Pareto posterior fronts

A. O. Hero

University of Michigan - Ann Arbor

http://www.eecs.umich.edu/~hero

Outline

1. Gene filtering problem
2. Posterior Pareto analysis
3. Application: Fred Wright's data
4. Application: development and aging in retina

Pareto Gene Filtering

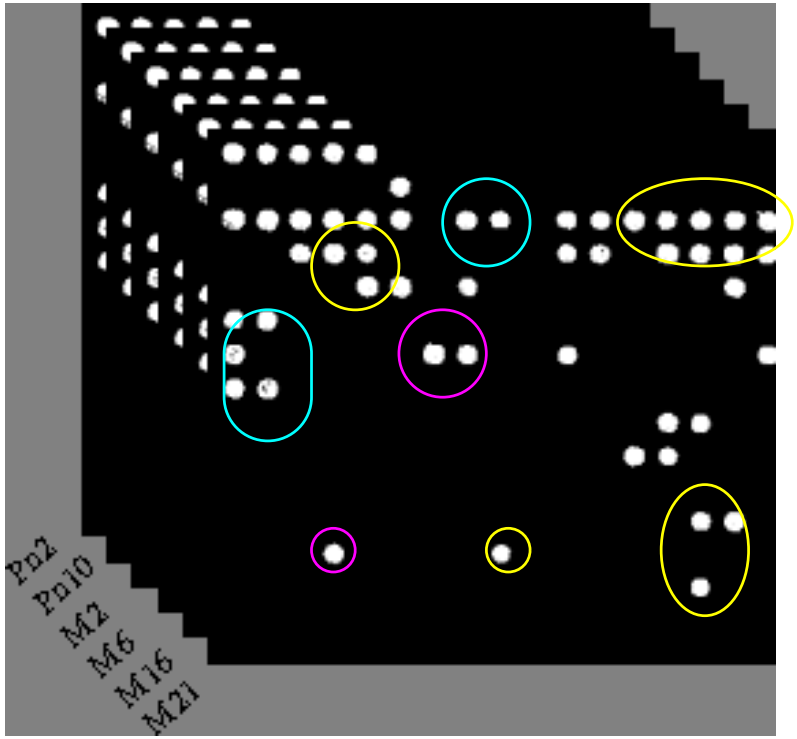

Figure 1: *Clustering on the Data Cube.*

Objective: Classify time trajectory of gene *i* into one of *K* classes

Gene Trajectory Classification

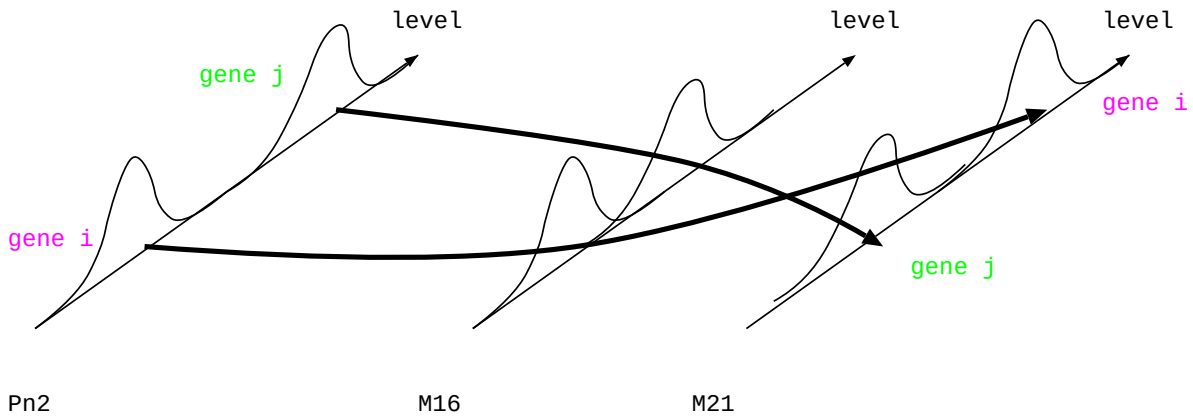

Figure 2: *Gene i is old dominant while gene j is young dominant*

Objective: extract gene trajectories (n) from sequence of repeated (m) microarray experiments over time samples (t)

