

# 主成分分析を用いた教師なし学習による変数選択を用いた ヒストン脱アセチル化酵素阻害剤の機能探索

田口 善弘<sup>1,a)</sup>

概要: 近年、エピジェネティック療法に注目が集まっている。その有効な1つとしてヒストンの脱アセチル化酵素阻害剤 (HDACi) が注目されている。恐怖記憶の消去は心的外傷後ストレス障害の治療に重要な療法だが、HDACi の投与が恐怖記憶の加速に貢献するというネズミを用いた実験が提案され、ヒストン修飾に対する効果が恐怖記憶の消去に貢献しているという報告がなされた。我々はこのデータを再解析し、HDACi の標的がヒストン修飾であるとは必ずしも言えないこと、また、他のプロセスが重要である可能性があることを報告した。

Y-H. TAGUCHI<sup>1,a)</sup>

## 1. はじめに

心的外傷後ストレス障害 (post-traumatic stress disorder, PTSD) [2] は生命に係るような強いストレス ( 戦場、多数の死者を伴う大災害 ) にさらされた場合、事後に発症する精神的な疾患である。症状としては、当時のことを思い出すだけで臨場的な恐怖感が再帰される、不眠、体験のフラッシュバックなど精神的なものが主たる症状であるが、その治療は困難であり、同時多発テロや、中東への派兵で多くの PTSD 患者をかかえるアメリカなどに於いてはその治療法の確立が焦眉の急となっている。しかし、PTSD 患者を実験対象とすることは道義的に難しく、治療法の確立もまた、困難を抱えている。

このような観点からいわゆる動物モデルの作成が試みられてきた [3]。その1つに正常なマウスに文脈的恐怖条件付けを行うことで擬似的な PTSD 状態を創りだして研究対象とする言う方法 [4] がある。文脈的恐怖条件付けにおいては、マウスをチャンバーに入れた状態で、特定の音を聞かせた後、電気ショックを与えるという条件付けを行う。このような条件付けをされたマウスは、音刺激なしでもチャンバーに入れたただけで恐怖によるフリーズを起こすようになる。この様に音という外的な刺激なしに、チャンバー

という環境だけでも反応を起こすようになることを「文脈的」と呼んでいる。この文脈的恐怖条件付けはチャンバーに入れられても電気刺激が与えられないという学習をすることで消去される (いわゆる消去学習 [5]) ことが知られているが、「チャンバーに入れられても電気刺激が与えられない」という学習は音刺激と電気刺激を結びつける学習より遥かに難しく、これが PTSD の治療の困難と強く結びついていると目されている。したがって、消去学習をいかに効率的に行うかが PTSD の治療に直接役立つものと思われる。

一方、近年、精神疾患とエピジェネティクスの関係が注目され [6]、治療対象としても扱われるようになってきた [7]。このような文脈から Gräffらはマウスの消去学習にヒストンの脱アセチル化酵素阻害剤 (HDACi) の投与が有効であることを報告している [4]。確かに、記憶という可塑的なプロセスに、同じく可塑的なプロセスであるエピジェネティクスが介在し、そこに調整を入れることで消去学習が加速されるというストーリーは美しく、わかりやすい。

本研究では Gräffら [4] の論文で解析された遺伝子発現プロファイルの元データを再解析し、この主張の妥当性を検証した。その結果、Gräffらがいうほどには HDACi の効果が消去学習に貢献しているとは必ずしも言えないのではないかと、という結論に至った [1] のでそれについてここに報告する。

<sup>1</sup> 中央大学理工学部物理学科  
東京都, 112-8551, 日本

<sup>a)</sup> tag@granular.com  
本研究内容は既に原著論文 [1] として公開済みである

## 2. 手法

### 2.1 mRNA プロファイル

本研究で用いた mRNA プロファイルはすべて GEO [8] からダウンロードした。詳しくは原著論文 [1] を見ていただくとして、具体的には Gräff らが用いた HDACi である CI-994 に加えて、別の研究グループが行った別の HDACi (FTY720) 投与実験の発現プロファイルも合わせて利用した。対象組織部位はいずれも消去学習と深い関係があるとみなされている海馬である。両実験共、文脈的恐怖条件付けの消去学習における投与薬剤の効果を確かめる目的で行われた実験である。各プロファイルはサンプルごとに平均0, 分散1 に規格化してから使用した。

