

# ゲノム編集技術を駆使した遺伝的リスク解釈モデルの構築

Department of Genetics, Stanford University

Gladstone-UCSF Institute of Genomic Immunology

太田 峰人

## 【はじめに】

ゲノムワイド関連解析に代表される大規模ゲノム研究は、これまでさまざまな形質や疾患に関わる遺伝的多型を多数同定してきた。その結果、例えば身長には 12,000 以上の遺伝的多型が関わることが明らかになった<sup>1</sup>。これは、ほぼすべての発現遺伝子が何らかの形で身長に影響を与えていることを示唆する。

遺伝学的研究の目的の一つは、因果関係が明確な遺伝的多型と形質の関連解析から開始することで、さまざまな形質や疾患の発症原因を特定することである。しかし、これ程多くの遺伝子が発症に関わる場合、その知見を生物学的に解釈するのは単純ではない。

留学中の指導者である Jonathan Pritchard らは、このようなパラドックスを世界に先駆けて報告し、その解釈には遺伝子発現制御ネットワークを介した多くの遺伝子間の相互作用が重要であるという Omnigenic モデルを提唱した<sup>2</sup>。この理論モデルは遺伝的研究結果の解釈を容易にするが、これまで実データでの証明がなされていなかった。

Omnigenic モデルを実データに応用するには、二つの大きな課題があった。一つ目は、これまで大規模な遺伝子制御ネットワークの構築を可能にするデータセットが不足していたことである。もう一つは、大規模ゲノム解析では遺伝子多型のレベルでの関連が定量化されるが、遺伝子 (gene) レベルに結びつけることが容易でない点が挙げられる。

今回、一つ目の課題については、Perturb-seq 法を用いることを計画した。Perturb-seq は、CRISPR 技術による遺伝子編集を単細胞 RNA-seq 解析と組み合わせることで、すべての遺伝子が遺伝子発現ネットワークに与える影響を定量化する手法である。二つ目の課題については、Whole Exome Sequencing から得られる Loss of Function (LOF) 多型 burden test を用いることにした。これにより、GWAS では得られない遺伝子レベルでの形質との関連を定量化することができる。

本報告では遺伝的リスク解釈モデル構築について報告する。

## 【方法】

これまで白血病由来の K562 株において、全遺伝子を対象とした大規模な Perturb-seq データが公開されている<sup>3</sup>。本データを用いて、各遺伝子-遺伝子間の制御ネットワークを推定した。さらに、non-negative matrix factorization 法を適用する事で単一細胞遺伝子発現マトリックスから遺伝子発現プログラムを同定し、プログラムレベルの制御についても定量化した。

一方、遺伝子関連解析については、UK biobank whole exome sequence データから、LOF 多型の情報を用いて遺伝子レベルの形質との関連を定量化した。LOF 多型は稀であるが、同一遺伝子における異なる LOF 多型が形質に与える効果量は同等と仮定できる。そのため、全ての LOF 多型を統合して遺伝子レベルで形質との関連解析を行う burden test を行った。burden test を用いる事で同一遺伝子上の LOF 多型の影響を統合して解析する事ができる。一方、LOF burden test は遺伝子サイズの影響を強く受けノイズも多いことがわかっている。今回我々は遺伝子発現などの事前情報を加味し、ベイズ推定法を用いる事で LOF burden test の推定効果量を改善する事に成功し、その情報を用いる事とした。

これらの異なる層の情報を、遺伝子を軸にして統合解析する事で形質の遺伝的差異に関わる遺伝子制御ネットワークを同定する事を試みた (図 1)。K562 株は特に赤血球の形態や量に関わる形質と強く関連する事を heritability enrichment 解析で確認した事から、これらの形質をモデルとして解析を進めた。

