The qPCR data statistical analysis

Ramon Goni[1]*, Patricia García[1] and Sylvain Foissac[1]

[1]Integromics SL, Madrid Science Park, 2 Santiago Grisolía, 28760 Tres Cantos, Spain
*To whom correspondence should be addressed. Email: ramon.goni@integromics.com

Abstract: Data analysis represents one of the biggest bottlenecks in qPCR experiments and the statistical aspects of the analysis are sometimes considered confusing for the non-expert. In this document we present some of the usual methods used in qPCR data analysis and a practical example using Integromics' RealTime StatMiner, the unique software analysis package specialized for qPCR experiments which is compatible with all Applied Biosystems Instruments. RealTime StatMiner (*http://www.integromics.com/StatMiner*) uses a simple, step-by-step analysis workflow guide that includes parametric, non-parametric and paired tests for relative quantification of gene expression, as well as 2-way ANOVA for two-factor differential expression analysis.

Keywords: qPCR, data analysis, RealTime StatMiner

I. INTRODUCTION

Since the invention of real-time PCR (qPCR), thousands of high-impact studies have been conducted and published using qPCR technique (Heid et al. 1996; Higuchi et al. 1993; VanGuilder, Vrana, and Freeman 2008). Because it is highly sensitive, qPCR is the preferred method for microarray data validation (Canales et al. 2006); however the most exciting applications have been in the discovery of new biomarkers and in diagnostic prediction (Gillis et al. 2007). Despite the fact that this technique has been widely used by researchers, there are several obstacles to analyzing the vast amounts of data generated.

Before data can be generated and analyzed, an hypothesis needs to be formed and the experiment designed. The success of a project depends on fundamental rules in the implementation of quality controls (review plates, filter outliers, removal of incorrect samples and flag genes undetected), selection of the optimal endogenous controls for normalization and the correct choice of the correct statistical method for the analysis. In this document we describe some of the crucial steps in qPCR data analysis and illustrate statistical notions with a concrete example using the RealTime StatMiner software.

II. BIOLOGICAL SAMPLES

As an example, consider the following experiment: to see the effect of a treatment on miRNA expression in mice, samples are extracted from two tissues (tissue Control-C and tissue Target-T). Additionally, three categories of mice are involved: untreated mice (NT), mice with a low dose of treatment (0.005gr - Low) and mice with a high dose (0.01gr -

High). As a result, this project has two experimental factors (see Figure 1):
1. Mouse tissue (C,T)
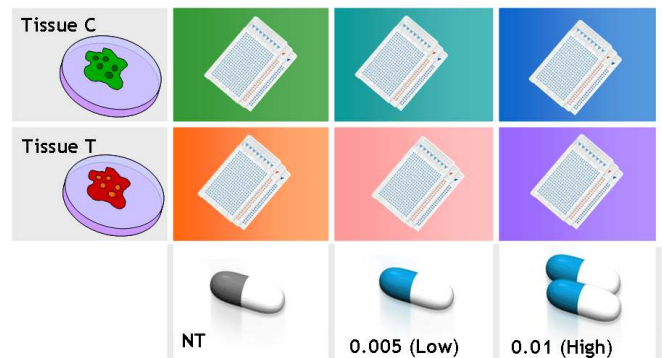2. Treatment dose (NT, Low, High)

Fig. 1. Experimental factors of the project.

III. SETTING THE EXPERIMENTAL DESIGN: FACTORS, GROUPS & SAMPLES

Prior to any experiment, an appropriate experimental design has to be established. Combining the two experimental factors in our previous example, there are six possible scenarios for a given sample (see Table 1):

1. A sample of Control tissue with no treatment: C.NT
2. A sample of Control tissue with low doses: C.Low
3. A sample of Control tissue with high doses: C.High
4. A sample of Target tissue with no treatment: T.NT
5. A sample of Target tissue with low doses: T. Low
6. A sample of Target tissue with high doses: T. High

