

平成 16年度大規模新規研究開発評価
第 1回評価検討会提出資料

ゲノムネットワーク研究の戦略的推進

平成 15年 9月 16日
文部科学省研究振興局
ライフサイエンス課

A. プロジェクトの概要

1. 名称、担当課室名、期間

名称	ゲノムネットワーク研究の戦略的推進
主管課	(主管課) 文部科学省研究振興局ライフサイエンス課 (課長: 戸谷一夫)
期間	5年間(平成16年度~平成20年度) 但し、開始後3年目に中間評価を実施。

2. 予算

(1) 総額

総事業費	平成16年度 80億円 (400億円/5年間)
国の負担額	平成16年度 80億円 (400億円/5年間)

注: 国以外については、共同研究等により参加を見込むが、金額については未定。

(2) 事業別予算額

ゲノム機能情報集中的解析	平成16年度 35億円 (175億円/5年間) (調整中)
ゲノム機能解析等の推進	平成16年度 10億円 (50億円/5年間) (調整中)
次世代ゲノム解析技術開発	平成16年度 10億円 (50億円/5年間) (調整中)
個別生命機能解析	平成16年度 15億円 (75億円/5年間) (調整中)
統合データベースの構築	平成16年度 10億円 (50億円/5年間) (調整中)

3. 目的: 背景と目指す方向

国際ヒトゲノム計画の達成にともない、ゲノム構造に関わる基盤的データが体系的に蓄積整備されつつあるなかで、ゲノム研究の方向性は世界的に機能解析へと向かいつつある。ゲノムの機能解析の成果は、ライフサイエンスのあらゆる研究の推進にあたって、重要な支柱となるものであり、一方、産業構造の改革及び国民の健康的な生活に重大な影響を及ぼすことから、我が国としては、我が国の研究資源等の強みを活かしながら、国際的動向を睨みつつ戦略的に取り組んでゆく必要がある。

国際的動向としては、本年4月に米国がENCODE計画としてヒトゲノムの全機能解析へ向けた計画を開始している。本プロジェクトが本格的に実施された場合には、国際ヒトゲノム計画に匹敵する規模のプロジェクトとなることが想定されるが、米国は当面自国のファンドのみによる計画とし、プロジェクトの参加者に対して守秘義務を課し、外部に対し一定期間その成果を公開しないこととしている。

これまで、発生等の基本的生命現象、疾病のメカニズムの解明、薬効等のメカニズム等について、これらを解明する試みが様々に行われている。これらの解明により、これらの複雑な生命現象のシステムの中で極めて効果的な分子標的が特定され、その成果は強力な知的財産権ともなり得るものである。このため、我が国としては、これらの研究を強力に推進していく必要がある。ヒトゲノム解読の完了後、これらの個別的研究 (small scale research) は全ゲノムを対象とした大量・包括的なデータ (large-scale science) との連携により実施されるようになってきており、このことは米科学アカデミーでも認識されている。従ってこのような研究を加速し、効果的に進めるためには、集中的に解析された網羅的な機能解析データの創出との有機的連携体制の構築が不可欠である。また、具体的には、遺伝子発現調整領域等のヒトゲノムの機能、あるいは、タンパク質 - タンパク質相互作用の解明等生体分子の相互作用の基礎データの網羅的創出が重要となっている。

これらの基礎データの創出にあたっては、ヒトあるいはマウスなどのヒトのメカニズムを研究する上で強力なツールとなる動物の完全長 cDNA クローンライブラリー、タンパク質の相互作用解析技術、プロモーターの構造情報などが重要となるが、我が国は他国と比較し、この分野で進んでおり、特定のヒトゲノムの機能あるいは生体分子の相互作用については、直ちに網羅的に解析を行う技術的基盤もある。

このため本プロジェクトは、今後のポストゲノムシーケンシング研究の一つのあり方として、個別生命現象のネットワークの解明、画期的な創薬の探索等の面においてヒトゲノム解読の成果を活用することを目指し、現在アメリカで進められている ENCODE 計画と異なり、ヒトゲノム機能の解明はネットワーク解明等に直接的に結びつくものについて、我が国の利点を生かし集中的に実施するとともに、ネットワーク解明等に資する生体分子相互作用等のデータについても技術的に可能なものについて集中的に解析を行おうとするものである。

また、その進め方に当たっては、提案公募により採択された新たな解析技術の開発も同時併行的に行い、厳格な評価の下に、常に最適の技術により可及的速やかに基盤データの創出を図ることを目指すものである。

更に、本プロジェクトにおいては、国際的優位性の高いもの、実用化において重要な意義を有するものの中から網羅的データの活用により、特に画期的な成果を見込み得る個別の生命機能の解明を行う研究を提案公募により実施し、集中的データの創出との有機的連携による、効果的な生命研究のシステムの構築をも目指すものである。

また、これらの情報をヒトゲノムを機軸とする様々な情報と有機的に関係付け、高度かつ有用なデータベースを構築し、今後のこの分野の研究の一層の発展に資することとする。

4. 本施策の位置付け

「平成16年度の科学技術に関する予算、人材等の資源配分の方針」(平成15年6月19日、総合科学技術会議)において、「ゲノム、タンパク質、糖鎖等の構造・機能及びそれらの形成するネットワークの解析とこれに必要な基盤的データベースの整備」はライフサイエンス分野の重点事項として位置づけられており、第30回の総合科学技術会議(平成15年7月23日)においても、「ライフサイエンス分野研究の新展開」に特に重視すべき分野として、「ゲノムネットワークの解明」が位置づけられている。

本施策は、このような総合科学技術会議の方針を踏まえつつ、科学技術・学術審議会ライフサイエンス委員会において取りまとめられた「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」を具体化したものである。

