

令和 2 年 7 月 6 日現在

機関番号：82610

研究種目：基盤研究(C) (一般)

研究期間：2016～2019

課題番号：16K07218

研究課題名(和文) 疾患関連SNPが引き起こす細胞のエピジェネティック病態—統計モデリングと実験検証

研究課題名(英文) Epigenetics of disease associated SNPs

研究代表者

竹内 史比古 (Takeuchi, Fumihiko)

国立研究開発法人国立国際医療研究センター・その他部局等・室長

研究者番号：50384152

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：ゲノムワイド関連解析により、様々な疾患と関連する多数の一塩基多型(SNP)が同定されたが、疾患機序の解明には殆どつながっていない。疾患機序を解明するためには、疾患関連SNPが、どの組織・細胞種において細胞状態に影響するかを明らかにする必要がある。本研究では、疾患関連SNPが、ある組織・細胞種においてエピジェネティックな状態にどのように影響するかを予測する統計モデルを構築し、実験検証した。ヒトにおけるT細胞・脳・胎盤の組織差を定量し、エピジェネティックな状態と疾患発症の因果関係モデルを構築し、細胞種ごとのエピジェネティック状態に対してSNPが与える影響を予測する統計手法を開発した。

研究成果の学術的意義や社会的意義

個人のDNAの違いにより、病気の罹りやすさが変わることが分かってきている。これはゲノムワイド関連解析という研究手法で解明されてきた。「DNAの違い 細胞の変化 病気」の両端の関係は分かっていたが、実は真ん中の「細胞の変化」が未解明である。本研究では、細胞の変化を推測するための統計解析法omicwasを開発した。ある組織・細胞種でどのような細胞の変化が起きているかが重要であり、これはomicwasにより推測できる。推測結果は、次段階の生物学実験をデザインするための指標になる。

研究成果の概要(英文)：Genome-wide association studies have identified many SNPs associated with various diseases, yet the disease mechanism has been elucidated for a small number of SNPs. In order to elucidate the disease mechanism, we need to understand in which tissue or cell type the SNPs function. Here, I developed a statistical model that predicts the effect of a SNP in a specific tissue or cell type, and performed experimental verification. I assessed tissue heterogeneity among human T cells, brain and placenta. I developed a causal model for epigenetic status and disease. I also developed a statistical method to predict the cell-type-specific epigenetic effect of a SNP.

研究分野：遺伝統計学

キーワード：エピジェネティクス 細胞種 エピゲノムワイド関連解析 遺伝子発現解析 QTL解析 遺伝統計学

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) ゲノムワイド関連解析 (GWAS) により、様々な疾患と関連する多数の一塩基多型 (SNP) が同定されたが、疾患機序の解明には殆どつながっていない。GWAS では、罹患者と健常者で遺伝子型頻度の異なる SNP を探索する。多くの場合、特定の染色体領域内の、連鎖不平衡を示す複数の SNP が疾患と関連し、どれが原因 SNP かは区別できない。関連 SNP のリストをなるべく絞り込むために、大人数・多人種の GWAS を組み合わせる試みが行われてきた。その結果、疾患関連 SNP の多くは、タンパクコード領域の外に位置することが分かり、遺伝子発現制御を通じた因果的関与が示唆された。

GWAS の結果を考察すると、多くの疾患関連 SNP において、遺伝子型依存的に、特定組織で、DNA メチル化などのエピジェネティック変化が起きて、疾患が発症すると推測される。例えば、糖尿病関連 SNP のリスクタイプの遺伝子型では、脂肪組織がインスリン存在下でも血中の糖を吸収せず、糖尿病を引き起こす可能性がある。このように疾患機序を解明するためには、疾患関連 SNP が、どの組織において細胞状態に影響するかを明らかにする必要がある。この問題に関して、研究開始当初 (2016 年 4 月) 以下の二つの方向性の研究が行われていたが、どちらも命題に直接答えてはいなかった。

(2) 疾患関連 SNP の解釈のために、これまで遺伝子発現の量的形質座位 (eQTL) と DNA メチル化の量的形質座位 (mQTL) 解析が行われてきた。これらの解析は、典型的には、100 人程度のリンパ芽球様細胞株で mRNA 発現ないしは DNA メチル化レベルを測定し、どの遺伝子の発現・メチル化が SNP と関連するかを探索する。そのようにして遺伝子が挙げられても、リンパ芽球様細胞株と当該疾患の病態に關する組織では、細胞種が異なるために、その解釈は限定的にならざるを得なかった。

こうした状況のなか、様々な細胞種の解析が、エピゲノム・ロードマップや国際エピゲノムコンソーシアムで進められている。幅広い細胞種の解析が行われつつあるが、細胞種あたりの解析自体は数人分に限られる。従って、SNP の細胞状態への影響は十分調査できない。

(3) 研究期間中の 2017 年 10 月に、米国の GTEx プロジェクトにより様々な臓器の合計 44 組織について eQTL の結果が公表された。これにより、上記 (2) の前半は解決された。

