bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DeepProfile: Deep learning of patient molecular profiles for precision medicine in acute myeloid leukemia

Ayse Dincer[1], Safiye Celik[1], Naozumi Hiranuma[1], and Su-In Lee[1,2]

[1] Paul G. Allen School of Computer Science and Engineering,

[2] Department of Genome Sciences, School of Medicine, University of Washington

Email: suinlee@cs.washington.edu

Abstract

Motivation: Learning robust prediction models based on molecular profiles (e.g., expression data) and phenotype data (e.g., drug response) is a crucial step toward the development of precision medicine. Extracting a meaningful low-dimensional feature representation from patient's molecular profile is the key to success in overcoming the high-dimensionality problems. Deep learning-based *unsupervised* feature learning has enormously improved image classification by enabling us to use large amounts of "unlabeled" images informative of the prediction task.

Approach: We present the *DeepProfile* framework that attempts to extract latent variables from publicly available expression data using the variational autoencoders (VAEs) and use these latent variables as features for phenotype prediction. *To our knowledge, DeepProfile is the first attempt to use deep learning to learn a feature representation from a large number of unlabeled (i.e, without phenotype) expression samples that are not incorporated to the prediction problem.* We apply DeepProfile to predicting response to hundreds of cancer drugs based on gene expression data. Most patients with advanced cancer continue to receive drugs that are ineffective. This is exemplified by acute myeloid leukemia (AML), a disease for which treatments and cure rates (in the range of 25%) have remained stagnant. Effectively deploying an ever-expanding array of cancer drugs holds great promise to improve prognoses but requires methods to predict how drugs will affect specific patients.

Result: We train the VAE model that represents a specific mapping from input variables (here, gene expression levels) into a much smaller number of latent variables, on the basis of gene expression data from AML patients available through the Gene Expression Omnibus (GEO). Our results show that the lower dimensional representation (i.e., latent variables) generated by using VAEs significantly outperform the original input feature representation (i.e., gene expression levels) in the drug response prediction problem.

Conclusion: We demonstrate the effectiveness of VAEs in extracting a low-dimensional feature representation from publicly available unlabeled gene expression data. We show that the learned features are relevant to drug response prediction, which indicates that the latent variables capture important processes relevant to the prediction problem.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Introduction

The number of potential cancer drugs are rapidly increasing – more than 1,200 cancer medicines are in clinical development in the U.S. [33]. However, cure rates of acute myeloid leukemia (AML) have remained stagnant (in the range of 25%) [22]. Cancers that are pathologically similar to each other often respond to the same drug regime differently. There is a great need to develop computational methods to match patients to drugs based on their molecular properties and to identify molecular markers for each drug which reflects the molecular basis for drug sensitivity.

Due to the importance of the problem, numerous studies focused on cancer drug response prediction and used various machine learning (ML) algorithms on a diverse range of biological and molecular data such as gene expression, mutations, and copy number aberrations. Many public database provides measurements of drug responses in cancer cell lines. Most prominent of them include Cancer Genome Project (CGP) [9] containing tests on 130 drugs in 639 cell lines and Cancer Cell Line Encyclopedia (CCLE) [4] containing 24 drugs tested in 479 cell lines. Both of these studies used elastic net to discover novel gene-drug associations. Jang et al. also showed that regression methods like elastic net and ridge regression seem to work well on the cancer drug response prediction problem [13]. Several other studies worked on more complex machine learning algorithms to improve the accuracy of the prediction. Methods like support vector machine (SVM), least squares SVM, and random forest were applied by various studies [8], [2], [32]. Ensemble methods and multitask learning were also used. Costello et al. found Bayesian multitask multiple kernel learning (MKL) method to be the best performing method among other machine learning algorithms and gene expression data to be the most useful data for prediction [7]. Yuan et al. used multitask learning across cancer drugs in order to increase both the accuracy and interpretability of the prediction problem [37]. Lee and Celik et al. developed MERGE algorithm that integrates multi-omic prior information to discover robust gene-drug associations [22].

Several studies used deep learning for similar purposes. Menden et al. used neural networks for the cancer drug sensitivity prediction [24]. Rampasek et al. built variational autoencoder (VAE) models [20] to improve drug response prediction accuracy using pre- and post- treatment cell lines [28]. Way and Greene have used VAEs to learn biologically relevant latent space from The Cancer Genome Atlas (TCGA) pan-cancer data [35]. Our approach, namely *DeepProfile*, is different from the past studies in that, to our knowledge, DeepProfile is the first attempt to use deep learning to learn a feature representation from a large number of unlabeled (i.e, without phenotype) expression samples that are *not* incorporated to the prediction problem and use the feature representation to solve prediction tasks. We showed that DeepProfile results in significant improvement in the prediction performance on AML drug sensitivity prediction problem, which is better than other dimensionality reduction methods.

