



Title	乳癌トランスクリプトームデータを用いた遺伝子発現と融合遺伝子の解析手法に関する研究
Author(s)	草田, 義昭
Citation	大阪大学, 2022, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/88158
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

論文内容の要旨

氏名（草田義昭）	
論文題名	乳癌トランスクリプトームデータを用いた遺伝子発現と融合遺伝子の解析手法に関する研究
論文内容の要旨	
<p>遺伝子転写産物（mRNA）全てを扱うトランスクリプトームデータは、マイクロアレイやRNAシーケンス（RNA-Seq）などで一度に大量に計測する事が可能となり、発現解析や構造変異解析などに応用されている。また、それらの解析データは日々公共のデータベース（Gene Expression Omnibus, GEO）に臨床データとともに登録されている。登録されたデータには、これまでのトランスクリプトーム解析が乳癌の治療成績に貢献したように新たな治療ターゲットであるバイオマーカーを見いだせる可能性がある。</p> <p>本研究では、マイクロアレイやRNA-Seqの臨床応用やバイオマーカーの検出を困難にしている問題点にそれぞれ焦点を当てた。マイクロアレイの検討では、臨床と同じ逐次的な検体に対応できる前処理法を提案した。そして、提案手法の適用可能性を調べるためにデータベースの臨床因子とトランスクリプトームの発現値との一致率を向上させることを目標とした。RNA-Seqでは、新たなバイオマーカーの一つである融合遺伝子の検出において前処理法を提案し、既存の手法による融合遺伝子の検出精度の向上を目標とした。本論文は以下の4章で構成されており、第1章では研究背景、目的、及び目標を述べた。</p> <p>第2章では、マイクロアレイデータを臨床で用いるために課題となっている実験室や実験者などの相違に伴う誤差であるバッチ効果に対して、臨床的需要のある1検体単位でのバッチ補正法の開発を行った。バッチ効果は、新たなバイオマーカーを見出すための大きな障壁となっているだけでなく、見出したバイオマーカーや診断法を1検体単位でそのまま臨床診断として適用することを困難にしている。本研究では、シングルアレイによるバッチ補正法であるノンパラメトリック-Z-スケーリング（NPZ）法を提案し、既存の手法に対して臨床因子との一致率の向上を目的とした。公共のデータベースGEOからエストロゲン受容体（ER）とヒト上皮増殖因子受容体2（HER2）の免疫組織学的染色（IHC）法の結果を有する2,817症例のマイクロアレイの発現データを抽出した。続いて、マイクロアレイデータから背景補正及び対数変換のみを行ったもの（正規化なし）と既存の3つの正規化法（MAS5、fRMA、及びνRMX）の4通りのデータに対して、バッチ補正なしと4つのシングルアレイによるバッチ補正（RANK、YuGene、Z、及びNPZ）とマルチアレイによるバッチ補正（ComBat）の6通りの処理を加えて、4×6で各々のERとHER2のIHCの結果とmRNAの発現の一致率を比較した。NPZ法のバッチ補正を行うことでIHCとmRNAの発現の一致率は向上した。さらにER陽性リンパ節転移陰性乳癌の内分泌療法患者におけるmRNAによるER陽性HER2陰性サブタイプとその他のサブタイプの再発曲線の解析においてバッチ補正なし（$P=0.01$）に対してNPZ法（$P=1.14E-04$）は予後をより明確に分けた。乳癌サブタイプによる治療効果の解析で最も化学療法感受性が高いとされるHER2タイプの病理学的完全寛解率の比較において、バッチ補正なし（41.1%）に対して、NPZ法（44.6%）はより高い病理学的完全寛解率を有し治療効果を強調した。これらの結果からNPZ法は腫瘍の特徴を維持した状態での逐次的なバッチ効果の補正に対して適用可能性を示した。</p> <p>第3章では、乳癌細胞のRNA-Seqデータから融合遺伝子の同定を行い、その検出精度の向上を目標とした。融合遺伝子は、遺伝子本来の機能を変化させるため腫瘍形成の原因の一つであるとともに治療のターゲットであるバイオマーカーとして注目されている。しかしながら、RNA-Seqによる融合遺伝子の同定の精度は十分ではない。その原因の一つがRNA-Seqのリード長は比較的短いため、参照ゲノムの複数の場所にマッピングされることである。RNA-Seqでは、発現量に応じて同一遺伝子から配列の一部が重複したオーバーラッピングリードが得られる。本研究では、オーバーラッピングリードの不一致部分を削除するためにリードの両端から数塩基をカットしたリード（シフトドリード）を作成し、シフトドリードの類似性を用いてリードをクラスタリングすることで、最終的に代表配列としてリードを伸長させるシフトドショートリードクラスタリング（SSC）法を提案した。そしてSSC法によって前処理されたリードを用いることで既存の融合遺伝子検出手法のリードのマッピング率が向上し、最終的に融合遺伝子の同定率が改善すると仮説を立てた。仮説を検証するために、4つの細胞株（BT-474、MCF-7、SKBR-3、及びBT-47D）からのRNA-SeqデータにSSC法を適用し、非適用例と比較した。SSC法にて1塩基だけシフトした場合にリードが平均201塩基から217塩基（108%）に伸張し、参照ゲノムへのマッピング率が89%から94%へ改善した。既存の融合遺伝子検出手法であるSTAR-Fusionによる融合遺伝子の同定率も49%から54%へと改善した。SSC法は、マッピング率を向上させ、融合遺伝子同定率も改善させる有効な手法である可能性が示唆された。</p> <p>第4章では本研究で得られた成果を総括し、本研究の寄与と今後の展望について述べた。マイクロアレイの解析では、臨床的需要に即したデータの補正と生物学的特徴を維持したデータ統合の両立が求められており、本研究で提案した前処理法のNPZ法は有力な一つの手法となりえることを示した。精度に課題のあるRNA-Seqの融合遺伝子検出においても予め前処理としてリードを伸長させるSSC法は、既存の手法の融合遺伝子検出の精度をさらに向上させその適用可能性を示した。日々蓄積されていくトランスクリプトームデータに対して本研究で提案した手法は、いずれも前処理として働くため既存の手法だけでなく既存の手法と共通した入力形式で新たに開発される手法にもそのまま適用可能である。本研究の提案手法を適用することでこれらの手法の精度をより向上させる可能性があり、今後のさらなる精度の高いトランスクリプトーム解析に貢献することが期待される。</p>	

