

するが、この対応は一般的に多対多関係であるため、そのままではデータ処理計算量が大幅に増大し、また解析結果に偏りや重複要素が発生して、データ処理が混乱する。さらにマウス、ラットのマイクロアレイの設計で、一部の遺伝子に複数のプローブセットが用意されていることも事態を一層複雑化させてしまう。このためまず、共通遺伝子名毎にデータを集約してデータを識別する方法を試行したが、複数の発現変動データを要約(summarize)するにしても、個々の発現データをそのままに識別子重複を許容するにしても、解析処理の混乱を回避できなかった。特に元の個別情報への参照が、要約過程での情報集約により困難になることが致命的であった。そこで次に、多対多対応のそれぞれにユニークな識別子(統合ID)を設定して処理し、必要に応じて summarize する方式を試行したところ、一部に相同遺伝子対応関係の多寡によるバイアスが残るものの、全般的には精度・速度のバランスの取れた解析処理が可能となった。

そこで順次 TGP ラットデータの Percellome 化を進めたが、TGP ではプロジェクト初期を除いて、Percellome 適用上必要な Wet レベルの品質管理ステップを行わなかったため、多くのデータで Percellome 専用の外部 RNA スパイク (GSC) 添加における精度不良が見られた。Percellome プロジェクト本体と同様の厳密さで精度管理を行うと統計処理等に必要の群内サンプル数を確保できなかったため、各群のデータ分散量を 50%CV まで許容して、データ変換処理を行った。

Percellome 化した TGP ラットデータは、

全て Surface グラフデータセットに変換し、先行研究で開発した RSort プログラム(個々の遺伝子の発現変動で、Surface グラフ(3次元波動面)の凹凸が少ないパターンを示す遺伝子、つまり、より強力もしくはよりユニークな発現制御を受ける遺伝子を自動抽出するソフトウェア)にてラットのプローブセット Id からなる候補遺伝子リストを生成した。これを上記の統合IDに変換してから、類似発現変動評価用ソフトウェア PercellomeExplorer の専用データベースにインポートして、解析性能(精度や処理速度)を評価した。その結果、種差の影響が最も強いものの、化学物質投与による種差を超えた共通影響も検出可能であり、計算処理速度も特段の劣化が無いなど、実用に耐える性能を有することが確認できた。

次に、Percellome データベース及び解析プログラムの公開に伴い、その汎用性を強化すべく、Percellome 法を適用していない外部のマイクロアレイデータとの比較を可能とする技術の開発を試みた。基礎調査として、Percellome データベースの全データを照会したところ、実験条件が類似したデータ間では全遺伝子の発現量はある範囲に収まり、またヒストグラムパターンも高〜中発現域において類似することが分かった。そこで、全遺伝子データを発現量順に並べ、ある一定の順位にあるプローブセットを中心に数百から数千の範囲のプローブセットをリストアップして基準値候補を得られないうか検討した。その結果、いわゆる Housekeeping 遺伝子を含め、安定した基準点になり得る遺伝子候補は無いものの、Percellome プロジェクトの実験条件設定の

場合、90%tile のプローブセットを中心に、前後 500 プローブセットずつ、計 1001 個のプローブセットを基準値候補として、Percellome データベース内の同条件データを対象に基準値候補のプローブセットの絶対量データを数万から数十万集め、その中央値を推定絶対量基準値とすることで、精度良く絶対量推定を行うことができることを見いだした。この推定プロセスは手作業で行うにはあまりにも煩雑であるため、これを自動化する絶対量推定ソフトウェア SnCalc.exe を開発し、さらに数百の絶対量推定プロセスを実行して推定計算の精度を確認した。

本推計アルゴリズムは原理的に全遺伝子の発現量分布が類似している Percellome データを照会する必要があるため、そのデータ照会が正しいかどうか、つまり全遺伝子の発現量分布が類似しているかどうかを検証するためのソフトウェア DDComp.exe も作成した。

一方、Percellome データベースにインターネット経由でアクセスし、絶対量データや Surface グラフを参照するために必要となるゲートウェイとして、現時点で最も普及しているプロトコルの 1 つ、REST (Representational State Transfer) に準拠した WebAPI を設計・構築し、Percellome 公開用 Web サーバーに実装した。これにより、ライフサイエンス研究用ソフトウェアの国際共通プラットフォーム GARUDA プロジェクト等、外部の学術・行政システムからのオンラインアクセスを可能にする本格的な一般公開の準備を整えた。

D. 考察

一般的に異種由来のデータ結合は困難とされているが、本年度生成したデータ統合アルゴリズムにより、統合データベースから種間共通要素や種差要素の検出など実用的な解析が可能となった。これは TGP ラットデータの Percellome マウスデータベースへの統合を実現し、より広範囲の化学物質-遺伝子発現情報が利用可能となったほか、種間でも保存されている毒性ネットワークの抽出など、従来にない成果が得られるものと期待される。また本技術は原理的にマウス-ラット間だけでなく、マウス-ヒト間のデータ結合も可能であるため、解析結果のヒトへの外挿を取り組むに当たっても有用な技術である。

絶対量推定ソフトウェア SnCalc.exe の適用には前提条件があるものの、Percellome 法に準拠せず取得されたマウスデータであっても絶対量推定し、Percellome データベースと数値レベルで直接比較できる可能性を提供した。これは膨大な量の既存データの有効活用に繋がる革新的な成果であり、Percellome データベースの有用性を飛躍的に高めるものである。また現在はマウスのみが対象となっているが、TGP ラットデータの完全統合の暁には、ラットデータについても推定計算が可能になる予定であり、創薬や毒性試験分野における膨大な既存ラットデータの有効活用が見込まれる。

これらの成果を外部の研究システムから自由に利用できるよう構築した Percellome WebAPI は、行政施策への貢献や学術研究、特に共同研究の推進効果が見込まれる。本 WebAPI には今後も新たな解析情報・機能の追加を予定しており、継続的により一層の

利用拡大と社会貢献を目指す。

373-380

E. 結論

平成 24 年度は研究計画に沿って開発研究を進め、予定通り、TGP ラットデータの Percellome 化及び統合データベース構築を開始し、Percellome データベース公開用 WebAPI の構築を行った。さらに追加の成果として Percellome 法を適用していない外部のマイクロアレイデータの絶対量推定ソフトウェア SnCalc.exe を開発した。

Percellome データベースに統合した TGP ラットデータの利用については、既に RSort.exe や PercellomeExplorer.exe 等、主要な解析ソフトウェアの対応アップデートをほぼ終えており、シームレスな解析操作が可能となっている。これらの成果は本研究班の反復暴露解析へ即時還元し、解析情報の精度向上や有効情報の生成に貢献する見込みである。Percellome データベース本体については外部への公開準備を整えた。今後、積極的な情報提供および共同研究の推進を行い、トキシコゲノミクス技術による毒性予測・評価システムの実用化に向けた取り組みを一層加速させる。