$$y_{tm}(n), \quad n = 1, \ldots, N,\ t = 1, \ldots, T,\ m = 1, \ldots, M.$$

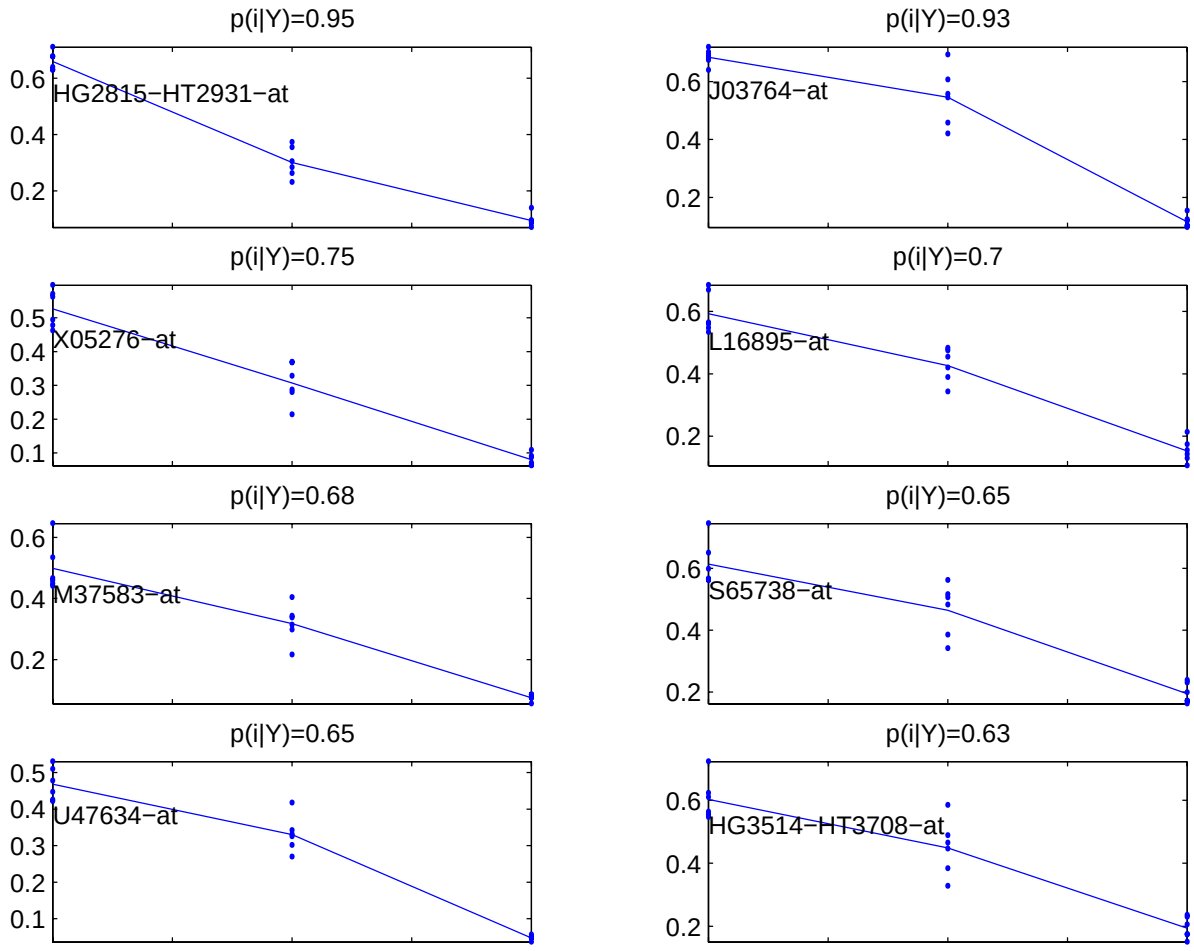

Figure 3: *8 ranked monotone decreasing gene profiles.*

Multi-objective Non-parametric Pareto Filtering

Define *trend vector*: $\psi(n) = [b_1, \ldots, b_6]$, $b_i \in \{0, 1\}$

- Old dominant filtering criteria:
  - Maximum end-to-end increase ($T = 6$)

$$\xi_1(Y(n)) = \overline{y}_{T*}(n) - \overline{y}_{1*}(n) = \max$$

  - high consistency over $6^4 = 4096$ possible combinations of trajectories

$$\xi_2(Y(n)) = \frac{\#\text{ trajectories having } \psi(n) = [1, \ldots, 1]}{4096}$$

Old Dominant Pareto Fronts

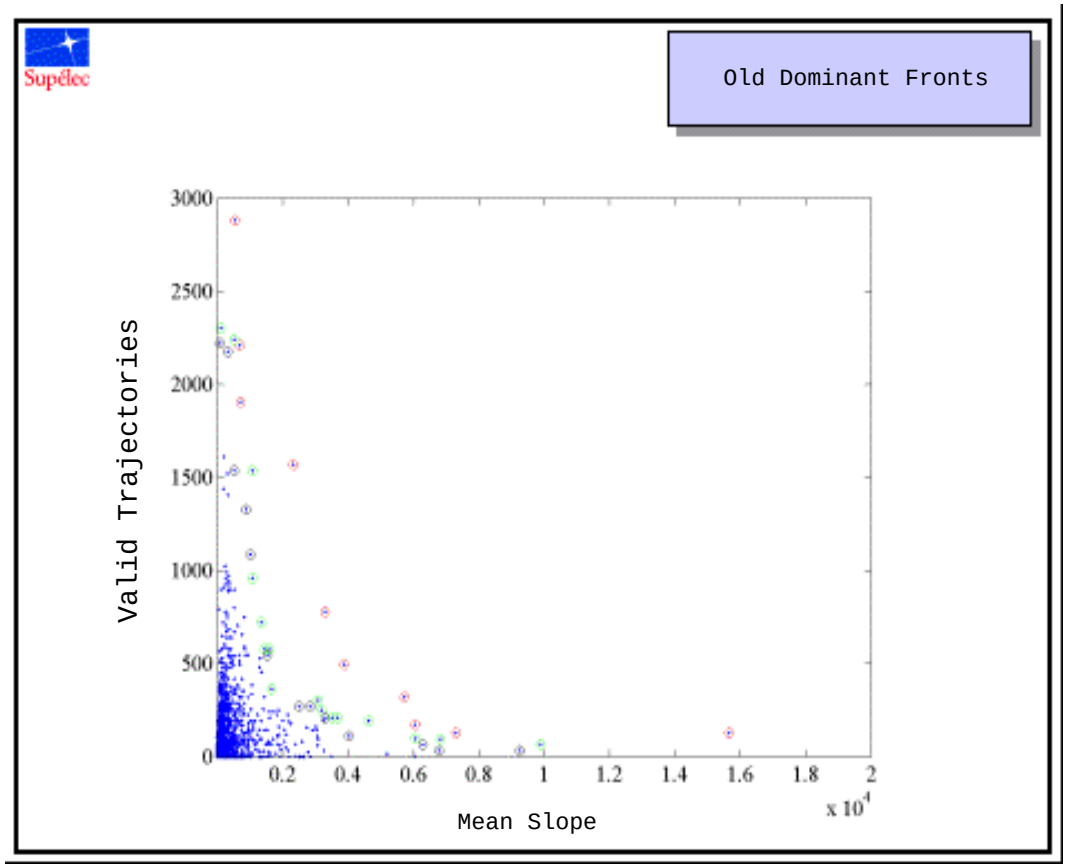

Figure 4: *Pareto fronts for old dominant genes.*

Confidence measures?

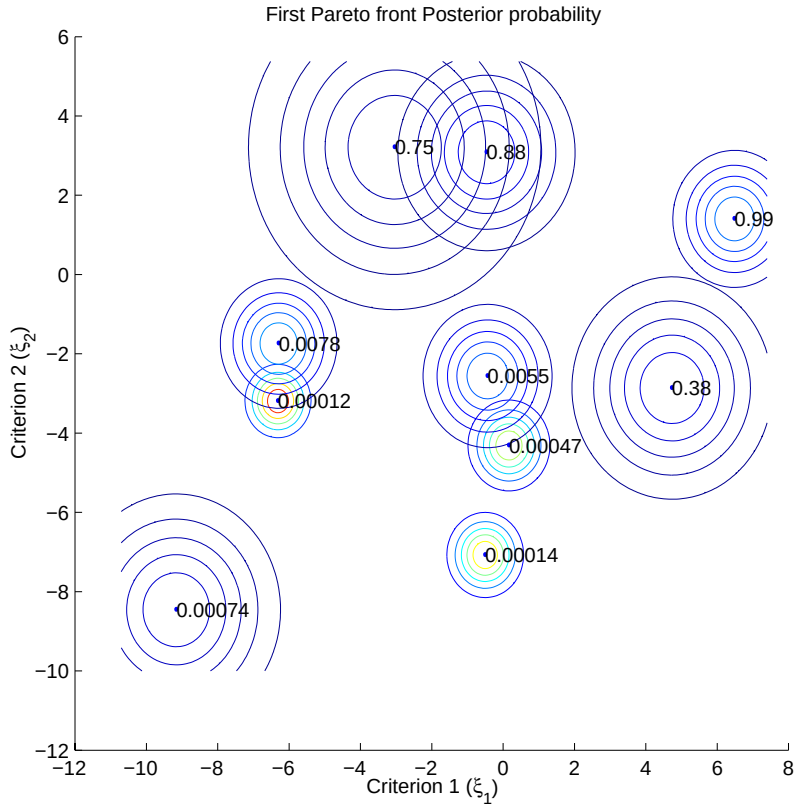

Figure 5: *PPF analysis over dual criteria to be maximized.*

Cross-validation approach

- Leave-one-out cross validation

Let $Y^{-m}(n)$ denote one possible set of $T \times (M-1) = 6 \times 3$ samples

Cross-validation Algorithm:

