Distribution of mRNA transcripts in single cells determined by quantitative RT-PCR

CHALMERS

TATAA Biocenter

BENGTSSON M[1,2*], STÅHLBERG A[2], RORSMAN P[1,3], KUBISTA M[2]

LUND UNIVERSITY

[1]Department of Experimental Medical Science, Lund University, 221 84 Lund, Sweden

[2]TATAA Biocenter and the Department of Chemistry and Bioscience, Chalmers University of Technology, Lundberg Laboratory, 413 90 Göteborg, Sweden

[3]The Oxford Centre for Diabetes, Endocrinology and Metabolism, The Churchill Hospital, Oxford, OX3 7LJ, England

* Contact information (Martin Bengtsson):
Phone: +46 (0)46 222 0629
Fax: +46 (0) 46 222 7763
e-mail: martin.bengtsson@med.lu.se

Introduction

Cells in a population are in many aspects unique in their characteristics, even in a seemingly homogenous culture or tissue. Numerous findings suggest that often cells respond very differently even if exposed to identical stimuli. Traditionally, gene expression analysis look at the average transcript levels of the whole population of cells. This is adequate for studying many biological problems but it has major limitations. For example, it is not possible to determine if two transcripts are mutually exclusive or co-expressed at the cellular level. Wide-spread responses in the whole population can not be distinguished from rare changes in expression taking place in a few cells. Also, cell heterogeneity in tissues makes cell-type specific analysis technically demanding, requiring sorting of cells.

In the advent of technologies allowing detection and quantification of very small amounts of mRNA (sub-femtograms), single cell gene expression analysis emerged. In this work we study the expression of five genes in individual mouse pancreatic islet cells and murine MIN6 insulinoma cells by reverse transcriptase quantitative real-time PCR (RT-QPCR). This technique offers superior sensitivity, accuracy, and dynamic range compared to alternative methods for gene expression analysis. The pancreatic islets of Langerhans are responsible for maintaining the balance in blood glucose levels by accurately releasing hormones affecting glucose uptake. The heterogeneity of the islets is high. Each islet contains on average a total of 1000 cells comprising at least four-five endocrine cell types. This diversity makes it difficult to study cell-type specific effects on gene transcription, which may have strong influence on the physiology of the endocrine pancreas.

Gene expression profiling in single pancreatic islet cells reveals lognormal distribution of mRNA levels

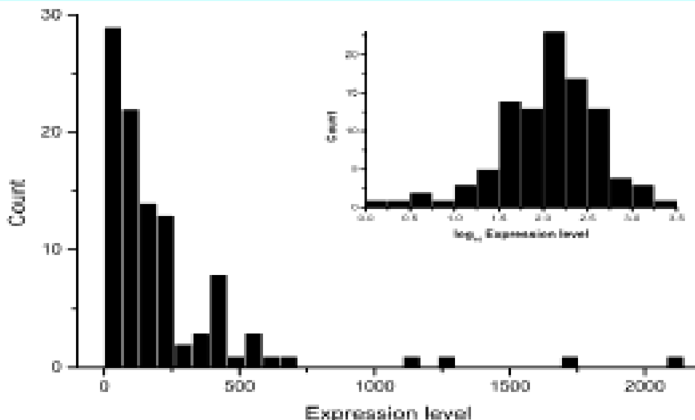

Fig: Histogram of 100 pancreatic β-cells expressing β-actin. The expression levels are skewed and the arithmetic mean would be heavily influenced by a few but highly expressing cells. Inset: Same data but with logarithmic scale, confirming lognormal distribution.

Material and methods

We have developed methodology to improve the fidelity and robustness of single cell gene expression analysis using reverse transcription real-time PCR. Special attention has been paid to adsorption phenomena, dilution effects, and inhibition of PCR by RT reaction component. The genes analyzed were β-actin (*ActB*), insulin I (*Ins1*), insulin II (*Ins2*), the K_{ATP}- channel subunits sulfonylurea receptor 1 (SUR1, a member of ATP-binding cassette family; *Abcc8*) and the inwardly rectifying channel Kir6.2 (subfamily J, member 11; *Kcnj11*). Out of a total of 121 cells, 60 were incubated in presence of 5 mM glucose (low) and 61 in 20 mM glucose (high).

MIN6-cells or primary cells of the Islets of Langerhans from mice were prepared according to standard protocols. Cells were collected using glass capillaries with a diameter of approximately 3-6 µm. Pipettes were emptied in lysis buffer (see figure at right), immediately followed by incubation on 80 ºC for 5 min and either freezing on dry ice for storage or brought directly to reverse transcription. Reverse transcription was optimized and carried out using both oligo(dT) and random hexamer priming. Real-time PCR was performed on Roche's LightCycler and Applied Biosystems' 7900 with SYBR Green I chemistry. Extremely high demands for specific PCR products were set on both primer design and PCR conditions.