### 2.2 主成分分析を用いた教師なし学習による変数選択

HDACi の投与のある無しで発現差がある mRNA のスクリーニングを行った。その手法としては当該目的に有効であることが示されている主成分分析を用いた教師なし学習による変数選択法 [9–26] を用いた。本手法についての詳細は原著論文 [1] を見ていただくとして概要を述べる。まず、第一に主成分分析を用いて遺伝子を低次元空間に埋め込む。この結果、サンプルに主成分得点が、遺伝子に主成分負荷量が割り当てられる通常のサンプル埋め込みとは逆に、遺伝子に主成分得点が、サンプルに主成分負荷量が付与される。通常のサンプル埋め込みとの違いは、サンプル埋め込みの場合は共分散行列、または、相関係数行列が対角化されるために遺伝子ごとに発現プロファイルの平均がゼロになるのに対して、遺伝子埋め込みの場合はグラム行列が対角化されるためにサンプルごとの平均プロファイルがゼロになる点である。また、一般に、サンプル平均、遺伝子ごと平均のそれぞれをゼロにした場合の効果は自明ではなく、この2つの主成分分析は一般には異った結果を与えることになり、どの様に異なるかを事前に計算することもできない点は注意を要する。また、サンプル平均と遺伝子平均を同時にゼロにすることはできないので、どちらか一方を選択することになる。主成分分析を用いた教師なし学習による変数選択において、サンプル平均をゼロにする遺伝子埋め込みの方を選択する理由は、一般に遺伝子発現プロファイルの計測では絶対値の計算は不可能であり、なんらかの正規化が必要なためである。遺伝子平均をゼロにするサンプル埋め込みの場合は、前処理として遺伝子発現プロファイルがサンプルごとに何らかの仮定に基づいてあらかじめ正規化されていることが多い。しかし、そこに任意性が入ってしまったのでは、せっかくの教師なし学習の意味がない。そこで、通常、主成分分析を用いた教師なし学習による変数選択ではサンプル平均をゼロにする方が選択される。

次に、遺伝子選択に用いる主成分を主成分負荷量を参考に選択する。今の場合、主成分負荷量はサンプルに紐付けられているために、主成分負荷量を観測することでどの主成分に生物学的に意味がある特徴が表れているかを知ることができる。本研究の場合には、HDACi 投与の有無で発現差がある主成分負荷量を選んだ。

最後に、選択された主成分負荷量に対応する主成分得点に多重ガウス分布を仮定し、 $\chi$  二乗分布を適用して  $P$  値を計算し、外れ値となっている遺伝子を選択する (多重比較補正された  $P$  値が 0.01 や 0.05 以下を外れ値とする)。なぜなら、このような遺伝子は注目する主成分に対して、偶然では説明できない大きな寄与を持っている、つまり、HDACi の投与に起因する発現差に寄与している遺伝子であるとみなすことができるからである。主成分得点がガウス分布に従うという仮定の是非には議論があるものと思われるが、この様な仮定は確率主成分分析と呼ばれる、主成分分析の統計学習的フレームでの再解釈でも採用されており [27]、それほど現実から離れたものではないと思われる。

## 3. 結果

### 3.1 文脈的恐怖条件付けマウスの消去学習実験に係る重要遺伝子の同定

マウスの文脈的恐怖条件付けの消去学習に2種類の HDACi を用いた場合、有意に発現量が変化している遺伝子を同定した。これは言葉で言うほど簡単ではない。2つの発現プロファイルにはそれぞれ、3組と4組(1組はそれぞれ HDACi 投与・非投与マウスのペアからなるので、サンプル数でいえば6サンプルと8サンプル)の実験が無い。これに対し、発現する可能性がある遺伝子の数は数万個に及ぶ。サンプル数を  $n$ 、変数の数を  $p$  とした場合の典型的な  $n \ll p$  問題になってしまっている。検定問題としてはこれは単純な二群比較であり、それほど工夫の余地はない。サンプル数の少なさのため、帰無仮説(二群に差はない)の棄却確率はどうしても大きくなり、棄却が困難になる。さらに数万個の多重比較であるために得られた  $P$  値の多重比較補正が必要になり、ただでさえ大きめの  $P$  値がさらに大きめに補正され、ますます帰無仮説の棄却、すなわち、有意に発現差のある遺伝子の同定、が困難になる。実際、このような極端な  $n \ll p$  問題において HDACi の投与の効果を見積もるといってささいな変化を検出しようとしてもうまく行かない場合が多い。Gräff らは数百個の有意に発現差がある遺伝子を  $p \leq 0.05$  の基準で選択しているが、論文には多重比較補正をしたという記述はなく、検定という意味では多数の擬陽性 (false positive, FP) を含んでいると思われる(それでも著名雑誌に論文として掲載されてしまうのが問題なのだ)。

