

# Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

T. R. Golub,<sup>1,2\*†</sup> D. K. Slonim,<sup>1†</sup> P. Tamayo,<sup>1</sup> C. Huard,<sup>1</sup>  
M. Gaasenbeek,<sup>1</sup> J. P. Mesirov,<sup>1</sup> H. Coller,<sup>1</sup> M. L. Loh,<sup>2</sup>  
J. R. Downing,<sup>3</sup> M. A. Caligiuri,<sup>4</sup> C. D. Bloomfield,<sup>4</sup>  
E. S. Lander<sup>1,5\*</sup>

Although cancer classification has improved over the past 30 years, there has been no general approach for identifying new cancer classes (class discovery) or for assigning tumors to known classes (class prediction). Here, a generic approach to cancer classification based on gene expression monitoring by DNA microarrays is described and applied to human acute leukemias as a test case. A class discovery procedure automatically discovered the distinction between acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) without previous knowledge of these classes. An automatically derived class predictor was able to determine the class of new leukemia cases. The results demonstrate the feasibility of cancer classification based solely on gene expression monitoring and suggest a general strategy for discovering and predicting cancer classes for other types of cancer, independent of previous biological knowledge.