

## Poster I-45

### Matrix Factorization for Molecular Pattern Recognition

Brunet, Jean-Philippe<sup>1</sup>, Tamayo, Pablo<sup>1</sup>, Golub, Todd R.<sup>1,2</sup>, Mesirov, Jill P.<sup>1</sup>

<sup>1</sup>Whitehead Institute/MIT Center for Genome Research, Cambridge, MA, USA; <sup>2</sup>Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA

The ability to generate large amounts of genomic information using DNA microarrays provides an opportunity to extract from these data previously unrecognized biological structure and meaning. The challenge, however, is that existing unsupervised clustering methods are often non-robust, and lack the ability to discover subtle, context-dependent biological patterns. We describe here the use of Non-negative Matrix Factorization (NMF), an algorithm based on decomposition-by-parts, and we demonstrate its ability to recover meaningful biological information from cancer-related microarray data without supervision. Coupled with a novel model selection mechanism, NMF is an efficient method for identification of distinct molecular patterns and provides a powerful method for class discovery. NMF appears to have higher resolution than other methods such as hierarchical clustering or self-organizing maps, and to be less sensitive to *a priori* selection of genes. Rather than separating gene clusters based on distance computation, NMF detects context-dependent patterns of gene expression in complex biological systems. This ability, similar to semantic polysemy in text, provides a general method for robust molecular pattern discovery.

*Federal grant support: R33 CA97556-01 (Mesirov) GenePattern: Methodologies and Software Tools for Molecular Pattern Recognition and Discovery.*