

2022年度 修士論文

テンソル分解による妊娠時カドミウム曝露が代謝に与える影響の推定

Estimation of Metabolic Effects upon
Cadmium Exposure during Pregnancy
Using Tensor Decomposition

2023年2月21日

指導教員 田口善弘 教授

中央大学大学院 理工学研究科 物理学専攻
所属

21N2100004C

天藏悠基^{*1}

^{*1} a17.rtps@g.chuo-u.ac.jp

要旨

母体妊娠中のカドミウム曝露が、生まれてくる子のインスリン代謝を破壊するという研究報告がある。この研究報告に基づき行われた実験データに対してテンソル分解という手法を用いて実際に影響を及ぼしている遺伝子を推定することと、それらの遺伝子によって考えられるインスリン代謝破壊のメカニズムを推測することを目的として本研究を行った。テンソル分解のうち、タッカー分解という方法を使うことにより、上記の生体现象が引き起こされる際に大きな発現量を持つ遺伝子を特定した。これにより、カドミウムの影響がたんぱく質の折り畳み阻害を引き起こし、小胞体ストレスを増幅し、小胞体関連分解を誘導していることが示唆された。また、小胞体関連分解の影響が何らかの機序で TNF- α の発現量の増大、または効果促進を誘導している可能性があることが明らかになった。TNF- α はスフィンゴ脂質の過剰発現を誘導することが知られている。本研究においても、スフィンゴ脂質の過剰発現を示唆する遺伝子群が特定された。スフィンゴ脂質の産生に関わる遺伝子の高発現に基づくセラミドの蓄積は、活性酸素の過剰産生を介した細胞内酸化ストレスを誘発する可能性があることも明らかになった。また、電子伝達系に関わる遺伝子群の高発現が酸化ストレスに寄与している可能性があることがわかった。これらの知見から細胞内酸化ストレスがインスリン受容体基質 1 のリン酸化を促進するメカニズムを明らかにした。インスリンシグナル伝達経路で重要な役割を担っているインスリン受容体基質 1 のリン酸化が、TNF- α 誘導のセラミド蓄積を介した細胞内酸化ストレスによって引き起こされるメカニズムを明らかにした。^{*2}

キーワード：機械学習、糖尿病

^{*2} この修士論文の内容は以下の原著論文で刊行済みです。

Amakura, Yuki, and Y-h. Taguchi. 2022. "Estimation of Metabolic Effects upon Cadmium Exposure during Pregnancy Using Tensor Decomposition" *Genes* 13, no. 10: 1698. <https://doi.org/10.3390/genes13101698>

目次

第 1 章	序論	2
1.1	研究背景	2
1.2	研究目的	5
第 2 章	用いたデータセットと解析方法	6
2.1	解析に用いたデータセットの概要	6
2.2	モデルの選択	6
2.3	データ処理	8
2.4	解析に用いるテンソルの構造	8
2.5	解析方法	9
第 3 章	結果	12
3.1	発現量が変動している遺伝子の同定	12
第 4 章	考察	17
4.1	同定された遺伝子の機能	17
4.2	同定した遺伝子によるインスリン代謝阻害機構の推定	18
第 5 章	結論	20
	謝辞	21
	付録	22
	参考文献	24

第1章

序論

1.1 研究背景

1.1.1 カドミウムの毒性

カドミウム (Cd) は、細胞内に存在する非必須の微量金属である [1]。しかし、食物からの摂取や暴露により、過剰な Cd が体内に入ると、様々な悪影響を及ぼす。Cd は鉱物中、土壌中など天然に存在する重金属である。日本においては、全国各地で一千年以上前から鉱山開発などにより地中から掘り起こされてきた。そのため、現在においても土壌や水田などからは Cd が検出される [2]。こういった環境中に排出された Cd が農作物に吸収されると、そのような農作物を長年摂取することにより、体内に徐々に蓄積される。自然環境中の Cd は、主に腎臓に蓄積する。このため、Cd 濃度の高い食品を長年にわたり摂取すると、腎機能障害を引き起こす可能性がある [2]。

一方で、Cd 中毒の一例として有名なイタイイタイ病は上述の場合に比べて、より高濃度な Cd を摂取した場合に引き起こされる [2]。富山県の神通川流域で起こった四大公害病の一つであるイタイイタイ病 [3] の原因は、鉱山から排出された Cd が一帯の流域や土壌を汚染したことであると言われている。腎障害によって、骨を形成する栄養素が体外に流出することで、骨が非常にもろくなる。これにより、軽微な衝撃であっても簡単に骨折するようになってしまい、全身を激しい痛みが襲う [4]。

このように、Cd は一定の摂取量を超えると、人体に悪影響を及ぼす。しかし、これらの例は人体への影響に関する報告の一部に過ぎず、他にも多くの報告がなされている。

1.1.2 細胞内代謝における Cd の悪影響

過剰な Cd の蓄積が細胞内代謝に与える主な悪影響は、たんぱく質のフォールディングの阻害である [5]。親電子性金属である Cd は、細胞のたんぱく質分子の構造に電氣的な影響を及ぼす [1]。たんぱく質がその機能を十分に発揮するためには、三次元構造が適切に維持されている必要がある。

細胞内には、核をはじめ、エネルギー産生のものであるミトコンドリアや、脂肪酸代謝のものであるペルオキシソーム、たんぱく質合成のものであるリボソーム、たんぱく質の修飾が行われるゴルジ体など、様々な細胞小器官（オルガネラ）が存在する。その中でも、たんぱく質の立体構造の形成、維持の役割を担うのが小胞体である [6]。リボソームなどで合成されたたんぱく質の折り畳みや切断の

他、糖鎖の付加、ジスルフィド結合の形成によって立体構造に安定性を与える [7]。

Cd の影響によって、折り畳みが阻害された異常たんぱく質（アンフォールディングたんぱく質）は小胞体に蓄積される。アンフォールディングたんぱく質の蓄積は小胞体ストレスを引き起こすことが知られている [8]。小胞体ストレスが比較的軽微な場合、小胞体内でたんぱく質の折り畳みの機能を実際に担っている分子シャペロンと呼ばれるたんぱく質群が、立体構造を正常なものに戻す [8]。一方で、ストレスが過剰になると、小胞体関連分解（ERAD : Endoplasmic Reticulum-Associated Degradation）を誘導する。ERAD の誘導に応答してユビキチン・プロテアソーム系が活性化され、アンフォールディングたんぱく質が分解される。蓄積されたアンフォールディングたんぱく質がユビキチン化を受け、プロテアソーム分解のターゲットとなることで、小胞体ストレスは緩和される。しかし、ストレスレベルが ERAD のストレス回避能力を超えると、細胞はアポトーシスを誘導する。また、ERAD は ATP 依存的に反応を進めることが知られている [8]。

1.1.3 インスリン代謝経路

インスリンは膵臓から分泌されるホルモンの一種である。膵臓のランゲルハンス島中の β 細胞から分泌される。生体内の血糖値を下げる働きを持つ唯一のホルモンであり、人間の生存のためには必要不可欠である [9]。

生体内の血糖値の上昇に反応して、膵臓からインスリンが分泌されることで、血液中のブドウ糖（グルコース）を細胞内に取り込み、細胞内でブドウ糖をエネルギーに変換する。これにより血糖値が抑えられる [9]。

インスリンが細胞膜の受容体に結合し、最終的にグルコースが細胞内に取り込まれる過程をインスリンシグナルという。インスリンシグナルでは、上記の通り、まず初めにインスリンが血液中を移動し、細胞膜上のインスリン受容体（IR : insulin receptor）に結合する。この結合によって、シグナル伝達を開始される。IR は細胞膜を貫通する形で局在しており、インスリンが膜の外側のユニットと結合することで膜内側のチロシンキナーゼが活性化され、IR は自己リン酸化される [10]。

