Quantitative aspects of real-time RT-qPCR: current and future

Third Scientific Meeting MolecularDiagnostics.be
Congress Centre Elzenveld, Antwerp
13th December 2011

Michael W. Pfaffl
Professor of Molecular Physiology
Technical University of Munich Weihenstephan, 85354 Freising, Germany
Michael.Pfaffl@wzw.tum.de
www.Gene-Quantification.info

Molecular diagnostic challenges for the future

• Minimal amount of tissue or biological matrix necessary
• Sensitive assay – on level of single molecules
• Quick diagnosis – in hours
• Wide quantification range
• Early detection of disease or micro-organism
• Stable and valid biomarker(s) or biomarker pattern
• Highly standardized and reproducible

⇒ RT-qPCR is the method of choice
⇒ Amplification, analysis, and quantification in one tube
⇒ Reliable biomarker pattern = Transcriptional biomarkers
⇒ MIQE compliant
Quantitative aspects of real-time RT-qPCR current and future

Outline:
1. MIQE guidelines
2. Single-cell analytics on DNA and RNA level
3. Transcriptional biomarker discovery

- MIQE is a set of guidelines that describe the minimum information necessary for evaluating qPCR experiments

⇒ The goal is:
  • To provide guidelines for authors, reviewers and editors to measure the technical quality of submitted manuscripts
  • To establish a clear framework to conduct quantitative RT-PCR experiments
  • To support experimental transparency
  • To increase reproducibility between laboratories worldwide
  • To promote more consistent, more comparable, and more reliable results
  • To standardize international qPCR nomenclature

⇒ To increase reliability of results to help to insure the integrity of scientific work, with major focus on biological relevance
### The MIQE checklist:
- **9 main-titles**
- **85 sub-titles**
- **57 essential information**
- **28 desirable information**

- Experimental design
- Sampling
- NA extraction
- RT
- qPCR target information
- qPCR primers
- qPCR protocol
- qPCR assay validation
- Data analysis & statistics

### FOCUS:

**“Absolute Quantification” and data analysis in single-cells**

<table>
<thead>
<tr>
<th>Sample type</th>
<th>tissue</th>
<th>sample</th>
<th>RNA</th>
<th>cDNA</th>
<th>PCR</th>
<th>RT-PCR product</th>
<th>quantification strategy</th>
<th>statistics</th>
<th>results or display and analysis tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>tissue sample</td>
<td>RNA</td>
<td>cDNA</td>
<td>PCR</td>
<td>RT</td>
<td>Real-time PCR amplification</td>
<td>detection &amp; software</td>
<td>Ct processing</td>
<td>mean or</td>
</tr>
</tbody>
</table>

### BioStatistics & BioInformatics:
- Cq vs. quantified molecules
- Normality of data (?
- t-Test (?
- ANOVA (on the ranks ?
- SAS, SPSS, Excel, Sigma Stat
- Permutation test
- Randomization test (REST)
- Bootstrapting (REST-2009)
- Multiple regression analysis
- Cluster analysis
- Multi-dimensional modeling

### Quantification strategy:
- **“absolute” quantification**
  - type of calibration curve
  - on single-cell level
  - Impact of biological matrix
  - add. normalization with RG
- **relative quantification**
  - total RNA, single-cell, mass
  - Normalization: RG vs. RG Index (> 3 RGs)
  - RefGenes.org pre-selection
  - geNorm, NormFinder, BestKeeper
  - REST, qBASE, GenEx, Kineffet
The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments

Stephen A. Bustin,1* Vladimir Benes,2 Jeremy A. Garson,2,4 Jan Hellemans,5 Jim Huggett,6 Mikael Kubista,7,8 Reinhold Mueller,9 Tanja Nolan,10 Michael W. Pfaffl,11 Gregory L. Shippy,12 Jo Vandecasteele,3 and Carl T. Wittwer11,14

Why focusing on single-cell analysis?

