

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 6 月 15 日現在

機関番号：82601

研究種目：基盤研究（C）

研究期間：平成 21 年度～平成 23 年度

課題番号：21500292

研究課題名（和文） 各々のマイクロアレイによるエクソン毎の遺伝子発現データの
絶対定量化技術の開発研究課題名（英文） Establishment of a method to measure absolute expression levels of
all exons using microarrays

研究代表者

相崎 健一（AISAKI KEN-ICHI）

国立医薬品食品衛生研究所・毒性部・第一室長

研究者番号：40322086

研究成果の概要（和文）：

エクソン用マイクロアレイを含め、mRNA の測定に關しての現状では相対的変動のみが把握可能である。これはスプライシングバリエーション研究の障害となっているため、各エクソンの発現量を細胞一個あたりのコピー数として精密測定する方法を検討した。その結果、半特異的ハイブリダイゼーション（核酸配列の部分的一致によるクロスハイブリダイゼーションの一種）による誤差を排除する必要を見だし、配列情報を基に in silico で補正する基本理論を生成した。

研究成果の概要（英文）：

Including exon-microarray, current microarray technology can detect only relative changes of mRNA expression. We tried to develop a new method to measure amount of each exon expression as copy number per one cell, for enhancing the performance study of splicing-variant in various situations. In the course of the study, we have revealed that serious errors are caused on exon microarray by “semi-specific hybridization”, a kind of cross-hybridization among nucleotides sharing partially common sequence. As a result, we generated a new basic theory which can improve quality of exon-microarray data based on the sequence data of probes and targets taking the semi-specific hybridization into account.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	1,800,000	0	1,800,000
2010 年度	1,300,000	0	1,300,000
2011 年度	500,000	0	500,000
総計	3,600,000	0	3,600,000

研究分野：総合領域

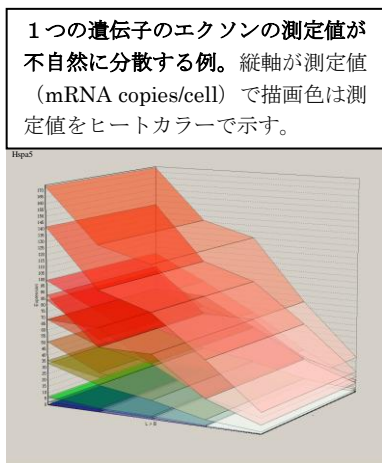
科研費の分科・細目：情報学、生体生命情報学

キーワード：バイオインフォマティクス、ゲノミクス

1. 研究開始当初の背景

ゲノム配列から1遺伝子あたりの平均エクソン数は30との推計され、遺伝子発現解析を行う際に多数のスプライシングバリエントが存在する可能性を無視する事は出来ない。このようなエクソン数の多い遺伝子の発現解析を行う場合、とりわけ複数の遺伝子を検討する場合や事前にバリエントの情報の無い場合、Northern blot や定量RT-PCR は、プローブの設定の問題などから、非効率的である。研究開始当初、最も現実的だったのは全遺伝子のエクソン配列に対する高密度エクソン用マイクロアレイを用いる方法である(2012年時点では次世代シーケンサによるRNA シークエンスが利用可能になってきたが、まだマイクロアレイのほうが低コストである)。しかし、残念ながら、標準的プロトコルでは相対的変動のみが把握可能であり、各エクソン発現量の絶対比較が困難である。このことは多くの遺伝子で、そのスプライシングバリエントが右図のように不自然に

分散することを意味し、網羅的なスプライシングバリエント解析の大きな障害と



なっていた。

2. 研究の目的

エクソン用アレイを用いてスプライシングバリエントの発現と機能を網羅的に研究するための基盤技術として、高精度測定技術の開発が必要とされている。そこで我々が3' Expression マイクロアレイ用に開発した絶対定量化手法(Percellome法:Kanno et al. BMC Genomics. 7, 64, 2006 / 特許第 4415079 号)をエクソン用アレイに適用し、さらにエクソン用アレイで発生している誤差を補正して、各エクソンの発現量を細胞一個あたりのコピー数として精密測定し、以てスプライシングバリエントの発現量を絶対量で解析する方法を開発する。

3. 研究の方法

(1)肝臓-脳混合 RNA サンプル(LBM)

プローブの基本性能評価や誤差補正アルゴリズムの性能評価用として、マウスの肝臓と脳から調整した混合サンプル(LBM)を使用した。具体的には、12週齢・雄のC57BL/6を麻酔下放血により安楽に屠殺し、肝臓と脳を摘出、約4mm 経に切り分け、直ちにRNAlater 溶液(Life Technologies 社)に投入し、RNA を保護した。4°C 12時間以上つけ込んだ後、RNAlater 溶液を除去したのち、臓器片にQIAGEN RNAeasy のホモジナイズ液を加えて破碎した。肝臓と脳それぞれの破碎液を10uL ずつ分取し、RNase,