F. 研究発表

1. 論文発表

1. Katsuhide Igarashi, Satoshi Kitajima, Ken-ichi Aisaki, Kentaro Tanemura, Yuhji Taquahashi, Noriko Moriyama, Eriko Ikeno, Nae Matsuda, Yumiko Saga, Bruce Blumberg, and Jun Kanno (2012). Development of Humanized Steroid and Xenobiotic Receptor Mouse by homologous knock-in of the human Steroid and Xenobiotic Receptor Ligand Binding Domain sequence. *J Toxicol Sci* 37:

2. Abe S, Kurata M, Suzuki S, Yamamoto K, Aisaki K, Kanno J, Kitagawa M. (2012). Minichromosome maintenance 2 bound with retroviral Gp70 is localized to cytoplasm and enhances DNA-damage-induced apoptosis. *PLoS One*. 7(6):e40129.

2. 学会発表

1. 種村健太郎、古川佑介、大塚まき、五十嵐勝秀、相崎健一、北嶋 聡、佐藤英明、菅野 純、発生-発達期の神経シグナルかく乱による遅発中枢影響解析-幼弱期マウスへのイボテン酸投与による成熟期の脳高次機能障害について-、第39回日本毒性学会学術年会 (2012. 7. 17)
2. 北嶋 聡、相崎健一、五十嵐勝秀、菅野 純、食品の安全性確認に向けた Percellome トキシコゲノミクスの適用-香料エストラゴールの場合-、第39回日本毒性学会学術年会 (2012. 7. 17)
3. 北嶋 聡、高橋 祐次、五十嵐 勝秀、相崎 健一、菅野 純、Percellome 網羅的定量的トキシコゲノミクス、平成24年度公益社団法人日本実験動物学会維持会員懇談会 (2012. 11. 16)

G. 知的所有権の取得状況

1. 特許取得

特許第 5177712 号、2013 年 1 月 18 日登録、特許権者：国立医薬品食品衛生研究所、NTT データ、発明者：菅野純、相崎健一ら、「競合的ハイブリダイゼーションにおける遺伝子データの補正方法及び補正装置」