5. 目標：具体的な到達点や期待される成果等

疾患等も含めた個別の生命現象を扱う研究と、網羅的ゲノム機能解析データの創出を行なう機関との有機的連携関係を構築し、我が国における新しい研究スキームを樹立する。

得られたデータは統合データベースにおいて相互の関連づけ(アノテーション)を行うことにより、様々な視点に基づくデータ利用を可能とし、ライフサイエンス研究の効果的推進に資することとなる。例えば、病気の原因遺伝子から発症に至るまでに関連する遺伝子やタンパク質の相互作用が明らかとなり、効率的なゲノム創薬及び画期的な病気の治療法の開発を行うことも可能となる。

このことによって、我が国の優位性の確立、維持、発展を図ることを目指すこととする。

6. 内容：個別の科学技術的な課題と研究開発の方法、課題毎の実施時期等

プロジェクトの目標	研究開発目標	研究開発方法	実施時期
ゲノム機能情報の集中的解析	A) 発現調節領域の機能解析 1) 転写開始点及び発現制御エレメントの検索 ハイスループット転写開始点解析法 どの組織、どのステージでどのプロモーター、どの転写開始点がいられるのかを同定する	1) ハイスループット転写開始点解析法 100,000,000Tagをゲノム上にマッピングする。	H16～H20

<p>ゲノム機能情報の集中的解析 (続き)</p>	<p>2) 転写調節因子間のカスケード解析</p> <ul style="list-style-type: none"> すべての転写調節因子のリストアップ 約 1000 個の unique な転写調節因子遺伝子を単離 上記の転写因子間のカスケード解析 転写調節因子の転写増強 転写減弱による下流遺伝子の同定 <p>3) 転写調節因子結合点の探索</p> <p>4) ゲノム保存領域探索による発現調節解析領域</p> <p>B) 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探索</p>	<p>2a) One Hybrid assay</p> <ul style="list-style-type: none"> FANTOM2 クローンセットの活用 どの転写因子がどの転写因子を調節するかという情報、つまりすべての転写因子間の転写カスケードを同定する <p>2b) マイクロアレイと転写因子トランスフェクション・ノックアウト</p> <p>3) chin on chip 転写調節因子をゲノム DNA に結合させたあとゲノム DNA と架橋後、抗体で沈殿させる。これらを microarray に供する。</p> <p>4) 多数の遺伝子間及びヒトと実験生物との種間での比較ゲノム解析</p> <p>Two Hybrid system</p> <p>細胞内で 2 つのタンパクが結合すると、レポーター遺伝子が働き蛍光を測定できる。 すべてのタンパク質間の相互作用を網羅的に解析できる。</p>	<p>H16 ~ H20</p>
-------------------------------	--	---	----------------------

<p>ゲノム機能解析等の推進(提案公募型の課題採択を行う)</p>	<p>ゲノム機能情報の解析</p>	<p>各機関より提案される特徴のあるゲノム機能解析を実施(課題例)</p> <ul style="list-style-type: none"> ・ 特定ゲノム領域の徹底した機能解析(米国 ENCODE 型研究) ・ ヒトゲノムの難解読領域の機能解析 ・ 各種細胞、臓器の遺伝子、タンパク質発現プロファイリング(標準データの確立) ・ 単一細胞遺伝子プロファイリングによる機能発現多様性の解析 ・ ゲノム機能解析のための siRNA ライブラリーの構築 ・ スプライシングバリエーションの徹底解析 ・ 修飾タンパク質の相互作用解析 ・ 我が国の実績を生かした酵母ツーハイブリッド系による疾患関連タンパク質相互作用ネットワークの解明 <p style="text-align: right;">など</p>	<p>H16 ~ H20</p>
	<p>次世代ゲノム解析技術の開発</p>	<p>遺伝子プロファイリングや、タンパク質相互作用解析法等に関する新たな技術開発を実施</p> <ul style="list-style-type: none"> ・ 絶対量測定型 DNA チップの開発 ・ In vitro virus 法を用いたタンパクチップの開発 ・ タグ付き複合体による in vivo タンパク相互作用の解析法 	<p>H16 ~ H20</p>

<p>ゲノム機能解析等の推進(提案公募型の課題採択を行う)(続き)</p>	<p>次世代ゲノム解析技術の開発(続き)</p>	<ul style="list-style-type: none"> ・ 標識ナノ粒子を用いたタンパク質分子間相互作用解析法 ・ マルチレーザー共焦点顕微鏡によるタンパク質細胞内局在解析法 ・ エネルギー転換を利用したタンパク質動態解析法 <p style="text-align: right;">など</p>	<p>H16 ~ H20</p>
	<p>個別生命機能解析</p>	<p>発生、幹細胞の分化、薬の標的分子、がん、高血圧等の疾病関連遺伝子等で、国際的な有意性の高いもの等を研究</p> <ul style="list-style-type: none"> ・ 脳発生のネットワーク解析 ・ 幹細胞分化の分子ネットワーク解析 ・ サーカディアンリズムのネットワーク解析 ・ 神経・記憶の分子ネットワーク ・ がん、免疫、糖尿病など疾患分子ネットワーク ・ 抗うつ剤、抗がん剤、環境ホルモンなど外的因子に対応する分子ネットワーク <p>など、多数のグループの応募が想定される。</p>	<p>H16 ~ H20</p>
<p>統合データベースの開発</p>	<p>ゲノム機能情報の集中的解析、個別生命機能解析等で得られたデータの有機的連携</p>	<ul style="list-style-type: none"> ・ 得られたデータのアノテーション(注釈付け)及び要素結合による、多様なニーズに応えられるデータベースの開発 	<p>H16 ~ H20</p>

(参考)

