# Finding disentangled representations using VAE

Raymond d'Anjou

Supervisors:
Marcel Reinders, Stavros Makrodimitris,
Tamim Abdelaal, Mohammed Charrout,
Mostafa elTager

Delft University of Technology

June 27, 2021

## 1 Abstract

This study presents a comparison of different Variational Autoencoder(VAE) models to see which VAE models are better at finding disentangled representations. Specifically their ability to encode biological processes into distinct latent dimensions. The biological processes that will be looked at are the cell cycle and differentiation state. The cell cycle is expressed as a S- and G2M-Score and the differentiation state is expressed as a number that quantifies the development time of the cells. First the models will be trained, after that the models will be evaluated. The evaluation is done by checking the latent dimensions for a correlation with the two aforementioned biological processes. From this it became quite clear that VAE and DIP-VAE performed the worst out of the four models tested. On the other hand $\beta$-VAE and $\beta$-TCVAE performed by far the best.

## 2 Introduction

An attractive class of deep learning models is Variational Autoencoder (VAE). The original VAE model was first defined by Kingma and Welling in 2012 [1]. These models can find a low dimensional representation of the original data, effectively compressing the data. However, VAE can not find a low-dimensional representation of the original data that is also disentangled. Disentangled representations are interesting because, each dimension of a disentangled representation encodes something meaningful and distinct. What this means is that each dimension captures a unique aspect of the dataset. Since 2012, numerous extensions for the original VAE have been created. Some of these extensions have been shown to be better at finding disentangled representations.

VAEs have been used a lot in the context of single cell data analysis [2]. This data is in the form of gene expressions and gene RNA counts measured in each cell. For this paper, the goal is to disentangle biological processes. However, benchmarking the performance of each model in finding these biological processes as a disentangled representation is difficult. This is due to the fact that in the data that is used for this study there are no true labels. Therefore, this study will mostly focus on data that we know how to get which in this case are the cell cycle score and differentiation state.

This leads to the research question: "How good are different VAE models at learning disentangled representations?" This will be done by comparing how good the different VAE models are at disentangling the cell cycle score and differentiation state from the data. The goal for the VAE models is to encode each biological process into a single latent dimension. The goal is for the correlation between this biological process and that single latent dimension to be as high as possible.

## 3 Methodology

### 3.1 Data collection

The dataset that is used for this research is cellular data from the mouse hippocampus. The data consists of both neuronal and non-neuronal cells. Due to the fact that these cells are still developing we can measure cells that are still in the progenitor phase. The data was obtained from "RNA velocity of single cells" by La Manno et al [3].

The data consists of 18219 cells and for each cell there is a measurement for every gene. However, a majority of this information about the genes are of bad quality. Therefore only the genes with a clear signal or high variance have been used. This brings the amount of genes down to 2239, the data is in the form of a gene expression. A gene expression is a numerical representation of the amount that a gene is currently expressing itself. Besides gene expression information there is also a pre-calculated cell cycle score for the S- and G2M-Phase. In addition to the cell cycle scores there is a number that expresses how far on a cell is in the differentiation process. Every cell can differentiate into other cells and at the beginning of this process the cells are dividing more actively. This number expresses how far they are into this process and is an indicator of how actively they are still dividing.

### 3.2 VAE models

The VAE extensions that are evaluated in this paper are: VAE, $\beta$-VAE, $\beta$-TCVAE, DIP-VAE. VAE is evaluated to be able to compare the other three extension with it and see how well they perform in comparison. In general each VAE model has the same layout, an encoder, sampling process and a decoder. The encoder transforms the high dimensional data to a low dimensional representation of the data, after which it is sampled to form the latent representation of the data. This latent representation is then the input

for the decoder which aims to reconstruct the original high dimensional data as accurately as possible.

#### 3.2.1 VAE

The original VAE model was the first one to be defined. The loss function is composed out of two parts the first part being the Kullback-Leibler divergence(KLD), which is important for the regularization of the model. The KLD is defined in the following way.

$$KLD = kld\_weight * KL(N(\sigma, \mu), N(0, 1)) \tag{1}$$

The second part is the reconstruction loss which is the mean square error(mse) between the reconstructed input and the actual input. Then the complete loss function is of the following form:

$$\begin{aligned} loss = mse(reconstructed, input) + \\ kld\_weight * KL(N(\sigma, \mu), N(0, 1)) \end{aligned} \tag{2}$$

#### 3.2.2 $\beta$-VAE

It has been shown that $\beta$-VAE is better at learning disentangled representations than vanilla VAE [4]. $\beta$-VAE attempts to adjust the balance between the reconstruction-loss(RL) and Kullback-Leibler divergence(KLD) by introducing a variable $\beta$ that is multiplied with the KLD. When $\beta > 1$ the VAE has the ability to start producing higher degrees of disentanglement [5]. The loss function of $\beta$-VAE:

$$loss = mse(reconstructed, input) + \beta * KLD \tag{3}$$

#### 3.2.3 DIP-VAE

What DIP-VAE tries to do is match the covariance of the prior distribution and the latent distribution [6]. It uses a variable lambda to control how much contribution should be made towards the disentanglement objective [7]. This could result into higher degrees of disentanglement.

#### 3.2.4 $\beta$-TCVAE

The $\beta$ total correlation VAE or $\beta$-TCVAE for short, is another extension based on the original VAE model. This model aims to maximize the mutual information between the data variables and latent variables. It tries doing this while minimizing the mutual information between the latent variables [8]. The loss function is of the following form:

$$\begin{aligned} loss = mse(reconstructed, input) \\ - distance(prior, posterior) - (\beta - 1) * KLD \end{aligned} \tag{4}$$

### 3.3 Experimental setup

Each VAE model that was evaluated had the same network layout. The encoder was a simple linear model with 1 hidden layer. This hidden layer was a fully connected layer and contained 400 neurons. The encoder leads into the variance and mean layer, which are both fully connected layer where the amount of neurons is the same as the latent space size. The latent space had a size of 128, therefore both of these layers contained 128 neurons.

The sampling of the results of these two layers results into a representation of the original data, but in 128 variables. The result of this forms the input of the decoder. The decoder contains two hidden layers that are fully connected. The first hidden layer contains 400 neurons, a batch normalization layer and a rectified linear unit. Batch normalization normalizes over each batch so that the values of that batch are between 0 and 1. The second hidden layer is just a fully connected layer which contains 400 neurons that scale down to the amount of output features.

Furthermore the data that was used for this research had a fair amount of noise in it. This resulted in the VAE models having a difficult time in properly learning from the data. This lead to something that is often described as KLD vanishing in literature [9]. Multiple solutions to this problem have been proposed such as using a weight for the KLD that increases overtime during training [9]. However, it was decided to keep the models as close to their original implementation as possible, therefore none of proposed solutions have been used.

Instead what was done to resolve this problem was only using 300 of the most informative genes present in the data. Using this alone did not result in the models being able to learn the data. The second part of the solution added a KLD weight with a small value. This is so that the KLD has a smaller effect on the model performance. To illustrate this, after modifying the loss function for Vanilla VAE it looks like this:

$$\begin{aligned} loss = mse(reconstructed, input) + \\ 0.00001 * KL(N(\sigma, \mu), N(0, 1)) \end{aligned} \tag{5}$$

These modifications were applied to all the different VAE models, this was done not only to make them work but also to keep the changes consistent across all the different VAE models.

