Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels

Martin Bengtsson,1,2,4 Anders Ståhlberg,2 Patrik Rorsman,1,3 and Mikael Kubista2

1Department of Experimental Medical Science, Lund University, 221 84 Lund, Sweden; 2Department of Chemistry and Biosciences, Molecular Biotechnology, Chalmers University of Technology and TATAA Biocenter, Lundberg Laboratory, 405 30 Göteborg, Sweden; 3The Oxford Centre for Diabetes, Endocrinology and Metabolism, The Churchill Hospital, Oxford, OX3 7LJ, United Kingdom

The transcriptional machinery in individual cells is controlled by a relatively small number of molecules, which may result in stochastic behavior in gene activity. Because of technical limitations in current collection and recording methods, most gene expression measurements are carried out on populations of cells and therefore reflect average mRNA levels. The variability of the transcript levels between different cells remains undefined, although it may have profound effects on cellular activities. Here we have measured gene expression levels of the five genes ActB, Ins1, Ins2, Abcc8, and Kcnj11 in individual cells from mouse pancreatic islets. Whereas Ins1 and Ins2 expression show a strong cell–cell correlation, this is not the case for the other genes. We further found that the transcript levels of the different genes are lognormally distributed. Hence, the geometric mean of expression levels provides a better estimate of gene activity of the typical cell than does the arithmetic mean measured on a cell population.

[Supplemental material is available online at www.genome.org.]
Table 1. Statistical parameters describing gene expression in single (insulin-expressing) β-cells at 5 and 20 mM glucose

<table>
<thead>
<tr>
<th>Glucose</th>
<th>Gene</th>
<th>N*</th>
<th>Arithmetic mean</th>
<th>Geometric mean</th>
<th>log_{10} Geometric mean (SD)</th>
<th>Shapiro-Wilk P-value</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 mM</td>
<td>ActB</td>
<td>51</td>
<td>130</td>
<td>61</td>
<td>1.79 (0.51)</td>
<td>0.56</td>
<td>−0.02</td>
</tr>
<tr>
<td></td>
<td>ins1</td>
<td>70</td>
<td>1700</td>
<td>190</td>
<td>2.29 (0.94)</td>
<td>0.32</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>ins2</td>
<td>44</td>
<td>5200</td>
<td>1100</td>
<td>3.03 (1.03)</td>
<td>0.005</td>
<td>−0.12</td>
</tr>
<tr>
<td></td>
<td>Abcc8</td>
<td>28</td>
<td>130</td>
<td>98</td>
<td>1.99 (0.34)</td>
<td>0.97</td>
<td>−0.05</td>
</tr>
<tr>
<td></td>
<td>Kcnj11</td>
<td>15</td>
<td>30</td>
<td>24</td>
<td>1.38 (0.22)</td>
<td>0.77</td>
<td>0.31</td>
</tr>
<tr>
<td>20 mM</td>
<td>ActB</td>
<td>45</td>
<td>420</td>
<td>300</td>
<td>2.47 (0.37)</td>
<td>0.53</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>ins1</td>
<td>49</td>
<td>7700</td>
<td>3200</td>
<td>3.51 (0.67)</td>
<td>0.09</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>ins2</td>
<td>47</td>
<td>16000</td>
<td>10000</td>
<td>4.01 (0.53)</td>
<td>0.05</td>
<td>−0.07</td>
</tr>
<tr>
<td></td>
<td>Abcc8</td>
<td>23</td>
<td>180</td>
<td>150</td>
<td>2.16 (0.33)</td>
<td>0.21</td>
<td>−0.01</td>
</tr>
<tr>
<td></td>
<td>Kcnj11</td>
<td>18</td>
<td>34</td>
<td>31</td>
<td>1.49 (0.23)</td>
<td>0.48</td>
<td>−0.19</td>
</tr>
</tbody>
</table>

*N is the number of cells expressing the tested gene. A total of 60 (84 for Ins1 and Ins2) cells were collected in 5 mM glucose, and 61 (85 for Ins1 and Ins2) cells in 20 mM glucose.

The geometric mean is derived from the formula $\mu_g = (\prod X_i)^{1/N}$.

Logarithm of the geometric mean and corresponding standard deviations.

For the Shapiro-Wilk test, a high P value reflects a good fit. If $P \geq 0.05$, then the null hypothesis that data are lognormal cannot be rejected.