One single-cell …
- contains a set of all genes
- can represent an organism
- might be the starting point of a population
  - reproductive biology (IVF)
  - cancer biology (CTC)
  - stem-cell research

⇒ Life starts with one single-cell …
⇒ To learn about life, you have to go back to the roots …
Why test cell-by-cell expression heterogeneity?

Current testing:
• Expression markers are based on features observed in tissue samples in clusters of cells, from 100s – 10,000s of cells

What if:
• Samples contain varying mix of normal and diseased cells (=> tissue heterogeneity)
• Expression pattern is changing between cell types and cellular layers (=> tissue heterogeneity)
• Expression pattern is changing cell-by-cell
• Expression dynamic is changing over time

Expression analysis of a cell population can be misleading

mRNA expression levels

qPCR of cell population

Single cell analysis

10 20 30 40 50 60

Rocha B., qPCR Event 2007, Freising
Analytical questions on the single-cell level:

- What is the dynamic quantification range for DNA and RNA?
- DNA and RNA assay sensitivity?
- Variability of quantification?
- What is the influence of the biological matrix?
- Does the cell sampling and handling have an impact on the quantification results?
- What is the transcriptional noise and what is the technical noise?

=> Validation of DNA and RNA quantification procedure on the single-cell level is necessary!

=> Adaptation of the MIQE guidelines!

"single-cell hardware"

48 reaction sites (AmpliGrid AG480F slides)
FACS sorting (DKFZ Heidelberg)
one-cell per reaction site
1µl reaction per reaction site
AmpliSpeed slide cycler (ASC200D)
Our single-cell workflow

1) single-cell (WBC) deposition by FACS and/or DNA or RNA spiking
2) cell-lysis
3) NA stabilization
4) RT
5) pre-amp multiplexing 10 cycles $2^{10} \sim 1000$-fold
6) dilution 1/100 in 96-well single-assays
7) qPCR single-assay 40 cycles
8) data analysis

RNA standard without WBC input copy

- log 0 1 2 3 4 5 6 7
- Cq 10 15 20 25 30 35 40

Input copy numbers vs Cq linear regression ($r = 0.84$) 95% confidence interval

DNA std.
DNA std. + pre-Amp
DNA std. + pre-Amp + LYM
RNA std.
RNA std. + LYM
mRNA expression profiling
Extraction Control Plasmid
artificial synthetic DNA and RNA standard sequence

- minimal homology to any DNA or RNA target gene
- any contamination can be excluded
- target compatible extraction efficiency
- exact known copy numbers (!)
- guaranteed DNA and RNA stability (!)

Validation of quantification on DNA level
(DNA spike-in experiment)

DNA standard + PreAmp

DNA standard (solely)

DNA standard + PreAmp + LYM

\[ r^2 = 0.997 \]

\[ r^2 = 0.986 \]

\[ r^2 = 0.981 \]
Technical variance induced by pre-amplification, RT, and cellular material (= “biological matrix effects”)

\( n = 4 \)

Table 2. Technical variations for DNA or RNA standard amplification curves induced by pre-amplification, reverse transcription and cellular material (\( n = 4 \), standard deviation = SD)

<table>
<thead>
<tr>
<th>DNA standard</th>
<th>DNA standard (solely)</th>
<th>DNA standard + PreAmp</th>
<th>DNA standard + PreAmp + LYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean copies (SD)</td>
<td>Variation (%)</td>
<td>Mean copies (SD)</td>
<td>Variation (%)</td>
</tr>
<tr>
<td>10^0</td>
<td>11.13 (1.83)</td>
<td>16.25</td>
<td>205.62 (145.78)</td>
</tr>
<tr>
<td>10^1</td>
<td>102.46 (8.17)</td>
<td>7.97</td>
<td>200.57 (62.67)</td>
</tr>
<tr>
<td>10^2</td>
<td>1058.81 (66.30)</td>
<td>6.28</td>
<td>2426.47 (1894.18)</td>
</tr>
<tr>
<td>10^3</td>
<td>10481.79 (170.02)</td>
<td>3.46</td>
<td>7385.76 (12926.16)</td>
</tr>
<tr>
<td>10^4</td>
<td>109308.79 (5957.06)</td>
<td>3.62</td>
<td>934585.01 (110475.66)</td>
</tr>
<tr>
<td>10^5</td>
<td>14047123.35 (800792.20)</td>
<td>4.61</td>
<td>2453015.76 (1305231.87)</td>
</tr>
</tbody>
</table>

