Title: Predicting genes associated with RNA methylation pathways using machine learning

Authors: Georgia Tsagkogeorga[1,2*], Helena Santos-Rosa[3], Andrej Alendar[3], Dan Leggate[1], Oliver Rausch[1], Tony Kouzarides[2,3], Hendrik Weisser[1†*] and Namshik Han[2,4†*]

Affiliations:
[1]STORM Therapeutics Ltd, Babraham Research Campus, Cambridge, UK
[2]Milner Therapeutics Institute, University of Cambridge, Puddicombe Way, Cambridge, UK
[3]The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, UK
[4]Cambridge Centre for AI in Medicine, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK
[†]Contributed equally
[*]Corresponding authors

ABSTRACT

RNA methylation plays an important role in functional regulation of RNAs, and has thus attracted an increasing interest in biology and drug discovery. Here, we collected and collated transcriptomic, proteomic, structural and physical interaction data from the Harmonizome database, and applied supervised machine learning to predict novel genes associated with RNA methylation pathways in human. We selected five types of classifiers, which we trained and evaluated using cross-validation on multiple training sets. The best models reached 88% accuracy based on cross-validation, and an average 91% accuracy on the test set. Using protein-protein interaction data, we propose six molecular sub-networks linking model predictions to previously known RNA methylation genes, with roles in mRNA methylation, tRNA processing, rRNA processing, but also protein and chromatin modifications. Our study exemplifies how access to large omics datasets joined by machine learning methods can be used to predict gene function.

INTRODUCTION

RNA modifications have been known since the 1960s, when the sequencing of the first transfer RNA (tRNA) from yeast revealed 10 chemically modified ribonucleosides, including pseudouridine (Ψ)[1]. Since then, the number of identified modifications has grown to over 150, found on both coding and non-coding RNAs across all three kingdoms of life[2]. Technological advances in the field have established that RNA modifications are widespread, reversible and dynamically regulated[1]. Methylation is the most abundant type, with methyl-groups decorating multiple RNA species, such as messenger RNA (mRNA), ribosomal RNA (rRNA) and tRNA, at different nucleosides and positions. So far, N6-methyladenosine (m^6A) is the most studied modification, commonly detected in mRNA, rRNA, long intergenic non-coding RNA (lincRNA), primary microRNA (pri-miRNA), and small nuclear RNAs (snRNA). Other methyl-marks include 5-methylcytosine (m^5C), N1-methyladenosine (m^1A), 7-methylguanosine (m^7G), 2'-O-dimethyladenosine (m^6Am) and 5-hydroxymethylcytosine (hm^5C)[3–5].

Deposition of methyl-marks on RNA is catalysed by writer enzymes, known as RNA methyltransferases. To date, there are 57 RNA methyltransferases identified in the human genome. Of these, five methylate mRNAs, six small RNAs, 14 rRNAs, and 22 tRNAs, whereas 12 remain with unknown substrates[6]. Most enzymes use S-adenosyl-methionine (SAM) as a methyl donor to the RNA substrate, while many also recruit accessory proteins, which are often essential for substrate binding, localization, and stability. The most well-studied examples of RNA methylation writers are by far the complex METTL3-METTL14 complex responsible for the deposition of m^6A, followed by a NOL1/NOP2/Sun (NSUN) domain-containing family of tRNA-modifying enzymes depositing m^5C on tRNAs[7].

Dynamic regulation of RNAs via chemical modifications has recently attracted a rising interest in RNA modifying enzymes as new potential therapeutic targets[8]. This is because multiple lines of evidence suggest that RNA methylation plays a far more important role in cell functioning than previously thought. In line with this, several studies have shown that RNA methylation is a key modulator of transcript stability, gene expression, splicing and translation efficiency[9–11]. Furthermore, a growing body of data has demonstrated that changes in RNA methylation processes can be linked to a range of cancers, neurological disorders and various other diseases[12]. Surprisingly, despite this critical role in cellular homeostasis and disease, RNA methylation pathways in general remain understudied[7]. Our current understanding of RNA modifications is also highly fragmentary, with an estimated 20% or more of RNA modifying enzymes still remaining unknown or unidentified[13].

Conventional approaches for studying novel gene functions include a range of labour-intensive wet-lab techniques, including mutagenesis, gene disruption or gene depletion (knocking-down/-out) for characterising gene-specific phenotypic effects, and chromatography and mass spectrometry for identifying molecular interactions. However, over the last two decades, access to large-scale omics data has enabled the use of "dry" computational methods for understanding biological functions. A wide array of bioinformatic tools have been developed under the umbrella of functional genomics, ranging from methods used to identify homologous genes with similar functionalities across species to genome-wide screens for specific sequence motifs and functional domains. Today, machine learning techniques are emerging as a powerful approach to harness the increasing wealth of large-scale biological data, allowing the discovery of hidden patterns and more reliable statistical predictions[14].

Here, we aimed to better understand the molecular pathways involved in RNA methylation in human using machine learning. To this end, we used publicly available human transcriptomic, proteomic, structural and protein-protein interaction data[15] and built a large machine learning dataset for supervised binary classification. We trained and evaluated five ensembles of predictive models: Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting (GB) models. We employed the best models to predict genes functionally associated with RNA methylation pathways in the human genome.

RESULTS AND DISCUSSION

Data engineering and feature selection

Mining functional annotation databases in conjunction with extensive literature searches allowed us to identify 92 proteins involved in RNA methylation (Table 1). These were either methyl-writers (known RNA methyltransferases[6] and their partner proteins in protein complexes), or enzymes previously annotated as putative RNA methyltransferases (see Methods). Genes encoding for these proteins constituted our positive class (Class 1) in machine learning analyses. To frame our predictive modelling as a binary classification problem, we assembled multiple stratified training and test datasets by randomly sampling a number of genes equal to our positive set from the remaining genome, ensuring that all genes of our initial dataset were sampled exactly once (Figure 1). Our rationale was that this would allow machine learning models to be trained and tested across a diverse range of other gene functions, instead of just choosing one function for the negative set. In addition, this approach alleviates any putative bias that may arise from sampling a single negative set of genes from the human genome.

We initially pooled 50,176 features collected from publicly available and previously curated transcriptomic, proteomic, functional annotation, structural and physical interaction datasets (Table 2). To identify features that were informative for classification and thereby useful for predicting genes associated with RNA methylation, we performed feature selection prior to model training, followed by feature ranking after training and cross-validation. To reduce the feature-to-sample ratio, first we eliminated features with excessive missing data in the training dataset. Second, we removed features with low variance, which resulted in a drastic dimensionality reduction to 1,505 features for the final dataset. Selected features used for classification were drawn from BioGPS[16] (35), Gene Ontology[17] (GO: 59), GTEx[18] (1,114), Human Protein Atlas[19] (HPA: 107), InterPro (1), Pathway Commons (PathCommons: 150) and TISSUES[20] (40) datasets.

During model training and cross-validation, we computed feature importance by using the GB importance measure as averaged across all training sets. The 50 most informative features and their relative importance in classification are shown in Figure 2. The features with the highest importance for the full feature set were mainly GO terms, such as GO:0032259, GO:0016740, GO:0003723, GO:0008168 and GO:0016070, all corresponding to methylation, transferase/methyltransferase activity and RNA metabolic processes. Equally, the InterPro domain IPR029063, which represents the S-adenosyl-L-methionine-dependent methyltransferase superfamily was ranked among the top 50 most informative features (Figure 2A). Although anticipated, the fact that the classifiers seemed to rely on RNA and methylation-related annotation features provides support that the models learn to classify genes with a strong link to RNA methylation processes.

Although GO annotations are informative, they may equally bias gene prediction towards pre-existing functional annotations. We assembled thus a second feature set of reduced dimensionality, by excluding GO and InterPro data types. When classifiers were trained on this reduced feature set, the most informative types of features were mainly GTEx expression profiles (Figure 2B). The GTEx project aims to provide a comprehensive public resource of tissue-specific gene expression and regulation, so far including samples from 54 non-diseased tissues across nearly 1000 individuals[18]. Tissue sample expression data as integrated in Harmonizome and thus sampled here, consist of one-hot-encoded sets of genes with high or low expression in each tissue sample relative to other tissue samples from the GTEx tissue expression profiles dataset.

A possible interpretation of the high ranking of such GTEx expression profile features is that under specific biological conditions, i.e., in certain tissues, RNA methylation genes tend to be collectively down- or up-regulated as compared to other processes. Alternatively, a high ranking of GTEx features may be due to the high proportion of GTEx features in the feature set and noise originating from the high dimensionality of the training dataset with respect to the feature-to-sample ratio. To investigate this further, we calculated the relative frequency of GTEx features in the top hundred most informative features across models from all training sets (Table 3). Notably, certain samples taken from the areas of blood, heart, pancreas, and brain were retrieved as informative by more than a hundred models.

Model performance

We selected five machine learning classifiers (LR, GNB, SVM, RF and GB) and trained each on training sets from the full and the reduced feature set, creating an ensemble of models per classifier and feature set. To evaluate model performance, we used 10-fold cross validation and standard performance quantification metrics, i.e., accuracy, precision, recall, F1 score, and Area Under the Curve of the Receiver Operating Characteristic (AUCROC). Overall, all five model ensembles showed very similar performance based on cross-validation (Table 4). Among classifiers trained using the full feature set, GB and RF models showed the highest average accuracy at 0.875 and 0.870, respectively, as well as a similarly high average precision of 0.895 and 0.870, respectively. The GB ensemble followed by that the RF models also yielded the highest AUROC score, with an average AUC estimated at 0.938 and 0.937, respectively.

The performance of the five classifiers for the reduced feature set without GO/InterPro annotations was diminished compared to the full dataset (Table 4). The model ensembles of SVM and RF outperformed the remaining three ensembles across almost all metrics. SVM models performed the best on the reduced feature set based on cross-validation, with an average prediction accuracy of 0.812, precision of 0.822 and AUROC of 0.864.

Based on the above results, we selected the best model ensembles to apply on previously unseen test data: GB for the full feature set and SVM for the reduced feature set. Accuracy, precision, recall and AUCROC for the test datasets were calculated by averaging the values obtained for each model in an ensemble. For the ensemble of GB models using the full feature set, the average test set accuracy was 0.905, precision 0.897 and recall 0.923 (Figure 3A). The average test set accuracy, precision and recall for SVM models trained on the reduced feature set were 0.830, 0.820 and 0.857, respectively (Figure 3). The average AUCROC was 0.973 for the GB model ensemble, and 0.899 for the SVM ensemble.

Model predictions and *in silico* validation

What do the models predict?

To evaluate results from different models and feature sets, we followed multiple approaches described in this and the following subsections. First, to get a high-level understanding of the predictions made by our models, we performed exploratory GO enrichment analyses of genes predicted with high confidence to be involved in RNA methylation. Here, we defined as high confidence all genes in the top 1% of the probability distribution for Class 1. For the GB ensemble trained on the full feature set, this comprised the top 269 predictions with an average probability score greater than 0.83. For the SVM models trained on the reduced feature set, 268 genes with a probability of 0.84 or higher were selected.

The top 50 enriched terms for GB and SVM models are shown in Figures 4A and B, respectively. Both model ensembles, independently of the dataset they derived from, yielded predictions enriched in GO terms associated with RNA biogenesis, localization, transport and processing. Note that top enrichment results for GB included additionally terms associated with DNA and protein methylation processes (Figure 4A). This may point to either a lack of specificity of the models with regards to the modification substrate, or a close functional link between RNA and other methylation pathways. Overall, the GO analyses provided a good qualitative control for model performance. The rationale here is that although we did not recover enrichment in the biological term "RNA methylation" *per se* (given that the models predict "novel" genes), features closely associated with the term should figure among the top GO results.

Do the models agree?

Our second analysis aimed to assess the degree of concordance between predictive models trained on the full and reduced feature sets. Figure 5 shows the predicted probability scores of each gene being assigned to Class 1, based on GB models derived from the full feature set versus the average probability obtained by the SVM models trained on the reduced feature set. Overall, the two ensembles yielded very similar predictions, as exemplified by the strong correlation between predicted probability scores ($r = 0.872$, $P < 2.2e-16$). Yet, for certain genes we observed a high degree of discordance between the GB/full and SVM/reduced models.

To further explore these discrepancies, we examined genes predicted to associate with RNA methylation pathways with a probability greater than 0.8 by one ensemble, but that were assigned to the negative class ($P < 0.5$) by the other ensemble. GO analysis of RNA methylation genes only predicted by SVM showed enrichment in the functions of anaphase-promoting complex-dependent catabolic process (P = 2.60E-07), antigen processing and presentation of peptide antigen via MHC class I (P = 7.69E-05), and mitochondrial translational elongation (P = 2.43E-04) among others (Figure 5). Given that gene expression constituted the most informative feature type for classifiers trained on the reduced feature set, it is likely that genes participating in the aforementioned processes exhibit highly similar expression profiles to RNA methylation genes - at least according to transcriptomic resources used here for learning.

On the opposite end of the distribution, considering genes recovered with a high probability score by GB models only, our analyses found significant enrichment in DNA, histone and protein methylation processes, as well as other RNA modification pathways ($P < 0.05$, Figure 5). This may represent a modelling artifact, i.e., predictions erroneously assigned to Class 1, that could be caused by the hierarchical nature of GO terms (e.g., “methylation” being the parent term of both “RNA methylation” and “DNA methylation” processes). An alternative interpretation is that our models capture a functional link between modification pathways operating at different substrates.

In silico validation of gene predictions

Of all classifiers, GB models that were trained on the full feature set showed the best performance based both on cross-validation and hold-out test datasets. We thus selected the top hundred genes predicted by the GB models to associate with RNA methylation pathways as candidates for further validation (Table 5). To evaluate these predictions with respect to previously known RNA methylation genes, we first performed a hierarchical clustering analysis of predicted plus positive (Class 1) genes based on the machine learning data used here (Figure 6). As anticipated, known and predicted genes were well clustered together, with no evident split between known and predicted RNA methylation genes.

Second, we interrogated the STRING database[21] for independent Protein-Protein Interaction (PPI) information on known RNA methylation genes and other genes of the human genome. We built a PPI network based on interactions with a confidence score of 400 or above, and performed Random Walks starting from proteins known to mediate methylation of RNAs (Class 1). This allowed us to weigh all other proteins in the network and rank them by their importance relative to our positive gene set. To evaluate whether genes predicted by our models were highly ranked among important interactors, we performed Gene Set Enrichment Analysis (GSEA) using the PageRank score as an input. We obtained a strong positive enrichment ($NES = 1.605$, $P = 0.0001$) for the model predictions (Table 6), corroborating their close functional association with RNA methylation pathways based on independent PPI evidence (Figure 7).

Insights into the role of new predictions

To gain functional insights into the role of newly predicted genes with regards to previously annotated RNA methyltransferases and associated proteins, we interrogated the STRING database for available PPI data connecting our model predictions to known RNA methylation genes. Our search unravelled a dense network of interactions (Figure 8A), comprising 2,450 edges (confidence ≥ 400). To further dissect these PPI data and identify subgroups of proteins associated with specific pathways, we employed the Louvain method of community detection[22]. We identified six communities in total (Figure 8B), which we annotated using a large collection of functional annotation resources[23].

Community 1 (C1, Figure 8B) groups most RNA methylation genes from the positive set, together with 10 model predictions: *CTU2*, *FARS2*, *HEMK1*, *KARS*, *MOCS3*, *MTO1*, *N6AMT1*, *PUS1*, *PUS3* and *TRNT1*. Functional analysis of community members showed that proteins comprising this sub-network are significantly enriched in the functions of tRNA modification (GO:0006400, P = 5.09E-70), tRNA methylation (GO:0030488, P = 6.31E-66), and tRNA processing (Reactome R-HSA-72306, P = 4.10E-45). Indeed, four predictions in the cluster, CTU2, MOCS3, PUS1 and PUS3, are RNA modifying enzymes mediating tRNA modifications. CTU2 and MOCS3 are involved in 2-thiolation of mcm^{5}S^{2}U at wobble positions of tRNAs, whereas PUS1 and PUS3 belong to the tRNA pseudouridine synthase TruA family and mediate the formation of pseudouridine at positions 27/28 and 38/39 of certain tRNAs, respectively[13]. Among other members of the same community, the gene *TRNT1* encodes the mitochondrial CCA tRNA nucleotidyltransferase 1 responsible for the addition of the conserved 3'-CCA sequence to tRNAs. It has been previously reported that the presence of the 3'-CCA tail on tRNA is required for target recognition by the tRNA methyltransferase NSUN6[24], which could underlie the functional link of TRNT1 with RNA methylation genes in our analyses.

Likewise, two aminoacyl-tRNA synthetases, FARS2 and KARS, were also predicted to be closely associated with RNA methylation pathways and were part of Community 1. FARS2 is a mitochondrial Phenylalanine-tRNA ligase, responsible for the charging of tRNA(Phe) with phenylalanine. *KARS* encodes a Lysin-tRNA ligase. Although, we have not found any orthogonal evidence linking FARS2 to RNA methylation, KARS has been previously inferred to physically interact with the RNA methyltransferase TRMT1, based on co-fractionation data (source BioGRID[25]).

The same sub-network also included two HemK methyltransferases, HEMK1 and N6AMT1. The former is a N5-glutamine methyltransferase responsible for the methylation of the glutamine residue in the GGQ motif of the mitochondrial translation release factor MTRF1L[26]. N6AMT1 methylates the eukaryotic translation termination factor 1 (eRF1) on Gln-185. Notably, it has been reported that N6AMT1 forms the catalytic subunit of a heterodimer with the RNA methyltransferase TRMT112[27], suggestive of a functional interplay between RNA methylation and post-translational modifications of translation factors.

Our models also predicted that *MTO1* is a gene functionally associated with RNA methylation pathways. Previous studies have shown that *MTO1* encodes for a mitochondrial protein which is indeed involved in the 5-carboxymethylaminomethyl modification (mnm^{5}s^{2}U34) of the wobble uridine base in mitochondrial tRNAs, with a crucial role in translation fidelity[28].

