

# 主成分分析を用いた教師なし学習による変数選択の Deng 出血熱患者血液遺伝子発現プロファイル解析への応用

田口 善弘<sup>1,a)</sup>

概要: 温暖化によって Deng 熱を媒介する蚊の日本での越冬が危惧されている。Deng 熱は通常、重篤には至らないが、まれに出血熱という致死性の疾患に発展することがある。出血熱は解熱後に起きるため、感染ではなく、患者の Deng 熱ウイルスに対する複雑な反応の結果起きると思われているが原因は不明である。本研究では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ることで、少数個の遺伝子で Deng 熱/Deng 出血熱の進行を記述できることを見出したので報告する。

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

## 1. はじめに

Deng 熱は蚊によって媒介される Deng ウイルスに感染することで発症する感染症の一種であるが熱帯病であるため、日本国内ではあまり問題になっておらず、まれに海外で感染した患者が発生したり、あるいは、飛行機で持ち込まれた蚊に刺されて感染したりする例 [2] があるのみであった。しかし、地球温暖化に伴って国内でも Deng ウイルスを媒介する蚊が越冬して繁殖する可能性が無視できなくなりつつあり、問題化しつつある。

Deng 熱は通常、命に係るような疾患ではないが、まれに Deng 出血熱 [3] という全身から出血して死に至る病態に発展する場合があります。通常、Deng 出血熱は解熱後に発症するため、Deng ウイルスの直接の作用というより、感染による宿主である人間の複雑な生体反応の結果起きるとみなされているが原因はよくわかっていない。

本研究では Deng 熱に感染して発症し、Deng 出血熱の発症に至った患者と至らなかった患者の血液の遺伝子発現プロファイルを再解析し、どのような遺伝子が Deng 出血熱発症の鍵となっているかを解明することを試みた。その結果、比較的少数個の遺伝子の発現で Deng 出血熱へと至る/至らない場合の病気の進行を記述できるという結果を得

た [1] のでここにそれを報告したい。

## 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 [4] 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、5 つのプロファイル、GSE5180833, GSE1305234, GSE2500135, および GSE43777-GPL570/201 を用いて解析を行った (以下、これらを順にデータセット 1 から 5 と呼ぶ)。

### 2.2 主成分分析を用いた教師なし学習による変数選択

5 つの発現プロファイルを別々に解析した。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 [5–22] 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

<sup>1</sup> 中央大学工学部物理学科  
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

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、発熱のある無しで差がある成分を選んだ。

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、発熱のある無しによ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 [23]、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 データセット 1 & 2

最初にデータセット 1 と 2 をそれぞれ独立に解析した(図 1)。図の様に第 2, 3 主成分負荷量が患者を発熱のある無しで二分できていることがわかった。そこで第 2, 3 主成分得点を用いて遺伝子に補正された  $P$  値を計算し、0.01 以下の遺伝子を選択したところ、それぞれ 879 個と 275 個の遺伝子が選択された。これらの遺伝子のうち、共通に選択されたものは全部で 46 遺伝子であった(表 1)。遺伝子の総数は 2 万個近くあり、数百個はその数%にすぎない。にも関わらず 46 個もの遺伝子が共通に選ばれる可能性はとても低い。遺伝子の選択に使われた第 2, 3 主成分得点の寄与率が数%以下であるため、信頼性に疑問を抱く

表 1 主成分分析を用いた教師なし学習による変数選択でデータセット 1 & 2 で共通に選ばれた 46 遺伝子

Table 1 Forty-six genes identified with PCA-based unsupervised FE between datasets 1 and 2.