\( r^2 \) Mean variation: \( r^2 = 0.997 \) 7.03 \( r^2 = 0.866 \) 28.19 \( r^2 = 0.981 \) 27.81

<table>
<thead>
<tr>
<th>RNA standard</th>
<th>RNA standard (solely)</th>
<th>RNA standard + LYM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean copies (SD)</td>
<td>Variation (%)</td>
<td>Mean copies (SD)</td>
</tr>
<tr>
<td>10</td>
<td>107.25 (134.03)</td>
<td>124.99</td>
</tr>
<tr>
<td>10^1</td>
<td>106.82 (141.53)</td>
<td>13.25</td>
</tr>
<tr>
<td>10^2</td>
<td>108.18 (394.42)</td>
<td>43.90</td>
</tr>
<tr>
<td>10^3</td>
<td>5794.50 (3835.54)</td>
<td>66.19</td>
</tr>
<tr>
<td>10^4</td>
<td>105940.66 (48630.58)</td>
<td>38.29</td>
</tr>
<tr>
<td>10^5</td>
<td>297710.24 (331477.18)</td>
<td>111.31</td>
</tr>
</tbody>
</table>

\( r^2 \) Mean variation: \( r^2 = 0.970 \) 66.34 \( r^2 = 0.962 \) 63.32

Validation of quantification on RNA level

(RNA spike-in experiment)
Sensitivity, variability and quantification range on DNA level

• Sensitivity down to 10 DNA molecules
• Quantification range $10^{-1} - 10^6$ DNA molecules
• Highest precision and trueness in "pure qPCR" experiment
• Sources of variation:
  qPCR ~7%  pre-Amp ~18-20%  biological matrix (LYM) ~1-2%
• PCR efficiencies:
  qPCR 96.7%  pre-Amp+qPCR 84.4%  pre-Amp+qPCR+LYM 74.2%

Sensitivity, variability and quantification range on RNA level

• Detection is possible for 10 RNA molecules, but with high variances!
• Quantification range is $10^2 - 10^6$ RNA molecules
• RT efficiency: ~67.7%  [ranged 47% - 129%]
• PCR efficiencies: qPCR ~77.4%  [average value]
• Sources of variation:
  RT+qPCR ~63-65%  RT ~55-58%  biological matrix (LYM) ~1-2%

LYM sampling and RNA handling effects

<table>
<thead>
<tr>
<th>standard</th>
<th>with evaporation</th>
<th>without evaporation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean copies</td>
<td>S.D.</td>
</tr>
<tr>
<td>10</td>
<td>107.23</td>
<td>134.03</td>
</tr>
<tr>
<td>$10^2$</td>
<td>106.82</td>
<td>14.15</td>
</tr>
<tr>
<td>$10^3$</td>
<td>898.38</td>
<td>394.42</td>
</tr>
<tr>
<td>$10^4$</td>
<td>5794.50</td>
<td>3835.54</td>
</tr>
<tr>
<td>$10^5$</td>
<td>105840.86</td>
<td>40630.58</td>
</tr>
<tr>
<td>mean</td>
<td>57.34</td>
<td>8.89</td>
</tr>
</tbody>
</table>