After the models have been trained for a set amount of epochs, the encoders will be used to encode the data into the compressed latent representation. Then for every latent dimension the correlation between that dimension and the cell cycle score or differentiation state is calculated. The correlation coefficient that will be used for calculating this is the Pearson correlation coefficient.

## 4 Results

The results are a collection of scatter plots, histograms and tables. These are used to depict correlations between either the cell cycle score or latent time. The higher the correlation the better, however it is also desirable that a biological process is encoded only once in the latent dimension.

### 4.1 Cell cycle scores

The cell cycle score is one of the two biological processes that will be used for this study to measure how good different VAE models are at finding disentangled representations. The purpose of using this as measurement is to see how well these VAE models can encode the S- and G2M-score into single latent dimensions. The higher the correlation between a score and a dimension the better, however it is also desired that it is encoded in as few dimensions as possible and optimally it is only encoded in one dimension. The best correlations found by every different VAE model can be found in table 1.

| Model | Parameters | S-Score | G2M-Score |
|---|---|---|---|
| VAE | N/A | -0.2620 | 0.3037 |
| $\beta$-VAE | $\beta = 1000$ | 0.5332 | 0.6518 |
| DIP-VAE | $\lambda$-diag = 1.0<br>$\lambda$-offdiag=1.5 | 0.2403 | 0.3121 |
| $\beta$-TCVAE | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ | -0.5405 | -0.9157 |

Table 1: The best correlations found in a single run by each model.

In general VAE seems to perform the worst as on average it finds the lowest correlations with the two different cell cycle scores. However, often VAE and DIP-VAE seemed to perform similarly and one could even argue that VAE was performing better. Besides this, both $\beta$-VAE and $\beta$-TCVAE performed a lot better than the other two VAE models. The difference in performance becomes even more clear when looking at figure 1. This figure shows the histogram of correlation values between every latent dimension and the cell cycle scores.

Furthermore, when looking at figure 1 it becomes clear that $\beta$-VAE and $\beta$-TCVAE perform much better than Vanilla VAE and DIP-VAE. In this figure it is also clearly visible that the two $\beta$ models clearly outperform the other two models. Moreover these two models found high correlations and encoded in as few dimensions as possible.

Additionally figures 2-5 depict the best correlations found by each model. For $\beta$-VAE and $\beta$-TCVAE it is clearly visible that there are correlation since the dots are starting to orientate more on a single line. However, quite the opposite is visible for Vanilla VAE and DIP-VAE where there seems to be no pattern at all in the scatter plots.

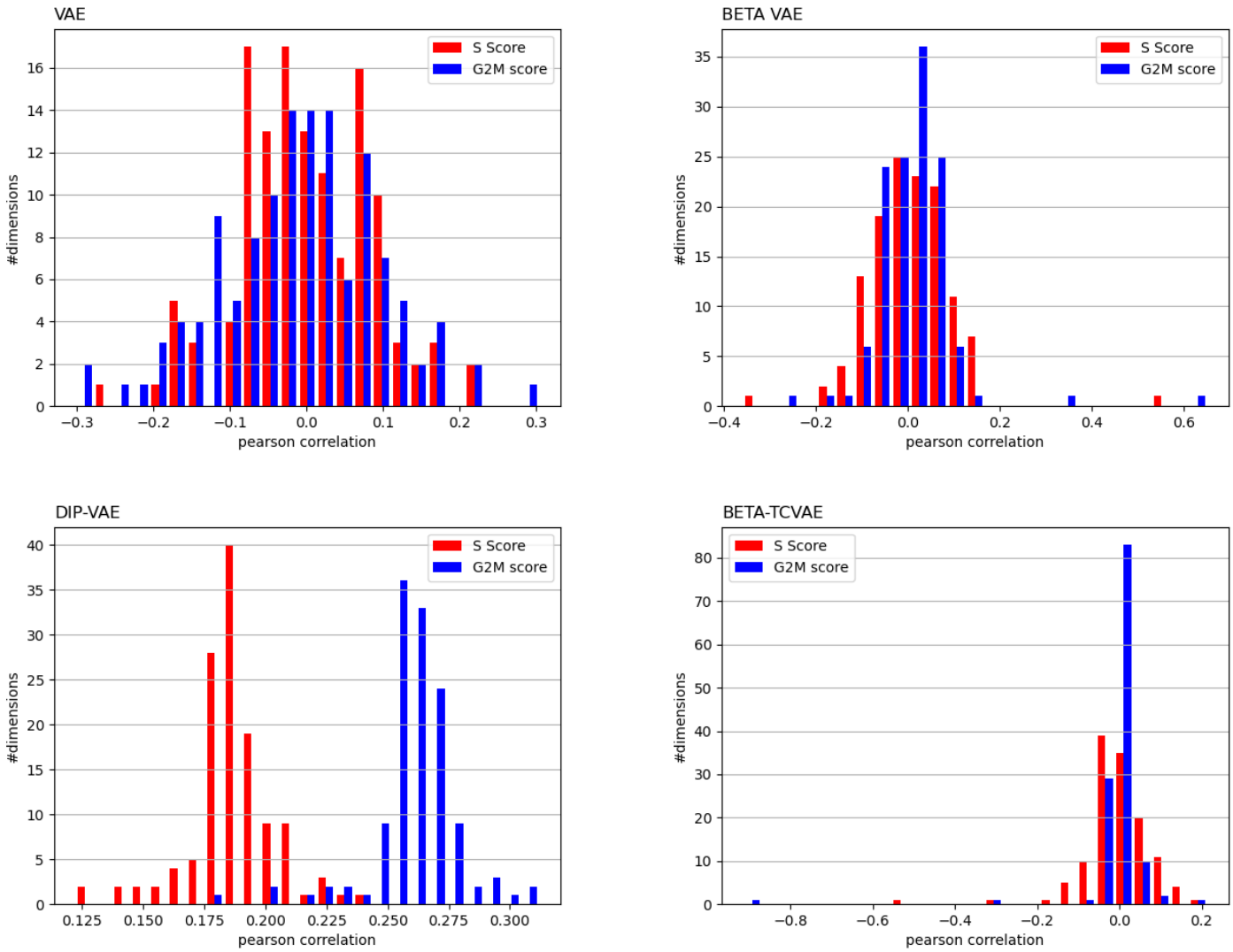


Figure 1: Depicts how many dimensions correlate a certain amount with the two different cell cycle scores, each figure is for a different model. Each model was ran for 50 epochs and with a latent dimension of 128. A: Vanilla VAE, B: $\beta$-VAE, with parameters: $\beta = 1000$, C: DIP-VAE, D: $\beta$-TCVAE, with parameters: $\alpha = 1.0$, $\beta = 1.0$, $\gamma = 1.0$

