Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
w Poznaniu

mgr Katarzyna Klonowska

Poszukiwanie genów związanych z predyspozycją do raka piersi ze szczególnym uwzględnieniem zmian liczby kopii odcinków DNA jako czynnika ryzyka

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

Poznań, 2017

Niniejsza praca doktorska była finansowana z grantów Narodowego Centrum Nauki:

OPUS 1 pt. „Identyfikacja nowych genów odgrywających ważną rolę w procesie nowotworzenia w częstych nowotworach człowieka” nr 2011/01/B/NZ5/02773

OPUS 9 pt. „Analiza roli genu *BARD1*, jako czynnika warunkującego genetyczną predyspozycję do raka piersi - zastosowanie unikalnej strategii, analizy asocjacji mutacji założycielskich” nr 2015/17/B/NZ2/01182

PRELUDIUM 10 pt. „Delecja genu *APOBEC3B* - analiza jej roli w rodzinnej predyspozycji do raka piersi oraz określenie jej funkcjonalności na poziomie mRNA” nr 2015/19/N/NZ5/02247

Praca została wykonana we współpracy z:

Katedrą i Zakładem Biologii i Genetyki Medycznej Gdańskiego Uniwersytetu Medycznego (prof. dr hab. n. med. Janusz Limon, dr Magdalena Ratajska)

Zakładem Genetyki i Patomorfologii Pomorskiego Uniwersytetu Medycznego w Szczecinie (prof. dr hab. n. med. Jan Lubiński)

Zakładem Ginekologii w Centrum Położnictwa i Ginekologii Uniwersyteckiego Szpitala Santaros Klinikos w Wilnie (dr Danuta Vasilevska)

W trakcie realizacji pracy doktorskiej, autorka otrzymywała specjalne stypendium naukowe z dotacji podmiotowej w ramach Krajowego Naukowego Ośrodka Wiodącego „Poznańskie Konsorcjum RNA” oraz stypendium Prezesa PAN za wybitne osiągnięcia uczestników studiów doktoranckich prowadzonych przez instytuty naukowe PAN

Pragnę serdecznie podziękować mojemu Promotorowi,
Panu dr hab. Piotrowi Kozłowskiemu, prof. IChB PAN,
za opiekę nad moim rozwojem naukowym,
wprowadzenie w interesującą tematykę badawczą,
wiele inspirujących dyskusji oraz cennych wskazówek,
a także wsparcie i wyrozumiałość.

Składam również podziękowania wszystkim Współpracownikom, a w szczególności:

Pani dr Magdalenie Ratajskiej oraz Panu prof. dr hab. n. med. Januszowi Limonowi z Katedry i Zakładu Biologii i Genetyki Medycznej Gdańskiego Uniwersytetu Medycznego

oraz

Panu dr Wojciechowi Kluźniakowi, Panu prof. dr hab. n. med. Cezaremu Cybulskiemu oraz Panu prof. dr hab. n. med. Janowi Lubińskiemu z Zakładu Genetyki i Patomorfologii Pomorskiego Uniwersytetu Medycznego w Szczecinie

za owocną współpracę oraz zaangażowanie w zrealizowane wspólnie projekty naukowe.

Wyrazy podziękowania kieruję również do Pracowników Europejskiego Centrum Bioinformatyki i Genomiki oraz Zakładu Genetyki Molekularnej IChB PAN za życzliwość i przyjazną atmosferę, która sprzyjała pracy naukowej.

Szczególne podziękowania kieruję do dr inż. Gosi Marcinkowskiej-Swojak za cenne porady, pomoc i poświęcony czas
oraz
dr Karola Czubaka za twórcze dyskusje i pomocne wskazówki.

Dziękuje również mojej Rodzinie
za okazywane wsparcie.

Niniejsza rozprawa doktorska składa się z następujących części:

STRESZCZENIE

SUMMARY

OPIS WYNIKÓW PRACY DOKTORSKIEJ

PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

1. **Klonowska K,** Ratajska M, Wojciechowska M, Kozlowski P
   Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene
   BioTechnologia, 2014, 95(3):203-214
2. Marcinkowska-Swojak M*, **Klonowska K***, Figlerowicz M, Kozlowski P
   An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs
   Gene, 2014, 546(2):257-262 (IF_{2014}= 2.14) *Autorzy mieli taki sam wkład w przygotowanie publikacji
3. **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, Irminger-Finger I, Limon J, 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_{2015}=5.23)
4. Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, **Klonowska K**, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I
   Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms
   Oncology Reports, 2015, 34(5):2609-2617 (IF_{2015}=2.49)
5. **Klonowska K**, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P
   The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population
   Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400 (IF_{2016}=5.17)
6. **Klonowska K**, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P
   MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency
   Oncotarget, 2016, 7(41):67266-67276 (IF_{2016}=5.17)

OŚWIADCZENIA WSPÓŁAUTORÓW

STRESZCZENIE

5 do 10% przypadków nowotworów piersi ma postać dziedziczną i ulega agregacji w rodzinie. Dziedziczne raki piersi są zwykle diagnozowane w młodszym wieku i często współwystępują w rodzinach z rakiem jajnika. Prowadzone dotąd badania doprowadziły do identyfikacji szeregu genów/*loci*, których mutacje związane są z predyspozycją do raka piersi, w tym genów *BRCA1* i *BRCA2*, których mutacje tłumaczą 16-40% rodzinnych przypadków raków piersi. Niemniej jednak, znaczna część czynników predysponujących do raka piersi (~50%) jest wciąż nieznana. Wśród genów-kandydatów wymienia się geny *BARD1* i *APOBEC3B*. W ostatnim czasie zasugerowano, że duże mutacje/zmiany liczby kopii odcinków DNA (CNV, ang. *copy number variants*) występujące w genach *BARD1* i *APOBEC3B* (jak również innych genach-kandydatach) mogą pełnić ważną rolę w dziedzicznej predyspozycji do nowotworów piersi i/lub jajnika.

Celem badań podjętych w ramach niniejszej pracy doktorskiej było pogłębienie wiedzy na temat czynników genetycznych związanych z predyspozycją do raka piersi, ze szczególnym uwzględnieniem dużych mutacji/CNV jako czynnika ryzyka.

W ramach niniejszej rozprawy opracowano strategię zależnej od ligacji multipleksowej amplifikacji sond (MLPA, ang. *Multiplex Ligation-dependant Probe Amplification*) do genotypowania dużych mutacji/CNV, którą następnie wykorzystano w analizie dużych mutacji w genie *BARD1*, która została przeprowadzona w dużej grupie kobiet z rakiem piersi i/lub jajnika (>800 próbek DNA). Wyniki otrzymane w powyższej analizie pozwoliły wykluczyć sugerowany wcześniej znaczący udział dużych mutacji w genie *BARD1* w predyspozycji do raka piersi i/lub jajnika. Szeroko zakrojona skala analizy, jak również wysoka jakość otrzymanych wyników, pozwoliła jednak zidentyfikować 7 mutacji punktowych genu *BARD1*, których potencjalne znaczenie funkcjonalne eksplorowano z użyciem narzędzi *in silico*.

W ramach kolejnego projektu, przeprowadzono analizę struktury i funkcji delecji genu *APOBEC3B* oraz podjęto próbę określenia jej udziału w predyspozycji do raka piersi w populacji europejskiej. Analiza pozwoliła określić dokładną strukturę delecji *APOBEC3B* i potwierdzić jej homogenność. Po raz pierwszy wykazała również wpływ delecji *APOBEC3B* na ekspresję wszystkich objętych przez nią genów, w tym genu hybrydowego *APOBEC3A/APOBEC3B*. Analiza asocjacji przeprowadzona w trzech niezależnych panelach próbek DNA (łącznie >6000 próbek) wykazała brak związku delecji *APOBEC3B* z predyspozycją do raka piersi w populacji europejskiej, co zostało dodatkowo potwierdzone w przeprowadzonej meta-analizie.

Dodatkowym wynikiem badań prowadzonych w ramach niniejszej pracy doktorskiej było opracowanie szeregu testów analitycznych, które mogą zostać wykorzystane w dalszych analizach. Wśród opracowanych testów/strategii jest oparty na MLPA test MTTE (ang. *Multipoint Test for Targeted-enrichment Efficiency*), służący do walidacji bibliotek DNA przygotowanych do sekwencjonowania eksomu na platformie NGS (WES, ang. *whole exome sequencing*).

Wyniki prezentowane w niniejszej rozprawie doktorskiej zostały opublikowane w pięciu recenzowanych artykułach eksperymentalnych. Zagadnienia dotyczące genetycznych czynników warunkujących występowanie rodzinnej postaci raka piersi i/lub jajnika oraz doniesienia naukowe na temat genów-kandydatów o sugerowanym udziale w predyspozycji do raka piersi, zgromadzone w wyniku przeglądu dostępnej literatury, zostały podsumowane w pracy przeglądowej, która również wchodzi w skład niniejszej rozprawy doktorskiej.

SUMMARY

Five to ten percent of all breast cancer cases are inherited and consequently aggregate in families. Hereditary breast cancer, on average, is diagnosed at a young age and often co-occurs in the family with ovarian cancer. Genetic studies led to the identification of different genes/*loci* predisposing to breast cancer, including *BRCA1* and *BRCA2*. It is estimated that germline mutations affecting *BRCA1* and *BRCA2* genes explain 16-40% of all familial breast cancer cases. It has to be noted, however, that a significant fraction of genetic factors predisposing to breast cancer (~50%) is still unknown. *BARD1* and *APOBEC3B* are among candidate genes, potentially associated with breast cancer predisposition. Recently, it was suggested that large mutations/copy number variants (CNV) in *BARD1* and *APOBEC3B* (and other candidate genes) may have a significant contribution to the inherited breast cancer susceptibility.

The aim of a study conducted within this doctoral research project was to expand the knowledge on genetic factors associated with breast cancer predisposition, focusing on large mutations/CNVs as a potential risk factor.

Within this doctoral project, Multiplex Ligation-dependant Probe Amplification (MLPA) strategy for large mutations/CNV genotyping was developed and subsequently utilized in an analysis of large mutations in the *BARD1* gene conducted in a large group of women with breast and/or ovarian cancer (>800 DNA samples). The obtained results allowed to exclude a significant role of large mutations in *BARD1* in breast and/or ovarian cancer predisposition. However, a large scale of the analysis and high quality of the obtained results allowed to identify seven point mutations in the *BARD1* gene. Functional significance of the identified mutations was tested using different *in silico* tools.

Further analyses were focused on a deletion of the *APOBEC3B* gene. The *APOBEC3B* deletion study was performed to determine its structure and function as well as to assess its potential association with breast cancer in the European population. The analysis allowed to determine the exact structure of the *APOBEC3B* deletion and to confirm its homogenous character. For the first time, it allowed to observe an association between the deletion and expression of all genes affected by the deletion, including the hybrid *APOBEC3A/APOBEC3B* gene. An association analysis performed in three independent panels of DNA samples (in total >6000 DNA samples) revealed lack of the association of the *APOBEC3B* deletion with breast cancer in the European population, which was additionally confirmed by the subsequently performed meta-analysis.

Development of various molecular assays, including an MLPA-based Multipoint Test for Targeted-enrichment Efficiency (MTTE) which can be used for the evaluation of targeted/exome enrichment efficiency in DNA libraries prepared for exome sequencing in NGS platform, was an additional result of the doctoral project. All molecular assays are available to the scientific community and can be utilized in further analyses.

The results presented in this doctoral thesis have been published in five peer-reviewed experimental articles. Current knowledge on genetic factors underlying familial breast and/or ovarian cancer and reports on candidate breast cancer genes were collected from the available literature and summarized in a review article, which is also included in this doctoral thesis.

OPIS WYNIKÓW PRACY DOKTORSKIEJ

WPROWADZENIE

Epidemiologia i etiologia raka piersi i jajnika

Rak piersi jest najczęstszym nowotworem złośliwym wśród kobiet w Polsce i w innych krajach rozwiniętych. Rak piersi stanowi około 22% wszystkich zachorowań na nowotwory złośliwe u kobiet w Polsce i po raku płuca stanowi drugą co do częstości przyczynę zgonów (14%) związanych z chorobami nowotworowymi. Rak jajnika, który często współwystępuje w rodzinie z rakiem piersi, jest piątym najczęstszym nowotworem występującym u kobiet w Polsce (stanowi około 5% wszystkich zachorowań na nowotwory złośliwe) i stanowi jedną z wiodących przyczyn zgonów wśród wszystkich nowotworów (wg raportu Krajowego Rejestru Nowotworów 2014, www.onkologia.org.pl). W ostatnich latach w Polsce corocznie rozpoznaje się kilkanaście tysięcy nowych przypadków raka piersi i kilka tysięcy raków jajnika (21036 przypadków zachorowań wg raportu Krajowego Rejestru Nowotworów 2015, www.onkologia.org.pl).

Rak piersi jest chorobą niejednorodną i jego poszczególne podtypy różnią się szeregiem cech histologicznych i molekularnych, których identyfikacja pozwala ocenić właściwości kliniczne nowotworu (w tym ich stopień złośliwości i potencjał przerzutowania) i ma duże znaczenie w doborze optymalnego postępowania terapeutycznego. Według ogólnie przyjętej klasyfikacji zaproponowanej przez Światową Organizację Zdrowia (WHO, ang. *World Health Organization*), raki piersi można podzielić na (i) raki nienaciekające (*in situ*/przedinwazyjne, nie dające przerzutów) oraz (ii) raki naciekające (inwazyjne, posiadające zdolność przerzutowania). W obu kategoriach można wyróżnić (i) raki przewodowe [rak przewodowy nienaciekający (DCIS, ang. *ductal cancer in situ*) i naciekający (ang. *infiltrating ductal carcinoma*)] i (ii) zrazikowe [rak zrazikowy nienaciekający (LCIS, ang. *lobular cancer in situ*) i naciekający (ang. *infiltrating lobular carcinoma*)]. Raki przewodowe naciekające stanowią 70-80% wszystkich diagnozowanych inwazyjnych raków piersi i na podstawie obrazu mikroskopowego dzieli się je na raki o różnym stopniu dojrzałości histopatologicznej/zróżnicowaniu (stopnie G1-G3). Niezależnej klasyfikacji raków piersi dokonano na podstawie wyników uzyskanych w analizach mikromacierzowych, które doprowadziły do identyfikacji kilku typów raków piersi o odrębnych cechach molekularnych i określonych, towarzyszących cechach histologicznych. Klasyfikacji tej dokonuje się między innymi na podstawie poziomu ekspresji/statusu receptorów estrogenowych (ER) i progesteronowych (PR) oraz receptora HER2 z rodziny receptorów nabłonkowego czynnika wzrostu (Perou, 2000; Malhotra, 2010).

Do tej pory nie stworzono uniwersalnego modelu wyjaśniającego patomechanizm raka piersi. Zaobserwowano jednak, że istnieje szereg czynników środowiskowych i genetycznych, których występowanie i wzajemne oddziaływanie może modyfikować ryzyko wystąpienia raka piersi. Do najistotniejszych czynników predysponujących do raka piersi należą wiek (>50) oraz wydłużona ekspozycja na estrogeny i progesteron, warunkowana wczesnym wiekiem pierwszej miesiączki i późnym wiekiem menopauzy (Hankinson, 2004; Lux, 2006). Długotrwałe przyjmowanie doustnej antykoncepcji hormonalnej, szczególnie przed 25 rokiem życia (Kotsopoulos, 2014), również znacząco podwyższa ryzyko raka piersi, w szczególności u nosicielek mutacji w genie *BRCA1* (Narod, 2002) oraz u kobiet z rodzin z agregacją przypadków raka piersi (Grabrick, 2000). Hormonalne środki antykoncepcyjne mogą mieć jednak pozytywny wpływ na zmniejszenie ryzyka raka jajnika (Narod, 1998; McLaughlin, 2007). Wykazano również związek pomiędzy nadmiernym spożywaniem tłuszczy i nadwagą oraz spożyciem alkoholu a zwiększonym ryzykiem raka piersi (Hankinson, 2004; Lux, 2006). Agregacja przypadków raka piersi i/lub jajnika w rodzinie, w szczególności ta występująca wśród krewnych pierwszego stopnia, jest również ważnym czynnikiem ryzyka raka piersi. Wskazuje to na istotną rolę podłoża genetycznego jako czynnika determinującego występowanie raka piersi.

Agregacja nowotworów w rodzinie i zespoły dziedzicznej predyspozycji do nowotworów

Znaczna większość przypadków chorób nowotworowych ma charakter sporadyczny. Nowotwór sporadyczny jest chorobą o złożonej etiologii. Rozwój nowotworów sporadycznych wynika z ekspozycji na czynniki środowiskowe, stylu życia i/lub kombinacji czynników genetycznych o niskiej penetracji. Frakcja niektórych typów nowotworów (między innymi raka piersi, jajnika i jelita grubego) występuje w postaci rodzinnych agregacji, tj. obserwowana jest u blisko spokrewnionych osób z częstością znacznie większą niż częstość występowania tych nowotworów w populacji ogólnej. Nowotwory dziedziczne, które podlegają agregacji w rodzinie, stanowią zazwyczaj od kilku do kilkunastu procent wszystkich przypadków określonych typów nowotworów. Rodzinne agregacje nowotworów są zazwyczaj warunkowane występowaniem pojedynczych mutacji[*] utraty funkcji (ang. *loss-of-function*) w specyficznym genie supresorowym. Mutacje w określonych genach supresorowych związane są z predyspozycją do określonego typu choroby nowotworowej lub

[*] Pojęcia mutacji, wariantu sekwencji i polimorfizmu genetycznego często definiowane są niejednoznacznie. Najczęściej definiuje się je w oparciu o arbitralne kryteria podziału, takie jak częstość występowania czy funkcja badanej zmiany na poziomie molekularnym, komórkowym lub klinicznym. W niniejszej pracy pojęcie mutacji będzie używane do określenia funkcjonalnych (zaburzających funkcję genu) lub potencjalnie funkcjonalnych zmian sekwencji obserwowanych w genach związanych z predyspozycją do nowotworów lub genach-kandydatach. W pozostałych przypadkach zostanie zastosowane pojęcie wariantu sekwencji lub polimorfizmu.

określonej grupy nowotworów. Do tej pory zidentyfikowano szereg genów, których mutacje związane są z występowaniem najczęstszych typów rodzinnych agregacji nowotworów, w tym gen *APC* związany z rodzinną polipowatością gruczolakowatą jelita grubego (FAP, ang. *familial adenomatous polyposis*) czy geny *BRCA1* i *BRCA2* związane z rodzinnymi agregacjami raka piersi i/lub jajnika. Identyfikacja tych genów jest niezwykle ważna w kontekście zrozumienia patomechanizmu agregacji nowotworów w rodzinie oraz odgrywa kluczową rolę w diagnostyce tych chorób. Nadal jednak nie rozpoznano znacznej części genów, których mutacje warunkują występowanie rodzinnej postaci nowotworów. W większości przypadków, nowotwory podlegające agregacji w rodzinie są dziedziczone w sposób autosomalny dominujący o niepełnej penetracji i ich rozwój warunkowany jest przez transmisję specyficznej mutacji w linii germinalnej. Ryzyko rozwoju nowotworu dziedzicznego może być jednak modyfikowane przez styl życia oraz czynniki środowiskowe, a także inne czynniki genetyczne obserwowane u danej osoby. Najczęściej obserwowane typy zespołów dziedzicznych, wraz z genami których mutacje odpowiedzialne są za występowanie określonych typów nowotworów, są wymienione w Table 1 zamieszczonej w pracy przeglądowej (Klonowska, 2014), która wchodzi w skład niniejszej rozprawy doktorskiej.

Podłoże genetyczne dziedzicznego raka piersi i jajnika

Na podstawie analizy zgodności zachorowań wśród bliźniaków mono- i dizygotycznych oszacowano, iż wskaźnik odziedziczalności (h^2) raka piersi wynosi ~0.3, a raka jajnika ~0.2 [np. (Lichtenstein, 2000)]. Około 5 do 10% przypadków nowotworów piersi ma postać dziedziczną i ulega agregacji w rodzinie. Dziedziczne raki piersi są zwykle diagnozowane w młodszym wieku i często współwystępują w rodzinach z rakiem jajnika.

Prowadzone dotąd badania genetyczne doprowadziły do identyfikacji szeregu genów/*loci*, których mutacje związane są z predyspozycją do raka piersi. Biorąc pod uwagę częstość występowania mutacji oraz miarę warunkowanego ryzyka raka piersi, dzieli się je na trzy główne grupy.

(i) Pierwsza grupa obejmuje geny o wysokiej penetracji, których mutacje występują rzadko (u <<1% kobiet w populacji ogólnej) i są związane z wysokim ryzykiem wystąpienia raka piersi i/lub jajnika. Według ogólnie przyjętych kryteriów (Foulkes, 2008; Easton, 2015; Couch, 2017), mianem genu wysokiego ryzyka określa się gen, którego mutacja warunkuje ryzyko raka piersi co najmniej 4-5 krotnie wyższe niż ryzyko raka piersi obserwowane w populacji ogólnej. Mutacje w genach wysokiego ryzyka są związane z występowaniem rzadkich rodzinnych zespołów dziedzicznych predysponujących do nowotworów złośliwych, w tym

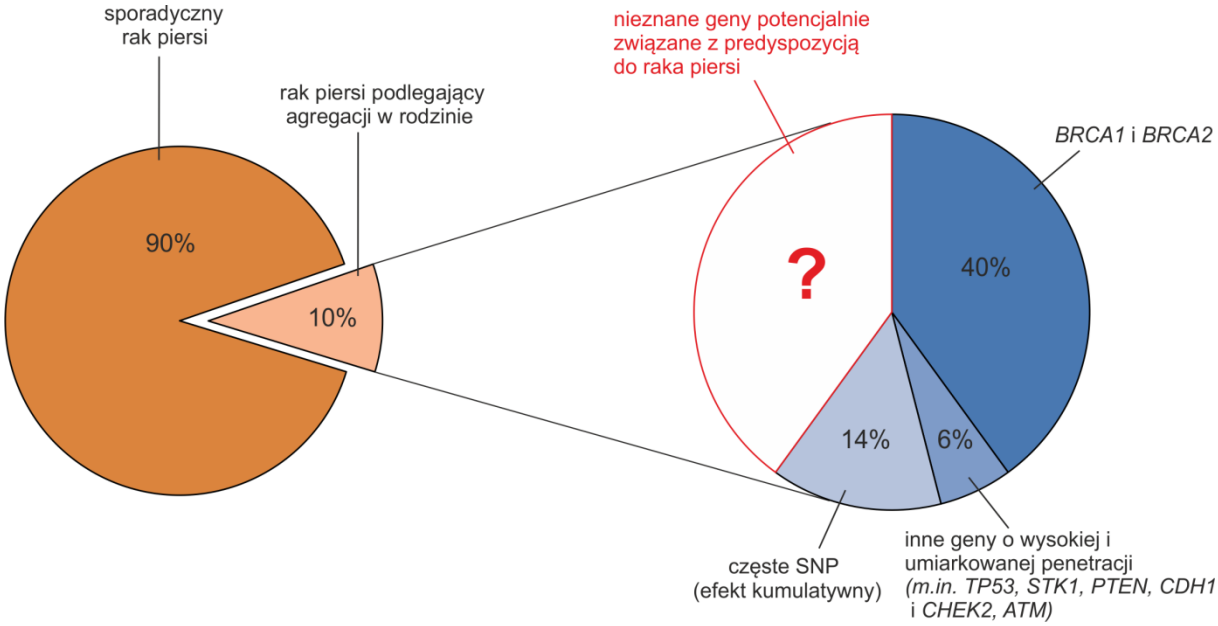
raka piersi. Mutacje genów *BRCA1* i *BRCA2* stanowią najczęstszą przyczynę występowania zespołów dziedzicznego raka piersi specyficznego narządowo (HBC, ang. *hereditary breast cancer – site specific syndrome*), dziedzicznego raka piersi-jajnika (HBOC, ang. *hereditary breast – ovarian cancer syndrome*) oraz dziedzicznego raka jajnika specyficznego narządowo (HOC, ang. *hereditary ovarian cancer – site specific syndrome*). Mutacje germinalne w genach *BRCA1* i *BRCA2* wiążą się z ryzykiem wystąpienia raka piersi w zakresie 40-90% oraz ryzykiem wystąpienia raka jajnika w zakresie 20%-50% (Ford, 1994; Antoniou, 2003; King, 2003; Chen, 2007) i tłumaczą 16-40% rodzinnych przypadków raków piersi. Mutacje w genach *TP53*, *CDH1*, *PTEN*, i *STK11* oraz genach *MSH6*, *MSH2*, *PMS2* i *MLH1* zaangażowanych w ścieżkę naprawy niesparowanych zasad (MMR, ang. *mismatch repair*) stanowią sumarycznie przyczynę <1% rodzinnych agregacji nowotworów piersi (Thompson, 2016; Slavin, 2017). Plejotropowy efekt mutacji w powyższych genach jest związany z występowaniem złożonych dziedzicznych zespołów nowotworowych, obejmujących różne typy nowotworów, w tym raka piersi [Table 1 w pracy przeglądowej (Klonowska, 2014)]. Powyższe geny klasyfikuje się jako geny wysokiego ryzyka, jednak nadal brakuje jednoznacznego, statystycznie przekonującego dowodu na temat ich roli w predyspozycji do raka piersi. Wynika to z bardzo niskiej częstości mutacji występujących w tych genach, jak również z faktu, że rak piersi stanowi tylko jeden z wielu typów nowotworów specyficznych dla poszczególnych zespołów dziedzicznych (Easton, 2015). Szeroko zakrojone analizy paneli genów-kandydatów (Thompson, 2016; Couch, 2017; Slavin, 2017), jak również analiza przeprowadzona w populacji polskiej (Cybulski, 2015a), wykazały, że mutacje genu *PALB2*, wcześniej często zaliczane do grupy umiarkowanego ryzyka, również związane są ze znacznym wzrostem ryzyka raka piersi (penetracja w zakresie 30-60%) (Antoniou, 2014; Couch, 2017) i jajnika (Norquist, 2016), i są odpowiedzialne za ~1% rodzinnych agregacji raka piersi (Thompson, 2016; Couch, 2017; Slavin, 2017). W 2015 roku dwie niezależne grupy badawcze (Cybulski, 2015b; Sun, 2015) przeprowadziły badania obejmujące analizę asocjacji genu *RECQL* z predyspozycją do raka piersi. Analiza asocjacji przeprowadzona w trzech różnych populacjach [polskiej i kanadyjskiej (Cybulski, 2015b) oraz chińskiej (Sun, 2015)], wykazała silny związek mutacji w genie *RECQL* z predyspozycją do raka piersi (5-30 krotny wzrost ryzyka raka piersi), zarówno w grupie pacjentek z nieselekcjonowanym rakiem piersi jak i wśród kobiet z agregacją raka piersi w rodzinie.

(ii) Druga grupa czynników predysponujących do raka piersi obejmuje geny, których mutacje występują z niewielką częstością (w większości u ≤1% kobiet w populacji ogólnej) i prowadzą do umiarkowanego wzrostu ryzyka raka piersi i jajnika, najczęściej 2 do 4 krotnie

wyższego niż ryzyko obserwowane w ogólnej populacji (Foulkes, 2008; Easton, 2015). Wśród genów umiarkowanej penetracji często wymienia się geny kodujące białka, które wraz z BRCA1/BRCA2 biorą udział w szlaku odpowiedzi na uszkodzenia DNA (DDR, ang. *DNA damage response*) oraz innych szlakach supresorowych. Na podstawie wstępnych doniesień szacuje się, że mutacje w genach umiarkowanego ryzyka mogą determinować występowanie około 5% przypadków rodzinnego raka piersi. W ostatnich latach opublikowano szereg prac [m.in. (Pennington, 2014; Couch, 2015; Tung, 2015; Norquist, 2016; Thompson, 2016; Buys, 2017; Couch, 2017; Kraus, 2017; Slavin, 2017)] podsumowujących wyniki analizy sekwencjonowania nowej generacji (NGS, ang. *next generation sequencing*) eksonów wybranych genów-kandydatów (w tym genów powszechnie zaliczanych do grupy genów o umiarkowanej penetracji, m.in. *ATM*, *CHEK2*, *BARD1*, *BRIP1*, *NBN*, *RAD51C*, *RAD51D*, *BLM*), przeprowadzonej w bardzo dużych grupach pacjentek z rakiem piersi i/lub jajnika (liczących 400-65000 kobiet). W ramach powyższych badań wykryto szereg funkcjonalnych zmian w sekwencji analizowanych genów, jednak doniesienia na temat miary warunkowanego przez nie efektu są często sprzeczne. Wskazuje to na potrzebę przeprowadzenia dodatkowych analiz, które pozwolą jednoznaczne określić związek genów o umiarkowanej penetracji z predyspozycją do rak piersi i/lub jajnika.

(iii) Trzecia grupa obejmuje częste polimorfizmy (w wielu przypadkach występujące z częstością >5% w ogólnej populacji) o niskiej penetracji, warunkujące niewielki wzrost ryzyka raka piersi [iloraz szans (OR, ang. *odds ratio*) ~ 1,2]. Ostatnio, szeroko zakrojone badania przeprowadzone z wykorzystaniem analizy asocjacji w skali całego genomu (GWAS, ang. *genome wide association study*), doprowadziły do identyfikacji związku licznych (94 *loci*) częstych polimorfizmów pojedynczych nukleotydów (SNP, ang. *single nucleotide polymorphismsm*) z predyspozycją do raka piersi. Szacuje się, że conajmniej 14% rodzinnych nowotworów piersi może wynikać z kumulatywnego efektu częstych polimorfizmów (Eccles, 2013; Melchor, 2013; Michailidou, 2013; Michailidou, 2015).

Podsumowując, wszystkie poznane dotąd czynniki genetyczne tłumaczą jedynie około 50% rozpoznanych rodzinnych raków piersi (Melchor, 2013) (Rycina 1). Zgodnie z założeniami poligenowej determinacji genetycznej, agregacja raka piersi i/lub jajnika w rodzinie, nie związana z mutacjami występującymi w genach wysokiej penetracji (w tym głównie *BRCA1* i *BRCA2)*, może wynikać z addytywnego efektu mutacji występujących w genach o umiarkowanej lub niskiej penetracji, który może być dodatkowo modyfikowany przez czynniki środowiskowe.

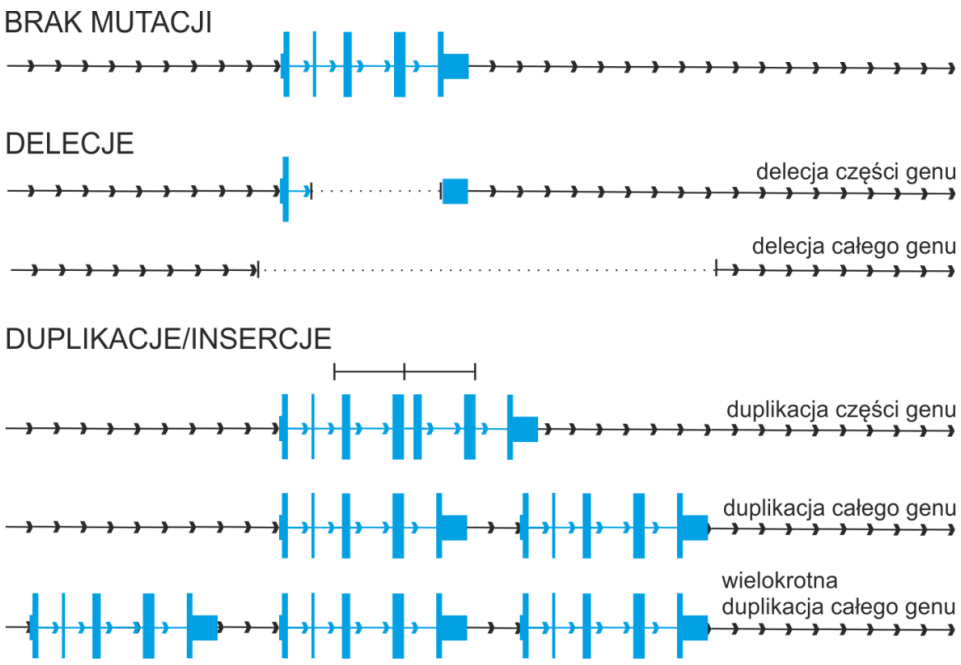


Rycina 1. *Schematyczne przedstawienie udziału czynników genetycznych predysponujących do raka piersi (stan wiedzy). Wykres z lewej strony pokazuje frakcję raków piersi występujących w postaci rodzinnej agregacji. Wykres z prawej strony, wskazuje, że mutacje w znanych genach, tłumaczą niewiele ponad 50% rodzinnych raków piersi [na podstawie Fig.1, (Klonowska, 2014)].*

Zmiany liczby kopii odcinków DNA – duże mutacje

W ostatnim czasie zasugerowano, że zmiany liczby kopii odcinków DNA (CNV, ang. *copy number variants*) występujące w genomie człowieka mogą pełnić ważną rolę w dziedzicznej predyspozycji do nowotworów (Krepischi, 2012; Long, 2013; Suehiro, 2013; Walker, 2017). Pod względem strukturalnym, CNV można umiejscowić pomiędzy małymi zmianami (zwanymi mutacjami punktowymi), obejmującymi od jednego do kilkunastu nukleotydów, a abberacjami chromosomowymi. CNV definiowane są jako delecje lub insercje segmentów DNA o wielkości od kilkuset do kilku milionów par zasad. Większość insercji to duplikacje znanych fragmentów genomu, występujące w postaci bezpośrednich, tandemowych powtórzeń (Rycina 2). CNV mogą występować zarówno w formie rzadkich mutacji (często występujących *de novo*) jak i w postaci powszechnych częstych polimorfizmów, segregujących w populacji zgodnie z prawami Mendla oraz podlegających dystrybucji zgodnej z zasadą Hardy'ego-Weinberga. Częsty (>1%) polimorfizm CNV obejmuje około 10% ludzkiego genomu (Korn, 2008; Conrad, 2010). Znane regiony CNV występują w obrębie wielu ważnych funkcjonalnych elementów, w tym setek genów kodujących białka. CNV zachodzące na geny mogą obejmować od jednego do kilku eksonów, jak również cały gen lub kilka genów. Delecje lub insercje fragmentów genomu występujące w genach związanych z chorobami o dziedziczeniu mendlowskim zalicza się do specjalnej kategorii germinalnych CNV, zwanych dużymi mutacjami/insercjami, które stanowią zazwyczaj od

kilku do kilkudziesięciu procent wszystkich mutacji obserwowanych w poszczególnych genach, (np. *DMD*, *NF1*, *TSC2*) (White, 2006; Wimmer, 2006; Kozlowski, 2007). Duże mutacje występują również w genach związanych z dziedzicznymi zespołami nowotworowymi, w tym szczególnie często w genie *BRCA1* związanym z dziedzicznym rakiem piersi i jajnika, gdzie indukowane są przez występowanie pseudogenu warunkującego niealleliczną rekombinację homologiczną (NAHR, ang. *non allelelic homologous recombination*) (Puget, 2002; Hogervorst, 2003; Montagna, 2003). Duże mutacje identyfikowane są również w genach-kandydatach o niesprecyzowanej roli w determinacji dziedzicznego raka piersi. W związku z powyższym sugeruje się istotną rolę dużych mutacji jako czynnika warunkującego występowanie nowotworów piersi, w tym frakcji rodzinnych raków piersi, których nie tłumaczą dotychczas poznane czynniki genetyczne (Frank, 2007; Sabatier, 2010; Krepischi, 2012; Long, 2013; Suehiro, 2013; Masson, 2014; Klonowska, 2017; Walker, 2017).



Rycina 2. *Najczęściej obserwowane rodzaje CNV (dużych mutacji) występujących w genach. Niebieski element oznacza hipotetyczny gen.*

Geny potencjalnie związane z predyspozycją do raka piersi – BARD1 i APOBEC3B

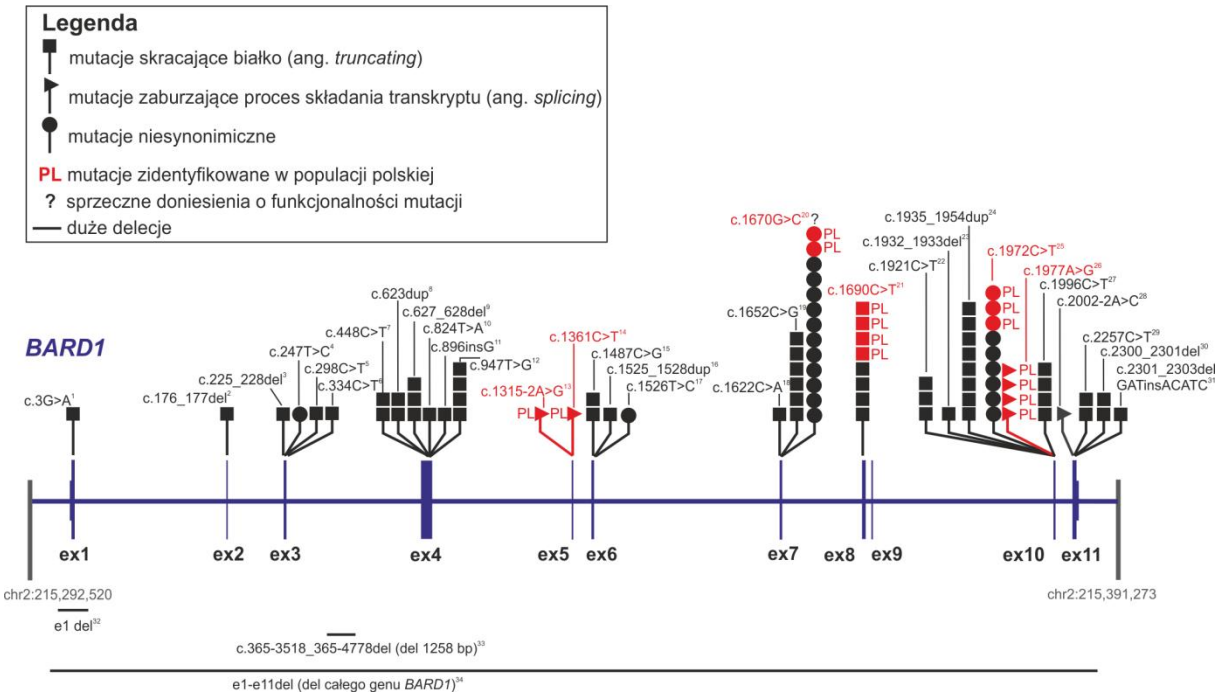
Istnieją obserwacje wskazujące, że geny kodujące białka, które wraz z BRCA1/BRCA2 biorą udział w szlaku odpowiedzi na uszkodzenia DNA oraz innych szlakach supresorowych mogą warunkować występowanie pewnej frakcji nowotworów piersi (Walsh, 2007; Foulkes, 2008). Gen *BARD1* (ang. *BRCA1 associated RING domain 1*) zaliczany jest do grupy genów-

kandydatów potencjalnie predysponujących do raka piersi. Gen *BARD1* jest zlokalizowany na długim ramieniu chromosomu 2 (2q34-35) i obejmuje 11 eksonów, kodujących białko o długości 777 aminokwasów, które wykazuje zarówno strukturalne jak i funkcjonalne podobieństwo do białka BRCA1. Oba białka posiadają domeny BRCT (ang. *BRCA1 carboxy terminal*) oraz domenę RING z motywem palca cynkowego. Domena ta umożliwia tworzenie heterodimeru BARD1/BRCA1, który stabilizuje oba białka i umożliwia BRCA1 pełnienie funkcji supresorowych (Wu, 1996). Wykazano, że mutacje zmiany sensu występujące w obrębie sekwencji kodującej domenę RING białka BRCA1 [m.in. c.181T>G (p.Cys61Gly)], zaburzające tworzenie i funkcję supresorową heterodimeru BARD1/BRCA1, są związane z wysokim ryzykiem raka piersi (Brzovic, 2001; Hashizume, 2001; Ruffner, 2001). Białko BARD1 odpowiedzialne jest również za szereg innych ważnych procesów w komórce, w których bierze udział niezależnie od BRCA1. Między innymi jest zaangażowane w proces indukcji apoptozy, który zachodzi za pośrednictwem interakcji BARD1-p53 (Irminger-Finger, 2001).

Analiza molekularna pacjentek z rakiem piersi i/lub jajnika, u których wykluczono występowanie mutacji w genach *BRCA1/2*, doprowadziła do identyfikacji różnych wariantów sekwencji genu *BARD1*. Wśród zidentyfikowanych wariantów można wyróżnić szereg funkcjonalnych i potencjalnie funkcjonalnych mutacji punktowych prowadzących do przedwczesnej terminacji translacji, zaburzeń struktury/funkcji białka lub alternatywnego procesu składania transkryptu (ang. *splicing*) [np. (Karppinen, 2004; De Brakeleer, 2010; Sabatier, 2010; Ratajska, 2012)] (Rycina 3).

Niemniej jednak, wciąż brakuje jednoznacznych, statystycznie przekonujących dowodów na rolę genu *BARD1* jako genu determinującego genetyczną predyspozycję do raka piersi. Wśród ograniczeń dotychczas przeprowadzonych analiz można wymienić (i) stosunkowo niewielkie grupy analizowanych pacjentów (w większości przypadków obejmujące 100-300 osób), a także (ii) prowadzenie badań w stosunkowo niewielkich rodzinach, uniemożliwiających wykazanie jednoznacznej (istotnej statystycznie) segregacji mutacji z chorobą.

W ostatnim czasie zasugerowano, że duże mutacje/delecje w genie *BARD1* (jak również w innych genach-kandydatach) mogą mieć znaczący udział w predyspozycji do raka piersi i w szeregu artykułów dyskutuje się potrzebę przeprowadzenia ich systematycznej analizy w odpowiednio dużej, reprezentatywnej grupie kobiet (Sabatier, 2010; Krepischi, 2012; Suehiro, 2013).



Rycina 3. *Mapa mutacji w genie BARD1 zidentyfikowanych u kobiet z rakiem piersi i/lub jajnika, w ramach następujących projektów/publikacji:*

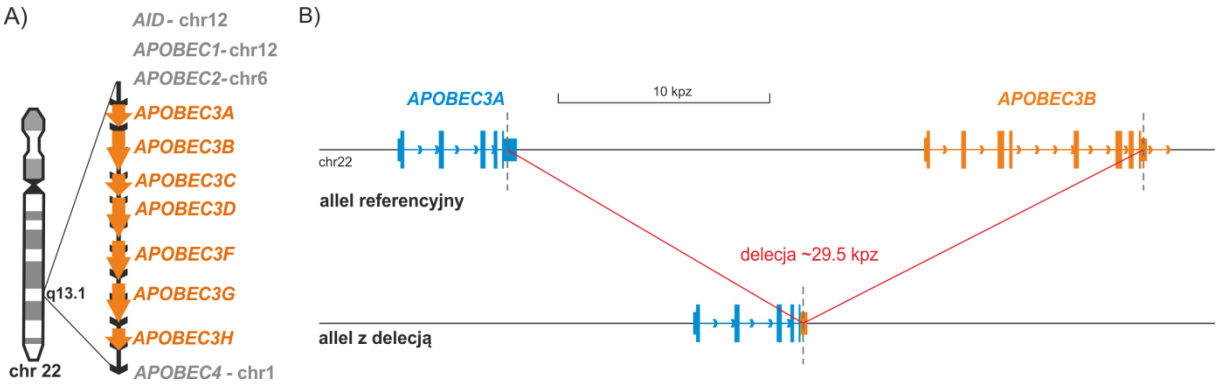
[1]Susswein i wsp., Genet Med, 2016 [2]Castera i wsp., Eur J Hum Genet, 2014 [3]Hamameh i wsp., Int J Cancer, 2017 [4]Hamameh i wsp., Int J Cancer, 2017 [5]Tung i wsp., Cancer, 2015 [6]Gonzalez-Rivera i wsp., Breast Cancer Res Treat, 2016 [7]Schoolmeester i wsp., Hum Pathol, 2017; Susswein i wsp., Genet Med, 2016 [8]Ramus i wsp., J Natl Cancer Inst, 2015; Susswein i wsp., Genet Med, 2016 [9]Ramus i wsp., J Natl Cancer Inst, 2015; Thompson i wsp., J Clin Oncol, 2016; Li i wsp., J Med Genet, 2016 [10]Couch i wsp., J Clin Oncol, 2015 [11]Couch i wsp., J Clin Oncol, 2015 [12]Tung i wsp., Cancer, 2015; Desmond i wsp., Jama Oncol, 2015; DeLeonardis, Breast J, 2017 [13]Ratajska i wsp., Br Can Res Treat, 2012 [14]Ratajska i wsp., Oncol Rep, 2015 [15]Ng i wsp., Clin Genet, 2016 [16]Couch i wsp., J Clin Oncol, 2015 [17]De Brakeleer i wsp., Hum Mutat, 2010 [18]Gonzalez-Rivera i wsp., Breast Cancer Res Treat, 2016 [19]Thompson i wsp., J Clin Oncol, 2016; Churpek i wsp., Br Can Res Treat, 2015; Maxwell i wsp., Genet Med., 2015; Tung i wsp., Cancer, 2015; Li i wsp., J Med Genet, 2016 [20]Thai i wsp., Hum Mol Genet, 1998; Ghimenti i wsp., Genes Chromosomes Cancer, 2002; Karpinnen i wsp., J Med Genet, 2004; Vahteristo i wsp., Eur J Hum Genet, 2006; Gorringe i wsp., Br Can Res Treat, 2008; Guenard i wsp., J Hum Genet, 2009; De Brakeleer i wsp., Hum Mutat, 2010; Walsh i wsp., Proc Natl Acad Sci U S A, 2011; Ratajska i wsp., Br Can Res Treat, 2012; Gonzalez-Hormazabal, Mol Biol Rep, 2012; De Brakeleer i wsp., Clin Genet, 2016; Ramus i wsp., J Natl Cancer Inst, 2015 [21]Ratajska i wsp., Br Can Res Treat, 2012; Ratajska i wsp., Oncol Rep, 2015; Klonowska i wsp., Sci Rep, 2015; Cybulski i wsp., Clin Genet, 2015; De Brakeleer i wsp., Clin Genet, 2016; Ramus i wsp., J Natl Cancer Inst, 2015; Susswein i wsp., Genet Med, 2016 [22]De Brakeleer i wsp., Clin Genet, 2016; Ramus i wsp., J Natl Cancer Inst, 2015; Gass i wsp., Clin Case Rep, 2017; Feliubadalo, Sci Rep, 2017 [23]Couch i wsp., J Clin Oncol, 2015 [24]De Brakeleer i wsp., Hum Mutat, 2010; Couch i wsp., J Clin Oncol, 2015; Tung i wsp., Cancer, 2015; Susswein i wsp., Genet Med, 2016 [25]Thai i wsp., Hum Mol Genet, 1998; Karpinnen i wsp., J Med Genet, 2004; Vahteristo i wsp., Eur J Hum Genet, 2006; De Brakeleer i wsp., Hum Mutat, 2010; De Brakeleer i wsp., Clin Genet, 2016; Ratajska i wsp., Oncol Rep, 2015; Klonowska i wsp., Sci Rep, 2015; Ramus i wsp., J Natl Cancer Inst, 2015 [26]Ratajska i wsp., Br Can Res Treat, 2012; Ratajska i wsp., Oncol Rep, 2015; Klonowska i wsp., Sci Rep, 2015 [27]Ramus i wsp., J Natl Cancer Inst, 2015; Tung i wsp., Cancer, 2015; Susswein i wsp., Genet Med, 2016; Crawford i wsp., Br Can Res Treat, 2017 [28]Couch i wsp., J Clin Oncol, 2015 [29]Hamameh i wsp., Int J Cancer, 2017 [30]Ramus i wsp., J Natl Cancer Inst, 2015; Couch i wsp., J Clin Oncol, 2015 [31]Ng i wsp., Clin Genet, 2016 [32]Tung i wsp., Cancer, 2015 [33]Rouleau i wsp., Br Can Res Treat, 2012 [34]Sabatier i wsp., Genes Chromosomes Cancer, 2010

Na mapie nie pokazano wariantów sekwencji o charakterze funkcjonalnie neutralnym lub trudnym do interpretacji [ang. variants of unknown significance (VUS)] oraz częstych SNP związanych z nowotworami [uaktualniona wersja Fig.3 z pracy przeglądowej (Klonowska, 2014)].

Ostatnio, wśród genów-kandydatów potencjalnie związanych z występowaniem raka piersi, wymienia się również geny *APOBEC3A* i *APOBEC3B*, zlokalizowane na długim ramieniu chromosomu 22 w klastrze genów *APOBEC3*. Białka APOBEC3A i APOBEC3B należą do

rodziny AID/APOBEC obejmującej 11 deaminaz cytydyny (Rycina 4A). Białka APOBEC3A i APOBEC3B posiadają zdolność do modyfikacji jednoniciowego DNA (ssDNA, ang. *single stranded DNA*) i biorą udział w wielu ważnych procesach komórkowych, w tym w procesie wrodzonej odpowiedzi odpornościowej przeciwko retrowirusom i wirusom DNA, regulacji przemieszczania się transpozonów w genomie oraz pośredniej regulacji procesu metylacji DNA. W ostatnich latach, w licznych zespołach badawczych można zaobserwować wyraźne wzmożenie wysiłków zmierzających do zrozumienia funkcjonalnego znaczenia genów *APOBEC3A* i *APOBEC3B*. Wzrost zainteresowania genami *APOBEC3A* i *APOBEC3B* wynika z (i) obserwacji, że są one odpowiedzialne za generowanie specyficznych wzorów somatycznej hipermutacji „*kataegis*" w genomach nowotworowych (między innymi w genomie raka piersi) (Nik-Zainal, 2012; Roberts, 2012; Burns, 2013) oraz (ii) wykazania związku częstej CNV w klastrze genów *APOBEC3* (germinalna delecja *APOBEC3B*) z predyspozycją do raka piersi (Long, 2013).

Delecja *APOBEC3B* obejmuje region genomu o długości około 29.5 kpz, zlokalizowany pomiędzy eksonem 5 (część niekodujaca) genu *APOBEC3A*, a eksonem ósmym genu *APOBEC3B* i prowadzi do usunięcia całej sekwencji kodującej genu *APOBEC3B*. Sugerowano, że w wyniku delecji *APOBEC3B* może powstać gen hybrydowy *APOBEC3A/APOBEC3B*, zbudowany z sekwencji kodującej genu *APOBEC3A* (ekson 1 – ekson 5) i regionu nieulegającego translacji 3' (3'UTR, ang. *3' untranslated region*) (ekson 8) genu *APOBEC3B* (Rycina 4B).



Rycina 4. *A) Schemat lokalizacji genów z rodziny AID/APOBEC w genomie człowieka [na podstawie Figure 1 w (Kitamura, 2011)]. B) Struktura delecji APOBEC3B. Eksony genów APOBEC3A oraz APOBEC3B są zaznaczone jako pionowe prostokąty (niebieskie w genie APOBEC3A i pomarańczowe w genie APOBEC3B). Wyższe prostokąty odpowiadają sekwencji kodującej, a niższe prostokąty reprezentują sekwencję UTR. Strzałki wzdłuż linii odpowiadającym sekwencjom intronów wskazują kierunek położenia genów [na podstawie Figure 1 w (Klonowska, 2017)].*

Częstość alleliczna delecji *APOBEC3B* jest wysoka w populacjach wschodniej Azji (37%), Ameryki Środkowej (58%) i Oceanii (93%) oraz umiarkowana i niska w populacji europejskiej (6%) i afrykańskiej (1%) [(Kidd, 2007), Figure 3B w (Klonowska, 2017)]. Istnieją wskazania, że delecja genu *APOBEC3B* związana jest z predyspozycją do różnych powszechnie występujących chorób człowieka. Asocjacja delecji *APOBEC3B* z predyspozycją do raka piersi (odpowiednio OR=1.3 i 1.8, dla delecji jednej i dwóch kopii genu *APOBEC3B*) została wykazana w szeroko zakrojonej analizie asocjacji (>5500 kobiet z rakiem piersi oraz >5500 kobiet z populacji kontrolnej) przeprowadzonej przez Long i współpracowników (Long, 2013) w populacji chińskiej. Próby potwierdzenia tej asocjacji przez innych badaczy (w różnych populacjach, w tym w populacji europejskiej/kaukaskiej) prowadziły jednak często do sprzecznych wyników (Xuan, 2013; Gohler, 2016). Należy również zauważyć, że analiza asocjacji delecji *APOBEC3B* prowadzona była w panelach próbek DNA pochodzących od kobiet z rakiem piersi nieselekcjonowanym pod kątem historii rodzinnej, co wskazuje na potrzebę rozpoznania roli delecji w agregacji raka piersi w rodzinie.

CEL PRACY

Ogólnym celem badań podjętych w ramach mojej pracy doktorskiej było pogłębienie wiedzy na temat czynników genetycznych związanych z predyspozycją do raka piersi. Najważniejszym aspektem pracy było podjęcie próby określenia udziału dużych mutacji/CNV w predyspozycji do raka piersi, ze szczególnym uwzględnieniem raka piersi podlegającego agregacji w rodzinie.

Powyższy cel realizowałam w ramach następujących celów szczegółowych:

- Opracowanie i optymalizacja strategii MLPA do genotypowania dużych mutacji/CNV związanych z predyspozycją do powszechnie występujących chorób człowieka
- Określenie roli dużych mutacji genu *BARD1* w predyspozycji do raka piersi i/lub jajnika - zastosowanie opracowanej strategii MLPA do analizy dużych mutacji w genie *BARD1* w dużej grupie kobiet (>800) z rakiem piersi i/lub jajnika
- Analiza struktury i funkcji delecji genu *APOBEC3B* oraz określenie jej udziału w predyspozycji do raka piersi w populacji europejskiej - analiza asocjacji typu case-control w trzech niezależnych panelach próbek DNA (łącznie >2900 próbek DNA kobiet z rakiem piersi i/lub jajnika i >3600 odpowiednio dobranych próbek kontrolnych)

Dodatkowym celem badań prowadzonych w ramach niniejszej pracy doktorskiej było opracowanie i zastosowanie innowacyjnej strategii MTTE (ang. *Multipoint Test for Targeted-enrichment Efficiency*) opartej na MLPA, służącej do walidacji bibliotek DNA przygotowanych do sekwencjonowania eksomu na platformie NGS (WES, ang. *whole exome sequencing*). Pomysł na opracowanie testu MTTE pojawił się w trakcie przygotowania rozprawy doktorskiej i jego realizacja była możliwa dzięki wiedzy na temat metody MLPA i doświadczeniu, które zdobyłam w trakcie realizacji zadań/celów szczegółowych pracy doktorskiej.

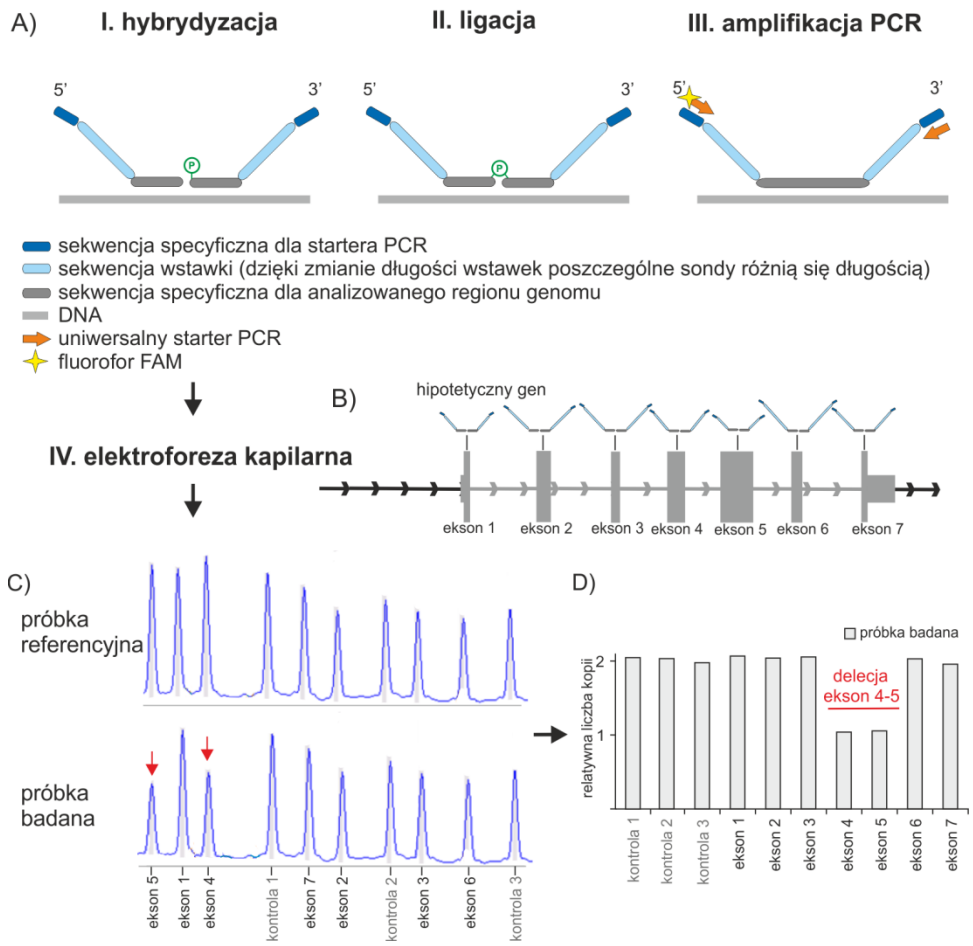
METODY

Ważną techniką wykorzystywaną w moich badaniach była zależna od ligacji multipleksowa amplifikacja sond (MLPA, ang. *Multiplex Ligation-dependant Probe Amplification*). Na potrzeby eksperymentów zaplanowanych w ramach mojej pracy doktorskiej, zaprojektowałam samodzielnie testy MLPA według strategii wcześniej opracowanej w naszym zespole (Kozlowski, 2007; Marcinkowska, 2010), która umożliwia wykorzystanie krótkich sond oligonukleotydowych, otrzymywanych w procesie syntezy chemicznej. Metoda MLPA została po raz pierwszy opisana przez Schouten i współpracowników w 2002 roku (Schouten, 2002) jako metoda służąca do analizy dużych mutacji. MLPA jest metodą opartą na multiplex PCR, która umożliwia jednoczesne zastosowanie ≤45 sond specyficznych dla różnych analizowanych regionów w genomie (na przykład eksonów, w analizie wybranego genu). Każda sonda MLPA składa się z dwóch pół-sond, specyficznych dla bezpośrednio przylegających do siebie sekwencji docelowych (ang. *target sequence*) występujących w obrębie analizowanego regionu genomu. Reakcja MLPA rozpoczyna się od denaturacji analizowanych próbek DNA oraz hybrydyzacji pół-sond do sekwencji docelowych. Pary pół-sond, które prawidłowo rozpoznały sekwencję docelową, podlegają w kolejnym etapie ligacji, a następnie amplifikacji PCR z zastosowaniem pary uniwersalnych starterów PCR. Jeden ze starterów jest wyznakowany fluorescencyjnie (fluorofor FAM), co umożliwia detekcję sygnału (produktu PCR). Jako, że sondy stosowane w jednym zestawie różnią się od siebie długością, możliwy jest ich rozdział (rozróżnienie) za pomocą elektroforezy kapilarnej. Wynikiem rozdziału produktów reakcji MLPA jest elektroferogram reprezentujący specyficzny układ pików odpowiadających sygnałom poszczególnych sond MLPA. Intensywność sygnału sond (zliczana na podstawie wysokości odpowiadających im pików) jest proporcjonalna do liczby kopii sekwencji w obrębie analizowanego regionu genomu. Relatywna liczba kopii analizowanego regionu może być określona poprzez przeprowadzenie odpowiedniej normalizacji sygnałów sond (odniesienie intensywności sygnału danej sondy do średniej intensywności sygnałów sond kontrolnych), a następnie porównanie znormalizowanych sygnałów odpowiadających poszczególnym sondom w próbkach badanych oraz próbkach referencyjnych (Rycina 5).

W moich badaniach wykorzystywałam również inne metody molekularne, takie jak PCR, RT PCR (ang. *reverse transcription PCR*), sekwencjonowanie metodą Sangera, ilościowy PCR (qPCR, ang. *quantitative PCR*) z zastosowaniem sond UPL (ang. *Universal Probe Library*), oraz ilościowa analiza PCR techniką emulsyjną (ddPCR, ang. *droplet digital PCR*).

Analiza statystyczna była bardzo istotnym elementem wszystkich etapów prowadzonych badań. Do analizy i interpretacji otrzymanych wyników stosowałam odpowiednie testy statystyczne, takie jak test Fishera, test Chi kwadrat, t-test, analiza wariancji ANOVA z testem dla trendu liniowego, czy korelacja Pearsona. W analizie asocjacji, wartości OR wraz z przedziałami ufności (95% CI) otrzymałam z wykorzystaniem jedno i wieloczynnikowej regresji logistycznej. Meta-analizę wyników badań asocjacji przeprowadziłam z wykorzystaniem metody Mantel-Haenszel.

Dodatkowo, w analizie funkcjonalności zidentyfikowanych mutacji punktowych używałam różnych narzędzi/programów *in silico*, w tym takich jak MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen2 oraz SIFT.



Rycina 5. *Schemat metody MLPA oraz analizy i interpretacji wyników. A) Etapy reakcji MLPA. Sekwencje wchodzące w skład sond MLPA zostały oznaczone odpowiednimi kolorami. B) Mapa hipotetycznego genu z zaznaczoną lokalizacją sond MLPA. W zestawie dodatkowo stosuje się sondy kontrolne, które są specyficzne dla regionów niezmiennych pod względem liczby kopii, zwykle występujących na innych chromosomach niż badany gen. C) Wyniki rozdziału produktów MLPA metodą elektroforezy kapilarnej – elektroferogramy próbki referencyjnej oraz badanej. Nazwy poszczególnych sond użytych w eksperymencie zostały umieszczone pod elektroferogramami. D) Wykres słupkowy przedstawiający stosunek intensywności sygnału poszczególnych sond w próbce badanej i referencyjnej (relatywna liczbę kopii w obrębie analizowanych regionów genomu). Przedstawiony przykład reprezentuje heterozygotyczną delecję dwóch kolejnych eksonów (4-5).*

SKRÓTOWY OPIS PUBLIKACJI WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

1. **Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P**
   Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene
   BioTechnologia, 2014, 95(3):203-214 (Klonowska, 2014)

Moją pracę doktorską rozpoczęłam od przeglądu literatury poruszającej szeroko pojętą tematykę predyspozycji do nowotworów piersi. Moje szczególne zainteresowanie koncentrowało się na zagadnieniach dotyczących genetycznych czynników warunkujących występowanie rodzinnej postaci raka piersi i/lub jajnika oraz rzadkich zespołów dziedzicznych predysponujących do nowotworów podlegających agregacji w rodzinie. W wyniku przeprowadzonego przeglądu literatury, zgromadziłam doniesienia naukowe dotyczące genów-kandydatów o sugerowanym udziale w predyspozycji do raka piersi, ze szczególnym uwzględnieniem genu *BARD1*, który stał się przedmiotem późniejszych analiz eksperymentalnych prowadzonych w ramach niniejszej pracy doktorskiej. Zgromadzoną wiedzę podsumowałam w artykule przeglądowym (Klonowska, 2014). Znaczna część zagadnień przedstawionych w niniejszym artykule przeglądowym została w skrótowej formie opisana w rozdziale WPROWADZENIE, w niniejszej pracy doktorskiej.

2. **Marcinkowska-Swojak M*, Klonowska K*, Figlerowicz M, Kozlowski P**
   An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs
   Gene, 2014, 546(2):257-262 (Marcinkowska-Swojak, 2014)
   ***Autorzy mieli taki sam wkład w przygotowanie publikacji**

Ważnym aspektem moich badań było opracowanie metody molekularnej, która spełni kryteria wysokiej precyzyjności oraz niskiej ceny, i tym samym umożliwi zastosowanie jej do analizy dużych mutacji/CNV w dużej liczbie próbek DNA w kontekście ich udziału w predyspozycji do nowotworów piersi. W związku z powyższym, w pierwszym etapie prac eksperymentalnych zaplanowanych w ramach mojej pracy doktorskiej, wraz z dr Małgorzatą Marcinkowską-Swojak brałam udział w opracowaniu i optymalizacji strategii MLPA do genotypowania dużych mutacji/CNV (Marcinkowska-Swojak, 2014). Jako model badawczy posłużyły nam trzy znane CNV o sugerowanym dużym potencjale funkcjonalnym i znaczeniu dla zdrowia człowieka, tj. dwualleliczna CNV obejmująca gen *UGT2B17* o pozytywnym wpływie na zmniejszenie ryzyka osteoporozy (Yang, 2008) oraz dwie wieloalleliczne CNV: (i) CNV obejmująca geny *CCL3L1* oraz *CCL4L1* i (ii) CNV w klastrze genów β-defensyn,

związane z podatnością na szereg chorób zakaźnych/zapalnych, w tym infekcję wirusem HIV (Gonzalez, 2005), malarię (Carpenter, 2012), łuszczycę (Hollox, 2008) i chorobę Crohn'a (Bentley, 2010). Moja rola w niniejszym projekcie polegała na przeprowadzeniu analizy MLPA wymienionych powyżej CNV w trzech panelach próbek DNA z projektu HapMap (Coriell Institute, NJ, USA) obejmujących odpowiednio (i) 48 próbek z populacji europejskiej [CEU, próbki z CEPH (Centre d'Etude du Polymorphisme)], (ii) 24 próbki z populacji chińskiej (CHB, ang. *Han Chinese in Bejing*) oraz (iii) 24 próbki z populacji afrykańskiej (YRI, ang. *Yoruba in Ibadan, Nigeria*). Analizy przeprowadziłam z wykorzystaniem testów MLPA zaprojektowanych przez dr Małgorzatę Marcinkowską-Swojak dla poszczególnych, analizowanych CNV. Każdy z zaprojektowanych testów MLPA poddałyśmy kilku etapom optymalizacji, w tym (i) walidacji/selekcji sond MLPA, (ii) selekcji optymalnej ilości sond MLPA oraz (iii) optymalizacji stosunku ilości sond kontrolnych do ilości sond specyficznych dla badanej CNV [Fig. 1 w (Marcinkowska-Swojak, 2014)]. W wyniku przeprowadzonej optymalizacji udało się wyodrębnić kilka czynników, które pozwoliły zmniejszyć zmienność sygnałów sond MLPA i jednocześnie zwiększyć rozdzielczość wyników genotypowania wielo-allelicznych CNV, przy zastosowaniu opracowanej strategii MLPA.

Podsumowując, w wyniku przeprowadzonej analizy opracowaliśmy uniwersalną metodę MLPA do genotypowania CNV występujących w dowolnym, wybranym regionie genomu człowieka. Ze względu na rzetelność generowanych wyników, a także relatywnie niski koszt analizy, proponowana metoda MLPA może zostać zastosowana do zakrojonych na szeroką skalę badań, takich jak badania asocjacji typu genotyp-fenotyp.

3. 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, Irminger-Finger I, Limon J, 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 (Klonowska, 2015)

W powyższej publikacji podsumowałam wyniki otrzymane w kolejnym etapie prac eksperymentalnych prowadzonych we współpracy z zespołem prof. Janusza Limona (Katedra i Zakład Biologii i Genetyki Medycznej, Gdański Uniwersytet Medyczny), których przedmiotem była analiza udziału dużych mutacji/CNV genu *BARD1* w predyspozycji do raka piersi i/lub jajnika. W celu przeprowadzenia analizy dużych mutacji w genie *BARD1* zaprojektowałam test MLPA z użyciem wcześniej opracowanej strategii (Marcinkowska-Swojak, 2014). Test MLPA został zaprojektowany w taki sposób, aby na każdy ekson genu

BARD1 przypadała co najmniej jedna sonda. W skład testu wchodzą również dwie sondy specyficzne dla sekwencji flankujących gen *BARD1* [Figure 1A i Supplementary Data w (Klonowska, 2015)]. Z wykorzystaniem zaprojektowanego testu, przeprowadziłam analizę próbek DNA pochodzących od >800 kobiet, w tym >500 kobiet z rodzinnym rakiem piersi i >300 kobiet z nieselekcjonowanym rakiem jajnika. Badania te były największymi dotychczas badaniami dużych mutacji w genie *BARD1*(liczba próbek większa o przynajmniej dwa rzędy wielkości). Przeprowadzona analiza nie wykazała obecności dużych mutacji w analizowanych grupach kobiet. Pozwoliła ona jednak na identyfikację trzech różnych, jednonukleotydowych substytucji [c.1690C>T (p.Gln564*), c.1972C>T (p.Arg658Cys) i c.1977A>G (p.Arg659Arg)] u siedmiu kobiet z rodzinnym rakiem piersi i/lub jajnika [Figure 1B-E w (Klonowska, 2015)]. Każda z mutacji została wykryta u przynajmniej dwóch niespokrewnionych pacjentek, co sugeruje ich założycielski charakter w populacji polskiej. Dodatkowo, mutacje te zostały wykryte w innych analizach prowadzonych przez naszych współpracowników (Ratajska, 2012; Cybulski, 2015c). W dostępnej literaturze, substytucje c.1690C>T (p.Gln564*) oraz c.1977A>G (p.Arg659Arg) określane są mianem mutacji o wysokim potencjale funkcjonalnym, natomiast c.1972C>T (p.Arg658Cys) określa się jako wariant o charakterze potencjalnie funkcjonalnym [Table 1 w (Klonowska, 2015)]. Funkcjonalność powyższych mutacji została potwierdzona przez naszych współpracowników z zespołu prof. Janusza Limona. Z zastosowaniem odpowiednich testów funkcjonalnych wykazali oni wpływ c.1977A>G (p.Arg659Arg) na alternatywne składanie transkryptu *BARD1* i przedwczesną terminację translacji (del ex2-9, p.Cys53_Trp635delinsfs*12) oraz związek wszystkich trzech mutacji z niestabilnością telomerów (Ratajska, 2012; Ratajska, 2015; Pilyugin, 2017).

Podsumowując, przeprowadzona analiza nie wykazała obecności dużych mutacji w genie *BARD1*, co pozwoliło wykluczyć sugerowany wcześniej znaczący udział dużych mutacji w genie *BARD1* w rodzinnej predyspozycji do raka piersi. Pomimo, iż identyfikacja małych mutacji w genie *BARD1* nie była bezpośrednim celem niniejszej analizy, udało nam się wykryć siedem mutacji (trzy różne zmiany), które stanowiły znaczny udział wszystkich zidentyfikowanych dotąd wariantów sekwencji genu *BARD1* o potencjale funkcjonalnym. Dodatkowym, istotnym wynikiem przeprowadzonych badań było opracowanie testu MLPA, którego szczegółowy projekt jest zamieszczony w powyższej publikacji [Supplementary Data i Supplementary Table S1 w (Klonowska, 2015)] i w przyszłości może zostać wykorzystany do dalszej analizy dużych mutacji genu *BARD1* (zarówno germinalnych jak i somatycznych) w raku piersi i/lub jajnika, jak również w innych typach nowotworów. Na przykład, w ramach

niedawno podjętej współpracy z dr Brittą Blümcke z Centrum Rodzinnego Raka Piersi i Jajnika w Szpitalu Uniwersyteckim w Kolonii, zaprojektowany test MLPA posłużył nam do niezależnej walidacji czterech dużych mutacji w genie *BARD1*, które zostały zidentyfikowane w Kolonii w analizie CNV przeprowadzonej w grupie >3500 kobiet z rakiem piersi i/lub jajnika z wykorzystaniem technologii NGS.

4. **Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I**
   Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms
   Oncology Reports, 2015, 34(5):2609-2617 (Ratajska, 2015)

Mój udział w powyższej publikacji, która została włączona do niniejszej rozprawy, jest wynikiem współpracy naszego zespołu z dr Magdaleną Ratajską i prof. Januszem Limonem (Katedra i Zakład Biologii i Genetyki Medycznej, Gdański Uniwersytet Medyczny), w ramach której byłam odpowiedzialna za przeprowadzenie jednego z etapów projektu badawczego podjętego w Gdańsku. Głównym celem projektu było określenie wpływu wybranych mutacji genu *BARD1* [w tym między innymi mutacji wcześniej zidentyfikowanych przez nas w ramach analizy opisanej w (Klonowska, 2015)] na alternatywny proces składania transkryptu *BARD1* u pacjentek z rakiem jajnika. W celu scharakteryzowania potencjału funkcjonalnego mutacji c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G genu *BARD1*, przeprowadziłam kompleksową analizę *in silico* z wykorzystaniem panelu programów, w tym MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły wykazać potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Testy funkcjonalne przeprowadzone przez naszych współpracowników potwierdziły sugerowany związek z alternatywnym procesem składania transkryptu *BARD1*, w przypadku trzech z czterech analizowanych mutacji. Wykazały one, że wszystkie trzy mutacje mają wpływ na alternatywne składanie transkryptu *BARD1* i/lub przedwczesną terminację translacji [c.1361C>T - del ex5, r.(=,1315_1395del), p.Gly439_Leu465del; c.1690C>T - del ex8, r.(=,1678_1810del), p.Gln564*; c.1977A>G – del ex2-9, r.(=,159_1903del), p.Cys53_Trp635delinsfs*12]. Wyniki przeprowadzonych przeze mnie analiz *in silico* zostały podsumowane w Table III zawartej w niniejszej publikacji (Ratajska, 2015).

5. Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P

The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population

Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400 (Klonowska, 2017)

W ramach kolejnego projektu, opisanego w powyższej publikacji wchodzącej w skład rozprawy doktorskiej, przeprowadziłam analizę struktury i funkcji delecji genu *APOBEC3B* oraz podjęłam próbę określenia jej udziału w predyspozycji do raka piersi w populacji europejskiej. Delecja *APOBEC3B* obejmuje region genomu, zlokalizowany pomiędzy eksonem 5 (część niekodujaca) genu *APOBEC3A*, a eksonem ósmym genu *APOBEC3B* i prowadzi do usunięcia całej sekwencji kodującej genu *APOBEC3B*. Sugerowano, że w wyniku delecji *APOBEC3B* może powstać gen hybrydowy *APOBEC3A/APOBEC3B*, zbudowany z sekwencji kodującej genu *APOBEC3A* (ekson 1 – ekson 5) i regionu 3'UTR (ekson 8) genu *APOBEC3B* [Figure 1A w (Klonowska, 2017)].

W pierwszym etapie zaplanowanych badań, przeprowadziłam analizę struktury delecji *APOBEC3B* z zastosowaniem 3 różnych metod molekularnych, tj. analizy sekwencjonowania Sangera oraz zaprojektowanych przez mnie testów A3Bdel_MLPA oraz A3Bdel_PCR. Przeprowadzona analiza pozwoliła określić punkty złamań delecji *APOBEC3B* z dokładnością do jednego nukleotydu. Punkty załamań zlokalizowane są w obrębie sekwencji o wysokiej homologii, co mogłoby indukować wielokrotne powstawanie delecji o podobnej, lecz nieidentycznej strukturze. Analiza przeprowadzona z zastosowaniem A3Bdel_MLPA ujawniła jednak homogenność struktury delecji [Figure 1D i Supplementary Figure 1 w (Klonowska, 2017)]. Rozpoznanie punktów złamań delecji pozwoliło mi w kolejnym kroku potwierdzić sugerowaną wcześniej obecność funkcjonalnego transkryptu hybrydowego *APOBEC3A/APOBEC3B* u nosicieli delecji *APOBEC3B* [Supplementary Figure 2 w (Klonowska, 2017)], a także odróżnić wysoce homologiczne transkrypty *APOBEC3A*, *APOBEC3B* i *APOBEC3A/APOBEC3B*. Przeprowadzona przeze mnie analiza z wykorzystaniem publicznie dostępnych danych pochodzących z profilowania mRNA w skali całego genomu oraz analiza eksperymentalna z użyciem ddPCR/UPL-qPCR ujawniły negatywną korelację pomiędzy delecją *APOBEC3B* a ekspresją genów *APOBEC3A* i *APOBEC3B* oraz pozytywną korelację pomiędzy delecją *APOBEC3B* a ekspresją hybrydowego transkryptu *APOBEC3A/APOBEC3B*. Analiza wykazała również, że ekspresja hybrydowego genu *APOBEC3A/APOBEC3B* występuje na poziomie ~10-krotnie niższym niż

poziom ekspresji kanonicznego genu *APOBEC3A*, co sugeruje udział sekwencji 3'UTR w alternatywnej regulacji tych genów [Figure 2 w (Klonowska, 2017)].

W ramach kolejnego etapu naszego projektu, przeprowadziliśmy szeroko zakrojoną analizę asocjacji delecji z ryzykiem raka piersi i jajnika, w trzech niezależnych panelach próbek DNA, obejmujących łącznie >2900 próbek DNA kobiet z rakiem piersi i/lub jajnika i >3600 odpowiednio dobranych próbek kontrolnych. Analiza asocjacji została przeprowadzona we współpracy z zespołem prof. Jana Lubińskiego (Zakład Genetyki i Patomorfologii, Pomorski Uniwersytet Medyczny w Szczecinie), dr Magdaleną Ratajską (Katedra i Zakład Biologii i Genetyki Medycznej, Gdański Uniwersytet Medyczny), oraz dr Danutą Vasilevską (Zakład Ginekologii w Centrum Położnictwa i Ginekologii, Uniwersytecki Szpital Santaros Klinikos w Wilnie). Przeprowadzona analiza wykazała brak asocjacji delecji z rakiem piersi i/lub jajnika w populacji europejskiej, w tym brak jej związku z predyspozycją do rodzinnego raka piersi [Table 1 i Table 2 w (Klonowska, 2017)]. W kolejnym kroku przeprowadziłam meta-analizę, która pozwoliła skonfrontować wyniki naszej analizy z wynikami analiz asocjacji delecji *APOBEC3B* z rakiem piersi i jajnika, a także innymi typami nowotworów, przeprowadzonymi wcześniej w różnych populacjach. Wyniki meta-analizy potwierdziły brak związku delecji z predyspozycją do raka piersi w populacji europejskiej/kaukaskiej [Figure 3 w (Klonowska, 2017)].

Podsumowując, nasza analiza pozwoliła określić dokładną strukturę delecji *APOBEC3B* i potwierdzić jej homogenność. Dostarczyła ona również bezpośredniego dowodu na to, że hybrydowy transkrypt *APOBEC3A/APOBEC3B* jest generowany z allelu z delecją *APOBEC3B*. Po raz pierwszy wykazała również wpływ delecji *APOBEC3B* na ekspresję wszystkich objętych przez nią genów, w tym *APOBEC3A/APOBEC3B*. Przeprowadzona przez nas analiza asocjacji wykazała brak związku delecji *APOBEC3B* z predyspozycją do raka piersi i jajnika w populacji Europejskiej, co zostało dodatkowo potwierdzone w przeprowadzonej meta-analizie.

6. Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P

MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency

Oncotarget, 2016, 7(41):67266-67276 (Klonowska, 2016a)

Niniejszy projekt nie wiąże się z tematyką predyspozycji do nowotworów. Pomysł na podjęcie tego projektu wyniknął z doświadczenia, które zdobyłam podczas realizacji mojej pracy doktorskiej, a w szczególności doświadczenia w projektowaniu i wykorzystaniu testów MLPA. Projekt ten polegał na opracowaniu metody MTTE, służącej do walidacji

efektywności wzbogacenia bibliotek DNA przygotowanych do analizy WES z zastosowaniem sekwencjonowania NGS [Figure 1 i Figure 2 w (Klonowska, 2016a)].

Do generowania bibliotek wzbogaconych w sekwencje eksonowe (ang. *exome enrichement*) powszechnie stosuje się kilka komercyjnie dostępnych zestawów, oferowanych między innymi przez firmy Agilent, Roche NimbleGen i Illumina. Zastosowanie dodatkowego etapu walidacji przygotowanych bibliotek DNA przed rozpoczęciem sekwencjonowania może uchronić badacza przed poniesieniem znacznych kosztów związanych z przeprowadzeniem analizy sekwencjonowania nieudanych, niewystarczająco wzbogaconych bibliotek DNA. Zaproponowana przez nas metoda MTTE pozwala scharakteryzować kilka czynników, m.in. efektywność wzbogacenia regionów kodujących (ang. *relative enrichment*), stopień wzbogacenia regionów kodujących (ang. *fold enrichment*) oraz stopień „wyczyszczenia" regionów niekodujących, niepodlegających procesowi wzbogacania (ang. *relative/overall clearance*) [Figure 3 w (Klonowska, 2016a)]. Ponieważ proponowana przez nas metoda MTTE jest nowa, przeprowadziliśmy bardzo restrykcyjną analizę walidacyjną stosując szereg technicznych, statystycznych, bioinformatycznych i genetycznych kryteriów. Porównanie wyników MTTE z wynikami niezależnej walidacji metodą qPCR oraz rzeczywistym poziomem wzbogacenia eksomu, określonym na podstawie frakcji odczytów NGS zmapowanych do regionów kodujących, wykazało wiarygodność i rzetelność proponowanej metody MTTE [Table 1 i Supplementary Figure S3 w (Klonowska, 2016a)]. Analiza wyników MTTE pozwoliła również po raz pierwszy wykazać pozytywną korelację pomiędzy zawartością puryn ($R=0.70$, $p=0.02$) w badanej sekwencji i poziomem jej wzbogacenia [Supplementary Figure S4 w (Klonowska, 2016a)].

Podsumowując, w niniejszej publikacji zaproponowaliśmy innowacyjną strategię MTTE, która może posłużyć do rzetelnej walidacji bibliotek DNA przygotowanych do sekwencjonowania eksomu. Zaproponowana strategia może zostać w łatwy sposób dostosowana do potrzeb różnych projektów NGS, prowadzonych z wykorzystaniem wzbogacania eksomu lub innych wybranych regionów genomu.

Inne publikacje, nie wchodzące w skład niniejszej rozprawy doktorskiej

Jestem również współautorką dwóch prac eksperymentalnych (Czubak, 2015; Lewandowska, 2015) oraz pracy przeglądowej (Klonowska, 2016b), będących wynikiem projektów również związanych z tematyką genetyki nowotworów, które jednak nie stanowiły głównego wątku moich badań i nie wchodzą w skład mojej pracy doktorskiej.

Prace eksperymentalne powstały w wyniku projektów, których przedmiotem była analiza somatycznej zmienności genomu raka płuca. Wynikiem pierwszego z projektów było opracowanie opartej na technikach Real-Time PCR i MLPA dwupoziomowej metody do badania zmienności somatycznej genu *EGFR* (Lewandowska, 2015). Rezultatem drugiego projektu było wykazanie bardzo dużej częstości zmian liczby kopii genów miRNA (m.in. *miR-30d*, *miR-21*, *miR-17* oraz *miR-155*), a także częstych amplifikacji genów biorących udział w biogenezie miRNA – *DROSHA* i *DICER1*. Badania wykazały, że zmienność liczby kopii genów *miR-200b*, *miR-30d* oraz *DROSHA* ma związek z przeżywalnością pacjentów z rakiem płuca (Czubak, 2015). W związku z tym, że w swojej pracy badawczej często wykorzystuję analizę danych generowanych w dużych projektach onkogenomicznych [np. The Cancer Genome Atlas (TCGA)], wraz z innymi współautorami przygotowałam również pracę przeglądową, która pełni rolę przewodnika po narzędziach bioinformatycznych (portalach) umożliwiających analizę całogenomowych danych nowotworowych oraz integrację tych danych z danymi epidemiologicznymi i klinicznymi (Klonowska, 2016b).

BIBLIOGRAFIA

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

Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, *et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet, 2003, 72(5):1117-1130.

Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, Tomiak E, Neuhausen SL, Teo ZL, Khan S, Aittomaki K, Moilanen JS, *et al.* Breast-cancer risk in families with mutations in PALB2. N Engl J Med, 2014, 371(6):497-506.

Bentley RW, Pearson J, Gearry RB, Barclay ML, McKinney C, Merriman TR, Roberts RL. Association of higher DEFB4 genomic copy number with Crohn's disease. Am J Gastroenterol, 2010, 105(2):354-359.

Brzovic PS, Meza JE, King MC, Klevit RE. BRCA1 RING domain cancer-predisposing mutations. Structural consequences and effects on protein-protein interactions. J Biol Chem, 2001, 276(44):41399-41406.

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, McDougle RM, Brown WL, Law EK, *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. Nature, 2013, 494(7437):366-370.

Buys SS, Sandbach JF, Gammon A, Patel G, Kidd J, Brown KL, Sharma L, Saam J, Lancaster J, Daly MB. A study of over 35,000 women with breast cancer tested with a 25-gene panel of hereditary cancer genes. Cancer, 2017, 123(10):1721-1730.

Carpenter D, Farnert A, Rooth I, Armour JA, Shaw MA. CCL3L1 copy number and susceptibility to malaria. Infect Genet Evol, 2012, 12(5):1147-1154.

Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol, 2007, 25(11):1329-1333.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, *et al.* Origins and functional impact of copy number variation in the human genome. Nature, 2010, 464(7289):704-712.

Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, Olson JE, Godwin AK, Pankratz VS, Olswold C, Slettedahl S, Hallberg E, Guidugli L, Davila JI, Beckmann MW, Janni W, *et al.* Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol, 2015, 33(4):304-311.

Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R, Thomas A, Lilyquist J, Feng B, McFarland R, Pesaran T, Huether R, LaDuca H, Chao EC, *et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol, 2017, [Epub ahead of print].

Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, Szwiec M, Byrski T, Debniak T, Gorski B, Sopik V, Akbari MR, Sun P, Gronwald J, Narod SA, Lubinski J, *et al.* Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. Lancet Oncol, 2015a, 16(6):638-644.

Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, Huzarski T, Gronwald J, Byrski T, Szwiec M, Jakubowska A, Rudnicka H, *et al.* Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet, 2015b, 47(6):643-646.

Cybulski C, Lubinski J, Wokolorczyk D, Kuzniak W, Kashyap A, Sopik V, Huzarski T, Gronwald J, Byrski T, Szwiec M, Jakubowska A, Gorski B, Debniak T, Narod SA, Akbari MR. Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland. Clin Genet, 2015c, 88(4):366-370.

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(27):23399-23416.

De Brakeleer S, De Greve J, Loris R, Janin N, Lissens W, Sermijn E, Teugels E. Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. Hum Mutat, 2010, 31(3):E1175-1185.

Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DG, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, *et al.* Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med, 2015, 372(23):2243-2257.

Eccles SA, Aboagye EO, Ali S, Anderson AS, Armes J, Berditchevski F, Blaydes JP, Brennan K, Brown NJ, Bryant HE, Bundred NJ, Burchell JM, Campbell AM, Carroll JS, Clarke RB, Coles CE, *et al.* Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Res, 2013, 15(5):R92.

Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet, 1994, 343(8899):692-695.

Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med, 2008, 359(20):2143-2153.

Frank B, Bermejo JL, Hemminki K, Sutter C, Wappenschmidt B, Meindl A, Kiechle-Bahat M, Bugert P, Schmutzler RK, Bartram CR, Burwinkel B. Copy number variant in the candidate tumor suppressor gene MTUS1 and familial breast cancer risk. Carcinogenesis, 2007, 28(7):1442-1445.

Gohler S, Da Silva Filho MI, Johansson R, Enquist-Olsson K, Henriksson R, Hemminki K, Lenner P, Forsti A. Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population. J Cancer Res Clin Oncol, 2016, 142(1):273-276.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell R J, *et al.* The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science, 2005, 307(5714):1434-1440.

Grabrick DM, Hartmann LC, Cerhan JR, Vierkant RA, Therneau TM, Vachon CM, Olson JE, Couch FJ, Anderson KE, Pankratz VS, Sellers TA. Risk of breast cancer with oral contraceptive use in women with a family history of breast cancer. JAMA, 2000, 284(14):1791-1798.

Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology - The lifelong interplay of genes, lifestyle, and hormones. Breast Cancer Res, 2004, 6(5):213-218.

Hashizume R, Fukuda M, Maeda I, Nishikawa H, Oyake D, Yabuki Y, Ogata H, Ohta T. The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. J Biol Chem, 2001, 276(18):14537-14540.

Hogervorst FB, Nederlof PM, Gille JJ, McElgunn CJ, Grippeling M, Pruntel R, Regnerus R, van Welsem T, van Spaendonk R, Menko FH, Kluijt I, Dommering C, Verhoef S, Schouten JP, van't Veer LJ, Pals G. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. Cancer Res, 2003, 63(7):1449-1453.

Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, Schalkwijk J. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet, 2008, 40(1):23-25.

Irminger-Finger I, Leung WC, Li J, Dubois-Dauphin M, Harb J, Feki A, Jefford CE, Soriano JV, Jaconi M, Montesano R, Krause KH. Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis. Mol Cell, 2001, 8(6):1255-1266.

Karppinen SM, Heikkinen K, Rapakko K, Winqvist R. Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. J Med Genet, 2004, 41(9):e114.

Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet, 2007, 3(4):e63.

King MC, Marks JH, Mandell JB, New York Breast Cancer Study G. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science, 2003, 302(5645):643-646.

Kitamura S, Ode H, Iwatani Y. Structural Features of Antiviral APOBEC3 Proteins are Linked to Their Functional Activities. Front Microbiol, 2011, 2:258.

Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P. Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene. BioTechnologia, 2014, 95(3):203-214.

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, Irminger-Finger I, Limon J, Kozlowski P. 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.

Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P. MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency. Oncotarget, 2016a, 7(41):67266-67276.

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, 2016b, 7(1):176-192.

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P. The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population. Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet, 2008, 40(10):1253-1260.

Kotsopoulos J, Lubinski J, Moller P, Lynch HT, Singer CF, Eng C, Neuhausen SL, Karlan B, Kim-Sing C, Huzarski T, Gronwald J, McCuaig J, Senter L, Tung N, Ghadirian P, Eisen A, *et al.* Timing of oral contraceptive use and the risk of breast cancer in BRCA1 mutation carriers. Breast Cancer Res Treat, 2014, 143(3):579-586.

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(3-4):389-400.

Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA, Krumbiegel M, Uebe S, Reuter M, Beckmann MW, Reis A. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. Int J Cancer, 2017, 140(1):95-102.

Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, Santos TM, Goncalves A, Nobrega AF, Pearson PL, Vianna-Morgante AM, Carraro DM, Brentani RR, Rosenberg C. Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res, 2012, 14(1):R24.

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(2):e0117983.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med, 2000, 343(2):78-85.

Long J, Delahanty RJ, Li G, Gao YT, Lu W, Cai Q, Xiang YB, Li C, Ji BT, Zheng Y, Ali S, Shu XO, Zheng W. A common deletion in the APOBEC3 genes and breast cancer risk. J Natl Cancer Inst, 2013, 105(8):573-579.

Lux MP, Fasching PA, Beckmann MW. Hereditary breast and ovarian cancer: review and future perspectives. J Mol Med (Berl), 2006, 84(1):16-28.

Malhotra GK, Zhao X, Band H, Band V. Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther, 2010, 10(10):955-960.

Marcinkowska-Swojak M, Klonowska K, Figlerowicz M, Kozlowski P. An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs. Gene, 2014, 546(2):257-262.

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.

Masson AL, Talseth-Palmer BA, Evans TJ, Grice DM, Hannan GN, Scott RJ. Expanding the genetic basis of copy number variation in familial breast cancer. Hered Cancer Clin Pract, 2014, 12(1):15.

McLaughlin JR, Risch HA, Lubinski J, Moller P, Ghadirian P, Lynch H, Karlan B, Fishman D, Rosen B, Neuhausen SL, Offit K, Kauff N, Domchek S, Tung N, Friedman E, Foulkes W, *et al.* Reproductive risk factors for ovarian cancer in carriers of BRCA1 or BRCA2 mutations: a case-control study. Lancet Oncol, 2007, 8(1):26-34.

Melchor L, Benitez J. The complex genetic landscape of familial breast cancer. Hum Genet, 2013, 132(8):845-863.

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast, *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet, 2013, 45(4):353-361, 361e1-2.

Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, Darabi H, Brand JS, Bojesen SE, *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet, 2015, 47(4):373-380.

Montagna M, Dalla Palma M, Menin C, Agata S, De Nicolo A, Chieco-Bianchi L, D'Andrea E. Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. Hum Mol Genet, 2003, 12(9):1055-1061.

Narod SA, Risch H, Moslehi R, Dorum A, Neuhausen S, Olsson H, Provencher D, Radice P, Evans G, Bishop S, Brunet JS, Ponder BA. Oral contraceptives and the risk of hereditary ovarian cancer. Hereditary Ovarian Cancer Clinical Study Group. N Engl J Med, 1998, 339(7):424-428.

Narod SA, Dube MP, Klijn J, Lubinski J, Lynch HT, Ghadirian P, Provencher D, Heimdal K, Moller P, Robson M, Offit K, Isaacs C, Weber B, Friedman E, Gershoni-Baruch R, Rennert G, *et al.* Oral contraceptives and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst, 2002, 94(23):1773-1779.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, *et al.* Mutational processes molding the genomes of 21 breast cancers. Cell, 2012, 149(5):979-993.

Norquist BM, Harrell MI, Brady MF, Walsh T, Lee MK, Gulsuner S, Bernards SS, Casadei S, Yi Q, Burger RA, Chan JK, Davidson SA, Mannel RS, DiSilvestro PA, Lankes HA, Ramirez NC, *et al.* Inherited Mutations in Women With Ovarian Carcinoma. JAMA Oncol, 2016, 2(4):482-490.

Pennington KP, Walsh T, Harrell MI, Lee MK, Pennil CC, Rendi MH, Thornton A, Norquist BM, Casadei S, Nord AS, Agnew KJ, Pritchard CC, Scroggins S, Garcia RL, King MC, Swisher EM. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. Clin Cancer Res, 2014, 20(3):764-775.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, *et al.* Molecular portraits of human breast tumours. Nature, 2000, 406(6797):747-752.

Pilyugin M, Andre PA, Ratajska M, Kuzniacka A, Limon J, Tournier BB, Colas J, Laurent G, Irminger-Finger I. Antagonizing functions of BARD1 and its alternatively spliced variant BARD1delta in telomere stability. Oncotarget, 2017, 8(6):9339-9353.

Puget N, Gad S, Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. Distinct BRCA1 rearrangements involving the BRCA1 pseudogene suggest the existence of a recombination hot spot. Am J Hum Genet, 2002, 70(4):858-865.

Ratajska M, Antoszewska E, Piskorz A, Brozek I, Borg A, Kusmierek H, Biernat W, Limon J. Cancer predisposing BARD1 mutations in breast-ovarian cancer families. Breast Cancer Res Treat, 2012, 131(1):89-97.

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I. Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms. Oncol Rep, 2015, 34(5):2609-2617.

Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell, 2012, 46(4):424-435.

Ruffner H, Joazeiro CA, Hemmati D, Hunter T, Verma IM. Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. Proc Natl Acad Sci U S A, 2001, 98(9):5134-5139.

Sabatier R, Adelaide J, Finetti P, Ferrari A, Huiart L, Sobol H, Chaffanet M, Birnbaum D, Bertucci F. BARD1 homozygous deletion, a possible alternative to BRCA1 mutation in basal breast cancer. Genes Chromosomes Cancer, 2010, 49(12):1143-1151.

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(12):e57.

Slavin TP, Maxwell KN, Lilyquist J, Vijai J, Neuhausen SL, Hart SN, Ravichandran V, Thomas T, Maria A, Villano D, Schrader KA, Moore R, Hu C, Wubbenhorst B, Wenz BM, D'Andrea K, *et al.* The contribution of pathogenic variants in breast cancer susceptibility genes to familial breast cancer risk. NPJ Breast Cancer, 2017, 3:22.

Suehiro Y, Okada T, Shikamoto N, Zhan Y, Sakai K, Okayama N, Nishioka M, Furuya T, Oga A, Kawauchi S, Maeda N, Tamesa M, Nagashima Y, Yamamoto S, Oka M, Hinoda Y, *et al.* Germline copy number variations associated with breast cancer susceptibility in a Japanese population. Tumour Biol, 2013, 34(2):947-952.

Sun J, Wang Y, Xia Y, Xu Y, Ouyang T, Li J, Wang T, Fan Z, Fan T, Lin B, Lou H, Xie Y. Mutations in RECQL Gene Are Associated with Predisposition to Breast Cancer. PLoS Genet, 2015, 11(5):e1005228.

Thompson ER, Rowley SM, Li N, McInerny S, Devereux L, Wong-Brown MW, Trainer AH, Mitchell G, Scott RJ, James PA, Campbell IG. Panel Testing for Familial Breast Cancer: Calibrating the Tension Between Research and Clinical Care. J Clin Oncol, 2016, 34(13):1455-1459.

Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K, Timms K, Garber JE, Herold C, Ellisen L, Krejdovsky J, DeLeonardis K, Sedgwick K, Soltis K, Roa B, Wenstrup RJ, *et al.* Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer, 2015, 121(1):25-33.

Walker LC, Pearson JF, Wiggins GA, Giles GG, Hopper JL, Southey MC. Increased genomic burden of germline copy number variants is associated with early onset breast cancer: Australian breast cancer family registry. Breast Cancer Res, 2017, 19(1):30.

Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell, 2007, 11(2):103-105.

White SJ, den Dunnen JT. Copy number variation in the genome; the human DMD gene as an example. Cytogenet Genome Res, 2006, 115(3-4):240-246.

Wimmer K, Yao S, Claes K, Kehrer-Sawatzki H, Tinschert S, De Raedt T, Legius E, Callens T, Beiglbock H, Maertens O, Messiaen L. Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. Genes Chromosomes Cancer, 2006, 45(3):265-276.

Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, Yang MC, Hwang LY, Bowcock AM, Baer R. Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet, 1996, 14(4):430-440.

Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, Kelley MC, Zheng W, Long J. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. Carcinogenesis, 2013, 34(10):2240-2243.

Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS, Xu XH, Yan H, Liu X, Qiu C, Zhu XZ, Chen T, *et al.* Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am J Hum Genet, 2008, 83(6):663-674.

PUBLIKACJE WCHODZĄCE W SKŁAD ROZPRAWY DOKTORSKIEJ

1

Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P

Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene

BioTechnologia, 2014, 95(3):203-214

BioTechnologia vol. 95(3) • pp. 203-214 • 2014
Journal of Biotechnology, Computational Biology and Bionanotechnology REVIEW PAPER

Genetic predisposition to breast and/or ovarian cancer – focus on the candidate *BARD1* gene

KATARZYNA KLONOWSKA [1], MAGDALENA RATAJSKA [2], MARZENA WOJCIECHOWSKA [1], PIOTR KOZLOWSKI [1]*

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

[2] Medical University of Gdansk, Gdańsk, Poland

* Corresponding author: kozlowp@yahoo.com

Abstract

Germline mutations affecting the *BRCA1* and *BRCA2* genes explain 16-40% of breast and/or ovarian cancers aggregated in families. Besides the *BRCA1/2* genes and several genetic factors associated with hereditary syndromes which increase the risk of breast cancer, a considerable fraction of potential breast cancer predisposing factors (~50%) remains unknown. It is presumed that candidate genes, functionally related to the *BRCA1/2* genes, may account for some of the missing heritability. The *BARD1* gene, which encodes a protein indispensable for BRCA1-mediated tumor suppression function and adequate apoptosis regulation, serves as a candidate breast cancer susceptibility gene. Some initial reports indicated that *BARD1* is a plausible target for several pathogenic mutations associated with increased breast and/or ovarian cancer risk. Nonetheless, further mutational studies are necessary to determine the penetrance and role of the *BARD1* gene in cancer predisposition.

Key words: *BARD1*, breast and/or ovarian cancer, familial cancer predisposition, DSB repair genes

Introduction

Breast cancer is the most frequent cancer and the primary cause of malignancy-associated deaths among women worldwide. Ovarian cancer, which frequently co-occurs with breast cancer in familial setting, is the fifth most common cancer (constitutes ~5% of all registered cancer cases) and the fourth leading cause of death attributed to cancer (constitutes cause of ~6% of cancer-related deaths) among Polish women. The highest breast and ovarian cancers rates are observed in western, developed countries (http://onkologia.org.pl/) (Youlden et al., 2012).

There is a number of risk factors associated with breast cancer. Age (> 40), early menarche (age of < 12), and late menopause (age of > 54) substantially increase the risk of breast cancer. Among other factors increasing the risk of breast cancer are: longstanding hormonal replacement therapy, high body mass index (BMI), and regular alcohol intake. Breast feeding and early age at first pregnancy are counted as protective factors, whereas hormonal contraceptives decrease the ovarian cancer risk and slightly increase breast cancer risk. Familial aggregation of breast cancer cases, especially among first-grade relatives, are important factors of breast cancer risk, indicating that genetic factors are essential determinants of breast and/or ovarian cancer risk (Hankinson et al., 2004; Lux et al., 2006).

Unified paradigms for breast and ovarian cancer etiology are difficult to elucidate. The longstanding exposure to hormones, as well as the interplay of environmental and genetic factors, modulate the probability of developing these complex diseases (Hankinson et al., 2004; Permuth-Wey and Sellers, 2009). Overall heritability of breast and ovarian cancer was estimated based on twins studies (monozygotic and dizygotic) for approximately 30% and 20%, respectively (Lichtenstein et al., 2000). Familial breast cancer constitutes 5-10% of all breast cancer cases. In the middle 1990's, studies conducted in families with strong aggregation of breast and/or ovarian cases led to the identification of *BRCA1* (**br**east **ca**ncer **1**) and *BRCA2* (**br**east **ca**ncer **2**) genes (Miki et al., 1994; Wooster et al., 1995). Germline mutations in these genes account for 16-40% of familial breast cancers (Beggs and Hodgson, 2009; Ripperger et al., 2009). Additionally, about 5% of breast cancer cases aggregated in family is attributed to mutations in genes associated with various hereditary syndromes and genes conferring moderate risk. Finally, it was reported that common

SNPs cumulatively explain 14% of the familial breast cancer cases (Michailidou et al., 2013). Other genetic breast cancer susceptibility factors (~50%) are unknown. There is a supposition that a fraction of inherited breast and/or ovarian cases can be attributed to heterozygous mutations in candidate genetic factors [e.g. *BARD1* (**B**RCA1 **a**ssociated **R**ING **d**omain **1**)], which contribution to breast and/or ovarian cancer remains to be evaluated – Figure 1 (Wooster and Weber, 2003).

Therefore, further investigation of hereditary genetic alterations which predispose to breast and/or ovarian cancer could reveal a fraction of missing heritability of breast and/or ovarian cancer and consequently may uncover new models for inherited susceptibility evaluation and contribute to the development of targeted preventive strategies (Manolio et al., 2009).

Hereditary cancer syndromes

The majority of cancers are sporadic. Sporadic cancer is a complex and multifactorial disease that is acquired owing to environmental exposures, lifestyle or multiple genetic factors (variants) of very low risk effects. A fraction of some cancers (especially breast, ovarian, and colorectal cancers) occurs in the form of familial aggregations, i.e. is observed in closely related individuals more frequently than it could be expected based on the frequency of the cancer in general population. It is estimated that familial cancers constitute up to 15% of particular cancers. Predominantly, familial aggregation of cancer cases is attributed to a single loss-of-function mutation in a specific tumor suppressor gene associated with a particular cancer type. Among genes, which mutations underlie the most common familial cancer types are 1) DNA mismatch repair genes [e.g. *MSH2* (**M**ut**S** **h**omolog **2**), *MSH6* (**M**ut**S** **h**omolog **6**), *MLH1* (**M**ut**L** **h**omolog **1**)] associated with hereditary non-polyposis colorectal cancer (HNPCC), 2) the *APC* (**a**denomatous **p**olyposis **c**oli) gene predisposing to familial adenomatous polyposis (FAP), and 3) *BRCA1* and *BRCA2* associated with breast and/or ovarian cancer aggregation. The identification of these genes was essential for the understanding of pathomechanism of familial cancer syndromes and laid the foundation for familial cancer genetic diagnostics. Most of hereditary cancers constitute autosomal dominant disorders that display incomplete penetrance (Nagy et al., 2004). Inherited cancers are initiated by the transmission of a genetic mutation in the germline. However, it must be noted that the risk of the development of inherited cancer may also be modulated by lifestyle and environmental exposures as well as other genetic factors. Multiple cases of genetically associated hereditary cancers are often aggregated within the family and can be related to a particular inherited cancer syndrome (see Table 1). The probability of the inheritance of cancer predisposition within a family increases with the number of individuals affected by cancer (Ellis, 2011; Heald and Church, 2011). Inherited cancer susceptibility can be also associated with the presence of multiple primary cancers or simultaneous occurrence of nonmalignant disorders in affected individual (Nagy et al., 2004). The occurrence of several generations with numerous cases of the early-onset, bilateral or multisynchronous cancers within a family can be counted as the hallmarks of hereditary cancer syndromes (Heald and Church, 2011).

Hereditary breast and ovarian cancer

Genetic variants predisposing to breast cancer can be divided into three major groups according to the breast and/or ovarian cancer risk conferred by these variants and their frequency in the population (Fig. 2) (Foulkes, 2008; Ripperger et al., 2009). Importantly, a substantial fraction of existing breast and/or ovarian cancer susceptibility genes with various degree of penetrance still remains unidentified (Wooster and Weber, 2003; Karppinen et al., 2004; Beggs and Hodgson, 2009).

The first group of breast and/or ovarian cancer susceptibility variants encompasses rare, high risk heterozygous mutations occurring in genes associated with several rare hereditary syndromes (Foulkes, 2008; Beggs and Hodgson, 2009). The major genes associated with susceptibility to hereditary breast and ovarian cancer syndrome (HBOC) are *BRCA1* and *BRCA2*. Germline mutations in these genes are associated with the risk of 50-80% and 30-50% for breast and ovarian cancers, respectively. Germline mutations in *BRCA1* and *BRCA2* genes explain approximately 16-40% of breast and/or ovarian cancer cases aggregated in families (Beggs and Hodgson, 2009; Ripperger et al., 2009; Roy et al., 2012). Other cancer syndromes listed below explain less than 5% of familial breast cancer aggregation. The probability

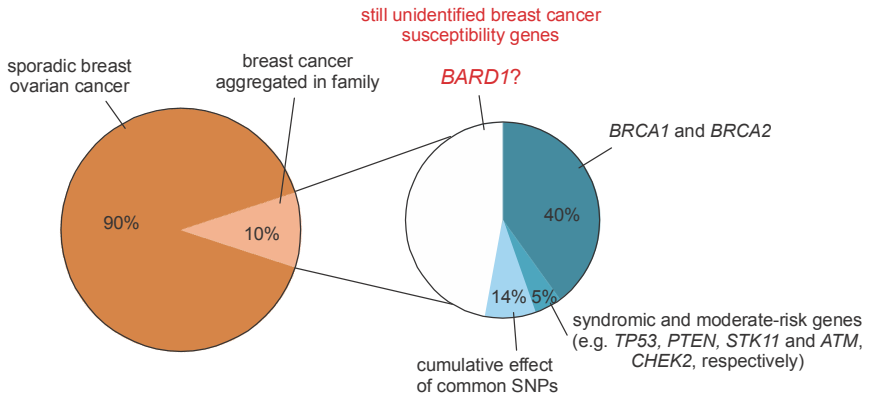


Fig. 1. Pie-chart schematically depicting the genes accounting for familial aggregations of breast cancer

Table 1. Hereditary cancer syndromes

Syndrome	Gene	Chromosomal localization	Mode of inheritance	Associated cancer	References
Hereditary breast and ovarian cancer (HBOC)	*BRCA1* *BRCA2*	17q21 13q12.3	autosomal dominant	predominantly breast and ovarian	(Miki et al., 1994; Wooster et al., 1995; Nagy et al., 2004)
Hereditary site-specific breast cancer	*BRCA1* *BRCA2*	17q21 13q12.3	autosomal dominant	predominantly breast	
Hereditary site – specific ovarian cancer	*BRCA1* *BRCA2*	17q21 13q12.3	autosomal dominant	predominantly ovarian also: prostate, fallopian tube, stomach, pancreatic, laryngeal	
Li-Fraumeni syndrome	*TP53*	17p13.1	autosomal dominant	breast, brain, sarcomas, leukemias	(Nagy et al., 2004)
Hereditary nonpolyposis colorectal cancer (HNPCC)/ Lynch syndrome	*MSH2* *MSH6* *MLH1* *PMS1* *PMS2*	2p21 2p16 3p21.3 2q31.1 7p22.2	autosomal dominant	endometrial colorectal, stomach, ovarian, pancreas, brain	(Nagy et al., 2004)
Familial adenomatous polyposis (FAP)	*APC*	5q21-q22	autosomal dominant	colorectal, colon, gastric, pancreatic, adenomas	(Nagy et al., 2004)
Cowden syndrome	*PTEN*	10q23.3	autosomal dominant	breast, thyroid, endometrial	(Nagy et al., 2004)
Ataxia-telangiectasia	*ATM*	11q22-q23	autosomal recessive	lymphomas, leukemias, breast	(Savitsky et al., 1995; Khanna, 2000)
Hereditary diffuse gastric cancer syndrome	*CDH1*	16q22.1	autosomal dominant	gastric, breast	(Berx et al., 1995; Pinheiro et al., 2010)
Fanconi anemia	*BRIP1* *PALB2* *BRCA2*	17q22.2 16p12.2 13q12.3	autosomal recessive	breast, leukemia	(Mathew, 2006; Walsh and King, 2007)
Peutz-Jeghers syndrome	*STK11 (LKB1)*	19p13.3	autosomal dominant	colon, small intestine, stomach, breast, pancreatic	(Hemminki et al., 1998; Nagy et al., 2004)
Nijmegen-breakage syndrome	*NBN*	8q21	autosomal recessive	lymphoma, breast, colorectal	(Matsuura et al., 1998; Steffen et al., 2004)

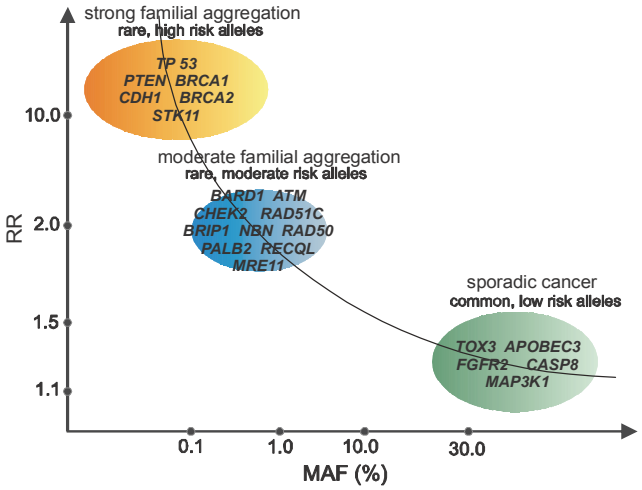


Fig. 2. The relation between the relative risk (RR) and the frequency (minor allele frequency, MAF) of genetic variants in breast and/or ovarian susceptibility genes

of the existence of genes with a population frequency and conferred risk comparable to *BRCA1* and *BRCA2* [e.g. predicted *BRCA3* (**br**east **ca**ncer **3**) gene] is very low (Narod and Foulkes, 2004; Foulkes, 2008). Li-Fraumeni syndrome is attributed to mutations in the *TP53* (**t**umor **p**rotein p**53**) gene encoding a protein "genome guardian" involved mainly in the control of cell cycle progression, repair of DNA damage, and apoptosis stimulation. Although less than 1% of early onset breast cancer cases clustered in families harbor germline mutations in *TP53*, women affected by Li-Fraumeni syndrome are at 49% risk of developing breast cancer by the age of 60 (Garber et al., 1991; Masciari et al., 2012). For women with germline mutations in the *CDH1* (**cadh**erin **1**) gene, associated with the hereditary diffuse gastric syndrome, or in the *STK11* (**s**erine/**t**hreonine **k**inase **11**) gene, associated with Peutz-Jeghers syndrome, the risk of developing breast cancer is approximately 30-40% (Hemminki et al., 1998; Pharoah et al., 2001; Lim et al., 2004). Recently, Tan and coworkers (Tan et al., 2012) have shown that pathogenic germline mutations in the promoter of the *PTEN* (**p**hosphatase and **ten**sin homolog) gene associated with Cowden syndrome can increase the lifetime risk of breast cancer up to 85%. Additionally, hereditary ovarian cancers can be attributed to variants in mismatch repair genes [*MLH1* (**mut**L **h**omolog **1**), *MSH2* (**mut**S **h**omolog **2**), *MSH6* (**mut**S **h**omolog **6**), and *PMS2* (**p**ost**me**iotic **s**egregation increased **2**)] associated with Lynch syndrome (Lynch et al., 2009).

Mutational analyses of candidate genes encoding proteins co-working with BRCA1 and BRCA2 in the same molecular pathways led to the identification of a second group of alterations that confer susceptibility to breast and/or ovarian cancer. This group comprises uncommon variants associated with moderate risk of breast and/or ovarian cancer (Beggs and Hodgson, 2009). It was reported that mutations in *BRIP1* (**BR**CA1 **i**nteracting **p**rotein C-terminal helicase **1**), *BARD1, RAD50, CHEK2* (**che**ckpoint **k**inase **2**), *NBN* (**nibrin**), *PALB2* (**p**artner **a**nd **l**ocalizer of **B**RCA**2**), and *ATM* (**a**taxia **t**elangiectasia **m**utated) are of intermediate penetrance and are associated with 2-4 fold increased risk of breast cancer. It is worth noting that very rare, bi-allelic mutations in breast cancer susceptibility genes *BRCA2, PALB2*, and *BRIP1* are associated with Fanconi's anemia. This suggests that some genes controlling DNA repair through homologous recombination and associated with this mostly recessive disorder may also contribute to the initiation of breast and/or ovarian cancer (Walsh and King, 2007; Foulkes, 2008; Beggs and Hodgson, 2009; van der Groep et al., 2011). Walsh and colleagues observed that mutations in Fanconi's anemia genes are involved in the development of hereditary ovarian cancer. The authors detected pathogenic heterozygous mutations in genes implicated in Fanconi's anemia pathway, [e.g. *BARD1, RAD50, NBN, PALB2, MRE11A* (**m**eiotic **re**combination **11** homolog **A**), *BRIP1,* and DNA repair protein gene *RAD51C* (**RAD51** homolog **C**)] in a group of patients with ovarian carcinoma not selected in terms of familial history of the disease (Walsh et al., 2011). Recently, Cybulski and coworkers have identified *RECQL* (**RecQ** helicase-**l**ike) as a new breast cancer susceptibility gene of moderate penetrance, with the use of the combination of a whole exome sequencing and a large-scale association study of recurrent mutations (Cybulski et al., 2015). The *RECQL,* similarly as other breast cancer susceptibility genes, is involved in DNA repair by resolving stalled DNA replication forks and thus preventing double-stranded DNA breaks.

The third group of breast and/or ovarian cancer predisposing variants comprises a common, low-penetrance polymorphisms, identified mainly in Genome Wide Association Studies (GWAS). Recently, 67 new and previously reported single nucleotide polymorphisms (SNPs) have been identified to be associated with a slightly increased breast cancer risk (odds ratio (OR) ~1,2) (Michailidou

et al., 2013). It was also presumed that the investigation of a copy number variation may uncover a substantial part of the still unidentified genetic *loci* related to the susceptibility to various complex diseases, including breast cancer. Until now, very few studies have assessed the association of CNV with breast cancer risk. Recently, a common large deletion in the *APOBEC3* (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3) gene cluster was correlated with an increased breast cancer risk ([OR] = 1.31 95% CI = 1.21 to 1.42 for one copy deletion) (Long et al., 2013).

In consonance with the polygenic model, the aggregation of breast and/or ovarian cases within a family, not attributed to mutations in *BRCA1* and *BRCA2*, can be caused by a combined effect of multiple genetic alterations in genes of low to moderate penetrance, presumably modified by environmental factors (Karppinen et al., 2004).

BRCA1 and BRCA2 – guardians of the genome integrity

The proteins encoded by *BRCA1* and *BRCA2* tumor suppressor genes work in concert to maintain the genome integrity through the interaction with a number of proteins, such as CHEK2, ATM, BARD1, NBN, RAD51, ATR (**a**taxia **t**elangiectasia and **R**AD3-related), p53, BRIP1, and PALB2. These genes act as guardians of the genome integrity and are involved in the pathways of DNA damage response (DDR), the regulation of transcription, cell cycle checkpoints, apoptosis, and ubiquitination (Narod and Foulkes, 2004; Roy et al., 2012).

Double strand breaks (DSB) constitute a threatening form of DNA damage, as unrepaired double strand lesions often lead to severe genomic rearrangements that contribute to cancer initiation. The main function of BRCA1 is to integrate the DSB repair mechanisms and checkpoint regulation that delay the cell cycle in order to provide time for DNA repair and to ensure that the genetic damage is not transmitted to the next generation whereas BRCA2 is responsible for the core mechanism of RAD51-mediated homologous recombination which was developed by mammals as one of the DDR systems (Hoeijmakers, 2001; Roy et al., 2012).

Mutations in *BRCA1* and *BRCA2*

According to the Breast Cancer Information Core (BIC) database (http://research.nhgri.nih.gov/bic/), approximately 3800 various genomic alterations in *BRCA1* and *BRCA2* genes have been detected so far (data of July, 2015). Missense and nonsense mutations, as well as small frameshift insertions/deletions and mutations affecting splice sites within introns (IVS), account for the most frequent alterations which occur in *BRCA1* and *BRCA2* genes (http:research.nhgri.nih.gov/bic/) (Thompson and Easton, 2004). Large genomic rearrangements in *BRCA1* and *BRCA2* have also been identified. It was reported that large mutations may account from 0% up to 36% of all mutations affecting *BRCA1* gene, across various populations. In *BRCA1* gene more than 80 various large mutations have been found, whereas in *BRCA2* gene much fewer large rearrangements have been reported (Hansen et al., 2009; Sluiter and van Rensburg, 2011).

It was estimated that mutations in *BRCA1* and *BRCA2* genes occur in about 1/400 individuals (Foulkes, 2008). In the majority of populations various mutations are located along the entire sequence of *BRCA1* and *BRCA2*. However, in some ethnic groups, owing to a founder effect, particular mutations in *BRCA1* and *BRCA2* occur with higher frequency. A founder effect can be defined as a loss of genetic variation which occurs due to interbreeding within a small group of individuals isolated from a larger group. As a consequence, relatively uncommon mutations become more frequent within such ethnic group (Ferla et al., 2007). For example in the Ashkenazi Jewish population, 1 in 40 individuals (Foulkes, 2008) is a carrier of 185delAG (c.68_69delAG) (Struewing et al., 1995) or 5382insC (c.5266dupC) (Roa et al., 1996) founder mutation in *BRCA1* or 6147delT (c.5946delT) (Neuhausen et al., 1996) founder mutation in *BRCA2*. In Iceland, a high frequency of founder 999del5 (c.771_775del5) *BRCA2* mutation was identified. This founder mutation was reported to cause the familial clustering of both female and male breast cancer cases. It was shown that 999del5 *BRCA2* mutation affects 40% of males with breast cancer from the Icelandic population (Thorlacius et al., 1996). Finally, in Poland, a high incidence of 5382insC, 300T > G (c.181T > G), and 4153delA (c.4034delA) *BRCA1* founder mutations has been identified (Sobczak et al., 1997; Gorski et al., 2000; Grzybowska et al., 2000; Gorski et al., 2004; Ratajska et al., 2008; Brozek et al., 2011). 3819del5 (c.3700_3704del5) and 185delAG mutations affecting the *BRCA1* gene were also reported to occur frequently in the Polish population (Gorski et al., 2000; Ratajska et al., 2008; Brozek et al., 2011).

BARD1 as a breast cancer susceptibility gene

Structure and functions of the *BARD1* gene and encoded protein

BARD1 gene is located at 2q34-35 and consists of 11 exons encoding protein of 777 amino acids. BARD1 was identified by yeast two-hybrid screening as a protein that associates with BRCA1 protein *in vivo*. BARD1 protein bears a striking structural resemblance to BRCA1 protein. Both proteins harbor a RING-finger motif and a nuclear export signal in the vicinity of their N-termini and two BRCA1 carboxy-terminal (BRCT) domains. BARD1 and BRCA1 proteins form a functional heterodimer through the binding of their RING-finger motifs. Apart from BRCT and RING domains, BARD1 contains three tandem ankyrin repeats (ANK) located in the central part of the protein. This structural motif is implicated in other protein-protein interaction (Wu et al., 1996). Neither BARD1 nor BRCA1 displays structural resemblance to BRCA2 (Irminger-Finger and Jefford, 2006).

Besides structural similarity, BARD1 and BRCA1 proteins share some common functions. Increased levels of these proteins are observed in spleen and testes, as well as in other proliferative tissues. Additionally, it was shown that the expression of *BRCA1* and *BARD1* in breast and ovaries is regulated hormonally and that the *in vitro* down-regulation of *BARD1* leads to the alteration of mammary epithelial cells phenotype (Irminger-Finger et al., 1998). Both *BARD1* and *BRCA1* deficiency is pathogenic for the cell. McCarthy and colleagues observed that *BARD1*$^{-/-}$ and *BRCA1*$^{-/-}$ as well as double *BARD1*$^{-/-}$*; BRCA1*$^{-/-}$ mice display phenotypic similarities. The deficiency of both *BARD1* and *BRCA1* leads to the deleterious genomic rearrangements and an early embryonic death which is attributed to the defective cell proliferation (McCarthy et al., 2003).

BARD1 forms a heterodimeric complex with BRCA1 through the interaction of domains comprising RING finger motifs (Wu et al., 1996). A Heterodimeric state is preferred by BARD1 and BRCA1, because this interaction is thought to stabilize both proteins (Meza et al., 1999) and is required for the nuclear localization of the complex (Irminger-Finger, 2010). It was shown that BRCA1-BARD1 heterodimeric complex has the E3 ubiquitin ligase activity (Ruffner et al., 2001; Baer and Ludwig, 2002; Morris and Solomon, 2004). Although the individual BRCA1 and BARD1 ubiquitin ligase activity is very low, it is considerably enhanced after the heterodimerization of the proteins (Hashizume et al., 2001). It was reported that mutations associated with breast cancer located in the RING domain of BRCA1 disrupt the ubiquitin ligase activity of BARD1-BRCA1 complex and abolish BRCA1 involvement in the mechanisms responsible for the protection of cell from ϒ-radiation (Hashizume et al., 2001; Ruffner et al., 2001). It is also suggested that BRCA1-BARD1 E3 ubiquitin ligase is implicated in DNA repair and that BARD1 is essential for BRCA1 tumor suppression functions.

A number of BRCA1-BARD1 targets have been identified, including CDC25C (**c**yclin B and cell **d**ivision **c**ycle **25C**) (Shabbeer et al., 2013), ϒ-tubulin (Starita et al., 2004), and H2AX (Chen et al., 2002). BRCA1-BARD1 E3 ligase was reported to ubiquitinate the proteins that orchestrate G2/M cell cycle checkpoint, i.e. cyclin B and CDC25C, what leads to their degradation and loss of control over the cell cycle progression (Shabbeer et al., 2013). Additionally, BARD1-BRCA1 heterodimer can control centrosome duplication, mediating the destruction of ϒ-tubulin (Starita et al., 2004). H2AX can also be ubiquitinated by E3 ligase, what indicates that BRCA1-BARD1 heterodimer can be implicated in chromatin remodeling (Chen et al., 2002).

Ryser and colleagues also observed an interaction of BARD1 and BRCA2 in mitosis. As full length BARD1 associates with BRCA1 at spindle poles in early mitosis, BARD1 β isoform (without RING domain), frequently found in gynecological cancers, interacts with BRCA2 in late mitosis. Accordingly, BARD1 isoforms have different functions in mitosis and may functionally associate with BRCA1 and BRCA2 proteins, which are responsible for the control of early and late phase of mitosis, respectively (Ryser et al., 2009).

Besides BRCA1/2-mediated functions, independent cellular activities of BARD1 were also reported. Irminger-Finger and colleagues proposed a paradigm of the "dual mode of action" for BARD1 activity in the cell. The authors distinguished the survival mode, in which BARD1 associates with BRCA1 and is implicated in the DNA damage response, and the death mode in which the excess of BARD1 over BRCA1, performs pro-apoptotic functions independently of BRCA1. It was observed that the interaction of BARD1 and BRCA1 diminishes the apoptosis induction. The study indicates that the geno-

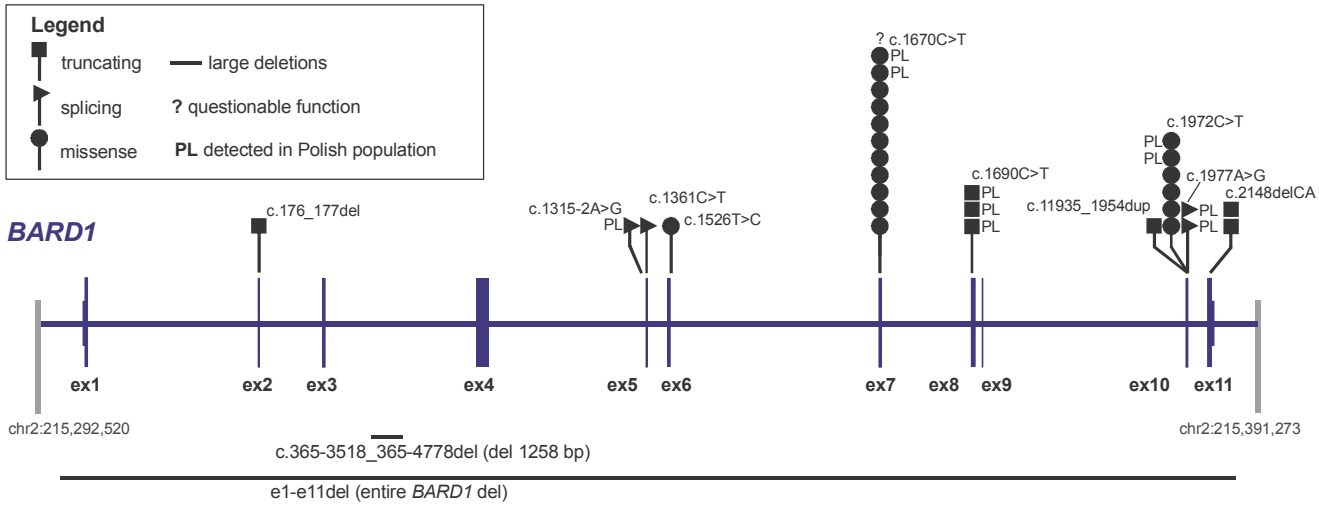


Fig. 3. Map of *BARD1* mutations identified in patients with breast and/or ovarian cancer. It does not show variants identified as neutral or of unknown significance, and common SNPs associated with cancers

toxic stress induces upregulation of *BARD1*. The increased accumulation of BARD1 has an impact on the stabilization of p53 through the association of BARD1 with p53 and a subsequent induction of apoptosis. Accordingly, the repression of BARD1 synthesis leads to an impaired apoptotic response to severe DNA damage (Irminger-Finger et al., 2001).

Mutations in the *BARD1* gene

The mutational analyses of *BARD1* in non-*BRCA* subjects with familial breast and/or ovarian cancer led to the identification of various *BARD1* sequence variants. These variants include deleterious and potentially deleterious mutations leading to premature termination of translation, disruption of protein structure/function, or alternative splicing (Karppinen et al., 2004; De Brakeleer et al., 2010; Sabatier et al., 2010; Ratajska et al., 2012; Castera et al., 2014; Cybulski et al., 2014; Pennington et al., 2014; Churpek et al., 2015; Couch et al., 2015; Klonowska et al., 2015) (summarized in Fig. 3). However, it has to be noted that some results of these studies are not consistent and inconclusive in terms of the functional significance of the mutation.

Some germline mutations, including missense, frameshift and silent alterations (Thai et al., 1998; Ghimenti et al., 2002; Ishitobi et al., 2003; Sauer and Andrulis, 2005; Huo et al., 2007) as well as a large heterozygous deletion (1258 bp) within intron 3 of the *BARD1* gene (Rouleau et al., 2012) were identified, however their impact on the protein structure and function is not elucidated.

The p.Cys557Ser (c.1670G > C) [rs28997576 in dbSNP database; http:ncbi.nlm.nih.gov/SNP] is the most commonly studied mutation in the *BARD1* gene. Sauer and coworkers (Sauer and Andrulis, 2005) showed that a defective growth suppression and impaired apoptotic functions are attributed to an ectopic expression of *BARD1* affected by the p.Cys557Ser mutation, indicating a possible deleterious effect of this variant. p.Cys557Ser was reported to slightly increase the breast cancer risk in Nordic (Finish, Icelandic, Danish, Swedish, and Norwegian) families with breast cancer aggregation (*BRCA1/2* positive – OR = 3.2; $p = 0.01$; 95% CI = 1.2-8.3; *BRCA1/2* negative – OR = 2.6; $p < 0.001$; 95% CI = 1.7-4.0) (Karppinen et al., 2004; Karppinen et al., 2006), however this finding was not confirmed by Vahteristo and colleagues (Vahteristo et al., 2006) who showed no association of the p.Cys557Ser with familial breast cancer susceptibility in Finland. Neither has the p.Cys557Ser mutation been associated with an increased risk of breast cancer in Australian and Polish populations (Gorringe et al., 2008; Jakubowska et al., 2008; Johnatty et al., 2009) whereas in Iceland, Stacey and colleagues (Stacey et al., 2006) showed a modest increase of the risk of breast cancer attributed to p.Cys557Ser and demonstrated that the risk of breast cancer among carriers of double mutations in *BARD1* p.Cys557Ser and *BRCA2* 999del5 is significantly increased (OR = 3.11; 95% CI = 1.16-8.4; p. = 0.046) (Stacey et al., 2006). However, in studies conducted in different European populations the role of p.Cys557Ser variant as a modifier of *BRCA1/2*-

associated cancer risk has not been confirmed (Jakubowska et al., 2008; Spurdle et al., 2011).

Sabatier and coworkers, identified a homozygous deletion of the entire *BARD1* gene by carrying out an analysis of an array-based comparative genomic hybridization (aCGH) profiles of breast cancer tumors from 330 patients with invasive breast adenocarcinoma. An additional aCGH analysis of DNA from a blood sample of the carrier of homozygous mutation revealed the presence of a heterozygous deletion of the entire *BARD1* gene. Interestingly, patients who harbor *BARD1* deletion (but are not affected by *BRCA1* mutations) displayed clinicopathological features which are specific for a phenotype associated with mutations occurring in *BRCA1* (Sabatier et al., 2010).

The mutational analysis of *BARD1* recently conducted in *BRCA1/2*-negative families with breast cancer aggregation has identified eleven intronic and fifteen exonic germline variants (one in-frame deletion, four silent, one frameshift and nine missense mutations). Three of these variants, i.e. p.Ile509Thr (c.1526T > C), p.Glu652fs (c.1935_1954dup) and p.Arg658Cys (c.1972C > T) have been shown to predispose to breast cancer and to co-segregate with the disease phenotype in the analyzed families. On the basis of *in silico* predictive analysis it was evaluated that the p.Ile509Thr missense variant, located within the ANK domain, causes the protein instability and improper protein folding. The p.Arg658Cys variant was found to have a "possible effect on function" based on protein modeling. Last but not least, de Brakeleer and colleagues have identified a novel protein truncating mutation, p.Glu652fs, which results in a loss of the entire second BRCT domain of BARD1, which may result in a defective DNA damage response (De Brakeleer et al., 2010).

Ratajska and colleagues conducted screening of germline variants of the *BARD1* gene in 109 of *BRCA1/2* -negative patients from families with the aggregation of breast and/or ovarian cancer via utilizing either denaturing high-performance liquid chromatography (DHPLC) or direct sequencing. Ten exonic and seven intronic variants, including five novel alterations were identified in this study. Three novel *BARD1* mutations [p.Gly439_Leu465del (c.1315-2A > G), p.Gln564X (c.1690C > T), p.Arg659Arg (c.1977A > G)] can be considered as possibly deleterious. The p.Gly439_Leu465del mutation is located in intron 4 within the consensus sequence of the splice acceptor site. The mutation causes skipping of exon 5 and disruption of two ANK repeats implicated in apoptosis and protein-protein association. *In silico* prediction suggests that this variant may cause an alteration of the BARD1 protein structure. Another nonsense, protein truncating mutation, p.Gln564X, which occurs in exon 8, leads to the loss of the BRCT domains. Finally, p.Arg659Arg *BARD1,* a presumably silent mutation located in exon 10 alters the exon splice enhancer motifs (ESE) and leads to exons 2-9 skipping (p.Cys53_Trp635delinsfsX12) (Ratajska et al., 2012).

The study performed recently on a large group (>800) of patients with breast and/or ovarian cancer indicated that large deletions are not common in *BARD1* and therefore may not contribute substantially to the breast cancer risk (Klonowska et al., 2015). The study also revealed that the p.Gln564X, p.Arg659Arg and p.Arg658Cys mutations are recurrent in the Polish population, what indicates their potential founder character (Klonowska et al., 2015). The founder character of these mutations is additionally supported by the fact that they were independently detected in other studies conducted in Polish population as well (Ratajska et al., 2012; Cybulski et al., 2014; Ratajska et al., 2015). The functional and *in silico* analyses suggested their possible deleterious character (Ratajska et al., 2012; Klonowska et al., 2015; Ratajska et al., 2015).

Recently, exome sequencing analyses focused on panels of breast cancer predisposing genes have also led to the identification of potentially deleterious *BARD1* mutations. Additionally, the study showed that *BARD1* is one of the most frequently mutated genes (after several moderate and highly penetrant genes, e.g. *PALB2*, *BRCA1* and *BRCA2)* in patients with breast and/or ovarian cancer (Walsh et al., 2011; Castera et al., 2014; Cybulski et al., 2014; Pennington et al., 2014; Churpek et al., 2015; Couch et al., 2015).

Conclusions

The genetic etiology of breast and/or ovarian cancer cases aggregated in families is only partially known. Apart from *BRCA1/2* and several other genes of moderate to high penetrance, a considerable fraction of breast cancer predisposing factors (~50%) is still unknown. It is presumed that DSB repair genes, encoding proteins that are involved in the same molecular pathway as BRCA1, may be candidate breast cancer susceptibility genes.

The findings that BARD1 is essential for BRCA1 tumor suppression functions and that it operates independently in the regultion of apoptosis, suggest that the *BARD1* gene may serve as a plausible target for mutations predisposing to breast and/or ovarian cancer. Although a number of mutational studies have already been conducted, a study on *BARD1* mutations in patients with breast and/or ovarian cancer is still in its infancy. Despite the fact that several potentially deleterious *BARD1* mutations have been identified, further studies should be carried out to evaluate their breast and/or ovarian cancer predisposing effect and to identify the unexplored mutations affecting the *BARD1* gene.

It is noteworthy that none of the studies conducted so far has provided a clear and statistically supported proof for the role of *BARD1* as a breast cancer susceptibility gene. Therefore, large-scale association studies of the selected *BARD1* mutations would be highly desirable to unequivocally confirm or reject the role that *BARD1* plays in breast and/or ovarian cancer susceptibility. Importantly, if breast cancer risk associated with *BARD1* mutations turns out to be considerably high, the inclusion of testing of the *BARD1* mutations into genetic diagnostics of breast cancer and other genetically associated cancers would be a far-reaching consequence.

Acknowledgements

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

References

Baer R., Ludwig T. (2002) *The BRCA1/BARD1 heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity*. Curr. Opin. Genet. Dev. 12: 86-91.

Beggs A.D., Hodgson S.V. (2009) *Genomics and breast cancer: the different levels of inherited susceptibility*. Eur. J. Hum. Genet. 17: 855-856.

Berx G., Staes K., van Hengel J., Molemans F., Bussemakers M.J., van Bokhoven A., van Roy F. (1995) *Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1)*. Genomics 26: 281-289.

Brozek I., Cybulska C., Ratajska M., Piatkowska M., Kluska A., Balabas A., Dabrowska M., Nowakowska D., Niwinska A., Pamula-Pilat J. et al. (2011) *Prevalence of the most frequent BRCA1 mutations in Polish population*. J. Appl. Genet. 52: 325-330.

Castera L., Krieger S., Rousselin A., Legros A., Baumann J.J., Bruet O., Brault B., Fouillet R., Goardon N., Letac O. et al. (2014) *Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes*. Eur. J. Hum. Genet. 22: 1305-1313.

Chen A., Kleiman F.E., Manley J.L., Ouchi T., Pan Z.Q. (2002) *Autoubiquitination of the BRCA1 *BARD1 RING ubiquitin ligase*. J. Biol. Chem. 277: 22085-22092.

Churpek J.E., Walsh T., Zheng Y., Moton Z., Thornton A.M., Lee M.K., Casadei S., Watts A., Neistadt B., Churpek M.M. et al. (2015) *Inherited predisposition to breast cancer among African American women*. Breast Cancer Res. Treat. 149: 31-39.

Couch F.J., Hart S.N., Sharma P., Toland A.E., Wang X., Miron P., Olson J.E., Godwin A.K., Pankratz V.S., Olswold C. et al. (2015) *Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer*. J. Clin. Oncol. 33: 304-311.

Cybulski C., Lubinski J., Wokolorczyk D., Kuzniak W., Kashyap A., Sopik V., Huzarski T., Gronwald J., Byrski T., Szwiec M. et al. (2014) *Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland*. Clin. Genet. [Epub ahead of print].

Cybulski C., Carrot-Zhang J., Kluzniak W., Rivera B., Kashyap A., Wokolorczyk D., Giroux S., Nadaf J., Hamel N., Zhang S. et al. (2015) *Germline RECQL mutations are associated with breast cancer susceptibility*. Nat. Genet. 47: 643-646.

De Brakeleer S., De Greve J., Loris R., Janin N., Lissens W., Sermijn E., Teugels E. (2010) *Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families*. Hum. Mutat. 31: E1175-1185.

Ellis C.N. (2011) *Obtaining and using genetic information*. In: *Inherited Cancer Syndromes. Current Clinical Management*. Ed. Ellis C.N., New York: Springer-Verlag: 1-4.

Ferla R., Calo V., Cascio S., Rinaldi G., Badalamenti G., Carreca I., Surmacz E., Colucci G., Bazan V., Russo A. (2007) *Founder mutations in BRCA1 and BRCA2 genes*. Ann. Oncol. 18 Suppl 6: vi93-98.

Foulkes W.D. (2008) *Inherited susceptibility to common cancers*. N. Engl. J. Med. 359: 2143-2153.

Garber J.E., Goldstein A.M., Kantor A.F., Dreyfus M.G., Fraumeni J.F., Jr., Li F.P. (1991) *Follow-up study of twenty-four families with Li-Fraumeni syndrome*. Cancer Res. 51: 6094-6097.

Ghimenti C., Sensi E., Presciuttini S., Brunetti I.M., Conte P., Bevilacqua G., Caligo M.A. (2002) *Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations*. Genes Chromosomes Cancer 33: 235-242.

Gorringe K.L., Choong D.Y., Visvader J.E., Lindeman G.J., Campbell I.G. (2008) *BARD1 variants are not associated with breast cancer risk in Australian familial breast cancer*. Breast Cancer Res. Treat. 111: 505-509.

Gorski B., Byrski T., Huzarski T., Jakubowska A., Menkiszak J., Gronwald J., Pluzanska A., Bebenek M., Fischer-Maliszewska L., Grzybowska E. et al. (2000) *Founder mutations in the BRCA1 gene in Polish families with breast-ovarian cancer*. Am. J. Hum. Genet. 66: 1963-1968.

Gorski B., Jakubowska A., Huzarski T., Byrski T., Gronwald J., Grzybowska E., Mackiewicz A., Stawicka M., Bebenek M., Sorokin D. et al. (2004) *A high proportion of founder BRCA1 mutations in Polish breast cancer families*. Int. J. Cancer 110: 683-686.

Grzybowska E., Zientek H., Jasinska A., Rusin M., Kozlowski P., Sobczak K., Sikorska A., Kwiatkowska E., Gorniak L., Kalinowska E. et al. (2000) *High frequency of recurrent mutations in BRCA1 and BRCA2 genes in Polish families with breast and ovarian cancer*. Hum. Mutat. 16: 482-490.

Hankinson S.E., Colditz G.A., Willett W.C. (2004) *Towards an integrated model for breast cancer etiology - The lifelong interplay of genes, lifestyle, and hormones*. Breast Cancer Res. 6: 213-218.

Hansen T., Jonson L., Albrechtsen A., Andersen M.K., Ejlertsen B., Nielsen F.C. (2009) *Large BRCA1 and BRCA2 genomic rearrangements in Danish high risk breast-ovarian cancer families*. Breast Cancer Res .Treat. 115: 315-323.

Hashizume R., Fukuda M., Maeda I., Nishikawa H., Oyake D., Yabuki Y., Ogata H., Ohta T. (2001) *The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation*. J. Biol. Chem. 276: 14537-14540.

Heald B., Church J.M. (2011) *Cancer and genetic counseling*. In: *Inherited Cancer Syndromes. Current Clinical Management*. Ed. Ellis C.N., New York: Springer-Verlag: 23-29.

Hemminki A., Markie D., Tomlinson I., Avizienyte E., Roth S., Loukola A., Bignell G., Warren W., Aminoff M., Hoglund P. et al. (1998) *A serine/threonine kinase gene defective in Peutz-Jeghers syndrome*. Nature 391: 184-187.

Hoeijmakers J.H. (2001) *Genome maintenance mechanisms for preventing cancer*. Nature 411: 366-374.

Huo X., Hu Z., Zhai X., Wang Y., Wang S., Wang X., Qin J., Chen W., Jin G., Liu J. et al. (2007) *Common non-synonymous polymorphisms in the BRCA1 Associated RING Domain (BARD1) gene are associated with breast cancer susceptibility: a case-control analysis*. Breast Cancer Res. Treat. 102: 329-337.

Irminger-Finger I., Soriano J.V., Vaudan G., Montesano R., Sappino A.P. (1998) *In vitro repression of BRCA1-associated RING domain gene, Bard1, induces phenotypic changes in mammary epithelial cells*. J. Cell Biol. 143: 1329-1339.

Irminger-Finger I., Leung W.C., Li J., Dubois-Dauphin M., Harb J., Feki A., Jefford C.E., Soriano J.V., Jaconi M., Montesano R. et al. (2001) *Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis*. Mol. Cell 8: 1255-1266.

Irminger-Finger I., Jefford C.E. (2006) *Is there more to BARD1 than BRCA1?* Nat. Rev. Cancer 6: 382-391.

Irminger-Finger I. (2010) *BARD1, a possible biomarker for breast and ovarian cancer*. Gynecol. Oncol. 117: 211-215.

Ishitobi M., Miyoshi Y., Hasegawa S., Egawa C., Tamaki Y., Monden M., Noguchi S. (2003) *Mutational analysis of BARD1 in familial breast cancer patients in Japan*. Cancer Lett. 200: 1-7.

Jakubowska A., Cybulski C., Szymanska A., Huzarski T., Byrski T., Gronwald J., Debniak T., Gorski B., Kowalska E., Narod S.A. et al. (2008) *BARD1 and breast cancer in Poland*. Breast Cancer Res. Treat. 107: 119-122.

Johnatty S.E., Beesley J., Chen X., Hopper J.L., Southey M.C., Giles G.G., Goldgar D.E., Chenevix-Trench G., Spurdle A.B., Australian Ovarian Cancer Study G. et al. (2009) *The BARD1 Cys557Ser polymorphism and breast cancer risk: an Australian case-control and family analysis*. Breast Cancer Res. Treat. 115: 145-150.

Karppinen S.M., Heikkinen K., Rapakko K., Winqvist R. (2004) *Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer*. J. Med. Genet. 41: e114.

Karppinen S.M., Barkardottir R.B., Backenhorn K., Sydenham T., Syrjakoski K., Schleutker J., Ikonen T., Pylkas K., Rapakko K., Erkko H. et al. (2006) *Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 patients with cancer: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies*. J. Med. Genet. 43: 856-862.

Khanna K.K. (2000) *Cancer risk and the ATM gene: a continuing debate*. J. Natl. Cancer Inst. 92: 795-802.

Klonowska K., Ratajska M., Czubak K., Kuzniacka A., Brozek I., Koczkowska M., Sniadecki M., Debniak J., Wydra D., Balut M. et al. (2015) *Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example*. Sci. Rep. 5: 10424.

Lichtenstein P., Holm N.V., Verkasalo P.K., Iliadou A., Kaprio J., Koskenvuo M., Pukkala E., Skytthe A., Hemminki K. (2000) *Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland*. N. Engl. J. Med. 343: 78-85.

Lim W., Olschwang S., Keller J.J., Westerman A.M., Menko F.H., Boardman L.A., Scott R.J., Trimbath J., Giardiello F.M., Gruber S.B. et al. (2004) *Relative frequency and morphology of cancers in STK11 mutation carriers*. Gastroenterology 126: 1788-1794.

Long J., Delahanty R.J., Li G., Gao Y.T., Lu W., Cai Q., Xiang Y.B., Li C., Ji B.T., Zheng Y. et al. (2013) *A common deletion in the APOBEC3 genes and breast cancer risk*. J. Natl. Cancer Inst. 105: 573-579.

Lux M.P., Fasching P.A., Beckmann M.W. (2006) *Hereditary breast and ovarian cancer: review and future perspectives*. J. Mol. Med. (Berl) 84: 16-28.

Lynch H.T., Casey M.J., Snyder C.L., Bewtra C., Lynch J.F., Butts M., Godwin A.K. (2009) *Hereditary ovarian carcinoma: heterogeneity, molecular genetics, pathology, and management*. Mol. Oncol. 3: 97-137.

Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon

L.R., Chakravarti A. et al. (2009) *Finding the missing heritability of complex diseases*. Nature 461: 747-753.

Masciari S., Dillon D.A., Rath M., Robson M., Weitzel J.N., Balmana J., Gruber S.B., Ford J.M., Euhus D., Lebensohn A. et al. (2012) *Breast cancer phenotype in women with TP53 germline mutations: a Li-Fraumeni syndrome consortium effort*. Breast Cancer Res. Treat. 133: 1125-1130.

Mathew C.G. (2006) *Fanconi anaemia genes and susceptibility to cancer*. Oncogene 25: 5875-5884.

Matsuura S., Tauchi H., Nakamura A., Kondo N., Sakamoto S., Endo S., Smeets D., Solder B., Belohradsky B.H., Der Kaloustian V.M. et al. (1998) *Positional cloning of the gene for Nijmegen breakage syndrome*. Nat. Genet. 19: 179-181.

McCarthy E.E., Celebi J.T., Baer R., Ludwig T. (2003) *Loss of Bard1, the heterodimeric partner of the Brca1 tumor suppressor, results in early embryonic lethality and chromosomal instability*. Mol. Cell Biol. 23: 5056-5063.

Meza J.E., Brzovic P.S., King M.C., Klevit R.E. (1999) *Mapping the functional domains of BRCA1. Interaction of the ring finger domains of BRCA1 and BARD1*. J. Biol. Chem. 274: 5659-5665.

Michailidou K., Hall P., Gonzalez-Neira A., Ghoussaini M., Dennis J., Milne R.L., Schmidt M.K., Chang-Claude J., Bojesen S.E., Bolla M.K. et al. (2013) *Large-scale genotyping identifies 41 new loci associated with breast cancer risk*. Nat. Genet. 45: 353-361, 361e351-352.