• Evaporation has a significant effect on later RNA quantification!
• Sensitivity dropped from 10 to 100 RNA molecules
• Quantification range dropped from $10^{-1} - 10^3$ to $10^2 - 10^5$ RNA molecules
• Variation increased from ~9% to ~58%
• RNA stabilization is highly recommended!
Gene expression profiling in LPS treated LYM

n = 192 single LYM

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPS +</td>
<td>A</td>
<td>St1</td>
<td>St2</td>
<td>St3</td>
<td>St4</td>
<td>St5</td>
<td>St6</td>
<td>St7</td>
<td>St8</td>
<td>St9</td>
<td>St10</td>
<td>St11</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPS -</td>
<td>E</td>
<td>St1</td>
<td>St2</td>
<td>St3</td>
<td>St4</td>
<td>St5</td>
<td>St6</td>
<td>St7</td>
<td>St8</td>
<td>St9</td>
<td>St10</td>
<td>St11</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pipetting scheme corresponding to lymphocyte deposition = 2 NTC

Gene specific standard curves in qPCR

<table>
<thead>
<tr>
<th>Gene</th>
<th>PCR eff. [%]</th>
<th>r²</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>108.2</td>
<td>0.981</td>
<td>-3.14</td>
</tr>
<tr>
<td>TLR4</td>
<td>97.2</td>
<td>0.990</td>
<td>-3.39</td>
</tr>
<tr>
<td>IL1β</td>
<td>118.0</td>
<td>0.989</td>
<td>-2.96</td>
</tr>
<tr>
<td>TNFα</td>
<td>94.2</td>
<td>0.998</td>
<td>-3.47</td>
</tr>
</tbody>
</table>

Box plot of quantified copies-per-cell (median, 10th, 25th, 75th and 90th percentiles) GAPDH (n = 144), TNFα (n = 56), IL-1β (n = 32), and TLR4 (n = 86) of 192 analyzed cells

LYM single-cell gene expression

1µg/ml LPS treatment 2h at 37°C

Box plot of quantified copies-per-cell (median, 10th, 25th, 75th and 90th percentiles) GAPDH (n = 144), TNFα (n = 56), IL-1β (n = 32), and TLR4 (n = 86) of 192 analyzed cells

p = 0.003
p = 0.014
Conclusions

- Reliable quantification down to 10 copies per cell is technically possible.
- Higher variation on RNA level and at low copy number.
- Major impact of sampling and handling (evaporation).
- Nearly NO inhibition or interference through LYM (=> has to be proven for each tissues or biological matrices !!!)

natural mRNA variance >> technical variability
cellular expression noise >> quantification noise

- To guarantee valid results, GLP and the application of the MIQE guideline is recommended.

Future questions:
- Valid sample integrity control?
- Cell dependent variability?
- Data analysis in single-cell expression profiling?

Biomarkers on various “-ome” layers

gene

transcription

splicing

translation

post-translatory modifications

DNA => pre-mRNA => mRNA => protein => function

DNA methylatation

DNA => pri-miR => pre-miR => miR

micro-transcriptome

transcriptome

proteome

metabolome

genome

epi-genome

spliceosome

micro-transcriptome
Gynecomastia: development of abnormally large mammary glands

Estrogens => development of primary and secondary sexual characteristics

Prolactin => breast development, milk production

Doping!
Classical analytical methods for drug/hormone residue analysis

Use of specific anabolic agents in agriculture is licensed in Canada, USA, Australia, South Africa ... Since 1988 the use of anabolic agents is prohibited in the EU!

Anabolica misuse => sensitive screening and test system are established in EU

- RIA
- ELISA (EIA)
- LC-MS/MS
- GC-MS
- GC-MS/MS

⇒ analysis of specific compounds
⇒ analysis of known anabolics

AIM

Establishment of a new hormone screening method

- Screening of physiological response
  - in various hormone sensitive tissues
  - in easy accessible tissues, e.g. body liquids, blood, and hair roots

- Regulative impact of hormones on transcriptome
  - qRT-PCR = “most sensitive” detection method
    (⇒ 1 molecule detection limit)

- Evaluation and analysis of regulated genes
  mRNA & microRNA transcript candidates

- Candidate verification
  - via statistical methods, e.g. HCA or PCA
  - via independent methods (NGS)

- Specification of “Expressed Biomarkers”
- Identification of “Gene Networks”
Functional target gene groups

