

博士論文

ゲノムと組織特異的発現情報の統合で明らかにする  
マイクロ RNA の疾患病態

坂上 沙央里

## 目 次

序文	1
方法	7
結果	26
考察	47
結語	53
引用文献	54
謝辞	81

# 序 文

## 【ヒトゲノム解析の歴史と現在の課題】

遺伝統計学は、ゲノムとヒトの多様な形質の間にどのような関連があるかを統計学的に明らかにすることを目的とした学問である。かつては、連鎖解析の手法でメンデル形式の遺伝病に代表されるような稀な疾患の原因変異を見出すことが遺伝統計学研究の主流だった。しかし、2001年にヒトゲノムプロジェクトが終了<sup>1</sup>し、個人の多様性を特徴づける一塩基多型（SNP; single nucleotide polymorphism）の情報がマイクロアレイ技術の成熟により大量に正確に得られるようになった。ゲノム情報が正確に安価に大量に得られることを背景とし、2000年代以降、ゲノムワイド関連解析研究（GWAS; genome-wide association study）が遺伝統計学研究の潮流となった<sup>2</sup>。2002年に世界で初めてのGWASが行われて以来<sup>3</sup>、4000件以上ものGWASがこれまでに行われてきた<sup>4</sup>。GWASでは、疾患の有無と遺伝子型の頻度との関連を全ゲノムのSNPに渡って一つずつ検定を行うが、GWASで得られるのは統計学的に有意に形質と関連があるSNPの染色体上の座標という一次元的な情報である。ヒトゲノム上の隣接するSNPどうしは連鎖不平衡関係にあり、それぞれの存在

が非独立な関係にある。また、ゲノムのうち多くの部分を占めるタンパク質非コード部位（ノンコーディング領域）の機能的な意義に関しては、タンパク質コード部位（コーディング領域）に比して網羅的解明は道半ばである。したがって、統計学的に有意に疾患と関連のある SNP の位置情報からは、①SNP によって影響が与えられる遺伝子機能②SNP による生物学的機能変化の種類・程度の両者を一意に決めることが困難である。実際、2 型糖尿病患者で行われた大規模 GWAS の結果に対して、ベイズ統計学を用いたファインマッピング手法を適用してどの SNP がどの遺伝子に作用して糖尿病の発症に寄与するかを調べた Mahajan らの研究<sup>5</sup>では、コーディング領域内に疾患感受性 SNP があっても、その場所と連鎖不平衡関係にある別の転写制御領域の SNP が真に発症の原因であることが多いことが報告されている。

さらに、大規模 GWAS の要約統計量の数理的解析により、複雑形質の遺伝構造(genetic architecture)は極めて多因子 (polygenic) であることが判明した。例えば身長は、最も遺伝率 (heritability; ゲノム情報で説明される形質の分散) の大きいヒト形質の 1 つであり、双生児研究などから 70% がゲノム情報のみに規定されていると考えられてきたが、統計学的に有意に身長と関連する SNP だけではそのうち 9% しか説明でき

ない<sup>6</sup>。一方で、統計学的な有意性に限らずゲノム全体に散在する小さな効果サイズの多型をすべて足し合わせるポリジェニックモデル (Polygenic model) を数理的に再現し、実際の GWAS 統計量に当てはめるとかなりの遺伝率が説明可能となることが判明した<sup>7</sup>。生活習慣病などのありふれた疾患 (common disease) は、個々の効果サイズ(effect size)は小さいがゲノム上に大量に存在するありふれた遺伝子多型 (common variant) の個人ごとの組み合わせ・足し合わせにより発症するということが分かってきた<sup>8</sup>。

すなわち、疾患病態の正確な理解には、①非コード領域の SNP②統計学的有意水準を満たさない効果サイズの小さい SNP に関しても機能的な解釈を行うことが重要である。その際には、先述した個々の細胞上の基本的に不可変・一次元的なゲノム情報が、空間的分布・外部組織からの刺激・環境因子への暴露により、いつ、どのような組織で、どのような刺激を受けて遺伝子発現量を変化させ疾患発症にいたるのかという文脈を理解する必要がある。この文脈理解に、ゲノム情報とマルチオミクス情報を統合して解析することが必要である。近年、ハイスループットな実験系の発達により、ヒトの組織・細胞での転写情報「トランスクリプトーム」<sup>9</sup>、タンパク質情報「プロテオーム」<sup>10</sup>、代謝物情報「メタ

ボローム」<sup>11</sup>といった生命現象の多様なレイヤーのスナップショットがカタログ化されており、それらとゲノムやエピゲノム情報<sup>12</sup>との統合的解析で新たな生物学的知見が得られるようになってきている。例えば、当時過去最大の BMI に対する GWAS 情報と、組織特異的な発現情報(トランスクリプトーム)との組み合わせにより、肥満に関連する遺伝子は主に脳での発現が顕著であること、「脳疾患としての肥満」という概念が明らかになった<sup>13</sup>。

このようなこれまでのゲノム研究の潮流から、新たなオミクス層としてマイクロ RNA (miRNA) の網羅的発現情報とゲノム解析との統合解析することとした。

### 【マイクロ RNA と疾患病態】

マイクロ RNA (miRNA) は、タンパク質非コード RNA すなわちノンコーディング RNA の一種であり、その長さは 21~25 塩基と短い。miRNA は、タンパク質コード RNA すなわちコーディング RNA であるメッセンジャー RNA (mRNA) と結合することで特定の遺伝子の転写産物の翻訳を抑制する<sup>14,15</sup>。近年、自己免疫疾患や悪性腫瘍など多くの疾患におい

て miRNA の関与が注目され、バイオマーカーや創薬対象としての研究が進展している<sup>16-18</sup>。例えば、全身の関節破壊を生じる自己免疫疾患である関節リウマチにおいては miR-146a や miR-125a-5p をはじめとして種々の miRNA を疾患病態理解や疾患活動性マーカーとして使用する試みが盛んである<sup>19</sup>。さらに、miRNA 自体を阻害する薬品や、miRNA の機能を模倣する薬品などの核酸医薬品の研究開発も盛んに行われている<sup>18</sup>。miRNA の発現量も遺伝子による調節を受けており、個人の遺伝子型と miRNA 発現量の関係を検討する cis-eQTL 解析も実施されてきた<sup>20</sup>。一方で、疾患病態に関与する miRNA のスクリーニングには、一部の miRNA を対象に少数のサンプルを用いた実験的な発現量計測が主に用いられており、網羅的に候補 miRNA を検出する手法は未開発であった。miRNA と mRNA の結合関係は部分相補的であり、単純な一対一対応ではなく多数対多数の結合を考慮する必要がある点も、詳細な検討の妨げとなっていた。

一方で、GWAS シグナルから miRNA の疾患への関与を導く試みも認められはじめてきている。例えば、腸管の炎症を伴う自己免疫疾患であるクローン病の GWAS では、欧米人集団のクローン病発症に強く関連する遺伝子 IRGM 内の同義変異 (synonymous variant) が miR-196 の結合

seed 部位内に位置しており、リスク変異の保有により miR-196 の結合が阻害されて IRGM 発現量の調節が失われることが示されている<sup>21</sup>。

2017 年、理化学研究所の FANTOM5 コンソーシアムは、世界に先駆けて 179 種類のヒト細胞における約 2000 種類の miRNA の網羅的な測定を行った<sup>22</sup>。これにより、各細胞で発現上位 5 個の miRNA が全体の発現量の半分を占めることで分かるように、miRNA の発現パターンが組織特異的であることが判明した。疾患と miRNA を解析するうえでも、組織という文脈を考慮することが重要と考えられた。

これまで述べた背景を考慮し、本研究において、大規模疾患ゲノム解析結果・miRNA-標的遺伝子ネットワーク・組織特異的 miRNA 発現測定カタログの 3 つの情報を計算機上で統合解析することによる網羅的 miRNA インシリコ・スクリーニング手法 (MIGWAS ; miRNA enrichment analysis in GWAS) を開発し、ゲノム情報のみでは分からなかった miRNA の疾患病態への組織特異的文脈での関与の理解と、バイオマーカー・創薬対象候補 miRNA の同定を目的とした。

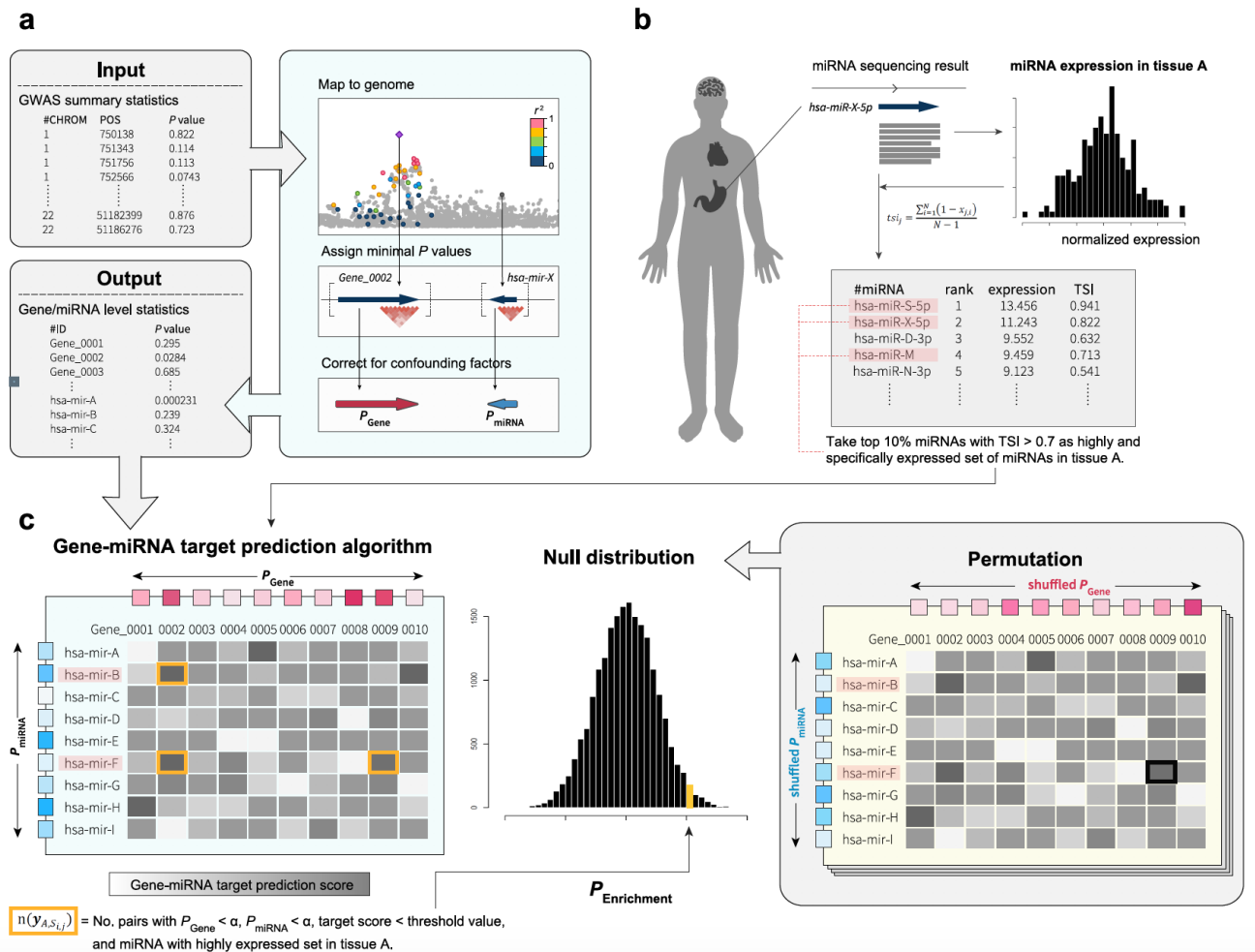


# 方 法

## MIGWAS の概要

MIGWAS では、ある miRNA が疾患発症に関与しているならば、その miRNA 自体の発現を調節する SNP にも、その miRNA が標的として翻訳後調節を行うターゲット遺伝子の発現を調節する SNP にも、疾患発症との関連が認められるだろうという仮説をおいた。これに基づき、miRNA と遺伝子の標的関係の中に GWAS の有意なシグナルが豊富に存在しているかどうか、そしてそれが特定の組織に特に集積しているかを帰無仮説と比較し統計学的に評価する手法を開発した (図 1)。本手法では、ゲノムワイドでの統計学的有意水準である  $P$  値  $< 5 \times 10^{-8}$  を超えない関連を持つ疾患関連 miRNA-遺伝子をペアワイズに網羅的に検出することができ、実験的検証の候補を見出すために有用である。

図 1



miRNA sequencing result

hsa-miR-X-5p

$$E_{Sij} = \frac{\sum_{j=1}^N (1 - x_{ij})}{N - 1}$$

miRNA expression in tissue A

図 1 : MIGWAS 解析の概観図

(a) : GWAS 統計量から遺伝子・miRNA レベルの P 値を求めた。(b) : 179 の細胞における miRNA の網羅的発現データを正規化し、細胞種特異的に高発現・特異的発現する miRNA セットを定義した。(c) : パーミュテーション法を用いて細胞種特異的な miRNA—遺伝子ネットワークへの GWAS シグナルのエンリッチメントを計算した。

## MIGWAS 手順(1) 遺伝子、miRNA レベル $P$ 値の計算

GWAS の要約統計量 (summary statistics) から得られる SNP と疾患との関連の強さの  $P$  値あるいは  $Z$  値を、MAGENTA ソフトウェア<sup>23</sup> を用いて連鎖不平衡関係等を考慮した遺伝子・miRNA レベルの  $P$  値へと変換した。具体的には、