Miki Y., Swensen J., Shattuck-Eidens D., Futreal P.A., Harshman K., Tavtigian S., Liu Q., Cochran C., Bennett L.M., Ding W. et al. (1994) *A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1*. Science 266: 66-71.

Morris J.R., Solomon E. (2004) *BRCA1 : BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair*. Hum. Mol. Genet. 13: 807-817.

Nagy R., Sweet K., Eng C. (2004) *Highly penetrant hereditary cancer syndromes*. Oncogene 23: 6445-6470.

Narod S.A., Foulkes W.D. (2004) *BRCA1 and BRCA2: 1994 and beyond*. Nat. Rev. Cancer 4: 665-676.

Neuhausen S., Gilewski T., Norton L., Tran T., McGuire P., Swensen J., Hampel H., Borgen P., Brown K., Skolnick M. et al. (1996) *Recurrent BRCA2 6174delT mutations in Ashkenazi Jewish women affected by breast cancer*. Nat. Genet. 13: 126-128.

Pennington K.P., Walsh T., Harrell M.I., Lee M.K., Pennil C.C., Rendi M.H., Thornton A., Norquist B.M., Casadei S., Nord A.S. et al. (2014) *Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas*. Clin. Cancer Res. 20: 764-775.

Permuth-Wey J., Sellers T.A. (2009) *Epidemiology of ovarian cancer*. Methods Mol. Biol. 472: 413-437.

Pharoah P.D., Guilford P., Caldas C., International Gastric Cancer Linkage C. (2001) *Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families*. Gastroenterology 121: 1348-1353.

Pinheiro H., Bordeira-Carrico R., Seixas S., Carvalho J., Senz J., Oliveira P., Inacio P., Gusmao L., Rocha J., Huntsman D. et al. (2010) *Allele-specific CDH1 downregulation and hereditary diffuse gastric cancer*. Hum. Mol. Genet. 19: 943-952.

Ratajska M., Brozek I., Senkus-Konefka E., Jassem J., Stepnowska M., Palomba G., Pisano M., Casula M., Palmieri G., Borg A. et al. (2008) *BRCA1 and BRCA2 point mutations and large rearrangements in breast and ovarian cancer families in Northern Poland*. Oncol. Rep. 19: 263-268.

Ratajska M., Antoszewska E., Piskorz A., Brozek I., Borg A., Kusmierek H., Biernat W., Limon J. (2012) *Cancer predisposing BARD1 mutations in breast-ovarian cancer families*. Breast Cancer Res. Treat. 131: 89-97.

Ratajska M., Matusiak M., Kuzniacka A., Wasag B., Brozek I., Biernat W., Koczkowska M., Debniak J., Sniadecki M., Kozlowski P. et al. (2015) *Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms.* Oncol. Rep. in press.

Ripperger T., Gadzicki D., Meindl A., Schlegelberger B. (2009) *Breast cancer susceptibility: current knowledge and implications for genetic counselling*. Eur. J. Hum. Genet. 17: 722-731.

Roa B.B., Boyd A.A., Volcik K., Richards C.S. (1996) *Ashkenazi Jewish population frequencies for common mutations in BRCA1 and BRCA2*. Nat. Genet. 14: 185-187.

Rouleau E., Jesson B., Briaux A., Nogues C., Chabaud V., Demange L., Sokolowska J., Coulet F., Barouk-Simonet E., Bignon Y.J. et al. (2012) *Rare germline large rearrangements in the BRCA1/2 genes and eight candidate genes in 472 patients with breast cancer predisposition*. Breast Cancer Res. Treat. 133: 1179-1190.

Roy R., Chun J., Powell S.N. (2012) *BRCA1 and BRCA2: different roles in a common pathway of genome protection*. Nat. Rev. Cancer 12: 68-78.

Ruffner H., Joazeiro C.A., Hemmati D., Hunter T., Verma I.M. (2001) *Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity*. Proc. Natl. Acad. Sci. U S A 98: 5134-5139.

Ryser S., Dizin E., Jefford C.E., Delaval B., Gagos S., Christodoulidou A., Krause K.H., Birnbaum D., Irminger-Finger I. (2009) *Distinct roles of BARD1 isoforms in mitosis: full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2*. Cancer Res. 69: 1125-1134.

Sabatier R., Adelaide J., Finetti P., Ferrari A., Huiart L., Sobol H., Chaffanet M., Birnbaum D., Bertucci F. (2010) *BARD1 homozygous deletion, a possible alternative to BRCA1 mutation in basal breast cancer*. Genes Chromosomes Cancer 49: 1143-1151.

Sauer M.K., Andrulis I.L. (2005) *Identification and characterization of missense alterations in the BRCA1 associated*

RING domain (BARD1) gene in breast and ovarian cancer. J. Med. Genet. 42: 633-638.

Savitsky K., Bar-Shira A., Gilad S., Rotman G., Ziv Y., Vanagaite L., Tagle D.A., Smith S., Uziel T., Sfez S. et al. (1995) *A single ataxia telangiectasia gene with a product similar to PI-3 kinase*. Science 268: 1749-1753.

Shabbeer S., Omer D., Berneman D., Weitzman O., Alpaugh A., Pietraszkiewicz A., Metsuyanim S., Shainskaya A., Papa M.Z., Yarden R.I. (2013) *BRCA1 targets G2/M cell cycle proteins for ubiquitination and proteasomal degradation*. Oncogene 32: 5005-5016.

Sluiter M.D., van Rensburg E.J. (2011) *Large genomic rearrangements of the BRCA1 and BRCA2 genes: review of the literature and report of a novel BRCA1 mutation*. Breast Cancer Res. Treat. 125: 325-349.

Sobczak K., Kozlowski P., Napierala M., Czarny J., Wozniak M., Kapuscinska M., Losko M., Koziczak M., Jasinska A., Powierska J. et al. (1997) *Novel BRCA1 mutations and more frequent intron-20 alteration found among 236 women from Western Poland*. Oncogene 15: 1773-1779.

Spurdle A.B., Marquart L., McGuffog L., Healey S., Sinilnikova O., Wan F., Chen X., Beesley J., Singer C.F., Dressler A.C. et al. (2011) *Common genetic variation at BARD1 is not associated with breast cancer risk in BRCA1 or BRCA2 mutation carriers*. Cancer Epidemiol. Biomarkers Prev. 20: 1032-1038.

Stacey S.N., Sulem P., Johannsson O.T., Helgason A., Gudmundsson J., Kostic J.P., Kristjansson K., Jonsdottir T., Sigurdsson H., Hrafnkelsson J. et al. (2006) *The BARD1 Cys557Ser variant and breast cancer risk in Iceland*. PLoS Med. 3: e217.

Starita L.M., Machida Y., Sankaran S., Elias J.E., Griffin K., Schlegel B.P., Gygi S.P., Parvin J.D. (2004) *BRCA1-dependent ubiquitination of gamma-tubulin regulates centrosome number*. Mol. Cell. Biol. 24: 8457-8466.

Steffen J., Varon R., Mosor M., Maneva G., Maurer M., Stumm M., Nowakowska D., Rubach M., Kosakowska E., Ruka W. et al. (2004) *Increased cancer risk of heterozygotes with NBS1 germline mutations in Poland*. Int. J. Cancer 111: 67-71.

Struewing J.P., Abeliovich D., Peretz T., Avishai N., Kaback M.M., Collins F.S., Brody L.C. (1995) *The carrier frequency of the BRCA1 185delAG mutation is approximately 1 percent in Ashkenazi Jewish individuals*. Nat. Genet. 11: 198-200.

Tan M.H., Mester J.L., Ngeow J., Rybicki L.A., Orloff M.S., Eng C. (2012) *Lifetime cancer risks in individuals with germline PTEN mutations*. Clin. Cancer Res. 18: 400-407.

Thai T.H., Du F., Tsan J.T., Jin Y., Phung A., Spillman M.A., Massa H.F., Muller C.Y., Ashfaq R., Mathis J.M. et al. (1998) *Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers*. Hum. Mol. Genet. 7: 195-202.

Thompson D., Easton D. (2004) *The genetic epidemiology of breast cancer genes*. J. Mammary Gland Biol. Neoplasia 9: 221-236.

Thorlacius S., Olafsdottir G., Tryggvadottir L., Neuhausen S., Jonasson J.G., Tavtigian S.V., Tulinius H., Ogmundsdottir H.M., Eyfjord J.E. (1996) *A single BRCA2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes*. Nat. Genet. 13: 117-119.

Vahteristo P., Syrjakoski K., Heikkinen T., Eerola H., Aittomaki K., von Smitten K., Holli K., Blomqvist C., Kallioniemi O.P., Nevanlinna H. (2006) *BARD1 variants Cys557Ser and Val507Met in breast cancer predisposition*. Eur. J. Hum. Genet. 14: 167-172.

van der Groep P., van der Wall E., van Diest P.J. (2011) *Pathology of hereditary breast cancer*. Cell Oncol. (Dordr) 34: 71-88.

Walsh T., King M.C. (2007) *Ten genes for inherited breast cancer*. Cancer Cell 11: 103-105.

Walsh T., Casadei S., Lee M.K., Pennil C.C., Nord A.S., Thornton A.M., Roeb W., Agnew K.J., Stray S.M., Wickramanayake A. et al. (2011) *Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing*. Proc. Natl. Acad. Sci. U S A 108: 18032-18037.

Wooster R., Bignell G., Lancaster J., Swift S., Seal S., Mangion J., Collins N., Gregory S., Gumbs C., Micklem G. (1995) *Identification of the breast cancer susceptibility gene BRCA2*. Nature 378: 789-792.

Wooster R., Weber B.L. (2003) *Breast and ovarian cancer*. N. Engl. J. Med. 348: 2339-2347.

Wu L.C., Wang Z.W., Tsan J.T., Spillman M.A., Phung A., Xu X.L., Yang M.C., Hwang L.Y., Bowcock A.M., Baer R. (1996) *Identification of a RING protein that can interact in vivo with the BRCA1 gene product*. Nat. Genet. 14: 430-440.

Youlden D.R., Cramb S.M., Dunn N.A., Muller J.M., Pyke C.M., Baade P.D. (2012) *The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality*. Cancer Epidemiol. 36: 237-248.

2

Marcinkowska-Swojak M*, Klonowska K*, Figlerowicz M, Kozlowski P
*Autorzy mieli taki sam wkład w przygotowanie publikacji

An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs

Gene, 2014, 546(2):257-262

Gene 546 (2014) 257–262

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs

Malgorzata Marcinkowska-Swojak [a,1], Katarzyna Klonowska [a,1], Marek Figlerowicz [a,b], Piotr Kozlowski [a,b,*]

[a] European Centre of Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
[b] Poznan University of Technology, Pl. Marii Sklodowskiej-Curie 5, 60-965 Poznan, Poland

ARTICLE INFO

Article history:
Received 23 April 2014
Accepted 31 May 2014
Available online 2 June 2014

Keywords:
MLPA
Multi-allelic CNV
CCL3L1
DEFB
UGT2B17

ABSTRACT

Copy number variation has recently been recognized as an important type of genetic variation that modifies human phenotypes. Copy number variants (CNVs) are being increasingly associated with various human phenotypes and diseases. However, the lack of an appropriate method that allows fast, inexpensive and, most importantly, accurate CNVs genotyping significantly hampers CNV analysis. This limitation especially affects the analysis of multi-allelic CNVs that frequently modify various phenotypes. Recently, we developed a multiplex ligation-dependent probe amplification (MLPA)-based strategy for multiplex copy number genotyping and the validation of candidate CNV-miRNAs. Here we present the adaptation and optimization of this recently developed method for high-resolution genotyping of individual disease-related multi-allelic CNVs. We developed appropriate assays for three well-known and extensively studied CNVs: CNV-CCL3L1, CNV-DEFB, and CNV-UGT2B17, which have been associated with various human phenotypes including inflammation-related and infectious diseases. With the use of these assays we identified several general factors that allow to increase the resolution of the copy number genotyping. Performed experiments confirmed the high reproducibility and accuracy of the obtained genotyping results. The reliability of the results and relatively low per-genotype cost makes this strategy an attractive method for large-scale experiments such as genotype–phenotype association studies.

1. Introduction

Copy number variants (CNVs) have recently become a well-recognized component of genetic variation (Iafrate et al., 2004; Sebat et al., 2004). CNVs are genomic regions (of roughly 1 kb to 1 Mb in length) that differ in copy number (CN) in compared genomes. CNVs include deletions, duplications, multiple duplications or more complex rearrangements. Common CNVs, also known as copy number polymorphisms (CNPs), account for approximately 10% of the human genome. Although CNVs are more frequent in intergenic regions, they overlap hundreds of protein-coding genes, regulatory sequences, and other functional genetic elements. Although the majority of CNVs are probably neutral, increasing numbers of CNVs are being associated with various human phenotypes, including diseases (reviewed recently (Cantsilieris and White, 2013)). CNVs that modify human phenotypes include the following examples: the simple bi-allelic CNV on chromosome 4q13.2 spanning *UGT2B17* (CNV-UGT2B17) associated with a decreased risk of osteoporosis and bone-structure-related phenotypes (Yang et al., 2008) and two complex multi-allelic CNVs: (i) the CNV on chromosome 17q12 spanning *CCL3L1* and *CCL4L1* (CNV-CCL3L1) and (ii) the CNV on chromosome 8p23.1 spanning the beta-defensin cluster (CNV-DEFB). The CNV-CCL3L1 and CNV-DEFB were associated with many infectious and inflammatory-related diseases including: HIV infection and AIDS development (Gonzalez et al., 2005), malaria (Carpenter et al., 2012), systemic lupus erythematosus (Mamtani et al., 2008), psoriasis (Hollox et al., 2008), and Crohn's disease (Bentley et al., 2010; Fellermann et al., 2006) (reviewed in (Colobran et al., 2010; Olsson and Holmdahl, 2012)). It has to be noted however that there are many controversies and conflicting results regarding these associations (Bhattacharya et al., 2009; Field et al., 2009; Urban et al., 2009) (discussed in Olsson and Holmdahl, 2012). It was suggested that observed discordances and problems with replication of reported CNV associations result mostly from the lack of appropriate method allowing reliable CNV genotyping.

A method with the potential for individual CNV genotyping is a multiplex ligation-dependent probe amplification (MLPA) (Garcia-Canas et al., 2011; Groth et al., 2008; Perne et al., 2009; Zhang et al., 2014) that is based on the principle that the signals of the MLPA probes are proportional to the CN of the targeted genomic regions. Briefly, MLPA is a multiplex

Abbreviations: CN, copy number; CNP, copy number polymorphism; CNV, copy number variant; CV, coefficient of variation; MAPH, multiplex amplifiable probe hybridization; MLPA, multiplex ligation-dependent probe amplification; PRT, paralog ratio test; QC, quality control; qPCR, quantitative PCR.

* Corresponding author at: Noskowskiego 12/14, 61-704 Poznan, Poland.
E-mail addresses: m-marcinkowska@o2.pl (M. Marcinkowska-Swojak), k.klonowska89@gmail.com (K. Klonowska), marekf@ibch.poznan.pl (M. Figlerowicz), kozlowp@yahoo.com (P. Kozlowski).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2014.05.072
0378-1119/

method that utilizes up to 45 probes that are specific for different genomic locations (Schouten et al., 2002). Each MLPA probe consists of two half-probes whose target-specific sequences hybridize to directly adjacent target sequences, allowing subsequent ligation and dosage-dependent amplification of probes that correctly recognize their targets. The products of MLPA probe amplification are separated by capillary electrophoresis. The signal of MLPA probes (peak heights or areas) is first normalized against the signals of the control probes that are located in CN-stable regions and are then compared (divided by) to the corresponding signals observed in a reference sample. The CN-genotypes may also be deduced from the distribution of the normalized signal without comparing it to a reference sample (Marcinkowska-Swojak et al., 2013; White et al., 2007). The normalized signal distribution may be visualized on a signal scatter plot on which different CN-genotypes form distinct signal clusters. Such an approach is more practical in cases of common CNVs, and it allows the simultaneous analysis of all samples. MLPA was originally designed and is commonly used for the detection of large mutations in disease-related genes (Aretz et al., 2005; Bunyan et al., 2007; De Luca et al., 2007; Kozlowski et al., 2007b; Montagna et al., 2003). Other proposed applications and modifications of MLPA include expression analysis (Eldering et al., 2003; Mykowska et al., 2011), methylation identification (Nygren et al., 2005) and multi-color labeling (White et al., 2004) as well as stuffer-free and multiple-oligo probe design (Serizawa et al., 2010; Shin et al., 2012) (reviewed in (Kozlowski et al., 2008)).

Recently, we have developed a new strategy for CNV genotyping that takes advantage of the general principles of the MLPA method (Marcinkowska-Swojak et al., 2013). However, in comparison to standard MLPA, in our strategy short oligonucleotide probes instead of long MLPA probes are used. The use of short MLPA probes for detection of large mutations in various human genes was shown before (Kozlowski et al., 2007b; Stern et al., 2004; White et al., 2004). It allows easy custom design and generation of assays for almost any genomic region of interest. We have successfully used this method for multiplex analyses and validation of candidate CNV-miRNAs, which were previously identified with the use of computational analysis (Marcinkowska et al., 2011). Although proposed strategy allowed us to have accurate and reliable genotyping of the most selected CNVs, it worked somewhat worse with complex multi-allelic CNVs, and did not always allowed discrete resolution of CN-genotypes. The genotyping errors affected mostly higher (CN > 4) CN-genotypes. Similar limitations of CN-genotyping were also observed before (Carpenter et al., 2012; Groth et al., 2008; Zhang et al., 2014).

In this study, we adapted a recently developed MLPA-based strategy for high-resolution (i.e., with integer CN calling) genotyping of multi-allelic CNVs associated with human phenotypes. We optimized this strategy to increase the resolution of the CN-specific signals and thus to improve the reliability of CN-genotyping. The developed assays confirmed the improved reliability and reproducibility of the identified CN-genotypes. Most of the proposed solutions are general and can be applied to almost any CNV of interest.

2. Materials and methods

A total of 96 DNA samples from three HapMap populations were purchased from Coriell Institute (Camden, NJ, USA): 48 samples from a European population (CEU) from the Centre d'Etude du Polymorphisme Humain (CEPH) Collection, representing 16 family trios (2 parents and one child); 24 unrelated samples from a Han Chinese population in Beijing, China (CHB); and 24 unrelated samples from a Yoruba population from Ibadan, Nigeria (YRI). The CHB and YRI samples were combined into one batch (CHB + YRI) and analyzed in one experiment. According to information provided by the Coriell Institute, all samples were diluted in deionized water to a working concentration of 50 ng/μl of which 5 μl was used for single MLPA reaction.

The MLPA probes and the general probe set layout were designed according to a previously proposed strategy (Kozlowski et al., 2007b; Marcinkowska et al., 2010). To reduce the disproportion between the signals of the CNV-specific probes and the control probes, the signals of the CNV-CCL3L1 and CNV-DEFB specific probes were lowered by the use of probe-specific competitors (short oligonucleotides that compete with the MLPA probes for access to the target sequences; for details see Kozlowski et al., 2007a; Marcinkowska-Swojak et al., 2013). Competitors were used in the same concentration as MLPA probes (1:1 ratio). The detailed characteristics and sequences of all probes used in this study are presented in Supplementary data (Table S1). The MLPA reactions were run and analyzed as described previously (Marcinkowska-Swojak et al., 2013; Schouten et al., 2002). For evaluation of assay performance each experiment was performed in duplicate. CN-genotypes were determined based on distribution of normalized probe signals visualized on the signal histograms (see Fig. 1c), 2D signal scatter plots (see Fig. 2a) or 3D signal scatter plots (see Supplementary data). Assuming almost linear relation between CN and MLPA-probe signal, the CN-genotype of the lowest signal cluster is estimated based on comparison of its distance to the '0,0' coordinate and the subsequent inter-cluster distance (for details see Marcinkowska-Swojak et al., 2013). All of the reagents, except for the MLPA probes, were purchased from MRC-Holland (Amsterdam, the Netherlands). The MLPA probes were synthesized by IDT (Skokie, IL, USA).

All statistical analyses were performed using Prism v. 4.0 (GraphPad, San Diego, CA). 3D scatter plots were generated with the use of function scatterplot3d() from package scatterplot3d (v. 0.3-33), program R version 2.13.1, all other graphs were generated with the use of Microsoft Excel.

All of the human genome coordinates indicated in this report refer to the Feb. 2009 (GRCh37/hg19) human reference sequence.

3. Results

For our analysis we selected three well-recognized CNVs mentioned above: CNV-CCL3L1, CNV-DEFB, and CNV-UGT2B17. Two of these CNVs, CNV-CCL3L1 and CNV-DEFB, are multi-allelic CNVs located in regions of complex genomic structure that contain multiple repeat elements and segmental duplications. CNV-UGT2B17 is a relatively simple bi-allelic polymorphism. For the MLPA probe and assay design, we applied the general strategy utilized before for the detection of large mutations (Kozlowski et al., 2007b). For all performed analyses we used three sets of HapMap samples: 48 European samples (CEU), 24 Asiatic samples (CHB) and 24 African samples (YRI). Each experiment was performed in duplicate.

The important factors that influence performance of all CN-genotyping methods are: signal resolution (distance between signals corresponding to subsequent CN-genotypes) and signal variation. In the case of MLPA, both these factors depend mostly on the reliability of CNV-specific probes. Therefore, in the first step of assay optimization, we carefully selected CNV-specific probes based on their experimentally evaluated performance. For example, to study CNV-CCL3L1, we designed seven MLPA probes located in different positions across the region spanning *CCL3L1* and *CCL4L1* (Fig. S1, Table 1, Table S1). All of these probes were designed to specifically recognize all of the reference copies of the *CCL3L1/CCL4L1* region but not its highly homologous pseudogenes and paralogs, such as the region containing *CCL3* and *CCL4* (see Fig. S1). All seven CNV-CCL3L1-specific probes, together with five control probes, were used for the optimization analyses. To test the reliability and reproducibility of the CNV-specific probes, we compared their normalized signal across all the analyzed samples. The representative pairwise correlation analysis is shown in Fig. S2, and the average correlation coefficient (R^2) values are shown in Fig. 1a. Consistently, the normalized signals of the five probes showed very high correlations with each other (R^2 ~ 0.98), whereas the signals of the two other probes (CCL3L1_5 and CCL3L1_6) showed distinctly lower correlation (R^2 ~ 0.90–0.95). Because such deviation may affect the quality of CN-genotyping, especially in cases of multi-allelic CNVs, we excluded the CCL3L1_5 and CCL3L1_6 probes from further analyses. The very high correlation between five probes covering CNV-CCL3L1

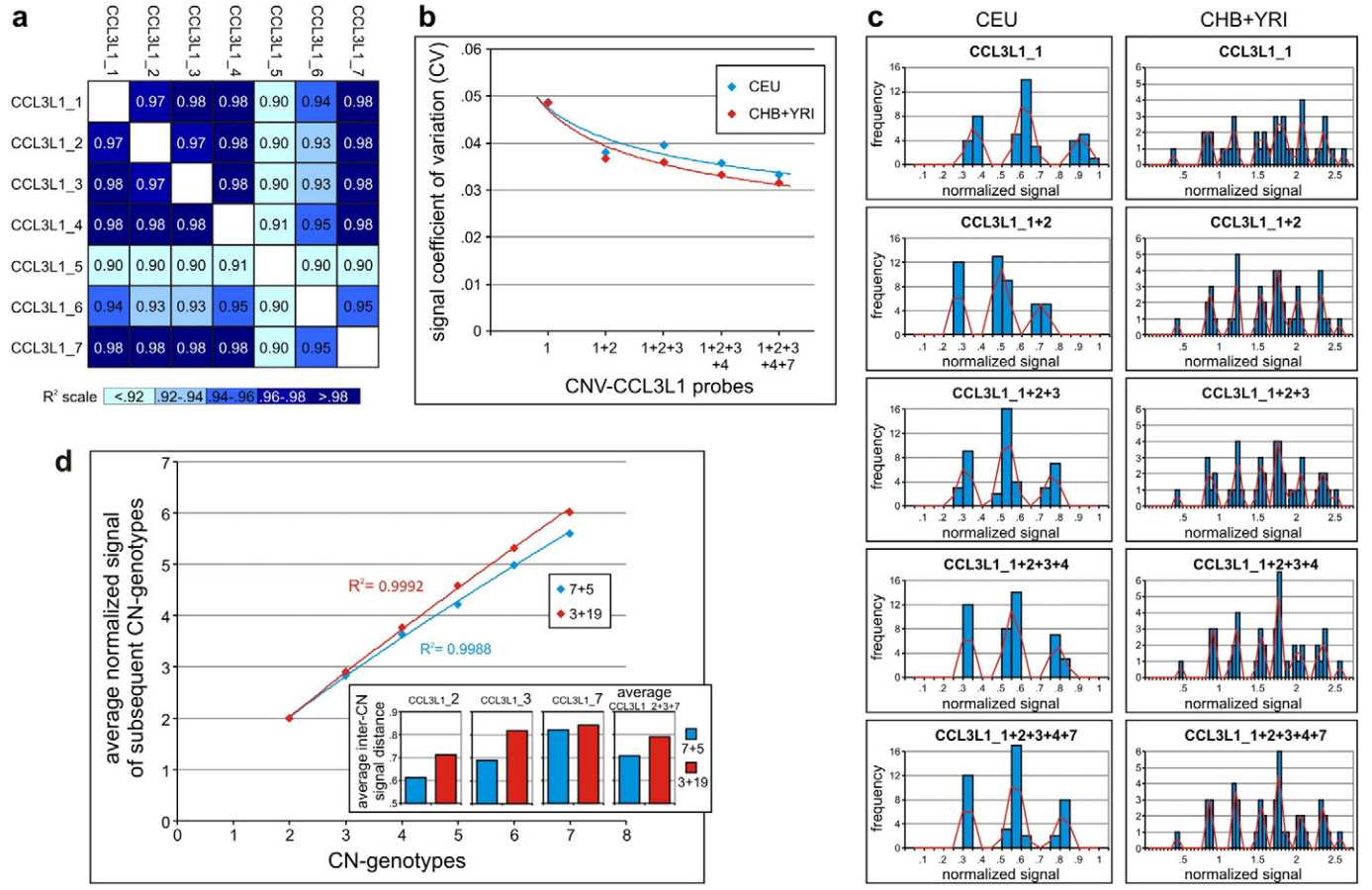


Fig. 1. Probe selection and optimization of the MLPA assay for CNV-CCL3L1. *a.* The result of pairwise probe-to-probe signal correlation analyses (average for two experiments for all of the analyzed samples, N = 96). The heat map shows the R^2 values, with darker blue and lighter blue indicating higher and lower R^2 values, respectively. Probes CCL3L1_1, 2, 3, 4, and 7 show a distinctly higher correlation than probes CCL3L1_5 and 6 (see also Fig. S2). *b.* The variation of the average signal of an increasing number of probes (from one to five probes) selected based on analyses presented in *a*. The y-axis shows the calculated values of the CV, and the x-axis shows the number and ID of the probes for which the signals were averaged. The CV presented on the graph is an average of the CV values calculated separately for each CN-genotype represented in an analyzed sample set by a substantial number of samples (>4). The CN values were calculated independently for the sample sets CEU (blue points) and CHB + YRI (red points), analyzed in separate experiments. *c.* The histograms showing distributions of the normalized signal for different numbers of probes [from 1 (top panels) to 5 (bottom panels)] combined as in *b*. Visibly distinct clusters represent the subsequent CN-genotypes. The red lines in each panel represent the moving averages (period 2) of the bar heights. The graphs for the CEU and CHB + YRI sample sets are shown on the left and right, respectively. Note that similar results to those shown in panels b and c were obtained when different combinations of CNV-CCL3L1 probes were applied (data not shown). *d.* The distance between the average signals of subsequent CN-genotype clusters measured with the use of the two MLPA assays developed for CNV-CCL3L1: (i) an optimization assay composed of 7 CNV-specific and 5 control probes (7 + 5, blue points) and (ii) a genotyping assay composed of 3 CNV-specific and 19 control probes (3 + 19, red points). The scatter plot shows the average values for the three probes (CCL3L1_2, 3 and 7) used in both assays. For the sake of visual comparison signals of 7 + 5 and 3 + 19 assays were normalized against the signal of CN-genotype 2. The bar plots show the average distance between the subsequent signal cluster for all three probes analyzed together and for each probe separately.

region indicates also that there is no substantial structural and CN variation along this region (see Fig. S2).

We then experimentally determined the optimal number of MLPA probes for genotyping of individual CNVs. We compared the variations [expressed as coefficient of variation (CV)] of the normalized signals of individual probes and the variation of the average normalized signal from increasing number of probes (from two to five probes). As shown in Fig. 1b, the signal variation decreases mostly when the signals of two probes are combined, whereas the addition of more probes does not substantially decrease the signal variation. Combination of probe signals also increases the resolution of CN-genotypes (Fig. 1c). The average signal from two or three probes separates CN-genotypes better than the signal from a single probe. A further increase in the number of probes does not substantially improve the resolution. Therefore based on the above results, we selected three probes to genotype the selected CNVs. Although the use of three probes instead of two does not substantially improve the resolution of CN-genotypes or decrease the signal variation, the additional advantages of using three probes include: the identification of CN-genotypes on three different 2D signal scatter plots and the probe signal presentation on 3D signal scatter plots (Fig. S3). The above mentioned strategies of signal analysis allow validation of genotyping results and may help to classify questionable genotypes.

Reducing the number of CNV-specific probes from seven to three also decreases the disproportion between the CNV-specific and control probes. The excessive influence of CNV-specific probes on the total signal (especially in cases of high CN-genotypes, CN >4) affects the linear relationship between the probe signal and the CN, thus reducing the separation between the signals of subsequent CN-genotypes. Therefore, to reduce the influence of CNV-specific probes on the total MLPA signal, we also increased the number of control probes (from 5 to 19). As shown in Fig. 1d, the magnitude of the difference in the signals (resolution) of subsequent CN-genotypes was higher in experiments with the CNV-CCL3L1 genotyping set (3 + 19; CNV-specific probes + control probes, respectively), compared to the CNV-CCL3L1 optimization set (7 + 5 probes) (Fig. 1d).

Similar optimization steps were applied for CNV-DEFB. The results obtained for CNV-DEFB were consistent with those for CNV-CCL3L1 (data not shown), and three CNV-DEFB-specific probes were selected for the genotyping set. Because CNV-UGT2B17 is a simple bi-allelic

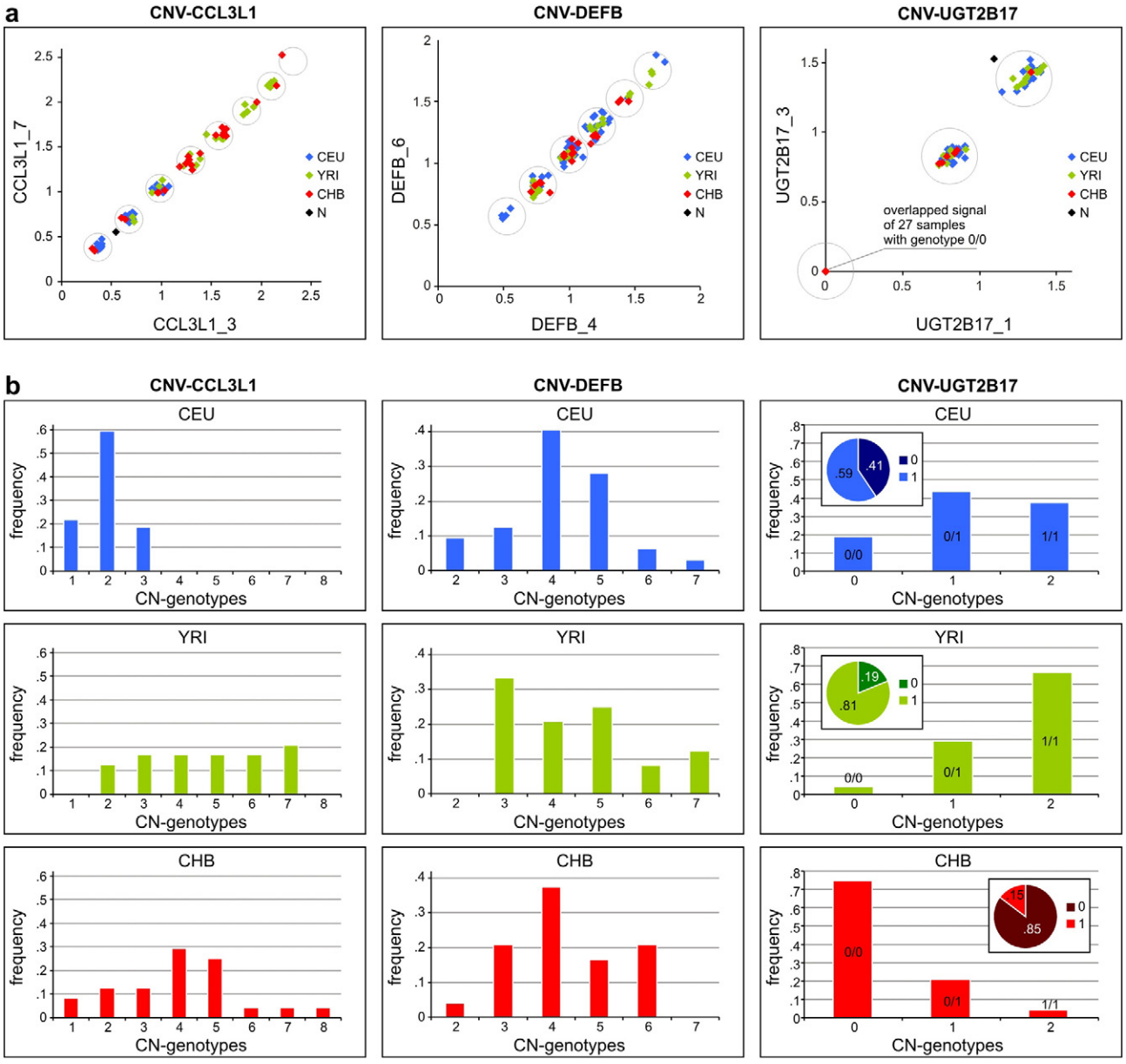


Fig. 2. Representative results of the selected CNV genotyping. *a.* From left to right: the selected 2D signal scatter plots of CNV-CCL3L1, CNV-DEFB, and CNV-UGT2B17 for the CEU (blue), CHB (red), and YRI (green) samples. Each point on the plot represents a single sample that is color-labeled according to the population. The black points represent no-call (N) genotypes. The number in the parentheses indicates the CN-genotype of the lowest signal cluster. The x and y coordinates indicate the normalized signals of the selected two (out of three) CNV-specific probes (probe IDs are indicated along the x and y axes). *b.* The CN-genotype frequency distribution of CNV-CCL3L1, CNV-DEFB, and CNV-UGT2B17 in the three tested populations. The blue, red, and green bar plots show the observed frequencies (y-axis) of the CN-genotypes (x-axis) in the CEU, CHB, and YRI samples, respectively. The CN-allele frequency was also calculated for CNV-UGT2B17 and is presented in the pie charts. The alleles constituting genotypes with 0, 1, and 2 copies are indicated on the bars.

polymorphism, the probes used for its genotyping were not validated experimentally. The characteristics of the designed MLPA sets are presented in Table 1.

To demonstrate the performance of the designed MLPA assays, we used them to genotype the CEU, CHB, and YRI sample sets. Representative examples of the CNV-CCL3L1, CNV-DEFB, and CNV-UGT2B17 genotyping analyses as 2D signal scatter plots are shown in Fig. 2a. The corresponding results as 3D signal scatter plots and different combinations of probes in the 2D signal scatter plots are shown in Fig. S3. In all of the cases, the normalized MLPA-probe signals formed distinctive clusters corresponding to CN-genotypes (Figs. 2a and S3).

Extensive quality control (QC) analyses demonstrated very high experiment-to-experiment reproducibility (99%) and very high concordance of CN-genotypes between our study and the reference genotypes previously determined with the use of CNV-dedicated high-density hybrid arrays (overall 97%; 100% for genotypes <7) (McCarroll et al., 2008). In addition, the genotypes of bi-allelic CNV-UGT2B17 showed agreement with the Hardy–Weinberg equilibrium (in all three tested populations) and perfect agreement with the Mendelian mode of inheritance (in the CEU parents–offspring trios) (Fig. S4). It has to be noted however, that QC estimations may vary depending on the number and quality of the analyzed DNA samples.

For all tested CNVs, we determined the frequency of integer CN-genotypes in all three tested populations (Fig. 2b, Table S2). The general CN-genotype distribution differed significantly between the analyzed populations (Fig. 2b). The CN-genotype distributions and the differences between the populations are consistent with previous reports (Conrad et al., 2010; Gonzalez et al., 2005; Hollox et al., 2008; McCarroll et al., 2008; Yang et al., 2008), additionally confirming the reliability of our assays. For example, an average CN-genotype of

Table 1
General characteristics of the designed MLPA sets used in this study.

Designed probes		Optimization set	Genotyping set
CNV-CCL3L1 assay			
CNV-specific probes	CCL3L1_1	+	
	CCL3L1_2	+	+
	CCL3L1_3	+	+
	CCL3L1_4	+	
	CCL3L1_5	+	
	CCL3L1_6	+	
	CCL3L1_7	+	+
Control probes	01–05	+	+
	06–16, 18–20		+
CNV-DEFB assay			
CNV-specific probes	DEFB_1	+	+
	DEFB_2	+	
	DEFB_3	+	
	DEFB_4	+	+
	DEFB_5	+	
	DEFB_6	+	+
Control probes	01–05	+	+
	06–15, 17–20		+
CNV-UGT2B17 assay			
CNV-specific probes	UGT2B17_1		+
	UGT2B17_3		+
	UGT2B17_4		+
Control probes	01–05		+

+ indicates probes used in the optimization or genotyping MLPA sets. The sequence and detailed characterization of each probe are presented in Table S1.

CNV-CCL3L1 is 2 (range 1–3) in the European population, 4 (range 1–8) in the Asiatic population and 5 (range 2–7) in the African population.

4. Discussion

Although several methods, both genome-wide and locus-specific, have been developed for CNV identification and analysis, a method that can serve as the gold standard for CNV genotyping is still needed. Such a method is especially needed for multi-allelic CNVs, which are frequently located in complex genomic regions. Among the methods used for locus-specific CNV genotyping are: quantitative PCR (qPCR) that although is the most commonly used method, it also raises the most concerns (Fernandez-Jimenez et al., 2011; Fode et al., 2011; Hosono et al., 2009; Pelak et al., 2011; Zhang et al., 2014), paralog ratio test (PRT) (Armour et al., 2007; Carpenter et al., 2012; Hollox et al., 2008; Walker et al., 2009), multiplex amplifiable probe hybridization (MAPH) (den Dunnen and White, 2006; Sellner and Taylor, 2004) and recently proposed droplet digital PCR (McDermott et al., 2013; Miotke et al., 2014). The advantages and limitations of the currently used methods, as well as the need for a method, allowing rapid and accurate CN determination in a large number of samples, have been previously discussed (Cantsilieris et al., 2012; McCarroll et al., 2008).

Here, we have developed MLPA-based assays for individual CN-genotyping of well-known disease-related CNVs. Two of these CNVs (CNV-CCL3L1 and CNV-DEFB) are multi-allelic and are located in complex segmentally duplicated genomic regions, whereas CNV-UGT2B17 is simple bi-allelic CNV. To improve the performance of the developed assays, we performed several optimization steps, including: (i) experimental selection/validation of the MLPA probes, (ii) selection of the optimal number of CNV-specific probes, and (iii) enhancement of the signal resolution by increasing the ratio of control to CNV-specific probes. The results of the two latter steps are general and may be applied to any MLPA-based assay for multi-allelic CNVs.

Additional advantage of CNV genotyping strategy presented here is its relatively low cost. The cost per genotype is ~$8, which consists of the capillary electrophoresis separation (~$3), MLPA reagents and consumables (~$1), and a starting cost of probe synthesis, which will vary depending on the scale of the experiment (~$4, assuming an analysis of 1000 samples). However, once the probes are synthesized, they can be used for thousands of reactions. The turnaround time required to run the assay and analyze the results is similar to standard MLPA analyses (1–2 days) and depends on the pipeline setup and the skill of the experimenter. Although the assays presented in this article were designed for individual CNVs, capacity of short-probe MLPA assays (39) allow combination and multiplex analysis of several (~5) CNVs of interest.

In conclusion, our study demonstrates a valuable approach for high accuracy CN-genotyping of multi-allelic CNVs. Such assays can be used in large-scale experiments, as for instance association studies.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgments

This work was supported by the National Science Centre (2011/01/B/NZ5/02773, 2011/01/B/NZ2/04816).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2014.05.072.

References

Aretz, S., Stienen, D., Uhlhaas, S., Loff, S., Back, W., Pagenstecher, C., McLeod, D.R., Graham, G.E., Mangold, E., Santer, R., Propping, P., Friedl, W., 2005. High proportion of large genomic STK11 deletions in Peutz–Jeghers syndrome. Hum. Mutat. 26, 513–519.
Armour, J.A., Palla, R., Zeeuwen, P.L., den Heijer, M., Schalkwijk, J., Hollox, E.J., 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res. 35, e19.
Bentley, R.W., Pearson, J., Gearry, R.B., Barclay, M.L., McKinney, C., Merriman, T.R., Roberts, R.L., 2010. Association of higher DEFB4 genomic copy number with Crohn's disease. Am. J. Gastroenterol. 105, 354–359.
Bhattacharya, T., Stanton, J., Kim, E.-Y., Kunstman, K.J., Phair, J.P., Jacobson, L.P., Wolinsky, S.M., 2009. CCL3L1 and HIV/AIDS susceptibility. Nat. Med. 15, 1112–1115.
Bunyan, D.J., Skinner, A.C., Ashton, E.J., Sillibourne, J., Brown, T., Collins, A.L., Cross, N.C., Harvey, J.F., Robinson, D.O., 2007. Simultaneous MLPA-based multiplex point mutation and deletion analysis of the dystrophin gene. Mol. Biotechnol. 35, 135–140.
Cantsilieris, S., White, S.J., 2013. Correlating multiallelic copy number polymorphisms with disease susceptibility. Hum. Mutat. 34, 1–13.
Cantsilieris, S., Baird, P.N., White, S.J., 2012. Molecular methods for genotyping complex copy number polymorphisms. Genomics (epub ahead of print).
Carpenter, D., Farnert, A., Rooth, I., Armour, J.A., Shaw, M.A., 2012. CCL3L1 copy number and susceptibility to malaria. Infect. Genet. Evol. 12, 1147–1154.
Colobran, R., Pedrosa, E., Carretero-Iglesia, L., Juan, M., 2010. Copy number variation in chemokine superfamily: the complex scene of CCL3L–CCL4L genes in health and disease. Clin. Exp. Immunol. 162, 41–52.
Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurles, M.E., 2010. Origins and functional impact of copy number variation in the human genome. Nature 464, 704–712.
De Luca, A., Bottillo, I., Dasdia, M.C., Morella, A., Lanari, V., Bernardini, L., Divona, L., Giustini, S., Sinibaldi, L., Novelli, A., Torrente, I., Schirinzi, A., Dallapiccola, B., 2007. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. J. Med. Genet. 44, 800–808.
den Dunnen, J.T., White, S.J., 2006. MLPA and MAPH: sensitive detection of deletions and duplications. Curr. Protoc. Hum. Genet. 51:7.14.1–51:7.14.20.
Eldering, E., Spek, C.A., Aberson, H.L., Grummels, A., Derks, I.A., de Vos, A.F., McElgunn, C.J., Schouten, J.P., 2003. Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. Nucleic Acids Res. 31, e153.
Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W., Teml, A., Schwab, M., Lichter, P., Radlwimmer, B., Stange, E.F., 2006. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predispose to Crohn disease in the colon. Am. J. Hum. Genet. 79, 439–448.
Fernandez-Jimenez, N., Castellanos-Rubio, A., Plaza-Izurieta, L., Gutierrez, G., Irastorza, I., Castano, L., Vitoria, J.C., Bilbao, J.R., 2011. Accuracy in copy number calling by qPCR and PRT: a matter of DNA. PLoS One 6, e28910.
Field, S.F., Howson, J.M.M., Maier, L.M., Walker, S., Walker, N.M., Smyth, D.J., Armour, J.A., Clayton, D.G., Todd, J.A., 2009. Experimental aspects of copy number variant assays at CCL3L1. Nat. Med. 15, 1115–1117.

Fode, P., Jespersgaard, C., Hardwick, R.J., Bogle, H., Theisen, M., Dodoo, D., Lenicek, M., Vitek, L., Vieira, A., Freitas, J., Andersen, P.S., Hollox, E.J., 2011. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PLoS One 6, e16768.

Garcia-Canas, V., Mondello, M., Cifuentes, A., 2011. Combining ligation reaction and capillary gel electrophoresis to obtain reliable long DNA probes. J. Sep. Sci. 34, 1011–1019.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W., Clark, R.A., Anderson, S.A., O'Connell, R.J., Agan, B.K., Ahuja, S.S., Bologna, R., Sen, L., Dolan, M.J., Ahuja, S.K., 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307, 1434–1440.

Groth, M., Szafranski, K., Taudien, S., Huse, K., Mueller, O., Rosenstiel, P., Nygren, A.O., Schreiber, S., Birkenmeier, G., Platzer, M., 2008. High-resolution mapping of the 8p23.1 beta-defensin cluster reveals strictly concordant copy number variation of all genes. Hum. Mutat. 29, 1247–1254.

Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M., Reis, A., Armour, J.A., Schalkwijk, J., 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nat. Genet. 40, 23–25.

Hosono, N., Kato, M., Kiyotani, K., Mushiroda, T., Takata, S., Sato, H., Amitani, H., Tsuchiya, Y., Yamazaki, K., Tsunoda, T., Zembutsu, H., Nakamura, Y., Kubo, M., 2009. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. Clin. Chem. 55, 1546–1554.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. Nat. Genet. 36, 949–951.

Kozlowski, P., Lin, M., Meikle, L., Kwiatkowski, D.J., 2007a. Robust method for distinguishing heterozygous from homozygous transgenic alleles by multiplex ligation-dependent probe assay. Biotechniques 42, 584 (586, 588).

Kozlowski, P., Roberts, P., Dabora, S., Franz, D., Bissler, J., Northrup, H., Au, K.S., Lazarus, R., Domanska-Pakiela, D., Kotulska, K., Jozwiak, S., Kwiatkowski, D.J., 2007b. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype–phenotype correlations. Hum. Genet. 121, 389–400.

Kozlowski, P., Jasinska, A.J., Kwiatkowski, D.J., 2008. New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis 29, 4627–4636.

Mamtani, M., Rovin, B., Brey, R., Camargo, J.F., Kulkarni, H., Herrera, M., Correa, P., Holliday, S., Anaya, J.M., Ahuja, S.K., 2008. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. Ann. Rheum. Dis. 67, 1076–1083.

Marcinkowska, M., Wong, K.-K., Kwiatkowski, D.J., 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. Sci. World J. 10, 2003–2018.

Marcinkowska, M., Szymanski, M., Krzyzosiak, W.J., Kozlowski, P., 2011. Copy number variation of microRNA genes in the human genome. BMC Genomics 12, 183.

Marcinkowska-Swojak, M., Uszczynska, B., Figlerowicz, M., Kozlowski, P., 2013. An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example. Hum. Mutat. 34, 763–773.

McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., Altshuler, D., 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat. Genet. 40, 1166–1174.

McDermott, G.P., Do, D., Litterst, C.M., Maar, D., Hindson, C.M., Steenblock, E.R., Legler, T.C., Jouvenot, Y., Marrs, S.H., Bemis, A., Shah, P., Wong, J., Wang, S., Sally, D., Javier, L., Dinio, T., Han, C., Brackbill, T.P., Hodges, S.P., Ling, Y., Klitgord, N., Carman, G.J., Berman, J.R., Koehler, R.T., Hiddessen, A.L., Walse, P., Bousse, L., Tzonev, S., Hefner, E., Hindson, B.J., Cauly 3rd, T.H., Hamby, K., Patel, V.P., Regan, J.F., Wyatt, P.W., Karlin-Neumann, G.A., Stumbo, D.P., Lowe, A.J., 2013. Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR. Anal. Chem. 85, 11619–11627.

Miotke, L., Lau, B.T., Rumma, R.T., Ji, H.P., 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.

Montagna, M., Dalla Palma, M., Menin, C., Agata, S., De Nicolo, A., Chieco-Bianchi, L., D'Andrea, E., 2003. Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. Hum. Mol. Genet. 12, 1055–1061.

Mykowska, A., Sobczak, K., Wojciechowska, M., Kozlowski, P., Krzyzosiak, W.J., 2011. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. Nucleic Acids Res. 39, 8938–8951.

Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., Errami, A., 2005. Methylation-specific MLPA (MS-MLPA):L simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res. 33, e128.

Olsson, L.M., Holmdahl, R., 2012. Copy number variation in autoimmunity—importance hidden in complexity? Eur. J. Immunol. 42, 1969–1976.

Pelak, K., Need, A.C., Fellay, J., Shianna, K.V., Feng, S., Urban, T.J., Ge, D., De Luca, A., Martinez-Picado, J., Wolinsky, S.M., Martinson, J.J., Jamieson, B.D., Bream, J.H., Martin, M.P., Borrow, P., Letvin, N.L., McMichael, A.J., Haynes, B.F., Telenti, A., Carrington, M., Goldstein, D.B., Alter, G., 2011. Copy number variation of KIR genes influences HIV-1 control. PLoS Biol. 9, e1001208.

Perne, A., Zhang, X., Lehmann, L., Groth, M., Stuber, F., Book, M., 2009. Comparison of multiplex-dependent probe amplification and real-time PCR accuracy for gene copy number quantification using the beta-defensin locus. Biotechniques 47, 1023–1028.

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

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-scale copy number polymorphism in the human genome. Science 305, 525–528.

Sellner, L.N., Taylor, G.R., 2004. MLPA and MAPH: new techniques for detection of gene deletions. Hum. Mutat. 23, 413–419.

Serizawa, R.R., Ralfkiaer, U., Dahl, C., Lam, G.W., Hansen, A.B., Steven, K., Horn, T., Guldberg, P., 2010. Custom-designed MLPA using multiple short synthetic probes: application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. J. Mol. Diagn. 12, 402–408.

Shin, G.W., Jung, S.H., Yim, S.H., Chung, B., Yeol Jung, G., Chung, Y.J., 2012. Stuffer-free multiplex ligation-dependent probe amplification based on conformation-sensitive capillary electrophoresis: a novel technology for robust multiplex determination of copy number variation. Electrophoresis 33, 3052–3061.

Stern, R.F., Robberts, R.G., Mann, K., Yau, S.C., Berg, J., Ogilvie, C.M., 2004. Multiplex ligation-dependent probe amplification using a completely synthetic probe set. Biotechniques 37, 399–405.

Urban, T.J., Wientrob, A.C., Fellay, J., Colombo, S., Shianna, K.V., Gumbs, C., Rotger, M., Pelak, K., Dang, K.K., Detels, R., Martinson, J.J., O'Brien, S.J., Letvin, N.L., McMichael, A.J., Haynes, B.F., Carrington, M., Telenti, A., Michael, N.L., Goldstein, D.B., 2009. CCL3L1 and HIV/AIDS susceptibility. Nat. Med. 15, 1110–1112.

Walker, S., Janyakhantikul, S., Armour, J.A., 2009. Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. Genomics 93, 98–103.

White, S.J., Vink, G.R., Kriek, M., Wuyts, W., Schouten, J., Bakker, B., Breuning, M.H., den Dunnen, J.T., 2004. Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. Hum. Mutat. 24, 86–92.

White, S.J., Vissers, L.E., Geurts van Kessel, A., de Menezes, R.X., Kalay, E., Lehesjoki, A.E., Giordano, P.C., van de Vosse, E., Breuning, M.H., Brunner, H.G., den Dunnen, J.T., Veltman, J.A., 2007. Variation of CNV distribution in five different ethnic populations. Cytogenet. Genome Res. 118, 19–30.

Yang, T.L., Chen, X.D., Guo, Y., Lei, S.F., Wang, J.T., Zhou, Q., Pan, F., Chen, Y., Zhang, Z.X., Dong, S.S., Xu, X.H., Yan, H., Liu, X., Qiu, C., Zhu, X.Z., Chen, T., Li, M., Zhang, H., Zhang, L., Drees, B.M., Hamilton, J.J., Papasian, C.J., Recker, R.R., Song, X.P., Cheng, J., Deng, H.W., 2008. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am. J. Hum. Genet. 83, 663–674.

Zhang, X., Muller, S., Moller, M., Huse, K., Taudien, S., Book, M., Stuber, F., Platzer, M., Groth, M., 2014. 8p23 beta-defensin copy number determination by single-locus pseudogene-based paralog ratio tests risk bias due to low-frequency sequence variations. BMC Genomics 15, 64.

Supplementary Materials for:

An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs

Malgorzata Marcinkowska-Swojak[a], Katarzyna Klonowska[a], Marek Figlerowicz[a,b], Piotr Kozlowski[a,b]*

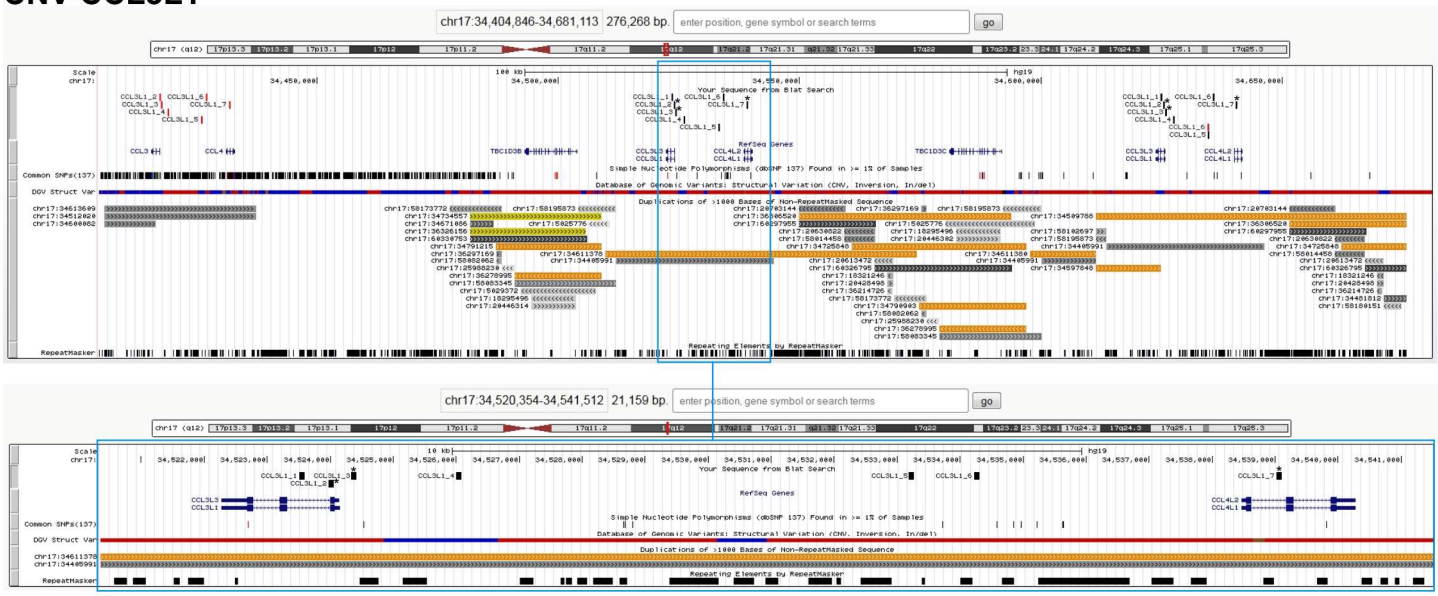
[a]European Centre of Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
[b]Poznan University of Technology, Pl. Marii Sklodowskiej-Curie 5, 60-965 Poznan, Poland

*corresponding author: Piotr Kozlowski - kozlowp@yahoo.com,

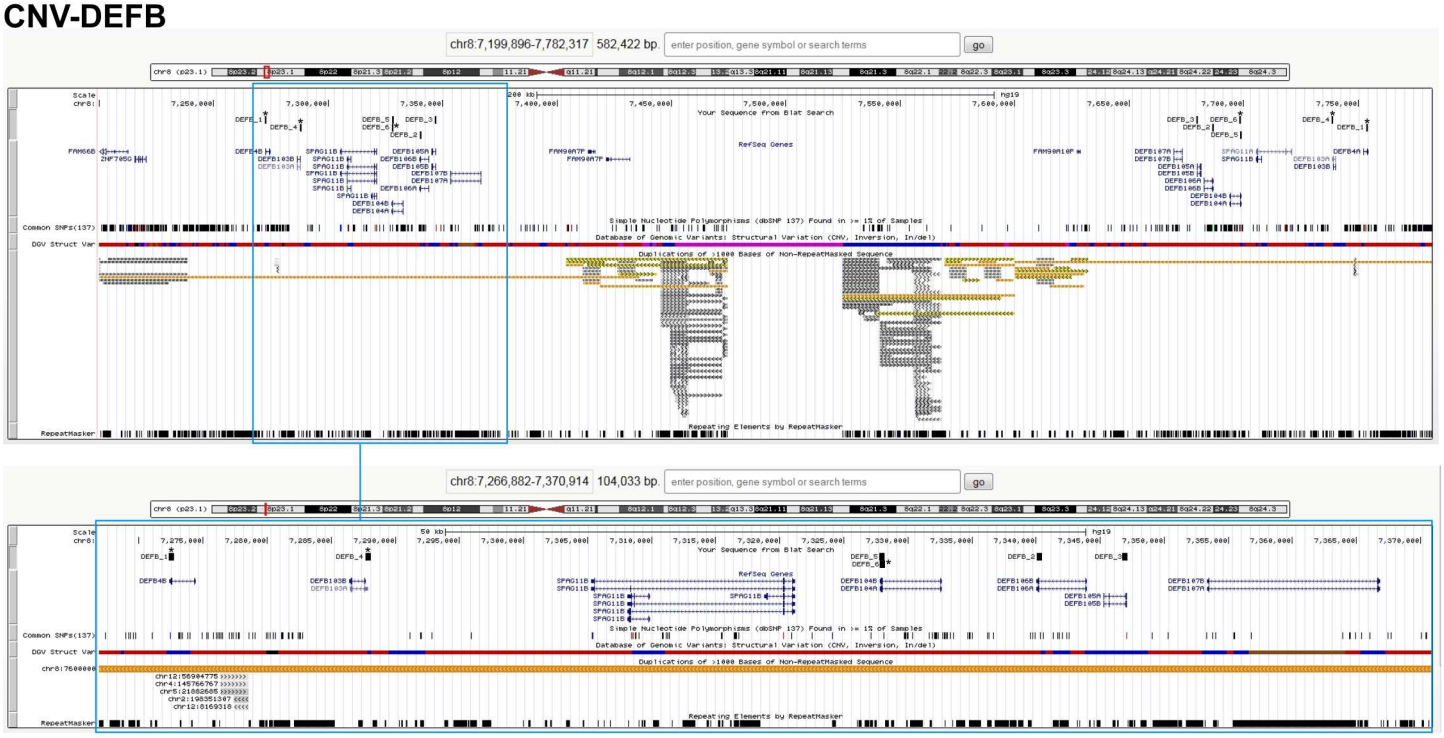
Tel: +48 616653049, Fax: +48 618520532, IBCh PAS, Noskowskiego 12/14, 61-704 Poznan, Poland

Contents:

CNV-CCL3L1



CNV-DEFB



CNV-UGT2B17

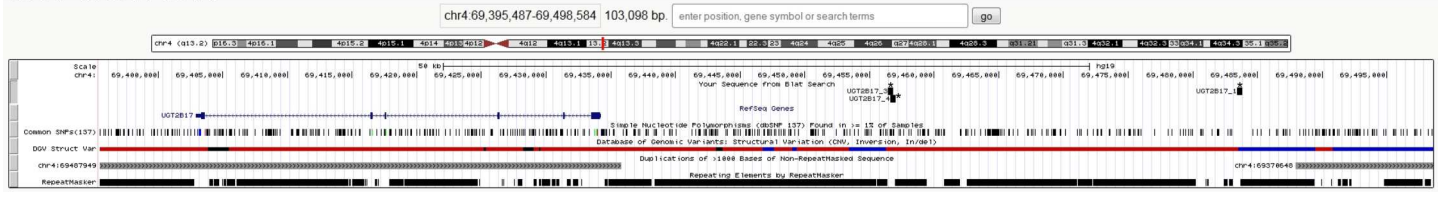


Fig.S1. Screenshots from the UCSC Genome Browser (hg19) depicting the location of the MLPA probes in three CNV regions tested in this study. The visualized UCSC tracks include the RefSeq genes, SNPs, CNV regions from DGV (dense), segmental duplications and repeat elements. The positions of the MLPA probes are indicated in the track “Your Sequence from Blat Search”. Note that in some cases, the MLPA probes map to more than one position in the reference genome. The black- and red-labeled probe positions indicate targets with perfect homology to the designed probes and targets with mismatches preventing probe ligation and subsequent amplification, respectively. The positions of the probes were intentionally selected to distinguish targets of interest from their highly homologous paralogs. The asterisks indicate probes selected for the genotyping experiments. The presented screenshots are intended for computer review, and printouts of this figure may be unreadable.

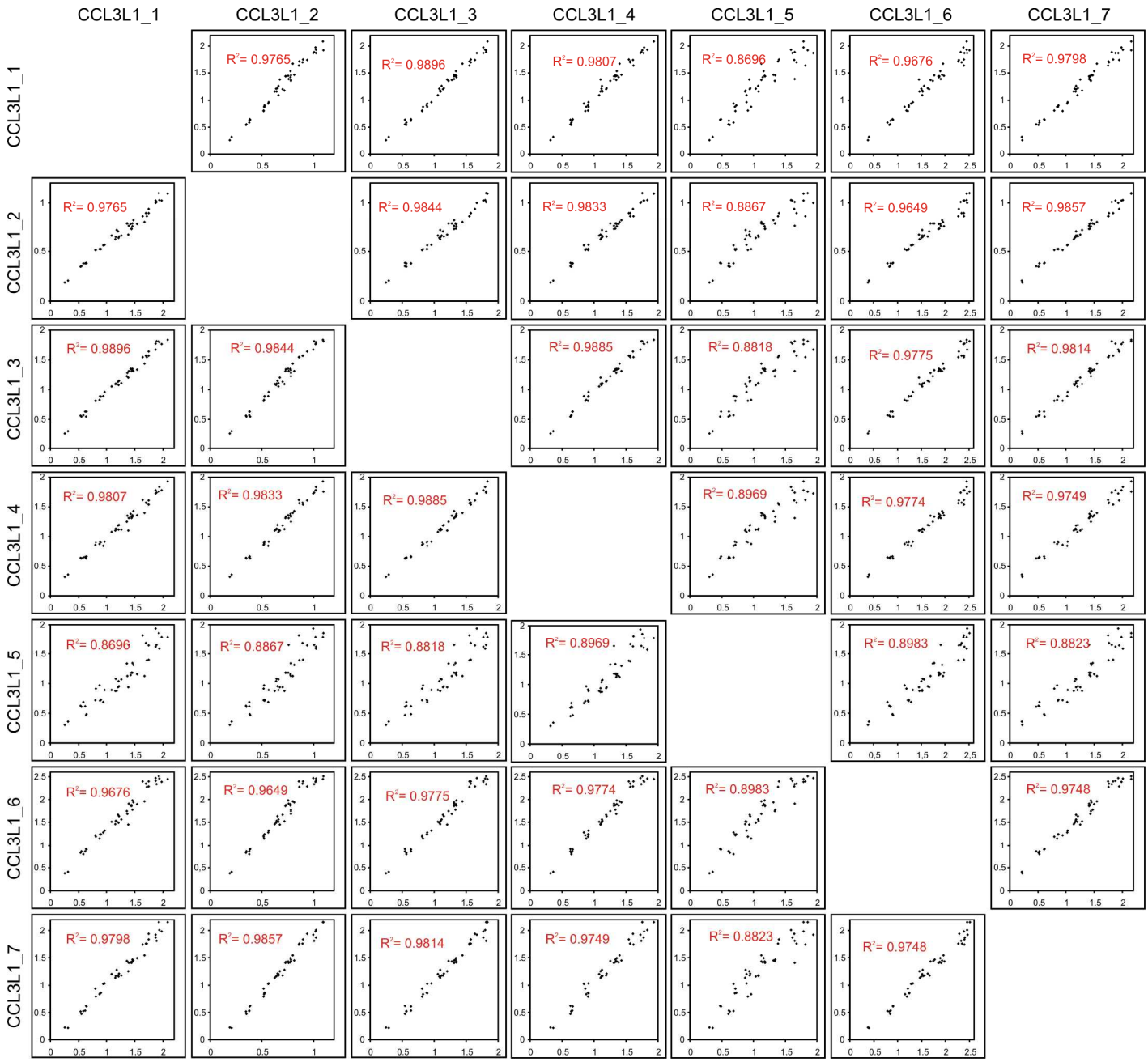


Fig.S2. The pairwise correlation analyses for the CNV-CCL3L1 probes. The probe IDs indicated over and on the left correspond to the normalized probe signals shown on x and y axes, respectively.

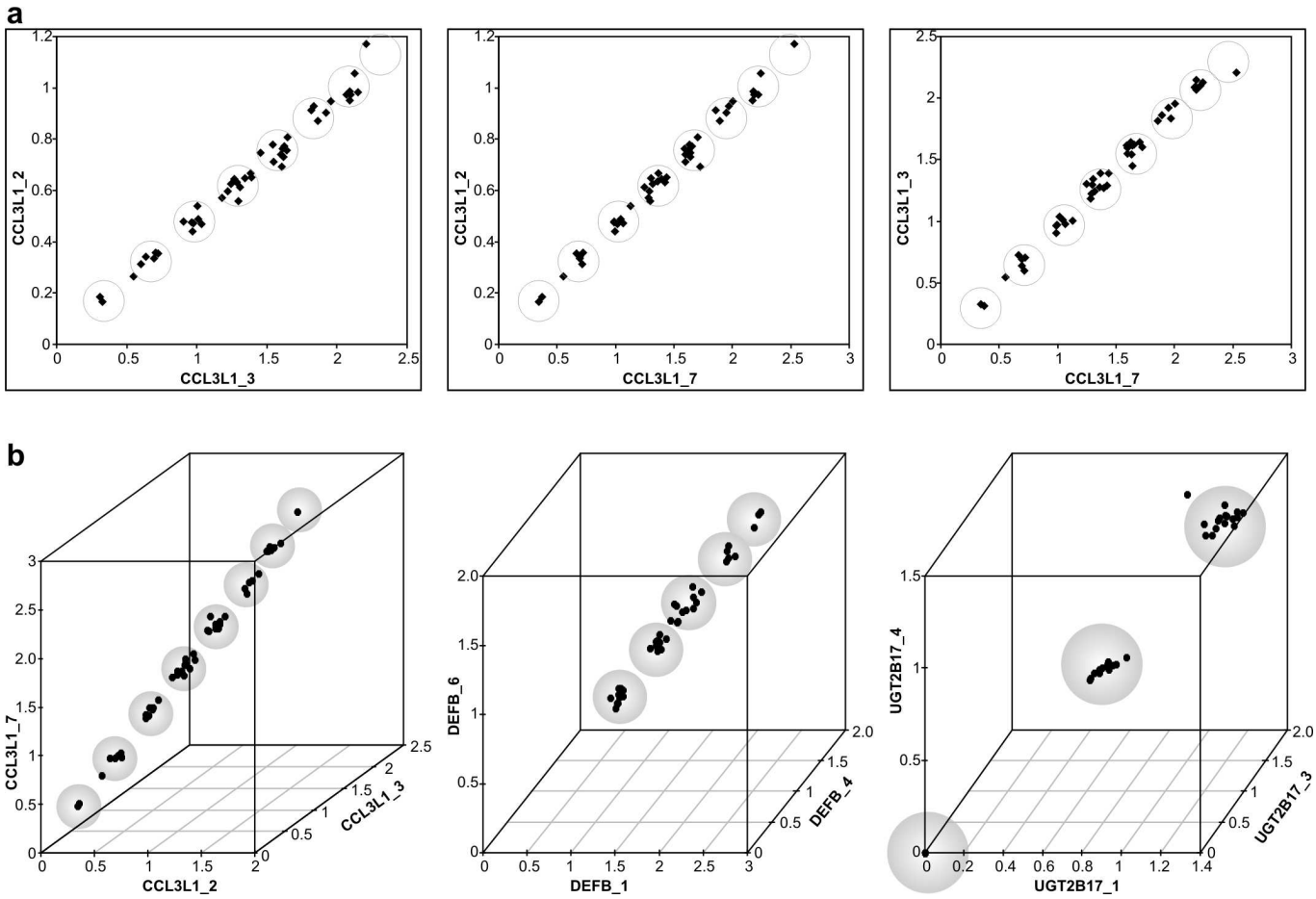
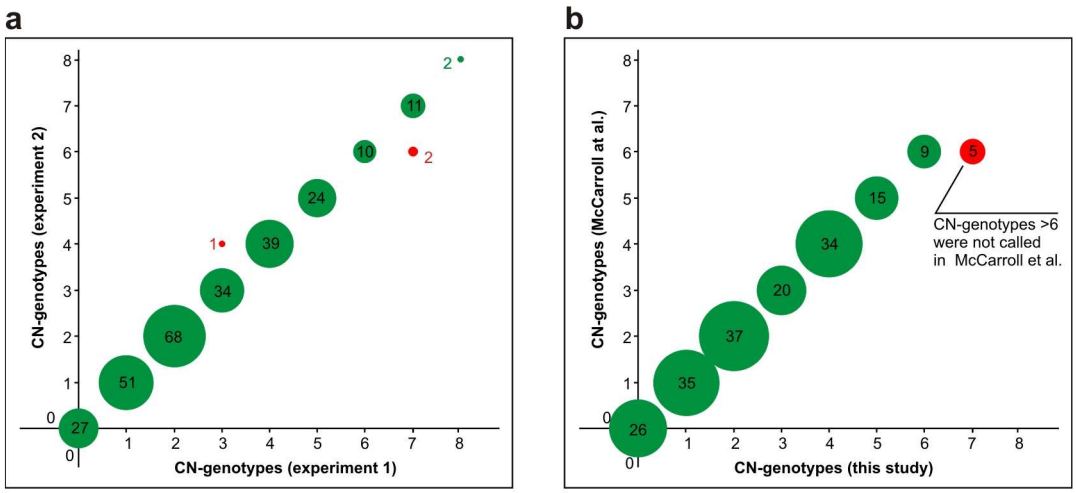
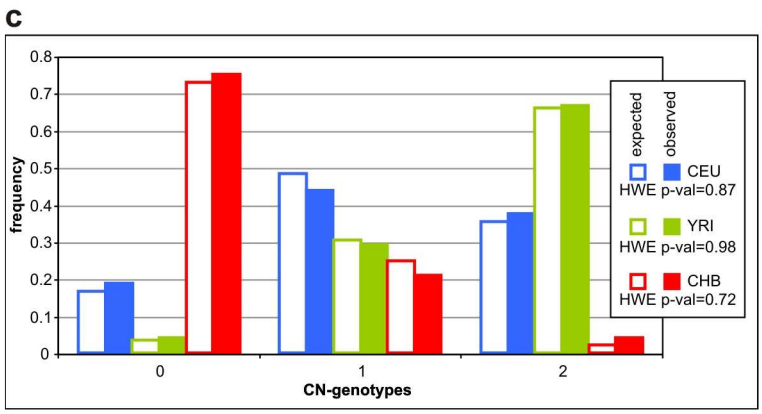


Fig.S3. The normalized probe signals on the 2D and 3D signal scatter plots. **a** The normalized signals of the three CNV-specific probes can be presented in three different complementary 2D signal scatter plots. The example presents the results of the CNV-CCL3L1 genotyping in the CHB+YRI sample set. The probe IDs are indicated along the graph axes. **b** The normalized signals of the three CNV-specific probes presented on a 3D signal scatter plot. The example presents the analyses of CNV-CCL3L1, CNV-DEFB and CNV-UGT2B17 in the CHB+YRI sample sets. The benefit from using 3D scatter plots is through the interactive version, in which each chart may be rotated to the position in which the separation of the clusters is best seen.





d

CEU trio	father	child	mother
	alleles constituting CN-genotypes		
1463	0/1	0/0	0/0
1375	0/0	0/1	1/1
1347	1/1	1/1	0/1
1454	1/1	0/1	0/0
1408	0/1	1/1	1/1

Fig.S4. The quality control analyses of the obtained genotyping results. **a** A comparison of all 268 CN-genotypes determined for all of the tested regions in the two subsequent experiments. The green and red circles indicate the concordant and discordant results, respectively, with the number of results noted in or next to the circle. **b** A comparison of 181 CN-genotypes determined in our study and previously described with the use of CNV-dedicated high-resolution microarray platform by (McCarroll et al. 2008). Note that the comparison was performed only for CNV-UGT2B17 and CNV-DEFB (the previous study did not cover CNV-CCL3L1). All of the discordant genotypes (CN=7 in our results and CN=6 in the previous study) result from the fact that in a previous study (McCarroll et al. 2008), genotypes higher than 6 CN were not called (genotypes with higher numbers were most likely rounded down to 6 copies). The color key is the same as in **a**. **c** Agreement with the Hardy-Weinberg equilibrium of the CNV-UGT2B17 genotypes for all of the populations (blue- CEU, red- CHB, green- YRI). The open and filled bars indicate the expected and the observed CNV-genotypes frequencies, respectively. **d** The representative CEU family trios depicting agreement of the observed CN-genotypes and inferred CN-alleles of CNV-UGT2B17 with the Mendelian inheritance.

