Identification of repurposed drugs targeting significant long non-coding RNAs in the cross-talk between diabetes mellitus and Alzheimer’s disease

Shokoofeh Ghiam¹, Changiz Eslahchi²³, Koorosh Shahpasand⁴, Mehran Habibi-Rezaei⁵ & Sajjad Gharaghani¹

The relationship between diabetes mellitus (DM) and Alzheimer’s disease (AD) is so strong that scientists called it “brain diabetes”. According to several studies, the critical factor in this relationship is brain insulin resistance. Due to the rapid global spread of both diseases, overcoming this cross-talk has a significant impact on societies. Long non-coding RNAs (lncRNAs), on the other hand, have a substantial impact on complex diseases due to their ability to influence gene expression via a variety of mechanisms. Consequently, the regulation of lncRNA expression in chronic diseases permits the development of innovative therapeutic techniques. However, developing a new drug requires considerable time and money. Recently repurposing existing drugs has gained popularity due to the use of low-risk compounds, which may result in cost and time savings. In this study, we identified drug repurposing candidates capable of controlling the expression of common lncRNAs in the cross-talk between DM and AD. We also utilized drugs that interfered with this cross-talk. To do this, high degree common lncRNAs were extracted from microRNA-lncRNA bipartite network. The drugs that interact with the specified lncRNAs were then collected from multiple data sources. These drugs, referred to as set D, were classified into positive (D⁺) and negative (D⁻) groups based on their effects on the expression of the interacting lncRNAs. A feature selection algorithm was used to select six important features for D. Using a random forest classifier, these features were capable of classifying D⁺ and D⁻ with an accuracy of 82.5%. Finally, the same six features were extracted for the most recently FDA-approved drugs in order to identify those with the highest likelihood of belonging to D⁺ or D⁻. The most significant FDA-approved positive drugs, chromium nicotinate and tapentadol, were presented as repurposing candidates, while cefepime and dihydro-alpha-ergocryptine were recommended as significant adverse drugs. Moreover, two natural compounds, curcumin and quercetin, were recommended to prevent this cross-talk. According to the previous studies, less attention has been paid to the role of lncRNAs in this cross-talk. Our research not only did identify important lncRNAs, but it also suggested potential repurposed drugs to control them.

Abbreviations
IR Insulin resistance
DM Diabetes mellitus
AD Alzheimer's disease
T3D Type 3 diabetes
APOE Apolipoprotein E

¹Laboratory of Bioinformatics and Drug Design, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. ²Department of Computer and Data Sciences, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran. ³School of Biological Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran. ⁴Department of Brain and Cognitive Sciences, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran. ⁵School of Biology, College of Science, University of Tehran, Tehran, Iran. *email: ch.eslahchi@sbu.ac.ir; s.gharaghani@ut.ac.ir
IDE  Insulin degradation enzyme
GLUT4  Glucose transporter type 4
AChE  Acetylcholinesterase
T2DM  Type 2 diabetes mellitus
NcRNA  Non-coding RNA
MiRNA  MicroRNA
LncRNA  Long non-coding RNAs
CADD  Computer aided drug design
FDA  Food and Drug Administration
OOB  Out of bagging error
SVM  Support vector machine
CV  Cross validation
Aβ  Amyloid beta

Recent studies\textsuperscript{1-2} have linked insulin resistance (IR) and impaired insulin signaling to the cross-talk between diabetes mellitus (DM) and Alzheimer’s disease (AD). Many scientists believe that Alzheimer’s disease is both a neurological and neuroendocrine disorder. Therefore, Type 3 Diabetes (T3D) is the most appropriate term to describe this complicated disease\textsuperscript{3}. Other cellular and molecular factors involved in the cross-talk between DM and AD are oxidative stress, obesity, inflammation, dysregulation of Apolipoprotein E (APOE), Insulin Degradation Enzyme (IDE), Glucose Transporter type 4 (GLUT4), and Acetylcholinesterase (AChE)\textsuperscript{4}. Nowadays, computational-based methods are very popular because they save time and money, so experts are using computational methods to introduce important genes, proteins, drugs etc.

To overcome this cross-talk, various computational-based studies have been performed so far. Mittal et al. extracted differentially expressed genes from Type 2 diabetes mellitus (T2DM) and Alzheimer’s patients. Then, they constructed a protein–protein interaction network using the proteins’ first neighbors. Pathway enrichment analysis was also performed, and notable proteins as well as significant pathways were introduced\textsuperscript{5}.

Using Biological Expression Language, Karki et al.\textsuperscript{6} proposed six shared pathways between T2DM and AD. They also discussed the dangers of certain anti-diabetic drugs in the progression of AD.

Hu et al.\textsuperscript{7} employed data from the ROSMAP Project to introduce common pathways between T2DM and AD. They analyzed multi-omic data using an inference method.

In addition to the importance of proteins, recent research suggests that non-coding RNAs (ncRNAs) play an important role in the progression of complex diseases\textsuperscript{8-11}. The reason is that they are encoded by the majority of the human genome. Furthermore, approximately two-thirds of the experimentally detected microRNAs (miRNAs) are expressed in the brain and are active participants in its functions\textsuperscript{12}. Long non-coding RNAs (lncRNAs) have also been shown to play a biomarker role in the detection of IR, inflammation, and DM, as well as the progression of AD\textsuperscript{13,14}.

Ghiam et al.\textsuperscript{15} investigated the role of the ncRNAs as potential biomarkers in the cross-talk between DM and AD.

In addition to the biomarker role of ncRNAs in early disease detection, they may also be used as pharmacological targets in the treatment of a variety of disorders\textsuperscript{16}. On the other hand, a new drug takes a long time to complete all phases of drug design (at least 10 years). This procedure is both time-consuming and expensive. To address the aforementioned problems, computer-aided drug design (CADD) has been proposed as a method for predicting new drugs for further study. Methods of drug repositioning (also known as drug repurposing) have been utilized to identify new therapeutic roles for existing drugs. For this purpose, Shakil used molecular docking to investigate the effects of two antidiabetic drugs, Ertugliflozin and Sotagliflozin, and their targets, Sodium glucose cotransporter 2 and AChE, on AD. Finally, he emphasized the sotagliflozin structure for future Type 3 diabetes treatment\textsuperscript{17}.