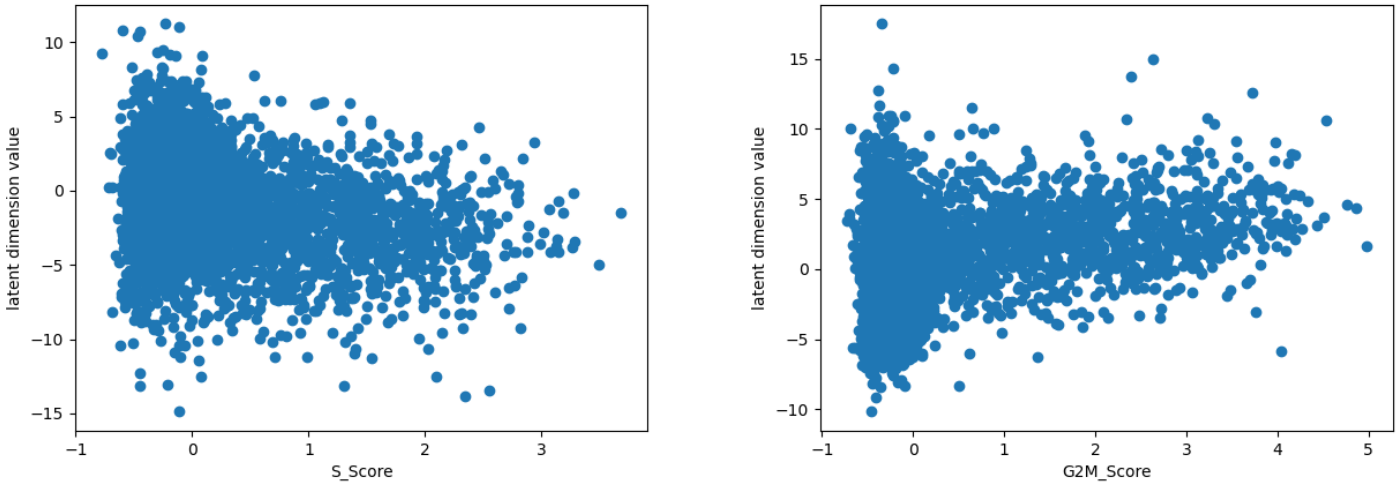


Figure 2: Scatterplots for the best correlations with the cell cycle scores that were encoded into a latent dimension by Vanilla VAE. The correlations found are, S-Score: -0.2620, G2M-Score: 0.3037

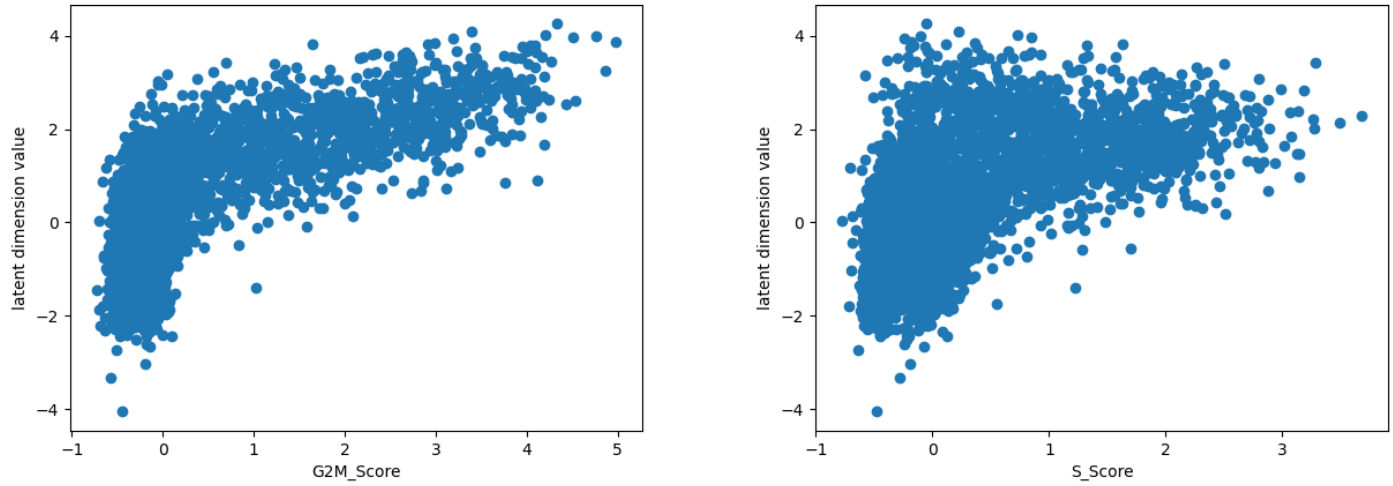


Figure 3: Scatterplots for the best correlations with the cell cycle scores that were encoded into a latent dimension by $\beta$-VAE. The correlations found are, S-Score: 0.5332, G2M-Score: 0.6518

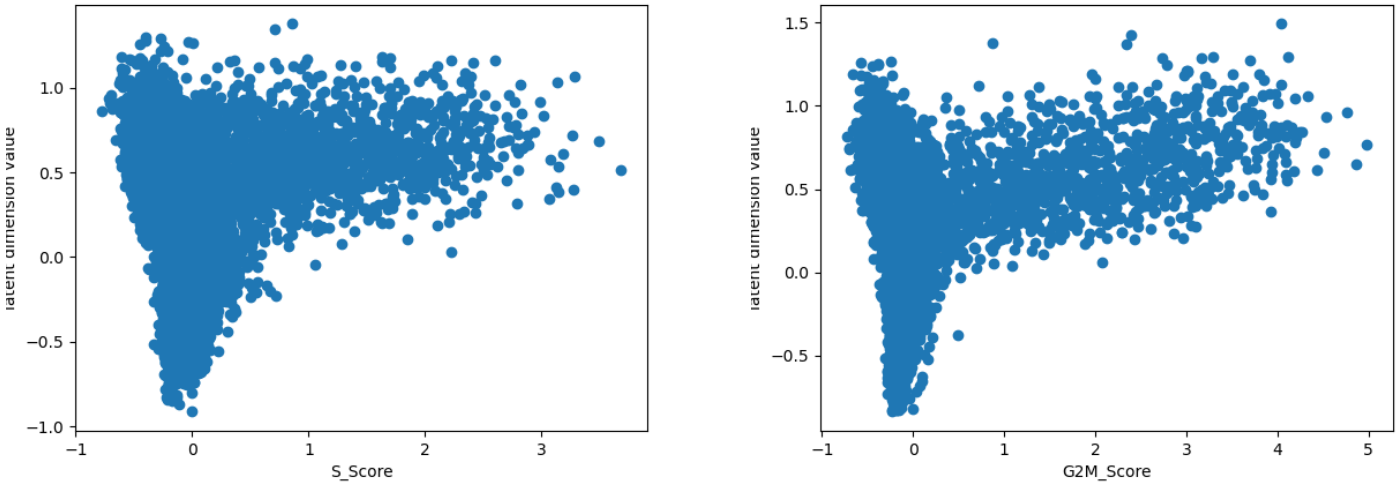


Figure 4: Scatterplots for the best correlations with the cell cycle scores that were encoded into a latent dimension by DIP-VAE. The correlations found are, S-Score: 0.2403, G2M-Score: 0.3121