この困難を乗り越えるために我々は従来から提案してきた主成分分析を用いた教師なし学習による変数選択

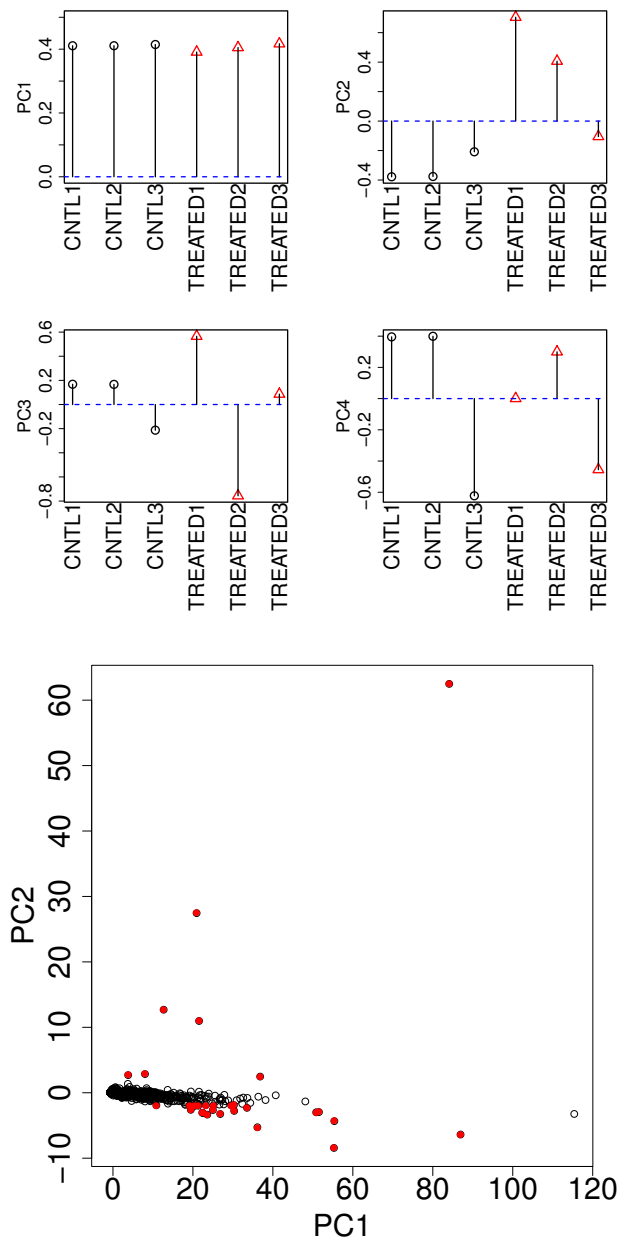


図 1 主成分分析の結果 (CI-994)。上: 第一~四主成分負荷量、下: 第1と第2主成分得点の散布図。赤い丸が選択された遺伝子。  
Fig. 1 The results of PCA(CI-994). Upper: The 1st to 4th PC loadings, lower: Scatter plot of the first and the second PC scores. Red circles are genes selected.

法 [1,9-26] を採用した。同手法は典型的な  $n \ll p$  問題であっても、多重比較補正後も有意 (0.05 とか 0.01 とかで帰無仮説を棄却できる) な遺伝子を選択できる能力が比較的高いことが知られている。まずは、Gräffらが行った CI-994 の投与実験の方から解析を行った。この発現プロファイルはアレイではなく NGS で計測されたものである。図 1 に 6 サンプル (3 組) に付与された第一から四主成分負荷量のサンプル依存性を示す。遺伝子の選択に用いるには HDACi の処理のある無しで有意に発現差がある主成分得点を同定する必要がある。2 群の  $t$  検定を行ったと

ころ、有意に差があるのは第二主成分負荷量だけあった ( $P = 0.05$ )。そこで第二主成分得点が高い 30 遺伝子を選択することができた (図 1 の赤丸, 補正後  $P$  値が 0.05 以下)。これでまずは第一関門である極端な  $n \ll p$  問題であるために有意差のある遺伝子が検出できないという問題は回避できた。

