

ゲノムネットワークプロジェクトの  
課題の報告

プログラム名	(1) <b>ゲノム機能情報の解析</b> (2) ヒトゲノムネットワークプラットフォームの構築 (3) 次世代ゲノム解析技術の開発 (4) 個別生命機能の解析 (5) 動的ネットワーク解析技術開発					
課題の分類	(1) 中核機関 (2) 指定課題 (3) <b>公募課題</b>					
課題名 (和英併記)	<i>In vitro virus</i> 法による転写因子複合体の大規模解析 (Large-scale Analyses of Transcription Factor Complexes by <i>in vitro Virus</i> Method)					
代表研究者名 (所属機関・職名)	柳川 弘志 (学校法人慶應義塾・教授)					
年度別研究費 (千円)	16年度	17年度	18年度	19年度	20年度	合計
	50,000	60,000	70,000	117,556	50,000	347,556
	研究者名	所属機関・部局・職			専門分野	
研究 組 織 図	代表機関	☆柳川 弘志	慶應義塾大学・理工学部・教授			分子生物学
		宮本 悦子	慶應義塾大学・理工学部・准教授			分子生物学
		斎藤 輪太郎	慶應義塾大学・環境情報学部・専任講師			バイオインフォマティクス
分担機関	(代表研究者)					
	(分担研究者)					
<b>【研究目的】</b>						
<p>ゲノムネットワークプロジェクトにおける「ゲノム機能情報の解析」プログラムの一環として、プロジェクトで必要な基盤的なゲノム機能データを産出し、横軸研究班が集中的に解析する転写因子関連タンパク質および縦軸研究班のリクエストタンパク質の相互作用に関するデータを質・量ともに強化・補完する。そのための技術開発を慶應大学理工学部と環境情報学部が連携して実施する。</p>						
<b>【課題の概要】</b>						
<p>上記目的のための技術開発を、慶應大学理工学部と環境情報学部が連携して実施する。理工学部では、主に大規模実験による転写因子関連タンパク質や縦軸研究班のリクエストタンパク質の相互作用に関するデータを産出し、環境情報学部では、バイオインフォマティクスによるデータの信頼性解析や複合体解析を実施する。具体的には、以下の4つの研究課題について研究を推進する。</p>						
<p>(1) IVV法による大規模解析システムの構築 (柳川)  (2) IVV法によるヒト転写因子のPPI解析 (宮本)  (3) IVV法より得られたPPIデータの検証および複合体解析 (斎藤)  (4) IVV法より得られるデータのサプライズ・バリエーション解析 (宮本)</p>						

## 【平成20年8月末までの研究成果】

### (1) IVV法による大規模解析システムの構築（柳川）

#### 1. IWAS (IVV Analysis System)の開発

ロボットを用いた、IVV法による96ウェルプレート使用の並列セレクションにより得られる大量の配列データの遺伝子の自動解析を行うために、平成16年度までに、IVV解析システム (IVV Analysis System: IWAS) を開発した。これにより、配列評価、遺伝子同定、信頼性評価、検出領域のマッピング等を自動で行うことができるようになった。平成17年度では、ソフトウェアを改善し、作業時間の短縮及びデータのダウンロード機能の追加等の高機能化を実現した。平成18年度以降、この開発したシステムを拡張することで、大規模解析システムの構築を行い、18年度には最初の大規模PPIデータ (約940PPIs) を、さらに19年度から20年度にかけては、縦軸連携の大規模PPIデータ (約600PPIs) をそれぞれ解析した。

#### 2. データ統合解析システムの開発

IVV法を用いて得られるタンパク質間相互作用データは、領域情報である。それらの相互作用領域をより詳細に解析するために、データ統合解析システムを開発した。具体的には、IVV法によって決定されたタンパク質相互作用領域に対して、バリエーション情報、SNPs情報 (NCBI dbSNP)、モチーフ/ドメイン情報 (InterPro) を統合した。バリエーション領域情報については、NCBI, DDBJ, UniProt から取得したアミノ酸配列の比較 (ペアワイズアライメント) によって相違が示されたものをバリエーション領域と定義した。これらの統合された情報は、リレーショナルデータベースの形でサーバーに蓄えられ、ユーザーはウェブブラウザを通じてアクセスが可能となっており、任意の条件 (例: ドメイン領域と相互作用領域が重なるPPIを抽出したい、等) を満たすPPIを抽出するためのフィルタリングを行うことも可能となっている。また、フィルタリングされた結果については、XML形式およびタブ区切りファイルとしてダウンロードすることも出来、これは後に続く二次的なインフォマティクスの解析などに利用し易い形式となっている。今回開発したシステムを用いることで、個別生物学的研究 (縦軸研究) に関わる遺伝子を中心としたアノテーション付きの詳細なPPIデータを迅速に出力することが可能となった。また、これらのアノテーション付きデータについて、平成20年8月にGNPプラットフォームにアップロードした。平成20年9月に、コンソーシアムに公開予定である。