次に、リン酸化された IR にインスリン受容体基質 1（IRS1 : insulin receptor substrate 1）が結合する。同様に IRS1 もリン酸化され、これが PI3 キナーゼ（phosphoinositide 3-kinase）を活性化する。活性化された PI3 キナーゼは PKB（protein kinase B）を細胞膜へ移動させ、活性化する。（PKB は別名 AKT としても知られ、アポトーシスを誘導する経路でも役割を担う）。PKB は細胞内へのグルコースの取り込み機構において決定的な役割を果たす GLUT4（glucose transporter type 4）を細胞膜へ移動させる [10]。GLUT4 はグルコースの細胞内濃度と血中濃度の間にある濃度勾配に反応し、血中から細胞内へとグルコースを取り込む。その後グルコースはリン酸化され、解糖系に入るか、グリコーゲンへと多量体化される [11]。このような過程で、インスリンの働きにより、血中グルコース濃度が抑えられる。

1.1.4 インスリン抵抗性と糖尿病

インスリン抵抗性とは、血糖値を正常範囲内に戻すために必要なインスリンの量が過剰な状態を意味する。インスリン抵抗性を引き起こす主な原因として、内臓脂肪型肥満が挙げられる [12]。こ

れはアディポカインというたんぱく質郡が脂肪細胞から分泌されることに起因する。特に TNF- α (Tumor Necrosis Factor- α) は大型の脂肪細胞から分泌され、インスリン抵抗性を引き起こすことが知られている [13]。一方で、アディポカインの中にはインスリンの感受性を促進するものも存在し、例えば、アディポネクチンと呼ばれるものは小型化した脂肪細胞から分泌され、インスリン感受性を亢進する [13]。

インスリン抵抗性は糖尿病の原因であり、2021 年の世界の糖尿病人口は 5 億 3,700 万人であるといわれており、これは全体の 10.5% に相当する [14]。日本においては、糖尿病罹患者と糖尿病予備軍の人口を合わせると 2000 万人に上るといわれている [15]。さらに、糖尿病には様々な合併症が存在し、それらの治療効果を鈍らせることもある。インスリン抵抗性の改善には生活習慣の改善が不可欠であり、特に上述の通り、肥満は TNF- α などのインスリン感受性に対して、悪影響を及ぼす。

医療費という観点で見ると、2017 年時点で日本の国民医療費は 43 兆 710 億円であり、このうちの 1 兆 2,239 億円は糖尿病によるものである (平成 29 年度 国民医療費の概況 (厚生労働省))。

このように、糖尿病は健康面、経済面の双方の点から一種の社会課題となっている。このため現在も世界中で糖尿病の予防、治療について、様々な角度から研究が行われている。

1.1.5 カドミウム毒性とインスリン抵抗性

Cd 毒性には酸化ストレスが大きく関与していることを示唆する報告がある。Shaikh らは、Cd 慢性投与後のマウスに抗酸化物質を投与することにより、Cd 毒性が軽減されることを報告している。このことは、Cd 毒性は酸化ストレスを介して様々な生物学的副作用を引き起こすことを示唆している [16]。

また、いくつかの研究で、Cd 毒性とインスリン代謝異常との関連が指摘されている。例えば、高重らは、高用量の Cd を 2 週間暴露したマウスは、脂肪細胞の分化に異常をきたす可能性があることを報告している。高重らは、脂肪細胞の分化、拡大、および機能に異常が生じ、インスリン抵抗性、高血圧、および心血管疾患を引き起こす可能性があることを指摘している [17]。

さらに、Cd やその他の有害金属汚染物質への出生前曝露は、生まれてくる子供の心血管疾患や 2 型糖尿病を含む肥満関連疾患のリスクを高めることを示唆する有力な証拠がある [18]。Al-Saleh らはまた、出生前の Cd 曝露が出生時体重に有意な影響を与えることを報告している [19]。

1.1.6 引用データ発行元の実験者等による報告

本研究で用いたデータは、上述した Cd の生体への影響に対する報告に基づいて、マウスで検証し、その詳細なメカニズムを調べるために行った実験 [20] から得られたものである。今回使用したデータに関して、実験を行った Jackson らのトランスクリプトーム解析の結果は、Cd 曝露が出生時体重に影響を与えることを示唆している。また、解析の結果から、Cd 曝露により ERAD および細胞内酸化ストレスが誘起され、ミトコンドリア機能障害が発生することが示唆されている。さらに Cd 曝露に伴い、脂質異常を示す遺伝子の発現の変化が観察されている。

Jackson らの解析では、発現変動遺伝子 (DEG : differentially expressed genes) を同定している。具体的には、母親が Cd に曝露された個体と、母親が Cd に曝露されていない個体の肝臓トラン

スクリプトーム解析データを、同じ時期、同じ性別で比較し、その遺伝子間のシグナル値の比率を比較して、発現レベルに有意差があるかどうかを判断する方法である。この方法によって、生後 42 日目 (PND42 : postnatal day 42) 雌個体のデータについて 5789 個の DEG が同定されており、これは解析に用いた遺伝子発現データ中の遺伝子数の 1/3 以上を占めている [20]。

このように、各時点で解析を行い、結果を比較しているが、全ての時点のデータを統合した解析や、PND の変化に対応して変化する遺伝子を同定することは試みられていない。また、今回独自に行った、全時点のデータを説明変数とした線形回帰法 (2.2.2 節参照) では、5865 個の遺伝子が過剰発現していることが判明した。しかし、メカニズムを理解するためには、決定的な役割を果たす遺伝子を特定することが望ましいため、全遺伝子の 3 分の 1 以上が過剰発現しているという結果は、本解析の目的に対して不適切である。

1.2 研究目的

本解析では、Jackson らの肝臓トランスクリプトーム解析データに対してテンソル分解を行い、PND に対して、ある程度の時間依存性を持つ遺伝子を特定し、全体の遺伝子数に対してより少ない数の過剰発現遺伝子を特定し、どの遺伝子が重要な役割を担っているかを推定する。また、同定した過剰発現している遺伝子のうち、どのような遺伝子がインスリン抵抗性獲得に決定的な役割を果たしているのか、そのメカニズムとして考えられる経路はどういったものなのかを明らかにすることを目的として本研究を行った。さらに、ERAD、細胞内酸化ストレス、ミトコンドリア機能不全が相互作用するメカニズムを解明し、これらの細胞内代謝経路がインスリン抵抗性の獲得にどう影響するのかを明らかにすることを目指した。

第2章

用いたデータセットと解析方法

2.1 解析に用いたデータセットの概要

“GSE150679”^{*1}は、米国 National Center for Biotechnology Information (NCBI) が提供するデータベースである、Gene Expression Omnibus (GEO) からダウンロードしたものである。

このデータセットは、Cd 曝露された雌の親マウス（妊娠 2 週間前から出生まで飲料水中の 500ppb CdCl₂ に曝露）と Cd 曝露されていない雌から生まれた子をそれぞれ PND1、PND21、PND42 で安楽死させ、肝臓トランスクリプトーム解析を行うことによって得られたものである。Cd を曝露された雌から産まれた子は、親マウス同様、PND10 まで Cd に曝露された。このトランスクリプトーム解析データは、Illumina (Illumina, Inc. San Diego, CA 92122 USA) HiSeq2500 シークエンサーで塩基配列を決定したものである。

親マウスが Cd に曝露されたかどうか、および肝臓トランスクリプトーム解析が行われたタイミング (PND) に関する各条件の区別について、生物学的複製 (バイオロジカルレプリケート) が存在している。また、このデータにおける各個体の遺伝子発現量は正規化されている。上述のデータセットは PND1、PND21、PND42 の各時点でのデータを含む 3 つの CSV ファイルから構成されている。

2.2 モデルの選択

2.2.1 テンソル分解を用いる意義

今回の解析では、テンソル分解のうち、タッカー分解を用い、アルゴリズムは高次特異値分解 (HOSVD : higher order singular value decomposition) [21] を採用した。この方法は「Large p Small n 問題」に対して有効である [21]。本解析のように、遺伝子数に対して検体数が非常に少なく、かつ、複数の遺伝子がある程度解釈可能な変数として選択する場合に有効である。生物実験では、倫理的な観点や被検体の入手の困難さから、一般にこのような少数の検体を利用することが多い。

前述の通り、Jackson らは PND ごとに DEG を同定し、それぞれ、PND1、PND21、PND42 を比較している。一方、本解析ではテンソルに整形したデータに対して、Tucker 分解を適用することで、PND の経過に伴って発現が変化する遺伝子を同定することができる。このことは、モデル選択

