

# 単変量 Cox 回帰にもとづく遺伝子選択と

## 複合共変量による生存期間の予測

国立中央大学（台湾）、統計研究所 江村剛志

名古屋大学大学院医学系研究科 松井茂之

Institute of Statistical Science, Academia Sinica Hsuan-Yu Chen

### 摘要

遺伝子発現量が、がん患者の生存期間の予後予測に有用であることは広く報告されている。遺伝子発現量と生存期間の関連を調べる研究において、単変量 Cox 回帰を使った遺伝子選択法を用いる予後予測法が生物医学研究者の間で広く利用されてきた。興味深いことに、単変量 Cox 回帰分析の結果を線形結合した複合共変量 (Compound covariate) 法は、リッジ回帰法や Lasso 法などの罰則付き多変量 Cox 回帰法に匹敵する予後予測性能を持つことが報告されている (Emura, Chen and Chen 2012)。また複合共変量は、遺伝子選択を取り入れた臨床試験のデザインに組み込むこと (Matsui et al. 2012; Matsui 2015) や、患者レベルの個別化予測を実現することが報告されており (Emura et al. 2017)、個別医療の発展に貢献することが期待される。しかしながら、単変量 Cox 回帰分析による最適な遺伝子選択法、False Discovery Rate (FDR) の計算法、選択された遺伝子の予測能力の評価を行う方法などを実行するためのソフトウェアは開発されていない。

本稿では、単変量 Cox 回帰分析にもとづく遺伝子選択法のいくつかの方法論を概説し、それら統計手法を実行する R パッケージ「*compound.Cox*」(Emura, Matsui and Chen 2018-) を紹介する。肺癌データ (Chen et al. 2007) を用いて遺伝子選択法を説明し、False discovery rate (FDR) の計算やクロスバリデーションによる予測能力の評価を行う統計手法を紹介する。また選択された遺伝子を使用した予後予測の手法、とりわけコピュラ (Emura and Chen 2016; 2018) や縮小推定法 (Emura et al. 2012) などの高度な手法も紹介する。本稿では審査中の原論文 (Emura, Matsui and Chen 2018-) の内容の一部を日本語で解説し、追加の解説も与えている。本稿では R パッケージ「*compound.Cox*」のプログラムの一部の使用例や実装アルゴリズムの考え方を提供するが、より詳細な議論については原論文を参照されたい。

**キーワード** ; 遺伝子発現量、高次元データ解析、コピュラ、個別医療、生存時間解析、多重検定、特徴選択、予後予測

---

責任著者 ; 江村剛志 (Takeshi Emura); Email: [takeshiemura@gmail.com](mailto:takeshiemura@gmail.com)

Graduate Institute of Statistics, National Central University,  
Zhongda Road, Zhongli District, Taoyuan 32001, Taiwan

## 1. 研究の背景

単変量 Cox 回帰分析にもとづく遺伝子選択法は医学研究においてよく広く用いられている手法である。個々の遺伝子と生存時間の関連性を単変量 Cox 回帰分析により調べ、有意性の高い順に遺伝子を選択していく単純な手法である (Witten and Tibshirani 2010)。具体的には、すべての遺伝子に対して生存期間との関連性の有無を検定し、ある閾値 ( $P$  値  $< 0.05$  や  $P$  値  $< 0.001$ ) 以上の有意性を持つ遺伝子を複数選択する。

単変量 Cox 回帰分析によって選択された複数の遺伝子を使用して構築した予測式が、癌患者の予後予測に有効であることは多数報告されている (Beer et al. 2002; Chen et al. 2007; Yoshihara et al. 2010 など)。予測式を構築する際には、予測に組み込む遺伝子数を設定する必要がある。固定された  $P$  値の閾値 ( $P$  値  $< 0.001$  など) をもとに選択する遺伝子数を設定する手法 (Simon 2003; Chen et al. 2007; Matsui et al. 2012; Emura et al. 2017) が医学研究において一般的に用いられている。得られたデータにクロスバリデーションを行い最適な閾値を求める統計手法 (Matsui 2006; Bøvelstad et al., 2007; van Wieringen et al. 2009) なども提案されている。