TABLE I
EXPERIMENTAL DESIGN

SampleName	Tissue	Treatment	Condition
C_NT_1	C	NT	C.NT
C_NT_2	C	NT	C.NT
C_NT_3	C	NT	C.NT
C_Low_1	C	Low	C.Low
C_Low_2	C	Low	C.Low
C_Low_3	C	Low	C.Low
C_High_1	C	High	C.High
C_High_2	C	High	C.High
C_High_3	C	High	C.High
T_NT_1	T	NT	T.NT
T_NT_2	T	NT	T.NT
T_NT_3	T	NT	T.NT
T_Low_1	T	Low	T.Low
T_Low_2	T	Low	T.Low
T_Low_3	T	Low	T.Low
T_High_1	T	High	T.High
T_High_2	T	High	T.High
T_High_3	T	High	T.High

Experimental design of 18 samples using 2 experimental factors. The last column summarizes the biological conditions of the sample

Because the goal is to assess the expression of miRNAs for all configurations, samples representing every scenario are needed. The question then is, are 6 samples enough or are more samples needed for this project? As statistical significance requires multiple measurements, biological replicates are necessary (multiple samples per configuration). The following points can help to estimate the number of required samples:

• No statistical significance can be obtained for a differential expression measurement if only one sample is available for one of the conditions.
• Three is the minimum number of samples per group that is required to detect outliers and to obtain statistical significance.
• If the expected differential expression is high (e.g. a

Fig. 2.   RealTime StatMiner experiment design section. To load the experiment design information (see Table 1) just click on Add Factor and fill every cell with the biological information. Then simply save the information on an external file and Apply the design.

knock-down experiment) three biological replicates can suffice for the test. Conversely, when low differential expression is expected (e.g. gene regulation by miRNAs), more biological replicates may be needed. As it is not always possible to know "a priori" the difference in the expression sometimes it is better to start with 3 biological replicates and add more later when needed.
• The variability of the expression values between measurements from the same condition is an important factor. The lower this variability is, the lower the number of required samples.
• Overall, increasing the number of samples increases the power of the statistical test.

In our example we use three samples per condition (or three biological replicates per group; see Box 1). The project contains 18 samples (six groups x three samples per group). The samples are named using the "experimental condition" as the prefix and the number of "biological replicates" as the suffix (see Table 1 and Figure 2).

IV.   FOLD CHANGE IN qPCR

In every well, the qPCR experiment measures the expression intensity of a certain gene from a sample under specific biological conditions. This measurement is expressed in Cycles to Threshold (Ct) of PCR, a relative value that represents the cycle number at which the amount of amplified DNA reaches the threshold level. Because of the technical variability between experiments the Ct needs to be normalized (see Box 2). Differential expression is done gene by gene by comparing the normalized Ct values (ΔCt) of all the biological replicates between two groups of samples (two biological conditions).

Figure 3 shows the differential expression of the miRNA mmu-miR-25 between the Control tissue without treatment (C.NT) and the Target tissue without treatment (T.NT). Because in every cycle of PCR (Ct value) the amount of DNA is approximately duplicated, the Ct is in the logarithmic scale

BOX 1- Biological replicates and technical replicates
Technical replicates are measurements that are done using exactly the same sample to test the reproducibility of the qPCR technology (instruments, reagents or protocols). Once this is done and potential outliers are removed, technical replicates are usually aggregated to a single measurement. Biological replicates on the other hand are designed to be representative of a general biological condition, therefore they are extracted from different sources (reproducing the experimental conditions). Extracting three different samples of the tissue C from the same mouse only represents a single animal. In order to obtain biological replicates that characterize the *Mus musculus* specie, samples should be extracted from different mice.

(A) qPCR Raw mesurements

Legend: mmu-miR-25, snoRNA135 (EC)