2. 実用新案登録

なし

3. その他

なし



化学物質の有害性評価手法の迅速化、高度化に関する研究

—網羅的定量的大規模トキシコゲノミクスデータベースの維持・拡充と
毒性予測評価システムの実用化の為にインフォマティクス技術開発—

【分担研究】

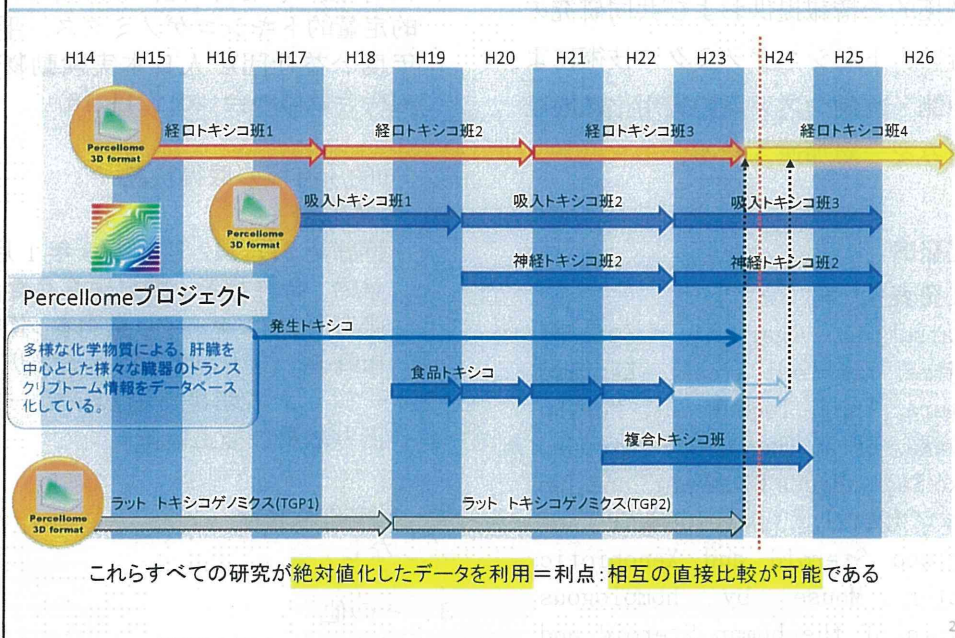
Percellome 3次元データ等の為の専用解析ソフトウェアの 開発研究

国立医薬品食品衛生研究所
安全性生物試験研究センター
毒性部 第一室
相崎 健一

1



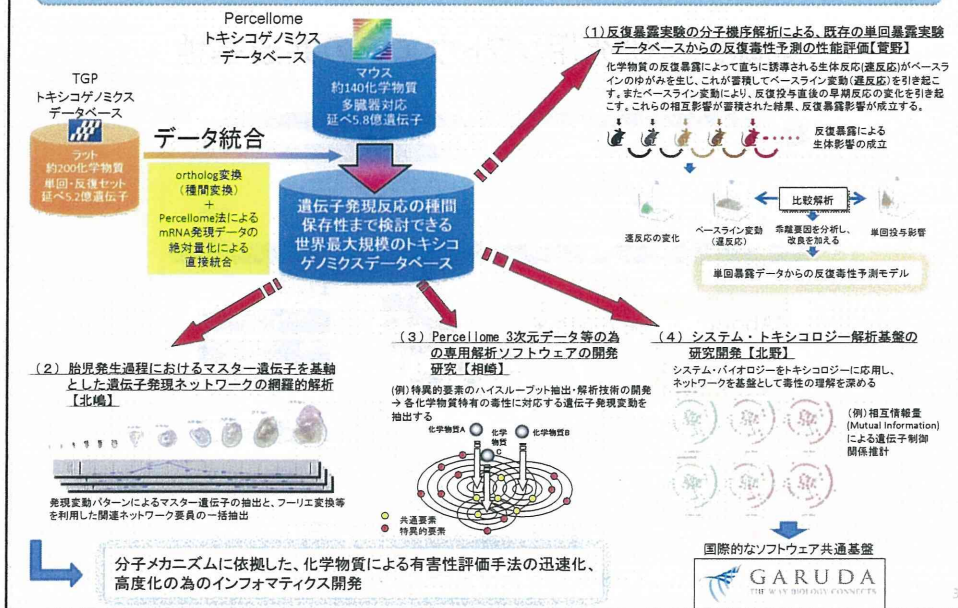
毒性部が展開するトキシコゲノミクス研究



2

化学物質の有害性評価手法の迅速化、高度化に関する研究

—網羅的定量的大規模トキシゲノミクスデータベースの維持・拡充と毒性予測評価システムの実用化の為のインフォマティクス技術開発—



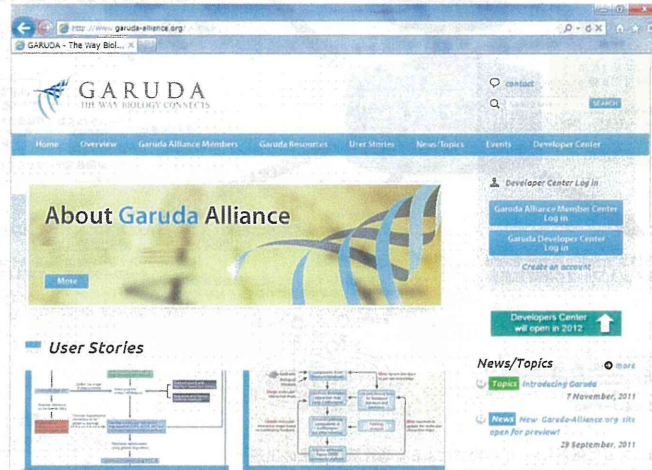
「公開」に向けた取り組み

1. REST API of PercellomeWeb
2. 一般データの絶対量推定技術
3. TGPラットデータ統合～相同遺伝子情報結合



GARUDA Alliance

国際的な研究用ソフトウェアの共通基盤



5



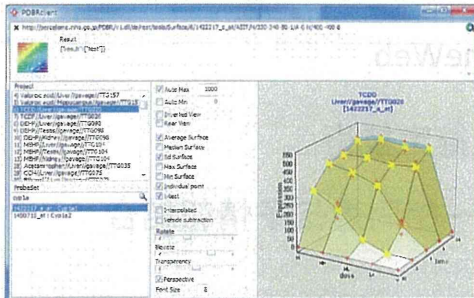
REST API of PercellomeWeb

REST (Representational State Transfer)
ハイパーメディアシステムのためのソフトウェア
アーキテクチャのスタイルのひとつ

http://percellome.nih.go.jp/PDBR/v1.dll/ds/rest/tools/Surface/6/1422217_a_at/ASIT/N/330-340-50-1/A-0-N/400-400-8



REST client 例 (GARUDA Application)



User manual



Functions

- Announcement
- PercellomeLogo
- Project
- ProbeSetId
- GeneInformation
- GroupAverageData(formatted)
- GroupSdData(formatted)
- GroupMedianData(formatted)
- GroupMaximumData(formatted)
- GroupMinimumData(formatted)
- IndividualData(formatted)
- SurfaceGraph

6



「公開」関連技術の開発①

Percellome非適用*の 測定済みマイクロアレイデータの 絶対量推定

*測定時にPercellome外部SpikeRNAを添加していないもの



Percellome絶対量化手法拡張の必要性

- 一般的な(外部Spikeが添加されていない)測定済みマイクロアレイデータの絶対量推定
 - >>> 数値レベルでの直接比較が可能となり、Percellomeの利用が促進される
 - Percellomeデータの一層の高精度化
 - >>> 外部Spikeの添加エラーによるデータ不安定化の抑制
- >>>> SnCalcの作成



SnCalc 基本コンセプト

- Percellomeデータベースの規模を活かし、Percellomeで絶対量化した溶媒群のデータから、各データの基準点の「あるべき」コピー数/細胞を求める。
 - ・・・前提) トランスクリプトームの分布がPercellomeDBの標準データと同様であること。

>>> DDComp.exeによるトランスクリプトーム分布レベルのQC実施
(Difference of Distribution Compare)

SnCalc.exe

Filename	Path	Time	Time	Code
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	02	M	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	02	M	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	02	M	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	02	M	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	V	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	V	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	V	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	V	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	L	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	04	L	
TT0119-010-08-0006.TOT	R:\MTT0119-U1User	02	V	

- 絶対量換算に用いる検量線は、
基準点と原点を通る直線を
採用する。



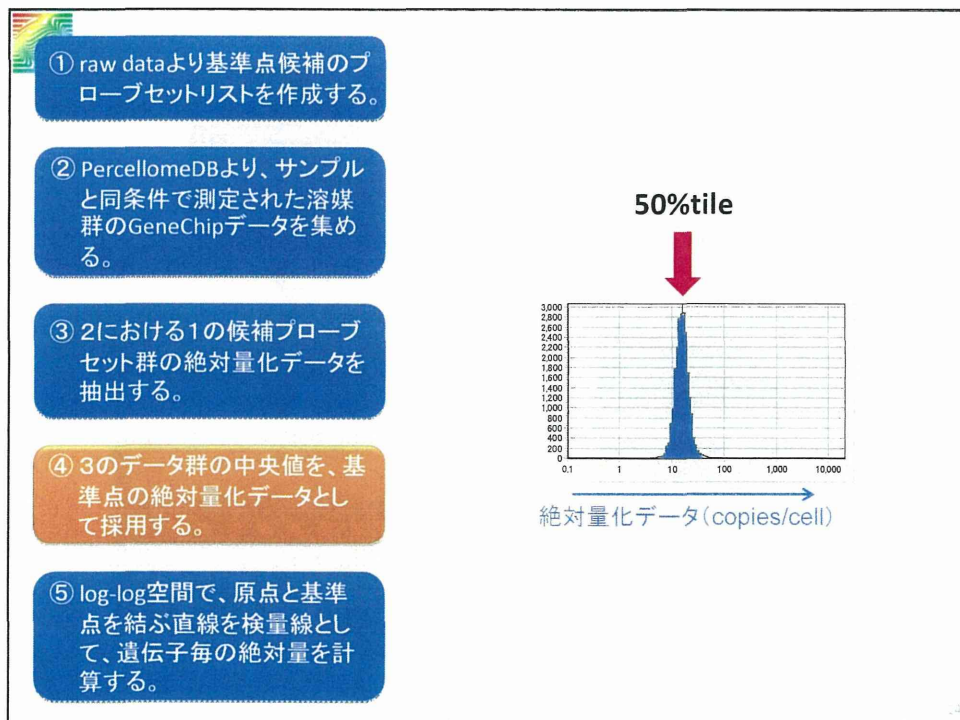
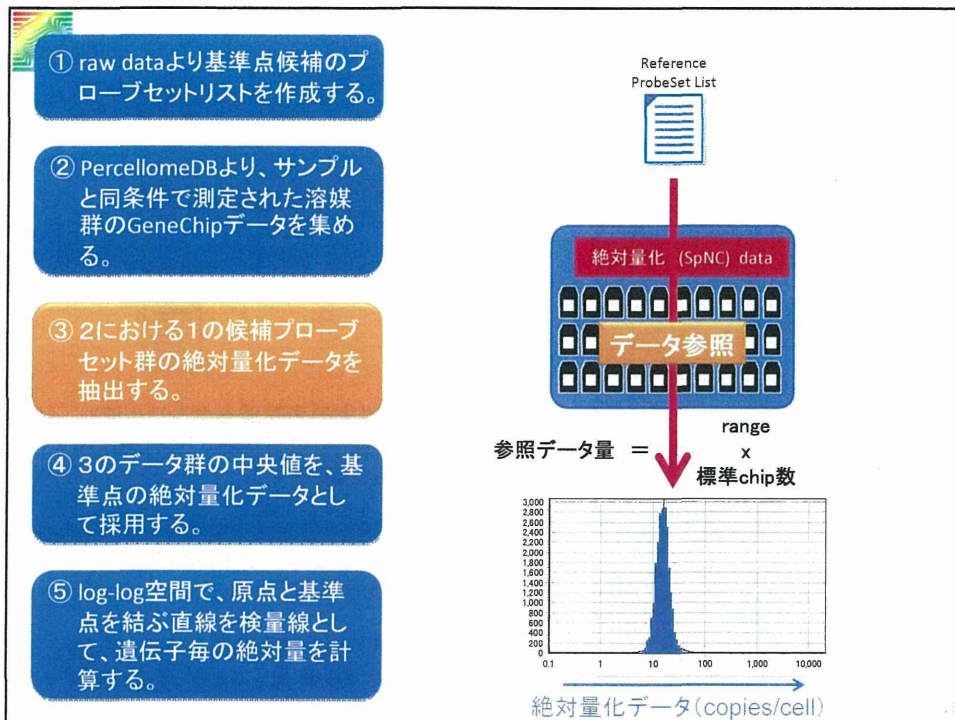
- ① raw dataより基準点候補のプロブセットリストを作成する。
- ② PercellomeDBより、サンプルと同条件で測定された溶媒群のGeneChipデータを集める。
- ③ 2における1の候補プロブセット群の絶対量化データを抽出する。
- ④ 3のデータ群の中央値を、基準点の絶対量化データとして採用する。
- ⑤ log-log空間で、原点と基準点を結ぶ直線を検量線として、遺伝子毎の絶対量を計算する。