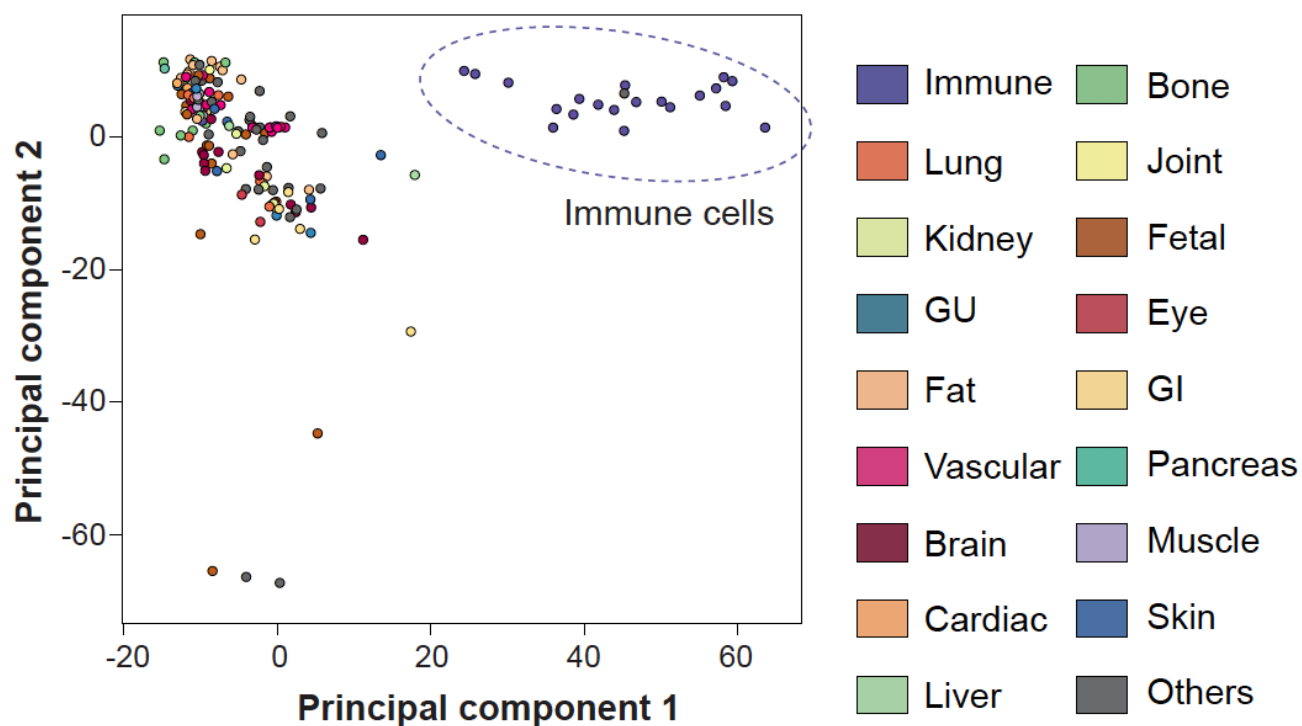
- ① まず、2 つ以上の転写産物が、同じ染色体上の 1Mb 以上離れて存在するか、別の染色体にある遺伝子は除いた。フィルター後の 18,343 の遺伝子に対して、スプライシングアイソソームから考えられる物理的な距離の開始位置の 110 kb 上流から、終了位置の 40 kb 下流までに存在する SNP を、各遺伝子・miRNA レベル統計量の対象範囲と定義した。この位置はこれまでの cis-eQTL による発現量調節 SNP の存在範囲に関する結果を元として先行文献にて決定された。
- ② 各遺伝子・miRNA に対し、 $P_g^{\text{BestSNP}}$  を、上記で定義した領域内の SNP のうち最小の  $P$  値と定義した。
- ③  $P_g^{\text{BestSNP}}$  に存在する交絡因子を除く目的で、Step-wise な多変量回帰を用いた結果、交絡因子として遺伝子サイズ・kb あたりの SNP 数、kb あたりの独立 SNP 数、kb あたりの組み換えホットスポット数、連鎖不平衡単位が同定された。したがって、Sidak の補正<sup>24</sup> を用いてこれらの交絡因子を調整し、各遺伝子・miRNA レベルの  $P$  値 ( $P_{\text{Gene}}$ ,

$P_{\text{miRNA}}$ )へと最終的な変換を行った。

### 組織特異的網羅的 miRNA 発現データの処理

FANTOM5 コンソーシアムデータベースより、179 細胞種における網羅的 miRNA 発現量データをダウンロードした。1,842 pre-miRNA 由来の mature miRNA のうち、新規同定 miRNA を除く発現量 Count per million (CPM)値を以後統計処理することとした。収集した生の CPM 値をまず統計ソフトウェア R の preprocessCore v1.34.0 パッケージを用いて quantile-normalization を行った。179 細胞種は、その由来組織により 18 組織に分類した。(bone ( $n = 10$ ), brain ( $n = 14$ ), cardiac ( $n = 3$ ), eye ( $n = 3$ ), fat ( $n = 15$ ), fetal ( $n = 14$ ), gastrointestinal (GI;  $n = 7$ ), genitourinary (GU;  $n = 6$ ), immune ( $n = 22$ ), joint ( $n = 1$ ), kidney ( $n = 7$ ), liver ( $n = 4$ ), lung ( $n = 10$ ), muscle ( $n = 8$ ), pancreas ( $n = 1$ ), skin ( $n = 6$ ), vascular ( $n = 15$ ), others ( $n = 33$ ))。これらの細胞種ごとの発現 miRNA プロファイルの主成分分析結果を、組織ごとに分類色分けした図を図 2 に示す。

図 2 : 細胞種ごとの網羅的 miRNA 測定結果の主成分分析



各点は miRNA 発現プロファイルに対する主成分分析の主成分 1・2 を示し、その色は細胞の収集された組織を示す。GU; genitourinary, GI; gastrointestinal.

次に、上記で処理した細胞種ごとの網羅的 miRNA 発現量データを元に、各細胞における特異的発現 miRNA セットを特定することを目的に、組織得意性インデックス (TSI; tissue specificity index) を下記のように定義<sup>25</sup>した。

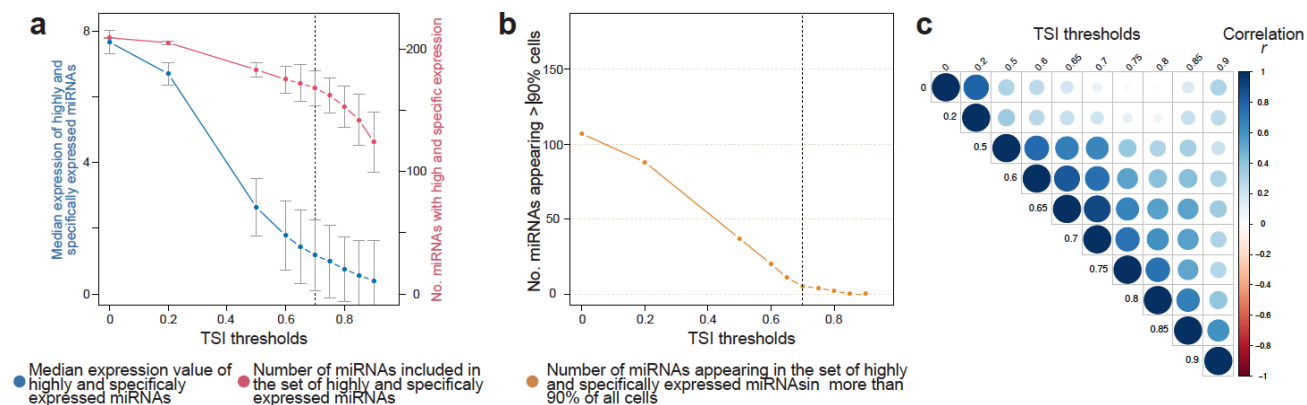
$$tsi_j = \frac{\sum_{i=1}^N (1-x_{ji})}{N-1}$$

ここで N は発現量の計測された総細胞数、 $x_{ji}$  は j 番目の miRNA の全細胞での最大発現量で正規化した i 番目の細胞での発現量を意味する。

各細胞種において、①全 miRNA 発現量のうち上位 10 パーセント以上の発現量をもち②TSI>0.7 を満たす miRNA を、上位特異的発現 miRNA セットとして定義することとした (図 1b に示す)。

TSI > 0.7 について、採用した根拠として図 3 に TSI をスライドさせたときの上位特異的発現 miRNA セットの発現量・数の変化と MIGWAS 統計量の変化を示した。TSI > 0.7 の基準をもとに高発現のみならず、特異的発現を示す miRNA セットを選び組織特異性の検討を行うことができ、かつ検定に使用する miRNA 数も検出力を保つレベルに維持することができた。

図 3 : 組織特異性インデックス (TSI) による組織特異性発現 miRNA セットの内容の変化と MIGWAS 解析統計量の変化



(a) 赤い点で TSI のしきい値により定義される高・特異的発現 miRNA 数の正中値と標準偏差を示した。青い点で TSI のしきい値により定義される高・特異的発現 miRNA の発現量カウントの発現正中値と標準偏差を示した。(b) TSI のしきい値により 90%以上の細胞に高・特異的発現 miRNA として定義される miRNA の数を示した。(c) TSI の条件により出力される MIGWAS 結果の統計量の相関係数をプロットした。色と円の大きさにより、相関係数とその有意性を示した。

## miRNA-ターゲット遺伝子データベースの収集

miRNA は特定の遺伝子に作用して翻訳後調節を行うが、結合する遺伝子の情報は完全ではなく、複数の予測アルゴリズムを用いた結果に基づく複数のデータベースが存在するのが現状である。これを考慮し、一般的に用いられている 4 つの miRNA ターゲット遺伝子予測データベース（TargetScan Human, DIANA-TarBase, PITA, and miRDB）の網羅的情報をダウンロードし、以後の解析に用いるために各遺伝子を行・各 miRNA を列とした行列形式への変換を行った（表 1）。



表 1 : miRNA-標的遺伝子予測データベースのリスト

Database	No. miRNAs	No. target genes	Reference PMID	Accessed date	Website
TargetScan Human	1,530	18,366	26267216	2012/7/11	<a href="http://www.targetscan.org/vert_72/">http://www.targetscan.org/vert_72/</a>
DIANA-TarBase	1,918	18,475	22135297	2014/1/23	<a href="http://diana.imis.athena-innovation.gr/DianaTools/index.php">http://diana.imis.athena-innovation.gr/DianaTools/index.php</a>
PITA	670	16,942	17893677	2014/12/15	<a href="https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html">https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html</a>
miRDB	1,916	16,655	25378301	2014/10/29	<a href="http://www.mirdb.org/">http://www.mirdb.org/</a>

本研究では上記の 4 アルゴリズムを用いて統合解析を行った。

## GWAS シグナルの組織特異的 miRNA への集積の定量

エンピリカルに組織特異的 miRNA 標的遺伝子ネットワークの中の GWAS シグナルのエンリッチメントを定量するために、以下の手法を用いた (図 1c)。

$\mathcal{Y}_{A,S_{i,j}}$  を以下の 4 条件(i)-(iv)を満たす miRNA-遺伝子ペアの集合とする。

- (i)  $i$  番目の予測アルゴリズムにおける miRNA, 遺伝子ペアのスコアが  $j$  番目のしきい値よりも大きい
- (ii)  $P_{\text{miRNA}}$  がしきい値  $\alpha=0.01$  よりも小さい
- (iii)  $P_{\text{gene}}$  がしきい値  $\alpha=0.01$  よりも小さい
- (iv) miRNA が、上記のように定義した細胞 A における上位特異的発現 miRNA セットに含まれる

今回、ゲノム上に散在する小さな effect size の多型も含めた polygenic な関与を検討するためにしきい値  $\alpha=0.01$  を採用した。その他のしきい値でのベンチマーク試験を行い、先行文献<sup>38</sup>や統計量のインフレーションの値をもとにこの値を選択することとした。

組織特異的 miRNA 標的遺伝子ネットワークの中の GWAS シグナルのエンリッチメントを定量するために、 $n(\mathcal{Y}_{A,S_{i,j}})$  を 20,000 回の帰無仮説下でのパーミュテーションにおける統計量を比較した。帰無仮説では、組織特異的

miRNA 発現および、miRNA-標的遺伝子ネットワークと GWAS 結果に関連がないことを仮定した。そのため、1 パーミュテーションごとに、 $P_{miRNA}$  および  $P_{gene}$  をランダムに入れ替えたうえで同様の計算を行い、 $n(\mathbf{y}_{A,S_{i,j}})$  を求めた。

エンリッチメントの P 値を示す  $P_{enrichment,A,S_{i,j}}$  は  $P(n(\mathbf{y}') \geq n(\mathbf{y}))$  と定義した。つ

ぎに  $i$  番目のアルゴリズムでのしきい値  $j$  を上位 1 パーセントから 0.1 パーセントまで、ログ尺度で 8 分割するようにスライドさせ、ここまでの統計量  $P$  値を以下のように Fisher 法でメタ解析した。

$$P_{enrichment,A} = \Phi \left( \frac{1}{\sqrt{n_{algorithm}}} \sum_{i=1}^{n_{algorithm}=4} \left( \frac{1}{\sum_j k_{S_{i,j}}} \sum_{j=1}^9 k_{S_{i,j}} \Phi^{-1}(P_{enrichment,A,S_{i,j}}) \right) \right),$$

$$\text{where } k_{S_{i,j}} = \begin{cases} 1, & \text{when mean } y' \geq 5 \\ 0, & \text{when mean } y' < 5 \end{cases}$$

ここで  $n(\mathbf{y}'_{(k)A,S_{i,j}})$  が疎 (sparse) であるとき帰無仮説の歪みが生じる可能性を考慮し、5 以上のときの統計量を採用することとしている。また、 $\Phi$  は正規分布の累積分布関数を示す。パーミュテーション  $P$  値  $< 0.05$  を有意水準として、以下の議論を行うこととした。

上記解析を行うソフトウェア MIGWAS を公共ソフトウェアリポジトリである GitHub 上で公開し、マルチスレッドオプションを実装するこ

とで高速計算を可能とした。

### **GWAS 統計量の収集**

公共ウェブリソースあるいは研究協力者より、49 個の複雑形質に関する要約統計量 (summary statistics) を収集し、表 2 にまとめた。おおまかに、4 疾患群 (immune/allergy-related [ $n = 4$ ], neuropsychiatric [ $n = 2$ ], cardiovascular [ $n = 1$ ], and genitourinary [ $n = 1$ ]) と 6 量的形質群 (anthropometric [ $n = 3$ ], metabolic [ $n = 3$ ], musculoskeletal [ $n = 1$ ], cardiovascular [ $n = 2$ ], kidney-related [ $n = 2$ ], and hematological [ $n = 2$ ]) に分類した。残りの 28 統計量に関しては、バイオバンクジャパンプロジェクト<sup>26</sup>のデータを使用し、10 疾患類 (immune/allergy-related [ $n = 4$ ], metabolic [ $n = 1$ ], musculoskeletal [ $n = 2$ ], neuropsychiatric [ $n = 1$ ], eye-related [ $n = 2$ ], cardiovascular [ $n = 5$ ], lung-related [ $n = 1$ ], liver-related [ $n = 2$ ], genitourinary [ $n = 2$ ], and malignancy [ $n = 7$ ]) と 1 量的形質群 (anthropometric [ $n = 2$ ]) に分類した。