Proteinase 処理の後、Picogreen 色素 (Life Technologies 社)を用いて DNA 含量を測定した。肝臓、脳それぞれの破碎液の DNA 含量を基準に、100%:0%, 75%:25%, 50%:50%, 25%:75%, 0%:100%の割合で混合した 5 種類の混合液を調整し、DNA 含量に比例した適量の Percellome 独自の外部スパイク RNA (Graded-dose spike-cocktail: GSC)を添加した。

(2)Affymetrix エクソン用アレイ解析

3' Expression アレイや定量 RT-PCR を主な対象とした Percellome 標準プロトコル (BMC Genomics,29,64,2006)を、ランダムプライマーを用いて cDNA 合成するエクソン用アレイに最適化し、これに則って各エクソンに対応したプローブセットごとに発現量を測定した。

(3)リアルタイム PCR による定量 RT-PCR

エクソン用アレイに用いたものと同じ混合サンプル由来の RNA 溶液を分取し、ABI PRISM 7900HT にて、ABI の標準プロトコルに従って、GSC と一部の遺伝子を対象とした RNA 定量測定を行った。測定データの数値化は定量 PCR 用の Percellome 標準プロトコルを適用して実施し、細胞 1 個あたりのコピー数を測定した。

(4)データ解析

マイクロアレイや定量 RT-PCR のデータは、raw 値と Percellome 法を適用した絶対量化値を用いた。解析やアルゴリズム実証に用いるプログラムは、Excel、R など市販もしくはオープンソースのソフトウェアのほか、研究者自身が Delphi 言語 (=object pascal 言語、開発環境としては Borland

Delphi7 もしくは Delphi2010, Delphi XE2)を利用してプログラム開発し利用した。1000 万件以下の小・中規模のデータ操作には組み込み型リレーショナルデータベースコンポーネントである DBISAM (ElevateSoftware 社)を利用して、プログラム開発の効率化を図った。またそれ以上の大規模計算が必要な場合は、専用計算サーバーシステム (MF サーバーシステム)にアプリケーションソフトウェアを移植して、大規模計算用グリッド PC や高速データベースエンジンである Teradata データベース (日本 Teradata 社)を活用して計算処理を行った。

4. 研究成果

(1)エクソン用アレイの定量性能検証

Affymetrix 社製のエクソン用アレイの基本性能を確認する為に、LBM サンプル 5 点 (肝臓 : 脳 = 100%:0%, 75%:25%, 50%:50%, 25%:75%, 0%:100%)を実際のエクソン用アレイを用いて測定した。LBM サンプルの測定結果は、忠実度の高い測定系ならば全ての mRNA 分子が混合比に応じた直線的な量的分布を呈するはずであり、エクソン用アレイについても、同社 3'expression アレイと同等の定量性能を有することを確認した。しかしながらこの測定データにおいても 1 つの遺伝子に属するエクソンの測定値がバリエーションの結果とは考えにくい不自然なパターンに分散する現象が多数見られた。

(2)定量 RT-PCR データや次世代シーケンサデータを利用した補正の試み

スプライシングバリエーションの存在が分かっ

ている一部遺伝子(エストロゲン受容体など)を対象として、LBM サンプルの定量 RT-PCR 測定を行い、エクソン用アレイデータとの誤差を補正する為の係数をエクソン毎に得た。LBM 以外のサンプルを用いて、補正性能を検証したところ、特に肺など肝臓・脳以外の臓器由来のサンプルにおいて十分な補正性能が得られなかった。

同様に LBM サンプルを次世代シーケンサ (イルミナ社 Genome Analyzer IIx) を用いて網羅的に mRNA 発現量を測定したデータを手に入れ、同様にエクソン毎の補正用係数を得たが、定量 RT-PCR データを参照した場合と同様、肺などでは十分な補正性能が認められなかった。

これらの結果は、肝臓・脳以外の臓器では LBM サンプルとは発現している遺伝子の種類や発現量が異なるためと考えられた。

(3)エクソン用アレイにおける半特異的ハイブリダイゼーションの確認

2 の結果より、1 つの遺伝子のエクソンの測定値が不自然に分散する現象が、個々のエクソン用プローブの性能に依るものではない可能性を見出した。LBM サンプルをエクソン用アレイで測定したデータを個々のプローブレベルで精査したところ、直線パターン、飽和パターンに加え、クロスハイブリダイゼーションを示唆するパターン(肝臓と脳を 75%:25%、50%:50%、25%:75%で混合したサンプルにおいて、肝臓 100%と脳 100%の両サンプルで測定したデータから予測される理論値より低い数値となるパターン。これはハイブリダイゼーション工程で誤結合した核酸断片が、