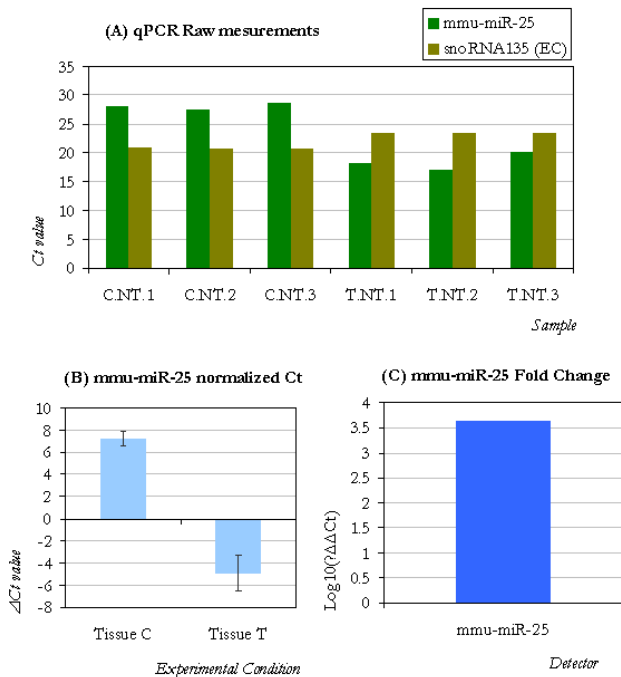**(B) mmu-miR-25 normalized Ct**

(C) mmu-miR-25 Fold Change

Fig. 3. Representation of the process from the measurement to the differential expression of tissue C (untreated), using as the control baseline the tissue T (untreated). (A) Cts for the Endogenous Control snoRNA135 and the detector mmu-mina-25 are calculated using qPCR. (B) Then the Cts are normalized using the Endogenous Control gene. (C) Finally the differential expression of mmu-mina-25 is calculated and represented in Log 10 scale.

fold change of the analysis: the efficiency of the PCR reaction and the absence of expression for a given gene.

• The efficiency of the PCR reaction. Although the number of generated molecules is supposed to double at each cycle of an ideal PCR experiment, in practice, this ratio may be lower. When different targets are not amplified with the same reaction efficiency, the comparison of their expression levels requires some adjustment. Using the TaqMan technology, the efficiency is assumed to be close to 100% (Applied Biosystems 2006), but in other technologies such as SYBR Green the fold change should be adjusted. RealTime StatMiner integrates, in the workflow analysis the functionality of efficiency correction (see the RealTime StatMiner manual; *http://www.integromics.com/StatMiner*).

• The absence of expression for a given gene. When the mRNA quantity of the gene does not exceed a detection threshold, the corresponding Ct value is undetermined or close to the upper limit of the possible range, raising the issue of reproducibility (Nolan, Hands, and Bustin 2006). In such cases the detector should be considered "not detected". The fold

and inversely proportional to the quantity of DNA/RNA. Therefore high ΔCts represent low expression while highly expressed genes have low ΔCts. Comparing the normalized expression (ΔCt) of the two conditions it is possible to calculate the fold change of the expression of the miRNA (−ΔΔCt). The fold change is the expression ratio: if the fold change is positive it means that the gene is upregulated; if the fold change is negative it means it is downregulated (Livak and Schmittgen 2001). There are two factors that can bias the

BOX 2 – Imputation of Ct values
Sometimes the Ct values are undetermined (not detected after certain Cycles) or absent (when no reaction takes place in the corresponding well), which raises a mathematical issue for the analysis of the project. To address this issue, RealTime StatMiner imputes Ct values. Undetermined values are set to a maximum Ct (e.g. 40). If the Ct value is totally absent, an imputation is performed by using the values of the other biological replicates. For example in this project, the Ct value of the detector mmu-miR-30c is 22.5 for C.NT.1, 20.4 for C.NT.2 ; there is no value for C.NT.3. After imputation (and the selection of the median as aggregation method between samples with the same experimental condition) the Ct is 21.4 for mmu-miR-30c in C.NT.3
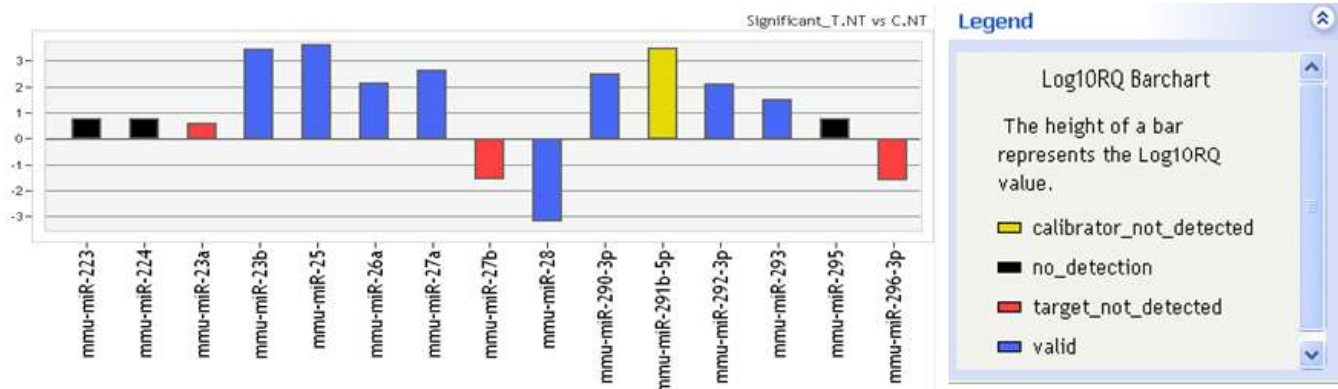