DeepProfile has three unique aspects compared to previous studies on drug sensitivity prediction or dimensionality reduction: (1) DeepProfile extracts a lower dimensional feature representation of a patient's gene expression data by transferring information from many other patients with the same cancer type captured by the VAE model. (2) DeepProfile uses deep learning in order to learn non-linear mappings between genes and latent variables which might reveal deeper structures within the data and potentially capture complex, nonlinear relationships between gene expression and their complex traits (drug sensitivity). (3) DeepProfile shows significantly better prediction performance compared to other dimensionality reduction methods.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2 Methods

2.1 Datasets

We trained our VAE model using publicly available gene expression data from different Affymetrix microarray platforms which we downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database. These data consist of 4,367 leukemia patient samples, which include 2,813 with AML and others with ALL (acute lymphoblastic leukemia), CML (chronic myelogenous leukemia), CLL (chronic lymphocytic leukemia), BPDCN (blastic plasmacytoid dendritic cell neoplasm), or MDS (myelodysplastic syndrome). The details of the datasets collected from GEO are provided in Table 1.

The data we used to test the learned VAE model had been collected by the University of Washington Medical Center (UWMC) and consists of genome-wide gene expression data from 30 AML patient samples and *in vitro* drug sensitivity of these patients to 160 chemotherapy drugs, as introduced by Lee and Celik et al. [22].

We chose to use publicly available data from GEO to train our VAE model because this enables us to utilize a large number of training samples to learn low-dimensional embedding of high-dimensional gene expression data. We also believe that the VAE model learned by a large set of publicly available samples is more generalizable to broader leukemia (or AML) populations.

In order to integrate data from various platforms, we used Bioconductor annotation databases to convert the probe IDs specific to the array platforms to the human gene IDs. There are 4,051 genes that are present in all datasets listed in Table 1. We also *standardized* (i.e., made zero-mean and unit variance) each gene in each dataset before combining the datasets for learning the VAE model. This is done to ensure that different features (here, gene expression levels) are on the same scale. We finally applied batch effect correction on the data [16] to minimize the effect of potential confounders resulting from experimental variations.

Table 1: Details of the gene expression datasets used for VAE learning

Data Source	GEO ID	Sequencing Platform	Number of Samples
MD Anderson Cancer Center [34, 17]	GSE6891	HG-U133 plus 2.0	537 AML
University of Munich [25]	GSE12417	HG-U133A	163 AML
University of Munich [25]	GSE12417	HG-U133B	163 AML
University of Munich [25]	GSE12417	HG-U133A	79 AML
Roche Molecular Systems [21, 11]	GSE13159	HG-U133 plus 2.0	1480 leukemia
St Jude Children's Research Hospital [27]	GSE14471	HG-U133A	111 AML
Brigham and Women's Hospital [10]	GSE16015	HG-U133 plus 2.0	107 AML
Erasmus MC [31]	GSE17061	HG-U133 plus 2.0	35 AML
Erasmus MC - Sophia Children's Hospital [3, 30]	GSE17855	HG-U133 plus 2.0	237 AML
Brigham and Women's Hospital [26]	GSE21261	HG-U133 plus 2.0	96 AML
Centre for Regenerative Medicine [29]	GSE30903	HG-U133 plus 2.0	24 AML
Centro Investigacion del Cancer [5]	GSE33075	HG-U133 plus 2.0	27 CML
European Institute of Oncology [1]	GSE34860	HG-U133A	78 AML
Charite Berlin	GSE37307	HG-U133A	30 AML
University Hospital Grosshadern, Ludwig-Maximilians-University [23, 12, 14, 15]	GSE37642	HG-U133A	422 AML
University Hospital Grosshadern, Ludwig-Maximilians-University [23, 12, 14, 15]	GSE37642	HG-U133B	422 AML
University Hospital Grosshadern, Ludwig-Maximilians-University [23, 12, 14, 15]	GSE37642	HG-U133 plus 2.0	140 AML
St. Jude Children's Research Hospital [36]	GSE43176	HG-U133A	104 AML
Université Lille II [6]	GSE89565	HG-U133 plus 2.0	12 BPDCN
Université Lille II [6]	GSE89565	HG-U133 plus 2.0	35 T-ALL
Université Lille II [6]	GSE89565	HG-U133 plus 2.0	65 AML

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.2 The DeepProfile framework

We adopt a deep learning approach to learn a low-dimensional feature representation (or 'embedding') for the gene expression data. A variational autoencoder (VAE) is an extension of a classical autoencoder and uses variational inference to infer the posterior of latent embeddings given input data (i.e., $p(z|x)$, where z refers to latent embeddings and x refers to input variables). Like a classical autoencoder, the VAE learns latent embeddings with the objective of minimizing reconstruction error. However, unlike a classical autoencoder, the VAE assumes that the posterior is Gaussian distributed with a standard Gaussian prior (i.e., $N(0, 1)$). This formulation enables us to learn network parameters using scalable optimization methods (such as adaptive moment estimation (Adam)) and reparameterization tricks. The learned decoder (i.e., $p(x|z)$) can then be used as a generative model to generate new samples from underlying latent embedding space. The standard normal prior forces the encoding and decoding networks to produce a generalizable, smooth latent space by learning meaningful features and embedding similar samples close together. We use a VAE model to learn meaningful latent features from the gene expression data of leukemia patients collected from publicly available datasets and use the learned latent features to predict the drug response of AML patients to various anti-cancer drugs. The DeepProfile framework is visualized in Figure 1.