<i>FBXO7</i>	<i>MX1</i>	<i>LY6E</i>	<i>IFI27</i>	<i>TNFSF10</i>
<i>OAS1</i>	<i>CDC20</i>	<i>GYPC</i>	<i>PI3</i>	<i>FCGR3A</i>
<i>HBA1</i>	<i>HBA2</i>	<i>HBG1</i>	<i>HBG2</i>	<i>IFI44L</i>
<i>IFIT3</i>	<i>CCR1</i>	<i>FPR1</i>	<i>STAT2</i>	<i>ISG15</i>
<i>OASL</i>	<i>CD38</i>	<i>TNFRSF17</i>	<i>CXCR1</i>	<i>ZBP1</i>
<i>HBB</i>	<i>IFI35</i>	<i>MKRN1</i>	<i>APOBEC3A</i>	<i>ALAS2</i>
<i>IL1RN</i>	<i>RSAD2</i>	<i>ASCC2</i>	<i>IFIT2</i>	<i>ADIPOR1</i>
<i>SLC25A37</i>	<i>OAS3</i>	<i>SDF2L1</i>	<i>TMEM140</i>	<i>FKBP11</i>
<i>HERC5</i>	<i>ITM2C</i>	<i>TXNDC5</i>	<i>STRADB</i>	<i>SLC25A39</i>
<i>EPSTI1</i>				

かもしれないが、全く独立の2つのデータセットでこれだけの一致度を見ていることから見ても遺伝子選択は成功していると見るべきだろう。むしろ、第 2, 3 主成分得点の寄与率の少なさは、患者の発熱のある無しを特徴づける遺伝子が数百個しか無いという結果と整合的だ、と見るべきだろう。46 遺伝子の選択が妥当であることを確認するため、この 46 遺伝子だけを使ってデータセット 1 と 2 を主成分分析した結果が図 2 である。発熱のある無しで患者が分かれるという配置は保存されており、遺伝子の選択は正しかったらしいとわかる。更に不完全ながら Deng 熱と Deng 出血熱の患者の区別もできていることもわかる。

#### 3.2 データセット 3 & 4 & 5

選択された 46 遺伝子が Deng 熱/Deng 出血熱の進行の記述に本質的であることを確認するために、更に独立な 3 つのデータセット(3 & 4 & 5)を用いて、46 遺伝子「だけ」で主成分分析を行ってみた(図 3)。これら 3 つのデータセットは、データセット 1 & 2 とは異なり、より細かい時系列で Deng 出血熱に至るプロセスがサンプリングされている。一見してわかるのは 3 つのデータセットで極めて類似した V 字型の軌跡を病気の進行がたどっていることである。このことから、データセット 1 & 2 から選択された 46 遺伝子は、全く独立なデータセット 3 & 4 & 5 についても病気の進行をよく記述する遺伝子だということがわかるだろう。更に特にデータセット 3 において顕著だが、これら 46 遺伝子は病気の治癒(解熱)に伴って、Deng 熱と Deng 出血熱の分離を明確に記述していることがわかる。このことから、46 遺伝子の選択は Deng 熱/Deng 出血熱の進行において極めて重要なキーとなる遺伝子であるということが明らかであろう。この様に主成分分析を用いた教師なし学習による変数選択は、対象となるサンプル群の特徴をよく記述する少数個の遺伝子の選択に優れているのである。

