



IMSUT & RCAST GCOE 特別セミナー 〈キャリアパス支援テクニカルセミナー〉

「HGC スパコン Web サービス利用法講習会」開催のご案内

(医科研ヒトゲノム解析センター・スーパーコンピュータ室 主催)

ヒトゲノム解析センター (HGC) で利用可能な Web サービスを紹介いたします。

広く皆様のご参加をお待ちしております。

日 時：2013 年 2 月 7 日 (木) 10:00 ~ 17:00

会 場：東京大学医科学研究所ヒトゲノム解析センター 4 階セミナー室

対 象：ヒトゲノム解析センター計算機システム利用者および利用予定者

定 員：35 名 (事前申込者を優先、空席があれば当日参加も可能)

参加費：無料

10:00 - 10:30 HGC スパコンサービスの紹介

10:30 - 11:00 統合型データベース GeneCards の紹介

11:00 - 11:30 Comprehensive and reliable protein-protein interactions
using HitPredict and Hintdb

11:30 - 12:00 Gene function identification using eF-site and COXPRESdb

13:00 - 13:30 転写制御領域の解析用データベースとツール DBTSS, DBTBS, &
Melina2 と細胞内局在部位予測サーバ PSORT

13:30 - 14:00 がんエキソームシーケンスデータ解析パイプライン
Genomon-exome の紹介

14:00 - 14:30 Transcriptome sequence data から融合遺伝子を抽出する
プラットフォーム Genomon-fusion についての紹介

14:40 - 15:10 発現モジュール抽出ソフトウェア EEM の紹介

15:10 - 15:40 生体内分子モデルのグラフィカルシミュレーションソフトウェア
Cell Illustrator Online

16:00 - 16:30 生命システム情報統合データベース KEGG の紹介

16:30 - 17:00 大規模遺伝子ネットワーク推定ソフトウェア SiGN

[HGC スパコンサービスの紹介]

* HGC システムエンジニア

HGC のスーパーコンピュータシステムは、ゲノム分野では国内最大級の計算機システムで、各種サーバや高速大容量のディスクアレイ装置によって多様なゲノム研究環境を提供している。講習会では、HGC スーパーコンピュータシステムの概要、有償利用サービスの概要について紹介する。

[統合型データベース GeneCards の紹介]

* 小池麻子 ((株) 日立製作所中央研究所)

GeneCards は Weizmann Institute で構築されている統合型データベースである。遺伝子の配列情報だけでなく、蛋白質相互作用、遺伝子機能情報、SNP 情報化合物情報など様々な情報を一度に検索することができる。GeneCards に含まれる情報と効率的な検索方法について解説する。

[Comprehensive and reliable protein-protein interactions using HitPredict and Hintdb]

* Ashwini Patil (助教)

With the availability of a large number of protein-protein interactions, it is imperative to assess their reliability prior to their use in further scientific studies. HitPredict is a database that provides quality-assessed interactions in 9 species. Interactions are assigned with a score based on the genomic evidence found in their support. Such evidence can be obtained in the form of homologous interactions across species. The database, Hintdb, provides such a list of homologous interactions for those already determined experimentally. Together, HitPredict and Hintdb provide a useful resource of protein-protein interactions and their properties.

[Gene function identification using eF-site and COXPRESdb]

* Ashwini Patil (助教)

Gene function may be identified through the study of gene expression patterns or protein structure. Coexpressdb is a database of coexpression relationships between genes for 11 species calculated using gene expression data from numerous microarray studies to help identify new gene functions.

eF-site (electrostatic-surface of Functional site) is a database of functional sites on the molecular surfaces of protein structures. It displays the electrostatic potentials and hydrophobic properties on the protein surface in order to facilitate the identification of functional sites in proteins.

[転写制御領域の解析用データベースとツール DBTSS, DBTBS, & Melina2 と細胞内局在部位予測サーバ PSORT]

* 中井謙太 (教授)

一般に転写制御領域のイン・シリコ解析から有用な情報を得ることは容易でないことが多く、出来る限り実験情報と組み合わせた方がよい。DBTSS には東大新領域の菅野・鈴木グループがいわゆる TSS-seq 実験によって得たヒトなどの遺伝子の転写開始点情報と付随情報が収められている。DBTBS は枯草菌の転写に関する情報を文献から収集したデータベースである。また、Melina2 はいくつかの de novo 共通モチーフ抽出ツールを同時に走らせて、結果を比較検討するためのツールである。最後に、PSORT はタンパク質のアミノ酸配列情報をもとに、その細胞内局在部位を予測するためのプログラム群である。いずれも簡単な使い方を含めて紹介する。

[がんエキソームシーケンスデータ解析パイプライン Genomon-exome の紹介]

* 千葉健一 (学術支援専門職員)

* 白石友一 (特任助教)