Our VAE model consists of encoder and decoder networks both with 4 dense layers. The encoder network for means and standard deviations share the first three dense layers which have 1,024, 256, and 64 hidden units, respectively. All layers use batch normalization and rectified linear unit (ReLU) activation. The fourth dense layers have 8 hidden units (latent variable count) and are separately trained for means and standard deviations. Similarly, the decoder has 3 dense layers with 64, 256, and 1,024 hidden units with ReLU activation. The final layer has 4,051 hidden units (original data dimension) with identity activation. We use reconstruction error (i.e., mean squared error) and Kullback-Leibler (KL) divergence of the posterior and prior as an objective function. The network is trained by Adam method with a learning rate of 0.0005 [19]. Furthermore, we applied the warm-up process to gradually introduce a KL divergence term in the objective [18], starting with a scaling factor of 0 (corresponding to standard autoencoder) and slowly reaching to 1 (corresponding to standard VAE). The model is built using Keras.

2.3 Training and testing of the DeepProfile framework

After learning the VAE model, we used the inferred weights to encode an 8-dimensional feature vector for each of the 30 AML patients from which we have the drug response data. We then used the encoded low-dimensional representation (LDR) in an L1-regularized linear regression (for drug response prediction) or L1-regularized logistic regression (for complete remission class prediction) setting and measured the prediction performance. We carried out the drug response prediction task separately for each drug. We used leave-one-out cross-validation (LOOCV) to compute prediction error and used 5-fold cross-validation (CV) on the training samples to select the regularization parameter λ.

Since the VAE model is non-convex, the learned LDR is not unique. To ensure that our results takes into account the potential variation in the prediction performance due to the variability of the learned LDR, we trained the VAE model 10 times and repeated the prediction tasks explained above for each of the 10 different learned 8-dimensional LDRs. We included the error bars that represent one standard deviation across 10 VAE runs when we presented our results (Figures 2 and 3)

