Supplementary information

Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression

Evgeny A. Moskalev [1]*, Mikhail G. Zavgorodnij [2], Svetlana P. Majorova [3], Ivan A. Vorobjev [4], Pouria Jandaghi [1], Irina V. Bure [1] and Jörg D. Hoheisel [1]

[1]Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany,

[2]Functional Analysis and Operational Equations, Voronezh State University, University square 1, 394006 Voronezh, Russia,

[3]Higher Mathematics and Physical and Mathematical Modeling, Voronezh State Technical University, Moskovsky avenue 14, 394026 Voronezh, Russia,

[4]Functional Morphology of Hemoblastoses, National Hematology Research Centre of Russian Academy of Medical Sciences, Novy Zykovsky passage 4a, 125167 Moscow, Russia.

Derivation of equation 2 of a hyperbolic best fit curve

Equation of a hyperbolic best fit curve as suggested by Warnecke *et al.* (9):

$$\textit{Equation 1:}\quad y = \frac{100bx}{bx - x + 100}$$

In order to find the family of hyperbolic curves, which pass through the extreme calibration points with abscissae $x = 0$ and $x = 100$, a general equation of a hyperbola is considered to be:

$$y = \frac{ax + c}{dx + e}$$

where *a, c, d, e* are arbitrary parameters. Fitting the parameters in order to include the 0% and 100% methylation points – A (0; y_0) and B (100; y_1) – and substitution of the coordinates of A and B to the equation leads to a system of two algebraic equations:

$$\begin{cases} y_0 = \frac{c}{e} \\ y_1 = \frac{100a + c}{100d + e} \end{cases} \quad \begin{cases} c = y_0 e \\ 100a + c = 100y_1 d + y_1 e \end{cases} \quad \begin{cases} c = y_0 e \\ a = 100y_1 d + \frac{(y_1 - y_0)e}{100} \end{cases}$$

Substitution of the parameters *a* and *c* results in the equation:

$$y = \frac{\frac{100y_1dx + (y_1 - y_0)ex}{100} + y_0e}{dx + e}$$

Further simplification leads to

$$y = \frac{\frac{100xy_1d + (y_1 - y_0)ex + 100y_0e}{100}}{dx + e}$$

or

$$y = \frac{100xy_1d + (y_1 - y_0)ex + 100y_0e}{100(dx + e)}$$

Division of nominator and denominator by *e*:

$$y = \frac{\left(\frac{100xy_1d}{e}\right) + (y_1 - y_0)x + 100y_0}{\frac{100dx}{e} + 100}$$