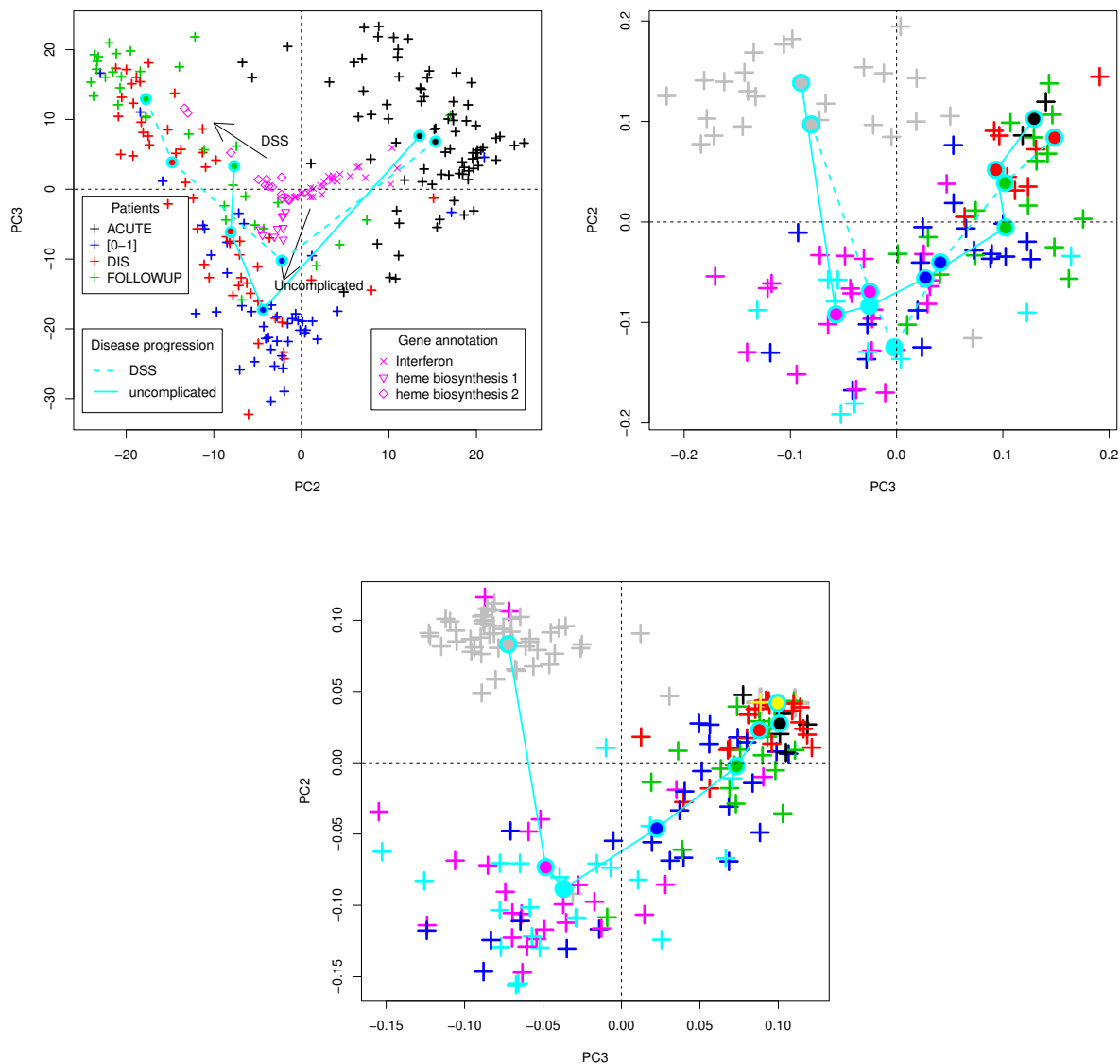


図 3 表 1 にある 46 遺伝子だけを用いて主成分分析した場合のバイプロット。上左: データセット 3。黒い+: 急性期、青い+: 発病後 0 から 1 日、赤い+: デング熱、緑の+: 治癒後。水色の実線と点線はデング熱になったかデング出血熱になったか。丸は+の重心位置。ピンクの記号は選択された 46 遺伝子。上右: データセット 4。+ は黒→赤→緑→青→水色→ピンク→グレーの順に病気の進行を示す。グレーは治癒後。実線と点線はデング熱になったかデング出血熱になったか。丸は+の重心。下: データセット 5。記号の意味はデータセット 4 と同じだが、デング出血熱のサンプルは含まれていない。

**Fig. 3** Biplots obtained by PCA using 46 genes in Table 1. upper left: data set 3. Black crosses: Acute, blue crosses: days 0 to 1, red crosses: disease, green crosses: convalescent patients. Magenta characters are 46 genes. Solid and broken lines are DF and DHF, respectively. Filled circles are center of mass. upper right: data set 4. black, red, green, blue, cyan, magenta and gray crosses represents disease progression (gray ones are convalescent). Solid and broken lines are DF and DHF, respectively. Filled circles are center of mass. lower: dataset 5. the same as dataset 4 excluding that it does not include DHF.

