APPLICATION NOTE

Gene expression analysis

Efficient analysis of gene expression differences using TaqMan arrays, QuantStudio 6 Pro and 7 Pro systems, and Relative Quantification software

Here we illustrate:

- How Applied Biosystems™ TaqMan® Gene Expression Arrays can be used to analyze expression levels in RNA derived from archived ovarian tumors
- The ease of using the Applied Biosystems™ QuantStudio™ 6 Pro and 7 Pro Real-Time PCR Systems
- The simplified analysis of gene expression data using the free, cloud-based Applied Biosystems™ Relative Quantification (RQ) application

Introduction

Analysis of gene expression levels is a critical part of understanding the molecular basis of the functions of genes, the differentiation of cells, and dysregulation of systems. Although microarrays and RNA sequencing (RNA-Seq) experiments provide enormous amounts of information and are best for very broad, discovery-based experiments, TaqMan Gene Expression Assays are the gold standard for analyzing expression levels in smaller sets of genes.

In many biological pathways and diseases, sets of genes are co-regulated at the transcriptional level. In order to fully understand a phenotype, disease, or the impact of a treatment, scientists often want to analyze a group of functionally linked messenger RNA (mRNA) transcripts. Alternatively, a family of sequence-related genes might show differential transcriptional behaviors that are important for a phenotype or disease.

In order to efficiently analyze groups of mRNAs for a set of genes that are important in biological pathways, phenotypes, or diseases, panels of TaqMan Gene Expression Assays have been developed. These panels allow investigators to focus on a set of genes related to their biological effect, instead of having to pursue more expensive and complicated RNA-Seq or microarray experiments.

Here we demonstrate the use of the Applied Biosystems™ TaqMan® Human Oncogenes-Tumor 96-Well Gene Expression Array in analyzing expression levels using RNA derived from archived ovarian tumor samples (Figure 1). For this analysis, we introduce the use of the QuantStudio 6 Pro and 7 Pro systems and Design and Analysis software, which simplifies setting up and running reactions on the instrument. Finally, we show how the RQ app simplifies QC analysis of gene expression data and provides options for visualizing trends and differences in gene expression levels.

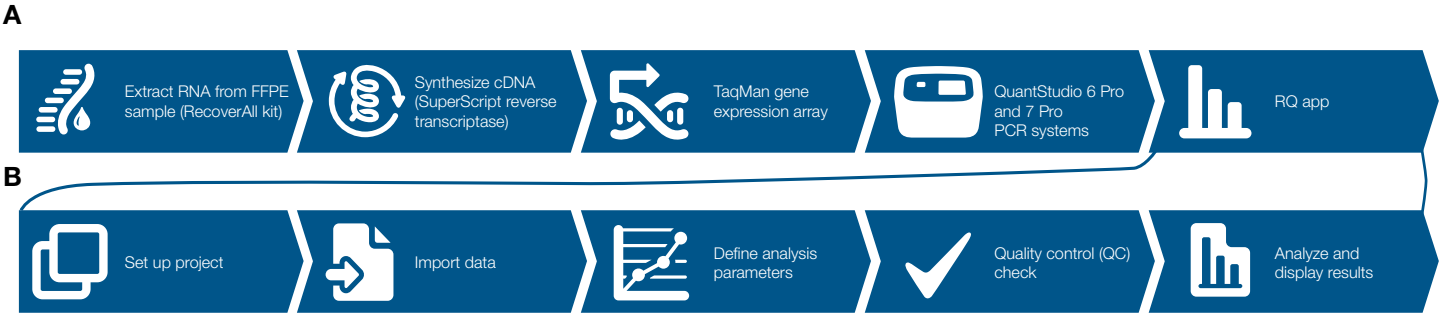


Figure 1. Workflows for gene expression analysis, from RNA extraction from formalin-fixed, paraffin-embedded (FFPE) samples to data interpretation. (A) Complete, end-to-end workflow using products from Thermo Fisher Scientific that facilitate all the steps of analyzing differences in RNA expression levels. **(B)** The freely available, cloud-based RQ app simplifies data analysis and visualization.