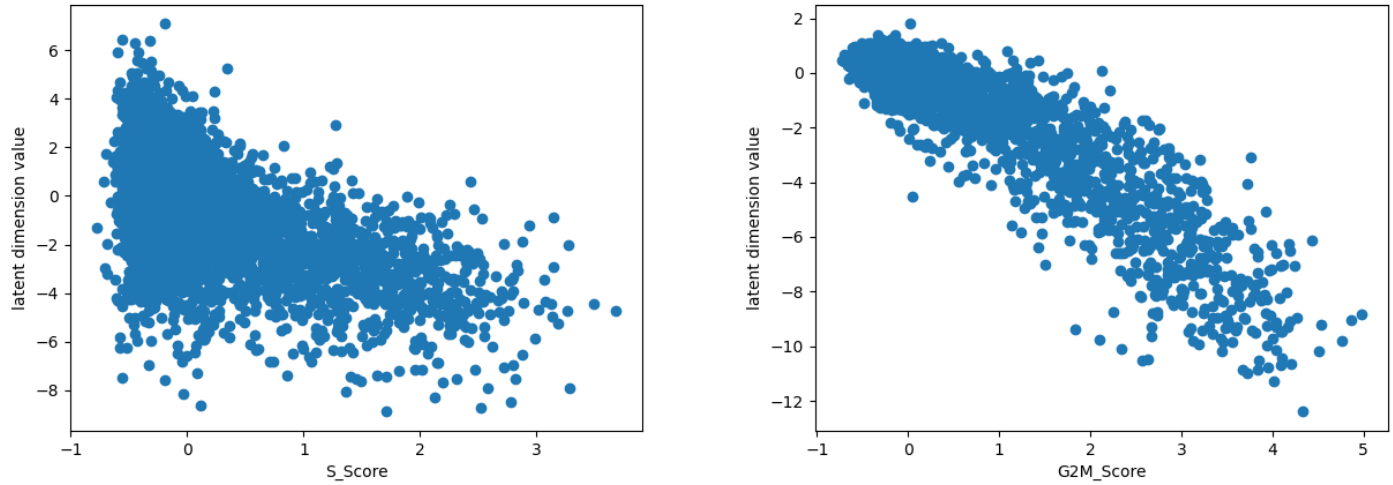


Figure 5: Scatterplots for the best correlations with the cell cycle scores that were encoded into a latent dimension by $\beta$-TCVAE. The correlations found are, S-Score: -0.5405, G2M-Score: -0.9157

## 4.2 Differentiation state

The differentiation state is the second and final biological process that is used to measure how good different VAE models are at finding disentangled representations. The differentiation state is expressed as the latent time.

### 4.2.1 Latent time over all cells

The goal is to see if the different VAE models are able to encode the latent time for all cells. This means that for this measurement no filter is applied to which cells are used to measure the correlation. The best correlation found by every different VAE model can be found in table 2.

| Model | Parameters | Latent time |
|---|---|---|
| VAE | N/A | -0.0297 |
| $\beta$-VAE | $\beta = 1000$ | 0.0345 |
| DIP-VAE | $\lambda$-diag = 1.0<br>$\lambda$-offdiag=1.5 | 0.0152 |
| $\beta$-TCVAE | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ | -0.0277 |

Table 2: The best correlations found in a single run by each model.

When looking at table 2 it becomes clear that none of the VAE models is exceptionally good at encoding the latent time for all cells into a single latent dimension. This is further reinforced when looking at figure 6. An important thing to notice in figure 6 is how small the correlation values are. In this figure it is also clearly visible that all the latent dimensions of every VAE model can only find a low correlation with the latent time for all cells.

### 4.2.2 Latent time per cell cycle phase

Now instead of just looking at all the cells, the cells will be filtered on which cell cycle phase they are currently in. This results in having to check correlations for cells that are in 3 different phases, the phases being: G1, S and G2M. The best correlations for each phase per model can be found in table 3.

| Model | Parameters | S | G1 | G2M |
|---|---|---|---|---|
| VAE | N/A | 0.0567 | 0.0345 | 0.0892 |
| $\beta$-VAE | $\beta = 1000$ | -0.0639 | -0.0329 | -0.0848 |
| DIP-VAE | $\lambda$-diag = 1.0<br>$\lambda$-offdiag=1.5 | -0.0526 | -0.0156 | 0.0537 |
| $\beta$-TCVAE | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ | -0.0595 | -0.0390 | 0.0787 |

Table 3: The best correlations found in a single run by each model.

The correlations found are slightly higher than when looking at all cells. However, looking at table 3 it seems as if the models still really did not learn anything about it. The correlations are still quite low. When looking at figure 7 it is quite clear that DIP-VAE performs the worst of all the VAE models. What also stands out is that for VAE, $\beta$-VAE and $\beta$-TCVAE is that the highest correlations are only captured in a single dimension.

A remarkable aspect of table 3 and figure 7 is that it is quite clear that the VAE models have the most difficult time encoding the latent time for the G1-phase. On average the VAE models had a much easier time encoding the latent time for the S-phase and the G2M-phase compared to the G1-phase.

### 4.2.3 Latent time over cell types

Here the cells will be split up even further. Now instead of filtering on which phase the cell is currently in, all the cells are filtered on the celltype. In total there are 14 different cell types present in the dataset, for each single one of these cell types the correlation will need to be checked. Due to the great amount of cell types present in the data, not all cell types will be displayed in figures. The cell types that are displayed in the histograms are mainly picked on the correlation values across all different VAE models, prioritizing high correlation values. The best correlations with the latent time for each cell type can be found in table 4.

| Model | VAE | $\beta$-VAE | DIP-VAE | $\beta$-TCVAE |
|---|---|---|---|---|
| Parameters | N/A | $\beta = 1000$ | $\lambda$-diag = 1.0<br>$\lambda$-offdiag= 1.5 | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ |
| nIPC | 0.1075 | 0.1006 | -0.0776 | 0.1160 |
| Nbl1 | -0.1852 | -0.1896 | -0.1302 | 0.1537 |
| Nbl2 | 0.1246 | -0.0640 | -0.0752 | 0.1079 |
| Imm-Granule1 | -0.0535 | 0.0601 | -0.0551 | 0.1005 |
| Imm-Granule2 | 0.0768 | -0.0675 | 0.0899 | -0.0740 |
| GlialProg | -0.2567 | -0.2045 | -0.1220 | 0.2089 |
| Granule | -0.1220 | -0.0995 | 0.0618 | -0.0947 |
| CA | -0.0619 | -0.0724 | 0.0441 | -0.0605 |
| CA1-Sub | -0.0753 | -0.0619 | -0.0775 | 0.0717 |
| CA2-3-4 | -0.0749 | 0.0400 | -0.0649 | -0.0784 |
| RadialGlia | -0.0849 | -0.1079 | 0.1004 | -0.0959 |
| RadialGlia2 | -0.1361 | 0.1364 | 0.0883 | -0.1627 |
| OPC | -0.1048 | 0.1513 | 0.0996 | -0.1293 |
| ImmAstro | -0.1079 | -0.1007 | -0.0642 | -0.0746 |

Table 4: The best correlations found for the latent time for every different kind of cell type in a single run by each model

Figures 8-11 show the histograms of correlation values between every latent dimension and the latent time. What can be noted is that some models encoded the best correlation multiple times. This can be seen in figure 11 when looking at GlialProg. This behaviour can also be seen in the other VAE models.