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

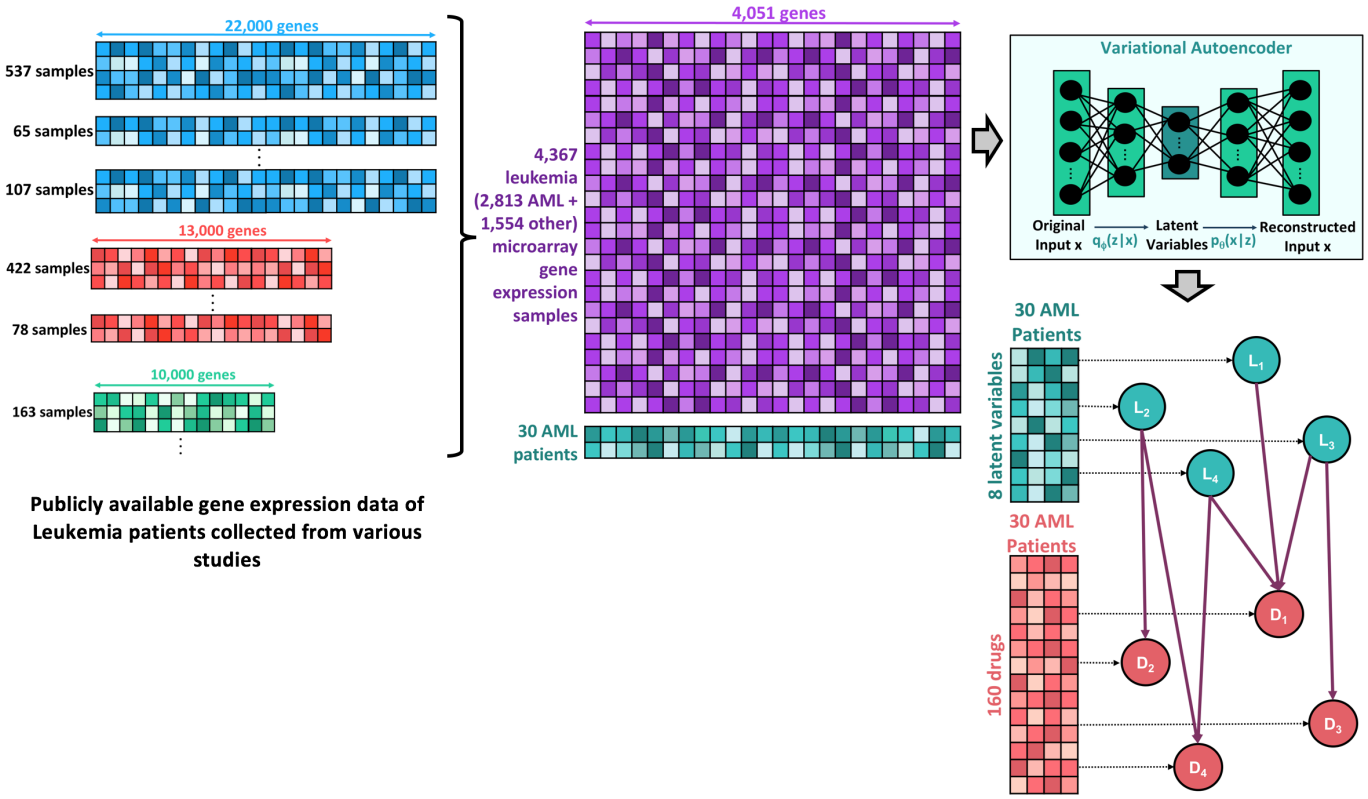

Figure 1: The DeepProfile framework. Microarray datasets gathered from various publicly available studies are combined after standardizing each dataset. The data is batch effect corrected and 4,051 genes present in all datasets are used in the further analysis. The combined data matrix contains 4,367 leukemia gene expression samples among which 2,813 are from AML. The VAE network is trained from these 4,367 samples, and using the trained network, an 8-dimensional latent variable vector is learned for each of 30 AML patients for which we have the response to 160 drugs. Afterwards, a Lasso regression that takes the learned VAE latent representation as the input is used to predict the response of the 30 AML patients to each of the 160 drugs.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3 Results

We compared the learned VAE embeddings to the 16,864 gene expression levels measured in 30 AML patients (Figure 2), as well as to LDR inferred by other dimensionality reduction methods including k-means clustering and Principal Component Analysis (PCA) (Figure 3a). We evaluated our methods by predicting (i) drug response and (ii) complete remission.

3.1 Drug response prediction results

We used the same Lasso regression tests (Section 2.3) for each method in comparison, and measured LOOCV mean-squared error (MSE) for each of the 160 anti-cancer drugs. We trained our VAE model in two different settings using gene expression data from a different set of samples; (I) 4,367 samples from different leukemia types besides AML, and (II) 2,813 AML samples. We call the VAE models in Setting I and Setting II "VAE leukemia" and "VAE AML", respectively. We used those different settings in order to examine how the diversity in the VAE training data affects the AML drug response prediction performance of the learned VAE latent representation. Each of the two settings makes use of 4,051 genes that are overlapping in all leukemia datasets (Table 1).

Figure 2a compares the average MSE over all drugs when we use the expression levels from 16,864 genes, VAE-Leukemia LDR, and VAE-AML LDR. We observed that both VAE-Leukemia LDR and VAE-AML LDR outperformed the gene expression levels in predicting drug response. The VAE-AML LDR led to a lower MSE than the VAE-Leukemia LDR, and reduced the MSE by 9.9% compared to the gene expression levels. We believe that the lower error we obtained from VAE-AML LDR compared to VAE-Leukemia LDR is because VAE can learn more AML-specific features in VAE-AML LDR that can be more useful for AML drug response prediction problem. Thus, even though eliminating other leukemia patients reduces the number of samples that VAE-AML LDR can use, the error is still reduced compared to VAE-Leukemia LDR.

Figure 2b shows the average MSE values for 44 drugs whose response is predicted well (i.e. MSE $\leq$ 0.7 achieved by at least one of the gene expression levels, VAE-Leukemia LDR, or VAE-AML LDR). For well-predicted drugs, both VAE-Leukemia LDR and VAE-AML LDR led to an average MSE lower than the one from the gene expression levels, and VAE-AML LDR reduced the average MSE by 15.2% compared to the gene expression levels.

Figure 2c compares the MSE values obtained by the gene expression levels and VAE-AML LDR for each of the 160 cancer drugs. For 68.1% of the drugs (109 out of 160 drugs), VAE-AML LDR outperforms the gene expression levels. When the MSE values are compared for only 44 well-predicted drugs (i.e. MSE $\leq$ 0.7 achieved by at least one of gene expression and VAE-AML), the VAE-AML LDR obtains a lower error than the gene expression for 65.9% of the drugs (29 out of 44 drugs). These results demonstrate that the DeepProfile model is successful at drug response prediction and especially VAE-AML LDR can reduce the prediction error significantly compared to the gene expression levels.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

