluded	1.93	2.17	1.94	2.13	2.32	
107	IIIA	R	M	54	FFPE	70%	1.40	1.45	1.79	1.94	1.39	1.47	1.53	excluded	1.57	1.40	1.78	excluded	1.95	1.37	2.11	2.08	2.43	2.88	
108	IV	M	W	64	FFPE	60%	1.22	1.45	1.55	3.17	1.34	2.19	1.79	excluded	2.75	2.19	2.52	excluded	3.42	2.41	3.40	2.47	2.45	2.56	
109			M	62	FFPE	40%	1.66	1.62	1.65	1.78	1.87	1.54	1.64	excluded	1.70	1.68	2.18	excluded	1.85	1.84	1.83	1.75	1.87	2.15	
110	IIIA	P	W	67	FFPE	15%	1.62	1.60	1.72	1.85	1.89	2.12	2.51	excluded	2.46	2.16	2.17	excluded	2.40	2.52	3.25	2.65	2.38	2.57	+
111			M	62	cyto	80%	2.07	2.05	excluded	1.67	excluded	1.22	1.54	2.19	1.40	1.56	2.34	1.79	2.12	1.94	2.27	2.69	1.73	2.29	
112	IIA		M	44	cyto	80%	2.56	2.22	excluded	1.35	excluded	1.42	1.54	2.17	2.37	2.36	2.24	1.98	excluded	2.76	3.43	2.67	2.75	2.52	
113			M	59	FFPE	50%	2.15	2.05	2.11	1.51	excluded	excluded	2.64	2.83	excluded	1.24	1.61	excluded	1.26	1.16	1.77	excluded	1.69	2.27	
114			M	66	FFPE	50%	2.04	1.65	excluded	excluded	excluded	2.14	3.07	2.77	excluded	3.02	2.29	excluded	excluded	2.95	2.98	excluded	2.26	2.22	
115	IIIB	P	W	38	cyto	55%	2.29	1.99	2.45	1.49	2.12	excluded	1.81	2.91	2.62	2.00	1.92	excluded	excluded	2.38	3.50	3.14	2.58	2.93	
116	IV	M	W	54	FFPE	25%	1.15	1.29	1.78	excluded	1.33	2.35	1.66	3.12	3.74	2.32	2.52	excluded	3.67	4.20	3.87	3.64	excluded	4.6	+
118			M	60	FFPE	50%	1.44	1.59	1.99	1.89	1.56	1.63	2.35	excluded	1.95	2.41	2.44	2.85	2.52	2.91	2.76	3.15	2.57	2.98	
119	IIA	R	M	73	FFPE	30%	1.66	1.74	1.70	2.37	2.42	1.20	2.43	1.83	2.23	2.53	1.72	2.60	4.01	2.19	4.29	2.44	2.19	2.01	
120	IIIA	P	M	66	FFPE	50%	2.22	2.09	1.66	1.48	1.84	1.88	2.16	2.04	2.31	2.54	2.44	2.76	excluded	2.64	2.80	2.62	2.43	2.45	
121			M	71	cyto	40%	2.90	2.60	2.01	1.32	1.98	1.65	1.93	1.65	2.20	2.73	2.13	2.56	excluded	2.73	2.91	2.55	excluded	2.65	
123	IIIB	P	M	55	cyto	65%	excluded	2.62	excluded	1.31	excluded	1.39	1.94	1.78	excluded	2.44	2.37	2.29	excluded	excluded	excluded	excluded	2.59	2.82	
124	IIIA	P	M	73	FFPE	65%	1.70	1.74	1.57	1.83	1.63	3.06	2.20	2.85	3.11	3.95	3.35	3.80	3.24	4.20	3.56	3.97	3.18	2.76	
125			M	68	FFPE	10%	1.15	1.09	1.41	2.32	1.27	3.62	2.26	2.74	4.26	5.03	4.26	4.88	4.11	4.61	4.67	3.97	3.07	2.71	
126			M	61	FFPE	15%	1.43	1.19	1.55	2.16	1.62	2.94	2.44	2.96	3.11	3.07	2.84	3.37	3.92	3.02	4.41	3.19	2.64	2.73	
127	IV	M	W	48	FFPE	15%	1.41	1.39	1.43	1.81	1.57	2.71	2.04	2.83	2.61	2.58	3.27	3.61	excluded	3.22	4.88	3.82	2.96	2.83	
128	IV	M	M	49	FFPE	15%	1.32	1.17	1.44	2.53	1.87	3.12	2.43	3.23	3.35	4.15	3.31	3.93	4.91	3.93	3.70	3.33	2.94	2.49	
129	IV	R	M	70	cyto	70%	2.85	1.54	2.12	1.79	1.46	2.33	2.59	2.36	2.38	1.77	2.85	2.72	1.79	1.78	3.47	2.67	3.34	2.14	
132			M	64	cyto	80%	1.91	2.39	1.74	1.97	1.77	1.05	2.01	1.57	1.55	3.24	2.64	2.26	3.06	2.16	2.25	3.08	2.30	2.26	
133			M	57	cyto	60%	1.71	1.67	1.64	1.23	1.41	0.78	1.74	1.26	1.31	1.82	2.83	1.94	excluded	1.65	1.65	2.08	1.93	2.19	
134	IIIA	P	M	58	cyto	70%	1.97	1.73	2.26	2.04	1.46	1.27	2.87	2.59	1.13	1.78	2.00	1.96	2.27	1.27	1.82	1.79	2.25	2.24	
135			M	57	cyto	80%	2.17	2.17	1.86	1.39	1.88	1.57	1.83	1.94	1.79	2.24	2.31	2.36	2.12	1.99	2.46	2.05	2.11	2.29	
136			W	42	cyto	80%	excluded	2.08	1.93	1.40	excluded	1.38	2.10	1.70	2.66	2.87	3.56	2.88	excluded	3.28	excluded	excluded	excluded	2.29	
137			W	62	cyto	50%	excluded	1.93	2.46	1.93	excluded	1.57	1.91	1.93	3.24	2.41	2.47	2.20	3.43	excluded	excluded	3.18	2.13	7.72	+
138			M	66	cyto	90%	excluded	excluded	2.24	excluded	excluded	excluded	excluded	excluded	4.20	excluded	excluded	excluded	excluded	excluded	excluded	3.53	excluded	2.98	
140			M	79	cyto	60%	excluded	1.34	1.68	0.84	excluded	0.68	1.16	0.94	1.75	0.83	1.53	1.21	2.27	1.65	excluded	excluded	1.41	0.83	
142			W	57	cyto	80%	1.84	2.98	2.32	1.44	1.60	excluded	1.87	1.61	1.82	1.55	excluded	2.14	2.30	2.73	2.15	2.94	2.27	2.28	
143			W	52	cyto	60%	2.35	2.47	2.12	2.08	2.05	1.41	1.85	1.62	excluded	1.63	2.20	1.94	1.88	2.24	1.59	3.14	2.01	2.63	
144	IIIA	P	W	80	FFPE	15%	1.80	1.36	1.40	1.84	1.91	2.68	1.64	2.11	3.00	3.32	2.56	2.93	3.94	3.99	excluded	excluded	2.28	2.15	
145			M	59	cyto	70%	2.00	1.73	1.53	1.24	1.64	1.65	1.36	1.42	1.85	2.65	1.93	2.87	1.81	3.42	excluded	excluded	excluded	2.09	
146			M	64	cyto	60%	1.74	1.67	2.58	2.22	excluded	3.04	4.30	excluded	3.85	4.18	3.89	2.61	excluded	6.75	excluded	4.51	3.18	2.09	
147			M	59	cyto	60%	2.86	2.63	2.47	1.40	2.81	1.40	2.11	1.81	2.04	2.61	2.26	2.74	2.75	2.83	excluded	excluded	excluded	2.56	
148			W	77	cyto	60%	1.96	1.52	1.69	1.64	2.28	2.38	2.02	1.73	3.06	3.27	2.55	3.01	excluded	4.04	excluded	excluded	2.68	2.36	

