

トキシコゲノミクスにおける技術の標準化:Percellome

“Percellome” a standardization method for toxicogenomics



菅野 純(写真) 相崎健一 中津則之 五十嵐勝秀

Jun Kanno, Ken-ichi Aisaki, Noriyuki Nakatsu and Katsuhide Igarashi

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

◎形質非依存型トキシコゲノミクスに適用するため、マイクロアレイなどから細胞1個当りの mRNA コピー数として発現量を得る方法(Percellome)を開発した。この結果、検体間はもとより実験間、マイクロアレイのバージョン間、異なったプラットフォーム間、あるいは異なった臓器や動物種間のデータ比較が容易となる。完全教師なしクラスタリングなどのデータ解析法の開発を加え、この手法を毒性メカニズム研究を進める手がかりとし、これに基づいた毒性評価、毒性予測のあらたな展開をめざしている。



Key word : マイクロアレイ技術, トキシコゲノミクス, 分子毒性学, 標準化, 化学物質安全性評価

毒性学は、それが現象の記述学に基礎をおいていた段階では、black box である生体に対して投与した化学物質とそれが引き起こした症状との関連性に基づいた化学物質の体系化を基盤として発達してきた。その過程で“不確実係数”や“LD₅₀”の概念がさまざまな経験を取り入れる形で利用され、現在まで有効に機能してきている。しかし、生命科学の進歩により毒性学も生体内における分子レベルから形態レベルまでのメカニズム記述を基礎とするものへと変わりつつある。著者らはこれに対応すべく transcriptome データのもっとも効率的な取得・解析法を検討してきた。その結果として“形質非依存型トキシコゲノミクス(phenotype-independent toxicogenomics)”を導入したのは、明瞭な形質発現を伴わない段階や状況において生体内に生じているシグナル伝達カスケードの変化を標的とした mRNA 解析が必要であると判断したことによる。

形質非依存型トキシコゲノミクス

分子毒性学の立場から一番知りたいことは生体

内で実際に起こっている一連の事象であり、transcriptome の場合にはすべての遺伝子の情報をもとにした遺伝子カスケードの全容解明である。これが分かれば究極的には膨大な時間と費用のかかる定型的な長期毒性試験(ラットなどを用いる)に比べ、より早く、安く、正確な評価、種差や個人差を勘案した正確なヒト毒性予測が可能となることが強く期待される。とくに胎児、新生児、小児、成人、老人の各発達段階における生体側の反応様式・感受性の変化や、複数の物質の進入による複合作用なども包括的に扱うことができるようになると考えられる。このようなことから、今後の毒性学における transcriptome 解析、すなわちトキシコゲノミクスは従来の“形質依存型”¹⁾のものから“形質非依存型”に発想を転換する時期にきているといえよう。

PercellomeとMillefeuille data

形質非依存型トキシコゲノミクスに適用するため、著者らは、細胞1個当りの mRNA コピー数として発現量を得る Percellome 法を、当時それに必

要な条件を満たしていたアフィメトリクス社の GeneChip を対象に開発した(特許出願中, 投稿中). このシステムは大きく 4 つの要素からなっている.

第 1 に RNA 用に準備したサンプル破碎液の極一部からその DNA 濃度を簡便に測定する方法, 第 2 に用量関係を考慮し工夫されたスパイク RNA 液の調整と, それの破碎液への添加法, 第 3 に応答特性に基づいた絶対量化アルゴリズム, そしてマイクロアレイの用量相関性能の検証や, データ変換に用いる標準サンプルセットとそのデータ変換アルゴリズムである. Percellome データは, 細胞 1 個当りの mRNA コピー数であるので, ゼロを起点とする均等目盛りで各遺伝子の発現量をマイクロアレイ間ではもとより, 実験間において直接比較することが可能となった. 後述するトキシコゲノミクスのように用量と時間などの実験要素を二次元に展開し, そのうえに Percellome データを表示する三次元表示(Millefeuille data)を行うことにより生物学者が遺伝子変動を直感的に把握しやすくなり(図 1-3), その後のデータ解析とインフォマティクス形成に大きく貢献することが示されつつある. この Percellome の特徴は実験結果を長きにわたり逐次蓄積する必要があるトキシコゲノミクス研究には重要なことである. また,

後述するように異なったプラットフォーム間でのデータ互換にも拡張可能であり, 共通の大型データベースやコンソーシアム構築にも貢献する可能性が高い.