Table S1. MLPA assays- detailed characteristics

CNV-CCL3L1 assay

probe ID	probe location (hg19)	chr band	probe type	5'PSS	length	5'SS	length	5'TSS	length	Tm	3'TSS	length	Tm	3'SS	length	3'PSS	length	5'HPL	3'HPL	TPL
control01	chr22:30069296-30069338	22q12.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75,6	GGCAAAACTTCTGGCCCAGAAG	22	71,0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	46	47	93
control 06	chr9:135787705-135787750	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	TCTCACCAAATCTCAGCCCGCTT	23	70,8	TCCTCATCGTTCAGCCGATGTCA	23	71,3	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	48	48	96
control09	chr16:2137895-2137936	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactac	9	CCCCGCTGGACTACGAGTGCA	21	75	ACCTGGTGTCCCTGCAGTGCA	21	72,5	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	49	50	99
control11	chr16:2166547-2166588	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	GACCCCTGACGCCCTCTGGCAC	21	76,9	AGCAGGACGGCCTCTCAGCCC	21	75,8	aatctac	7	TCTAGATTGGATCTTGCTGGCGC	23	51	51	102
control15	chr7:55223529-55223572	7p11.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	ATGTGGTGACAGATCACGGCTCG	23	71,9	TGCGTCCGAGCCTGTGGGGCC	21	80,6	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	53	52	105
control07	chr9:135777042-135777086	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	ACGAAGCTGGAGGACTGCAGGAA	23	72	CATGATTGCGGAGCTGCGGATA	22	72,0	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	54	54	108
control02	chr1:156105818-156105862	1q22	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72,8	CTGGACATCAAGCTGGCCCTG	21	72,7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	55	56	111
CCL3L1_2	chr17:34523990-34524034/ chr17:34625564-34625608	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	GACTCACGTGGTGCAGAGAGGAC	23	72,5	CTGGTTGCAGAGAGCCATGGTG	22	72,5	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	57	57	114
CCL3L1_2_K			competitor					GACTCACGTGGTGCAGAGAGGAC	23	72,5										
control18	chr17:37863311-37863352	17q12	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	CACCTCTACCAGGGCTGCCAG	21	73,1	GTGGTGCAGGGAAACCTGGAA	21	71,7	caaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	59	58	117
CCL3L1_3	chr17:34524359-34524410/ chr17:34625933-34625984	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CATGACTGGTTCTAGCTTCATGGGGTT	27	72,3	TCTCCTGTGAGTGTGAAGAGGGGTG	25	72,7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	60	60	120
CCL3L1_3_K			competitor					CATGACTGGTTCTAGCTTCATGGGGTT	27	72,3										
control05	chr2:109545794-109545837	2q12.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72,8	AGACGAGGACTACGGCTGCGTC	22	71,8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	62	124
control17	chr7:55210033-55210083	7p11.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	TTCTCAGCCTCCAGAGGATGTTCAATAA	28	70,2	CTGTGAGGTGGTCCTTGGGAATT	23	70,7	tggtcaaactaaatctac	18	TCTAGATTGGATCTTGCTGGCGC	23	64	64	128
control08	chr9:135772645-135772706	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	GGTATGAGGCTCAGAAAAGGATAACCCA	28	71,9	GGTGTTTGAATTGGAGATCTTAGATTTATATGGC	34	70,2	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	66	66	132
CCL3L1_7	chr17:34539030-34539083/ chr17:34640596-34640649	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaatt	22	CTGCTCATGTTAGGTGGGAATGGATAC	27	71,1	AAGGGACCATATTTGGGGTTCTGGTAG	27	71,1	tggtcaaactaaatctac	18	TCTAGATTGGATCTTGCTGGCGC	23	68	68	136
CCL3L1_7_K			competitor					CTGCTCATGTTAGGTGGGAATGGATAC	27	71,1										
control19	chr17:37883704-37883746	17q12	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccac	30	CCCAGCCCTCTACAGCGGTAC	21	72,1	AGTGAGGACCCCACAGTACCCC	22	71,8	tatctaatggtcaaactaaatctac	25	TCTAGATTGGATCTTGCTGGCGC	23	70	70	140
control03	chr17:3397657-3397712	17p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70,6	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70,9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	72	144
control20	chr7:116436117-116436165	7q31.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	CAGAAGATAACGCTGATGATGAGGTGG	27	71,7	ACACACGACCAGCCTCCTTCTG	22	70,7	aatgtatctaatggtcaaactaaatctac	29	TCTAGATTGGATCTTGCTGGCGC	23	74	74	148
control12	chr16:2187439-2187512	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgc	27	ATTGTTTGAAGGCGGTTGTGGAATAAAATG	30	70,9	CAATTGTTTAGAAAGGATTTCTTAAATAGAAAAGACATATGTTT	44	69,1	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	76	76	152

control10	chr16:2134982-2135025	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttc	36	CTTCTTTGGCGACGAGTCAAACA	23	69,6	AGCCAATCCTGCTGCCCAATG	21	71,0	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	78	78	156
control13	chr4:88959518-88959566	4q22.1	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttt	34	GAATACGGCAACTCCGAGTCAGAAATG	27	72,2	GATCCTGCTCTATCCCCCAGGA	22	71,5	ttgcgaaatgtatctaatggtcaaactaaatctac	35	TCTAGATTGGATCTTGCTGGCGC	23	80	80	160
control14	chr4:88967835-88967894	4q22.1	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttt	34	CTTGGCAATTTCAGCCTTTAAAGCTGATC	29	72,6	CGATATGTCACAACTTTTGATTTCTTCCTGG	31	71,4	atgtatctaatggtcaaactaaatctac	28	TCTAGATTGGATCTTGCTGGCGC	23	82	82	164
control04	chr11:14515205-14515256	11p15.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagctcgcg	44	TGCATGTTTGGAGCATCGACACA	23	70,4	GCTATGTTAGAAGAAATGCTGTTTTGGCC	29	70,5	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	86	86	172
CNV-specific probes used only in optimization set:																				
CCL3L1_4	chr17:34526013-34526062/ chr17:34627584-34627633	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	GAAATGAAAGCTCAGAACCCCATGAC	26	70,9	GAAGAAGTGACCTTCCACCAGAGCC	24	70,9	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	59	58	117
CCL3L1_4_K			competitor					GAAATGAAAGCTCAGAACCCCATGAG	26	70,9										
CCL3L1_1	chr17:34523522-34523582/ chr17:34625096-34625156	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	GATCCAGATACCTGAATGGACTGTTCTCTTAC	32	71,5	CTCTCTTCATGGAATTTTGTCGGTTCAAG	29	71,5	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	64	64	128
CCL3L1_1_K			competitor					GATCCAGATACCTGAATGGACTGTTCTCTTAG	32	71,5										
CCL3L1_5	chr17:34533209-34533257/ chr17:34634757-34634805	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattg	23	CCAGGATGTCCAGATTGACTGCCT	24	72,5	CTCTACCCTGCTCCTCAGCACTTTC	25	72,4	tggtcaaactaaatctac	18	TCTAGATTGGATCTTGCTGGCGC	23	66	66	132
CCL3L1_5_K			competitor					CCAGGATGTCCAGATTGACTGCCT	24	72,5										
CCL3L1_6	chr17:34534232-34534280/ chr17:34635780-34635828	17q12	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	CTTCTTGGCTTTCCGGTCTCCAC	23	72,5	GTTTGCCCTCCAGGAGTTCTAAATGC	26	72,2	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	70	70	140
CCL3L1_6_K			competitor					CTTCTTGGCTTTCCGGTCTCCAC	23	72,5										

CNV-DEFB assay

probe ID	probe location (hg19)	chr band	probe type	5'PSS	length	5'SS	length	5'TSS	length	Tm	3'TSS	length	Tm	3'SS	length	3'PSS	length	5'HPL	3'HPL	TPL
control01	chr22:30069296-30069338	22q12.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75,6	GGCAAAACTTCTGGCCCAGAAG	22	71,0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	46	47	93
control 06	chr9:135787705-135787750	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	TCTCACCAAATCTCAGCCCGCTT	23	70,8	TCCTCATCGTTCAGCCGATGTCA	23	71,3	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	48	48	96
control09	chr16:2137895-2137936	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactac	9	CCCCGCTGGACTACGAGTGCA	21	75	ACCTGGTGTCCCTGCAGTGCA	21	72,5	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	49	50	99
control11	chr16:2166547-2166588	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	GACCCCTGACGCCTCTGGCAC	21	76,9	AGCAGGACGGCCTCTCAGCCC	21	75,8	aatctac	7	TCTAGATTGGATCTTGCTGGCGC	23	51	51	102
control15	chr7:55223529-55223572	7p11.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	ATGTGGTGACAGATCACGGCTCG	23	71,9	TGCGTCCGAGCCTGTGGGGCC	21	80,6	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	53	52	105
control07	chr9:135777042-135777086	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	ACGAAGCTGGAGGACTGCAGGAA	23	72	CATGATTGCGGAGCTGCGGATA	22	72,0	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	54	54	108
control02	chr1:156105818-156105862	1q22	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72,8	CTGGACATCAAGCTGGCCCTG	21	72,7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	55	56	111
control16	chr7:55218989-55219034	7p11.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	CAAAAGTGTGATCCAAGCTGTCCCA	25	72	ATGGGAGCTGCTGGGGTGCAG	21	74,6	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	57	114

control18	chr17:37863311-37863352	17q12	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	CACCTCTACCAGGGCTGCCAG	21	73,1	GTGGTGCAGGGAAACCTGGAA	21	71,7	caaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	59	58	117
DEFB_1	chr8:7272500-7272551/ chr8:7754064-7754115	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	ATTTTGTTCCAGGGAGACCACAGGT	25	70,4	GCCAATTTGTTTATACCTTCTAGGGCA	27	68,6	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	60	60	120
DEFB_1_K			competitor					ATTTTGTTCCAGGGAGACCACAGGT	25	70,4										
control05	chr2:109545794-109545837	2q12.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72,8	AGACGAGGACTACGGCTGCGTC	22	71,8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	62	124
DEFB_6	chr8:7327988-7328045/ chr8:7698549-7698606	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	AACGGGCAGTCCCATAACCACATAT	25	69,9	TCTGTCCAATTCAAATTCGCTTCTCACTAGATA	33	70,4	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	64	64	128
DEFB_6_K			competitor					AACGGGCAGTCCCATAACCACATAT	25	69,9										
control08	chr9:135772645-135772706	9q34.13	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	GGTATGAGGCTCAGAAAAGGATAACCCA	28	71,9	GGTGTTTGAATTGGAGATCTTAGATTTATATGGC	34	70,2	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	66	66	132
DEFB_4	chr8:7287687-7287739/ chr8:7738857-7738909	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	GAAAAGGTGTGCTTGGTCACTTTATAAAGG	30	70,3	TTCCAGCCACAGCTGCAATTCTT	23	69,3	ctaatggtcaaactaaatctac	22	TCTAGATTGGATCTTGCTGGCGC	23	68	68	136
DEFB_4_K			competitor					GAAAAGGTGTGCTTGGTCACTTTATAAAGG	30	70,3										
control19	chr17:37883704-37883746	17q12	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaaattgatgccac	30	CCCAGCCCTCTACAGCGGTAC	21	72,1	AGTGAGGACCCCACAGTACCCC	22	71,8	tatctaatggtcaaactaaatctac	25	TCTAGATTGGATCTTGCTGGCGC	23	70	70	140
control03	chr17:3397657-3397712	17p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70,6	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70,9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	72	144
control20	chr7:116436117-116436165	7q31.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	CAGAAGATAACGCTGATGATGAGGTGG	27	71,7	ACACACGACCAGCCTCCTTCTG	22	70,7	aatgtatctaatggtcaaactaaatctac	29	TCTAGATTGGATCTTGCTGGCGC	23	74	74	148
control12	chr16:2187439-2187512	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgc	27	ATTGTTTGAAGGCGGTTGTGGAATAAAATG	30	70,9	CAATTGTTTAGAAAGGATTTCTTAAATAGAAAAGACATATGTTT	44	69,1	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	76	76	152
control10	chr16:2134982-2135025	16p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttc	36	CTTCTTTGGCGACGAGTCAAACA	23	69,6	AGCCAATCCTGCTGCCCAATG	21	71,0	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	78	78	156
control13	chr4:88959518-88959566	4q22.1	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttt	34	GAATACGGCAACTCCGAGTCAGAAATG	27	72,2	GATCCTGCTCTATCCCCCAGGA	22	71,5	ttgcgaaatgtatctaatggtcaaactaaatctac	35	TCTAGATTGGATCTTGCTGGCGC	23	80	80	160
control14	chr4:88967835-88967894	4q22.1	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttt	34	CTTGGCAATTTCAGCCTTTAAAGCTGATC	29	72,6	CGATATGTCACAACTTTTGATTTCTTCCTGG	31	71,4	atgtatctaatggtcaaactaaatctac	28	TCTAGATTGGATCTTGCTGGCGC	23	82	82	164
control04	chr11:14515205-14515256	11p15.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagctcgcg	44	TGCATGTTTGGAGCATCGACACA	23	70,4	GCTATGTTAGAAGAAATGCTGTTTTGGCC	29	70,5	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	86	86	172
CNV-specific probes used only in optimization set:																				
DEFB_2	chr8:7340131-7340179/ chr8:7686422-7686470	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CTATAATGCTCCCACATGGCTGGAT	24	70	GGTCCGACAGCATTTCAGAGACTT	25	69,9	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	57	57	114
DEFB_2_K			competitor					CTATAATGCTCCCACATGGCTGGAT	24	70										
DEFB_3	chr8:7346888-7346938/ chr8:7679665-7679715	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	CATGTTCCCCATTTCTAGACATTTCCTG	28	70,9	GAAAGCTCCCTCAGCAGCTCTTT	23	69,6	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	66	66	132

DEFB_3_K			competitor					CATGTTCCCCATTTCTAGACATTTCCTG	28	70,9										
DEFB_5	chr8:7327838-7327886/ chr8:7698708-7698756	8p23.1	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	TCCTACTTCCAGCGACTCTAGGGAC	25	70,4	CAGCACTACTGCGTTTCAGGGTTT	24	70,2	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	58	59	117
DEFB_5_K			competitor					TCCTACTTCCAGCGACTCTAGGGAC	25	70,4										

CNV-UGT2B17 assay

probe ID	probe location (hg19)	chr band	probe type	5'PSS	length	5'SS	length	5'TSS	length	Tm	3'TSS	length	Tm	3'SS	length	3'PSS	length	5'HPL	3'HPL	TPL
control01	chr22:30069296-30069338	22q12.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75,6	GGCAAAACTTCTGGCCCAGAAG	22	71,0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	46	47	93
UGT2B17_1	chr4:69483594-69483642	4q13.2	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	GGAAGGGCAGTCTCGACTCAAAC	23	71,4	CCCTCTTGCTACATTCAAAGAACTGC	26	69,7	c	1	TCTAGATTGGATCTTGCTGGCGC	23	49	50	99
UGT2B17_4	chr4:69456808-69456857	4q13.2	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	TGTGCACGATTTAGCCCAGTAGCA	24	70,1	ATTATGCCAATCTCCCATGGGTGAAC	26	70,7	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	51	51	102
UGT2B17_3	chr4:69456624-69456676	4q13.2	CNV-specific	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	TCCTCATTTGTCTTTGGATTAGTGGCC	27	70,4	CTTTACTTGTGTGCTTCGGCATGTTG	26	71,0	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	54	54	108
control02	chr1:156105818-156105862	1q22	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72,8	CTGGACATCAAGCTGGCCCTG	21	72,7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	55	56	111
control05	chr2:109545794-109545837	2q12.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72,8	AGACGAGGACTACGGCTGCGTC	22	71,8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	62	124
control03	chr17:3397657-3397712	17p13.3	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70,6	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70,9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	72	144
control04	chr11:14515205-14515256	11p15.2	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttcagctcgcg	44	TGCATGTTTGGAGCATCGACACA	23	70,4	GCTATGTTAGAAGAAATGCTGTTTTGGCC	29	70,5	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	86	86	172

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 length

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

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

Table S2. Genotypes observed in presented study.

sample ID	sex	HapMap pop.	family ID	family relation	CNV-CCL3L1		CNV-DEFB		CNV-UGT2B17	
					this study	previous study	this study	previous study (SMC)	this study	previous study (SMC)
NA12878	W	CEU	1463	daughter	1	NG	4	4	0	1
NA12892	W	CEU	1463	mother	2	NG	4	4	0	1
NA12891	M	CEU	1463	father	2	NG	4	4	1	1
NA12740	W	CEU	1444	daughter	2	NG	4	4	1	1
NA12751	W	CEU	1444	mother	2	NG	4	4	1	1
NA12750	M	CEU	1444	father	2	NG	5	5	1	2
NA10835	M	CEU	1416	son	1	NG	5	5	2	2
NA12249	W	CEU	1416	mother	3	NG	4	4	2	2
NA12248	M	CEU	1416	father	1	NG	5	5	1	0
NA10863	W	CEU	1375	daughter	1	NG	4	4	1	2
NA12234	W	CEU	1375	mother	1	NG	3	3	2	1
NA12264	M	CEU	1375	father	2	NG	4	4	0	2
NA12707	M	CEU	1358	son	2	NG	3	3	1	1
NA12717	W	CEU	1358	mother	2	NG	4	4	1	1
NA12716	M	CEU	1358	father	2	NG	2	2	0	1
NA10854	W	CEU	1349	daughter	3	NG	4	4	2	2
NA11840	W	CEU	1349	mother	3	NG	6	6	2	2
NA11839	M	CEU	1349	father	3	NG	3	3	1	1
NA10859	W	CEU	1347	daughter	3	NG	4	4	2	2
NA11882	W	CEU	1347	mother	2	NG	5	5	1	1
NA11881	M	CEU	1347	father	2	NG	5	5	2	1
NA10857	M	CEU	1346	son	3	NG	7	6	1	2
NA12044	W	CEU	1346	mother	3	NG	4	4	2	0
NA12043	M	CEU	1346	father	2	NG	5	5	1	0
NA07348	W	CEU	1345	daughter	1	NG	5	5	1	1
NA07345	W	CEU	1345	mother	1	NG	4	4	1	2
NA07357	M	CEU	1345	father	2	NG	6	6	2	2
NA10851	M	CEU	1344	son	2	NG	4	4	0	1
NA12057	W	CEU	1344	mother	2	NG	4	4	0	2
NA12056	M	CEU	1344	father	2	NG	5	5	0	0
NA12864	M	CEU	1459	son	2	NG	2	2	1	1
NA12873	W	CEU	1459	mother	2	NG	2	2	1	0
NA12872	M	CEU	1459	father	2	NG	4	4	2	1
NA12801	M	CEU	1454	son	1	NG	3	3	1	1
NA12813	W	CEU	1454	mother	1	NG	4	4	0	1
NA12812	M	CEU	1454	father	2	NG	4	4	2	1
NA12753	W	CEU	1447	daughter	2	NG	5	N	2	2
NA12763	W	CEU	1447	mother	2	NG	7	6	2	2
NA12762	M	CEU	1447	father	1	NG	3	3	2	2
NA10830	M	CEU	1408	son	1	NG	4	4	2	1
NA12236	W	CEU	1408	mother	1	NG	5	N	2	2
NA12154	M	CEU	1408	father	2	NG	3	3	1	0
NA10861	W	CEU	1362	daughter	3	NG	3	3	1	1
NA11995	W	CEU	1362	mother	3	NG	2	2	1	2
NA11994	M	CEU	1362	father	2	NG	5	5	2	1
NA07048	M	CEU	1341	son	1	NG	4	4	1	0
NA07055	W	CEU	1341	mother	3	NG	5	5	1	1
NA07034	M	CEU	1341	father	1	NG	4	4	1	0
NA18526	W	CHB	-	unrelated	3	NG	3	3	0	0
NA18529	W	CHB	-	unrelated	8	NG	2	2	1	1
NA18532	W	CHB	-	unrelated	5	NG	3	3	0	0
NA18537	W	CHB	-	unrelated	5	NG	4	4	0	0
NA18540	W	CHB	-	unrelated	4	NG	N	N	0	N
NA18542	W	CHB	-	unrelated	4	NG	6	6	1	1
NA18561	M	CHB	-	unrelated	5	NG	4	4	0	0
NA18562	M	CHB	-	unrelated	4	NG	6	6	0	0
NA18563	M	CHB	-	unrelated	7	NG	6	6	0	0
NA18570	W	CHB	-	unrelated	3	NG	6	6	0	0
NA18571	W	CHB	-	unrelated	5	NG	6	6	0	0
NA18572	M	CHB	-	unrelated	4	NG	4	4	0	0
NA18573	W	CHB	-	unrelated	1	NG	5	N	1	1

NA18576	W	CHB	-	unrelated	2	NG	3	3	0	0
NA18577	W	CHB	-	unrelated	6	NG	4	4	1	1
NA18579	W	CHB	-	unrelated	3	NG	5	5	0	0
NA18603	M	CHB	-	unrelated	4	NG	5	N	0	0
NA18605	M	CHB	-	unrelated	5	NG	4	4	0	0
NA18612	M	CHB	-	unrelated	4	NG	3	3	1	1
NA18620	M	CHB	-	unrelated	5	NG	4	4	0	0
NA18621	M	CHB	-	unrelated	2	NG	4	4	2	2
NA18622	M	CHB	-	unrelated	4	NG	3	3	0	0
NA18623	M	CHB	-	unrelated	2	NG	5	N	0	0
NA18624	M	CHB	-	unrelated	1	NG	4	4	0	0
NA18501	M	YRI	Y004	unrelated	3	NG	4	4	2	2
NA18507	M	YRI	Y009	unrelated	5	NG	3	3	2	2
NA18516	M	YRI	Y013	unrelated	6	NG	3	3	2	2
NA18522	M	YRI	Y016	unrelated	7	NG	4	4	2	N
NA18856	M	YRI	Y023	unrelated	4	NG	4	4	1	1
NA18859	M	YRI	Y012	unrelated	5	NG	3	3	2	2
NA18863	M	YRI	Y024	unrelated	6	NG	7	6	1	1
NA19100	W	YRI	Y105	unrelated	6	NG	5	N	1	1
NA19103	M	YRI	Y042	unrelated	3	NG	4	4	2	2
NA19127	W	YRI	Y077	unrelated	2	NG	4	4	2	2
NA19132	W	YRI	Y101	unrelated	5	NG	5	5	2	2
NA19138	M	YRI	Y043	unrelated	4	NG	5	5	2	2
NA19140	W	YRI	Y071	unrelated	7	NG	6	6	1	1
NA19141	M	YRI	Y071	unrelated	4	NG	6	6	2	2
NA19152	W	YRI	Y072	unrelated	6	NG	5	5	2	2
NA19153	M	YRI	Y072	unrelated	2	NG	7	6	2	2
NA19159	W	YRI	Y056	unrelated	2	NG	5	5	1	1
NA19172	W	YRI	Y047	unrelated	4	NG	3	3	0	0
NA19192	M	YRI	Y112	unrelated	3	NG	3	3	2	2
NA19193	W	YRI	Y112	unrelated	7	NG	3	3	2	2
NA19202	W	YRI	Y045	unrelated	7	NG	5	5	1	1
NA19204	W	YRI	Y048	unrelated	5	NG	7	6	2	2
NA19222	W	YRI	Y058	unrelated	3	NG	3	3	2	2
NA19240	W	YRI	Y117	unrelated	7	NG	3	3	1	1

LEGEND:

W – woman
M – man
CEU – European population, Utah residents with ancestry from northern and western Europe
CHB – Asiatic population, Han Chinese from Beijing, China
YRI – African population, Yoruba from Ibadan, Nigeria
NG – not genotyped in the study
SMC – genotyped in previous study by McCarroll et al. 2008 (ref.)
Red font indicates discordant genotypes

3

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, Irminger-Finger I, Limon J, 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

SCIENTIFIC REPORTS

OPEN

Analysis of large mutations in *BARD1* in patients with breast and/or ovarian cancer: the Polish population as an example

Received: 24 September 2014
Accepted: 13 April 2015
Published: 21 May 2015

Katarzyna Klonowska[1], Magdalena Ratajska[2], Karol Czubak[1], Alina Kuzniacka[2], Izabela Brozek[2], Magdalena Koczkowska[2], Marcin Sniadecki[2], Jaroslaw Debniak[2], Dariusz Wydra[2], Magdalena Balut[2], Maciej Stukan[3], Agnieszka Zmienko[1,4], Beata Nowakowska[5], Irmgard Irminger-Finger[6], Janusz Limon[2] & Piotr Kozlowski[1,4]

Only approximately 50% of all familial breast cancers can be explained by known genetic factors, including mutations in *BRCA1* and *BRCA2*. One of the most extensively studied candidates for breast and/or ovarian cancer susceptibility is *BARD1*. Although it was suggested that large mutations may contribute substantially to the deleterious variants of *BARD1*, no systematic study of the large mutations in *BARD1* has been performed. To further elucidate the role of large mutations in *BARD1*, we designed a multiplex ligation-dependent probe amplification (MLPA) assay and performed an analysis of 504 women with a familial breast and/or ovarian cancer and 313 patients with ovarian cancer. The investigation did not reveal any large mutations in the *BARD1* gene. Although the analysis was not focused on identification of small mutations, we detected seven deleterious or potentially deleterious point mutations, which contribute substantially to the total number of *BARD1* mutations detected so far. In conclusion, although we cannot exclude the presence of large mutations in *BARD1*, our study indicates that such mutations do not contribute substantially to the risk of breast and/or ovarian cancer. However, it has to be noted that our results may be specific to the Polish population.

Five to ten percent of all breast cancer (MIM#114480) cases are inherited and consequently aggregate in families. Hereditary breast cancer, on average, is diagnosed in a young age and/or co-occurs with ovarian cancer (MIM#167000). It is estimated that germline mutations affecting the highly susceptible *BRCA1* (MIM*113705) and *BRCA2* (MIM*600185) genes explain 16–40% of all familial breast cancer cases[1]. Moreover, highly penetrant mutations in genes such as *TP53* (MIM*191170), *STK11* (MIM*602216), *CDH1* (MIM*192090), and *PTEN* (MIM + 601728) are associated with various hereditary cancer syndromes and account for approximately 1% of all breast cancer cases that aggregate in families[2]. Another ~5% of familial breast cancers may be explained by mutations in moderately susceptible genes such as *ATM* (MIM*607585), *CHEK2* (MIM + 604373), *NBN* (MIM*602667), *RAD50* (MIM*604040), *RAD51B* (MIM*602948), and *RAD51D* (MIM*602954) and the genes implicated in Fanconi anemia. Finally, a

[1]European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland. [2]Medical University of Gdansk, Marii Sklodowskiej-Curie 3a, 80-211 Gdansk, Poland. [3]Gdynia Oncology Centre, Powstania Styczniowego 1, 81-519 Gdynia, Poland. [4]Poznan University of Technology, Pl. Marii Sklodowskiej-Curie 5, 60-965 Poznan, Poland. [5]Department of Medical Genetics, Institute of Mother and Child, Kasprzaka 17A, 01-211 Warsaw, Poland. [6]University Hospitals of Geneva, HUG, Geneva, Switzerland. Correspondence and requests for materials should be addressed to P.K. (email: kozlowp@yahoo.com)

significant proportion of breast cancer aggregation may result from the cooperative effect of common polymorphisms (primarily SNPs) or by their interaction with highly susceptible genes. Recently, a large, cooperative genome-wide association study identified the association of 67 new and previously reported SNPs with breast cancer[3]. It was estimated that these SNPs cumulatively explain 14% of the familial heritability of breast cancer, and a further 9% may be explained by yet unknown common SNPs[3]. Altogether, only approximately 50% of all familial breast cancer cases can be explained by known genetic factors[1,4]. Therefore, the identification of additional cancer-susceptibility genes is highly sought-after[1,2,4,5].

Among the candidate breast and/or ovarian cancer susceptibility genes are those encoding proteins that interact with BRCA1/2 in DNA damage response and other tumor suppressor pathways[2,4]. One such gene that has been intensively studied is *BARD1* (BRCA1 associated RING domain 1; MIM #601593). *BARD1* is composed of 11 exons spread out over an 85-kb region at 2q35 and encodes a 777-amino-acid protein that shares both structural and functional similarities with BRCA1. Both proteins possess an amino-terminal RING-finger motif that facilitates BARD1/BRCA1 heterodimer formation. This in turn stabilize both proteins and is essential for the expression of the tumor suppressor functions of BRCA1[6]. It has been shown that missense mutations in the BRCA1 RING-finger domain that hamper heterodimer formation are highly penetrant deleterious mutations.

The analysis of *BARD1* in subjects with increased genetic risk of breast and/or ovarian cancer led to the identification of dozens of sequence alterations, including definitively damaging, frameshift and nonsense mutations[7–30]. Detected mutations are distributed over almost entire sequence of *BARD1* and no strong hot-spot mutation or hot-spot region was identified so far. The exception may be c.1670G>C (p.Cys557Ser) which is relatively frequent in European populations however its role in breast cancer predisposition is equivocal[7,10–13,15–17,19,22,23,25]. Depending on a type of tested samples and a criteria of mutation definition, the *BARD1* mutation rate (ratio of the number of mutations and the number of analyzed patients) varies between 2.8% and 6.1%[10,19,23]. Recently, highly deleterious *BARD1* mutations were also detected with the use of exome sequencing of cancer predisposing genes[21,26–30]. These more objective approaches recurrently show that *BARD1* belongs to the group of the most frequently mutated genes, after *BRCA1* and *BRCA2*. For example, recent analysis of 1824 patients with triple-negative breast cancer unselected for family history led to the identification of 9 definitive *BARD1* mutations (more mutations was identified only in *BRCA1*, *BRCA2* and *PALB2*)[29]. It was shown that a number of the identified point mutations in *BARD1* co-segregate in families with cancer[10,19]. *BARD1* small-size mutations were also analyzed as modifiers of *BRCA1*/*BRCA2* attributed risk[31–33].

However, the knowledge of large mutations in the *BARD1* gene is still very limited. One of such alterations is a deletion of 1260-bp in intron 3 of *BARD1*[24]. Additionally, a germline deletion of the entire *BARD1* gene was detected in a non-*BRCA* patient with triple-negative breast cancer. Consequently, it has been suggested that large mutations (multi-exon deletions or insertions) in *BARD1* as well as in other breast cancer susceptibility genes may substantially contribute to familial breast/ovarian cancer risk[20,34,35]. This corresponds with the previous studies demonstrating that large rearrangements may account for a substantial fraction of all of the disease-related mutations in a particular gene. Normally this fraction accounts for ~5% of all detected mutations; however, it is strongly dependent on both the gene and population background, and in some cases this number well exceeds 10%. However, to our knowledge, no systematic analysis focusing solely on large germline mutations in *BARD1* has been performed. Therefore, to unequivocally elucidate this issue, we performed a comprehensive analysis of the large mutations in *BARD1* in over 800 samples with either familial breast cancer or unselected ovarian cancer.

Results and Discussion

MLPA is the method of choice for the detection of large mutations; however, commercial MLPA assays are available only for a limited number of the most intensively studied genes, and there is no assay for *BARD1*. Therefore, as a first step, we designed and generated a new MLPA assay covering all 11 exons (12 probes; one probe in each exon, two probes in exon 4) as well as the 5'- and 3'-flanking sequences (2 probes) of *BARD1* (Fig. 1). Additionally our assay was comprised of 3 control probes (located in copy-number-stable regions in chromosomes 1, 17, and 22) and 3 probes located in *ARID1A* (MIM*603024) (not used in this study). To prove the dosage-sensitivity of the designed MLPA probes, we performed an analysis of three types of positive control samples: (i) anonymous control sample with the large-scale duplication of the 2q34-37 region in which *BARD1* is located; (ii) DNA control sample, digested with *Hind*III, and mixed (1:1) with undigested sample to simulate heterozygous deletion (*Hind*III cuts target sequences of two consecutive probes, BARD1_e03 and BARD1_e04.1, and does not cut target sequences of any other MLPA probes used in the assay); and (iii) control sample, in which target sequences of two consecutive probes, BARD1_e06 and BARD1_e07, were masked by specific masking-oligonucleotides, complementary to target sequences of the selected probes (upon hybridization, the masking-oligonucleotides, prevent target recognition and subsequent ligation of MLPA probes). All the tests confirmed the dosage-sensitivity of the designed MLPA assay/probes (Fig. 1).

The designed MLPA assay was used for *BARD1* large-mutation analysis of 504 patients from families with breast and/or ovarian cancer aggregation and 313 patients with unselected ovarian cancer. The conducted analysis did not show any MLPA patterns indicating the presence of a large mutation in the analyzed samples. However, in seven samples (one unselected ovarian and six familial breast cancer cases), we observed a 28–45% reduction of the individual probe signal: two samples with a reduced signal

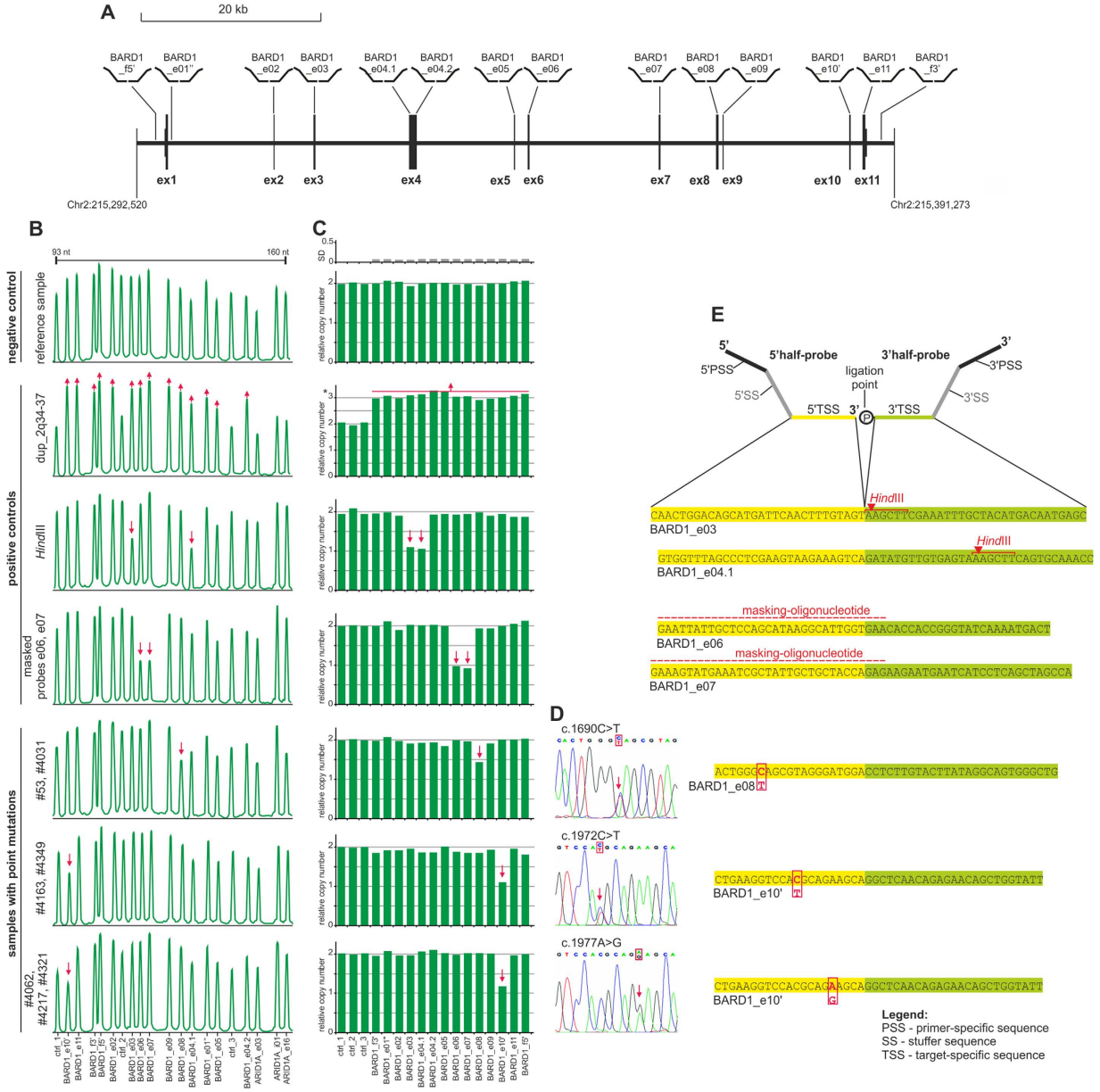


Figure 1. Analysis of the large mutations in *BARD1*, conducted with the use of a homemade MLPA assay. (**a**) A schematic map of the *BARD1* gene and the flanking genomic regions, with the positions and IDs of the MLPA probes indicated. The exons are presented as vertical rectangles with proportional size and spacing based on the NM_000465 *BARD1* sequence (reverse complement) retrieved from the UCSC Genome Browser (human genome reference sequence Mar 2006 NCBI36/hg18 assembly). The upper and lower rectangles correspond to the protein coding and untranslated sequences, respectively. In panels (**b–e**) there are representative results of the following control samples and samples with different point mutations (from the top): (i) the representative negative result without any mutations; (ii) the positive control sample with duplication of 2q34-37 in which *BARD1* is located; (iii) artificial positive control sample, composed of 1:1 mixture of *Hind*III digested and undigested genomic DNA sample; (iv) artificial positive control sample, generated by masking the target sequences of BARD1_e06 and BARD1_e07 probes with probe-specific masking-oligonucleotides; (v) samples #53 and #4031 with the mutation c.1690C>T in exon 8; (vi) samples #4163 and #4349 with the mutation c.1972C>T in exon 10 and (vii) samples #4062, #4217, and #4321 with the mutation c.1977A>G in exon 10. (**b**) The MLPA electropherograms of the representative MLPA results. The probe IDs are shown under the electropherograms. An arrowhead indicates a reduced signal of the MLPA probe. (**c**) The bar plots (corresponding to the electropherograms shown in panel b) representing the normalized copy number value (y-axis) of each probe (x-axis). The gray bar plot (above) indicates standard deviation values (SD; ranged between 0.066 and 0.086) of test MLPA probes, calculated based on signal variation of particular probes in each analyzed sample (except the samples with mutation). (**d**) Sequencing results of the exons showing a reduced signal in the MLPA analysis. (**e**) The target sequences of the affected probes. A schematic representation of the MLPA probe is shown above (for details see[37,41]). The 5'- and 3'-target sequences are indicated in yellow and green, respectively. The positions of *Hind*III sites, the masking-oligonucleotides and corresponding mutations are indicated in red.

sample ID	sample type	type of family	nucleotide change	canonical AA translation of nt change	predicted effect of the mutation
#53	unselected ovarian	Br/Ov	c.1690C>T	p.Gln564*	**deleterious nonsense mutation**[23,27]
#4031	familial	Br	c.1690C>T	p.Gln564*	
#4163	familial	Br/Ov	c.1972C>T	p.Arg658Cys	**missense mutation**, described either as deleterious, potentially deleterious or neutral[7,10,12,19,38] PANTHER: change in conserved AA, score -3.030/-10; PolyPhen2: probably damaging, score 0.995/1
#4349	familial	Br	c.1972C>T	p.Arg658Cys	
#4062	familial	Br/Ov	c.1977A>G	p.Arg659Arg	**deleterious splice mutation** (exons 2-9 deletion; p.Cys53_Trp635delinsfs*12)[23]
#4217	familial	Br/Ov	c.1977A>G	p.Arg659Arg	
#4321	familial	Br	c.1977A>G	p.Arg659Arg	

Table 1. The point mutations detected in this study in breast and/or ovarian cancer susceptible patients. Br – site specific breast cancer family, Br/Ov – breast and ovarian cancer family; The variation sites are defined based on NM_000465 *BARD1* sequence.

in exon 8 and five samples with a reduced signal in exon 10 (Fig. 1). None of the rest of the analyzed samples had a MLPA probe signal reduced or increased by more than 10%.

As large heterozygous deletions lead to an approximately 50% signal reduction and commonly affect subsequent MLPA probes, we assumed that the observed reductions of the single-exon signals may have resulted from small-size sequence variants present in the target sequences of the corresponding probes. It was previously shown that such sequence variants may affect probe hybridization and/or ligation and, in consequence, lead to a relative signal reduction[36,37]. In all cases, the sequence analysis revealed heterozygous single nucleotide substitutions located at different distances (3–15 nucleotides) from the ligation point of the MLPA probes. In both samples that had the reduced signal in exon 8, we found the nonsense mutation c.1690C>T (p.Gln564*) located 15-nucleotides downstream of the probe ligation point, while in two of the samples with the reduced signal in exon 10 we found the missense mutation c.1972C>T (p.Arg658Cys) located 10-nucleotides downstream of the probe ligation point. In the three remaining samples with the reduced signal in exon 10 we found the silent mutation c.1977A>G (p.Arg659Arg) located 5-nucleotides downstream of the probe ligation point. Two of the mutations, c.1690C>T and c.1977A>G, were previously reported as definitely pathological. Although c.1977A>G is an apparently silent mutation, it affects several exonic splicing enhancer (ESE) motifs, resulting in the deletion of exons 2–9 and leading to a frameshift and the premature termination of translation (p.Cys53_Trp635delinsfs*12)[23]. The third mutation, c.1972C>T, was reported either as a potentially pathological or as an unclassified variant (Table 1)[7,10,12,19]. It was also shown that this rare sequence variant (<1%) is a risk allele associated with lung cancer (OR = 1.55)[38]. Further computational analysis of the potential functional consequences of this mutation showed that it causes substitution of a very-conserved arginine in position 658 (e.g., PANTHER, http://www.pantherdb.org/) and has a highly deleterious and destabilizing effect on protein structure (e.g., PolyPhen2, http://genetics.bwh.harvard.edu/pph2/) (Table 1). The high resolution melting (HRM) screening of c.1977A>G and c.1972C>T in a panel of 1000 unselected control samples led to the identification of one and three cases with these mutations, respectively (Ratajska M *et al.* unpublished).

Six of the identified point mutations were detected in patients with familial breast cancer, and one mutation, c.1690C>T, was detected in patient #53, who was originally enrolled as a patient with unselected ovarian cancer (Table 1). However, further analysis of the family of patient #53 revealed an aggregation of the disease within the family (Supplementary Materials: Supplementary Fig. S1 online).

Although all three of the detected single nucleotide substitutions are either deleterious or potentially deleterious mutations, the detailed explanation of their role in the predisposition of breast and/or ovarian cancer requires further functional and epidemiological analysis, which was not the subject of this study. Nonetheless, the identification of these single nucleotide variants with the use of an assay that is dedicated to the detection of large heterozygous mutations makes it highly unlikely that the lack detection of large mutations was due to false-negative errors.

An additional result of our study is the production of a homemade MLPA assay that can be used in any further analyses of large mutations (both germline and somatic) in *BARD1* in both breast/ovarian cancer as well as in other types of cancer. Our analysis was conducted on large number of samples, which further helped to prove the robustness and high reliability of this test.

Conclusions

In summary, our study, conducted on a group of 817 patients, did not lead to the detection of any large mutations in *BARD1*. Although we cannot exclude the presence of such mutations in *BARD1*, our

results clearly indicate that these mutations do not contribute substantially (>>10% of the total *BARD1* mutations) to *BARD1* sequence variation and, subsequently, to familial breast and/or ovarian cancer aggregation. However, it does not deny the role of *BARD1* as the breast cancer susceptibility gene. It has to be also noted that our results may be specific to the Polish population.

Methods

The study comprises DNA samples (extracted from whole blood) from 504-non-BRCA patients (tested for the 5 most common *BRCA1* mutations in the Polish population, c.68_69delAG, c.181T>G, c.3700_3704del5, c.4034delA, c.5266dupC, cumulatively accounting for >90% of all *BRCA* mutations[39]) from families with breast and/or ovarian cancer aggregation (as previously defined[40]) and 313 patients with ovarian cancer that was unselected in terms of the familial history of the disease. The patients' blood samples were collected between 1999 and 2012. Informed consent was obtained from all of the patients, and the study was approved by the medical review board of Medical University of Gdansk (NKEBN/399/2011-2012). The methods were carried out in accordance with the approved guidelines.

The multiplex ligation-dependent probe amplification (MLPA) probes and the probe-set layout were designed according to a previously proposed and well validated strategy[37,41]. This strategy exclusively utilizes short oligonucleotide probes that can easily be generated via standard chemical synthesis. The sequences and detailed characteristics of all of the probes as well as their exact position in the *BARD1* sequence are shown in Supplementary Materials: Supplementary Table S1 and Supplementary Data online, respectively. The MLPA analysis was performed with the use of a homemade *BARD1* assay (combined with reagents purchased from MRC-Holland, Amsterdam, The Netherlands), according to general recommendations published in previous studies[37,42]. The products of the MLPA reactions were diluted 20× in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated by size with capillary electrophoresis (POP7 polymer; ABI Prism 3130XL apparatus; Applied Biosystems, Carlsbad, CA, USA). The obtained electropherograms were analyzed using GeneMarker software (version 2.2.0; SoftGenetics, State College, PA, USA). The normalized signal of each probe (peak height divided by the average peak height of the control probes) was divided by the corresponding signal of a reference probe and multiplied by 2. The obtained values that correspond to the copy number of particular exons/regions were visualized in bar graphs. The analysis of samples with detected aberrant MLPA pattern (with mutations) was repeated at least two times.

The *Hind*III digested positive control sample was generated as follows; 1 μg of genomic DNA was incubated overnight with 20 U of *Hind*III in conditions recommended by manufacturer (Thermo Fisher Scientific, Lafayette, CO, USA) and then mixed with equal amount of undigested DNA. To generate artificial control sample with masked target sequences of the BARD1_e06 and BARD1_e07 probes, 7 fmol of each of two masking-oligonucleotides: GAA TTA TTG CTC CAG CAT AAG GCA TTG GTG AA (specific to BARD1_e06) and GAA AGT ATG AAA TCG CTA TTG CTG CTA CCA GAG (specific to BARD1_e07) were added to the MLPA reaction together with the MLPA probe mix at the hybridization step.

The mutation sequencing was performed on ABI Prism 3130 genetic analyzer; Applied Biosystems, Carlsbad, CA, USA, according to the manufacturer's general recommendations.

References

1. Melchor, L. & Benitez, J. The complex genetic landscape of familial breast cancer. *Hum. Genet.* **132,** 845–863 (2013).
2. Foulkes, W. D. Inherited susceptibility to common cancers. *New Engl. J. Med.* **359,** 2143–2153 (2008).
3. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45,** 353–361, 361e351-352 (2013).
4. Walsh, T. & King, M. C. Ten genes for inherited breast cancer. *Cancer Cell* **11,** 103–105 (2007).
5. Eccles, S. A. *et al.* Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. *Breast Cancer Res.* **15,** R92 (2013).
6. Wu, L. C. *et al.* Identification of a RING protein that can interact *in vivo* with the BRCA1 gene product. *Nat. Genet.* **14,** 430–440 (1996).
7. Thai, T. H. *et al.* Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. *Hum. Mol. Genet.* **7,** 195–202 (1998).
8. Ghimenti, C. *et al.* Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Gene. Chromosome. Canc.* **33,** 235–242 (2002).
9. Ishitobi, M. *et al.* Mutational analysis of BARD1 in familial breast cancer patients in Japan. *Cancer Lett.* **200,** 1–7 (2003).
10. Karppinen, S. M., Heikkinen, K., Rapakko, K. & Winqvist, R. Mutation screening of the BARD1 gene: evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. *J. Med. Genet.* **41,** e114 (2004).
11. Sauer, M. K. & Andrulis, I. L. Identification and characterization of missense alterations in the BRCA1 associated RING domain (BARD1) gene in breast and ovarian cancer. *J. Med. Genet.* **42,** 633–638 (2005).
12. Vahteristo, P. *et al.* BARD1 variants Cys557Ser and Val507Met in breast cancer predisposition. *Eur. J. Hum. Genet.* **14,** 167–172 (2006).
13. Karppinen, S. M. *et al.* Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 patients with cancer: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies. *J. Med. Genet.* **43,** 856–862 (2006).
14. Huo, X. *et al.* Common non-synonymous polymorphisms in the BRCA1 Associated RING Domain (BARD1) gene are associated with breast cancer susceptibility: a case-control analysis. *Breast Cancer Res. Tr.* **102,** 329–337 (2007).
15. Jakubowska, A. *et al.* BARD1 and breast cancer in Poland. *Breast Cancer Res. Tr.* **107,** 119–122 (2008).
16. Gorringe, K. L., Choong, D. Y., Visvader, J. E., Lindeman, G. J. & Campbell, I. G. BARD1 variants are not associated with breast cancer risk in Australian familial breast cancer. *Breast Cancer Res. Tr.* **111,** 505–509 (2008).

17. Johnatty, S. E. *et al.* The BARD1 Cys557Ser polymorphism and breast cancer risk: an Australian case-control and family analysis. *Breast Cancer Res. Tr.* **115,** 145–150 (2009).
18. Guenard, F. *et al.* Genetic sequence variations of BRCA1-interacting genes AURKA, BAP1, BARD1 and DHX9 in French Canadian families with high risk of breast cancer. *J. Hum. Genet.* **54,** 152–161 (2009).
19. De Brakeleer, S. *et al.* Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. *Hum. Mutat.* **31,** E1175–1185 (2010).
20. Sabatier, R. *et al.* BARD1 homozygous deletion, a possible alternative to BRCA1 mutation in basal breast cancer. *Gene. Chromosome. Canc.* **49,** 1143–1151 (2010).
21. Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *P. Natl. Acad. Sci. U S A* **108,** 18032–18037 (2011).
22. Ding, D. P. *et al.* Lack of association between BARD1 Cys557Ser variant and breast cancer risk: a meta-analysis of 11,870 cases and 7,687 controls. *J. Cancer Res. Clin.* **137,** 1463–1468 (2011).
23. Ratajska, M. *et al.* Cancer predisposing BARD1 mutations in breast-ovarian cancer families. *Breast Cancer Res. Tr.* **131,** 89–97 (2012).
24. Rouleau, E. *et al.* Rare germline large rearrangements in the BRCA1/2 genes and eight candidate genes in 472 patients with breast cancer predisposition. *Breast Cancer Res. Tr.* **133,** 1179–1190 (2012).
25. Gonzalez-Hormazabal, P. *et al.* The BARD1 Cys557Ser variant and risk of familial breast cancer in a South-American population. *Mol. Biol. Rep.* **39,** 8091–8098 (2012).
26. Castera, L. *et al.* Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. *Eur. J. Hum. Genet.* **22,** 1305–1313 (2014).
27. Cybulski, C. *et al.* Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland. *Clin. Genet.* e-pub ahead of print Oct 20; doi:10.1111/cge.12524 (2014).
28. Pennington, K. P. *et al.* Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clin. Cancer Res.* **20,** 764–775 (2014).
29. Couch, F. J. *et al.* Inherited Mutations in 17 Breast Cancer Susceptibility Genes Among a Large Triple-Negative Breast Cancer Cohort Unselected for Family History of Breast Cancer. *J. Clin. Oncol.* e-pub ahead of print Dec 1; doi:10.1200/JCO.2014.57.1414 (2014).
30. Churpek, J. E. *et al.* Inherited predisposition to breast cancer among African American women. *Breast Cancer Res. Tr.* e-pub ahead of print Nov 27; doi:10.1007/s10549-014-3195-0 (2014).
31. Stacey, S. N. *et al.* The BARD1 Cys557Ser variant and breast cancer risk in Iceland. *PLOS Med.* **3,** e217 (2006).
32. Rebbeck, T. R. *et al.* Modification of ovarian cancer risk by BRCA1/2-interacting genes in a multicenter cohort of BRCA1/2 mutation carriers. *Cancer Res.* **69,** 5801–5810 (2009).
33. Spurdle, A. B. *et al.* Common genetic variation at BARD1 is not associated with breast cancer risk in BRCA1 or BRCA2 mutation carriers. *Cancer Epidem. Biomar.* **20,** 1032–1038 (2011).
34. Krepischi, A. C. *et al.* Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res.* **14,** R24 (2012).
35. Suehiro, Y. *et al.* Germline copy number variations associated with breast cancer susceptibility in a Japanese population. *Tumor Biol.* **34,** 947–952 (2013).
36. Wimmer, K. *et al.* Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Gene. Chromosome. Canc.* **45,** 265–276 (2006).
37. Kozlowski, P., Jasinska, A. J. & Kwiatkowski, D. J. New applications and developments in the use of multiplex ligation-dependent probe amplification. *Electrophoresis* **29,** 4627–4636 (2008).
38. Rudd, M. F. *et al.* Variants in the GH-IGF axis confer susceptibility to lung cancer. *Genome Res.* **16,** 693–701 (2006).
39. Szwiec, M. *et al.* Recurrent mutations of BRCA1 and BRCA2 in Poland: an update. *Clin. Genet.* e-pub ahead of print Feb 14; doi: 10.1111/cge.12360 (2014).
40. Ratajska, M. *et al.* BRCA1 and BRCA2 point mutations and large rearrangements in breast and ovarian cancer families in Northern Poland. *Oncol. Rep.* **19,** 263–268 (2008).
41. Marcinkowska, M., Wong, K. K., Kwiatkowski, D. J. & Kozlowski, P. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. *Sci. World J.* **10,** 2003–2018 (2010).
42. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30,** e57 (2002).

Acknowledgements

This work was supported by National Science Centre 2011/01/B/NZ5/02773 and 2011/02/A/NZ2/00017, Ministry of Science and Higher Education N407 627740, and KNOW program of the Polish Ministry of Science and Higher Education.

Author Contributions

K.K. – designed MLPA assay, performed MLPA validation, analyzed MLPA test, point mutations analysis, prepared artificial positive control samples, participated in manuscript preparation; M.R. – participated in conceiving the study and manuscript preparation, coordinated samples selection and participated in DNA extraction and samples characterization; K.C. – participated in MLPA analysis; A.K. – participated in DNA extraction and samples characterization; I.B. – provided and characterized familial breast and ovarian cancer samples; M.K. – participated in samples selection and DNA isolation; M.S., J.D., D.W., M.S. – provided unselected ovarian cancer samples; M.B. – participated in mutation characterization; A.Z. – participated in MLPA analysis; B.N. – provided positive control sample; I.I.F.- participated in data interpretation and manuscript preparation; J.L. – supervised clinical samples characterization and selection, participated in conceiving the study and manuscript preparation; P.K. – supervised MLPA analysis and interpretation, participated in conceiving the study and manuscript preparation, coordinated the study. All authors read and approved the final draft.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Klonowska, K. *et al.* Analysis of large mutations in *BARD1* in patients with breast and/or ovarian cancer: the Polish population as an example. *Sci. Rep.* **5**, 10424; doi: 10.1038/srep10424 (2015).

Supplementary Materials for:

Analysis of large mutations in *BARD1* in patients with breast and/or ovarian cancer: the Polish population as an example

Katarzyna Klonowska[a], Magdalena Ratajska[c], Karol Czubak[a], Alina Kuzniacka[c], Izabela Brozek[c], Magdalena Koczkowska[c], Marcin Sniadecki[c], Jaroslaw Debniak[c], Dariusz Wydra[c], Magdalena Balut[c], Maciej Stukan[d], Agnieszka Zmienko[a,b], Beata Nowakowska[e], Irmgard Irminger-Finger[f], Janusz Limon[c] and Piotr Kozlowski[a,b]*

[a]European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland
[b]Poznan University of Technology, Pl. Marii Sklodowskiej-Curie 5, 60-965 Poznan, Poland
[c]Medical University of Gdansk, Marii Sklodowskiej-Curie 3a, 80-211 Gdansk, Poland
[d]Gdynia Oncology Centre, Powstania Styczniowego 1, 81-519 Gdynia, Poland
[e]Department of Medical Genetics, Institute of Mother and Child, Kasprzaka 17A, 01-211 Warsaw, Poland
[f]University Hospitals of Geneva, HUG, Geneva, Switzerland

*corresponding author: Piotr Kozlowski - kozlowp@yahoo.com
tel: +48 616653100, fax: +48 618520532, IChB PAS, Noskowskiego 12/14, 61-704 Poznan, Poland

Supplementary data

The *BARD1* sequence with the positions of the MLPA probes marked.

The NM_000465 *BARD1* sequence (Mar 2006 NCBI36/hg18 assembly) with the locations of the MLPA probes marked. The 5' half-probes are marked in yellow, and the 3' half-probes are marked in green. The ID of each probe is shown next to the *BARD1* sequence. The exons are indicated in blue, the low complexity/repetitive regions are labeled with lower case letters and the positions of the SNPs are indicated in red.

BARD1
>chr2:215296520-215387673 (reverse complement)

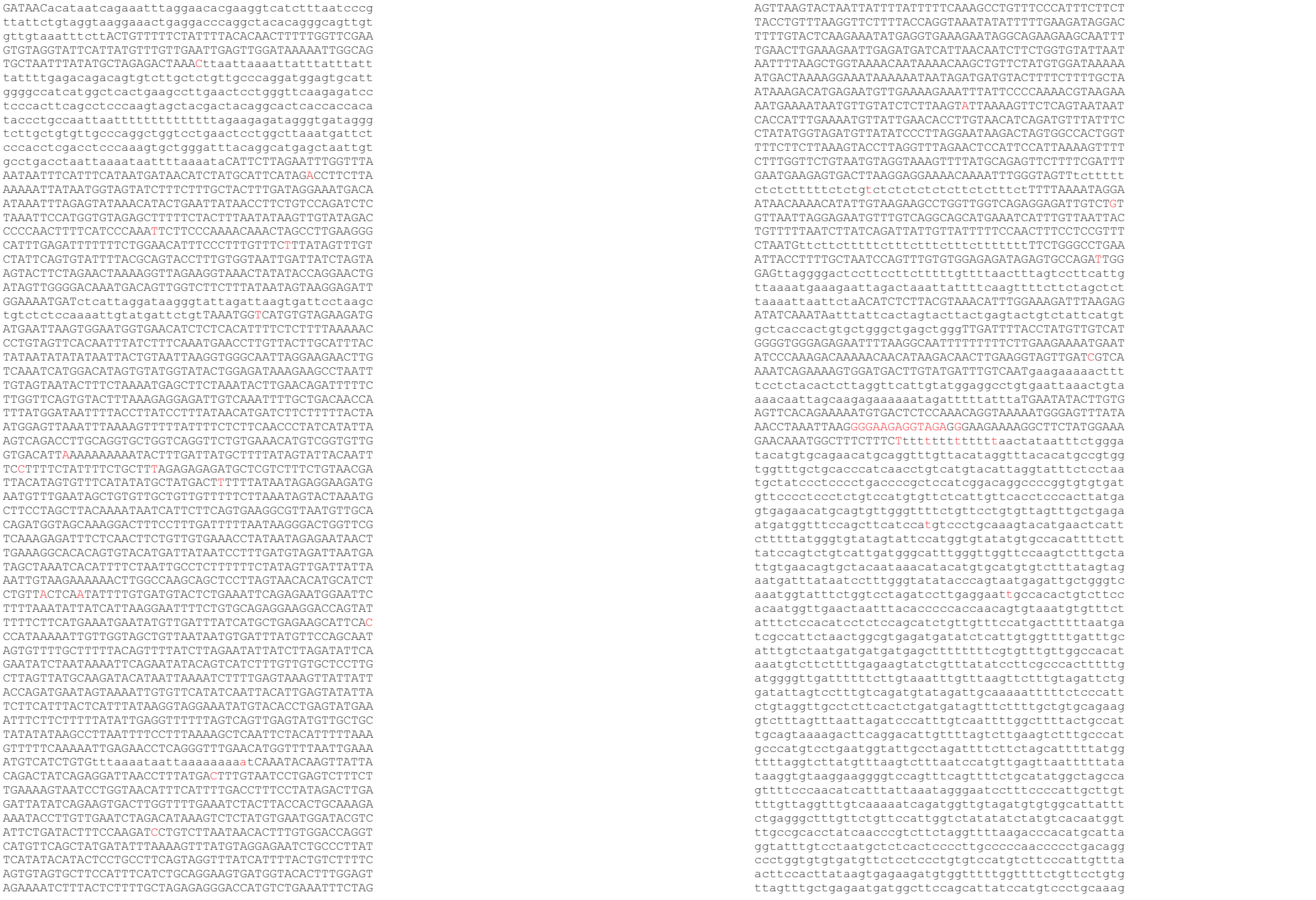
GAAGTGGCCTTTCCAGTGCATAATACCATAATTAAAAATCAGAGAAGAA
TTAATATAAAGGTAAGAATTTGTGAGATATCTTTACACTCCTTGGGAGTA
GAGTCTGGGTATTATGGATATCCCATTTTGCTTGTTATATATGTTTATCC
AAGTGAATATGAGGAAGATTTACCTCTTTGTAAAACACTTCATAACACAA
CAGAAGAGAAAACAGCATTTTGACTGTTTTCAACAAACCATTATATAAGC
CTGAAGATCAGAAAAGAATTCAAGGCTGGAGATGCCTTCATAGCATGGTA
ACACTACAGAGGGTAACTGAAGCAGTCGGGGGGATGAGGGTTTAAGTGGG
CCTGGAGAGAGGAGGATAGGTAAATAAGGCTGAAGCAAGCTTATCTGCAT
GCCTCAGAATCCACCCAGGACTACAGTGTTACAGTGGGGGAGGGATGGAT
GATTTTGAGTCTAGGATACTTGGGAATACTGCTTACTAGGAAGAGAACAG
ACTAGGAGTGAAATGGAGCAACAAGTCTAAACTTTTATGTACACAGGAGG
TCCTTGCAGGGAAAACTCTGAAAAAGAAATGATACCATTTAGCCCTTTTT
AAAGTGAGAAGAAGAGGACATTGTTCATGTCCTAAACCATGGCTTTACTG
AGACCCTGCTGATGGCTGGAGTTCCCTGCAGAACTTTTTCTCTGGGACAA
GAATTCCAGGATTTTACACATTGTTGGATGTACCCCATGGGAGAAGGAAT
GTGCCCTTGAAACTCAGGATCCAGGGTTAAGAGGAAGAGTTGGTTCAAAA
CAAACATCCCATCTCCACTAAAGGCAAAGCACCAGCAAGCTAAGCTGTAG
AAAAGCAAGGCTGGAAGGAAGAAAATGGAATGGGAAGTTGCACATCGGAG
AGTAACCCTGCTTATTCTCTTTATGAGTTTCATTATGGCCACTATTAAAC
AGCTGCCACGAATGTCATGACCCCTGGGGAGATGGAACAGGACAATGGCA
GAAGAAAGGAAAAAAGCAAACAAACAGTAGAAAGAACTAATTGTATCTGC
ATCATAATTCATTTAATAGCATGGTATGAATACGCCAGTTATAAATGTTT
GGTCCCAAATCATTTTACCAAAATGTGAAATTTTCATTTACTCTGATCAC
AGTTAAATTGTGCCccaagtatagtggcgcacgcctgtagccccagctac
tcggcagggtgaggcacgagaatcgcttgaacccttgaacgcaggaggc
agaggccacactgagctgagattgtgccactgcactccagcctgggtgac
agagcaagactctgtctcCCCCGCCCCCCCGCAAAAAAAAAACTGTGCCC
TTAATAAAAAAAAATTGATAAAAATGTTCAGTGTTATTCTTTAGTAAATA
AAGCATTAACAGTACTGATTATATATAATCATCCTCAATCagttaccatt
tactgagtacttatggtgtatgtgattttacatttaatcttatttaatag
tcctccaaatctacccgataaatattatccccattttgcagacaaggaaa
ctaatgctcagagagcttaaggcttggtttaaatcacacaattatttagt
gaaagtgggattataagtaggacaatctatctcctgaaatcctcaattta
ccccattaacaccatagtatctctcCAGGTTTTAACGTTTTAAGTTTTCG
TTCTTTTTCTTTTTCTTTCTTTTCCttttgtgttttttttttttttttt
ttttgagacagtctcgctccgttgcccaggctggagggcagcgacacca
tcttgcctcactgcaaccttcacctcctggcttccagctatttttctgcc
tcagcctcctgaggaactgggattacaggtgcatgccaccacacctggct
aatttttatatttttagtagagacggggtttcaccatgttagccaggctg
gccttgaactcctgacctcaagtgatccacctgccttggcctcccaaagt
gctgggattacaagtgtgagccaccgtgcctgaccCATTTTAAGTTTTCT
AGAACTTGATTTCTCCCACGCTCAACAAAGAGTAAGCACTGGCTTAACAG
GACTAATAAGGACTAACTTCTTAGGTGCCTCTGAGTTAAAAAAAAAAATT
GGTTCGTTTGTGTGCTATCATAGTTTCAGTTAAAACATTTAATAAAGCCT
TATTTAGTTGTGCTTATTCTAactaaccaagtgtgtgattttggccaagt
catttaaactccttgagtttTCATTATTTAGTGGAGCTACAAggcacagt
ggctcacacctgtaatcccagcattttgggaggctgaggcagaagaattg
cttgaggtcaggagttctagagcagcctggccaacatagagagaccctgt
ctctacaaataattaaaaaaaaaattatccaggcatggtggggtgcacct
gcacaggtcccagctactcgggaggctgaggtgggagggtcacttgagcc
caggagttcaaagctgcagtgagctgagatcgtaccactgtactccatcc
tgggcaacagagtgagaccttgtcccaaaaaaCCCCTCAAAAGACCAAAA
ATATAATTTACTGgggccattaataatcaaaaagatctgggttaatattc
tgccactgttgttatctgtgcctccatcctggacaagtATGCttttcttt
ttacaaatcttacttttctctaaagatgggggtcttgctatgttgcctag
gctggccttgaaccccggggctgaagtgattctcctgcctcagcctccca
agtagctaggactacaggcatgtgcTGAGCCccttgggcaagtaaattaa
gatctctgagccacacctgcctcctctTATTAAAGAGCActgtcttagct
gttataacaaaatactataatctagatggcttaagcaacagaaatttgtc
atatttctggagtctagaagtctgagattagcgtgccagtatgactgggt
tttggggagggccctcttcttttgcgttcttgctgtgtcctcatgtggtg
gagagagagagaactctctcttcttacaaggtcattaaccccatcatggg
ggctctgccctataaccccatctaaatctaactaccaccccaaaggtccc
ctccaaatactatcacgttggggttacaacttttaatatatgaatttggg
ggtgacacTACAGatactaatgcccatttcatagggtccctataaggctt
aaggcaggtattaacataggaaagcacttaaagctgggtctggcTTGGGT
AGGTAGttcaatgattcaacaaacactgagcacctacctggagccaagca
ctgcatgtgccacatgaagcgatattgggaaatgagtcacatgcagccaa
tctctggccttttggaggttttgaactagaagaggggacacgcacataatcg
tatgtgtgtgtatctacatacacaggtgatatgatgcacctgagagaatc
cagtctaggaactaggaaaaccttaaggagtgatacttcagctgtattc
tgaaggatgaggaatggagaggaggccaattccaggttccaaagtgaatc
tttgcgcaaaagccatgaggcagcaaggtgcagggggcttttaatgaccta
agggaacgcactgtggggttgggaccatgatggccagaggaggttttgac
aagaggctagtcaaagagCAGAGAAAAACTTTAAGGAGTTTTAAGAAAGG
GAAGTGCCACGATGAGTTTTGTGTTTGGAAATGTTTTAGGCGGCCACACC
GCAGTCTCGGGACTGGCTGGGACCTGGATAGACACTTGGATATCAGCTAG BARD1_f5'
AAAGCTACGACAGGAAACCAGGCGAGGACGAGGCAACTGGGGATGTTGTG
GAGCAGAGACGGAGAGAAAATGGGTGCATTCGAGACAAGTTAGGGGGAAA
AAATGCAAGGACTTGGAAATGAACTTGGGGCGCGGCAGGAAGGCATGACG
GGTTGCTCTGTAGGTCTTATCTGTAAATTACGGCGATCAGTGAAAGATCT
GGAGGAGGAAGGTGGACACACTCTCTAACAAAAAAAACCCTTTTTGAAAT
TTTATACCAATATTTTAAAAGTAAACCAGATCTTTTCAGACATGCCTTTG
AGCTGATATTTGTTAACTAGTTAGAATTAGAAACTTTCCTTATTTTTACT
CAGTTACAATATACGCCACAGCTGAGGTGAGAGGAAAGAAAAGGTTGCTT
TCTTAGGAACAAAGAGTGGTACCTTCAGTATCGTGGGCAAAGCTTTTCCA
AGTCCAACAGCAGTCAAAACAGCGCTTTTTATAAATAACACTCAGCTAAA
AGTTTCTGGGTTTGTGATTGTTCCAACGGTTAAGCTCGGATGAGGGTCCC
TGGAGTCGTAGCTCCCGGGAAACGTCGACTGGCTTTCCACCTGGACTTCA
TCCGTCCAGGCAGCCCAGAGGGGCTTCAGGCCCCGCCCGCTCTCCTGCCA
ACTACAGCCTCGCGACTGCGCTCAGCCTTCAGGCCCGCCCCTTCGGTCA
AGCGGCGTGCTCTCACTGCACGGCGCCTGGGCCCCGCGCGCCGGGACCTC
GGTTTCAGCCGTCCTGTCCTGCCCCGAGGCCCCTAGGCCCCGCCCCTGGG
CCCCGCGCGCCAGGACTTCGGTTTCGACCGTCCTGTCCCGCCCCGAGGCT
CCTAGGCCCCGCCCCCTCTGTCCCCGGCGTGTTCTCGCGGCTCCGCCCCT
AGGACCCGCGCGCCGGGACTTTGGCAAGTTTCAGCCGTCCGGCCCCGCCC
CCTCGGTCCCACGGCTCTCGCGGCCCCTCCCCTAAGTCCCACACGCCGGG
ACTTTGGCAAGTTTCAGCCTCCAGCCCCACCCCTAGGTCCCGCCCACTCG
GCCAGCGGCTGGCTCTCGCGGCCCCGCCCCTGTGCCCTGCGAGTCCCTAT
TTTGGGAGCATTGCGGccgccgtgccccgccccctccccgcgcgccccgcc
ccTCTGGCGGCCCGCCGTCCCAGACGCGGGAAGAGCTTGGCCGGTTTCGA exon 1
GTCGCTGGCCTGCAGCTTCCCTGTGGTTTCCCGAGGCTTCCTTGCTTCCC
GCTCTGCGAGGAGCCTTTCATCCGAAGGCGGGACGATGCCGGATAATCGG
CAGCCGAGGAACCGGCAGCCGAGGATCCGCTCCGGGAACGAGCCTCGTTC
CGCGCCCGCCATGGAACCGGATGGTCGCGGTGCCTGGGCCCACAGTCGCG
CCGCGCTCGACCGCCTGGAGAAGCTGCTGCGCTGCTCGCGTTGGTAAAGA
CGGAGCTTCTTGGGGGTGGCTGCGAGGGCACGGGTCGCACAGTTTCTGGG BARD1_e01''
GGCGGCAGAATCTTTTCAAATCTTCCGTTTCCTCCTTCCGTTCCCGCGCT
GCAGTCGGGTCGGCGTGCGGTTAGCACCTGCCGGGGGATATAGTATTAAC
AACTTCTGCTTCTCATTCACTTTATTTTTGGGCGACTTACCGGCCTCCCC
TTGCCCTGAATCCAACTGAAAcggtagtttttgaacttcagcgggctgaa
gaaccgtctggaggtgtggctaaaaaaatgttcatcccggtcgcgcctcc
agagtttgaatcgggctggggtggggctgaggcttctgcatttttttacCC
GGCCCTGGATTACCCCGCTGCTTTCCGGGAGCTGTGGCGAATTGGGCTGG

