

Streszczenie

Tytuł: Analiza związku mutacji nonsensownych z wybranymi fenotypami w populacji polskiej.

Autor: Konrad Podsiadło

Mutacje odgrywają kluczową rolę w adaptacji do nowych warunków środowiskowych, w tym w zmianach funkcjonalności białek, które są niezbędne do adaptacji w szlakach biochemicznych. Jednak mutacje mogą także prowadzić do chorób, takich jak mukowiscydoza, dystrofia mięśniowa Duchenne'a, β -talasemia i nowotwory. Występowanie przedwczesnych kodonów stop jest powszechną przyczyną takich chorób, a mogą one wynikać z mutacji germlinej lub somatycznej DNA, niedokładnego splicingu pre-mRNA lub braku optymalizacji edycji RNA. Naturalna selekcja może prowadzić do nowych alleli z wysoką częstością pochodnych alleli i zróżnicowanym poziomem zróżnicowania populacji. W niniejszym badaniu przeanalizowano 141 SNPs prowadzących do przedwczesnych kodonów stop, z których większość ma wystarczająco wysokie występowanie ($> 5\%$) by być uważane za podlegające niemal neutralnej selekcji. PTCs zostały wybrane z badania Fujikury i wsp., a niektóre z genów zawierających mutację zostały ontologicznie sklasyfikowane jako związane z metabolizmem, metabolizmem leków, układem odpornościowym, palcami cynkowymi i keratyną. Na podstawie tej ontologii w niniejszym badaniu analizowane były następujące fenotypy: otyłość, nadwaga, płodność i długość życia.

Konkretne cele niniejszego badania to analiza możliwych związków między kodonami przedterminacyjnymi a wiekiem (długość życia), analiza możliwych związków między kodonami przedterminacyjnymi a liczbą dzieci (płodność) oraz analiza możliwych związków między kodonami przedterminacyjnymi a wskaźnikiem masy ciała (otyłość i nadwaga).

Analiza skojarzeń została przeprowadzona na bazie danych uzyskanej na podstawie umowy z Uniwersytetem Łódzkim, która zawierała 5 600 próbek zdrowych ludzi w Polsce, z 500 000 SNP od każdego badanego, w tym 141 SNP PTC. Wszystkie analizy skojarzeń statystycznych w tej pracy doktorskiej zostały przeprowadzone za pomocą platformy statystycznej R. Standaryzowane wielkości efektów używane w tej pracy to: d Cohena, r Spearmana i współczynnik szans. Do analizy skojarzeń z zmiennymi ilościowymi wykorzystano modele regresji liniowej i logistycznej, zaimplementowane w pakiecie R SNPassoc.

Wszystkie testy statystyczne były dwustronne i ustalono dwa progi istotności statystycznej: nominalny $p < 0,05$ oraz skorygowany Bonferroniem $p < (0,05 / 141 \text{ SNP analizowanych}) = 3,55 \times 10^{-4}$. Dla zmiennej jakościowej kategorii BMI względem SNP dla modeli dominujących i recesywnych użyto testu exact dokładnego Fishera oraz testu trendy Cochran-Armitage'a.

Celem badania było zbadanie możliwych związków między PTC a trzema różnymi fenotypami: długowiecznością, płodnością i ryzykiem otyłości. W badaniu przeanalizowano 141 SNP PTC (z 5095 pacjentów) i stwierdzono, że 21 z nich wykazywało statystycznie istotne związki z jednym z trzech fenotypów. Jednym obiecującym kandydatem do dalszej analizy jest gen ERV3-1, ponieważ SNP na tym genie (rs67047829) wykazał potencjalne znaczenie kliniczne w związku z BMI. Mimo ciekawych wyników, badanie ma kilka ograniczeń, a te wyniki należy traktować z rozwagą.

Abstract

Title: Analysis of the relationship between nonsense mutations and selected phenotypes in the Polish population

Author: Konrad Podsiadło

Mutations play a crucial role in adaptation to new environmental conditions, including changes in protein functionality that are necessary for adaptation in biochemical pathways. However, mutations can also lead to diseases such as cystic fibrosis, Duchenne muscular dystrophy, β -thalassemia, and cancer. The occurrence of premature stop codons is a common cause of such diseases, and these can arise from germline or somatic DNA mutations, inaccurate pre-mRNA splicing, or lack of optimization of RNA editing. Natural selection can result in new alleles with high frequencies of derived alleles and varying levels of population diversity. In the present study, 141 SNPs which lead to premature stop codons were studied, most of which have high enough frequencies (>5%) to be regarded as being subject to near-neutral selection. These PTCs were selected from a study by Fujikura et al, and some of the genes harboring the mutation have been ontologically categorized as being involved in metabolism, drug metabolism, the immune system, zinc fingers, and keratin. Based on this ontology, the phenotypes analyzed in the present study were: obesity, overweight, fertility, and life expectancy.

The specific aims of the present study were the analysis of possible associations between pretermination codons and age (life span), analysis of possible associations between pretermination codons and the number of children (fertility) and analysis of possible associations between pretermination codons and body mass index (obesity and overweight).

Association analysis was performed on a database obtained through an agreement with the University of Lodz, which contained 5,600 samples from healthy people in Poland, with 500,000 SNPs from each subject, including 141 PTC SNPs. All statistical association analyses for this thesis were performed using the R statistical platform. The standardized effect sizes used in this thesis were: Cohen's d, Spearman's r and odds ratio. For association analysis with quantitative variables, linear and logistic regression models, implemented in the R SNPassoc package, were used. All statistical tests were two-tailed and two statistical significance p.value thresholds were set, the nominal $p < 0.05$ and after Bonferroni correction $< (0.05/141 \text{ SNPs analyzed}) = 3.55 \times 10^{-4}$. For the qualitative variable of BMI categories versus SNP for dominant and recessive models, the Fisher exact test and Cochran Armitage trend test were used.

The study aimed to investigate possible associations between PTCs and three different phenotypes: longevity, fertility, and obesity risk. The study analyzed 141 PTC SNPs (from 5095 patients) and found that 21 of them showed statistically significant associations with one of the three phenotypes. One promising candidate for further analysis is the ERV3-1 gene, as a SNP on this gene (rs67047829) showed a potential clinical significance in its association with BMI. Despite the interesting findings, the study has several limitations, and these results should be treated with caution.