

# ExpressCluster v1 (RC1)

**Description:** Interactive GUI-based K-Means clustering of gene expression profiles.

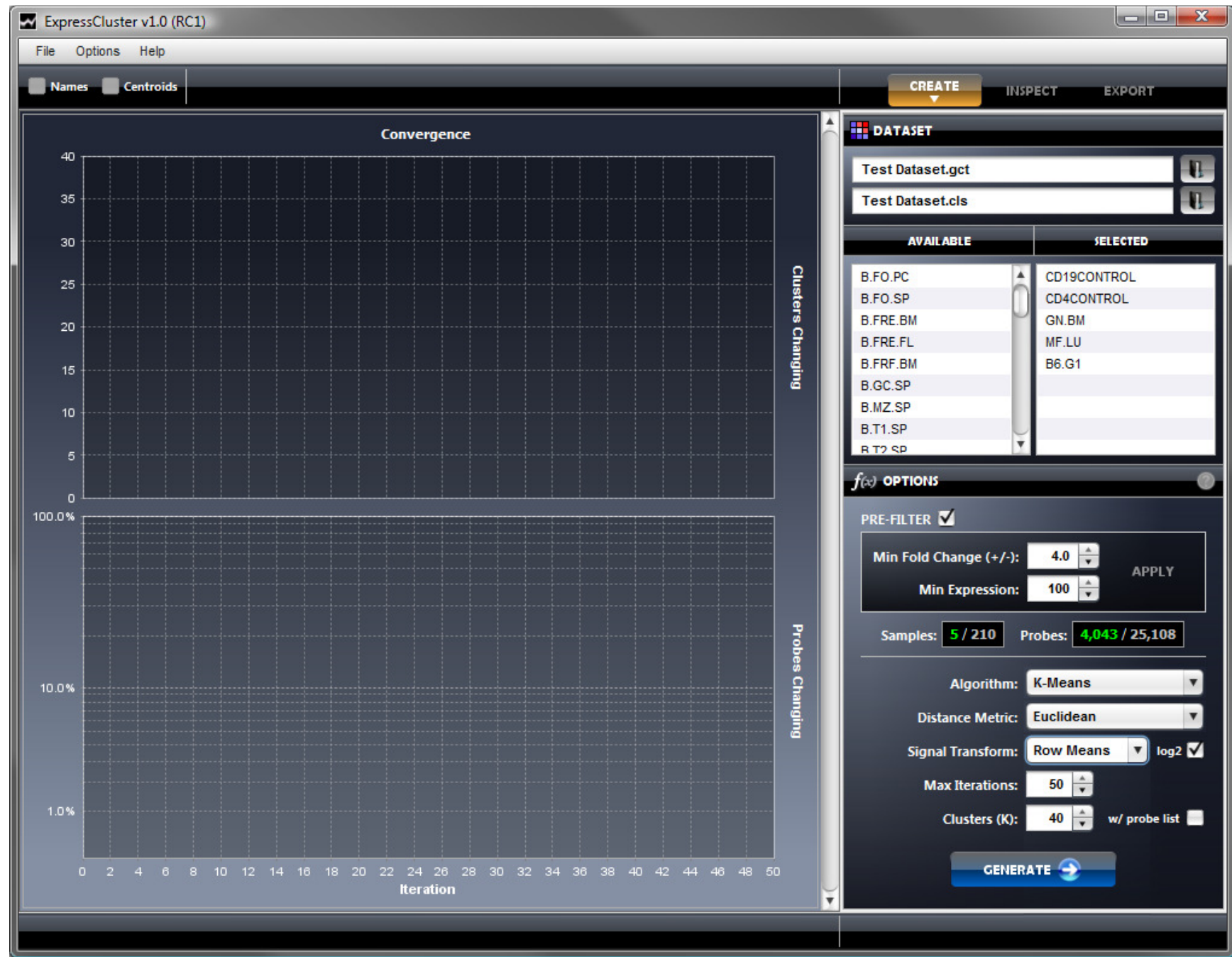
**Features Overview:**

- Input: GCT (can add RES support if needed).
- Input: CLS (optional) – if provided, it will cluster on the class-means values instead.
- User can select which samples within a GCT they would like to cluster on, and drag/drop them to rearrange their ordering.
- User can pre-filter probesets based on minimum differentiation and/or minimum expression. This is carried out prior to clustering so that the user knows how many probes they are starting with (allowing for more appropriate determination of K).
- Multiple parameter settings for distance metric, signal transformation.
- Real-time convergence plots: users can see how quickly their clusters have converged (very useful for identifying over-fitting and determining the most optimal starting parameters).
- Ability to specify desired centroids (in addition to random) via standard probe ID list (useful for clustering around genes of interest). A priori centroids are fixed and are not recomputed over successive iterations.
- Clusters can be ordered using a number of different criteria: number of probes, mean correlation to centroids, cluster name, variance of the centroid, or based on how similar they are to a particular cluster. The default is variance.
- Ability to merge or split (deterministically) clusters. Modified clusters appear in a different color.
- Interactive highlighting of specific probes within a cluster.
- Alternate "relative"/"global" scaling for expression profiles.