This study aims to introduce (reuse) Food and Drug Administration (FDA) approved drugs for regulating the expression of common lncRNAs in the cross-talk between DM and AD. Additionally, drugs with side effects were reported.

The following section discusses data collection, machine learning, and network-based methods. Positive drugs (candidates for repurposing) and negative drugs (those with adverse effects) were proposed in the results section, and we will elaborate on them in the discussion section.

**Materials and methods**

**Materials.**  *Collection of data.* Insulin is required for normal cognitive functions of the hippocampus. According to scientists, the primary factor in the cross-talk between DM and AD is insulin resistance in the brain\textsuperscript{18}. Consequently, significant mRNAs were extracted from DM, AD, and IR. The set of common mRNAs was selected as follows:

\[
P = \text{DM}_p \cap \text{AD}_p \cap \text{IR}_p
\]

where \(\text{DM}_p\), \(\text{AD}_p\), and \(\text{IR}_p\) denote the sets of curated proteins associated with DM, AD and IR in the Disgenet database respectively.

The information of miRNAs and IncRNAs were collected from miRTarBase\textsuperscript{19}, LncBase\textsuperscript{20}, and Starbase\textsuperscript{21} databases. We attempted to select the most reliable mRNA–miRNA, mRNA–lncRNA, and miRNA–drug interactions with experimental evidence. To do so, we looked for mRNA–miRNA and mRNA–lncRNA interactions in the Starbase database with CLIP-Data ≥ 3 (strict selection), mRNA–miRNA interactions in the miRTarBase
Identifying significant miRNAs based on the mRNA-miRNA network. A mRNA–miRNA bipartite network was built in the first state. The nodes in one part of this network represent a set of selected mRNAs, while the nodes in the other part represent miRNAs which have interacted with at least one of the selected mRNAs retrieved from databases. Assume that \( P = \{p_1, p_2, \ldots, p_k\} \) and \( M = \{m_1, m_2, \ldots, m_l\} \) represent the set of mRNAs and miRNAs, respectively. If miRNA \( m_i \) interacts with mRNA \( p_j \), there is an edge between \( p_j \) and \( m_i \).

High degree miRNAs were chosen as significant in the constructed network. These miRNAs are denoted by \( M' = \{m'_1, m'_2, m'_3, \ldots, m'_s\} \).

Identification of significant lncRNAs based on the miRNA-lncRNA network. In this state, the miRNA-lncRNA bipartite network was constructed. Nodes in one part are the set of significant miRNAs (\( M' \)) and nodes in the other part are the set of lncRNAs received from databases which have interacted with at least one of the selected miRNAs. Assume that \( L = \{l_1, l_2, \ldots, l_t\} \) is the set of mentioned lncRNAs. If lncRNA \( l_j \) interacts with miRNA \( m'_i \), there is an edge between \( m'_i \) and \( l_j \). High degree IncRNAs were chosen as significant in this network. These IncRNAs are denoted by \( L' = \{l'_1, l'_2, l'_3, \ldots, l'_t\} \).
Drug collection based on the IncRNA-drug network. At first, the regulation (upregulation or downregulation) effect of members of L’ in both DM and AD were diagnosed. Let \( D = \{d_1, d_2, d_3, \ldots, d_n\} \) denoted the set of drugs which interacted with at least one of the elements in L’. Suppose that \( d_i \) interacts with both \( l_i \) and \( l’_i \) which are upregulated and downregulated respectively in both DM and AD. It means that \( d_i \) inhibits both of them therefore, the effect of \( d_i \) on \( l_i \) is positive (therapeutic effect) but on \( l’_i \) is negative (exacerbation effect). In this case we remove \( d_i \) from D. Accordingly, drugs with contradictory effect were omitted. Let \( D’ = \{d’_1, d’_2, d’_3, \ldots, d’_n\} \) are denoted the set of remaining drugs. Now we constructed the IncRNA-drug bipartite network which nodes in one part are L’ and the other part are D’. There is an edge between \( l_i \) and \( d’_j \) if drug \( d’_j \) interacts with IncRNA \( l_i \). Next, based on the therapeutic effect or side effect of D’ on the regulation of members in L’, D’ was divided in to \( D^+ = \{d’_1, d’_2, d’_3, \ldots, d’_n\} \) which cure the intended diseases and \( D^- = \{d’_{j+1}, d’_{j+2}, d’_{j+3}, \ldots, d’_n\} \) which severely increase the negative effects of the intended diseases. For instance, suppose that \( l_i \) is upregulated in both DM and AD and \( d’_i \) inhibits it therefore, \( d’_i \) will be categorized in \( D^+ \). On the other hand, suppose that \( l_i \) is upregulated too but \( d’_i \) causes overexpression of this IncRNA (inverse effect) so, by using this drug, the negative effects of both DM and AD will increase. As a result, \( d’_i \) will be categorized in \( D^- \).

The regulation (upregulation or downregulation) effect of L’ members in both DM and AD was initially identified. \( D = \{d_1, d_2, d_3, \ldots, d_n\} \) represents the set of drugs that interacts with at least one element in L’.

