

2004年8月3日
独立行政法人 理化学研究所

疾患・表現形質マップの解析システムを公開

独立行政法人理化学研究所（野依良治理事長）は、疾患を中心とした表現形質の遺伝学的マッピング情報（マップ情報^{*1}）を文献から収集して電子化し、ゲノムや生体内パスウェイ情報と統合的に利用できるデータベース：『TraitMap（トレイトマップ）』を開発しました。ゲノム科学総合研究センター（GSC:榊佳之センター長）ゲノム知識ベース研究開発チームの豊田哲郎チームリーダーらによる成果です。

今までの疾患関連遺伝子座^{*2}を収録したデータベースと異なる点として、TraitMapでは原因遺伝子が存在する可能性（対数オッズ=LOD値^{*3}）をゲノム上のマップ情報として記録してあるのが特徴です。このためマップ情報と生体反応の流れ（生体内パスウェイ^{*4}）の関連性を定量的に評価したり、複数のマップ情報を合成して解析するなど、コンピュータ上での幅広い利用が可能になりました。

さらにGSCでは「生命戦略統合データベースシステム(<http://omicspace.riken.jp>)」の開発を進めており、ヒトを含む動植物においてゲノムからフェノーム（表現形質）^{*5}にわたる全体像（オーミックススペース^{*6}）の解明を進めています。この統合システムには生体内パスウェイなど様々な情報が統合的に収録されていますが、TraitMapを加えることにより、原因遺伝子が存在する可能性の高い生体内パスウェイをTraitMapのレコードから直接検索することが可能になりました。

このようにさまざまな生命現象のつながりを、コンピュータ上でも調べることにより、ライフサイエンス研究の促進、革新的な医療技術等産業の発展が期待されます。本研究成果は、バイオインフォマティクス分野では最も有名な学会「ISMB/ECCB2004」^{*7}（8月4日、スコットランドにて開催）で発表され、論文はBioinformatics誌^{*8}の特集号に掲載されます。

1. 背景

塩基配列解析技術の進歩とともにヒトゲノムの塩基配列解析が行われ、現在、その全塩基配列はほとんど解析されています。また、ヒトは全て同じゲノム情報を持っているのではなく、個人間で異なる部分を多数持ち、個人の特徴を表しています。今後は、これら塩基配列情報を効率的かつ効果的に活用することにより、個人の体質に合わせた革新的な医療技術等産業の発展に結び付けるための技術開発が期待されていますが、特に、疾患を中心とした表現形質を収集して電子化したデータベースなど、その前提となる情報利用環境の整備及び高度化を特に図る必要があります。

2. 研究成果と手法

多くの疾患研究では疾患原因遺伝子とその変異を見つけ出すことを目的として、数多くの患者やモデル動物を調べてデータを集め、それを遺伝学的マッピングという方法で対数オッズ（LOD値）のマップを作成します。このLOD値はゲノム上の

ある領域内に原因遺伝子が存在する可能性を反映する指標であり、この値が大きいほどそこに原因遺伝子が存在する可能性が高いことが分かります。このようなマップ情報はこれまで数多くの文献に報告されてきましたが、ほとんどが図1のような画像データであるため、この LOD 値の情報をコンピュータに解釈させて利用することが難しい状況でした。

研究チームでは、このマップ情報を電子化したデータベース (TraitMap) を作成し、コンピュータが直接解釈できるようにしました。これにより初めてマップ情報が定量的に他の情報と統合化して利用できるようになり、マップ情報と生体内パスウェイの関連性を定量的に評価したり、複数のマップ情報を合成して解析するなど、コンピュータ上での幅広い利用が可能になりました。その簡単な例をあげると、図2に示すように、TraitMap に記録されている LOD 値のグラフとゲノムデータを照らし合わせることで、各遺伝子座における LOD 値を計算できます。そして、その遺伝子を代謝経路などの生体内パスウェイと照合させることで、パスウェイに乗っている遺伝子の LOD 値の合計値を算出することができ、各パスウェイが疾患に関わる領域にどれだけ含まれているかを計算してランク付けすることができ、疾患原因遺伝子探索の際に参考にすることができます。

