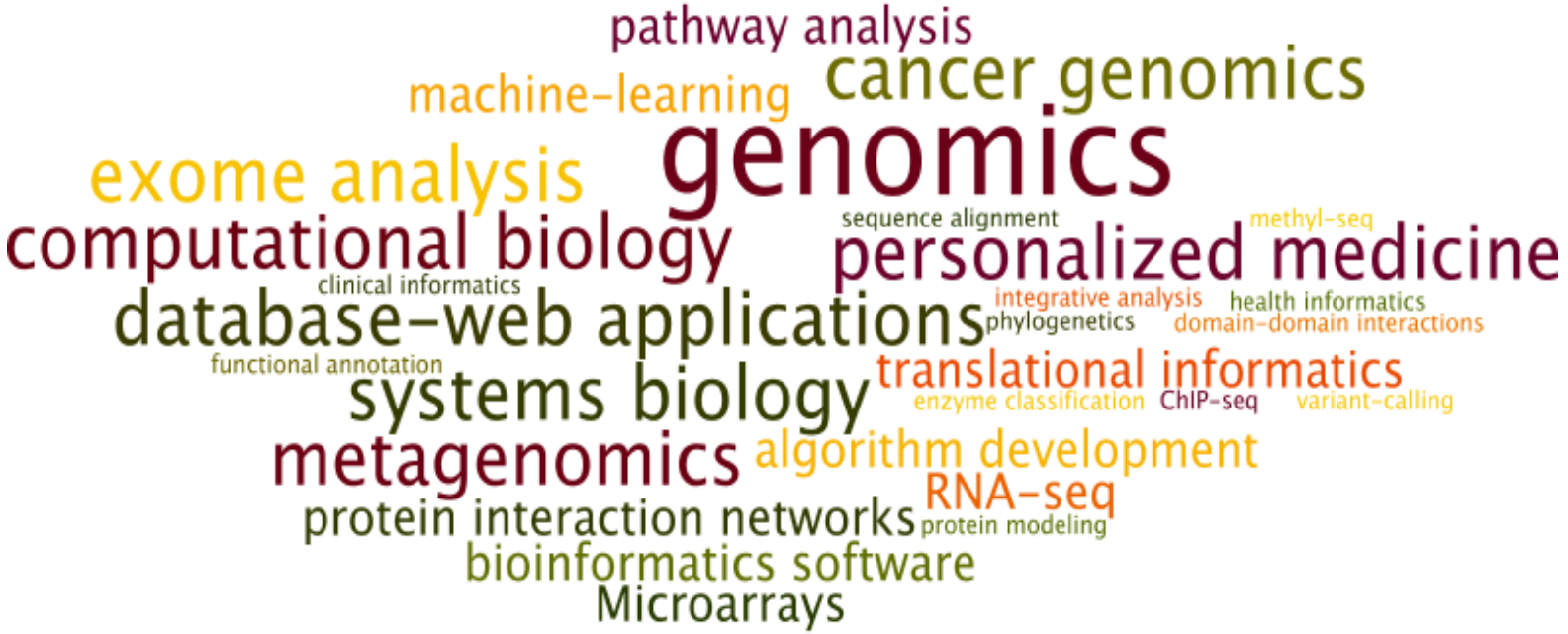
Accessing Biomedical Informatics Resources for Clinical and Translational Research

Babu Guda, Ph.D.

babu.guda@unmc.edu

Professor, Genetics, Cell Biology & Anatomy
Chief Bioinformatics & Research Computing Officer
Director, Bioinformatics and Systems Biology Core
University of Nebraska Medical Center

Resources for Biomedical Informatics Research



Bioinformatics and Systems Biology Core

- Bioinformatics Infrastructure (computing servers, databases, software tools, INBRE-supported computing and storage servers)
- Services, data analysis and application development
  - Support an array of data analysis software tools
  - Develop custom solutions to bioinformatics analysis, develop software and web-based tools
- Educational and training activities
  - Multi-omics Journal club
  - Summer workshop on bioinformatics
  - Bioinformatics courses
  - CTR seminar series

Bioinformatics Services offered

- Next-Gen Sequencing data analysis
  - Whole genome, whole exome, single-cell exome sequencing data analysis (SNVs, indels, copy numbers, splice junctions, etc.)
  - RNA-seq: Gene expression, miRNAs, fusion genes, isoform expressions, etc)
  - ChIP: Transcriptional factor binding analysis
  - Methyl-Seq: Identificaiton of the methylation sites (CpG sites and CpG islands) in the genome
  - Microbiome and Metagenomics data analysis
- Array-based and Systems Biology data analysis
  - Microarray and proteomics data analysis
  - Downstream analysis of gene expression: pathways, interaction networks, etc.
- Pre-grant and post-award bioinformatics support
  - Provide consultation with experimental design, preliminary data generation
  - Provide letter of support for grant proposals
  - Data analysis services for basic, clinical and translational research projects
- Database and web-based bioinformatics applications
- Machine-learning and other custom bioinformatics support

List of Bioinformatics programs (Public)

- Affymetrix Annotation Converter
- BLAST
- BLAT
- BRB-Array Tools
- BioPerl
- Bioconductor
- Bowtie
- Clustal2
- Ensembl
- Erlang
- FASTX-Toolkit
- Git
- Glimmer
- HMMER
- I-TASSER
- In-Silico PCR
- MATLAB
- MEME Suite
- MaxQuant
- Mfold
- Microarray Analysis in R
- Muscle
- PHYLIP
- PERL Modules
- R
- RiboSW
- SQLite
- Samtools
- Weka