#### 3. 実験支援システムの開発

複数の個別生命現象研究機関 (縦軸研究機関) との連携のための実験支援環境の強化として、次の三点を実施した。

一点目は、ベイトタンパク質作製を支援するためにベイト設計支援システムとIWASを連携させ、ベイト設計から解析までをスムーズに進行できるようにしたことである。二点目として、縦軸研究機関との連携による実験において、連携機関ごとに異なる複数種のライブラリを区別するために個別に設計されたタグ配列情報の登録およびその解析を可能にするためのIWASの改良を行った。三点目は、これまでヒトに特化されていたシステムをマウスにも対応可能にしたことである。またこれに関連して、マウスのベイトおよびライブラリを用いて行われた実験結果はヒトへの転用を考慮に入れ、ヒト・オーソログへの対応付けを自動的に行うためのツールも開発した。実際のマウスのPPIデータにヒト・オーソログへの対応付けを応用したところ、約85%変換出来た。これらの実験支援機能の強化によって、大規模なIVV実験をより円滑に遂行できる環境が整備され、縦軸研究機関との連携をスムーズに行うことが出来た。

## (2) IVV法によるヒト転写因子のPPI解析 (宮本)

平成 16 年度に、1) IVV 法の大規模解析のためのベイト作製の自動化、2) ライブラリ量産の自動化を確立した。平成 17 年度は、96 ウェルプレートでの並列セレクションの自動化を実現後、1) ベイト作製におけるロボットを用いた多種類ベイトの準備、2) ロボットによる二段階セレクションの条件決定、3) 先行実験として 96 ウェルプレートの一部のウェル (16 ウェル) によるセレクションを行った。この先行実験で行った 12 個のベイトタンパク質における相互作用検出実験の結果を解析したところ、200 以上の相互作用を検出することができた。平成 18 年度では、前年までの確立した 96 ウェルプレート全体を使用した相互作用検出実験システムを用いて、1) 縦軸班のリクエストである 18 転写因子を含むヒト転写因子に関する大規模タンパク質間相互作用実験の並列セレクションによる実施、2) タンパク質間相互作用検証実験であるプルダウン実験のロボットを用いた大規模化の検討、3) プルダウンによるタンパク質間相互作用の信頼性の検証を行った。並列セレクションにより得られた配列について、バイオインフォマティクスの専門家と協力して解析を行い、68 ベイト (50 転写因子) の結合相手として、約 940 相互作用を抽出した。さらに抽出した配列を、a) 擬陽性と考えられる配列、その出現頻度が、b) 一回であるもの、c) 複数回であるものに分類、各分類からプルダウン実験に用いる配列を選択、合計約 100 配列についてプルダウン実験により検証した。その結果、a は約 1 割、b は約 7 割、c は約 8 割の確率で相互作用が確認された。平成 19 年度は、縦軸研究機関との連携研究を進めるために、1) 大規模タンパク質間相互作用検出実験システムの自由度拡大のための改良、2) 東京大学医科学研究所 (井上純一郎)、九州大学 (福井宣規)、福岡大学 (白澤専二)、京都大学 (眞貝洋一)、大阪バイオサイエンス研究所 (裏出良博)、東京医科歯科大学 (高柳広) 等の 6 研究機関からのリクエスト、合計 43 ベイト (転写因子を含む 39 遺伝子) におけるマウスタンパク質間相互作用検出実験を行った。平成 20 年度に、それらの実験結果として、合計約 600 相互作用 (1 ベイト当たり換算して約 14 の相互作用) を検出した。得られた相互作用に関して、プルダウンアッセイによる検証実験による信頼性は、約 79%であった。ヒトに変換したデータについての信頼性は、約 75%であった。これらのデータは、平成 20 年度 8 月末に、遺伝研のプラットフォームへアップロードした。平成 20 年 9 月に、コンソーシアムに公開予定である。

