# Pathformer: biological pathway informed Transformer model integrating multi-modal data

Xiaofan Liu[1,2], Yuhuan Tao[1,2], Zilin Cai[1], Pengfei Bao[1,2], Hongli Ma[1,2], Kexing Li[1], Yunping Zhu[3,] *, Zhi John Lu[1,2,] *

[1]MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China.

[2]Institute for Precision Medicine, Tsinghua University, Beijing 100084, China.

[3]State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, 38 Life Science Park, Changping District, Beijing 102206, China

*To whom correspondence should be addressed: Zhi John Lu, Tel: +86 10 62789217, E-mail: zhilu@tsinghua.edu.cn；Yunping Zhu, Tel.: +86 10 61777058, E-mail: zhuyunping@ncpsb.org.cn.

## Abstract

Multi-modal biological data can provide comprehensive views of gene regulation and cell development. However, conventional integration methods ignore prior biological knowledge and lack interpretability. Therefore, we developed Pathformer, a biological pathway informed deep learning model based on Transformer with bias to integrate multi-modal data. Pathformer leverages criss-cross attention mechanism to capture crosstalk between different biological pathways and modalities, and utilizes SHapley Additive Explanation method to reveal key pathways, genes, and regulatory mechanisms. Through benchmark studies on 28 TCGA datasets, we demonstrated the superior performance and interpretability of Pathformer on various biomedical classification tasks, compared to other models. Subsequently, Pathformer was successful applied to various multi-modal data from tissue, cells and blood for cancer subtyping, cell annotation and diseases noninvasive diagnosis. Pathformer identified many interesting disease-related deregulation events, such as scavenger receptors and DAP12 signaling transmission in cancer patients' body fluid, which are related to extracellular vesicle transport and immune response.

**Keywords:** Multi-modal integration; Transformer; Pathway crosstalk network; Cell annotation; Liquid biopsy.

# Introduction

The rapid progress in high-throughput technologies has made it possible to curate multi-modal data for disease studies using genome-wide platforms. These platforms can analyze different molecular alterations in the same samples, such as DNA variances (e.g., mutation, methylation, and copy number variance) and RNA alterations (e.g., expression, alternative promoter, splicing, and editing). Integrating these multi-modal data offers a more comprehensive view of gene regulation in diseases than analyzing single type of data[1]. For instance, multi-modal data integration is helpful in addressing certain key challenges of cancer diagnosis and prognosis, such as heterogeneity of intra- and inter-cancer, and complex molecular interactions[2]. Therefore, there is a pressing need for advanced computational methods that uncover interactions of multi-modal data.

Current algorithms for integrating multi-modal data can be broadly categorized into three groups: early integration models that merge multi-modal data into a single matrix[3,4], late integration models that process each modality separately and then combine their outputs through averaging or maximum voting[5,6], and intermediate integration models that dynamically merge multi-modal data[7,8]. Recently, instead of previous methods that mainly focus on unsupervised problems, several supervised algorithms have been proposed for different biological classification problem. For example, mixOmics uses latent component analysis to find common features among multi-modal data[9]. Wang et al. proposed multi-omics graph convolutional networks (MOGONet), a late integration model that uses graph convolutional networks for modal-specific learning and view correlation discovery network for multi-modal integration[10]. Moon et al. proposed two modal data integration and interpretation algorithm (MOMA) that utilizes attention mechanisms to extract important modules[11]. These methods rely on computational inference to capture relationship between modalities, but ignore the immensely informative prior biological knowledge such as regulatory networks.

To improve the interpretability, several studies have attempted to incorporate prior biological knowledge into deep learning models for multi-modal data integration. For instance, Ma et al. proposed a visible neural network that combines with biological pathways to model the impact of gene interactions on yeast cell growth[12]. Meanwhile, pathway-associated sparse deep neural network (PASNet) was utilized to accurately predict the prognosis of glioblastoma multiforme (GBM) patients[13]. Recently, a sparse neural network integrating multiple molecular features based on a multilevel view of biological pathways, P-net, was published for the classification of prostate cancer patients[14]. Another method, PathCNN, was developed to predict survival of GBM patient by using principal component analysis (PCA) algorithms to define multi-modal pathway images and a convolutional neural network[15]. However, these algorithms rarely considered the synergy and nonlinear relationships between pathways. Given the

complexity of biological systems, understanding the pathway crosstalk is crucial for comprehending more complex biological procession[16], which can help deep learning models better capture multi-modal interactions.

Inspired by these prior works, we propose Pathformer, which combines pathway crosstalk networks and the Transformer encoder with bias for the interpretation and classification of multi-modal data. Recently, Transformer has demonstrated its capability in handling multi-modal tasks in computational fields[17]. It hasn't been applied to the biological multi-modal data for lack of reliable biological embedding methods and solutions to the memory explosion posed by the vast amount of gene inputs. These challenges are addressed by Pathformer. First, Pathformer uses multiple statistical indicators of multi-modal data as gene embedding, which comprehensively describes different perspectives of gene information. Second, Pathformer utilizes a sparse neural network based on prior pathway knowledge to transform gene embeddings into pathway embeddings, which not only captures valuable information but also addresses memory explosion issue. Third, Pathformer incorporates pathway crosstalk networks into the Transformer model with bias to enhance the exchange of information between different modalities and pathways.

