Volume 7, Issue 11 November 2014

Bioinformatics up to Date

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

Inside…..

BIOINFORMATICS INDUSTRIAL TRAINING PROGRAMME (BIITP) 2014-15

http://www.bcil.nic.in/biitp2014-15/index.asp

Cat Genome Project

The cat genome sequencing project, funded by the National Human Genome Research Institute, part of the National Institutes of Health (NIH), began in 2007. The project's initial goal was to study hereditary diseases in domestic cats, which are similar in some cases to those that afflict humans, including neurological disorders, and infectious and metabolic diseases.

To obtain the high-quality reference genome needed for this research, the team sequenced a domestic female Abyssinian cat named Cinnamon. They chose this particular cat because they could trace its lineage back several generations. This cat's family also had a particular degenerative eye disorder the researchers wanted to study.

To better understand characteristics of domestication, the researchers sequenced the genomes of select purebred domestic cats. Hallmarks of their domestication include features such as hair color, texture and patterns, as well as facial structure and how docile a cat is. Cats are bred for many of these types of characteristics. In fact, most modern breeds are the result of humans breeding cats for their favorite hair patterns.

The Birman breed has characteristic white paws. Comparing the Birman to other breeds' genomes reveals that humans likely bred cats for this quality.

The team also looked at a breed called Birman, which has characteristic white paws. The researchers traced the white pattern to just two small changes in a gene associated with hair color. They found that this genetic signature appears in all Birmans, likely showing that humans selectively bred these cats for their white paws and that the change to their genome happened in a remarkably short period of time.

The group also compared the cat genome with those of other mammals -- including a tiger, cow, dog and human -- to understand more about the genetics of cat biology.

"We looked at the underlying genetics to understand why certain abilities to survive in the wild evolved in cats and other carnivores," said Michael Montague, PhD, the study's first author and a postdoctoral research associate at The Genome Institute.

The differences they found in the cat genome help explain characteristics such as why cats are almost exclusively carnivorous and how their vision and sense of smell differ from other animals like dogs.

[http://www.biologynews.net/archives/2014/11/10/the_cats_meow_genome_reveals_clues_to_domestication.html]

GeneIndex: iPhone/iPad

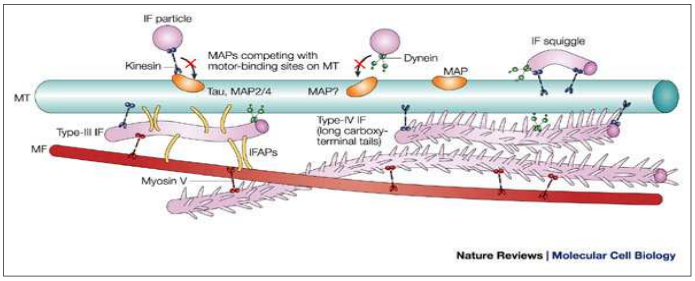
GeneIndex provides a convenient and portable way to lookup gene symbols while at a seminar, conference, or lab meeting. Genes are linked to common life science websites such as NCBI, COSMIC, KEGG, PubMed, SymAtlas, UCSC genome browser, Pathway Commons, Genatlas, Wikipedia, HUGO, and OMIM. GeneRIF gene interactions can also be queried.

Getting a read on structural variation in cancer

Tumor genomes can have high structural variation caused by, for example, deletions, insertions and translocations. Detecting these variants in sequence data is challenging, and most tools specialize in finding certain mutation classes. Moncunill *et al.* take an unusual approach in their somatic mutation finder（SMUFIN），which compares tumor and matched normal genomic sequence reads directly and uses a tree structure to identify tumor-specific reads that are likely to bear mutations.

SMUFIN, a methodology for the identification of somatic variation in tumor genomes from their direct comparison with their corresponding normal samples. SMUFIN also provides an integrated solution for the identification, in a single run, of somatic SNVs and structural variants（insertions, deletions, inversions and translocations of any size），which can currently be partially achieved only by combining several independent programs and in-house filtering schemes into complex computational pipelines. SMUFIN can uncover single-base changes as well as structural variation of any size at base-pair resolution and does not rely on sequence alignment to a reference genome, making it fast. The researchers used SMUFIN to detect changes in blood and solid tumors, including highly complex variation indicative of chromothripsis.