TraitMap の開発において、マップ情報は糖尿病などの疾患を中心として文献から集め、XML というデータ記述言語を用いることでコンピュータが解釈可能なデータ形式を定義した結果、表1に示すようにヒト、マウス、シロイヌナズナで120のマップデータのレコードからなるデータベースになりました。図3ではヒトの染色体上に TraitMap レコードに記録されている LOD 値のグラフを赤で表示させた様子です。図4はオーミックススペースの構造をベースに構築された統合データベースシステムのアーキテクチャを概念的にあらわしています。各層には多数のデータセットが各平面として収納されており、必要なデータセットを選択して統合的に検索できるようになっています。このデータ構造を横から眺めると従来型のゲノムブラウザ⁹と似た画面になりますが、上から眺めると、ゲノム内に存在する遺伝子間のネットワークが点で表示されており、これは TraitMap の画面で、マップ情報と生体内ネットワーク (パスウェイ) の情報を重ね合わせて比較することができることをあらわしています。

TraitMap 使用例として、2型糖尿病患者では心疾患の合併症を引き起こしやすく、非 HDL コレステロールの血中濃度が高い人ほど心疾患になりやすいことが知られていますが、これら2つの疾患に関するマップ情報を使って、共通するパスウェイが存在するかどうかを検索してみました。マウスにおける2型糖尿病のリスク因子として、インスリン抵抗性へのなりやすさの指標の1つである血中インスリン濃度に基づくマップ情報を使用し、また、心疾患へのなりやすさの指標の1つである非 HDL コレステロールの値に基づくマップ情報の2つを使用しました。

図5・図6に示すようにこれら2つのマップの共通領域において、血中インスリン濃度のマップ領域内にある(1番染色体上に存在する)遺伝子 *Irs1* と、非 HDL リポタンパクコレステロール濃度のマップ領域内にある(4番染色体上に存在する) *Lepr*、(6番染色体上に存在する) *Pparg* との関連性情報の存在が最上位にランキングされ、これら遺伝子は2つの疾患に共通して影響を与えている可能性が高いパスウェイであるという推定結果が得られました。このような推定は擬陽性の場合も

あるので解釈に注意が必要ですが、このような推定が迅速に行えるシステムは疾患原因遺伝子探索研究に非常に有用なツールとなります。

3. 今後の展開

今後は、この TraitMap を含む図 4 の統合データベースを用いることで、ヒトやマウスやシロイヌナズナを使った表現形質解析から原因遺伝子を探索する研究で遺伝子推定を効率的に行うことができ、研究全体の効率化が期待されます。特にゲノム分野では膨大なデータが公開されており、それを研究現場で効率的に活用することが求められています。また、TraitMap を含む図 4 の統合データベースはゲノム関連データの高度な利用環境を世界中の研究者が容易に享受できるようにするものです。

(問い合わせ先)

独立行政法人理化学研究所 横浜研究所
ゲノム科学総合研究センター
ゲノム知識ベース研究開発チーム

チームリーダー 豊田 哲郎

Tel : 045-503-9610 / Fax : 045-503-9553

(報道担当)

独立行政法人理化学研究所 広報室

Tel : 048-467-9272 / Fax : 048-462-4715

Mail : koho@riken.jp

<補足説明>

※1 遺伝学的マッピング情報(マップ情報)

表現形質の遺伝学的マッピング情報とは、原因遺伝子が存在する可能性を確率的指標 (LOD 値) で示すマップ。このマップが明らかになっても、原因遺伝子までたどり着いたわけではない。マップ領域内には多数の遺伝子が存在し、その中から真の原因遺伝子を突き止めるにはさらに膨大な研究時間がかかる。一般的に研究者はあたりをつけながら可能性の高そうな遺伝子の周囲に変異がないか調べている。このため、マップ情報をデータベース化し、生体内パスウェイ情報などと照らし合わせて、原因遺伝子の候補を推定する情報技術は非常に重要である。