- ① raw dataより基準点候補のプローブセットリストを作成する。
- ② PercellomeDBより、サンプルと同条件で測定された溶媒群のGeneChipデータを集める。
- ③ 2における1の候補プローブセット群の絶対量化データを抽出する。
- ④ 3のデータ群の中央値を、基準点の絶対量化データとして採用する。
- ⑤ log-log空間で、原点と基準点を結ぶ直線を検量線として、遺伝子毎の絶対量を計算する。

raw data Center 90%tile / Range 1000ps = 90%tile ± 5%

- 疑似適用対象の測定済みサンプルにPercellome用の外部Spike RNAが添加されていないとも良い！
- 基準点候補を、特定のプローブセットに固定しない為、より多くのマイクロアレイデータに適用可能である。
- PercellomeDB全体を参照する為、個々の参照データの誤差による影響を排除できる。

- ① raw dataより基準点候補のプローブセットリストを作成する。
- ② PercellomeDBより、サンプルと同条件で測定された溶媒群のGeneChipデータを集める。
- ③ 2における1の候補プローブセット群の絶対量化データを抽出する。
- ④ 3のデータ群の中央値を、基準点の絶対量化データとして採用する。
- ⑤ log-log空間で、原点と基準点を結ぶ直線を検量線として、遺伝子毎の絶対量を計算する。

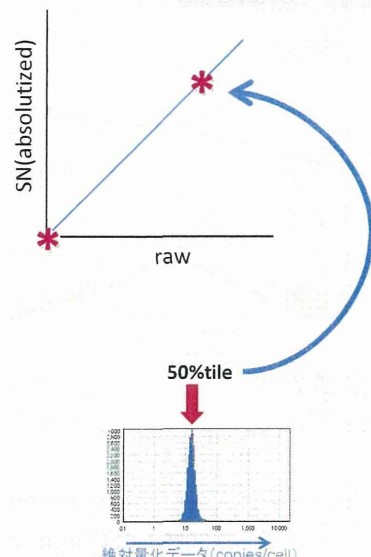




- ① raw dataより基準点候補のプロープセットリストを作成する。
- ② PercellomeDBより、サンプルと同条件で測定された溶媒群のGeneChipデータを集める。
- ③ 2における1の候補プロープセット群の絶対量化データを抽出する。
- ④ 3のデータ群の中央値を、基準点の絶対量化データとして採用する。
- ⑤ log-log空間で、原点と基準点を結ぶ直線を検量線として、遺伝子毎の絶対量を計算する。

SN data


原点と基準点を通る直線を検量線とする



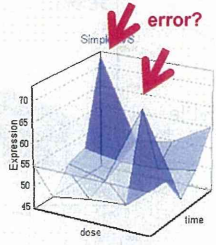
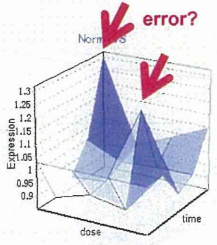
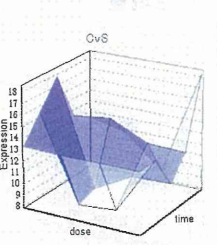
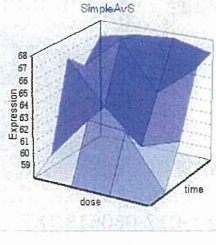
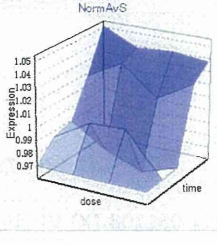
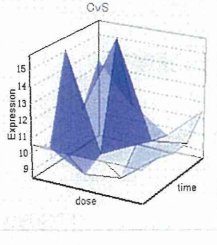
50%tile

絶対量化データ (copies/cell)

