

RESEARCH

Open Access



Single-cell gene expression analysis reveals β -cell dysfunction and deficit mechanisms in type 2 diabetes

Lichun Ma¹ and Jie Zheng^{2*}

From 29th International Conference on Genome Informatics
Yunnan, China. 3-5 December 2018

Abstract

Background: Type 2 diabetes (T2D) is one of the most common chronic diseases. Studies on T2D are mainly built upon bulk-cell data analysis, which measures the average gene expression levels for a population of cells and cannot capture the inter-cell heterogeneity. The single-cell RNA-sequencing technology can provide additional information about the molecular mechanisms of T2D at single-cell level.

Results: In this work, we analyze three datasets of single-cell transcriptomes to reveal β -cell dysfunction and deficit mechanisms in T2D. Focused on the expression levels of key genes, we conduct discrimination of healthy and T2D β -cells using five machine learning classifiers, and extracted major influential factors by calculating correlation coefficients and mutual information. Our analysis shows that T2D β -cells are normal in insulin gene expression in the scenario of low cellular stress (especially oxidative stress), but appear dysfunctional under the circumstances of high cellular stress. Remarkably, oxidative stress plays an important role in affecting the expression of insulin gene. In addition, by analyzing the genes related to apoptosis, we found that the TNFR1-, BAX-, CAPN1- and CAPN2-dependent pathways may be crucial for β -cell apoptosis in T2D. Finally, personalized analysis indicates cell heterogeneity and individual-specific insulin gene expression.

Conclusions: Oxidative stress is an important influential factor on insulin gene expression in T2D. Based on the uncovered mechanism of β -cell dysfunction and deficit, targeting key genes in the apoptosis pathway along with alleviating oxidative stress could be a potential treatment strategy for T2D.

Keywords: Single-cell, Hyperglycaemia, Type 2 diabetes, β -cell dysfunction, β -cell deficit, Insulin expression, Apoptosis, Oxidative stress

Background

Type 2 diabetes (T2D) is one of the major causes of death worldwide [1]. It is characterized by insulin resistance and impaired insulin secretion [2, 3]. Insulin resistance denotes declined insulin sensitivity in insulin targeted cells or tissues, while insufficient insulin secretion is related to pancreatic β -cell dysfunction and the loss of β -cell mass [4, 5]. β -cells are located in the islets of

Langerhans, i.e. endocrine regions of pancreas. The main function of β -cells is to synthesize, store and secrete insulin, which is a peptide hormone and takes effects in decreasing the blood glucose level. It is reported that β -cell function declines even before the diagnosis of T2D [6]. In addition, β -cell deficit of about 20%~65% was demonstrated for T2D in several studies [7–9]. Kahn [10] investigated the contribution of insulin resistance and β -cell dysfunction to the pathophysiology of T2D. Yoon et al. [11] measured β -cell mass in T2D. Although these works intended to study the β -cell dysfunction and deficit

*Correspondence: zhengjie@shanghaitech.edu.cn

²School of Information Science and Technology, ShanghaiTech University, 201210 Shanghai, China

Full list of author information is available at the end of the article



mechanisms in T2D, they were mainly built upon bulk-cell analysis which can only provide average information about a population of cells.

Since the transcriptomes were firstly measured at single-cell level by Tang et al. in 2009, the technique of single-cell RNA sequencing (scRNA-seq) has experienced an explosive development in the past 10 years [12–15]. Compared with bulk-based approaches, scRNA-seq can provide crucial insights into cellular heterogeneity and bring profound new discoveries in biology [16–21]. For example, Deng et al. reported stochastic expression of monoallelic genes in mammalian cells [22]; Buettner et al. detected hidden subpopulations of cells by analyzing scRNA-seq data [23]. The technique of scRNA-seq has also been applied to transcriptome profiling of human pancreatic cells for both healthy and T2D donors [24–27]. Xin et al. [28] and Segerstolpe et al. [29] showed the expression heterogeneity of human islet cells (e.g. α -cells, β -cells and δ -cells). They also analyzed the alterations of gene expression patterns as well as the enriched signaling pathways in T2D compared with healthy people.

In this study, we aim to unravel the β -cell dysfunction and deficit mechanisms in T2D by analyzing the single-cell transcriptomic data of β -cells. Three single-cell transcriptomic datasets were adopted because each of the datasets contains more than 100 of β -cells (we have found but not used a few other available datasets because they only contain limited numbers of β -cells). We named the three datasets as dataset 1, dataset 2 and dataset 3, respectively. All the three datasets consist of β -cells obtained from T2D donors and healthy donors. The analysis was carried out from three aspects, i.e. β -cell dysfunction, β -cell deficit and personalized analysis. Firstly, we focused on the mechanisms of β -cell dysfunction in T2D. It is well known that the major role of pancreatic β -cells is to produce insulin. Thus, we analyzed the expression levels of INS (i.e. the gene that encodes the proinsulin precursor of active insulin) in β -cells belonging to healthy and T2D donors of each dataset. Different patterns of INS expression were detected in the three datasets. To explore the reasons, we examined the cellular stress in β -cells of the three datasets, and applied different machine learning algorithms to discriminate healthy and T2D β -cells by using the stress related features. Modeling the vulnerability of T2D β -cells to cellular stress, we found that oxidative stress could be a major influential factor on INS expression. Secondly, to study the mechanisms of β -cell deficit in T2D, we investigated the expression levels of the genes in the apoptosis pathway, conducted principle component analysis and carried out mutual information calculation. As a result, genes and pathways that are crucial for β -cell apoptosis in T2D were detected. In the last part, we performed personalized analysis of INS expression and the expression of death executioner caspases.

Based on the analysis of the three datasets of β -cell transcriptomes, we obtained the following main results. Some β -cells in T2D donors have comparable INS expression levels with those in healthy donors; β -cells in T2D have normal INS expression under low cellular stress, but they have dysfunction under high cellular stress; β -cells in healthy people can deal with the cellular stress, maintaining normal INS expression; oxidative stress could be a major influential factor on INS expression; TNFR1-, BAX-, CAPN1- and CAPN2-dependent pathways may be crucial for β -cell apoptosis in T2D; INS and death executioner caspases are differentially expressed among donors. Note that some of the above results could hardly be obtained from bulk-cell analysis.