Locally Installed Bioinformatics Databases

Databases	Description
BioLiP	Biologically-relevant Ligand-Protein interactions
BRENDA	A comprehensive enzyme database
dbSNP	Single Nucleotide Polymorphism Database
GenBank	Multiple BLAST databases and tools from GenBank
GO	Gene Ontology database for functional annotation
IntAct	Protein-protein interaction database
InterPro	Provides functional analysis of protein domains
KEGG	A database of curated pathways, enzymes, substrates
miRBase	Database of published microRNA sequences
OMIM	Compendium of human genes and genetic phenotypes
PATRIC	Pathosystems Resource Integration Center
PDB	Protein data bank for 3-D structure information
Pfam/Rfam	Protein and RNA family/domain database
PharmGKB	Pharmacogenomics knowledge resource
RepBase	Database of repetitive elements from eukaryotes
STRING	Database of known and predicted protein interactions
UniProt/ SwissProt	Extensively curated annotations of gene products

Locally Installed Bioinformatics Software Tools

Tools	Description
BLAT	Performs rapid mRNA/DNA or protein alignments
ChAMP	An R package for Chip Analysis Methylation Pipeline for Illumina HumanMethylation450 arrays
CLC Genomics*	A software suite with a number of molecular biology software tools including NGS data analysis
COHCAP	COHCAP (City of Hope CpG Island Analysis Pipeline)
CUPSAT I-Mutant	These tools predict changes in protein stability upon point mutation
Cytoscape	An open source platform for complex network analysis
ECemble[$]	A tool for predicting enzymes and enzymatic reactions
HMMER	A software suite to search the Pfam/Rfam databases
IPA*	Ingenuity pathway analysis tool and Knowledgebase
I-TASSER	Tool for large-scale protein structure prediction, refinement, and structure-based function annotations
LaserGene*	Comprehensive DNA/protein sequence analysis tool
MAUVE	Multiple genome alignment tools for gapped genomes
NCBI-Toolkit	BLAST programs and utilities for sequence searching
ngLOC[$]	Protein subcellular localization prediction tool
RaptorX	A protein structure prediction server
RnBeads	R package for analysis of DNA methylation data
Vector NTI*	A suite of sequence analysis and design tools

*Licensed software, [$]In-house software developed in the Guda lab.
URL – Online tool.

Locally Installed Next-Gen Sequence Analysis Tools

NGS Tools	Description
ABySS	De novo, parallel, paired-end sequence assembler
ALLPATHS-LG	Whole-genome shotgun assembler
ANNOVAR	Tool for functional annotation of genetic variants
BFAST	Blat-like Fast Accurate Search Tool (BFAST)
BWA	Maps sequence reads to a large reference genomes
FASTQC	A quality control tool for NGS data
FASTX-Toolkit	Command line tools for FASTA/FASTQ file preprocessing.
GATK	The Genome Analysis Toolkit for NGS data analysis
Maq	Mapping and Assembly with Qualities (Maq)
MEME Suite	A suite of Motif-based sequence analysis tools
MetaID	A tool for taxonomic profiling of metagenomics data
MuTect	Identifies somatic point mutations in NGS data
Oases	*De novo* transcriptome assembler using NGS data
Picard	Picard is a JAVA-based tool to manipulate SAM files
QUAKE	Corrects substitution sequencing errors in NGS reads
RNA-STAR	Aligns RNA-seq reads to a reference genome
SAMtools	Utilities for manipulating alignments including sorting, merging, indexing and formatting alignments
SOAP	A set of tools for assembly and analysis of NGS reads
Trinity	Trinity is a de novo transcriptome assembler
Tuxedo Suite	Consists of Bowtie, Tophat, and Cufflinks, used in the RNASeq analysis pipeline
Velvet	Velvet is a de novo genomic assembler for NGS data

INBRE-supported Journals at the McGoogan Library, UNMC

Bioinformatics.
Cancer cell.
Cell.
Cell Stem Cell
Current biology.
Development.
Developmental cell
The EMBO journal.
Genes & Development
Genome research.
Human molecular genetics
Immunity.
Journal of medical genetics.
Mechanisms of development
Modern pathology
Molecular and cellular biology.
Molecular cell.
Nature.
Nature biotechnology.
Nature cell biology.
Nature genetics.
Nature immunology.
Nature materials
Nature medicine.
Nature methods
Nature nanotechnology
Nature neuroscience.
Nature protocols
Nature structural and molecular biology
Neuron.
Science translational medicine
Structure.