Fig. 4. RealTime StatMiner Fold change results comparing C.NT versus T.NT. Upregulated detectors take positive values while repressed detectors are negative. Detectors in blue are expressed in both tissues, Detectors in yellow are not expressed in C.NT, detectors in red are not expressed in T.NT and those in black are not expressed in either of the two tissues. Regardless of the fold change sign detectors in yellow are upregulated and those in red downregulated (see Box 3 for an explanation of mmu-miR-23a fold change).

change value of a gene that is not expressed in some of the biological conditions may not be reliable and it may produce misleading results, as exemplified in Box 3. RealTime StatMiner can detect and discard unexpressed genes to avoid false results (see Figure 4).

V.   STATISTICAL SIGNIFICANCE

Having a positive fold change suggests that a certain miRNA is upregulated but is this extensible to any other mice? In other words, is the differential expression of this miRNA statistically significant or was the result achieved by chance? The statistical test calculates the p-value of every detector compared in the analysis. According to the gold standard in statistics, a p-value lower than 0.05 is considered significant (Fisher 1925), although some authors set the cut-off at 0.01. Statistics are widely utilized in most of the works published; even so it is unclear to some qPCR users how to apply these methods.

VI.   PARAMETRIC OR NON-PARAMETRIC TEST

A statistical test can be parametric or non-parametric. To know which of the two types of tests to choose one question needs to be answered: does the Ct value of every detector in the project follow a "normal" distribution? In other words, would the distribution of the Ct values for a single detector results in a histogram similar to the plot A in Figure 5 if the experiment was done with an large number of mice? The

BOX 3 – Apparently inverted fold change signal from Ct values analysis

In Figure 4 the miRNA mmu-miR-23a indicates a positive fold change for tissue T (over-expression) despite of the average Ct values in tissue T [35.1, 35.3, 35.3] being higher (less expressed) than those of tissue C [34.8, 34.5, 34.6]. The reason for this apparent contradiction lies in the normalization process (ΔCt computation) which is a key step in the analysis. Ct values without normalization correction are meaningless. The reference gene used for normalization in this project is snoRNA135, with a Ct of [21.0, 20.8, 20.8] in tissue C and a Ct of [23.5, 23.5, 23.4] in tissue T. The computed ΔCt values are [13.8, 13.7, 13.8] for tissue C and [11.6, 11.8, 11.9] for tissue T, producing a positive −ΔΔCt of 2.0 or Log10RQ= 0.602.

RealTime StatMiner flags the detector in tissue T as "not detected" with a red color because its Ct value is higher than the cut-off of 35, hence it is not considered as a reliable Ct expression value.

As a general rule, conclusions regarding differential expression can only be drawn when the compared Ct values are produced by reliable measurements (blue color in the fold change bar).
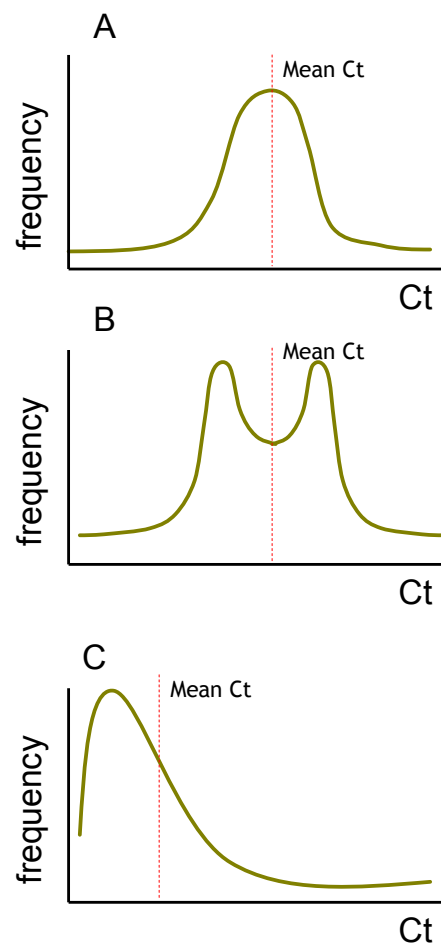

