



**WYDZIAŁ BIOLOGII
I OCHRONY ŚRODOWISKA**
Uniwersytet Łódzki

Marcin Słomka

**Zmienność genów kodujących białka
z nadrodziny ABC (ATP-binding cassette)
w populacji polskiej w odniesieniu do
innych populacji**

Variation of genes encoding proteins from the ABC
superfamily (ATP-binding cassette) in the Polish
population in comparison to other populations

Praca doktorska

wykonana w Pracowni Biobank w
Katedrze Biologii Nowotworów i
Epigenetyki

Promotor:

- Prof. dr hab. Grzegorz Bartosz

Promotor pomocniczy:

- Dr Dominik Strapagiel, prof. UŁ

Pragnę serdecznie podziękować Promotorowi rozprawy
Panu prof. dr hab. Grzegorzowi Bartoszowi
za zaangażowanie, życzliwość, cierpliwość, udzielone wskazówki
i nieustającą gotowość do udzielenia koniecznej pomocy.
Był to dla mnie ogromny zaszczyt i przyjemność.

Bardzo dziękuję również Kierownikowi Pracowni Biobank
Dr Dominikowi Strapagielowi, prof. UŁ
za stworzenie szansy i możliwości rozwoju,
powierzone zaufanie i wszelką udzieloną pomoc, której nie da się opisać.

Chciałbym szczególnie podziękować koledze
dr Michałowi Sewerynowi,
bez którego wiedzy, pomysłowości i wsparcia analitycznego
ta praca by w ogóle nawet nie powstała.

Składam serdecznie podziękowania
Kolegium Dziekańskiemu Wydziału Biologii i Ochrony Środowiska Uniwersytetu Łódzkiego
za stworzenie możliwości realizacji tej pracy i mojego rozwoju naukowego.

Dziękuję wszystkim tym, którzy w jakikolwiek sposób przez te wszystkie lata
przyczynili się do powstania tej pracy:
Kolegom i Koleżankom z projektu Testoplek, z Pracowni Biobank,
Zespołowi z Instytutu Biologii Medycznej PAN w Łodzi,
Bliskim i każdemu kto mnie wsparł w tym dziele, a nie sposób Go tutaj wymienić.

„Dać świadectwo Prawdzie” (J 18, 37)
Szczególnie zaś dziękuję Bogu w Trójcy Jedynemu, bo wszystko jest jedynie Jego darem i łaską,
to dzięki Niemu i dla Niego powstała ta praca
oraz dziękuję za niezawodne wstawiennictwo Najświętszej Maryi Pannie Królowej Polski.

SPIS TREŚCI:

ŹRÓDŁA FINANSOWANIA BADAŃ PROWADZONYCH W RAMACH ROZPRAWY DOKTORSKIEJ	3
WSPÓŁPRACA	3
WYKAZ ELEMENTÓW WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ	4
STRESZCZENIE	5
ABSTRACT	7
1. WSTĘP	9
1.1. Ogólna charakterystyka nadrodziny białek ABC.....	9
1.2. Znaczenie białek ABC i szkodliwy wpływ niektórych wariantów genetycznych na wybranych przykładach	10
1.3. Udział białek ABC w farmakoterapii i związana z tym rola zmienności genetycznej	13
1.4. Współczesne badania populacyjne dotyczące zmienności genetycznej.....	15
2. CEL PRACY I HIPOTEZY BADAWCZE	20
3. MATERIAŁY I METODY	21
3.1. Syntetyczny opis materiałów i metod przedstawionych w cyklu publikacji wchodzących w skład rozprawy doktorskiej.....	21
3.1.1. Próbkki genomowego DNA	21
3.1.2. Metoda skanowania genów <i>ABCC1</i> i <i>ABCG2</i> z wykorzystaniem HRM	21
3.1.3. Metoda genotypowania wariantów w <i>ABCC1</i> z wykorzystaniem HRM.....	22
3.1.4. Weryfikacja wyników HRM z wykorzystaniem sekwencjonowania metodą Sangera.....	22
3.1.5. Analiza zmienności genetycznej <i>ABCC1</i> i <i>ABCG2</i>	23
3.1.6. Analiza nierównowagi sprzężeń i bloków haplotypowych <i>ABCC1</i> i <i>ABCG2</i>	23
3.1.7. Analiza potencjalnego efektu wykrytych wariantów missensownych na funkcjonalność białka	24
3.2. Opis materiałów i metod dla badań nieopublikowanych, wchodzących w skład rozprawy doktorskiej.....	25
3.2.1. Próbkki genomowego DNA	25
3.2.2. Genotypowanie z wykorzystaniem mikromacierzy.....	25
3.2.3. Dane dla populacji referencyjnych	26
3.2.4. Imputacja genotypu	27
3.2.5. Analiza skalowania wielowymiarowego.....	27
3.2.6. Analiza IBS i redukcji wymiaru.....	28
3.2.7. Analiza ilości klastrów	29
3.2.8. Identyfikacja wariantów z różną częstością alleli pomiędzy klastrami	29

3.2.9. Analiza nierównowagi sprzężeń	29
3.2.10. Analiza potencjalnego efektu wykrytych wariantów missensownych na funkcjonalność białka	29
3.2.11. Analiza asocjacyjna całego genomu	29
3.2.12. Analiza asocjacyjna całego transkryptomu	30
3.2.13. Analiza fenotypowa na podstawie zróżnicowania populacji polskiej	30
4. WYNIKI	31
4.1. Syntetyczne omówienie wyników przedstawionych w publikacjach wchodzących w skład rozprawy doktorskiej.....	31
4.1.1. Załącznik nr 1.....	31
4.1.2. Załącznik nr 2.....	35
4.2. Omówienie wyników badań nieopublikowanych wchodzących w skład rozprawy doktorskiej	39
4.2.1. Zróżnicowanie populacji z bazy <i>1000 Genomes Projects</i> oraz populacji polskiej na podstawie zmienności genów ABC.....	39
4.2.2. Zróżnicowanie populacji europejskich oraz populacji polskiej pod względem zmienności genów ABC	40
4.2.3. Analiza wariantów związanych z klastrowaniem	43
4.2.4. Analiza grupowania próbek dla elementów ontologii, z którymi związane są geny ABC ...	58
4.2.5. Zróżnicowanie genów kodujących nadrodzinę białek ABC w populacji polskiej.....	82
4.2.6. Identyfikacja zmienności w genomie związanej z otrzymanym klastrowaniem	85
4.2.7. Różnice w ekspresji genów w narządach lub ich strukturach w poszczególnych klastrach	86
4.2.8. Analiza związku klastrowania populacji polskiej z fenotypem	87
5. DYSKUSJA	88
6. WNIOSKI I WERYFIKACJA HIPOTEZ BADAWCZYCH.....	96
7. BIBLIOGRAFIA	97
POZOSTAŁY DOROBIEK NIEWCHODZĄCY W SKŁAD ROZPRAWY DOKTORSKIEJ.....	104
ZAŁĄCZNIK NR 1.....	110
SUPLEMENT DO ZAŁĄCZNIKA NR 1	127
ZAŁĄCZNIK NR 2.....	208
SUPLEMENT DO ZAŁĄCZNIKA NR 2	230
OŚWIADCZENIA WSPÓŁAUTORÓW O UDZIALE W PUBLIKACJACH.....	261

ŹRÓDŁA FINANSOWANIA BADAŃ PROWADZONYCH W RAMACH ROZPRAWY DOKTORSKIEJ

„TESTOPLEK – Rola transporterów oporności wielolekowej w farmakokinetyce i toksykologii – testy in vitro w praktyce farmaceutycznej i klinicznej”. Projekt POIG.01.01.02-10-005/08 współfinansowany przez Unię Europejską z Europejskiego Funduszu Rozwoju Regionalnego.



UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



„Cyfrowe udostępnianie zasobów biomolekularnych i opisowych Biobanku i Katedry Antropologii Uniwersytetu Łódzkiego – charakterystyka populacji zamieszkujących tereny dzisiejszej Polski na przestrzeni dziejów. Platforma informacyjna e-Człowiek.pl”. Projekt POPC.02.03.01-00-0012/17 współfinansowany przez Unię Europejską w ramach Programu Operacyjnego Polska Cyfrowa.



UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



Projekt „Utworzenie sieci biobanków w Polsce w obrębie Infrastruktury Badawczej Biobanków i Zasobów Biomolekularnych BBMRI-ERIC”, na podstawie Decyzji Ministra Nauki i Szkolnictwa Wyższego nr DIR/WK/2017/01.



Ministerstwo Nauki
i Szkolnictwa Wyższego

WSPÓŁPRACA

Pracownia Genetyki i Fizjologii Mycobacterium, Instytut Biologii Medycznej Polskiej Akademii Nauk
ul. Lodowa 106, 93-232 Łódź



WYKAZ ELEMENTÓW WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

- **Opublikowane artykuły oryginalne:**

Załącznik nr 1:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Bartosz, G., Dziadek, J. Strapagiel, D., 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

Załącznik nr 2:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G., Strapagiel, D., 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. Genes (Basel), 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

Łącznie 2 opublikowane artykuły o sumarycznym IF = 6,428 oraz sumą punktów MEiN = 125

- **Badania nieopublikowane w formie dysertacji**

STRESZCZENIE

Białka ABC, zawierające kasetę wiążącą ATP (ang. ATP-Binding Cassette), stanowią największą nadrodzinę transporterów błonowych. U ludzi kodowane są przez 48 genów, które zostały podzielone na siedem rodzin, od ABCA do ABCG (geny ABC). Białka te biorą udział w wielu podstawowych procesach komórkowych i regulują homeostazę związków, transportując je zarówno przez błony wewnątrzkomórkowe organelli, jak i przez błonę komórkową. Wśród nich jest wiele endogennych i egzogennych substratów, m.in. lipidy, w tym cholesterol i steroidy, hormony, aminokwasy i peptydy, cukry, nukleozydy, jony oraz szerokie spektrum ksenobiotyków i duża grupa skoniugowanych metabolitów. Niektóre mutacje w genach ABC prowadzą zatem do wielu poważnych chorób wrodzonych. Ze względu na swoją kluczową funkcję w detoksykacji organizmu, wiele transporterów ABC ma także szczególne znaczenie w procesie usuwania leków, a ich aktywność jest jedną z głównych przyczyn powstania niekorzystnego zjawiska oporności wielolekowej (MDR, ang. *Multi-Drug Resistance*), najczęściej obserwowanej w leczeniu nowotworów. Niektóre warianty w genach ABC istotnie więc wpływają na różnice w odpowiedzi na farmakoterapię. Co więcej, częstość występowania wielu z nich istotnie różni się pomiędzy populacjami o różnym pochodzeniu etnicznym. Ostatnie badania wykazały duże międzyosobnicze i międzypopulacyjne zróżnicowanie genetyczne całej nadrodziny białek ABC. Nie stwierdzono jednak czy jest ono na tyle istotne, aby wyodrębnić populacje z tego samego regionu geograficznego, jak polską spośród pozostałych europejskich. Do tej pory jednak, dla populacji polskiej wykonano badania tylko dla pojedynczych wariantów w genach ABC. Głównym celem pracy była zatem analiza zmienności tych genów w populacji polskiej i porównanie jej z populacjami, zarówno o odrębnym pochodzeniu etnicznym, jak i również tylko z tymi o pochodzeniu europejskim.

W publikacjach, wchodzących w zakres rozprawy, opracowano metodykę badania zmienności genów ABC z wykorzystaniem metody topnienia produktów PCR w wysokiej rozdzielczości (HRM). Wykazano, że jest ona użyteczna zarówno w podejściu celowanego genotypowania, jak i przeszukiwania sekwencji kodujących w poszukiwaniu wariantów (skanowanie). Przeskanowano z jej wykorzystaniem prawie całe sekwencje genów kodujące dwa transportery ABC związane ze zjawiskiem MDR (*ABCC1*, *ABCG2*) w 190 próbkach od ochotników z polskiej populacji. Dokonano analizy otrzymanej zmienności i porównano uzyskane wyniki z danymi dla innych populacji.

W badaniach nieopublikowanych, wchodzących w skład rozprawy, wykorzystano dane uzyskane podczas genotypowania kohorty ochotników z populacji polskiej z użyciem mikromacierzy Illumina. Otrzymaną zmienność genów ABC porównano z populacjami referencyjnymi z bazy *1000 Genomes Project* według analizy grupowania próbek na podstawie podobieństwa alleli.

Zdecydowaną większość wykrytych wariantów w genach ABC stanowiły warianty rzadkie, z częstością występowania rzadszego allelu w badanej populacji (MAF) $<0,05$, zarówno

w skanowanych sekwencjach kodujących genów *ABCC1* i *ABCG2*, jak i w przekroju ogólnej zmienności wszystkich genów ABC poddanych analizie. Nie było zatem podstaw do odrzucenia hipotezy badawczej nr 1, w myśl której geny ABC są wysoce zakonserwowane i większość obserwowanej zmienności w ich sekwencjach stanowi zmienność o charakterze rzadkim ($MAF < 0,05$). Wyniki te świadczą o silnym zakonserwowaniu ewolucyjnym genów ABC, zwłaszcza w obrębie sekwencji kodujących kluczowe domeny wiążące ATP.

Porównanie wyników wykrytej zmienności *ABCC1* i *ABCG2* wykazało istotne różnice w częstości występowania specyficznych wariantów, między populacją polską, a populacjami o innym pochodzeniu etnicznym jak wschodnioazjatyckimi i afrykańskimi. Nie zaobserwowano jednak istotnych różnic w stosunku do wyników dla innych populacji pochodzenia europejskiego. Odzwierciedlenie tych wyników uzyskano w analizie klastrowania, gdzie zaobserwowano wyraźne grupowanie poszczególnych populacji, w tym polskiej, odpowiadające regionowi geograficznemu, z którego etnicznie pochodzą, co w sumie nie dało podstaw do odrzucenia hipotezy badawczej nr 2 zakładającej, że zmienność genów ABC w populacji polskiej istotnie różni się od populacji o innym pochodzeniu etnicznym.

Analiza tylko populacji polskiej i europejskich, wykazała klastrowanie w ogóle nie związane z pochodzeniem populacyjnym. Populacja polska, pod względem zmienności genów ABC, nie różni się więc istotnie od innych populacji referencyjnych o pochodzeniu europejskim, przez co hipoteza badawcza nr 3 postulująca, że zmienność genów ABC w populacji polskiej istotnie różni się od innych populacji referencyjnych o pochodzeniu europejskim, została odrzucona.

Celem kolejnych analiz była zatem próba wyjaśnienia podłoża zaobserwowanego klastrowania populacji europejskich. Wykazano, że istotnie odpowiadają za nie warianty częste ($MAF \geq 0,05$), w większości w sekwencjach niekodujących kilku genów, związanych z homeostazą kwasów żółciowych, transportem cAMP oraz gospodarką glukozową. Zaproponowano na tej podstawie niezależny od populacji, utrwalony mechanizm współistnienia różnych fenotypów, związanych z gospodarką lipidowo-glukozową, który jest regulowany przez kwasy żółciowe, w czym istotny udział mogą mieć transportery z nadrodziny ABC. Prawdopodobnie, równoległym czynnikiem klastrującym jest także zmienność genu *ABCC11*, związanego z funkcjami wydzielniczymi gruczołów apokrynowych. W analizie GWAS dla populacji polskiej wykazano z kolei, że obserwowane klastrowanie jest unikatowe bo istotnie związane w głównej mierze ze zmiennością genów kodujących białka ABC.

Prace badawcze w niniejszej rozprawie są pierwszymi, które w tak kompleksowy sposób podjęły temat zmienności genów kodujących całą nadrodzinę białek ABC w polskiej populacji.

ABSTRACT

ATP-Binding Cassette (ABC) proteins are the largest superfamily of membrane transporters. In humans, they are encoded by 48 genes, which have been classified into seven families, namely ABCA to ABCG (ABC genes). These proteins are involved in many crucial functions and cellular homeostasis maintaining by transporting different compounds through the intracellular membranes of organelles and the cell membrane as well. Among them, there is a plethora of many endogenous and exogenous substrates, e.g. lipids, including cholesterol and steroids, hormones, amino acids and peptides, sugars, nucleosides, ions but also a wide spectrum of xenobiotics and a large group of conjugated metabolites. Therefore, specific mutations in these genes are associated with many serious congenital diseases. Due to their key function of detoxification, many ABC transporters are also associated with the process of drug efflux, so their activity is one of the main causes of the adverse phenomenon called multi-drug resistance (MDR), mostly observed during cancer treatment. Moreover, some genetic variants also significantly affect the differences in pharmacotherapy efficacy and their prevalence varies significantly between populations of different ethnic origin. Recent high-throughput studies have demonstrated large inter-individual and inter-population genetic variability across the entire ABC protein superfamily. However, the results have not revealed whether observed variation is so significantly different to separate populations from the same geographical region, as Polish population from the other ones of European descent. So far, only few variants in the genes encoding ABC proteins have been studied in the Polish population. The main goal of the dissertation was variation analysis of the genes encoding ABC proteins in the Polish population and its comparison with other populations of a different ethnicity, including populations of European origin.

In the original publications included in the dissertation, an efficient methodology of studying variation of ABC genes was developed, based with the method of PCR products high-resolution melting (HRM). Presented results have demonstrated that it can be useful in both strategies, targeted genotyping and scanning coding sequences for the search of genetic variants. The method was used to scan almost the entire genes coding sequences of two ABC transporters associated with the MDR phenomenon (*ABCC1*, *ABCG2*) in 190 samples from Polish volunteers. The obtained variability was analysed and the results were compared with corresponding data for other populations.

In the unpublished data included in the dissertation, results from the genotyping method with the Illumina microarrays of a Polish volunteer's cohort have been applied. The obtained variability of genes encoding ABC proteins was compared with reference populations from the *1000 Genomes Project* database according to analysis of clustering samples based on allele similarity.

The vast majority of detected variants in ABC genes have been observed as the rare ones, with the minor allele frequency (MAF) in population <0.05 . It concerns both approaches, scanning

with HRM of *ABCC1* and *ABCG2* coding sequences, and the study of the entire variability of ABC genes included in the analysis. These results have not revealed the basis to reject of the research hypothesis no. 1 according to which ABC genes are highly conserved and majority of the observed variants in their sequences is rare ($MAF < 0.05$). These results indicate high evolutionary conservation of ABC genes, especially within sequences encoding the crucial ATP-binding domains.

The results of the detected *ABCC1* and *ABCG2* variation demonstrated significant differences of specific variants frequencies between the Polish population and other populations of different ethnic origin like East Asian and African. However, no significant differences have been observed in comparison with other European populations. These results have been confirmed in the clustering analysis, in which a distinct grouping of populations from the same geographical region was obtained, which was also consistent for the Polish population. There was no basis to reject the research hypothesis no. 2 assuming that the variability of ABC genes in the Polish population is significantly different from the population of other ethnic origin.

The clustering result for the Polish and European populations has not correlated with the origin of samples donors. In conclusion, the Polish population has not differed significantly from the other reference populations of European origin in term of variability of ABC genes, which led to the rejection of the research hypothesis no. 3 postulating that the variability of ABC genes in the Polish population is significantly different from other European populations.

The main goal of the followed analyses was attempted to explanation of the observed European populations clustering. The pool of common variants ($MAF \geq 0.05$) associated with clustering was identified, mostly in non-coding sequences of several genes, related to bile acid homeostasis, cAMP transport and glucose metabolism. On this basis, a population origin-independent and conserved mechanism was proposed of various phenotypes coexistence related to lipid-glucose metabolism, which is regulated by bile acids. ABC superfamily transporters would play a significant role in this process. Probably, an additional clustering factor is also the *ABCC11* gene variation, associated with the secretory functions of apocrine glands. In turn, the GWAS analysis for the Polish population demonstrated that the observed clustering is unique because of significant association mainly with the variability of ABC genes.

The research summarized in this dissertation is the first to address the subject of variability of genes encoding the entire superfamily of ABC proteins in the Polish population in such a comprehensive manner.

1. WSTĘP

1.1. Ogólna charakterystyka nadrodziny białek ABC

Homeostaza komórkowa znacząco uzależniona jest od kontrolowanej wymiany substratów endogennych i egzogennych przez błony. Transport aktywny, który umożliwia ruch substancji wbrew gradientowi stężeń lub potencjałowi elektrochemicznemu, możliwy jest dzięki różnym białkom transbłonowym. Wśród nich, białka ABC – zawierające kasetę wiążącą ATP (ang. *ATP-Binding Cassette*) – stanowią największą nadrodzinę transporterów błonowych (Juan-Carlos i in., 2021). U ludzi białka te kodowane są przez 48 genów, które zostały podzielone na siedem rodzin od ABCA do ABCG, na podstawie homologii sekwencji, podobieństwa struktury oraz kolejności domen w białku; geny te określone zostały jako geny ABC (Tab. 1) (Juan-Carlos i in., 2021; Wang i in., 2021).

Charakterystycznym schematem architektury białek z tej nadrodziny jest czterodomenowa struktura składająca się z dwóch cytoplazmatycznych domen wiążących nukleotydy (NBD, ang. *Nucleotide-Binding Domains*), które wiążą i hydrolizują ATP oraz dwóch domen transbłonowych (TMD, ang. *Transmembrane Domains*), które rozpoznają i transportują substraty. Podczas gdy struktura i funkcja NBD są podobne we wszystkich rodzinach, TMD są wysoce heterogenne, co pozwala transporterom rozpoznawać różne substraty i wykorzystywać energię z hydrolizy ATP do przemieszczania cząsteczek przez błony, niezależnie od gradientu stężeń (Robey i in., 2018). Specyfika struktury zależy jednak od poszczególnych białek ABC i przykładowo, niektóre zwane są półtransporterami (ang. *half-transporters*), ponieważ składają się tylko z jednej domeny TMD i jednej NBD. Z kolei białka z rodzin ABCE i ABCF nie mają domen TMD, dlatego nie uznaje się ich za transportery, chociaż przypisuje się im ważną rolę regulatorową (Nobili i in., 2020). Domeny NBD zawierają konserwatywne motywy konieczne do związania i hydrolizy ATP. Należą do nich: motyw Walker A (lub też pętla P, ang. *P-loop motif*), motyw Walker B oraz typowy wyłącznie dla tej nadrodziny motyw sygnaturowy ABC (ang. *ABC signature motif* lub też motyw C, ang. *C-loop motif*), określany także jako motyw LSGGQ (ang. *motif LSGGQ*), który znajduje się tuż powyżej motywu Walker B i jest związany z przyłączaniem nukleotydu. Ponadto inne motywy w NBD zostały opisane jako ważne w tym procesie: pętla Q (ang. *Q-loop region*), pętla D (ang. *D-loop*) i motyw H (ang. *H motif*) (Choi i Yu, 2014; Juan-Carlos i in., 2021; Liu, 2019).

Białka ABC biorą udział w wielu podstawowych procesach komórkowych regulując poziom związków na różnych poziomach – wyrzucają je na zewnątrz błony plazmatycznej, jak i też utrzymują fizjologiczną homeostazę ich wewnątrzkomórkowych stężeń dzięki transportowi przez błony wewnątrzkomórkowych organelli, m.in. mitochondriów, lizosomów, retikulum endoplazmatycznego czy aparatu Golgiego (He i in., 2021; Wang i in., 2021). Transportowi podlega szereg endogennych substratów, ale też związków pochodzenia egzogenne, m.in. lipidy, w tym cholesterol i steroidy, hormony, aminokwasy i peptydy, cukry, nukleozydy, jony oraz szerokie spektrum ksenobiotyków

i duża grupa metabolitów, w tym wiele koniugatów siarczanowych, koniugatów glutationu i glukuronidów (Tab. 1) (Leslie i in., 2005; Liu, 2019; Nobili i in., 2020). Dlatego też ekspresja genów ABC zachodzi w wielu rodzajach tkanek, m.in. w wątrobie, jelicie cienkim i grubym, płucach, nerkach, sercu, mięśniach szkieletowych, gruczołach, komórkach jednojądrzastych krwi obwodowej, ale też i naturalnych barierach: łożyskowej, krew-mózg i krew-jądro. Wzór ekspresji dla poszczególnych transporterów jest jednak wysoce tkankowo specyficzny (Juan-Carlos i in., 2021; Nobili i in., 2020; Wang i in., 2021). Również położenie w błonie plazmatycznej komórki jest zależne od rodzaju transportera i może być szczytowe (apikalne) czyli od strony światła tkanki lub podstawno-boczne (bazolateralne) (Vasiliou i in., 2009).

1.2. Znaczenie białek ABC i szkodliwy wpływ niektórych wariantów genetycznych na wybranych przykładach

Transportery ABC pełnią kluczową rolę w prawidłowym funkcjonowaniu wielu procesów komórkowych i szlaków metabolicznych. Mutacje w genach je kodujących prowadzą zatem do poważnych chorób wrodzonych lub zaburzeń o podłożu genetycznym (Tab. 1). Duża ich część związana jest z utrzymaniem homeostazy szeroko rozumianych lipidów, w tym cholesterolu czy kwasów żółciowych. Cholesterol w celu rozkładu musi zostać przetransportowany z tkanek obwodowych przez krew do wątroby, która jest jedynym narządem zdolnym do jego przetworzenia, degradacji i ostatecznie wydzielenia do żółci. Cholesterol, jako wysoce nierozpuszczalny, musi być pakowany i transportowany przez lipoproteiny osocza w procesie określanym jako „odwrotny transport cholesterolu” (ang. *reverse cholesterol transport*). ABCA1 pośredniczy w transporcie cholesterolu i fosfolipidów do ubogich w lipidy apolipoprotein (apoA1 i apoE), które następnie tworzą lipoproteiny o dużej gęstości (HDL, ang. *High-Density Lipoproteins*). Warianty genetyczne w *ABCA1* powodujące utratę funkcji białka prowadzą do niezwykle rzadkiej choroby tangerskiej (<100 przypadków, wg bazy *MedlinePlus*), charakteryzującej się bardzo niskimi poziomami HDL i apoA1. U chorych estry cholesterolu gromadzą się w wysokich stężeniach w makrofagach i różnych tkankach obwodowych, co powoduje miażdżycową chorobę sercowo-naczyniową i może przyczyniać się do wczesnego wystąpienia choroby Alzheimera i dysfunkcji nerek (Quazi i Molday, 2011). Z chorobą Alzheimera były również łączone inne geny ABC o podobnym mechanizmie transportu cholesterolu jak *ABCA2* i *ABCA7* (Liu, 2019).

Mutacje w *ABCA4* prowadzą do choroby Stargardta (częstość 1:10 000, wg bazy *MedlinePlus*), związanej z wieloma stanami zwyrodnieniowymi siatkówki. Przyczyną jest nieprawidłowy transport w siatkówce witaminy A i przyspieszone tworzenie toksycznych ubocznych produktów degradacji retinalu w procesie zwanym „cyklem widzenia”. W efekcie w nabłonku barwnikowym siatkówki

pojawiają się fluorescencyjne granulki lipofuscyny, charakterystyczne dla zwyrodnienia plamki żółtej (Tarling i in., 2013).

Jaskrawym przykładem konsekwencji niewłaściwej gospodarki lipidowej związanej z białkami ABC, są mutacje w *ABCA12* powodujące utratę funkcji białka transportującego lipidy w ziarnistościach błonowych keratynocytów powodujące rybią łuskę arlekinową (ang. *Harlequin ichthyosis*) (częstość 1:200 000, wg bazy *MedlinePlus*). Nieprawidłowy rozwój ziarnistości błonowych keratynocytów prowadzi do poważnego upośledzenia bariery skórnej w formie łusek, co powoduje predyspozycje do sepsy, odwodnienia i nieprawidłowej regulacji termicznej, a w efekcie prowadzi do śmierci (Tarling i in., 2013).

Szczególne znaczenie mają geny kodujące transportery umożliwiające prawidłową gospodarkę kwasami żółciowymi. Powstają jako produkty katabolizmu wysoce nierozpuszczalnego cholesterolu. Odgrywają wiele ról w metabolizmie lipidów: tworzą micelle w jelicie cienkim, które ułatwiają rozpuszczanie, trawienie oraz wchłanianie lipidów i witamin rozpuszczalnych w tłuszczach, indukują przepływ żółci z hepatocytów do pęcherzyka żółciowego, a wydalenie tych miceli stanowi główną drogę usuwania cholesterolu (de Aguiar Vallim i in., 2013). *ABCB11* koduje główny transporter zwany BSEP (pompa eksportu soli żółciowych, ang. *Bile-Salt Export Pump*), który ulega ekspresji wyłącznie na wierzchołkowej błonie hepatocytów i odpowiada za transport koniugatów kwasów żółciowych z hepatocytu do kanalików żółciowych. Mutacje genu kodującego ten transporter i powodujące zaburzenia jego funkcji, są przyczyną postępującej rodzinnej cholestazy wewnątrzwątrobowej (PFIC-2, ang. *Progressive Familial Intrahepatic Cholestasis type 2*; częstość 1:100 000, wg bazy *MedlinePlus*) prowadzącej do gromadzenia się soli żółciowych w hepatocytach i w efekcie do uszkodzenia wątroby (Liu, 2019; Tarling i in., 2013). Z kolei w wyniku ekspresji w wątrobie i jelicie cienkim dwóch genów *ABCG5* i *ABCG8* powstaje heterodimer, którego głównymi funkcjami są wydzielanie cholesterolu i steroli roślinnych do żółci oraz ograniczenie wchłaniania steroli i wydalenie ich z enterocytów z powrotem do jelita krętego. Mutacje, które uszkadzają którykolwiek z półtransporterów, skutkują bardzo rzadką sitosterolemią (<100 przypadków, wg bazy *MedlinePlus*), która charakteryzuje się gromadzeniem steroli roślinnych i cholesterolu w wielu tkankach i krwi, co ostatecznie prowadzi do przedwczesnego rozwoju miażdżycy (Liu, 2019; Neumann i in., 2017; Tarling i in., 2013). Całości procesu tworzenia żółci dopełnia produkt genu *ABCB4*, transporter odpowiadający za wydzielanie fosfatydylocholiny z wątroby do żółci (Neumann i in., 2017).

Niektóre transportery ABC zapewniają inną, kluczową funkcję homeostatyczną – utrzymanie równowagi jonowej. Gen *ABCC7* koduje niezwykle specyficzny błonowy regulator przewodnictwa (CFTR, ang. *Cystic Fibrosis Transmembrane conductance Regulator*). Jest to bowiem jedyny kanał anionowy dla jonów chlorkowych w błonie komórkowej wielu komórek oraz jedyny kanał

bramkowany ligandem (ATP), który jest zużywany w tym procesie. CFTR zawiera również dodatkową domenę zwaną regulacyjną (R) po wewnętrznej stronie błony plazmatycznej. Jej fosforylacja sprzężona z hydrolizą ATP przez domeny NBD otwiera kanał na zewnątrz umożliwiając transport jonów zgodnie z ich gradientem elektrochemicznym (Hwang i Kirk, 2013). Licznie odnotowane mutacje w genie *ABCC7* prowadzą do mukowiscydozy, objawiającej się zaburzeniami wydzielania zbyt lepkiego śluzu przez gruczoły głównie płuc i trzustki, co prowadzi do ich uszkodzenia i niewydolności (Dean i in., 2022). To najczęstsza choroba dziedziczna związana z mutacjami w genach kodujących białka ABC i dotyczy zwłaszcza odmiany białej człowieka (częstość 1:2500, wg bazy *MedlinePlus*) (Dean i in., 2001)

Gen *ABCC8* koduje natomiast białko zwane receptorem sulfonilomocznika (SUR1, ang. *SulfonylUrea-Receptor*), które tworzy inny, wyjątkowy kanał potasowy (K_{ATP}) w komórkach beta w trzustce. Składa się on z tworzących centralny por czterech podjednostek Kir6.2 (prostownicze do wewnątrz podjednostki kanału potasowego) oraz otaczających je czterech regulacyjnych podjednostek SUR1, tworzących razem heterooktameryczny kompleks. Wewnątrzkomórkowy metabolizm glukozy prowadzi do zmian stężenia nukleotydów adeninowych, w wyniku czego ATP zamyka kanał, wiążąc się z Kir6.2. W rezultacie dochodzi do depolaryzacji błony, co z kolei aktywuje zależne od napięcia kanały wapniowe (Ca^{2+}) i prowadzi do wzrostu wewnątrzkomórkowego stężenia Ca^{2+} a ostatecznie wywołuje wydzielanie insuliny. Natomiast magnezowe kompleksy nukleotydów (Mg-ADP i Mg-ATP) stymulują aktywność kanału poprzez interakcję z SUR1, w wyniku której kanały ulegają otwarciu, a wydzielanie insuliny zostaje zahamowane. Mutacje w obrębie *ABCC8* mogą powodować zaburzenia wydzielania insuliny i różne formy wrodzonego hiperinsulinizmu (częstość najpowszechniejszej z nich 1:50 000, wg bazy *MedlinePlus*). Sulfonilomoczniki stymulują wydzielanie insuliny w cukrzycy typu 2, wiążąc się z SUR1 i zamykając kanały potasowe przez mechanizm niezależny od ATP. Aktywność kanału potasowego jest zatem krytyczna w utrzymaniu prawidłowej gospodarki glukozowo-insulinowej (Gloyn i in., 2004; Liu, 2019).

Geny *ABCB2* i *ABCB3* pełnią z kolei fundamentalną rolę w prawidłowym funkcjonowaniu swoistej odpowiedzi immunologicznej poprzez rozpoznawanie i eliminację komórek transformowanych wirusowo lub złośliwie. Kodują odpowiednio białka TAP1 i TAP2 (ang. *Transporter associated with Antigen Processing*), które razem tworzą heterodimer w błonie retikulum endoplazmatycznego. Antygen wiązany jest z łańcuchem ciężkim głównego układu zgodności tkankowej klasy I (MHC I, ang. *Major Histocompatibility Complex class I*) przez złożony białkowy kompleks ładujący peptyd (ang. *Peptide-Loading Complex*). Dzięki transportowi przez kompleks TAP, cząsteczka MHC I z przyłączonym peptydem przemieszcza się na powierzchnię komórki, gdzie jest prezentowana cytotoksycznym limfocytom T $CD8^{+}$. Nieprawidłowo funkcjonujące białka TAP

prowadzą do rozwoju chorób zakaźnych lub nowotworowych (<100 przypadków, wg bazy *MedlinePlus*) (Hulpke i Tampé, 2013).

1.3. Udział białek ABC w farmakoterapii i związana z tym rola zmienności genetycznej

Ze względu na swoją kluczową funkcję w detoksykacji organizmu i usuwaniu ksenobiotyków, transportery ABC mają szczególne znaczenie w procesie wyrzutu leków określanym jako efluks (ang. *drug efflux*) oraz ich metabolizmie, wpływając przez to na odpowiedź wobec zastosowanej terapii poprzez zmianę parametrów farmakokinetycznych: wchłaniania, dystrybucji, metabolizmu, wydalania oraz toksyczności leków (ADMET, ang. *Absorption, Distribution, Metabolism, Excretion, Toxicity*) (Robey i in., 2018). Z tego powodu uważane są za ważne czynniki prognostyczne kierunku rozwoju choroby i skuteczności farmakoterapii (Choi i Yu, 2014). Nadekspresja transporterów ABC jest więc jedną z głównych przyczyn powstania niekorzystnego zjawiska oporności wielolekowej (MDR, ang. *Multi-Drug Resistance*). Rozumiana jest ona jako jednoczesna oporność na kilka leków, które nie mają ze sobą związku strukturalnego ani funkcjonalnego (Juan-Carlos i in., 2021). MDR może być efektem kilku mechanizmów, o charakterze wrodzonym lub nabytym w wyniku chemioterapii, takich jak: zmniejszone wchłanianie leków, zmiany w punktach kontrolnych cyklu komórkowego lub jego zatrzymanie, modyfikacje celów leków czy wychwytywanie leków przez lizosomy albo inne organelle wewnątrzkomórkowe i pęcherzyki międzykomórkowe (Choi i Yu, 2014; He i in., 2021; Li i in., 2016). Transportery ABC prowadzą natomiast do powstania MDR poprzez zwiększony efluks leków, co najczęściej jest obserwowane w komórkach nowotworowych (Juan-Carlos i in., 2021). Oporność wielolekowa jest więc istotną przeszkodą w skutecznej chemioterapii, a transportery ABC stały się głównym celem w walce z nią, zwłaszcza u pacjentów onkologicznych (He i in., 2021). Substratami transporterów ABC jest bowiem szerokie spektrum powszechnie stosowanych leków przeciwnowotworowych czyli cytostatyków o różnym mechanizmie działania: inhibitorów mitozy, jak alkaloidy barwinka różowego czy taksany, inhibitorów topoizomerazy (np. topotekan, irynotekan), antymetabolitów (np. metotreksat), antybiotyków cytostatycznych (np. mitoksantron), (He i in., 2021). Najczęściej wiązane ze zjawiskiem MDR były głównie geny *ABCB1* – kodujący transporter określany jako MDR1 (białko oporności wielolekowej 1, ang. *Multidrug Resistance Protein 1*) lub glikoproteina P (P-gp, ang. *P-glycoprotein*), *ABCC1* – kodujący transporter nazywany MRP1 (białko 1 związane z opornością wielolekową, ang. *Multidrug Resistance-associated Protein 1*) oraz *ABCG2* – kodujący transporter nazywany BCRP (białko oporności raka piersi, ang. *Breast Cancer Resistance Protein*) (Tab. 1) (Choi i Yu, 2014).

W wielu badaniach wykazano, że niektóre z germinalnych (dziedzicznych) wariantów w genach kodujących transportery ABC istotnie wpływają na różnice w odpowiedzi na farmakoterapię raka, jej toksyczność, nawrót choroby lub jej progresję czy przeżywalność (Xiao i in.,

2020b). Różne mechanizmy molekularne mogą wyjaśniać związek między wariantami genetycznymi transporterów ABC a obserwowanym fenotypem, takie jak: zmniejszona dostępność aminoacylo-tRNA, obniżona stabilność mRNA, wpływ na konformację białka, aktywność transportera lub specyficzność substratowa (Arrigoni i in., 2016). Szacuje się, że indywidualna zmienność genetyczna wyjaśnia 20–30% różnic międzypersonalnych w odpowiedzi na lek (Zhou i Lauschke, 2022). Założenia farmakoterapii spersonalizowanej opierają się zatem na odejściu od jednego schematu leczenia, a w zamian na doborze najbardziej odpowiedniego leku i dawki w oparciu o profil genetyczny pacjenta (Arrigoni i in., 2016). Dzięki temu możliwe byłoby nie tylko zwiększenie skuteczności terapeutycznej, ale także obniżenie kosztów leczenia (Li i in., 2016). Dlatego zrozumienie roli zmienności genetycznej związanej z odpowiedzią na leczenie może zapewnić w przyszłości bezpieczniejszą i efektywniejszą farmakoterapię (Juan-Carlos i in., 2021).

Obserwowana międzypersonalna zmienność genetyczna, a co za tym idzie też, ewentualny związek z farmakoterapią, jest w znacznej mierze spowodowany różnym rozpowszechnieniem wariantów genetycznych w poszczególnych populacjach (Zhou i Lauschke, 2022). W genach interesujących z punktu widzenia farmakogenetyki, w tym kodujących transportery ABC, stwierdzano istotne różnice w częstości występowania niektórych alleli w wariantach określanych jako częste – czyli o częstości występowania rzadszego allelu w badanej populacji lub puli próbek $\geq 0,05$ (MAF, ang. *Minor Allele Frequency*) (Ono i in., 2013; Zhou i Lauschke, 2022). Różnice obserwowane były pomiędzy populacjami z różnych regionów geograficznych i dotyczyły zwłaszcza tych o innym pochodzeniu etnicznym (europejskim, wschodnioazjatyckim lub afrykańskim) (Cropp i in., 2008). Badania te, w dużej mierze przeprowadzone na pojedynczych grupach etnicznych, sugerowały także, że mogą istnieć istotne, etniczne różnice w skuteczności lub toksyczności leków zależne od genotypu (Cropp i in., 2008; Xiao i in., 2020b). Profilowanie genów związanych z farmakogenetyką, jak kodujących transportery ABC, w różnych populacjach może zatem ujawnić zmienność genetyczną, która predysponuje do różnic w odpowiedzi na lek (Zhou i Lauschke, 2022). Badania genów ABC zaczęły się więc skupiać na przeszukiwaniu ich sekwencji eksonowych, w celu określenia funkcjonalnej zmienności genetycznej w populacjach o różnym pochodzeniu etnicznym (Bosch i in., 2005; Fukushima-Uesaka i in., 2007; Lee i in., 2007; Leschziner i in., 2006; Maekawa i in., 2006; Saito i in., 2002; Wang i in., 2004). W populacji polskiej takim badaniom w genach kodujących białka ABC poddano jednak tylko pojedyncze warianty (Dudarewicz i in., 2012; Jaźwińska-Tarnawska i in., 2015; Koszarska i in., 2014; Niebudek i in., 2019; Salagacka-Kubiak i in., 2016). W trakcie projektu „TESTOPLEK – Rola transporterów oporności wielolekowej w farmakokinetyce i toksykologii – testy *in vitro* w praktyce farmaceutycznej i klinicznej”, realizowanego w Katedrze Biofizyki Molekularnej Uniwersytetu Łódzkiego (obecnie Katedra Biologii Nowotworów i Epigenetyki), zbadano, że w populacji polskiej warianty w genach ABC, a związanych z MDR są reprezentowane (Majewska i in.,

2012; Majewska i in., 2013). Dlatego przeszukanie ich sekwencji kodujących w celu zbadania zmienności genetycznej w populacji polskiej, należy uznać za jak najbardziej uzasadnione.

1.4. Współczesne badania populacyjne dotyczące zmienności genetycznej

Ogromny postęp technologiczny w sekwencjonowaniu i genotypowaniu zapoczątkował nową erę w badaniach genetyki człowieka i farmakogenetyki (Yee i in., 2018). Wysokoprzepustowe technologie takie jak genotypowanie z wykorzystaniem mikromacierzy czy sekwencjonowanie następnej generacji (NGS, ang. *Next-Generation Sequencing*) umożliwiają szybkie i masowe uzyskanie danych dla całych ludzkich genomów, eksomów czy setek tysięcy specyficznych wariantów genetycznych równolegle (Gulilat i in., 2019). W rezultacie, zarówno częste jak i rzadkie warianty wielu genów, w tym genów transporterów błonowych, mogą zostać zidentyfikowane jako potencjalnie związane z parametrami ADMET lub pomóc w szacowaniu ryzyka odpowiednich chorób w badaniach epidemiologicznych. W ostatnich latach coraz częściej wykorzystywano wspomniane technologie do badania wariantów w farmakogenach w dużych i zróżnicowanych populacjach etnicznych, z różnych regionów świata (Xiao i in., 2020a; Yee i in., 2018). Wraz ze wzrostem dostępności informacji o genotypie, różnice genetyczne między grupami etnicznymi są ujawniane z coraz większą rozdzielczością. W badaniach stwierdzano silne różnice w dystrybucji wariantów i funkcjonalności genów między głównymi grupami etnicznymi, a co więcej, dane te sugerują, że wyraźne różnice genetyczne mogą istnieć nawet w stosunkowo małych regionach geograficznych (Zhou i Lauschke, 2022). W ostatnim dużym badaniu charakteryzującym geny kodujące nadrodzinę białek ABC, w kilku dużych grupach etnogeograficznych, wykazano m.in. duże międzyosobnicze jak i międzypopulacyjne zróżnicowanie pod tym względem (Xiao i in., 2020a). Nie zostało jednak wykazane czy jest ono na tyle istotne, że pozwala zróżnicować populacje o podobnym pochodzeniu etnicznym, jak populację polską od innych europejskich. Wskazywałoby to wówczas na indywidualny profil genetyczny nadrodziny białek ABC w każdej populacji, także polskiej, co mogłoby wpłynąć na schematy we wprowadzaniu medycyny spersonalizowanej w różnych populacjach. Prace badawcze w niniejszej rozprawie są pierwszymi, które w tak kompleksowy sposób podejmują temat zmienności genów kodujących całą nadrodzinę białek ABC w polskiej populacji.

Tabela 1. Geny kodujące nadrodzinę białek ABC u człowieka.

Gen	Lokalizacja w chromosomie ^a	Położenie w sekwencji ^b	Aliasy ^c	Funkcje białka ^{d, e}	Związek wariantów genetycznych z chorobami lub opornością wielolekową ^d
<i>ABCA1</i>	9q31.1	104781006-104928155	ABC-1, ABC1, CERP, HDLDT1, TGD	transport fosfolipidów i cholesterolu; tworzenie HDL	choroba tangierska
<i>ABCA2</i>	9q34.3	137007234-137028915	ABC2	transporter lipidów i homeostaza cholesterolu; regulacja wytwarzania amyloidu-beta	zaburzenia rozwoju intelektualnego ze słabym wzrostem z lub bez drgawek lub ataksją
<i>ABCA3</i>	16p13.3	2275881-2340746	ABC-C, ABC3, EST111653, LBM180, SMDP3	transport fosfolipidów; usuwanie cholesterolu; homeostaza surfaktanta płucnego	śródmięszowa choroba płuc
<i>ABCA4</i>	1p22.1	93992834-94121148	ABC10, ABCR, ARMD2, CORD3, FFM, RMP, RP19, STGD, STGD1	transport pochodnych retinylowych	choroba Stargarda; barwnikowe zwyrodnienie siatkówki; dystrofia czopkowo-pręcikowa
<i>ABCA5</i>	17q24.3	69244311-69327244	ABC13, EST90625	transport cholesterolu; tworzenie HDL	włóknistość dziąseł z nadmiernym owłosieniem
<i>ABCA6</i>	17q24.2-q24.3	69078702-69141895	EST155051	transport lipidów i homeostaza cholesterolu w makrofagach	
<i>ABCA7</i>	19p13.3	1039997-1065572	ABCA-SSN, ABCX, AD9	transport fosfolipidów i cholesterolu; organizacja tratw lipidowych; fagocytoza makrofagów; regulacja wytwarzania amyloidu-beta	choroba Alzheimer
<i>ABCA8</i>	17q24.2	68867289-68955392		transport cholesterolu; produkcja sfingomieliny	
<i>ABCA9</i>	17q24.2	68974488-69060949	EST640918	transport lipidów oraz homeostaza cholesterolu; różnicowanie monocytów	
<i>ABCA10</i>	17q24.3	69147214-69244846	EST698739	transport lipidów oraz homeostaza cholesterolu, różnicowanie monocytów	
<i>ABCA12</i>	2q35	214931542-215138626	ARCI4A, ARCI4B, ICR2B, LI2	transport lipidów i homeostaza cholesterolu	rybia łuska arlekinowa
<i>ABCA13</i>	7p12.3	48171458-48647497		transport cholesterolu	
<i>ABCB1</i>	7q21.12	87503017-87713323	ABC20, CD243, CLCS, GP170, MDR1, P-GP, PGY1	transport ksenobiotyków; transport lipidów	nieswoiste zapalenie jelit; oporność wielolekowa

Gen	Lokalizacja w chromosomie ^a	Położenie w sekwencji ^b	Alias ^c	Funkcje białka ^{d, e}	Związek wariantów genetycznych z chorobami lub opornością wielolekową ^d
<i>TAP1 (ABCB2)</i>	6p21.32	32845209-32853816	ABC17, APT1, D6S114E, PSF-1, PSF1, RING4, TAP1N*0102N	transport antygenów peptydowych MHC I	zespół nagich limfocytów
<i>TAP2 (ABCB3)</i>	6p21.32	32821833-32838739	ABC18, APT2, D6S217E, PSF-2, PSF2, RING11	transport antygenów peptydowych MHC I	zespół nagich limfocytów
<i>ABCB4</i>	7q21.12	87401696-87480435	ABC21, GBD1, ICP3, MDR2, MDR2/3, MDR3, PFIC-3, PGY3	transport fosfolipidów do żółci	postępująca rodzinna wewnątrzwątrobową cholestaza
<i>ABCB5</i>	7p21.1	20615667-20777038	ABCB5alpha, EST422562	rozwój i naprawa rogówki	
<i>ABCB6</i>	2q35	219209772-219218994	ABC, LAN, MTABC3, PRP, umat	transport porfiryn; transport jonów metali; fenotyp grupy krwi Lan	dyschromatoza uniwersalna; mikroftalmia izolowana, z colobomą; rodzinna pseudohiperkaliemia
<i>ABCB7</i>	Xq13.3	75051048-75156732	ABC7, ASAT, Atm1p, EST140535	transport i homeostaza żelaza	niedokrwistość syderoblastyczna z ataksją
<i>ABCB8</i>	7q36.1	151028422-151047782	EST328128, M-ABC1, MABC1	kanał potasowy w mitochondrium; transport żelaza	
<i>ABCB9</i>	12q24.31	122920951-122981649	EST122234, TAPL	transport peptydów do lizosomów	
<i>ABCB10</i>	1q42.13	229516582-229558707	EST20237, M-ABC2, MTABC2	transport mitochondrialny; biosynteza hemu i hemoglobiny	
<i>ABCB11</i>	2q31.1	168915498-169031324	ABC16, BRIC2, BSEP, PFIC-2, PFIC2, PGY4, SPGP	transport soli żółciowych do żółci; homeostaza lipidów, cholesterolu, kwasów żółciowych	postępująca rodzinna cholestaza wewnątrzwątrobową
<i>ABCC1</i>	16p13.11	15949138-16143257	ABC29, ABCC, GS - X, MRP, MRP1	transport anionów organicznych i ksenobiotyków; transport koniugatów glutationu i glukuronidów, leukotrienu C4	głuchota; oporność wielolekowa
<i>ABCC2</i>	10q24.2	99782640-99852594	ABC30, CMOAT, DJS, MRP2, cMRP	transport ksenobiotyków; transport koniugatów siarczanu, glukuronidów i glutationu; transport bilirubiny; transport soli żółciowych	zespół Dubina-Johnsona; oporność wielolekowa
<i>ABCC3</i>	17q21.33	50634777-50692253	ABC31, EST90757, MLP2, MOAT - D, MRP3, cMOAT2	transport ksenobiotyków; transport koniugatów glukuronianu i glutationu, leukotrienu C4; transport soli żółciowych	oporność wielolekowa

Gen	Lokalizacja w chromosomie ^a	Położenie w sekwencji ^b	Alias ^c	Funkcje białka ^{d, e}	Związek wariantów genetycznych z chorobami lub opornością wielolekową ^d
<i>ABCC4</i>	13q32.1	95019835-95301475	MOAT-B, MOATB, MRP4	transport cyklicznych nukleotydów (cAMP, cGMP), kwasów żółciowych, koniugatów steroidowych, leukotrienów, moczanu i prostaglandyn	oporność wielolekowa
<i>ABCC5</i>	3q27.1	183919934-184017939	ABC33, EST277145, MOAT-C, MOATC, MRP5, SMRP, pABC11	transport ksenobiotyków; transport cyklicznych nukleotydów (cAMP, cGMP), kwasu foliowego, hemu	oporność wielolekowa
<i>ABCC6</i>	16p13.11	16149565-16223522	ABC34, ARA, EST349056, GACI2, MLP1, MOAT-E, MOATE, MRP6, PXE, PXE1, URG7	transport ksenobiotyków; transport koniugatów glutationu, leukotrienu C4; uwalnianie ATP do krwiobiegu; hamuje apoptozę, w której pośredniczy TNF-alfa	kępki żółte rzekome (<i>Pseudoxanthoma elasticum</i>); zwapnienie tętnic w okresie niemowlęcym; oporność wielolekowa
<i>CFTR (ABCC7)</i>	7q31.2	117287120-117715971	ABC35, CF/MRP, MRP7, TNR-CFTR, dJ760C5.1	transport jonów chlorkowych; homeostaza płynów w drogach oddechowych	mukowiscydoza; wrodzony obustronny brak nasieniowodów
<i>ABCC8</i>	11p15.1	17392498-17476894	ABC36, HHF1, HI, HRINS, MRP8, PHHI, SUR, SUR1, SUR1delta2, TNDM2	regulator transportu jonów potasowych w trzustce; uwalnianie insuliny	rodzinna hipoglikemia hiperinsulinemiczna; permanentna cukrzyca noworodkowa; przejściowa cukrzyca noworodków; hipoglikemia wywołana leucyną
<i>ABCC9</i>	12p12.1	21797389-21942543	ABC37, ATFB12, CANTU, CMD10, SUR2	kanał potasowy w mięśniu sercowym, mięśniach szkieletowych, naczyniowych i gładkich	kardiomiopatia rozstrzeniowa; rodzinne migotanie przedsionków; osteochondrodysplazja hipertrychotyczna; niepełnosprawność intelektualna i zespół miopatii
<i>ABCC10</i>	6p21.1	43427366-43450427	EST182763, MRP7, SIMRP7	transport ksenobiotyków; transport glukuronidów i koniugatów glutationu, w tym leukotrienu C4	oporność wielolekowa
<i>ABCC11</i>	16q12.1	48165773-48247568	EWWD, MRP8, WW	transport ksenobiotyków; transport cyklicznych nukleotydów, S-koniugatów glutationu, leukotrienu C4, siarczanów steroidów, glukuronidów, koniugatów kwasów żółciowych	oporność wielolekowa
<i>ABCC12</i>	16q12.1	48080882-48156018	MRP9	funkcja nieokreślona	
<i>ABCD1</i>	Xq28	153724856-153744755	ABC42, ALD, ALDP, AMN	transport długołańcuchowych kwasów tłuszczowych do peroksysomu; udział w procesie mielinizacji; odpowiedzi komórkowej na stres oksydacyjny i odpowiedzi zapalnej	adrenoleukodystrofia

Gen	Lokalizacja w chromosomie ^a	Położenie w sekwencji ^b	Aliasy ^c	Funkcje białka ^{d,e}	Związek wariantów genetycznych z chorobami lub opornością wielolekową ^d
<i>ABCD2</i>	12q12	39550033-39619803	ABC39, ALDL1, ALDR, ALDRP, hALDR	transport długołańcuchowych kwasów tłuszczowych do peroksysomu	
<i>ABCD3</i>	1p21.3	94418389-94518666	ABC43, CBAS5, PMP70, PXMP1, ZWS2	transport długołańcuchowych kwasów tłuszczowych i żółciowych do peroksysomu	wrodzony defekt syntezy kwasów żółciowych
<i>ABCD4</i>	14q24.3	74285269-74303055	ABC41, EST352188, MAHCl, P70R, P79R, PMP69, PXMP1L	transport kobalaminy z lizosomów do cytozolu	kwasyca metylomalonowa i homocystynuria
<i>ABCE1</i>	4q31.21	145098288-145129524	ABC38, OABP, RLI, RLI1, RNASEL1, RNASELI, RNS4I	kontrola translacji - uwalnianie niefunkcjonalnych rybosomów i degradacja uszkodzonych mRNA w szlaku "No-Go Decay"	
<i>ABCF1</i>	6p21.33	30571393-30597179	ABC27, ABC50	regulacja wydajnej inicjacji translacji mRNA	
<i>ABCF2</i>	7q36.1	151211484-151227205	ABC28, EST133090, HUSSY-18, HUSSY18	funkcja nieokreślona	
<i>ABCF3</i>	3q27.1	184186095-184194012	EST201864	udział w obronie przeciwwirusowej; przyłączanie kadheryny	
<i>ABCG1</i>	21q22.3	42199689-42297244	ABC8, WHITE1	transport fosfolipidów i cholesterolu	
<i>ABCG2</i>	4q22.1	88090150-88231628	ABC15, ABCP, BCRP, BCRP1, BMDP, CD338, CDw338, EST157481, GOUT1, MRX, MXR, MXR-1, MXR1, UAQTL1	transport ksenobiotyków; transport steroidów, koniugatów siarczanowych, moczanu; homeostaza porfiryn i hemu; fenotyp grupy krwi Jr	wpływ na stężenie kwasu moczowego; oporność wielolekowa
<i>ABCG4</i>	11q23.3	119149052-119162653	WHITE2	transport steroli, cholesterolu; tworzenie HDL; związany z transportem amyloidu-beta; udział w indukcji apoptozy	
<i>ABCG5</i>	2p21	43812472-43838865	STSL	transport steroli roślinnych i cholesterolu; usuwanie ich do żółci	sitosterolemia
<i>ABCG8</i>	2p21	43831942-43882988	GBD4	transport steroli roślinnych i cholesterolu; usuwanie ich do żółci	sitosterolemia

^a – według bazy HUGO Gene Nomenclature Committee (HGNC); ^b – według sekwencji referencyjnej GRCh38/hg38; ^c – według bazy NCBI Gene; ^d – według bazy UniProtKB/Swiss-Prot; ^e – według bazy Gene Ontology.

2. CEL PRACY I HIPOTEZY BADAWCZE

Głównym celem pracy była charakterystyka zmienności genów kodujących białka ABC w populacji polskiej i porównanie otrzymanych wyników z dostępnymi danymi dla innych populacji, zarówno o odrębnym pochodzeniu etnicznym, jak i również dla innych populacji o pochodzeniu europejskim. Prace realizowane w ramach obecnej rozprawy wpisują się w obecny nurt badań obejmujący genetyczne podstawy zróżnicowania międzypopulacyjnego w obrębie genów kodujących transportery o znaczeniu farmakogenetycznym i epidemiologicznym.

Do realizacji celu głównego pracy postawiono następujące hipotezy badawcze:

Hipoteza badawcza nr 1

Ze względu na wiele kluczowych funkcji jakie pełnią białka ABC, geny je kodujące są wysoce zakonserwowane i większość obserwowanej zmienności w ich sekwencjach stanowi zmienność o charakterze rzadkim (MAF <0,05).

Hipoteza badawcza nr 2

Populacja polska, pod względem zmienności w genach kodujących białka ABC, istotnie różni się od populacji o innych pochodzeniu etnicznym, jak populacji wschodnioazjatyckich czy afrykańskich.

Hipoteza badawcza nr 3

Populacja polska, pod względem zmienności w genach kodujących białka ABC, istotnie różni się od innych populacji referencyjnych o pochodzeniu europejskim.

Powyższą hipotezę zweryfikowano poprzez osiągnięcie kolejnych celów szczegółowych:

1. Opracowanie metodyki skanowania sekwencji kodujących białka ABC z wykorzystaniem metody topnienia produktów PCR w wysokiej rozdzielczości (HRM).
2. Przeskanowanie sekwencji kodujących dwa transportery ABC związane ze zjawiskiem MDR (*ABCC1*, *ABCG2*) w próbkach od ochotników z polskiej populacji i porównanie uzyskanych wyników z danymi dla innych populacji.
3. Klastrowanie próbek od ochotników z polskiej populacji i z innych referencyjnych populacji na podstawie zmienności genów całej nadrodziny białek ABC.

W związku z weryfikacją i odrzuceniem hipotezy badawczej nr 3 postawiono kolejny cel:

4. Identyfikacja wariantów, genów lub procesów z jakimi związane są białka ABC, które odpowiadają za uzyskany wynik klastrowania próbek od ochotników z populacji polskiej i z referencyjnych populacji europejskich.

3. MATERIAŁY I METODY

3.1. Syntetyczny opis materiałów i metod przedstawionych w cyklu publikacji wchodzących w skład rozprawy doktorskiej

3.1.1. Próbkę genomowego DNA

W latach 2010-2012, w ramach projektu „TESTOPLEK – Rola transporterów oporności wielolekowej w farmakokinetyce i toksykologii – testy *in vitro* w praktyce farmaceutycznej i klinicznej”, realizowanego w Katedrze Biofizyki Molekularnej Uniwersytetu Łódzkiego, zebrano kolekcję ponad 10 000 próbek śliny od niespokrewnionych ochotników z terenu całej Polski. W 2013 została ona zarejestrowana w katalogu sieci Infrastruktura Badawcza Biobanków i Zasobów Biomolekularnych (BBMRI, ang. *Biobanks and BioMolecular resources Research Infrastructure*) pod akronimem POPULOUS (POPulation — LOdz UniverSity Biobank) i zdeponowana w Pracowni Biobank Katedry Biofizyki Molekularnej Uniwersytetu Łódzkiego (Dobrowolska i in., 2019; Strapagiel i in., 2016). Każdy ochotnik wyraził pisemną, świadomą zgodę na udział w badaniu i wypełnił kwestionariusz. Badania uzyskały zgody Komisji Bioetyki Badań Naukowych Uniwersytetu Łódzkiego.

Genomowe DNA zostało wyizolowane z próbek śliny przy użyciu platformy MagNA Pure LC 2.0 (Roche, Bazylea, Szwajcaria) i po wewnętrznej kontroli ilościowej oraz jakościowej, metodą PCR w celu określenia płci (Strapagiel i in., 2019), a jego stężenie zostało znormalizowane dla wszystkich próbek do wartości 200 pg/μl.

Do badania zmienności genu *ABCC1* metodą skanowania (ang. *scanning/screening method*), z wykorzystaniem metody topnienia produktów PCR w wysokiej rozdzielczości (HRM, ang. *High Resolution Melting*), została wybrana grupa 190 losowych próbek DNA z kolekcji POPULOUS. Do oceny częstości występowania specyficznych wariantów w *ABCC1* z wykorzystaniem HRM metodą celowanego genotypowania (ang. *targeted genotyping method*) wybrano z kolei 380 innych próbek (Załącznik nr 1). Do badania zmienności genu *ABCG2* metodą skanowania wybrano następną pulę 190 próbek DNA z kolekcji POPULOUS (Załącznik nr 2). W każdym badaniu próbki DNA zebrane były odpowiednio na płytkach 96-łolkowych (każda z płytek zawierała również pojedynczą próbkę z wodą jako kontrolą negatywną reakcji).

3.1.2. Metoda skanowania genów *ABCC1* i *ABCG2* z wykorzystaniem HRM

Metoda HRM opiera się na precyzyjnym pomiarze profilu topnienia dwuniciowego DNA. Temperatura topnienia zależy od długości i składu nukleotydowego produktu PCR, a zatem zmiana nawet pojedynczej zasady generuje inną krzywą topnienia w porównaniu do pozostałych próbek, dzięki czemu można wykryć jednocześnie wiele wariantów genetycznych tworzących odrębne klastry topnienia. PCR w czasie rzeczywistym z następującym bezpośrednio po nim etapem topnienia przeprowadzano na płytkach 384-dołkowych z użyciem systemu Bio-Rad CFX384™ (Bio-Rad

Laboratories Inc., Hercules, USA), w duplikacie dla każdej próbki. Skanowaniu w poszukiwaniu wariantów genetycznych podlegały sekwencje eksonów i ich granice z intronami, uważane za najważniejsze obszary genów, jako kodujące oraz mogące mieć wpływ na poziom ekspresji lub alternatywny splicing. Aby zapewnić akceptowalną rozdzielczość topnienia odpowiednią dla wykrywania wariantów, w przypadku skanowania maksymalna długość topionego produktu PCR nigdy nie przekraczała 250 par zasad. W związku z tym, wiele długich eksonów musiało zostać podzielonych na osobno skanowane i zazębiające się obszary, co w przypadku skanowania *ABCC1* dało w sumie 51 fragmentów do analizy (Załącznik nr 1), a w przypadku *ABCG2* – 24 (Załącznik nr 2).

3.1.3. Metoda genotypowania wariantów w *ABCC1* z wykorzystaniem HRM

Równoległe do badania zmienności genu *ABCC1* metodą skanowania, z wykorzystaniem HRM przeprowadzono także genotypowanie wyselekcjonowanych wariantów w tym genie, aby porównać wyniki z obu metod i ocenić w ten sposób poprawność działania metody skanowania. Na podstawie baz danych lub doniesień literaturowych wybrano w tym celu 17 polimorfizmów pojedynczego nukleotydu (SNP, ang. *Single Nucleotide Polymorphism*), interesujących ze względu na wysoką polimorficzność lub potencjalne znaczenie kliniczne. Wykorzystana metodyka była taka sama jak w przypadku podejścia opartego na skanowaniu, jednak w tym przypadku starano się skrócić długość analizowanego fragmentu jak to tylko było możliwe, aby zredukować ryzyko detekcji wariantu innego niż celowany. Maksymalna długość badanego produktu PCR zatem nie przekraczała 125 par zasad (Załącznik nr 1).

3.1.4. Weryfikacja wyników HRM z wykorzystaniem sekwencjonowania metodą Sanger

Analizę krzywych topnienia prowadzono stosując oprogramowanie Bio-Rad Precision Melt Analysis Software, wersja 1.2 (Bio-Rad Laboratories Inc., Hercules, CA, USA). Wyniki topnienia dla każdej płytki z próbkami DNA analizowano oddzielnie, aby uniknąć fałszywie pozytywnych wyników (Słomka i in., 2017). Aby zweryfikować uzyskane wyniki zmienności genetycznej na podstawie HRM, za każdym razem wybierano po kilka próbek, reprezentujących każdy klaster topnienia, do sekwencjonowania metodą Sanger. W przypadku gdy liczność klastra była mniejsza niż cztery próbki w metodzie skanowania (Załącznik nr 1 i 2), lub mniejsza niż trzy próbki w metodzie genotypowania (Załącznik nr 1), sekwencjonowano metodą Sanger wszystkie próbki z klastra.

Wyselekcjonowane próbki przygotowano metodą PCR, poddano reakcji sekwencjonowania z użyciem BigDye Terminator V3.1 (Applied Biosystems, USA), a produkt sekwencjonowania oczyszczono stosując zestaw BigDye X-Terminator (Applied Biosystems, USA). Następnie każdy oczyszczony produkt analizowano używając 3500 Genetic Analyzer (Applied Biosystems, USA). Ten etap prac wykonano we współpracy z Instytutem Biologii Medycznej PAN w Łodzi.

3.1.5. Analiza zmienności genetycznej ABCC1 i ABCG2

Wyniki sekwencjonowania analizowano za pomocą oprogramowania CodonCode Aligner (CodonCode Corporation, Centerville, NY, USA), stosując jako referencje odpowiednie sekwencje genomowego DNA dla *ABCC1* (Załącznik nr 1) i *ABCG2* (Załącznik nr 2) pozyskane z bazy *GenBank*. Wyniki sekwencjonowania wybranych próbek porównano z odpowiednimi klastrami topnienia i na tej podstawie ustalono genotyp dla każdego klastra, a dzięki temu, dla każdej próbki. Każdemu z wykrytych wariantów przypisano odpowiednią nomenklaturę, tj. identyfikatory z bazy *dbSNP* (numery rs) oraz, w oparciu o sekwencje referencyjne, pozycję w DNA względem kodującej sekwencji, a dla wariantów wykrytych w eksonach, także pozycję kodowanego aminokwasu w białku (Załącznik nr 1 i 2).

Zgodność zaobserwowanego rozkładu genotypów z prawem Hardy'ego-Weinberga została zweryfikowana dla wszystkich wykrytych wariantów testem dokładnym (ang. *Hardy-Weinberg equilibrium exact test*), zgodnym jeśli $p > 0,001$.

Dla każdego z wykrytych wariantów została obliczona częstość występowania jego rzadszego allelu. Tego parametru używano przy porównaniu wyników uzyskanych z metod skanowania i genotypowania *ABCC1*, z wykorzystaniem testu chi-kwadrat dla liczby alleli, z poprawką Yatesa jeśli jakkolwiek ilość alleli była mniejsza niż pięć (istotne, jeśli wartość $p < 0,05$) (Załącznik nr 1). W ten sam sposób porównano uzyskane wyniki zmienności genetycznej *ABCC1* i *ABCG2* z dostępnymi danymi literaturowymi dla innych populacji (Załącznik nr 1 i 2).

3.1.6. Analiza nierównowagi sprzężeń i bloków haplotypowych ABCC1 i ABCG2

Analizę nierównowagi sprzężeń (LD, ang. *linkage disequilibrium*) i bloków haplotypów dla *ABCC1* i *ABCG2* przeprowadzono za pomocą programu Haploview 4.2. Ocenę LD wykonano dla każdej pary wykrytych wariantów używając parametrów $|D'|$ i r^2 (Załącznik nr 1 i 2). Uzyskane wyniki obu tych parametrów w trakcie analizy *ABCG2* porównano z dostępnymi danymi populacyjnymi za pomocą narzędzia LDmatrix z pakietu LDlink (Załącznik nr 2).

Identyfikację bloków haplotypowych w *ABCC1* przeprowadzono z wykorzystaniem metody opisaną przez Gabriela i in. (2002), zgodnie z którą podczas tworzenia bloków wykluczono polimorfizmy z $MAF < 0,05$ i uwzględniono ostatecznie 16 wariantów (Załącznik nr 1). Z kolei identyfikację bloków haplotypowych w *ABCG2* wykonano na podstawie tzw. testu czterech gamet (ang. *four-gamete test*) (Wang i in., 2002), ze względu na ogółem niewielką ilość wykrytych wariantów, a dla haplotypów z każdego bloku przeprowadzono analizę sieci za pomocą narzędzia Network 10 (Załącznik nr 2).

3.1.7. Analiza potencjalnego efektu wykrytych wariantów missensownych na funkcjonalność białka

Przewidywanie efektu wariantów missensownych (zmiany sensu), wykrytych w sekwencji *ABCC1*, na funkcjonalność i strukturę białka wykonano za pomocą narzędzia *in silico* PolyPhen-2 (Załącznik nr 1). Do takiej analizy wariantów, wykrytych w *ABCG2*, posłużono się ponadto narzędziem SIFT oraz złożonymi metodami przewidywania patogeniczności REVEL i MetaLR (Załącznik nr 2).

3.2. Opis materiałów i metod dla badań nieopublikowanych, wchodzących w skład rozprawy doktorskiej

3.2.1. Próbkę genomowego DNA

Próbki z polskiej populacji pochodziły od 6481 ochotników z powyżej opisanej kolekcji POPULOUS oraz od 179 ochotników, pochodzących z wytypowanych regionów geograficznych, grup kulturowych lub mniejszości etnicznych na terenie Polski, zebranych w ramach projektu „Cyfrowe udostępnianie zasobów biomolekularnych i opisowych Biobanku i Katedry Antropologii Uniwersytetu Łódzkiego – charakterystyka populacji zamieszkujących tereny dzisiejszej Polski na przestrzeni dziejów. Platforma informacyjna e-Człowiek.pl”, realizowanego w Pracowni Biobank i Katedrze Antropologii Uniwersytetu Łódzkiego. Ta mniejsza grupa próbek była reprezentowana przez przedstawicieli społeczności: Karaimów, polskich Tatarów, polskich Ormian, Podhalan, Kurpiów, Ślązaków i Hrubieszowian. Zgodę na badania wyraziła Komisja Bioetyki Badań Naukowych Uniwersytetu Łódzkiego. Wszyscy badani wyrazili pisemną, świadomą zgodę na udział w badaniu oraz wypełnili kwestionariusze.

Próbki śliny, zebrane były w fiolkach do długoterminowego przechowywania DNA Oragene OG-500 (DNA Genotek, Kanata, Kanada). Dla każdej przeprowadzono manualną izolację genomowego DNA z 500 µl śliny, zgodnie z dedykowanym protokołem producenta (PrepitL2P, PD-PR-052, DNA Genotek, Kanata, ON, Kanada). Wszystkie próbki DNA, zakwalifikowane do genotypowania z wykorzystaniem mikromacierzy, przeszły najpierw wnikliwą, wewnętrzną kontrolę jakości, tj. stężenie DNA ≥ 50 ng/µl (mierzone zestawem Quant-iT™ dsDNA Broad Range Assay Kit (Invitrogen™, Carlsbad, CA, USA), zgodnie z protokołem producenta) oraz określenie płci metodą PCR (Strapagiel i in., 2019). Probki DNA ostatecznie normalizowano do stężenia 50 ng/µl.

3.2.2. Genotypowanie z wykorzystaniem mikromacierzy

5625 próbek DNA włączonych do badania z kolekcji POPULOUS zgenotypowano przy użyciu mikromacierzy 24x1 Infinium HTS Human Core Exome (Illumina Inc., San Diego, CA, USA). Z kolei na mikromacierzach Infinium Multi-Ethnic Global-8 v1.0 (Illumina Inc.) zgenotypowano kolejnych 856 próbek DNA z kolekcji POPULOUS oraz 179 próbek DNA od przedstawicieli polskich regionów geograficznych, grup kulturowych lub mniejszości etnicznych. Te dwa zestawy mikromacierzy różniły się pulą oraz ilością wariantów jakie były możliwe do zgenotypowania z użyciem każdego z nich (551 945 wariantów w Infinium HTS Human Core Exome oraz 1 779 819 wariantów w Infinium Multi-Ethnic). Procedurę genotypowania przeprowadzono zgodnie z protokołami dostarczonymi przez producenta. W skrócie, DNA poddano amplifikacji, pofragmentowano enzymatycznie i zhybrydowano do szkiełek mikromacierzowych, na których przeprowadzono następnie syntezę nukleotydową w pozycjach wariantów i wybarwienie. Szkiełka zostały zeskanowane za pomocą iScan

(Illumina Inc.). Surowe pomiary intensywności fluorescencji, w celu przeprowadzenia kontroli jakości, wprowadzono do oprogramowania GenomeStudio 2.0 (Illumina, Inc.) z wykorzystaniem modułu Genotyping (Illumina Inc.). Wszystkie próbki spełniły kryteria kwalifikujące, jeśli wykazały parametr Call Rate $\geq 0,98$ oraz parametr 10% GenCall $\geq 0,4$. Warianty z nieprawidłowo sklastrowanym genotypem, wieloalleliczne SNP, nie spełniające prawa Hardy’ego-Weinberga na podstawie testu dokładnego ($p < 0,001$) oraz polimorfizmy zlokalizowane na chromosomach płciowych, zostały wykluczone z dalszej analizy.

3.2.3. Dane dla populacji referencyjnych

Dane dla populacji referencyjnych z całego świata pozyskano z bazy danych *1000 Genomes Project*, na podstawie zintegrowanych plików Variant Call Format (VCF) z fazy trzeciej projektu (Auton i in., 2015). W bazie łącznie 26 różnych populacji zaklasyfikowane zostało do pięciu regionów geograficznych z całego świata, przy czym do przynależności populacji brane było pod uwagę pochodzenie etniczne (Tab. 2).

Tabela 2. Populacje referencyjne z bazy *1000 Genomes Project* wzięte do analiz w rozprawie.

Kod populacji	Opis populacji	N	Kod wspólnego regionu geograficznego	Wspólny region geograficzny
CHB	Chińczycy Han z Pekinu, Chiny	103	EAS	Azja Wschodnia
JPT	Japończycy z Tokio, Japonia	104		
CHS	południowi Chińczycy Han	105		
CDX	Chińczycy Dai w Xishuangbanna, Chiny	93		
KHV	Wietnamczycy Kinh z Ho Chi Minh, Wietnam	99		
CEU	Mieszkańcy Utah o rodowodzie z północnej lub zachodniej Europy	99	EUR	Europa
TSI	Toskańczycy we Włoszech	107		
FIN	Finowie z Finlandii	99		
GBR	Brytyjczycy z Anglii i Szkocji	91		
IBS	Iberyjczycy z Hiszpanii	107		
YRI	Jorubowie z Ibadan, Nigeria	108	AFR	Afryka
LWK	Luhja z Webuye, Kenia	99		
GWD	Gambijczycy z zachodniej Gambii	113		
MSL	Mende z Sierra Leone	85		
ESN	Esan z Nigerii	99		
ASW	Amerykanie z USA o rodowodzie afrykańskim	61		
ACB	Mieszkańcy Barbadosu o rodowodzie afrykańskim	96		
MXL	Meksykanie z Los Angeles, USA	64	AMR	Mieszane populacje z Ameryki
PUR	Mieszkańcy Puerto Rico	104		
CLM	Kolumbijczycy z Medellin, Kolumbia	94		
PEL	Peruwiańczycy z Limy, Peru	85		
GIH	Hindusi Gudźarati, z Houston w Teksasie, USA	103	SAS	Azja Południowa
PJL	Pendźabczycy z Lahore, Pakistan	96		
BEB	Bengalczycy z Bangladeszu	86		
STU	Tamilowie ze Sri Lanki, z W. Brytanii	102		
ITU	Hindusi Telugu z W. Brytanii	102		

3.2.4. Imputacja genotypu

Dane uzyskane z genotypowania z wykorzystaniem mikromacierzy w celu imputacji przygotowano zgodnie z wytycznymi dostępnymi na stronie serwera *TopMed*. Allele w badanych wariantach zostały uzgodnione według referencji zgodnej z GRCh37/hg19. Następnie dane zostały podzielone na chromosomy i sfazowane. Dane z genotypowania dla obu rodzajów mikromacierzy zostały przygotowane jako oddzielne zbiory. Tak przygotowane dane przesłano na serwer *TopMed*, gdzie nastąpiła docelowa imputacja z użyciem narzędzia „minimac” (<https://genome.sph.umich.edu/wiki/Minimac/>). W pierwszej fazie imputacji została wykonana kontrola jakości danych, w tym usunięcie wariantów monomorficznych oraz takich, których rozkład częstości alleli był niezgodny z prawem Hardy’ego-Weinberga. Następnie zidentyfikowane zostały warianty referencyjne, względem których budowany był model imputacji. W *loci*, w których nie było wariantów referencyjnych, imputacja nie została wykonana z sukcesem, w związku z czym takich genów ABC niestety nie można było uwzględnić w kolejnych analizach (Tab. 3). Wytworzony został raport, w którym zidentyfikowane zostały warianty usunięte z analizy. Dane zimputowane dostępne były w formie dawek allelu alternatywnego, co umożliwiło zrandomizowane wywołanie genotypu.

Dla ujednolicenia metody identyfikacji genotypu, dane całogenomowe pochodzące z bazy *1000 Genomes Project* zostały również zimputowane z użyciem serwera *TopMed* według opisanego dalej schematu. Z plików VCF zostały wywołane konkretne pozycje zgenotypowane z wykorzystaniem mikromacierzy Infinium Multi-Ethnic. Na podstawie referencji GRCh37/hg19 zostały uzgodnione allele i dane poddano fazowaniu. Po podzieleniu na chromosomy, pliki VCF zostały wysłane jako oddzielne zbiory danych na serwer w celu imputacji. Finalnie, dane do analizy populacyjnej zostały uzyskane poprzez wyekstrahowanie ze zbioru danych konkretnych *loci*, w których znajdują się badane geny. Na tej podstawie wyfiltrowano wszystkie warianty z sekwencji genów ABC, doliczając także takie, które znajdowały się we flankujących fragmentach genów o długości 5000 pz. W przypadku jeśli warianty znajdowały się w sekwencjach zachodzących na siebie dwóch genów, były brane pod uwagę dla obu z nich.

3.2.5. Analiza skalowania wielowymiarowego

Do analizy skalowania wielowymiarowego (MDS, ang. *MultiDimensional Scale*) wykorzystano próbki ze wszystkich populacji z bazy *1000 Genomes Project* oraz próbki zgenotypowane z wykorzystaniem mikromacierzy Infinium Multi-Ethnic, ze względu na większą ilość pokrywających się wariantów, niż Infinium HTS Human Core Exome. Z dostępnych plików VCF dla *1000 Genomes Project* wywołane zostały warianty zgenotypowane dla próbek z populacji polskiej. Następnie dla połączonych dwóch zbiorów danych została wykonana redukcja wymiaru z użyciem metody MDS zaimplementowanej w narzędziu PLINK.

Tabela 3. Liczba wariantów w genach ABC po imputacji dla obu rodzajów mikromacierzy. Pozostałych genów ABC, dla których imputacja nie została zakończona z sukcesem, nie podano.

Gen	Liczba wariantów po imputacji dla mikromacierzy Infinium Multi-Ethnic	Liczba wariantów po imputacji dla mikromacierzy HTS Human Core Exome
ABCA2	3106	2992
ABCA3	8517	8022
ABCA7	3395	3213
ABCA12	20289	18852
ABCB5	18643	17429
ABCB6	1090	1011
ABCB8	2419	2276
ABCB9	6642	6261
ABCB11	11627	10847
ABCC2	6340	5950
ABCC4	29549	27535
ABCC8	9434	8996
ABCC9	14275	13124
ABCC10	2410	2262
ABCC11	8517	8055
ABCC12	7649	7221
ABCD2	7547	6948
ABCD3	10593	9623
ABCD4	2215	2069
ABCF1	2321	2157
ABCF2	1688	1556
ABCG1	11735	11106
ABCG4	1391	1308
ABCG5,ABCG8	9916	9322

3.2.6. Analiza IBS i redukcji wymiaru

Do analizy IBS (ang. *Identity By State*), z wariantów otrzymanych po imputacji (Tab. 3) wybrano tylko te, dla których allel rzadszy w badanej puli próbek pojawił się co najmniej pięć razy. Do estymacji IBS został użyty algorytm snpgdsIBS zaimplementowany w pakiecie SNPRelate. Następnie dla tak zbudowanej macierzy IBS zdefiniowana została odległość między punktami danych jako ujemny $\log_{10}(\text{IBS})$. Ta macierz odległości została następnie użyta w analizie redukcji wymiaru z użyciem funkcji `umap` w pakiecie `uwot`, który jest kompletną implementacją metody UMAP. Macierz rzutowana została na trzy wymiary, przy czym dla dokładnego określenia sąsiedztwa, dla każdego punktu danych rozważanych było 20 najbliższych sąsiednich punktów. Wizualizacja danych po „zanurzeniu” odbyła się z użyciem pakietu `plotly`.

3.2.7. Analiza ilości klastrow

Do modelowania ilości klastrow użyty został trójwymiarowy model Gaussowski, z wyestymowaną macierzą korelacji, zaimplementowany w pakiecie mclust. Do wyznaczenia ilości klastrow zostało użyte Bayesowskie kryterium informacyjne (ang. *Bayes Information Criterion*) – przy czym maksymalna ilość klastrow ustalona została na 10. Górne ograniczenie było konieczne, ponieważ algorytm miał tendencję do przeuczenia ze względu na ilość punktów danych i definiowania bardzo dużej (>40) ilości skupień.

3.2.8. Identyfikacja wariantów z różną częstością alleli pomiędzy klastrami

Do detekcji wariantów o różnej częstości alleli pomiędzy klastrami użyto nieparametrycznego testu Kruskalla-Wallisa. Nie wprowadzono poprawek związanych z wielokrotnym testowaniem z uwagi na potrzebę detekcji całych haplotypów (bloków haplotypowych), które mają różną częstość alleli pomiędzy skupieniami danych.

3.2.9. Analiza nierównowagi sprzężeń

Do analizy nierównowagi sprzężeń został użyty algorytm snpgdsLDMat zaimplementowany w pakiecie SNPRelate. Macierz LD dla par wariantów na wspólnym chromosomie obliczono używając metody „złożonej miary” (ang. *composite measure*) nierównowagi sprzężeń.

3.2.10. Analiza potencjalnego efektu wykrytych wariantów missensownych na funkcjonalność białka

Do przewidywania efektu wariantów missensownych, wykrytych w puli wariantów o różnej częstości alleli pomiędzy klastrami na funkcjonalność białka, wykorzystano program Variant Effect Predictor (VEP), dostępny w bazie danych Ensembl. W tym celu użyto narzędzi PolyPhen-2 oraz SIFT.

3.2.11. Analiza asocjacyjna całego genomu

Analiza została przeprowadzona z użyciem pakietu GENESIS w środowisku R. Po przygotowaniu danych w odpowiednim formacie została wyestymowana macierz pokrewieństwa (ang. *kinship*) z użyciem metody KING-robust. Nie zostały wykonane żadne poprawki na istotne komponenty analizy redukcji wymiaru (PC-AiR, ang. *Principal Components Analysis in Related Samples*) z uwagi na to, że wówczas zostałby usunięty sygnał związany z podziałem na klastry. Docelowa analiza asocjacyjna została wykonana w oparciu o mieszane modele liniowe z macierzą pokrewieństwa jako macierz korelacji modelu. Istotność każdego wariantu w kontekście modelu addytywnego została oceniona z użyciem „Score test”.

3.2.12. Analiza asocjacyjna całego transkryptomu

Do oszacowania znaczenia genetycznej komponenty ekspresji zostały użyte bazy referencyjne pochodzące z modeli wykorzystujących pakiet glmnet dla danych z bazy *GTEX*, dostępne w repozytorium IMlabtools (<https://github.com/hakylimlab/MetaXcan>). Do oszacowania zostały wykorzystane dane zimputowane oraz wszystkie dostępne w bazie modele narządów i ich struktur. Wykorzystano narzędzie PrediXcan (po uprzedniej harmonizacji danych) oraz standardowy model liniowy zaimplementowany w pakiecie limma w R – oznacza to wykonanie poprawki na wielokrotne testowanie zgodne z metodą zaimplementowaną w funkcji eBayes w pakiecie limma.

3.2.13. Analiza fenotypowa na podstawie zróżnicowania populacji polskiej

Związek pomiędzy otrzymanym klastrowaniem populacji polskiej a poszczególnymi chorobami cywilizacyjnymi, na podstawie danych ankietowych ochotników z kwestionariusza, sprawdzono testem chi-kwadrat w tabeli 2x9, istotne jeśli $p < 0,05$. Związek z chorobą konkretnych klastrow był oceniany na podstawie porównania różnic pomiędzy udziałem osób zdrowych a tych deklaruujących chorobę, liczonych dla każdego klastra.

4. WYNIKI

4.1. Syntetyczne omówienie wyników przedstawionych w publikacjach wchodzących w skład rozprawy doktorskiej

4.1.1. Załącznik nr 1

Słomka, M., Sobalska-Kwapis, M., Korycka-Machala, M., Bartosz, G., Dziadek, J. and Strapagiel, D., 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. *BMC Genet.*, 16: 114.

Dane uzyskane z badania metodą celowanego genotypowania posłużyły do weryfikacji zaobserwowanych wyników zmienności genetycznej *ABCC1* otrzymanych metodą skanowania i dzięki temu do porównania dokładności obu metod wykorzystujących HRM. Spośród 17 SNP poddanych celowanemu genotypowaniu, dla siedmiu nie wykazano żadnej zmienności tą metodą, ani też nie zostały one znalezione podczas skanowania. Chociaż warianty te mogą mieć potencjalne znaczenie kliniczne, to zidentyfikowano je wcześniej tylko podczas badań *in vitro* lub w populacjach innych niż odmiany białej człowieka, stąd ich niewykrycie w tym badaniu nie było zaskakujące. Z kolei wartości MAF uzyskane dla ośmiu SNP obiema metodami były dość podobne i nieistotne statystycznie. Jedynie w przypadku dwóch polimorfizmów stwierdzono marginalnie istotne różnice w wartościach MAF ($0,05 > p > 0,01$) pomiędzy obiema metodami, co jednak prawdopodobnie było spowodowane wielkością grupy badanej i niezależnością grup próbek badanych obiema metodami. Dane te pozwoliły stwierdzić, że HRM jest wydajną i dokładną techniką zarówno dla ukierunkowanego genotypowania wybranych polimorfizmów, jak i do badania zmienności genetycznej w większych sekwencjach, w tym także genach kodujących białka z nadrodziny ABC.

W badaniu przeprowadzono kompletne skanowanie sekwencji 30 z 31 eksonów *ABCC1* w poszukiwaniu wariantów polimorficznych w 190 próbkach pochodzących z populacji polskiej. Jedynie ekson pierwszy nie został ostatecznie poddany badaniu z powodu problemów z amplifikacją w trakcie PCR, jako region wyjątkowy bogaty w pary GC. W badaniu metodą skanowania znaleziono w sumie 46 wariantów polimorficznych. Duża ich część, bo aż 30, reprezentowała warianty rzadkie (MAF $< 0,05$), w tym siedem nie zgłoszonych jeszcze nigdy wcześniej w bazach danych. 21 wariantów typu SNP znajdowało się w sekwencjach eksonów, a 11 z nich zidentyfikowano jako niesynonimiczne, w tym jeden nowy SNP znaleziony w tym badaniu. Spośród wszystkich wykrytych niesynonimicznych wariantów tylko jeden, c.2012G>T (p.Gly671Val), występował homozygotycznie i to jako wariant częsty (MAF = 0,077). Pozostałe SNP, które zostały znalezione w eksonach jako homozygotyczne były wariantami synonimicznymi. W sumie w eksonach znaleziono sześć SNP, które można uznać za warianty częste (MAF $\geq 0,05$). Oprócz polimorfizmów w regionach kodujących, 23 SNP znajdowały się w intronach, a jeden w sekwencji 3' UTR. Ponadto podczas skanowania znaleziono jedną delecję

w sekwencji intronu, a dwa inne warianty polimorficzne zlokalizowane w intronach występowały we wszystkich próbkach wyłącznie jako homozygoty allelu niereferencyjnego.

Na podstawie dostępnych danych literaturowych dokonano porównania zaobserwowanej w niniejszej pracy zmienności genetycznej w *ABCC1* z tą odnotowaną w badaniach nad innymi populacjami. W populacji polskiej wykryto zatem 18 wariantów, które występowały również w populacji japońskiej (Fukushima-Uesaka i in., 2007) oraz 13 wariantów, które wykryto w populacji chińskiej (Wang i in., 2004). Dla niektórych wariantów znalezionych w niniejszym badaniu stwierdzono istotne różnice między wartościami MAF – trzy SNP wykryto z wyższym MAF niż we wspomnianych populacjach azjatyckich, z kolei siedem innych polimorfizmów miało wartość MAF istotnie niższą od co najmniej jednej z tych populacji azjatyckich. Porównując natomiast uzyskane wyniki z dostępnymi danymi publikacyjnymi dla innych populacji odmiany białej, zidentyfikowano 25 wspólnych wariantów, których wartość MAF w zdecydowanej większości nie była istotnie różna niż w populacji polskiej. Dodatkowo, ta sama grupa wariantów zidentyfikowana również w niniejszym badaniu różnicowała opisywaną populację odmiany białej od populacji pochodzenia azjatyckiego (Wang i in., 2006). Wyniki te pozwoliły wyciągnąć wniosek, że chociaż duża grupa wariantów ma podobne wartości MAF u różnych populacji, to istnieją również takie polimorfizmy, których wartości MAF są istotnie różne. Przyczyną wykrytych różnic było zatem najprawdopodobniej pochodzenie etniczne donorów próbek (Oselin i in., 2003), ponieważ opisywane rozbieżności pomiędzy samymi populacjami azjatyckimi były mniejsze, niż między nimi a populacjami odmiany białej, w tym populacją polską, jak w niniejszym badaniu.

W przeprowadzonej analizie nierównowagi sprzężeń zostały wykryte pary polimorfizmów o silnym LD, a wyniki te okazały się być bardzo podobne do tych uzyskanych dla populacji japońskiej (Fukushima-Uesaka i in., 2007). Na tej podstawie można wnioskować, że sprzężenia między specyficznymi wariantami są wysoce zakonserwowane ewolucyjnie i występują niezależnie od populacji. Analiza LD na podstawie danych dla dziewięciu SNP umożliwiła zidentyfikowanie w niniejszym badaniu 12 częstych haplotypów (częstość $>0,01$) w czterech blokach. Osiem z tych SNP zostało określonych *in silico* jako „haplotype tag SNP” (htSNP), których oznaczenie wystarcza do wskazania haplotypu. Niewielkie zróżnicowanie haplotypów *ABCC1* zaobserwowane w niniejszym badaniu było spójne z danymi literaturowymi dla populacji pochodzenia europejskiego (Leschziner i in., 2006). Wykryte bloki haplotypowe w populacji polskiej nie obejmowały długich fragmentów genu, a liczba SNP je tworząca była niewielka, co było spowodowane ogółem niewielką ilością wariantów częstych (MAF $>0,05$) znalezionych w całym badaniu oraz brakiem sprzężeń pomiędzy wariantami już wykrytymi. Mimo to, główne rdzenie bloków i ich granice odzwierciedlały jednak te, które zostały zidentyfikowane w badaniach nad innymi populacjami. To prowadzi do wniosku, że

miejsca rekombinacji w *ABCC1* są wysoce zakonserwowane bez względu na populację (Fukushima-Uesaka i in., 2007; Leschziner i in., 2006; Wang i in., 2004).

Analiza *in silico* wpływu wszystkich niesynonimicznych wariantów wykrytych w tym badaniu na potencjalną funkcjonalność białka wskazała, że aż sześć z nich, w tym jeden nowy zidentyfikowany SNP, w zależności od użytego modelu, ma efekt albo potencjalnie szkodliwy (ang. *possibly damaging effect*), albo prawdopodobnie szkodliwy (ang. *probably damaging effect*). Dane literaturowe nie potwierdziły jednak jednoznacznie wyników takich obserwacji otrzymanych *in silico*. Letourneau i in. (2005) przebadali bowiem 10 missensownych SNP w *ABCC1* i prawie wszystkie z nich, zaklasyfikowane przez PolyPhen-2 jako szkodliwe, nie wpływały ani na ekspresję, ani na funkcjonalność białka. Podobny wynik opisano również wcześniej dla SNP c.2012G>T (p.Gly671Val), wykrytego i w niniejszym badaniu. Mimo tego, że ta substytucja zlokalizowana jest pobliżu sekwencji, która koduje wysoce konserwatywny motyw Walker A (sześć aminokwasów powyżej) i zmniejszyła ekspresję mRNA, to nie zmieniła ona właściwości transportowych białka, chociaż analiza PolyPhen-2 wyraźnie przewidywała jej szkodliwy wpływ (Conrad i in., 2001). Inne dane potwierdziły jeszcze te obserwacje, bowiem żadna z substytucji p.Arg633Gln, p.Gly671Val, p.Arg723Gln, wykrytych również w niniejszym badaniu, nie zmieniła funkcjonalności transportera MRP1 (Conseil i Cole, 2013). Z drugiej strony istnieją dane literaturowe dowodzące związku wyżej wymienionych wariantów ze skutecznością farmakoterapii lub ich eksperymentalnego wpływu na funkcjonalność białek. Opisywano m.in., że wariant c.2168G>A (p.Arg723Gln) redukował *in vitro* oporność komórek z nadekspresją MRP1 na niektóre leki przeciwnowotworowe (Yin i in., 2009) i wpływał na odpowiedź na chemioterapię u pacjentów z rakiem jajnika (Obata i in., 2006), choć jednak nie u pacjentów z rakiem płuc (Yin i in., 2011), a dwa inne warianty, c.825T>C (p.Val275=) i c.2012G>T (p.Gly671Val), były skorelowane z występowaniem gorączki neutropenicznej u pacjentek z rakiem piersi (Vulsteke i in., 2013). Rozbieżność obserwowana w różnych badaniach dotyczących znaczenia klinicznego wariantów mogła wynikać z niedoskonałości modelu komórkowego *in vitro*, który może być niewystarczający do ujawnienia wszystkich funkcji jakie te białka pełnią w organizmie (Conrad i in., 2001). Sugeruje to zatem, że aby wyjaśnić wpływ poszczególnych SNP w *ABCC1* na fenotyp, należałoby przeprowadzić wszechstronną charakterystykę biochemiczną, farmakologiczną i molekularną odpowiednich mutantów MRP1, ponieważ efekty te mogą być złożone i dość selektywne (Conseil i Cole, 2013). To z kolei prowadzi do wniosku, że tylko kompleksowe badanie polimorfizmów i ich fenotypów może wyjawiać rzeczywiste znaczenie tych wariantów. Dlatego warianty zakwalifikowane *in silico* jako szkodliwe dla działania białka powinny zostać jeszcze przebadane eksperymentalnie. Ponadto w niniejszym badaniu nie znaleziono żadnego nowego wariantu w kluczowych domenach białka, ani żadnej mutacji w tych sekwencjach, zidentyfikowanych wcześniej podczas badań *in vitro* jako szkodliwe. To pozwoliło wyciągnąć wniosek, że regiony te

w *ABCC1* są wysoce konserwatywne i jest mało prawdopodobne, aby znaleźć występujące w nich warianty niesynonimiczne w populacji zdrowych ochotników.

Przeprowadzone badanie było pierwszym, które skupiało się całościowo na populacyjnej analizie jakiegokolwiek genu kodującego białko ABC w populacji polskiej i umożliwiło opisanie zmienności genetycznej *ABCC1* Polaków.

4.1.2. Załącznik nr 2

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G. and Strapagiel, D., 2020. Comprehensive Analysis of ABCG2 Genetic Variation in the Polish Population and Its Inter-Population Comparison. *Genes (Basel)*, 11(10)

W oparciu o opracowaną na pierwszym etapie rozprawy metodologię wykrywania wariantów polimorficznych w genach metodą skanowania z wykorzystaniem HRM, poddano badaniu 14 z 16 eksonów *ABCG2* wraz z okalającymi je sekwencjami intronowymi dla 190 próbek z populacji polskiej. Z całej sekwencji kodującej białko nie udało się zeskanować tylko jej ostatniego fragmentu w eksonie 16, ze względu na problemy z rozdzielczością topnienia produktów PCR, które nawet przy niewielkiej długości zawierały wiele topiących się niezależnie domen (ang. *multiple melting domains*). Podobnie nie udało się zeskanować niekodujących regionów genu w eksonie 1 i 16 ze względu na długie sekwencje bogate w pary GC, co skutkowało trudnościami w amplifikacji PCR.

W badaniu zidentyfikowano w sumie 17 polimorfizmów. Ponieważ dwa z nich znajdowały się w tej samej pozycji nukleotydowej, ale wykazywały różne efekty, nonsensowny c.706C>T (p.Arg236Ter) i synonimiczny c.706C>A (p.Arg236=), policzono je jako oddzielne SNP. W sekwencjach eksonów znaleziono więc łącznie osiem SNP, pięć z nich to były warianty niesynonimiczne oraz jeden wspomniany nonsensowny. Tylko jeden z nich został znaleziony jako homozygotyczny, c.421C>A (p.Gln141Lys) z oszacowaną wartością MAF = 0,117. W sekwencjach intronowych zidentyfikowano osiem SNP, jeden z nich homozygotycznie oraz wykryto jedną delecję. Jedynie dwa wykryte SNP, w tym jeden eksonowy, prezentowały wartość MAF $\geq 0,05$, co sugeruje, że zdecydowana większość obserwowanej zmienności w *ABCG2* jest zmiennością rzadką.

Porównanie otrzymanych wyników z dostępnymi danymi literaturowymi dla innych populacji wykazało istotnie niższe wartości MAF w populacji polskiej, niż w azjatyckich: chińskiej (Wang i in., 2004), japońskiej (Maekawa i in., 2006) i koreańskiej (Lee i in., 2007), aż dla pięciu z ośmiu współdzielonych SNP. Natomiast nie wykazano żadnych istotnych różnic w wartościach MAF dla pięciu wspólnych SNP pomiędzy populacją polską, a holenderską (Bosch i in., 2005). Obserwacje te okazały się zgodne z wynikami badań skupiających się na dwóch, najczęstszych SNP w sekwencji kodującej *ABCG2*, mianowicie c.34G>A (p.Val12Met) oraz c.421C>A (p.Gln141Lys). Populacja polska prezentowała istotnie niższe wartości MAF dla tych SNP niż dwie populacje azjatyckie: wietnamska i chińska (Lee i in., 2007), natomiast wartości te nie różniły się od wartości MAF dla innych populacji pochodzenia europejskiego: szwedzkiej (Backstrom i in., 2003), węgierskiej (Fischer i in., 2007), czeskiej (Stiburkova i in., 2017) oraz niemieckiej (Muller i in., 2009; Stark i in., 2009). Spostrzeżenia te korelują również z danymi dla generalnych grup etnicznych, dostępnymi jedynie dla c.421C>A (p.Gln141Lys) – stwierdzono na ich podstawie: istotnie niższą wartość MAF w populacji polskiej niż

w populacji chińskiej Han, brak różnic w stosunku do badanych grup z odmiany białej oraz istotnie wyższą wartość od tej w badanych grupach o pochodzeniu afrykańskim (de Jong i in., 2004).

Uzyskane w niniejszym badaniu wyniki jednoznacznie zatem wskazują, że częstości tych dwóch polimorfizmów w *ABCG2* w populacji polskiej, nie różnią się znacząco od innych populacji odmiany białej i są istotnie niższe niż w populacjach pochodzenia azjatyckiego. Co zaskakujące zatem, wniosek ten stał w sprzeczności z wcześniejszymi i jedynymi dotychczas dostępnymi wynikami zmienności genetycznej *ABCG2* dla populacji polskiej (Niebudek i in., 2019; Salagacka-Kubiak i in., 2016), badającymi oba te SNP. Uzyskane wartości MAF w przytoczonych pracach dla c.34G>A (p.Val12Met) i c.421C>A (p.Gln141Lys) wynosiły odpowiednio 0 i 0,010, były więc znacznie niższe niż wartości ustalone w badaniu wchodzącym w skład niniejszej rozprawy: odpowiednio 0,032 ($p = 0,027$) i 0,117 ($p < 0,001$). Bardzo prawdopodobnym wyjaśnieniem takiego niedoszacowania byłby m.in. niereprezentatywny dobór grupy badanej. Autorzy wspomnianych prac wykorzystali metodę polimorfizmu długości fragmentów restrykcyjnych (RFLP), która jednak nie obejmuje żadnego dodatkowego kroku weryfikującego dokładność wyników, jakim w niniejszym badaniu było sekwencjonowanie metodą Sangera. Ponadto w pierwszym badaniu składającym się na niniejszą rozprawę wykazano również, że stosowana metodologia z wykorzystaniem HRM jest skuteczna i czuła zarówno dla skanowania jak i celowanego genotypowania wariantów polimorficznych. Zatem, mimo pewnych ograniczeń tej metody diskutowanych w poprzedniej publikacji (Słomka i in., 2017), wykazano, że jest ona użyteczna w badaniu zmienności genetycznej. W niniejszej rozprawie określono częstości występowania SNP c.34G>A (p.Val12Met) i c.421C>A (p.Gln141Lys) w populacji polskiej i stwierdzono, że obserwowana zmienność w *ABCG2* nie różni się istotnie w populacji polskiej i w innych przebadanych populacjach odmiany białej.

Ze względu na najwyższą częstość występowania w populacji, SNP c.34G>A (p.Val12Met) i c.421C>A (p.Gln141Lys) były też najczęściej badanymi *in vitro* oraz *in vivo* polimorfizmami *ABCG2*, o czym świadczą doniesienia literaturowe. Chociaż znaczenie c.34G>A (p.Val12Met) nie zostało do końca wyjaśnione, to często jednak donoszono, że polimorfizm c.421C>A (p.Gln141Lys) był związany ze zmienioną funkcjonalnością białka *in vitro*. Obserwowano m.in.: niższy poziom dojrzewania mRNA, zmiany fałdowania białek i transportu, w porównaniu z ekspresją białka typu dzikiego w błonie komórkowej (Hira i Terada, 2018; Zábó i in., 2020). Obniżenie wydajności tych procesów prawdopodobnie zachodzi w efekcie zmiany aminokwasu zlokalizowanego w obrębie ważnego regionu wiążącego ATP, pomiędzy motywami Walker A i B białka, co wpływa na poziom aktywności transportera (Noguchi i in., 2009). Obserwacje te znajdowały potwierdzenie w danych klinicznych, niezależnie od badanej populacji, leków czy rodzaju nowotworu – redukcja aktywności *ABCG2* opisywana była jako związana z niepożądaną odpowiedzią na chemioterapię przeciwnowotworową, choć efekt farmakogenomiczny SNP c.421C>A (p.Gln141Lys) był zależny od substratu i genotypu (Hira

i Terada, 2018; Jani i in., 2014; Zhang i in., 2018). Odnotowano także istotne różnice w odpowiedzi na leczenie, które były zależne od pochodzenia etnicznego – stwierdzone je między populacjami pochodzenia wschodnioazjatyckiego i europejskiego (Liu i in., 2017). Co więcej, liczne badania na populacjach różnego pochodzenia etnicznego, w tym badania asocjacyjne całego genomu (GWAS, ang. *Genome Wide Association Study*), wykazały również, że c.421C>A (p.Gln141Lys) jest głównym czynnikiem ryzyka choroby naczyniowej i hiperurykემii (Chen i in., 2018; Woodward i in., 2009; Yamagishi i in., 2010; Zhang i in., 2013), natomiast inne badania sugerowały z kolei rolę ochronną SNP c.34G>A (p.Val12Met) przed chorobą naczyniową (Yu i in., 2017; Zhou i in., 2014). Zatem określenie właściwej częstości obu tych wariantów w populacji polskiej wydaje się być ważnym krokiem w kierunku przyszłych zastosowań klinicznych i terapii celowanej.

Mimo opisywanego znaczenia klinicznego obu wspomnianych powyżej wariantów, zaskakująco nie zostały one wskazane w analizie *in silico* jako potencjalnie szkodliwe dla funkcjonowania białka. Z kolei dwa inne rzadkie warianty wykryte w tym badaniu, zostały uznane za potencjalnie szkodliwe. Dane literaturowe potwierdziły takie przewidywanie dla SNP c.1714A>C (p.Ser572Arg), gdyż w jego efekcie wykazano obniżony poziom białka *ABCG2* w błonie plazmatycznej (Toyoda i in., 2019), natomiast brak było takich doniesień dla drugiego SNP c.335C>A (p.Pro112Gln). To sugeruje, że niektóre rzadkie warianty niesynonimiczne mogą mieć także znaczenie kliniczne, jednak w celu określenia ich realnego wpływu wymagane byłyby badania funkcjonalne. Chociaż więc tego typu SNP mogłyby się okazać ważne dla ich nosicieli, to jednak ich faktyczny wpływ na skuteczną terapię jest trudny do oszacowania, ponieważ obserwowana ich częstość jest zbyt niska.