Licensed Software for INBRE Network Members

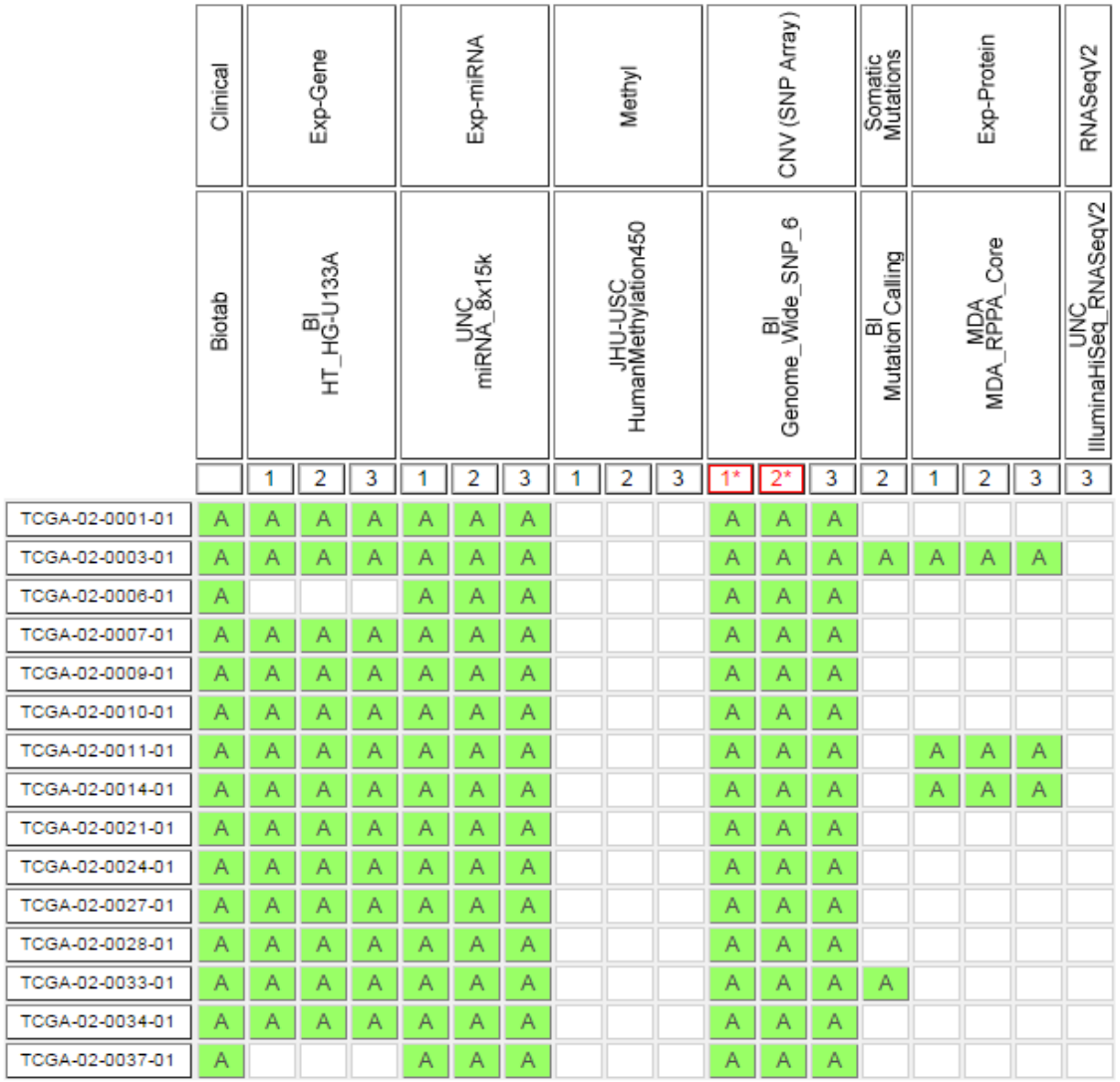
- Vector NTI
  - Comprehensive DNA/RNA/protein sequence analysis suite
- CLC Genomics Workbench
  - Collection of tools for NGS-based data analysis that include experimental data from genomics, transcriptomics, epigenomics, etc.
- Ingenuity Pathway Analysis (available only at UNMC)
  - Contains a large literature curated knowledgebase and tools for pathway and network analysis.

Summer Workshop on Bioinformatics

- Workshop taught by Kiran Bastola and Mark Pauley at UNO
- Workshop Format
  - Dates: July 2017
  - Four consecutive Fridays from 9am to Noon
  - Taught at Peter Keweit Institute (PKI)
  - Four modules, one on each day
  - Topics covered:
    - Gquery Entrez
    - Biological database search
    - Vector NTI
    - Vector NTI/Ingenuity Pathway Analysis

Resources for Basic, Clinical and Translational Research Projects

Heterogeneous data from TCGA (Glioblastoma)



	Clinical	Exp-Gene			Exp-miRNA			Methyl			CNV (SNP Array)			Somatic Mutations	Exp-Protein			RNASeqV2
	Biotab	BI HT_HG-U133A			UNC miRNA_8x15k			JHU-USC HumanMethylation450			BI Genome_Wide_SNP_6			BI Mutation Calling	MDA MDA_RPPA_Core			UNC IlluminaHiSeq_RNASeqV2
		1	2	3	1	2	3	1	2	3	1*	2*	3	2	1	2	3	3
TCGA-02-0001-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0003-01	A	A	A	A	A	A	A				A	A	A	A	A	A	A	
TCGA-02-0006-01	A				A	A	A				A	A	A					
TCGA-02-0007-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0009-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0010-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0011-01	A	A	A	A	A	A	A				A	A	A		A	A	A	
TCGA-02-0014-01	A	A	A	A	A	A	A				A	A	A		A	A	A	
TCGA-02-0021-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0024-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0027-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0028-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0033-01	A	A	A	A	A	A	A				A	A	A	A				
TCGA-02-0034-01	A	A	A	A	A	A	A				A	A	A					
TCGA-02-0037-01	A				A	A	A				A	A	A					