As far as we are aware, Pathformer is the first biological multi-modal integration model that combines prior pathways knowledge and Transformer encoder model. We evaluated Pathformer on 28 benchmark datasets of the Cancer Genome Atlas (TCGA)[18] and demonstrated its superior performance and biological interpretability on various classification tasks, compared to other integration models. Furthermore, we applied Pathformer to tissue, cells and liquid biopsy multi-modal data integration in cancer subtyping, cell annotation and cancer diagnosis with high accuracy. Finally, we utilized biological interpretability of Pathformer to reveal interesting pathways, genes and regulatory mechanisms related to diseases in human tissue, cells and plasma.

# Results

## Framework of Pathformer

Pathformer is an interpretable multi-modal integration model that can be applied in different biomedical classification tasks, including cancer diagnosis, cancer subtype classification, cancer stage classification, cancer survival analysis, and single cell annotation (**Fig. 1a**). Pathformer utilizes biological pathway network and a Transformer encoder to allow better information fusion and better classification for downstream tasks. Specifically, Pathformer uses biological multi-modal data and biological pathway information as input, and define biological multi-modal embedding (gene embedding and pathway embedding). It then enhances the fusion of information between various modalities and pathways by combining pathway crosstalk networks with Transformer encoder.

Finally, a fully connected layer serves as the classifier for different downstream classification tasks. In addition, Pathformer uses biological interpretable module to identity important features and regulatory processes for enhancing biological significance of the model (**Fig. 1a,** see **Methods** for details).

We curated all pathways from four public databases, then selected 1,497 pathways based on the criterion of gene number, overlap ratio with other pathways, and the number of pathway subsets. Next, we used *BinoX*[19], a classic tool for crosstalk analysis, to calculate the crosstalk relationships among the 1,497 pathways. Based on these relationships, we created a pathway crosstalk network as Pathformer's input (see **Methods** and **Supplementary Notes**).

Multi-modal biological data preprocessing and embedding are crucial components of Pathformer (**Fig. 1b**). Pathformer supports any type and quantityof multi-modal data as input. We preprocessed the raw sequence reads into multi-modal data, including different omics or different variant features of the same omics (see **Methods** and **Supplementary Notes**). These multi-modal data are on different levels, such as nucleotide level, fragment level, and gene level, which significantly influence data integration. To address this, we used multiple statistical indicators as gene embeddings to retain the gene diversity across different modalities (see **Fig. 1b** and **Methods**). Subsequently, we used the known gene-pathway mapping relationship to develop a sparse neural network based on prior pathway knowledge (PSNN) to transform gene embedding into pathway embedding. The PSNN has two layers representing genes and pathways, respectively. These two layers are not fully connected, but rather share a connection pruned based on the pathway and gene inclusion relationships. If there is no correlation between a given gene and a given pathway, the connection weight between two neurons is set to be 0; otherwise, it is learned through training (see **Methods**). Therefore, pathway embedding is a dynamic embedding method. The PSNN can not only restore the mapping relationship between genes and pathways, but also identify important genes in different pathways through trained weights, and can transfer the complementarity of modalities at the gene level to the pathway level. Additionally, this biological multi-modal embedding step does not require additional gene selection, thereby avoiding bias and overfitting problems resulting from artificial feature selection.

Transformer module with pathway crosstalk network bias is the key module of Pathformer model (**Fig. 1c**). Inspired by the Evoformer model used in AlphaFold2[20] for processing multiple sequences, we developed the Transformer module based on criss-cross attention (CC-attention) with bias for data fusion of pathways and modalities. Particularly, multi-head column-wise self-attention (col-attention) is used to enhance the exchange of information between pathways, with the pathway crosstalk network matrix serving as the bias for col-attention to guide the flow of information. Multi-head row-wise self-attention (row-attention) is employed to facilitate

116 information exchange between different modalities, and the updated multi-modal embedding matrix is used to
117 update the pathway crosstalk network matrix by calculating the correlation between pathways. More details of the
118 Transformer module are described in **Methods**.

119 In addition, to comprehend Pathformer's decision-making process, we used averaging attention maps in row-
120 attention to represent the contributions of different modalities, and SHapley Additive exPlanations[21] (SHAP value)
121 to decipher the important pathways and their key genes (see **Methods**). SHAP value is a post hoc model
122 interpretation method that assigns an importance value to each feature to explain the relationship between features
123 and classification[21]. In addition, the z-score of SHAP values of different modalities for each pathway and gene can
124 demonstrate modal complementarity at the gene level and the pathway level. Finally, the hub module of the updated
125 pathway crosstalk network represents the most critical regulatory mechanism in classification, and is screened by
126 sub-network scores based on SHAP values of pathways. Links of the updated network indicate crosstalk
127 relationships that affect classification tasks (see **Methods**).

## 128 Pathformer outperforms existing multi-modal integration methods in various classification 129 tasks using benchmark datasets

130 To evaluate the performance of Pathformer, we tested model on various cancer classification tasks as benchmark
131 studies: cancer early- and late- stage classification (10 TCGA cancer datasets), low- and high- survival risk
132 classification (10 TCGA cancer datasets), and cancer subtype classification (8 TCGA cancer datasets) (see
133 **Supplementary Fig. 1** and **Supplementary Notes**). For these tasks, DNA methylation, DNA copy nunumber
134 variation (CNV), and RNA expression were used as input. For model training and test, we performed 2 times 5-fold
135 cross-validation that divided the data into a discovery set (75%) and a validation set (25%) for each test (see
136 **Supplementary Fig. 1** and **Methods**). We first optimized hyperparameters using 5-fold cross-validation on the
137 discovery set, with macro-averaged F1 score as the criterion for grid search. The results of optimal hyperparameter
138 combination for each dataset are listed in **Supplementary Fig. 2** and **Supplementary Table 1**. Then, we trained
139 Pathformer using the discovery set with early stopping and tested it on the validation set.