Fig. 5. Possible distributions of the values of a given variable. For a given variable (e.g. gene expression) possible values (e.g. Cts) are represented on the X-axis while the frequency of such value is represented on the Y-axis. A represents a normal distribution; B represents a symmetric distribution with no single pick−therefore not normal; and C represents asymmetric distribution−again, also therefore not normal.

normal distribution (see Plot A in Figure 5) is symmetric and has a bell-shaped curve with a single peak. The parametric test runs under the assumption that the distribution of gene or miRNA Ct value is normal while non-parametric tests do not make such an assumption.

Parametric: moderated t-test

One of the most popular parametric tests is the "Student's t-test". The Student's t-test (many times referred simply as t-test) needs normally distributed variables and is based in the statistical parameters mean and standard deviation. Frequently, the number of biological replicates available is low (three biological replicates in this case) and as a result the standard deviation is not well represented. RealTime StatMiner integrates the "moderated t-test" (see Figure 6), which is a variant of the "t-test" oriented to experiments with few biological replicates. The primary difference between the moderate t-test and the Student t-test is in the calculation of the standard deviation (Smyth 2004).

10.Relative Quantification   [What is this?]

Fig. 6. RealTime StatMiner parametric and non-parametric statistical tests. The test is conducted by selecting the two experimental conditions and setting the p-value cut-off. If the number of detectors is high you can select a multi-hypothesis correction method (see the False Discovery Rate section).

The remaining question is the following: can it be proven that the assumption of normal distribution of the Ct values is correct? A possible method is the Kolmogorov-Smirnov Goodness-of-Fit Test (Chakravarti, Laha, and Roy 1967). But if there are only a few samples the Kolmogorov-Smirnov test cannot be conclusive. When normality is not proven, using a non-parametric test (not assuming normality) reduces the risk of misinterpretation of the results. However, a non-parametric test is less powerful than the parametric test when the data is normally distributed.

In conclusion, the "moderate t-test" is a powerful statistical method that has been widely and successfully used in gene expression analysis and microarray data where normality and other mathematical assumptions are not exact (Smyth et al. 2007). The "moderate t-test" uses a variant of linear models with an empirically moderated estimate of the standard error— effectively borrowing information from the ensemble variance of genes to aid inference about individual genes. This gives improved statistical power for even small sample sizes as referred in Henderson et. al. (2005).

Non-Parametric: Wilcoxon rank sum test

The "Wilcoxon rank-sum test", also referred to in the bibliography as the "Mann-Whitney U-test" (Glover and Mitchell 2008), is the non-parametric alternative of the t-test available in RealTime StatMiner (see Figure 6) and is used when it's been proven that the distribution of the Ct values is

not normal. In the same way that the normality of a distribution can be proven, it is also possible to demonstrate the non-normality of the distribution. One possible method is the Shapiro–Wilk test (Shapiro and Wilk 1965). The assumptions of normality or not normality are difficult to demonstrate in many qPCR projects given the low number of biological replicates available. In the case where normality is uncertain, the Wilcoxon test can be used. Another alternative is to base the assumption on previously published works (Dondrup et al. 2009).

VII. ONE-TAILED OR TWO-TAILED TEST

In differential gene expression experiments, one hypothesis cannot include more than one detector. In other words, there is one hypothesis is per detector (e.g. mmu-miR-25) and per experimental condition mate pair (e.g. tissue T versus tissue C). Every hypothesis can be formulated in three ways:

A)   mmu-miR-25 is more expressed in tissue T than in tissue C.

B)   mmu-miR-25 is less expressed in tissue T than in tissue C.

C)   mmu-miR-25 is more or less (but not equally) expressed in tissue T than in tissue C.