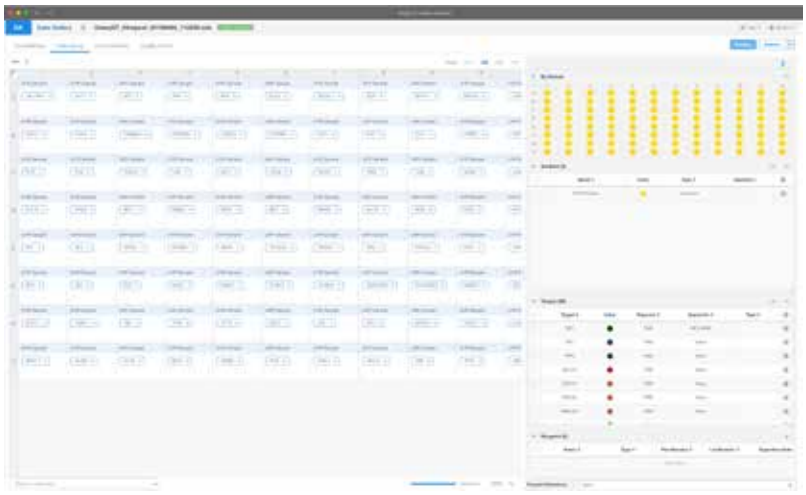
ThermoFisher SCIENTIFIC

Complete, end-to-end workflow

Gene expression analysis involves many different steps. Thermo Fisher has a complete portfolio of products that have been designed to work together to help simplify the workflow, from RNA extraction from FFPE samples through cDNA synthesis and PCR to data analysis. The PCR instrument software was redesigned for the QuantStudio 6 Pro and 7 Pro systems with a cleaner user interface that facilitates plate setup and can run in Microsoft™ Windows™ or Mac™ operating environments. The cloud-based RQ app further simplifies analysis by providing options for visualizing data, normalizing to a gene or group of genes, and calculating expression levels.

We demonstrated this integrated workflow by performing a simple gene expression analysis of tumor samples. Five different ovarian tumor samples were obtained as FFPE samples sliced and mounted on microscope slides. The estimated tumor content in these samples ranged from 55% to 98%. RNA was extracted from single slices using the Invitrogen™ RecoverAll™ kit for FFPE samples. We recovered 0.32–25 µg of total RNA from these samples. No adjacent normal tissue was available from these collections, so universal human reference (UHR) RNA was used as a control. cDNA was synthesized using 100 ng of total RNA and the Invitrogen™ SuperScript™ IV VILO kit. The entire cDNA synthesis reaction was combined with Applied Biosystems™ TaqMan® Gene Expression Master Mix, and dispensed (20 µL/well) onto TaqMan Human Oncogenes-Tumor 96-Well Gene Expression Array Plates. Reactions were run on the QuantStudio 7 Pro system with a 96-well block. The QuantStudio Pro system features cloud connectivity, so data were directly imported into the cloud-based RQ app when the PCR program was complete.

A



B

C

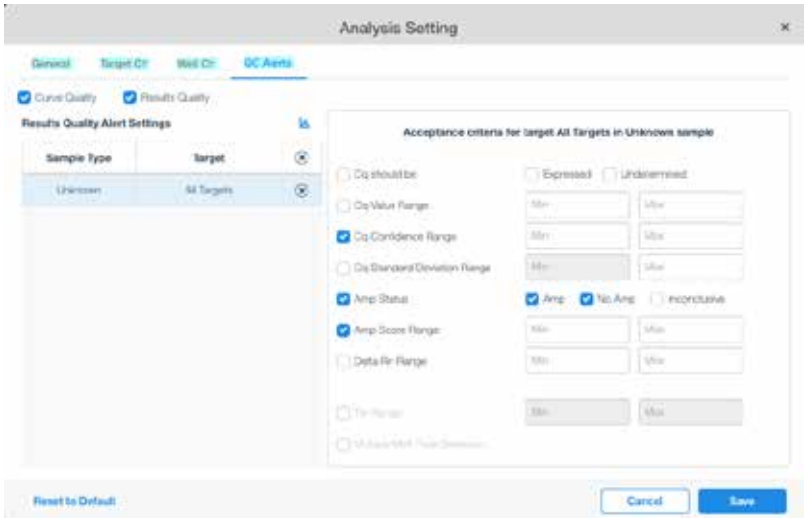


Figure 2. Setting up plate runs on the QuantStudio 6 Pro and 7 Pro systems using the Design and Analysis tool. (A) Assay layouts on predesigned plates can be directly imported into the instrument if the plate contains an RFID tag, or the information can be entered manually. Sample disposition (if more than one is used) and refinements of assay and reagent information can also be performed on this screen. **(B)** Run profiles can easily be modified. **(C)** Quality control choices can be made before initiating the run. The results using these parameters will appear in the final data output.