149	IV	M	M		FFPE	40%	1.86	1.88	1.93	1.78	1.85	1.98	2.44	1.75	2.41	2.99	2.92	2.74	3.65	2.29	3.48	4.26	2.40	2.24		
151			M	60	cyto	90%	1.93	1.92	1.89	2.29	2.34	1.80	1.37	1.33	2.01	2.31	2.47	2.35	excluded	2.48	excluded	14.55	2.07	1.92		
152	IIIB	M	M	59	FFPE	2%	1.89	1.58	1.76	2.59	2.85	2.98	2.24	2.79	4.60	5.07	2.80	3.58	6.30	5.15	3.74	excluded	2.79	2.21		
153	IIIA		M	70	FFPE	0%	2.41	2.28	1.87	1.57	2.33	1.51	1.85	1.96	1.44	1.61	1.76	1.57	1.84	1.65	excluded	2.21	1.96	2.09		
154			M	64	cyto	70%	2.06	2.39	1.94	1.69	1.59	1.33	1.76	1.12	2.39	2.17	1.65	2.30	excluded	3.38	2.60	excluded	2.33	2.17		
155			M	69	cyto	80%	2.77	2.23	2.22	1.77	2.00	2.03	2.08	1.77	2.20	2.72	2.02	2.36	2.33	2.85	3.76	2.52	2.33	2.33		
156			M	59	cyto	70%	1.89	2.16	2.16	1.51	2.04	1.80	2.70	1.51	2.27	1.88	2.59	1.80	excluded	2.04	3.38	2.45	2.34	2.77		
157	IV	M	M	56	FFPE	35%	1.51	1.93	1.41	1.81	1.43	1.44	1.40	1.59	1.15	1.53	1.80	1.39	2.09	1.61	2.13	3.15	1.73	2.93		
158		M	M	59	FFPE	40%	1.39	1.86	1.72	1.70	1.84	1.14	1.87	1.79	1.77	2.81	2.18	2.11	2.46	2.63	excluded	2.56	2.29	1.99		
159			M	58	FFPE	60%	1.72	1.63	1.66	1.60	2.08	1.64	2.00	2.22	2.11	1.92	1.86	2.00	2.51	2.09	2.22	2.34	1.90	2.10		
160	IV	P	M	54	FFPE	30%	1.91	1.89	1.66	1.54	1.99	1.65	2.00	1.86	1.77	1.73	1.81	1.95	2.05	2.08	2.20	2.36	1.96	2.21		
161			M	74	cyto	80%	1.93	1.92	1.82	1.65	1.72	1.49	1.80	1.75	1.85	2.03	2.68	1.87	1.96	2.18	1.86	2.07	1.80	1.84		
162			M	76	cyto	60%	1.90	2.03	2.10	0.97	1.45	1.02	1.69	0.89	1.02	1.25	1.56	1.57	excluded	1.34	1.98	2.53	1.88	2.70		
164	IIIB	P	M	55	cyto	90%	1.59	2.18	2.19	1.89	1.78	1.64	1.47	2.17	2.75	1.97	2.76	2.84	2.90	2.20	3.14	2.96	2.43	2.37		
165			M	66	FFPE	30%	2.30	excluded	2.05	1.67	2.06	2.52	1.74	2.22	3.01	3.13	2.61	3.00	2.43	3.53	3.05	3.30	2.36	2.24		
166	IV	M	M	56	FFPE	25%	1.77	1.69	1.65	1.67	2.23	1.70	1.94	2.43	2.17	2.01	2.02	1.81	2.66	2.29	2.94	2.85	1.99	2.11		
167	IIIA	M	M	74	FFPE	50%	1.62	1.30	1.32	1.04	1.68	0.91	1.81	1.57	1.67	2.35	2.03	1.60	3.31	2.02	2.14	1.80	1.92	1.96		
168	IIIB	M	M	61	FFPE	25%	1.34	1.14	1.71	1.84	1.89	1.15	1.33	2.21	1.53	1.45	1.43	1.67	2.07	1.80	3.83	2.29	2.45	2.00		
169			W	64	FFPE	75%	1.70	excluded	1.78	1.80	1.41	1.21	1.73	2.23	1.70	1.62	2.17	2.67	2.85	2.05	2.38	3.75	3.18	2.20		
170	IIIA	R	W	77	FFPE	25%	2.07	2.05	2.03	1.50	1.94	1.29	1.81	1.88	1.58	1.23	1.64	1.84	2.25	1.89	2.27	2.24	2.10	2.23		
171	IV	P	W	65	cyto	85%	2.23	1.99	2.03	2.34	2.05	1.99	2.18	2.39	2.17	2.18	2.13	2.35	2.39	3.49	2.22	2.39	2.00	1.95		
172			M	54	FFPE	70%	1.41	excluded	1.59	2.04	1.96	1.48	2.07	2.49	2.78	2.20	2.39	2.42	3.52	3.13	2.65	2.79	2.00	1.95		
173	IIIB	P	M	67	cyto	40%	1.57	1.30	1.51	2.38	1.52	1.42	1.58	1.65	2.10	2.73	1.94	2.48	2.71	2.75	3.19	3.20	1.91	2.97		
174		M	M	59	FFPE	50%	1.99	1.51	1.96	2.20	1.83	1.66	1.88	2.50	2.59	2.54	2.77	3.40	3.65	2.93	3.78	2.60	1.96	2.20		
175	IV	R	W	36	FFPE	20%	1.80	1.45	1.81	2.10	1.77	2.26	2.09	2.58	2.68	2.82	2.44	2.94	3.53	3.30	3.00	3.04	2.52	2.31		
176	IIIA	P	W	62	FFPE	10%	1.92	excluded	1.87	2.02	2.34	2.23	3.25	2.45	2.30	2.05	2.83	3.68	3.32	2.81	3.71	3.46	2.57	2.26		
177			W	47	cyto	80%	2.72	excluded	2.16	1.41	2.21	1.92	2.17	excluded	2.59	2.98	2.28	2.97	excluded	3.26	3.48	excluded	excluded	2.81		
178			W	59	FFPE	80%	1.86	1.72	2.18	1.99	1.91	1.44	1.70	1.74	1.92	excluded	2.07	excluded	3.09	1.59	2.26	3.53	2.56	2.59		
180	IV	P	M	73	cyto	80%	2.27	excluded	2.25	excluded	2.37	excluded	excluded	excluded	3.53	excluded	excluded	excluded	4.37	4.15	excluded	5.09	3.41	2.56		
181	IIIB	P	M	60	FFPE	25%	1.35	1.62	1.62	1.40	2.05	1.77	1.67	1.71	excluded	1.51	1.82	1.87	excluded	1.58	excluded	1.92	2.29	2.27		
182			M	56	cyto	90%	1.70	1.59	1.79	2.39	1.79	1.64	2.47	1.30	1.75	1.93	2.24	2.14	2.26	2.12	2.70	2.67	1.88	2.62		
183			M	75	cyto	80%	1.69	1.74	1.85	1.84	1.65	2.52	1.40	1.85	1.67	1.97	3.32	2.06	2.23	1.75	3.64	3.04	2.26	2.23		
184	IV	P	M	61	FFPE	25%	1.59	1.42	1.56	1.65	2.10	1.58	2.05	1.88	1.93	1.73	2.06	1.94	3.42	2.01	2.91	2.31	2.47	2.39		
185	IIIA	M	W	69	FFPE	20%	1.73	1.79	1.71	2.25	1.91	1.16	2.80	1.77	1.19	1.23	excluded	1.45	excluded	excluded	1.34	2.38	2.29	2.38		
186			W	76	cyto	60%	1.60	2.02	2.07	2.44	2.36	2.99	2.06	2.35	2.51	2.58	2.44	2.93	2.79	2.85	2.12	2.80	2.08	2.07		
187	IV	M	W	62	cyto	90%	1.80	excluded	2.22	excluded	2.33	excluded	excluded	excluded	1.82	excluded	excluded	excluded	2.12	2.27	1.83	3.91	2.27	11.90	+	
188	IV	M	M	51	FFPE	25%	1.83	1.74	1.78	1.72	2.63	2.39	2.76	excluded	1.93	2.29	2.31	2.22	3.03	2.68	2.44	2.18	1.96	1.90		
189	IV	M	M	60	cyto	80%	1.35	1.29	1.98	1.80	2.29	1.78	2.32	excluded	1.33	1.82	2.45	2.22	1.79	1.54	1.81	2.08	1.91	1.71		
190			M	65	FFPE	40%	1.64	1.47	1.69	1.44	1.97	excluded	2.34	excluded	0.97	0.91	2.55	1.44	1.50	1.14	2.10	1.69	1.92	2.14		
191			M	66	cyto	80%	1.49	excluded	1.65	excluded	1.99	excluded	excluded	excluded	1.22	excluded	excluded	excluded	1.97	1.47	1.75	1.89	1.76	2.05		
192			M	77	cyto	80%	1.09	1.99	1.70	1.86	1.37	1.44	1.58	1.31	1.95	1.90	2.41	2.09	2.09	2.03	2.48	2.14	2.22	2.18	+	
193	IV	M	M	59	FFPE	30%	1.30	1.54	1.60	1.60	2.13	1.54	1.80	1.93	1.32	1.88	3.13	1.98	2.24	2.00	2.73	2.16	1.81	2.24		
194	IV	P	M	66	cyto	80%	1.54	1.64	1.52	1.60	1.25	1.81	1.77	1.78	2.55	3.01	3.39	2.87	2.27	2.04	2.65	2.62	2.46	2.13		
195	IV	P	W	59	cyto	80%	2.25	2.17	2.11	1.81	2.12	1.62	1.98	1.85	1.70	1.72	1.83	1.68	2.06	2.06	1.77	1.99	1.73	2.22		
196	IV	P	W	58	FFPE	60%	1.61	1.76	1.86	1.89	1.89	1.54	1.82	1.92	1.63	1.58	2.21	1.91	2.34	1.82	2.15	2.26	2.00	2.85	+	
197			M	66	cyto	80%	1.47	2.20	2.00	1.80	1.79	1.44	1.74	excluded	2.63	2.69	2.54	3.29	excluded	4.10	3.30	excluded	excluded	2.34		
199	IIIA	M	M	53	FFPE	70%	1.52	1.38	1.52	1.60	2.05	1.79	1.89	2.41	2.00	2.16	2.84	2.28	2.73	2.55	2.77	2.40	2.51	2.66		
200	IIIA	R	W	51	FFPE	20%	2.06	1.95	1.86	1.57	2.05	1.70	1.89	1.83	1.58	1.65	1.98	1.94	2.04	1.92	2.10	2.65	2.06	2.25		
201			M	66	cyto	80%	1.94	1.95	2.00	1.42	1.80	1.80	2.00	excluded	2.48	2.61	2.39	2.65	excluded	3.18	2.70	2.77	2.89	2.58		
202			M	71	FFPE	30%	1.78	1.69	1.85	1.83	1.76	excluded	1.81	2.16	2.32	2.01	2.48	2.74	2.29	2.47	2.60	2.75	2.51	2.77	+	
204			M	58	FFPE	60%	1.72	1.81	1.77	1.81	1.89	1.78	2.07	excluded	1.47	1.56	2.19	1.82	2.06	1.54	1.77	2.30	1.97	2.32		
205	IV	P	W	61	cyto	10%	1.85	2.08	2.03	1.55	1.77	2.15	1.82	excluded	2.75	3.28	2.50	3.28	excluded	4.22	3.66	excluded	excluded	3.64		
206	IV	P	W	56	FFPE	45%	1.45	1.76	1.70	2.03	1.48	2.43	1.74	excluded	3.16	3.04	2.94	3.40	4.01	2.32	3.89	3.28	2.81	2.40		
207			M	60	cyto	5%	2.16	2.12	2.74	2.05	1.49	1.81	1.88	excluded	2.63	2.74	2.21	3.19	excluded	3.43	2.76	2.72	excluded	3.18		
208			M	72	cyto	90%	1.29	1.73	2.07	1.30	1.02	1.45	4.00	excluded	2.01	1.64	3.38	2.26	excluded	3.06	2.95	excluded	excluded	2.88		
209			W	71	cyto	90%	1.32	2.34	1.58	1.41	2.00	1.19	2.07	1.52	2.96	3.62	3.02	2.22	1.75	2.46	3.05	3.95	2.13	2.54	+	
210			M	58	cyto	80%	1.35	1.23	1.16	1.65	1.38	1.43	1.92	1.57	3.55	4.66	2.80	3.76	excluded	4.75	3.81	excluded	excluded	2.99		
211			M	68	cyto	80%	1.29	excluded	1.56	excluded	1.30	excluded	excluded	excluded	3.19	excluded	excluded	excluded	excluded	4.69	4.01	excluded	excluded	excluded		
212	IV	P	M	76	FFPE	70%	1.31	1.36	1.55	1.81	1.90	2.04	1.60	2.22	1.96	2.14	2.45	2.67	2.58	2.45	2.76	2.85	2.64	2.60		
213			W	78	cyto	90%	1.58	2.14	1.83	1.80	1.96	1.00	1.99	2.05	2.36	2.49	2.43	2.62	2.92	2.47	2.62	3.21	3.95	2.89		
214			M	62	FFPE	20%	1.65	2.01	1.53	1.42	1.57	excluded	1.40	excluded	0.76	0.85	excluded	1.17	1.59	0.64	1.21	1.46	1.49	2.27	+	
215	IV	M	M	64	FFPE	25%	1.79	1.86	1.84	2.01	2.20	2.22	1.97	excluded	2.50	2.73	2.64	2.83	3.11	2.60	3.00	3.54	2.64	2.50		
216			W	59	cyto	80%	2.21	2.43	2.00	1.86	2.15	1.54	2.06	excluded	1.91	2.21	2.13	2.58	2.06	2.38	2.28	2.30	1.92	2.19		
217			M	48	cyto	80%	2.49	1.64	2.25	2.34	2.03	1.67	2.78	excluded	1.64	2.43	2.45	3.66	2.52	2.84	2.67	5.42	2.62	2.58		
219			M	79	FFPE	30%	1.83	2.12	1.75	1.69	2.08	1.45	1.70	1.92	1.29	1.45	2.00	1.79	2.05	1.59	2.45	1.99	7.72	2.48		
220			W	55	FFPE	20%	1.70	2.48	1.85	1.69	2.16	1.66	1.68	2.01	1.61	2.02	2.00	2.55	2.50	2.18	2.60	2.28	2.17	2.24		
222	IIIA	P	M	65	FFPE	20%	1.60	1.85	1.63	1.37	1.74	1.59	1.68	1.93	1.90	1.84	2.14	2.15	2.12	2.25	2.47	2.28	2.20	2.10		
223	IV	P	M	70	cyto	70%	1.92	2.29	2.17	1.60	2.25	1.64	1.34	1.51	2.69	3.22	3.14	2.24	2.36	3.90	3.33	6.41	2.81	3.51	+	
224	IV	M	M	60	cyto	60%	2.04	3.02	2.57	1.40	1.64	1.33	1.94	excluded	2.02	2.23	2.08	1.97	excluded	2.21	2.73	2.52	2.19	2.23		
225			W	76	FFPE	20%	1.50	1.61	1.62	2.42	1.93	3.97	2.38	2.87	4.83	5.07	3.20	4.96	excluded	5.14	4.48	excluded	2.85	2.29		
226			W	59	cyto	80%	1.60	2.65	1.82	2.11	1.73	1.04	1.65	1.46	1.80	2.06	1.92	2.01	2.34	2.01	2.90	2.29	2.17	2.13		
227			W	64	cyto	80%	1.97	excluded	2.18	excluded	1.99	excluded	excluded	excluded	3.12	excluded	excluded	excluded	excluded	4.18	3.25	excluded	excluded	3.99	+	
228			M	72	cyto	70%	2.55	2.89	2.16	1.29	2.12	1.89	2.70	1.39	3.03	4.49	2.41	3.65	excluded	3.95	3.83	excluded	excluded	2.68		
229			W	65	cyto	80%	2.22	excluded	1.94	excluded	2.10	excluded	excluded	excluded	5.10	excluded	excluded	excluded	excluded	5.86	5.02	excluded	excluded	excluded		
230	IIB	P	M	44	FFPE	50%	1.90	1.75	2.10	2.05	1.85	2.77	2.38	2.37	2.68	2.39	2.69	3.12	3.50	2.96	3.60	2.56	2.96	2.43		
231			W	53	cyto	60%	2.06	2.16	1.81	2.05	1.71	2.40	2.07	1.93	2.30	2.95	2.91	2.39	2.79	2.35	2.79	2.69	2.48	2.44		
232			W	70	cyto	50%	2.41	2.35	2.14	2.35	2.21	2.38	2.06	2.50	2.36	2.46	2.59	2.91	2.80	4.14	2.96	2.80	2.38	2.41		
233			M	47	cyto	60%	2.05	2.07	1.70	1.85	1.84	2.63	2.09	2.21	2.83	3.77	3.10	3.25	2.71	3.09	3.28	3.09	2.60	2.34		
234			M	60	cyto	60%	2.10	2.08	1.89	1.70	2.05	2.56	2.15	1.98	2.70	3.03	2.85	2.88	2.52	3.21	2.85	3.48	2.78	2.89		
235		M	W	72	cyto	60%	1.90	2.15	1.92	1.98	1.91	2.24	1.33	2.31	2.14	1.53	2.47	2.49	2.38	2.40	2.17	2.36	2.64	2.31		
236	IIIA	M	W	66	FFPE	30%	1.53	1.85	1.61	1.83	1.82	2.31	1.83	2.46	1.92	2.35	2.69	2.55	2.46	2.93	2.69	3.58	2.29	2.39		
237		M	M	56	FFPE	30%	1.90	1.73	1.86	2.17	1.87	1.79	2.28	1.96	2.16	3.20	2.59	2.35	2.84	3.66	3.21	3.26	2.56	2.29		
238	IIIB		W	59	cyto	80%	excluded	excluded	excluded	1.97	excluded	excluded	2.14	2.34	excluded	excluded	3.96	excluded	excluded	excluded	excluded	excluded	5.51	3.34		
240					FFPE		1.46	1.57	excluded	3.03	excluded	excluded	2.58	3.53	8.99	9.71	excluded	6.72	15.17	11.82	9.34	excluded	2.93	excluded		
241	IIIA	P			FFPE		excluded	2.12	1.80	1.61	1.90	0.95	1.82	1.76	1.75	1.94	1.73	1.81	1.69	1.81	1.49	2.96	2.16	2.88		
242					FFPE	20%	1.59	1.55	1.99	2.02	1.59	2.43	2.46	2.28	2.75	2.52	3.23	2.66	2.56	2.80	3.43	3.08	2.74	2.98		
243					FFPE	40%	1.86	2.14	1.19	excluded	1.74	1.25	1.98	excluded	1.06	1.52	1.97	1.43	1.46	0.71	1.31	1.40	excluded	excluded		
244					FFPE	75%	1.69	1.56	1.42	1.18	1.82	excluded	excluded	excluded	1.03	1.00	1.41	1.08	1.16	1.09	1.43	1.62	1.42	2.10		
245					FFPE	15%	1.98	1.73	1.55	1.28	1.82	excluded	2.09	excluded	1.13	1.01	2.11	1.09	1.21	1.10	2.24	1.95	1.64	2.42		
246		M	M	59	cyto	80%	1.87	1.13	1.56	excluded	2.03	3.10	excluded	4.85	4.74	4.92	4.04	3.35	6.02	5.52	4.79	5.46	3.35	2.39		
247					FFPE	65%	1.61	1.54	1.60	1.84	2.05	1.47	1.82	1.77	1.61	2.33	2.76	1.80	2.00	1.70	1.92	2.05	2.09	2.22		
248					FFPE	80%	2.10	2.07	1.78	1.34	excluded	excluded	excluded	1.76	excluded	1.10	excluded	1.21	1.41	1.10	1.44	excluded	1.74	1.88		
249	IIIA	P			FFPE	70%	1.88	2.28	excluded	1.77	excluded	excluded	1.38	1.88	0.86	0.95	1.63	excluded	1.65	1.04	2.01	2.25	2.01	2.54		
250	IIIA	P			FFPE	5%	1.52	1.42	1.60	1.83	1.59	3.58	2.04	2.38	4.15	4.61	3.20	4.24	3.94	4.60	4.42	3.80	2.99	2.65		
251	IV	P			FFPE	20%	2.50	2.27	2.18	1.39	excluded	excluded	1.54	1.77	1.12	1.53	1.77	1.44	1.22	1.06	1.92	excluded	2.05	excluded		
252					cyto	90%	1.92	1.76	3.35	1.46	excluded	1.92	2.16	1.14	1.73	1.88	2.29	2.72	excluded	3.58	3.34	5.43	3.36	2.93		
253					cyto	80%	1.53	1.80	2.38	1.65	1.83	1.92	2.44	1.38	2.31	3.31	2.60	3.38	excluded	2.62	3.22	3.66	excluded	2.53		
254					FFPE	10%	2.12	2.19	1.56	1.55	1.80	excluded	1.89	1.64	1.09	1.50	excluded	1.35	1.50	1.06	1.70	1.51	1.67	2.14		
255					FFPE	25%	2.36	excluded	excluded	1.26	excluded	excluded	1.45	excluded	excluded	1.09	excluded	1.35	1.61	1.01	1.44	excluded	1.50	2.07		
256					FFPE	10%	2.22	2.69	1.09	excluded	2.25	excluded	1.70	excluded	0.82	1.05	excluded	0.93	1.81	0.82	0.94	excluded	excluded	excluded		
257					FFPE	35%	2.17	2.46	1.51	1.36	1.47	0.95	excluded	1.69	0.90	1.40	2.16	1.23	1.28	0.91	1.59	1.63	1.81	2.49		
258					FFPE	70%	1.77	2.82	1.44	0.79	2.46	excluded	excluded	1.19	1.07	1.17	2.63	0.75	2.25	0.96	1.39	excluded	2.35	2.72		
259			M	78	FFPE	60%	1.03	1.80	2.39	1.67	excluded	1.72	2.24	2.17	2.62	2.16	2.29	2.03	2.84	3.94	1.83	excluded	1.97	2.83		
260					Cyto	80%	excluded	3.23	2.43	1.56	2.37	2.23	2.83	excluded	4.63	5.02	3.27	3.26	excluded	5.07	4.97	excluded	3.71	excluded		
261					FFPE	5%	2.13	2.22	1.58	excluded	2.01	excluded	2.65	excluded	1.40	1.25	1.60	1.61	1.73	1.37	1.61	1.89	1.92	2.16		
262					FFPE	15%	2.34	excluded	1.33	excluded	2.43	excluded	excluded	excluded	1.73	1.57	excluded	1.72	1.99	1.04	1.44	1.67	1.66	1.96		
263					FFPE	25%	1.51	1.51	1.57	2.19	1.61	3.99	2.14	2.43	4.72	5.54	4.02	5.40	4.74	5.42	4.81	excluded	3.38	excluded		
264					cyto	30%	1.87	2.05	2.05	2.25	1.72	1.68	2.11	2.05	1.78	1.62	2.64	1.82	2.69	2.03	2.39	2.87	2.24	1.97		
265					cyto	80%	2.48	2.71	2.07	1.37	1.99	2.62	2.13	excluded	4.67	4.93	2.92	4.13	excluded	5.61	4.91	excluded	excluded	2.69		
266					cyto	80%	3.23	4.97	2.25	2.42	3.76	2.76	2.68	excluded	3.88	4.91	2.97	4.35	excluded	4.72	3.43	3.57	3.24	2.54		
267					FFPE	30%	1.80	2.07	1.59	1.49	1.92	1.17	1.82	1.69	1.07	1.28	1.76	1.38	1.51	1.15	1.28	1.62	1.63	2.06		
268					FFPE	80%	1.84	2.45	1.36	1.54	1.70	excluded	excluded	1.69	1.11	1.48	1.61	1.31	2.14	0.98	1.58	1.62	2.02	2.18		
269					FFPE	70%	2.29	2.48	2.15	1.44	2.49	excluded	2.50	1.80	1.20	1.02	1.84	0.97	1.55	1.25	1.91	2.07	2.04	2.54		
270					FFPE	60%	2.35	2.46	2.05	1.70	2.05	excluded	1.67	2.12	1.29	1.72	1.92	1.46	1.38	1.43	1.70	1.91	1.95	2.32		
271					FFPE		0.99	excluded	excluded	2.57	1.85	4.58	2.56	3.72	4.97	5.65	3.64	5.12	5.67	6.25	4.35	3.48	2.93	2.52		
272					FFPE		1.25	1.64	1.69	2.11	2.06	2.70	1.88	2.52	3.72	3.49	2.96	3.27	3.66	3.80	2.89	3.13	2.58	2.43		
273					FFPE		1.83	1.70	2.05	2.75	2.49	3.57	excluded	3.44	4.10	4.12	3.49	4.37	6.07	4.93	4.41	4.58	excluded	2.60		
274					FFPE		1.76	1.93	1.85	2.23	2.22	2.85	2.19	2.86	3.97	3.80	3.17	3.45	4.80	4.58	3.41	3.32	2.69	2.35		

LEGEND

PTC - percentage of tumor cells

cyto - sample isolated from cytological smear

FFPE - formalin-fixed, paraffin-embedded samples

copy number category color coding:

normal

hom del

loss

gain

amplification

excluded

3

Klonowska K., Czubak K., Wojciechowska M., Handschuh L., Żmieńko A., Figlerowicz M., Dams-Kozłowska H., Kozłowski P.

