Bioinformatic Services

SEQ LL

SEQUENCE THE LOWER LIMIT

About SeqLL

As we push the limits of Next Gen Sequencing (NGS) methods and protocols we are able to explore broader and more sophisticated applications that address complex biological questions. The corresponding data set generated becomes larger and more complex. The old standard of applying a basic data set review may not adequately address subtle differences that your research requires.

Bioinformatics

Maximizing the value of your work begins with creating a protocol that specifies all areas from specimen collection, sample processing and finally optimizing the data set review.

SeqLL bioinformatics specialists are available to work with your team from the initial planning of the methods and protocols through the sequencing stages and finally to designing the type and level of bioinformatics review that will be applied to the project. Our experienced experts combine a deep technical knowledge with experience and the latest in bioinformatics analysis programs.

Why True Single Molecule Sequencing?

True Single Molecule Sequencing (tSMS) allows a new level of sensitivity to be obtained without bias. Combing tSMS with the appropriate bioinformatics review can make the difference in determining the significance of low level expression molecules. This allows the researcher to then decide what parameters such as reads per sample should be increased or attention focused on other areas to determine the functional relevancy of the molecule.

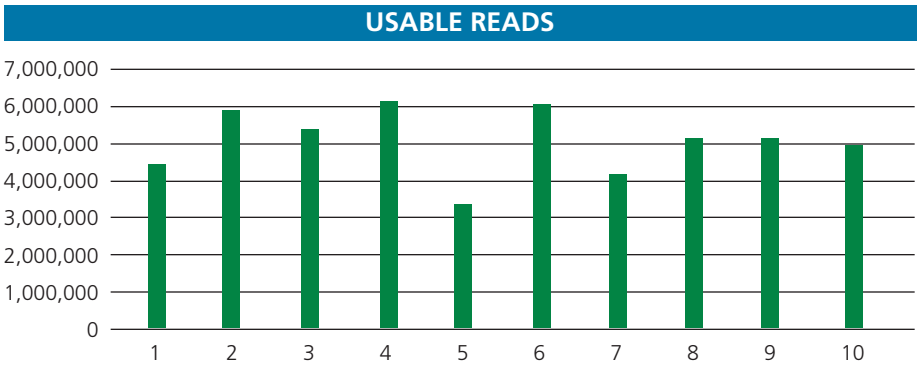
The SeqLL bioinformatics group offers multiple levels of data evaluation to accommodate a wide range from basic requirements to complex needs.

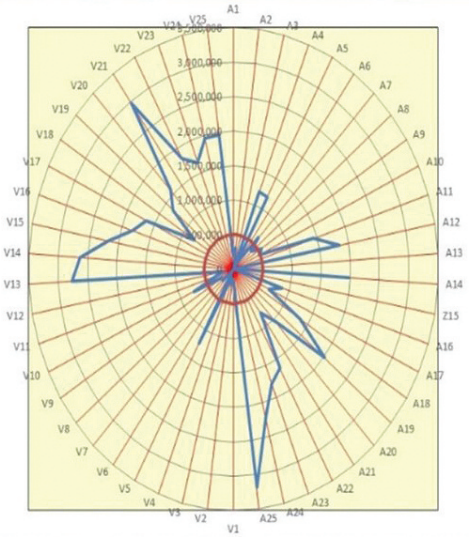
Basic Bioinformatics

Data interpretation starts with informed reads, quality control charts, Digital Gene Expression (DGE), box plots and bar graphs. These include raw reads, filtered reads, aligned reads (when applicable).

Standard Final Sequencing Report

- The results from each project are presented in a clear format (hard copy and electronic) explaining the protocol, sequencing parameters, the raw data and basic data review.
- Files are available in standard formats such as BAM, FASTA or FASTQ.
- Reads are quality filtered data reduced to mapped reads, a list of unique sequences and their copy number.
- Alignments being checked and reported as the quality of the run is assessed, reviewed and discussed.





Usable Reads

Informative reads—In a typical experiment it is expected that we can expect from 500,000 – 4,000,000 informative reads. The number is dependent on a number of factors including if the RNA was degraded.