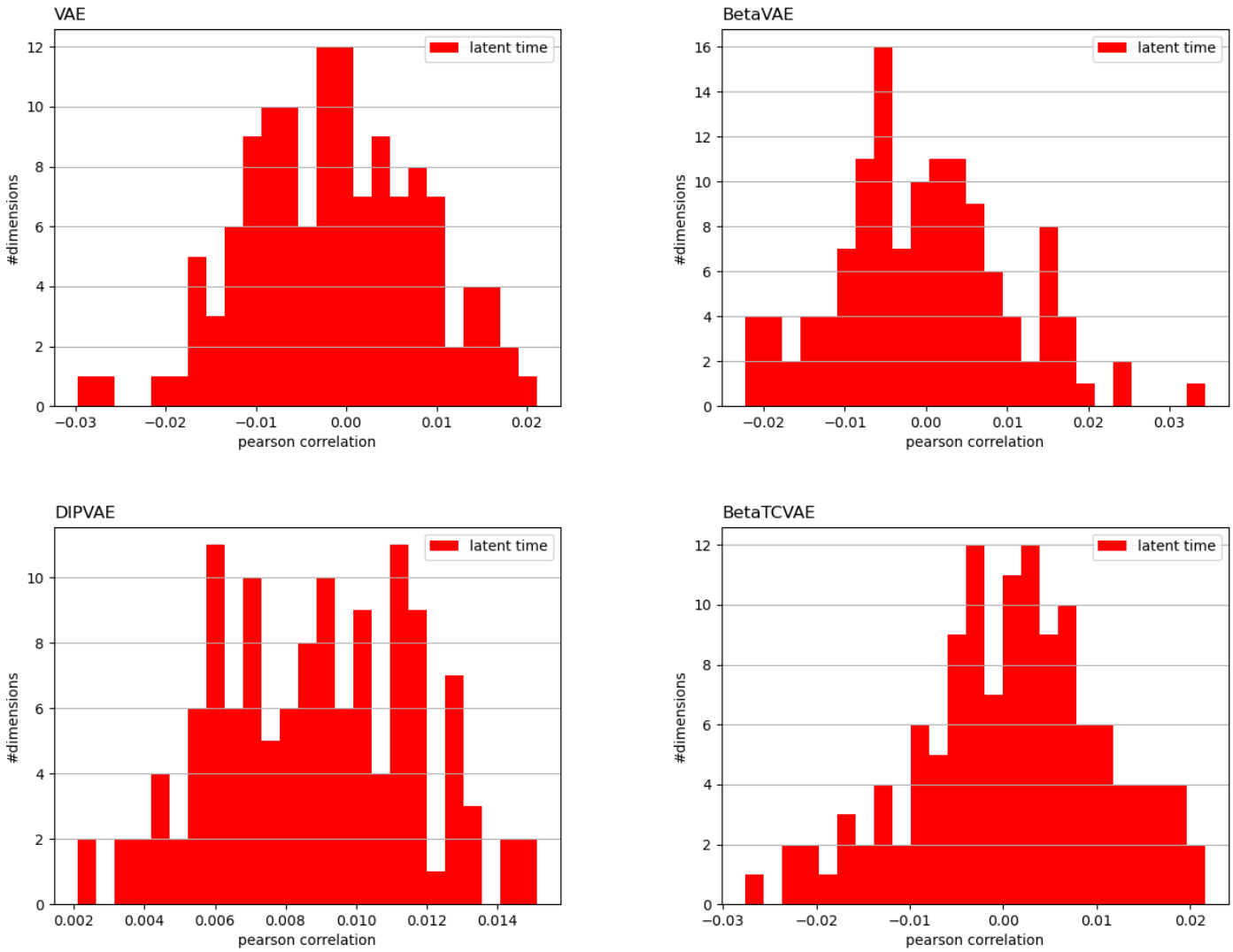


Figure 6: Depicts how many dimensions correlate a certain amount with the latent time when checking over all cells. Each histogram depicts the performance of a different VAE model.

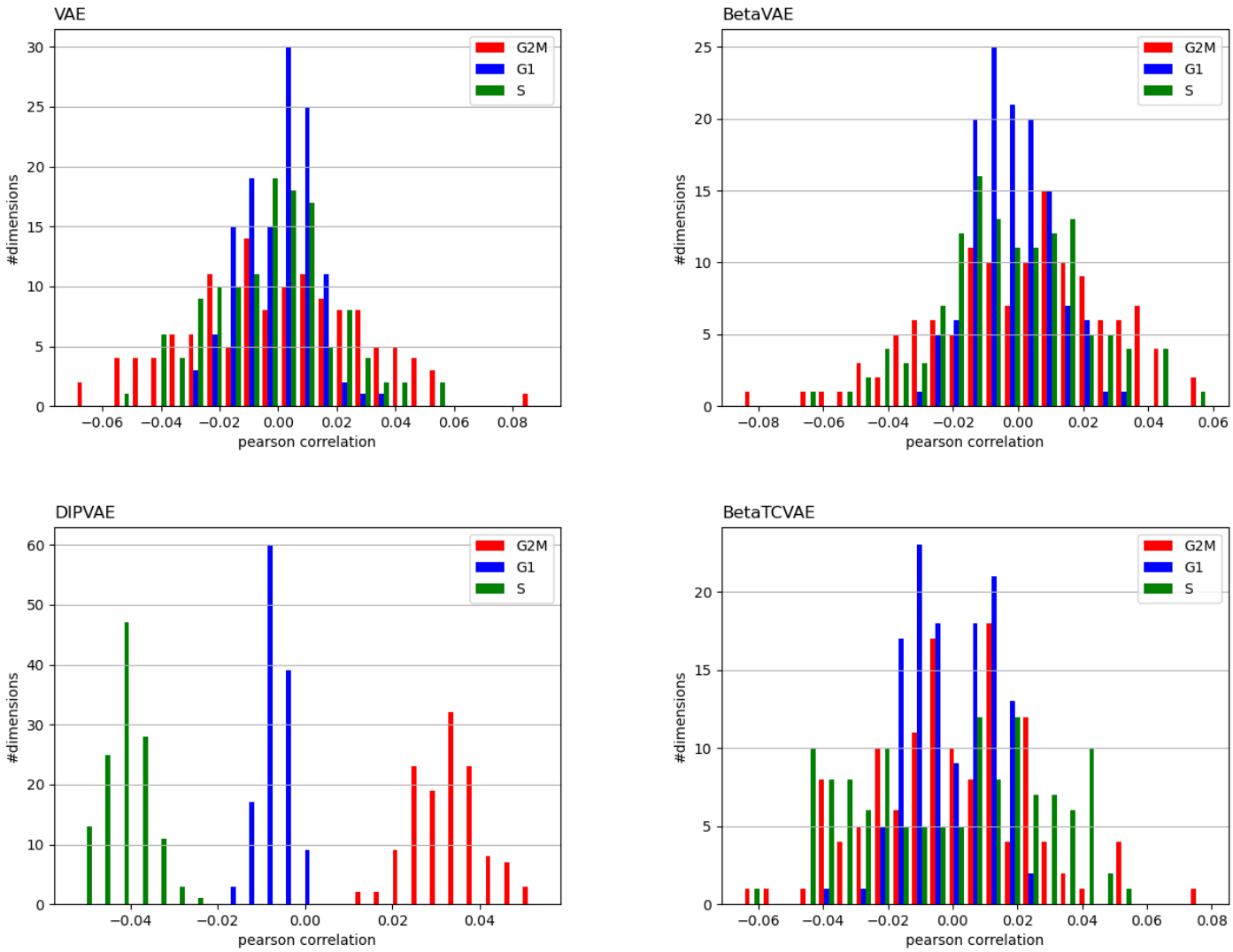


Figure 7: Depicts how many dimensions correlate a certain amount with the latent time per cell cycle phase. Each histogram depicts the performance of a different VAE model.

