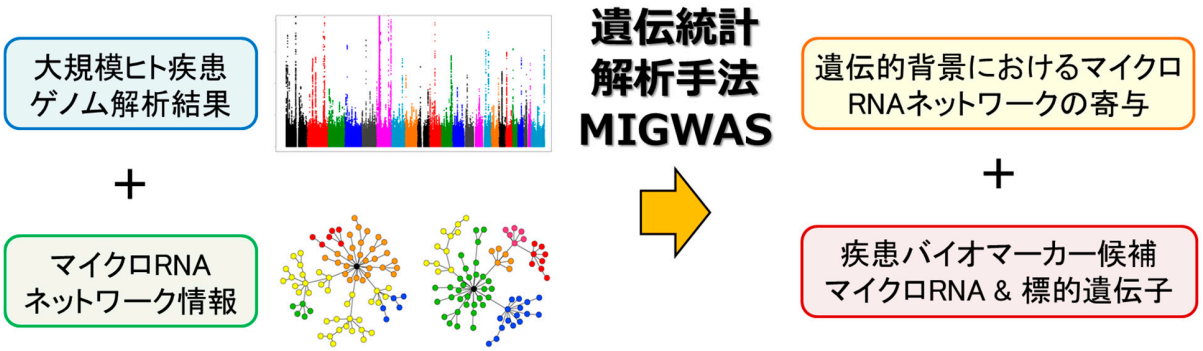
92　遺伝統計解析手法MIGWASに基づく核酸ゲノム創薬	岡田　随象

【目的】マイクロ RNA（miRNA）は生体内に存在する機能性小分子 RNA であり、多彩なヒト疾患において、バイオマーカーや治療標的としての役割が期待されている。既存の miRNA スクリーニング手法は、少数サンプルを対象に一部の miRNA に対する発現定量比較験などの機能的実験により実施されており、より大規模かつ網羅的なスクリーニング手法の開発、特にゲノムビッグデータ解析により同定された疾患感受性遺伝子の情報を活用した新たな miRNA スクリーニング手法の開発に期待が高まっていた。本研究は、遺伝統計解析により大規模疾患ゲノム解析の成果を miRNA-標的遺伝子ネットワークと融合する、解析パイプライン MIGWAS（miRNA enrichment analysis in GWAS）の実装を進めることにより、疾患ゲノム情報に基づく miRNA および標的遺伝子のインシリコ・スクリーニング手法の確立を目的とする。

【方法】次の 5 点において MIGWAS 解析パイプラインの改良を行った。1. miRNA-標的遺伝子におけるスコアの画一的な再計算、2. データ格納方式の変更、3. プログラム言語の変更、4. 解析パイプラインの一元化、5. 並列計算処理の実装。公的データベースより組織特異的 miRNA 発現情報を取得し、細胞組織間および miRNA 間の分散を考慮した正規化処理を実施した。日本人集団 60 名を対象に、次世代シークエンサーHiSeq2500 を用いて末梢血単核球由来の miRNA 発現定量解析を実施した。

【結果】MIGWAS 解析パイプラインの改良の結果、大幅な計算時間の短縮と解析操作の簡素化を実現した。細胞組織特異的な miRNA 発現情報に対して正規化処理を実施し、異なる細胞組織間であっても、適切な正規化処理を実施することで組織特異性の miRNA 発現量の相対的な比較検討が可能となることを確認した。次世代シークエンサーのリード上におけるクオリティ・スコアの評価を実施したところ、成熟した miRNA の塩基長である 22 塩基までの範囲において良好なスコアを得ることができた。取得・正規化された組織特異的 miRNA 発現情報を MIGWAS 解析パイプラインに統合することによって、組織特異的なバイオマーカーmiRNA のインシリコ・スクリーニングの実装を進めていく。