Group	Reference	Usable	Aligned	% Filtered	% Aligned	Mean Length	Length Filtered	Ctrl Filtered	Lead T Filtered	IncRate Filtered	BAO Filtered	AT Filtered	Del	Ins	Sub	Error
Bene 1	Oligo BFSC12	6,965,280	4,457,779	45.3%	64.0%	34.55	40.30%	3.88%	1.25%	1.10%	21.01%	0.31%	2.09%	2.37%	0.45%	4.04%
Bene 2	Oligo BFSC12	5,298,049	3,343,069	40.0%	63.1%	34.25	46.20%	4.75%	2.54%	2.31%	22.95%	0.41%	2.10%	2.06%	0.47%	4.35%

Advanced Bioinformatic Data Analysis

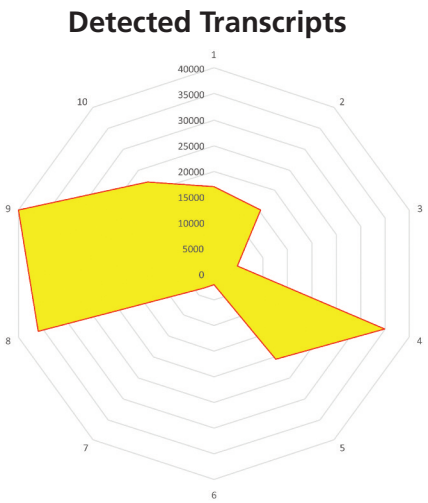
The bioinformatics team can provide an advanced evaluation of the data set generated. To maximize the value of the data set, the team will work with research groups to design the final evaluation into the initial protocol.

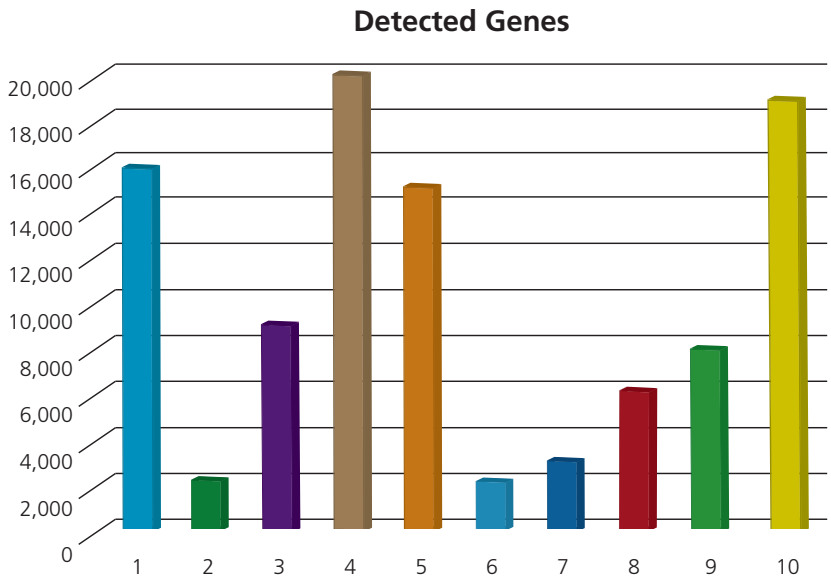
In depth analysis of the data set may provide an insight into aspects of the data that are typically overlooked or not obvious to many researchers.

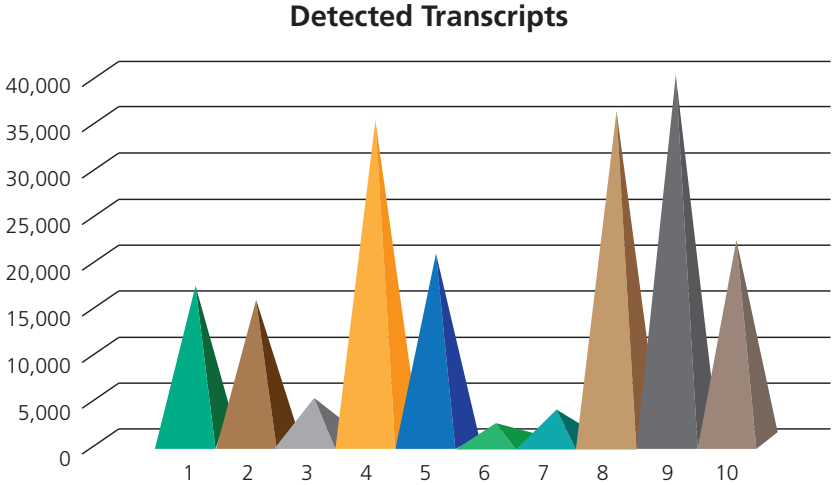
Typical Plots from a standard data review include:

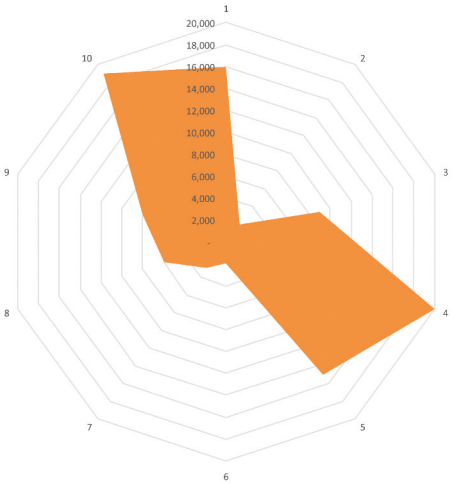
- Digital Gene Expression tables (DGE)
- Differentially expressed genes can be located and reported.
- MW and MA plots vs Controls to provide insight into the effect
- Principal Component Analyses
- The biological function of differentially expressed genes can be analyzed and pathways analyses plots determined.

Bar plots show the numbers of detected transcripts (upper plot) and detected genes (lower plot) — those transcripts/genes whose expression was not null in examined samples.





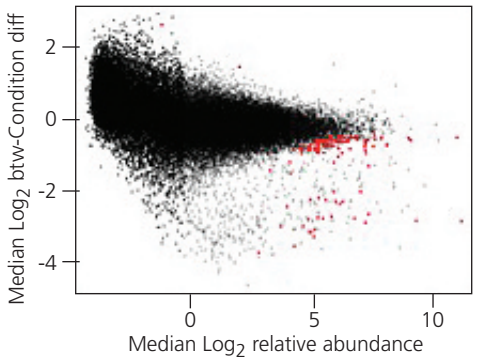




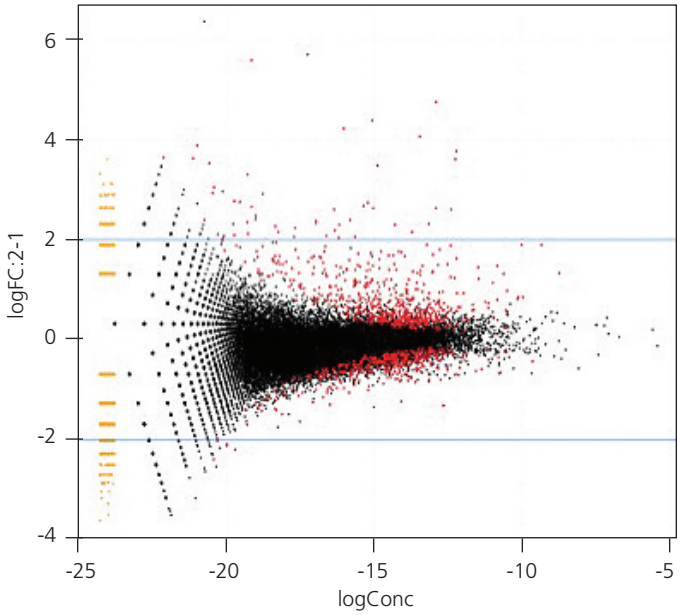
MA plots of Treated Specimens versus controls indicate the efficacy of the treatment.

The MA plot is a plot of median $\log_2$ expression difference between examined conditions vs. median $\log_2$ relative abundance — in a usual MA plot these are mean values.