140 We compared the classification performance of Pathformer with several existing multi-modal integration
141 methods, including early integration methods based on base classifiers, i.e., nearest neighbor algorithm (KNN),
142 support vector machine (SVM), logistic regression (LR), random forest (RF), and extreme gradient boosting
143 (XGBoost); late integration methods based on KNN, SVM, LR, RF, and XGBoost; partial least squares-discriminant
144 analysis (PLSDA) and sparse partial least squares-discriminant analysis (sPLSDA) of mixOmics[9]; two deep

145 learning-based integration methods, MOGONet[10] and PathCNN[15]. MOGONet is a multi-modal integration method
146 based on graph convolutional neural network. PathCNN is a representative multi-modal integration method that
147 combines pathway information. During comparison methods, the multi-modal data were preprocessed with the
148 statistical indicators and features were prefiltered with ANOVA as input (see **Supplementary Notes**).

149 Pathformer consistently outperformed the other integration methods in most classification tasks, evaluated by
150 macro-averaged F1 score (F1score_macro) (**Fig. 2**), as well as area under the receiver operating characteristic curve
151 (AUC) and average F1 score weighted by support (F1score_weighted) (**Supplementary Fig. 3** and **Supplementary**
152 **Table 2**). We showed F1score_macro in the main figure because it is a more robust measurement than the other two
153 scores for the imbalanced classes. In the cancer stage classification and survival classification tasks, Pathformer
154 achieved the best F1score_macro and F1score_weighted in all the 10 datasets, and the best AUC in 8 of 10 datasets.
155 In cancer subtype classification of TCGA, Pathformer achieved the best F1score_macro in 7 of 8 datasets, the best
156 F1score_weighted in 6 of 8 datasets, and the best AUC in 6 of 8 datasets. Notably, Pathformer substantially
157 outperformed the other methods in the challenging classification tasks like cancer early- and late- stage classification
158 and low- and high- survival risk classification, showing average increases of 11% and 15% in F1score_marco
159 compared with XGBoost, respectively. This highlights Pathformer's exceptional learning ability. Moreover, in terms
160 of stability, Pathformer also showed significantly better generalization ability than the other deep learning
161 algorithms, as indicated by the cross-validation variances (**Supplementary Fig. 4**).

## 162 Ablation analysis shows that Pathformer benefits from multi-modal integration, attention 163 mechanism and pathway crosstalk network

164 We used ablation analysis to evaluate the essentialities of each type of data and each module of model in the multi-
165 model data integration of Pathformer, based on nine datasets of cancer early- and late- stage classification. First, we
166 evaluated the essentialities of seven different data inputs, including RNA expression, DNA methylation, DNA CNV,
167 and a combination thereof (**Fig. 3a**). By comparing the classification performances of seven models, we discovered
168 that the model with all three modalities as input achieved the best performance, followed by RNA expression-only
169 and DNA methylation-only model. Furthermore, we observed that the performances of models with single modality
170 can vary greatly between datasets. For example, DNA methylation-only model performed better than RNA
171 expression-only and DNA CNV-only in the KIRC dataset, but the opposite performances were observed in the
172 LUAD dataset. These findings suggest that different modalities have disparate behaviors in different cancer types,
173 and emphasized the necessity of multi-modal data integration in various cancer classification tasks.

Next, we also evaluated the essentialities of different modules in Pathformer. We developed 4 models, namely CC-attention, Transformer, PSNN, and NN, which successively remove one to multiple modules of Pathformer. CC-attention is a model without pathway crosstalk network bias. Transformer is a model without either pathway crosstalk network bias or row-attention. PSNN is a model that directly uses classification module with pathway embedding as input. NN is a model that directly uses classification module with gene embedding as input. As shown in **Fig. 3b**, the complete Pathformer model achieved the best classification performance, while the performance of CC-Attention, Transformer, PSNN, and NN decreased successively. Transformer had a significantly lower classification performance compared to CC-Attention, but no significant improvement compared to PSNN. This indicates that the criss-cross attention mechanism (**Fig. 1c**) plays a key role in Pathformer, with respect to information fusion and crosstalk between different biological pathways and between different modalities (i.e., multi-omics).

## Application of Pathformer to breast cancer subtype classification

From above results of benchmark tasks, we knew that applying Pathformer to breast cancer tissue multi-modal data for subtype classification can achieve better classification results than other methods. Besides, model interpretability is often more important in practical applications. Next, we demonstrated that Pathformer has good interpretability in breast cancer subtype classification, using attention weights and SHAP value to reveal breast cancer subtype related modalities, pathways, and genes (**Fig. 4**).

First, at the modality level, we visualized the contributions of different modalities for breast cancer subtype classification by the attention weights (**Fig. 4a**). The contribution of transcriptomic data was greater than 50% in breast cancer subtype classification, which is consistent with the fact that PAM50 is defined based on transcriptomic data[22]. Combining with the results of stage classification and survival classification using breast cancer tissue multi-modal data (**Supplementary Figs. 5a, 6a**), we observed that transcriptome always played a crucial role in various classification tasks; DNA CNV had greater contribution in subtype classification; and DNA methylation contributed more substantially in early- and late- stage classification. In addition, the contributions of various statistical indicators in the same modality were also different for different classification tasks. These findings further validated the necessity of multi-modal integration and biological multi-modal embedding.