Co więcej, w prezentowanym badaniu potwierdzono również występowanie heterozygotycznego, nonsensownego SNP c.706C>T (p.Arg236Ter), który był w literaturze wiązany z grupą krwi Jr(a-). Fenotyp taki jak dotąd był częściej opisywany w populacji japońskiej niż wśród badanych z odmiany białej (Haer-Wigman i in., 2014), jednak takie dane nie były zbierane od dawców próbek wykorzystanych w niniejszym badaniu i tym samym zweryfikowanie fenotypu próbki nie było możliwe. Jednak wykrycie takiego wariantu w stosunkowo niewielkiej puli 190 próbek może sugerować, że warianty związane z fenotypem Jr(a-) w populacji polskiej nie są tak rzadkim zjawiskiem, jak wcześniej zakładano i aby to zweryfikować, należałoby poddać genotypowaniu dużo większy zbiór próbek. Z drugiej jednak strony, może to być odkrycie incydentalne co sugerowałoby, że wpływ tak rzadkich polimorfizmów na całą polską populację wydaje się być raczej niewielki.

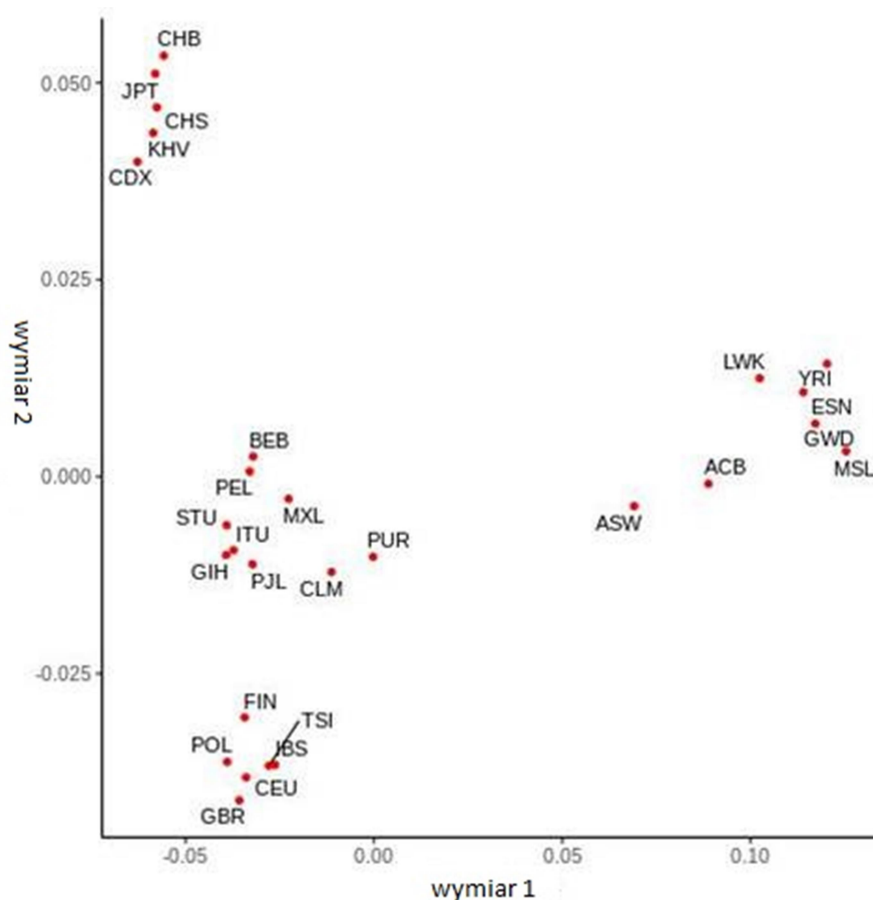
Przeprowadzono także analizę LD przy użyciu parametrów r^2 i $|D'|$ – sprzężenie wykazano jedynie dla trzech SNP. Wynik ten był zgodny z danymi innych populacji z bazy *1000 Genomes Project*, gdzie takie same zależności stwierdzono wśród populacji pochodzenia europejskiego, jak i ogółem światowej populacji. Sugerowałoby to zatem, że model dziedziczenia poszczególnych wariantów jest raczej ewolucyjnie zakonserwowany, niezależnie od badanej populacji. Wniosek ten wsparła analiza

bloków haplotypowych. Zdefiniowano bowiem miejsce rekombinacji w sekwencji genu i na tej podstawie wyznaczono dwa bloki ze zidentyfikowanymi w sumie siedmioma haplotypami, których częstości oszacowano na co najmniej 1%. Pięć SNP zostało oznaczonych jako htSNP. Analiza częstości otrzymanych haplotypów między populacjami, dla pokrywających się wariantów, wykazała zwłaszcza w bloku pierwszym bardzo duże podobieństwo między populacjami polską a holenderską (Bosch i in., 2005), natomiast w populacji japońskiej częstości poszczególnych haplotypów zdecydowanie się różniły (Maekawa i in., 2006). Blok drugi przejawiał już dużo większe podobieństwo między wszystkimi populacjami. Wyniki te potwierdzają wniosek, że różnice między populacjami nie dotyczą pojedynczych wariantów *ABCG2*, tylko raczej dziedziczonych wspólnie *loci*.

4.2. Omówienie wyników badań nieopublikowanych wchodzących w skład rozprawy doktorskiej

4.2.1. Zróżnicowanie populacji z bazy *1000 Genomes Projects* oraz populacji polskiej na podstawie zmienności genów ABC

Na podstawie zmienności genów ABC, które udało się zimputować, dokonano porównania pomiędzy wszystkimi populacjami referencyjnymi z bazy *1000 Genomes Project* oraz kohorty 1035 ochotników z populacji polskiej. Dla lepszej rozdzielczości rozdziału, podobieństwo między wszystkimi próbkami z każdej populacji zostało zredukowane do jednego punktu i przedstawione na wykresie MDS (Ryc. 1). W ten sposób populacje utworzyły wyraźne cztery zgrupowania, odpowiadające regionowi świata, z którego etnicznie pochodzą, z wyjątkiem mieszanych populacji amerykańskich, które były zbliżone do populacji południowoazjatyckich i tworzyły jedno zgrupowanie. Osobno skupione były natomiast populacje pochodzenia europejskiego, wschodnioazjatyckiego oraz afrykańskiego – w tym konkretnym przypadku populacje wywodzące się etnicznie z Afryki, ale mieszkające wiele pokoleń w innej części świata (USA, Barbados), zdecydowanie jednak odstawały od rdzennych populacji afrykańskich. Populacja polska (POL) na tym tle zgodnie z oczekiwaniami grupowała razem z pozostałymi populacjami pochodzenia europejskiego.



Rycina 1. Grupowanie populacji polskiej oraz populacji referencyjnych z bazy *1000 Genomes Project* na podstawie zmienności genów ABC. Legenda populacji w Tabeli 2.

4.2.2. Zróźnicowanie populacji europejskich oraz populacji polskiej pod względem zmienności genów ABC

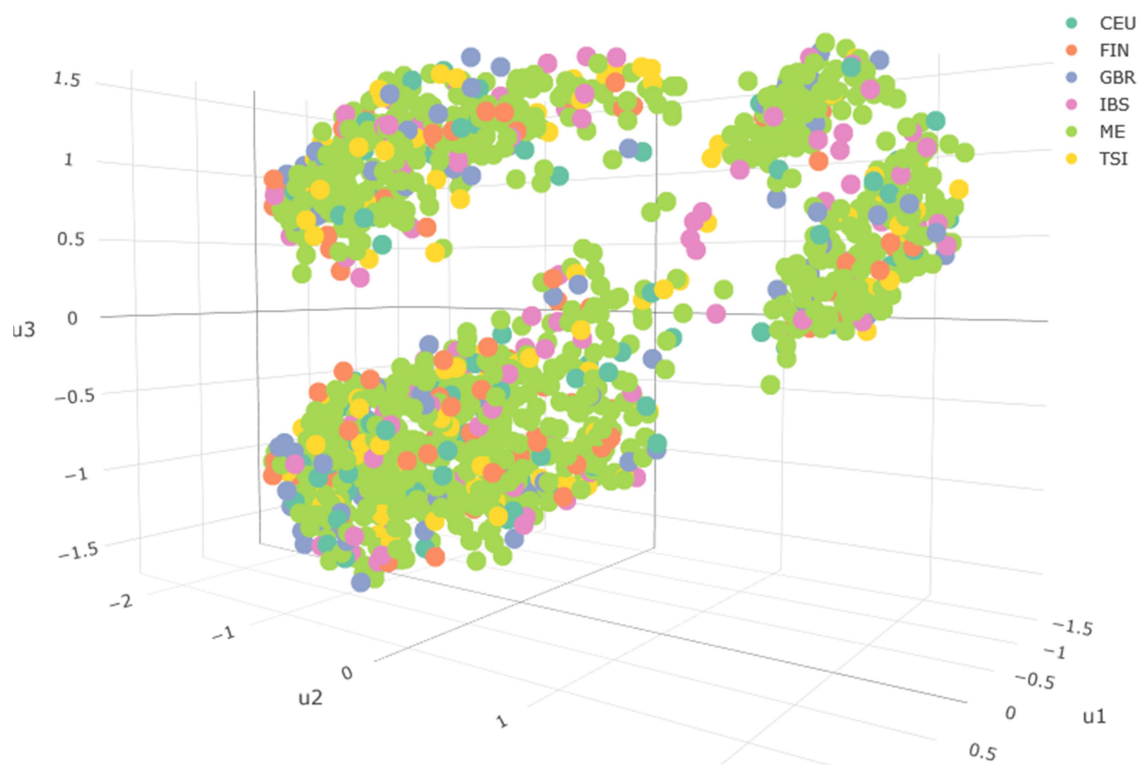
Analiza IBS rzutowana w przestrzeni przez macierz odległości UMAP dla 19153 współdzielonych wariantów, pomiędzy pięcioma populacjami europejskimi z bazy *1000 Genomes Project*, a zgenotypowaną kohortą (ME) 1035 próbek z populacji polskiej, pogrupowała wszystkie próbki na cztery klastry, przy czym dwa nieznacznie od siebie oddzielone (Ryc. 2). Co ciekawe, otrzymane klastrowanie nie było w ogóle związane z pochodzeniem populacyjnym próbek, ponieważ w każdym klastrze zaobserwowano równomierne rozmieszczenie próbek z każdej populacji wziętej do analizy.

Przeprowadzono następnie analogiczną analizę, ale jedynie dla 3962 współdzielonych wariantów częstych ($MAF \geq 0,05$) (Ryc. 3). Otrzymano wynik podobny jak wcześniej, tj. cztery główne zgrupowania próbek odpowiadające tym samym, co we wcześniejszej analizie. W porównaniu jednak do klastrowania opartego na wszystkich wariantach, w tym podejściu zaobserwowano znacznie wyraźniejsze rozdzielanie pomiędzy poszczególnymi zgrupowaniami i mniej próbek na ich granicach. Otrzymane w tej analizie cztery zgrupowania próbek dodatkowo poklastrowano i metodą najlepszego dopasowania wyznaczono w sumie dziewięć odrębnych klastrów, pomiędzy którymi można wskazać granice (Tab. 4, Ryc. 4). Największe dwa zgrupowania próbek obejmowały zatem klastry ponumerowane odpowiednio 1, 3, 8, 9 oraz 2 i 6. Z kolei na zgrupowania próbek słabiej od siebie rozdzielone składały się klastry o numerach 5 i 7 oraz samodzielnie odseparowany klaster nr 4.

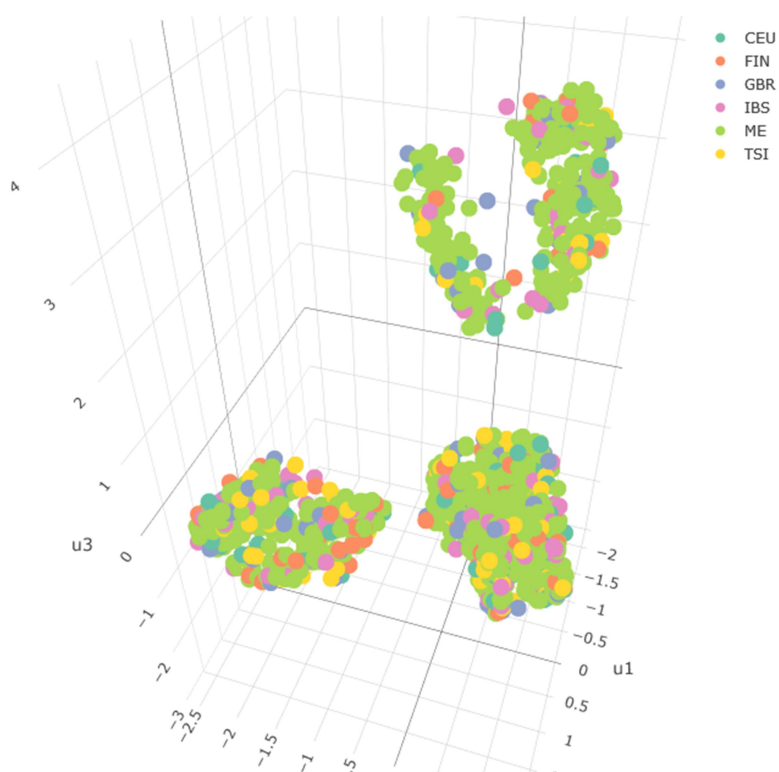
Ten sam rodzaj analizy dla 15191 wariantów rzadkich ($MAF < 0,05$) ujawnił zupełnie inny obraz grupowania próbek, z jedną, główną koncentracją punktów i co najmniej trzema oddzielnymi klastrami (Ryc. 5). Także w tym przypadku klastrowanie nie było związane z pochodzeniem populacyjnym próbek. Największa grupa próbek wykazała duże rozproszenie w jednym kierunku, co wskazuje na niski poziom współdzielenia haplotypów między badanymi próbkami. Pozostałe klastry wydają się być gęściej rozmieszczone wokół swoich centrów, jednakże jest to najprawdopodobniej jedynie efekt znacznie mniejszej liczności próbek jakie je tworzą. Na separację tych kilku klastrów od siebie prawdopodobnie wpływają warianty z MAF nieznacznie poniżej progu uznania ich za częste, których efekt zdominował typowo rzadkie warianty. W związku z brakiem korelacji zaobserwowanego rozkładu klastrów dla wariantów rzadkich, w porównaniu do wszystkich wariantów oraz do wariantów częstych, jak i marginalną częstością w populacji ($MAF < 0,05$), w kolejnych analizach efekt zmienności rzadkiej nie odgrywał znaczącej roli.

Tabela 4. Liczność poszczególnych dziewięciu klastrów na podstawie współdzielonych wariantów częstych w analizie klastrowania populacji polskiej i europejskich.

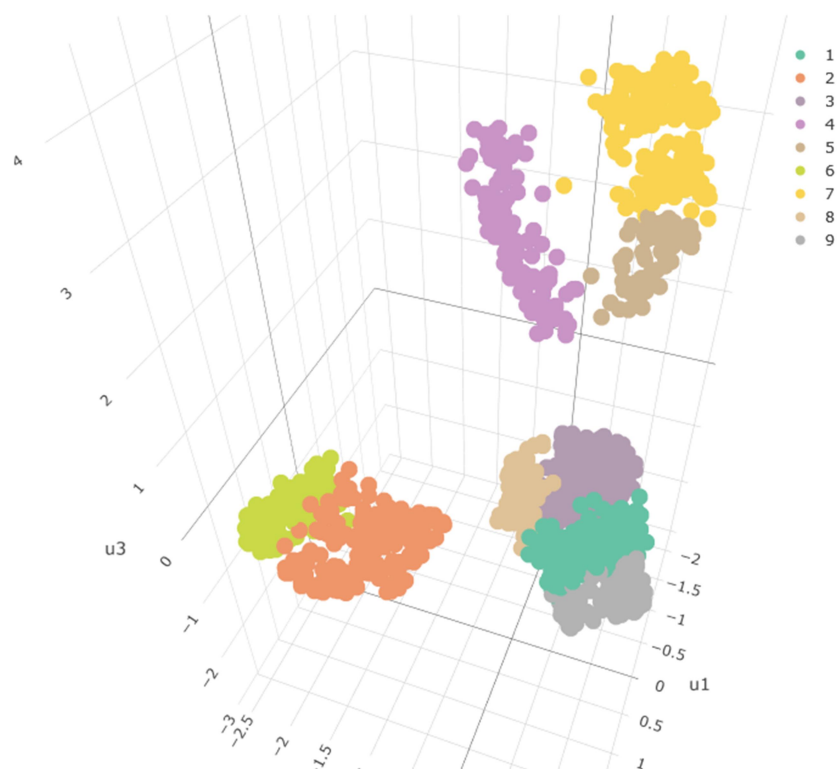
klaster	1	2	3	4	5	6	7	8	9
liczność	247	193	271	125	64	193	175	141	129



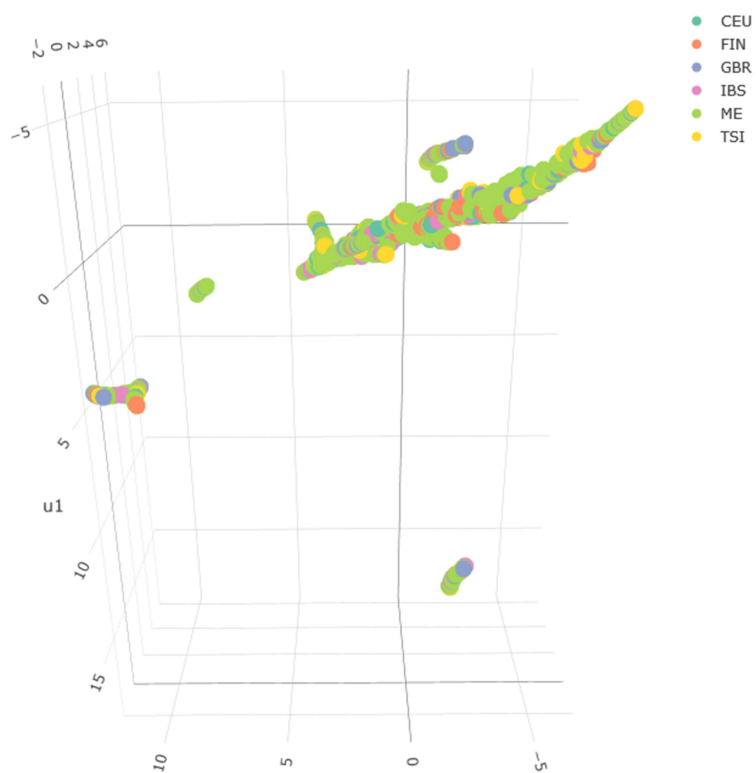
Rycina 2. Klastrowanie populacji polskiej i europejskich dla wszystkich współdzielonych wariantów.



Rycina 3. Klastrowanie populacji polskiej i europejskich dla współdzielonych wariantów częstych.



Rycina 4. Klastrowanie populacji polskiej i europejskich na dziewięć odrębnych klastrów na podstawie współdzielonych wariantów częstych.



Rycina 5. Klastrowanie populacji polskiej i europejskich dla współdzielonych wariantów rzadkich.

4.2.3. Analiza wariantów związanych z klastrowaniem

Na podstawie wyniku grupowania próbek, spośród wszystkich wariantów zidentyfikowano pulę 647 polimorfizmów istotnie związanych z dystrybucją próbek w poszczególnych dziewięciu klastrach. Co ciekawe, 33 warianty z tej puli stanowiły zmiany typu insercja/delecja. Wszystkie te warianty zlokalizowano w ośmiu genach kodujących białka ABC, na pięciu chromosomach (Tab. 5 i Tab. 6). Analiza nierównowagi sprzężeń dla poszczególnych chromosomów wykazała obecność bloków haplotypowych, które nie podlegały zdarzeniom rekombinacji i warianty w ich obrębie są współdziedziczone razem ($r^2 \geq 0,7$). W chromosomie 2 zaobserwowano dwa bloki (Ryc. 6); w chromosomie 7 – siedem bloków (Ryc. 7); w chromosomie 11 – dwa bloki (Ryc. 8); w chromosomie 13 – pięć bloków (Ryc. 9); w chromosomie 16 – trzy bloki (Ryc. 10).

ABCB11, *ABCG5* i *ABCG8*, zlokalizowane na tym samym chromosomie, skupiały więc w sumie tylko 11 takich wariantów czyli wielokrotnie mniej od pozostałych genów. Może być to jednak efekt niedostatecznego wzbogacenia w analizie akurat w tym *locus* o warianty sprzężone z pojedynczymi wariantami, zidentyfikowanymi jako istotne. Z kolei nadmiar dużej ilości wariantów w genach *ABCB5*, *ABCC4*, *ABCC8*, *ABCC11* przeciwnie, prawdopodobnie jest efektem wspólnego dziedziczenia w blokach haplotypowych dużej grupy wariantów sprzężonych z tymi rzeczywiście istotnie związanymi z klastrowaniem. Ilości wariantów w genie nie można więc bezpośrednio wiązać z istotnością jego wpływu na klastrowanie.

Trzy z tych genów, *ABCB11*, *ABCG5* i *ABCG8*, bezpośrednio związane są z homeostazą kwasów żółciowych, którą opisano we wstępie. Białko kodowane przez *ABCC4* związane jest z transportem ksenobiotyków, kwasów żółciowych i koniugatów glutationu, takich jak leukotrieny, ale i też szeregu endogennych związków, które odgrywają kluczową rolę w transdukcji sygnału komórkowego, jak cykliczny AMP (cAMP) i cykliczny GMP (cGMP) (Russel i in., 2008). Podobnie, białko kodowane przez *ABCC11* bierze udział w transporcie ksenobiotyków, kwasów żółciowych, sprzężonych steroidów, S-koniugatów glutationu oraz cyklicznych nukleotydów (Liu, 2019). Co ciekawe, *ABCC11* odpowiada też za funkcje wydzielnicze gruczołów apokrynowych i warunkuje fenotyp woskowiny usznej. Woskowina mokra występuje dominująco w populacjach wschodnioazjatyckich, a sucha w europejskich i afrykańskich (Yoshiura i in., 2006). *ABCC8* koduje podjednostkę współtworzącą kanał potasowy związany z uwalnianiem insuliny w trzustce, co szerzej opisano we wstępie. Niewiele natomiast jest wiadomo o funkcjach białek kodowanych przez *ABCB5* i *ABCC12*. Stwierdzono jedynie udział *ABCB5* w rozwoju i naprawie komórek rąbka rogówki (Ksander i in., 2014), ale najczęściej opisywano jego nadekspresję w komórkach czerniaka złośliwego, co może świadczyć o jego związku z melanogenezą (Lin i in., 2013; Setia i in., 2012). *ABCC12* z kolei powstał w wyniku duplikacji *ABCC11*, więc ich funkcje mogą być podobne, co sugeruje opisywany związek transportera *ABCC12* z idiopatyczną przewlekłą

cholestazą u dzieci i utratą wewnątrzwątrobowych dróg żółciowych w organizmach modelowych (Pham i in., 2021).

Tabela 5. Summaryczna ilość wariantów w genach kodujących białka ABC zidentyfikowanych jako istotnie związanych z klastrowaniem populacji europejskich.

chromosom	2	2	2	7	11	13	16	16
gen	<i>ABCB11</i>	<i>ABCG5</i>	<i>ABCG8</i>	<i>ABCB5</i>	<i>ABCC8</i>	<i>ABCC4</i>	<i>ABCC11</i>	<i>ABCC12</i>
warianty związane z klastrowaniem	6	2	3	266	97	125	146	2

Tabela 6. Lista wariantów zidentyfikowana jako istotnie związanych z grupowaniem próbek na dziewięć klastrow.

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
<i>ABCG5</i>				
NC_000002.12:g.43833693T>C	rs6716451	0,330	blok 1	1
NC_000002.12:g.43835757G>T	rs4953020	0,327	blok 1	2
<i>ABCG8</i>				
NC_000002.12:g.43842642A>T	rs4148206	0,400	blok 1	3
NC_000002.12:g.43844713G>C	rs4148213	0,400	blok 1	4
NC_000002.12:g.43867015C>T	rs149956799	0,256	-	5
<i>ABCB11</i>				
NC_000002.12:g.168919611T>A	rs580670	0,371	blok 2	6
NC_000002.12:g.168924308A>G	rs575671	0,456	blok 2	7
NC_000002.12:g.168937773G>A	rs565412	0,476	-	8
NC_000002.12:g.168943931A>T	rs853790	0,474	blok 2	9
NC_000002.12:g.168954234_168954235insT	rs34477999	0,485	blok 2	10
NC_000002.12:g.168956861G>A	rs853775	0,470	blok 2	11
<i>ABCB5</i>				
NC_000007.14:g.20615769C>T	rs73076550	0,085	blok 3	1
NC_000007.14:g.20616328G>T	rs117519129	0,153	blok 7	2
NC_000007.14:g.20616431G>A	rs7809674	0,315	blok 6	3
NC_000007.14:g.20616627C>G	rs74989588	0,077	blok 4	4
NC_000007.14:g.20616682T>C	rs79217266	0,077	blok 4	5
NC_000007.14:g.20617335A>C	rs12667934	0,085	blok 3	6
NC_000007.14:g.20617386T>G	rs55973507	0,077	blok 4	7
NC_000007.14:g.20618360G>C	rs77959491	0,152	blok 7	8
NC_000007.14:g.20619596G>T	rs4446638	0,316	blok 6	9
NC_000007.14:g.20619922T>A	rs12667658	0,316	blok 6	10
NC_000007.14:g.20621499C>G	rs79207698	0,153	blok 7	11
NC_000007.14:g.20621700C>T	rs56080217	0,316	blok 6	12
NC_000007.14:g.20622901A>G	rs78293595	0,077	blok 4	13
NC_000007.14:g.20623446C>T	rs111872870	0,077	blok 4	14
NC_000007.14:g.20623447A>T	rs77363367	0,153	blok 7	15
NC_000007.14:g.20623473C>T	rs75494098	0,077	blok 4	16
NC_000007.14:g.20624007A>G	rs78057207	0,153	blok 7	17
NC_000007.14:g.20624061G>C	rs7777864	0,256	blok 6	18
NC_000007.14:g.20625864G>A	rs11773775	0,077	blok 4	19
NC_000007.14:g.20625917G>A	rs75558592	0,153	blok 7	20
NC_000007.14:g.20626009T>C	rs75285484	0,153	blok 7	21
NC_000007.14:g.20626158A>G	rs73684577	0,230	blok 6	22
NC_000007.14:g.20626159T>A	rs73684578	0,230	blok 6	23
NC_000007.14:g.20626965T>G	rs75495305	0,077	blok 4	24
NC_000007.14:g.20627189G>A	rs12154301	0,152	blok 7	25
NC_000007.14:g.20627690dupT	rs34730550	0,358	blok 6	26
NC_000007.14:g.20627804A>G	rs76568631	0,077	blok 4	27

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000007.14:g.20628654C>A	rs78414512	0,152	blok 7	28
NC_000007.14:g.20629010A>C	rs11543624	0,077	blok 4	29
NC_000007.14:g.20629261T>G	rs76522145	0,153	blok 7	30
NC_000007.14:g.20629785G>A	rs187883871	0,077	blok 4	31
NC_000007.14:g.20629787delGAA	rs35282965	0,085	blok 3	32
NC_000007.14:g.20629930C>G	rs1476484	0,316	blok 6	33
NC_000007.14:g.20630006T>C	rs1476485	0,164	blok 5	34
NC_000007.14:g.20630075delC	rs35608196	0,164	blok 5	35
NC_000007.14:g.20630720G>A	rs16872578	0,300	blok 6	36
NC_000007.14:g.20630783T>C	rs13309246	0,085	blok 3	37
NC_000007.14:g.20630849_20630857delAATTTTCT	rs57419347	0,164	blok 5	38
NC_000007.14:g.20631100G>A	rs17143197	0,086	blok 3	39
NC_000007.14:g.20631507T>C	rs17804960	0,137	blok 7	40
NC_000007.14:g.20631625A>G	rs79682588	0,137	blok 7	41
NC_000007.14:g.20631658A>T	rs75353400	0,077	blok 4	42
NC_000007.14:g.20633098C>T	rs6963630	0,163	blok 5	43
NC_000007.14:g.20633111T>C	rs6945713	0,163	blok 5	44
NC_000007.14:g.20633375G>A	rs13247791	0,085	blok 3	45
NC_000007.14:g.20633620G>A	rs6964702	0,300	blok 6	46
NC_000007.14:g.20634114T>A	rs34322449	0,300	blok 6	47
NC_000007.14:g.20634420A>G	rs55653237	0,300	blok 6	48
NC_000007.14:g.20634505G>A	rs34259677	0,300	blok 6	49
NC_000007.14:g.20635145C>T	rs6946005	0,300	blok 6	50
NC_000007.14:g.20635215G>T	rs6946293	0,086	blok 3	51
NC_000007.14:g.20635258C>A	rs6946164	0,086	blok 3	52
NC_000007.14:g.20635290A>G	rs6946054	0,300	blok 6	53
NC_000007.14:g.20635393T>G	rs56144764	0,077	blok 4	54
NC_000007.14:g.20635414T>C	rs6966528	0,137	blok 7	55
NC_000007.14:g.20635672G>A	rs67238775	0,085	blok 3	56
NC_000007.14:g.20635840C>A	rs75277359	0,137	blok 7	57
NC_000007.14:g.20636287G>T	rs13238912	0,086	blok 3	58
NC_000007.14:g.20636467T>G	rs13226131	0,222	blok 7	59
NC_000007.14:g.20636658A>C	rs111343887	0,077	blok 4	60
NC_000007.14:g.20637196T>C	rs79259809	0,137	blok 7	61
NC_000007.14:g.20637361C>G	rs77726188	0,077	blok 4	62
NC_000007.14:g.20637933G>A	rs74565940	0,137	blok 7	63
NC_000007.14:g.20638133G>A	rs35484201	0,085	blok 3	64
NC_000007.14:g.20638374_20638378delATTTA	rs140658888	0,077	blok 4	65
NC_000007.14:g.20638771G>T	rs2893005	0,163	blok 5	66
NC_000007.14:g.20639090C>T	rs6461510	0,086	blok 3	67
NC_000007.14:g.20639222G>T	rs6967789	0,086	blok 3	68
NC_000007.14:g.20639325T>A	rs76285172	0,137	blok 7	69
NC_000007.14:g.20639632C>G	rs35613245	0,085	blok 3	70
NC_000007.14:g.20639663G>A	rs2188594	0,164	blok 5	71
NC_000007.14:g.20639848A>G	rs11761598	0,077	blok 4	72
NC_000007.14:g.20640384T>C	rs1004060	0,223	blok 7	73
NC_000007.14:g.20640426A>G	rs77923898	0,077	blok 4	74
NC_000007.14:g.20640507G>A	rs77072593	0,077	blok 4	75
NC_000007.14:g.20640699T>C	rs11772341	0,077	blok 4	76
NC_000007.14:g.20640730T>C	rs11773053	0,077	blok 4	77
NC_000007.14:g.20640928A>C	rs12700225	0,085	blok 3	78
NC_000007.14:g.20640960T>C	rs7793193	0,300	blok 6	79
NC_000007.14:g.20641012T>C	rs34219383	0,085	blok 3	80
NC_000007.14:g.20641930A>T	rs1004226	0,085	blok 3	81
NC_000007.14:g.20642119A>T	rs76711429	0,137	blok 7	82
NC_000007.14:g.20642566A>G	rs73688305	0,216	blok 6	83
NC_000007.14:g.20642890A>C	rs7783959	0,203	blok 6	84
NC_000007.14:g.20643581G>T	rs7349983	0,125	blok 7	85
NC_000007.14:g.20643824A>G	rs12669250	0,080	blok 3	86
NC_000007.14:g.20644137G>A	rs117557136	0,125	blok 7	87
NC_000007.14:g.20644316C>T	rs74547036	0,077	blok 4	88
NC_000007.14:g.20644673A>C	rs6957205	0,206	blok 7	89

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000007.14:g.20644712C>A	rs6957382	0,080	blok 3	90
NC_000007.14:g.20645628A>G	rs17216606	0,080	blok 3	91
NC_000007.14:g.20645885T>G	rs11983326	0,283	blok 6	92
NC_000007.14:g.20646163G>A	rs11769236	0,202	blok 6	93
NC_000007.14:g.20646277T>C	rs11772926	0,203	blok 6	94
NC_000007.14:g.20646488A>C	rs1858948	0,300	blok 6	95
NC_000007.14:g.20647558C>T	rs2893006	0,202	blok 6	96
NC_000007.14:g.20647709C>T	rs75784515	0,201	blok 6	97
NC_000007.14:g.20647981A>G	rs61732039	0,125	blok 7	98
NC_000007.14:g.20648419C>G	rs6949755	0,299	blok 6	99
NC_000007.14:g.20648486G>C	rs6950224	0,282	blok 6	100
NC_000007.14:g.20648512G>A	rs6950237	0,282	blok 6	101
NC_000007.14:g.20648519C>T	rs6949934	0,080	blok 3	102
NC_000007.14:g.20648538G>A	rs6950370	0,282	blok 6	103
NC_000007.14:g.20648828T>C	rs6974437	0,080	blok 3	104
NC_000007.14:g.20649392C>T	rs55663200	0,077	blok 4	105
NC_000007.14:g.20649442A>C	rs6955287	0,283	blok 6	106
NC_000007.14:g.20649797A>C	rs916737	0,203	blok 6	107
NC_000007.14:g.20649872C>T	rs916738	0,159	blok 5	108
NC_000007.14:g.20650312C>A	rs13244876	0,080	blok 3	109
NC_000007.14:g.20650314T>A	rs11973722	0,159	blok 5	110
NC_000007.14:g.20650326G>T	rs12674033	0,080	blok 3	111
NC_000007.14:g.20650528T>C	rs10950825	0,159	blok 5	112
NC_000007.14:g.20650550C>T	rs10950826	0,159	blok 5	113
NC_000007.14:g.20650957C>T	rs76160218	0,125	blok 7	114
NC_000007.14:g.20650991G>T	rs78181077	0,125	blok 7	115
NC_000007.14:g.20651153A>G	rs77707831	0,078	blok 4	116
NC_000007.14:g.20651218T>C	rs75047349	0,125	blok 7	117
NC_000007.14:g.20651236G>A	rs80029421	0,125	blok 7	118
NC_000007.14:g.20651262T>C	rs3213622	0,283	blok 6	119
NC_000007.14:g.20651361A>G	rs79943066	0,125	blok 7	120
NC_000007.14:g.20651390C>T	rs3213623	0,159	blok 5	121
NC_000007.14:g.20651424T>C	rs34603556	0,202	blok 6	122
NC_000007.14:g.20651596G>A	rs61227829	0,125	blok 7	123
NC_000007.14:g.20651907delCTT	rs200280589	0,077	blok 4	124
NC_000007.14:g.20652237C>A	rs17816709	0,202	blok 6	125
NC_000007.14:g.20652305C>T	rs138002473	0,125	blok 7	126
NC_000007.14:g.20652324G>T	rs2285554	0,283	blok 6	127
NC_000007.14:g.20652377C>G	rs2285555	0,202	blok 6	128
NC_000007.14:g.20652393A>C	rs2285556	0,202	blok 6	129
NC_000007.14:g.20652624C>A	rs79719780	0,125	blok 7	130
NC_000007.14:g.20652761C>A	rs77926987	0,202	blok 6	131
NC_000007.14:g.20652790T>G	rs2285557	0,202	blok 6	132
NC_000007.14:g.20652909C>T	rs2285558	0,283	blok 6	133
NC_000007.14:g.20652945_20652946insA	rs138475823	0,125	blok 7	134
NC_000007.14:g.20653014A>C	rs2285559	0,202	blok 6	135
NC_000007.14:g.20654189delT	rs149803992	0,125	blok 7	136
NC_000007.14:g.20654224T>C	rs57779704	0,125	blok 7	137
NC_000007.14:g.20654452_20654453insTTC	rs56740661	0,125	blok 7	138
NC_000007.14:g.20654535C>G	rs58835955	0,125	blok 7	139
NC_000007.14:g.20654645T>C	rs73684658	0,125	blok 7	140
NC_000007.14:g.20654866G>A	rs6958493	0,158	blok 5	141
NC_000007.14:g.20654879G>C	rs6958496	0,080	blok 3	142
NC_000007.14:g.20655022C>G	rs79760827	0,125	blok 7	143
NC_000007.14:g.20655102A>G	rs13243492	0,283	blok 6	144
NC_000007.14:g.20655112C>T	rs6958596	0,158	blok 5	145
NC_000007.14:g.20655203C>A	rs12700226	0,159	blok 5	146
NC_000007.14:g.20655282G>T	rs2097681	0,125	blok 7	147
NC_000007.14:g.20655333T>C	rs2078905	0,284	blok 6	148
NC_000007.14:g.20655798C>T	rs7796207	0,159	blok 5	149
NC_000007.14:g.20655845delA	rs34183859	0,159	blok 5	150
NC_000007.14:g.20656075A>T	rs7796543	0,302	blok 6	151

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000007.14:g.20656096_20656097insA	rs34900454	0,157	blok 5	152
NC_000007.14:g.20656111T>C	rs7779761	0,284	blok 6	153
NC_000007.14:g.20656236G>A	rs78982843	0,125	blok 7	154
NC_000007.14:g.20656279C>T	rs35359563	0,284	blok 6	155
NC_000007.14:g.20656425C>T	rs58797611	0,159	blok 5	156
NC_000007.14:g.20656480G>A	rs67368461	0,284	blok 6	157
NC_000007.14:g.20656526dupT	rs35911127	0,159	blok 5	158
NC_000007.14:g.20656551G>A	rs68180882	0,158	blok 5	159
NC_000007.14:g.20656695delT	rs35398667	0,284	blok 6	160
NC_000007.14:g.20656764T>G	rs35120147	0,159	blok 5	161
NC_000007.14:g.20656863C>T	rs73278668	0,078	blok 4	162
NC_000007.14:g.20656966G>C	rs1079907	0,284	blok 6	163
NC_000007.14:g.20657121C>A	rs59488551	0,159	blok 5	164
NC_000007.14:g.20657289T>C	rs6951722	0,125	blok 7	165
NC_000007.14:g.20657326_20657327insA	rs34999687	0,158	blok 5	166
NC_000007.14:g.20657343A>G	rs7802093	0,159	blok 5	167
NC_000007.14:g.20657416C>T	rs34125782	0,080	blok 3	168
NC_000007.14:g.20657471T>C	rs7785407	0,284	blok 6	169
NC_000007.14:g.20657732C>G	rs7806279	0,203	blok 6	170
NC_000007.14:g.20657908A>G	rs17143253	0,159	blok 5	171
NC_000007.14:g.20658371C>T	rs76913103	0,125	blok 7	172
NC_000007.14:g.20658454C>G	rs7807360	0,203	blok 6	173
NC_000007.14:g.20658647A>G	rs2301641	0,306	blok 6	174
NC_000007.14:g.20659181C>A	rs73278683	0,204	blok 6	175
NC_000007.14:g.20659393C>T	rs73278684	0,126	blok 7	176
NC_000007.14:g.20659403delT	rs142562743	0,204	blok 6	177
NC_000007.14:g.20659435A>G	rs73684672	0,078	blok 4	178
NC_000007.14:g.20659738C>T	rs73278686	0,126	blok 7	179
NC_000007.14:g.20659803C>T	rs112762761	0,126	blok 7	180
NC_000007.14:g.20659856G>C	rs11763916	0,078	blok 4	181
NC_000007.14:g.20660000A>T	rs11770673	0,078	blok 4	182
NC_000007.14:g.20660043C>G	rs6947250	0,204	blok 6	183
NC_000007.14:g.20660056C>T	rs11763925	0,078	blok 4	184
NC_000007.14:g.20660299T>G	rs6967746	0,204	blok 6	185
NC_000007.14:g.20660433A>G	rs11770822	0,078	blok 4	186
NC_000007.14:g.20660465C>T	rs11764054	0,079	blok 4	187
NC_000007.14:g.20660490C>T	rs11764057	0,078	blok 4	188
NC_000007.14:g.20660494A>G	rs58540449	0,078	blok 4	189
NC_000007.14:g.20660523G>A	rs60844047	0,078	blok 4	190
NC_000007.14:g.20660567T>C	rs13223438	0,097	blok 3	191
NC_000007.14:g.20660624G>A	rs6948334	0,126	blok 7	192
NC_000007.14:g.20660703T>C	rs58976477	0,078	blok 4	193
NC_000007.14:g.20660738A>C	rs6952128	0,302	blok 6	194
NC_000007.14:g.20660890T>A	rs73684676	0,078	blok 4	195
NC_000007.14:g.20660904A>G	rs73684678	0,078	blok 4	196
NC_000007.14:g.20660991G>A	rs5011447	0,078	blok 4	197
NC_000007.14:g.20661065A>T	rs2893007	0,078	blok 4	198
NC_000007.14:g.20661258G>T	rs2390348	0,204	blok 6	199
NC_000007.14:g.20661462G>T	rs74809645	0,078	blok 4	200
NC_000007.14:g.20661948G>T	rs75731479	0,078	blok 4	201
NC_000007.14:g.20662210delAG	rs140558196	0,078	blok 4	202
NC_000007.14:g.20662578G>C	rs7792341	0,206	blok 6	203
NC_000007.14:g.20662591delA	rs141084193	0,078	blok 4	204
NC_000007.14:g.20662671T>C	rs77663125	0,078	blok 4	205
NC_000007.14:g.20662939T>G	rs75171064	0,078	blok 4	206
NC_000007.14:g.20663238C>G	rs12154879	0,128	blok 7	207
NC_000007.14:g.20663549T>C	rs66744319	0,098	blok 3	208
NC_000007.14:g.20663586T>C	rs12155194	0,206	blok 6	209
NC_000007.14:g.20663855A>G	rs111905715	0,078	blok 4	210
NC_000007.14:g.20663963G>T	rs146448884	0,128	blok 7	211
NC_000007.14:g.20663977A>G	rs13223600	0,097	blok 3	212
NC_000007.14:g.20664035T>C	rs13238268	0,175	blok 5	213

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000007.14:g.20664160A>G	rs7801455	0,303	blok 6	214
NC_000007.14:g.20664325A>C	rs1011559	0,128	blok 7	215
NC_000007.14:g.20664500T>C	rs1011560	0,123	blok 7	216
NC_000007.14:g.20664514G>A	rs916739	0,173	blok 5	217
NC_000007.14:g.20664627C>T	rs916740	0,174	blok 5	218
NC_000007.14:g.20664648C>G	rs916741	0,174	blok 5	219
NC_000007.14:g.20664821T>G	rs1011561	0,125	blok 7	220
NC_000007.14:g.20665257C>T	rs6949607	0,121	blok 7	221
NC_000007.14:g.20665655C>T	rs1059006	0,175	blok 5	222
NC_000007.14:g.20665674C>A	rs73280614	0,119	blok 7	223
NC_000007.14:g.20665896C>G	rs4476902	0,121	blok 2	224
NC_000007.14:g.20665972C>G	rs10950827	0,176	blok 5	225
NC_000007.14:g.20667278A>T	rs12540337	0,120	blok 2	226
NC_000007.14:g.20667621A>C	rs28656113	0,298	blok 1	227
NC_000007.14:g.20667794G>A	rs28603008	0,297	blok 1	228
NC_000007.14:g.20667846C>G	rs370062232	0,175	blok 5	229
NC_000007.14:g.20670240T>C	rs62456151	0,296	blok 1	230
NC_000007.14:g.20670404C>T	rs1548642	0,121	blok 2	231
NC_000007.14:g.20670686C>G	rs1972525	0,121	blok 2	232
NC_000007.14:g.20670766A>G	rs7455102	0,121	blok 2	233
NC_000007.14:g.20670841G>A	rs10266328	0,298	blok 1	234
NC_000007.14:g.20670863C>T	rs6943237	0,121	blok 2	235
NC_000007.14:g.20671022delAAAG	rs111363721	0,121	blok 7	236
NC_000007.14:g.20671536A>G	rs12700228	0,175	blok 5	237
NC_000007.14:g.20671622C>T	rs1859777	0,297	blok 1	238
NC_000007.14:g.20671635T>C	rs13220972	0,175	blok 5	239
NC_000007.14:g.20672085A>T	rs7782507	0,121	blok 2	240
NC_000007.14:g.20672670T>C	rs6970012	0,121	blok 2	241
NC_000007.14:g.20672872A>G	rs13240297	0,395	-	242
NC_000007.14:g.20673296C>G	rs10275105	0,121	blok 2	243
NC_000007.14:g.20673823G>C	rs2097941	0,121	blok 2	244
NC_000007.14:g.20675024G>A	rs7806967	0,121	blok 2	245
NC_000007.14:g.20675510C>T	rs2158852	0,125	blok 2	246
NC_000007.14:g.20676586A>C	rs10236314	0,124	blok 2	247
NC_000007.14:g.20676621T>C	rs10281505	0,125	blok 2	248
NC_000007.14:g.20676693G>A	rs10266137	0,301	blok 1	249
NC_000007.14:g.20677634C>G	rs34258758	0,174	blok 5	250
NC_000007.14:g.20679039G>A	rs4721933	0,120	blok 2	251
NC_000007.14:g.20679627_20679634delTGAGAAAA	rs138864104	0,175	blok 5	252
NC_000007.14:g.20679778A>C	rs7795777	0,120	blok 2	253
NC_000007.14:g.20681461T>C	rs11764760	0,177	blok 5	254
NC_000007.14:g.20681935G>C	rs6974315	0,140	blok 2	255
NC_000007.14:g.20682257G>C	rs61529054	0,112	blok 7	256
NC_000007.14:g.20682929A>T	rs2158851	0,123	blok 2	257
NC_000007.14:g.20683439C>T	rs7812159	0,229	blok 5	258
NC_000007.14:g.20683469C>A	rs7812181	0,139	blok 2	259
NC_000007.14:g.20683498T>C	rs10215589	0,292	blok 1	260
NC_000007.14:g.20685203G>C	rs17817117	0,304	blok 6	261
NC_000007.14:g.20686625C>T	rs76022765	0,059	-	262
NC_000007.14:g.20688302T>C	rs67937303	0,179	blok 5	263
NC_000007.14:g.20688403G>A	rs77320744	0,051	-	264
NC_000007.14:g.20689035A>G	rs2214917	0,171	blok 5	265
NC_000007.14:g.20689204G>C	rs2190411	0,244	-	266
ABCC8				
NC_000011.10:g.17406010T>C	rs739688	0,371	-	1
NC_000011.10:g.17409101C>T	rs10766397	0,357	-	2
NC_000011.10:g.17424391T>C	rs4281466	0,457	-	3
NC_000011.10:g.17437560G>A	rs11024286	0,330	-	4
NC_000011.10:g.17437571A>T	rs17775256	0,346	blok 2	5
NC_000011.10:g.17438096A>G	rs3781950	0,401	blok 2	6
NC_000011.10:g.17438382_17438383insCACA	rs3831596	0,400	blok 2	7

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000011.10:g.17438384T>C	rs60105962	0,400	blok 2	8
NC_000011.10:g.17438537T>A	rs12417170	0,400	blok 2	9
NC_000011.10:g.17438634A>G	rs10832788	0,400	blok 2	10
NC_000011.10:g.17438732A>G	rs2237983	0,400	blok 2	11
NC_000011.10:g.17438755T>C	rs10832789	0,400	blok 2	12
NC_000011.10:g.17439165C>T	rs2237982	0,400	blok 2	13
NC_000011.10:g.17439352T>C	rs2237981	0,400	blok 2	14
NC_000011.10:g.17439782C>G	rs2237980	0,400	blok 2	15
NC_000011.10:g.17439996T>C	rs2237979	0,345	blok 2	16
NC_000011.10:g.17440757A>C	rs2283261	0,401	blok 2	17
NC_000011.10:g.17441180G>A	rs757109	0,401	blok 2	18
NC_000011.10:g.17441662C>T	rs2237978	0,344	blok 2	19
NC_000011.10:g.17442987T>C	rs2073587	0,402	blok 2	20
NC_000011.10:g.17444498A>C	rs886292	0,349	blok 2	21
NC_000011.10:g.17444721A>C	rs7117162	0,349	blok 2	22
NC_000011.10:g.17444976A>G	rs2040653	0,348	blok 2	23
NC_000011.10:g.17445441T>G	rs10766399	0,350	blok 2	24
NC_000011.10:g.17445506T>G	rs10766400	0,349	blok 2	25
NC_000011.10:g.17445656T>C	rs11024290	0,402	blok 2	26
NC_000011.10:g.17445790T>C	rs11821723	0,349	blok 2	27
NC_000011.10:g.17446770T>C	rs7950189	0,478	blok 1	28
NC_000011.10:g.17446946A>C	rs7947326	0,349	blok 2	29
NC_000011.10:g.17447065G>T	rs7932043	0,480	blok 1	30
NC_000011.10:g.17447558T>C	rs12794627	0,349	blok 2	31
NC_000011.10:g.17447919T>C	rs4459286	0,400	blok 2	32
NC_000011.10:g.17449037A>G	rs7934185	0,402	blok 2	33
NC_000011.10:g.17450455T>C	rs12786125	0,400	blok 2	34
NC_000011.10:g.17451875T>C	rs2283259	0,399	blok 2	35
NC_000011.10:g.17451890C>T	rs2283258	0,309	blok 2	36
NC_000011.10:g.17452771T>G	rs2057661	0,400	blok 2	37
NC_000011.10:g.17453736G>A	rs2073586	0,394	blok 2	38
NC_000011.10:g.17453896G>A	rs2073585	0,397	blok 2	39
NC_000011.10:g.17454841C>T	rs2237977	0,393	blok 2	40
NC_000011.10:g.17454969T>G	rs2237976	0,399	blok 2	41
NC_000011.10:g.17456283C>T	rs2237975	0,347	blok 2	42
NC_000011.10:g.17456610C>T	rs2237974	0,343	blok 2	43
NC_000011.10:g.17459288T>C	rs7105391	0,347	blok 2	44
NC_000011.10:g.17460401T>C	rs886291	0,348	blok 2	45
NC_000011.10:g.17460781T>C	rs886290	0,347	blok 2	46
NC_000011.10:g.17461072A>G	rs2074319	0,348	blok 2	47
NC_000011.10:g.17462767T>A	rs7943461	0,348	blok 2	48
NC_000011.10:g.17462799_17462800insAG	rs36065130	0,400	blok 2	49
NC_000011.10:g.17463081C>G	rs4148610	0,348	blok 2	50
NC_000011.10:g.17463184C>T	rs4148609	0,348	blok 2	51
NC_000011.10:g.17463234C>G	rs3815066	0,348	blok 2	52
NC_000011.10:g.17463424G>A	rs2301703	0,400	blok 2	53
NC_000011.10:g.17465081T>G	rs3819524	0,344	blok 2	54
NC_000011.10:g.17465097G>C	rs3819523	0,344	blok 2	55
NC_000011.10:g.17465131T>C	rs3819522	0,349	blok 2	56
NC_000011.10:g.17465190C>T	rs3819521	0,344	blok 2	57
NC_000011.10:g.17465306A>G	rs3819520	0,346	blok 2	58
NC_000011.10:g.17465421C>G	rs3819519	0,400	blok 2	59
NC_000011.10:g.17465532C>T	rs3819518	0,344	blok 2	60
NC_000011.10:g.17465731C>A	rs3819517	0,400	blok 2	61
NC_000011.10:g.17465802G>A	rs12285976	0,344	blok 2	62
NC_000011.10:g.17465898C>T	rs11024295	0,399	blok 2	63
NC_000011.10:g.17466260G>A	rs10832791	0,344	blok 2	64
NC_000011.10:g.17466286T>C	rs10766401	0,400	blok 2	65
NC_000011.10:g.17466340G>A	rs10766402	0,400	blok 2	66
NC_000011.10:g.17466478A>G	rs10766403	0,400	blok 2	67
NC_000011.10:g.17466554C>T	rs11024296	0,344	blok 2	68
NC_000011.10:g.17466864G>C	rs12803384	0,398	blok 2	69

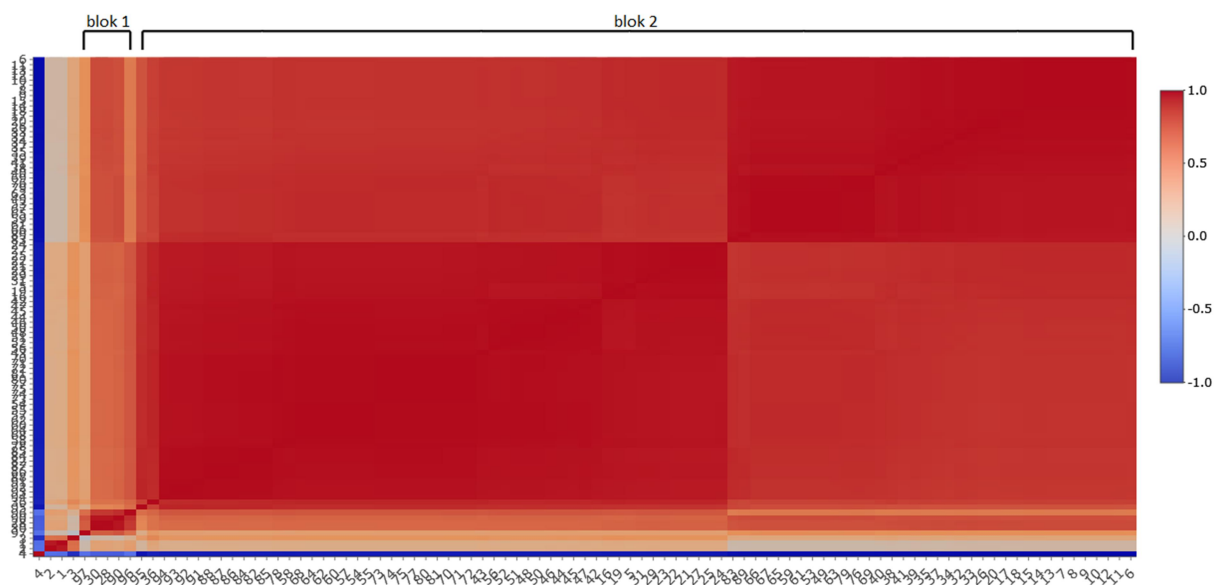
Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000011.10:g.17467114A>C	rs7105832	0,343	blok 2	70
NC_000011.10:g.17467139G>C	rs7119716	0,343	blok 2	71
NC_000011.10:g.17467156C>T	rs7120682	0,343	blok 2	72
NC_000011.10:g.17467559C>T	rs11826760	0,343	blok 2	73
NC_000011.10:g.17467720T>C	rs4148608	0,343	blok 2	74
NC_000011.10:g.17467732C>T	rs4148607	0,343	blok 2	75
NC_000011.10:g.17467743T>G	rs4148606	0,398	blok 2	76
NC_000011.10:g.17467746T>C	rs4148605	0,343	blok 2	77
NC_000011.10:g.17467790G>C	rs4148604	0,345	blok 2	78
NC_000011.10:g.17467816T>C	rs4148603	0,398	blok 2	79
NC_000011.10:g.17468145G>A	rs4148602	0,343	blok 2	80
NC_000011.10:g.17468717T>C	rs2283256	0,343	blok 2	81
NC_000011.10:g.17468954C>T	rs7106250	0,338	blok 2	82
NC_000011.10:g.17469024C>T	rs2237973	0,389	blok 2	83
NC_000011.10:g.17469326C>T	rs11024297	0,338	blok 2	84
NC_000011.10:g.17469582T>A	rs2237971	0,338	blok 2	85
NC_000011.10:g.17469652G>T	rs2237970	0,339	blok 2	86
NC_000011.10:g.17469789T>G	rs2237969	0,339	blok 2	87
NC_000011.10:g.17469839A>T	rs2237968	0,339	blok 2	88
NC_000011.10:g.17470024C>T	rs10400389	0,395	blok 2	89
NC_000011.10:g.17471117T>C	rs2237967	0,475	blok 1	90
NC_000011.10:g.17471635A>G	rs2237966	0,342	blok 2	91
NC_000011.10:g.17472277T>C	rs11024299	0,342	blok 2	92
NC_000011.10:g.17472515A>T	rs4148601	0,342	blok 2	93
NC_000011.10:g.17473633G>A	rs2283254	0,342	blok 2	94
NC_000011.10:g.17473883G>T	rs2283253	0,386	blok 2	95
NC_000011.10:g.17474969A>G	rs1048099	0,485	blok 1	96
NC_000011.10:g.17476247C>G	rs985136	0,497	blok 1	97
ABCC4				
NC_000013.11:g.95045626C>T	rs1151471	0,349	blok 4	1
NC_000013.11:g.95051721dupC	rs34749414	0,350	blok 4	2
NC_000013.11:g.95052023C>T	rs1189450	0,374	blok 4	3
NC_000013.11:g.95054888C>T	rs1189445	0,350	blok 4	4
NC_000013.11:g.95057141A>G	rs1189442	0,179	blok 3	5
NC_000013.11:g.95059578G>A	rs1751031	0,177	blok 3	6
NC_000013.11:g.95060813C>T	rs1612388	0,176	blok 3	7
NC_000013.11:g.95061256A>G	rs1630901	0,176	blok 3	8
NC_000013.11:g.95062604A>C	rs1751033	0,176	blok 3	9
NC_000013.11:g.95062722C>T	rs1751034	0,177	blok 3	10
NC_000013.11:g.95069727A>G	rs1189452	0,176	blok 3	11
NC_000013.11:g.95069832G>A	rs1189451	0,176	blok 3	12
NC_000013.11:g.95070105A>G	rs9524790	0,176	blok 3	13
NC_000013.11:g.95070428G>T	rs2619311	0,321	blok 4	14
NC_000013.11:g.95070785C>T	rs2619312	0,176	blok 3	15
NC_000013.11:g.95103327T>C	rs76658873	0,116	-	16
NC_000013.11:g.95179748T>G	rs4148493	0,361	-	17
NC_000013.11:g.95186530G>T	rs2148529	0,139	blok 1	18
NC_000013.11:g.95192701T>C	rs1678377	0,138	blok 1	19
NC_000013.11:g.95193351T>G	rs1751014	0,138	blok 1	20
NC_000013.11:g.95196413T>C	rs9516530	0,289	-	21
NC_000013.11:g.95203246T>C	rs4773844	0,225	-	22
NC_000013.11:g.95205881G>A	rs1617844	0,138	blok 1	23
NC_000013.11:g.95206432G>A	rs3818493	0,372	blok 2	24
NC_000013.11:g.95206450G>C	rs3818494	0,375	blok 2	25
NC_000013.11:g.95206742T>C	rs2274406	0,375	blok 2	26
NC_000013.11:g.95209550A>G	rs899494	0,162	blok 1	27
NC_000013.11:g.95209727C>G	rs1678403	0,188	blok 1	28
NC_000013.11:g.95209743G>T	rs2389203	0,201	blok 1	29
NC_000013.11:g.95209942A>G	rs899495	0,188	blok 1	30
NC_000013.11:g.95210155A>T	rs943288	0,164	blok 1	31
NC_000013.11:g.95210642A>G	rs899496	0,187	blok 1	32

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000013.11:g.95210889A>G	rs899497	0,188	blok 1	33
NC_000013.11:g.95210980A>G	rs4148471	0,187	blok 1	34
NC_000013.11:g.95211027C>T	rs4148470	0,187	blok 1	35
NC_000013.11:g.95211104A>G	rs4148469	0,187	blok 1	36
NC_000013.11:g.95211192T>C	rs3782963	0,187	blok 1	37
NC_000013.11:g.95211251A>G	rs3782965	0,187	blok 1	38
NC_000013.11:g.95211405G>T	rs3782966	0,187	blok 1	39
NC_000013.11:g.95211517A>G	rs3782968	0,187	blok 1	40
NC_000013.11:g.95211526A>G	rs1678400	0,164	blok 1	41
NC_000013.11:g.95211901C>T	rs9524828	0,187	blok 1	42
NC_000013.11:g.95211906G>A	rs9516537	0,187	blok 1	43
NC_000013.11:g.95212147T>C	rs9516538	0,187	blok 1	44
NC_000013.11:g.95212224G>A	rs7332663	0,187	blok 1	45
NC_000013.11:g.95212278G>A	rs7332836	0,187	blok 1	46
NC_000013.11:g.95212279A>G	rs7334994	0,187	blok 1	47
NC_000013.11:g.95212554G>T	rs7333281	0,163	blok 1	48
NC_000013.11:g.95212764C>T	rs7335147	0,187	blok 1	49
NC_000013.11:g.95212918G>A	rs9516539	0,187	blok 1	50
NC_000013.11:g.95213065T>A	rs9524830	0,187	blok 1	51
NC_000013.11:g.95213118T>C	rs7322825	0,187	blok 1	52
NC_000013.11:g.95213321G>A	rs7139533	0,187	blok 1	53
NC_000013.11:g.95213645C>A	rs9524831	0,187	blok 1	54
NC_000013.11:g.95213813A>G	rs7983336	0,187	blok 1	55
NC_000013.11:g.95214213C>T	rs7987653	0,187	blok 1	56
NC_000013.11:g.95214326A>G	rs7988494	0,187	blok 1	57
NC_000013.11:g.95214360G>A	rs9524833	0,187	blok 1	58
NC_000013.11:g.95214524A>C	rs4334136	0,187	blok 1	59
NC_000013.11:g.95214788A>G	rs1824913	0,187	blok 1	60
NC_000013.11:g.95214830G>A	rs1824912	0,188	blok 1	61
NC_000013.11:g.95214981A>T	rs1824911	0,187	blok 1	62
NC_000013.11:g.95215148T>C	rs7999175	0,187	blok 1	63
NC_000013.11:g.95215375G>A	rs9524835	0,187	blok 1	64
NC_000013.11:g.95215378G>A	rs9524836	0,187	blok 1	65
NC_000013.11:g.95215755C>T	rs4773847	0,187	blok 1	66
NC_000013.11:g.95216076C>T	rs4148467	0,187	blok 1	67
NC_000013.11:g.95216726G>C	rs4771907	0,187	blok 1	68
NC_000013.11:g.95216877G>A	rs4148465	0,187	blok 1	69
NC_000013.11:g.95217048A>G	rs4148464	0,187	blok 1	70
NC_000013.11:g.95217154G>C	rs4148463	0,187	blok 1	71
NC_000013.11:g.95217498A>G	rs7321744	0,187	blok 1	72
NC_000013.11:g.95217768C>T	rs6492769	0,187	blok 1	73
NC_000013.11:g.95218578A>G	rs9524839	0,187	blok 1	74
NC_000013.11:g.95218668G>A	rs9524840	0,187	blok 1	75
NC_000013.11:g.95219697G>A	rs7321623	0,187	blok 1	76
NC_000013.11:g.95220157T>C	rs4771908	0,187	blok 1	77
NC_000013.11:g.95220189_95220193delTTATA	rs59941084	0,187	blok 1	78
NC_000013.11:g.95220261A>C	rs2766476	0,164	blok 1	79
NC_000013.11:g.95221140A>G	rs4773848	0,187	blok 1	80
NC_000013.11:g.95222529G>A	rs4773849	0,187	blok 1	81
NC_000013.11:g.95222688G>A	rs1926656	0,187	blok 1	82
NC_000013.11:g.95222702T>C	rs1926657	0,187	blok 1	83
NC_000013.11:g.95223264C>T	rs7325019	0,188	blok 1	84
NC_000013.11:g.95223360T>C	rs4612933	0,188	blok 1	85
NC_000013.11:g.95223418_95223419insTC	rs10648065	0,188	blok 1	86
NC_000013.11:g.95223779T>G	rs4371040	0,188	blok 1	87
NC_000013.11:g.95223847C>G	rs4303338	0,188	blok 1	88
NC_000013.11:g.95223883G>A	rs9516545	0,188	blok 1	89
NC_000013.11:g.95223894T>C	rs2892707	0,188	blok 1	90
NC_000013.11:g.95224521G>A	rs7330776	0,164	blok 5	91
NC_000013.11:g.95225631C>T	rs12428470	0,187	blok 1	92
NC_000013.11:g.95225852G>T	rs12854605	0,187	blok 1	93
NC_000013.11:g.95226168T>C	rs61965944	0,187	blok 1	94

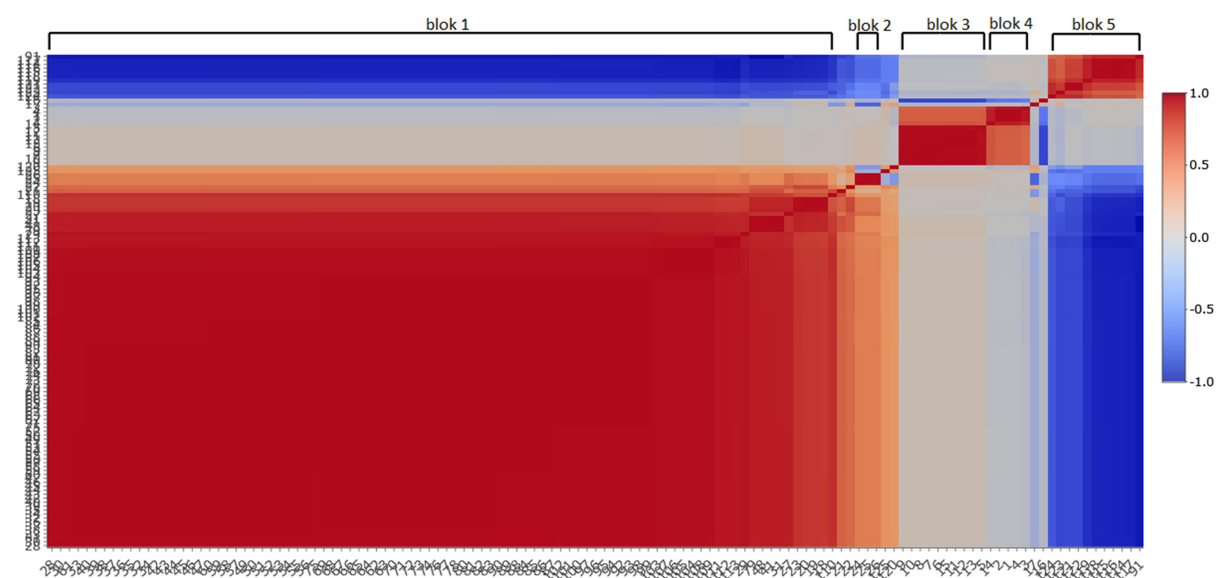
Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000013.11:g.95226387T>C	rs9524845	0,187	blok 1	95
NC_000013.11:g.95226588G>A	rs4636781	0,187	blok 1	96
NC_000013.11:g.95226598G>A	rs4505186	0,187	blok 1	97
NC_000013.11:g.95226686A>G	rs4520712	0,187	blok 1	98
NC_000013.11:g.95226963G>A	rs4773853	0,187	blok 1	99
NC_000013.11:g.95227156G>A	rs4773854	0,187	blok 1	100
NC_000013.11:g.95227187A>C	rs4773855	0,187	blok 1	101
NC_000013.11:g.95227587C>T	rs6492771	0,187	blok 1	102
NC_000013.11:g.95228501G>A	rs4238313	0,185	blok 1	103
NC_000013.11:g.95229016C>T	rs7324706	0,186	blok 1	104
NC_000013.11:g.95229077T>G	rs7335233	0,186	blok 1	105
NC_000013.11:g.95229118C>T	rs7329432	0,186	blok 1	106
NC_000013.11:g.95229486G>A	rs2389234	0,186	blok 1	107
NC_000013.11:g.95230558C>T	rs9524850	0,186	blok 1	108
NC_000013.11:g.95231794C>T	rs4773857	0,186	blok 1	109
NC_000013.11:g.95232158A>G	rs7326711	0,232	blok 1	110
NC_000013.11:g.95233088A>G	rs7333234	0,188	blok 1	111
NC_000013.11:g.95234022T>C	rs9524856	0,189	blok 1	112
NC_000013.11:g.95235733delG	rs4148457	0,191	blok 1	113
NC_000013.11:g.95236366delCTT	rs35347010	0,166	blok 5	114
NC_000013.11:g.95236987C>T	rs2892713	0,168	blok 5	115
NC_000013.11:g.95238110A>C	rs9561816	0,167	blok 5	116
NC_000013.11:g.95239162T>A	rs12430658	0,165	blok 5	117
NC_000013.11:g.95239763delT	rs35059505	0,166	blok 5	118
NC_000013.11:g.95243974T>C	rs11616679	0,142	blok 5	119
NC_000013.11:g.95244398C>T	rs4771909	0,422	-	120
NC_000013.11:g.95253051T>C	rs9590216	0,182	blok 5	121
NC_000013.11:g.95254045T>C	rs9590218	0,182	blok 5	122
NC_000013.11:g.95272074C>T	rs9561824	0,201	blok 5	123
NC_000013.11:g.95273879G>A	rs9561826	0,156	blok 5	124
NC_000013.11:g.95278431G>A	rs10508017	0,473	-	125
ABCC12				
NC_000016.10:g.48142336G>C	rs77309533	0,101	blok 3	1
NC_000016.10:g.48143151G>T	rs8049005	0,101	blok 3	2
ABCC11				
NC_000016.10:g.48167220C>G	rs41280939	0,095	blok 3	3
NC_000016.10:g.48198625A>G	rs7203695	0,233	-	4
NC_000016.10:g.48201666C>G	rs8050120	0,192	blok 2	5
NC_000016.10:g.48202755A>C	rs7206909	0,193	blok 2	6
NC_000016.10:g.48204140C>T	rs75797074	0,137	blok 1	7
NC_000016.10:g.48204206G>T	rs7187233	0,191	blok 2	8
NC_000016.10:g.48205972G>A	rs5017203	0,127	blok 1	9
NC_000016.10:g.48207176A>G	rs9932850	0,181	blok 2	10
NC_000016.10:g.48207637T>C	rs8046814	0,181	blok 2	11
NC_000016.10:g.48212065T>C	rs13336264	0,181	blok 2	12
NC_000016.10:g.48214457A>T	rs8057475	0,182	blok 2	13
NC_000016.10:g.48215007T>C	rs8047091	0,182	blok 2	14
NC_000016.10:g.48215814G>T	rs11863133	0,127	blok 1	15
NC_000016.10:g.48215938C>A	rs11866312	0,127	blok 1	16
NC_000016.10:g.48216100G>A	rs11863233	0,127	blok 1	17
NC_000016.10:g.48216115G>T	rs11863236	0,127	blok 1	18
NC_000016.10:g.48216307T>C	rs28654935	0,127	blok 1	19
NC_000016.10:g.48216310G>A	rs28376623	0,127	blok 1	20
NC_000016.10:g.48216336C>A	rs28626252	0,127	blok 1	21
NC_000016.10:g.48216525C>G	rs78716244	0,127	blok 1	22
NC_000016.10:g.48216650A>C	rs74481009	0,127	blok 1	23
NC_000016.10:g.48216918C>T	rs28684834	0,127	blok 1	24
NC_000016.10:g.48216931C>A	rs28503414	0,127	blok 1	25
NC_000016.10:g.48217224C>A	rs28505535	0,127	blok 1	26
NC_000016.10:g.48217461G>T	rs28696036	0,127	blok 1	27
NC_000016.10:g.48217564T>C	rs73540901	0,127	blok 1	28

Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000016.10:g.48217967G>A	rs73542903	0,127	blok 1	29
NC_000016.10:g.48217990A>G	rs73542904	0,127	blok 1	30
NC_000016.10:g.48218170C>T	rs112345293	0,181	blok 2	31
NC_000016.10:g.48218609T>G	rs13338524	0,206	blok 2	32
NC_000016.10:g.48219062C>T	rs13335504	0,128	blok 1	33
NC_000016.10:g.48219223T>G	rs13338757	0,182	blok 2	34
NC_000016.10:g.48219239A>G	rs28571687	0,128	blok 1	35
NC_000016.10:g.48219294T>C	rs28481895	0,182	blok 2	36
NC_000016.10:g.48219363A>G	rs59758263	0,182	blok 2	37
NC_000016.10:g.48219607A>T	rs73546403	0,182	blok 2	38
NC_000016.10:g.48219670T>A	rs76976241	0,128	blok 1	39
NC_000016.10:g.48219772G>A	rs113372282	0,128	blok 1	40
NC_000016.10:g.48219810G>A	rs13332872	0,128	blok 1	41
NC_000016.10:g.48219932G>C	rs9927515	0,128	blok 1	42
NC_000016.10:g.48220075A>G	rs9939742	0,128	blok 1	43
NC_000016.10:g.48220201G>A	rs9927756	0,128	blok 1	44
NC_000016.10:g.48220225C>T	rs9935854	0,128	blok 1	45
NC_000016.10:g.48221224G>T	rs13334480	0,128	blok 1	46
NC_000016.10:g.48221628C>T	rs9941123	0,128	blok 1	47
NC_000016.10:g.48221738delGT	rs71973142	0,128	blok 1	48
NC_000016.10:g.48221778A>G	rs9924823	0,128	blok 1	49
NC_000016.10:g.48222121G>A	rs74582156	0,127	blok 1	50
NC_000016.10:g.48222171C>T	rs28602441	0,128	blok 1	51
NC_000016.10:g.48222573A>G	rs16945972	0,128	blok 1	52
NC_000016.10:g.48222691T>C	rs16945974	0,128	blok 1	53
NC_000016.10:g.48223089G>A	rs7195606	0,128	blok 1	54
NC_000016.10:g.48223885T>C	rs28605399	0,128	blok 1	55
NC_000016.10:g.48224470T>A	rs73546424	0,127	blok 1	56
NC_000016.10:g.48224508C>A	rs9927238	0,127	blok 1	57
NC_000016.10:g.48224513T>C	rs9933821	0,127	blok 1	58
NC_000016.10:g.48224821G>A	rs113099408	0,127	blok 1	59
NC_000016.10:g.48224895C>T	rs117437537	0,126	blok 1	60
NC_000016.10:g.48225169C>T	rs113761362	0,127	blok 1	61
NC_000016.10:g.48225192G>A	rs7185951	0,127	blok 1	62
NC_000016.10:g.48225200G>A	rs117892325	0,126	blok 1	63
NC_000016.10:g.48225539T>C	rs7196373	0,182	blok 2	64
NC_000016.10:g.48225657G>A	rs73546434	0,127	blok 1	65
NC_000016.10:g.48225847C>T	rs73546437	0,127	blok 1	66
NC_000016.10:g.48226352A>G	rs8055128	0,127	blok 1	67
NC_000016.10:g.48226528T>C	rs8060696	0,127	blok 1	68
NC_000016.10:g.48226566T>C	rs8060722	0,127	blok 1	69
NC_000016.10:g.48226572C>T	rs8047397	0,127	blok 1	70
NC_000016.10:g.48226719G>A	rs8056100	0,127	blok 1	71
NC_000016.10:g.48226787G>A	rs8056143	0,127	blok 1	72
NC_000016.10:g.48227040A>C	rs73546439	0,127	blok 1	73
NC_000016.10:g.48227330C>T	rs113855339	0,127	blok 1	74
NC_000016.10:g.48227418C>G	rs113647089	0,127	blok 1	75
NC_000016.10:g.48228106T>C	rs7359477	0,127	blok 1	76
NC_000016.10:g.48228412T>C	rs112408414	0,127	blok 1	77
NC_000016.10:g.48228551A>G	rs112037239	0,127	blok 1	78
NC_000016.10:g.48229128G>A	rs9938414	0,127	blok 1	79
NC_000016.10:g.48229622A>G	rs8046378	0,126	blok 1	80
NC_000016.10:g.48229682T>A	rs8051546	0,127	blok 1	81
NC_000016.10:g.48229808T>C	rs8051852	0,127	blok 1	82
NC_000016.10:g.48230337G>T	rs8043751	0,127	blok 1	83
NC_000016.10:g.48230706A>G	rs13336759	0,127	blok 1	84
NC_000016.10:g.48230864T>C	rs13338105	0,127	blok 1	85
NC_000016.10:g.48231149C>T	rs11863414	0,127	blok 1	86
NC_000016.10:g.48231866C>T	rs16945988	0,127	blok 1	87
NC_000016.10:g.48232278C>T	rs16945990	0,127	blok 1	88
NC_000016.10:g.48232555A>T	rs28570577	0,126	blok 1	89
NC_000016.10:g.48232569A>C	rs11866225	0,127	blok 1	90

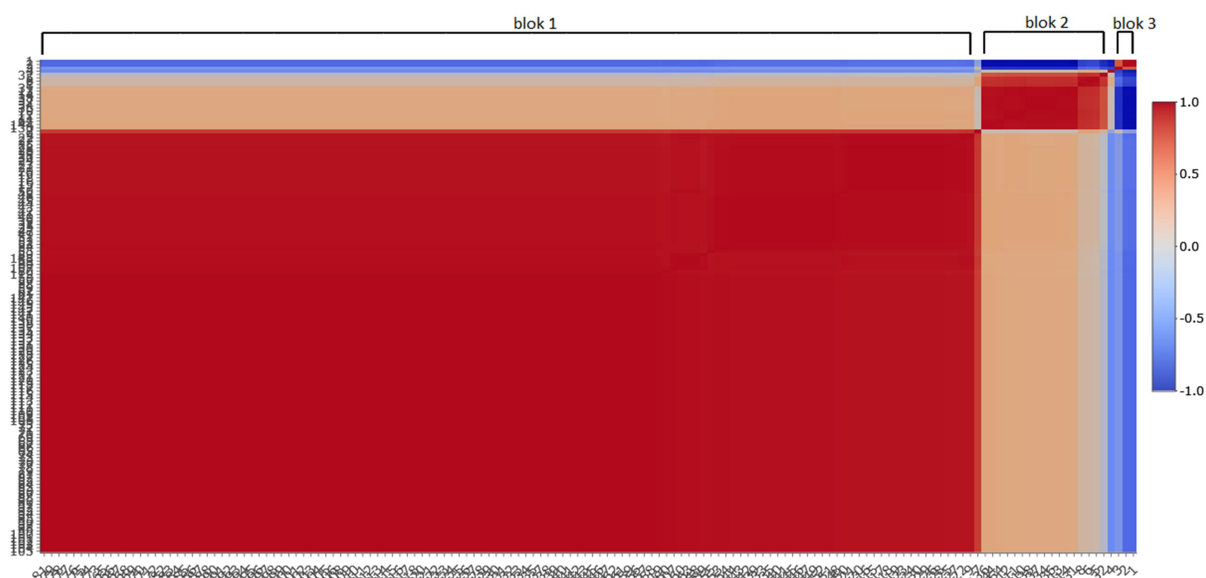
Wariant	nr rs wariantu	MAF	blok w chr.	nr w chr.
NC_000016.10:g.48232628T>C	rs11859911	0,127	blok 1	91
NC_000016.10:g.48232920C>T	rs11865596	0,127	blok 1	92
NC_000016.10:g.48233419G>T	rs13334052	0,127	blok 1	93
NC_000016.10:g.48233436C>A	rs13337955	0,127	blok 1	94
NC_000016.10:g.48233591T>G	rs13332304	0,127	blok 1	95
NC_000016.10:g.48233667A>G	rs13331031	0,127	blok 1	96
NC_000016.10:g.48234376T>G	rs11861346	0,127	blok 1	97
NC_000016.10:g.48234377T>A	rs11861347	0,127	blok 1	98
NC_000016.10:g.48234384T>A	rs11861348	0,127	blok 1	99
NC_000016.10:g.48234407T>C	rs11861379	0,127	blok 1	100
NC_000016.10:g.48234808G>C	rs13336542	0,127	blok 1	101
NC_000016.10:g.48234948_48234949dupGA	rs112734273	0,127	blok 1	102
NC_000016.10:g.48235018G>A	rs13336625	0,127	blok 1	103
NC_000016.10:g.48235210delAGAGATGCAA	rs398088092	0,127	blok 1	104
NC_000016.10:g.48235650A>G	rs10521167	0,127	blok 1	105
NC_000016.10:g.48236007T>C	rs16946006	0,127	blok 1	106
NC_000016.10:g.48236518C>T	rs9926206	0,126	blok 1	107
NC_000016.10:g.48236597T>C	rs9934833	0,127	blok 1	108
NC_000016.10:g.48236663A>G	rs9932328	0,127	blok 1	109
NC_000016.10:g.48237551G>A	rs8043685	0,127	blok 1	110
NC_000016.10:g.48237641G>A	rs8058396	0,127	blok 1	111
NC_000016.10:g.48237770A>G	rs8062407	0,127	blok 1	112
NC_000016.10:g.48237777C>T	rs8054117	0,127	blok 1	113
NC_000016.10:g.48237907C>T	rs8054444	0,127	blok 1	114
NC_000016.10:g.48238441G>C	rs16946011	0,127	blok 1	115
NC_000016.10:g.48238475C>T	rs16946014	0,127	blok 1	116
NC_000016.10:g.48238612delA	rs111614375	0,127	blok 1	117
NC_000016.10:g.48238805T>C	rs4340327	0,127	blok 1	118
NC_000016.10:g.48238915C>T	rs4325548	0,127	blok 1	119
NC_000016.10:g.48239284T>C	rs112904913	0,127	blok 1	120
NC_000016.10:g.48239288C>T	rs7203378	0,127	blok 1	121
NC_000016.10:g.48239387C>T	rs7203432	0,127	blok 1	122
NC_000016.10:g.48239433A>G	rs58938063	0,127	blok 1	123
NC_000016.10:g.48240101C>T	rs112433800	0,127	blok 1	124
NC_000016.10:g.48240213G>A	rs113241663	0,127	blok 1	125
NC_000016.10:g.48240511T>C	rs7202089	0,127	blok 1	126
NC_000016.10:g.48240707C>T	rs73555628	0,127	blok 1	127
NC_000016.10:g.48240870G>A	rs73555630	0,127	blok 1	128
NC_000016.10:g.48241016C>T	rs112070127	0,127	blok 1	129
NC_000016.10:g.48241221C>T	rs28849445	0,127	blok 1	130
NC_000016.10:g.48241258T>G	rs28842015	0,127	blok 1	131
NC_000016.10:g.48241658T>C	rs8049398	0,127	blok 1	132
NC_000016.10:g.48241664C>A	rs8056692	0,127	blok 1	133
NC_000016.10:g.48242119T>A	rs112624830	0,127	blok 1	134
NC_000016.10:g.48242368A>C	rs7186126	0,127	blok 1	135
NC_000016.10:g.48242373A>G	rs7186129	0,181	blok 2	136
NC_000016.10:g.48242610G>A	rs7186982	0,127	blok 1	137
NC_000016.10:g.48242671A>G	rs7191181	0,127	blok 1	138
NC_000016.10:g.48243063T>A	rs111465091	0,127	blok 1	139
NC_000016.10:g.48243107A>G	rs112059727	0,127	blok 1	140
NC_000016.10:g.48243269C>T	rs9930843	0,127	blok 1	141
NC_000016.10:g.48243273T>A	rs113449465	0,127	blok 1	142
NC_000016.10:g.48243597_48243598insT	rs112913257	0,127	blok 1	143
NC_000016.10:g.48243939T>C	rs9940038	0,181	blok 2	144
NC_000016.10:g.48244070C>T	rs73555643	0,127	blok 1	145
NC_000016.10:g.48244714A>G	rs111886126	0,127	blok 1	146
NC_000016.10:g.48244751T>C	rs7187166	0,127	blok 1	147
NC_000016.10:g.48245181C>G	rs60431412	0,126	blok 1	148



Rycina 8. Analiza nierównowagi sprzężeń wariantów w chromosomie 11.



Rycina 9. Analiza nierównowagi sprzężeń wariantów w chromosomie 13.



Rycina 10. Analiza nierównowagi sprzężeń wariantów w chromosomie 16.

Z puli 647 wariantów jedynie 16 zostało zlokalizowanych w sekwencjach kodujących (tak rozumianych, jeśli kodowały jakikolwiek transkrypt), z czego dziewięć były to zmiany synonimiczne (Tab. 7). Przewidywany efekt *in silico* na funkcjonalność białka pozostałych siedmiu wariantów missensownych wykazał ich potencjalny szkodliwy efekt tylko dla SNP: rs16945988 (*ABCC11*), rs3819519, rs3819518 (oba *ABCC8*) i rs61732039 (*ABCB5*) (jedynie według algorytmu SIFT), przy czym oba warianty w *ABCC8* zlokalizowane są wyłącznie w niekodujących transkryptach. Warianty eksonowe, w tym missensowne, wydaje się zatem mają raczej marginalny wpływ na obserwowane klastrowanie. Zdecydowana większość wykrytej zmienności w puli 647 wariantów związanych z obrazem klastrowania była zatem zlokalizowana w sekwencjach intronowych.

Sprawdzono potencjalny efekt 647 wariantów na poziom ekspresji genów w 54 poszczególnych rodzajach narządów lub ich struktur z bazy danych GTEx (Tab. 8). Nie wykryto wpływu żadnego wariantu na poziom ekspresji jakiegokolwiek genów w dziewięciu narządach lub strukturach: nerka kora, nerka rdzeń, jajnik, macica, pochwa, szyjka macicy – endocervix, szyjka macicy – ectocervix, jajowód, pęcherz moczowy. W pozostałych 45 zidentyfikowano wpływ na poziom ekspresji 25 różnych genów lub niekodujących transkryptów, osiem z nich obejmowało jednak pseudogeny (cztery przypadki) lub niekodujące transkrypty jak lncRNA (ang. *long non-coding RNA*) (trzy przypadki) lub snoRNA (ang. *small nucleolar RNAs*) (jeden przypadek). Stwierdzono wpływ wspomnianych wariantów na ekspresję czterech genów kodujących białka ABC, przy czym najczęściej na *ABCC8*, bo efekt jakiegokolwiek wariantu został zaobserwowany aż w 35 narządach lub strukturach. Najwięcej wariantów oddziaływało zatem na jego ekspresję w mięśniach szkieletowych, nerwie piszczelowym, tarczycy, okrężnicy poprzecznej oraz niektórych strukturach mózgu. Z kolei wpływ największej ilości wariantów na ekspresję *ABCC4* dotyczył głównie tarczycy i krwi pełnej,

w przypadku *ABCC11* jedynie nadnerczy i hodowli komórkowych fibroblastów, a w przypadku *ABCG5* ślędziony. Odnotowano także wpływ tych wariantów zlokalizowanych w genach ABC na ekspresję innych genów. Warianty w sekwencji *ABCC8* wywoływały efekt na ekspresję aż dziewięciu innych genów lub transkryptów, warianty w *ABCC11* – czterech, w *ABCC4* – trzech, w *ABCB11* – trzech, *ABCG8* – jednego i *ABCB5* – jednego. Największa liczba spośród tych wariantów okazuje się może mieć wpływ na poziom ekspresji genu *LONP2*, aż w 21 narządach lub ich strukturach, głównie: tkankach tłuszczowych, tętnicach, przełyku, płucu, mięśniach szkieletowych, nerwie piszczelowym, skórze i ślędzionie. Rolą tego białka jest prawdopodobnie utrzymanie homeostazy peroksysomów podczas katalizowania beta-oksydacji kwasów tłuszczowych (według bazy *UniProtKB/Swiss-Prot*). Godne odnotowania są też geny *G6PC2*, kodujący podjednostkę glukozy-6-fosfatazy – kluczowego enzymu w glukoneogenezie, dla którego stwierdzono wpływ na ekspresję w 12 strukturach oraz *KCNJ11*, kodujący podjednostkę Kir6.2 kanału potasowego, dla którego stwierdzono wpływ na ekspresję w dziewięciu strukturach. Dla pozostałych genów kodujących, na których ekspresję wpływają te warianty, nie udało się zidentyfikować powiązań funkcjonalnych. Co ciekawe jednak, w wątrobie stwierdzono też wpływ wielu wariantów na ekspresję lncRNA *ENSG00000261538*.

4.2.4. Analiza grupowania próbek dla elementów ontologii, z którymi związane są geny ABC

Przeprowadzono szereg grupowań próbek na podstawie procesów biologicznych lub zlokalizowania w kompartmentach komórkowych (w sumie zostały określone jako elementy ontologii, ang. *ontology term*) z bazy *Gene Ontology*, w które zaangażowane są geny ABC, aby niezależnie w ten sposób określić podłoże opisanego powyżej klastrowania wszystkich genów ABC uwzględnionych w analizie po imputacji. Do grupowania wzięto pod uwagę w sumie 33 procesy biologiczne lub kompartmenty komórkowe, w które według bazy zaangażowanych jest co najmniej osiem genów, a także elementy ontologii arbitralnie wybrane jako interesujące, jeśli liczność genów w nie zaangażowana była mniejsza niż osiem (Tab. 9). Z analizy wykluczono elementy ontologii, z którymi według bazy związanych było ponad 40 genów czyli zdecydowana większość nadrodziny ABC, gdyż nie ujawniłoby to podłoża obserwowanego klastrowania, a zatem: wiązanie ATP (ang. *ATP binding*) (48 genów), błona komórkowa (ang. *membrane*) (48 genów), wiązanie nukleotydów (ang. *nucleotide binding*) (47 genów), hydroliza ATP (ang. *ATP hydrolysis activity*) (44 geny), aktywność transportera typu ABC (ang. *ABC-type transporter activity*) (44 geny), transport transbłonowy (ang. *transmembrane transport*) (44 geny), aktywność transbłonowego transportera sprzężonego z aktywnością ATPazową (ang. *ATPase-coupled transmembrane transporter activity*) (43 geny), integralny składnik błony (ang. *integral component of membrane*) (41 genów). Podczas grupowania próbek uwzględniono warianty zarówno częste jak i rzadkie.

Tabela 7. Warianty z puli 647 związanych z klastrowaniem, zlokalizowane w kodujących sekwencjach genów, z przewidywanym wpływem *in silico* SNP missensownych na funkcjonalność białka.

pozycja wariantu	gen	numer rs	efekt wariantu	zmiana aminokwasu	SIFT	PolyPhen
NC_000007.14:g.20643581G>T	ABCB5	rs7349983	synonimiczny	p.Val209=	-	-
NC_000007.14:g.20647558C>T	ABCB5	rs2893006	synonimiczny	p.Ser335=	-	-
NC_000007.14:g.20647981A>G	ABCB5	rs61732039	missensowny	p.Asp370Gly	szkodliwy (0,01)	łagodny (0,149)
NC_000007.14:g.20651424T>C	ABCB5	rs34603556	missensowny	p.Met446Thr	nieszkodliwy (1,00)	łagodny (0)
NC_000007.14:g.20651596G>A	ABCB5	rs61227829	synonimiczny	p.Ala58= p.Ala503=	-	-
NC_000007.14:g.20658647A>G	ABCB5	rs2301641	missensowny	p.Lys115Glu p.Lys560Glu	nieszkodliwy (1,00)	-
NC_000011.10:g.17465421C>G	ABCC8	rs3819519	missensowny, NMD	p.Arg188Pro	szkodliwy (0)	łagodny (0,001)
NC_000011.10:g.17465532C>T	ABCC8	rs3819518	missensowny, NMD	p.Ser151Asn	szkodliwy (0,02)	łagodny (0,195)
NC_000011.10:g.17474969A>G	ABCC8	rs1048099	synonimiczny	p.Pro69= p.Pro29= p.Pro64= p.Pro67= p.Pro35=	-	-
NC_000013.11:g.95062722C>T	ABCC4	rs1751034	synonimiczny	p.Lys1116= p.Lys1069= p.Lys1073=	-	-
NC_000013.11:g.95206742T>C	ABCC4	rs2274406	synonimiczny	p.Arg242= p.Arg317= p.Arg330=	-	-
NC_000013.11:g.95209550A>G	ABCC4	rs899494	synonimiczny	p.Ile148= p.Ile223= p.Ile236=	-	-
NC_000016.10:g.48215007T>C	ABCC11	rs8047091	synonimiczny	p.Lys374=	-	-
NC_000016.10:g.48216115G>T	ABCC11	rs11863236	missensowny	p.Ala317Glu	nieszkodliwy (0,27)	łagodny (0)
NC_000016.10:g.48222691T>C	ABCC11	rs16945974	synonimiczny	p.Gln228=	-	-
NC_000016.10:g.48231866C>T	ABCC11	rs16945988	missensowny	p.Arg19His	szkodliwy (0,01)	łagodny (0,003)

NMD – proces niszczenia nonsensownych transkryptów (ang. *Nonsense-Mediated Decay*); predykcja algorytmu SIFT: szkodliwy <0,05, nieszkodliwy ≥0,05; predykcja algorytmu PolyPhen: łagodny ≤0,446, 0,446< potencjalnie szkodliwy ≤0,908, prawdopodobnie szkodliwy >0,908.

Dla sześciu elementów ontologii otrzymano wyraźne cztery, niezależne zgrupowania próbek o składzie klastrow, odpowiadającym tym jakie otrzymano podczas klastrowania próbek dla wariantów częstych. W tych przypadkach zaobserwowano nawet dużo wyraźniejsze rozdzielanie zgrupowania składającego się z klastrow 5 i 7 od klastru nr 4. Były to więc elementy ontologii opisane jako: błona plazmatyczna szczytowa (12 genów) (Ryc. 11), aktywność transbłonowego transportera anionów nieorganicznych sprzężona z aktywnością ATPazową (10 genów) (Ryc. 12), transport anionów nieorganicznych (8 genów) (Ryc. 13), aktywność transportera typu ABC S-koniugatów glutationu (7 genów) (Ryc. 14), transport leukotrienów (7 genów) (Ryc. 15), transport kwasów żółciowych i soli żółciowych (6 genów) (Ryc. 16). Dla procesu biologicznego „transport lipidów” grupowanie na cztery opisane powyżej skupienia były ciągle zachowane, chociaż granice między nimi

były już dużo bardziej zbliżone, prawdopodobnie dlatego, że aż 27 genów jest opisanych w bazie jako związane z tym procesem, co obniża rozdzielczość klastrowania (Ryc. 17). Tendencja taka jest widoczna zwłaszcza dla elementu ontologii „błona plazmatyczna”, uwarunkowanej od aż 31 genów, gdzie granice pomiędzy poszczególnymi klastrami w większości ulegają zatarciu (Ryc. 18). W przypadku procesu biologicznego „aktywność transportera ksenobiotyków typu ABC” (13 genów) widoczne było wyraźniejsze rozdzielanie się poszczególnych klastrów w zgrupowaniach próbek, tak np. klastry nr 1 i 9 wyraźniej były odseparowane od dalej złączonych klastrów nr 3 i 8, podobnie klastry nr 2 i 6 zaczęły od siebie odstawać, choć bez wyraźnie zaznaczonej granicy (Ryc. 19). Natomiast w przypadku elementów ontologii: wiązanie białek (25 genów) (Ryc. 20), aparat Golgiego (10 genów) (Ryc. 21), transport przez barierę krew-mózg (9 genów) (Ryc. 22), uzyskano rozdział na jedynie dwa zgrupowania, gdyż granice między klastrami nr 2, 6 i nr 4 oraz między nr 1, 3, 8, 9, a nr 5, 7 uległy całkowitemu zatarciu. Efekt ten był tym bardziej widoczny im mniej genów było w niego zaangażowanych. Podobny obraz uzyskano dla elementów ontologii: aktywność transbłonowego transportera ksenobiotyków (8 genów) (Ryc. 23) i błona plazmatyczna podstawno-boczna (7 genów) (Ryc. 24), choć skupienia próbek w tych przypadkach były znacznie bardziej rozciągnięte świadcząc o większym znaczeniu w tym grupowaniach wariantów rzadkich lub przebijającym się efekcie innych genów na taki model klastrowania. Dla pozostałych elementów ontologii uzyskano już zupełnie inne obrazy grupowania próbek, świadczące o braku ich związku z klastrowaniem wszystkich genów ABC uwzględnionych w analizie. Mianowicie, poszczególne klastry w zgrupowaniach były wymieszane pomiędzy sobą albo tworzyły wiele odstających od siebie małych skupień (Ryc. 25-43).

Powyższe wyniki sugerują, że istnieje gen lub ich grupa, którego zmienność zasadniczo wpływa na rozróżnienie czterech zgrupowań, a nie dwóch. Traktując zatem każde klastrowanie jako jedynie efekt zmienności genetycznej specyficznego składu genów jaki go kształtują, zbliżony trzon klastrowania na dwa i cztery zgrupowania stanowią, z niewielkimi fluktuacjami, głównie geny z rodziny ABCC. W przypadku wszystkich klastrowań, w których otrzymano tylko dwa zgrupowania próbek, w żadnym z nich nie był uwzględniony gen *ABCC11*, co wskazuje na jego bezpośredni udział w podziale na kolejne dwa zgrupowania i sumarycznie tworzenie czterech grupowań.

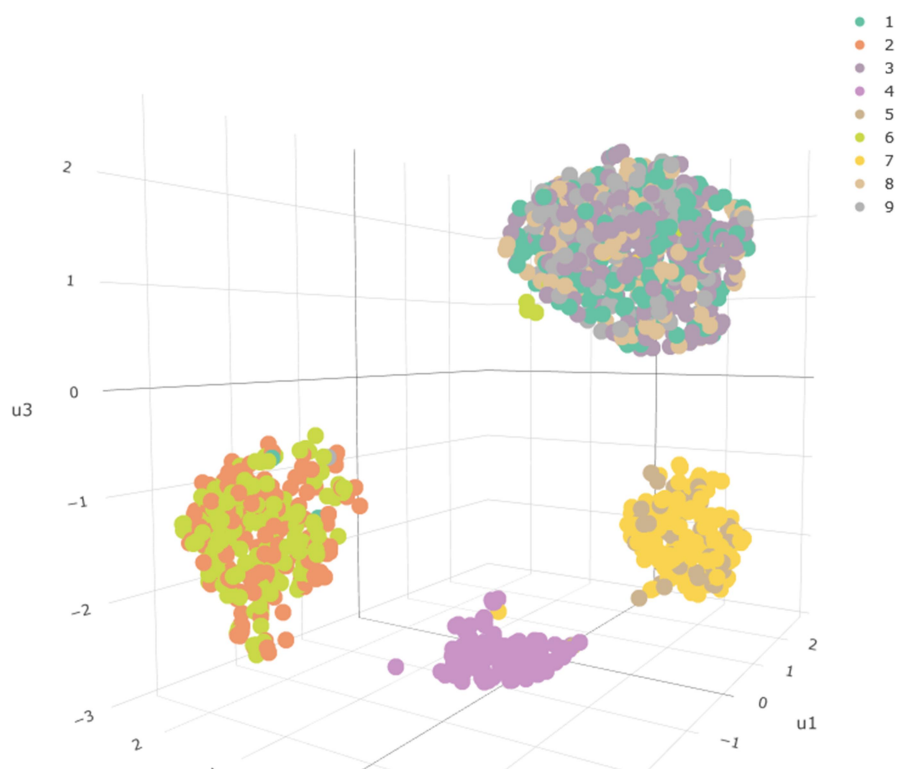
[illegible]

Oryginalna nazwa tkanki w bazie GTEx ↓	Nazwa tkanki ↓	Gen ↓	ABCC4	ABCC8	ABCC11	ABCG5	CLDN10	DZIP1	ENSG000000260052	ENSG000000260196	ENSG000000261538	G6PC2	HPS5	KCNJ11	LONP2	LRPPRC	NCR3LG1	NOSTRIN	OR7E14P	RNY3P8	RPL23P8	RPS13	SAA4	SDHCP4	SIAH1	SNORD14B	SPC25
Muscle - Skeletal	Mięsień szkieletowy		1	95											140												
Nerve - Tibial	Nerw piszczelowy		2	94								4			137			1									
Pancreas	Trzustka														19	1		1									1
Pituitary	Przysadka mózgowa		1																								
Prostate	Prostata			1																							
Skin - Not Sun Exposed Suprapubic	Skóra - nieeksponowana na słońce (podbrzusze)			1								5		4	17	1									8	1	
Skin - Sun Exposed Lower leg	Skóra - eksponowana na słońce (dolna część nogi)			4								3		2	36	1		2				21					
Small Intestine - Terminal Ileum	Jelito cienkie - koniec jelita krętego			2																							
Spleen	Śledziona					4									134												
Stomach	Żołądek			3								5							34								
Testis	Jądro		4	5														5									
Thyroid	Tarczycza		103	76								5		1	17												
Whole Blood	Krew pełna		89						1																7		

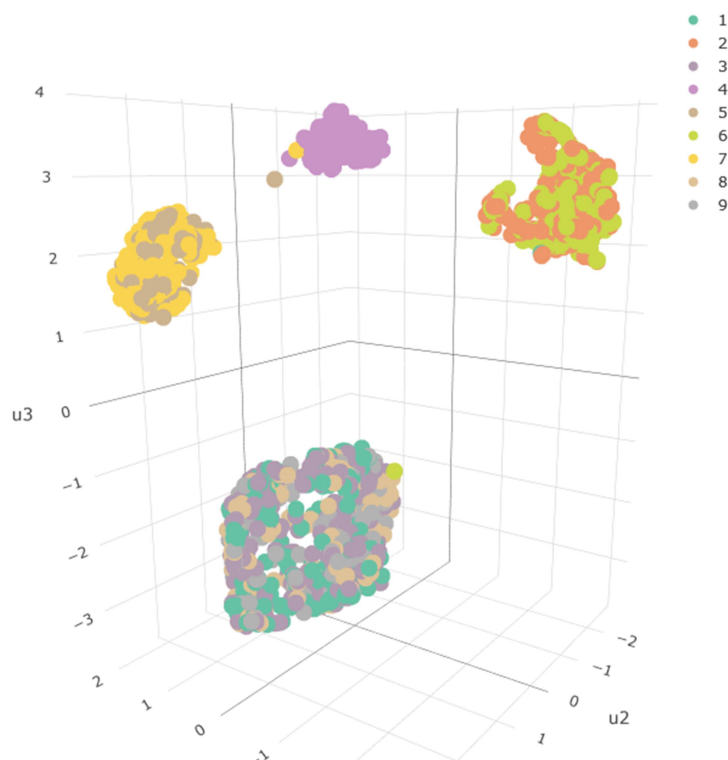
CLDN10 - ang. claudin 10; **DZIP1**- ang. DAZ interacting zinc finger protein 1; **ENSG000000260052** - niekodujący transkrypt (lncRNA); **ENSG000000260196** - niekodujący transkrypt (lncRNA); **ENSG000000261538** - niekodujący transkrypt (lncRNA); **G6PC2** - ang. glucose-6-phosphatase catalytic subunit 2; **HPS5** - ang. HPS5 biogenesis of lysosomal organelles complex 2 subunit 2; **KCNJ11** - ang. potassium inwardly rectifying channel subfamily J member 11; **LONP2** - ang. lon peptidase 2, peroxisomal; **LRPPRC** - ang. leucine rich pentatricopeptide repeat containing; **NCR3LG1** - ang. natural killer cell cytotoxicity receptor 3 ligand 1; **NOSTRIN** - ang. nitric oxide synthase trafficking inductor; **OR7E14P** - ang. olfactory receptor family 7 subfamily E member 14 pseudogene; **RNY3P8** - ang. RNY3 pseudogene 8; **RPL23P8** - ang. ribosomal protein L23 pseudogene 8; **RPS13** - ang. ribosomal protein S13; **SAA4** - ang. serum amyloid A4; **SDHCP4** - ang. succinate dehydrogenase complex subunit C pseudogene 4; **SIAH1** - ang. siah E3 ubiquitin protein ligase 1; **SNORD14B** - ang. small nucleolar RNA, C/D box 14B (snoRNA); **SPC25** - ang. SPC25 component of NDC80 kinetochore complex

Tabela 9. Klastrowanie 33 procesów biologicznych i kompartmentów komórkowych (elementów ontologii) na podstawie bazy *Gene Ontology*, z którymi związane są geny ABC. Geny zaciemnione nie zostały uwzględnione w analizie: *ABCB7* i *ABCD1* ze względu na położenie w chromosomie X; pozostałe – brak wyniku imputacji.