Community 2 (C2, Figure 8B) consists mainly of newly predicted genes, associated with four genes from the positive set: *C7orf60*, *HENMT1*, *RRNAD1* and *RSAD1*. The gene *C7orf60* or *BMT2* encodes a probable S-adenosyl-L-methionine-dependent methyltransferase. Recent studies have suggested that BMT2 (also known as SAMTOR) acts as an inhibitor of mTOR complex 1 (mTORC1) signalling in human, a SAM sensor signalling methionine sufficiency[29]. In yeast, BMT2 is responsible for the m^{1}A2142 modification of 25S rRNA[30]. Two other methyltransferase genes in the same cluster were *RRNAD1* and *HENMT1*. The former encodes for ribosomal RNA adenine dimethylase domain containing 1, but little is known about its function. HENMT1 is a small RNA methyltransferase that adds a 2'-O-methyl group at the 3'-end of piRNAs, contributing to the maintenance of Transposable Element (TE) repression in adult germ cells[31]. Functional annotation of this community indicated an enrichment in peptidyl-lysine methylation function (GO:0018022, P = 1.92E-06), albeit this was based on only four proteins out the 23 forming this cluster (SETD4, VCPKMT, METTL21A, and METTL18). Among members of this community, we identified proteins with a role in methylation of other substrates. For example, FAM86A catalyses the trimethylation of the elongation factor 2 (eEF2) at Lys-525[32]. METTL13 is also a methyltransferase responsible for the dual post-translational methylation of the elongation factor 1-alpha (eEF1A) at two positions (Gly-2 and Lys-55), modulating mRNA translation in a codon-specific manner[33]. Both genes are involved in modifying translation elongation factor residues, same as N6AMT1 mentioned above. Our results hence suggest that post-translational modifications of translation factors and epitranscriptomic changes on RNAs could be interconnected in modulating translational efficiency.

Community 3 (C3, Figure 8B) comprises 48 protein members, of which 10 are part of our positive set and 38 were predicted by the models. Overall, we found a strong enrichment for functional terms linked to ncRNA processing (GO:0034470, P = 6.79E-40) and rRNA processing (R-HSA-72312, P = 1.03E-39). For example, among Community 3 members, our predictions include five genes encoding for members of the nuclear RNA exosome, *DIS3*, *EXOSC2*, *EXOSC5*, *EXOSC8* and *EXOSC9*. The exosome is known to participate in a wide variety of cellular RNA processing and degradation events preventing nuclear export and/or translation of aberrant RNAs. Exosome function is thus likely to be interlinked with epitranscriptomic marks on RNAs.

We also identified a sub-cluster within the community connecting DIMT1, EMG1, FBL and NOP2 with 15 proteins predicted by our models. All members of the sub-cluster are RNA-binding proteins involved in rRNA modification in the nucleus (R-HSA-6790901, P = 5.44E-36). *EMG1* encodes for an RNA methyltransferase that methylates pseudouridine at position 1248 in 18S rRNA[34]. Pathway annotation data further suggest that EMG1 together with eight new predictions (CIRH1A, DCAF13, HEATR1, NOL11, UTP3, UTP6, UTP20 and WDR3) are required in pre-18S rRNA processing and ribosome biogenesis. Of these, the *NOL11* gene encodes a nucleolar protein contributing to pre-rRNA transcription and processing[35]. Partial evidence furthermore suggests that NOL11 interacts with the rRNA 2’-O-methyltransferase fibrillarin, FBL, which is involved in pre-rRNA processing by catalysing the site-specific 2'-hydroxyl methylation of pre-ribosomal RNAs[35]. FBL together with RRP9 and NOP56 are part of the box C/D RNP complex catalysing the ribose-2'-O-methylation of target RNAs.

Finally, three novel gene predictions within this community, *DPH5*, *TPMT* and *RRP8*, were previously reported to have SAM-dependent methyltransferase activity. *DPH5* is coding for a methyltransferase that catalyses the tri-methylation of the eEF2 as part of the diphthamide biosynthesis pathway, whereas *TPMT* encodes an enzyme that metabolizes thiopurine drugs. We cannot rule out that these may be false positives cases, i.e., erroneous predictions that stem from the presence of the SAM-binding domain in the protein. Yet genes mediating post-translational modifications were repeatedly classified as components of RNA methylation pathways by our machine learning models (e.g., *FAM86A* in Community 2). A noteworthy case is RRP8, which in human is reported to bind to H3K9me2 and to probably act as a methyltransferase, yet studies in yeast have shown that the RRP8 homologue is responsible for installing m1A in the peptidyl transfer centre of the ribosome (m^1A645 in 25S)[36].

Community 4 (C4, Figure 8B) constitutes a large cluster of 42 proteins. Functional analysis of the group indicates that most community members are chromatin modifying enzymes (R-HSA-3247509, P = 8.74E-29), or are associated in general with chromatin organization (R-HSA-4839726, P = 8.74E-29) and histone modification (WP2369, P = 1.08E-23). Previously known RNA methylation genes in this community were mainly involved in RNA-capping pathways, e.g., *RNMT*, *CMTR1*, *CMTR2*, *FAM103A1*, *TGS1* and *RNGTT*. Recent studies have suggested that there is indeed extensive crosstalk between RNA modifications and epigenetic mechanisms of gene regulation[7,37,38].

Community 5 (C5) and Community 6 (C6) encompass fewer members than the other communities. Community 5 consists of 10 proteins creating a small sub-network of RNA methyltransferases and partner proteins involved in RNA methylation (GO:0001510, P = 1.91E-17) and mRNA methylation, in particular (GO:0080009, P = 6.26E-16). Notably, this community captures proteins involved in the m6A pathway, including the m^{6}A writer complex of METTL3-METTL14 with co-factor WTAP, METTL16 and ZC3H13, as well as the m^{6}Am writer METTL4[39]. Community 6 is the smallest of all communities with only four protein members, two previously annotated RNA methylation genes, *HSD17B10* and *KIAA0391*, and two predicted genes *POP1* and *POP4*. Functional analysis suggests that all four proteins contribute to tRNA processing (R-HSA-72306, P = 5.97E-09) and three of them are involved in tRNA 5'-end processing (GO:0099116, P = 5.32E-08). The *HSD17B10* gene encodes the 3-hydroxyacyl-CoA dehydrogenase type-2, which is involved in mitochondrial fatty acid beta-oxidation. *HSD17B10* is involved in tRNA processing as it also forms a subcomplex of the mitochondrial ribonuclease P together with TRMT10C/MRPP1[40]. This subcomplex, named MRPP1-MRPP2, catalyses the formation of N1-methylguanine and N1-methyladenine at position 9 (m^{1}G9 and m^{1}A9, respectively) in tRNAs. *KIAA0391*, also known as *PRORP*, encodes a catalytic ribonuclease component of mitochondrial ribonuclease P. It appears that POP1 and POP2 are also components of ribonuclease P and contribute to tRNA maturation via 5'-end cleavage.

Potential drawbacks

Our machine learning models and analyses have provided a wealth of new information on putative gene networks underpinning RNA methylation in human. However, it is worth noting the limitations of our approach. First, because only few writer enzymes are to date known to deposit methyl-marks on RNA[6], we started from a very limited number of positive (and by consequence negative) samples to use for machine learning. Even though model performance based on test data was good, the small sample sizes may have hampered how well our models generalise. In addition, our models overpredicted genes associated with RNA methylation pathways, as a large number of genes obtained a high probability score for Class 1. This is because we followed a modelling approach using balanced positive and negative classes to optimise model performance.

Second, it is uncertain whether employing previous knowledge from functional annotations may have biased model predictions. We addressed this caveat to an extent by using a reduced feature set without annotation features, such as GO terms. When looking at predictions based on models trained on this dataset, we identified genes previously known to be involved in cell differentiation, G2/M cell cycle, antigen presentation and mitochondrial translation ($P < 0.05$, Figure 5). Even based on this unbiased set of classifiers, machine learning models point to a recurrent theme of this study: that RNA methylation is functionally interconnected to a range of other core cellular functions. For example, we repeatedly found genes encoding protein methyltransferases among the top model predictions. The key question here is whether these genes represent false positives, spurred by the hierarchical structure of GO terms or the shared SAM binding domain. These ambiguous predictions should be interpreted with caution, although multiple lines of evidence suggest that this could well be a biologically meaningful result echoing the crosstalk between DNA, RNA and post-transcriptional modification processes.

CONCLUSIONS

RNA methylation is a key modulator of transcript stability, splicing and translation efficiency, playing a critical role in cellular homeostasis and disease[4]. Yet, its molecular underpinnings remain to date poorly understood[11]. Here, we aimed to gain novel insights into genes associated with RNA methylation pathways in human using machine learning approaches. Specifically, we analysed available transcriptomic, proteomic, structural and protein-protein interaction data in a supervised machine learning framework.

Our machine learning models showed very good performance on unseen test data, reaching high accuracy (91%), precision (90%) and recall (92%). *A priori* gene knowledge (e.g., GO annotations) together with expression data constituted the most informative data types in predictive modelling. Notably, in certain tissues, such as blood, heart, pancreas and brain, genes mediating RNA methylation seemed to show an up- or down-regulated expression profile.

Using independent PPI data, we orthogonally validated top model predictions by corroborating close functional links to previously known RNA methylation genes. Community detection delineated six molecular subnetworks, with distinct roles in tRNA processing (C1, C6), rRNA processing (C3), mRNA methylation (C5), but also protein (C2) and chromatin modifications (C4). Network analyses suggested that deposition of methyl marks on tRNAs is co-orchestrated with other modification processes, such as 2-thiolation and pseudouridine formation. Similarly, rRNA methyltransferases appeared functionally linked to several genes involved in rRNA processing and ribosomal biogenesis. Intriguingly, RNA-capping enzymes were clustered with chromatin modifiers, raising the hypothesis of a crosstalk between the two processes. Our results further indicate that post-translational modifications of translation factors and epitranscriptomic changes on RNAs are intertwined in modulating translational efficiency. Overall, our study exemplifies how access to omics datasets joined by machine learning methods can be used to infer molecular pathways and novel gene function.

METHODS

Dataset assembly and pre-processing

To assemble a machine learning dataset for predicting genes involved in RNA methylation process in the human genome, we first curated a list of previously known RNA methylation genes. For this, we performed searches in standard functional annotation resources, such as ExPASy ENZYME (https://enzyme.expasy.org/), InterPro (https://www.ebi.ac.uk/interpro/) and the GO Resource (http://geneontology.org/), in conjunction with a comprehensive literature review for annotated RNA methyltransferases following up on the pioneering paper of Schapira[6]. This allowed us to identify 92 proteins involved – or putatively involved – in RNA methylation to use for machine learning modelling (Table 1).

To obtain informative features for classifying gene functions, we interrogated the Harmonizome database[15]. Harmonizome provides a large collection of the pre-processed datasets for genes and proteins, with ~72 million attributes (functional associations) from over 70 major online resources. We selected 15 one-hot-encoded datasets from four broad categories: (i) transcriptomics; (ii) proteomics; (iii) structural or functional annotations; and (iv) physical interactions (Table 2). In particular, from omics experiments, we sampled BioGPS[16], GTEx[18], HPA[19] and TISSUES[20] gene and protein expression profile data. From functional datasets, we considered GO annotations and InterPro structural domains. Finally, from physical interactions datasets, we selected KEGG and Reactome Pathways, as well as Hub Proteins and Pathway Commons. Collating these data yielded an initial matrix of 26,935 genes and 50,176 one-hot-encoded features ("full feature set"). In addition, we compiled a second dataset of reduced dimensionality, by excluding all 5,148 GO and InterPro annotation features ("reduced feature set").

Problem framing, model definition, training and evaluation

To estimate the probability of a gene being associated with RNA methylation, we used standard machine learning approaches for binary classification. We labelled the 92 previously known RNA methylation genes as positive samples (Class 1), and split them into two sets comprising: (i) 80% of the data for training and cross-validation (n=74) and (ii) 20% kept unseen for model testing (n=18). We considered the remaining genes of the human genome as negative samples (Class 0) and performed an analogous 80/20 split into training/cross-validation (n=21,476) and test sets (n=5,368). The underlying assumption here is that the vast majority of genes in the human genome serve other functions, thus the number of false negatives in the training data should be very small.

To produce balanced sets of training samples, and to later reduce the variance of our final models through averaging, negative genes kept for training (n=21,476) were further divided into sets of 74 – equal to the number of positive samples for training. We thus generated 290 training sets, where the positive class remained fixed and the negative class was represented by a random draw of an equal number of genes from the rest of the genome, sampling each gene once.

Starting with 290 training sets and our unprocessed Harmonizome data comprising 50,176 features, we next performed filtering to remove low-information features. We removed features with (i) zero values in more than 70% of the samples in each training set, or (ii) less than 16% variance in at least one training set. The selected features for each of the 290 training sets were then merged into a final list of features for model training and testing. We followed the exact same selection process for the reduced feature set as well.

We next considered five types of machine learning models for binary classification: Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting (GB) models. We used grid search and 3-fold cross-validation on each training set for the SVM hyperparameter tuning of the kernel function (linear or RBF), cost parameter, and kernel bandwidth (RBF kernel only). For RF, we used grid search to determine the optimal number of trees in the forest, followed by a randomized search to select the best parameters for maximum number of features considered for splitting a node, maximum number of levels in each decision tree, minimum number of data points placed in a node before the node is split, and minimum number of data points allowed in a leaf node. Likewise, for the GB model, we performed grid search to optimise the learning rate and number of trees in the forest, and subsequently performed a randomized search to tune the remaining decision tree parameters (see RF). We trained all five predictive models on each of the training sets from the full and reduced feature sets, respectively. The performance of all classifiers was estimated using 10-fold cross-validation, i.e., the dataset was split into 10 folds, of which nine were used for the training process and one for testing. The process was repeated ten times, and model performance was estimated using standard performance metrics: accuracy, precision, recall (sensitivity), F1 score and Area Under the Receiver Operating Characteristic Curve (AUROC), averaged across the ten repeats. Finally, we used GB feature ranking to determine the top 100 most informative features across the ensemble of training sets for the full and reduced feature sets, respectively.

Final model testing on test dataset and genome-wide prediction

Once the best classifiers for the full and reduced datasets were selected based on cross-validation, we tested the performance of the model ensembles on unseen data. Analogous to the procedure described above for training data, we generated 298 testing datasets, by splitting the negative genes kept for testing into equal sets of 18 genes, and combining them with the 18 of positive samples previously retained. Each model from the classifier ensemble was evaluated on each of the test datasets using accuracy, precision, recall, F1 score and AUROC. Overall performance was calculated by averaging results of all models across test sets.

Likewise, the prediction probability of each human gene was calculated by averaging probability scores for Class 1 across all models of the best ensemble for the full and reduced feature sets, respectively. Most non-Class 1 genes (all except the test cases) were part of the negative samples in the training data of exactly one model in the ensemble; however, due to the high number of models (290) the effects of this on the final predictions is expected to be negligible.

All visualisations and meta-analyses were performed using the R software environment (v. 4.0.5)[41]. A heatmap of known and predicted RNA methylation genes across all features used for machine learning was generated using the R package pheatmap. Further *in silico* validation of model predictions was performed using GO enrichment analyses of predicted genes within the domain “Biological Process” using the package clusterProfiler[42]. Protein-Protein Interaction (PPI) data for human were obtained from STRING (v.11.0)[21] and filtered to interactions with a combined score of 400 and above. All network analyses were performed using the igraph R package[43]. Functional annotation of PPI communities was performed using EnrichR[23].

ACKNOWLEDGEMENTS

The authors are thankful to Adrián Rodríguez Bazaga for his valuable input on the machine learning analyses, and Woochang Hwang for his feedback on the network analyses.

COMPETING INTERESTS

GT, DL, OR and HW are employees of Storm Therapeutics. TK is a co-founder of Abcam and Storm Therapeutics.

REFERENCES

1. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200 (2017).
2. Boccaletto, P. *et al.* MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
3. Barbieri, I. & Kouzarides, T. Role of RNA modifications in cancer. *Nat. Rev. Cancer* 1–20 (2020) doi:10.1038/s41568-020-0253-2.
4. Huang, H., Weng, H., Deng, X. & Chen, J. RNA Modifications in Cancer: Functions, mechanisms, and therapeutic implications. *Annu. Rev. Cancer Biol.* **4**, 221–240 (2020).
5. Delatte, B. *et al.* Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* **351**, 282–285 (2016).
6. Schapira, M. Structural chemistry of human RNA methyltransferases. *ACS Chem. Biol.* **11**, 575–582 (2016).
7. Tzelepis, K., Rausch, O. & Kouzarides, T. RNA-modifying enzymes and their function in a chromatin context. *Nat. Struct. Mol. Biol.* **26**, 858–862 (2019).
8. Copeland, R. A., Olhava, E. J. & Scott, M. P. Targeting epigenetic enzymes for drug discovery. *Curr. Opin. Chem. Biol.* **14**, 505–510 (2010).
9. Shi, H., Chai, P., Jia, R. & Fan, X. Novel insight into the regulatory roles of diverse RNA modifications: Re-defining the bridge between transcription and translation. *Mol. Cancer* **19**, 78 (2020).
10. Chou, H.-J., Donnard, E., Gustafsson, H. T., Garber, M. & Rando, O. J. Transcriptome-wide Analysis of Roles for tRNA Modifications in Translational Regulation. *Mol. Cell* (2017) doi:10.1016/j.molcel.2017.11.002.
11. Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? *Nat. Rev. Genet.* **17**, 365–372 (2016).
12. Jonkhout, N. *et al.* The RNA modification landscape in human disease. *RNA* **23**, 1754–1769 (2017).
13. de Crécy-Lagard, V. *et al.* Matching tRNA modifications in humans to their known and predicted enzymes. *Nucleic Acids Res.* **47**, 2143–2159 (2019).
14. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, 35 (2017).
15. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, (2016).
16. Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10**, R130 (2009).
17. The Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
18. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
19. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, (2015).
20. Palasca, O., Santos, A., Stolte, C., Gorodkin, J. & Jensen, L. J. TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database* **2018**, (2018).

21. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
22. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
23. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
24. Haag, S. *et al.* NSUN6 is a human RNA methyltransferase that catalyzes formation of m5C72 in specific tRNAs. *RNA N. Y. N* **21**, 1532–1543 (2015).
25. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
26. Ishizawa, T., Nozaki, Y., Ueda, T. & Takeuchi, N. The human mitochondrial translation release factor HMRF1L is methylated in the GGQ motif by the methyltransferase HMPrmC. *Biochem. Biophys. Res. Commun.* **373**, 99–103 (2008).
27. Li, W., Shi, Y., Zhang, T., Ye, J. & Ding, J. Structural insight into human N6amt1–Trm112 complex functioning as a protein methyltransferase. *Cell Discov.* **5**, 1–13 (2019).
28. Tischner, C. *et al.* MTO1 mediates tissue specificity of OXPHOS defects via tRNA modification and translation optimization, which can be bypassed by dietary intervention. *Hum. Mol. Genet.* **24**, 2247–2266 (2015).
29. Gu, X. *et al.* SAMTOR is an S-adenosylmethionine sensor for the mTORC1 pathway. *Science* **358**, 813–818 (2017).
30. Sharma, S., Watzinger, P., Kötter, P. & Entian, K.-D. Identification of a novel methyltransferase, Bmt2, responsible for the N-1-methyl-adenosine base modification of 25S rRNA in Saccharomyces cerevisiae. *Nucleic Acids Res.* **41**, 5428–5443 (2013).
31. Lim, S. L. *et al.* HENMT1 and piRNA stability are required for adult male germ cell transposon repression and to define the spermatogenic program in the mouse. *PLOS Genet.* **11**, e1005620 (2015).
32. Davydova, E. *et al.* Identification and characterization of a novel evolutionarily conserved lysine-specific methyltransferase targeting eukaryotic translation elongation factor 2 (eEF2) *. *J. Biol. Chem.* **289**, 30499–30510 (2014).
33. Jakobsson, M. E. *et al.* The dual methyltransferase METTL13 targets N terminus and Lys55 of eEF1A and modulates codon-specific translation rates. *Nat. Commun.* **9**, 1–15 (2018).
34. Meyer, B. *et al.* The Bowen–Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Ψ1191 in yeast 18S rRNA. *Nucleic Acids Res.* **39**, 1526–1537 (2011).
35. Freed, E. F., Prieto, J.-L., McCann, K. L., McStay, B. & Baserga, S. J. NOL11, implicated in the pathogenesis of North American Indian childhood cirrhosis, is required for pre-rRNA transcription and processing. *PLOS Genet.* **8**, e1002892 (2012).
36. Shima, H. & Igarashi, K. N1-methyladenosine (m1A) RNA modification: the key to ribosome control. *J. Biochem. (Tokyo)* (2020) doi:10.1093/jb/mvaa026.
37. Kan, R. L., Chen, J. & Sallam, T. Crosstalk between epitranscriptomic and epigenetic mechanisms in gene regulation. *Trends Genet.* **0**, (2021).
38. Huang, H. *et al.* Histone H3 trimethylation at lysine 36 guides m6A RNA modification co-transcriptionally. *Nature* **567**, 414–419 (2019).
39. Chen, H. *et al.* METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing. *Cell Res.* **30**, 544–547 (2020).

40. Vilardo, E. *et al.* A subcomplex of human mitochondrial RNase P is a bifunctional methyltransferase—extensive moonlighting in mitochondrial tRNA biogenesis. *Nucleic Acids Res.* **40**, 11583–11593 (2012).
41. R Core Team. *R: A Language and environment for statistical computing*. (R Foundation for Statistical Computing, 2019).
42. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
43. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).

TABLES

Table 1. Known RNA methyltransferases and related proteins used as positive set (Class 1).

Table 2. Gene-feature omics datasets used in machine learning analyses (source Harmonizome).

Table 3. Highly informative features based on models trained on the reduced feature set, and their frequency in the top100 features across all models of the classifier ensemble.

Table 4. Model performance based on 10-fold cross-validation.

Table 5. Top 100 gene predictions based on the GB model ensemble of the full feature set.

Table 6. Personalised PageRank score of top 100 model predictions based on PPI data (source: STRING).

FIGURES

Figure 1. Schematic representation of the analysis workflow. Previously known RNA methylation genes were used as positive samples (Class 1) and split into two sets comprising 80% of the data for training and 20% kept unseen for model testing. An analogous 80/20 split was performed for the remaining genes of the human genome, which were further divided into sets of equal size to the positive samples and used as negative samples (Class 0) to generate stratified sets for training and testing. Following feature pre-filtering, five types of machine learning models for binary classification - Logistic Regression (LR), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting (GB) - were trained on each of the training sets resulting in a classifier ensemble. Each model from the classifier ensemble was evaluated on each of the test datasets and overall performance was calculated by averaging results of all models across test sets. The best-performing ensemble was used to make predictions for the whole genome.

Figure 2. Feature importance. Top 50 most informative features ranked by their relative importance in predictive modelling based on the **A.** full and **B**. reduced feature sets.

Figure 3. Model performance based on test data. Accuracy, precision, recall and AUC score distributions as estimated across test datasets for the best model ensembles: **A.** GB models for the full feature set; and **B.** SVM models for the reduced feature set.

Figure 4. Functional enrichment analyses of high-confidence predictions. GO enrichment analysis of all genes in the top 1% of the probability distribution for Class 1 based on **A.** GB models, full feature set and **B.** SVM models, reduced feature set. Top enriched terms include functions such as RNA biogenesis, localization, transport, and processing. For GB predictions, additional functions were associated with DNA and protein methylation processes.

Figure 5. Concordance between predictive models. Middle panel: Scatterplot of the predicted probability score of each gene being assigned to Class 1, based on GB models trained on the full feature set versus SVM models trained on the reduced feature set. Side panels: Top 15 enriched GO terms associated with genes assigned to Class 1 with a probability greater than 0.8 by one ensemble only (right: SVM models only; left: GB models only). Enriched terms are represented as a network with edges connecting overlapping gene sets.

Figure 6. Heatmap of predicted and known RNA methylation genes. Hierarchical clustering analysis of predicted plus positive genes shows no evident split between predictions (yellow) and known RNA methylation genes (green). Features (columns) used for machine learning are shown in different colours based on the data source.

Figure 7. GSEA analysis of model predictions based on PageRank score. Personalised PageRank score of all human genes was computed using PPI data from STRING, starting from previously known RNA methylation genes. A strong positive enrichment (NES = 1.605, P = 0.0001) was obtained for model predictions, corroborating a close functional association with RNA methylation pathways.

Figure 8. PPI network of known and predicted genes involved in RNA pathways. A. Network based on available PPI data connecting newly predicted genes with previously annotated RNA methyltransferases and associated proteins. **B.** Subgroups of proteins associated with specific pathways, as inferred using the Louvain method of community detection.

Figure 1

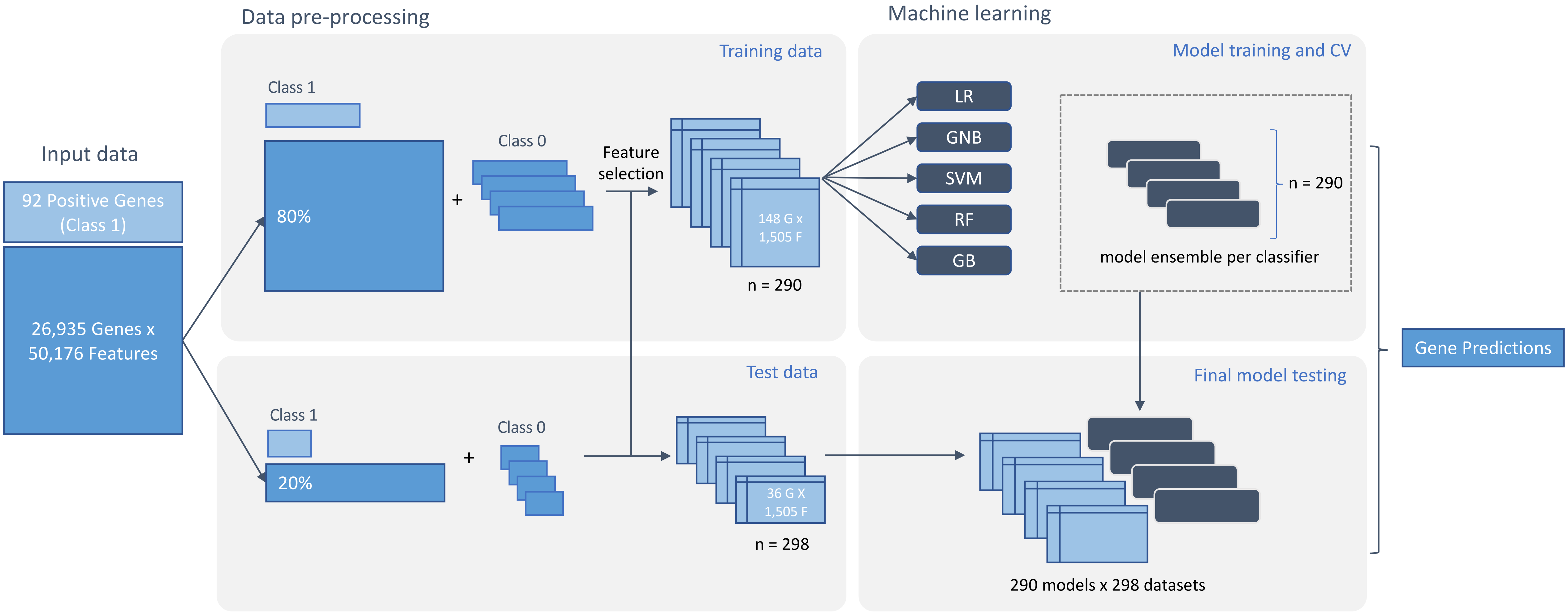
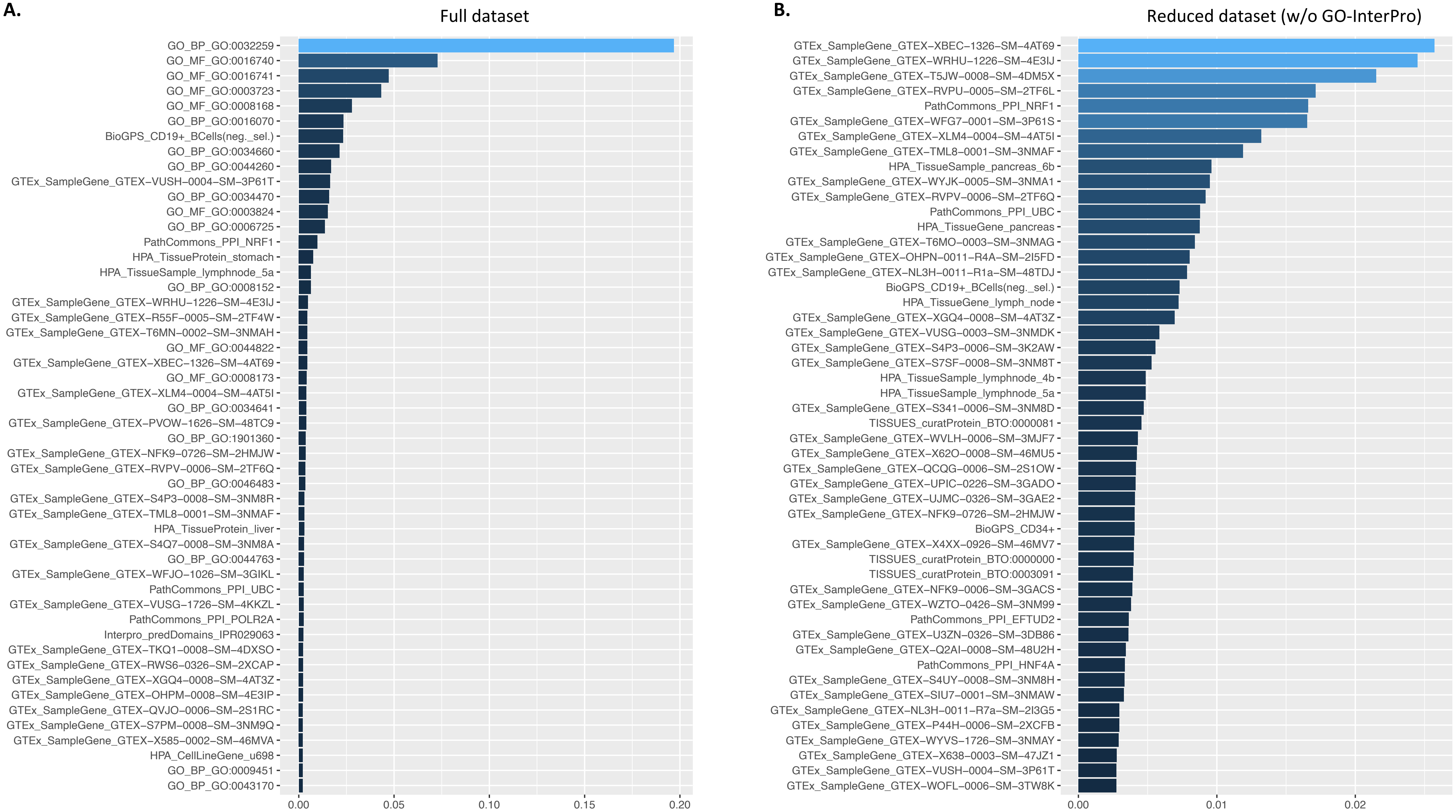

Input data
92 Positive Genes (Class 1)
26,935 Genes x 50,176 Features
Data pre-processing
Training data
Class 1
80%
+
Class 0
Feature selection
148 G x 1,505 F
n = 290
Test data
Class 1
20%
+
Class 0
36 G X 1,505 F
n = 298
Machine learning
Model training and CV
LR
GNB
SVM
RF
GB
n = 290
model ensemble per classifier
Final model testing
290 models x 298 datasets
Gene Predictions

Figure 2



A.
Full dataset
GO_BP_GO:0032259
GO_MF_GO:0016740
GO_MF_GO:0016741
GO_MF_GO:0003723
GO_MF_GO:0008168
GO_BP_GO:0016070
BioGPS_CD19+_BCells(neg._sel.)
GO_BP_GO:0034660
GO_BP_GO:0044260
GTEx_SampleGene_GTEX−VUSH−0004−SM−3P61T
GO_BP_GO:0034470
GO_MF_GO:0003824
GO_BP_GO:0006725
PathCommons_PPI_NRF1
HPA_TissueProtein_stomach
HPA_TissueSample_lymphnode_5a
GO_BP_GO:0008152
GTEx_SampleGene_GTEX−WRHU−1226−SM−4E3IJ
GTEx_SampleGene_GTEX−R55F−0005−SM−2TF4W
GTEx_SampleGene_GTEX−T6MN−0002−SM−3NMAH
GO_MF_GO:0044822
GTEx_SampleGene_GTEX−XBEC−1326−SM−4AT69
GO_MF_GO:0008173
GTEx_SampleGene_GTEX−XLM4−0004−SM−4AT5I
GO_BP_GO:0034641
GTEx_SampleGene_GTEX−PVOW−1626−SM−48TC9
GO_BP_GO:1901360
GTEx_SampleGene_GTEX−NFK9−0726−SM−2HMJW
GTEx_SampleGene_GTEX−RVPV−0006−SM−2TF6Q
GO_BP_GO:0046483
GTEx_SampleGene_GTEX−S4P3−0008−SM−3NM8R
GTEx_SampleGene_GTEX−TML8−0001−SM−3NMAF
HPA_TissueProtein_liver
GTEx_SampleGene_GTEX−S4Q7−0008−SM−3NM8A
GO_BP_GO:0044763
GTEx_SampleGene_GTEX−WFJO−1026−SM−3GIKL
PathCommons_PPI_UBC
GTEx_SampleGene_GTEX−VUSG−1726−SM−4KKZL
PathCommons_PPI_POLR2A
Interpro_predDomains_IPR029063
GTEx_SampleGene_GTEX−TKQ1−0008−SM−4DXSO
GTEx_SampleGene_GTEX−RWS6−0326−SM−2XCAP
GTEx_SampleGene_GTEX−XGQ4−0008−SM−4AT3Z
GTEx_SampleGene_GTEX−OHPM−0008−SM−4E3IP
GTEx_SampleGene_GTEX−QVJO−0006−SM−2S1RC
GTEx_SampleGene_GTEX−S7PM−0008−SM−3NM9Q
GTEx_SampleGene_GTEX−X585−0002−SM−46MVA
HPA_CellLineGene_u698
GO_BP_GO:0009451
GO_BP_GO:0043170
0.00
0.05
0.10
0.15
0.20
B.
Reduced dataset (w/o GO-InterPro)
GTEx_SampleGene_GTEX−XBEC−1326−SM−4AT69
GTEx_SampleGene_GTEX−WRHU−1226−SM−4E3IJ
GTEx_SampleGene_GTEX−T5JW−0008−SM−4DM5X
GTEx_SampleGene_GTEX−RVPU−0005−SM−2TF6L
PathCommons_PPI_NRF1
GTEx_SampleGene_GTEX−WFG7−0001−SM−3P61S
GTEx_SampleGene_GTEX−XLM4−0004−SM−4AT5I
GTEx_SampleGene_GTEX−TML8−0001−SM−3NMAF
HPA_TissueSample_pancreas_6b
GTEx_SampleGene_GTEX−WYJK−0005−SM−3NMA1
GTEx_SampleGene_GTEX−RVPV−0006−SM−2TF6Q
PathCommons_PPI_UBC
HPA_TissueGene_pancreas
GTEx_SampleGene_GTEX−T6MO−0003−SM−3NMAG
GTEx_SampleGene_GTEX−OHPN−0011−R4A−SM−2I5FD
GTEx_SampleGene_GTEX−NL3H−0011−R1a−SM−48TDJ
BioGPS_CD19+_BCells(neg._sel.)
HPA_TissueGene_lymph_node
GTEx_SampleGene_GTEX−XGQ4−0008−SM−4AT3Z
GTEx_SampleGene_GTEX−VUSG−0003−SM−3NMDK
GTEx_SampleGene_GTEX−S4P3−0006−SM−3K2AW
GTEx_SampleGene_GTEX−S7SF−0008−SM−3NM8T
HPA_TissueSample_lymphnode_4b
HPA_TissueSample_lymphnode_5a
GTEx_SampleGene_GTEX−S341−0006−SM−3NM8D
TISSUES_curatProtein_BTO:0000081
GTEx_SampleGene_GTEX−WVLH−0006−SM−3MJF7
GTEx_SampleGene_GTEX−X62O−0008−SM−46MU5
GTEx_SampleGene_GTEX−QCQG−0006−SM−2S1OW
GTEx_SampleGene_GTEX−UPIC−0226−SM−3GADO
GTEx_SampleGene_GTEX−UJMC−0326−SM−3GAE2
GTEx_SampleGene_GTEX−NFK9−0726−SM−2HMJW
BioGPS_CD34+
GTEx_SampleGene_GTEX−X4XX−0926−SM−46MV7
TISSUES_curatProtein_BTO:0000000
TISSUES_curatProtein_BTO:0003091
GTEx_SampleGene_GTEX−NFK9−0006−SM−3GACS
GTEx_SampleGene_GTEX−WZTO−0426−SM−3NM99
PathCommons_PPI_EFTUD2
GTEx_SampleGene_GTEX−U3ZN−0326−SM−3DB86
GTEx_SampleGene_GTEX−Q2AI−0008−SM−48U2H
PathCommons_PPI_HNF4A
GTEx_SampleGene_GTEX−S4UY−0008−SM−3NM8H
GTEx_SampleGene_GTEX−SIU7−0001−SM−3NMAW
GTEx_SampleGene_GTEX−NL3H−0011−R7a−SM−2I3G5
GTEx_SampleGene_GTEX−P44H−0006−SM−2XCFB
GTEx_SampleGene_GTEX−WYVS−1726−SM−3NMAY
GTEx_SampleGene_GTEX−X638−0003−SM−47JZ1
GTEx_SampleGene_GTEX−VUSH−0004−SM−3P61T
GTEx_SampleGene_GTEX−WOFL−0006−SM−3TW8K
0.00
0.01
0.02