„Oncogenomic Portals for the Visualization and Analysis of Genome-wide Cancer Data”

Oncotarget 2015, doi: 10.18632/oncotarget.6128

Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska[1], Karol Czubak[1], Marzena Wojciechowska[1], Luiza Handschuh[1,2], Agnieszka Zmienko[1,3], Marek Figlerowicz[1,3], Hanna Dams-Kozlowska[4,5] and Piotr Kozlowski[1]

[1] European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

[2] Department of Hematology and Bone Marrow Transplantation, Poznan University of Medical Sciences, Poznan, Poland

[3] Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland

[4] Department of Diagnostics and Cancer Immunology, Greater Poland Cancer Centre, Poznan, Poland

[5] Chair of Medical Biotechnology, Poznan University of Medical Sciences, Poznan, Poland

Correspondence to: *Piotr Kozlowski*, ***email***: *kozlowp@yahoo.com*

Keywords: cBioPortal, COSMIC, IntOGen, PPISURV/MIRUMIR, Tumorscape

Received: *June 18, 2015* ***Accepted***: *September 28, 2015* ***Published***: *October 15, 2015*

ABSTRACT

Somatically acquired genomic alterations that drive oncogenic cellular processes are of great scientific and clinical interest. Since the initiation of large-scale cancer genomic projects (e.g., the Cancer Genome Project, The Cancer Genome Atlas, and the International Cancer Genome Consortium cancer genome projects), a number of web-based portals have been created to facilitate access to multidimensional oncogenomic data and assist with the interpretation of the data. The portals provide the visualization of small-size mutations, copy number variations, methylation, and gene/protein expression data that can be correlated with the available clinical, epidemiological, and molecular features. Additionally, the portals enable to analyze the gathered data with the use of various user-friendly statistical tools. Herein, we present a highly illustrated review of seven portals, i.e., Tumorscape, UCSC Cancer Genomics Browser, ICGC Data Portal, COSMIC, cBioPortal, IntOGen, and BioProfiling. de. All of the selected portals are user-friendly and can be exploited by scientists from different cancer-associated fields, including those without bioinformatics background. It is expected that the use of the portals will contribute to a better understanding of cancer molecular etiology and will ultimately accelerate the translation of genomic knowledge into clinical practice.

INTRODUCTION

Cancer encompasses a broad spectrum of diseases (>100) that arise from somatically acquired genetic, epigenetic, transcriptomic, and proteomic alterations that have accumulated in the genomes of cancer cells [1]. These alterations are implicated in hallmark oncogenic cellular processes that are characterized by, e.g., sustained proliferative signaling, resistance to apoptosis, induction of invasion and metastasis, and neoangiogenesis [2]. The somatic loss-of-function or gain-of-function alterations are overrepresented in specific genomic regions, which could indicate their potential suppressive or oncogenic roles, respectively. However, it must be noted that somatic mutations occur on different genetic backgrounds and can sometimes interact with germline mutations, which could modify predisposition to cancer when such mutations occur in cancer-associated genes.

Recent advances in technologies for high-throughput genome analysis, such as microarray-based methods and next-generation sequencing (NGS), have enhanced progress in the field of oncogenomics [3]. These tools were fundamental for the initiation and development of multi-centered cancer genomic projects, such as (i) the

Wellcome Trust Sanger Institute's Cancer Genome Project (CGP) [4, 5], (ii) The Cancer Genome Atlas (TCGA) [6-8], and (iii) the International Cancer Genome Consortium (ICGC) cancer genome projects [9, 10]. These projects have been launched for genome-wide analyses of genetic, epigenetic, transcriptomic, and proteomic alterations in hundreds or even thousands of cancer samples. Their general aim is to provide publicly available oncogenomic datasets for the better understanding of the molecular mechanisms that underlie cancer and for the assessment of the influence of specific alterations on clinical phenotypes. Application of the appropriate pipeline for computational interpretation and thought-provoking visualization of the results of oncogenomic projects is crucial to exploring the multidimensional character of genome-wide cancer data [11]. In response to this need, a number of oncogenomic portals were created to assist with accessing the abundant cancer datasets. These portals gather and facilitate the analysis of data with regard to small-size mutation, copy number variation (CNV), methylation, and gene/protein expression. Moreover, they offer a wide range of analysis tools that include the testing of correlations of specific genomic alterations with available clinical information.

Herein, we provide a highly illustrated guide through several web-based oncogenomic portals that were generated to facilitate scientists from different cancer-associated fields, including molecular and clinical oncologists, epidemiologists, and bioinformaticians, with the extraction of meaningful information from expanding oncogenomic sources. Browsing through the portals, prospective users will find a variety of data regarding cancer types and subtypes, oncogenic molecular pathways and cancer-associated genes of interest. All of the portals described below are user-friendly and provide intuitive integration as well as interactive oncogenomic dataset visualizations, and thus, bioinformatics skills and knowledge are not essential to exploring and using these tools. The individual paragraphs listed below present the characteristics and possible utilization of selected web portals. Descriptions and figures that present specific portals were prepared according to their versions from the first half of 2015, and they are summarized in Table 1.

Tumorscape

Tumorscape [12, 13] was developed at The Broad Institute of MIT and Harvard in Cambridge, MA USA. This website was one of the first oncogenomic portals to provide information about cancer copy number changes in a format that was easily accessible to non-bioinformaticians. With this portal, the copy number profiles of over 3,700 cancers (both primary cancers and cell lines) are mapped to the human genome reference sequence and are visualized as heatmap tracks, with the use of the Integrative Genomics Viewer (The Broad Institute). Genomic regions with increased (>2) and decreased (<2) copy number are marked, respectively, in red and blue colors, the intensity of which indicates the amplitude of the copy number changes (Figure 1). The tracks that represent all of the analyzed samples are shown next to one another, forming a panel that allows direct comparison and visualization of all of the analyzed samples. In addition, Tumorscape provides tools that allow "cancer-centric" and "gene-centric" data analyses. The "cancer-centric" analysis (Figure 1A) provides a list of genomic regions that are either significantly amplified or deleted in a specific cancer along with information about the genes that are located in the altered regions. The "gene-centric" (Figure 1B) analysis provides summary statistics of the copy number alterations that affect a gene of interest in a specific cancer type and/or across all cancer types. This summary enables the interpretation of the role of an analyzed gene as a potential oncogene or tumor suppressor.

UCSC cancer genomics browser

The University of California at Santa Cruz (UCSC) Cancer Genomics Browser [14-19] integrates oncogenomic CNV, small-size mutations, methylation, transcriptomic, and proteomic datasets that were obtained in a variety of experiments that were conducted with the use of samples from different cancer types and subtypes. With this portal, all of the oncogenomic information is mapped to the human genome reference sequence and presented as color-coded heatmap tracks. As in Tumorscape, the data from specific experiments are visualized as panels of heatmap tracks in which each track represents an individual sample. Using this portal, the required data can be browsed from the perspective of the whole genome, the exome, a specific chromosome, or a gene. Additionally, there is also the possibility of viewing PARADIGM datasets to gather a sample-specific "gene activity level." This parameter (obtained using the PARADIGM method) [20] provides the incorporation of pathway interactions (which are deposited in the NCI Pathway Interaction Database) [21] and the integration of data with regard to different types of oncogenomic alterations, e.g., changes in the expression or copy number of a given gene [16]. In the UCSC Cancer Genomics Browser, multiple panels can be simultaneously displayed to visualize different categories of oncogenomic information for a specific cancer type and/or the same category of oncogenomic data for different cancer types (Figure 2A-2D). With this browser, analyses can be concurrently conducted for thousands of samples (oncogenomic datasets) that are sorted by different clinical, epidemiological, and molecular features (Figure 2). These features include survival, histological type, tumor nuclei percent, followup treatment success, new tumor event after initial treatment, neoplasm histologic grade, and tumor necrosis percent, as well as gender, age

Table 1: Main characteristics of the selected oncogenomic portals.

database	data source	sites of analysed cancer[1]	organisation of data[2]	oncogenomic data/ analyses	link/literature
Tumorscape	Broad Institute	Bd; Bld; Br; Bra; Clr; Eso; GIST; HN; Htp; Kd; Lng; Lvr; Lymph; Msh; Ov; Pnc; Prst; Sk; ST; Stc; Swn; Thr; Utr; also in: cancer cell lines	level i-iii	copy number alterations	http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf; [12]
UCSC Cancer Genomics Browser	TCGA, SU2C Breast Cell Line, Cancer Cell Line Encyclopedia, The Connectivity Map, TARGET, cancer data from literature	Bd; Bld; Br; Bra; Chl; Col; Clr; EG; Eso; HN; Kd; Lng; Lvr; Lymph; Msh; Ov; Pan; Pnc; Prc / Prn; Prst; Rc; Sk; ST; Stc; Thm; Thr; Utr; also: cancer cell lines; cancer data from mouse models	level i-iii	DNA copy number, miRNA/exon/gene/ protein expression, DNA methylation, gene-level mutations, PARADIGM pathway activity; clinical, epidemiological, and molecular information	https://genome-cancer.ucsc.edu; [14-18]
ICGC Data Portal	ICGC, TCGA ,TARGET	Bd; Bld; Bo; Br; Bra; Clr; Col; Eso; HN; Kd; Lng; Lvr; Lymph; Nb; Ov; Pnc; Prst; Rc; Sk; ST; Stc; Thr; Utr;	level i-iv	simple somatic mutations, copy number somatic alterations, structural somatic mutations, simple germline variants, DNA methylation, gene/ protein expression, miRNA expression, exon junction; epidemiological and clinical data	https://dcc.icgc.org; [32]
COSMIC	TCGA, ICGC, cancer data from literature	Bo; Br; EA; Eso; GIST; Htp; Kd; Lvr; Lng; Ov; Pnc; Prst; Sk; Stc; Tst; Thm; Thr; Utr	level iii-iv	somatic mutations, copy number alterations, gene expression	http://www.sanger.ac.uk/genetics/CGP/cosmic; [39-43]
cBioPortal	AMC, BCCRC, BGI, British Columbia, Broad, Broad/ Cornell, CCLE, CLCGP, Genentech, ICGC, JHU, Michigan, MKSCC, MKSCC/ Broad, NCCS, NUS, PCGP, Pfizer UHK, Riken, Sanger, Singapore, TCGA, TSP, UTokyo, Yale	ACC; Bd; Bld; Br; Bra; Chl; Clr; Eso; HN; Kd; Lng; Lvr; Lymph; MM; Npx; Ov; Pnc; Prst; Sk; ST; Stc; Thr; Utr; also: cancer cell lines	level iii-iv	mutations, putative copy number alterations; mRNA expression, protein/ phosphoprotein level; survival analyses	http://www.cbioportal.org; [57, 58]
IntOGen (2014.12)	TCGA, ICGC, cancer data from literature	Bd; Bld; Br; Bra; Clr; Eso; HN; Kd; Lng; Lvr; Lymph; Ov; Pnc; Prst; Sk; Stc; Thr; Utr	level iii-iv	results of the analyses indicating driver alterations and genes; therapies tailored to the mutation profiles of the analyzed patients	http://www.intogen.org/; [67-70]
BioProfiling.de					
PPISURV	for gene expression: Gene Expression Omnibus; for interactome: IntAct, HPRD, Reactome, HumanCyc, NCI_NATURE, PhosphoSitePlus	Bd; Bld; Br; Bra; Col; Htp; Lng; Lvr; Lymph; Ov; Prst; ST; Utr	level iv	survival analyses	http://bioprofiling.de/GEO/PPISURV/ppisurv.html; [81]
MIRUMIR	Gene Expression Omnibus	Br; Eso; Lvr; Lng; Npx; Ov; Prst; Sk	level iv	survival analyses	http://www.bioprofiling.de/GEO/MIRUMIR/mirumir.html; [83]
DRUGSURV	for gene expression: Gene Expression Omnibus; for drugs modulating a gene of interest: DrugBank, Pubchem Bioassay	Bld; Br; Bd; Col; Bra; Lng; Lvr; Lymph; Prst;; ST; Utr	level iv	list of drugs targeting specific genes/ cancer types; survival analyses	http://www.bioprofiling.de/GEO/DRUGSURV/index.html; [85]

[1]List of abbrieviations of cancer sites. In the brackets there are exemplary cancer subtypes included in the portals.
ACC – adenoid cystic carcinoma; Bd – bladder; Bld – blood; Bo – bone; Br – breast; Bra – brain; Chl – cholangiocarcinoma; Clr – colorectal; Col – colon; EA – eye and adnexa; EG - endocrine glands; Eso – esophagus; GIST – gastrointestinal; HN – head and neck; Htp – hematopoietic; Kd – kidney; Lng – lung; Lvr – liver and biliary tract; Lymph – Lymphoma; Msh – mesothelioma; Mth – mouth; Nb – neuroblastoma; Npx – nasopharynx; Ov – ovary; Pan – pancancer; Pnc – pancreas; Pnx – pharynx; Prc/Prn - pheochromocytoma and paraganglioma; Prst – prostate; Rc – rectum; Sk – skin; ST – soft tissues; Stc – stomach; Swn – schwannoma; Thm – thymus; Thr – thyroid; Tst – testis; Utr – uterine (cerxix and corpus).
[2]In oncogenomic portals cancer resources are arranged in different levels of organisation, including: (i) raw, (ii) computationally processed/normalized, (iii) interpreted and (iv) summarized data [3].