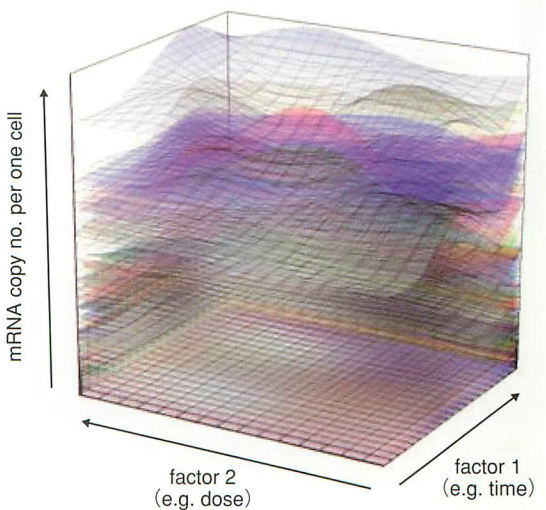


図 1 PercellomeデータのMillefeuille表示

x 軸に時間, y 軸に用量などの, 実験条件を展開し, z 軸に細胞 1 個当りの Percellome データをプロットし, 地図のような曲面を描画することにより単一の三次元グラフにすべての probeset データを表示することができる[多層であることから, これを Millefeuille(MF)データと名づけた]. MF 表示により生物学的反応を直感的に把握しやすくなり, データ解析に有用であることを経験している.

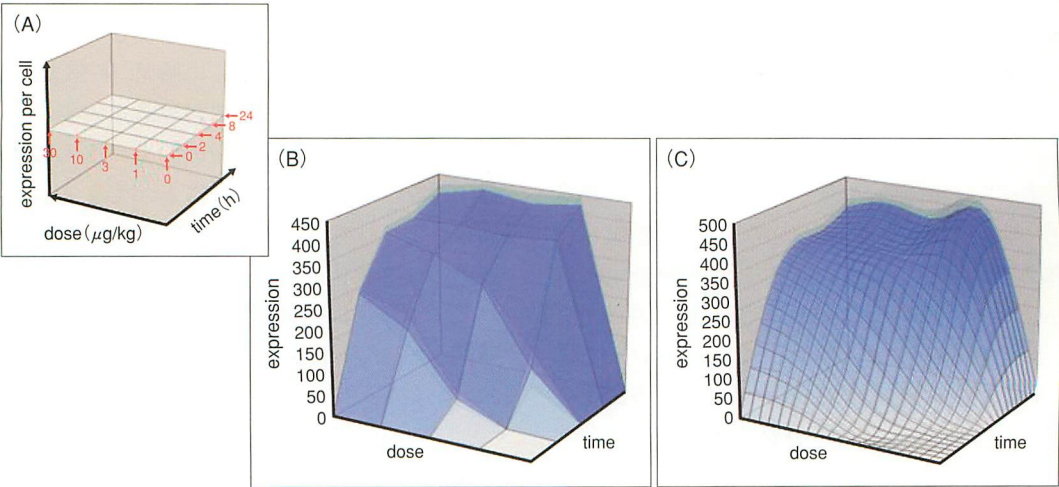


図 2 遺伝子発現の表示例(1)

2,3,7,8-TCDD(溶媒: コーン油)を単回経口曝露した際の, C57BL/6 雄マウス肝における遺伝子発現の表示例(用量と投与後時間と表示の関係を A のグラフに示す). Cyp1a1 遺伝子の初期応答を格子点表示(B)とスプライン表示(C)にて表示した. 2 時間後に用量依存的に Cyp1a1 遺伝子発現が誘導されている(平均曲面±1sd 曲面).