Clinical: all 499

Exp-Gene: 440

Exp-miRNA: 474

Methyl: 130

SNP Array: 493

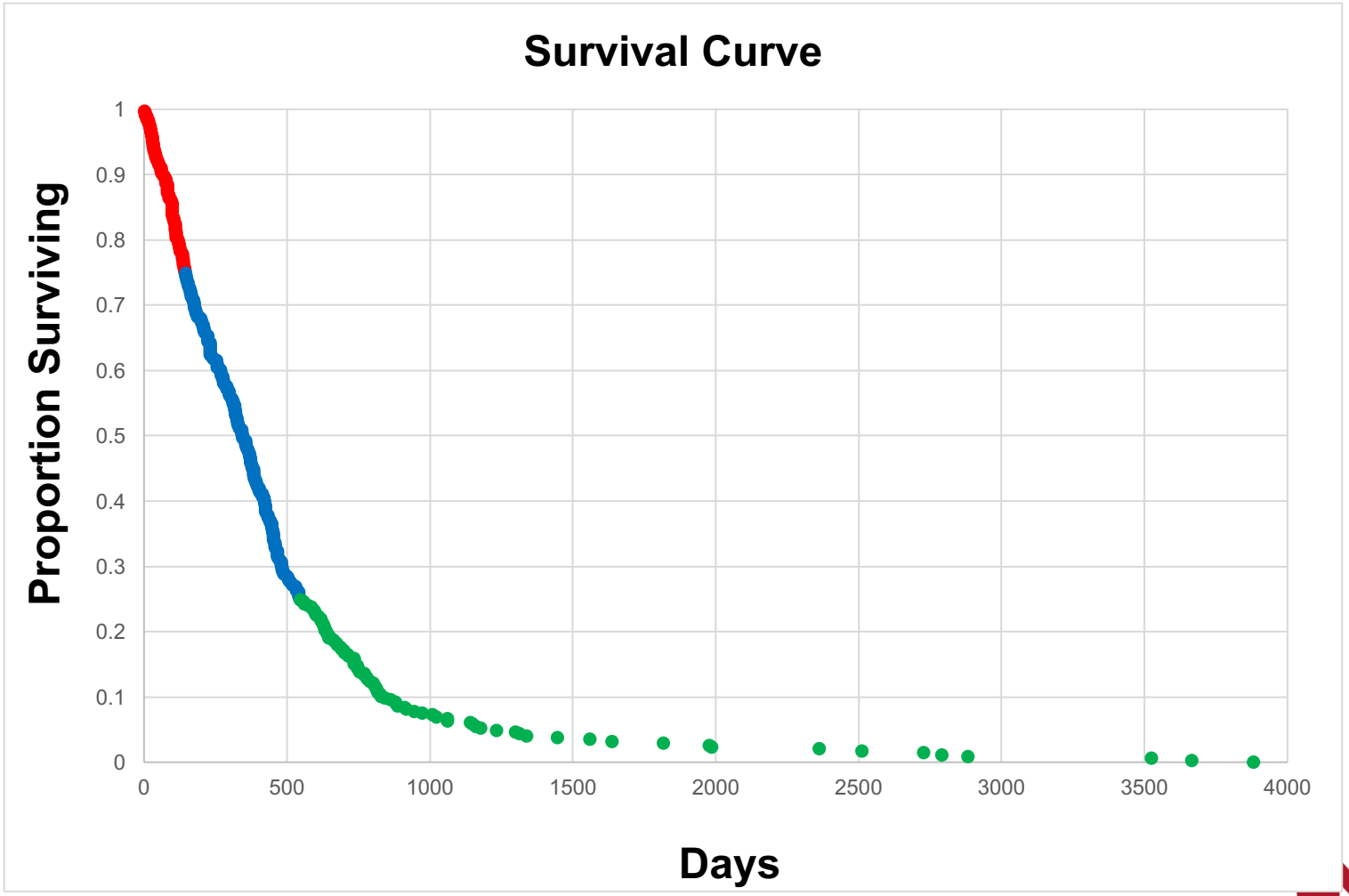
Somatic Mutations: 278

Exp-Protein: 210

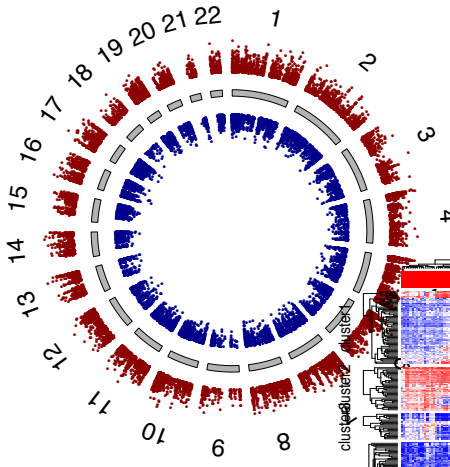
RNASeqV2: 159

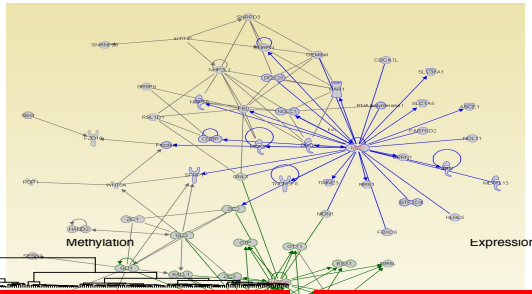
Data Type	Platform	# Patients	Format	Analysis Method
Gene Expression	Affymetrix HT Human Genome U133	440	CEL files	Bioconductor
miRNA Expression	Agilent Human miRNA Microarray 8x15k	474	BAM files	GATK
Methylation	Illumina Human Methylation 450	130	idat files	RnBeads
SNP Array	Affymetrix Genome-Wide Human SNP Array 6.0	493	CEL files	Bioconductor
Somatic Mutations	Agilent exome capture, Illumina HiSeq, Picard/Firehouse alignment, QC, variant calling	278	vcf files	Various filtering steps (filterVCF in bioconductor, cosmic, etc.)
Protein Expression	MDA Reverse Phase Protein Array	210	txt file (relative protein expression)	Statistics
RNASeqV2	Illumina HiSeq 2000	159	tab-delimited text (raw expression signal)	Statistics