Next, at the pathway and gene level, we identified key pathways with top 15 SHAP value and key genes with top 5 SHAP value of each pathway in breast cancer subtype classification (**Fig. 4b**). Then, we presented a hub module of the updated pathway crosstalk network (**Supplementary Figs. 7**). Here, *complex I biogenesis* pathway

203 which participates in cancer cell proliferation and metastasis[23] was identified as the most critical pathway in the
204 classification and a key node in the hub module of the updated pathway crosstalk network. Five mitochondrial genes
205 (MT-ND3, MT-ND1, MT-ND4, MT-ND2, and MT-ND6) were identified as key genes of the *complex I biogenesis*
206 pathway in breast cancer subtype classification by Pathformer, which is consistent with the literatures[24]. In addition,
207 in the hub module of the updated pathway crosstalk network (**Supplementary Figs. 7**), *complex I biogenesis*
208 pathway, *TP53-regulated metabolic genes* pathway[25] and and *signaling by ERBB4* pathway[26], these famous
209 biomarkers, has been identified as the most critical regulatory mechanism for breast cancer subtype classification.

210 Moreover, many other important pathways identified by Pathformer for breast cancer subtype classification
211 have also already been reported previously (**Fig. 4b**). For example, the expression of *nucleotide excision repair*
212 pathway is reduced in TNBC, which may affect survival after platinum chemotherapy of patients[27]. RFC4 is the key
213 gene of this pathway, and DNA CNV of RFC4 was reported to play a crucial role in determining individual breast
214 cancer subtypes[28], which is consistent with the prediction of the gene's pillar module by Pathformer. E2F1, the key
215 gene of *transcription of E2F targets under negative control by p107 and p130 in complex with HDAC1* pathway, its
216 expressions are significantly up-regulated in TNBC [29], and its DNV CNV is also a susceptibility factor for breast
217 cancer[30], again consistent with Pathformer.

218 Subsequently, to facilitate a more intuitive comparison between each subtype, we present the updated pathway
219 embeddings of all samples, where pathways are composed of the top 10 SHAP value-ranked pathways for each
220 subtype (**Fig. 4c**). From the heatmap, we can observe several representative subtype-specific pathway clusters. For
221 instance, among the HER2 positive subtype-specific pathways, the most prominent ones are the *PI3K events in*
222 *ERBB2 signaling* pathway and *downregulation of ERBB2 signaling* pathway, which is consistent with previous
223 reports[31,32]. Additionally, the data also show that Luminal A subtype-specific pathway cluster exhibit a high
224 correlation with DNA damage and repair processes[33].

## 225 Application of Pathformer to single-cell multi-omics data

226 With the development of single-cell sequencing technology, there is an increasing demand for researchers to
227 annotate cell types by integrating single-cell multi-omics data. As a universal multi-modal integration framework,
228 Pathformer can effectively solve the challenges posed by high sparsity and high noise with prior pathway knowledge.
229 Therefore, Pathformer can be applied not only to bulk data, but also to single-cell multi-omics data for cell
230 annotation. We applied Pathformer to a single-cell benchmark dataset consisting of single-cell chromatin
231 accessibility data (scATAC-seq data) and single-cell gene expression (scRNA-seq data) from a mixture of human

232 cell lines, i.e., BJ, H1, GM12878 and K562 (see **Methods**).

233 Pathformer utilizes a sparse neural network based on prior pathway knowledge to integrate genes from the
234 same pathway, thereby converting gene embeddings into pathway embeddings. This not only resolves the memory
235 overflow issue of the Transformer module caused by long inputs but also reduces data loss rate and heterogeneity.
236 In the tested single-cell dataset, we compared dimensionality reduction based on pathway embedding and gene
237 embedding, demonstrating that the pathway embedding updated by Pathformer better annotated different types of
238 cells than the gene embedding (**Fig. 5a**). We also compared the performance between Pathformer based on the
239 pathway embedding and neural networks based on the gene embedding, further confirming that pathway embedding
240 enhances the accuracy of cell type prediction (**Fig. 5b**).

241 To understand the decision-making process of classification for BJ, H1, K562, and GM12878 cell mixtures,
242 we can employ Pathformer's biological interpretability module to extract the most critical pathways, genes and
243 regulatory mechanisms that contribute to cell annotation (**Fig. 5c-d**). *ADP signaling through P2Y purinoceptor 1*
244 pathway is evidently the most crucial pathway in human cell mixtures classification and serves as a significant node
245 in the core regulatory network. According to the literature, there is a physical interaction between P2Y receptor and
246 K562 cell-specific G protein[34], which participates in the process of hematopoietic differentiation and K562 cell
247 proliferation and growth, aligning with the pathway network relationship predicted by Pathformer. Additionally, in
248 the hub module (**Fig. 5d**), the crosstalk relationship between *ADP signaling through P2Y purinoceptor 1* pathway
249 and *AP-1* pathway, as the main skeleton of network, is generated by Pathformer. Although there is no direct evidence
250 of a regulatory relationship between P2Y receptor and AP-1, both are associated with cell proliferation and
251 migration processes[34,35], indicating a potential synergistic role in these cellular behaviors. In summary, Pathformer
252 visualizs key pathways in cell annotation and identifies potential crosstalk relationships.