Suppose that \( d_i \) interacts with \( l_i \) and \( l’_i \), which are respectively upregulated and downregulated in both DM and AD. It indicates that \( d_i \) inhibits both \( l_i \) and \( l’_i \), therefore, the effect of \( d_i \) on \( l’_i \) is beneficial (therapeutic effect), while its effect on \( l_i \) is detrimental (exacerbation effect). In this instance, we eliminate \( d_i \) from D. Thus, drugs with contradictory effects were excluded. Let \( D’ = \{d’_1, d’_2, d’_3, \ldots, d’_n\} \) represents the remaining drugs. Now we have constructed the IncRNA-drug bipartite network, with L’ nodes in one part and D’ nodes in the other. If \( d’_j \) interacts with IncRNA \( l’_i \), there is an edge between \( l’_i \) and \( d’_j \). Next, based on the therapeutic effect or side effect of D’ on the regulation of members in L’, D’ was divided into \( D’^+ = \{d’_{j+1}, d’_{j+2}, d’_{j+3}, \ldots, d’_n\} \) which cure the intended diseases and \( D’^- = \{d’_1, d’_2, d’_3, \ldots, d’_{j-1}\} \) which significantly intensify the negative effects of the intended diseases. For example, if \( l’_i \) is upregulated in both DM and AD and \( d’_i \) inhibits it, then \( d’_i \) will be classified as \( D’^+ \). Alternatively, if \( l’_i \) is also upregulated but \( d’_i \) causes overexpression of this IncRNA, then the use of this drug will exacerbate the negative effects of both DM and AD. Consequently, \( d’_i \) will be classified as \( D’^- \).

Features of drugs. For drugs in D’, we converted sdf files to mol files for using Open Babel,25 a tool for converting chemical file formats. PaDEL26, a software for calculating molecular descriptors and fingerprints, was also employed to extract 1D–2D features based on the structural properties of D’ members. Only those features that best classified D’ into D’+ and D’− were selected. The algorithm for feature selection is as follows:

1. Eliminating constant features: Features that have the same value for all drugs in D’ were omitted because they had no impact on the classification process.
2. Removing redundant features: To improve performance, highly correlated features should be removed. To achieve this, we computed the Pearson correlation coefficient using the "findcorrelation" function from the ‘caret’ package in R, with a cutoff of 0.9. This function takes the absolute values of pair-wise correlations into account. The function eliminates the variable with the highest mean absolute correlation when two variables have a high correlation.
3. Discarding features with low entropy: To find efficient variables for the model, entropy was calculated for each feature using the entropy function in R, and features with entropy less than \( \theta \) were removed.
4. Removing features with a p-value greater than 0.05: For the remaining attributes, the t-student test was used to select features with significant differences (p value < 0.05) between D’+ and D’−.

Finally, significant attributes \( F = \{f_1, f_2, f_3, \ldots, f_n\} \) were chosen.

Model designing. A random forest classifier was constructed using features extracted from the previous state.

Out of bagging error (OOB) was calculated using Python’s random forest library to determine the optimal number of trees. OOB is a type of random forest mean prediction error in which trees without the intended sample in their bootstrap are utilized. The model was then evaluated using five-fold cross validation, and performance metrics were reported.

Choosing FDA-approved drugs that are relevant to members of D’. In the same manner as in previous states, PaDEL was used to extract 1D–2D features for FDA-approved drugs, and the same set of extracted features, F, was selected. Let \( D_{FDA} = \{df_1, df_2, df_3, \ldots, df_v\} \) denote drugs with selected features. Then, for DFDA, outlier detection was utilized to select the most pertinent FDA-approved drugs for D’. The procedure for detecting outliers is as follows:

1. Number of atoms: very small drugs with nATOM ≤ μ was removed from the study since they have no impact on the classification process.
2. Applicability of domain: the warning leverage for each compound was calculated using the formula: \( h_i = x_i^T (X^T X)^{-1} x_i \), where:
3. \( x_i \) is the query molecule’s description-row vector
4. \( X \) is the k×n matrix (k descriptor values for each of the n training molecules)
5. $h^*$ (warning leverage) is usually set to $3p/n$, where $p$ is the number of model variables ($|F|$) plus one and $n$ is the number of training compounds. Outliers were identified as drugs with a $hi$ value greater than the warning leverage and were removed from the dataset.

6. One class classification: although relevant drugs were selected from previous steps by removing outliers, to further optimize the running time of the classification algorithm, it is necessary to select highly relevant drugs to $D'$. To achieve this, for the remaining drugs in the $DFDA$, a single class classification method employing Support Vector Machine (SVM) classifier with radial basis was utilized, and other outliers were excluded. Finally, the most relevant compounds to $D'$ were chosen from all FDA-approved drugs. These drugs are denoted by $D'_{FDA} = \{df'_{1}, df'_{2}, df'_{3}, ..., df'_{q}\}$. Now, this set was fed into the newly constructed random forest model, and for each drug the probability of belonging to $D'$ or $D^-$ was calculated. The drugs with probabilities of 0.95 or higher were selected. Those with a positive effect (belongs to $D'$) were suggested as repurposing candidates and those with a negative effect (belongs to $D^-$) were introduced as warning drugs that could simultaneously increase the risk of DM and AD.

**Results**

**Network analysis.** Based on the data collection section above, 9 mRNAs shared by DM, AD, and IR were extracted from the curated Disgenet database (Table 1, first column). There are 412 miRNAs which have interacted with at least one of these mRNAs. Then we constructed the mRNA-miRNA bipartite network and significant miRNAs were extracted. In this network, no miRNA interacts with the INS and LEP proteins, so these two proteins were omitted from our investigation. Then, miRNAs with the highest degree (degree 4) were chosen as significant. By this strategy, 7 miRNAs were identified (Table 1, second column). Similarly, we constructed the miRNA-lncRNA bipartite network. This network contains 1700 lncRNAs, only three of them have the highest degree of 7, and eight of them have degree 6. These 11 lncRNAs were deemed significant (Table 1, forth column). These degrees are chosen so that approximately 10 miRNAs and 10 lncRNAs are considered as significant. For instance, if we considered degree 3 for the first network, we would select more than 40 miRNAs, and if we considered degree 5 for lncRNAs, we would select more than 30 lncRNAs. Supplementary Tables S1 and S2 list all mRNA-miRNA and miRNA-lncRNA interactions in these two networks. However, the final tripartite network was built using 7 selected mRNAs, 7 selected miRNAs, and 11 selected lncRNAs. Results have been shown in Fig. 2 and Table 1.