- Dynamic heatmaps for each cluster. Color scale reflects either the relative or global transformed signal intensity, or the original expression values as found in the GCT. Several pre-defined color schemes to choose from.
- Search – can search for a probe ID or gene symbol (if that is in the 'description' column of GCT). Any clusters containing a search term are automatically highlighted, and the specific matching probes within each clusters are automatically selected.
- Saving images – Users can copy any cluster plot or heatmap to the clipboard, or save to disk.
- Batch export – Users can batch export cluster plots, probe ID lists, GCTs, or a PDF containing a matrix of plots for the selected clusters.
- Multiple GCT export options. For each cluster, can export; clustered samples, the clustered class-means, all samples, or all class means (class-means option only available if a CLS has been specified).
- Users can optionally compress all exported files to a single ZIP file.

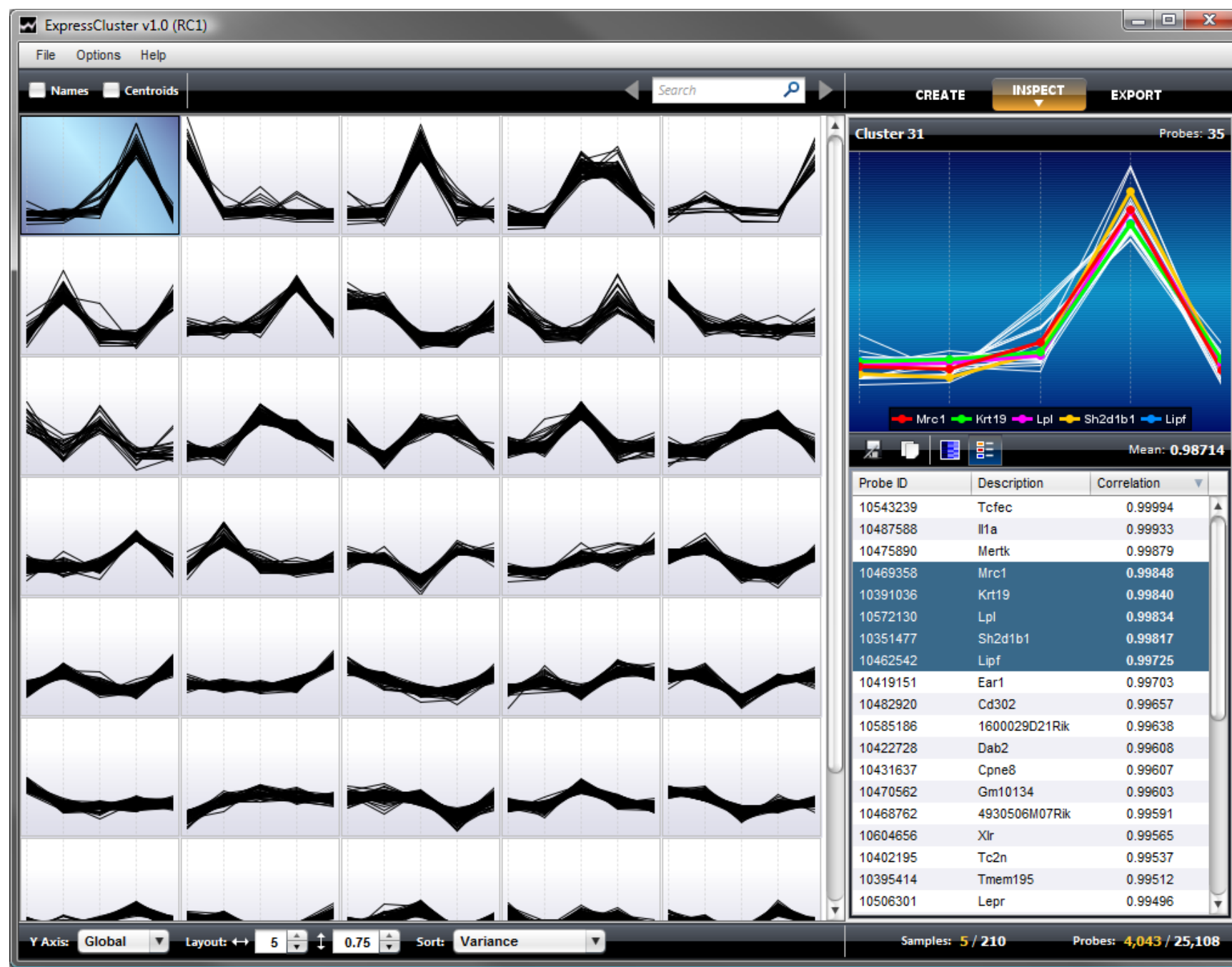
**Development Notes:**

- 1) This version is a "release candidate", meaning beyond beta, but not quite final. All comments/suggestions are welcome.
- 2) Full documentation and "Help" menu options will be available in the final release.
- 3) PCA-based K-Means clustering (i.e. k-means clustering on the component scores) is technically already implemented, but is not currently enabled so you cannot select it. I need to polish off a few things still and I've not yet decided on whether it will be in the first release or the next.
- 4) ExpressCluster was written in such a way that it allows for additional clustering algorithms to be used in the future (i.e. PAM etc).

