

研究課題	研究期間	平成16年度～20年度	所属機関	研究課題代表者
ヒトゲノムネットワーク情報システムの構築			国立遺伝学研究所 生命情報・DDBJ研究センター	五條堀 孝

1.研究目的

ゲノムネットワークプロジェクトにおける本研究課題の目標は、ヒトゲノムから複雑な生体反応がどのように産み出されているのかを、多種多様な遺伝子やタンパク質間の相互作用の協調的ネットワークの解明を通じて明らかにするために、本プロジェクトで産生されるデータを統合整理し、更なる研究(生命機構の解明や創薬研究など)のためにフィードバックする、データのプラットフォームを構築していくことにある。そのため、共同研究環境としてのヒトゲノムネットワークプラットフォームを実現し、ヒト生体分子ネットワーク情報の収集、解析と利用者への公開を実施するため、パイプライン技法を中心とした、高速なデータの連携・解析・可視化手法の実現を目指し、基礎生物学から医・薬学の応用分野に至るまで統合的な情報の提供を可能とするデータベースを構築・運用し、そのデータベースからの発見的情報利用技術の開発を、東京大学、長浜バイオ大学、特定非営利活動法人システム・バイオロジー研究機構と共同で実施する。

また、ゲノムネットワークプロジェクトにおける共同研究環境としてのゲノムネットワークプラットフォームを構築のため、プロジェクト参加機関の産出するデータやパブリックデータなど、関連する種々の情報を統合化してヒト生体分子ネットワークの全貌解明に向けた研究の基盤となるデータベースの構築・提供を行うとともに、発生・分化、免疫・生体防御、生体恒常性(内分泌、脳・神経系)などの複雑な事象の解明に向けて、統合化した情報を利用する新しいデータ解析手法やツールを発明・開発する。それらのデータや解析環境を用いて、生命情報科学の立場から生体分子骨格データベースの完成と提供を目指す。

2.研究概要

具体的には、上記の目標を達成するために、

- (1)各研究グループから産出される大規模な生体分子間情報のデータを統合し、既に公開されている情報を含めた付加情報を与え、データベース化するシステム(これをパイプラインと呼ぶ)を開発する。
- (2)プロジェクト参加者間の情報交換を促進するべく、プロジェクトの成果情報をはじめプロジェクト内部の情報交換システムを構築し、プロジェクトメンバーに提供する。
- (3)ゲノムネットワークプロジェクトにおける研究成果を公開し、様々な基礎および応用研究に提供するためのインターフェイスを開発し公開する。
- (4)プロジェクト内外で得られたデータを活用し、新しい知識発見につなげるため、データベースからゲノムネットワークに関する情報および知識を抽出するアルゴリズム等独自の技術を開発する。

以上を大きな柱として、ヒトゲノムネットワーク解明のための基礎情報の維持、提供のための情報システムの構築を進めると共に、生命情報科学の立場から、得られたデータに付加情報を与えると共に新しい知識の発見を目指す。

3.研究成果

ゲノムネットワークプロジェクトは「ゲノム機能情報の解析」、「ヒトゲノムネットワークプラットフォームの構築」、「次世代ゲノム解析技術の開発」、「個別生命機能の解析」、「動的ネットワーク解析技術開発」の戦略的な研究プログラムから構成される。ヒトゲノムネットワークプラットフォームでは、これらの相互連携を助け、得られた成果を統合することにより、より相乗的かつ有用な共有知識として基盤となるシステムの確立を進めている。また、プロジェクトにおいて産出されるデータだけでなく、ゲノムネットワークに関する、既に公開されている関連データをも格納し、様々な情報をゲノムワイドに統合・提供するインターフェイスを開発し、ゲノムネットワークプロジェクト内外の研究者に公開している。また、ゲノムネットワークビューワーとして、ゲノム、PPI(タンパク質相互作用)、遺伝子発現それぞれのビューワーを通じて、比較ゲノム、遺伝子発現解析、PPIネットワーク解析などの多彩な機能を開発し、提供(図)も同時に行っている。これにより、我が国で行われるゲノム機能の解明において、基盤となる情報システムを確立するとともに、基盤となる知識情報を維持し提供していく。更に、当該インターフェイス構築の過程で、ゲノムネットワーク解明のためのさまざまな手法と解析データを産出する。これは、生物学的な新発見・生命現象の解明に貢献することが期待される。

