

## Zastosowanie metod statystycznych w lokalizacji loci sprzężonych z cechą ilościową

Mianem lokalizacji genów wpływających na pewne ilościowe fenotypowe cechy osobnicze (skrótowo zagadnienie to jest nazywane mapowaniem QTLi — *Quantitative Trait Loci*) określa się zadania pozwalające na wskazanie tych genów, które istotnie wpływają na badaną cechę ilościową. Lokalizacja genów sprowadza się do wskazania loci (czyli pozycji na chromosomie), których stan jest istotnie sprzężony z badaną cechą ilościową. Taką cechą może być twardość skorupki jajka, masa mięśniowa, wydajność mleczna krów lub inna cecha osobnicza. Metody te są również stosowane w badaniach cech ludzkich, przykładem jest zeszłoroczny warsztat GAW14 (Genetic Analysis Workshop <http://www.gaworkshop.org/>), podczas którego poszukiwano genów odpowiedzialnych za łatwość wpadania w nałogi (głównie alkoholizm).

Genetycy podczas badań są w stanie określić stan (genotyp) dla wielu loci dla każdego badanego osobnika (w loci najczęściej występuje jeden z dwóch, trzech możliwych genotypów). Na to zagadnienie można patrzeć jak na testowanie wielu hipotez o istotności wpływu stanu danego loci na badaną cechę. Zagadnienie wygląda pozornie prosto, wymaga jednak wykorzystywania bardzo zaawansowanych metod statystycznych.

Pierwsze prace nad metodami lokalizacji QTLi zapoczątkował Soller [5]. Wykorzystywał on podstawowe metody statystyczne (analiza ANOVA), do wskazania zmiennych (odpowiadających stanom w różnych loci), które istotnie wpływają na badaną cechę. Od tego czasu proponowanych było wiele modyfikacji, usprawnień oraz zupełnie nowych metod. W ostatnich latach największym zainteresowaniem cieszą się metody wyboru liczby istotnych zmiennych. Wielu statystyków proponowało różne metody wykorzystujące popularne kryteria wyboru rozmiaru modelu (AIC lub BIC, patrz [2]), procesy stochastyczne ([4], [3]) lub technikę łańcuchów Markowa (RJMCMC) do oceny prawdopodobieństwa a posteriori dla modeli o różnych rozmiarach.

Powodem, dla którego standardowe metody zawodzą, jest olbrzymia liczba hipotez (rzędu dziesiątek tysięcy), przy znacznej, ale dużo mniejszej, liczbie obserwacji (najczęściej nie przekraczającej tysiąca). Powoduje to konieczność stosowania korekt na poziom istotności, uwzględniających tak znaczną liczbę hipotez. Dodatkowo statystyki testowe dla poszczególnych hipotez są silnie skorelowane i nieuwzględnienie tych korelacji prowadzi do znaczącego spadku mocy.

Badania, które prowadzę, dotyczą korekty poziomu istotności uwzględniającej strukturę korelacji pomiędzy statystykami testowymi oraz wyznaczania efektywnej liczby testów pozwalających na efektywniejszą kontrolę błędu pierwszego rodzaju [1]. Podczas referatu przedstawię metodę pozwalającą na kontrolę błędu FWER w przypadku silnie skorelowanych statystyk testowych. Metoda ta jest uniwersalna i można ją stosować również w innych zagadnieniach, w których występuje wielokrotne testowanie (np. analizach mikromacierzowych).

### Literatura

- [1] P. Biecek, M. Bogdan, R. W. Doerge, J. K. Ghosh, *Locating multiple interacting quantitative trait loci with the modified version of Bayesian Information Criterion* (w przygotowaniu).
- [2] M. Bogdan, R. W. Doerge, J. Ghosh, *Modifying the Schwarz Bayesian Information Criterion to Locate Multiple Interacting Quantitative Trait Loci*, *Genetics* 167 (2004), 989-999.
- [3] J. Dupuis, D. Siegmund, *Statistical Methods for Mapping Quantitative Trait Loci From a Dense Set of Markers*, *Genetics* 151 (1999), 373-386.
- [4] D. Siegmund, *Model selection in irregular problems: Applications to mapping quantitative trait loci*, *Biometrika* 91:4 (2004), 785-800.
- [5] M. Soller, T. Brody, A. Genizi, *On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines*. *Theor. Appl. Genet.* 47 (1976), 35-39.