`Do` $m = 1, \ldots, 4^6$:

`Compute` $\left(\xi_1(Y^{-m}(n)), \xi_2(Y^{-m}(n))\right)$

`Find Genes in First 3 Pareto fronts:` G^{-m}

`End`

Resistant `Genes =` $\cap_{m=1}^{4^6} G^{-m}$

Posterior Pareto Front (PPF) approach

Given prior on mean expression levels $\overline{\xi}_p(n) = E[\xi_p(Y(n))]$ find

$$
\begin{aligned}
& p(i|Y) \stackrel{\text{def}}{=} P(\text{gene } i \text{ on Pareto front}|Y) \\
= \quad & P(\overline{\xi}_1(i) \geq \max_j \overline{\xi}_1(j) \text{ or } \ldots \text{ or } \overline{\xi}_P(i) \geq \max_j \overline{\xi}_P(j)|Y) \\
= \quad & \sum_{k=1}^{P} P(E_k(i)|Y) - \sum_{k_1<k_2} P(E_{k_1}(i), E_{k_2}(i)|Y) + \ldots \\
& + (-1)^{p+1} \sum_{k_1<\ldots<k_p} P(E_{k_1}(i), \ldots, E_{k_p}(i)|Y) \\
& + (-1)^{P+1} P(E_{k_1}(i), \ldots, E_{k_P}(i)|Y)
\end{aligned}
$$

E_i denotes the event $\xi_1(\mu(i)) \geq \max_j \xi_1(\mu(j))$

Gaussian observations with noninformative prior

1. Assume conditionally linear Gaussian model $\varepsilon_{tm}(n) \sim N(0, \sigma_t^2(n))$

$$y_{tm}(n) = \mu_t(n) + \varepsilon_{tm}(n)$$

2. Assume non-informative prior

$$f_{\mu_t(n),\sigma_t^2(n)}(u,s) = \frac{c}{s^{a/2}}, \; u \in \mathbf{R}, \; s \in \mathbf{R}^+$$

3. Adopt *Profile contrasts* as selection criteria:

$$\begin{bmatrix} \overline{\xi}_1(n) \\ \vdots \\ \overline{\xi}_P(n) \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1T} \\ \vdots & \ddots & \vdots \\ a_{P1} & \cdots & a_{PT} \end{bmatrix} \begin{bmatrix} \mu_1(n) \\ \vdots \\ \mu_T(n) \end{bmatrix}$$

Example contrasts

$$A_2 = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$A_2' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

$$A_3' = \begin{bmatrix} -1 & 1 & 0 \\ -1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix},$$

Application: Fred Wright's Affy data

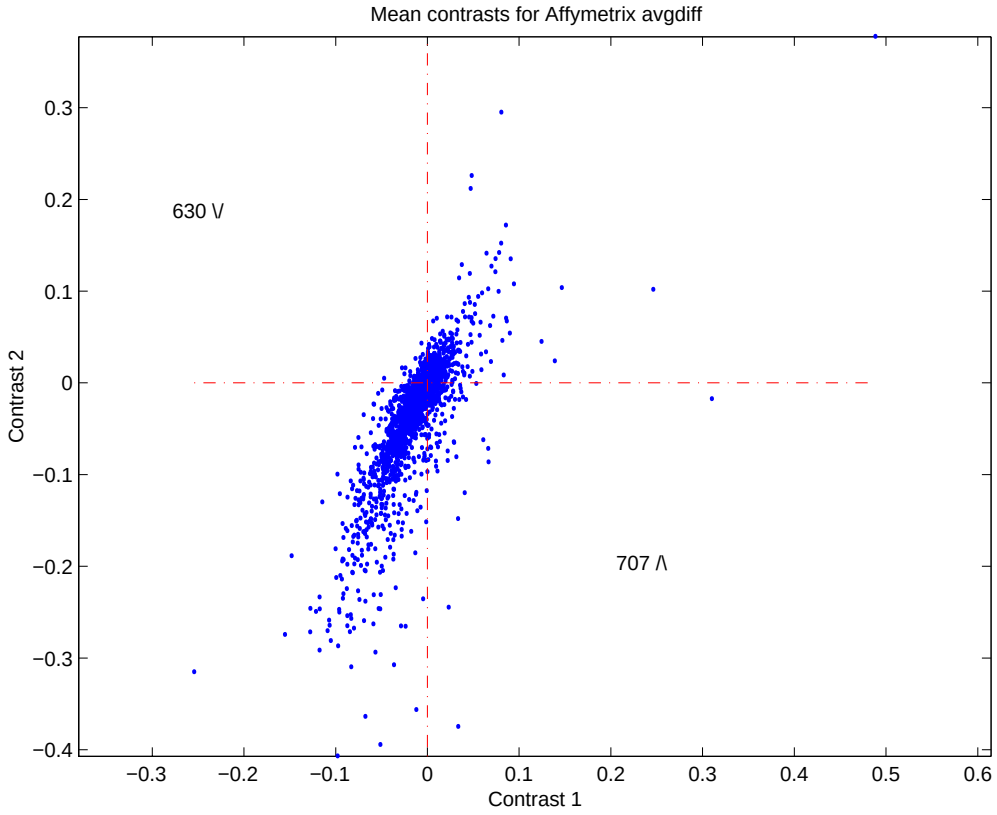

Figure 6: *Scatterplot of slope contrasts (Sample mean contrasts defined from the first two rows of A_3') for avgdiff indices for Fred Wright's HuGeneFL mixture study. Annotations are the number of non-monotone genes with convex cup (upper left) and convex cap (lower right) profiles.*