253 Finally, to validate the robustness of Pathformer, we tested the model on an independent single-cell multi-
254 omics dataset, including 74 K562 cells and 475 other cell types (see **Methods**). Utilizing the trained Pathformer
255 model, we accurately predicted the K562 cells in this independent validation dataset (**Fig. 5e**).

## 256 Application of Pathformer to liquid biopsy

257 Liquid biopsy is a non-invasive detection way with important clinical applications such as disease (e.g., cancer)
258 diagnosis and status monitoring[36]. RNA alterations reflect the complementarity between different levels of
259 information and help to overcome missed detection results of single data to further improve the accuracy of disease
260 diagnosis[37]. Therefore, we used Pathformer to integrate multi-modal data of cell-free RNA-seq (cfRNA-seq) in

261 blood for classifying cancer patients from healthy controls. We curated three types of cfRNA-seq datasets derived
262 from three different blood components: plasma, extracellular vesicle (EV), and platelet (see **Methods**).
263 We first calculated seven RNA-level modalities as Pathformer's input, including RNA expression, RNA
264 splicing, RNA editing, RNA alternative promoter (RNA alt. promoter), RNA allele-specific expression (RNA ASE),
265 RNA single nucleotide variations (RNA SNV), and chimeric RNA. From results of 5-fold cross-validation in
266 **Supplementary Fig. 8**, we found that the model with all modalities as input had the best comprehensive
267 performance on three datasets, followed by RNA expression-only model and RNA alt. promoter-only model, and
268 some models with other modalities exhibited great fluctuations on different datasets. In order to effectively integrate
269 information without redundancy, we performed further feature selection based on different modality combinations
270 evaluated by Pathformer. First, we calculated the contributions of each modality and its corresponding statistical
271 indicators (**Fig. 6a**). Similar to the results of cross-validation, RNA expression was the core modality across all
272 datasets. Next, we performed 5-fold cross-validation find an optimal modality combination for each dataset (**Fig.
273 6b**). We found that plasma dataset with 7 modalities, EV dataset with 3 modalities, and platelet dataset with 3
274 modalities obtained the best performance with AUCs above 0.9 and sensitivity above 40% on 99% specificity. It is
275 worth noting that Pathformer maintained the same level of predictive accuracy even for early-stage cancer plasma
276 samples (**Supplementary Fig. 9**). In conclusion, Pathformer effectively integrated multi-modal data from human
277 plasma, and accurately classified cancer patients from healthy controls.

## 278 Pathformer reveals deregulated pathways and genes in cancer patients' plasma

279 Because the Pathformer model has biological interpretability, we used Pathformer to predict cancer related pathways
280 and genes in the above liquid biopsy data, with a particular focus on the plasma data (**Fig. 7**). Then, we can gain
281 insight into the deregulated alterations in body fluid (i.e., plasma) for cancer patients vs. healthy controls.
282 First, in comparison to cancer tissue data (**Fig. 4, Supplementary Fig. 6-7**), we found that vesicle transport
283 and coagulation related pathways occupied an important position in plasma datasets, which is consistent with the
284 characteristics of body fluid (**Fig. 7a**). Furthermore, we also observed that active pathways and key genes of plasma
285 dataset were more similar to those in platelet dataset (**Fig. 7a, Supplementary Fig. 10a**), which is consistent with
286 a recent report showing platelet is a major origin in the plasma cell-free transcriptome[38].
287 Next, we examined there interesting pathways. The first is *binding and uptake of ligands* (e.g., oxidized low-
288 density lipoprotein, oxLDL) *by scavenger receptors* pathway was identified as the most active pathway in both
289 plasma and EV datasets (**Fig. 7a, Supplementary Fig. 11a**). It is well established that scavenger receptors play a

290 crucial role in cancer prognosis and carcinogenesis by promoting the degradation of harmful substances and
291 accelerating the immune response[39]. Scavenger receptors are also closely related to the transport process of vesicles,
292 such as stabilin-1[40]. Meanwhile, HBB, HBA1, HBA2, FTH1, HSP90AA1 were identified as key genes in this
293 pathway. HBB, as a known biomarker, has been demonstrated to be downregulated in gastric cancer blood
294 transcriptomics[41]. HSP90AA1 is also a potential biomarker for various cancers[42], especially in the blood[43]. The
295 other interesting pathways are *DAP12 signaling* pathway and *DAP12 interactions* pathway revealed in both plasma
296 and platelet datasets (**Fig. 7a, Supplementary Fig. 10a**). DAP12 triggers natural killer cell immune responses
297 against certain tumor cells[44], which is regulated by platelet[45]. Among the top 5 key genes of DAP12 related pathway,
298 B2M was reported as a serum protein encoding gene and a widely recognized tumor biomarker[46]; HLA-E and HLA-
299 B were reported as cancer biomarkers in tissue and plasma[47,48].

300 In addition, Pathformer provides insight into the interplay between various biological processes and their
301 impact on cancer progression by updating pathway crosstalk network (**Fig. 7b**). In the plasma data, the link between
302 *binding and uptake of ligands by scavenger receptors* pathway and *iron uptake and transport* pathway was a novel
303 addition to the updated network. The crosstalk between two pathways was amplified by Pathformer in plasma
304 dataset, probably because they were important for classification and shared the same key gene, FTH1, one of two
305 intersecting genes between the two pathways. In summary, Pathformer's updated pathway crosstalk network
306 visualizes the information flow between pathways related to cancer classification task in the liquid biopsy data,
307 providing novel insight into the cross-talk of biological pathways in cancer patients' plasma.