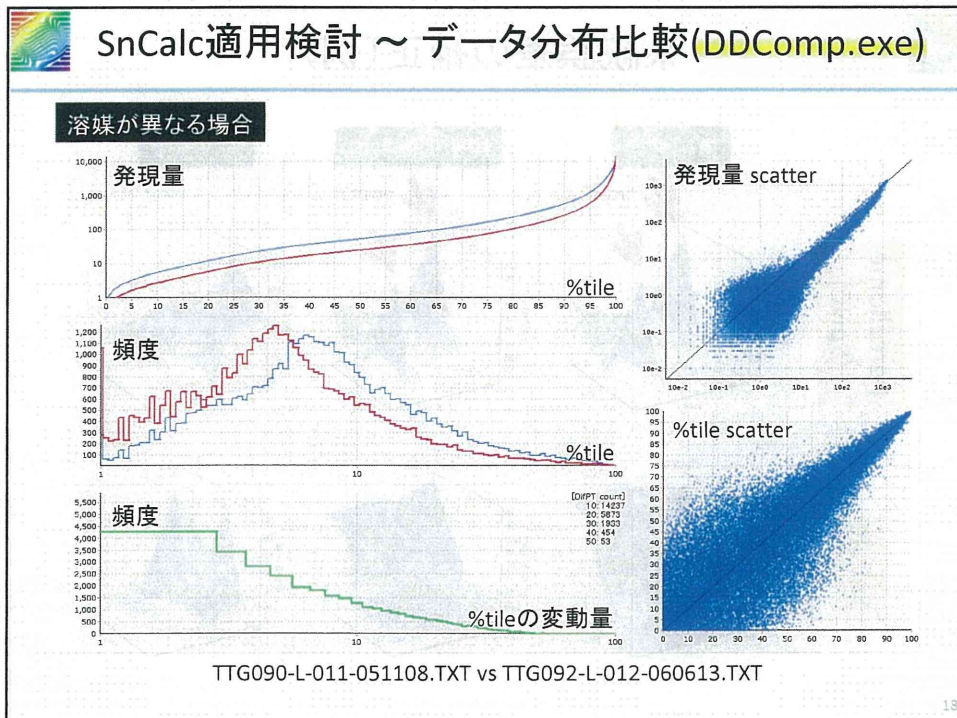
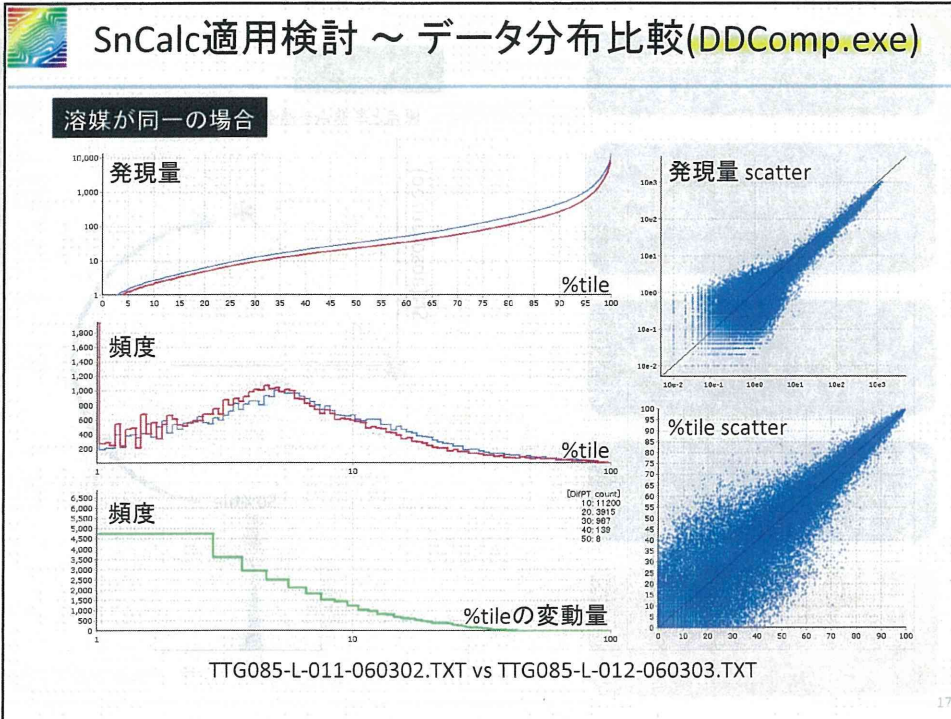
15



系統誤差の補正 (例)

	単純平均	標準化後平均	%CV平均
Percellome (SpNC)			
SnCalc			

16





新型反復暴露実験による 毒性(成立)機序の経時的解析

• 反復暴露による慢性毒性影響の分解

A. 概日リズム1周期内(24時間以内)

反復投与毎に繰り返される早い反応

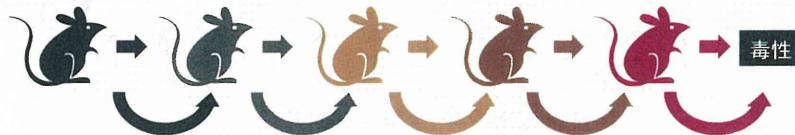
急性毒性、慢性暴露で毎日繰り返され、少しずつ変化してゆく過渡反応

B. 概日リズム周期単位(日単位)

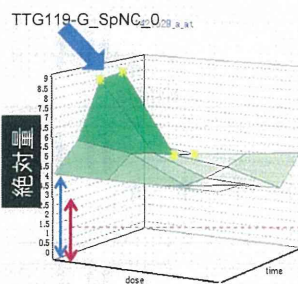
ベースラインの変化を起こす反応

AIによって生じた**恒常性維持機構**の歪み、
および その蓄積、記憶の機序(エピジェネティクスなどによる?)