^{*1} <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150679>

の理由の一つである。

また、次節で示すように、本解析で用いるデータの構造を考慮し解析手法として利用できるものの中でも、最も単純なモデルである重回帰モデルは、このような「Large p Small n 問題」を分析する際には性能が低い。テンソル分解も比較的単純なモデルであるが、生物学的データは一般的に複雑な構造をしているため、テンソル分解のような単純なモデルを選択することが重要である。

2.2.2 重回帰モデルを適用した際の結果

この節では、重回帰モデルを用いた結果を示す。説明変数は、長さ 18 のベクトル、 t_l と s_l であり、各個体はここでの添え字 l で指定される。 t_l は PND の説明変数、 s_l は母体の Cd 曝露の有無を 1 (Cd 曝露あり) と 0 (Cd 曝露なし) で表す説明変数である。例えば、 $l = 1$ で指定される個体は $t_l = 1$ 、 $s_l = 1$ であるから、Cd 曝露された母体から生まれていて、且つ、PND1 のタイミングで安楽死させたサンプルを表す。

$$t_l = \begin{pmatrix} 1 \\ 21 \\ 42 \\ \vdots \\ 1 \\ 21 \\ 42 \end{pmatrix} \quad s_l = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

今回の解析対象になっている遺伝子の総数は 12,795 個であり、添え字 i は遺伝子を表している。また、 $v_{i'l}$ の要素は、 $l = 1, \dots, 9$ の Cd 曝露時の遺伝子 i' の発現量であり、それ以降の要素 ($l = 10, \dots, 18$) は Cd 曝露なしの発現量である。また $l = 1, l = 2, l = 3$ はそれぞれ、PND = 1, PND = 21, PND = 42 の発現量である。それ以降の要素にはバイオロジカルレプリケートを考慮したサンプルの発現量が格納されている。これらの変数を用いて、各遺伝子 i について、以下の重回帰分析を 12,795 回繰り返した。

$$v_{i'l} = a_{i'} + b_{i'}t_l + c_{i'}s_l$$

次に、重回帰分析で得られた係数 b_i と c_i のそれぞれについて、 t 検定を行い、 t 分布により p 値を付与した。そして、両係数で $p < 0.05$ が得られた遺伝子 i を特定した。この方法で得られた i の数は 5865 個であった。しかし、この方法では、全遺伝子の約 46% が DEG であるという結果になり、決定的な役割を担う遺伝子候補を絞り込むことができないため、不適切な方法であると考えられる。

2.3 データ処理

上述のデータをテンソル分解が適用できるようにするため、テンソルに整形した。まず、性差については、条件区別されたバイオロジカルレプリケートのサンプル数が不十分であったため、この解析から除外した（これは、PND42 ファイルのみが性別を条件付きで区別しており、PND1 および PND21 データではこの区別が不十分で、テンソルを形成することができなかったためである）。同一条件下での個体数が 3 つのファイル間で異なるため、

さらに、同一条件下での個体の 3 つのデータセットを選択し、CSV ファイルの列を以下の順序で並べ替えた。

PND1: GSM4556351 GSM4556352 GSM4556353
GSM4556347 GSM4556348 GSM4556349

PND21: GSM4556358 GSM4556359 GSM4556360
GSM4556355 GSM4556356 GSM4556357

PND42: GSM4556365 GSM4556366 GSM4556367
GSM4556361 GSM4556362 GSM4556363

また、解析対象の遺伝子について、本解析では、3 つのファイルで遺伝子発現量が共通して得られている 12,795 個の遺伝子に着目した（PND42 のみ他の 2 つのファイルより多くの遺伝子の発現レベルのデータを含んでいた）。

2.4 解析に用いるテンソルの構造

各テンソル次元の変数は、元データが整形された後に記述される。その構造を図 2.1 に示す。

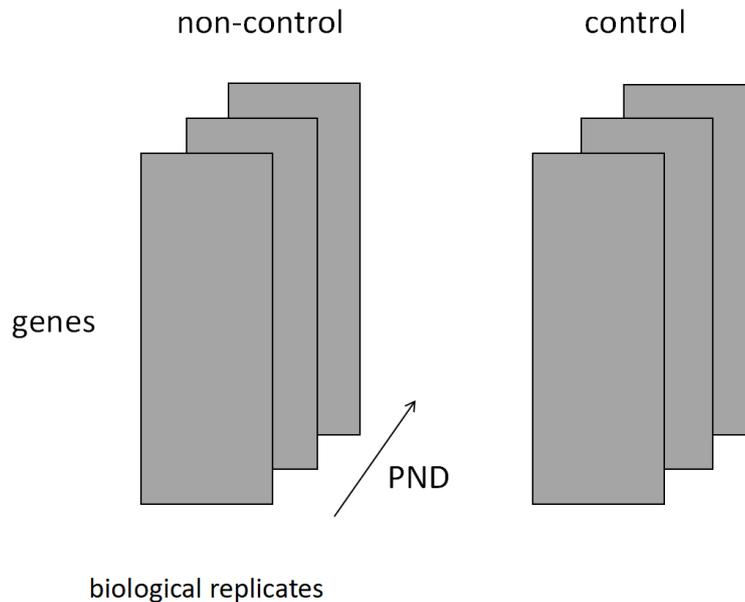


図 2.1: 図 2.1 は、解析に用いた 4 次テンソルを表したものである。1 次元目は遺伝子、2 次元目はバイオロジカルレプリケート、3 次元目は時間経過を生後日数 (PND) で表し、4 次元目は Cd 曝露の有無を示す (“non-control”は Cd 曝露母から生まれた子、“control”は Cd 非曝露母から生まれた子)。各次元のランクを N, M, K, L とすると、 $(N, M, K, L) = (12795, 3, 3, 2)$ となる。

テンソルの 1 次元目には 12,795 個の遺伝子が並んでいる。2 次元目は、前述のように元のデータから選択した同じ条件のマウスのサンプルで構成されている。3 次元目は、産後からトランスクリプトーム解析によって安楽死させるまでの経過日数を表している。PND1、PND21、PND42 の順で並べている。4 次元目は、母親の Cd 曝露の有無を表す。図 1 では、“non-control”が Cd 曝露集団を、“control”が非曝露集団を表している。

これらの変数を用いて 4 次テンソル x_{ijkl} を作成した。各次元における行列のランクを表すパラメータは $(N, M, K, L) = (12795, 3, 3, 2)$ となっている。

2.5 解析方法

2.5.1 テンソル分解を用いた教師なし学習

本分析では、テンソル分解のうち、タッカー分解を用い、アルゴリズムは HOSVD を使用した。例えば、各次元にランク N, M, K を持つ 3 次のテンソル x_{ijk} を考える (本解析で扱ったテンソルは 4 次テンソルであるが、3 次から 4 次へのタッカー分解の拡張は容易である)。そして、このテンソルに対してタッカー分解を適用すると、直交行列 U を用いて以下のように分解を表すことができる。

$$x_{ijk} = \sum_{l_1=1}^N \sum_{l_2=1}^M \sum_{l_3=1}^K G(l_1, l_2, l_3) u_{l_1 i} u_{l_2 j} u_{l_3 k}$$

タッカー分解は、元のデータテンソルをコアテンソル $G(l_1, l_2, l_3)$ と 3 つの特異値行列 $u_{l_1 i}$ 、 $u_{l_2 j}$ 、 $u_{l_3 k}$ に分解する。これらの行列は、それぞれ変数 i 、 j 、 k で指定される依存性を表している。

例えば、行列 $u_{l_1 i}$ の中で大きな値を持つ変数 i を見つけて分析することが目的であるとする。まず、 $u_{l_2 j}$ と $u_{l_3 k}$ にそれぞれ何らかの関心のある依存性に対応するパラメータ l_2' と l_3' を選択する。例えば、 j が時間変化を表す変数であれば、選択した l_2' に対応ベクトル $u_{l_2' j}$ は、単調な増減ベクトルなどわかりやすい変化があるベクトルが好ましい（解析の目的や選択対象になっている変数が何を表しているのかによる）。

次に、選択された l_2' と l_3' をコアテンソル $G(l_1, l_2, l_3)$ に代入する。コアテンソルの各要素は、パラメータ (l_1, l_2, l_3) で指定された各ベクトルの積に対する重みを表すため、結果として得られるベクトル $G(l_1', l_2', l_3')$ において大きな絶対値を持つパラメータ l_1' を選択することで元のテンソルに対して大きく寄与する (l_1', l_2', l_3') を得られる。この考え方にに基づき、 l_1' を選択する。