表 2 : 本研究で用いた GWAS

Trait category	Trait	Study model	No. samples	No. SNPs	Ethnicity	Reference PMID	
Anthropometric	Height	quantitative	253,288	2,550,858	European	25282103	
	Height	quantitative	159,148	5,961,600	Japanese	-	
	BMI	quantitative	339,224	2,555,510	European	25673413	
	BMI	quantitative	158,284	5,961,600	Japanese	28892062	
Immune/Allergy	Rheumatoid arthritis	case/control	80,799	9,739,303	Trans-ethnic	24390342	
	Systemic lupus erythematosus	case/control	6,899	7,793,762	Japanese	-	
	Graves' disease	case/control	9,929	6,020,808	Japanese	-	
	Crohn's disease	case/control	30,740	953,241	European	23128233	
	Ulcerative colitis	case/control	26,897	1,428,749	European	23128233	
	Atopic dermatitis	case/control	9,438	6,031,258	Japanese	26482879	
	Asthma	case/control	36,440	5,961,330	Japanese	-	
	Pollinosis	case/control	34,406	5,961,329	Japanese	-	
	Metabolic	Type 2 diabetes	case/control	110,452	2,915,011	European	24509480
		Type 2 diabetes	case/control	65,702	5,961,329	Japanese	-
LDL cholesterol		quantitative	188,577	2,692,564	European	20686565	
HDL cholesterol		quantitative	188,577	2,692,429	European	20686565	
Triglycerides		quantitative	188,577	2,692,560	European	20686565	
Musculoskeletal	Bone mineral density	quantitative	32,961	2,554,591	European	22504420	
	Osteoporosis	case/control	35,969	5,961,329	Japanese	-	
Neuropsychiatric	Schizophrenia	case/control	81,080	9,444,230	European	25056061	
	Epilepsy	case/control	30,979	5,961,329	Japanese	-	
	Alzheimer's disease	case/control	54,162	7,055,881	European	24162737	
Eye-related	Age-related macular degeneration	case/control	56,494	2,399,382	European	23455636	
	Glaucoma	case/control	22,795	5,961,599	Japanese	-	
Cardiovascular	Arteriosclerosis obliterans	case/control	32,252	5,961,329	Japanese	-	
	Myocardial infarction	case/control	4,864	4,773,769	Japanese	24916648	
	Coronary artery disease	case/control	194,427	9,455,778	European	23202125	
	Atrial fibrillation	case/control	36,792	5,018,048	Japanese	28416822	
	Cerebral aneurysm	case/control	31,467	5,961,329	Japanese	-	
	Ischemic stroke	case/control	43,550	5,961,330	Japanese	-	
	Systolic blood pressure	quantitative	69,395	2,650,287	European	21909115	
	Diastolic blood pressure	quantitative	69,395	2,673,126	European	21909115	
Lung-related	COPD	case/control	31,920	5,961,329	Japanese	-	
Kidney-related	Uric acid	quantitative	110,347	2,207,610	European	23263486	
	eGFR	quantitative	67,093	2,724,035	European	20383146	
Liver-related	Chronic hepatitis B	case/control	29,962	5,961,329	Japanese	-	

Genitourinary	Chronic hepatitis C	case/control	33,858	5,961,329	Japanese	-
	Age at menarche	quantitative	182,416	2,441,815	European	25231870
	Uterine fibroid	case/control	23,212	5,961,329	Japanese	-
Hematological	Urolithiasis	case/control	35,116	5,961,329	Japanese	-
	Red blood cell	quantitative	71,861	2,589,454	European	23222517
	Platelet	quantitative	66,867	2,703,393	European	22139419
Malignancy	Lung cancer	case/control	31,052	5,932,052	Japanese	-
	Esophageal cancer	case/control	28,403	5,932,052	Japanese	-
	Gastric cancer	case/control	33,349	5,932,052	Japanese	-
	Colorectal cancer	case/control	33,870	5,932,052	Japanese	-
	Breast cancer	case/control	21,768	5,932,052	Japanese	-
	Endometrial cancer	case/control	19,423	5,961,329	Japanese	-
	Prostate cancer	case/control	15,770	5,932,052	Japanese	-

BMI; body mass index, LDL cholesterol; low-density lipoprotein cholesterol, HDL cholesterol; high-density lipoprotein cholesterol, COPD; chronic obstructive pulmonary disease, eGFR; estimated glomerular filtration rate.

これらの要約統計量から、各 SNP の UCSC hg19 リファレンスにおける位置(base pair position)とその形質関連  $P$  値を収集し、フォーマット整形を行った。

### ハイスループット miRNA シークエンス対象検体

今回の疾患対照研究においては、日本人の早期関節リウマチ (RA) 患者 30 名と、日本人健常人参加者 33 名を対象とした。全参加者から、事前に研究同意書を取得し、今研究は大阪大学倫理委員会より承認 (承認番号 734-5) された。参加者の詳細情報を表 3 にまとめた。

表 3 : 発現変動解析実験の参加者（関節リウマチ患者・健常人）の情報

	RA ( <i>n</i> = 30)	HC ( <i>n</i> = 33)
Age, years	67 (57-75)	37 (33-55)
Sex, % female	83.3	39.4
Duration of symptoms, months	12 (6-12)	NA
RF positive, %	83.3	NA
ACPA positive, %	56.7	NA
DAS28-CRP	3.66 (2.98-4.80)	NA
miRNA-seq raw read count, million	6.29 (5.47-8.12)	5.34 (3.98-6.37)
Mapped rate to miRNA, %	4.81 (3.97-6.44)	9.10 (5.62-11.22)

正中値と四分位範囲を示す。

RA; rheumatoid arthritis, HC; healthy control, RF; rheumatoid factor, ACPA; anti-citrullinated protein antibody, DAS; disease activity score. CRP; C-reactive protein.



## miRNA シークエンス・Differential expression 解析

上記患者および健常人検体より、全血を採取し、EDTA チューブに回収した。Ficoll-Paque density gradient medium を用いて末梢血単核細胞 (PBMC) を分離し、miRNeasy Micro Kit (Qiagen)を用いて全 RNA を単離した。SMARTer smRNA-Seq Kit (Takara)とその指定のプロトコルを用いて、miRNA シークエンスライブラリを作成した。HiSeq 2500 (Illumina, read length of 100 bp, single-end)により RNA シークエンスを実施した。クオリティコントロールとして、Cutadapt ソフトウェアを用いてアダプターシーケンスを除去、低クオリティリード (Phred quality score < 20 in > 80% of total bases) の除去、長リードの除去 (length > 50bp) を行った。

クオリティコントロールパスしたリードを、miRBase に登録された既知の miRNA 配列に対して Bowtie を用いて推奨条件により<sup>27</sup>アライメントした。リウマチ患者・健常人での Differential expression 解析を R ソフトウェアの DESeq2 パッケージを用いて行った。発現変動 miRNA は、False discovery rate (FDR-q) < 0.05 で定義した。

## 発現変動 miRNA と MIGWAS 結果のオーバーラップエンリッチメント

### 解析

上記で検出した、RA 患者と健常人との発現変動 miRNA が、MIGWAS 解析で検出された候補 RA 関連 miRNA と偶然よりも有意に重複しているかどうか検証するために、パーミュテーションによるエンリッチメント解析を行った。ここでは、シーケンス実験による miRNA 発現量テーブルのうちサンプル ID を 1 パーミュテーションごとに入れかえ、それぞれに帰無仮説化での発現変動 miRNA セットを求めた。発現変動 miRNA セットと MIGWAS 解析で検出された候補 RA 関連 miRNA セットとのオーバーラップ数を計測し、これを元に経験に基づく (empirical)  $P$  値を求めた。

## 関節リウマチ候補 miRNA のインシリコ・レプリケーション研究

上記で検出した、RA 患者と健常人との発現変動 miRNA かつ MIGWAS 解析による候補 miRNA 周辺領域に関して、最も関連の強い SNP の新規 GWAS メタ解析を実行した。複数人種間関節リウマチメタ解析 (19,234 関節リウマチ症例対 61,565 対照群) に、さらに 2 つのコ

ホートにおける GWAS (3,308 関節リウマチ症例対 8,357 対照群) を加え、固定効果を仮定した *inverse-variance* メタ解析を行った。

### **Summary-level Mendelian Randomization 解析**

MIGWAS で検出された miRNA は、標的遺伝子の転写物に影響を与えて形質に関与することを仮定していることから、標的遺伝子の形質への関与に関しても、対象臓器での発現量調節が重要であることを仮定し、検証することとした。まず、GTEx データベースを用いて eGene を周囲に  $FDR-q < 0.05$  で cis-eQTL バリエーションをもつ遺伝子として定義した。そして、MIGWAS により検出された miRNA の標的遺伝子が eQTL 解析での eGene に集積しているかどうかを二項分布で検定した。次に、Summary-level Mendelian Randomization (SMR)<sup>36</sup> を実行した。本 SMR 解析では、GWAS と eQTL の要約統計量を用いて、発現量変動が疾患発症に linkage でなく causal あるいは pleiotropy を介して関与しているかを、Mendelian randomization と同様の数理モデルにより検定することができる。今回、関節リウマチの eGene および SMR 解析を行うに当たり、発現変動を調べる対象臓器として免疫細胞が有効であることが仮定されたが、GTEx データベースでの情報量から代替として全血データを使用した。

## 結 果

### MIGWAS の概要

今回の研究概要図を図 1 に示した。MIGWAS では、ある miRNA が疾患発症に関与しているならば、その miRNA 自体の発現を調節する SNP にも、その miRNA が標的として翻訳後調節を行うターゲット遺伝子の発現を調節する SNP にも、疾患発症との関連が認められるという仮説をおいた。これに基づき、miRNA と遺伝子の標的関係の中に GWAS の有意なシグナルが豊富に存在しているかどうか、そしてそれが特定の組織に特に集積しているかを帰無仮説と比較し統計学的に評価する手法を開発した。まず、miRNA と遺伝子の標的関係ネットワークにおける各大規模 GWAS シグナルのエンリッチメントを評価した。次に、網羅的組織特異的 miRNA 発現データベース FANTOM5 コンソーシアムのデータを使用し、そのシグナルが特にどの細胞種に集積しているかを評価した。

### ヒト複雑形質への miRNA-標的遺伝子の関与

我々は、過去に世界で行われた延べ 350 万人以上を対象とした 49 の GWAS 結果を MIGWAS で解析した。対象形質はヒト複雑形質で多岐にわたり、大まかに anthropometric ( $n = 4$ ), immune/allergy ( $n = 8$ ), metabolic ( $n = 5$ ), musculoskeletal ( $n = 2$ ), neuropsychiatric ( $n = 3$ ), eye-related ( $n = 2$ ), cardiovascular ( $n = 8$ ), lung-related ( $n = 1$ ), kidney-related ( $n = 2$ ), liver-related ( $n = 2$ ), genitourinary ( $n = 3$ ), hematological ( $n = 2$ ), and malignancy traits ( $n = 7$ )と分類を行った (表 2)。

この解析により、身長・関節リウマチ・2型糖尿病が有意( $P < 0.05$ )に病態への miRNA の関与が重要であることが判明した (図 4 a)。さらに、身長<sup>28</sup>・2型糖尿病<sup>29</sup>においては欧米人・日本人と人種をまたぎ共通して miRNA の関与が示し、本解析の頑健性も確認された。

図 4

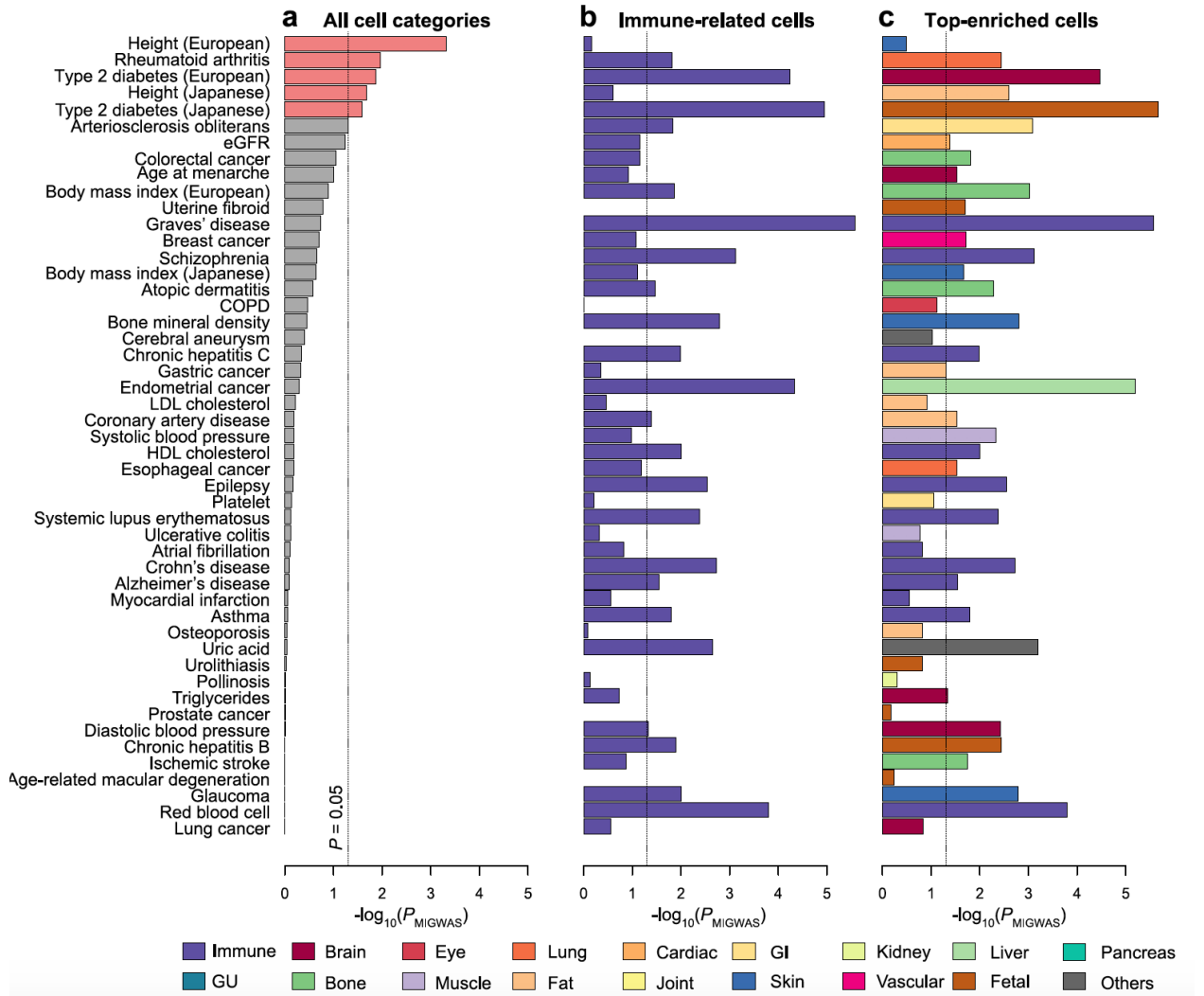


図 4 : MIGWAS 解析実施結果

49 のヒト複雑形質に対して行われた GWAS に MIGWAS を実行した結果。  
**(a)** : 組織特異性を加味しない疾患全体での解析結果。棒グラフにより関連の強さを  $\log_{10}(P_{MIGWAS})$  で示している。ピンクで色をつけた形質は、統計学的に *nomimally significant* であることを示す。  
**(b)** : 免疫細胞での MIGWAS 解析実施結果。  
**(c)** : 最も関連の強い細胞種を示した解析結果。棒グラフの色はその細胞種の組織分類 (図下に示す) により定義した。