## (3) IVV法より得られたPPIデータの検証および複合体解析 (斎藤)

平成 16 年度は、IVV 法より得られた PPI データの検証のための初期検討を、マニュアル実験で得られたデータで、インターログをベースとした検証を行い、IVV データの信頼性検証法を確立した。平成 17~18 年度は、ヒト転写因子の 6 ベイトの並列セレクションで得られたデータ (5 千シーケンス) を用いて、6 つのベイトタンパク質で 120 相互作用を解析した。その結果、IVV データの中にはいくつかヒトの既知の複合体が含まれることが示され、同時に、酵母菌の TAP 法で取得された複合体とよい一致が見られた。これらを土台として、平成 19~20 年度は、IVV 法により得られた縦軸リクエストを含む 50 転写因子の大規模解析 (3~4 万シーケンス) について、信頼性解析と複合体解析を行った。実験データの妥当性の検証 (信頼性解析) は、(a)既知の知見と合致するもの、(b)インターログ (相互作用するタンパク質 A, B のそれぞれのホモログ A', B' の間の相互作用の解析)、(c)既知のモチーフ間相互作用、という 3 つの指標を用いて行った。複合体解析について、公共データベースに登録されている相互作用ネットワークより、相互作用が密な部分を複合体が形成される可能性が高いクラスター (機能モジュール) として抽出し、さらに IVV 配列を貼付けることによって、互いに同時に結合し得るタンパク質の組み合わせとして、既知の複合体 13 個を含む 61 個を複合体の候補として抽出した。また、平成 19 年度末に、相互作用の信頼性指標をもとに、ユーザが任意の指標を用いて信頼性の高い相互作用を絞り込んだ後、データの再現性などを視覚的に確認できるような GUI (graphical user interface) を

整備し、遺伝研へアップロードした。また、複合体抽出の開発・改良に力を入れ、そのアルゴリズムの定式化を行うと共に、公共データも取り入れて複合体予測を行った。

#### (4) IVV 法より得られるデータのサプライズ・バリエーション解析 (宮本)

##### 1. IVV 法より得られたデータを用いたサプライズ・バリエーション解析の検討

平成 16 年度には、IVV 法より得られたデータを用いたサプライズ・バリエーション解析方法を検討するために、その時点までに得られていた IVV データの中から、サプライズ・バリエーションをもった遺伝子の抽出を試みた。IVV データのクラスタリング、サプライズ・バリエーションのデータベースの選択、IVV データとサプライズ・バリエーションのマッチング、抽出基準の調整など、解析の土台作りを行った。平成 17 年度は、前年度の検討結果をもとに IVV のプレ実験データ (6 ベイト ; 5 千シーケンス) を用いて解析を行い、タンパク質間相互作用領域がサプライズ・バリエーションによって変化している遺伝子を 20 前後抽出することができた。この過程で、複数のサプライズ・バリエーションデータベースの統合、IVV データとサプライズ・バリエーションのマッチング工程の半自動化、抽出された候補遺伝子の画像化などを行い、より大量のデータへの対応が可能な解析方法を確立した。

##### 2. 大規模データへの解析方法の適用と解析の自動化

平成 18 年度は、平成 17 年度までに確立した解析方法を大規模実験データ (50 ベイト ; 3~4 万シーケンス) に適用した。その結果、スプライシングのパターンによっては、タンパク質間相互作用領域が欠損しているアイソフォームが生産される遺伝子が 20 個程度抽出されてきた。この内の数個について、実際に相互作用能に違いがあるのかを *in vitro* の実験によって検証したところ、異なるアイソフォーム間で相互作用に違いがあることが判った。この結果は選択的スプライシングが相互作用の on/off、あるいは強弱を決定する機構として実際に生体内で働いている可能性を示唆するものである。以上のように、大規模 IVV 実験によって決定された相互作用部位とサプライズ・バリエーションデータの統合、データの解析、検証実験という一連の流れが完成させることができた。