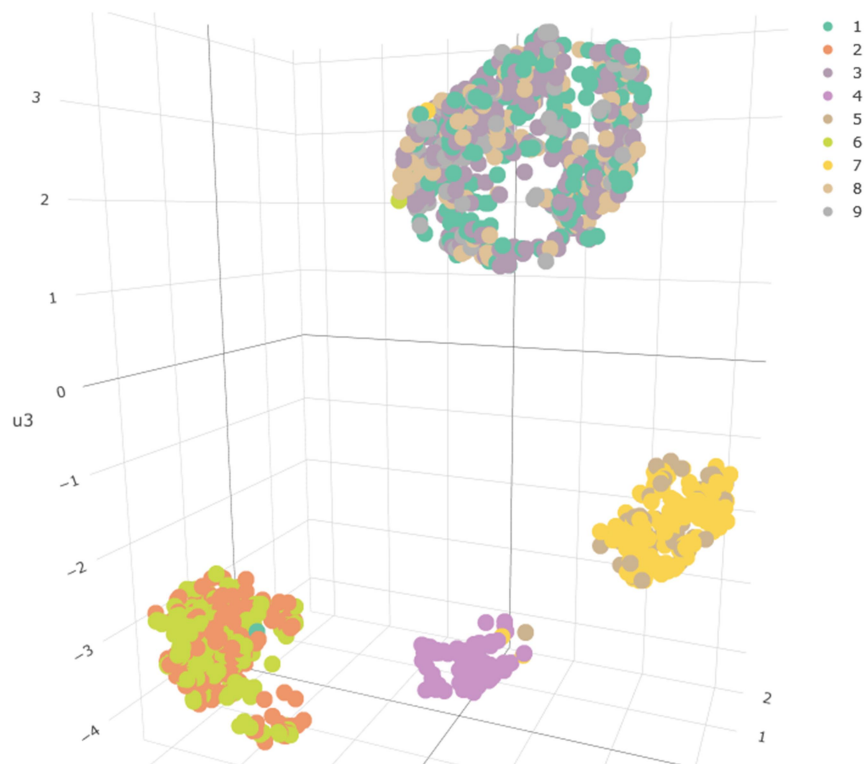
liczba genów	Originalna nazwa elementu ontologii w bazie <i>Gene Ontology</i>	Nazwa elementu ontologii, dla którego przeprowadzono klastrowanie
12	apical plasma membrane	błona plazmatyczna szczytowa
10	ATPase-coupled inorganic anion transmembrane transporter activity	aktywność transblonowego transportera anionów nieorganicznych sprzężona z aktywnością ATPazową
8	inorganic anion transport	transport anionów nieorganicznych
7	ABC-type glutathione S-conjugate transporter activity	aktywność transportera typu ABC S-koniugatów glutationu
7	leukotriene transport	transport leukotrienów
6	bile acid and bile salt transport	transport kwasów żółciowych i soli żółciowych
27	lipid transport	transport lipidów
13	ABC-type xenobiotic transporter activity	aktywność transportera ksenobiotyków typu ABC
25	protein binding	wiązanie białek
10	Golgi apparatus	aparat Golgiego
9	transport across blood-brain barrier	transport przez barierę krew-mózg
8	xenobiotic transmembrane transporter activity	aktywność transblonowego transportera ksenobiotyków
7	basolateral plasma membrane	błona plazmatyczna podstawno-boczna
31	plasma membrane	błona plazmatyczna
8	cytoplasmic vesicle	pęcherzyk cytoplazmatyczny
14	intracellular membrane-bounded organelle	organelle otoczone błoną wewnątrzkomórkową
11	endoplasmic reticulum	retikulum endoplazmatyczne
12	lipid transporter activity	aktywność transportera lipidów
10	endoplasmic reticulum membrane	błona retikulum endoplazmatycznego
11	endosome	endosom
8	protein homodimerization activity	aktywność homodimeryzacji białek
10	response to xenobiotic stimulus	odpowiedź na bodziec ksenobiotyczny
10	xenobiotic transport	transportu ksenobiotyków
8	mitochondrion	mitochondrium
8	cytosol	cytozol
9	cholesterol efflux	wyrzut cholesterolu
8	lysosomal membrane	błona lizosomalna
8	nucleoplasm	nukleoplazma
7	cytoplasm	cytoplazma
8	cholesterol homeostasis	homeostaza cholesterolu
5	metal ion binding	wiązanie jonów metali
4	organic anion transmembrane transporter activity	aktywność transblonowego transportera anionów organicznych
3	MHC class I protein binding	wiązanie z białkami MHC klasy I
<i>ABCA1</i>		
<i>ABCA2</i>		
<i>ABCA3</i>		
<i>ABCA4</i>		
<i>ABCA5</i>		
<i>ABCA6</i>		
<i>ABCA7</i>		
<i>ABCA8</i>		
<i>ABCA9</i>		
<i>ABCA10</i>		
<i>ABCA12</i>		
<i>ABCA13</i>		
<i>ABCB1</i>	X	
<i>TAP1</i>		



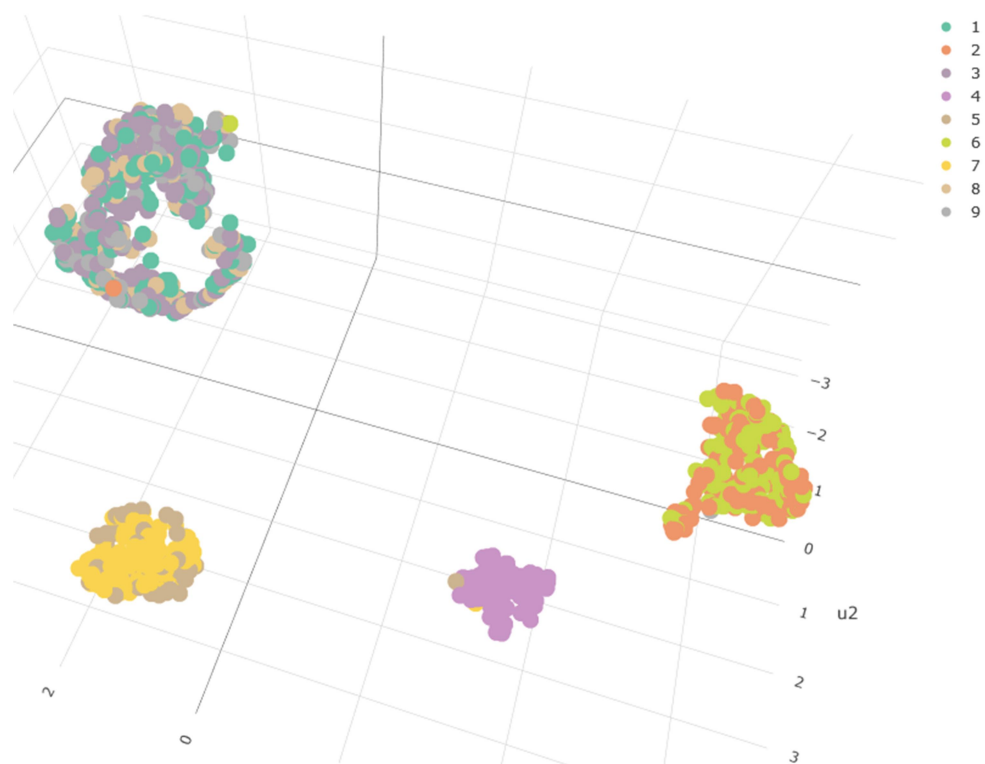
Rycina 11. Klastrowanie dla elementu ontologii „błona plazmatyczna szczytowa”.



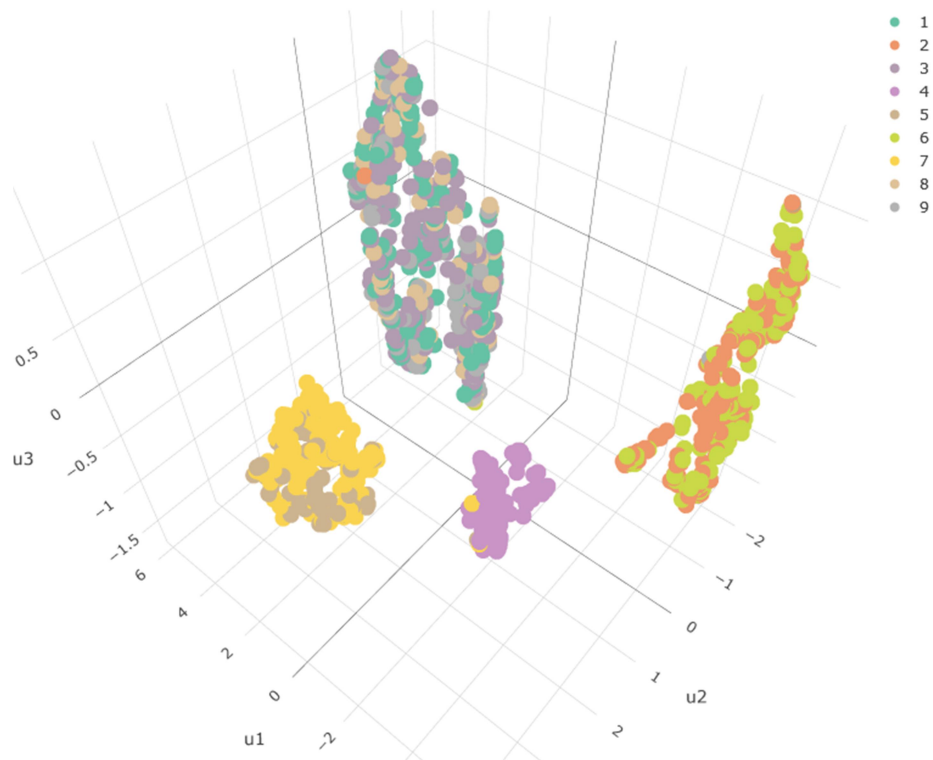
Rycina 12. Klastrowanie dla elementu ontologii „aktywność transbłonowego transportera anionów nieorganicznych sprzężona z aktywnością ATPazową”.



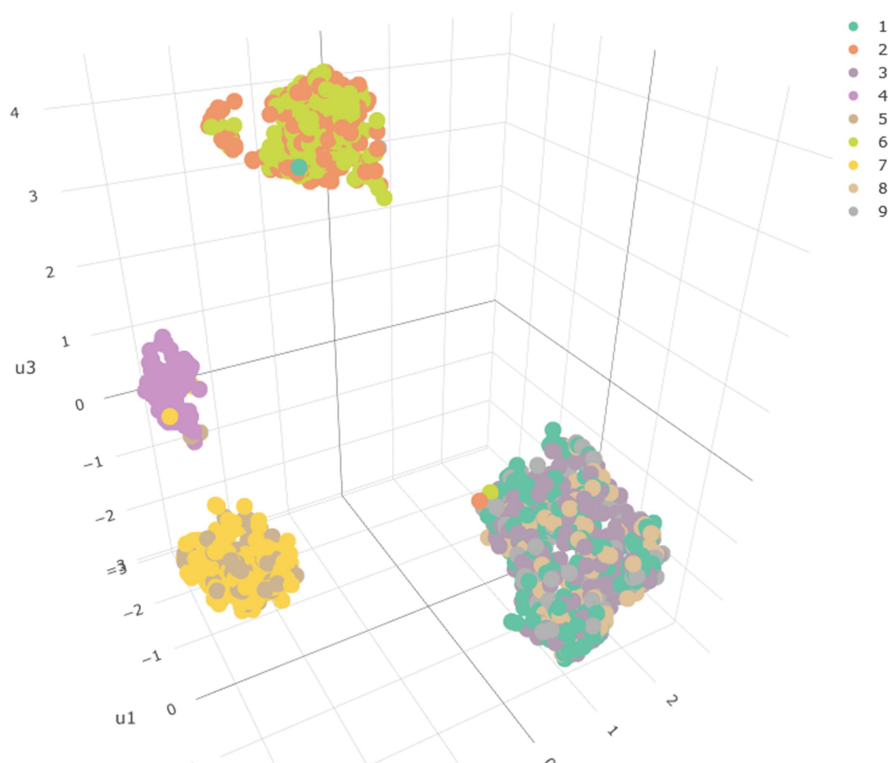
Rycina 13. Klastrowanie dla elementu ontologii „transport anionów nieorganicznych”.



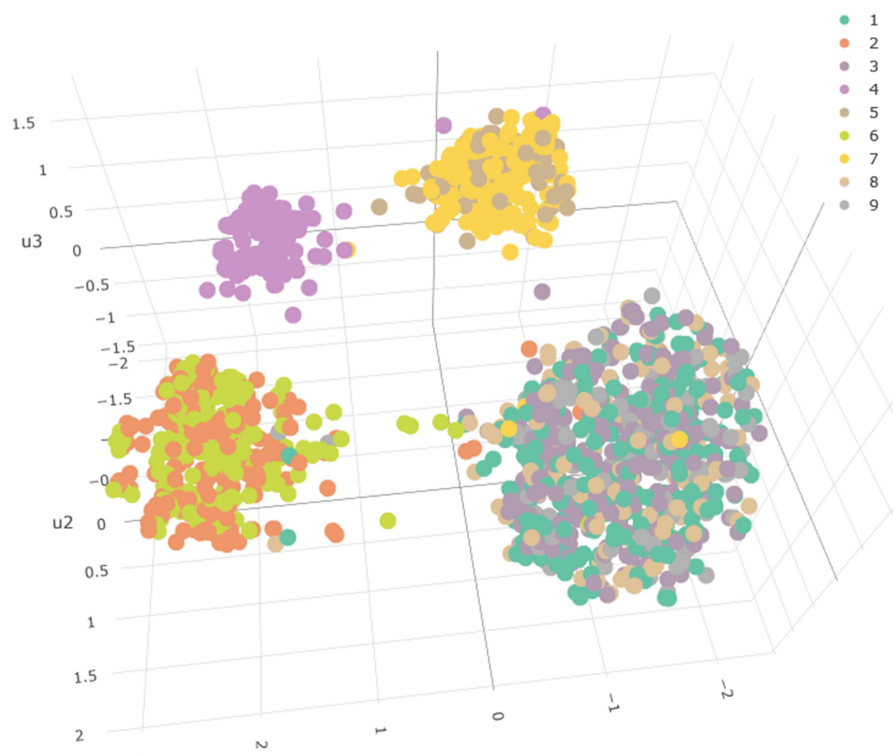
Rycina 14. Klastrowanie dla elementu ontologii „aktywność transportera typu ABC S-koniugatów glutationu”.



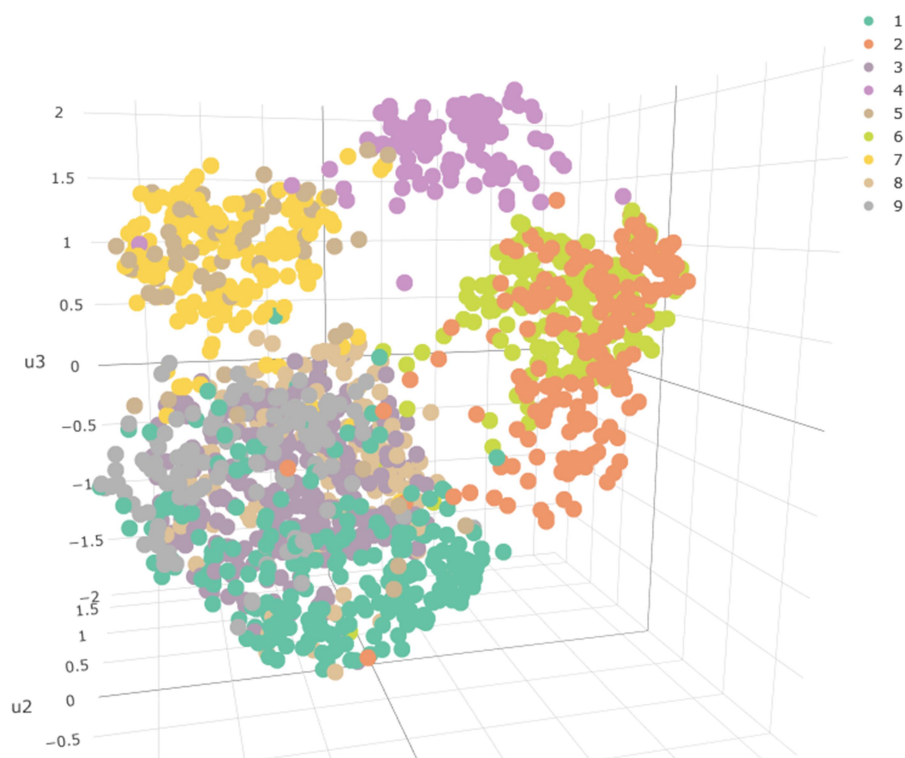
Rycina 15. Klastrowanie dla elementu ontologii „transport leukotrienów”.



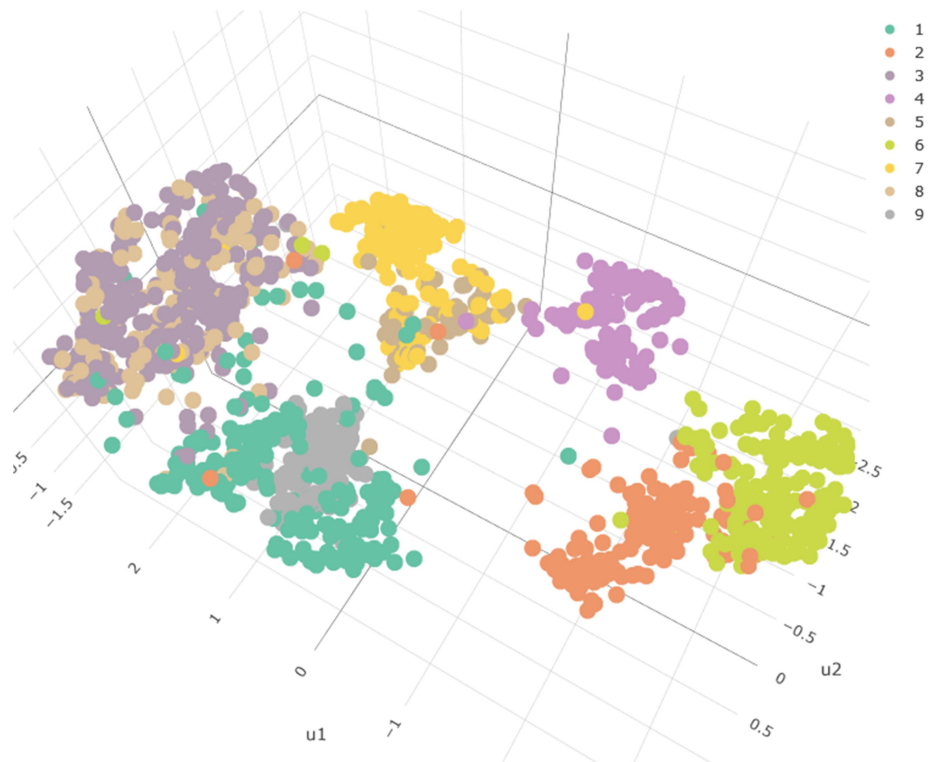
Rycina 16. Klastrowanie dla elementu ontologii „transport kwasów żółciowych i soli żółciowych”.



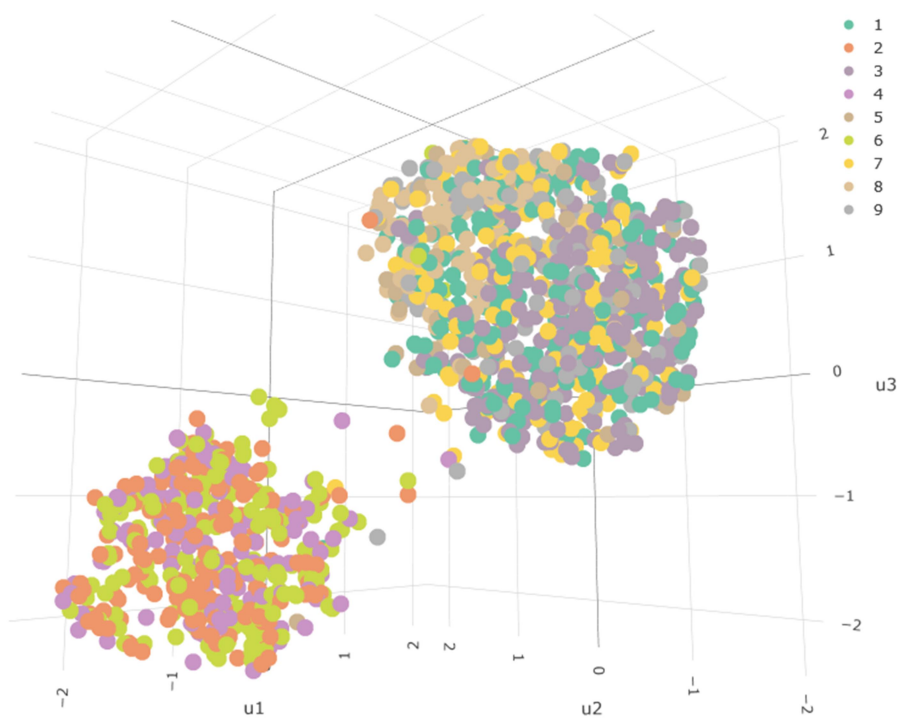
Rycina 17. Klastrowanie dla elementu ontologii „transport lipidów”.



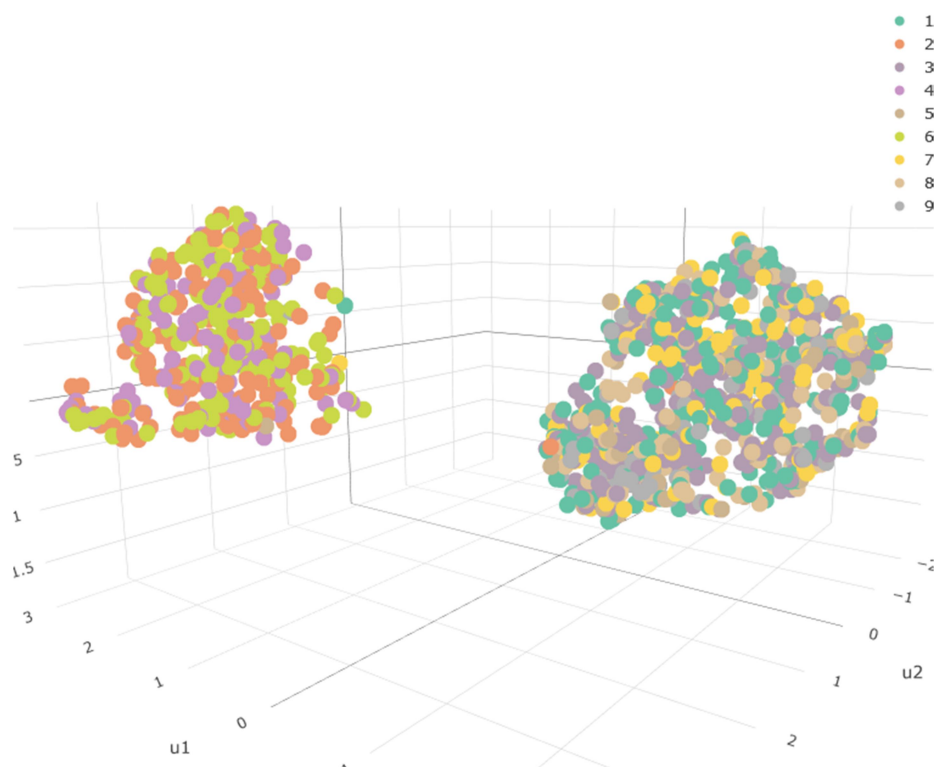
Rycina 18. Klastrowanie dla elementu ontologii „błona plazmatyczna”.



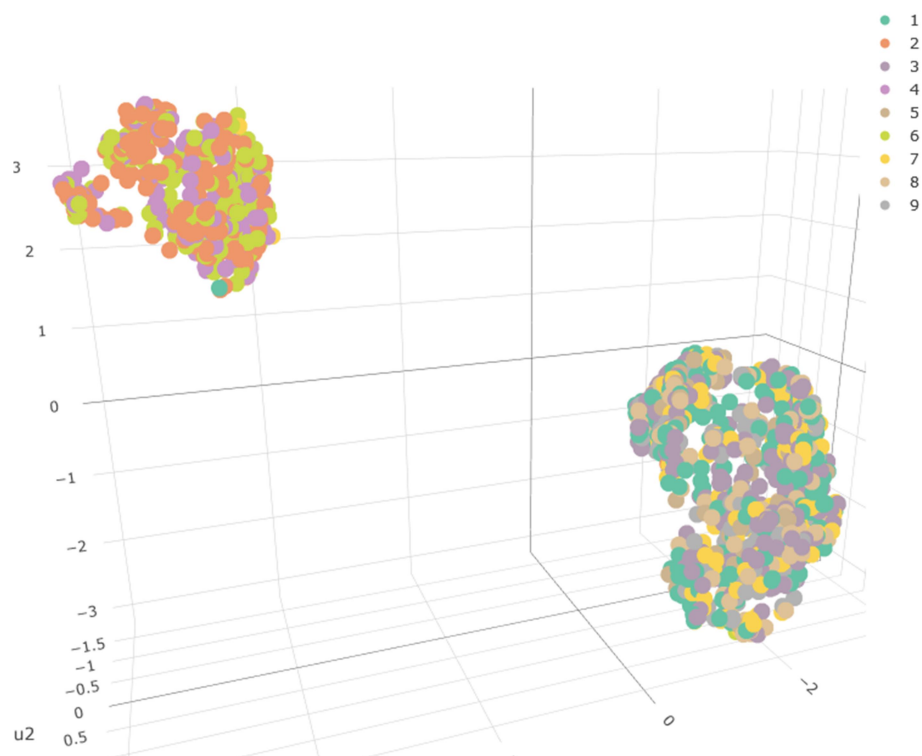
Rycina 19. Klastrowanie dla elementu ontologii „aktywność transportera ksenobiotyków typu ABC”.



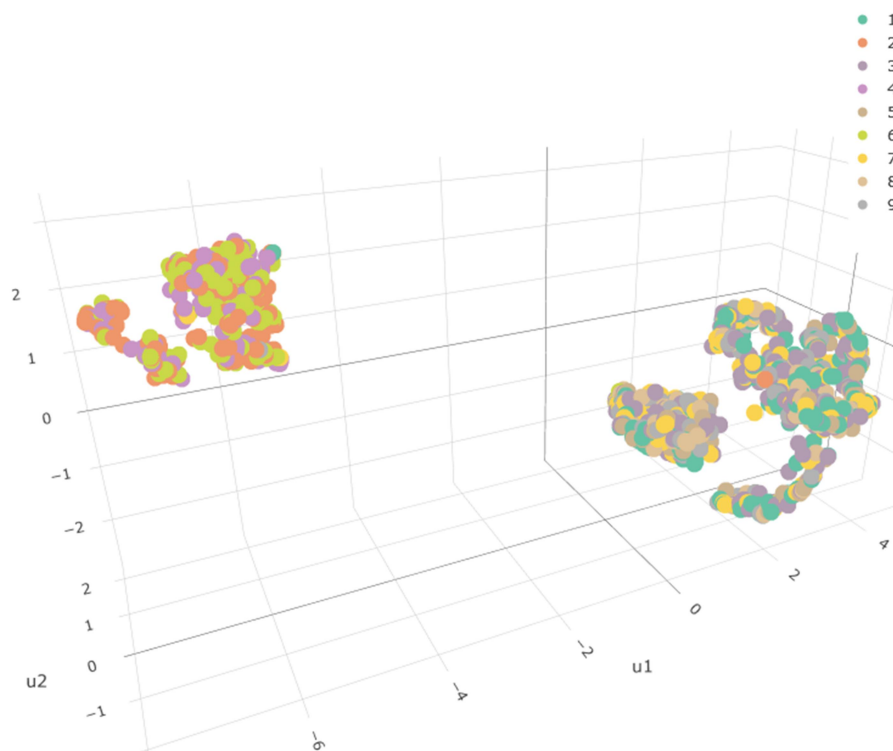
Rycina 20. Klastrowanie dla elementu ontologii „wiązanie białek”.



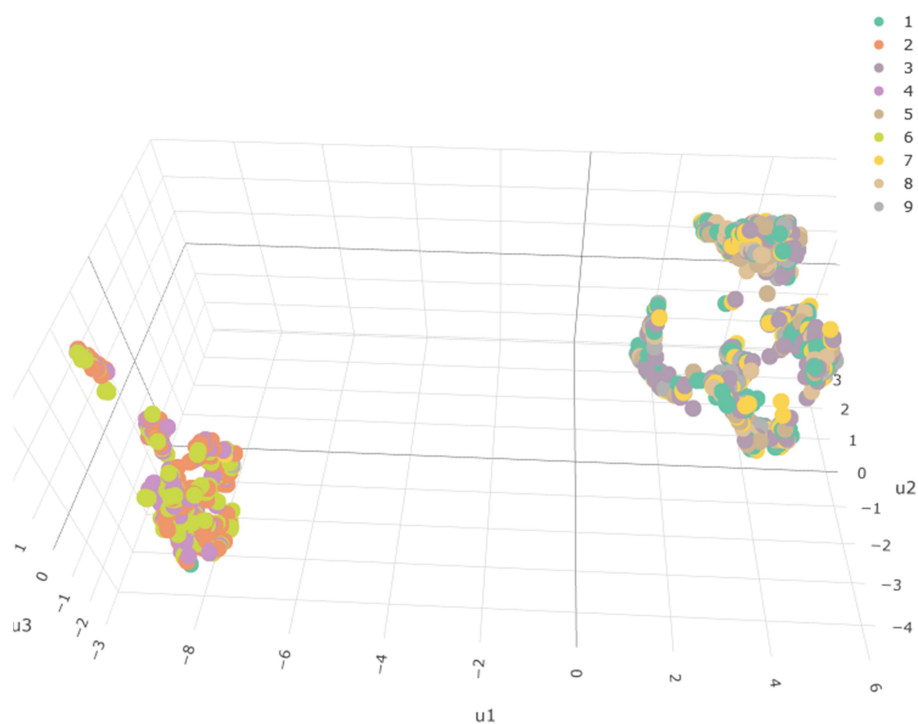
Rycina 21. Klastrowanie dla elementu ontologii „aparat Golgiego”.



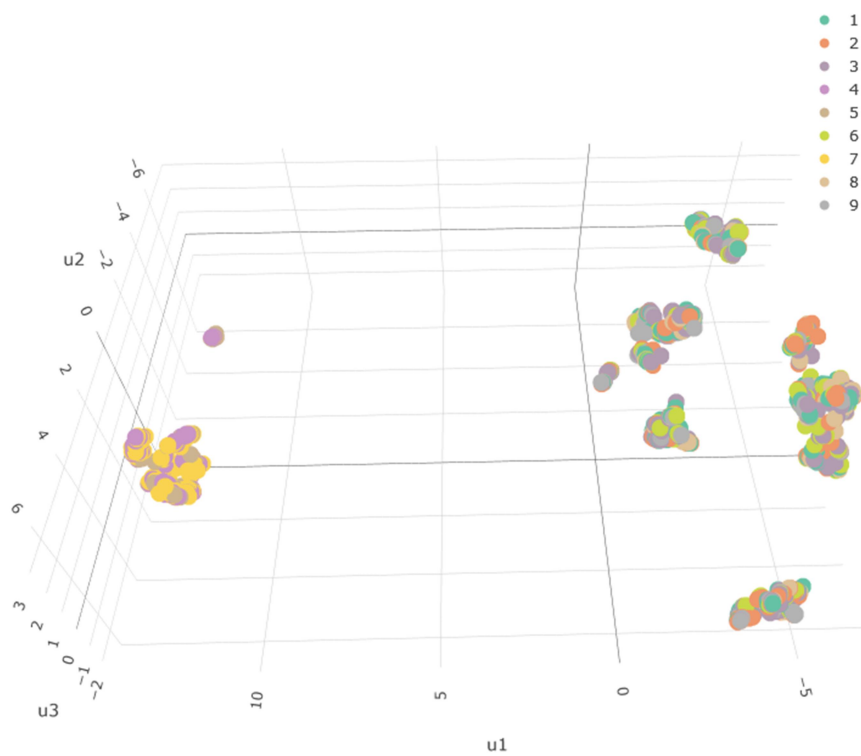
Rycina 22. Klastrowanie dla elementu ontologii „transport przez barierę krew-mózg”.



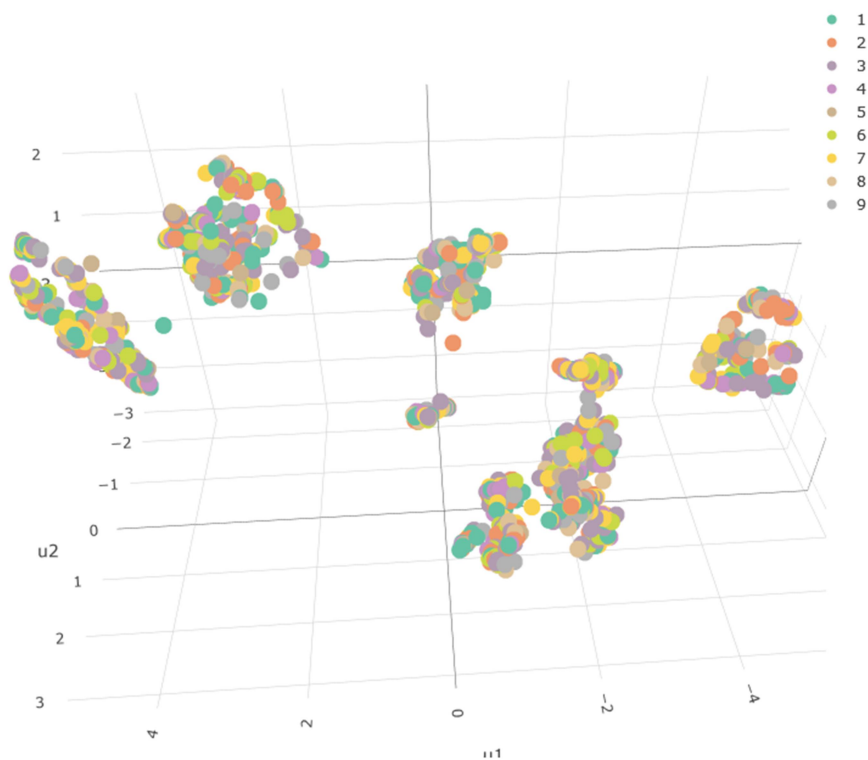
Rycina 23. Klastrowanie dla elementu ontologii „aktywność transbłonowego transportera ksenobiotyków”.



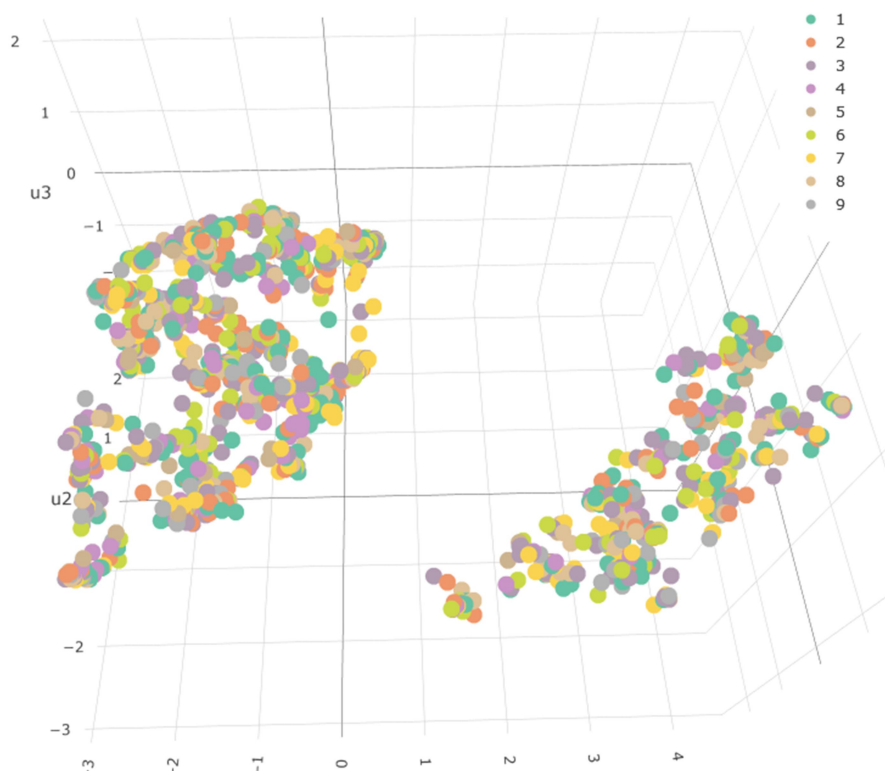
Rycina 24 . Klastrowanie dla elementu ontologii „błona plazmatyczna podstawno-boczna”.



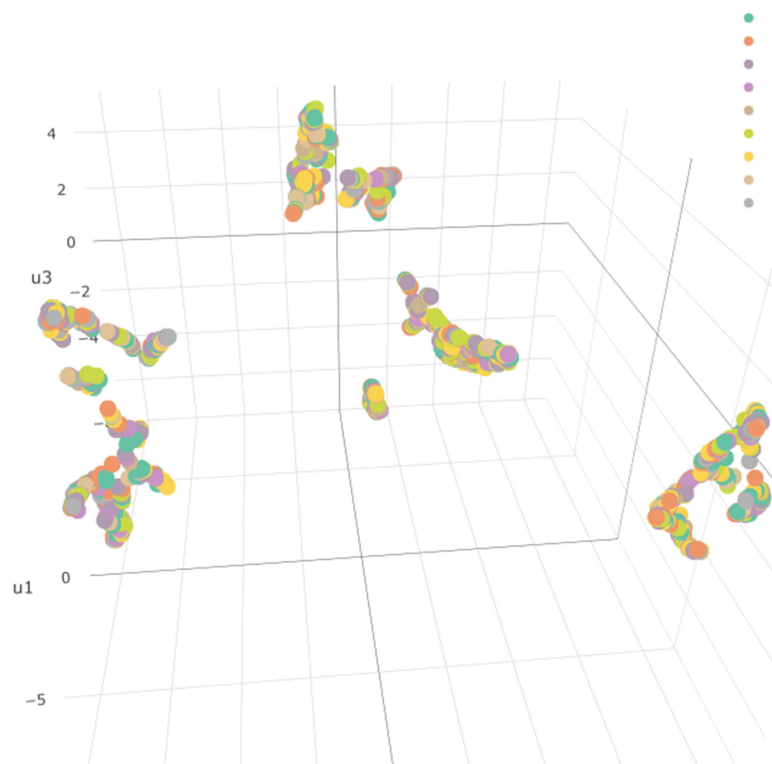
Rycina 25. Klastrowanie dla elementu ontologii „pęcherzyk cytoplazmatyczny”.



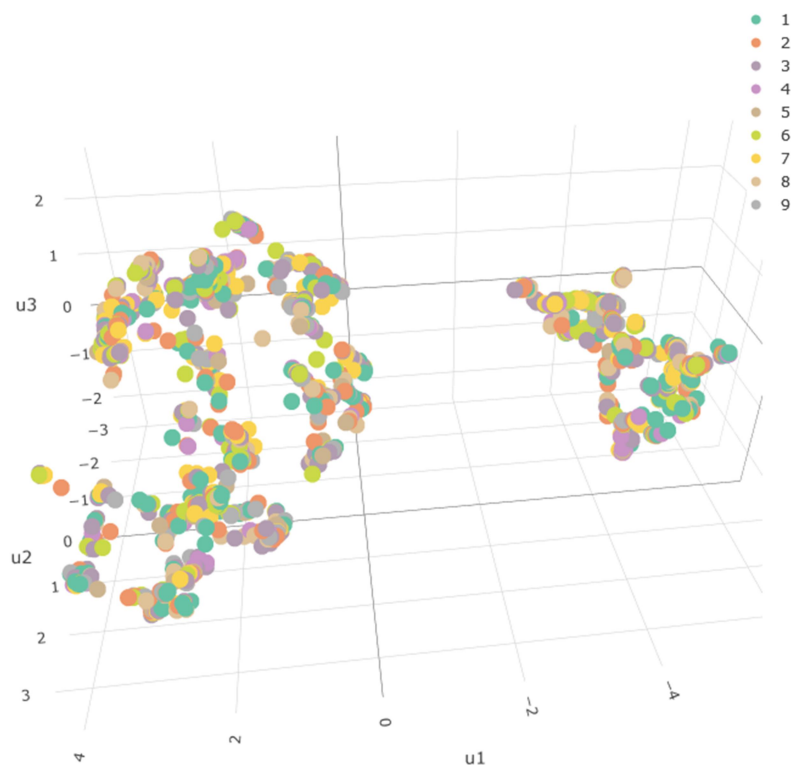
Rycina 26. Klastrowanie dla elementu ontologii „organelle otoczone błoną wewnątrzkomórkową”.



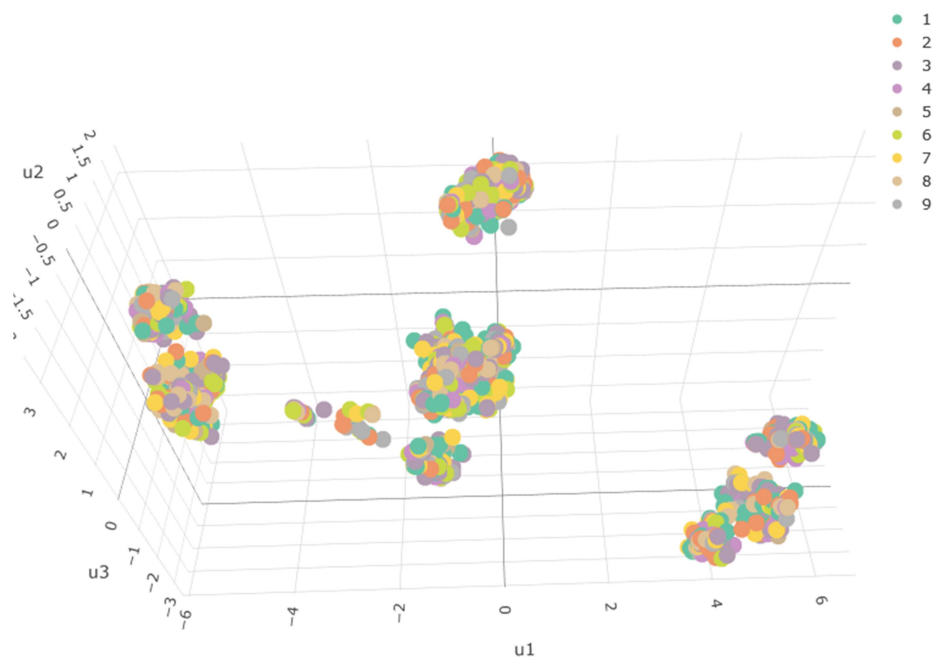
Rycina 27. Klastrowanie dla elementu ontologii „retikulum endoplazmatyczne”.



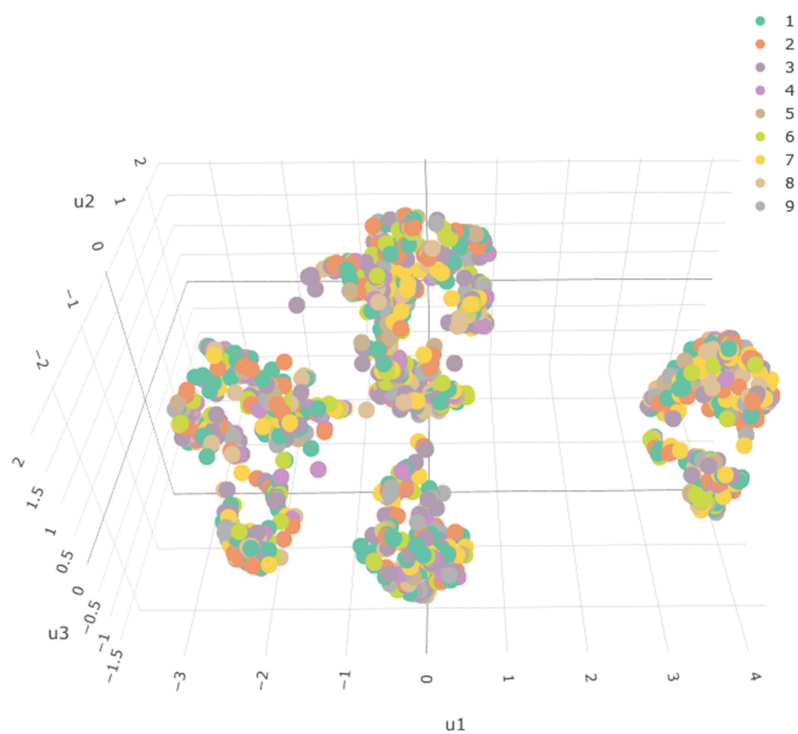
Rycina 28. Klastrowanie dla elementu ontologii „aktywność transportera lipidów”.



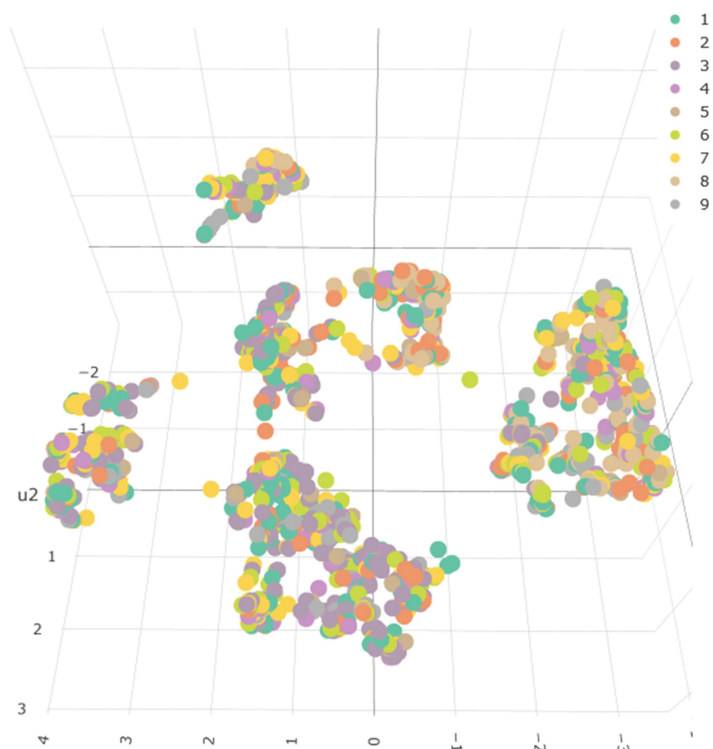
Rycina 29. Klastrowanie dla elementu ontologii „błona retikulum endoplazmatycznego”.



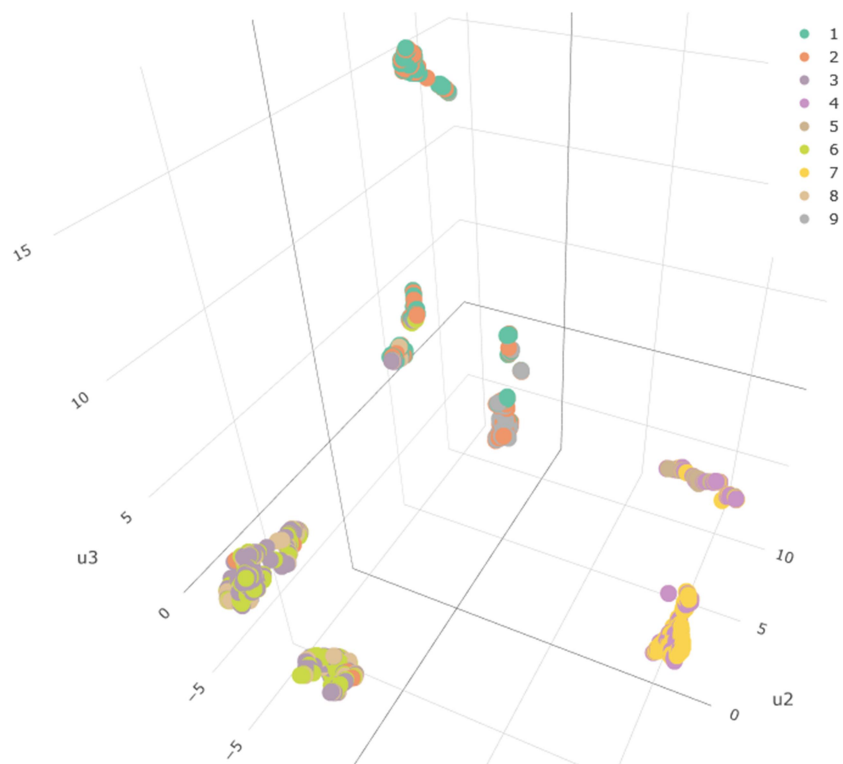
Rycina 30. Klastrowanie dla elementu ontologii „endosom”.



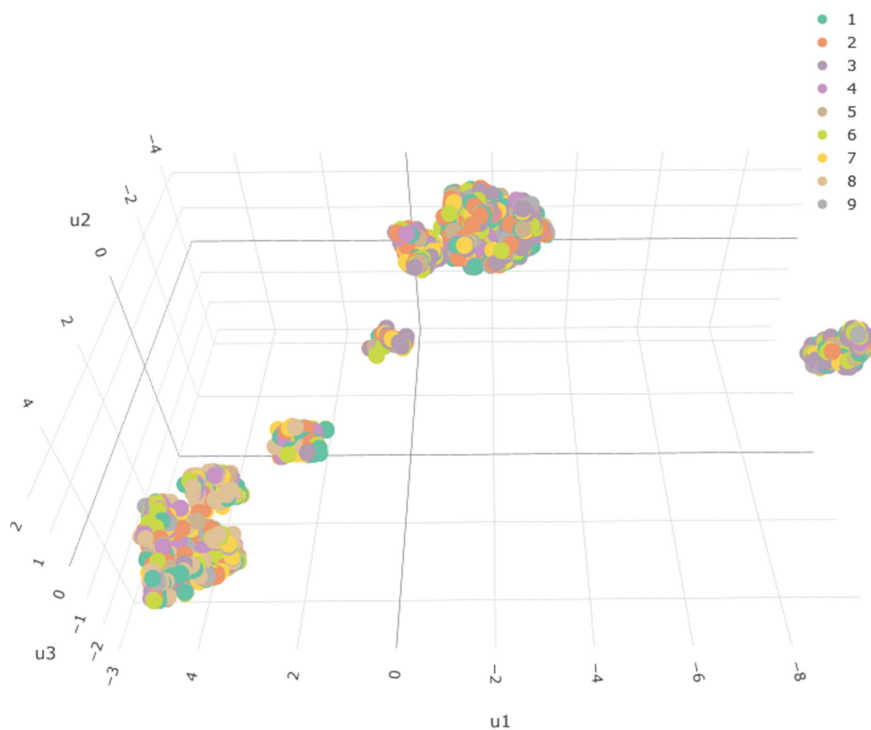
Rycina 31. Klastrowanie dla elementu ontologii „aktywność homodimeryzacji białek”.



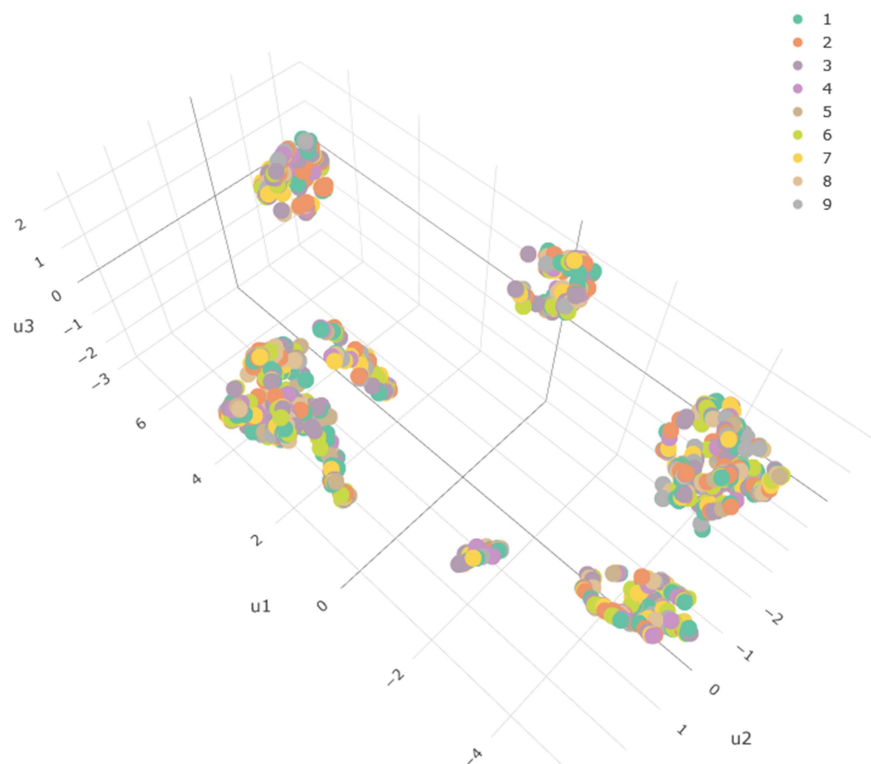
Rycina 32. Klastrowanie dla elementu ontologii „odpowiedź na bodziec ksenobiotyczny”.



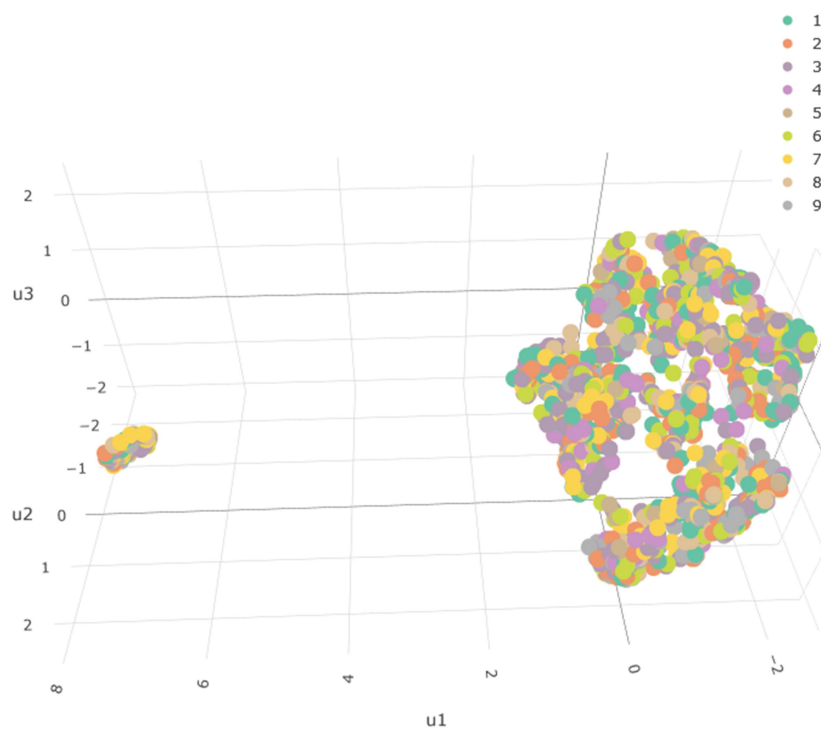
Rycina 33. Klastrowanie dla elementu ontologii „transport ksenobiotyków”.



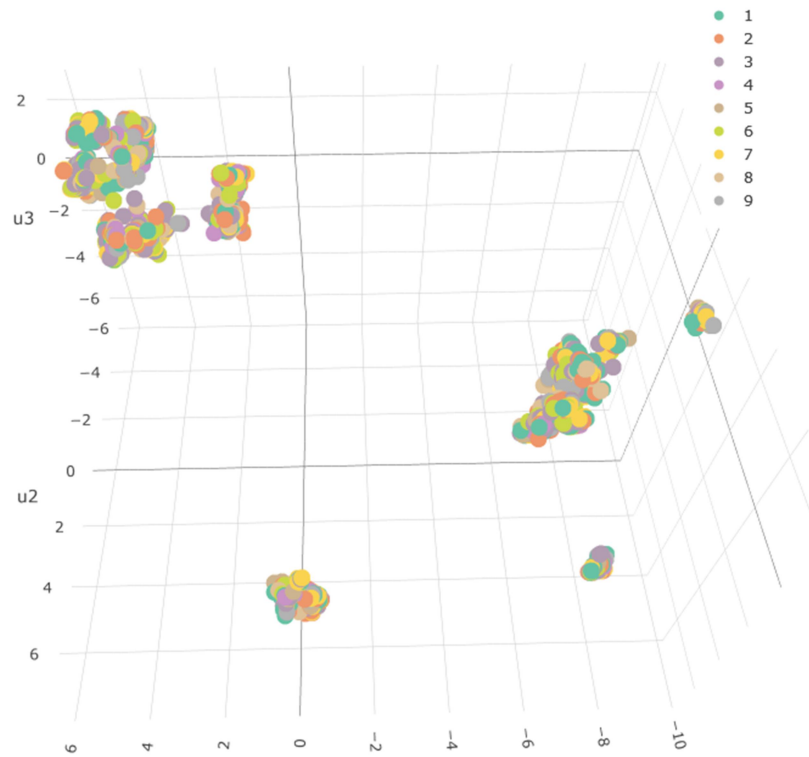
Rycina 34. Klastrowanie dla elementu ontologii „mitochondrium”.



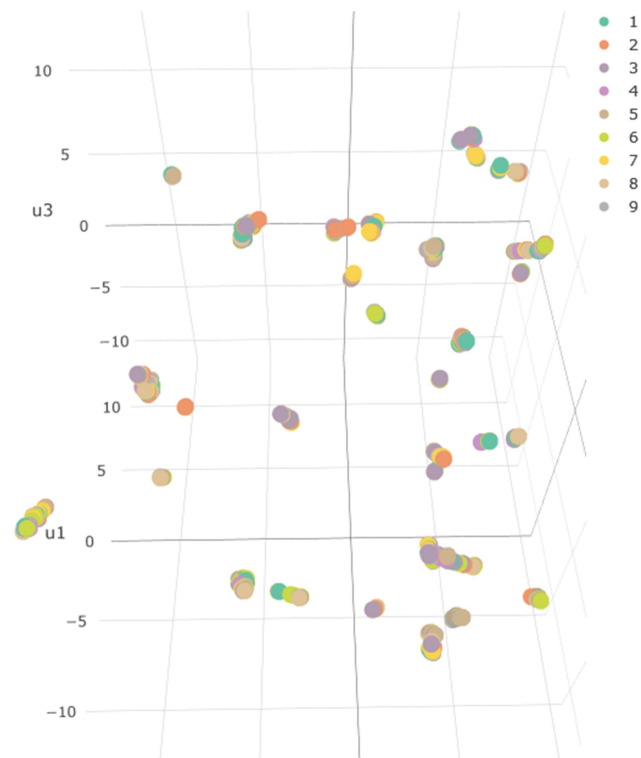
Rycina 35. Klastrowanie dla elementu ontologii „cytozol”.



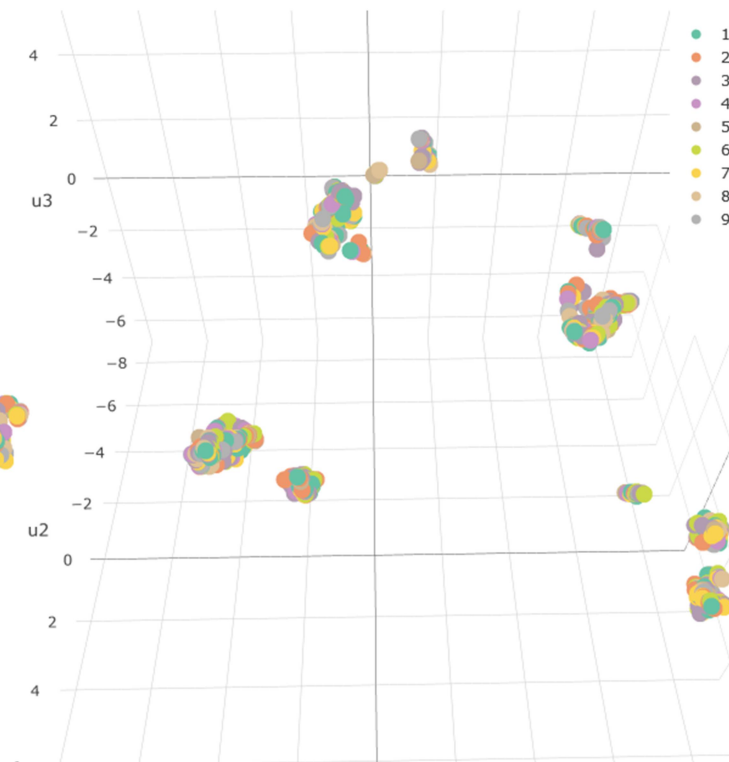
Rycina 36. Klastrowanie dla elementu ontologii „wyrzut cholesterolu”.



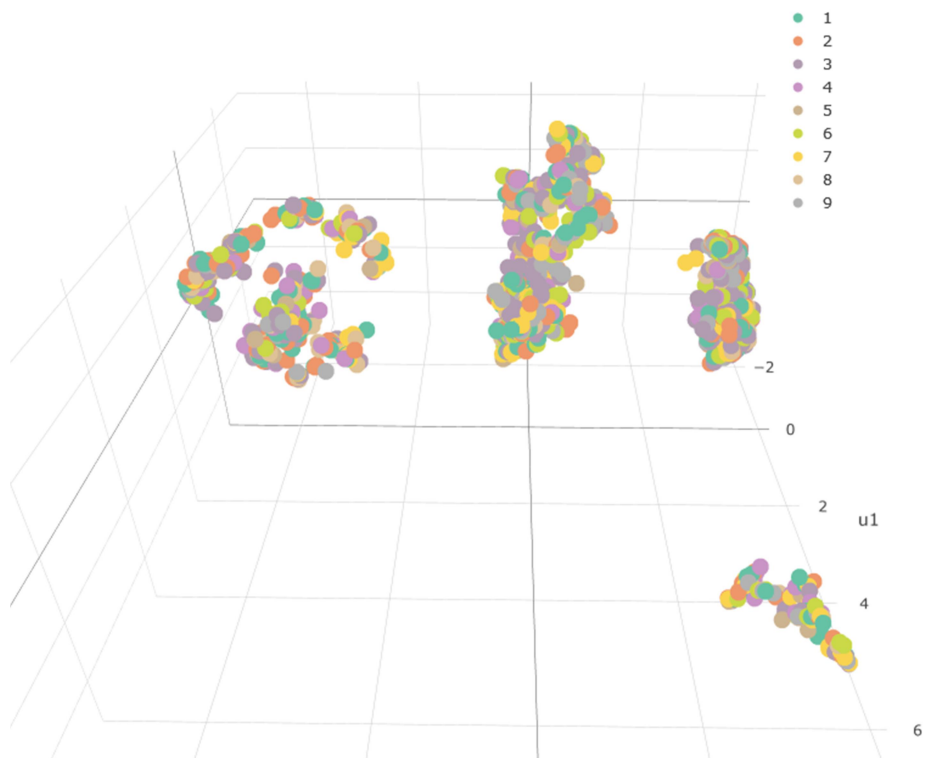
Rycina 37. Klastrowanie dla elementu ontologii „błona lizosomalna”.



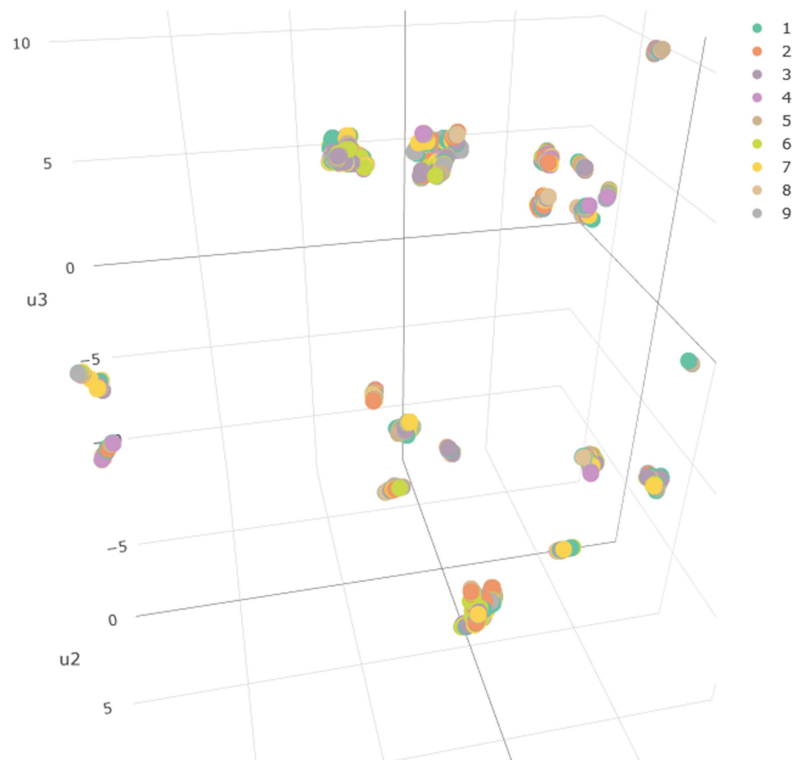
Rycina 38. Klastrowanie dla elementu ontologii „nukleoplazma”.



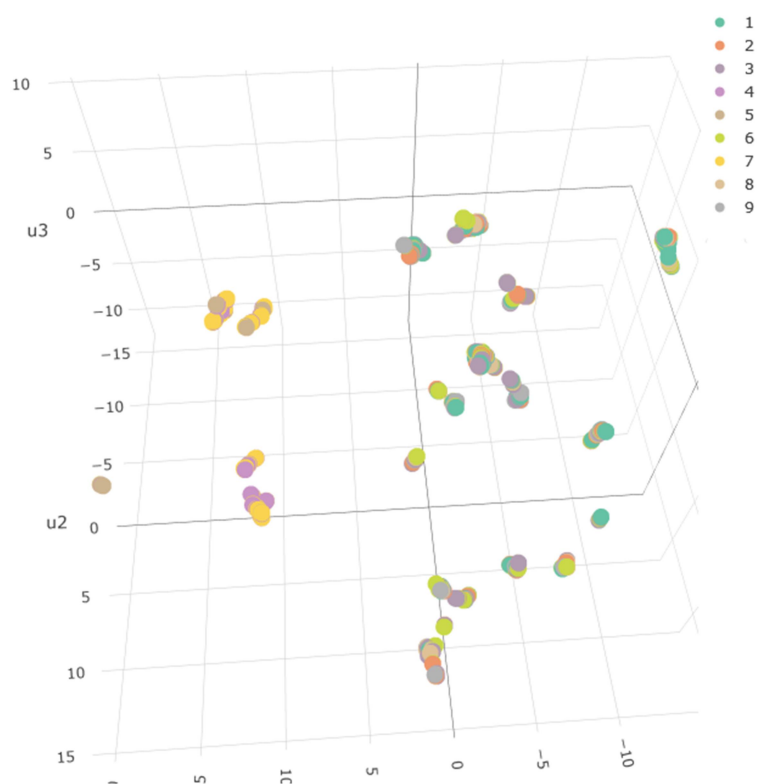
Rycina 39. Klastrowanie dla elementu ontologii „cytoplazma”.



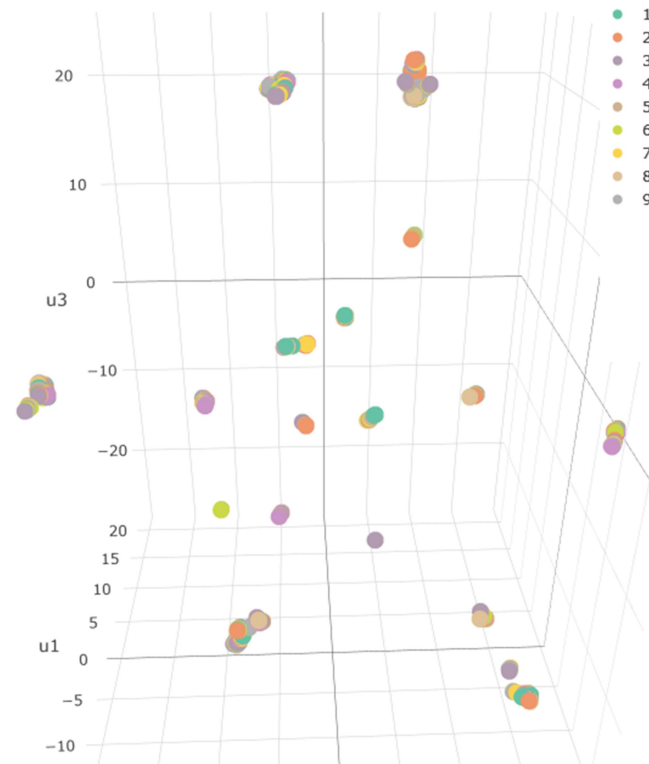
Rycina 40. Klastrowanie dla elementu ontologii „homeostaza cholesterolu”.



Rycina 41. Klastrowanie dla elementu ontologii „wiązanie jonów metali”.



Rycina 42. Klastrowanie dla elementu ontologii „aktywność transbłonowego transportera anionów organicznych”.



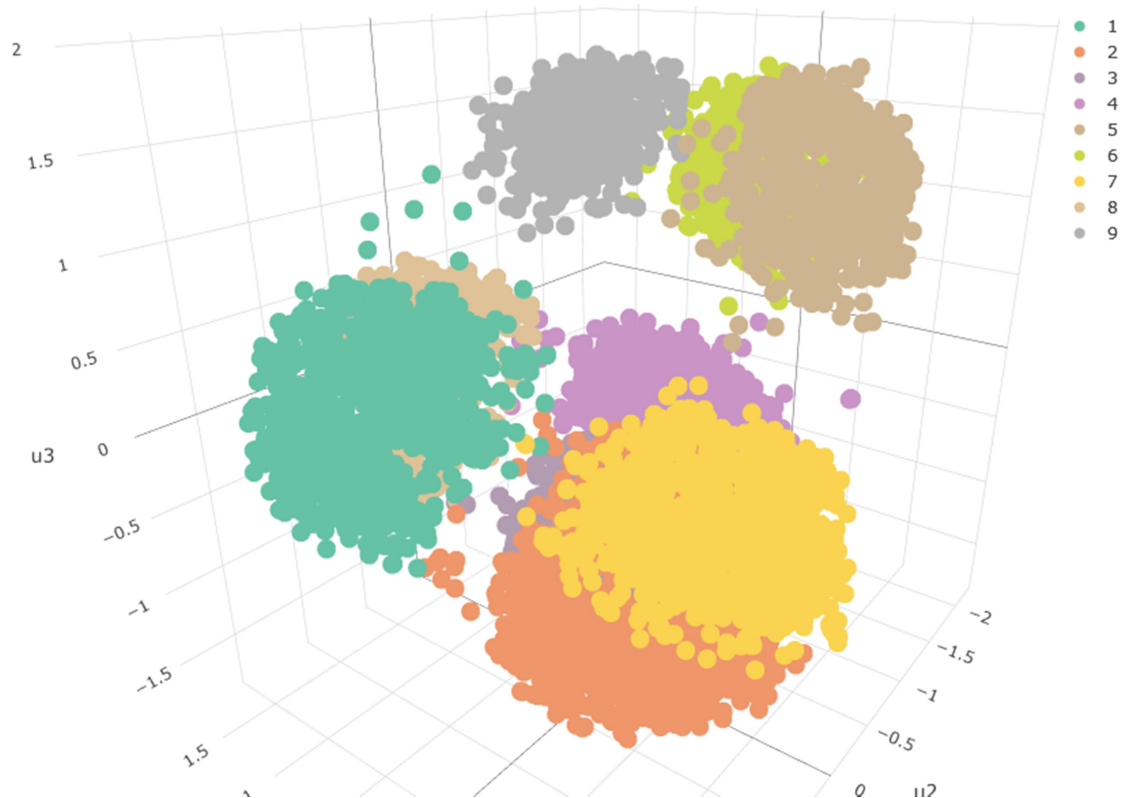
Rycina 43. Klastrowanie dla elementu ontologii „wiązanie z białkami MHC klasy I”.

4.2.5. Zróżnicowanie genów kodujących nadrodzinę białek ABC w populacji polskiej

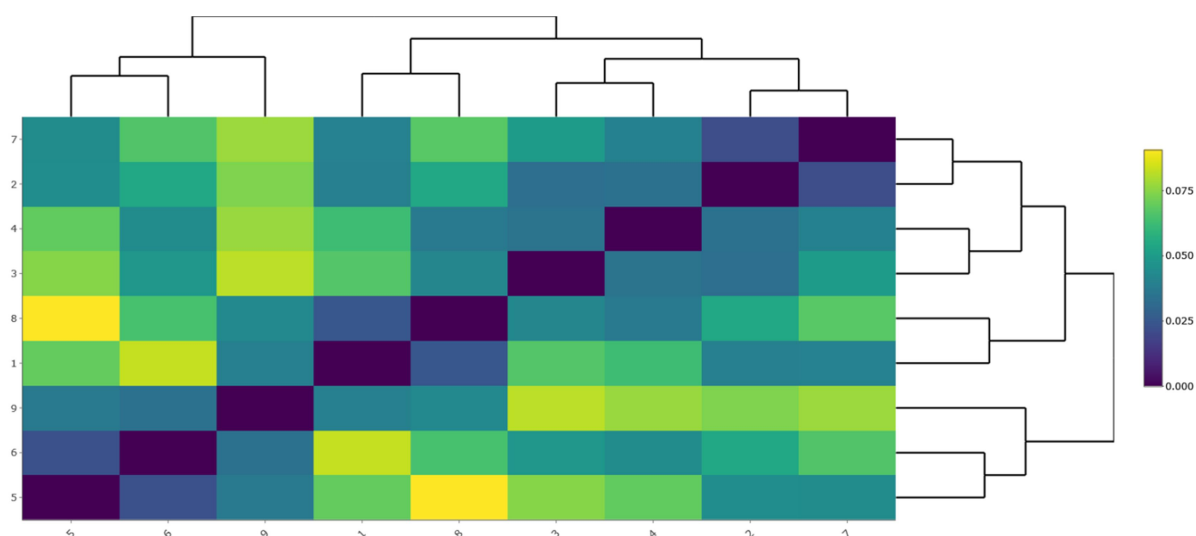
Klastrowaniu poddano kohortę zgenotypowanych 6660 próbek pochodzących od ochotników reprezentujących populację polską. Analiza 24824 współdzielonych wariantów wykazała grupowanie na cztery główne skupienia próbek, co odpowiadało wynikowi dla wariantów częstych w populacjach europejskich. Już na tym etapie możliwe było więc wyznaczenie optymalnej liczby klastrow, których ilość najlepiej oddawała model grupowania próbek. I w tym przypadku wyznaczono ich dziewięć (Ryc. 44). Zróżnicowanie genetyczne pomiędzy poszczególnymi klastrami pokazała dodatkowo analiza F_{st} , w której klastry istotnie ($\geq 0,05$) różniły się od tych położonych na przeciwnych płaszczyznach (Ryc. 45). W przypadku analizy próbek na podstawie 3602 wariantów częstych ($MAF \geq 0,05$) otrzymano wynik odpowiadający powyższemu dla wszystkich wariantów, choć zgrupowania były wyraźniej oddzielone przez mniejszą ilość próbek na ich granicach (Ryc. 46). Porównując jednak ten wynik z analogicznym dla populacji europejskich, to granice pomiędzy poszczególnymi zgrupowaniami w tym przypadku nie były aż tak wyraźnie zarysowane, ale jest to prawdopodobnie jedynie efekt ogromnego zwiększenia skali próbek wziętej do analizy, co obniża rozdzielczość rozdziału. Podobnie i tym razem, otrzymano model optymalnej ilości dziewięciu klastrow. Jedno zgrupowanie tworzyły cztery klastry: nr 7 i 2, z zaznaczoną granicą pomiędzy klastrami nr 3 i 4. Wyraźniej od niego

oddzielone znajdowało się zgrupowanie klastrów nr 1 i 8. Z kolei pozostałe zgrupowania tworzyły klastry nr 5 i 6 oraz samodzielne z klastrem nr 9.

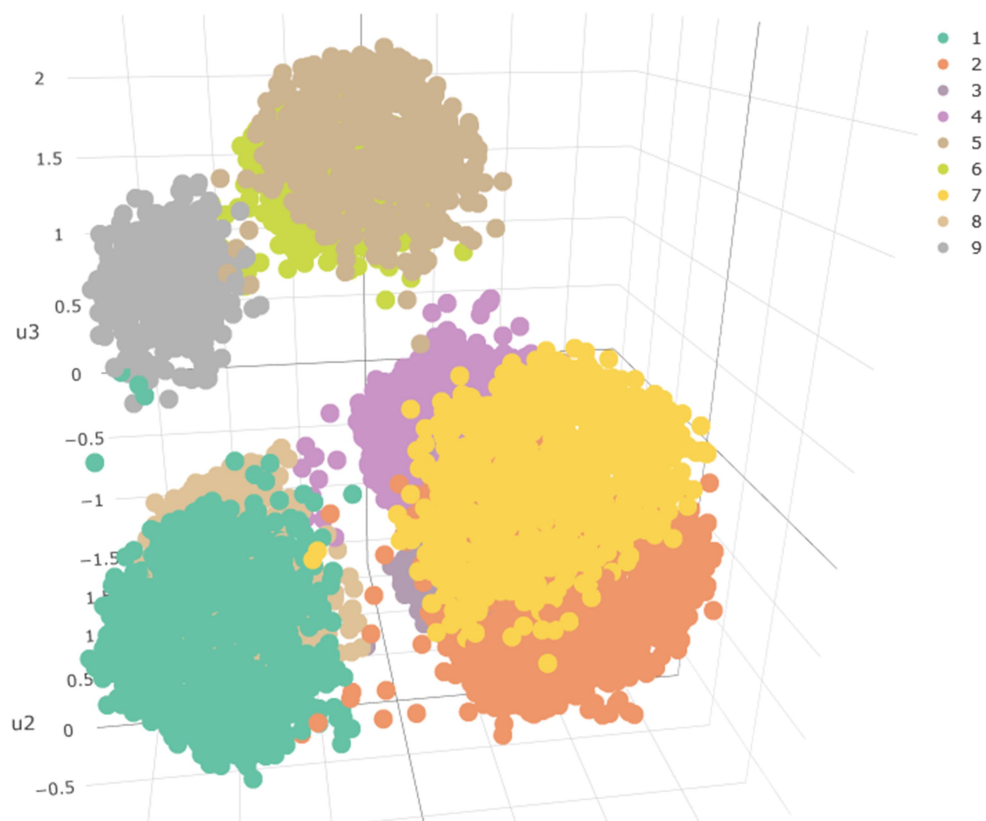
Klastrowanie 21222 wariantów typu rzadkiego (MAF <0,05), podobnie jak przy porównaniu populacji europejskich, zupełnie nie odwzorowało przyporządkowania próbek do klastrów z grupowania wariantów częstych (Ryc. 47). Otrzymano zatem jedno główne zgrupowanie próbek oraz kilka zdecydowanie mniejszych, daleko odseparowanych.



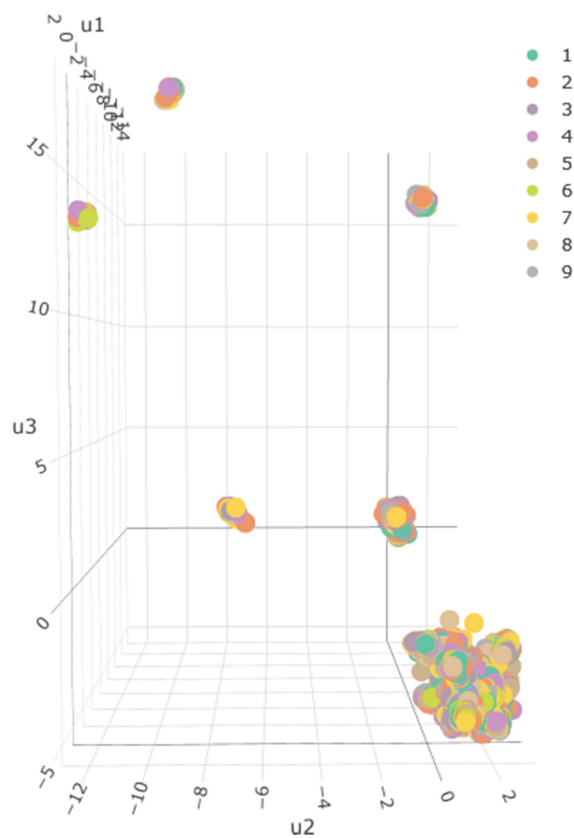
Rycina 44. Klastrowanie populacji polskiej dla wszystkich współdzielonych wariantów.



Rycina 45. Analiza F_{st} różnicowania genetycznego pomiędzy zidentyfikowanymi klastkami w populacji polskiej.



Rycina 46. Klastrowanie populacji polskiej dla współdzielonych wariantów częstych.



Rycina 47. Klastrowanie populacji polskiej dla współdzielonych wariantów rzadkich.

4.2.6. Identyfikacja zmienności w genomie związanej z otrzymanym klastrowaniem

Dla wyników klastrowania populacji polskiej przeprowadzono badanie asocjacyjne całego genomu (GWAS, ang. *Genome-Wide Association Study*), w celu zidentyfikowania zmienności w specyficznych *loci*, jaka odpowiada za otrzymany obraz klastrowania. Pozwoliło to też jednoznacznie wyjaśnić kwestię unikatowości wyniku klastrowania wyłącznie dla genów ABC. Zidentyfikowano zatem warianty ($MAF \geq 0,05$) i geny, w jakich one lokują, które są związane z wyodrębnieniem każdego z dziewięciu klastrow (Tab. 10). W analizie stwierdzono największą ilość wariantów zlokalizowanych w genach ABC, która miała taki efekt: *ABCB5*, *ABCC4*, *ABCC8*, *ABCC11*, *ABCC12*. Co więcej, udział tych wariantów był zdecydowanie najbardziej istotny w klastrowaniu, osiągając dla niektórych często wartości rzędu $p < 10^{-100}$. Warianty zlokalizowane w innych genach, zazwyczaj w tym samym *locus* co geny kodujące białka ABC, były pojedynczymi zdarzeniami i najczęściej miały marginalnie istotne znaczenie ($p < 10^{-5} - 10^{-6}$). Wyjątkowo istotne na tym tle (p dochodzące rzędu 10^{-216}) były jedynie po dwa warianty zlokalizowane w genach *LONP2* i *SIAH1*, różniące każdy klaster. Pozostałe przypadki stanowiły warianty w przestrzeni międzygenowej, osiągając jednak w pojedynczych przypadkach istotną wartość $p < 10^{-20}$.

Tabela 10. Ilość wariantów ($MAF \geq 0,05$) w genomie na kolejnych chromosomach, które istotnie ($p < 10^{-5}$) różnicują poszczególne klastry między sobą dla populacji polskiej. Wyszczególniono geny ABC.

Chr. ↓	Ilość wariantów ($MAF \geq 0,05$), które istotnie ($p < 10^{-5}$) różnicują poszczególne klastry (1 – 9) między sobą								
	1	2	3	4	5	6	7	8	9
N →	1051	1160	510	723	595	584	991	516	530
1		3	4					2	2
2		2	2		2	1	1	2	1
3	1		6	6			2	2	2
4		2			6	3	2	2	
5	1		3		3		1	2	
6	1	4	1	2	1	1	4	1	1
7	28 <i>ABCB5</i> : 20	29 <i>ABCB5</i> : 19	55 <i>ABCB5</i> : 28	46 <i>ABCB5</i> : 29	32 <i>ABCB5</i> : 22	26 <i>ABCB5</i> : 21	25 <i>ABCB5</i> : 21	33 <i>ABCB5</i> : 23	1
8	1	1			2	1	2	4	1
9	1		2	1		3	3	2	1
10		2	1		1		1	2	3
11		25 <i>ABCC8</i> : 14	2	1	2	3	28 <i>ABCC8</i> : 13	7	2
12		3 <i>ABCC9</i> : 1	2 <i>ABCC9</i> : 2	3	5		1		2
13	50 <i>ABCC4</i> : 33	30 <i>ABCC4</i> : 23	14 <i>ABCC4</i> : 14	20 <i>ABCC4</i> : 20	27 <i>ABCC4</i> : 19	21 <i>ABCC4</i> : 19	22 <i>ABCC4</i> : 20	29 <i>ABCC4</i> : 25	40 <i>ABCC4</i> : 26
14		1	1	1	2	1	3		
15	1						1		2
16	33 <i>ABCC11</i> : 11 <i>ABCC12</i> : 2	41 <i>ABCC11</i> : 11 <i>ABCC12</i> : 2	19 <i>ABCC11</i> : 10 <i>ABCC12</i> : 2	21 <i>ABCC11</i> : 10 <i>ABCC12</i> : 2	76 <i>ABCC11</i> : 14 <i>ABCC12</i> : 2	74 <i>ABCC11</i> : 14 <i>ABCC12</i> : 2	37 <i>ABCC11</i> : 10 <i>ABCC12</i> : 2	19 <i>ABCC11</i> : 10 <i>ABCC12</i> : 2	60 <i>ABCC11</i> : 14 <i>ABCC12</i> : 2
17			2						1
18		1	2	3				4	
19								1	1
20						2		2	
21								2	1
22			1		1				1

4.2.7. Różnice w ekspresji genów w narządach lub ich strukturach w poszczególnych klastrach

W analizie asocjacyjnej całego transkryptomu (TWAS, ang. *Transcriptome-Wide Association Study*) zidentyfikowano geny, które wykazały istotnie ($p < 0,05$) zmienioną ekspresję w każdym z dziewięciu klastrów w stosunku do pozostałych, w specyficznych narządach lub ich strukturach z 44 rodzajów z bazy danych *GTEX* (Tab. 11). Spośród genów ABC najbardziej specyficznie w poszczególnych klastrach zachodziła ekspresja *ABCC4*, najistotniej w tarczycy, krwi pełnej i okrężnicy esicy, oraz *ABCC8*, dla którego istotne zmiany w ekspresji były obserwowane tylko w klastrach nr 2 i 7, najczęściej skorelowana z tymi dla *KCNJ11*, dla tkanek okrężnicy, przetyku, serca, nerwu, mięśni szkieletowych, niektórych struktur mózgu, tętnic, tarczycy i tkanki tłuszczowej. Ponadto we wszystkich klastrach istotnie zmieniona była ekspresja genów z tego samego *locus*: *ABCC11*, *LONP2* i *SIAH1*. Prawdopodobnie są to więc geny ulegające konstytutywnej ekspresji w większości rodzajów tkanek i są ze sobą sprzężone. Warianty jednak, które wpływają na ich ekspresję bardzo dobrze rozdzielają poszczególne klastry. Pozostałe geny, których zmiany ekspresji odnotowano w poszczególnych klastrach, pełnią różnorodne funkcje komórkowe, bezpośrednio niepowiązane ze sobą jak: regulacja ubikwitynacji, tworzenie cytoszkieletu, proliferacja lub różnicowanie komórek albo też ich funkcje nie zostały jeszcze dokładnie opisane.

Tabela 11. Ilość narządów i struktur z bazy *GTEX*, w których ekspresja genów była istotnie ($p < 0,05$) zmieniona w każdym klastrze. Z wyjątkiem genów kodujących białka ABC podano, jeśli liczba ≥ 4 .

Chr. ↓	Gen ↓	Ilość struktur, w których ekspresja genów była istotnie (p <0,05) zmieniona w każdym klastrze (1 – 9)									
		Klaster →	1	2	3	4	5	6	7	8	9
13	ABCC4		12	6		3	4	5	4	6	8
11	ABCC8			23					24		
16	ABCC11		4	4	4	4	4	4	4	4	4
13	DZIP1										5
22	FBXO7				4						
11	KCNJ11			12					14		
16	KRBOX5			4			11	10			5
16	LONP2		31	31	31	31	31	31	31	31	31
17	MIS12						6				
2	NBAS				6						
11	NCR3LG1								7		
11	PLEKHA7			6					5		
14	PLEKHH1					4					
12	PRDM4						5				
7	SGCE								4		
16	SHCBP1		4	4		4	4	4	4		4
16	SIAH1		5	5	4	4	5	5	4	4	5
6	SMIM8					15					
13	UGGT2										5

DZIP1- ang. DAZ interacting zinc finger protein 1; **FBXO7** - ang. F-Box Protein 7; **KCNJ11** - ang. potassium inwardly rectifying channel subfamily J member 11; **KRBOX5**- ang. KRAB box domain containing 5; **LONP2** - ang. lon peptidase 2, peroxisomal; **MIS12** - ang. kinetochore complex component; **NBAS** - ang. NBAS subunit of NRZ tethering complex; **NCR3LG1** - ang. natural killer cell cytotoxicity receptor 3 ligand 1; **PLEKHA7** - ang. pleckstrin homology domain containing A7; **PLEKHH1** - ang. pleckstrin homology, MyTH4 And FERM domain containing H1; **PRDM4** - ang. PR/SET domain 4; **SGCE** - ang. sarcoglycan epsilon; **SHCBP1** - ang. SHC binding and spindle associated 1; **SIAH1** - ang. siah E3 ubiquitin protein ligase 1; **SMIM8** - ang. small integral membrane protein 8; **UGGT2** - ang. UDP-glucose glycoprotein glucosyltransferase 2.

4.2.8. Analiza związku klastrowania populacji polskiej z fenotypem

Dysponując jednorodnymi danymi ankietowymi od 5656 ochotników z populacji polskiej, w tym na temat chorób cywilizacyjnych, dokonano analizy ich związku z obserwowanym klastrowaniem (Tab. 12). Stwierdzono, że dystrybucja wariantów w poszczególnych dziewięciu klastrach była istotnie związana z cukrzycą i dotyczyła zwłaszcza klastrów nr 4 i 7. Na granicy istotności zaobserwowano również związek z chorobami reumatycznymi, szczególnie w klastrze nr 5. Nie stwierdzono natomiast korelacji z klastrowaniem dla miażdżycy i astmy, a w przypadku kilku innych interesujących chorób jak choroba Alzheimera, dotyczyły one jednostkowych przypadków i nie można było na ich podstawie dokonać wiarygodnej oceny istotności, dlatego zostały pominięte.

Tabela 12. Częstość dawców próbek z populacji polskiej w odpowiednich klastrach, deklarujących lub nie deklarujących poszczególne choroby.

Odpowiedzi ankietowe ↓	Częstość dawców próbek deklarujących chorobę lub nie, w poszczególnych klastrach (1-9)								
	1	2	3	4	5	6	7	8	9
N →	903	971	439	620	496	504	847	440	436
Cukrzyca (test chi-kwadrat = 18,9860; p = 0,0149)									
tak	0,0421	0,0391	0,0410	0,0677	0,0423	0,0357	0,0661	0,0318	0,0482
nie	0,9579	0,9609	0,9590	0,9323	0,9577	0,9643	0,9339	0,9682	0,9518
Miażdżycza (test chi-kwadrat = 2,7626; p = 0,9484)									
tak	0,0377	0,0278	0,0364	0,0323	0,0363	0,0258	0,0319	0,0341	0,0298
nie	0,9623	0,9722	0,9636	0,9677	0,9637	0,9742	0,9681	0,9659	0,9702
Astma (test chi-kwadrat = 7,9589; p = 0,4375)									
tak	0,0520	0,0505	0,0387	0,0403	0,0585	0,0337	0,0531	0,0341	0,0505
nie	0,9480	0,9495	0,9613	0,9597	0,9415	0,9663	0,9469	0,9659	0,9495
Choroby reumatyczne (test chi-kwadrat = 15,5280; p = 0,0497)									
tak	0,0952	0,1020	0,1139	0,0774	0,1411	0,0913	0,1051	0,0864	0,0986
nie	0,9048	0,8980	0,8861	0,9226	0,8589	0,9087	0,8949	0,9136	0,9014

5. DYSKUSJA

Białka ABC są największą i bardzo starą ewolucyjnie rodziną białek transbłonowych, a geny je kodujące wykazują niezwykle zakonserwowanie od bakterii po ludzi (Dean i in., 2022). Uważa się, że wyewoluowały w odpowiedzi na substancje toksyczne i inne warunki środowiskowe, na które narażony jest każdy gatunek (Silverton i in., 2011). Sugeruje to, że prawidłowe funkcjonowanie transporterów ABC ma kluczowe znaczenie w utrzymaniu właściwej kondycji organizmu, co najlepiej potwierdza fakt, że do tej pory zidentyfikowano mutacje powodujące choroby jednogenowe w aż ponad 20 genach ABC (Dean i in., 2022; Silverton i in., 2011).

W publikacjach wchodzących w skład niniejszej rozprawy przeprowadzono analizę zmienności genetycznej prawie całych sekwencji kodujących w genach *ABCC1* i *ABCG2* u ochotników z polskiej populacji. Jedynie kilka wariantów znalezionych w eksonach obu genów występowało z częstością $MAF \geq 0,05$, a spośród nich tylko pojedyncze były zmianami niesynonimicznymi. W obu przypadkach większość ogółem wykrytych wariantów stanowiły zatem te zlokalizowane w intronach, mimo że skanowaniu poddano tylko krótkie ich fragmenty okalające eksony. Z kolei w badaniach nieopublikowanych włączonych do niniejszej rozprawy, wariantów częstych w sekwencjach kodujących jak i niekodujących, a pokrywających się między populacją polską i danymi z bazy *1000 Genomes Project*, zidentyfikowano 3962, natomiast rzadkich 15191 ($MAF < 0,05$). Należy mieć jednak na uwadze, że do tego badania były wykorzystane tylko specyficzne warianty naniesione przez producenta na mikromacierzach, a nie zmienność z pełnych sekwencji genów, stąd szacunek zwłaszcza zmienności rzadkiej, która jest typowo osobnicza, może być mocno zaniżony.

Zmienność typu rzadkiego w genach kodujących białka ABC w populacji polskiej jest zatem zdecydowanie bardziej powszechna niż zmienność typu częstego więc tym samym nie było podstaw do odrzucenia hipotezy badawczej nr 1. Wniosek ten jest zgodny z wynikiem jedyne do tej pory badania, charakteryzującego geny całej nadrodziny białek ABC u 138632 dawców z siedmiu różnych populacji etnicznych, a głównie ich eksomy (pełne sekwencje kodujące). Stwierdzono, że aż 98,5% zmienności stanowiła ta rzadka, przy czym autorzy przyjęli próg $MAF < 0,01$. To dało ostatecznie wynik tylko 917 wariantów częstych ($MAF \geq 0,01$) w sekwencjach kodujących wszystkie te geny (Xiao i in., 2020a). Także w badaniu większej grupy farmakogenów, w którym uwzględniono 22 kodujące transportery ABC, zmienność rzadka ($MAF < 0,01$) dotyczyła 92,9% wariantów (Kozyra i in., 2017). Taki wynik sumarycznie pokrywa się z ogólnym obrazem zmienności pełnego genomu wskazującym, że wszystkie populacje wykazują wzbogacenie rzadkich wariantów, odzwierciedlające ewolucyjnie stosunkowo niedawny, gwałtowny wzrost wielkości populacji, skutki zróżnicowania geograficznego i środowiskowego (Abecasis i in., 2012). Nie jest zaskakujące zatem, że obserwowana zmienność w genach kodujących białka ABC i innych farmakogenach podlegała tym samym procesom. Stopień zróżnicowania rzadkich wariantów, w wyniku takich zdarzeń jak efekt wąskiego gardła lub mieszanie

się z innymi populacjami, może być jednak specyficzny w zależności od populacji (Abecasis i in., 2012). Zazwyczaj zmienność rzadka jest jednak wynikiem niedawnych mutacji, stąd przypuszcza się, że są one skupione geograficznie lub nawet prywatne dla określonych populacji (Nelson i in., 2012). Taki wynik zaobserwowano we wspomnianym wcześniej badaniu Xiao i in. (2020a), w którym aż 76% funkcjonalnych, przypuszczalnie szkodliwych wariantów (ang. *putatively deleterious variants*) w genach ABC, było specyficznych tylko dla jednej populacji, a prawie połowę z nich wykryto w europejskiej. Ten rodzaj zmienności może być zatem przyczyną osobniczej odpowiedzi na farmakoterapię (Ingelman-Sundberg i in., 2018; Kozyra i in., 2017).

Wy tłumaczeniem zaobserwowanej dystrybucji wariantów może być rzadsze występowanie w populacji tych, które wpływają niekorzystnie na funkcjonalność białek niż tych, które nie mają takiego efektu (Silverton i in., 2011). Ten wniosek wspierałaby ograniczona zmienność w tak zakonserwowanych i kluczowych domenach dla prawidłowego działania białek jak NBD (Dean i Annilo, 2005). Potwierdza to zwłaszcza fakt, że mutacje w wielu genach ABC są związane z tak poważnymi chorobami lub wpływem na odpowiedź farmakoterapii. Rzadka częstość występowania wariantów redukujących prawidłowe działania białka sugeruje, że takie warianty nie mają korzystnego ewolucyjnie fenotypu, a zatem nie podlegają pozytywnej presji selekcyjnej (Silverton i in., 2011). To ostatecznie prowadzi do wniosku, że zwłaszcza sekwencje genów ABC, są mocno zakonserwowane ewolucyjnie.

Porównanie wyników wykrytej zmienności *ABCC1* i *ABCG2* w populacji polskiej z danymi literaturowymi podejmującymi ten temat dla innych populacji wykazało istotne różnice w częstości występowania specyficznych wariantów. Ujawniły się one szczególnie pomiędzy wartościami MAF w populacji polskiej, a tymi uzyskanymi w populacjach o innym pochodzeniu etnicznym czyli wschodnioazjatyckimi takimi jak japońska, chińska i koreańska, oraz tymi wykrytymi w kohortach dawców o pochodzeniu afrykańskim. Nie zaobserwowano jednak istotnych różnic w stosunku do wyników dla innych populacji odmiany białej. Jest to zgodne z innymi, wcześniejszymi badaniami porównującymi częstość wariantów u tych populacji (Ileiri, 2012; Kurose i in., 2012; Wang i in., 2006). Wynik ten ponadto został potwierdzony w badaniach nieopublikowanych, wchodzących w skład niniejszej rozprawy. W analizie klastrowania populacji polskiej oraz populacji referencyjnych z całego świata uzyskano wyraźne grupowanie poszczególnych populacji w obrębie regionu geograficznego, z którego pochodziły. Populacja polska na tym tle była istotnie oddzielona od populacji z innych regionów geograficznych, a skupiona razem z pozostałymi europejskimi, a zatem nie było podstaw do odrzucenia hipotezy badawczej nr 2. Otrzymany wynik dla genów ABC oddaje zatem naturalny obraz zmienności całogenomowej między populacjami o różnym pochodzeniu, która wskazuje na efekt silnego i długotrwałego wąskiego gardła w przeszłości, po którym nastąpił niezwykle szybki wzrost

populacji spoza Afryki i genetyczne wyodrębnienie się populacji europejskich, azjatyckich i amerykańskich (Auton i in., 2015).

W kolejnej analizie przeprowadzono podobne klastrowanie, jednak zawężone już tylko do populacji polskiej i pozostałych europejskich. Otrzymany wynik nie był jednak w ogóle związany z pochodzeniem populacyjnym dawców próbek ponieważ byli oni segregowani do zupełnie pod tym względem niezależnych klastów, przez co hipotezę badawczą nr 3 należy odrzucić. Obserwowana zmienność nie jest zatem zależna od populacji, a raczej ma charakter osobniczy. Wynik ten był nieoczekiwany ponieważ opisywane były przesłanki sugerujące, że różne populacje mają stosunkowo dużą liczbę wariantów, które zazwyczaj są genetycznie lub geograficznie odrębne w obrębie ich grupy kontynentalnej, co świadczyłoby o stosunkowo niedawnych adaptacjach do lokalnych zmian w środowisku (Auton i in., 2015; Wang i in., 2007). Niemniej jednak podkreślano też, że liczba genów wykazujących silne zróżnicowanie między blisko spokrewnionymi populacjami nie jest wielka (Auton i in., 2015). W odniesieniu do takich genów można było jednak przypuszczać, że ich lista obejmuje te kodujące nadrodzinę białek ABC, ze względu na pełną gamę kluczowych funkcji jakie pełnią, także tych o znaczeniu adaptacyjnym jak chociażby presja selekcyjna związana z usuwaniem ksenobiotyków. Można było zatem oczekiwać, że analiza szczegółowej zmienności genetycznej całej rodziny genów wykaże specyficzne różnice między populacjami europejskimi. Klastrowanie tego jednak nie potwierdziło, mimo nawet stosunkowo znacznej rozbieżności geograficznej i historycznej między badanymi populacjami europejskimi. To zaskakujący wynik chociażby dla populacji fińskiej, wykazującej znacznie niższy poziom współdzielenia wariantów rzadkich z innymi populacjami, szczególnie południowoeuropejskimi (Nelson i in., 2012).

W celu wyjaśnienia potencjalnego podłoża takiego klastrowania wśród populacji europejskich, przeprowadzono zatem najpierw odrębne klastrowania w oparciu o zmienność częstą ($MAF \geq 0,05$) i rzadką ($MAF < 0,05$). Uzyskany wynik dla zmienności częstej uwydatniał wyraźnie lepszą rozdzielczość widocznych wcześniej zgrupowań dla pełnej zmienności. Natomiast zmienność rzadka, nie wykazała takiego związku ani także populacyjnie specyficznego klastrowania, mimo swojego prawdopodobnie stosunkowo niedawnego pochodzenia i zwiększonego poziomu zróżnicowania w poszczególnych populacjach (Abecasis i in., 2012). Ponadto próbki tworzące każdy z klastów dla zmienności rzadkiej były rozlokowane w innych klastach zmienności częstej, świadcząc o niezależnym dziedziczeniu tych wariantów. Mimo więc swojego opisywanego, potencjalnie istotnego wpływu na prawidłowe funkcjonowanie białka i farmakoterapię (Ingelman-Sundberg i in., 2018; Kozyra i in., 2017; Zhou i in., 2021), to zmienność rzadka nie jest związana z otrzymanym obrazem populacyjnego klastrowania.

Warianty, które istotnie wpływały na obraz klastrowania populacji europejskich, zidentyfikowano z różnym rozkładem w ośmiu genach ABC. Niezależne od tego wykonane

klastrowania próbek na podstawie elementów ontologii, z jakimi są związane odpowiednie geny ABC, odzwierciedlają ten sam model klastrowania dla funkcji dotyczących transportu kwasów żółciowych, anionów nieorganicznych i S-koniugatów glutationu. Wyniki z obu podejść spójnie sugerują zatem, że za otrzymanym obrazem klastrowania genów ABC w populacjach europejskich w dużej mierze stoi zmienność tych, które są zaangażowane w procesy związane z homeostazą kwasów żółciowych. Odpowiadają one za podział na dwa zgrupowania próbek, gdyż prawdopodobnie to zmienność genu *ABCC11* jest czynnikiem, który powoduje dodatkowy rozdział populacji europejskich z dwóch zgrupowań, na cztery. Oznaczałoby to współistnienie dwóch, niezależnych mechanizmów klastrujących populacje europejskie na podstawie zmienności genów ABC, a *ABCC11* mógłby zatem być wówczas związany z oboma z nich.

Literatura źródłowa obszernie opisuje mechanizmy, w których kwasy żółciowe są nie tylko trawinnymi związkami powierzchniowo czynnymi, ale także ważnymi cząsteczkami sygnalizacyjnymi, które stymulują kilka szlaków sygnałowych w celu regulacji niektórych, kluczowych procesów biologicznych, w tym metabolizmu lipidów i glukozy oraz homeostazy energetycznej (Ding i in., 2015; Jin i in., 2019; Prawitt i Staels, 2010; Staels i in., 2010). Głównie działają jako ligandy dla jądrowego receptora farnesoidowego X (FXR, ang. *Farnesoid X Receptor*) oraz receptora kwasu żółciowego 1 sprzężonego z białkiem G (GPBAR1, ang. *G Protein-coupled Bile Acid Receptor 1*), zwanego najczęściej TGR5 (ang. *Takeda G protein Receptor 5*) (Chávez-Talavera i in., 2017; Shapiro i in., 2018). W skrócie, aktywowany w hepatocytach i enterocytach FXR, w złożonej pętli ujemnego sprzężenia zwrotnego, hamuje syntezę kwasów żółciowych, stymuluje ekspresję transporterów kodowanych przez *ABCB11*, *ABCB4*, *ABCG5/ABCG8* i dzięki temu wpływa odpowiednio soli żółciowych, fosfolipidów i cholesterolu do kanałów żółciowych, a także ogranicza wychwyt zwrotny soli żółciowych z krwi (de Aguiar Vallim i in., 2013). Badania sugerowały też, że aktywacja FXR obniża poziom glukozy w osoczu poprzez uwrażliwienie na działanie insuliny, a także zmniejsza poziom glukozy-6-fosfatazy czym hamuje glukoneogenezę i zwiększa syntezę glikogenu w wątrobie (Ding i in., 2015; Shapiro i in., 2018). Z kolei aktywacja TGR5 powoduje internalizację receptora, zwiększenie wewnątrzkomórkowego poziomu cAMP i aktywację kinazy białkowej A, co prowadzi do fosforylacji białek docelowych, w tym odpowiednich czynników transkrypcyjnych (Ding i in., 2015; Kuipers i in., 2014; Staels i in., 2010). Skutki są tkankowo specyficzne i obejmują działanie przeciwzapalne, rozluźnienie pęcherzyka żółciowego, przyspieszoną perystaltykę jelit, zwiększony wydatek energetyczny w brunatnej tkance tłuszczowej i mięśniach poprzez uwalnianie glukagonopodobnego peptydu 1 (GLP-1) w komórkach L jelit i stymulację produkcji insuliny w komórkach trzustki (Kuipers i in., 2014; Molinaro i in., 2018). U pacjentów z hipercholesterolemią podawanie sekwestrantów wiążących kwasy żółciowe, w celu zwiększenia poziomu wydalania ich z kałem, powodowało zwiększenie ich syntezy i tym samym obniżenie poziomu lipoprotein o małej gęstości (LDL, ang. *Low-Density Lipoprotein*) w osoczu,

a u pacjentów z cukrzycą typu 2, obniżenie poziomu glukozy w osoczu i zwiększoną wrażliwość na insulinę (Staels i in., 2010). W doświadczeniach na myszach wykazano, że dieta wysokotłuszczowa i zwiększenie we krwi poziomu kwasów żółciowych aktywowały TGR5, co poprzez stymulację cAMP prowadziło do zwiększenia poziomu aktywnego hormonu tarczycy i powodowało podniesienie poziomu intensywności metabolizmu glukozy oraz zmniejszenie otyłości i insulinooporności (de Aguiar Vallim i in., 2013; Kuipers i in., 2014; Shapiro i in., 2018).