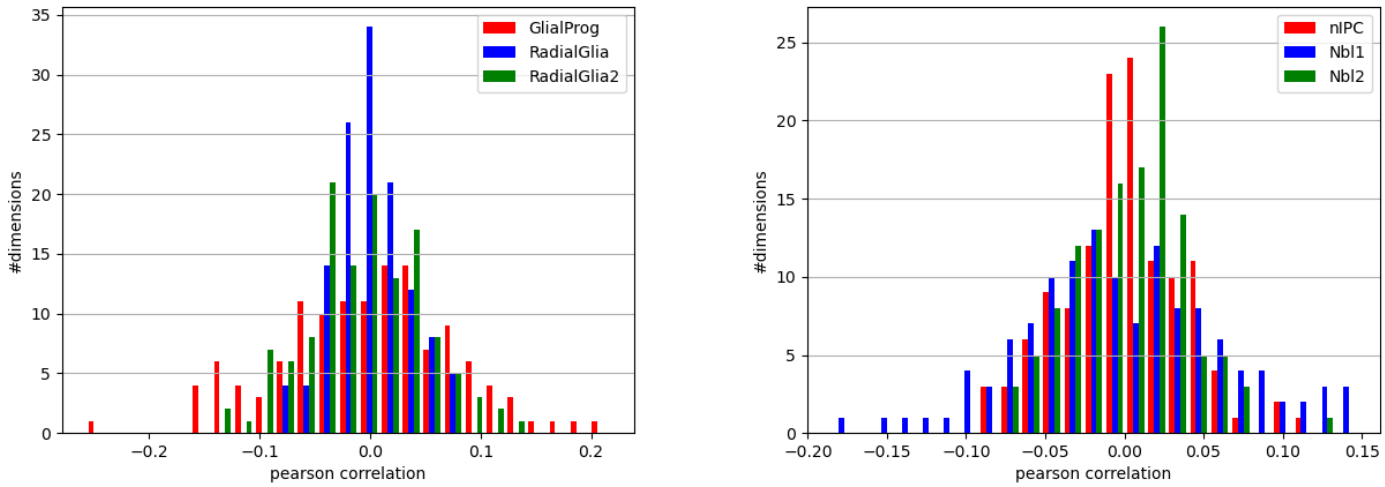


Figure 8: Histograms depicting how many dimensions correlate a certain amount with the latent time per cell type as found by VAE.

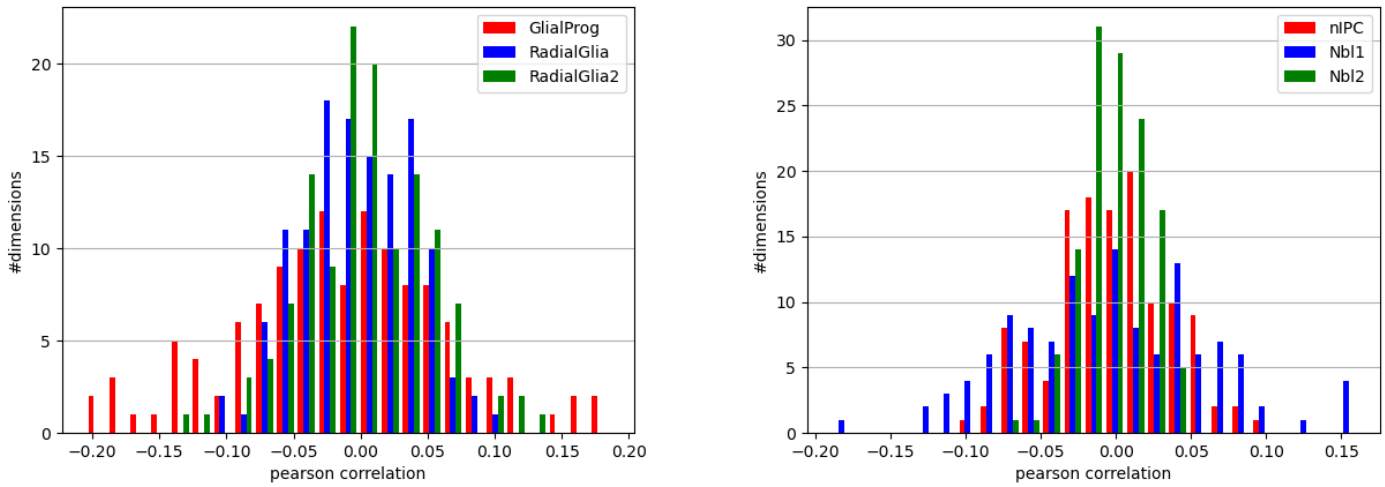


Figure 9: Histograms depicting how many dimensions correlate a certain amount with the latent time per cell type as found by $\beta$-VAE.

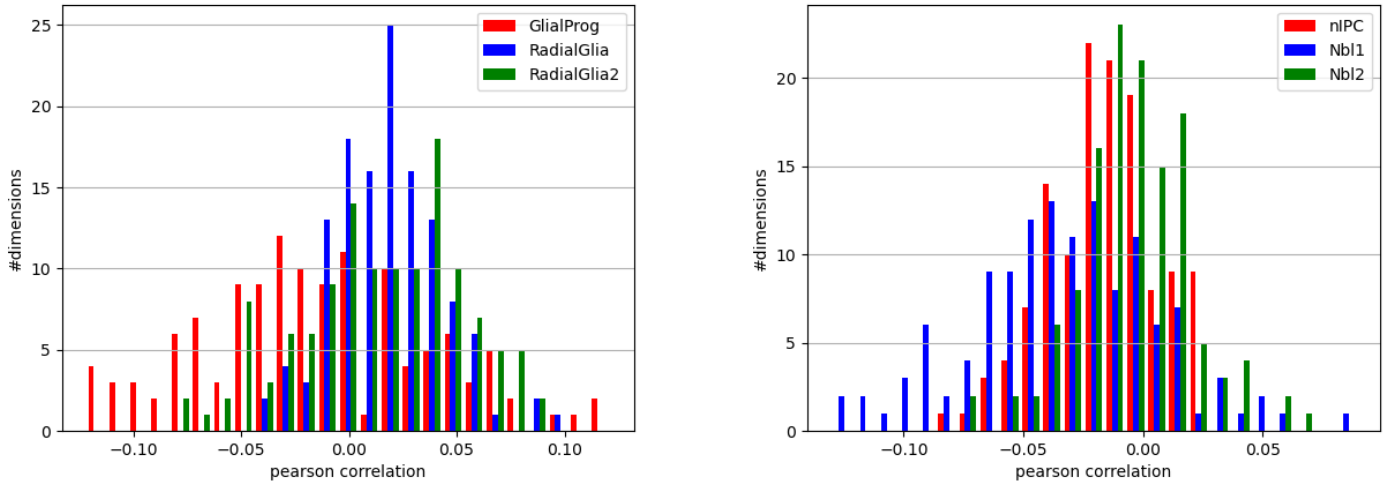


Figure 10: Histograms depicting how many dimensions correlate a certain amount with the latent time per cell type as found by DIP-VAE.

