

Article

Estimation of Metabolic Effects upon Cadmium Exposure during Pregnancy Using Tensor Decomposition

Yuki Amakura ^{1,†} and Y-h. Taguchi ^{2,*,†} 

¹ Graduate School of Science and Engineering, Chuo University, Tokyo 112-8551, Japan

² Department of Physics, Chuo University, Tokyo 112-8551, Japan

* Correspondence: tag@granular.com

† These authors contributed equally to this work.

Abstract: A simple tensor decomposition model was applied to the liver transcriptome analysis data to elucidate the cause of cadmium-induced gene overexpression. In addition, we estimated the mechanism by which prenatal Cd exposure disrupts insulin metabolism in offspring. Numerous studies have reported on the toxicity of Cd. A liver transcriptome analysis revealed that Cd toxicity induces intracellular oxidative stress and mitochondrial dysfunction via changes in gene expression, which in turn induces endoplasmic reticulum-associated degradation via abnormal protein folding. However, the specific mechanisms underlying these effects remain unknown. In this study, we found that Cd-induced endoplasmic reticulum stress may promote increased expression of tumor necrosis factor- α (TNF- α). Based on the high expression of genes involved in the production of sphingolipids, it was also found that the accumulation of ceramide may induce intracellular oxidative stress through the overproduction of reactive oxygen species. In addition, the high expression of a set of genes involved in the electron transfer system may contribute to oxidative stress. These findings allowed us to identify the mechanisms by which intracellular oxidative stress leads to the phosphorylation of insulin receptor substrate 1, which plays a significant role in the insulin signaling pathway.

Keywords: cadmium exposure; insulin metabolism disruption; tensor decomposition



Citation: Amakura, Y.; Taguchi, Y.-h. Estimation of Metabolic Effects upon Cadmium Exposure during Pregnancy Using Tensor Decomposition. *Genes* **2022**, *13*, 1698. <https://doi.org/10.3390/genes13101698>

Academic Editors: Yuval Itan and Stefano Lonardi

Received: 5 July 2022

Accepted: 13 September 2022

Published: 22 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cadmium is a non-essential trace metal found in cells [1]. However, when excessive amounts of Cd enter the body owing to inoculation through food or exposure, it accumulates in the cells. As an electrophilic metal, Cd exerts a negative effect on cellular protein molecules [1]. An adverse effect of excessive Cd accumulation is the inhibition of protein folding. For a protein to fully perform its function, its three-dimensional structure must be maintained properly. Thus, unfolded proteins accumulate in the endoplasmic reticulum, the organelle responsible for proper protein folding. The accumulation of unfolded proteins is known to cause endoplasmic reticulum stress, and excessive stress induces endoplasmic reticulum-associated degradation (ERAD). The ubiquitin-proteasome system is activated in response to ERAD induction to degrade unfolded proteins. Conversely, endoplasmic reticulum stress is alleviated when accumulated unfolded proteins undergo ubiquitination, making them targets for proteasome degradation. However, when stress levels exceed the stress-avoidance capacity of ERAD, cells induce apoptosis. ERAD also promotes ATP-dependent responses. Additionally, insulin signaling is initiated when insulin binds to the insulin receptor (IR) on the cell membrane, which ultimately translocates glucose transporter type 4 (GLUT4) to the cell membrane, allowing glucose to enter the cell. The glycolytic system in cells utilizes glucose taken up by the body to produce ATP. When insulin signaling is inhibited, symptoms, such as elevated blood sugar levels, are observed, which also cause type 2 diabetes.

Some reports have suggested that oxidative stress plays a major role in Cd toxicity. Shaikh et al. reported that Cd toxicity was reduced by administering antioxidants to mice

after chronic Cd administration. This suggests that Cd toxicity causes various adverse biological effects via oxidative stress [2].

Several studies have linked Cd toxicity to abnormal insulin metabolism. For example, Takashige et al. reported that mice exposed to high doses of Cd for two weeks might develop abnormal adipocyte differentiation, expansion, and function, leading to insulin resistance, hypertension, and cardiovascular disease [3].

In addition, compelling evidence suggests that prenatal exposure to Cd and other toxic metal contaminants increases the risk of cardiovascular disease and obesity-related morbidity, including type II diabetes, in unborn children [4]. Al-Saleh et al. also reported that prenatal Cd exposure significantly affected birth weight [5]. The data used in this study are based on such reports and are derived from experiments conducted to verify them in mice and to investigate the detailed mechanisms.

The transcriptome analysis results of Jackson et al. (who obtained the data used in this analysis) suggest that Cd exposure induces ERAD and intracellular oxidative stress and mitochondrial dysfunction may cause insulin resistance. Additionally, changes in the expression of genes indicative of lipid abnormalities have been observed in response to Cd exposure [6].

The analysis by Jackson et al. identified differentially expressed genes (DEGs). Specifically, the method was used to determine whether there was a significant difference in expression levels by comparing the ratio of signal values between the genes in the liver transcriptome analysis data of individuals whose mothers were exposed to Cd and those whose mothers were not exposed to Cd at the same point in their lives and of the same gender. In this method, 5789 DEGs were identified for the female postnatal day 42 (PND42) data, which represent more than one-third of the total number of genes in the data identified as DEGs [6].

Thus, although analyses were performed at each time point and the results were compared, no attempts have been made to conduct an integrated analysis of data from all time points or to identify genes that change in response to PND changes.