洗浄工程で洗い流される結果、最終的な測定値が低下する為に発生する)を検出した。

該当プローブの配列情報を確認したところ、部分的に核酸配列が類似する遺伝子が複数存在した。つまり、ハイブリダイゼーション工程において、これらの遺伝子由来の cDNA がプローブに誤結合し、アレイ設計上結合すべき遺伝子由来の cDNA の結合を阻害していることが示唆された。

(4)半特異的ハイブリダイゼーション影響の除去

3' expression アレイにおける半特異的ハイブリダイゼーション影響の補正を行う為のアルゴリズム MLANG (論文準備中、特願 2010-294175)は、実測データを用いて補正係数学習を行い、補正係数を取得するシステムであり、3' expression アレイの設計特性に最適化して設計されている為、そのままエクソン用アレイに用いることはできなかった。特に①ハイブリダイゼーション工程においては DNA-RNA 結合ではなく、DNA-DNA 結合が起こること、②プローブ設計時の参照配列は 3'端領域だけではなく、遺伝子全長に及ぶこと、による影響が大きかったため、①については学習ルーチンのパラメータ変更により、また②については補正係数の初期値算出ルーチンの拡張で対応した。具体的には、まず全プローブ配列の総当たり比較を行って類似配列対応を抽出した上で、プローブ毎に補正係数の初期値を計算し、次に実際のエクソン用アレイ測定データを用いて学習ルーチンによる補正係数の最適化を試みた。

しかしながら、遺伝子の 3'端領域のみを対象とする 3' expression アレイと異なり、エクソン用アレイ

イは全遺伝子領域を対象とするため、プローブと半特異的に結合する遺伝子配列領域が極めて多く、半特異的ハイブリダイゼーション発生の推計に必要な計算量が既存の計算用サーバーにとって過大となってしまう計算工程を完遂できなかった。

そこでエクソン用アレイのプローブ設計上の特徴(1つのエクソン辺り4つのプローブを設計する為、設計対象領域が狭い事が多く、一部が重複した隣接プローブが多数存在することなど)を活かして疑似的にプローブ数を減少させ、半特異的ハイブリダイゼーション発生推計の計算回数を省いて計算効率を上げた。さらにエクソン領域情報の確度から計算対象とするプローブを選択し、総計算量を低減させた。これらの結果、エクソン用アレイにおける計算量を3'expressionアレイにおける計算量の数十倍まで減らすことに成功し、エクソン用アレイデータの *in silico* 補正の実用化に目処を付けた。

(5)結論

エクソン用マイクロアレイにおける主要な誤差発生要因を同定し、基本的な補正方法も考案した。実際の計算を完遂できていないが、これは利用可能な計算サーバーの記憶容量に依存する問題であり、より大規模なシステムを利用すれば解決される。またリソースの少ない計算サーバーでも計算を完遂できるよう、計算アルゴリズムの改良を進め、実現の目処も立っている。

今後、計算アルゴリズムの改良を進め、現行の計算サーバーでも実行可能なエクソン用マイクロアレイデータの補正アルゴリズムを完成させる。

本アルゴリズムは、測定済みデータについても補正計算が可能である点に特徴があり、今まで蓄積されてきたエクソン用アレイデータの高精度化が期待され、スプライシングバリエーション研究を始め、各エクソンの発現量変動情報を利用する幅広い応用研究の推進に貢献するものと期待される。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計7件)

[学会発表](計14件)

①相崎 健一、Percellome プロジェクト・オンライン解析システム、第38回日本トキシコロジー学会学術年会、2011.7.13、横浜

②相崎 健一、Percellome Project における Bioinformatics の進展、日本バイオインフォマテイクス学会応用システムバイオロジー研究会第3回応用システムバイオロジー研究会(招待講演)、2011.6.27、福岡

[図書](計2件)

①DARRELL R. BOVERHOF、Applications of Toxicogenomics in Safety Evaluation and Risk Assessment、John Wiley & Sons, Inc.、323-330、2011

② Ingemar Pongratz、Hormone-Disruptive Chemical Contaminants in Food、Royal Society

of Chemistry, 184-198, 2011

[その他]

ホームページ等

①<http://percellome.nihs.go.jp>

6. 研究組織

(1)研究代表者

相崎 健一 (AISAKI KEN-ICHI)

国立医薬品食品衛生研究所・毒性部・第一

室長

研究者番号:40322086