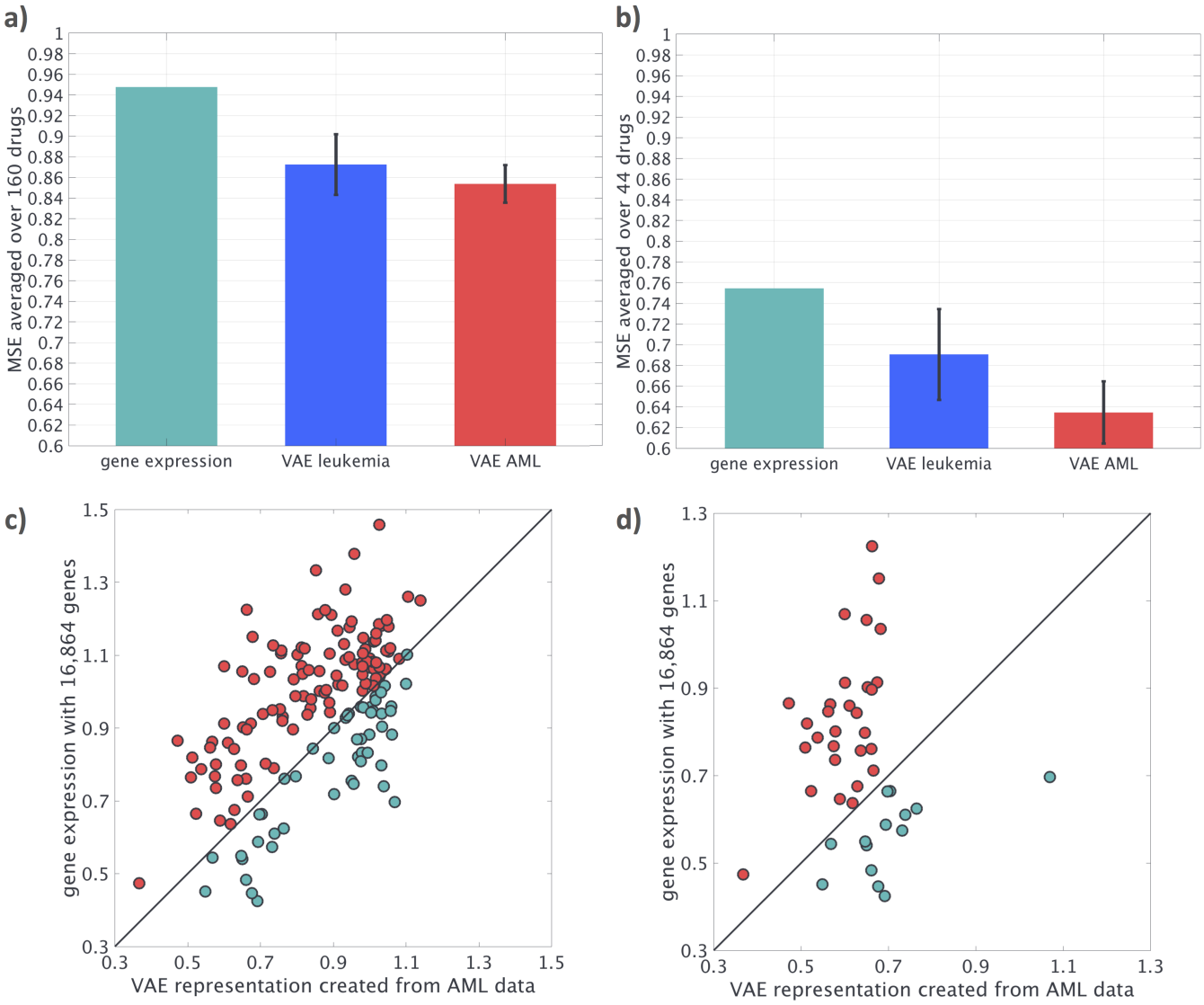

Figure 2: (a) Lasso regression MSE values averaged over all 160 anticancer drugs obtained by three methods; 16,864 gene expression levels of 30 AML patients, the VAE-Leukemia LDR, and VAE-AML LDR. The error bars represent one standard deviation of error values across 10 different runs of VAE. (b)Lasso regression MSE values obtained by the same three methods averaged over 44 drugs after the drugs for which all three methods have an MSE > 0.7 are excluded. (c) Scatter plot comparing the MSE values obtained by the gene expression levels and VAE-AML. d) Scatter plot comparing the MSE values obtained by the gene expression and VAE-AML when we exclude the dots for the drugs for which both gene expression and VAE-AML have an MSE > 0.7. Each dot in (c) and (d) represents a drug and the dots above the diagonal line correspond to the drugs for which VAE-AML outperforms the gene expression levels.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.2 Additional drug response prediction results

We further investigated the drug response prediction performance of our DeepProfile framework by comparing to two other dimensionality reduction algorithms — k-means clustering and PCA. For k-means clustering, we learned 8 gene clusters and used the cluster centroids as LDR, while for PCA, we used top 8 principal components as LDR. We also analyzed in this section how the results are affected from the depth of the VAE model and the training data size.

Figure 3a compares the performance of the VAE-AML LDR with PCA and k-means clustering. VAE-AML LDR can outperform both PCA and k-means algorithms for the same training data and the same size of latent dimensions. This is potentially because non-linear dimensionality reduction of VAE produces more informative LDR relative to the linear methods.

Figure 3b illustrates the effect of using deeper VAE-AML models for the drug response prediction problem. Adding more layers to VAE models led to a higher performance, which is not surprising because deeper networks are able to discover complex non-linear associations among genes. Yet, when the networks are too deep, the learned VAE-AML LDR performs worse due to insufficient sample size.