Figure 3

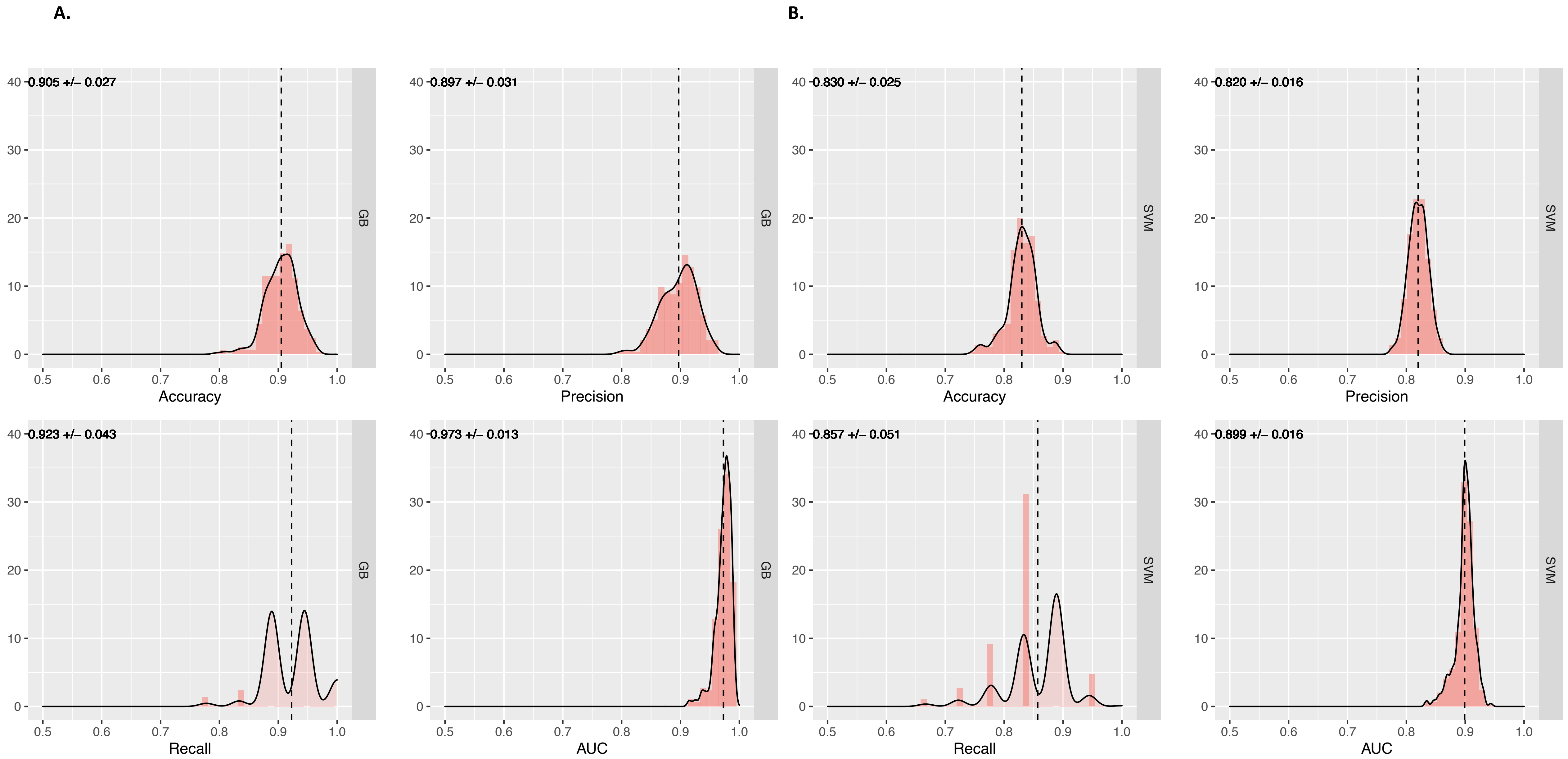

A.
B.
0.905 +/– 0.027
0.897 +/– 0.031
0.830 +/– 0.025
0.820 +/– 0.016
0.923 +/– 0.043
0.973 +/– 0.013
0.857 +/– 0.051
0.899 +/– 0.016
GB
SVM
40
30
20
10
0
0.5
0.6
0.7
0.8
0.9
1.0
Accuracy
Precision
Recall
AUC

Figure 4

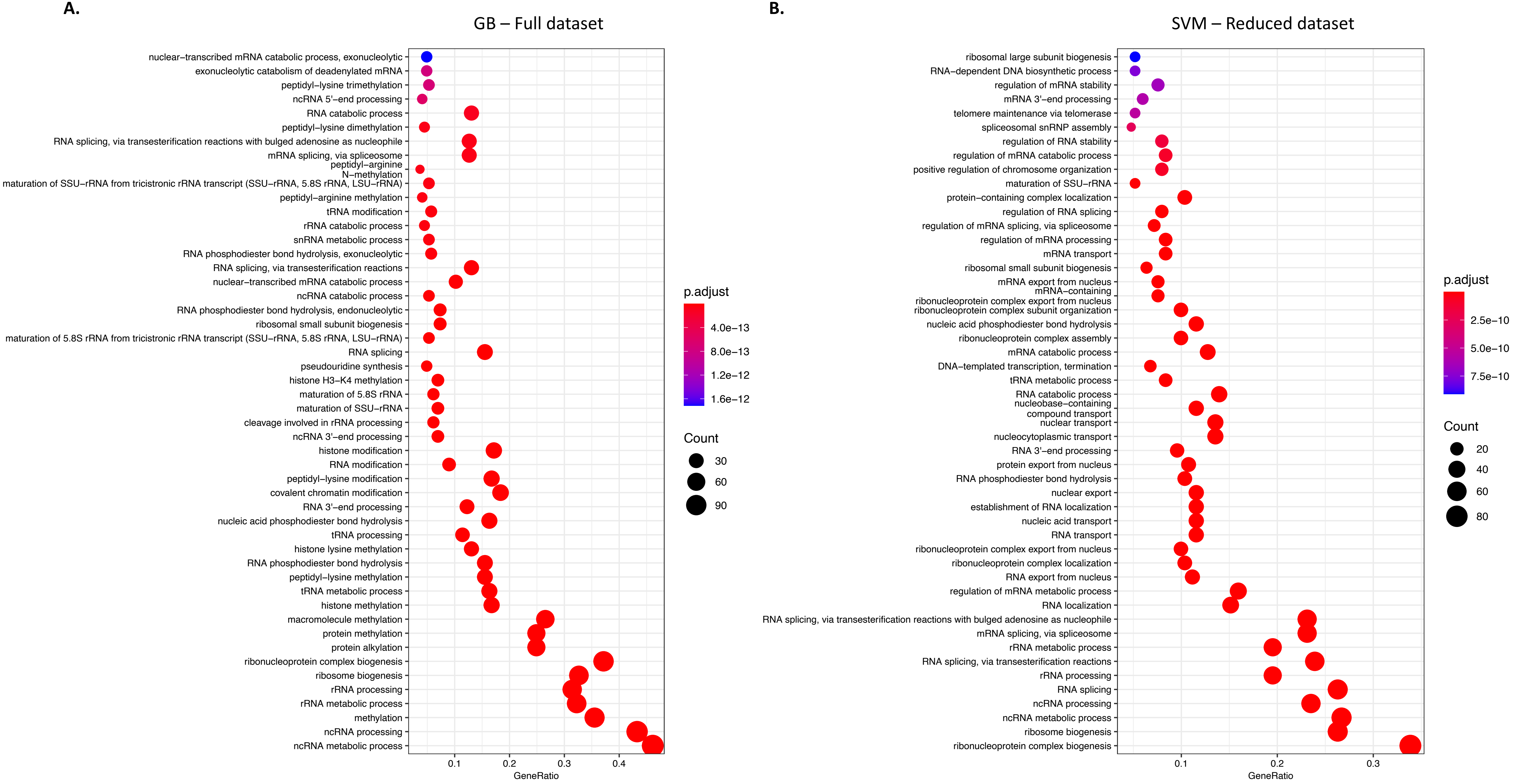

A.
GB – Full dataset
nuclear−transcribed mRNA catabolic process, exonucleolytic
exonucleolytic catabolism of deadenylated mRNA
peptidyl−lysine trimethylation
ncRNA 5'−end processing
RNA catabolic process
peptidyl−lysine dimethylation
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
mRNA splicing, via spliceosome
peptidyl−arginine N−methylation
maturation of SSU−rRNA from tricistronic rRNA transcript (SSU−rRNA, 5.8S rRNA, LSU−rRNA)
peptidyl−arginine methylation
tRNA modification
rRNA catabolic process
snRNA metabolic process
RNA phosphodiester bond hydrolysis, exonucleolytic
RNA splicing, via transesterification reactions
nuclear−transcribed mRNA catabolic process
ncRNA catabolic process
RNA phosphodiester bond hydrolysis, endonucleolytic
ribosomal small subunit biogenesis
maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU−rRNA, 5.8S rRNA, LSU−rRNA)
RNA splicing
pseudouridine synthesis
histone H3−K4 methylation
maturation of 5.8S rRNA
maturation of SSU−rRNA
cleavage involved in rRNA processing
ncRNA 3'−end processing
histone modification
RNA modification
peptidyl−lysine modification
covalent chromatin modification
RNA 3'−end processing
nucleic acid phosphodiester bond hydrolysis
tRNA processing
histone lysine methylation
RNA phosphodiester bond hydrolysis
peptidyl−lysine methylation
tRNA metabolic process
histone methylation
macromolecule methylation
protein methylation
protein alkylation
ribonucleoprotein complex biogenesis
ribosome biogenesis
rRNA processing
rRNA metabolic process
methylation
ncRNA processing
ncRNA metabolic process
0.1
0.2
0.3
0.4
GeneRatio
p.adjust
4.0e−13
8.0e−13
1.2e−12
1.6e−12
Count
30
60
90
B.
SVM – Reduced dataset
ribosomal large subunit biogenesis
RNA−dependent DNA biosynthetic process
regulation of mRNA stability
mRNA 3'−end processing
telomere maintenance via telomerase
spliceosomal snRNP assembly
regulation of RNA stability
regulation of mRNA catabolic process
positive regulation of chromosome organization
maturation of SSU−rRNA
protein−containing complex localization
regulation of RNA splicing
regulation of mRNA splicing, via spliceosome
regulation of mRNA processing
mRNA transport
ribosomal small subunit biogenesis
mRNA export from nucleus
mRNA−containing ribonucleoprotein complex export from nucleus
ribonucleoprotein complex subunit organization
nucleic acid phosphodiester bond hydrolysis
ribonucleoprotein complex assembly
mRNA catabolic process
DNA−templated transcription, termination
tRNA metabolic process
RNA catabolic process
nucleobase−containing compound transport
nuclear transport
nucleocytoplasmic transport
RNA 3'−end processing
protein export from nucleus
RNA phosphodiester bond hydrolysis
nuclear export
establishment of RNA localization
nucleic acid transport
RNA transport
ribonucleoprotein complex export from nucleus
ribonucleoprotein complex localization
RNA export from nucleus
regulation of mRNA metabolic process
RNA localization
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
mRNA splicing, via spliceosome
rRNA metabolic process
RNA splicing, via transesterification reactions
rRNA processing
RNA splicing
ncRNA processing
ncRNA metabolic process
ribosome biogenesis
ribonucleoprotein complex biogenesis
0.1
0.2
0.3
GeneRatio
p.adjust
2.5e−10
5.0e−10
7.5e−10
Count
20
40
60
80

Figure 5

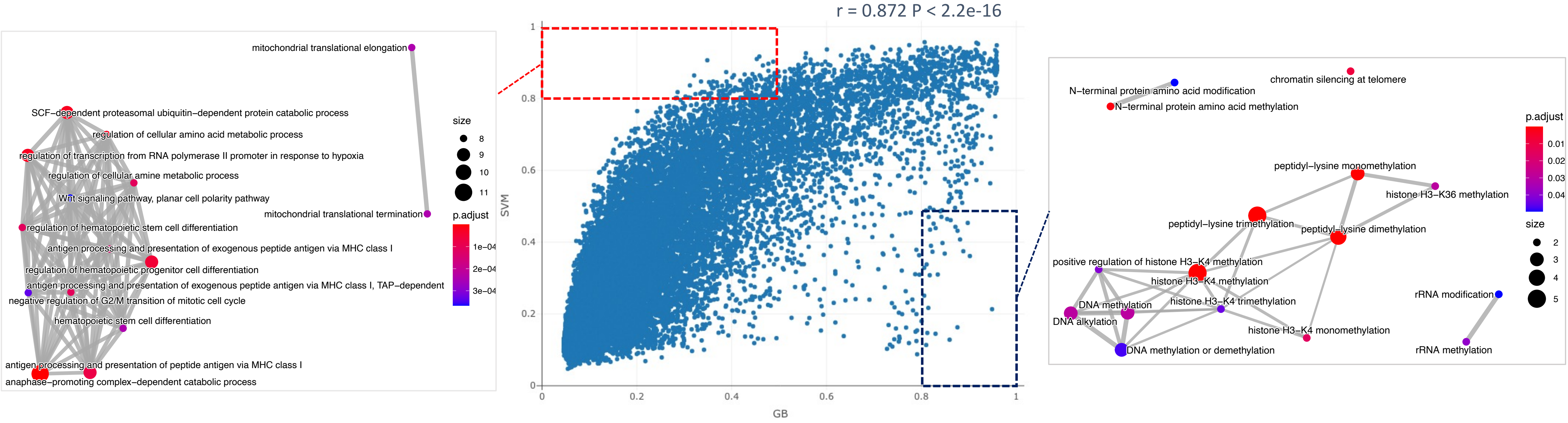

mitochondrial translational elongation
SCF−dependent proteasomal ubiquitin−dependent protein catabolic process
regulation of cellular amino acid metabolic process
regulation of transcription from RNA polymerase II promoter in response to hypoxia
regulation of cellular amine metabolic process
Wnt signaling pathway, planar cell polarity pathway
mitochondrial translational termination
regulation of hematopoietic stem cell differentiation
antigen processing and presentation of exogenous peptide antigen via MHC class I
regulation of hematopoietic progenitor cell differentiation
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP−dependent
negative regulation of G2/M transition of mitotic cell cycle
hematopoietic stem cell differentiation
antigen processing and presentation of peptide antigen via MHC class I
anaphase−promoting complex−dependent catabolic process
size
8
9
10
11
p.adjust
1e−04
2e−04
3e−04
r = 0.872 P < 2.2e-16
SVM
0
0.2
0.4
0.6
0.8
1
GB
0
0.2
0.4
0.6
0.8
1
N−terminal protein amino acid modification
chromatin silencing at telomere
N−terminal protein amino acid methylation
p.adjust
0.01
0.02
0.03
0.04
peptidyl−lysine monomethylation
histone H3−K36 methylation
peptidyl−lysine trimethylation
peptidyl−lysine dimethylation
size
2
3
4
5
positive regulation of histone H3−K4 methylation
histone H3−K4 methylation
DNA methylation
histone H3−K4 trimethylation
rRNA modification
DNA alkylation
histone H3−K4 monomethylation
DNA methylation or demethylation
rRNA methylation

Figure 6

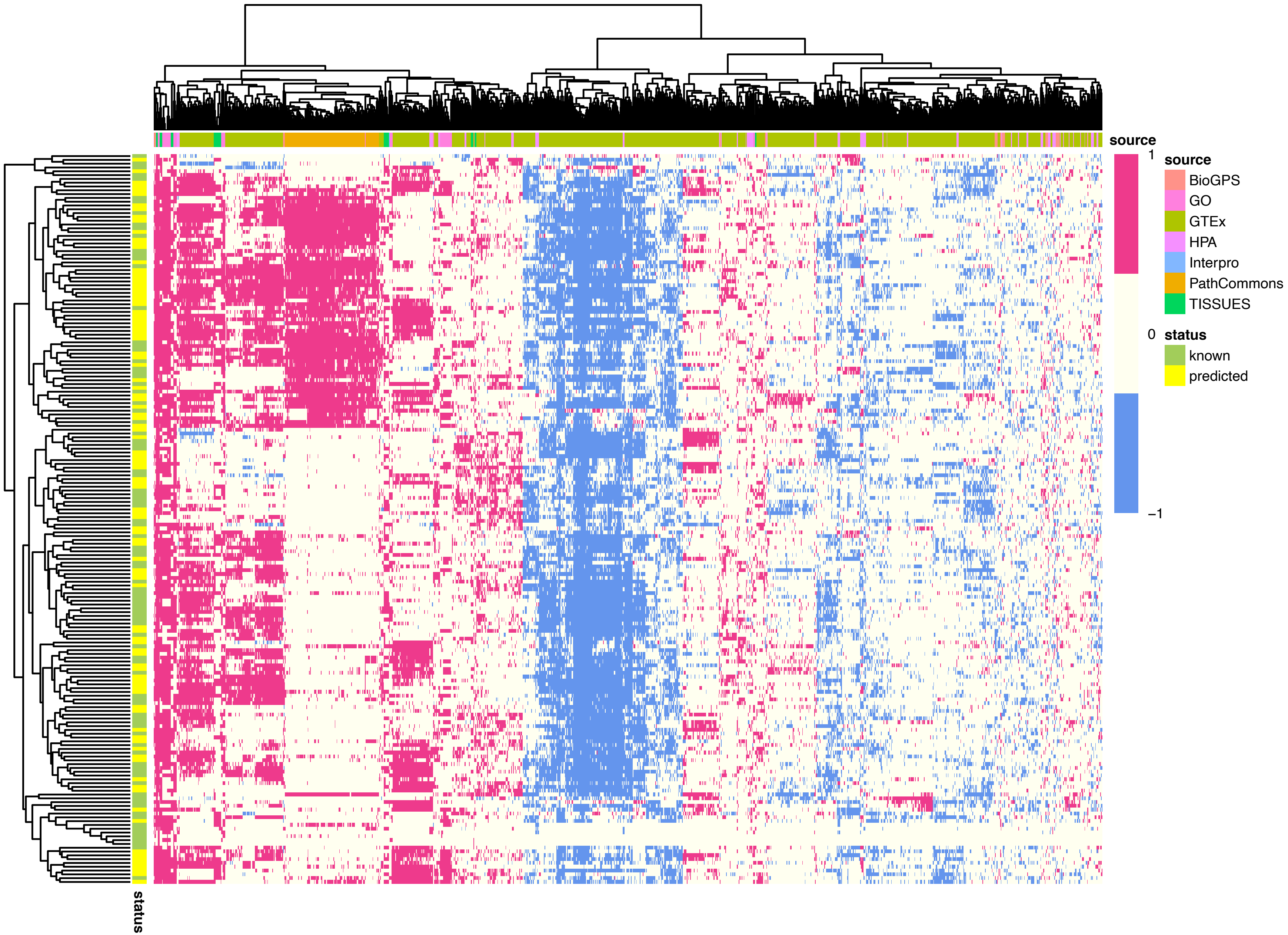

source
1
0
−1
source
BioGPS
GO
GTEx
HPA
Interpro
PathCommons
TISSUES
status
known
predicted
status