In the case of A) and B) the statistical test should be configured as a one-tailed test, while in the case of C) the test should be two-tailed. The two-tailed statistical test can be implemented by following a standard differential expression

A)

B)

Fig. 7. RealTime StatMiner paired tests. A) For paired experiments select the option "Are your samples paired?" in RealTime StatMiner and associate every sample with the mouse of origin. Samples will be paired according to the animal information. B) Paired t-test is available in the drop-down menu of the differential expression step, where the "Pair Column" of the experimental design is mandatory.

experiment in RealTime StatMiner. Although the "A" hypothesis should ideally not be tested using a two-tailed test, it is still possible to derive some conclusions from that case (see Box 4).

VIII. PAIRED OR NON-PAIRED TEST

There is a special experiment design that requires a specific type of test: the paired test. In the following example, there are two ways to design the project:

A)   Using 18 mice: each sample has an independent source animal.

B)   Using three mice: Samples of two tissues (T & C) are extracted from each mouse before any treatment (NT), after a low dose (0.005gr) and after using a high dose (0.01gr).

In case of "A", the samples are independent and not-paired.

For this type of design a non-paired test is used like the moderated t-test or the Wilcoxon rank sum test (described before). In the case of "B" the samples are dependant and moreover they are paired. In such a scenario RealTime StatMiner integrates the parametric "paired t-test" (Glover and Mitchell 2008).

Paired t-test: The experiment design

If the experiment is paired, every sample belongs to one of the three mice (M1 or M2 or M3). The association between the sample and the mouse of origin is done in the experimental design (see Figure 7). When running the paired t-test in RealTime StatMiner it is mandatory to select the column in the experimental design related to the sample mouse source (see Figure 7).
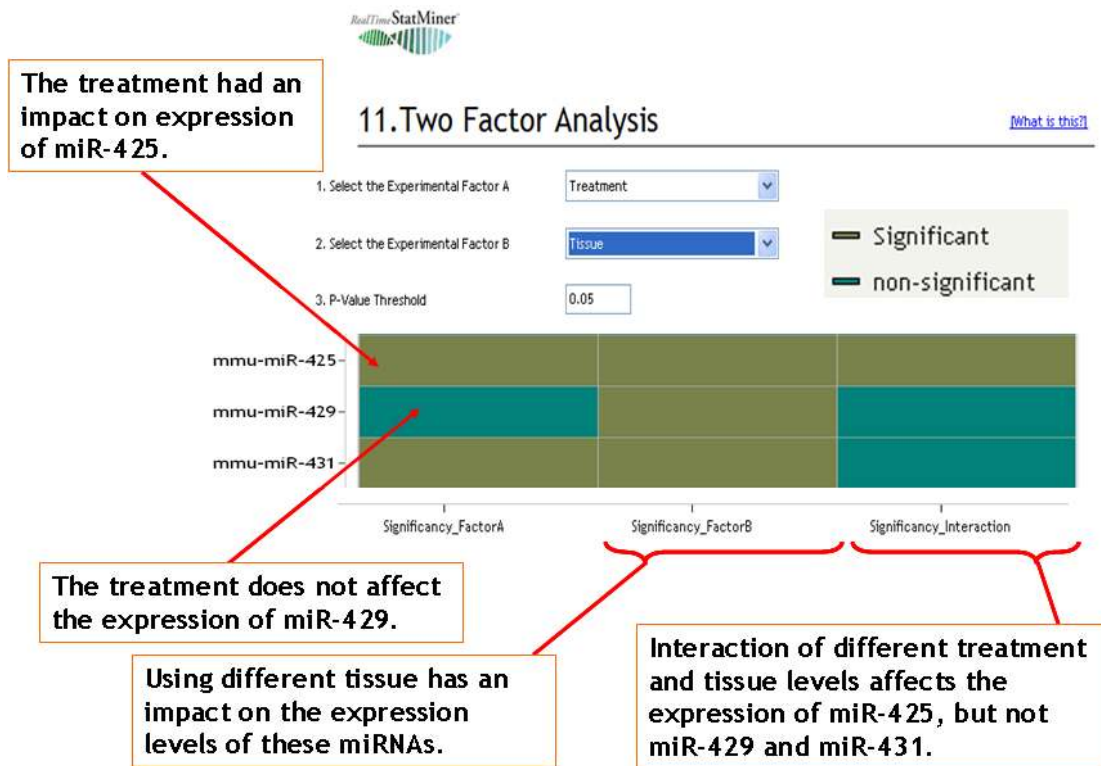
Fig. 8. RealTime StatMiner 2-way ANOVA. Selecting two experimental factors (e.g. treatment dose and tissue type) allows you testing the effect of each of them independently (first and second column) and the effect of the interaction (third column). The legend of the resulting figure shows whether or not the p-values are significant.

