

STRESZCZENIE

Nadwaga i otyłość stanowią złożone zaburzenia o poligenicznej architekturze. Genetyczne podłoże nadwagi i otyłości obejmuje setki a nawet miliony wariantów genetycznych (SNP), z których każdy ma niewielki wpływ na ryzyko tej choroby. Ustalenie genetycznych uwarunkowań otyłości wielogenowej stanowi w dalszym ciągu cel dużych badań międzyośrodkowych.

W bieżącej pracy przeprowadzono badania asocjacyjne całego genomu (GWAS), towarzyszącą im analizę PheGWAS oraz badania asocjacyjne całego transkryptomu (TWAS), w celu zidentyfikowania wariantów i potwierdzenia powiązań genotyp – fenotyp związanych z nadmierną masą ciała w populacji polskiej. W tym celu, przeprowadzono trzy niezależne GWAS obejmujące > 500 000 i >1 700 000 polimorfizmów pojedynczego nukleotydu (SNP) zlokalizowanych na przestrzeni całego genomu na platformie Illumina (odpowiednio dla użytych zestawów mikromacierzy: Infinium Core Exome i Infinium Multi-Ethnic). Badania te wykonano na trzech kohortach pochodzenia polskiego: jednej dziecięcej, DZ (N = 431) i dwóch dorosłych (N=350), wydzielonych ze względu na zastosowaną mikromacierz (CE, ME). GWAS przeprowadzono w każdej z badanych grup dla trzech różnych cech związanych z nadmierną masą ciała: BMI [kg/m²] wyznaczonego na drodze pomiarów antropometrycznych oraz masy tłuszczowej FM [kg] i kąta fazowego PA [°] wyznaczonych za pomocą impedancji bioelektrycznej (BIA). W każdym GWAS uwzględniono również interakcję SNP z płcią. Pierwszy GWAS bada powiązania z BMI w każdej z wydzielonych kohort. Drugi GWAS bada powiązania z FM, a analiza uwzględnia masę ciała jako współzmienną w modelu statystycznym. W trzecim GWAS analizowano asocjacje pomiędzy badanymi SNP a wartościami PA oznaczonymi w BIA. W obrębie badanych kohort, wykonano wizualizację pheGWAS, w celu eksploracji badanego fenotypu otyłości, na który globalnie składają się wszystkie testowane w bieżącej pracy zmienne (BMI, FM i PA). Na koniec przeprowadzono analizę TWAS dla grupy dzieci, mającą na celu ocenę wpływu komponenty genetycznej ekspresji (w podskórnej – SAT i wiscelarnej – VAT tkance tłuszczowej) na wartość BMI, FM i PA przy kontrolowanym wpływie płci i wieku badanych osób.

Przeprowadzona analiza wykazała istnienie *loci*, na przestrzeni całego genomu mających wpływ na badane zmienne. Wykazany za pomocą GWAS efekt wpływu łącznie 500 SNP (we wszystkich badanych grupach) na BMI, FM i PA był w większości przypadków sugestywny (ang. suggestive association, SA, $p < 1 \times 10^{-5}$). Spośród 500 wspomnianych wyżej SNP, 80 wykazywało poziom istotności dla badań całego genomu (ang. genome-wide significance, GWS, $p \leq 5 \times 10^{-8}$). *Loci*, dla których obserwowano najniższe p-wartości w badanej grupie osób pochodzenia polskiego, były w większości, we wcześniejszych badaniach, związane z rozmiarem ciała. Uwzględnienie w analizie interakcji

badanych SNP z płcią uwidocznili dymorfizm płciowy wpływu SNP na wartości BMI, FM i PA. Efekt zależny od płci obserwowano dla 183/500 zidentyfikowanych wariantów (SA, $p < 1 \times 10^{-5}$), a allele ryzyka w większości tych przypadków wywierały silniejszy efekt na BMI, FM i PA w grupie kobiet niż mężczyzn. Zidentyfikowano jedno nowe *loci* na chr 15 w obrębie genu *PCDHGA1* (rs2240700, Joint.pval = $1,90 \times 10^{-7}$) związane z masą tłuszczową w grupie dzieci, które nie było wcześniej powiązane z nadwagą ani otyłością.

Wykazano, że na wartość kąta fazowego (PA) mają wpływ częste warianty genetyczne. Zidentyfikowano 6 *loci*, które osiągnęły poziom GWS ($p \leq 5 \times 10^{-8}$): rs112877691 w genie *TMEM132B* (chr 12, Joint.pval 2.99×10^{-11}), rs13079964 w obrębie genu *ALCAM* (chr 3, Joint.pval = 1.61×10^{-10}), rs7640978 w genie *CMTM6* (chr 3, Joint.pval = 3.34×10^{-9}), rs151111787 w *ME1* (chr 6, Joint.pval 3.61×10^{-9}) oraz *locus* na chr 2 w obrębie genu *CCDC85A* i chr 8 w niekodującym obszarze międzygenowym pomiędzy *RFPL4AP7* i *RN7SKP294*. Pięć z powyższych *loci* wykazywało efekt zależny od płci, silniejszy u płci żeńskiej w porównaniu z płcią męską (poza *locus* na chr 8, w przypadku którego, efekt płci nie był istotny).

Ponadto w toku prowadzonych analiz, przy użyciu narzędzia PheGWAS, zidentyfikowano regiony na chromosomach, w obrębie których statystycznie znamienne SNP, wywołują wspólny efekt na BMI, FM i PA. Dzięki zastosowanej w niniejszej pracy analizie TWAS, przeprowadzonej z wykorzystaniem sumarycznych statystyk opisanych w GWAS, na kohorcie dziecięcej, w dwóch rodzajach tkanki tłuszczowej SAT i VAT, zidentyfikowano łącznie 55 unikalnych genów ($p.val < 0.01$), których ekspresja zależna była od wariantów genetycznych związanych z analizowanymi parametrami masy ciała.