エキソーム領域のシーケンスを行い、網羅的にエキソン領域の遺伝子変異を調べることは、シーケンス解析でスタンダードな解析となりました。しかし、たくさんのサンプルをシーケンスすることによって出力されるデータは膨大であり、大規模な計算リソースを利用したシーケンスデータ解析の技術が、効率的に研究を進める上で重要になります。このような現状のなか、Genomon-exome では HGC スーパーコンピュータを用いた並列処理によるデータ解析の高速化を実現いたしました。

本講習では、Genomon-exome を用いてエキソームシーケンスデータのマッピングから、データ解析までの流れを実演でご紹介いたします。

[Transcriptome sequence data から融合遺伝子を抽出するプラットフォーム Genomon-fusion についての紹介]

* 白石友一 (特任助教)

* 千葉健一 (学術支援専門職員)

融合遺伝子は、癌の発病機序に大きな役割を果たしている。近年のハイスループットシーケンス技術の発展により、網羅的に融合遺伝子を検出することが可能になりつつある。しかし、リピートや相同配列により生じる偽陽性が未だに多いことなどの数多くの問題があり、性能の良いアルゴリズムの開発は未だに重要な課題である。本講習では、融合遺伝子検出のための新しい

プラットフォーム Genomon-fusion について紹介する。まず、本手法のアルゴリズムの概要を説明し、その上で、実際の FASTQ ファイルから融合遺伝子検出までの解析の流れを実演の形で紹介する。

[発現モジュール抽出ソフトウェア EEM の紹介]

* 新井田厚司（特任助教）

EEM 法は遺伝子セット情報に基づいて発現データ中で共発現している遺伝子群、発現モジュールを抽出する手法である。公共データベースから取得したがん細胞のマイクロレイ発現データおよび遺伝子セット情報を EEM web service にアップロードし、スパコン上で解析、結果を取得するまでの流れを紹介する。

[生体内分子モデルのグラフィカルシミュレーションソフトウェア Cell Illustrator Online]

* 齊藤あゆむ（技術専門職員）

Cell Illustrator Online (CIO) は、生体内パスウェイを描く機能とシミュレーションの機能が一体となっているソフトウェアです。開発当初から「生物系の研究室に所属する研究者が実験をするかたわら、パスウェイの情報を知識整理し、その知識整理したモデルをシミュレーションすることで次の実験に何をするかを決められる」ことを目標として開発してきました。講習会では、CIO を用いて基本的な生体内反応のモデル化の方法を学習するとともに、サーカディアンなどの生体内パスウェイモデルのシミュレーションを行います。

[生命システム情報統合データベース KEGG の紹介]

* 矢野亜津子（Pathway Solutions）

KEGG (Kyoto Encyclopedia of Genes and Genomes) は、分子レベルの情報から細胞、個体、エコシステムといった高次生命システムレベルの機能や有用性を理解するためのリソースである。とくにゲノムやメタゲノムをはじめとしたハイスループットデータの生物学的意味解釈に広く利用されている。講習会では KEGG の基本的な検索法とパスウェイマッピング等のツールの利用法、および医薬品・疾患関連の KEGGMEDICUS 利用法を紹介する。

[大規模遺伝子ネットワーク推定ソフトウェア SiGN]

* 玉田嘉紀（東京大学大学院情報理工学系研究科コンピュータ科学専攻助教）

SiGN (サイン) は HGC のスーパーコンピュータを用いることにより、マイクロレイなどで得られる遺伝子発現データから遺伝子間の発現の依存関係を表す遺伝子ネットワークを推定することのできるソフトウェアです。ベイジアンネットワーク・状態空間モデル・グラフィカルガウシアンモデル・ベクトル自己回帰モデルが利用可能（一部実装中）で遺伝子ノックダウン実験データ、患者サンプル由来のデータや薬剤投与時系列データなどを利用して、静的・動的な遺伝子ネットワークを推定することが可能です。講習会では各遺伝子ネットワークモデルおよび推定法の紹介を行い、実際のデータファイル準備からセルイラストレータオンラインを用いたネットワーク解析までの流れを紹介します。

本講習会でご用意しているパソコンは MacBook です。講習会にはご自身のパソコンをお持ちいただくことも可能ですが、ネットワークセキュリティ誓約書にサインをいただく必要があります。

申込方法：下記フォーマットにて、2月5日（火）17:00 までに E-mail にてお申込みください。

E-mail: hgc-lect@hgc.jp 東京大学医科学研究所ヒトゲノム解析センター・スーパーコンピュータ室

「HGC スパコン Web サービス利用法講習会(2月7日)」に参加します。

氏名：
所属：
E-mail：
TEL：
ご要望：

本講習会についてご要望がございましたら、お申し込みの際にご記入をお願いします。

Dear All Users,

On February 7 (Thursday), we will be holding a workshop titled "Using Web Services", in which we will introduce, in omnibus format, the web services that can be used at the Human Genome Center. The workshop will be held in Japanese. If you are interested in attending, please contact us at hgc-lect@hgc.jp.

Supercomputing Services
Human Genome Center