Table 1. Selected mRNAs, miRNAs and lncRNAs. Significant mRNAs, miRNAs, and lncRNAs are summarized in Table 1. No interacting miRNAs have been reported in the existing databases for INS and LEP among the 9 mRNAs. The hub miRNAs and lncRNAs are shown in columns 2 and 4, and the associated degrees of the intended miRNAs and lncRNAs are shown in columns 3 and 5.

| mRNAs (P) | MiRNAs (M) | Degree | LncRNAs (L') | Degree |
|-----------|------------|--------|--------------|--------|
| INS       | hsa-miR-3184-5p | 4      | NEAT1        | 7      |
| NOS3      | hsa-miR-17-5p  | 4      | XIST         | 7      |
| PPARG     | hsa-miR-20a-5p | 4      | EBLN3P       | 7      |
| SLCA4     | hsa-miR-93-5p  | 4      | H19          | 6      |
| SOD2      | hsa-miR-106b-5p| 4      | MALAT1       | 6      |
| HMOX1     | hsa-miR-20b-5p | 4      | HCG18        | 6      |
| TNF       | hsa-miR-200c-3p| 4      | PSMA3-AS1    | 6      |
| INS       | –           | 4      | AC021078.1   | 6      |
| LEP       | –           | 4      | AC085261.1   | 6      |
| INS       | –           | 4      | AC024940.6   | 6      |
| LEP       | –           | 4      | NORAD        | 6      |

Finally, the most relevant compounds to $D'$ were chosen from all FDA-approved drugs. These drugs are denoted by $D'_{FDA} = \{df'_{1}, df'_{2}, df'_{3}, ..., df'_{q}\}$. Now, this set was fed into the newly constructed random forest model, and for each drug the probability of belonging to $D'$ or $D^-$ was calculated. The drugs with probabilities of 0.95 or higher were selected. Those with a positive effect (belongs to $D'$) were suggested as repurposing candidates and those with a negative effect (belongs to $D^-$) were introduced as warning drugs that could simultaneously increase the risk of DM and AD.

**Significant mRNAs, miRNAs, and lncRNAs are summarized in Table 1.** No interacting miRNAs have been reported for INS and LEP among the 9 mRNAs. The hub miRNAs and lncRNAs are shown in columns 2 and 4, and the associated degrees of the intended miRNAs and lncRNAs are shown in columns 3 and 5.
In this table, 7 mRNAs in the cross-talk between DM and AD (achieved from curated Disgenet database) and their mechanisms in the relationship between the two intended diseases are introduced using literature review.

For the 7 significant miRNAs identified by the miRNA-lncRNA bipartite network, we also conduct a literature review to confirm their role in the cross-talk between DM and AD.

In both diseases, hsa-miR-106b-5p, hsa-miR-17-5p, hsa-miR-20a-5p, hsa-miR-93-5p, and hsa-miR-200c-3p have been downregulated, whereas hsa-miR-20b-5p has been upregulated. Downregulation of hsa-miR-106b-5p, hsa-miR-17-5p, hsa-miR-20a-5p and hsa-miR-93-5p as well as overexpression of hsa-miR-20a-5p may alter the expression of amyloid precursor protein (APP) and subsequently influence Aβ, which is responsible for hyperphosphorylation of tau protein. In addition, altering the expression of hsa-miR-200c-3p may increase the expression of GSK3β, which results in hyperphosphorylation of tau. Moreover, hsa-miR-3184-5p is a novel miRNA in the cross-talk between DM and AD, which is proposed in this study for future experimental investigation. Table 3 shows the summary of these results.

This table describes the mechanisms of 7 significant miRNAs in the cross-talk between DM and AD. Only miR-3184-5p introduced in this study as a novel miRNA in this cross-talk which needs additional experimental research.

Similarly, we investigated the role of obtained IncRNAs in DM and AD cross-talk through literature review. Recent studies suggest that NEAT1 and XIST, two important biomarkers in DM and AD, are upregulated in both diseases, which not only activate the Akt/mTOR signaling pathway but also dysregulate the expression of miR-124, resulting in an increase in BACE1 concentration and Aβ plaque deposition in the brain. H19, another

---

**Table 2.** mRNAs interpretation. In this table, 7 mRNAs in the cross-talk between DM and AD (achieved from curated Disgenet database) and their mechanisms in the relationship between the two intended diseases are introduced using literature review.

| mRNA   | Expression | Mechanism                               |
|--------|------------|-----------------------------------------|
| HMOX1  | Upregulation | Formation of neurofibrillary tangles   |
| INSR   | Downregulation | Insulin resistance                      |
| NOS3   | Upregulation | Apoptosis in cortical neurons           |
| PPARG  | Upregulated  | Targeting Wnt/Beta-Catenin Pathway      |
| SLC2A4 | Downregulation | Disrupting glucose transportation in the brain |
| SOD2   | Downregulation | Mitochondrial dysfunction               |
| TNF    | Upregulation | Increasing inflammation                  |

---

**Figure 2.** Final mRNA-miRNA-IncRNA network. The first layer (purple nodes) consists of 7 shared mRNAs between DM and AD that were extracted from curated Disgenet database. The second layer (blue nodes) consists of 7 high degree miRNAs obtained from mRNA-miRNA bipartite network and the last layer (yellow nodes) consists of 11 high degree IncRNAs achieved from miRNA-IncRNA bipartite network.
significant lncRNA, is upregulated in DM and AD, resulting in hippocampal neuron apoptosis. Meanwhile, low levels of MALAT1 (downregulation) have been found in Alzheimer’s patients compared to controls, reducing the lncRNA’s anti-apoptotic effect. Abdulle et al. also demonstrated the significance of MALAT1 as a target in diabetic-related diseases. Additionally, HCG18 and BPLN3P are upregulated, which increases the risk of insulin resistance and type 1 diabetes, respectively. The remaining obtained lncRNAs are novel in this cross-talk which are proposed in this study for future experimental analysis. Table 4 displays the summary of these results.