## 組織特異的な miRNA-標的遺伝子の複雑形質への関与

ここまでの組織特異性を加味しない解析からヒト形質に様々な程度で miRNA-標的遺伝子の関与が認められることが分かった。そこで、それらの解析に 179 細胞種特異的発現 miRNA データを統合することで、GWAS-miRNA-標的遺伝子ネットワークの細胞種特異的な関与を検証した。FANTOM5 コンソーシアムの 179 細胞種における 1,842 pre-miRNA 由来の mature miRNA 発現量データを正規化後、179 細胞種をその由来組織により 18 組織に分類した。これらの細胞種ごとの発現 miRNA プロファイルの主成分分析結果（図 2）から、特に免疫細胞が特定のクラスターに分類されることが判明した。興味深いことに、疾患ごとに miRNA が強く関与している組織は極めて多様性が高かった(図 4 c)。バセドウ病、C 型肝炎、関節リウマチ、全身性エリテマトーデス、クローン病などの免疫疾患では免疫細胞が miRNA の活動の場（図 4 b）であり、LDL コレステロール・心筋梗塞発症においては脂肪細胞が miRNA の活動の場であることが分かった。表 4 に、各形質で最も強いエンリッチメントを示した細胞種の名称を示す。なお、組織特異性の検討においては複数人種をまたいで共通の組織が検出されることはなかった。組織特異的エピゲノムデータと GWAS とを統合して、形質の遺伝基盤において重要な組

織を特定する解析<sup>33</sup>を用いた先行研究では GWAS の人種により重要な組織が異なる場合も報告されており、今回の結果のさらなる検証が求められる。



表 4 : 組織特異的 MIGWAS 解析の結果

Category	Trait	Immune-related cells	Top-associated tissue	Top-associated cell name	
Anthropometric	Height (European)	-	skin	hair follicle dermal papilla cell	
	Height (Japanese)	-	fat	preadipocyte of the breast	
	BMI (European)	neutrophil	bone	human chondrocyte de-differentiated sample	
	BMI (Japanese)	-	skin	light melanocyte	
Immune/Allergy	Rheumatoid arthritis	phagocyte	lung	bronchial smooth muscle cell	
	Systemic lupus erythematosus	mast cell	immune	mast cell	
	Graves' disease	CD14-positive CD16-negative classical monocy	immune	CD14-positive CD16-negative classical monocy	
	Crohn's disease	mast cell	immune	mast cell	
	Ulcerative colitis	-	muscle	skeletal muscle myoblast	
	Atopic dermatitis	mesenchymal stem cell of the bone marrow	bone	bone cell	
	Asthma	neutrophil	immune	neutrophil	
	Pollinosis	-	kidney	human renal cortical epithelial cell sample	
Metabolic	Type 2 diabetes (European)	fibroblast of lymphatic vessel	brain	Schwann cell	
	Type 2 diabetes (Japanese)	mast cell- stimulated sample	fetal	human amniotic membrane cell- Amnion sample	
	Bone mineral density	fibroblast of lymphatic vessel	skin	keratinocyte	
	Osteoporosis	-	fat	fibroblast of mammary gland	
	LDL cholesterol	-	fat	subcutaneous fat cell	
	HDL cholesterol	natural killer cel	immune	natural killer cel	
	Triglycerides	-	brain	central nervous system pericyte	
Neuropsychiatric	Schizophrenia	CD14-positive CD16-negative classical monocyte	immune	CD14-positive CD16-negative classical monocyte	
	Epilepsy	myeloid leukocyte	immune	myeloid leukocyte	
	Alzheimer's disease	B cell	immune	B cell	
Eye-related	Age-related macular degeneration	-	fetal	fibroblast of villous mesenchyme	
	Glaucoma	mature alpha-beta T cell	skin	keratinocyte	
Cardiovascular	Arteriosclerosis obliterans	fibroblast of lymphatic vessel	gastrointestinal	smooth muscle cell of the esophagus	
	Myocardial infarction	-	immune	mesenchymal stem cell of the bone marrow	
	Coronary artery disease	immature conventional dendritic cell	fat	subcutaneous fat cell	
	Atrial fibrillation	-	immune	CD4-positive alpha-beta T cell	
	Cerebral aneurysm	-	fetal	iPS	
	Ischemic stroke	-	bone	human chondrocyte re-differentiated sample	
	Systolic blood pressure	-	muscle	skeletal muscle satellite cell	
	Diastolic blood pressure	B cell	brain	Schwann cell	
	Lung-related	COPD	-	eye	fibroblast of the conjunctiva
	Kidney-related	Uric acid	CD14-positive CD16-negative classical monocyte	others	GAG secreting cell
eGFR		-	cardiac	fibroblast of cardiac tissue	
Liver-related	Chronic hepatitis B	phagocyte	fetal	human amniotic membrane cell- Amnion sample	
	Chronic hepatitis C	mast cell- stimulated sample	immune	mast cell- stimulated sample	
Genitourinary	Age at menarche	-	brain	human pituitary gland - adult sample	
	Uterine fibroid	-	fetal	fibroblast of villous mesenchyme	
	Urolithiasis	-	fetal	fibroblast of villous mesenchyme	
Hematological	Red blood cell	human CD14-positive monocyte sample	immune	human CD14-positive monocyte sample	
	Platelet	-	gastrointestinal	keratinized cell of the oral mucosa	
Malignancy	Lung cancer	-	brain	human diencephalon - adult sample	
	Esophageal cancer	-	lung	bronchial smooth muscle cell	
	Gastric cancer	-	fat	perirenal preadipocyte	
	Colorectal cancer	-	bone	nucleus pulposus cell of intervertebral disc	
	Breast cancer	-	vascular	smooth muscle cell of the subclavian artery	
	Endometrial cancer	neutrophil	liver	endothelial cell of hepatic sinusoid	
Prostate cancer	-	fetal	embryonic stem cell		

これまでの文献検討から、miRNA のヒト免疫疾患への関与が推定されている<sup>30</sup>ことから、ここでさらに免疫形質について着目した解析を行った(図5)。ここでも、各免疫疾患に miRNA-標的遺伝子ネットワークを介して関与する細胞種は極めて特異的であることが分かった。例えば関節リウマチでは、肺、骨、免疫細胞に特に強い関連(それぞれ  $P = 3.6 \times 10^{-3}$ ,  $9.6 \times 10^{-3}$ , and  $1.5 \times 10^{-2}$ )を認めた。喫煙などの環境因子への暴露を契機に肺において病的自己免疫反応が惹起され、それが骨びらんへ進展するというこれまでの関節リウマチの生物学的理解とも一致するものであった<sup>31</sup>。

図 5

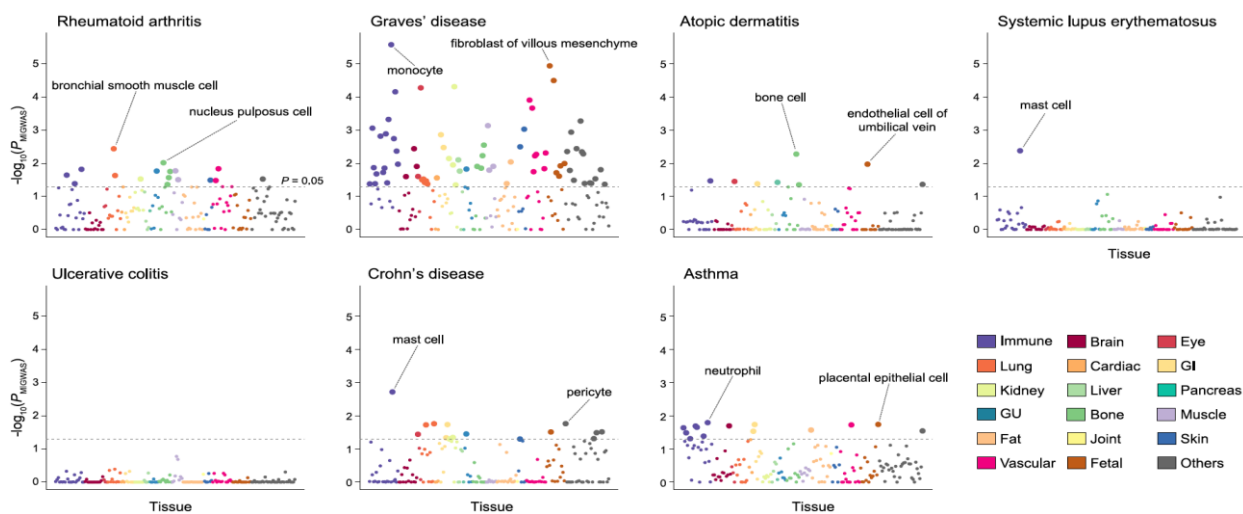


図 5 : 免疫疾患における MIGWAS 解析結果

7つの免疫アレルギー疾患 GWAS について MIGWAS を適用した結果を細胞種ごとに示した。各点は、関連の強さを  $\log_{10}(P_{\text{MIGWAS}})$  で示しており、横軸に全 179 細胞種での結果を並べた。点線は有意水準の  $P \text{ value} < 0.05$  を示す。有意かつ上位 2 細胞の名前を点の脇に示した。GI; gastrointestinal, GU; genitourinary.

## MIGWAS 結果の実験的検証

我々は、このようにして MIGWAS が同定した miRNA が実際に疾患病態に寄与していることを実証するために、これまで文献的にも miRNA が病態に関与していることが推察されている関節リウマチ<sup>30</sup>を例にとり、関節リウマチ患者と健常人とで MIGWAS で同定された関節リウマチ関与候補 miRNA に発現量の差異を認めるかの実証実験を行った。30名の早期関節リウマチ患者・33名の健常人の末梢血単核球細胞における網羅的 miRNA シークエンス(次世代シーケンサーを用いた miRNAseq)を行い、発現量変動解析をした。今回、末梢血単核球細胞を対象細胞種として選択した根拠は、臨床的・実験的に得ることが可能であり、これまでのゲノム研究でも疾患の遺伝率の集積していることが示されてきた細胞種だからである<sup>32,33</sup>。

この実験により、False Discovery Rate < 0.05 を基準としたとき、94 の発現変動 miRNA が検出された(図 6 a)。このうち、4 つの miRNA は MIGWAS により候補 miRNA として同定されていたものであった。パーミュテーション法を用いたオーバーラップエンリッチメント解析を実行したところ、MIGWAS により同定した miRNA は偶然より 98 倍高い確率( $P = 0.0010$ )で発現変動 miRNA を予測していることが判明した(図

7)。特に、今回オーバーラップを認めた hsa-miR-762 は有意に関節リウマチ患者で高く発現しており ( $\log_2$  fold change = 1.15 and FDR-q = 0.043)、さらに極めて組織特異的(tissue specificity index [TSI] = 0.982、図 6 b および図 6 c)に免疫細胞に発現を認めることが分かった。

