SLqPCR: Functions for analysis of real-time quantitative PCR data at SIRS-Lab GmbH

Dr. Matthias Kohl
SIRS-Lab GmbH (Jena, Germany)

October 22, 2008

Contents

1 Introduction ........................................... 1
2 Selection of most stable reference/housekeeping genes 1
3 Normalization by geometric averaging ............... 7

1 Introduction

The package "SLqPCR" was designed for the analysis of real-time quantitative RT-PCR data. In this short vignette we describe and demonstrate the available functions.

2 Selection of most stable reference/housekeeping genes

We describe the selection of the best (most stable) reference/housekeeping genes using method and data set of Vandesompele et al (2002) [1] (in the sequel: Vand02). We load library and data

\[
> \text{library(SLqPCR)} \\
> \text{data(vandesompele)} \\
> \text{str(vandesompele)}
\]

'data.frame': 85 obs. of 10 variables:
$ ACTB : num 0.0425 0.0192 0.1631 0.5726 0.037 ...
We start by ranking the selected reference/housekeeping genes. The function `selectHKgenes` proceeds stepwise; confer Section “Materials and methods” in Vand02. That is, the gene stability measure M of all candidate genes is computed and the gene with the highest M value is excluded. Then, the gene stability measure M for the remaining gene is calculated and so on. This procedure is repeated until two respectively `minNrHK` is reached.

```r
> tissue <- as.factor(c(rep("BM", 9), rep("POOL", 9), rep("FIB", + 20), rep("LEU", 13), rep("NB", 34)))
> res.BM <- selectHKgenes(vandesompele[tissue == "BM", ], method = "Vandesompele", + geneSymbol = names(vandesompele), minNrHK = 2, trace = TRUE, + na.rm = TRUE)
```

### Step 1:

<table>
<thead>
<tr>
<th>gene expression stability values M:</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPRT1</td>
</tr>
<tr>
<td>0.5160313</td>
</tr>
<tr>
<td>B2M</td>
</tr>
<tr>
<td>0.7747634</td>
</tr>
</tbody>
</table>
average expression stability M: 0.6362855
gene with lowest stability (largest M value): ACTB
Pairwise variation, (9 / 10): 0.076469

### Step 2:

```r
Step 2:
gene expression stability values M:
   HPRT1  RPL13A  YWHAZ  UBC   GAPD  SDHA  TBP   HMBS
  0.4705664 0.5141375 0.5271169 0.5554718 0.5575295 0.5738460 0.6042110 0.6759176
B2M 0.7671985
average expression stability M: 0.5828883
gene with lowest stability (largest M value): B2M
Pairwise variation, (8 / 9): 0.0776543
```
### Step 3:
Gene expression stability values M:

<table>
<thead>
<tr>
<th></th>
<th>HPRT1</th>
<th>RPL13A</th>
<th>SDHA</th>
<th>YWHAZ</th>
<th>UBC</th>
<th>GAPD</th>
<th>TBP</th>
<th>HMBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>0.439</td>
<td>0.473</td>
<td>0.524</td>
<td>0.525</td>
<td>0.540</td>
<td>0.556</td>
<td>0.562</td>
<td>0.621</td>
</tr>
</tbody>
</table>

Average expression stability M: 0.5302283

Gene with lowest stability (largest M value): **HMBS**

Pairwise variation, (7 / 8): 0.067112

### Step 4:
Gene expression stability values M:

<table>
<thead>
<tr>
<th></th>
<th>HPRT1</th>
<th>RPL13A</th>
<th>YWHAZ</th>
<th>UBC</th>
<th>SDHA</th>
<th>GAPD</th>
<th>TBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>0.439</td>
<td>0.469</td>
<td>0.488</td>
<td>0.504</td>
<td>0.518</td>
<td>0.525</td>
<td>0.556</td>
</tr>
</tbody>
</table>

Average expression stability M: 0.4999437

Gene with lowest stability (largest M value): **TBP**

Pairwise variation, (6 / 7): 0.06813202

### Step 5:
Gene expression stability values M:

<table>
<thead>
<tr>
<th></th>
<th>HPRT1</th>
<th>RPL13A</th>
<th>UBC</th>
<th>YWHAZ</th>
<th>GAPD</th>
<th>SDHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>0.430</td>
<td>0.445</td>
<td>0.460</td>
<td>0.473</td>
<td>0.501</td>
<td>0.557</td>
</tr>
</tbody>
</table>

Average expression stability M: 0.4773775

Gene with lowest stability (largest M value): **SDHA**

Pairwise variation, (5 / 6): 0.08061944

### Step 6:
Gene expression stability values M:

<table>
<thead>
<tr>
<th></th>
<th>UBC</th>
<th>RPL13A</th>
<th>HPRT1</th>
<th>YWHAZ</th>
<th>GAPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>0.420</td>
<td>0.420</td>
<td>0.422</td>
<td>0.422</td>
<td>0.484</td>
</tr>
</tbody>
</table>