- steroid receptors
- steroid hormone metabolism
- proliferation / apoptosis
- angiogenesis
- tissue remodelling
- immune relevant factors
**Analytical work flow**

mRNA and microRNA quantification MIQE compliant

- RNA extraction
- RNA integrity control
- RT-qPCR
- Gene regulation
- Statistical evaluation
- Data analysis

### mRNA expression profiling results

from heifers ovarian samples

<table>
<thead>
<tr>
<th>functional group</th>
<th>gene</th>
<th>x-fold regulation</th>
<th>p-value</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>receptors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>1.67</td>
<td>0.028</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ERβ</td>
<td>8.19</td>
<td>0.010</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>LHR</td>
<td>0.29</td>
<td>0.004</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>FSHR</td>
<td>3.61</td>
<td>0.026</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Aik-6</td>
<td>0.56</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flt-1</td>
<td>0.64</td>
<td>0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGF-1IR</td>
<td>2.03</td>
<td>0.003</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>PFR</td>
<td>2.20</td>
<td>0.026</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td><strong>angiogenesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VEGF 120</td>
<td></td>
<td>0.34</td>
<td>0.021</td>
<td>*</td>
</tr>
<tr>
<td>VEGF 164</td>
<td></td>
<td>0.21</td>
<td>0.015</td>
<td>*</td>
</tr>
<tr>
<td>FGF-2</td>
<td>2.06</td>
<td>0.036</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>ANGPT-2</td>
<td>2.32</td>
<td>0.015</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>MMP-2</td>
<td>2.46</td>
<td>0.007</td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>TIMP-2</td>
<td></td>
<td>0.24</td>
<td>0.026</td>
<td>*</td>
</tr>
<tr>
<td><strong>proliferation/apoptosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snaid 2</td>
<td>1.27</td>
<td>0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c-jun</td>
<td>2.00</td>
<td>0.052</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p53</td>
<td>1.68</td>
<td>0.021</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>IGF-1</td>
<td>2.34</td>
<td>0.026</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>BMP-2</td>
<td>2.64</td>
<td>0.005</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td><strong>steroid synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSA1</td>
<td>0.53</td>
<td>0.021</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>HSD17B3</td>
<td>0.61</td>
<td>0.024</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td><strong>others</strong></td>
<td>Inhibit A</td>
<td>4.67</td>
<td>0.003</td>
<td>**</td>
</tr>
</tbody>
</table>

Becker C. et al., HMBCI 2010
### microRNA expression profiling results

**microRNA qPCR array – 2x 384 miR set from Exiqon**

<table>
<thead>
<tr>
<th>miR-</th>
<th>x-fold regulation (2$^{-ΔΔCq}$)</th>
<th>p-value</th>
<th>significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR12*</td>
<td>3.16</td>
<td>0.006</td>
<td>**</td>
</tr>
<tr>
<td>miR5</td>
<td>2.16</td>
<td>0.010</td>
<td>*</td>
</tr>
<tr>
<td>miR6</td>
<td>1.23</td>
<td>0.010</td>
<td>*</td>
</tr>
<tr>
<td>miR7</td>
<td>0.92</td>
<td>0.029</td>
<td>*</td>
</tr>
<tr>
<td>miR7-5p</td>
<td>2.05</td>
<td>0.030</td>
<td>*</td>
</tr>
<tr>
<td>miR10</td>
<td>1.20</td>
<td>0.032</td>
<td>*</td>
</tr>
<tr>
<td>miR103</td>
<td>1.50</td>
<td>0.030</td>
<td>*</td>
</tr>
<tr>
<td>hsa-miR-let-7f-2</td>
<td>1.23</td>
<td>0.032</td>
<td>*</td>
</tr>
<tr>
<td>miR152</td>
<td>1.62</td>
<td>0.041</td>
<td>*</td>
</tr>
<tr>
<td>miR49</td>
<td>1.97</td>
<td>0.036</td>
<td>*</td>
</tr>
<tr>
<td>miR744*</td>
<td>1.71</td>
<td>0.046</td>
<td>*</td>
</tr>
<tr>
<td>miR215</td>
<td>1.61</td>
<td>0.049</td>
<td>*</td>
</tr>
<tr>
<td>miR572</td>
<td>39.49</td>
<td>0.054</td>
<td>*</td>
</tr>
<tr>
<td>miR138</td>
<td>3.42</td>
<td>0.090</td>
<td>*</td>
</tr>
<tr>
<td>miR192</td>
<td>1.49</td>
<td>0.091</td>
<td>*</td>
</tr>
<tr>
<td>miR1247</td>
<td>1.79</td>
<td>0.096</td>
<td>*</td>
</tr>
<tr>
<td>miR194</td>
<td>1.00</td>
<td>0.087</td>
<td>*</td>
</tr>
<tr>
<td>miR505</td>
<td>1.52</td>
<td>0.094</td>
<td>*</td>
</tr>
<tr>
<td>miR146-5p</td>
<td>1.27</td>
<td>0.090</td>
<td>*</td>
</tr>
<tr>
<td>miR125a-3p</td>
<td>4.43</td>
<td>0.089</td>
<td>*</td>
</tr>
</tbody>
</table>