Większość z tych wyników pozyskano jednak z badań na modelach mysich, u ludzi mechanizmy te nie zostały jeszcze wystarczająco dobrze zbadane (Hansen i in., 2014; Molinaro i in., 2018). Na podstawie molekularnych zależności jakie uzyskano w wynikach rozprawy można jednak przypuszczać, że istnieją uzasadnione przesłanki do funkcjonowania takich procesów. Bezpośrednio połączone z nimi może być bowiem funkcjonowanie białek ABC i zmienność genów je kodujących, w tym wykryte w niniejszej rozprawie jako związane z klastrowaniem populacji europejskich. *ABCB11*, *ABCG5/ABCG8* poprzez wydzielanie soli żółciowych i cholesterolu, *ABCC8* przez stymulowanie wydzielania insuliny, a *ABCC4* i *ABCC11* przez udział w procesie wydzielania żółci lub poprzez transport cAMP w szlakach sygnałowych. To z kolei sugerowałoby wniosek, że u Europejczyków, bez względu na populację, utrwalone są dwa różne fenotypy związane z homeostazą lipidowo-glukozową. Oznaczałoby to zatem, że ich zakonserwowanie jest starsze niż nastąpił rozdział na poszczególne populacje odmiany białej człowieka. Ten przypuszczalny model może świadczyć o doborze, który zachowuje w populacji allele dostosowane do poziomu metabolizmu, np. w zależności od rodzaju diety lub dostępności pokarmu. Taki rodzaj selekcji można by określić zbalansowanym lub równoważącym (ang. *balancing selection*) i polega na utrzymywaniu w puli genowej populacji alleli z częstością większą niż to wynikałoby z samego dryfu genetycznego (Minias i Vinkler, 2022). Wyjaśnienie jednak mechanizmu jaki stałby u podłoża tej selekcji jest w tym momencie niestety trudne do ustalenia. Co więcej, ten wniosek potwierdzałaby także przeprowadzona analiza fenotypowa dla poszczególnych klastrow w populacji polskiej, na podstawie danych ankietowych dotyczących cywilizacyjnych jednostek chorobowych i wykazany istotny związek z cukrzycą. W dobie współczesnego braku powszechnych problemów z zaspokojeniem potrzeb żywieniowych u populacji pochodzenia europejskiego, a nawet przeciwnie, wobec stylu odżywiania, który prowadzi do demograficznej plagi otyłości i cukrzycy typu 2 (Fang i in., 2022), postulowana zmienność genów kodujących niektóre białka ABC może być więc niespodziewanie istotnym czynnikiem związanym z tymi chorobami cywilizacyjnymi.

Co ciekawe też, polimorfizmy istotnie związane z opisywanym klastrowaniem zlokalizowano niemal wyłącznie w intronach. Warianty w regionach niekodujących mogą jednak wpływać na elementy regulatorowe, takie jak promotory, wzmacniacze, wyciszacze i izolatory, które z kolei mogą zmieniać ich powinowactwo do czynnika transkrypcyjnego lub zmieniać lokalną strukturę

chromatyny. Z kolei warianty w miejscach splicingu mogą go zakłócać, prowadząc do nieprawidłowego składania transkryptów i ich usuwania w procesie NMD (ang. *Nonsense-Mediated Decay*) lub skutkować produkcją dysfunkcyjnych białek (Zhou i in., 2018). Dobrym przykładem jest analiza miejsc wiązania opisywanego receptora FXR, obejmująca cały genom, która ujawniła, że tylko 10% specyficznych elementów odpowiedzi DNA (FXRE) jest obecnych w proksymalnych promotorach docelowych genów, podczas gdy większość (90%) znajdowała się w intronach lub regionach międzygenowych (de Aguiar Vallim i in., 2013).

W niniejszej rozprawie stwierdzono również, że warianty te mogą mieć wpływ na ekspresję genów zależną od narządu lub jego struktury, nie tylko tych kodujących białka ABC, ale też innych np. związanych z gospodarką glukozową jak glukozo-6-fosfataza czy podjednostka Kir6.2. O ile dla niektórych z pozostałych genów ciężko jest wykazać takie powiązanie z funkcją jaką pełnią i może to być jedynie efekt wspólnych *loci* z genami zawierającymi te warianty, tak wpływ na ekspresję w wątrobie, zwłaszcza jednego lncRNA, może już być nieprzypadkowy. Takie niekodujące transkrypty odgrywają ważne role w wielu procesach, takich jak transkrypcja, splicing, translacja mRNA i modyfikacje epigenetyczne, ze względu na swoją specyficzność interakcji zarówno z DNA, jak i białkami (Arrighi i in., 2016; Hlaváč i in., 2020a). Efekt specyficznej ekspresji genów ABC w poszczególnych narządach lub ich strukturach został ponadto potwierdzony w niniejszej rozprawie w niezależnej analizie TWAS na populacji polskiej. Uzyskano bardzo dobre odwzorowanie analizy przeprowadzonej dla wariantów istotnie związanych z klastrowaniem populacji europejskich, potwierdzając tym samym miejscowo specyficzną modulację ekspresji tych genów.

Narządowo specyficzny wpływ zmienności genów ABC na ekspresję może być ponadto też związany z miejscami ekspresji receptorów kwasów żółciowych. Wspierałoby to postulowany udział tej zmienności w regulacji gospodarki lipidowo-glukozowej. Ekspresja *TGR5* zachodzi bowiem nie tylko w komórkach nieparenchymalnych wątroby, takich jak: komórki nabłonkowe pęcherzyka żółciowego, cholangiocyty (komórki nabłonkowe przewodu żółciowego) czy komórki Kupffera (makrofagi), ale także w wielu innych tkankach, w tym w brązowej tkance tłuszczowej, jelitowych komórkach enteroendokrynnych, które wydzielają inkretyny jak GLP-1, neuronach jelitowych i mózgu (de Aguiar Vallim i in., 2013; Kuipers i in., 2014). Z kolei wysoką ekspresję *FXR* stwierdzono także w nadnerczach, choć nie do końca poznano mechanizm działania receptora w tym gruczole, oraz białej tkance tłuszczowej, sercu, nerkach, żołądku, trzustce, komórkach śródbłonna, mięśniach gładkich naczyń i komórkach układu odpornościowego (de Aguiar Vallim i in., 2013; Kuipers i in., 2014; Shapiro i in., 2018).

Klastrowanie próbek, wykonane dla jedynie polskiej populacji na dużo większej kohorcie, zgodne było z wynikiem uzyskanym w przypadku porównania z populacjami europejskimi. Przede wszystkim jednak, analiza GWAS wykonana dla pełnego genomu wykazała unikatowy model

otrzymanego obrazu klastrowania, który był istotnie związany głównie ze zmiennością genów ABC. W poszczególnych klastrach wykazano przede wszystkim efekt *ABCB5*, *ABCC4*, *ABCC8* i *ABCC11*. Potwierdzałoby to zatem udział w klastrowaniu genów kodujących transportery związane z przekazywaniem sygnału komórkowego przez cAMP i gospodarką glukozową. Roli *ABCB5* w tych zależnościach nie da się jednak jednoznacznie wytłumaczyć, jak to już było opisane, ze względu na obecnie nikłą wiedzę na jego temat. Z kolei w badaniach nad *ABCC11*, porównanie międzypopulacyjne SNP rs17822931 wykazało jego związek z typem woskowiny usznej, co przemawia za geograficznie ograniczoną selekcją tej cechy. Allel częstszy w populacjach azjatyckich w stosunku do pozostałych prawdopodobnie wiązał się z przystosowaniem do wówczas chłodnego klimatu, poprzez mniejszą potliwość (Ohashi i in., 2011; Yoshiura i in., 2006). Inne badania dotyczące wariantów w genach ABC, pod kątem zdarzeń ostatnich pozytywnych selekcji (ang. *recent positive selection*), potwierdzały taki efekt dla *ABCC11*, ale wyniki dla pozostałych genów, w tym *ABCC4* i *ABCC12*, nie dały jednoznacznych rezultatów (Wang i in., 2007). W niniejszej rozprawie nie zidentyfikowano jednak wspomnianego polimorfizmu jako istotnie związanego z klastrowaniem, mimo prawdopodobnego wpływu tego genu na otrzymany obraz grupowania populacji. Możliwy więc jest potencjalny inny efekt jego działania w odmianie białej człowieka, uwarunkowany przez inne warianty, jako że powiązany jest on także z innymi funkcjami gruczołów apokrynowych, jak np. zapachem potu (Yoshiura i in., 2006)

Potwierdzony model selekcji zbalansowanej u człowieka stanowi zmienność genów głównego układu zgodności tkankowej (Pierini i Lenz, 2018). Mimo to, w związku z brakiem wyniku imputacji dla genów kodujących transportery TAP1 i TAP2, a związanych bezpośrednio z tym układem, nie udało się ocenić ich wpływu na przeprowadzone w rozprawie klastrowanie populacji. W innych badaniach stwierdzono, że zmienność w tych genach była bardzo ograniczona lub nawet najniższa spośród całej nadrodziny, co uzasadnia problem metodyczny w niniejszej rozprawie z uzyskaniem wystarczającej ilości wariantów do imputacji (Hlaváč i in., 2020b; Xiao i in., 2020a). To jednak też oznacza, że selekcja genów *MHC* nie idzie w parze z selekcją genów *ABCB2* i *ABCB3*, które wykazały mocne zakonserwowanie, co wyjaśnia negatywny wpływ ich wariantów genetycznych na odporność i procesy zapalne (Hlaváč i in., 2020b). Z kolei najwięcej wariantów było wykrywanych w genach kodujących białka z rodziny ABCA, zaangażowanych w homeostazę lipidów, ale mógł to być jednak jedynie efekt długości tych genów (Hlaváč i in., 2020b; Xiao i in., 2020a). W niniejszej rozprawie nie stwierdzono jednak związku genów z tej rodziny z klastrowaniem populacji, co w połączeniu z ilością powiązanych z nimi chorób sugeruje, że procesy za jakie odpowiadają są niezwykle ważne i zakonserwowane.

Podsumowując, należy też wskazać ograniczenia związane ze skomplikowaniem podjętego w rozprawie zagadnienia oraz metodą jaka była możliwa do zastosowania. Białka ABC są

w zdecydowanej większości transporterami „wielozadaniowymi”, co oznacza, że wykazują działanie pleiotropowe i każdy z nich odpowiada za szereg różnorodnych procesów, a nie tylko pojedynczą funkcję (Cole, 2014; He i in., 2021). Co więcej, wiele z białek ABC pełni tę samą lub zbliżoną funkcję, nawzajem się dopełniając (Juan-Carlos i in., 2021). W związku z tym, niezwykle trudno jest wskazać efekt konkretnego genu lub wariantów jakie odpowiadają za obserwowany fenotyp lub na ich podstawie ten fenotyp jednoznacznie sprecyzować. Ponadto, zastosowana metodyka w badaniach nieopublikowanych bazowała na ogromnej puli specyficznych wariantów. Nie była to jednak pełna zmienność genomowa ani eksomowa, przez co niektóre rejony genów były nie dość wzbogacone w warianty, co wręcz uniemożliwiło zbadanie wpływu niektórych genów na klastrowanie populacji. Ta być może istotna informacja genetyczna mogła więc zostać pominięta w prezentowanych analizach. Szczególnie dotkliwy był zwłaszcza brak możliwości przeanalizowania najważniejszych genów ABC związanych z opornością wielolekową, choć dwa z nich w prezentowanej rozprawie udało się przeanalizować w populacji polskiej z wykorzystaniem metody HRM. Dlatego wydaje się zasadne w przyszłości przeprowadzić reanalizę otrzymanych wyników z wykorzystaniem pełnych sekwencji genów, w celu zwłaszcza dokładniejszego opracowania modelu klastrowania populacji europejskich, pod kątem zmienności w genach kodujących całą nadrodzinę białek ABC.

6. WNIOSKI I WERYFIKACJA HIPOTEZ BADAWCZYCH

W prezentowanej rozprawie zrealizowano wszystkie postawione cele szczegółowe. Pozwoliły one sformułować wnioski i zweryfikować hipotezy badawcze.

Hipoteza badawcza nr 1 – nie było podstaw do odrzucenia

Zdecydowaną większość zidentyfikowanej zmienności w genach kodujących białka ABC stanowi ta o charakterze rzadkim (MAF <0,05). Świadczy to o wysokim zakonserwowaniu ich sekwencji, ze względu na wiele istotnych funkcji dla prawidłowego funkcjonowania organizmu.

Hipoteza badawcza nr 2 – nie było podstaw do odrzucenia

Populacja polska, pod względem zmienności w genach kodujących białka ABC, wraz z pozostałymi populacjami o pochodzeniu europejskim, istotnie różni się od populacji o innych pochodzeniu etnicznym, jak populacje wschodnioazjatyckie, afrykańskie czy amerykańskie.

Hipoteza badawcza nr 3 – została odrzucona

Populacja polska, pod względem zmienności w genach kodujących białka ABC, nie różni się istotnie od innych populacji referencyjnych o pochodzeniu europejskim. Obserwowane klastrowanie populacji europejskich nie jest związane z pochodzeniem. Wykazano, że istotnie odpowiadają za nie warianty częste, zlokalizowane głównie w sekwencjach niekodujących kilku genów związanych z homeostazą kwasów żółciowych, transportem cAMP oraz gospodarką glukozową. Zaproponowano na tej podstawie niezależny od populacji utrwalony mechanizm współistnienia różnych fenotypów, związanych z gospodarką lipidowo-glukozową i regulowany przez kwasy żółciowe, w czym istotny udział miałyby transportery z nadrodziny ABC. Wykazano, że obserwowane klastrowanie populacyjne jest unikatowe bo istotnie związane w głównej mierze ze zmiennością genów kodujących białka ABC.

7. BIBLIOGRAFIA

- Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. i McVean, G.A., 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422): 56-65.
- Arrigoni, E., Galimberti, S., Petrini, M., Danesi, R. i Di Paolo, A., 2016. ATP-binding cassette transmembrane transporters and their epigenetic control in cancer: an overview. *Expert Opin. Drug Metab. Toxicol.*, 12(12): 1419-1432.
- Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. i Abecasis, G.R., 2015. A global reference for human genetic variation. *Nature*, 526(7571): 68-74.
- Backstrom, G., Taipalensuu, J., Melhus, H., Brandstrom, H., Svensson, A.C., Artursson, P. i Kindmark, A., 2003. Genetic variation in the ATP-binding cassette transporter gene ABCG2 (BCRP) in a Swedish population. *Eur. J. Pharm. Sci.*, 18(5): 359-64.
- Bosch, T.M., Kjellberg, L.M., Bouwers, A., Koeleman, B.P., Schellens, J.H., Beijnen, J.H., Smits, P.H. i Meijerman, I., 2005. Detection of single nucleotide polymorphisms in the ABCG2 gene in a Dutch population. *Am. J. Pharmacogenomics*, 5(2): 123-31.
- Chávez-Talavera, O., Tailleux, A., Lefebvre, P. i Staels, B., 2017. Bile Acid Control of Metabolism and Inflammation in Obesity, Type 2 Diabetes, Dyslipidemia, and Nonalcoholic Fatty Liver Disease. *Gastroenterology*, 152(7): 1679-1694.e3.
- Chen, C.J., Tseng, C.C., Yen, J.H., Chang, J.G., Chou, W.C., Chu, H.W., Chang, S.J. i Liao, W.T., 2018. ABCG2 contributes to the development of gout and hyperuricemia in a genome-wide association study. *Sci. Rep.*, 8(1): 3137.
- Choi, Y.H. i Yu, A.M., 2014. ABC transporters in multidrug resistance and pharmacokinetics, and strategies for drug development. *Curr. Pharm. Des.*, 20(5): 793-807.
- Cole, S.P., 2014. Multidrug resistance protein 1 (MRP1, ABCC1), a "multitasking" ATP-binding cassette (ABC) transporter. *J. Biol. Chem.*, 289(45): 30880-8.
- Conrad, S., Kauffmann, H.M., Ito, K., Deeley, R.G., Cole, S.P. i Schrenk, D., 2001. Identification of human multidrug resistance protein 1 (MRP1) mutations and characterization of a G671V substitution. *J. Hum. Genet.*, 46(11): 656-63.
- Conseil, G. i Cole, S.P., 2013. Two polymorphic variants of ABCC1 selectively alter drug resistance and inhibitor sensitivity of the multidrug and organic anion transporter multidrug resistance protein 1. *Drug. Metab. Dispos.*, 41(12): 2187-96.
- Cropp, C.D., Yee, S.W. i Giacomini, K.M., 2008. Genetic variation in drug transporters in ethnic populations. *Clin. Pharmacol. Ther.*, 84(3): 412-6.
- de Aguiar Vallim, T.Q., Tarling, E.J. i Edwards, P.A., 2013. Pleiotropic roles of bile acids in metabolism. *Cell Metab.*, 17(5): 657-69.
- de Jong, F.A., Marsh, S., Mathijssen, R.H., King, C., Verweij, J., Sparreboom, A. i McLeod, H.L., 2004. ABCG2 pharmacogenetics: ethnic differences in allele frequency and assessment of influence on irinotecan disposition. *Clin. Cancer Res.*, 10(17): 5889-94.
- Dean, M. i Annilo, T., 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu. Rev. Genomics Hum. Genet.*, 6: 123-42.
- Dean, M., Moitra, K. i Allikmets, R., 2022. The human ATP-binding cassette (ABC) transporter superfamily. *Hum. Mutat.*, 43(9): 1162-1182.
- Dean, M., Rzhetsky, A. i Allikmets, R., 2001. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.*, 11(7): 1156-66.
- Ding, L., Yang, L., Wang, Z. i Huang, W., 2015. Bile acid nuclear receptor FXR and digestive system diseases. *Acta Pharm. Sin. B*, 5(2): 135-44.
- Dobrowolska, S., Michalska-Madej, J., Słomka, M., Sobalska-Kwapis, M. i Strapagiel, D., 2019. Biobank Łódź – population based biobank at the University of Łódź, Poland. *Eur. J. Transl. Clin. Med.*, 2(1): 85-95.

- Dudarewicz, M., Barańska, M., Rychlik-Sych, M., Trzciński, R., Dziki, A. i Skrzętkowicz, J., 2012. C3435T polymorphism of the ABCB1/MDR1 gene encoding P-glycoprotein in patients with inflammatory bowel disease in a Polish population. *Pharmacol. Rep.*, 64(2): 343-50.
- Fang, X., Miao, R., Wei, J., Wu, H. i Tian, J., 2022. Advances in multi-omics study of biomarkers of glycolipid metabolism disorder. *Comput. Struct. Biotechnol. J.*, 20: 5935-5951.
- Fischer, S., Lakatos, P.L., Lakatos, L., Kovacs, A., Molnar, T., Altorjay, I., Papp, M., Szilvasi, A., Tulassay, Z., Osztoivits, J., Papp, J., Demeter, P., Schwab, R., Tordai, A. i Andrikovics, H., 2007. ATP-binding cassette transporter ABCG2 (BCRP) and ABCB1 (MDR1) variants are not associated with disease susceptibility, disease phenotype response to medical therapy or need for surgery in Hungarian patients with inflammatory bowel diseases. *Scand. J. Gastroenterol.*, 42(6): 726-33.
- Fukushima-Uesaka, H., Saito, Y., Tohkin, M., Maekawa, K., Hasegawa, R., Kawamoto, M., Kamatani, N., Suzuki, K., Yanagawa, T., Kajio, H., Kuzuya, N., Yasuda, K. i Sawada, J., 2007. Genetic variations and haplotype structures of the ABC transporter gene ABCC1 in a Japanese population. *Drug Metab. Pharmacokinet.*, 22(1): 48-60.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J. i Altshuler, D., 2002. The structure of haplotype blocks in the human genome. *Science*, 296(5576): 2225-9.
- Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., Howard, N., Srinivasan, S., Silva, J.M., Molnes, J., Edghill, E.L., Frayling, T.M., Temple, I.K., Mackay, D., Shield, J.P., Sumnik, Z., van Rhijn, A., Wales, J.K., Clark, P., Gorman, S., Aisenberg, J., Ellard, S., Njølstad, P.R., Ashcroft, F.M. i Hattersley, A.T., 2004. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N. Engl. J. Med.*, 350(18): 1838-49.
- Gulilat, M., Lamb, T., Teft, W.A., Wang, J., Dron, J.S., Robinson, J.F., Tirona, R.G., Hegele, R.A., Kim, R.B. i Schwarz, U.I., 2019. Targeted next generation sequencing as a tool for precision medicine. *BMC Med. Genomics*, 12(1): 81.
- Haer-Wigman, L., Ait Soussan, A., Ligthart, P., de Haas, M. i van der Schoot, C.E., 2014. Molecular analysis of immunized Jr(a-) or Lan- patients and validation of a high-throughput genotyping assay to screen blood donors for Jr(a-) and Lan- phenotypes. *Transfusion*, 54(7): 1836-46.
- Hansen, M., Sonne, D.P. i Knop, F.K., 2014. Bile acid sequestrants: glucose-lowering mechanisms and efficacy in type 2 diabetes. *Curr. Diab. Rep.*, 14(5): 482.
- He, J., Fortunati, E., Liu, D.X. i Li, Y., 2021. Pleiotropic Roles of ABC Transporters in Breast Cancer. *Int. J. Mol. Sci.*, 22(6).
- Hira, D. i Terada, T., 2018. BCRP/ABCG2 and high-alert medications: Biochemical, pharmacokinetic, pharmacogenetic, and clinical implications. *Biochem. Pharmacol.*, 147: 201-210.
- Hlaváč, V., Holý, P. i Souček, P., 2020a. Pharmacogenomics to Predict Tumor Therapy Response: A Focus on ATP-Binding Cassette Transporters and Cytochromes P450. *J. Pers. Med.*, 10(3).
- Hlaváč, V., Václavíková, R., Brynychová, V., Koževnikovová, R., Kopečková, K., Vrána, D., Gatěk, J. i Souček, P., 2020b. Role of Genetic Variation in ABC Transporters in Breast Cancer Prognosis and Therapy Response. *Int. J. Mol. Sci.*, 21(24).
- Hulpke, S. i Tampé, R., 2013. The MHC I loading complex: a multitasking machinery in adaptive immunity. *Trends Biochem. Sci.*, 38(8): 412-20.
- Hwang, T.C. i Kirk, K.L., 2013. The CFTR ion channel: gating, regulation, and anion permeation. *Cold Spring Harb. Perspect. Med.*, 3(1): a009498.
- Ieiri, I., 2012. Functional significance of genetic polymorphisms in P-glycoprotein (MDR1, ABCB1) and breast cancer resistance protein (BCRP, ABCG2). *Drug. Metab. Pharmacokinet.*, 27(1): 85-105.
- Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y. i Lauschke, V.M., 2018. Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum. Genomics*, 12(1): 26.

- Jani, M., Ambrus, C., Magnan, R., Jakab, K.T., Beery, E., Zolnerciks, J.K. i Krajcsi, P., 2014. Structure and function of BCRP, a broad specificity transporter of xenobiotics and endobiotics. *Arch. Toxicol.*, 88(6): 1205-48.
- Jaźwińska-Tarnawska, E., Jęskowiak, I., Waszczuk, E., Mulak, A., Głowacka, K., Hurkacz, M., Paradowski, L., Zaleska, Z. i Wiela-Hojeńska, A., 2015. Genetic polymorphism of ABCB1 gene (C3435T) in patients with inflammatory bowel diseases. Is there any gender dependency? *Pharmacol. Rep.*, 67(2): 294-8.
- Jin, L.H., Fang, Z.P., Fan, M.J. i Huang, W.D., 2019. Bile-ology: from bench to bedside. *J. Zhejiang Univ. Sci. B*, 20(5): 414-427.
- Juan-Carlos, P.M., Perla-Lidia, P.P., Stephanie-Talia, M.M., Mónica-Griselda, A.M. i Luz-María, T.E., 2021. ABC transporter superfamily. An updated overview, relevance in cancer multidrug resistance and perspectives with personalized medicine. *Mol. Biol. Rep.*, 48(2): 1883-1901.
- Koszarska, M., Kucsma, N., Kiss, K., Varady, G., Gera, M., Antalffy, G., Andrikovics, H., Tordai, A., Studzian, M., Strapagiel, D., Pulaski, L., Tani, Y., Sarkadi, B. i Szakacs, G., 2014. Screening the expression of ABCB6 in erythrocytes reveals an unexpectedly high frequency of Lan mutations in healthy individuals. *PLoS One*, 9(10): e111590.
- Kozyra, M., Ingelman-Sundberg, M. i Lauschke, V.M., 2017. Rare genetic variants in cellular transporters, metabolic enzymes, and nuclear receptors can be important determinants of interindividual differences in drug response. *Genet. Med.*, 19(1): 20-29.
- Ksander, B.R., Kolovou, P.E., Wilson, B.J., Saab, K.R., Guo, Q., Ma, J., McGuire, S.P., Gregory, M.S., Vincent, W.J., Perez, V.L., Cruz-Guilloty, F., Kao, W.W., Call, M.K., Tucker, B.A., Zhan, Q., Murphy, G.F., Lathrop, K.L., Alt, C., Mortensen, L.J., Lin, C.P., Zieske, J.D., Frank, M.H. i Frank, N.Y., 2014. ABCB5 is a limbal stem cell gene required for corneal development and repair. *Nature*, 511(7509): 353-7.
- Kuipers, F., Bloks, V.W. i Groen, A.K., 2014. Beyond intestinal soap--bile acids in metabolic control. *Nat. Rev. Endocrinol.*, 10(8): 488-98.
- Kurose, K., Sugiyama, E. i Saito, Y., 2012. Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: implications in the clinical trials for novel drug development. *Drug Metab. Pharmacokinet.*, 27(1): 9-54.
- Lee, S.S., Jeong, H.E., Yi, J.M., Jung, H.J., Jang, J.E., Kim, E.Y., Lee, S.J. i Shin, J.G., 2007. Identification and functional assessment of BCRP polymorphisms in a Korean population. *Drug Metab. Dispos.*, 35(4): 623-32.
- Leschziner, G., Zabaneh, D., Pirmohamed, M., Owen, A., Rogers, J., Coffey, A.J., Balding, D.J., Bentley, D.B. i Johnson, M.R., 2006. Exon sequencing and high resolution haplotype analysis of ABC transporter genes implicated in drug resistance. *Pharmacogenet. Genomics*, 16(6): 439-50.
- Leslie, E.M., Deeley, R.G. i Cole, S.P., 2005. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol. Appl. Pharmacol.*, 204(3): 216-37.
- Letourneau, I.J., Deeley, R.G. i Cole, S.P., 2005. Functional characterization of non-synonymous single nucleotide polymorphisms in the gene encoding human multidrug resistance protein 1 (MRP1/ABCC1). *Pharmacogenet. Genomics*, 15(9): 647-57.
- Li, W., Zhang, H., Assaraf, Y.G., Zhao, K., Xu, X., Xie, J., Yang, D.H. i Chen, Z.S., 2016. Overcoming ABC transporter-mediated multidrug resistance: Molecular mechanisms and novel therapeutic drug strategies. *Drug Resist. Updat.*, 27: 14-29.
- Lin, J.Y., Zhang, M., Schatton, T., Wilson, B.J., Alloo, A., Ma, J., Qureshi, A.A., Frank, N.Y., Han, J. i Frank, M.H., 2013. Genetically determined ABCB5 functionality correlates with pigmentation phenotype and melanoma risk. *Biochem. Biophys. Res. Commun.*, 436(3): 536-42.
- Liu, X., 2019. ABC Family Transporters. *Adv. Exp. Med. Biol.*, 1141: 13-100.
- Liu, X., Fiocco, M., Swen, J.J. i Guchelaar, H.J., 2017. Assessment of ethnic differences in sunitinib outcome between Caucasian and Asian patients with metastatic renal cell carcinoma: a meta-analysis. *Acta Oncol.*, 56(4): 582-589.

- Maekawa, K., Itoda, M., Sai, K., Saito, Y., Kaniwa, N., Shirao, K., Hamaguchi, T., Kunitoh, H., Yamamoto, N., Tamura, T., Minami, H., Kubota, K., Ohtsu, A., Yoshida, T., Saijo, N., Kamatani, N., Ozawa, S. i Sawada, J., 2006. Genetic variation and haplotype structure of the ABC transporter gene ABCG2 in a Japanese population. *Drug Metab. Pharmacokinet.*, 21(2): 109-21.
- Majewska, M., Sobalska, M., Siewierska, A., Strapagiel, D. i Bartosz, G., 2012. Single nucleotide polymorphisms of selected ABCC genes in the Polish population, 4th FEBS special meeting ATP-binding cassette (ABC) proteins: from multidrug resistance to genetic diseases, Innsbruck, Austria, pp. 84.
- Majewska, M., Strapagiel, D. i Bartosz, G., 2013. Single nucleotide polymorphisms of ABCB1 and ABCG2 genes in the Polish cancer patients, Role of multidrug resistance proteins in pharmacokinetics and toxicology, Ryn, Poland, pp. 61.
- Minias, P. i Vinkler, M., 2022. Selection Balancing at Innate Immune Genes: Adaptive Polymorphism Maintenance in Toll-Like Receptors. *Mol. Biol. Evol.*, 39(5).
- Molinaro, A., Wahlström, A. i Marschall, H.U., 2018. Role of Bile Acids in Metabolic Control. *Trends Endocrinol. Metab.*, 29(1): 31-41.
- Muller, P.J., Dally, H., Klappenecker, C.N., Edler, L., Jager, B., Gerst, M., Spiegelhalder, B., Tuengerthal, S., Fischer, J.R., Drings, P., Bartsch, H. i Risch, A., 2009. Polymorphisms in ABCG2, ABCC3 and CNT1 genes and their possible impact on chemotherapy outcome of lung cancer patients. *Int. J. Cancer*, 124(7): 1669-74.
- Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M.D., Nangle, K., Wang, J., Abecasis, G., Cardon, L.R., Zöllner, S., Whittaker, J.C., Chisoe, S.L., Novembre, J. i Mooser, V., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, 337(6090): 100-4.
- Neumann, J., Rose-Sperling, D. i Hellmich, U.A., 2017. Diverse relations between ABC transporters and lipids: An overview. *Biochim. Biophys. Acta Biomembr.*, 1859(4): 605-618.
- Niebudek, K., Balcerczak, E., Mirowski, M., Pietrzak, J., Zawadzka, I. i Zebrowska-Nawrocka, M., 2019. The contribution of ABCG2 G34A and C421A polymorphisms to multiple myeloma susceptibility. *Onco Targets Ther.*, 12: 1655-1660.
- Nobili, S., Lapucci, A., Landini, I., Coronello, M., Roviello, G. i Mini, E., 2020. Role of ATP-binding cassette transporters in cancer initiation and progression. *Semin. Cancer Biol.*, 60: 72-95.
- Noguchi, K., Katayama, K., Mitsuhashi, J. i Sugimoto, Y., 2009. Functions of the breast cancer resistance protein (BCRP/ABCG2) in chemotherapy. *Adv. Drug Deliv. Rev.*, 61(1): 26-33.
- Obata, H., Yahata, T., Quan, J., Sekine, M. i Tanaka, K., 2006. Association between single nucleotide polymorphisms of drug resistance-associated genes and response to chemotherapy in advanced ovarian cancer. *Anticancer Res.*, 26(3B): 2227-32.
- Ohashi, J., Naka, I. i Tsuchiya, N., 2011. The impact of natural selection on an ABCC11 SNP determining earwax type. *Mol. Biol. Evol.*, 28(1): 849-57.
- Ono, C., Kikkawa, H., Suzuki, A., Suzuki, M., Yamamoto, Y., Ichikawa, K., Fukae, M. i Ieiri, I., 2013. Clinical impact of genetic variants of drug transporters in different ethnic groups within and across regions. *Pharmacogenomics*, 14(14): 1745-64.
- Oselin, K., Mrozikiewicz, P.M., Gaikovitch, E., Pakla, R. i Roots, I., 2003. Frequency of MRP1 genetic polymorphisms and their functional significance in Caucasians: detection of a novel mutation G816A in the human MRP1 gene. *Eur. J. Clin. Pharmacol.*, 59(4): 347-50.
- Pham, D.H., Kudira, R., Xu, L., Valencia, C.A., Ellis, J.L., Shi, T., Evason, K.J., Osuji, I., Matuschek, N., Pfuhler, L., Mullen, M., Mohanty, S.K., Husami, A., Bull, L.N., Zhang, K., Wali, S., Yin, C. i Miethke, A., 2021. Deleterious Variants in ABCC12 are Detected in Idiopathic Chronic Cholestasis and Cause Intrahepatic Bile Duct Loss in Model Organisms. *Gastroenterology*, 161(1): 287-300.e16.

- Pierini, F. i Lenz, T.L., 2018. Divergent Allele Advantage at Human MHC Genes: Signatures of Past and Ongoing Selection. *Mol. Biol. Evol.*, 35(9): 2145-2158.
- Prawitt, J. i Staels, B., 2010. Bile acid sequestrants: glucose-lowering mechanisms. *Metab. Syndr. Relat. Disord.*, 8 Suppl 1: S3-8.
- Quazi, F. i Molday, R.S., 2011. Lipid transport by mammalian ABC proteins. *Essays Biochem.*, 50(1): 265-90.
- Robey, R.W., Pluchino, K.M., Hall, M.D., Fojo, A.T., Bates, S.E. i Gottesman, M.M., 2018. Revisiting the role of ABC transporters in multidrug-resistant cancer. *Nat. Rev. Cancer*, 18(7): 452-464.
- Russel, F.G., Koenderink, J.B. i Masereeuw, R., 2008. Multidrug resistance protein 4 (MRP4/ABCC4): a versatile efflux transporter for drugs and signalling molecules. *Trends Pharmacol. Sci.*, 29(4): 200-7.
- Saito, S., Iida, A., Sekine, A., Miura, Y., Ogawa, C., Kawauchi, S., Higuchi, S. i Nakamura, Y., 2002. Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABCC/MRP/CFTR). *J. Hum. Genet.*, 47(4): 147-71.
- Salagacka-Kubiak, A., Zebrowska, M., Wosiak, A., Balcerczak, M., Mirowski, M. i Balcerczak, E., 2016. ABCG2 in peptic ulcer: gene expression and mutation analysis. *J. Appl. Genet.*, 57(3): 335-42.
- Setia, N., Abbas, O., Sousa, Y., Garb, J.L. i Mahalingam, M., 2012. Profiling of ABC transporters ABCB5, ABCF2 and nestin-positive stem cells in nevi, in situ and invasive melanoma. *Mod. Pathol.*, 25(8): 1169-75.
- Shapiro, H., Kolodziejczyk, A.A., Halstuch, D. i Elinav, E., 2018. Bile acids in glucose metabolism in health and disease. *J. Exp. Med.*, 215(2): 383-396.
- Silverton, L., Dean, M. i Moitra, K., 2011. Variation and evolution of the ABC transporter genes ABCB1, ABCC1, ABCG2, ABCG5 and ABCG8: implication for pharmacogenetics and disease. *Drug Metabol. Drug Interact.*, 26(4): 169-79.
- Słomka, M., Sobalska-Kwapis, M., Wachulec, M., Bartosz, G. i Strapagiel, D., 2017. High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies. *Int. J. Mol. Sci.*, 18(11).
- Staels, B., Handelsman, Y. i Fonseca, V., 2010. Bile acid sequestrants for lipid and glucose control. *Curr. Diab. Rep.*, 10(1): 70-7.
- Stark, K., Reinhard, W., Grassl, M., Erdmann, J., Schunkert, H., Illig, T. i Hengstenberg, C., 2009. Common polymorphisms influencing serum uric acid levels contribute to susceptibility to gout, but not to coronary artery disease. *PLoS One*, 4(11): e7729.
- Stiburkova, B., Pavelcova, K., Zavada, J., Petru, L., Simek, P., Cepek, P., Pavlikova, M., Matsuo, H., Merriman, T.R. i Pavelka, K., 2017. Functional non-synonymous variants of ABCG2 and gout risk. *Rheumatology (Oxford)*, 56(11): 1982-1992.
- Strapagiel, D., Majewska, M., Słomka, M., Janik, K., Sobalska, M. i Bartosz, G., 2019. Method for determination of gender. In: U.o. Lodz (Editor), *Wiadomości Urzędu Patentowego*. University of Lodz, Poland.
- Strapagiel, D., Sobalska, M., Słomka, M. i Marciniak, B., 2016. Biobank Lodz - DNA Based Biobank at the University of Lodz, Poland. *Open J. Bioresour.*, 3(1): e6.
- Tarling, E.J., de Aguiar Vallim, T.Q. i Edwards, P.A., 2013. Role of ABC transporters in lipid transport and human disease. *Trends Endocrinol. Metab.*, 24(7): 342-50.
- Toyoda, Y., Mančíková, A., Krylov, V., Morimoto, K., Pavelcová, K., Bohatá, J., Pavelka, K., Pavlíková, M., Suzuki, H., Matsuo, H., Takada, T. i Stiburkova, B., 2019. Functional Characterization of Clinically-Relevant Rare Variants in ABCG2 Identified in a Gout and Hyperuricemia Cohort. *Cells*, 8(4).
- Vasiliou, V., Vasiliou, K. i Nebert, D.W., 2009. Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics*, 3(3): 281-90.
- Vulsteke, C., Lambrechts, D., Dieudonne, A., Hatse, S., Brouwers, B., van Brussel, T., Neven, P., Belmans, A., Schoffski, P., Paridaens, R. i Wildiers, H., 2013. Genetic variability in the multidrug resistance associated protein-1 (ABCC1/MRP1) predicts hematological toxicity in

- breast cancer patients receiving (neo-)adjuvant chemotherapy with 5-fluorouracil, epirubicin and cyclophosphamide (FEC). *Ann. Oncol.*, 24(6): 1513-25.
- Wang, H., Hao, B., Zhou, K., Chen, X., Wu, S., Zhou, G., Zhu, Y. i He, F., 2004. Linkage disequilibrium and haplotype architecture for two ABC transporter genes (ABCC1 and ABCG2) in Chinese population: implications for pharmacogenomic association studies. *Ann. Hum. Genet.*, 68(Pt 6): 563-73.
- Wang, J.Q., Wu, Z.X., Yang, Y., Teng, Q.X., Li, Y.D., Lei, Z.N., Jani, K.A., Kaushal, N. i Chen, Z.S., 2021. ATP-binding cassette (ABC) transporters in cancer: A review of recent updates. *J. Evid. Based Med.*, 14(3): 232-256.
- Wang, N., Akey, J.M., Zhang, K., Chakraborty, R. i Jin, L., 2002. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.*, 71(5): 1227-34.
- Wang, Z., Sew, P.H., Ambrose, H., Ryan, S., Chong, S.S., Lee, E.J. i Lee, C.G., 2006. Nucleotide sequence analyses of the MRP1 gene in four populations suggest negative selection on its coding region. *BMC Genomics*, 7: 111.
- Wang, Z., Wang, J., Tantoso, E., Wang, B., Tai, A.Y., Ooi, L.L., Chong, S.S. i Lee, C.G., 2007. Signatures of recent positive selection at the ATP-binding cassette drug transporter superfamily gene loci. *Hum. Mol. Genet.*, 16(11): 1367-80.
- Woodward, O.M., Köttgen, A., Coresh, J., Boerwinkle, E., Guggino, W.B. i Köttgen, M., 2009. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl. Acad. Sci. U S A*, 106(25): 10338-42.
- Xiao, Q., Zhou, Y. i Lauschke, V.M., 2020a. Ethnogeographic and inter-individual variability of human ABC transporters. *Hum. Genet.*, 139(5): 623-646.
- Xiao, Q., Zhou, Y. i Lauschke, V.M., 2020b. Impact of variants in ATP-binding cassette transporters on breast cancer treatment. *Pharmacogenomics*, 21(18): 1299-1310.
- Yamagishi, K., Tanigawa, T., Kitamura, A., Köttgen, A., Folsom, A.R. i Iso, H., 2010. The rs2231142 variant of the ABCG2 gene is associated with uric acid levels and gout among Japanese people. *Rheumatology (Oxford)*, 49(8): 1461-5.
- Yee, S.W., Brackman, D.J., Ennis, E.A., Sugiyama, Y., Kamdem, L.K., Blanchard, R., Galetin, A., Zhang, L. i Giacomini, K.M., 2018. Influence of Transporter Polymorphisms on Drug Disposition and Response: A Perspective From the International Transporter Consortium. *Clin. Pharmacol. Ther.*, 104(5): 803-817.
- Yin, J.Y., Han, L.F., Huang, Q., Xu, X.J., Zhou, H.H. i Liu, Z.Q., 2011. ABCC1 polymorphism Arg723Gln (2168G> A) is associated with lung cancer susceptibility in a Chinese population. *Clin. Exp. Pharmacol. Physiol.*, 38(9): 632-7.
- Yin, J.Y., Huang, Q., Yang, Y., Zhang, J.T., Zhong, M.Z., Zhou, H.H. i Liu, Z.Q., 2009. Characterization and analyses of multidrug resistance-associated protein 1 (MRP1/ABCC1) polymorphisms in Chinese population. *Pharmacogenet. Genomics*, 19(3): 206-16.
- Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., Bannai, M., Tokunaga, K., Sonoda, S., Komaki, R., Ihara, M., Saenko, V.A., Alipov, G.K., Sekine, I., Komatsu, K., Takahashi, H., Nakashima, M., Sosonkina, N., Mapendano, C.K., Ghadami, M., Nomura, M., Liang, D.S., Miwa, N., Kim, D.K., Garidkhuu, A., Natsume, N., Ohta, T., Tomita, H., Kaneko, A., Kikuchi, M., Russomando, G., Hirayama, K., Ishibashi, M., Takahashi, A., Saitou, N., Murray, J.C., Saito, S., Nakamura, Y. i Niikawa, N., 2006. A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat. Genet.*, 38(3): 324-30.
- Yu, K.H., Chang, P.Y., Chang, S.C., Wu-Chou, Y.H., Wu, L.A., Chen, D.P., Lo, F.S. i Lu, J.J., 2017. A comprehensive analysis of the association of common variants of ABCG2 with gout. *Sci. Rep.*, 7(1): 9988.
- Zámbó, B., Móznér, O., Bartos, Z., Török, G., Várady, G., Telbisz, Á., Homolya, L., Orbán, T.I. i Sarkadi, B., 2020. Cellular expression and function of naturally occurring variants of the human ABCG2 multidrug transporter. *Cell. Mol. Life Sci.*, 77(2): 365-378.

- Zhang, L., Spencer, K.L., Voruganti, V.S., Jorgensen, N.W., Fornage, M., Best, L.G., Brown-Gentry, K.D., Cole, S.A., Crawford, D.C., Deelman, E., Franceschini, N., Gaffo, A.L., Glenn, K.R., Heiss, G., Jenny, N.S., Kottgen, A., Li, Q., Liu, K., Matise, T.C., North, K.E., Umans, J.G. i Kao, W.H., 2013. Association of functional polymorphism rs2231142 (Q141K) in the ABCG2 gene with serum uric acid and gout in 4 US populations: the PAGE Study. *Am. J. Epidemiol.*, 177(9): 923-32.
- Zhang, W., Sun, S., Zhang, W. i Shi, Z., 2018. Polymorphisms of ABCG2 and its impact on clinical relevance. *Biochem. Biophys. Res. Commun.*, 503(2): 408-413.
- Zhou, D., Liu, Y., Zhang, X., Gu, X., Wang, H., Luo, X., Zhang, J., Zou, H. i Guan, M., 2014. Functional polymorphisms of the ABCG2 gene are associated with gout disease in the Chinese Han male population. *Int. J. Mol. Sci.*, 15(5): 9149-59.
- Zhou, Y., Arribas, G.H., Turku, A., Jürgenson, T., Mkrtchian, S., Krebs, K., Wang, Y., Svobodova, B., Milani, L., Schulte, G., Korabecny, J., Gastaldello, S. i Lauschke, V.M., 2021. Rare genetic variability in human drug target genes modulates drug response and can guide precision medicine. *Sci. Adv.*, 7(36): eabi6856.
- Zhou, Y., Fujikura, K., Mkrtchian, S. i Lauschke, V.M., 2018. Computational Methods for the Pharmacogenetic Interpretation of Next Generation Sequencing Data. *Front. Pharmacol.*, 9: 1437.
- Zhou, Y. i Lauschke, V.M., 2022. Population pharmacogenomics: an update on ethnogeographic differences and opportunities for precision public health. *Hum. Genet.*, 141(6): 1113-1136.

INTERNETOWE SERWERY I BAZY DANYCH:

1000 Genomes Project	https://www.internationalgenome.org/
Ensembl Variant Effect Predictor (VEP)	https://www.ensembl.org/Tools/VEP
Gene Ontology Resource	http://geneontology.org/
GTEx Portal	https://gtexportal.org/home/
HUGO Gene Nomenclature Committee (HGNC)	https://www.genenames.org/
MedlinePlus	https://medlineplus.gov/
NCBI GenBank	https://www.ncbi.nlm.nih.gov/genbank/
NCBI Gene	https://www.ncbi.nlm.nih.gov/gene/
NCBI dbSNP	https://www.ncbi.nlm.nih.gov/snp/
TopMed	https://imputation.biodatacatalyst.nhlbi.nih.gov/
UniProtKB/Swiss-Prot	https://www.uniprot.org/

POZOSTAŁY DOROBIEK NIEWCHODZĄCY W SKŁAD ROZPRAWY DOKTORSKIEJ

Publikacje:

1. Strapagiel, D., Sobalska-Kwapis, M., Słomka, M. i Marciniak, B., 2016. Biobank Lodz – DNA Based Biobank at the University of Lodz, Poland. *Open J. Bioresour.*, 3(1): e6. doi: <http://doi.org/10.5334/ojb.22>
IF= BRAK / MEiN = 5 (Recenzowana publikacja naukowa w języku innym niż polski, zamieszczona w zagranicznym czasopiśmie naukowym niezamieszczonym w wykazie czasopism naukowych).
2. Ratajewski, M., Słomka, M., Karaś, K., Sobalska-Kwapis, M., Korycka-Machata, M., Sałkowska, A., Dziadek, J., Strapagiel, D. i Dastych, J. 2017. Functional Analysis of the rs774872314, rs116171003, rs200231898 and rs201107751 Polymorphisms in the Human RORγT Gene Promoter Region. *Genes (Basel)*, 8(4): 126. doi: 10.3390/genes8040126.
IF=3,191 / MEiN= 25 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
3. Sobalska-Kwapis, M., Suchanecka, A., Słomka, M., Siewierska-Górska, A., Kępka, E. i Strapagiel, D. 2017. Genetic association of FTO/IRX region with obesity and overweight in the Polish population. *PLoS One*, 12(6): e0180295. doi: 10.1371/journal.pone.0180295.
IF=2,766 / MEiN= 40 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
4. Sobalska-Kwapis, M., Smolarz, B., Słomka, M., Szaflik, T., Kępka, E., Kulig, B., Siewierska-Górska, A., Polak, G., Romanowicz, H., Strapagiel, D. i Szyłło, K. 2017. New variants near RHOJ and C2, HLA-DRA region and susceptibility to endometriosis in the Polish population – The genome-wide association study. *Eur. J. Obstet. Gynecol. Reprod. Biol.*, 217: 106-112. doi: 10.1016/j.ejogrb.2017.08.037.
IF= 1,809 / MEiN=25 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
5. Nedoszytko, B., Siemińska, A., Strapagiel, D., Dąbrowski, S., Słomka, M., Sobalska-Kwapis, M., Marciniak, B., Wierzba, J., Skokowski, J., Fijałkowski, M., Nowicki, R. i Kalinowski, L. 2017. High prevalence of carriers of variant c.1528G>C of HADHA gene causing long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (LCHADD) in the population of adult Kashubians from North Poland. *PLoS One*, 12(11): e0187365. doi: 10.1371/journal.pone.0187365.
IF=2,766 / MEiN= 40 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
6. Słomka, M., Sobalska-Kwapis, M., Wachulec, M., Bartosz, G. i Strapagiel, D. 2017. High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies. *Int. J. Mol. Sci.*, 18(11): 2316. doi: 10.3390/ijms18112316.
IF= 3,687 / MEiN= 30 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
7. Romanowicz, H., Strapagiel, D., Słomka, M., Sobalska-Kwapis, M., Kępka, E., Siewierska-Górska, A., Zadrożny, M., Bieńkiewicz, J. i Smolarz, B. 2017. New single nucleotide polymorphisms (SNPs) in homologous recombination repair genes detected by microarray analysis in Polish breast cancer patients. *Clin. Exp. Med.*, 17(4): 541-546. doi: 10.1007/s10238-016-0441-2.
IF= 2,787 / MEiN= 25 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
8. Kukla-Bartoszek, M., Pośpiech, E., Spólnicka, M., Karłowska-Pik, J., Strapagiel, D., Żądzińska, E., Rosset, I., Sobalska-Kwapis, M., Słomka, M., Walsh, S., Kayser, M., Sitek, A. i Branicki, W. 2018. Investigating the impact of age-depended hair colour darkening during childhood on DNA-based

- hair colour prediction with the HirisPlex system. *Forensic Sci. Int. Genet.*, 36: 26-33. doi: 10.1016/j.fsigen.2018.06.007.
IF= 4.884 / MEiN= 45 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
9. Pruszkowska-Przybylska, P., Sitek, A., Rosset, I., Sobalska-Kwapis, M., Słomka, M., Strapagiel, D. i Żądzińska, E. 2018. Association of the 2D:4D digit ratio with body composition among the Polish children aged 6-13 years. *Early Hum. Dev.*, 124: 26-32. doi: 10.1016/j.earlhumdev.2018.08.001.
IF= 1,853 / MEiN= 35 pkt (Komunikat MNiSW z dnia 25 stycznia 2017 r. ws. wykazu czasopism naukowych ustalonych na podstawie wykazów ogłoszonych w latach 2013-2016).
 10. Pruszkowska-Przybylska, P., Sitek, A., Rosset, I., Żądzińska, E., Sobalska-Kwapis, M., Słomka, M. i Strapagiel, D. 2019. The association between socioeconomic status, duration of breastfeeding, parental age and birth parameters with BMI, body fat and muscle mass among prepubertal children in Poland. *Anthropol. Anz.*, 76(5): 409-419. doi: 10.1127/anthranz/2019/0955.
IF= 0,679 / MEiN= 70 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 11. Jarczak, J., Grochowalski, Ł., Marciniak, B., Lach, J., Słomka, M., Sobalska-Kwapis, M., Lorkiewicz, W., Pułaski, Ł. i Strapagiel D. 2019. Mitochondrial DNA variability of the Polish population. *Eur. J. Hum. Genet.*, 27(8): 1304-1314. doi: 10.1038/s41431-019-0381-x.
IF= 3,657 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 12. Płociński, P., Macios, M., Houghton, J., Niemiec, E., Płocińska, R., Brzostek, A., Słomka, M., Dziadek, J., Young, D. i Dziembowski, A. 2019. Proteomic and transcriptomic experiments reveal an essential role of RNA degradosome complexes in shaping the transcriptome of *Mycobacterium tuberculosis*. *Nucleic Acids Res.*, 47(11): 5892-5905. doi: 10.1093/nar/gkz251.
IF= 11,502 / MEiN= 200 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 13. Dobrowolska, S., Michalska-Madej, J., Słomka, M., Sobalska-Kwapis, M. i Strapagiel, D. 2019. Biobank Łódź – population based Biobank at the University of Łódź, Poland. *Eur. J. Transl. Clin. Med.*, 2(1): 85-95.
IF= BRAK / MEiN= 40 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 14. Pruszkowska-Przybylska, P., Sitek, A., Rosset, I., Sobalska-Kwapis, M., Słomka, M., Strapagiel, D., Żądzińska, E. i Morling, N. 2020. Association of saliva 25(OH)D concentration with body composition and proportion among pre-pubertal and pubertal Polish children. *Am. J. Hum. Biol.*, 32(5): e23397. doi: 10.1002/ajhb.23397.
IF= 1,937 / MEiN= 70 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 15. Bialik, P., Wysokiński, D., Słomka, M., Morawiec, Z., Strapagiel, D. i Woźniak, K. 2020. HRM screening of the UBC9 gene encoding the SUMO-E2-conjugating enzyme - case-control study in breast cancer. *Exp. Oncol.*, 42(2): 130-134. doi: 10.32471/exp-oncology.2312-8852.vol-42-no-2.14739.
IF= BRAK / MEiN= 40 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).
 16. Nedoszytko, B., Sobalska-Kwapis, M., Strapagiel, D., Lange, M., Górka, A., Elberink, J.N.G.O., van Doormaal, J., Słomka, M., Kalinowski, L., Gruchała-Niedoszytko, M., Nowicki, R.J., Valent, P. i Niedoszytko, M. 2020. Results from a Genome-Wide Association Study (GWAS) in Mastocytosis

Reveal New Gene Polymorphisms Associated with WHO Subgroups. Int. J. Mol. Sci., 21(15): 5506. doi: 10.3390/ijms21155506.

IF= 5,924 / MEiN= 140 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

17. Grochowalski, Ł., Jarczak, J., Urbanowicz, M., Słomka, M., Szargut, M., Borówka, P., Sobalska-Kwapis, M., Marciniak, B., Ossowski, A., Lorkiewicz, W. i Strapagiel, D. 2020. Y-Chromosome Genetic Analysis of Modern Polish Population. Front. Genet., 11: 567309. doi: 10.3389/fgene.2020.567309.

IF= 4,599 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

18. Pruszkowska-Przybylska, P., Sitek, A., Rosset, I., Sobalska-Kwapis, M., Słomka, M., Strapagiel, D., Żądzińska, E. i Morling, N. 2021. Associations between second to fourth digit ratio, cortisol, vitamin D, and body composition among Polish children. Sci. Rep., 11(1): 7029. doi: 10.1038/s41598-021-86521-7.

IF= 4,997 / MEiN= 140 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

19. Buda, P., Chyb, M., Smorczewska-Kiljan, A., Wieteska-Klimczak, A., Paczesna, A., Kowalczyk-Domagała, M., Okarska-Napierała, M., Sobalska-Kwapis, M., Grochowalski, Ł., Słomka, M., Sitek, A., Książyk, J. i Strapagiel, D. 2021. Association Between rs12037447, rs146732504, rs151078858, rs55723436, and rs6094136 Polymorphisms and Kawasaki Disease in the Population of Polish Children. Front. Pediatr., 9: 624798. doi: 10.3389/fped.2021.624798.

IF= 3,569 / MEiN= 70 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

20. Brzostek, A., Płociński, P., Minias, A., Ciszewska, A., Gąsior, F., Pawełczyk, J., Dziadek, B., Słomka, M. i Dziadek, J. 2021. Dissecting the RecA-(In)dependent Response to Mitomycin C in *Mycobacterium tuberculosis* Using Transcriptional Profiling and Proteomics Analyses. Cells., 10(5): 1168. doi: 10.3390/cells10051168.

IF= 7,666 / MEiN= 140 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

21. Koj, N., Grochowalski, Ł., Jarczak, J., Wójtowicz, W., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. 2021. The association between polymorphisms near TMEM18 and the risk of obesity: a meta-analysis. BMC Med. Genomics., 14(1): 179 doi: 10.1186/s12920-021-01025-7.

IF= 3,622 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

22. Pruszkowska-Przybylska, P., Sitek, A., Rosset, I., Sobalska-Kwapis, M., Słomka, M., Strapagiel, D., Żądzińska, E. i Morling, N. 2021. Cortisol concentration affects fat and muscle mass among Polish children aged 6-13 years. BMC Pediatr., 21(1): 365. doi: 10.1186/s12887-021-02837-3.

IF= 2,567 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

23. Płocińska, R., Wasik, K., Płociński, P., Lechowicz, E., Antczak, M., Błaszczuk, E., Dziadek, B., Słomka, M., Rumijowska-Galewicz, A. i Dziadek, J. 2022. The Orphan Response Regulator Rv3143 Modulates the Activity of the NADH Dehydrogenase Complex (Nuo) in *Mycobacterium tuberculosis* via Protein-Protein Interactions. Front. Cell Infect. Microbiol., 12: 909507. doi: 10.3389/fcimb.2022.909507.

IF= wartość wskaźnika określona będzie na przełomie 2023 roku (IF₂₀₂₁= 6,073) / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

24. Rahmioglu, N., Mortlock, S., Ghiasi, M., Møller, P.L., Stefansdottir, L., Galarneau, G., Turman, C., Danning, R., Law, M.H., Sapkota, Y., Christofidou, P., Skarp, S., Giri, A., Banasik, K., Krassowski, M., Lepamets, M., Marciniak, B., Nöukas, M., Perro, D., Sliz, E., Sobalska-Kwapis, M., Thorleifsson, G., Topbas-Selcuki, N.F., Vitonis, A., Westergaard, D., Arnadottir, R., Burgdorf, K.S., Campbell, A., Cheuk, C.S.K., Clementi, C., Cook, J., De Vivo, I., DiVasta, A., Dorien, O., Donoghue, J.F., Edwards, T., Fontanillas, P., Fung, J.N., Geirsson, R.T., Girling, J.E., Harkki, P., Harris, H.R., Healey, M., Heikinheimo, O., Holdsworth-Carson, S., Hostettler, I.C., Houlden, H., Houshdaran, S., Irwin, J.C., Jarvelin, M.R., Kamatani, Y., Kennedy, S.H., Kepka, E., Kettunen, J., Kubo, M., Kulig, B., Kurra, V., Laivuori, H., Laufer, M.R., Lindgren, C.M., MacGregor, S., Mangino, M., Martin, N.G., Matalliotaki, C., Matalliotakis, M., Murray, A.D., Ndungu, A., Nezhat, C., Olsen, C.M., Opoku-Anane, J., Padmanabhan, S., Paranjpe, M., Peters, M., Polak, G., Porteous, D.J., Rabban, J., Rexrode, K.M., Romanowicz, H., Saare, M., Saavalainen, L., Schork, A.J., Sen, S., Shafir, A.L., Siewierska-Górska, A., Słomka, M., Smith, B.H., Smolarz, B., Szaflik, T., Szyłto, K., Takahashi, A., Terry, K.L., Tomassetti, C., Treloar, S.A., Vanhie, A., Vincent, K., Vo, K.C., Werring, D.J., Zeggini, E., Zervou, M.I., Adachi, S., Buring, J.E., Ridker, P.M., D'Hooghe, T., Goulielmos, G.N., Hapangama, D.K., Hayward, C., Horne, A.W., Low, S.K., Martikainen, H., Chasman, D.I., Rogers, P.A.W., Saunders, P.T., Sirota, M., Spector, T., Strapagiel, D., Tung, J.Y., Whiteman, D.C., Giudice, L.C., Velez-Edwards, D.R., Uimari, O., Kraft, P., Salumets, A., Nyholt, D.R., Mägi, R., Stefansson, K., Becker, C.M., Yurttas-Beim, P., Steinhorsdottir, V., Nyegaard, M., Missmer, S.A., Montgomery, G.W., Morris, A.P. i Zondervan, K.T., 2023. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat. Genet.*, 55(3): 423-436. doi: 10.1038/s41588-023-01323-z.

IF= wartość wskaźnika określona będzie na przełomie 2024 roku (IF₂₀₂₁= 41,376) / MEiN= 200 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych).

Sumaryczny punkty IF= 128,189

Sumaryczne punkty MEiN= 2005

Patenty:

1. Strapagiel, D., Majewska, M., Słomka, M., Janik, K., Sobalska, M. i Bartosz, G. Sposób ustalania płci (ang. *Method for determination of gender*). Zgłaszający: Uniwersytet Łódzki. Numer aplikacji: PL40656913A, 2013-12-18. Numer publikacji: PL406569A1, 2015-06-22; PL232652B1, 2019-07-31. 75 pkt. (Punktacja przyznawana przez Urząd Patentowy Rzeczypospolitej Polskiej (UPRP) źródło informacji SEDN).
2. Strapagiel, D., Majewska, M., Słomka, M., Janik, K., Sobalska, M. i Bartosz, G. Sposób ustalania płci (ang. *Method for determination of sex*). Zgłaszający: Uniwersytet Łódzki. Numer aplikacji: PL42342013A, 2013-12-18. Numer publikacji: PL423420A1, 2018-03-26; PL232653B1, 2019-07-31. 75 pkt. (Punktacja przyznawana przez Urząd Patentowy Rzeczypospolitej Polskiej (UPRP) źródło informacji SEDN).

Doniesienie konferencyjne – plakaty:

1. Słomka, M., Korycka-Machała, M., Bartosz, G. i Strapagiel, D. Genetic variation in the ATP-binding cassette transporter gene ABCC1 (MRP1) in the Polish population. Konferencja "FEBS-ABC2014 - 5th FEBS Special Meeting on ATP-Binding Cassette (ABC) Proteins: From Multidrug Resistance to Genetic Diseases" organizowana przez Federation of European Biochemical Societies (FEBS) i Uniwersytet Medyczny w Wiedniu, 8 – 14 marca 2014, Innsbruck, Austria
2. Słomka, M., Korycka-Machała, M., Dziadek, J. i Strapagiel, D. Genetic variation of the ATP-binding cassette transporter gene ABCG2 in the Polish population. Sympozjum "14th Annual Horizons in Molecular Biology Symposium" organizowane przez Uniwersytet w Getyndze, 11 – 14 września 2017, Instytut Chemii Biofizycznej Maxa Plancka w Getyndze, Niemcy

3. Słomka M., Marciniak B., Sobalska-Kwapis M., Grochowalski Ł., Pułaski Ł., Strapagiel D. Genetic divergence of ABC human genes across worldwide reference populations with comparison of Polish population. Konferencja "FEBS Special Meeting 2018: ATP-Binding Cassette (ABC) Proteins: From Multidrug Resistance to Genetic Disease" organizowana przez Federation of European Biochemical Societies (FEBS) i Uniwersytet Medyczny w Wiedniu, 6 – 12 marca 2018, Innsbruck, Austria
4. Janik-Superson, K., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. Genetic determinants of brown eye pigmentation in the Polish population based on GWAS. Konferencja "Towards Comprehensive Population Studies – BIAŁYSTOK PLUS" organizowana przez Uniwersytet Medyczny w Białymstoku, 13 – 14 kwietnia 2018, Białystok.
5. Polak M, Janik-Superson K, Sobalska-Kwapis M, Shrubkovskiy I, Koj N, Słomka M, Marciniak B, Strapagiel D, Grochowalski Ł. Obesity and overweight associated SNPs in Polish population. Konferencja "Towards Comprehensive Population Studies – BIAŁYSTOK PLUS" organizowana przez Uniwersytet Medyczny w Białymstoku, 13 – 14 kwietnia 2018, Białystok.
6. Polak, M., Koj, N., Shrubkovskiy, I., Janik-Superson, K., Sobalska-Kwapis, M., Słomka, M., Borówka, P., Marciniak, B. i Strapagiel, D. Analysis of modern DNA as a part of platform e-Człowiek.pl. Konferencja "Polish Scientific Network" organizowana jest przez Akademię Młodych Uczonych PAN, Ministerstwo Nauki i Szkolnictwa Wyższego i Klub Stypendystów Fundacji na rzecz Nauki Polskiej, we współpracy z Uniwersytetem Medycznym w Łodzi, 21 – 23 czerwca 2018, Łódź.
7. Janik-Superson, K., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. Big Data Lab Management Systems. Konferencja „Polish Scientific Networks” organizowana jest przez Akademię Młodych Uczonych PAN, Ministerstwo Nauki i Szkolnictwa Wyższego i Klub Stypendystów Fundacji na rzecz Nauki Polskiej, we współpracy z Uniwersytetem Medycznym w Łodzi, 21 – 23 czerwca 2018, Łódź.
8. Janik-Superson, K., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. Genetic determinants of blond hair in the Polish Population based on GWAS. Eur J Trans Clin Med. 2018;1(Suppl.4):70.
9. Jarczak, J., Grochowalski, Ł., Marciniak, B., Słomka, M., Sobalska-Kwapis, M., Lorkiewicz, W., Pułaski, Ł. i Strapagiel, D. "Mitochondrial DNA variability of the Polish population - including divisions into voivodeships". Konferencja "Towards Comprehensive Population Studies - BIAŁYSTOK PLUS" organizowana przez Uniwersytet Medyczny w Białymstoku, 13 – 14 kwietnia 2018, Białystok.
10. Grochowalski, Ł., Jarczak, J., Marciniak, B., Słomka, M., Sobalska-Kwapis, M. i Strapagiel, D. "Clustering – an useful tool used in population studies". Konferencja "Polish Scientific Network" organizowana jest przez Akademię Młodych Uczonych PAN, Ministerstwo Nauki i Szkolnictwa Wyższego i Klub Stypendystów Fundacji na rzecz Nauki Polskiej, we współpracy z Uniwersytetem Medycznym w Łodzi, 21 – 23 czerwca 2018, Łódź.
11. Koj, N., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. Identification of important missense mutation in KCNK10 gene in individuals with myocardial infarction incidence by SNP microarrays analysis. Konferencja "Towards Comprehensive Population Studies - BIAŁYSTOK PLUS" organizowana przez Uniwersytet Medyczny w Białymstoku, 13 – 14 kwietnia 2018, Białystok.

12. Koj, N., Grochowalski, Ł., Sobalska-Kwapis, M., Słomka, M., Marciniak, B. i Strapagiel, D. Two polymorphisms of TMEM18 rs6548238 and rs4854344 and the risks of obesity in children and adults: A meta-analysis. II Krajowa Naukowo-Szkoleniowa Konferencja Biobanków Polskich organizowana przez BBMRI.pl, 11 – 12 października 2018, PORT - Polski Ośrodek Rozwoju Technologii, Wrocław.
13. Słomka, M., Jarczak, J., Grochowalski, Ł., Marciniak, B., Lach, J., Sobalska-Kwapis, M., Lorkiewicz, W., Pułaski, Ł. i Strapagiel, D. Genetic diversity of Polish population based on the analysis of mitochondrial DNA. III Krajowa Naukowo Szkoleniowa Konferencja Biobanków Polskich organizowana przez BBMRI.pl, 06-08 listopada 2019, Uniwersytet Łódzki, Łódź.

Wystąpienia ustne:

1. Sobalska-Kwapis, M., Suchanecka, A., Słomka, M., Siewierska-Górska, A., Kępka, E. i Strapagiel, D. Genetic association of FTO/IRX region with obesity and overweight in the Polish population. VII Sympozjum Polskich Użytkowników Platformy Illumina, 12 – 13 października 2017, Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Biologii.

ZAŁĄCZNIK NR 1

Słomka, M., Sobalska-Kwapis, M., Korycka-Machala, M., Bartosz, G., Dziadek, J. Strapagiel, D., 2015. Genetic variation of the ABC transporter gene ABCC1 (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

RESEARCH ARTICLE

Open Access



Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1 – MRP1) in the Polish population

Marcin Ślomka¹, Marta Sobalska-Kwapis¹, Małgorzata Korycka-Machała², Grzegorz Bartosz³, Jarosław Dziadek² and Dominik Strapagiel^{1*}

Abstract

Background: Multidrug resistance-associated protein 1 (MRP1), encoded by the *ABCC1* gene, is an ATP-binding cassette transporter mediating efflux of organic anions and xenobiotics; its overexpression leads to multidrug resistance. In this study, 30 exons (from 31 in total) of the *ABCC1* gene as well as and their flanking intron sequences were screened for genetic variation, using the High Resolution Melting (HRM) method, for 190 healthy volunteers representing the Polish population. Polymorphism screening is an indispensable step in personalized patient therapy. An additional targeted SNP verification study for ten variants was performed to verify sensitivity of the scanning method.

Results: During scanning, 46 polymorphisms, including seven novel ones, were found: one in 3' UTR, 21 in exons (11 of them non-synonymous) and 24 in introns, including one deletion variant. These results revealed some ethnic differences in frequency of several polymorphisms when compared to literature data for other populations. Based on linkage disequilibrium analysis, 4 haplotype blocks were determined for 9 detected polymorphisms and 12 haplotypes were defined. To capture the common haplotypes, haplotype-tagging single nucleotide polymorphisms were identified.

Conclusions: Targeted genotyping results correlated well with scanning results; thus, HRM is a suitable method to study genetic variation in this model. HRM is an efficient and sensitive method for scanning and genotyping polymorphic variants. Ethnic differences were found for frequency of some variants in the Polish population compared to others.

Thus, this study may be useful for pharmacogenetics of drugs affected by MRP1-mediated efflux.

Keywords: High resolution melting, *ABCC1*, MRP, Gene scanning, Genotyping

Background

The human multidrug resistance-associated protein 1 (MRP1) is a member of the ATP-binding cassette (ABC) transporter superfamily and is encoded by the *ABCC1* gene [1]. MRP1 was first described by Cole et al. [2] who cloned the overexpressed transporter from lung cancer cell line H69AR. The gene is mapped to chromosome 16p13.1. Its length is around 200,000 base pairs; the gene contains 31 exons encoding a protein of 1,531 amino acids, with molecular weight of 190 kDa [3, 4]. The protein has 3 hydrophobic transmembrane domains (TMDs), also called

membrane-spanning domains (MSDs), named TMD0 (N-terminal), TMD1 (middle) and TMD2 (C-terminal). They contain 17 transmembrane helices (TM), 5 in TMD0 and 6 in both TMD1 and TMD2. The intracellular region of the protein contains 2 hydrophilic nucleotide binding domains (NBDs), NBD1 associated with TMD1 and NBD2 associated with TMD2 [5, 6]. They are both involved in binding and hydrolyzing ATP, which is indispensable for substrate transport. Each of NBDs contains highly conserved motifs: Walker A motif which binds the β -phosphate of ATP and Walker B motif that interacts with Mg^{2+} ions. There is also a third, 13 amino acid sequence motif, between Walker A and Walker B called the ABC signature [1, 7].

MRP1 is ubiquitously expressed in many human tissues and physiological barriers like blood–brain or blood–testis

* Correspondence: strapag@biol.uni.lodz.pl

¹Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Piłarskiego 14/16, 90-231 Łódź, Poland

Full list of author information is available at the end of the article



© 2015 Ślomka et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

barriers. High levels of MRP1 expression were detected in lung, kidney, heart, testis, placenta, adrenal glands and skeletal muscles, while lower levels of expression were found in intestine, colon, brain, peripheral blood mononuclear cells and liver [1, 3, 8]. With the exception of brain cells, MRP1, in contrast to other ABC transporters, is expressed at the basolateral membrane of polarized epithelial cells where it plays a protective role against xenobiotics and their metabolites [3, 9].

Many drugs are good substrates for MRP1, so its overexpression leads to multidrug resistance (MDR), especially during cancer chemotherapy. This effect can be observed in many types of cancer cells including solid tumors (lung cancer, breast cancer, gastric and colon carcinomas, melanoma, prostate cancer, neuroblastoma) as well as in various types of leukemias [3, 9]. MRP1-overexpressing cells are capable of transporting a large group of substrates, including anticancer drugs like folate-based antimetabolites, anthracyclines, *Vinca* alkaloids, antiandrogens and even some inorganic anions such as arsenite and arsenate. MRP1 is also an active transporter of a broad range of endogenous compounds including glutathione (GSH), glucuronate and sulfate-conjugates of organic anions, such as inflammatory mediator leukotriene C₄ or estradiol-17- β -D-glucuronide (E₂17 β G). Therefore, MRP1 plays a crucial role in efflux of organic anions, anionic drugs, xenobiotic conjugates and additionally in the maintenance of redox balance by participation in cellular response to oxidative stress [3, 7, 9–11].

A large number of single nucleotide polymorphisms (SNPs) in *ABCC1* gene were detected so far in studies of various populations, especially Asian and Caucasian. Saito et al. [12] reported 779 genetic variations in eight ABC genes in Japanese population, 95 among them concerning *ABCC1*. Subsequently, Fukushima-Uesaka et al. [13], screening *ABCC1* in the same population, identified 86 genetic variations including 31 novel ones. Similar screening of Chinese population revealed 32 SNPs [14]. Leschziner et al. [15] screened five ABC transporter genes in Caucasian individuals and detected 221 variants, including 61 for *ABCC1*, among them 22 novel for this gene. They identified also large blocks of high linkage disequilibrium and low haplotype diversity across all the genes in the examined population. Comparative study including polymorphisms from four different populations: Chinese, Malay, Indian and Caucasian showed that apart from population-specific SNPs, only 6 of 71 detected SNPs altered amino acid sequence, so *ABCC1* is considered to be a highly conserved gene [16]. Majority of previously reported genetic variations were found in intron sequences. However, there is evidence in literature that some SNPs change protein functioning and effectiveness of cancer chemotherapy. Such cases were reported for neuroblastoma [17], breast cancer [18], ovarian cancer [19] and other

serious diseases like: chronic obstructive pulmonary disease (COPD) [20, 21] cystic fibrosis [22] and major depression [23].

Previously mentioned studies have shown importance of *ABCC1* polymorphisms and its genetic variation. These data can have great relevance for pharmacogenetics in personalized patient therapy. Based on the data described above for several populations, there exist ethnic differences in the frequency of *ABCC1* polymorphic variants. Some of them are so strongly linked to each other that they are always inherited together, forming haplotypes. Haplotype is a combination of polymorphic variants inherited on the same chromosome and sometimes has stronger clinical relevance for drug response or adverse events than a single polymorphism [24]. Moreover, there is a large haplotype diversity across different populations [25]. Therefore, prior comprehensive analysis of the whole gene is required in the specific population before application of genetic variation data in pharmacotherapy.

Our previous studies performed during the project TESTOPLEK in which we have examined genetic variation of multidrug transporter genes in the Polish population indicated that *ABCC1* polymorphisms were present at significant level in a sample of Polish individuals. However, various SNPs in the *ABCC1* gene differed significantly in their frequency in tested DNA samples [26]. This prompted us to verify the whole coding sequence of *ABCC1* gene. We conducted high throughput analysis and screened 30 exons (from 31 in total) of *ABCC1* gene for polymorphism variants in the Polish population (which is exclusively Caucasians). This is the first study of genetic variation of *ABCC1* gene in the Polish population.

Methods

Materials and genomic DNA samples

Human genomic DNA samples were derived from anonymous Polish unrelated volunteers who declared to be healthy. Samples were randomly selected from the “normal Polish population” genetic collection at the Biobank Lab, Department of Molecular Biophysics, University of Lodz. Genetic material for this collection was sampled in 2011 – 2012 within the EU-funded TESTOPLEK project. All subjects gave their written informed consent to participate in the study. This study was approved by the relevant regional ethical committee (Research Bioethics Commission, University of Lodz - Decision no. 8/KBBN-UL/II/2014 and Statement of the Research Bioethics Commission, University of Lodz from 17th of June, 2010) and all procedures were performed in accordance with the Declaration of Helsinki.

Saliva was collected into Oragene OG-500 DNA collection/storage receptacles (DNA Genotek, Kanata, Canada) and genomic DNA was subsequently isolated by the

MagNA Pure LC DNA Isolation Kit - Large Volume (Roche, Basel, Switzerland) with final concentration normalized to 200 pg/μl [27]. A total of 190 samples were enrolled in the scanning study and other 380 in genotyping.

Screening of *ABCC1* by High Resolution Melting (HRM)

Method

Investigation of *ABCC1* genetic variation was conducted using High Resolution Melting method. The method is based on precise measurements of DNA melting profile. The dsDNA melting temperature depends on length and nucleotide composition of the PCR product. Even a single base variation generates a different melting curve, and many variants can be detected in this way. The list of primers used for HRM scanning of all the areas is presented in Table 1. The single reaction mixture (10 μl) was prepared using Janus® Automated Workstation (Perkin Elmer Inc., Waltham, USA) and composed of GoTaq® Colorless Master Mix (2×) (Promega, Madison, USA), LC Green Plus® dye (10×) (BioFire Defense Inc., Salt Lake City, USA), 0.5 μl of 10 μM primers mixture, 3 μl DNA, and filled up to the final volume with water. Reaction was performed on 384-micro well plates using CFX384™ real-time PCR system (Bio-Rad Laboratories Inc., Hercules, USA) on duplicate samples. The reaction conditions were as follows: initial denaturation at 95 °C for 3 min, 50 amplification cycles of denaturation at 95 °C for 30 s and annealing at specific temperature depending on the primers used, for 30 s. The plate was read after each cycle. Directly afterwards, melting curve was determined, the plate being incubated at 90 °C for 60 s, 40 °C for 60 s and from 65 °C to 95 °C with an increment of 0.2 °C for 10 s with plate reading. The obtained data were analysed with the Bio-Rad Precision Melt Analysis Software, version 1.2 (Bio-Rad Laboratories Inc., Hercules, USA). Based on HRM melting curve analyse, confirmation of genetic variation for every melting cluster was obtained by direct sequencing method for several samples selected from each cluster (or all samples from clusters with less than 4 samples). Study workflow design for all described methods is presented in Fig. 1.

HRM genotyping

Independently of the scanning study, we performed an additional genotyping study. We focused on highly polymorphic regions of the *ABCC1* gene existing in databases, often examined in recent publications. We also tested some interesting variants observed in literature that may have clinical significance. The selection provided us with 17 SNPs to check. To identify common genetic variation of selected SNPs in the *ABCC1* gene, we used the HRM analysis method according to conditions described above. Sequences of primers used for genotyping are described in Table 1. Based on HRM

melting curve analysis, confirmation of genetic variation for every melting cluster was obtained by direct sequencing method for several samples selected from each cluster (or all samples from clusters less than 3 samples). We used the data obtained from genotyping to compare them with HRM scanning results (comparison of minor allele frequencies - MAFs) and thereby to validate the accuracy of the screening method.

DNA sequencing

For sequencing, selected samples were prepared by PCR. Reaction mixture filled with water to a final volume of 50 μl per sample was composed using GoTaq® Green Master Mix (2×) (Promega, Madison, USA), 5 μl of 10 μM primers mix and 5 μl DNA. Reaction conditions were as follows: an initial denaturation at 95 °C for 3 min, 45 amplification cycles of denaturation at 95 °C for 30 s, annealing at specific temperature depending on the primers used for 30 s and extension at 72 °C for 45 s, followed by final extension at 72 °C for 5 min (list of primers for exon sequencing is shown in Table 1). Afterwards, PCR products were purified using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). DNA concentration and size of PCR product was determined by agarose electrophoresis of 2 μl of each sample. Based on the intensity of bands the samples were diluted from 5 to 50 times and applied for sequencing reaction using BigDye Terminator V3.1 (Applied Biosystems, USA) according to the manufacturer's protocol. The PCR-sequencing product was purified using BigDye X-Terminator kit following the manufacturer's protocol (Applied Biosystems, USA). Further, 30 μl of each purified sample was applied to 96 wells of titration plates and analyzed in 3500 Genetic Analyzer (Applied Biosystems, USA).

Polymorphism detection

The *ABCC1* genomic DNA sequence (NG_028268.1) was obtained from GenBank (<http://www.ncbi.nlm.nih.gov>) and used as a reference sequence during analysis of sequencing results by CodonCode Aligner software (CodonCode Corporation, Centerville, USA). Sequencing results of selected samples were compared with respective HRM clusters. Based on this, genotype for each melting cluster was established and genetic variation for each sample was verified.

Availability of supporting data

For each of the polymorphisms detected in this study in *ABCC1* gene, the following parameters were assigned: dbSNP IDs (rs numbers) (<http://www.ncbi.nlm.nih.gov/SNP/>), nucleotide position within or relative to the coding sequence based on the NM_004996.3, and amino acid position in the protein for SNPs in the coding sequence

Table 1 Pairs of primers used for HRM screening, sequencing and genotyping ABCC1

Exon	Number of scanned bp from the exon (without primer)	Forward primer (5' → 3')	Reverse primer (5' → 3')	Number of scanned bp downstream from the exon (without primer)	Size (bp) including primers
Scanning	1	104*	CAGCGCTAGCGCCAGCAG	CCCGTGCTTTCCTCC	234
	2	31	GTCTCTGGGGTGTCTCT	TCTGAATGTAGCTCGGTCA	201
	3	20	GCTTTCAGAACACGGTCCTC	TTATCACCACACCACTCC	203
	4	129	GGGGGCTCTGTGTAGGATA	CCACTGAGCCAGCAGAT	215
			AAGCTGAGGCAAGGAGATCA	AAGGTAGCAAGCAGCTGAGG	163
			AGCCTGGGTGACAAGAGTGA	TGGATCTCAGGATGGCTAGG	189
			ATGCTCACTTCTGGCTGGT	GGCAACGATGACTTCTACA	170
	5	27	CAGCCCCAGAATGTGATCTT	CACACACGACACACACT	212
	6	20	TTTCCTCTCTCCCAAC	AGCTGAGCATGTTTATTGTT	182
	7	33	CCCTCCTCTGTCTTGAATC	GAGTGAAGGAGGTGACCA	240
			GGAAACAAGTCGTGCTGTTT	ATCTTGCCAGAACCAAAA	184
	8	74	TTCCCTGAAGGTGACATTC	AGGGTTCACTCTTCTGT	219
			GGAGGCTTTGATCGTCAAGT	AAAGCCAGGAGGGAATG	199
	9	14	AGTATTCAGGCCCTCTCT	TGAGGAAGCAGATGTGGAAG	165
			TACACCGTGTCTGTTTGT	CAGGAGGGATGTGGAAGT	175
	10	27	TCTTCAGTGCACACTCAC	GCCACAGGAGGTAGAGACA	207
			TCTGTGGACGCTCAGAGTT	CTGCCACACGTAGAAGTCC	157
	11	21	TCTTCTGTCTGGTGGTGTGAA	TGGACTAAATCTCATGGAGAG	209
	12	41	GTTGAGTGTGGGTGATCC	GGCAGACTTCTCAGCACCT	217
			GCATTCAAGGACAAGGTGCT	CCTGGCAACATAGTGACCT	206
	13	62	GTCTCAGGGCTGTCACT	ACCGAGGATGTTGAACAAG	184
			CCAGACAGCCTTCGTCTTT	GCTTTCGTGGCTAGAACC	149
14	49		TTAACCTTGGTTGTTTTC	CAAGTAGGGAGGACCCAGTG	228
15	41		GTTCTGTGCATGTGGAGTCG	GGCTCTCTCCACACCATCAC	179
16	73		TTCTCTACTGGGGTAAATTGAGG	CTGAGACAGGGACGACTTT	177
			ATCCCGAAGTGCTTTG	GCTTTTCTCAGACCACCG	171
17	12		CTGTCTCACTGTTCTCCA	CTGACCTTCTCGCCAATCTC	214
			CTCCAGACCTTGAAATCCT	CAGCCCACTGAGATTGTGAG	134
18	20		ATTTCAGGAAACCCACTC	GGCCAATCACATTTTCAAGA	185
			CTTCGATGATCCCTCTCAG	CCTAAAGGGACAGTTCTG	166
19	26		CTCACATGTGCACTCAG	TCTGTGCTGGCATAGTAGG	204

Table 1 Pairs of primers used for HRM screening, sequencing and genotyping ABCC1 (Continued)

		GCCAAAGCTAGGCAGTCTCAC	GGCAAGTAGCTCATGCTGTG	99
		TCTGAGATGGGCTCTACCA	CACCTCAAAGAGGCCAATG	182
	48	TTGTTGCCCTTGTTTTAGC	CTCTCTGTCTCCCACTG	223
20		GCATCTGTACGTTGACACC	TGGCGCTCTATAAACTCCA	228
21		GGTGGTGTCATGTGCTAAG	ATGGGGTCATCAGTCCAGAG	184
22	25	GGCTTCCAACATATTGGTCA	CAAAACGCTGAGGACTCTAAGG	204
23		CCCTCTCTGCATTGTGGAGT	GTCCACTGTGTCAGCTCTT	214
24	21	ATGAGCTTTCTTTGAGCGGAC	GGCTTGTCCCAGAACACTAAG	244
25	22	GAGGGAACCTTCATCAACTCC	TCTGGTTCTCTGTCACCTTC	212
26		CCCATTTCAACGAGACCTTG	GAAAAAGATCGAGGCAAGA	208
27	26	AGAGCTGACTCCATGCCTGT	GGCTCCCCATTAAACAGACT	225
28	24	CGCCGCTTACTCTAGAAAT	CCAAAGACAAGAGGTCAG	184
29	45	GGGGAGTCACAGCTTTACCA	GGGAATGGGTGAGGGAAT	248
30	6	ACACTGGGCCCTTCTGTG	TGATCTCTCTTCGGCAGAC	113
31		CTGGGCTTATTTTCGGATCAA	CAGTGCAATCATAGGGCTTG	174
32	55	AGTCTTAGGTGCGCTCCAT	CCTTCTGCACATTATGGTC	227
33		CCCACCTGAGGACTTCGT	CAAAACCCCCCTACCGAGATG	143
34	74	GTGTCTCTTTTCGTTCTCC	ATGGTGACTGGATGAGGTC	217
35		CCTGCTGAGGAAGACGAAGA	AGGCAAACTCCAAAGCCTA	204
36	15	CCCTTCCCCTCATGTCTGA	ATAGGCCCTGCAGTCTGAC	185
Sequencing	1	-	-	-
2		TCTCGAACTCCAGCCTAAA	TAAGCGGAGAGCAAAAGATT	749
3		GAGCATGTGTACACAGACAAA	CTCTGACCTCAGGTGATCC	638
4		GTGGTGA AACCCCGTCTTTA	CCTTGGAGCAACACAGACAA	604
5		CCCGAGTAGCTGGGATTACA	ACCATGCCAAGTGAGAAACC	606
6		GGGTTGTTGTGGGATAGAA	AAGCAAAAGCAGTCCAGCAT	622
7		CTGCATGACCAACAGAATG	CAAACTCCCGACCAAGTGAT	651
8		CTTTGACTCTGCTTCCTG	TAGGGGAGCCTAGTGGGTTT	691
9		AGCAATTAGTGCAACACCCC	AAGACATCTACACGCCACC	613
10		ACAGCCTGGAAGGTAGAAA	AATGCCAGGGAGCTCTACA	664
11		GAGCGTGACCTGCTTATTC	CACCGCCTACAAGGTCTAA	608
12		GATGTTGAGTGATGGGCTGA	CCTGGGCAACATAGTGACCT	363
13		ATACTGCCCCAGGTTTTTCC	AAAAAGGATAGGAATGGCGG	611

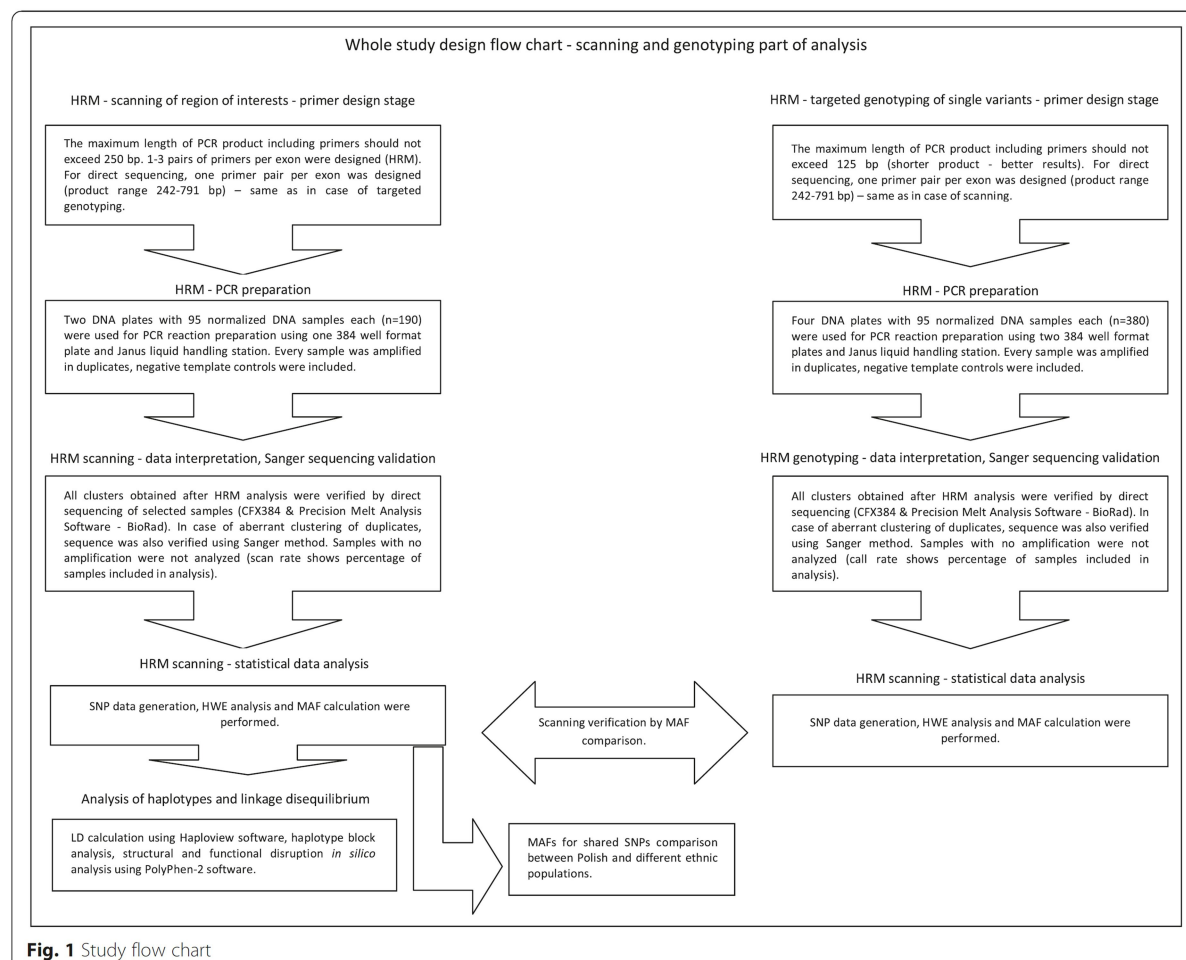
Table 1 Pairs of primers used for HRM screening, sequencing and genotyping ABCC1 (Continued)

14	CTGGGAAACCTTGAAGT	CCAAGGAAAGAAATGCAAG	263
15	CCCAAGATGATGTCAGAT	TAGCAAGCCAAGAATCTCC	632
16	TCCTCTTCAGACCTTGGT	CCTGCCCTTAGGACAGCAC	676
17	ACTCTTTGGAAGCAGGGTT	AGTGAGACCTGAGCCACCC	693
18	CTCAAGCCATCTTCCACTC	GAAAGCCAGCCCTGTGACTTA	717
19	CATGTCCTCCACCTTCAGACT	CCAGCTTAACCTCCGTGCTTC	748
20	CTGCACAGTTCAAGCACT	GCCAAGAGACCTGAGCAAG	620
21	CCGTCTCTTATGCCATGGT	GGTTCAATCCCACTTCTGC	461
22	GCATGTGCTAAGCTGCCTAT	TGTAAATGGGCACACTGGA	485
23	CTCAAGTGATCACCACCT	TGCTTCAAAAGCACCACCT	617
24	GAGGTAGCACTTTGAGCC	ATGATTCACTGTCTCTCCGC	791
25	AAGGCAAGCTTCAGATTCCA	AAAGCATTCCTCTGTACCT	674
26	AAATGCCACGTGACTCTCC	CCTGTGAGGTATCCAGCTC	242
27	AGGGACAGAGGGACACAG	AAATCTGTGGGCTCATTTG	618
28	TCATGATGGAAACTCACCA	GAAGATGAAGTAGGGCAA	669
29	AGACAGGTGTGCCATGTT	TCAACTGAATGGAGCTGGTG	689
30	GGTGTGAGCTGTGTTTGA	AAGAGAGGATCCACCCACTG	556
31	CCATGATTGATGTGGGTAG	AAGCACCAGGAAACCCTTG	640
Genotyping	SNP	Reverse primer (5' → 3')	Size (bp) including primers
c.128G > C	TCAGAACACGGTCTCTCG	ACAGGCCAGAGGTAAAA	49
c.616-7942A > G	AGCAACAGGGCAACAATC	GACGGAGCTAAATGTCCAG	76
c.816G > A	CTTGTGTCAGGAGC	GCAGGATCCTGGAGGAGTA	51
c.825 T > C			
c.1057G > A	GCTCCTTTGCAGGTTGCTCATC	GGGCGCTTCGTGTCATTCA	48
c.1062 T > C	GGTTGCTCATCAAGTTCGTGA	GTCTGGGCGCTTCGTGTC	41
c.1218+8A > G	CATGAGGATCAAGACCGCTG	GTGGAAGTCGGGCCACAT	84
c.1299G > T	CCAGAAATCTCTCAGGTT	GTACGTGGCCAAAGTCCATGA	80
c.1684 T > C	GTTCTGCTTCAGAGGCC	TTCAGGGGAAACCCGAGGAT	125
c.1704C > T			
c.2012G > T	TCACCTTCTCATCCCCG	CACTTTGTCCATCTCAGCCA	104
c.2168G > A	CAGGCTGGATTAGAATGA	ATGGTTCCTCCAGCTGACAT	67

Table 1 Pairs of primers used for HRM screening, sequencing and genotyping ABCC1 (Continued)

c.2581G > A	CAGGAGCTGCTGGCTCG	TCTGTGCTGGCATAGGTACG	62
c.2965G > A	TCCTTTTCATGTGTAACCATGTG	CAATAGTTGGAAGCCAGCG	46
c.3140G > C	CTACTCCATGGCCGTGTCC	AGGATGCTGTGCAGCAGGT	75
c.3436G > A	GAGGTCTACGTGGCTTCC	GAATAGACCGGGGAGCG	69
c.4002G > A	GACAAGTCGGATGCCAG	CGAAATAAGCCCCAGGGTC	112
c.4009A > G	AGCTGGGAAGTCGTCCCT	TGATCTCTCCTTCGGCAG	65

104^b - upstream from the START codon. 21^{**} - downstream from the STOP codon. Bold variants means these ones which genetic variation was detected during genotyping method



based on the reference sequence NP_004987.2. These data were obtained from GenBank (<http://www.ncbi.nlm.nih.gov>) and were used in this paper as the nomenclature of variants.

The obtained data as supporting materials are presented in details in Additional files 1, 2, 3 and 4.

Prediction of impact of SNPs upon protein functionality

Using PolyPhen-2 v2.2.2r398 tool (<http://genetics.bwh.harvard.edu/pph2/>), we performed automatic prediction of effects of missense SNPs and amino acid substitutions for protein functionality and structure. PolyPhen-2 uses damaging alleles datasets, like HumDiv or HumVar, to calculate naïve Bayes posterior probability and to classify a mutation deleterious effect as benign, possibly damaging or probably damaging. This ternary classification depends on thresholds of false positive rate (FPR) denoting the chance that mutation would be falsely qualified as deleterious. For HumDiv-trained model, the thresholds are 5 %

and 10 % FPR while for HumVar-trained model they are 10 % and 20 % FPR, respectively [28].

Linkage disequilibrium and haplotype blocks analysis

The observed genotype distribution was determined for all the detected polymorphisms by performing the Hardy-Weinberg equilibrium (HWE) exact test assuming consistency for P-value higher than 0.001. Linkage disequilibrium (LD) and haplotype block analysis was performed by Haploview 4.2 software (<http://www.broad.mit.edu/mpg/haploview/>). The LD analysis was made using $|D'|$ and r^2 parameters for each variation pair. $|D'|$ coefficient was preferentially used to model recombination rates in examined population, and r^2 parameter to model association power [29, 30].

Studying haplotypes in the human genome requires definition of the location of haplotype blocks, which are regions with little evidence for historical recombination and within which only a few common haplotypes are observed [31]. To determine haplotype blocks in *ABCC1*

for the Polish population, we used the method described previously by Gabriel et al. [31] based on confidence bounds on D' . For pairs of SNPs being in strong linkage, 95 % upper confidence bound D' was higher than 0.98 and the lower bound was above 0.7 while for pairs of SNPs with strong evidences of historical recombination upper bound on D' is lower than 0.9. According to this model, polymorphisms with MAF [22] lower than 0.05 were excluded during the creation of blocks and 16 polymorphic variants with higher MAF were included. Haplotypes with frequency lower than 1 % were grouped as “others”.

Results

ABCC1 genetic variation detected in this study

We screened coding regions and exon/intron boundaries as the most important gene areas (Additional file 1). To assure optimal detection level, the maximal length of tested PCR product was never longer than 250 and 125 base pairs for scanning and genotyping, respectively. For scanning, most of exons had to be divided into separate scanned areas, which provided 51 fragments for analysis. We performed complete polymorphism screening in 30 exons in total. We were unable to scan exon 1 because of problems with amplification during PCR, caused by extremely GC-rich sequence region.

In the scanning study on 190 Polish volunteers, we found 46 different SNPs in the scanned areas, and seven of them being novel and not reported yet (Table 2). 21 SNPs were located in exons and 11 of them change amino acid sequence as non-synonymous variants, including one novel polymorphic variant detected in this study: c.596C > T (p.Ser199Leu) with MAF = 0.003. Among all detected non-synonymous variants, only one, c.2012G > T (p.Gly671Val), occurred as a homozygote with estimated MAF = 0.077. The rest of SNPs found homozygously in exons were synonymous variants which resided at: c.825 T > C (p.Val275=), c.1062 T > C (p.Asn354=), c.1684 T > C (p.Leu562=), c.1704C > T (p.Tyr568=) and c.4002G > A (p.Ser1334=) with MAF values equal to 0.309, 0.332, 0.130, 0.088 and 0.277, respectively. Apart from polymorphisms in coding regions, 23 SNPs were located in introns and a last one (c.*3C > G) in the 3' UTR. Furthermore, one intron deletion variant c.2461-39_2461-38delAT was found during the scanning. Two polymorphic variants located in introns, c.2871 + 26C > T and c.3079 + 10G > A, occurred only as non-reference homozygotes in all subjects.

We used the data from genotyping study to validate our scanning results and compare accuracy of HRM methods (Additional file 2). For genotyping, we tested a subset of 17 SNPs, selected according to the literature or detected during scanning. Seven of them have

not been found during scanning in examined samples and also did not show any genetic variation by genotyping method. Genotype validation by sequencing was performed for all of examined SNPs (Table 3). We calculated MAFs to enable us to assess the accuracy of scanning method compared to genotyping. Results obtained for eight of SNPs were quite similar and the differences were not statistically significant (Table 3). We found statistically significant difference in MAF values obtained in the both our studies for only two *loci* – c.1062 T > C (p.Asn354+) and c.2012G > T (p.Gly671Val).

All the amino acids altered by non-synonymous variants were located in cytoplasmic domains of the protein. Variants c.596C > T (p.Ser199Leu) and c.814C > T (p.Pro272Ser) were located in the third intracellular loop (between TM5 and TM6), variant c.1299G > T (p.Arg433Ser) in the fourth intracellular loop, variant c.3196C > T (p.Arg1066Trp) in the seventh intracellular loop (between TM7 and TM8). Additional four variants located in the loop containing NBD1 alter amino acids sequence: c.1898G > A (p.Arg633Gln) and c.2012G > T (p.Gly671Val) are located 44 and 6 amino acids upstream of the Walker A motif, respectively, while c.2168G > A (p.Arg723Gln) and c.2876A > G (p.Lys959Arg) are located 37 amino acids downstream of this motif, respectively. Similarly, three variants changing amino acids in the loop containing NBD2 were detected: c.3886C > T (p.Arg1296Trp) and c.3901C > T (p.Arg1301Cys) are located 30 and 25 amino acids upstream of the Walker A motif, respectively, while c.4093G > A (p.Asp1365Asn) is located 30 amino acids downstream of the motif. Detected variant c.4002G > A (p.Ser1334=) occurred in the last position of Walker A motif in NBD2, however it does not change the amino acid sequence.

Analysis of all the non-synonymous variants detected in this study by the PolyPhen-2 tool (data in Additional file 3) showed for HumDiv-trained model that five of them: c.814C > T (p.Pro272Ser), c.1898G > A (p.Arg633Gln), c.2168G > A (p.Arg723Gln), c.2876A > G (p.Lys959Arg), the novel one c.4093G > A (p.Asp1365Asn), probably have benign influence on the functioning of the protein. For one polymorphism, c.3196C > T (p.Arg1066Trp) a possibly damaging effect on protein activity was expected. The novel variant c.596C > T (p.Ser199Leu) was estimated as a probably damaging substitution, likewise as four others: c.1299G > T (Arg433Ser), c.2012G > T (p.Gly671Val), c.3886C > T (p.Arg1296Trp) and c.3901C > T (p.Arg1301Cys). On the other hand, analysis for HumVar-trained model indicated that three polymorphisms: c.1299G > T (Arg433Ser), c.2012G > T (p.Gly671Val), c.3901C > T (p.Arg1301Cys), lead to probably damaging substitutions and two others, c.596C > T (p.Ser199Leu) and c.3886C > T (p.Arg1296Cys), are possibly damaging variants.

Table 2 Summary of ABCC1 variants detected during scanning by HRM

Exon scanned by HRM	dbSNP ID	Variant position NM_004996.3:	Intron/amino acid residue NP_004987.2:	Observed genotypes ^{a, b} (n)			HWE exact test P-value ^c	MAF ^d
				R/R	R/V	V/V		
2	rs8187843	c.225 + 26G > A	Intron	164	25	0	1	(A) 0.066
4	rs587783373*	c.352-79G > A	Intron	185	1	0	1	(A) 0.003
4	rs4148337	c.352-66 T > C	Intron	15	80	91	0.727	(T) 0.296
5	rs483352860*	c.596C > T	p.Ser199Leu	186	1	0	1	(T) 0.003
6	rs8187846	c.677 + 17C > T	Intron	188	1	0	1	(T) 0.003
7	rs483352864*	c.809 + 16C > T	Intron	188	1	0	1	(T) 0.003
7	rs45609533	c.809 + 31G > T	Intron	183	5	0	1	(T) 0.013
7	rs903880	c.809 + 54C > A	Intron	112	65	11	0.684	(A) 0.231
7	rs246232	c.809 + 64C > G	Intron	84	90	14	0.174	(G) 0.314
8	rs546943313	c.810-73C > T	Intron	187	1	0	1	(T) 0.003
8	rs200194736	c.814C > T	p.Pro272Ser	187	1	0	1	(T) 0.003
8	rs2230669	c.816G > A	p.Pro272=	172	16	0	1	(A) 0.043
8	rs246221	c.825 T > C	p.Val275=	84	92	12	0.059	(C) 0.309
8	rs587783372*	c.855G > A	p.Pro285=	187	1	0	1	(A) 0.003
9	rs35587	c.1062 T > C	p.Asn354=	78	91	16	0.185	(C) 0.332
9	rs35588	c.1218 + 8A > G	Intron	82	91	16	0.245	(G) 0.327
9	rs483352877*	c.1218 + 9C > T	Intron	188	1	0	1	(T) 0.003
10	rs60782127	c.1299G > T	p.Arg433Ser	186	2	0	1	(T) 0.005
12	rs17265551	c.1677 + 56C > T	Intron	162	27	0	0.604	(T) 0.072
13	rs35604	c.1678-37G > A	Intron	2	45	142	0.745	(G) 0.130
13	rs483352863*	c.1678-34G > A	Intron	188	1	0	1	(A) 0.003
13	rs35605	c.1684 T > C	p.Leu562=	2	45	142	0.745	(T) 0.130
13	rs8187858	c.1704C > T	p.Tyr568=	157	31	1	1	(T) 0.088
14	rs112282109	c.1898G > A	p.Arg633Gln	187	1	0	1	(A) 0.003
16	rs8187863	c.2001C > T	p.Ser667=	187	1	0	1	(T) 0.003
16	rs45511401	c.2012G > T	p.Gly671Val	161	25	2	0.296	(T) 0.077
17	rs4148356	c.2168G > A	p.Arg723Gln	181	9	0	1	(A) 0.024
19	rs45607032	c.2461-39_2461-38delAT	Intron	179	9	0	1	(delAT) 0.024
19	rs2074087	c.2461-30C > G	Intron	0	44	144	0.083	(C) 0.117
19	rs45492500	c.2461-27G > A	Intron	172	14	2	0.056	(A) 0.048
21	rs11075296	c.2871 + 26C > T	Intron	0	0	189	1	-
22	rs768191257	c.2876A > G	p.Lys959Arg	187	1	0	1	(G) 0.003
22	rs3851716	c.3079 + 10G > A	Intron	0	0	188	1	-
22	rs34794353	c.3079 + 24C > T	Intron	187	1	0	1	(T) 0.003
22	rs3887893	c.3079 + 62 T > C	Intron	67	96	25	0.358	(C) 0.388
23	rs191017838	c.3171G > A	p.Leu1057=	187	2	0	1	(A) 0.005
23	rs199773531	c.3196C > T	p.Arg1066Trp	188	1	0	1	(T) 0.003
25	rs41278168	c.3591-5C > T	Intron	187	1	0	1	(T) 0.003
27	rs200922662	c.3886C > T	p.Arg1296Trp	187	1	0	1	(T) 0.003
27	rs201533167	c.3901C > T	p.Arg1301Cys	187	1	0	1	(T) 0.003

Table 2 Summary of ABCC1 variants detected during scanning by HRM (Continued)

28	rs2230671	c.4002G > A	p.Ser1334=	102	68	18	0.202	(A) 0.277
28	rs188980645	c.4093G > A	p.Asp1365Asn	187	1	0	1	(A) 0.003
29	rs212087	c.4126-45G > A	Intron	62	85	42	0.239	(A) 0.447
30	rs212088	c.4487 + 18G > A	Intron	136	47	5	0.775	(A) 0.148
31	rs587783374*	c.4551G > A	p.Gln1517=	186	1	0	1	(A) 0.003
31	rs373453875	c.*3C > G	3' UTR	186	1	0	1	(G) 0.003

*Novel polymorphic variants detected in this study. ^aNumber of genotypes detected during this study, R – reference allele, V – variant allele. ^bTotal number of examined samples was 190, however during scanning single samples were excluded because of accidental problems with amplification during PCR and hence, total number of genotypes is not equal precisely 190. ^cP-value is consistent with Hardy-Weinberg equilibrium if $P > 0.001$. ^dMinor allele shown in brackets with its frequency. Bold variants signifies the ones which were validated by genotyping results

Linkage disequilibrium analysis

Based on full genotype sets of 44 polymorphic variants confirmed by Hardy-Weinberg equilibrium exact test (Table 2), linkage disequilibrium analysis using r^2 and $|D'|$ statistics was performed (Additional file 4). Accordingly to the r^2 parameter, perfect linkage ($r^2 = 1$) was detected between the new variant c.809 + 16C > T and c.3196C > T, between c.1678-37G > A and c.1684 T > C, between c.2168G > A and c.2461-39_2461-38delAT. Almost perfect linkage was observed between c.1062 T > C and c.1218 + 8A > G ($r^2 = 0.928$), between c.825 T > C (p.Val275=) and c.1062 T > C (p.Asn354=) ($r^2 = 0.822$) and between c.825 T > C (p.Val275=) and c.1218 + 8A > G ($r^2 = 0.894$). Apart from perfect linkage between variants

c.1678-37G > A and c.1684 T > C, they were both strongly linked ($r^2 = 0.705$) with c.2461-30C > G. Other linkages ($r^2 = 0.624$ and $r^2 = 0.658$, respectively) were found between c.225 + 26G > A and c.816G > A, and between c.809 + 54C > A and c.809 + 64C > G.

