

Permutation test の検証

2016.05.21

ファイル名 :

PG_only_simu.permtest_20160420.R

読み込みデータファイル名 (RData) :

Selevsek_2015_all_Selevsek_20160123.RData

読み込みデータファイル名 (関数) :

func_for_permutest_of_Rho_20151124.R

OldFunc_20151125.R

func_for_calcEScore_20160123.R

系列内相関(時点間相関)と 系列間相関(mRNA・タンパク間)

- 時点間相関 : $\gamma(x_{g,,t}, x_{g,,t-1})$
- 系列間相関 : $\gamma(x_{g,m,t}, x_{g,p,t-u})$

ただし、

- g: 遺伝子ID
- m: mRNA, p: タンパク
- t: 時点, u: 時間差(遺伝子ごとに算出)

時点間相関の大小が、系列間相関の有意性の判定に影響を与えていないか調べる。

方針

観測データにランダムノイズを加えて、
シミュレーションデータとする。

時点間相関、系列間相関それぞれの、
順位相関係数を算出し、
観測データとの差(下表)の程度
によって、シミュレーションデータを分ける

観測データとの差 (± 0.05)

時点間相関	0.0	0.1	0.2	0.3
系列間相関	0.0	-0.1	-0.2	-0.3

16通り

16通りの各組み合わせにつき、
100個のシミュレーションデータを抽出する。

16通りの各組み合わせにつき、
100個のシミュレーションデータの内、
いくつが有意と判定されるか調べ、
時点間相関の増加によって、
有意と判定される割合が増えないことを
確認する。