「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」における基盤データ創出のための技術項目の評価

1. 現時点において技術的に実行に移せる項目

「発現制御領域解析」、「トランスクリプトームの徹底解析」、「RNAレベルでの発現プロファイル」、「タンパク質 - タンパク質相互作用(非修飾タンパク)」

2. 一部技術開発を伴いながら、大規模化への取り組みが可能なもの

「プロテオーム解析」、「タンパク質 - タンパク質相互作用(リン酸化、脱リン酸化及びその他の修飾)」等

3. さらなる技術開発が優先されるもの

「より高度化されたタンパク質の相互作用(Ub化、SUMO化、膜タンパク質の相互作用)等

7. 実施体制：実施機関、組織、推進委員等

ゲノム機能情報の集中的解析及び統合データベースの開発については、大規模解析技術、設備、人材等の十分なリソースを持ち、解析・収集データを提供する能力と意欲があることが必要であり、現在のところゲノム機能情報の集中的解析については、理化学研究所、統合データベースの開発については国立遺伝学研究所を想定している。

また、ゲノム機能解析等の推進については、提案公募型の研究を予定しており、各機関における研究ポテンシャルを活かした研究推進を行う予定である。

これらの個別の研究をコーディネートし、研究全体の推進を図るためにはプログラム内に強力な中央推進組織が不可欠であり、プロジェクトの進捗状況の把握、提案公募課題の事前評価、新技術開発の成果の評価、集中解析等を行なう中核機関の事業内容の評価等を行うこととする。

8. 研究者：プロジェクトリーダーと主要研究者

プロジェクトリーダー候補者と略歴

候補者名	(調整中)
所属、役職	(調整中)
略歴	(調整中)

主要研究者・候補者と略歴

候補者名	榎 佳之
所属、役職	理化学研究所ゲノム科学総合研究センターゲノム構造情報研究グループプロジェクトディレクター
略歴	1971年 東京大学大学院理学系研究科博士課程修了 1971年 米国カリフォルニア大学ウィルス研究所研究員 1973年 三菱化成生命科学研究所副主任研究員 1981年 九州大学医学部附属遺伝情報実験施設助教授 1985年 九州大学遺伝情報実験施設教授

	1992年 東京大学医科学研究所ヒトゲノム解析センター教授 (～現在) 1998年 理化学研究所ゲノム科学総合研究センター ゲノム構造情報研究グループ プロジェクトディレクター
--	---

候補者名	林崎 良英
所属、役職	理化学研究所ゲノム科学総合研究センター遺伝子構造・機能研究グループ プロジェクトディレクター
略歴	1986年 大阪大学医学部大学院医学研究科内科系博士課程卒業 1988年 国立循環器病センター研究所バイオサイエンス部研究員 1992年 理化学研究所ライフサイエンス筑波研究センター ジーンバンク室研究員 1994年 理化学研究所ライフサイエンス筑波研究センター ゲノム科学研究室主任研究員 1998年 理化学研究所ゲノム科学総合研究センター 遺伝子・機能研究グループ プロジェクトディレクター (～現在) 2001年 スウェーデン王立カロリンスカ研究所客員教授兼務 (～現在)

候補者名	五條堀 孝
所属、役職	国立遺伝学研究所生命情報・DDBJ 研究センター長
略歴	1979年 九州大学大学院理学研究科博士課程修了 1980年 アメリカテキサス大学ヒューストン校上級研究員 1983年 国立遺伝学研究所生理遺伝部門研究員 1988年 国立遺伝学研究所集団遺伝研究系助教授 1990年 国立遺伝学研究所遺伝情報研究センター教授 1990年 総合研究大学院大学生命科学研究科教授併任(～現在) 1994年 国立遺伝学研究所遺伝情報研究センター教授 2001年 国立遺伝学研究所生命情報・DDBJ 研究センター長 (～現在) 2001年 (独)産業技術総合研究所生物情報解析研究センター 副センター長(～現在)

候補者名	岩柳 隆夫
所属、役職	(株)日立製作所 ライフサイエンス推進事業部CTO
略歴	1977年 東京大学工学系応用化学大学院博士課程修了 1977年 (株)日立製作所中央研究所入所

	1998年 (株)日立製作所基礎研究所副所長
	1999年 (株)日立製作所中央研究所副所長
	1999年 (株)日立製作所ライフサイエンス推進事業部CTO

9. 評価体制：委員会名簿、予定評価時期等

中央推進組織において、プロジェクトの進捗状況の把握、提案公募課題の事前評価、新技術開発の成果の評価、集中解析等を行なう中核機関の事業内容の評価等を行うこととする。またプロジェクト開始3年目(平成18年度)には中間評価として、ライフサイエンス委員会等、外部の有識者委員会を活用して、提案公募型の研究については、それ以降の継続を検討し、中核機関における研究についてはその研究体制等の検討を行う。

(中央推進組織の委員名簿については調整中)

10. その他：既存プロジェクトとの関係、府省連携、産学連携等

既存プロジェクトとの関係

府省名	予算規模	役割分担
文部科学省	778.7億円 (平成14~19年度までの総事業費)	「タンパク3000プロジェクト」は、タンパク質の基本構造等を解析するプロジェクトであり、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に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、具体的にそのタンパク質が機能するか否かに影響を与える物質(すなわち薬)の設計を行うことができる。このような化学物質を得ることはゲノム創薬の大幅な進捗につながる。
文部科学省	192.9億円 (平成14年度補正予算から平成19年度までの総事業費)	「テーラーメイド医療実現化プロジェクト」は、SNP(一塩基多型)解析を通じ、個人の遺伝情報と疾患、薬剤応答性との関係の解析を行うプロジェクトであり、この過程で疾患遺伝子に関係する情報も多く得られる。この情報をゲノムネットワーク研究に提供することにより、ゲノムネットワーク研究では、疾患等関係遺伝子同士の関係を調べるのが可能となる。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ただけでなく、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出来る。