Results

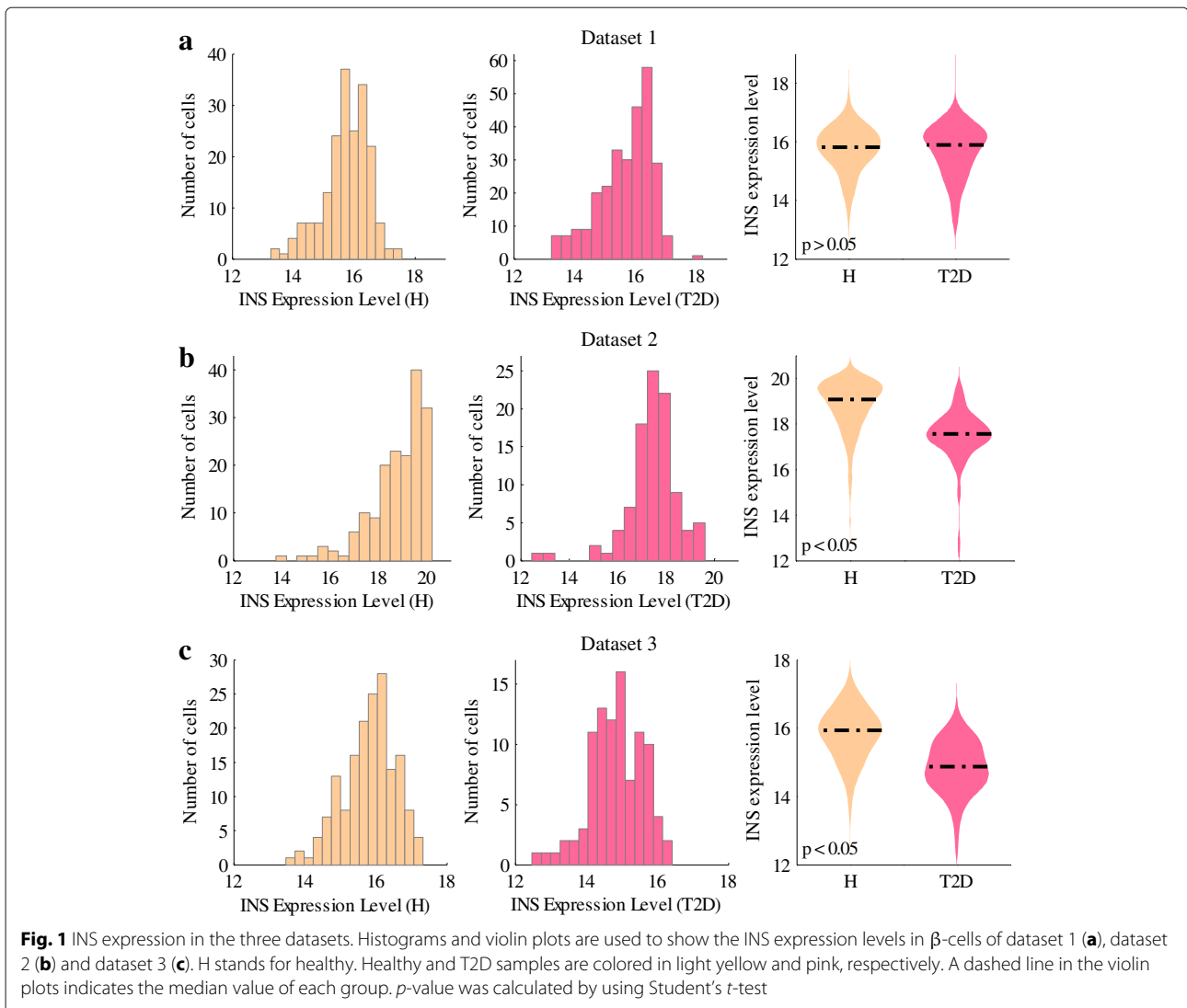
Pancreatic β -cell dysfunction in T2D

Single-cell INS expression

In this work, we employed three datasets of β -cell transcriptomes, which were obtained from the published works of Xin et al. [28], Segerstolpe et al. [29] and Lawlor et al. [27]. All the three datasets consist of β -cells obtained from T2D donors and healthy donors. They were named as dataset 1, dataset 2 and dataset 3, respectively. The numbers of donors and β -cells of each dataset are shown in Table 1. Overall, the three datasets comprise gene expression levels of 1,006 β -cells from 36 donors. Among the β -cells, 472, 270 and 264 belong to the first, second, and third datasets, respectively. Figure 1 presents the INS expression levels of healthy and T2D β -cells of the three datasets. As shown in the figure, the single-cell data can reveal cellular heterogeneity in the INS expression, which cannot be obtained from bulk-cell analysis. For both datasets 2 and 3, the median value of INS expression in T2D β -cells is lower than that of the healthy β -cells, although some T2D β -cells have comparable INS expression levels with the healthy ones. In dataset 1, however, the median values of INS expression of healthy and T2D β -cells are almost equal. It is well known that one characteristic of T2D is a relative deficiency of insulin. However, this characteristic is not shown in dataset 1. To uncover the regulatory factors of the different INS expression patterns, we investigated the cellular stress in β -cells of each

Table 1 The numbers of donors and β -cells of each dataset

Gene expression dataset of β -cells		Dataset 1	Dataset 2	Dataset 3
Number of donors	T2D	6	4	3
	Healthy	12	6	5
Number of β -cells	T2D	278	99	96
	Healthy	194	171	168



dataset, as the increase of cellular stress was detected in T2D by many studies.

Cellular stress

Plenty of evidence indicates that prolonged exposure of β -cells to hyperglycemia and high free fatty acids (FFA) causes deleterious effects of endoplasmic reticulum (ER) stress, oxidative stress, and increase of β -cell apoptosis [6, 30–34]. ER stress is developed as the continuous demand of insulin, leading to the increased burden of β -cell and the accumulation of misfolded proteins in the ER lumen. ER stress is mediated by IRE1, EIF2AK3 and ATF6 [35, 36]. Reactive oxygen species (ROS) are accumulated to cytotoxic level during chronic glucose and fatty acids metabolism [5, 37–39]. Besides, hyperglycemia may disrupt the electron transport chain in mitochondria, which is also a main source of free radicals [40, 41]. The oxidative stress (cumulative ROS) promotes the activation of

ASK1, JNK and P38ALPHA. Hyperglycemia and high FFA induce increased β -cell apoptosis by several mechanisms, including promoting proapoptotic gene expression, and increasing ER stress, oxidative stress as well as inflammation stress. We use the expression levels of CASP3, CASP6, and CASP7 (i.e. death executioner caspases) to represent the rate of apoptosis.

Figure 2 shows the expression of aforementioned stress-related genes, including ER stress, oxidative stress and apoptosis-related genes. As shown in the figure, cellular stress is at a low level for both healthy and T2D β -cells in dataset 1 (Fig. 2a), while high stress exists in both groups of cells in dataset 3 (Fig. 2c). In dataset 2, healthy β -cells have weak stress whereas T2D β -cells experience strong stress (Fig. 2b). By comparing the cellular stresses and the INS expressions of the three datasets (Figs. 1 and 2), the following hypotheses may be proposed: T2D β -cells perform similarly to the healthy ones in INS expression under