[読み込みデータ]
Selevsek,2015

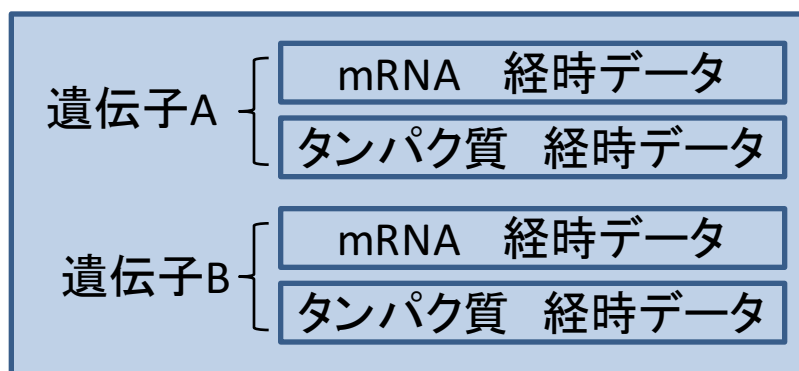
C_args 第1引数で
ファイル名を指定

C_args 第2引数で指定した
遺伝子の経時データを抽出
例：

遺伝子A:YDL110C

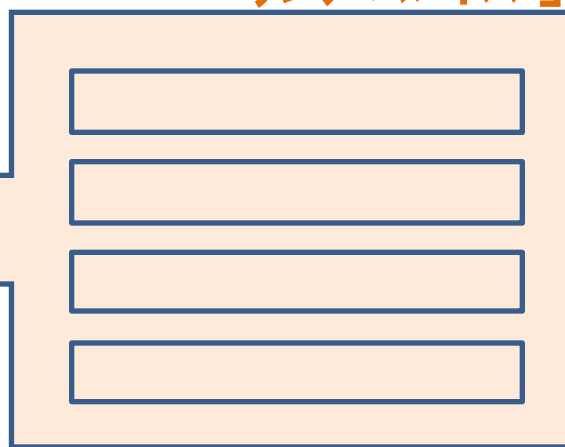
遺伝子B:YDR122W

「タネ データ」



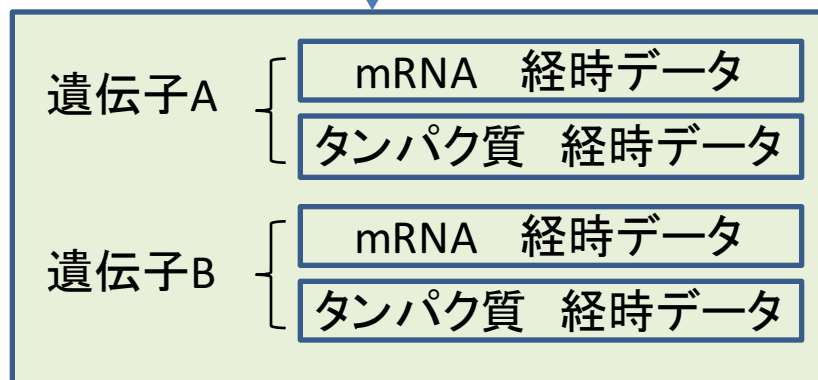
sim.dat.seed

「ランダムノイズ」



足す

シミュレーション
データ

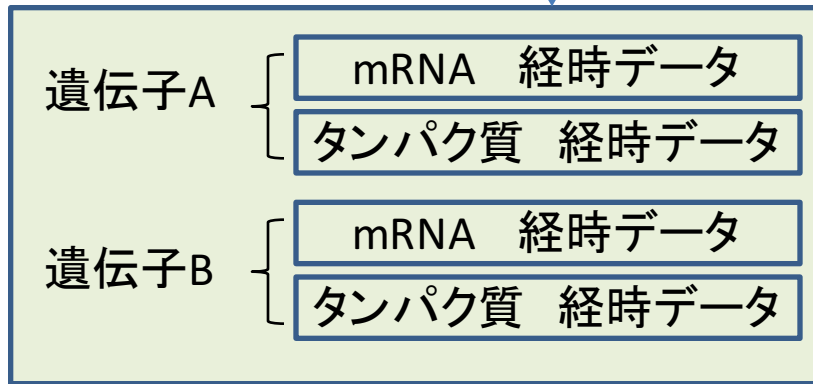


sim.dat

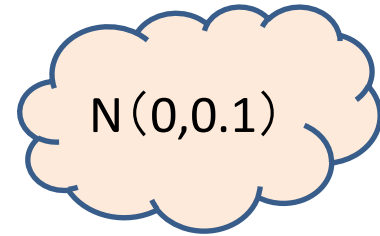
ランダム
サンプリング

$N(0,0.1)$

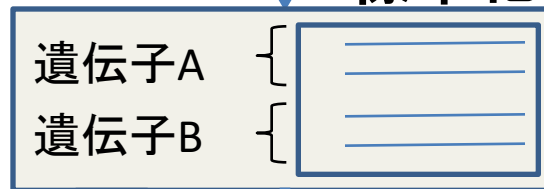
シミュレーション
データ



ランダム
サンプリング



標準化



`sim.dat $ norm`

系列間相関、及び
時点間相関について
Observed Data と比較
(`rho.which` ,
`rho.which.bw.tp`)

それぞれの順位相関係数について、Obs. との
差を $0.1(\pm 0.05)$ 刻みで変数として保持

系列間相関の差 → `sim.dat $ rho.which`

時点間相関の差 → `sim.dat $ rho.which.bw.tp`

`sim.dat $ rho_combi <- paste(
"_" , rho.which, rho.which.bw.tp)`

シミュレーションデータ
を行結合

sim.dat.set

16通り, 全て
 $N \geq 100$

ループ

NO

YES

permutation test

1

時点情報をシャッフル

→ mRNA経時データとタンパク質
経時データの順位相関

10,000回繰り返した結果の順位相関の分布
を帰無分布として、帰無分布中での元データ
の順位相関のパーセンタイル点をp値とする。

p 値

2

....

p 値

⋮

⋮

100

....

p 値

時点間相関、系列間相関の16通りの
組み合わせ全てについて、
有意となる割合を算出。