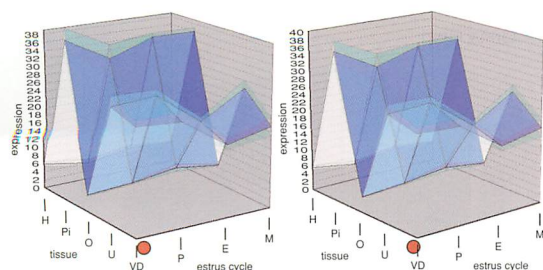


図 3 遺伝子発現の表示例(2)

C57BL/6 雌マウス(12 週齢)の性周期による視床下部, 下垂体, 卵巣, 子宮および陰の ER α 遺伝子の発現変化のステレオ三次元表示例(交差視線で赤丸を重ねる。H: hypothalamus, Pi: Pituitary, O: ovary, U: uterus, V: vagina, D: diestrus, P: proestrus, E: estrus, M: metestrus)。縦軸は細胞 1 個当りの平均発現量(Percellome)。ER α は, 下垂体, 子宮, 陰に多く発現し, 後二者では estrus に軽度発現が減少することが示される(平均曲面 ± 1 sd 曲面)。

方法の概略

① DNA測定……細胞 1 個当りの mRNA 情報を得るために, サンプルを構成する総細胞数を測定する。実際に細胞数を計測することはとくに実質臓器の場合には困難であるため, その代替指標として細胞核内のゲノム DNA 量を用いる。サンプルを DNA 測定専用消費することを避けるため, RNA 調整用の組織破碎液の極一部(通常, 10 μ l)を DNA 測定に用いるプロトコルを確立した。

② 多段階濃度スパイクカクテル(GSC: dose-graded spike cocktail)……細胞 1 個当りの mRNA の標準として組織破碎液に添加するスパイク RNA にはアフィメトリクス社の GeneChip が使用者のために用意していた 5 種類の枯草菌由来遺伝子の RNA を用いた。5 種類の枯草菌 RNA をそれぞれ適切な長さに合成し, 5 段階の用量に配合したカクテルを作製した。これにより広い濃度範囲をカバーする標準用量作用曲線をすべてのサンプルに導入することが可能となった。

③ 絶対量化プログラム……アフィメトリクス GeneChip は蛍光シグナルと mRNA 量と両対数軸上で直線関係にあることを後述の LBM 標準サンプルなどにより確認した。その直線化式により GSC を直線化して絶対量化を行う変換アルゴリズムを開発し, それを自動実行するプログラムを独自に開発した(相崎)。

④ GeneChipの用量相関性確認およびバージョン間・プラットフォーム間データ変換対応のためのLBM(liver-brain mix)標準サンプルおよびデータ変換アルゴリズム……遺伝子発現プロファイルが大きく異なる一対の組織を一定の比率で相互に希釈しあったサンプルセットを表記の目的のために用意した。ここでは肝と脳を用い, 100:0, 75:25, 50:50, 25:75, および 0:100 の混合比の 5 サンプルからなるセットを用意した。

絶対量化の原理

基本的原理はサンプルの細胞数(ゲノム DNA 濃度で代替)に比例した分子数のスパイク RNA を添加することで, サンプルの細胞 1 個当りの mRNA 絶対量(コピー数)の指標をサンプル中に導入するものである。ただし, スパイク RNA は 1 点を規定するものではなく, 5 種類の枯草菌遺伝子に対する RNA(哺乳類の配列と交差しない)を適切な公比をもたせて 5 段階の濃度に割り振ったカクテルとして用いることが特徴である^{2,3)}。これにより絶対コピー数の指標となると同時に, 広い用量範囲について検量線を各サンプルに導入したことになり, mRNA 抽出から GeneChip の蛍光測光までの過程で生じるデータ全体の歪みを補正する際に威力を発揮するとともに, すべての GeneChip の発現値を統一基準下で安定的に絶対量化する効果を有している。サンプルに由来するすべての測定値は直線化された GSC 検量線に基づいて絶対量に変換される。

データ互換性他システムへの拡張

本システムの GSC を添加したサンプルはスパイク RNA を検出するプライマーセットを用意することで定量 PCR においても容易に絶対量化データを得ることができる。詳細は他に譲るが, Percellome 定量 PCR システムを完成し, 現段階までに比較検討した約 40 遺伝子のうちの 90%以上が GeneChip データとほぼ同一の結果を示している(図 4)。また, アフィメトリクス GeneChip 以外のプラットフォームとの間, あるいは異なった増幅プロトコル間のデータ互換も可能である。本システムが適応可能なプラットフォームの条件

