



Bioinformatics up to Date

(Bioinformatics Center, Biotechnology Division)
 North-East Institute of Science & Technology
 Jorhat - 785 006, Assam

CONTENTS

| | |
|--------------------------|---|
| Cover Story | 1 |
| Bioinfy Quiz | 1 |
| Computational Chemistry | 2 |
| Proteomics | 2 |
| Bioserver | 3 |
| Computers for Biologists | 4 |
| Genomics | 4 |
| Bioinfy Animator | 5 |
| Contact Us | 5 |
| Answers to quiz | 5 |

Adviser:
 Dr.P.G.Rao

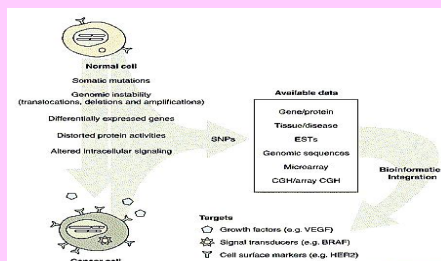
Editors:
 Dr. R.L.Bezbaruah
 Pompi Sharma
 Dhruvjyoti Gogoi

COVER STORY

Bioinformatics in cancer therapeutics—hype or hope?

Development and use of bioinformatics is essential for the future of cancer therapeutics. Most cancer treatments work for only a subset of patients and this is likely to remain true for many molecularly-targeted drugs. This results in a large proportion of patients receiving ineffective treatments and is a huge financial burden on our health care system. It is essential that we develop accurate tools for delivering the right treatment to the right patient based on biological characterization of each patient’s tumor. Gene-expression profiling of tumors using DNA microarrays is a powerful tool for pharmacogenomics targeting of treatments.

A good example is the Oncotype DX™ assay (Genomic Health) recently described for identifying the subset of node-negative estrogen-receptor-positive breast cancer patients who do not require adjuvant chemotherapy.¹ Development of genomic tests that are sufficiently validated for broad clinical application requires the sustained effort of a team that includes clinical investigators, biologic scientists and bio statisticians. With proper focus and support, gene-expression-based diagnostic tests could be developed today to assist patients and physicians with a wide range of difficult decisions regarding the use of currently existing treatments. Many tumors consist of mixtures of subclones containing different sets of mutated, overexpressed and silenced genes. This heterogeneity makes the process of identifying good molecular targets very challenging. Most overexpressed genes and mutated genes may not represent good molecular targets because resistant subclones are present. The best target is a ‘red dot’ gene whose mutation occurs early in oncogenesis and dysregulates a key pathway that drives tumor growth in all of the subclones. Examples include mutations in the genes *ABL*, *HER-2*, *KIT*, *EGFR* and probably *BRAF*, in chronic myelogenous leukemia, breast cancer, gastrointestinal stromal tumors, non-small-cell lung cancer and melanoma, respectively. Effective development of therapeutics requires identification of red-dot targets, development of drugs that inhibit the red-dot targets, and diagnostic classification of the pathways driving the growth of each patient’s tumor..



A.K. Verma and S.B. Prasad
 North Eastern Hill University
 Deptt. of Zoology

BIOINFY QUIZ

- Of the organisms that follow, what has the largest genome size?
 A) *Drosophila melanogaster*
 B) *Oryza sativa*
 C) *Helicobacter pylori*
- What step of DNA sequencing is skipped during shotgun sequencing?
 A) mapping step
 B) cloning of DNA fragment
 C) computer analysis
- The translated genes of genomes that encode proteins are referred to as
 A) introns.
 B) codons
 C) open reading frame
- The first researcher to sequence a genome, in 1977, was
 A) Venter
 B) Sanger
 C) Fodor
- Genes for typical single-character Mendelian traits are called
 A) single-copy genes
 B) segmental duplications
 C) tandem clusters

Answers on page 5

COMPUTATIONAL CHEMISTRY

SPARTAN 10

Spartan has a wide range of features including conformational searching, calculation of structure, energies, and properties, and quantifying 3-D molecular similarity.

Spartan is powerful chemistry software with an elegantly designed graphical interface that makes it easy to learn and use. It is effective for learning (and teaching) the subtleties of organic and physical chemistry. For professional chemists and researchers Spartan accelerates research in both academic and commercial environments. The new Odyssey program provides an electronic learning environment with rich interactive 3D molecular simulations and structured chemistry content ideally suited to first year undergraduate chemistry and pharmaceutical science courses.

www.computational-chemistry.co.uk

PROTEOMICS

MASCOT

Mascot is a powerful search engine which uses mass spectrometry data to identify proteins from primary sequence databases.

