

平成 1 6 年度大規模新規研究開発評価  
第 2 回評価検討会提出資料

「ゲノムネットワーク研究」  
府省への質問事項回答

平成 1 5 年 1 0 月 1 日

文部科学省

質問 1 の回答の一部、質問 2 の回答は、未発表の研究内容、または  
知的所有権に関わる内容を含むため、

(問) ゲノムネットワークの定義を示されたい。

(回答)

ゲノムは、遺伝子の発現を時間的、空間的、量的に調節し、多数の遺伝子の相互作用を生み出し、全体として複雑な生命活動が成り立っている。この遺伝子発現の時間的、空間的、量的な調節と、遺伝子間の相互作用は、多数の転写調節因子と転写調節エレメントの相互作用を基本とし、調節因子の働きをコントロールするシグナル伝達系の多数のたんぱく質、さらにそれらを修飾し活性化・不活性化させる様々なたんぱく質やホルモンなどの複雑な相互作用がかかわって、生み出されている。このようなゲノムの全遺伝子の働きを調和させ、生命活動を成り立たせている仕組みをゲノムネットワークと呼ぶ。

## 1. 対象とするゲノム等の選択理由

(問) この研究の対象生物は何か、その理由とともに示されたい。理研及びその他日本の機関が有する cDNA がカバーしている翻訳枠は、全遺伝子の何パーセントに相当するのか、理研とその他に分けて、ヒト、マウスごとに示されたい。また、どの遺伝子、どのタンパク質から研究に着手するのか。プライオリティ及びその基準を明白に示されたい。例えば創薬/医療分野に成果が応用される遺伝子、細胞、細胞内パスウェイなどを優先的に行わないのか。

(回答)

研究の対象生物は、ヒトおよびそのモデルとしてのマウス(プロモーターの解析では一部の霊長類も利用)である。

文部科学省科学技術・学術審議会研究計画・評価部会ライフサイエンス委員会ゲノム領域小委員会の報告にあるように、本プロジェクトは、ヒトゲノムシーケンスの解読完了を受けて打ち出されたものである。そのため、ヒトを対象にすることはもちろんのこと、ゲノムシーケンス解読が終了し、cDNAクローンの収集が充実している生物種をその対象とすべきである。現状でのヒトゲノムシーケンスの解読完了及びマウス完全長cDNAのシーケンスの充実を考えると、最大の研究対象であるヒトとそのモデル動物であるマウスが、主たる研究対象生物としてふさわしいと考えられる。とりわけマウスは、ヒトで解析できない部分において、極めて大きな役割を担う予定である。

cDNAの中のCoding geneについて、理研及びその他の国内の機関が有する、翻訳枠を持つcDNAクローン数と確定遺伝子数に対する割合は下記の表のとおりである。遺伝子あるいはタンパク質の面から見た研究のプライオリティは、ゲノムネットワークの根底にある転写調節因子とそれに繋がるシ

グナル伝達系関連のタンパク質が最も高く、次いで創薬 / 医療分野への応用が期待されるGPCR / 酵素等が続き、さらにわが国で解明されたものを中心に重要な疾患の原因・関連遺伝子も研究対象とする。しかし、本プロジェクトは、最終的には全遺伝子 / タンパク質間の相互作用の検討を目指すものである。

(問) ゲノムネットワーク研究が、当面ポストヒトゲノムとして推進されることは、理解できる。ただ、植物や、魚、無脊椎動物等のヒト以外のゲノム研究やポストゲノム研究も、生物学として重要なばかりでなく、ヒトゲノム研究と意外な関連をもつこともあるとおもわれるが、それらとの関係はどのように考えているのか示されたい。

(回答)

前述のとおり、本プロジェクトは、文部科学省科学技術・学術審議会研究計画・評価部会ライフサイエンス委員会ゲノム領域小委員会の報告である「ポストヒトゲノムにおける重点的取り組み方策について」において、ヒトゲノムシーケンスの解読完了を受けて打ち出されたものである。

そのため現時点では本プロジェクトは、ヒトのゲノム機能解析を目指すこととしており、ヒト及びそのモデルとしてのマウスのゲノムについて体系的に研究することとしている。またその他の生物のゲノム情報についても、ヒトゲノムのネットワーク解明のために、比較ゲノム解析等の手法で活用する予定である。

2. 網羅的にゲノムネットワークを解析する場合に、研究成果の到達目標をどう設定するのか？

(問) 転写調節因子のネットワーク解明の最終目標には、ゲノム上を網羅的に解析した場合、いくつ位の転写調節因子の解明が予想されるのか？ また、その推定の根拠は何か？ そもそもこの研究にゲノム解析のような完了はあるのか？ 計画年度内に解明できる転写調節因子は、ゲノムに存在する転写調節因子の何パーセントをカバーするのか？ それぞれについて示されたい。

(回答)

未発表の研究内容、または知的所有権に関わる内容を含むため、一般には非公開とする。

### 3. オールジャパン体制の構築する仕組みについて

(問) 事業別予算額は技術の改良やブレイクスルーによって変動するはずである。5年という期間を考えると、集中的解析の部分も方法論の進展によっては遂行グループの変更など大胆な見直しも必要である。中核的機関をおくとしても予算のアロケーションは中央推進機関で適切かつ柔軟に考慮されるべきであるが、中央推進機関はそのような構造・権限になっているか。

(回答)

中央推進機関については、グループ全体の研究開発目標を設定し、最も効果的・効率的に計画を実施するために、必要な全体調整を行うこととしている。そのため、予算の割り振り、事業計画の立案の基本方針を含めた、総合的な統括を行う予定である。

そのための組織として、中央推進機関の下に、提案公募課題の採択に係る委員会、ゲノム機能情報の集中的解析に関する委員会、知的財産の取り扱いに関する委員会、統合データベースに関する委員会等の設置を予定しており、これらの組織は、外部の評価委員の参画を得て透明度の高い運営を心がけることとしている。