しかし、サンプル数も少なく、単一の実験に基づく遺伝子の同定には不安が残る。そこでもう 1 つの HDACi (FTY720) を用いた、同様の文脈的恐怖条件付けマウスの消去学習実験の遺伝子発現プロファイルを解析して比較することを試みた。もし、同定した遺伝子が文脈的恐怖条件付けマウスの消去学習に固有の遺伝子であるならば、他の HDACi の投与でも遺伝子発現の変化が見られるはずである。2 番目の実験は NGS ではなくアレイによるものであり、実験グループや動物も同じではないという条件なのでそれなりのロバストネスがない限り、一致は望めない。さらに数万個の遺伝子からわずか 30 個を選んだという条件なので相当に偶然に恵まれない限り、独立した実験で同じ遺伝子がたまたま同定されるとは考えにくい。図 2 に 8 サンプル (4 組) に付与された第一から四主成分負荷量のサンプル依存性を示す。CI-994 の時と異なり、今回は残念ながら、HDACi の投与の有無で有意に差がある主成分負荷量は第四主成分までの中には見つからなかった。そこで図 3 にあるように、第三、四主成分をまとめて一個の変量と見て有意差を検定した見たところ、 $P = 0.04$  となり、かろうじて有意差を検出することができた。そこで改めて第三、四主成分得点を用いて図 2 の赤丸で示される 364 個の遺伝子を同定できた (補正した  $P$  値が 0.01 以下)。364 個と言うと多そうに思われるが、遺伝子の総数が数万であることを考えればたかだか数%である。にも関わらずのこの中には CI-994 を用いた実験データの解析で同定された 30 個の遺伝子のうち 11 個が含まれていた。これが偶然に起きる確率は  $P = 5.92 \times 10^{-12}$  であり (フッシャーの正確確率検定による) 事実上 0 である。したがって、CI-994 で同定された 30 個の遺伝子は独立した実験の結果と偶然ではありえないほどの高い一致率をもっており、文脈的恐怖条件付けマウスの消去学習に本質的な役割を担っていることが期待できる遺伝子であると言える。

### 3.2 同定された遺伝子の生物学的な妥当性

次に同定された 30 個の遺伝子の生物学的な妥当性を検証する。30 個という数はいかにも少なく、本当に意味がある選択であれば喜ばしいがたまたま発現の大きな遺伝子が 30 個選ばれただけであるとすると意味がない。この様な場合によく試みられるのが、enrichment analysis である。これは予め機能ごとにグループ化された遺伝子を用いて、そのグループ内の遺伝子が偶然ではありえないほどの有意差で同定された 30 個の遺伝子に集中して選択されていれば、

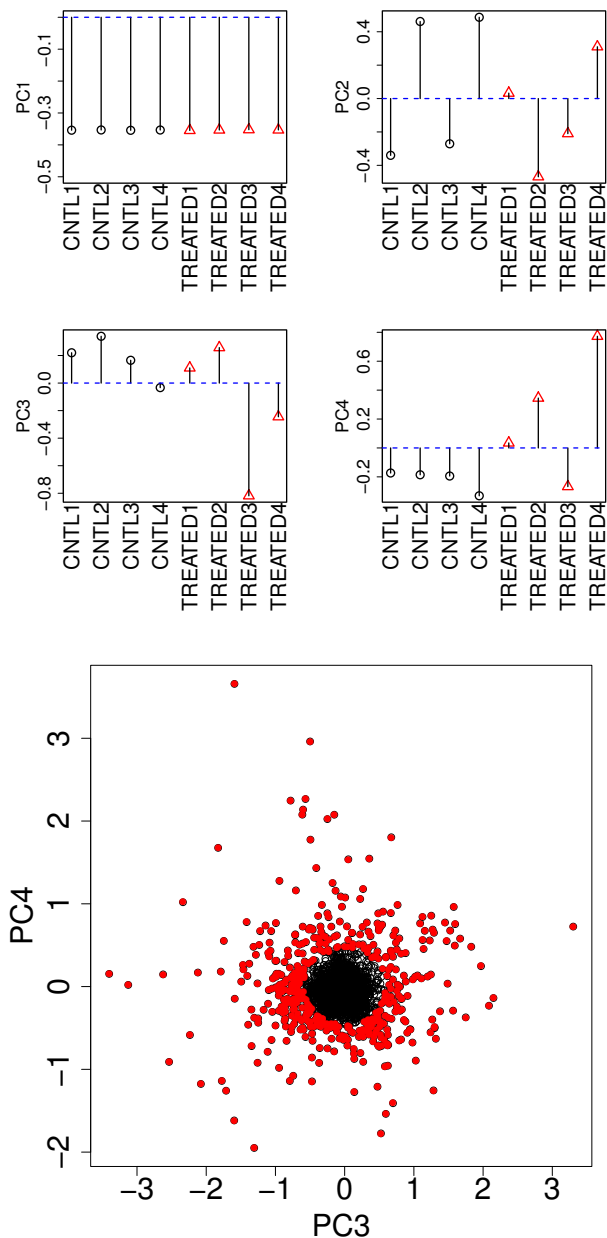


図 2 主成分分析の結果 (FTY720)。上: 第一~四主成分負荷量、  
下: 第3と第4主成分得点の散布図。赤い丸が選択された遺  
伝子。

Fig. 2 The results of PCA(FTY720). Upper:The 1st to 4th PC  
loadings, lower:Scatter plot of the third and the fourth  
PC scores. Red circles are genes selected.