単変量 Cox 回帰を用いるということは、個々の遺伝子が単一でどれだけ生存期間に関与しているかによって、遺伝子選択をしていることになる。多変量 Cox のモデルのように、複数の遺伝子が全体でどれだけ生存期間に関与しているかによって、遺伝子選択を行う手法と異なる点に注意する。多変量 Cox モデルは遺伝子間の相関を考慮できるという利点があるが、多変量回帰モデルが単変量回帰モデルに比べて必ずしも優れているというわけではない。例えば、Witten and Tibshirani (2010) のデータ解析例では、単変量 Cox 回帰によって選ばれた 4 つの遺伝子を使用した予測式が、多変量回帰モデルによって構築された複雑な予測式より高いことが示されている。

興味深いことに、単変量 Cox 回帰分析の結果を線形結合した複合共変量 (Compound covariate) 法は、リッジ回帰法や Lasso 法などの罰則付き多変量 Cox 回帰法に匹敵する予測精度を持つことが報告されている (Emura, Chen and Chen, 2012)。このような単変量回帰にもとづく手法が多変量回帰にもとづく手法より優れているケースは高次元変数を扱うさまざまな統計手法の中で示されている (Dudoit, Fridlyand and Speed, 2002 など)。また複合共変量は、遺伝子選択を含む臨床試験のデザインに組み込むこと (Matsui et al. 2012; Matsui 2015) や、患者レベルの個別化予測を実現することが報告されており (Emura et al. 2017)、個別医療の発展に貢献することが期待される。しかしながら、単変量 Cox 回帰による最適な遺伝子選択、False Discovery Rate (FDR) の計算、選択された遺伝子の予測能力の評価などをするための包括的なソフトウェアは開発されていない。

## 2. 単変量 Cox 回帰にもとづく遺伝子選択

本節では単変量 Cox 回帰にもとづく遺伝子選択法と、選択された遺伝子を用いた生存期間の予後予測手法の概要を与える。

### 2.1 P 値の閾値

医学研究者は伝統的な P 値の閾値である 0.05 (Chen et al. 2007) や 0.01 (Rosenwald et al. 2002; Yau et al. 2010; Yoshihara et al. 2010) を用いる傾向がある。Simon (2003) は P 値の閾値 0.001 を高次元遺伝子解析で提案している。またこの閾値 0.001 は生存期間データにより遺伝子選択を行う際に用いられている (Vasselli et al. 2003; Matsui et al. 2012; Emura et al. 2017)。このような固定された P 値の閾値は便利な遺伝子選択基準であるが、選択された遺伝子が全体として予測に有効かどうかの保証や、閾値 0.001 の数理的妥当性はなく、閾値を決めるガイドラインの 1 つに過ぎない。

予測に最適な遺伝子選択を行うため、Matsui (2006) は単変量 Cox 回帰法に適した cross-validated likelihood (CVL) を定義し、CVL 値を最適化する閾値をデータから求めることを提案している。Matsui (2006) はこの手法を応用して、最適な数の遺伝子を選択した後、複合共変量法で予後予測法を構築することを提案している。

似たような CVL 基準が Bøvelstad et al. (2007) によって提案されているが、こちらは多変量 Cox 回帰法に適した CVL 値であるため、単変量 Cox 回帰によって選ばれた遺伝子の予測性能を評価するのに適さない。実際、Bøvelstad et al. (2007) の数値実験では、CVL 値を最適化する遺伝子は数個ほどしか選ばれず、その予測性能もリッジ回帰法 や Lasso 法よりもだいぶ劣るといった結果になった。このような誤った CVL 値の使用は、Matsui (2006) で提案された CVL の計算がかなり複雑であり、下記で考察するクロスバリデーションアルゴリズムと合わせて再考、再定義が必要であることを示している。以下で、用いるべきクロスバリデーションの基本的な考え方を紹介する。