The linear regression method using data from all time points as explanatory variables (see Section 2.2.1) resulted in the overexpression of 5865 genes. However, to comprehend the mechanism, it is preferable to identify genes that play a decisive role; therefore, the overexpression of more than one-third of all genes is inappropriate.

In this analysis, we used tensor decomposition on the liver transcriptome analysis data of Jackson et al. to identify genes that have some level of time dependence on PND, to identify a smaller number of overexpressed genes relative to the total number of genes, and to estimate which genes play an important role. The study also identified a number of genes that are overexpressed in the PND that are not yet understood. We also aimed to elucidate the mechanisms by which ERAD, intracellular oxidative stress, and mitochondrial dysfunction interact, and to determine how insulin resistance is acquired.

2. Materials and Methods

2.1. Preparation of the Dataset

“GSE150679” (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150679> (accessed on 16 September 2022)) was downloaded from Gene Expression Omnibus (GEO), a database provided by the National Center for Biotechnology Information (NCBI), USA.

This dataset was obtained by euthanizing offspring born from Cd-exposed female parental mice (exposed to 500 ppb CdCl₂ in drinking water from two weeks prior to gestation until birth) and offspring born from control females at PND1, PND21, and PND42, respectively, and performing liver transcriptome analysis. The offspring of Cd-exposed females were similarly exposed to Cd at up to PND 10. For this analysis, they were sequenced with an Illumina (Illumina, Inc. San Diego, CA 92122 USA) HiSeq2500 sequencer. Biological replicates existed for each conditional distinction regarding whether the parent mice were exposed to Cd and when the liver transcriptome analysis was performed. Gene expression levels for each individual in this data were normalized.

The dataset also consisted of three Excel files containing data from PND1, PND21, and PND42.

2.2. Model Selection

2.2.1. Significance of Using Tensor Decomposition

In this analysis, we used the Tucker decomposition among the tensor decompositions, and the algorithm was higher-order singular value decomposition (HOSVD [7]). This method is effective for the “Large p Small n problem”. It is useful in situations where the number of specimens is very small relative to the number of genes, as in this analysis, and where multiple genes are selected as variables with some interpretability. Biological experiments, in general, utilize such small numbers of specimens due to ethical considerations and difficulties in obtaining specimens.

As noted above, Jackson et al. identified DEGs based on postnatal days and compared PND1, PND21, and PND42. By transforming the data into a tensor and applying Tucker decomposition, it is possible to identify genes whose expression changes with the passage of PND. This is another reason for our model selection.

In addition, as we will demonstrate in the next section, the simplest imaginable model, the multiple regression model, performs poorly in analyzing such a “large p small n problem”. Tensor decomposition is also a relatively simple model; however, because biological data are typically complex, it is important to select a simple model such as tensor decomposition.

2.2.2. Inappropriate Variable Selection with Multiple Regression Models

The results of using a multiple regression model are used. We set the explanatory variables as the vectors of length 18 and t_l and s_l , where t_l is the explanatory variable for PND, and s_l is the variable for the presence of Cd exposure.

$$t_l = \begin{pmatrix} 1 \\ 21 \\ 42 \\ \vdots \\ 1 \\ 21 \\ 42 \end{pmatrix} \quad s_l = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The target variable was set to $v_{i'l}$, and the same value was assigned for a vector of length 18, containing the expression levels of gene i' . The elements of $v_{i'l}$ are the expression levels with Cd exposure at $l = 1, \dots, 9$, and subsequent elements are without Cd exposure. Additionally, $l = 1, l = 2$, and $l = 3$ were the expression levels of PND = 1, PND = 21, and PND = 42, respectively. Following that, the elements considered to be biological replicates were stored.

$$v_{i'l} = a_i + b_i t_l + c_i s_l$$

The following multiple regression analysis was repeated 12,795 times for each gene i using these variables. Next, for each of the coefficients b_i and c_i obtained through a multiple regression analysis, we performed the χ -square test and assigned a p value. Then, we identified the gene i for which $p < 0.05$ was obtained for both coefficients. The number of i 's obtained using this method was 5865. However, this method is considered inappropriate because it cannot shortlist the candidate causative genes, since it resulted in the high expression of about 46% of the total genes.

2.3. Data Processing

The data were formatted into tensors for use in tensor analysis. First, sex differences were excluded from this analysis owing to insufficient data from conditionally differentiated littermates to formulate a tensor (this is because only the PND42 file has conditional distinctions for sex, and these distinctions were not sufficient in the PND1 and PND21 data to form the tensor).

Since the number of individuals under the same conditions varied between the three files, three datasets of individuals under the same conditions were selected for further analysis, and the columns of the Excel file were rearranged in the following order. The sample ID clearly indicates the order.

PND1: GSM4556351 GSM4556352 GSM4556353
GSM4556347 GSM4556348 GSM4556349

PND21: GSM4556358 GSM4556359 GSM4556360
GSM4556355 GSM4556356 GSM4556357

PND42: GSM4556365 GSM4556366 GSM4556367
GSM4556361 GSM4556362 GSM4556363

The analysis focused on the 12,795 genes whose gene expression levels were consistently obtained in the three files (only PND42 contained data on the expression levels of more genes than the other two).

2.4. Structure of the Tensor Used in the Analysis

The variables in each tensor dimension are described after the original data have been formatted. Its structure is illustrated in Figure 1.