IX. THE FALSE DISCOVERY RATE (FDR)

As previously mentioned, the hypothesis is atomic and therefore there is a single hypothesis per detector and per experimental condition comparison (e.g. mmu-miR-25 is more or less expressed in tissue T than in tissue C). Every statistical test is related to a single hypothesis. As a result, when comparing the expression of miRNAs between tissue T and C, the number of tests is equal to the number of miRNAs in the project (hundreds). Increasing the number of tests increases the probability of finding a significant p-value by chance (e.g. performing 100 tests we expect 5 p-values under 0.05 just by chance). To avoid this effect the False Discovery Rate (FDR) negatively adjusts the p-values according to the number of total tests. There are several methodologies to implement the FDR. The most popular are:

• Benjamini-Hochberg (Benjamini and Hochberg 1995): This method assumes that the multi-tests are dependent—that the expression of the detectors has some level of correlation. This assumption is correct in many projects as it is known that genes are biologically co-regulated.
• Bonferroni (Westfall and Young 1993). This approach does not make the assumption that the genes of the project could be directly or indirectly co-regulated. If the dependency between tests (between the expression of the genes) is not assumed, the evidence of the differential expression is more

BOX 4 – What if a two-tailed test is used for a one-tailed hypothesis?
If the following (one-tailed) hypothesis: "mmu-miR-25 is expressed more in tissue T than in tissue C" is tested using RealTime StatMiner (which uses a two-tailed test), there are three possible scenarios. Setting the p-value cutoff to 0.05, the conclusions of the different scenarios are:

1. If the p-value in the two-tailed test is lower than 0.05 and the fold change is positive, then the result of the one-tailed test is also significant for that hypothesis.
2. If the p-value in the two-tailed test is lower than 0.05 and the fold change is negative, then the result of the one-tailed test is always not-significant for that hypothesis.
3. If the p-value in the two-tailed test is higher than or equal to 0.05, then results of the one-tailed test may or may not be significant. The only way to know is to configure the test as one-tailed.

In summary: in some cases when the one-tailed is not available, it is still possible to generate evidence of differential expression for a one-tailed hypothesis using the two-tailed test.

difficult to demonstrate.

In summary using the Bonferroni method the p-values is higher than when using Benjamini-Hochberg. RealTime StatMiner integrates these two methods, as well as the Holm method (Holm 1979)− a sequentially rejective Bonferroni procedure that is much less conservative but that still maintains the experiment-wise error rate. Another alternative method in RealTime StatMiner is the Hochberg method (Hochberg 1988), a more powerful procedure than Holm's method. Instead of testing sequentially, starting with the smallest p-value, Hochberg starts with the largest p-value and stopps when a significant result is obtained and declaring all untested results to be significant.

X. MULTI-FACTOR ANALYSIS

The moderate t-test can compare "experimental conditions" (e.g., tissue C versus tissue T) and perform multi-comparisons (e.g. "not-treated" versus "low treatment and high treatment"). However, it does not allow the comparison of "experimental factors". In other words the moderate t-test cannot prove that the effect of the treatment differs across tissues.

The 2-way ANOVA test (Glover and Mitchell 2008) can demonstrate whether the tissue, the treatment or the interaction between both factors has an impact on gene expression. In this example (as in most of the cases) the comparison is made between two factors. The 2-way ANOVA test is implemented in RealTime StatMiner and produces a p-value for every detector and for each of the following three cases (see Figure 8):

1. The effect of the interaction between tissue and treatment in every detector.

2. The effect of the tissue type in the expression of every detector.

3. The effect of the dose of the treatment in the expression of every detector.

The 2-way ANOVA implemented in RealTime StatMiner will identify those miRNAs that have a significantly different response to the treatment depending on the tissue where they are expressed. In our example the input data is balanced because for every experimental condition there are exactly three biological replicates. The 2-way ANOVA can be balanced or unbalanced (where the number of biological replicates is not equal in every condition). RealTime StatMiner supports both balanced and unbalanced datasets.