(問) 研究当初において、理研に年間35億円の予算がいくことの妥当性について、現状のタンパク間相互作用解析能力とも関連し、具体的な各項目・内容について、積算算出根拠等を示されたい。

(回答)

現行のほ乳細胞2-ハイブリッド法については、目標とする大規模タンパク質間相互作用解析に適合することを確認済みである。従って機器と人員を整備することにより、タンパク質の組み合わせで、年間に $2 \times 10^8$ の組み合わせ(想定総組み合わせ数の20%強)を初年度からスクリーニングすることが可能である。(参考資料2参照)

(問) ゲノム機能情報の集中的解析に記されている「研究開発方法」については理研以外で実施することがより効率的と考えられるものもあると思われる。その実施に於いて、当面は理研を拠点とした上で他の機関と連携してはどうか？ 同様に総合データベースの開発に於いても、その実施について当面は遺伝研を拠点とした上で他と連携してはどうか？との意見があるが、理研、遺伝研それぞれが実施拠点として位置付けられる理由は何か示されたい。

(問) 現状のタンパク相互間解析技術では、完全にはゲノムネットワークを解析できない可能性が考慮され、そのために公募型の研究が検討されることとなっているが、より広汎な公募を中心とする競争的資金の大半を当て、新しい発想や知恵を導入するのが妥当ではないのか、との意見がある。この研究を理化学研究所主導で行う理由について示されたい。

(回答)

文部科学省科学技術・学術審議会研究計画・評価部会ライフサイエンス委員会ゲノム領域小委員会の報告である「ポストヒトゲノムにおける重点的取り組み方策について」において、ゲノム機能情報の集中的解析に必要なポテンシャルについて、「集中的な解析を行う機関は、ヒトなどのゲノムネットワーク解析に必要な十分なリソース、大規模解析技術、設備、人材を持ち、解析のための技術開発をも行い、解析・収集データを提供する能力と意欲があること、また、それらが国際的に優位性を持っていることが必要である。」とされている。

理化学研究所はオリジナルの完全長cDNA作成技術を開発し、6万種のマウス完全長cDNAや14千種のヒトcDNAを収集保有していること、遺伝子カバー率世界最大級のDNAマイクロアレイ開発等を行っており、ヒトなどのゲノムネットワーク解析に必要な十分なリソース、大規模解析技術、設備、人材を持ち、

解析のための技術開発をも行っている。また、ゲノム機能の解析・収集データを提供する能力があり、それらが国際的な有意性を持っている。さらに、ゲノム機能情報の集中的解析を行うための基盤が整備されているため、日本独自の強みを活かして、ゲノム機能情報の集中的解析を行うには最適な研究機関と考える。

また、統合的データベースの構築を行う機関についても、質量ともに十分なインフォマティックスの設備、技術及び人材を有し国際的な優位性を持つ必要がある。遺伝学研究所は DDBJ 等の運用実績からみても、これらの条件を十分満たすものと考えられる。

なお各機関の具体的に研究の推進にあたっては、他の機関との共同・連携は大いに進めていく必要があると考えている。例えば理化学研究所の場合には、現在技術的に可能なゲノム機能情報の集中的解析を体系的に進め、その過程で新たなゲノム機能情報解析に関する技術開発及び導入を図ることとしており、その際には、提案公募により募集しているゲノム機能情報の解析を行っている機関・組織との連携が考えられる。

(参考資料 3参照)

#### 4 . 具体的な制度 運用について

(問) わが国のゲノム・タンパク関連のバイオベンチャーも海外に負けない技術を擁するものも出てきた。大企業だけでなく、これらバイオベンチャーの国家的大型プロジェクトに参画できる機会を多く与えて欲しいと思うが、バイオベンチャーなどが本プロジェクトに参加できる仕組みはどのようなになっているか示されたい。

(回答)

次世代ゲノム解析技術開発等の提案公募型の研究課題に対してバイオベンチャー企業が参加することは、なんら妨げられるものではなく、バイオベンチャー発の技術が、評価を受けた上で、将来的に網羅的な解析等に用いられる可能性も考えられる。この他にも、本プロジェクトに採択された研究機関との共同研究を行うことにより、技術の移転を受ける形で本プロジェクトに参加することも十分あり得ると考えている。

(問) 昨年、理研の発生・再生化学総合研究センターに27歳の若い研究者がリーダーに抜擢されて注目された。わが国の大学では若手研究者が登用されることが困難なことから、国の大型プロジェクトにおいては若い優秀な研究者に大きく育つチャンスを与えてもらいたく思うが、本プロジェクトにおいて、若手研究者登用の仕組みはどのようになっているか示されたい。

(回答)

本プロジェクトの提案公募型の研究課題に対しては、自由な応募が可能であり、中央推進機関及び傘下の組織における審査の上で、採択を行うこととしている。公募に当たっては、若手研究者についても広く応募できるように配慮されることが望ましいと考えている。

(問) 放医研から提案されている「ハイセツプ法」は発現遺伝子を網羅的に解析できるものではないかと思われる。また、大規模新規研究開発のひとつである「先端機器計測分析技術 機器開発事業」の中などからでも早期に成果があがり、ゲノムネットワーク研究に応用される技術成果が得られるのではないかと期待される。このようなプロジェクト外で有効な手法が開発された場合に柔軟にそれらの技術を導入する仕組みはどのようなになっているか示されたい。

(回答)

本プロジェクトについては、中央推進機関において随時研究体制を見直すこととしており、ゲノムネットワーク研究に極めて有効と考えられる技術が、プロジェクト外で開発された場合には、中央推進機関の下で検討した上で、具体的にプロジェクトに取り込むこととしている。

(問) 本プロジェクトの提案公募研究制度で推進する部分と科学研究費など他の競争的資金制度との関係はどのように示されたい。

(回答)