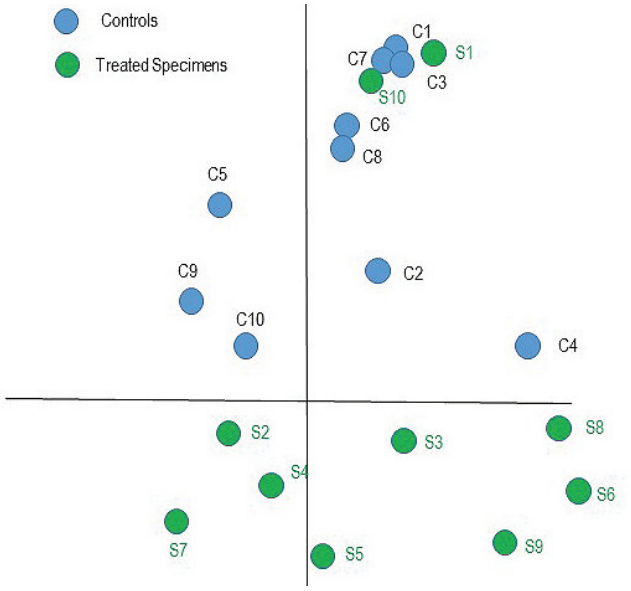
The MW plot is a plot of the between-condition fold change (M) vs. with-in-condition fold change (W) (MW). Features where the fdr statistic is <0.01, are shown in red. Features where the median centered log ratio (clr) value is below the geometric mean are highlighted in black if they are not significant. Those where the median clr value is greater than the geometric mean are shown in gray.



Plot Using Tagwise Dispersion



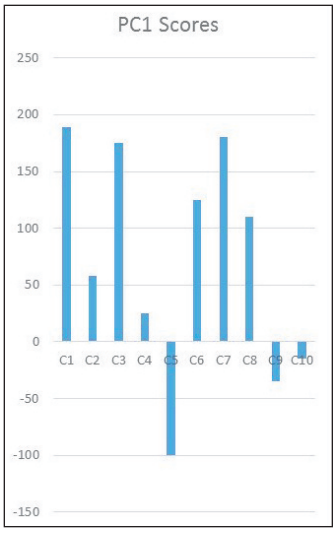
PCA Mapping based on RPMK Expression

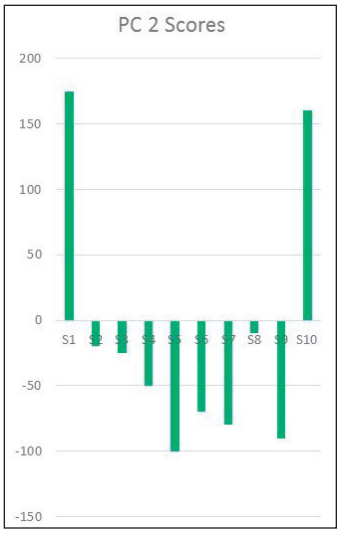


PC1-PC2 projection of the Controls versus the treated samples

This is a picture of PC1-PC2 projection with samples that have informative reads above the threshold level. PC1 and PC2 scores are of the samples are computed from their DGE. Samples with low informative reads are normally excluded. The level is determined by the bioinformatic protocol. The green points correspond to the treated samples and the blue to untreated samples. A number near the point encodes the corresponding sample: e.g. blue point with label "C1" is sample #1 and a green one with label "5" is "treated sample #5", etc.

Thus, the treated samples that can't be separated from controls were excluded and the differential gene expression analysis is repeated.