Table 4. LncRNAs interpretation. This table explains the role of the obtained lncRNAs in the cross-talk between DM and AD. Among 11 significant lncRNAs, 5 of them are novel, proposed in this study for future experimental analysis.

| lncRNA | Expression | Mechanism |
|--------|------------|-----------|
| NEAT1  | Upregulation | Biomarkers in DM and AD |
| XIST   | Upregulation | Insulin resistance |
| HCG18  | Upregulation | Increasing the risk of type 1 diabetes mellitus |
| EBLN3P | Upregulation | Apoptosis of hippocampal neurons |
| H19    | Upregulation | |
| MALAT1 | Downregulation | |
| PSMA3-AS1 | Novel | |
| AC021078.1 | Novel | |
| AC005261.1 | Novel | |
| AC024940.6 | Novel | |
| NORAD  | Novel | |

Table 3. miRNAs interpretation. This table describes the mechanisms of 7 significant miRNAs in the cross-talk between DM and AD. Only miR-3184-5p introduced in this study as a novel miRNA in this cross-talk which needs additional experimental research.

| miRNA     | Expression | Mechanism                                                                 |
|-----------|------------|---------------------------------------------------------------------------|
| hsa-miR-106b-5p | Downregulation | Regulating Aβ induced tau phosphorylation via regulating the expression of Fyn |
| hsa-miR-17-5p | Downregulation | Regulating APP expression                                                                 |
| has-miR-20a-5p | Downregulation | Regulating Aβ plaques                                                                 |
| hsa-miR-20b-5p | Upregulation | Attenuating apoptosis induced by Aβ25–35                                    |
| hsa-miR-93-5p | Downregulation | |
| hsa-miR-200c-3p | Downregulation | Increasing tau phosphorylation (p-tau) by increasing p-GSK-3β               |
| hsa-miR-3184-5p | Novel miRNA | -                                                                        |

Feature description. PaDEL was utilized to extract 1444 features from the 40 drugs that were chosen from the lncRNA-drug bipartite network. There were 293 omitted constant features. Then, 745 highly correlated features with cutoff = 0.9 were also removed (Fig. 3A). Low entropy features (θ < 1.5) that had no effect on classifying D− and D+ were removed as well (Fig. 3B).

Consequently, 293 features remained. The t-student test was then performed to identify the most significant features, and 6 of them with a p value less than 0.05 that best classified D− and D+ were selected. The final drug-feature matrix, which includes 40 drugs and 6 significant features, is presented in Table S3 of the Supplementary materials.
Table 5. Drugs and the corresponding lncRNAs. The drugs in the first column (D−) and the third column (D+) have a negative and positive effect on the common significant lncRNAs in DM and AD respectively. Furthermore, the interaction of lncRNAs with drugs is reported in columns 2 and 4, respectively.

| Drugs (D−)     | lncRNAs | Drugs (D+)     | lncRNAs     |
|----------------|---------|----------------|-------------|
| Panobinostat   | NEAT1, H19 | Adriamycin | NEAT1, MALAT1 |
| Diamorphine    | NEAT1   | Dexamethasone | NEAT1       |
| Deferasirox    | NEAT1   | Arachidonic acid | NEAT1       |
| Ubiquinol      | NEAT1   | Sorafenib    | NEAT1       |
| Nilotinib      | NEAT1, XIST | Estradiol | NEAT1       |
| Rosiglitazone  | NEAT1   | Cilnidipine  | NEAT1       |
| Trovafloxacin  | NEAT1   | Fluorouracil | NEAT1       |
| Sangivamycin   | NEAT1, XIST | Oxymatrine | MALAT1       |
| Busphenol A    | NEAT1   | Quercetin    | XIST, MALAT1 |
| Apratoxin A    | NEAT1, XIST | pma      | MALAT1       |
| Bexarotene     | NEAT1   | Curcumin     | H19         |
| beta-Asarone   | MALAT1  | Plx4720      | H19         |
| Bleomycin      | MALAT1  | Mk-886       | H19         |
| 3–3′–4′ tetra-chlorobiphenyl | MALAT1 | Tamoxifen | H19         |
| Oxaliplatin    | H19, MALAT1 | Fulvestrant | H19         |
| Vincristine    | H19, MALAT1 | Diclofenac | XIST        |
| Etanercept     | H19     |               |             |
| Doxycycline    | H19     |               |             |
| Azathioprine   | H19     |               |             |
| Decitabine     | XIST, H19 |               |             |
| Etoposide      | H19     |               |             |
| Verapamil      | H19     |               |             |
| Epirubicin     | H19     |               |             |
| Pirarubicin    | H19     |               |             |
| Bortezomib     | H19     |               |             |
| Azacitidine    | XIST    |               |             |

Figure 3. The feature selection process. (A) The correlation matrix of 1444 features achieved from PaDEL has been displayed. The values range from −1 to 1, with −1 denoting a negative correlation and 1 denoting a positive correlation. Features that were highly correlated (cutoff = 0.9) were removed. (B) Features with low entropy (below the red dashed line) were omitted.
Random forest classifier. A random forest classifier based on the drug-feature matrix was used to construct a model for further analysis. The optimal number of trees was determined to be approximately 100 based on the tradeoff between OOB error and accuracy. Figure 4 depicts the findings.

Furthermore, fivefold cross validation (cv) was used to evaluate the constructed random forest classifier, and several performance measures were reported. During the evaluation 80% of the data was selected for the training dataset and 20% for the test dataset. As shown in Tables 6 and 7, the accuracy of each fold in a fivefold cv ranges between 0.75 and 0.875%, and the overall accuracy is 0.825. In addition, the precision and recall values for D+ and D− ranges from 0.75 to 1. The F1 measures for negative and positive sets are 0.89 and 0.86, respectively. Results have been shown in Tables 6 and 7.