※2 疾患関連遺伝子座

疾患の発症やかかりやすさに影響を与える遺伝子を疾患関連遺伝子といい、その遺伝子のゲノム上の存在位置を遺伝子座という。

※3 対数オッズ

本文中で説明されている LOD 値と同じ意味。遺伝学的マッピング計算で算出される。遺伝現象において疾患形質とそこにあるマーカーが強く連鎖する場合に大きい値となり、原因遺伝子の存在領域を絞り込む際の指標となる。

※4 生体内パスウェイ

代謝反応やシグナル伝達など、生体内分子に媒介される一連の生体反応の流れをいう。

※5 フェノーム(表現形質)

フェノタイプの集合概念である。ちなみに生命がもつ1セットの遺伝子集合がゲノム。

※6 オーミックスペース

図7はオーミックスペース (Omic Space) の概念図です。下からゲノム、トランスクリプトーム、プロテオーム、メタボローム、フェノームの階層構造になっており、各平面における分子間の相互作用が点で表示されています。例えば、タンパクAが遺伝子Bの転写を促進するという相互関係がトランスクリプトーム平面上の点ABで表現されています。そして、各階層を経由して縦方向に繋がるネットワークがあり表現形質としてフェノーム層に現れる様子かプロットされています。

※7 ISMB/ECCB2004

正式名称は、「International Conference on Intelligent Systems for Molecular Biology / European Conference on Computational Biology 2004」で、バイオインフォマティクス分野では最も有名な学会。日本でも「ISMB」と呼ばれている。

※8 Bioinformatics 誌

英国の学術専門誌。インパクトファクター（雑誌の影響力を表す一つの指標）は4.615であるが生物情報科学の代表誌される。

※9 ゲノムブラウザ

ゲノムデータを閲覧するアプリケーションソフトのこと。

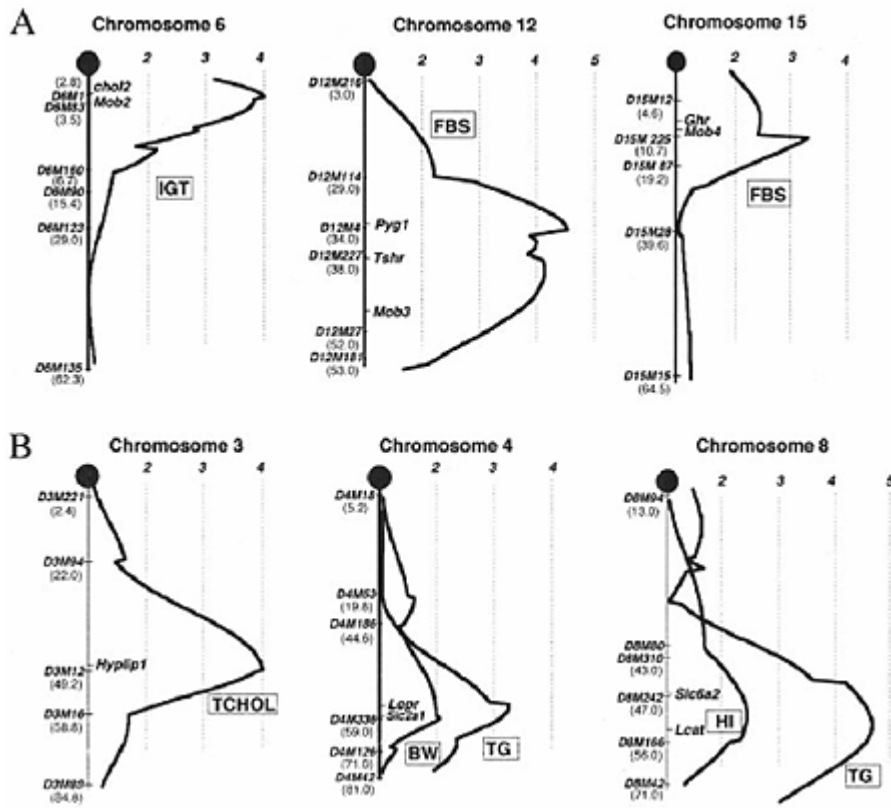


図1 文献に記載されているマッピング結果の例

(Susceptibility and Negative Epistatic Loci Contributing to Type 2 Diabetes and Related Phenotypes in a KK/Ta Mouse Model" Toshihide Shike et al. 2001)