While a number of similar programs available, Mascot is unique in that it integrates all of the proven methods of searching. These different search methods can be categorised as follows:

Peptide Mass Fingerprint in which the only experimental data are peptide mass values

Sequence Query in which peptide mass data are combined with amino acid sequence and composition information. A super-set of a sequence tag query

MS/MS Ion Search using uninterpreted MS/MS data from one or more peptides.

www.matrixscience.com

BIOSERVER

PSA

The Protein Sequence Analysis (PSA) server predicts protein secondary and tertiary structure based on sequence, and is available for researchers who have amino acid sequences for proteins of unknown structure and for which no homologous sequences are known. To use PSA, one submits a single amino acid sequence to the server, which may be instructed to analyze the sequence in one of three ways: using **Type-1**, **Type-2**, or **WD-repeat** DSMs. DSMs are **D**iscrete **S**tate-space **M**odels for patterns of alpha-helices, strands, tight turns, and loops in specific structural classes.

The PSA System determines the probable placement of secondary structural elements along the sequence. In addition, when using Type-1 models, it also determines the probable tertiary structural class of the protein. In fact, it uses knowledge of this structural class when it computes the probabilities for secondary structural elements. The exact nature of the output depends on the type of analysis requested.

The analysis algorithm is based on probabilistic Discrete State-space Models (DSMs) and optimal filtering and smoothing algorithms as described in the paper "Structural analysis based on state-space modeling" by C.M. Stultz, J.V. White, and T.F. Smith, *Protein Science* (1993), 2:305-314.

Submit a PSA Sequence Analysis Request

[BMERC : psa-request](#) : Submit request

Please . . .

- We ask that you **refrain from submitting large numbers of requests** between the hours of 9 a.m. to 6 p.m. Eastern time, Monday thru Friday.
- There is a **limit of one request per month from [unlicensed commercial users](#)**. There is no limit for educational and government users.

| E-mail address: | <input type="text"/> | <small>Don't remember the location of your Web results? Click here to have it sent to your e-mail address.</small> | <input type="button" value="E-mail URL"/> | | | | | | | | | | | | |
|--|---|--|---|---------------|---|---------------------------------|--------------------------------|--|---------------------------------|-----------------|---------------------------------|------------------------------------|-----------------|--|--|
| Subject: | <input type="text"/> | <small>(Ideally, this should be 15 characters or less. Used to label graphs and e-mail messages.)</small> | | | | | | | | | | | | | |
| Comments: (E.g. your name, affiliation, technical comments, etc.) | <input type="text"/> | | | | | | | | | | | | | | |
| Analysis assumptions (choose one) | <table><thead><tr><th>Choice</th><th>Description</th><th>Length limits</th></tr></thead><tbody><tr><td><input type="radio"/> Monomeric-Soluble</td><td>Type-1 analysis</td><td>At least 40 and at most 350 AA</td></tr><tr><td><input checked="" type="radio"/> Minimal</td><td>Type-2 analysis</td><td>At most 1000 AA</td></tr><tr><td><input type="radio"/> WD-repeat</td><td>WD-repeat analysis</td><td>At most 1000 AA</td></tr></tbody></table> | Choice | Description | Length limits | <input type="radio"/> Monomeric-Soluble | Type-1 analysis | At least 40 and at most 350 AA | <input checked="" type="radio"/> Minimal | Type-2 analysis | At most 1000 AA | <input type="radio"/> WD-repeat | WD-repeat analysis | At most 1000 AA | | |
| Choice | Description | Length limits | | | | | | | | | | | | | |
| <input type="radio"/> Monomeric-Soluble | Type-1 analysis | At least 40 and at most 350 AA | | | | | | | | | | | | | |
| <input checked="" type="radio"/> Minimal | Type-2 analysis | At most 1000 AA | | | | | | | | | | | | | |
| <input type="radio"/> WD-repeat | WD-repeat analysis | At most 1000 AA | | | | | | | | | | | | | |
| Sequence: | <input type="text"/> | | | | | | | | | | | | | | |
| Plot format: | PostScript <input type="button" value="v"/> | | | | | | | | | | | | | | |
| Return results via: | private Web <input type="button" value="v"/> | | | | | | | | | | | | | | |
| <input type="button" value="Submit Sequence"/> | | | | | | | | | | | | | | | |

www.bmerc-www.bu.edu

COMPUTERS FOR BIOLOGISTS

CLC MAIN WORKBENCH

A comprehensive workbench for advanced DNA, RNA, and protein analyses

CLC Main Workbench aggregates user friendly gene expression analysis with all features of CLC protein Workbench, CLC RNA Workbench, and CLC DNA Workbench in one integrated software package.

CLC Main Workbench creates a software environment enabling users to make a large number of advanced DNA, RNA, and protein sequence analyses, combined with gene expression analysis, smooth data management, and excellent graphical viewing and output options.