Table 6. Performance measures of the constructed random forest model with fivefold cv. *Accuracy: Acc.

| Fold 1 Acc | Fold 2 Acc | Fold 3 Acc | Fold 4 Acc | Fold 5 Acc | Overall Acc |
|------------|------------|------------|------------|------------|-------------|
| 0.875      | 0.75       | 0.875      | 0.875      | 0.75       | 0.825       |

Table 7. Confusion matrix.

|            | Precision | Recall | F1-score |
|------------|-----------|--------|----------|
| Negative   | 0.8       | 1      | 0.89     |
| Positive   | 1         | 0.75   | 0.86     |

Random forest classifier. A random forest classifier based on the drug-feature matrix was used to construct a model for further analysis. The optimal number of trees was determined to be approximately 100 based on the tradeoff between OOB error and accuracy. Figure 4 depicts the findings.

Furthermore, fivefold cross validation (cv) was used to evaluate the constructed random forest classifier, and several performance measures were reported. During the evaluation 80% of the data was selected for the training dataset and 20% for the test dataset. As shown in Tables 6 and 7, the accuracy of each fold in a fivefold cv ranges between 0.75 and 0.875%, and the overall accuracy is 0.825. In addition, the precision and recall values for D+ and D− range from 0.75 to 1. The F1 measures for negative and positive sets are 0.89 and 0.86, respectively. Results have been shown in Tables 6 and 7.

Selecting relevant FDA approved drugs. The structural files of 2474 FDA-approved drugs (D_{FDA}) were downloaded from DrugBank, and 1D-2D features were extracted from PaDEL. Drugs with a nATOM (number of atoms) less than 7 were eliminated (Fig. 5A). The 6 significant features obtained in the previous step were chosen for the remaining 2317 drugs. However, for 13 drugs, the value of some selected features was unavailable, so they were eliminated. The total number of drugs was thus reduced to 2304. To detect outliers, first, the applicability of the domain was determined, and then the warning leverage was computed as follows:

\[ \text{Warning leverage} = \frac{3p}{n} = 3 \times (6 + 1)/2304 = 0.0091 \]

Each drug with a warning leverage equal to or exceeding the aforementioned threshold was omitted (Fig. 5B). Consequently, 94 drugs were eliminated.

The remaining 2210 drugs were classified using a one-class classification SVM method, and 1221 drugs (D’_{FDA}) were selected as the most relevant to at least one of the 40 drugs.
Then, $D'_{FDA}$ was given as input into the constructed random forest model, and the probability of each drug belonging to the $D^+$ or $D^-$ was calculated. The drugs with a probability of 0.95 or higher were deemed the most relevant. Due to their positive effect on the regulation of lncRNAs associated with DM and AD, certain $D^+$ drugs have been proposed as repositioning candidates. Those in $D^-$ have been classified as adverse drugs as they have a negative effect on the regulation of corresponding lncRNAs. Table 8 lists all selected FDA-approved drugs discovered in this study.

The final FDA-approved negative and positive drugs are listed in this table. Positive drugs are suggested for repurposing. Negative drugs, on the other hand, have been introduced as warning drugs that may disrupt the expression of common lncRNAs and significantly increase the negative effects of both diseases simultaneously.

Discussion
In this work, using network and machine learning approaches, candidate drugs for repurposing were introduced based on common mRNAs, miRNAs, and lncRNAs associated with the cross-talk between DM and AD. Drugs that may exacerbate the negative effects of both diseases were also discussed. The first 11 significant lncRNAs, as listed in Table 1, are taken into account. Only four of these lncRNAs have been associated with drug interactions in lncRNA-drug databases: NEAT1, XIST, H19, and MALAT1. The interactions with the remaining seven lncRNAs have not yet been identified. Therefore, this study focuses on the four lncRNAs listed above in order to identify suitable drugs for the repositioning objective.

Scientists believe that inhibiting NEAT1 and XIST reduces Aβ plaques in patients' brains by regulating the expression of interacted ncRNAs and proteins in specific pathways. Therefore, silencing of the aforementioned lncRNAs may represent a novel potential therapeutic target in the treatment of both diseases. Furthermore, researchers show that inhibiting H19 prevents apoptosis in hippocampal neurons in both diseases. Additionally, experiments indicate that overexpression of MALAT1 increases the lncRNA's neuroprotective function in the brains of patients and decreases neuronal injury in the hippocampus.

The discovery of drugs that can silence NEAT1, XIST, or H19 while also inducing MALAT1 overexpression could be considered a new treatment for DM and AD cross-talk.

Conversely, drugs that have a negative effect on the targeted lncRNAs must be avoided. According to Table 8, there are two categories of FDA-approved drugs. Positive and negative drugs are listed, respectively, in the first and third columns. Positive drugs have been recommended for diabetic and Alzheimer’s patients. Estradiol, for instance, is effective in preventing Aβ plaques and tau hyperphosphorylation in the brains of Alzheimer's patients, in addition to reducing the risk of diabetes. It has been demonstrated that drugs in the chromium family have a curable effect on DM, AD, and IR, as they not only improve cognitive function in older adults but also regulate glucose excretion in patients with DM and IR. Consequently, it may be one of the most significant candidate drugs in this cross-talk. Tapentadol, a drug used to treat pain in the elderly, is another positive drug. According to researchers, it may improve the behavior of elderly dementia patients. Additionally, it has been proposed as a novel therapy for diabetic neuropathy. Niacin is a B vitamin that facilitates the transformation of food into energy. According to research on the effect of niacin on AD, scientists believe that including niacin in one's diet can aid in disease prevention. Volunteers who received less niacin were up to 70 percent more likely
to develop the disease than those who received more. However, it has the potential to increase glucose levels in diabetic patients, so dosage must be closely monitored\(^54\). Sibutramine is used for weight loss. Additionally, it reduces LDL cholesterol, thereby reducing the cardiovascular risk in diabetic patients\(^55\). Moreover, since high cholesterol levels may increase the risk of AD\(^56\), it may be a new drug capable of curing Alzheimer’s patients, particularly by controlling APOE.