# 308 Discussion

309 Pathformer utilizes a biological multi-modal embedding (**Fig. 1b**) based on pathway-based sparse neural network,
310 providing a demonstration of applying Transformer model on biological multi-modal data integration. Particularly,
311 we showed that the criss-cross attention mechanism (**Fig. 1c**) contributed to the classification tasks by capturing
312 crosstalk between biological pathways and potential regulation between modalities (i.e., multi-omics).

313 ***Applications of Pathformer.*** Pathformers will be widely applied to various biomedical classification
314 problems, handling diverse types of multi-modal data, such as tissue, blood, and single-cell data. We have
315 demonstrated that Pathformer leverages prior pathway knowledge to alleviate high sparsity and noise in single-cell
316 data, enabling accurate cell annotation, and capturing core regulatory mechanisms for classification, such as G
317 protein-coupled processes in leukemia cells. Furthermore, Pathformer plays a particularly prominent role in clinical
318 applications, such as cancer subtyping, staging, prognosis, and diagnosis. For instance, we have demonstrated

excellent performance of Pathformer on noninvasive diagnosis of cancer based on multi-modal data of liquid biopsy. The accuracies (AUC scores) of cancer classification in plasma, EV, and platelet cfRNA-seq datasets were all higher than 90%. Moreover, the interpretability of the Pathformer model can help researchers gain insights into the complex regulation processes involved in cancer, such as scavenger receptors related pathway, DAP12 related pathway and other active pathways consistent with the characteristics of body fluid data.

***Limitations of Pathformer and future directions.*** Pathformer used genes involved in pathways from four public databases, all of which consist of protein-coding genes. However, a substantial body of literature has reported that noncoding RNAs are also crucial in cancer prognosis and diagnosis[49]. Therefore, incorporating noncoding RNAs and their related functional pathways into Pathformer would be a potential future work. Another flaw of Pathformer is the computing memory issue. Pathway embedding of Pathformer has prevented memory overflow of Transformer module caused by long inputs. However, when adding more pathways or gene sets (e.g., transcription factors), Pathformer still faces the issue of memory overflow. In the future work, we may introduce linear attention to further improve computational speed.

# Methods

## Data collection and preprocessing

For benchmark testing, we collected 28 datasets across different cancer types from TCGA to evaluate classification performance of Pathformer and existing comparison methods, which consists of 8 datasets for cancer subtype classification, 10 datasets for cancer early- and late- stage classification, and 10 datasets for cancer low- and high-survival risk classification. To verify the effect of Pathformer in cell annotation, we collected tow single cell multi omics benchmark datasets: one was comprised1030 cells generated by SNARE-seq[50], the other as independent validation set was comprised 549 cells generated by scCAT-seq[51]. Besides, to further verify the effect of Pathformer in cancer diagnosis, we also collected three types of body fluid datasets: the plasma dataset (comprising 373 samples assayed by total cell-free RNA-seq[52,53]), the extracellular vesicle (EV) dataset (comprising 477 samples from two studies assayed by exosomal RNA-seq[54,55]), and the platelet dataset (comprising 918 sample from two studies assayed by tumor-educated blood platelet RNA-seq[56,57]). Through our biological information pipeline, totally 4, 2 and 7 biological modalities are obtained for TCGA dataset, single-cell dataset and liquid biopsy dataset, respectively. More details are described in **Supplementary Fig. 1** and **Supplementary Notes**.

## The Pathformer model

As shown in **Fig. 1**, Pathformer model consists of the following six modules: biological pathway input, pathway crosstalk network calculation, multi-modal data input, biological multi-modal embedding, Transformer module with pathway crosstalk network bias, and classification module.

***Biological pathways and crosstalk network***

We collected 2,289 pathways of four public databases including Kyoto Encyclopedia of Genes and Genomes database (KEGG)[58], Pathway Interaction database (PID)[59], Reactome database (Reactome)[60], and BioCarta Pathways database (BioCarta)[61]. Then, we filtered these pathways by three criteria: gene number, the overlap ratio with other pathways (the proportion of genes in the pathway that are also present in other pathways), and the number of pathway subsets (the number of pathways included in the pathway). Following the principle of moderate size and minimal overlap with other pathway information, we selected 1,497 pathways with gene number between 15 and 100, or gene number greater than 15 and overlap ratio less than 1, or gene number greater than 15 and the number of pathway subsets less than 5. Next, we used *BinoX* to calculate the crosstalk relationship of pathways and build a pathway crosstalk network with adjacency matrix $\boldsymbol{P} \in \mathbb{R}^{N_p \times N_p}$, $N_p$=1,497 (more details in **Supplementary Notes**).

***Biological multi-modal data input and embedding***

Pathformer supports any number of modalities as input which may have different dimensions, including nucleotide level, fragment level, and gene level. Modalities and their dimensions level for different datasets are described in **Supplementary Table 3**. We represented multi-modal input matrix of a sample as $\boldsymbol{M}$ , and converted matrix $\boldsymbol{M}$ into gene encoding $\boldsymbol{E}_G$ and pathway encoding $\boldsymbol{E}_P$. First, we used a series of statistical indicators in different modalities as gene embedding. These statistical indicators include gene level score, count, entropy, minimum, maximum, mean, weighted mean in whole gene, and weighted mean in window. Gene embedding is calculated as follows:

$$\boldsymbol{E}_G = \boldsymbol{F}_E(\boldsymbol{M}) = \left[f_{\boldsymbol{E}_1}(\boldsymbol{G}_1),\ f_{\boldsymbol{E}_2}(\boldsymbol{G}_2), \cdots, f_{\boldsymbol{E}_m}(\boldsymbol{G}_m)\right] \in \mathbb{R}^{N_g \times D_g}$$