### 2.2 クロスバリデーション

選択された複数の遺伝子から成る予後予測式の予測性能をクロスバリデーションを用いて厳密に評価することが提案されている (Simon 2005)。クロスバリデーション法では、まず  $n$  個の標本を  $K$  個のグループにランダムかつ均等に分ける。次に、1 つのグループの標本を除いた  $n - n/K$  個の訓練標本を用いて、遺伝子選択を行い、予後予測式を構築する。その予測式の予測精度を、除いていた 1 つのグループの標本 (テスト標本) をもとに測定する。これを全てのグループに対して行い、予測精度の指標を  $K$  個のグループ全てで平均化し、予測式の全体としての正確さの指標を計算する。

上述のクロスバリデーションによって得られた予測精度の解釈を十分に理解することが重要である。各グループごとに、遺伝子選択と予測式の構築のプロセスが独立に行われるため、1つのグループで選択される遺伝子は、全標本を使って選択される遺伝子と全く異なる場合がある。すなわち、全てのグループが異なる遺伝子を用いて予測式を構築し、それらの予測性能を計算していることになる。結果として、クロスバリデーションでは、全標本で選ばれた遺伝子の予測性能を評価しているのではなく、与えられた遺伝子選択アルゴリズムの予測性能を評価していることになる。

### 2.3 複数遺伝子による予測式

通常、予測式は選択された遺伝子の発現量の重み付き和（線形結合）として定義される。医学研究において、予後が悪いことは高リスクであることに対応するので、予測式の値が大きいほど高リスクになるように重みを付けるのが通常である。Cox 回帰分析を使用している場合、予測式の値が大きいほど、ハザードの値が大きくなる（高リスクになる）ような重み付けを考える。

遺伝子選択の際に使用した単変量検定の結果を再利用して重みを計算した予測式を、複合共変量（Compound covariate）という。複合共変量は元来 Tukey（1993）が臨床試験における治療群と対照群の間の共変量分布のバランスをとるための手法として提唱された。複合共変量は単変量解析の“複合”であるため、多変量回帰モデルを介さずに、遺伝子選択後にダイレクトに予測式を構築できるという点が面白い。複合共変量法は Tukey の元来の意図から離れ、高次元遺伝子発現量を利用した生存期間の予測に有用であることが多数報告されてきた（Beer et al. 2002; Wang et al. 2005; Matsui 2006; Chen et al. 2007; Emura et al. 2012, 2017; Matsui et al. 2012; Waldron et al. 2014, Zhao et al. 2014）。複合共変量法は単変量回帰を利用した簡便な手法であるため、生物医学研究者の間では、“複合共変量”の名を使用せずに利用されてきた。実際、単変量 Cox 回帰にもとづく遺伝子選択法で選ばれた遺伝子の予後予測手法として、複合共変量法を考えることは極めて自然である。

共線性の考慮の理由などから、多変量 Cox 回帰モデルを当てはめて不必要な遺伝子を除くことが医学研究では広く用いられている（Lossos et al. 2004 など）が、このような遺伝子“再選択”の手法の理論的、数理的、数値的な裏付けはなされていない。単変量解析で選ばれた遺伝子に、多変量 Cox 回帰を当てはめて、有意でない遺伝子を除いた予測式を構築すると予測性能は悪くなるようである（Bøvelstad et al. 2007）。単変量解析で選ばれた遺伝子に、多変量 Cox 回帰モデルを当てはめる場合は、リッジ回帰法（Yoshihara et al. 2010）や Compound shrinkage 法（Emura, Chen and Chen 2012）などの縮小推定法で回帰係数を推定することが推奨される。Emura, Matsui and Chen (2018-) のシミュレーション解析では、

複合共変量法、リッジ回帰法、Compound shrinkage 法がどれも優れた予測能力を示したが、Lasso 法は前述の 3 手法よりやや予測能力が劣っていた。

