Partek Genomics Suite 6.2
Summary of Features

*Partek Genomics Suite Specific Features*

**Exon Analysis**
- **Import & Normalization**
  - Import from Affymetrix CEL files
  - Exon-on-Gene Visualizations

**Statistical Analysis**
- Special Adaptation of Mixed Model ANOVA to detect Alternative Splicing

**Miscellaneous**
- Gene-level summarization
- Automatically link to genomic databases
  - NetAffx
  - UCSC Genome Browser
- Memory- and speed-optimized easily analyze over 1 million exons for large scale experiments including several hundred arrays.

**Chromosome Copy Number Analysis**
- **Import**
  - Import from Affymetrix CNAT files
  - Import from Affymetrix CEL files
  - Import from CNAG, dChip
- **Copy Number Estimation**
  - Affymetrix Chip
    - 10K
    - 100K
    - 500K
  - Expression
  - SNP
  - Exon
  - Tiling

**Miscellaneous**
- Normalize data with supplied reference or user supplied reference data
- Analyze Copy Number from Genotype Mapping, Expression, Exon, and Tiling Arrays

**Robust Multi-chip Analysis (RMA)**
- Fast RMA computation
- Uses constant memory (~300MB) for any number of arrays
- Import and normalize hundreds of chips on a desktop or notebook computer
- Simple graphic user interface

**NCBI GEO Importer/ Exporter**
- Download experiments from the NCBI Gene Expression Omnibus
- Publish your data to the NCBI GEO from your Partek spreadsheets using the Partek® GEO Publisher™

**Gene Lists**
- Easily create and export annotated gene lists
- Conveniently filter data using gene lists

**Annotations**
- Add annotations to results such as:
  - GenBank accession #
  - Common gene name
  - UniGene ID
  - Chromosome and location

**Built-in External Links**
- Built-in external links for the following external web databases:
  - Nucleotide
  - Genome
  - PubMed
  - LocusLink
  - GenBank
  - UniGene
  - NetAffx
  - GeneCards
  - UCSC Genome Browser

**Other Gene Expression-Specific Visualizations**
- Volcano Plot
- M vs. A Plot
- Easily create images and thumbnails from:
  - Original .CEL files
  - Summarized .CEL files
  - Difference/residual .CEL files

*All Partek Software Features*

**Analytical Spreadsheet®**
- No limit to the number of rows or the number of columns
- Import from flat files, Microsoft® Excel™
- Copy & paste to/from Microsoft® Windows Clipboard
- Automatic File Merging
- Import any ODBC compliant database including Oracle, SQL Server, Sybase, Informix, and Microsoft® Access™

Data Filtering
Row Filters
- Include/exclude data points based on one or more properties
- Randomly shuffle the observations
- Resample with replacement for bootstrap simulations
- Interactive data filtering using sliders
- Programmable filtering

Column Filters
- Include/exclude variables for variable selection
- Include/exclude variable groups

Summary Statistics
Measures of Central Tendency and Distribution
- Min, Max, Range
- Mean, Median
- Variance
- Standard Deviation
- Average Deviation
- Coefficient of Variation
- Root-mean Square
- Harmonic Mean
- Tukey’s Biweight
- Skewness, Kurtosis

Measures of Association
- Linear (Pearson’s) Correlation
- Rank (Spearman’s rho) Correlation
- Covariance
- Kendall’s Tau
- Apply to variables or observations

Data Transformations
Scaling
- Linear scaling
- Standardization (z-scores)
- Mean Centering
- Minimum Centering
- Maximum Centering
- Absolute Value Centering
- Apply to variables or observations

Normalization
- Log, Anti-log
- Power, Absolute Power
- Replacement by Ranks
- Box-Cox Auto Power
- Inversion
- Apply to variables or observations
- Variance stabilization
- Quantile normalization

Ranks & Scores
- Median
- Wilcoxon (Rank Order)
- Blom
- Tukey
- van der Waerden
- Savage

Smoothing
- Mean
- Weighted mean
- Median
- Moving average

Distance & Similarity Measures
- Euclidean
- Euclidean Squared
- Average Euclidean
- Mahalanobis
- Minkowski
- Average Minkowski
- Maximum Value
- Minimum Value
- Absolute Value (“City Block”)
- Shape Similarity
- Cosine Similarity
- Pearson’s Correlation
- Absolute Pearson’
- Spearman’s Correlation
- Absolute Spearman’s
- Kendall’s Tau
- Absolute Kendall’s Tau
- Canberra
- Bray-Curtis
- Tanimoto

List Manager
- Intersect multiple lists
- Create union of multiple lists
- Concatenate multiple lists
- Save & export annotated lists
- Venn Diagrams
- Query from the spreadsheet

Anomaly Detection
- Singular variables for class-labeled data
- Automatic detection of zero-variance variables
- Automatic detection of variables with no non-missing values
- Identify collinear variables