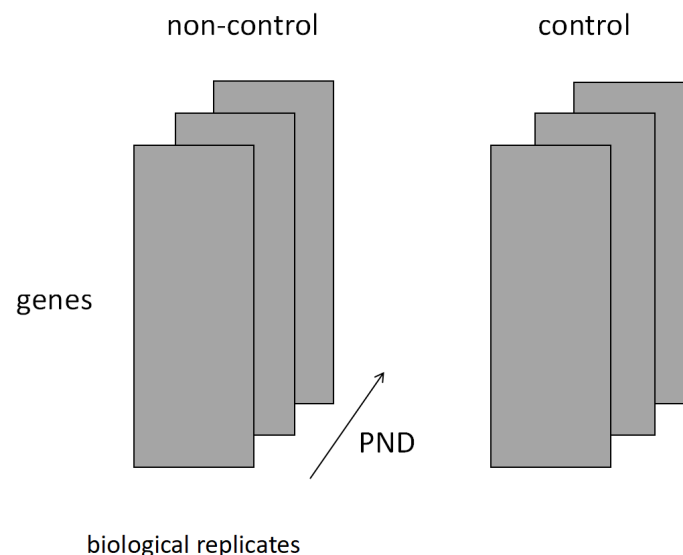


Figure 1. The following figure depicts the four order tensors used in the analysis. The first dimension describes the genes, biological replicates are provided in the second dimension, time course is expressed as postnatal day (PND) in the third dimension, and presence or absence of Cd exposure is indicated in the fourth dimension (non-control refers to offspring born from Cd-exposed mothers, while control refers to offspring born from Cd-unexposed mothers.). Let N , M , K , and L denote the ranks of each dimension, $(N, M, K, L) = (12,795, 3, 3, 2)$.

The first dimension of the tensor was lined with 12,795 genes.

The second dimension consists of a sample of mice under the same conditions selected from the original data as described previously.

The third dimension represents the passage of time since birth. The above samples were in the order of PND1, PND21, and PND42.

The fourth dimension represents the presence or absence of maternal Cd exposure. In Figure 1, 'non-control' represents the Cd-exposed population and 'control' represents the unexposed population.

A fourth-order tensor x_{ijkl} was created using these variables. The parameters representing the rank of the matrix in each dimension were set to $(N, M, K, L) = (12,795, 3, 3, 2)$.

2.5. How to Analysis

2.5.1. Unsupervised Learning Using Tensor Decomposition

In this analysis, we used Tucker decomposition among the tensor decompositions, and the algorithm used was higher order singular value decomposition (HOSVD [7]).

For example, there was a third-order tensor x_{ijk} with ranks N , M , and K in each dimension (the tensor treated in this analysis was a fourth-order tensor, but it was easy to extend the Tucker decomposition from a third- to fourth-order tensor). Tucker decomposition was then performed on this tensor using the orthogonal matrix U and decomposed as follows:

$$x_{ijk} = \sum_{l_1=1}^N \sum_{l_2=1}^M \sum_{l_3=1}^K G(l_1, l_2, l_3) u_{l_1 i} u_{l_2 j} u_{l_3 k}$$

Tucker decomposition decomposes the original data tensor into the core tensor $G(l_1, l_2, l_3)$ and three singular value matrices $u_{l_1 i}$, $u_{l_2 j}$, $u_{l_3 k}$. These matrices represent the dependencies on variables i , j , and k , respectively.

For example, suppose we want to locate a variable i with a large value in the matrix $u_{l_1 i}$ for analysis. The first step is to select the parameters l_2' and l_3' that correspond to some interest dependence in $u_{l_2 j}$ and $u_{l_3 k}$, respectively. For example, if j is a variable that represents time variation, it would be preferable that the vector $u_{l_2 j}$ for the chosen l_2' be straightforward, such as a monotonic increase or decrease vector.

Next, we would substitute the selected l_2' and l_3' into core tensor $G(l_1, l_2, l_3)$. Because each element of the core tensor represents the weight of the product of each vector specified by the parameter (l_1, l_2, l_3) , the parameter l_1 exhibits a large absolute value in the resulting vector $G(l_1', l_2', l_3')$, which exhibits a large absolute value in the vector $G(l_1', l_2', l_3')$, which has a considerable contribution to the original tensor. Based on this idea, we selected l_1' .

Next, we obtain the vector $u_{l_1' i}$ by substituting l_1' with $u_{l_1 i}$. This vector follows the dependency of the chosen parameters l_2', l_3' and has a considerable contribution to the original tensor.

2.5.2. Variable Selection Using χ -Square Test

Finally, the χ -square test was performed using the normalized version of this vector to identify i with a particularly high value among the elements of the obtained vector $u_{l_1' i}$, and a p -value was assigned.

$$p_i = p_{\chi^2} \left[> \sum_{l_1'} \left(\frac{u_{l_1' i}}{\sigma_{l_1'}} \right)^2 \right]$$

The above-mentioned objective was achieved by applying a multiple comparison correction to obtain i with a p -value of less than 0.05.

Figure 2 shows a flowchart of the overall flow of the selection of overexpressed genes in this analysis.