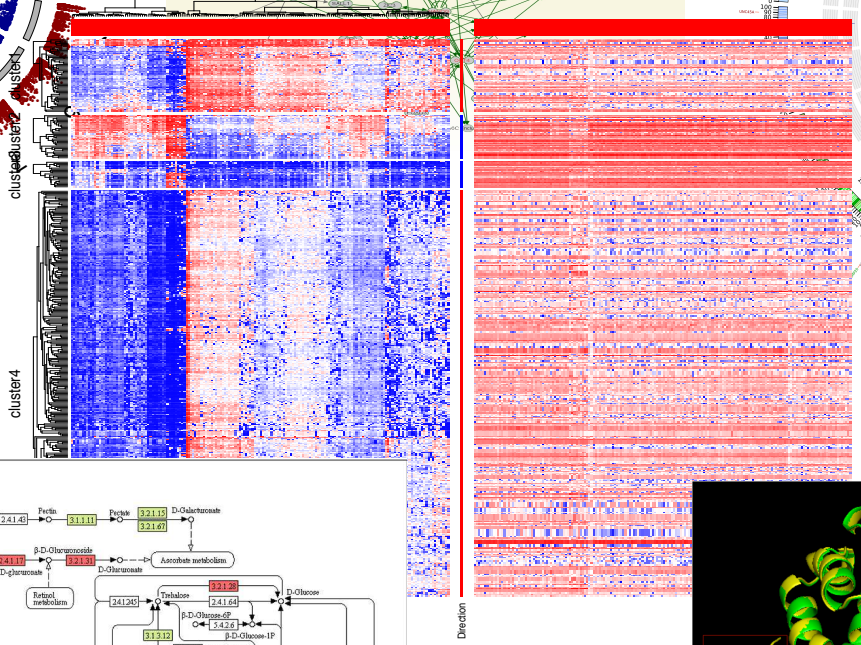
Survival analysis of TCGA Glioblastoma patients

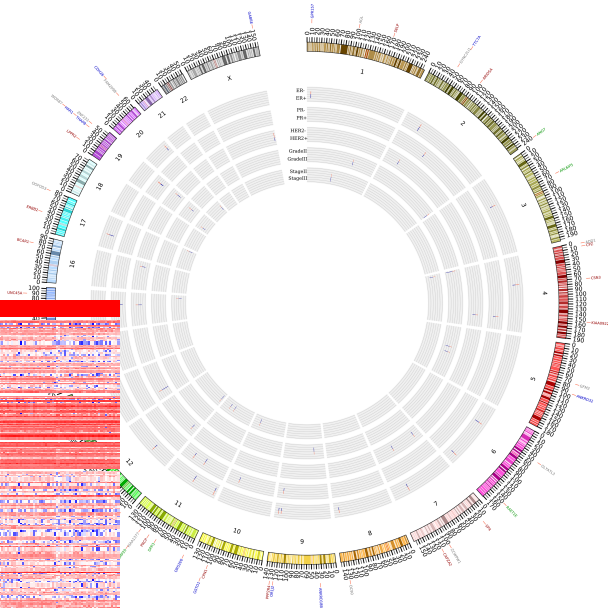


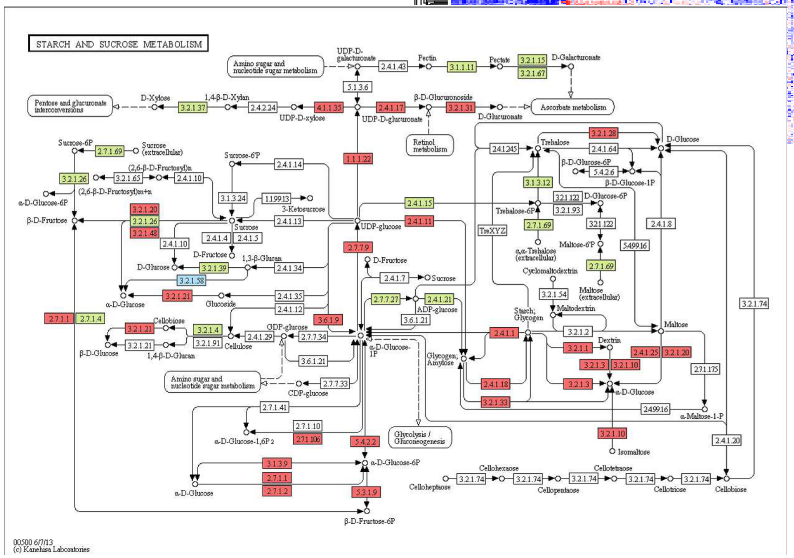
A collage of illustrations from our recent publications