Missing Data
Partek supports missing values in your data using a variety of techniques.
- Summarization of missing cases on an observation basis or measurement basis
- Select or filter data based on the amount of missing data
- Case-wise deletion (ignore observations with missing data)
• Impute missing data using mean, median, floor, ceiling, constant, and K-nearest neighbor method

Random Number Generators
Partek provides a variety of random generators to support a variety of simulations.
- Normal
- Uniform
- Exponential
- Gamma
- Binomial
- Poisson

External Link Manager
- Link external applications, databases, and web resources
- Interactive link triggering from spreadsheet or graphics

Scripting Interface
- Embedded Tcl/Tk command interpreter
- Automate analyses
- Create custom user interfaces
- Record processing steps

Statistical Inference
Parametric Tests
- One-sample t-test
- One-sample z-test
- Two-sample t-test, equal variance
- Two-sample t-test, unequal variance
Analysis of Variance (ANOVA)
- N-way ANOVA (no limit to number of factors)
- No limit to the number of levels for each factor
- Handles balanced & unbalanced designs
- Random & fixed effects (mixed model)
- Variance component estimation
- Type III/Type IV sums of squares
- Repeated Measures
- Analysis of Covariance (ANCOVA)
- Any number of numeric covariates
- Supports flexible linear contrasts
- Pre-planned and post-hoc contrasts

Survival Analysis
- Cox regression
  - No limit to the number of predictors
  - Predictors can be numeric or categorical
  - Likelihood ratio test
  - Wald’s test
  - Score test
- Kaplan-Meier
- Log-Rank
- Wilcoxon-Gehan

Nonparametric Tests
- Mann-Whitney
- Kruskal-Wallis
- Chi-Square
- Fisher Exact
- McNemar
- Friedman
- Quade
- Kolmogorov-Smirnov
- Multiple test correction for p-values

Reporting
Flexible result tabulation includes:
- Corrected and uncorrected p-values
- Descriptive statistics (mean, min, max, standard deviation, standard error)
- Fold-change
- Ratio
- Difference in means

False Discovery Rate
- Dunn-Sidak
- Bonferroni
- Bootstrap & permutation tests for Type I error
- Benjamini & Hochberg step-up and step-down false discovery rate

Principal Components Analysis (PCA)
- R & Q-mode analysis
- Three dispersion methods
- Factor loadings
- Handles millions of variables
- View variable and principal component relations in Bi-plot
- Three types of eigenvector scaling (normalized, V-, W-vector)

Automatically detects nested/nesting relationships

Multidimensional Scaling (MDS)
- Metric & Non-metric MDS
- Twenty distance measures
- Three initialization methods (PCs, Orthogonal, and Random)
- Configurable power transformation
- Configurable learning rate and inhibition factor
- Early stopping by minimum error or number of iterations
- Export mapping, input set distance/dissimilarity, and output set Euclidean distances

Correspondence Analysis
- Graphically analyze the similarity of categorical variables from a contingency table
- 1, 2, and 3 dimensional correspondence analysis bi-plots of contingency tables
- Export frequencies, row/column profiles, percents, row/column residuals
- Contingency table analyses statistics including Chi-square, total inertia, Cramer's V, degrees of freedom, total inertia, contingency coefficient, and uncertainty

Cluster Analysis
- K-means and Fuzzy C-Means partitioning methods
- Divisive & agglomerative hierarchical clustering
- Interactive dendograms & intensity plots for clustering observations & variables simultaneously
- Twenty measures of similarity
- Find the best partition by evaluating multiple number of cluster partitions using the Davies-Bouldin metric
- Evaluate the clustering results using Davies-Bouldin and Modified Hubert internal criteria or Rand and Jaccard external criteria

- Export Projection scores
- Export eigenvalues & eigenvectors
- SCREE plots
- Export projection errors
- Export residuals and reconstructed data matrix

External Link Manager
- Link external applications, databases, and web resources
- Interactive link triggering from spreadsheet or graphics

Scripting Interface
- Embedded Tcl/Tk command interpreter
- Automate analyses
- Create custom user interfaces
- Record processing steps

Statistical Inference
Parametric Tests
- One-sample t-test
- One-sample z-test
- Two-sample t-test, equal variance
- Two-sample t-test, unequal variance
Analysis of Variance (ANOVA)
- N-way ANOVA (no limit to number of factors)
- No limit to the number of levels for each factor
- Handles balanced & unbalanced designs
- Random & fixed effects (mixed model)
- Variance component estimation
- Type III/Type IV sums of squares
- Repeated Measures
- Analysis of Covariance (ANCOVA)
- Any number of numeric covariates
- Supports flexible linear contrasts
- Pre-planned and post-hoc contrasts

Survival Analysis
- Cox regression
  - No limit to the number of predictors
  - Predictors can be numeric or categorical
  - Likelihood ratio test
  - Wald’s test
  - Score test
- Kaplan-Meier
- Log-Rank
- Wilcoxon-Gehan