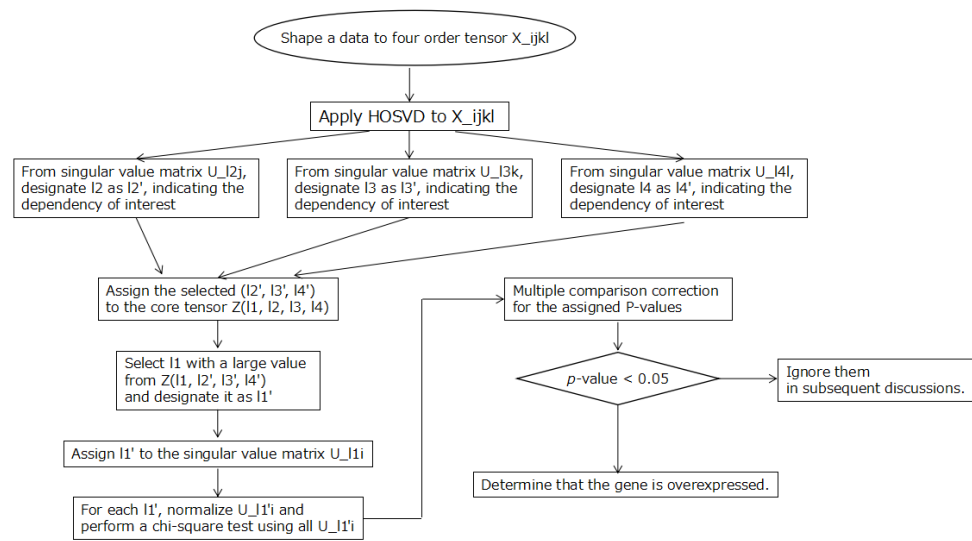


Figure 2. The entire flow of this analysis is represented in the flowchart diagram.

3. Results

3.1. Identification of Genes with High Expression Levels

When we decomposed the fourth-order tensor used in this analysis in R using HOSVD, we obtained a list of Z containing the elements of the core tensor and a list of U consisting of the four singular value matrices.

The tensor used was as described previously, where the subscripts $i, j, k,$ and l denote the gene, biological replicate, time course, and the presence or absence of maternal Cd-exposure, respectively. The purpose of this analysis was to identify the gene i , whose expression level is elevated due to Cd exposure; we investigated whether we could identify a parameter l_1 in the matrix u_{l_1i} that is dependent on Cd exposure, invariant to biological replicates, and sometimes time-dependent. If we could locate a parameter l_1 in the matrix u_{l_1i} that is equivalent to the dependence on Cd exposure, invariance to biological replicates, and some degree of time dependence, we assigned l_1 to u_{l_1i} to obtain the gene i of interest. To accomplish this, we first selected $l_2, l_3,$ and l_4 .

First, u_{l_2j} was extracted from the object U . We chose $l_1 = 1$ because it is a biological replicate; the smaller the individual variation under the identical condition, the better (Figure 3).

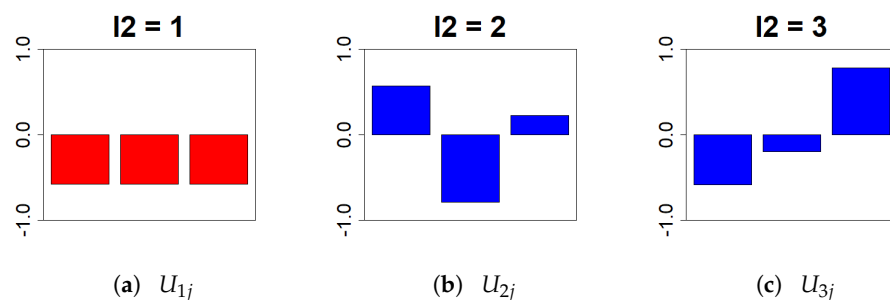


Figure 3. This bar chart represents the values of the singular value matrix U_{l_2j} on the column vector. (First column: (a), second column: (b), and third column: (c)). For the analysis, there should be no variation in gene expression between individuals under identical conditions. Therefore, we chose $l_2 = 1$.

Next, we removed u_{l_3k} . We chose $l_3 = 2$, which follows a monotonic decrease, because extracting a time dependence that follows a monotonic increase or decrease provides a better interpretation of whether the expression level rises or falls as the days pass from birth (Figure 4).

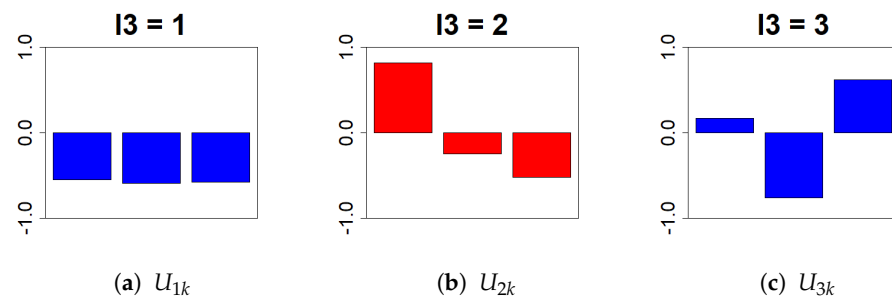


Figure 4. This bar chart represents the values of the singular value matrix $U_{l_3 k}$ on the column vector. (First column: (a), second column: (b), and third column: (c)). The objective is to isolate genes whose expression varies time-dependently due to Cd exposure. Therefore, we selected $l_3 = 2$, which shows a monotonic decrease relative to PND.

Next, we removed $u_{l_4 l}$. Because this study aimed to identify the genes with differential expression in the presence or absence of Cd exposure, $l_4 = 2$ was selected (Figure 5).