遺伝統計解析手法 MIGWAS による疾患ゲノム情報を活用した miRNA インシリコ・スクリーニング



92. 遺伝統計解析手法 MIGWAS に基づく核酸ゲノム創薬

岡田　随象

大阪大学 大学院医学系研究科 遺伝統計学

Key words : 遺伝統計学，マイクロ RNA，核酸ゲノム創薬，インシリコ・スクリーニング

緒　言

マイクロ RNA（miRNA）は生体内に存在する機能性小分子 RNA であり、特定の標的遺伝子への結合を介してタンパク質への翻訳過程を制御している。miRNA は悪性腫瘍や免疫関連疾患など多彩なヒト疾患において、バイオマーカーや治療標的としての役割が期待されている。生体内には数千種類もの miRNA が存在するため、有用なスクリーニング手法の確立が重要視されている。既存のスクリーニング手法は、比較的少数のサンプルを対象に一部の miRNA に対する発現定量比較験などの機能的実験により実施されており、より大規模かつ網羅的なスクリーニング手法の開発が望まれていた。特に、商用アレイに搭載された一部の miRNA のみがスクリーニング対象となっていることは、解析に網羅性を下げるという観点において大きな課題となっていた。次世代シークエンサーに代表されるヒトゲノム配列解読技術の発展により、数十万人規模もの大規模でヒト疾患ゲノム解析が行われる時代となっており、ゲノムビッグデータ解析により同定された疾患感受性遺伝子の情報を活用した、新たな miRNA スクリーニング手法の開発に期待が高まっている。

本研究は、遺伝統計解析により、既に数百万人規模で多彩なヒト疾患に対して実施されたゲノム解析の成果を、miRNA-標的遺伝子ネットワークと融合することにより、疾患ゲノム情報に基づく miRNA および標的遺伝子のインシリコ・スクリーニング手法の確立を目的とする。

miRNA データ解析の主要なボトルネックとして、個々の miRNA が翻訳制御する標的遺伝子との対応関係の複雑さと曖昧さが挙げられる。筆者らはこれまでに、大規模ヒト疾患ゲノム解析の一種であるゲノムワイド関連解析（genome-wide association study; GWAS）と、複数のバイオインフォマティクス技術により予測された miRNA と標的遺伝子のペアが構成する高次元ネットワークとのつながりを、スーパーコンピューター上で網羅的に検討する遺伝統計解析手法 MIGWAS（**mi**RNA-target gene enrichment analysis in GWAS）を新規に開発した［1］。MIGWAS の開発により、疾患ゲノム情報に基づく miRNA と標的遺伝子のペアの網羅的なインシリコ・スクリーニングが初めて可能になった。本研究計画においては、MIGWAS における解析パイプラインの更なる改良による網羅性の拡大と計算時間の短縮、組織特異的 miRNA 発現情報の MIGWAS 解析パイプラインへの統合を実施し、バイオマーカーとしての miRNA と標的遺伝子の同定を図る。筆者らは、大規模疾患ゲノム解析の実施［2］と、疾患ゲノム情報から直接的にゲノム創薬およびドラッグ・リポジショニングを行う遺伝統計解析を展開しており［3］、本研究計画においても情報解析通じた疾患ゲノム核酸創薬の推進を進めていく。

方　法

1．MIGWAS 解析パイプラインの改良

MIGWAS 解析パイプラインの改良を実施した。まず、miRNA が標的とする標的遺伝子のリストを、複数の推定アルゴリズムから取得した。miRDB、MiRmap、PITA、TargetScan の 4 つの代表的なアルゴリズムを選定し、最新版のヒトゲノム標準情報（hg19）上の遺伝子産物に対する、miRNA データベース情報（r22）上の miRNA の標的結合スコアを同じ条件下で網羅的に計算した。従来の MIGWAS においては、miRNA と標的遺伝子のスコアを 1 対 1 対応の行の集合としてデータを格納していたため、数千万にものぼる miRNA-標的遺伝子対応関係における該当情報のアク

セスに時間を要していた。本研究においては新たに、miRNA と標的遺伝子の行列形式にデータを格納する方針を採用することで、行列上の該当行列へと変更を行い、データアクセスの高速化を図った。また、実装するプログラムソースコードを perl から python へと変更した。複数のソースコードにまたがっていた処理プロセスを一元化し、単一のソースコードのみで実行可能とした。さらに、既存の計算ライブラリパッケージを用いた並列計算処理機能（マルチスレッド機能）を実装することで、解析時間の大幅な短縮を図った。

2．組織特異的 miRNA 発現情報の取得

疾患ゲノム情報とエピゲノム情報との統合において、数多くの組織から取得されたエピゲノム情報に対して別個に統合を実施し、細胞組織間で解析結果を比較検討することにより、疾患における組織特異性を明らかにすることが可能である。本研究において、公開データベースから組織特異的 miRNA 発現プロファイルを取得し、MIGWAS 解析パイプラインへの統合を実施した。細胞組織間における miRNA 発現プロファイルの差を補正する目的で、正規化処理を実施した。