Figure 7

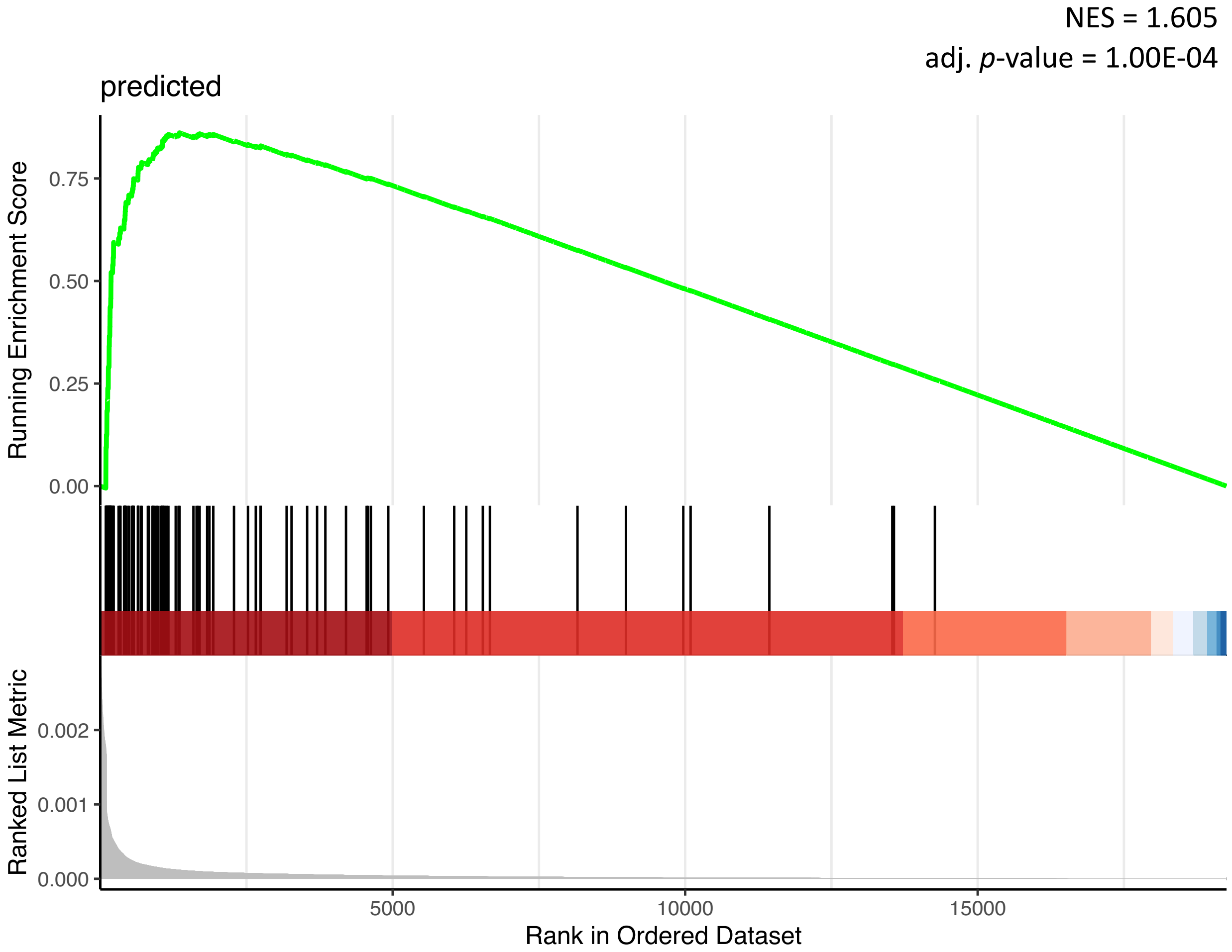

NES = 1.605
adj. *p*-value = 1.00E-04
predicted
Running Enrichment Score
0.75
0.50
0.25
0.00
Ranked List Metric
0.002
0.001
0.000
5000
10000
15000
Rank in Ordered Dataset

Figure 8

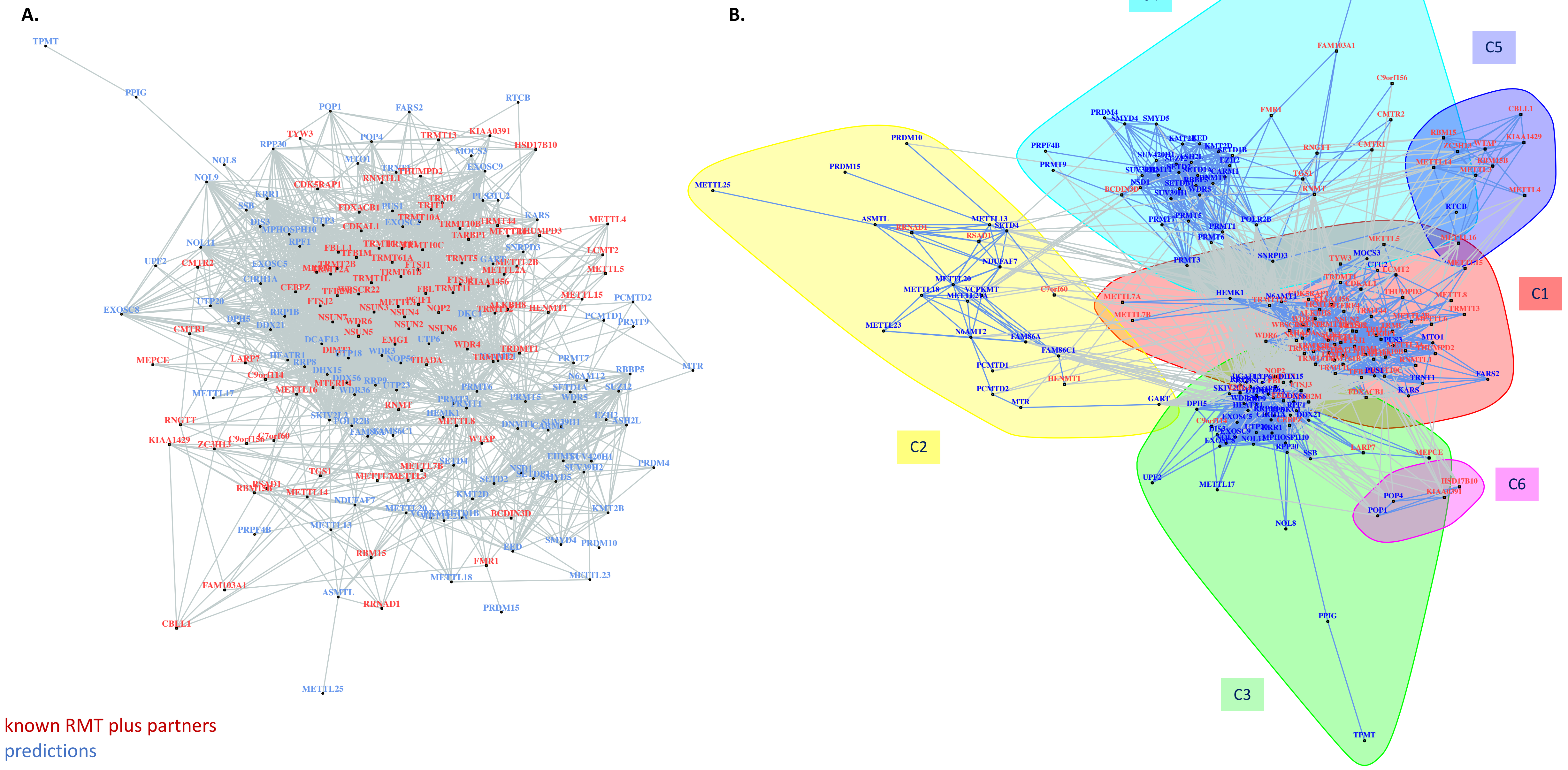

A.
B.
TPMT
PPIG
NOL8
NOL9
POP1
FARS2
RTCB
TYW3
RPP30
POP4
TRMT13
KIAA0391
HSD17B10
MOCS3
MTO1
EXOSC9
KRR1
SSB
CDK5RAP1
FDXACB1
RNMTL1
THUMPD2
METTL4
KARS
THUMPD3
LCMT2
METTL5
CMTR2
UPF2
NOL11
EXOSC5
EXOSC8
METTL15
PCMTD2
PCMTD1
PRMT9
MTR
HENMT1
CMTR1
MEPCE
LARP7
C9orf114
METTL16
RRP8
DHX15
METTL17
RNGTT
KIAA1429
ZC3H13
C9orf156
C7orf60
RNMT
WTAP
TGS1
RSAD1
RBM15B
METTL14
NDUFAF7
METTL13
PRPF4B
RBM15
FMR1
EED
SMYD4
PRDM10
KMT2B
PRDM4
METTL18
METTL23
PRDM15
ASMTL
RRNAD1
FAM103A1
CBLL1
METTL25
BCDIN3D
PRMT7
RBBP5
N6AMT2
SUZ12
EZH2
ASH2L
PCIF1
FAM103A1
C9orf156
FMR1
CMTR2
CBLL1
RBM15
ZC3H13
WTAP
KIAA1429
METTL14
RBM15B
METTL4
RTCB
RNGTT
CMTR1
TGS1
RNMT
PRDM4
SMYD4
SMYD5
PRDM10
PRPF4B
PRMT9
PRDM15
METTL25
ASMTL
RRNAD1
RSAD1
METTL13
SETD4
NDUFAF7
METTL20
METTL18
VCPKMT
METTL21A
METTL23
N6AMT2
FAM86A
FAM86C1
PCMTD1
PCMTD2
MTR
HENMT1
C7orf60
METTL7A
METTL7B
HEMK1
GART
DPH5
SNRPD3
PRMT3
PRMT6
PRMT1
PRMT5
PRMT7
POLR2B
METTL5
MOCS3
CTU2
LCMT2
METTL16
METTL15
TYW3
THUMPD3
METTL8
TRMT13
MTO1
THUMPD2
RNMTL1
TRNT1
KARS
FARS2
FDXACB1
UPF2
METTL17
NOL8
PPIG
TPMT
SSB
RPP30
LARP7
MEPCE
HSD17B10
KIAA0391
POP4
POP1
C1
C2
C3
C4
C5
C6
known RMT plus partners
predictions

Table 1

HGNC symbol	Approved name	HGNC ID	NCBI gene ID	Ensembl	UCSC gene ID	RefSeq accession	Location	Modification	Synonyms
ALKBH8	alkB homolog 8, tRNA methyltransferase	HGNC:25189	91801	ENSG00000137760	uc009yxp.4	NM_138775	11q22.3	mchm5U, mcm5s2U, mcm5U, mcm5Um	
BCDIN3D	BCDIN3 domain containing RNA methyltransferase	HGNC:27050	144233	ENSG00000186666	uc001rvh.4	NM_181708	12q13.12	mm(pN)	
BMT2	base methyltransferase of 25S rRNA 2 homolog	HGNC:26475	154743	ENSG00000164603	uc003vgo.2	NM_152556	7q31.1		C7orf60
BUD23	BUD23 rRNA methyltransferase and ribosome maturation factor	HGNC:16405	114049	ENSG00000071462	uc003tyt.4	NM_001202560	7q11.23	m7G	WBSCR22
CBLL1	Cbl proto-oncogene like 1	HGNC:21225	79872	ENSG00000105879	uc003veq.4	NM_024814	7q22.3		
CDK5RAP1	CDK5 regulatory subunit associated protein 1	HGNC:15880	51654	ENSG00000101391	uc002wyz.5	NM_016408	20q11.21	ms2i6A	
CDKAL1	CDK5 regulatory subunit associated protein 1 like 1	HGNC:21050	54901	ENSG00000145996	uc003ndd.3	NM_017774	6p22.3	ms2t6A	
CEBPZ	CCAAT enhancer binding protein zeta	HGNC:24218	10153	ENSG00000115816	uc002rpz.5	NM_005760	2p22.2		
CMTR1	cap methyltransferase 1	HGNC:21077	23070	ENSG00000137200		NM_015050	6p21.2	m7GpppNm	
CMTR2	cap methyltransferase 2	HGNC:25635	55783	ENSG00000180917		NM_018348	16q22.2	m7GpppNmNm	
DIMT1	DIMT1 rRNA methyltransferase and ribosome maturation factor	HGNC:30217	27292	ENSG00000086189	uc003jta.4	NM_014473	5q12.1	m6,6A	
EMG1	EMG1 N1-specific pseudouridine methyltransferase	HGNC:16912	10436	ENSG00000126749	uc031ysa.2	NM_006331	12p13.31		
FBL	fibrillarin	HGNC:3599	2091	ENSG00000105202	uc002omn.4	NM_001436	19q13.2	Xm	
FBLL1	fibrillarin like 1	HGNC:35458	345630	ENSG00000188573	uc011dep.3	NM_001355274	5q34		
FDXACB1	ferredoxin-fold anticodon binding domain containing 1	HGNC:25110	91893	ENSG00000255561	uc001pmc.5	NM_138378	11q23.1		
FMR1	fragile X mental retardation 1	HGNC:3775	2332	ENSG00000102081	uc010nst.4	NM_002024	Xq27.3		
FTSJ1	FtsJ RNA 2'-O-methyltransferase 1	HGNC:13254	24140	ENSG00000068438	uc004djo.3	NM_001282157	Xp11.23	Cm,Um,Gm, f5Cm, hm5Cm, mcm5Um	
FTSJ3	FtsJ RNA 2'-O-methyltransferase 3	HGNC:17136	117246	ENSG00000108592	uc002jca.3	NM_017647	17q23.3	m	
HENMT1	HEN methyltransferase 1	HGNC:26400	113802	ENSG00000162639	uc001dvu.5	NM_144584	1p13.3		
HSD17B10	hydroxysteroid 17-beta dehydrogenase 10	HGNC:4800	3028	ENSG00000072506	uc004dsl.2	NM_004493	Xp11.22	m1G,m1A	
LARP7	La ribonucleoprotein 7, transcriptional regulator	HGNC:24912	51574	ENSG00000174720	uc003iay.5	NM_016648	4q25		
LCMT2	leucine carboxyl methyltransferase 2	HGNC:17558	9836	ENSG00000168806	uc001zrg.4	NM_014793	15q15.3	o2Yw, yW	
MEPCE	methylphosphate capping enzyme	HGNC:20247	56257	ENSG00000146834	uc003uuw.3	NM_001194990	7q22.1	m7Gpp(pN)	
METTL1	methyltransferase like 1	HGNC:7030	4234	ENSG00000037897	uc010ssd.3	NM_005371	12q14.1	m7G	
METTL14	methyltransferase like 14	HGNC:29330	57721	ENSG00000145388	uc003icf.4	NM_020961	4q26		
METTL15		HGNC:26606	196074	ENSG00000169519	uc001msh.3	NM_152636	11p14.1		
METTL16	methyltransferase like 16	HGNC:28484	79066	ENSG00000127804	uc002fut.4	NM_024086	17p13.3		
METTL2A	methyltransferase like 2A	HGNC:25755	339175	ENSG00000087995	uc002izv.3	NM_181725	17q23.2		
METTL2B	methyltransferase like 2B	HGNC:18272	55798	ENSG00000165055	uc003vnf.3	NM_018396	7q32.1		
METTL3	methyltransferase like 3	HGNC:17563	56339	ENSG00000165819	uc001wbc.4	NM_019852	14q11.2	m6A	
METTL4	methyltransferase like 4	HGNC:24726	64863	ENSG00000101574	uc002klh.5	NM_022840	18p11.32	m6Am	
METTL5	methyltransferase like 5	HGNC:25006	29081	ENSG00000138382	uc002ufp.4	NM_014168	2q31.1		
METTL6	methyltransferase like 6	HGNC:28343	131965	ENSG00000206562	uc062hcc.1	NM_152396	3p25.1	m3C	
METTL7A	methyltransferase like 7A	HGNC:24550	25840	ENSG00000185432	uc058nys.1	NM_014033	12q13.12		
METTL7B	methyltransferase like 7B	HGNC:28276	196410	ENSG00000170439	uc010spr.3	NM_152637	12q13.2		
METTL8	methyltransferase like 8	HGNC:25856	79828	ENSG00000123600	uc032ojq.2	NM_024770	2q31.1		
MRM1	mitochondrial rRNA methyltransferase 1	HGNC:26202	79922	ENSG00000278619	uc032ggy.3	NM_024864	17q12	Gm	
MRM2	mitochondrial rRNA methyltransferase 2	HGNC:16352	29960	ENSG00000122687	uc003slm.3	NM_013393	7p22.3	Um	FTSJ2
MRM3	mitochondrial rRNA methyltransferase 3	HGNC:18485	55178	ENSG00000171861	uc002frw.4	NM_018146	17p13.3	Gm	RNMTL1
MTERF4	mitochondrial transcription termination factor 4	HGNC:28785	130916	ENSG00000122085		NM_182501	2q37.3		
NOP2	NOP2 nucleolar protein	HGNC:7867	4839	ENSG00000111641	uc058kgw.1	NM_006170	12p13.31		
NSUN2	NOP2/Sun RNA methyltransferase 2	HGNC:25994	54888	ENSG00000037474	uc003jdu.4	NM_017755	5p15.31	m5C	
NSUN3	NOP2/Sun RNA methyltransferase 3	HGNC:26208	63899	ENSG00000178694	uc003drl.2	NM_022072	3q11.2	f5C	
NSUN4	NOP2/Sun RNA methyltransferase 4	HGNC:31802	387338	ENSG00000117481	uc001cpr.3	NM_199044	1p33	m5C	
NSUN5	NOP2/Sun RNA methyltransferase 5	HGNC:16385	55695	ENSG00000130305	uc011kev.4	NM_148956	7q11.23		
NSUN6	NOP2/Sun RNA methyltransferase 6	HGNC:23529	221078	ENSG00000241058	uc010qcp.2	NM_182543	10p12.31	m5C	
NSUN7	NOP2/Sun RNA methyltransferase family member 7	HGNC:25857	79730	ENSG00000179299	uc003gvj.4	NM_024677	4p14		
PCIF1	PDX1 C-terminal inhibiting factor 1	HGNC:16200	63935	ENSG00000100982	uc002xqs.4	NM_022104	20q13.12		
PRORP	protein only RNase P catalytic subunit	HGNC:19958	9692	ENSG00000100890	uc001wsy.3	NM_014672	14q13.2		KIAA0391
RAMAC	RNA guanine-7 methyltransferase activating subunit	HGNC:31022	83640	ENSG00000169612	uc002bjl.3	NM_031452	15q25.2		
RBM15	RNA binding motif protein 15	HGNC:14959	64783	ENSG00000162775	uc021orn.2	NM_022768	1p13.3		
RBM15B	RNA binding motif protein 15B	HGNC:24303	29890	ENSG00000259956	uc003dbd.4	NM_013286	3p21.2		
RNGTT	RNA guanylyltransferase and 5'-phosphatase	HGNC:10073	8732	ENSG00000111880	uc003pmr.4	NM_003800	6q15	m7Gpp(pN)	
RNMT	RNA guanine-7 methyltransferase	HGNC:10075	8731	ENSG00000101654	uc002ksl.2	NM_003799	18p11.21	m7Gpp(pN)	
RRNAD1	ribosomal RNA adenine dimethylase domain containing 1	HGNC:24273	51093	ENSG00000143303	uc001fpu.4	NM_015997	1q23.1		
RSAD1	radical S-adenosyl methionine domain containing 1	HGNC:25634	55316	ENSG00000136444	uc002iqw.2	NM_018346	17q21.33		
SPOUT1	SPOUT domain containing methyltransferase 1	HGNC:26933	51490	ENSG00000198917	uc004bwd.3	NM_016390	9q34.11		C9orf114
TARBP1	TAR (HIV-1) RNA binding protein 1	HGNC:11568	6894	ENSG00000059588	uc001hwd.3	NM_005646	1q42.2	Gm	
TFB1M	transcription factor B1, mitochondrial	HGNC:17037	51106	ENSG00000029639	uc003qqj.5	NM_001350501	6q25.3	m6,6A	
TFB2M	transcription factor B2, mitochondrial	HGNC:18559	64216	ENSG00000162851	uc001ibn.4	NM_022366	1q44		
TGS1	trimethylguanosine synthase 1	HGNC:17843	96764	ENSG00000137574	uc003xsj.5	NM_024831	8q12.1	m2,2,7Gpp(pN)	
THADA	THADA armadillo repeat containing	HGNC:19217	63892	ENSG00000115970	uc002rsx.4	NM_022065	2p21		
THUMPD2	THUMP domain containing 2	HGNC:14890	80745	ENSG00000138050	uc002rru.3	NM_025264	2p22.1		
THUMPD3	THUMP domain containing 3	HGNC:24493	25917	ENSG00000134077	uc003brn.5	NM_015453	3p25.3		
TRDMT1	tRNA aspartic acid methyltransferase 1	HGNC:2977	1787	ENSG00000107614	uc001iop.4	NM_004412	10p13	m5C	
TRIT1	tRNA isopentenyltransferase 1	HGNC:20286	54802	ENSG00000043514	uc057fcv.1	NM_017646	1p34.2	i6A	
TRMO	tRNA methyltransferase O	HGNC:30967	51531	ENSG00000136932		NM_016481	9q22.33	m6t6A	C9orf156
TRMT1	tRNA methyltransferase 1	HGNC:25980	55621	ENSG00000104907	uc060ugy.1	NM_017722	19p13.13	m2,2G	
TRMT10A	tRNA methyltransferase 10A	HGNC:28403	93587	ENSG00000145331	uc003hva.5	NM_152292	4q23	m1G	
TRMT10B	tRNA methyltransferase 10B	HGNC:26454	158234	ENSG00000165275	uc004aai.5	NM_144964	9p13.2	m1G	
TRMT10C	tRNA methyltransferase 10C, mitochondrial RNase P subunit	HGNC:26022	54931	ENSG00000174173	uc003duz.5	NM_017819	3q12.3	m1G,m1A	
TRMT11	tRNA methyltransferase 11 homolog	HGNC:21080	60487	ENSG00000066651	uc003qam.4	NM_021820	6q22.32		
TRMT112	tRNA methyltransferase subunit 11-2	HGNC:26940	51504	ENSG00000173113	uc001nzt.5	NM_016404	11q13.1	m7G	
TRMT12	tRNA methyltransferase 12 homolog	HGNC:26091	55039	ENSG00000183665	uc003yra.5	NM_017956	8q24.13	o2Yw, yW	
TRMT13	tRNA methyltransferase 13 homolog	HGNC:25502	54482	ENSG00000122435	uc001dsv.4	NM_019083	1p21.2		
TRMT1L	tRNA methyltransferase 1 like	HGNC:16782	81627	ENSG00000121486	uc001grf.5	NM_030934	1q25.3		
TRMT2A	tRNA methyltransferase 2 homolog A	HGNC:24974	27037	ENSG00000099899	uc002zrk.3	NM_022727	22q11.21	m5U	
TRMT2B	tRNA methyltransferase 2 homolog B	HGNC:25748	79979	ENSG00000188917	uc004egq.4	NM_024917	Xq22.1		
TRMT44	tRNA methyltransferase 44 homolog	HGNC:26653	152992	ENSG00000155275	uc003glg.3	NM_152544	4p16.1	Um	
TRMT5	tRNA methyltransferase 5	HGNC:23141	57570	ENSG00000126814	uc001xff.5	NM_020810	14q23.1	m1G, m1I	
TRMT6	tRNA methyltransferase 6	HGNC:20900	51605	ENSG00000089195	uc002wmh.3	NM_001281467	20p12.3	m1A	
TRMT61A	tRNA methyltransferase 61A	HGNC:23790	115708	ENSG00000166166	uc001yng.4	NM_152307	14q32	m1A	
TRMT61B	tRNA methyltransferase 61B	HGNC:26070	55006	ENSG00000171103	uc002rmm.5	NM_017910	2p23.2	m1A	
TRMT9B	tRNA methyltransferase 9B (putative)	HGNC:26725	57604	ENSG00000250305	uc010lsq.4	NM_001099677	8p22		KIAA1456
TRMU	tRNA 5-methylaminomethyl-2-thiouridylate methyltransferase	HGNC:25481	55687	ENSG00000100416	uc003bhp.4	NM_018006	22q13.31	tm5s2	
TYW3	tRNA-yW synthesizing protein 3 homolog	HGNC:24757	127253	ENSG00000162623	uc001dgn.4	NM_138467	1p31.1		
VIRMA	vir like m6A methyltransferase associated	HGNC:24500	25962	ENSG00000164944	uc003ygo.3	NM_015496	8q22.1		KIAA1429
WDR4	WD repeat domain 4	HGNC:12756	10785	ENSG00000160193	uc002zci.5	NM_001260474	21q22.3		
WDR6	WD repeat domain 6	HGNC:12758	11180	ENSG00000178252	uc062jnu.1	NM_001320546	3p21.31	Cm, Gm,f5Cm, hm5Cm	
WTAP	WT1 associated protein	HGNC:16846	9589	ENSG00000146457	uc003qsl.6	NM_152857	6q25.3		
ZC3H13	zinc finger CCCH-type containing 13	HGNC:20368	23091	ENSG00000123200	uc001vas.3	NM_015070	13q14.13		
ZCCHC4	zinc finger CCHC-type containing 4	HGNC:22917	29063	ENSG00000168228	uc003grl.5	NM_001318148	4p15.2		