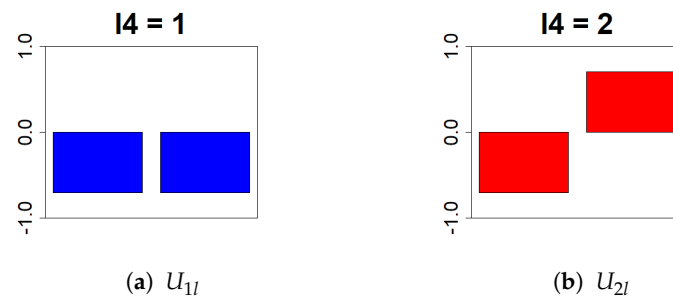


Figure 5. This bar chart represents the values of the singular value matrix $U_{l_4 l}$ on the column vector. (First column: (a), second column: (b)). The objective is to isolate genes that show differences in gene expression levels due to Cd exposure. Therefore, we selected $l_4 = 2$, which has distinct positive and negative values.

Because $l_2 = 1$, $l_3 = 2$, and $l_4 = 2$ were selected, we sought to identify the genes that exhibit small individual differences under the same conditions, differ with respect to maternal Cd exposure, and exhibit a decreasing expression trend over time.

Next, a vector of length 12,795 was obtained by taking the list of objects Z and specifying $G(l_1, 1, 2, 2)$ and the parameters in the core tensor G . The element denoted by l_1 , which has a large absolute value on this vector, is dependent on the parameters $l_2 = 1$, $l_3 = 2$, and $l_4 = 2$, and contributes significantly to the original tensor. Because of the dimensionality of the original tensor, the 19th and subsequent elements on the vector $G(l_1, 1, 2, 2)$ were very small; therefore, only the values of the elements from 1 to 18 are shown in Figure 6. We then selected $l_1 = 3, 11, 14$ with large absolute values as the values of l_1 to be assigned to $u_{l_1 i}$. These were substituted into $u_{l_1 i}$ to generate three vectors.

To identify the genes with relatively high expression levels, we assumed that u_{3i} , u_{11i} , and u_{14i} followed a normal distribution independently, and assigned p -values using the χ -square test. Then, using these three vectors, we executed the `p.adjust` command.

Additionally, the p -value for each gene i was adjusted for multiple comparisons using the `p.adjust` function; then, i with a p -value of less than 0.05 was selected. The selected gene i followed the dependency of interest and corresponded to a gene with a very high expression level. From the total of 12,795 genes, 204 were selected for this analysis.

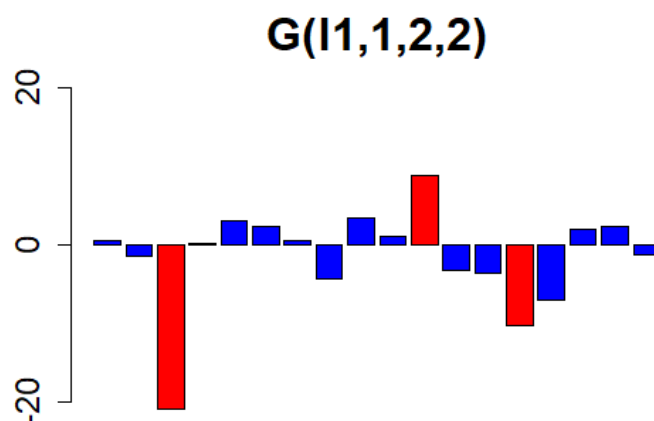


Figure 6. The values of the vectors with ($l_2 = 1$, $l_3 = 2$, and $l_4 = 2$) on the core tensor, from the first element to the eighteenth element, are represented using a bar chart. The top three with the highest values ($l_1 = 3, 11, 14$) were selected. Owing to the rank of the tensor before decomposition, the nineteenth and subsequent elements are omitted because their values are very small.

3.2. Checking the Ontology with gProfiler

“gProfiler” (<https://biit.cs.ut.ee/gprofiler/> (accessed on 16 September 2022)) is a web tool developed by an Estonian research group, the Bioinformatics, Algorithmics and Data Mining Group (BIIT). Ontologies were identified using enrichment analysis on the 204 genes obtained in the above analysis, as shown in Table 1.

Table 1. Enrichment analysis results using gProfiler. The ontology of hits, database IDs, and p -values for each gene were recorded for the 204-gene enrichment analysis.

Database	Term Name	Term ID	p -Value
KEGG	Metabolic pathway	KEGG01100	1.661×10^{-6}
KEGG	Activation of platelets	KEGG04611	2.294×10^{-2}
REAC	Lipid metabolism	R-HSA-556833	1.107×10^{-2}
GO BP	Ionic transport	GO0006811	2.741×10^{-8}
GO BP	Cell adhesion	GO0007155	4.702×10^{-6}
GO BP	Intracellular signaling	GO0035556	4.207×10^{-6}
GO BP	Cellular protein modification process	GO0036211	5.519×10^{-9}
GO BP	Regulation of gene expression	GO0010468	1.241×10^{-8}
GO BP	RNA metabolism	GO0016070	9.895×10^{-11}
GO CC	Mitochondria	GO0005739	1.990×10^{-4}
GO CC	Endoplasmic reticulum	GO0005783	1.990×10^{-3}

Table 1 shows that ontologies include “Lipid metabolism”, “Intracellular signaling”, “Regulation of gene expression”, “Mitochondria”, and “Endoplasmic reticulum”. This could provide a similar interpretation to the results of Jackson et al. In addition, one justification for model selection is the that enrichment analysis has multiple ontology hits. This is because multiple ontology hits are rare, as was the case here when the model was most likely inappropriate and the analysis is failed.