## 【結果及び考察】

今回特に K562 細胞株に遺伝的リスクの集積が強かった MCH (mean corpuscular hemoglobin, 赤血

球あたりのヘモグロビン量)に影響を及ぼす遺伝的パスウェイについての解析結果について報告する。

まず初めに LOF burden test で得られた MCH に関わる遺伝子の効果量を、Perturb-seq から得られた遺伝子の遺伝子発現に対する効果量と比較した。興味深い事に、MCH の遺伝的リスクは、HBA1 遺伝子の発現量に対する制御効果と非常に強く相関する事が分かった ( $P = 3 \times 10^{-7}$ 、図 2)。つまり、例えば HBA1 を正に制御する遺伝子は MCH と正に関連し、負に制御する遺伝子は MCH と負に関連するといった傾向が見られた。また、HBA1 遺伝子自体も MCH と最も強く関連する遺伝子の一つとして同定された。HBA1 タンパクはヘモグロビンの構成要素の一つであり、MCH の値に直接関わる事が生物学的にも予想される。

これらの結果は、omnigenic model で提唱した遺伝子発現制御ネットワークと遺伝的リスクの関わりを実データで証明した初めての例と言える。さらに、Perturb-seq に内包された遺伝子発現制御ネットワークと LOF burden test で同定された遺伝子レベルの関連統計量を比較解析する事で疾患・形質に関わる“コア・パスウェイ”を同定する事が可能である事を示唆する。

そこで我々はさらに HBA1 以外の MCH に関わるコア・パスウェイを同定するため、non-negative matrix factorization 法を用いて遺伝子発現マトリックスをパスウェイレベルに分解した。これにより、60 の細胞内パスウェイと、それを構成する遺伝子を同定し、さらに各パスウェイに対して遺伝子をもつ制御効果量を推定する事が可能となる。

興味深いことに、MCH については細胞周期パスウェイとヘモグロビン生成パスウェイにおいて、パスウェイを構成する遺伝子及びパスウェイを制御する遺伝子が共に強く遺伝的関連に集積していた。さらにオートファジーパスウェイを制御する遺伝子も強い集積が見られた。これらは、MCH という形質を決定する上で遺伝的に重要なコア・パスウェイと呼ぶ事ができる。

これらのコア・パスウェイを同定した事で、遺伝学的解析で同定された“形質関連遺伝子”がパスウェイ及びその制御ネットワークの中でどこに位置し、どのように形質と関わるかという地図を作成する事が出来る(図 3)。この中で、例えば *SUPT5H* 遺伝子は形質に関わる全てのパスウェイを正または負に制御し、結果全てのパスウェイ制御を介して、*SUPT5H* 遺伝子の発現増加は MCH の上昇につながる事が分かる。これまで GWAS など大規模ゲノム解析では”疾患関連遺伝子”の候補を多数同定して来たが、それらがどの様に相互に関わりながら、形質に関わるかを同定する手段は確立されていなかった。今回の我々のアプローチを用いる事で、疾患関連遺伝子がどの様にパスウェイやその制御に関わり、さらに形質に関わっていくかを同定する事が可能である事を示した。

