

Next Gen Sequencing and Analysis Services



NCGR offers Illumina CS-Pro™ certified **Next Generation Sequencing** and **Analysis Services** providing clonal, shotgun sequences and automated alignment, variant detection, expression analysis based on read count, and visualization. The National Center for Genome Resources (NCGR, www.ncgr.org) has a longstanding reputation for developing effective bioinformatics tools for scientific research evident by **two recent awards: 2009 Bio-IT World Best Practices in Basic Research: Schizophrenia studies** and **Analysis Software Alpheus awarded 2009 Laureate by Computerworld.**

• Sequencing Services



- **Illumina Certified Service Provider (CS-Pro™) & first in North America**
- Eight Illumina Genome Analyzer Ix instruments
- Whole Genome/Transcriptome Shotgun
- mRNA Seq, ChIP, Small RNA, DGE tagged transcriptome
- ~ 7 to 20 million reads / channel (7 channels/run)
- Singleton and mate-paired reads 1 & 2 x 36, 54, 72, and 90.
- DNA or RNA
- FASTQ file output delivered through a secure FTP site
- [Price](#) (please contact Faye Schilkey at fds@ncgr.org or 505-995-4449)

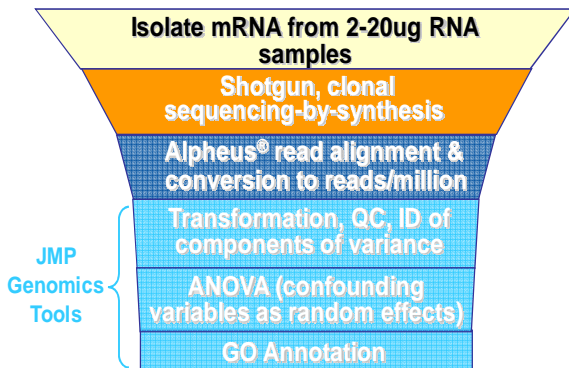
• Award Winning Data Analysis Services



THE COMPUTERWORLD HONORS PROGRAM

- **Alpheus® web-based analysis system for Next-Gen sequencing data.**
- GSNAP -based read alignment to reference genome or transcriptome
- Solexa, 454, SOLiD, Sanger reads
- Tools for identification and enumeration of nucleotide variants, indels, premature stop codons, splice isoforms, genomic rearrangements, and gene/transcript read counts for expression profiling
- Case-control sample comparison
- Secure, web-based platform for visualization and analysis
- **JMP Genomics interface for downstream statistical analysis**
- [Price](#) (please contact Faye Schilkey at fds@ncgr.org or 505-995-4449)

• Digital Transcript Expression (DTE) Analysis Pipeline Service



- 2009 Bio-IT World Best Practices winner
- Single molecule sensitivity
- Detects all transcripts – known or novel
- Detects all isoforms – known or novel
- Sequence verification for each measurement
- Any species
- Absolute measurement of transcript abundance
- Precise variation between runs
- Extensible to concomitant detection of nucleotide and structural variation

Sequencing Services – Illumina Genome Analyzer Ix



X 8

NCGR boasts eight Illumina Genome Analyzer II Instruments based on Solexa Technology

The [Illumina Genome Analyzer II](#) is a next-generation sequencing platform based on massively parallel, shotgun, clonal sequencing-by-synthesis (SBS). DNA samples are fragmented and millions of resultant fragments are sequenced using Solexa's proprietary Clonal Single Molecule Array technology with novel reversible terminator-based sequencing chemistry. Solexa sequences are relatively short (36 to 130+ nucleotides), but robust and accurate. Among next-generation sequencing technologies, this system is currently the most cost-effective.

The Solexa method employs attachment of randomly fragmented DNA to a planar, optically transparent surface and solid phase amplification to create an ultra-high density sequencing flow cell with > 150 million clusters, each containing ~1,000 copies of template per sq. cm. Clusters are sequenced using four-color DNA sequencing-by-synthesis with reversible terminators and removable fluorescence. This approach is accurate and avoids artifacts at homopolymer repeats. Fluorescence is detected following laser excitation and total internal reflection optics. Alternative sample preparation methods allow the Illumina/Solexa system to be used for other genetic analysis applications, including gene expression and small RNA discovery.