次に、 l_1' を $u_{l_1 i}$ に代入し、ベクトル $u_{l_1' i}$ を得る。このベクトルは選択したパラメータ l_2' と l_3' の依存性に従うものであり、元のテンソルに対して大きな寄与を持つ。

2.5.2 χ 二乗検定による変数選択

最後に、得られたベクトル $u_{l_1' i}$ の要素のうち、特に大きな値を持つ遺伝子 i を特定するために、このベクトルを正規化したものを用いて χ 二乗検定を行い、 χ 二乗分布によって p 値を付与する。

$$p_i = p_{\chi^2} \left[> \sum_{l_1'} \left(\frac{u_{l_1' i}}{\sigma_{l_1'}} \right)^2 \right]$$

p 値が 0.05 未満の i を選択し、BH 法を用いた多重比較補正（付録参照）を適用することで、任意の依存性に従う過剰発現遺伝子を特定する。

図 2.2 に、本解析における過剰発現遺伝子の選択に関する全体の流れをフローチャートで示す (図 2.2)。

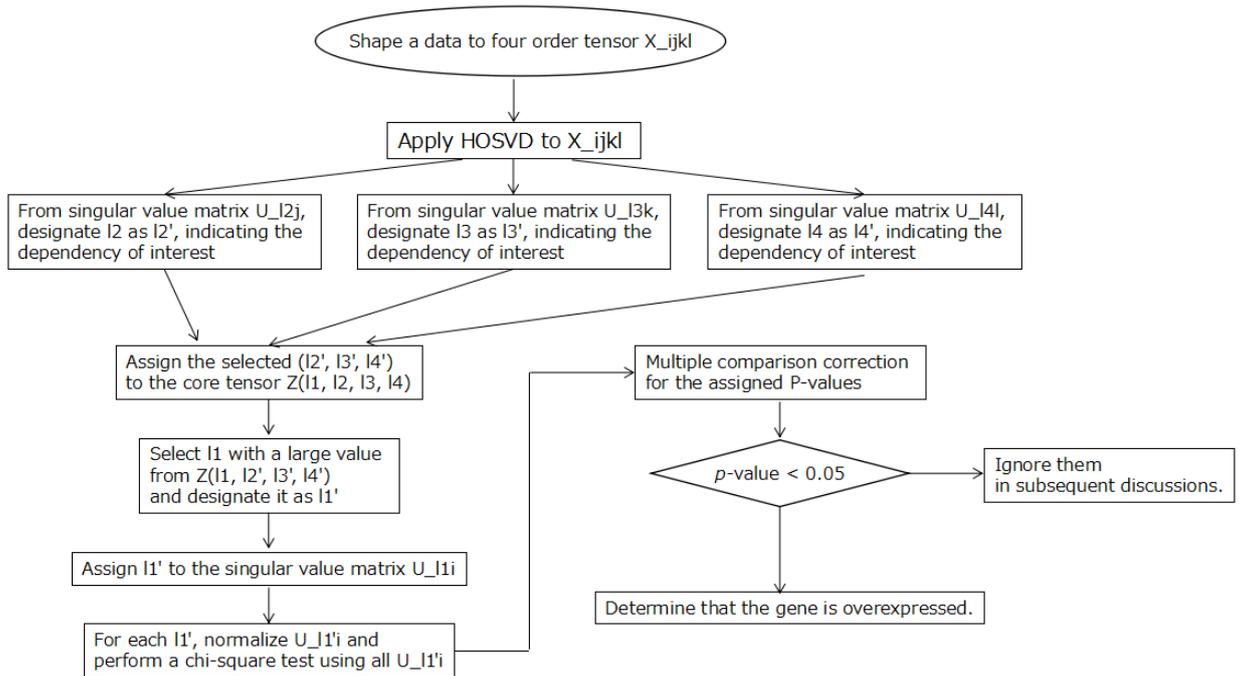


図 2.2: 本解析の全体の流れをフローチャート図に表した。このフローに従って解析を進めることで最終的に、選択した条件下で大きな発現量を持つ遺伝子を特定することができる。

第3章

結果

3.1 発現量が変動している遺伝子の同定

本解析で用いた4階テンソルを統計解析言語 R [22] で HOSVD により分解した。R には他の言語同様、複数のデータ型が存在するが、この分解によって得られるのは、コアテンソルの要素を含む Z と、4つの特異値行列からなるリスト U という二つのリスト型のデータである。R のリスト型は、どのような型のデータでも格納することができる [22]。例えば、ベクトル型のデータには同じ型のデータを格納することしかできないが、これに比較してリスト型のデータはその要素に数値、ベクトル等、また、リスト自身をも持つことができ、複雑なデータを一つのオブジェクトに保有することができる。本解析では特に、リスト U の要素として、各添え字、 i, j, k, l に対応した4つの特異値ベクトル（ベクトル型のデータ）が格納されている。

本解析の目的は、Cd 曝露により発現量が上昇する遺伝子 i を同定することでありその中でも特に大きな発現量を同定することである。我々は、行列 $u_{l_1 i}$ において、母体の Cd 曝露に対して差異があり、バイオロジカルレプリケートに対して不変性を示し、何らかの時間依存的なパラメータ l_1' を決定することから始めた。このようなパラメータ l_1 をコアテンソル Z 中の要素の絶対値をもとに見出すことができれば、 l_1 を $u_{l_1 i}$ に代入して目的の遺伝子 i を得ることができる。そのために、まず、 l_2, l_3, l_4 を選択した。

まず、特異値行列のリスト U から $u_{l_2 j}$ を抽出し、列ベクトル上の値を確認した。 $l_2 = 1$ としたのは、バイオロジカルレプリケートにおいては、同一条件下で実験が行われており、互いに個体差がないのが理想であり、小個体差が小さいほど好ましいからである（図 3.1）。

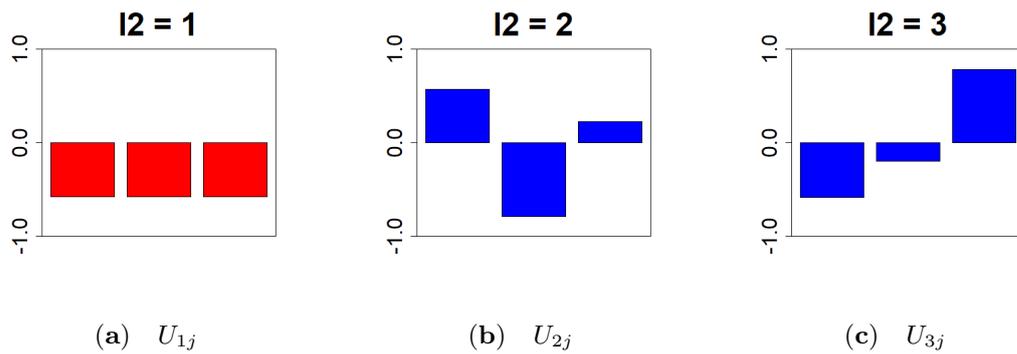


図 3.1: この棒グラフは、特異値行列 $u_{l_2 j}$ の列ベクトル上の値を表したものである。(1 列目: (a)、2 列目: (b)、3 列目: (c)). 本解析においては、同一条件下での個体間の遺伝子発現のばらつきがないことが望ましい。そこで、 $l_2 = 1$ を選択した。

次に、 $u_{l_3 k}$ から l_3 を選択した。単調な増減に従う時間依存性を抽出することで、生後日数が経過するにつれて発現量が上昇するのか下降するのか、より適切に解釈できるため、単調な減少に従う $l_3 = 2$ を選択した (図 3.2)。