(4) エピゲノムワイド関連解析や遺伝子発現解析では病変組織と正常組織が比較されるが、組織は複数の細胞種から成り、細胞種ごとに比較できれば理想的である。2017 年から組織検体を構成する単一細胞ごとの RNA 網羅解読 (scRNA-seq) が可能になり、組織を構成する様々な細胞種を直接測定できるようになった。scRNA-seq は一検体につき約 100 万円とまだ高価なため、多数の検体の解析はできない。

### 2. 研究の目的

疾患関連 SNP が、どの組織・どの細胞種において細胞状態に影響するかを明らかにする。疾患関連 SNP が各種の細胞のエピジェネティックな状態に与える影響を予測する統計モデルを構築し、実験検証する。

### 3. 研究の方法

#### (1) 組織差の定量

単独 SNP による各種組織のエピジェネティック状態変化を予測するモデルを構築し、分散分析により評価した。

#### (2) エピジェネティックな状態と疾患発症の因果関係

因果関係のモデル化については、Pearl らの Structural Causal Model を基本とし、Mendelian Randomization も用いた。メタボリックシンドロームの各種疾患を対象とした。また、CpG メチル化に影響しうる環境として、喫煙も対象とした。これらの疾患および環境について、Epigenome-wide association study の文献を調査した。

#### (3) 細胞種ごとのエピジェネティック変動

細胞種ごとのエピジェネティック状態に対して、疾患ないしは疾患関連 SNP が与える影響を予測する統計手法を開発した。組織検体の遺伝子発現・DNA メチル化データから、細胞種ごとの発現変動・メチル化変動を推定する非線形リッジ回帰による R 言語パッケージ omicwas を開発した [文献]。また実際にラット肝臓でシングルセル ATAC-seq 実験を行なった。

### 4. 研究成果

#### (1) 組織差の定量

ヒトにおける T 細胞、脳、胎盤の CpG メチル化率の分散については、71% が組織差、2% が人種差、0~15% が近傍の SNP (mQTL) で説明できた。従って、仮に単一の組織のみを解析できた場合には、SNP で説明できる割合が 0~52% に上昇するはずである。この結果は、CpG メチル化率と SNP の研究において、不均一な検体が悪影響を及ぼすことを示している。

### (2) エピジェネティックな状態と疾患発症の因果関係

CpG との関連については、確度の高い結果が得られているが、因果関係については、研究方法も含めてまだ不確定なところがあった。例えば、糖尿病と関連する CpG について、1) CpG メチル化が糖尿病を引き起こした、2) 糖尿病の影響により CpG メチル化が変化した、3) 双方向に影響、4) 相互に因果関係はないのに別の交絡因子により相関が起きている、の 4 通りが考えられる。Mendelian Randomization および大規模 GWAS からは、2 と 4 が多いことが示唆された。末梢血で測定したエピジェネティック状態が個別の疾患を正確に反映していないこと、また喫煙が交絡していることを示す結果が得られた[文献 ]。

### (3) 細胞種ごとのエピジェネティック変動

実データとシミュレーションで検証したところ、どの細胞種においても最適な手法との違いが小さく、従来法よりも頑健だった。セルソートした細胞種で実際に行った実験には及ばないものの、組織の実験データからもある程度は細胞種ごとの遺伝子発現変動・DNA メチル化変動を推定することができた[文献 ]。

#### < 引用文献 >

<https://github.com/fumi-github/omicwas>

竹内 史比古「生活習慣病 GWAS と DNA メチル化」実験医学(0288-5514)38 巻 4 号 Page560-565(2020.03)

Fumihiko Takeuchi, Norihiro Kato, Nonlinear ridge regression improves robustness of cell-type-specific differential expression studies, bioRxiv 158758.

<https://doi.org/10.1101/2020.06.18.158758>

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 0件 / うち国際共著 1件 / うちオープンアクセス 1件）

1. 著者名 Fumihiko Takeuchi, Norihiro Kato	4. 巻 158758
2. 論文標題 Nonlinear ridge regression improves robustness of cell-type-specific differential expression studies	5. 発行年 2020年
3. 雑誌名 bioRxiv	6. 最初と最後の頁 1
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/2020.06.18.158758	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 3件）

1. 発表者名 Fumihiko Takeuchi
2. 発表標題 Finer analysis of the correlation between SNP, CpG methylation and gene expression
3. 学会等名 アメリカ人類遺伝学会（国際学会）
4. 発表年 2016年

1. 発表者名 Fumihiko Takeuchi
2. 発表標題 New software for cell-type-specific EWAS & DE analysis; evaluation in real data
3. 学会等名 CSHL Single cell analyses（国際学会）
4. 発表年 2019年

1. 発表者名 Fumihiko Takeuchi
2. 発表標題 New software for cell-type-specific EWAS & DE analysis; evaluation in real data
3. 学会等名 International Human Epigenome Consortium（国際学会）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<https://github.com/fumi-github/omicwas>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----