

# Fast, user-friendly, and powerful SOLiD™ data analysis

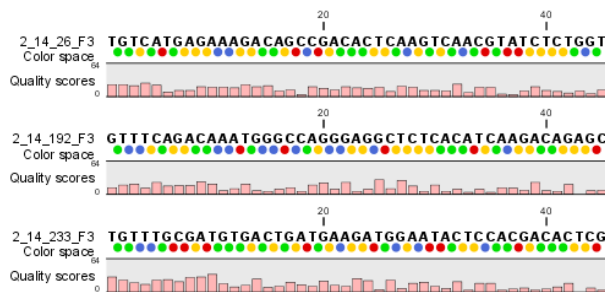
Powered by CLC Genomics Workbench

CLC Genomics Workbench enables everyone to rapidly analyze and visualize the huge amounts of data generated by the SOLiD™ System. The user-friendly and intuitive interface essentially takes High Throughput Analysis away from hardcore bioinformatics programmers doing command-line scripts, and hands it to scientists searching for biological results.

Furthermore, the versatile nature of CLC Genomics Workbench allows it to blend seamlessly into existing sequencing analysis workflows, easing implementation and maximizing return on investment.

CLC Genomics Workbench includes a wide range of SOLiD™ System data analyses carried out in native color space, including:

- Genomics analyses
- Transcriptomics analyses
- Epigenomics analyses



Part of a sequence list showing sequences in color space and base space, as well as their quality scores.

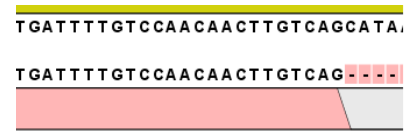
## Genomics Features

CLC Genomics Workbench Incorporates SIMD based algorithms to yield a significant speed-up of data analyses related to Targeted Resequencing, Whole Genome Resequencing, and de novo Sequencing.

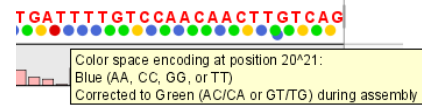
The reference assembly functionality as well as the de novo assembly functionality of CLC Genomics Workbench support both short (>50bp) and long (<50bp) reads, they support paired-ends reads, and they support gapped and ungapped alignments.

Reference assembly / mapping of SOLiD™ System data is carried out in native color space. Up to 80% more hits have been found when assembling 35mer SOLiD data

in color space, compared to assembling the same data in base space.



Reference assembly as well as de novo assembly support the analysis of different kinds of data at the same time. An example would be the de novo assembly of Sanger data and SOLiD™ System paired end data in the same analysis.



One of the dots have both a blue and a green color. This is because this color has been corrected during assembly. Putting the mouse on the dot displays the small explanatory message.

Other genomics features include support for multiplexing of several analyses in batches, SNP detection, DIP detection, and identification of genomic rearrangements.

## Transcriptomics Features

CLC Genomics Workbench includes mRNA seq for Whole Transcriptome Analyses based on the RPKM model (Reads Per Kilobase exon Model per million mapped reads). This is a good and easy way for normalizing values for the expression level of a gene when using Digital Gene Expression.

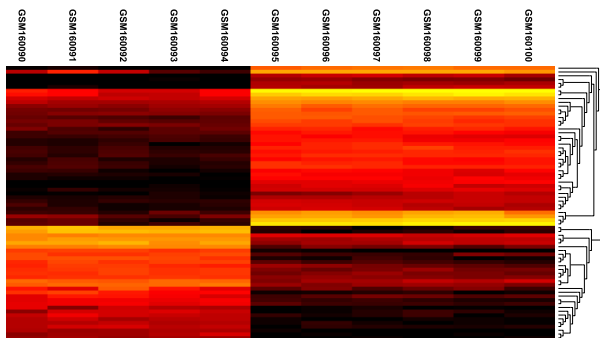
Feature ID	Expression values	Transcripts	Unique gene reads	Unique exon reads	Total gene reads
ABHD8	5.416,87	1	656	595	695
ABHD9	21,02	1	18	2	32
AKAP8	673,58	1	222	124	361
AKAP8L	4.311,30	1	772	478	697
ANKRD41	125,49	1	27	20	31
AP1M1	2.749,13	1	426	326	468
ARMC6	1.238,56	1	201	149	230
ARRDC2	1.034,80	2	236	160	333
ATP13A1	1.332,95	1	325	244	341
B3GNT3	0,00	1	36	0	76
BRD4	1.427,11	2	656	554	693
BST2	1.479,23	1	67	60	80
C19orf42	720,63	1	91	51	107
C19orf44	943,97	1	92	13	316
C19orf50	2.653,11	1	264	195	307
C19orf60	5.254,14	2	346	242	359