For the $|D'|$ values, perfect linkage ($|D'| = 1$) was observed for the great majority of variants pairs (760 for 946). Strong linkage ($1 > |D'| \geq 0.8$) was detected for other 17 pairs. Small areas of low $|D'|$ values, and hence weak or lacking of linkage, were defined between polymorphisms from group: c.225 + 26G > A, c.352-66 T > C, c.809 + 31G > T, c.809 + 54C > A, c.809 + 64C > G, c.816G > A, c.825 T > C, c.1062 T > C, c.1218 + 8A > G, c.1677 + 56C > T, and those from group: c.1678-37G > A, c.1684 T > C,

Table 3 Summary of ABCC1 selected SNPs genotyping by HRM and comparing them with scanning results

dbSNP ID	Variant residue NM_004996.3:	Intron/amino acid residue NP_004987.2:	Observed genotypes ^a (n)			HWE exact test P-value ^b	MAF ^c (genotyping)	MAF ^c (scanning)	Chi-square test P-value ^d
			R/R	R/V	V/V				
rs41395947	c.128G > C	p.Cys43Se	380	0	0	1	–	–	–
rs2230669	c.816G > A	p.Pro272=	362	18	0	1	(A) 0.024	(A) 0.043	0.079
rs246221	c.825 T > C	p.Val275	197	160	23	0.243	(C) 0.271	(C) 0.309	0.187
rs8187852	c.1057G > A	p.Val353Met	379	0	0	1	–	–	–
rs35587	c.1062 T > C	p.Asn354=	204	142	33	0.247	(C) 0.274	(C) 0.332	0.044
rs35588	c.1218 + 8A > G	Intron	190	160	30	0.709	(G) 0.289	(G) 0.325	0.214
rs60782127	c.1299G > T	p.Arg433Ser	373	6	0	1	(T) 0.008	(T) 0.005	0.623
rs35605	c.1684 T > C	p.Leu562=	13	105	262	0.588	(T) 0.172	(T) 0.130	0.063
rs8187858	c.1704C > T	p.Tyr568=	325	55	0	0.242	(T) 0.072	(T) 0.087	0.374
rs45511401	c.2012G > T	p.Gly671Val	346	28	3	0.007	(T) 0.045	(T) 0.077	0.038
rs4148356	c.2168G > A	p.Arg723Gln	360	19	0	1	(A) 0.025	(A) 0.024	0.888
rs45517537	c.2581G > A	p.Ala861Thr	380	0	0	1	–	–	–
rs35529209	c.2965G > A	p.Ala989Thr	378	0	0	1	–	–	–
rs13337489	c.3140G > C	p.Cys1047Ser	380	0	0	1	–	–	–
rs28706727	c.3436G > A	p.Val1146Ile	380	0	0	1	–	–	–
rs2230671	c.4002G > A	p.Ser1334=	204	140	32	0.296	(A) 0.271	(A) 0.277	0.850
rs28364006	c.4009A > G	p.Thr1337Ala	380	0	0	1	–	–	–

^aNumber of genotypes detected during this study, R – reference allele, V – variant allele. ^bP-value is consistent with Hardy-Weinberg equilibrium if $P > 0.001$. ^cMinor allele shown in brackets with its frequency. ^dP-value of Chi-square test with Yates' correction, no significant difference between MAFs if $P > 0.05$

c.1704C > T, c.2012G > T, c.2168G > A, c.2461-39_2461-38delAT, c.2461-30C > G, c.2461-27G > A, c.3079 + 62 T > C, c.4002G > A, c.4126-45G > A, c.4487 + 18G > A.

Haplotype block analysis

Based on the model of Gabriel et al. [31] a total of 12 common haplotypes were identified and their frequencies were defined in four blocks (Table 4). In block 1, variants c.809 + 54C > A and c.809 + 64C > G were included, and the most frequent haplotype in this block was that containing major alleles (frequency: 0.686), the next most frequent one was the haplotype with both minor alleles (frequency: 0.231). Block 2 spanned SNPs c.825 T > C (p.Val275=), c.1062 T > C (p.Asn354=) and c.1218 + 8A > G at a distance of 1 kb, with the most frequent haplotype containing all major alleles (frequency: 0.659), subsequently the haplotype with all minor alleles (frequency: 0.300), two other ones had low frequency (set alleles TCG: 0.021 and set alleles TCA: 0.011). Block 3 had two haplotypes composed of SNPs c.1678-37G > A and c.1684 T > C (p.Leu562=). The first haplotype was the set with major alleles (frequency: 0.870) and the second one the set with minor alleles (frequency: 0.130). Block 4 was created of SNPs c.4126-45G > A and c.4487 + 18G > A at a distance of around 2 kb. The most frequent haplotype was that of alleles combination AG (frequency: 0.447), the next ones with major alleles (frequency: 0.405) and allele combination GA (frequency: 0.148). The frequency of recombination was also calculated between blocks as a value of multiallelic D' coefficient, and equaled to 0.65 between block 1 and 2, 0.55 between block 2 and 3 and 0.27 between block 3 and 4. HaploView predicted 51 possible connections of haplotypes for recombination between blocks higher than 1 %. However, recombination between blocks higher than 10 % occurred only between the most frequent haplotype from block 1 and the most frequent one from block 2, between three most frequent haplotypes from block 2 and the most frequent one from block 3, and between the most frequent one from block 3 and three haplotypes from block 4. Apart from c.1684 T > C (p.Leu562=), the remaining SNPs in blocks were tagged *in silico* as haplotype tag SNPs (htSNPs).

Discussion

In this paper we highlighted the fact that the High Resolution Melting method is a powerful technique both for targeted genotyping of selected SNPs and for scanning of long parts of gene sequences. The obtained results for targeted genotyping and scanning of *ABCC1* gene regions did not show statistically significant differences in MAFs values for most SNPs. Marginal statistical significance ($0.05 > P > 0.01$) of difference between MAF values for scanning and targeted genotyping methods for two SNPs c.1062 T > C (p.Asn354=) and c.2012G > T (p.Gly671Val)

is probably caused by sample size effect (there was no overlap between groups of individuals selected for both stages of the study). Seven SNPs from a subset of 17 individually genotyped ones were not detected during the scanning and did not show any genetic variation during genotyping either. Although these variants were previously indicated to may have clinical significance and were therefore included in our study, they were only identified during *in vitro* studies or in non Caucasian populations (Exome Aggregation Consortium (ExAC), Cambridge, MA (<http://exac.broadinstitute.org>), date accessed: Aug 2015). Thus, it was expected not to find them in our study for 380 genotyped samples, where technical or methodological problems with detection should not have caused any discrepancy.

During the scanning study we identified 46 polymorphic variants by screening 30 exons of the *ABCC1* gene with adjacent intronic fragments in 190 unrelated Polish individuals. A large fraction of variants represented rare or even completely novel (not previously published) ones. We were unable to compare them with the ones from the only other publication on *ABCC1* variant distribution in Caucasians because of unavailability of the raw data from that study [15].

Comparing our genetic variation results obtained using the scanning method with data for Asian populations, 18 variants were detected in the Polish population which were present also in the Japanese one ($n = 153$) [13], while 13 variants from the Polish population were also detected in the Chinese one ($n = 27$) [14]. Some of these variants exhibited significant differences in MAF values, confirmed by chi-square test for alleles (with Yates' correction if any allele quantity was below 5) and showed as their P-values, assumed significant if < 0.05 . Higher MAF for the Polish population was detected for variant c.809 + 54C > A (MAF = 0.231, compared to 0.059 for Japanese ($P < 0.001$) and 0.093 for Chinese ($P = 0.020$) populations), synonymous SNP c.4002G > A (MAF = 0.277, compared to 0.196 for Japanese ($P = 0.014$) and only 0.111 for Chinese ($P = 0.009$) populations) and for variant c.4126-45G > A between Polish (MAF = 0.447) and Japanese (MAF = 0.304) populations ($P < 0.001$). On the other hand, lower MAF for Polish population was noticed for variants: c.809 + 64C > A (0.313 for Polish, compared 0.418 ($P = 0.004$) for Japanese and 0.481 ($P = 0.014$) for Chinese populations), c.1678-37G > A and synonymous SNP c.1684 T > C (for both SNPs, MAF = 0.130 for Polish, 0.245 ($P < 0.001$) for Japanese but not significant for Chinese 0.204 ($P = 0.141$)), non-synonymous SNP c.2168G > A and deletion variant c.2461-39_2461-38delAT (for both variants, MAF = 0.024 for Polish, 0.065 ($P = 0.007$) for Japanese and not significant for Chinese 0.065 ($P = 0.108$)), c.2461-30C > G (for Polish MAF = 0.117, not significant for Chinese 0.167 ($P = 0.299$) but for Japanese 0.245

Table 4 Haplotype block analysis according to Gabriel et al. [31]

Variant position ^{a,b}	Block 1		Block 2		Block 3		Block 4	
	Haplotype frequency	Recombination between blocks ^d	Haplotype frequency	Recombination between blocks ^d	Haplotype frequency	Recombination between blocks ^d	Haplotype frequency	Recombination between blocks ^d
	c.809 + 54C >A	c.809 + 64C >G	c.825 T >C	c.1062 T >C	c.1218 + 8A >G	c.1678- 37G >A	c.4126- 45G >A	c.4487 + 18G N A
Haplotypes ^c	C	C	T	T	A	A	A	G
	A	G	C	C	G	G	G	G
	C	G	T	C	G		G	A
others	0.001	0.001	T	C	A			
			others					
					0.009			

^aposition according to reference sequences for coding nucleotides NM_004996.3 and amino acid residues NP_004987.2. ^bBold variants signify haplotype tag SNPs (htSNPs). ^cMajor alleles in white boxes, minor alleles in shaded boxes; all the haplotypes with their frequencies in population; haplotypes below 1 % grouped as "others". ^dLevel of recombination between blocks as a value of multiallelic D'

($P < 0.001$) and c.4487 + 18G > A between Polish (MAF = 0.148) and Japanese (MAF = 0.291, $P < 0.001$) populations. These results suggest that although a wide group of common variants for different populations have had similar MAF, there are also other polymorphisms which MAF values were significantly different. Since differences between Asian populations were consistently smaller than between them and the Polish population, we can conclude that ethnic ancestry is probable underlying cause of these variations. Similar conclusions were described previously during research of four polymorphisms in *ABCC1* gene in Caucasians: c.816G > A, c.825 T > C, c.1684 T > C, c.4002G > A [32]. All these results correspond to those obtained by Wang et al. [16] who studied 60 SNPs and 10 deletion variants for four populations. During our *ABCC1* scanning, we detected 25 common variants and obtained similar results for Polish population as those authors for Caucasian representatives. A slight difference in MAF was observed for few polymorphisms: the highest were seen for c.809 + 64C > G (MAF = 0.313 for Polish and 0.200 ($P = 0.043$) for Caucasians) and for c.825 T > C (p.Val275=) (MAF = 0.309 for Polish population and 0.139 ($P = 0.003$) for Caucasians), while the difference for c.4002G > A (MAF = 0.277 and 0.375 ($P = 0.093$) for Polish and Caucasians, respectively) was not significant. Furthermore, their results confirmed our previous observations concerning ethnic differences in occurrence of the same polymorphic variants between Caucasian and Asian populations. These differences relate mainly to the SNPs mentioned previously as most ethnically significant: c.809 + 54C > A, c.809 + 64C > A, c.3079 + 62 T > C, c.4002G > A and c.4126-45G > A.

We performed analysis of linkage disequilibrium (which is defined as non-random allelic association of different *loci*) to summarize our results for those 44 polymorphisms which were found to be consistent with the Hardy-Weinberg model. It turned out that, as expected, the r^2 coefficient was most useful for analysis, and we detected polymorphic variant pairs with strong LD: our results are very similar even to those obtained for Japanese population [13]. It can be concluded on this basis that the specific linkages between particular variants are highly evolutionary conserved and occur in every population. Human genome is organized in haplotype blocks with high LD inside them, separated one from another by regions of low LD. Blocks show no evidence of historical recombination inside. This definition of haplotype blocks could be applied to 17 SNPs which had MAF higher than 0.05, which was the lower threshold in this analysis. We found four blocks and 12 common haplotypes including nine SNPs detected during this study. We determined htSNPs which should be sufficient to identify common haplotypes in population. Low haplotype diversity of *ABCC1* in Polish population

is consistent with the conclusion of another study on Caucasians [15]. Although blocks for our results are small and the number of SNPs used for haplotype construction is low, the main core of blocks and their borders are retained across populations regardless of the blocks definitions used. Our blocks 1 and 2, usually connected as one in other studies, extend from intron 7 to intron 9. Our block 3 include SNPs from vicinity of exon 13 and block 4 contains SNPs from intron 28 to intron 30. It corresponds to block borders detected in other studies and suggest points of recombination which had taken place in intron 12 and in intron 27 or 28 [13–15].

Analysis using PolyPhen-2 which predicts the influence of detected non-synonymous SNPs on protein functionality suggested that six of them are suspect. However, each analysis is only a statistical calculation based on a full dataset and some experimental data do not support these results. Létourneau et al. [33] examined 10 missense SNPs of *ABCC1* and almost all of them, classified by PolyPhen-2 software as deleterious, affected neither expression nor functionality of the protein. A similar observation was also reported previously for the SNP c.2012G > T (p.Gly671Val), detected also by us. Although this substitution affects a residue near the highly conserved motif Walker A, it reduced mRNA expression but did not change transport properties of the protein, although PolyPhen-2 analysis clearly showed its damaging impact [34]. Recently, other data confirmed these observations and none of the amino acid substitutions: p.Arg633Gln, p.Gly671Val, p.Arg723Gln, detected also in our study, was found to change functionality of MRP1 transporter. Two substitutions altered drug resistance and inhibitor sensitivity and we found one of them, c.4002G > A (p.Ser1334=) [35].

On the other hand, there are other papers showing contrasting data for above-mentioned variants, proving their clinical association or experimental impact on protein functionality. Thus, it was demonstrated that variant c.2168G > A (p.Arg723Gln) reduced resistance activity of MRP1-overexpressing cells to drugs like daunorubicin, doxorubicin, etoposide, vinblastine and vincristine *in vitro* [36]. Although correlation between carrying this SNP and chemotherapy response was not observed in lung cancer patients [37], this variant was linked to the response to chemotherapy in patients with ovarian cancer [19]. Two other polymorphic variants detected in this study c.825 T > C (p.Val275=) and c.2012G > T (p.Gly671Val), were correlated with febrile neutropenia as an effect of FEC-induced hematological toxicity in breast cancer patients [38]. The major allele variant c.825 T > C (p.Val275=), in combination with another one c.*543C > T (beyond our scanning), was also associated with anthracycline-induced cardiotoxicity after chemotherapy in children with acute lymphoblastic leukemia [39]. The

discrepancy observed in different studies on clinical significance of c.2012G > T (p.Gly671Val) and c.2168G > A (p.Arg723Gln) is of unclear origin. Cellular conditions *in vitro* in which mutants were expressed could be inadequate to reveal all their real functions, e.g. a change in mRNA structure could be more essential than the amino acid change caused by SNP [34]. Conseil and Cole [35] concluded from this discordance that comprehensive biochemical and pharmacological characterization of the respective MRP1 mutants should be performed to identify consequences of *ABCC1* SNPs on phenotype because these effects might be quite selective. Additional nucleic acid-based methods should be also used to clarify the impact of individual SNPs.

These examples show that only comprehensive examination of polymorphisms and studying their phenotypes can tell more about their real influence. Therefore, novel variants like c.596C > T (p.Ser199Leu), detected in this study and qualified *in silico* as damaging, need to be checked in an experimental investigation. Additionally, the fact that we found no novel mutation in important nucleotide binding motifs, nor any of the mutations identified previously *in vitro* as deleterious in NBD loops, proved that these highly conserved regions in the protein structure are unlikely to be affected by polymorphisms occurring in the healthy population.

Conclusions

We scanned 30 exons and short flanking intron sequences of *ABCC1*. HRM is an efficient and sensitive method for scanning and genotyping polymorphic variants. We compared results of long-range scanning to targeted genotyping of selected SNPs showing that both methods are highly sensitive tools to study genetic variation. During our study we found 45 polymorphic variants, some of them rare and eight of them novel not previously reported. This is the first study about *ABCC1* genetic variation in the Polish population and we found some ethnic differences in frequency of genetic variants compared to other populations. These results need further complementation with a larger number of probands and new data including also introns and untranslated regions as sequences with polymorphic variants which may have clinical relevance as well.

Additional files

Additional file 1: Supplemental materials for scanning study (include HRM plots and sequencing traces). (DOCX 12174 kb)

Additional file 2: Supplemental materials for targeted genotyping study (include HRM plots and sequencing traces). (DOCX 5582 kb)

Additional file 3: Supplemental materials for structural and functional disruption - *in silico* analysis using PolyPhen-2 software. (DOCX 13 kb)

Additional file 4: Supplemental materials for linkage disequilibrium analysis (include pairwise plots with r^2 and $|D'|$ parameters). (DOCX 327 kb)

Abbreviations

ABC transporters: ATP-binding cassette transporters; E₂17βG: Estradiol-17-β-D-glucuronide; GSH: Glutathione; HRM: High resolution melting; htSNPs: Haplotype tag SNPs; HWE: Hardy Weinberg equilibrium; LD: Linkage disequilibrium; MAF: Minor allele frequency; MRP1: Multidrug resistance-associated protein 1; NBD: Nucleotide binding domain; TM: Transmembrane helix; TMD: Transmembrane domain.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived and designed the experiments: DS, MSI. Performed the experiments: MSI, MSo, MKM. Analysed data: MSI, MSo, DS. Contributed reagents/materials/analysis tools DS, GB. Contributed to the writing of manuscript: MSI, DS, MSo, JD, GB. All authors approved the final manuscript.

Acknowledgments

The authors would like to thank the reviewers for their comments that help to improve the manuscript. In addition we would like to thank Łukasz Pułaski for fruitful discussion and language editing. This work was supported by the Polish POIG grant 01.01.02-10-005/08 TESTOPEK from the European Regional Development Fund.

Author details

¹Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Piłarskiego 14/16, 90-231 Łódź, Poland. ²Institute of Medical Biology, Polish Academy of Sciences, Lodowa 106, 93-232 Łódź, Poland. ³Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Pomorska 141/143, 90-236 Łódź, Poland.

Received: 14 April 2015 Accepted: 11 September 2015

Published online: 23 September 2015

References

- Leslie EM, Deeley RG, Cole SP. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. *Toxicol Appl Pharmacol*. 2005;204(3):216–37.
- Cole SP, Bhardwaj G, Gerlach JH, Mackie JE, Grant CE, Almquist KC, et al. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science*. 1992;258(5088):1650–4.
- Bakos E, Homolya L. Portrait of multifaceted transporter, the multidrug resistance-associated protein 1 (MRP1/ABCC1). *Pflugers Arch*. 2007;453(5):621–41.
- Chang XB. A molecular understanding of ATP-dependent solute transport by multidrug resistance-associated protein MRP1. *Cancer Metastasis Rev*. 2007;26(1):15–37.
- Choudhuri S, Klaassen CD. Structure, function, expression, genomic organization, and single nucleotide polymorphisms of human ABCB1 (MDR1), ABCC (MRP), and ABCG2 (BCRP) efflux transporters. *Int J Toxicol*. 2006;25(4):231–59.
- DeGorter MK, Conseil G, Deeley RG, Campbell RL, Cole SP. Molecular modeling of the human multidrug resistance protein 1 (MRP1/ABCC1). *Biochem Biophys Res Commun*. 2008;365(1):29–34.
- Leslie EM, Deeley RG, Cole SP. Toxicological relevance of the multidrug resistance protein 1, MRP1 (ABCC1) and related transporters. *Toxicology*. 2001;167(1):3–23.
- Flens MJ, Zaman GJ, van der Valk P, Izquierdo MA, Schroeijs AB, Scheffer GL, et al. Tissue distribution of the multidrug resistance protein. *Am J Pathol*. 1996;148(4):1237–47.
- Sharom FJ. ABC multidrug transporters: structure, function and role in chemoresistance. *Pharmacogenomics*. 2008;9(1):105–27.
- Deeley RG, Cole SP. Substrate recognition and transport by multidrug resistance protein 1 (ABCC1). *FEBS Lett*. 2006;580(4):1103–11.

11. Munoz M, Henderson M, Haber M, Norris M. Role of the MRP1/ABCC1 multidrug transporter protein in cancer. *IUBMB Life*. 2007;59(12):752–7.
12. Saito S, Iida A, Sekine A, Miura Y, Ogawa C, Kawauchi S, et al. Identification of 779 genetic variations in eight genes encoding members of the ATP-binding cassette, subfamily C (ABCC/MRP/CFTR). *J Hum Genet*. 2002;47(4):147–71.
13. Fukushima-Uesaka H, Saito Y, Tohkin M, Maekawa K, Hasegawa R, Kawamoto M, et al. Genetic variations and haplotype structures of the ABC transporter gene ABCC1 in a Japanese population. *Drug Metab Pharmacokinet*. 2007;22(1):48–60.
14. Wang H, Hao B, Zhou K, Chen X, Wu S, Zhou G, et al. Linkage disequilibrium and haplotype architecture for two ABC transporter genes (ABCC1 and ABCG2) in Chinese population: implications for pharmacogenomic association studies. *Ann Hum Genet*. 2004;68(Pt 6):563–73.
15. Leschzner G, Zabaneh D, Pirmohamed M, Owen A, Rogers J, Coffey AJ, et al. Exon sequencing and high resolution haplotype analysis of ABC transporter genes implicated in drug resistance. *Pharmacogenet Genomics*. 2006;16(6):439–50.
16. Wang Z, Sew PH, Ambrose H, Ryan S, Chong SS, Lee EJ, et al. Nucleotide sequence analyses of the MRP1 gene in four populations suggest negative selection on its coding region. *BMC Genomics*. 2006;7:111.
17. Pajic M, Murray J, Marshall GM, Cole SP, Norris MD, Haber M. ABCC1 G2012T single nucleotide polymorphism is associated with patient outcome in primary neuroblastoma and altered stability of the ABCC1 gene transcript. *Pharmacogenet Genomics*. 2011;21(5):270–9.
18. Kunicka T, Vaclavikova R, Hlavac V, Vrana D, Pecha V, Raus K, et al. Non-coding polymorphisms in nucleotide binding domain 1 in ABCC1 gene associate with transcript level and survival of patients with breast cancer. *PLoS One*. 2014;9(7):e101740.
19. Obata H, Yahata T, Quan J, Sekine M, Tanaka K. Association between single nucleotide polymorphisms of drug resistance-associated genes and response to chemotherapy in advanced ovarian cancer. *Anticancer Res*. 2006;26(3B):2227–32.
20. Siedlinski M, Boezen HM, Boer JM, Smit HA, Postma DS. ABCC1 polymorphisms contribute to level and decline of lung function in two population-based cohorts. *Pharmacogenet Genomics*. 2009;19(9):675–84.
21. Budulac SE, Postma DS, Hiemstra PS, Kunz LI, Siedlinski M, Smit HA, et al. Multidrug resistance-associated protein-1 (MRP1) genetic variants, MRP1 protein levels and severity of COPD. *Respir Res*. 2010;11:60.
22. Mafficini A, Ortombina M, Sermet-Gaudelus I, Lebecque P, Leal T, Iansa P, et al. Impact of polymorphism of Multidrug Resistance-associated Protein 1 (ABCC1) gene on the severity of cystic fibrosis. *J Cyst Fibros*. 2011;10(4):228–33.
23. Lee SH, Lee MS, Lee JH, Kim SW, Kang RH, Choi MJ, et al. MRP1 polymorphisms associated with citalopram response in patients with major depression. *J Clin Psychopharmacol*. 2010;30(2):116–25.
24. Judson R, Stephens JC, Windemuth A. The predictive power of haplotypes in clinical response. *Pharmacogenomics*. 2000;1(1):15–26.
25. Wang Z, Wang B, Tang K, Lee EJ, Chong SS, Lee CG. A functional polymorphism within the MRP1 gene locus identified through its genomic signature of positive selection. *Hum Mol Genet*. 2005;14(14):2075–87.
26. Majewska M, Sobalska M, Siewierska A, Strapagiel D, Bartosz G. Single nucleotide polymorphisms of selected ABCC genes in the Polish population. In 4th FEBS special meeting ATP-binding cassette (ABC) proteins: from multidrug resistance to genetic diseases; March 3–9, 2012; Innsbruck, Austria: FEBS; 2012. pp. 84.
27. Koszarska M, Kucsmas N, Kiss K, Varady G, Gera M, Antalffy G, et al. Screening the expression of ABCB6 in erythrocytes reveals an unexpectedly high frequency of Lan mutations in healthy individuals. *PLoS One*. 2014;9(10):e111590.
28. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–9.
29. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet*. 2003;12(7):771–6.
30. Mueller JC. Linkage disequilibrium for different scales and applications. *Brief Bioinform*. 2004;5(4):355–64.
31. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576):2225–9.
32. Oselin K, Mrozikiewicz PM, Gaikovitch E, Pakla R, Roots I. Frequency of MRP1 genetic polymorphisms and their functional significance in Caucasians: detection of a novel mutation G816A in the human MRP1 gene. *Eur J Clin Pharmacol*. 2003;59(4):347–50.
33. Letourneau IJ, Deeley RG, Cole SP. Functional characterization of non-synonymous single nucleotide polymorphisms in the gene encoding human multidrug resistance protein 1 (MRP1/ABCC1). *Pharmacogenet Genomics*. 2005;15(9):647–57.
34. Conrad S, Kauffmann HM, Ito K, Deeley RG, Cole SP, Schrenk D. Identification of human multidrug resistance protein 1 (MRP1) mutations and characterization of a G671V substitution. *J Hum Genet*. 2001;46(11):656–63.
35. Conseil G, Cole SP. Two polymorphic variants of ABCC1 selectively alter drug resistance and inhibitor sensitivity of the multidrug and organic anion transporter multidrug resistance protein 1. *Drug Metab Dispos*. 2013;41(12):2187–96.
36. Yin JY, Huang Q, Yang Y, Zhang JT, Zhong MZ, Zhou HH, et al. Characterization and analyses of multidrug resistance-associated protein 1 (MRP1/ABCC1) polymorphisms in Chinese population. *Pharmacogenet Genomics*. 2009;19(3):206–16.
37. Yin JY, Han LF, Huang Q, Xu XJ, Zhou HH, Liu ZQ. ABCC1 polymorphism Arg723Gln (2168G > A) is associated with lung cancer susceptibility in a Chinese population. *Clin Exp Pharmacol Physiol*. 2011;38(9):632–7.
38. Vulsteke C, Lambrechts D, Dieudonne A, Hatse S, Brouwers B, van Brussel T, et al. Genetic variability in the multidrug resistance associated protein-1 (ABCC1/MRP1) predicts hematological toxicity in breast cancer patients receiving (neo-)adjuvant chemotherapy with 5-fluorouracil, epirubicin and cyclophosphamide (FEC). *Ann Oncol*. 2013;24(6):1513–25.
39. Semsei AF, Erdelyi DJ, Ungvari I, Csagoly E, Hegyi MZ, Kiszal PS, et al. ABCC1 polymorphisms in anthracycline-induced cardiotoxicity in childhood acute lymphoblastic leukaemia. *Cell Biol Int*. 2012;36(1):79–86.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



SUPLEMENT DO ZAŁĄCZNIKA NR 1

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Bartosz, G., Dziadek, J. Strapagiel, D., 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

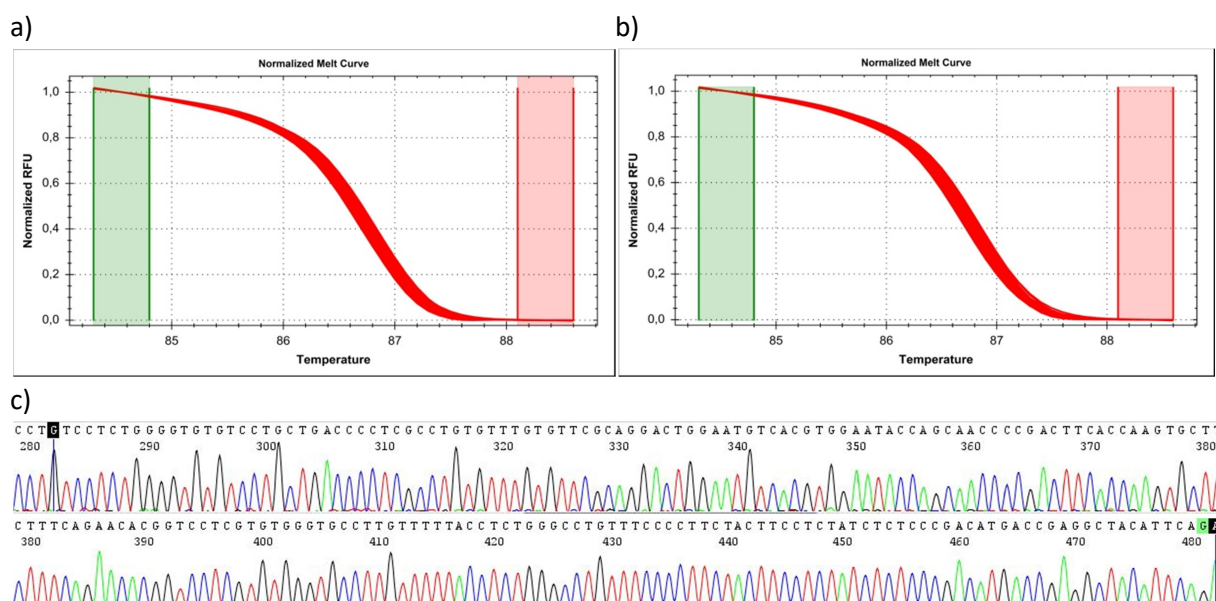


Figure S1 Summary of *ABCC1* exon 2 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16007766 – 16007966, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S1 Summary of *ABCC1* exon 2 (the first scanned area, NC_00016.10: 16007766 – 16007966) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

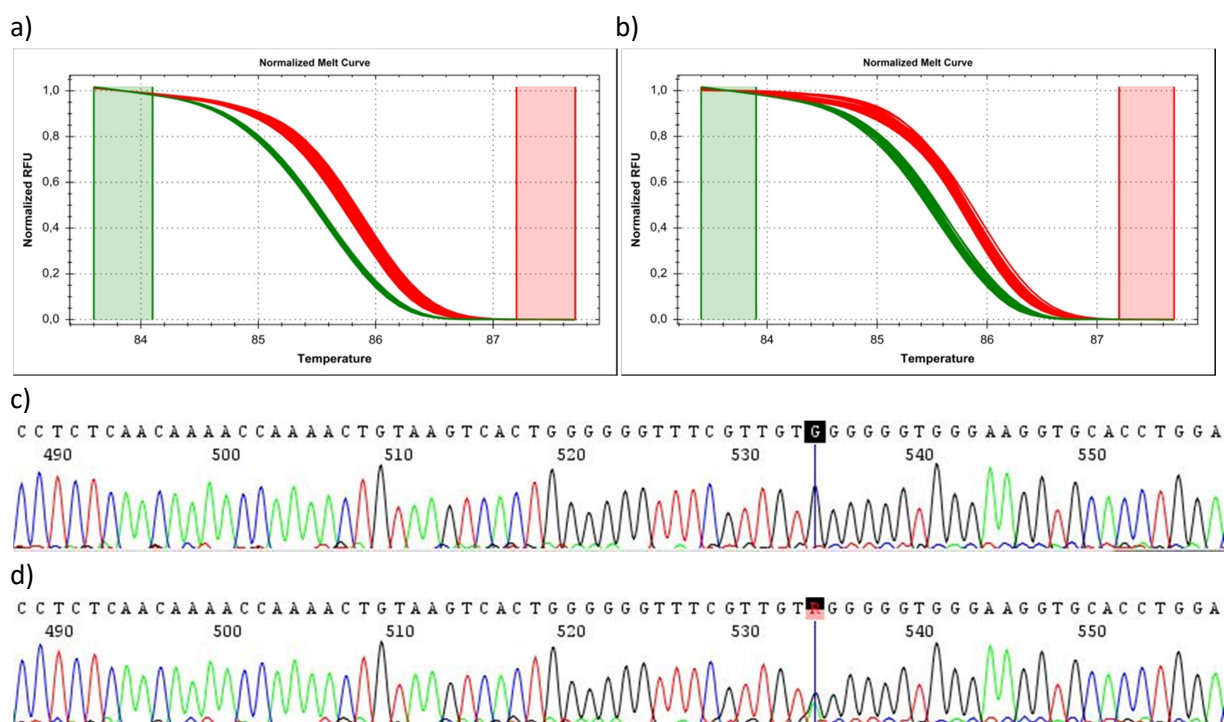


Figure S2 Summary of *ABCC1* exon 2 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16007862 – 16008064, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.225+26G>A) (polymorphic base in black boxes).

Table S2 Summary of *ABCC1* exon 2 (the second scanned area, NC_00016.10: 16007862 – 16008064) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	86	78	164
Samples in cluster 2 (green)	9	16	25
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	3	5
Scan rate	-	-	0,9947

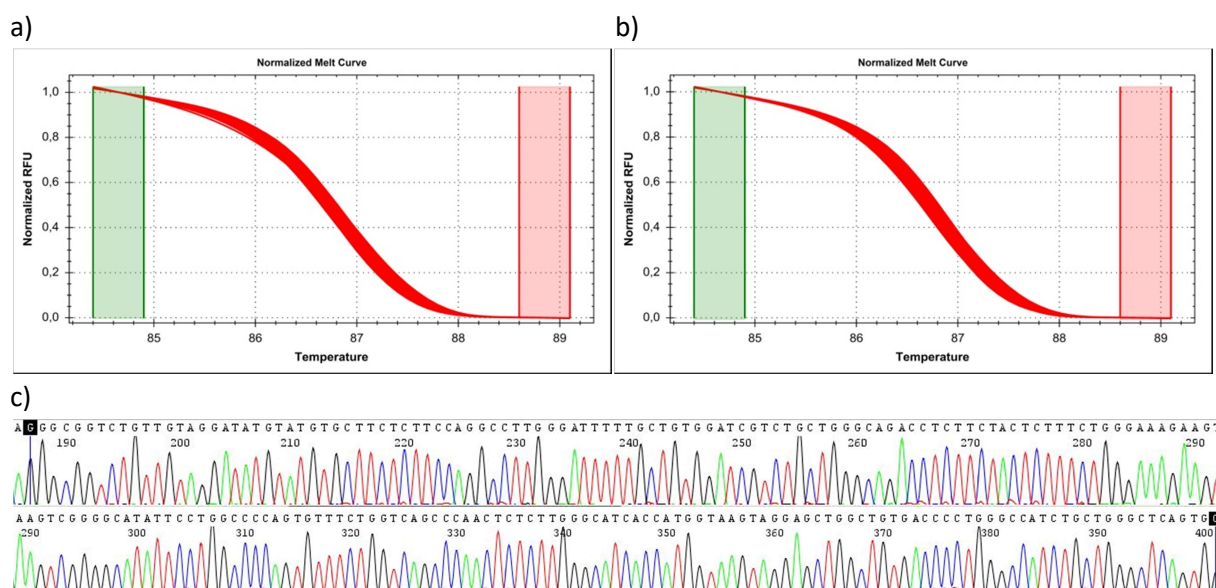


Figure S3 Summary of *ABCC1* exon 3 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16009736 – 16009950, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S3 Summary of *ABCC1* exon 3 (NC_00016.10: 16009736 – 16009950) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

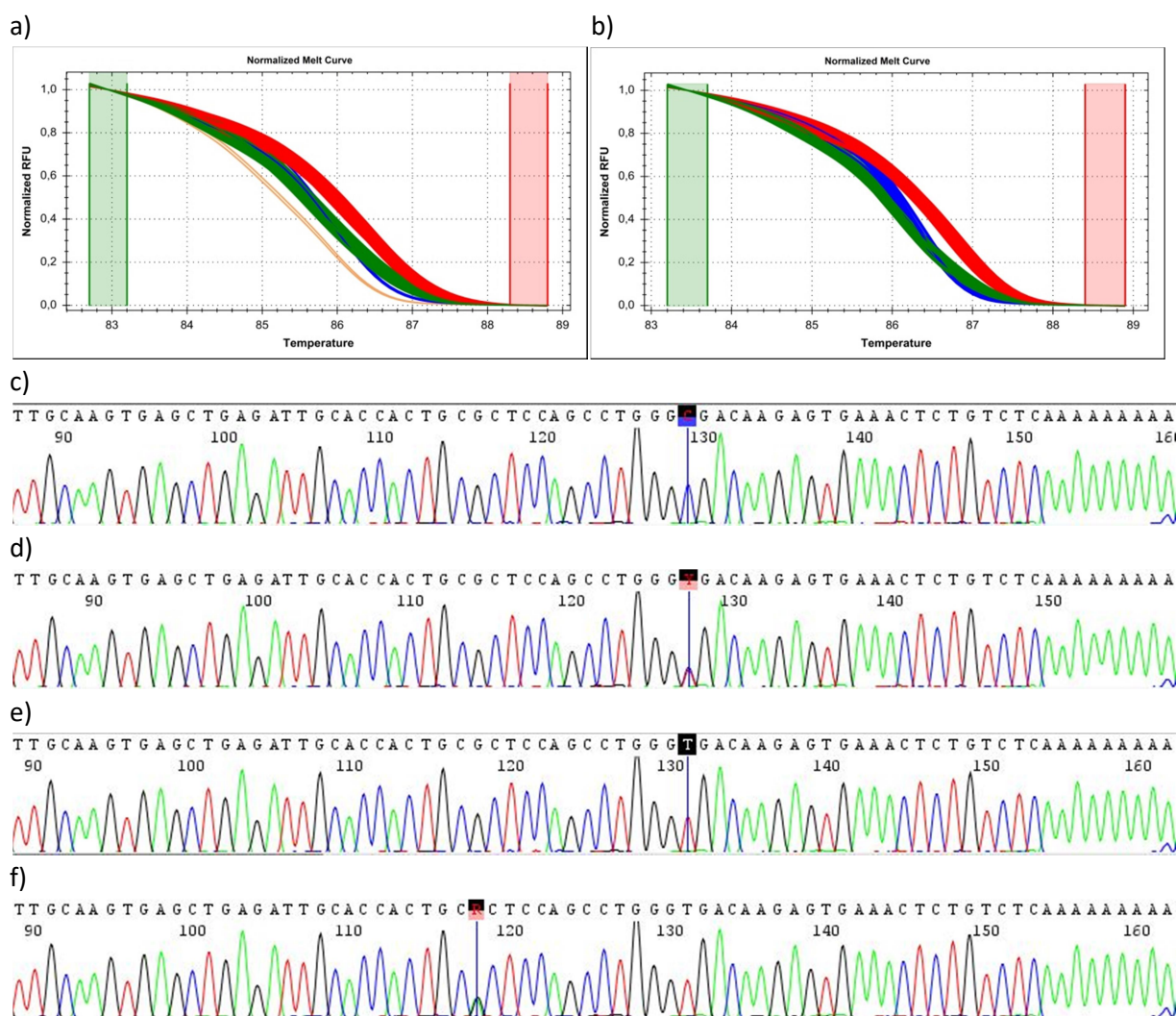


Figure S4 Summary of *ABCC1* exon 4 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16014342 – 16014504, including primers). Examples of sequencing traces for the samples: c) from the red clusters (homozygous variant c.352-66T>C), d) from the green clusters (heterozygous variant c.352-66T>C), e) from the blue clusters (no variants), f) from the orange cluster (heterozygous variant c.352-79G>A, the novel one) (polymorphic bases in black boxes).

Table S4 Summary of *ABCC1* exon 4 (the first scanned area, NC_00016.10: 16014342 – 16014504) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	4	3	7
Samples in cluster 1 (red)	46	45	91
Samples in cluster 2 (green)	42	38	80
Samples in cluster 3 (blue)	3	11	14
Samples in cluster 4 (orange)	1	-	1
Samples excluded	3	1	4
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	3	2	5
Samples verified by sequencing from cluster 4	1	-	1
Scan rate	-	-	0,9789

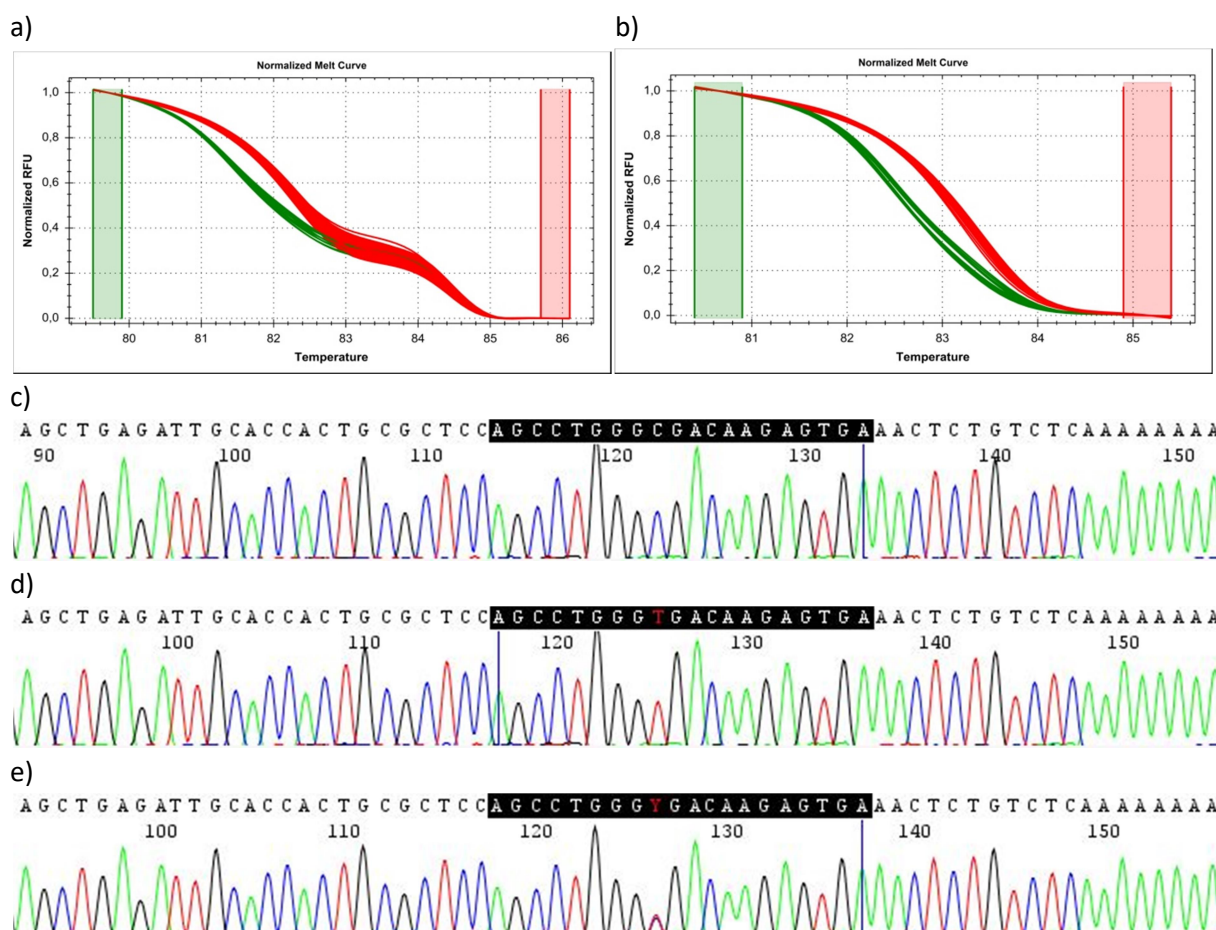
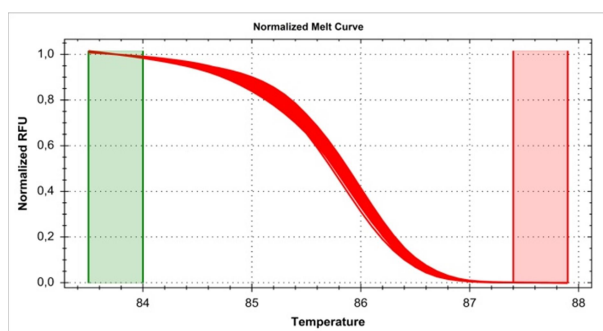


Figure S5 Summary of *ABCC1* exon 4 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16014417 – 16014605, including primers), plots clustering inefficient with regard to the variant inside the forward primer region. Examples of sequencing traces for the samples: c) from the red clusters (homozygous variant detected in the forward primer region), d) from the green clusters (no variants), e) from the green clusters (heterozygous variant detected in the forward primer region) (sequence highlighted in black indicates the primer region, variants inside) - additional scanning area flanking this region was created (see above, Figure S4 and Table S4), no additional variants differentiating clusters were detected for this area.

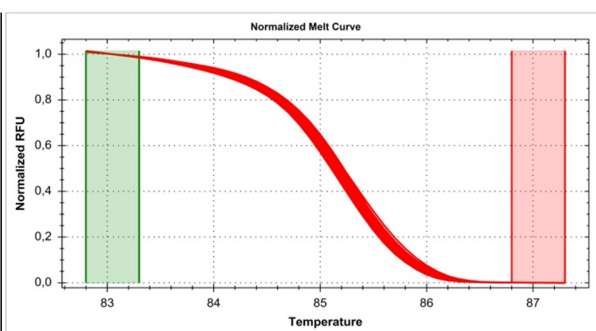
Table S5 Summary of *ABCC1* exon 4 (the second scanned area, NC_00016.10: 16014417 – 16014605) scanning by HRM. Scan rate is the ratio between positive clustered samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	88	88	176
Samples in cluster 2 (green)	6	7	13
Samples excluded	1	0	1
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	5	7
Scan rate	-	-	0,9947

a)



b)



c)

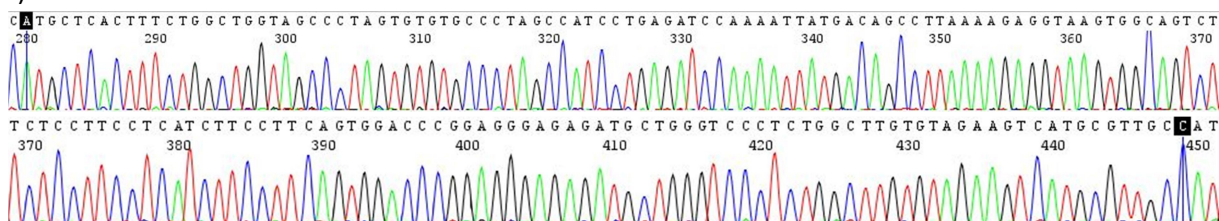


Figure S6 Summary of *ABCC1* exon 4 scanning by HRM: a), b) melting plots for the third scanned area of the exon (NC_00016.10: 16014551 – 16014720, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S6 Summary of *ABCC1* exon 4 (the third scanned area, NC_00016.10: 16014551 – 16014720) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

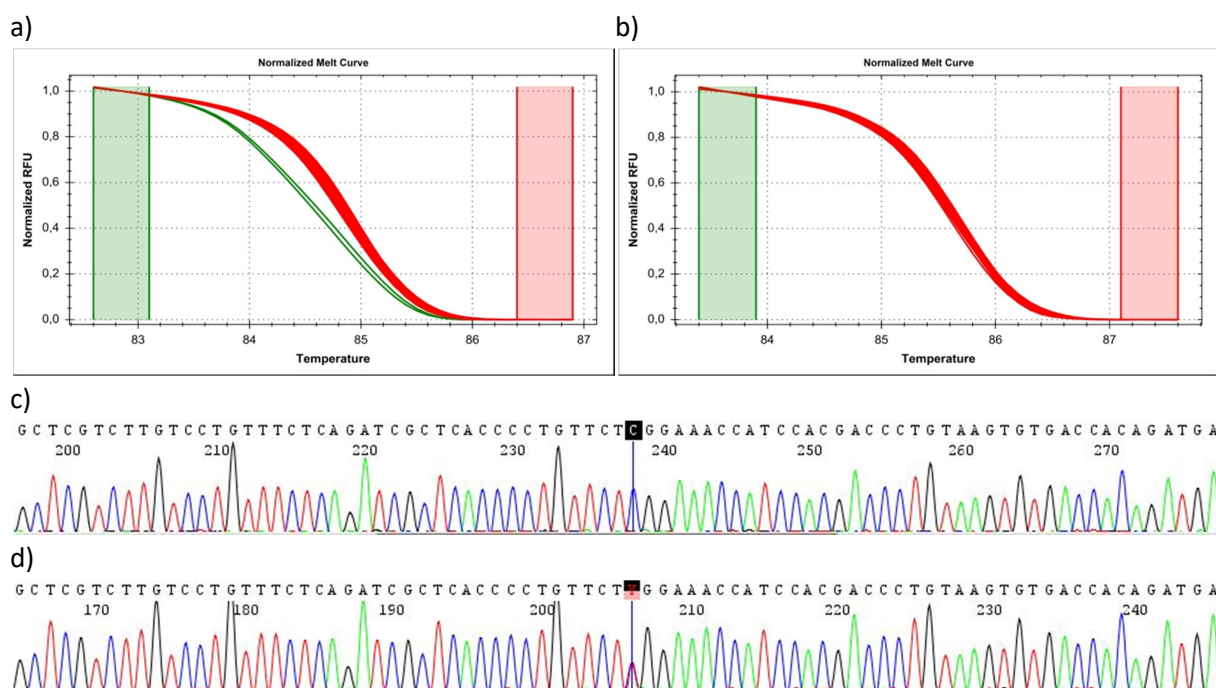


Figure S7 Summary of *ABCC1* exon 5 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16016449 – 16016660, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.596C>T, the novel one) (polymorphic base in black boxes).

Table S7 Summary of *ABCC1* exon 5 (NC_00016.10: 16016449 – 16016660) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	93	93	186
Samples in cluster 2 (green)	1	-	1
Samples excluded	1	2	3
Samples verified by sequencing from cluster 1	2	3	5
Samples verified by sequencing from cluster 2	1	-	1
Scan rate	-	-	0,9842

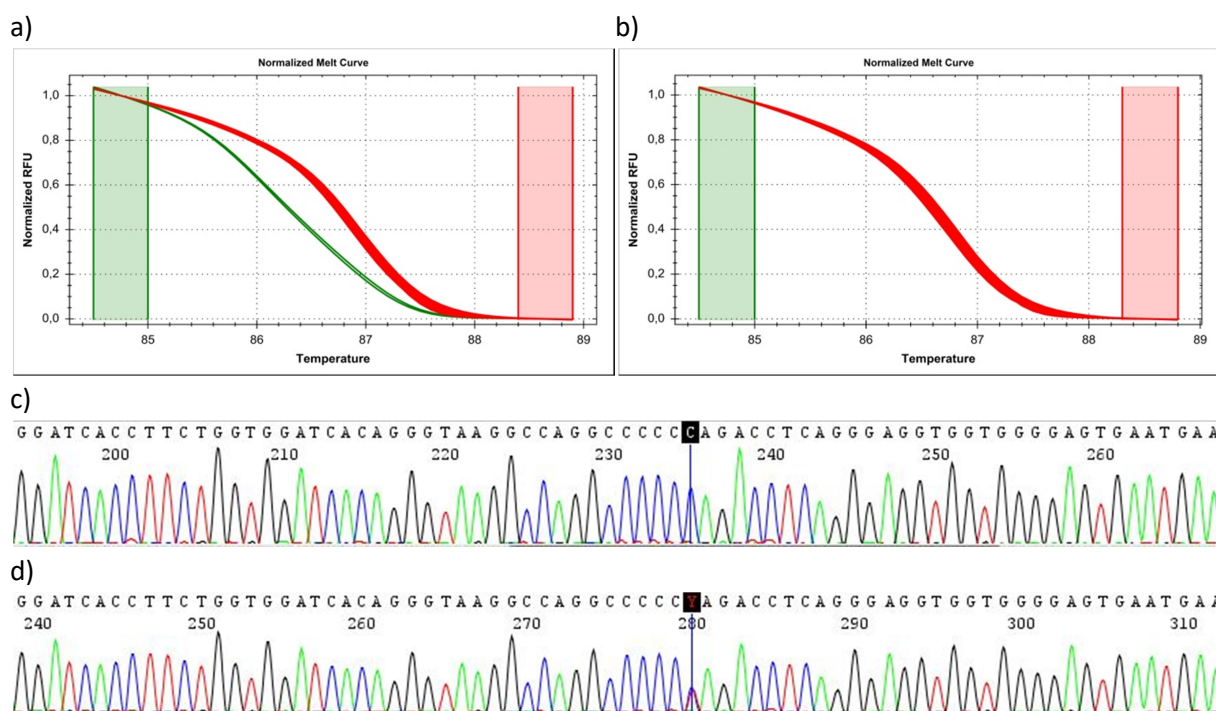


Figure S8 Summary of *ABCC1* exon 6 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16033069 – 16033220, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.677+17C>T) (polymorphic base in black boxes).

Table S8 Summary of *ABCC1* exon 6 (NC_00016.10: 16033069 – 16033220) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	93	94	187
Samples in cluster 2 (green)	1	-	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	3	5
Samples verified by sequencing from cluster 2	1	-	1
Scan rate	-	-	0,9895

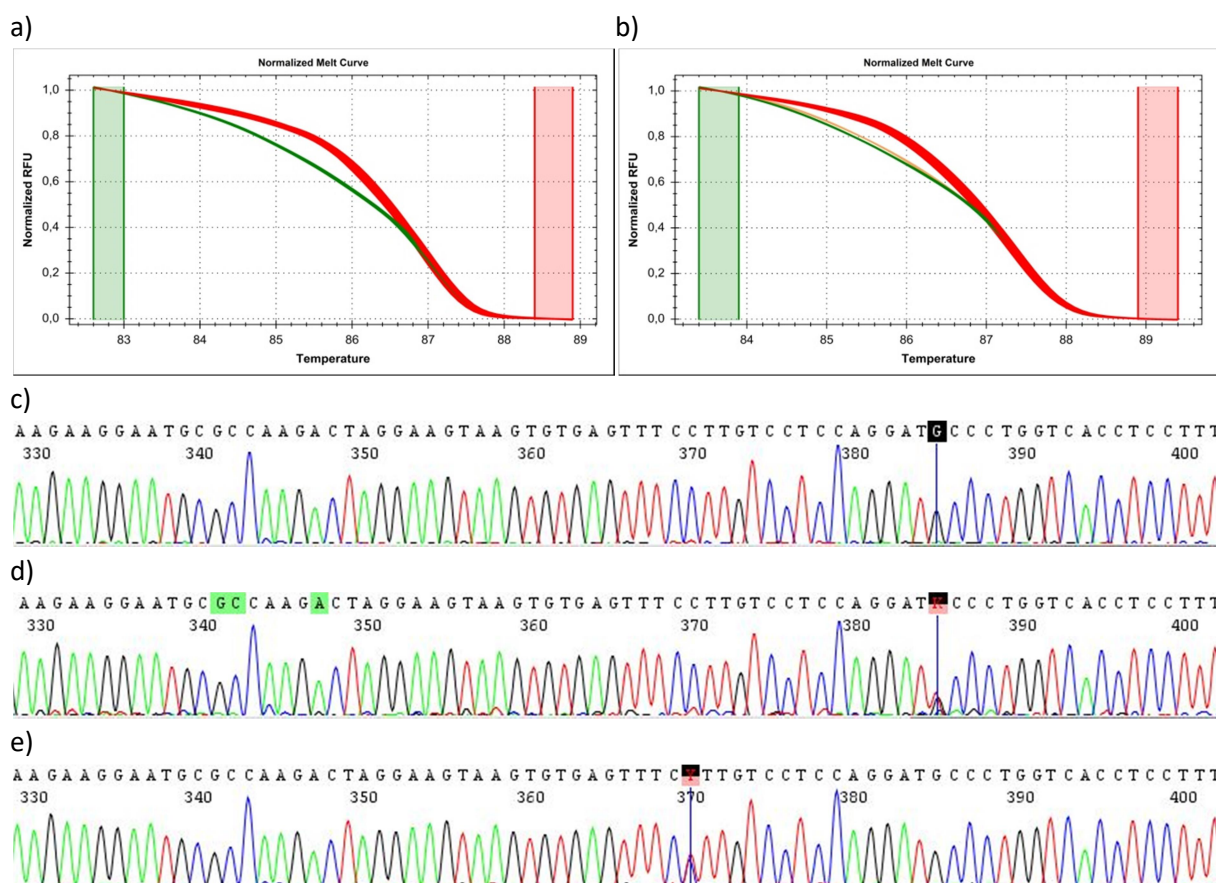


Figure S9 Summary of *ABCC1* exon 7 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16036418 – 16036657, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.809+31G>T), e) from the orange cluster (heterozygous variant c.809+16C>T, the novel one) (polymorphic bases in black boxes).

Table S9 Summary of *ABCC1* exon 7 (the first scanned area, NC_00016.10: 16036418 – 16036657) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	3	5
Samples in cluster 1 (red)	92	91	183
Samples in cluster 2 (green)	3	2	5
Samples in cluster 3 (orange)	-	1	1
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	2	5
Samples verified by sequencing from cluster 2	3	2	5
Samples verified by sequencing from cluster 3	-	1	1
Scan rate	-	-	0,9947

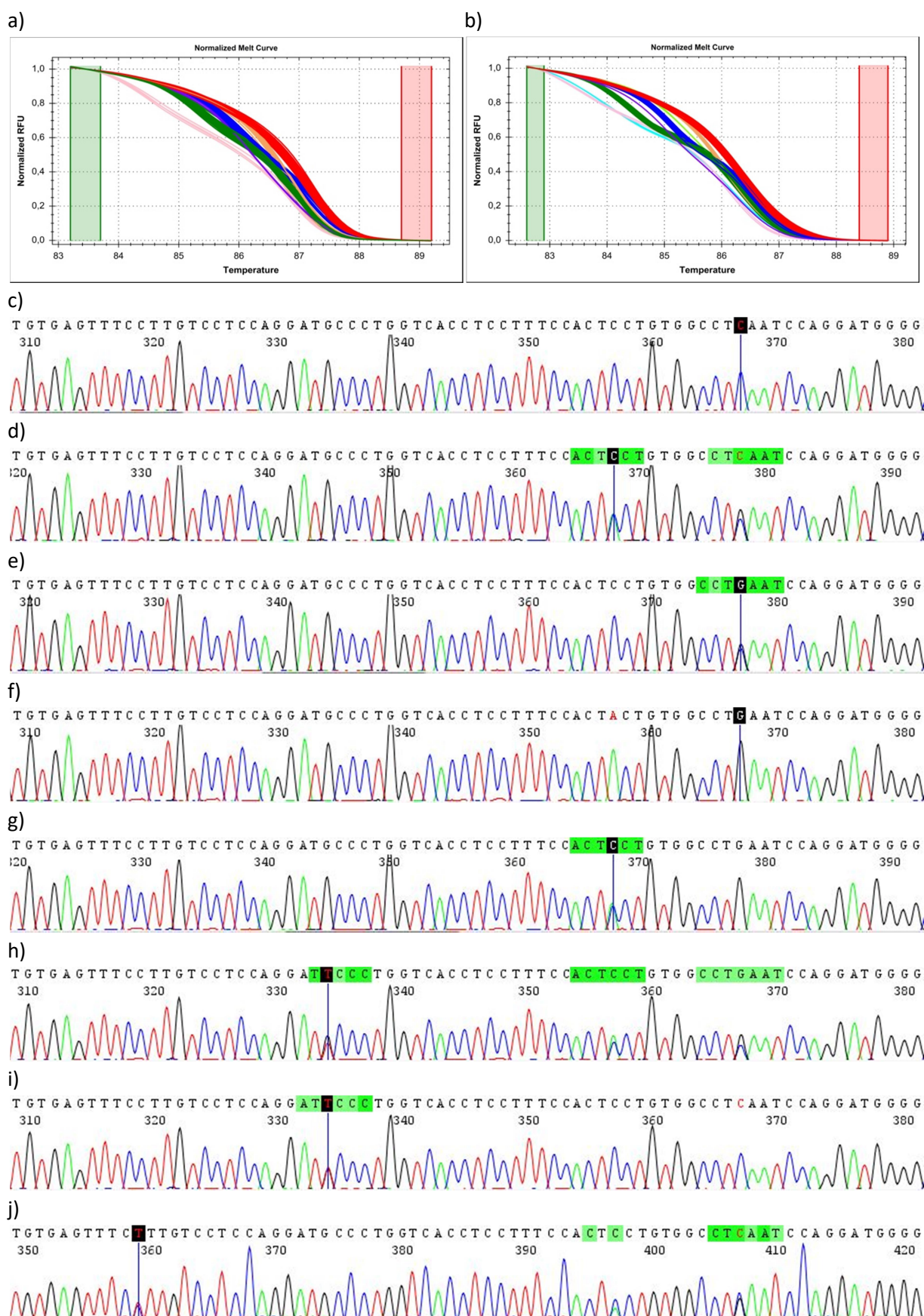


Figure S10 Summary of *ABCC1* exon 7 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16036544 – 16036727, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters

(heterozygous variants c.809+54C>A (in black box) and c.809+64C>G), e) from the blue clusters (heterozygous variant c.809+64C>G (in black box)), f) from the orange clusters (homozygous variants c.809+54C>A and c.809+64C>G (in black box)), g) from the lime cluster (heterozygous variant c.809+54C>A (in black box) and homozygous variant c.809+64C>G), h) from the pink clusters (heterozygous variants c.809+31G>T (in black box) and c.809+54C>A and c.809+64C>G), i) from the violet clusters (heterozygous variant c.809+31G>T (in black box)), j) from the cyan cluster (heterozygous variants c.809+16C>T (in black box, the novel one) and c.809+54C>A and c.809+64C>G).

Table S10 Summary of *ABCC1* exon 7 (the second scanned area, NC_00016.10: 16036544 – 16036727) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	6	8	14
Samples in cluster 1 (red)	46	36	82
Samples in cluster 2 (green)	24	34	58
Samples in cluster 3 (blue)	14	14	28
Samples in cluster 4 (orange)	7	4	11
Samples in cluster 5 (lime)	-	3	3
Samples in cluster 6 (pink)	2	1	3
Samples in cluster 7 (violet)	1	1	2
Samples in cluster 8 (cyan)	-	1	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	2	2	4
Samples verified by sequencing from cluster 5	-	3	3
Samples verified by sequencing from cluster 6	2	1	3
Samples verified by sequencing from cluster 7	1	1	2
Samples verified by sequencing from cluster 8	-	1	1
Scan rate	-	-	0,9895

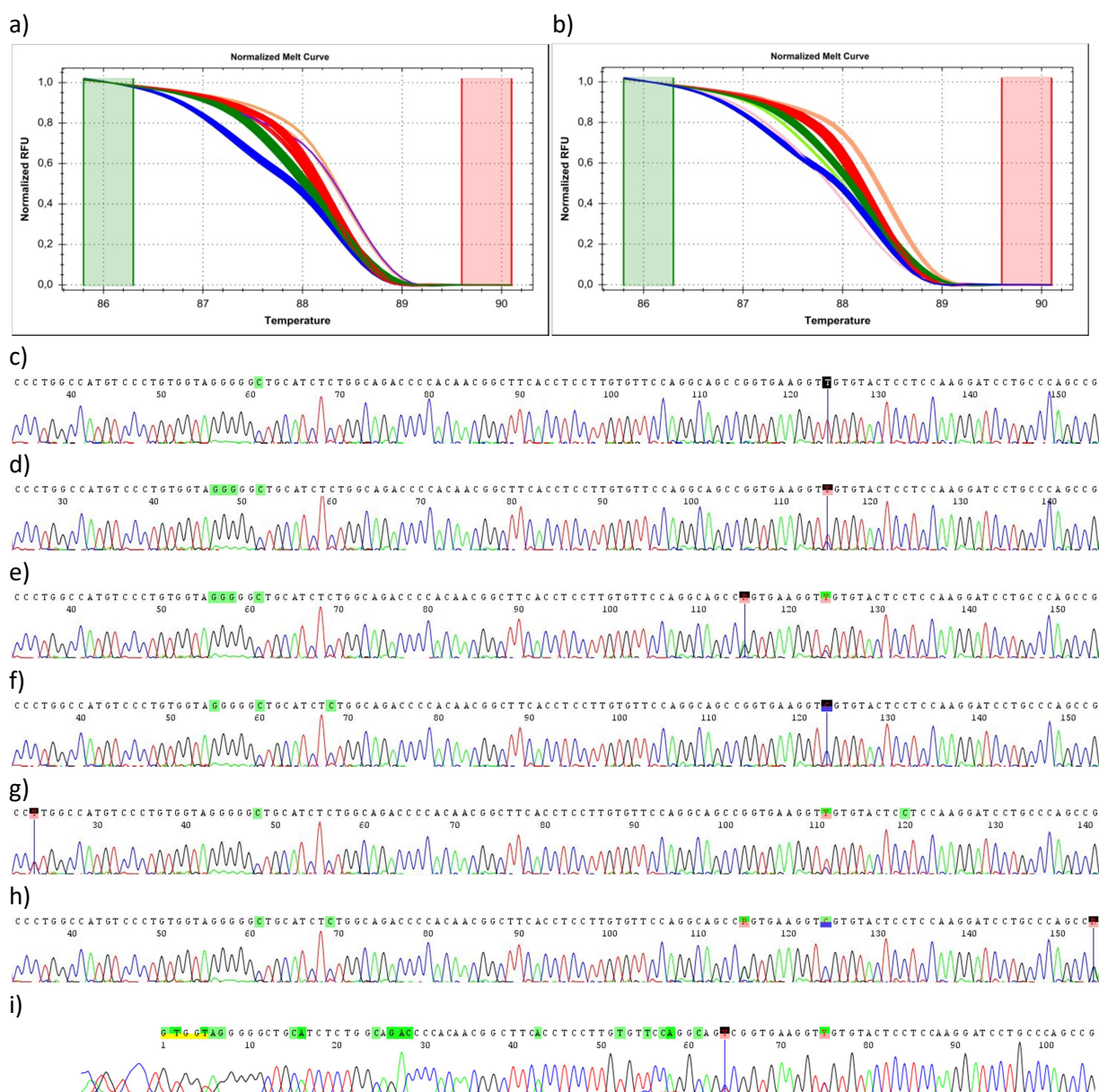


Figure S11 Summary of *ABCC1* exon 8 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16044356 – 16044574, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.825T>C (in black box)), e) from the blue clusters (heterozygous variants c.816G>A (in black box) and c.825T>C), f) from the orange clusters (homozygous variant c.825T>C (in black box)), g) from the lime cluster (heterozygous variants c.810-73C>T (in black box) and c.825T>C), h) from the pink cluster (heterozygous variant c.816G>A and homozygous variant c.825T>C and heterozygous variant c.855G>A (the novel one, in black box)), i) from the violet cluster (heterozygous variants c.814C>T (in black box) and c.825T>C).

Table S11 Summary of *ABCC1* exon 8 (the first scanned area, NC_00016.10: 16044356 – 16044574) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	5	6	11
Samples in cluster 1 (red)	44	40	84
Samples in cluster 2 (green)	40	35	75
Samples in cluster 3 (blue)	7	8	15
Samples in cluster 4 (orange)	2	9	11
Samples in cluster 5 (lime)	-	1	1
Samples in cluster 6 (pink)	-	1	1
Samples in cluster 7 (violet)	1	-	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	2	2	4
Samples verified by sequencing from cluster 5	-	1	1
Samples verified by sequencing from cluster 6	-	1	1
Samples verified by sequencing from cluster 7	1	-	1
Scan rate	-	-	0,9895

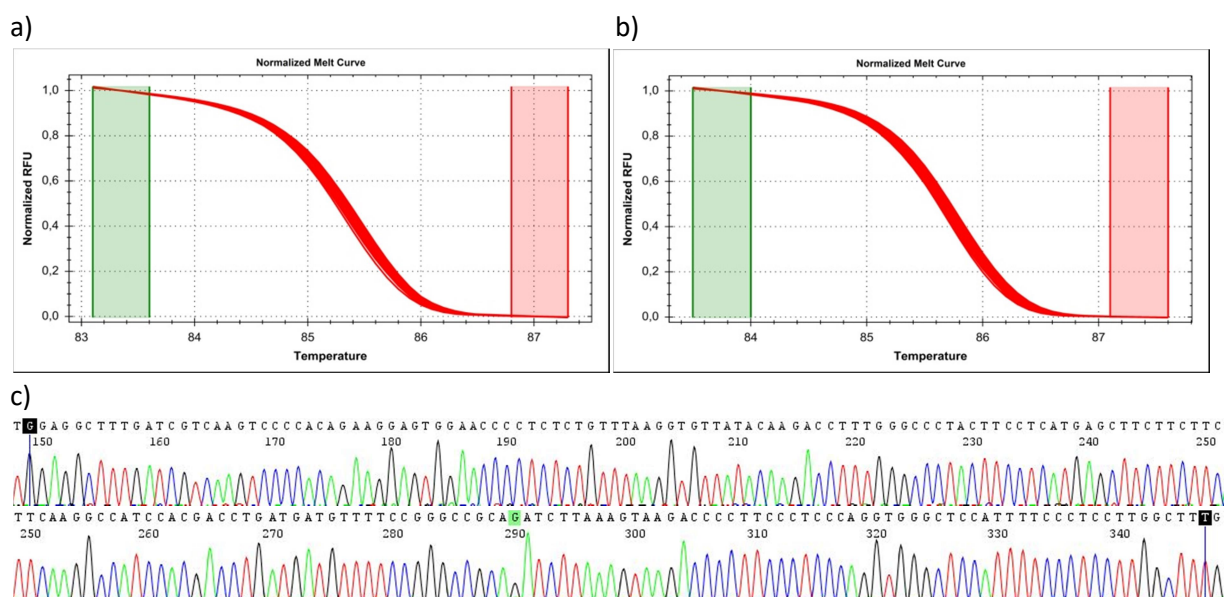


Figure S12 Summary of *ABCC1* exon 8 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16044531 – 16044729, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S12 Summary of *ABCC1* exon 8 (the second scanned area, NC_00016.10: 16044531 – 16044729) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

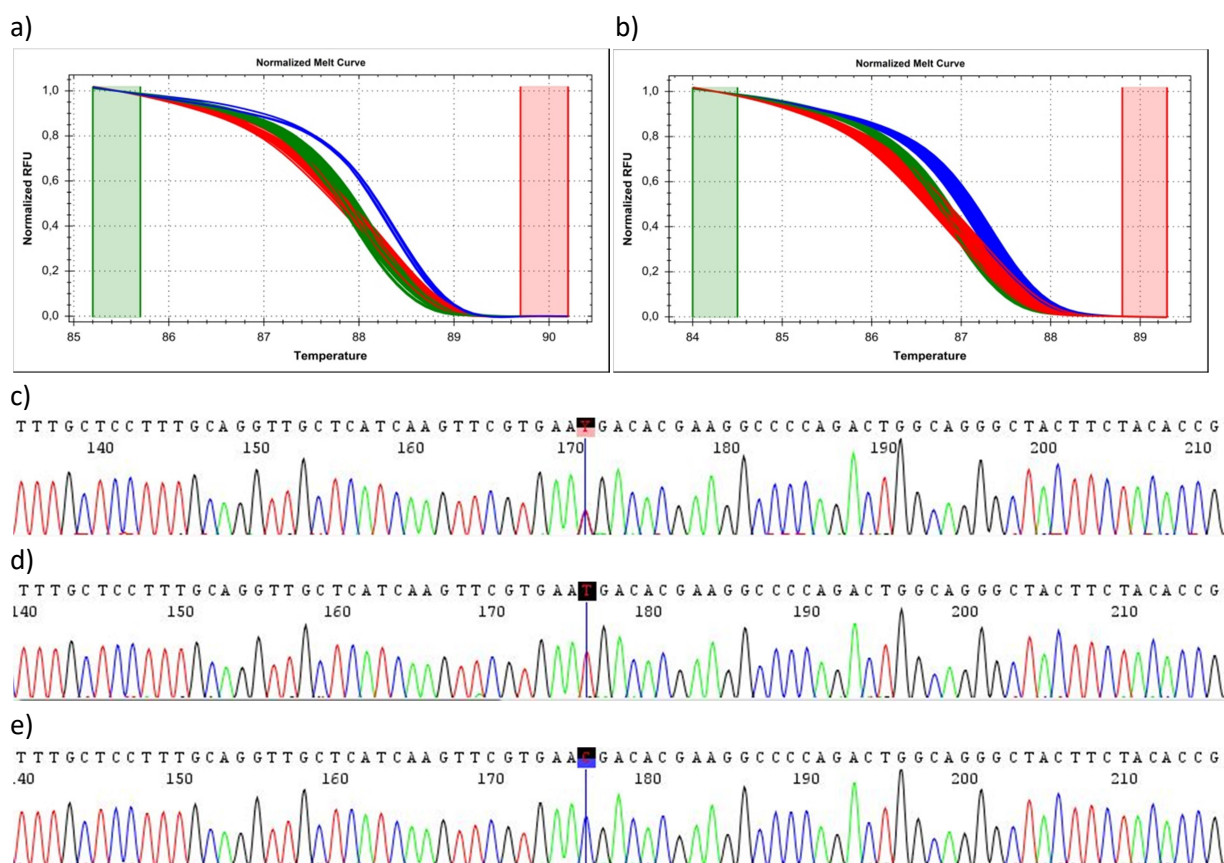


Figure S13 Summary of *ABCC1* exon 9 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16045802 – 16045966, including primers). Examples of sequencing traces for the samples: c) from the red clusters (heterozygous variant c.1062T>C), d) from the green clusters (no variants), e) from the blue clusters (homozygous variant c.1062T>C) (polymorphic base in black boxes).

Table S13 Summary of *ABCC1* exon 9 (the first scanned area, NC_00016.10: 16045802 – 16045966) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	3	3	6
Samples in cluster 1 (red)	47	44	91
Samples in cluster 2 (green)	44	34	78
Samples in cluster 3 (blue)	2	14	16
Samples excluded	2	3	5
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Scan rate	-	-	0,9737

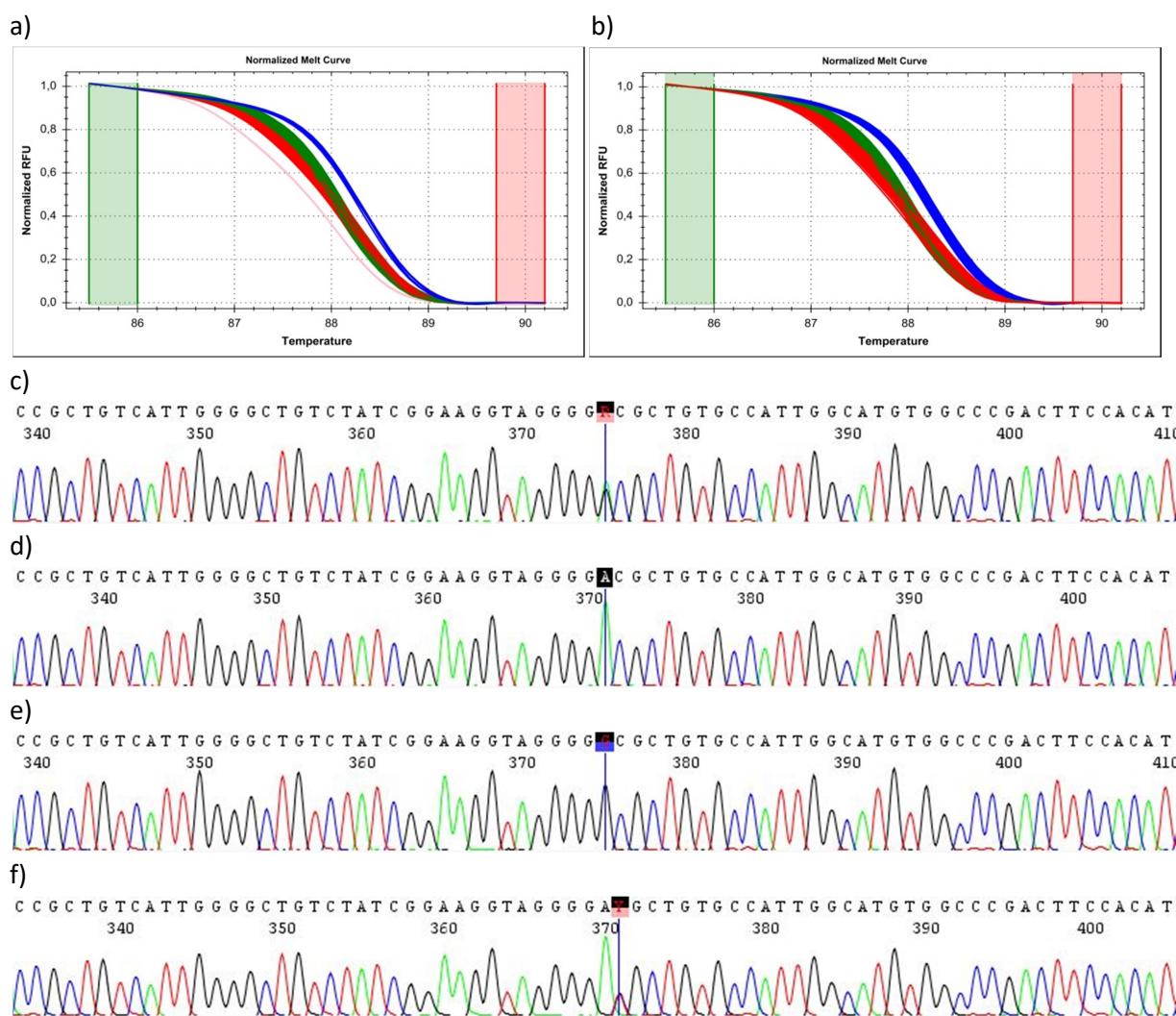


Figure S14 Summary of *ABCC1* exon 9 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16045891 – 16046065, including primers). Examples of sequencing traces for the samples: c) from the red clusters (heterozygous variant c.1218+8A>G (in black box)), d) from the green clusters (no variants), e) from the blue clusters (homozygous variant c.1218+8A>G (in black box), f) from the pink cluster (heterozygous variant c.1218+9C>T, the novel one (in black box)).

Table S14 Summary of *ABCC1* exon 9 (the second scanned area, NC_00016.10: 16045891 – 16046065) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	4	3	7
Samples in cluster 1 (red)	47	44	91
Samples in cluster 2 (green)	43	38	81
Samples in cluster 3 (blue)	4	12	16
Samples in cluster 4 (pink)	1	-	1
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	1	2	3
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	1	-	1
Scan rate	-	-	0,9947

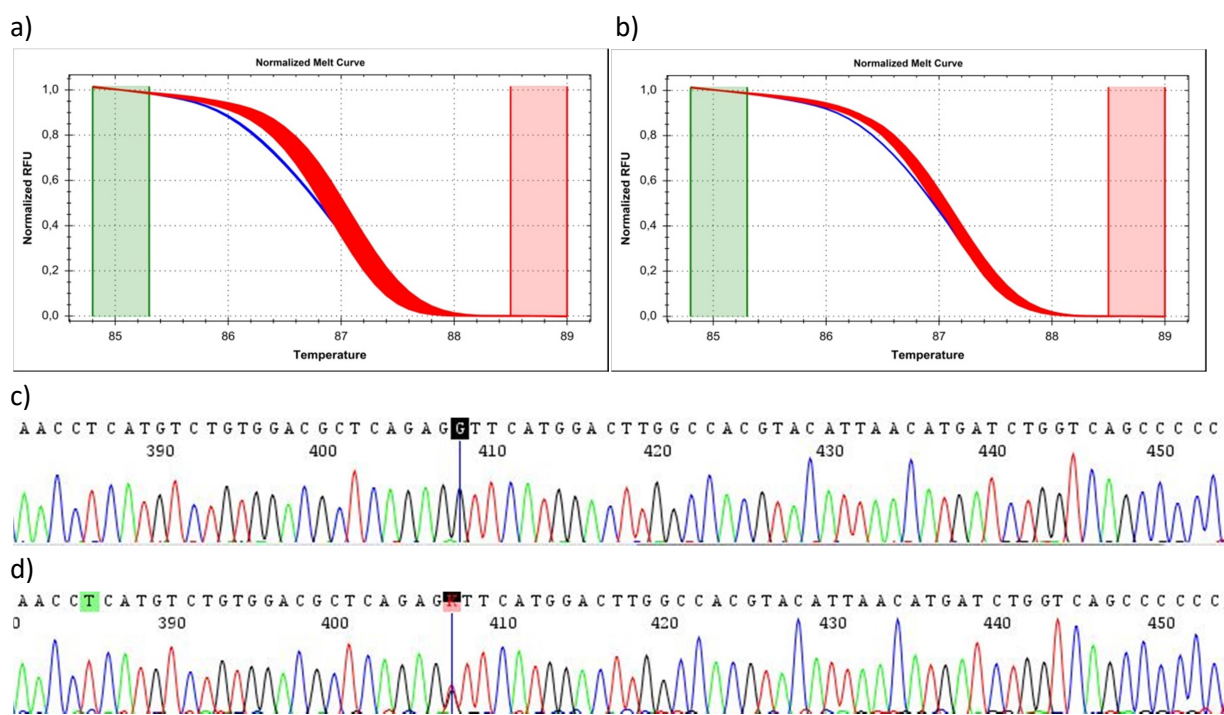


Figure S15 Summary of *ABCC1* exon 10 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16048095 – 16048301, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the blue clusters (heterozygous variant c.1299G>T) (polymorphic base in black boxes).

Table S15 Summary of *ABCC1* exon 10 (the first scanned area, NC_00016.10: 16048095 – 16048301) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	93	93	186
Samples in cluster 2 (blue)	1	1	2
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	4	6
Samples verified by sequencing from cluster 2	1	1	2
Scan rate	-	-	0,9895

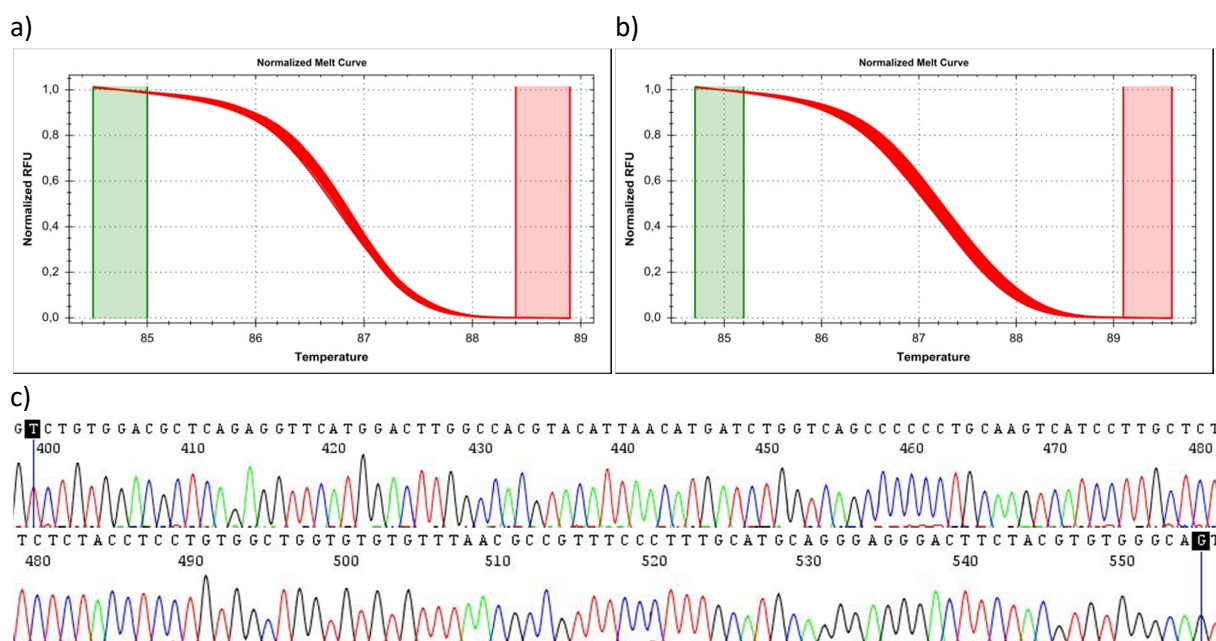


Figure S16 Summary of *ABCC1* exon 10 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16048205 – 16048361, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S16 Summary of *ABCC1* exon 10 (the second scanned area, NC_00016.10: 16048205 – 16048361) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	93	187
Samples excluded	1	2	3
Samples verified by sequencing from cluster 1	2	3	5
Scan rate	-	-	0,9842

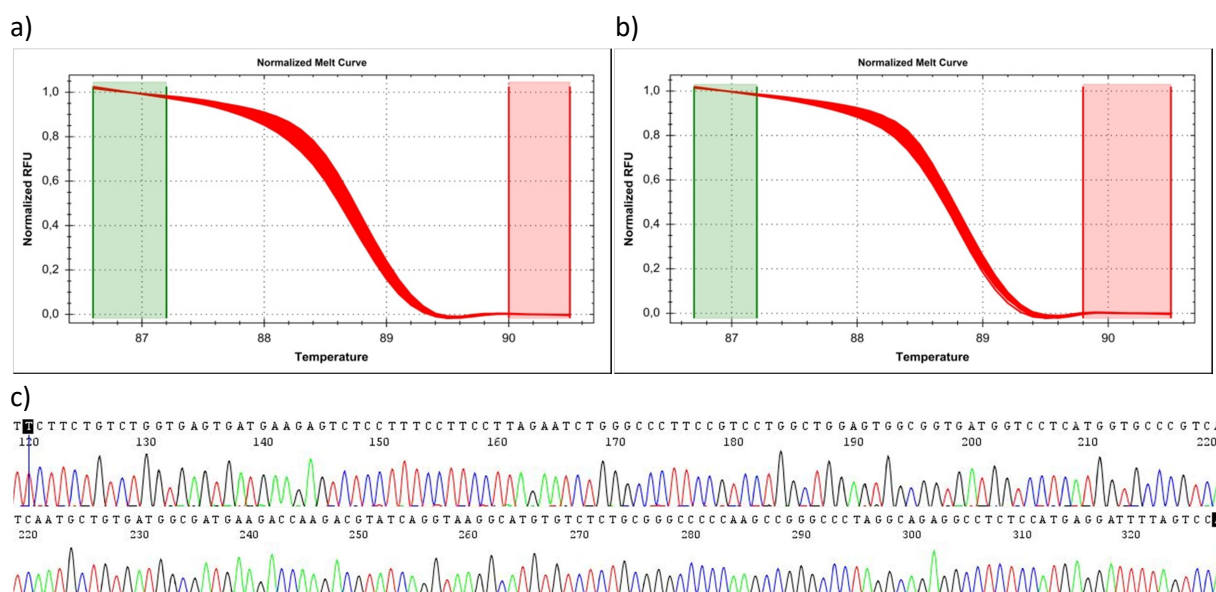


Figure S17 Summary of *ABCC1* exon 11 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16052680 – 16052888, including primers). An example of the sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S17 Summary of *ABCC1* exon 11 (NC_00016.10: 16052680 – 16052888) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	95	188
Samples excluded	2	0	2
Samples verified by sequencing from cluster 1	3	4	7
Scan rate	-	-	0,9895

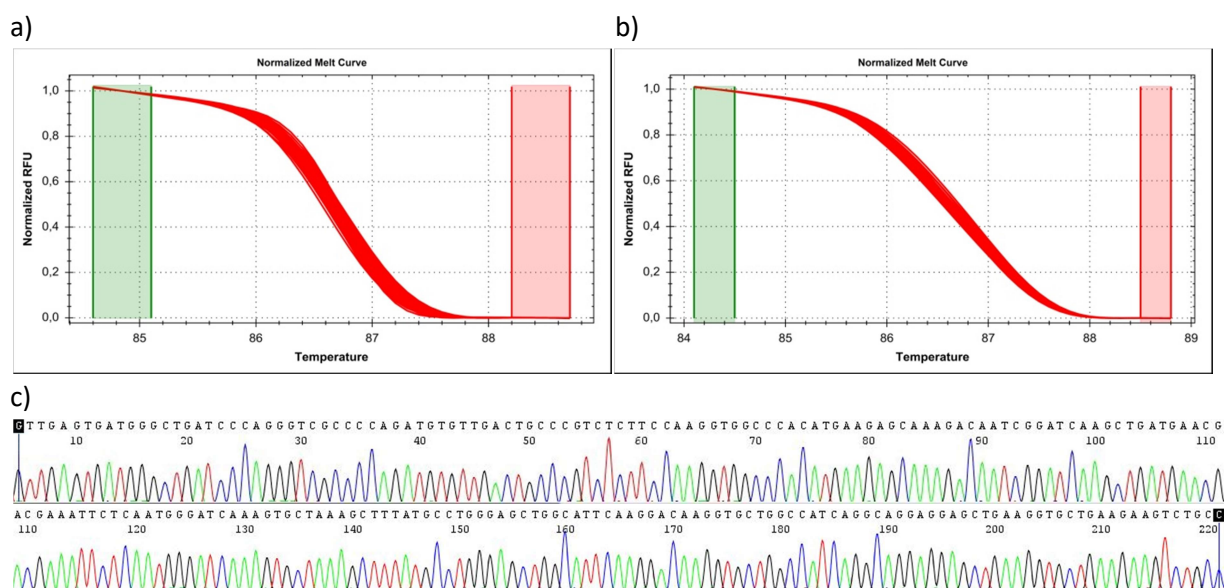


Figure S18 Summary of *ABCC1* exon 12 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16056031 – 16056247, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S18 Summary of *ABCC1* exon 12 (the first scanned area, NC_00016.10: 16056031 – 16056247) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

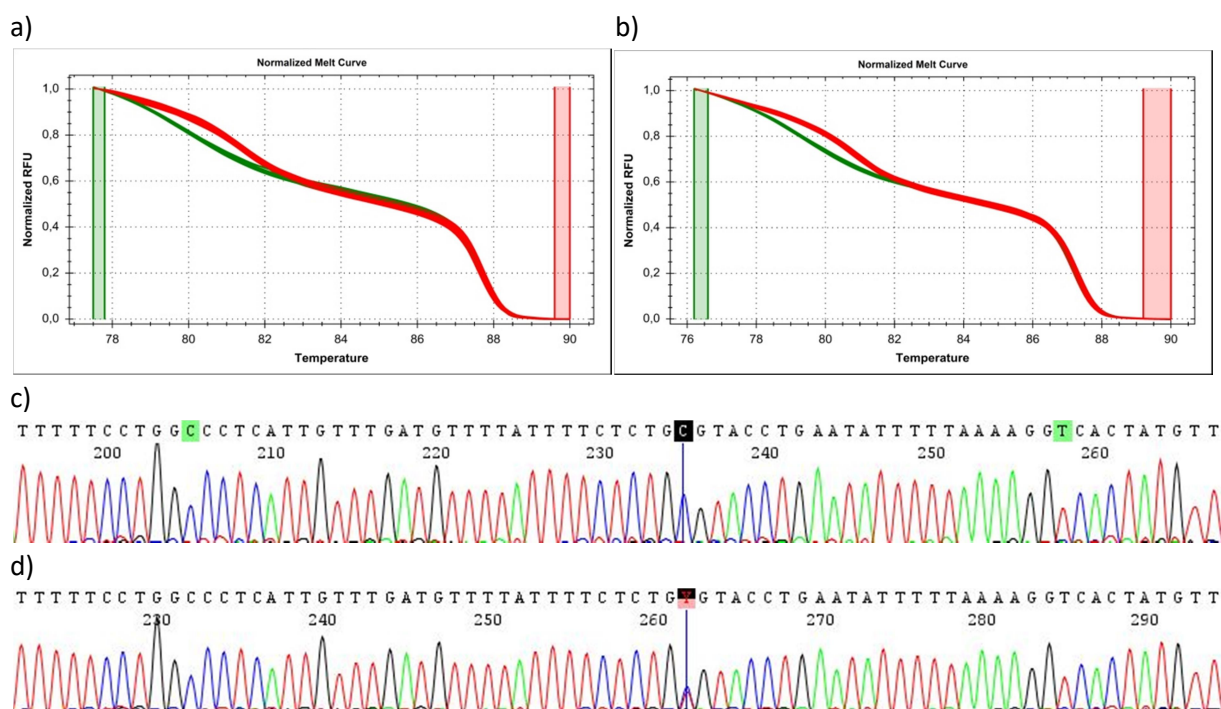


Figure S19 Summary of *ABCC1* exon 12 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16056185 – 16056390, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.1677+56C>T) (polymorphic base in black boxes).

Table S19 Summary of *ABCC1* exon 12 (the second scanned area, NC_00016.10: 16056185 – 16056390) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	82	80	162
Samples in cluster 2 (green)	13	14	27
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Scan rate	-	-	0,9947

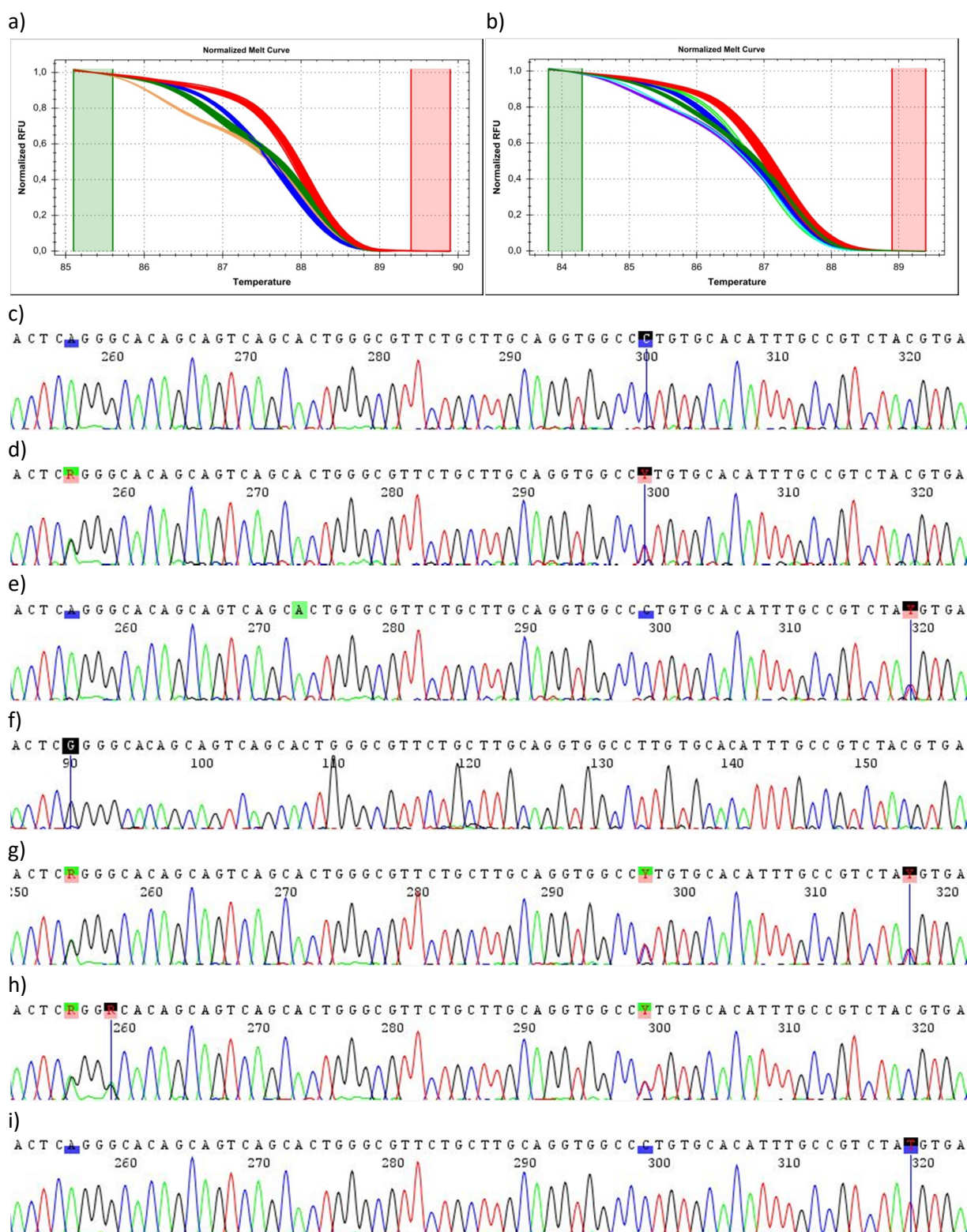


Figure S20 Summary of *ABCC1* exon 13 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16068075 – 16068258, including primers). Examples of sequencing traces for the samples: c) from the red clusters (homozygous variants c.1678-37G>A and c.1684T>C (in black box)), d) from the green clusters (heterozygous variants c.1678-37G>A and c.1684T>C (in black box)), e) from the blue clusters (homozygous variants c.1678-37G>A and c.1684T>C and heterozygous variant c.1704C>T (in black box)), f) from the orange cluster (no variants), g) from the

violet cluster (heterozygous variants c.1678-37G>A and c.1684T>C and c.1704C>T (in black box)), h) from the cyan cluster (heterozygous variants c.1678-37G>A and 1678-34G>A (the novel one, in black box) and c.1684T>C), i) from the lime cluster (homozygous variants c.1678-37G>A and c.1684T>C and 1704C>T (in black box)).

Table S20 Summary of *ABCC1* exon 13 (the first scanned area, NC_00016.10: 16068075 – 16068258) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	4	6	10
Samples in cluster 1 (red)	57	55	112
Samples in cluster 2 (green)	22	20	42
Samples in cluster 3 (blue)	13	16	29
Samples in cluster 4 (orange)	2	-	2
Samples in cluster 5 (violet)	-	2	2
Samples in cluster 6 (cyan)	-	1	1
Samples in cluster 7 (lime)	-	1	1
Samples excluded	1	0	1
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	2	-	2
Samples verified by sequencing from cluster 5	-	2	2
Samples verified by sequencing from cluster 6	-	1	1
Samples verified by sequencing from cluster 7	-	1	1
Scan rate	-	-	0,9947

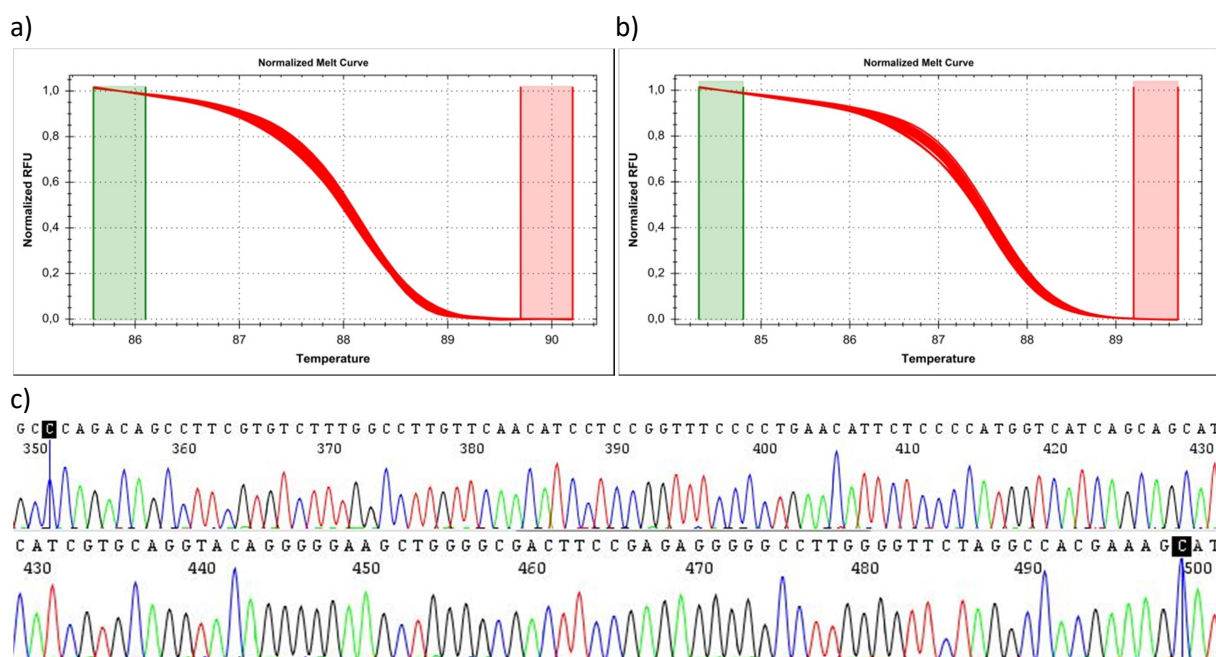


Figure S21 Summary of *ABCC1* exon 13 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16068215 – 16068363, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S21 Summary of *ABCC1* exon 13 (the second scanned area, NC_00016.10: 16068215 – 16068363) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	92	185
Samples excluded	2	3	5
Samples verified by sequencing from cluster 1	4	3	7
Scan rate	-	-	0,9737

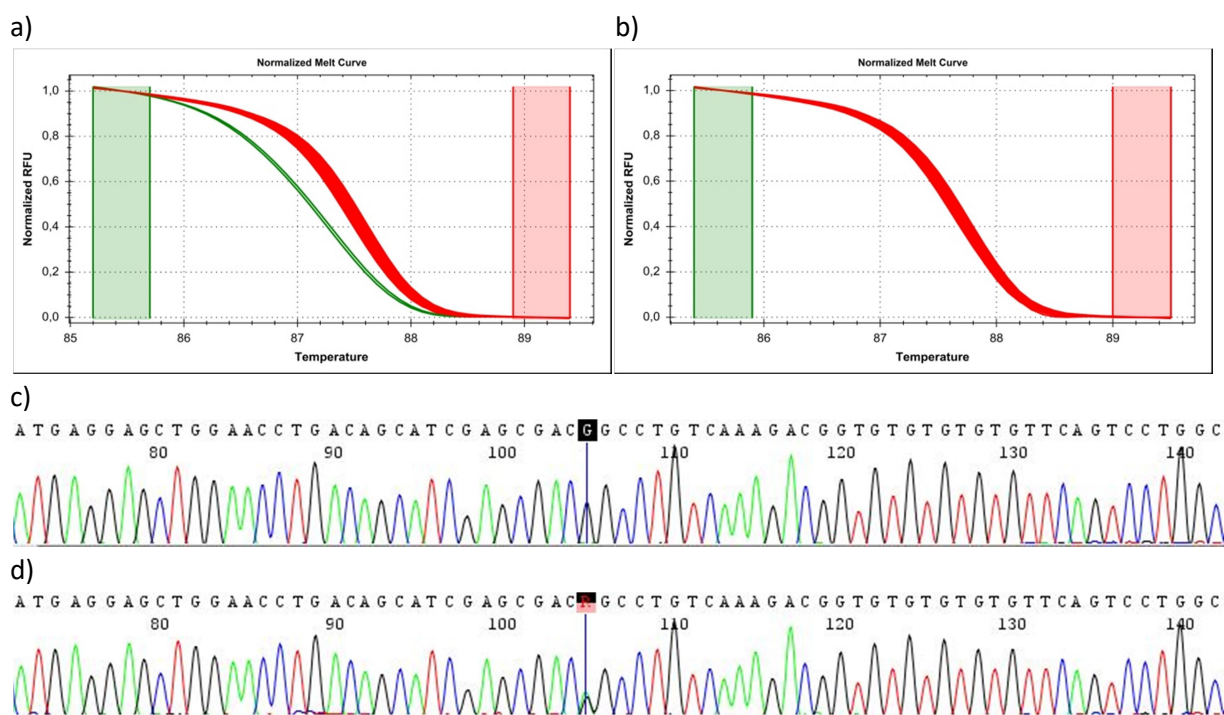


Figure S22 Summary of *ABCC1* exon 14 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16071573 – 16071800, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.1898G>A) (polymorphic base in black boxes).