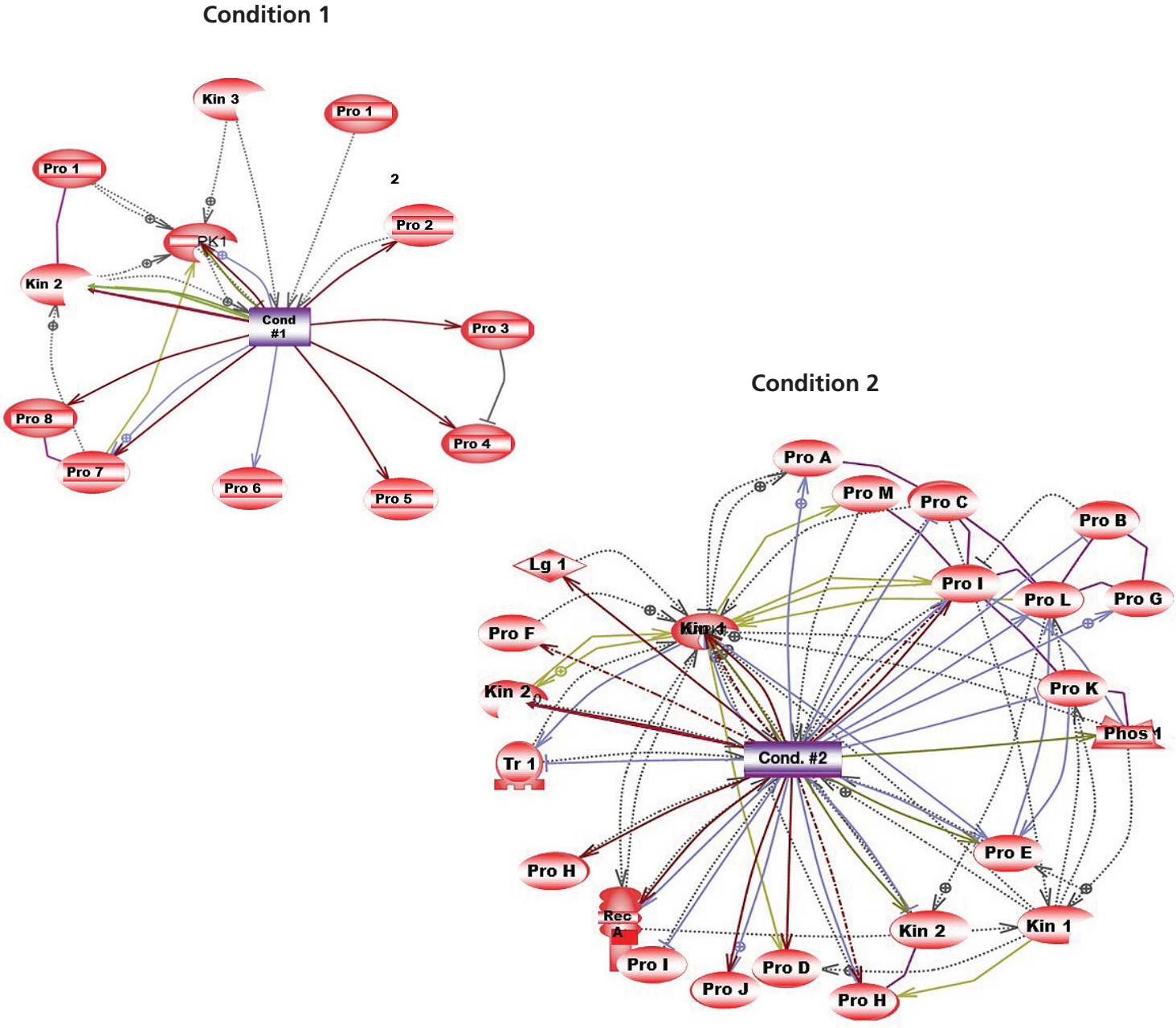


Relation Plots

The list of differentially expressed genes is used to search for all known relations . A printout is provided of each DEG with the known relationships, reference locations and other relevant information.

In	Rel	Out	Type	Sentence	TextRef	# of Ref	DOI	
XXX	--->	Depressive Disorder, Major	Regulation	Finally, ceftriaxone has an antidepressant behavioral profile in the tail suspension and forced swim tests supporting the possibility that the enhancement of GLT1 activity represents a new pharmacological strategy for treating major depressive disorder. The ability of the N-methyl-D-aspartate antagonist riluzole (Frizzo et al, 2004) to upregulate EAAT2 gene and protein expression, and the antidepressant effects of riluzole in treatment-resistant patients (Sanacora et al, 2007; Zarate et al, 2005), is consistent with the EAAT2 decrease in major depressive disorder and suicide.	info:pmid/19376090#body:115, in-fo:pmid/18923405#body:538	2	10.1016/j.bcp.2009.04.008	NEW PATH

Relationship Plots allow key proteins to be identified and more easily followed and understood for individual disease conditions.