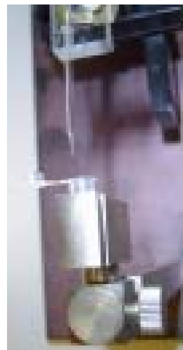

Fig: Glass pipette containing a single intact cell is emptied in 2-3 µl lysis buffer containing a weak, non-chaotropic detergent, and brought to 80 ºC for 5 min.

Models of gene induction

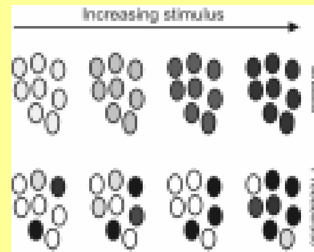

Fig: Four groups of cells exposed to an inducer respond by expressing higher levels of transcript (filled circles). In the graded model all cells respond equally and in proportion to the stimulus. The probabilistic, or binary, model suggests that the level of stimulus is proportional to the probability of a cell expressing a gene at full strength. This would result in a bimodal population distribution.

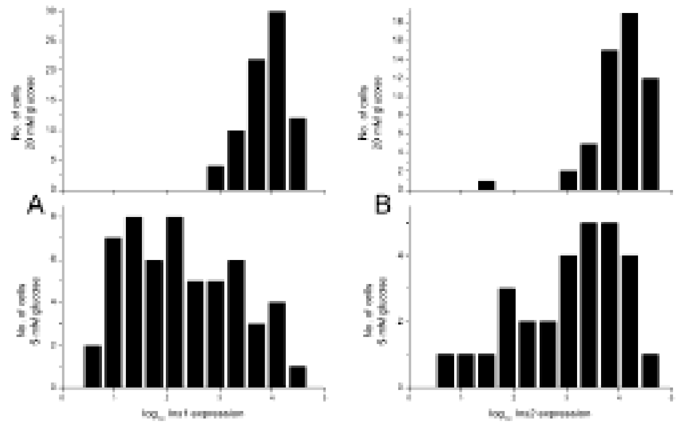

Fig: **A**: Histograms of *Ins1* expression levels in cells incubated in 5 mM (bottom) and 20 mM (top) glucose. Horizontal axis has logarithmic scale, which is identical in the two histograms. Scale on the vertical axis indicates the cell count in each histogram bin. **B**: Corresponding histograms for *Ins2* expression levels.

The geometric mean of expression levels provides a better estimate of gene activity of the average cell than the arithmetic mean measured on a cell population.

	ActB	Ins1	Ins2	Abcc8	Kcnj11
ActB	1				
Ins1	0.15	1			
Ins2	0.12	**0.90**	1		
Abcc8	-0.02	-0.01	0.06	1	
Kcnj11	0.11	-0.02	0.24	-0.15	1

Table: Pearson correlation coefficients based on logarithms of expression levels. r close to 1 signifies strong positive correlation, while a value close to zero means no correlation. Negative r value would be anti-correlation.

Ins1 and Ins2 are highly correlated

Arithmetic mean: $x = (\Sigma X_n)/N$

Geometric mean: $x^* = (\Pi X_n)^{1/N}$

	Arithmetic[1]	*Geometric[1]*
ActB	3.3	4.9
Ins1	**4.6**	**17**
Ins2	**3.0**	**9.5**
Abcc8	1.4	1.5
Kcnj11	1.1	1.3

Table: Ratios of mean expression levels in cell populations incubated in 20 mM and 5 mM glucose.
[1]Arithmetic and geometric mean value of the expression level at high glucose concentration divided with the corresponding mean value at low glucose concentration.

Conclusions

- Single cell mRNA harvest followed by quantitative RT-PCR is in our hands a robust method to quantify mRNA in single cells.
- Cell populations are heterogeneous at the mRNA level.
- The distribution of gene expression levels in a cell population are lognormally distributed, and thus the geometric mean is a better estimate of the average cell in a population than the arithmetic mean.
- Glucose stimulation enhanced the expression of the insulin genes.
- Of the genes analyzed, Ins1 and Ins2 were the only ones that correlated at the single cell level.

Supported by the Crafoord Foundation, the Royal and Hvitfeldtska Foundation, Swedish Research Council (8647), the Swedish Diabetes Association, the NovoNordisk Foundation, the Swedish Strategic Foundation, the Goran Gustafsson Foundation for Research in the Natural Sciences and Medicine and the Medical Faculty Lund University.