A table view of an expression sample generated from a sequence file of NGS mRNA reads.

For studies of expression levels of pre-defined genes, CLC Genomics Workbench includes Tag Based Transcriptomics for counting and analyzing sequence tags relative to the genes of interest.

From the initial identification of expression levels, CLC Genomics Workbench has tools to support a full workflow of downstream analyses including visual quality control tools, such as principal component blots and box plots, transformation and normalization tools, tools for statistical testing and false discovery rate control, clustering algorithms, heat-map visualization, and tests on gene annotations, such as Hyper Geometric tests and Gene Set Enrichment analysis.

In addition to the above, the mRNA seq functionality can be used for discovering putative exons (intronic regions containing a relatively high number of matches). This could indicate that this region is present in a novel transcript variant. Using existing primer design tools in the Workbench, primers can be designed to verify putative exons transcripts.

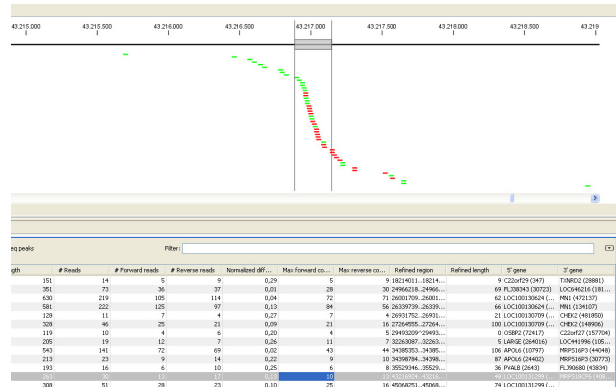


Heatmap: Hierarchical clustering of features.

Digital gene expression analyses in CLC Genomics Workbench are carried out in native color space data format for optimizing the scientific results of the SOLiD™ System.

### Epigenomics Features

When the SOLiD™ System is used for determining DNA binding sites on the genome through Chromatin immunoprecipitation (ChIP) sequencing, CLC Genomics Workbench can be used for a range of analyses - from visualization and the initial mapping of reads to finding statistically significant peaks, identifying and analyzing proteins or interest, and a range of other downstream analyses.



A list of detected ChIP seq peaks found on a human chromosome 22, and a graphical display that shows the match reads of a peak selected in the table

The ChIP seq analyses are carried out in native color space data format for optimizing the scientific results of the SOLiD™ System.

### Other Bioinformatics Features

CLC Genomics Workbench includes more than 100 downstream bioinformatics analyses, all fully integrated in one single user-friendly, graphically based scientific application.

Some features are BLAST including graphical display of results, primer design, molecular cloning, several types of alignments, phylogenetic analyses, RNA secondary structure prediction, restriction enzyme analyses and management, NCBI and Swiss prot database searches, 3D protein views, antigenicity and hydrophobicity analyses, and much more.

### Purchasing Information

For CLC Genomics Workbench purchasing information, technical support, or to download a trial, visit [www.clcbio.com/genomics](http://www.clcbio.com/genomics) or contact CLC bio at [info@clcbio.com](mailto:info@clcbio.com), 1.800.208.5981 (U.S.), or +45.7022.5509 (outside U.S.).



### System requirements

- Mac OSX 10.4 or later (including Intel-based Macs)
- Windows 2000, Windows XP, or Windows Vista
- Linux: Redhat or SuSE
- 32 bit version and 64 bit versions of operating system on all platforms
- 4 GB RAM required
- 4 to 16 GB required for large assemblies
- 1024x768 display recommended