30個の遺伝子は生物学的にこの機能を担っている確率が高いとみなすやり方である。たまたま発現の大きな遺伝子が30個選ばただけであればこのようなことは起き得ない。これを確認するため我々は3つの独立した enrichment analysis server [28-30] に30個の遺伝子をアップロードしてみた。詳細は原著論文 [1] に譲るが、非常に多数の gene ontology term (GO term) や KEGG や REACTOME のパスウェイが有意に enrichment されていると同定されただけでなく、3種類の enrichment analysis server 間の一致度

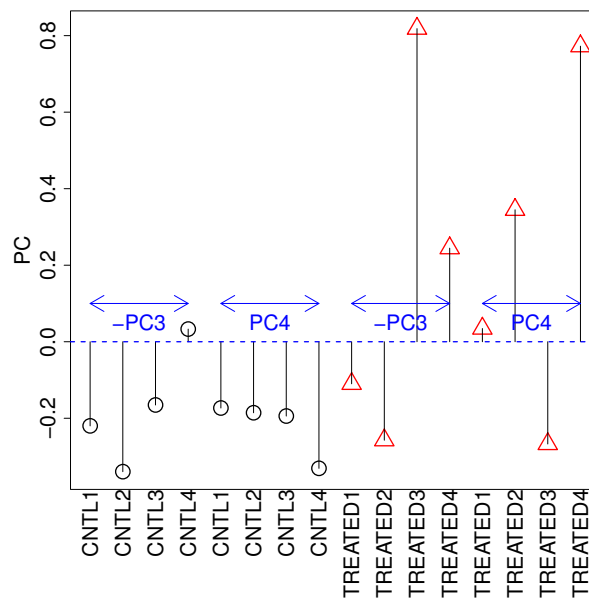


図 3 図 2 に示された第三、四主成分負荷量をまとめて一個の変量と見た場合

Fig. 3 Merged PC loadings of the third and the fourth PC loadings shown in 図 2

も非常に高かった。30個という遺伝子の数は enrichment analysis に用いるには決して多い数ではなく、ヒット数が0であってもやむを得ない少なさである。それに関わらず、非常に多数個の GO term やパスウェイが検出されたことはこの解析方法の妥当性を示している。さらに3つのサーバ間で一致率が高かったこともどのサーバでも重要とみなされる確度の高い遺伝子が30個の遺伝子の中に豊富に含まれているということを示しており、結果の確からしさをサポートするものである。

### 3.3 他の手法との比較

次に他の手法との比較について議論する。本研究では教師なし学習を用いることで既知のサンプルのラベリング (HDACi 使用の有無) をあえて無視している。確かにどの主成分を遺伝子の選択に用いるかという主成分の選定にはラベリング情報を用いてはいるが、主成分負荷量の構成にはラベリングを一切用いていないため、算出された主成分負荷量がラベリングと有意な相関をもっているという保証は皆無である (実際、FTY720 を用いた場合には個々の主成分負荷量には有意なサンプルラベリング依存性は存在しなかった)。このような場合、「ラベリングを使ったほうがよりよい結果が得られるのでは」という考え方は当然にありうる。数万個の遺伝子に対して行われた数個のサンプルの情報から、有意にラベル間差異がある遺伝子を同定することは確かに困難だが、今回の場合はたまたま簡単に同定できる場合だったかもしれない、その場合、わざわざ教師なし学習を使う意味はないし、使わないほうがよい結果を与え

るかもしれない。このような疑念をはらすために、我々は（より少ない数の遺伝子しか検出できなかったしたがってより同定が難しいと思われる）CI-994 の場合を例にとって、いくつかの伝統的な教師あり学習による遺伝子選択を行ってみた。