研究者の自由な発想に基づく科研費等競争的資金による研究とは異なり、本プロジェクトの提案公募を行う部分については、ゲノムネットワーク研究、あるいはゲノム機能の集中的解析への寄与という観点から研究を推進する必要がある。このような研究課題は定められた機関の中で実施されることから、より明確な目標の設定、ある程度の実現性を具備していることが求められる。例えば次世代ゲノム解析技術開発においては、3年程度で実用化の目処を立て、集中的解析に供することができる研究開発を行う必要がある。

なお本プロジェクトの提案公募による研究実施にあたっては、競争的資金の下で実施をされている研究との間で事前に十分調整を図る必要がある。

## 5. 他省庁等のプロジェクトとの協力体制について

(問) 予算規模は NIH に比べてはるかに小さいことから、他の省庁の同様なプロジェクトも巻き込んだシステムティックネットワークが必要となってくると思われる。タンパク3000等各省庁との予算の重複の防止や協力関係など、どのように仕組みが考えられているか示されたい。

(回答)

タンパク3000 プロジェクトは、タンパク質の基本構造を解析するプロジェクトであり、その活性部位を明らかにして、そこに結合する物質を解析することにより、タンパク質の機能を知ることができる。そのため、ゲノムネットワーク研究の成果によって重要な役割を持つことが明らかとなったタンパク質については、タンパク3000 プロジェクトにおいて、構造等の解析を行うことにより、画期的な新薬の設計等への道が拓かれ、ゲノム創薬の推進につながるものと考えている。

また、ゲノムネットワーク研究では、疾患等関係遺伝子同士の間関係を調べることが可能となる。そのため、テーラーメイド医療実現化プロジェクトにおける遺伝子解析をより詳細にすることができるだけでなく、ゲノムネットワーク研究において、どのタンパク質を網羅的に調べるかを示唆することが可能となり、相補的な効果を期待することができる。

また提案公募型の採択にあたっては、当プロジェクトの主旨にのっとり、他の機関のプログラムの研究課題と重複が生じないよう十分調整・配慮したい。

## 6.技術背景について

(問) Glycobiology、Epigenetics はどのような形で含まれているのか示されたい。

(回答)

Glycobiology、Epigenetics については、よい提案があれば、個別生命機能解析等の提案公募型の課題として取り入れて参りたい。

(問)公募型の個別生命機能解析の課題に於いて「分子ネットワーク解析」の意味が明瞭とは言えない。転写制御系(トランスクリプトーム)とシグナル伝達系(シグナローム)を共に解析するものなのか具体的に説明されたい。

(回答)

個別生命機能解析に関する提案公募課題では、発生、細胞分化、細胞周期、免疫応答、サーカデアンリズム、薬物応答など様々な生命現象の解析をゲノムネットワークという視点にたって解き明かすことを想定している。したがって、転写制御系(トランスクリプトーム)とシグナル伝達系(シグナローム)のどちらもが対象になる。これらの研究ではゲノム機能情報の集中的解析によって得られたデータをフルに活用し、情報を共有しながら進めることが求められる。

(問)タンパク質間の相互作用の分析は主に“Two hybrid system”によるようだが、現在考えられているその他の技術(タンパクチップ、イーストのツーハイブリッド系、タグ付きのコンプレックス解析など)について、それぞれの特徴、技術補完性などについて対比して示されたい。また、タンパク・タンパクの相互作用を広汎に調べるにはより巾広い技術の採用が必要と思われるが、現在の技術で充分解析できるのか、また不足であれば問題点等を示されたい。

(回答)

タンパク質間相互作用を分析する手法に関して、それぞれの方法の概要、特徴(長所と短所)、技術補完性について別表にまとめた(参考資料4参照)。網羅的にタンパク質間相互作用を調べるためには、大きな障害となる技術的困難さがないことが必須である。網羅性からは2-ハイブリッド法が優れており、更にヒト(及びマウス)を対象とすることを考えると、ほ乳細胞2-ハイブリッド法をスクリーニングの主力に置くのが最も適していると考えられる。また、技術補完性を考慮すると、その他の各手法を適宜取り入れることも重要と考えられる。例えば、ほ乳細胞2-ハイブリッド法と酵母2-ハイブリッド法、ほ乳細胞2-ハイブリッド法とタグ付きコンプレックス解析のように、2つ以上の解析手法を組み合わせると、より精度が高く、カバーする領域が広いデータが得られるものと考えられる。

タンパク質間相互作用をさらに広汎に調べるためには、タンパク質の局在や、修飾を考慮に入れたアッセイが必要となる。膜タンパク質を対象とした相互作用、リン酸化されたタンパク質の相互作用などは特に重要と考えられる。これら特殊な相互作用をアッセイする方法に関してはプロトタイプが、ほ乳細胞2-ハイブリッド法に既に存在するので、それを改良しハイスループット化する研究が必要となる。「ポストヒトゲノムにおける重点的取り組み方策について」では、こ

これらの技術について「一部技術開発を伴いながらも、大規模化への取り組みが可能なもの」と扱っており、本プロジェクトでは、これら技術開発について組みつつ研究を進捗することとしている。

(問) タンパク質発現プロファイリング(標準データの確立)と関連させ、次世代ゲノム解析技術の開発に於いて「抗体チップ」「擬態チップ」が、考えられると思われるが、これらの検討状況などについて示されたい。

(回答)

「抗体チップ」「擬態チップ」に関しては、ゲノム機能情報の解析、次世代ゲノム解析技術開発等の提案公募課題において、よい提案があれば、採択することが考えられる。

(問) 数千、数万の遺伝子が関与した大規模なネットワークを推定し、生物学的に意味のある結果が得られた研究はこれまでに例があるのか。もしあるのなら、示されたい。もしないのなら、その理由を示されたい。

(回答)