組織特異的遺伝子発現情報を充実させる目的で、末梢血単核球における miRNA 発現情報を取得した。60 名の日本人集団から取得された末梢血より、末梢血単核球を遠心分離で採取し、miRNA を含む短鎖 RNA 分画を抽出した。SMARTer smRNA-Seq Kit（タカラバイオ社）を用いて、抽出した短鎖 RNA 分画からライブラリを作成し、次世代シークエンサー機器 HiSeq2500（イルミナ社）を用いて miRNA 発現量シークエンスを高深度にて実施した（1 サンプルあたり 500 万リード）。

結果および考察

1．MIGWAS 解析パイプラインの改良

MIGWAS 解析パイプラインにおける改良を実施した。1. miRNA-標的遺伝子におけるスコアの画一的な再計算、2. データ格納方式の変更、3. プログラム言語の変更、4. 解析パイプラインの一元化、5. 並列計算処理の実装、を行ったことで、一つのゲノムワイド関連解析結果に対する適用時間を、10 日程度から、数時間へと大幅に短縮した（96 CPU コアを有する大型計算機サーバーを使用した場合）。将来的に予定されている組織特異的 miRNA 発現情報の MIGWAS 解析パイプラインへの実装においては、数十～数百種類の細胞組織の各において、MIGWAS を並列に実行する必要があり、データ計算量の増大が懸念されるため、本研究におけるデータ計算時間の短縮が効果的に働くと期待される。

2．組織特異的 miRNA 発現情報の MIGWAS 解析パイプラインへの統合

公的データベースから取得した細胞組織特異的な miRNA 発現情報に対して、細胞組織間および miRNA 間の分散を考慮した正規化処理を実施した。異なる細胞組織間であっても、適切な正規化処理を実施することで組織特異性の miRNA 発現量の相対的な比較検討が可能となることを確認した。

次世代シークエンサー機器を用いて末梢血単核球由来の miRNA 発現量定量を実施した。次世代シークエンサーのリード上におけるクオリティ・スコアの評価を実施したところ、成熟した miRNA の塩基長である 22 塩基までの範囲において良好なクオリティ・スコアが得られたことが確認できた（図 1）。メッセンジャーRNA を対象とした場合と異なり、次世代シークエンスを用いた miRNA の発現定量解析については明確な解析プロトコルが定まっていなかったが、本成果により、より広汎な細胞組織、より大規模なサンプル数に対してのシークエンスの方向性を定めることができた。

3．今後の研究方針

本研究により、MIGWAS 解析パイプラインの高速化と、組織特異的 miRNA 発現情報の取得が実施された。今後は、各細胞組織別に、特異的に発現する miRNA を対象とした MIGWAS を実施することで、組織特異的な miRNA インシリコ・スクリーニング機能の実装を進めたい。さらに、同定したバイオマーカーmiRNA に対しては臨床サンプル由来の実証実験および機能解析実験を実施し、新たな疾患ゲノム核酸創薬の推進を図っていきたい。

