

Gene Filtering by Posterior Pareto Fronts

A. Hero[†], G. Fleury[◊], S. Yoshida[‡], T. Carter[#], C. Barlow[#] and A. Swaroop[‡]

[◊]Ecole Supérieure d'Electricité, Service des Mesures, 91192 Gif-sur-Yvette, France

[†]Depts. of EECS, BioMedical Eng., and Statistics, University of Michigan, Ann Arbor MI 49109, USA

[‡]Depts. of Ophthalmology and Human Genetics, University of Michigan, Ann Arbor MI 48105, USA

[#]The Salk Institute for Biological Studies, La Jolla CA 92037, USA ¹

Microarray analysis of gene expression profiles offers one of the most promising avenues for exploring genetic factors underlying disease, regulatory pathways controlling cell function, organogenesis and development. The promise of microarrays is that the technology could allow researchers to accurately quantify expression in RNA levels of thousands of genes in a tissue sample, thereby providing valuable information about complex gene expression patterns. Recent advances in bioinformatics have brought us closer to realizing this promise. However, the massive scale and variability of microarray gene data creates new and challenging problems of clustering and data mining: the so-called *gene filtering problem*.

We recently introduced a new approach to gene filtering, called Pareto gene filtering, which is based on multicriterion optimization and cross-validation. Pareto gene filtering allows the experimenter to isolate genes that achieve a good compromise between several competing gene-ranking criteria. Such genes lie on the so called *Pareto front* and are called non-dominated genes. In this paper we present a Bayes posterior analysis approach to Pareto gene filtering which we call the method of *posterior Pareto fronts* (PPF). The main advantage of the PPF approach over the Pareto gene filtering approach is that it ranks each gene according to its posterior probability that it belongs to the Pareto front.