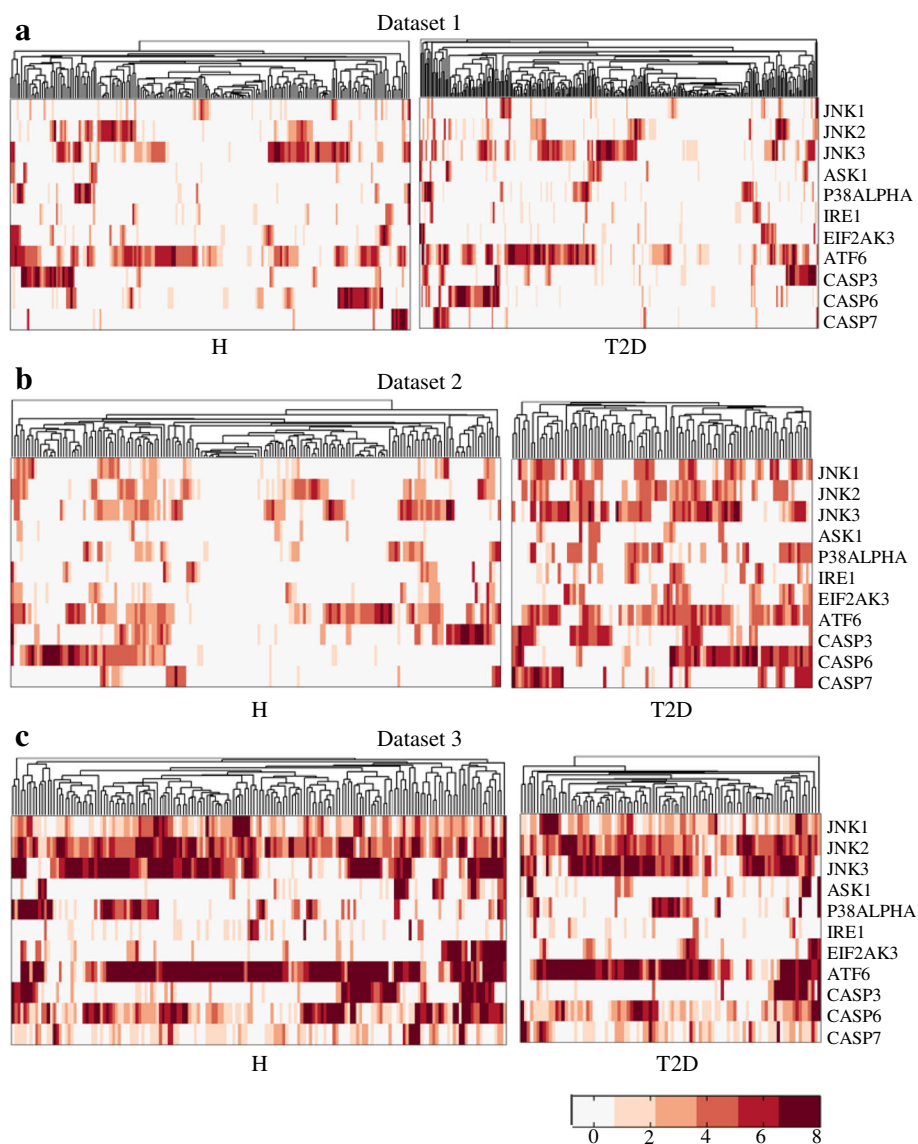


Fig. 2 Cellular stress related genes. **a**, **b** and **c** are heatmaps plotting the expression levels of the stress related genes of datasets 1, 2 and 3, respectively. H and T2D represent healthy and T2D β -cells. Each row in a heatmap corresponds to a gene and each column represents a cell. Colors in the heatmap denote the log₂ expression values

low cellular stress (dataset 1); T2D β -cells tend to be dysfunctional under the circumstances of high cellular stress (dataset 2); healthy β -cells can partly deal with high cellular stress, maintaining INS expression at normal level (dataset 3, Fig. 1c). To further test these hypotheses, we employed different classifiers to discriminate healthy and T2D β -cells.

Discrimination of healthy and T2D β -cells

In order to test the above analysis results, we employed different classifiers to discriminate healthy and T2D β -cells for each dataset, using expression data of genes that

are related to ER stress, oxidative stress and apoptosis. In addition, the INS expression related genes were also included as features of the two groups of cells [42, 43]. Overall, 45 genes were selected (Additional file 1: Table S1) [35–41]. Then, we chose genes that were expressed in more than 35% of all the cells in each dataset (40% for dataset 1, these values were derived according to the proportion of healthy β -cells in all the cells of each dataset). We conducted this step because a feature (gene) cannot contribute to distinguishing healthy and T2D β -cells if it is barely expressed in the cells. Here, expressed genes are those with expression levels no less than 1. Then 17, 25

and 31 genes met the conditions in datasets 1, 2 and 3, respectively. For fair comparison, we also applied the 25 genes selected in dataset 2 to datasets 1 and 3.

Afterwards, five classifiers, i.e. Bayesian network, support vector machine (SVM), random forest, logistic regression and neural network (NN), were used to predict the cellular conditions (i.e. healthy or T2D) of β -cells in each dataset. The performance is shown in Fig. 3. Besides prediction accuracy, we also adopted F-measure to evaluate the prediction performance, as the numbers of healthy and T2D cells are not balanced in each dataset. As shown in Fig. 3, all the five classifiers could hardly distinguish healthy and T2D β -cells in dataset 1. It implies that healthy and T2D β -cells behave similarly in the scenario of low cellular stress. For dataset 2, the prediction accuracy and F-measure are both high, indicating that β -cells in T2D have dysfunction under the circumstances of high cellular stress. For dataset 3, the prediction performance is lower than that of dataset 2, probably due to the heavy cellular stress in both of the groups of β -cells. However, healthy β -cells maintain normal INS expression under high stress (Fig. 1c), suggesting that the cells can partly deal with the cellular stress.