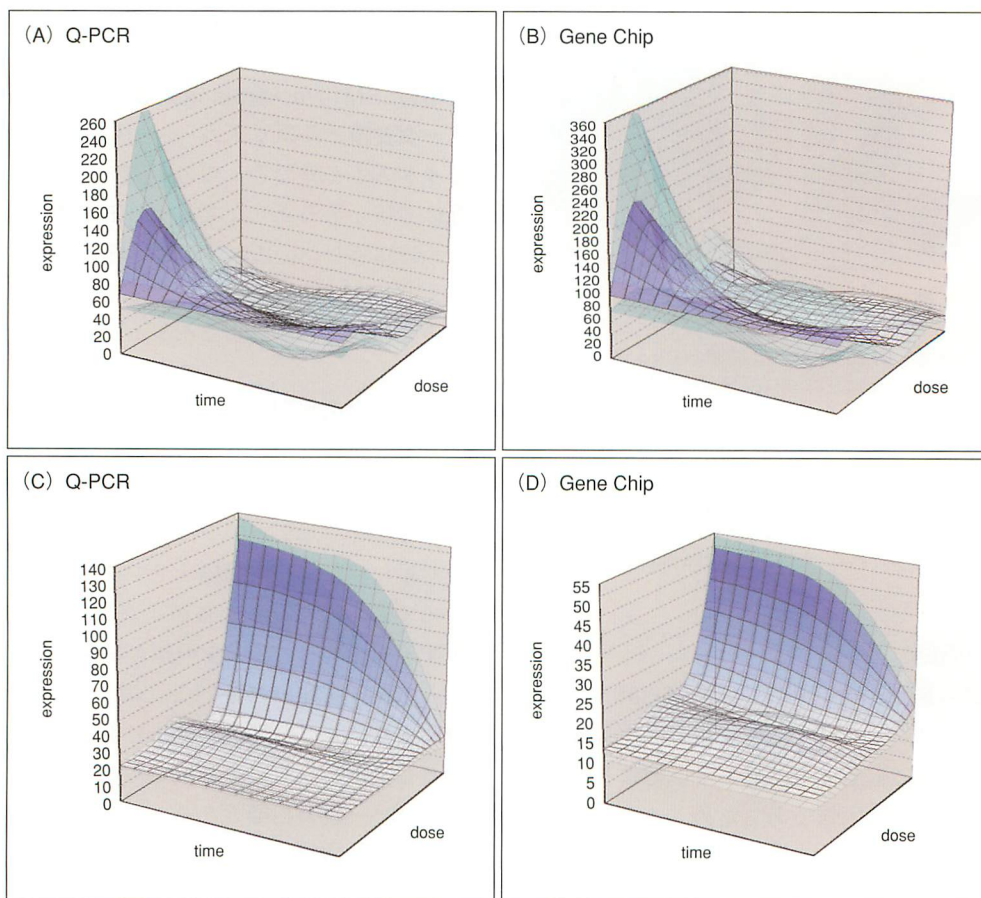


図 4 定量PCR PercellomeとGeneChip (Affymetrix) の比較

図 2 の TCDD 実験の 60 サンプルについて SYBR Green を用いた定量 PCR (ABI PRISM7900HT) に Percellome 法を適応し MF 表示した (A, C)。ほぼ同等の反応曲面が得られる。

としては、GSC を受け付けるプローブセットが用意されていること、および用量相関性が確保されていることの 2 点を満たしている必要がある (図 5, 6)。

遺伝子カスケード解析をめざした形質非依存型トキシコゲノミクスへの適用

厚生労働科学研究費補助金のプロジェクトにこの Percellome システムが採用され、進行中であり [厚生労働科学研究費補助金 H14-トキシコ-001 (創薬支援トキシコゲノミクス) および H15-化学-002 (化学物質トキシコゲノミクス)], 両者とも 4~5 段階の用量について 4 時点での遺伝子発現を観測する 16~20 群 (一群 3 匹) の構成からなるプロトコールにより、ひとつの化合物について 48~60 匹の動物のサンプルから Percellome デー

タを生成している。とくに後者の化学物質トキシコゲノミクスプロジェクトでは X 軸に時間、Y 軸に用量、Z 軸に発現量 (ゼロからの均等目盛り表示) をプロットして得られる三次元曲面表示を行い、GeneChip が 45,000 のプローブセットからなる場合、ひとつの化合物のデータは 45,000 枚の局面の層状集合体からなる [多層構造からなる菓子などに準えミルフィーユ・データ (Millefeuille (MF) データ) と名づけた]。