Average expression stability M: 0.4377282

Gene with lowest stability (largest M value): **GAPD**

Pairwise variation, (4 / 5): 0.08416531

### Step 7:
Gene expression stability values M:

<table>
<thead>
<tr>
<th></th>
<th>RPL13A</th>
<th>UBC</th>
<th>YWHAZ</th>
<th>HPRT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values</td>
<td>0.370</td>
<td>0.398</td>
<td>0.417</td>
<td>0.442</td>
</tr>
</tbody>
</table>

Average expression stability M: 0.4067706

Gene with lowest stability (largest M value): **HPRT1**

Pairwise variation, (3 / 4): 0.09767827
Step 8:
gene expression stability values M:
  UBC  RPL13A  YWHAZ
0.3559286 0.3761358 0.3827933
average expression stability M: 0.3716192
gene with lowest stability (largest M value): YWHAZ
Pairwise variation, (2/3): 0.1137450

Step 9:
gene expression stability values M:
  RPL13A   UBC
0.3492712 0.3492712
average expression stability M: 0.3492712

```r
> res.POOL <- selectHKgenes(vandesompele[tissue == "POOL",], method = "Vandesompele",
+ geneSymbol = names(vandesompele), minNrHK = 2, trace = FALSE,
+ na.rm = TRUE)
> res.FIB <- selectHKgenes(vandesompele[tissue == "FIB",], method = "Vandesompele",
+ geneSymbol = names(vandesompele), minNrHK = 2, trace = FALSE,
+ na.rm = TRUE)
> res.LEU <- selectHKgenes(vandesompele[tissue == "LEU",], method = "Vandesompele",
+ geneSymbol = names(vandesompele), minNrHK = 2, trace = FALSE,
+ na.rm = TRUE)
> res.NB <- selectHKgenes(vandesompele[tissue == "NB",], method = "Vandesompele",
+ geneSymbol = names(vandesompele), minNrHK = 2, trace = FALSE,
+ na.rm = TRUE)

We obtain the following ranking of genes (cf. Table 3 in Vand02)

```r
> ranks <- data.frame(c(1, 1:9), res.BM$ranking, res.POOL$ranking,
+ res.FIB$ranking, res.LEU$ranking, res.NB$ranking)
> names(ranks) <- c("rank", "BM", "POOL", "FIB", "LEU", "NB")
> ranks

<table>
<thead>
<tr>
<th>rank</th>
<th>BM</th>
<th>POOL</th>
<th>FIB</th>
<th>LEU</th>
<th>NB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RPL13A</td>
<td>GAPD</td>
<td>GAPD</td>
<td>UBC</td>
<td>GAPD</td>
</tr>
<tr>
<td>2</td>
<td>UBC</td>
<td>SDHA</td>
<td>HPRT1</td>
<td>YWHAZ</td>
<td>HPRT1</td>
</tr>
<tr>
<td>3</td>
<td>YWHAZ</td>
<td>HMBS</td>
<td>YWHAZ</td>
<td>B2M</td>
<td>SDHA</td>
</tr>
<tr>
<td>4</td>
<td>HPRT1</td>
<td>HPRT1</td>
<td>UBC</td>
<td>GAPD</td>
<td>UBC</td>
</tr>
<tr>
<td>5</td>
<td>GAPD</td>
<td>TBP</td>
<td>ACTB</td>
<td>RPL13A</td>
<td>HMBS</td>
</tr>
<tr>
<td>6</td>
<td>SDHA</td>
<td>UBC</td>
<td>TBP</td>
<td>TBP</td>
<td>YWHAZ</td>
</tr>
<tr>
<td>7</td>
<td>TBP</td>
<td>RPL13A</td>
<td>SDHA</td>
<td>SDHA</td>
<td>TBP</td>
</tr>
</tbody>
</table>
```
Remark 1:

(a) Since the computation is based on gene ratios, the two most stable control genes in each cell type cannot be ranked.

(b) In praxis the selection of reference/housekeeping genes may require an additional step which is the computation of relative quantities via relQuantPCR; e.g.

```r
> exa1 <- apply(vandesompele[tissue == "BM", , 2, relQuantPCR, +   E = 2)
```

We plot the average expression stability $M$ for each cell type (cf. Figure 2 in Vand02).