図 6

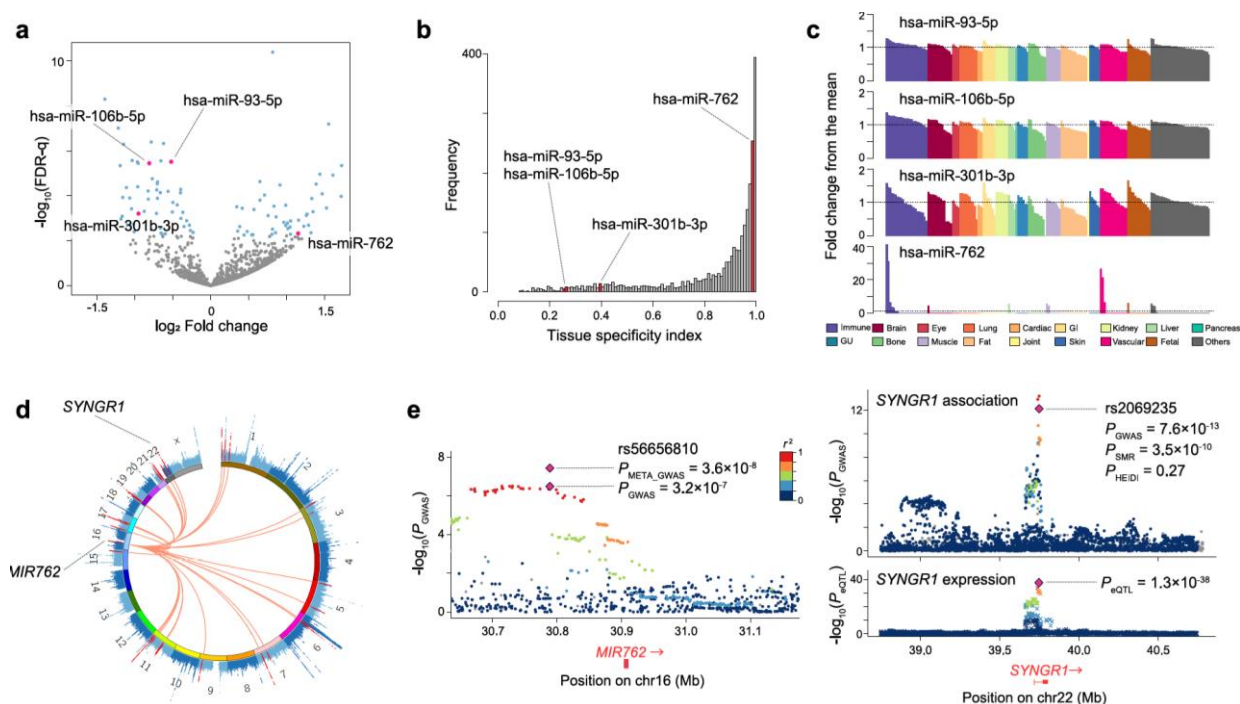
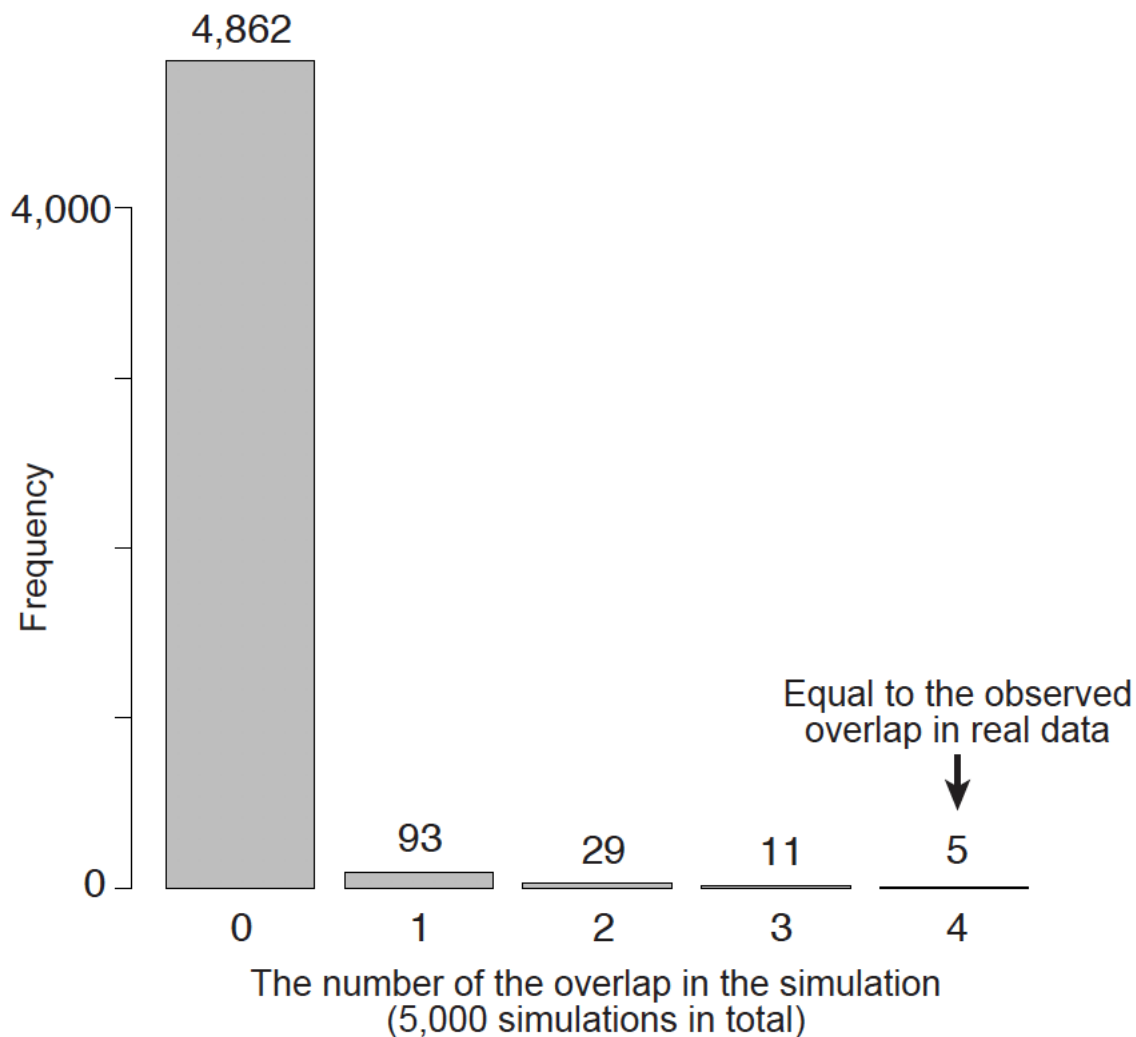


図 6 : 関節リウマチ患者と健常人との発現変動 miRNA 実験結果とその後方解析

(a) : 関節リウマチ患者と健常人との発現変動 miRNA 実験結果をボルケーノプロットで示した。X 軸は  $\log_2$  fold change を示し、Y 軸は  $-\log_{10}(\text{FDR-q})$  を示す。青色で示した点は  $\text{FDR-q} < 0.05$  を示した発現変動 miRNA である。ピンクで示した点は、そのうち MIGWAS で同定された疾患関与 miRNA 候補と重複していた点を示す。(b) : 今回解析対象とした miRNA の TSI (組織特異性インデックス) を棒グラフで示したもの。ピンクで示した棒は、(a) で述べた発現変動 miRNA のうち MIGWAS で同定された疾患関与 miRNA 候補と重複していた miRNA が属する TSI の bin を示したものである。(c) : オーバーラップ miRNA について、各 179 種の細胞での正規化済発現量を全細胞での平均発現量との比で示したもの。y 軸は全細胞での平均発現量からの fold change の値を示している。(d) : 関節リウマチ GWAS のマンハッタン CIRCOS プロット。MIR762 とその標的遺伝子を線で結んだ。(e) : MIR762 と SYNGR1 の領域プロット。メタ解析結果もともに示す。右のパネルには SYNGR1 の eQTL 解析結果と SMR の結果をともに示す。

図 7 : MIGWAS 結果と miRNA 発現変動解析実験結果とのオーバーラップエンリッチメント解析における、パーミュテーション結果の概要



各棒は、5,000 パーミュテーション工程ごとに発現変動 miRNA として定義された miRNA と MIGWAS から出力された miRNA との重複数を記録し、ヒストグラムとして示したものである。最も右の bin (重複数 4) が実際のデータにて観測された重複数である。

## オーバーラップ miRNA のインシリコ・レプリケーション研究

上記の解析によりオーバーラップを認めた 4 miRNA に関しては、すでに既報の大規模 GWAS<sup>34</sup> でゲノムワイド水準は超えないが suggestive な関連が示されていた (19,234 RA cases and 61,565 controls,  $P_{\text{GWAS}} \geq 3.2 \times 10^{-7}$ )。そこで、新規 GWAS メタ解析を実行し、該当 miRNA に強いシグナルが再現されるかどうかを検証 (インシリコレプリケーション研究) した。新たに 3,308 RA cases と 8,357 controls を加えて、上記の既報 GWAS の要約統計量とのメタ解析を行うと、4 miRNA のうち 3 miRNA をコードしている 2 領域において、ゲノムワイド水準を超えて関節リウマチ発症への関連が認められた (22,119 cases and 69,363 controls;  $P_{\text{GWAS\_META}} = 3.3 \times 10^{-8}$  for rs34130487 at *MIR95-MIR106B* at 7q22 and  $P = 3.6 \times 10^{-8}$  for rs56656810 at *MIR762* at 16q11; 図 6 d, 6 e, 表 5)。これらの領域の関節リウマチへの関与は、今回 MIGWAS により候補領域として上がったことでメタ解析を行ったために検出することができた新規領域である。



表 5 : 関節リウマチメタ解析によるインシリコ再現研究結果

rsID	Chr	Position	Allele 1/2	Gene	Stage	Cohort	No. subjects		Allele 1 frequency		OR (95%CI)	P
							RA cases	Controls	RA cases	Controls		
rs34130487	7	99,759,205	C/T	MIR95- MIR106B	GWAS	Brass	483	1,631	0.739	0.739	1.05 (0.88-1.25)	0.57
						Canada	589	1,554	0.725	0.737	0.94 (0.81-1.10)	0.44
						Eira	1,097	1,044	0.741	0.723	1.10 (0.96-1.26)	0.16
						Narac1	863	1,191	0.726	0.746	0.91 (0.77-1.08)	0.29
						Narac2	896	6,603	0.739	0.729	1.11 (0.99-1.25)	0.081
						Wtccc	1,520	10,507	0.727	0.709	1.10 (1.00-1.20)	0.039
						ReAct	275	804	0.752	0.715	1.20 (0.97-1.50)	0.099
						Dutch	1,172	1,684	0.720	0.693	1.18 (1.04-1.33)	0.0098
						ACR-Ref	β47	264	0.731	0.743	0.98 (0.71-1.33)	0.88
						Corrona	894	1,838	0.725	0.725	0.99 (0.87-1.13)	0.91
						Vandy	739	2,247	0.709	0.724	0.93 (0.82-1.06)	0.27
						iChip_ES	-	-	-	-	-	-
						iChip_NL	-	-	-	-	-	-
						iChip_SE-E	-	-	-	-	-	-
						iChip_SE-U	-	-	-	-	-	-
						iChip_UK	-	-	-	-	-	-
						iChip_US	-	-	-	-	-	-
						iChip_i2b2	-	-	-	-	-	-
						BBJ	2,414	14,245	0.532	0.512	1.08 (1.02-1.16)	0.011
						Kyoto	1,237	2,087	0.545	0.505	1.19 (1.07-1.33)	0.0015
						Korea	799	751	0.475	0.459	1.07 (0.93-1.24)	0.35
Replication	IORRA2	2,354	5,252	0.550	0.516	1.13 (1.06-1.22)	5.0.E-04					
BBJ3	954	3,105	0.532	0.518	1.06 (0.95-1.18)	0.29						
Meta-analysis	-	16,633	54,807	0.627	0.620	1.08 (1.05-1.11)	3.3.E-08					
rs56656810	16	30,788,759	A/C	MIR762	GWAS	Brass	483	1,631	0.287	0.261	1.13 (0.94-1.35)	0.18
						Canada	589	1,554	0.287	0.265	1.10 (0.95-1.29)	0.21
						Eira	1,097	1,044	0.283	0.263	1.11 (0.97-1.27)	0.15
						Narac1	863	1,191	0.263	0.270	1.01 (0.86-1.20)	0.89
						Narac2	896	6,603	0.256	0.271	0.95 (0.85-1.07)	0.42
						Wtccc	1,520	10,507	0.279	0.258	1.12 (1.02-1.22)	0.015
						ReAct	275	804	0.294	0.253	1.28 (1.02-1.59)	0.030
						Dutch	1,172	1,684	0.261	0.251	1.08 (0.95-1.22)	0.24
						ACR-Ref	347	264	0.285	0.268	1.12 (0.83-1.51)	0.46
						Corrona	894	1,838	0.283	0.262	1.12 (0.98-1.27)	0.10
						Vandy	739	2,247	0.256	0.255	1.00 (0.88-1.15)	0.9562
						iChip_ES	397	399	0.283	0.263	1.11 (0.86-1.42)	0.42
						iChip_NL	303	2,001	0.264	0.251	1.06 (0.85-1.33)	0.60
						iChip_SE-E	740	1,117	0.276	0.265	1.05 (0.91-1.22)	0.52
						iChip_SE-U	522	962	0.306	0.273	1.20 (1.00-1.43)	0.044
						iChip_UK	1,645	6,082	0.275	0.252	1.11 (1.01-1.22)	0.023
						iChip_US	997	2,132	0.292	0.277	1.09 (0.96-1.23)	0.20
						iChip_i2b2	882	1,863	0.274	0.261	1.08 (0.95-1.23)	0.23
						BBJ	2,414	14,245	0.087	0.087	1.00 (0.90-1.12)	1.00
						Kyoto	1,237	2,087	0.089	0.069	1.33 (1.09-1.62)	0.0042
						Korea	799	751	0.091	0.082	1.16 (0.88-1.53)	0.30
Replication	IORRA2	2,354	5,252	0.093	0.084	1.14 (1.00-1.29)	0.046					
BBJ3	954	3,105	0.094	0.089	1.08 (0.90-1.29)	0.43						
Meta-analysis	-	22,119	69,363	0.211	0.196	1.09 (1.05-1.12)	3.6.E-08					
rs11089637	22	21,979,096	C/T	MIR301B	GWAS	Brass	483	1,631	0.172	0.165	1.12 (0.91-1.37)	0.28
						Canada	589	1,554	0.171	0.157	1.17 (0.96-1.42)	0.12
						Eira	1,097	1,044	0.193	0.189	1.05 (0.89-1.23)	0.57
						Narac1	863	1,191	0.186	0.157	1.27 (1.03-1.57)	0.027
						Narac2	896	6,603	0.166	0.163	1.04 (0.90-1.20)	0.62
						Wtccc	1,520	10,507	0.182	0.166	1.14 (1.03-1.28)	0.017
						ReAct	275	804	0.168	0.159	1.08 (0.81-1.44)	0.60
						Dutch	1,172	1,684	0.191	0.187	1.03 (0.88-1.21)	0.68
						ACR-Ref	347	264	0.174	0.170	0.92 (0.60-1.40)	0.70
						Corrona	894	1,838	0.173	0.157	1.12 (0.96-1.31)	0.14
						Vandy	739	2,247	0.167	0.154	1.10 (0.94-1.30)	0.23
						iChip_ES	397	399	0.188	0.158	1.15 (0.85-1.55)	0.36
						iChip_NL	303	2,001	0.173	0.162	1.13 (0.87-1.45)	0.36
						iChip_SE-E	740	1,117	0.183	0.183	1.00 (0.84-1.18)	0.95
						iChip_SE-U	522	962	0.178	0.179	1.02 (0.83-1.26)	0.85
						iChip_UK	1,645	6,082	0.175	0.150	1.20 (1.08-1.33)	0.0010
						iChip_US	997	2,132	0.171	0.160	1.10 (0.94-1.28)	0.23
						iChip_i2b2	882	1,863	0.174	0.166	1.15 (0.99-1.35)	0.071
						BBJ	2,414	14,245	0.472	0.464	1.03 (0.97-1.10)	0.33
						Kyoto	1,237	2,087	0.459	0.454	1.02 (0.91-1.14)	0.73
						Korea	799	751	0.440	0.419	1.09 (0.95-1.26)	0.23
Replication	IORRA2	2,354	5,252	0.484	0.468	1.06 (0.99-1.14)	0.12					
BBJ3	954	3,105	0.473	0.468	1.02 (0.92-1.13)	0.66						
Meta-analysis	-	22,119	69,363	0.280	0.273	1.07 (1.05-1.10)	3.7.E-07					