この MF data は曲面 (平均曲面と ± 1 sd 曲面) の各格子点が 3 匹の動物に由来する 3 つのデータに基づいており、格子点のデータの信頼性の評価、artifact の除去や、生物学的な蓋然性を有する変化であるかどうかの判別に適しているうえに、類似の反応を示す遺伝子の選別に威力を発揮する。生体反応の分子メカニズム解析 (カスケード解析) を

図 5 異なったマイクロアレイ間のデータ比較

現在評価中の2社のマイクロアレイによるLBM標準サンプルセット(liver-brain mix)のPercellomeデータ、中央のliver:brain=50:50のサンプルに対する比率で表示し、直線性のよい(R^2 が1に近い)トップ250遺伝子(プローブセット)を示す。直線性のよい遺伝子が多く存在することが確認されたことから、各社間のデータ互換が可能となることが期待される(互換法については図6参照)。今後、4×4群実験(16格子点からなる)のサンプルを用いた評価を行う予定である。

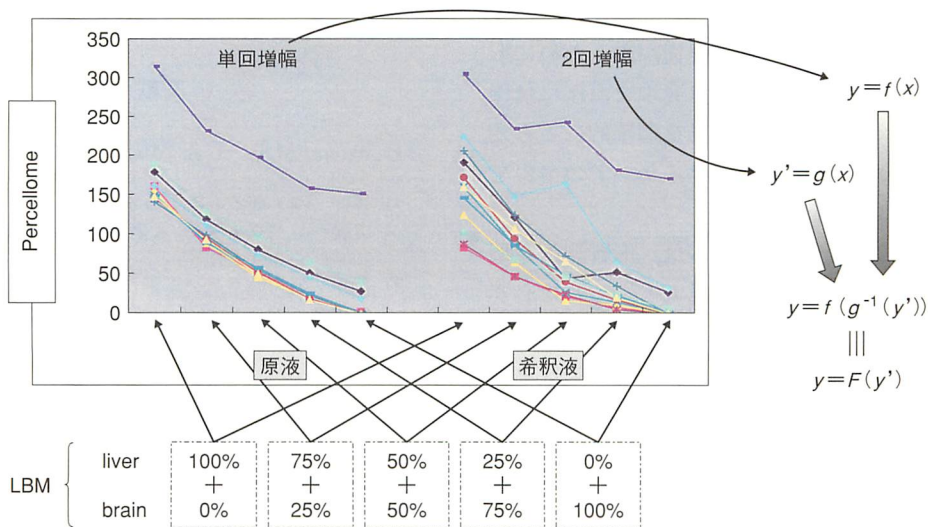
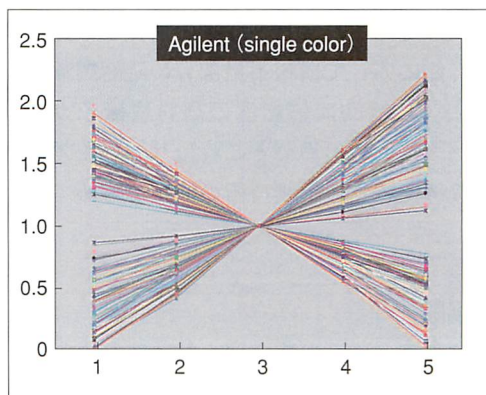
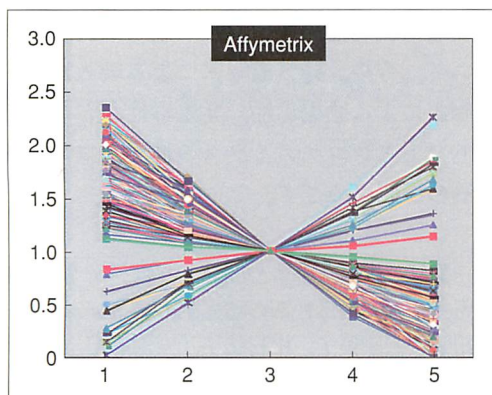
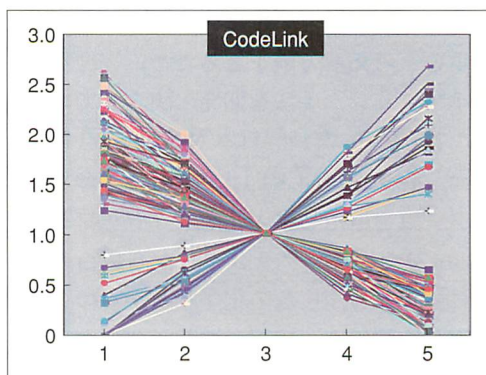