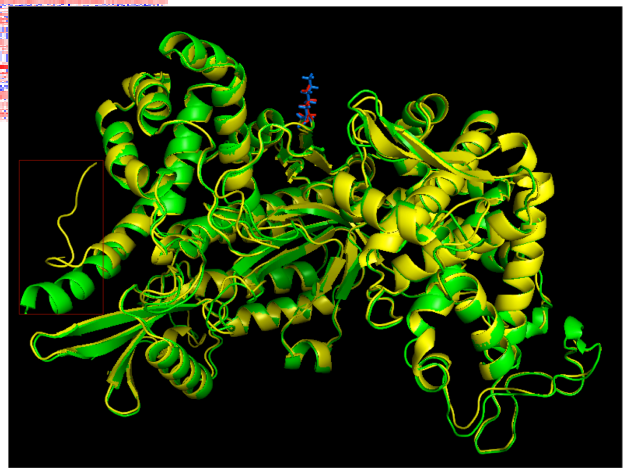




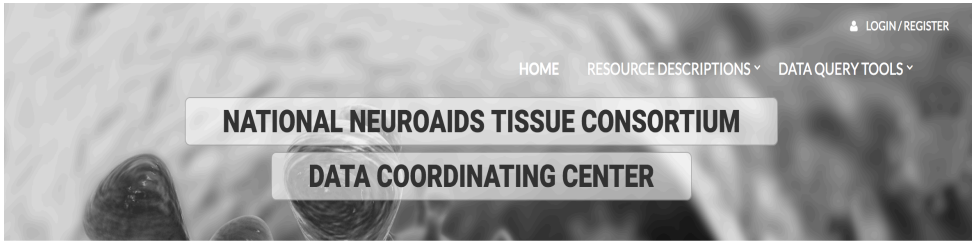








Database and Web Applications



DESCRIPTION OF RESOURCES

NNTC

The NNTC was established in 1998 to facilitate access to antemortem and postmortem tissues `and fluids (blood, cerebrospinal fluid) for the international neuroAIDS research community. The consortium's goals included establishment of a network of brain banks, collection of nervous system tissues in a standardized fashion, and maximization of the information gleaned from the scientific studies of these tissues. In addition, a goal of the NNTC was to link the experimental and clinical data pertaining to each sample.

See More

CHARTER

The CNS HIV Anti-Retroviral Therapy Effects Research (CHARTER) began in 2002 to explore the changing presentation of HIV neurological complications in the context of emerging antiviral treatments such as combination antiretroviral therapy. CHARTER's study aims are to determine how central and peripheral nervous system complications of HIV are affected by different histories and regimens of antiretroviral therapy (ART). Many of the data metrics gathered are quite similar to the ones gathered by the NNTC, and biofluids (e.g. plasma and CSF, but not tissues) are similarly banked.

See More

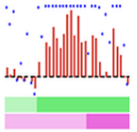
NCBI Gene Search Form

Gene #1 : A1BG
Gene #2 : BAALC
Gene #3 : C1QC
Gene #4 :
Gene #5 :

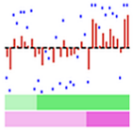
Add Gene | Remove Gene | Search | Clear

ne name for a drop-down list of suggested gene names from the HUGO database.
genes to query by pressing "Add Gene" button. Up to 5 genes can be analyzed.

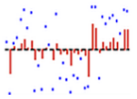
Antiretroviral therapy effect on brain of patients with HIV-associated
gnitive disorders
n: **C1QC**, complement component 1, q subcomponent, C chain
: Homo sapiens
GPL570, 225353_s_at (ID_REF), GDS4231, 714 (Gene ID)
ype: Expression profiling by array, transformed count, 35 samples
411
aSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors



- Antiretroviral therapy effect on brain of patients with HIV-associated
gnitive disorders
n: **BAALC**, brain and acute leukemia, cytoplasmic
: Homo sapiens
GPL570, 222780_s_at (ID_REF), GDS4231, 79870 (Gene ID), AI870583
ype: Expression profiling by array, transformed count, 35 samples
860
aSets Gene UniGene Profile neighbors Chromosome neighbors Homologene neighbors



- Antiretroviral therapy effect on brain of patients with HIV-associated
gnitive disorders
n: **BAALC**, brain and acute leukemia, cytoplasmic
: Homo sapiens



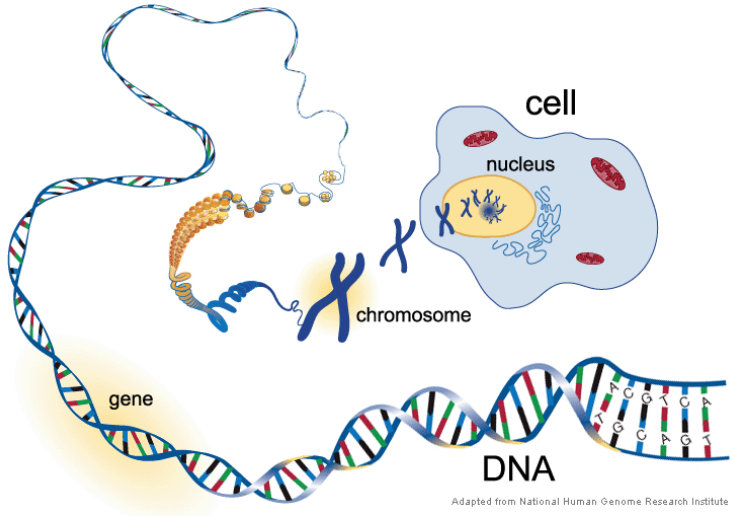
2011: Jeopardy Contest: Man versus Machine
IBM Watson Genomics Software