Let $100\frac{d}{e} = b - 1$, then

$$y = \frac{(b - 1)xy_1 + (y_1 - y_0)x + 100y_0}{(b - 1)x + 100}$$

Simplification results in

$$y = \frac{bxy_1 + xy_1 - xy_1 - y_0x + 100y_0}{bx - x + 100}$$

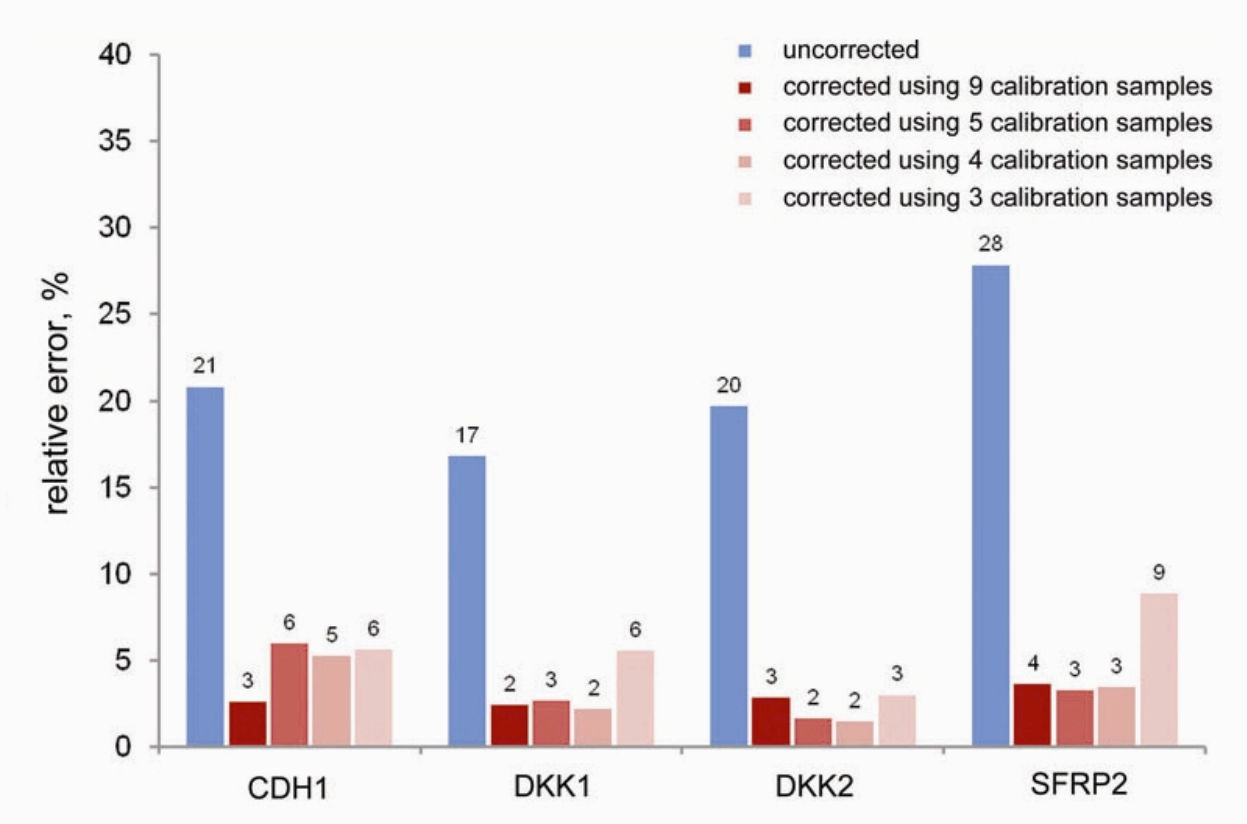
The result is equation 2:

$$y = \frac{(by_1 - y_0)x + 100y_0}{bx - x + 100}$$

If $y_0 = 0$ and $y_1 = 100$ – that is A (0; 0) and B (100; 0) – equation 2 is simplified to equation 1:

$$y = \frac{(100b - 0)x + 100 * 0}{bx - x + 100} \quad \text{or} \quad y = \frac{100bx}{bx - x + 100}$$

Supplementary Figure S1. Comparison of the average relative errors of the bias-correction procedure for *CDH1, DKK1, DKK2* and *SFRP2* based on a different number of calibration samples. The blue bars represent the relative errors of the raw data. The red bars show the corrected data; left to right, the following sets of calibration samples were used: 0, 12.5, 25, 37.5, 50, 62.5, 75, 87.5 and 100% methylation (nine-sample set); 0, 25, 50, 75, 100% (five-sample set); 0, 25, 75, 100% (four-sample set) and 0, 50, 100% (three-sample set).



Supplementary Figure S2. Correction of methylation degrees for *CDH1, DACT1, DKK1, DKK2, DKK3, DKK4, SFRP2, SFRP3* in leukemic cell lines MEC-1, EHEB and CD19^{+} B cells of healthy donors by means of cubic polynomial regression using nine control DNA samples.
(a) CpG maps of the interrogated regions. Vertical bars indicate the positions of CpG dinucleotides. The positions of the first exons are shown as black rectangles. The arrows indicate transcriptional start sites. The red bars (denoted "PYRO") specify the CpG sites quantified by pyrosequencing. **(b)** The diagrams show uncorrected (blue) and corrected (red) average methylation values of the eight selected gene fragments in cell lines MEC-1 (top), EHEB (middle) and CD19^{+} B cells of five healthy individuals (bottom; the average methylation percentages of the five CD19^{+} samples are shown). Some genes are aberrantly hypermethylated in both cell lines and essentially unmethylated in normal CD19+ B cells. Although reportedly hypermethylated in other tumour entities, *DACT1* was unmethylated in all the samples studied.