3.3. Evaluation of Analysis Results

This analysis used a simple tensor decomposition model to identify 204 genes that were monotonically expressed in response to PND and whose expression levels varied depending on the presence or absence of Cd exposure. The dependence of a monotonic

decrease on PND was chosen for clarity of interpretation, indicating that we identified a gene whose expression is maximally elevated when PND1 is overexpressed and then decreases gradually. As previously stated, the analysis by Jackson et al. did not attempt to identify genes that change over time, and we believe our analysis is novel in this respect.

The 204 genes identified represent approximately 1.6% of the total, and compared to the expression variation analysis by Jackson et al. and the multiple regression model presented above, the number of overexpressed genes considerably reduced.

The enrichment analysis also included ontologies that appeared to be highly relevant, which lends credence to this analysis.

4. Discussion

4.1. Functions of the Identified Gene

This study aimed to identify the mechanism by which insulin signaling is inhibited. We began by determining whether the ERAD-related genes were overexpressed.

As mentioned previously, the ubiquitin-proteasome system promotes protein degradation during ERAD development. The obtained genes included ubiquitin-conjugating enzyme E2 E3 (UBE2E3) with E2 ubiquitin ligase activity [8], autocrine motility factor receptor (AMFR) with E3 ubiquitin ligase activity [9], and others with ubiquitin ligase activity, such as transcription factor E4F1 (E4F1) [10] and tripartite motif-containing protein 32 (TRIM32) [11]. In addition, the COMM domain-containing protein 9 (COMMD9) regulates these activities [12]. Proteasome-related genes include PSMC2, which encodes a subunit of the 26s proteasome [13], and proteasome activator complex subunit 1 (PSME1), which is part of the proteasome activator complex PA28 the 11S regulator, known as PA28 [14]. The elevated expression of multiple genes involved in the ubiquitin–proteasome system may be responsible for the development of ERAD.

The genes involved in the nuclear factor- κ B (NF- κ B) pathway, which plays a central role in the immune response, were TANK-binding kinase 1 (TBK1), which indirectly interacts with NF- κ B [15]; caspase recruitment domain-containing protein 11 (CARD11), which is an activator [16]; and COMMD9, which is a negative regulator [17], suggesting that NF- κ B may be highly expressed. It is activated by the cytokine TNF- α , and TNF- α may be highly expressed due to the high expression of NFAT activating protein with ITAM motif 1 (NFAM1), which encodes a receptor that activates the promoter of TNF- α [18]. Previous studies have shown a positive correlation between TNF- α expression and development [19], and the results of this study are consistent with this finding. Given that ERAD is ATP-dependent, we verified whether any genes that upregulate ATP production were highly expressed and identified ATP synthase membrane subunit c locus 1 (ATP5G1), which encodes a subunit of the enzyme that catalyzes ATP synthesis.

The genes involved in the electron transfer system were NADH ubiquinone oxidoreductase complex assembly factor 7 (NDUFA7), which is involved in stabilizing the assembly of electron-transfer complex enzymes [20], and electron transfer flavoprotein subunit beta lysine methyltransferase (ETFBKMT), which negatively regulates the electron-transfer system function [21]. The enzyme dihydrolipoamide *s*-acetyltransferase (DLAT), which links the electron-transfer system to the glycolytic system [22], was also highly expressed. The metabolites of glycolysis and fatty acid metabolism are transported to the electron transport system to promote ATP production in mitochondria. The genes involved in fatty acid metabolism were glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1 (GPIHBP1), which promotes fatty acid degradation [23], and peroxisome proliferative activated receptor alpha (PPARA), which encodes a growth factor-activated receptor for peroxisomes [24], the site of fatty acid metabolism. Considering that these metabolites are transferred to the electron transfer system, the activation of the electron transfer system may be responsible for the high expression of genes involved in glycolysis and fatty acid metabolism. The activation of the electron transfer system stimulates ATP production directly.

Additionally, genes involved in sphingolipid biosynthesis, serine palmitoyltransferase, long chain base subunit 2 (SPTLC2) [25] and sterile alpha motif domain containing 8 (SAMMD8) [26], were highly expressed. This suggests that cells may over-synthesize sphingolipids.

4.2. Elucidation of the Mechanism of Insulin Metabolism Inhibition from the Obtained Genes

Cd exposure of the mother also exposes the fetus to Cd [27]. Excessive cellular Cd uptake inhibits protein folding, which leads to ERAD development. This analysis suggests that the high expression of genes involved in ubiquitin–proteasome system glycolysis, fatty acid metabolism, and electron transfer systems promote ATP synthesis, which may assist ERAD development. The positive correlation between the expression levels of ERAD and TNF- α , and the high expression of genes related to TNF- α and NF- κ B, suggest that TNF- α may be over-secreted in the cells. TNF- α is known to activate ASMase (oxidized sphingomyelinase), and ASMase promotes the production of sphingolipids. In this analysis, a high expression of genes involved in sphingolipid synthesis was also observed, suggesting that sphingolipid accumulation may occur. The accumulation of ceramide, a sphingolipid, leads to the overproduction of reactive oxygen species (ROS). Additionally, ROS are produced during the electron transfer system reaction. ROS-mediated oxidative stress activates the apoptosis signal-regulating kinase 1 (ASK1) [28]. Phosphatidylserine decarboxylase proenzyme (PISD) was one of the highly expressed genes identified in this study. It is also involved in the activation of ASK1. ASK1 activates c-Jun N-terminal kinases (JNK), and JNK inactivates it by phosphorylating 307 serine residues of IRS1, which are essential for insulin signaling. Consequently, a pathway for the development of insulin resistance due to oxidative stress via TNF- α -induced ceramide accumulation can be inferred from this analysis. However, from this analysis, we could not elucidate the mechanism underlying the positive correlation between ERAD and the high expression of TNF- α .