B．府省における考え方

国際関係上の意義

平成14年4月に国際ヒトゲノム計画が達成されたことにより、ゲノム研究は機能解明を中心とした本格的な国際競争の時代に入った。ゲノムの機能解析の成果は、我が国の知的財産権の保護にもつながることから、ゲノム研究の進め方については、国際競争の側面を十分に認識する必要がある。このような流れの中で、アメリカは、NIHを中心にゲノム塩基配列情報上の機能部位を網羅的に同定するENCODE（Encyclopedia of Human DNA Elements）計画を含む国家プロジェクトを公表した。この計画は全ゲノム上の機能アノテーションを目標とするものであり、当面の三年間はパイロットプロジェクトとしてスタートする。

本プロジェクトにおいては、このような国際的に激化する知的財産取得に関する競争において優位に立つことを目標として、ヒト、マウスcDNAライブラリー、高速塩基配列決定設備などのノウハウ等、我が国独自の研究ポテンシャルを活用した研究を展開する必要がある。

社会経済上の意義・効果

得られた研究成果はライフサイエンス分野における様々な研究開発の共通データとして提供されるようになる。また、疾患等の原因遺伝子から発現に至るまでに関係する遺伝子やタンパク質の相互作用が明らかになることにより、新薬の効果的な創出に資する。

これにより、疾患の原因から発症までのメカニズム解明が実現し、高齢化社会を迎える国民の健康な生活が保持できるとともに、医薬品産業を通じた経済の活性化が可能となる。

運営の効率性

本プロジェクトにおいては、ゲノム機能情報の集中的解析を進めるとともに、解析技術の高度化及び実用化を行い、集中的解析に移行させる。これにより解析の効率化を図ることを目指す。

また、ゲノム解析の技術開発、個別的生命機能の解析等については、提案公募型とし、プロジェクトの推進にあたっては、中間評価により、体制を見直す。

平成16年度大規模新規研究開発評価
第1回評価検討会提出資料

ゲノムネットワーク研究の戦略的推進 (参考資料)

平成15年9月16日
文部科学省研究振興局
ライフサイエンス課

ゲノムネットワーク研究の戦略的推進 - 塩基配列解読から機能解明へ -

平成16年度概算要求額 80億円
(運営費交付金中の推計額を含む)

国際ヒトゲノム 計画の達成

(平成15年4月14日)

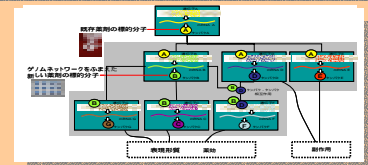


ゲノムの構造(塩基配列)が解読され、
今後はその機能の解明へ

ゲノム研究は
機能解明を中心とした
本格的国際競争
の時代に突入

米国: ENCODE計画発表
ヒトゲノムの全機能解明へ

ゲノムネットワークとは



生命現象を表現する様々な遺伝子や生体分子の相互作用を統合することによって明らかになる、生命の統合的なシステムのこと。

研究開発の目標

ヒトを対象として個別の生命現象の分子ネットワークを統合し、生命をひとつの統合したシステムとして包括的に解明するためのフレームの構築を目指し、ヒトゲノム機能等の基本的な情報の創出、より高次の情報を得るための技術や方法論の開発を行う。

ゲノムネットワーク研究推進方策

我が国の強みを活かす研究

ヒト及びマウスcDNAライブラリー、高速塩基配列決定設備やノウハウなどのリソースの活用

効果的な研究推進体制(バイオプラットフォーム)の構築

集中的なゲノム解析とその各種疾患等の個別のネットワーク研究との有機的連携を確保

集中的解析の実施と平行して、ゲノムネットワーク解析のための新規の技術開発を実施

期待できる成果

病因から発症までのメカニズム解明

新たな治療法
創薬の開発

健康な生活
の実現

経済活性化
の実現



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

ゲノムの徹底解析等
遺伝子の特定
領域の徹底解析
など

生命現象等
脳の発生、概日
リズム、幹細胞
の分化など

薬の標的
分子等
病変関連
タンパク質など

疾患関連
遺伝子等
がん、高血圧
糖尿病、免疫系
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**様々な視点に基づく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)