A Analysis by Cancer Type

Summary | Amplifications | Deletions

Cancer Type: Lung NSC **Available Data:** 62 peaks **Page:** 1 2 3 4

Peak Region	#Genes in Peak	Residual Q-value	Frequency of Amplification		
			Overall	Focal	High-level
chr14:35755817-35835900	0	5.34E-78	0.438	0.243	0.1
chr12:67484379-68036772	4	1.39E-26	0.323	0.153	0.053
chr8:129271608-129288419	0	3.86E-25	0.557	0.196	0.087
chr7:54995340-55363609	1	7.19E-25	0.45	0.101	0.067
chr12:56419523-56444825	6	1.03E-22	0.322	0.153	0.041

Genes in Selected Peak: EGFR

(truncated for presentation purpose)

B Analysis by Gene

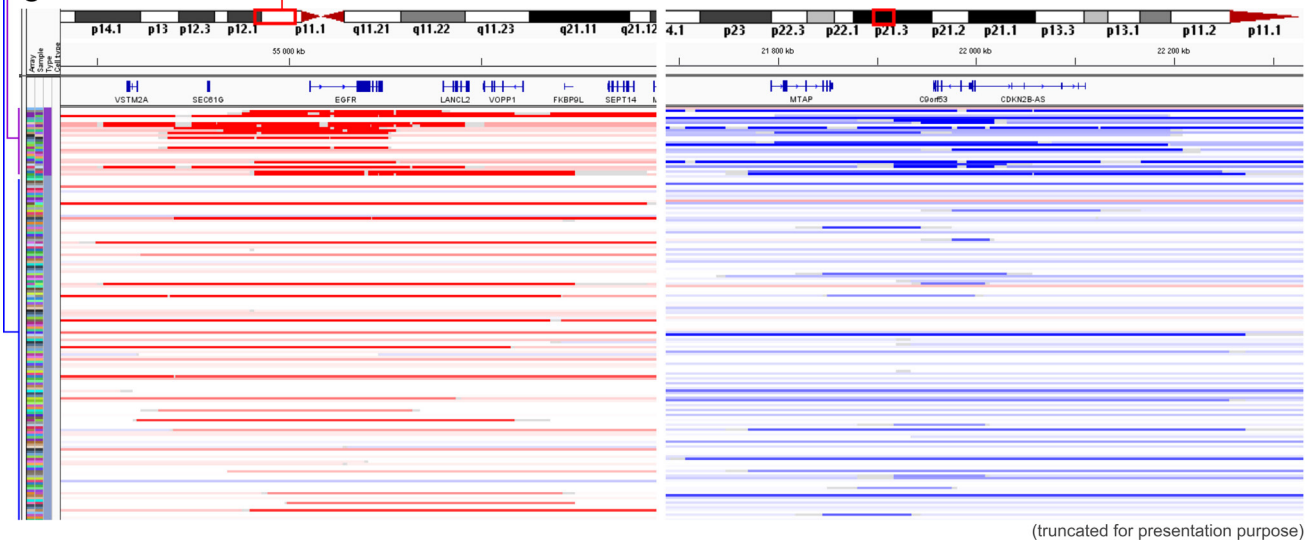
Summary | Amplifications | Deletions

EGFR (chr7:55054218-55206232)

Cancer Subset	In Peak?	Nearest Peak	#Genes in Peak	Q-value	Frequency of Amplification		
					Overall	Focal	High-level
all_cancers	Yes	chr7:54911575-55275039	1	1.57E-46	0.3159	0.0562	0.0438
all_epithelial	Yes	chr7:54911575-55363609	1	5.62E-37	0.4161	0.0725	0.0575
Lung NSC	Yes	chr7:54995340-55363609	1	9.29E-25	0.4529	0.0996	0.0668
all_lung	Yes	chr7:54946418-55363609	1	4.39E-23	0.4496	0.0956	0.0633
Esophageal squamous	Yes	chr7:54762764-55366816	2	7.55E-8	0.7273	0.3409	0.2273
Breast	Yes	chr7:54662395-57367942	14	0.0318	0.3498	0.0617	0.0494
Glioma	No	chr7:55196509-55258847	0	3.4E-36	0.8293	0.4634	0.439
all_neural	No	chr7:55196509-55258847	0	1.79E-23	0.3917	0.1014	0.0922
Hepatocellular	No	No peak on chromosome	0	0.895	0.3636	0.0248	0.0413

(truncated for presentation purpose)

C



(truncated for presentation purpose)

Figure 1: Examples of Tumorscape data analysis and visualization. A. An example of the results that were obtained with the "cancer-centric" analysis. The table shows a list of genomic regions that were most frequently amplified in lung adenocarcinoma. The q-value represents the likelihood of a random occurrence of the specific amplification/deletion that is calculated based on the background copy number variation. The fourth most frequently amplified region that spans *EGFR* is highlighted. **B.** Results obtained with "gene-centric" analysis; the table depicts a list of cancers in which the representative gene (*EGFR)* is located in or near the frequently amplified region (orange and yellow rows, respectively). **C.** Visualization of chromosomal regions that span the exemplary *EGFR* and *CDKN2A* genes, which are undergoing frequent amplifications and deletions, respectively. The heatmaps show copy number variations of glioma and lung adenocarcinoma samples. Each row represents an individual sample, and red and blue indicate amplification and deletion, respectively.

at initial pathologic diagnosis, tobacco smoking history, cytogenetic abnormalities, and expression subtypes. Apart from the heatmap tracks (Figure 2B), the data presented in specific panels can be summarized and plotted as box-and-whiskers or proportions (Figure 2C).

The datasets can also be statistically processed and depicted with the use of a number of tools, such as the hgSignature, which enables the simultaneous analysis of the expression of several genes, to incorporate an algebraic expression signature as a clinical feature. The inclusion of such a feature to the statistical analysis of cancer data could allow the correlation of the molecular and clinical phenotypes or the subdivision of the clinical phenotypes based on the molecular data [15]. Additionally, a correlation of the available clinical, epidemiological, and molecular features with a patient's survival can be depicted in a Kaplan-Meier plot (Figure 2E). Subgroups of samples (distinguished based on the associated features or genomic signatures) can be compared in terms of the obtained oncogenomic data with the use of various statistical tests [i.e., differences in mean, Wilcoxon, Fisher's exact, Fisher's linear discriminant, Jarque Bera normality, Levene homogeneity of variances (HOV), Brown - Forsythe HOV, and Student's T-tests], which can be adjusted for multiple hypotheses p-values through the Bonferonni and Benjamini-Hochberg false discovery rate (FDR) corrections. Importantly, all of the genomic information that is stored in the UCSC database can be easily downloaded for external analyses.

Successful applications of the UCSC Cancer Genomics Browser in cancer-associated research are described in many papers [22-30]. For example, Wu and colleagues [22] used the statistical tool for the generation of a Kaplan-Meier plot to support the significance of their experimental data. Their study revealed that the up-regulated expression level of *HNF1A-AS1* in lung adenocarcinoma is significantly correlated with the TNM stage, tumor size, and lymph node metastasis. These results are in line with the Kaplan-Meier plot, which indicates that patients with high *HNF1A-AS1* expression overall experienced worse survival compared to patients with low *HNF1A-AS1* expression. The UCSC Cancer Genomics Browser is also broadly used for downloading genomic and clinical data for external analyses [24-26, 30].

It is also noteworthy that the authors of the UCSC Cancer Genomics Browser are currently developing a new oncogenomic platform called UCSC Xena [31], which allows users to upload, visualize, and analyze a custom genomic dataset in the context of the large projects data stored in the web browser. Although the UCSC Cancer Genomics Browser and the UCSC Xena currently coexist, it is anticipated that after adding some vital functionalities, UCSC Xena will replace the UCSC Cancer Genomics Browser [18].

ICGC data portal

The ICGC Data Portal [32, 33] provides integration and visualization of the results of 55 cancer projects. This portal was created for the analysis of genomic sequence alterations in relation to clinical patient characteristics, such as ethnicity and epidemiological information. With this portal, the oncogenomic data can be analyzed using four interactive entry points: "Cancer Projects," "Advanced Search," "Data Analysis" and "Data Repository" (Figure 3A). The "Cancer Projects" (Figure 3B) enables data browsing from distinct projects that focus on the oncogenomic analysis of specific cancer types and subtypes. For each dataset, the provided summary includes a list of available oncogenomic data types, most affected donors, genes most frequently affected by cancer alterations, and most common mutations. It is also possible to use the "keyword search" tool to browse all of the gathered oncogenomic data in terms of a specific gene, mutation, donor, or molecular pathway that is of interest. The integration of external databases, such as the Ensembl [34], OMIM [35], Reactome [36], and COSMIC [37], enables the user to look more broadly at a specific gene, molecular pathway, or mutation in terms of its role in carcinogenesis. The "Advanced Search" (Figure 3C) allows extending the analysis and correlating data with additional clinical (e.g., tumor stage, relapse type, disease status), epidemiological (e.g., gender, age at diagnosis, vital status), molecular (e.g., type of the mutation and its consequence), and technical (e.g., type of sequencing platform used for the analysis) information. The "Data Analysis" entry point allows launching three types of analyses: "Enrichment Analysis," "Phenotype Comparison," and "Set Operations." The "Enrichment Analysis" permits the user to identify groups of gene sets from the selected "universe," i.e., Reactome Pathways, Gene Ontology (GO) Molecular Function, GO Biological Process or GO Cellular Component, which appear to be statistically significantly over-represented when compared with a custom gene set that is uploaded by the user. The uploaded custom gene set can consist of up to 10,000 genes. The "Enrichment Analysis" is based on a hypergeometric test and Benjamini-Hochberg adjustment for multiple test corrections with the FDR value threshold selected by the user. The "Phenotype Comparison" analysis allows the user to compare some clinical and epidemiological characteristics across patients with various cancer types, whereas the "Set Operations" can be used to distinguish the shared fraction of the analyzed sets, which are depicted in a Venn diagram (e.g., mutations that are causative across several cancer types). "Data Repository" allows all of the ICGC Cancer Project data to be downloaded and analyzed with the use of external programs and tools of interest. An example of ICGC Data Portal utilization for downloading oncogenomic data has already been published [38].

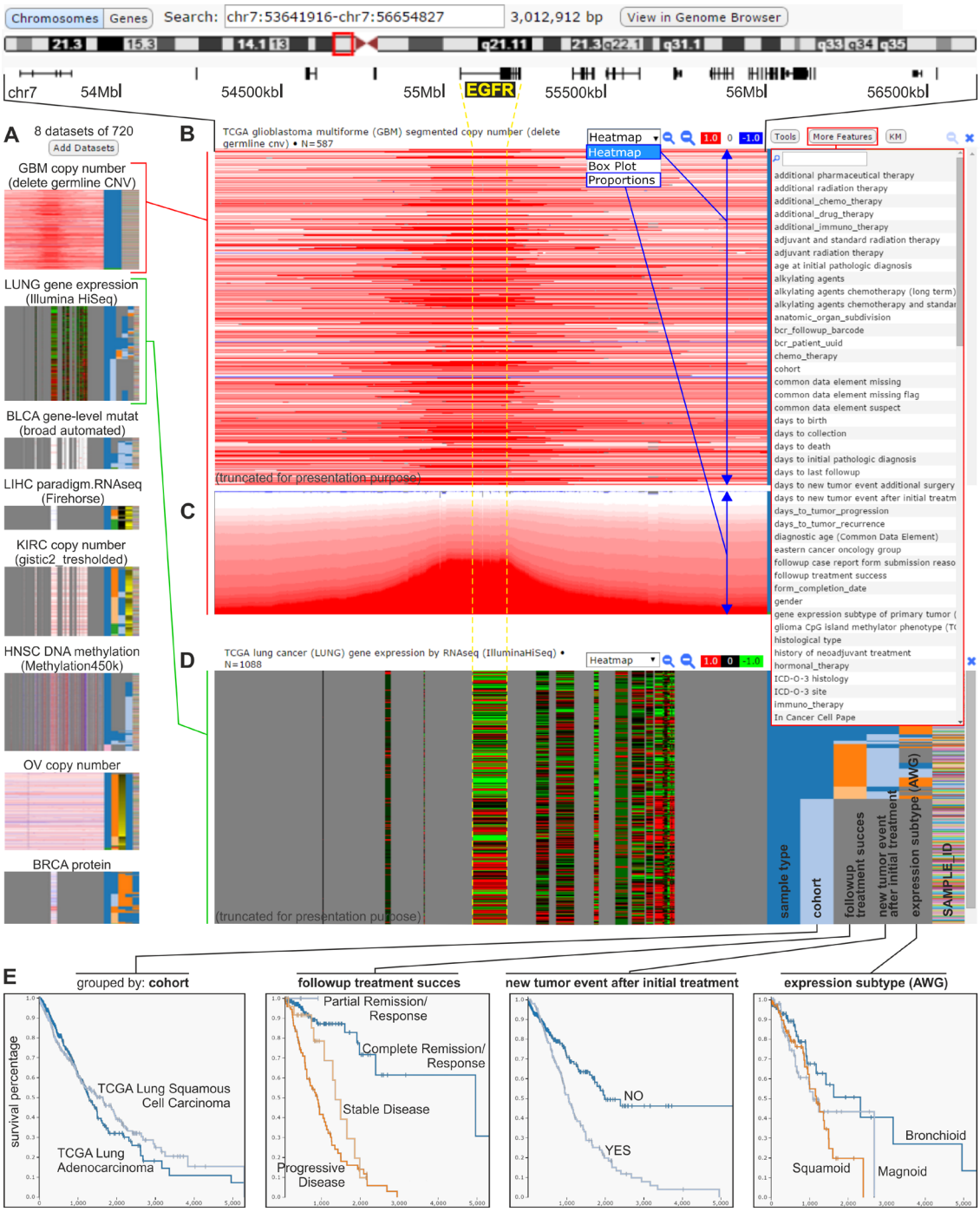


Figure 2: The UCSC Cancer Genomics Browser. An example of analysis focused on the *EGFR* genomic region that is conducted concurrently on various oncogenomic data across different cancer types and subtypes. **A.** Small-scale images (icons) of selected datasets that are simultaneously visualized in the browser. Datasets represented by icons are displayed in a column, similar to the datasets from panels B-D. **B.** A heatmap panel that presents the results of the TCGA genome-wide copy number analysis of glioblastoma multiforme (GBM) samples. A screenshot of the GBM dataset was used for presentation, based on the presence of considerable amplification of the genomic region that spans the representative *EGFR*. Each horizontal line (track) represents a specific sample. The red or blue colors indicate, respectively, a gain or loss in the copy number. On the right side of panel B, there is a drop-down list with epidemiological, clinical, and molecular attributes that can be used to sort the presented data (as shown in panel D). **C.** The TCGA copy number data identified in patients with GBM visualized as a proportions plot. **D.** A heatmap panel showing the results of TCGA analysis of gene expression in lung cancer samples in the genes that are indicated above (e.g., *EGFR*). Red and green colors indicate, respectively, upregulation and downregulation of the relative gene expression. The samples are sorted by epidemiological, clinical, and molecular attributes (selected from a drop-down list of attributes), as in panels B and C, shown on the right side of the expression panel. The copy number and expression data presented in panels B-D correspond to the same genomic region indicated above panel B. **E.** Kaplan-Meier plots generated using the attributes of lung cancer samples (shown in the right side of panel D).

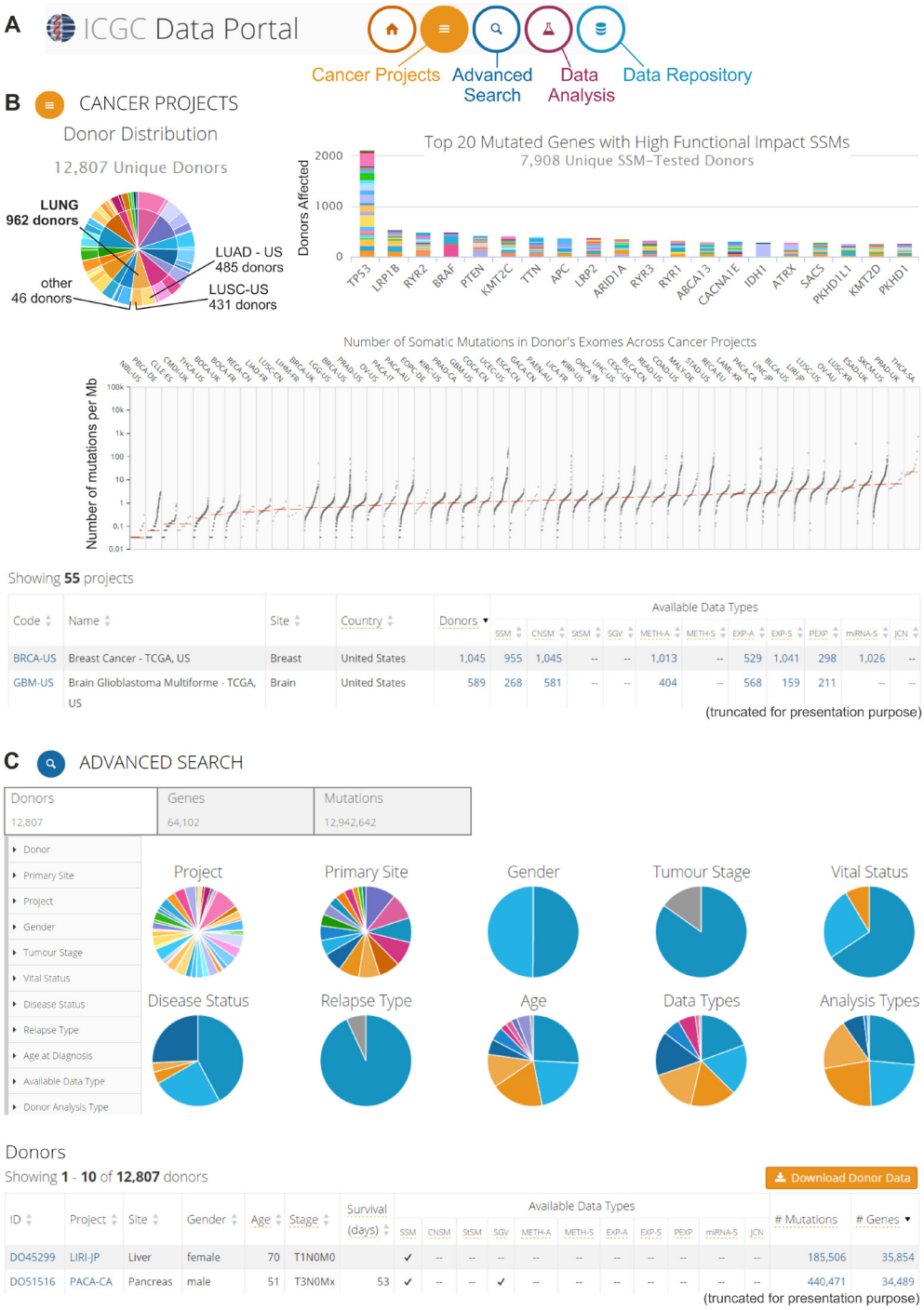


Figure 3: The ICGC Data Portal. An example of possible data analyses and visualizations. **A.** Three interactive entry points to the ICGC Data Portal. **B.** The "Cancer Projects" entry point. Screenshot of summary results from all 55 cancer projects. The upper left-hand panel: pie chart that depicts the distribution of cancer types (internal circle) and cancer subtypes/projects (external circle) among the donors, e.g., different lung cancer types and subtypes/projects (indicated in the pie chart). The upper right-hand panel: bar plot that represents the top 20 most frequently mutated genes. Different colors indicate different projects. The middle panel: scatter plot that depicts the distribution of the number of somatic mutations in the donors' exomes across cancer projects. Each dot represents the number of somatic mutations (per 1 Mb) that are identified in the analyzed sample. Vertical lines indicate the median number of mutations. The bottom part of panel B shows a summary of each project. More information about the specific project (types of experimental analyses, available genomic data, most commonly mutated genes, most common mutations, and most affected donors) can be found by clicking at specific project code. **C.** The "Advanced Search" entry point, which enables extended analysis of the oncogenomic data. This screenshot shows the browsing of donor features. The upper left-hand panel depicts features that can be used for filtering the donor data. The middle panel (pie charts) provides a summary of the clinical, epidemiological, and molecular attributes of the donors. The bottom panel represents summary data about specific donors. More information (clinical and genetic) can be found by clicking at the donor ID.

