

# 大規模ゲノムデータを用いた日本人における運動能力の遺伝的特徴の決定および 遺伝的運動能力予測システムの開発

中井 琢

東北大学大学院 医学系研究科 酸素医学分野

## 1. 緒言

ヒトの運動能力のおよそ70%は、遺伝的に決定されている<sup>1</sup>。さまざまなスポーツ関連遺伝子が同定され、スポーツパフォーマンスに影響を及ぼす1塩基多型 (single nucleotide polymorphism: SNP) が数多く見つかっている<sup>2</sup>。また、筋力や最大酸素摂取量などの運動能力は、多数のSNPによって規定される量的形質であり、運動能力の遺伝的特徴の決定にはゲノムワイドなSNP解析が必須である。近年、SNPに基づいたスポーツ指導が普及しつつあるが、量的形質であることは加味せずに限られた数のSNPを指標としている。本研究では、東北メディカルメガバンク機構のデータベースを用いて、ゲノムワイドに運動能力に関与するSNP群を同定したのち、機械学習による量的形質の予測モデルを構築することとした。加えて、ゲノムに基づいたスポーツ指導を行う上で重要となるSNPの新規同定も目指す。

東北メディカル・メガバンク機構 (ToMMo) は、2011年に発生した東日本大震災をきっかけに東北地方の地域医療の支援および未来型医療の創成を目的に設立された。ToMMoは、妊婦の方を中心とした三世代コホート調査および東北沿岸部を対象とした地域住民コホート調査の参加者約15万人のゲノムデータおよび生体試料を保有している大規模なバイオバンクである。保管されている試料および情報は疾患歴から生化学検査情報、さらには全血における網羅的遺伝子発現情報まで多岐に渡る。これにより、ToMMoのデータバンクを用いることで、SNPが関与する疾患や表現型の間接形質を解明することが可能である<sup>3</sup>。

ゲノムワイドに形質と関連のあるSNPを統計的に明らかにする一般的な手法として、ゲノムワイド関連解析 (Genome Wide Association Study: GWAS) がある<sup>4</sup>。GWASは、単一遺伝性疾患の原因SNPの同定には強力である一方で、多数のSNPが相乗的に表現型を決定する量的形質の解析においてはその精度は低い。量的形質に関与するSNPの相関スコアを算出して、スコア値の総和から表現型を予測するポリジェニックスコア (polygenic score: PGS)<sup>5</sup>や機械学習技術を応用することでこれまで判別不可能だったSNP保有パターンの多様性を視覚的に明らかにすることができる手法も開発されている<sup>6</sup>。これらの手法の問題点として、それぞれのSNPの機能を加味されておらず、あくまで統計学的に表現型を予測する手法となっている。本研究では、各SNPの機能的な分類に基づいて、表現型を予測するモデルの開発に挑む。

赤血球は、全身への酸素運搬を担う細胞であり、持久性運動能力と密接に関係している<sup>7</sup>。

そこで、赤血球に関する GWAS を行い、SNP の保有パターンからその表現型の予測を試みる。さらに、全血トランスクリプトームデータと SNP 情報を統合することで、予測精度に改善につながるか検証する。また、全血トランスクリプトームや生化学検査値からその中間形質を明らかにすることで、ゲノム情報に基づいたスポーツ指導において有益な情報を提供することを目指す。

## 2. 研究方法

### 2.1 被験者

ToMMo におけるすべての研究は、東北大学医学系研究科および ToMMo の倫理委員会の審査の承認を経て実施されている。本研究は、ToMMo の三世代コホート調査および地域住民コホート調査の参加者のうち、解析基準を満たしたゲノムデータ (38KJPN) および血液検査・生化学情報を用いて行われた。コホート調査の参加者からは、研究への情報提供を許諾するインフォームドコンセントを取得している<sup>8</sup>。

### 2.2 ジェノタイピング・クオリティコントロール (QC)

コホート参加者のゲノム情報は、ジャポニカアレイ v2 を用いてジェノタイピングを行なった<sup>9</sup>。ジャポニカアレイより得られたデータは、659,326 個のバリエーションを持つ 90,565 人であった。ジャポニカアレイ v2 から得られたゲノム情報から、IMPUTE2 (ver.2.32) を用いたインプテーションにより全ゲノム配列を推定した。検出された SNP の割合が 0.95、~~を示す DishQC が 0.82 以下、の検体を取り除いた。さらに、Hardy-Weinberg 平衡が  $1.00 \times 10^{-5}$  の SNP を取り除いた。その後、QCTOOL (<https://www.well.ox.ac.uk/gav/qctool>) を用いて、以下のオプション: `-omit-chromosome-compare-variants-by position, -flip-to-match-chort1` によりジェノタイプデータを統合した。QC 後のデータは、54,034,112 個のバリエーション、90,565 人となった。さらに、PLINK (v1.90b5.1) を用いて、以下のオプション: `-maf 0.05, -hwe 0.05, -geno 0.01, -indep-pairwise 1500 150 0.03 -genome` により、SNP の QC を行い、8,595,665 とした。