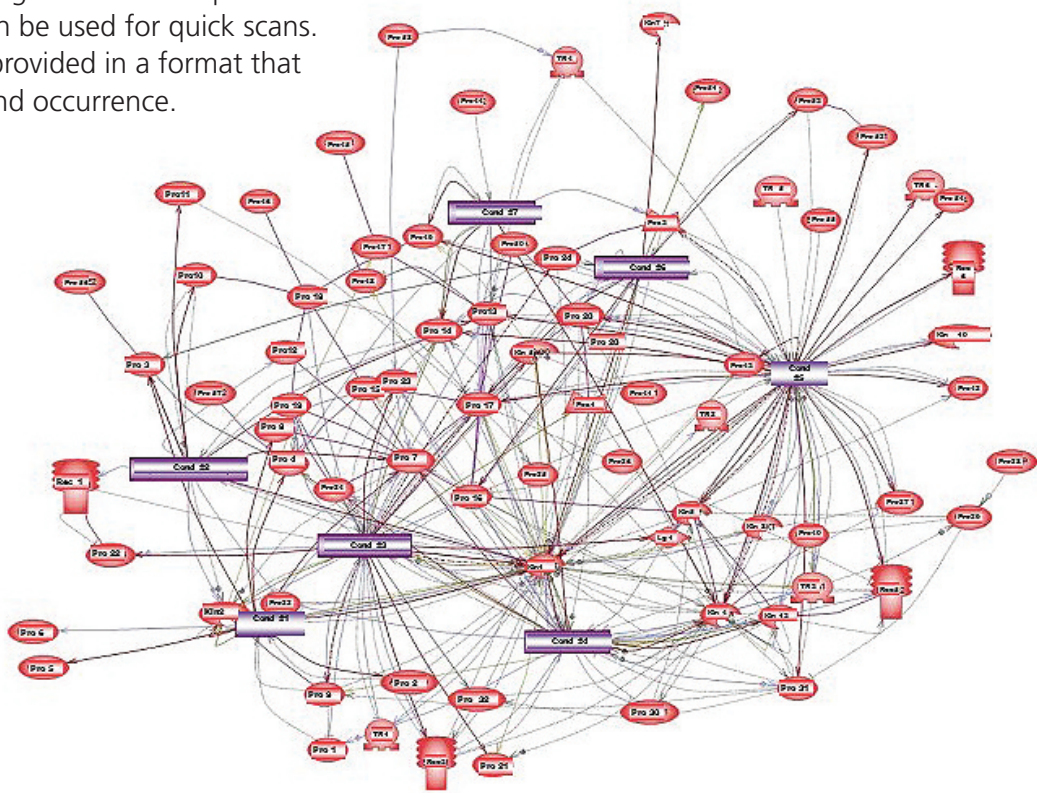


Each Protein and mechanism of operation is color coded for easy identification.

PROTEINS ARE COLORED BY GROUP:

- Protein kinases
- Kinases
- Phosphatases
- miRNAs
- Ligands
- Transcription Factors
- Receptors
- Protein
- Disease
- Cell Process
- Functional Class
- Small Molecule
- Complex
- Clinical Parameter
- Treatment
- Regulation
- Mol Transport
- Chemical Reaction
- Direct Regulation
- Expression
- Mol Synthesis
- Promoter Binding
- Binding
- Prot Modification
- miRNA Effect
- State Change
- Quantitative Change
- Genetic Change
- Functional Association
- Biomarker
- Clinical Trial

A graphical representation describing the relationships between multiple disease states can be used for quick scans. The actual detailed information is provided in a format that can be easily sorted by relevance and occurrence.



Custom Data Analysis

If additional, more complex bioinformatics needs such as functional gene information mining or comparable genomic analysis may be required. A custom bioinformatics review can be created allowing more in depth mining of the data set. Typically this expands the report to include Differentially Expressed Gene (DEG), gene ontology and pathway analysis to gain additional insight.

Custom reports created that are submissions grade and publications ready. Identification of key elements inside the data sets can be undertaken. Biomarkers; key If additional, more complex bioinformatics needs such as functional gene information mining or comparable genomic analysis.

317 New Boston Street, S210
Woburn, MA 01801 USA
+1 781.460.6016
custserv@SeqLL.com
www.seqll.com