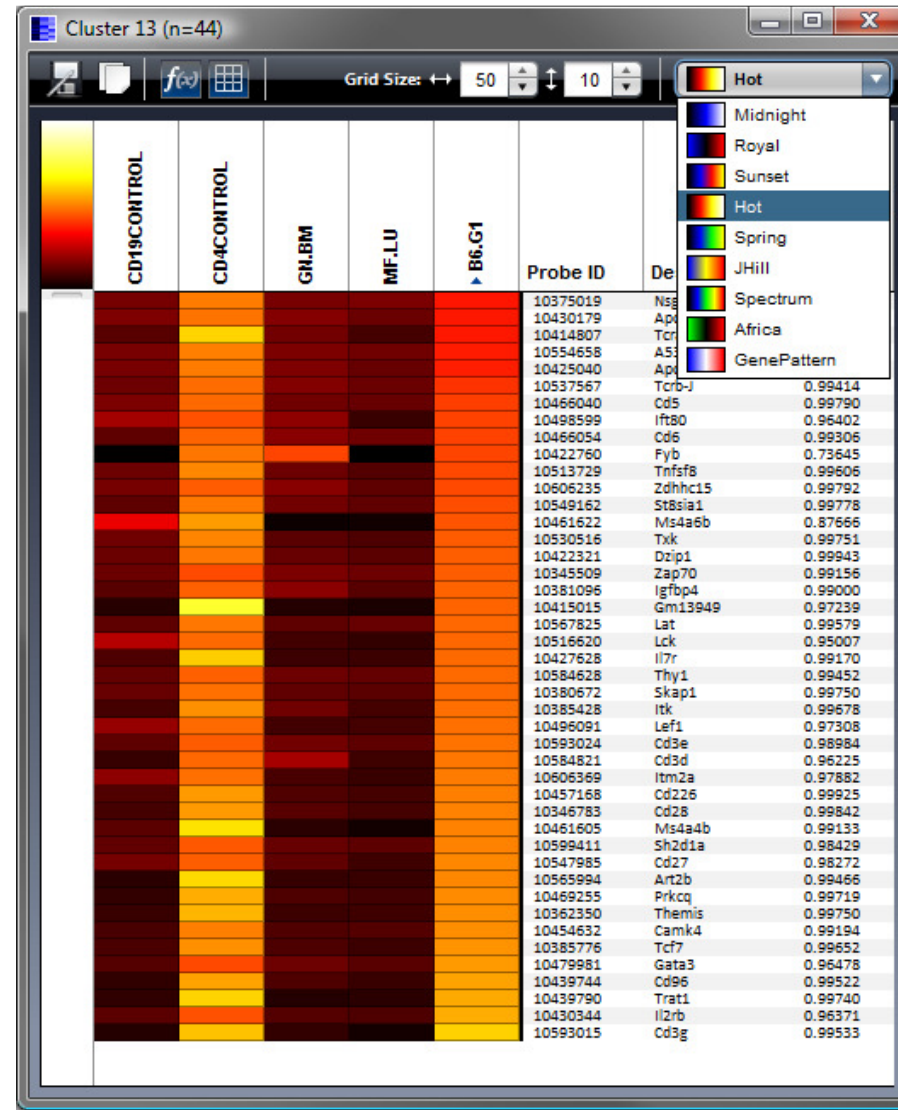
## Main Screen – Cluster Creation



## Exploring Results



## Cluster Heatmaps



## Exporting Plots & Data

ExpressCluster v1.0 (RC1)

File Options Help

Names Centroids

CREATE INSPECT EXPORT

DETAILS

☐ SELECTED ☐ ALL

DATA: ☒ Probe ID lists (TXT)  
☒ Datasets (GCT)  
↳ Current Classes

PLOTS: ☒ Individual plots (PNG)  
☒ Cluster Matrix (PDF)  
↳ 3 × 3

Total Files: 17 Create .ZIP file: ☐

Analysis Name (prefix for all output files):  
Test Dataset  
C:\

EXPORT...

SELECTED CLUSTERS 5 / 40

| Name       | Probes | Mean Correlation |
|------------|--------|------------------|
| Cluster 31 | 35     | 0.98714          |
| Cluster 21 | 64     | 0.98913          |
| Cluster 13 | 44     | 0.97799          |
| Cluster 11 | 33     | 0.94320          |
| Cluster 29 | 116    | 0.97270          |

Y Axis: Global Layout: 5 0.75 Sort: Variance

Samples: 5 / 210 Probes: 4,043 / 25,108