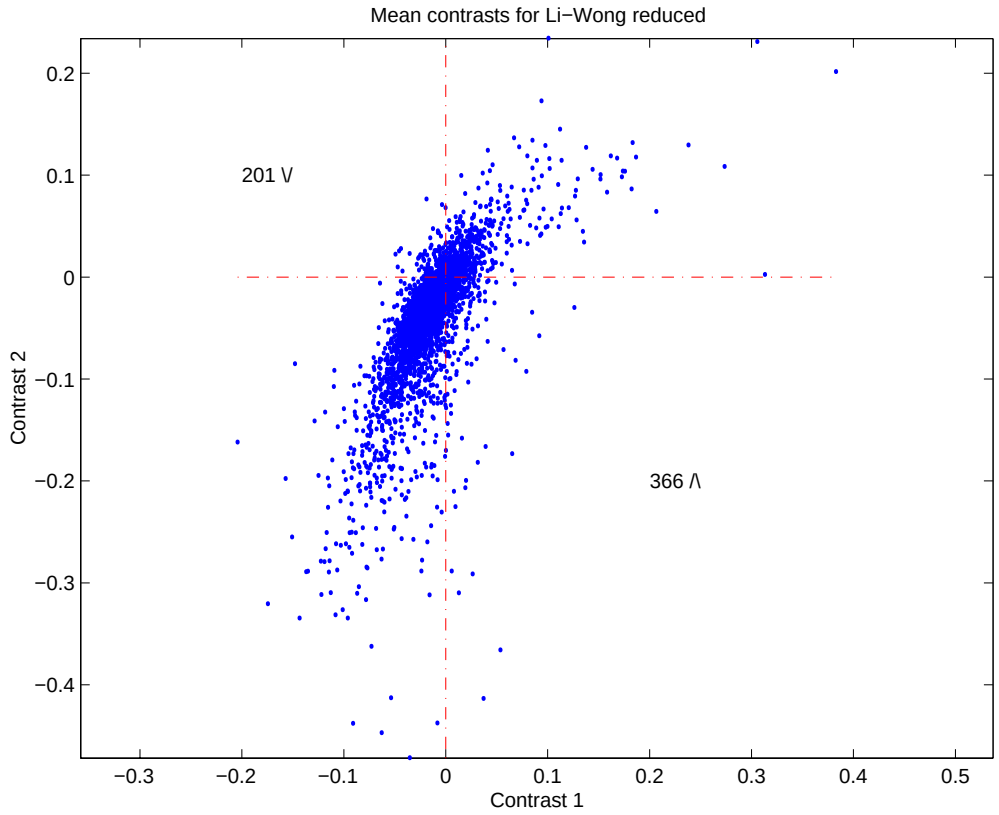

Figure 7: *Scatterplot of slope contrasts (Sample mean contrasts defined from the first two rows of A_3') for Li-Wong reduced indices for Fred Wright's HuGeneFL mixture study. Annotations are the number of non-monotone genes with convex cup (upper left) and convex cap (lower right) profiles.*

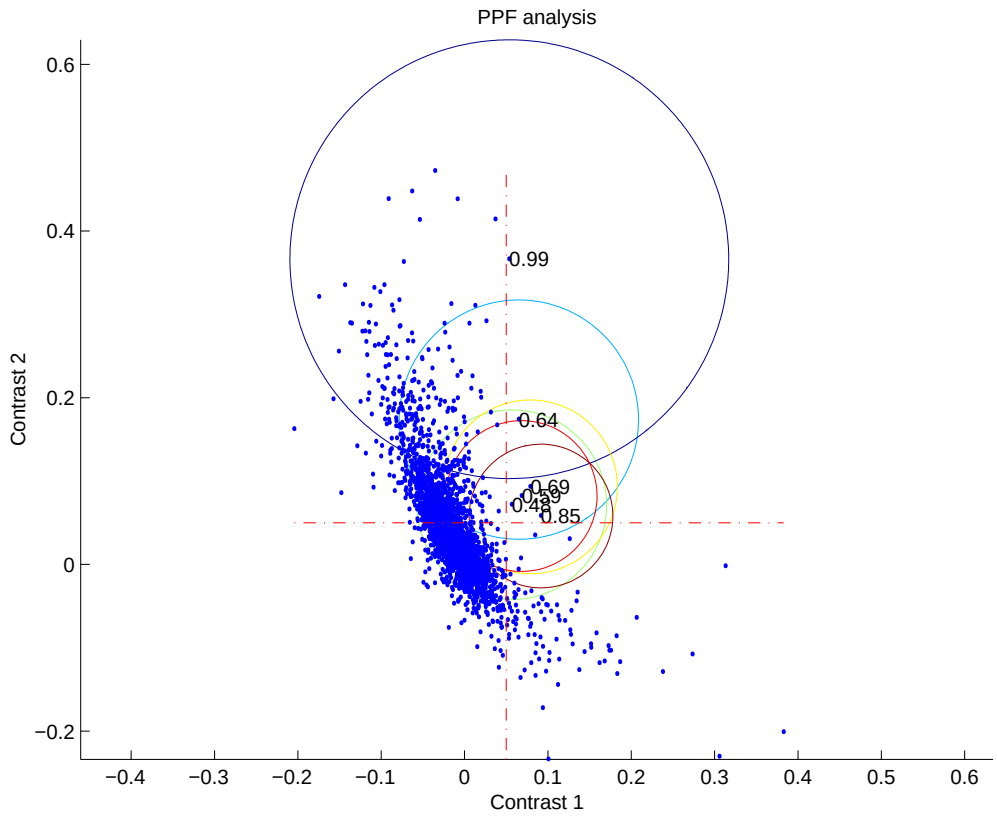

Figure 8: *The 6 top scoring genes resulting from PPF analysis of the most non-monotone convex cap profiles for Fred Wright's data using Li-Wong reduced indices (*$A = [-1, 1, 0; 1, 1, -1]$*).*

