

Integração de Multi-ômicas e Aprendizado de Máquina: Revolucionando a Identificação de Biomarcadores na Medicina de Precisão

Mariana Tozzi Rodrigues, Karen Ruth Michio Barbosa, Gabriel da Cruz Simões, Matheus Correia Casotti, Débora Dummer Meira.

Introdução: A integração de dados provenientes de diferentes camadas moleculares — como genômica, transcriptômica, proteômica e metabolômica — permite uma compreensão mais abrangente e sistêmica dos mecanismos biológicos subjacentes a doenças complexas. Nesse contexto, abordagens integrativas aliadas a algoritmos de aprendizado de máquina (*machine learning, ML*) têm se mostrado fundamentais para a identificação de padrões moleculares, predição de interações biológicas e geração de novos *insights* em estudos biomédicos de larga escala.

Objetivo: Analisar como modelos de *ML* podem contribuir para a compreensão de sistemas biológicos complexos em estudos multi-ômicos, com ênfase na descoberta e validação de novos biomarcadores com aplicabilidade clínica.

Metodologia: Realizou-se busca sistemática de artigos científicos nas bases de dados PubMed, Scopus e Web of Science, utilizando descritores relacionados à integração de dados multi-ômicos e algoritmos de *ML* em estudos biomédicos.

Resultados: A revisão evidenciou que, embora cada abordagem de inteligência artificial (IA) e *ML* apresente limitações específicas dependendo do desenho do estudo, o uso combinado de múltiplos algoritmos potencializa a robustez dos achados. Em um estudo de análise integrativa de dados transcriptômicos *bulk*, 159 genes diferencialmente expressos (DEGs) foram submetidos a quatro algoritmos de *ML* — LASSO, SVM-RFE, XGBoost e Boruta —, resultando na seleção consensual de quatro *core* genes : COL1A1, EMP1, MYH11 e SASH1. A validação multi-ômica subsequente demonstrou que o SASH1 constitui um biomarcador prognóstico robusto e um potencial preditor de resposta terapêutica no carcinoma de células escamosas de cabeça e pescoço (HNSCC). Notavelmente, sua expressão mostrou-se especificamente reduzida em populações de células malignas, com distribuição espacial restrita às regiões estromais fibróticas com alta expressão de COL1A1, evidenciando uma interação biologicamente relevante no microambiente tumoral.

Relevância Científica: O uso integrado de IA e *ML* em estudos multi-ômicos representa uma abordagem promissora para a aceleração da descoberta de alvos terapêuticos, o aprimoramento da precisão diagnóstica e o avanço da medicina personalizada. A identificação de biomarcadores como o SASH1 exemplifica o potencial translacional dessas ferramentas na oncologia de precisão.

Conclusão: O avanço das tecnologias ômicas de alto rendimento, combinado à capacidade analítica dos modelos de *ML*, amplia substancialmente a possibilidade de decifrar sistemas biológicos complexos. A aplicação integrada dessas abordagens permite identificar padrões moleculares consistentes, validar biomarcadores em múltiplas camadas ômicas e pavimentar o caminho para novas estratégias diagnósticas e terapêuticas na medicina de precisão.