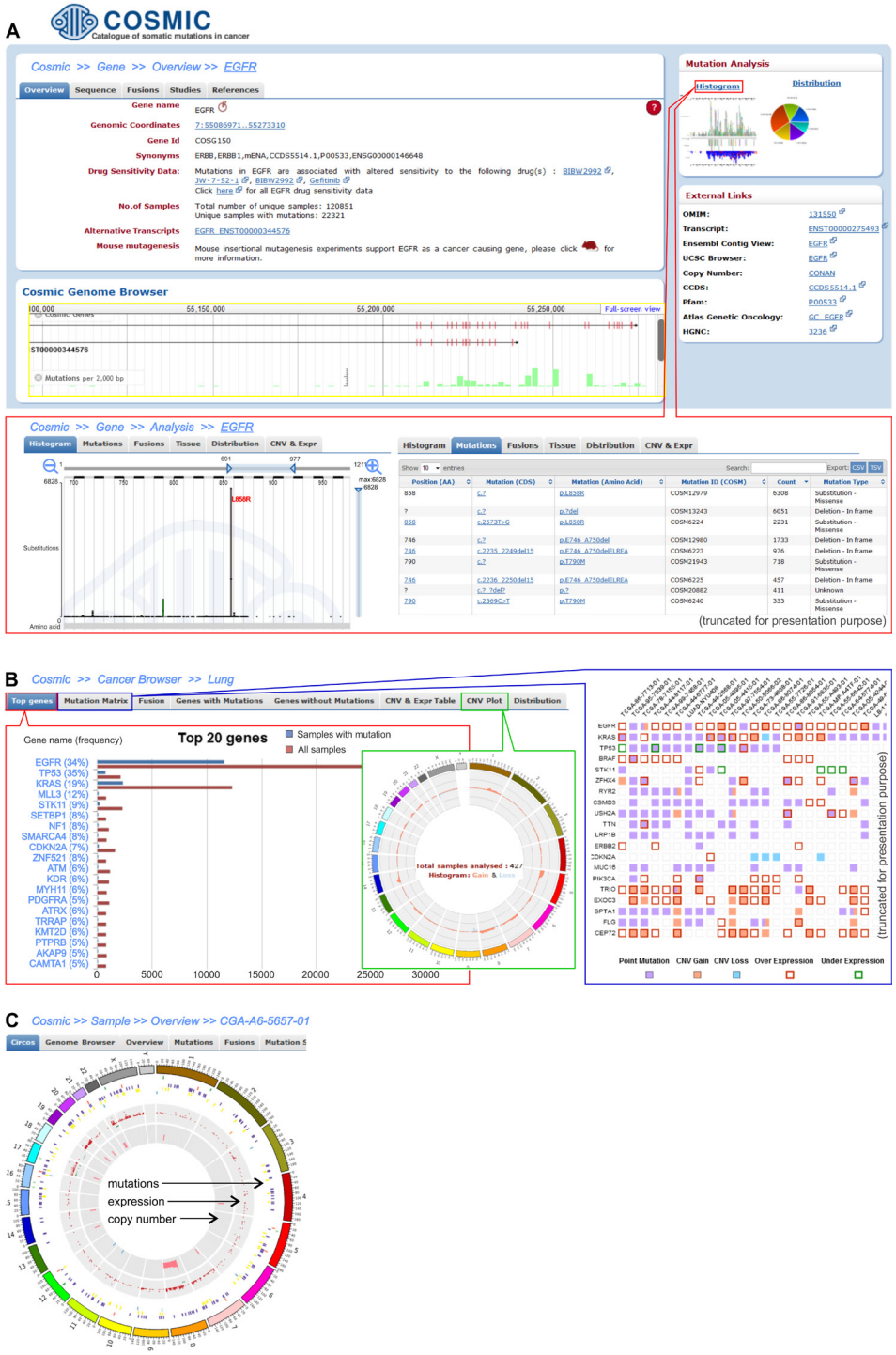


Figure 4: Three levels of data analysis in the COSMIC browser. A. Screenshot shows exemplary *EGFR* gene data. The upper left-hand panel demonstrates basic information about the gene, whereas the right-hand panel of "Mutation analysis" provides links to the detailed data of mutations that were detected in the *EGFR*. Within the panel, there is a "Histogram" link that allows detailed analysis of the gene alterations, whose features are shown in the framed panel. One of the histograms shows the distribution of *EGFR* tyrosine kinase domain mutations, with the most frequently occurring mutation being L858R. The distribution can also be visualized as a table (on the right). **B.** The screenshots present the results for the representative lung adenocarcinoma cancer type. The left framed panel shows a list of the 20 most frequently mutated genes, whereas the middle and right framed panels display a CNV plot and the Mutation Matrix, respectively. The CNV circular plot shows a summary of the copy number variations across the whole genome of the lung adenocarcinoma. The height of the corresponding bars shows the total number of samples with CNV in a specific region. The Mutation Matrix presents alterations in the most frequently mutated genes (*y*-axis) in the adenocarcinoma samples that have the highest number of alterations (*x*-axis). **C.** Circular plot of all of the alterations (coding mutations, gene expression and CNV) that are detected in an individual exemplary sample (TCGA-A6-5657-01) of adenocarcinoma.

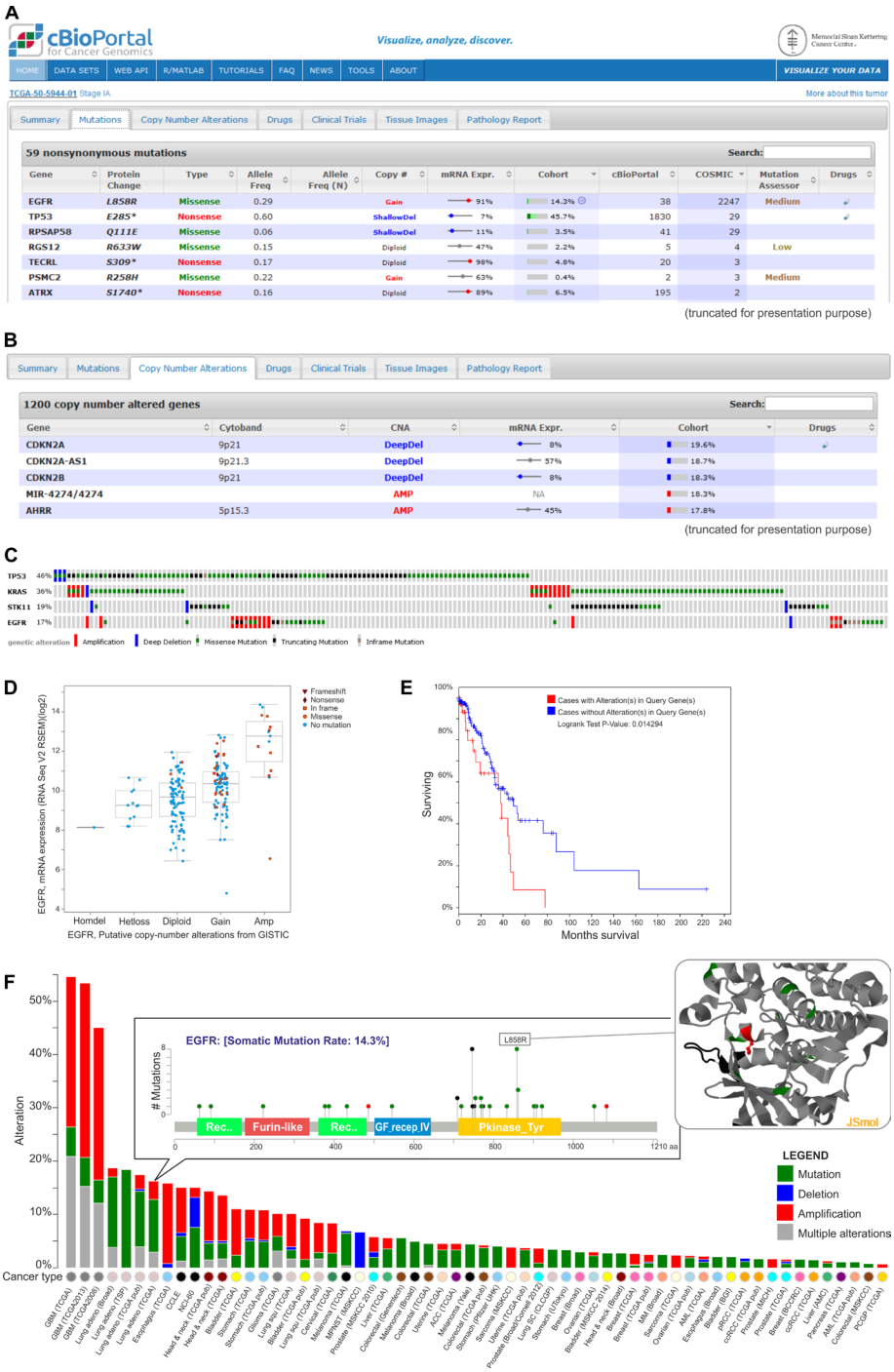


Figure 5: Exemplary data analysis and visualization available in the cBioPortal. A. The table shows nonsynonymous mutations in the TCGA-50-5944-01 sample of lung adenocarcinoma. They are characterized by the mutation name, its type, its frequency and its effect on the expression of the mutated gene. Additional information on the frequency of specific mutations can be found under the "cBioPortal" and "Cosmic" columns. The table also provides the information about the predictable impact of a given mutation on the gene function (under the Mutation Assessor tool). **B.** Genes with copy number alterations (CNAs) in the TCGA-50-5944-01 sample are shown. The table also contains the information on the frequency of CNA in a specific gene and the effect of the alterations on the gene expression. **C.** Summary of the genomic alterations in four selected genes of lung adenocarcinoma samples. Each column shows an individual tumor sample in which homozygous deletions (blue), amplifications (red), missense mutations (green squares), truncating mutations (black squares) and no mutation changes (grey) were found. **D.** A plot of the correlation between copy number alterations and mRNA expression of the exemplary *EGFR* gene. **E.** Kaplan-Meier plot of overall survival shown for patients with (red) and without (blue) changes in *EGFR*. F. Summary graph of *EGFR* alterations (shown in different colors) in individual studies deposited in the portal. For a selected study, the distribution of the mutations is shown in the inset. For a selected mutation (here L858R), a 3D interactive protein structure can be displayed (the position of the mutation is indicated in red).

COSMIC

The Catalogue of Somatic Mutations in Cancer (COSMIC) was developed at the Wellcome Trust Sanger Institute in Hinxton, UK [37, 39-43]. It is the most comprehensive database of somatic mutations in cancer. The portal provides information about the CNV and the expression level of cancer-associated genes that is obtained via the analysis of all of the samples that were tested for specific mutations (both positive and negative results are reported). This tool enables the calculation of the objectivized frequency of mutations in different types of tumors. The records included in COSMIC are derived from two sources: (i) a literature review of over 21,000 research papers and (ii) two projects: TCGA and ICGC. Together, these sources provide information that is obtained from more than a million samples. For almost 20,000 samples, whole-genome sequencing was conducted, which provided complete information about alterations in their genomes. In addition to the above, literature curation allowed the generation of the Cancer Gene Census, which is available under the COSMIC external links; thus far, it is the most reliable list of cancer-associated genes.

The data integrated in COSMIC can be searched by sample name, by gene name, and via cancer browser (Figure 4). Searching by the sample name allows the user to obtain a genome-wide overview of all of the cancer-associated events (e.g., mutations, gene fusions, and CNV) associated with a sample of interest. The second approach enables the user to overview all of the data that is related to a specific gene, such as its sequence, mutations, fusions, copy number variations, and expression. The data that refer to a specific cancer type (mutations, fusion and copy number and expression alterations of genes) can be retrieved via the cancer browser.

Due to its comprehensiveness, COSMIC is widely used and has been cited in hundreds of publications (e.g., [44-56]). For example, Chen et al. [46] used this database to confirm the presence of specific mutations in the *KRAS*, *NRAS*, and *BRAF* genes in myeloma cell lines. In another study, Ostrow and colleagues [48] took advantage of the Cancer Gene Census to select well-known cancer-associated genes for further analyses of the dynamics of the evolutionary process within tumors, with a focus on breast cancer.

cBioPortal

The cBioPortal [57-59] was developed at the Memorial Sloan-Kettering Cancer Center in New York City, NY USA. This portal contains genomic data, including copy number alterations, mRNA and microRNA expression, DNA methylation and protein and phosphoprotein abundance, which were obtained for multiple types of cancer. Currently, the portal collects records that were derived from 91 individual cancer studies, in which 31 types of cancer were analyzed with the use of over 21,000 samples. Because the tools that were integrated in the portal perform different types of analyses, different statistical tests can be used to assess the significance in specific analysis (for example, Fisher's exact test can be used to calculate the significance of mutual exclusivity of two genes or the log-rank test can be used to calculate survival analysis significance). All of the portal data can be retrieved in a format that is compatible with the R framework for statistical computing and graphics.

Cancer-associated alterations deposited in the cBioPortal can be browsed as (i) the overview of all of the genomic events that were detected in an individual cancer sample (Figure 5A, 5B), (ii) alterations in a specific gene across all of the samples that were included in one study (Figure 5C-5E), and (iii) a comparison of the frequency of the alterations in a given gene across all 91 studies (Figure 5F). For each study, it is also possible to inquire which genes are most frequently altered in the analyzed set of samples. In the cBioPortal, the genomic data are integrated with clinical outcomes, which allows determining whether a specific gene plays a potentially oncogenic role in a given cancer type. Apart from the on-line analysis of data deposited in the portal, there is also the possibility to download the results that were obtained for a specific study. Additionally, the browser enables the visualization of data that is uploaded by the user.

A wide range of tools that are available makes the portal useful in various types of analyses, which has resulted in its popularity and applicability (e.g., [51, 60-65]). For example, the authors of this paper used this portal to determine the correlation between copy number changes and expression level of two miRNA biogenesis genes (*DROSHA* and *DICER1*) that were found to be frequently amplified in lung cancer [63]. Other authors used the cBioPortal for the analysis of the *PARK2* deletion in low-grade glioma and glioblastoma and for the analysis of the correlation between *PARK2* mRNA expression and prognosis in patients [60]. Lu and colleagues used the portal to retrieve copy number data for the design of the model that predicts genetic interactions in human cancer [61].