## 2.4 独立でないセンサーへの対処

Cox 回帰分析は次の 2 つの仮定の下で妥当な手法である。

- (1) 生存時間と遺伝子発現量の関係が Cox 比例ハザードモデルに従う。
- (2) 生存時間とセンサー時間が条件付き独立である（遺伝子を与えた条件で）。

仮定 (2) は“独立センサー”の仮定として知られており、Cox 回帰分析だけでなく、Kaplan-Meier 法、Log-rank 検定法、などほとんどの生存時間解析の手法で必要とされる重要な仮定である。Emura and Chen (2016) は独立センサーの仮定 (2) がモデルの仮定 (1) より重要であることを解析的に示している。とくに、脱落 (dropout) によって引き起こされるセンサーは生存時間と深く関連しているため、システムティックなバイアスが入るので注意が必要である (Emura and Chen 2018)。仮定 (2) が成立しない場合、Cox 回帰分析 (スコア検定、ワルド検定を含む) の結果にバイアスが生じる。この“従属センサー”に起因するバイアスに対応する多くの手法が生物統計学者の間で広く議論されている (Rivest and Wells 2001; Chen 2010; Collett 2015; Emura and Chen 2016; 2018; Emura and Michimae 2017; Staplin et al. 2015; Moradian et al. 2017)。遺伝子選択の際に、従属センサーがみられる場合、コピュラモデルを使用してバイアスに対応した手法 (Emura and Chen 2016) がより正確に遺伝子を選択することが示されている。しかしながら、この Copula を使用する手法は計算に膨大な時間がかかるために、通常の単変量 Cox 回帰法で遺伝子選択した後に、予測式の再計算に用いることが推奨されている (Emura, Matsui and Chen 2018-)。

## 3. パッケージ「*compound.Cox*」による遺伝子選択

上に述べた単変量 Cox 回帰にもとづく遺伝子選択法は簡潔な手法ではあるが、計算面において考慮すべき要因がいくつかあり、追加の計算アルゴリズムが必要になることがある。例えば、どの検定統計量を使って P 値を計算するのか決める必要がある。通常の Cox 回帰分析において、ワルド検定、スコア検定、尤度比検定の結果に間にあまり差は無く、一般的に尤度比検定の小標本性能が良いとされている。遺伝子などの高次元変量に対して多重検定を行う場合、尤度比検定はほとんど使用されず、ワルド検定かスコア検定のいずれかを使用する。

スコア検定の計算式が極めて単純な代数であることは、高次元変量下での膨大な計算量の観点から留意すべきである (Witten and Tibshirani 2010)。またワルド検定から導出される P 値は Cox モデルの回帰係数  $\beta$  の推定値と直結しているため、抽象的に  $\beta=0$  の妥当性を検定するスコア検定より理解が容易であるという利点がある。Emura, Mastui and Chen (2018-) はスコア検定を全遺伝子に実行するための単純な行列計算法を提案しており、この計算法を用いて高次元変量下での計算時間の大幅な短縮が可能である。パッケージ「*compound.Cox*」のスコア検定ではこの行列計算法を実装した。

検定統計量を決めたとしても、次に閾値をどのように定めるか (固定 P 値かデータ適合 P 値) という問題がある。医学研究における実践において、ある研究者はいろいろな P 値の閾値を試して、どの P 値で選択した遺伝子の集まりが予測モデルの構築に便利なのかを考えるかもしれない。また、ある研究者は、予測モデルの構築よりも、予測能力のある新たな遺伝子を発見するために全遺伝子に有意性検定を行いたいと考えるかもしれない。前者では、選択された遺伝子の予測能力の評価、後者では選択された遺伝子内の偽陽性の数を評価するアルゴリズムが有効である。しなしながら、これらのアルゴリズムは複雑なクロスバリデーションや permutation test の実行が必要であり、多くの生物医学研究者にとって困難である。