Table 2

Dataset	Description	Measurement	Association	Category	Resource	Genes	Attributes	Associations
BioGPS Human Cell Type and Tissue Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by microa	Gene-cell type or tissue associations by differential expression o	Transcriptomics	BioGPS	16379	84 cell type or tissues	205445 gene-cell type or tissue associations
GTEx Tissue Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by RNA-s	Gene-tissue associations by differential expression of gene acros	Transcriptomics	Genotype Tissue Expression	25557	29 tissues	112583 gene-tissue associations
GTEx Tissue Sample Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by RNA-s	Gene-tissue sample associations by differential expression of ge	Transcriptomics	Genotype Tissue Expression	19247	2918 tissue samples	8421199 gene-tissue sample associations
HPA Cell Line Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by RNA-s	Gene-cell line associations by differential expression of gene acr	Transcriptomics	Human Protein Atlas	15372	43 cell lines	102943 gene-cell line associations
HPA Tissue Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by RNA-s	Gene-tissue associations by differential expression of gene acros	Transcriptomics	Human Protein Atlas	17423	31 tissues	81082 gene-tissue associations
HPA Tissue Sample Gene Expression Profiles Dataset	mRNA expression prof	Gene expression by RNA-s	Gene-tissue sample associations by differential expression of ge	Transcriptomics	Human Protein Atlas	16657	121 tissue samples	303267 gene-tissue sample associations
GO Biological Process Annotations Dataset	Curated annotations o	Association by literature cu	Gene-biological process associations from curated gene annotati	Structural or functional annotations	Gene Ontology	15717	13212 biological processs	969303 gene-biological process associations
GO Molecular Function Annotations Dataset	Curated annotations o	Association by literature cu	Gene-molecular function associations from curated gene annota	Structural or functional annotations	Gene Ontology	15777	4162 molecular functions	223181 gene-molecular function associations
InterPro Predicted Protein Domain Annotations Dataset	Protein domains predi	Association by computatio	Protein-protein domain associations by sequence similarity to do	Structural or functional annotations	InterPro	18002	11015 protein domains	62614 gene-protein domain associations
KEGG Pathways Dataset	Sets of proteins partici	Association by literature cu	Protein-pathway associations from curated pathways	Structural or functional annotations	Kyoto Encyclopedia of Genes	3947	200 pathways	9324 gene-pathway associations
Reactome Pathways Dataset	Sets of proteins partici	Association by literature cu	Protein-pathway associations from curated pathways	Structural or functional annotations	Reactome	7535	1638 pathways	83680 gene-pathway associations
TISSUES Curated Tissue Protein Expression Evidence Scores Dataset	Protein tissue expressi	Association by literature cu	Protein-tissue associations by integrating evidence from manual	Proteomics	TISSUES	16215	643 tissues	357442 gene-tissue associations
HPA Tissue Protein Expression Profiles Dataset	Semiquantitative prote	Protein expression by imm	Protein-tissue associations by differential expression of protein a	Proteomics	Human Protein Atlas	15704	44 tissues	138576 gene-tissue associations
Hub Proteins Protein-Protein Interactions Dataset	Sets of proteins intera	Association by data aggreg	Protein-hub protein associations from aggregated protein-protei	Physical interactions	Hub Proteins	9362	289 hub proteins	58320 gene-hub protein association
Pathway Commons Protein-Protein Interactions Dataset	Protein-protein interac	Association by data aggreg	Protein-protein associations from low-throughput or high-throug	Physical interactions	Pathway Commons	15747	15747 interacting proteins	3527164 gene-interacting protein associations

Table 3

Data source	Feature ID	Tissue (if applicable)	Nb Sets	Frequency
PathCommons_PPI	NRF1		233	80.9
PathCommons_PPI	UBC		193	67.0
GTEx_SampleGene	GTEX-RVPV-0006-SM-2TF6Q	Whole Blood	172	59.7
HPA_TissueSample	pancreas_6b	Pancreas	170	59.0
GTEx_SampleGene	GTEX-WYJK-0005-SM-3NMA1	Whole Blood	169	58.7
GTEx_SampleGene	GTEX-WRHU-1226-SM-4E3IJ	Heart - Left Ventricle	155	53.8
HPA_TissueGene	lymph_node	Lymph Node	152	52.8
HPA_TissueSample	lymphnode_5a	Lymph Node	144	50.0
HPA_TissueSample	lymphnode_4b	Lymph Node	139	48.3
GTEx_SampleGene	GTEX-T5JW-0008-SM-4DM5X	Cells - Cultured fibroblasts	137	47.6
GTEx_SampleGene	GTEX-XLM4-0004-SM-4AT5I	Cells - EBV-transformed lymphocytes	133	46.2
BioGPS	CD19+_BCells(neg._sel.)	B Cells	132	45.8
GTEx_SampleGene	GTEX-RVPU-0005-SM-2TF6L	Whole Blood	129	44.8
GTEx_SampleGene	GTEX-NFK9-0726-SM-2HMJW	Thyroid	128	44.4
HPA_TissueGene	pancreas	Pancreas	126	43.8
GTEx_SampleGene	GTEX-XBEC-1326-SM-4AT69	Heart - Left Ventricle	125	43.4
GTEx_SampleGene	GTEX-OHPN-0011-R4A-SM-2I5FD	Brain - Amygdala	122	42.4
GTEx_SampleGene	GTEX-VUSG-0003-SM-3NMDK	Cells - EBV-transformed lymphocytes	121	42.0
GTEx_SampleGene	GTEX-T6MO-0003-SM-3NMAG	Cells - EBV-transformed lymphocytes	113	39.2
GTEx_SampleGene	GTEX-Q2AI-0008-SM-48U2H	Cells - Cultured fibroblasts	112	38.9
GTEx_SampleGene	GTEX-WFG7-0001-SM-3P61S	Cells - EBV-transformed lymphocytes	111	38.5
GTEx_SampleGene	GTEX-WZTO-0426-SM-3NM99	Lung	111	38.5
GTEx_SampleGene	GTEX-X62O-0008-SM-46MU5	Cells - Cultured fibroblasts	111	38.5
TISSUES_curatProtein	BTO:0003091	Urogenital System	101	35.1
GTEx_SampleGene	GTEX-S7SF-0008-SM-3NM8T	Cells - Cultured fibroblasts	100	34.7
GTEx_SampleGene	GTEX-NL3H-0011-R1a-SM-48TDJ	Brain - Hippocampus	98	34.0
TISSUES_curatProtein	BTO:0000000		96	33.3
PathCommons_PPI	HNF4A		94	32.6
BioGPS	CD8+_Tcells	T Cells	93	32.3
TISSUES_curatProtein	BTO:0000081	Reproductive System	90	31.3
TISSUES_curatProtein	BTO:0000042		89	30.9
BioGPS	CD34+		88	30.6
GTEx_SampleGene	GTEX-S4UY-0008-SM-3NM8H	Cells - Cultured fibroblasts	88	30.6
GTEx_SampleGene	GTEX-UJMC-0326-SM-3GAE2	Thyroid	86	29.9
GTEx_SampleGene	GTEX-XGQ4-0008-SM-4AT3Z	Cells - Cultured fibroblasts	86	29.9
BioGPS	CD105+_Endothelial		85	29.5
GTEx_SampleGene	GTEX-WYVS-1726-SM-3NMAY	Breast - Mammary Tissue	85	29.5
HPA_CellLineGene	karpas707		81	28.1
GTEx_SampleGene	GTEX-WZTO-0006-SM-3NM9T	Whole Blood	80	27.8
GTEx_SampleGene	GTEX-S3XE-0006-SM-3K2AA	Whole Blood	78	27.1
GTEx_SampleGene	GTEX-TML8-0001-SM-3NMAF	Cells - EBV-transformed lymphocytes	78	27.1
GTEx_SampleGene	GTEX-X638-0003-SM-47JZ1	Cells - EBV-transformed lymphocytes	77	26.7
GTEx_SampleGene	GTEX-NL3H-0011-R7a-SM-2I3G5	Brain - Putamen (basal ganglia)	76	26.4
GTEx_SampleGene	GTEX-QDVJ-0008-SM-48U2E	Cells - Cultured fibroblasts	76	26.4
GTEx_SampleGene	GTEX-UPK5-0003-SM-3NMDI	Cells - EBV-transformed lymphocytes	75	26.0
HPA_TissueSample	testis_7a	Testis	75	26.0
GTEx_SampleGene	GTEX-QCQG-0006-SM-2S1OW	Whole Blood	73	25.3
PathCommons_PPI	EFTUD2		73	25.3
GTEx_SampleGene	GTEX-NL4W-0006-SM-2I3GH	Whole Blood	72	25.0
HPA_CellLineGene	u698		72	25.0
GTEx_SampleGene	GTEX-S7PM-0008-SM-3NM9Q	Cells - Cultured fibroblasts	71	24.7
GTEx_SampleGene	GTEX-U3ZN-0326-SM-3DB86	Thyroid	71	24.7
GTEx_SampleGene	GTEX-XQ8I-0006-SM-4BOQ5	Whole Blood	71	24.7
GTEx_SampleGene	GTEX-X4XX-0926-SM-46MV7	Thyroid	70	24.3
HPA_TissueGene	tonsil	Tonsil	70	24.3
GTEx_SampleGene	GTEX-S4P3-0008-SM-3NM8R	Cells - Cultured fibroblasts	69	24.0
GTEx_SampleGene	GTEX-S4Q7-0006-SM-3K2AT	Whole Blood	67	23.3
GTEx_SampleGene	GTEX-WHSB-1826-SM-3TW8M	Muscle - Skeletal	67	23.3
PathCommons_PPI	BCLAF1		67	23.3
GTEx_SampleGene	GTEX-UPIC-0226-SM-3GADO	Thyroid	65	22.6
GTEx_SampleGene	GTEX-WOFL-0006-SM-3TW8K	Whole Blood	65	22.6
GTEx_SampleGene	GTEX-X261-0011-R7A-SM-4E3JJ	Brain - Putamen (basal ganglia)	65	22.6
HPA_TissueSample	testis_7e	Testis	65	22.6
GTEx_SampleGene	GTEX-RVPU-0011-R1A-SM-2XCAI	Brain - Hippocampus	64	22.2
GTEx_SampleGene	GTEX-S341-0006-SM-3NM8D	Whole Blood	64	22.2
GTEx_SampleGene	GTEX-T6MN-0002-SM-3NMAH	Cells - EBV-transformed lymphocytes	63	21.9
GTEx_SampleGene	GTEX-NFK9-0006-SM-3GACS	Whole Blood	62	21.5
GTEx_SampleGene	GTEX-P44H-0006-SM-2XCFB	Whole Blood	62	21.5
GTEx_SampleGene	GTEX-UPIC-1526-SM-4IHLU	Uterus	62	21.5
GTEx_SampleGene	GTEX-POMQ-0008-SM-48TE7	Cells - Cultured fibroblasts	61	21.2
GTEx_SampleGene	GTEX-VUSH-0004-SM-3P61T	Cells - EBV-transformed lymphocytes	61	21.2
GTEx_SampleGene	GTEX-X8HC-0726-SM-46MWG	Thyroid	61	21.2
GTEx_SampleGene	GTEX-QESD-0006-SM-2I5G6	Whole Blood	60	20.8
GTEx_SampleGene	GTEX-S4P3-0006-SM-3K2AW	Whole Blood	60	20.8
HPA_TissueProtein	rectum	Rectum	60	20.8
PathCommons_PPI	NOP56		60	20.8
GTEx_SampleGene	GTEX-T5JC-0001-SM-3NMAK	Cells - EBV-transformed lymphocytes	59	20.5
GTEx_SampleGene	GTEX-X585-0002-SM-46MVA	Cells - EBV-transformed lymphocytes	59	20.5
GTEx_SampleGene	GTEX-WHSE-0126-SM-3NMBT	Skin - Not Sun Exposed (Suprapubic)	58	20.1
PathCommons_PPI	RPS9		58	20.1
GTEx_SampleGene	GTEX-RTLS-0006-SM-2TF58	Whole Blood	57	19.8
GTEx_SampleGene	GTEX-T2IS-0426-SM-32QPE	Heart - Left Ventricle	57	19.8
GTEx_SampleGene	GTEX-UPIC-0926-SM-4IHLV	Liver	57	19.8
TISSUES_curatProtein	BTO:0001489	Whole Body	57	19.8
GTEx_SampleGene	GTEX-RWS6-0326-SM-2XCAP	Heart - Left Ventricle	56	19.4
PathCommons_PPI	RPL7A		56	19.4
HPA_TissueSample	tonsil_8b1	Tonsil	55	19.1
HPA_TissueSample	skeletalmuscle_d	Muscle - Skeletal	54	18.8
HPA_TissueSample	testis_7b	Testis	54	18.8
GTEx_SampleGene	GTEX-PVOW-1626-SM-48TC9	Esophagus - Mucosa	53	18.4
GTEx_SampleGene	GTEX-WFON-0001-SM-3P61W	Cells - EBV-transformed lymphocytes	53	18.4
GTEx_SampleGene	GTEX-XGQ4-0005-SM-4AT5U	Whole Blood	53	18.4
HPA_TissueSample	testis_4a	Testis	53	18.4
PathCommons_PPI	RPS13		53	18.4
GTEx_SampleGene	GTEX-TSE9-2626-SM-4DXV2	Uterus	52	18.1
TISSUES_curatProtein	BTO:0000534	Gonad	52	18.1
GTEx_SampleGene	GTEX-U8T8-0008-SM-4DXSP	Cells - Cultured fibroblasts	51	17.7
HPA_TissueSample	pancreas_6a	Pancreas	51	17.7
GTEx_SampleGene	GTEX-P78B-0008-SM-48TE1	Cells - Cultured fibroblasts	50	17.4
GTEx_SampleGene	GTEX-SIU7-0001-SM-3NMAW	Cells - EBV-transformed lymphocytes	50	17.4