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。

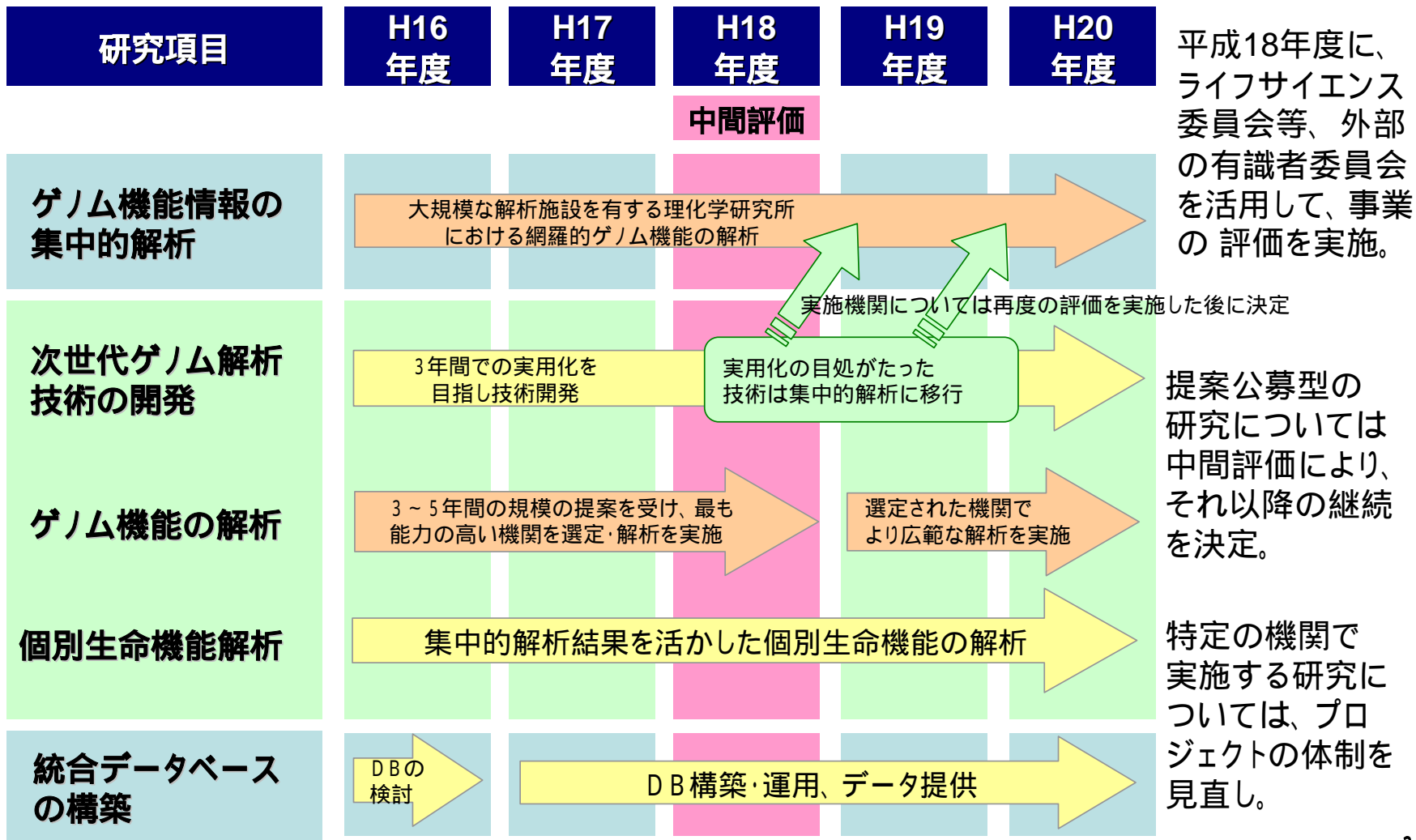
統合データベースの構築

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ート/推進を図る中央推進機関を設置

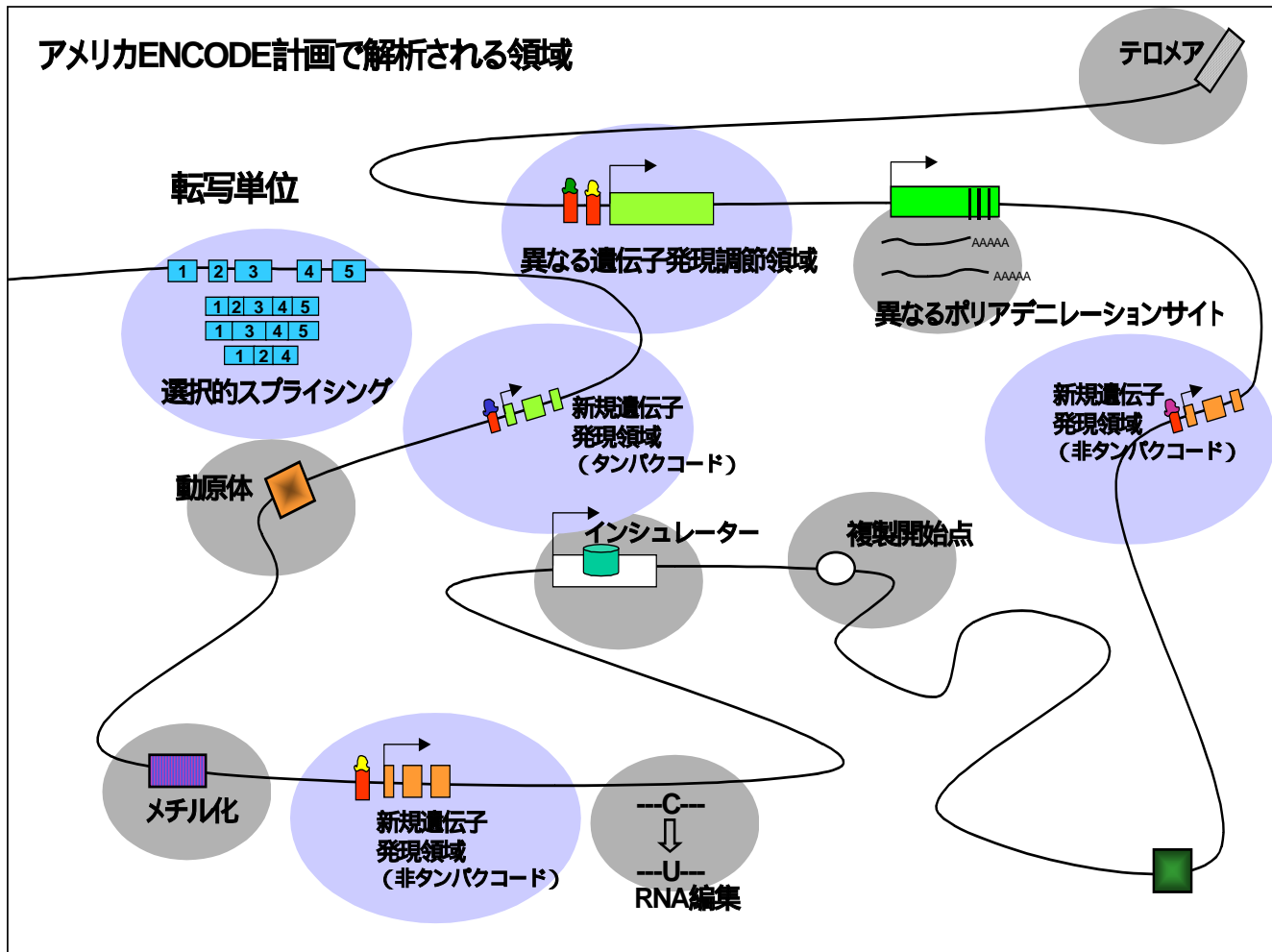
ゲノムネットワーク研究の研究計画



本構想とENCODE計画の違い

	本構想	ENCODE計画
ヒトゲノムの解析対象	遺伝子、発現調節領域等ゲノムネットワークに直接関係する機能に限定	全機能 左の機能に加え、動原体、テロメア、複製開始点、メチル化サイト等
解析の規模	特定機能についてゲノム全体について解析	44領域について解析(ゲノム全体の1%) 3年間のパイロットスタディー後は、全ゲノムに対象を拡大予定
進め方	ネットワーク研究をプロジェクトの一部として推進し、相互の連携関係を構築しながら推進。	ネットワーク研究とは独立したプロジェクトとして推進

アメリカENCODE計画で解析される領域



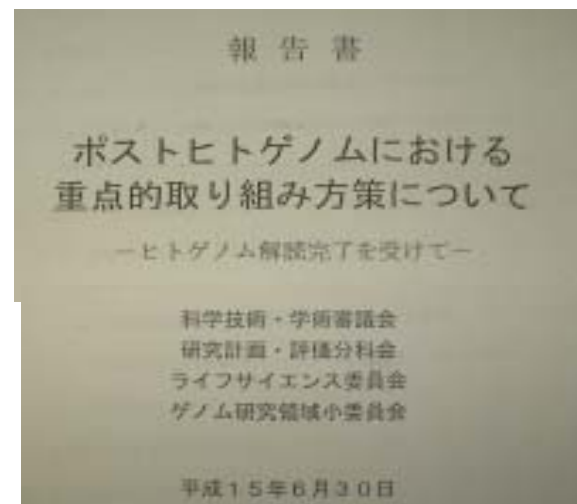
文部科学省ライフサイエンス委員会 ゲノム小委員会答申

- **基盤となるデータの創出**
技術的にすぐに実行に移せる
項目

- トランスクリプトームの徹底解析
- RNAレベルでの発現プロファイル
- タンパク質-タンパク質相互作用(非修飾タンパク)

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

さらなる技術開発が優先される
べきもの



わが国の有利な点

- トランスクリプトーム(プラットフォーム資源)が整備されている
- 技術が、すでに整備されている
- ゲノム小委員会答申から
 - 発現制御領域や新規RNA分子種の発見、収集など、ゲノムDNAの塩基配列を中心とした情報抽出については、緻密で網羅的な実験的アプローチや比較ゲノム解析など様々な手法がある。
 - 完全長cDNAは、わが国独自の技術として国際的な優位性があり、ヒト及びマウスのcDNAに関しては、我が国は世界に先駆け、総合的・包括的・系統的に収集したcDNAバンクを創出した。
 - タンパク質-タンパク質相互作用の分野においては、我が国には完全長cDNAのセットが他にないリソースとして存在し、また、その包括的系統的解析においても系統的なデータを世界に先駆け創出した実績をもつ。

プラットフォーム資源

articles

Functional annotation of a full-length mouse cDNA collection

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium*

The RIKEN Mouse Gene Encyclopaedia Project, a systematic approach to determining the full coding potential of the mouse genome, involves collection and sequencing of full-length complementary DNAs and physical mapping of the corresponding genes to the mouse genome. We organized an international