

# Comprehensive and User-Friendly Analysis of Illumina Sequencing Data

CLC bio's Genomics Workbench provides a complete workflow for turning massive amounts of raw data from the Genome Analyzer into world-class scientific results—on a laptop.

## INTRODUCTION

The Illumina Genome Analyzer represents a breakthrough in the volume of sequence data that can routinely be generated by any lab. CLC bio's Genomics Workbench is a user-friendly application that provides cutting-edge analysis and visualization tools for researchers to get the most out of their data, whether they generate their own sequences in-house, or use a service provider or core lab (Figure 1).

Like the Genome Analyzer, Genomics Workbench is streamlined and easy to use. The software runs on standard computers and operating systems, and provides users a well-organized and intuitive graphical interface. Genomics Workbench leverages capabilities common in multimedia computers to accelerate assembly. *De novo* assembly, reference-guided assembly, SNP detection, digital gene

expression, and other analysis and visualization tasks are all available in a single software package.

## WIZARD-BASED TOOLBOX

Genomics Workbench provides a graphical environment and a wide range of tools. Data analysis tools are selected from a toolbox, which launch in a wizard-based interface for intuitive customized analysis and display options. Each step in the wizard is supported with online help.

## Import Data

Genomics Workbench directly imports quality-calibrated reads generated by Genome Analyzer Pipeline software. Thus, all downstream analysis takes full advantage of the high-quality output of the Genome Analyzer. Sequence.txt files exported from the Pipeline software for single read or paired-end sequenc-

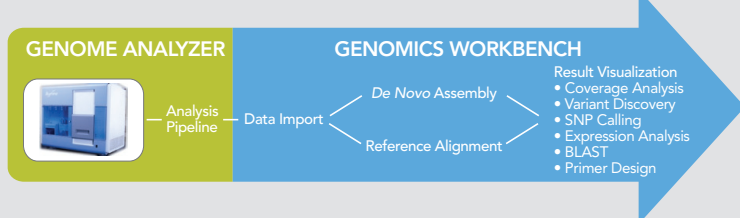
ing experiments of any read length are parsed by Genomics Workbench. If multiplexed sequencing is used, separate samples are automatically disambiguated before analysis.

## Reference-Assisted or *De Novo* Assembly

Genomics Workbench supports two different workflows to assemble genomic sequencing data: *de novo* or reference-assisted. Single or paired-end reads (or a combination of both) can be imported along with a reference sequence using the Reference Assembly Wizard. Assembled sequences can be viewed alongside annotations in the genome browser and subjected to further analyses or summarized in statistics reports to inform follow-up experiments or sequencing runs. In-depth analysis of sample sequences is facilitated by a wide range of informative display options. CDS, genes, non-coding RNA, and other annotation features from the reference can be visualized on the genome browser, along with individual reads and identified variants from the sample.

When non-standard species are being analyzed, or when experiments demand assemblies be independent from reference sequences, the algorithm behind the *De Novo* Assembly Wizard builds contigs and assembles the genome from only sequence reads. Regions with low coverage and

FIGURE 1: GENOMICS WORKBENCH PROVIDES COMPREHENSIVE TOOLS FOR ANALYZING ILLUMINA SEQUENCING DATA



Scientists using Illumina sequencing are supported by the wide range of data analysis tools in Genomics Workbench. Reads are assembled with or without a reference, and downstream analysis options generate results visualized on a genome browser and in tables.

contig conflicts are automatically identified. The number and position of contigs are displayed in tables or on the genome browser for analysis or design of the next experiment.

Genomics Workbench software is optimized for using 75–100+ bp reads generated by the Genome Analyzer. The assembly algorithm also integrates Illumina's unique combination of long-insert mate pair and standard paired-end sequencing. The Genome Analyzer's unparalleled read diversity delivers the highest coverage and largest mappable genome, even across complex rearrangements or repeats.

### Variant Detection

After assembly, a SNP detection algorithm can be run on the sequence with the SNP Detection Wizard. The annotated genome browser can be used to visually scan the sample to identify polymorphic regions of interest. Individual SNPs are listed on a table (Figure 2), which can be exported to tools for association or targeted genotyping experiments.

TABLE 1: PERFORMANCE METRICS

#### SYSTEM REQUIREMENTS

Intel or AMD CPU

Assembly and Analysis of genomes smaller than 10 Mb:

- 2 GB RAM required, 4 GB RAM recommended

Assembly and analysis of genomes larger than 10 Mb:

- 2 GB RAM required, 8 GB RAM recommended

64 bit computer and OS recommended when using more than 2 GB RAM

#### RESULTS

De Novo ( <i>E. coli</i> )	Reference Assisted (human)
22,170 N50	86 million reads, 2x35 bp paired-end
433 contigs	3 Gb, ~1x coverage
99% genomic coverage	8 hours on a standard desktop computer

### SUMMARY

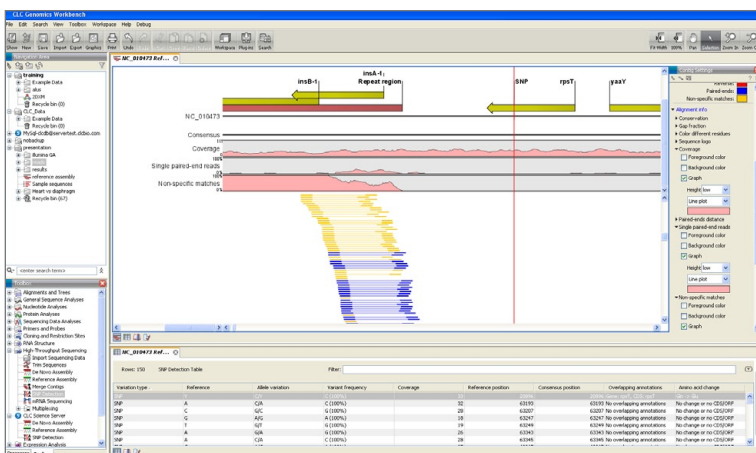
Genomics Workbench from CLC bio is a highly tailored analysis application for the Illumina Genome Analyzer. This software package provides comprehensive and easy-to-use methods for any scientist to analyze massive sequencing output. Genome Analyzer sequence reads can be used for *de novo* or reference-assisted assembly and a variety of downstream analyses. This collection of high-

performance tools is fast and can run on typical lab computers. Genomics Workbench facilitates the biological interpretation of vast amounts of sequence data made available by the Illumina Genome Analyzer.

### GENOMICS WORKBENCH

For 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.).

FIGURE 2: GENOME BROWSER VIEW OF ALIGNED SEQUENCE



A view of genome-wide results of aligned sequence reads is readily accessible. Samples and analysis tools are selected in the left panes, and results are shown on a genome browser and tables (middle). Browser display options are selected using the control panel on the right.

### GENOME ANALYZER INFORMATION

For more information about Illumina's Genome Analyzer, visit [www.illumina.com](http://www.illumina.com) or contact us at the address below.

Illumina, Inc.

Customer Solutions

9885 Towne Centre Drive  
San Diego, CA 92121-1975

1.800.809.4566 (toll free)

1.858.202.4566 (outside North America)

[techsupport@illumina.com](mailto:techsupport@illumina.com)

[www.illumina.com](http://www.illumina.com)

### FOR RESEARCH USE ONLY

© 2009 Illumina, Inc. All rights reserved.

Illumina, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPRO, and GenomeStudio are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.  
Pub. No. 970-2008-023 Current as of 16 February 2009