**miRNA regulation in bovine liver under the influence of TBA+E2 up-regulated genes**

- miR 378: promotes cell survival, tumor growth, angiogenesis
- miR 412: regulates uterine leiomyomas growth, benign uterine smooth muscle tumor
- miR 15a: tumorgenesis, unarrested cell cycle

---

### microRNA expression profiling

**single assays (Qiagen) vs. microRNA qPCR array (Exiqon)**

<table>
<thead>
<tr>
<th>miR-</th>
<th>x-fold regulation (2$^{-ΔΔCq}$)</th>
<th>p-value</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR15a</td>
<td>0.49</td>
<td>0.023</td>
<td>*</td>
</tr>
<tr>
<td>miR20a</td>
<td>0.49</td>
<td>0.023</td>
<td>*</td>
</tr>
<tr>
<td>miR27b</td>
<td>1.97</td>
<td>0.079</td>
<td>*</td>
</tr>
<tr>
<td>miR29c</td>
<td>1.30</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>miR34a</td>
<td>1.30</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>miR103</td>
<td>1.40</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>miR106a</td>
<td>1.74</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>miR138</td>
<td>0.75</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>miR181c</td>
<td>0.72</td>
<td>0.094</td>
<td></td>
</tr>
<tr>
<td>miR320d</td>
<td>1.15</td>
<td>0.094</td>
<td></td>
</tr>
</tbody>
</table>

**P < 0.05**

**P < 0.1 (trend of regulation)**
How to infer gene networks from expression profiles

Mukesh Banai,1,2, Vincenzo Belcastro,2,3, Alberto Ambesi-Impiombato2,4 and Diego di Bernardo1,2.

1 Translational and Computational Biology Unit, Laboratory of Molecular Medicine, National Institute of Genetics, Casaccia Research Centre, CNR, Rome, Italy
2 Department of Biomedical Sciences, Faculty of Medicine and Surgery, University of Bologna, Bologna, Italy
3 Department of Information Engineering, University of Bologna, Bologna, Italy
4 Department of Pharmacology and Biophysics, Faculty of Medicine and Surgery, University of Bologna, Bologna, Italy

Infering, or ‘reverse-engineering’, gene networks can be defined as the process of identifying gene interactions from experimental data through computational analysis.

**Coexpression networks and clustering algorithms**

Clustering, although not properly a network inference algorithm, is the current method of choice to visualise and analyse gene expression data. Clustering is based on the idea of grouping genes with similar expression profiles in clusters (Eisen et al., 1998). Similarity is measured by a distance metric, as for example the correlation coefficient among a pair of genes. The number of clusters can be set either automatically (transcriptional regulations; gene regulations).