functional annotation meeting (FANTOM) to annotate the first 21,076 cDNAs to be analysed in this project. Here we describe the first RIKEN clone collection, which is one of the largest described for any organism. Analysis of these cDNAs extends known gene families and identifies new ones.

In mammals and higher plants, interpreting the genome sequence is not straightforward: coding regions are interspersed with noncoding DNA, and an individual gene may give rise to many products. Thus, genomic sequence cannot be reliably decoded to identify the spectrum of messenger RNAs (the transcriptome) and their corresponding protein products (the proteome). This problem is illustrated by the different estimates of the number of human genes (30,000, 35,000 and 120,000)^{1,2,3}. Although gene prediction programs have become more accurate and sensitive, the sequence of a full-length cDNA clone provides more reliable evidence for the existence and structure of a gene. The Mouse Gene Encyclopaedia Project aims to identify and sequence every transcript encoded by the mouse genome. Here, we report the characterization of our first cDNA set of 21,076 mouse clones (some of which are derived from the same transcripts).

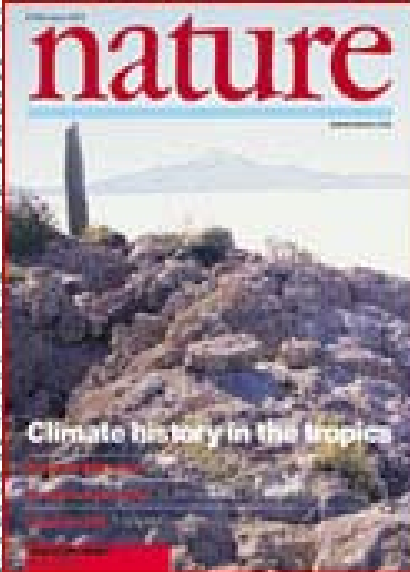
Strategies

In the first phase of the project, we prepared around 160 full-length, enriched⁴, normalized and subtracted⁵ cDNA libraries from various tissues and developmental stages. From these, we collected and clustered 930,000 3' ends and sequenced to produce about 128,000 groups that were targeted for sequencing.