NCGR
National Center for Genome Resources

Sequencing Center

- **Eight (8) Illumina GA II's**
 - Over 550 runs / Over 50 customers
 - 1st Certified Service Provider in North America
 - On-site full-time Illumina Field Service Engineer
- **IT Bottleneck solved w/millions \$ worth of compute Infrastructure**
 - IT and Bioinformatics Leadership/Expertise
- **LIMS delimited process - Grindstone**
- **Alpheus[®] variant and expression detection pipeline**
 - Over 50 instances/customers
 - Web-based for worldwide analysis
 - Alignment to any species + transcriptome/genome alignments
 - JMP-Genomics interface for statistics
 - GSNAP alignments for long reads SE/PE
- **Fee for service and internal-research operation**

illumina CSPro
CERTIFIED SERVICE PROVIDER

ALPHEUS
Sequence Variant Detection Pipeline

jmp
THE COMPUTERIZED HONOR PROGRAM

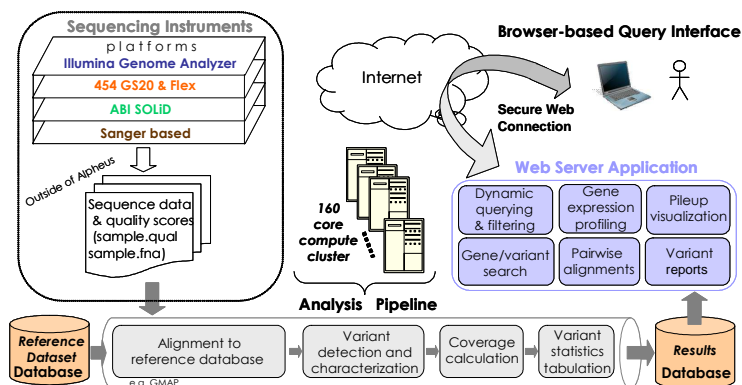
Bio IT World Best Practices AWARD 2009

Data Analysis Services - Alpheus®

We provide cutting-edge analysis, secure data management, an automated alignment pipeline, and an easy to use web interface <http://alpheus.ncgr.org>

KEY ANALYSIS FEATURES

- **GMAP**/GSNAP read alignment to reference genome or transcriptome
- Solexa, 454, SOLiD, Sanger reads
- Tools for identification and enumeration of nucleotide variants, splice isoforms and genomic rearrangements
- Case-control sample comparisons
 - Identification and enumeration of nucleotide variants, splice isoforms, structural variants
 - Search gene/transcript expression
- Reports sSNPs, nsSNPs, indels, premature stop codons, and splice isoforms. Read coverage statistics are reported by gene or transcript together with a visualization module based upon an individual transcript or genomic segment
- Result export capability in excel and SAS JMP-Genomics formats



DATA MANAGEMENT Alpheus® includes automated database ETL (extract, transfer, and load) tools to import metadata, raw data, and FASTA reads/quality automatically from the sequencing service.

ANALYSIS PIPELINE The Alpheus® pipeline is agnostic as to read type or the sequence library and provides alignment of reads onto reference and sequence variant detection, sequence variant characterization, and statistics reporting. Pipeline characteristics include:

- Pair-wise alignments use BioJava MegaBLAST and Java GMAP parsers
- Alignments to reference databases
- Variant detection (SNPs and indels)

USER INTERFACE AND VISUALIZATION The Alpheus® user interface permits the researcher to:

- Search genes by sequence read criteria
- View list of candidate genes defined for a project and sequencing statistics for each sample
- Visualize large data sets in an intuitive manner

NCGR CYBERINFRASTRUCTURE

The screenshot shows the Alpheus web interface for transcript analysis. The top section displays 'transcript: NM_005514.5' and 'case: Solexa-41-2'. Below this is a 'variant filter' section with various options. The main area is divided into three parts: 'overview' (a genomic map), 'closeup' (a detailed view of a region from 266bp to 1266bp), and 'Pileup Visualization' (a visualization of reads at a specific position). On the right, there is a 'Pairwise Alignment Detail' section showing sequence quality scores and a pairwise alignment between a read and a reference sequence. At the bottom, there is a 'Variant Analysis Results' section with a 'Single Nucleotide Variant Analysis Results' table.

Variant ID	Variant	NCBI Variants	Strands	Context	Characterization	Case: Reads with variant/total reads
261295	t142c		Both	CDS	S->A BLOSUM62: 1 (conservative)	NA01489-GS20: 0/28 NA01489-GS20: 0/17 NA01490-GS20: 0/22 reads alignments Solexa-41: 0/0
261325	g412a	rs16856510	Both	CDS	D->N BLOSUM62: 1 (conservative)	NA01489-GS20: 1/1 reads alignments NA01489-GS20: 2/2 reads alignments NA01490-GS20: 6/6 reads alignments Solexa-41: 19/66 reads alignments

Alpheus® is provided as a software service, meaning that the application is hosted and maintained by NCGR.