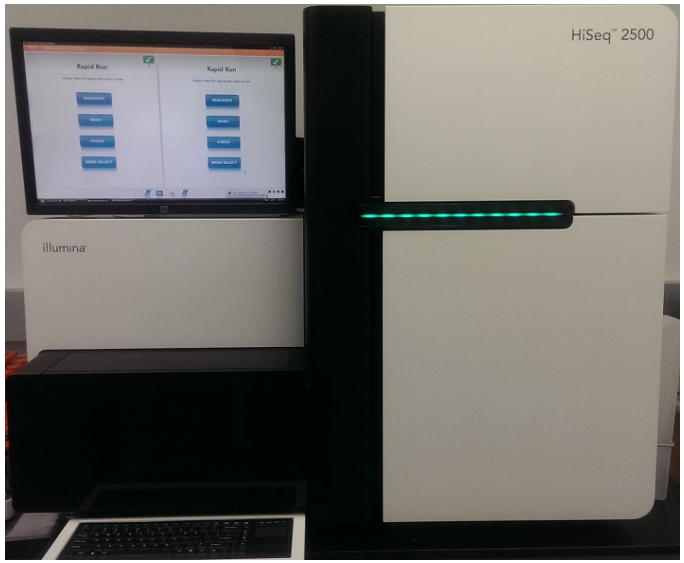
UNMC-IBM Collaboration

Precision Medicine-1000 Breast Tumor Genomes Project

- All the patients are registered through the Breast Cancer Registry Program at UNMC
- Samples include all major subtypes of breast cancer (Luminal-A, Luminal-B, Her2, Basal)
- Sequencing of normal/tumor DNA done at UNMC
- Tumor DNA is compared against normal DNA for the same patient to identify the somatic mutations
- Identified mutations, copy number variations and gene expression data are tested using the Watson Genomics Software

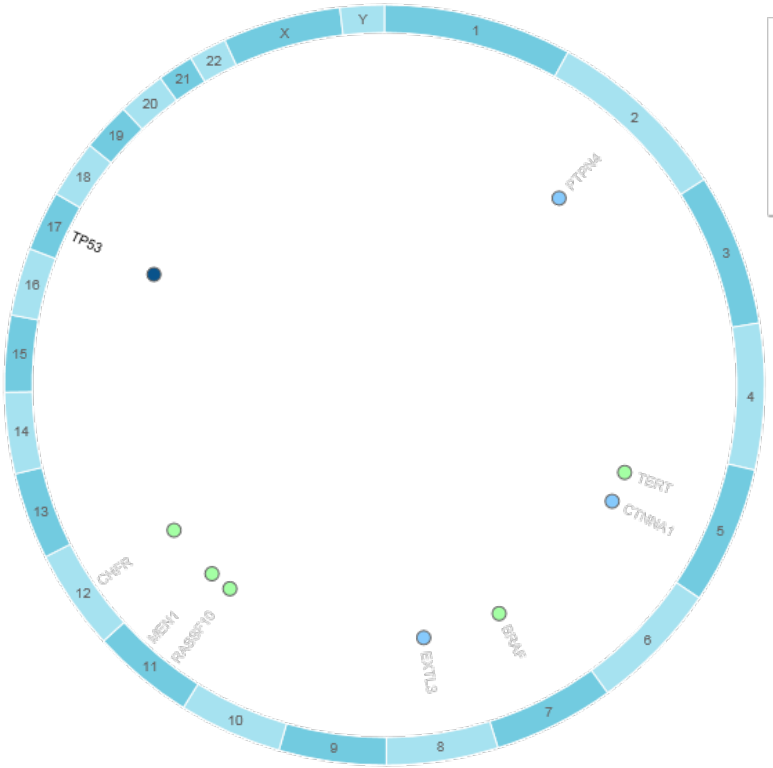
Collaboration with Ken Cowan, UNMC


cell
nucleus
chromosome
gene
DNA
Adapted from National Human Genome Research Institute

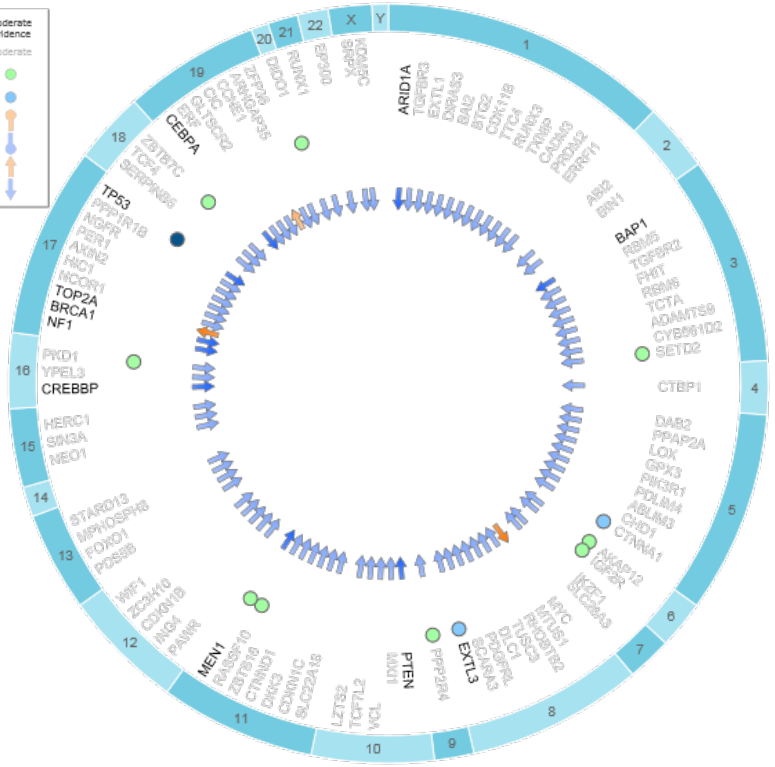

HiSeq™ 2500
illumina


HOLLAND
COMPUTING CENTER

Molecular Profile Report From Watson Genomics



Patient_1234
Only Mutation Data



Patient_1234
Mutation + Expression Data

How to contact us?

