Estimation of Power and Analysis of qPCR Data with Normal Mixed Models

Auli Partanen, Éva Tas, Juha Akkila, Sami Hokkanen
Orion Corporation Orion Pharma, Finland

Statististikot Suomen Lääketeollisuudessa syysseminaari
Espoo, Finland on November 3th, 2008
qPCR

- Is used to quantify DNA or messenger RNA (mRNA) in a sample

- When combined with reverse transcriptase (RT-PCR), relative gene expressions between tissues or genes can be compared

- Measure fluorescence reporter

http://www.rt-pcr.com/
http://www.gene-quantification.info/
Standardizing measurements

• Significant problems caused by
  – variability of RNA templates, assay designs and protocols
  – various data normalization
  – data analysis strategies
• Which are tried to control by
  – consistently using standard chemistries, protocols and reaction conditions
  – pipetting robot
  – repeated measurements (technical repeats)
  – all measurements (to be compared) at the same time
  – controlling efficiency in each run
Aim of study and Study design

• **Aim of study is to evaluate relative expression ratio between genes and treatments**
• **18 animal tissue samples in 4 groups**
  1. control (n=4)
  2. disease model (n=5)
  3. disease model with study drug treatment (n=4)
  4. disease model with reference drug treatment (n=5)
• **Expression levels of 3 target genes and 12 potential reference genes** were analyzed by quantitative RT-PCR in **3 replicates** for each sample.
## Study design

<table>
<thead>
<tr>
<th></th>
<th>Original data</th>
<th>Selected data</th>
<th>Analysis data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Samples</strong></td>
<td>18</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td><strong>Treatments</strong></td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Genes</strong></td>
<td>3 targets</td>
<td>X 3 technical repeats</td>
<td>1 target</td>
</tr>
<tr>
<td></td>
<td>12 references</td>
<td>3 references</td>
<td>3 references</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>810</td>
<td>96</td>
<td>16</td>
</tr>
</tbody>
</table>

*) Arithmetic mean
## Analysis Data

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Gene</th>
<th>Sample</th>
<th>Ct</th>
<th>Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>Target A</td>
<td>28</td>
<td>35.6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
<td>38.1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>37.3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td>35.7</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>Reference</td>
<td>28</td>
<td>24.7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32</td>
<td>28.3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>26.5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td>25.1</td>
<td>2</td>
</tr>
<tr>
<td>Study drug treatment</td>
<td>Target A</td>
<td>66</td>
<td>37.1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>69</td>
<td>36.3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>36.6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71</td>
<td>37.1</td>
<td>3</td>
</tr>
<tr>
<td>Study drug treatment</td>
<td>Reference</td>
<td>66</td>
<td>24.5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>69</td>
<td>24.9</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
<td>23.8</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71</td>
<td>25.4</td>
<td>4</td>
</tr>
</tbody>
</table>
Definitions

- For reference gene: \( \text{mean } r_{\text{control}} \) and \( \text{mean } r_{\text{treatment}} \)
- \( \Delta C_{t_{\text{ref}}} = \text{mean } r_{\text{control}} - \text{mean } r_{\text{treatment}} \)

- For target gene: \( \text{mean } t_{\text{control}} \) and \( \text{mean } t_{\text{treatment}} \)
- \( \Delta C_{t_{\text{target}}} = \text{mean } t_{\text{control}} - \text{mean } t_{\text{treatment}} \)

- \( \Delta \Delta C_t = \Delta C_{t_{\text{ref}}} - \Delta C_{t_{\text{target}}} \)

- \( (E_{\text{target}})^{\Delta C_{t_{\text{target}}}} \)

- Expression ratio = \( \frac{(E_{\text{ref}})^{\Delta C_{t_{\text{ref}}}}}{(E_{\text{target}})^{\Delta C_{t_{\text{target}}}}} \) = 2 \( -\Delta \Delta C_{t_{\text{adjusted}}} \)
PAE and adjusted $\Delta\Delta C_t$

- qPCR data analysis is based on the assumption that PCR products double each cycle ($AE=2$).
- When the AE (Amplification Efficiency) is not 2, Ct -values are recommended to be adjusted.
- We used percentile AE (PAE) instead of AE

$$AE = 2^{PAE}$$

$$PAE = \log_2(AE)$$

$$\Delta\Delta C_t_{adjusted} = PAE_{ref} \times \Delta C_t_{ref} - PAE_{target} \times \Delta C_t_{target}$$

- Efficiency can be estimated for a group of reactions or a single reaction by simple regression model.
Normal Mixed Model with Gene as a repeated factor

- Effects
  - gene (target, reference)
  - treatment (study drug, control)
  - gene by treatment interaction
  - sample (sample number)
  - residual

- Mixed Model in SAS

```sas
PROC MIXED; CLASS gene treatment sample;
MODEL  Ct = gene treatment gene*treatment;
REPEATED gene / SUBJECT = sample TYPE = UN;
```
Estimation of $\Delta\Delta$Ct based on Mixed Model parametrization (A) and (B) ?

Mixed Model in SAS (parametrization A)

PROC MIXED; CLASS gene treatment sample;
MODEL Ct = gene treatment gene*treatment;
REPEATED gene / SUBJECT = sample TYPE = UN;

$\Delta\Delta$Ct: ESTIMATE gene*treatment +1 -1 -1 +1;

Mixed Model in SAS (parametrization B)

PROC MIXED; CLASS subgroup sample;
MODEL Ct = subgroup;
REPEATED / SUBJECT = sample TYPE = UN;

$\Delta\Delta$Ct_{adjusted}: ESTIMATE subgroup '+PAE_{target} -PAE_{target} -PAE_{ref} +PAE_{ref}';

• (subgroup is a categorical variable with 4 classes)
Estimation of PAE

PAE estimates are based on the data of 5 different dilutions (and 3 technical repeats per dilution) over the pooled samples treated by 4 treatments.

<table>
<thead>
<tr>
<th>Gene</th>
<th>PAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference 1</td>
<td>0.831</td>
</tr>
<tr>
<td>Reference 2</td>
<td>0.929</td>
</tr>
<tr>
<td>Reference 3</td>
<td>0.903</td>
</tr>
<tr>
<td>Target A</td>
<td>0.924</td>
</tr>
</tbody>
</table>

Mean PAE of reference genes: 0.888
## Table with descriptive $\Delta\Delta C_t$ and model estimates

<table>
<thead>
<tr>
<th></th>
<th>$\Delta\Delta C_t$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Descriptive statistics</td>
</tr>
<tr>
<td>PAE Adjusted</td>
<td>1.437 (0.708)</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>1.613 (0.786)</td>
</tr>
</tbody>
</table>
\( \Delta \Delta Ct \) and expression ratio with 95% CIs

\[ \text{Expression ratio} = 2^{-\Delta \Delta Ct} \]
Power curves for the expression ratio between 0 and 10 when type I error is 0.05

Power calculation is based on
- noncentral t-distribution and the log2 transformed expression ratio ($\Delta\Delta$Ct)
- variance was estimated as the sum of four equal variances
- degrees of freedom was estimated as a sum of $n$ subtracted by a number of groups
Conclusion

- Importance of power calculation in study planning phase.
- Normal mixed model works well with qPCR data and enables the dependence between genes.
- With balanced, complete data these two methods give similar results.
- Mixed model gives more accurate estimates with unbalanced data.
- Data transfer from instrument to analysis software is challenging.
References

Thank you!

Questions?