### 2.3 血液検査

被験者の末梢血は、EDTA-2Na 入りチューブ、ヘパリンチューブ、通常のチューブ、それぞれに収集したのち、全血球計算および-80°Cでの保管を行なった。血清鉄・フェリチン・総鉄結合能・不飽和鉄の測定は、東北大学病院検査部にて実施された。

## 2.4 ゲノムワイド関連解析 (Genome-wide association study: GWAS)

BOLT-LMM v2.3.2 を用いて、線形混合モデルにより常染色体上の各 SNP の赤血球数・ヘマトクリット値・ヘモグロビン値・平均赤血球容積 (MCV)・平均赤血球血色素量 (MCH) に対する効果量を導いた。5.0×10<sup>-8</sup>以下の p 値を満たす SNP を統計的に有意であるとした。連鎖不平衡関係にある SNP の中から p 値の最も小さな SNP を各領域における原因 SNP とした (表 1)。

## 2.5 全血トランスクリプトーム解析

ヘパリンチューブに回収した血液サンプルから、室温下で 1,700×g にて 20 分間遠心を行い、末梢血単核球を分離した。血小板および血漿の混入を防ぐため、2 mM EDTA 入りのリン酸緩衝生理食塩水 30 mL を用いて洗浄し、250×g での 10 分間の遠心を 3 度繰り返した。その後、Maxwell 16 LEV simply RNA Blood Kit にて RNA を抽出した。RNA の完全性を示す RNA integrity number (RIN)は、Agilent 2100 Bioanalyzer にて RNA 6000 Nano Kit を用いて解析した。トランスクリプトーム解析のための cDNA ライブラリは、TruSeq RNA Sample Prep Kit v2 (Illumina)を用いて、250 ng の RNA から精製した。Poly(A)+ RNA を選別したのち、両端にアダプターを付加したフラグメント長 306-355 bp の cDNA のライブラリを得た。シーケンス解析は、TruSeq PE Cluster Kit v3-cBot-HS、TruSeq SBS Kit v3-HS、HiSeq2500 を用いて行なった。シーケンスクオリティを FastQC で確認したのち、TopHat v2.0.9 を用いてヒトゲノムリファレンス hs37d5 にマッピングした。最後に、Cutdiff を用いて各遺伝子の発現量 (FPKM: fragments per kilobase of transcript per million fragments mapped read)を定量した<sup>10</sup>。

## 3. 結果

### GWAS による赤血球関連表現型に対する感受性遺伝子の同定

ToMMo コホート参加者 90565 人のゲノムデータを用いて、赤血球数・ヘマトクリット値・ヘモグロビン値・平均赤血球容積・平均赤血球血色素量に関する GWAS を行なった。その結果、すべての赤血球関連表現型において有意な多型を多数認めた (RBC: 4001 個, Hct: 527, Hgb: 474, MCV: 1948, MCH: 1135)。続いて、SNP の保有パターンからの表現型の予測を目的に、赤血球数に関連する多型データを用いたクラスタリング解析および主成分分析を行なった。GWAS 解析では、連鎖不平衡関係にある SNP の有意であると判定されてしまうため、有意な多型が密集する領域内で p 値の最も小さな代表的な 50 多型を用いて解析を行なった。しかし、明確な分類は認めず、各集団における赤血球数の差を認めなかった (データ非公開)。この結果から、SNP 保有パターンによる赤血球数の予測は困難であることが示唆された。

## GWAS と全血トランスクリプトームデータの統合による表現型予測精度の向上

遺伝子多型は、タンパク質の機能に影響を与えるものと転写および mRNA の安定性に関与するものに分けることができる。全血トランスクリプトームデータとの統合により、赤血球数と有意に関与する全 50 多型のうち、16 多型が近傍の遺伝子発現と関与していることが明らかとなった。また、いずれの多型もおおよそハーディ・ワインベルグ平衡が成立したが、rs [REDACTED] (SNPa) のみホモ接合子が存在しなかった (図 2)。それぞれの転写量と赤血球数は相関関係にあり (データ非公開)、これらの多型が遺伝子発現制御の変化を介して、赤血球数を規定していることがわかった。そこで、全血トランスクリプトームデータと GWAS データの統合により、赤血球数の予測精度の向上に試みた。先ほど同定した 16 多型が影響を及ぼしている遺伝子の転写量に基づいて、クラスタリング解析を行なったところ、多血を示す集団を認めた (データ非公開)。この集団における多血の程度は、16 のそれぞれの多型による赤血球数への影響よりも顕著に大きく、遺伝子発現が相乗的に多血の表現型を決定していることが明らかとなった。

