Genedata Expressionist®

Biomarker Edition

Biomarkers are set to catalyze biomedical research. Critical for targeted therapeutics and increasingly used to predict drug safety, biomarkers find their application not just in diagnostics but also as research tools. Discovering and developing them pose technical and scientific challenges comparable to drug discovery.

Genedata Expressionist Biomarkers Edition is an enterprise-level scientific computing solution tailored to meet this challenge. Users are placed at the center of a powerful analysis platform which tightly integrates statistical processing tools with the biological knowledge needed to develop effective markers of disease.

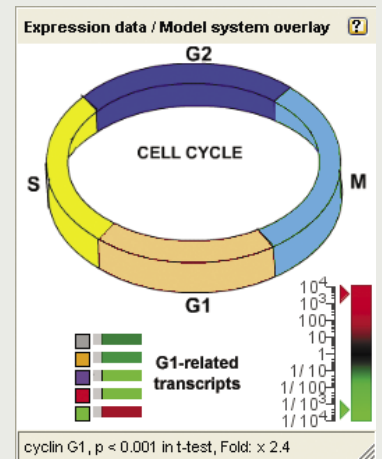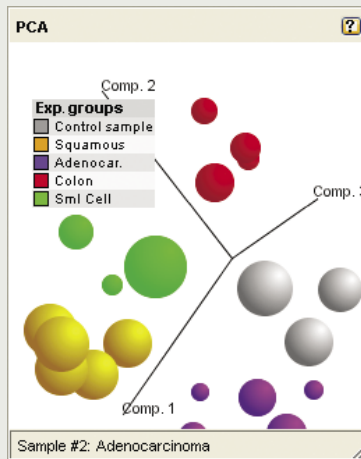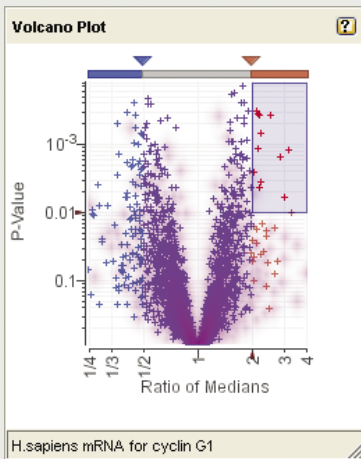| Expressionist® Biomarkers Edition	Advantages for biomarker discovery
An enterprise-level scientific computing solution for biomarker discovery	⤑ Combined analysis of transciptomic, proteomic and metabolomic experiments
	⤑ Integrated workflow streamlines data management and result interpretation
	⤑ Advanced analysis functions to characterize disease

Integrated and highly scalable platform for data analysis and interpretation
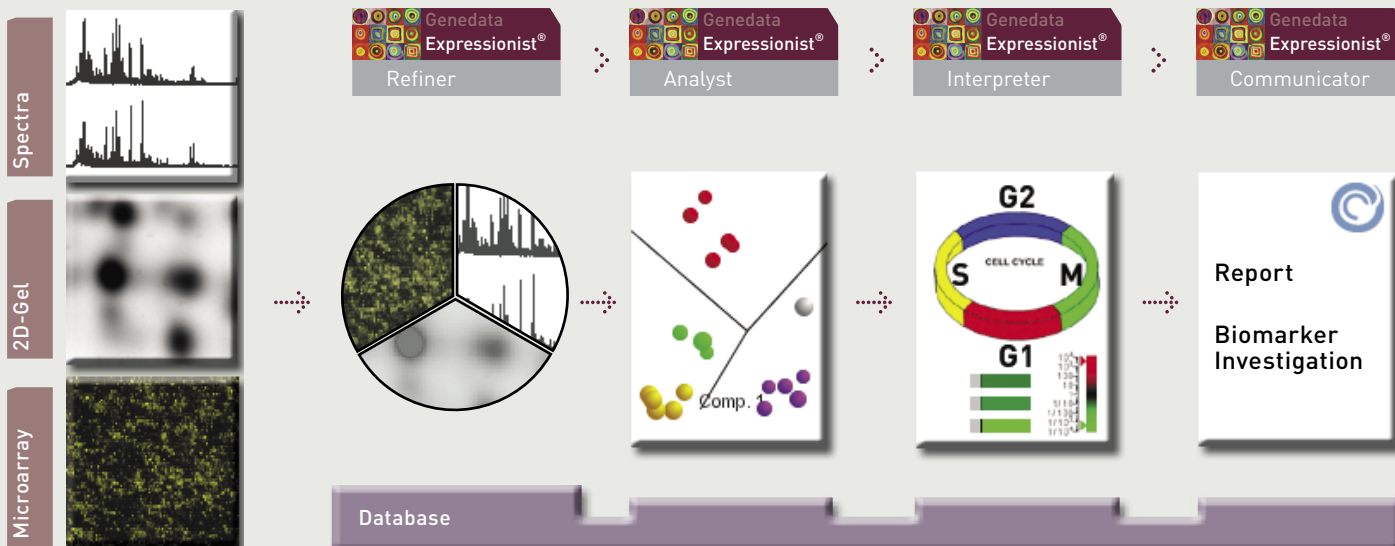
Overview →→→

Transcriptomic, proteomic and metabolomic data are loaded into Expressionist via Refiner and evaluated for quality issues. Data management, analysis and result reporting functions are tightly integrated. Sophisticated analysis tools can be applied efficiently and consistently as a workflow.