まず最初に試みたのは普通の  $t$  検定を用いた同定である。この場合、各3個からなる二群の平均値に有意差があるかという厳しい検定になってしまう。案の定、数万個の数を考慮した多重比較補正を算出された  $P$  値に対して施した結果、有意差のある遺伝子は1つも検出されなかった。次に我々はカテゴリ回帰（あるいは別名 ANOVA）を用いた遺伝子選択をおこなってみた。二群を仮定したカテゴリ回帰を実施し、回帰が有意である  $P$  値を検出し、多重比較補正しても有意な  $P$  値をもつ遺伝子を選択してみた。その結果、我々は23個の遺伝子が多重比較補正後も有意な  $P$  値を伴っていることを見出した。この数は主成分分析を用いた教師なし学習で選択された30個よりは少ないが、無視できる数ではない。そこでこの23個の生物学的な妥当性を判定するために、さきの3つの enrichment analysis server にアップロードしてみた。その結果、たった一個の enrichment さえないことがわかった。このことは、たとえ有意な差がある遺伝子が選んでも結果に生物学的な意味があるとはかぎらないという当り前の結果であるが、逆に主成分分析を用いた教師なし学習で選択された30個がどれほど画期的であったかを示している。更に確認として我々は遺伝子発現プロファイルの解析によく用いられる limma [31] も試してみた。limma は数こそ24個とカテゴリ回帰よりわずかに多い遺伝子を検出したが、それらを3つの enrichment analysis server にアップロードしても一個も enrichment が見いだせないのは同じであった。さらに悪いことに、limma とカテゴリ回帰で検出された遺伝子は23個が重なっていた。つまり、教師あり学習を用いる限り、同定された遺伝子は不可避免的に生物学的に意味はないものになってしまう可能性が高いのである。

この様な結論は一見、矛盾しているようにみえるだろう。しかし、全ての教師あり学習はカテゴリ間でもっとも差の大きい遺伝子を選ぶという性質上、不可避免的にカテゴリ内ではばらつきの少ない遺伝子を選ぶことになる。しかし、図1で見るようにカテゴリ内のばらつきは決して無視できる大きさではない。そしてこれらが主成分負荷量として検出される以上、カテゴリ内のばらつき、つまり、HDACi による処理がある物同士、無い物同士の差異は、なんらかの外的な要因（例えば実験手続きの系統的差異）によってもたらされた必然的なものであるかもしれない、ペナルティとして排除すべきものではない可能性がある。教師あり学習はカテゴリ内差異のうち、系統系なものやゆらぎであるものを区別することはできない。この結果、生物学的には必ずしも妥当性のない、カテゴリ内の差異が少ない遺伝子を

恣意的に選んでしまった結果、生物学的な意味が劣ってしまったと考えられるのである。

### 3.4 HDACi の機能

Gräffら [4] は「ヒストン修飾というエピジェネティックな効果が恐怖記憶の確立に関与しており、HDACi はそこに介在することで消去学習の加速に関与している」というストーリーを組み、実験をし、論文を書いた。しかし、論文をよく読むと肝心の HDACi によるヒストン脱アセチル化抑制の効果は、選択された少数個の遺伝子に対して検証されているに過ぎない。実際、今回我々が同定した30個の遺伝子が特に強くヒストンのアセチル化をされているというデータは見いだせなかった。脱アセチル化の抑制、なのだから、もともとヒストンがアセチル化していないのであれば、そもそも HDACi は発現に寄与しようがない。それではなぜ、これらの遺伝子の発現が HDACi の投与によって変化したのだろうか。HDAC というのは人間がつけたただの名前に過ぎず、HDAC がヒストンの脱アセチル化にしか寄与できない、というのは実は勝手な思い込み過ぎない。実際、HDAC の一種である HDAC4 はヒストンの脱アセチル化とは別に多数のタンパクと直接結合すること [32] が知られている。HDACi はあくまで HDAC の機能を阻害する薬剤に過ぎないのであるから、HDACi がヒストンの脱アセチル化ではなく、HDAC4 と他のタンパクとの結合を阻害することで遺伝子発現制御に関与してもおかしくはない。実際、30個の遺伝子のうち8個ほどは実験的に HDAC4 の標的であると目されている遺伝子であった。また、これら30個の遺伝子は複数の転写因子の共通の標的であることも判明した。HDAC4 の結合先には複数の転写因子が含まれており、HDAC4 はこれら転写因子との相互作用を通じて遺伝子の発現に関与していることがわかっている。30個の遺伝子を標的とする転写因子と HDAC4 との結合が実験的に確認されたわけではないが、HDACi は、30個の遺伝子を標的とする転写因子と HDAC4 の結合を阻害することで30個の遺伝子の発現に関与しているという可能性も考えられる。このように HDACi がヒストンの修飾に関与することで消去学習に関与する遺伝子の発現に寄与し、消去学習に貢献していると結論付けるのはやや早計なのではないか、というのが我々の結論である。

## 4. 結言

実験の前に仮説を立て、それに沿って実験を行い、ストーリー通りの結果を得て論文を書くのは美しい。しかし、一方でそれは仮説やストーリーに沿わない結論を排除する危険性も内包する。これに対し、本研究のような、教師なし学習を用いた網羅的な解析はデータ駆動型の仮説提示型のプログラムであるため、当初想定しなかったシナリオの検出も可能とする。科学の正しい進歩のためには後者のよう

