

## 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 współlistnienia 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.

Marcin Tombs

## 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.

Marcin Stankiewicz