CGGGCCGCCCCGGAGACCCTCTAAATTAGAAGCAGCTGCCACTCTAAGTT
AAACTGGCCTTTTTGACATTTTCTCCGTGCCAGCTTTTTCGAGTGAGATG
GGATGGAGCATCGGATATCTACCATAGTTGTAGATTGAAGATGGCACGGA
ATTTCTCATTTTCTTAGTTTGCTCAAAAGACTGTATGTCTGGTGTCCCCG
CTCTTAGTGATGCTGTTTATTGTTTTCCTTCATGCTGTCACATTATGGGA
GTCCTCTCAGGGTTTTCATTCCGTCCAAAGACATTTTCATGAGTGTTACT
TCTTGCTTTTTTTCGTGTAGCTCTTTCTAGTATACAAACCAAGAAGGTTT
GTAAAACCCAAGAAAGGTTTGTCTCATTTTCATTTCTTCTGAAAAGCATT
TTTATTGAGCACTTGCTGTGTGTACAGCTAAGTATTTCTGTATGAAGATC
ACTGGCAGTTTTTTTGTTATTTGTAGCTGTCAAAAGTTTATATCCTTTTT
GGACTTTTTTCCCCTTTGATTTCTAACTTGCAGAACTTAAGTTACCTAGT
TGCATAAAGTATAAGTAACACAACttttttttttttttcctttttgaga
cggagtctcgctctgtcgcccaggctggagtgcagtggcacgatcttgtc
tcactgcaagctccgcctcccgggttcacgccattctcctgcttcagcct
cccaagtagctgggactgcaggcgcccgccatcatgcccggctaatttttt
tgtatttttagtacagatggggtttcaccgtgttagccaggatggtctcc
gtctcctgacctcgtgatccgcccgcctctgcctcccaaagtgctgggat
tacaggcgtgagccacggtgcccggccAAACTTTTTCTTTAAAAATGAAA
AGGACTTAACTGTCAGAGTTTTGCACTTATATTTTCTTTGACGAAAACTT
TTAATATCTGTGACAAAATGACATTGCACTTCTTAGAAAAGTGACACTCA
GATAAGTGAGTACATGAAGACATTTGAGTTGTGTGTTTGTTTTCCTGTAG
CACATGGGTCATCTGAGAATAGGAAATTTAAAACTGGCACTGCTAGTTGC
AATAACAGAAGCATTATAGTTCACACCTTGATAGTCCAACGAAAGTTAGA
CATTCCCTTGGCACATACCTATgtggaagcatcgggcttcacttagactg
tggtcagaatgcctgagttcaggttctagctttgacactattgagtaggc
caagttattcatctctcattttccggatatgagtaatactaatgattagc
ttattttcttgtgaggattcattaataatcataaatttattgagcaccca
taaagtgccagaccctattctgggtattgggggatttgttagtgaacaaag
cagacaataaattccatctcttagggagcttttatggaagaaacaaAATG
AAATTTTACATGTATGTAAAGTATTGAGTTTGGTGGTTGTGATAGAATAA
ATGATAAATAATGTGTGTCTTAATCTAAAAGAATGTGGTGTTTGCTTGCT
ACATGATTTGGGAGAGACAAAATAGATAATATATATGAATATATGTTGTC
CTGAAAAATAACAAAGCCGGTTTCACTTTTTATCCCCTACCAAATATTTT
TCCAGTTTCTAACTGAAGTGAGTAGACAGTAAGAGGCAACGGAGTGTAAA
CAGACATCAGTAACCAGGCCCAGACTTTTAGCATGGCTCACCCATTTGAC
TACAGAATGCTATTGAATGCTATTGATTTTTTTTTTTTTTAATCCGAACCT
CTATAGATGAACTGAATCTTAACTCTTAGACTGCATGTTTTCATGTaggg
gtctgcaaatttttcggtaaagggccaggtaattaatattttaggcctt
tctgcccgcaaatagtttttgttcttatttatttttacatctcttaaaaa
tgtaaaaaccattctttccttaagggctgtacagaaacagactgctggct
tggtttggcccacaggcaatagttGCTAAACTGTCGATAAAGTCTTGATA
AACTTAATGTCTCAAGTTTAAAAATCCTTTTTGGTATACAATTTTATATC
ACATGGGGTGCCtttttttttttttttttccctaggcaaagtttcactctg
tcacccaagctggagtgcagtggcatgatttcagctcactgcaacctcct
tcccccgggttcaagcaattctcttgcctcagcctcctgagtagctggaa
ctacaggtgcatgccaccatgcccggctaatttttgtatttttactagag
acggggtttcatcatgttggccaggctggtctccaactcctgacctcaag
tgatccgcccacttcagcctcccaaagtgctgggattacaggcgtgagcc
accatgcctggtcTTTACATGGGCTTTAATGTTTAGAAATGATTTTGCTT
TCCTCTGTAGGTGTCTCTGCAACTTTGGCACTATCTATAGTGAAAAAAGC
ATGGCCAGCTGCACCTTACTCAAGGAGGAAAAGCATTGTTCCTACCTCTG
CCTTTGCCCTGGTAGCCTCCTGACATTGCCGGCAAGGAGCAACTCAGAAT
GGTCTGCTTCTTCCATTTGCTACCCTCCATTCAAAGCAGTGTCATATTAT
TTGAAATTGTGTCAGTATGTCCTTTTACAGAGTTTCAGAGTTCTACCCAG
TGGGCTAGTGATTACCCCGATTACCCTCACATTCAGCTTGCAGCTGACAT
CTGCCTTGTGTTGCACATGGATCTGTTTTGCAGCAAGCATTACTTGCTGG
TGAGGGTGGATTGGTACATAACCGGGCATCCTTTCTTGACCTTTAACATG
AATCTGGCTTCAAGGTGAAGCCGATGAAAGTGTTCAAGAGGCCAGATGTG
ACATTTATATTAGGACCAAGCCTCACAACCAGAAAGAAGCTATTTAAGGA
TTTGGAAAAATGAGACAGGAAGACCATATTATTTGAGAGGCTGGCAGAAA
GCACAGTAATAATGTGTTCCTAAAGGAAGCAGGAAATGGTAATAGGAGCT
GCAGTGTAGAAGCCCCAGGTGTCATATAAACACTGTTTTGTGAATTAATAG
GTACCTCTTGGCATATTTTTGAGGGTGAGAGTGGGATTAGACATCTATGA
CAATTGTAACGTGGGCCTTTTTAATCAACTTTTTATGAAAAGTTAGAAAG

TAATCTTTTAACCTTTATGATAGGTATTTTTTTTTTTCTGAGTTCTTGTT
TAGGAGATGATCAAGCTTAATTTTGTATTCTCTCATACTTTGAGGcagca
tagtgtagatgttaaatgtgtggaatctggagttagattccctgggttca
aatcctggctccattatttaggttattttatctttttatgcttccttttc
ttcatctgcacaaaatgagtggatattcataAGATGCCTTGAGCTGTGTC
TGTTATGTACTAAATACTATATGAATATTTGTGAAATAAAGGATAGGTTC
TCTGACTCCGGGTTCAGCTTTTGCCTTTCTTCATGAGTTTTGAATGTTCA
GAGGCATTCTCACAGAGCAGTGATCCCCAAATGGAATGATTTCTCCCTGG
CACTGGTTATTCATTAGAATAACGAGAAATCACTGAGACACTGTGTAAGG
GGTGATAAGGTGATGCATTTTCTACTTCTGCCTTTAAGGAGTGTATGCTT
AATTATAAGGGATAAAATAAATACAAGTTCAGGGTAAATTGTCGTGCAGG
TCGACATAGTGCTGAGAAGAGTACTTGGGTAGGGCCAGTGGAAGGAAGCT
TCATGGGCTTTTAAATTGAAATATGAATAGAtttttcttttttctgagacg
gagtctccctctgtcacccaggctggagtgcggtggtgtgatctcggctc
actgcaacctccgcctgccaggttcaagtgattctcctgcctcagcctcc
tgagtagctgggattacaggtgagcgccaccaaacccggctaatttttgt
atttttagtagacacggggtttcaccatgttgctcaggctggtctcgaac
tcctgaccttgtgatccgcacttcttggcctcctaaagtgctaggattac
aggcgtgagccaccacaccaggccGAGATGCGAATAGATTTTTAAAATGG
AGCCAGGTTGCAGGGAGCTGGGGTGATGGGTGTTATTTGCAAATAACTAG
CTATAAAGAATTGTAGGGTAAAGGGGGACTGGGTCAGGTGAACAGAGAGT
TAAGACAGGGTCAAATCAGTGATAGTAGCCTAACAACAGAATACTCAGTT
GGTTTATATAGTTAACTTCCTAAAGAgagcaatttgtatagctttcctgg
actctggttcctcatctataaaatgaggaaaatggactcgttgatttgta
aagttctctccagtttgcagatttcatagttATTTGTGTCCACTGGAtga
agtttaacatacatacagaaaaggcacaaatcattggtgttcagcttgTA
TCGTATCTTCCAAGTTAGAATTTATTTGTGAGATTTATTTATTTATGGAA
TATTTTGAGTGATGTCTGTCTCCTTCAGTTATATATTTAGTAAAACTAGG
TTCACGTTTGTAGCCACAATGCCAGACACATAGTTTCTTAGTGTATGTTA
AATTGATGTATGAAAGAACTTAGGTTAAGGTAGGATCACATTTGAAGGGC
CTTTGCAAGTGGTAATACCAGAAGATCTAGTTAATTTACATGAATTGCTA
AGAATCCCCTATAGGTTCCATGTCATTATCCATATACTTTACTTGTTAAA
CACTTTATTCTTCTAAAGATAAACACAAGAAAAATTGTTTAATTGGCTTG
GTGAGCCTGCACATTTAGATGTTCAACATGTTGGTCCACACGTTAAAATT
ACCTTGAGGATTCTTTAAAATGTAAGAGCACTTTGGGAGCTGTCTTGCCT
AATATTAAATAGCAGTTGATATTTCATGGAGATACAGGGGAGCCATTCTA
TTAAGAAATAGATCTGCCTTGGTATTATCATCTCAGATGGATTTACCTTC
CTTTTTTTCTTTTCGAAGAAAATTTTGTATCAGTAAGAACTTTAAAAAAA
ATTGTTCTCACTACTTAATTCTTCCCTAAGTTATCCTCTGACAGCACTGA
TCtgaaggggtgggttgcccctccacacctgtgggtgtttctcgtaaggt
ggaacgagagacttggaaaagaaaaagacacagatacaaagtatagagaa
agaaataagggacccagggaaccagcgttcagcatatggaggatcccgc
cagcctctgagttcccttagtatttattgatcattcatgggtgtttctcc
gagaggggatgtgtcagggtcacaagacaatagtggggagagggtcagc
agacaaacacgtgaacaaaggtctttgcaacatagacaaggtaaaggatt
aagtactgtgcttttagatatgcatacacagaaacatctcaatgctttac
aaagcagtattgctgcccgcacgtcccacctccagccctaaggcggtttt
tccctatctcagtagatggaacctacaatcgggttttatactgagacatt
ccattgcccagggacgggcaggagacagatgccttcctcttgtctcaact
gcaagaggcttgccttcctcttatactaatcctcctcagcacagaccctt
tacgggtgtcgggctgggggacggtcaggtctttcccttccacaaggcc
atatttcagactatcacatggggagaaaccttggacaatacctggctttc
ctaggcaggggtccttccgcagtgtttgtgtccctgggtacttgagatta
ggagtggtgatgactcttaaggagcaagctgccttcaagcatctgttta
acaaagcacatcttgcacaaccctaatccatttaaccctgagtttgaca
cagcacatgtttcagagagcaccgggttgggggtaggtcatagattaac
agaatctcaaggcagaagaatttttcttagtacagaataaaatggagtct
cctatgtctacttctttctatacagacacagtaacaatctgatctctctt
gcttttccccacaCTGATCTTTCTAAATCAACTACATGTAAAGTTACAGA
AGTATGACTGTAGATGGGGAAAAAGTAGAGTAGCTGAATTTTCAAGGTAC
AGAAATCCTGTTACCATGCAAATGTCTTCTCTTATTTGTTTGGATTTTTT
CTATCTAATATATATGTATATCAAACAGAAAATAGGGTAAGGATAAAATG
GAATTGTTAATACAAATGCATTTTTGTATATAGGCTTCCATTTGGTTCAT
TCACTAAGAACTTATTGAATCCTGCAGTTAATCTAGCACCATGCTAGGT


exon 2
BARD1_e02


exon 3
BARD1_e03




exon 4
BARD1_e04.1
BARD1_e04.2

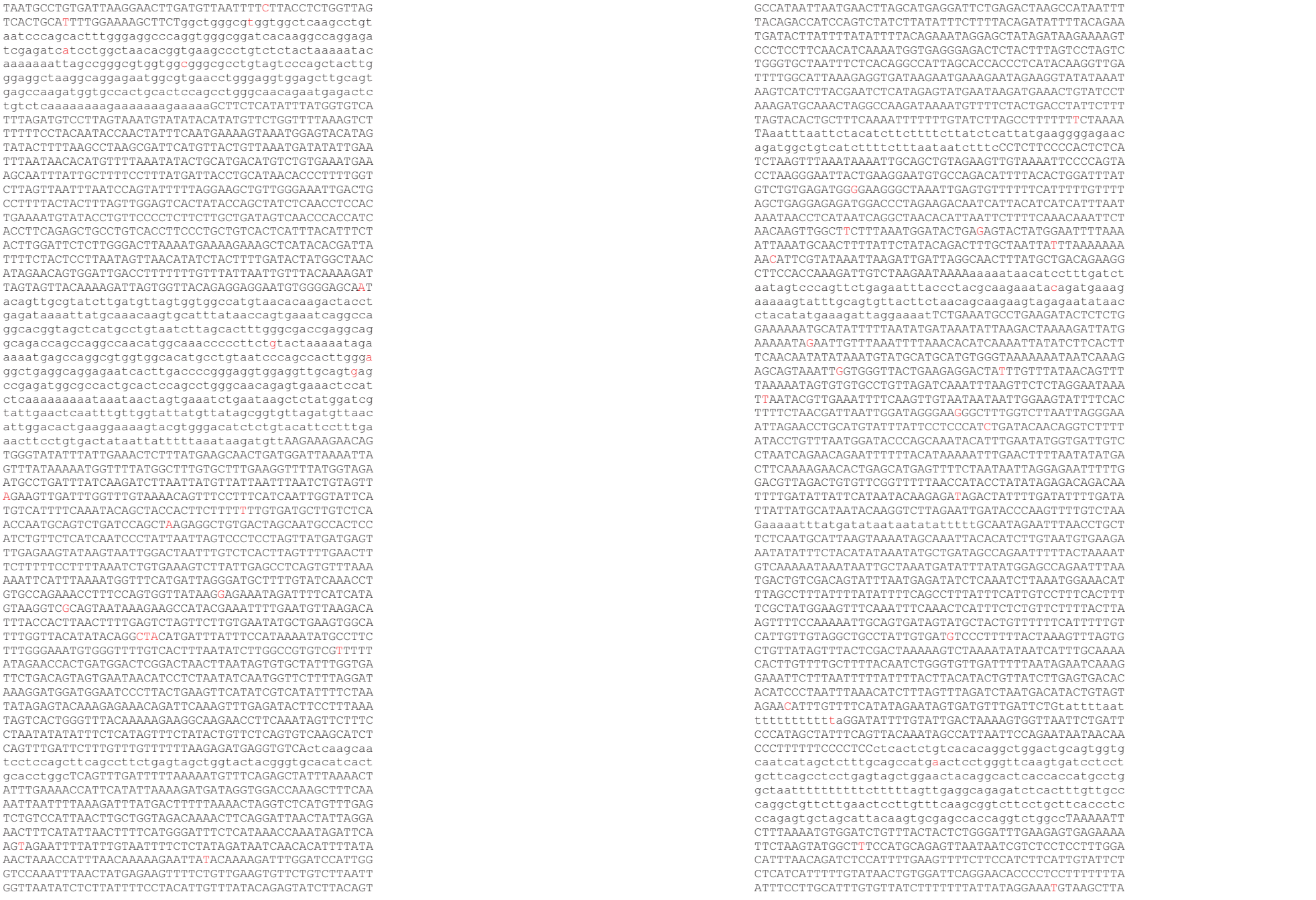


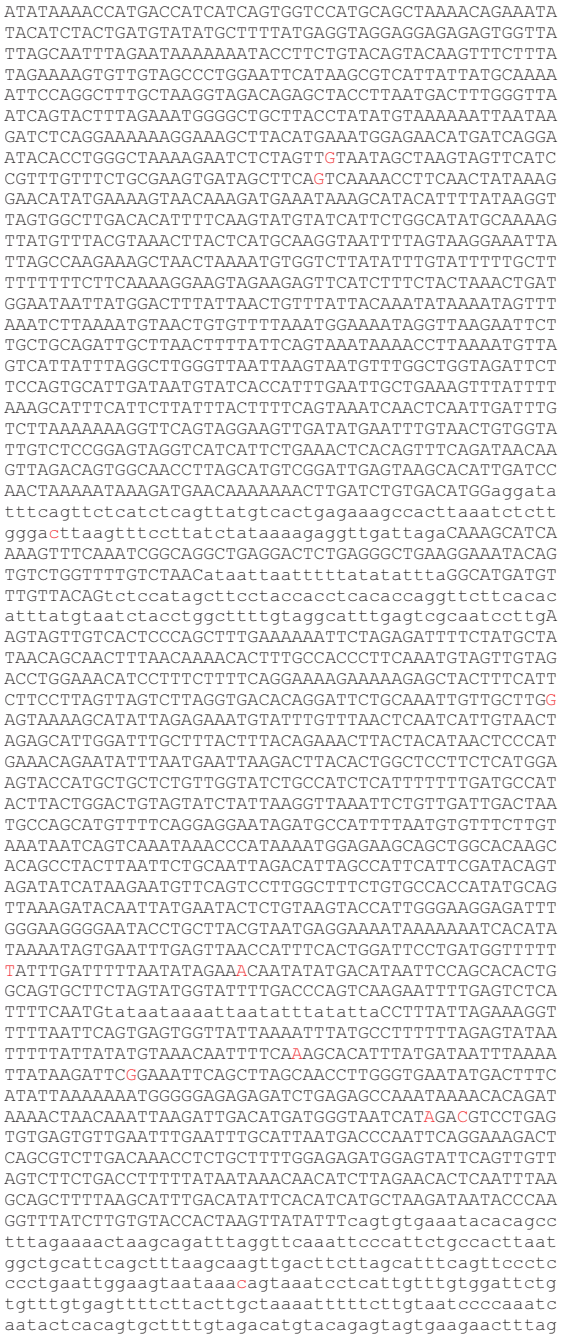
exon 5
BARD1_e05
exon 6
BARD1_e06



exon 7
BARD1_e07


exon 8
BARD1_e08
exon 9
BARD1_e09







exon 10
BARD1_e10’
exon 11
BARD1_e11
BARD1_f3’

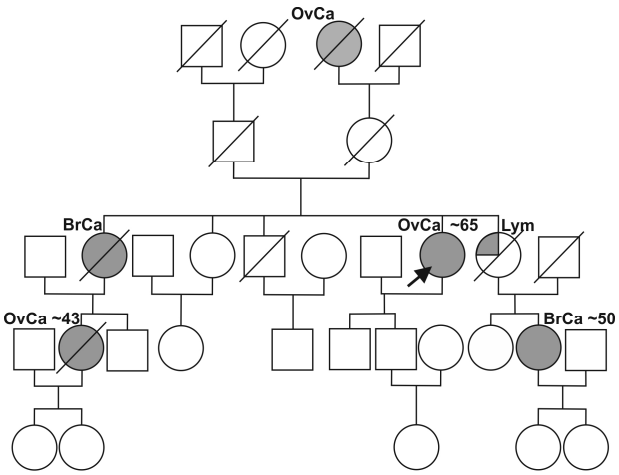
TGGTAGGAACTAATGCTTGTCTTATTTAATTACTAAAAGCCACCATTTTC
TTTGATAGATTGAGCTACAGATTGTAAACTTCATGTATTTCTTTATAAGT
CAACCCTTTTCAAAGATACGCACATCAAACTGAATGAATAAATAAATATT
GAGAAGTTGTCTTCGTTTAAGAGCTGTAATTTTTTTTTTTTAGTCAACT
AATAGTCTTATGCCTGGCTAATTAAGCTGCAGACATTCTTATGAATTAGG
GATTGTTGATTATTTGTGTAGCTTTCTAAGGCACAAGGTGCACTTTCTGA
TACCTGTAAGTTTCTATTAGAATGGTGATGTGAGAGAAGTGGGAAAAGCA
TGTTCTTTGACTAGGCTTAAAGGGTTAATGACTAATACTCCAGCCCCAGG
GAGGATCCTGAATCGGCTTTGAAGTTAGACATGGGTTTGTTAACAATCTA
ACCTGTATCATCAGTGATAAAGGGGGGGAATACAGTTCATTTCACCTGAG
TATAAAGATGTGTGAAAGTGCTTAGCACACTGTAGGCATTCACAACTCGC
CTTCCAGGACCCCTGTGAGCAAGCTCATGAAAAAGTGACTGTTGGTAGCA
CAGCTTGTGTCAGTCCTGAGTATCCTGGCTCTTGTATCCCATGCACAAGA
CTGCTGGGTGTCCTCCTGTATTCACTTTACTGGTTTTAGCCTGAACAACA
ACCAGAGCCAGGCCCTAGTTTCTTATTTCTGACTTCTCACCATCCCATCT
GCACTGGTAAAGTTTGCCCATCTATGTAGCCACTCCTGCTCATTTAACTA
AAATGTTTAGCTTTCATCCCCTCAGTGTATTTACTGTGGATACCAAGATT
GCTTAAGCCTCCACACAGAAGTGGGGCACCTTGGACAGCACTGGCAAAGC
TTCTACAGTTACCATAATATCATAAACCCTTCCGCCACCTCCACTTGAGG
CCTAAAACAAGGATGTTTATCTTTTCTGCCTTAGTTCATAGTTTGTGAGA
CTCATATTTTCATTAATACCACTAATTTGTCAAAACACACCTTATTTGCC
TGCTGTGTAAGATAATGTAACCTAAAAATAAAGGTCTTTTTCTGAATCTT
ACTACTGTAAATGTTTGAGCTTTTAAACAAATCTGTTTACAAAAATATTA
CATAGTTCAGAAATACAAAAAGAATTTCTGCAGAATACAAAAGCACAAGG
AAGAAAACAGGCACAAAATCCAGGGACCGTTAAACTGACAAACTATTTTG
TGTGGCCTTCATAGTGTTAAAAGTACATCAAATGCTGTTGACAGGGGACA
TGAGCTTCCATTCACCACATCCTCACCACTTGTCTCTCACCCCTTACATT
TATTCTTTTAGTTGGGCCTTGGAAACATTGACTTAATCCACAACTCAGGA
ACAACAACTATTAACATTCTTGTCttttttctttccccataagttattggg
gtacaggtggtattgggttacatgagtaaattagttgtgatttgtgagat
tttggtgcacccatcacccgagcagtgtacactgcaccatatttgtagtc
tttatccttcgccccccctcccactcttcctcccaagtctccaaagtccat
tgtattattcttatgcctttgcatcttcatagtttagcttccacatatca
atgagaacatactgtgtttggttttccattcctgagttacttcacttaga
ataatagtctccaatgtcatccaggtcgctgcaaatgctcttcattcatt
gctttttatggctgcatagtattgcatcacacacacacacacacacacac
acacacacacacacacacacgtatatgtatatatacacacaccacagttt
atccactcattgactgatgagcatttgggttggttccatgattctgcaat
tgtgaattgtgctgctataaacatgggtgtgcgagtatgttttccagta
atgacttatttcctctgggtagatacccagtagtgggattgctggatca
aatggtagttctacttttagttctttaaggaatcttcactgttttccata
gtggctatactagtttgccttcccaccagcagtgtagaagtgttccctgt
tcactgcatgcatgccaacatctgttttttgattgtttaattatgaccat
tcttgcaggagcaaggtggtatcacattgtggttttgatttgctgatcat
ttctgatgttgagcatttttttcatatgtttgttggccatttgtgtatctt

Supplementary Table S1

The MLPA probe set designed for the analysis of large mutations in the *BARD1* gene.

probe	5′ half-probe							3′ half-probe							5' HPL	3' HPL	total probe length
	5' PSS	length	5' SS	length	5' TSS	length	Tm	3' TSS	length	Tm	3' SS	length	3' PSS	length			
ctrl_1 [chr22]	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	46	47	93
BARD1_e10'	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	CTGAAGGTCCACGCAGAAGCA	21	70.8	GGCTCAACAGAGAACAGCTGGTATT	25	69.1		0	TCTAGATTGGATCTTGCTGGCGC	23	48	48	96
BARD1_e11	GGGTTCCCTAAGGGTTGGA	19	cgcta	5	CTCAGTAGAAAGCCCAAGCCAGACA	25	71.7	GTGACGTGACTCAGACCATCAATACAG	27	70.5		0	TCTAGATTGGATCTTGCTGGCGC	23	49	50	99
BARD1_f3'	GGGTTCCCTAAGGGTTGGA	19	cgctactac	9	GTGAGAGTGGGGCCACATTAGCT	23	71.4	GTTATTGTTTCCATGGGTCAGTGTGGA	27	71.5	c	1	TCTAGATTGGATCTTGCTGGCGC	23	51	51	102
BARD1_f5'	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	CTGGGACCTGGATAGACACTTGGATAT	27	70.5	CAGCTAGAAAGCTACGACAGGAAACCA	27	71.1	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	52	53	105
BARD1_e02	GGGTTCCCTAAGGGTTGGA	19	cgc	3	CAGTACTAACATTCTGAGAGAGCCTGTGTGTT	32	71.5	TAGGAGGATGTGAGCACATCTTCTGTAGG	29	71.3	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	54	54	108
ctrl_2 [chr1]	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	55	56	111
BARD1_e03	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	CAACTGGACAGCATGATTCAACTTTGTAGT	30	70.6	AAGCTTCGAAATTTGCTACATGACAATGAGC	31	71.1	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	57	57	114
BARD1_e06	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	GAATTATTGCTCCAGCATAAGGCATTGGT	29	71.4	GAACACCACCGGGTATCAAAATGACT	26	71.0	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	58	59	117
BARD1_e07	GGGTTCCCTAAGGGTTGGA	19	cgctactacta	11	GAAAGTATGAAATCGCTATTGCTGCTACCA	30	70.3	GAGAAGAATGAATCATCCTCAGCTAGCCA	29	71.9	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	60	60	120
BARD1_e09	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	ACTCATGTTGTTGTTCCTGGTGATGCA	27	71.6	GTTCAAAGTACCTTGAAGTGTATGCTTGGG	30	71.5	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	62	62	124
BARD1_e08	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	ACTGGGCAGCGTAGGGATGGA	21	71.8	CCTCTTGTACTTATAGGCAGTGGGCTG	27	71.4	caaactaaatctac	14	TCTAGATTGGATCTTGCTGGCGC	23	64	64	128
BARD1_e04.1	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	GTGGTTTAGCCCTCGAAGTAAGAAAGTCA	29	71,5	GATATGTTGTGAGTAAAGCTTCAGTGCAAACC	32	71.6	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	66	66	132
BARD1_e01''	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TTTCTGGGGGCGGCAGAATCTTT	23	72.3	TCAAATCTTCCGTTTCCTCCTTCCG	25	71.3	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	68	68	136
BARD1_e05	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	GCGACATACCTTCTGTTGAATACCTTTTACAA	32	70.5	AATGGAAGTGATCCAAATGTTAAAGACCATGC	32	71.5	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	70	70	140
ctrl_3 [chr17]	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	72	144
BARD1_e04.2	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcc	28	GAGTGATGTCTAGTCCCTCAGCAATGA	27	70.9	AGCTGTTGCCCAATATGGCTGTGAAA	26	71.6	tatctaatggtcaaactaaatctac	25	TCTAGATTGGATCTTGCTGGCGC	23	74	74	148
ARID1A_e03	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttt	35	CAGTCTCAACCACCACAGCTCC	22	71.0	AGTCCTCTCAGCCTCCATACTCCC	24	70.7	aatgtatctaatggtcaaactaaatctac	29	TCTAGATTGGATCTTGCTGGCGC	23	76	76	152
ARID1A_i01	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttt	34	AGCTTCCACTTCTGTGGACTGCTTC	25	70.5	AGGTGTGTGGTAGTAGTCTAGGTGAGGG	28	70.9	tgtatctaatggtcaaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	78	78	156
ARID1A_e16	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttcagct	40	CCACAGCCGAATCTCATGCCT	21	71.1	TCCAACCCAGACTCGGGGATG	21	72.6	tttgcgaaatgtatctaatggtcaaactaaatctac	36	TCTAGATTGGATCTTGCTGGCGC	23	80	80	160

PSS – primer specific sequence, SS – stuffer sequence, TSS - target specific sequence, HPL – half probe length



Supplementary Figure S1 A pedigree of proband #53 (carrier of the c.1690C>T mutation), who is indicated by the arrow. The family members with breast or ovarian cancer are shown in grey. The circle that is ¼ grey indicates a different cancer type. OvCa – ovarian cancer, BrCa – breast cancer, Lym – lymphoma. The numbers near the patient's cancer type indicate the age of diagnosis.

4

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I

Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms

Oncology Reports, 2015, 34(5):2609-2617

Cancer predisposing *BARD1* mutations affect exon skipping and are associated with overexpression of specific *BARD1* isoforms

MAGDALENA RATAJSKA[1,7*], MAGDALENA MATUSIAK[1*], ALINA KUZNIACKA[1], BARTOSZ WASAG[1], IZABELA BROZEK[1,2], WOJCIECH BIERNAT[3], MAGDALENA KOCZKOWSKA[1], JAROSLAW DEBNIAK[4], MARCIN SNIADECKI[4], PIOTR KOZLOWSKI[5], KATARZYNA KLONOWSKA[5], MAXIM PILYUGIN[6], DARIUSZ WYDRA[4], GEOFF LAURENT[7,8], JANUSZ LIMON[1,2] and IRMGARD IRMINGER-FINGER[6,7]

[1]Department of Biology and Genetics, Medical University of Gdansk, 80-211 Gdansk; [2]Regional Oncological Outpatient Clinic, 80-219 Gdansk; [3]Department of Pathology, Medical University of Gdansk, 80-214 Gdansk; [4]Department of Gynecology, Medical University of Gdansk, 80-402 Gdansk; [5]European Centre of Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-138 Poznan, Poland; [6]Laboratory of Molecular Gynecology and Obstetrics, University Hospital Geneva-Belle-Idée, 2, Chemin du Petit Bel Air, Chêne Bourg CH-1225, Switzerland; [7]Centre for Cell Therapy and Regenerative Medicine and [8]Lung Institute of Western Australia, School of Medicine and Pharmacology, University of Western Australia, Harry Perkins Institute of Medical Research, M519, QEII Medical Centre, Nedlands WA 6009, Australia

Received May 27, 2015; Accepted July 2, 2015

DOI: 10.3892/or.2015.4235

Abstract. BARD1 is the main binding partner of BRCA1 and is required for its stability and tumor-suppressor functions. In breast cancer and other epithelial cell carcinomas, alternatively spliced isoforms of BARD1 are highly upregulated and correlated with poor outcome. Recent data indicate that germline mutations of *BARD1* may predispose to breast and/or ovarian cancer. To evaluate the role of *BARD1* germline mutations in predisposition to ovarian cancer we scanned a cohort of 255 patients for the presence of previously reported mutations located in exons 5, 8 and 10 using high-resolution melting analysis. Within this group we identified single-patients carrying mutation in exon 8 (c.1690C>T, p.Gln564Ter), two different variants in exon 10 (c.1972C>T, p.Arg658Tyr; c.1977A>G, p.=) and a carrier of novel missense mutation located in exon 5 (c.1361C>T, p.Pro454Leu). Three out of four identified mutations alter exonic splicing enhancing motives and result in expression of incorrect splicing skipping of exons 5, 8, and 2-9, respectively. Our data indicate that *BARD1* variants may predispose to ovarian cancer in limited number of patients although based on actual data it is difficult to estimate its actual penetrance.

Correspondence to: Dr Magdalena Ratajska, Department of Biology and Genetics, Medical University of Gdansk, Debinki 1, 80-211 Gdansk, Poland
E-mail: mratajska@gumed.edu.pl; magda.ratajska@uwa.edu.au

*Contributed equally

Key words: BARD1, ovarian cancer, mutations, splicing

Introduction

Ovarian cancer is a heterogeneous group of tumors, where epithelial ovarian cancer accounts for the majority of cases. The 5-year survival for invasive ovarian cancer strongly depends on the stage at time of diagnosis and varies from 30 to 90% (1,2). In the majority of patients late diagnosis and high morbidity result mainly from asymptomatic manifestation of the disease (3). Although multiple genetic and epigenetic changes have been studied, and some are characteristic for ovarian cancer, it is still not clear how these changes affect tumorigenesis. To date at least 15 oncogenes and 16 potential tumor-suppressor genes in several signaling pathways have been associated with ovarian cancer (4).

In the majority of populations, ~5-15% of all ovarian cancer cases are caused by inheritance of mutations in genes with an autosomal dominant pattern of transmission (such as *BRCA1/2*). Although the actual proportion strongly depends on the studied population and may be much higher, reaching 30% in the Ashkenazi Jews (5-8). In Poland, the portion of hereditary ovarian cancer due to *BRCA1/2* mutation ranges from 13.5 to 14.9% (9-11).

Another ~6% of constitutive ovarian cancer cases are related to alterations in low and moderate-penetrant genes (12). These genes, products of which are known to interact with BRCA1/2, are involved in DNA repair and cell cycle regulation and therefore are good candidates for possible breast and ovarian susceptibility genes (13,14).

The human *BARD1* gene (BRCA1-associated RING domain 1) is located at the long arm of chromosome 2 (2q34-35) and encodes a nuclear protein of 777 amino acids that shares many structural and functional similarities with BRCA1 (15-17). Like BRCA1, BARD1 has an amino-terminal RING-finger motif and two carboxy-terminal BRCT

domains (15). The RING-finger motif is known to be essential for BRCA1/BARD1 heterodimer formation, while BRCT domains are found in proteins maintaining genome integrity (18). Proper interaction between these proteins is crucial for BRCA1 stability and E3 ubiquitin ligase activity of the BRCA1/BARD1 complex (16,17,19).

BARD1 has additionally three ankyrin (ANK) repeats potentially mediating protein-protein interactions and present in many proteins of various functions (20,21). Interestingly, no other proteins comprising RING, ANK and BRCT motifs together are known (16).

Like BRCA1, BARD1 shows highest expression in actively proliferating cells and those that undergo apoptosis (17,22). Loss of BARD1, as well as BRCA1, leads to developmental retardation and early embryonic lethality of corresponding knockout mice; cells from these mice are characterized by chromosomal instability (23). In addition, it was observed that tumors with homozygous deletions of the entire *BARD1* gene displayed a BRCA1 mutation-like expression profile (24).

Germline mutations in the *BARD1* gene, although detected with low frequency and in a limited number of patients, can be qualified as novel candidates for ovarian cancer susceptibility in a subset of families negative for *BRCA1/2* mutations. The first study linking *BARD1* with ovarian cancer presented a patient with clear cell ovarian carcinoma in who the missense c.1692G>C (p.Gln564His) mutation was identified (25). The p.Gln564His reduces binding of BARD1 to the polyadenylation cleavage specification complex (CstF-50) and abrogates p53-dependent apoptosis (22,26). Most recent publications reported three *BARD1* mutations: One affecting splicing, c.1977A>G, p.=, one non-sense, p.Gln715Ter and one frameshift c.2148delCA; p.Thr716fs*12 (12,27,28). Finally, germline *BARD1* mutations were identified in patients with a breast cancer (24,27,29-31), cervical cancer (32), and neuroblastoma (33).

In the present study, we present four different possibly pathogenic *BARD1* variants identified in a group of unselected ovarian cancer patients. Three of the identified alterations result in incorrect splicing of the corresponding exons, which may prompt expression of specific BARD1 isoforms and promote carcinogenesis.

Materials and methods

Study population. The study comprised 255 unselected ovarian cancer patients referred to the Department of Gynaecological Oncology of Medical University of Gdansk between 1995 and 2009. Within the study group 162 (63.5%) patients were diagnosed with serous ovarian cancer. The remaining 36.5% of tumors were classified as endometroid (n=30/255; 12%); mucinous (n=25/255; 10%); clear cell (n=17/255; 7%); and non-differentiated (n=17/255; 7%). Four tumors (n=4/255; 1%) did not have complete histopatological classification. Average age at diagnosis was 58 (range, 20-88 years). Informed consent was obtained from all of the patients and the study was approved by the Medical Review Board of Medical University of Gdansk (NKEB/399/2011-2012). The frequency of identified *BARD1* variants was investigated in an unselected population-based control group of 1,000 anonymous samples collected at birth (dried blood spots) and in a group of 200 healthy females matched by age.

Blood samples. Patient samples: Genomic DNA was extracted from the whole blood using the Genomic Midi AX kit (A&A Biotechnology, Poland).

In addition, from selected patients a blood sample was collected into Tempus™ Blood RNA Tubes and total RNA was isolated with Tempus™ Spin RNA Isolation kit (Life Technologies, USA). cDNA was synthesized using the Go Script™ Reverse Transcriptase according to manufacturer's instructions (Promega, USA).

Population-based control group: Genomic DNA was extracted from 1,000 dried blood spots using Kapa Express Extract kit (Kapa Biosystems, USA).

Tissue samples. Ovarian cancer: DNA from ovarian tumors was extracted from formalin-fixed, paraffin-embedded (FFPE) tissue blocks using Kapa Express Extract kit (Kapa Biosystems). Total RNA from FFPE tissue blocks was isolated using the High Pure FFPE RNA Micro kit (Roche Diagnostics, Switzerland).

Chronic myeloid leukemia (CML) controls: In order to verify a presence of identified BARD1 isoforms in a hormone-independent cancer, we examined cDNA from bone marrow cells from 120 CML patients.

Mutation screening. All samples were tested for the presence of three previously described *BARD1* mutations (c.1315-2A>G in intron 4; c.1690C>T in exon 8 and c.1977A>G in exon 10) by using high-resolution melting analysis (LightScanner® System; BioFire Defense, USA). Primers listed in Table I were designed based on the *BARD1* gene sequence obtained for the Ensemble database (http://www.ensembl.org; *BARD1*: ENSG00000138376). In order to simplify subsequent sequencing analysis, M13 adaptor sequences (indicated by capital letters) were added to the 5'-end of each primer. All amplicons demonstrating melting profiles distinct from those of the wild-type samples were subsequently sequenced (ABI PRISM 3130; Life Technologies). More detailed information on screening protocols, can be obtained from the corresponding author upon request.

Bioinformatics analysis and sequence variation nomenclature. *BARD1* mutations were numbered according to the Human Genome Variation Society guidelines (34). The *BARD1* sequence was in accordance with GenBank NM_000465.2 All mutations were analyzed for potential pathogenic effect using the following *in silico* software: *Alamut Mutation Interpretation software*, *ESE finder*, *Human Splicing Finder*, *MutPred Splice*, *Mutation Tester*, *PolyPhen 2*, *Rescue ESE* and *SIFT*.

Results

Here we investigated a role of *BARD1* germline mutations in predisposition to ovarian cancer. Frequency of *BARD1* recurrent mutations was estimated in a group of 255 unselected ovarian cancer patients. Exons 5, 8 and 10 together with flanking intron sequences were analyzed by HRM technique followed by bi-directional sequencing. A summary of the identified *BARD1* alterations is presented in Table II.

BARD1 germline mutations. In patient #109, diagnosed with advanced serous ovarian cancer at the age of 70, a novel

Table I. Sequences of the primers used for amplification of exons 5, 8 and 10 together with PCR and HRM conditions.

No.	Primer name	Primer sequence	PCR annealing temperature (˚C)	Product length (bp)	HRM melting range (˚C)
1	BARD1_ex5Frw	TGTAAAACGACGGCCAGTttttcctttctttccttaatgctt	64	236	78-88
2	BARD1_ex5Rev	CAGGAAACAGCTATGACCaagagtatatgtggcagaggatga			
3	BARD1_ex8Frw	TGTAAAACGACGGCCAGTtcgtctaatgtttttaacactggt	68	210	80-90
4	BARD1_ex8Rev	CAGGAAACAGCTATGACCtctaccccacctcccaaaat			
5	BARD1_ex10Frw	TGTAAAACGACGGCCAGTtcgtctaatgtttttaacactggt	63	293	80-90
6	BARD1_ex10Rev	CAGGAAACAGCTATGACCagctgttgaaagggcagaag			

BARD1, BRCA1-associated RING domain 1.

Table II. *BARD1* gene variants identified in intron and exon sequences in 255 patients with ovarian cancer.

					Patients (n=255)		Unselected controls (n=1,000)		Healthy controls (n=200)	
No.	Intron/exon	Nucleotide change	Effect	Status	n	%	n	%	n	%
Sequence variants identified in intron sequences										
1	4	c.1315-19A>G	NE	rs6704780	130	51	NA	NA	NA	NA
2	10	c.2001+66A>G	NE	De Brakeleer *et al* (29)	7	2.75	NA	NA	NA	NA
Sequence variants identified in exon sequences										
1	5	c.1361C>T r.[=,1315_1395del] p.Gly439_Leu465del	p.Pro454Leu	Novel	1	0.39	0	0	0	0
2	8	c.1690C>T r.[=,1678_1810del] p.Met560Ter	p.Gln564Ter	Ratajska *et al* (27)	1	0.39	0	0	0	0
3	10	c.1972C>T	p.Arg658Cys	rs3738888	1	0.39	3	0.3	2	1
4		c.1977A>G r.[=, 159_1903del] p.Cys53_Trp635delinsfs*12	p.=	Ratajska *et al* (27)	1	0.39	1	0.1	2	1

BARD1, BRCA1-associated RING domain 1.

substitution c.1361C>T located in exon 5 was identified. The substitution results in a change of a highly conserved amino acid (p.Pro454Leu), located within the ANK repeats of the protein which suggest possible alteration of the BARD1 structure and/or interactions with other proteins. The information regarding the patient family was limited and it was unclear whether additional breast and/or ovarian cases were present in other family members (Fig. 1A). The germline mutation was found heterozygous in the tested tumor sample. The c.1361C>T was not detected in the control groups.

The second identified genetic variant was a non-sense mutation in exon 8 (c.1690C>T, p.Gln564Ter) previously reported by us (27). This mutation is located between the ANK repeats and the BRCT domains. The alteration was found in patient #53 diagnosed with serous ovarian cancer at the age of 65. The patients' family showed a strong aggregation of the disease (Fig. 1B). The sisters of the proband were affected with breast cancer and lymphoma. In addition, three nieces were diagnosed with ovarian cancer at the age of 43 and a breast cancer at the age of ~50. This variant was not found in the studied control groups.

The third, possibly deleterious alteration, located in exon 10 (c.1972C>T) was identified in patient #150 affected with ovarian cancer at the age of ~50. The proband's paternal aunt and grandmother were diagnosed with hepatic metastases of unknown primary site (Fig. 1C). The substitution c.1972C>T results in an amino acid change at position 658 (p.Arg658Cys), which is located between the two BRCT domains. This genetic variant was also detected in both control groups (unselected control group: n=3/1,000; 0.3% and healthy individuals:

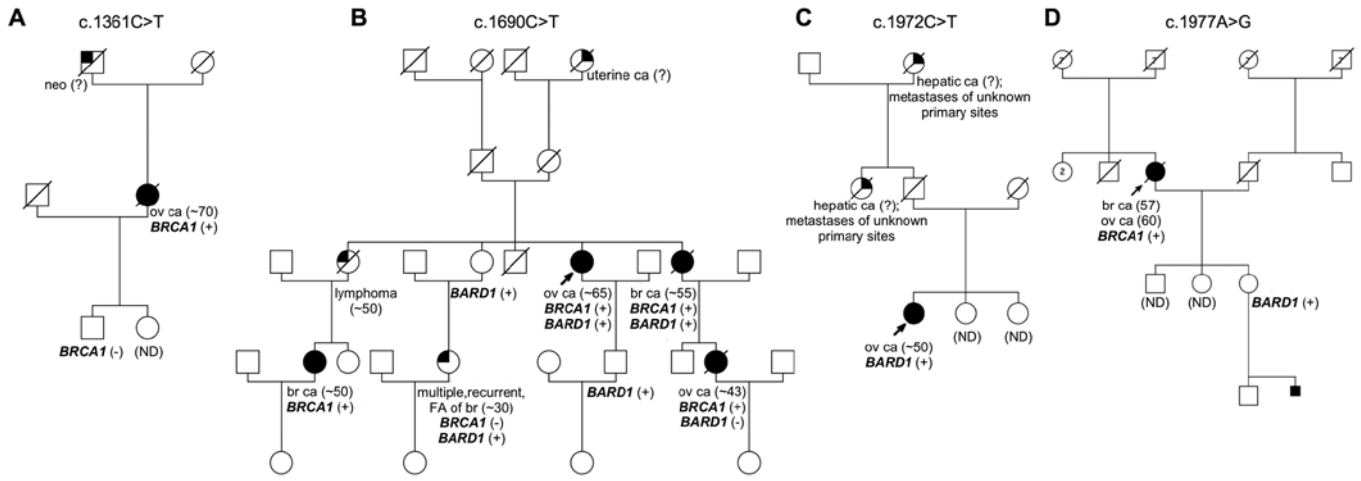


Figure 1. Co-segregation analysis. (A) Pedigree of family #109 carrying mutation c.1361C>T; (B) pedigree of family #53 carrying mutation c.1690C>T (*BARD1*) and c.5266dupC (*BRCA1*); (C) pedigree of family #150 carrying mutation c.1972C>T; (D) pedigree of family #4321 carrying mutation c.1977A>G. (♂) proband; (~70) age at diagnosis; (/) dead; (-) wild-type individuals; (+) individuals carrying mutation; (ND), not done. BARD1, BRCA1-associated RING domain 1.

n=2/200; 1%) suggesting its limited significance. Analysis of the tumor sample revealed the heterozygous status of this substitution.

Finally, we identified a carrier (#4321) of a previously described c.1977A>G mutation in exon 10 (27). The patient was diagnosed with both breast and ovarian cancer at the age of 57 and 60, respectively. There was no information on other cancer cases within the proband's family. Although co-segregation analysis revealed that one of the proband's daughters was a carrier of this alteration (Fig. 1D). Tumor tissue was not available for the study. c.1977A>G variant was also detected among control samples (unselected control group: n=1/1,000; 0.1% and healthy individuals: n=2/200; 1%).

Analysis of amplicons spanning the flanking intronic sequence of exons 5, 8 and 10 lead to identification of two frequent intronic variants c.1315-19A>G (rs6704780) and c.2001+66A>G (rs75237746), which are well-known polymorphisms (18). These were present in a studied group of patients with a frequency of 49 and 2.6%, respectively. Because of the high frequency, and therefore presumably neutral character of these SNPs, we did not investigate their frequency in control groups.

Identification of a BARD1/ BRCA1 double mutation carrier. Interestingly, mutational analysis of the *BRCA1* gene in patient #53, positive for *BARD1* alteration (c.1690C>T), revealed the presence of a deleterious mutation c.5266dupC in exon 20. Further segregation analysis showed that both *BRCA1* and *BARD1* mutations were present in other family members (Fig. 1B). Investigation of the mutation status in tumor tissue of patient #53 reveled a heterozygotic character of both alterations.

Cancer-associated alterations of the BARD1 gene affect splicing. In order to evaluate the possible effect of the identified *BARD1* sequence alterations on protein function and splicing, *in silico* analysis was performed. The synonymous substitution c.1977A>G, previously reported to affect splicing (27), was used as indicator of accuracy in predicting potential splicing disruption. Although the employed software packages individually generated inconsistent results, the combined analyses indicated a possible influence of the tested mutations on the splicing process (Table III).

In order to confirm the *in silico* results, RT-PCR using RNA isolated from patients' peripheral blood cells was applied. In one case, due to the death of the patient with *BARD1* c.1361C>T substitution, RNA was extracted from the normal tissue macrodissected from the resected tumor sample. RT-PCR was performed with primers located within exons 4 and 6. Agarose gel electrophoresis showed two prominent bands. The upper band had the expected wild-type size (218 bp) and the lower band of 137 bp corresponded to a fragment lacking exon 5. Sequencing of the shorter band confirmed exon 5 skipping, resulting in in-frame deletion from c.1315 to c.1395 [r.(=, 1315_1395del); p.Gly439_Leu465del] (Fig. 2A-D). On the protein level, this deletion results in disruption of the 1st and 2nd ANK repeat.

Similarly, the c.1690C>T mutation in exon 8 was examined. The RT-PCR experiment was performed using forward and reverse primers in exons 7 and 9, respectively. Once again agarose gel electrophoresis exhibited two bands: One band corresponding to the wild-type fragment (264 bp) and a lower band. Sequencing of the smaller band confirmed the presence of frame-shift deletion of exon 8 [r.(=,1678_1810del)], resulting in formation of premature stop codon at position p.Met560 (Fig. 2E-H).

Finally, c.1972C>T substitution located in exon 10 was analyzed. RNA was extracted and RT-PCR using previously described primers was applied (27). Alteration c.1972C>T does not affect the splicing process.

Ovarian cancer-associated BARD1 alterations might be cancer-specific. Ultimately, we analyzed 120 samples of patients with CML to assess the frequency of the newly identified BARD1 isoforms (lacking exon 5 and 8, respectively) in other, hormone-independent types of cancer. None of the

Table III. Prediction of possible pathogenicity of identified mutations performed by combining seven different *in silico* tools.

Exon	Nucleotide change	Effect	Predicted effect on splicing				Predicted pathogenicity		
			MutPred Splice	Human Splicing Finder	Alamut	Rescue ESE	PolyPhen 2	SIFT	Mutation Tester
Sequence variants identified in exon sequences									
5	c.1361C>T	p.Pro454Leu	Neutral	Probably affects splicing	Eliminates binding motifs for SC35	Eliminates one binding motif	Probably damaging	Deleterious	Disease-causing
8	c.1690C>T	p.Gln564Ter	Affecting	Probably affects splicing	Eliminates binding motifs for SF2/ASF; creates new SRp55 binding motif	No effect on existing binding motifs	Not relevant	Not relevant	Not relevant
10	c.1972C>T	p.Arg658Cys	Neutral	Probably no effect on splicing	Eliminates binding motifs for SF2/ASF; affects SC35 and SRp55 binding motifs	No effect on existing binding motifs	Probably benign	Deleterious	Disease-causing
10	c.1977A>G	p.=	Affecting	Probably affects splicing	Eliminates binding motifs for SF2/ASF and SRp55; creates new SRp40 binding motif; affects second SF2/ASF binding motif	Eliminates five binding motifs	Not relevant	Not relevant	Not relevant

analyzed sample showed the presence of *BARD1* alterations resulting in formation of the isoforms lacking exons 5 or 8 (data not shown).

Discussion

All initial *BARD1* studies described a limited number of sequence variants, including several missense mutations and one in-frame deletion of 21 bp with unknown consequence for the protein function. However, more recent studies reported several truncating mutations, which were segregating with the disease. Interestingly, *BARD1* germline variants were identified not only in patients with breast and ovarian cancer (12,24,27-29,31), but also among individuals with familial neuroblastoma (33).

In the present study, in a cohort of 255 unselected ovarian cancer cases, one novel and three previously reported genomic *BARD1* alterations in exons 5, 8 and 10 (c.1361C>T; c.1690C>T, c.1972C>T, c.1977A>G) were identified. All four genetic variants were absent or infrequent in the control group (0.1-0.3%), indicating their pathogenic potential.

c.1361C>T mutation disrupts the binding motif for splicing factor SC35 and results in an in-frame deletion of exon 5 and disruption of two presumably important ANK repeats. Based on previously published data, we assume that lack of 27 amino acids at positions p.Gly439_Leu465 may diminish the ability of BARD1 to induce apoptosis (35,36). BARD1 mRNAs lacking exon 5 could be a full length (FL) BARD1 or mRNA isoforms lacking also other exons, namely BARD1β, BARD1α, BARD1κ, BARD1ϕ, BARD1γ (Fig. 3B), which have been described before for breast, ovarian, lung, colon cancer and neuroblastoma (37-40).

c.1361C>T is the second *BARD1* mutation resulting in exon 5 skipping. In a previous study, we identified an intronic variant located in the donor site of intron 4 (c.1315-2A>G) that also resulted in in-frame deletion of exon 5 (27). Exon 5 deletion was first observed in the ovarian cancer cell line NuTu-19 (41), derived from spontaneous mutation of rat ovarian cancer cells that recapitulates human ovarian cancer when injected intraperitoneally into mice. This cell line expressed no FL BARD1, but BARD1β which had an additional deletion of exon 5 (41). The NuTu-19 cells were resistant to apoptosis induction, but became sensitive when expressing exogenous FL BARD1 suggesting that loss of exon 5 leads to isoforms that have lost tumor-suppressor functions affecting the apoptosis pathway (41). BARD1β with an additional exon 5 deletion was also observed in mouse spermatogenesis (42). However, impact of exon 5 deletion on the function of FL BARD1 or BARD1β has not been determined.

The second identified alteration, c.1690C>T, was located in exon 8. *In silico* analysis showed that this alteration significantly changed the binding motif for splicing factors. Splicing alteration was confirmed by RT-PCR [r.(=,1678_1810del), p.Met560*]. Translation of the c.1690C>T mutation results in a truncated protein lacking the BRCT domains, which play an integral role in the DNA damage response (43). *BRCA1* mutations located in BRCT domains are correlated with an increased risk for both breast and ovarian cancer (44,45). It was demonstrated that BRCA1 truncated only by 10 amino acids is less stable, does not accumulate in the nucleus and

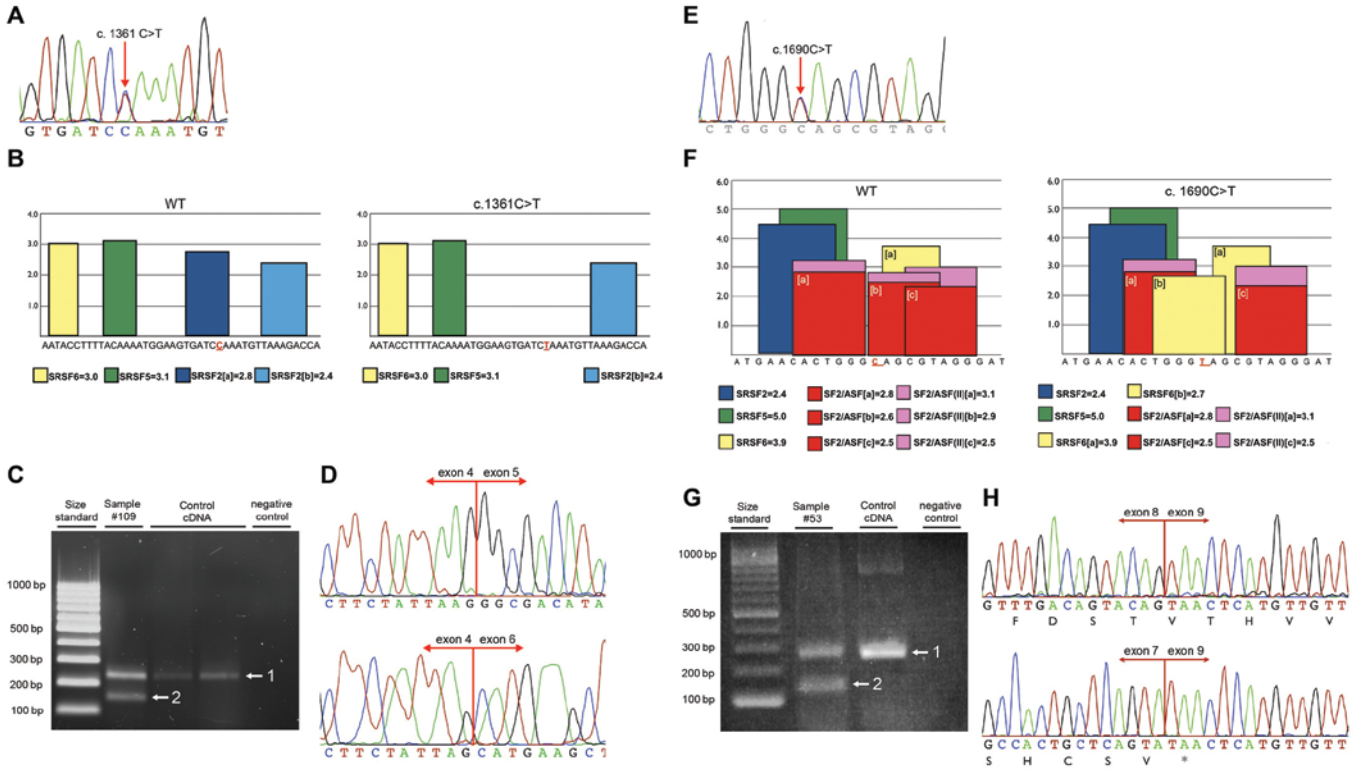


Figure 2. (A-D) Analysis of variant c.1361C>T. (A) Sequence analysis of exon 5 of the *BARD1* gene. The mutation is marked with an arrow. (B) *In silico* tools analysis shows left panel with original ESE motifs located within exon 5 and right panel with ESE motifs status in sequence that harbors the mutation. (C) mRNA expression analysis: Agarose gel electrophoresis of PCR spanning exon 5 revealed two bands: upper band (1) the expected wild-type allele (218 bp) and lower band (2) of 137 bp corresponding to deletion of exon 5. (D) Sequence of the wild-type band (top panel) and the bottom panel exhibit an in-frame deletion of exon 5 of *BARD1* gene. (E-H) Analysis of mutation c.1690C>T in *BARD1*. (E) Sequencing results demonstrating alterations located in exon 8 of *BARD1* gene (arrows). (F) *In silico* tool analysis shows original ESE motifs located within exon 8 of wild-type *BARD1* and right panel with ESE motifs status after in *BARD1* sequence with c.1690C>T mutation. (G) Agarose gel electrophoresis revealed two bands: Upper band (1) of the expected wild-type allele (264 bp) and lower band (2) with a size of 131 bp. (H) cDNA sequence analysis confirmed sequence of the wild-type *BARD1* (top panel) and deletion of exon 8 of *BARD1* gene (bottom panel). BARD1, BRCA1-associated RING domain 1.

fails to colocalize with BARD1 and BRIP1 (44). It is therefore likely that lack of BARD1 BRCT domains have a similar impact on protein function.

In addition, this mutation could not only produce a truncated product of FL BARD1 mRNA, but also truncated proteins from BARD1 isoform mRNAs (Fig. 3C). Previously we reported the *BARD1* c.1690C>T alteration as a non-sense mutation (27), but additional studies showed its impact on splicing alteration.

Moreover, the carrier of c.1690C>T mutation in *BARD1* was also diagnosed with mutation c.5266dupC in the *BRCA1* gene. The duplication of the cytosine at position c.5266 is the second most frequent *BRCA1* mutation all over the world and the most common in Poland (10,11,18). The c.1690C>T mutation in *BARD1* was previously described in a family with an aggregation of breast, colon, and uterine cancer and was further identified in four different families with aggregation of breast and ovarian cancer (27, and Ratajska unpublished data). Co-existence of *BARD1* and *BRCA1* could have either an aggravating or a mitigating affect. However, co-segregation of both mutations may indicate the latter. Another case of co-occurrence of germline *BARD1* (p.Gln715Ter) and *BRCA1* (c.3600del11) mutations was described in a patient with serous carcinoma (28). Moreover, co-occurrence of a *BARD1* (p.Cys557Ser) germline alteration with a *BRCA2* (c.771_775del) truncating mutation was reported as a risk amplifying factor to carriers of both mutations (46). As BRCA1, BRCA2 and BARD1 act together in several tumor suppressor pathways, it is likely that double mutations increase the risk of cancer.

Within this study group, we identified a carrier of recurrent *BARD1* missense mutation (c.1972C>T; p.Arg658Cys) located in exon 10. Several studies have classified c.1972C>T substitution as possibly deleterious (29) or with unknown significance (25,30,47). Rudd *et al* suggested its correlation with a risk of lung cancer with an odds ratio 1.55 (47). In our study, *in silico* analysis performed by using *Mutation Tester*, *PolyPhen*, and *SIFT* indicated its potentially damaging character. However, RT-PCR showed that this variant does not alter the splicing of *BARD1* gene.

Finally, we described another case of synonymous change (c.1977A>G) that results in aberrant splicing, leading to a transcript lacking exons 2-9. This mRNA is identical to BARD1η (Fig. 3A), which can either be translated from the first methionine and end with a premature stop codon in exon 10 (p.Cys53_Trp635delinsfs*12), or translation could start with a first methionine in an alternative open reading frame (ORF) (39,48). This c.1977A>G alteration was previously identified and characterized in a patient with clear cell ovarian cancer with familial aggregation of the disease (27). Additionally, BARD1η isoform was found by Li *et al* in human

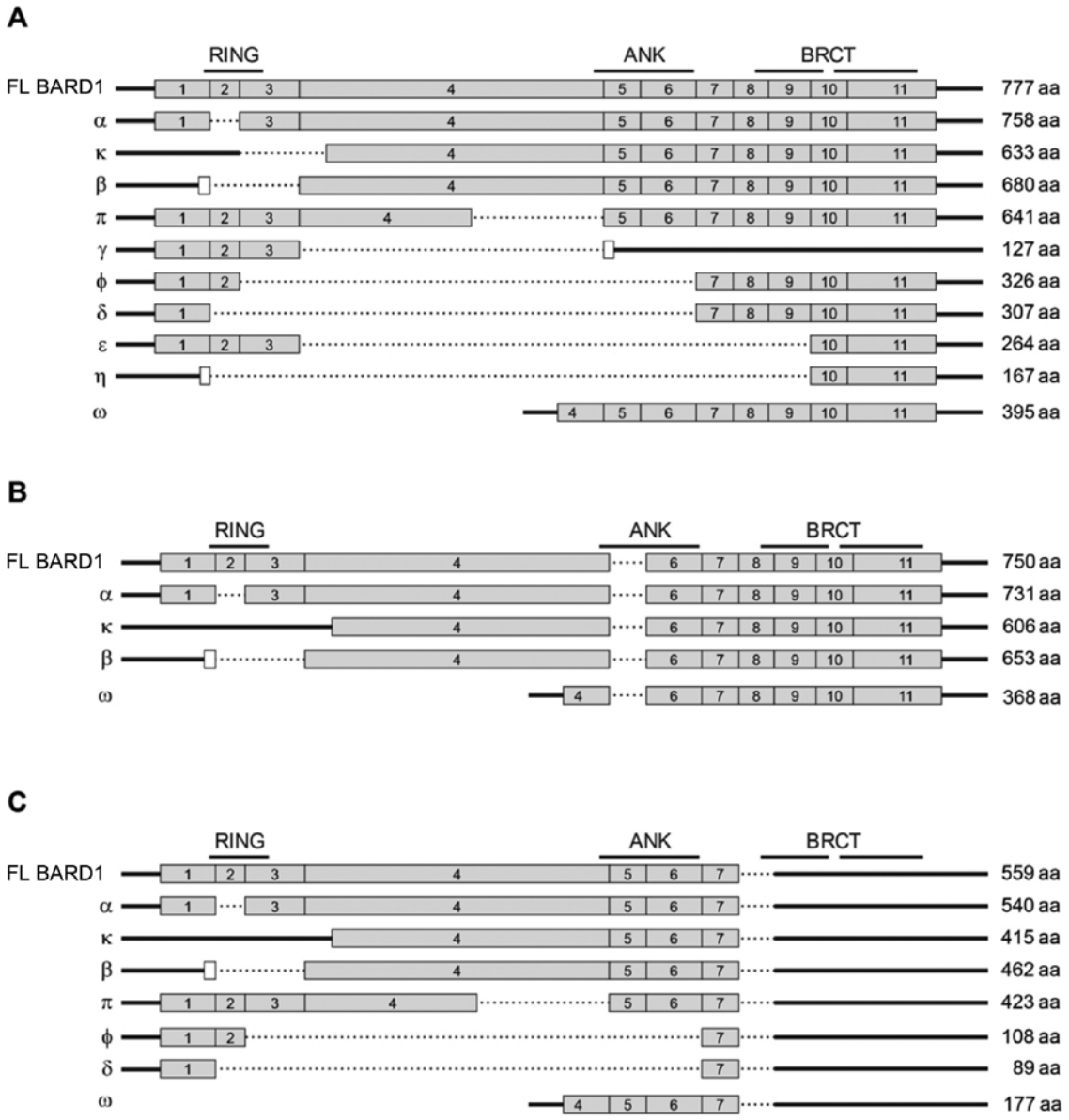


Figure 3. The schematic structure of protein coding FL BARD1 mRNA and splice isoforms is presented (A). Thick line, non-coding sequences; dashed bar, exon skipping; gray bars, protein coding sequence corresponding the main BARD1 reading frame. White bars, alternative, in-frame ORFs. The positions of RING domain, ANK repeats and BRCT domains are shown on the top. The expected length of the polypeptides encoded by the isoforms is shown on the right. (B) The structure of the BARD1 splice isoforms with the exon 5 skipping. (C) The structure of the BARD1 splice isoforms with exon 8 skipping. BARD1, BRCA1-associated RING domain 1; FL, full length; ORF, open reading frame; ANK, ankyrin.

cytotrophoblast invasion and in gynecological cancers, and authors showed the presence of a 21-kDa protein, consistence with the predicted molecular weight of BARD1η (39,48).

Importantly, our results underscore the necessity of multi-level *BARD1* mutation screening, as the molecular analysis limited to DNA may result in high-level misclassification of the detected genetic variants. This phenomenon was widely studied in *NF1* and *ATM*, where authors demonstrated that ~13% of patients may be incorrectly diagnosed (49-51).

Several studies demonstrated that BARD1 isoforms are widely expressed in different types of cancer and that spliced isoforms are often more abundant than FL BARD1 (37-40). Moreover, RNA interference experiments suggested that BARD1 splice variants have functional roles and are the driving force of tumorigenesis (37,39,40,52-54). It was shown that the expression of alternatively spliced BARD1 isoforms was associated with poor outcome and short survival of breast and lung cancer patients. The link between sequence alterations and alternative splicing causing tumorigenesis was convincingly demonstrated for neuroblastoma (40). SNPs that are significantly associated with aggressive neuroblastoma were identified in intronic sequences of *BARD1*. Increased expression of BARD1β was linked to the disease-associated SNP and to functions in malignant transformation (40).

Thus, we suggest that germline *BARD1* mutations are responsible for a portion of hereditary ovarian cancers and *BARD1* should be included in gene panels that are used for molecular diagnosis of breast and ovarian patients. The actual pathogenic role of *BARD1* sequence variants is rather due to activation of alternative splicing and enhanced expression

of certain isoforms than mutations *per se*. Even apparently harmless variants may lead to incorrect splicing process and expression of oncogenic dominant negative forms of BARD1.

Acknowledgements

The authors wish to thank all the patients and their families who participated in this study for their invaluable contribution. This study was supported by the Polish National Science Centre projects: N407 627740, 2011/02/A/NZ2/00017, 2011/01/B/NZ5/02773.

References

1. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ and Thun MJ: Cancer statistics, 2005. CA Cancer J Clin 55: 10-30, 2005.
2. Urban N and Drescher C: Potential and limitations in early diagnosis of ovarian cancer. Adv Exp Med Biol 622: 3-14, 2008.
3. Ruddon RW: Cancer Biology. Oxford University Press, USA, 2007.
4. Bast RC Jr, Hennessy B and Mills GB: The biology of ovarian cancer: New opportunities for translation. Nat Rev Cancer 9: 415-428, 2009.
5. Berchuck A, Heron KA, Carney ME, Lancaster JM, Fraser EG, Vinson VL, Deffenbaugh AM, Miron A, Marks JR, Futreal PA, *et al*: Frequency of germline and somatic BRCA1 mutations in ovarian cancer. Clin Cancer Res 4: 2433-2437, 1998.
6. Malander S, Ridderheim M, Måsbäck A, Loman N, Kristoffersson U, Olsson H, Nilbert M and Borg A: One in 10 ovarian cancer patients carry germ line BRCA1 or BRCA2 mutations: Results of a prospective study in Southern Sweden. Eur J Cancer 40: 422-428, 2004.
7. Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Kwan E, Jack E, Vesprini DJ, Kuperstein G, Abrahamson JL, *et al*: Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. Am J Hum Genet 68: 700-710, 2001.
8. Hirsh-Yechezkel G, Chetrit A, Lubin F, Friedman E, Peretz T, Gershoni R, Rizel S, Struewing JP and Modan B: Population attributes affecting the prevalence of BRCA mutation carriers in epithelial ovarian cancer cases in Israel. Gynecol Oncol 89: 494-498, 2003.
9. Menkiszak J, Gronwald J, Górski B, Jakubowska A, Huzarski T, Byrski T, Foszczyńska-Kłoda M, Haus O, Janiszewska H, Perkowska M, *et al*: Hereditary ovarian cancer in Poland. Int J Cancer 106: 942-945, 2003.
10. Brozek I, Ochman K, Debniak J, Morzuch L, Ratajska M, Stepnowska M, Stukan M, Emerich J and Limon J: High frequency of BRCA1/2 germline mutations in consecutive ovarian cancer patients in Poland. Gynecol Oncol 108: 433-437, 2008.
11. Ratajska M, Krygier M, Stukan M, Kuźniacka A, Koczkowska M, Dudziak M, Śniadecki M, Dębniak J, Wydra D, Brozek I, *et al*: Mutational analysis of BRCA1/2 in a group of 134 consecutive ovarian cancer patients. Novel and recurrent BRCA1/2 alterations detected by next generation sequencing. J Appl Genet 56: 193-198, 2015.
12. Walsh T, Casadei S, Lee MK, Pennil CC, Nord AS, Thornton AM, Roeb W, Agnew KJ, Stray SM, Wickramanayake A, *et al*: Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. Proc Natl Acad Sci USA 108: 18032-18037, 2011.
13. Apostolou P and Fostira F: Hereditary breast cancer: The era of new susceptibility genes. Biomed Res Int 2013: 747318, 2013.
14. Walsh T and King MC: Ten genes for inherited breast cancer. Cancer Cell 11: 103-105, 2007.
15. Wu LC, Wang ZW, Tsan JT, Spillman MA, Phung A, Xu XL, Yang MC, Hwang LY, Bowcock AM and Baer R: Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nat Genet 14: 430-440, 1996.
16. Irminger-Finger I and Jefford CE: Is there more to BARD1 than BRCA1? Nat Rev Cancer 6: 382-391, 2006.
17. Irminger-Finger I, Soriano JV, Vaudan G, Montesano R and Sappino AP: In vitro repression of Brca1-associated RING domain gene, Bard1, induces phenotypic changes in mammary epithelial cells. J Cell Biol 143: 1329-1339, 1998.
18. Williams RS, Bernstein N, Lee MS, Rakovszky ML, Cui D, Green R, Weinfeld M and Glover JN: Structural basis for phosphorylation-dependent signaling in the DNA-damage response. Biochem Cell Biol 83: 721-727, 2005.
19. Baer R and Ludwig T: BRCA1/BARD1 heterodimer, a tumor suppressor complex with ubiquitin E3 ligase activity. Curr Opin Genet Dev 12: 86-91, 2002.
20. Sedgwick SG and Smerdon SJ: The ankyrin repeat: A diversity of interactions on a common structural framework. Trends Biochem Sci 24: 311-316, 1999.
21. Li J, Mahajan A and Tsai MD: Ankyrin repeat: A unique motif mediating protein-protein interactions. Biochemistry 45: 15168-15178, 2006.
22. Irminger-Finger I, Leung WC, Li J, Dubois-Dauphin M, Harb J, Feki A, Jefford CE, Soriano JV, Jaconi M, Montesano R, *et al*: Identification of BARD1 as mediator between proapoptotic stress and p53-dependent apoptosis. Mol Cell 8: 1255-1266, 2001.
23. McCarthy EE, Celebi JT, Baer R and Ludwig T: Loss of Bard1, the heterodimeric partner of the Brca1 tumor suppressor, results in early embryonic lethality and chromosomal instability. Mol Cell Biol 23: 5056-5063, 2003.
24. Sabatier R, Adélaïde J, Finetti P, Ferrari A, Huiart L, Sobol H, Chaffanet M, Birnbaum D and Bertucci F: BARD1 homozygous deletion, a possible alternative to BRCA1 mutation in basal breast cancer. Genes Chromosomes Cancer 49: 1143-1151, 2010.
25. Thai TH, Du F, Tsan JT, Jin Y, Phung A, Spillman MA, Massa HF, Muller CY, Ashfaq R, Mathis JM, *et al*: Mutations in the BRCA1-associated RING domain (BARD1) gene in primary breast, ovarian and uterine cancers. Hum Mol Genet 7: 195-202, 1998.
26. Kleiman FE and Manley JL: The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. Cell 104: 743-753, 2001.
27. Ratajska M, Antoszewska E, Piskorz A, Brozek I, Borg Å, Kusmierek H, Biernat W and Limon J: Cancer predisposing BARD1 mutations in breast-ovarian cancer families. Breast Cancer Res Treat 131: 89-97, 2012.
28. Pennington KP, Walsh T, Harrell MI, Lee MK, Pennil CC, Rendi MH, Thornton A, Norquist BM, Casadei S, Nord AS, *et al*: Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. Clin Cancer Res 20: 764-775, 2014.
29. De Brakeleer S, De Grève J, Loris R, Janin N, Lissens W, Sermijn E and Teugels E: Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. Hum Mutat 31: E1175-E1185, 2010.
30. Karppinen SM, Heikkinen K, Rapakko K and Winqvist R: Mutation screening of the BARD1 gene: Evidence for involvement of the Cys557Ser allele in hereditary susceptibility to breast cancer. J Med Genet 41: e114, 2004.
31. Castéra L, Krieger S, Rousselin A, Legros A, Baumann JJ, Bruet O, Brault B, Fouillet R, Goardon N, Letac O, *et al*: Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. Eur J Hum Genet 22: 1305-1313, 2014.
32. Narayan G, Pulido HA, Koul S, Lu XY, Harris CP, Yeh YA, Vargas H, Posso H, Terry MB, Gissmann L, *et al*: Genetic analysis identifies putative tumor suppressor sites at 2q35-q36.1 and 2q36.3-q37.1 involved in cervical cancer progression. Oncogene 22: 3489-3499, 2003.
33. Pugh TJ, Morozova O, Attiyeh EF, Asgharzadeh S, Wei JS, Auclair D, Carter SL, Cibulskis K, Hanna M, Kiezun A, *et al*: The genetic landscape of high-risk neuroblastoma. Nat Genet 45: 279-284, 2013.
34. den Dunnen JT and Antonarakis SE: Nomenclature for the description of human sequence variations. Hum Genet 109: 121-124, 2001.
35. Jefford CE, Feki A, Harb J, Krause KH and Irminger-Finger I: Nuclear-cytoplasmic translocation of BARD1 is linked to its apoptotic activity. Oncogene 23: 3509-3520, 2004.
36. Fox D III, Le Trong I, Rajagopal P, Brzovic PS, Stenkamp RE and Klevit RE: Crystal structure of the BARD1 ankyrin repeat domain and its functional consequences. J Biol Chem 283: 21179-21186, 2008.