IntOGen

The Integrative Oncogenomics Cancer Browser (IntOGen) [66] was developed by the Biomedical Genomics Group integrated in the Research Unit on Biomedical Informatics of the University Pompeu Fabra, Biomedical Research Park in Barcelona. The browser contains the results of computational secondary analyses of oncogenomic data from several large genome-wide projects. The analyses were focused on the selection

Figure 6: Exemplary results generated in the PPISURV and MIRUMIR databases. A. Results generated with the PPISURV. Survival analysis shown for representative *EGFR* and its interactome. From the top: the first table depicts the summary of *EGFR* interactions that are annotated according to different interactomes across the available datasets. The last column of the table provides a link for more detailed characteristics of a selected interactome (shown in the second table). It includes the results of the analysis of the influence of the particular interactome on survival determined for all of the available datasets. The third table presents datasets on the direct correlation between *EGFR* expression and survival. The last column of the table is a link for the visualization of the data in the Kaplan-Meier graph. The exemplary graph shows the influence of *EGFR* expression on survival in lung adenocarcinoma patients. **B.** Results generated with MIRUMIR. The table shows a summary analysis for a representative microRNA-21 on the influence of its expression on survival in a specific cancer type. The inset represents the Kaplan-Meier graph of the effect of the microRNA expression on disease-free survival in breast cancer.

of cancer-associated genes that are known as drivers. IntOGen is one of the most dynamically developing and updating oncogenomic browsers.

In the initial release of the browser, catalogued cancer data were provided in a set of three integrated web-based sub-portals, namely, the IntOGen Arrays [67], IntOGen TCGA [68], and IntOGen Mutations [69], which allowed the browsing of visualized cancer data from different perspectives. The first sub-portal, i.e., the IntOGen Arrays, exploited cancer data on genome-wide expression and copy number for analyses aimed at selecting genes and molecular pathways that are associated with specific cancer types and subtypes [67]. Analyses provided by the other two IntOGen sub-portals were performed on a partially different set of oncogenomic data but with the use of a similar rationale. In the IntOGen TCGA, the set of somatic sequence alterations identified by exome sequencing of over 3,000 tumors from 12 cancer types (TCGA pan cancer data) was used for analyses focused on the identification of cancer-associated genes, i.e., drivers [68]. The IntOGen Mutations was focused on the evaluation of the role of somatic sequence variants in carcinogenesis and the identification of cancer drivers. In addition to the TCGA data, this sub-portal took advantage of the results from other large projects, e.g., the ICGC. The portal provided results obtained via the analyses of over 4,500 cancer exomes/genomes from 13 cancer types [69]. The results previously gathered in the interactive web-based platforms are currently available in the form of downloadable databases at the IntOGen site [66].

The introduction of a new release of IntOGen (release 2014.12) was aimed at building a bridge between molecular oncogenomics and clinical practice (the personalization of medicine) [70]. Nuria Lopez-Bigas and other co-authors of the browser proposed a strategy of "*in silico* prescription" of tailored anticancer therapy. In the first stage of the strategy, a secondary computational analysis of oncogenomic data from 6,792 patients of 28 different cancer types was performed. The analysis was focused on the evaluation of the role of somatic sequence alterations (including simple somatic variants, copy number alterations and fusion events) in carcinogenesis and the identification of cancer drivers. The drivers were selected when focusing on the following factors: mutation frequency in comparison to background (MutSigCV tool [47]), the presence of highly functional mutations (Oncodrive FM tool [71]), and regional clustering of mutations (Oncodrive CLUST tool [72]) [68, 70]. Although, all of the above tools take advantage of various algorithms and statistical methods, they all are based on similar principles and utilize similar oncogenic gene features. It is important to note that all of the implemented algorithms are supported by appropriate statistical tests. Information about the 459 identified driver genes, including their "mode of action" [loss-of-function (LoF), gain-of-function (GoF) or switch-of-function (SoF)] as assessed with the use of the OncodriveROLE tool [73], is deposited in the Cancer Drivers Database. It can be either interactively visualized in the IntOGen web site [66] or downloaded for external analysis. In further stages of the strategy, Rubio-Perez and colleagues created the Cancer Drivers Actionability Database, which catalogues the already available and candidate therapies (under preliminary research or clinical trials) that are tailored to the cancer genomes of patients who were analyzed in the first stage. The Cancer Drivers Actionability Database can also be downloaded from the IntOGen website [66]. Additionally, the IntOGen portal can be exploited for the analysis of external data in the context of a single tumor or a cohort of tumors.

IntOGen is increasingly used by scientists from various cancer-associated fields for confirmation or identification of a potential driver role of genes of interest (e.g., selected based on experimental results) [74-78]. For example, Kovac and colleagues used IntOGen and MutSigCV programs for computational validation of 20 candidate papillary renal cell carcinoma (pRCC)-specific driver genes, which were selected based on the sequencing analysis of 31 exomes or genomes of pRCCs. The computational analysis of TCGA pRCC data for somatic single nucleotide variants (SNVs) in the candidate genes revealed significantly mutated genes and confirmed *SETD2, BAP1, NFE2L2* and *CUL3* as drivers, with a more modest degree of support for some other genes from a set of experimentally predefined candidates [74].

BioProfiling.de portal

The BioProfiling.de portal [79, 80] contains three distinct databases: PPISURV [81, 82], MIRUMIR [83, 84], and DRUGSURV [85, 86]. The main purpose of PPISURV [81, 82] is the identification of important cancer-associated genes that do not have direct impact on the cancer survival outcome but nevertheless affect cancer by various interactions with other genes. Such a map of connections is called a "gene interactome"; it is created based on several external databases, which deposit information about the following: direct protein interactions (deposited in the IntAct Molecular Interaction Database [87]), regulatory and signaling pathways (Reactome, NCI Pathway Interaction Database, and HumanCyc databases) [21, 36, 88], and protein post-translational modifications (PhosphoSitePlus database) [89]. PPISURV allows users to analyze the influence of the gene interactome as well as a gene of interest on survival (Figure 6A). These analyses are performed with the use of over 40 whole transcriptome expression studies that were performed with the use of approximately 8,000 samples that represent 17 types of cancer.

The MIRUMIR provides a similar type of analysis as the PPISURV; however, it is focused on the impact of specific microRNA gene expression on survival in specific

cancer types. Either MIRUMIR or PPISURV enable the visualization of survival data via Kaplan-Meier graphs, showing the influence of the expression of a gene of interest on survival in a specific cancer type (Figure 6B).

The third database that is incorporated in the BioProfiling.de portal is DRUGSURV [85]. DRUGSURV provides the opportunity to explore the survival effect of expression alterations of genes that are known to be modulated by a selected drug. This database includes information about approximately 1,700 drugs that were approved by the Food and Drug Administration (FDA), along with approximately 5,000 experimental drugs. A specific drug, cancer type or gene can be queried and investigated in terms of its anticancer potential.

The advantage of the tools that are available in the BioProfiling.de portal is that all of them provide results that are supported by appropriate statistical analysis (the R statistical package), which is not always available for the tools in the other oncogenomic portals. A false discovery rate control procedure is implemented to adjust the p-values when there is multiple testing.

The usefulness of the above-mentioned databases has been confirmed in a number of publications (e.g., [63, 90-98]). For example, Schittek et al., [90] used the PPISURV to perform survival analysis on patients who were stratified based on the expression of *CK1* gene isoforms (*CSNK1A1*, *CSNK1D*, and *CSNK1E*) in different cancers. In another study [99], MIRUMIR was used to evaluate the potential of miR-200c and miR-141 to serve as biomarkers in breast cancer.

CONCLUSIONS

Since the initiation of large-scale oncogenomic projects, a variety of databases and web-based portals have been created to enable the interactive visualization and interpretation of the abundant genome-wide cancer data. The range of available web-based portals is not limited to those described in our review. Among other noteworthy portals that provide sets of visualization tools that are helpful for oncogenomic data analysis are Oasis [100, 101], Oncomine [102, 103], Cancer Genetics Web [104], and CaSNP [105, 106]. In short, Oasis is a recently launched open-access web portal for explanatory analysis of cancer data. This portal was developed based on a custom version of the BioMart framework that was designed for oncogenomics data analysis, and it provides a unique set of visualization tools. Oncomine is another portal that provides useful visualization and analytical tools, which can browse and analyze over 715 expression and sequence alteration datasets. The Cancer Genetics Web is a web-based tool that can be used to gather literature that is related to a specific cancer type/predisposing syndrome or a gene of interest that is potentially associated with cancer. This tool provides a short summary about a disease and gene of interest, as well as a list of the latest publications and useful external links. Another interesting feature of the Cancer Genetics Web portal is a colorful panel of summarizing keywords that are available for each gene. The fourth portal is CaSNP, which gathers the results of genome-wide CNA profiling that was performed with the use of SNP arrays across 34 different cancer types. In most of the portals, the datasets and methods that are applied in their analyses and graphical presentations are continually updated. As a result, the portals deliver complex pictures of cancer genome alterations and their potential impact on cancer molecular pathogenesis. Importantly, the portals are very intuitive and address a wide community of researchers, who are not necessarily familiar with advanced computational methods. The users can take advantage of oncogenomic portals to further explore the cancer molecular basis and select new candidate cancer-associated genes for experimental validation. Regardless of current interest in exploring data that is gathered in the portals, the usefulness of the tools that are available in the oncogenomic portals will be verified in time by the users. Ultimately, it is expected that the utilization of the portals for the analysis of expanding oncogenomic data will make a substantial contribution to our understanding of cancer molecular etiology and the translation of extended cancer genomic knowledge into clinical practice.

ACKNOWLEDGMENTS

This work was supported by the National Science Centre grant 2011/01/B/NZ5/02773.

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

REFERENCES

1. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458: 719-724.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011; 144: 646-674.
3. Chin L, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. Genes Dev. 2011; 25: 534-555.
4. Dickson D. Wellcome funds cancer database. Nature. 1999; 401: 729.
5. Cancer Genome Project. https://www.sanger.ac.uk/research/projects/cancergenome/.
6. Mc Lendon R, Friedman A, Bigner D, Van Meir EG, Brat DJ, Mastrogianakis GM, Olson JJ, Mikkelsen T, Lehman N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455: 1061-1068.
7. Collins FS, Barker AD. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart

a new course across the complex landscape of human malignancies. Sci Am. 2007; 296: 50-57.

8. The Cancer Genome Atlas. http://cancergenome.nih.gov/.
9. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, et al. International network of cancer genome projects. Nature. 2010; 464: 993-998.
10. The International Cancer Genome Consortium. https://icgc.org/.
11. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N. Visualizing multidimensional cancer genomics data. Genome Med. 2013; 5: 9.
12. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010; 463: 899-905.
13. Tumorscape. http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf.
14. Zhu J, Sanborn JZ, Benz S, Szeto C, Hsu F, Kuhn RM, Karolchik D, Archie J, Lenburg ME, Esserman LJ, Kent WJ, Haussler D, Wang T. The UCSC Cancer Genomics Browser. Nat Methods. 2009; 6: 239-240.
15. Sanborn JZ, Benz SC, Craft B, Szeto C, Kober KM, Meyer L, Vaske CJ, Goldman M, Smith KE, Kuhn RM, Karolchik D, Kent WJ, Stuart JM, et al. The UCSC Cancer Genomics Browser: update 2011. Nucleic Acids Res. 2011; 39: D951-959.
16. Goldman M, Craft B, Swatloski T, Ellrott K, Cline M, Diekhans M, Ma S, Wilks C, Stuart J, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2013. Nucleic Acids Res. 2013; 41: D949-954.
17. Cline MS, Craft B, Swatloski T, Goldman M, Ma S, Haussler D, Zhu J. Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. Sci Rep. 2013; 3: 2652.
18. Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, Haussler D, Zhu J. The UCSC Cancer Genomics Browser: update 2015. Nucleic Acids Res. 2015; 43: D812-817.
19. UCSC Cancer Genomics Browser. https://genome-cancer.ucsc.edu.
20. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics. 2010; 26: i237-245.
21. NCI Pathway Interaction Database. http://pid.nci.nih.gov/.
22. Wu Y, Liu H, Shi X, Yao Y, Yang W, Song Y. The long non-coding RNA HNF1A-AS1 regulates proliferation and metastasis in lung adenocarcinoma. Oncotarget. 2015; 6: 9160-9172.
23. Mahauad-Fernandez WD, Borcherding NC, Zhang W, Okeoma CM. Bone marrow stromal antigen 2 (BST-2) DNA is demethylated in breast tumors and breast cancer cells. PLoS One. 2015; 10: e0123931.
24. Sokhi UK, Bacolod MD, Emdad L, Das SK, Dumur CI, Miles MF, Sarkar D, Fisher PB. Analysis of global changes in gene expression induced by human polynucleotide phosphorylase (hPNPase(old-35)). J Cell Physiol. 2014; 229: 1952-1962.
25. Brown DV, Daniel PM, D'Abaco GM, Gogos A, Ng W, Morokoff AP, Mantamadiotis T. Coexpression analysis of CD133 and CD44 identifies proneural and mesenchymal subtypes of glioblastoma multiforme. Oncotarget. 2015; 6: 6267-6280.
26. Yin S, Yang J, Lin B, Deng W, Zhang Y, Yi X, Shi Y, Tao Y, Cai J, Wu CI, Zhao G, Hurst LD, Zhang J, et al. Exome sequencing identifies frequent mutation of MLL2 in non-small cell lung carcinoma from Chinese patients. Sci Rep. 2014; 4: 6036.
27. Chen X, Meng J, Yue W, Yu J, Yang J, Yao Z, Zhang L. Fibulin-3 suppresses Wnt/beta-catenin signaling and lung cancer invasion. Carcinogenesis. 2014; 35: 1707-1716.
28. Zhao C, Qiao Y, Jonsson P, Wang J, Xu L, Rouhi P, Sinha I, Cao Y, Williams C, Dahlman-Wright K. Genome-wide profiling of AP-1-regulated transcription provides insights into the invasiveness of triple-negative breast cancer. Cancer Res. 2014; 74: 3983-3994.
29. Poaty H, Coullin P, Peko JF, Dessen P, Diatta AL, Valent A, Leguern E, Prevot S, Gombe-Mbalawa C, Candelier JJ, Picard JY, Bernheim A. Genome-wide high-resolution aCGH analysis of gestational choriocarcinomas. PLoS One. 2012; 7: e29426.
30. Veeraraghavan J, Tan Y, Cao XX, Kim JA, Wang X, Chamness GC, Maiti SN, Cooper LJ, Edwards DP, Contreras A, Hilsenbeck SG, Chang EC, Schiff R, et al. Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. Nat Commun. 2014; 5: 4577.
31. Xena. http://xena.ucsc.edu/.
32. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, Liang Y, Rivkin E, Wang J, Whitty B, Wong-Erasmus M, Yao L, Kasprzyk A. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. Database (Oxford). 2011; 2011: bar026.
33. ICGC Data Portal. https://dcc.icgc.org.
34. Ensembl. http://www.ensembl.org/index.html.
35. OMIM. http://www.omim.org/.
36. Reactome. http://www.reactome.org/.
37. COSMIC. http://www.sanger.ac.uk/genetics/CGP/cosmic.
38. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494: 366-370.
39. Bamford S, Dawson E, Forbes S, Clements J, Pettett R,

Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer. 2004; 91: 355-358.

40. Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, Flanagan A, Teague J, Wooster R, Futreal PA, Stratton MR. Cosmic 2005. Br J Cancer. 2006; 94: 318-322.

41. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet. 2008; Chapter 10: Unit 10 11.

42. Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. 2010; 38: D652-657.

43. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011; 39: D945-950.

44. Gaston D, Hansford S, Oliveira C, Nightingale M, Pinheiro H, Macgillivray C, Kaurah P, Rideout AL, Steele P, Soares G, Huang WY, Whitehouse S, Blowers S, et al. Germline mutations in MAP3K6 are associated with familial gastric cancer. PLoS Genet. 2014; 10: e1004669.

45. Kelley RK, Magbanua MJ, Butler TM, Collisson EA, Hwang J, Sidiropoulos N, Evason K, McWhirter RM, Hameed B, Wayne EM, Yao FY, Venook AP, Park JW. Circulating tumor cells in hepatocellular carcinoma: a pilot study of detection, enumeration, and next-generation sequencing in cases and controls. BMC Cancer. 2015; 15: 206.