Skewness is calculated as $\gamma = \Sigma(X_i - \mu)^2/(N - 1)\sigma^2$.

five genes indicate that whereas the number of insulin transcripts per cell is in the order of several thousand copies, ActB and Abcc8 transcripts are present in a few hundred copies. The number of transcripts of the K_{ATP}-channel subunit Kcnj11 is $< 300$ copies per cell. Based on the presence of Ins1 or Ins2 transcripts, it was concluded that at least 123 cells (73%) were β-cells. The fraction of β-cells in islet preparations is known to show large variations, and the average is 70%–80% (Barg et al. 2000).

To visualize the gene expression profile in a population of cells, distribution plots were used. Figure 1 shows histograms of ActB expression levels in both logarithmic and linear scale. As confirmed by the Shapiro-Wilk normality test, the transcript distribution is lognormal at 95% significance level ($P = 0.05$) (Table 1). The transcript distribution of the other genes is also lognormal at the same significance level. Corresponding data for the MIN6 cells are in Supplemental Table 1.

Lognormal distributions are common in nature. They are occasionally mistaken for normal (Gaussian) distributions, although the difference is fundamental (Limpert et al. 2001). In lognormal distribution, random, independent effects are multiplicative, while they are additive in a normal distribution. Bacteria in exponential growth are a classic example of a biological system where the numbers of organisms in the different colonies show lognormal distribution. Many more examples are known (Koch 1966; Limpert et al. 2001). The parameters describing the lognormal distributions of the expression levels of the studied genes at low and high glucose concentrations are given in Table 1. Whereas the arithmetic mean is representative for a normally distributed population, the geometric mean is a better indicator for a lognormal population. The differences between the geometric and arithmetic means of cellular gene expression levels for the Ins1 and Ins2 genes in the non-glucose-stimulated cells were substantial and amounted to nine- and fivefold, respectively. This is consistent with lognormal distribution of expression levels.

We ascertained that the observed lognormal distribution of expression levels reflects true biological variability and is not an artifact of the technology or the approach used (see Supplemental information; Methods; Supplemental Fig. 1). The finding that cellular transcript levels are lognormally distributed has implications on the interpretation of gene expression data in general. If mRNA expression levels among cells are lognormally rather than normally distributed, then the average expression measured on a cell population does not reflect the expression in the typical cell in the population. The average value is strongly biased by a small population of cells with very active transcription of the particular gene. Accordingly, it may not be valid to extrapolate results of gene expression measurements on cell populations to the single-cell level. We analyzed this aspect by measuring the distribution of the ratios between the expression levels at high and low glucose concentration for Ins1 and Ins2 (see Table 3). Glucose stimulation has been reported to increase Ins1 and Ins2 expression two-

Table 2. Pearson correlation coefficients based on logarithms of expression levels

<table>
<thead>
<tr>
<th></th>
<th>ActB</th>
<th>Ins1</th>
<th>Ins2</th>
<th>Abcc8</th>
<th>Kcnj11</th>
</tr>
</thead>
<tbody>
<tr>
<td>ActB</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ins1</td>
<td>0.15</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ins2</td>
<td>0.12</td>
<td>0.90</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abcc8</td>
<td>0.02</td>
<td>0.01</td>
<td>0.06</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Kcnj11</td>
<td>0.11</td>
<td>−0.02</td>
<td>0.24</td>
<td>0.04</td>
<td>1</td>
</tr>
</tbody>
</table>

Pearson coefficient is calculated as

$$-1 \leq r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2}\sqrt{\Sigma(y_i - \bar{y})^2}} \leq 1,$$

where $x$ and $y$ are the logarithms of expression levels of the two genes, $i$ indexes the cell, and the bar indicates average. $r$ close to 1 signifies strong positive correlation, while a value close to zero means no correlation. Negative $r$ value would be anti-correlation.
and fivefold (Nielsen et al. 1985), and this is in the range of the arithmetic means of our single cell expression data. By contrast, the ratios of the geometric means of the single cell expression levels are 17 and 9.5 for Ins1 and Ins2, respectively. Thus, glucose may have a much stronger effect on insulin gene transcription in some β-cells than is indicated by measurements on populations of cells.

The expression levels of Ins1 and Ins2 are affected similarly by glucose, suggesting that the two genes are regulated by similar or perhaps even a common mechanism (Wicksteed et al. 2001). By comparing at single-cell level, we find the expressions of the insulin genes highly correlated (Pearson coefficient of 0.90) (Table 2). In contrast, their expression does not correlate with that of the other genes studied (Table 2). Thus, there are not two populations with over-all high and low transcriptional activity, but some cells have low and others have high Ins1 and Ins2 expression independently of the expression of the other genes. This conclusion provides additional support of the notion that the observed variation of gene expression reflects true biological variation in cellular mRNA levels. Our analyses also provide the novel observation that the expression of ActB is also stimulated by glucose (Table 3), which casts doubt on the use of ActB as expression standard (Vandesompele et al. 2002). Interestingly, we find no correlation between ActB and Ins1/Ins2 transcription, suggesting that the molecular mechanisms behind the stimulations of ActB and Ins1/Ins2 are different.

