

# 回帰分析を用いたマウス活動量データの遺伝解析

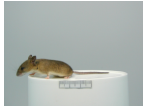
加藤 昇吾 数理・推論研究系 助教

## はじめに

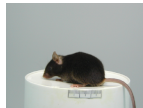
### 【背景】

本研究では、以下の2系統のマウスの遺伝子型に関するデータを扱う。

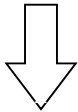
野生由来マウス系統 (MSM)



実験用マウス系統 (B6)



Nishi *et al.* (2010)は、2系統のマウスのホームケージ活動量(下図参照)に有意な差があることを明らかにした。また、染色体領域 Chr6の遺伝子型の違いが、活動量に大きく影響していることも示した。



### 【問題】

染色体領域 Chr6のどの部分が活動量に影響しているのか？

### 【目的】

この問題を解決するため、私たちは次のステップで研究を行った (Kato *et al.*, to appear)。

1. 染色体領域 Chr6の一部をMSMに置換したB6のコンジェニック系統を樹立。
2. 得られたコンジェニック系統とB6系統の活動量データから、Chr6のどの部分が影響しているのかを統計解析により推定。

このポスターでは、発表者が携わった2の内容について紹介する。

ホームケージ活動量の測定に関する装置

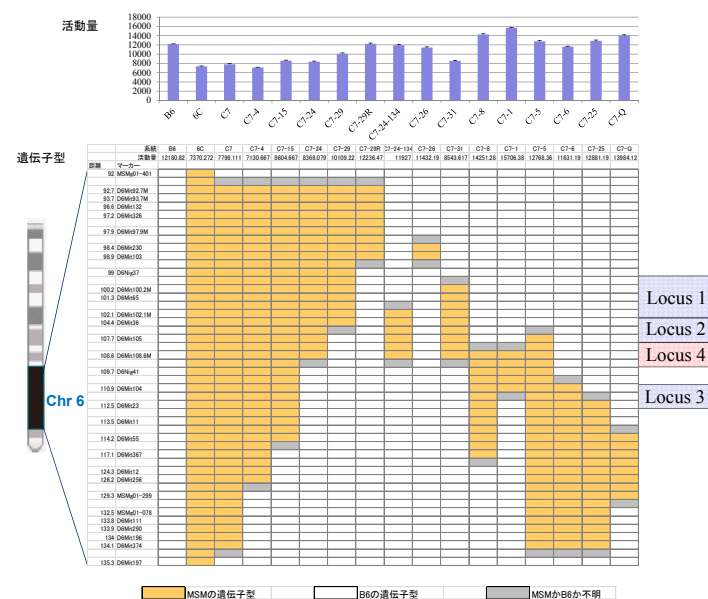
ホームケージ 行動解析装置 赤外線センサー



## 統計解析

上記の目的2を達成するため、私たちは回帰分析を用いて得られた活動量データの統計解析を行った。

【データ】 活動量および遺伝子型のデータ(B6系統とコンジェニック系統)



### 【解析における問題点】

今回の統計解析にあたり、最初に考えたのが、染色体領域Chr6の遺伝子型の情報から、活動量を予測する以下の回帰モデルを用いることだった。

$$Y_i = \beta_0 + \sum_{k=1}^{30} \beta_k x_{ik} + \sum_{k=1}^{29} \beta_{k+30} x_k : x_{k+1} + \varepsilon_i \quad (1)$$

ここで、

$Y_i$ : 系統  $i$  の活動量 ( $i = 1, \dots, 17$ ),

$x_{ik} = \begin{cases} 1, & \text{系統 } i \text{ の } k \text{ 番目のマーカーの遺伝子型がMSMと等しい} \\ 0, & \text{系統 } i \text{ の } k \text{ 番目のマーカーの遺伝子型がB6と等しい} \end{cases}, k = 1, \dots, 30,$

$x_k : x_{k+1} = x_{ik} x_{ik+1}, \beta_k$ : 回帰係数,  $\varepsilon_i \sim i.i.d. N(0, \sigma^2)$ .

回帰モデル(1)で冗長な説明変数を除き、説明変数の数を59から18まで削減した。しかし依然として、

(説明変数の数) > (系統数)

となっているため、通常の最小二乗法では回帰モデル(1)の推定値を一意に求めることができない ( $n \ll p$  問題)。

### 【解析方法】

この問題を克服する方法として、私たちは以下の2つのステップにて解析を行った。

#### Step 1: lasso

$$Y_i = \beta_0 + \sum_{k=1}^{30} \beta_k x_{ik} + \sum_{k=1}^{29} \beta_{k+30} x_k : x_{k+1} + \varepsilon_i \quad \text{subject to} \quad \sum_{k=0}^{59} |\beta_k| \leq c.$$

lassoは「(説明変数の数) > (系統数)」の場合にも一意に解が求まり、かつ、変数選択に有効であることが知られている。

#### Step 2: all subsets

lassoで選択されなかった説明変数を除き、「(説明変数の数) < (系統数)」となったデータに対して、all subsetsによるさらなる次元削減を行う。

### 【結果】

上記の解析手法を用いて推定されたモデルの回帰係数とp値は次の通り。

Marker	lasso	all subsets	p-value
Intercept	13176.1	12823.8	$1.28 \times 10^{-12}$
D6Mit230 - D6Mit103	-1035.4	-659.3	0.183
D6Nig100.2M - D6Mit65	-2397.8	-2389.1	0.00312
D6Nig65 : D6Nig102.1M			
D6Mit36 - D6Mit105	-1573.9	-2119.8	0.0370
D6Mit105			
D6Mit105 : D6Nig108.6M	-744.5	-1738.2	0.0669
D6Nig108.6M	700.1	2627.3	0.00123
D6Nig108.6M : D6Nig41			
D6Nig41	771.5	*	*
D6Nig41 : D6Mit104			
D6Mit104 : D6Mit23	-803.5	-944.6	0.0378
D6Mit367 : D6Mit12			
D6Mit12 - D6Mit256	-76.9	*	*
D6Nig01-299 : D6Nig01-078			
D6Nig01-078 - D6Mit374	-253.6	*	*

□はp値が0.05以下

- lassoのチューニングパラメータ ( $\lambda \approx 43.2$ ) は、100個のクロスバリデーションサンプルからの推定値のminimum modeかつ最頻値。
- all subsetsはAICが最小になるモデルを選択。

### 【結論】

この結果から以下の結論が導かれる。

MSMの遺伝子型がB6の遺伝子型と比べて：

- Locus 1-3: 活動量が有意に小さくなる領域。
- Locus 4: 活動量が有意に大きくなる領域。

特にLocus 1・2はLocus 3に比べ、遺伝子型の違いがより大きな活動量の差を生じさせる。

### 【参考文献】

1. Kato, S., Ishii, A., Nishi, A., Kuriki, S. and Koide, T. Segregation of a QTL cluster for home-cage activity using a new mapping method based on regression analysis of congenic mouse strains. *Heredity*, to appear.
2. Nishi, A., Ishii, A., Takahashi, A., Shiroishi, T. and Koide, T. (2010). QTL analysis of measures of mouse home-cage activity using B6/MSM consomic strains. *Mamm Genome*, 21, 477-485.