46. Chen Y, Huang R, Ding J, Ji D, Song B, Yuan L, Chang H, Chen G. Multiple myeloma acquires resistance to EGFR inhibitor via induction of pentose phosphate pathway. Sci Rep. 2015; 5: 9925.

47. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499: 214-218.

48. Ostrow SL, Barshir R, DeGregori J, Yeger-Lotem E, Hershberg R. Cancer evolution is associated with pervasive positive selection on globally expressed genes. PLoS Genet. 2014; 10: e1004239.

49. Li Y, Liang C, Wong KC, Jin K, Zhang Z. Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. Nucleic Acids Res. 2014; 42: e76.

50. Gulukota K, Helseth DL, Jr., Khandekar JD. Direct observation of genomic heterogeneity through local haplotyping analysis. BMC Genomics. 2014; 15: 418.

51. Arcila ME, Drilon A, Sylvester BE, Lovly CM, Borsu L, Reva B, Kris MG, Solit DB, Ladanyi M. MAP2K1 (MEK1) Mutations Define a Distinct Subset of Lung Adenocarcinoma Associated with Smoking. Clin Cancer Res. 2015; 21: 1935-1943.

52. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, Shin A, Jee SH, Kim DH, Cai Q, Long J, Shi J, Wen W, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. Nat Genet. 2014; 46: 533-542.

53. Hu B, Ying X, Wang J, Piriyapongsa J, Jordan IK, Sheng J, Yu F, Zhao P, Li Y, Wang H, Ng WL, Hu S, Wang X, et al. Identification of a tumor-suppressive human-specific microRNA within the FHIT tumor-suppressor gene. Cancer Res. 2014; 74: 2283-2294.

54. Shtivelman E, Davies MQ, Hwu P, Yang J, Lotem M, Oren M, Flaherty KT, Fisher DE. Pathways and therapeutic targets in melanoma. Oncotarget. 2014; 5: 1701-1752.

55. Riester M, Werner L, Bellmunt J, Selvarajah S, Guancial EA, Weir BA, Stack EC, Park RS, O'Brien R, Schutz FA, Choueiri TK, Signoretti S, Lloreta J, et al. Integrative analysis of 1q23.3 copy-number gain in metastatic urothelial carcinoma. Clin Cancer Res. 2014; 20: 1873-1883.

56. Li K, Liu Y, Zhou Y, Zhang R, Zhao N, Yan Z, Zhang Q, Zhang S, Qiu F, Xu Y. An integrated approach to reveal miRNAs' impacts on the functional consequence of copy number alterations in cancer. Sci Rep. 2015; 5: 11567.

57. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2: 401-404.

58. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C, Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6: pl1.

59. cBioPortal. http://www.cbioportal.org.

60. Lin DC, Xu L, Chen Y, Yan H, Hazawa M, Doan N, Said JW, Ding LW, Liu LZ, Yang H, Yu S, Kahn M, Yin D, et al. Genomic and Functional Analysis of the E3 Ligase PARK2 in Glioma. Cancer Res. 2015; 75: 1815-1827.

61. Lu X, Megchelenbrink W, Notebaart RA, Huynen MA. Predicting human genetic interactions from cancer genome evolution. PLoS One. 2015; 10: e0125795.

62. Medico E, Russo M, Picco G, Cancelliere C, Valtorta E, Corti G, Buscarino M, Isella C, Lamba S, Martinoglio B, Veronese S, Siena S, Sartore-Bianchi A, et al. The molecular landscape of colorectal cancer cell lines unveils clinically actionable kinase targets. Nat Commun. 2015; 6: 7002.

63. Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M, Kozlowski P. High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer. Oncotarget. 2015; 6: 23399-416.

64. Powers S. Cooperation between MYC and companion 8q genes in hepatocarcinogenesis. Hepatology. 2015; 61: 757-758.

65. Chen H, Song R, Wang G, Ding Z, Yang C, Zhang J, Zeng Z, Rubio V, Wang L, Zu N, Weiskoff AM, Minze LJ, Jeyabal PV, et al. OLA1 regulates protein synthesis and integrated stress response by inhibiting eIF2 ternary complex formation. Sci Rep. 2015; 5: 13241.

66. IntOGen. http://www.intogen.org/.

67. Gundem G, Perez-Llamas C, Jene-Sanz A, Kedzierska A, Islam A, Deu-Pons J, Furney SJ, Lopez-Bigas N. IntOGen: integration and data mining of multidimensional oncogenomic data. Nat Methods. 2010; 7: 92-93.

68. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, Lawrence MS, Getz G, Bader GD, Ding L, Lopez-Bigas N. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep. 2013; 3: 2650.

69. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A, Lopez-Bigas N. IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods. 2013; 10: 1081-1082.

70. Rubio-Perez C, Tamborero D, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, Mestres J, Gonzalez-Perez A, Lopez-Bigas N. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell. 2015; 27: 382-396.

71. Gonzalez-Perez A, Lopez-Bigas N. Functional impact bias reveals cancer drivers. Nucleic Acids Res. 2012; 40: e169.

72. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics. 2013; 29: 2238-2244.

73. Schroeder MP, Rubio-Perez C, Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. Bioinformatics. 2014; 30: i549-555.

74. Kovac M, Navas C, Horswell S, Salm M, Bardella C, Rowan A, Stares M, Castro-Giner F, Fisher R, de Bruin EC, Kovacova M, Gorman M, Makino S, et al. Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution. Nat Commun. 2015; 6: 6336.

75. Pickering CR, Zhou JH, Lee JJ, Drummond JA, Peng SA, Saade RE, Tsai KY, Curry JL, Tetzlaff MT, Lai SY, Yu J, Muzny DM, Doddapaneni H, et al. Mutational landscape of aggressive cutaneous squamous cell carcinoma. Clin Cancer Res. 2014; 20: 6582-6592.

76. Moncunill V, Gonzalez S, Bea S, Andrieux LO, Salaverria I, Royo C, Martinez L, Puiggros M, Segura-Wang M, Stutz AM, Navarro A, Royo R, Gelpi JL, et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. Nat Biotechnol. 2014; 32: 1106-1112.

77. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, Pooley KA, Pritchard AL, Tiffen JC, Petljak M, Palmer JM, Symmons J, Johansson P, et al. POT1 loss-of-function variants predispose to familial melanoma. Nat Genet. 2014; 46: 478-481.

78. Yang CY, Lu RH, Lin CH, Jen CH, Tung CY, Yang SH, Lin JK, Jiang JK, Lin CH. Single nucleotide polymorphisms associated with colorectal cancer susceptibility and loss of heterozygosity in a Taiwanese population. PLoS One. 2014; 9: e100060.

79. Antonov AV. BioProfiling.de: analytical web portal for high-throughput cell biology. Nucleic Acids Res. 2011; 39: W323-327.

80. BioProfiling.de. http://bioprofiling.de/.

81. Antonov AV, Krestyaninova M, Knight RA, Rodchenkov I, Melino G, Barlev NA. PPISURV: a novel bioinformatics tool for uncovering the hidden role of specific genes in cancer survival outcome. Oncogene. 2014; 33: 1621-1628.

82. PPISURV. http://bioprofiling.de/GEO/PPISURV/ppisurv.html.

83. Antonov AV, Knight RA, Melino G, Barlev NA, Tsvetkov PO. MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. Cell Death Differ. 2013; 20: 367.

84. MIRUMIR. http://www.bioprofiling.de/GEO/MIRUMIR/mirumir.html.

85. Amelio I, Gostev M, Knight RA, Willis AE, Melino G, Antonov AV. DRUGSURV: a resource for repositioning of approved and experimental drugs in oncology based on patient survival information. Cell Death Dis. 2014; 5: e1051.

86. DRUGSURV. http://www.bioprofiling.de/GEO/DRUGSURV/index.html.

87. IntAct Molecular Interaction Database. http://www.ebi.ac.uk/intact/.

88. HumanCyc. http://humancyc.org/.

89. PhosphoSitePlus. http://www.phosphosite.org.

90. Schittek B, Sinnberg T. Biological functions of casein kinase 1 isoforms and putative roles in tumorigenesis. Mol Cancer. 2014; 13: 231.

91. Althubiti M, Lezina L, Carrera S, Jukes-Jones R, Giblett SM, Antonov A, Barlev N, Saldanha GS, Pritchard CA, Cain K, Macip S. Characterization of novel markers of senescence and their prognostic potential in cancer. Cell Death Dis. 2014; 5: e1528.

92. Antonov A, Agostini M, Morello M, Minieri M, Melino G, Amelio I. Bioinformatics analysis of the serine and glycine pathway in cancer cells. Oncotarget. 2014; 5: 11004-11013.

93. Salah Z, Arafeh R, Maximov V, Galasso M, Khawaled S, Abou-Sharieha S, Volinia S, Jones KB, Croce CM, Aqeilan RI. miR-27a and miR-27a* contribute to metastatic properties of osteosarcoma cells. Oncotarget. 2015; 6: 4920-4935.

94. Suh SS, Yoo JY, Cui R, Kaur B, Huebner K, Lee TK, Aqeilan RI, Croce CM. FHIT suppresses epithelial-mesenchymal transition (EMT) and metastasis in lung cancer through modulation of microRNAs. PLoS Genet. 2014; 10: e1004652.

95. Braun FK, Mathur R, Sehgal L, Wilkie-Grantham R, Chandra J, Berkova Z, Samaniego F. Inhibition of methyltransferases accelerates degradation of cFLIP and sensitizes B-cell lymphoma cells to TRAIL-induced apoptosis. PLoS One. 2015; 10: e0117994.

96. Li XJ, Ren ZJ, Tang JH. MicroRNA-34a: a potential therapeutic target in human cancer. Cell Death Dis. 2014; 5: e1327.

97. Tsimokha AS, Kulichkova VA, Karpova EV, Zaykova JJ, Aksenov ND, Vasilishina AA, Kropotov AV, Antonov A, Barlev NA. DNA damage modulates interactions between microRNAs and the 26S proteasome. Oncotarget. 2014; 5: 3555-3567.

98. Bongiorno-Borbone L, Giacobbe A, Compagnone M, Eramo A, De Maria R, Peschiaroli A, Melino G. Anti-tumoral effect of desmethylclomipramine in lung cancer stem cells. Oncotarget. 2015; 6: 16926-16938.

99. Antolin S, Calvo L, Blanco-Calvo M, Santiago MP, Lorenzo-Patino MJ, Haz-Conde M, Santamarina I, Figueroa A, Anton-Aparicio LM, Valladares-Ayerbes M. Circulating miR-200c and miR-141 and outcomes in patients with breast cancer. BMC Cancer. 2015; 15: 297.

100. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, Arnaiz O, Awedh MH, Baldock R, Barbiera G, Bardou P, Beck T, Blake A, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015; 43: W589-598.

101. OASIS. http://www.oasis-genomics.org/.

102. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004; 6: 1-6.

103. Oncomine. https://www.oncomine.org.

104. Cancer Genetics Web. http://www.cancerindex.org/geneweb/.

105. Cao Q, Zhou M, Wang X, Meyer CA, Zhang Y, Chen Z, Li C, Liu XS. CaSNP: a database for interrogating copy number alterations of cancer genome from SNP array data. Nucleic Acids Res. 2011; 39: D968-974.

106. CaSNP. http://cistrome.dfci.harvard.edu/CaSNP/.

OŚWIADCZENIA WSPÓŁAUTORÓW

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Poznań, 7 grudnia 2013

Dr hab. Piotr Kozłowski, prof. IChB PAN

OŚWIADCZENIA

Dotyczy rozprawy doktorskiej mgr Karola Czubaka:

Mgr Karol Czubak wykonywał pracę doktorską w Instytucie Chemii Bioorganicznej PAN od 2012 roku. Jego praca doktorska jest częścią projektu badawczego (grantu) Narodowego Centrum Nauki pod tytułem „Identyfikacja nowych genów odgrywających ważną rolę w procesie nowotworzenia w częstych nowotworach człowieka"

Jako, że od początku byłem opiekunem naukowym mgr Karola Czubaka, a od października 2015 roku również jego promotorem, miałem możliwość zauważyć, że mimo iż jest on młodym naukowcem, w pracy charakteryzuje się cechami właściwymi dojrzałym badaczom, takimi jak pracowitość, samodzielność i systematyczność. Pozwoliło mu to zdobyć doświadczenie i wiedzę odpowiednie do tego, aby wyniki swoich badań mógł z powodzeniem wpisać w szeroki kontekst ogólnego stanu wiedzy odnośnie genetyki nowotworów.

Skuteczność pracy mgr Karola Czubaka pozwoliła osiągnąć już znaczne postępy w ramach realizowanej pracy doktorskiej, czego wynikiem jest przygotowanie kilku publikacji, w powstaniu których miał on znaczący udział (w dwóch z nich jest pierwszym autorem). Moja rola, jako głównego autora wszystkich publikacji wchodzących w skład rozprawy doktorskiej mgr Karola Czubaka, polegała na zaplanowaniu i koordynacji badań, pozyskaniu środków i przygotowaniu manuskryptów.

Poniżej przedstawiam zakres prac wykonanych przez mgr Karola Czubaka, oraz mój udział w poszczególnych publikacjach:

- Lewandowska M. A.*, Czubak K.*, Klonowska K., Jozwicki W., Kowalewski J., Kozlowski P.
  The use of a two-tiered testing strategy for the simultaneous detection of small EGFR mutations and EGFR amplification in lung cancer
  PLoS One, 2015, 10:e0117983 (IF 3.23)

  Mgr Karol Czubak przeprowadził w tej pracy wszystkie eksperymenty obejmujące analizę mutacji w genie *EGFR* z wykorzystaniem testu MLPA (EGFRmut+), sekwencjonowanie, a także reakcje droplet digital PCR i Real-Time PCR mające na celu określenie liczby kopii badanych genów. Ponadto, doktorant wykonał wszystkie analizy statystyczne oraz znaczną część analizy związku mutacji genu *EGFR* z danymi klinicznymi. Mgr Karol Czubak brał również aktywny udział w przygotowaniu manuskryptu, z niewielką pomocą merytoryczną przygotował wszystkie ryciny i materiały uzupełniające.

  Mój udział w niniejszej publikacji polegał na zaplanowaniu koncepcji badań oraz nawiązaniu współpracy z Centrum Onkologii w Bydgoszczy. Nadzorowałem wszystkie eksperymenty i analizy. Moja główna rola polegała na przygotowaniu tekstu manuskryptu. Ponadto, zapoznałem doktoranta z zagadnieniami będącymi tłem i bezpośrednim przedmiotem publikacji, oraz nadzorowałem jego pracę.

- Czubak K., Lewandowska M. A., Klonowska K., Roszkowski K., Kowalewski J., Figlerowicz M., Kozlowski P.
  High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer
  Oncotarget, 2015, 6(27):23399-416 (IF 6.35)

  Rola mgr Karola Czubaka w przygotowaniu tej pracy polegała na selekcji genów do analizy, zaprojektowaniu sond i testów MLPA, przeprowadzeniu wszystkich eksperymentów, analizie statystycznej i opracowaniu rezultatów, które w celu pogłębienia analizy skonfrontował z danymi pochodzącymi z dużych projektów sekwencjonowania genomów nowotworowych. Przeprowadził on również analizę danych klinicznych oraz brał udział w przygotowaniu manuskryptu, włączając w to przygotowanie wszystkich rycin i materiałów suplementarnych.

  Mój udział w tej publikacji polegał na zaplanowaniu badań oraz koordynacji i nadzorowaniu eksperymentów i analiz. Ponadto, zapoznałem doktoranta z zagadnieniami będącymi bezpośrednim przedmiotem publikacji oraz nadzorowałem

jego pracę. Wykonałem również największą część pracy związanej z przygotowaniem tekstu manuskryptu.

- Klonowska K., Czubak K., Wojciechowska M., Handschuh L., Zmienko A., Figlerowicz M., Dams-Kozlowska H., Kozlowski P.
  Oncogenomic Portals for the Visualization and Analysis of Genome-wide Cancer Data
  Oncotarget, 2015, doi: 10.18632/oncotarget.6128 (IF 6.35)

  Rola mgr Karola Czubaka w przygotowaniu niniejszego artykułu polegała na zebraniu informacji, przygotowaniu opisów oraz sporządzeniu ilustracji dla czterech z siedmiu omawianych w nim portali onkogenomicznych, tj. Tumorscape, COSMIC, cBioPortal oraz BioProfiling.de. Wymienione portale opisane są w odpowiadających im podrozdziałach pracy. Niezbędnym w tym celu środkiem było dokonanie przez doktoranta przeglądu literatury oraz zapoznanie się ze strukturą portali oraz dostępnymi w nich narzędziami. Praca nad tym artykułem była dla doktoranta doskonałą okazją do dogłębnego zapoznania się z możliwościami wykorzystania danych generowanych w dużych projektach badania genomów nowotworowych.