Figure 3c demonstrates that the performance of VAE LDR increases with larger sample size, as expected. This indicates that our framework can further reduce the error with more samples.

3.3 Complete remission prediction results

In order to demonstrate that the LDR learned by VAE can effectively predict other phenotypes, we trained L1 regularized logistic regression on the learned VAE LDR to predict the complete remission phenotype of 30 AML patients. Complete remission for a cancer patient means that all signs of the cancer are removed by the therapy. We note that the patients are treated using a few common AML drugs in clinic while the drug response data that we use for the prediction problem we tackled in Section 3.1 and 3.2 are from *in vitro* testing of the tumor samples taken from the patients for 160 chemotherapy drugs.

Figure 4 compares VAE LDR to the gene expression levels and two other dimensionality reduction algorithms — PCA and k-means clustering — for predicting CR. The larger area under the ROC curve for VAE LDR shows that it outperforms the two other LDRs and the gene expression levels for CR prediction. This result demonstrates that the LDR learned by VAE can generalize to other prediction tasks.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

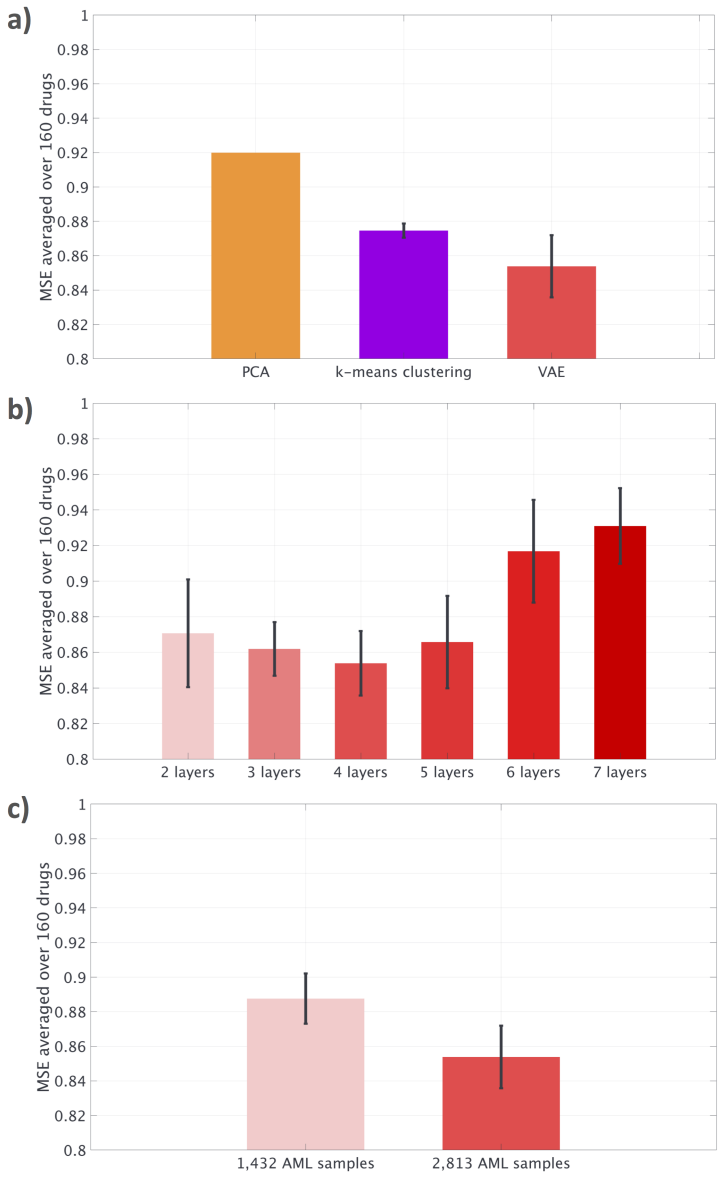

Figure 3: (a) Comparison of VAE with other dimensionality reduction algorithms — PCA and k-means clustering. MSE values are averaged over all 160 anti-cancer drugs obtained by the LDR generated by each method. All three methods use the same data with 2,843 AML samples and all three methods reduce the dimensionality to 8. The error bars represent one standard deviation of error values obtained by 10 different runs of VAE or k-means clustering. (b) The effect of depth of the VAE network on the drug response prediction performance. The bars represent the average MSE values obtained by a VAE with two, three, four, five, six, and seven layers respectively in the encoder and decoder. (c) The effect of the sample size on the drug response prediction performance. The plot compares the Lasso regression MSE obtained by the VAE trained from all available AML samples (2,813 samples) and VAE trained from half as many (1,432) AML samples.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