表 1. トraitマップデータベースの記録リスト

種類	疾患区分	記録No.
シロイヌナズナ	circadian rhythm	1
	hormone response	2
	inflorescence development	3
	light response	5
ヒト	disease	12
	risk factor for disease	11
マウス	behavior	2
	disease	9
	fitness component	2
	mandible shape	4
	paradigm for antidepressants	4
	resistibility against disease	4
	susceptibility to insulin resistance	2
	risk factor for disease	59
	計	120

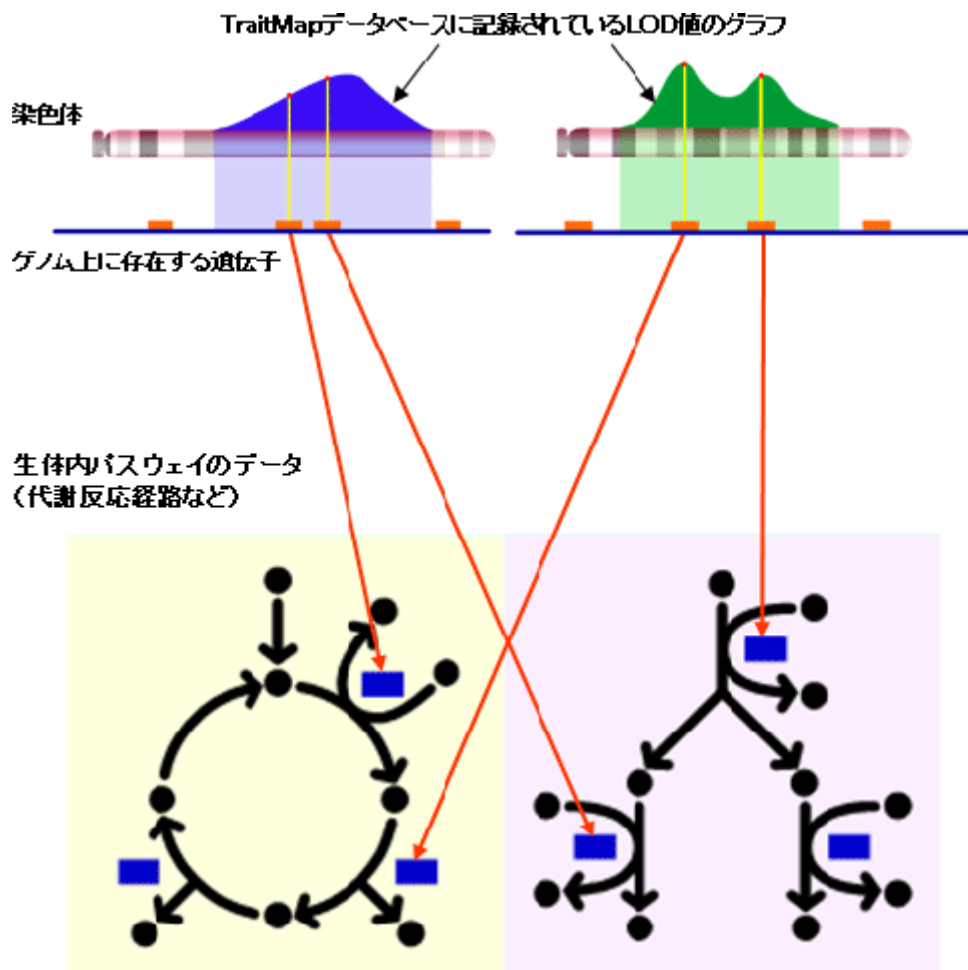


図2 LOD値と生体内パスウェイ

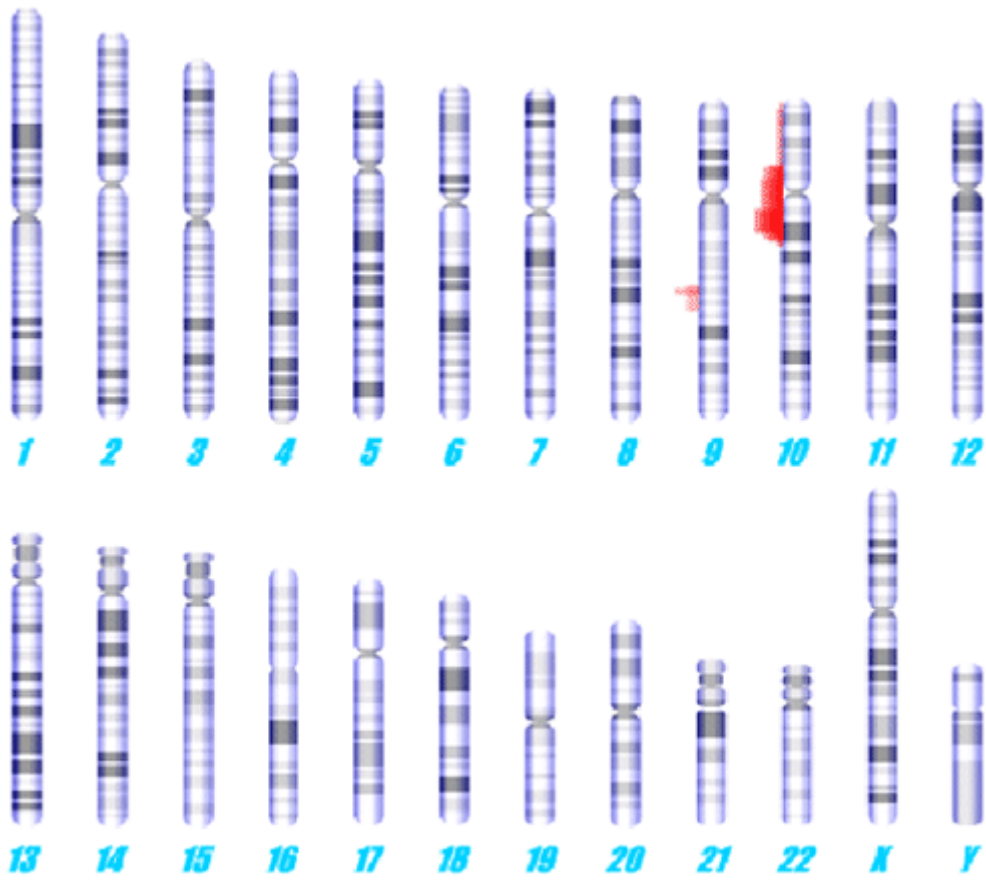


図3 ヒト染色体上に表示されたマップ情報 (TraitMap レコードのデータ)