### 3.3 他の手法との比較

他の手法で同じことができないかを確認した。比較手法としては遺伝子発現プロファイルから遺伝子の選択を行う

のごく一般的な手法である SAM [24] と limma [25] を採用した。詳細は原著論文 [1] に譲るが、ごく簡単に言ってしまうと、データセット 1 & 2 に関する限り、補正した  $P$  値

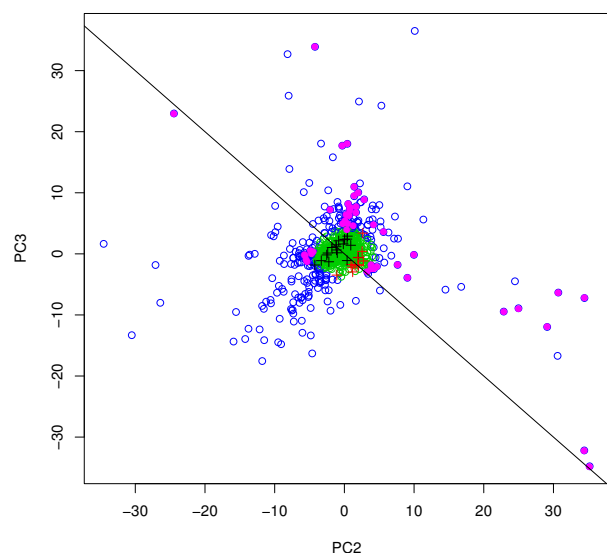
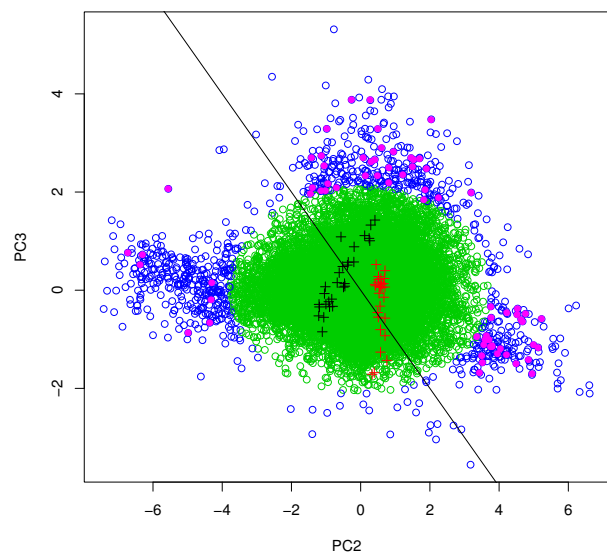


図 1 主成分分析の結果(第2, 3主成分によるバイプロット)。上: データセット1 (第2, 3主成分得点の寄与率はそれぞれ1.45%と0.45%)、下: データセット2 (第2, 3主成分得点の寄与率はそれぞれ3.39%と2.90%)。緑の○: 選択されなかった遺伝子。青の○: 選択された遺伝子。ピンクの○: データセット1と2で共通に選択された遺伝子。黒の+: 発熱のある患者、赤の+: 発熱の無い患者。

Fig. 1 Principal component analysis (PCA) applied to dataset 1 (upper: PC2 and PC3 have 1.45% and 0.45% contributions, respectively.) and 2 (lower: PC2 and PC3 have 3.29% and 2.90% contributions, respectively.). Biplots using the second and the third PCs. Green and blue open circles are not selected and selected genes, respectively. Magenta circles are genes selected commonly in both datasets 1 and 2. Black and red crosses are patients with and without symptom (fever), respectively.

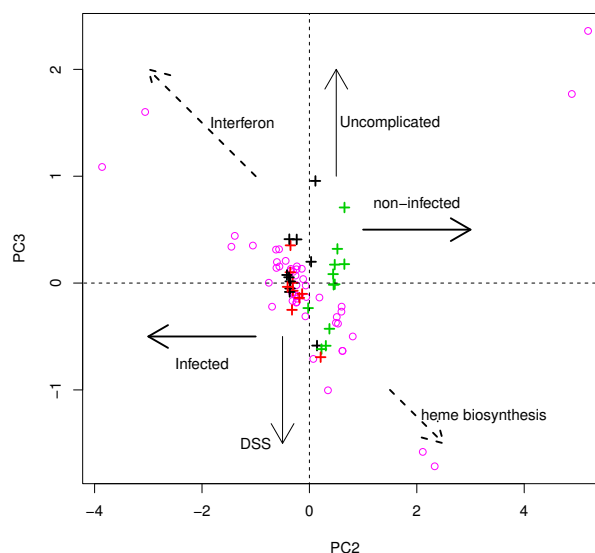
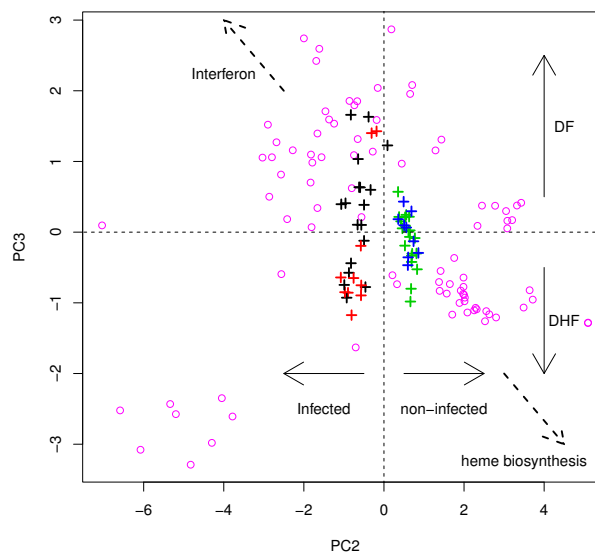