## 鉄代謝への影響を及ぼす多型の同定

rs [REDACTED] (SNPb) は、フェリチン分解を介して鉄代謝を制御する遺伝子 A のエクソン上に存在するアミノ酸同義置換である。鉄は赤血球造血に必須な因子であることから、SNPb が鉄代謝の変化を介して赤血球数を決定付けていると仮説を立てた。そこで、血液検査における鉄関連の測定値を解析した。血清鉄 (FE)、総鉄結合能 (TIBC)、不飽和鉄結合能 (UIBC) においてはバリエーション間の差を認めなかったが、血清フェリチン値 (FER) が多型をホモに有する被験者において顕著に亢進していた (データ非公開)。血清フェリチン値は貯蔵鉄量を反映していることから、SNPb が遺伝子 A のフェリチン分解機能を抑制して、貯蔵鉄利用不全が生じていると考えられた。全血中の遺伝子 A の転写量は、CC ホモ接合に比べて、ヘテロ接合および TT ホモ接合において有意に高かった (データ非公開)。

## 4. 考察

赤血球数は、複数の遺伝子多型によって規定される量的形質である。SNP は一塩基置換によって生じる多型であり、多型が存在する箇所によってその機能は異なる。エクソン上に存在してアミノ酸置換を生じさせるものやイントロンや制御領域に存在してその遺伝子の転写量に変化を及ぼすもの (expression quantity trait loci: eQTL) がある。しかし、eQTL 解析のためには、健全なヒトから組織 RNA を取得する必要性があり、侵襲性の高い研究である。本研究では、低侵襲的に組織を取得できる全血においてトランスクリプトーム解析を行うことで、遺伝子発現変化を介して、赤血球数を決定づける 16 多型を同定した。ただ、本研究で解析できた

eQTL はあくまで血球細胞において機能的な多型に限られる。遺伝子の転写制御機構は組織および細胞ごとに異なるため、他にも遺伝子発現量に關与する多型が存在すると考えられる。実際、赤血球造血因子エリスロポエチンをコードする *EPO* 遺伝子のプロモーター領域にある rs1617640 が、*EPO* 遺伝子の転写を抑制して、赤血球造血抑制性にはたらくことがわかっている<sup>11</sup>。

一般的に、eQTL は遺伝子あたりの表現型への貢献は低く、相乗的に表現型を決定付けているとされている<sup>12</sup>。本研究においても、各多型の赤血球数の影響は小さかったものの、多血にはたらく多型を複数持つ集団は顕著に多血を示した。また、eQTL の表現型の影響度の低さから HW 平衡に従って集団内に存在すると考えられたが、SNPa はホモ接合子が存在せず、致命的な表現型を示すことが示唆された。これまでにまったく報告のない多型であるため、今後解析を進めていく必要がある。

SNPb は、フェリチン値との關与がみられ、機能的鉄欠乏により貧血を呈す多型であることが示唆された。SNPb による貧血の表現型は、日常生活を送る上ではまったく問題のないレベルであるが、持久性アスリートにとっては看過できない。アスリートの貧血治療では、一般的な戦略として鉄剤投与が行われる。しかし、SNPb を持つアスリートは、機能的鉄欠乏状態にあるため、体内には十分量の鉄が存在すると予測される。このようなアスリートへの鉄剤投与は、貧血改善効果に乏しく、組織への更なる鉄の沈着によりヘモクロマトーシスを引き起こす危険性もある。このように、ゲノム情報は、アスリートのパフォーマンス向上だけでなく、アスリートの適切な健康管理にも活用できると期待される。

## 6. 参考文献

1. Guth LM, et al. Genetic influence on athletic performance. *Curr Opin, Pediatr*, 2013;25(6):653-658
2. Wang G, et al. Genomics of elite sporting performance: what little we know and necessary advances. *Adv Genet*, 2013;84:123-149
3. Kikuya M, et al. Protocol and research perspectives of the ToMMo child health study after the 2011 great east Japan earthquake. *Tohoku J Ecp Med*, 2015;236(2):123-130
4. Tomfohrde J, et al. Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. *Science*, 1994;264(5162):1141-1145
5. Lewis CM, et al. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*, 2020;12(1):44
6. Sakaue S, et al. Dimensionality reduction reveals fine-scale structure in the Japanese population with consequences for polygenic risk prediction. *Nat Commun*, 2020;11(8):1569
7. MairBaur H. Red blood cells in sports: effects of exercise and training on oxygen supply by red blood cells. *Front Physiol*, 2013;4:332
8. Yamada M, et al. Genetic loci for lung function in Japanese adults with adjustment for exhaled nitric oxide levels as airway inflammation indicator. *Commun Biol*, 2021;4(8):1288
9. Tadaka S, et al. 3.5KJPNv2: an allele frequency panel of 3552 Japanese individuals including the X chromosome. *Hum Genome Var*, 2019;6:28
10. Ohmomo H, et al. Reduction of systematic bias in transcriptome data from human peripheral blood mononuclear cells for transportation and biobanking. *PLoS One*, 2014;9(8):e104283
11. Amanzada A, et al. Erythropoietin rs1617640 G allele associates with an attenuated rise of serum erythropoietin and a marked decline of hemoglobin in hepatitis C patients undergoing antiviral therapy. *BMC Infect Dis*, 2014;14:503
12. Peters JE, et al. Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. *PLoS Genet*, 2016;12(3):e1005908