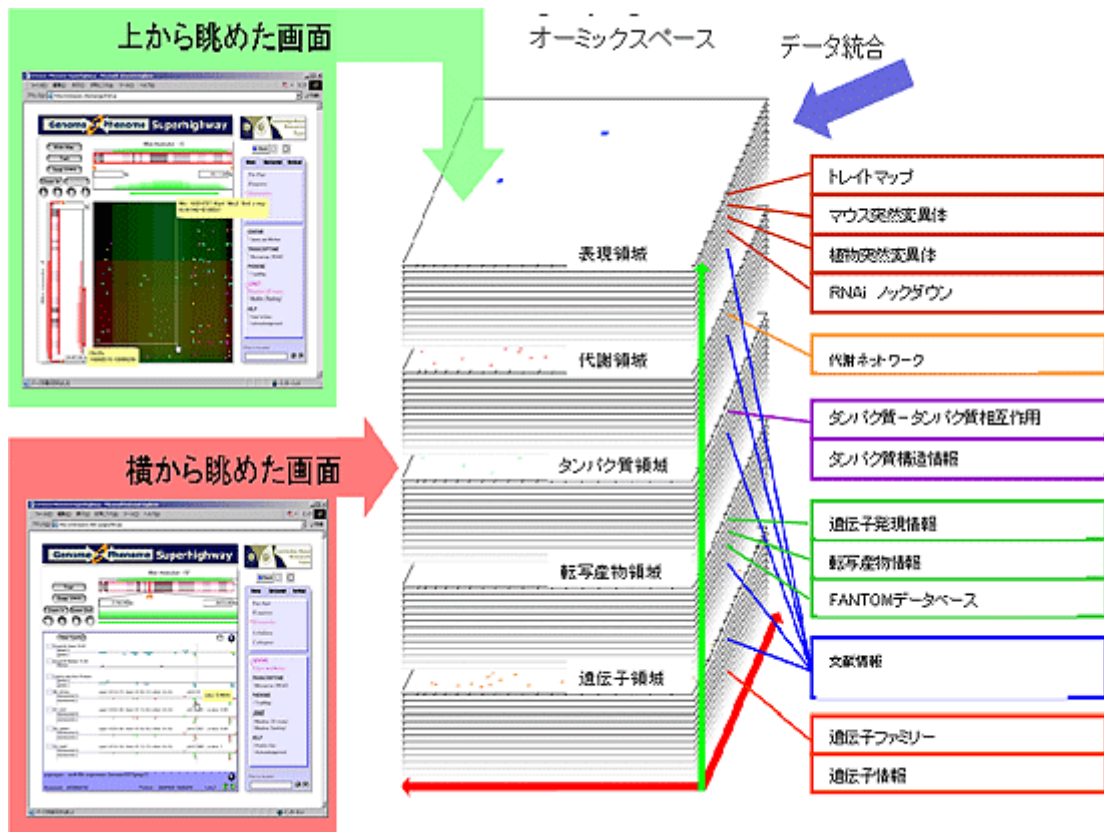


図 4 統合データベースシステムのアーキテクチャー概念図