➡ **※(準)定常状態の変化は、絶対量化データでなければ評価できない**

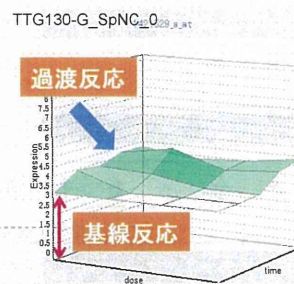


19



単回投与

同スケール



反復投与

Txnrd1

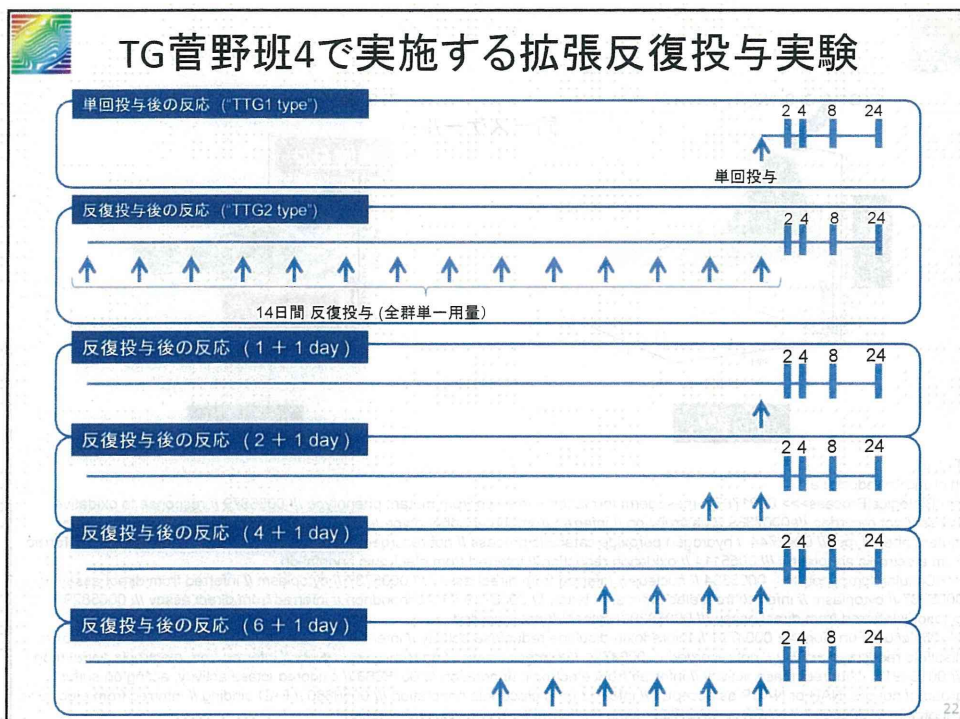
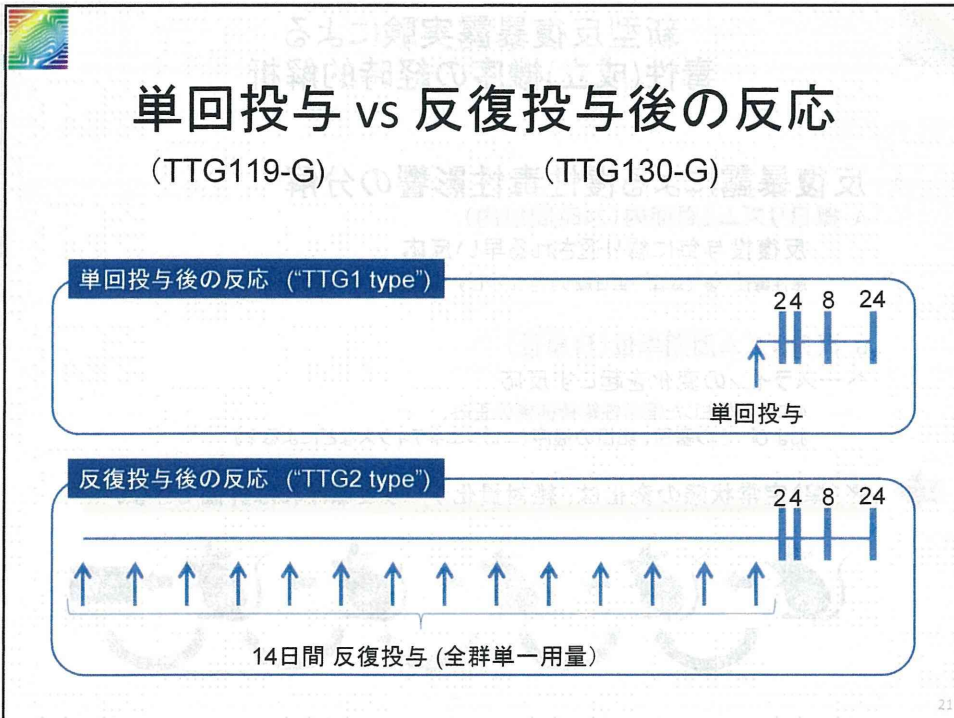
thioredoxin reductase 1


<<<BiologicalProcess>>> 0001707 // mesoderm formation // inferred from mutant phenotype /// 0006979 // response to oxidative stress // not recorded /// 0007369 // gastrulation // inferred from mutant phenotype /// 0008283 // cell proliferation // inferred from mutant phenotype /// 0042744 // hydrogen peroxide catabolic process // not recorded /// 0045454 // cell redox homeostasis // inferred from electronic annotation /// 0055114 // oxidation reduction // inferred from electronic annotation

<<<CellularComponent>>> 0005634 // nucleus // inferred from direct assay /// 0005737 // cytoplasm // inferred from direct assay /// 0005737 // cytoplasm // inferred from electronic annotation /// 0005739 // mitochondrion // inferred from direct assay /// 0005829 // cytosol // inferred from direct assay /// 0005829 // cytosol // not recorded

<<<MolecularFunction>>> 0004791 // thioredoxin-disulfide reductase activity // inferred from direct assay /// 0004791 // thioredoxin-disulfide reductase activity // inferred from electronic annotation /// 0016491 // oxidoreductase activity // inferred from electronic annotation /// 0016668 // oxidoreductase activity, acting on sulfur group of donors, NAD or NADP as acceptor // inferred from electronic annotation /// 0050660 // FAD binding // inferred from electronic annotation /// 0050661 // NADP or NADPH binding // inferred from electronic annotation

20






「公開」関連技術の開発②

ラットデータ(TGPデータ)の統合

23



化学物質の有害性評価手法の迅速化、高度化に関する研究

網羅的定量的大規模トキシゲノミクスデータベースの維持・拡充と毒性予測評価システムの実用化の為に
インフォマティクス技術開発

TGP
トキシゲノミクス
データベース

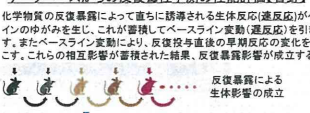
ラット
約2000化学物質
単回・反復セット
延べ5.2億遺伝子

Percellome
トキシゲノミクス
データベース

マウス
約140化学物質
多臓器対応
延べ5.8億遺伝子

(1) 反復暴露実験の分子機序解析による、既存の単回暴露実験データベースからの反復毒性予測の性能評価【北野】

化学物質の反復暴露によって直ちに誘導される生体反応(速反応)がベースラインのゆがみを生じ、これが蓄積してベースライン変動(遅反応)を引き起こす。またベースライン変動により、反復投与直後の早期反応の変化を引き起こす。これらの相互影響が蓄積された結果、反復暴露影響が成立する。



速反応の変化 → ベースライン変動(遅反応) → 単回投与影響

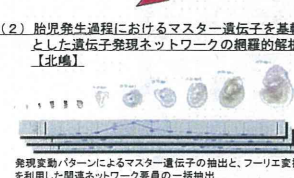
比較解析 → 単回暴露データからの反復毒性予測モデル

データ統合

ortholog変換(種間変換)
+
Percellome法によるmRNA発現データの絶対量化による直接統合

遺伝子発現反応の種間保存性まで検討できる世界最大規模のトキシゲノミクスデータベース

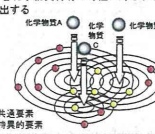
(2) 胎児発生過程におけるマスター遺伝子を基盤とした遺伝子発現ネットワークの網羅的解析【北嶋】