![WBC expression data analysis](image)

**WBC expression data analysis**

*mRNA data*

Significantly regulated genes in heifer blood (16 days after treatment)
Vaginal epithelial cell expression data analysis

PCA

HCA

Combined WBC & vaginal epithelial cell data analysis

=> integrative functional informatics
NGS => Conformational study
Finding new biomarker candidates

Experimental setup

Alteration of gene regulation in bovine liver by treatment with anabolic steroids

- Group I: control animals (n=2)
- Group II: treated animals (n=3)

RNA isolation

sequencing

data analysis

* mapping
* expression analysis

NGS results - sequence statistics

- number of sequences: > 25 Mio.
- sequence length: 76 bases
- successfully mapped bovine genes: 16,947
- successfully mapped transcripts: 45,832
- significantly regulated genes: 9,000 (control vs. treatment group)
- gene regulation at least 2-fold: 306

- screening for significantly regulated genes
  - with high relative expression change, more than 2-fold
  - and high expression value

<table>
<thead>
<tr>
<th>Genes (ex. 10^4 total)</th>
<th>Transcripts (ex. 10^7 total)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>NE=0</td>
<td>1,315</td>
</tr>
<tr>
<td>NE=0.5</td>
<td>1,064</td>
</tr>
<tr>
<td>NE=0.1</td>
<td>4,103</td>
</tr>
<tr>
<td>NE=1</td>
<td>510</td>
</tr>
<tr>
<td>NE=10</td>
<td>567</td>
</tr>
</tbody>
</table>
Verification of 20 regulated genes via RT-qPCR (out of 40 selected)

<table>
<thead>
<tr>
<th>no</th>
<th>acc. no.</th>
<th>gene name</th>
<th>x-fold NGS</th>
<th>sign NGS</th>
<th>x-fold qPCR</th>
<th>p-value PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation x-fold NGS (n=2/3) and fold PCR (n=9)

\[ y = 0.301x + 3.315 \]

PCA - qPCR data

obtained from 11 significantly regulated genes chosen by literature

Blue diamonds: control animals
Red triangles: treated animals

PCA and HCA - qPCR data
obtained from 20 significantly regulated genes chosen by NGS

Blue diamonds = control animals
Red triangles = treated animals

All 9 untreated control animals vs. one treated animal

treated animal no. 5
treated animal no. 8
Proof of concept analysis!

Heifers treated with Synovex plus
200mg Trenbolon Azetat + 28mg Östradiol Benzoat
sample from an animal trial from 1998

Pour-on application of steroid hormone cocktail on calves

Red stars: 3-fold treatment
Pink diamonds: 1-fold treatment
Blue triangles: controls

BiblioSphere
Cocitation based network of genes regulated at least 2 fold between control and treatment. (and min NE of 0.5 in at least one condition)

Genes from input list are colored based on ratio between control and treatment. Red – induced, blue – repressed with treatment.

White boxes represent transcription factors that are co-cited with at least 10 input genes.
Green lines indicate potential binding sites in the promoter region.
Conclusion

- Tissues that are directly influenced by steroid hormones with abundant receptor concentrations are more sensitive and show higher gene regulations => hence better for the biomarker identification
- transcriptomics analysis discovered promising expression changes
  - mRNA served as first biomarker candidates
  - additional candidates via microRNA profiling
  - verification of existing candidates via NGS RNA-Seq
  - new mRNA candidates via NGS RNA-Seq

**Integrative biomarker discovery approach on transcriptome level:**
- Significantly regulated transcripts (mRNA and microRNA)
- in various tissues (WBC, liver, and vaginal smear)
- by different analysis methods (qRT-PCR, qPCR array, and NGS RNA-Seq)
- by various statistical algorithms (PCA and HCA)

⇒ The more integrative biomarker candidates are discovered the better a prediction of hormone treatment is possible

Acknowledgements

Martina Reiter
Benedikt Kirchner
Helena Müller
Claudia Holzhauer
Wolfgang Mann
Irmgard Riedmaier
Vladimir Benes
Heinrich HD Meyer

Free GenEx download => [GenEx.gene-quantification.info]