な研究ももっと精力的に行われるべきだと考える。

謝辞 本研究は科研費 26120528 でサポートされた。

#### 参考文献

- [1] Taguchi, Y.-H.: Principal component analysis based unsupervised feature extraction applied to publicly available gene expression profiles provides new insights into the mechanisms of action of histone deacetylase inhibitors, *Neuroepigenetics*, Vol. 8, pp. 1 – 18 (online), DOI: <http://dx.doi.org/10.1016/j.nepig.2016.10.001> (2016).
- [2] 伊東杏里: PTSD について, 杏林医学会雑誌, Vol. 47, No. 1, pp. 73–76( オンライン), DOI: 10.11434/kyorin-med.47.73 (2016).
- [3] 山田一之: 遺伝子改変マウスによるストレス性精神障害モデル実験系の構築, 動物心理学研究, Vol. 54, No. 1, pp. 41–47( オンライン), DOI: 10.2502/janip.54.41 (2004).
- [4] Gräff, J., Joseph, N. F., Horn, M. E., Samiei, A., Meng, J., Seo, J., Rei, D., Bero, A. W., Phan, T. X., Wagner, F., Holson, E., Xu, J., Sun, J., Neve, R. L., Mach, R. H., Haggarty, S. J. and Tsai, L. H.: Epigenetic priming of memory updating during reconsolidation to attenuate remote fear memories, *Cell*, Vol. 156, No. 1-2, pp. 261–276 (2014).
- [5] 天野大樹: 恐怖条件付け記憶の想起, 消去に対する扁桃体外核の役割, 日本薬理学雑誌, Vol. 140, No. 1, pp. 14–18 ( オンライン), DOI: 10.1254/fpj.140.14 (2012).
- [6] 松澤大輔: 不安とエビジェネティクス, 心身医学, Vol. 56, No. 4, pp. 333–339( オンライン), DOI: 10.15064/jjpm.56.4.333 (2016).
- [7] 芳原輝之, 内田周作, 山形弘隆, 大舘孝治, 渡邊義文: ヒストン脱アセチル化酵素阻害剤の早期抗うつ効果発現における分子機構の解析, 日本生物学的精神医学会誌, Vol. 23, No. 1, pp. 35–40( オンライン), DOI: 10.11249/jsbpjpp.23.1.35 (2012).
- [8] NCBI: Gene Expression Omnibus, NCBI (online), available from <https://www.ncbi.nlm.nih.gov/geo/> (accessed 2016-11-08).
- [9] Taguchi, Y. H.: microRNA-mRNA Interaction Identification in Wilms Tumor Using Principal Component Analysis Based Unsupervised Feature Extraction, *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 71–78 (online), DOI: 10.1109/BIBE.2016.14 (2016).
- [10] Taguchi, Y. H.: Identification of More Feasible MicroRNA-mRNA Interactions within Multiple Cancers Using Principal Component Analysis Based Unsupervised Feature Extraction, *Int J Mol Sci*, Vol. 17, No. 5, p. E696 (2016).
- [11] Taguchi, Y. H., Iwadate, M. and Umeyama, H.: SFRP1 is a possible candidate for epigenetic therapy in non-small cell lung cancer, *BMC Med Genomics*, Vol. 9 Suppl 1, p. 28 (2016).
- [12] Taguchi, Y. H., Iwadate, M. and Umeyama, H.: Heuristic principal component analysis-based unsupervised feature extraction and its application to gene expression analysis of amyotrophic lateral sclerosis data sets, *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, *2015 IEEE Conference on*, pp. 1–10 (online), DOI: 10.1109/CIBCB.2015.7300274 (2015).
- [13] Taguchi, Y. H., Iwadate, M. and Umeyama, H.: Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for post-traumatic stress disorder-mediated heart disease, *BMC Bioinformatics*, Vol. 16, p. 139 (2015).
- [14] Taguchi, Y. H.: Principal component analysis based unsupervised feature extraction applied to budding yeast temporally periodic gene expression, *BioData Min*, Vol. 9, p. 22 (2016).
- [15] Taguchi, Y. H.: Identification of aberrant gene expression associated with aberrant promoter methylation in primordial germ cells between E13 and E16 rat F3 generation vinclozolin lineage, *BMC Bioinformatics*, Vol. 16 Suppl 18, p. S16 (2015).
- [16] Taguchi, Y. H.: Integrative Analysis of Gene Expression and Promoter Methylation during Reprogramming of a Non-Small-Cell Lung Cancer Cell Line Using Principal Component Analysis-Based Unsupervised Feature Extraction, *Intelligent Computing in Bioinformatics* (Huang, D.-S., Han, K. and Gromiha, M., eds.), Lecture Notes in Computer Science, Vol. 8590, Springer International Publishing, Heidelberg, pp. 445–455 (2014).
- [17] Taguchi, Y. H., Iwadate, M., Umeyama, H., Murakami, Y. and Okamoto, A.: Heuristic principal component analysis-based unsupervised feature extraction and its application to bioinformatics, *Big Data Analytics in Bioinformatics and Healthcare* (Wang, B., Li, R. and Perrizo, W., eds.), pp. 138–162 (2015).
- [18] Umeyama, H., Iwadate, M. and Taguchi, Y. H.: TINAGL1 and B3GALNT1 are potential therapy target genes to suppress metastasis in non-small cell lung cancer, *BMC Genomics*, Vol. 15 Suppl 9, p. S2 (2014).
- [19] Murakami, Y., Kubo, S., Tamori, A., Itami, S., Kawamura, E., Iwaisako, K., Ikeda, K., Kawada, N., Ochiya, T. and Taguchi, Y. H.: Comprehensive analysis of transcriptome and metabolome analysis in Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma, *Sci Rep*, Vol. 5, p. 16294 (2015).
- [20] Murakami, Y., Tanahashi, T., Okada, R., Toyoda, H., Kumada, T., Enomoto, M., Tamori, A., Kawada, N., Taguchi, Y. H. and Azuma, T.: Comparison of Hepatocellular Carcinoma miRNA Expression Profiling as Evaluated by Next Generation Sequencing and Microarray, *PLoS ONE*, Vol. 9, No. 9, p. e106314 (2014).
- [21] Murakami, Y., Toyoda, H., Tanahashi, T., Tanaka, J., Kumada, T., Yoshioka, Y., Kosaka, N., Ochiya, T. and Taguchi, Y. H.: Comprehensive miRNA expression analysis in peripheral blood can diagnose liver disease, *PLoS ONE*, Vol. 7, No. 10, p. e48366 (2012).
- [22] Taguchi, Y. H. and Murakami, Y.: Universal disease biomarker: can a fixed set of blood microRNAs diagnose multiple diseases?, *BMC Res Notes*, Vol. 7, p. 581 (2014).
- [23] Taguchi, Y. H. and Murakami, Y.: Principal component analysis based feature extraction approach to identify circulating microRNA biomarkers, *PLoS ONE*, Vol. 8, No. 6, p. e66714 (2013).
- [24] Kinoshita, R., Iwadate, M., Umeyama, H. and Taguchi, Y. H.: Genes associated with genotype-specific DNA methylation in squamous cell carcinoma as candidate drug targets, *BMC Syst Biol*, Vol. 8 Suppl 1, p. S4 (2014).
- [25] Ishida, S., Umeyama, H., Iwadate, M. and Taguchi, Y. H.: Bioinformatic Screening of Autoimmune Disease Genes and Protein Structure Prediction with FAMS for Drug Discovery, *Protein Pept. Lett.*, Vol. 21, No. 8, pp. 828–39 (2014).
- [26] Taguchi, Y. H. and Okamoto, A.: Principal Component