, where $\boldsymbol{G}_i$ is modality $i$, $D_g$ is length of gene embedding for all modalities, $\boldsymbol{F}_E$ is a series of gene embedding functions. $\boldsymbol{F}_E$ uses a series of statistical indicators to uniformly convert the multi-modal data into the gene level, and the embedding functions corresponding to different modalities and different datasets are different (more details in **Supplementary Table 3** and **Supplementary Notes**). Then, we used the known biological pathways to construct a sparse neural network for converting the gene embedding $\boldsymbol{E}_G$ into the pathway embedding $\boldsymbol{E}_P$, as described below:

$$\boldsymbol{E_P} = \boldsymbol{W}_{\boldsymbol{sparse}}^{\boldsymbol{T}}\boldsymbol{E_G} + \boldsymbol{B}, \boldsymbol{E_P} \in \mathbb{R}^{N_p \times D_p}$$

, where $N_p$ is the number of pathways, $D_p = D_g$ is the length of pathway embedding, $\boldsymbol{W}_{sparse} \in \mathbb{R}^{N_g \times N_p}$ is a learnable sparse weight matrix, and $\boldsymbol{B}$ is a bias term. $\boldsymbol{W}_{sparse}$ is constructed based on the known relationship between pathways and genes. When the given gene and the pathway are irrelevant, the corresponding element of $\boldsymbol{W}_{sparse}$ will always be 0. Otherwise, it needs to be learned through training.

***Transformer module with pathway crosstalk network bias***

We employed the Transformer module based on criss-cross attention with pathway crosstalk network bias, which has 3 blocks. Each block of Transformer module contains the following processes: multi-head column-wise self-attention (col-attention), multi-head row-wise self-attention (row-attention), layer normalization, GELU activation, residual connection, and network update.

Multi-head column-wise self-attention contains 8 heads, each head is a mapping of $\boldsymbol{Q_1}$,$\boldsymbol{K_1}$,$\boldsymbol{V_1}$,$\boldsymbol{P}$, which are query vector, key vector, and value vector of multi-modal embedding and pathway crosstalk network matrix, respectively. First, we represented the $h$th column-wise self-attention by $\boldsymbol{A}_{col}^{(h)}$, calculated as follows:

$$\boldsymbol{A}_1^{(h)} = (\boldsymbol{Q}_1\boldsymbol{K}_1^T)/\sqrt{d}$$

$$\boldsymbol{A}_{col}^{(h)} = \mathrm{dropout}_{0.2}(\mathrm{softmax}(\boldsymbol{A}_1^{(h)} + \boldsymbol{P})) \cdot \boldsymbol{V}_1^{(h)}$$

, where $h = 1,2,\cdots,H$ is the $h$th head; $H$ is the number of heads; $\boldsymbol{Q}_1 = \boldsymbol{E}_P\boldsymbol{W}_{Q_1}^{(h)}$ , $\boldsymbol{K}_1 = \boldsymbol{E}_P\boldsymbol{W}_{K_1}^{(h)}$, $\boldsymbol{V}_1 = \boldsymbol{E}_P\boldsymbol{W}_{V_1}^{(h)}$ are

linear transformations of the input $\boldsymbol{E}_P$; $\boldsymbol{W}_{Q_1}^{(h)} \in \mathbb{R}^{D_p \times d}$, $\boldsymbol{W}_{K_1}^{(h)} \in \mathbb{R}^{D_p \times d}$, $\boldsymbol{W}_{V_1}^{(h)} \in \mathbb{R}^{D_p \times d}$ are the weight matrices as parameters; *d* is the attention dimension; $\mathrm{dropout}_{0.2}$ is a dropout neural network layer with a probability of 0.2; and softmax is the normalized exponential function.

Next, we merged multi-head column-wise self-attention and performed a series of operations as follows:

$$\boldsymbol{g}_1^{(h)} = sigmoid(\boldsymbol{E}_P \boldsymbol{W}_{g_1}^{(h)})$$

$$\boldsymbol{U}_1 = \sum_{h=1}^{H} (\boldsymbol{g}_1^{(h)} \circ \boldsymbol{A}_{col}^{(h)}) \cdot \boldsymbol{W}_{U_1}^{(h)}$$

$$\boldsymbol{U}_1' = \boldsymbol{U}_1 + \boldsymbol{E}_P$$

$$\boldsymbol{O}_1 = \mathrm{dropout}_{0.2}(\mathrm{GELU}(\mathrm{LayerNorm}(\boldsymbol{U}_1') \cdot \boldsymbol{W}_{O_{11}})) \cdot \boldsymbol{W}_{O_{12}} + \boldsymbol{U}_1'$$

, where $h = 1,2,\cdots,H$ is the *h*th head; *H* is the number of heads; $\circ$ is the matrix dot product; $\boldsymbol{W}_{g_1}^{(h)} \in \mathbb{R}^{D_p \times d}$, $\boldsymbol{W}_{U_1}^{(h)} \in \mathbb{R}^{d \times D_p}$, $\boldsymbol{W}_{O_{11}} \in \mathbb{R}^{D_p \times o}$, $\boldsymbol{W}_{O_{12}} \in \mathbb{R}^{o \times D_p}$ are the weight matrices as parameters; *o* is a constant; LayerNorm is the layer normalization function; GELU is the distortion of RELU activation function; and $\mathrm{dropout}_{0.2}$is a dropout neural network layer with a probability of 0.2.