図 6 LBMを用いたデータ互換法

単回増幅プロトコルと2回増幅プロトコルとの間での例を示す。LBMの濃厚原液を単回増幅プロトコルにて希釈液を2回増幅プロトコルにてPercellome測定し、両方において良好な用量相関性が得られたプローブセットについては $y=f(x)$ 、 $y'=g(x)$ なる関数を割り当てることができ、 x が共通であることを利用して、 y が y' の関数として一義的に決定される。

目標に、遺伝子欠失動物の活用を見込んで、マウスを用いた実験を重ね、2005年1月現在で約50

の既知化合物についてのデータ収集を終え、向こう1年のうちに目標の90化合物データを蓄積す

る予定である。解析においては、すべての遺伝子が平等に重要と考える方策をとるため、完全な教師なしクラスタリング手法を開発・実施している (NTT コムウェア株式会社/日本 NCR 株式会社との共同開発, Teradata(NCR) 上に搭載)。この手法はクラスター数を指定せず、45,000 プローブセット (MOE430v2) を小さいクラスターから数百のクラスターに分類する。複数の化学物質からのクラスターデータの解析と、適切な遺伝子欠失マウスによる MF data により客観的な遺伝子カスケードの描出を試みている。これと既知の情報との比較を行い、必要に応じて情報不足部分の確認実験を追加実施し、最終的に信頼性の高い遺伝子カスケードデータベースの構築とこれに基づいた効率的で正確な毒性評価・予測技術の開発を進めている。

今後の展開

毒性学は毒性という形質発現を基に成り立っているが、それを用いた毒性評価の効率化と正確性向上のための分子毒性学的なメカニズム解析の導入を計るにあたり形質が発現する以前の段階、あるいはフィードバック機構が働くために形質発現が乏しい状態など、形質発現を直接には伴わない状況での遺伝子発現変動を網羅的にとらえて毒性にかかわる遺伝子発現カスケードの全容を解明する必要性に迫られている。

これに応えるために、著者らは、形質非依存型トキシコゲノミクスの概念の導入と、それに必要な技術であるマイクロアレイや定量 PCR から細胞 1 個当りの mRNA 絶対量情報を生成する Per-

cellome システムを開発し、生物学者にとってわかりやすい三次元 MF data として表示する方法との組合せにより生物学的蓋然性に基づいた検証を進めている。本システムは大型のトキシコゲノミクス・プロジェクトを対象として開発したものであるが、実際には小規模の実験サンプルに対しても、さらに、global 標準化手法の適用が難しい少数遺伝子の発現変動をみるシステムに対しても有用性が示されている。現在、完全教師なしクラスタリングのつぎのステップとしてカスケード描出に必要な情報を抽出するアルゴリズムの開発を進めている。今後、経気道曝露による吸入毒性、経胎盤曝露による胎児毒性、発生毒性などを加えた総合的なトキシコゲノミクスの構築とコンソーシアムへの展開をめざす。

謝辞：本システムの開発とプロジェクトの遂行にあたっては、当毒性部の諸先生方および安東朋子、森山紀子、近藤優子、中村祐子、安部麻紀、吉木健太、松田菜恵、森田紘一、今井あや子、青柳千百合、相原妃佐子の各氏に深謝する。本研究は、厚生労働科学研究費補助金 H13-生活-012, H13-生活-013, H14-トキシコ-001 および H15-化学-002 による。

文献

- 1) Waters, M. D. et al. : *Mutat. Res.*, **544** : 415-424, 2003.
- 2) 菅野 純・他：トキシコゲノミクス。ゲノム研究実験ハンドブック (辻本豪三編)。実験医学別冊, 2004, pp.329-337.
- 3) 菅野 純・他：ゲノム毒性学形質非依存型トキシコゲノミクスの導入。細胞工学, **23** : 685-693, 2004.

* * *