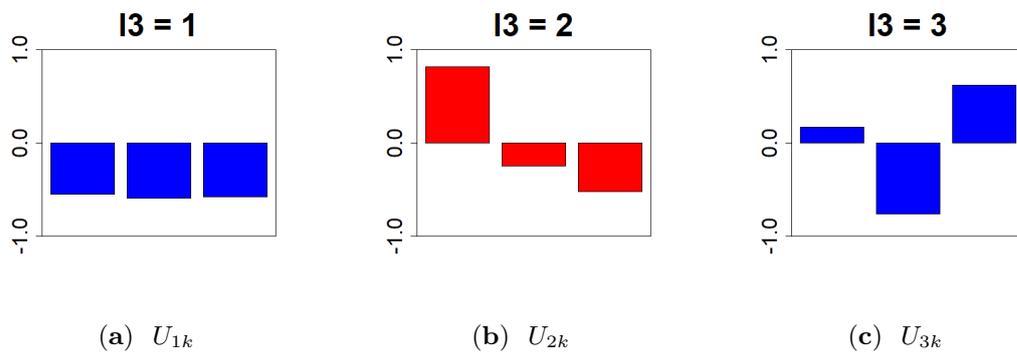


図 3.2: この棒グラフは、特異値行列 $u_{l_3 k}$ の列ベクトル上の値を表したものである。(1 列目: (a)、2 列目: (b)、3 列目: (c)). 目的は、Cd 曝露により発現が時間依存的に変化する遺伝子を選択することである。そこで、PND に対して単調な減少を示す $l_3 = 2$ を選択した。

次に、 $u_{l_4 l}$ から l_4 を選択した。本研究では、Cd 曝露の有無に対して、発現差異の認められる遺伝子を同定することを目的としているため、 $l_4 = 2$ を選択した (図 3.3)。

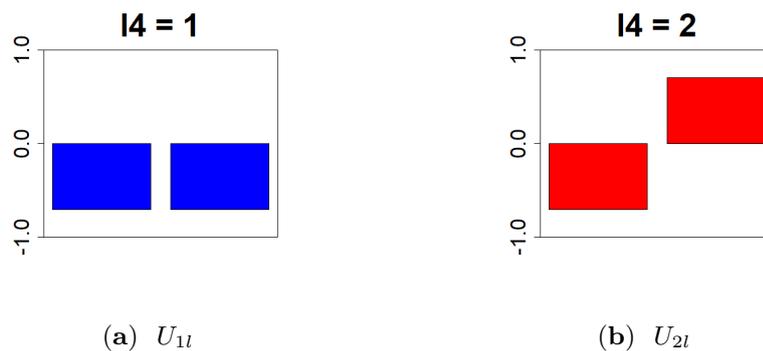


図 3.3: この棒グラフは、特異値行列 $u_{l_4 l}$ の列ベクトル上の値を表したものである。(1 列目: (a)、2 列目: (b)). 本解析の目的は、Cd 曝露による遺伝子発現レベルの違いを示す遺伝子を選択することである。そこで、明確に正負で値が異なっている $l_4 = 2$ を選択した。

ここまでで、 $l_2 = 1$ 、 $l_3 = 2$ 、 $l_4 = 2$ を選択したことにより、同一条件下で個体差が小さく、母親の Cd 曝露量に対して差があり、経時的に発現量が減少する傾向を示す遺伝子を同定する準備ができた。

次に、コアテンソルのリスト Z を取り出し、 $G(l_1, 1, 2, 2)$ というように、コアテンソル G のパラメータを指定することで、長さ 12,795 のベクトルを得た。このベクトル上で絶対値の大きい 11 で示される要素は、パラメータ $l_2 = 1$ 、 $l_3 = 2$ 、 $l_4 = 2$ の依存性に従い、元のテンソルに大きく寄与している。

元のテンソルの次元が大きいため、ベクトル $G(l_1, 1, 2, 2)$ 上の 19 番目以降の要素は非常に小さい値となっている。図 6 では 1~18 番目の要素の値のみを示している。そこで、 $u_{l_1 i}$ に代入する l_1 の値として、絶対値の大きな $l_1 = 3, 11, 14$ を選択した。これらを $u_{l_1 i}$ に代入することで、3 つのベクトル、 u_{3i} 、 u_{11i} 、 u_{14i} を得る。

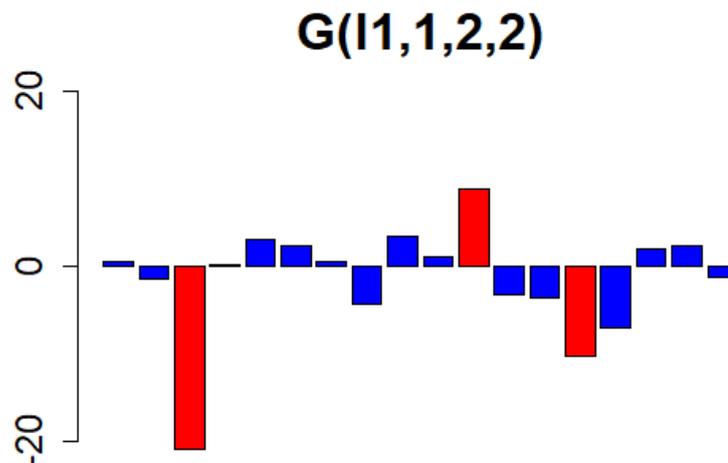


図 3.4: コアテンソル上の $(l_2 = 1, l_3 = 2, l_4 = 2)$ を持つベクトルの値を、第 1 要素から第 18 要素まで棒グラフで表したものである。最も値が大きいもの、上位 3 つ ($l_1 = 3, 11, 14$) を選択した。19 番目以降の要素は、分解前のテンソルのランクの関係で、値が非常に小さいため省略した。

比較的発現量の大きい遺伝子を特定するため、 u_{3i} 、 u_{11i} 、 u_{14i} がそれぞれ独立に正規分布に従うと仮定し、 χ 二乗検定を用いて、 χ 二乗分布によって p 値を割り当てるために、この 3 つのベクトルを用いて、R で `pchisq` コマンドを実行した。さらに、各遺伝子 i の p 値を `p.adjust` コマンドで BH 法を用いて多重比較補正し、 p 値が 0.05 未満の i を選択した。ここで選択された遺伝子 i は、関心のある依存性に従っており、発現量が非常に大きい遺伝子に相当する。ここまでの過程で、合計 12,795 個の遺伝子から、全体の約 1.6 % に相当する 204 個が選択された。

3.1.1 gProfiler によるオントロジー確認

「gProfiler」^{*1}は、主にバイオインフォマティクス、アルゴリズム、データマイニングについての研究を行っているエストニアの研究グループである Bioinformatics Algorithmics and Data Mining Group (BIIT) が開発したウェブツールである。上記の解析で得られた 204 個の遺伝子に対して、gProfiler を用いてエンリッチメント解析を行い、オントロジーを特定したところ、表 3.1 に示すような結果が得られた。

*1 <https://biit.cs.ut.ee/gprofiler>

表 3.1: エンリッチメント解析結果。gProfiler を用いた、204 遺伝子のエンリッチメント解析について、ヒットしたオントロジー、データベース ID、各オントロジーの p 値を記録している。この p 値は補正 p 値である。

Database	Term Name	Term ID	P-Value
KEGG	Metabolic pathway	KEGG01100	1.661E-6
KEGG	Activation of platelets	KEGG04611	2.294E-2
REAC	Lipid metabolism	R-HSA-556833	1.107E-2
GO BP	Ionic transport	GO0006811	2.741E-8
GO BP	Cell adhesion	GO0007155	4.702E-6
GO BP	Intracellular signaling	GO0035556	4.207E-6
GO BP	Cellular protein modification process	GO0036211	5.519E-9
GO BP	Regulation of gene expression	GO0010468	1.241E-8
GO BP	RNA metabolism	GO0016070	9.895E-11
GO CC	Mitochondria	GO0005739	1.990E-4
GO CC	Endoplasmic reticulum	GO0005783	1.990E-3

表 3.1 より、ヒットしたオントロジーの中に、「脂質代謝」、「細胞内シグナル伝達」、「遺伝子発現制御」、「ミトコンドリア」、「小胞体」などがあることがわかる。エンリッチメント解析において、複数のオントロジーがヒットすることが、モデル選択を正当化する一つの理由となっている。解析に失敗した場合、モデルが不適切であった可能性が高く、今回のように複数のオントロジーがヒットすることは稀である。