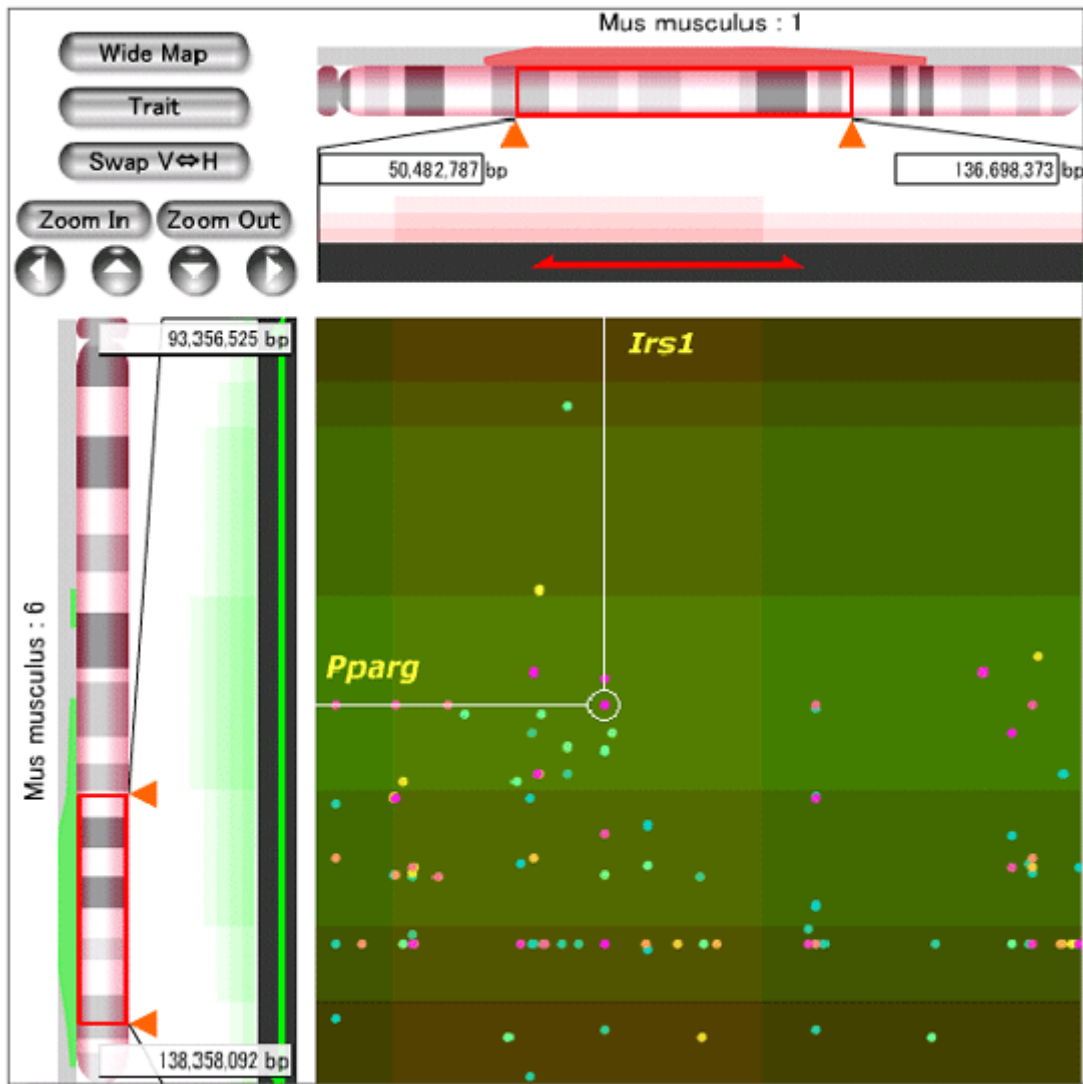


図5 TraitMap データベースを使った解析例 (1)