37. Zhang YQ, Pilyugin M, Kuester D, Leoni VP, Li L, Casula G, Zorcolo L, Schneider-Stock R, Atzori L and Irminger-Finger I: Expression of oncogenic BARD1 isoforms affects colon cancer progression and correlates with clinical outcome. Br J Cancer 107: 675-683, 2012.
38. Zhang YQ, Bianco A, Malkinson AM, Leoni VP, Frau G, De Rosa N, André PA, Versace R, Boulvain M, Laurent GJ, *et al*: BARD1: An independent predictor of survival in non-small cell lung cancer. Int J Cancer 131: 83-94, 2012.
39. Li L, Ryser S, Dizin E, Pils D, Krainer M, Jefford CE, Bertoni F, Zeillinger R and Irminger-Finger I: Oncogenic BARD1 isoforms expressed in gynecological cancers. Cancer Res 67: 11876-11885, 2007.
40. Bosse KR, Diskin SJ, Cole KA, Wood AC, Schnepp RW, Norris G, Nguyen le B, Jagannathan J, Laquaglia M, Winter C, *et al*: Common variation at BARD1 results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. Cancer Res 72: 2068-2078, 2012.
41. Feki A, Jefford CE, Berardi P, Wu JY, Cartier L, Krause KH and Irminger-Finger I: BARD1 induces apoptosis by catalysing phosphorylation of p53 by DNA-damage response kinase. Oncogene 24: 3726-3736, 2005.
42. Feki A, Jefford CE, Durand P, Harb J, Lucas H, Krause KH and Irminger-Finger I: BARD1 expression during spermatogenesis is associated with apoptosis and hormonally regulated. Biol Reprod 71: 1614-1624, 2004.
43. Leung CC and Glover JN: BRCT domains: Easy as one, two, three. Cell Cycle 10: 2461-2470, 2011.
44. Nelson AC and Holt JT: Impact of RING and BRCT domain mutations on BRCA1 protein stability, localization and recruitment to DNA damage. Radiat Res 174: 1-13, 2010.
45. di Masi A, Gullotta F, Cappadonna V, Leboffe L and Ascenzi P: Cancer predisposing mutations in BRCT domains. IUBMB Life 63: 503-512, 2011.
46. Stacey SN, Sulem P, Johannsson OT, Helgason A, Gudmundsson J, Kostic JP, Kristjansson K, Jonsdottir T, Sigurdsson H, Hrafnkelsson J, *et al*: The BARD1 Cys557Ser variant and breast cancer risk in Iceland. PLoS Med 3: e217, 2006.
47. Rudd MF, Webb EL, Matakidou A, Sellick GS, Williams RD, Bridle H, Eisen T and Houlston RS; GELCAPS Consortium: Variants in the GH-IGF axis confer susceptibility to lung cancer. Genome Res 16: 693-701, 2006.
48. Li L, Cohen M, Wu J, Sow MH, Nikolic B, Bischof P and Irminger-Finger I: Identification of BARD1 splice-isoforms involved in human trophoblast invasion. Int J Biochem Cell Biol 39: 1659-1672, 2007.
49. Ars E, Kruyer H, Gaona A, Serra E, Lazaro C and Estivill X: Prenatal diagnosis of sporadic neurofibromatosis type 1 (NF1) by RNA and DNA analysis of a splicing mutation. Prenat Diagn 19: 739-742, 1999.
50. Baralle D and Baralle M: Splicing in action: Assessing disease causing sequence changes. J Med Genet 42: 737-748, 2005.
51. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengüt S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA, *et al*: Splicing defects in the ataxia-telangiectasia gene, ATM: Underlying mutations and consequences. Am J Hum Genet 64: 1617-1631, 1999.
52. Ryser S, Dizin E, Jefford CE, Delaval B, Gagos S, Christodoulidou A, Krause KH, Birnbaum D and Irminger-Finger I: Distinct roles of BARD1 isoforms in mitosis: Full-length BARD1 mediates Aurora B degradation, cancer-associated BARD1beta scaffolds Aurora B and BRCA2. Cancer Res 69: 1125-1134, 2009.
53. Woditschka S, Evans L, Duchnowska R, Reed LT, Palmieri D, Qian Y, Badve S, Sledge G Jr, Gril B, Aladjem MI, *et al*: DNA double-strand break repair genes and oxidative damage in brain metastasis of breast cancer. J Natl Cancer Inst 106: pii: dju145, 2014.
54. Chen J and Weiss WA: Alternative splicing in cancer: Implications for biology and therapy. Oncogene 34: 1-14, 2015.

5

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P

The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population

Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400

The 30 kb deletion in the *APOBEC3* cluster decreases *APOBEC3A* and *APOBEC3B* expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population

Katarzyna Klonowska[1], Wojciech Kluzniak[2], Bogna Rusak[2], Anna Jakubowska[2], Magdalena Ratajska[3], Natalia Krawczynska[3], Danuta Vasilevska[4], Karol Czubak[1], Marzena Wojciechowska[1], Cezary Cybulski[2], Jan Lubinski[2] and Piotr Kozlowski[1]

[1]Department of Molecular Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

[2]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland

[3]Department of Biology and Medical Genetics, Medical University of Gdansk, Gdansk, Poland

[4]Department of Gynecology, Centre of Obstetrics and Gynecology, Vilnius University Hospital Santariskiu Klinikos, Vilnius, Lithuania

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

Keywords: *APOBEC3B, hereditary breast cancer, MLPA, copy number variation (CNV), meta-analysis*

Received: *April 25, 2017* ***Accepted:*** *June 10, 2017* ***Published:*** *July 19, 2017*

ABSTRACT

***APOBEC3B*, in addition to other members of the *APOBEC3* gene family, has recently been intensively studied due to its identification as a gene whose activation in cancer is responsible for a specific pattern of massively occurring somatic mutations. It was recently shown that a common large deletion in the *APOBEC3* cluster (the *APOBEC3B* deletion) may increase the risk of breast cancer. However, conflicting evidence regarding this association was also reported. In the first step of our study, using different approaches, including an in-house designed multiplex ligation-dependent probe amplification assay, we analyzed the structure of the deletion and showed that although the breakpoints are located in highly homologous regions, which may generate recurrent occurrence of similar but not identical deletions, there is no sign of deletion heterogeneity. This knowledge allowed us to distinguish transcripts of all affected genes, including the highly homologous canonical *APOBEC3A* and *APOBEC3B*, and the hybrid *APOBEC3A/APOBEC3B* gene. We unambiguously confirmed the presence of the hybrid transcript and showed that the *APOBEC3B* deletion negatively correlates with *APOBEC3A* and *APOBEC3B* expression and positively correlates with *APOBEC3A/APOBEC3B* expression, whose mRNA level is >10-fold and >1500-fold lower than the level of *APOBEC3A* and *APOBEC3B*, respectively. In the next step, we performed a large-scale association study in three different cohorts (2972 cases and 3682 controls) and showed no association of the deletion with breast cancer, familial breast cancer or ovarian cancer. Further, we conducted a meta-analysis that confirmed the lack of the association of the deletion with breast cancer in non-Asian populations.**

INTRODUCTION

Breast cancer is the primary cause of cancer-associated death among women worldwide. The probability of developing breast cancer is modulated by an interplay of lifestyle, environmental, and genetic factors. The overall heritability (h^2) of breast cancer was estimated at approximately 30% [1]. Inherited breast cancer cases that aggregate in families constitute five to ten percent of all breast cancer cases. Highly penetrant germline mutations in *BRCA1* and *BRCA2* and in several genes associated with various hereditary cancer syndromes explain 16-40% of all breast cancer cases that aggregate in families [2–4]. Moreover, it is estimated that mutations in several susceptible genes of moderate penetrance, e.g., *ATM*, *CHEK2* or *NBN*, account for another 5% of all familial breast cancer cases [2, 5, 6]. Recently, international collaborative analyses involving genome-wide association studies (GWASs) have revealed common low-penetrance single nucleotide polymorphisms (SNPs) in 94 loci that are individually associated with breast cancer. It is assumed that the cooperative effect of the identified SNPs may underlie more than 20% of breast cancer heritability [7–9]. Overall, the genetic background of breast cancer predisposition in approximately 50% of breast cancer cases aggregated in families still remains to be explained [4, 7].

It was presumed that investigating copy number variants (CNVs) may uncover a substantial part of still unidentified genetic loci related to the susceptibility to various complex diseases [10]. CNVs have been shown to be associated with several complex diseases, including HIV infection and AIDS development [11], osteoporosis [12], Crohn's disease [13] and autism [14]. Recently, it was also suggested that CNVs may underlie hidden susceptibility to breast cancer [15–17]. One common CNV that potentially increases the risk of breast cancer is the deletion of the *APOBEC3B* gene, which occurs with a high allelic frequency in East Asian (37%), Amerindian (58%) and Oceanic populations (93%) and with a moderate (6%) or low (1%) allelic frequency in European and African populations, respectively [18]. The germline *APOBEC3B* deletion, comprising an ~30 kb genomic region, extends between the last noncoding exon of *APOBEC3A* and the eighth exon of *APOBEC3B* and leads to the complete removal of the *APOBEC3B* protein-coding region. It was suggested that as a result of the *APOBEC3B* deletion, a hybrid gene, *APOBEC3A/APOBEC3B*, which possesses a coding sequence of *APOBEC3A* (exon 1 – exon 5) and 3' untranslated region (3'UTR) from *APOBEC3B* (exon 8), is generated. It can be assumed that transcript generated from the *APOBEC3B* deletion allele is subjected to different cellular regulation. Although the function of the two distinct APOBEC3A and APOBEC3B proteins is being intensively studied (reviewed in [19–24]), little is known about the influence of the germline *APOBEC3B* deletion on the expression of affected genes (genes overlapped by the *APOBEC3B* deletion, i.e., *APOBEC3A*, *APOBEC3B*, and the *APOBEC3A/APOBEC3B* hybrid). Therefore, direct evidence that the presumed *APOBEC3A/APOBEC3B* hybrid transcript actually arises and a detailed elucidation of its structure are of high importance. The knowledge of the exact structure of the hybrid transcript is vital for the design of a comprehensive tests for analysis of the influence of the *APOBEC3B* deletion genotype on the expression of *APOBEC3B*, *APOBEC3A* and the *APOBEC3A/APOBEC3B* hybrid gene, which would deepen the current knowledge of the functional consequences of the *APOBEC3B* deletion.

Several associations of the *APOBEC3B* deletion with different complex human diseases have already been reported [17, 25–42], including an association of the *APOBEC3B* deletion with breast cancer risk in a Chinese population (OR=1.31/1.76 for one/two copies of the deletion) identified by Long and coworkers [17] and in two smaller studies in two other Asiatic populations, i.e., Iranian [36] and Malaysian [40]. In contrast, association of the *APOBEC3B* deletion with breast cancer is much less conclusive in European populations. The association was confirmed by another study of the Long group in a European-American population [31] but not in a Swedish population [37]. The association was also not confirmed in two smaller studies in Indian [38] and Moroccan [39] populations. It was concluded that in the Caucasian population, the relationship of the *APOBEC3B* deletion with increased breast cancer risk cannot be convincingly stated; therefore, further large-scale comprehensive association studies are necessary [43]. The currently available results report risk related to the *APOBEC3B* deletion in groups of unselected breast cancer cases. The role of the *APOBEC3B* deletion in familial breast cancer predisposition remains to be elucidated.

The APOBEC3A and APOBEC3B proteins belong to Activation-Induced Cytidine Deaminase (AID)/Apolipoprotein B mRNA Editing Enzyme, Catalytic Polypeptide-like (APOBEC) family, which consists of 11 cytidine deaminases. APOBEC3A and APOBEC3B possess the capability of introducing sequence alterations in single-stranded DNA (ssDNA) and are involved in various vital cellular processes (reviewed in [19, 20]), including innate immune response against retroviruses (e.g., HTLV1) [44] and DNA viruses [e.g., HBV (e.g., [32, 45, 46]) and HPV (e.g., [47, 48])], regulation of retrotransposon element movement (e.g., [49–51]), and regulation of DNA methylation (e.g., [52–54]). More recently, APOBEC3B and APOBEC3A were also reported to be mutagenic enzymes whose activation in cancer is responsible for specific patterns of massively occurring somatic mutations [55, 56], referred to as *kataegis* (from the Greek for "thunderstorm") [57] or "mutation clusters" [58]. These patterns were observed in several cancer types, including breast cancer [59–62]. However, newer reports

indicate that some controversies regarding specificities and the role of particular APOBEC3s in *kataegis* also exist [63, 64].

In this study, we determined the exact structure of the hybrid *APOBEC3A*/*APOBEC3B* gene and provided direct evidence of the presence of the hybrid *APOBEC3A*/*APOBEC3B* transcript in individuals carrying allele with the *APOBEC3B* deletion. We also analyzed the relationship between the *APOBEC3B* deletion genotypes and the expression of the affected genes, i.e., *APOBEC3A, APOBEC3B* and hybrid *APOBEC3A*/*APOBEC3B*, and showed that the *APOBEC3B* deletion negatively correlates with the expression of *APOBEC3A* and *APOBEC3B* and positively correlates with expression of *APOBEC3A/APOBEC3B*. We also performed a large-scale, case-control study of the association of the *APOBEC3B* deletion with breast and ovarian cancer in three different European cohorts (encompassing >6500 samples), which revealed the lack of association of the *APOBEC3B* deletion with breast and ovarian cancer in European populations. To obtain a more global view of the role of the *APOBEC3B* deletion in cancer predisposition, we also conducted a comprehensive meta-analysis, considering all association studies of the deletion with breast and other types of cancer.

RESULTS

Design of the A3Bdel_PCR assay and comprehensive analysis of the structure of the *APOBEC3B* deletion

As breakpoints of the deletion overlap with extended highly homologous regions (Figure 1, segmental duplications of 95% similarity cover almost entire *APOBEC3A* and 3'-half of *APOBEC3B*), determination of their exact positions may be redundant. Therefore, in the first step we designed a simple single-tube A3Bdel_PCR assay to distinguish the reference (A3B+) and the deletion (A3B-) alleles and to unequivocally confirm the exact size and position of the deletion. The test takes advantage of nucleotide positions specific for particular duplicated regions defined based on a careful analysis of the reference sequence (hg19) and on previous results [18]. The assay consists of three PCR primers, i.e., one forward primer (F) and two distinct reverse primers (R1 and R2) (Figure 1). F is located on the border of *APOBEC3A* intron 3 and exon 4, upstream of the presumed 5'-breakpoint of the *APOBEC3B* deletion; R1 is located in the *APOBEC3A* exon 5 downstream of the presumed 5'-breakpoint of the deletion; and R2 is specific to the sequence within the *APOBEC3B* exon 8 downstream of the presumed 3'-breakpoint of the deletion (Figure 1A). R1 and R2 primers distinguish the A3B+ and A3B- alleles, respectively. The primers are localized in such a way that the amplicons corresponding to the A3B+ and the A3B- are of different lengths, which distinguish them and identify the *APOBEC3B* deletion genotypes (Figure 1). With the use of the designed A3Bdel_PCR primers, we performed a sequencing analysis that determined the deletion breakpoints at a single-nucleotide resolution (Figure 1). The sequencing analysis refined the *APOBEC3B* deletion to a 29 936 bp genomic region. It has to be noted, however, that the 5' or 3' breakpoints of the *APOBEC3B* deletion lie within a 350 bp sequence that is identical on both sides of the deletion; therefore, the exact position of the deletion depends on the assumed convention/nomenclature (according to the HGVS nomenclature: GRCh37/hg19: g.chr22:39358631_g.chr22:39388566del or APOBEC3A:c.717_APOBEC3B:c.1265del).

The location of the breakpoints of the deletion in almost identical, segmentally duplicated regions may induce recurrent occurrence of similar but not identical deletions arising due to non-allelic homologous re-combination (NAHR). Therefore, to exclude the potential heterogeneity of the *APOBEC3B* deletion, we utilized multiplex ligation-dependent probe amplification (MLPA), which is the method of choice for the analysis of large deletions. Because commercial MLPA assays are available only for a limited number of genes and there is no such assay for the *APOBEC3A* and *APOBEC3B* genes, we designed a homemade A3Bdel_MLPA assay. The assay was designed and generated according to a strategy previously developed in our group [65–68]. The A3Bdel_MLPA assay is composed of 12 probes, i.e., 4 probes located in close proximity to the breakpoints within flanking sequences of the deletion, 5 probes located in the presumed deletion region, and 3 control probes specific to copy-number stable regions located in chromosomes 1, 2, and 22 (for details, see Materials and Methods, Figure 1A and Supplementary Table 1). The A3Bdel_MLPA probe set was verified to provide robust, high-quality results in a series of optimization experiments performed using a set of reference gDNA samples. The optimized A3Bdel_MLPA assay was used to analyze two panels of gDNA samples, i.e., a panel of 31 samples from the HapMap project and a panel of 17 samples derived from women with breast and/or ovarian cancer with different *APOBEC3B* deletion genotypes that were previously determined with the use of the A3Bdel_PCR assay (Figure 1D and Supplementary Figure 1). We observed a perfect correlation between the obtained MLPA patterns and the *APOBEC3B* deletion genotypes identified in the A3Bdel_PCR analysis, which indicates the lack of heterogeneity in the structure of the *APOBEC3B* deletion.

Effect of the *APOBEC3B* deletion on the expression of the affected genes

In the first step, to assess the effect of the *APOBEC3B* deletion on *APOBEC3A* and *APOBEC3B* expression, we took advantage of whole genome mRNA

profiling datasets regarding panels of HapMap samples derived from LCLs (lymphoblast cell lines) from B-lymphocytes. The genome-wide expression datasets were generated for 270 and 45 samples from basic (phase I/II) HapMap panel, with the use of microarray [69] and RNAseq [70] technology, respectively. From the datasets, we extracted data regarding the expression levels of *APOBEC3A* and *APOBEC3B* and compared them with the *APOBEC3B* deletion genotypes determined for these samples by [18] and independently determined by us (data not shown). As shown in Figure 2A, the *APOBEC3B* deletion genotype (presence of the deletion) negatively

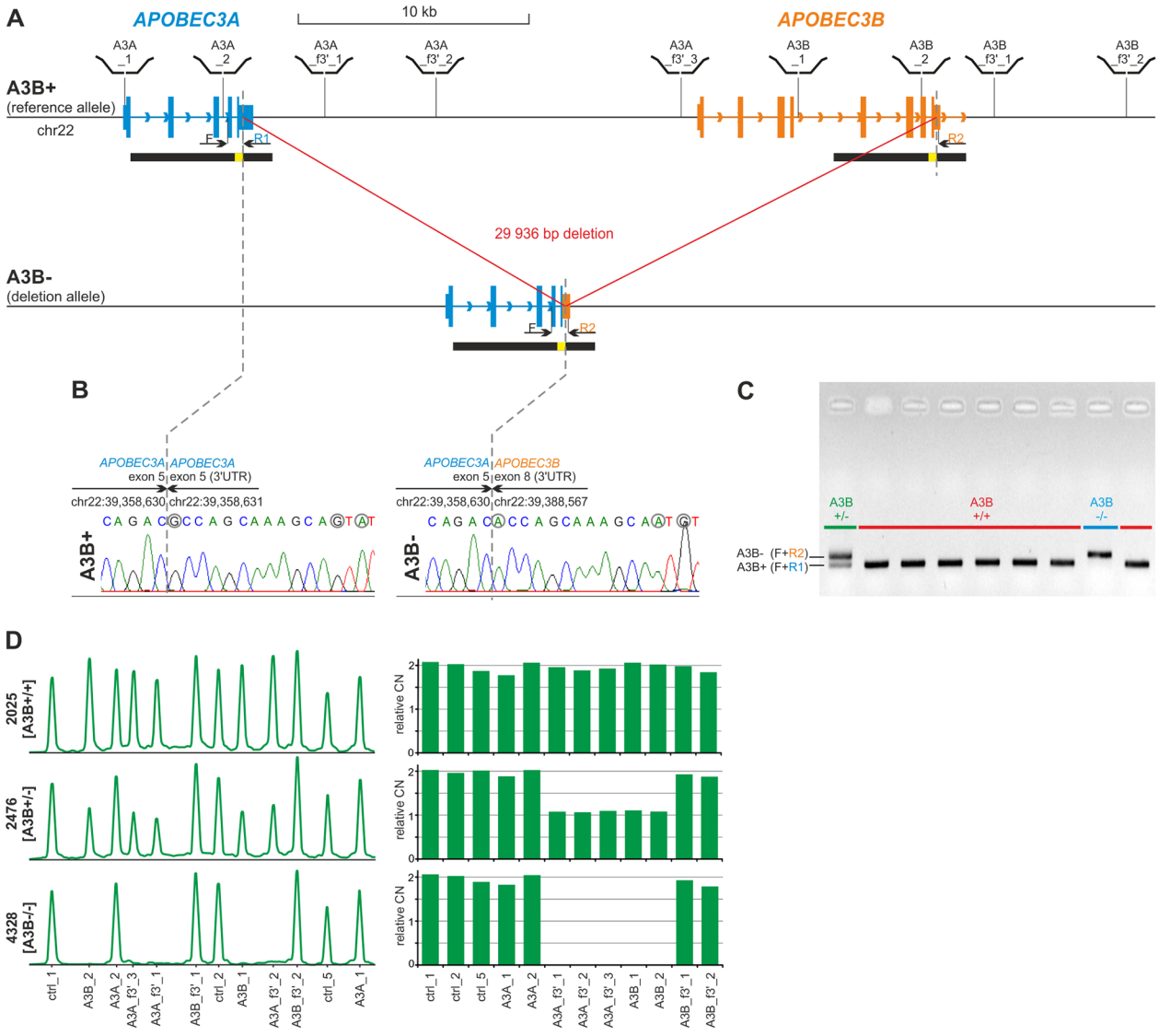


Figure 1: Structure of the *APOBEC3B* deletion and design of the PCR-based and MLPA-based genotyping assays. **(A)** Structure of A3B+ and A3B- alleles. Blue and orange vertical rectangles indicate exons of *APOBEC3A* and *APOBEC3B*, respectively. Higher and lower rectangles indicate coding and UTR sequences, respectively. Arrowheads along intron lines indicate the direction of the genes. Black horizontal bars under the gene schemes indicate highly homologous (~95%) segmentally duplicated regions, with a yellow inset indicating 350 bp redundant fragment of 100% homology. Positions of MLPA probes used for A3Bdel_MLPA assay are schematically depicted above the region map; positions of F, R1 and R2 primers used for sequencing and in A3Bdel_PCR assay are indicated under the map. **(B)** Results of Sanger sequencing of deletion breakpoints amplified with the F-R1 (left-hand side) and F-R2 (right-hand side) pairs of primers, respectively. Dotted lines indicate exact breakpoint positions according to HGVS recommendation (after the last nucleotide of the redundant sequence). Sequencing was performed using homozygous A3B+/+ and A3B-/- samples. Nucleotides that distinguish the A3B+ and A3B- alleles are circled in gray. **(C)** Visualization (agarose gel with EtBr staining) of PCR products of the A3Bdel_PCR assay. A3B+/+ genotypes without deletion; A3B-/- genotype with homozygous deletion; A3B+/- genotype with heterozygous deletion. **(D)** Exemplary results of the A3Bdel_MLPA analysis. The left-hand side panel: the MLPA electropherograms of the representative samples with A3B+/+, A3B+/-, and A3B-/- genotypes. The probe IDs are shown under the electropherograms. The right-hand side panel: bar plots (corresponding to the electropherograms shown on left-hand side) representing the normalized copy number value (y-axis) of each probe (x-axis).

correlates with *APOBEC3B* expression [both microarray (R=-0.74, p<0.001) and RNAseq (R=-0.65, p<0.001) data] but not with *APOBEC3A* expression. It has to be noted, however, that lack of correlation of *APOBEC3A* expression with the deletion genotype may result from (i) very low levels of *APOBEC3A* expression (in comparison to *APOBEC3B*) in some tissues and cell subsets, including T- and B-lymphocytes and breast cancer cell lines [55, 64, 71–73], or (ii) high homology between *APOBEC3A* and *APOBEC3B* and other *APOBEC* family members that might lead to the mismapping of RNAseq reads or cross-hybridization of *APOBEC3A*-specific probes [55, 64]. The lack of consistency in the measurement of the *APOBEC3A* expression level between the studies (in [70], it is lower than the level of *APOBEC3B*, but in [69], it is higher than the level of *APOBEC3B*) strongly suggests the occurrence of the cross-hybridization.

Therefore, to further investigate the effect that the *APOBEC3B* deletion has on the expression of *APOBEC3A* and *APOBEC3B* as well as the *APOBEC3A/APOBEC3B* hybrid gene, occurring in the presence of the *APOBEC3B* deletion, we performed our own experimental analysis. First, we performed a sequencing analysis of the presumed *APOBEC3B/APOBEC3A* hybrid transcript. The analysis unequivocally confirmed that the transcript is actually generated from the allele with the *APOBEC3B* deletion and defined its structure at a single-nucleotide resolution (Supplementary Figure 2). Taking advantage of the gathered information about the precise sequence structure of the *APOBEC3B* deletion and the hybrid transcript, we designed A3A_exp, A3B_exp and A3A/A3B_exp assays for expression analysis that distinguished the hybrid *APOBEC3A/APOBEC3B* transcript from the canonical *APOBEC3A* and *APOBEC3B* transcripts (for details,

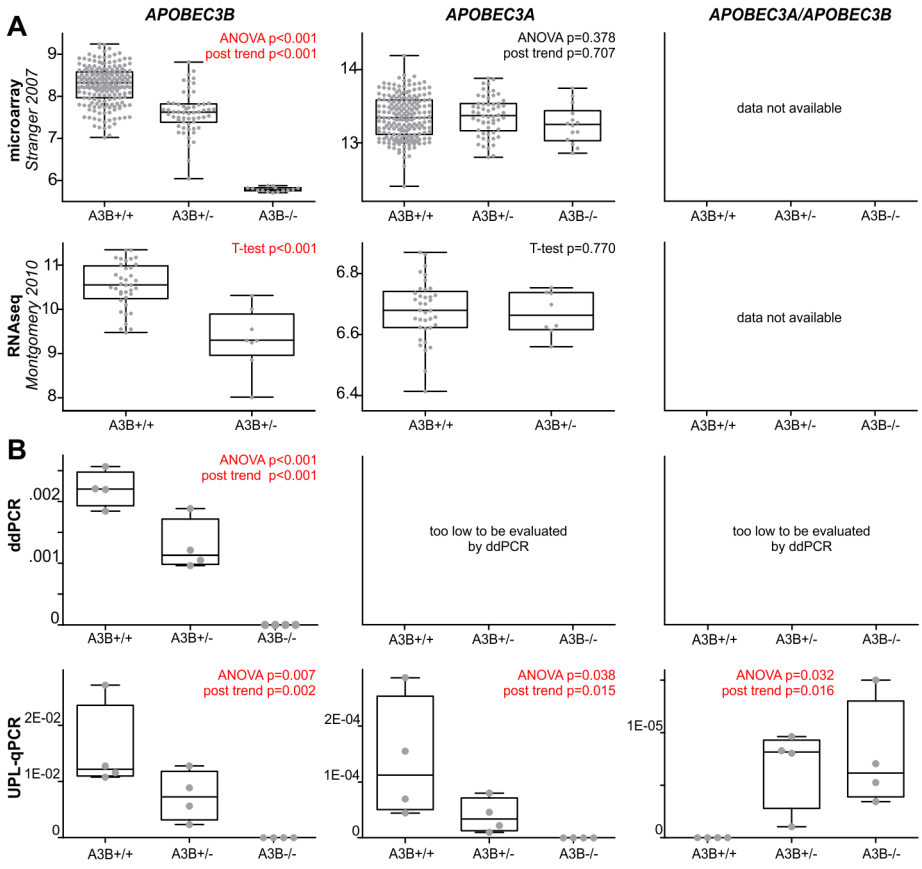


Figure 2: Effect of the *APOBEC3B* deletion on the expression of the affected genes, i.e., *APOBEC3B*, *APOBEC3A*, and *APOBEC3A/APOBEC3B*. (A) Comparison of the *APOBEC3B* deletion genotypes (A3B+/+, A3B+/-, A3B-/-; x-axes) in HapMap samples with relative expression (y-axes) level of the affected genes (indicated above) retrieved from microarray (upper row, [69]) and RNAseq (lower row, [70]) whole genome mRNA profiling datasets. It has to be noted that panel of HapMap samples analyzed in RNAseq study did not comprise samples with A3B-/- genotype. **(B)** Experimental analysis of expression of the affected genes in 12 HapMap cell lines with A3B+/+ (n=4), A3B+/- (n=4) and A3B-/- (n=4). The expression analysis was performed with ddPCR (upper row) and UPL-qPCR (lower row). Note that due to very low levels of *APOBEC3A* and *APOBEC3A/APOBEC3B* expression (>>1000 lower than *APOBEC3B*), they could not be reliably evaluated with ddPCR. The box-and-whisker plots summarize the distribution of the relative expression data points determined by microarray, RNAseq, ddPCR, and UPL-qPCR. The band inside of each box represents the median, and the upper and lower edges of the box represent 1st and 3rd quartile of distribution. Whiskers indicate the lowest and the highest observed values.

see Materials and Methods). For the purpose of the analysis, we selected 12 LCLs from the HapMap project representing different *APOBEC3B* genotypes, i.e., 4 cell lines with a reference *APOBEC3B* genotype (A3B+/+; 2 *APOBEC3B* copies), 4 cell lines with a heterozygous deletion of *APOBEC3B* (A3B+/−; 1 *APOBEC3B* copy), and 4 cell lines with a homozygous deletion of *APOBEC3B* (A3B−/−; 0 *APOBEC3B* copies). The A3A_exp, A3B_exp and A3A/A3B_exp assays were utilized for the evaluation of the expression levels in all the LCLs using droplet digital PCR (ddPCR) method. The ddPCR analysis revealed a negative correlation between the *APOBEC3B* deletion genotype and the expression of the *APOBEC3B* gene. In cell lines with A3B+/+, A3B+/− and A3B−/− genotypes, *APOBEC3B* expression decreased gradually in a nearly linear manner (R=-0.96, p<0.001) (Figure 2B and Supplementary Figure 3). However, due to the very low level signal of *APOBEC3A* and *APOBEC3A/APOBEC3B*, we were not able to reliably evaluate their expression with ddPCR. Therefore, in the next step, we used a UPL-qPCR technique that has a higher dynamic range than ddPCR. With the use of UPL-qPCR, we confirmed the negative correlation between the *APOBEC3B* deletion genotype and *APOBEC3B* expression (R=-0.82, p=0.002) and revealed a negative and positive correlation between the *APOBEC3B* deletion genotype and the expression of *APOBEC3A* (R=-0.70, p=0.015) and the expression of *APOBEC3A/APOBEC3A* (R=0.67, p=0.016), respectively (Figure 2B). As shown in Figure 2B, the expression of *APOBEC3A* and *APOBEC3A/APOBEC3B* is very low and is, respectively, ~130-fold and >1500-fold lower than the expression of *APOBEC3B*. Surprisingly, our analysis revealed that *APOBEC3A/APOBEC3B* hybrid expression is substantially lower (~12-fold) than the expression of canonical *APOBEC3A*, which indicates the role of the 3'UTR in the differential regulation of these genes and that *APOBEC3A/APOBEC3B* does not compensate for the lack of *APOBEC3A* dosage.

Analysis of the association of the *APOBEC3B* deletion with breast and ovarian cancer risk

For association analysis, we used several case-control set-ups from three different cohorts, i.e., GDANSK (523 BC-cases; 343 OC-cases; 853 controls), SZCZECIN (2009 BC-cases; 2005 controls; 615 NH-controls), and VILNIUS (97 OC-cases; 209 controls) (for details see Materials and Methods). The size of cumulative breast cancer case-control groups was estimated based on the frequency of the deletion in the European population (~11-13% - based on our preliminary results and [18]) and expected effect (OR~1.3 – estimated based on previous studies, i.e., [17, 31]) of the *APOBEC3B* deletion in order to obtain adequate statistical power (>90%) of the analysis. All samples were genotyped using a simple single tube A3Bdel_PCR assay (for details see "Design of the A3Bdel_ PCR assay and comprehensive analysis of the structure of the *APOBEC3B* deletion" and Materials and Methods). The distribution of the *APOBEC3B* deletion in control samples in all three groups was in good agreement with that expected under Hardy–Weinberg equilibrium (HWE) (p>>0.05), which indicates the high quality of the obtained genotyping results and the lack of bias in the detection of homozygous and heterozygous deletions. Logistic regression was used to derive odds ratios (ORs) and 95% confidence intervals (CIs) for the associations between the *APOBEC3B* deletion and cancer risk.

Numbers and frequencies of alleles and genotypes identified in the case and control groups are summarized in Table 1 (breast cancer association study) and Table 2 (ovarian cancer association study). In the association analyses, we focused mostly on the dominant model of inheritance (A3B+/− and A3B−/− vs. A3B+/+) (Table 1 and Table 2), but we also performed association analysis assuming additive and recessive models of inheritance (Supplementary Table 2 and Supplementary Table 3). The latter two models have much lower statistical power due to the low frequency (≤1%) of homozygous deletions (A3B−/− genotypes).

As shown in Table 1, the frequency of the *APOBEC3B* deletion in breast cancer cases does not significantly differ from that observed in population controls in GDANSK[OR(95%CI)=1.31(0.94-1.82), p=0.11)] or SZCZECIN [OR(95%CI)=0.89(0.74-1.07), p=0.20]. It also does not differ from cancer-free controls (SZCZECIN-NH controls; negative for any cancer and negative for family history of cancer) [OR(95%CI)=1.03(0.78-1.36), p=0.83]. To increase the power of the analysis, we combined SZCZECIN and GDANSK cohorts but observed no association [OR(95%CI)=0.98(0.83-1.15), p=0.79]. Adjusting for the origin of the samples did not significantly influence the results. We also did not see an association of the *APOBEC3B* deletion with familial breast cancer cases, which were selected from the GDANSK and SZCZECIN breast cancer cases [OR (95%CI)=1.06(0.82-1.36), p=0.67].

The *APOBEC3B* deletion also does not show association with ovarian cancer in either GDANSK [OR(95%CI)=0.77(0.50-1.19), p=0.24] or VILNIUS [OR(95%CI)=0.66(0.27-1.61), p=0.36]. It also does not associate with ovarian cancer in a combined GDANSK/VILNIUS cohort [OR(95%CI)=0.75(0.51-1.10), p=0.14]. Adjusting for the origin of the samples did not significantly influence the results.

Meta-analysis of association studies of the *APOBEC3B* deletion with cancer

The association of the *APOBEC3B* deletion with cancer has been analyzed within eleven case-control studies conducted in populations of different ethnicities,

Table 1: Analysis of the association of the *APOBEC3B* deletion with breast cancer risk using dominant model of inheritance (A3B+/− and A3B−/− vs. A3B+/+)

group	genotypes	No of cases (%)	No of controls (%)	OR(95%CI)	adjusted OR(95%CI)
GDANSK BC cases (n=523) vs. unselected controls (n=853)	A3B+/+	450 (86.04%)	759 (88.98%)	1.31(0.94-1.82) p=0.11	-
	A3B+/−	71 (13.58%)	91 (10.67%)		
	A3B−/−	2 (0.38%)	3 (0.35%)		
	A3B+/− and A3B−/−	73 (13.96%)	94 (11.02%)		
SZCZECIN BC cases (n=2009) vs. unselected controls (n=2005)	A3B+/+	1764 (87.80%)	1733 (86.43%)	0.89(0.74-1.07) p=0.20	0.90(0.75-1.08)[a] p=0.26
	A3B+/−	235 (11.70%)	267 (13.32%)		
	A3B−/−	10 (0.50%)	5 (0.25%)		
	A3B+/− and A3B−/−	245 (12.20%)	272 (13.57%)		
SZCZECIN BC cases (n=2009) vs. NH-controls (n=615)	A3B+/+	1764 (87.80%)	542 (88.13%)	1.03(0.78-1.36) p=0.83	0.95(0.70-1.29)[a] p=0.73
	A3B+/−	235 (11.70%)	68 (11.06%)		
	A3B−/−	10 (0.50%)	5 (0.81%)		
	A3B+/− and A3B−/−	245 (12.20%)	73 (11.87%)		
GDANSK+SZCZECIN BC cases (n=2532) vs. unselected controls (n=2858)	A3B+/+	2214 (87.44%)	2492 (87.19%)	0.98(0.83-1.15) p=0.79	0.97(0.83-1.14)[b] p=0.73
	A3B+/−	306 (12.09%)	358 (12.53%)		
	A3B−/−	12 (0.47%)	8 (0.28%)		
	A3B+/− and A3B−/−	318 (12.56%)	366 (12.81%)		
GDANSK+SZCZECIN familial BC cases (n=640) vs. unselected controls (n=2858)	A3B+/+	554 (86.56%)	2492 (87.19%)	1.06(0.82-1.36) p=0.67	1.15(0.87-1.52)[b] p=0.32
	A3B+/−	84 (13.13%)	358 (12.53%)		
	A3B−/−	2 (0.31%)	8 (0.28%)		
	A3B+/− and A3B−/−	86 (13.44%)	366 (12.81%)		

[a]adjusted for age; [b]adjusted for the origin of the study

different cancer types and different sizes [published and available in PubMed up to February 2017 [17, 25, 31, 32, 34, 36–41]. It has to be noted that the results obtained in the analyses are inconsistent and even conflicting. To obtain a more global view on the effect of the *APOBEC3B* deletion on cancer, we conducted a comprehensive meta-analysis that considered all case-control studies of the deletion (including our own) performed using cases with different cancer types (predominantly breast cancer). In total, 17637 cases and 19387 controls were enrolled in our meta-analysis (not including single studies of the deletion association with cervical, oral [38] and hepatocellular [32] cancer) (Figure 3A). As shown in Figure 3A, our meta-analysis revealed the consistent association of the *APOBEC3B* deletion with breast cancer in Asian populations [OR(95%CI)=1.367(1.282-1.458), p<0.001], but in European populations, the effect of the *APOBEC3B* deletion was much smaller and not significant [OR(95%CI)=1.102(0.995-1.221), p=0.063]. A single study of a north African population also did not show an association of the *APOBEC3B* deletion with breast cancer. Consequently, meta-analysis in the general population showed significant but modest association of the *APOBEC3B* deletion with breast cancer [OR(95%CI)=1.193(1.055-1.348), p=0.005]. Additionally, our meta-analysis revealed the lack of the association of the *APOBEC3B* deletion with ovarian cancer [OR(95%CI)=1.070(0.558-2.052), p=0.839] and opposite association (protective effect) of the deletion with bladder cancer [OR(95%CI)=0.834(0.734-0.948), p=0.005]. It has to be noted, however, that these latter two meta-analyses have much lower statistical power (each composed of only two studies). Additionally, Middlebrooks and colleagues [41] showed that the effect of the *APOBEC3B* deletion on bladder cancer is mostly driven by SNP rs1014971 [being in linkage disequilibrium (LD) with the deletion], and mostly disappears after adjustment for this SNP genotype [OR(95%CI)=0.88(0.72-1.07), p=0.21 and 0.96(0.79-1.16), p=0.67 in European and Japanese populations, respectively]. In the forest plot summarizing

Table 2: Analysis of the association of the *APOBEC3B* deletion with ovarian cancer risk using dominant model of inheritance (A3B+/− and A3B−/− vs. A3B+/+)

group	genotypes	No of cases (%)	No of controls (%)	OR(95%CI)	adjusted OR(95%CI)
GDANSK: OC cases (n=343) vs. unselected controls (n=853)	A3B+/+	313 (91.25%)	759 (88.98%)	0.77(0.50-1.19) p=0.24	-
	A3B+/−	28 (8.16%)	91 (10.67%)		
	A3B−/−	2 (0.58%)	3 (0.35%)		
	A3B+/− and A3B−/−	30 (8.75%)	94 (11.02%)		
VILNIUS: OC cases (n=97) vs. unselected controls (n=209)	A3B+/+	90 (92.78%)	187 (89.47%)	0.66(0.27-1.61) p=0.36	-
	A3B+/−	6 (6.19%)	22 (10.53%)		
	A3B−/−	1 (1.03%)	0 (0%)		
	A3B+/− and A3B−/−	7 (7.22%)	22 (10.53%)		
GDANSK + VILNIUS: OC cases (n=440) vs. unselected controls (n=1062)	A3B+/+	403 (91.59%)	946 (89.08%)	0.75(0.51-1.10) p=0.14	0.75(0.51-1.11)[b] p=0.15
	A3B+/−	34 (7.73%)	113 (10.64%)		
	A3B−/−	3 (0.68%)	3 (0.28%)		
	A3B+/− and A3B−/−	37 (8.41%)	116 (10.92%)		

[b]adjusted for the origin of the study

the *APOBEC3B* deletion association studies (Figure 3A), we also included the results of Revathidevi and colleagues [38] and Zhang and colleagues [32], which were only studies of cervical/oral and HBV-related hepatocellular cancer, respectively, and were therefore not included in the meta-analysis. The frequency of the deletion allele varies significantly across different ethnic groups/geographic regions. Drastic differences in the *APOBEC3B* deletion frequency among populations suggest that it is subjected to different selective pressures in human populations, it is functionally important, and it possesses the potential to modify phenotypes. It transpires that the worldwide distribution of the deletion frequency among control groups from studies enrolled in our meta-analysis resembles the distribution previously determined by Kidd and colleagues [18], with a mean frequency of 28.5% in Asia and 7.8% in Europe (worldwide frequency: 19.33%), which indicates that the studies enrolled in our meta-analysis are reliable and devoid of evident genotyping errors/biases (Figure 3B).

DISCUSSION

The two major reasons for the recent significant intensification of efforts on understanding the functional importance of the *APOBEC3B* gene are (i) the observation that activation of *APOBEC3B* in cancer leads to the generation of specific hypermutation signatures in breast and other cancer genomes [55, 57, 58] and (ii) the association of the *APOBEC3B* gene deletion with breast cancer risk [17, 25].

The aim of our study was to extend the current knowledge of the structure of the *APOBEC3B* deletion, its influence on the expression of the affected genes, and its association with breast and ovarian cancer predisposition in the European population. We determined the exact structure/breakpoints of the *APOBEC3B* deletion (1 nt resolution) and showed that although the breakpoints are located in highly homologous regions that may induce NAHR ([74, 75] and references within) and trigger recurrent occurrence of similar but not identical deletions, there is no sign of deletion heterogeneity. Even a small heterogeneity in the breakpoint positions would prevent the A3Bdel_PCR test from detecting the deletion and would cause discordance between A3Bdel_PCR and A3Bdel_MLPA results (not observed in our study in the panel of HapMap samples from European, African and Asiatic populations). This result strongly suggests that the deletion occurred in a single event (most likely in Africa) and then spread with the migration throughout the world, becoming common in European and Asiatic/Oceanic populations. Similar conclusions could not be derived from haplotype analysis due to low LD of flanking SNPs with the deletion [18]. We also delivered evidence of the presence of the hybrid *APOBEC3A*/*APOBEC3B* transcript in individuals carrying allele with the *APOBEC3B*

deletion and confirmed its suggested structure [18, 27] with 1 nt resolution. These findings allowed us to distinguish canonical *APOBEC3A* and *APOBEC3B* genes/transcripts from the hybrid *APOBEC3A/APOBEC3B* and develop several molecular assays, i.e., A3Bdel_PCR, A3Bdel_MLPA, and A3A_exp, A3B_exp and A3A/A3B_exp, which the scientific community can use in further analyses of the complex genomic region encompassing the *APOBEC3B* deletion. It has to be noted that assays A3A_exp, A3B_exp and A3A/A3B_exp allow for an efficient distinction of the highly homologous canonical (*APOBEC3A* and *APOBEC3B*) and hybrid *APOBEC3A/APOBEC3B* transcripts and constitute an important extension of the available qPCR assays (either UPL- or TaqMan/SYBR Green-based) [55, 64, 72, 76–79] dedicated only for the analysis of canonical transcripts. E.g., in contrast to the previously developed assays, our A3A/A3B_exp assay allows to detect *APOBEC3A/APOBEC3B* transcript not only if present alone in samples with the homozygous A3B−/− deletion (e.g., [64]) but also if present on background of the other genotypes. The accuracy of transcript distinguishing of the A3A_exp, A3B_exp, and A3A/A3B_exp assays was validated using sequence analysis of PCR products (e.g., Supplementary Figure 2, lower panel). The lack of signal of the A3A_exp and A3B_exp assays in samples with homozygous deletion and lack of signal of the A3A/A3B_exp assay in samples without deletion additionally confirm the specificity of the developed assays and unequivocally show that the assays allow distinguishing the canonical and hybrid transcripts. Additionally, the developed expression assays are free of missmapping/cross-hybridization effects, affecting previously reported microarray- and RNAseq-based results and hampering distinguishing the highly homologous *APOBEC3A* and *APOBEC3B* transcripts (see also Figure 2A). This problem was reported and described with more details previously [64]. The proposed assays are cost effective [several cents (PCR genotyping) to ~5$ (MLPA) per sample)], provide reliable results and can be utilized in large-scale association and functional studies of the *APOBEC3B* deletion.

In our study, we showed for the first time the association of the *APOBEC3B* deletion with the expression of all affected genes, including the *APOBEC3A/APOBEC3B* hybrid. We performed expression analysis using the ddPCR and UPL-qPCR methods in the lymphoblastoid cell lines (in total: 12) with the naturally occurring genotypes in a natural genetic background. Additionally, we supported our results with data extracted from genome-wide datasets. Our analysis showed that the *APOBEC3B* deletion negatively correlates with the expression of *APOBEC3A* and *APOBEC3B* and positively correlates with the expression of *APOBEC3A/APOBEC3B*. Additionally, it showed that the *APOBEC3B* expression level is much higher (>100x) than the expression of *APOBEC3A*, which is still higher (>10x) than the expression of the *APOBEC3A/APOBEC3B* hybrid. Much higher expression of the canonical *APOBEC3A* than of the *APOBEC3A/APOBEC3B* hybrid gene indicates that the loss of *APOBEC3A* expression that goes with the deletion is not compensated by the increased level of the hybrid. As *APOBEC3A* and *APOBEC3A/APOBEC3B* transcripts differ only by their 3'UTRs, it strongly supports the role of the 3'UTR in differential regulation of these two sister genes, suggesting the loss of a positive regulatory element or acquisition of a negative regulatory element (decreasing transcription efficiency or stability of hybrid transcript) specific for the 3'UTR of *APOBEC3B*. This result clearly contrasts with the results of Caval et al. [73] who with the use of artificially created model genes (constructed in plasmids and transfected into human cells) demonstrated increased expression of the transcript with the *APOBEC3B* 3'UTR. The discordance between our results and the results of Caval et al. [73] is most likely a consequence of the difference between natural and plasmid-based expression system and/or different types of cell lines used in the experiments. On the other hand, our expression results are generally inline with the results of Starrett et al. [64] who, also using artificial model (CRISPR/Cas9 generated deletion), showed high expression of *APOBEC3B* and very low but similar expression of *APOBEC3A* and *APOBEC3A/APOBEC3B*. The previous studies of the *APOBEC3B* deletion genotype-expression relationship (i) did not distinguish the canonical *APOBEC3A* and the hybrid *APOBEC3A/APOBEC3B* transcripts, and/or (ii) were conducted using artificially created cell line models with transfected reporter constructs, and/or (iii) were performed with the use of single tissue or cell line samples, and/or (iv) were often inconclusive, at least partially due to high homology between *APOBEC3A* and *APOBEC3B*, as well as other members of the *APOBEC3* gene family [32, 33, 40, 63, 64, 73, 78, 80, 81].

Unequivocal confirmation of the presence of hybrid transcript (generated from A3B- allele) may strengthen the notion/hypothesis that the APOBEC3A enzyme generated from the *APOBEC3A/APOBEC3B* hybrid transcript may play an important role in the induction of *kataegis* [73]. This notion is inline with observations showing that amount of somatic mutations occurring in APOBEC3-specific sequence context is on average higher in cancer samples with the homozygous *APOBEC3B* deletion [63, 64]. On the other hand, the very low expression level of *APOBEC3A* and even lower expression of the *APOBEC3A/APOBEC3B* hybrid argues against this hypothesis and may suggest that some other member of the APOBEC3 family may play the role in generation of the APOBEC3-dependent somatic mutations. Recently, it was shown that such a gene may be one of the several variants (haplotypes) of *APOBEC3H* (i.e., *APOBEC3H-I*), which occurrence across human population additionally correlates with the occurrence of the *APOBEC3B* deletion

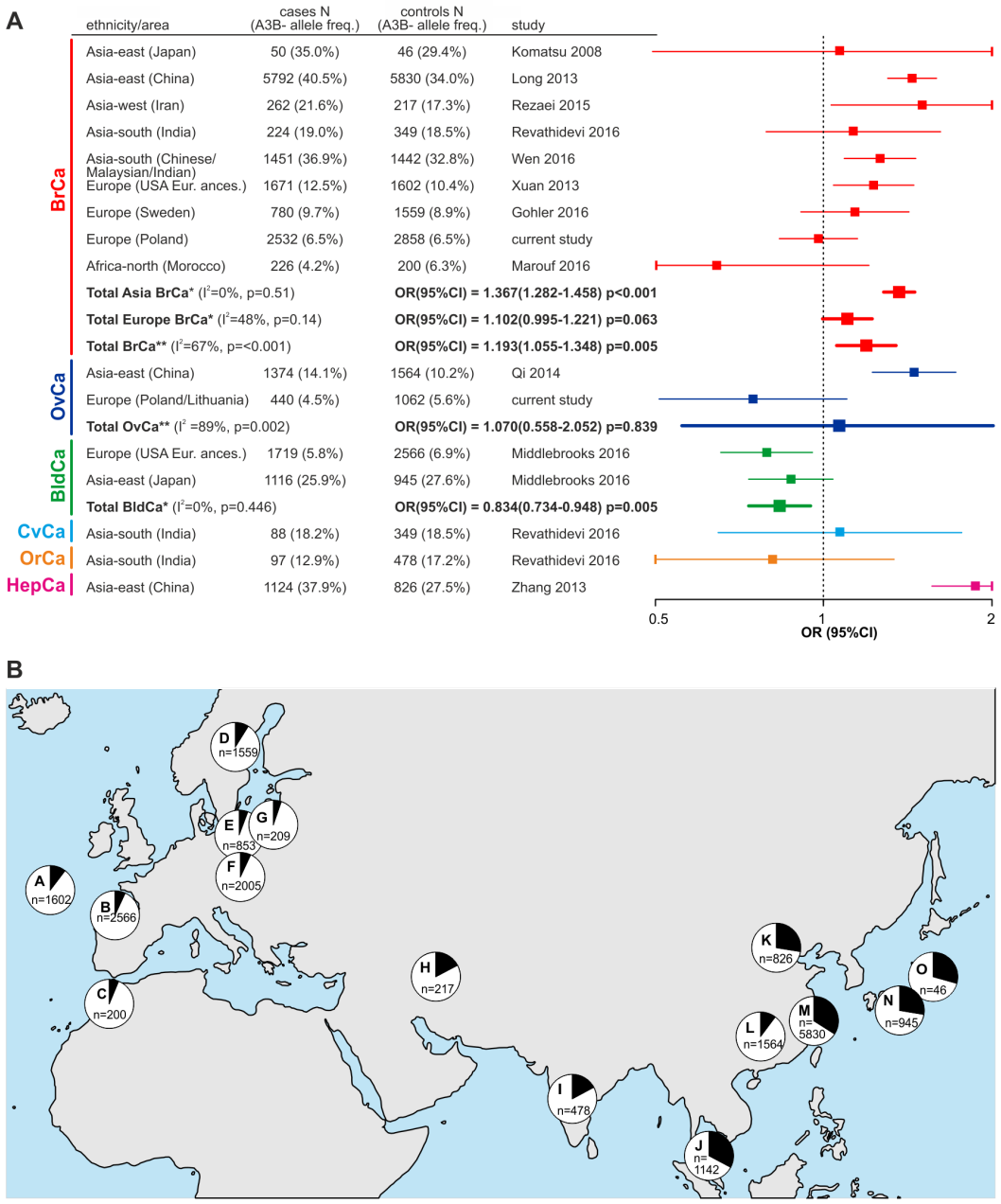


Figure 3: Meta-analysis of association studies of the *APOBEC3B* deletion with breast and other types of cancer. (A) Forest plot summarizing results of the meta-analysis. Characteristics of the studies used in the meta-analysis are shown on the left side of the plot. The plot illustrates the measure of the effect of the *APOBEC3B* deletion on predisposition to particular cancer types, i.e., OR (square) with the corresponding 95% CI (horizontal lines), observed in different studies enrolled in the meta-analysis. The meta-analysis was performed under the dominant model of inheritance (A3B+/+ and A3B+/- vs. A3B-/-). Weighted odds ratios with the corresponding 95%Cis (squares with horizontal lines in bold) were obtained using Mantel-Haenszel method under fixed* or random** effects models, depending on results of heterogeneity tests (p>0.10 for the Q test and I<50% were considered to indicate a lack of significant heterogeneity). The vertical dotted line indicates no effect (OR=1) of the *APOBEC3B* deletion on cancer predisposition. The OR values are plotted on a logarithmic scale to obtain symmetrical CIs and equivalent visualization of ORs of values greater and lower than 1. Horizontal lines corresponding to CIs that are out of range of 0.5-2 have been cut (vertical line) for visualization purposes. BrCa - breast cancer; OvCa - ovarian cancer; BldCa - bladder cancer; CvCa - cervical cancer; OrCa - oral cancer; HepCa - hepatocellular cancer. **(B)** Worldwide distribution of the allelic frequencies (black segments in pie charts) of the *APOBEC3B* deletion in control groups used in studies that were enrolled in the meta-analysis. Each letter indicates particular geographic region/population from a particular study, i.e., A - USA European ancestry [31]; B - Spain/USA European ancestry [41]; C - Morocco [39]; D - Sweden [37]; E - Poland (GDANSK) (current study); F - Poland (SZCZECIN) (current study); G - Lithuania (VILNIUS) (current study); H - Iran [36]; I - India [38]; J - Malaysia [40]; K - China [32]; L - China [34]; M - China [17]; N - Japan [41]; O - Japan [25]. "n" indicates the number of samples in the control groups in each study.

[64]. It has to be noted, however, that the above part of the discussion is very speculative and based on often conflicting pieces of evidence reported in different studies. It indicates that much more has to be done to explain the role of the *APOBEC3* genes and the *APOBEC3B* deletion in the induction of somatic mutations in cancer.

In the next step, we performed a large-scale association study of the *APOBEC3B* deletion with breast and/or ovarian risk, which encompassed either separate or appropriately combined analyses of three European cohorts, i.e., GDANSK, SZCZECIN and VILNIUS. Our association study comprised 2972 cases and 3682 controls; it was the largest *APOBEC3B* deletion association study performed in European populations and the first *APOBEC3B* deletion association study to use the familial form of breast cancer. Our analysis revealed the lack of association of the *APOBEC3B* deletion with breast cancer risk and additionally did not show association of the deletion with familial breast cancer risk. As familial breast cancer represents a more extreme phenotype, one could expect it to show a stronger effect of the association. Therefore, the lack of association of the *APOBEC3B* deletion with familial breast cancer additionally confirms no effect of the deletion in the European population. Similarly, we also show the lack of the association of the deletion with ovarian cancer. To obtain a more global view of the relationship between the *APOBEC3B* deletion and cancer predisposition, we performed a comprehensive multilayer meta-analysis of all available studies on the association of the *APOBEC3B* deletion with cancer predisposition conducted in different populations and geographic regions. Although the meta-analysis showed substantial heterogeneity in the results obtained by different groups and a moderate global effect of the *APOBEC3B* deletion, it revealed a consistent association of the *APOBEC3B* deletion with breast cancer in Asian populations but a lack of this association in European populations. It has to be noted, however, that the association in Asian populations is driven mostly by a large seminal study [17]. The general overlap of the geographical distribution of the *APOBEC3B* deletion frequency observed in the meta-analyzed association studies with that observed before [18] exclude substantial genotyping inaccuracy in the association studies as the main source of the discordances of the *APOBEC3B* deletion effect in Asiatic and European populations. Therefore, we believe that the main source of the discordances may be differences in genetic background, such as the presence of some other causative genetic variants that in Asiatic populations share the haplotype with the *APOBEC3B* deletion but are absent or not in LD with the deletion in European populations. Some support for this hypothesis may be the analysis of HapMap Phase I SNPs that revealed different pattern of LD in regions flanking the deletion in European and Asiatic populations [18]. This analysis showed SNPs in regions directly adjacent to the deletion being in moderate LD with the deletion in Asiatic but not in European populations. Additionally, analysis of the Northern European population showed a lack of SNPs in strong LD with the deletion [37]. Later, analysis of 1000 Genomes Project data identified only one surrogate SNP, i.e., rs12628403, that showed complete LD with the deletion in a European population, lower LD in a Chinese population, and very low LD in an African population [17, 41]. Additionally, the presence of interfering environmental factors specific to particular geographic regions/populations cannot be excluded. The observed drastic differences of the frequency of the *APOBEC3B* deletion in different populations strongly support the notion of an interaction between environmentally driven selective pressure and the deletion. The function of APOBEC3B that may play a role in selective pressure may be its potential involvement in response against viral infections, e.g., HBV or HTLV1 infection [32, 44–46], or its suggested role in innate immunity against malaria [29]. However, there are some controversies and a lack of consensus on the involvement of APOBEC3B in the restriction of particular types of viruses, e.g., HIV1 (e.g., [26, 28, 33, 42]; reviewed in [20]).

Our study is not free of limitations. Our association analysis involved relatively small numbers of samples from women with ovarian cancer, which limits its power to detect potential associations with ovarian cancer risk. The performed meta-analysis is very heterogeneous in terms of proband ethnicity, geographical regions, cancer types and cancer characteristics. Additionally, some ethnic groups or geographical regions are not represented at all; e.g., there is only one small study of an African population (including no study of African-Americans) and no study of South American populations. It limits the power of some more specific (e.g., ethnic-specific) conclusions. Some studies included in the meta-analysis are very small (<<500 samples) with limited statistical power. Moreover, our expression analyses were performed with the use of only one type of cell lines. We cannot exclude that the *APOBEC3B* deletion effect may be different in different cell lines or tissues.

In conclusion, in this study, we determined the exact breakpoints of the *APOBEC3B* deletion (1 nt resolution) and confirmed its homogeneity. Additionally, we provided direct evidence for the generation of the transcriptionally active hybrid gene *APOBEC3A/APOBEC3B* from the allele with the *APOBEC3B* deletion and confirmed the suggested structure of *APOBEC3A/APOBEC3B*, which allowed us to distinguish *APOBEC3A*, *APOBEC3B*, and *APOBEC3A/APOBEC3B* expression levels. For the first time, we showed the association of the *APOBEC3B* deletion with the expression of all affected genes, including the *APOBEC3A/APOBEC3B* hybrid. We observed that the expression of *APOBEC3A/APOBEC3B* is ~10x lower than the expression of *APOBEC3A*, which

implies the role of the 3'UTR in the differential regulation of these two genes coding for the same protein. We showed a lack of association of the *APOBEC3B* deletion with breast and/or ovarian risk (for the first time including familial breast cancer), which was independently validated in three European cohorts (in total: 2972 cases and 3682 controls). We also performed a comprehensive summary/visualization of all available reports on the association of the *APOBEC3B* deletion with cancer predisposition, which was obtained in our meta-analysis of association studies performed in various populations and geographic regions. It also has to be noted that within our study, we developed a variety of molecular assays that can be used for further analysis of the complex genomic region encompassing the *APOBEC3B* deletion.

MATERIALS AND METHODS

DNA samples

To compare the *APOBEC3B* deletion genotype with the expression of the affected genes, we used 270 reference DNA samples from the basic (phaseI/II) HapMap panel, including (i) 90 African samples from the Yoruba in Ibadan, Nigeria (YRI); (ii) 90 samples from individuals from Utah, USA, from the Centre d'Etude du Polymorphisme Humain collection (CEU); (iii) 45 samples from Han Chinese in Beijing, China (CHB); (iv) 45 samples from Japanese in Tokyo, Japan (JPT). All DNA samples were purchased from Coriell Institute for Medical Research (NJ, USA).

The *APOBEC3B* deletion association study was performed using genomic DNA samples from case-control groups collected at the Medical University of Gdansk (GDANSK group), at the International Hereditary Cancer Center in Szczecin (SZCZECIN group) and at Vilnius University Hospital Santariskiu Klinikos, Lithuania (VILNIUS group). All subjects from GDANSK, SZCZECIN, and VILNIUS provided informed written consent, and the study was approved by the appropriate local ethic committees. All subjects were Caucasians of European ancestry. Subjects from GDANSK and SZCZECIN were ethnically Poles. Subjects from VILNIUS were of mixed ethnicity, composed mostly of Lithuanians (~60%) but also Poles and Russians, with similar fractions of the ethnicities in case and control groups. In all groups, the control samples were derived from the same geographical region as the case samples.

The GDANSK group comprised 523 BC-cases [women with breast cancer from families with breast and/or ovarian cancer aggregation, as defined previously [82], negative for the 5 most common *BRCA1* mutations in the Polish population, c.68_69delAG, c.181T> G, c.3700_3704del5, c.4035delA, c.5266dupC [83, 84], 343 OC-cases (women with ovarian cancer unselected in terms of the familial history of the disease), and 853 controls (unselected population control samples).

The SZCZECIN group comprised 2009 BC-cases (women with breast cancer, unselected for familial history of the disease) and two distinct sets of control samples: 2005 SZCZECIN controls (cancer-free unselected women) and 615 SZCZECIN NH-controls (cancer-free women with a negative family history of cancer). Histological subtypes were determined for 1397 SZCZECIN BC-cases, of which 1117 (79.96%) were invasive *ductal carcinoma*, 128 (9.16%) were invasive *lobular carcinoma*, 47 (3.36%) were *ductal carcinoma* in situ, 24 (1.72%) were *carcinoma medullare*, 22 (1.57%) were *carcinoma mucinosum*, and 1-17 (0.07-1.22%) were of other subtypes of breast cancer. Breast cancer grade was determined for 1047 SZCZECIN BC-cases, of which 737 (70.39%) and 310 (29.61%) were classified as G1/G2 and G3, respectively. Status of estrogen receptor (ER) was determined for 1372 SZCZECIN BC-cases, of which 416 (30.32%) and 956 (69.68%) were ER negative and ER positive, respectively. Status of progesterone receptor (PgR) was determined for 1365 SZCZECIN BC-cases, of which 450 (32.97%) and 915 (67.03%) were PgR negative and PgR positive (PgR+), respectively. Status of human epidermal growth factor receptor 2 (HER2) was determined for 1291 SZCZECIN BC-cases, of which 1037 (80.33%) and 254 (19.67%) were HER2 negative and HER2 positive, respectively. Triple-negative (TN) status of breast cancer was determined for 218 SZCZECIN BC-cases (17.30% of cases in which all three receptors were tested). Information about familial history of cancer was recorded for 1590 cases, of which 117 (7.36%) were classified as familial breast cancer cases (according to criteria previously defined in [85]). The mean age of breast cancer diagnosis in SZCZECIN BC-cases was 54 years (range 21-92), with 187 (9.31%), 712 (35.44%) and 1110 (55.25%) women diagnosed before age 40, between the ages 40 and 50, and after age 50, respectively. A total of 52 (2.6%) SZCZECIN BC-cases were positive for one of three major *BRCA1* founder mutations in the Polish population, i.e., c.181T>G, c.4035delA, and c.5266dupC. SZCZECIN controls included women with a mean age of 52 years (range 17-94), and SZCZECIN NH-controls included women with a mean age of 67 years (range 36-95).

The VILNIUS group comprised 97 OC-cases (women with unselected ovarian cancer) and 209 controls (samples from unselected control population). Mean age of ovarian cancer diagnosis in VILNIUS OC-cases was 53 years (range 18-78). VILNIUS controls comprised women with mean age of 44 years (range 18-85).

Analysis of the structure of the *APOBEC3B* deletion

Sequencing analyses were performed using ABI Prism 3130XL apparatus (Applied Biosystems, Carlsbad,

CA, USA). The obtained sequences were analyzed with the use of Finch TV (v.1.4.0) (Geospiza Inc.).

MLPA analysis was performed using the in-house designed A3Bdel_MLPA assay. The probe-set layout was designed according to a previously proposed [65, 66] and well-validated strategy (e.g., [67, 86–88]). This strategy exclusively utilized short oligonucleotide probes that can easily be generated via standard chemical synthesis. Each probe consists of two half-probes of equal size, and the total probe length ranges from 93 to 128 nt. The target sequences for the probes were selected to avoid single nucleotide polymorphisms (SNPs), insertions/deletions, and sequences of extremely low or high guanosine-cytosine (GC) content. The sequences and detailed characteristics of all probes are shown in Supplementary Table 1. The MLPA probes were synthesized by IDT (Skokie, IL, USA). The MLPA reactions were run according to the manufacturer's general recommendations and previously published [65, 89], with reagents purchased from MRC-Holland, Amsterdam, The Netherlands. The products of the MLPA reactions were diluted 20× in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated by size with capillary electrophoresis (POP7 polymer; ABI Prism 3130XL apparatus; Applied Biosystems, Carlsbad, CA, USA). The obtained electropherograms were analyzed using GeneMarker software (version 2.2.0; SoftGenetics, State College, PA, USA). The normalized signal of each probe (peak height divided by the average peak height of the control probes) was divided by the corresponding signal of a reference probe and multiplied by 2. The obtained values that correspond to the copy number of particular regions were visualized in bar graphs.

Analysis of the *APOBEC3B* deletion genotype–expression relationship

Publically available genome-wide mRNA profiling data of different panels of HapMap samples that were used in our study are deposited at Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo) database (series accession number: GSE6536), and Array Express database (https://www.ebi.ac.uk/arrayexpress/; accession number: E-MTAB-197 and E-MTAB-198) as well as at http://jungle.unige.ch/rnaseq_CEU60/ (file with normalized data "RNASEQ60_array_rep_expr.txt.gz"). The *APOBEC3B* deletion genotypes of the analyzed HapMap samples were determined before [18] and independently confirmed in our study.

Human lymphoblastoid cell lines (LCLs) GM18532, GM18537, GM18540, GM18542, GM18561, GM18570, GM18572, GM18573, GM18577, GM18579, GM18603, and GM18612 from the HapMap panel were purchased from the Coriell Institute for Medical Research (USA; http://www.coriell.org). All cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM; Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich, St. Louis, MO, USA), 1% Glutamax (Cellgro, Mediatech Inc., Manassas, VA, USA), 1x antibiotic/antimycotic solution (Sigma-Aldrich), and 1x MEM non-essential amino acid solution (Sigma-Aldrich). Mycoplasma infection in cell cultures was controlled with the use of the MycoFluor™ Mycoplasma Detection Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol.

Total RNA was extracted from cells with TRIzol (Invitrogen) according to the manufacturer's instructions. Equal amounts of RNA (2 μg) were reverse-transcribed with random primers (Invitrogen) using SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol.

ddPCR was performed according to the manufacturer's general recommendations (Bio-Rad, CA, USA). ddPCR reactions (containing 1× EvaGreen Supermix, 200 nM primers, one-twentieth of cDNA to a final volume of 20 μL) were mixed with seventy microliters of droplet generation oil and used to form droplets in a QX200 droplet generator. The partitioned emulsion was then slowly transferred to a 96-well PCR plate (Eppendorf, Hamburg, Germany). After being heat-sealed with foil, the plates containing the droplets were PCR cycled to final point under conditions at 95°C for 5 min, then 95°C for 30 s, and 56°C for 30 s, and 72°C for 45 s for 40 cycles, 72°C for 2 min, 4°C for 5 min, 90°C for 5 min, then held at 12°C. Following PCR, samples were read on a Bio-Rad QX200 reader and data were analyzed using Quantasoft software v.1.7.4.0917 (Bio-Rad, CA, USA). Primers used in the ddPCR analyses were designed at the exon-exon junctions or in different, adjacent exons spanning a long intron to avoid amplification of the potential gDNA traces. For each analyzed cDNA sample, the following sets of PCR primers were used: (i) A3B_exp, i.e., test amplicon for *APOBEC3B*: Fwd primer 5'GACCTACGATGAGTTTGAGT3', Rev primer 5'TTAGAGACTGAGGCCCAT3' (amplicon length: 163 bp); (ii) A3A_exp, i.e., test amplicon for *APOBEC3A*: Fwd primer 5'CATTCTCCAGAATCAGGG3', Rev primer 5'CTTGATCGGGAGCATAC 3' (amplicon length: 170 bp); (iii) A3A/A3B_exp, i.e., test amplicon for *APOBEC3A/APOBEC3B* hybrid: Fwd primer 5'TGACCTACGATGAATTTAAGC3', Rev primer 5'ATCTACTTGATCAGGAGCAC3' (amplicon length: 293 bp), (iv) reference amplicon for *GAPDH*: Fwd primer 5'CACCACCAACTGCTTAGC3', Rev primer 5'CATGGACTGTGGTCATGAG3' (amplicon length: 87 bp).

UPL-qPCR analyses were performed with the use of LightCycler 480 system with probes from the Universal Probe Library (UPL) (Roche Applied Science, Penzberg, Germany), according to the manufacturer's protocol. For test amplicons, the same primer sets were used as in the ddPCR analyses. The following set of PCR primers was used as a reference amplicon for *ACTB*: Fwd

primer 5'CCAACCGCGAGAAGATGA3', Rev primer 5'CCAGAGGCGTACAGGGATAG3' (amplicon length: 97 bp). The applied primers-UPL probes combinations were as follows: (i) *APOBEC3B* test amplicon, probe #11; (ii) *APOBEC3A* test amplicon, probe #13; (iii) *APOBEC3A/APOBEC3B* hybrid test amplicon, probe #13; (iv) *ACTB* reference amplicon, probe #64. qPCR assays were run in triplicates for 45 cycles and normalized tested gene expression level was calculated by the formula: $2^{-\Delta Ct}$, where $\Delta Ct = Ct_{\text{test amplicon}} - Ct_{\text{reference amplicon}}$, using the LightCycler 480 instrument software.

A3Bdel_PCR assay for genotyping of the *APOBEC3B* deletion

Genotyping of the *APOBEC3B* deletion was performed with the use of A3Bdel_PCR assay that consists of three PCR primers, i.e., one forward primer F: 5'CCTGTCCCTTTTCAGAATTTAAGC3', and two reverse primers: R1: 5'CTTGATCGGGAGCATAC3' (complementary to reference allele; F+R1 amplicon length: 572 bp), and R2: 5'TGGAGCCAATTAATCACTTCAT3' (complementary to deletion allele, F+R2 amplicon length: 707 bp). All three primers were used simultaneously in each single reaction. PCR was performed in a 6.25-µl reactions composed of 0.3 µl of 10 µM dilution of each primer (synthesized by Sigma-Aldrich), 0.125 µl of dNTPs mix (Promega, Madison, WI, USA), 0.05 µl of GoTaq DNA Polymerase (concentration 5 u/µl) (Promega), 1.25 µl of 5X colorless GoTaq reaction buffer (containing 7.5 mM $MgCl_2$) (Promega), 2.925 µl of deionized water, and 1 µl of DNA (~30 ng). The following cycling conditions were used: 2 min at 95°C, followed by 35 cycles at 95°C for 20 sec, 56°C for 20 sec, and 72°C for 5 min, followed by 5 min at 72°C. The obtained PCR products were visualized on a standard 1.5% agarose gel.

Statistical methods

The *APOBEC3B* deletion genotype-expression relationship was analyzed with the use of one-way analysis of variance (ANOVA) with post-test for linear trend, and t-test using Prism v. 4.0 (GraphPad, San Diego, CA).

HWE was assessed in the control groups using the chi-square test. Associations between the *APOBEC3B* deletion and breast or ovarian cancer risk were assessed using ORs and 95%CIs derived from logistic regression models. ORs(95%CIs) were estimated in the analyses assuming different models of inheritance, i.e., dominant (A3B+/− and A3B−/− vs. A3B+/+), recessive (A3B−/− vs. A3B+/− and A3B+/+), and additive (A3B−/− vs. A3B+/− vs. A3B+/+) models. Appropriate adjustments for the origin of the study and age (when data regarding age were available for particular case-control groups) were applied. Meta-analysis was performed with the use of Mantel-Haenszel method assuming dominant model of inheritance. MedCalc Statistical Software version 14.8.1 (MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org) was used for all logistic regression analyses and meta-analysis. All statistical tests were two-sided, and a p-value less than 0.05 was considered statistically significant.

Author contributions

KK - designed and optimized MLPA test and performed MLPA analyses, developed genotyping test and conducted genotyping analyses, designed expression assay and performed genotype-expression analyses, performed all statistical analyses and meta-analysis, participated in manuscript preparation, including tables, figures and supporting materials; WK - performed genotyping of samples from the SZCZECIN cohort; BR - performed genotyping of samples from the SZCZECIN cohort; AJ - participated in the SZCZECIN samples characterization/selection and participated in manuscript preparation; MR - characterized samples from the GDANSK group, participated in manuscript preparation, coordinated the Poznan-Gdansk cooperation; NK - participated in genotyping of samples from the VILNIUS cohort; DV - provided, characterized and genotyped samples from the VILNIUS cohort; KC - performed cell cultures and RNA extraction from the lymphoblastoid HapMap cell lines; MW - performed cell cultures and RNA extraction the lymphoblastoid HapMap cell lines; CC - coordinated the Poznan-Szczecin cooperation and genotyping analysis of samples from the SZCZECIN cohort and participated in manuscript preparation; JL - founder of Szczecin International Hereditary Cancer Centre (IHCC) samples collection, selected samples for analysis, discussed the concept of manuscript; PK - conceived and coordinated the study and collaboration with the Gdansk and Szczecin research groups, supervised all analyses, prepared (with KK) the manuscript including tables, figures and supporting materials.

ACKNOWLEDGMENTS

We thank the Laboratory of Subcellular Structures Analysis of the Institute of Bioorganic Chemistry, Polish Academy of Sciences in Poznan, for the use LightCycler 480 system (Roche). All capillary electrophoresis separations were conducted in Laboratory of Molecular Biology Techniques, Faculty of Biology, Adam Mickiewicz University, Poznan.

CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

FUNDING

This work was supported by research grants 2015/19/N/NZ5/02247 (recipient: KK) and 2015/17/B/

NZ2/01182 (recipient: PK) from the National Science Centre, Poland (http://www.ncn.gov.pl/) and KNOW program of the Polish Ministry of Science and Higher Education.

REFERENCES

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000; 343:78-85.
2. Beggs AD, Hodgson SV. Genomics and breast cancer: the different levels of inherited susceptibility. Eur J Hum Genet. 2009; 17:855-856.
3. Ripperger T, Gadzicki D, Meindl A, Schlegelberger B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. Eur J Hum Genet. 2009; 17:722-731.
4. Melchor L, Benitez J. The complex genetic landscape of familial breast cancer. Hum Genet. 2013; 132:845-863.
5. Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell. 2007; 11:103-105.
6. Foulkes WD. Inherited susceptibility to common cancers. N Engl J Med. 2008; 359:2143-2153.
7. Eccles SA, Aboagye EO, Ali S, Anderson AS, Armes J, Berditchevski F, Blaydes JP, Brennan K, Brown NJ, Bryant HE, Bundred NJ, Burchell JM, Campbell AM, et al. Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast Cancer Res. 2013; 15:R92.
8. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353-361, 361e351-352.
9. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015; 47:373-380.
10. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747-753.
11. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005; 307:1434-1440.
12. Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS, Xu XH, Yan H, Liu X, et al. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am J Hum Genet. 2008; 83:663-674.
13. Bentley RW, Pearson J, Gearry RB, Barclay ML, McKinney C, Merriman TR, Roberts RL. Association of higher DEFB4 genomic copy number with Crohn's disease. Am J Gastroenterol. 2010; 105:354-359.
14. Celestino-Soper PB, Shaw CA, Sanders SJ, Li J, Murtha MT, Ercan-Sencicek AG, Davis L, Thomson S, Gambin T, Chinault AC, Ou Z, German JR, Milosavljevic A, et al. Use of array CGH to detect exonic copy number variants throughout the genome in autism families detects a novel deletion in TMLHE. Hum Mol Genet. 2011; 20:4360-4370.
15. Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, Santos TM, Goncalves A, Nobrega AF, Pearson PL, Vianna-Morgante AM, Carraro DM, Brentani RR, et al. Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res. 2012; 14:R24.
16. Suehiro Y, Okada T, Shikamoto N, Zhan Y, Sakai K, Okayama N, Nishioka M, Furuya T, Oga A, Kawauchi S, Maeda N, Tamesa M, Nagashima Y, et al. Germline copy number variations associated with breast cancer susceptibility in a Japanese population. Tumour Biol. 2013; 34:947-952.
17. Long J, Delahanty RJ, Li G, Gao YT, Lu W, Cai Q, Xiang YB, Li C, Ji BT, Zheng Y, Ali S, Shu XO, Zheng W. A common deletion in the APOBEC3 genes and breast cancer risk. J Natl Cancer Inst. 2013; 105:573-579.
18. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE. Population stratification of a common APOBEC gene deletion polymorphism. PLoS Genet. 2007; 3:e63.
19. Conticello SG. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 2008; 9:229.
20. Moris A, Murray S, Cardinaud S. AID and APOBECs span the gap between innate and adaptive immunity. Front Microbiol. 2014; 5:534.
21. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov. 2015; 5:704-712.
22. Rebhandl S, Huemer M, Greil R, Geisberger R. AID/APOBEC deaminases and cancer. Oncoscience. 2015; 2:320-333.
23. Henderson S, Fenton T. APOBEC3 genes: retroviral restriction factors to cancer drivers. Trends Mol Med. 2015; 21:274-284.
24. Burns MB, Leonard B, Harris RS. APOBEC3B: pathological consequences of an innate immune DNA mutator. Biomed J. 2015; 38:102-110.
25. Komatsu A, Nagasaki K, Fujimori M, Amano J, Miki Y. Identification of novel deletion polymorphisms in breast cancer. Int J Oncol. 2008; 33:261-270.
26. An P, Johnson R, Phair J, Kirk GD, Yu XF, Donfield S, Buchbinder S, Goedert JJ, Winkler CA. APOBEC3B

deletion and risk of HIV-1 acquisition. J Infect Dis. 2009; 200:1054-1058.

27. Abe H, Ochi H, Maekawa T, Hatakeyama T, Tsuge M, Kitamura S, Kimura T, Miki D, Mitsui F, Hiraga N, Imamura M, Fujimoto Y, Takahashi S, et al. Effects of structural variations of APOBEC3A and APOBEC3B genes in chronic hepatitis B virus infection. Hepatol Res. 2009; 39:1159-1168.
28. Itaya S, Nakajima T, Kaur G, Terunuma H, Ohtani H, Mehra N, Kimura A. No evidence of an association between the APOBEC3B deletion polymorphism and susceptibility to HIV infection and AIDS in Japanese and Indian populations. J Infect Dis. 2010; 202:815-816; author reply 816-817.
29. Jha P, Sinha S, Kanchan K, Qidwai T, Narang A, Singh PK, Pati SS, Mohanty S, Mishra SK, Sharma SK, Awasthi S, Venkatesh V, Jain S, et al. Deletion of the APOBEC3B gene strongly impacts susceptibility to falciparum malaria. Infect Genet Evol. 2012; 12:142-148.
30. Ezzikouri S, Kitab B, Rebbani K, Marchio A, Wain-Hobson S, Dejean A, Vartanian JP, Pineau P, Benjelloun S. Polymorphic APOBEC3 modulates chronic hepatitis B in Moroccan population. J Viral Hepat. 2013; 20:678-686.
31. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, Kelley MC, Zheng W, Long J. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. Carcinogenesis. 2013; 34:2240-2243.
32. Zhang T, Cai J, Chang J, Yu D, Wu C, Yan T, Zhai K, Bi X, Zhao H, Xu J, Tan W, Qu C, Lin D. Evidence of associations of APOBEC3B gene deletion with susceptibility to persistent HBV infection and hepatocellular carcinoma. Hum Mol Genet. 2013; 22:1262-1269.
33. Imahashi M, Izumi T, Watanabe D, Imamura J, Matsuoka K, Ode H, Masaoka T, Sato K, Kaneko N, Ichikawa S, Koyanagi Y, Takaori-Kondo A, Utsumi M, et al. Lack of association between intact/deletion polymorphisms of the APOBEC3B gene and HIV-1 risk. PLoS One. 2014; 9:e92861.
34. Qi G, Xiong H, Zhou C. APOBEC3 deletion polymorphism is associated with epithelial ovarian cancer risk among Chinese women. Tumour Biol. 2014; 35:5723-5726.
35. Prasetyo AA, Sariyatun R, Reviono, Sari Y, Hudiyono, Haryati S, Adnan ZA, Hartono, Kageyama S. The APOBEC3B deletion polymorphism is associated with prevalence of hepatitis B virus, hepatitis C virus, Torque Teno virus, and Toxoplasma gondii co-infection among HIV-infected individuals. J Clin Virol. 2015; 70:67-71.
36. Rezaei M, Hashemi M, Hashemi SM, Mashhadi MA, Taheri M. APOBEC3 Deletion is Associated with Breast Cancer Risk in a Sample of Southeast Iranian Population. Int J Mol Cell Med. 2015; 4:103-108.
37. Gohler S, Da Silva Filho MI, Johansson R, Enquist-Olsson K, Henriksson R, Hemminki K, Lenner P, Forsti A. Impact of functional germline variants and a deletion polymorphism in APOBEC3A and APOBEC3B on breast cancer risk and survival in a Swedish study population. J Cancer Res Clin Oncol. 2016; 142:273-276.
38. Revathidevi S, Manikandan M, Rao AK, Vinothkumar V, Arunkumar G, Rajkumar KS, Ramani R, Rajaraman R, Ajay C, Munirajan AK. Analysis of APOBEC3A/3B germline deletion polymorphism in breast, cervical and oral cancers from South India and its impact on miRNA regulation. Tumour Biol. 2016; 37:11983-11990.
39. Marouf C, Gohler S, Filho MI, Hajji O, Hemminki K, Nadifi S, Forsti A. Analysis of functional germline variants in APOBEC3 and driver genes on breast cancer risk in Moroccan study population. BMC Cancer. 2016; 16:165.
40. Wen WX, Soo JS, Kwan PY, Hong E, Khang TF, Mariapun S, Lee CS, Hasan SN, Rajadurai P, Yip CH, Mohd Taib NA, Teo SH. Germline APOBEC3B deletion is associated with breast cancer risk in an Asian multi-ethnic cohort and with immune cell presentation. Breast Cancer Res. 2016; 18:56.
41. Middlebrooks CD, Banday AR, Matsuda K, Udquim KI, Onabajo OO, Paquin A, Figueroa JD, Zhu B, Koutros S, Kubo M, Shuin T, Freedman ND, Kogevinas M, et al. Association of germline variants in the APOBEC3 region with cancer risk and enrichment with APOBEC-signature mutations in tumors. Nat Genet. 2016; 48:1330-1338.
42. Singh H, Marathe SD, Nain S, Nema V, Ghate MV, Gangakhedkar RR. APOBEC3B deletion impacts on susceptibility to acquire HIV-1 and its advancement among individuals in western India. APMIS. 2016; 124:881-887.
43. Han Y, Qi Q, He Q, Sun M, Wang S, Zhou G, Sun Y. APOBEC3 deletion increases the risk of breast cancer: a meta-analysis. Oncotarget. 2016; 7:74979-74986.
44. Ooms M, Krikoni A, Kress AK, Simon V, Munk C. APOBEC3A, APOBEC3B, and APOBEC3H haplotype 2 restrict human T-lymphotropic virus type 1. J Virol. 2012; 86:6097-6108.
45. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP. Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases *in vitro* and *in vivo*. Proc Natl Acad Sci U S A. 2005; 102:8321-8326.
46. Zhang W, Zhang X, Tian C, Wang T, Sarkis PT, Fang Y, Zheng S, Yu XF, Xu R. Cytidine deaminase APOBEC3B interacts with heterogeneous nuclear ribonucleoprotein K and suppresses hepatitis B virus expression. Cell Microbiol. 2008; 10:112-121.
47. Vartanian JP, Guetard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science. 2008; 320:230-233.
48. Warren CJ, Xu T, Guo K, Griffin LM, Westrich JA, Lee D, Lambert PF, Santiago ML, Pyeon D. APOBEC3A functions as a restriction factor of human papillomavirus. J Virol. 2015; 89:688-702.

49. Stenglein MD, Harris RS. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem. 2006; 281:16837-16841.

50. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C. APOBEC3 proteins inhibit human LINE-1 retrotransposition. J Biol Chem. 2006; 281:22161-22172.

51. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K. All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res. 2007; 35:2955-2964.

52. Carpenter MA, Li M, Rathore A, Lackey L, Law EK, Land AM, Leonard B, Shandilya SM, Bohn MF, Schiffer CA, Brown WL, Harris RS. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. J Biol Chem. 2012; 287:34801-34808.

53. Wijesinghe P, Bhagwat AS. Efficient deamination of 5-methylcytosines in DNA by human APOBEC3A, but not by AID or APOBEC3G. Nucleic Acids Res. 2012; 40:9206-9217.

54. Fu Y, Ito F, Zhang G, Fernandez B, Yang H, Chen XS. DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. Biochem J. 2015; 471:25-35.

55. 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.

56. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, Kim J, Kwiatkowski DJ, Fargo DC, Mieczkowski PA, Getz G, Gordenin DA. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet. 2015; 47:1067-1072.

57. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell. 2012; 149:979-993.

58. Roberts SA, Sterling J, Thompson C, Harris S, Mav D, Shah R, Klimczak LJ, Kryukov GV, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. Mol Cell. 2012; 46:424-435.

59. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat Genet. 2013; 45:977-983.

60. Leonard B, Hart SN, Burns MB, Carpenter MA, Temiz NA, Rathore A, Vogel RI, Nikas JB, Law EK, Brown WL, Li Y, Zhang Y, Maurer MJ, et al. APOBEC3B upregulation and genomic mutation patterns in serous ovarian carcinoma. Cancer Res. 2013; 73:7222-7231.

61. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet. 2013; 45:970-976.

62. Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife. 2013; 2:e00534.

63. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, Bolli N, Davies HR, Knappskog S, Martin S, Papaemmanuil E, Ramakrishna M, Shlien A, Simonic I, et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat Genet. 2014; 46:487-491.

64. Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, Feng Y, Adolph MB, Chelico L, Law EK, Carpenter MA, Harris RS. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. Nat Commun. 2016; 7:12918.

65. 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.

66. 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.

67. Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P. An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example. Hum Mutat. 2013; 34:763-773.

68. Marcinkowska-Swojak M, Klonowska K, Figlerowicz M, Kozlowski P. An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs. Gene. 2014; 546:257-262.

69. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavare S, Deloukas P, et al. Population genomics of human gene expression. Nat Genet. 2007; 39:1217-1224.

70. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermitzakis ET. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010; 464:773-777.

71. Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J Virol. 2009; 83:9474-9485.

72. Refsland EW, Stenglein MD, Shindo K, Albin JS, Brown WL, Harris RS. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues:

implications for HIV-1 restriction. Nucleic Acids Res. 2010; 38:4274-4284.

73. Caval V, Suspene R, Shapira M, Vartanian JP, Wain-Hobson S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3'UTR enhances chromosomal DNA damage. Nat Commun. 2014; 5:5129.

74. Stankiewicz P, Lupski JR. The genomic basis of disease, mechanisms and assays for genomic disorders. Genome Dyn. 2006; 1:1-16.

75. Weckselblatt B, Rudd MK. Human Structural Variation: Mechanisms of Chromosome Rearrangements. Trends Genet. 2015; 31:587-599.

76. Sieuwerts AM, Willis S, Burns MB, Look MP, Meijer-Van Gelder ME, Schlicker A, Heideman MR, Jacobs H, Wessels L, Leyland-Jones B, Gray KP, Foekens JA, Harris RS, et al. Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. Horm Cancer. 2014; 5:405-413.

77. Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM, Temiz NA, Vogel RI, Meijer-van Gelder ME, Sweep FC, Span PN, Foekens JA, Martens JW, et al. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. Sci Adv. 2016; 2:e1601737.

78. Liu J, Sieuwerts AM, Look MP, van der Vlugt-Daane M, Meijer-van Gelder ME, Foekens JA, Hollestelle A, Martens JW. The 29.5 kb APOBEC3B Deletion Polymorphism Is Not Associated with Clinical Outcome of Breast Cancer. PLoS One. 2016; 11:e0161731.

79. Sieuwerts AM, Schrijver WA, Dalm SU, de Weerd V, Moelans CB, Ter Hoeve N, van Diest PJ, Martens JW, van Deurzen CH. Progressive APOBEC3B mRNA expression in distant breast cancer metastases. PLoS One. 2017; 12:e0171343.

80. Cescon DW, Haibe-Kains B, Mak TW. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. Proc Natl Acad Sci U S A. 2015; 112:2841-2846.

81. Tsuboi M, Yamane A, Horiguchi J, Yokobori T, Kawabata-Iwakawa R, Yoshiyama S, Rokudai S, Odawara H, Tokiniwa H, Oyama T, Takeyoshi I, Nishiyama M. APOBEC3B high expression status is associated with aggressive phenotype in Japanese breast cancers. Breast Cancer. 2016; 23:780-788.

82. Ratajska M, Brozek I, Senkus-Konefka E, Jassem J, Stepnowska M, Palomba G, Pisano M, Casula M, Palmieri G, Borg A, Limon J. BRCA1 and BRCA2 point mutations and large rearrangements in breast and ovarian cancer families in Northern Poland. Oncol Rep. 2008; 19:263-268.

83. Gorski B, Jakubowska A, Huzarski T, Byrski T, Gronwald J, Grzybowska E, Mackiewicz A, Stawicka M, Bebenek M, Sorokin D, Fiszer-Maliszewska L, Haus O, Janiszewska H, et al. A high proportion of founder BRCA1 mutations in Polish breast cancer families. Int J Cancer. 2004; 110:683-686.

84. Szwiec M, Jakubowska A, Gorski B, Huzarski T, Tomiczek-Szwiec J, Gronwald J, Debniak T, Byrski T, Kluzniak W, Wokolorczyk D, Birkenfeld B, Akbari MR, Narod SA, et al. Recurrent mutations of BRCA1 and BRCA2 in Poland: an update. Clin Genet. 2015; 87:288-292.

85. Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, Giroux S, Nadaf J, Hamel N, Zhang S, Huzarski T, Gronwald J, Byrski T, et al. Germline RECQL mutations are associated with breast cancer susceptibility. Nat Genet. 2015; 47:643-646.

86. 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.

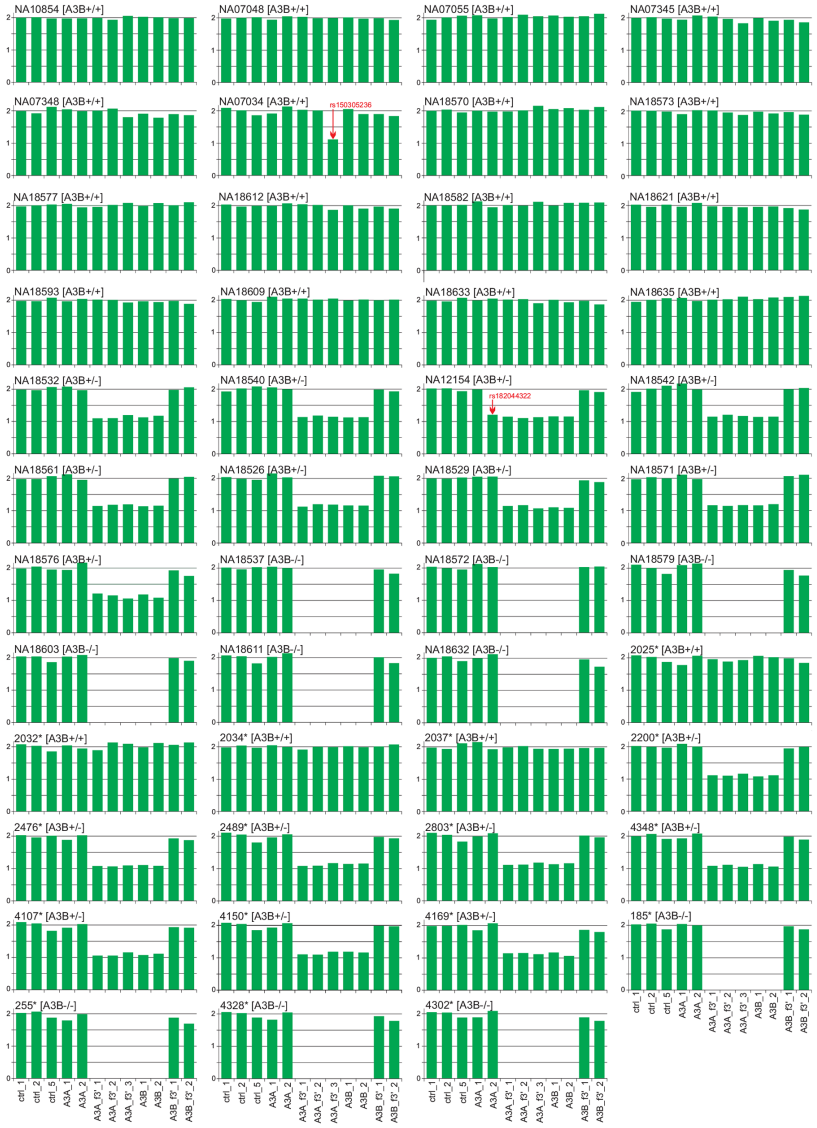
87. Czubak K, Lewandowska MA, Klonowska K, Roszkowski K, Kowalewski J, Figlerowicz M, Kozlowski P. High copy number variation of cancer-related microRNA genes and frequent amplification of DICER1 and DROSHA in lung cancer. Oncotarget. 2015; 6:4716.

88. Marcinkowska-Swojak M, Szczerbal I, Pausch H, Nowacka-Woszuk J, Flisikowski K, Dzimira S, Nizanski W, Payan-Carreira R, Fries R, Kozlowski P, Switonski M. Copy number variation in the region harboring SOX9 gene in dogs with testicular/ovotesticular disorder of sex development (78,XX; SRY-negative). Sci Rep. 2015; 5:14696.

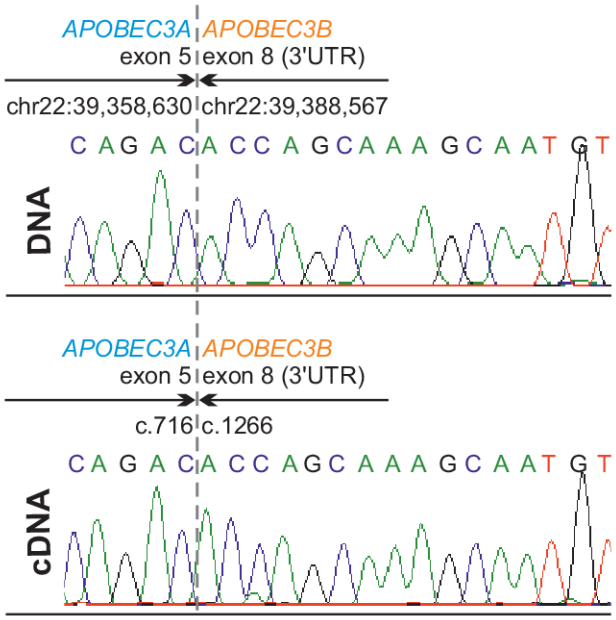
89. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002; 30:e57.

The 30 kb deletion in the *APOBEC3* cluster decreases *APOBEC3A* and *APOBEC3B* expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population

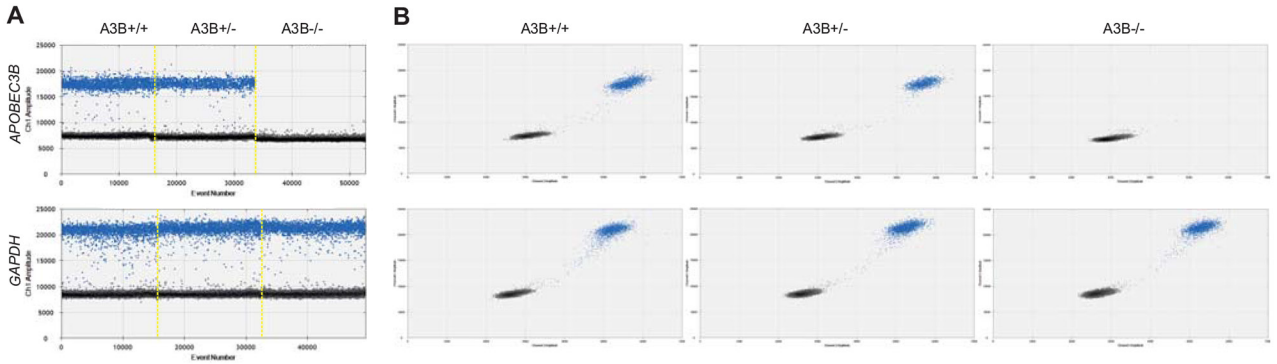
SUPPLEMENTARY MATERIALS



Supplementary Figure 1: Results of MLPA analyses of 31 samples from HapMap panel and 17 samples from women with breast and/or ovarian cancer (indicated with asterisks). The analysis was performed using the in-house designed A3Bdel_MLPA assay. The sample genotypes determined with the use of the A3Bdel_PCR assay are indicated in brackets. Bar plots represent normalized copy number values (y-axis) of each probe (x-axis). In two analyzed samples, i.e., NA07034 and NA12154, we observed an ~40% reduction of the individual probe signals (marked with red arrows). Sequencing analysis revealed that the reduction of signals in these samples resulted from rare SNPs (indicated in red) occurring in close proximity to the ligation sites of the respective MLPA probes.



Supplementary Figure 2: Sanger sequencing of the deletion breakpoint (dotted line) in genomic DNA (upper panel) and cDNA (lower panel) confirms occurrence of the hybrid transcript in samples with the deletion; here, a sample with a homozygous A3B−/− deletion genotype.



Supplementary Figure 3: Exemplary ddPCR results of expression (transcript) level analysis of *APOBEC3B* and reference *GAPDH* in three representative HapMap samples with the A3B+/+, A3B+/−, and A3B−/− genotypes. Note that to obtain comparable numbers of positive signals (droplets), prior to the *GAPDH* analysis, the samples were additionally 400x diluted. The ddPCR analysis and data visualization was performed with the use of BIO-RAD QuantaSoft. The expression level (number of cDNA copies of particular gene in analyzed sample) is proportional to the number of positive droplets. Positive and negative droplets are marked in blue and black, respectively. **(A)** Visualizations in 1D Amplitude view; y-axis – channel 1 amplitude (fluorescent signal intensity), x-axis – sample with the particular genotype (indicated above). **(B)** Visualizations in 2D Amplitude view (corresponding to panel A); y-axis – channel 1 amplitude; x-axis – channel 2 amplitude.

Supplementary Table 1: Detailed characteristics and sequences of the probes included in the A3Bdel_MLPA assay.[a]

See Supplementary File 1

Supplementary Table 1: Detailed characteristics and sequences of the probes included in the A3B+ MLPA assay

probe id	probe location (hg19)	5'PSS	length	5'SS	length	5'TSS	length	Tm	3'TSS	length	Tm	3'SS	length	3'PSS	length	5'HPL	3'HPL	TPL
ctrl_1	chr22:30,069,296-30,069,338	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	46	47	93
A3B_2	chr22:39,388,023-39,388,071	GGGTTCCCTAAGGGTTGGA	19	cgct	4	CCCTGTCCCTTTTCAGAGTTTGAGT	25	70.2	ACTGCTGGGACACCTTTGTGTACC	24	70.4	c	1	TCTAGATTGGATCTTGCTGGCGC	23	48	48	96
A3A_2	chr22:39,357,843-39,357,886	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	AGATGTGGGCCCAGGGAGGGC	21	77.1	AGGGAGAGTGGCTGGAAGTGGAA	23	71.67	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	50	49	99
A3A_f3'_3	chr22:39,377,756-39,377,804	GGGTTCCCTAAGGGTTGGA	19	cgcta	5	CTGTTTATAGACACCCTTCTCCCTTCC	27	70.04	TTGGTGTCGCAGACATCGCAGT	22	70.95	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	51	51	102
A3A_f3'_1	chr22:39,362,163-39,362,209	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	CAGACCCCAGAATCAGAGACTGAG	24	70.1	GAGTCAAAGACAGGCAGGAGACG	23	70.8	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	53	52	105
A3B_f3'_1	chr22:39,391,228-39,391,279	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	CTCCCCATTCATCTGCAAAGTGAGT	25	70.8	CTAGGGAGGTTCTTGTGATTCTTTGGC	27	71.6	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	54	54	108
ctrl_2	chr1:156,105,818-156,105,862	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	55	56	111
A3B_1	chr22:39,382,626-39,382,676	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CTCTAAGGCACCTGGGTCTTGGACTA	26	72.4	GAGGAACAAACAGCAAGACAGGGTG	25	72.3	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	57	57	114
A3A_f3'_2	chr22:39,367,085-39,367,132	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	GAGATGAGCTTTCCATGACGACCAA	25	71.4	GGTGGAGGAGTCAGATTCTGGGT	23	71.34	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	59	58	117
A3B_f3'_2	chr22:39,397,215-39,397,269	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	CCACTCTTACTCCCTTTCATCCCGTA	26	71.3	ACACGTAGCTCTACTAGCCTGCTATTGGG	29	71.48	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	60	60	120
ctrl_5	chr2:109,545,794-109,545,837	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	71.8	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	62	124
A3A_1	chr22:39,353,593-39,353,648	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	GCTCCTGAATCCTAAGAGAATGTTGGTG	28	71.4	AAGATCTTAACACCACGCCTTGAGCAAG	28	71.6	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	64	64	128

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 length

TPL - total probe length

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA

SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

AC# V00604 Phage M13 genome

position: 3-99

5'-cgctactactattagtagaattgatgccacctttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccatttgcgaaatgtatctaatggtcaaactaaatctac-3'

Supplementary Table 2: Analysis of the association of the *APOBEC3B* deletion with breast cancer risk using different models of inheritance

group	model of inheritance	OR(95%CI)	adjusted OR(95%CI)
GDANSK BC cases (n=523) vs. unselected controls (n=853)	dominant A3B+/− and A3B−/− vs. A3B+/+	1.31(0.94-1.82) p=0.11	-
	recessive* A3B−/− vs. A3B+/− and A3B+/+	1.09(0.18-6.53) p=0.93	-
	additive A3B−/− vs. A3B+/− vs. A3B+/+	1.28(0.94-1.75) p=0.12	-
SZCZECIN BC cases (n=2009) vs. unselected controls (n=2005)	dominant A3B+/− and A3B−/− vs. A3B+/+	0.89(0.74-1.07) p=0.20	0.90(0.75-1.08)[a] p=0.26
	recessive* A3B−/− vs. A3B+/− and A3B+/+	2.00(0.68-5.87) p=0.21	2.14(0.73-6.3)[a] p=0.17
	additive A3B−/− vs. A3B+/− vs. A3B+/+	0.91(0.76-1.09) p=0.31	0.93(0.78-1.11)[a] p=0.41
SZCZECIN BC cases (n=2005) vs. NH-controls (n=615)	dominant A3B+/− and A3B−/− vs. A3B+/+	1.03(0.78-1.36) p=0.83	0.95(0.70-1.29)[a] p=0.73
	recessive* A3B−/− vs. A3B+/− and A3B+/+	0.61(0.21-1.79) p=0.37	0.34(0.10-1.13)[a] p=0.08
	additive A3B−/− vs. A3B+/− vs. A3B+/+	1.00(0.77-1.30) p=1.00	0.90(0.68-1.21)[a] p=0.49
GDANSK+SZCZECIN BC cases (n=2532) vs. unselected controls (n=2858)	dominant A3B+/− and A3B−/− vs. A3B+/+	0.98(0.83-1.15) p=0.79	0.97(0.83-1.14)[b] p=0.73
	recessive* A3B−/− vs. A3B+/− and A3B+/+	1.70(0.69-4.16) p=0.25	1.70(0.69-4.19)[b] p=0.25)
	additive A3B−/− vs. A3B+/− vs. A3B+/+	1.0(0.85-1.16) p=0.96	0.99(0.85-1.16)[b] p=0.90
GDANSK+SZCZECIN familial BC cases (n=640) vs. unselected controls (n=2858)	dominant A3B+/− and A3B−/− vs. A3B+/+	1.06(0.82-1.36) p=0.67	1.15(0.87-1.52)[b] p=0.32
	recessive* A3B−/− vs. A3B+/− and A3B+/+	1.12(0.24-5.27) p=0.89	0.88(0.17-4.71)[b] p=0,88
	additive A3B−/− vs. A3B+/− vs. A3B+/+	1.06(0.83-1.35) p=0.66	1.14(0.87-1.48)[b] p=0.36

[a]adjusted for age; [b]adjusted for the origin of the study; *in the recessive model we obtained relatively low statistical power due to the low frequency of homozygous deletions.

Supplementary Table 3: Analysis of the association of the *APOBEC3B* deletion with ovarian cancer risk using different models of inheritance

group	model of inheritance	OR(95%CI)	adjusted OR(95%CI)
GDANSK OC cases (n=343) vs. unselected controls (n=853)	dominant A3B+/− and A3B−/− vs. A3B+/+	0.77(0.50-1.19) p=0.24	-
	recessive* A3B−/− vs. A3B+/− and A3B+/+	1.66(0.28-9.99) p=0.58	-
	additive A3B−/− vs. A3B+/− vs. A3B+/+	0.82(0.54-1.22) p=0.32	-
VILNIUS OC cases (n=97) vs. unselected controls (n=209)	dominant A3B+/− and A3B−/− vs. A3B+/+	0.66(0.27-1.61) p=0.36	-
	additive A3B−/− vs. A3B+/− vs. A3B+/+	0.78(0.34-1.77) p=0.55	-
GDANSK + VILNIUS OC cases (n=440) vs. unselected controls (n=1062)	dominant A3B+/− and A3B−/− vs. A3B+/+	0.75(0.51-1.10) p=0.14	0.75(0.51-1.11)[b] p=0.15
	recessive* A3B−/− vs. A3B+/− and A3B+/+	2.42(0.49-12.05) p=0.28	2.44(0.49-12.14)[b] p=0.28
	additive A3B−/− vs. A3B+/− vs. A3B+/+	0.81(0.56-1.16) p=0.25	0.81(0.56-1.16)[b] p=0.25

[b]adjusted for the origin of the study; *in the recessive model we obtained relatively low statistical power due to the low frequency of homozygous deletion, the recessive model was not tested in the VILNIUS group due to lack of homozygous deletions in the control group.

6

Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P

MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency

Oncotarget, 2016, 7(41):67266-67276

Research Paper

MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency

Katarzyna Klonowska[1], Luiza Handschuh[1,2], Aleksandra Swiercz[1,3], Marek Figlerowicz[1,3], Piotr Kozlowski[1,4]

[1]European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland

[2]Department of Hematology and Bone Marrow Transplantation, Poznan University of Medical Sciences, 60-569 Poznan, Poland

[3]Institute of Computing Science, Poznan University of Technology, 60-965 Poznan, Poland

[4]Institute of Technology and Chemical Engineering, Poznan University of Technology, 60-965 Poznan, Poland

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

Keywords: *MTTE, exome enrichment, next generation sequencing, MLPA, capillary electrophoresis*

Received: *May 15, 2016* ***Accepted:*** *July 28, 2016* ***Published:*** *August 27, 2016*

ABSTRACT

Although currently available strategies for the preparation of exome-enriched libraries are well established, a final validation of the libraries in terms of exome enrichment efficiency prior to the sequencing step is of considerable importance. Here, we present a strategy for the evaluation of exome enrichment, i.e., the Multipoint Test for Targeted-enrichment Efficiency (MTTE), PCR-based approach utilizing multiplex ligation-dependent probe amplification with capillary electrophoresis separation. We used MTTE for the analysis of subsequent steps of the Illumina TruSeq Exome Enrichment procedure. The calculated values of enrichment-associated parameters (i.e., relative enrichment, relative clearance, overall clearance, and fold enrichment) and the comparison of MTTE results with the actual enrichment revealed the high reliability of our assay. Additionally, the MTTE assay enabled the determination of the sequence-associated features that may confer bias in the enrichment of different targets. Importantly, the MTTE is low cost method that can be easily adapted to the region of interest important for a particular project. Thus, the MTTE strategy is attractive for post-capture validation in a variety of targeted/exome enrichment NGS projects.

INTRODUCTION

Recently, next-generation sequencing (NGS) has become the leading method for analyzing the architecture of human genomes. Although the cost of whole genome sequencing (WGS) has decreased significantly in recent years, it still substantially hampers the use of WGS for large-scale studies involving abundant DNA sample sets [1]. However, it has to be noted that resistance to biases in coverage of some genomic regions (e.g., rich in GC nucleotides) and complete coverage of the genome (especially using most recent PCR-free WGS) are among advantages of WGS [2]. Nevertheless, targeted sequencing, based on the capture and enrichment of a restricted part of a genome (i.e., multiple genomic loci of interest), is currently commonly applied to reduce the costs and the amount of data that requires time-consuming analysis [3]. The use of a targeted enrichment sequencing strategy focused on well-characterized coding sequences, i.e., whole exome sequencing (WES), yields informative results that are easier to interpret [4–6]. Targeted/exome sequencing may be favorable especially for applications that require high-coverage of the analyzed regions for identification of low–frequency sequence variants. Such applications include: identification of somatic mutations in cancer genome, identification of mosaic mutations in disease-related genes, identification of mitochondrial DNA heteroplasmy, or identification of sequence variants in mixed DNA samples (e.g., in forensic genetics). Currently, several popular ready-to-use kits

for the preparation and sequencing of exome-enriched libraries (mostly from Agilent, Roche NimbleGen, and Illumina) are commercially available. Studies focused on the comparison of their performance revealed that generally all kits are well established and provide results of comparable quality. However, it has to be noted that drawbacks and differences in some aspects of enrichment technologies, e.g., accuracy of variant detection and presence of enrichment biases associated with sequence characteristics, were also identified [1, 7–13]. For example, Meienberg and colleagues have revealed that currently available exome-enrichment platforms cannot efficiently capture all known coding exons and emphasized the need of constant evaluation of the updated platform versions [11].

Due to the revealed differences in the performance of the exome-enrichment platforms and the high costs associated with the downstream sequencing analysis, a quality control for capture performance and exome enrichment efficiency is highly desirable. The introduction of a post-capture validation step preceding the sequencing analysis may prevent the sequencing of unsuccessfully enriched libraries [14–16].

Here, we developed a new strategy and propose an exome enrichment validation assay, the Multipoint Test for Targeted-enrichment Efficiency (MTTE). MTTE is based on and utilizes the standard well-validated protocol of Multiplex Ligation-dependent Probe Amplification (MLPA) method (Figure 1). The general concept and principle of the MLPA method are discussed in [17–19]. Our assay comprises multiple probes located both in targeted (exome-enriched) and non-targeted genomic regions. In this report, we show that the MTTE is a robust and cost-effective assay that allows the effective and reliable assessment of several enrichment-associated parameters.

RESULTS

Design of the MTTE assay

Our assay for post-capture exome enrichment validation is composed of 20 MLPA probes, including 10 probes located in targeted genomic regions (mostly exons of protein coding genes and one region overlapping an annotated miRNA sequence), 9 probes located in non-targeted genomic regions (introns and intergenic regions), and one probe located in flank of the targeted regions (49 bp from exon 1 of the *BARD1* gene). The probe located in flank of targeted region was used to test the enrichment of sequences located in close vicinity (≤150bp) to the targeted regions that are enriched together with targeted sequences. The MLPA probe set was designed according to a strategy developed previously in our group, allowing easy design and generation of the assay for the analysis of almost any region of interest [19, 20]. Selected regions were approximately evenly distributed over the genome. To allow the direct comparison of the enrichment efficiency of targeted and non-targeted regions situated in close proximity to each other, in two cases, probes of different types were located in the same gene, i.e., *BARD1* and *ARID1A* (Figure 2A). The designed MLPA probe set was verified to provide robust high quality results in a series of optimization experiments performed using a set of reference gDNA samples (Figure 2B).

MTTE evaluation of the enrichment-associated parameters

We then used our MTTE assay to analyze the enrichment efficiency in one normal (normal_1) and two acute myeloid leukemia samples (leukemia_1 and leukemia_2). The MTTE analysis of the relative amount of targeted and non-targeted regions was performed during five consecutive steps of the exome enrichment procedure, i.e., (i) untreated gDNA, (ii) fragmented gDNA, (iii) the PCR-amplified gDNA library with ligated adapters, (iv) the PCR-amplified library after the first enrichment, and (v) the PCR-amplified library after the second enrichment (Figure 2C).

Representative MTTE results (electropherograms and bar graphs) from leukemia_1 are presented in Figure 3. Bar graphs for all three analyzed samples are shown in Supplementary Figure S1. As shown in Figure 3 and Supplementary Figure S1, DNA fragmentation (ii) and adapter ligation followed by PCR amplification (iii) do not influence the pattern of MLPA signals substantially when compared with the pattern of signals from untreated gDNA samples (i). Therefore, specimens from any of these steps (i-iii) may be considered as a reference for monitoring the exome enrichment. For subsequent analysis, we used the sample from step (iii), which bears structural resemblance to exome-enriched specimens (iv-v), as reference.

As shown in Figure 3 and Supplementary Figure S1, the signal intensities of probes located in the non-targeted regions were drastically reduced after the first enrichment and almost completely disappeared after the second enrichment step.

To quantify the exome enrichment efficiency in steps after the first (iv) and the second (v) enrichment, we calculated the following enrichment-associated parameters (for details, see Materials and Methods): relative enrichment (RE); relative clearance (RC), i.e., 1-RE, calculated for each non-targeted region as well as for target flank region; overall clearance (OC), i.e., average of RC values of non-targeted regions; and fold enrichment (FE), i.e., increase of the fraction of signal intensity of the probes located in targeted regions, weighted by the fraction of the genome covered by the targeted regions. As the regions targeted by the TruSeq Exome Enrichment Kit cover 62 Mb, i.e., 0.02 of the human genome, the theoretical

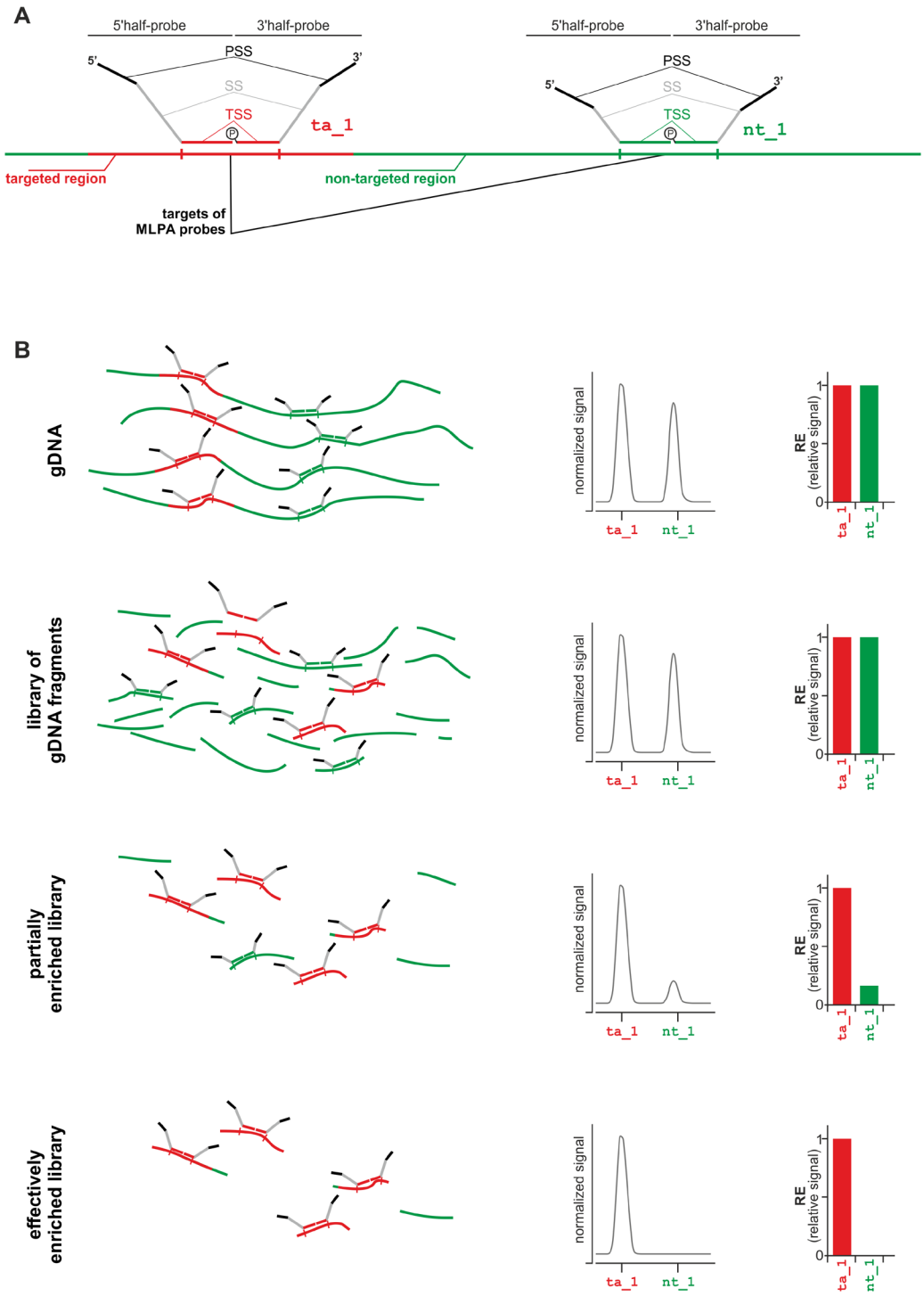


Figure 1: Strategy of MTTE analysis. A. Schematic representation of targeted region-specific (ta_1; left-hand side) and non-targeted region-specific (nt_1; right-hand side) MLPA probes. Each MLPA probe is composed of two half-probes: a 5'half-probe and a 3'half-probe. Each half-probe is composed of a target-specific sequence (TSS), a primer-specific sequence (PSS), and a stuffer sequence (SS) that allows differentiating MLPA probes by size. More details about the design of MLPA probes may be found in [18, 20]. The first step of the MLPA reaction is hybridization of MLPA probes with the input DNA. Only probes which were correctly hybridized to their targets are subsequently ligated and then amplified with a pair of universal primers. The products of the MLPA reaction are separated in capillary electrophoresis and their relative signals are proportional to the dosage of their targets in the input DNA. **B.** The MTTE analysis of (from the top) gDNA, non-enriched gDNA library (reference), partially enriched library, and effectively enriched library. From the left, schematic representation of the MLPA probes hybridizing to targeted- and non-targeted regions in the input DNA (for simplicity, adapter sequences attached to DNA fragments during library preparation are not indicated in the scheme), electropherograms with signals (peaks) of ta_1 and nt_1 probes, bar-graphs showing relative signals of ta_1 and nt_1 probes.

maximum FE is 50. This parameter is comparable to the fold enrichment calculated using the SeqCap qPCR assay from Roche NimbleGen. Analysis of the enrichment-associated parameters revealed the robustness and high reliability of the MTTE assay. The targeted regions were captured with high efficiency. After the first enrichment, the OC reached a value ranging between 0.91 and 0.92, whereas FE was in a range between 10.5 and 11.2. These parameters were significantly improved after the second round of enrichment, with an OC value ranging between 0.98 and 1 and FE reaching a value of up to 50. RC values calculated for individual probes located in the non-targeted regions ranged from 0.83 to 1 and from 0.94 to 1 for the first and the second steps of enrichment, respectively. It should be noted that the clearance of the target flank (the tf_2q35 region), located close to the targeted region (exon 1 of *BARD1*) is much less effective (RC ranging between 0.87 and 0.88 after the second enrichment) than clearance of other non-targeted regions. This result is in agreement with the hypothesis that sequences adjacent to the targeted regions (≤150nt) are also captured by the enrichment procedure. Comparison of the RC values of probes located in the non-targeted regions with their distance to the nearest targeted genomic region further confirmed the positive correlation between the clearance efficiency and increasing distance from the targeted regions (R=0.77, p=0.006). These observations further confirm the specificity of our MTTE assay and correspond to average coverage of sequences surrounding the targeted sequences, which gradually decreases with increasing distance from the targeted sequences and reaches the minimum (~0) at ~500 nt upstream and downstream from the targeted regions (Supplementary Figure S2).

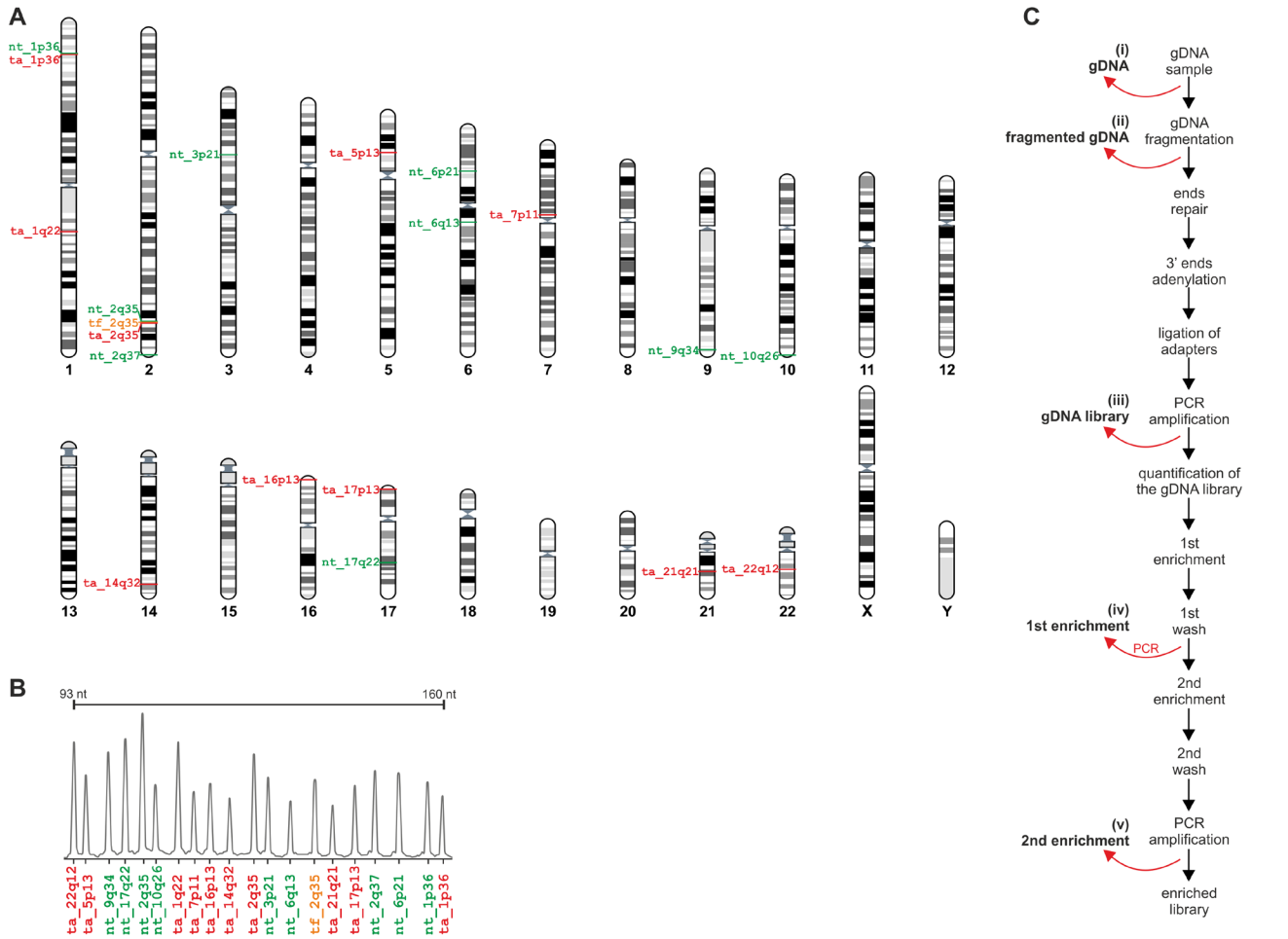


Figure 2: Design of the MTTE assay. A. The positions of MTTE probes in the human genome. The positions of particular probes are indicated on the chromosome ideograms (left-hand side). IDs of probes located in the targeted regions, in close vicinity (<150) to the targeted regions, and the non-targeted regions are indicated in red, orange and green, respectively. The figure was prepared using the "Ensembl karyotypes" tool available on the Ensembl portal [30]. **B.** Electropherogram showing representative MTTE results from a control gDNA sample from the HapMap panel. Each peak corresponds to the signal of the particular probe indicated below [color coded as in A]. **C.** The workflow of exome enrichment procedure with indicated points at which trace amounts of specimens were obtained for analysis.

Exome enrichment efficiency: comparison of the MTTE-based evaluation and actual NGS enrichment

For comparison of the MTTE results with actual enrichment, we calculated the fraction of NGS reads mapping to the targeted sequences and the fold enrichment of these values (FE^{NGS}), after the first (iv) and second (v) steps of enrichment. As shown in Table 1, the FE calculated based on the MTTE results corresponded well with the FE^{NGS}. After the first round of enrichment, both FE and FE^{NGS} reached a value of ~10, which corresponds to the borderline enrichment value recommended by Roche NimbleGen, manufacturer of the SeqCap qPCR assay. Slightly higher difference between the FE and FE^{NGS} values observed after the second round of enrichment is due to the high disproportion between the fractions of targeted and non-targeted regions (here 1:50). The

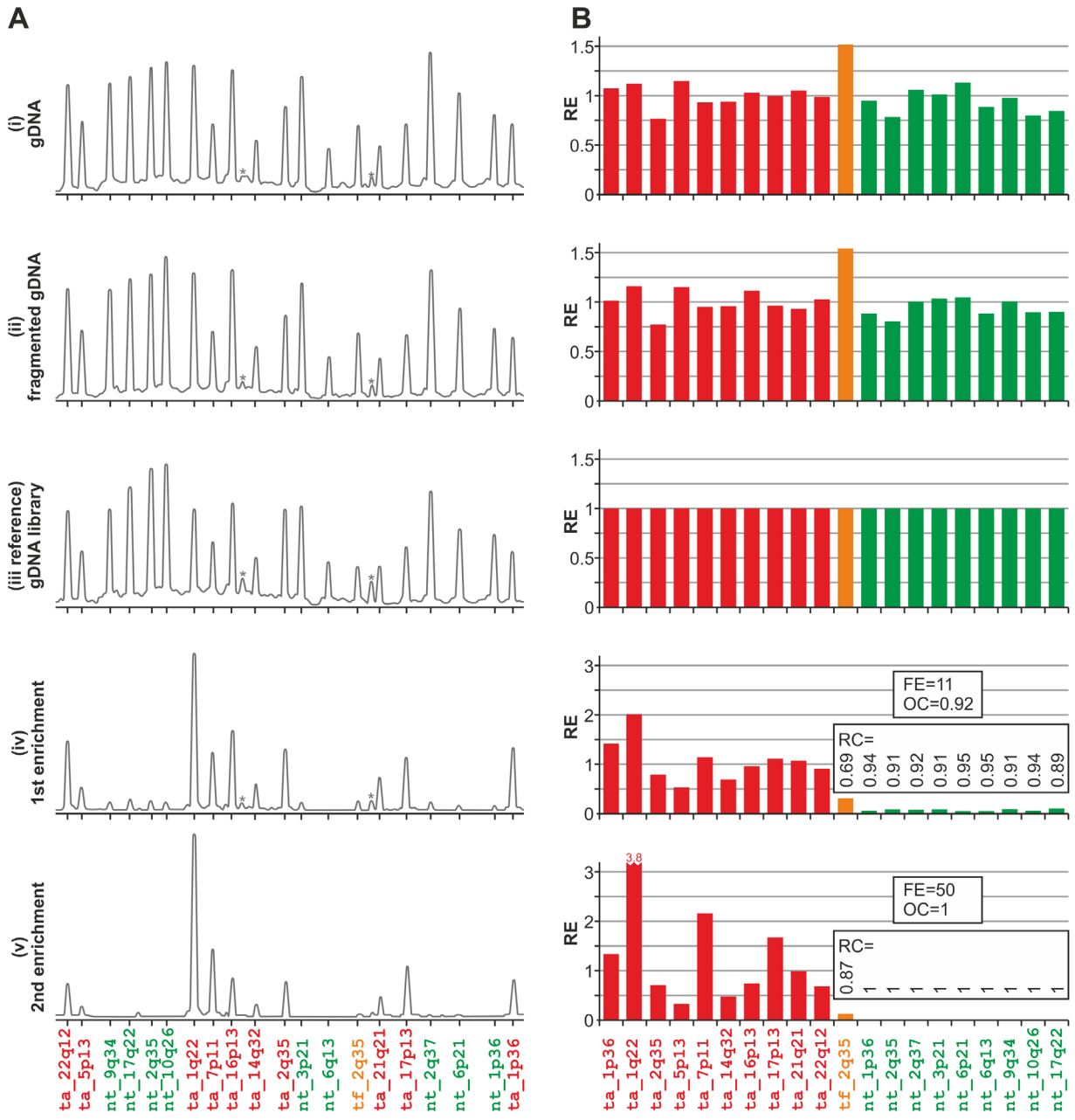


Figure 3: Representative results of the analysis conducted using the in-house designed MTTE assay. A. The electropherograms of the MLPA results obtained in the analysis of specimens from distinct steps of the exome-enriched library preparation, performed using the leukemia_1 sample. The probe IDs are shown under the electropherograms. Asterisks indicate background signals (unspecific peaks). **B.** The bar plots (corresponding to the electropherograms shown in panel A) representing the relative enrichment (y-axis) of each analyzed region (x-axis). The corresponding RC as well as FE and OC values are indicated on the graphs (steps iv-v).

Table 1: Enrichment efficiency (OC, FE, FE^{NGS}, and FE^{qPCR}) after 1st and 2nd steps of enrichment of the three analyzed samples

	OC		FE		FE^{NGS}		FE^{qPCR}	
	1st enrichment	2nd enrichment	1st enrichment	2nd enrichment	1st enrichment	2nd enrichment	1st enrichment RN/ratio	2nd enrichment RN/ratio
Normal_1	0.91	1	11.2	50	10.9	23.1	8.3/10.5	13.0/34.9
Leukemia_1	0.92	1	11.0	50	9.6	23.7	9.5/11.6	14.0/45.1
Leukemia_2	0.92	0.98	10.5	32.2	11.3	29.5	12.1/11.9	20.5/32.9

Normal_1, Leukemia_1, and Leukemia_2 – three samples used in the experiment; OC – overall clearance; FE – fold enrichment; FE^{NGS} – fold enrichment of the fraction of NGS reads mapping to the targeted sequences; FE^{qPCR} – fold enrichment calculated based on qPCR analyses; RN – calculated as proposed by the Roche NimbleGen protocol and [16]; ratio – calculated based on ratio of enrichment of targeted and non-targeted regions weighted by proportion of targeted/non-targeted regions in the genome. Note that some differences between FE^{qPCR} (RN) and FE^{qPCR} (ratio) may result from imprecision of measurement of the DNA concentration. Amount of input DNA is assumed to be a normalization factor in FE^{qPCR} (RN) calculation. According to Roche NimbleGen (SeqCap EZ Library SR User's Guide, v4.2), two-fold differences in the FE^{qPCR} (RN) measures should be considered as the same.

precision of estimation of fold enrichment decreases when the fold enrichment reaches maximum. As proposed in the SeqCap qPCR validation strategy (SeqCap EZ Library SR User's Guide, v4.2), two-fold differences in the fold enrichment measures should be considered as the same. The higher FE values calculated based on the MTTE results are a consequence of the lower sensitivity of MTTE to detect very low traces of non-targeted sequences, which is due to the limited dynamic range of capillary electrophoresis (very low signals may not be detected and are classified as 0). On the other hand, slightly lower values of FE^{NGS} may be due to the fact that among NGS reads overlapping targeted regions are also those extending to the non-targeted regions. As a result, the calculated fraction of target-associated reads may be higher than the actual fraction of TruSeq targeted regions (>2%).

Comparison of qPCR- and MTTE-based evaluation of targeted enrichment

To compare MTTE results with an evaluation of enrichment based on qPCR analysis (proposed before by Roche NimbleGen), we designed and optimized four qPCR assays for targeted regions showing different enrichment efficiency in MTTE experiments. Additionally, we designed three qPCR assays for non-targeted regions, previously analyzed with the use of MTTE. In both cases, the PCR amplicons were designed to maximally overlap target sequences of corresponding MTTE probes. As expected, after the first and second rounds of enrichment threshold cycle (Ct) values decreased for all targeted regions and increased for all non-targeted regions in comparison to the values for the non-enriched library (Supplementary Figure S3). The fold enrichment values calculated based on qPCR results (FE^{qPCR}) confirm borderline (~10) and high quality enrichment (>10) after the first and the second round of enrichment, respectively (Table 1).

Additionally, the direct comparison of the enrichment efficiency of individual probes shows that RE values calculated for the MTTE probes correspond well with FE^{qPCR} values, calculated for the individual qPCR assays (after the first enrichment: $R>0.95$, $p<0.001$; after the second enrichment: $R>0.85$; $p<0.01$) (Supplementary Figure S3).

Determination of the structural and sequence-associated features related to bias in enrichment

The results of our MTTE analysis of exome-enriched libraries [steps (iv) and (v)] revealed a considerable difference (up to sixteen-fold) between the RE values of particular targeted probes. This results from the uneven enrichment of different regions. However, it should be noted that the enrichment pattern was recurrent across all analyzed samples (Supplementary Figure S1) and was confirmed in the regions analyzed by qPCR (Supplementary Figure S3). This finding implies that the observed bias in enrichment may result from specific structural and/or sequence-associated features. Thus, we analyzed the potential correlation between the RE values and several features of the targeted regions such as, the length of the targeted region, the occurrence of repetitive elements in the flanking sequences of targeted regions, and the content of different nucleotides in the targeted plus DNA strand (Supplementary Figure S4). The analyses did not indicate the influence of targeted

region length and repetitive sequences on the RE of the targeted probes [correlation coefficient (R)~0]; however, a moderate to high positive correlation was observed between the RE values and the fractions of (i) GC (first enrichment: R=0.59, p=0.06; second enrichment: R=0.40, p=0.23), (iii) purines (first enrichment: R=0.69, p=0.02; second enrichment: R=0.70, p=0.02), and (ii) guanine (G) (first enrichment: R=0.77, p=0.005; second enrichment: R=0.69, p=0.02) nucleotides in the sequence targeted by the TruSeq exome enrichment probes. Consequently, a negative correlation was observed between the RE values and the fraction of thymine (T) (first enrichment: R=0.63, p=0.04; second enrichment: R=0.55, p=0.08) in the targeted regions.

DISCUSSION

The use of targeted/exome enrichment NGS is becoming increasingly popular. Beside the above mentioned platforms from Agilent, Roche NimbleGen and Illumina, there are also many alternative more tailored approaches, including: MYbaits Target Enrichment Kit (MYcroarray), xGen (IDT), Custom targeted sequencing oligo pools (CustomArray), The RainDance ThunderStorm system (RainDance technologies), Access Array system (Fluidigm), and Quest 5-hmC DNA Enrichment Kit (Zymo Research). Although currently available ready-to-use kits for the preparation of exome-enriched libraries are well established and provide reliable results, a final evaluation of their capture performance and exome enrichment efficiency is of considerable interest due to the high costs associated with the downstream sequencing analysis. Here, we propose an innovative multipoint MTTE strategy that enables the complex evaluation of exome enrichment efficiency using a comprehensively selected and optimized MLPA probe-set. We used the MTTE assay for the calculation of several enrichment-associated parameters that reflect the actual enrichment with high accuracy. The main advantages of our MTTE assay are as follows: (i) MTTE is composed of multiple probes located in targeted sequences of different characteristics. Enrichment of a particular target type (e.g., with higher GC content) may be interpreted in the context of a particular project. (ii) MTTE probes are specific for both targeted and non-targeted regions, therefore allowing the evaluation of enrichment of targeted regions and the evaluation of clearance of non-targeted regions. These measures are complementary and additionally validate each other. (iii) MTTE does not require optimization or generation of standard curves. It takes advantage of a standard protocol (standard reaction conditions, easily accessible reagent set) of MLPA. The standard MLPA setup was validated and successfully used in hundreds of research and clinical studies for the analysis of large mutations in disease-related genes [19, 21–23]. (iv) The MTTE strategy can be easily adapted to the region of interest important for a particular project, e.g., if particular gene is sequenced, target-specific probes may be located in each exon of the gene. (v) The MTTE test is cost-effective (~5 USD per sample, including cost of the capillary electrophoresis separation, except the starting cost of probes synthesis). (vi) Finally, MTTE may also be used for optimization of enrichment procedures (testing the effect of multiple conditions).

Compared to the commercially available SeqCap qPCR assay, MTTE allows the analysis of more genomic regions (also non-targeted) and utilizes a much simpler setup. MTTE requires the analysis of only two samples, i.e., two MLPA reactions [one reference (not enriched) and one enriched sample] performed in standard conditions, whereas the SeqCap qPCR assay requires 24 reactions, averaging the results (preparation of standard curve) and optimizing the conditions for all fragments to be analyzed. Additionally, our own experience with qPCR analysis indicates that it is much more labor intensive and analysis of each sample requires numerous reactions. The number of required reactions also substantially increases the cost of qPCR analysis (20-50 USD, depending on the used system and number of evaluated regions). However, the MTTE assay is limited by dynamic range of capillary electrophoresis and may miss very low traces of non-targeted probe signals in some samples after enrichment. Nevertheless, this limitation does not affect the reliability of our test to detect poorly enriched libraries (FE<10, recommended by Roche NimbleGen). MTTE also detects libraries with FE=10-25 that do not represent the highest quality of enrichment. It is worth noting that in exome enrichment, FE values of 10 and 25 correspond to 80% and 50% of reads mapping out of the targeted regions, respectively.

Additionally, we took advantage of our MTTE results to determine the potential sequence-associated features that may confer bias in the enrichment of different capture targets. The identified sequence-associated features that increase or decrease the efficiency of targeted enrichment generally overlap with features identified before [e.g., [1, 7, 11, 12]]. For example, we observed a positive correlation between the enrichment efficiency and the fraction of GC in the enriched region (Supplementary Figure S4) as well as between the clearance efficiency and the increasing distance to the nearest targeted region (Supplementary Figure S2). Additionally, our analysis revealed a significant positive correlation between RE values and fraction of purines in targeted sequences that, to our knowledge, has not been reported before. The analysis of the fraction of individual nucleotides showed that the main driver of both purines and GC effect on the enrichment is presence of G in targeted sequences. It may suggest that not only thermodynamic parameters but also sequence composition of either a probe or a targeted sequence may influence enrichment efficiency. It has to be noted, however, that these results should be interpreted carefully due to (i) the non-random selection of probed

regions that does not cover the full range of the analyzed parameter (e.g., GC content in a range of 39-60%), and (ii) the low power of our analysis (low number of analyzed regions).

In conclusion, we designed an innovative MTTE strategy for the evaluation of exome-enriched libraries that may be easily adapted to any set of selected targets (e.g., exome, miRNome, methylome (methyl-seq) or panel of genes of interest). The strategy allows not only reliable estimation of general sample enrichment but also allows the evaluation of enrichment of regions of specific characteristics or location that may be of special interest for a project.

MATERIALS AND METHODS

Samples

The performance of MLPA probes and the MTTE assay was validated using reference gDNA samples from the HapMap panel purchased from the Coriell Institute for Medical Research [24]. According to the information from the Coriell Institute, all samples were diluted to a working concentration of 50 ng/µl.

The enrichment analysis was performed on 3 samples [one normal (normal_1) and two acute myeloid leukemia (leukemia_1 and leukemia_2) samples] sequenced at the ECBaG (European Centre for Bioinformatics and Genomics) in the framework of other projects.

Preparation of libraries and sequencing analysis

Three genomic DNA samples (1 µg per sample) were fragmented through sonication [Bioruptor NextGen (Diagenode, Denville, NJ, USA)] at a low power and for 45 cycles (30 s on, 30 s off) and used for the construction of the indexed gDNA libraries using a TruSeq DNA Sample Preparation Kit (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. Whole exome enrichment was performed from 500 ng of each gDNA library using a TruSeq Exome Enrichment Kit (Illumina). For testing of MTTE performance, we compared the efficiency of enrichment after the first and second rounds of enrichment. For this purpose, 10% (3 µl) of the library from the first round of enrichment was pulled out and subsequently treated as the corresponding library after the second step of enrichment (PCR amplification in 10 µl volume). Size distribution of the exome-enriched libraries was validated using the High Sensitivity DNA Assay in a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The libraries were quantified with a Qubit™ 1.0 Fluorometer (Invitrogen, Life Technologies, Grand Island, NY, USA) and real-time PCR [Rotor-Gene Q (Qiagen, Venlo, The Netherlands)], performed according to the qPCR Quantification Guide (Illumina TruSeq Enrichment Guide) using primers complementary to the adapter sequences and MESA Green qPCR MasterMix (Eurogentec). For sequencing with the Genome Analyzer GAIIx (Illumina), three libraries were combined per lane of a single-read flow-cell in the following way: lane 1 - three libraries before enrichment; lane 2 – three libraries after the first round of enrichment; lane 3 - three libraries after the second round of enrichment. Approximately 9 million 101 bp-long reads were collected for each library. The raw data were submitted to adapter trimming and quality filtering with the FASTX-Toolkit. The remaining reads were mapped to the human reference genome (hg19) using Bowtie 2. To calculate the number of reads mapped to the gene coding and non-coding regions, specialized Linux bash scripts were prepared, taking into account the coordinates of the regions targeted by the TruSeq Exome Enrichment Kit (Illumina).

MTTE assay and MLPA analysis

MLPA analysis was performed using the in-house designed MTTE assay. The MLPA probes and the probe-set layout were designed and generated according to a previously proposed [19, 20] and well-validated [e.g.,[23, 25–27]] strategy. This strategy exclusively utilizes short oligonucleotide probes that can be easily generated via standard chemical synthesis. Each probe consists of two half-probes of equal size, and the total probe length ranges from 93 to 160 nt. The target sequences for the probes were selected to avoid single nucleotide polymorphisms (SNPs), insertions/deletions, copy number variable regions (CNVRs), segmentally duplicated sequences, repeat elements, and sequences of extremely low or high guanosine-cytosine (GC) content [18, 20]. The designed probes are complementary to the targeted and non-targeted regions, selected based on the genomic coordinates of the regions targeted by the TruSeq Exome Enrichment kit. The sequences and detailed characteristics of all probes are shown in Supplementary Table S1. All probes were selected from MLPA sets designed before and validated in hundreds of samples in our previous projects [19, 20, 23, 25, 26, 28]. The MLPA probes were synthesized by IDT (Skokie, IL, USA). Samples for the MTTE analysis were obtained at several steps during the preparation of exome-enriched libraries. In the MTTE analysis of each sample, we used a 5 µl aliquot of the gDNA/library collected at the following steps of the enrichment procedure: (i) gDNA (50 ng/µl), (ii) 4-fold dilution of 1.25 µl of fragmented gDNA (up to ~25 ng/µl), (iii) 5-fold dilution of 1 µl of gDNA library, (iv) 5-fold dilution of 1 µl of the exome enriched library after the first enrichment, and (v) 5-fold dilution of 1 µl of the exome enriched library after the second enrichment. The MLPA reactions were run according to the manufacturer's general recommendations (MRC-Holland, Amsterdam, The Netherlands), and as

described earlier [17, 19]. The products of the MLPA reactions were diluted 100× in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated by size using capillary electrophoresis (POP7 polymer; ABI Prism 3130XL apparatus; Applied Biosystems, Carlsbad, CA, USA). The electropherograms were analyzed using GeneMarker software v 2.2.0 (SoftGenetics, State College, PA, USA). The signal intensities (peak heights) were retrieved and transferred to prepared Excel sheets (available upon request). The signal intensity of each probe was divided by the geometric mean of signal intensities of the probes specific for the targeted regions to normalize the values and to minimize run-to-run variation. To calculate RE for each targeted and non-targeted region, the normalized signal intensity of each probe was divided by the corresponding normalized signal of the probe in the reference sample to avoid the biased effect that results from various efficiency of the probes amplification. FE for samples after the first and the second enrichment was calculated according to the following equation: $FE=[(TP_E*TR)/(TP_E*TR+NP_E*NR)]/[(TP*TR)/(TP*TR+NP*NR)]$, where: TP_E – the average RE of target probes in the sample after enrichment; TR – the fraction of targeted regions (here 0.02); NP_E – the average RE of non-target probes in the sample after enrichment; NR – the fraction of non-targeted regions (here 0.98); TP – the average RE of target probes in the reference sample, which equals 1; and NP – the average RE of non-target probes in the reference sample, which equals 1.

qPCR analyses

qPCR analyses were performed with the use of LightCycler 480 system with probes from the Universal Probe Library (UPL) (Roche), following the protocol proposed by the manufacturers. qPCR analyses were conducted for all three samples (i.e., normal_1, leukemia_1 and leukemia_2) before enrichment and after the 1st and 2nd steps of enrichment, with the use of four qPCR assays for targeted regions and three qPCR assays for non-targeted regions. The concentration of all samples on each step was measured with the use of the Qubit™ 1.0 Fluorometer (Invitrogen, Life Technologies, Grand Island, NY, USA). For the analyses of leukemia_1, leukemia_2, and normal_1 samples, 1 ng of input DNA from each step of enrichment were used. The sequences of the primers and numbers of probes applied within the individual assays were as follows: ta_1p36 assay – Fwd: 5'-AATAGGGCCTGAGGGAAAC-3', Rev: 5'-GTAGCGGCTAGGAGAATACAT-3', probe #57; ta_5p13 assay – Fwd: 5'-ACAGGCCTC TTGGTCTTGT-3', Rev: 5'-TGTTGCGAAGCTCT TTGGT-3', probe #17; ta_7p11 assay – Fwd: 5'-GCCAAAAGTGTGATCCAAG-3', Rev: 5'-GTTTCT GGCAGTTCTCCTC-3', probe #3; ta_21q21 assay – Fwd: 5'-GCATTAACAGTGTATGATGCC-3', Rev: 5'-TTATCCAGCAGGGTGACTC-3', probe #88; nt_2q37 assay – Fwd: 5'- CCCCAAAAAAATCCTCAGA-3', Rev: 5'- TGGGCTGAAGTTGCTGTAG-3', probe #60; nt_9q34 assay – Fwd: 5'-ATACTGAGAGGGAAACAGCAG-3', Rev: 5'-CATAAGCTCCACTTACTGGC-3', probe #41; nt_17q22 assay – Fwd: 5'- AAAGTC CTGACTCCCTCACT-3', Rev: 5'- AGAAGTGG GACCAGTGTCT-3', probe #49. qPCR assays were run for 45 cycles and if particular assay was amplified in or after 36 cycle, we assumed that the Ct value equals 36. For each assay PCR efficiencies (E) were measured based on a standard curve analysis. Fold enrichment (FE^{qPCR}) was calculated as $E^{\Delta Ct}$, where ΔCt is the difference between Ct values of a non-enriched library (reference) and an enriched library (either after the 1st or 2nd enrichment), as proposed by Roche NimbleGen and [16].