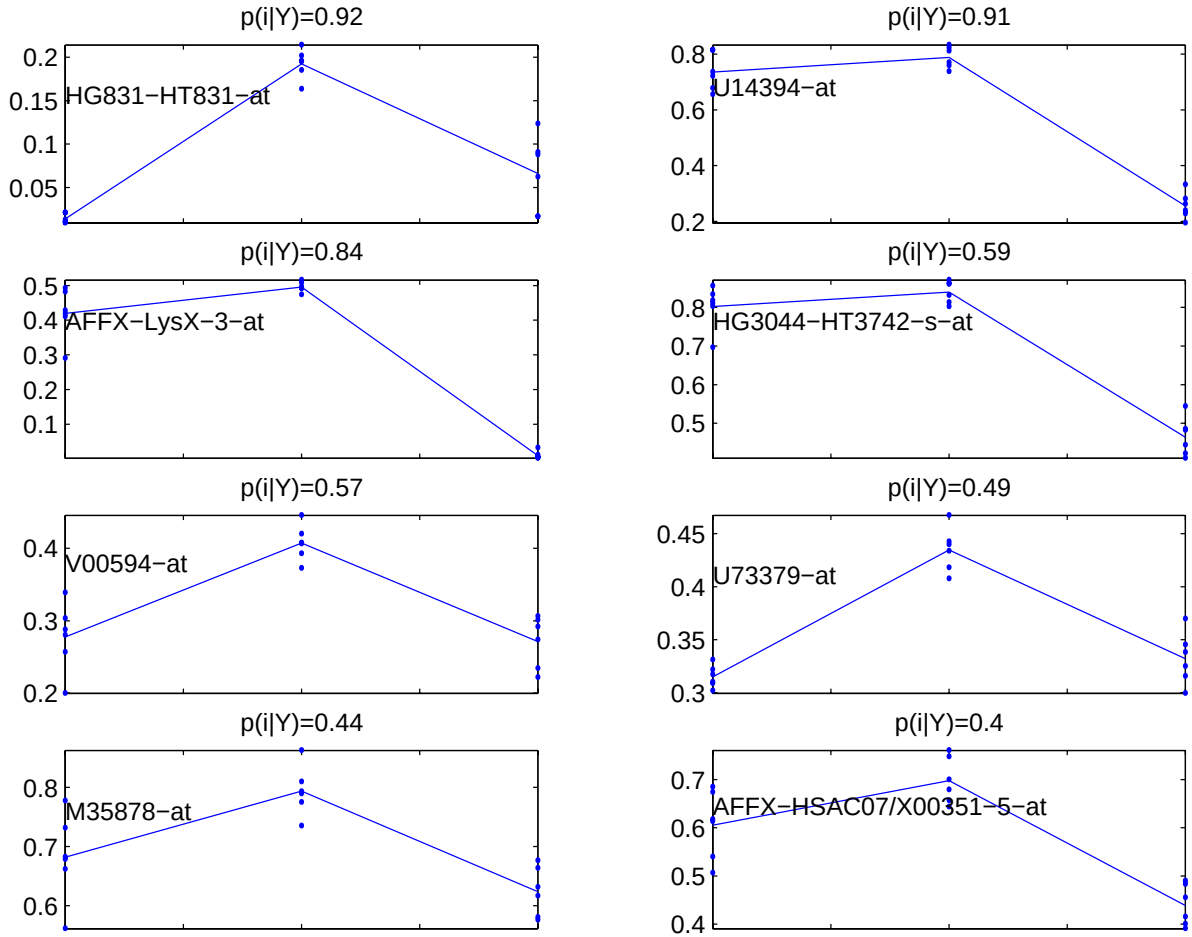

Figure 9: *First 8 rank ordered convex cup genes profiles from Li-Wong indices.*

Application: Affy human retina study

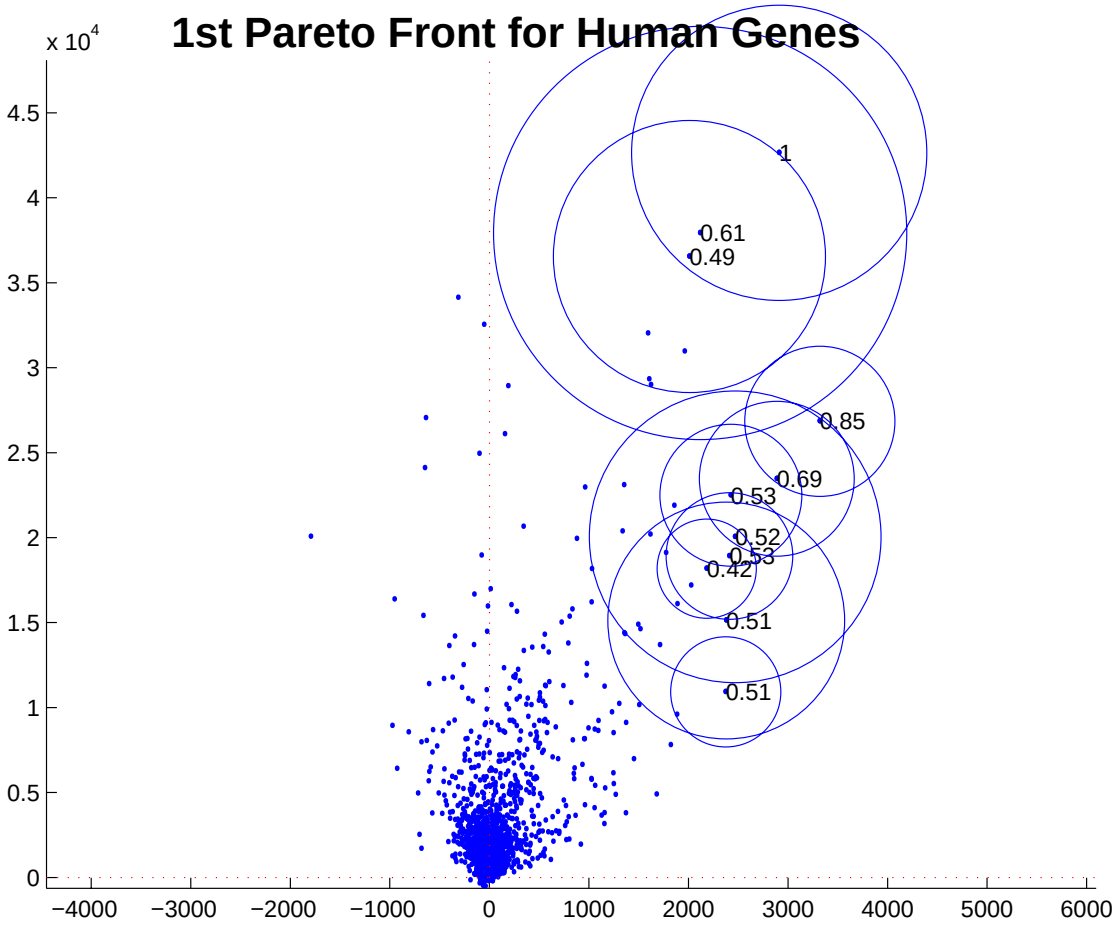