Table S22 Summary of *ABCC1* exon 14 (NC_00016.10: 16071573 – 16071800) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	93	94	187
Samples in cluster 2 (green)	1	-	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	3	5
Samples verified by sequencing from cluster 2	1	-	1
Scan rate	-	-	0,9895

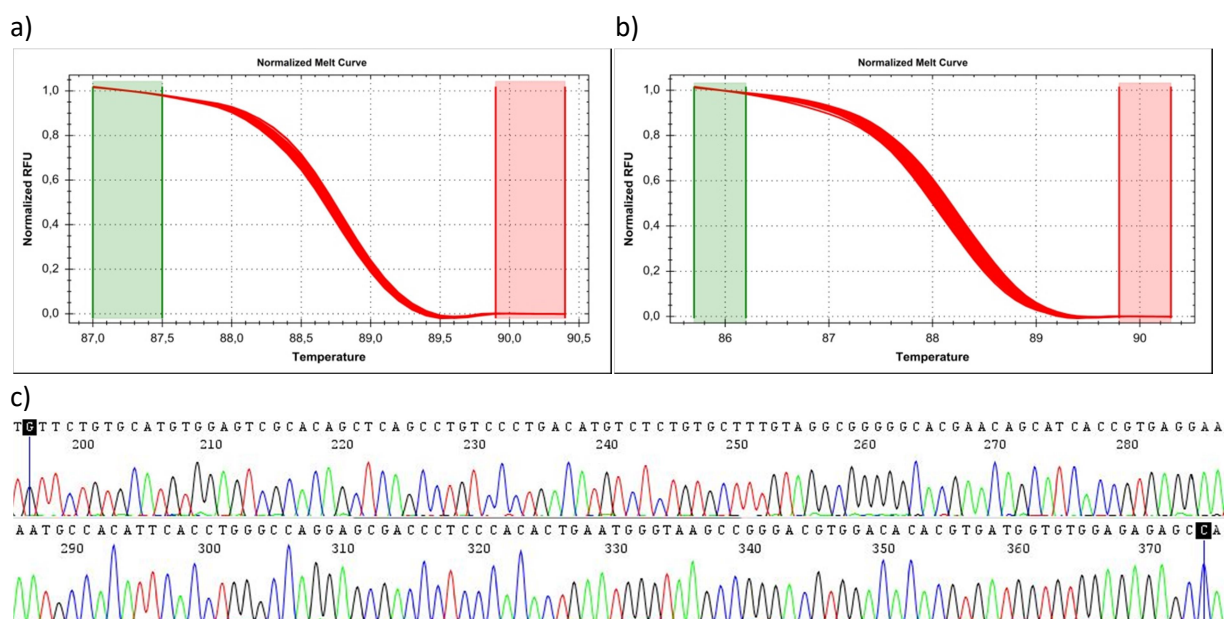


Figure S23 Summary of *ABCC1* exon 15 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16076265 – 16076443, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S23 Summary of *ABCC1* exon 15 (NC_00016.10: 16076265 – 16076443) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

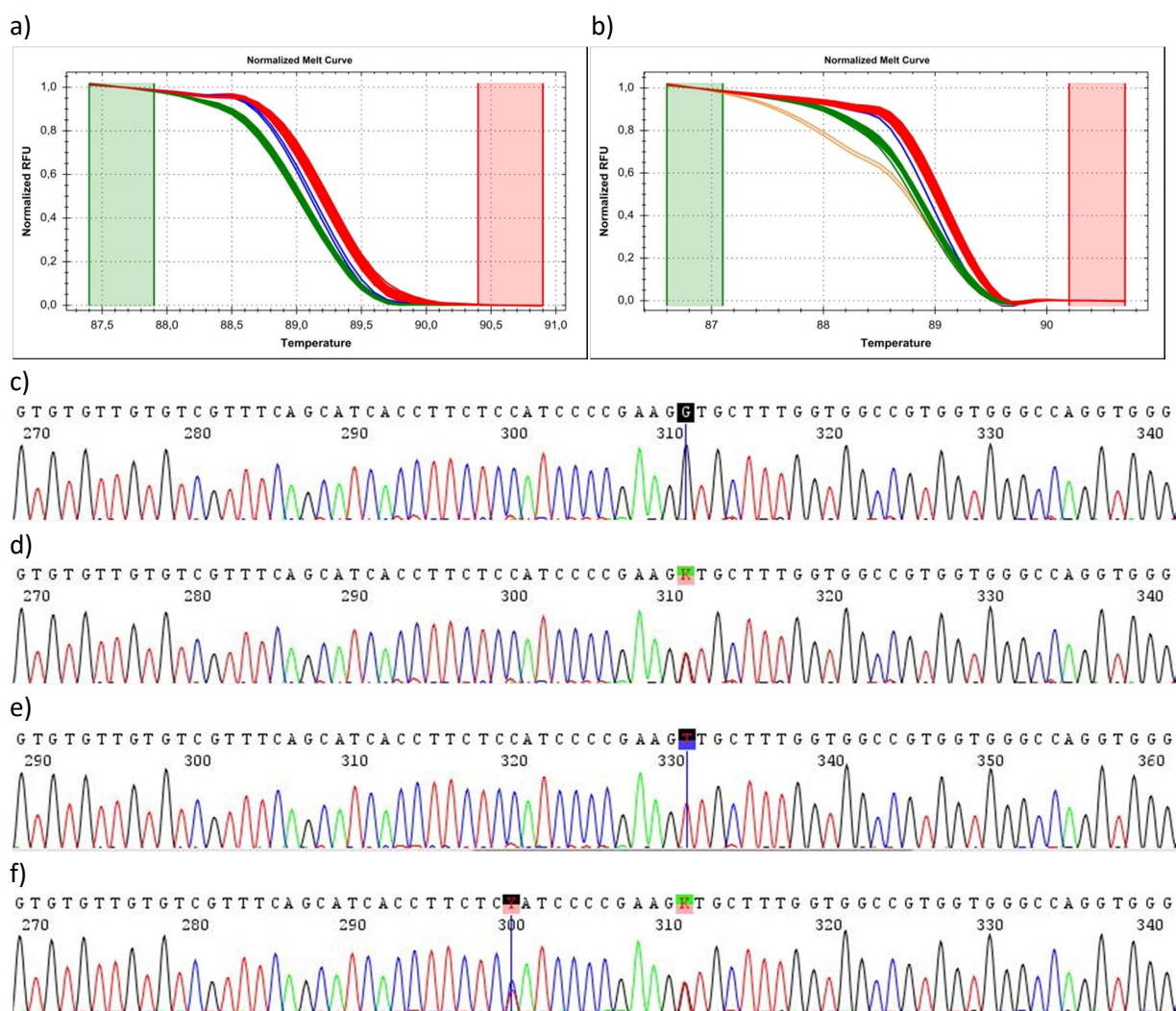


Figure S24 Summary of *ABCC1* exon 16 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16079255 – 16079431, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.2012G>T (in black box)), e) from the blue clusters (homozygous variant c.2012G>T (in black box)), f) from the orange cluster (heterozygous variant c.2001C>T (in black box) and heterozygous variant c.2012G>T).

Table S24 Summary of *ABCC1* exon 16 (the first scanned area, NC_00016.10: 16079255 – 16079431) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	3	4	7
Samples in cluster 1 (red)	82	79	161
Samples in cluster 2 (green)	11	13	24
Samples in cluster 3 (blue)	1	1	2
Samples in cluster 4 (orange)	-	1	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	1	1	2
Samples verified by sequencing from cluster 4	-	1	1
Scan rate	-	-	0,9895

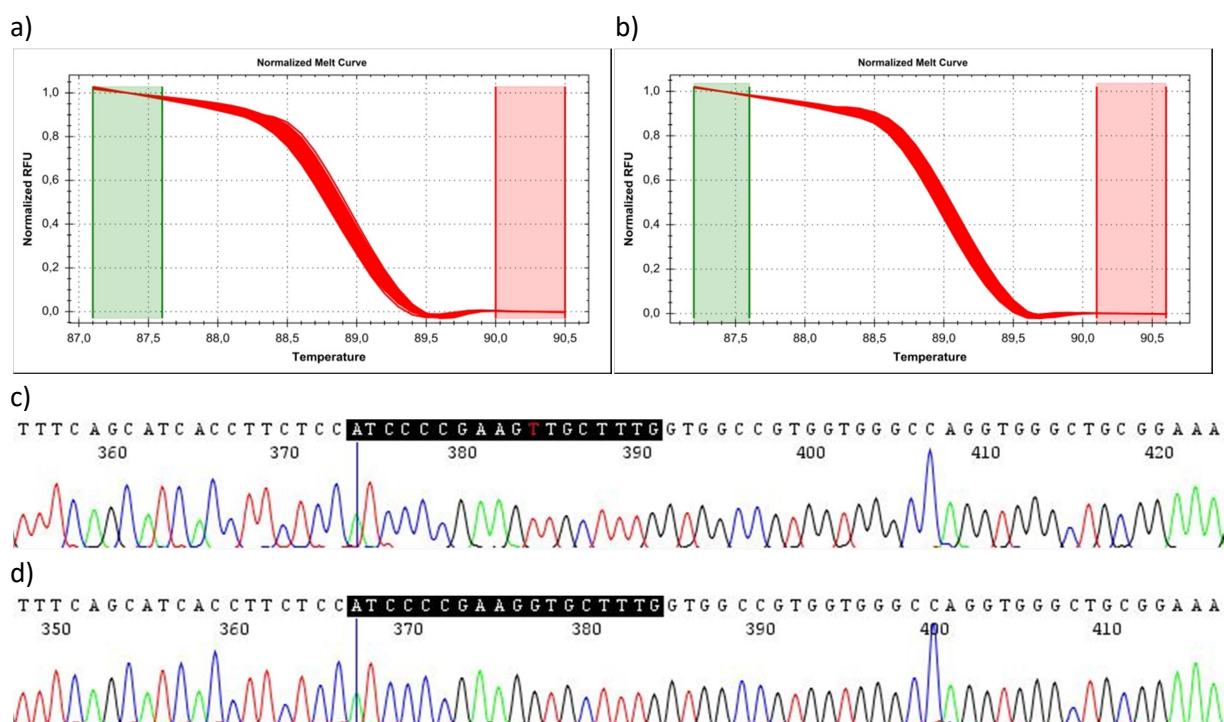


Figure S25 Summary of *ABCC1* exon 16 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16079365 – 16079535, including primers). Examples of sequencing traces for the samples: c) from the red clusters, homozygous variant detected in the forward primer region (black boxes indicate the primer region and variant inside), d) from the red clusters, no variants – no additional variants were detected in this area. The region containing the forward primer sequence was scanned in the previous area (see above Figure S24 and Table S24).

Table S25 Summary of *ABCC1* exon 16 (the second scanned area, NC_00016.10: 16079365 – 16079535) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

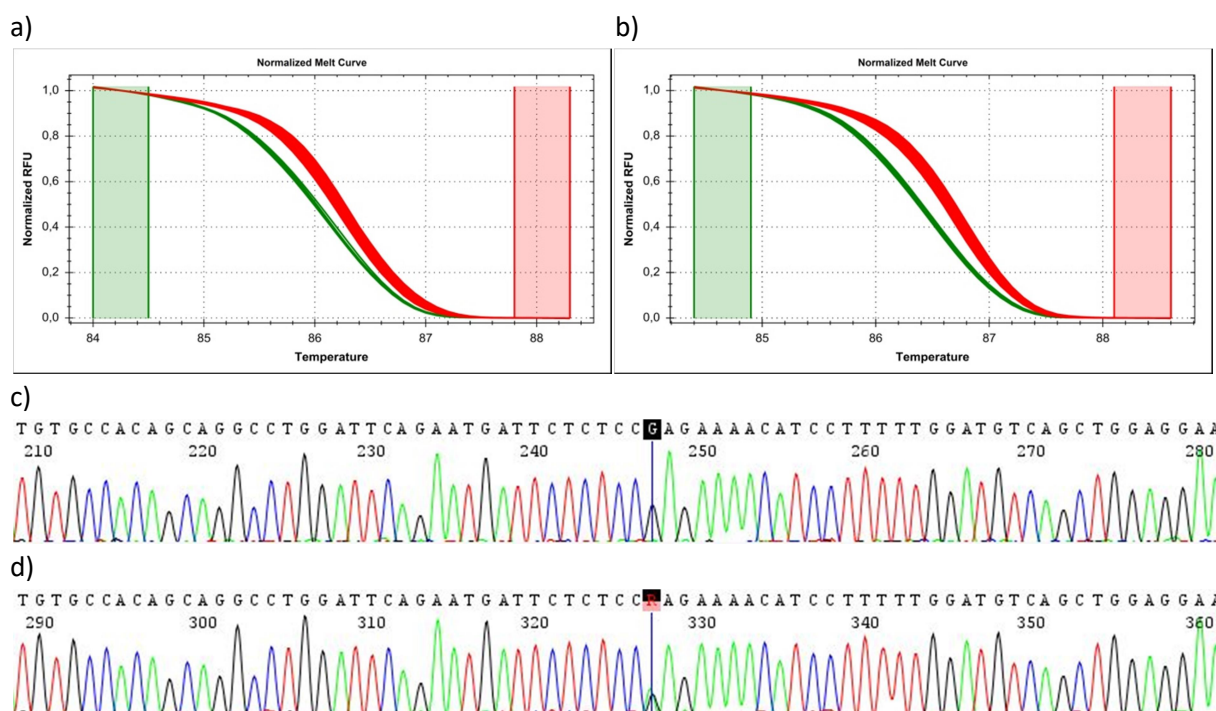


Figure S26 Summary of *ABCC1* exon 17 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16083334 – 16083547, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.2168G>A) (polymorphic base in black boxes).

Table S26 Summary of *ABCC1* exon 17 (the first scanned area, NC_00016.10: 16083334 – 16083547) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	91	90	181
Samples in cluster 2 (green)	4	5	9
Samples excluded	0	0	0
Samples verified by sequencing from cluster 1	3	3	6
Samples verified by sequencing from cluster 2	2	2	4
Scan rate	-	-	1

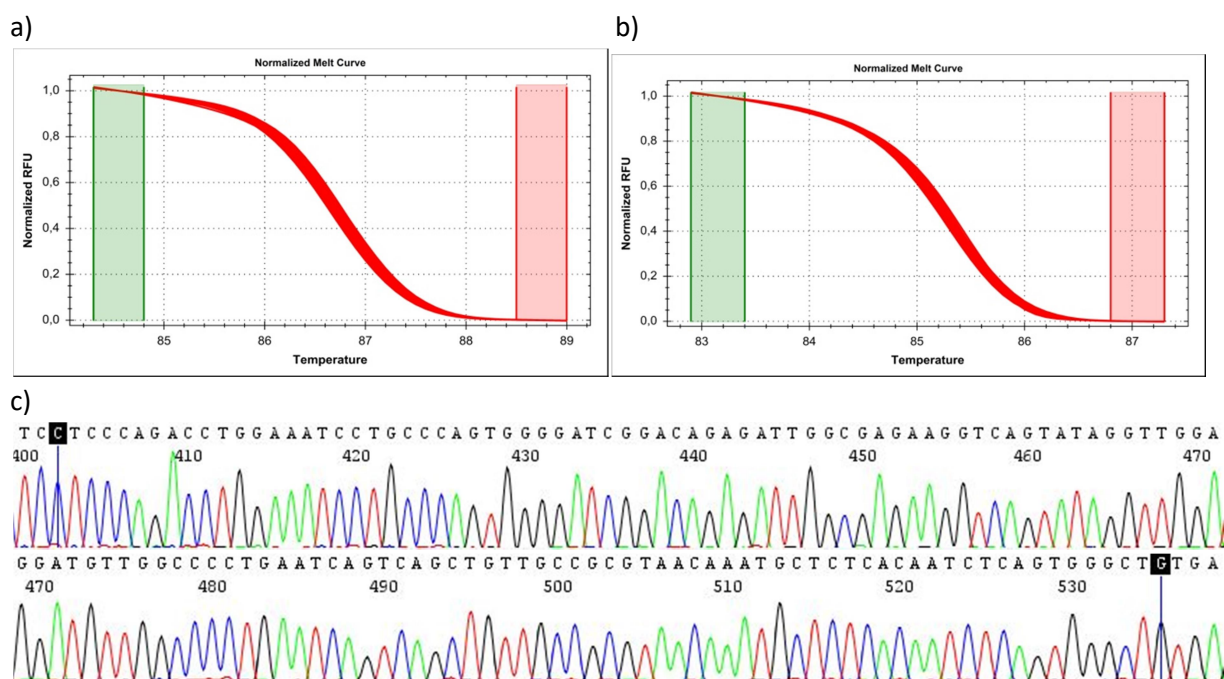


Figure S27 Summary of *ABCC1* exon 17 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16083489 – 16083622, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S27 Summary of *ABCC1* exon 17 (the second scanned area, NC_00016.10: 16083489 – 16083622) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

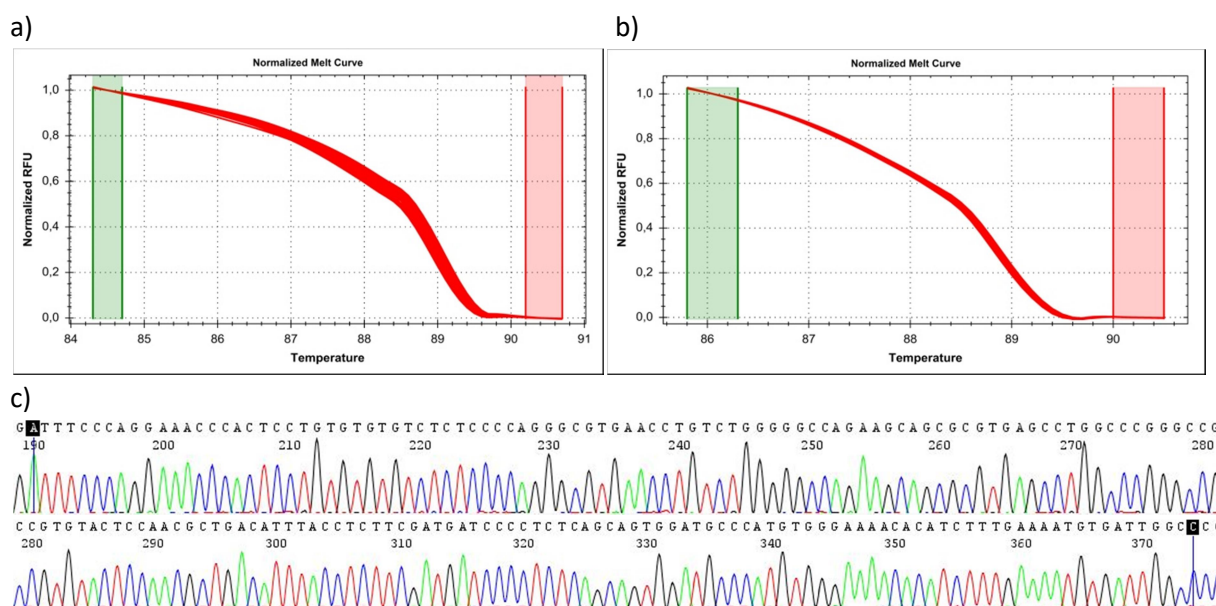


Figure S28 Summary of *ABCC1* exon 18 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16086784 – 16086968, including primers). A example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S28 Summary of *ABCC1* exon 18 (the first scanned area, NC_00016.10: 16086784 – 16086968) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

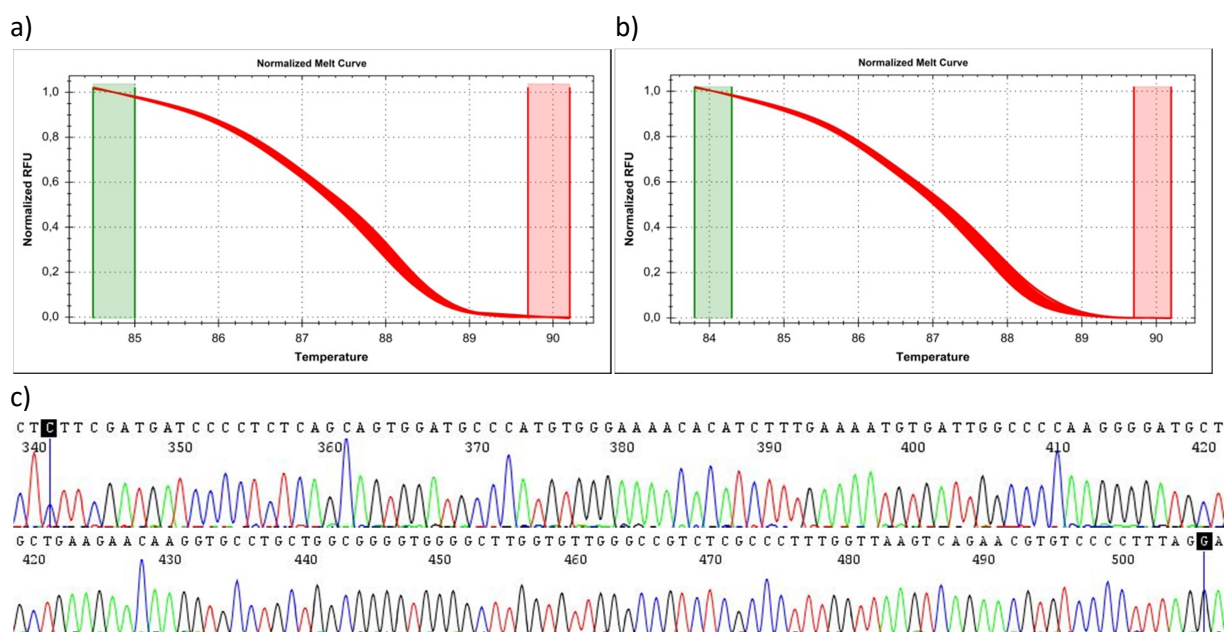


Figure S29 Summary of *ABCC1* exon 18 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16086901 – 16087066, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S29 Summary of *ABCC1* exon 18 (the second scanned area, NC_00016.10: 16086901 – 16087066) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

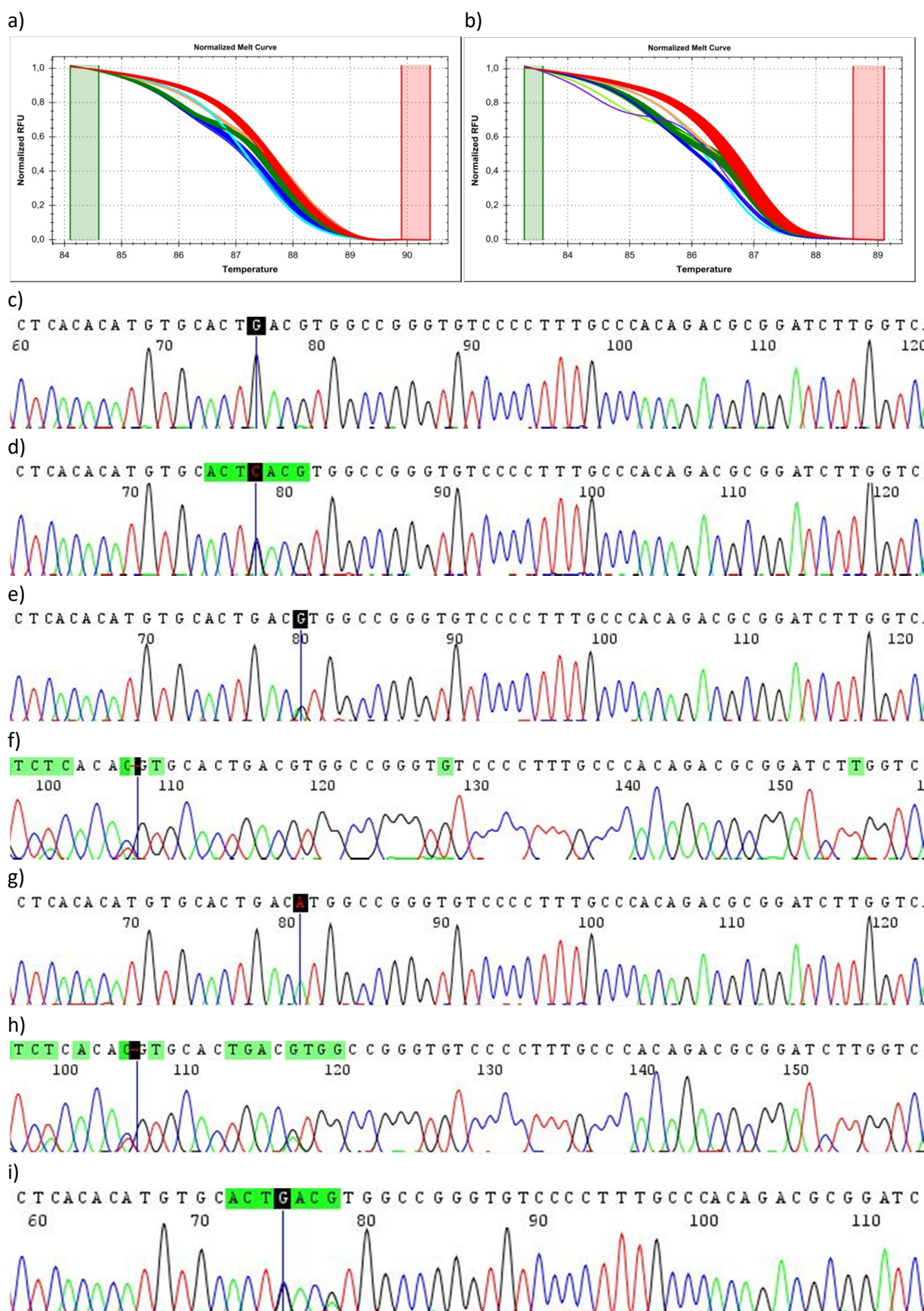


Figure S30 Summary of *ABCC1* exon 19 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16090344 – 16090442, including primers). Examples of sequencing traces for the samples: c) from the red clusters (homozygous variant c.2461-30C>G (in black box)),

d) from the green clusters (heterozygous variant c.2461-30C>G (in black box)), e) from the blue clusters (homozygous variant c.2461-30C>G and heterozygous variant c.2461-27G>A (in black box)), f) from the orange clusters (deletion variant c.2461-39_2461-38delAT (deletion point highlighted in black) and homozygous variant c.2461-30C>G), g) from the cyan clusters (homozygous variant c.2461-30C>G and homozygous variant c.2461-27G>A (in black box)), h) from the lime cluster (deletion variant c.2461-39_2461-38delAT (deletion point highlighted in black) and homozygous variant c.2461-30C>G and heterozygous variant c.2461-27G>A), i) from the violet cluster (heterozygous variant c.2461-30C>G (in black box) and heterozygous variant c.2461-27G>A).

Table S30 Summary of *ABCC1* exon 19 (the first scanned area, NC_00016.10: 16090344 – 16090442) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	5	7	12
Samples in cluster 1 (red)	62	59	121
Samples in cluster 2 (green)	22	21	43
Samples in cluster 3 (blue)	5	7	12
Samples in cluster 4 (orange)	4	4	8
Samples in cluster 5 (cyan)	1	1	2
Samples in cluster 6 (lime)	-	1	1
Samples in cluster 7 (violet)	-	1	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	2	2	4
Samples verified by sequencing from cluster 5	1	1	2
Samples verified by sequencing from cluster 6	-	1	1
Samples verified by sequencing from cluster 7	-	1	1
Scan rate	-	-	0,9895

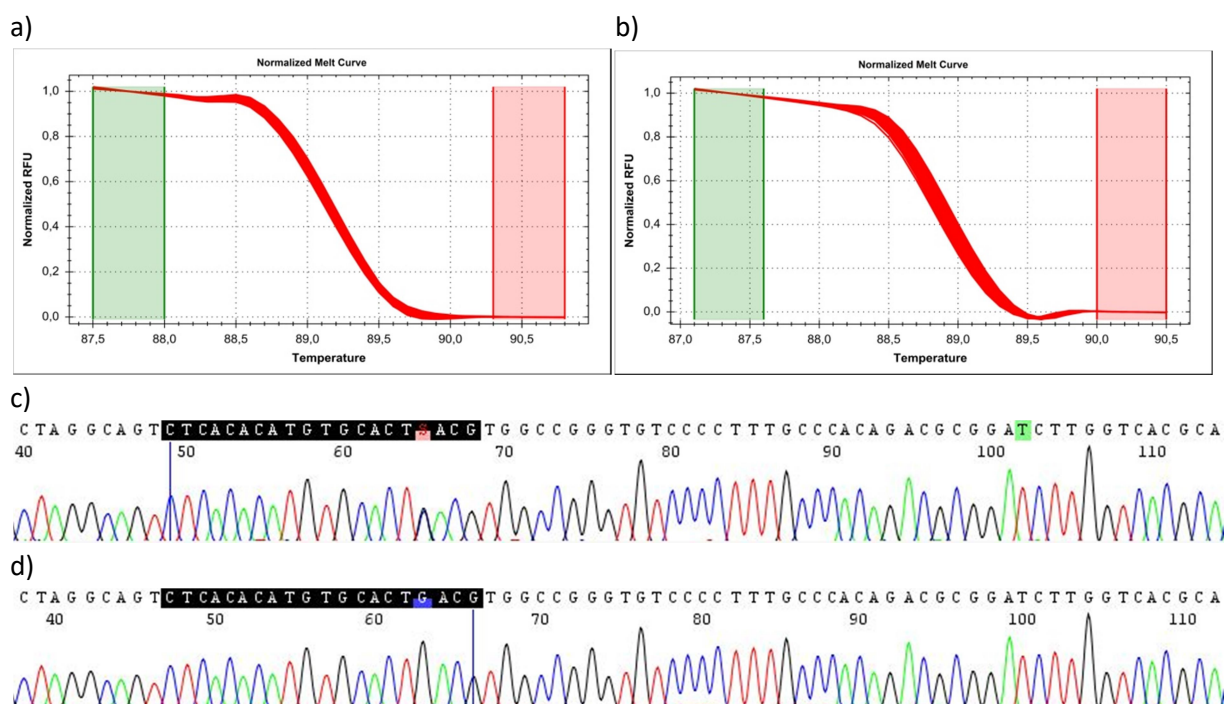


Figure S31 Summary of *ABCC1* exon 19 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16090344 – 16090562, including primers). Examples of sequencing traces for the samples: c) from the red clusters, heterozygous variants detected in the forward primer region (black boxes indicate the primer region and variant inside), d) from the red clusters, homozygous variant detected in the forward primer region – additional scanning area flanking this region was created (see above, Figure S30 and Table S30), no additional variants were detected in this area.

Table S31 Summary of *ABCC1* exon 19 (the second scanned area, NC_00016.10: 16090344 – 16090562) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	92	186
Samples excluded	1	3	4
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9789

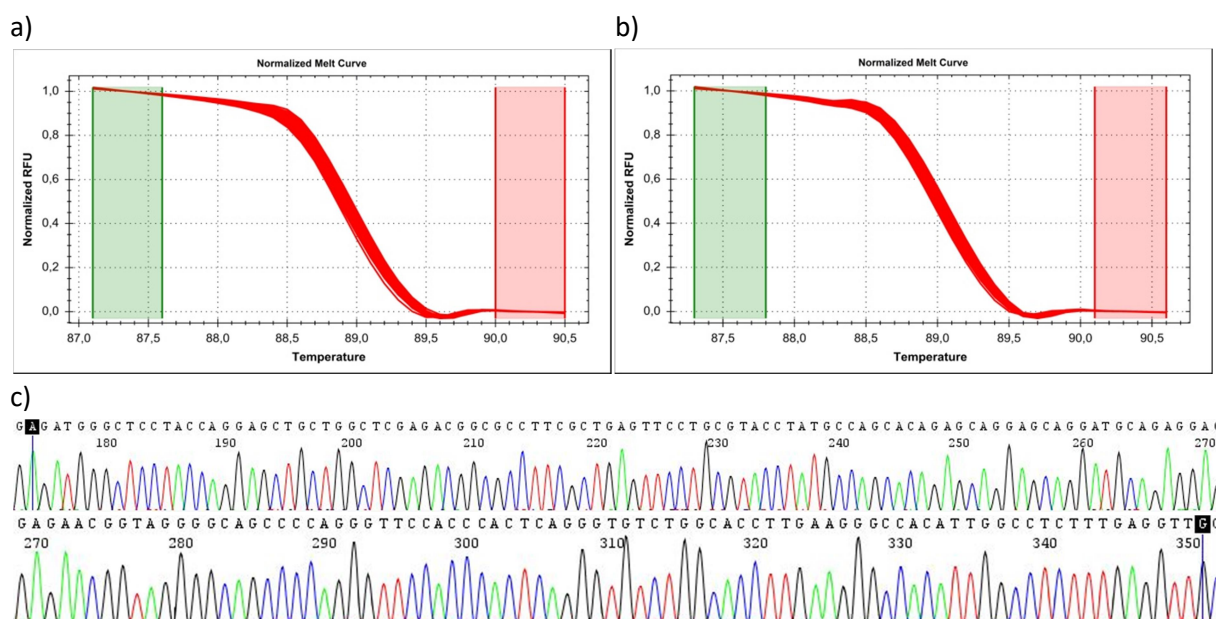


Figure S32 Summary of *ABCC1* exon 19 scanning by HRM: a), b) melting plots for the third scanned area of the exon (NC_00016.10: 16090483 – 16090664, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S32 Summary of *ABCC1* exon 19 (the third scanned area, NC_00016.10: 16090483 – 16090664) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

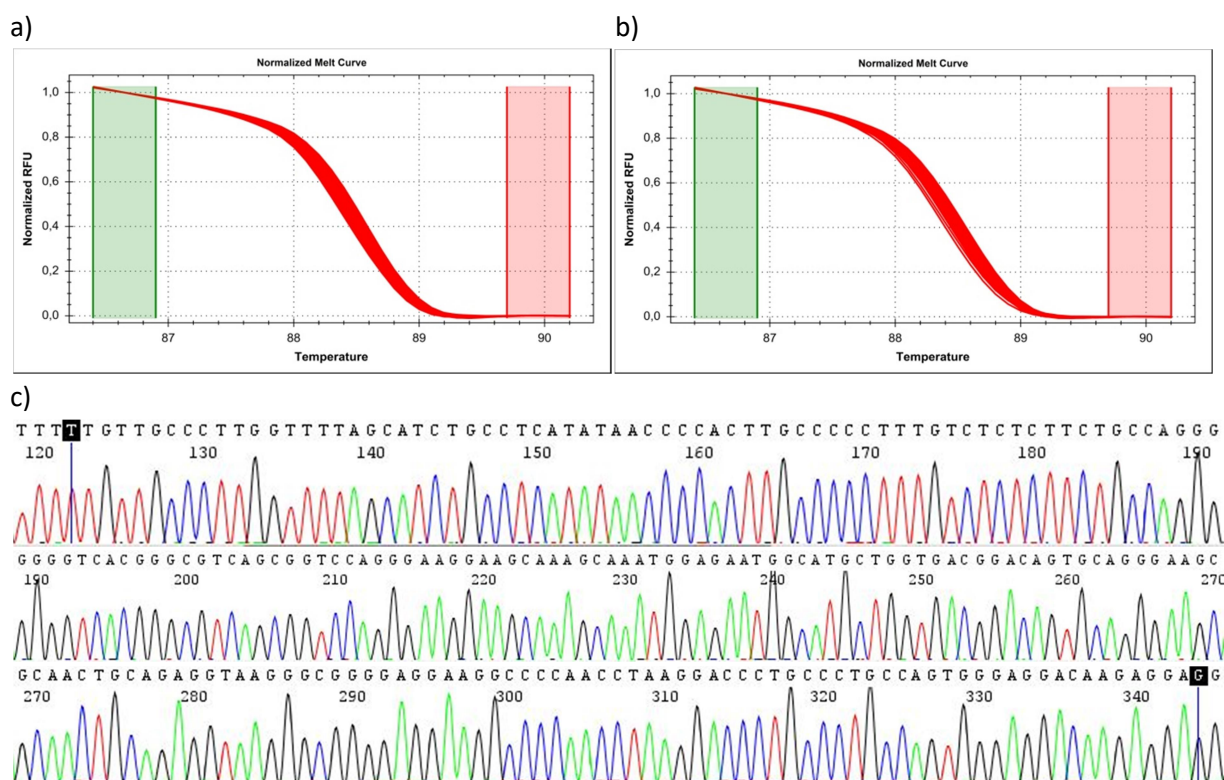


Figure S33 Summary of *ABCC1* exon 20 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16102559 – 16102781, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S33 Summary of *ABCC1* exon 20 (NC_00016.10: 16102559 – 16102781) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

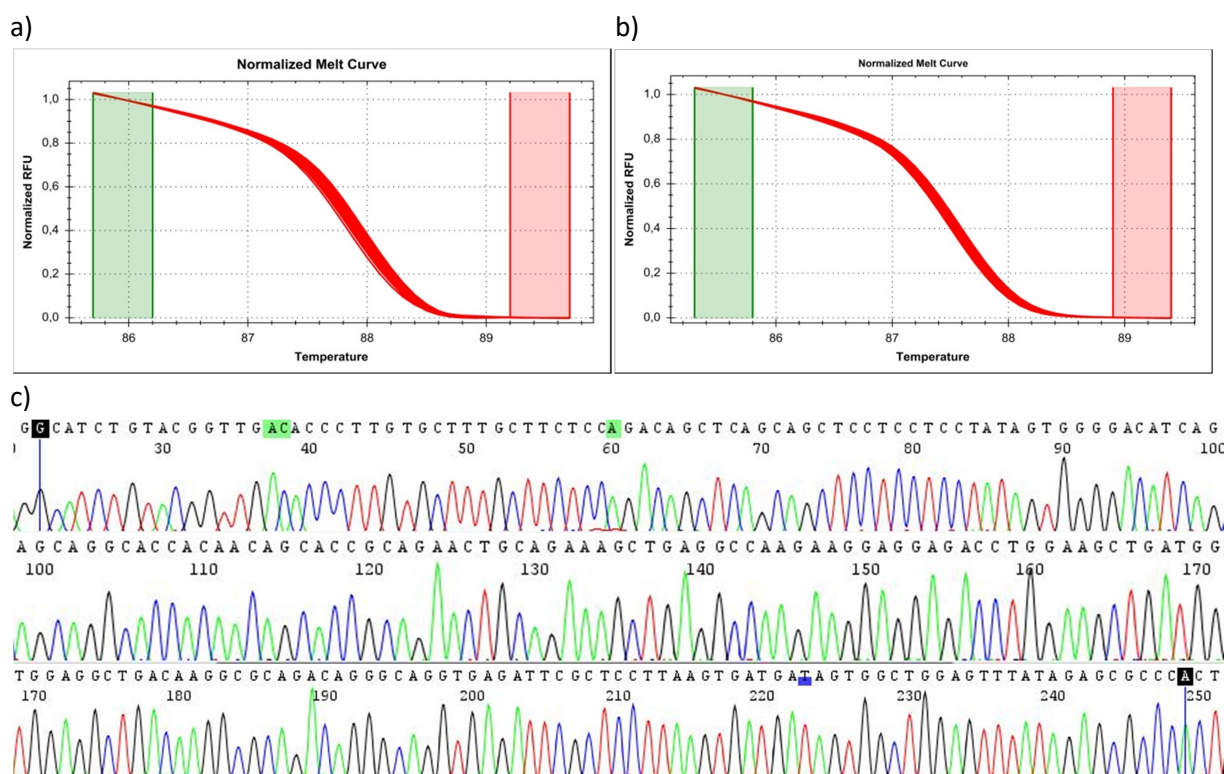


Figure S34 Summary of *ABCC1* exon 21 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16106698 – 16106925, including primers). An example of sequencing trace for the samples: c) from the red clusters, homozygous variant 2871+26C>T (highlighted in blue) detected for all the samples (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S34 Summary of *ABCC1* exon 21 (NC_00016.10: 16106698 – 16106925) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	95	189
Samples excluded	1	0	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

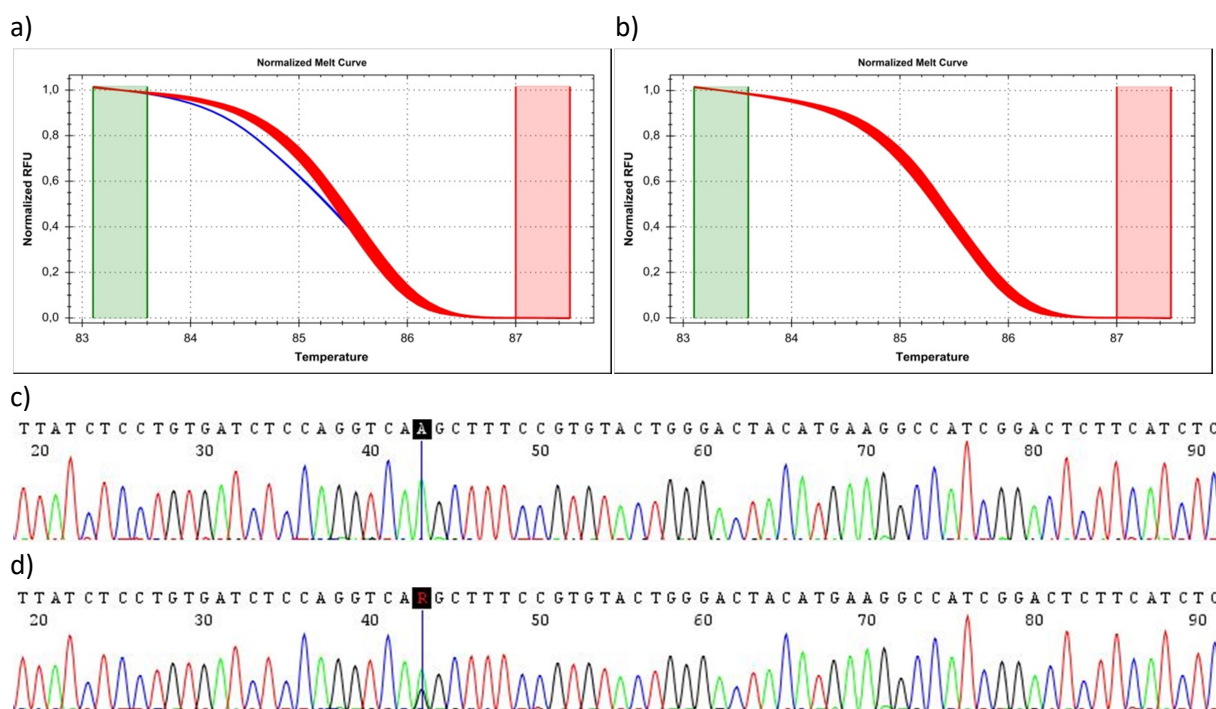


Figure S35 Summary of *ABCC1* exon 22 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16111331 – 16111514, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the blue cluster (heterozygous variant c.2876A>G) (polymorphic base in black boxes).

Table S35 Summary of *ABCC1* exon 22 (the first scanned area, NC_00016.10: 16111331 – 16111514) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	93	94	187
Samples in cluster 2 (blue)	1	-	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Samples verified by sequencing from cluster 2	1	-	1
Scan rate	-	-	0,9895

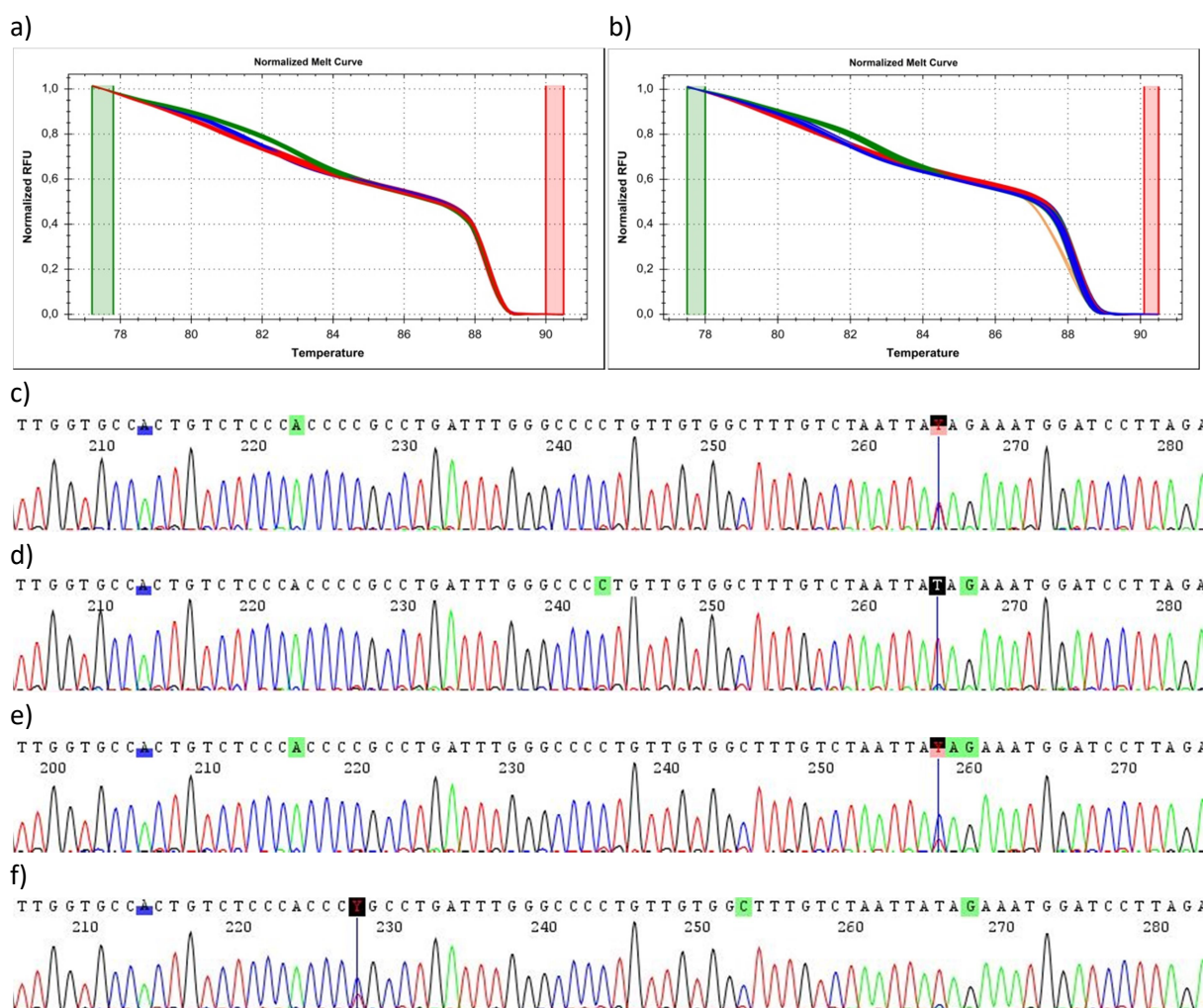


Figure S36 Summary of *ABCC1* exon 22 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16111473 – 16111676, including primers). Examples of sequencing traces for the samples: c) from the red clusters (heterozygous variant c.3079+62T>C (in black box)), d) from the blue clusters (no variants), e) from the green clusters (homozygous variant c.3079+62T>C (in black box)), f) from the orange cluster (heterozygous variant c.3079+24C>T (in black box)), homozygous variant 3079+10G>A (highlighted in blue) detected for all the samples.

Table S36 Summary of *ABCC1* exon 22 (the second scanned area, NC_00016.10: 16111473 – 16111676) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	3	4	7
Samples in cluster 1 (red)	58	38	96
Samples in cluster 2 (blue)	24	42	66
Samples in cluster 3 (green)	12	13	25
Samples in cluster 4 (orange)	-	1	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Samples verified by sequencing from cluster 4	-	1	1
Scan rate	-	-	0,9895

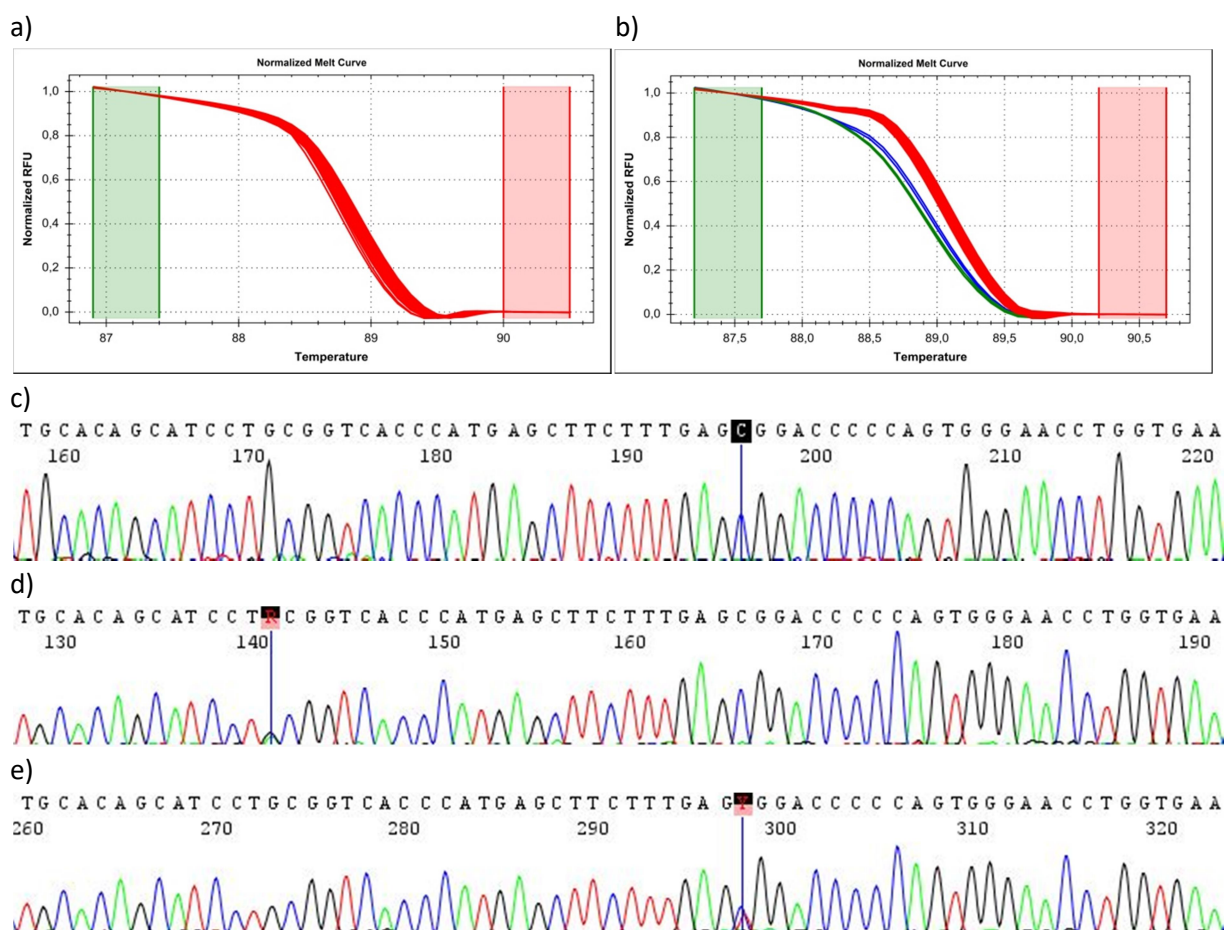


Figure S37 Summary of *ABCC1* exon 23 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16114725 – 16114938, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.3171G>A (in black box)), e) from the blue cluster (heterozygous variant c.3196C>T (in black box)).

Table S37 Summary of *ABCC1* exon 23 (the first scanned area, NC_00016.10: 16114725 – 16114938) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	3	4
Samples in cluster 1 (red)	95	91	186
Samples in cluster 2 (green)	-	2	2
Samples in cluster 3 (blue)	-	1	1
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Samples verified by sequencing from cluster 2	-	2	2
Samples verified by sequencing from cluster 3	-	1	1
Scan rate	-	-	0,9947

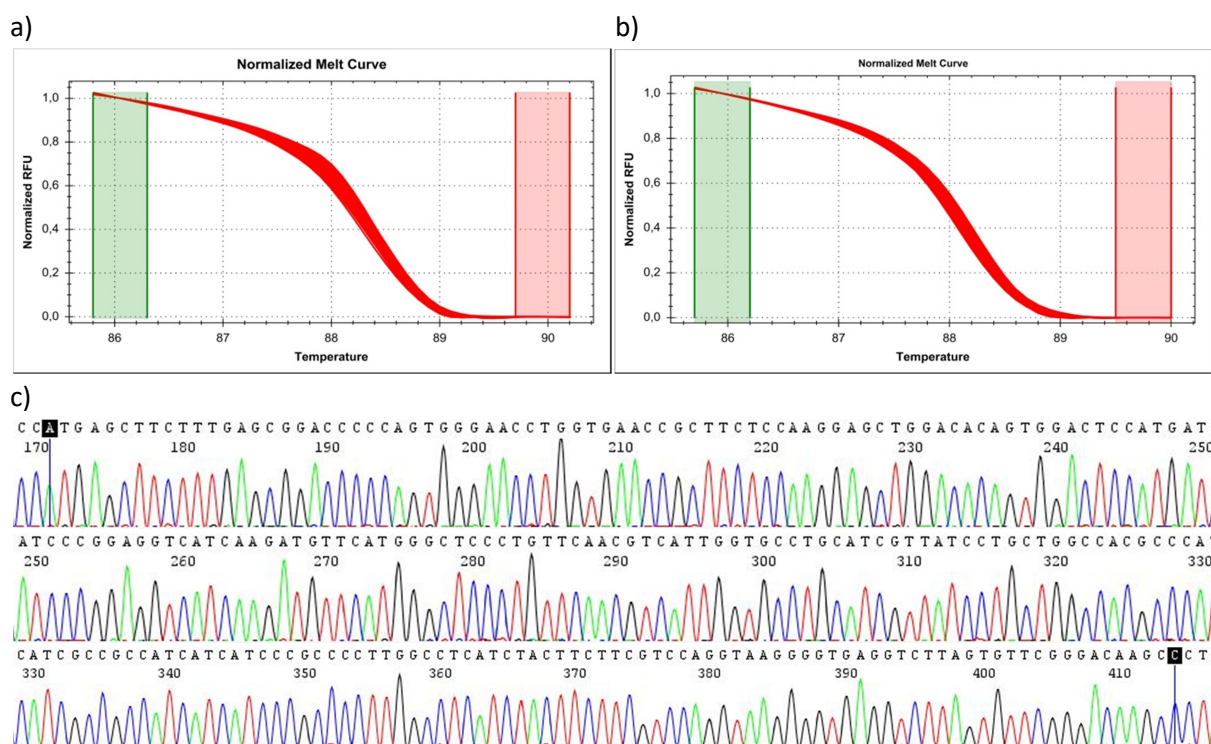


Figure S38 Summary of *ABCC1* exon 23 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16114867 – 16115110, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S38 Summary of *ABCC1* exon 23 (the second scanned area, NC_00016.10: 16114867 – 16115110) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	93	186
Samples excluded	2	2	4
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9789

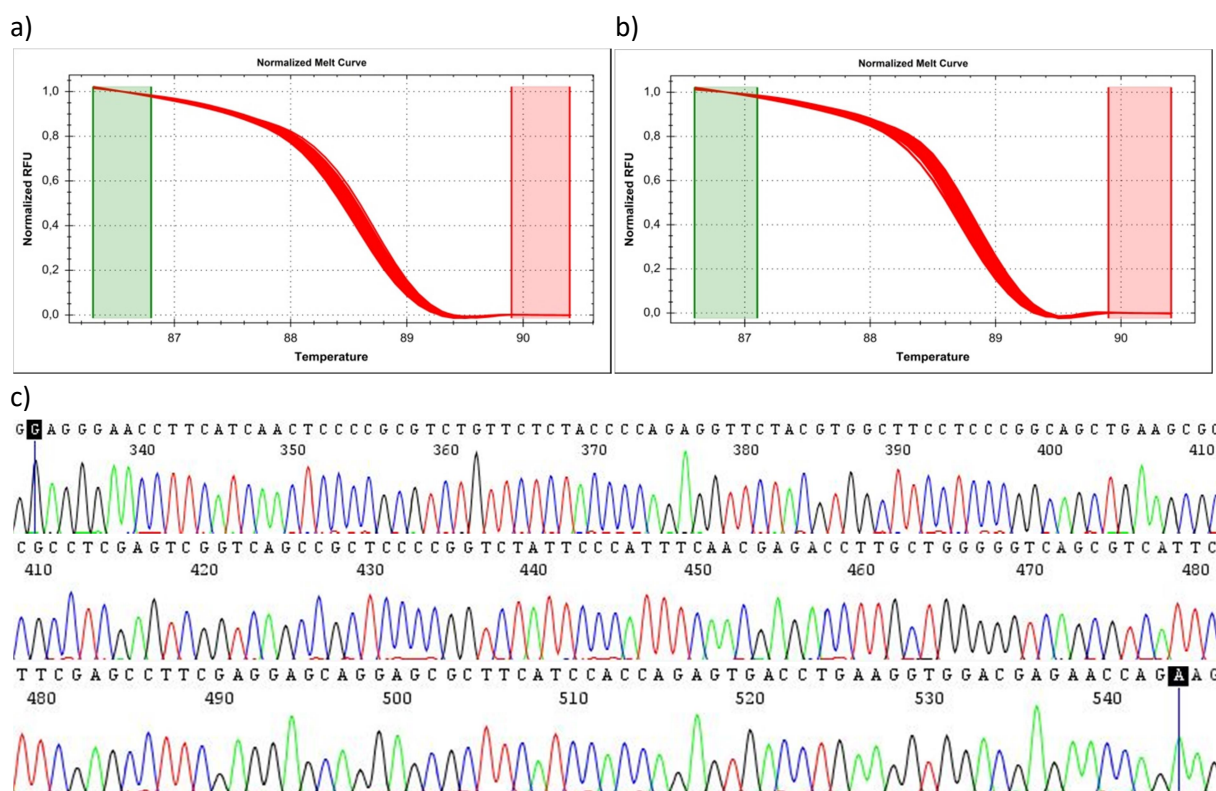


Figure S39 Summary of *ABCC1* exon 24 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16121932 – 16122143, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S39 Summary of *ABCC1* exon 24 (the first scanned area, NC_00016.10: 16121932 – 16122143) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

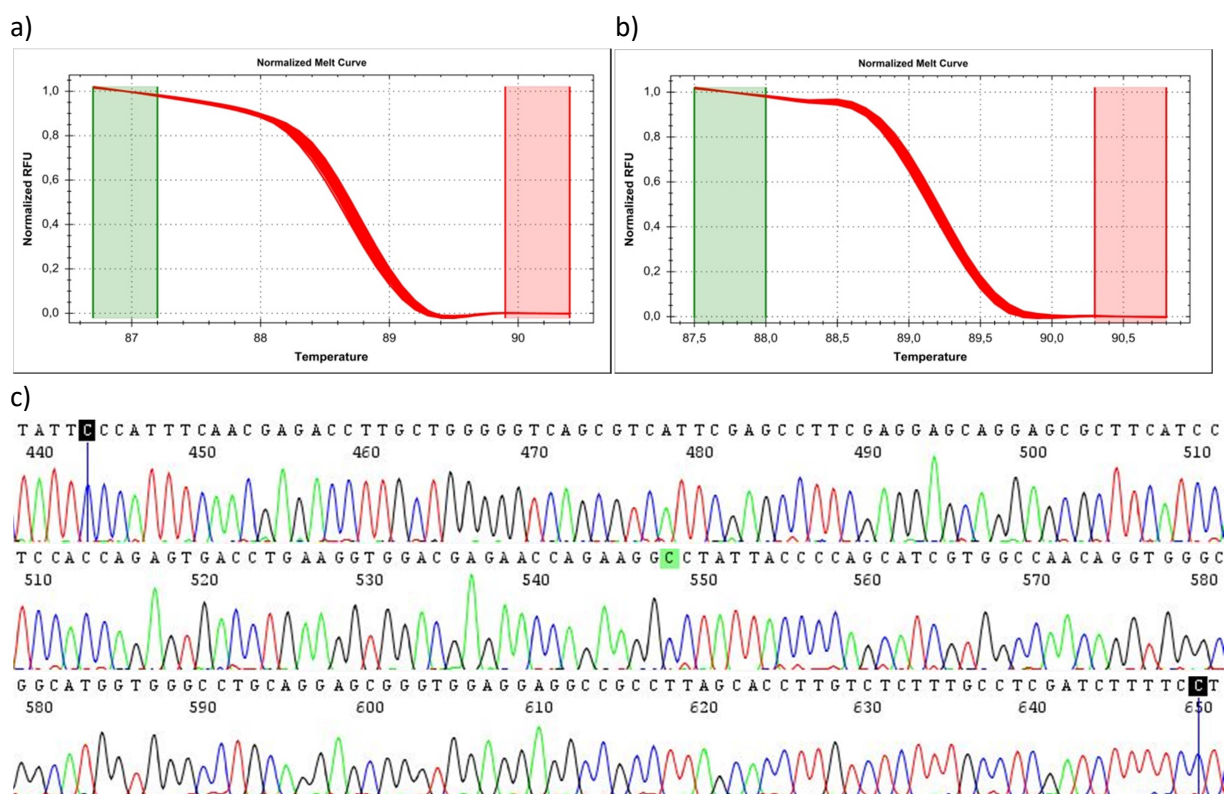


Figure S40 Summary of *ABCC1* exon 24 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16122042 – 16122249, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S40 Summary of *ABCC1* exon 24 (the second scanned area, NC_00016.10: 16122042 – 16122249) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

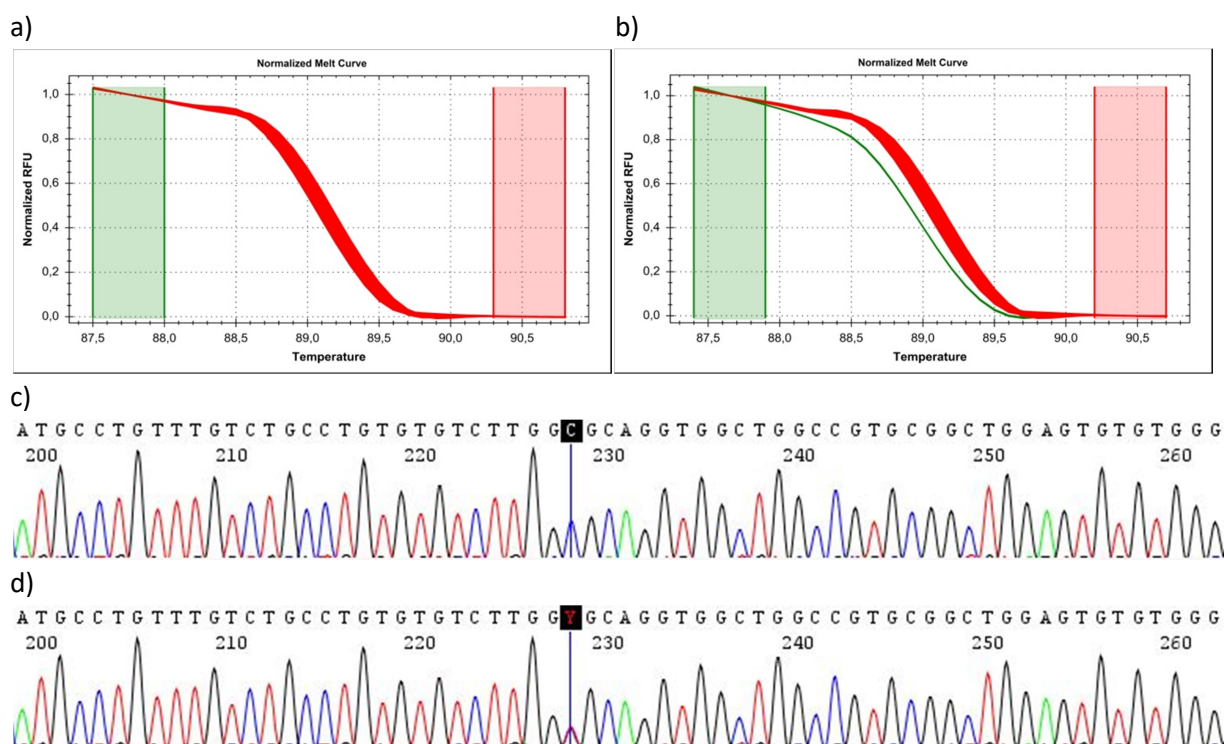


Figure S41 Summary of *ABCC1* exon 25 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16124743 – 16124967, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.3591-5C>T) (polymorphic base in black boxes).

Table S41 Summary of *ABCC1* exon 25 (NC_00016.10: 16124743 – 16124967) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	2	3
Samples in cluster 1 (red)	94	93	187
Samples in cluster 2 (green)	-	1	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	2	5
Samples verified by sequencing from cluster 2	-	1	1
Scan rate	-	-	0,9895

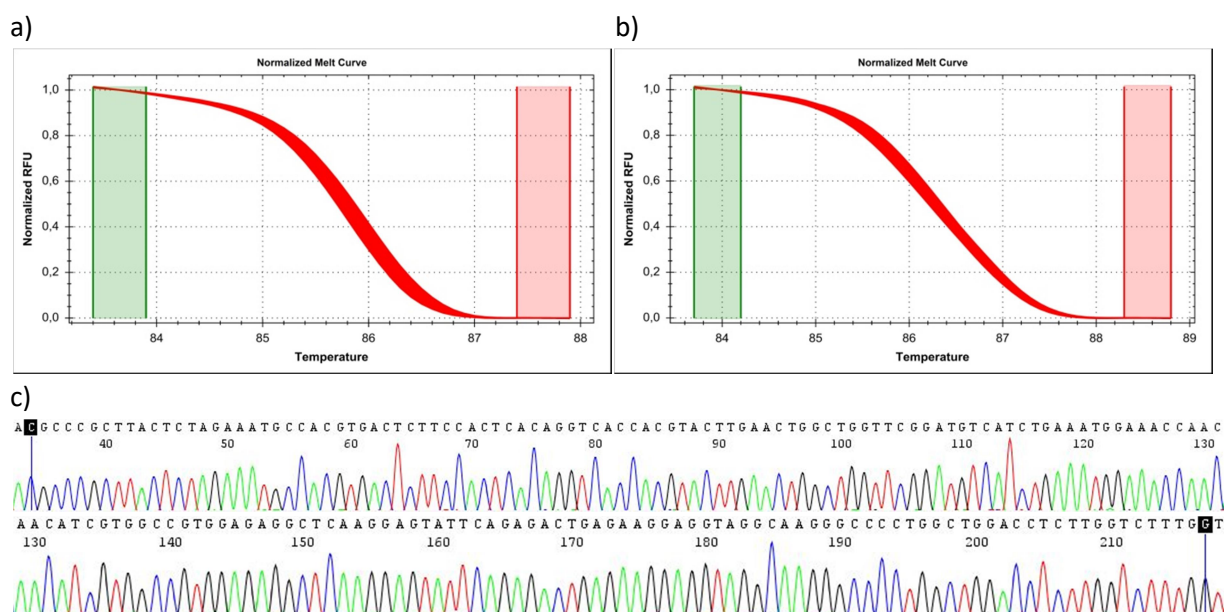


Figure S42 Summary of *ABCC1* exon 26 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16125766 – 16125949, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S42 Summary of *ABCC1* exon 26 (NC_00016.10: 16125766 – 16125949) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	93	188
Samples excluded	0	2	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

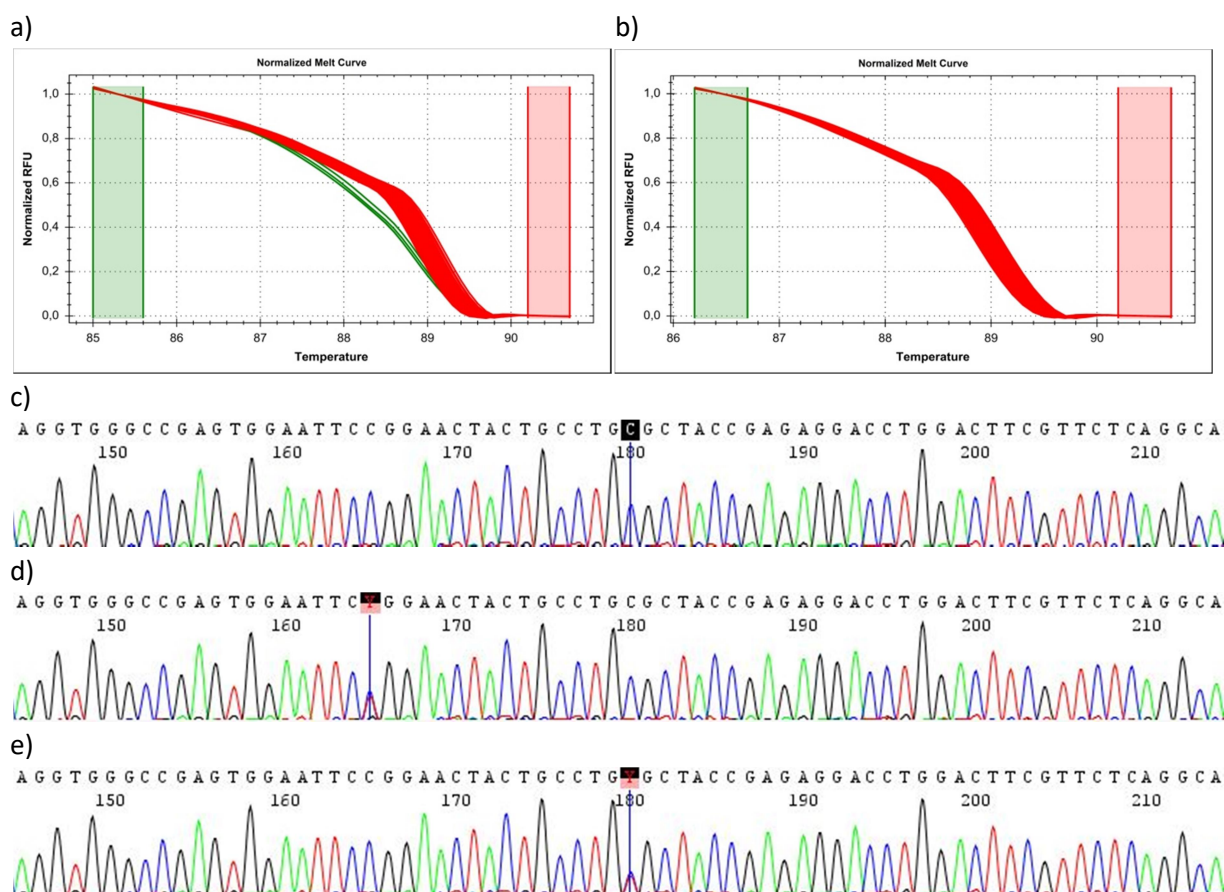


Figure S43 Summary of *ABCC1* exon 27 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16131724 – 16131971, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster, one sample (heterozygous variant c.3886C>T), e) from the green cluster, one sample (heterozygous variant c.3901C>T) (polymorphic bases in black boxes).

Table S43 Summary of *ABCC1* exon 27 (NC_00016.10: 16131724 – 16131971) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	92	94	186
Samples in cluster 2 (green)	2	-	2
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	3	5
Samples verified by sequencing from cluster 2	2	-	2
Scan rate	-	-	0,9895

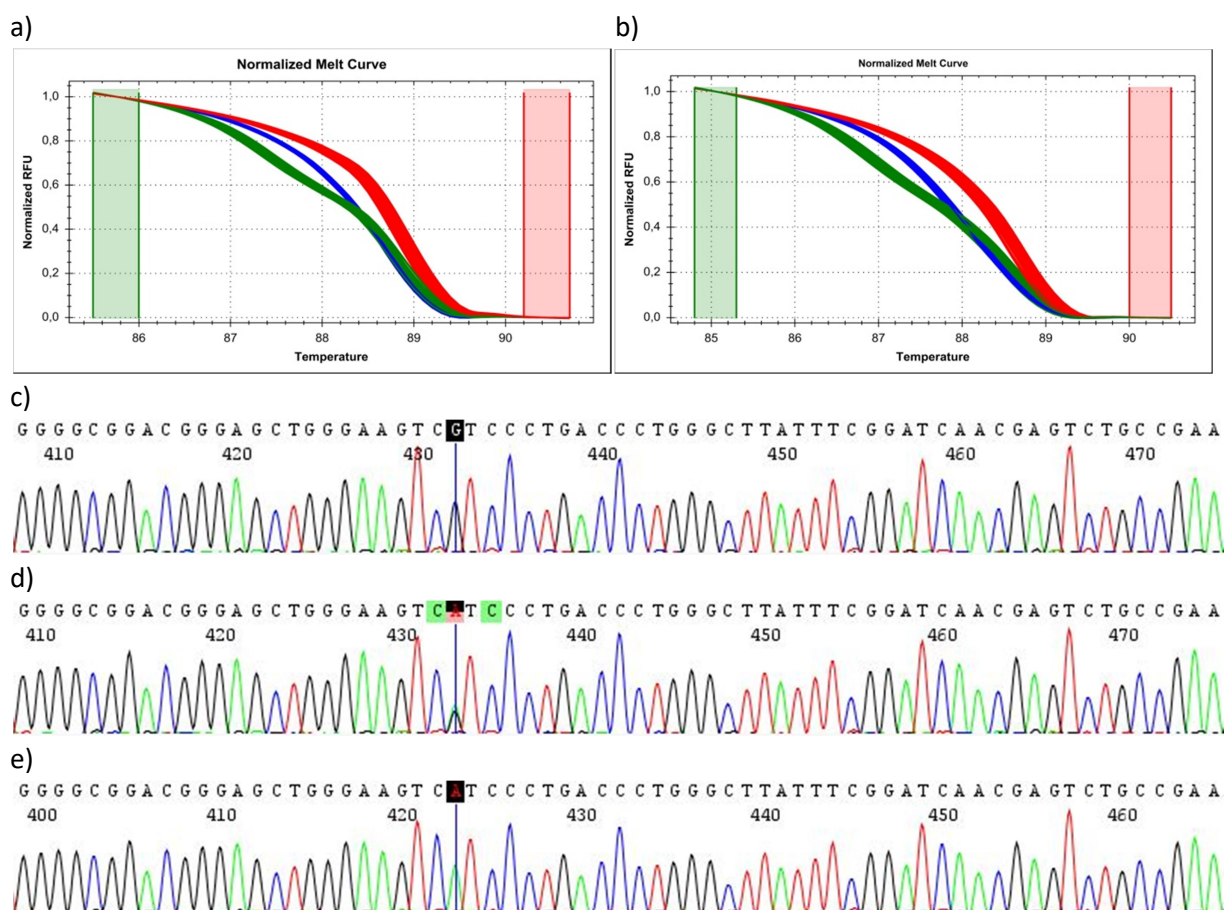


Figure S44 Summary of *ABCC1* exon 28 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16134325 – 16134437, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.4002G>A), e) from the blue cluster (homozygous variant c.4002G>A) (polymorphic base in black boxes).

Table S44 Summary of *ABCC1* exon 28 (the first scanned area, NC_00016.10: 16134325 – 16134437) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	53	49	102
Samples in cluster 2 (green)	32	36	68
Samples in cluster 3 (blue)	9	9	18
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Scan rate	-	-	0,9895

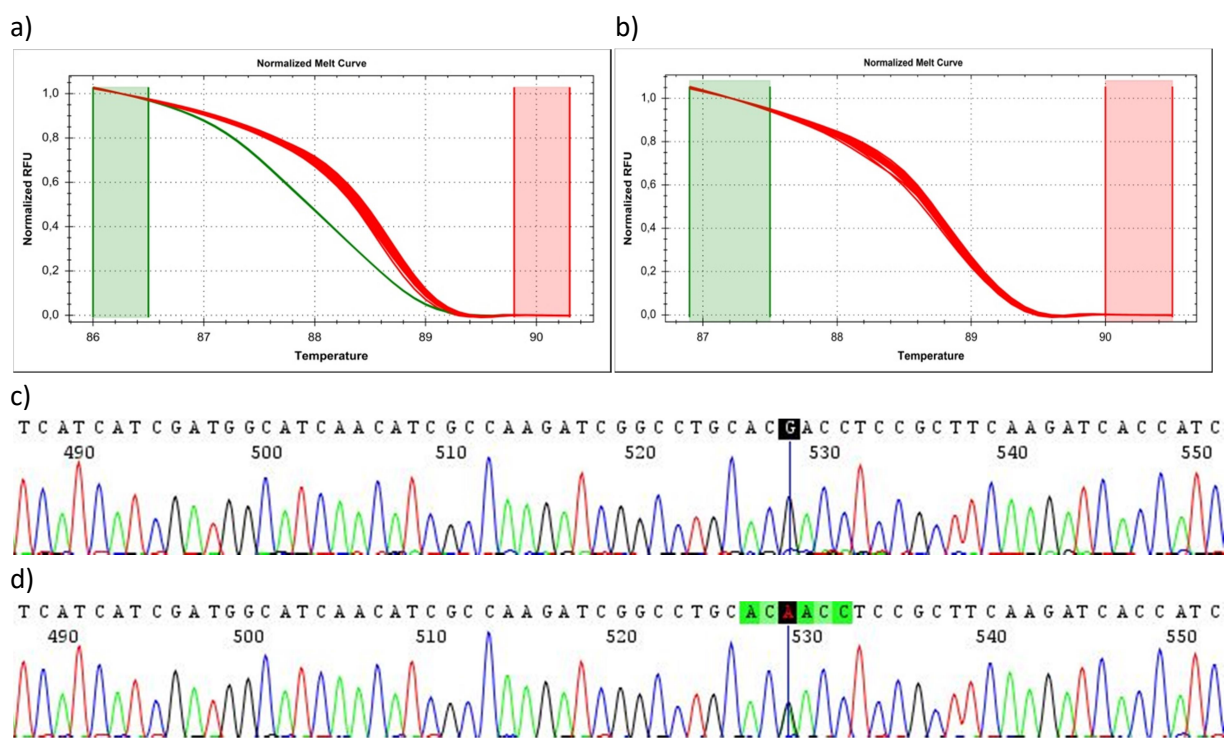


Figure S45 Summary of *ABCC1* exon 28 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16134395 – 16134568, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.4093G>A) (polymorphic bases in black boxes).

Table S45 Summary of *ABCC1* exon 28 (the second scanned area, NC_00016.10: 16134395 – 16134568) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	1	3
Samples in cluster 1 (red)	93	94	187
Samples in cluster 2 (green)	1	-	1
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	2	3	5
Samples verified by sequencing from cluster 2	1	-	1
Scan rate	-	-	0,9895

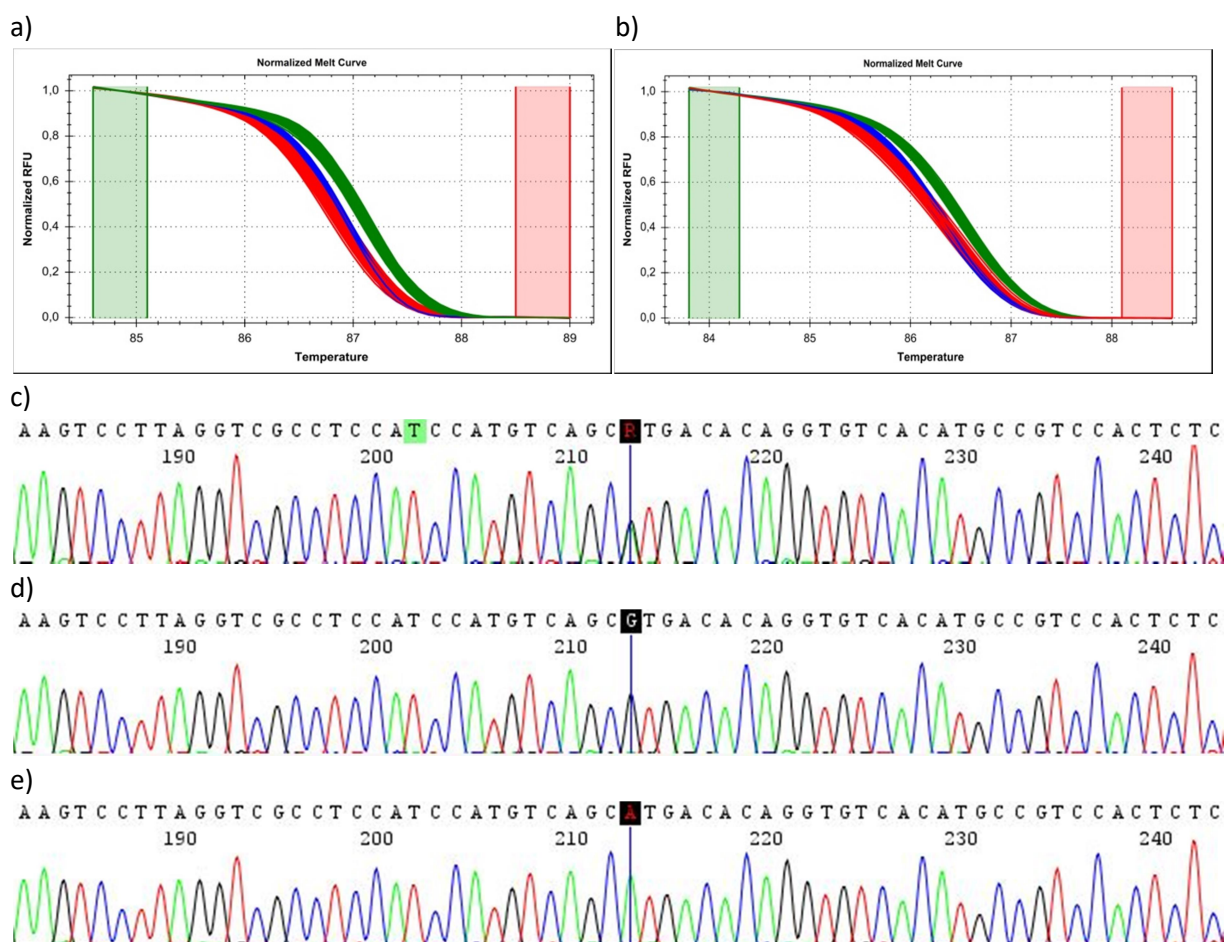


Figure S46 Summary of *ABCC1* exon 29 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16136403 – 16136629, including primers). Examples of sequencing traces for the samples: c) from the red clusters (heterozygous variant c.4126-45G>A), d) from the green cluster (no variants), e) from the blue cluster (homozygous variant c.4126-45G>A) (polymorphic bases in black boxes).

Table S46 Summary of *ABCC1* exon 29 (the first scanned area, NC_00016.10: 16136403 – 16136629) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	3	3	6
Samples in cluster 1 (red)	40	45	85
Samples in cluster 2 (green)	35	27	62
Samples in cluster 3 (blue)	20	22	42
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	2	2	4
Scan rate	-	-	0,9947

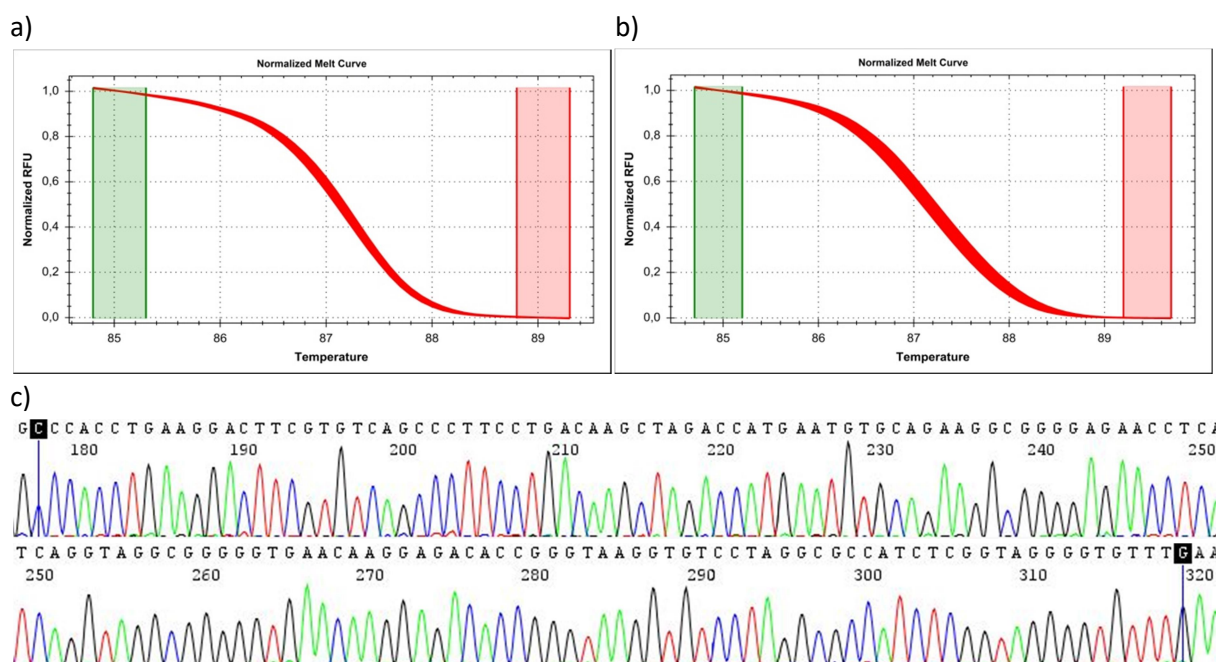


Figure S47 Summary of *ABCC1* exon 29 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16136569 – 16136711, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S47 Summary of *ABCC1* exon 29 (the second scanned part, NC_00016.10: 16136569 – 16136711) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	94	94	188
Samples excluded	1	1	2
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9895

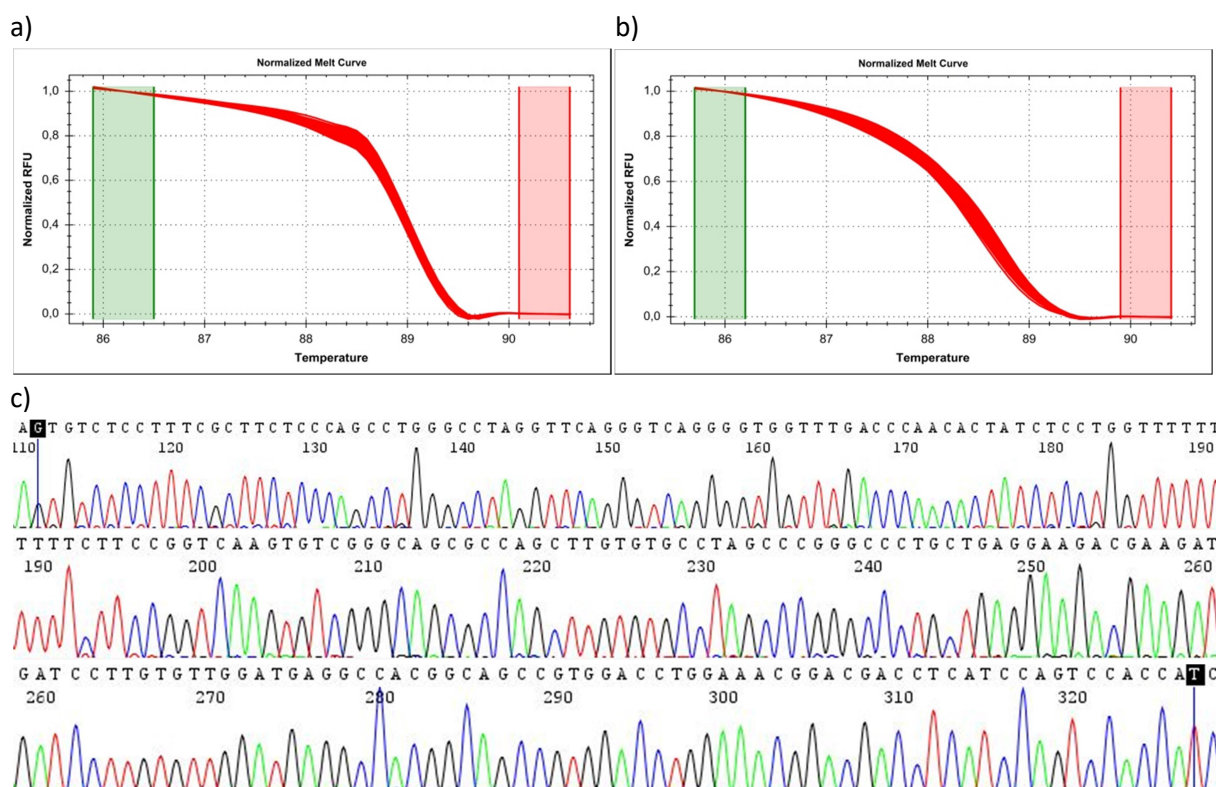


Figure S48 Summary of *ABCC1* exon 30 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_00016.10: 16138270 – 16138486, including primers). An example of sequencing trace for the samples: c) from the red clusters, no variants were detected (black boxes indicate the beginning and the end of the HRM product, including primers).

Table S48 Summary of *ABCC1* exon 30 (the first scanned part, NC_00016.10: 16138270 – 16138486) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Samples verified by sequencing from cluster 1	3	3	6
Scan rate	-	-	0,9947

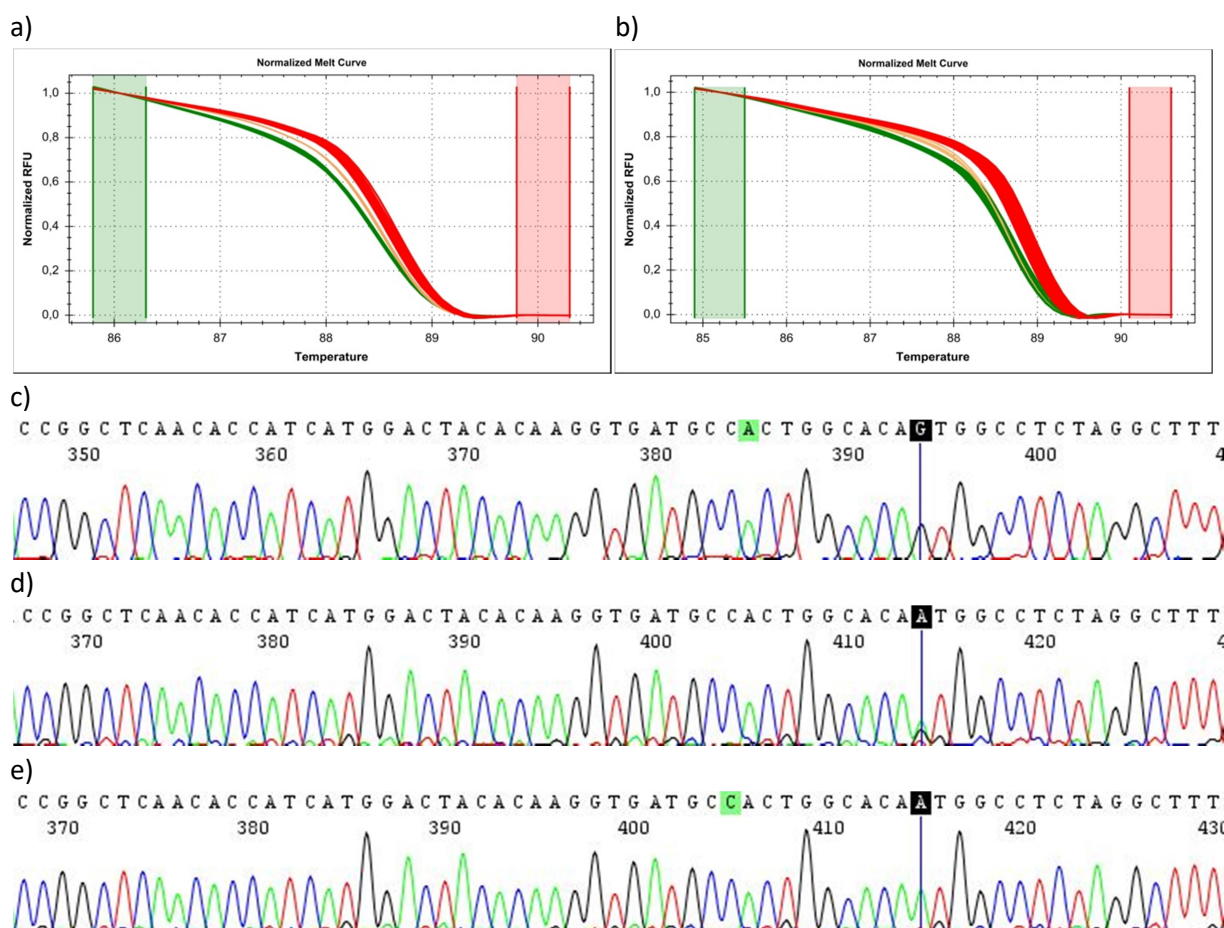


Figure S49 Summary of *ABCC1* exon 30 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_00016.10: 16138400 – 16138603, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green clusters (heterozygous variant c.4487+18G>A), e) from the orange clusters (homozygous variant c.4487+18G>A) (polymorphic bases in black boxes).

Table S49 Summary of *ABCC1* exon 30 (the second scanned area, NC_00016.10: 16138400 – 16138603) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	3	3	6
Samples in cluster 1 (red)	68	68	136
Samples in cluster 2 (green)	24	23	47
Samples in cluster 3 (orange)	1	4	5
Samples excluded	2	0	2
Samples verified by sequencing from cluster 1	2	2	4
Samples verified by sequencing from cluster 2	2	2	4
Samples verified by sequencing from cluster 3	1	2	3
Scan rate	-	-	0,9895

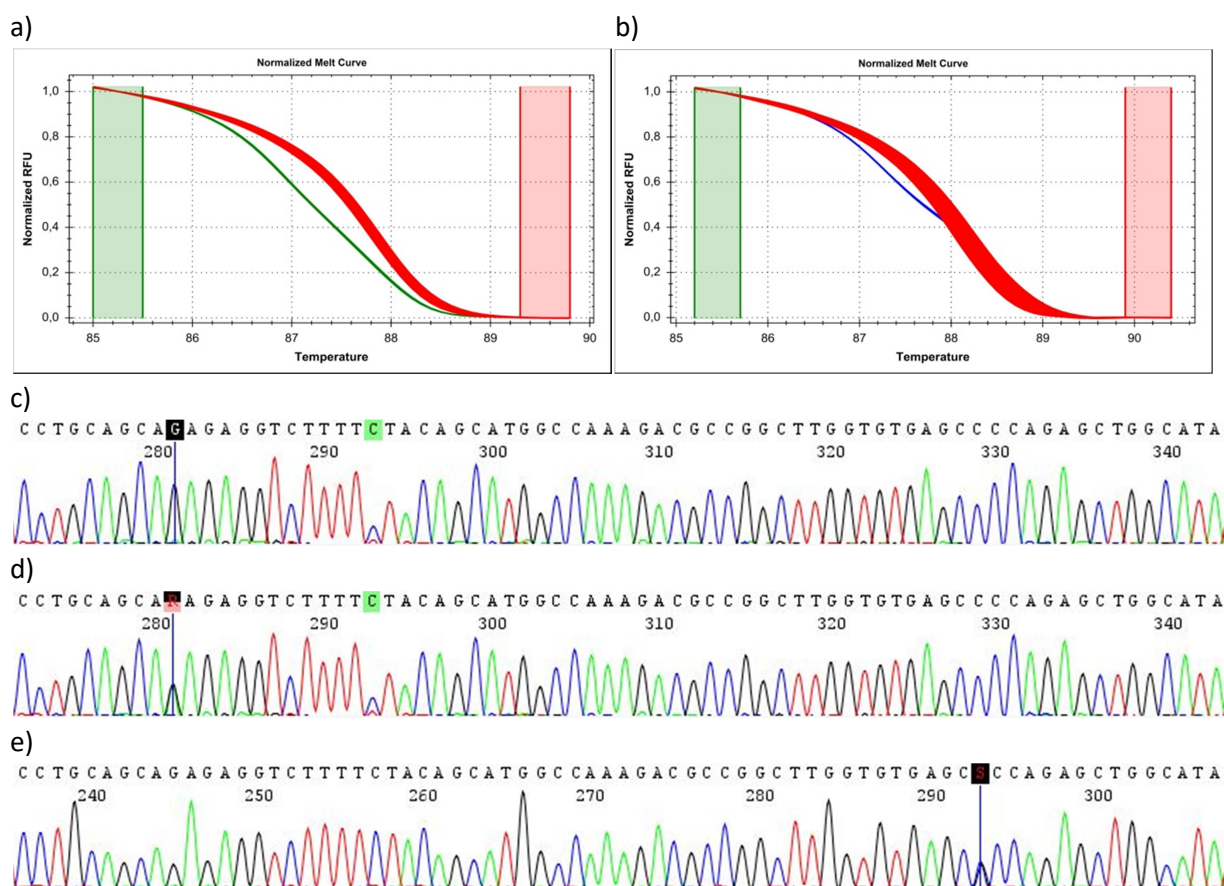


Figure S50 Summary of *ABCC1* exon 31 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_00016.10: 16141138 – 16141322, including primers). Examples of sequencing traces for the samples: c) from the red clusters (no variants), d) from the green cluster (heterozygous variant c.4551G>A, the novel one), e) from the blue cluster (homozygous variant c.*3C>G) (polymorphic bases in black boxes).

Table S50 Summary of *ABCC1* exon 31 (NC_00016.10: 16141138 – 16141322) scanning by HRM. Scan rate is the ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	2	2	4
Samples in cluster 1 (red)	93	92	185
Samples in cluster 2 (green)	1	-	1
Samples in cluster 3 (blue)	-	1	1
Samples excluded	1	2	3
Samples verified by sequencing from cluster 1	3	3	6
Samples verified by sequencing from cluster 2	1	-	1
Samples verified by sequencing from cluster 3	-	1	1
Scan rate	-	-	0,9842

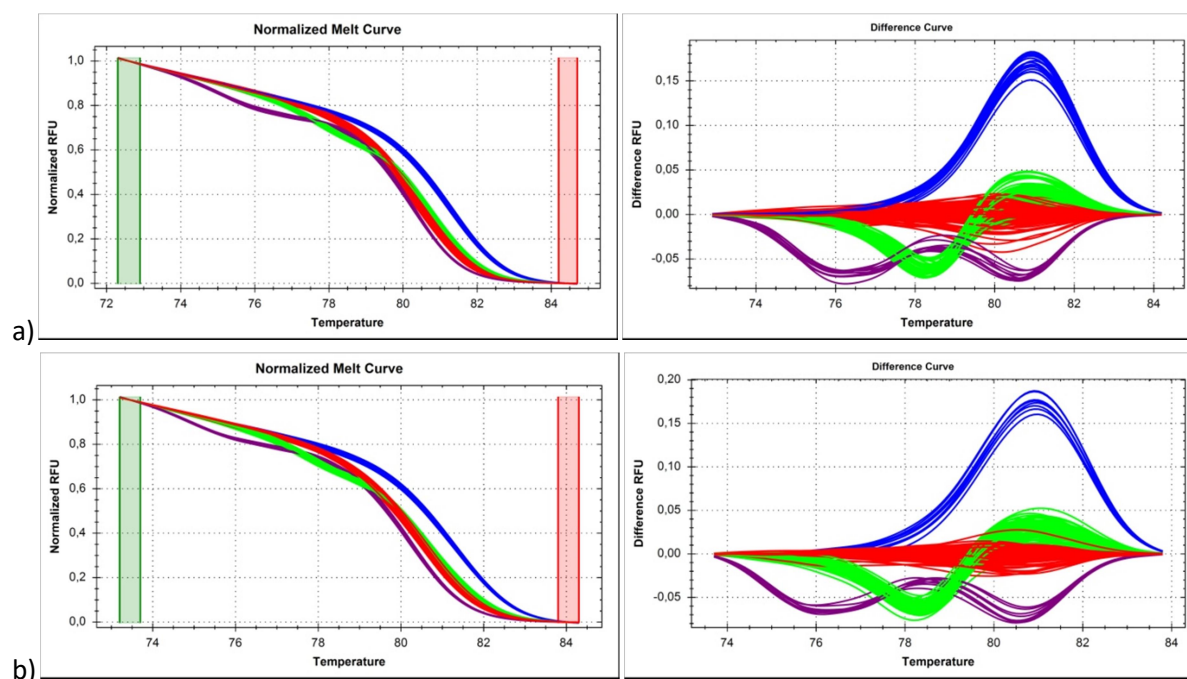
Table S51 Summary of *ABCC1* scanning by HRM method. The average scan rate for the all the *ABCC1* scanned areas and the number of total sequenced samples are on the bottom of the table.