大規模なネットワークを推定し、生物学的に意味のある結果が得られた研究例を下記に示す。Reference は結果の発表論文である。

1. 酵母の 6,000 の全遺伝子をもとに、全てのタンパク質間相互作用を明らかにしようとする試みの中では、1,004 個のタンパク質が関与する 957 ペアの相互作用が検出され、タンパク質間ネットワークの網羅的解析に成功している(ref. 1)。この試みにおけるネットワーク解析の一例を挙げると、オートファジー過程に機能するタンパク質と細胞周期のチェックポイントに重要なタンパク質との相互作用を見だし、二つの異なるパスウェイのクロスリンクが明らかにされた。
2. タンパク質間相互作用は、少数の限られた数の遺伝子 (タンパク質) がネットワークを形成しているというタイプではなく、スケールフリー型と呼ばれる、限られたハブに当たるタンパク質を中心に、多くのタンパク質との相互作用が認められていて、このようなハブでは莫大なリンクがあることが明らかにされた (ref. 2)。
3. 酵母の転写調節因子を用いて、調節因子間の制御ネットワークが書けた例がある(ref. 3)。本研究では細胞周期を転写調節因子の制御関係で解析できた。
4. *C. elegans* で大規模に解析されたタンパク質間相互作用(interactome)と表現形質解析(Phenome)を組み合わせて、DNA damage response に関与

する遺伝子を多く同定することができた (ref. 4)。

5. 酵母で大規模に解析されたタンパク質間相互作用(interactome)と表現形質解析を組み合わせることにより、DNA の Repair 系と Cell cycle 系、細胞骨格制御系、クロマチンのリモデリング系、蛋白や RNA の分解系、脂質代謝系などが相互に関係していることが示された(ref. 5)。
6. 酵母で大規模に解析されたタンパク質間相互作用(interactome)と表現形質解析(Phenome(knockout lethality))を組み合わせることにより、生存に必須なタンパク質 (essential protein) はお互いに相互作用しやすいこと、lethal protein を介した情報伝達のクロストークはほとんど存在しないことが示された(ref. 6)。
7. マウス遺伝子の大規模解析から得られた情報を元に行った個別解析研究例である。マウスの 3,500 個のタンパク質をコードする完全長 cDNA を用いたパイロットスタディにより約 500 ペアの相互作用を検出し(ref. 7, 8)、これらの中の機能未知タンパク質について既知タンパク質との結合やタンパク質相互作用ネットワークにより、その生物学的機能を同定することができた (ref. 9, 10)。

T2BP TRAF2 binding protein, TRAF2, TRAF6 と結合することから TNF シグナルカスケード、IL-1 シグナルカスケードに関与する事がわかった (ref. 9)。

TIP-1 beta-catenin に結合し、その転写活性可能を抑制して Wnt シグナルカスケードに関与している事がわかった (ref. 10)。

(1) Proc Natl Acad Sci U S A. (2001) 98(8):4569-74.

(2) Nature (2001) 411: 41-42.

- (3) Science (2002) 298, 799-804.
- (4) Science (2002) 295, 127-131.
- (5) Mol. Cancer Res. (2002) 1: 103-112.
- (6) Biochem. Biophys. Res. Commu. (2003) 301(3):633-40.
- (7) Genome Res. (2001) 11(10):1758-1765.
- (8) Genome Res. (2003) 13(6B):1534-1541.
- (9) Biochem. Biophys. Res. Commu. (2002) 290(3) 1108-1113
- (10) J Biol. Chem. (2003) In press

このように大規模に解析されたゲノム関連データより、生命をグローバルに捕らえることは、すでにいくつかの生物種ですでに試みられており、個別研究から見出すことができない知見を得ることが実際に可能となっている。

本プロジェクトでは、複数のゲノム機能解析の技法を用い、転写制御、情報伝達等の幅広いゲノム機能データの作成を目指している。これらゲノム機能データを組み合わせることにより、単一の技法によって得られたゲノム機能データからは見つけることができない、さらに高次の知見を得ることが可能と考えられる。また、大規模解析データが個別研究に大きく貢献していることは7)であげた実証例のみならず、1)で掲げた大規模解析についての論文だけでも300以上の引用があり、新しい生物学的知見の発見においてコミュニティ全体に貢献していることは明白である。(参考資料5)

(問) ヒトゲノム計画は当初の様々な技術的問題点を克服して、プロダクション(大量データ生産)フェーズに入ることができた。大規模ネットワーク推定に技術的問題点(とくに情報科学的な)は存在しないのか。あるとすればそれを示されたい。また、技術的問題点はなく現時点でプロダクションフェーズが可能ならば、何個の遺伝子が関与するネットワークでどのような結果を得ることができるのか、ネットワークの規模としての数値目標を示されたい。

(回答)

ゲノム機能の集中的解析において技術的な問題はどの時点においてもあり「ポストヒトゲノムにおける重点的取り組み方策について」では、ゲノム機能の集中的解析における技術開発に関して、「技術的に実行に移せる項目」、「一部技術開発を伴いながらも、大規模化への取り組みが可能なもの」、「更なる技術開発が優先されるべきもの」の3つのカテゴリーに分類している。

これに従い、本プロジェクトにおいては、「技術的に実行に移せる項目」に関して、プロジェクト開始当初から大規模解析をスタートさせることとしている。

この「技術的に実行に移せる項目」に関しては、遺伝子の転写制御ネットワークと細胞内タンパク質ネットワークを解析するための基本技術であるハイスループット転写開始点解析法、one-hybrid法やchip-on-chip法などのタンパク質-DNA相互作用(PDI)解析法、各種のタンパク質間相互作用解析法があげられる。ネットワークを構築するための基本的ゲノム機能データベースを作成する上で直ちに使える技術である。また、より完成されたネットワークを構築するためには、さらに高度な解析手法が必要であり、これらについても並行して技術の開発行っていくこととしている。

ネットワークの規模について現段階で想定している目標は、主なデータとして1億個程度の遺伝子転写開始点の情報(ハイスループット転写開始点解析法)