発現変動パターンによるマスター遺伝子の抽出と、ファミリー変換等を利用した関連ネットワーク要素の一括抽出

(3) Percellome 3次元データ等のための専用解析ソフトウェアの開発研究【相崎】


(例) 特異的要素のハイスループット抽出・解析技術の開発
→ 各化学物質特有の毒性に対応する遺伝子発現変動を抽出する



● 共通要素
● 特異的要素

(4) システム・トキシコロジー解析基盤の研究開発【北野】


システム・バイオロジーをトキシコロジーに応用し、ネットワークを基盤として毒性の理解を深める



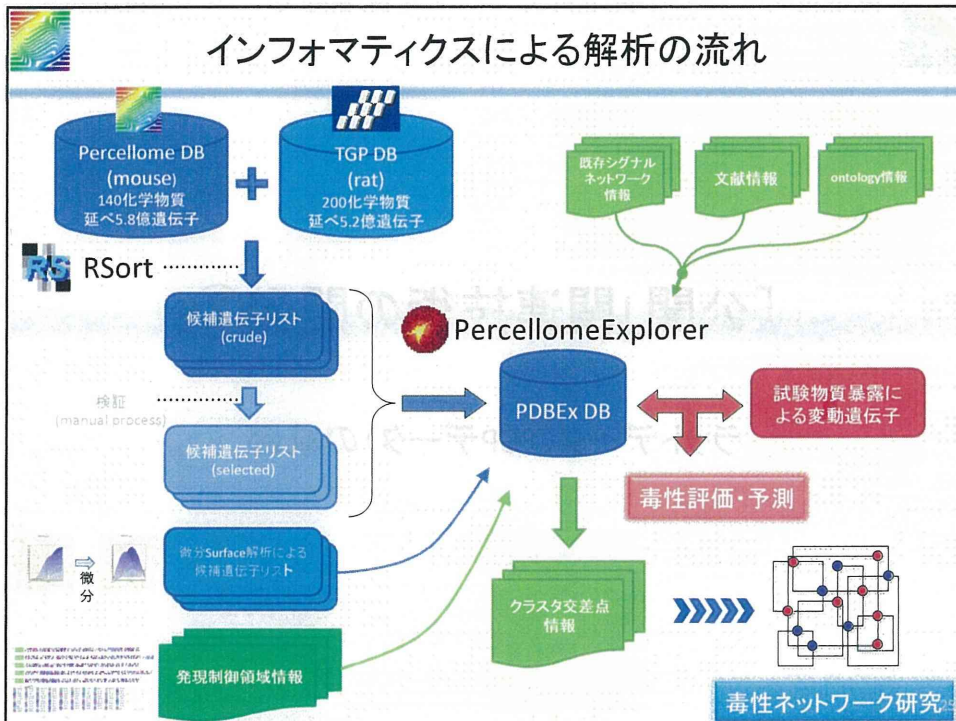
(例) 相互情報量(Mutual Information)による遺伝子制御関係推計

分子メカニズムに依拠した、化学物質による有害性評価手法の迅速化、高度化の為にインフォマティクス開発

国際的なソフトウェア共通基盤



24



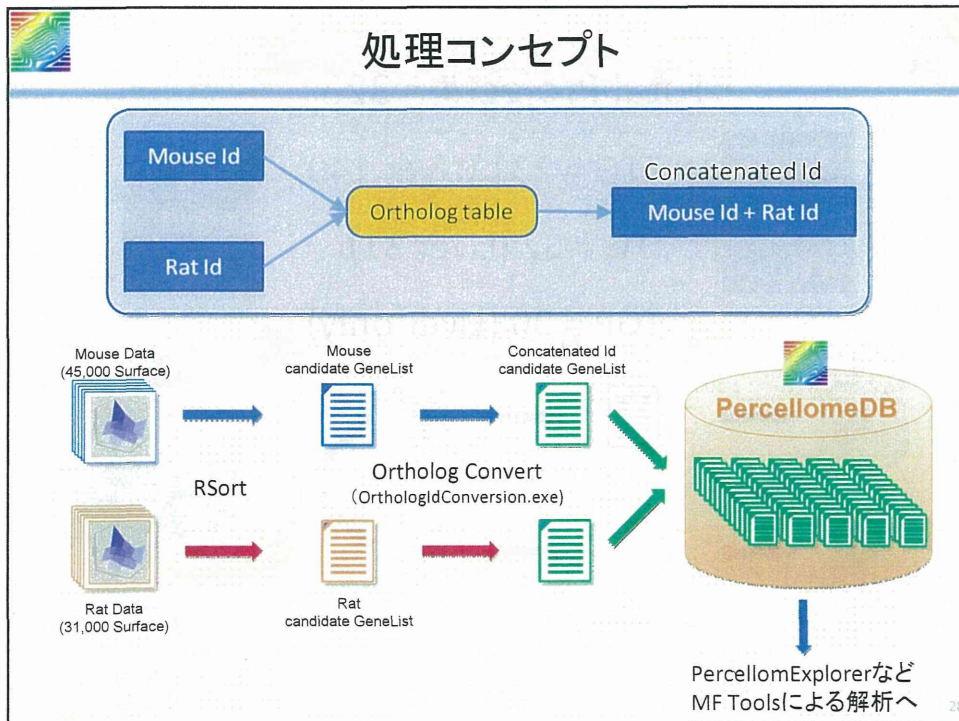
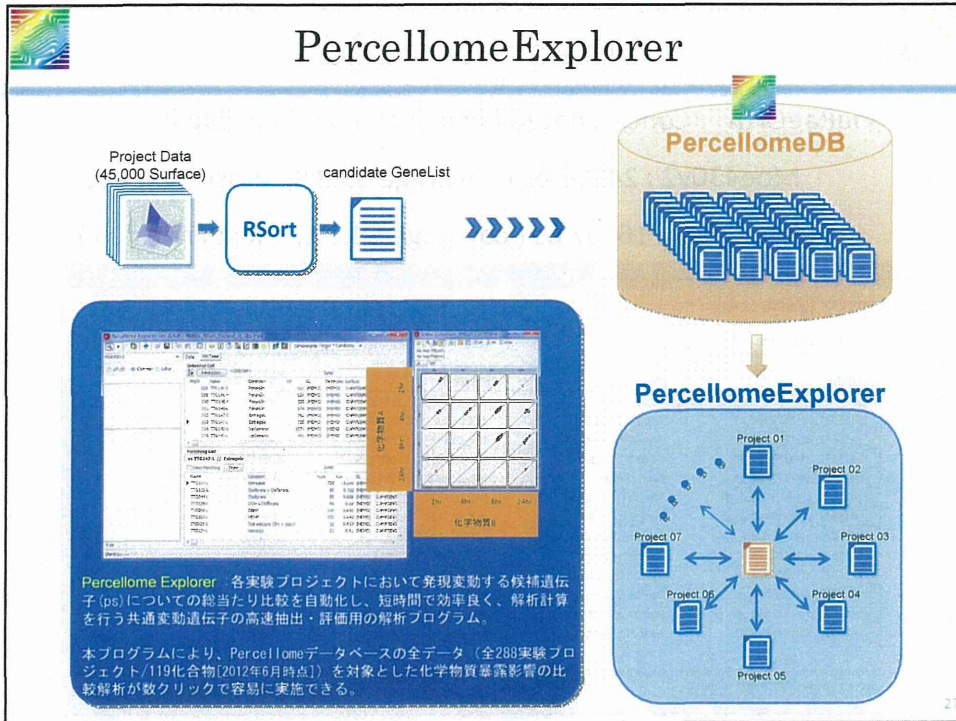
RSort