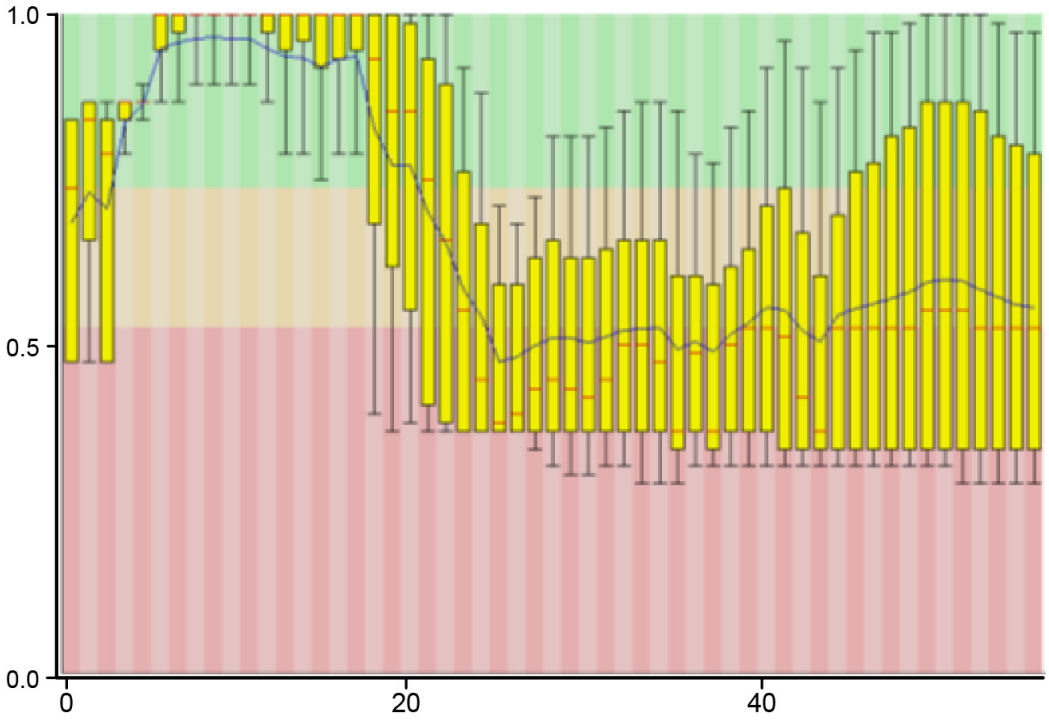


図 1. 次世代シークエンサーと用いた抹消血単核球由来の miRNA シークエンス結果
末梢血単核球から抽出した miRNA を、イルミナ社の次世代シークエンサーHieq2500 を用いて発現定量解析した。成熟 miRNA の塩基長である 22 塩基を含む範囲で、良好なシークエンススコアが得られている。横軸にシークエンス結果の塩基長、縦軸に各塩基長に対応するシークエンススコアを配置した。

共同研究者・謝辞

本研究の共同研究者は、東京大学大学院医学系研究科アレルギー・リウマチ学教室の坂上沙央里と、理化学研究所生命医科学研究センター自己免疫疾患研究チームの山本一彦である。

文　献

1) Okada Y, Muramatsu T, Suita N, Kanai M, Kawakami E, Iotchkova V, Soranzo N, Inazawa J, Tanaka T. Significant impact of miRNA-target gene networks on genetics of human complex traits. Sci Rep. 2016 6:22223. PMID: 26927695 DOI: 10.1038/srep22223.

2) Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, Iwata N, Ikegawa S, Hirata M, Matsuda K, Kubo M, Okada Y, Kamatani Y. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. Nat Genet. 2018 50:390-400. PMID: 29403010 DOI: 10.1038/s41588-018-0047-6.

3) Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A, Ortmann W, Bhangale T, Denny JC, Carroll RJ, Eyler AE, Greenberg JD, Kremer JM, Pappas DA, Jiang L, Yin J, Ye L, Su DF, Yang J, Xie G, Keystone E, Westra HJ, Esko T, Metspalu A, Zhou X, Gupta N, Mirel D, Stahl EA, Diogo D, Cui J, Liao K, Guo MH, Myouzen K, Kawaguchi T, Coenen MJ, van Riel PL, van de Laar MA, Guchelaar HJ, Huizinga TW, Dieudé P, Mariette X, Bridges SL Jr, Zhernakova A, Toes RE, Tak PP, Miceli-Richard C, Bang SY, Lee HS, Martin J, Gonzalez-Gay MA, Rodriguez-Rodriguez L, Rantapää-Dahlqvist S, Arlestig L, Choi HK, Kamatani Y, Galan P, Lathrop M; RACI consortium; GARNET consortium,

Eyre S, Bowes J, Barton A, de Vries N, Moreland LW, Criswell LA, Karlson EW, Taniguchi A, Yamada R, Kubo M, Liu JS, Bae SC, Worthington J, Padyukov L, Klareskog L, Gregersen PK, Raychaudhuri S, Stranger BE, De Jager PL, Franke L, Visscher PM, Brown MA, Yamanaka H, Mimori T, Takahashi A, Xu H, Behrens TW, Siminovitch KA, Momohara S, Matsuda F, Yamamoto K, Plenge RM. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014 506:376-381 PMID: 24390342. DOI: 10.1038/nature12873.