図 2 表 1 にある46 遺伝子だけを用いて主成分分析した場合のバイプロット。上: データセット1, 下: データセット2。赤い+: デング出血熱患者。黒い+: デング出血熱以外のデング熱患者。青の+: 健康者。緑の+: 治療後の患者。ピンクの丸は選択された46 遺伝子。

Fig. 2 Biplots obtained by PCA using 46 genes in Table 1. Upper: data set 1, lower: data set 2. red crosses: DHF, black crosses: DF, blue crosses: healthy controls, green crosses: convalescent patients. Magenta circles: selected 46 genes.

が0.01以下という同じ基準だけでは、非常に多数個の遺伝子が選ばれてしまい、主成分分析を用いた教師なし学習による変数選択の場合のように少数個の遺伝子がデータセット1 & 2で共通に選ばれるという結果は出せなかった。通常、このような場合には更にいわゆる fold change を考慮し

て変数選択を行って遺伝子を絞り込むことが多いが、そもそも今回のデータセットは単純な2群ではなく(発熱のある無し以外に、 Deng 出血熱になったかどうかという区別もある) fold change で絞り込むと言っても単純にはできない。いずれにせよ、 $P$  値だけで十分な絞り込みができ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に比べて SAM や limma が変数選択の性能的に劣った手法なのは明らかだろう。

しかし、それでも更に慎重を記するならば以下のような2つの反論はありうる。

- SAM や limma で  $P$  値のしきい値をもっとずっと小さくすれば絞り込みは行えるので主成分分析を用いた教師なし学習による変数選択と同じようにデータセット 1 & 2 で共通な十分少数個の遺伝子が選択できるのではないか。
- SAM や limma でより多くの遺伝子が選択できるのは悪いことではなく、むしろ、主成分分析を用いた教師なし学習による変数選択に比べてこれらの手法が高性能すぎるだけなのではないか。例えば、サンプル数を意図的に半減させた場合、SAM や limma ではしかるべく少数個の遺伝子が選択される一方、主成分分析を用いた教師なし学習による変数選択では補正された  $P$  値が 0.01 以下の遺伝子が減ってしまうのではないか。

これらの批判は実際、原著論文投稿時に担当エディタによって提起された。これらの批判は、しかし、余り妥当なものであるとは言えない。第一点については、そもそも、補正された  $P$  値は擬陽性の割合を示していると理解されている。選択された遺伝子が数百個の場合に、補正された  $P$  値が 0.001 とか 0.0001 とかであることを要求することはそもそも、擬陽性の数を 1 以下にすることに相当するので無意味である。また、第二点についても、そもそも補正された  $P$  値はサンプルサイズの影響を除去した値であるはずであり、サンプルサイズを変えたら選択される遺伝子の数が変わってしまうようでは、 $P$  値の補正がうまく行っていないということと同義である。だから、仮に第二点が生じたとしてもそれは SAM や limma が主成分分析を用いた教師なし学習による変数選択に比べて優れているという結論には全く結びつかない。