Podsumowując, przeprowadzone analizy asocjacyjne umożliwiły zidentyfikowanie polimorfizmów genetycznych i regionów genomu przyczyniających się do rozwoju nadmiernej masy ciała u Polaków, wykazując ich związek z BMI i FM. Wykazano ponadto, że na PA mają wpływ częste warianty genetyczne, a sam PA może być ważnym parametrem do badania stanu zdrowia związanego ze składem ciała. Otrzymane wyniki stanowią podstawę do dalszych badań mających na celu ustalenie mechanizmów wpływu czynników genetycznych na wartości zmiennych wykorzystywanych do charakterystyki fenotypu otyłości.

ABSTRACT

Overweight and obesity are common, complex disorders with a polygenic architecture. The genetic basis of overweight and obesity includes hundreds or even millions of genetic variants (SNPs), each of which has little effect on the risk of the disease. The genetic architecture of polygenic obesity is still the basis of large multi-centre studies conducted on a global scale.

In the current work, genome-wide association studies (GWAS) and accompanying PheGWAS visualisation and transcriptome-wide association studies (TWAS) were performed to identify variants and confirm genotype-phenotype associations with excessive body weight in the Polish population. For this purpose, first, three independent GWAS were performed covering > 500,000 and > 1,700,000 (microarray kits used: Infinium Core Exome and Infinium Multi-Ethnic, respectively) single nucleotide polymorphisms (SNP) on the Illumina platform for three cohorts of Polish origin: two groups of adults (ME, N = 223 and CE, N = 127) and one to which schoolchildren were included (DZ, N=431). GWAS was carried out in each of the study groups (ME, CE and DZ) for three different variables related to excessive body weight: body mass index BMI [kg/m²] determined by anthropometric measurements, fat mass FM [kg] and phase angle PA [°] determined by bioelectrical impedance. The interaction of SNPs with sex was also considered in each GWAS. The first GWAS examines associations with BMI in each of the identified cohorts, focusing solely on body mass index values. The second GWAS examines associations with FM, and the analysis considers body weight as a covariate in the statistical model. In the third GWAS, associations between the tested SNPs and PA values determined in the BIA, were attempted. Within the studied cohorts, pheGWAS visualization was performed in order to explore the studied obesity phenotype, which globally consists of all variables tested in the current study (BMI, FM and PA). Finally, a TWAS was performed for a group of children. The aim of TWAS was to assess the impact of genetic expression components (in subcutaneous - SAT and visceral – VAT adipose tissue) on the BMI, FM and PA values, taking into account the sex and age of the subjects.

The analysis presented in this dissertation showed the existence of many *loci* throughout the genome, affecting the studied variables. A total of 500 SNPs (in all study groups) reached suggestive association threshold (SA, $p < 5 \times 10^{-5}$). From them, as many as 80, reached the genome-wide significance level (GWS, $p \leq 5 \times 10^{-8}$). Nevertheless, the *loci* for which the lowest p-values of statistical significance were observed in the Polish population were associated with body size in previous studies. To identify variants with sex – specific effects, we examined SNP – sex interaction. The analysis revealed the sexual dimorphism of the analyzed BMI, FM and PA values associated with excessive body weight. A gender dependent effect was observed for 183/500 variants identified (SA, $p < 1 \times 10^{-5}$). Risk

alleles in most of these cases had a stronger effect on BMI, FM and PA in females than in males. One new *locus* was identified on chr 15, in *PCDHGA1* gene (rs2240700, Joint.pval = $1,90 \times 10^{-7}$) which was associated with FM in the group of children and was not previously reported as related to overweight or obesity, neither in adults nor the childhood.

A total of 6 *loci* have been shown to influence the value of the phase angle (PA) at the GWS level ($p \leq 5 \times 10^{-8}$): rs112877691 in *TMEM132B* (chr 12, Joint.pval = 2.99×10^{-11}), rs13079964 in *ALCAM* (chr 3, Joint.pval = 1.61×10^{-10}), rs7640978 in *CMTM6* (chr 3, Joint.pval = 3.34×10^{-9}), rs151111787 in *ME1* (chr 6, Joint.pval 3.61×10^{-9}) and two other *loci*: on chr 2 which overlapped *CCDC85A* gene and on chr 8 mapped to non-coding, intergenic regions, in which the nearest were two pseudogenes: *RFPL4AP7* and *RN7SKP294*. Five of them showed a sex-specific effect at SA ($p < 1 \times 10^{-5}$), which was stronger in females compared to males. However, there was no evidence of significant sex interaction for the chr 8 *locus*.

The graphical visualisation using the PheGWAS tool, revealed the regions on the chromosomes within which the identified SNPs cause a common effect on BMI, FM and PA in tested groups. A total of 55 unique genes ($p.val < 0.01$) were identified in the TWAS, the expression of which is associated with the analyzed body weight parameters in two types of adipose tissue, SAT and VAT .

In conclusion, the association analyses made it possible to identify genetic polymorphisms and genome regions contributing to excessive body weight in Poles. The body components, i.e. BMI and FM, which are commonly used to classify overweight and obesity were affected by frequencies of risk alleles of presented SNPs. PA has been shown to be influenced by genetic variants, and PA itself can be an important parameter to investigate nutritional status related to body composition. The obtained results underline the need for a deeper understanding the genetic factors influence on the values describing the variables used to characterize the obesity phenotype.