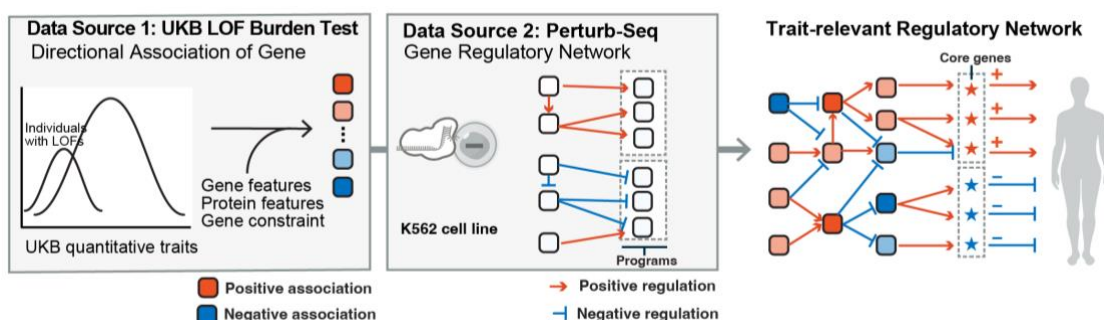


図 1. 研究の概要. ゲノムデータと Perturb-seq データを組み合わせる事で形質に関わる遺伝子制御ネットワークを同定する事を目標とした。

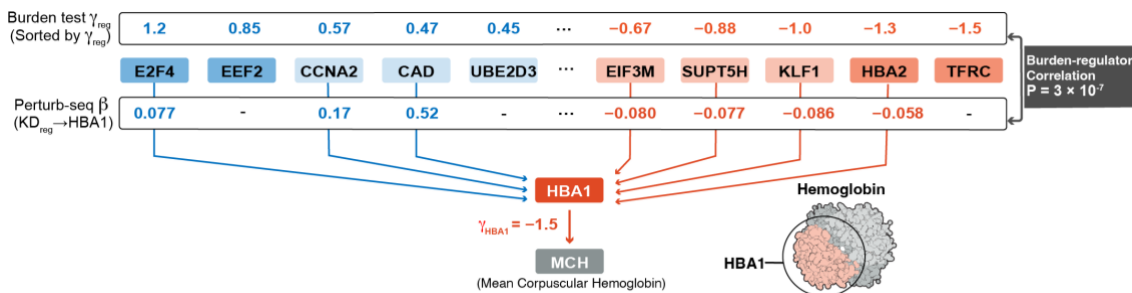


図 2. HBA1 制御遺伝子の MCH との関わり。多くの遺伝子が HBA1 の遺伝子発現制御を介して MCH という形質に関わる事を示した。

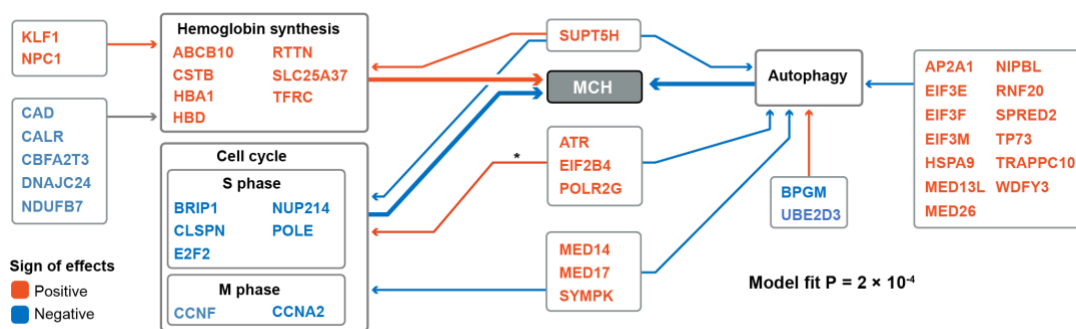


図 3. MCH に関連する遺伝子発現制御ネットワーク。各遺伝子の色は MCH との関連の方向性を示す。それぞれの遺伝子がどのように形質に関わるかが解釈可能となる。

### 【おわりに】

赤血球関連形質をモデルとして、大規模遺伝子編集技術と遺伝子関連解析を組み合わせる事で形質に関わる遺伝子制御ネットワークを同定する事が可能である事を示した。このアプローチは同様のデータを様々な細胞種において作成していく事で広く応用が可能である。現在様々な他の細胞種において大規模な Perturb-seq データを作成するプロジェクトが大規模コンソーシアムなどにおいて進行中である。我々のアプローチをこの様なデータと、それが関わる疾患、特に自己免疫疾患に応用していく事で、疾患に関わる遺伝子発現ネットワークを同定していく事を今後の目標としている。

### 【謝辞】

本留学をサポートしていただいたアステラス病態代謝研究会に深謝致します。

### 【引用文献】

1. <https://doi.org/10.1038/s41586-022-05275-y>, 2022.
2. <https://doi.org/10.1016/j.cell.2017.05.038>, 2017.
3. <https://doi.org/10.1016/j.cell.2022.05.013>, 2022.