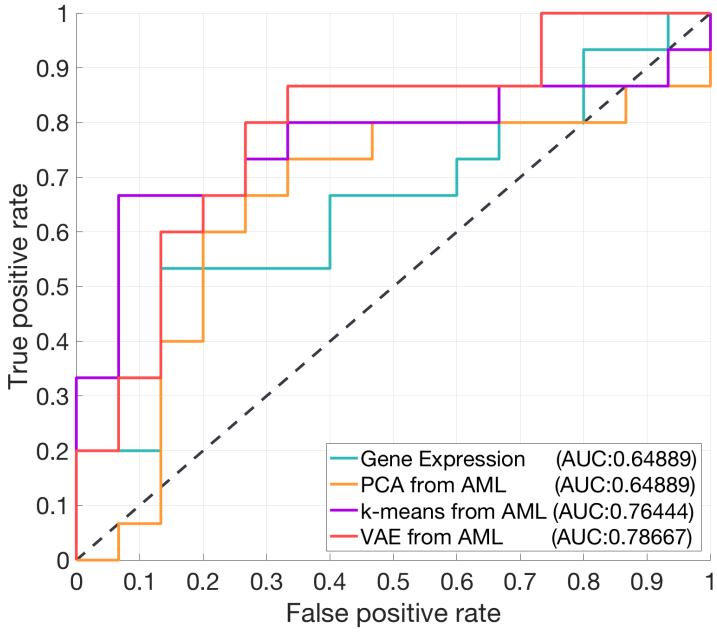

Figure 4: ROC curves for comparing the results of complete remission prediction accuracy obtained using L1-regularized logistic regression trained using 4 different inputs: the gene expression levels, top 8 principal components, 8 cluster centroids learned by k-means, and 8 dimensional VAE-AML LDR.

4 Conclusion

In this paper, we present the DeepProfile framework that adopts the variational autoencoder (VAE) to learn low dimensional representation (LDR) from publicly available unlabeled (i.e., without phenotype data) gene expression datasets and uses the extracted LDR to predict sensitivity to anti-cancer drugs and complete remission for AML patients. We observed that the LDR generated by VAE better predicted the drug response and complete remission than the original gene expression data and two other commonly used LDR learning methods. When we used samples from only AML patients, DeepProfile reduced the average error obtained by the gene expression by 9.9% for all 160 drugs and by 15.2% for the 44 best-performing drugs. Despite that the samples used in VAE training are obtained from many different studies carried in different countries and labs with different sequencing technologies, VAE is quite successful at disentangling the discrepancies in the data and creating an LDR that can be used for different cancer phenotype prediction purposes.

It is interesting to note that the performance of VAE does not only depend on the sample size but is also highly affected by the nature of data. We observed that when we added samples from patients with other types of leukemia, the prediction performance deteriorated. We hypothesized that, since different cancer subtypes have different characteristic and each cancer subtype shows specific molecular properties, adding more data from different leukemia types may not help extracting features important for AML.

Our future directions include: (1) improving our learning algorithm using semi-supervised VAE which benefits from labels of data while training the network, (2) incorporating RNA-seq data along with microarray data to increase the sample size for training VAE to allow it to discover further hidden characteristics from the data, and (3) extending the framework to different cancer types and

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

building a generic tool that is useful for extracting latent features specific to different cancer types.

References

[1] Myriam Alcalay et al. "Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance". In: *Blood* 106.3 (2005), pp. 899–902.

[2] Samirkumar B Amin et al. "Gene Expression Profile Alone Is Inadequate In Predicting Complete Response In Multiple Myeloma". In: *Leukemia* 28.11 (2014), pp. 2229–34.

[3] Brian V Balgobind et al. "Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia". In: *Haematologica* 96.2 (2011), pp. 221–230.

[4] Jordi Barretina et al. "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity". In: *Nature* 483 (2012), pp. 603–607.

[5] Rocio Benito et al. "Imatinib therapy of chronic myeloid leukemia restores the expression levels of key genes for DNA damage and cell-cycle progression". In: *Pharmacogenetics and Genomics* 22.5 (2012), pp. 381–8.

[6] Adam Ceroi et al. "LXR agonist treatment of blastic plasmacytoid dendritic cell neoplasm restores cholesterol efflux and triggers apoptosis". In: *Blood* 128.23 (2016), pp. 2694–2707.

[7] James C Costello et al. "A community effort to assess and improve drug sensitivity prediction algorithms". In: *Nature Biotechnology* 32 (2014), pp. 1202–1212.

[8] Anneleen Daemen et al. "Modeling precision treatment of breast cancer". In: *Genome Biology* 14.10 (2013), R110.

[9] Mathew J Garnett et al. "Systematic identification of genomic markers of drug sensitivity in cancer cells". In: *Nature* 483 (2012), pp. 570–575.

[10] Claudia Haferlach et al. "AML with mutated NPM1 carrying a normal or aberrant karyotype show overlapping biologic, pathologic, immunophenotypic, and prognostic features". In: *Blood* 114.14 (2009), pp. 3024–3032.

[11] Torsten Haferlach et al. "Clinical Utility of Microarray-Based Gene Expression Profiling in the Diagnosis and Subclassification of Leukemia: Report From the International Microarray Innovations in Leukemia Study Group". In: *Journal of Clinical Oncology* 28.15 (2010), pp. 2529–2537.