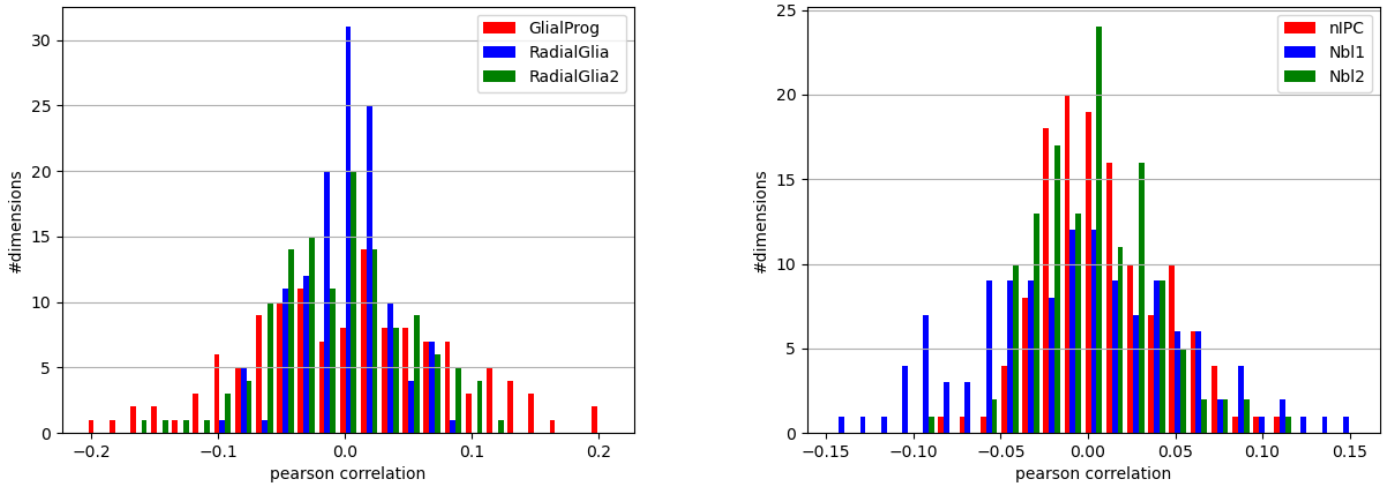


Figure 11: Histograms depicting how many dimensions correlate a certain amount with the latent time per cell type as found by $\beta$-TCVAE.

### 4.3 Regression model

To test whether the VAE models encoded the biological processes more accurately in 2 dimensions, a linear regression model was used. The correlations found by a fitted regression model are compared to the correlations found by each model.

First a look was taken at the cell cycle score, the formula that was used for this was the following:

$$cell\_cycle\_score = v_{01} + v_{02} + v_{01} * v_{02} \tag{6}$$

Table 5 shows the correlations found when just using the VAE models, this is used to compare the correlations found by using a regression model to a baseline. Table 6 shows the correlations found by the linear regression model. It is important to note that the results from table 5 and 6 were obtained in the same run, this was done to keep the comparison as fair as possible.

| Model | Parameters | S-Score | G2M-Score |
|---|---|---|---|
| VAE | N/A | -0.2466 | -0.2663 |
| $\beta$-VAE | $\beta = 1000$ | 0.4365 | 0.5012 |
| dip-vae | $\lambda$-diag = 1.0<br>$\lambda$-offdiag=1.5 | -0.2545 | -0.3686 |
| $\beta$-TCVAE | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ | -0.3756 | -0.9241 |

Table 5: Regular correlations for the cell cycle scores found by each model in a single run

| Model | S-Score | G2M-Score |
|---|---|---|
| VAE | 0.3547 | 0.3907 |
| $\beta$-VAE | 0.5795 | 0.7220 |
| DIP-VAE | 0.5378 | 0.6054 |
| $\beta$-TCVAE | 0.5405 | 0.9290 |

Table 6: The correlations found by using a regression model over two different latent dimensions

When looking at tables 5 and 6 it becomes quite obvious that for the cell cycle scores, all the VAE models find better correlations if a regression model is run afterwards. What is remarkable is the great jump in performance that can be seen when looking at DIP-VAE.

For the latent time, the formula for the regression model was the same as the one that was used for the cell cycle score:

$$latent\_time = v_{01} + v_{02} + v_{01} * v_{02} \tag{7}$$

Table 7 displays the correlations found when only using the VAE models, this was again done to be able to compare the regression model to correlations found by the VAE model. Table 8 shows the correlations found when using a linear regression model over the latent space.

| Model | Parameters | Latent time |
|---|---|---|
| VAE | N/A | 0.0281 |
| $\beta$-VAE | $\beta = 1000$ | 0.0275 |
| dip-vae | $\lambda$-diag = 1.0<br>$\lambda$-offdiag=1.5 | -0.0151 |
| $\beta$-TCVAE | $\alpha = 1.0$<br>$\beta = 1.0$<br>$\gamma = 1.0$ | -0.0318 |

Table 7: Regular correlations for the latent time found by each model in a single run

| Model | Latent time |
|---|---|
| VAE | 0.0411 |
| $\beta$-VAE | 0.0405 |
| dip-vae | -0.0366 |
| $\beta$-TCVAE | -0.0429 |

Table 8: The correlations found by using a regression model over two different latent dimensions

When looking at tables 7 and 8 it becomes clear that the improvement in the correlation found when using a regression model is only minor.

## 5 Responsible Research

Reproducibility is important when discussing the results that have been found. However, reproducing results from a paper is often far from easy [10]. Therefore this section will be dedicated to making the reproduction process as smooth as possible. The code will be open source and available on GitLab. Besides this all the configurations that were used have been mentioned in the paper so that the exact models can be recreated.

On the GitLab page there will be some additional information. Along side this it will also contain all the code that was used to obtain the results that have been used in this study.

## 6 Discussion

### 6.1 Cell cycle scores

Almost all of the VAE models performed as expected at encoding the S- and G2M-score into a single latent dimension. When looking at table 1 it becomes quite clear that there is a performance gap between the different VAE models. This is also clearly visible in figure 1, the $\beta$-VAE and $\beta$-TCVAE models find much higher correlations than VAE and DIP-VAE. What stands out in figure 1 is that both for $\beta$-VAE and $\beta$-TCVAE the highest correlation was found only once, this means that it encoded it in a single dimension.

The VAE model performed the worst on average at encoding the cell cycle scores into the latent dimension. However,

whilst it performed the worst at finding correlations it usually performed the best when looking at reconstruction loss alone.

One thing that can be observed from the results is that nearly all of the VAE models seem to be having a harder time encoding the S-Score than they do for encoding the G2M-Score. This is not something that came as unexpected. A possible explanation for this might be because during the G2M-phase the cell is actively dividing. Which could make it a lot clearer in the data as the genes that are involved in the cell cycle are much more active. The most extreme example of a model finding a better correlation for the G2M score is the $\beta$-TCVAE. This model consistently found very high correlations for the G2M-score. The only exception to this seemed to be the regular $\beta$-VAE which found similar but good correlations for both the S- and G2M-scores.