発現パターンの凹凸の程度・特徴点の位置(=実験条件)を評価する。
 → 特定の実験条件において、明確な発現誘導パターンを呈する遺伝子を抽出できる。

(peak数)

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

26



Concatenated Id

MoeRaeOrth : Concatenated Id (= Ortholog Pair, 39246)

Moe430v2 : 24950 ps (coverage 55.3%, most Freq =20)

Rae230v2 : 16867 ps (coverage 54.2%, most Freq =26)

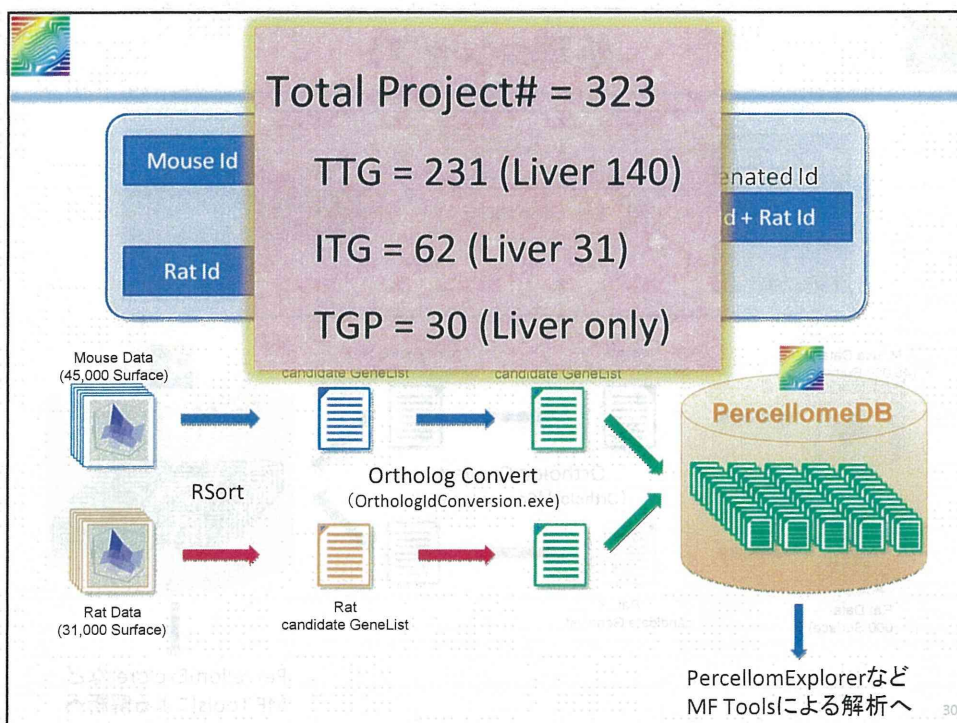
Table - MoeRaeOrth (C:\MilleFeuille\MoeRaeOrth)

Browse Data | Structure | Indexes | Full Text Indexing


Select Visible Fields... Copy On Append

AffyId	Common	GenBank	Description
1415670_at-1371810_at	Copg Copg	BC024686 BF284093	(MEMO)
1415670_at-1372464_at	Copg Copg	BC024686 BE109050	(MEMO)
1415671_at-1388365_at	At6v0d1 At6v0d1	NM_013477 BI276424	(MEMO)
1415672_at-1398895_at	Golga7 Golga7	NM_020585 BM389419	(MEMO)
1415673_at-1375964_at	PspH PspH	NM_133900 BF282282	(MEMO)
1415674_a_at-1373907_at	Trappc4 Trappc4	NM_021789 BI288890	(MEMO)
1415676_a_at-1371348_at	Psmb5 Psmb5	NM_011186 BI285627	(MEMO)
1415676_a_at-1385531_at	Psmb5 Psmb5	NM_011186 BI281524	(MEMO)
1415676_a_at-1385532_x_at	Psmb5 Psmb5	NM_011186 BI281524	(MEMO)
1415677_at-1389407_at	Dhrs1 Dhrs1	NM_026819 BE099453	(MEMO)
1415678_at-1368859_at	Ppm1a Ppm1a	BC008595 NM_017038	(MEMO)
1415679_at-1371338_at	Psenen Psenen	NM_025498 AI232272	(MEMO)
1415680_at-1373744_at	Anapc1 Anapc1	NM_008569 BF548006	(MEMO)

If the current table contains BLOB fields, simply double-click on the field to view and/or modify the field contents Last Updated On 2012年12月15日 1:08:01

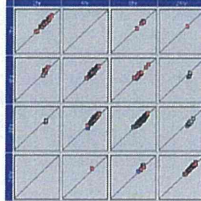


PercellomeExplorer



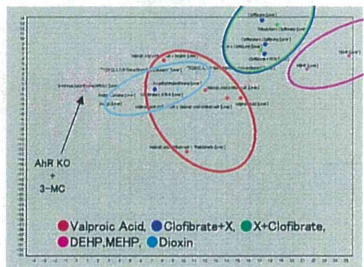
既知のネットワーク情報やGO情報との照合処理も可能

時系列を考慮した誘導遺伝子比較
(例)食品添加物Sodium Dehydroacetateと
Acephateの高い類似性を発見



Acephate

Sodium Dehydroacetate




誘導遺伝子情報による
化学物質の主成分解析図

AHR KO
3-MC

- Valproic Acid, ● Clofibrate+X, ● X+Clofibrate,
- DEHP,MEHP, ● Dioxin

>>>> 全データベースを対象とした時間経過と用量を加味した解析が容易に実施できるようになった。
また解析経過及び結果の可視化により、データ検証の精度とスピードが格段に向上した。

31



PE-PCA

①

Expand-G3_Std-Av_MoeRaeOrth

Liver Only

Non-Strict, Target mode

32