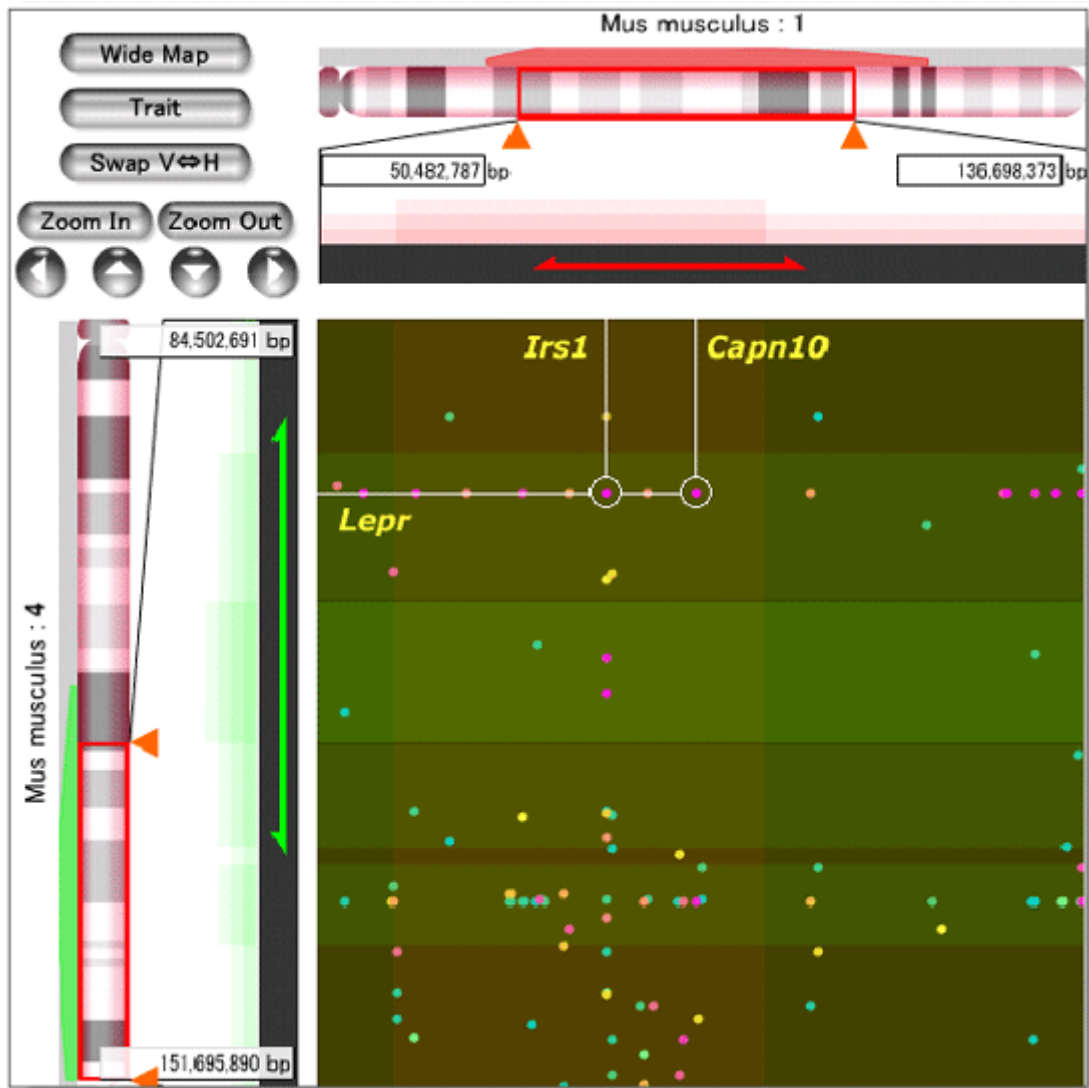


図6 TraitMap データベースを使った解析例 (2)

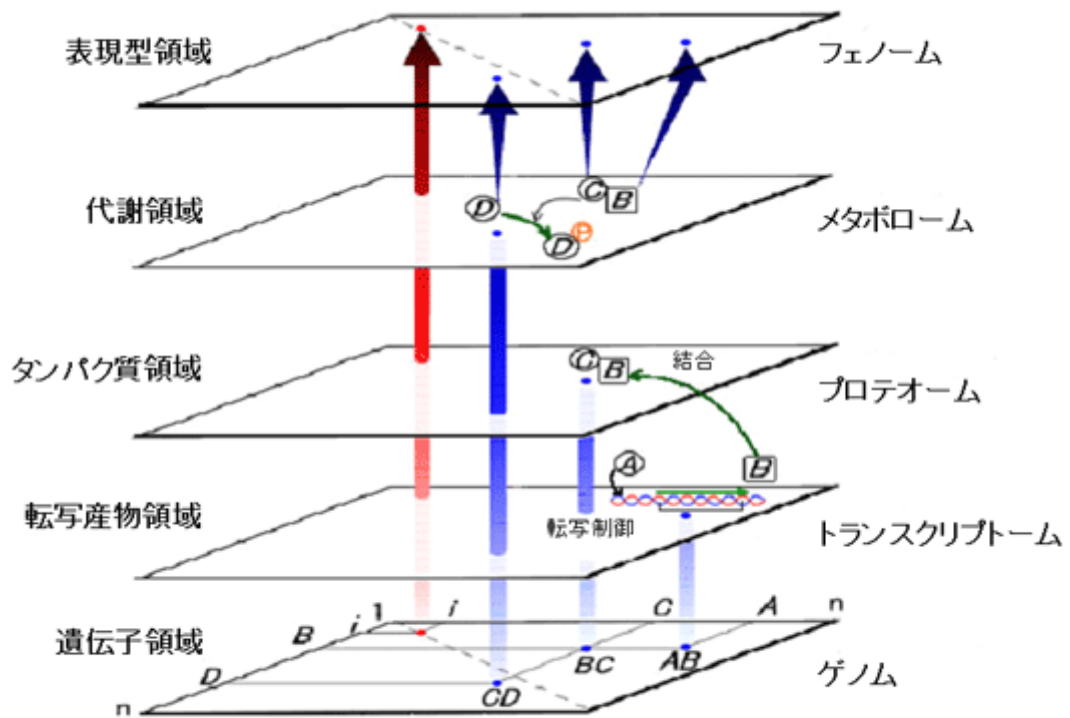


図7 オーミックススペース (Omic Space) の概念図