

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)

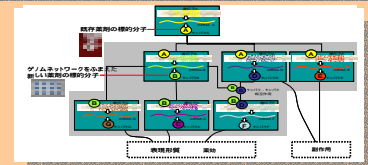


ゲノムの構造(塩基配列)が解読され、
今後はその機能の解明へ

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>

現在の技術を遙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。

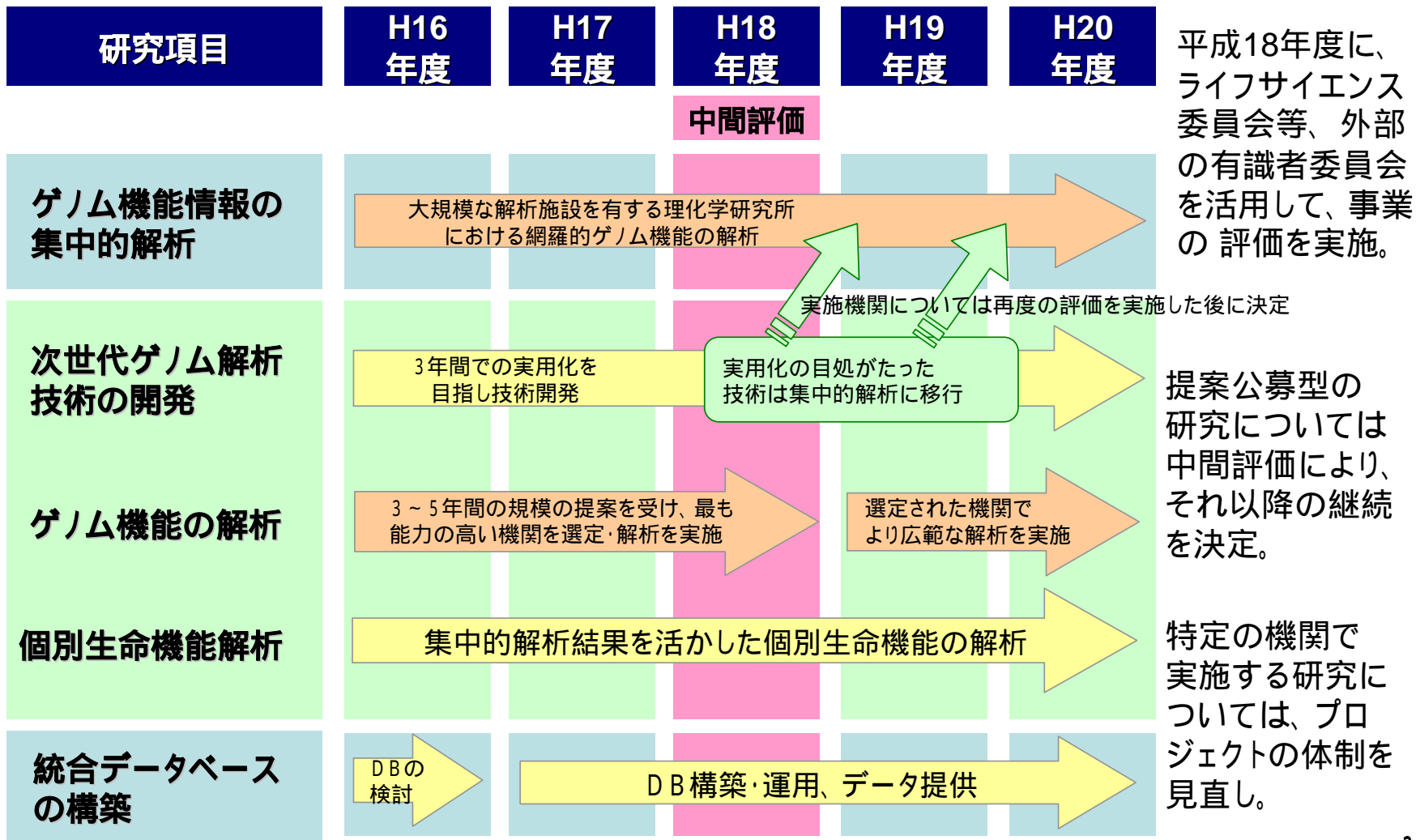
統合データベースの構築

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

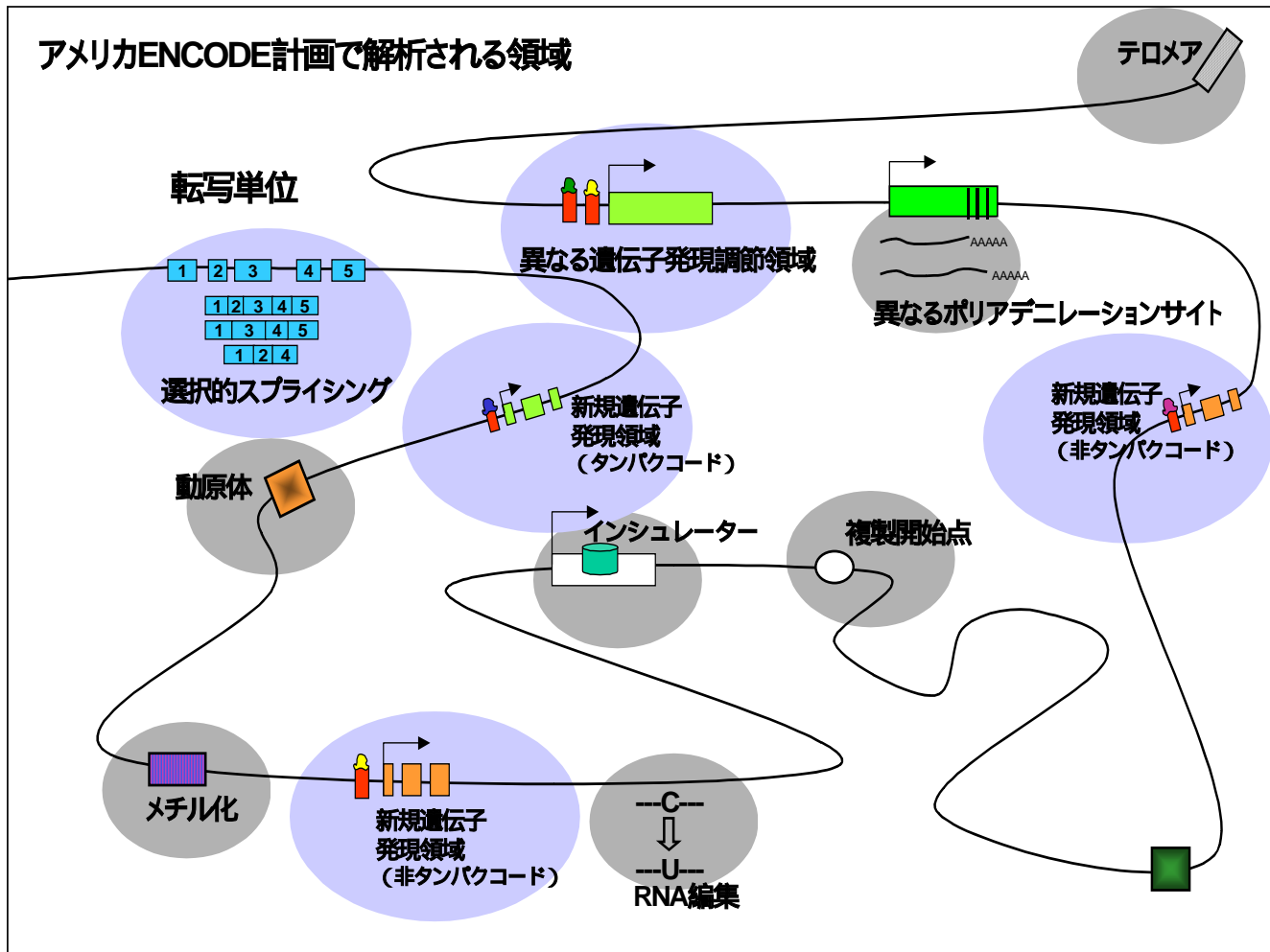
ゲノムネットワーク研究の研究計画



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



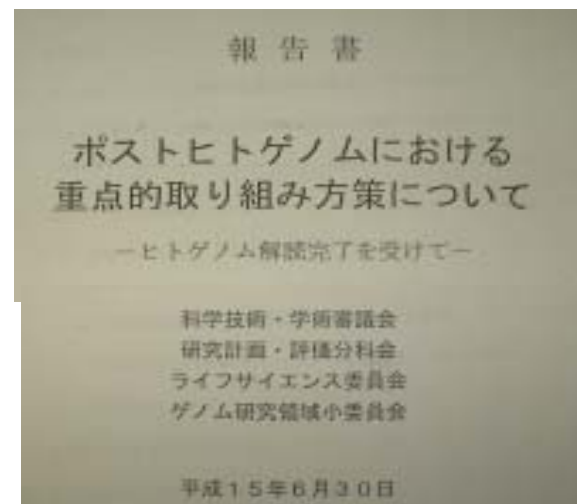
文部科学省ライフサイエンス委員会 ゲノム小委員会答申

- **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).

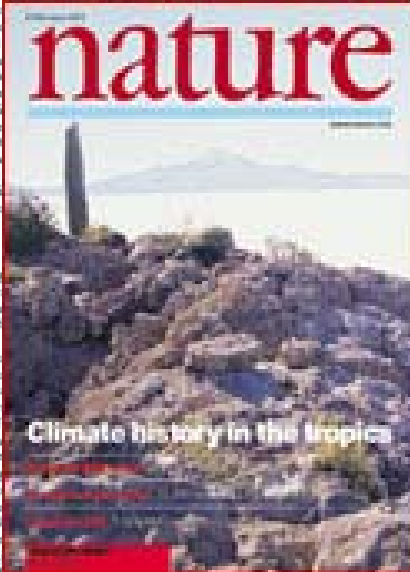
Strategies