3.1.2 解析結果の評価

本解析では、単純なテンソル分解モデルを用いて、PND に対して単調減少し、Cd 曝露の有無によって発現量が変化する遺伝子 204 個を同定した。単調減少の PND 依存性は、解釈を明確にするために選んだもので、PND1 から PND42 にかけて、徐々に減少する遺伝子の中でも特に大きな発現量を持つものを同定したことを示す。前述のように、Jackson らの解析では経時的に変化する遺伝子の同定は試みられておらず、この点で、我々の解析は新規性があるものだと考えられる。さらに、同定された 204 個の遺伝子は全体の約 1.6 % に相当し、Jackson らの発現変動解析や、前述の重回帰モデルと比較すると、過剰発現遺伝子の数はかなり減少している。また、エンリッチメント解析では、関連性が高いと思われるオントロジーが複数含まれており、この解析結果は信憑性があるものであると考えることができる。

第4章

考察

4.1 同定された遺伝子の機能

本研究は、インスリンシグナル伝達が阻害されるメカニズムを明らかにすることを目的とした。まず、ERAD 関連遺伝子が過剰発現しているかどうかを確認した。

前述したように、ERAD 誘導時にはユビキチン-プロテアソーム系がたんぱく質の分解を促進する。得られた遺伝子の中には、E2 ユビキチンリガーゼ活性を持つ UBE2E3 (ubiquitin-conjugating enzyme E2 E3) [23]、E3 ユビキチンリガーゼ活性を持つ AMFR (autocrine motility factor receptor) [24]、転写因子 E4F1 (E4F1) [25] や、TRIM32 (tripartite motif-containing protein 32) [26] などユビキチンリガーゼ活性に関与する遺伝子が含まれていた。また、これらの活性を制御している COMMD9 (COMM domain-containing protein 9) [27] も含まれていた。

プロテアソーム関連遺伝子としては、26s プロテアソームのサブユニットをコードする PSMC2 [28] や、プロテアソーム活性化因子複合体 PA28 の 11S 調節因子の一部である PSME1 (proteasome activator complex subunit 1) [29] が高発現していた。このように、ユビキチン-プロテアソーム系に関与する複数の遺伝子が高発現していることは、ERAD が誘導されていることの根拠になると考えられる。

免疫反応に中心的な役割を果たす NF- κ B (Nuclear factor- κ B) 経路に関与する遺伝子の高発現も認められた。NF- κ B と間接的に相互作用する TBK1 (TANK-binding kinase 1) [30]、NF- κ B 経路の活性化因子である CARD11 (caspase recruitment domain-containing protein 11) [31]、負の調節因子である COMMD9 [32] などの遺伝子が含まれていることは、NF- κ B の発現量にも変化があり、何らかの作用を加えている可能性を示唆している。

NF- κ B はサイトカインである TNF- α によって活性化されることが知られている。本解析結果では、TNF- α のプロモーターを活性化する受容体である NFAM1 (NFAT activating protein with ITAM motif 1) [33] の高発現が認められた。このことから、TNF- α も高発現している可能性がある。これまでの研究で、TNF- α の発現と ERAD の発生という二つの生体現象の間には、正の相関があることが示されており [34]、本研究の結果はこの知見と一致するものであると考えることができる。

ERAD 反応過程は ATP 依存的であることから、ATP 産生を上昇させる遺伝子が高発現しているかどうかを確認した。解析結果には ATP 合成を触媒する酵素のサブユニットをコードする ATP5G1

(ATP synthase membrane subunit c locus 1) が含まれていた。

次に、解糖系や脂肪酸代謝の代謝産物は電子伝達系に運ばれ、ミトコンドリアでの ATP 産生を促進するという知見をもとに、これらに関わる遺伝子が解析結果に含まれているか確認した。電子伝達系に関わる遺伝子としては、電子伝達複合体酵素の組み立ての安定化に関わる NDUFAF7(NADH ubiquinone oxidoreductase complex assembly factor 7) [35] と、電子伝達系機能を負に制御する ETFBKMT (electron transfer flavoprotein subunit beta lysine methyltransferase) [36] が高発現していた。また、電子伝達系と解糖系をつなぐ酵素である DLAT (dihydrolipoamide S-acetyltransferase) [37] も高発現していた。

脂肪酸代謝に関わる遺伝子としては、脂肪酸分解を促進する GPIHBP1 (glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1) [38] と、脂肪酸代謝の場であるペルオキシソームの成長因子活性化受容体をコードする PPARA (peroxisome proliferative activated receptor alpha) [39] が高発現していた。これらの代謝物が電子伝達系に移行することを考えると、解糖や脂肪酸代謝に関与する遺伝子の活性化と、電子伝達系関連遺伝子の高発現は互いに関与している可能性がある。また、これらの解糖系や脂肪酸代謝に関わる遺伝子の高発現により電子伝達系が活性化されている可能性が十分に考えられる。

さらに、スフィンゴ脂質の生合成に関わる遺伝子である、Serine SPTLC2(serine palmitoyltransferase, long chain base subunit 2) [40] と SAMD8 (Sterile alpha motif domain containing 8) [41] が高発現していた。このことは、細胞がスフィンゴ脂質を過剰に合成している可能性を示唆している。

4.2 同定した遺伝子によるインスリン代謝阻害機構の推定

母体の Cd 曝露の影響は胎児にも及ぶことが知られている [42]。細胞内に過剰に取り込まれた Cd はたんぱく質のフォールディングを阻害し、これが ERAD の発生につながる。前述のとおり、本解析結果で得られたユビキチン-プロテアソーム系に関わる遺伝子の高発現が ERAD 発生の裏付けとなり、また、解糖系、脂肪酸代謝系、電子伝達系に関わる遺伝子の高発現が ATP 合成を促進し、ERAD の反応機構を活性化している可能性が示唆された。ERAD と TNF- α の発現量が正の相関を示し、TNF- α と NF- κ B の関連遺伝子が高発現していることから、TNF- α が細胞内で過剰分泌されている可能性も示唆された。

TNF- α は ASMAse (酸化スフィンゴミエリナーゼ) を活性化することが知られており、ASMAse はスフィンゴ脂質の産生を促進することが知られている。本解析では、スフィンゴ脂質合成に関わる遺伝子の高発現も確認され、スフィンゴ脂質の蓄積が起こっている可能性が示唆された。スフィンゴ脂質の一つであるセラミドの蓄積は、ROS(Reactive Oxygen Species) の過剰産生を引き起こす。さらに、電子伝達系の反応過程においても ROS が発生する。ROS を介した酸化ストレスは、ASK1 (apoptosis signal-regulating kinase 1) を活性化することが知られている [43]。PISD(Phosphatidylserine decarboxylase proenzyme) は、本研究で同定された高発現遺伝子の一つであるが、PISD もまた、ASK1 の活性化に役割を果たす。

ASK1 は JNK を活性化し、JNK (c-Jun N-terminal kinases) は、インスリンシグナルにおいて決

定的な役割を果たす IRS1 の 307 セリン残基をリン酸化することで不活性化させる。従って、TNF- α によるセラミド蓄積を介した酸化ストレスによるインスリン抵抗性発症の経路を本解析結果から推測することができる。しかし、本解析では、ERAD と TNF- α の高発現がどのような機序によって正の相関を示すのかについては、推測することができなかった。

また、上記の遺伝子以外にも、表 1 に示すように、gProfiler によるエンリッチメント解析において、「遺伝子発現の制御」や「RNA 代謝」などのオントロジーもヒットしていた。これは、遺伝子の発現量は何らかの機序で変化していることを示唆しているが、その具体的なメカニズムは本研究では明らかにすることができなかった。また、ERAD 発症時のストレス回避の最終手段であるアポトーシスに関連する遺伝子も多数見つかったが、今回推測したインスリン抵抗性獲得経路とアポトーシスが実際にどのように関連しているのかは不明である。今後は、これらの因子がどのように関与しているのか、特に遺伝子発現変化の原因や影響に関する部分や、ERAD と TNF- α の関連性について、重点的に解明していく必要があると考えている。

第5章

結論

妊娠中の Cd 曝露が子孫のインスリン代謝を阻害するという研究報告に基づき、この生物学的現象の原因を解明するために、肝臓トランスクリプトーム解析データを用いたテンソル分解により 204 個の過剰発現遺伝子を同定した。