Figure 2 shows the logarithm of Ins1 and Ins2 expression at low and high glucose levels in histograms. The large shift toward higher expression levels at elevated glucose concentration reflects 100-fold increase in insulin expression triggered by the sugar. Asymmetry, or skewness, of the distribution constitutes evidence of deviation from lognormal behavior. The skewness values for the genes analyzed, quantified as \( \sum (X - \mu)^2 \) / \( N - 1 \) for \( N = 1 \times 10^4 \), are presented in Table 1. A positive value suggests that the distribution is skewed toward higher values, and vice versa. Ins1 distributions exhibited high skewness of opposite signs at low and high glucose concentrations. This may suggest the existence of two subpopulations of cells: one active to secrete insulin and one dormant, exhibiting a bimodal distribution. The bimodal gene induction model (Ko 1992; Fiering et al. 2000) suggests that a subpopulation of the β-cells has enhanced transcription levels characterized by a high mean value, while another subpopulation has low transcriptional activity characterized by a much lower mean value. Elevated glucose level thus increases the probability that individual cells are activated. The currently prevailing model of the action of enhancers is consistent with the binary model for gene activation (Fiering et al. 2000; Paldi 2003).

An alternative interpretation of the data shown in Figure 2 is that high glucose “locks” the insulin gene in a high-expression state, whereas its expression at low glucose concentrations is more random. It should be noted that a fairly large portion of the cells at low glucose concentration exhibit high insulin gene expression as that observed for most cells during stimulation with 20 mM glucose.

When the expression levels of two genes are correlated, a common regulatory mechanism is often assumed. It may be through a mechanism that actually affects the two genes the same way, such as a common transcription factor, resulting in correlation at the single cell level, but it can also be a general increase in transcriptional activity due to, for example, environmental factors. The latter would also give rise to correlation in expression between genes, but not necessarily at the level of the individual cell. In our system, Ins1 and Ins2 are correlated on the cell level, while Ins1/Ins2 and ActB are correlated only on the population level. Our technology offers means to distinguish between these two cases and is expected to become especially useful for studies of molecular mechanisms underlying complex biological processes as well as disease.

### Methods

#### Preparation and culture of cells

Animals used in this study were healthy female National Maritime Research Institute (NMRI) mice aged 3–4 mo that were obtained from a commercial breeder (Bomholtgaard, Ry, Denmark) and fed a normal diet ad libitum. Care and use of animals were approved by the ethical committee of Lund University. The mice were sacrificed by cervical dislocation, and pancreatic islets were isolated by collagenase P digestion (Roche) (Olofsson et al. 2002).

---

**Table 3.** Ratios of mean expression levels in cell populations incubated in 20 mM and 5 mM glucose

<table>
<thead>
<tr>
<th>Gene</th>
<th>Arithmetic</th>
<th>Geometric</th>
</tr>
</thead>
<tbody>
<tr>
<td>ActB</td>
<td>3.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Ins1</td>
<td>4.6</td>
<td>17</td>
</tr>
<tr>
<td>Ins2</td>
<td>3.0</td>
<td>9.5</td>
</tr>
<tr>
<td>Abcc8</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>Kcnj11</td>
<td>1.1</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Arithmetic and geometric mean value of the expression level at high glucose concentration divided with the corresponding mean value at low glucose concentration. The mean values are calculated according to legends in Table 1.
Single islet cells were then prepared by gently shaking the collected islets at low extracellular [Ca$^{2+}$] as described (Eliasson et al. 1997). Dispersed cells were plated on plastic Petri dishes (Nunc) in RPMI 1640 (SVA) supplemented with 10% fetal calf serum, 100 U/mL penicillin, and 10 µg/mL streptomycin (all from Invitrogen) in the presence of either 5 or 20 mM glucose (Sigma-Aldrich). The cells were maintained in culture for 20–24 h. The data presented are from four batches of cells from different animals that were exposed to either 5 (two animals) or 20 (one animal) mM glucose. To account for individual variation between the animals, a fourth batch of cells were prepared from one animal and incubated in both 5 and 20 mM glucose. The individual variation was found to be negligible compared with the effect of glucose.