```r
> library(RColorBrewer)
> mypalette <- brewer.pal(5, "Set1")
> matplot(cbind(res.BM$meanM, res.POOL$meanM, res.FIB$meanM, res.LEU$meanM, +   res.NB$meanM), type = "b", ylab = "Average expression stability $M$", +   xlab = "Number of remaining control genes", axes = FALSE, +   pch = 19, col = mypalette, ylim = c(0.2, 1.22), lty = 1, +   lwd = 2, main = "Gene stability measure")
> axis(1, at = 1:9, labels = as.character(10:2))
> axis(2, at = seq(0.2, 1.2, by = 0.2), labels = as.character(seq(0.2, +   1.2, by = 0.2)))
> box()
> abline(h = seq(0.2, 1.2, by = 0.2), lty = 2, lwd = 1, col = "grey")
> legend("topright", legend = c("BM", "POOL", "FIB", "LEU", "NB"), +   fill = mypalette)
```
Second, we plot the pairwise variation for each cell type (cf. Figure 3 (a) in Vand02)

```r
> mypalette <- brewer.pal(8, "YlGnBu")
> barplot(cbind(res.BM$variation, res.POOL$variation, res.FIB$variation,
+             res.LEU$variation, res.NB$variation), beside = TRUE, col = mypalette,
+             space = c(0, 2), names.arg = c("BM", "POOL", "FIB", "LEU",
+             "NB"))
> legend("topright", legend = c("V9/10", "V8/9", "V7/8", "V6/7",
+             "V5/6", "V4/5", "V3/4", "V2/3"), fill = mypalette, ncol = 2)
> abline(h = seq(0.05, 0.25, by = 0.05), lty = 2, col = "grey")
> abline(h = 0.15, lty = 1, col = "black")
```
Remark 2:
Vand02 recommend a cut-off value of 0.15 for the pairwise variation. Below this bound the inclusion of an additional housekeeping gene is not required.

3 Normalization by geometric averaging

To normalize your data by geometric averaging of multiple reference/housekeeping genes you can proceed as follows

```r
> data(SLqPCRdata)
> SLqPCRdata
```

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
<th>HK1</th>
<th>HK2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>26.6</td>
<td>25.6</td>
<td>12.8</td>
</tr>
<tr>
<td>A2</td>
<td>26.9</td>
<td>25.8</td>
<td>13.2</td>
</tr>
<tr>
<td>A3</td>
<td>27.4</td>
<td>26.1</td>
<td>13.1</td>
</tr>
</tbody>
</table>
> (relData <- apply(SLqPCRdata, 2, relQuantPCR, E = 2))

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
<th>HK1</th>
<th>HK2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.21763764</td>
<td>0.008974206</td>
<td>0.9330330</td>
</tr>
<tr>
<td>A2</td>
<td>0.17677670</td>
<td>0.007812500</td>
<td>0.7071068</td>
</tr>
<tr>
<td>A3</td>
<td>0.12500000</td>
<td>0.006345722</td>
<td>0.7578583</td>
</tr>
<tr>
<td>A4</td>
<td>0.10153155</td>
<td>0.004487103</td>
<td>0.6155722</td>
</tr>
<tr>
<td>B1</td>
<td>0.20306310</td>
<td>0.007812500</td>
<td>0.8705506</td>
</tr>
<tr>
<td>B2</td>
<td>1.00000000</td>
<td>0.153893052</td>
<td>0.7578583</td>
</tr>
<tr>
<td>B3</td>
<td>0.23325825</td>
<td>0.017948412</td>
<td>0.8705506</td>
</tr>
<tr>
<td>B4</td>
<td>0.43527528</td>
<td>0.038473263</td>
<td>0.4665165</td>
</tr>
<tr>
<td>C1</td>
<td>0.04736614</td>
<td>0.004487103</td>
<td>0.7578583</td>
</tr>
<tr>
<td>C2</td>
<td>1.00000000</td>
<td>0.153893052</td>
<td>0.7578583</td>
</tr>
<tr>
<td>C3</td>
<td>0.06698584</td>
<td>0.012691444</td>
<td>0.8705506</td>
</tr>
<tr>
<td>C4</td>
<td>0.53588673</td>
<td>0.287174589</td>
<td>0.6597540</td>
</tr>
<tr>
<td>D1</td>
<td>0.03349292</td>
<td>0.004809158</td>
<td>0.8705506</td>
</tr>
<tr>
<td>D2</td>
<td>0.81225240</td>
<td>1.000000000</td>
<td>1.0000000</td>
</tr>
<tr>
<td>D3</td>
<td>0.13397168</td>
<td>0.203063099</td>
<td>0.8122524</td>
</tr>
<tr>
<td>D4</td>
<td>0.13397168</td>
<td>0.176776695</td>
<td>0.7578583</td>
</tr>
</tbody>
</table>

> geneStabM(relData[, c(3, 4)])

    HK1   HK2
0.2574717 0.2574717

> (exprData <- normPCR(SLqPCRdata, c(3, 4)))

<table>
<thead>
<tr>
<th>Gene1</th>
<th>Gene2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1.728585</td>
</tr>
</tbody>
</table>

8
A2  1.689720  1.620623
A3  1.727684  1.645714
A4  1.713602  1.645553
B1  1.714500  1.656708
B2  1.558954  1.373669
B3  1.706201  1.583870
B4  1.564586  1.436241
C1  1.820707  1.681626
C2  1.561410  1.228651
C3  1.826986  1.620401
C4  1.587369  1.292483
D1  1.871526  1.692677
D2  1.615795  1.229836
D3  1.755636  1.356920
D4  1.758402  1.371940

References