Statistical analysis

All statistical analyses were performed using GraphPad QuickCalcs [29]. All statistical tests were two-sided and p-values less than 0.05 were considered statistically significant.

ACKNOWLEDGMENTS

This work was supported by the National Science Centre [2015/17/B/NZ2/01182, 2014/12/W/NZ2/00466]; and Polish Ministry of Science and Higher Education KNOW program. We would like to thank the Laboratory of Subcellular Structures Analysis of the Institute of Bioorganic Chemistry, Polish Academy of Sciences in Poznan, for the opportunity to use LightCycler 480 system (Roche).

CONFLICTS OF INTEREST

All authors declare no conflicts of interest.

Authors' contributions

KK – designed MTTE probes/assay, performed all but NGS experiments and analyses, elaborated the enrichment-associated parameters, substantially participated in manuscript, tables, figures and supplementary materials preparation; LH – prepared libraries and performed NGS experiments, participated in manuscript preparation; AS – performed computational analyses of NGS data; MF – participated in manuscript preparation and study design; PK – conceived the study, supervised and coordinated all experiments, participated in all analyses, elaborated the enrichment-associated parameters, prepared manuscript. All authors read and approved the final draft.

REFERENCES

1. Chilamakuri CS, Lorenz S, Madoui MA, Vodak D, Sun J, Hovig E, Myklebost O, Meza-Zepeda LA. Performance comparison of four exome capture systems for deep sequencing. BMC Genomics. 2014; 15:449.
2. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet. 2016; 135:359-362.
3. Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ. Targeted enrichment of genomic DNA regions for next-generation sequencing. Brief Funct Genomics. 2011; 10:374-386.
4. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010; 42:30-35.
5. Samuels DC, Han L, Li J, Quanghu S, Clark TA, Shyr Y, Guo Y. Finding the lost treasures in exome sequencing data. Trends Genet. 2013; 29:593-599.
6. Wang Z, Liu X, Yang BZ, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. Front Genet. 2013; 4:160.
7. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol. 2011; 29:908-914.
8. Asan, Xu Y, Jiang H, Tyler-Smith C, Xue Y, Jiang T, Wang J, Wu M, Liu X, Tian G, Wang J, Wang J, Yang H, et al. Comprehensive comparison of three commercial human whole-exome capture platforms. Genome Biol. 2011; 12:R95.
9. Parla JS, Iossifov I, Grabill I, Spector MS, Kramer M, McCombie WR. A comparative analysis of exome capture. Genome Biol. 2011; 12:R97.
10. Sulonen AM, Ellonen P, Almusa H, Lepisto M, Eldfors S, Hannula S, Miettinen T, Tyynismaa H, Salo P, Heckman C, Joensuu H, Raivio T, Suomalainen A, et al. Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol. 2011; 12:R94.
11. Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, Xu Z, Steinmann B, Carrel T, Rothlisberger B, Schlapbach R, Bruggmann R, Matyas G. New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res. 2015; 43:e76.
12. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, Ashikawa K, Kubo M, Tsunoda T. Performance comparison of four commercial human whole-exome capture platforms. Sci Rep. 2015; 5:12742.
13. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. Hum Mutat. 2015; 36:815-822.
14. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, Richard McCombie W, Hannon GJ. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat Protoc. 2009; 4:960-974.
15. Henry IM, Nagalakshmi U, Lieberman MC, Ngo KJ, Krasileva KV, Vasquez-Gross H, Akhunova A, Akhunov E, Dubcovsky J, Tai TH, Comai L. Efficient Genome-Wide Detection and Cataloging of EMS-Induced Mutations Using Exome Capture and Next-Generation Sequencing. Plant Cell. 2014; 26:1382-1397.
16. Salmon A, Udall JA, Jeddeloh JA, Wendel J. Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. G3: Genes| Genomes| Genetics. 2012; 2:921-930.
17. 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.
18. Kozlowski P, Jasinska AJ, Kwiatkowski DJ. New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis. 2008; 29:4627-4636.
19. 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.
20. 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.
21. Gatta V, Scarciolla O, Gaspari AR, Palka C, De Angelis MV, Di Muzio A, Guanciali-Franchi P, Calabrese G, Uncini A, Stuppia L. Identification of deletions and duplications of the DMD gene in affected males and carrier females by multiple ligation probe amplification (MLPA). Hum Genet. 2005; 117:92-98.
22. De Luca A, Bottillo I, Dasdia MC, Morella A, Lanari V, Bernardini L, Divona L, Giustini S, Sinibaldi L, Novelli A, Torrente I, Schirinzi A, Dallapiccola B. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. J Med Genet. 2007; 44:800-808.
23. 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; 6:4716. doi: 10.18632/oncotarget.4351.
24. Coriell Institute for Medical Research https://www.coriell.org/.

25. 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.

26. Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P. An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example. Hum Mutat. 2013; 34:763-773.

27. Marcinkowska-Swojak M, Szczerbal I, Pausch H, Nowacka-Woszuk J, Flisikowski K, Dzimira S, Nizanski W, Payan-Carreira R, Fries R, Kozlowski P, Switonski M. Copy number variation in the region harboring SOX9 gene in dogs with testicular/ovotesticular disorder of sex development (78, XX; SRY-negative). Sci Rep. 2015; 5:14696.

28. 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.

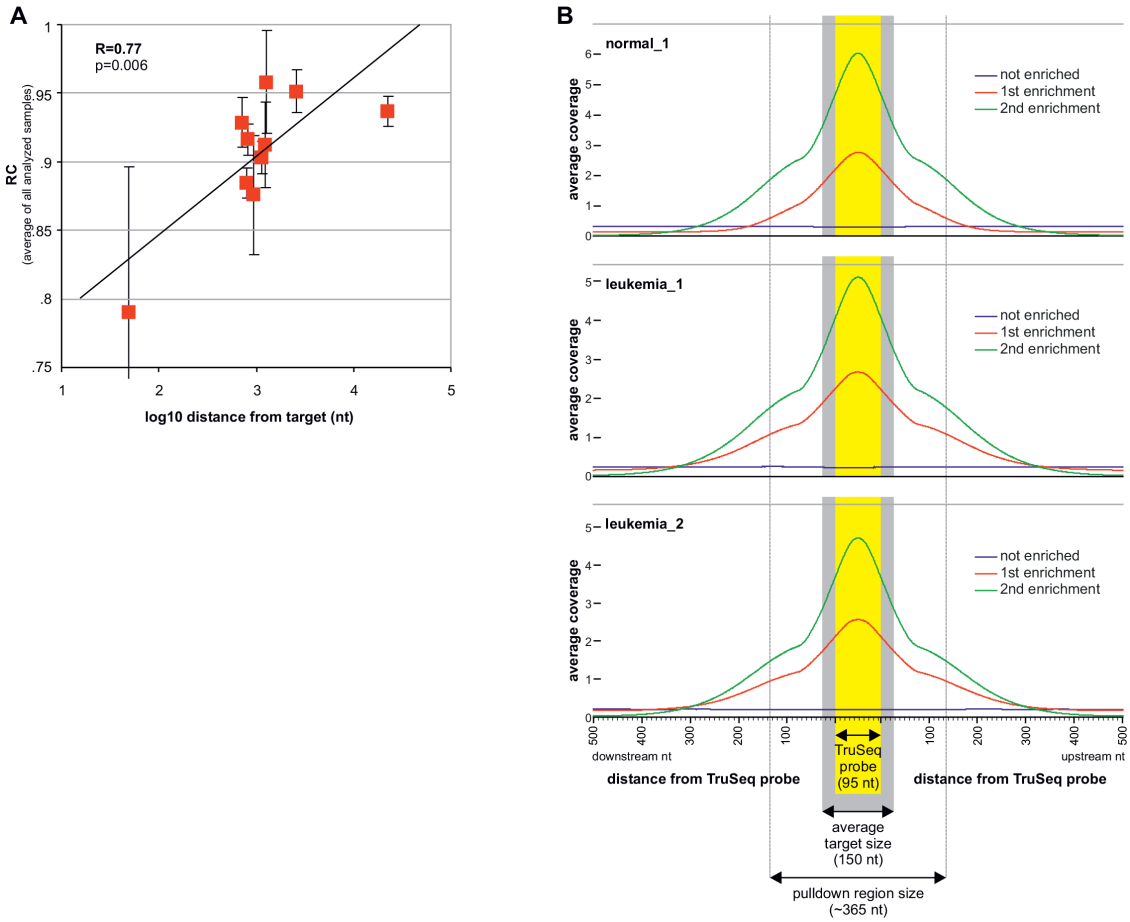
29. GraphPad QuickCalcs http://www.graphpad.com/quickcalcs/.

30. Ensembl http://www.ensembl.org.

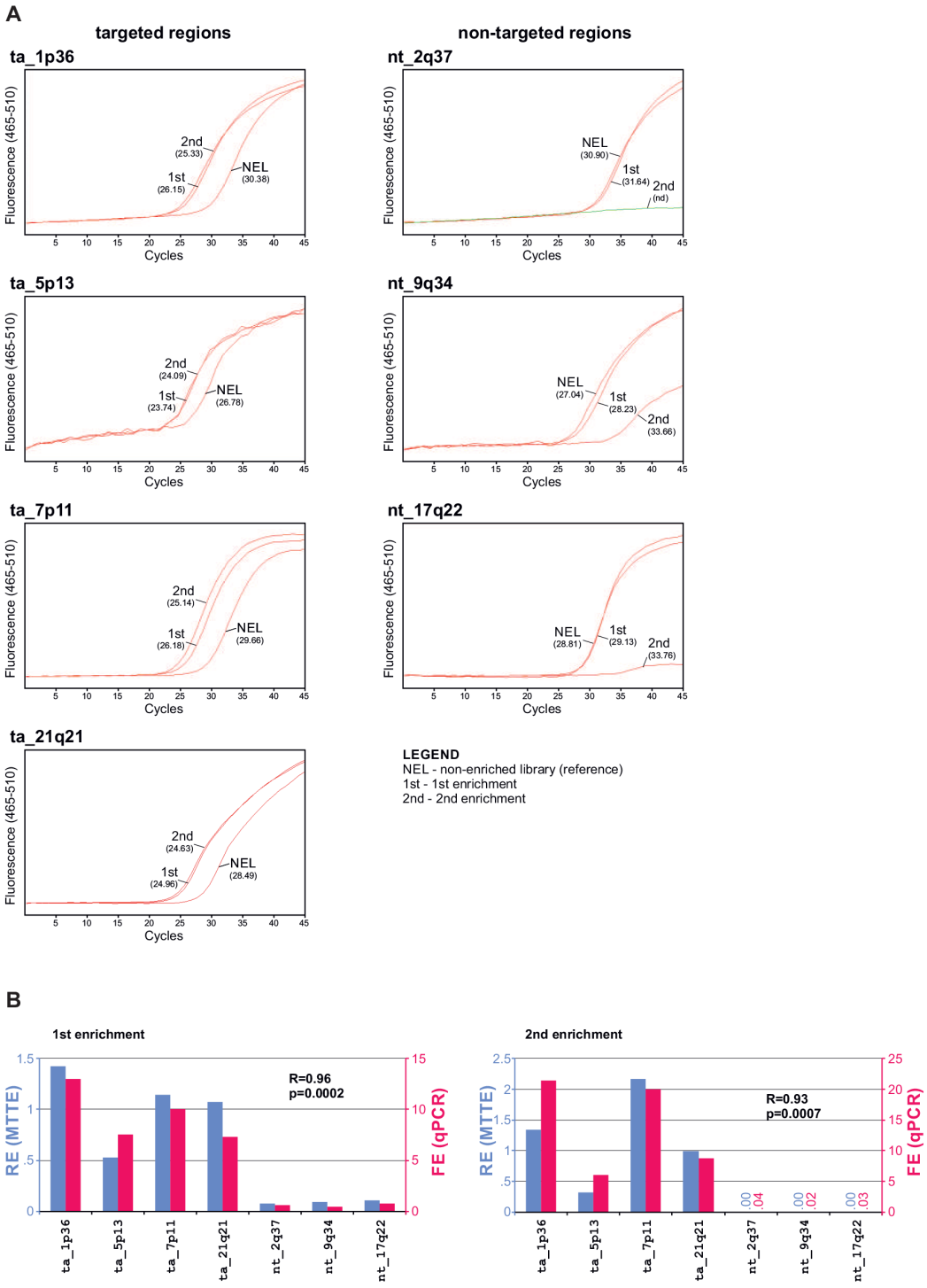
MTTE: an innovative strategy for the evaluation of targeted/ exome enrichment efficiency

SUPPLEMENTARY FIGURES AND TABLE

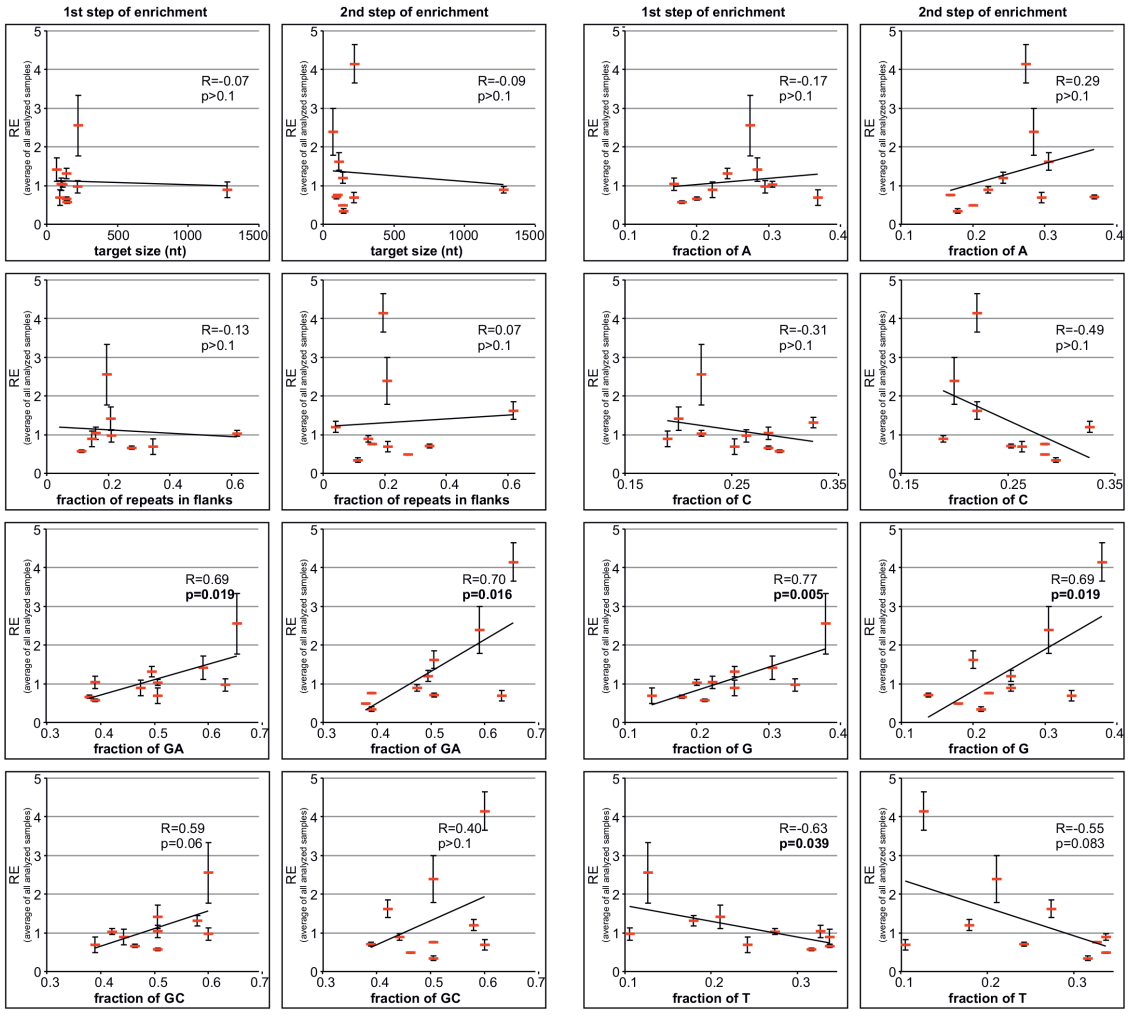
Supplementary Figure S1: Results of the analysis conducted using the in-house designed MTTE assay. The bar graphs represent the relative enrichment (y-axis) of each analyzed region (x-axis), obtained in the analysis of specimens from steps (i-v) of exome enrichment of (from the left) normal_1, leukemia_1, and leukemia_2 samples. The corresponding RC values are indicated on the graphs (steps iv-v).



Supplementary Figure S2: Average coverage of the sequences surrounding the targeted sequences. A. The relationship between the clearance efficiency calculated based on the MTTE results and increasing distance from the targeted regions (log10 values). Average (red mark) and standard deviation (SD, error bars) of RC in all analyzed samples (n=3) are indicated. Correlation coefficient (R), p-value for correlation, and linear correlation trend line are shown on the graph. **B.** Plots illustrating an average NGS coverage in relation to the position of the TruSeq exome enrichment probe. For the analysis, we selected 77000 probes that are distant to each other by at least 1000 nucleotides. A targeted region and a pulldown region are indicated on the graph. The pulldown region represents the sequence that may be captured from the library assuming the average size of the library fragments (inserts) to be 230 bp (see technical notes regarding the TruSeq Exome Enrichment Kit).



Supplementary Figure S3: Evaluation of the enrichment efficiency with the use of qPCR analysis. A. Representative results of qPCR assays designed for the selected targeted (left-hand side) and non-targeted (right-hand side) regions. The curves representing non-enriched library and libraries after the first and second round of enrichments are indicated on the graphs (corresponding Ct values are indicated in brackets). **B.** Comparison of the RE and FE values (calculated based on MTTE and qPCR results, respectively) corresponding to the selected targeted and non-targeted regions analyzed by both MTTE and q PCR. The figure illustrates representative results obtained for the leukemia_1 sample. The results of the other analyzed samples are very similar.



Supplementary Figure S4: Correlation of RE values (y-axis) of targeted probes with potential enrichment-affecting factors (x-axis). Average (red mark) and SD (error bars) of RE in all analyzed samples (n=3) are indicated.R, p-value for correlation, and linear correlation trend line are shown on the graph. Graphs in the left and right columns show the results after the first and second rounds of enrichment, respectively. For the correlation analysis involving fraction of repetitive elements, we took into account a region of 2.5 kb upstream and downstream of the center of TruSeq exome enrichment probes.

Supplementary Table S1: Detailed characteristics and sequences of the probes included in the MTTE assay

See Supplementary File 1

Supplementary Table S1. Detailed characteristics and sequences of the probes included in the MTTE assay.

probe id	former probe id	use of probe in former projects	gene	probe location (hg19)	5'PSS	length	5'SS	length	5'TSS	length	Tm	5'HPL	3'TSS	length	Tm	3'SS	length	3'PSS	length	3'HPL	TPL
ta_22q12	Control1 [chr22]	(19)	*NF2*, exon 12	chr22:30,069,296-30,069,338	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	46	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	47	93
ta_5p13	DROSHA_1	(21)	*DROSHA*, exon 30	chr5:31,410,867-31,410,908	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	GCCAAGGTCTTGGTGCGAAGC	21	73.2	48	GCCACAGGCCTCTTGGTCTTG	21	72.3	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	48	96
nt_9q34	3'TSC1_2	(19)	*C9orf9*, intron 2	chr9:135,761,779-135,761,825	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	AAATTCCAGCGTGGGCCACCCAT	23	74.7	49	ACTGAGAGGGAAACAGCAGGGACT	24	70.5	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	50	99
nt_17q22	mir-142_4	(23)	*BZRAP1-AS1*, intron 2	chr17:56,409,474-56,409,517	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	CAGGAGCCCAAGGCTATCCCAA	22	73.1	51	CAAAAATGGTGGCCATGTTGGG	22	71.3	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	51	102
nt_2q35	BARD1_f5'	(22)	*LOC101928103*, intron 1	chr2:215,675,559-215,675,612	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	CTGGCACCTGGATAGACACTTGGATAT	27	70.5	52	CAGCTAGAAAGCTACGACAGGAAACCA	27	71.1	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	53	105
nt_10q26	mir-202_2	(23)	intergenic region	chr10:135,062,340-135,062,384	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	AAGGAACAAGGCTGAGGCCTCA	22	70.0	54	GAGGCCTCTTCAGTGACCATGTC	23	70.8	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	54	108
ta_1q22	Control2 [chr1]	(19)	*LMNA*, exon 6	chr1:156,105,818-156,105,862	GGGTTCCCTAAGGGTTGGA	19	cgctactact	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	55	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	56	111
ta_7p11	EGFR_e5	(20,24)	*EGFR*, exon 5	chr7:55,218,989-55,219,034	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	CAAAAGTGTGATCCAAGCTGTCCCA	25	72.0	57	ATGGGAGCTGCTGGGGTGCAG	21	74.6	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	57	114
ta_16p13	TSC2_exon23	(19)	*TSC2*, exon 23	chr16:2,126,095-2,126,137	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	CCCATCACGTCATAGCCATGTG	22	70.2	58	GTTCATCAGGTGCCGCCTGCC	21	76.0	tcaaactaaatctac	15	TCTAGATTGGATCTTGCTGGCGC	23	59	117
ta_14q32	DICER1_2	(21)	*DICER1*, exon 14	chr14:95,577,662-95,577,708	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	GGTAGCACTGCCTTCGTTTCGTG	23	71.5	60	GAACCTGGTCTTCCTGGAACACTG	24	71.5	aaactaaatctac	13	TCTAGATTGGATCTTGCTGGCGC	23	60	120
ta_2q35	BARD1_e09	(22)	*BARD1*, exon 9	chr2:215,609,825-215,609,881	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagt	16	ACTCATGTTGTTGTTCCTGGTGATGCA	27	71.6	62	GTTCAAAGTACCTTGAAGTGTATGCTTGGG	30	71.5	taaatctac	9	TCTAGATTGGATCTTGCTGGCGC	23	62	124
nt_3p21	mir-566_2	(23)	*SEMA3F*, intron	chr3:50,210,365-50,210,410	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattg	23	ACCCTGAGCCTGGGAGCATTAA	22	70.0	64	CCCCCACCTGAGCAAAAACTCTTT	24	71.7	ggtcaaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	64	128
nt_6q13	miR-30a_2	(21)	intergenic region	chr6:72,114,572-72,114,637	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	CTTAGTACTTGGCATTCATATACCACTGCTGAC	33	71.2	66	TAGAGCTTGAGTCAATCCACCAAACTCAGCTAT	33	72.3	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	66	132
tf_2q35	BARD1_e01''	(22)	*BARD1*, intron 1	chr2:215,674,039-215,674,086	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TTTCTGGGGGGCGGCAGAATCTTT	23	72.3	68	TCAAATCTTCCGTTTCCTCCTTCCG	25	71.3	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	68	136
ta_21q21	miR-155_2	(21)	*MIR155HG*, exon 3	chr21:26,946,376-26,946,424	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	GCATTCACATGGAACAAATTGCTGC	25	70.9	70	CGTGGGAGGATGACAAAGAAGCAT	24	71.4	tctaatggtcaaactaaatctac	23	TCTAGATTGGATCTTGCTGGCGC	23	70	140
ta_17p13	Control3 [chr17]	(19)	*ASPA*, exon 5	chr17:3,397,657-3,397,712	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	72	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	144
nt_2q37	mir-149_1	(23)	*GPC1*, intron 1	chr2:241,394,664-241,394,713	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcca	29	CAAAAAAATCCTCAGAGCACTGGGGA	26	72.3	74	AGGGAAGTGTCCCAGTGACTGAGG	24	72.3	tgtatctaatggtcaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	74	148
nt_6p21	mir-1275_1	(23)	intergenic region	chr6:33,967,092-33,967,143	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacc	31	GAAGCTATAGATGCTAAGGCAACCGG	26	70.3	76	GGCTCAGGTTCTCTTCTTCAAAGCTC	26	71.1	tgtatctaatggtcaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	76	152
nt_1p36	ARID1A_i01	(22)	*ARID1A*, intron 1	chr1:27,026,605-27,026,657	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctt	34	AGCTTCCACTTCTGTGGACTGCTTC	25	70.5	78	AGGTGTGTGGTAGTAGTCTAGGTGAGGG	28	70.9	tgtatctaatggtcaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	78	156
ta_1p36	ARID1A_e16	(22)	*ARID1A*, exon 16	chr1:27,100,117-27,100,158	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttcagct	40	CCACAGCCGAATCTCATGCCT	21	71.1	80	TCCAACCCAGACTCGGGGATG	21	72.6	tttgcgaaatgtatctaatggtcaactaaatctac	36	TCTAGATTGGATCTTGCTGGCGC	23	80	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 length

TPL - total probe length

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA

SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

AC# V00604 Phage M13 genome

position: 3-99

5'-cgctactactattagtagaattgatgccacctttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccatttgcgaaatgtatctaatggtcaaactaaatctac-3'

OŚWIADCZENIA WSPÓŁAUTORÓW

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 4 sierpnia 2017

dr hab. Piotr Kozłowski, prof. IChB PAN
Kierownik Zakładu Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu

OŚWIADCZENIA

Dotyczy rozprawy doktorskiej mgr Katarzyny Klonowskiej:

Mgr Katarzyna Klonowska wykonywała pracę doktorską w Instytucie Chemii Bioorganicznej PAN od 2014 roku. Praca doktorska mgr Katarzyny Klonowskiej jest częścią kierowanych przeze mnie projektów badawczych finansowanych w ramach grantów Narodowego Centrum Nauki, tj. OPUS1 pt. „Identyfikacja nowych genów odgrywających ważną rolę w procesie nowotworzenia w częstych nowotworach człowieka" oraz OPUS 9 pt. „Analiza roli genu *BARD1*, jako czynnika warunkującego genetyczną predyspozycję do raka piersi - zastosowanie unikalnej strategii, analizy asocjacji mutacji założycielskich", a także kierowanego przez mgr Katarzynę Klonowską grantu Narodowego Centrum Nauki PRELUDIUM 10 pt. „Delecja genu *APOBEC3B* - analiza jej roli w rodzinnej predyspozycji do raka piersi oraz określenie jej funkcjonalności na poziomie mRNA".

Jako że od początku byłem opiekunem naukowym mgr Katarzyny Klonowskiej, a od 21 października 2015 roku również jej promotorem, miałem możliwość zauważyć, iż mimo że jest ona młodym naukowcem, w pracy charakteryzuje się cechami właściwymi dojrzałym badaczom. Mgr Katarzyna Klonowska jest bardzo silne zaangażowana w pracę naukową, w której wykazuje się dużą samodzielnością. Chociaż zaangażowanie i samodzielność uważam za najistotniejsze cechy w kontekście charakterystyki młodego naukowca, chciałbym wymienić również: pracowitość, zdolność szybkiego uczenia się i rozpoznawania nowej tematyki oraz systematyczność. Pozwoliło jej to zdobyć doświadczenie i wiedzę do tego, aby wyniki swoich badań mogła w powodzeniem wpisać w szeroki kontekst ogólnego stanu wiedzy dotyczącej genetycznej predyspozycji do nowotworów.

Wynikiem rozprawy doktorskiej mgr Katarzyny Klonowskiej jest sześć publikacji, w których przygotowaniu miała ona znaczący udział (w pięciu z nich jest pierwszym autorem). Moja

rola, jako głównego autora pięciu z publikacji wchodzących w skład rozprawy doktorskiej mgr Katarzyny Klonowskiej, polegała za zaplanowaniu i koordynacji badań, pozyskaniu środków oraz przygotowaniu manuskryptów.

Współautorstwo moje i mgr Katarzyny Klonowskiej w publikacji Ratajska i wsp., Oncology Reports, 2015, która również została włączona do rozprawy, jest wynikiem współpracy z dr Magdaleną Ratajską i prof. Januszem Limonem z Gdańskiego Uniwersytetu Medycznego w ramach której byliśmy odpowiedzialni na przeprowadzenie jednego z etapów większego projektu, tematycznie związanego z przedmiotem rozprawy doktorskiej mgr Katarzyny Klonowskiej.

Poniżej przedstawiam zakres prac wykonanych przez mgr Katarzynę Klonowską, oraz mój udział w poszczególnych publikacjach:

- Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P.
  Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene
  BioTechnologia, 2014, 95(3):203-214.

  Rola mgr Katarzyny Klonowskiej w przygotowaniu niniejszej pracy przeglądowej polegała na przygotowaniu manuskryptu oraz sporządzeniu tabeli i rycin. Niezbędnym w tym celu środkiem było dokonanie przez doktorantkę przeglądu adekwatnej literatury oraz wybór najważniejszych zagadnień związanych z czynnikami ryzyka raka piersi i/lub jajnika, a także selekcja genów związanych z ryzykiem tych nowotworów oraz syndromów dziedzicznych predysponujących do nowotworów podlegających agregacji w rodzinie.
  Praca nad tym artykułem była dla doktorantki doskonałą okazją do dogłębnego zapoznania się z szeroko pojętą tematyką genetycznej predyspozycji do nowotworów piersi i/lub jajnika, ze szczególnym uwzględnieniem doniesień dotyczących związku genu *BARD1* z tymi nowotworami. Mgr Katarzyna Klonowska samodzielnie zgromadziła większość materiału i w znacznym stopniu samodzielnie przygotowała manuskrypt i ryciny. Jej całkowity udział przygotowaniu pracy szacuję na ~75%.

  Mój udział w przygotowaniu niniejszej pracy polegał głównie na zaplanowaniu koncepcji pracy oraz jej zawartości merytorycznej i koordynacji pracy nad manuskryptem. Ponadto, zapoznałem doktorantkę z zagadnieniami będącymi tłem i bezpośrednim przedmiotem publikacji, oraz nadzorowałem jej pracę.

- Marcinkowska-Swojak M*, Klonowska K*, Figlerowicz M, Kozlowski P.
  An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs
  Gene, 2014, 546(2):257-62.
  *Autorzy mieli taki sam wkład w przygotowanie publikacji

  Mgr Katarzyna Klonowska przeprowadziła w tej pracy eksperymenty MLPA i analizy wyników dotyczące wielo-allelicznych polimorfizmów liczby kopii (CNV) związanych z powszechnymi chorobami człowieka, w tym CNV genów *CCL3L1* i *CCL4L1*, wiązanego z podatnością na infekcję wirusem HIV, CNV genu *UGT2B17* związanego z występowanie osteoporozy, oraz CNV grupy genów β-defensyn, związanych z występowaniem

łuszczycy. Eksperymenty wykonywane były przez mgr Katarzynę Klonowską we współpracy z dr inż. Małgorzatą Marcinkowską-Swojak, która zaprojektowała używane w badaniach testy. We współpracy z dr inż. Małgorzatą Marcinkowską-Swojak, mgr Katarzyna Klonowska przygotowała ryciny i materiały suplementarne do pracy oraz brała udział w przygotowaniu manuskryptu. Udział mgr Katarzyny Klonowskiej (jak również dr inż. Małgorzaty Marcinkowskiej-Swojak) w części eksperymentalnej oceniam na 50% (każda), co zostało odzwierciedlone we wspólnym pierwszym autorstwie, a całkowity wkład mgr Katarzyny Klonowskiej szacuję na 25%.

Mój udział w niniejszej publikacji polegał na zaplanowaniu koncepcji badań. Nadzorowałem wszystkie eksperymenty i analizy. Pomagałem również od strony merytorycznej w interpretacji wyników, przeprowadzeniu i wyborze testów statystycznych i przygotowaniu manuskryptu.

- 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, Irminger-Finger I, Limon J, 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.

  Mgr Katarzyna Klonowska jest autorem niemal wszystkich eksperymentów i analiz, brała również udział w merytorycznej stronie przygotowania publikacji. Bardziej szczegółowo, na potrzeby powyższej publikacji mgr Katarzyna Klonowska zaprojektowała test MLPA do analizy dużych mutacji w genie *BARD1*, przeprowadziła walidację testu MLPA oraz wykorzystała zoptymalizowany test MLPA do analiz, a także zinterpretowała otrzymane wyniki. Dodatkowo, wykonała analizę mutacji punktowych w genie *BARD1* oraz przygotowała panel „sztucznych" kontroli pozytywnych przy użyciu enzymów restrykcyjnych oraz oligonukleotydów kompetycyjnych. Brała również udział w przygotowaniu manuskryptu, w tym rycin oraz materiałów suplementarnych.
  Wkład Katarzyny Klonowskiej w prace eksperymentalne i analizę danych oceniam na >90%, zaś całkowity wkład szacuję na ~50%.

  Mój udział w przygotowaniu tej publikacji polegał na zaplanowaniu koncepcji badań oraz nawiązaniu i koordynacji współpracy z Katedrą i Zakładem Biologii i Genetyki Medycznej Gdańskiego Uniwersytetu Medycznego. Nadzorowałem wszystkie eksperymenty i analizy. Brałem również udział w przygotowaniu manuskryptu.

- Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I.
  Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms
  Oncology Reports, 2015, 34(5):2609-17.

  Rola mgr Katarzyny Klonowska w przygotowaniu powyższej publikacji polegała na udziale w charakterystyce potencjalnej patogenności mutacji c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G w genie *BARD1*. W celu opracowania charakterystyki mutacji, mgr

Katarzyna Klonowska przeprowadziła analizę *in silico* z wykorzystaniem programów MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji. Udział mgr Katarzyny Klonowskiej w tej publikacji szacuję na około 5%, jednak udział ten stanowił bezpośrednią kontynuację prac związanych z mutacjami wykrytymi w publikacji powyżej (Klonowska i wsp., Sci Rep, 2015).

Moja rola w przygotowaniu powyższej publikacji polegała na zaplanowaniu koncepcji analiz *in silico* prowadzonych przez mgr Katarzynę Klonowską oraz nadzorowaniu pracy doktorantki.

- Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P.
  MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency
  Oncotarget, 2016, 7(41):67266-67276.

Na potrzeby powyższej publikacji, mgr Katarzyna Klonowska wykonała i zinterpretowała wszystkie analizy eksperymentalne poza eksperymentami z użyciem sekwencjonowania nowej generacji (NGS). Zaprojektowała zestaw sond MLPA wchodzących w skład testu MTTE, przeprowadziła jego optymalizację i wykorzystała zoptymalizowany test MTTE do analiz. W celu przeprowadzenia niezależnej walidacji otrzymanych wyników, zaprojektowała i zoptymalizowała zestaw starterów, który wykorzystała do analiz z użyciem metody qPCR z użyciem sond UPL. Brała również udział w opracowaniu strategii MTTE, w tym sposobu obliczania parametrów określających poziom wzbogacenia regionów kodujących genomu w bibliotekach przygotowanych do analizy NGS. Brała także udział w napisaniu manuskryptu, jak również przygotowaniu tabel, rycin i materiałów suplementarnych.
Wkład Katarzyny Klonowskiej w prace eksperymentalne i analizę danych oceniam na 80%, zaś całkowity wkład szacuję na 45%.

Mój udział w przygotowaniu powyższej publikacji polegał na zaplanowaniu koncepcji badań oraz nadzorowaniu i koordynacji wszystkich eksperymentów. Wspólnie z mgr Katarzyną Klonowską opracowałem strategię MTTE oraz sposób obliczania parametrów do walidacji wzbogacenia bibliotek NGS, a także przygotowałem manuskrypt.

- Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P.
  The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population
  Oncotarget, 2017, [Epub ahead of print]

Na potrzeby powyższej publikacji, mgr Katarzyna Klonowska zaprojektowała i zoptymalizowała test A3Bdel_MLPA oraz zastosowała go do analizy struktury delecji *APOBEC3B*. Mgr Katarzyna Klonowska zaprojektowała również test A3Bdel_PCR, który wykorzystała w analizie genotypowania delecji w próbkach z grupy GDANSK. Zaprojektowała również testy A3A_exp, A3B_exp oraz A3A/A3B_exp służące do analizy

ekspresji genów objętych przez delecję oraz zastosowała je w analizie związku genotyp – ekspresja z wykorzystaniem metod ddPCR oraz UPL-qPCR. Mgr Katarzyna Klonowska opracowała rezultaty wszystkich eksperymentów oraz przeprowadziła wszystkie niezbędne analizy statystyczne. Przygotowała również meta-analizę podsumowującą wszystkie dotychczasowe badania asocjacji delecji *APOBEC3B* z nowotworami, włączając w nią naszą analizę. Brała także udział w przygotowaniu tekstu manuskryptu wraz z rycinami, tabelami i materiałami suplementarnymi.
Wkład Katarzyny Klonowskiej w prace eksperymentalne i analizę danych oceniam na 70%, zaś całkowity wkład w przygotowanie pracy szacuję na 50%.

Mój udział w przygotowaniu powyższej publikacji polegał na zaplanowaniu koncepcji badań oraz nadzorowaniu wszystkich analiz, a także koordynacji współpracy z Katedrą i Zakładem Biologii i Genetyki Medycznej Gdańskiego Uniwersytetu Medycznego oraz Zakładem Genetyki i Patomorfologii Pomorskiego Uniwersytetu Medycznego w Szczecinie. Ponadto zapoznałem doktorantkę z zagadnieniami będącymi bezpośrednim przedmiotem publikacji oraz przygotowałem z nią manuskrypt.

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

Piotr Kozłowski

P.S. Udział pozostałych współautorów w poszczególnych publikacjach został określony w indywidualnych oświadczeniach załączonych do rozprawy mgr Katarzyny Klonowskiej. Oświadczenia zostały zebrane od wszystkich współautorów, z wyjątkiem Irmgard Irminger-Finger oraz Geoffa Laurenta, którzy nie odpowiedzieli na nasze wielokrotne prośby w tej sprawie.

Gdańsk, 15.04.2017

dr n. med. Magdalena Ratajska
Katedra i Zakład Biologii i Genetyki
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
mratajska@gumed.edu.pl
tel. +48 58 349-15-33

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P. ***"Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene"*** *Biotechnologia 2014; 95(3): 203-214.*

Powyższa praca powstała w ramach współpracy między Katedrą i Zakładem Biologii i Genetyki Gdańskiego Uniwersytetu Medycznego a Instytutem Chemii Bioorganicznej PAN w Poznaniu

Moja rola w przygotowaniu powyższej pracy przeglądowej polegała na konsultowaniu z głównymi autorami (KK i PK) jej struktury oraz zawartości merytorycznej. Brałam udział w selekcji genów związanych z predyspozycją do raka piersi i/lub jajnika oraz syndromów dziedzicznych predysponujących do nowotworów podlegających agregacji w rodzinie, stanowiących część zagadnień opisanych w powyższej pracy przez mgr Katarzynę Klonowską.

Z poważaniem,

dr n. med. Magdalena Ratajska

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 4.08.2017

Dr hab. Marzena Wojciechowska
Zakład Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
ul. Z. Noskowskiego 12/14
61-704 Poznań
Email: mwojska@gmail.com

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P. **"Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene"** BioTechnologia 2014; 95(3): 203-214.

Moja rola w przygotowaniu powyższej pracy przeglądowej polegała na redagowaniu tekstu manuskryptu oraz jego korekcie pod kątem językowym. Brałam również udział w przygotowaniu rycin zawartych w manuskrypcie.

Z poważaniem,

dr hab. Marzena Wojciechowska

Poznań, 28.04.2015

OŚWIADCZENIE

Oświadczam, iż mój udział w publikacji:

Marcinkowska-Swojak M, Klonowska K, Figlerowicz M, Kozlowski P „An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs" Gene 2014, 546: 257–262.

polegał na zaprojektowaniu sond i testów MLPA do analizy wielo-allelicznych CNV związanych z powszechnymi chorobami człowieka (CNV genu *CCL3L1* wiązanego z zakażeniami wirusem HIV, CNV genu *UGT2B17* związanego z występowaniem osteoporozy i CNV grupy genów β-defensyn, związanych z występowaniem łuszczycy) oraz na zaplanowaniu analiz optymalizacyjnych MLPA. Wszystkie eksperymenty i analizy z wykorzystaniem zaprojektowanych przeze mnie testów wykonała mgr Katarzyna Klonowska. Ponadto, wspólnie z mgr Katarzyną Klonowską, brałam udział w przygotowaniu manuskryptu, rycin i materiałów suplementarnych. Mój udział, jak również udział mgr Katarzyny Klonowskiej w części eksperymentalnej wynosi 50%, co zostało odzwierciedlone we wspólnym pierwszym autorstwie.

Z poważaniem

Marcinkowska-Swojak

dr inż. Małgorzata Marcinkowska-Swojak

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 28.04.2015

prof. dr hab. Marek Figlerowicz
Kierownik Zakładu Biologii Molekularnej i Systemowej
Dyrektor Instytutu Chemii Bioorganicznej PAN w Poznaniu

OŚWIADCZENIE

Dotyczy udziału w publikacji:

Marcinkowska-Swojak M, Klonowska K, Figlerowicz M, Kozlowski P. An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs
Gene 2014, 546: 257–262.

Jako współautor powyższej publikacji oświadczam, iż brałem udział w dyskusji nad koncepcją i strukturą pracy oraz w końcowym etapie przygotowania manuskryptu.

Z poważaniem

prof. dr hab. Marek Figlerowicz

Gdańsk, 17.07.2017

dr n. med. Magdalena Ratajska
Katedra i Zakład Biologii i Genetyki
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
mratajska@gumed.edu.pl
tel. +48 58 349-15-33

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

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, Irminger-Finger I, Limon J, 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.*

Powyższa praca powstała w ramach współpracy między Katedrą i Zakładem Biologii i Genetyki Gdańskiego Uniwersytetu Medycznego a Instytutem Chemii Bioorganicznej PAN w Poznaniu.

Na potrzeby powyższej publikacji, której jestem współautorem, brałam udział w zaplanowaniu badań i przygotowaniu manuskryptu. Byłam również odpowiedzialna za koordynację selekcji oraz scharakteryzowanie próbek DNA pochodzących od pacjentek z rakiem piersi i/lub jajnika, które zostały poddane dalszej analizie pod kątem obecności dużych mutacji w genie *BARD1*. Brałam również udział w izolacji części próbek DNA od wyselekcjonowanych pacjentek z rakiem piersi i/lub jajnika.
Oświadczam, iż mgr Katarzyna Klonowska wykonała wszystkie analizy eksperymentalne, w tym zaprojektowała test MLPA do analizy dużych mutacji w genie *BARD1*, przeprowadziła walidację testu MLPA oraz wykorzystała zoptymalizowany test MLPA do analiz, a także zinterpretowała otrzymane wyniki. Wykonała również analizę mutacji punktowych w genie *BARD1* oraz przygotowała panel „sztucznych" kontroli pozytywnych przy użyciu enzymów restrykcyjnych oraz oligonukleotydów kompetycyjnych. Wykonała również analizy bioinformatyczne i statystyczne, których wyniki zostały opisane w powyższej publikacji. Mgr Katarzyna Klonowska brała również udział w napisaniu manuskryptu oraz przygotowaniu tabel, rycin i materiałów suplementarnych.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

dr n. med. Magdalena Ratajska

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 5 sierpnia 2017

Dr Karol Czubak
Zakład Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu
email: kczubak@ibch.poznan.pl
tel. 61 665 30 34

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:
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, Irminger-Finger I, Limon J, 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.

Jako współautor powyższej publikacji, oświadczam, iż brałem udział w interpretacji części wyników analizy MLPA dużych mutacji w genie *BARD1* (niezależnie od analiz przeprowadzonych przez Katarzynę Klonowską, jako drugi ekspert).

Z poważaniem,

Karol Czubak

dr Karol Czubak

Gdańsk, 18.07.2017

dr n. med. Alina Kuźniacka
Katedra i Zakład Biologii i Genetyki
Gdański Uniwersytet Medyczny
ul. Skłodowskiej-Curie 3a
80-211 Gdańsk

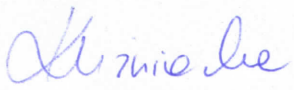
Oświadczenie o współautorstwie

Oświadczam, iż mój udział w publikacji:

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, Irminger-Finger I, Limon J, 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.

polegał na przeprowadzeniu izolacji i przygotowaniu charakterystyki części próbek DNA od pacjentek z nieselekcjonowanym rakiem jajnika które zostały poddane analizie pod kątem występowania dużych mutacji w genie *BARD1* przeprowadzonej w ramach powyższej pracy przez mgr Katarzynę Klonowską.

Z poważaniem,



Gdańsk., 20.07.2017

dr n. med. Izabela Brożek
Katedra i Zakład Biologii i Genetyki
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
izabro@gumed.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

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, Irminger-Finger I, Limon J, 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.*

Jako współautor powyższej publikacji, oświadczam iż moja rolą było udostępnienie i przygotowanie charakterystyki części pacjentek z rodzinnym rakiem piersi i/lub jajnika, które zostały włączone do analizy dużych mutacji w genie *BARD1*.

Z poważaniem,

M. Brożek

dr n. med. Magdalena Koczkowska
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
magda.koczkowska@gumed.edu.pl

Gdańsk, 24.07.2017

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy współautorstwa w publikacji:

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, Irminger-Finger I, Limon J, 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.

Jako współautor powyższej publikacji oświadczam, iż brałam udział w wyselekcjonowaniu i izolacji części próbek DNA od pacjentek z rakiem piersi i/lub jajnika do analizy dużych mutacji w genie *BARD1* wykonanej przez Katarzynę Klonowską.

Z poważaniem,

Magdalena Koczkowska

dr n. med.Magdalena Koczkowska

Gdańsk, 5.07.2017

lek. med. Marcin Śniadecki
Gdański Uniwersytet Medyczny
ul. M. Skłodowskiej-Curie 3a
80-211 Gdańsk

OŚWIADCZENIE

Oświadczam, iż mój udział w publikacji:

„Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example"

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, Irminger-Finger I, Limon J, Kozlowski P. Scientific Reports 2015; 5: 10424.

polegał na wyselekcjonowaniu pacjentek z rakiem jajnika do badań genu *BARD1* prowadzonych w ramach powyższej pracy.

Z poważaniem,

lek. med. Marcin Sniadecki

Gdańsk, 5.07.2017

dr n. med. Jarosław Dębniak
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy udziału w publikacji:

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, Irminger-Finger I, Limon J, 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.

polegał na wyselekcjonowaniu pacjentek z rakiem jajnika do badań genu *BARD1* prowadzonych w ramach powyższej pracy.

Z poważaniem,

Dr n. med. Jarosław Dębniak
specjalista położnictwa i ginekologii
Gdańsk, ul. Hoffa 11A
ZUS 8621880

dr n. med. Jarosław Dębniak

Gdańsk, 11.07.2017

dr hab. n. med. Dariusz Wydra
Katedra i Klinika Ginekologii,
Ginekologii Onkologicznej i Endokrynologii Ginekologicznej
Gdański Uniwersytet Medyczny
dwydra@uck.gda.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

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, Irminger-Finger I, Limon J, 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.

polegał na wyselekcjonowaniu pacjentek z rakiem jajnika do badań genu *BARD1* prowadzonych w ramach powyższej pracy.

Z poważaniem,

Dr hab. n. med. Dariusz Wydra
specjalista położnictwa i ginekologii
endokrynologii ginekologicznej
onkologii ginekologicznej
4850990550[illegible] 2736096

dr hab. n. med. Dariusz Wydra

Gdańsk, 2.08.2017

Magdalena Bałut
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy udziału w publikacji:

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, Irminger-Finger I, Limon J, 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.

Jako współautor powyższej publikacji, oświadczam, iż brałam udział w charakterystyce jednonukleotydowych substytucji zidentyfikowanych w ramach analizy mutacji w genie *BARD1*.

Z poważaniem,

Magdalena Bałut

Magdalena Bałut

Gdynia, 13.07.2017

dr n. med. Maciej Stukan
Gdyńskie Centrum Onkologii
ul. Powstania Styczniowego 1
81-519 Gdynia

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Oświadczam, iż mój udział w publikacji:

„Analysis of large mutations in BARD1 in patients with breast and/or ovarian cancer: the Polish population as an example"

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, Irminger-Finger I, Limon J, Kozlowski P. Scientific Reports 2015; 5: 10424.

polegał na wyselekcjonowaniu pacjentek z rakiem jajnika do badań genu *BARD1* prowadzonych w ramach powyższej pracy.

Z poważaniem,

dr n. med. Maciej Stukan

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 14.08.2017

Dr Agnieszka Żmieńko
Zakład Biologii Molekularnej i Systemowej
Instytut Chemii Bioorganicznej PAN w Poznaniu
Email: akisiel@ibch.poznan.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:
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, Irminger-Finger I, Limon J, 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.

Jako współautor powyższej publikacji, oświadczam, iż brałam udział w interpretacji części wyników analizy MLPA dużych mutacji w genie *BARD1* (niezależnie od analiz przeprowadzonych przez Katarzynę Klonowską, jako drugi ekspert).

Z poważaniem,

Agnieszka Żmieńko

dr Agnieszka Żmieńko

Warszawa, 17.04.2017

dr n. med. Beata Nowakowska
Zakład Genetyki Medycznej
Instytut Matki i Dziecka
ul. Kasprzaka 17a
01-211 Warszawa
tel. 22 32 77 131
beata.nowakowska@imid.med.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

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, Irminger-Finger I, Limon J, 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.

Na potrzeby powyższej publikacji, udostępniłam do analizy próbkę DNA z duplikacją obejmującą gen *BARD1*, która posłużyła jako kontrola pozytywna w badaniach genu *BARD1*.

Z poważaniem,

dr n.med. Beata Nowakowska

Gdańsk, 17 lipca 2017

prof. dr hab. n. med. Janusz Limon
Katedra i Zakład Biologii i Genetyki
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
jlimon@gumed.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

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, Irminger-Finger I, Limon J, 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.

Jako współautor powyższej publikacji, oświadczam, iż moja rola polegała na nadzorowaniu selekcji oraz charakterystyki próbek klinicznych wykorzystanych do analizy dużych mutacji w genie *BARD1*. Brałem również udział w zaplanowaniu koncepcji badań oraz przygotowaniu manuskryptu.

Z poważaniem,

prof. dr hab. n. med. Janusz Limon

Gdańsk, 19.06.2017

dr n. med. Magdalena Ratajska
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
Dębinki 1,
80-211 Gdańsk
mratajska@gumed.edu.pl
tel. +48 58 349-15-33

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako głównej autorki powyższej publikacji polegała na zaprojektowaniu doświadczeń oraz ich przeprowadzeniu. Ponadto byłam zaangażowana w analizę danych i przygotowanie manuskryptu.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

Ratajska

dr n. med. Magdalena Ratajska

Data i miejsce:
02.08.2017 Stanford

Magdalena Matusiak
Stanford School of Medicine
Pathology Department
300 Pasteur Dr.
Lane Building, Room 216
Stanford, CA 94305
USA

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I **"Cancerpredisposing BARD1 mutationsaffectexonskipping and areassociated with overexpression of specific BARD1 isoforms"**OncologyReports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako głównej autorki powyższej publikacji polegała na przeprowadzeniu analizy mutacji genu *BARD1* w grupie pacjentek z nieselekcjonowanym rakiem jajnika.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

Magdalena Matusiak

Gdańsk, 19.06.2017

dr n. med. Alina Kuźniacka
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
Dębinki 1,
80-211 Gdańsk
akuzniacka@gumed.edu.pl
tel. +48 58 349-15-33

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako współautora powyższej publikacji polegała na izolacji materiału genetycznego, analizie części pacjentów oraz pomocy prze analizie uzyskanych wyników.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

dr n. med. Alina Kuźniacka

Gdańsk, 13.06.2017

dr hab. n. med. Bartosz Wasąg
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
bwasag@gumed.edu.pl
tel. +48 58 349 15 31

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako autora powyższej publikacji polegała na analizie mutacji w genie *BARD1* oraz pomocy przy przygotowaniu manuskryptu.

Jednocześnie informuję, że mgr Katarzyna Klonowska przeprowadziła analizę *in silico* wariantów genetycznych c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G w genie *BARD1* z wykorzystaniem następujących programów, *MutPred Splice*, *Human Splicing Finder*, *Rescue ESE*, *PolyPhen 2* oraz *SIFT*. Pozwoliło to ocenić potencjalny wpływ wymienionych substytucji na alternatywny proces składania transkryptu BARD1 i/lub strukturę białka. Otrzymane wyniki wykonanych analiz zostały przedstawione w *Table III*, zawartej w publikacji.

Z poważaniem,

Katedra i Zakład Biologii
i Genetyki Medycznej
Gdańskiego Uniwersytetu Medycznego
dr hab. n. med. Bartosz Wasąg
adiunkt

Gdańsk, 19.06.2017

dr n. med. Izabela Brozek
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
Dębinki 1,
80-211 Gdańsk
izabro@gumed.edu.pl
tel. +48 58 349-15-31

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako współautora powyższej publikacji polegała na przeprowadzeniu analizy kliniczno-rodowodowej wybranych pacjentów. Ponadto byłam zaangażowana w przygotowanie manuskryptu.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

MBrozek

dr n. med. Izabela Brozek

Gdańsk, 02.08.2017

prof. dr hab. Wojciech Biernat
Katedra i Zakład Patomorfologii
Gdański Uniwersytet Medyczny
ul. Mariana Smoluchowskiego 17
80-214 Gdańsk
biernat@gumed.edu.pl
tel. +48 58 349-37-40

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I **"Cancerpredisposing BARD1 mutationsaffectexonskipping and areassociated with overexpression of specific BARD1 isoforms"**OncologyReports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako współautora w/w pracy polegała na przeprowadzeniu oceny histopatologicznej przypadków. Ponadto byłem zaangażowany w analizę wyników i przygotowanie manuskryptu.

W przypadku dodatkowych pytań proszę o kontakt.

Z poważaniem,

prof. dr hab. Wojciech Biernat

Gdańsk, 26.07.2017

dr n. med. Magdalena Koczkowska
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
Ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
magda.koczkowska@gumed.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy współautorstwa w publikacji:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I. ***"Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms"*** *Oncology Reports 2015; 34(5):2609-17.*

Oświadczam, iż moja rola jako współautora powyższej publikacji polegała na izolacji materiału genetycznego, analizie części pacjentów oraz pomocy przy analizie uzyskanych wyników.

Z poważaniem,

Magdalena Koczkowska

dr n. med. Magdalena Koczkowska

Gdańsk, 19.06.2017

Dr n. med. Jarosław Dębniak
Katedra i Ginekologii, Ginekologii Onkologicznej i Endokrynologii Ginekologicznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
jdebniak@wp.pl
tel. +48 58 349-34-86

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako autora powyższej publikacji polegała na sekcji pacjentek do badania, analizie danych klinicznych oraz pomocy w przygotowaniu manuskryptu.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

Dr n. med. Jarosław Dębniak
specjalista położnictwa i ginekologii
Gdańsk, ul. Heweliusza 10A
PWZ 3621880

Jarosław Dębniak

Gdańsk, 19.06.2017

Lek. Marcin Śniadecki
Katedra i Ginekologii, Ginekologii Onkologicznej i Endokrynologii Ginekologicznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
marcinsniadecki@gumed.edu.pl
tel. +48 58 349-34-86

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako autora powyższej publikacji polegała na sekcji pacjentek do badania, analizie danych klinicznych oraz pomocy w przygotowaniu manuskryptu.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

MARCIN ŚNIADECKI
LEKARZ
[illegible]
tel. 501-337-941

Marcin Śniadecki

Maxim Pilyugin, PhD
Senior scientific collaborator
Laboratory of Molecular Gynecology and Obstetrics
University Hospitals of Geneva, Switzerland
T. +41 22 372 44 53
e-mail: maxim.pilyugin@unige.ch

To whom it may concern:

Dear Sir, Madam,

As a co-author of the article ***"Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms"*** (*Oncol Rep. 2015 Nov;34(5):2609-17. doi: 10.3892/or.2015.4235. Epub 2015 Sep)* I was involved in the design of PCR experiments, results discussion, article writing and editing and figures preparation.

If you have any additional questions do not hesitate to contact me.

Sincerely,

Maxim Pilyugin, PhD  Geneva, 27.06.2017

Gdańsk, 19.06.2017

Dr hab. n. med. Dariusz Wydra
Katedra i Ginekologii, Ginekologii Onkologicznej i Endokrynologii Ginekologicznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
dwydra@uck.gda.pl
tel. +48 58 349-34-86

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako autora powyższej publikacji polegała na sekcji pacjentek do badania, analizie danych klinicznych oraz pomocy w przygotowaniu manuskryptu.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

p.o. KIEROWNIK
Katedry i Kliniki Ginekologii,
Ginekologii Onkologicznej
i Endokrynologii Ginekologicznej GUMed

Dariusz Wydra

dr hab. n. med. Dariusz Wydra

Gdańsk, 19.06.2017

prof. dr hab. Janusz Limon
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
Dębinki 1,
80-211 Gdańsk
jlimon@gumed.edu.pl
tel. +48 58 349-15-31

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I "Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms" Oncology Reports 2015; 34(5):2609-17.

Oświadczam, iż moja rola jako współautora powyższej publikacji polegała na przeprowadzeniu analizy rodowodowo-klinicznej pacjentów, pomocy przy analizie uzyskanych wyników oraz udziale w przygotowaniu tekstu niniejszej publikacji.

Jednocześnie informuję iż mgr Katarzyna Klonowska przeprowadziła analizę *in silico* następujących wariantów genetycznych genu *BARD1*: c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G. W celu opracowania charakterystyki mutacji, mgr Katarzyna Klonowska użyła następujących programów: MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

prof. dr hab. Janusz Limon

Pomorski Uniwersytet Medyczny w Szczecinie

Szczecin,15.08.2017

dr Wojciech Kluźniak
Zakład Genetyki i Patomorfologii
Międzynarodowe Centrum Nowotworów Dziedzicznych (MCND)
Pomorski Uniwersytet Medyczny w Szczecinie
ul. Unii Lubelskiej 1
71-252 Szczecin
email: kluzniak.w@gmail.com

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:
Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Jako współautor powyższej publikacji oświadczam, iż moją rolą w przygotowaniu pracy było przeprowadzenie analizy genotypowania delecji genu *APOBEC3B* w panelu próbek DNA z grupy SZCZECIN. Do genotypowania wykorzystałem test A3Bdel_PCR opracowany przez Katarzynę Klonowską.

Z poważaniem

Pomorski Uniwersytet Medyczny w Szczecinie

Szczecin, 20.08.2017

Bogna Rusak
Zakład Genetyki i Patomorfologii
Międzynarodowe Centrum Nowotworów Dziedzicznych (MCND)
Pomorski Uniwersytet Medyczny w Szczecinie
ul. Unii Lubelskiej 1
71-252 Szczecin
email: b_rusak@yahoo.com

OŚWIADCZENIE

Oświadczam, iż mój udział w przygotowaniu poniższej publikacji:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

polegał na przeprowadzeniu analizy genotypowania delecji genu *APOBEC3B* z użyciem testu A3Bdel_PCR w panelu próbek DNA z grupy SZCZECIN.

Z poważaniem,

Bogna Rusak

Zakład Genetyki i Patomorfologii
Pomorski Uniwersytet Medyczny (PUM) w Szczecinie
ul. Unii Lubelskiej 1
71-252 Szczecin
email: aniaj@pum.edu.pl

Szczecin, 08. 08. 2017

Dotyczy współautorstwa w publikacji:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P.
The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population.
Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Jako współautor powyższej pracy oświadczam, że mój udział w przygotowaniu w/w pracy polegał na wyselekcjonowaniu próbek DNA z grupy badanej „SZCZECIN" do analizy delecji *APOBEC3B* oraz opracowaniu charakterystyki klinicznej pacjentek. Brałam także udział w dyskusji nad koncepcją pracy.

Z poważaniem,

dr hab. n. med. Anna Jakubowska, prof. PUM

Gdańsk, 30.07.2017

dr n. med. Magdalena Ratajska
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
mratajska@gumed.edu.pl
tel. +48 58 349-15-33

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Kłonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017

Jako współautor powyższej publikacji oświadczam, iż moją rolą była koordynacja współpracy pomiędzy Zakładem Genetyki Molekularnej, IChB PAN w Poznaniu i Katedrą i Zakładem Biologii i Genetyki Medycznej Gdańskiego Uniwersytetu Medycznego. Byłam również odpowiedzialna za selekcję i scharakteryzowanie próbek DNA od pacjentów z grupy GDANSK.

Z poważaniem,

Ratajska Magdalena

dr n. med. Magdalena Ratajska

Gdańsk, 5.08.2017

Mgr Natalia Krawczyńska
Katedra i Zakład Biologii i Genetyki Medycznej
Gdański Uniwersytet Medyczny
ul. Marii Skłodowskiej-Curie 3a
80-211 Gdańsk
nataliakrawczynska@gumed.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. ***The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population.*** Oncotarget, 2017

Oświadczam, ż na potrzeby powyższej publikacji przeprowadziłam część analiz genotypowania delecji genu *APOBEC3B* w panelu próbek DNA od pacjentów z grupy VILNIUS.

Z poważaniem,

Krawczyńska Natalia

Natalia Krawczyńska

Wilno, 1.09.2017

Oświadczam, iż mój udział w poniższej pracy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

polegał na udostępnieniu próbek DNA z grupy LITHUANIA wraz z ich charakterystyką kliniczną. Brałam również udział w analizie próbek z grupy LITHUANIA pod kątem występowania delecji w genie *APOBEC3B*. Do analizy wykorzystywałam test A3Bdel_PCR oraz procedurę opracowaną w Poznaniu przez Katarzynę Klonowską.

Z poważaniem,

Danuta Vasilevska
Department of Gynecology,
Centre of Obstetrics and Gynecology,
Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 5 sierpnia 2017

dr Karol Czubak
Zakład Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu
email: kczubak@ibch.poznan.pl
tel. 61 665 30 34

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Jako współautor powyższej publikacji oświadczam, iż moją rolą w przygotowaniu pracy był (razem z dr hab. Marzeną Wojciechowską) udział w hodowli linii komórkowych HapMap oraz przeprowadzenie izolacji RNA i reakcji RT PCR w celu pozyskania cDNA do analizy ekspresji przeprowadzonej przez Katarzynę Klonowską.

Z poważaniem,

Karol Czubak

dr Karol Czubak

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 9.08.2017

dr hab. Marzena Wojciechowska
Zakład Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
ul. Z. Noskowskiego 12/14
61-704 Poznań
Email: mwojska@gmail.com

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Jako współautorka powyższej publikacji oświadczam, iż moją rolą w przygotowaniu pracy był (razem z dr Karolem Czubakiem) udział w hodowli linii komórkowych HapMap oraz przeprowadzenie izolacji RNA i reakcji RT PCR w celu pozyskania cDNA do analizy ekspresji przeprowadzonej przez Katarzynę Klonowską.

Z poważaniem,

Wojciechowska Marzena

dr hab. Marzena Wojciechowska

Szczecin, 20.08.2017

Prof. Cezary Cybulski
Zakład Genetyki i Patomorfologii
Międzynarodowe Centrum Nowotworów Dziedzicznych
Pomorski Uniwersytet Medyczny w Szczecinie
ul. Unii Lubelskiej 1
71-252 Szczecin
email: cezarycy@sci.pum.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P. **"The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population."** Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Jako współautor powyższej publikacji byłem odpowiedzialny za koordynację współpracy pomiędzy Zakładem Genetyki Molekularnej IChB PAN w Poznaniu oraz Zakładem Genetyki i Patomorfologii PUM w Szczecinie oraz nadzorowałem analizę genotypowania delecji genu *APOBEC3B* prowadzonej w panelu próbek z grupy SZCZECIN. Brałem również udział w przygotowaniu manuskryptu.

Proszę o kontakt w przypadku dodatkowych pytań.

Z poważaniem,

Szczecin, 20.08.2017

Prof. dr hab. Jan Lubiński
Zakład Genetyki i Patomorfologii
Międzynarodowe Centrum Nowotworów Dziedzicznych
Pomorski Uniwersytet Medyczny w Szczecinie
ul. Unii Lubelskiej 1
71-252 Szczecin
tel. +48 91 441 7250, fax: +49 91 441 7251
email: lubinski@pum.edu.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J and Kozlowski P.
The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population.
Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400.

Oświadczam, iż brałem udział w wyselekcjonowaniu próbek DNA pochodzących z repozytorium kierowanego przeze mnie Międzynarodowego Centrum Nowotworów Dziedzicznych (MCND) w Szczecinie, udostępnionych do analizy delecji genu *APOBEC3B*. Brałem również udział w dyskusjach, których celem było ustalenie koncepcji badań oraz struktury manuskryptu.

Proszę o kontakt w przypadku dodatkowych pytań.

KIEROWNIK
Zakładu Genetyki i Patomorfologii
prof. dr hab. n. med. Jan Lubiński

Z poważaniem,

Jan Lubiński

POMORSKI UNIWERSYTET MEDYCZNY W SZCZECINIE
Katedra Onkologii
ZAKŁAD GENETYKI I PATOMORFOLOGII
70-115 Szczecin, ul. Połabska 4
tel. +48 91 466 1532, tel./fax +48 91 466 1533
Centrum Nowych Technologii Medycznych
71-252 Szczecin, ul. Unii Lubelskiej 1
tel. +48 91 441 7250, fax +48 91 441 7251

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 28.04.2017

dr Luiza Handschuh
Europejskie Centrum Bioinformatyki i Genomiki
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań
tel. 61 665 3052
email: luizahan@ibch.poznan.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:

Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P.

MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency

Oncotarget, 2016, 7(41):67266-67276. doi: 10.18632/oncotarget.11646.

Oświadczam, iż na potrzeby powyższej publikacji przygotowałam biblioteki NGS oraz przeprowadziłam analizę sekwencjonowania z wykorzystaniem platformy Genome Analyzer GAIIx (Illumina).

Z poważaniem,

dr Luiza Handschuh

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 5 sierpnia 2017

Dr Aleksandra Świercz
Europejskie Centrum Bioinformatyki i Genomiki
Instytut Chemii Bioorganicznej PAN
Noskowskiego 12/14
61-704 Poznań
E-mail: aleksandra.swiercz@cs.put.poznan.pl

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Dotyczy:
Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P.

MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency

Oncotarget, 2016, 7(41):67266-67276. doi: 10.18632/oncotarget.11646.

Oświadczam, iż moja rola w przygotowaniu powyższej publikacji polegała na przeprowadzeniu analizy bioinformatycznej (filtracja odczytów, mapowanie odczytów do sekwencji referencyjnych, ocena głębokości pokrycia) danych otrzymanych w analizie sekwencjonowania NGS.

Z poważaniem,

Aleksandra Świercz

Dr Aleksandra Świercz

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 28.04.2017

prof. dr hab. Marek Figlerowicz
Kierownik Zakładu Biologii Molekularnej i Systemowej
Dyrektor Instytutu Chemii Bioorganicznej PAN w Poznaniu

OŚWIADCZENIE

Dotyczy udziału w publikacji:

Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P.
MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency
Oncotarget, 2016, 7(41):67266-67276. doi: 10.18632/oncotarget.11646.

Jako współautor powyższej publikacji oświadczam, iż brałem udział w dyskusji nad koncepcją i strukturą pracy oraz redagowaniu tekstu manuskryptu.

Z poważaniem,

prof. dr hab. Marek Figlerowicz

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Poland
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61852 05 32, e-mail: ibch@ibch.poznan.pl

REGON 000849327

Poznań, 13.08.2017

mgr Katarzyna Klonowska
Zakład Genetyki Molekularnej
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu
email: kklonowska@man.poznan.pl

OŚWIADCZENIA - UDZIAŁ W PRZYGOTOWANIU PUBLIKACJI

1. **Klonowska K, Ratajska M, Wojciechowska M, Kozlowski P**

   Genetic predisposition to breast and/or ovarian cancer – focus on the candidate BARD1 gene

   BioTechnologia, 2014, 95(3):203-214

Mój udział w niniejszej pracy przeglądowej polegał na przygotowaniu manuskryptu oraz sporządzeniu tabeli i rycin. Na potrzeby publikacji dokonałam przeglądu literatury dotyczącej genetycznej predyspozycji do nowotworów piersi i/lub jajnika, ze szczególnym uwzględnieniem doniesień dotyczących związku genu *BARD1* z tymi nowotworami. Dokonałam wyboru najważniejszych zagadnień związanych z czynnikami ryzyka raka piersi i/lub jajnika, a także selekcji genów związanych z ryzykiem tych nowotworów oraz syndromów dziedzicznych predysponujących do nowotworów podlegających agregacji w rodzinie.

2. **Marcinkowska-Swojak M*, Klonowska K*, Figlerowicz M, Kozlowski P**

   An MLPA-based approach for high-resolution genotyping of disease-related multi-allelic CNVs

   Gene, 2014, 546(2):257-262

   ***Autorzy mieli taki sam wkład w przygotowanie publikacji**

W ramach projektu podsumowanego w niniejszej publikacji przeprowadziłam eksperymenty MLPA i analizy wyników dotyczące wielo-allelicznych polimorfizmów liczby kopii (CNV) związanych z powszechnymi chorobami człowieka. Eksperymenty wykonywałam we współpracy z dr inż. Małgorzatą Marcinkowską-Swojak, która zaprojektowała używane w badaniach zestawy sond MLPA. We współpracy z dr inż. Małgorzatą Marcinkowską-Swojak przygotowałam ryciny i materiały suplementarne do pracy oraz brałam udział w przygotowaniu manuskryptu.

3. **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, Irminger-Finger I, Limon J, 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

Na potrzeby powyższej publikacji zaprojektowałam test MLPA do analizy dużych mutacji w genie *BARD1*, przeprowadziłam jego walidację oraz wykorzystałam zoptymalizowany test MLPA do analiz, a także zinterpretowałam otrzymane wyniki. Dodatkowo, wykonałam analizę mutacji punktowych w genie *BARD1* oraz przygotowałam panel „sztucznych" kontrol pozytywnych przy użyciu enzymów restrykcyjnych oraz oligonukleotydów kompetycyjnych. Brałam również udział w przygotowaniu manuskryptu, w tym rycin oraz materiałów suplementarnych.

4. **Ratajska M, Matusiak M, Kuzniacka A, Wasag B, Brozek I, Biernat W, Koczkowska M, Debniak J, Sniadecki M, Kozlowski P, Klonowska K, Pilyugin M, Wydra D, Laurent G, Limon J, Irminger-Finger I**

   Cancer predisposing BARD1 mutations affect exon skipping and are associated with overexpression of specific BARD1 isoforms

   Oncology Reports, 2015, 34(5):2609-2617

Moja rola w przygotowaniu niniejszej publikacji polegała na charakterystyce mutacji c.1361C>T, c.1690C>T, c.1972C>T oraz c.1977A>G w genie *BARD1*. Na potrzeby publikacji przeprowadziłam analizę *in silico* z wykorzystaniem programów MutPred Splice, Human Splicing Finder, Rescue ESE, PolyPhen 2 oraz SIFT, które pozwoliły ocenić potencjalny wpływ analizowanych mutacji na alternatywny proces składania transkryptu *BARD1*, zachowawczość ewolucyjną zmienionego aminokwasu i/lub strukturę białka. Otrzymane wyniki analizy zostały podsumowane w "Table III", zawartej w publikacji.

5. **Klonowska K, Kluzniak W, Rusak B, Jakubowska A, Ratajska M, Krawczynska N, Vasilevska D, Czubak K, Wojciechowska M, Cybulski C, Lubinski J, Kozlowski P**

   The 30 kb deletion in the APOBEC3 cluster decreases APOBEC3A and APOBEC3B expression and creates a transcriptionally active hybrid gene but does not associate with breast cancer in the European population

   Oncotarget, 2017, [Epub ahead of print], doi: 10.18632/oncotarget.19400

Na potrzeby niniejszej publikacji zaprojektowałam i zoptymalizowałam test A3Bdel_MLPA oraz użyłam go do analizy struktury delecji *APOBEC3B*. Zaprojektowałam także test A3Bdel_PCR, który wykorzystałam w analizie genotypowania delecji w próbkach z grupy GDANSK. Zaprojektowałam również testy A3A_exp, A3B_exp oraz A3A/A3B_exp do analizy ekspresji genów objętych przez delecję oraz zastosowałam je w analizie związku genotyp – ekspresja z wykorzystaniem metod ddPCR oraz UPL-qPCR. Opracowałam rezultaty eksperymentów oraz przeprowadziłam niezbędne analizy statystyczne. Przygotowałam również meta-analizę podsumowującą wszystkie dotychczasowe badania asocjacji delecji *APOBEC3B* z nowotworami, włączając w nią naszą analizę. Brałam także udział w przygotowaniu manuskryptu wraz z rycinami, tabelami i materiałami suplementarnymi.

6. **Klonowska K, Handschuh L, Swiercz A, Figlerowicz M, Kozlowski P**
   MTTE: an innovative strategy for the evaluation of targeted/exome enrichment efficiency
   Oncotarget, 2016, 7(41):67266-67276

W ramach projektu podsumowanego w niniejszej publikacji wykonałam i zinterpretowałam wszystkie analizy eksperymentalne poza eksperymentami z użyciem sekwencjonowania nowej generacji (NGS). Zaprojektowałam zestaw sond MLPA wchodzących w skład testu MTTE, przeprowadziłam jego optymalizację i wykorzystałam zoptymalizowany test MTTE do analiz. W celu przeprowadzenia niezależnej walidacji otrzymanych wyników, zaprojektowałam i zoptymalizowałam zestaw starterów, który wykorzystałam do analiz z użyciem metody qPCR z użyciem sond UPL. Brałam również udział w opracowaniu strategii MTTE, w tym sposobu obliczania parametrów określających poziom wzbogacenia regionów kodujących genomu w bibliotekach przygotowanych do analizy NGS. Brałam także udział w przygotowaniu manuskryptu, jak również sporządzeniu tabel, rycin i materiałów suplementarnych.

Bardziej szczegółowy opis powyższych prac przedstawiony jest w rozdziale „OPIS WYNIKÓW PRACY DOKTORSKIEJ" niniejszej rozprawy.

K. Klonowska

Katarzyna Klonowska