

Análise de *biclustering* de dados de microarranjos

Isabel Luzia Nori dos Santos¹

Roberto Hiroshi Higa²

O estudo de perfis de expressão gênica relacionados a manifestações de diferentes fenótipos pode fornecer informações importantes para a compreensão da biologia envolvida nestes processos. Em particular, na agricultura, a identificação dos genes mais relevantes para manifestações de fenótipos de interesse econômico constitui uma etapa importante do processo de melhoramento genético animal e vegetal. No projeto “Rede Genômica Animal”, que integra a carteira de projetos do Macroprograma 1 do Sistema Embrapa de Gestão (SEG), a tecnologia escolhida para estudos de expressão gênica é a de microarranjos, que permite mensurar simultaneamente a expressão gênica de milhares de genes de um organismo.

Uma parte importante na análise de dados de expressão gênica utilizando a tecnologia de microarranjos consiste em agrupar genes ou experimentos em grupos com perfis similares. Em geral, para realizar essas análises são utilizadas técnicas de agrupamentos mais conhecidas, tais como k-means, SOM e agrupamento hierárquico³, que permitem identificar grupos de genes com perfil de expressão

¹ Unicamp; isabellns@cnptia.embrapa.br

² Embrapa Informática Agropecuária; roberto@cnptia.embrapa.br

³ Disponível em: <http://cran.r-project.org/web/views/Cluster.html>

similar através do conjunto de condições experimentais e grupos de condições experimentais com perfil de expressão similares.

Uma dificuldade encontrada ao se aplicar essas técnicas é que elas implementam um modelo global, embora seja conhecido que os padrões de ativação de um grupo de genes seja comum apenas sob condições experimentais específicas. Neste sentido, as técnicas de *biclustering* (MADEIRA; OLIVEIRA, 2004), que implementam um modelo local, podem ser utilizadas para encontrar simultaneamente subgrupos de genes e condições experimentais.

O objetivo deste trabalho é desenvolver *scripts* de análise de *bi-clustering* para análise de dados de expressão gênica baseados na tecnologia de microarranjos, utilizando a ferramenta estatística R¹.

No momento, encontram-se em desenvolvimento *scripts* R para *two-way clustering*, uma forma simplificada para obtenção de *biclustering* que utiliza algoritmos de agrupamentos para agrupar linhas (genes) e colunas (condições experimentais) da matriz de dados separadamente. As técnicas de agrupamentos usadas nesta tarefa são o PAM(1) e Hopach².

Na sequência, serão estudados algoritmos específicos de *biclustering* e suas respectivas implementações em R para, então, desenvolver-se os *scripts* de análise.

Os dados utilizados para validação dos *scripts*, tanto para *two-way clustering* quanto para *biclustering* serão os dados de experimentos de resistência a carrapatos, realizados no escopo do projeto “Rede Genômica Animal”.

¹ R-project. Disponível em: <http://www.r-project.org/>

² Disponível em: <http://cran.r-project.org/web/packages/hopach/index.html>

Referências

MADEIRA, S. C.; OLIVEIRA, A. L. Biclustering algorithms for biological data analysis: a survey. *IEEE Transactions on Computational Biology and Bioinformatics*, v. 1, n.1, p. 24-45, Jan./Mar. 2004.