Figure 10: *First posterior Pareto front (Affy human study).*

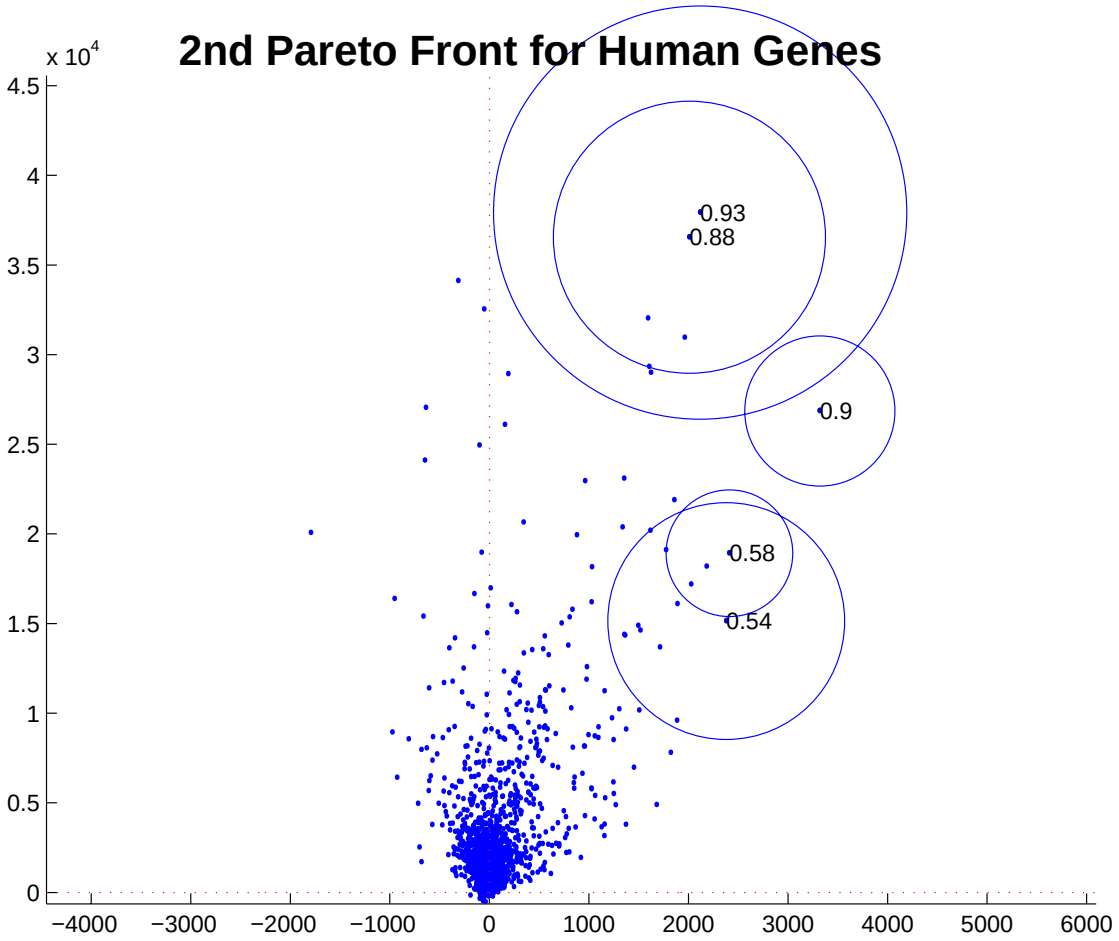

Figure 11: *Second posterior Pareto front (Affy human study).*

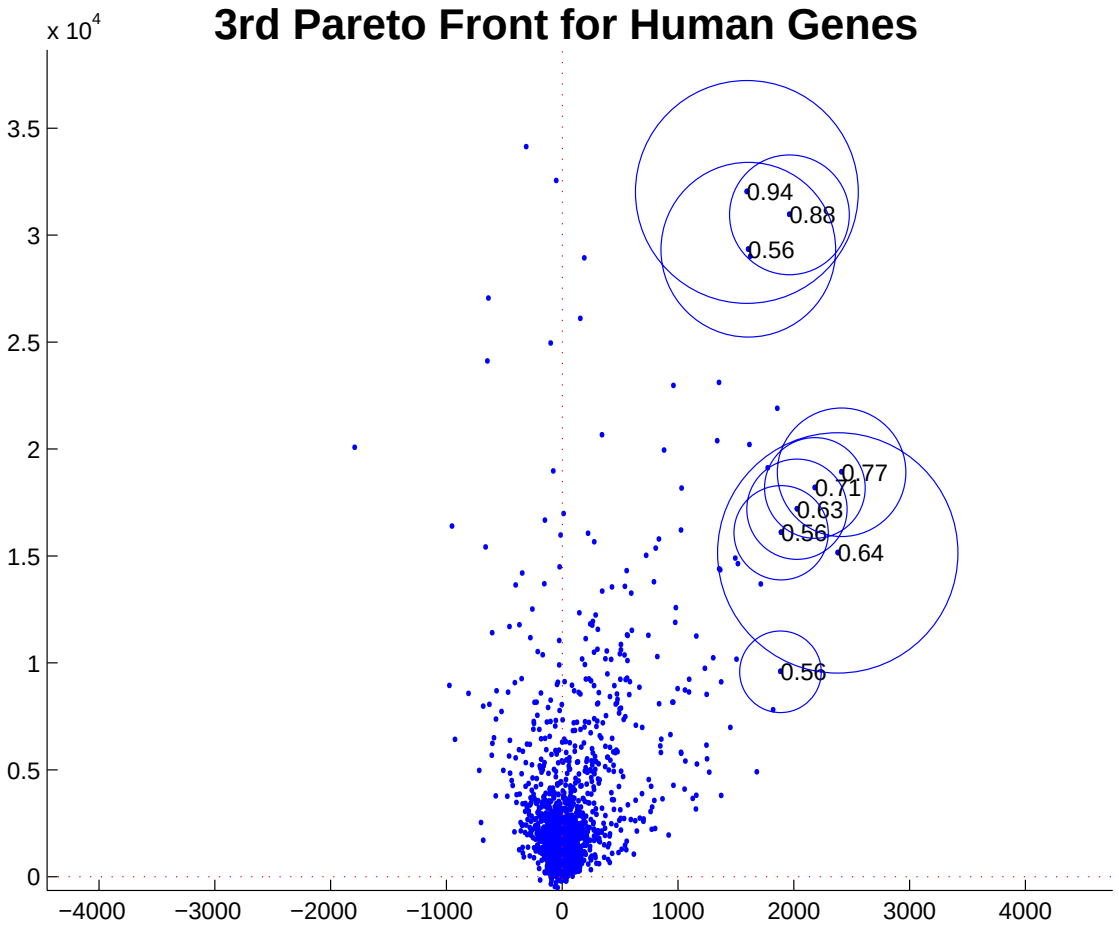