- Babu Guda, PhD
  - BMI-KCA Co-Director: babu.guda@unmc.edu
- Peng Xiao, PhD
  - Operational Director, Bioinformatics Core: peng.xiao@unmc.edu
- https://www.unmc.edu/gudalab/
- https://www.unmc.edu/bsbc/
- https://www.unmc.edu/inbre/cores/bioinformatics/index.html

Acknowledgements

Core supported members

Sanjit Pandey, M.S.
Jasjit K. Banwait. Ph.D.
Peng Xiao, Ph.D.
Yeong Kim, Ph.D.
Matyas Cserhati, Ph.D.
Duc Le, B.S.
Meng Niu, Ph.D.
Navodita Upadhyay, MS

Funding

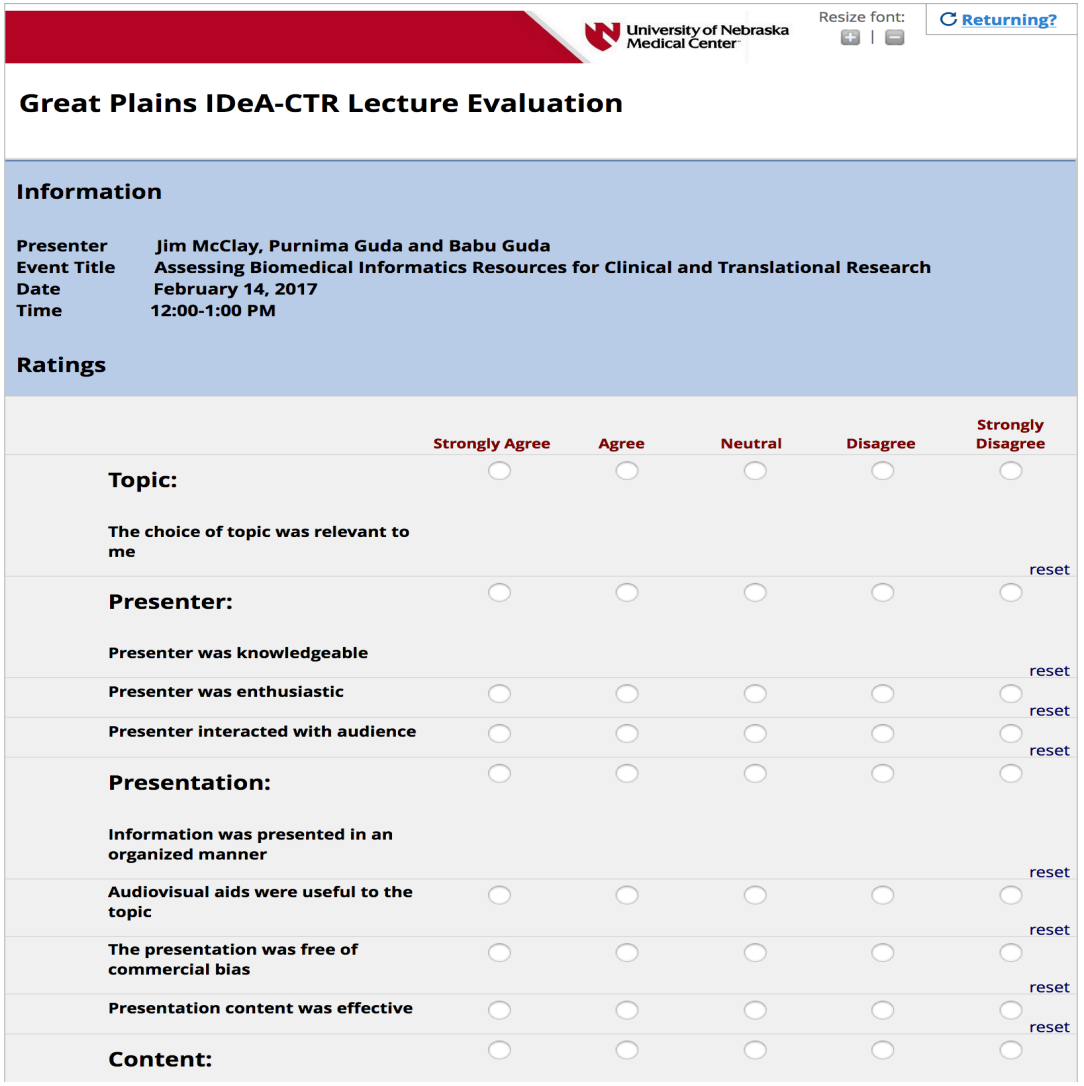
Nebraska Research Initiative (NRI)
NIH funding (INBRE, CCSG, NNTC)
IDeA-CTR

UNMC Cores

Genomics, Bioinformatics, RITO

Evaluation Form

Link: https://unmcredcap.unmc.edu/redcap/surveys/?s=J3AE8AR43T



University of Nebraska Medical Center

Resize font:

Returning?

Great Plains IDeA-CTR Lecture Evaluation

Information

Presenter	Jim McClay, Purnima Guda and Babu Guda
Event Title	Assessing Biomedical Informatics Resources for Clinical and Translational Research
Date	February 14, 2017
Time	12:00-1:00 PM

Ratings

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	
Topic: The choice of topic was relevant to me	○	○	○	○	○	reset
Presenter: Presenter was knowledgeable	○	○	○	○	○	reset
Presenter was enthusiastic	○	○	○	○	○	reset
Presenter interacted with audience	○	○	○	○	○	reset
Presentation: Information was presented in an organized manner	○	○	○	○	○	reset
Audiovisual aids were useful to the topic	○	○	○	○	○	reset
The presentation was free of commercial bias	○	○	○	○	○	reset
Presentation content was effective	○	○	○	○	○	reset
Content:	○	○	○	○	○	