

**SUPPLEMENTARY MATERIALS**

**Software, Calculations, Graphical Interface Program overview**
- The data is imported from Excel spreadsheet using special wizard; no template for submitted data needed. Data can be raw Ct values, or processed, e.g. normalized data;
- Several groups of patients (Target groups) can be specified. Any group can be designated as “Control”; differentiation of all other groups from the Control group can be evaluated;
- ROC parameters are calculated for any miRNA combination, including complex biomarkers, comprising more than one miRNA as numerator, denominator or both;
- Several kinds of customized graphs are displayed, including Points and Box Plots for a set of data, 2D graphs for two markers, ROC graphs for several markers and their combinations, histograms.

**Data Flow and Technology**
- The software is created in .Net V4.0 as a WinForm application, using C# and VS 2010 / VSS as Development environment. For Math and Statistic calculations, MS Excel COM interface is utilized. For histogram support, MathIridium library is used.
- After importing, data converted to internal format and saved for future use in this format, preventing any inadvertent data altering. Along with source data, normalization settings & results, the last status of data analysis (Control and Target groups), are saved. For Markers and Targets, aliases are allowed.

**Normalization**
Calculations of normalized values is performed according to the following equation:
\[ C = 2^{\left(\text{Ct [miRNA1]} - \text{Ct [miRNA2]}\right) } \times M, \]
where \( \text{Ct [miRNA2]} \) – experimentally determined Ct value for a potential denominator;
\( \text{Ct [miRNA1]} \) – experimentally determined Ct value for a potential numerator;
\( M \) is a constant multiplier, chosen, so that to keep the normalized values C in a computationally convenient range, 10 – 1000.

**Statistics**
Along with regular data set parameters (average, median, standard deviation, finding outliers), some parametric and non-parametric statistics are used:
- Chi-square test with histograms – to define normality of data set;
- T – test (Student) of significance for normally distributed data;
• Pearson correlation coefficient for normally distributed data;
• Mann-Whitney test of significance for other distributions;
• Spearman’s (Rank) association coefficient for other distributions.

For calculation of ROC parameters probabilistic approach is used. The calculated parameters include AUC, Sensitivity, Specificity, Accuracy, Matthews correlation coefficient (MCC), F1 score, and combined value as a weighted sum of all parameters.

Conversions
Linear approximation is used for each data set probability integral function. Every sample gets two probability values: to belong to Control, and to belong to Target group. These probability values hold information about a position of a given data point with respect to other points in the dataset and, rather than experimental raw data, are used for statistical analysis (correlation, significance of separation, combination). The use of these probability values allows one to easily perform operations with values, which are all in the range of 0 to 1.0; to conveniently make “once for all” changes in the decision algorithm, such as shifting cutoff points; and to create a complex function of biomarkers, comprising several miRNA pairs, etc.

Decision
Binary decision is based on comparing an actual value and value of cutoff point. In probabilistic approach, we are using probability of the sample to belong to the Target group with probability assigned to cutoff point. “Equilibrium” point, i.e. point where probabilities to belong to Control and Target groups are equal, is selected as the initial value for the cutoff point on ROC curve. Moving this point right or left means creating preference for Specificity over Sensitivity, or vice versa. An important aspect here is that this preference can be introduced for all biomarkers by including one cutoff value in the decision making.

Graphical Interface: a Screenshot
Figure S1: Ratios of miRNA levels (biomarker pairs) in plasma of MCI patients and age-matched controls. The ratios of concentrations of miRNA pairs in plasma samples of MCI and age-matched donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratios of various miRNA were calculated as $2^{-\Delta\Delta Ct} \times 100$. See the Fig. 1 legend for the description of the statistical analysis.
Figure S2. Comparison of miRNA biomarker pairs in plasma of Group 1 (30-50 years old, “CY”) and Group 2 (70-80 years old, “AMC”) individuals with normal cognitive functions. The ratios of concentrations of miRNA in plasma samples of Group1 (30-50 years old, CY) and Group2 (70-80 years old, AMC) donors with normal cognitive function, 20 samples in each group, were measured by RT-PCR and the ratio of various miRNA was calculated as $2^{-\Delta Ct} \times 100$. See the legend to Fig. 1 for the description of the statistical analysis.