- ment Analysis for Bacterial Proteomic Analysis, *Pattern Recognition in Bioinformatics* (Shibuya, T., Kashima, H., Sese, J. and Ahmad, S., eds.), Vol. 7632, Springer International Publishing, Heidelberg, pp. 141–152 (2012).
- [27] C.M. ビシヨップ: パターン認識と機械学習 上, 丸善出版 (2012).
- [28] Huang, d. a. W., Sherman, B. T. and Lempicki, R. A.: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat Protoc*, Vol. 4, No. 1, pp. 44–57 (2009).
- [29] Reimand, J., Arak, T. and Vilo, J.: g:Profiler—a web server for functional interpretation of gene lists (2011 update), *Nucleic Acids Res.*, Vol. 39, No. Web Server issue, pp. W307–315 (2011).
- [30] Chen, Y. A., Tripathi, L. P. and Mizuguchi, K.: TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery, *PLoS ONE*, Vol. 6, No. 3, p. e17844 (2011).
- [31] Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W. and Smyth, G. K.: limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, Vol. 43, No. 7, p. e47 (2015).
- [32] Sando, R., Gounko, N., Pieraut, S., Liao, L., Yates, J. and Maximov, A.: HDAC4 governs a transcriptional program essential for synaptic plasticity and memory, *Cell*, Vol. 151, No. 4, pp. 821–834 (2012).