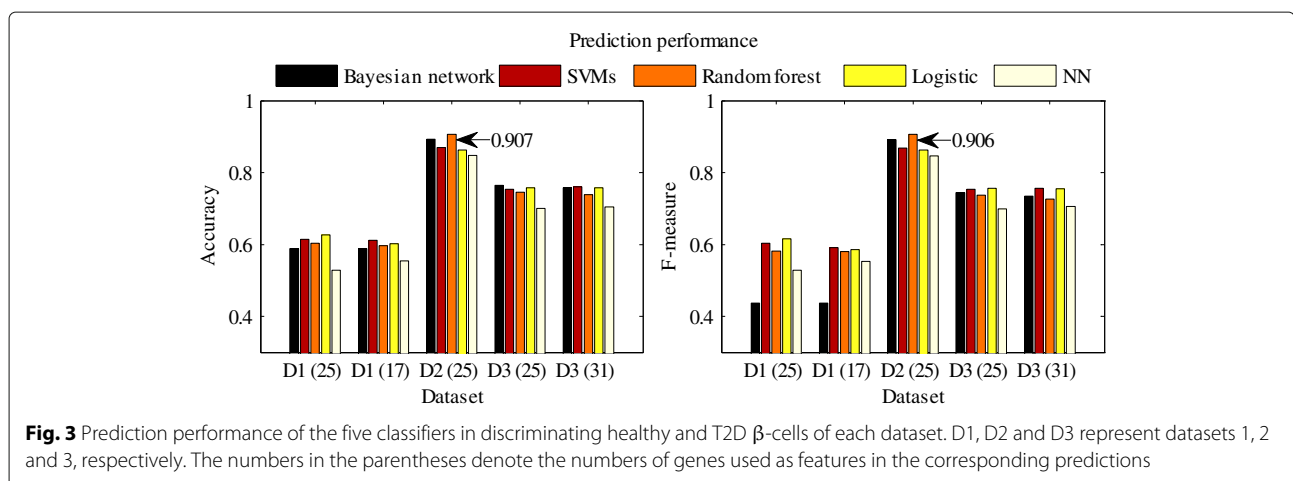
Vulnerability of T2D β -cells

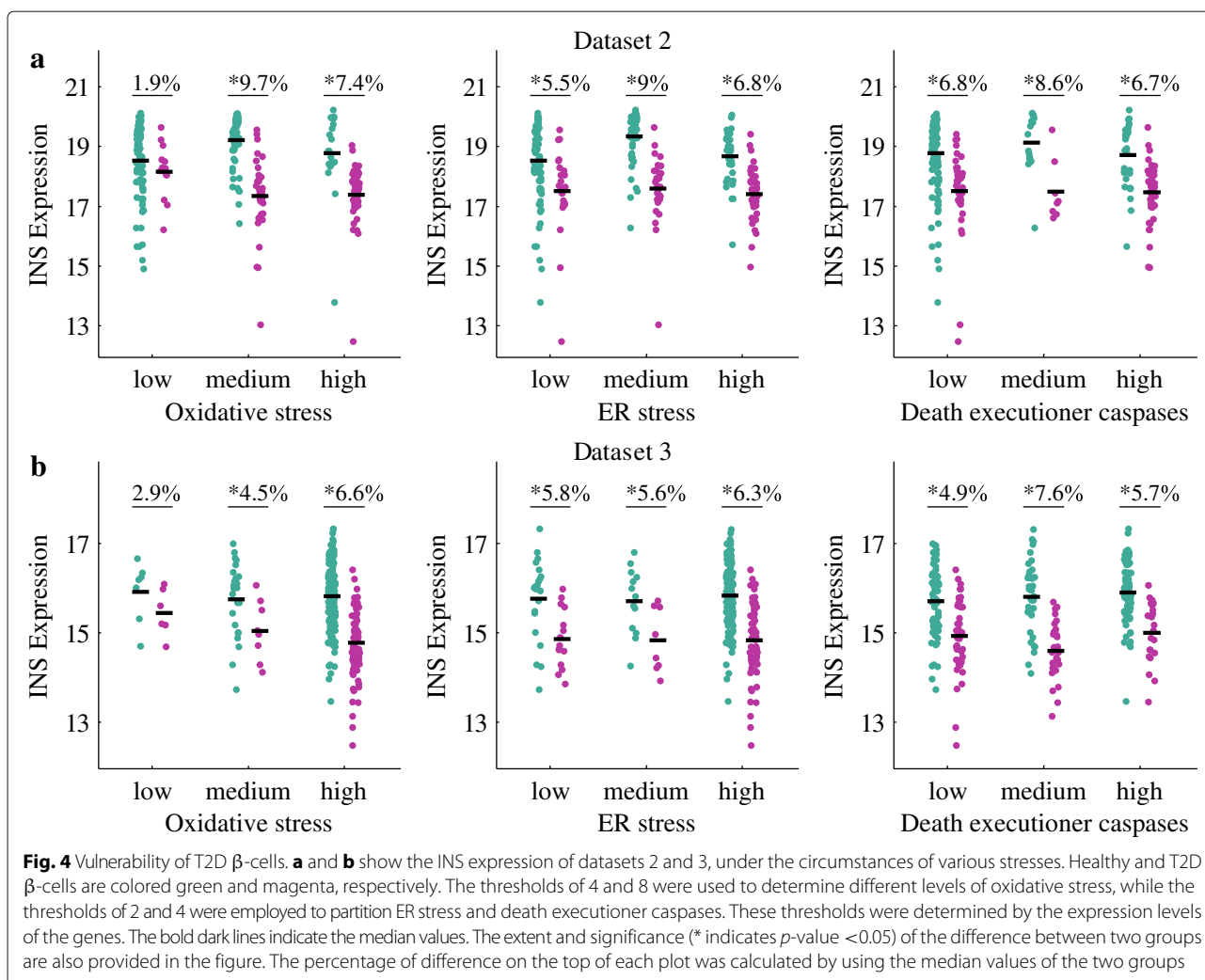
In dataset 3, both the healthy and T2D groups of β -cells suffer from high cellular stress. However, INS is more highly expressed in healthy β -cells than in the T2D ones (Fig. 1c). This is likely due to the fact that pancreatic β -cells in T2D are vulnerable to dysfunction as the toxic effects of hyperglycaemia and high FFA (i.e. glucolipotoxicity). In other words, healthy β -cells can deal with high cellular stress but T2D β -cells cannot, and thereby INS expression is lower in T2D β -cells than that in healthy cells under similar cellular stress (Fig. 4). Dataset 1 is not analyzed here, due to the low cellular stress and comparable INS expression between the healthy and T2D groups

of β -cells. In Fig. 4, oxidative stress, ER stress and death executioner caspases are represented by the expression levels of JNK, ATF6 and CASP6, respectively. We used the three genes as they expressed in more than 35% (the method of deriving the value has been mentioned previously) of all the β -cells in each dataset. As shown in Fig. 4a, INS expression in T2D β -cells is slightly lower than that in healthy ones when oxidative stress is weak. However, under medium or high oxidative stress, there is a significant difference in INS expression between the two groups. For ER stress and death executioner caspases, INS expression is significantly different between the healthy and T2D cells even when cellular stress is at a low level. Figure 4b shows similar patterns as Fig. 4a, indicating the vulnerability of T2D β -cells (as reflected by the reduced INS expression) compared with the healthy β -cells under similar cellular stress.

Major influential factors for INS expression

To further determine the major influential factors for INS expression in T2D, we computed the correlation and mutual information between INS expression and the three aforementioned factors (ER stress, oxidative stress and death executioner caspases) for T2D β -cells of datasets 2 and 3 (Fig. 5a). As shown in Fig. 5a, the correlation between oxidative stress and INS expression is the strongest, and the mutual information between them is the largest among the three factors. Table 2 lists the detailed values of entropy, joint entropy and mutual information, where INS, ER, OXID and CASP denote INS expression, ER stress, oxidative stress and death executioner caspases, respectively. Entropy and joint entropy measure the average information content of one or a set of variable(s), while mutual information measures the information of one variable obtained through another. Figure 5b shows INS expression in T2D β -cells under low, medium and high levels of oxidative stress. For both





datasets 2 and 3, there is a significant difference of INS expression between the β -cells under low and high levels of oxidative stress. These results reveal that oxidative stress could be a major factor that affects the INS expression.

β -cell apoptosis in T2D

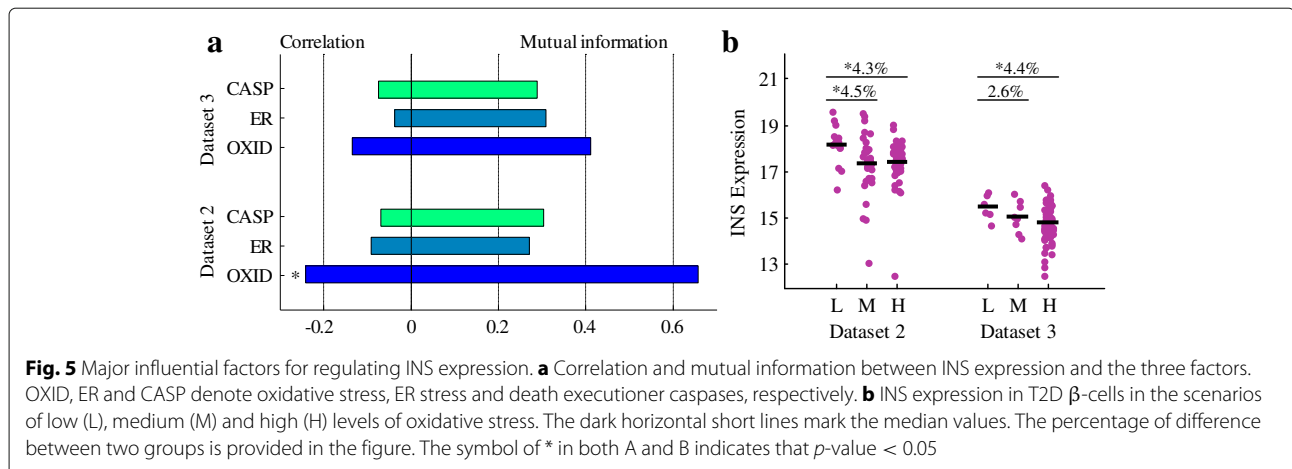
Expression of death executioner caspases

In apoptosis, the caspase proteins of CASP3, CASP6 and CASP7 act as death executioner enzymes. Thus, we used the total expression levels of these genes encoding the three caspases to measure the rate of β -cell apoptosis. Figure 6 shows the distribution of the total expression of the death executioner caspases (TEDECs) in each dataset. Increased apoptosis in T2D is only observed in dataset 2. In dataset 1, the median values of TEDECs of both groups of cells (i.e. healthy and T2D) approach 0, while the values of the two groups are both high in dataset 3. These patterns are consistent with the cellular stress (i.e., ER stress and oxidative stress) of datasets 1 and 3 (Fig. 2), where low

cellular stress is observed in dataset 1 and high stress in dataset 3. Surprisingly, the median value of TEDECs of the healthy β -cells is higher than that of the T2D β -cells in dataset 3, despite the increased β -cell apoptosis and β -cell deficit in T2D as reported in the literature. This pattern needs further analysis and explanation in the future.