Scanned exon	Area of the exon	Clustered samples	Excluded samples	Sequenced samples	Scan rate
2	1	189	1	6	0,9947
2	2	189	1	9	0,9947
3	whole	188	2	6	0,9895
4	1	186	4	14	0,9789
4	2	189	1	11	0,9947
4	3	188	2	6	0,9895
5	whole	187	3	6	0,9842
6	whole	188	2	6	0,9895
7	1	189	1	11	0,9947
7	2	188	2	25	0,9895
8	1	188	2	19	0,9895
8	2	189	1	6	0,9947
9	1	185	5	12	0,9737
9	2	189	1	12	0,9947
10	1	188	2	8	0,9895
10	2	187	3	5	0,9842
11	whole	188	2	7	0,9895
12	1	189	1	6	0,9947
12	2	189	1	8	0,9947
13	1	189	1	18	0,9947
13	2	185	5	7	0,9737
14	whole	188	2	6	0,9895
15	whole	189	1	6	0,9947
16	1	188	2	11	0,9895
16	2	189	1	6	0,9947
17	1	190	0	10	1
17	2	189	1	6	0,9947
18	1	188	2	6	0,9895
18	2	188	2	6	0,9895
19	1	188	2	20	0,9895
19	2	186	4	6	0,9789
19	3	189	1	6	0,9947
20	whole	188	2	6	0,9895
21	whole	189	1	6	0,9947
22	1	188	2	7	0,9895
22	2	188	2	13	0,9895
23	1	189	1	9	0,9947
23	2	186	4	6	0,9789
24	1	189	1	6	0,9947
24	2	188	2	6	0,9895
25	whole	188	2	6	0,9895
26	whole	188	2	6	0,9895
27	whole	188	2	7	0,9895
28	1	188	2	12	0,9895
28	2	188	2	6	0,9895
29	1	189	1	12	0,9947

29	2	188	2	6	0,9895
30	1	189	1	6	0,9947
30	2	188	2	11	0,9895
31	whole	187	3	8	0,9842
Overall		9405	95	437	0,9900

Table S52 Summary of *ABCC1* c.816G>A (p.Pro272=, rs2230669) and c.825T>C (p.Val275=, rs60782127) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	4	4	4	4	-
Samples in cluster 1 (blue – GG/CC variant)	9	4	5	5	23
Samples in cluster 2 (green – GG/CT variant)	27	40	38	37	142
Samples in cluster 3 (red – GG/TT variant)	54	47	48	48	197
Samples in cluster 4 (violet – GA/CT variant)	5	4	4	5	18
Samples excluded	0	0	0	0	0
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	1	1	1	4
Samples verified by sequencing from cluster 4	1	1	1	1	4
Call rate	-	-	-	-	1



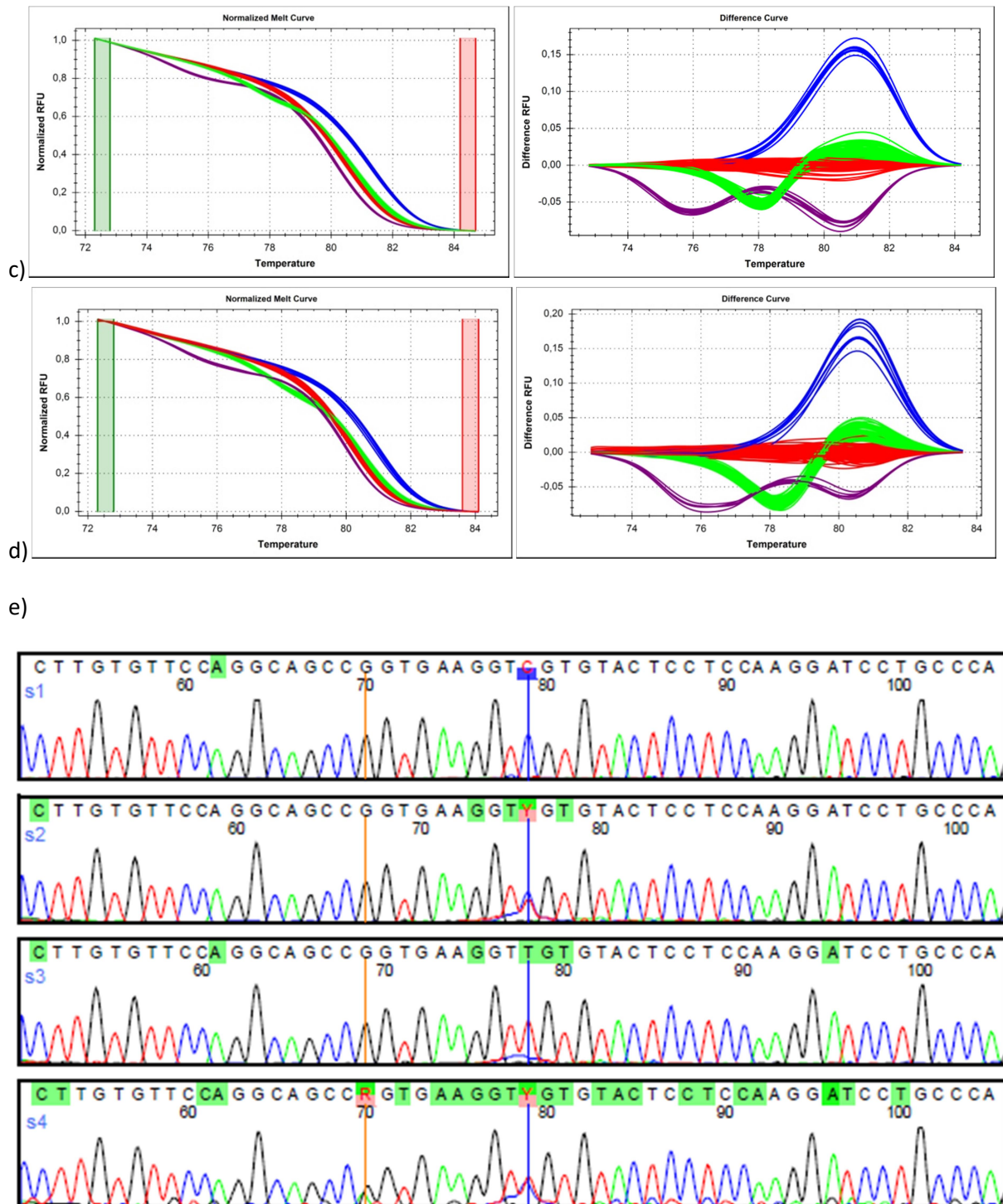
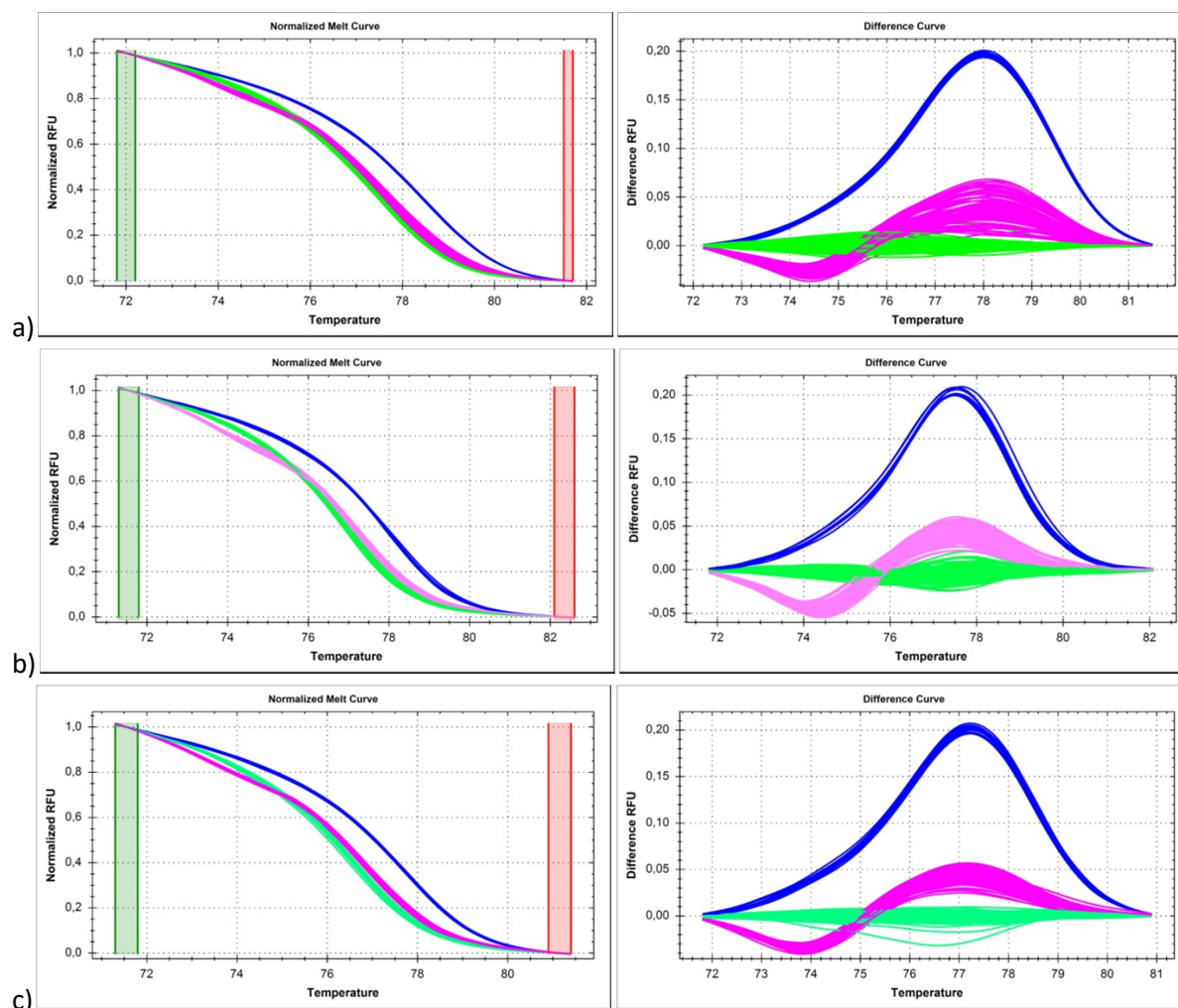


Figure S51 Summary of *ABCC1* c.816G>A (p.Pro272=, rs2230669) and c.825T>C (p.Val275=, rs60782127) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 51 bp including primers, 14 bp without primers), blue – cluster 1 (GG/CC), green – cluster 2 (GG/CT), red – cluster 3 (GG/TT), violet – cluster 4 (GA/CT); e – Example of the sequencing traces for the samples from four different clusters: s1 – cluster 1 (GG/CC), s2 – cluster 2 (GG/CT), s3 – cluster 3 (GG/TT); s4 – cluster 4 (GA/CT), red and blue lines indicate c.816G>A and c.825T>C genotype, respectively.

Table S53 Summary of *ABCC1* c.1062T>C (p.Asn354=, rs35587) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	3	3	3	3	-
Samples in cluster 1 (blue – CC variant)	8	5	12	8	33
Samples in cluster 2 (pink – CT variant)	35	36	37	34	142
Samples in cluster 3 (green – TT variant)	51	54	46	53	204
Samples excluded	1	0	0	0	1
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	1	1	1	4
Call rate	-	-	-	-	0,9974



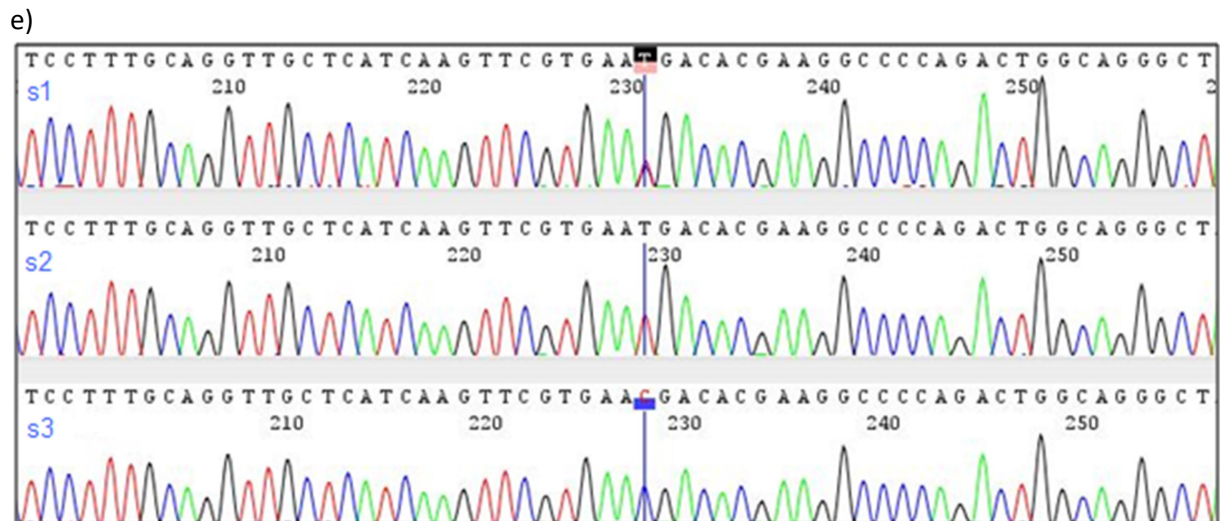
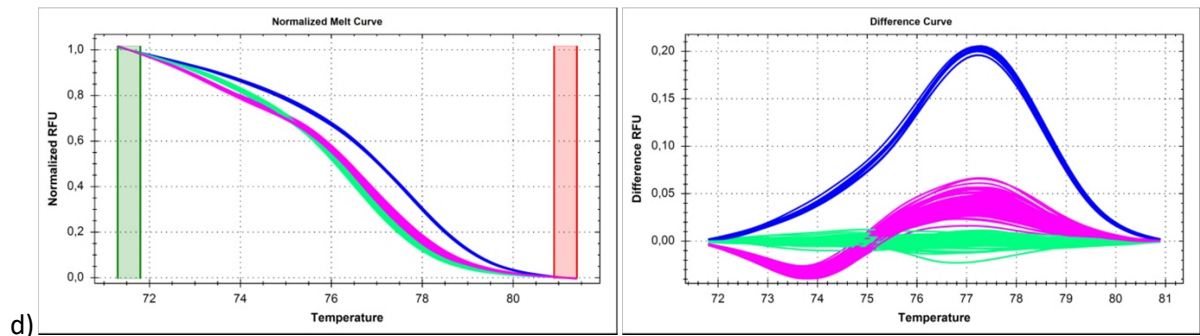
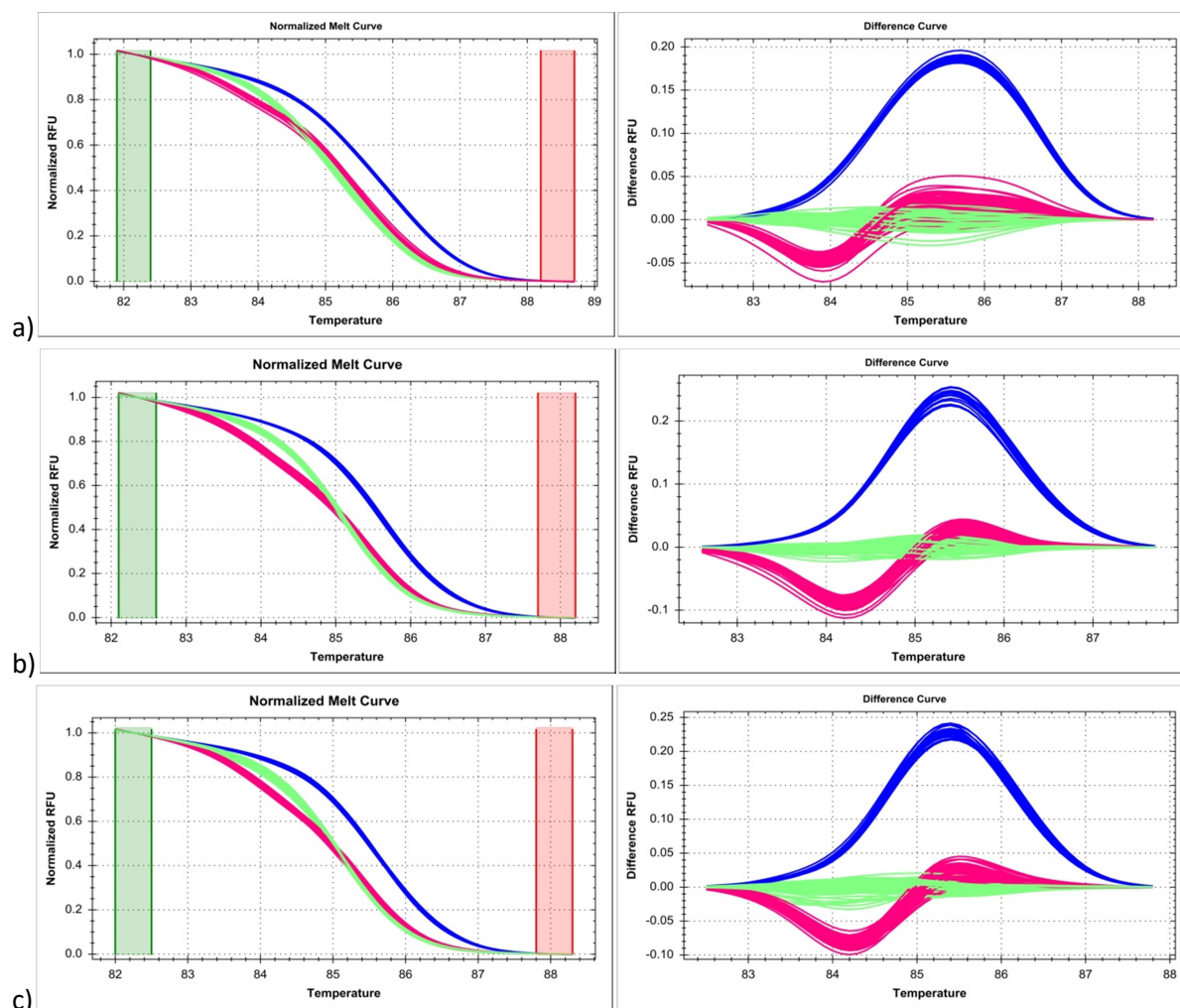


Figure S52 Summary of *ABCC1* c.1062T>C (p.Asn354=, rs35587) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 41 bp including primers, 2 bp without primers), blue – cluster 1 (CC), pink – cluster 2 (CT), green – cluster 3 (TT); e – Example of the sequencing traces for the samples from three different clusters: s1 – cluster 2 (CT), s2 – cluster 3 (TT), s3 – cluster 1 (CC), blue lines indicate c.1062T>C genotype.

Table S54 Summary of *ABCC1* c.1218+8A>G (intron, rs35588) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	3	3	3	3	-
Samples in cluster 1 (blue – GG variant)	8	10	6	6	30
Samples in cluster 2 (pink – AG variant)	39	41	40	40	160
Samples in cluster 3 (green – AA variant)	48	44	49	49	190
Samples excluded	0	0	0	0	0
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	1	1	1	4
Call rate	-	-	-	-	1



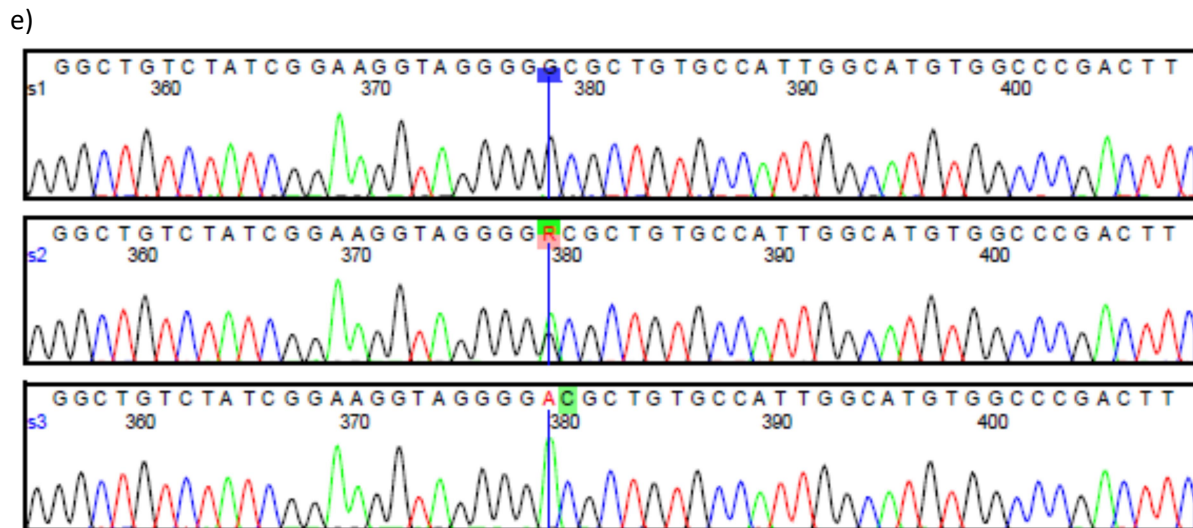
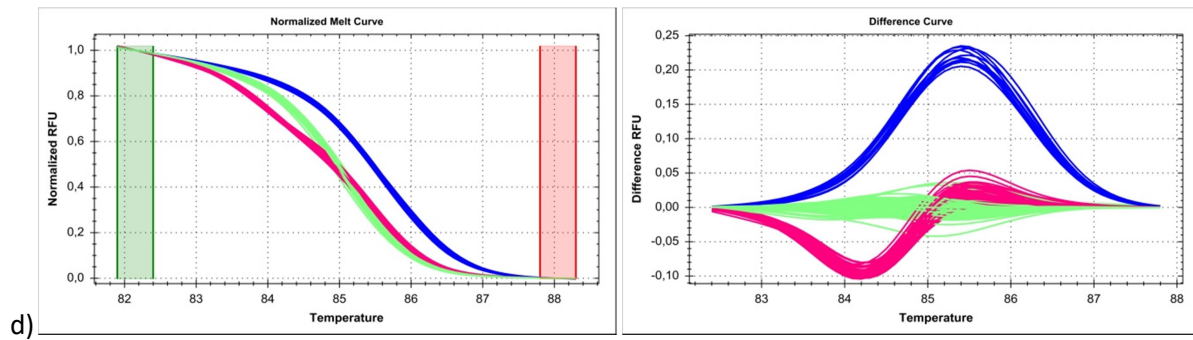
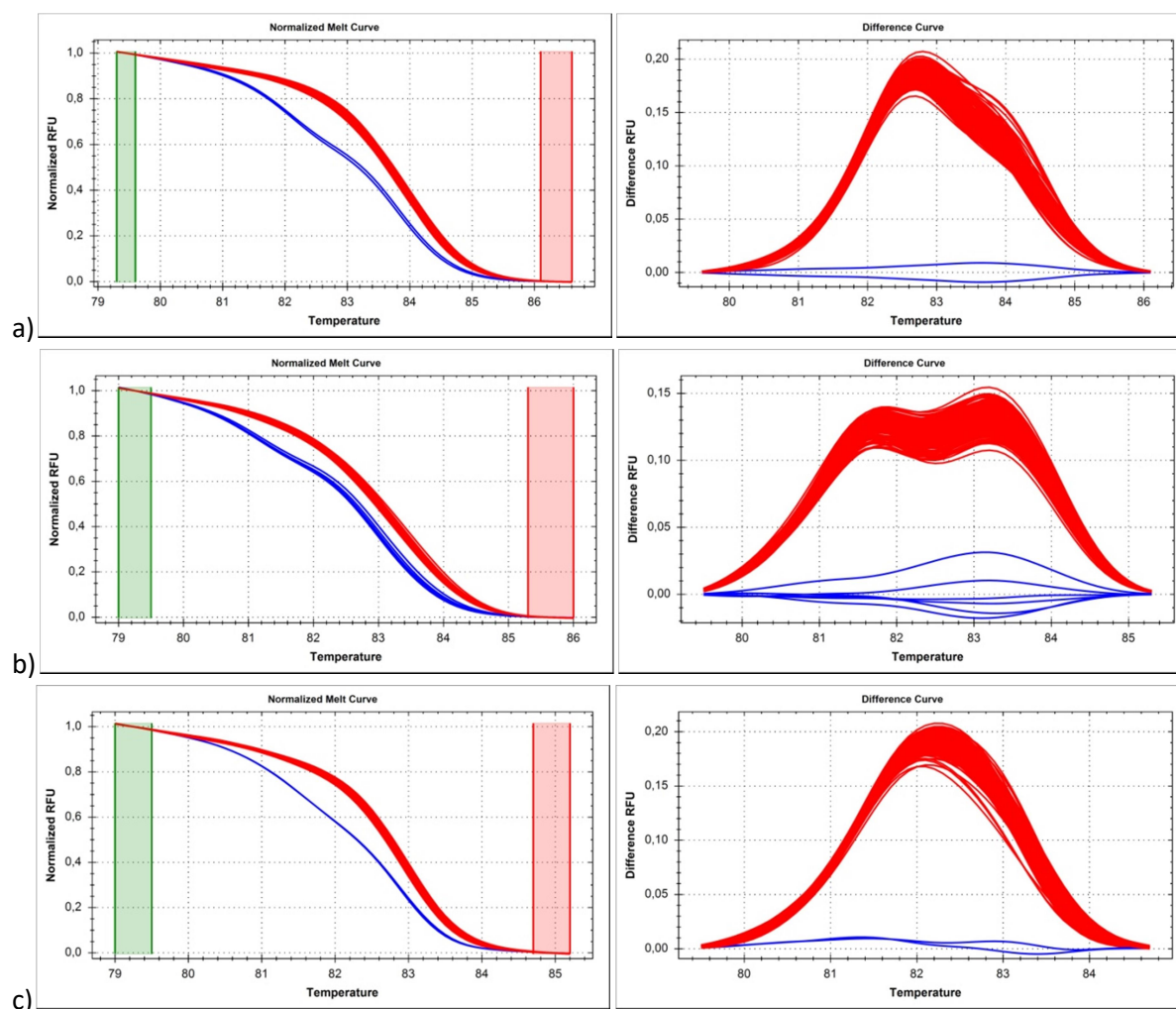


Figure S53 Summary of *ABCC1* c.1218+8A>G (intron, rs35588) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 84 bp including primers, 46 bp without primers), blue – cluster 1 (GG), pink – cluster 2 (AG), green – cluster 3 (AA); e – Example of the sequencing traces for the samples from three different clusters: s1 – cluster 1 (GG), s2 – cluster 2 (AG), s3 – cluster 3 (AA), blue lines indicate c.1218+8A>G genotype.

Table S55 Summary of *ABCC1* c.1299G>T (p.Arg433Ser, rs60782127) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	2	2	2	2	-
Samples in cluster 1 (red – GG variant)	94	92	94	93	373
Samples in cluster 2 (blue – GT variant)	1	3	1	1	6
Samples excluded	0	0	0	1	0
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	3	1	1	6
Call rate	-	-	-	-	0,9974



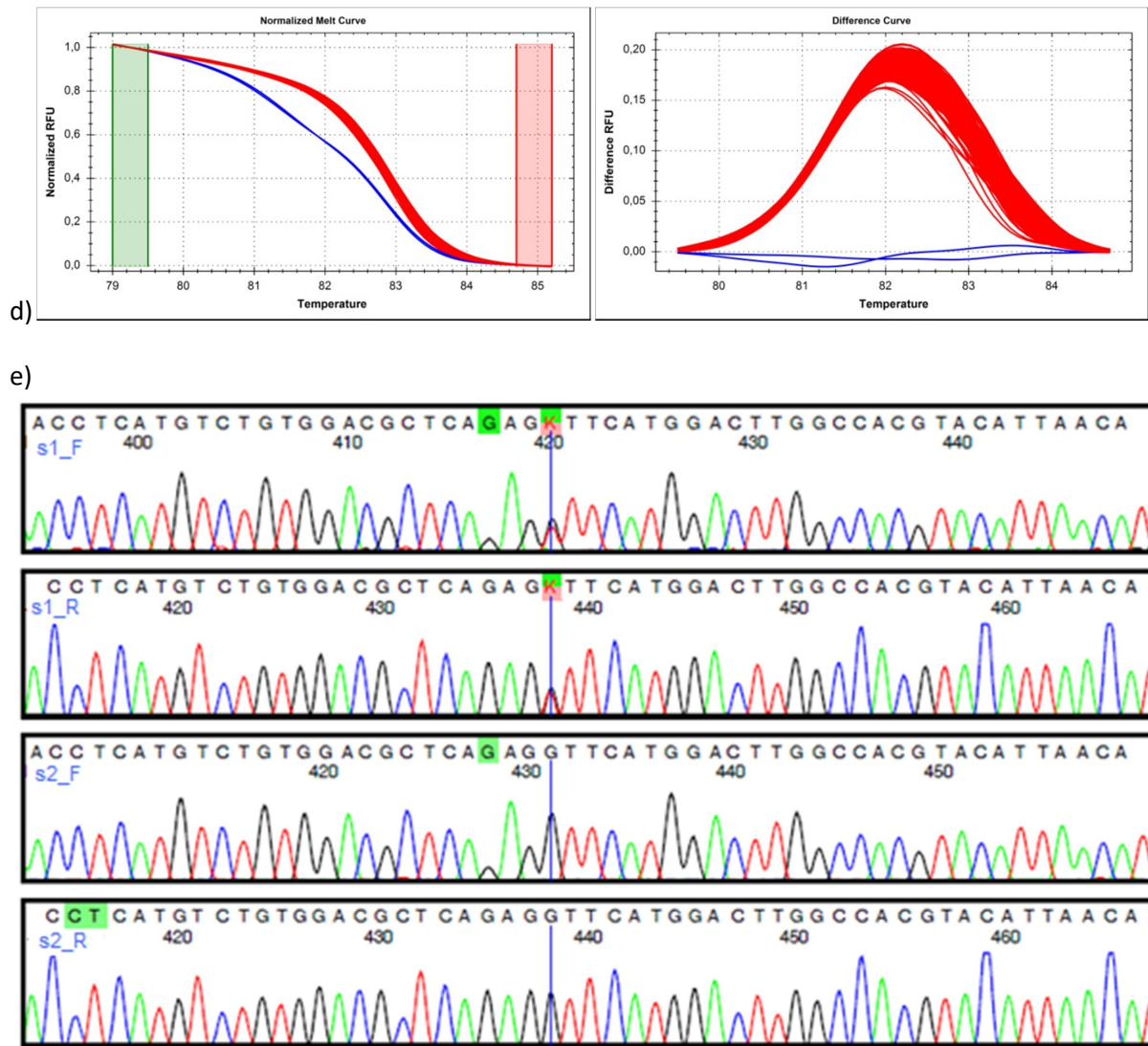
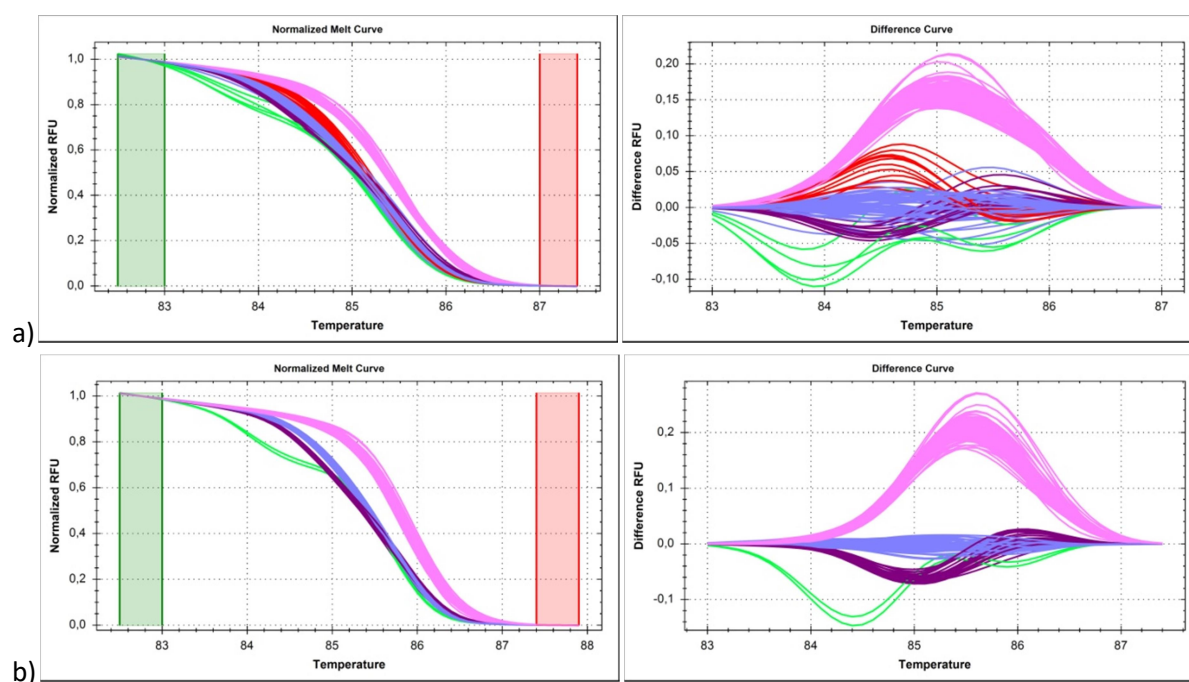


Figure S54 Summary of *ABCC1* c.1299G>T (p.Arg433Ser, rs60782127) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 80 bp including primers, 41 bp without primers), red – cluster 1 (GG), blue – cluster 2 (GT) ; e – Example of the sequencing traces for the samples from three different clusters: s1_F&R – cluster 2 (GT), s2_F&R – cluster 1 (GG), blue lines indicate c.1299G>T genotype.

Table S56 Summary of *ABCC1* c.1684T>C (p.Leu562=, rs35605) and c.1704C>T (p.Tyr568=, rs8187858) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	5	4	5	5	-
Samples in cluster 1 (pink – CC/CC variant)	55	54	51	53	213
Samples in cluster 2 (violet – CC/CT variant)	5	12	16	16	49
Samples in cluster 3 (blue – CT/CC variant)	27	28	19	25	99
Samples in cluster 4 (green– CT/CT variant)	2	1	2	1	6
Samples in cluster 5 (red– TT/CC variant)	5	0	7	1	13
Samples excluded	0	0	0	0	0
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	1	1	1	4
Samples verified by sequencing from cluster 4	2	1	2	1	6
Samples verified by sequencing from cluster 5	2	0	2	1	5
Call rate	-	-	-	-	1



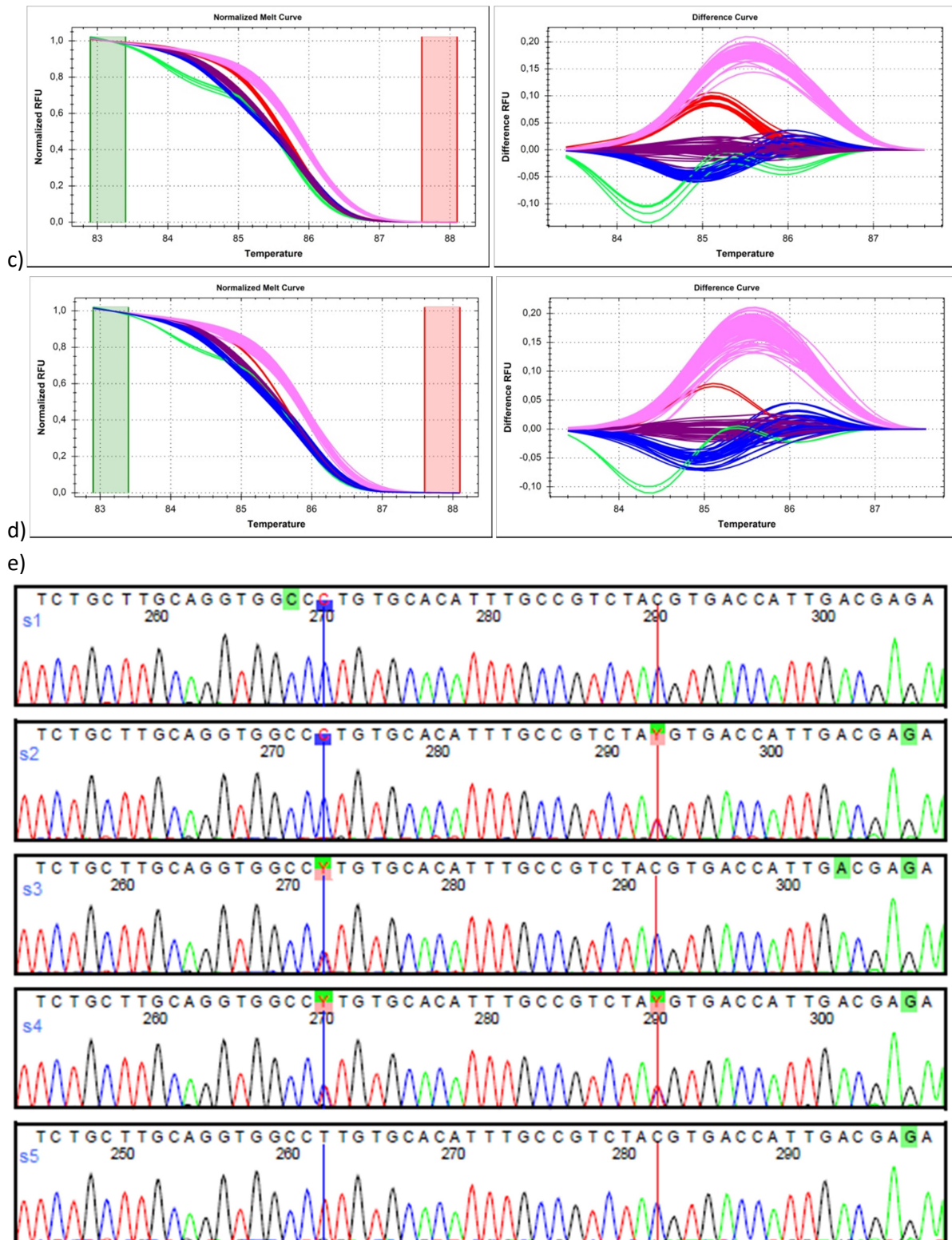
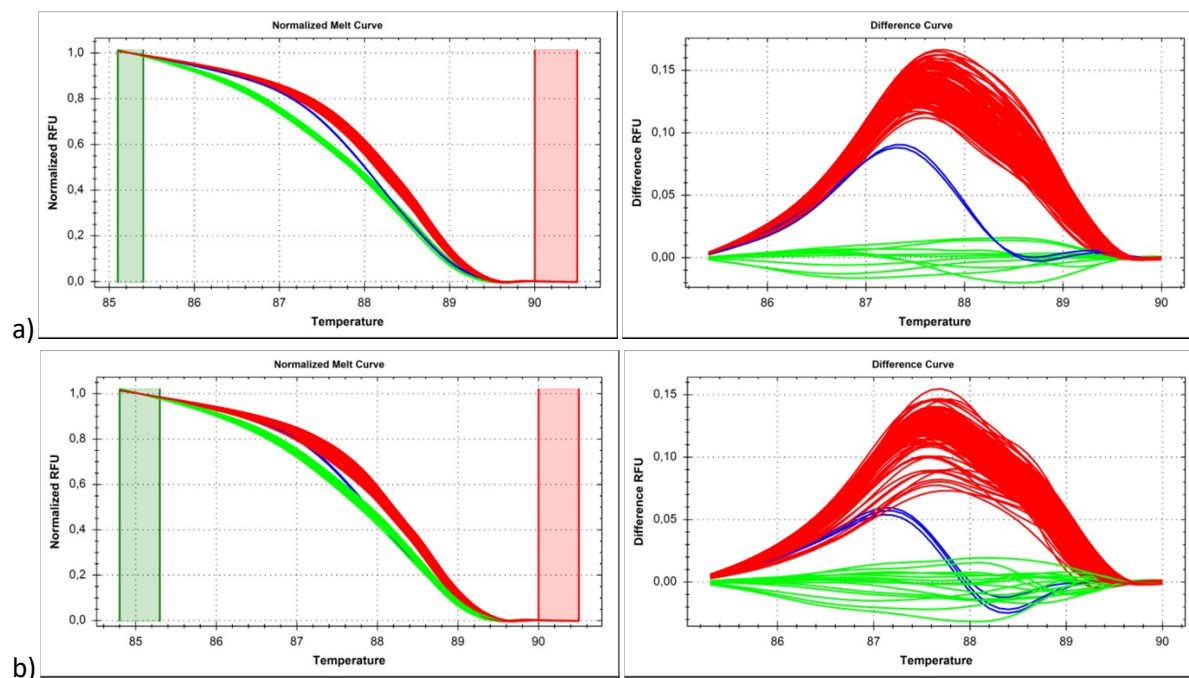


Figure S55 Summary of *ABCC1* c.1684T>C (p.Leu562=, rs35605) and c.1704C>T (p.Tyr568=, rs8187858) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 125 bp including primers, 89 bp without primers), pink – cluster 1 (CC/CC), violet – cluster 2 (CC/CT), blue – cluster 3 (CT/CC), green – cluster 4 (CT/CT), red – cluster 5 (TT/CC); e – Example of the sequencing traces for the samples from five different clusters:

s1 – cluster 1 (CC/CC), s2 – cluster 2 (CC/CT), s3 – cluster 3 (CT/CC), s4 – cluster 4 (CT/CT), s5 – cluster 5 (TT/CC), blue and red lines indicate c.1684T>C and c.1704C>T genotype, respectively.

Table S57 Summary of *ABCC1* c.2012G>T (p.Gly671Val, rs45511401) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	3	3	3	3	-
Samples in cluster 1 (red – GG variant)	88	85	90	83	346
Samples in cluster 2 (green – GT variant)	5	8	5	10	28
Samples in cluster 3 (blue – TT variant)	1	2	0	0	3
Samples excluded	1	0	0	2	3
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	2	0	0	3
Call rate	-	-	-	-	0,9921



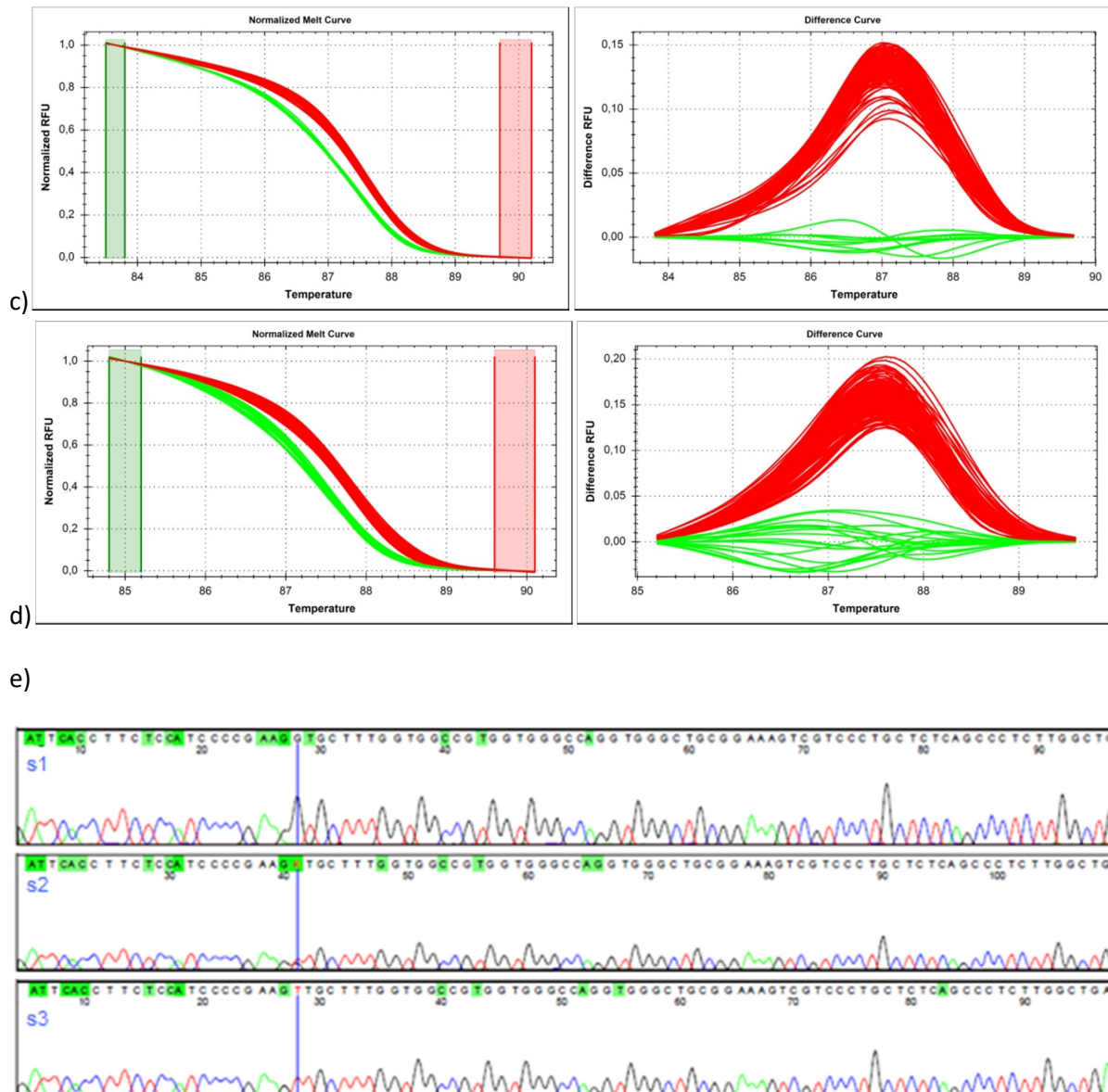
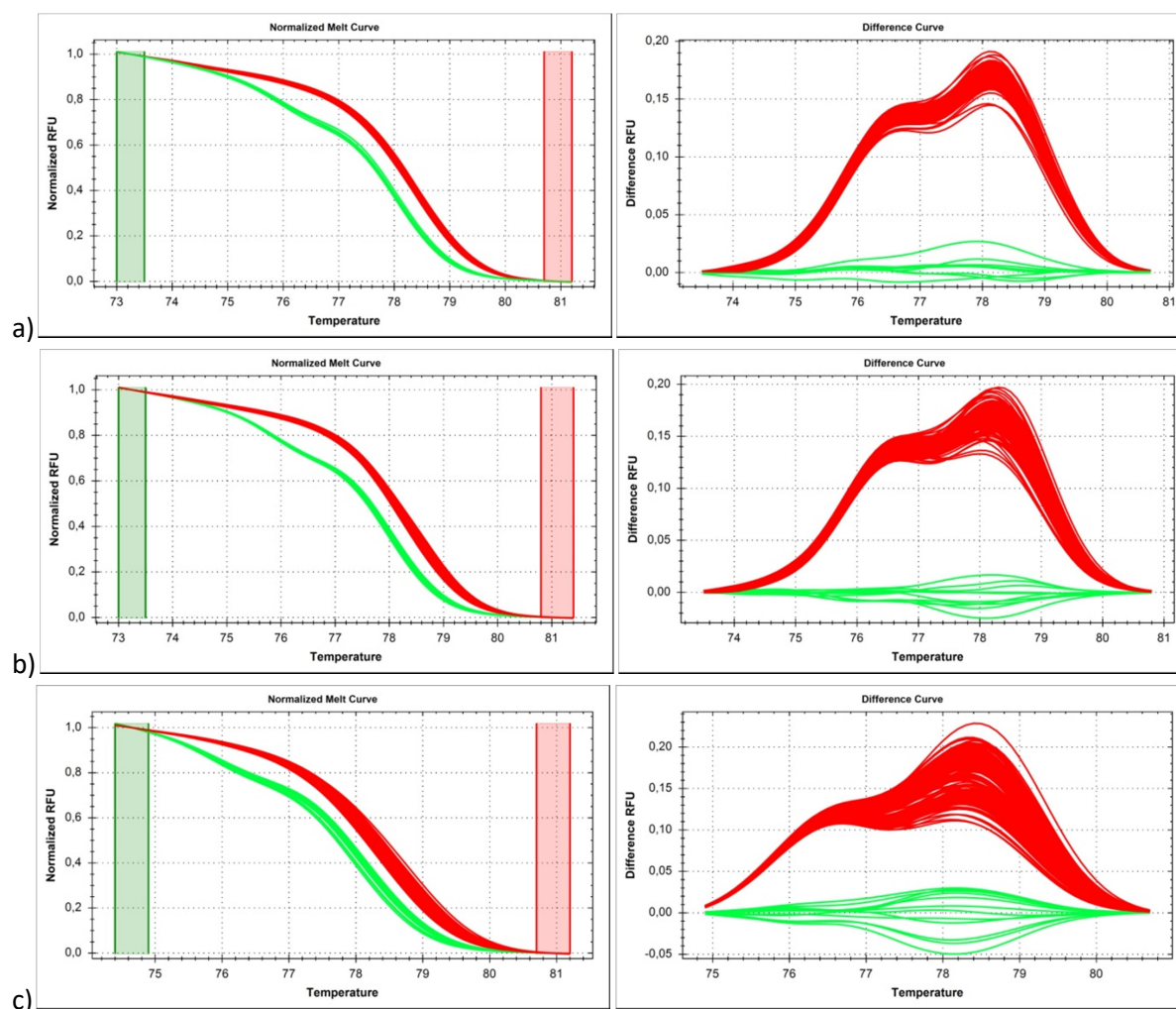


Figure S56 Summary of *ABCC1* c.2012G>T (p.Gly671Val, rs45511401) genotyping by HRM: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 104 bp including primers, 66 bp without primers); e – Example of the sequencing traces for the samples from the three different clusters: s1 – cluster 1 (GG), s2 – cluster 2 (GT); s3 – cluster3 (TT), blue lines indicate c.2012G>T genotype.

Table S58 Summary of *ABCC1* c.2168G>A (p.Arg723Gln, rs4148356) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	2	2	2	2	-
Samples in cluster 1 (red – GG variant)	91	89	90	90	360
Samples in cluster 2 (green – GA variant)	4	6	4	5	19
Samples excluded	0	0	1	0	1
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Call rate	-	-	-	-	0,9974



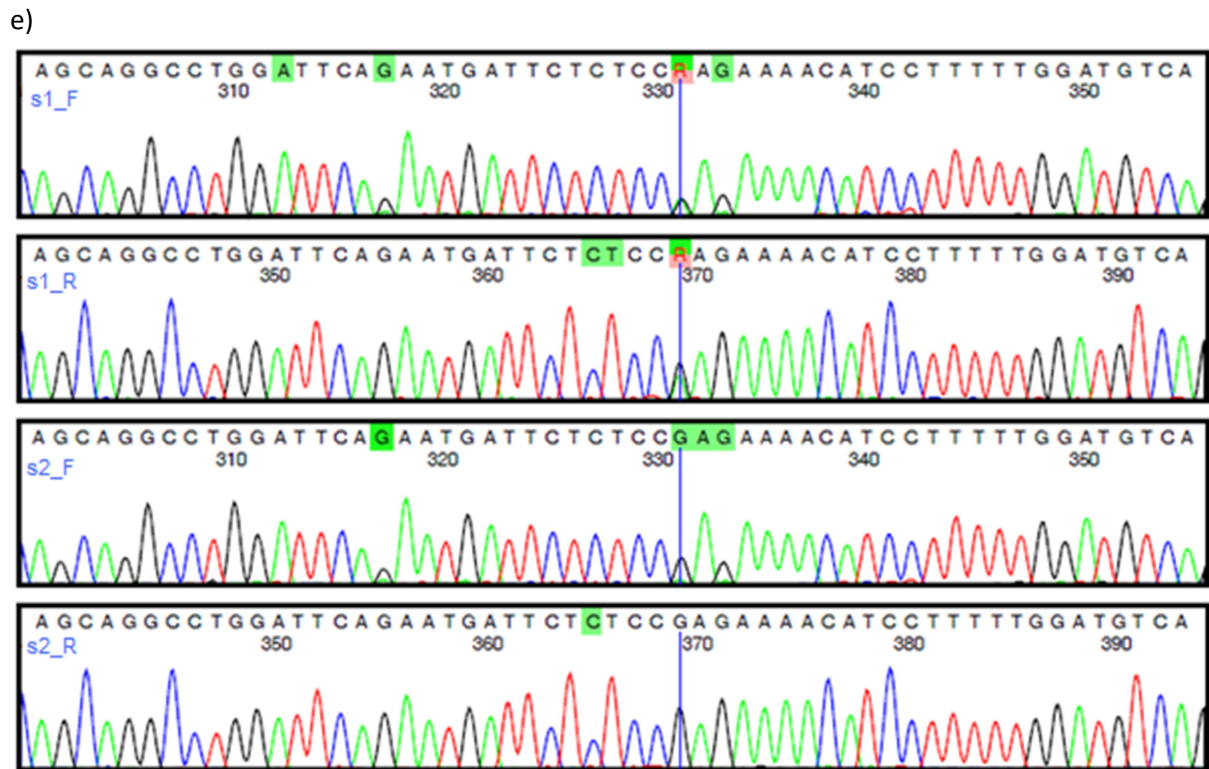
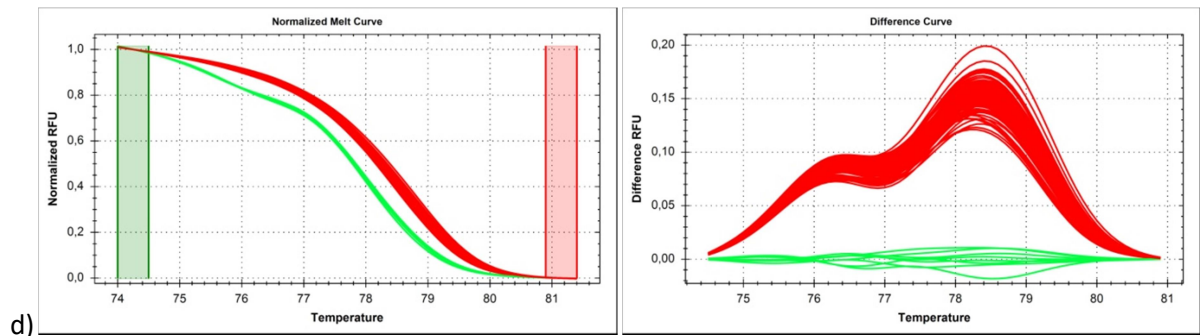
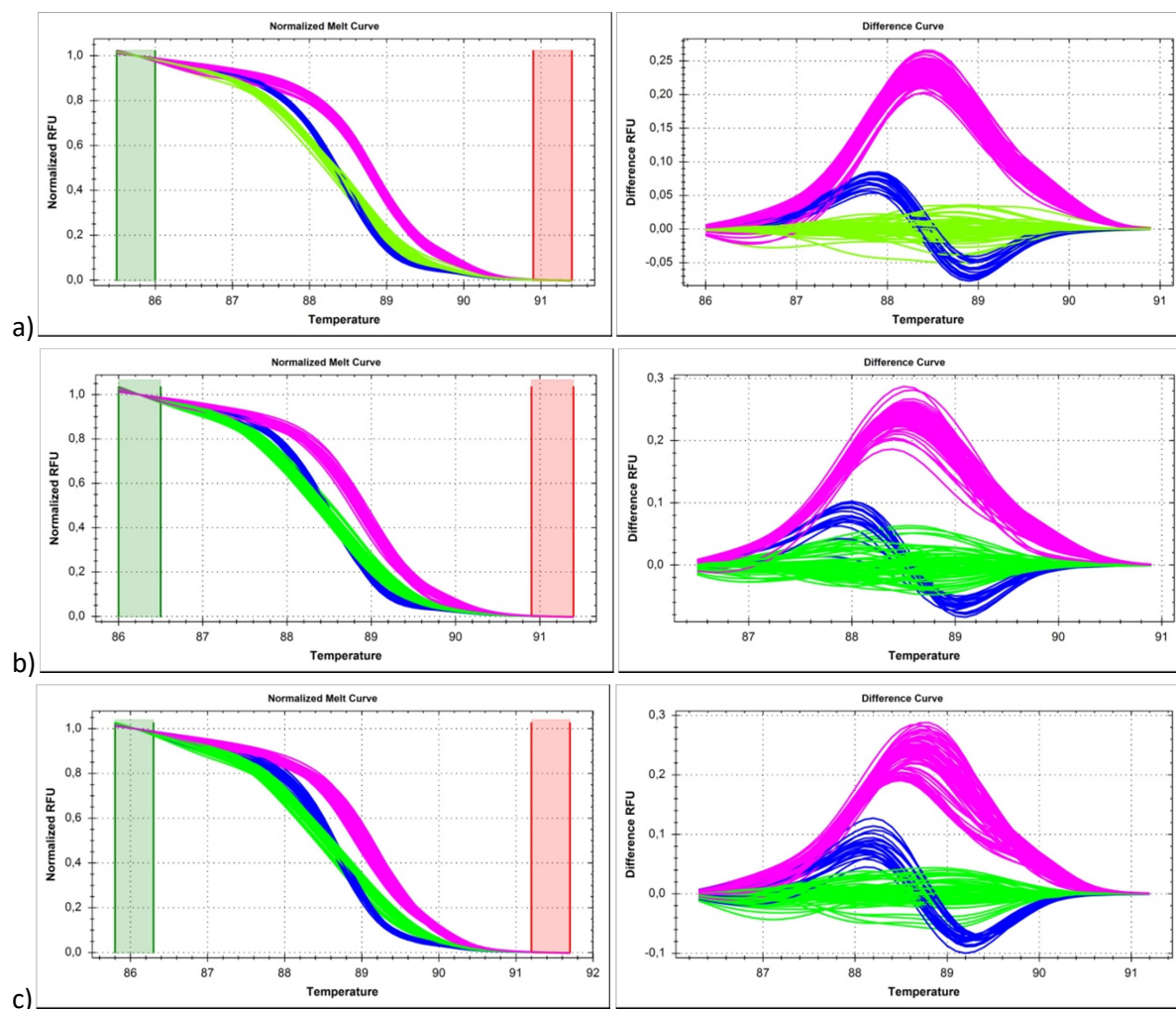


Figure S57 Summary of *ABCC1* c.2168G>A (p.Arg723Gln, rs4148356) genotyping by HRM.: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 67 bp including primers, 27 bp without primers, red – cluster 1 (GG), green – cluster 2 (GA); e – Example of the sequencing traces for the samples from the two different clusters: s1_F&R – cluster 2 (GA), s2_F&R – cluster 1 (GG), blue lines indicate c.2168G>A genotype.

Table S59 Summary of *ABCC1* c.4002G>A (p.Ser1334=, rs2230671) genotyping by HRM. Call rate is the ratio between positive clustered to the all genotyped samples.

	Plate 1	Plate 2	Plate 3	Plate 4	Overall
Samples genotyped	95	95	95	95	380
Melting clusters	3	3	3	3	-
Samples in cluster 1 (pink – GG variant)	60	48	52	44	204
Samples in cluster 2 (green – GA variant)	26	37	39	38	140
Samples in cluster 3 (blue – AA variant)	9	8	4	11	32
Samples excluded	0	2	0	2	4
Samples verified by sequencing from cluster 1	1	1	1	1	4
Samples verified by sequencing from cluster 2	1	1	1	1	4
Samples verified by sequencing from cluster 3	1	1	1	1	4
Call rate	-	-	-	-	0,9895



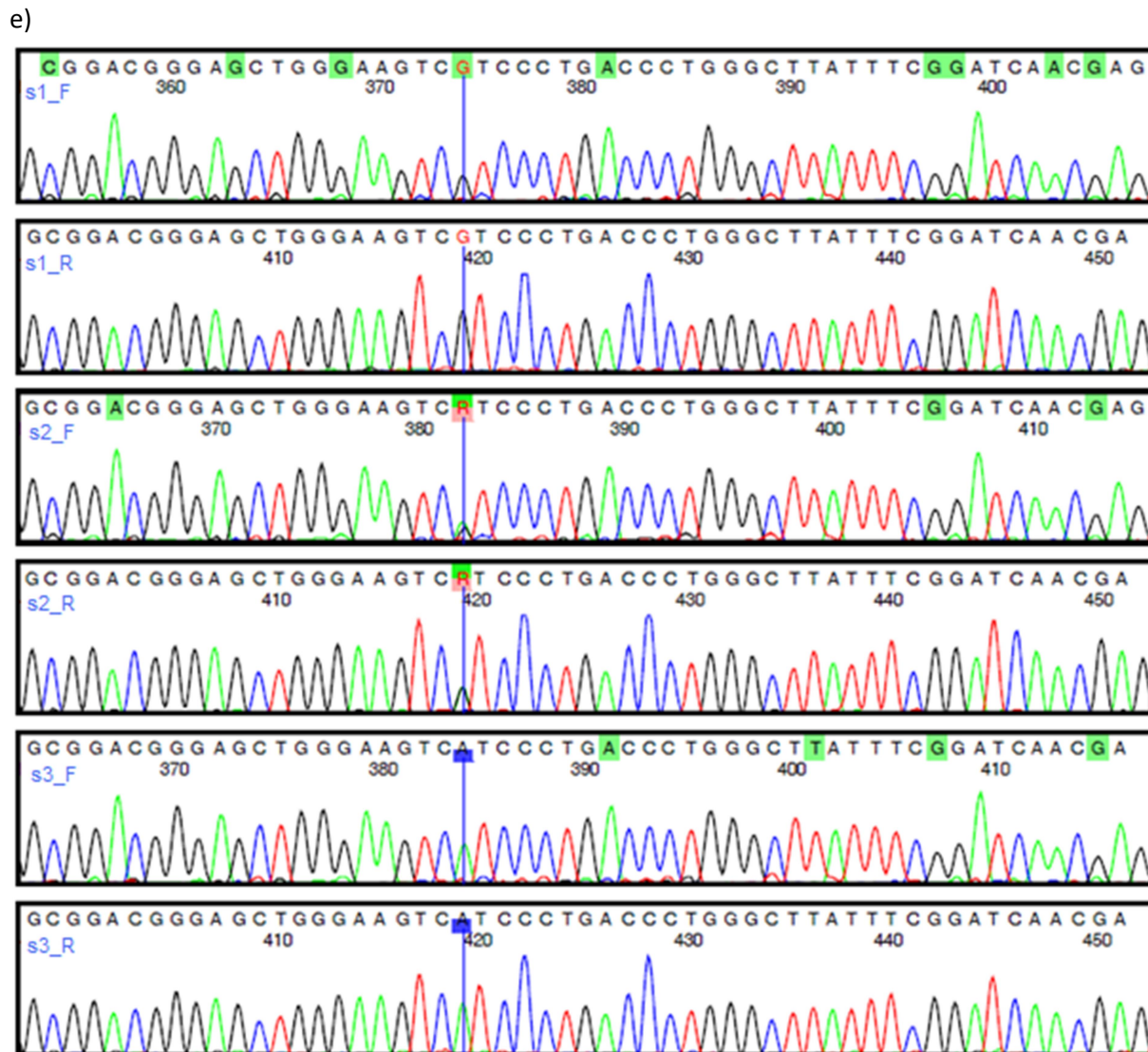
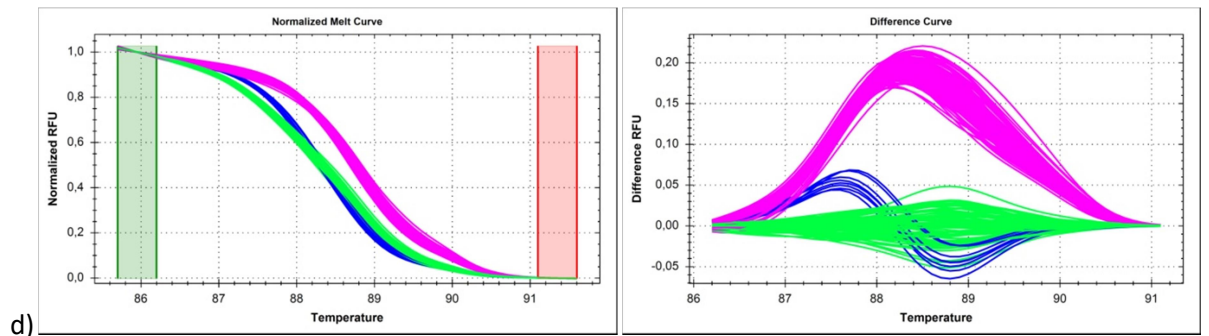


Figure S58 Summary of *ABCC1* c.4002G>A (p.Ser1334=, rs2230671) genotyping by HRM.: a,b,c,d – High resolution melting profiles of four different DNA plates (analysed PCR product length: 112 bp including primers, 76 bp without primers), pink - cluster 1 (GG), green – cluster 2 (GA), blue – cluster 3 (AA); e – Example of the sequencing traces for the samples from the three different clusters: s1_F&R – cluster 1 (GG), s2_F&R – cluster 2 (GA), s3_F&R – cluster 3 (AA), blue lines indicate c.4002G>A genotype.

Table S60 Overall summary of HRM genotyping method

dbSNP ID	Variant residue NM_004996.3:	Intron/amino acid residue NP_004987.2:	Clustered samples	Sequenced samples	Excluded samples	All samples	Genotyping rate
rs2230669	c.816G>A	p.Pro272=	380	16	0	380	1
rs246221	c.825T>C	p.Val275					
rs35587	c.1062T>C	p.Asn354=	379	12	1	380	0,9974
rs35588	c.1218+8A>G	Intron	380	12	0	380	1
rs60782127	c.1299G>T	p.Arg433Ser	379	10	1	380	0,9974
rs35605	c.1684T>C	p.Leu562=	380	23	0	380	1
rs8187858	c.1704C>T	p.Tyr568=					
rs45511401	c.2012G>T	p.Gly671Val	377	11	3	380	0,9921
rs4148356	c.2168G>A	p.Arg723Gln	379	8	1	380	0,9974
rs2230671	c.4002G>A	p.Ser1334=	376	12	4	380	0,9895
Overall summary			3030	104	10	3040	---
Overall genotyping rate			---	---	---	---	0,9967

Table S61 Nonsynonymous variants detected in this study and their predicted effect on protein functioning classified by PolyPhen-2 tool as: benign influence, possibly or probably damaging

Nonsynonymous variant	HumDiv value	HumDiv result	HumVar value	HumVar result
c.596C>T (p.Ser199Leu)	0.982	probably damaging	0.520	possibly damaging
c.814C>T (p.Pro272Ser)	0.000	benign	0.001	benign
c.1299G>T (p.Arg433Ser)	1.000	probably damaging	0.999	probably damaging
c.1898G>A (p.Arg633Gln)	0.001	benign	0.002	benign
c.2012G>T (p.Gly671Val)	1.000	probably damaging	0.992	probably damaging
c.2168G>A (p.Arg723Gln)	0.014	benign	0.013	benign
c.2876A>G (p.Lys959Arg)	0.001	benign	0.005	benign
c.3196C>T (p.Arg1066Trp)	0.786	possibly damaging	0.343	benign
c.3886C>T (p.Arg1296Trp)	0.995	probably damaging	0.816	possibly damaging
c.3901C>T (p.Arg1301Cys)	1.000	probably damaging	0.999	probably damaging
c.4093G>A (p.Asp1365Asn)	0.003	benign	0.013	benign

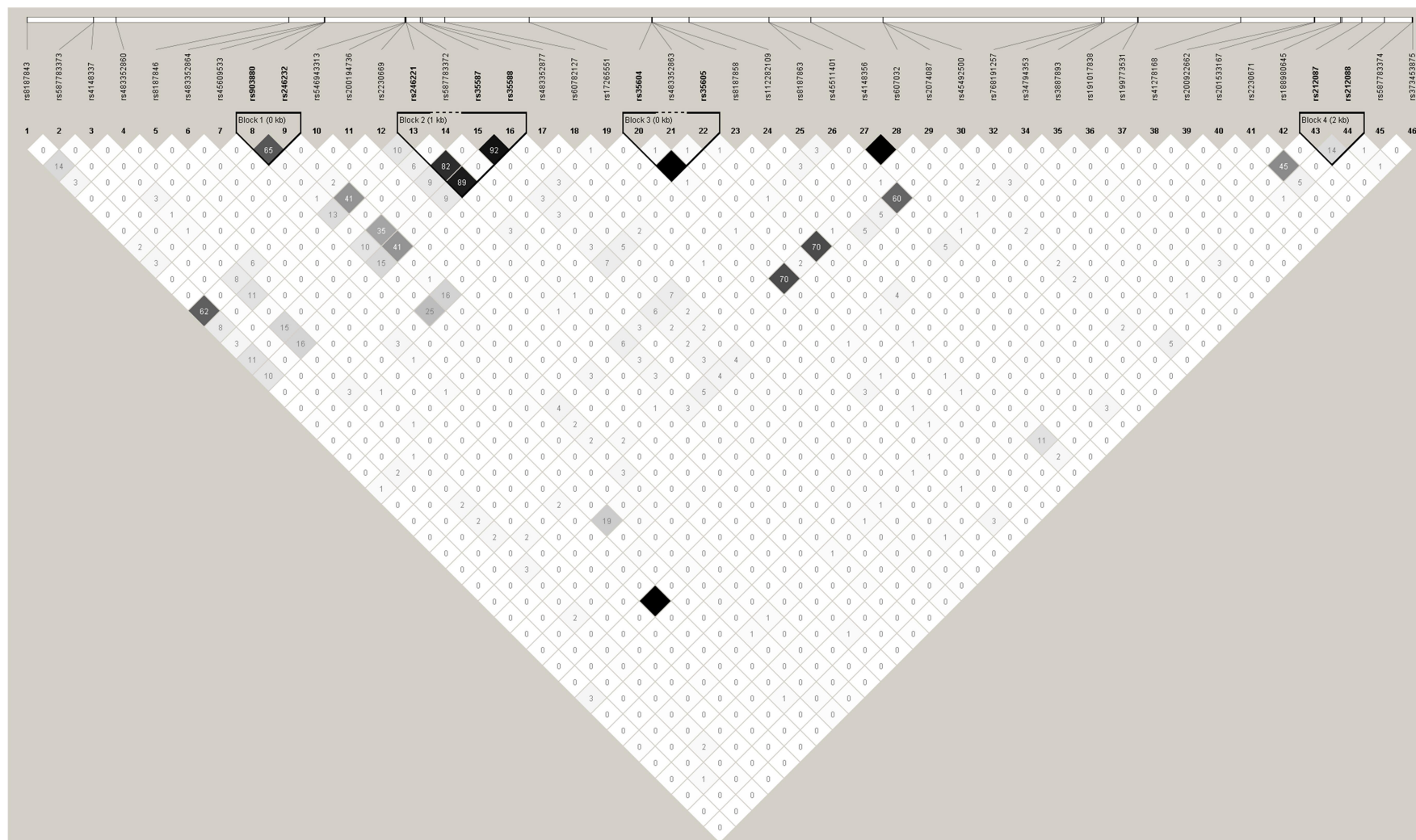


Figure S59 Linkage disequilibrium analysis of ABCC1 SNPs detected in this study using the r^2 parameter. Pairwise linkage shown as r^2 values (x100) by graded grey colors; darker color means closer linkage and lack of value means $r^2=1$. Triangles with bold margins show haplotype blocks.

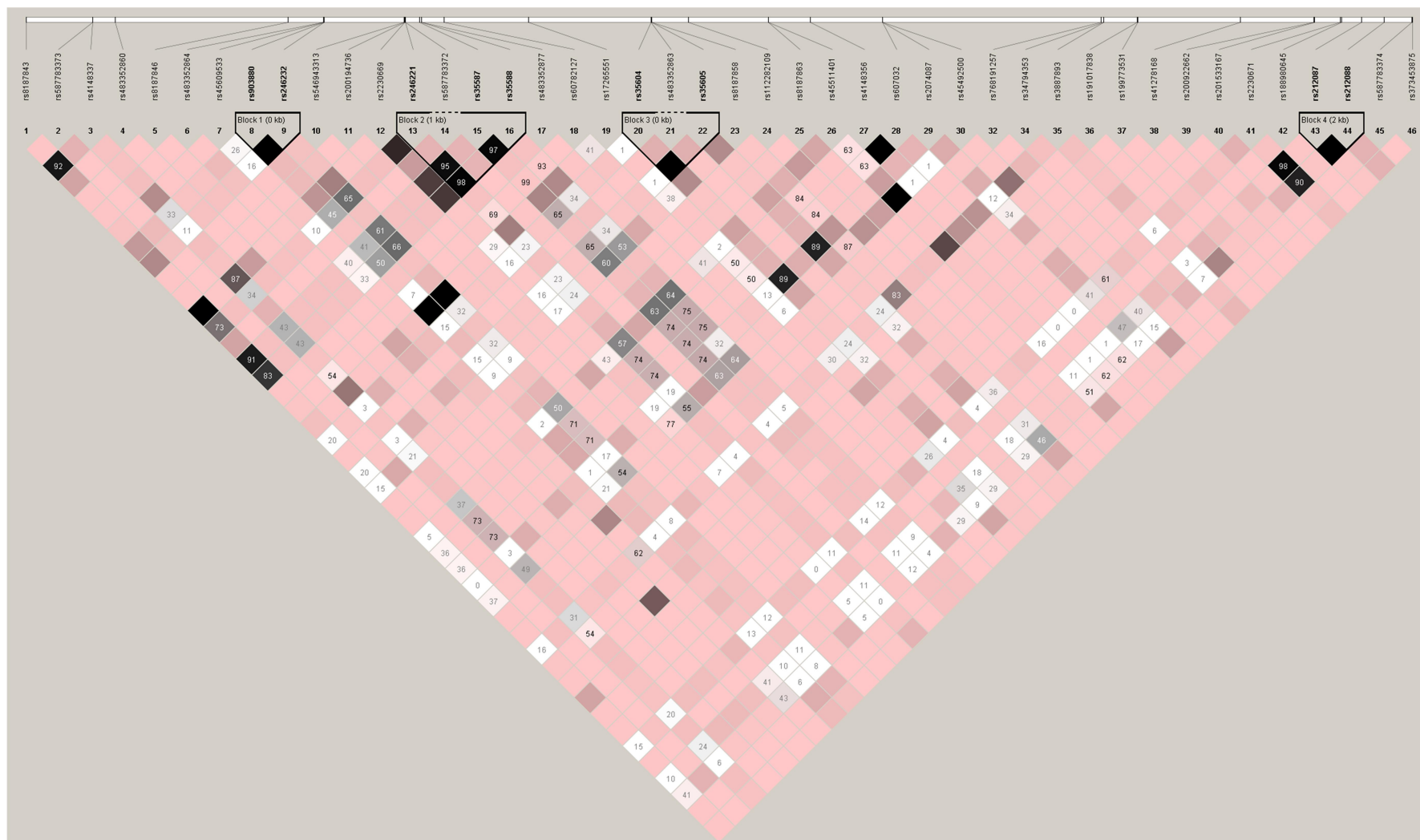


Figure S60 Linkage disequilibrium analysis of ABCC1 SNPs detected in this study using $|D'|$ parameter. Pairwise linkage shown as $|D'|$ values (x100) by graded pink and black colors, darker color means closer linkage and lack of value means $|D'|=1 \times 100$. Triangles with bold margins show haplotype blocks.

ZAŁĄCZNIK NR 2

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G., Strapagiel, D., 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. *Genes (Basel)*, 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

Article

Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison

Marcin Słomka ^{1,2}, Marta Sobalska-Kwapis ^{1,2}, Małgorzata Korycka-Machala ³,
Jarosław Dziadek ³, Grzegorz Bartosz ⁴ and Dominik Strapagiel ^{1,2,*}

¹ Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Pilarskiego 14/16, 90-231 Łódź, Poland; marcin.slomka@biol.uni.lodz.pl (M.S.); marta.sobalska@biol.uni.lodz.pl (M.S.-K.)

² BBMRI.pl Consortium, 54-066 Wrocław, Poland

³ Institute of Medical Biology, Polish Academy of Sciences, Lodowa 106, 93-232 Łódź, Poland; mkorycka@cbm.pan.pl (M.K.-M.); jdziadek@cbm.pan.pl (J.D.)

⁴ Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Pomorska 141/143, 90-236 Łódź, Poland; grzegorz.bartosz@biol.uni.lodz.pl

* Correspondence: dominik.strapagiel@biol.uni.lodz.pl; Tel.: +48-42-6655702; Fax: +48-42-6354070

Received: 17 August 2020; Accepted: 26 September 2020; Published: 29 September 2020



Abstract: ATP-binding cassette sub-family G member 2 (*ABCG2*), also known as breast cancer resistance protein (BCRP), is one of the key efflux ATP-binding cassette (ABC) transporters of xenobiotics, their metabolites and endogenous compounds such as urate. Some of its genetic variants have been found to influence protein functioning, resulting in serious clinical implications concerning chemotherapy response, as well as gout or blood group phenotype Jr(a-). Previous reports have suggested that the frequencies of certain crucial polymorphisms, such as c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys) differ significantly between the Polish population and other Caucasian populations. Thus, to clarify this issue, the present study performs a complete analysis of the genetic variation of *ABCG2* coding sequence in the Polish population. The genetic variation in 14 out of 15 coding exons of the *ABCG2* gene, as well as their flanking intron sequences, were examined among 190 healthy representatives of the Polish population using scanning with High Resolution Melting (HRM). HRM scanning revealed 17 polymorphisms: eight in the exons (including five missense variants and one point-nonsense mutation) and nine in the intron sequences (eight single nucleotide polymorphisms (SNPs) and one deletion variant). These included variants correlating with the presence of gout and phenotype Jr(a-). Linkage disequilibrium, haplotype blocks and haplotype analyses were also performed. The frequencies of the most common polymorphisms in the Polish population did not differ significantly to those observed for other Caucasian populations, but demonstrated divergence from non-Caucasian populations. We hope that our findings may be helpful for other researchers and clinicians, evaluating the pharmacogenetic role of *ABCG2*.

Keywords: High Resolution Melting; *ABCG2*; BCRP; gene scanning; Polish population

1. Introduction

Among the members of the ATP-binding cassette (ABC) transporters family, *ABCG2* functions as one of the key efflux transporters of xenobiotics and their metabolites [1]. Due to it being first isolated from multidrug-resistant human breast cancer cells, the transporter has also been given the name “breast cancer resistance protein” (BCRP) [2]. The *ABCG2* gene is located on chromosome 4q22, where it spans over 66 kbp and contains 16 exons encoding a 655 amino acid protein of molecular weight 72.6 kDa [3,4].

ABCG2 is also known as the “half ABC transporter” because has only one ATP-binding cassette and six transmembrane domains, and hence requires dimerization or oligomerization to transport substrates [5].

ABCG2 plays a protective role and hence is widely expressed in normal cells and tissues, such as capillary endothelial cells, gastrointestinal tract, and hematopoietic stem cells, as well as the tissue of the liver, kidney, and testes; it has also been identified in the maternal glands, the maternal–fetal barrier of the placenta, and the blood-brain barrier [1,5]. However, it has been suggested that ABCG2 expression may be associated with a poor response to cancer chemotherapy and clinical drug resistance [6] associated with the efflux of drugs from cancer cells and multidrug resistance (MDR) to a variety of chemotherapeutic agents, small-molecule inhibitors like tyrosine kinase inhibitors (TKIs) or cyclin-dependent kinase 9 (CDK9) inhibitors [7]. Overexpression of human ABCG2 is shown to confer resistance against various, clinically-relevant compounds, including mitoxantrone, methotrexate, topotecan, irinotecan, imatinib, and flavopiridol [8]. This mechanisms has been associated with various types of leukemia and cancers of the breast, lung, and gastrointestinal system, among others [9,10]. MDCKII-BCRP cells with ABCG2 overexpression were also found to be more vulnerable to exogenous oxidative stress induced by several oxidants. This finding can be important in the process of selective death induction within cells overexpressing ABCG2 [11].

Beside detoxification, ABCG2 also plays other important transport roles. It regulates heme/porphyrins homeostasis by allowing the efflux of porphyrins from cells [12] and also the transport of endogenous folates, mostly mono-, di-, and tri-glutamates of folic acid [10]. In addition, ABCG2 is known as “urate transporter” because it regulates serum uric acid level by the excretion of urate [13].

Many studies suggest that the presence of certain polymorphic variants could alter the transport functions of ABCG2, thus influencing the pharmacokinetics and resistance associated with various chemotherapeutic agents [5,14]. In addition, a dysfunction of the transporter may also lead to the development of gout, a common disease characterized by sodium urate salt deposition, typically in or around the joints and the kidney, which can potentially lead to inflammatory arthritis, nephropathy and kidney stones. Such deposition of urate crystals occurs as a direct consequence of elevated urate serum levels, i.e., hyperuricemia, which itself also increases the risk of other diseases of the excretory and circulatory systems [15].

Furthermore, several ABCG2 nonsense and frameshift mutations, identified in the last decade, are demonstrated to be responsible for a lack of the erythrocyte blood group antigen Jr^a, which is a condition known in medicine for over 40 years. All of the null alleles introduce premature stop codons in the ATP-binding domain of ABCG2, resulting in the development of the Jr(a-) blood group phenotype [16,17]. Such individuals, i.e., those null for ABCG2, produce an ABCG2 alloantibody upon blood transfusion that can cause transfusion reactions, which can lead to fatal hemolytic diseases of the fetus and newborn in extreme cases [18].

The observed frequencies of ABCG2 polymorphisms greatly differ between populations [19,20]. The most widely represented across different populations, and hence the most extensively studied due to their potential clinical relevance, have been the missense variants c.34G>A (p.Val12Met) and, especially, c.421C>A (p.Gln141Lys) [21]. Although the pharmacogenetic role of c.34G>A (p.Val12Met) single nucleotide polymorphisms (SNP) remains unclear, c.421C>A (p.Gln141Lys) has often been associated with altered protein function [22]. It appears to be very common in Asian populations, with reported allelic frequencies between 27% and 34%. Its frequency in Caucasian populations is approximately 10%, while is rare in sub-Saharan African and African-American populations, with frequencies of only 5% [23]. The genetic variation of the complete coding sequence has been most commonly described for Asian populations [24–26] and, among Caucasians, the Swedish [27] and Dutch populations have been fully studied [28].

The only ABCG2 polymorphisms studied so far in the Polish population were the most common ones, viz. c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys), described recently in two case-control studies of the same team, examining their association with peptic ulcers development [29] and

multiple myeloma susceptibility [30]. However, their observed allelic frequencies, 0% of c.34G>A (p.Val12Met) and only 1% of c.421C>A (p.Gln141Lys), were significantly lower from those described previously in other Caucasian populations. In our opinion, this unexpected discrepancy should be clarified, whether actually frequencies of these SNPs in the Polish population significantly differ from other Caucasians, especially taking into account their clinical associations. Our preliminary analysis of the genetic variation of *ABCG2* in a Polish population revealed the occurrence of not only the most frequently-described polymorphisms but also a number of rare or clinically important variants. These premises all together prompted us to take a look more carefully at *ABCG2* genetic variation in the Polish population. Therefore, to present the first complete genetic profile of *ABCG2* in the Polish population we examine not only the common variants of this gene but all its coding fragments using High Resolution Melting (HRM). The findings describe virtually the almost entire coding sequence of *ABCG2* among 190 individuals representing the Polish population.

2. Material and Methods

2.1. Genomic DNA Samples

Human genomic DNA samples were derived from anonymous unrelated volunteers recruited during the years 2010–2012 from the whole of Poland as part of the TESTOPLEK research project. In 2013, the samples were registered in the BBMRI catalog as the POPULOUS collection (POPULATION—LOdz UniverSity Biobank) at Biobank Lab, Department of Molecular Biophysics, University of Lodz [31,32]. Each subject gave written informed consent to participate in the study and completed a questionnaire. The study was approved by the Research Bioethics Commission, University of Lodz (Decision no. 8/KBBN-UL/II/2014 and Statement of the Research Bioethics Commission, University of Lodz from 17 June 2010). All procedures were performed in accordance with the Declaration of Helsinki.

Saliva samples were collected from each individual into Oragene OG-500 DNA collection/storage receptacles (DNA Genotek, Kanata, ON, Canada). Genomic DNA was subsequently isolated using a MagNA Pure LC DNA Isolation Kit—Large Volume (Roche, Basel, Switzerland) and quantified using broad range Quant-iT™ dsDNA Broad Range Assay Kit (Invitrogen™, Carlsbad, CA, USA). All the DNA samples also underwent internal quality control by PCR for sex determination [33], which is a standard procedure in our laboratory. The final DNA concentration was normalized to 200 pg/μL. A group of 190 representative samples from the POPULOUS collection were randomly enrolled in the study and their DNA was collected on two 96-vial plates (each of plate also included single sample with molecular grade water as negative control).

2.2. *ABCG2* Scanning Using HRM

ABCG2 genetic variation was studied using HRM which is an efficient and sensitive method for scanning sequences [34]. The reaction was performed as was described previously [34], regarding mixture composition, conditions of PCR and melting, using 384-micro well plates and CFX384™ real-time PCR system (Bio-Rad Laboratories Inc., Hercules, CA, USA) in duplicate for each sample. The primers are presented in Table 1. The HRM melting curve analysis was performed in Bio-Rad Precision Melt Analysis Software, version 1.2 (Bio-Rad Laboratories Inc., Hercules, CA, USA). Melting results for each DNA plate have been analyzed separately to avoid false distinction [35]. To verify the obtained genetic variation results, several samples representing each melting cluster were selected for direct sequencing. In the event that fewer than four samples were obtained, all samples from cluster were used.

Table 1. Pairs of primers used for High Resolution Melting (HRM) screening and sequencing of ABCG2.

Exon	Number of Scanned bp Upstream from the Exon (without Primer)	Forward Primer (5'–3')	Reverse primer (5'–3')	Number of Scanned bp Downstream from the Exon (without Primer)	Size (bp) Including Primers	Polymorphism Detected in HRM Scanning
HRM scanning						
1	-	-	-	-	-	-
2	9 (28*)	TGGACTATCAACTTACTATGCTTTC	CAGCTCCTTCAGTAAATGCC	-	161	c.34G>A (rs2231137)
		GACAGCTTCATGACCTGA	TGGAATAGCCAAACCTGTG	54	213	c.203+36A>G (rs148152)
3	23	GAACTGACTTAAATGGAATAGTAAA	AAAAGTCACAGAAACCGCT	23	153	c.263+10A>G (rs2231138)
4	16	TCTATAGGTATAGACCCACAAGAT	GAACAGAAAATGCAAAACCA	11	190	c.335C>A (rs19473672)
5	32	CATGCTCTAGAAAAGACTCATATCA	TGACCTCTTAAATCCGTTCC	37	174	c.421C>A (rs2231142)
		GAACCTTACAGTCTCAGCAGCTC	CCACACAGGGAAGCTCTACT	178	-	-
6	22	GATAATGACTGGTGTATACATGAC	GTCCATCTCTATATAGTCCCTTTT	49	115	c.532+16A>G (rs2231144)
		TTATCCGTGGTGTCTCTGG	CCCCAAGATATCTGGGACA	215	-	-
7	51	GGCAAGAATAGATATTATTACTAGAA	TATCCAGGCTGATGAATGGA	129	-	c.706C>T/C>A (rs140207606) c.690+19_690+17delTGT (rs753256449)
		GCAGGGACGAACAATCATCT	AGCCCGGTGGAACAATAAGTC	109	-	-
		TGATAGCCTCACTTATTTGGC	CTATTATGAAGCATTTTACAGCATAA	125	-	-
8	32	TCTTTGTTTCCAAAGACATCA	AGCAAAATCTGGGACTGT	15	191	-
9	22	TGTGTTATCTGTGTGTTCC	GGAGTGCATAAATCTCCGC	15	120	-
		CCAAGCAGGATAAGCCACTC	GCTGATCTCTCTTGAAGACTGTG	26	140	c.1060G>A (rs138606116)
		AATTACATCAACTTTTCCGGGG	AACCAATGTGTCCCATTTGA	199	-	-
10	24	TCTCTAATGAAACTCTTCCCTTT	TGAAGAAAGTACAGCATTTCTGA	24	181	-
11	60	TGTGAAAGAGTTTGTGGGTA	CTAACCAATAGCCCCTGCTG	37	229	c.1278+28G>A (rs771435451) c.1302G>A (rs781367109) c.1367+20G>A (rs2231153)
12	37	TGCCCTGTGGCTCTTAAAC	CTTGTAACATCTCTCATGGGT	40	147	c.1368+21A>T (rs77229662)
		TCAGGGATATACAGAGTGCA	CTGACTTCACCATAGGCAA	174	-	c.1492+38G>A (rs2231155)
13	26	GTGACCTAGGCAGTTGGGTT	AACCACTCTGACCTGCTG	156	-	-
		TGTTTACCCTTATGATGGTGGC	TCAGAGCAAAACAGTTCAGAC	145	-	-
		AGCAGGTCAGAGTGTGTTT	TGTGCAGGAAGAAATGAAGGA	42	127	c.1647+40T>C (rs2231157)
14	10	GCAGGCTGACTTTTATGATTG	TGTTCTAGCTTGGGAATGC	29	172	c.1714A>C (rs200894058)
15	15	CTGTTTACCTTGCCCTGCTC	CAAAACCAATTTTGAACCTGAA	11	151	-
16	-	-	-	-	-	-

28*—upstream from the START codon. Due to observation of multiple melting domains for coding fragment of exon 16 in HRM scanning, despite several attempts, resulting no possibility to successively study its genetic variation, primers sequences for this fragments have been omitted.

Table 1. Contd.

Exon	Number of Scanned bp Upstream from the Exon (without Primer)	Forward Primer (5'→3')	Reverse primer (5'→3')	Number of Scanned bp Downstream from the Exon (without Primer)	Size (bp) Including Primers	Polymorphism Detected in HRM Scanning
HRM scanning						
1	-	-	-	-	-	-
2		GTTCGCTGCTTCCCAIT	TGACTTCATGATCGCTGC		719	
3		ATGCATGACCTGTTTGTGTT	AAAGCAAGACACCATGGCT		650	
4		AGCCAATGCTGCTGCTTT	ACCAGCAGACATTCCAAACT		716	
5		AGTTGGAAATGCTGCTGGT	ACGTACAAACACCAITGGC		649	
6		AGTACAGTATGGCCGGAAT	CCCCTACACCTCATCACAG		718	
7		TGCAGATTACTGGGTGCT	CACATGCATGCACATGAAA		721	
8		CGTGGGAAGAAGAGAAAGAA	GCAAATGCATATGCTCAAG		535	
9		GGCCTGTATCTTTCTAATGACAG	CTGAGTTCATGCCACACAC		967	
10		TTTTCCACGAGTGTCTTG	AGGCAGGAGAAITGCTTGAA		575	
11		AGGCCCATCTCAAAATAC	TTGCTTCTCTCTCCACATT		732	
12		GAGAGGGAGGATGTGTGAA	GGTTCAGGCTGCAGTGAGAT		727	
13		TTCTGTGAAAAGAGAGGCA	TCAAGAGAATCTGCCGATCA		797	
14		CAGTGCTCTTGACCTTCCA	TGAACGTAAATGCAACAATGTG		709	
15		ATTGGGGAGAGAGAGGAA	CAGGAACCATAGCCCTTGA		731	
16		-	-		-	

28*—upstream from the START codon. Due to observation of multiple melting domains for coding fragment of exon 16 in HRM scanning, despite several attempts, resulting no possibility to successively study its genetic variation, primers sequences for this fragments have been omitted.

2.3. Verification of Scanning Results by DNA Sequencing

Preparation of samples for sequencing including reaction mixture and the conditions used for PCR, the samples purification and quality control procedures, have been described previously [34]. The list of primers used for direct sequencing is presented in Table 1. The samples were applied for sequencing reaction using BigDye Terminator V3.1 (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. The PCR-sequencing product was purified using BigDye X-Terminator kit following the manufacturer's protocol (Applied Biosystems, Foster City, CA, USA). Then, 30 µL of each purified sample was applied to the 96 wells of the titration plates and analyzed using a 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

2.4. Detection of ABCG2 Polymorphisms

Sequencing results were analyzed by CodonCode Aligner software (CodonCode Corporation, Centerville, NY, USA) using ABCG2 genomic DNA sequence NG_032067.2 obtained from GenBank (<http://www.ncbi.nlm.nih.gov>) as a reference sequence. Each of detected ABCG2 polymorphisms were assigned the following parameters from GenBank, which were then used as nomenclature in present study: dbSNP IDs (rs numbers), the coding DNA nucleotide position based on the NM_004827.3 reference sequence, the amino acid position in the protein for SNPs in exons based on the NP_004818.2 reference sequence.

2.5. Inter-Population Comparison of Frequency of ABCG2 Polymorphisms

The population specific differences have been revealed and expressed as MAF values, confirmed by the chi-square test for alleles, with Yates' correction if any allele quantity was below 5. The significance of the difference was shown as *p*-values, with values < 0.05 being regarded as significant. Comparison of obtained genetic variation results with literature data from ABCG2 scanning for other populations was performed for overlapped polymorphisms.

2.6. Predicting the Effect of Missense Variants

The effect of missense variants on protein functionality was predicted with common in silico tools: SIFT—which predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between alternate amino acids (<https://sift.bii.a-star.edu.sg/>); PolyPhen-2—which uses 5%/10% false positive rates for the HumDiv model and 10%/20% false positive rates for the HumVar model as the thresholds for classification (<http://genetics.bwh.harvard.edu/pph2/dokuwiki/start>); REVEL—an ensemble method for predicting the pathogenicity of missense variants which integrates scores from other tools (MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP, and phastCons) (<https://sites.google.com/site/revelgenomics/>); MetaLR—an ensemble method that uses logistic regression to integrate the results of other components (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP) (<https://sites.google.com/site/jpopgen/dbNSFP>).

2.7. Linkage Disequilibrium and Haplotype Blocks Analysis

The consistency of genotype distribution was verified for all the detected polymorphisms by the Hardy–Weinberg equilibrium (HWE) exact test assuming a *p*-value higher than 0.001. Linkage disequilibrium (LD) and haplotype block analysis was performed by Haploview 4.2 software (<http://www.broad.mit.edu/mpg/haploview/>). LD analysis was performed for each pair of polymorphisms using |D'| and *r*² parameters. The obtained data for both these linkage parameters was compared with available population statistics using the LDmatrix Tool from the LDlink package (<https://ldlink.nci.nih.gov/?tab=ldmatrix>).

Haplotype block identification was performed based on the four-gamete test which indicates the occurrence of at least one historical recombination event and identify a set of contiguous and ordered SNP markers which demonstrate no evidence for recombination [36].

The consistency of the haplotype distribution was additionally verified by the MACH 1.0 tool (<http://csg.sph.umich.edu/abecasis/MACH/index.html>). Based on this data, network analysis was performed for haplotypes from each block with the Network 10 tool using the media-joining algorithm which was chosen as the most suitable approach (<https://www.fluxus-engineering.com/sharenet.htm>).

3. Results

3.1. Detected ABCG2 Genetic Variation

The ABCG2 coding regions and exon/intron boundaries were scanned using HRM, as these were the most important gene areas. The maximal length of the melted PCR products was not longer than 250 bp to ensure the acceptable melting resolution suitable for polymorphisms detection. Therefore, several long exons were divided into separate fragments for analysis, giving a total of 24 areas successively scanned by HRM (Figures S1–S24 and Tables S1–S24). A total of 14 coding exons were fully scanned for each of the 190 subjects. Only the final coding fragment in exon 16 could not be scanned due to problems with the melting resolution caused by the presence of multiple melting domains and this was true even for very shortened length of PCR products. Likewise, exon 1 and the major part of exon 16 had to be omitted due to them containing long, non-coding sequences, resulting in difficulties with PCR amplification caused by GC-rich composition among others.

Table 2. Summary of detected ATP-binding cassette sub-family G member 2 (ABCG2) genetic variation using HRM.

Exon Scanned by HRM	dbSNP ID	Variant Position NM_004827.3:	Intron/Amino Acid Residue NP_004818.2:	Observed Genotypes ^{a,b} (n)			HWE Exact Test ^c p-Value	MAF ^d
				R/R	R/V	V/V		
2	rs2231137	c.34G>A	p.Val12Met	175	12	0	1	(A) 0.032
2	rs4148152	c.203+36A>G	Intron	175	14	0	1	(C) 0.037
3	rs2231138	c.263+10A>G	Intron	177	10	0	1	(G) 0.027
4	rs199473672	c.335C>A	p.Pro112Gln	184	1	0	1	(A) 0.003
5	rs2231142	c.421C>A	p.Gln141Lys	147	38	3	0.954	(A) 0.117
6	rs2231144	c.532-16A>G	Intron	174	13	0	1	(G) 0.035
7	rs140207606	c.706C>T	p.Arg236Ter	186	1	0	1	(T) 0.003
7	rs140207606	c.706C>A	p.Arg236=	186	1	0	1	(A) 0.003
7	rs753256449	c.690-19_690-17delTGT	Intron	186	1	0	1	(delTGT) 0.003
9	rs138606116	c.1060G>A	p.Gly354Arg	187	1	0	1	(A) 0.003
11	rs771435451	c.1278-28G>A	Intron	187	1	0	1	(A) 0.003
11	rs781367109	c.1302G>A	p.Thr434=	187	1	0	1	(A) 0.003
11	rs2231153	c.1367+20G>A	Intron	0	11	177	1	(G) 0.029
12	rs772239662	c.1368-21A>T	Intron	188	1	0	1	(T) 0.003
12	rs2231155	c.1492+38G>A	Intron	187	2	0	1	(A) 0.005
13	rs2231157	c.1647+40T>C	Intron	68	93	27	0.734	(C) 0.391
14	rs200894058	c.1714A>C	p.Ser572Arg	186	1	0	1	(C) 0.003

^a Number of genotypes detected during this study, R—reference allele, V—variant allele. ^b Total number of examined samples was 190, however during scanning single samples were excluded due to accidental problems with amplification during PCR and hence, total number of genotypes does not equal precisely 190. ^c p-value is consistent with Hardy–Weinberg equilibrium if $p > 0.001$. ^d Minor allele shown in brackets with its frequency.

HRM scanning identified 17 variants. Two of these were located in the same nucleotide position but presented different molecular effects (c.706C>T, p.Arg236Ter, and c.706C>A, p.Arg236=) so they were counted in present study as separate SNPs (Table 2). Eight SNPs were found in the exons (Figure 1) and five of them are missense variants which change the amino acid sequence: c.34G>A (p.Val12Met), c.335C>A (p.Pro112Gln), c.421C>A (p.Gln141Lys), c.1060G>A (p.Gly354Arg), c.1714A>C (p.Ser572Arg). One SNP was also heterozygously detected as point-nonsense mutation resulting in a premature stop codon (c.706C>T, p.Arg236Ter) with the minor allele frequency (MAF) estimated as 0.003. Among the SNPs located in the exons, only one was found homozygously, c.421C>A (p.Gln141Lys) with MAF = 0.117. Eight SNPs were identified in the intron sequences, one of them

homozygously c.1647+40T>C with MAF = 0.391. Moreover, one deletion variant was detected in an intron at c.690-19_690-17delTGT with MAF = 0.003.

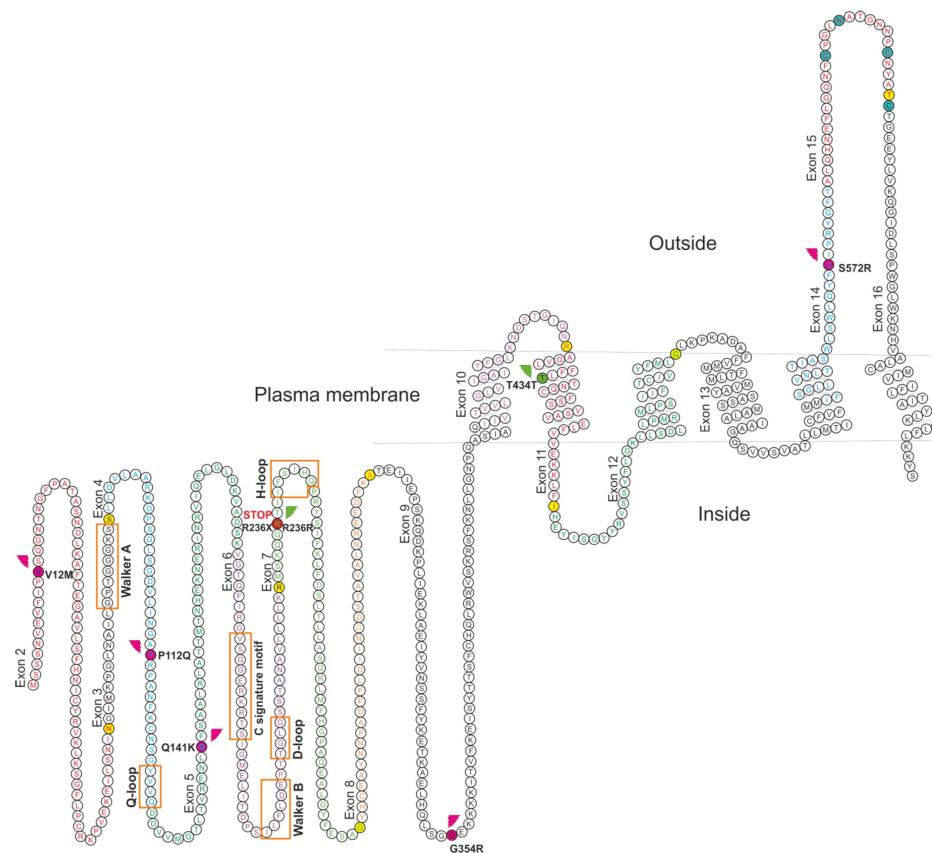


Figure 1. Structure of ABCG2 transporter in plasma membrane with highlighted amino acids for which genetic variants were detected in this study (based on the NP_004818.2 reference sequence). Amino acids encoded by subsequent exons are described and typed by different font color, and amino acids located on exons borders as splicing effect have background colored in yellow. Conserved motifs in Nucleotide Binding Domain: Walker A and B, the ABC signature C motif, Q-, D- and H-loops are described and indicated with orange squares. Amino acids involved in forming intra-molecular (C592-C608) and inter-molecular disulfide bonds (C603), and functional N-glycosylation site (N596) have background colored in blue. Amino acids with colored background and marks mean variants detected in this study: purple—missense, green—silent, red—nonsense mutation.

The cytoplasmic nucleotide binding domain (NBD) of the protein, encompassing amino acids 1–395, was found to include four amino acids altered by missense variants and one premature stop variant: c.335C>A (p.Pro112Gln) located 25 amino acids downstream of the Walker A motif and 11 upstream of the Q-loop motif; variant c.421C>A (p.Gln141Lys), located 15 amino acids downstream of the Q-loop motif and 45 upstream of the C signature motif; SNP c.706C>T (p.Arg236Ter), located 25 amino acids downstream of Walker B motif, and three upstream of the H-loop motif. All of the motifs have been described as highly conserved and form the ATP catalytic center of the protein. SNP c.1714A> (p.Ser572Arg) was located in the extracellular domain, 20 amino acids upstream of the first cysteine which is crucial for forming intra-molecular disulfide bonds (Figure 1).

In silico analysis of all the missense variants detected in this study, incorporating several of the most common predictors, surprisingly classified three SNPs, *viz.* c.34G>A (p.Val12Met), c.421C>A (p.Gln141Lys) and c.1060G>A (p.Gly354Arg), as benign in protein functioning. However, two other variants, *viz.* c.335C>A (p.Pro112Gln) and c.1714A> (p.Ser572Arg), were assessed to be potentially damaging and having a possible deleterious effect on protein activity (Table S25).

3.2. Inter-Population Comparison of ABCG2 Genetic Variation

The studied Polish population demonstrated lower MAF values than those obtained in Asians, represented by Chinese [24], Japanese [26], and Korean populations [25], for the following SNPs: c.34G>A (p.Val12Met); c.203+36A>G; c.263+10A>G; c.421C>A (p.Gln141Lys); c.1367+20G>A (Table 3). Only SNP c.1647+40T>C did not significantly differ between the Polish population (MAF = 0.391) and the Japanese (MAF = 0.398, $p = 0.840$) and the Korean populations (MAF = 0.326, $p = 0.135$). It was nevertheless found to be significantly different to the Chinese population (MAF = 0.241, $p = 0.033$), however, this was probably only an effect of the small sample size included in that particular study. The Polish population demonstrated a higher MAF for c.532–16A>G (0.035); however, this SNP was only detected in the Japanese population (MAF = 0.003, $p = 0.002$). The comparison of a similar set of overlapping SNPs, *viz.* c.34G>A (p.Val12Met), c.203+36A>G, c.421C>A (p.Gln141Lys), c.1367+20G>A and c.1647+40T>C, did not reveal any significant differences between two Caucasian populations: the Polish and the Dutch [28] (Table 3).

Table 3. Inter-population comparison of minor allele frequency (MAF) values of ABCG2 polymorphisms.

Population	<i>n</i>	rs2231137 c.34G>A (p.Val12Met)	rs4148152 c.203+56A>G	rs2231138 c.263+10A>G	rs2231142 c.421C>A (p.Gln141Lys)	rs2231144 c.532-16A>G	rs138606116 c.1060G>A (p.Gly354Arg)	rs2231153 c.1367+20G>A	rs2231157 c.1647+40T>C
Polish (current study)	190	0.032	0.037	0.027	0.117	0.035	0.003	0.029	0.391
Chinese [24]	27	0.204 (<i>p</i> < 0.001)	0.259 (<i>p</i> < 0.001)	0.204 (<i>p</i> < 0.001)	0.222 (<i>p</i> < 0.032)	-	-	0.315 (<i>p</i> < 0.001)	0.241 (<i>p</i> = 0.033)
Japanese [26]	177	0.192 (<i>p</i> < 0.001)	0.192 (<i>p</i> < 0.001)	0.068 (<i>p</i> = 0.009)	0.319 (<i>p</i> < 0.001)	0.003 (<i>p</i> = 0.002)	0.003 (<i>p</i> = 0.508)	0.161 (<i>p</i> < 0.001)	0.398 (<i>p</i> = 0.840)
Korean [25]		0.230 (<i>p</i> < 0.001, <i>n</i> = 275)	0.250 (<i>p</i> < 0.001, <i>n</i> = 92)	0.098 (<i>p</i> < 0.001, <i>n</i> = 92)	0.280 (<i>p</i> < 0.001, <i>n</i> = 275)	-	-	0.240 (<i>p</i> < 0.001, <i>n</i> = 92)	0.326 (<i>p</i> = 0.135, <i>n</i> = 92)
Dutch [28]	100	0.065 (<i>p</i> = 0.066)	0.065 (<i>p</i> = 0.130)	-	0.120 (<i>p</i> = 0.916)	-	-	0.045 (<i>p</i> = 0.336)	0.465 (<i>p</i> = 0.086)
Vietnamese [25]	140	0.360 (<i>p</i> < 0.001)	-	-	0.310 (<i>p</i> < 0.001)	-	-	-	-
Chinese [25]	191	0.330 (<i>p</i> < 0.001)	-	-	0.290 (<i>p</i> < 0.001)	-	-	-	-
Swedish [27]	60	0.020 (<i>p</i> = 0.376)	-	-	0.100 (<i>p</i> = 0.608)	-	-	-	-
Hungarian [37]	149	0.040 (<i>p</i> = 0.570)	-	-	0.094 (<i>p</i> = 0.336)	-	-	-	-
Czech [38]	100	0.015 (<i>p</i> = 0.221)	-	-	0.075 (<i>p</i> = 0.113)	-	-	-	-
German		0.030 (<i>p</i> = 0.862, <i>n</i> = 348) [39]	-	-	0.104 (<i>p</i> = 0.440, <i>n</i> = 1552) [40]	-	-	-	-
Han Chinese [20]	95	-	-	-	0.342 (<i>p</i> < 0.001)	-	-	-	-
American Caucasian [20]	88	-	-	-	0.119 (<i>p</i> = 0.938)	-	-	-	-
European Caucasian [20]	84	-	-	-	0.107 (<i>p</i> = 0.738)	-	-	-	-
African American [20]	94	-	-	-	0.053 (<i>p</i> = 0.015)	-	-	-	-
African (sub-Saharan) [20]	938	-	-	-	0.009 (<i>p</i> < 0.001)	-	-	-	-

Shaded cells mean significantly higher MAF than in the Polish population; bolded values mean significantly lower MAF than in the Polish population.

These observations are consistent with those of studies from other populations, which were focused only on the most common variants c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys). The Polish population demonstrated significantly lower MAFs for these SNPs (0.032 and 0.117, respectively) than in Asian populations, such as those from like Vietnam (MAF = 0.360 and MAF = 0.310, respectively) and another studied group from China (MAF = 0.330 and MAF = 0.290, respectively) [25]. However, the values were not significantly different to those from other Caucasian populations, such as those in Sweden (MAF = 0.020 and MAF = 0.100, respectively) [27], Hungary (MAF = 0.040, and MAF = 0.094, respectively) [37], the Czech Republic (MAF = 0.015, and MAF = 0.075, respectively) [38] and Germany (MAF = 0.030, [39] and MAF = 0.104 [40], respectively) (Table 3).

These observations also correspond with comparison of ethnic groups in general, available only for c.421C>A (p.Gln141Lys). The MAF among the Polish population (0.117) was found to be significantly lower than that for the Han Chinese population (MAF = 0.342, $p < 0.001$), although its frequency was not significantly different than that observed for Caucasian Americans (MAF = 0.119, $p = 0.938$) and Europeans (MAF = 0.107, $p = 0.738$) and it was higher than African Americans (MAF = 0.053, $p = 0.015$) or Africans from the sub-Sahara region (MAF = 0.009, $p < 0.001$) [20] (Table 3).

