Supplementary Material

A real time PCR assay for comparative quantification of methylated alleles (QAMA): analysis of the retinoblastoma locus [retinoblastomas]

Zeschnigk et al.

Statistical model

From the RT-PCR experiments the ratio of two specimen is determined via the ΔCT method. From this ratio the fraction of, say, methylated DNA is to be inferred. For any PCR experiment of two DNA species we have:

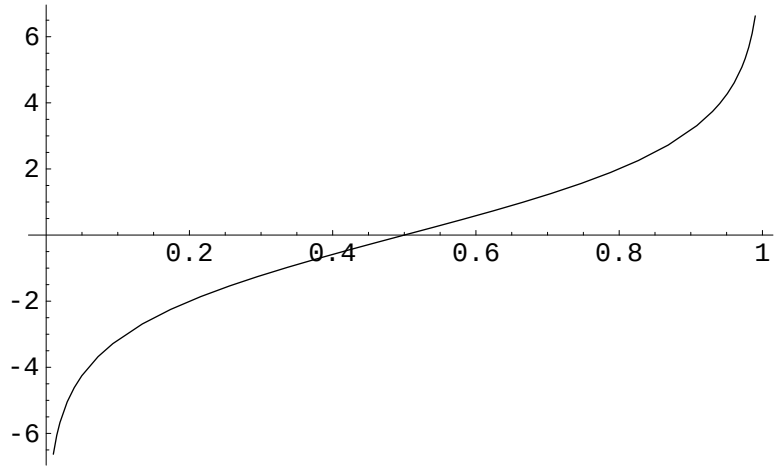
$$r = \frac{dc_0e^t}{dc_1e^t} \tag{1}$$

Here r is the ratio of concentrations, c_0, c_1 are the initial DNA concentrations of two specimen, e is the PCR efficiency (hence $1 < e < 2$), d is a constant of proportionality and t represents a cycle number in the exponential phase of amplification. Obviously, to determine the ratio of two DNA concentrations both only have to be measured up to a common constant of proportionality. In the situtation of the experiments we consider methylated and unmethylated alleles. c_0 gives the fraction of methylated alleles and $c_0 = 1 - c_1$. For a fixed t formula 1 is plotted in figure 1 for varying c_0.

Since different probes are used we may assume different constants of proportionality in measuring the signals from these specimen, say d_0, d_1. Furthermore we assume nonlinear effects that become larger as c_0 tends either to 0 or 1 in the following formula:

$$r = \frac{d_0}{d_1}\left(\frac{c_0e^t}{(1-c_0)e^t}\right)^b \tag{2}$$

Figure 1: Relationship of fraction c of one DNA species with another DNA species having concentration $1-c$ to the ratio of concentrations $c/(1-c)$



As c_0 deviates from $\frac{1}{2}$ the exponent b weights the ratio of concentrations (for $c_0 = \frac{1}{2}$ the quotient is 1). This weighting might represent interactions of PCR amplification of the two DNA species, different efficiencies or non-linear effects in the measurement process. We do not try to model any individual effect, since the limited amount of data allows only to infer few parameters. As we see below a good fit is achieved with the present model.

Finally, we perform the analysis on a logarithmic scale as such an analysis best fits the data model. With $c = c_0$, $a = ln(d_0) - ln(d_1)$ we have:

$$ln(r) = ln\left(\frac{d_0}{d_1}\left(\frac{c_0 e^t}{(1-c_0)e^t}\right)^b\right) \tag{3}$$

$$= ln(d_0) - ln(d_1) + ln\left(\frac{c_0 e^t}{(1-c_0)e^t}\right)^b \tag{4}$$

$$= a + b\, ln\left(\frac{c}{(1-c)}\right) \tag{5}$$

Derived from the experiments used for the standard curve, we get a list of pairs of

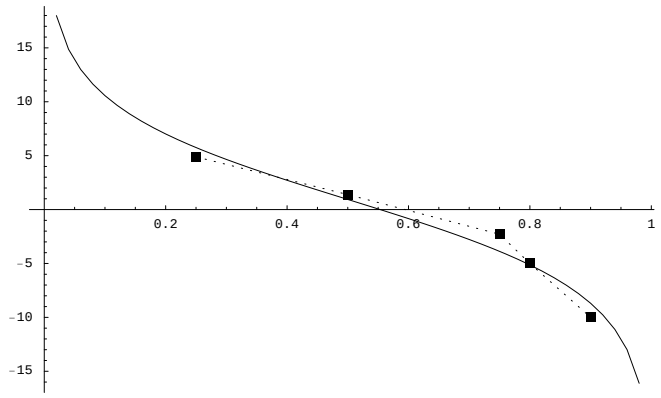
ratios and fractions is given, denoted $D = ((r_0, C_0), ..., (r_n, C_N))$ we now choose a and b such that formula 3 best fits to the data D. The fitting is done by minimizing sum of squares of differences.

Example

We have implemented a software package in Mathematica (you need the Mathematica software package from Wolfram Research: `http://www.wolfram.com`). Evaluate the cells in the notebook `regressMethylation.nb`. Open a new notebook and enter your data in pairs as explained above and shown in the `In[13]` line of figure 2. The next part shows the fitted curve for the data together with the data. Finally, `Out[14]` gives the formula to used to compare fractions from ratios (with y being a ratio) and further diagnostic output from the statistical procedure.

Figure 2: Example Session to determine parameters *a* and *b* of a experiment


```
In[13]:= d = {{0.25, 4.89}, {0.5, 1.39}, {0.75, -2.29}, {0.8, -5}, {0.9, -10}};
         regressExpression[d]
```




```
Out[14]= { 1/(1 + 2^(-0.329224 (0.938677 - y))), {a → 0.938677, b → -3.03745},

                                  Estimate     SE           TStat        PValue
 {ParameterTable → 1              0.938677     0.693855     1.35284      0.269032 ,
                   Log[x/(1-x)]/Log[2]  -3.03745     0.355262     -8.54989     0.00336205

  RSquared → 0.960578, AdjustedRSquared → 0.947438, EstimatedVariance → 1.7325,

                          DF     SumOfSq     MeanSq      FRatio      PValue
  ANOVATable → Model      1      126.647     126.647     73.1006     0.00336205},
               Error      3      5.1975      1.7325
               Total      4      131.844

                                  Estimate     SE           TStat        PValue
 {ParameterTable → 1              14.5321      2.94117      4.94093      0.0159016 ,
                   x              -27.7627     4.31314      -6.43677     0.00760261

  RSquared → 0.932481, AdjustedRSquared → 0.909975, EstimatedVariance → 5.15308,

                          DF     SumOfSq     MeanSq      FRatio      PValue
  ANOVATable → Model      1      213.502     213.502     41.432      0.00760261},
               Error      3      15.4592     5.15308
               Total      4      228.962

 {-3.23599, 57.2404}}
```