In addition to the genes mentioned above, there were also hits for ontologies such as “Regulation of gene expression” and “RNA metabolism”, as shown in Table 1. This suggests that gene expression levels were altered, but the specific mechanism could not be determined in this study.

We also found a number of genes related to apoptosis, which is a last resort for stress avoidance in ERAD development, but it is unclear how apoptosis is actually related to the pathway of insulin resistance acquisition, which we speculated in this study.

In future studies, we plan to focus on determining how these factors are involved and, in particular, what causes and influences gene expression changes.

5. Conclusions

Based on a study reporting that Cd exposure during pregnancy disrupts insulin metabolism in offspring, 204 overexpressed genes were identified using tensor decomposition with liver transcriptome analysis data to elucidate the cause of this biological phenomenon.

Included among the 204 genes were those involved in ERAD development and the ubiquitin-proteasome system, a stress avoidance pathway, implying that Cd toxicity inhibits protein folding. A group of genes involved in sphingolipid production and the electron transfer system was also highly expressed, indicating excessive intracellular oxidative stress. The ERAD and intracellular oxidative stress results were consistent with those of Jackson et al. [6].

It has been suggested that TNF- α may serve as a link between ERAD and high sphingolipid expression caused by Cd toxicity. The high expression of genes encoding the promoter of TNF- α and related to NF- κ B supported this hypothesis. However, this analysis was not able to determine the precise mechanism of the association between ERAD and TNF- α . Based on the association between TNF- α and sphingolipids, we could deduce a pathway by which TNF- α activates ASMase.

An intracellular oxidative stress-based pathway was hypothesized regarding the onset of insulin resistance.

The effects of cadmium on changes in gene expression at the transcriptional level must continue to be investigated.

Author Contributions: Y.T. planned the research. Y.A. performed analyses. Y.T. and Y.A. evaluated the results, discussions, and outcomes, and wrote and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Japan Society for the Promotion of Science, KAKENHI [grant numbers 19H05270, 20K12067, 20H04848] to Y.T.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Gene Expression Omnibus (GEO), a database provided by the National Center for Biotechnology Information (NCBI), USA at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150679> (accessed on 16 September 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ohba, K.I. Transport and Toxicity of Cadmium. *Nichieishi* **2006**, *73*, 269–274. [[CrossRef](#)] [[PubMed](#)]
2. Shaikh, Z.A.; Vu, T.T.; Zaman, K. Oxidative stress as a mechanism of chronic cadmium-induced hepatotoxicity and renal toxicity and protection by antioxidants. *Toxicol. Appl. Pharmacol.* **1999**, *154*, 256–263. [[CrossRef](#)] [[PubMed](#)]
3. Takashige, K.; Hiroyuki, S.; Rie, F.; Yoshito, K.; Masahisa, I.; Kojun, S.; Shinya, S.; Masao, S. Cadmium reduces adipocyte size and expression levels of adiponectin and Peg1/Mest in adipose tissue. *Toxicology* **2010**, *267*, 20–26.
4. Heindel, J.J.; Bruce, B.; Mathew, C.; Ronit, M.; Alberto, M.; Michelle, A.M.; Angel, N.; Paola, P.; Giancarlo, P. Metabolism disrupting chemicals and metabolic disorders. *Reprod. Toxicol.* **2017**, *68*, 3–33. [[CrossRef](#)]
5. Luo, Y.; McCullough, L.E.; Tzeng, J.Y.; Darrah, T.; Vengosh, A.; Maguire, R.L.; Maity, A.; Samuel-Hodge, C.; Murphy, S.K.; Mendez, M.A.; et al. Maternal blood cadmium, lead and arsenic levels, nutrient combinations, and offspring birth weight. *BMC Public Health* **2017**, *17*, 354.
6. Thomas, W.J.; Garret, L.R.; Chris, M.S.; Aubrey, L.S.; Guillette, T.C.; Scott, M.B. Gestational Cd Exposure in the CD-1 Mouse Induces Sex-Specific Hepatic Insulin Insensitivity, Obesity, and Metabolic Syndrome in Adult Female Offspring. *Toxicol. Sci.* **2020**, *178*, 264–280.
7. Taguchi, Y. *Unsupervised Feature Extraction Applied to Bioinformatics: A PCA Based and TD Based Approach*; Springer International Publishing AG: Cham, Switzerland, 2019; 329p.
8. Ito, K.; Kato, S.; Matsuda, Y.; Kimura, M.; Okano, Y. cDNA cloning, characterization, and chromosome mapping of UBE2E3(alias Ubch9), encoding an N-terminally extended human ubiquitin-conjugating enzyme. *Cytogenet. Cell Genet.* **1999**, *84*, 99–104.
9. Marilyn, R.; Jacky, G.G.; Pascal, S.P.; Hao, P.; Monique, L.; Michel, B.; Phuong, U.L.; Ivan, R.N. The gene product of the gp78/AMFR ubiquitin E3 ligase cDNA is selectively recognized by the 3F3A antibody within a subdomain of the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1316–1322.
10. Laurent, L.C.; Laëtitia, K.L.; Conception, P.; Eric, J.; Matthieu, L.; Elodie, H.; Robinson, T.; Guillaume, B.; Ayelet, S.; Manuel, S.R.; et al. E4F1 Is an Atypical Ubiquitin Ligase that Modulates p53 Effector Functions Independently of Degradation. *Cell* **2006**, *127*, 775–788.
11. Annie, P.C.; John, S.B.; Hsan-Jan, Y.; Marwan, K.T.; Todd, E.S.; Ruth, E.S.; Darryl, Y.N.; Terry, A.B.; Kwang-Youn, A.K.; Jian, H.; et al. Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet–Biedl syndrome gene (BBS11). *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6287–6292.
12. Zhan, W.; Wang, W.; Han, T.; Xie, C.; Zhang, T.; Gan, M.; Wang, J.B. COMMD9 promotes TFDP1/E2F1 transcriptional activity via interaction with TFDP1 in non-small cell lung cancer. *Cell Signal.* **2017**, *30*, 59–66. [[CrossRef](#)]
13. Tanahashi, N.; Suzuki, M.; Fujiwara, T.; Takahashi, E.I.; Shimbara, N.; Chung, C.H.; Tanaka, K. Chromosomal Localization and Immunological Analysis of a Family of Human 26S Proteasomal ATPases. *Biochem. Biophys. Res. Commun.* **1998**, *243*, 229–232. [[CrossRef](#)]
14. Ma, C.P.; Slaughter, C.A.; DeMartino, G.N. Identification, purification, and characterization of a protein activator (PA28) of the 20 S proteasome (macropain). *J. Biol. Chem.* **1992**, *267*, 10515–10523. [[CrossRef](#)]
15. Pomerantz, J.L.; Baltimore, D. NF-kappaB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **1999**, *18*, 6694–6704. [[CrossRef](#)]
16. Bertin, J.; Wang, L.; Guo, Y.; Jacobson, M.D.; Poyet, J.L.; Srinivasula, S.M.; Merriam, S.; DiStefano, P.S.; Alnemri, E.S. CARD11 and CARD14 Are Novel Caspase Recruitment Domain (CARD)/Membrane-associated Guanylate Kinase (MAGUK) Family Members that Interact with BCL10 and Activate NF- κ B. *J. Biol. Chem.* **2001**, *276*, 11877–11882. [[CrossRef](#)]