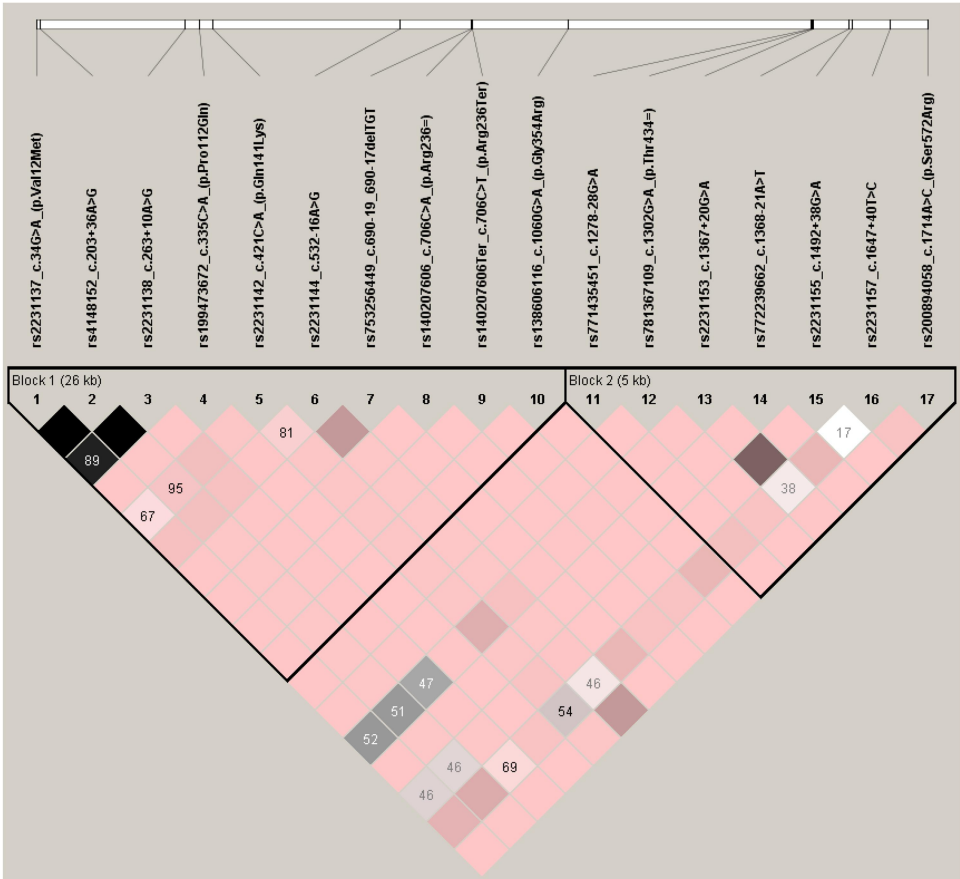


Figure 2. Linkage disequilibrium analysis of *ABCG2* polymorphisms detected in this study using $|D'|$ parameter. Pairwise linkage demonstrated as $|D'|$ values (100 $\times$) by graded pink and black colors, denser color means closer linkage and lack of value means $|D'| = 1 \times 100$. Bolded triangles show haplotype blocks.

3.3. Linkage Disequilibrium Analysis

Linkage disequilibrium analysis was performed using r^2 and $|D'|$ statistics based on consistency with the HWE exact test for all detected variants. For the $|D'|$ parameter, perfect linkage ($|D'| = 1$) was observed for the great majority of polymorphism pairs (122 from 136), while strong linkage ($1 > |D'| \geq 0.8$) was detected for the next three pairs. Weak or absent linkage, as indicated by low $|D'|$ values ($|D'| < 0.8$) was observed mostly between pairs of polymorphisms from c.34G>A (p.Val12Met), c.203+36A>G, c.263+10A>G, c.421C>A (p.Gln141Lys), c.532-16A>G and those from c.1367+20G>A, c.1492+38G>A, c.1647+40T>C (Figure 2).

Regarding r^2 parameter, no examples of perfect linkage ($r^2 = 1$) were detected between any polymorphism pair. However, strong linkage was revealed between three SNPs: c.34G>A (p.Val12Met) and c.203+36A>G ($r^2 = 0.852$); c.203+36A>G and c.263+10A>G ($r^2 = 0.706$); c.34G>A (p.Val12Met) and c.263+10A>G ($r^2 = 0.665$) (Figure 3).

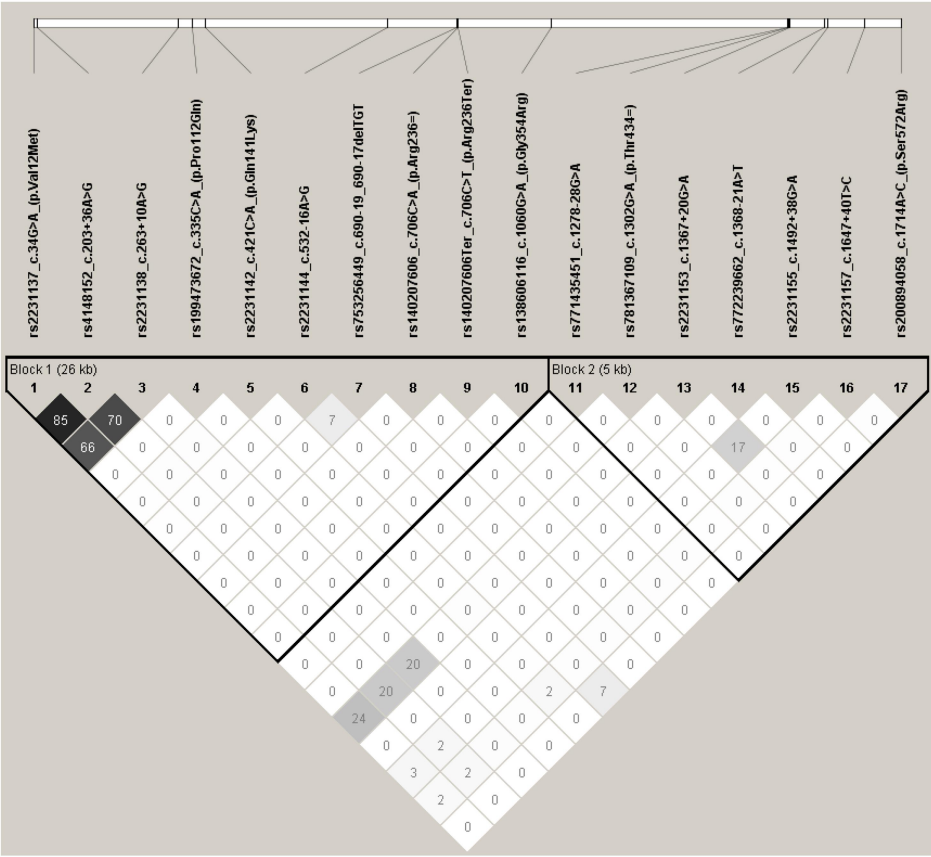


Figure 3. Linkage disequilibrium analysis of *ABCG2* polymorphisms detected in this study using r^2 parameter. Pairwise linkage demonstrated as r^2 values ($\times 100$) by graded black colors, denser color means closer linkage. Bolded triangles show haplotype blocks.

These observations are in line with the linkage statistics obtained by LDlink for the 1000 Genomes Project database. Most polymorphism pairs demonstrated maximal $|D'|$ values, with weaker values identified for the same pairs given above, regardless of the chosen 1000 Genomes population. Likewise, only SNP c.34G>A (p.Val12Met), c.203+36A>G and c.263+10A>G were strongly linked in terms of r^2 parameter among Europeans (Figure S25a) and world population in general (Figure S25b).

3.4. Haplotype Block Estimation and Haplotype Analysis

The analysis of haplotype blocks revealed at least one recombination spot, which was clearly indicated upstream of SNP c.1367+20G>A (scanned exon 11). However, two other SNPs were detected closely upstream of the estimated recombination spot and because these were observed only in single samples so were arbitrarily included to the second haplotype block. Hence, two haplotype blocks were defined, the first one including detected polymorphisms from exon 2 to 9, and the second one, polymorphisms from exon 11 to 14. A total of seven haplotypes were identified whose frequencies were estimated as at least 1% (Table 4).

Block 1, spanning 26 kb, included 10 polymorphisms and four common haplotypes were established for them. The most frequent haplotype comprised only the major alleles of all polymorphisms (frequency: 0.804), followed by haplotype with minor allele A for SNP c.421C>A (p.Gln141Lys) (frequency: 0.113). The frequencies of the next two haplotypes were much lower, i.e., 0.031 for c.532-16A>G and 0.024 for c.203+36A>G (both, with minor alleles G). Block 2 encompassed seven SNPs over a distance of around 5 kb. The most frequent haplotype consisted of only major alleles of all SNPs (frequency: 0.584), followed by one with the minor allele C for SNP c.1647+40T>C (frequency: 0.376) and another with the minor allele G for c.1367+20G>A (frequency: 0.020). These five SNPs used as the basis for distinguishing the major haplotypes, were tagged as haplotype tag SNPs (htSNPs). A complete set of haplotypes and their frequencies are presented in Table S26.

The most common haplotype from block 1 and the two most frequent from block 2 demonstrated greater than 10% crossing. Other possible crossings between haplotypes with frequencies higher than 1% were observed between the second most common haplotype of block 1 and the two most common from block 2, between the third most common haplotype of block 1 and the most common from block 2, and between the fourth most common haplotype of block 1 and the second and third from block 2. The frequency of recombination between blocks was estimated as 0.40, calculated as the multiallelic D' coefficient.

Network analysis found the most common haplotype in block 1 to act as core for the others. However, three closely-linked variants, *viz.* c.34G>A (p.Val12Met), c.203+36A>G and c.263+10A>G, demonstrated separate paths of divergence (Figure S26a). In block 2, the most common haplotype was found to act as the root for the two next most common but together, they all determine the origin for the remaining minor haplotypes (Figure S26b).

3.5. Inter-Population Comparison of ABCG2 Haplotypes

The adjusted haplotype profiles of Japanese [26] and Dutch populations [28] were also compared with those of the present study. Differences in the frequencies of the common haplotypes G-A-A-C, G-A-A-A, A-G-A-C and A-G-G-C for the overlapping SNPs in block 1, *viz.* c.34G>A (p.Val12Met), c.203+36A>G, c.263+10A>G, c.421C>A (p.Gln141Lys), were observed between the Polish (frequencies of 0.848, 0.115, 0.008, and 0.024, respectively) and Japanese populations (frequencies of 0.489, 0.319, 0.130, and 0.062, respectively). In contrast, haplotypes: G-A-C, G-A-A, A-G-C for the overlapping SNPs c.34G>A (p.Val12Met), c.203+36A>G, and c.421C>A (p.Gln141Lys) in block 1, were very similar between the Polish (0.848, 0.115 and 0.032) and Dutch populations (0.816, 0.119, and 0.065). Smaller differences were observed for common haplotype frequencies from block 2: haplotypes A-T, A-C, and G-T for the overlapping SNPs c.1367+20G>A and c.1647+40T>C, were 0.585, 0.386, and 0.024 in the present study and 0.441, 0.398, and 0.161 for the Japanese population and 0.493, 0.463, and 0.045 for the Dutch population.

Table 4. Haplotype block analysis.

Variant residue ^{a,b}	Block 1										Recombination ^d between blocks	Block 2					Haplotype frequency
	c.34G>A p.Val12Met	c.203+36A>G	c.263+10A>G	c.335C>A p.Pro112Gln	c.421C>A p.Gln141Lys	c.532-16A>G	c.706C>T p.Arg236Ter	c.706C>A p.Arg236=	c.690-19_690-17delTCT ^e	c.106G>A p.Gly534Arg		Haplotype frequency					
Haplotypes ^c	G	A	A	C	C	A	C	C	TGT	G	0.804	G	A	G	T	A	0.584
	G	A	A	C	A	A	C	C	TGT	G	0.113	G	A	A	C	A	0.376
	G	A	A	C	C	G	C	C	TGT	G	0.031	G	G	G	T	A	0.020
	A	G	G	C	C	A	C	C	TGT	G	0.024	others	others	others	others	others	0.020

^a Residues followed by reference sequences for coding nucleotide NM_004827.3 and amino acid position NP_004818.2. ^b Bolded polymorphisms mean haplotype tag SNPs (htSNPs). ^c Major alleles in white boxes; minor alleles in shaded boxes; all the haplotypes with their frequencies in population; haplotypes below 1% not shown and grouped like "others". ^d Level of recombination between blocks as a value of multiallelic D'. ^e TGT allele in haplotype means no deletion.

4. Discussion

Almost the entire coding sequence of the *ABCG2* gene, together with its neighboring intronic fragments, was scanned using HRM in samples taken randomly from a population of 190 unrelated Polish individuals. The results identified the presence of 17 different polymorphisms within the gene, including rare ones and some of which have possible clinical significance.

Our results unambiguously demonstrate that the frequencies of the most common polymorphisms of *ABCG2* in the Polish population, which is exclusively Caucasian, do not differ significantly to other Caucasian populations and are lower than in Asian populations. This conclusion contradicts previous, and the only available so far results of *ABCG2* genetic variation in the Polish population, obtained by Niebudek, et al. [30] and Salagacka-Kubiak, et al. [29] for SNPs c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys), with both studies probably using the same control group ($n = 97$). The obtained MAF values for the SNPs were 0 and 0.010, respectively, which were significantly lower than those established in our study: 0.032 ($p = 0.027$) and 0.117 ($p < 0.001$), respectively. Despite their results appearing surprisingly low compared to the current state of knowledge, the authors did not comment on these differences at all. It is possible that these inconsistencies could be attributed to sample size or bias effects such as too small study group or insufficiently randomized sample selection. However, very possible explanation of such underestimation would be also not enough sensitive genotyping method. Authors performed genotyping by method of restriction fragment length polymorphism (RFLP) which is still commonly used in laboratory practice, but does not include any additional step verifying results accuracy. Greater sensitivity can be achieved by methods based on qPCR such as HRM, used in the present study, which examine genetic variation in throughout the entire tested sequence, not only at selected points. In our previous study, HRM has also been demonstrated to be an efficient and sensitive method for scanning and genotyping polymorphic variants. We have successfully scanned the coding sequence of one of the most important ABC genes, *ABCC1*, for representatives of the Polish population, with the results being confirmed by targeted genotyping of selected SNPs [34]. Moreover, both HRM strategies ensure greater reliability, by including an additional control step, such as verification by direct sequencing. Although HRM also has some limitations, which we also described previously [35], which prevented us from scanning the final *ABCG2* coding part in exon 16, our findings demonstrate it to be an efficient technique to study *ABCG2* genetic variation. This conclusion is also supported by other research groups, which have successively used HRM for association studies of *ABCG2* polymorphisms with gout [41] and Jr(a-) phenotype [42]. Therefore, we determine the occurrence frequencies of SNPs c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys) in the Polish population, which have been clearly defined in this study.

Due to their potential clinical relevance, the common *ABCG2* SNPs, c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys), are the most extensively studied in vitro and in vivo research. Although the significance of c.34G>A (p.Val12Met) remains unclear, c.421C>A (p.Gln141Lys) has often been reported to be associated with altered protein functionality. Generally, lower levels of mRNA maturation, protein folding and membrane trafficking were observed compared to the wild-type protein expression in the cellular membrane [43] but this also occurred as a result of ubiquitination-mediated proteasomal degradation, which the protein undergoes when it is recognized as misfolded [44]. This processes probably occur as an effect of an amino acid change located within the functionally important ATP-binding region between the Walker A and B motifs of the protein which likely affects its ATPase activity levels [6]. Reduced *ABCG2* expression as consequence of the presence of SNP c.421C>A (p.Gln141Lys) was associated with an adverse response to cancer chemotherapy, but its pharmacogenomic effect appears to be dependent on substrate and correlated with its genotype [6,15] which has been also well summarized recently elsewhere [23,44]. Among 67 Japanese patients with chronic phase of chronic myeloid leukemia (CML) with genotype CA or AA the dose-adjusted imatinib mesylate (IM) trough concentration was significantly higher than in those with genotype CC [45]. Authors suggested reduction of IM dose from 400 mg for patients with the wild-type CC genotype to 300 mg for genotype CA or AA to obtain a plasma threshold of approximately 1000 ng/mL [46].

In case of 67 Caucasian patients with solid malignancies and gastrointestinal stromal tumors, a 22% reduction in plasma clearance (imatinib and CGP74588) was observed in heterozygous CA (genotype AA not found) compared with wild-type CC patients [47]. In another study for 349 Caucasian patients with primary lung cancer (161 small cell lung cancer (SCLC), 187 nonsmall cell lung cancer (NSCLC) and 1 mixed), significantly worse overall survival was observed for carriers with A allele treated with platinum-based drugs [39]. In Kurdish case-control study in which 100 breast cancer patients and 200 healthy controls were enrolled, A allele were significantly more frequent in patients and those with AA genotype had higher risk of progressing breast cancer. In addition, patient with A allele had complete response to chemotherapeutic agents (treatments with anthracyclines and Paclitaxol) [48]. Furthermore, in Chinese study of 1169 breast cancer patients (anthracycline-based chemotherapy), genotype GA/AA of variant c.34G>A (p.Val12Met), genotype AA of variant c.421C>A (p.Gln141Lys) and their haplotypes A-C and G-A were significantly associated with increased risk for developing breast carcinoma [49]. The conclusion seems to be common for all of mentioned studies, that clinical effects of c.421C>A (p.Gln141Lys) are significant and frequently observed, regardless of populations, drugs or cancers. Hence, c.421C>A (p.Gln141Lys) is considered to be one of the most important variant, both in terms of cancer chemotherapy response and also drug toxicity and resistance [6]. However, there were suggested ethnic differences in clinical response between East Asians and Caucasians [44], e.g., the TKIs sunitinib and erlotinib induced more frequent adverse events at normal doses in East Asians than Caucasians, resulting in higher rates of treatment failure [50]. Unfortunately, neither clinical study has been performed yet for the Polish population, however, abovementioned examples clearly suggest, that targeted genotyping together with treatment response monitoring could be very important in choosing the best strategy of therapy and should be implemented in clinical approach.

Furthermore, a number of studies of different ethnic populations, including genome-wide association (GWA) studies, have also indicated c.421C>A (p.Gln141Lys) to be the major risk factor for gout and hyperuricemia, which is correlated with increased urate concentration [13,51–53]. Regarding SNP c.34G>A (p.Val12Met), although it did not alter the protein expression and urate transport activity [43], some high-throughput studies [41,54], supported by meta-analysis results [38], suggest that it plays a protective role against gout. Thus, determining the appropriate frequency of both of these *ABCG2* variants in each particular population appears to be an important step in determining the full clinical implications of the gene.

Moreover, the scan also found rare SNPs, including several missense ones. The in silico analysis based on common predictors suggested that two of them, c.335C>A (p.Pro112Gln) and c.1714A>C (p.Ser572Arg), may be deleterious, while interestingly, the most common *ABCG2* SNPs were indicated to be likely benign for protein functioning, in spite their abovementioned clinical associations. A recent functional study, examining SNP c.1714A>C (p.Ser572Arg), confirmed this prediction and reported that when present, the cells did not produce a matured glycoprotein and demonstrated decreased levels of *ABCG2* protein in the plasma membrane [55]. This could result from the fact that the SNP is located near to conserved cysteine responsible for forming intra-molecular disulfide bound. Although similar effect could also be observed for variant c.335C>A (p.Pro112Gln) located near to important conserved motifs in NBD, to the best of our knowledge, does not appear to have been the subject of similar study. This may suggest, that some rare missense variants could have also clinical relevance and require functional studies to determine their assumed impact. Although such SNPs may have significant consequences for their carriers, their true influence in effective therapy is difficult to estimate as they observed frequency in population is quite low. However, for this reason, impact of such rare polymorphisms for the entire Polish population seems to be rather a little. Another missense variant c.1060G>A (p.Gly354Arg) was also found in the Japanese population with the same frequency ($n = 177$, MAF = 0.003, $p = 0.508$ – insignificant) [26] as in our study; however, it has indicated as being likely benign and has yet to be subjected to a functional study.

Furthermore, our present findings also indicate the presence of a heterozygous nonsense mutation c.706C>T (p.Arg236Ter) (MAF = 0.003) which was associated with the Jr(a-) blood group phenotype.

So far, c.376C>T (p.Gln126Ter, rs72552713) nonsense mutation was reported the most frequently but is rather typical for Japanese population. Among Caucasians, the nonsense mutations c.706C>T (p.Arg236Ter), observed in the present study, c.730C>T (p.Gln244Ter, lack dbSNP ID) and c.736C>T (p.Arg246Ter, rs200190472) are more commonly identified [56]. Unfortunately, it was not possible to collect the medical data of the donors and hence to verify the sample phenotype. Nevertheless, this finding may suggest that variants involved with the Jr(a-) phenotype in the Polish population are no such phenomenon as we previously thought and may should be considered more seriously in transfusion.

To gain a fuller picture of the *ABCG2* in Polish population, linkage disequilibrium (LD) and haplotype block analysis were performed. The LD results indicated the presence of very similar dependencies to those observed for Europeans and general worldwide population, suggesting that its inheritance pattern is rather evolutionary conserved, regardless of population type. This conclusion is consistent with haplotype block analysis which defined the two studied blocks, the first spanning from exon 2 to exon 9 and the second from exon 11 to exon 14 to be retained across populations. These also contained the same cores: between exon 2 and 5 and between exon 11 and 15, with a recombination spot between them. The polymorphisms located between exon 5 and 11 have been variously classified, probably due to their rare inheritance in this sequence fragment: they were classified in the first block, between the 5' flanking region and exon 9, as in the present study and in the Japanese population [26]; however, they were found to be classified in the second block, between exon 5 and 15, in the Chinese study [24] and they were not classified in either block in the Korean study [25].

Population specific differences have been also observed in common haplotypes comparison. Their frequencies in the Polish population looked very similar as in Dutch population [28], whereas they greatly differed compared to Japanese population [26], especially in haplotype block 1. These findings support the conclusion that differences between populations do not concern single or common polymorphisms of *ABCG2*, but the great, inherited parts of the gene. For that reason, linkage disequilibrium and haplotype analysis can reveal potentially significant population-specific differences which can be used in identification of SNPs in routine screening to obtain precision treatment based on *ABCG2* polymorphisms genotype or haplotype [44].

5. Conclusions

In conclusion, in the present study almost entire *ABCG2* coding sequence was scanned using an efficient HRM method for 190 representatives of the Polish population. The findings indicate that the frequencies of most common SNPs in Polish populations did not differ significantly from those in other Caucasians, thus providing a clarification this issue of the distribution of the clinically-relevant SNPs c.34G>A (p.Val12Met) and c.421C>A (p.Gln141Lys). Additionally, a comprehensive analysis was performed of the genetic variation of *ABCG2* in the Polish population. Clinical implementation of *ABCG2* genotyping is crucial for the development of individualized therapy and the determination of accurate dosing guidelines in future. We believe that our findings may be helpful for other researchers evaluating *ABCG2* clinical associations and be of value for clinicians working to improve diagnostic or therapeutic efficacy.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/10/1144/s1>, Figures S1–S24: HRM plots of *ABCG2* scanning for each scanned area, Figure S25: Linkage disequilibrium statistics r^2 and $|D'|$ from LDlink for 1000 Genomes Project database, for Europeans (a) and general world population (b), Figure S26: Network analysis of *ABCG2* haplotypes detected in this study for block 1 (a) and block 2 (b), Tables S1–S24: Summary of *ABCG2* scanning for each scanned area, Table S25: Nonsynonymous variants detected in this study and their predicted effect on protein functioning classified by in silico tools, Table S26: All the observed haplotypes in this study and their frequencies in *ABCG2* blocks.

Author Contributions: Conceived and designed the experiments: D.S. and M.S. Performed the experiments: M.S., D.S., M.S.-K., and M.K.-M. Data analysis: M.S. and D.S. Contributed reagents/materials/analysis tools: D.S., G.B. Contributed to the writing of manuscript: M.S., D.S., M.S.-K., G.B., and J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish POIG grant 01.01.02-10-005/08 TESTOPLEK from the European Regional Development Fund and by Polish Ministry of Science and Higher Education no. DIR/WK/2017/01: “Biobank network in Poland, within the BBMRI-ERIC Research Infrastructure of Biobanks and Biomolecular Resources”. The APC was funded by the Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź.

Acknowledgments: We would like to thank Justyna Jarczak for her comments that help to improve the manuscript. We would also like to thank Edward Lowczowski for professional linguistic improvement of the text.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Robey, R.W.; To, K.K.; Polgar, O.; Dohse, M.; Fetsch, P.; Dean, M.; Bates, S.E. ABCG2: A perspective. *Adv. Drug Deliv. Rev.* **2009**, *61*, 3–13. [\[CrossRef\]](#) [\[PubMed\]](#)
- Doyle, L.A.; Yang, W.; Abruzzo, L.V.; Krogmann, T.; Gao, Y.; Rishi, A.K.; Ross, D.D. A multidrug resistance transporter from human MCF-7 breast cancer cells. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15665–15670. [\[CrossRef\]](#) [\[PubMed\]](#)
- Allikmets, R.; Schriml, L.M.; Hutchinson, A.; Romano-Spica, V.; Dean, M. A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance. *Cancer Res.* **1998**, *58*, 5337–5339. [\[PubMed\]](#)
- Allen, J.D.; Brinkhuis, R.F.; Wijnholds, J.; Schinkel, A.H. The mouse Bcrp1/Mxr/Abcp gene: Amplification and overexpression in cell lines selected for resistance to topotecan, mitoxantrone, or doxorubicin. *Cancer Res.* **1999**, *59*, 4237–4241.
- Stacy, A.E.; Jansson, P.J.; Richardson, D.R. Molecular pharmacology of ABCG2 and its role in chemoresistance. *Mol. Pharmacol.* **2013**, *84*, 655–669. [\[CrossRef\]](#)
- Noguchi, K.; Katayama, K.; Mitsuhashi, J.; Sugimoto, Y. Functions of the breast cancer resistance protein (BCRP/ABCG2) in chemotherapy. *Adv. Drug Deliv. Rev.* **2009**, *61*, 26–33. [\[CrossRef\]](#)
- Doyle, L.; Ross, D.D. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). *Oncogene* **2003**, *22*, 7340–7358. [\[CrossRef\]](#)
- Robey, R.W.; Ierano, C.; Zhan, Z.; Bates, S.E. The challenge of exploiting ABCG2 in the clinic. *Curr. Pharm. Biotechnol.* **2011**, *12*, 595–608. [\[CrossRef\]](#)
- Robey, R.W.; Polgar, O.; Deeken, J.; To, K.W.; Bates, S.E. ABCG2: Determining its relevance in clinical drug resistance. *Cancer Metastasis Rev.* **2007**, *26*, 39–57. [\[CrossRef\]](#)
- Natarajan, K.; Xie, Y.; Baer, M.R.; Ross, D.D. Role of breast cancer resistance protein (BCRP/ABCG2) in cancer drug resistance. *Biochem. Pharmacol.* **2012**, *83*, 1084–1103. [\[CrossRef\]](#)
- Krzyżanowski, D.; Bartosz, G.; Grzelak, A. Collateral sensitivity: ABCG2-overexpressing cells are more vulnerable to oxidative stress. *Free Radic. Biol. Med.* **2014**, *76*, 47–52. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ishikawa, T.; Nakagawa, H. Human ABC transporter ABCG2 in cancer chemotherapy and pharmacogenomics. *J. Exp. Ther. Oncol.* **2009**, *8*, 5–24. [\[PubMed\]](#)
- Woodward, O.M.; Köttgen, A.; Coresh, J.; Boerwinkle, E.; Guggino, W.B.; Köttgen, M. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10338–10342. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gradhand, U.; Kim, R.B. Pharmacogenomics of MRP transporters (ABCC1-5) and BCRP (ABCG2). *Drug Metab. Rev.* **2008**, *40*, 317–354. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jani, M.; Ambrus, C.; Magnan, R.; Jakab, K.T.; Beery, E.; Zolneric, J.K.; Krajcsi, P. Structure and function of BCRP, a broad specificity transporter of xenobiotics and endobiotics. *Arch. Toxicol.* **2014**, *88*, 1205–1248. [\[CrossRef\]](#) [\[PubMed\]](#)
- Saison, C.; Helias, V.; Ballif, B.A.; Peyrard, T.; Puy, H.; Miyazaki, T.; Perrot, S.; Vayssier-Taussat, M.; Waldner, M.; Le Pennec, P.Y.; et al. Null alleles of ABCG2 encoding the breast cancer resistance protein define the new blood group system Junior. *Nat. Genet.* **2012**, *44*, 174–177. [\[CrossRef\]](#)
- Zelinski, T.; Coghlan, G.; Liu, X.Q.; Reid, M.E. ABCG2 null alleles define the Jr(a-) blood group phenotype. *Nat. Genet.* **2012**, *44*, 131–132. [\[CrossRef\]](#)
- Cleophas, M.C.; Joosten, L.A.; Stamp, L.K.; Dalbeth, N.; Woodward, O.M.; Merriman, T.R. ABCG2 polymorphisms in gout: Insights into disease susceptibility and treatment approaches. *Pharmgenom. Pers. Med.* **2017**, *10*, 129–142. [\[CrossRef\]](#)

19. Cervenak, J.; Andrikovics, H.; Ozvegy-Laczka, C.; Tordai, A.; Nemet, K.; Varadi, A.; Sarkadi, B. The role of the human ABCG2 multidrug transporter and its variants in cancer therapy and toxicology. *Cancer Lett.* **2006**, *234*, 62–72. [\[CrossRef\]](#)
20. de Jong, F.A.; Marsh, S.; Mathijssen, R.H.; King, C.; Verweij, J.; Sparreboom, A.; McLeod, H.L. ABCG2 pharmacogenetics: Ethnic differences in allele frequency and assessment of influence on irinotecan disposition. *Clin. Cancer Res.* **2004**, *10*, 5889–5894. [\[CrossRef\]](#)
21. Kurose, K.; Sugiyama, E.; Saito, Y. Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: Implications in the clinical trials for novel drug development. *Drug Metab. Pharmacokinet.* **2012**, *27*, 9–54. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Ieiri, I. Functional significance of genetic polymorphisms in P-glycoprotein (MDR1, ABCB1) and breast cancer resistance protein (BCRP, ABCG2). *Drug Metab. Pharmacokinet.* **2012**, *27*, 85–105. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Zhang, W.; Sun, S.; Zhang, W.; Shi, Z. Polymorphisms of ABCG2 and its impact on clinical relevance. *Biochem. Biophys. Res. Commun.* **2018**, *503*, 408–413. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Wang, H.; Hao, B.; Zhou, K.; Chen, X.; Wu, S.; Zhou, G.; Zhu, Y.; He, F. Linkage disequilibrium and haplotype architecture for two ABC transporter genes (ABCC1 and ABCG2) in Chinese population: Implications for pharmacogenomic association studies. *Ann. Hum. Genet.* **2004**, *68*, 563–573. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Lee, S.S.; Jeong, H.E.; Yi, J.M.; Jung, H.J.; Jang, J.E.; Kim, E.Y.; Lee, S.J.; Shin, J.G. Identification and functional assessment of BCRP polymorphisms in a Korean population. *Drug Metab. Dispos.* **2007**, *35*, 623–632. [\[CrossRef\]](#)
26. Maekawa, K.; Itoda, M.; Sai, K.; Saito, Y.; Kaniwa, N.; Shirao, K.; Hamaguchi, T.; Kunitoh, H.; Yamamoto, N.; Tamura, T.; et al. Genetic variation and haplotype structure of the ABC transporter gene ABCG2 in a Japanese population. *Drug Metab. Pharmacokinet.* **2006**, *21*, 109–121. [\[CrossRef\]](#)
27. Backstrom, G.; Taipalensuu, J.; Melhus, H.; Brandstrom, H.; Svensson, A.C.; Artursson, P.; Kindmark, A. Genetic variation in the ATP-binding cassette transporter gene ABCG2 (BCRP) in a Swedish population. *Eur. J. Pharm. Sci.* **2003**, *18*, 359–364. [\[CrossRef\]](#)
28. Bosch, T.M.; Kjellberg, L.M.; Bouwers, A.; Koeleman, B.P.; Schellens, J.H.; Beijnen, J.H.; Smits, P.H.; Meijerman, I. Detection of single nucleotide polymorphisms in the ABCG2 gene in a Dutch population. *Am. J. Pharm.* **2005**, *5*, 123–131. [\[CrossRef\]](#)
29. Salagacka-Kubiak, A.; Zebrowska, M.; Wosiak, A.; Balcerczak, M.; Mirowski, M.; Balcerczak, E. ABCG2 in peptic ulcer: Gene expression and mutation analysis. *J. Appl. Genet.* **2016**, *57*, 335–342. [\[CrossRef\]](#)
30. Niebudek, K.; Balcerczak, E.; Mirowski, M.; Pietrzak, J.; Zawadzka, I.; Zebrowska-Nawrocka, M. The contribution of ABCG2 G34A and C421A polymorphisms to multiple myeloma susceptibility. *Onco. Targets Ther.* **2019**, *12*, 1655–1660. [\[CrossRef\]](#)
31. Strapagiel, D.; Sobalska, M.; Słomka, M.; Marciniak, B. Biobank Lodz-DNA Based Biobank at the University of Lodz, Poland. *Open J. Bioresour.* **2016**, *3*, e6. [\[CrossRef\]](#)
32. Dobrowolska, S.; Michalska-Madej, J.; Słomka, M.; Sobalska-Kwapis, M.; Strapagiel, D. Biobank Łódź-population based biobank at the University of Łódź, Poland. *Eur. J. Transl. Clin. Med.* **2019**, *2*, 85–95. [\[CrossRef\]](#)
33. Strapagiel, D.; Majewska, M.; Słomka, M.; Janik, K.; Sobalska, M.; Bartosz, G. Method for Determination of Gender. Patent Number: PL 232652 B1, 31 July 2019.
34. Słomka, M.; Sobalska-Kwapis, M.; Korycka-Machala, M.; Bartosz, G.; Dziadek, J.; Strapagiel, D. Genetic variation of the ABC transporter gene ABCC1 (Multidrug resistance protein 1-MRP1) in the Polish population. *BMC Genet.* **2015**, *16*, 114. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Słomka, M.; Sobalska-Kwapis, M.; Wachulec, M.; Bartosz, G.; Strapagiel, D. High Resolution Melting (HRM) for High-Throughput Genotyping-Limitations and Caveats in Practical Case Studies. *Int. J. Mol. Sci.* **2017**, *18*, 2316. [\[CrossRef\]](#)
36. Wang, N.; Akey, J.M.; Zhang, K.; Chakraborty, R.; Jin, L. Distribution of recombination crossovers and the origin of haplotype blocks: The interplay of population history, recombination, and mutation. *Am. J. Hum. Genet.* **2002**, *71*, 1227–1234. [\[CrossRef\]](#)

37. Fischer, S.; Lakatos, P.L.; Lakatos, L.; Kovacs, A.; Molnar, T.; Altorjay, I.; Papp, M.; Szilvasi, A.; Tulassay, Z.; Osztoivits, J.; et al. ATP-binding cassette transporter ABCG2 (BCRP) and ABCB1 (MDR1) variants are not associated with disease susceptibility, disease phenotype response to medical therapy or need for surgery in Hungarian patients with inflammatory bowel diseases. *Scand. J. Gastroenterol.* **2007**, *42*, 726–733. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Stiburkova, B.; Pavelcova, K.; Zavada, J.; Petru, L.; Simek, P.; Cepek, P.; Pavlikova, M.; Matsuo, H.; Merriman, T.R.; Pavelka, K. Functional non-synonymous variants of ABCG2 and gout risk. *Rheumatology* **2017**, *56*, 1982–1992. [\[CrossRef\]](#)
39. Muller, P.J.; Dally, H.; Klappenecker, C.N.; Edler, L.; Jager, B.; Gerst, M.; Spiegelhalter, B.; Tuengerthal, S.; Fischer, J.R.; Drings, P.; et al. Polymorphisms in ABCG2, ABCC3 and CNT1 genes and their possible impact on chemotherapy outcome of lung cancer patients. *Int. J. Cancer* **2009**, *124*, 1669–1674. [\[CrossRef\]](#)
40. Stark, K.; Reinhard, W.; Grassl, M.; Erdmann, J.; Schunkert, H.; Illig, T.; Hengstenberg, C. Common polymorphisms influencing serum uric acid levels contribute to susceptibility to gout, but not to coronary artery disease. *PLoS ONE* **2009**, *4*, e7729. [\[CrossRef\]](#)
41. Zhou, D.; Liu, Y.; Zhang, X.; Gu, X.; Wang, H.; Luo, X.; Zhang, J.; Zou, H.; Guan, M. Functional polymorphisms of the ABCG2 gene are associated with gout disease in the Chinese Han male population. *Int. J. Mol. Sci.* **2014**, *15*, 9149–9159. [\[CrossRef\]](#)
42. Tanaka, M.; Kamada, I.; Takahashi, J.; Kimura, K.; Matsukura, H.; Tani, Y. Defining the Jr(a-) phenotype in the Japanese population. *Transfusion* **2014**, *54*, 412–417. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Zámbo, B.; Móznér, O.; Bartos, Z.; Török, G.; Várady, G.; Telbisz, Á.; Homolya, L.; Orbán, T.I.; Sarkadi, B. Cellular expression and function of naturally occurring variants of the human ABCG2 multidrug transporter. *Cell. Mol. Life Sci.* **2020**, *77*, 365–378. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Hira, D.; Terada, T. BCRP/ABCG2 and high-alert medications: Biochemical, pharmacokinetic, pharmacogenetic, and clinical implications. *Biochem. Pharmacol.* **2018**, *147*, 201–210. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Takahashi, N.; Miura, M.; Scott, S.A.; Kagaya, H.; Kameoka, Y.; Tagawa, H.; Saitoh, H.; Fujishima, N.; Yoshioka, T.; Hirokawa, M.; et al. Influence of CYP3A5 and drug transporter polymorphisms on imatinib trough concentration and clinical response among patients with chronic phase chronic myeloid leukemia. *J. Hum. Genet* **2010**, *55*, 731–737. [\[CrossRef\]](#)
46. Takahashi, N.; Miura, M. Therapeutic drug monitoring of imatinib for chronic myeloid leukemia patients in the chronic phase. *Pharmacology* **2011**, *87*, 241–248. [\[CrossRef\]](#)
47. Petain, A.; Kattygnarath, D.; Azard, J.; Chatelut, E.; Delbardo, C.; Geoerger, B.; Barrois, M.; Séronie-Vivien, S.; LeCesne, A.; Vassal, G. Population pharmacokinetics and pharmacogenetics of imatinib in children and adults. *Clin. Cancer Res.* **2008**, *14*, 7102–7109. [\[CrossRef\]](#)
48. Ghafouri, H.; Ghaderi, B.; Amini, S.; Nikkhoo, B.; Abdi, M.; Hoseini, A. Association of ABCB1 and ABCG2 single nucleotide polymorphisms with clinical findings and response to chemotherapy treatments in Kurdish patients with breast cancer. *Tumour. Biol.* **2016**, *37*, 7901–7906. [\[CrossRef\]](#)
49. Wu, H.; Liu, Y.; Kang, H.; Xiao, Q.; Yao, W.; Zhao, H.; Wang, E.; Wei, M. Genetic Variations in ABCG2 Gene Predict Breast Carcinoma Susceptibility and Clinical Outcomes after Treatment with Anthracycline-Based Chemotherapy. *BioMed Res. Int.* **2015**, *2015*, 279109. [\[CrossRef\]](#)
50. Liu, X.; Fiocco, M.; Swen, J.J.; Guchelaar, H.J. Assessment of ethnic differences in sunitinib outcome between Caucasian and Asian patients with metastatic renal cell carcinoma: A meta-analysis. *Acta Oncol.* **2017**, *56*, 582–589. [\[CrossRef\]](#)
51. Chen, C.J.; Tseng, C.C.; Yen, J.H.; Chang, J.G.; Chou, W.C.; Chu, H.W.; Chang, S.J.; Liao, W.T. ABCG2 contributes to the development of gout and hyperuricemia in a genome-wide association study. *Sci. Rep.* **2018**, *8*, 3137. [\[CrossRef\]](#)
52. Zhang, L.; Spencer, K.L.; Voruganti, V.S.; Jorgensen, N.W.; Fornage, M.; Best, L.G.; Brown-Gentry, K.D.; Cole, S.A.; Crawford, D.C.; Deelman, E.; et al. Association of functional polymorphism rs2231142 (Q141K) in the ABCG2 gene with serum uric acid and gout in 4 US populations: The PAGE Study. *Am. J. Epidemiol.* **2013**, *177*, 923–932. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Yamagishi, K.; Tanigawa, T.; Kitamura, A.; Köttgen, A.; Folsom, A.R.; Iso, H. The rs2231142 variant of the ABCG2 gene is associated with uric acid levels and gout among Japanese people. *Rheumatology* **2010**, *49*, 1461–1465. [\[CrossRef\]](#) [\[PubMed\]](#)

54. Yu, K.H.; Chang, P.Y.; Chang, S.C.; Wu-Chou, Y.H.; Wu, L.A.; Chen, D.P.; Lo, F.S.; Lu, J.J. A comprehensive analysis of the association of common variants of ABCG2 with gout. *Sci. Rep.* **2017**, *7*, 9988. [[CrossRef](#)] [[PubMed](#)]
55. Toyoda, Y.; Mančíková, A.; Krylov, V.; Morimoto, K.; Pavelcová, K.; Bohatá, J.; Pavelka, K.; Pavlíková, M.; Suzuki, H.; Matsuo, H.; et al. Functional Characterization of Clinically-Relevant Rare Variants in ABCG2 Identified in a Gout and Hyperuricemia Cohort. *Cells* **2019**, *8*, 363. [[CrossRef](#)] [[PubMed](#)]
56. Haer-Wigman, L.; Ait Soussan, A.; Ligthart, P.; de Haas, M.; van der Schoot, C.E. Molecular analysis of immunized Jr(a-) or Lan- patients and validation of a high-throughput genotyping assay to screen blood donors for Jr(a-) and Lan- phenotypes. *Transfusion* **2014**, *54*, 1836–1846. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

SUPLEMENT DO ZAŁĄCZNIKA NR 2

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G., Strapagiel, D., 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. *Genes (Basel)*, 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

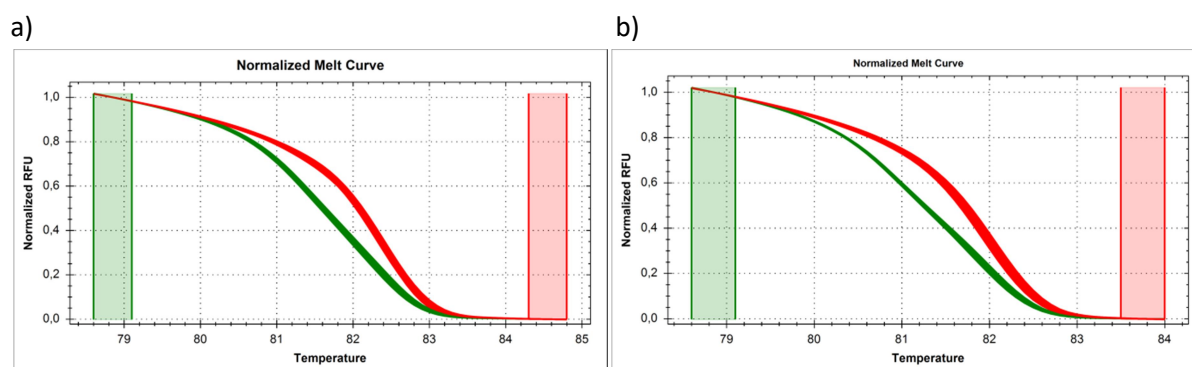


Figure S1 Summary of *ABCG2* exon 2 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89061202 - 89061042, including primers).

Table S1 Summary of *ABCG2* exon 2 (the first area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.34G>A	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	2	4
Samples in cluster 1 (red)	GG	86	89	175
Samples in cluster 2 (green)	GA	7	5	12
Samples excluded		2	1	3
Scan rate		-	-	0.9842

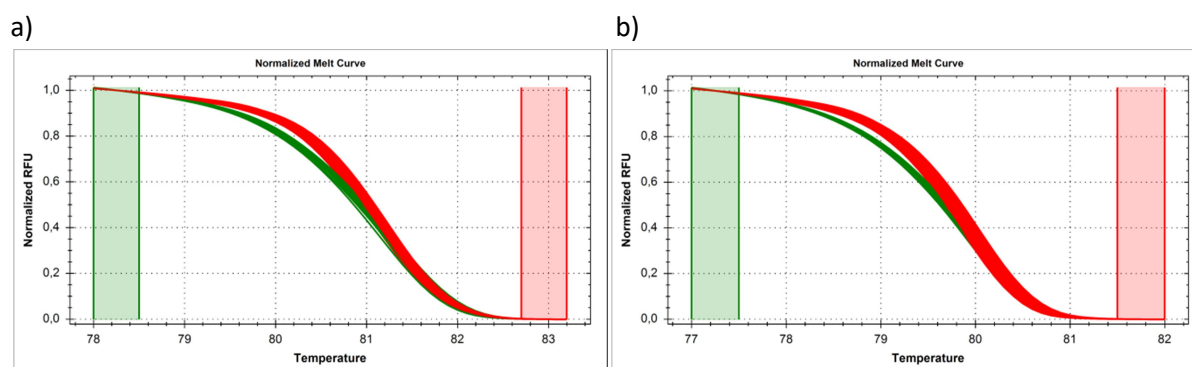


Figure S2 Summary of *ABCG2* exon 2 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_000004.11: 89061082 - 89060870, including primers).

Table S2 Summary of *ABCG2* exon 2 (the second area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.203+36A>G	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	2	4
Samples in cluster 1 (red)	AA	86	89	175
Samples in cluster 2 (green)	AG	9	5	14
Samples excluded		0	1	1
Scan rate		-	-	0.9947

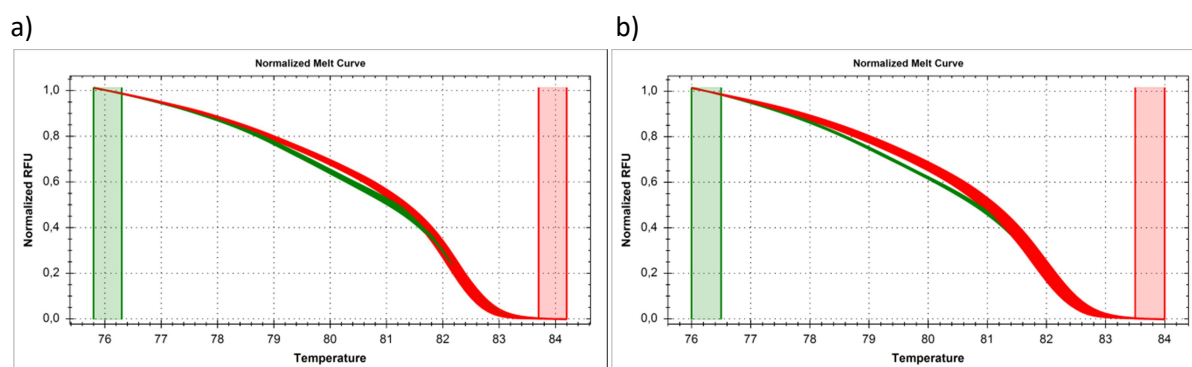


Figure S3 Summary of *ABCG2* exon 3 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89053837 - 89053685, including primers).

Table S3 Summary of *ABCG2* exon 3 scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.263+10A>G	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	2	4
Samples in cluster 1 (red)	AA	86	91	177
Samples in cluster 2 (green)	AG	7	3	10
Samples excluded		2	1	3
Scan rate		-	-	0.9842

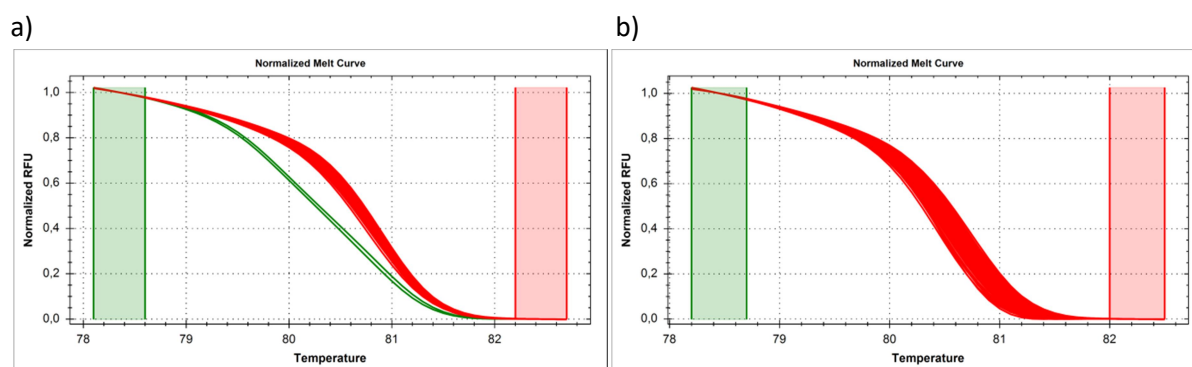
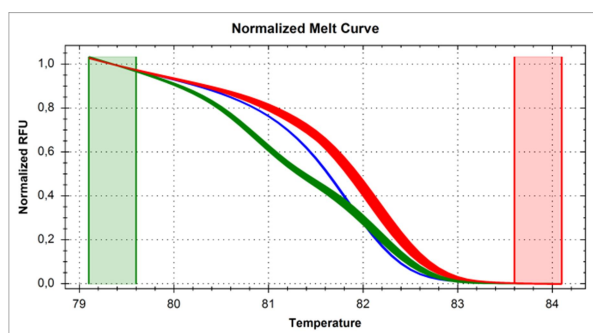


Figure S4 Summary of *ABCG2* exon 4 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89053112 - 89052923, including primers).

Table S4 Summary of *ABCG2* exon 4 scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.335C>A	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	1	3
Samples in cluster 1 (red)	CC	92	92	184
Samples in cluster 2 (green)	CA	1	0	1
Samples excluded		2	3	5
Scan rate		-	-	0.9737

a)



b)

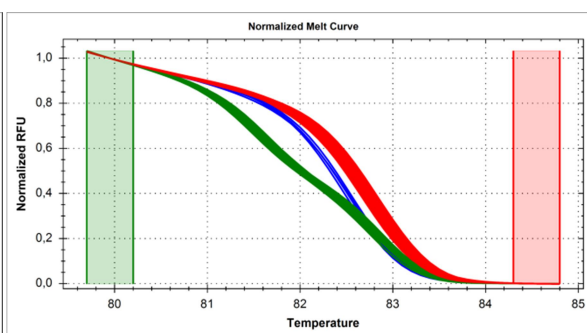
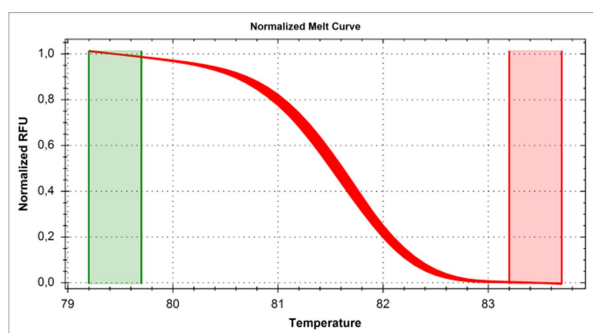


Figure S5 Summary of *ABCG2* exon 5 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89052424 - 89052251, including primers)

Table S5 Summary of *ABCG2* exon 5 (the first area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.421C>A	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		3	3	6
Samples in cluster 1 (red)	CC	79	68	147
Samples in cluster 2 (green)	CA	15	23	38
Samples in cluster 3 (blue)	AA	1	2	3
Samples excluded		0	2	2
Scan rate		-	-	0.9895

a)



b)

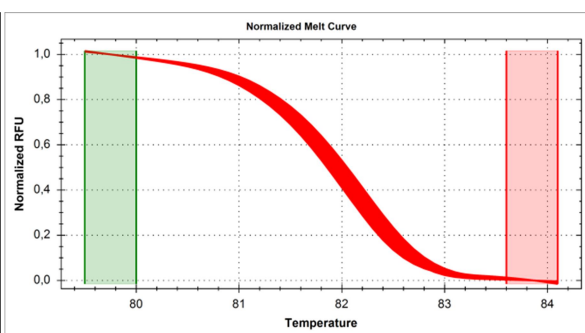


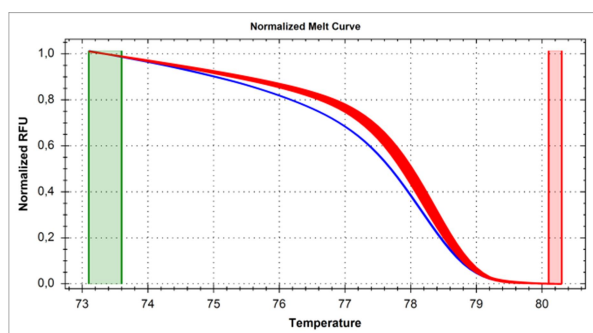
Figure S6 Summary of *ABCG2* exon 5 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_000004.11: 89052332 - 89052155, including primers)

Table S6 Summary of *ABCG2* exon 5 (the second area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	94	187
Samples excluded	2	1	3
Scan rate	-	-	0.9842

wild type

a)



b)

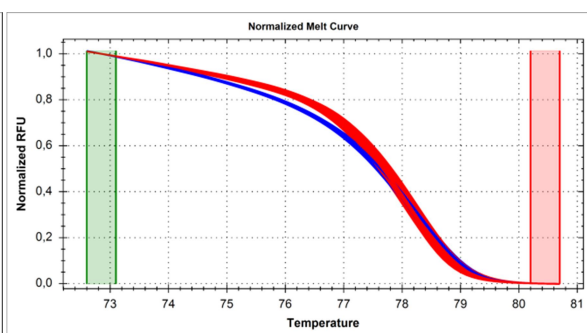
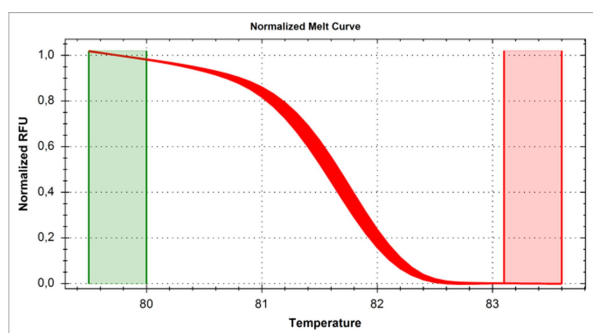


Figure S7 Summary of *ABCG2* exon 6 scanning by HRM: a), b) melting for the first scanned area of the exon (NC_000004.11: 89042992 - 89042878, including primers).

Table S7 Summary of *ABCG2* exon 6 (the first area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.532-16A>G	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	2	4
Samples in cluster 1 (red)	AA	92	82	174
Samples in cluster 2 (blue)	AG	1	12	13
Samples excluded		2	1	3
Scan rate		-	-	0.9842

a)



b)

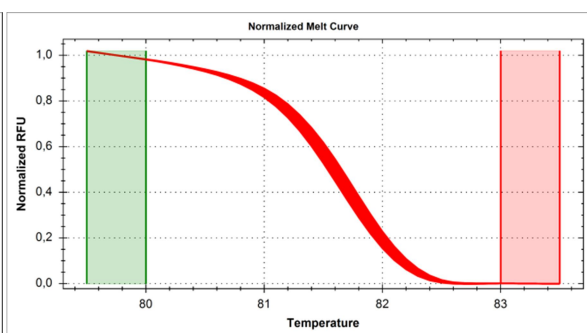


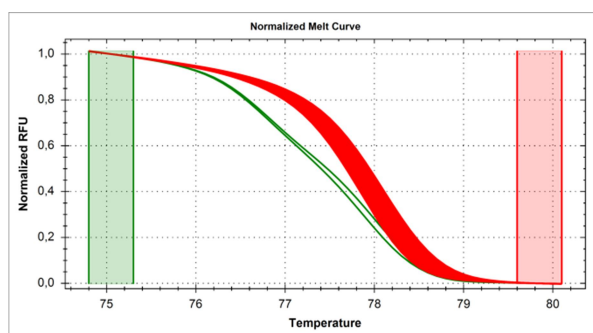
Figure S8 Summary of *ABCG2* exon 6 scanning by HRM: a), b) melting for the second scanned area of the exon (NC_000004.11: 89042932 - 89042718, including primers).

Table S8 Summary of *ABCG2* exon 6 (the second area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	92	94	186
Samples excluded	3	1	4
Scan rate	-	-	0.9789

wild type

a)



b)

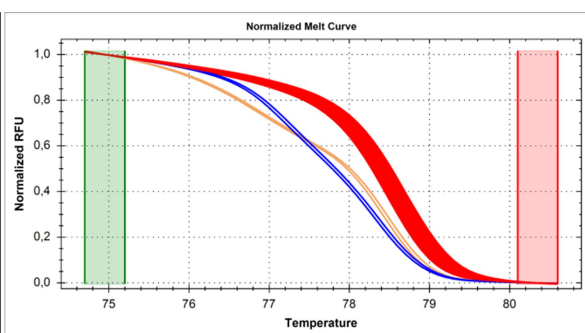
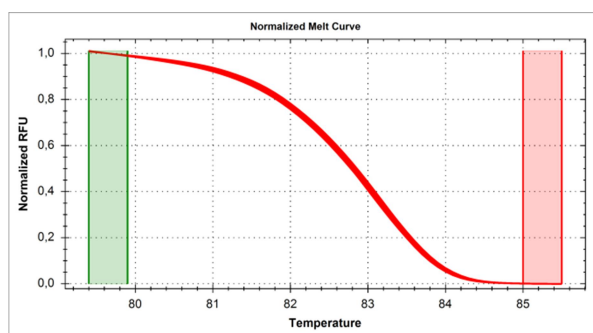


Figure S9 Summary of *ABCG2* exon 7 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89039490 - 89039362, including primers).

Table S9 Summary of *ABCG2* exon 7 (the first area) scanning by HRM. Variants position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.706C>T c.706C>A	c.690-19_690- 17delTGT	Plate 1	Plate 2	Overall
Samples scanned			95	95	190
Melting clusters			2	3	5
Samples in cluster 1 (red)	CC	TGT	92	92	184
Samples in cluster 2 (green)	CT	TGT	1	0	1
Samples in cluster 2 (blue)	CA	TGT	0	1	1
Samples in cluster 2 (orange)	CC	delTGT	0	1	1
Samples excluded			2	1	3
Scan rate			-	-	0.9842

a)



b)

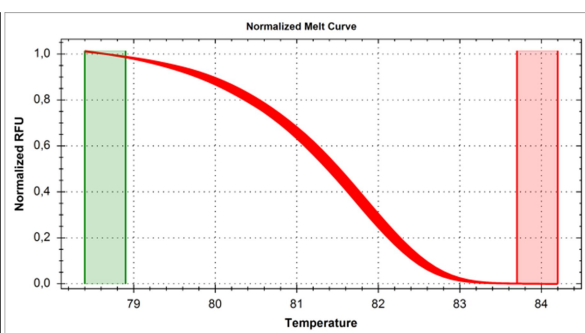


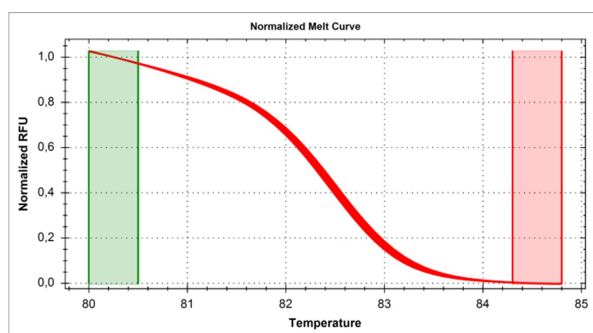
Figure S10 Summary of *ABCG2* exon 7 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_000004.11: 89039403 - 89039295, including primers).

Table S10 Summary of *ABCG2* exon 7 (the second area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	93	188
Samples excluded	0	2	2
Scan rate	-	-	0.9895

wild type

a)



b)

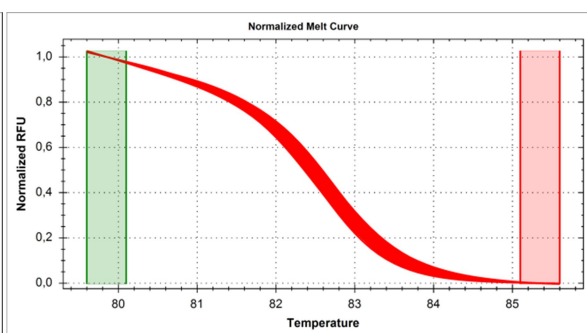


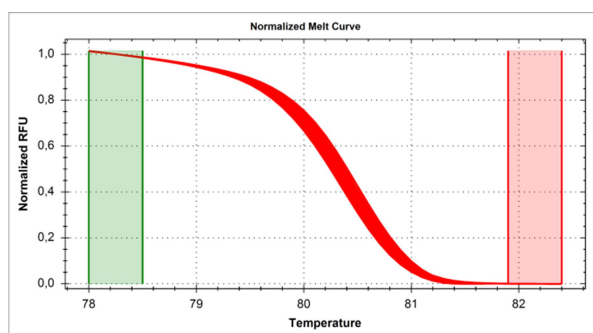
Figure S11 Summary of *ABCG2* exon 7 scanning by HRM: a), b) melting plots for the third scanned area of the exon (NC_000004.11: 89039343 - 89039219, including primers).

Table S11 Summary of *ABCG2* exon 7 (the third area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	94	187
Samples excluded	2	1	3
Scan rate	-	-	0.9842

wild type

a)



b)

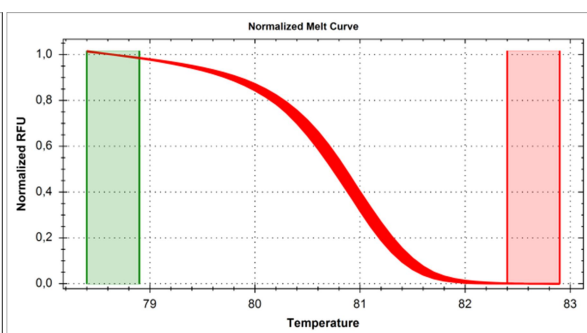


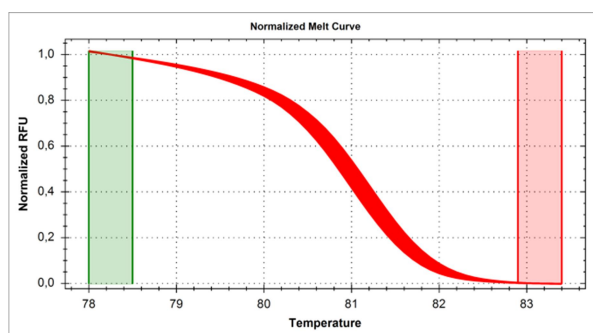
Figure S12 Summary of *ABCG2* exon 8 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89036264 - 89036074, including primers).

Table S12 Summary of *ABCG2* exon 8 scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	94	187
Samples excluded	2	1	3
Scan rate	-	-	0.9842

wild type

a)



b)

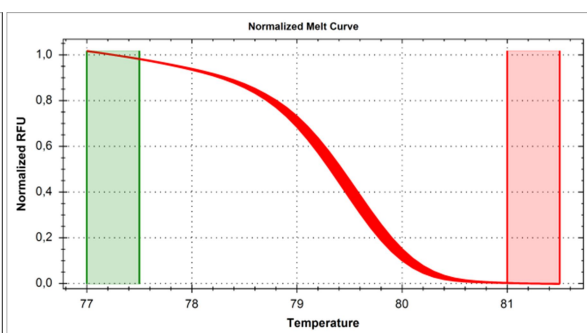


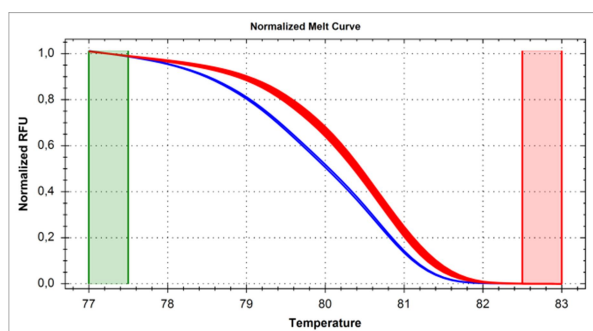
Figure S13 Summary of *ABCG2* exon 9 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89034751 - 89034632, including primers).

Table S13 Summary of *ABCG2* exon 9 (the first area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Scan rate	-	-	0.9947

wild type

a)



b)

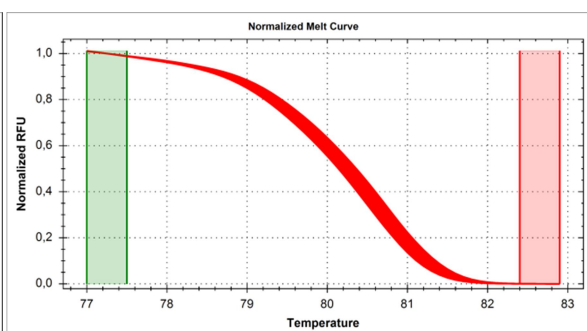
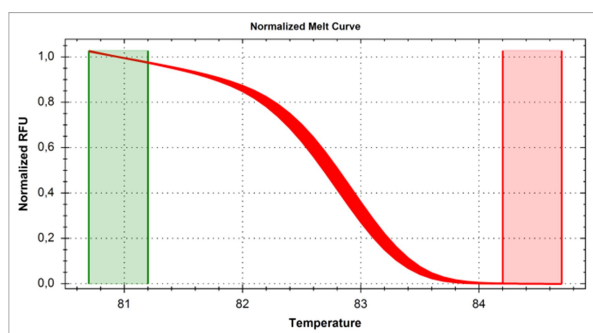


Figure S14 Summary of *ABCG2* exon 9 scanning by HRM: a), b) melting for the second scanned area of the exon (NC_000004.11: 89034684 - 89034545, including primers).

Table S14 Summary of *ABCG2* exon 9 (the second area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1060G>A	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	1	3
Samples in cluster 1 (red)	GG	94	93	187
Samples in cluster 2 (blue)	GA	1	0	1
Samples excluded		0	2	2
Scan rate		-	-	0.9895

a)



b)

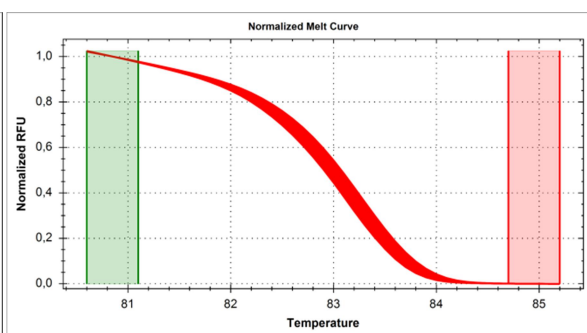


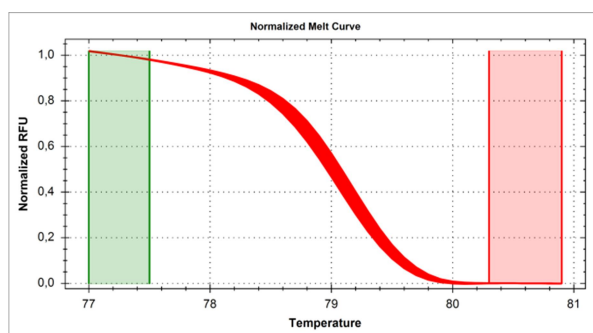
Figure S15 Summary of *ABCG2* exon 9 scanning by HRM: a), b) melting plots for the third scanned area of the exon (NC_000004.11: 89034606 - 89034408, including primers).

Table S15 Summary of *ABCG2* exon 9 (the third area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	94	187
Samples excluded	2	1	3
Scan rate	-	-	0.9842

wild type

a)



b)

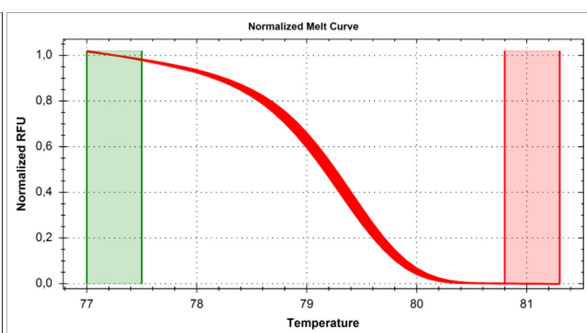


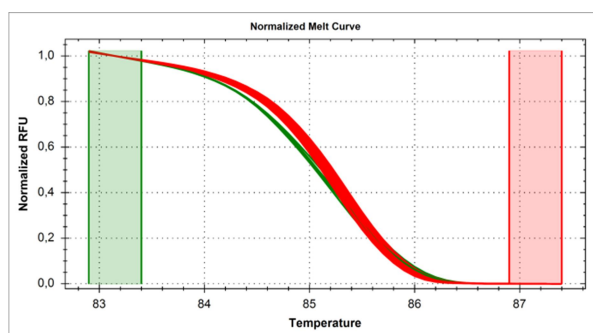
Figure S16 Summary of *ABCG2* exon 10 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89028467 - 89028287, including primers).

Table S16 Summary of *ABCG2* exon 10 scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	94	187
Samples excluded	2	1	3
Scan rate	-	-	0.9842

wild type

a)



b)

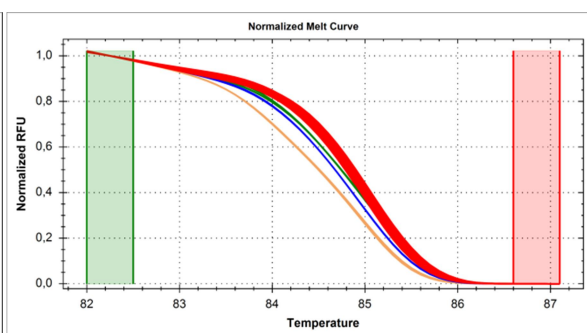
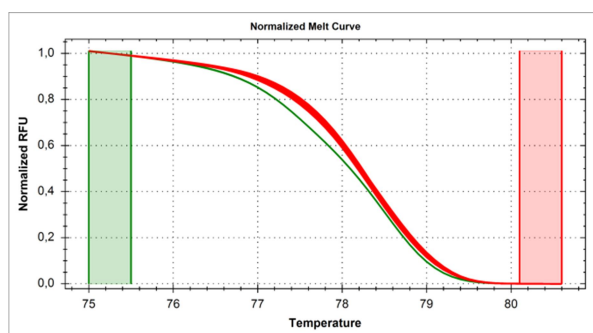


Figure S17 Summary of *ABCG2* exon 11 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89022553 - 89022325, including primers).

Table S17 Summary of *ABCG2* exon 11 scanning by HRM. Variants position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1278-28G>A	c.1302G>A	c.1367+20G>A	Plate 1	Plate 2	Overall
Samples scanned				95	95	190
Melting clusters				2	4	6
Samples in cluster 1 (red)	GG	GG	AA	90	85	175
Samples in cluster 2 (green)	GG	GG	GA	5	6	11
Samples in cluster 2 (blue)	GA	GG	AA	0	1	1
Samples in cluster 2 (orange)	GG	GA	AA	0	1	1
Samples excluded				0	2	2
Scan rate				-	-	0.9895

a)



b)

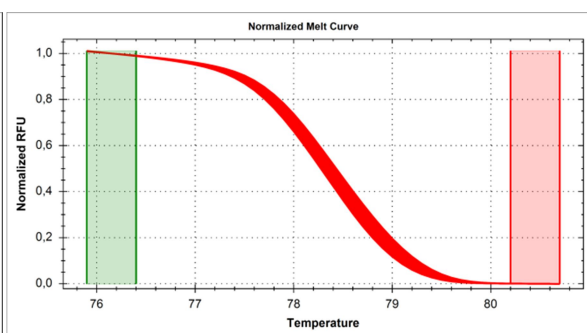
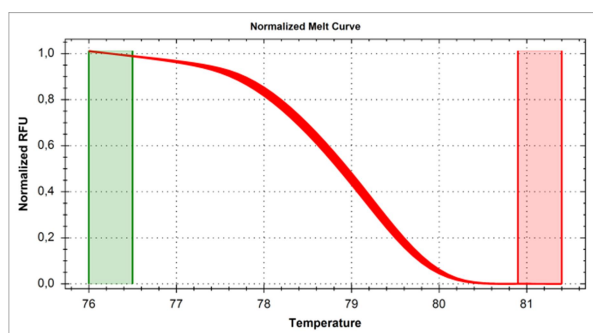


Figure S18 Summary of *ABCG2* exon 12 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89020657 - 89020511, including primers).

Table S18 Summary of *ABCG2* exon 12 (the first area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1368-21A>T	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		2	1	3
Samples in cluster 1 (red)	AA	94	94	188
Samples in cluster 2 (green)	AT	1	0	1
Samples excluded		0	1	1
Scan rate		-	-	0.9947

a)



b)

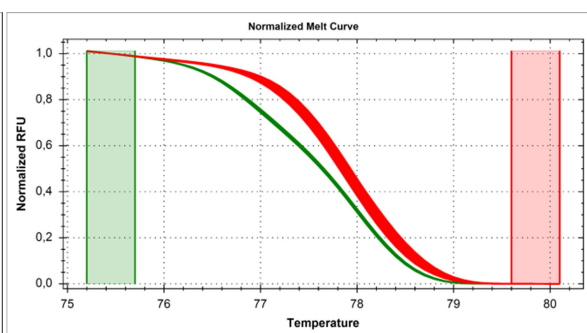
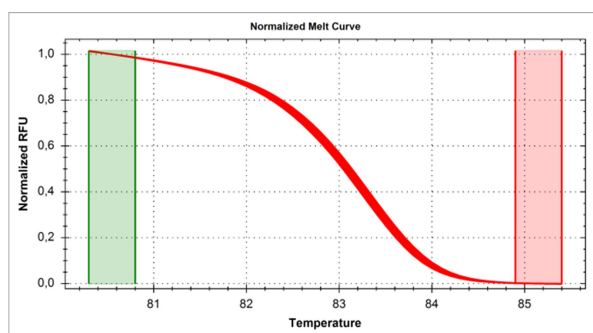


Figure S19 Summary of *ABCG2* exon 12 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_000004.11: 89020589 - 89020416, including primers).

Table S19 Summary of *ABCG2* exon 12 (the second area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1492+38G>A	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		1	2	3
Samples in cluster 1 (red)	GG	95	92	187
Samples in cluster 2 (green)	GA	0	2	2
Samples excluded		0	1	1
Scan rate		-	-	0.9947

a)



b)

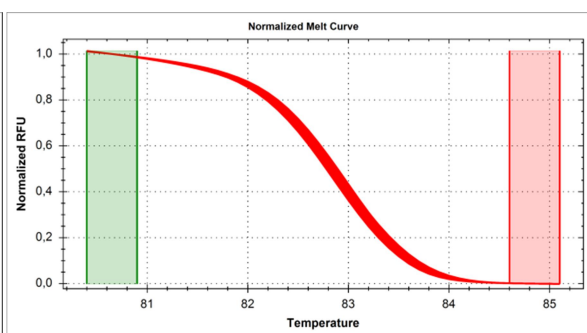


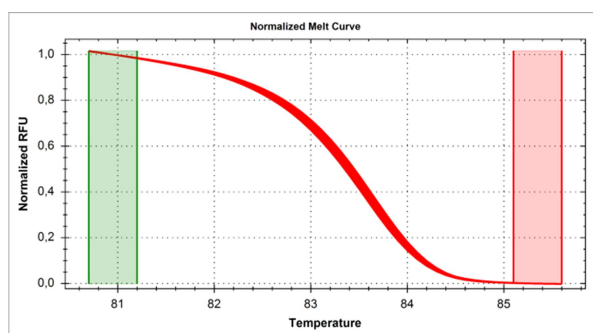
Figure S20 Summary of *ABCG2* exon 13 scanning by HRM: a), b) melting plots for the first scanned area of the exon (NC_000004.11: 89018805 - 89018650, including primers).

Table S20 Summary of *ABCG2* exon 13 (the first area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	94	189
Samples excluded	0	1	1
Scan rate	-	-	0.9947

wild type

a)



b)

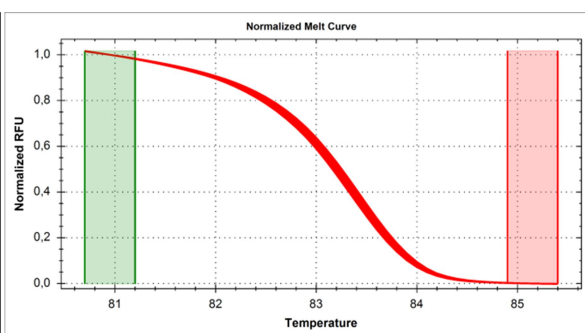


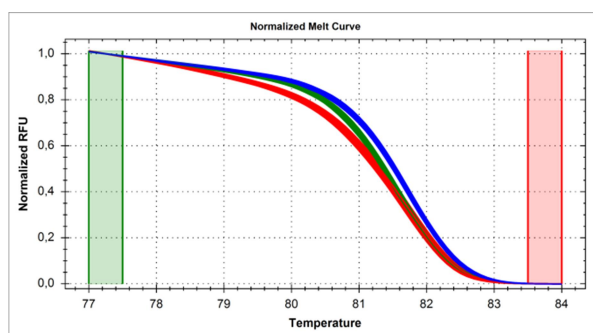
Figure S21 Summary of *ABCG2* exon 13 scanning by HRM: a), b) melting plots for the second scanned area of the exon (NC_000004.11: 89018723 - 89018579, including primers).

Table S21 Summary of *ABCG2* exon 13 (the second area) scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	95	93	188
Samples excluded	0	2	2
Scan rate	-	-	0.9895

wild type

a)



b)

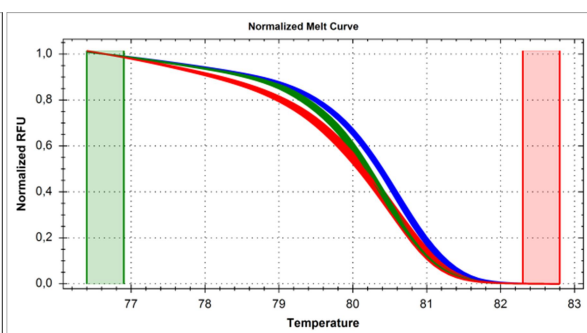
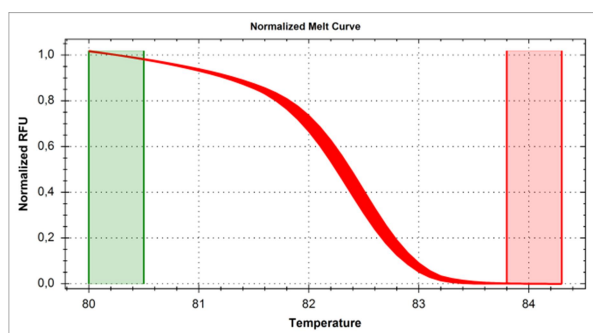


Figure S22 Summary of *ABCG2* exon 13 scanning by HRM: a), b) melting plots for the third scanned area of the exon (NC_000004.11: 89018668 - 89018542, including primers)

Table S22 Summary of *ABCG2* exon 13 (the third area) scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1647+40T>C	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		3	3	6
Samples in cluster 1 (red)	TC	48	45	93
Samples in cluster 2 (green)	TT	30	38	68
Samples in cluster 3 (blue)	CC	17	10	27
Samples excluded		0	2	2
Scan rate		-	-	0.9895

a)



b)

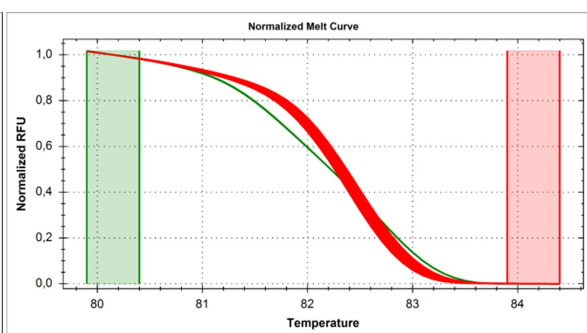
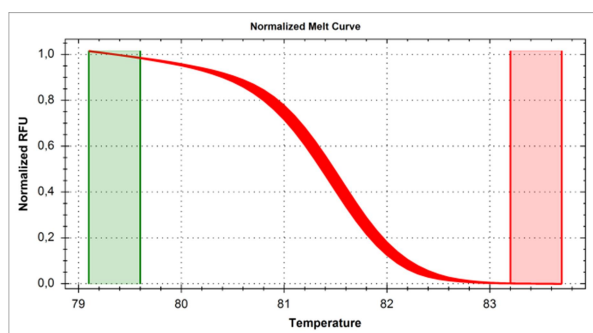


Figure S23 Summary of *ABCG2* exon 14 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89016794 - 89016623, including primers).

Table S23 Summary of *ABCG2* exon 14 scanning by HRM. Variant position according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	c.1714A>C	Plate 1	Plate 2	Overall
Samples scanned		95	95	190
Melting clusters		1	2	3
Samples in cluster 1 (red)	AA	93	93	186
Samples in cluster 2 (green)	AC	0	1	1
Samples excluded		2	1	3
Scan rate		-	-	0.9842

a)



b)

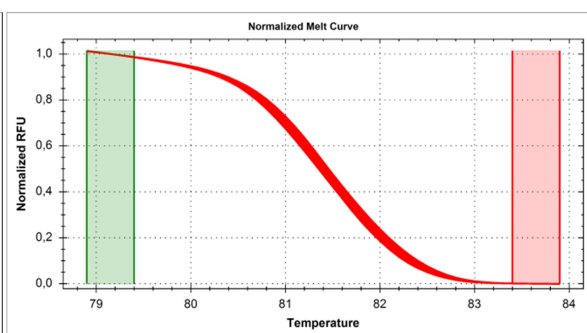


Figure S24 Summary of *ABCG2* exon 15 scanning by HRM: a), b) melting plots for the whole exon scanned area (NC_000004.11: 89015846 - 89015696, including primers).

Table S24 Summary of *ABCG2* exon 15 scanning by HRM. No variant detected according to reference sequence NM_004827.3. Scan rate is ratio between positive clustered and verified samples to all the scanned samples.

	Plate 1	Plate 2	Overall
Samples scanned	95	95	190
Melting clusters	1	1	2
Samples in cluster 1 (red)	93	89	182
Samples excluded	2	6	8
Scan rate	-	-	0.9579

wild type

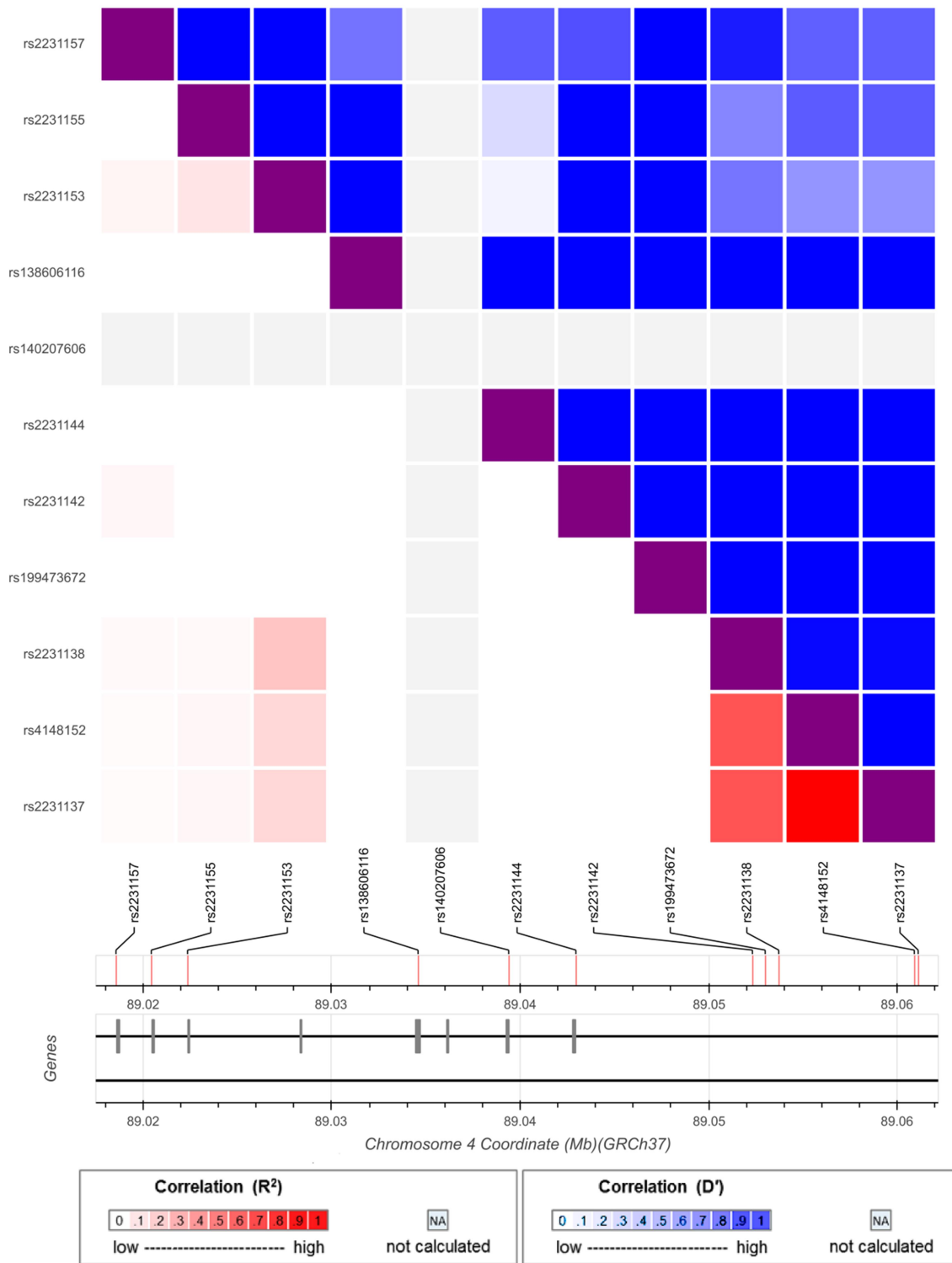
Figure S25. Linkage disequilibrium statistics r^2 and $|D'|$ from LDlink for 1000 Genomes Project database, for Europeans (a) and average world population (b).

Due to absence in database some rare variants, including detected in this study: c.690-19_690-17delTGT, c.1278-28G>A, c.1302G>A (p.Thr434=), c.1368-21A>T, c.1714A>C (p.Ser572Arg), they were omitted in this comparison.

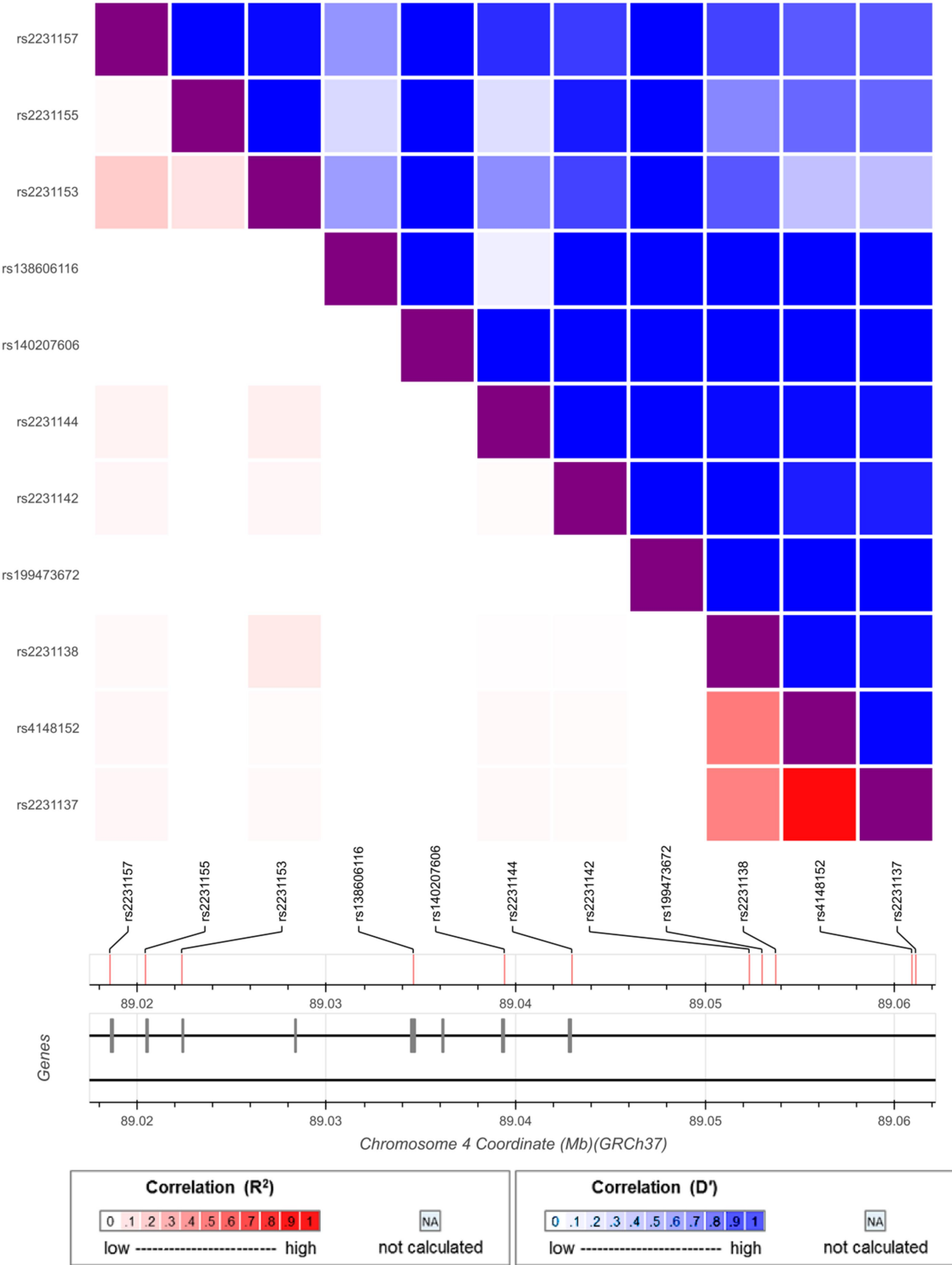
Figure S26. Network analysis of *ABCG2* haplotypes detected in this study for block 1 (a) and block 2 (b).

Each circle represents particular haplotype and its area corresponds with defined frequency. Haplotypes description base on numeration from Table S26. Variants determining subsequent haplotypes marked in each connection, htSNPs are bolded.

a)



b)



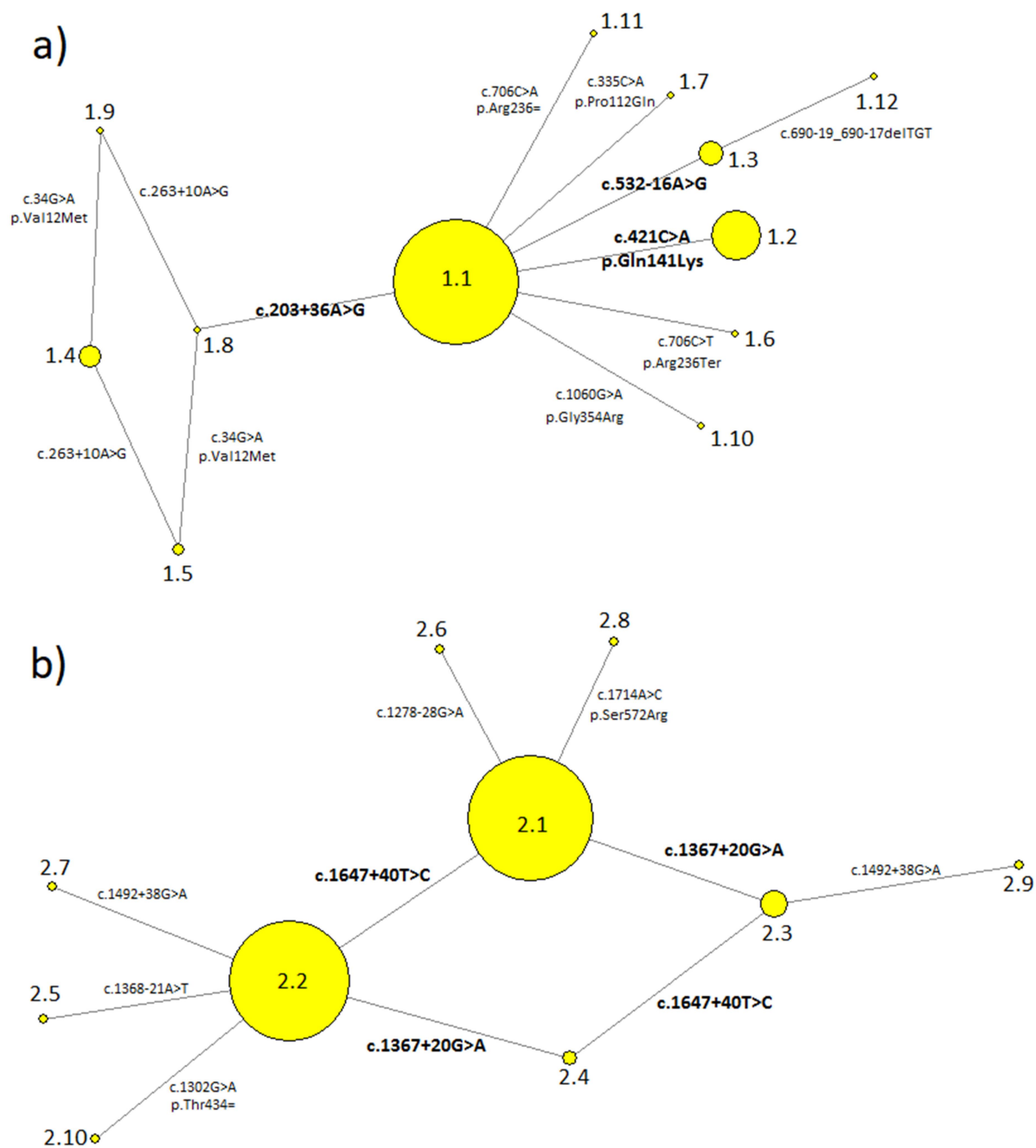


Table S25 Nonsynonymous variants detected in this study and their predicted effect on protein functioning classified by *in silico* tools

	SIFT		PolyPhen-2 HumDiv		PolyPhen-2 HumVar		REVEL		MetaLR	
c.34G>A (p.Val12Met)	1	tolerated	0	benign	0.001	benign	0.180	likely benign	0	tolerated
c.335C>A (p.Pro112Gln)	0	deleterious	0.995	probably damaging	0.978	probably damaging	0.425	likely benign (suspected)	0.774	damaging
c.421C>A (p.Gln141Lys)	0.19	tolerated	0.372	benign	0.216	benign	0.106	likely benign	0	tolerated
c.1060G>A (p.Gly354Arg)	0.53	tolerated	0.016	benign	0.006	benign	0.088	likely benign	0.191	tolerated
c.1714A>C (p.Ser572Arg)	0	deleterious	0.997	probably damaging	0.989	probably damaging	0.897	likely disease causing	0.687	damaging

SIFT version 5.2.2: a score < 0.05 are called 'deleterious' and all others are called 'tolerated'

PolyPhen-2 v2.2.2r398 HumDiv train model: 0 - 0.452 (benign) < 0.453 – 0.956 (possibly damaging) < 0.957 - 1 (probably damaging)

PolyPhen-2 v2.2.2r398 HumVar train model: 0 - 0.446 (benign) < 0.447 – 0.908 (possibly damaging) < 0.909 - 1 (probably damaging)

REVEL - scores above 0.5 called as 'likely disease causing' and scores below 0.5 as 'likely benign' (recommended the actual score is used when assessing a variant and a cut-off appropriate to chosen requirements)

MetaLR - scores above 0.5 called as 'damaging' and scores below 0.5 as 'tolerated'

Table S26 All the observed haplotypes in this study and their frequencies in *ABCG2* blocks

Block 1			Block 2		
Name	Haplotype	Frequency	Name	Haplotype	Frequency
1.1	GAACCACC(TGT)G	0.804	2.1	GGAAGTA	0.584
1.2	GAACAACC(TGT)G	0.113	2.2	GGAAGCA	0.376
1.3	GAACCGCC(TGT)G	0.031	2.3	GGGAGTA	0.020
1.4	AGGCCACC(TGT)G	0.024	2.4	GGGAGCA	0.006
1.5	AGACCACC(TGT)G	0.009	2.5	GGATGCA	0.003
1.6	GAACCATC(TGT)G	0.003	2.6	AGAAGTA	0.003
1.7	GAAACACC(TGT)G	0.003	2.7	GGAAACA	0.003
1.8	GGACCACC(TGT)G	0.003	2.8	GGAAGTC	0.003
1.9	GGGCCACC(TGT)G	0.003	2.9	GGGAATA	0.003
1.10	GAACCACC(TGT)A	0.003	2.10	GAAAGCA	0.003
1.11	GAACCACA(TGT)G	0.003			
1.12	GAACCGCC(delTGT)G	0.003			

OŚWIADCZENIA WSPÓŁAUTORÓW O UDZIALE W PUBLIKACJACH

Oświadczenie współautorów o udziale w publikacji

mgr Marcin Słomka

Pracownia Biobank, Katedra Biologii Nowotworów i Epigenetyki

Wydział Biologii i Ochrony Środowiska

ul. Pomorska 139, 90-235 Łódź

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Bartosz, G., Dziadek, J. Strapagiel, D., 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. *BMC Genet.*, 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 60% i polegał na:

Współtworzeniu koncepcji pracy i projektowaniu eksperymentu skanowania genu metodą HRM; wykonaniu prac eksperymentalnych polegających na opracowaniu metody skanowania *ABCC1* metodą HRM, wykonaniu skanowania tego genu, przygotowania próbek do sekwencjonowania Sangera; analizie wyników skanowania metodą HRM, sekwencjonowania Sangera oraz opracowaniu danych (porównanie dwóch podejść wykorzystujących HRM czyli skanowania genu i celowanego genotypowania, przewidywanie *in silico* potencjalnego efektu wariantów na funkcjonalność białka, analiza nierównowagi sprzężeń, bloków haplotypowych, porównaniu z danymi dla innych populacji), a także przygotowaniu dla nich grafik i suplementów; pisaniu manuskryptu i jego redagowaniu; odpowiedziach dla recenzentów.

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G., Strapagiel, D., 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. *Genes (Basel)*, 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 60% i polegał na:

Współtworzeniu koncepcji pracy i projektowaniu eksperymentu skanowania genu metodą HRM; wykonaniu prac eksperymentalnych polegających na skanowaniu *ABCG2* metodą HRM, przygotowania próbek do sekwencjonowania Sangera; analizie wyników skanowania metodą HRM, sekwencjonowania Sangera oraz opracowaniu danych (przewidywanie *in silico* potencjalnego efektu wariantów na funkcjonalność białka, analiza nierównowagi sprzężeń, bloków haplotypowych,

porównaniu z danymi dla innych populacji), a także przygotowaniu dla nich grafik i suplementów;
pisanu manuskryptu i jego redagowaniu; odpowiedziach dla recenzentów.

.....*Marcin Stankiewicz*.....

podpis

Oświadczenie współautorów o udziale w publikacji

mgr inż. Marta Sobalska-Kwapis

Pracownia Biobank, Katedra Biologii Nowotworów i Epigenetyki

Wydział Biologii i Ochrony Środowiska

ul. Pomorska 139, 90-235 Łódź

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., **Sobalska-Kwapis, M.**, Korycka-Machała, M., Bartosz, G., Dziadek, J., Strapagiel, D., 2015.

Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 / MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 10% i polegał na:

Wykonaniu prac eksperymentalnych polegających na przeprowadzeniu celowanego genotypowania metodą HRM wytypowanych wariantów w genie *ABCC1* i przygotowaniu próbek do sekwencjonowania Sangera; analizie wyników genotypowania i sekwencjonowania Sangera oraz opracowaniu danych – przygotowaniu dla nich suplementu; redagowaniu manuskryptu.

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., **Sobalska-Kwapis, M.**, Korycka-Machała, M., Dziadek, J., Bartosz, G., Strapagiel, D., 2020.

Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. Genes (Basel), 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 5% i polegał na:

Wykonaniu prac eksperymentalnych polegających na skanowaniu *ABCG2* metodą HRM i przygotowaniu próbek do sekwencjonowania Sangera; redagowaniu manuskryptu.



podpis

Oświadczenie współautorów o udziale w publikacji

dr Małgorzata Korycka-Machała

Pracownia Genetyki i Fizjologii *Mycobacterium*

Instytut Biologii Medycznej Polskiej Akademii Nauk

ul. Lodowa 106, 93-232 Łódź

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., **Korycka-Machała, M.**, Bartosz, G., Dziadek, J. Strapagiel, D., 2015.

Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 / MEiN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 10% i polegał na:

Wykonaniu prac eksperymentalnych polegających na sekwencjonowaniu Sanger; redagowaniu manuskryptu.

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., **Korycka-Machała, M.**, Dziadek, J., Bartosz, G., Strapagiel, D., 2020.

Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. Genes (Basel), 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 10% i polegał na:

Wykonaniu prac eksperymentalnych polegających na sekwencjonowaniu Sanger; redagowaniu manuskryptu.



podpis

Oświadczenie współautora o udziale w publikacjach

prof. dr hab. Grzegorz Bartosz

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., **Bartosz, G.**, Dziadek, J. Strapagiel, D., 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEiN= 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 5% i polegał na:

Współudziale w planowaniu badań; kierowaniu projektem obejmującym pozyskanie materiałów do badań i wykonującym badania metodą HRM w niniejszej pracy; redagowaniu manuskryptu.

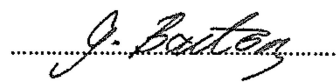
Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., **Bartosz, G.**, Strapagiel, D., 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. Genes (Basel), 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 5% i polegał na:

Współudziale w planowaniu badań; kierowaniu projektem obejmującym pozyskanie materiałów do badań i wykonującym badania metodą HRM w niniejszej pracy; redagowaniu manuskryptu.


.....
podpis

Oświadczenie współautorów o udziale w publikacji

prof. dr hab. Jarosław Dziadek

Pracownia Genetyki i Fizjologii *Mycobacterium*

Instytut Biologii Medycznej Polskiej Akademii Nauk

ul. Lodowa 106, 93-232 Łódź

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Bartosz, G., **Dziadek, J.** Strapagiel, D., 2015.

Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. BMC Genet., 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 / MEiN = 25 pkt (Komunikat MNIŚW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 5% i polegał na:

Planowaniu badań obejmujących wykonanie sekwencjonowania Sangera i redagowaniu manuskryptu.

Oświadczam, że mój szacunkowy udział w publikacji:

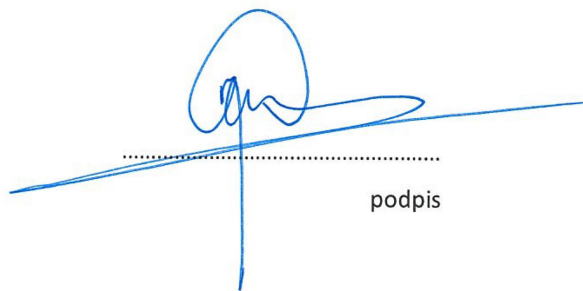
Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., **Dziadek, J.**, Bartosz, G., Strapagiel, D., 2020.

Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. Genes (Basel), 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEiN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 5% i polegał na:

Planowaniu badań obejmujących wykonanie sekwencjonowania Sangera i redagowaniu manuskryptu.



podpis

Oświadczenie współautorów o udziale w publikacji

Dr Dominik Strapagiel, prof. UŁ

Pracownia Biobank, Katedra Biologii Nowotworów i Epigenetyki

Wydział Biologii i Ochrony Środowiska

ul. Pomorska 139, 90-235 Łódź

Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Bartosz, G., Dziadek, J. **Strapagiel, D.**, 2015. Genetic variation of the ABC transporter gene *ABCC1* (Multidrug resistance protein 1-MRP1) in the Polish population. *BMC Genet.*, 16: 114. doi:10.1186/s12863-015-0271-3

IF= 2.152 /MEIN = 25 pkt (Komunikat MNiSW z 18.12.2015 r. wykaz czasopism z IF- część A)

wynosi 10% i polegał na:

Tworzeniu koncepcji pracy i projektowaniu eksperymentu z wykorzystaniem obu metod HRM; wsparcie w analizie wyników skanowania i genotypowania metodą HRM, sekwencjonowania Sangera oraz opracowaniu danych (porównanie dwóch podejść wykorzystujących HRM czyli skanowania genu i celowanego genotypowania, porównaniu z danymi dla innych populacji); redagowaniu manuskryptu; odpowiedziach dla recenzentów; autor korespondujący.

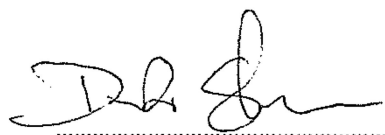
Oświadczam, że mój szacunkowy udział w publikacji:

Słomka, M., Sobalska-Kwapis, M., Korycka-Machała, M., Dziadek, J., Bartosz, G., **Strapagiel, D.**, 2020. Comprehensive Analysis of *ABCG2* Genetic Variation in the Polish Population and Its Inter-Population Comparison. *Genes (Basel)*, 11(10): 1144. doi:10.3390/genes11101144

IF= 4.096 / MEIN= 100 pkt (Załącznik do komunikatu z dnia 21 grudnia 2021: wykaz czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych)

wynosi 15% i polegał na:

Tworzeniu koncepcji pracy i projektowaniu eksperymentu skanowania genu metodą HRM; wykonaniu prac eksperymentalnych polegających na skanowaniu *ABCG2* metodą HRM, przygotowania próbek do sekwencjonowania Sangera; analizie wyników skanowania metodą HRM, sekwencjonowania Sangera; redagowaniu manuskryptu; odpowiedziach dla recenzentów; autor korespondujący.



podpis