また、独自手法の開発として、CAGEタグマッピングのための独自アルゴリズムの開発、ネガティブフィードバック予測アルゴリズム開発やシスエレメント予測ツールの実装など各種機能の強化に努めている。また、タンパク質相互作用情報解析のためのテキストマイニング手法、スプライスバリエーションデータベースなどの開発も進めている。同時に、これらの新規技術を含めてヒトゲノムネットワークの解析も進めており、得られた成果は、順次プラットフォーム上で、また論文として公開して行く予定である。

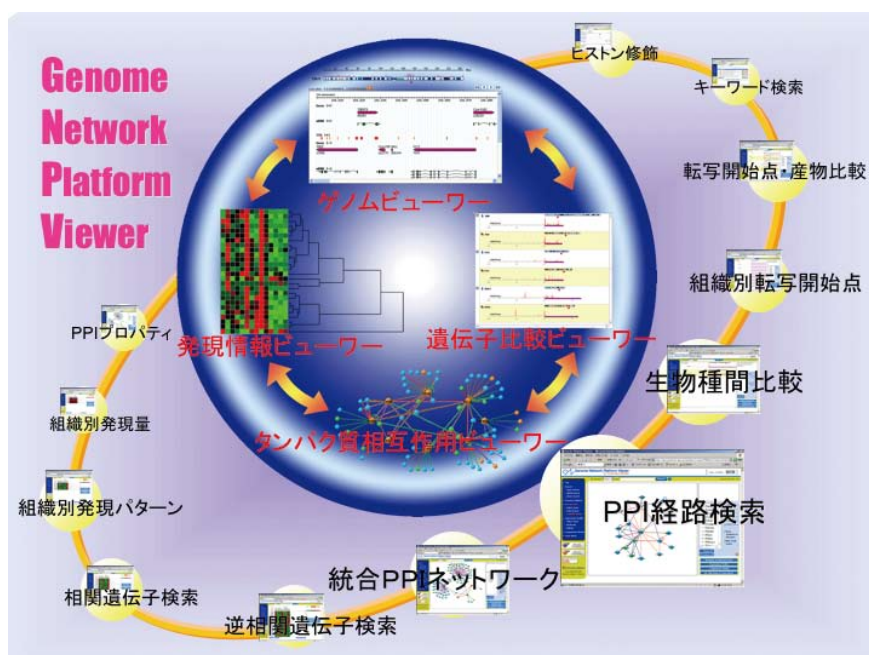
表 ゲノムネットワークプラットフォームにて一般公開中のデータセット

(2008年5月現在)

DB 分類	データ種別	レコード単位	レコード数	データ産生元
GNP Data	CAGE(human)	Tags	19,233,780	理化学研究所
		RNA samples	103	
	PPI(Y2H)	interactions	1,307	日立製作所
		proteins	997	
	qRT-PCR	genes	1,911	理化学研究所
		RNA samples	30	
	Tiling Array(titech)	locations	1,741	東京工業大学
		samples	4	
	Tiling Array(RIKEN)	locations	616	理化学研究所
		samples	2	
Illumina microarray	probes	47,293	理化学研究所	
	RNA samples	15		
Public DB	Gene(EntrezGene human)	genes	36,458	—
	Gene(EntrezGene mouse)	genes	63,977	—
	PPI	interactions	49,887	—
		proteins	10,175	—
	Expression(H-ANGEL)	RNAs	48,147	—
		tissues	40	—
	CAGE (mouse)	Tags	11,567,973	—
		RNA samples	209	—
	UniProt	proteins	164,040	—
	InterPro	domains	14,768	—
GO	GO terms	25,985	—	
OMIM	MIM IDs	19,198	—	

4.研究展望

プロジェクト開始から既存技術の調査をもとにデータベースおよび基本システムの設計を中心に進めてきたが、現在、「ゲノム機能情報の解析」の中核機関である理化学研究所からのデータを中心に関連する公開データを取り込み、コンソーシアム、一般向けそれぞれの公開を行っている。今後は、縦軸機関から産出される個別研究の成果を各種データと統合し、各種機能の開発を進めていく。データの定期的な更新提供と、並行して、独自のシステムを用いたデータの解析及び公開データを含めたデータ統合を進めると共に、新規技術の展開とそれに基づく新知識発見に力を入れていく。また、プロジェクト全体を統合したデータの供給下として、プロジェクト内のみならず、国内外のライフサイエンス研究に寄与できるように努力を続ける。



ゲノムネットワークプラットフォーム利用システム概要
<http://genomenetwork.nig.ac.jp>