Pathway analysis of apoptosis in T2D

To further understand the regulation of abnormal β -cell death in T2D, we conducted pathway analysis of apoptosis. Among the three datasets analysed here, the increased level of β -cell apoptosis has been only observed in dataset 2 (Fig. 6), thus we focused on dataset 2 to detect the major apoptosis pathway in T2D β -cells. According to the apoptosis pathways in KEGG Pathway Database [44] and GeneGO MetaCore [45], a total of 24 apoptosis-related genes were involved in our study. They constitute the death receptor-, mitochondria-, TP53- as well as the ER stressed-mediated apoptosis pathways. The expression levels of these genes are presented in Fig. 7. The



genes encoding RIPK1, RAIDD, TNFR1, TRADD, BAX, BCL2L1, CAPN1 and CAPN2 are more highly expressed in T2D β -cells than the healthy ones.

We then computed the mutual information between the TEDECs and the proteins in the apoptosis pathway (Fig. 8a, Additional file 2: Table S2). In T2D, much information about the TEDECs can be obtained through the genes encoding TNFR1, DAXX, BAX, BCL2L1, DIABLO, CAPN1 and CAPN2. Besides, we conducted principal component analysis (PCA), and projected the cells (characterized by the 24 apoptosis-related genes) to the first two principal components (Fig. 8b). Although the healthy and T2D groups of cells are not completely separated, they have difference in the new projection. By checking the first two principal components, we found that the values associated with the genes encoding RAIDD, TNFR1, BAX, BCL2L1, DIABLO, CAPN1 and CAPN2 are larger than others (Additional file 3: Figure S1). It implies that these genes contribute significantly to the new projection.

To summarize the above analyses, the death receptor TNFR1-mediated pathway, mitochondrial BAX-related

pathway, as well as the CAPN1- and CAPN2-dependent pathway may be crucial in T2D.

We also analyzed the genes related to apoptosis in dataset 3. Different from dataset 2, in dataset 3 TP53 is highly expressed. The results are provided in the supplementary documents (Additional file 2: Table S3, Additional file 3: Figures S2 and S3).

Personalized analysis of β -cells

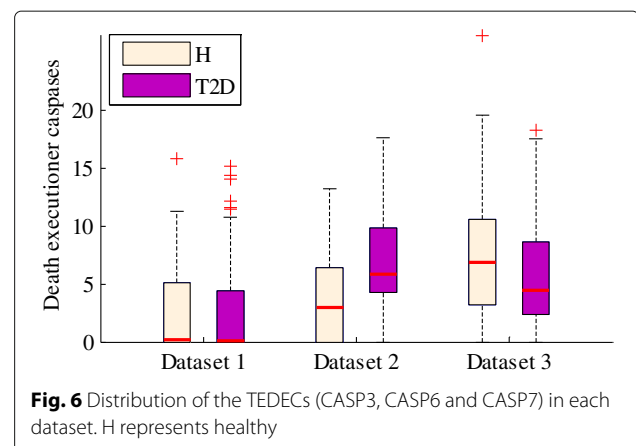
In addition to comparing the gene expression in healthy and T2D β -cells, we also conducted personalized analysis. Figure 9 shows the INS expression of β -cells from each donor. Healthy and T2D donors are arranged sequentially in rows. Each vertical bar represents INS expression level of a β -cell. The symbol of * indicates the position of the median INS expression level of each donor. As shown in the figure, INS expression is different among donors. In datasets 2 and 3, most of the T2D patients have lower median INS expression levels than the healthy donors.

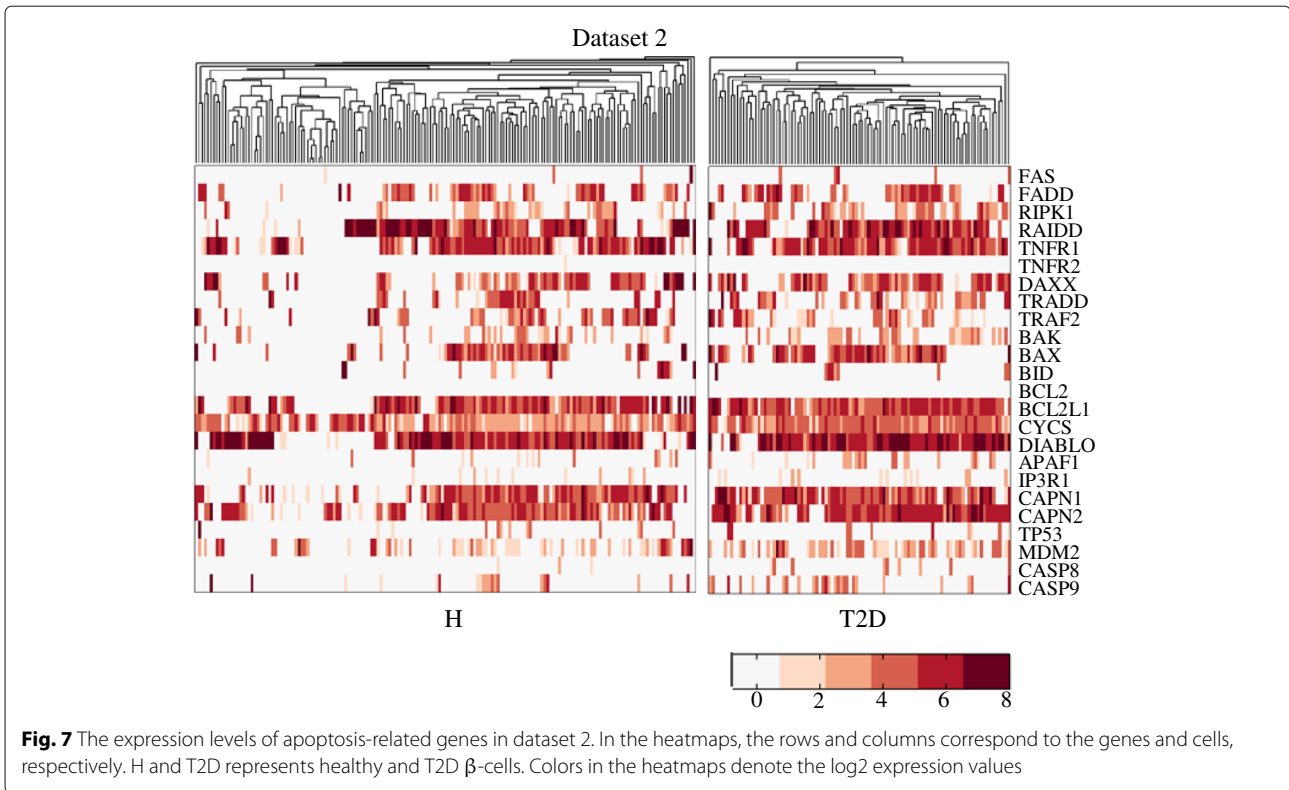
We also analyzed the TEDECs of each donor (Additional file 3: Figure S4). Similar to the INS expression, the TEDECs are also different among donors. Note that,

Table 2 Entropy, joint entropy and mutual information

Dataset		2	3
Entropy	INS	1.95	1.65
	OXID	3.68	3.93
	ER	2.44	3.23
	CASP	2.41	2.85
Joint entropy	INS and OXID	4.96	5.17
	INS and ER	4.12	4.57
	INS and CASP	4.06	4.21
Mutual information	INS and OXID	0.66	0.41
	INS and ER	0.27	0.31
	INS and CASP	0.30	0.29

The unit is bit. The bold numbers are the highest values among those compared in the same group (i.e. the same measure and the same dataset)