XI. RESULTS

The expression of 376 miRNAs is compared between two mouse tissues using the qPCR Applied Biosystems rodent miRNA panel A. 161 detectors of the plate were flagged as non-expressed in Tissue C and 172 were flagged as not expressed in Tissue T, using a cut-off CT of 35 (134 were flagged in both tissues). The project was normalized using the endogenous control snoRNA135. The column "Condition" is used to impute missing values. 62 out of the 178 non-flagged miRNAs where found significantly (p-value < 0.05) differentially expressed (56 upregulated and 6 downregulated), between these two tissues using the parametric paired t-test and adjusting the p-value with the Benjamini-Hochberg FDR method. Finally some samples were treated with a compound using a low dose (0.005gr) and others using a high dose (0.01gr). Running a 2-way ANOVA analysis, we found that for 110 out of the 178 non-flagged miRNAs the treatment significantly affects the expression, (p-value 0.05) depending on the tissue.

SUPPLEMENTARY INFORMATION

You can find the dataset of this project and other relevant information at:

http://www.integromics.com/WhitePaper_qPCR_Statistics

REFERENCES

Applied Biosystems. 2006. Amplification Efficiency of TaqMan Gene Expression Assays. Applied Biosystems Application Note.

Benjamini, Yoav, and Hochberg, Yosef. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. http://dx.doi.org/10.2307/2346101.

Canales, Roger D., Yuling Luo, James C Willey, Bradley Austermiller, Catalin C. Barbacioru, Cecilie Boysen, Kathryn Hunkapiller, et al. 2006. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nature Biotechnology* 24, no. 9 (September): 1115-1122. doi:10.1038/nbt1236.

Chakravarti, I.M., R.G. Laha, and J Roy. 1967. "Handbook of Methods of Applied Statistics." *Handbook of Methods of Applied Statistics*, 1:392-394. John Wiley and Sons.

Dondrup, Michael, Andrea T Hüser, Dominik Mertens, and Alexander Goesmann. 2009. An evaluation framework for statistical tests on microarray data. *Journal of Biotechnology* 140, no. 1-2 (March 10): 18-26.

Fisher, R.A. 1925. Statistical Methods for Research Workers . Oliver and Boyd.

Gillis, A J M, H J Stoop, R Hersmus, J W Oosterhuis, Y Sun, C Chen, S Guenther, et al. 2007. High-throughput microRNAome analysis in human germ cell tumours. *The Journal of Pathology* 213, no. 3 (November): 319-328. doi:10.1002/path.2230.

Glover, Thomas, and Kevin Mitchell. 2008. An Introduction to Biostatistics. 2nd ed. Waveland Pr. Inc, September 30.

Heid, C.A., J Stevens, K. J .Livak, and P.M. Williams. 1996. Real time quantitative PCR. *Genome Research* 6, no. 10 (October): 986-994.

Higuchi, R., C. Fockler, G. Dollinger, and R. Watson. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Bio/Technology* (Nature Publishing Company) 11, no. 9 (September): 1026-1030.

Hochberg, Yosef. 1988. A Sharper Bonferroni Procedure For Multiple Tests Of Significance. *Biometrika* 75: 800-803.

Holm, S. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6: 65-70.

Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T))

Method. *Methods* (San Diego, Calif.) 25, no. 4 (December): 402-408. doi:10.1006/meth.2001.1262.

Nolan, Tania, Rebecca E. Hands, and Stephen A. Bustin. 2006. Quantification of mRNA using real-time RT-PCR. Nat. *Protocols* 1, no. 3 (November): 1559-1582. doi:10.1038/nprot.2006.236.

Shapiro, S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52: 591-611.

Smyth, Gordon K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3: Article3. doi:10.2202/1544-6115.1027.

Smyth, Gordon K, Matthew Ritchie, Natalie Thorne, and James Wettenhall. 2007. Linear Models for Microarray Data . The Walter and Eliza Hall Institute of Medical Research Melbourne.

VanGuilder, Heather D., Kent E. Vrana, and Willard M. Freeman. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques* 44, no. 5 (April): 619-626. doi:10.2144/000112776.

Westfall, Peter H., and Young, S. Stanley . 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. 1st ed. *Wiley-Interscience*, January 5.