

Algorytm normalizacji

Stosowany algorytm normalizacji może działać zarówno w oparciu o wszystkie peptydy, jak i pewien ich podzbiór, charakteryzujący się małymi rozrzutami w nieznormalizowanym zbiorze danych. Wykorzystuje on prosty model, mówiący, że logarytm zmierzonego poziom ekspresji i -tego peptydu w j -tej próbce może być przedstawiony jako:

$$x_{ij} = \mu_i + \alpha_j(\mu_i) + \epsilon_{ij} \quad ,$$

gdzie μ_i jest rzeczywistą wartością ekspresji w skali logarytmicznej, α_j jest nieliniową ciągłą funkcją modelującą zależny od poziomu ekspresji efekt j -tej próbki, a ϵ_{ij} niezależnym błędem losowym o zerowej wartości średniej. Oszacowanie efektu próbki następuje przy użyciu lokalnie ważonej regresji wielomianowej LOESS. Zaletą użycia tej nieparametrycznej metody jest możliwość estymowania zależności $\alpha_j(\mu_i)$ bez wstępnych założeń dotyczących jej postaci funkcyjnej.

Normalizacja odbywa się względem pseudoprobki referencyjnej, utworzonej jako wektor wartości średnich \bar{x} macierzy danych X :

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j \quad .$$

Estymatą efektu j -tej próbki na i -ty peptyd $\hat{\alpha}_j(\mu_i)$ staje się wartość krzywej dopasowania LOESS wyznaczonej na podstawie wykresu zależności $x_j - \bar{x}$ od \bar{x} . Znormalizowana wartość obliczana jest jako różnica:

$$\tilde{x}_{ij} = x_{ij} - \hat{\alpha}_j(x_{ij}) \quad .$$

Proces powtarzany jest w sposób iteracyjny, aż do momentu, w którym wartość średnia kwadratów różnic pomiędzy krzywymi normalizacyjnymi otrzymanymi w dwóch kolejnych iteracjach spadnie poniżej zadanego progu.

Opisana powyżej procedura stosowana jest przy normalizacji powtórzeń technicznych. W przypadku normalizacji próbek pochodzących z różnych grup badanych, normalizacja odbywa się z użyciem zbioru peptydów wybranych w oparciu o ich wariancję we wszystkich próbkach zbioru danych. Podstawą podejścia jest założenie, że efekt związany z pojedynczą próbką zawiera składową „techniczną” oraz składową związaną z prawdziwym efektem biologicznym, która w wypadku idealnego peptydu nieróżnicującego powinna być równa 0. Jeśli obie te składowe są niezależnymi efektami losowymi, to peptydy mające tylko wariancję techniczną, bez efektu biologicznego, powinny charakteryzować się mniejszą wariancją całkowitą. Ponieważ wariancja jest zależna od wartości ekspresji, to aby zapewnić równomierny rozkład peptydów w całym zakresie dynamicznym, wybór peptydów odbywa osobno dla pewnej liczby przedziałów wartości ekspresji. Dla każdego przedziału wybierany jest peptyd o najmniejszej wariancji. Normalizacja odbywa się według podanego powyżej schematu, z tą jednak różnicą, że w każdej iteracji postać

krzywej LOESS wyznaczana jest na podstawie wybranego zbioru peptydów, a następnie używana dla wszystkich peptydów.