Figure 12: *Third posterior Pareto front (Affy human study).*

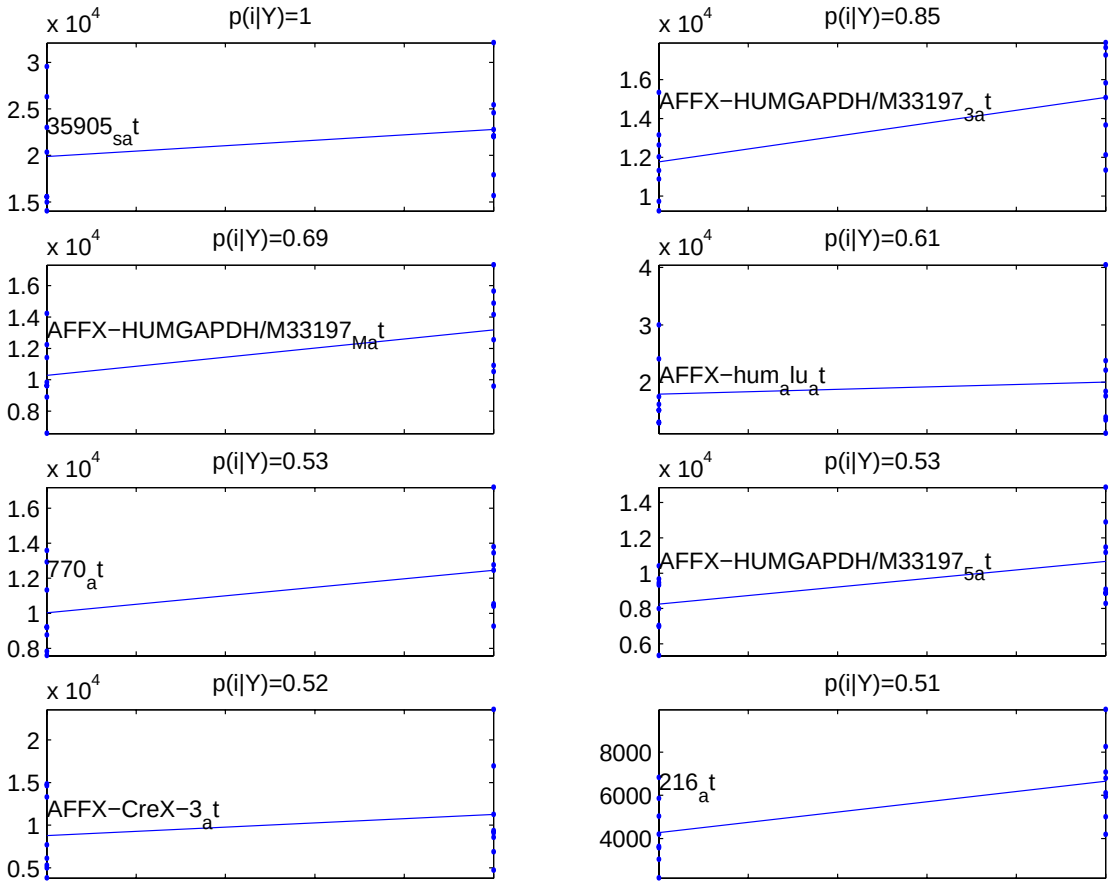

Figure 13: *Ranked first posterior Pareto front gene trajectories (Affy human study).*