Nonparametric Tests
- Mann-Whitney
- Kruskal-Wallis
- Chi-Square
- Fisher Exact
- McNemar
- Friedman
- Quade
- Kolmogorov-Smirnov
- Multiple test correction for p-values

Reporting
Flexible result tabulation includes:
- Corrected and uncorrected p-values
- Descriptive statistics (mean, min, max, standard deviation, standard error)
- Fold-change
- Ratio
- Difference in means

False Discovery Rate
- Dunn-Sidak
- Bonferroni
- Bootstrap & permutation tests for Type I error
- Benjamini & Hochberg step-up and step-down false discovery rate

Principal Components Analysis (PCA)
- R & Q-mode analysis
- Three dispersion methods
- Factor loadings
- Handles millions of variables
- View variable and principal component relations in Bi-plot
- Three types of eigenvector scaling (normalized, V-, W-vector)
Predictive (Diagnostic) Modeling

- Train, validate, and test on hundreds of models consisting of different variable selection and classification methods
- Uni-biased accuracy estimate from 2-level cross-validation
- Model selection based on either Correct Rate, Normalized Correct Rate, or Kappa
- Variable selection filter approaches: ANOVA and Shrinking Centroids
- Variable selection wrapper approaches: forward selection, backward elimination, exhaustive and genetic algorithm
- Linear Discriminant analysis
- Bayesian (Quadratic) Classifier
- Multi-prototype classifier
- Neural Networks
- K-nearest neighbor configurable with multiple neighbors and distance measures
- Nearest centroid classifier
- K-nearest neighbor and nearest centroid classifiers are not limited by the "curse of dimensionality"
- Diagnostic model validation using built-in cross-validation, nested cross-validation, jackknife and bootstrap
- Built-in cross validation, nested cross-validation/variable selection for all predictive models for use with building models from very high dimensional data
- Automatic variable selection search methods for all predictive models include:
  - Forward selection
  - Backward elimination
  - Exhaustive evaluation of all combinations
  - Genetic algorithm
- Diagnostic model deployment using Partek Black Boxes
- Receiver Operating Characteristic (ROC) allow trade off of Sensitivity vs. Specificity

Variable Selection

- Backward elimination
- Forward selection
- Genetic Algorithm
- Exhaustive evaluation
- Discriminant Analysis
- Analysis of Variance (ANOVA)
- Linear regression
- Nonparametric classifier selection
- Create custom evaluation criteria
- Stop at a specified # of variables
- Stop at a specified error threshold
- Automatically filter original data to any evaluated subset of variables

Performance Reporting

- Confusion matrix
- Classification summary including percent correct and standard errors
- Receiver Operating Characteristic (ROC)
- R2
- Rank & Linear correlation

Receiver Operating Characteristics (ROC)

- Evaluate the classifier performance over the entire range of operation
- Visual and quantitative assessment of a classifier’s performance
- Built-in to automatic cross-validation
- Area Under Curve (AUC)

K-Nearest Neighbor Classifier

- Twenty distance measures

Nearest Centroid Classifier

- Similar to PAM (Prediction Analysis for Microarray from Stanford University)

Discriminant Analysis

- Linear model
- Quadratic (Bayes) model
- Export data with class assignments, posterior probabilities
- Export Mahalanobis distances

Multi-layer Perceptron (MLP)

- Network adapts input/output biases, activation steepness, dynamic range, and weights
- Batch or on-line updating
- Export weights, biases, range, and steepness parameters
- Export predicted categories or response variable
- Back-propagation
- Scaled Conjugate Gradient (SCG) (Hestenes-Stiefel, Polak-Ribiere)
- Adjustable weight decay
- Built-in stopped training with validation set
- Automatically scale inputs and outputs
- Categorical and continuous variable modeling

Multiple Prototype Classifier

- Model & classify data using prototypes
- No software imposed limits on the number of prototypes
- LVQ1, LVQ2.1, LVQ3 training methods
- Three learning rate functions
- Twenty measures of distance/dissimilarity
- Discrete & Fuzzy classifications
- Export classified data and learned prototypes
- Batch and on-line updating

Connectivity

Database Connectivity

- Import from any ODBC-compliant database:
  - Oracle
  - SQL Server
  - Sybase
  - Informix
• Microsoft® Access
• Fully programmable via Tcl scripting language
• Can generate dynamic queries
• Customizable by the user

Web Integration
• Load data dynamically from the web
• Load and run command scripts dynamically from the web
• Issue dynamic HTTP requests, process results

Inter-Process Communication (IPC)
• Use TCP/IP for bulk data transfer
• DDE Client / DDE Server
• Load Java Virtual Machine
• COM Client

Help
• Tutorials
• User Manuals
• Command Reference Manuals

Licensing Models
• FLEXlm® license manager
• Node locked user licenses
• Concurrent user licenses
• Temporary license check-out

System Requirements
• 128 MB RAM
• 100 MB free disk space
• 16-bit color minimum
• 800x600 screen resolution

Platforms Supported
• Windows XP/Me/2000/NT/98
• Linux
• Macintosh – coming soon