CLC Main Workbench is available on Windows, Mac OS X, and Linux platforms.

www.clcbio.com

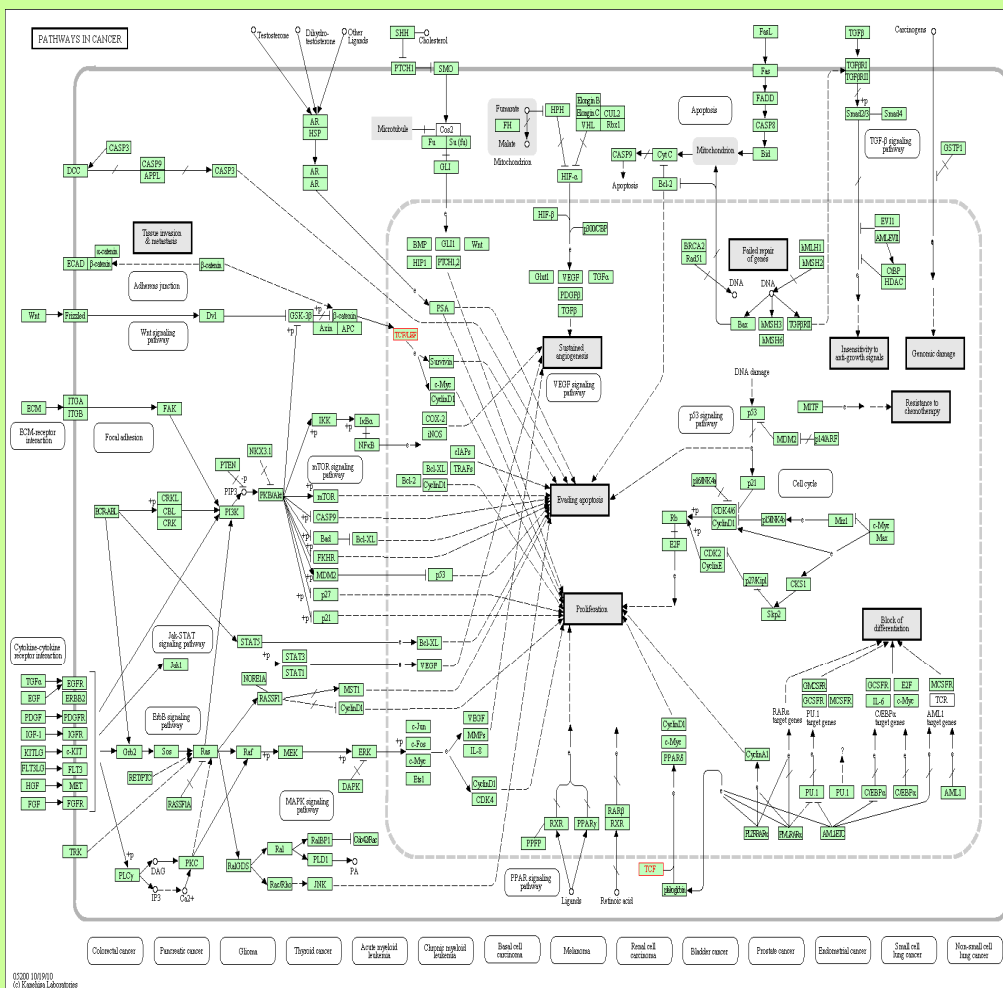
GENOMICS

UCSC Genome Bioinformatics

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC).

The screenshot shows the UCSC Genome Browser Gateway interface. At the top is a navigation menu with links: Home, Genomes, Blat, Tables, Gene Sorter, PCR, Session, FAQ, Help. Below the menu is the title "Human (*Homo sapiens*) Genome Browser Gateway". A copyright notice states: "The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz. Software Copyright (c) The Regents of the University of California. All rights reserved." The main search area contains a form with the following fields: "clade" (Mammal), "genome" (Human), "assembly" (Feb. 2009 (GRCh37/hg19)), "position or search term" (chr21:33,031,597-33,041,570), "gene" (empty), and "image width" (800). There is a "submit" button. Below the form are buttons for "track search", "add custom tracks", "configure tracks and display", and "clear position". A link says "Click here to reset the browser user interface settings to their defaults." Below the search area is a section titled "About the Human Feb. 2009 (GRCh37/hg19) assembly (sequences)". It states: "The February 2009 human reference sequence (GRCh37) was produced by the Genome Reference Consortium." Under "Sample position queries", it explains that a genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. It lists examples: "chr7" (Displays all of chromosome 7), "chrUn_gl000212" (Displays all of the unplaced contig gl000212), "chr3:1-1000000" (Displays first million bases of chr 3, counting from p-arm telomere), and "chr3:1000000+2000" (Displays a region of chr3 that spans 2000 bases, starting with position 1000000). To the right of this text is a graphic of a human figure with a DNA double helix and various genomic tracks overlaid. The letters "U C S C" are displayed at the bottom of the graphic. The text "Homo sapiens" is visible at the bottom right of the graphic.

PATHWAY IN CANCER



www.genome.jp

Answers of Bioinfy Quiz

1) B 2)A 3)C 4)B 5)A

CONTACT US:

Pompi Sharma, Project Assistant (Level – II)

E-mail: pompi.sharma86@gmail.com

Dhrubajyoti Gogoi, Project Assistant (Level – II)

E-mail: dhruba.bio.du@gmail.com