Analysis of results

Design and Analysis software simplifies the setup of plates and PCR conditions and can be accessed either on the instrument or remotely. For this experiment, the complete assay layout of the TaqMan Human Oncogenes-Tumor 96-well array plate was manually entered into the software (Figure 2A). The QuantStudio 6 Pro and 7 Pro systems can automatically import this information by reading the RFID tag on a plate, but the plates used for this demonstration did not have RFID tags. Sample and probe attributes can be entered or modified, the run profile set up and modified (Figure 2B), and primary analysis and QC settings established (Figure 2C). An .edt file can then be uploaded to the instrument if necessary, via the cloud or USB drive. Once the PCR run has been completed, the primary data can be viewed and checked based on QC settings set up in the file. The features displayed here are the same as on other QuantStudio instruments, so the primary analysis will be familiar to established users (Figure 3).

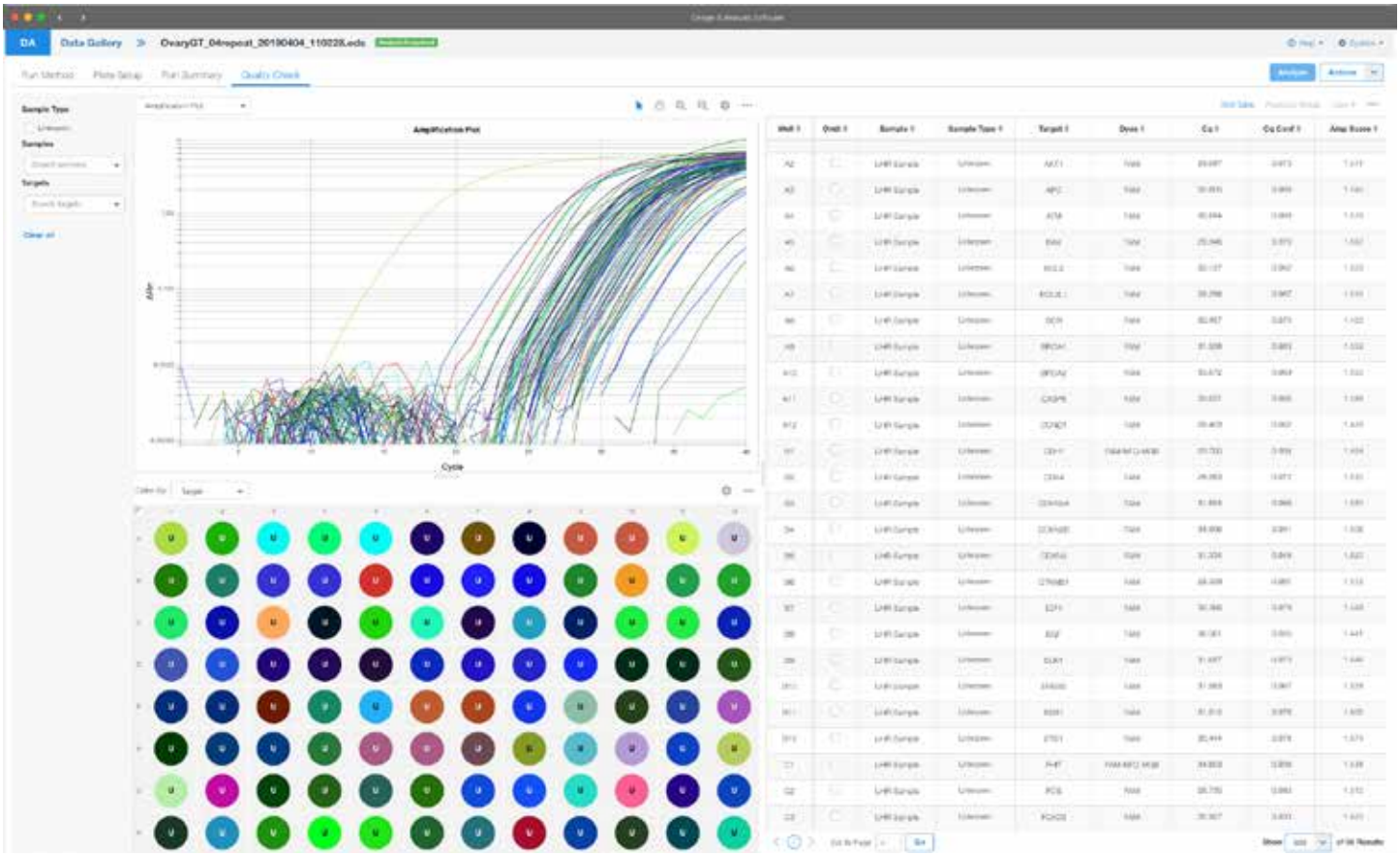


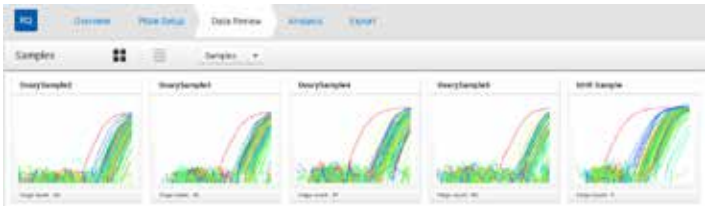
Figure 3. Primary analysis of a completed run. The amplification curves, assay maps, and results including QC flags are displayed after a run is complete. The curves can be adjusted as needed, similar to on other QuantStudio instruments.