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 program⁶ that the average accuracy of the 21,076 cDNA clones was 99.14%; 72% (15,236 clones) of clones shorter than 1,000 bp and 6,739 sequences (32%) were determined to be novel (see Supplementary Information Fig. 1B).

We extracted the open reading frame (ORF) from each cDNA sequence using the RIKEN DECODER program⁷. The DECODER program corrected frame-shifts in 3,376 (16%) of the cDNA sequences. The likelihood that the sequence selected was a true ORF (score $\geq 1/2$) calculated in light of the Kozak consensus usage and position of the initiation codon was determined. The likelihood that a frame shift occurred was determined (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, 419, 473-477, 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 describing 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 Ishikiri,³ Hidetoshi Kawamura,³ Akemi Yoshimura,³ Junichiro Mura,³ Takahiro Kusumagi,³ Mitsuru Oka,³ Risa Iryo,³ Harika Ueda,³ Kenichi Matsukura³]; RICE² Jun Kawai,⁴ Piero Carninci,⁴ Jun Adachi,⁴ Katsunori Aizawa,⁴ Takahiro Arakawa,⁴ Shiro Fukuda,⁴ Ayako Hara,⁴ Wataru Hashikura,⁴ Norihito Hayata,⁴ Keiichi Imotani,⁴ Yoshiki Ishii,⁴ Masayuki Itoh,⁴ Rieko Kagawa,⁴ Shinji Kondo,⁴ Hideoaki Kuroki,⁴ Ai Miyazaki,⁴ Naoki Oono,⁴ Yoshinari Ota,⁴ Rintaro Saita,⁴ Daisuke Senaki,⁴ Kenjiro Seto,⁴ Kazuhiko Shibata,⁴ Akira Shinagawa,⁴ Toshiyuki Shiraki,⁴ Masayasu Yoshino,⁴ Yoshihide Hayashizaki⁴]