In the second phase of the project, we selected a single clone from each cluster for sequencing. Preference was given to clones from libraries estimated to contain the highest representation of full-length transcripts. To expedite sequencing, we focused on relative short cDNAs (Fig. 1), which are probably biased in favour of truncated clones. To increase the likelihood of discovering new genes, we also biased our selection towards clones with novel 3' end sequences. We sequenced 21,076 cDNA clones, with average length 1,257 base pairs (bp); the longest clone sequenced was 6,327 bp (Supplementary Information Fig. 1A). All sequences have been registered in the public sequence database DDBJ, except for 13 cDNAs assembled using sequences from public expressed sequence tag (EST) databases (available at <http://genome.gsc.riken.go.jp/genome/fantom/viewer/est/>). We estimated using the PHRED⁶ base-calling program⁷ that the average accuracy of the sequences was 99.14% (72% (15,236 clones) of clones shorter than 1,000 bp and 6,739 sequences (32%) were determined by PHRED scores) (see Supplementary Information Fig. 1B).

We extracted the open reading frame (ORF) from each cDNA sequence using the RIKEN DECODER program⁸. The DECODER program corrected frame-shifts in 3,376 (16%) of the cDNAs. The likelihood that the sequence selected was a true ORF (score (1/4)ⁿ calculated in light of the Kozak consensus usage and position of the initiation codon) that a frame shift occurred was determined using PHRED scores.

Annotation of cDNAs



マウス完全長cDNA
Nature Vol. 409,
pp685-pp690,2001

Nature, 420, 520-562, 2002

articles

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team*

*A full list of authors appears at the end of this paper

Only a small proportion of the mouse genome is transcribed into mature messenger RNA transcripts. There is an international collaborative effort to identify all full-length mRNA transcripts from the mouse, and to ensure that each is represented in a physical collection of clones. Here we report the manual annotation of 60,770 full-length mouse complementary DNA sequences. These are clustered into 33,409 transcriptional units, contributing 96.1% of a newly established mouse transcriptome database. Of these transcriptional units, 4,259 are new protein-coding and 11,865 are new non-coding messages, indicating that non-coding RNA is a major component of the transcriptome. 41% of all transcriptional units showed evidence of alternative splicing. In protein-coding transcripts, 79% of splice variations altered the protein product. Whole-transcriptome analysis resulted in the identification of 2,431 sense-antisense pairs. The present work, completely supported by physical clones, provides the most comprehensive survey of a mammalian transcriptome so far, and is a valuable resource for functional genomics.

理研・マウスゲノムコンソーシアム
共同

articles

Nature, 420, 520-562, 2002

Initial sequencing and comparative analysis of the mouse genome

Mouse Genome Sequencing Consortium*

*A list of authors and their affiliations appears at the end of this paper

The sequence of the mouse genome is a key informational tool for understanding the complex experimental tool for biomedical research. Here, we report the results of an international collaborative effort to sequence the mouse genome. We also present an initial comparative analysis of the mouse genome. We describe some of the insights that can be gleaned from the two sequences. We discuss the evolutionary forces shaping the size, structure and sequence of the genomes; the conservation of the genomes; the much lower extent of sequence orthology covering less than half of the genomes under selection; the number of protein-coding genes; the expansion of gene families; the evolution of proteins; and the identification of intraspecific polymorphisms.





イネ完全長cDNA
28,000種
Science, 2003

Collection, Mapping, and
Annotation of Over 28,000 cDNA
Clones from *japonica* Rice

The Rice Full-Length cDNA Consortium

[National Institute of Agrobiological Sciences Rice Full-Length cDNA Project Team: Shoshi Kikuchi,¹ Kouji Satoh,¹ Toshifumi Nagata,² Nobuyuki Kawagashira,¹ Keiji Doi,¹ Naoki Kihimata,¹ Junshi Yazaki,¹ Masahiro Ishikawa,¹ Hiromi Yamada,¹ Hisako Ouka,¹ Inaeko Hotta,¹ Keiichi Kajima,² Takahiro Nareiki,² Eisuke Ohnoda,² Wataru Yahagi,² Kohji Suzuki,² Chao Jin Li,² Kenji Ohtsuki,² Teru Shikiki¹]; Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Yasuhiko Ohtera,³ Kazuo Murakami,³ Yoshiharu Iida,³ Saezumi Sugano,³ Tatsuro Fujimura,³ Yutaka Suzuki,³ Yuki Tsunoda,³ Takashi Karasaki,³ Takako Kodama,³ Hiromi Masuda,³ Michio Kobayashi,³ Qilong Xie,³ Min Lu,³ Ryuya Narikawa,³ Akio Sugiyama,³ Keiichi Mizuno,³ Satoko Yokemizo,³ Junko Nishura,³ Rieko Bada,³ Jinya Ishikawa,³ Hidetoshi Kawamura,³ Akemi Yoshimura,³ Junichiro Mura,³ Takahiro Kusumagi,³ Mitsuru Oka,³ Risa Iryo,³ Harika Ueda,³ Kenichi Matsukawa,³ RIKEN Jun Kawai,³ Piero Carninci,³ Jun Adachi,³ Katsunori Aizawa,³ Takahiro Arakawa,³ Shiro Fukuda,³ Ayako Hara,³ Wataru Hashidume,³ Norihito Hayata,³ Keiichi Imotani,³ Yoshiki Ishii,³ Masayuki Itoh,³ Rieko Kagawa,³ Shinji Kondo,³ Hideoaki Kuroki,³ Ai Miyazaki,³ Naoki Oono,³ Yoshinari Ota,³ Rintaro Saita,³ Daisuke Senaki,³ Kenjiro Seto,³ Kazuhiko Shibata,³ Akira Shinagawa,³ Toshiyuki Shiraki,³ Masayasu Yoshino,³ Yoshihide Hayashizaki³]