しかし、念の為これらの点についても検討を行った。まず第一点について。SAM と limma のそれぞれについて、 $P$  値が小さい順という基準で、主成分分析を用いた教師なし学習による変数選択と同じ数をデータセット 1 & 2 で無理やり選んでどれくらいオーバーラップがあるものか調べることを試みた。まず、SAM についてはそもそも、それ以上の数の遺伝子に確率 0 が付与されており、主成分分析を用いた教師なし学習による変数選択と同じ数の上位の遺伝子を選択すること自身が無理であった。limma については主成分分析を用いた教師なし学習による変数選択と同じ数

の遺伝子を選択することは可能だったが、その場合データセット 1 & 2 で一、二個の遺伝子しか共通に選べなかった。これは極端に小さな補正  $P$  値には統計的な意味はないだろう、という予想通りの結果であり、第一点の危惧は考慮に値しないことがわかった。また、第二点については、試行的にサンプル数を半分にして、SAM、limma、及び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て補正した  $P$  値が 0.01 以下という条件で何個の遺伝子が選択されるのかを、半数のサンプルの選択をランダムに何度も繰り返して平均を取ってみた。SAM や limma の場合、サンプル数が半分になると補正した  $P$  値が 0.01 以下という基準ではあまり数が変わらなかったり(つまり、多すぎる数が選択されたまま)であったり、あるいはいきなり有意な差がある遺伝子の数が 0 になったりでうまくは行かなかった。これはそもそも補正された  $P$  値がサンプル数の効果を打ち消すように補正されていることを考えれば妥当な結果だろう。一方、主成分分析を用いた教師なし学習による変数選択の場合にはサンプル数を半分にしても選択される遺伝子数はほぼ変わらなかった。これは奇妙に思えるかもしれないが、主成分分析を用いた教師なし学習による変数選択の場合には  $P$  値が遺伝子に付与された主成分得点に対して計算されていることを考えるとそれほど不思議ではない。主成分分析の観点からは、サンプル数をランダムに半減させることは、高次元空間からランダムに半分の次元数の低次元空間を選んでから主成分分析を行っているのと同じである。元の次元(=サンプル数)が十分に大きければランダムに次元を半減させても遺伝子に付与される主成分得点が極端に変わることは考えにくい。となれば補正  $P$  値の値があまり変わらなくても驚くには値しない。したがって、第二点の反論も否定されたと言えるだろう。

#### 4. 結言

本報告では Deng 出血熱の進行を記述する少数個の遺伝子を選択するという難しい問題に主成分分析を用いた教師なし学習による変数選択が非常に有効であること、また、伝統的な教師あり学習に基づく変数選択法である SAM や limma が主成分分析を用いた教師なし学習による変数選択に大きく劣った性能しか出せないことを示した。主成分分析を用いた教師なし学習による変数選択は今までもいろいろな問題に非常に有効であることが示されており、もっといろいろな問題で広く活用されるべきであると考えている。

原著論文 [1] では人工データによるデモンストレーションや、より生物学的な議論も行ったが本稿では割愛した。興味ある読者は原著論文 [1] を参照頂きたい。

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 Components Analysis Based Unsupervised Feature Extraction Applied to Gene Expression Analysis of Blood from Dengue Haemorrhagic Fever Patients, *Sci Rep*, Vol. 7, p. 44016 (2017).
- [2] 石川洋文, 下河原理江子, 笛田 薫: 代々木公園ではどのように 2014 年デング熱アウトブレイクが進展したのか? —数理モデルによる解析—, *日本衛生学雑誌*, Vol. 72, No. 1, pp. 55–65( オンライン), DOI: 10.1265/jjh.72.55 (2017).
- [3] 倉根一郎: デング出血熱の病態解明の進展とワクチン開発の現状, *ウイルス*, Vol. 52, No. 1, pp. 15–20( オンライン), DOI: 10.2222/jsv.52.15 (2002).
- [4] NCBI: Gene Expression Omnibus, NCBI (online), available from (<https://www.ncbi.nlm.nih.gov/geo/>) (accessed 2016-11-08).
- [5] 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).
- [6] 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).
- [7] 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).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 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).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] 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).
- [21] 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).
- [22] Taguchi, Y. H. and Okamoto, A.: Principal Component 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).
- [23] C.M. ビシヨップ: パターン認識と機械学習 上, 丸善出版 (2012).
- [24] Schwender, H.: *siggens: Multiple testing using SAM and Efron's empirical Bayes approaches* (2012).
- [25] 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).