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

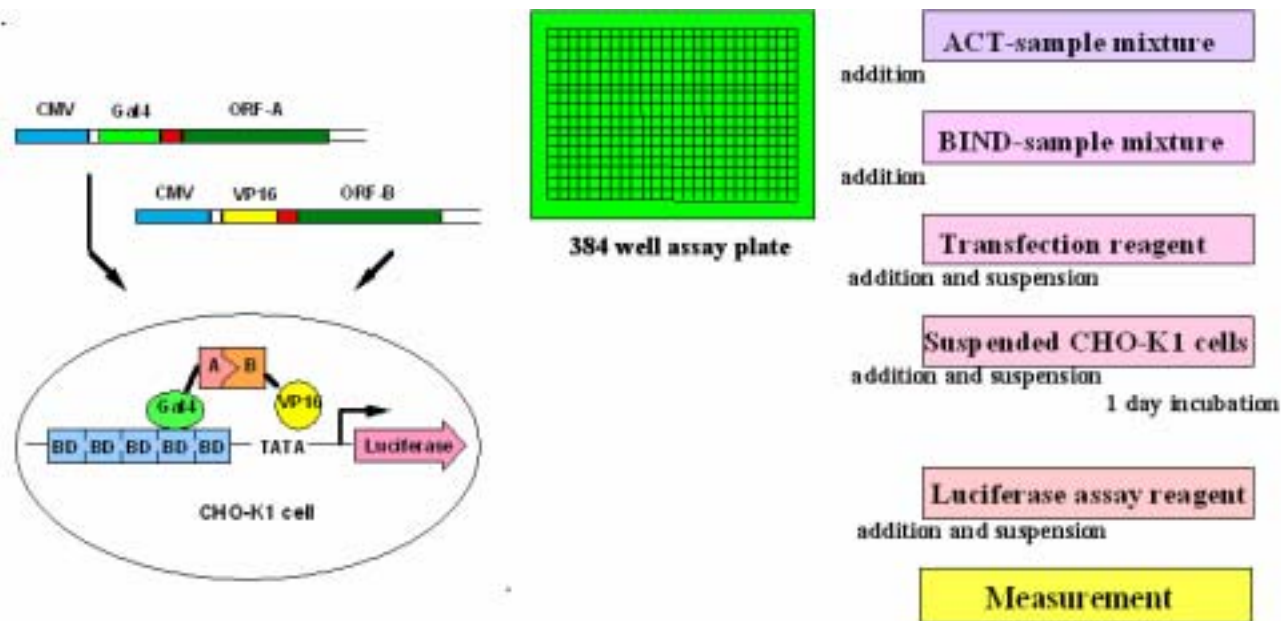
T2BP, a novel TRAF2 binding protein, can activate NF- κ B Ap-1 without TNF stimulation

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

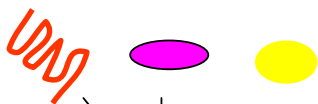
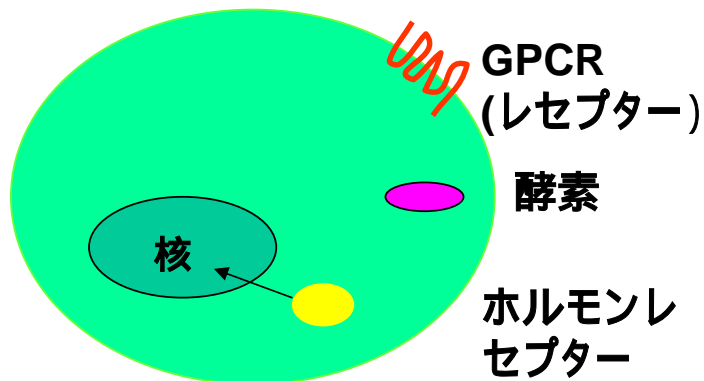
- ホルモン
- 神経伝達
- 細胞間信号伝達
- 細胞内信号伝達
- 転写ネットワーク

- 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。

蛋白質-蛋白質相互作用と創薬

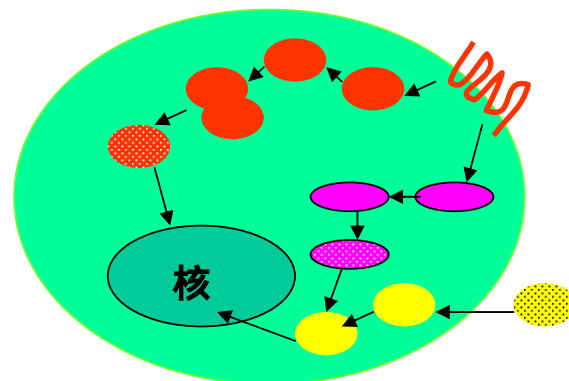
これまでの創薬



個別ターゲットバリデーション

リード化合物探索,
最適化

ゲノム/プロテオーム創薬



- 新規特異的ターゲット
- ターゲット
- 新規ドラッグアブルターゲット

ターゲットプール