1千～3千個程度の転写制御因子からなる転写カスケード(PDI)、現在の使用できる翻訳枠数(20千)とパイロットスタディにおける検出率の算定より、最低5万個のタンパク質間相互作用を同定する(解析対象となるこの翻訳枠数はその後増えることも予想される)。これらの情報を組み合わせることで、転写制御ネットワークと細胞内タンパク質ネットワークからなる大規模ネットワークの骨格が明らかになるものと期待される。

なお、その他の技術開発が必要な手法については、次世代ゲノム解析技術開発に関する公募枠が設けられており、公募された課題のうち実用化のレベルまで達したものから集中的解析に利用することが考えられる。

(問) 個別生命機能解析の内容はたかだか数十程度の遺伝子が関与する小規模なネットワーク (新しい情報技術はなくても手作業で可能) であり、従来型の分子細胞生物学研究である。これが混在しているために「ゲノムネットワーク」という曖昧な言葉がますます曖昧になっている。大規模ネットワークの部分と個別研究の部分は完全に分離して、別個に評価を行うべきである。それぞれ研究の「ゲノムネットワーク研究」における位置づけについてより明確に示されたい。

(回答)

ヒトゲノムシーケンスの解読完了にあたっては、ポストゲノム時代の鍵となる事項を把握し、ライフサイエンスの新しい展開を見据えることが重要である。生体分子の化学構造を収集した「ゲノム構造データベース」として、ヒトゲノム配列、完全長 cDNA、タンパク質の三次元構造、SNPs、HAPMAP などが産出され、さらに各々の生体分子における機能の一側面を網羅的に収集した「ゲノム機能データベース」が極めて重要な位置を締めるようになる。一つ一つの生命現象に焦点を当てた小さな単位のグループによる研究目標は従来と変わらないが、現在完成しつつある「ゲノム構造データベース」に加え、大規模網羅的に収集された「ゲノム機能データベース」を有効に活用するという、きわめて効率的な研究方式が出現し、そのアプローチは激変するものと考えられる。

また、網羅的に収集された機能データベースは、極めて有用なスタンダードのデータベースである一方、必ずしも個々の生命現象を全て対象として作られたものではない。そのため、そのデータベースでは尽くせない部分を、大規模センターがゲノムワイドスクリーニングを行うことにより補う必要がある。

このように大規模センターと小グループがお互いに相互作用しながらつくる縦横のマトリックス様式は、21 世紀のライフサイエンス先進国のさらなる飛躍につながるものと考えている。

(問) 転写調節因子を網羅的に解析する場合、転写が誘導される細胞の分化段階や細胞の種類、外部環境からの刺激とその後の時間経過によって、活性化される転写調節因子が異なることが考えられるが、どの程度まで、転写調節因子のネットワークを網羅的に解明するのか示されたい。

(回答)

網羅的解析の対象となる転写調節因子セットを設定し、以下の代表的かつ重要な生物事象における各転写調節因子遺伝子及びそれらの被制御遺伝子群の制御関係を調べることにより、遺伝子発現制御のカスケードを描き、また、各カスケード間のノードを探索してネットワークを解明していく。

解明予定の生物現象としては、

1. 発生における転写調節因子ネットワーク

各発生段階のマウスの組織をなるべく細かく分けて解析する。

2. 代表的な細胞株における転写調節因子

ヒト・マウスにおける代表的実験系となっている細胞株をつかう

ES細胞に関してもヒト・マウスで検討している。

3. 生体内細胞における転写調節因子ネットワーク

体内から単離できる細胞 (血球系・ガン細胞など)

以上により、条件を様々に変えた転写調節系ネットワークの基本的パターンを抽出しデータベース化する。個別研究でさらに特定の生物現象に関して知見を深める。

(問) 網羅的研究先行型ではなく、個別生命現象のゲノムネットワーク解明を先行させ、そこから得られた転写調節因子の情報をゲノムワイドに解析、ゲノムネットワークのデータ蓄積を図る、個別生命現象解明先行型に、プロジェクトを再編成できないか、との意見があるが網羅的な研究を先行させる理由は何を示されたい。

(回答)

前述のとおり、このプロジェクトの最終的に目指す体制とは、網羅的にゲノム機能データベースを作り、さらにそのデータベースと解析システムと、個別生命現象を研究するグループとが有機的に連携し、相互作用しながら研究を推進していくのが、ポストゲノムシーケンス時代の新しい研究システムであると考えている。そのため、網羅的研究と個別生命現象解明研究を並行させて実施することが重要である。ただし現実的には、既存のプロジェクト等で実施している個別生命現象の解明を行っている研究グループに、集中的解析によるゲノム機能の網羅的解析データをより有効に供給できるよう後者をやや先行させた形でスタートすることとなる。

## 7.到達目標及びその管理

(問)国としてゲノム関連研究全体としての方向性の中で、全体として5年後の到達点を示し、その中でこの課題がどのように位置づけられているかしめされたい。

(回答)

ゲノムネットワーク研究は我が国のゲノム分野の大規模研究の中では、DNA-タンパク質、タンパク質-タンパク質間のネットワーク解析に集中した研究であり、その成果をもとに個別的な疾患等における生命現象の解明を試みるプロジェクトである。

本プロジェクトによって、ヒト、マウスの DNA-タンパク質、タンパク質-タンパク質間相互作用の基礎的なデータベースが構築され、その情報が応用されることにより他のゲノム研究の大幅な加速が見込まれる。

たとえば、ここ5年で疾患の SNPs 研究が進み、多くの生活習慣病のリスク遺伝子が見つかることが想定される。しかし、これまでの研究結果をみると、それらの効果はリスクをせいぜい2~3倍あげるものであり、それらがなぜ病気のリスクに繋がっているのか、またそれをコントロールするにはどうすればよいのかを理解するために、ゲノムネットワーク研究による生体システム全体の理解は必要不可欠なものとなっているであろう。