Microtubule-associated protein



Microtubule-associated proteins（MAPs）are proteins that interact with the microtubules of the cellular cytoskeleton.MAPs bind to the tubulin subunits that make up microtubules to regulate their stability. A large variety of MAPs have been identified in many different cell types, and they have been found to carry out a wide range of functions. These include both stabilizing and destabilizing microtubules, guiding microtubules towards specific cellular locations, cross-linking microtubules and mediating the interactions of microtubules with other proteins in the cell.

Within the cell, MAPs bind directly to the tubulin dimers of microtubules. This binding can occur with either polymerized or depolymerized tubulin, and in most cases leads to the stabilization of microtubule structure, further encouraging polymerization. Usually, it is the C-terminal domain of the MAP that interacts with tubulin, while the N-terminal domain can bind with cellular vesicles, intermediate filaments or other microtubules. MAP-microtubule binding is regulated through MAP phosphorylation. This is accomplished through the function of the microtubule-affinity-regulating-kinase（MARK）protein. Phosphorylation of the MAP by the MARK causes the MAP to detach from any bound microtubules [2]. This detachment is usually associated with a destabilization of the microtubule causing it to fall apart. In this way the stabilization of microtubules by MAPs is regulated within the cell through phosphorylation.

GlycomeDB—a unified database for carbohydrate structures

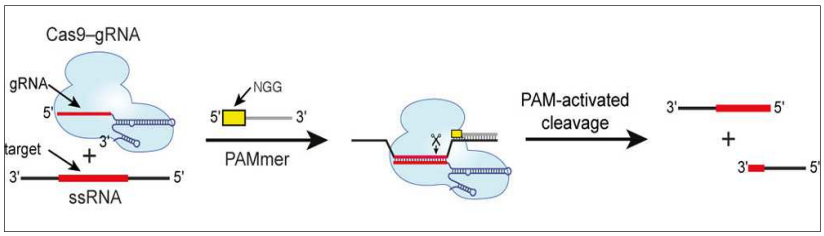
GlycomeDB integrates the structural and taxonomic data of all major public carbohydrate databases, as well as carbohydrates contained in the Protein Data Bank, which renders the database currently the most comprehensive and unified resource for carbohydrate structures worldwide. GlycomeDB retains the links to the original databases and is updated at weekly intervals with the newest structures available from the source databases. Hyperlinks to the original source of the data are established, so users can use the GlycomeDB Web-portal to access efficiently relevant additional information, which is only available in the original databases. The complete database can be downloaded freely or accessed through a Web-interface (www.glycome-db.org) that provides flexible and powerful search functionalities.

GlycomeDB contains the unified carbohydrate sequences of all publicly accessible databases that contain carbohydrates structures. In total 121 - 766 original sequences were parsed and integrated. Currently (August 2010) there are 35 - 873 unique carbohydrate sequences—with taxonomic annotations if available—stored in GlycomeDB, 11 - 822 of which are fully determined carbohydrates.

[Nucleic Acids Res. Jan 2011; 39(Database issue): D373–D376.]

RCas9: A Programmable RNA Editing Tool

A powerful scientific tool for editing the DNA instructions in a genome can now also be applied to RNA, the molecule that translates DNA's genetic instructions into the production of proteins.



