

Utilização da metodologia WGCNA na construção de redes gênicas e identificação de mecanismos envolvidos na resposta de vacas à mastite

Vinícius Campidelli Gozzo¹
Poliana Fernanda Giachetto²

A técnica de microarranjos é uma metodologia de análise da expressão gênica, que permite investigar o nível de expressão de milhares de transcritos em uma amostra, de maneira simultânea. Por meio dessa tecnologia, pode-se examinar comparativamente a atividade transcricional global que ocorre entre diferentes amostras, sejam elas diferentes tipos celulares, tecidos ou condição experimental, por exemplo. No âmbito da Embrapa, a tecnologia de microarranjos tem sido utilizada na investigação de mecanismos que levam à determinação de fenótipos de interesse, como resistência a doenças e parasitas, melhor qualidade de carne ou maior produção de leite.

Dentre as análises que podem ser realizadas a partir dos dados gerados pelos microarranjos, a análise de expressão diferencial tem sido bastante aplicada e se mostrado importante para descoberta dos genes que influenciam o processo estudado. Apesar de tratar cada gene como uma entidade independente, e não considerando que haja interação entre eles, a identificação dos genes diferencialmente expressos tem trazido resultados importantes na busca pelos principais genes envolvidos na manifestação dos fenótipos de interesse na agricultura, pecuária e saúde. Por outro lado, as células exibem um comportamento de interação complexo, que geralmente não pode ser completamente predito a partir das propriedades de componentes individuais do sistema, uma vez que a regulação da expressão

¹ Universidade Estadual de Campinas, campidelligozzo@gmail.com

² Embrapa Informática Agropecuária, poliana@cnptia.embrapa.br

gênica e, essencialmente, todas as funções biológicas são resultantes do efeito coordenado de múltiplos genes. Assim, para se compreender como uma característica se manifesta, muitas vezes é necessária a identificação das vias biológicas envolvidas, o que inclui identificar os genes participantes e explicitar as relações entre eles. Várias metodologias podem ser utilizadas na construção de redes gênicas a partir dos dados de microarranjos. Abordagens que utilizam co-expressão na identificação de redes gênicas baseiam-se na premissa de que genes com perfis de expressão similares podem possuir funções similares (EISEN et al., 1998) e, portanto, participar de um mesmo processo biológico. Podem ser empregadas quando o objetivo é a identificação de grupos de genes envolvidos no fenótipo estudado. Além disso, com base no princípio conhecido como “princípio da culpa por associação”, que afirma que se um gene desconhecido é co-expresso com genes conhecidos, envolvidos em um mecanismo biológico particular, o gene desconhecido em questão está provavelmente envolvido no mesmo processo (SAITO et al., 2008), esse tipo de análise permite a predição da função de genes ainda não caracterizados.

Assim, tendo em vista a importância da predição das funções biológicas de genes ainda não caracterizados e da identificação de redes gênicas, esse trabalho teve por objetivo implantar uma metodologia baseada em co-expressão para análise dos dados gerados nos experimentos de microarranjos da Rede Genômica Animal. Nesse trabalho, dados de um experimento com microarranjos da Embrapa Gado de Leite, realizado para se identificar genes e mecanismos biológicos envolvidos na resposta de vacas leiteiras a infecção por microorganismos causadores da mastite foram utilizados. A mastite é uma das principais doenças que acometem os rebanhos leiteiros no Brasil e no mundo e leva a sérios prejuízos, decorrentes da queda da qualidade e da produção de leite.

Foram utilizados dois conjuntos de dados, gerados pela hibridização de chips com RNA extraído de amostras de leite de animais não infectados (n=5, grupo CONTROLE) e amostras de leite de animais colhidas 3 horas após infecção artificial com *Staphylococcus agalactiae* (n=5, grupo INFECTADO). Os dados foram pré-processados utilizando-se os pacotes *affy* e *affyQCReport* do R/Bioconductor e as redes de co-expressão construídas utilizando a metodologia WGCNA - *Weighted Gene Co-expression Network Analysis* que, a partir de medidas de correlação (Pearson), calcula a conectividade dos genes e a sobreposição topológica, possibilitando a

determinação de módulos pelo agrupamento (*clustering*) dos genes mais similares. Uma vez construídas as redes, os módulos (grupos de genes fortemente correlacionados) identificados foram comparados entre elas, a fim de se descobrir quais grupos não eram conservados, e que podem evidenciar os mecanismos biológicos distintos entre os dois grupos. Tais análises foram realizadas por meio dos pacotes do R/Bioconductor, Biobase, cluster, gcrma e WGCNA. A função molecular mais representativa dos genes de cada módulo foi identificada pela ferramenta Blast2GO (<http://www.blast2go.com/b2ghome>).

Foram construídas as 2 redes e identificados 17 módulos, sendo 3 deles não conservados entre os grupos (Figura 1 e Tabela 1). Esses módulos representam mecanismos alterados entre os 2 grupos, e que podem estar

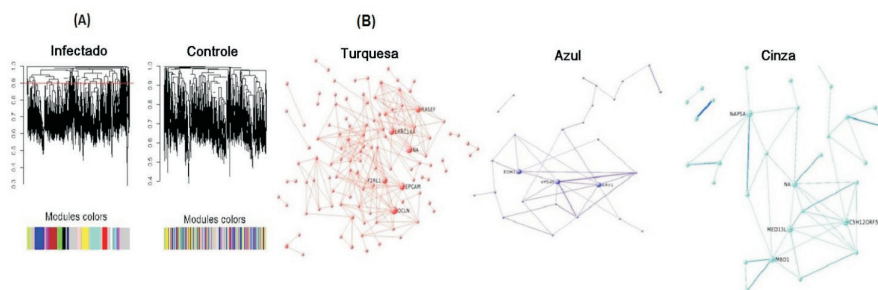


Figura 1. Dendogramas com os respectivos módulos identificados (A) e visualização dos módulos não conservados entre as redes (B).

Tabela 1. Módulos não conservados, genes componentes e funções moleculares.

Módulo	Genes	Função molecular
Cinza	BTG3, CD3E, MBD1, CHIC2, PLXNA3, MOCS3, NEIL1, VPS45, BCL2	Apoptose e reconhecimento de antígenos
Turquesa	FGL1, GJA1, F2RL1, PTPRF, S100A2, TGFB2	Mediadores de inflamação e genes marcadores de mastite
Azul	CD97, MAD2L1, ZFP106, CDKN2C, LOC514364, NOP14, PCBD1, LOC100139798, AP1S1, EDN1, IL1B, ANXA11	Divisão celular e resposta inflamatória

envolvidos na resposta dos animais à infecção inicial por *S. agalactiae*. No momento, outras análises estão sendo conduzidas, a partir desses resultados, com o objetivo de se aprofundar o conhecimento acerca dos genes identificados, que podem levar a um melhor entendimento da resposta dos animais à mastite e identificar candidatos que serão usados em programas de melhoramento animal.

Referências

EISEN, M. B.; SPELLMAN, P. T.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression pattern. **PNAS**, v. 95, p.14863-14868, 1998.

SAITO, K; HIRAI, M. Y.; SAKAKIBARA, K. Y. Decoding genes with coexpression networks and metabolomics – majority report by precogs. **Trends in Plant Science**, v. 13, n.1, p.36-43, 2008.