また、ここ5年間にDNAチップ技術を使って多くの遺伝子の発現プロファイルのデータが蓄積されると考えられる。しかし、これらのデータの意味付けにはゲノムネットワークの理解が不可欠となる。

このように日々発展する生命科学において、生物、生命現象をひとつのシステムとして理解することに対して、ゲノムネットワーク研究は重要な意義を担っていることが想定される。

(問)ゲノムネットワーク研究という大きな課題の3年後の到達目標、5年後の到達目標を、一般国民(tax payer)用だけでなく scientist 用に具体的に、しっかりしたものを示されたい。研究推進の方策(人材、年次計画、スケール、サンプル、etc)についても、より具体的な計画を示されたい。また、個別目標への段階的到達点(いわゆるマイルストーン)と到達時期を設定は、この研究では半年が適当である(年度では不十分)との意見があるが、どのような目標設定等が検討されているか示されたい。

(回答)

基本的に、横断的プロジェクトは、逐次検証しながら進めていく予定である。技術開発の公募 個別生命現象ネットワーク研究に関しては、申請者に提案してもらってから、到達点が判明する。

また、網羅的ゲノム機能データベースの作成に関する年次計画は以下のよう  
に考えている。

1.

一年目：プロジェクトの立ち上げと基本的ゲノム機能データの作成

既に存在する技術を用いて基幹となるゲノム機能データの作成を開始する。すなわち、ハイスループット転写開始点解析法による転写調節領域解析、転写因子制御フレームワーク(タンパク DNA 相互作用)、非修飾状態のタンパク質間相互作用解析を中心とするゲノム機能データ作成を行う  
応用的ゲノム機能データベース作成のための技術開発を開始する。

2.

三年目まで：基本的ゲノム機能データの作成と応用的ゲノム機能データ作成  
一年目に立ち上がった基幹となるゲノム機能データの作成の量産化とデータベース作成を行い、解析開始を目指す。この年度でもハイスループット転写開始点解析法による転写調節領域解析、比較ゲノム解析によるブ

ロモーター領域の詳細な解析、転写因子制御フレームワーク (タンパク DNA 相互作用)、非修飾状態のタンパク質間相互作用解析が中心となる。

- (1) 量産化されてきた基幹となるゲノム機能データについてデータベースを作成して解析を開始し、ほぼ網羅的なものとする。
- (2) 引き続き応用的ゲノム機能データベース作成のための技術開発を行う。実用化のめどが立った技術については小規模なパイロット実験を行う。
- (3) 技術開発が完了した研究について、応用的ゲノム機能データの作成を開始し、この時期には本格化させる。具体的には、転写制御因子とプロモーターとの相互作用の動的変化、リン酸化などの修飾状態のタンパク質間相互作用解析、従来解析が難しかった膜たんぱく質間相互作用解析、遺伝子ノックアウト細胞を用いた表現形質解析、タンパク質や核酸の細胞内局在解析などが中心となる。

3.

五年目まで：応用的ゲノム機能データ作成の量産化と完成

- (1) 前年度に立ち上がった応用的ゲノム機能データ作成の量産化を目指す。
- (2) 量産化されてきた応用的ゲノム機能データについてデータベースを作成して解析を開始し、ほぼ網羅的に作成する。
- (3) 基本的ゲノム機能データ作成において未完成部分があれば完成させる。
- (4) プロジェクトを通じて得られる全ゲノム機能データの統合的解析を開始する。さらに、プロジェクトのまとめを行う。

網羅的ゲノム機能解析には、先行する既存技術を用いた基本的ゲノム機能データ解析と、技術開発を伴った応用的ゲノム機能データ解析という二本の柱がある。公募 個別生命現象ネットワーク研究に貢献するためには、できる限り早急に基本的ゲノム機能データをそろえること、蓄積された技術を基盤として、のより付加価値の高いゲノム機能データを取りそろえることが重要と考えている。

## 8. 統合データベースについて

(問) ゲノムネットワークの統合データベースを目指すためには、システムティックな研究体制が望まれる。またヒトゲノムの解読終了が宣言されたとしても、実際は約 1% の難解読領域が残っている。これらの観点から次の技術の参画を調査 検討し示されたい。システムバイオロジー、プロテオーム、メタボローム、遺伝子をノックアウトしての遺伝子ネットワーク、転写シーケンスによる難解読領域。

(回答)

ヒトゲノムの難解読領域の解明については、よい提案があれば、個別生命機能解析等の提案公募型の課題として取り入れて参りたい。

(問) インフォマティクスの約10億円の予算の使用目的と本プロジェクトの関係をより具体的に説明されたい。

(回答)

ゲノムネットワーク統合データベースプロジェクトの費用として、以下を試算している。

**1. データベース構築/インフォマティクスサービス提供 (計 520 百万円)**

ゲノムネットワーク統合データベースにおける、データ収集、データベース構築を行う。また、データ収集の一環として、データ作成グループに対する情報処理技術の提供を行う。

- (1) 基幹プロジェクト対応 (216 百万円)  
目的：本プロジェクトの基幹データによるゲノムネットワーク構築
- (2) 公募プロジェクト対応 (224 百万円)  
目的：情報処理技術提供
- (3) パブリックデータベース対応 (80 百万円)  
目的：公共データのゲノムネットワークへの取り込みを行う

**2. バイオインフォマティクス(アルゴリズム)研究開発/システム開発 (計 424 百万円)**

ゲノムネットワーク統合データベースにおける、バイオインフォマティクス技術の研究開発、データベースシステム開発、及びその運用を行う

- (1) バイオインフォマティクス研究開発 (52 百万円)
- (2) 解析アルゴリズム研究開発 (36 百万円)
- (3) マッピングアルゴリズム研究開発(36 百万円)
- (4) ネットワーク情報検索アルゴリズム研究開発( 36 百万円)
- (5) サブネットワーク抽出アルゴリズム研究開発( 24 百万円)
- (6) データベースシステム開発 (36 百万円)
- (7) データ標準化システム開発(24 百万円)
- (8) アノテーションシステム開発(60 百万円)
- (9) 情報表示システム開発(120 百万円)