Curcumin, a bright yellow chemical produced by curcuma, and quercetin, a plant pigment found in grapefruit, onion, and other fruits and vegetables, are two positive natural compounds. According to numerous studies, curcumin plays an important role in reducing Aβ plaques and inhibiting tau aggregation, as well as controlling blood sugar and managing diabetes complications. In addition, quercetin reduces oxidative stress by regulating signaling pathways, as well as lowering hyperglycemia and improving IR. In AD, it also has neuroprotective properties.

On the other hand, negative drugs have not been suggested for patients with DM and AD. Diamorphine, for instance, increases tau hyperphosphorylation in AD patients by targeting the FYN protein\(^57\). The FDA has not approved zaleplon, a sleeping pill, for Alzheimer’s patients because this class drugs have been linked to an increased risk of dementia in the elderly\(^58\). The anti-inflammatory drug, sulfasalazine, is used to treat Crohn’s disease and rheumatoid arthritis. Despite being an anti-inflammatory, chou et al. discovered that it may increase the risk of dementia in patients with rheumatoid arthritis. Hypoglycemia has been added to the list of side effects of sulfasalazine in T2DM patients\(^59,60\). Geodon is the brand name for the antipsychotic drug ziprasidone, which is used to treat schizophrenia and bipolar disorder. Experimental studies indicate that it increases the risk of death in Alzheimer’s patients, so it is neither recommended nor approved for dementia.

---

| Positive drugs          | Probability | Negative drugs          | Probability |
|-------------------------|-------------|-------------------------|-------------|
| Dexamethasone           | 1           | Panobinostat            | 1           |
| Sorafenib               | 1           | Diamorphine             | 1           |
| Estradiol               | 1           | Deferasirox             | 1           |
| Fluorouracil            | 1           | Nilotinib               | 1           |
| Curcumin                | 1           | Rosiglitazone           | 1           |
| Quercetin               | 1           | Trovafloxacin           | 1           |
| Tamoxifen               | 1           | Bisphenol A            | 1           |
| Fulvestrant             | 1           | Hexarotene              | 1           |
| Diclofenac              | 1           | Oxaliplatin             | 1           |
| Chromium nicotinate     | 0.97        | Vindesine               | 1           |
| Tapentadol              | 0.97        | Doxycycline             | 1           |
| Niacin                  | 0.96        | Azothiaprine            | 1           |
| Chlorophenesin          | 0.95        | Decitabine              | 1           |
| Methanesin              | 0.95        | Etoposide               | 1           |
| Homosulate              | 0.95        | Verapamil               | 1           |
| Sibutramine             | 0.95        | Epirubicin              | 1           |
| 2-Mercaptobenzothiazole | 0.95        | Bortezomib              | 1           |
| 2-Mercaptobenzothiazole |             | Azacitidine             | 1           |
| Dihydro-alpha-ergocryptine |         | 0.99                    |             |
| Cefepime                |             | 0.99                    |             |
| Epicriptide             | 0.984       | 0.98                    |             |
| Alectinib               | 0.98        | 0.98                    |             |
| Dihydroergocristine     | 0.974       | 0.974                   |             |
| Deferoxamine            | 0.974       | 0.974                   |             |
| Glipizide               | 0.974       | 0.974                   |             |
| Ziprasidone             | 0.97        | 0.97                    |
| Betrixaban              | 0.97        | 0.97                    |
| Remimazolam             | 0.97        | 0.97                    |
| Glimepiride             | 0.97        | 0.97                    |
| Dihydroergotamine       | 0.97        | 0.97                    |
| Tazemetostat            | 0.97        | 0.97                    |
| Zaleplon                | 0.96        | 0.96                    |
| Cocarboxylase           | 0.96        | 0.96                    |
| Sulfasalazine           | 0.954       | 0.954                   |             |

Table 8. Final selected FDA approved drugs with positive and negative effects. The final FDA-approved negative and positive drugs are listed in this table. Positive drugs are suggested for repurposing. Negative drugs, on the other hand, have been introduced as warning drugs that may disrupt the expression of common lncRNAs and significantly increase the negative effects of both diseases simultaneously.
patients. Furthermore, it increases the risk of hyperglycemia and diabetes. Cefepime, an antibiotic used to treat bacterial infections, has been linked to disruption of brain function; therefore, Alzheimer’s patients should only take it on a doctor’s advice.

In DM or AD, drugs such as cocarboxylase, glipizide, glimepiride, and the ergotamine family are being investigated. They are, however, classified as negative drugs in our study because they have an adverse effect on the regulation of common IncRNAs, and further experimental research into their impact on DM and AD is recommended.

There is insufficient information regarding the associations between the remaining drugs and the intended diseases; therefore, more research is required.

In addition to the obtained drugs, we investigated the impact of well-known drugs on the expression of the significant mRNAs, miRNAs, and IncRNAs that are currently used for the treatment of DM and AD from DrugBank database. There are 52 and 11 DM and AD drugs, respectively. In our results, the anti-diabetic drug Rosiglitazone is categorized as a negative drug. Based on our findings, the use of this drug may increase the risk of AD by elevating NEAT1 expression (see Tables 4, 5). Therefore, this drug may be an important candidate for future experimental research on the role of anti-diabetic drugs in the progression of AD.

However, there is no other overlap between drugs that we introduced in this study and the remaining 62 DM or AD drugs in the DrugBank database.

In future work, we will increase the reliability of our study by evaluating the expression of introduced IncRNAs using human samples. Then, we intend to conduct a laboratory experiment at the Royan Institute for Stem Cells to test introduced candidate drugs on the obtained IncRNAs using mouse samples. On the other hand, we proposed that dysregulation of certain mRNAs and miRNAs causes cross-talk between DM and AD. Since the focus of this paper was on IncRNA-drug repurposing, one of the future studies could concentrate on drug repurposing based on introduced mRNAs and miRNAs.