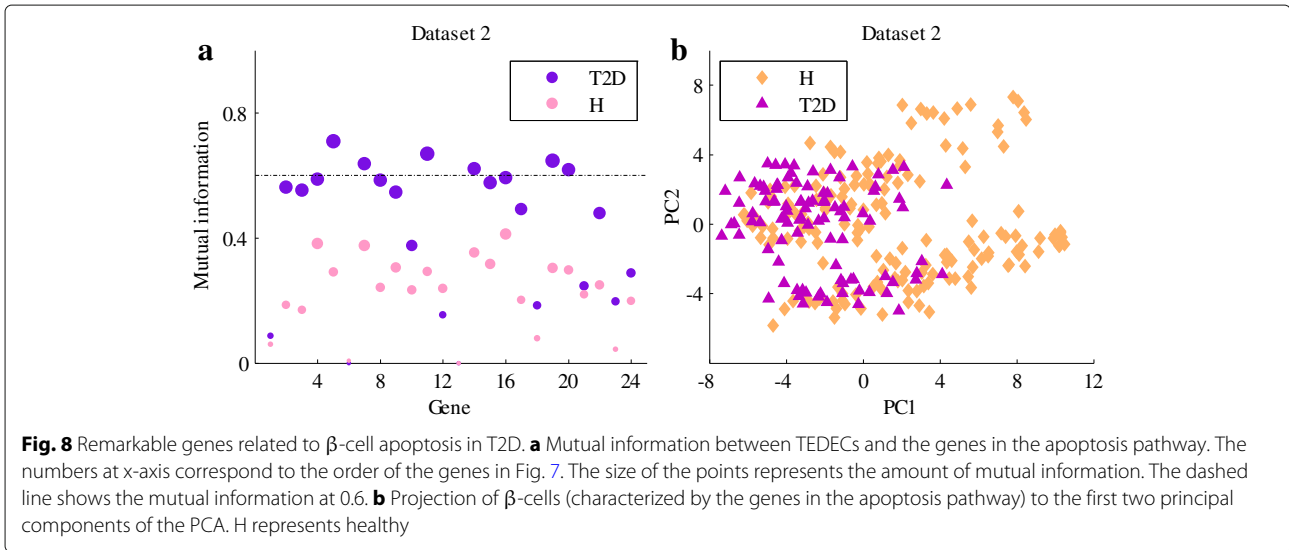


in dataset 2, the median values of the TEDECs of T2D patients are all larger than those of the healthy donors.

Discussion

In this work, we conducted single-cell data analysis to decipher pancreatic β -cell dysfunction and deficit mechanisms in T2D. Three single-cell transcriptomic datasets were employed in our study. Different from bulk-cell data analysis, single-cell data analysis allows us to capture

inter-cell heterogeneity and explore the data deeply to unravel the mechanisms of diseases. It is well known that a major function of pancreatic β -cells is to produce secretory insulin. Thus, we firstly examined the INS expression levels in the β -cells of each dataset. In datasets 2 and 3, INS expression in the T2D β -cells is generally lower than that in the healthy cells, but the expression is similar in healthy and T2D β -cells of dataset 1. To explain the observation of INS expression, we checked genes that are



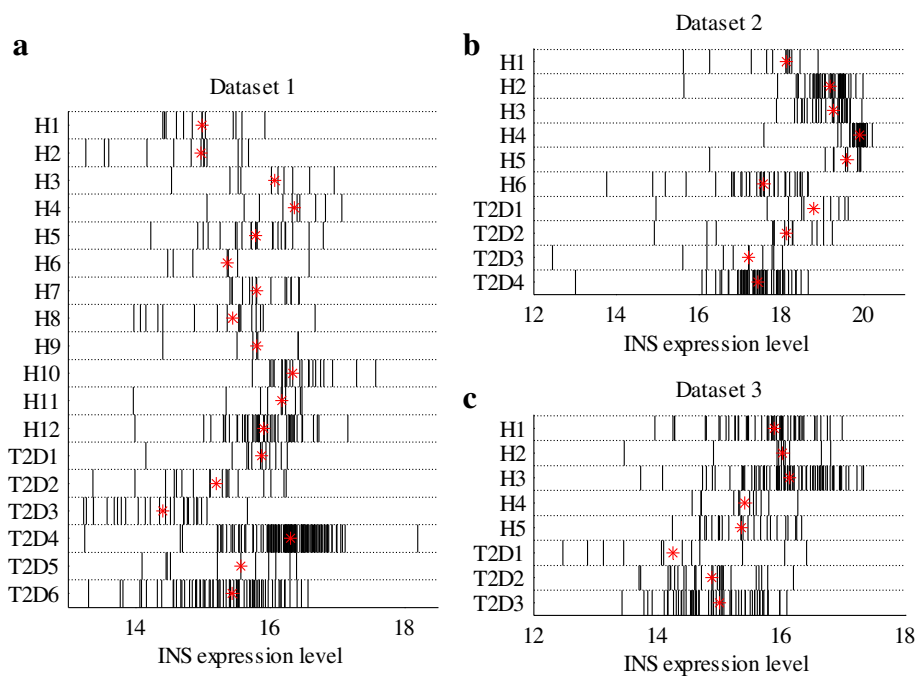


Fig. 9 INS expression levels of β -cells of each donor. **a**, **b** and **c** show the INS expression in donors of datasets 1, 2 and 3, respectively. H_i and $T2D_i$ represent healthy and T2D donors, where i stands for the index number of a donor

related to cellular stress, and found that these genes were lowly expressed in both the healthy and T2D groups of β -cells of dataset 1. In dataset 2, T2D β -cells had high cellular stress while healthy β -cells experienced low stress. Moreover, in dataset 3, the cellular stress in both groups of cells was high.

Considering the INS expression levels and cellular stresses of the three datasets, we obtained the following results. T2D β -cells perform normally in INS expression under low cellular stress (dataset 1), but they behave dysfunctionally under high cellular stress (dataset 2); healthy β -cells can deal with high cellular stress, maintaining INS expression at normal levels despite the stress (dataset 3). To further validate our analysis results, we employed five classifiers to predict the cellular conditions (i.e. healthy or T2D) of the β -cells, using the expression levels of stress- and INS-related genes. We also proposed that β -cells in T2D are vulnerable to stress-induced dysfunction. In other words, under similar cellular stresses, T2D β -cells have abnormal INS expression while healthy β -cells perform normally. This may be caused by the toxic effects of hyperglycaemia and high FFA. Besides, our analysis showed that oxidative stress could be an important influential factor on INS expression. This is consistent with the experimental results in [46], which show that MAFA and PDX1 are inactivated under oxidative stress, resulting in the decrease of insulin secretion of T2D β -cells. Meanwhile, the impaired β -cell function can be repaired

by relieving oxidative stress. For instance, as reported in [47, 48], insulin secretion can be improved in vitro upon treatment with an antioxidant of bis (1-hydroxy-2,2,6,6-tetramethyl-4-piperidinyl) decanoate di-hydrochloride (IAC) in T2D.

T2D is also characterized by a relative deficit of pancreatic β -cells [9, 33]. It has long been demonstrated that β -cell apoptosis would increase in T2D patients and T2D mouse models [33, 34]. As the apoptosis measurements are not available for the three single-cell datasets, we used the TEDECs (i.e. CASP3, CASP6 and CASP7) to estimate the rate of β -cell apoptosis. However, increased apoptosis of T2D β -cells is only observed in dataset 2, whereas in dataset 3 the median value of TEDECs of the healthy β -cells is higher than that of the T2D β -cells. This striking observation needs further clarification as a future work. In addition, we conducted personalized analysis of INS and TEDECs, and showed that INS and TEDECs are different among donors, with T2D patients having lower INS expression and higher apoptosis in β -cells than healthy donors.