さらに、平成 19 年度からは、上記の解析方法を PDB に登録されている複合体情報にまで拡張させ、相互作用が選択的スプライシングによって変化すると考えられるノード (タンパク質) の網羅的な抽出を試みた。同一遺伝子に由来する異なるタンパク質配列間の比較からバリエーション領域を決定し、そこからさらにバリエーション領域とタンパク質相互作用領域とが重なる場合の抽出を行ったところ、相互作用領域情報を持つ 1,842 遺伝子中、1,085 遺伝子が相互作用領域とバリエーション領域の重なり (1 残基以上) を持つことが明らかとなり、予想以上に多く遺伝子がバリエーションに応じて PPI ネットワークを変化させている可能性が示唆された。これらの「IVV データとサプライズ・バリエーション解析」に関する解析過程については、(1) の課題で構築されたシステムへの組み込みを行い、完全な自動可を実現した。

#### 【GNP との関連・GNP への貢献】

##### a) 転写・制御の解明に関連する研究成果

##### (1) IVV 法による大規模解析システムの構築 (柳川)

大規模 IVV 実験データから得られる大量の転写因子を中心とした相互作用領域データを円滑に処理し、さらに、バリエーション情報、SNPs 情報 (NCBI dbSNP)、モチーフ/ドメイン情報 (InterPro)、疾患情報 (OMIM) との関わりについても多角的な解析を可能にするシステムを開発した。

(1) IVV法によるヒト転写因子のPPI解析 (宮本)

IVVを用いて、転写因子の相互作用検出に適した大規模タンパク質間相互作用検出システムを確立し、50転写因子の結合相手として、約940相互作用を抽出した。このデータにより、ヒト既知50転写因子の公共データを約2倍に増やすことが出来た。また、IVV法が相互作用領域を同時に解析出来ることから、転写因子の相互作用領域の詳細な解析が実現し、転写因子のネットワークの特徴として、その相互作用領域に非構造領域を積極的に採用していることが明らかとなった。転写因子は非構造領域の割合が高いと言われていたが、今回、初めて、相互作用領域に顕著に集中している事実が分かった。このことは、転写因子が多く相互作用相手と頻繁に複合体形成を交換するのに適した構造であることを示唆する。

(3) IVV法より得られたPPIデータの検証および複合体解析 (斎藤)

転写因子をベイトにすることで、転写制御に関わるタンパク質複合体を予測することができ、どのタンパク質とどのタンパク質が組み合わさって転写制御を行っているか、推測できた。今後複合体中の各タンパク質の役割を調べることで、転写制御に関する知見がより深まると考えられる。

(4) IVV法より得られるデータのサプライズ・バリエーション解析 (宮本)

相互作用領域が選択的スプラインシングによって欠損した場合、PPIネットワーク (相互作用ネットワーク) が変化すると考えられ、特に変化を起こすノードがハブである場合には、相互作用ネットワークに大きな変化をもたらすと考えられる。本研究課題の目的は、上記のような選択的スプラインシングによって変化する可能性を持つPPIを網羅的に抽出し、その結果生じる相互作用ネットワークの変化が生物学的にどのような意味を持つのかを明らかにすることであった。IVV法で大量に得られたデータは転写因子を中心とした相互作用領域データであり、本解析で得られる結果、すなわち相互作用ネットワークに影響を及ぼすサプライズ・バリエーションは、転写制御ネットワークに直接関わる重要なものだと考えられる。

b) 縦軸研究 (または横軸研究) への貢献に繋がる研究成果

(1) IVV法による大規模解析システムの構築 (柳川)

縦軸研究機関のリクエストに応じて行ったIVV実験によって決定されたPPIデータに対して、疾患情報、バリエーション情報、SNPs情報、モチーフ/ドメイン情報、遺伝子発現情報を統合するためのシステム開発を行った。開発されたシステムによって得られた結果を付加したIVVの相互作用データは、現在ゲノムネットワークプラットフォームにて公開されている。また、異なる研究機関ごとの多様なcDNAライブラリーやベイトの設計・管理などに対応するため、IVVデータ解析システム (IWAS) の実験支援に関する機能を強化し、縦軸研究機関との連携を円滑に進めることに貢献した。

(2) IVV法によるヒト転写因子のPPI解析 (宮本)

縦軸研究機関への貢献として、平成18年度に、縦軸研究機関リクエストである18転写因子について相互作用検出実験を実施し、約260相互作用を検出し、遺伝研へアップした。平成19~20年度に縦軸研究機関リクエスト39遺伝子について相互作用検出実験を行い、約600相互作用を検出し、遺伝研へアップし、コンソーシアムへ公開した。また、横軸研究への貢献として、ヒトタンパク質へ変換した後、検証実験を行い、マウスとヒトの両データを遺伝研のプラットフォームへアップした。