The cloud-based RQ app was designed to take TaqMan gene expression data from any instrument and perform calculations that would highlight meaningful differences in transcript abundance. The cloud connectivity of the QuantStudio 6 Pro and 7 Pro systems allows direct import of data from the instrument to the app, remotely. And because the RQ app architecture is project-based, all files related to a single project can be analyzed at one time. For example, the amplification curves from all five ovarian tumor/UHR plates can be visualized at one time (Figure 4A). Alternatively, the amplification curves of each TaqMan assay can be visualized for all of the samples (Figure 4B).

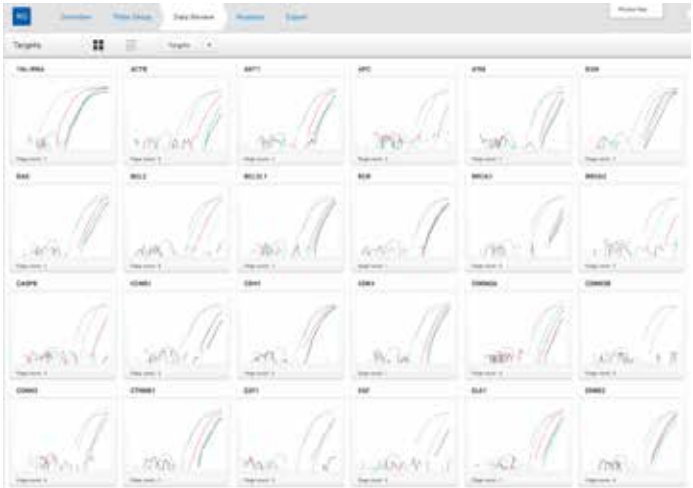
Since it was designed to analyze gene expression results, the RQ app has various options for normalizing the data. On the Overview screen (Figure 4C), one sample can be chosen as the reference sample; for this experiment, the UHR sample was chosen as the reference. Sample replicates, either technical or biological, are defined here. Additionally, similar types of samples can be grouped and ultimately analyzed together if defined in the Biogroups field. Further normalization options are accessed on the Data Analysis screen (Figure 4D). Here, the endogenous control, housekeeping, or reference gene(s) are defined. The software provides guidance as to which candidate control genes are most useful by providing scores (low scores reflect more stable expression across samples), allowing poorly performing genes to be ignored. Alternatively, the data sets can be normalized using global mean normalization, if appropriate.