Conclusions

In this work, to uncover the mechanisms of β -cell dysfunction and deficit in T2D, we conducted single-cell transcriptomic data analysis. By analyzing the INS expression levels and cellular stresses of three β -cell transcriptomic datasets, we observed that the T2D β -cells perform

normally in INS expression in the condition of low cellular stress but behave dysfunctionally under high stress. The healthy β -cells can deal with high cellular stress and keep INS expression at normal levels. In addition, analyses of correlation and mutual information showed that oxidative stress could be a critical influential factor on INS expression in T2D. This is consistent with some experimental results in the literature. Moreover, we analysed genes related with β -cells death regulation and observed increased apoptosis in T2D cells only in dataset 2, when adopting the TEDECs as an estimation of apoptosis rate. The TNFR1-mediated pathway, mitochondrial BAX-related pathway, as well as the CAPN1- and CAPN2-dependent pathway may play important roles in T2D. Finally, personalized analysis showed some diversity of INS expression among donors.

Materials and methods

Experimental data of single-cell transcriptomes

The data we analysed here were obtained from three published works of Xin et al. [28], Segerstolpe et al. [29] and Lawlor et al. [27]. The gene expression levels in [28] and [29] were reported in reads per kilobase of transcript per million mapped reads (RPKM), while the records in [27] was quantified as transcripts per million (TPM). Due to the different measurements, we only compared gene expression of cells within the same dataset.

Classification of cells

In order to predict cellular states (i.e. healthy or T2D β -cells), we employed five classifiers: Bayesian network, support vector machine (SVM), random forest, logistic regression and neural network (NN). Bayesian network is a probabilistic graphical model represented by a directed acyclic graph. It contains a set of variables as well as their conditional probability distributions. SVM maps the features into a high-dimensional space and conducts classification using hyperplane(s). Random forest is composed of an ensemble of decision trees, and a voting strategy is employed for the final prediction. In logistic regression, a logistic function is used to compute the probability of the dependent variable and to determine the potential class of a sample. NN is constructed with a group of interconnected neurons, which are organized as the input layer, hidden layer and output layer. Detailed information of these algorithms is provided in [49–54]. We employed the algorithms in Weka 3.8.1 to conduct the classification [55], and leave-one-out cross-validation was used for model validation.

The performance of the algorithms is evaluated using the measurements of accuracy and F-measure. Given the number of instances of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), the accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

To address the issue that positive and negative samples are not balanced in this study, we also used F-measure which is the harmonic mean of precision and recall. It is calculated as:

$$F\text{-measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \quad (2)$$

where precision = $TP/(TP + FP)$ and recall = $TP/(TP + FN)$.

Mutual information

Given two discrete random variables X and Y , the mutual information provides a measure of the mutual dependence between them [56]. In terms of information, it measures the obtained information about X obtained through Y , or the uncertainty about X reduced through Y , and vice versa. The mutual information between X and Y is defined as follows,

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (3)$$

where $p(x)$ and $p(y)$ represent the marginal probability distributions, and $p(x, y)$ denotes the joint probability distribution. In this work, we calculated the mutual information by using the entropy and joint entropy:

$$I(X; Y) = H(X) + H(Y) - H(X, Y), \quad (4)$$

where

$$\begin{cases} H(X) = - \sum_{x \in X} p(x) \log p(x) \\ H(Y) = - \sum_{y \in Y} p(y) \log p(y) \\ H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y). \end{cases} \quad (5)$$

$H(X)$ and $H(Y)$ are the entropies of X and Y , while $H(X, Y)$ stands for the joint entropy of the two variables. The derivation of Eq. (4) from Eq. (3) can be found in [56]. We discretized the gene expression data by taking the floor of the values, as we calculated the entropy in a discrete way. In addition, base 2 was employed for the logarithms to compute entropy, implying that the unit of bit was used for measuring the mutual information.

Spearman's rank correlation coefficient

To measure the monotonic relationship between cellular stress and INS expression, we calculated the Spearman's rank correlation coefficient, following the steps given in [57].

Principal component analysis (PCA)

PCA was implemented based on an orthogonal linear transformation, which decorrelates samples of possibly correlated variables. After the transformation, the first principle component has the largest variance, the second

one holds the second largest variance, and so on. Thus, the fundamental goal of PCA is the change of basis, after which a small number of principal components can be identified to provide a reasonable description of the original data. The derivation and instructions for implementation of PCA are available in [58].

Comparison of INS expression

We compared the INS expression levels between two groups of cells using Student's *t*-test. The difference is considered as statistically significant if the *p*-value is less than 0.05.

Additional files

Additional file 1: Table S1. Genes used in cellular condition (i.e. healthy or T2D) prediction. (XLSX 10 kb)

Additional file 2: Tables S2-S3. Entropy, joint entropy and mutual information between TEDECs and the genes in the apoptosis pathway. (XLSX 13 kb)

Additional file 3: Figures S1-S4. Supplementary figures. (DOCX 742 kb)

Abbreviations

ER: Endoplasmic reticulum; FFA: Free fatty acids; FN: False negative; FP: False positive; NN: Neural network; PCA: Principal component analysis; ROS: Reactive oxygen species; RPKM: Reads per kilobase of transcript per million mapped reads; scRNA-seq: single-cell RNA sequencing; SVM: Support vector machine; T2D: Type 2 diabetes; TEDECs: Total expression of the death executioner caspases; TN: True negative; TP: True positive; TPM: Transcripts per million

Acknowledgements

Not applicable.

Funding

Publication costs are funded by Start-Up Grant of ShanghaiTech University.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 19 Supplement 19, 2018: Proceedings of the 29th International Conference on Genome Informatics (GIW 2018): bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-19>.

Authors' contributions

L.M. performed β -cell gene expression analysis, decoded β -cell dysfunction and deficit mechanisms in T2D, and drafted the manuscript. J.Z. initiated the project, participated in the design of the study, and helped draft the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable. We used three published datasets in this article.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Biomedical Informatics Lab, School of Computer Science and Engineering, Nanyang Technological University, 639798 Singapore, Singapore. ²School of Information Science and Technology, ShanghaiTech University, 201210 Shanghai, China.

Published: 31 December 2018

References

- Organization WH, et al. Global Report on Diabetes. France: World Health Organization; 2016.
- Weyer C, Bogardus C, Mott DM, Pratley RE. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest*. 1999;104(6):787–94.
- Bergman RN, Finegood D, Kahn S. The evolution of β -cell dysfunction and insulin resistance in type 2 diabetes. *Eur J Clin Invest*. 2002;32:35–45.
- Kahn S. The importance of β -cell failure in the development and progression of type 2 diabetes. *J Clin Endocrinol Metab*. 2001;86(9):4047–58.
- Rhodes CJ. Type 2 diabetes—a matter of β -cell life and death?. *Sci*. 2005;307(5708):380–4.
- Popa S, Mota M. Type 2 Diabetes. In: Masuo K, editor. Rijeka: IntechOpen; 2013.
- Meier JJ, Bonadonna RC. Role of reduced β -cell mass versus impaired β -cell function in the pathogenesis of type 2 diabetes. *Diabetes Care*. 2013;36(Supplement 2):113–9.
- Sakuraba H, Mizukami H, Yagihashi N, Wada R, Hanyu C, Yagihashi S. Reduced beta-cell mass and expression of oxidative stress-related DNA damage in the islet of Japanese Type II diabetic patients. *Diabetol*. 2002;45(1):85–96.
- Rahier J, Guiot Y, Goebbels R, Sempoux C, Henquin J-C. Pancreatic β -cell mass in European subjects with type 2 diabetes. *Diabetes Obes Metab*. 2008;10:32–42.
- Kahn S. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetol*. 2003;46(1):3–19.
- Yoon KH, Ko SH, Cho JH, Lee JM, Ahn YB, Song KH, Yoo SJ, Kang MI, Cha BY, Lee KW, et al. Selective β -cell loss and α -cell expansion in patients with type 2 diabetes mellitus in Korea. *J Clin Endocrinol Metab*. 2003;88(5):2300–8.
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods*. 2009;6(5):377.
- Svensson V, Natarajan KN, Ly L-H, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell RNA-sequencing experiments. *Nat Methods*. 2017;14(4):381.
- Stegle O, Teichmann SA, Marioni JC. Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet*. 2015;16(3):133.
- Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet*. 2013;14(9):618.
- Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. *Nat Methods*. 2011;8(4 Suppl):S6–11.
- Saadatpour A, Lai S, Guo G, Yuan G-C. Single-cell analysis in cancer genomics. *Trends Genet*. 2015;31(10):576–86.
- Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, Grimes HL. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nat*. 2016;537(7622):698.
- Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*. 2016;166(5):1308–23.
- Janiszewska M, Liu L, Almendro V, Kuang Y, Pawelczak C, Sakr RA, Weigelt B, Hanker AB, Chandralapaty S, King TA, et al. In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat Genet*. 2015;47(10):1212.
- Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betscholtz C, et al. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Sci*. 2015;347(6226):1138–42.