**3. その他費用 (計 56 百万円)**

出張費、機材更新費その他

(問)「統合データベース構築」を作成するためにアルゴリズム、その哲学をつくる人の育成が重要である。現場をまわって考えをきく、勉強会を開くなどして考えを吸い上げ、データベースをつくり上げる制度が必須と思われる。公募によって即戦力となる人材を確保するのは困難と思われるが、これらの人材等を確保し、育成するために、どのような検討がなされているか示されたい。

(回答)

ゲムネットワークの統合データベース開発にあたり、人材の確保及び育成は非常に重要なものと認識している。その達成のため、中央推進機関の下に関連研究分野の専門家を中心とした統合データベースに関する委員会を設置することを考えている。この委員会では、データベースデザイン 最新技術の導入検討・アルゴリズム開発方針などの技術的な問題だけでなく、人材活用などについても検討を行い、中央推進機関はその検討結果に基づいて、統合データベースの研究開発・構築・提供体制に関する基本方針を決定する予定である。

(問) データの公開・共有についてのポリシーについて

- 1)本計画から得られるデータは知財権との関わりがあるため、ゲノムシーケンスの場合と同様な公開ポリシーはとれないとは考えられるが、データの公開についてどのような戦略・方策をとるのか示されたい。
- 2)中核機関が多額の経費を使っておこなういわゆる横軸研究については、データの速やかな公開が必要であり、少なくとも本計画に参加する研究者(コンソーシアム)に対しては一次データを即時に利用できる体制が必要であると考え、構想を示されたい。

(回答)

本プロジェクトにおける解析の結果得られたデータについては、迅速に流通される体制を構築し、最終的にはすべて公開すべきものと考えている。また、プロジェクトに参加している機関間においては、共同研究を含め、外部に対して守秘義務を負う契約をプロジェクト参加に当たって締結する等の方策により、事業全般にわたって、知的財産権が適切に保護されるよう検討すべきものと考えている。

特にゲノム機能の集中的解析によって得られたデータについては、参加機関外に将来的に公開は行うものの、公開の具体的な時期及びその方法等、今後、中央推進機関等において検討することとしたい。(参考資料6)

## 9.生命倫理 教育について

(問) 米国のヒトゲノムプロジェクトでは ELSI(Ethical Legal and Social Implications)に予算の4 - 5%を割いてきた。また、本年度からはじまるポストゲノムビジョンの3本柱のひとつは、Genome to Health となっている。米国に比べてわが国のヒトゲノム予算における生命倫理・教育に回される予算は極端に少ない様に思われるが、これらの関連予算の文科省、及び他省庁における予算、検討状況などを示されたい。

(回答)

ヒトゲノム予算あるいは当該ゲノムネットワーク研究においては、現在のところ生命倫理・教育に対する経費は計上されていないが、国の指針、他のプロジェクトにおける生命倫理等に対する取り組み等にならって、今後とも引き続き検討し、シンポジウムの開催等による情報の公開に努めたい。

また、生命倫理に関しては科学技術振興調整費を活用し、人材養成の取り組みを行っている。(参考資料7)

## 10.アウトプット・アウトカムについて

(問)「ヒト・ゲノム解析」の結果が人々の世界観や哲学に大きな影響を与えたことは明らかである。では「ゲノムネットワーク研究」は、一般社会の人々まで、知的興奮を呼び起こすような、どのようなアウトカム(社会的、産業的波及効果)が期待できるのか。それらのアウトカムを社会が摩擦なく享受するためにプロジェクトの中でどんな配慮が行われる計画か。「ゲノムネットワーク研究」の成果に直接リンクする社会貢献についてどのようなものが想定されるか示されたい。

(問)日本が国費を使って、「ゲノムネットワーク研究」を行うことの意義、得られる国益について示されたい。また、得られた成果はどのような形で公開し、もしくは公開せず、日本の利点、国益とするのか示されたい。

(回答)

本プロジェクトによって、ヒト、マウスのDNA-タンパク質、タンパク質-タンパク質間相互作用の基礎的なデータベースが構築される。

ゲノムネットワーク研究は各種遺伝子と生命現象の関係とそのメカニズムの解明に直結するものであり、ここで創出される体系的かつ精度の高いデータは各々の生命現象の研究にとって重要な基盤データとなり、生命全体をシステムとしてとらえる生命科学の新しい枠組みを生み出すことが期待される。

このことにより、国民健康の増進、福祉の向上等に大きく寄与することが期待されており、具体的には病態に関連する遺伝子やタンパク質等の生体分子の相互作用を明らかにすることにより、新しい治療法や薬の標的分子を得られ、ゲノム創薬のさらなる推進が可能となることも考えられる。

なお、成果については最終的に公開を行うものの、当面の間は得られたデータについて、外部に対して守秘義務を負う契約を、プロジェクト参加にあたって締結することにより知的財産権を確保し、ひいては我が国の国益を確保することとしている。

平成16年度大規模新規研究開発評価  
第2回評価検討会提出資料

「ゲノムネットワーク研究」  
府省への質問事項回答  
(参考資料)