Once the analysis parameters are defined, the RQ software will perform the appropriate calculations, including C_T and $\Delta\Delta C_T$, endogenous control normalization, mean and standard error if replicates are defined, quantification relative to the control sample, and others. The software also provides various tools for visualizing and exporting the processed data. For complete instructions and details, online help is available in the upper right corner of the screen. Data can also be exported for analysis in third-party software such as Microsoft™ Excel™ or SAS JMP™ software.

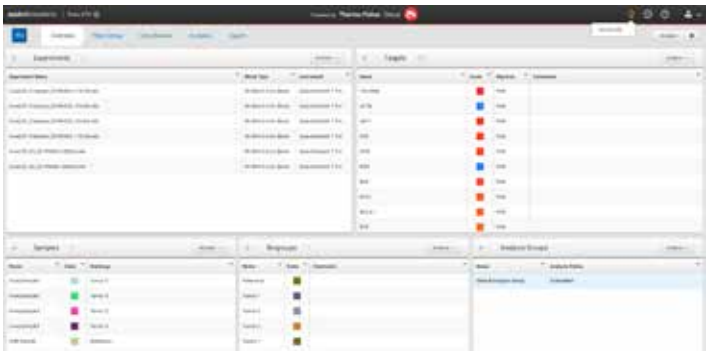
A



B



C



D

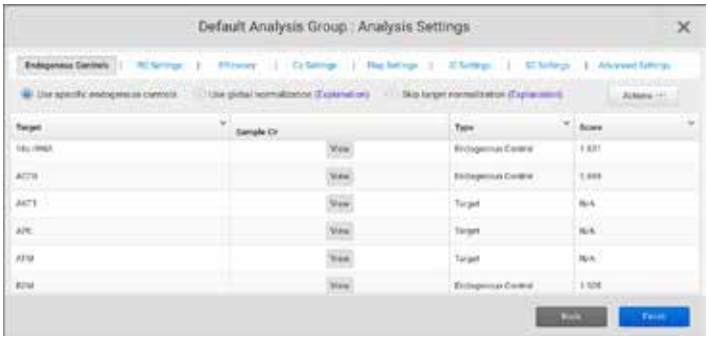


Figure 4. Visualization of primary data and setup of secondary analysis in the RQ app. Once primary data are input into the RQ app, they can analyzed several ways: **(A)** display of all assay amplification curves for each sample; **(B)** display of amplification curves for each individual assay in all samples; **(C)** definition and labeling of control samples and biological groups (part of the normalization process of analysis); **(D)** definition of gene normalization methods, including defining endogenous controls or global normalizations. Several other options for analysis can be defined on this screen.

For example, one way to perform a QC check of the entire data set is to compare box plots of the C_T of each of the samples (Figure 5A). The RQ app can take the population of C_T values for each assay in a sample and show average, median, and significant values, including outliers, in each sample. Hovering over the box plot will bring up the summary; hovering over an outlier will give the identity of that gene. For the example data set, it is clear that the overall average C_T value of the UHR sample is lower than that of the tumor samples, suggesting that there was more amplifiable cDNA in that sample than in the FFPE samples.

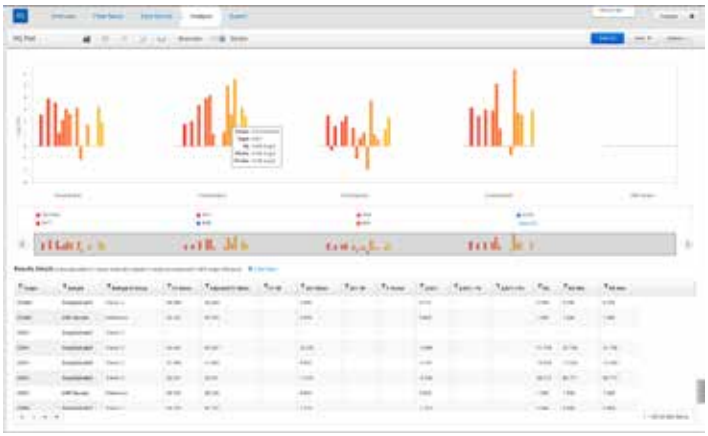
Another way to analyze the data is to visualize transcript abundance differences by comparing simple bar graphs. The RQ app can take the relative C_T for each gene and plot it relative to the sample (Figure 5B) or to other genes (not shown). In this example, only a subset of genes is shown; there are more screens showing the rest of the 96 genes in the panel. Note that the relative expression of the endogenous control genes and all the genes in the UHR sample is 1 (0 on a $\log_2$ scale), since these were used for normalization. A table with the numerical values is shown below the graphs; the order can be changed by clicking on the headers of each column in the table. This is an effective way to visualize the expression differences among single genes.