22. Deng Q, Ramsköld D, Reinius B, Sandberg R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Sci*. 2014;343(6167):193–6.
23. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol*. 2015;33(2):155.
24. Lawlor N, Khetan S, Ucar D, Stitzel ML. Genomics of islet (dys) function and type 2 diabetes. *Trends Genet*. 2017;33(4):244–55.
25. Li J, Klughammer J, Farlik M, Penz T, Spittler A, Barbieuc C, Berishvili E, Bock C, Kubicek S. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep*. 2016;17(2):178–87.
26. Wang YJ, Schug J, Won K-J, Liu C, Naji A, Avrahami D, Golson ML, Kaestner KH. Single cell transcriptomics of the human endocrine pancreas. *Diabetes*. 2016;65(10):3028–38.
27. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, Kycia I, Robson P, Stitzel ML. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res*. 2017;27(2):208–22.
28. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, Murphy AJ, Yancopoulos GD, Lin C, Gromada J. Rna sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab*. 2016;24(4):608–15.
29. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab*. 2016;24(4):593–607.
30. Eizirik DL, Cardozo AK, Cnop M. The role for endoplasmic reticulum stress in diabetes mellitus. *Endocr Rev*. 2007;29(1):42–61.
31. Del Prado S. Role of glucotoxicity and lipotoxicity in the pathophysiology of Type 2 diabetes mellitus and emerging treatment strategies. *Diabetic Med*. 2009;26(12):1185–92.
32. Evans JL, Goldfine ID, Maddux BA, Grodsky GM. Oxidative stress and stress-activated signaling pathways: a unifying hypothesis of type 2 diabetes. *Endocr Rev*. 2002;23(5):599–622.
33. Butler AE, Janson J, Bonner-Weir S, Ritzel R, Rizza RA, Butler PC. β -cell deficit and increased β -cell apoptosis in humans with type 2 diabetes. *Diabetes*. 2003;52(1):102–10.
34. Butler AE, Janson J, Soeller WC, Butler PC. Increased β -cell apoptosis prevents adaptive increase in β -cell mass in mouse model of type 2 diabetes: evidence for role of islet amyloid formation rather than direct action of amyloid. *Diabetes*. 2003;52(9):2304–14.
35. Cnop M, Ladriere L, Hekerman P, Ortis F, Cardozo AK, Dogusan Z, Flamez D, Boyce M, Yuan J, Eizirik DL. Selective inhibition of eukaryotic translation initiation factor 2 α dephosphorylation potentiates fatty acid-induced endoplasmic reticulum stress and causes pancreatic β -cell dysfunction and apoptosis. *J Biol Chem*. 2007;282(6):3989–97.
36. Karaskov E, Scott C, Zhang L, Teodoro T, Ravazzola M, Volchuk A. Chronic palmitate but not oleate exposure induces endoplasmic reticulum stress, which may contribute to INS-1 pancreatic β -cell apoptosis. *Endocrinol*. 2006;147(7):3398–407.
37. Seifert EL, Estey C, Xuan JY, Harper M-E. Electron transport chain-dependent and-independent mechanisms of mitochondrial H₂O₂ emission during long-chain fatty acid oxidation. *J Biol Chem*. 2010;285(8):5748–58.
38. Piro S, Anello M, Di Pietro C, Lizzio MN, Patan G, Rabuazzo AM, Vigneri R, Purrello M, Purrello F. Chronic exposure to free fatty acids or high glucose induces apoptosis in rat pancreatic islets: possible role of oxidative stress. *Metab Clin Exp*. 2002;51(10):1340–7.
39. Lowell BB, Shulman GI. Mitochondrial dysfunction and type 2 diabetes. *Sci*. 2005;307(5708):384–7.
40. Nishikawa T, Edelstein D, Du XL, Yamagishi S-i, Matsumura T, Kaneda Y, Yorek MA, Beebe D, Oates PJ, Hammes H-P, et al. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. *Nat*. 2000;404(6779):787.
41. Circo ML, Aw TY. Reactive oxygen species, cellular redox systems, and apoptosis. *Free Radic Biol Med*. 2010;48(6):749–62.
42. Hagman DK, Hays LB, Parazzoli SD, Poitout V. Palmitate inhibits insulin gene expression by altering PDX-1 nuclear localization and reducing MafA expression in isolated rat islets of Langerhans. *J Biol Chem*. 2005;280(37):32413–8.
43. Kaneto H, Xu G, Fujii N, Kim S, Bonner-Weir S, Weir GC. Involvement of c-jun N-terminal kinase in oxidative stress-mediated suppression of insulin gene expression. *J Biol Chem*. 2002;277(33):30010–8.
44. Apoptosis KEGG Pathway Database. http://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map04210&keyword=apoptosis Accessed 15 Mar 2017.
45. GeneGo. <https://portal.genego.com> Accessed 02 Feb 2017.
46. Guo S, Dai C, Guo M, Taylor B, Harmon JS, Sander M, Robertson RP, Powers AC, Stein R. Inactivation of specific β cell transcription factors in type 2 diabetes. *J Clin Invest*. 2013;123(8):3305–16.
47. Lupi R, Del Guerra S, Mancarella R, Novelli M, Valgimigli L, Pedulli G, Paolini M, Soletti A, Filippini F, Mosca F, et al. Insulin secretion defects of human type 2 diabetic islets are corrected in vitro by a new reactive oxygen species scavenger. *Diabetes Metab*. 2007;33(5):340–5.
48. D'Aleo V, Del Guerra S, Martano M, Bonamassa B, Canistro D, Soletti A, Valgimigli L, Paolini M, Filippini F, Boggi U, et al. The non-peptidyl low molecular weight radical scavenger IAC protects human pancreatic islets from lipotoxicity. *Mol Cell Endocrinol*. 2009;309(1-2):63–6.
49. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learning*. 1997;29(2-3):131–63.
50. Cortes C, Vapnik V. Support-vector networks. *Mach Learning*. 1995;20(3):273–97.
51. Breiman L. Random forests. *Mach Learning*. 2001;45(1):5–32.
52. Freedman DA. *Statistical Models: Theory and Practice*. Cambridge: cambridge university press; 2009.
53. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev*. 1958;65(6):386.
54. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nat*. 1986;323(6088):533.
55. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann; 2016.
56. Cover TM, Thomas JA. *Elements of Information Theory*. New Jersey: Wiley; 2012.
57. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15(1):72–101.
58. Shlens J. A tutorial on principal component analysis. 2014. arXiv preprint arXiv:1404.1100.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