論文審査の結果の要旨及び担当者

氏 名 (草 田 義 昭)			
	(職)	氏 名	
論文審査担当者	主 査	教授	松田 秀雄
	副 査	教授	松田 史生
	副 査	准教授	戸谷 吉博
	副 査	准教授	小蔵 正輝

論文審査の結果の要旨

遺伝子転写産物 (mRNA) 全てを扱うトランスクリプトームデータは、マイクロアレイやRNAシークエンス (RNA-Seq) などで一度に大量に計測する事が可能となり、発現解析や構造変異解析などに応用されている。本研究では、貴重な臨床データが付随しているにも関わらず、マイクロアレイやRNA-Seqの臨床応用やバイオマーカーの検出を困難にしている問題点にそれぞれ焦点を当てた。マイクロアレイの検討では、臨床と同じ逐次的な検体に対応できる前処理法を提案して、その適用可能性を調べるためにデータベースの臨床因子とトランスクリプトームの発現値との一致率の改善を最終的な目標とした。RNA-Seqでは、新たなバイオマーカーの一つである融合遺伝子の検出において、既存の手法の精度を向上しえる前処理法を提案し融合遺伝子の検出精度の向上を目標とした。本論文は以下の4章で構成されており、第1章では研究背景、目的、及び目標を述べた。

第2章では、マイクロアレイデータを臨床で用いるために課題となっている実験室や実験者などの相違に伴う誤差であるバッチ効果に対して、臨床的需要のある1検体単位でのバッチ補正法の開発を行った。本研究では、シングルアレイによるバッチ補正法であるノンパラメトリック-Z-スケーリング (NPZ) 法を提案し、既存の手法に対して臨床因子との一致率の向上を目的とした。公共のデータベースGEOからエストロゲン受容体 (ER) とヒト上皮増殖因子受容体2 (HER2) の免疫組織学的染色 (IHC) 法の結果を有する2,817症例のマイクロアレイの発現データを抽出しバッチ補正を行った結果から、NPZ法は、腫瘍の特徴を維持した状態での逐次的なバッチ効果の補正に対して適用可能性を示した。

第3章では、乳癌細胞のRNA-Seqデータから融合遺伝子の同定を行い、その検出精度の改善を目標とした。RNA-Seqによる融合遺伝子の同定の精度が十分ではない原因の一つがRNA-Seqのリード長は比較的短いため、参照ゲノムの複数の場所にマッピングされることである。マッピングする前に数塩基シフトしたリードを作成し、その類似性を高速に識別しクラスタリングすることで、代表配列として伸長するシフトドショートリードクラスタリング (SSC) 法を提案した。SSC法は、マッピング率を向上させ、融合遺伝子同定率も改善させる有効な手法である可能性が示唆された。

第4章では本研究で得られた成果を総括し、本研究の寄与と今後の展望について述べた。マイクロアレイの解析では、臨床的需要に即したデータの補正と生物学的特徴を維持したデータ統合の両立が求められており、本研究で提案した前処理法のNPZ法は有力な一つの手法となりえることを示した。精度に課題のあるRNA-Seqの融合遺伝子検出においても予め前処理としてリードを伸長させるSSC法は、既存の手法の融合遺伝子検出の精度をさらに向上させその適用可能性を示した。

以上のことから、本論文は乳癌トランスクリプトームデータを用いた遺伝子発現と融合遺伝子の解析手法を提案することで、トランスクリプトームのデータ解析についての研究の発展に貢献するものである。よって、博士 (情報科学) の学位論文として価値のあるものと認める。