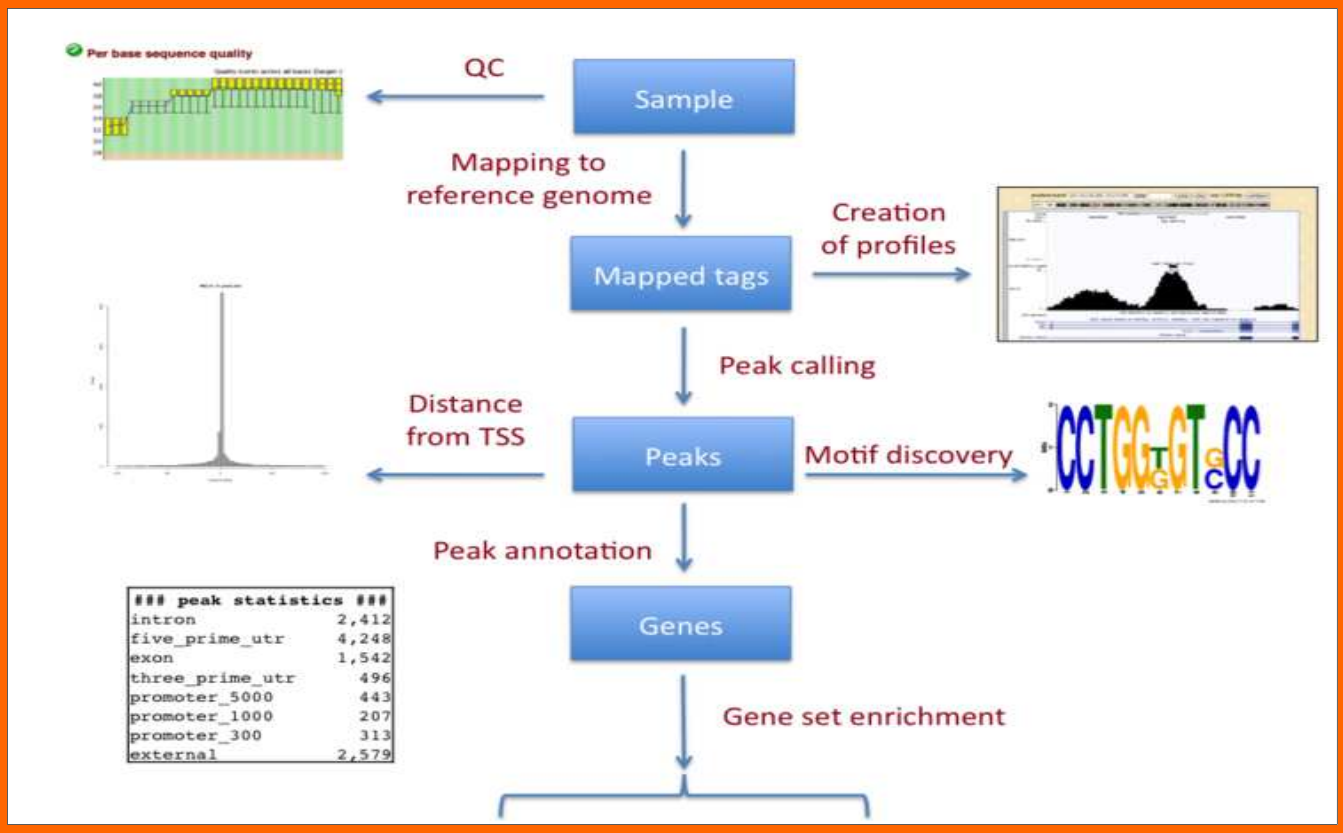
A team of researchers with Berkeley Lab and the University of California (UC) Berkeley has demonstrated a means by which the CRISPR/Cas9 protein complex can be programmed to recognize and cleave RNA at sequence-specific target sites. This finding has the potential to transform the study of RNA function by paving the way for direct RNA transcript detection, analysis and manipulation.

."Using specially designed PAM-presenting oligonucleotides, or PAMmers, RCas9 can be specifically directed to bind or cut RNA targets while avoiding corresponding DNA sequences, or it can be used to isolate specific endogenous messenger RNA from cells," says Doudna, who holds joint appointments with Berkeley Lab's Physical Biosciences Division and UC Berkeley's Department of Molecular and Cell Biology and Department of Chemistry.

In recent years, the CRISPR/Cas complex has emerged as one of the most effective tools for doing this. CRISPR, which stands for Clustered Regularly Interspaced Short Palindromic Repeats, is a central part of the bacterial immune system and handles sequence recognition. Cas9 - Cas stands for CRISPR-assisted - is an RNA-guided enzyme that handles the sniping of DNA strands at the specified sequence.