[12] Tobias Herold et al. "Isolated trisomy 13 defines a homogeneous AML subgroup with high frequency of mutations in spliceosome genes and poor prognosis". In: *Blood* 124.8 (2014), pp. 1304–1311.

[13] In S Jang et al. "Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data". In: *Pacific Symposium on Biocomputing* (2014), pp. 63–74.

[14] Hanna Janke et al. "Activating FLT3 Mutants Show Distinct Gain-of-Function Phenotypes In Vitro and a Characteristic Signaling Pathway Profile Associated with Prognosis in Acute Myeloid Leukemia". In: *PLOS ONE* 9.3 (Mar. 2014), pp. 1–14.

[15] Xi Jiang et al. "Eradication of Acute Myeloid Leukemia with FLT3 Ligand–Targeted miR-150 Nanoparticles". In: *Cancer Research* 76.15 (2016), pp. 4470–4480.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[16] W. Evan Johnson, Cheng Li, and Ariel Rabinovic. "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics* 8.1 (2007), pp. 118–127.

[17] Hendrik J. M. de Jonge et al. "High VEGFC expression is associated with unique gene expression profiles and predicts adverse prognosis in pediatric and adult acute myeloid leukemia". In: *Blood* 116.10 (2010), pp. 1747–1754.

[18] C. Kaae Sønderby et al. "Ladder Variational Autoencoders". In: *ArXiv e-prints* (Feb. 2016). arXiv: `1602.02282 [stat.ML]`.

[19] Diederik P Kingma and Jimmy Ba. "Adam: a Method for Stochastic Optimization". In: *International Conference on Learning Representations* (2015).

[20] Diederik P Kingma and Max Welling. "Auto-Encoding Variational Bayes". In: *ArXiv e-prints* (Dec. 2013). arXiv: `1312.6114 [stat.ML]`.

[21] Alexander Kohlmann et al. "An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in Leukemia study prephase". In: *British Journal of Haematology* 142.5 (2008), pp. 802–807.

[22] Su-In Lee et al. "A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia". In: *Nature Communications* 9.1 (2014), pp. 42–54.

[23] Zejuan Li et al. "Identification of a 24-Gene Prognostic Signature That Improves the European LeukemiaNet Risk Classification of Acute Myeloid Leukemia: An International Collaborative Study". In: *Journal of Clinical Oncology* 31.9 (2013), pp. 1172–1181.

[24] Michael P Menden et al. "Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties". In: *PLOS ONE* 8.4 (Apr. 2013), pp. 1–7.

[25] Klaus H. Metzeler et al. "An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia". In: *Blood* 112.10 (2008), pp. 4193–4201.

[26] Miriam Miesner et al. "Multilineage dysplasia (MLD) in acute myeloid leukemia (AML) correlates with MDS-related cytogenetic abnormalities and a prior history of MDS or MDS/MPN but has no independent prognostic relevance: a comparison of 408 cases classified as "AML not otherwise specified" (AML-NOS) or "AML with myelodysplasia-related changes" (AML-MRC)". In: *Blood* 116.15 (2010), pp. 2742–2751.

[27] Ina Radtke et al. "Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia". In: *Proceedings of the National Academy of Sciences* 106.31 (2009), pp. 12944–12949.

[28] Ladislav Rampasek et al. "Dr.VAE: Drug Response Variational Autoencoder". In: *ArXiv e-prints* (June 2017). arXiv: `1706.08203 [stat.ML]`.

[29] Valentina Salvestrini et al. "Purinergic signaling inhibits human acute myeloblastic leukemia cell proliferation, migration, and engraftment in immunodeficient mice". In: *Blood* 119.1 (2012), pp. 217–226.

[30] Julie Damgaard Sandahl et al. "t(6;9)(p22;q34)/DEK-NUP214-rearranged pediatric myeloid leukemia: an international study of 62 patients". In: *Haematologica* 99.5 (2014), pp. 865–872.

[31] Fernando P G Silva et al. "Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status". In: *Blood* 114.14 (2009), pp. 3001–3007.

[32] Lindsay C Stetson et al. "Computational identification of multi-omic correlates of anticancer therapeutic response". In: *BMC Genomics* 15.7 (2014), S2.

bioRxiv preprint doi: https://doi.org/10.1101/278739; this version posted March 8, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[33] "Summer 2016 chart pack of the Pharmaceutical Research and Manufacturers of America". In: *PhRMA* (2016).

[34] Roel G.W. Verhaak et al. "Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling". In: *Haematologica* 94.1 (2009), pp. 131–134.

[35] Gregory P Way and Casey S Greene. "Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders". In: *Pac Symp Biocomput.* 23 (2018), pp. 80–91.

[36] Jin Xu et al. "Dominant Role of Oncogene Dosage and Absence of Tumor Suppressor Activity in Nras-Driven Hematopoietic Transformation". In: *Cancer Discovery* 3.9 (2013), pp. 993–1001.

[37] Han Yuan et al. "Multitask learning improves prediction of cancer drug sensitivity". In: *Scientific reports* 6 (2016).