  Mój udział w przygotowaniu niniejszej pracy przeglądowej polegał głównie na zaplanowaniu koncepcji pracy. Ponadto koordynowałem i nadzorowałem prace nad manuskryptem. Brałem również udział w tworzeniu ilustracji.

Proszę o kontakt w przypadku jakichkolwiek pytań odnośnie przedstawionych powyżej oświadczeń.

Piotr Kozłowski

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 2.04.2015

dr hab. n med. Marzena Anna Lewandowska, MBA
Kierownik Zakładu Genetyki i Onkologii Molekularnej
Innowacyjne Forum Medyczne
Centrum Onkologii im.Prof.F.Łukaszczyka
Ul. Dr. Romanowskiej 2
85-796 Bydgoszcz
lewandowskam@co.bydgoszcz.pl

CERTYFIKAT AKREDYTACYJNY CENTRUM MONITOROWANIA JAKOŚCI

Dotyczy współautorstwa w publikacji:

The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer

Marzena A. Lewandowska, Karol Czubak, Katarzyna Klonowska, Wojciech Jóźwicki, Janusz Kowalewski i Piotr Kozłowski

PLoS One. 2015, 10: e0117983.

TÜV NORD PN-EN ISO 9001 PN-EN ISO 14001 PN-N-18001

Oświadczam, że w powyższej pracy uczestniczyłam, jako równorzędny pierwszy autor, wspólnie z mgr Karolem Czubakiem. Powyższa praca powstała w ramach współpracy miedzy Centrum Onkologii im. Prof. Franciszka Łukaszczyka w Bydgoszczy a Instytutem Chemii Bioorganicznej PAN w Poznaniu.

EMAS Zweryfikowany system zarządzania środowiskowego REGON:PL 2.04-001-2

Moja główna rola w przygotowaniu powyższej publikacji, polegała na koordynacji zebrania, charakterystyki oraz logistyki przekazania do badań genetycznych próbek raka płuca (NSCLC ang. non-small cell lung cancer). Ponadto, w ramach rutynowych badań diagnostycznych prowadzonych w Centrum Onkologii w Bydgoszczy przeprowadziłam analizę mutacji punktowych w genie *EGFR* łącznie z ich interpretacją z wykorzystaniem komercyjnego testu RT-PCR (ang. real time PCR), EGFR-RT52 firmy Entrogen, Inc .

Wszystkie pozostałe, przedstawione w publikacji, eksperymenty i analizy, w tym analizę genu *EGFR* z wykorzystaniem metody MLPA, sekwencjonowanie, a także reakcje ddPCR i RT-PCR mające na celu określenie liczby kopii genów *EGFR*, *MET* i *ERBB2*, zostały przeprowadzone w Instytucie Chemii Bioorganicznych PAN, przez mgr Karola Czubaka.

OECI

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

dr hab. Marzena A. Lewandowska

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Mgr Katarzyna Klonowska

Oświadczenie o współautorstwie w publikacji:

Lewandowska MA*, Czubak K*, Klonowska K, Jozwicki W, Kowalewski J, Kozlowski P. *The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small EGFR Mutations and EGFR Amplification in Lung Cancer.* **PLoS One**. 2015 10: e0117983.
[*autorzy równorzędni]

Jako współautorka powyższej publikacji oświadczam, że moja rola polegała na pomocy w zaplanowaniu, przeprowadzeniu i analizie wyników ddPCR (ang. *droplet digital PCR*). Analizę tą przeprowadziłam wspólnie z mgr Karolem Czubakiem. Rezultaty przedstawione zostały w materiałach uzupełniających (Figure S1).

K. Klonowska

Katarzyna Klonowska

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 10.04.2015

Dr n med. Wojciech Jóźwicki
Koordynator Zakładu Patologii Nowotworów i Patologii
Centrum Onkologii im.Prof.F.Łukaszczyka
Ul. Dr. Romanowskiej 2
85-796 Bydgoszcz
jozwickiw@co.bydgoszcz.pl

CERTYFIKAT AKREDYTACYJNY CENTRUM MONITOROWANIA JAKOŚCI

TUV NORD PN-EN ISO 9001 PN-EN ISO 14001 PN-N-18001

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer
Lewandowska MA, Czubak K, Klonowska K, Jozwicki W, Kowalewski J, Kozlowski P.
PLoS One. 2015, 10 (2): e0117983.

Moja rola w przygotowaniu powyższej publikacji polegała na charakterystyce patomorfologicznej i selekcji próbek nowotworowych wykorzystanych w badaniach.

EMAS Zweryfikowany system zarządzania środowiskowego REGON PL 2.04-001-2

Dr n med. Wojciech Jóźwicki

OECI

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 8.04.2015

Prof. zw. dr hab. n. med. Janusz Kowalewski
Katedra Chirurgii Klatki Piersiowej i Nowotworów
Collegium Medicum, UMK w Toruniu;
Centrum Onkologii im.Prof.F.Łukaszczyka
Ul. Dr. Romanowskiej 2
85-796 Bydgoszcz

CERTYFIKAT AKREDYTACYJNY CENTRUM MONITOROWANIA JAKOŚCI

PN-EN ISO 9001
PN-EN ISO 14001
PN-N-18001

EMAS
Zweryfikowany system zarządzania środowiskowego
REGON.PL 2.04-001-2

OECI

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

The Use of a Two-Tiered Testing Strategy for the Simultaneous Detection of Small *EGFR* Mutations and *EGFR* Amplification in Lung Cancer

Lewandowska MA, Czubak K, Klonowska K, Jozwicki W, Kowalewski J, Kozlowski P.

PLoS One. 2015 Feb 26;10(2):e0117983.

Moja rola w przygotowaniu powyższej publikacji polegała na charakterystyce klinicznej pacjentów z niedrobnokomórkowym rakiem płuca.

Prof. zw. dr hab. n. med. Janusz Kowalewski

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 23.10.2015

Dr hab. n med. Marzena Anna Lewandowska, MBA
Kierownik Zakładu Genetyki i Onkologii Molekularnej
Innowacyjne Forum Medyczne
Centrum Onkologii im. Prof. F. Łukaszczyka
Ul. Dr I. Romanowskiej 2
85-796 Bydgoszcz
lewandowskam@co.bydgoszcz.pl

Dotyczy współautorstwa w publikacji:

Karol Czubak, Marzena Anna Lewandowska, Katarzyna Klonowska, Krzysztof Roszkowski, Janusz Kowalewski, Marek Figlerowicz, Piotr Kozlowski
High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer
Oncotarget. 2015, 6:23399-416.

Powyższa praca powstała w ramach współpracy między Centrum Onkologii im. Prof. Franciszka Łukaszczyka w Bydgoszczy a Instytutem Chemii Bioorganicznej PAN w Poznaniu.
Moja główna rola w przygotowaniu powyższej publikacji polegała na selekcji, zebraniu, charakterystyce oraz koordynacji i logistyce wymiany próbek i informacji kliniczno-histopatomorfologicznej próbek raka płuca (NSCLC ang. non-small cell lung cancer) zakończona zintegrowaniem powyższych informacji. Ponadto brałam udział w analizie zmian w genomie oraz ich związku z charakterystyka kliniczną, w szczególności z przeżywalnością pacjentów.
W przypadku dodatkowych pytań, proszę o kontakt.

Z poważaniem,

dr hab.n med. Marzena A. Lewandowska

Poznań 07.12.2015

Mgr Katarzyna Klonowska

Oświadczenie o współautorstwie

High copy number variation of cancer-related microRNA genes and frequent amplification of *DICER1* and *DROSHA* in lung cancer

Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M, Kozlowski P.

Oncotarget. 2015, 6:23399-416.

Jako współautorka powyższej pracy oświadczam, iż moja rola w jej przygotowaniu polegała na pomocy w wykorzystaniu portalu cBioPortal w celu określenia zależności między zmianą liczby kopii genów *DICER1* i *DROSHA,* a zmianą poziomu ekspresji tych genów.

K. Klonowska

Katarzyna Klonowska

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 23.10.2015

Dr hab. n. med. Krzysztof Roszkowski
Koordynator Oddziału Klinicznego Radioterapii
Centrum Onkologii im. Prof. F. Łukaszczyka
Ul. Dr Romanowskiej 2
85-796 Bydgoszcz

CERTYFIKAT AKREDYTACYJNY CENTRUM MONITOROWANIA JAKOŚCI

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

TÜV NORD PN-EN ISO 9001 PN-EN ISO 14001 PN-N-18001

High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer

Karol Czubak, Marzena Anna Lewandowska, Katarzyna Klonowska, Krzysztof Roszkowski, Janusz Kowalewski, Marek Figlerowicz, Piotr Kozlowski

Oncotarget. 2015, 6:23399-416.

Moja rola w przygotowaniu powyższej publikacji polegała na uaktualnieniu charakterystyki klinicznej 120 na 245 pacjentów z niedrobnokomórkowym rakiem płuca oraz wstępnej analizie przeżycia u 120 pacjentów dalej weryfikowanej i ponownie statystycznie opracowywanej na potrzeby publikacji przez głównych autorów pracy.

EMAS Zweryfikowany system zarządzania środowiskowego REGON,PL 2.04-001-2

OECI

Dr hab. n. med. Krzysztof Roszkowski

Dr hab. med. Krzysztof Roszkowski
specjalista onkologii radioterapii
4766570

CENTRUM ONKOLOGII
IM. PROF. FRANCISZKA ŁUKASZCZYKA
85-796 BYDGOSZCZ, UL. DR I. ROMANOWSKIEJ 2

tel. (52) 37 43 000; fax. (52) 37 43 301
e-mail: co@co.bydgoszcz.pl

konto: PKO Bank Polski
87 1440 1215 0000 0000 0615 4387

REGON 001255363
NIP 554-22-17-419

Bydgoszcz 23.10.2015

Prof. zw. dr hab. n. med. Janusz Kowalewski
Katedra Chirurgii Klatki Piersiowej i Nowotworów
Collegium Medicum, UMK w Toruniu;
Centrum Onkologii im. Prof. F. Łukaszczyka
Ul. Dr Romanowskiej 2
85-796 Bydgoszcz

CERTYFIKAT AKREDYTACYJNY CENTRUM MONITOROWANIA JAKOŚCI

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

TUV NORD PN-EN ISO 9001 PN-EN ISO 14001 PN-N-18001

High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer

Karol Czubak, Marzena Anna Lewandowska, Katarzyna Klonowska, Krzysztof Roszkowski, Janusz Kowalewski, Marek Figlerowicz, Piotr Kozlowski

Oncotarget. 2015, 6:23399-416.

EMAS Zweryfikowany system zarządzania środowiskowego REGON PL 2.04-001-2

Moja rola w przygotowaniu powyższej publikacji polegała na charakterystyce klinicznej pacjentów z niedrobnokomórkowym rakiem płuca.

Prof. zw. dr hab. n. med. Janusz Kowalewski

ORDYNATOR
Oddziału Klinicznego Chirurgii
Klatki Piersiowej i Nowotworów
prof. zw. dr hab. Janusz Kowalewski

OECI

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Poznań 15.11.2015

prof. dr hab. Marek Figlerowicz

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

Oświadczenie o współautorstwie w publikacji:

Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M, Kozlowski P. *High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer.* Oncotarget. 2015, 6:23399-416.

Oświadczam, że moja rola jako współautora powyższej publikacji polegała na pomocy w przygotowywaniu manuskryptu oraz dyskusji wyników otrzymanych dla genów *DICER1* i *DROSHA*.

Prof. dr hab. Marek Figlerowicz

Poznań 07.12.2015

Mgr Katarzyna Klonowska

Oświadczenie o współautorstwie

Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski

Oncotarget. 2015, [e-pub ahead of print] doi: 10.18632.

Jako współautorka powyższej publikacji oświadczam, że moja główna rola w jej przygotowaniu polegała na zebraniu informacji oraz na przygotowaniu opisów i ilustracji do 3 z 7 omówionych portali, tj. UCSC Cancer Genomics Browser, ICGC Data Portal oraz IntOGen. Wymienione portale opisane są w odpowiadających im podrozdziałach pracy. Brałam również udział w poszukiwaniu i czytaniu literatury na temat opisywanych przeze mnie portali oraz zapoznawałam się z ich zasobami i dostępnymi w nich narzędziami.

K. Klonowska

mgr Katarzyna Klonowska

Poznań 15.11.2015

dr Marzena Wojciechowska

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

Oświadczenie o współautorstwie

Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski, Oncotarget. 2015, doi: 10.18632.

Moja rola jako współautorki powyższej publikacji polegała na pomocy w przygotowaniu manuskryptu, głównie w redagowaniu tekstu pod względem językowym i strukturalnym. Do mojej roli należała również pomoc w przygotowaniu ilustracji.

dr Marzena Wojciechowska

Poznań 15.11.2015

dr Luiza Handschuh

Pracownia Mikromacierzy i Głębokiego Sekwencjonowania

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

Oświadczenie o współautorstwie

Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka, Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski

Oncotarget. 2015, doi: 10.18632.

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą była konsultacja z głównymi autorami (KK i KC) struktury niektórych omawianych baz danych oraz prezentowanych wyników.

dr Luiza Handschuh

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/ 14, 61-704 Poznań, Polska
tel.: +48-61 852 85 03, sekretariat 852 89 19
fax: +48-61 8520532 e-mail: office@ibch.poznan.pl

Poznań, 15.11.2015

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Oncogenomic portals for the visualization and analysis of genome-wide cancer data

Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski

Oncotarget. 2015, doi:10.18632/oncotarget.6128

Jako współautorka powyższej pracy oświadczam, iż mój wkład w jej powstanie polegał na konsultowaniu z głównymi autorami (KK i KC) struktury niektórych omawianych baz danych oraz prezentowanych wyników.

Agnieszka Żmieńko

dr Agnieszka Żmieńko

Zakład Biologii Molekularnej i Systemowej
Instytut Chemii Bioorganicznej PAN

e-mail: akisiel@ibch.poznan.pl
tel. 61-6653052

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Poznań 15.11.2015

prof. dr hab. Marek Figlerowicz
Zakład Biologii Molekularnej i Systemowej
Instytut Chemii Bioorganicznej PAN
Ul. Z. Noskowskiego 12/14
61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Oncogenomic portals for the visualization and analysis of genome-wide cancer data
Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski
Oncotarget. 2015, doi: 10.18632.

Oświadczam, że moja rola jako współautora powyższej publikacji polegała na krytycznym czytaniu oraz zgłaszaniu uwag w trakcie powstawania manuskryptu.

prof. dr hab. Marek Figlerowicz

WIELKOPOLSKIE CENTRUM ONKOLOGII

ZAKŁAD DIAGNOSTYKI I IMMUNOLOGII NOWOTWORÓW
PRACOWNIA IMMUNOLOGII NOWOTWORÓW

wielkopolskie centrum onkologii

ul. Garbary 15
61-866 Poznań

tel. 061 885-06-65
fax 061 852-85-02

Poznań 15.11.2015

Hanna Dams-Kozłowska, dr n. biol.
Kierownik
Pracowni Immunologii Nowotworów
Wielkopolskie Centrum Onkologii w Poznaniu
oraz
Katedra Biotechnologii Medycznej
Uniwersytet Medyczny w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy publikacji: **„Oncogenomic portals for the visualization and analysis of genome-wide cancer data”**, Katarzyna Klonowska, Karol Czubak, Marzena Wojciechowska, Luiza Handschuh, Agnieszka Zmienko, Marek Figlerowicz, Hanna Dams-Kozlowska, Piotr Kozlowski, Oncotarget. 2015, doi: 10.18632/oncotarget.6128.

Oświadczam, że moja rola w przygotowaniu powyższej publikacji polegała na udziale w opracowaniu koncepcji artykułu oraz w planowaniu wstępnej struktury manuskryptu.

Hanna Dams-Kozłowska, dr n. biol.

Kierownik Pracowni
Immunologii Nowotworów
dr n. biol. Hanna Dams-Kozłowska