[http://scicasts.com/bio-it/1856-molecular-biology/8463-rcas9-a-programmable-rna-editing-tool/]

Bioinfo. Animation



Chip-Seq flowchart

Patent News

Database for comparing and displayinggenomic information

WO 1998026407 A3

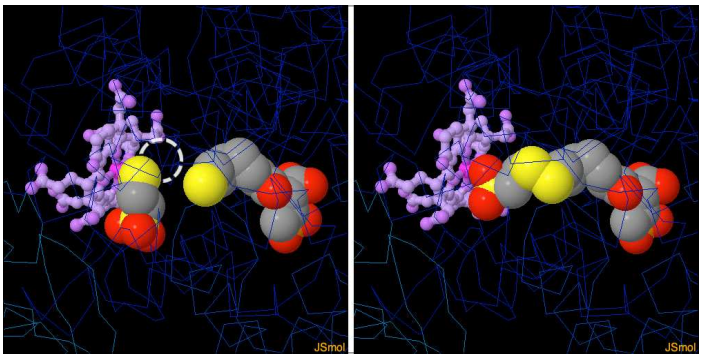
Publication date 8 Oct 1998
Inventors Cathryn E Sabatini,
Applicant Incyte Pharma Inc

ABSTRACT

Disclosed is a relational database system (146) for storing and manipulating biomolecular sequence information, the database including genomic libraries for a plurality of types of organisms, the libraries having multiple genomicsequences, at least some of which represent open reading frames located along a contiguous sequence on each the plurality of organ-isms' genomes, and a user interface capable of receiving a selection (662) of two or more of the genomiclibraries for comparison and displaying the results of the comparison (666). The system also provides a user interface capable of receiving a selection of one or more probe open reading frames for use in determining homologous matches between such probe open reading frame (s) and the open reading frames in thegenomic libraries, and displaying the results of the determination.

Methyl-coenzyme M Reductase

Methanogens are tiny microbes similar to bacteria that colonize anaerobic environments such as the bottom of lakes and swamps or the gut of cows and humans. They feed on molecules like carbon dioxide, methanol and acetic acid that are produced by fermenting bacteria, and release methane as their waste product, bubbling up as marsh gas, or in the case of our own resident methanogens, less socially-acceptable gases.

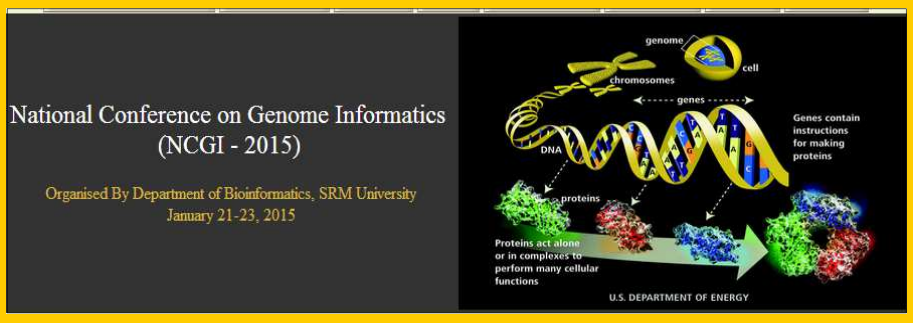


This tricky operation requires a series of enzymes that progressively strip oxygen atoms off the carbon and add hydrogen. The enzyme methyl-coenzyme M reductase, performs the last reaction in the process, releasing methane. It is a large complex composed of six chains: two copies of three different types. The enzyme has two active sites buried at the base of deep tunnels, protecting the reaction from the surrounding water.

Two structures give a glimpse of the enzyme before and after its reaction. Before the reaction, the enzyme binds to two cofactors: coenzyme M, which carries the methyl group, and coenzyme B. The enzyme releases the methyl group as methane, and connects the two coenzymes together with a disulfide linkage. This unusual disulfide-linked molecule is then used by the cell to produce energy.

[November 2014 Molecule of the Month by David Goodsell doi: 10.2210/rcsb_pdb/mom_2014_11 (ePub Version)]

Upcoming Events



SRM University, Intra College Rd
Potheri, Kattankulathur
Tamil Nadu - 603203.

Web: http://www.srmuniv.ac.in/ncgi/ncgi.html

Dibrugarh University, Assam, India

Web: http://dibru.ac.in/largefile/Chemistry%20Cheminar2015/index.html

Kindly send us your feedback to

Robin Das
Project Fellow; BIF, Biotech Division.
CSIR-North East Institute of Science and Technology, Jorhat, Assam
E-mail: robindas460@gmail.com
Ph No-07399923578