解析結果として得られた遺伝子中には、ERAD の形成やストレス回避経路であるユビキチン・プロテアソーム系に関わる遺伝子も含まれており、Cd 毒性によってたんぱく質のフォールディングが阻害されていることが示唆された。また、スフィンゴ脂質の産生、電子伝達系に関わる遺伝子群、また、電子伝達系に代謝産物を提供する代謝経路上の遺伝子群も高発現しており、細胞内の酸化ストレスが過剰であることが示唆された。上記の ERAD と細胞内酸化ストレスの結果は、Jackson らの結果と一致した [20]。

また、TNF- α は、Cd 毒性による ERAD とスフィンゴ脂質の高発現をつなぐ役割を担っている可能性が示唆されている。TNF- α のプロモーターをコードする遺伝子や NF- κ B に関連する遺伝子が高発現していることはこの仮説を支持するものであった。しかし、この解析では、ERAD と TNF- α の関連性の具体的なメカニズムを明らかにすることはできなかった。

TNF- α とスフィンゴ脂質の関連性から、TNF- α が ASMase を活性化することによりセラミドの過剰産生を引き起こしているという経路を推定することができた。インスリン抵抗性の獲得経路に関しては、細胞内酸化ストレスが寄与している可能性が推定できた。転写レベルでの遺伝子発現の変化に及ぼす Cd の影響については、引き続き調査する必要がある。

謝辞

本研究は、中央大学大学院理工学研究科物理学専攻の非線形物理学研究室で行われたものです。

本論文執筆にあたっては、田口善弘教授に多大なるご協力を頂きました。研究開始当初は物理学科所属ということもあり、何もかも一から始めなければなりませんでした。そんな状況の中、生物学的な知識、機械学習の基本、Rでのプログラミング、テンソル分解という解析手法に至るまで、非常に多くのことを勉強させていただきました。また、学術的な知識だけではなく、学術誌への論文投稿の際には、論文の執筆の仕方をはじめ多くのことをご教授いただきました。

田口教授のおかげで、大学4年時からの3年間、非常に充実した時間を過ごすことができましたし、大きく成長することができたと確信を持って言えます。田口教授に心から御礼申し上げます。

本研究を副査していただき、本論文をより良いものにしてくださった坪井陽子教授、米満賢治教授に深く感謝申し上げます。本研究及び大学院での生活を送る上で日本学生支援機構には、多大な支援を頂きました。深く感謝申し上げます。

最後に、日常生活を支えてくださった両親に心より御礼申し上げます。

付録

多重比較補正

仮説検定を複数回実行する際、一般に繰り返した検定の数が多いほど、偶然仮説が棄却される可能性は高くなる。このような複数回繰り返された検定全体において帰無仮説が棄却される可能性のことを「familywise error rate」という。例えば、有意水準 $\alpha = 0.05$ の検定を 20 回繰り返したとき、一回でも帰無仮説が棄却される可能性 (familywise error rate) は 0.642 であるのに対して、これを 100 回繰り返したときの familywise error rate は、0.994 である。検定を繰り返し複数回行う (多重比較する) 際は、回数の増加に応じて familywise error rate が大きくなるという事実を考慮して、検定の有意水準を当初の設定値よりも小さくしなければ、本来棄却されることがない帰無仮説を棄却してしまうということが起こりうる。

この問題を是正する方法、つまり、familywise error rate を調整する方法を総称して多重比較補正と呼ぶ。これには複数の方法が考えられるが、大きく二つに分けられる。一つは、F 統計量や t 統計量などの統計量に基づいた方法である。もう一つは、P 値のみを調整する方法である。このうち、本解析で用いた BH 法は後者にあたるものである。特に BH 法は false discovery rate (FDR) を調整する方法である。FDR とは、「検定結果が有意になった数の中の、偽陽性の割合」を示す値である。帰無仮説が棄却された検定数を R 、実際には真である帰無仮説を棄却した検定数を v と置くと、FDR の定義は以下ようになる。

$$FDR = \frac{v}{R}$$

R は検定結果から得られる観測可能な値であるが、一方で v は観測不可能である。そこで FDR は、目的に応じた値に設定される [44]。

v は観測不可能であることから、 R を調整する、つまり棄却される帰無仮説の数を調整することで FDR が調整できる。

BH 法 (Benjamini & Hochberg 法)

BH 法は、1995 年に Benjamini と Hochberg によって発表された FDR を調整する方法である [44]。ここでは、実際のアлゴリズムを示す [45]。

- (1) 任意の FDR を定める。この値を α とする。
 - (2) m 個の帰無仮説それぞれに対して検定を行い、 p 値を計算する。
 - (3) 検定によって得られた m 個の p 値を昇順に並べて $p_1 \leq p_2 \leq \dots \leq p_m$ とする。ここで、帰無仮説 i に対応する p 値を p_i とする。
 - (4) $j = m$ とおく。
 - (5) $p_j \leq \alpha j / m$ ならば、(6) に進む。そうでなければ、 $j = j - 1$ として、この手順を繰り返す。ただし、 $j = 1$ となったとき、どの帰無仮説も棄却せずに、終了する。
 - (6) 帰無仮説 1、帰無仮説 2、 \dots 、帰無仮説 j を棄却して、終了する。
- (6) の過程によって、棄却される帰無仮説の数を減らし、FDR を調整する。

参考文献

- [1] Ohba, K.I. Transport and Toxicity of Cadmium. *Nichieishi* 2006, 73, 269–274.
- [2] MHLW. 「食品に含まれるカドミウム」に関するQ&A.
<https://www.mhlw.go.jp/houdou/2003/12/h1209-1c.html#01>.
- [3] 富山県. くらし・健康・教育. イタイイタイ病とは.
<https://www.pref.toyama.jp/1291/kurashi/kenkou/iryuu/1291/100025/100026.html>.
- [4] 富山県. くらし・健康・教育. イタイイタイ病の特徴.
<https://www.pref.toyama.jp/1291/kurashi/kenkou/iryuu/1291/100035/virtual/virtual02/virtual02-3.html>.
- [5] Quynh, G.L; Yuki, I; Kenji, K.; Yukio, K. Cadmium impairs protein folding in the endoplasmic reticulum and induces the unfolded protein response. *FEMS Yeast Reseach*. 2016. 16(5).
- [6] biology. 細胞生物学. 細胞小器官.
<https://bio.biopapyrus.jp/cell/organelle.html>.
- [7] 公益社団法人日本生化学会. レドックス制御による小胞体恒常性維持機構の解明～還元反応の場としての小胞体～.
<https://seikagaku.jbsoc.or.jp/10.14952/SEIKAGAKU.2020.920536/data/index.html>.
- [8] 公益社団法人日本生化学会. 小胞体ストレスと疾患.
<https://seikagaku.jbsoc.or.jp/10.14952/SEIKAGAKU.2018.900051/data/index.html>.
- [9] 厚生労働省. e-ヘルスネット. 健康用語辞典. 生活習慣病. インスリン.
<https://www.e-healthnet.mhlw.go.jp/information/dictionary/metabolic/ym-011.html>.
- [10] 公益社団法人日本薬学会. 薬学用語解説. インスリン受容体.
<https://www.pharm.or.jp/dictionary/wiki.cgi?%E3%82%A4%E3%83%B3%E3%82%B9%E3%83%AA%E3%83%B3%E5%8F%97%E5%AE%B9%E4%BD%93>.
- [11] ニュートリ株式会社. キーワードでわかる臨床栄養. 2-2：グリコーゲンの代謝.
<https://www.nutri.co.jp/nutrition/keywords/ch2-2/keyword5/>.
- [12] ヤクルト中央研究所. 健康用語の基礎知識. インスリン抵抗性.
https://institute.yakult.co.jp/dictionary/word_3121.php.
- [13] 公益社団法人日本薬学会. 薬学用語解説. アディポカイン.
<https://www.pharm.or.jp/dictionary/wiki.cgi?%E3%82%A2%E3%83%87%E3%82%A3%E3%83%9D%E3%82%AB%E3%82%A4%E3%83%B3>.