Volcano Plot [?]

P-Value

10^{-3}

0.01

0.1

1/4 1/3 1/2 1 2 3 4
Ratio of Medians

H.sapiens mRNA for cyclin G1

PCA [?]

Comp. 2

Exp. groups
- Control sample
- Squamous
- Adenocar.
- Colon
- Sml Cell

Comp. 3

Comp. 1

Sample #2: Adenocarcinoma

Expression data / Model system overlay [?]

G2

CELL CYCLE

S   M

G1

10^4
10^3
100
10
1
1/10
1/100
1/10^3
1/10^4

G1-related transcripts

cyclin G1, p < 0.001 in t-test, Fold: x 2.4

→→→
PCA plot (middle) shows principal components analysis performed on transcript profiles of 4 major cancer types (plus controls). Each cancer type (colored spheres) is confined to a discrete cluster. The clustering indicates that there are robust differences in expression behavior between each cancer.

An early phase of biomarker discovery requires integration of results from different technologies. Statistical tools are needed to identify robust biological correlates of disease. Later, results must be interpreted in the context of disease models, guiding the development of markers for diagnostic and research applications.

Expressionist delivers these requirements as an integrated workflow. Scalable data processing and quality evaluation tools combine results from different technologies. Q-value (false discovery) estimation and sophisticated techniques such as PCA (above, middle) are performed easily. Information about protein-protein interactions and gene regulatory networks can be subjected to quantitative analysis. An advanced data management system automates storage and handling of results.

Multidimensional Scaling [?]

- Control sample
- Squamous
- Adenocar.
- Colon
- Sml Cell
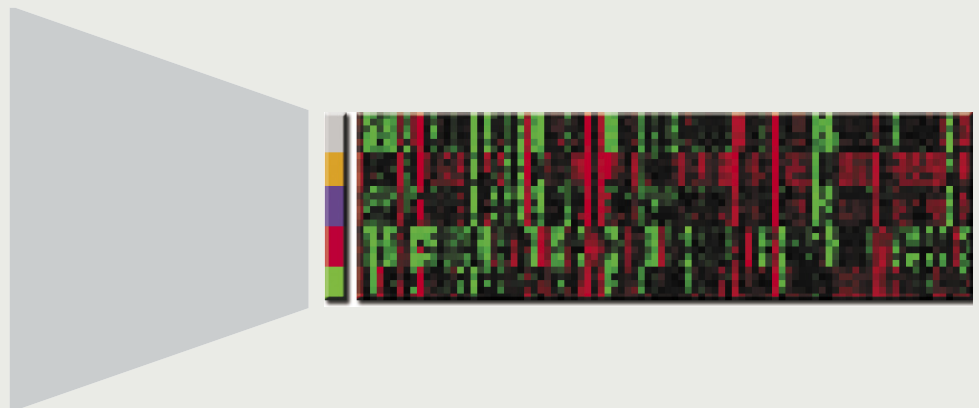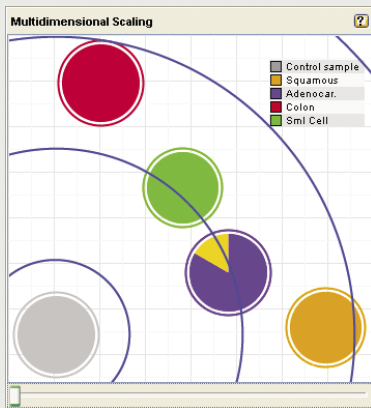
→→→
Supervised learning algorithms (eg SVM) identify 300 genes that optimally discriminate between 4 cancer types. Using this set of markers, samples are then classified on the basis of their gene expression profiles (right). Each cancer type, represented as a pie plot, is positioned on a map (left), a map representing overall expression activity. The yellow segment in the purple pie indicates an adenocarcinoma sample that has been classified as squamous (yellow pie). Classifications such as this may indicate data quality issues. The samples are classified on the basis of expression pattern alone.

Microarray data and other large scale technologies require advanced analysis tools. Expressionist incorporates powerful supervised learning algorithms (eg SVM, LDA) to refine the search for marker genes whose activity discriminate between different disease states.

The algorithms recover expression signatures of disease. Once characterized, these signatures can be used as powerful classification tools (above) to detect disease and evaluate treatments. Similar analysis can classify treatment responses according to known mode-of-action classes.

Expressionist's Biomarkers Edition incorporates integrated modules for data quality evaluation, analysis and result management. Based on industry-standard database tech-

nology, the platform integrates seamlessly with your existing IT infrastructure and expert support is available through Genedata's world-wide network of branches and subsidiaries.

Contact us at info@genedata.com or visit our website at www.genedata.com