Mouse insulinoma MIN6-cells (passage 30 and above) were cultured in DMEM medium (10 mM glucose, Invitrogen) supplemented with 10% fetal calf serum, 100 U/mL penicillin, and 10 µg/mL streptomycin (all from Invitrogen) to −50% confluence by using standard culture techniques.

**Single cell isolation and cDNA synthesis**

Cells that adhered to the dish were thoroughly washed with extracellular solution containing 138 mM NaCl, 5.6 mM KCl, 1.2 mM MgCl$_2$, 2.6 mM CaCl$_2$, and 5 mM HEPES (pH 7.4 with NaOH). Cell content was collected with a borosilicate glass patch-clamp pipette with an average diameter of 5 mm mounted on a hydraulic micromanipulator. By controlling the pressure inside the pipette, it was possible to collect intact or nearly intact cells with minimum volume of extracellular solution. Pipettes were emptied in 0.2-mL tubes initially containing 2 µL of nonchao-tropic lysis solution. Tubes were then vortexed for 10 sec. The lysis buffer contained 0.5% Igepal CA-630, 50 mM Tris-Cl (pH 8.0), 140 mM NaCl, 1.5 mM MgCl$_2$ (all Sigma), and 1 U/µL Prime RNase inhibitor (Eppendorf). Samples were immediately frozen in ethanol cooled with dry ice (temperature: −78°C), and stored at −25°C for subsequent reverse transcription. A reverse transcription master mix contained dNTP (Sigma-Aldrich) and oligo(dT) primers (TAGC Copenhagen) at final concentrations of 0.5 mM and 10 µM, respectively. It was also supplemented with 150 ng of random hexamer primers (TAGC Copenhagen) and 15 U of Prime RNase inhibitor (Eppendorf); 150 ng Linear Polymeramide (GenElute LPA, Sigma-Aldrich) was added to some samples, but this did not have an effect on the cDNA synthesis efficiency. It was also found that the cell could be emptied directly into the RT-base, without the lysis buffer, with maintained reaction efficiency. After incubation for 5 min at 80°C and subsequent cooling on ice, the cDNA synthesis was initiated by addition of reverse transcription enzyme (SuperScript III, Invitrogen) and 2-h incubation at 50°C. The reaction was terminated by heating for 15 min at 70°C. Final volume for all reverse transcription reactions was 10 µL. The cDNA was either immediately quantified by real-time PCR or stored at −20°C pending later analysis. A total of 169 cells were analyzed, of which 84 were incubated in 5 mM glucose and 85 cells in 20 mM glucose. Twenty-four of the cells in each group were analyzed with primers for Ins2 only; the remaining cells, with primers for ActB, Ins1, Ins2, Abcc8, and Kcnj11.

**Quantitative real-time PCR**

Two instruments were used for real-time PCR measurements: the LightCycler (Roche) and the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems). Ten-microliter reactions were used on the LightCycler containing 3 mM MgCl$_2$, 0.3 mM dNTP, 0.1 mg/mL BSA (all Sigma-Aldrich), 400 µM primer (MWG and TAGC Copenhagen), 0.5× SYBR Green 1 (Molecular Probes/Invitrogen), and 1 U JumpStart Taq polymerase (Sigma-Aldrich). Primer sequences are available upon request. A similar master mix was used with the ABI PRISM 7900HT system, with ROX as passive reference dye but without BSA and a final volume of 20 µL. Single-cell cDNA (1 µL) was added to the reaction mixture when using the LightCycler, while 2 µL cDNA was analyzed in the 7900. Absolute quantification of each cDNA species was performed by dilution series of purified PCR products (QiAg/Quick PCR purification reagent set, Qiagen), and concentrations were measured with a NanoDrop ND-1000 spectrophotometer (Nano-drop Technologies). Real-time PCR data analysis was performed according to established procedures (Bustin 2000; Stahlberg et al. 2005).

Some cells appeared to lack at least one transcript; 96% of all collected samples contained detectable transcripts from at least one gene, and 83% had detectable levels of ActB transcript. For the analysis of primary cells in this article, only the insulin-producing β-cells were used. The detection limit of the method is 10–20 copies, which is below the range of expected levels of ActB, Ins1, Ins2, and Abcc8. It is possible that some of the cells that lack insulin gene transcripts are members of an inert subpopulation of β-cells, but the extent would be limited to a maximum 10% of the cells.

**Acknowledgments**

We thank Kristina Borglid, Britt-Marie Nilsson, and Lena Thiman for technical assistance with cell preparations. Research was 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. We also thank Prof. Miyazaki for providing us with the MIN6 cell line. P.R. is a Royal Society-Wolfson Research Fellow.

**References**


Received February 11, 2005; accepted in revised form June 28, 2005.