Application: Affy mouse retina study

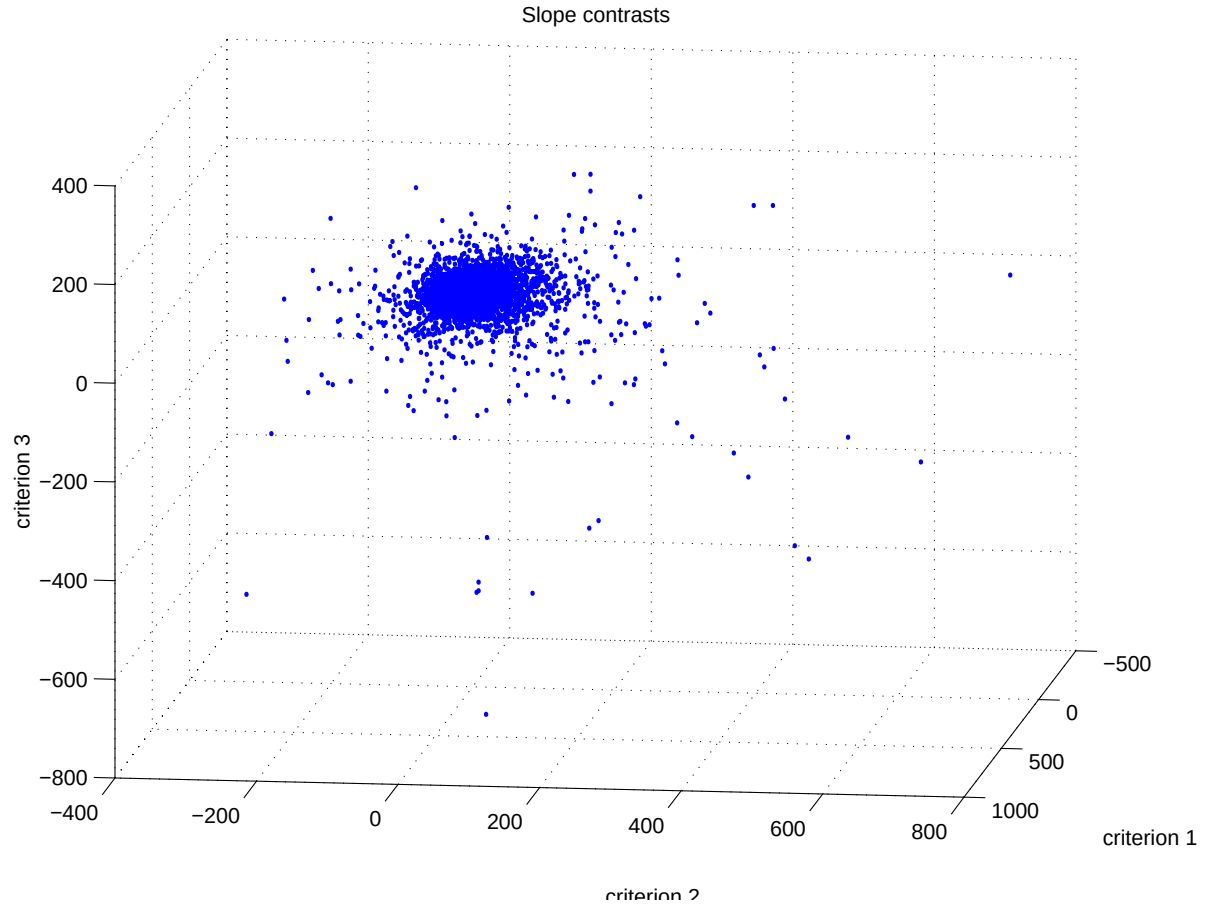

Figure 14: *Avgdiff indices for Affy mouse study.*

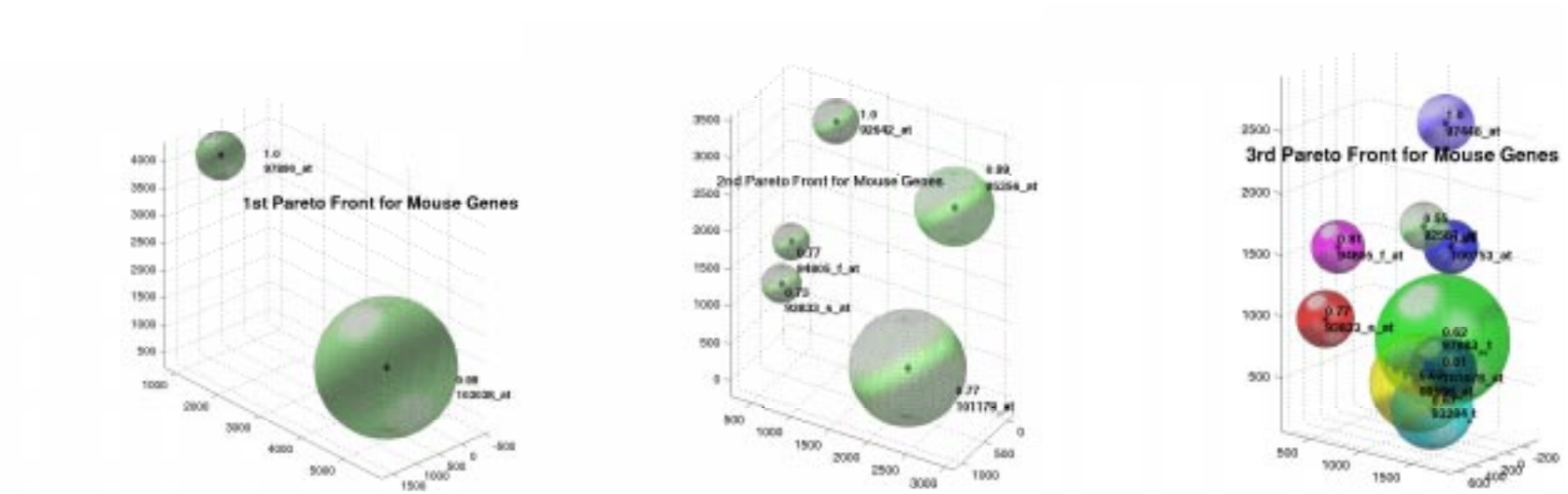

Figure 15: Pareto fronts for (affy mouse study)

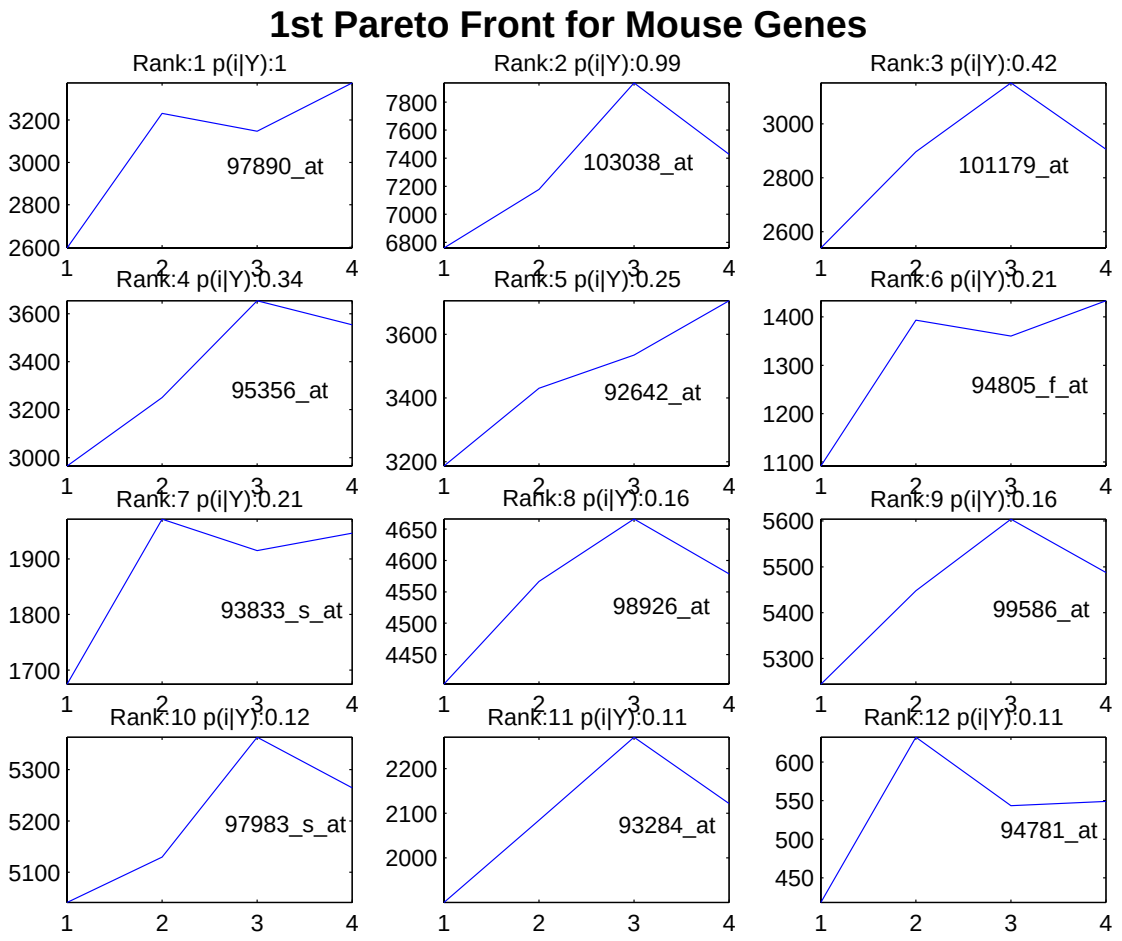

Figure 16: *Ranked first posterior Pareto front gene trajectories (Affy mouse study).*