Surprisingly, DIP-VAE performed quite bad at encoding the scores into the latent dimension. It seemed to have encoded it slightly in every dimension but not enough for it to actually mean anything. This was surprising because other papers found great results when it came to finding disentangled representations using DIP-VAE [8].

## 6.2 Differentiation state

All of the VAE models had a difficult time encoding the latent time into a single latent dimension. Therefore different ways of filtering to create subsets were used to check if the VAE models had encoded the latent time for specific subsets.

When looking at all the cells, none of the VAE models managed to encode it. All of the VAE models found around the same correlation for the latent time. The exception to this was DIP-VAE which seemed to perform worse than the other three models by relatively a big margin, this can be seen in table 2. It was expected that none of the VAE models would be able to encode the latent time when looking at all the cells. This is due to the fact that not all cells in the data are actively differentiating. For example, cells that are in the G2M phase are much more actively differentiating, making them more easy to recognize from the data.

However, when creating subsets based on which phase of the cell cycle each cell is in, the results are slightly better. When looking at table 3, the correlation values are a bit higher than they are in table 2. However, these values are still quite low and seem so low that nothing useful was encoded. A striking aspect of table 3 is that again DIP-VAE seems to be performing the worst by relatively a big margin. The weird behaviour of DIP-VAE becomes even more clear when looking at figure 7. It is the only model that isolates every phase as separate peaks. What also stands out in figure 7 is that the best correlations for VAE, $\beta$-VAE and $\beta$-TCVAE are captured in very little dimensions.

Furthermore, a different kind of subsets has also been used, one based on cell types. This resulted in higher correlations found than the other two methods. Whilst these correlations are higher than the previous two methods, they are still not significant. One trend that seems to continue when looking at table 4 is that DIP-VAE is performing the worst out of the 4 models. The other three models performed similarly as they did for the first two methods as well. Another thing that stands out from table 4 is that some cell types have on average a higher correlation than others, take for example the GlialProg type. On average this cell type has the highest correlation value, this was quite interesting because, these types of cells are mostly dividing. Therefore it was expected for the models to have the most difficulty with encoding this.

What is also noticeable in figure 8-11 is that the highest correlations found by each model are sometimes encoded in multiple dimensions. What also happens is that the model finds other correlations that are close to the best one found. This means that the VAE models did not encode the latent time per cell type into a single latent dimension.

## 6.3 VAE models

Considering the difficulty all the VAE models had with learning the data, a different form of preprocessing of the data might have to be used in the future. One that either reduces the noise in the data or makes it easier for the models to learn. However, eventually the models were able to learn the data after reducing the importance of the KLD term.

Looking at all the results, it's quite clear that there is a gap in the performance of the VAE models when measuring their ability to find disentangled representations. All models performed as expected, except DIP-VAE.

No matter what configuration was used, DIP-VAE continued to perform worse than the other three models. Therefore, DIP-VAE seemed cumbersome to configure when trying to get it to perform well for any of experiments done in this study. The fact that DIP-VAE performed this bad was quite unexpected. This was because there are other papers that found that DIP-VAE was quite good when it came to learning disentangled representations.

Another quite interesting finding was that at least for the cell cycle score both $\beta$-VAE and $\beta$-TCVAE performed by far the best. Other papers have found similar results, in the sense of that these papers also found that these two models are great at finding disentangled representations. Both of these VAE models found high correlations for the cell cycle score and encoded it into a single dimension.

For $\beta$-VAE when using a value for $\beta$ that was too low, it would result in the model performing similar to VAE. When making $\beta$ higher, it would result in better correlations found, but it also had a side effect. The side effect was that more than one latent dimensions captured the thing that was supposed to be captured.

In general VAE did not perform great when it came to finding disentangled representation. However, when solely looking at the reconstruction loss, VAE outperformed all the other models by quite a big margin. This is due to the fact that VAE aims to minimize the information loss.

### 6.4 Regression model

When using a linear regression model for two different latent dimensions at a time, it found better correlations when looking at the cell cycle scores. In fact it found better correlations for every model. This was the most clear when looking at the performance of DIP-VAE. When looking at tables 5 and 6 it is clear that the jump in performance for DIP-VAE is quite huge.

However, quite the opposite can be seen when using a linear regression model for the latent time. The correlations found by the regression model are barely any higher than the correlations found by just the VAE models. So while it might encode the cell cycle score better into two dimensions in some cases, the VAE models do not seem to consistently do this for everything.

## 7 Conclusions and Future Work

This study has shown that the Vanilla VAE model performed the worst at finding the cell cycle score as a disentangled representation. However, the performance of Vanilla VAE is almost on par with DIP-VAE. Both of these models seem to have a difficult time when it comes to encoding complex processes from the data into the latent representation. One of the most significant findings is the models that performed the best when it comes to encoding the cell cycle score into a latent dimension were $\beta$-VAE and $\beta$-TCVAE.

When looking at the latent time, it became quite clear that it is too complicated to be captured in one dimension. However, what is still noticeable is that DIP-VAE performed the worst in all cases. Besides this the other three models performed very similar when it came to encoding the latent time into a single latent dimension.

It would also be interesting to see what the effect would be of data that is preprocessed in a different manner where all the genes might be able to be used. This way it might be able to encode the latent time as well and through this the models could be compared to each other more fairly. It would also be interesting to see this study repeated with the proper configuration for DIP-VAE, if it exists. This would again create a fairer comparison between the models.

## 8 Supplementary

An interesting thing to note is that all of the models had difficulty learning the data when using all of the gene expressions. This resulted in the models usually only being able to find low correlations with the biological processes. Therefore it was decided to only use the most informative genes present in the data.

Furthermore, all of the VAE models seemed to perform a lot worse when using a small latent space. This resulted in the latent space being 128, going below that had a severe effect on the performance of the VAE models. Besides this it also resulted in lower correlations being found for the biological processes and distinct latent dimensions.

## References


[1] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[2] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef, "Deep generative modeling for single-cell transcriptomics," *Nature methods*, vol. 15, no. 12, pp. 1053–1058, 2018.

[3] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan *et al.*, "Rna velocity of single cells," *Nature*, vol. 560, no. 7719, pp. 494–498, 2018.

[4] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.

[5] R. T. Chen, X. Li, R. Grosse, and D. Duvenaud, "Isolating sources of disentanglement in variational autoencoders," *arXiv preprint arXiv:1802.04942*, 2018.

[6] J. Choi, G. Hwang, and M. Kang, "Discond-vae: Disentangling continuous factors from the discrete," *CoRR*, vol. abs/2009.08039, 2020. [Online]. Available: https://arxiv.org/abs/2009.08039

[7] A. Kumar, P. Sattigeri, and A. Balakrishnan, "Variational inference of disentangled latent concepts from unlabeled observations," *arXiv preprint arXiv:1711.00848*, 2017.

[8] A. H. Abdi, P. Abolmaesumi, and S. Fels, "A preliminary study of disentanglement with insights on the inadequacy of metrics," *arXiv preprint arXiv:1911.11791*, 2019.

[9] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.

[10] M. Baker, "Reproducibility crisis," *Nature*, vol. 533, no. 26, pp. 353–66, 2016.