## オーバーラップ miRNA の eQTL データとの統合解析

最後に、上記の 4 つの miRNA (表 6) のうち、最も TSI が高く、さらに関節リウマチの病態において最も重要な免疫細胞における特異的発現を認めた hsa-miR-762 について、組織特異的トランスクリプトームカタログである GTEx データベース<sup>35</sup>との統合解析を行った。MIGWAS で検出された miRNA は標的遺伝子の転写物に影響を与えて形質に関与することを仮定していることから、hsa-miR-762 の 41 の標的遺伝子の関節リウマチへの関与に関しても、対象臓器での発現量調節が重要であることを検証することとした。

表 6 : MIGWAS と miRNA 実験での発現変動とのオーバーラップを認めた miRNA と、その標的遺伝子

Candidate biomarker miRNA	RA cases vs controls		TSI	Target genes
	log <sub>2</sub> fold change in expression	FDR-q		
hsa-miR-93-5p	-0.522	0.00022	0.254	<b>ANKH</b> , <b>ANKRD52</b> , <b>ARCN1</b> , <b>BCL2L15</b> , <b>C7orf43</b> , <b>CASP8</b> , CDKN1A, DENND1B, FAM126B, FAM133B, FYCO1, IKZF4, KIAA1109, <b>KLF2</b> , <b>LDLR</b> , MAGI3, PEX13, <b>PHTF1</b> , <b>PRDX5</b> , RAB5B, RAG1, RSBN1, <b>SAR1B</b> , SEC24A, SLC12A5, SNN, STK38, TAGAP, TRAK2, ZBTB10
hsa-miR-106b-5p	-0.811	0.00022	0.269	<b>ANKH</b> , <b>ANKRD52</b> , <b>BCL2L15</b> , <b>C7orf43</b> , <b>CASP8</b> , CDK6, CDKN1A, <b>CEP76</b> , FAM126B, <b>FAM167A</b> , FAM65B, FYCO1, ICOS, <b>KLF2</b> , <b>LDLR</b> , MAGI3, PA2G4, PAPOLG, <b>PDGFB</b> , RAB5B, RSBN1, <b>RTKN2</b> , <b>SAR1B</b> , STK38, TMEM151B, TNFAIP3, VPS37C, <b>ZFP36L1</b>
hsa-miR-301b-3p	-0.954	0.0090	0.397	C5orf30, CDK6, DDX6, <b>ITSN1</b> , <b>LDLR</b> , PAN3, <b>SAR1B</b> , SERBP1, SNRPE, SRSF3, <b>TMEM50B</b> , ZBTB10
hsa-miR-762	1.154	0.043	0.982	ARHGAP20, C11orf20, C11orf9, C1orf93, C7orf59, CPNE5, <b>DAP</b> , <b>FADS2</b> , <b>GATS</b> , <b>GNAI2</b> , HDAC5, <b>INPP5B</b> , <b>IQGAP1</b> , <b>IRF5</b> , <b>ITSN1</b> , <b>PADI2</b> , <b>PFKL</b> , <b>PGAP3</b> , PHF15, PHKG2, PHLDB1,

---

*PKNOX2, **PPIL2**, RAF1,*  
*RAVER1, SLC25A23,*  
***SLC44A2**, STAC2, **SYNGR1**,*  
*TAB1, TMEM151B,*  
***TMPRSS3**, TNFRSF14,*  
*TRAF1, **TSPAN33**,*  
***UBASH3A**, UPK2, **UTP11L**,*  
*VPS37C, **YDJC**, **ZNF594***

---

標的遺伝子は 2 つ以上のアルゴリズムで上位 1 パーセントイルのスコアに分類されるものを抽出した。太字の遺伝子は、GTEx データで全血での eGene に含まれるもの。TSI; tissue specificity index, RA; rheumatoid arthritis.

まず、免疫細胞の代用としての全血細胞での eQTL 解析結果の *eGene*(周囲の遺伝子による cis-eQTL 調節によって発現量に変化する遺伝子)の中に、hsa-miR-762 の 41 の標的遺伝子は偶然よりも多く含まれることを示した(1.46-fold enrichment with binomial  $P$  value = 0.014)。これは、全血細胞において、hsa-miR-762 の標的遺伝子は周囲の遺伝子多型による発現量調節を受けていることを意味している。なお、この 41 の標的遺伝子の生物学的特徴を見出すために EnrichR(URL を参照)を使用したパスウェイ解析を実行したところ、関節リウマチ疾患パスウェイ・自己免疫疾患パスウェイ・TNF-alpha signaling パスウェイ・TNFR induced NFkappaB signaling パスウェイに特にエンリッチしていた。

つぎに、hsa-miR-762 の標的遺伝子の発現量を調節する多型と、関節リウマチへの関与について検証するために、summary data-based Mendelian randomization (SMR)<sup>36</sup>を実施した。この解析により、hsa-miR-762 の標的遺伝子周囲に発現量変動を介して linkage でなく causal あるいは pleiotropy を介して関節リウマチに関与している多型が存在するかどうかを検証し、検出することができた(表7)。例えば、hsa-miR-762 の標的遺伝子の一つである *SYNGRI* (シナプス可塑性に関わることが知られ、これまで GWAS では関節リウマチの他、知能指数、原発性胆汁性

肝硬変、炎症性腸疾患といった形質への関与が報告されている)周囲の遺伝子多型は、SMR 解析にて最も強いシグナルを示し関節リウマチへも強い関連を示した(図 6 e)。特に rs2069235 の関節リウマチ GWAS シグナル( $P_{\text{GWAS}} = 7.6 \times 10^{-13}$ )は、eQTL 効果( $P_{\text{eQTL}} = 1.3 \times 10^{-38}$ )を介していることが分かった( $P_{\text{SMR}} = 3.5 \times 10^{-10}$ )。すなわち、関節リウマチ患者では hsa-miR-762 の発現が特に免疫細胞で亢進しており、このターゲット遺伝子である *SYNGRI* もまた、発現量調節を介して関節リウマチの発症に強く関与していることが示された。

表 7 : Summary-level data Mendelian Randomization の結果

Gene	Top SNP	Allele 1	Allele 2	Freq	b_GWAS	se_GWAS	p_GWAS	b_eQTL	se_eQTL	p_eQTL	b_SMR	se_SMR	p_SMR	p_HEIDI
<i>CPNE5</i>	rs1064827	T	C	0.251	0.033	0.018	6.5E-02	0.191	0.031	1.3E-09	0.171	0.097	7.7E-02	3.8E-01
<i>DAP</i>	rs5745297	A	G	0.080	-0.023	0.048	6.3E-01	0.873	0.051	3.6E-66	-0.027	0.055	6.3E-01	6.2E-01
<i>FADS2</i>	rs968567	T	C	0.150	-0.104	0.021	7.3E-07	0.458	0.038	7.8E-33	-0.227	0.050	4.8E-06	5.1E-01
<i>GATS</i>	rs2528900	A	G	0.268	-0.057	0.017	6.6E-04	0.911	0.032	4.3E-184	-0.062	0.018	7.2E-04	7.1E-02
<i>INPP5B</i>	rs28525112	C	G	0.445	-0.093	0.018	1.4E-07	-0.595	0.028	2.2E-100	0.156	0.031	3.1E-07	4.5E-02
<i>IQGAP1</i>	rs2074585	G	A	0.466	-0.056	0.018	1.5E-03	-0.522	0.027	2.2E-80	0.108	0.034	1.7E-03	1.2E-01
<i>IRF5</i>	rs10954215	T	C	0.368	-0.055	0.013	4.6E-05	1.173	0.027	0.0E+00	-0.047	0.011	4.9E-05	1.9E-12
<i>ITSN1</i>	rs2300375	T	C	0.304	-0.054	0.017	1.0E-03	-0.525	0.030	1.6E-68	0.104	0.032	1.3E-03	2.0E-02
<i>PADI2</i>	rs11576552	C	A	0.368	-0.089	0.016	5.2E-08	-0.517	0.029	3.7E-73	0.172	0.033	1.8E-07	5.0E-04
<i>PFKL</i>	rs2838547	C	G	0.161	0.070	0.020	4.1E-04	0.340	0.041	5.4E-17	0.205	0.063	1.1E-03	1.3E-01
<i>PGAP3</i>	rs2934951	A	G	0.318	-0.083	0.014	2.3E-09	0.471	0.030	3.0E-57	-0.177	0.032	2.2E-08	2.0E-03
<i>RAF1</i>	rs12629082	G	A	0.432	0.052	0.018	3.8E-03	0.280	0.028	8.1E-24	0.187	0.067	5.3E-03	3.0E-01
<i>SLC25A23</i>	rs393375	A	T	0.469	-0.007	0.021	7.3E-01	-0.309	0.028	9.9E-29	0.024	0.069	7.3E-01	1.2E-01
<i>SLC44A2</i>	rs78671428	T	G	0.087	0.046	0.029	1.1E-01	-0.972	0.053	2.5E-75	-0.048	0.030	1.2E-01	5.9E-01
<i>SYNGR1</i>	rs2069235	A	G	0.299	0.113	0.016	7.6E-13	0.394	0.030	1.3E-38	0.288	0.046	3.5E-10	2.7E-01
<i>TMPRSS3</i>	rs2277798	A	G	0.374	-0.010	0.014	4.6E-01	-0.290	0.028	1.6E-25	0.035	0.048	4.6E-01	7.5E-05
<i>TNFRSF14</i>	rs7544646	C	G	0.531	0.057	0.013	2.3E-05	-0.445	0.028	1.5E-57	-0.128	0.031	4.3E-05	2.3E-04
<i>TRAF1</i>	rs4310279	G	A	0.278	0.063	0.015	2.5E-05	-0.418	0.031	9.6E-42	-0.152	0.038	5.7E-05	5.6E-02
<i>TSPAN33</i>	rs12667978	C	T	0.458	0.046	0.021	2.7E-02	0.293	0.029	3.7E-24	0.158	0.073	3.0E-02	1.1E-02
<i>UBASH3A</i>	rs1893592	C	A	0.276	-0.106	0.015	3.7E-12	0.464	0.031	1.1E-50	-0.228	0.036	3.0E-10	5.8E-07
<i>UTP11L</i>	rs35129114	T	C	0.264	-0.101	0.020	2.8E-07	-0.243	0.030	9.5E-16	0.417	0.096	1.5E-05	5.8E-02

Freq; allele frequency of allele 1, b\_GWAS; beta coefficient of allele 1 in GWAS, se\_GWAS; standard error in GWAS, p\_GWAS; P value in GWAS, b\_eQTL; beta coefficient of allele 1 in eQTL, se\_eQTL; standard error in eQTL, p\_eQTL; P value in eQTL, b\_SMR; beta coefficient of allele 1 in SMR, se\_SMR; standard error in SMR, p\_SMR; P value in SMR, p\_HEIDI; P value in HEIDI analysis.



## 考 察

我々は、大規模 GWAS データ、miRNA-標的遺伝子ネットワーク、網羅的細胞特異的 miRNA 発現データの 3 者を MIGWAS により統合解析し、組織特異的な miRNA の疾患病態への関与を示した。MIGWAS を多様なヒト複雑形質に応用することで、miRNA の疾患病態への関与は非常に組織特異性が高いことがはじめて明らかになった。さらに、MIGWAS を応用することで、どの miRNA が疾患に関与しているかの情報も得ることができる。この情報については、関節リウマチ患者・健常人とでの発現量変動解析により実証実験を行うことでバリデーションされ、さらに新規 GWAS メタ解析を行うことでインシリコにもレプリケーションされた。新規 GWAS メタ解析では関節リウマチの新規関連 2 領域を同定した。また大規模 eQTL データベースとの統合解析により、候補 miRNA が候補標的遺伝子を介してどのように関節リウマチの発症病態に関与しているかも推測することができた。

MIGWAS は特に 4 つの特性が重要と考えられる。

- (i) miRNA-標的遺伝子ネットワークにおける GWAS シグナルの集積

を定量化すること

- (ii) (i)がどの細胞種に特に集積しているかを定量することで、形質において重要な組織を特定すること
- (iii) フォローアップ研究のために、GWAS シグナルの中から候補領域を特定すること
- (iv) 疾患病態に重要と考えられる候補 miRNA を特定すること

我々の研究結果から、miRNA は標的遺伝子の転写物に影響を与え、さらに標的遺伝子は発現量調節を介して形質に関与しているという病態の一例を検出することができた。また、候補 miRNA に関しては、バイオマーカーや創薬ターゲットのための実験検証のために prioritization をできるという意味で、GWAS 単独ではできなかった意義深いものと考えている。