Table 4

Full Dataset

Model	Accuracy	+/-	Precision	+/-	Recall	+/-	F1	+/-	AUC	+/-
Gradient Boosting (GB)	0.875	0.025	0.895	0.033	0.865	0.031	0.872	0.025	0.938	0.015
Gaussian Naïve Bayes (GNB)	0.851	0.025	0.821	0.032	0.924	0.021	0.863	0.021	0.862	0.023
Logistic Regression (LR)	0.859	0.021	0.870	0.025	0.859	0.023	0.857	0.021	0.921	0.015
Random Forest (RF)	0.870	0.021	0.870	0.026	0.886	0.032	0.871	0.022	0.937	0.014
Support Vector Machine (SVM)	0.856	0.022	0.876	0.028	0.845	0.027	0.852	0.023	0.921	0.017
Dataset w/o GO/InterPro										
Gradient Boosting (GB)	0.799	0.029	0.800	0.035	0.819	0.032	0.801	0.029	0.860	0.031
Gaussian Naïve Bayes (GNB)	0.781	0.022	0.765	0.028	0.840	0.043	0.792	0.024	0.800	0.021
Logistic Regression (LR)	0.795	0.030	0.797	0.035	0.814	0.030	0.797	0.029	0.857	0.032
Random Forest (RF)	0.805	0.024	0.802	0.033	0.833	0.023	0.809	0.022	0.867	0.025
Support Vector Machine (SVM)	0.812	0.027	0.822	0.036	0.816	0.032	0.811	0.027	0.864	0.026

Table 5

Gene	Mean Prob	UniProt Entry	Entry Name	Gene Names	Protein Names
METTL13	0.944	Q8N6R0	EFNMT_HUMAN	*EEF1AKNMT KIAA0859 METTL13 CGI-01*	eEF1A lysine and N-terminal methyltransferase (eEF1A-KNMT) (Methyltransferase-like protein 13) [Includes: eEF1A lysine methyltransferase (EC 2.1.1.-); eEF1A N-terminal methyltransferase (EC 2.1.1.-)]
PRMT5	0.943	O14744	ANM5_HUMAN	*PRMT5 HRMT1L5 IBP72 JBP1 SKB1*	Protein arginine N-methyltransferase 5 (PRMT5) (EC 2.1.1.320) (72 kDa ICln-binding protein) (Histone-arginine N-methyltransferase PRMT5) (Jak-binding protein 1) (Shk1 kinase-binding protein 1 homolog) (SKB1 homolog) (SKB1Hs) [Cleaved into: Protein arginine N-methyltransferase 5, N-terminally processed]
RRP8	0.940	O43159	RRP8_HUMAN	*RRP8 KIAA0409 NML hucep-1*	Ribosomal RNA-processing protein 8 (EC 2.1.1.-) (Cerebral protein 1) (Nucleomethylin)
METTL18	0.933	O95568	MET18_HUMAN	*METTL18 ASTP2 C1orf156*	Histidine protein methyltransferase 1 homolog (EC 2.1.1.-) (Arsenic-transactivated protein 2) (AsTP2) (Methyltransferase-like protein 18)
SETD2	0.933	Q9BYW2	SETD2_HUMAN	*SETD2 HIF1 HYPB KIAA1732 KMT3A SET2 HSPC069*	Histone-lysine N-methyltransferase SETD2 (EC 2.1.1.359) (HIF-1) (Huntingtin yeast partner B) (Huntingtin-interacting protein 1) (HIP-1) (Huntingtin-interacting protein B) (Lysine N-methyltransferase 3A) (Protein-lysine N-methyltransferase SETD2) (EC 2.1.1.-) (SET domain-containing protein 2) (hSET2) (p231HBP)
RBBP5	0.930	Q15291	RBBP5_HUMAN	*RBBP5 RBQ3*	Retinoblastoma-binding protein 5 (RBBP-5) (Retinoblastoma-binding protein RBQ-3)
SETDB1	0.929	Q15047	SETB1_HUMAN	*SETDB1 ESET KIAA0067 KMT1E*	Histone-lysine N-methyltransferase SETDB1 (EC 2.1.1.366) (ERG-associated protein with SET domain) (ESET) (Histone H3-K9 methyltransferase 4) (H3-K9-HMTase 4) (Lysine N-methyltransferase 1E) (SET domain bifurcated 1)
PRDM15	0.929	P57071	PRD15_HUMAN	*PRDM15 C21orf83 ZNF298*	PR domain zinc finger protein 15 (EC 2.1.1.-) (PR domain-containing protein 15) (Zinc finger protein 298)
SUZ12	0.928	Q15022	SUZ12_HUMAN	*SUZ12 CHET9 JJAZ1 KIAA0160*	Polycomb protein SUZ12 (Chromatin precipitated E2F target 9 protein) (ChET 9 protein) (Joined to JAZF1 protein) (Suppressor of zeste 12 protein homolog)
SUV39H1	0.927	O43463	SUV91_HUMAN	*SUV39H1 KMT1A SUV39H*	Histone-lysine N-methyltransferase SUV39H1 (EC 2.1.1.355) (Histone H3-K9 methyltransferase 1) (H3-K9-HMTase 1) (Lysine N-methyltransferase 1A) (Position-effect variegation 3-9 homolog) (Suppressor of variegation 3-9 homolog 1) (Su(var)3-9 homolog 1)
KRR1	0.927	Q13601	KRR1_HUMAN	*KRR1 HRB2*	KRR1 small subunit processome component homolog (HIV-1 Rev-binding protein 2) (KRR-R motif-containing protein 1) (Rev-interacting protein 1) (Rip-1)
GART	0.926	P22102	PUR2_HUMAN	*GART PGFT PRGS*	Trifunctional purine biosynthetic protein adenosine-3 [Includes: Phosphoribosylamine--glycine ligase (EC 6.3.4.13) (Glycinamide ribonucleotide synthetase) (GARS) (Phosphoribosylglycinamide synthetase); Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1) (AIR synthase) (AIRS) (Phosphoribosyl-aminoimidazole synthetase); Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (5'-phosphoribosylglycinamide transformylase) (GAR transformylase) (GART)]
SNRPD3	0.926	P62318	SMD3_HUMAN	*SNRPD3*	Small nuclear ribonucleoprotein Sm D3 (Sm-D3) (snRNP core protein D3)
DIS3	0.922	Q9Y2L1	RRP44_HUMAN	*DIS3 KIAA1008 RRP44*	Exosome complex exonuclease RRP44 (EC 3.1.13.-) (EC 3.1.26.-) (Protein DIS3 homolog) (Ribosomal RNA-processing protein 44)
SUV39H2	0.922	Q9H5I1	SUV92_HUMAN	*SUV39H2 KMT1B*	Histone-lysine N-methyltransferase SUV39H2 (EC 2.1.1.355) (Histone H3-K9 methyltransferase 2) (H3-K9-HMTase 2) (Lysine N-methyltransferase 1B) (Suppressor of variegation 3-9 homolog 2) (Su(var)3-9 homolog 2)
WDR5	0.922	P61964	WDR5_HUMAN	*WDR5 BIG3*	WD repeat-containing protein 5 (BMP2-induced 3-kb gene protein)
PRDM4	0.920	Q9UKN5	PRDM4_HUMAN	*PRDM4 PFM1*	PR domain zinc finger protein 4 (EC 2.1.1.-) (PR domain-containing protein 4)
EXOSC2	0.920	Q13868	EXOS2_HUMAN	*EXOSC2 RRP4*	Exosome complex component RRP4 (Exosome component 2) (Ribosomal RNA-processing protein 4)
PRMT1	0.918	Q99873	ANM1_HUMAN	*PRMT1 HMT2 HRMT1L2 IR1B4*	Protein arginine N-methyltransferase 1 (EC 2.1.1.319) (Histone-arginine N-methyltransferase PRMT1) (Interferon receptor 1-bound protein 4)
SKIV2L2	0.917	P42285	MTREX_HUMAN	*MTREX DOB1 KIAA0052 MTR4 SKIV2L2*	Exosome RNA helicase MTR4 (EC 3.6.4.13) (ATP-dependent RNA helicase DOB1) (ATP-dependent RNA helicase SKIV2L2) (Superkiller viralicidic activity 2-like 2) (TRAMP-like complex helicase)
UTP23	0.917	Q9BRU9	UTP23_HUMAN	*UTP23 C8orf53*	rRNA-processing protein UTP23 homolog
FAM86A	0.917	Q96G04	EF2KT_HUMAN	*EEF2KMT FAM86A SB153*	Protein-lysine N-methyltransferase EEF2KMT (EC 2.1.1.-) (eEF2-lysine methyltransferase) (eEF2-KMT)
RPP30	0.917	P78346	RPP30_HUMAN	*RPP30 RNASEP2*	Ribonuclease P protein subunit p30 (RNaseP protein p30) (EC 3.1.26.5) (RNase P subunit 2)
EHMT1	0.917	Q9H9B1	EHMT1_HUMAN	*EHMT1 EUHMTASE1 GLP KIAA1876 KMT1D*	Histone-lysine N-methyltransferase EHMT1 (EC 2.1.1.-) (Euchromatic histone-lysine N-methyltransferase 1) (Eu-HMTase1) (G9a-like protein 1) (GLP) (GLP1) (Histone H3-K9 methyltransferase 5) (H3-K9-HMTase 5) (Lysine N-methyltransferase 1D)
METTL17	0.917	Q9H7H0	MET17_HUMAN	*METTL17 METT11D1*	Methyltransferase-like protein 17, mitochondrial (EC 2.1.1.-) (False p73 target gene protein) (Methyltransferase 11 domain-containing protein 1) (Protein RSM22 homolog, mitochondrial)
EXOSC9	0.917	Q06265	EXOS9_HUMAN	*EXOSC9 PMSCL1*	Exosome complex component RRP45 (Autoantigen PM/Scl 1) (Exosome component 9) (P75 polymyositis-scleroderma overlap syndrome-associated autoantigen) (Polymyositis/scleroderma autoantigen 1) (Polymyositis/scleroderma autoantigen 75 kDa) (PM/Scl-75)
N6AMT2	0.916	Q8WVE0	EFMT1_HUMAN	*EEF1AKMT1 N6AMT2*	EEF1A lysine methyltransferase 1 (EC 2.1.1.-) (N(6)-adenine-specific DNA methyltransferase 2) (Protein-lysine N-methyltransferase N6AMT2)
DDX56	0.916	Q9NY93	DDX56_HUMAN	*DDX56 DDX21 NOH61*	Probable ATP-dependent RNA helicase DDX56 (EC 3.6.4.13) (ATP-dependent 61 kDa nucleolar RNA helicase) (DEAD box protein 21) (DEAD box protein 56)
TPMT	0.916	P51580	TPMT_HUMAN	*TPMT*	Thiopurine S-methyltransferase (EC 2.1.1.67) (Thiopurine methyltransferase)
DPH5	0.915	Q9H2P9	DPH5_HUMAN	*DPH5 AD-018 CGI-30 HSPC143 NPD015*	Diphthine methyl ester synthase (EC 2.1.1.314) (Diphthamide biosynthesis methyltransferase)
SETD1A	0.915	O15047	SET1A_HUMAN	*SETD1A KIAA0339 KMT2F SET1 SET1A*	Histone-lysine N-methyltransferase SETD1A (EC 2.1.1.354) (Lysine N-methyltransferase 2F) (SET domain-containing protein 1A) (hSET1A) (Set1/Ash2 histone methyltransferase complex subunit SET1)
UTP3	0.915	Q9NQZ2	SAS10_HUMAN	*UTP3 CRLZ1 SAS10*	Something about silencing protein 10 (Charged amino acid-rich leucine zipper 1) (CRL1) (Disrupter of silencing SAS10) (UTP3 homolog)
SUV420H1	0.914	Q4FZB7	KMT5B_HUMAN	*KMT5B SUV420H1 CGI-85*	Histone-lysine N-methyltransferase KMT5B (Lysine N-methyltransferase 5B) (Lysine-specific methyltransferase 5B) (Suppressor of variegation 4-20 homolog 1) (Su(var)4-20 homolog 1) (Suv4-20h1) ([histone H4]-N-methyl-L-lysine20 N-methyltransferase KMT5B) (EC 2.1.1.362) ([histone H4]-lysine20 N-methyltransferase KMT5B) (EC 2.1.1.361)
EED	0.912	O75530	EED_HUMAN	*EED*	Polycomb protein EED (hEED) (Embryonic ectoderm development protein) (WD protein associating with integrin cytoplasmic tails 1) (WAIT-1)
DKC1	0.912	O60832	DKC1_HUMAN	*DKC1 NOLA4*	H/ACA ribonucleoprotein complex subunit DKC1 (EC 5.4.99.-) (CBF5 homolog) (Dyskerin) (Nopp140-associated protein of 57 kDa) (Nucleolar protein NAP57) (Nucleolar protein family A member 4) (snoRNP protein DKC1)
METTL23	0.911	Q86XA0	MET23_HUMAN	*METTL23 C17orf95*	Methyltransferase-like protein 23 (EC 2.1.1.-)
HEMK1	0.911	Q9Y5R4	HEMK1_HUMAN	*HEMK1 HEMK*	MTRF1L release factor glutamine methyltransferase (EC 2.1.1.297) (HemK methyltransferase family member 1) (M.HsaHemKP)
PRDM10	0.910	Q9NQV6	PRD10_HUMAN	*PRDM10 KIAA1231 PFM7 TRIS*	PR domain zinc finger protein 10 (EC 2.1.1.-) (PR domain-containing protein 10) (Tristanin)
POP1	0.910	Q99575	POP1_HUMAN	*POP1 KIAA0061*	Ribonucleases P/MRP protein subunit POP1 (hPOP1) (EC 3.1.26.5)
NSD1	0.910	Q96L73	NSD1_HUMAN	*NSD1 ARA267 KMT3B*	Histone-lysine N-methyltransferase, H3 lysine-36 specific (EC 2.1.1.357) (Androgen receptor coactivator 267 kDa protein) (Androgen receptor-associated protein of 267 kDa) (H3-K36-HMTase) (Lysine N-methyltransferase 3B) (Nuclear receptor-binding SET domain-containing protein 1) (NR-binding SET domain-containing protein)
KMT2D	0.910	O14686	KMT2D_HUMAN	*KMT2D ALR MLL2 MLL4*	Histone-lysine N-methyltransferase 2D (Lysine N-methyltransferase 2D) (EC 2.1.1.354) (ALL1-related protein) (Myeloid/lymphoid or mixed-lineage leukemia protein 2)
SMYD4	0.909	Q8IYR2	SMYD4_HUMAN	*SMYD4 KIAA1936*	SET and MYND domain-containing protein 4 (EC 2.1.1.-)
MOCS3	0.909	O95396	MOCS3_HUMAN	*MOCS3 UBA4*	Adenylyltransferase and sulfurtransferase MOCS3 (Molybdenum cofactor synthesis protein 3) (Molybdopterin synthase sulfurylase) (MPT synthase sulfurylase) [Includes: Molybdopterin-synthase adenylyltransferase (EC 2.7.7.80) (Adenylyltransferase MOCS3) (Sulfur carrier protein MOCS2A adenylyltransferase); Molybdopterin-synthase sulfurtransferase (EC 2.8.1.11) (Sulfur carrier protein MOCS2A sulfurtransferase) (Sulfurtransferase MOCS3)]
MTR	0.907	Q99707	METH_HUMAN	*MTR*	Methionine synthase (MS) (EC 2.1.1.13) (5-methyltetrahydrofolate--homocysteine methyltransferase) (Cobalamin-dependent methionine synthase) (Vitamin-B12 dependent methionine synthase)
RPF1	0.906	Q9H9Y2	RPF1_HUMAN	*RPF1 BXDC5*	Ribosome production factor 1 (Brix domain-containing protein 5) (Ribosome biogenesis protein RPF1)
PPIG	0.906	Q13427	PPIG_HUMAN	*PPIG*	Peptidyl-prolyl cis-trans isomerase G (PPIase G) (Peptidyl-prolyl isomerase G) (EC 5.2.1.8) (CASP10) (Clk-associating RS-cyclophilin) (CARS-Cyp) (CARS-cyclophilin) (SR-cyclophilin) (SR-cyp) (SRcyp) (Cyclophilin G) (Rotamase G)
PUS1	0.905	Q9Y606	TRUA_HUMAN	*PUS1 PP8985*	tRNA pseudouridine synthase A (EC 5.4.99.12) (tRNA pseudouridine(38-40) synthase) (tRNA pseudouridylate synthase I) (tRNA-uridine isomerase I)
SETD4	0.904	Q9NVD3	SETD4_HUMAN	*SETD4 C21orf18 C21orf27*	SET domain-containing protein 4 (EC 2.1.1.-) (EC 2.1.1.364)
MTO1	0.904	Q9Y2Z2	MTO1_HUMAN	*MTO1 CGI-02*	Protein MTO1 homolog, mitochondrial
PRMT3	0.903	O60678	ANM3_HUMAN	*PRMT3 HRMT1L3*	Protein arginine N-methyltransferase 3 (EC 2.1.1.-) (Heterogeneous nuclear ribonucleoprotein methyltransferase-like protein 3)
CTU2	0.903	Q2VPK5	CTU2_HUMAN	*CTU2 C16orf84 NCS2*	Cytoplasmic tRNA 2-thiolation protein 2 (Cytosolic thiouridylase subunit 2)
EZH2	0.903	Q15910	EZH2_HUMAN	*EZH2 KMT6*	Histone-lysine N-methyltransferase EZH2 (EC 2.1.1.356) (ENX-1) (Enhancer of zeste homolog 2) (Lysine N-methyltransferase 6)
WDR3	0.902	Q9UNX4	WDR3_HUMAN	*WDR3*	WD repeat-containing protein 3
FAM86C1	0.902	Q9NVL1	F86C1_HUMAN	*FAM86C1P FAM86C FAM86C1*	Putative protein FAM86C1P (EC 2.1.1.-) (Protein FAM86C)
PCMTD2	0.901	Q9NV79	PCMD2_HUMAN	*PCMTD2 C20orf36*	Protein-L-isoaspartate O-methyltransferase domain-containing protein 2
SSB	0.901	P05455	LA_HUMAN	*SSB*	Lupus La protein (La autoantigen) (La ribonucleoprotein) (Sjoegren syndrome type B antigen) (SS-B)
MPHOSPH10	0.900	O00566	MPP10_HUMAN	*MPHOSPH10 MPP10*	U3 small nucleolar ribonucleoprotein protein MPP10 (M phase phosphoprotein 10)
HEATR1	0.900	Q9H583	HEAT1_HUMAN	*HEATR1 BAP28 UTP10*	HEAT repeat-containing protein 1 (Protein BAP28) (U3 small nucleolar RNA-associated protein 10 homolog) [Cleaved into: HEAT repeat-containing protein 1, N-terminally processed]
ASH2L	0.900	Q9UBL3	ASH2L_HUMAN	*ASH2L ASH2L1*	Set1/Ash2 histone methyltransferase complex subunit ASH2 (ASH2-like protein)
METTL20	0.899	Q8IXQ9	ETKMT_HUMAN	*ETFBKMT C12orf72 METTL20*	Electron transfer flavoprotein beta subunit lysine methyltransferase (EC 2.1.1.-) (ETFB lysine methyltransferase) (ETFB-KMT) (Protein N-lysine methyltransferase METTL20)
POP4	0.899	O95707	RPP29_HUMAN	*POP4 RPP29*	Ribonuclease P protein subunit p29 (hPOP4) (EC 3.1.26.5)
RRP9	0.899	O43818	U3IP2_HUMAN	*RRP9 RNU3IP2 U355K*	U3 small nucleolar RNA-interacting protein 2 (RRP9 homolog) (U3 small nucleolar ribonucleoprotein-associated 55 kDa protein) (U3 snoRNP-associated 55 kDa protein) (U3-55K)