It is often helpful to visualize the correlation of expression across all the samples. The RQ app has an option for displaying a correlation plot of the C_T or $\Delta\Delta C_T$ values of all samples (Figure 5C). Hovering over the square will give the correlation coefficient. Here, a plot of the correlations of expression differences between samples shows that the ovarian tumor samples 3 and 5 are most divergent from the UHR (correlation coefficient $r = 0.2$); not surprisingly, ovarian tumors 3 and 5 are most similar to each other (correlation coefficient $r = 0.67$).

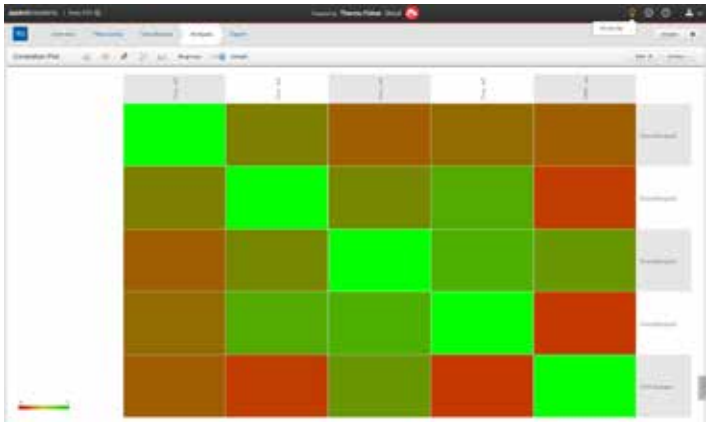
Groupings of similarly expressing genes and samples are often visualized using unsupervised clustering algorithms and heat maps. The RQ app has the ability to construct such heat maps based on the normalized data or C_T values using a variety of different, commonly used algorithms. For the example data, we used average linkage to cluster genes and plot distance using the Euclidean distance algorithm (Figure 5D). The results from this data set confirm that ovarian tumors 3 and 5 are similar to each other, and begin to point to groups of genes that are commonly differentially regulated in these samples relative to the others (for example, *TFRC* and *TP53* are downregulated and *NF1, CTNNB1,* and *NFKBIA* are upregulated relative to the other three samples). Other clustering options provided

A

B



C



D

Figure 5. The RQ app contains several modules for displaying results of normalized and analyzed data. (A) Box plot of C_T values for sample, showing maximum and minimum values, mean, median, quartiles, and outliers. **(B)** Portion of bar graph analysis. Here the relative expression values for a subset of genes in the plate are displayed. More results can be displayed by using the scroll arrows. **(C)** Correlation plot for the expression patterns of the different samples. **(D)** Unsupervised clustering and heat map of the expression levels of the different genes in the samples.

applied
biosystems

by the software include complete linkage and single linkage clustering, Pearson's coefficient in addition to Euclidean distance for plotting distance, and clustering using relative quantification or C_T values. Each of these methods has advantages, and the software allows the user to choose the method most appropriate for the biology.

Conclusions

Here we have demonstrated a workflow for analyzing expression of specific genes in FFPE tumor samples. Starting from slices of archived tissues on microscope slides, RNA can be extracted and analyzed using TaqMan gene expression array plates. The QuantStudio 6 Pro and 7 Pro systems simplify plate set-up and running the plate. Finally, the new cloud-based RQ app facilitates analysis of gene expression data. Together, these tools help investigators efficiently focus on discrete sets of transcriptional differences in cells.

Referenced products

Connect cloud platform—**thermofisher.com/connect**

QuantStudio 6 and 7 Pro instruments—**thermofisher.com/quantstudiopro**

SuperScript IV VILO kit—**thermofisher.com/superscript**

TaqMan array plates—**thermofisher.com/taqmanarrays**

TaqMan master mixes—**thermofisher.com/taqmanmm**

Find out more at **thermofisher.com/quantstudiopro**

ThermoFisher SCIENTIFIC