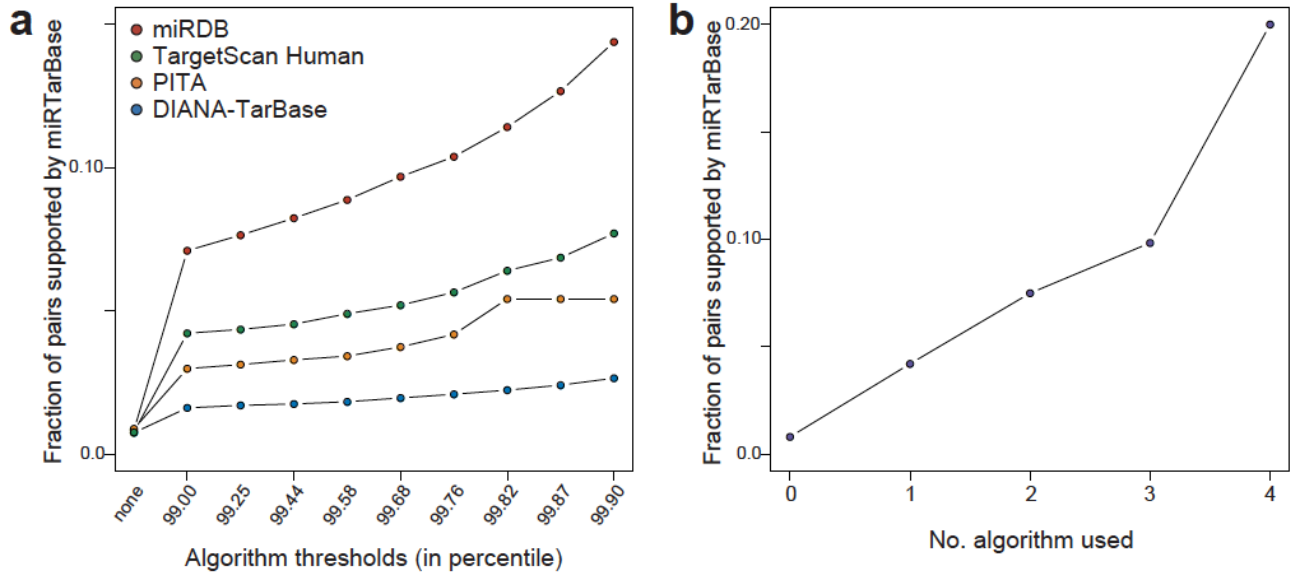
最後に、本研究の limitation と将来への展望を数点挙げる。

第一に、今回、網羅的な細胞種特異的 miRNA 発現データは健常人由来のものであったが、疾患においては miRNA プロファイルが異なる可能性は考慮される。これまで、トランスクリプトームやエピゲノムを用いて重要な細胞種・組織を GWAS から検出する研究<sup>12,32</sup>もまた、健常

人由来のデータを使用していることからそれに倣ったが、もし将来的に疾患由来の網羅的発現データが得られることがあれば、検討の余地はあるものと思われる。

第二に、miRNA と標的遺伝子の予測アルゴリズムには不確実性があることが挙げられる。これに対応するため、4つのアルゴリズムからの統計量の統合を行うという手法を取っており、実際、最近公開された miRTarBase という実験的な miRNA と標的遺伝子の結合関係の情報データベースと照らし合わせる<sup>37</sup>と、複数のデータベースの統合解析により正確な予測が可能となったことが判明した(図8)。

図 8 : miRNA-標的遺伝子のインシリコ予測アルゴリズムのインビトロデータベース (miRTarBase) による再現確認



The fraction of *in silico* predicted miRNA-gene target pairs supported by miRTarBase (either reporter assay, western blot, microarray or next-generation sequencing) by the thresholds in percentile for each algorithm (a) and by the number of algorithms used (b).

In silico に予測された miRNA- 遺伝子 標的 ペアの うち、miRTarBase(reporter assay, western blot, microarray あるいは次世代シーケンサー)でも実験的にペアであることが示されている割合を (a)各アルゴリズムのしきい値ごとにまたは (b)統合解析するうえで使用したアルゴリズム数ごとにプロットしたもの

第三に、GWAS を元にしたエンリッチメント解析では GWAS の統計量自体のインフレーションで MIGWAS の統計量のインフレーションが引き起こされる危険性も考慮された。これに対応するため、今回エンリッチメント解析においてはパーミュテーション法を用いた帰無分布を使用しており、人為的にインフレーションさせた GWAS でのシミュレーションでも統計量のインフレーションを認めないことを以前示している<sup>38</sup>。

第四に、今回の手法に組織特異的遺伝子発現の情報を組み入れていないことが挙げられる。miRNA×遺伝子ごとの組織特異的発現の組み合わせでの計算が膨大であることからそれについての詳細な検討を行わなかったが、将来的にさらなる統合解析を行うことは興味深いと考えられる。

第五に、本解析では、miRNA の発現量が周囲の cis-eQTL 効果を持つ SNP により遺伝子発現と同様の調節を受けていることを仮定した。現時点で miRNA eQTL の網羅的データベースが存在しないのが現状である。また、今回同定した 4 つの miRNA の周囲にも、*GAL3ST4*, *RNF40*, *PHKG2*, *UBE2L3* 等の遺伝子が存在することから、これらの遺伝子の影響を完全に排除できない。miRNA-cis-eQTL 効果の仮定に関しては今後次世代シ

ークエンサーを用いた網羅的発現データと遺伝子多型情報の解析により明らかにする必要がある。

## 結 語

本研究で開発した手法 MIGWAS により、細胞種特異的な文脈で miRNA とその標的遺伝子がヒトの複雑形質に関与していることが示され、またバイオマーカー候補 miRNA を同定した。

この 20 年で GWAS に代表されるヒトゲノム情報の解析がもたらした生命現象への知見は大きいですが、その一方で未だその解釈に課題が残っていることも事実である。ここ数年はハイスループットな実験系の発展が花盛りを迎えており、マルチオミクス情報は急速に蓄積されつつある。今回の研究で実施したような多層の生命情報の統合解析を行うことで、生まれ持ったゲノム情報が、どのようにしてその後の時空間的な文脈や環境因子への暴露によって我々の多様性を形作るかを理解すること、それによりゲノム研究を臨床現場にまで還元することに貢献したいと考えている。

## 引用文献

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock



GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C,  
Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W,  
Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR,  
Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal  
A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang  
G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola  
AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR,  
Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans  
GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H,  
Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer  
K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S,  
Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp  
M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG,  
Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W,  
Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P,  
Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T,  
Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G,  
Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L,

- Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium.. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
  3. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654 (2002).
  4. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001-6 (2014).
  5. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, Gan W, Kitajima H, Taliun D, Rayner NW, Guo X, Lu Y, Li M, Jensen RA, Hu Y, Huo S, Lohman KK, Zhang W, Cook JP, Prins BP, Flannick J, Grarup N, Trubetsky

VV, Kravic J, Kim YJ, Rybin DV, Yaghootkar H, Müller-Nurasyid M, Meidtner K, Li-Gao R, Varga TV, Marten J, Li J, Smith AV, An P, Ligthart S, Gustafsson S, Malerba G, Demirkan A, Tajes JF, Steinthorsdottir V, Wuttke M, Lecoeur C, Preuss M, Bielak LF, Graff M, Highland HM, Justice AE, Liu DJ, Marouli E, Peloso GM, Warren HR; ExomeBP Consortium; MAGIC Consortium; GIANT Consortium, Afaq S, Afzal S, Ahlqvist E, Almgren P, Amin N, Bang LB, Bertoni AG, Bombieri C, Bork-Jensen J, Brandslund I, Brody JA, Burtt NP, Canouil M, Chen YI, Cho YS, Christensen C, Eastwood SV, Eckardt KU, Fischer K, Gambaro G, Giedraitis V, Grove ML, de Haan HG, Hackinger S, Hai Y, Han S, Tybjærg-Hansen A, Hivert MF, Isomaa B, Jäger S, Jørgensen ME, Jørgensen T, Käräjämäki A, Kim BJ, Kim SS, Koistinen HA, Kovacs P, Kriebel J, Kronenberg F, Läll K, Lange LA, Lee JJ, Lehne B, Li H, Lin KH, Linneberg A, Liu CT, Liu J, Loh M, Mägi R, Mamakou V, McKean-Cowdin R, Nadkarni G, Neville M, Nielsen SF, Ntalla I, Peyser PA, Rathmann W, Rice K, Rich SS, Rode L, Rolandsson O, Schönherr S, Selvin E, Small KS, Stančáková A, Surendran P, Taylor KD, Teslovich TM, Thorand B, Thorleifsson G, Tin A, Tönjes A, Varbo A, Witte DR, Wood AR, Yajnik P, Yao J, Yengo L, Young R, Amouyel P, Boeing H, Boerwinkle E, Bottinger EP, Chowdhury R, Collins FS, Dedoussis G,

Dehghan A, Deloukas P, Ferrario MM, Ferrières J, Florez JC, Frossard P, Gudnason V, Harris TB, Heckbert SR, Howson JMM, Ingelsson M, Kathiresan S, Kee F, Kuusisto J, Langenberg C, Launer LJ, Lindgren CM, Männistö S, Meitinger T, Melander O, Mohlke KL, Moitry M, Morris AD, Murray AD, de Mutsert R, Orho-Melander M, Owen KR, Perola M, Peters A, Province MA, Rasheed A, Ridker PM, Rivadineira F, Rosendaal FR, Rosengren AH, Salomaa V, Sheu WH, Sladek R, Smith BH, Strauch K, Uitterlinden AG, Varma R, Willer CJ, Blüher M, Butterworth AS, Chambers JC, Chasman DI, Danesh J, van Duijn C, Dupuis J, Franco OH, Franks PW, Froguel P, Grallert H, Groop L, Han BG, Hansen T, Hattersley AT, Hayward C, Ingelsson E, Kardia SLR, Karpe F, Kooner JS, Köttgen A, Kuulasmaa K, Laakso M, Lin X, Lind L, Liu Y, Loos RJJ, Marchini J, Metspalu A, Mook-Kanamori D, Nordestgaard BG, Palmer CNA, Pankow JS, Pedersen O, Psaty BM, Rauramaa R, Sattar N, Schulze MB, Soranzo N, Spector TD, Stefansson K, Stumvoll M, Thorsteinsdottir U, Tuomi T, Tuomilehto J, Wareham NJ, Wilson JG, Zeggini E, Scott RA, Barroso I, Frayling TM, Goodarzi MO, Meigs JB, Boehnke M, Saleheen D, Morris AP, Rotter JI, McCarthy MI. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes article. *Nat. Genet.* **50**, 559–571

- (2018).
6. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21 (2008).
  7. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
  8. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
  9. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science.* **348**, 648–60 (2015).
  10. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szigyaró CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen

- M, von Heijne G, Nielsen J, Pontén F. Tissue-based map of the human proteome. *Science*. **347**, 1260419–1260419 (2015).
11. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology* vol. 17 451–459 (2016).
  12. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
  13. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Zhao JH, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Leach IM, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge

RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J,  
Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E,  
Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C,  
Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL,  
Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Chen YI, Clarke R, Daw  
EW, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K,  
Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B,  
Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H,  
Grammer TB, Gräßler J, Grönberg H, Groves CJ, Gusto G, Haessler J, Hall P,  
Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL,  
Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM,  
Johansson Å, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M,  
Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L,  
Lindström J, Lo KS, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PKE,  
Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani  
L, Moayyeri A, Monda KL, Morcken MA, Mulas A, Müller G, Müller-Nurasyid  
M, Musk AW, Nagaraja R, Nöthen MM, Nolte IM, Pilz S, Rayner NW,  
Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S,

Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Smith AV, Smolonska J, Stanton AV, Steinhorsdottir V, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen AC, Tan ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q; LifeLines Cohort Study, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gådin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, van 't Hooft FM, Vinkhuyzen AAE, Westra HJ, Zheng W, Zondervan KT; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN,



Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A,  
Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh  
J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson  
JG, Farrall M, Felix SB, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester  
T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason  
V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorf LA,  
Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC,  
Hyppönen E, Illig T, Jacobs KB, Jarvelin MR, Jöckel KH, Johansen B, Jousilahti  
P, Jukema JW, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM,  
Kiemeny LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari  
M, Kuusisto J, Lakka TA, Langenberg C, Marchand LL, Lehtimäki T, Lyssenko  
V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL,  
Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong  
KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller  
PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM,  
Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski  
MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP,  
Strauch K, Tönjes A, Trégouët DA, Tremblay A, Tremoli E, Virtamo J, Vohl

MC, Völker U, Waeber G, Willemssen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJJ, Speliotes EK. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

14. Lee, R., Feinbaum, R. & Ambros, V. A short history of a short RNA. *Cell* **116**, S89-92, 1 p following S96 (2004).
15. Lee, R. C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–4 (2001).
16. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–8 (2005).
17. Krützfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M, Stoffel M. Silencing of microRNAs in vivo with ‘antagomirs’. *Nature* **438**, 685–9 (2005).
18. Titze-de-Almeida, R., David, C. & Titze-de-Almeida, S. S. The Race of 10 Synthetic RNAi-Based Drugs to the Pharmaceutical Market. *Pharm. Res.* **34**, 1339–1363 (2017).
19. Murata K, Furu M, Yoshitomi H, Ishikawa M, Shibuya H, Hashimoto M, Imura Y, Fujii T, Ito H, Mimori T, Matsuda S. Comprehensive microRNA analysis identifies miR-24 and miR-125a-5p as plasma biomarkers for rheumatoid arthritis. *PLoS One* **8**, e69118 (2013).

20. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, Chen BH, Murabito JM, Yao C, Courchesne P, Munson PJ, O'Donnell CJ, Cox N, Johnson AD, Larson MG, Levy D, Freedman JE. Genome-wide identification of microRNA expression quantitative trait loci. *Nat. Commun.* **6**, 1–9 (2015).
21. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, Mari B, Barbry P, Mosnier JF, Hébuterne X, Harel-Bellan A, Mograbi B, Darfeuille-Michaud A, Hofman P. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat. Genet.* **43**, 242–245 (2011).
22. de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, Åström G, Babina M, Bertin N, Burroughs AM, Carlisle AJ, Daub CO, Detmar M, Deviatiiarov R, Fort A, Gebhard C, Goldowitz D, Guhl S, Ha TJ, Harshbarger J, Hasegawa A, Hashimoto K, Herlyn M, Heutink P, Hitchens KJ, Hon CC, Huang E, Ishizu Y, Kai C, Kasukawa T, Klinken P, Lassmann T, Lecellier CH, Lee W, Lizio M, Makeev V, Mathelier A, Medvedeva YA, Mejhert N, Mungall CJ, Noma S, Ohshima M, Okada-Hatakeyama M, Persson H, Rizzu P, Roudnický F, Sætrum P, Sato H, Severin J, Shin JW, Swoboda RK, Tarui H, Toyoda H, Vitting-Seerup K, Winteringham L, Yamaguchi Y, Yasuzawa K, Yoneda M,

- Yumoto N, Zabierowski S, Zhang PG, Wells CA, Summers KM, Kawaji H, Sandelin A, Rehli M; FANTOM Consortium, Hayashizaki Y, Carninci P, Forrest ARR, de Hoon MJL. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* **35**, 872–878 (2017).
23. Ayellet, V. S., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glyceic traits. *PLoS Genet.* **6**, (2010).
24. Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, Zhao J, Zhou X, Reveille JD, Jin L, Amos CI, Xiong M. Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur. J. Hum. Genet.* **18**, 111–117 (2010).
25. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, Rheinheimer S, Meder B, Stähler C, Meese E, Keller A. Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* **44**, 3865–3877 (2016).
26. Nagai A, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, Ninomiya T, Tamakoshi A, Yamagata Z, Mushiroda T, Murakami Y, Yuji K, Furukawa Y, Zembutsu H, Tanaka T, Ohnishi Y, Nakamura Y; BioBank Japan Cooperative Hospital Group, Kubo M. Overview of the BioBank Japan Project: Study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).

27. Ziemann, M., Kaspi, A. & El-Osta, A. Evaluation of microRNA alignment techniques. *RNA* **22**, 1120–1138 (2016).
28. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AA, Westra HJ, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney AS,

Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LC, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Å, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PK, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen AC, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FV, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R,

Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF,  
Zhang Q, Arveiler D, Bakker SJ, Beilby J, Bergman RN, Bergmann S, Biffar R,  
Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ,  
Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC,  
Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J,  
Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester  
T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V,  
Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC,  
Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh  
GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin MR, Jousilahti  
P, Jula AM, Kaprio J, Kastelein JJ, Kayser M, Kee F, Keinanen-Kiukaanniemi  
SM, Kiemeny LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT,  
Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T,  
Lupoli S, Madden PA, Männistö S, Manunta P, Marette A, Matise TC, McKnight  
B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC,  
Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G,  
Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC,  
Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski



MA, Schwarz PE, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinhorsdottir V,  
Stolk RP, Tardif JC, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl MC;  
Electronic Medical Records and Genomics (eMEMERGE) Consortium;  
MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study, Amouyel P,  
Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger  
EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PI,  
Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA,  
Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC,  
Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus  
S, Munroe PB, Njølstad I, Oostra BA, Palmer CN, Pedersen NL, Perola M,  
Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa  
E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D,  
Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J,  
Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H,  
Wichmann HE, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM,  
Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE,  
Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI,  
Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ,

Price AL, Lettre G, Loos RJ, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM.. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–86 (2014).

29. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Nex-generation sequencing in muylti-Ethnic Samples (T2D-GENES) Consortium, Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, Horikoshi M, Johnson AD, Ng MC, Prokopenko I, Saleheen D, Wang X, Zeggini E, Abecasis GR, Adair LS, Almgren P, Atalay M, Aung T, Baldassarre D, Balkau B, Bao Y, Barnett AH, Barroso I, Basit A, Been LF, Beilby J, Bell GI, Benediktsson R, Bergman RN, Boehm BO, Boerwinkle E, Bonnycastle LL, Burt N, Cai Q, Campbell H, Carey J, Cauchi S, Caulfield M, Chan JC, Chang LC, Chang TJ, Chang YC, Charpentier G, Chen CH, Chen H, Chen YT, Chia KS, Chidambaram M, Chines PS, Cho NH, Cho YM, Chuang LM, Collins FS, Cornelis MC, Couper DJ, Crenshaw AT, van Dam RM, Danesh

J, Das D, de Faire U, Dedoussis G, Deloukas P, Dimas AS, Dina C, Doney AS, Donnelly PJ, Dorkhan M, van Duijn C, Dupuis J, Edkins S, Elliott P, Emilsson V, Erbel R, Eriksson JG, Escobedo J, Esko T, Eury E, Florez JC, Fontanillas P, Forouhi NG, Forsen T, Fox C, Fraser RM, Frayling TM, Froguel P, Frossard P, Gao Y, Gertow K, Gieger C, Gigante B, Grallert H, Grant GB, Grrop LC, Groves CJ, Grundberg E, Guiducci C, Hamsten A, Han BG, Hara K, Hassanali N, Hattersley AT, Hayward C, Hedman AK, Herder C, Hofman A, Holmen OL, Hovingh K, Hreidarsson AB, Hu C, Hu FB, Hui J, Humphries SE, Hunt SE, Hunter DJ, Hveem K, Hydrie ZI, Ikegami H, Illig T, Ingelsson E, Islam M, Isomaa B, Jackson AU, Jafar T, James A, Jia W, Jöckel KH, Jonsson A, Jowett JB, Kadowaki T, Kang HM, Kanoni S, Kao WH, Kathiresan S, Kato N, Katulanda P, Keinanen-Kiukaanniemi KM, Kelly AM, Khan H, Khaw KT, Khor CC, Kim HL, Kim S, Kim YJ, Kinnunen L, Klopp N, Kong A, Korpi-Hyövälti E, Kowlessur S, Kraft P, Kravic J, Kristensen MM, Krithika S, Kumar A, Kumate J, Kuusisto J, Kwak SH, Laakso M, Lagou V, Lakka TA, Langenberg C, Langford C, Lawrence R, Leander K, Lee JM, Lee NR, Li M, Li X, Li Y, Liang J, Liju S, Lim WY, Lind L, Lindgren CM, Lindholm E, Liu CT, Liu JJ, Lobbens S, Long J, Loos RJ, Lu W, Luan J, Lyssenko V, Ma RC, Maeda S, Mägi R, Männistö S,

Matthews DR, Meigs JB, Melander O, Metspalu A, Meyer J, Mirza G, Mihailov E, Moebus S, Mohan V, Mohlke KL, Morris AD, Mühleisen TW, Müller-Nurasyid M, Musk B, Nakamura J, Nakashima E, Navarro P, Ng PK, Nica AC, Nilsson PM, Njølstad I, Nöthen MM, Ohnaka K, Ong TH, Owen KR, Palmer CN, Pankow JS, Park KS, Parkin M, Pechlivanis S, Pedersen NL, Peltonen L, Perry JR, Peters A, Pinidiyapathirage JM, Platou CG, Potter S, Price JF, Qi L, Radha V, Rallidis L, Rasheed A, Rathman W, Rauramaa R, Raychaudhuri S, Rayner NW, Rees SD, Rehnberg E, Ripatti S, Robertson N, Roden M, Rossin EJ, Rudan I, Rybin D, Saaristo TE, Salomaa V, Saltevo J, Samuel M, Sanghera DK, Saramies J, Scott J, Scott LJ, Scott RA, Segrè AV, Sehmi J, Sennblad B, Shah N, Shah S, Shera AS, Shu XO, Shuldiner AR, Sigurdsson G, Sijbrands E, Silveira A, Sim X, Sivapalaratnam S, Small KS, So WY, Stančáková A, Stefansson K, Steinbach G, Steinhorsdottir V, Stirrups K, Strawbridge RJ, Stringham HM, Sun Q, Suo C, Syvänen AC, Takayanagi R, Takeuchi F, Tay WT, Teslovich TM, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tikkanen E, Trakalo J, Tremoli E, Trip MD, Tsai FJ, Tuomi T, Tuomilehto J, Uitterlinden AG, Valladares-Salgado A, Vedantam S, Veglia F, Voight BF, Wang C, Wareham NJ, Wennauer R, Wickremasinghe AR, Wilsgaard T, Wilson JF, Wiltshire S, Winckler W,

- Wong TY, Wood AR, Wu JY, Wu Y, Yamamoto K, Yamauchi T, Yang M, Yengo L, Yokota M, Young R, Zabaneh D, Zhang F, Zhang R, Zheng W, Zimmet PZ, Altshuler D, Bowden DW, Cho YS, Cox NJ, Cruz M, Hanis CL, Kooner J, Lee JY, Seielstad M, Teo YY, Boehnke M, Parra EJ, Chambers JC, Tai ES, McCarthy MI, Morris AP. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–44 (2014).
30. Chen, J.-Q., Papp, G., Szodoray, P. & Zeher, M. The role of microRNAs in the pathogenesis of autoimmune diseases. *Autoimmun. Rev.* **15**, 1171–1180 (2016).
31. Malmström, V., Catrina, A. I. & Klareskog, L. The immunopathogenesis of seropositive rheumatoid arthritis: from triggering to targeting. *Nat. Rev. Immunol.* **17**, 60–75 (2017).
32. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shores N, Genovese G, Saunders A, Macosko E, Pollack S; Brainstorm Consortium, Perry JRB, Buenrostro JD, Bernstein BE, Raychaudhuri S, McCarroll S, Neale BM, Price AL. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

33. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, Ripke S, Day FR; ReproGen Consortium; Schizophrenia Working Group of the Psychiatric Genomics Consortium; RACI Consortium, Purcell S, Stahl E, Lindstrom S, Perry JR, Okada Y, Raychaudhuri S, Daly MJ, Patterson N, Neale BM, Price AL. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
34. 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* **506**, 376–381 (2014).
35. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
36. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
37. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, Hsieh PJ, Wu YS, Chen WL, Li MJ, Wu YC, Huang XY, Ng FL, Buddhakosai W, Huang PC, Lan KC, Huang CY, Weng SL, Cheng YN, Liang C, Hsu WL, Huang HD. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic*

*Acids Res.* **46**, D296–D302 (2018).

38. 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.* 1–9 (2016) doi:10.1038/srep22223.



## 解析ソフトウェア・ウェブサイト

- MIGWAS source code, <https://github.com/saorisakaue/MIGWAS>
- FANTOM5 consortium, <http://fantom.gsc.riken.jp/5/>
- The BioBank Japan Project (BBJ),  
<https://biobankjp.org/english/index.html>
- miRBase, <http://www.mirbase.org/>
- miRDB, <http://www.mirdb.org/>
- TargetScan Human, [http://www.targetscan.org/vert\\_72/](http://www.targetscan.org/vert_72/)
- DIANA-TarBase, <http://diana.imis.athena-innovation.gr/DianaTools/index.php>
- PITA, [https://genie.weizmann.ac.il/pubs/mir07/mir07\\_data.html](https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html)
- GTEx consortium, <https://www.gtexportal.org/home/>
- CAGE eQTL data, <http://cnsgenomics.com/shiny/CAGE/>
- R DESeq2 package,  
<https://bioconductor.org/packages/release/bioc/html/DESeq2.html>
- R preprocessCore package,  
<https://bioconductor.org/packages/release/bioc/html/preprocessCore.html>
- SMR software, [cnsgenomics.com/software/smr/](http://cnsgenomics.com/software/smr/)

- CIRCOS, <http://circos.ca/>
- Cutadapt, [cutadapt.readthedocs.io/en/stable/](http://cutadapt.readthedocs.io/en/stable/)
- Bowtie, <http://bowtie-bio.sourceforge.net/index.shtml>
- EnrichR, <https://amp.pharm.mssm.edu/Enrichr/>

## 謝 辞

本研究の遂行にあたっては、多くの方々のご指導とご協力を頂きました。

はじめに、本研究の趣旨に快く賛同いただき末梢血をご提供いただいた関節リウマチの患者さん、健常協力者の方々に感謝を申し上げます。これまでの大規模 GWAS ゲノムデータ構築にあたり DNA を提供いただきました多くの方々にも感謝申し上げます。

本論文は、東京大学大学院医学系研究科アレルギー・リウマチ学教室山本一彦名誉教授（現理化学研究所生命医科学センター副センター長）および、東京大学大学院医学系研究科アレルギー・リウマチ学教室藤尾圭志教授のご指導、ご鞭撻のもとにまとめられたものです。心からの感謝の意を申し上げます。研究にあたっては終始あたたかいご指導・ご助言および深い理解を賜りました大阪大学大学院医学系研究科遺伝統計学教室、岡田随象教授に心よりの感謝を申し上げます。

論文審査にあたっては、東京大学医学部医学系研究科内科学 血液腫瘍病態学黒川峰夫教授、同社会医学 衛生学 石川俊平教授、同内科学 代謝・栄養病態学 岩部真人特任准教授、同国際保健学 人類遺伝学 藤本明洋教授・東京大学先端科学技術研究センターゲノムサイ

エンス分野 油谷浩幸教授の貴重なご助言を賜りました。深く感謝申し上げます。

大阪大学大学院医学系研究科呼吸器・免疫内科学の熊ノ郷淳教授、前田悠一先生、新居卓朗先生には関節リウマチの患者さんの検体をご提供いただきましたことに感謝申し上げます。大阪大学大学院医学系研究科遺伝統計学教室の平田潤先生には、末梢血検体処理、マイクロ RNA シークエンスの施行に尽力いただいたことに感謝いたします。理化学研究所生命医科学研究センター統計解析研究チームの鎌谷洋一郎チームリーダー（現東京大学大学院新領域創成科学研究科複雑形質ゲノム解析分野教授）、秋山雅人研究員（現九州大学医学部眼科学教室講師）、石垣和慶研究員（現ハーバード大学研究員）には、日頃からゲノミクス研究や統計解析に関する多大なご指導をいただき、研究における基礎的な思考方法を教育していただきましたことを深く感謝申し上げます。東京女子医科大学膠原病リウマチ痛風センター山中寿前教授、猪狩勝則准教授、谷口敦夫教授には、関節リウマチの患者さんの新規ゲノムデータを快くご提供いただきましたことに感謝申し上げます。理化学研究所生命医科学研究センターFANTOM コンソーシアムの Michiel de Hoon 先生、Piero Carninci 先生、林崎良英先生、Alistair Forrest 先生、伊藤昌可先生、

Timo Lassmann 先生、川路英哉先生には、組織特異的・網羅的マイクロ RNA 発現データをご提供いただき、またデータ解析にあたっても多くのご助言をいただいたことに感謝申し上げます。京都大学の 大村浩一郎先生、伊藤宣先生、吉富啓之先生には関節リウマチの病態研究とマイクロ RNA 実験解析計画にご助言をいただいたことに感謝申し上げます。