17. Maine, G.N.; Burstein, E. COMMD proteins: COMMMing to the scene. *Cell. Mol. Life Sci.* **2007**, *64*, 1997–2005. [[CrossRef](#)]
18. Martha, L.S.; Abbie, L.; Kristina, L.B.; Roger, K.W. Tumor necrosis factor-related genes and colon and rectal cancer. *Int. J. Mol. Epidemiol. Genet.* **2011**, *2*, 328–338.
19. Liu, Q.; Yang, X.; Long, G.; Hu, Y.; Gu, Z.; Boisclair, Y.R.; Long, Q. Endoplasmic Reticulum Associated Degradation(ERAD) Deficiency Promotes Mitochondrial Dysfunction and Transcriptional Rewiring in Human Hepatic Cells. *J. Biol. Chem.* **2020**, *295*, 16743–16753. [[CrossRef](#)]
20. Olga, Z.R.; Lissiene, S.N.; Florin, S.; Eric, A.S. The arginine methyltransferase NDUFAF7 is essential for complex I assembly and early vertebrate embryogenesis. *Hum. Mol. Genet.* **2014**, *23*, 5159–5170.
21. Małeckı, J.; Jakobsson, M.E.; Ho, A.Y.; Moen, A.; Rustan, A.C.; Falnes, P.Ø. Uncovering human METTL12 as a mitochondrial methyltransferase that modulates citrate synthase activity through metabolite-sensitive lysine methylation. *J. Biol. Chem.* **2017**, *292*, 17950–17962. [[CrossRef](#)]
22. Rommel, A.M.; Todd, M.G.; Adam, O.; Hanna, G.B.; Rumela, C.; Elizabeth, A.R.; Yibin, K.; Thomas, S.; Ileana, M.C. Sirtuin 4 is a lipoamidase regulating pyruvate dehydrogenase complex activity. *Cell* **2014**, *159*, 1615–1625.
23. Young, S.G.; Davies, B.S.J.; Fong, L.G.; Gin, P.; Weinstein, M.M.; Bensadoun, A.; Beigneux, A.P. GPIHBP1: An endothelial cell molecule important for the lipolytic processing of chylomicrons. *Curr. Opin. Lipidol.* **2007**, *18*, 389–396. [[CrossRef](#)] [[PubMed](#)]
24. Sander, K. Integrated physiology and systems biology of PPAR α . *Mol. Metab.* **2014**, *3*, 354–371.
25. Akio, K. Various Metabolisms, Physiological Functions and Related Diseases of Fatty Acids. *Biochemistry* **2010**, *82*, 591–605.
26. Fikadu, G.T.; Ana, M.V.; Elleke, F.B.; Martin, H.; Amrita, J.; Angelika, H.; Pentti, S.; Joost, C.M.H. Sphingomyelin synthase-related protein SMSr is a suppressor of ceramide-induced mitochondrial apoptosis. *J. Cell Sci.* **2014**, *127*, 445–454.
27. Atsuhiro, N.; Masaji, O.; Yukiko, O.; Ichiro, W. Placental Transfer of Cadmium in Pregnant Women. *Jpn. J. Hyg.* **1989**, *44*, 579–586.
28. Tsuruzoe, K.; Nishikawa, T.; Araki, E. Oxidative Stress and Insulin Resistance. *Diabetes* **2006**, *49*, 845–848.