本研究では、単変量 Cox 回帰にもとづく遺伝子選択法と、複合共変量にもとづく予後予測を包括的に実行する R パッケージ「*compound.Cox*」(Emura, Matsui and Chen 2018-) を提案する。ワルド検定もしくはスコア検定を全遺伝子に適用し、遺伝子の選択を行い、選択された遺伝子の予測性能を測定するクロスバリデーションアルゴリズム(CVL 値)を実行するプログラムを実装している。このようなツールは、予後予測に最適な遺伝子の選択する際に有用である。また選択された遺伝子の中に存在する偽陽性の数を推定 (FDR 値) するツールも実装している。さらに、選択された遺伝子を使用した 3 種の予後予測式の構築法 (複合共変量法, compound shrinkage 法, コピュラ法) を実装している。肺がんデータ (Chen et al. 2007) を用いて遺伝子選択法を説明し、False discovery rate (FDR) の計算やクロスバリデーションによって選択された遺伝子の予測能力の評価をするための CVL 値の計算法を紹介する。また選択された遺伝子を使用した予後予測の手法およびその予測精度の比較を行う。コピュラを使用した手法 (Emura and Chen 2016) や縮小推定法 (Emura, Chen and Chen, 2012) などの数学的に高度な手法の使用例も紹介する。

## 肺がんデータへのパッケージ使用例

肺がんデータにRパッケージ「*compound.Cox*」を適用し、遺伝子選択を行う例を下に示す。

```
library(compound.Cox)
data("Lung")

train=Lung$train ## index for training samples
t.vec=Lung$t.vec[train]
d.vec=Lung$d.vec[train]
X.mat=as.matrix(Lung[,-c(1,2,3)][train,])

uni.selection(t.vec,d.vec,X.mat,K=20,P.value=0.05,score=FALSE,permutation=TRUE) ## Wald test
uni.selection(t.vec,d.vec,X.mat,K=20,P.value=0.05,score=TRUE,permutation=TRUE) ## Score test
```

上のコードでは、まずパッケージ内のオブジェクト *Lung* にある  $n=125$  の患者データから  $n=63$  の訓練標本を取り出している。*Lung* 内にある “train” は訓練標本 ( $n=63$ ) とテスト標本 ( $n=62$ ) を識別するための変数である。*Lung* 内にある “t.vec” はセンサーされた生存期間を、“d.vec” はセンサーの指標変数 (0=センサー、1=死亡) を表す変数である。*Lung* 内には 97 の異なる遺伝子の発現量があり、行列 “X.mat” に代入している。この生存時間データ (t.vec, d.vec, X.mat) にワルド検定もしくはスコア検定を行い、“ $P$  値 $<0.05$ ” の閾値で遺伝子選択を行っている。上記コードを実行すると、選択された遺伝子の詳細情報 (遺伝子名、回帰係数、 $Z$  値、 $P$  値) が表示され、CVL 値と FDR 値が表示される。“ $K=20$ ” は CVL 値をクロスバリデーションで計算する際に、 $n=63$  個の標本を  $K=20$  個のグループにランダムかつ均等に分けることを意味している。“permutation=TRUE” は FDR 値を計算するために permutation test を実行することを意味している。Permutation test での計算には時間を要するが、“permutation=FALSE” の場合でも FDR 値は簡便法により計算される。スコア検定とワルド検定の計算スピードを比較すると、スコア検定が圧倒的に速いことがわかる。上記の例はパッケージの説明のほんの一部であり、詳細は原論文 Emura, Mastui and Chen (2018-) を参照されたい。

## 謝辞

本研究は科研費 (MOST 103-2118-M-008-MY2)、CREST (番号: JPMJCR1412)、科研費 (番号: 16H06299) の助成を受けたものである。