This study can be expanded to include any other diseases with a correlation, such as DM and fatty liver or AD and Parkinson’s disease. The pipeline remains unchanged; only the extracted data changes. Consequently, this work can serve as a guideline for similar work on two other diseases that interact with each other.

Conclusion

Drug repurposing is one of the most important methods for developing new drugs from existing ones for the treatment of diseases. This paper proposed a network and machine learning-based method that not only suggests drugs to inhibit DM and AD cross-talk, but also introduces drugs that exacerbate the negative effects of the aforementioned diseases.

For instance, the drugs chromium nicotinate and tapentadol, which are used to treat IR and pain, respectively, may regulate the expression of NEAT1, XIST, MALAT1, and H19 in the brains of diabetic and Alzheimer’s patients. Consequently, these two drugs can be considered a novel treatment for this cross-talk. In addition, drugs such as ziprasidone and sulfasalazine, which are used to treat crohn’s disease and schizophrenia, may increase the risk of both DM and AD. The list of repurposing candidate drugs and adverse drugs for which we wish to conduct experimental research to confirm our findings is presented in Table 8. Two of the drugs in this table, curcumin and quercetin, are natural compounds that can regulate the expression of H19, XIST, and MALAT1. Therefore, long-term use of these natural products may reduce the risk of DM and AD.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 25 May 2022; Accepted: 19 October 2022
Published online: 31 October 2022

References

1. Baglietto-Vargas, D., Shi, J., Yaeger, D. M., Ager, R. & LaFerla, F. M. Diabetes and Alzheimer's disease crosstalk. *Neurosci. Biobehav. Rev.* **64**, 272–287 (2016).
2. McIlroy, E. C. & Pearson-Leary, I. GliTα: A central player in hippocampal memory and brain insulin resistance. *Exp. Neurol.* **323**, 113076 (2020).
3. Kandimalla, R., Thirumala, V. & Reddy, P. H. Is Alzheimer's disease a Type 3 diabetes? A critical appraisal. *Biochim. Biophys. Acta* **1863**(5), 1078–1089 (2017).
4. Mittal, K. & Katare, D. P. Shared links between type 2 diabetes mellitus and Alzheimer’s disease: A review. *Diabetes Metab. Syndr.* **10**(Suppl 1), S144–149 (2016).
5. Mittal, K., Mani, R. J. & Katare, D. P. Type 3 diabetes: Cross talk between differentially regulated proteins of type 2 diabetes mellitus and Alzheimer's disease. *Sci. Rep.* **6**(1), 25889 (2016).
6. Karki, R., Kodamullil, A. T. & Hofmann-Apitius, M. Comorbidity analysis between Alzheimer's disease and type 2 diabetes mellitus (T2DM) based on shared pathways and the role of T2DM drugs. *J. Alzheimer's Dis.* **60**(2), 721–731 (2017).
7. Hu, Z. et al. Shared causal pathways underlying Alzheimer's dementia and type 2 diabetes. *Sci. Rep.* **10**(1), 4107 (2020).
8. DiStefano, I. J. The emerging role of long noncoding RNAs in human disease. *Methods Mol. Biol. (Clifton, NJ)* **1706**, 91–110 (2018).
9. Lekka, E. & Hall, J. Noncoding RNAs in disease. *FEBS Lett* **592**(17), 2884–2900 (2018).
10. Li, Y. & Kowdley, K. V. MicroRNAs in common human diseases. *Genomics Proteomics Bioinform.* **10**(5), 246–253 (2012).
11. Natarajan, S. K., Smith, M. A., Wehrkamp, C. J., Mohr, A. M. & Mott, J. L. MicroRNA function in human diseases. *Med. Epigenet.* **1**(1), 106–115 (2013).
12. Idda, M. L., Munk, R., Abdelmohsen, K. & Gorospe, M. Noncoding RNAs in Alzheimer's disease. *Wiley Interdiscip. Rev. RNA* **9**(2), e1463 (2018).
13. He, X. et al. LncRNAs: Key players and novel insights into diabetes mellitus. *OncoTarget* **8**(41), 71325–71341 (2017).
56. Sáiz-Vazquez, O., Puente-Martínez, A., Uhillos-Landa, S., Pacheco-Bonrostro, J. & Santabárbara, J. Cholesterol and Alzheimer’s disease risk: A meta-meta-analysis. Brain Sci. 10(6), 386 (2020).
57. Egervari, G. et al. Chromatin accessibility mapping of the striatum identifies tyrosine kinase FYN as a therapeutic target for heroin use disorder. Nat. Commun. 11(1), 4634 (2020).
58. Shih, H. I. et al. An increased risk of reversible dementia may occur after zolpidem derivative use in the elderly population: A population-based case-control study. Medicine 94(17), e809 (2015).
59. Chou, R. C., Kane, M., Ghimire, S., Gautam, S. & Gui, J. Treatment for rheumatoid arthritis and risk of Alzheimer’s disease: A nested case-control analysis. CNS Drugs 30(11), 1111–1120 (2016).
60. Stamatiades, G. A., Echouffo-Tcheugui, J. B. & Garber, J. R. Sulfasalazine-induced hypoglycemia in a patient with type 2 diabetes and end-stage renal disease. AAACE Clin. Case Rep. 4(6), e493–e496 (2018).
61. Singh, S. & Wooltorton, E. Increased mortality among elderly patients with dementia using atypical antipsychotics. CMAJ Can. Med. Assoc. J. 173(3), 252 (2005).

Acknowledgements
The financial support of the Research Council of the University of Tehran is acknowledged.

Author contributions
S.G. designed and performed experiments, analyzed data and wrote the paper. C.E., S.G., K.S. and M.H.-R. supervised the research. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22822-9.

Correspondence and requests for materials should be addressed to C.E. or S.G.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022