PRMT6	0.899	Q96LA8	ANM6_HUMAN	*PRMT6 HRMT1L6*	Protein arginine N-methyltransferase 6 (EC 2.1.1.319) (Heterogeneous nuclear ribonucleoprotein methyltransferase-like protein 6) (Histone-arginine N-methyltransferase PRMT6)
UPF2	0.899	Q9HAU5	RENT2_HUMAN	*UPF2 KIAA1408 RENT2*	Regulator of nonsense transcripts 2 (Nonsense mRNA reducing factor 2) (Up-frameshift suppressor 2 homolog) (hUpf2)
PRMT7	0.898	Q9NVM4	ANM7_HUMAN	*PRMT7 KIAA1933*	Protein arginine N-methyltransferase 7 (EC 2.1.1.321) (Histone-arginine N-methyltransferase PRMT7) ([Myelin basic protein]-arginine N-methyltransferase PRMT7)
TRNT1	0.898	Q96Q11	TRNT1_HUMAN	*TRNT1 CGI-47*	CCA tRNA nucleotidyltransferase 1, mitochondrial (EC 2.7.7.72) (Mitochondrial tRNA nucleotidyl transferase, CCA-adding) (mt CCA-adding enzyme) (mt tRNA CCA-diphosphorylase) (mt tRNA CCA-pyrophosphorylase) (mt tRNA adenylyltransferase)
SETD1B	0.898	Q9UPS6	SET1B_HUMAN	*SETD1B KIAA1076 KMT2G SET1B*	Histone-lysine N-methyltransferase SETD1B (EC 2.1.1.354) (Lysine N-methyltransferase 2G) (SET domain-containing protein 1B) (hSET1B)
UTP6	0.898	Q9NYH9	UTP6_HUMAN	*UTP6 C17orf40 HCA66 MHAT*	U3 small nucleolar RNA-associated protein 6 homolog (Hepatocellular carcinoma-associated antigen 66) (Multiple hat domains protein)
WDR36	0.898	Q8NI36	WDR36_HUMAN	*WDR36*	WD repeat-containing protein 36 (T-cell activation WD repeat-containing protein) (TA-WDRP)
NOL9	0.897	Q5SY16	NOL9_HUMAN	*NOL9*	Polynucleotide 5'-hydroxyl-kinase NOL9 (EC 2.7.1.-) (Nucleolar protein 9)
FARS2	0.897	O95363	SYFM_HUMAN	*FARS2 FARS1 HSPC320*	Phenylalanine--tRNA ligase, mitochondrial (EC 6.1.1.20) (Phenylalanyl-tRNA synthetase) (PheRS)
VCPKMT	0.896	Q9H867	MT21D_HUMAN	*VCPKMT C14orf138 METTL21D*	Protein-lysine methyltransferase METTL21D (EC 2.1.1.-) (Methyltransferase-like protein 21D) (VCP lysine methyltransferase) (VCP-KMT) (Valosin-containing protein lysine methyltransferase)
EXOSC8	0.896	Q96B26	EXOS8_HUMAN	*EXOSC8 OIP2 RRP43*	Exosome complex component RRP43 (Exosome component 8) (Opa-interacting protein 2) (OIP-2) (Ribosomal RNA-processing protein 43) (p9)
NOP56	0.896	O00567	NOP56_HUMAN	*NOP56 NOL5A*	Nucleolar protein 56 (Nucleolar protein 5A)
ASMTL	0.896	O95671	ASML_HUMAN	*ASMTL*	Probable bifunctional dTTP/UTP pyrophosphatase/methyltransferase protein [Includes: dTTP/UTP pyrophosphatase (dTTPase/UTPase) (EC 3.6.1.9) (Nucleoside triphosphate pyrophosphatase) (Nucleotide pyrophosphatase) (Nucleotide PPase); N-acetylserotonin O-methyltransferase-like protein (ASMTL) (EC 2.1.1.-)]
SMYD5	0.895	Q6GMV2	SMYD5_HUMAN	*SMYD5 RAI15*	SET and MYND domain-containing protein 5 (EC 2.1.1.-) (Protein NN8-4AG) (Retinoic acid-induced protein 15)
DNMT1	0.895	P26358	DNMT1_HUMAN	*DNMT1 AIM CXXC9 DNMT*	DNA (cytosine-5)-methyltransferase 1 (Dnmt1) (EC 2.1.1.37) (CXXC-type zinc finger protein 9) (DNA methyltransferase HsaI) (DNA MTase HsaI) (M.HsaI) (MCMT)
PRMT9	0.895	Q6P2P2	ANM9_HUMAN	*PRMT9 PRMT10*	Protein arginine N-methyltransferase 9 (Protein arginine N-methyltransferase 10) (EC 2.1.1.320)
PUS3	0.894	Q9BZE2	PUS3_HUMAN	*PUS3 FKSG32*	tRNA pseudouridine(38/39) synthase (EC 5.4.99.45) (tRNA pseudouridine synthase 3) (tRNA pseudouridylate synthase 3) (tRNA-uridine isomerase 3)
NDUFAF7	0.894	Q7L592	NDUF7_HUMAN	*NDUFAF7 C2orf56 PRO1853*	Protein arginine methyltransferase NDUFAF7, mitochondrial (EC 2.1.1.320) (NADH dehydrogenase [ubiquinone] complex I, assembly factor 7) (Protein midA homolog)
RTCB	0.894	Q9Y3I0	RTCB_HUMAN	*RTCB C22orf28 HSPC117*	RNA-splicing ligase RtcB homolog (EC 6.5.1.8) (3'-phosphate/5'-hydroxy nucleic acid ligase)
RRP1B	0.893	Q14684	RRP1B_HUMAN	*RRP1B KIAA0179*	Ribosomal RNA processing protein 1 homolog B (RRP1-like protein B)
N6AMT1	0.893	Q9Y5N5	N6MT1_HUMAN	*N6AMT1 C21orf127 HEMK2 KMT9 PRED28*	Methyltransferase N6AMT1 (HemK methyltransferase family member 2) (M.HsaHemK2P) (Lysine N-methyltransferase 9) (EC 2.1.1.-) (Methylarsonite methyltransferase N6AMT1) (EC 2.1.1.-) (Protein N(5)-glutamine methyltransferase) (EC 2.1.1.-)
DDX21	0.893	Q9NR30	DDX21_HUMAN	*DDX21*	Nucleolar RNA helicase 2 (EC 3.6.4.13) (DEAD box protein 21) (Gu-alpha) (Nucleolar RNA helicase Gu) (Nucleolar RNA helicase II) (RH II/Gu)
POLR2B	0.892	P30876	RPB2_HUMAN	*POLR2B*	DNA-directed RNA polymerase II subunit RPB2 (EC 2.7.7.6) (DNA-directed RNA polymerase II 140 kDa polypeptide) (DNA-directed RNA polymerase II subunit B) (RNA polymerase II subunit 2) (RNA polymerase II subunit B2)
DCAF13	0.892	Q9NV06	DCA13_HUMAN	*DCAF13 WDSOF1 HSPC064*	DDB1- and CUL4-associated factor 13 (WD repeat and SOF domain-containing protein 1)
NOL11	0.892	Q9H8H0	NOL11_HUMAN	*NOL11 L14*	Nucleolar protein 11
DHX15	0.891	O43143	DHX15_HUMAN	*DHX15 DBP1 DDX15*	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 (EC 3.6.4.13) (ATP-dependent RNA helicase #46) (DEAH box protein 15)
PRPF4B	0.890	Q13523	PRP4B_HUMAN	*PRPF4B KIAA0536 PRP4 PRP4H PRP4K*	Serine/threonine-protein kinase PRP4 homolog (EC 2.7.11.1) (PRP4 kinase) (PRP4 pre-mRNA-processing factor 4 homolog)
UTP18	0.890	Q9Y5J1	UTP18_HUMAN	*UTP18 WDR50 CDABP0061 CGI-48*	U3 small nucleolar RNA-associated protein 18 homolog (WD repeat-containing protein 50)
KARS	0.889	Q15046	SYK_HUMAN	*KARS1 KARS KIAA0070*	Lysine--tRNA ligase (EC 2.7.7.-) (EC 6.1.1.6) (Lysyl-tRNA synthetase) (LysRS)
METTL21A	0.889	Q8WXB1	MT21A_HUMAN	*METTL21A FAM119A HCA557B*	Protein N-lysine methyltransferase METTL21A (EC 2.1.1.-) (HSPA lysine methyltransferase) (HSPA-KMT) (Hepatocellular carcinoma-associated antigen 557b) (Methyltransferase-like protein 21A)
EXOSC5	0.889	Q9NQT4	EXOS5_HUMAN	*EXOSC5 CML28 RRP46*	Exosome complex component RRP46 (Chronic myelogenous leukemia tumor antigen 28) (Exosome component 5) (Ribosomal RNA-processing protein 46) (p12B)
NOL8	0.889	Q76FK4	NOL8_HUMAN	*NOL8 C9orf34 NOP132*	Nucleolar protein 8 (Nucleolar protein Nop132)
PCMTD1	0.888	Q96MG8	PCMD1_HUMAN	*PCMTD1*	Protein-L-isoaspartate O-methyltransferase domain-containing protein 1
KMT2B	0.888	Q9UMN6	KMT2B_HUMAN	*KMT2B HRX2 KIAA0304 MLL2 MLL4 TRX2 WBP7*	Histone-lysine N-methyltransferase 2B (Lysine N-methyltransferase 2B) (EC 2.1.1.354) (Myeloid/lymphoid or mixed-lineage leukemia protein 4) (Trithorax homolog 2) (WW domain-binding protein 7) (WBP-7)
UTP20	0.888	O75691	UTP20_HUMAN	*UTP20 DRIM*	Small subunit processome component 20 homolog (Down-regulated in metastasis protein) (Novel nucleolar protein 73) (NNP73) (Protein Key-1A6)
CIRH1A	0.888	Q969X6	UTP4_HUMAN	*UTP4 CIRH1A cPERP-E KIAA1988*	U3 small nucleolar RNA-associated protein 4 homolog (Cirhin) (UTP4 small subunit processome component)
CARM1	0.887	Q86X55	CARM1_HUMAN	*CARM1 PRMT4*	Histone-arginine methyltransferase CARM1 (EC 2.1.1.319) (Coactivator-associated arginine methyltransferase 1) (Protein arginine N-methyltransferase 4)
METTL25	0.887	Q8N6Q8	MET25_HUMAN	*METTL25 C12orf26*	Methyltransferase-like protein 25 (EC 2.1.1.-)

Table 6

Gene	Mean Pr GB	Mean Pr SVM	PPagerank Score	Rank
METTL13	0.944	0.904	0.000142	998
PRMT5	0.943	0.880	0.000235	476
RRP8	0.940	0.813	0.000637	75
METTL18	0.933	0.926	0.000047	4535
SETD2	0.933	0.838	0.000139	1022
RBBP5	0.930	0.898	0.000134	1074
SETDB1	0.929	0.843	0.000092	1841
PRDM15	0.929	0.697	0.000010	13477
SUZ12	0.928	0.768	0.000107	1502
SUV39H1	0.927	0.620	0.000103	1577
KRR1	0.927	0.916	0.000568	92
GART	0.926	0.909	0.000295	339
SNRPD3	0.926	0.916	0.000311	316
DIS3	0.922	0.920	0.000246	447
SUV39H2	0.922	0.827	0.000075	2568
WDR5	0.922	0.871	0.000165	818
PRDM4	0.920	0.724	0.000010	13449
EXOSC2	0.920	0.952	0.000522	121
PRMT1	0.918	0.855	0.000300	333
SKIV2L2	0.917	0.917	0.000840	15
UTP23	0.917	0.930	0.000500	136
FAM86A	0.917	0.759	0.000058	3616
RPP30	0.917	0.885	0.000290	348
EHMT1	0.917	0.922	0.000094	1775
METTL17	0.917	0.872	0.000060	3446
EXOSC9	0.917	0.849	0.000270	388
N6AMT2	0.916	0.702	0.000065	3179
DDX56	0.916	0.955	0.000710	47
TPMT	0.916	0.691	0.000017	10001
DPH5	0.915	0.775	0.000137	1042
SETD1A	0.915	0.696	0.000120	1263
UTP3	0.915	0.936	0.000598	88
SUV420H1	0.914	0.833	0.000066	3095
EED	0.912	0.911	0.000101	1608
DKC1	0.912	0.914	0.000686	60
METTL23	0.911	0.778	0.000024	8067
HEMK1	0.911	0.616	0.000296	336
PRDM10	0.910	0.664	0.000032	6570
POP1	0.910	0.917	0.000158	876
NSD1	0.910	0.754	0.000045	4833
KMT2D	0.910	0.677	0.000122	1247
SMYD4	0.909	0.684	0.000014	11347
MOCS3	0.909	0.834	0.000168	799
MTR	0.907	0.716	0.000048	4483
RPF1	0.906	0.843	0.000647	73
PPIG	0.906	0.908	0.000073	2649
PUS1	0.905	0.929	0.000500	137
SETD4	0.904	0.774	0.000242	459
MTO1	0.904	0.890	0.000180	723
PRMT3	0.903	0.887	0.000234	480
CTU2	0.903	0.749	0.000149	941
EZH2	0.903	0.675	0.000213	553
WDR3	0.902	0.865	0.000891	6
FAM86C1	0.902	0.780	0.000056	3757
PCMTD2	0.901	0.662	0.000033	6449
SSB	0.901	0.886	0.000197	616
MPHOSPH10	0.900	0.916	0.000571	91
HEATR1	0.900	0.888	0.000684	61
ASH2L	0.900	0.775	0.000104	1555
METTL20	0.899	0.596	0.000145	973
POP4	0.899	0.918	0.000166	812
RRP9	0.899	0.922	0.000790	23
PRMT6	0.899	0.700	0.000161	848
UPF2	0.899	0.893	0.000155	890
PRMT7	0.898	0.746	0.000039	5441
TRNT1	0.898	0.838	0.000213	555
SETD1B	0.898	0.454	0.000145	970
UTP6	0.898	0.917	0.000878	7
WDR36	0.898	0.917	0.000758	33
NOL9	0.897	0.689	0.000212	557
FARS2	0.897	0.801	0.000096	1737
VCPKMT	0.896	0.679	0.000077	2434
EXOSC8	0.896	0.894	0.000211	561
NOP56	0.896	0.929	0.000898	5
ASMTL	0.896	0.595	0.000145	974
SMYD5	0.895	0.721	0.000021	8896
DNMT1	0.895	0.743	0.000177	741
PRMT9	0.895	0.563	0.000018	9876
PUS3	0.894	0.840	0.000563	94
NDUFAF7	0.894	0.598	0.000199	607
RTCB	0.894	0.890	0.000036	5960
RRP1B	0.893	0.906	0.000505	130
N6AMT1	0.893	0.696	0.000385	222
DDX21	0.893	0.801	0.000372	242
POLR2B	0.892	0.916	0.000628	77
DCAF13	0.892	0.883	0.000669	65
NOL11	0.892	0.900	0.000236	472
DHX15	0.891	0.928	0.000741	37
PRPF4B	0.890	0.921	0.000052	4110
UTP18	0.890	0.881	0.000796	22
KARS	0.889	0.912	0.000267	396
METTL21A	0.889	0.638	0.000083	2195
EXOSC5	0.889	0.894	0.000308	320
NOL8	0.889	0.930	0.000048	4463
PCMTD1	0.888	0.375	0.000034	6166
KMT2B	0.888	0.669	0.000073	2651
UTP20	0.888	0.808	0.000360	255
CIRH1A	0.888	0.842	0.000708	48
CARM1	0.887	0.619	0.000125	1199
METTL25	0.887	0.580	0.000009	14177