平成15年10月1日  
文部科学省

参考資料1及び2は、未発表の研究内容、または知的所有権に関わる内容を含むため、一般には非公開とする。

# ゲノムネットワーク研究の構成

## ゲノム機能情報の集中的解析

### 発現調節領域の機能解析 (タンパク質-DNA相互作用等)

- ヒトと実験生物との間での比較ゲノム解析
- 転写開始点及び発現制御エレメントの検索

### 組織・細胞別の遺伝子発現解析

- 各種臓器、細胞、発生過程の組織・細胞における遺伝子の発現解析
- 病変生物を用いた疾患関連遺伝子の発現解析

### タンパク質-タンパク質相互作用

- タンパク質間の結合を網羅的に検索

### 次世代ゲノム解析技術開発

- 遺伝子プロファイリング、トランスクリプトーム解析、細胞内局在等に  
関する新たな技術の開発

## 個別生命機能解析

- |  |                                      |                                |   |
|--|--------------------------------------|--------------------------------|---|
| ゲノムの<br>徹底解析等<br>遺伝子の特定<br>領域の徹底解析<br>など | 生命現象等<br>脳の発生、概日<br>リズム、幹細胞<br>の分化など | 薬の標的<br>分子等<br>病変関連<br>タンパク質など | 疾患関連<br>遺伝子等<br>がん、高血圧<br>糖尿病、免疫系<br>など |
|--|--------------------------------------|--------------------------------|---|

## 統合データベース

ゲノム機能情報の集中的解析及び個別ネットワーク解析の結果得られたデータの**アノテーション**(注釈付け)を通じて、相互の関連づけを行い、ゲノム研究や、代謝マップ、解剖学等様々な視点に基づき**情報利用を可能にし、あらゆるライフサイエンス研究の効果的推進に資する。**

プロジェクト内の迅速なデータ流通  
(実施機関外部に対して守秘義務)

## 独 理化学研究所において実施

平成16年度概算要求額 35億円  
(運営費交付金により対応するものを含む)

### ゲノム機能情報集中的解析

大規模な解析施設を有する理化学研究所において、網羅的な解析を集中的に実施。

## ゲノム機能解析等の推進 《提案公募》

平成16年度概算要求額 35億円

### ゲノム機能情報の解析 《10億円》

各機関における研究ポテンシャルを活かして、特徴のあるゲノム機能の解析を実施。  
3~5年間の解析規模の提案を受け、最も能力が高い機関を選定。

### 次世代ゲノム解析技術開発 《10億円》

現在の技術を遥かに凌駕するようなネットワーク解析技術(遺伝子プロファイル等に関する新たな技術等)の開発  
3年間での実用化を目指し、公募で選定。

### 個別生命機能解析 《15億円》

個別の生命現象に焦点を当てたネットワーク解析が対象。  
国際的な有意性の高いもの、実用化において重要な意義を持つ等の研究課題を採択。

## 統合データベースの構築

平成16年度概算要求額 10億円

ゲノム機能情報及びゲノムネットワークに関する情報を総合したデータベースの構築

研究全体のコーディネイト/推進を図る中央推進機関を設置

## 各種タンパク質間相互作用解析法の対比

方法名	哺乳細胞2 - ハイブリッド法	酵母2 - ハイブリッド法	タンパクチップ	タグタンパクでのコンプレックス解析
方法の概要	培養細胞に2つの融合タンパク質を発現させ、融合タンパク質間の相互作用があった場合に発現するレポーター遺伝子の活性を測定する。	酵母細胞に2つの融合タンパク質を発現させ、融合タンパク質間の相互作用があった場合に発現するレポーター遺伝子によって生存した酵母細胞をコロニーとして検出する。	発現させたタンパク質 (多数) をガラススライドに貼り付けたチップを用意し、ラベルしたタンパク質をふりかけ、結合した相手方タンパク質をラベルで検出する。	タグ付きタンパク質を特定の細胞に発現させた後に細胞を溶解し、タグに対する抗体で沈降させた結合複合体をマスマスペクトロメトリーで解析する。
長所	現在のところ最も感度の高い相互作用アッセイ法である。 ・哺乳細胞を用いたアッセイであるため、他の方法と比較して、より生理的な相互作用を検出する。 ・レポーター活性を数値化することにより、相互作用の信頼性を見積もることが可能。 ・サンプル調製から相互作用の検出まで最短2日で実行できる。 一過性のタンパク質発現を利用するため、多少の細胞毒性がある遺伝子もアッセイ可能。	現在のところ最も感度の高い相互作用アッセイ法である。 ・片方の融合タンパク質を発現する酵母細胞を準備すれば、交配 (mating) を利用してアッセイを行うことができる。 ・相互作用する相手方タンパク質をDNAシーケンスで同定できる。	・ひとたびチップを作成したら、一度に多くのタンパク質をスクリーニングすることが可能。 ・細胞毒性のあるタンパク質間の相互作用を調べることが可能。	・タンパク質複合体の構成タンパク質を一度に解析することが可能。
短所	転写活性化機能を持つタンパク質同士の相互作用を全長タンパク質を用いて解析できない。	・特定ベクター上に遺伝子をくみこんだコンストラクトが必要。 ・アッセイ結果が出るまでに1週間程度かかる。 ・酵母細胞の増殖に影響を与える遺伝子、毒性のある遺伝子のアッセイはできない。 ・転写活性化機能を持つタンパク質同士の相互作用を全長タンパク質を用いて解析できない。	・スライドに貼り付けるためのタンパク質を調製するのが非常に困難。 ・チップを安定して保存するのが困難。 ・人工的な細胞外環境での結合実験である。	個々のタンパク質間の相互作用はわからない。 ・発現量の少ないタンパク質は感度の問題により解析できない。 ・結合する相手方は培養細胞で天然に発現しているタンパク質であり、細胞により発現の有無があるため、多くの培養細胞での試行が必要である。 ・結合強度の推定が困難。
技術補完性	タグタンパクでのコンプレックス法による複合体データと組み合わせると複合体の構成タンパク質と構築様式がわかる。	タグタンパクでのコンプレックス法による複合体データと組み合わせると複合体の構成タンパク質と構築様式がわかる。	タグタンパクでのコンプレックス法による複合体データと組み合わせると複合体の構成タンパク質と構築様式がわかる。	哺乳細胞2 - ハイブリッド法、酵母2 - ハイブリッド法、タンパクチップ法などの1対1結合データと組み合わせると複合体の構成タンパク質と構築様式がわかる。