## 参考文献

- Beer DG, Kardia SLR., Huang CC., Giordano TJ, Levin AM, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine* 2002; **8**: 816-824.
- Bøvelstad HM, Nygård S, Storvold HL, Aldrin M., Borgan Ø, et al. Predicting survival from microarray data – a comparative study. *Bioinformatics* 2007; **23**: 2080-2087.
- Chen YH. Semiparametric marginal regression analysis for dependent competing risks under an assumed copula, *Journal of the Royal Statistical Society, Ser. B* 2010; **72**: 235-51.
- Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 2007; **356**: 11-20.
- Collett D. Modelling Survival Data in Medical Research, 3rd edition. CRC press, London, 2015
- Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. *J Am Stat Assoc* 2002; **97**: 77-87.
- Emura T, Chen HY, Matsui S, Chen YH. compound.Cox: univariate feature selection and compound covariate for predicting survival. *CRAN* 2018; Version **3.11**.
- Emura T, Chen YH, Gene selection for survival data under dependent censoring, a copula-based approach, *Statist Method Med Res* 2016; **25**(6): 2840-2857.
- Emura T, Chen YH, *Analysis of Survival Data with Dependent Censoring, Copula-Based Approaches*, JSS Research Series in Statistics, Springer, Singapore; 2018.
- Emura T, Chen YH, Chen HY. Survival prediction based on compound covariate under Cox proportional hazard models. *PLoS One* 2012; **7**(10): e47627. DOI:10.1371/journal.pone.0047627.
- Emura T, Michimae H. A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae, *Environ Ecol Stat* 2017; **24**(1): 151-73.
- Emura T, Nakatochi M, Matsui S, Michimae H, Rondeau V, Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: meta-analysis with a joint model, *Stat Method Med Res* 2017, doi:10.1177/0962280216688032
- Emura T, Matsui S, Chen HY, compound.Cox: univariate feature selection and compound covariate for predicting survival, in revision *Computer Methods and Programs in Biomedicine*.
- Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R, Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med* 2004; **350**(18): 1828-37.
- Matsui S. Predicting survival outcomes using subsets of significant genes in prognostic marker studies with microarrays. *BMC Bioinformatics* 2006; **7**:156.
- Matsui S. Statistical issues in clinical development and validation of genomic signatures. In Matsui S, Buyse M, and Simon R. (ed.) *Design and Analysis of Clinical Trials for Predictive Medicine* 2015, Boca Raton; CRC Press, pp. 207-226.



- Matsui S, Simon RM, Qu P, Shaughnessy JD, Barlogie B, Crowley J. Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine. *Clinical Cancer Research* 2012; **18**(21): 6065-6073.
- Rivest LP, Wells MT. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J Multivar Anal* 2001; **79**: 138-55
- Moradian H, Denis Larocque D, Bellavance F. Survival forests for data with dependent censoring, *Stat Method Med Res* 2017, DOI:10.1177/0962280217727314
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002; **346**(25):1937-47.
- Simon R. *Design and Analysis of DNA Microarray Investigations*. Springer, New York, 2003.
- Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005; **23**(29): 7332-41.
- Staplin ND, Kimber AC, Collett D, Roderick PJ. Dependent censoring in piecewise exponential survival models. *Stat Methods Med Res* 2015; **24**(3): 325-41.
- Tukey JW, Tightening the clinical trial. *Controlled Clinical Trials* 1993; **14**: 266-285.
- van Wieringen WN, Kun D, Hampel R, Boulesteix AL. Survival prediction using gene expression data: A review and comparison. *Comp Stat Data Anal*. 2009; **53**: 1590-1603.
- Vasselli JR, Shih JH, Iyengar SR, Maranchie J, Riss J, et al. Predicting survival in patients with metastatic kidney cancer by gene-expression profiling in the primary tumor. *Proc Natl Acad Sci* 2003; **100**(12): 6958-63.
- Waldron L, Haibe-Kains B, Culhane AC, Riester M, Ding J, et al. Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer. *J Natl Cancer Inst* 2014 **106**(5): dju049
- Wang Y, Klijn JG, Zhang Y, Sieuwerts, AM, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet* 2005; **365**(9460): 671-79.
- Witten DM, Tibshirani R. Survival analysis with high-dimensional covariates. *Stat Methods Med Res* 2010; **19**: 29-51.
- Yau C, Esserman L, Moore DH, Waldman, Sninsky FJ, BenzE CC. A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *BMC Breast Cancer Research* 2010; **12**:R85.
- Yoshihara K, Tajima A, Yahata T, Kodama S, Fujiwara H et al (2010). Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PloS One* **5**(3), e9615
- Zhao SD, Parmigiani G, Huttenhower C, Waldron L. Más-o-menos: a simple sign averaging method for discrimination in genomic data analysis. *Bioinformatics* 2014; **30**(21):3062-69.