Functional Annotation of a Full-Length *Arabidopsis* cDNA Collection

Motoaki Seki,^{1,2} Mari Narusaka,¹ Asako Kamiya,¹ Junko Ishida,¹ Masakazu Satou,¹ Tetsuya Sakurai,¹ Maiko Nakajima,¹ Akiko Enju,¹ Kenji Akiyama,¹ Youko Oono,^{2,3} Masami Muramatsu,^{4,5} Yoshihide Hayashizaki,^{4,5} Jun Kawai,^{4,5} Piero Carninci,^{4,5} Masayoshi Itoh,^{4,5} Yoshiyuki Ishii,^{4,5} Takahiro Arakawa,^{4,5} Kazuhiro Shibata,^{4,5} Akira Shinagawa,^{4,5} Kazuo Shinozaki^{1,2*}



シロイヌナズナ完全長cDNA
Science 2002

タンパク質 - タンパク質相互作用

システムは既に確立 20000 wells/day

- * Harukazu Suzuki et.al, *Genome Research*, **10**, 1758-65, 2001
Protein-protein interaction Panel Using Mouse Full-Length cDNAs

データ処理システム (ノイズは消去)

- * Rintaro Saito et.al, *Bioinformatics*, **19**, 756-63, 2003
Constriction of reliable protein-protein interaction networks with a new interaction generality measures

データベース

- * Harukazu Suzuki et.al, *Genome Research*, **13(6b)**, 1534-41, 2003
The Mammalian Protein-Protein Interaction Database and its Viewing System That Is Linked to the Main FANTOM2 Viewer

タンパク質-タンパク質相互作用の応用例

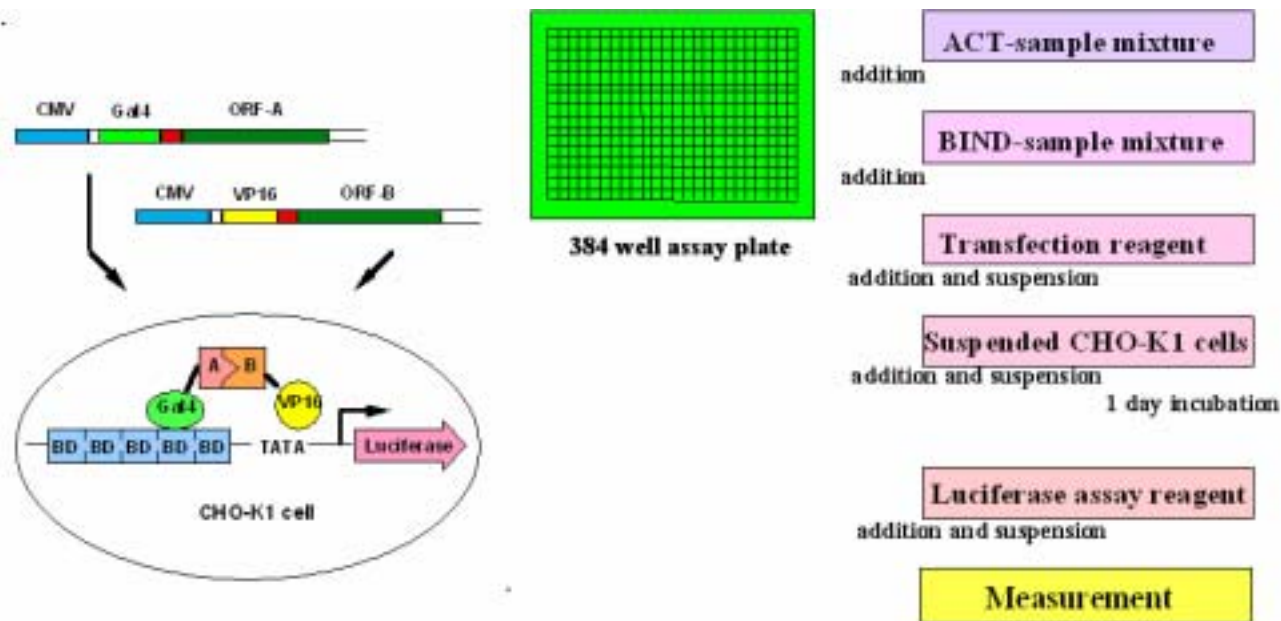
- * Mutsumi Kanamori et.al, *FEBS letter*, **532**, 241-46, 2002
NF- κ B activator Act1 associates with IL-1/Toll pathway adaptor molecule TRAF6
- * Mutsumi Kanamori et.al, *J Biol Chem*, in press
The PDZ protein TIP-1 inhibit β -catenin transcriptional activity and growth of colorectal cancer cells
- * Mutsumi Kanamori et.al, *Biochem Biophys Res Commun*, **290(3)**, 1108-13
T2BP, a novel TRAF2 binding protein, can activate NF- κ B Ap-1 without TNF stimulation

タンパク質-タンパク質相互作用システムの特徴

すべての試薬を加えるだけ
その後 well から発光を測定するのみ

Primer でタンパクの coding 領域を増幅

すべてのタンパクのコード領域のcDNAセットが既にそろっている



現在進行中のプロジェクト

ゲノム構造データベース

ゲノムシーケンスデータベース
完全長cDNAデータベース
タンパク3次構造データベース
SNPデータベース
その他

… 特許は構造情報のみの問題ではない

将来のプロジェクト

ゲノム機能データベース

発現制御領域
発現プロファイル
タンパク-タンパク相互作用
細胞内発現プロファイル
タンパク-DNA相互作用
タンパク-RNA相互作用
タンパク-RNA細胞内局在
タンパク-RNA細胞内定量的動態解析
ハイスループットin situ hybridization
バリエントトランスクリプトーム解析
Non-coding RNA 解析
Sense-Antisense RNA
など

… ゲノム機能はIP(知的財産)に近い、しかしそれだけでは特許性が弱い

ゲノムネットワークデータベース

ホルモン
神経伝達
細胞間信号伝達
細胞内信号伝達
転写ネットワーク

遠位(細胞外)
遠位(細胞外)
近位(細胞外)
細胞内
細胞内

… ゲノムネットワークはまさにIPに直結している。