- [14] International Diabetes Federation. IDF Diabetes Atlas Tenth Edition. <https://diabetesatlas.org/>.
- [15] 日本生活習慣病予防協会. 生活習慣病の調査・統計. 糖尿病. <https://seikatsusyukanbyo.com/statistics/disease/diabetes/>.
- [16] Shaikh, Z.A.; Vu, T.T.; Zaman, K. Oxidative stress as a mechanism of chronic cadmium-induced hepatotoxicity and renal toxicity and protection by antioxidants. *Toxicol. Appl. Pharmacol.* 1999, 154, 256–263.
- [17] Takashige, K.; Hiroyuki, S.; Rie, F.; Yoshito, K.; Masahisa, I.; Kojun, S.; Shinya, S.; Masao, S. Cadmium reduces adipocyte size and expression levels of adiponectin and Peg1/Mest in adipose tissue. *Toxicology* 2010, 267, 20–26 .
- [18] Heindel, J.J.; Bruce, B.; Mathew, C.; Ronit, M.; Alberto, M.; Michelle, A.M.; Angel, N.; Paola, P.; Giancarlo, P. Metabolism disrupting chemicals and metabolic disorders. *Reprod. Toxicol.* 2017, 68, 3–33.
- [19] Luo, Y.; McCullough, L.E.; Tzeng, J.Y.; Darrah, T.; Vengosh, A.; Maguire, R.L.; Maity, A.; Samuel-Hodge, C.; Murphy, S.K.; Mendez, M.A.; et al. Maternal blood cadmium, lead and arsenic levels, nutrient combinations, and offspring birth weight. *BMC Public Health* 2017, 17, 354.
- [20] Thomas, W.J.; Garret, L.R.; Chris, M.S.; Aubrey, L.S.; Guillette, T.C.; Scott, M.B. Gestational Cd Exposure in the CD-1 Mouse Induces Sex-Specific Hepatic Insulin Insensitivity, Obesity, and Metabolic Syndrome in Adult Female Offspring. *Toxicol. Sci.* 2020, 178, 264–280.
- [21] Taguchi Y-h. Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach; Springer International Publishing AG: Cham, Switzerland, 2019; 329p.
- [22] 富永大介. Rをはじめよう生命科学のための RStudio 入門; 羊土社; 2019
- [23] Ito, K.; Kato, S.; Matsuda, Y.; Kimura, M.; Okano, Y. cDNA cloning, characterization, and chromosome mapping of UBE2E3(alias UbCH9), encoding an N-terminally extended human ubiquitin-conjugating enzyme. *Cytogenet. Cell Genet.* 1999, 84, 99–104.
- [24] Marilyn, R.; Jacky, G.G.; Pascal, S.P.; Hao, P.; Monique, L.; Michel, B.; Phuong, U.L.; Ivan, R.N. The gene product of the gp78/AMFR ubiquitin E3 ligase cDNA is selectively recognized by the 3F3A antibody within a subdomain of the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.* 2004, 320, 1316–1322.
- [25] Laurent, L.C.; Laëtitia, K.L.; Conception, P.; Eric, J.; Matthieu, L.; Elodie, H.; Robinson, T.; Guillaume, B.; Ayelet, S.; Manuel, S.R.; et al. E4F1 Is an Atypical Ubiquitin Ligase that Modulates p53 Effector Functions Independently of Degradation. *Cell* 2006, 127, 775–788.
- [26] Annie, P.C.; John, S.B.; Hsan-Jan, Y.; Marwan, K.T.; Todd, E.S.; Ruth, E.S.; Darryl, Y.N.; Terry, A.B.; Kwang-Youn, A.K.; Jian, H.; et al. Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet–Biedl syndrome gene (BBS11). *Proc.*

- Natl. Acad. Sci. USA 2006, 103, 6287–6292.
- [27] Zhan, W.; Wang, W.; Han, T.; Xie, C.; Zhang, T.; Gan, M.; Wang, J.B. COMMD9 promotes TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer. *Cell Signal.* 2017, 30, 59–66.
- [28] Tanahashi, N.; Suzuki, M.; Fujiwara, T.; Takahashi, E.I.; Shimbara, N.; Chung, C.H.; Tanaka, K. Chromosomal Localization and Immunological Analysis of a Family of Human 26S Proteasomal ATPases. *Biochem. Biophys. Res. Commun.* 1998, 243, 229–232.
- [29] Ma, C.P.; Slaughter, C.A.; DeMartino, G.N. Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). *J. Biol. Chem.* 1992, 267, 10515–10523.
- [30] Pomerantz, J.L.; Baltimore, D. NF-kappaB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* 1999, 18, 6694–6704.
- [31] Bertin, J.; Wang, L.; Guo, Y.; Jacobson, M.D.; Poyet, J.L.; Srinivasula, S.M.; Merriam, S.; DiStefano, P.S.; Alnemri, E.S. CARD11 and CARD14 Are Novel Caspase Recruitment Domain (CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members that Interact with BCL10 and Activate NF- κ B. *J. Biol. Chem.* 2001, 276, 11877–11882.
- [32] Maine, G.N.; Burstein, E. COMMD proteins: COMMMing to the scene. *Cell. Mol. Life Sci.* 2007, 64, 1997–2005.
- [33] Martha, L.S.; Abbie, L.; Kristina, L.B.; Roger, K.W. Tumor necrosis factor-related genes and colon and rectal cancer. *Int. J. Mol. Epidemiol. Genet.* 2011, 2, 328–338.
- [34] Liu, Q.; Yang, X.; Long, G.; Hu, Y.; Gu, Z.; Boisclair, Y.R.; Long, Q. Endoplasmic Reticulum Associated Degradation(ERAD) Deficiency Promotes Mitochondrial Dysfunction and Transcriptional Rewiring in Human Hepatic Cells. *J. Biol. Chem.* 2020, 295, 16743–16753.
- [35] Olga, Z.R.; Lissiene, S.N.; Florin, S.; Eric, A.S. The arginine methyltransferase NDUFAF7 is essential for complex I assembly and early vertebrate embryogenesis. *Hum. Mol. Genet.* 2014, 23, 5159–5170.
- [36] Małeckı, J.; Jakobsson, M.E.; Ho, A.Y.; Moen, A.; Rustan, A.C.; Falnes, P.Ø. Uncovering human METTL12 as a mitochondrial methyltransferase that modulates citrate synthase activity through metabolite-sensitive lysine methylation. *J. Biol. Chem.* 2017, 292, 17950–17962.
- [37] Rommel, A.M.; Todd, M.G.; Adam, O.; Hanna, G.B.; Rumela, C.; Elizabeth, A.R.; Yibin, K.; Thomas, S.; Ileana, M.C. Sirtuin 4 is a lipoamidase regulating pyruvate dehydrogenase complex activity. *Cell* 2014, 159, 1615–1625.
- [38] Young, S.G.; Davies, B.S.J.; Fong, L.G.; Gin, P.; Weinstein, M.M.; Bensadoun, A.; Beigneux, A.P. GPIHBP1: An endothelial cell molecule important for the lipolytic processing of chylomicrons. *Curr. Opin. Lipidol.* 2007, 18, 389–396.
- [39] Sander, K. Integrated physiology and systems biology of PPAR α . *Mol. Metab.* 2014, 3, 354–371.

-
- [40] Akio, K. Various Metabolisms, Physiological Functions and Related Diseases of Fatty Acids. *Biochemistry* 2010, 82, 591–605.
- [41] Fikadu, G.T.; Ana, M.V.; Elleke, F.B.; Martin, H.; Amrita, J.; Angelika, H.; Pentti, S.; Joost, C.M.H. Sphingomyelin synthase-related protein SMSr is a suppressor of ceramide-induced mitochondrial apoptosis. *J. Cell Sci.* 2014, 127, 445–454.
- [42] Atsuhiko, N.; Masaji, O.; Yukiko, O.; Ichiro, W. Placental Transfer of Cadmium in Pregnant Women. *Jpn. J. Hyg.* 1989, 44, 579–586.
- [43] Tsuruzoe, K.; Nishikawa, T.; Araki, E. Oxidative Stress and Insulin Resistance. *Diabetes* 2006, 49, 845–848.
- [44] 大阪大学大学院医学研究科. Clinical Journal Club 1. 多重比較.
<https://www.med.osakau.ac.jp/pub/kid/clinicaljournalclub1.html>.
- [45] biostatistics. 基礎統計. Benjamini-Hochberg 法.
<https://bestbiostatistics.com/multiple/fdr.html>.