Multi-head row-wise self-attention enables information exchange between different modalities. It is a regular dot-product attention without pathway crosstalk network bias. It also contains 8 heads, and the *h*th row-wise self-attention, i.e., $\boldsymbol{A}_{row}^{(h)}$, is calculated as follows:

$$\boldsymbol{A}_2^{(h)} = (\boldsymbol{Q}_2 \boldsymbol{K}_2^T)/\sqrt{d}$$

$$\boldsymbol{A}_{row}^{(h)} = \mathrm{dropout}_{0.2}(\mathrm{softmax}(\boldsymbol{A}_2^{(h)})) \cdot \boldsymbol{V}_2^{(h)}$$

, where $h = 1,2,\cdots$, h is the *h*th head; *H* is the number of heads; $\boldsymbol{Q}_2 = \boldsymbol{E}_P^T \boldsymbol{W}_{Q_2}^{(h)}$ , $\boldsymbol{K}_2 = \boldsymbol{E}_P^T \boldsymbol{W}_{K_2}^{(h)}$, $\boldsymbol{V}_2 = \boldsymbol{E}_P^T \boldsymbol{W}_{V_2}^{(h)}$ are linear transformations of the input $\boldsymbol{E}_P^T$; $\boldsymbol{W}_{Q_2}^{(h)} \in \mathbb{R}^{N_p \times d}$, $\boldsymbol{W}_{K_2}^{(h)} \in \mathbb{R}^{N_p \times d}$, $\boldsymbol{W}_{V_2}^{(h)} \in \mathbb{R}^{N_p \times d}$ are the weight matrices as parameters; *d* is the attention dimension; $\mathrm{dropout}_{0.2}$ is a dropout neural network layer with a probability of 0.2; and softmax is the normalized exponential function.

Subsequently, we merged multi-head row-wise self-attention and performed a series of operations. The formulas are as follows:

$$\boldsymbol{g}_2^{(h)} = sigmoid(\boldsymbol{E}_P^T \boldsymbol{W}_{g_2}^{(h)})$$

$$\boldsymbol{U}_2 = \sum_{h=1}^{H} (\boldsymbol{g}_2^{(h)} \circ \boldsymbol{A}_{row}^{(h)}) \cdot \boldsymbol{W}_{U_2}^{(h)}$$

$$\boldsymbol{U}_2' = \boldsymbol{\beta} * \boldsymbol{U}_2 + \boldsymbol{E}_P^T$$

$$\boldsymbol{O}_2 = dropout_{0.2}(\mathrm{GELU}(\mathrm{LayerNorm}(\boldsymbol{U}_2') \cdot \boldsymbol{W}_{O_{21}})) \cdot \boldsymbol{W}_{O_{22}} + \boldsymbol{U}_2'$$

, where $h = 1,2,\cdots,$ h is the *h*th head; *H* is the number of heads; $\circ$ is the matrix dot product; $\boldsymbol{W}_{g_2}^{(h)} \in \mathbb{R}^{N_p \times d}$, $\boldsymbol{W}_{U_2}^{(h)} \in \mathbb{R}^{d \times N_p}$, $\boldsymbol{W}_{O_{21}} \in \mathbb{R}^{N_p \times o}$, $\boldsymbol{W}_{O_{22}} \in \mathbb{R}^{o \times N_p}$ are the weight matrices as parameters; $o$ is a constant; $\boldsymbol{\beta}$ is a constant coefficient for row-attention; LayerNorm is the layer normalization function; GELU is the distortion of RELU activation function; and $dropout_{0.2}$ is a dropout neural network layer with a probability of 0.2. $\boldsymbol{O}_2$ is pathway embedding input of the next Transformer block. In other words, when $\boldsymbol{E}_P$ is $\boldsymbol{E}_P^{(0)}$, $\boldsymbol{O}_2$ is $\boldsymbol{E}_P^{(1)}$, superscripts with parenthesis represent data at different block.

Then, we used the updated pathway embedding $\boldsymbol{O}_2$ to update the pathway crosstalk network. We exploited the correlation between embedding vectors of two pathways to update the corresponding element of the pathway crosstalk network matrix. The formula is as follows:

$$\boldsymbol{P}' = (\boldsymbol{P} \cdot \boldsymbol{P}^T)/N_p$$

, where $\boldsymbol{P}'$ is the updated pathway crosstalk network matrix of next Transformer block. In other words, when $\boldsymbol{P}'$ is $\boldsymbol{P}^{(1)}$, $\boldsymbol{P}$ is $\boldsymbol{P}^{(0)}$, superscripts with parenthesis represent data at different block.

***Classification module***

In order to solve the classification tasks, we used the fully connected neural network as the classification module to transform pathway embedding encoded by the Transformer module into the probability for each label. Three fully connected neural networks each have 300, 200, and 100 neurons, with dropout probability $dropout_c$, which is hyperparameter. More details of the classification module are described in **Supplementary Notes**.

## Model training and test

In this study, we implemented Pathformer's network architecture using the "PyTorch" package in *Python* v3.6.9, and our codes can be found in the GitHub repository (https://github.com/lulab/Pathformer). For model training and test, we divided the labeled dataset into the discovery set (75%) and the validation set (25%) hierarchically. We implemented model training, hyperparameter optimization and model early stopping on the discovery set and tested on the validation set (**Supplementary Fig. 1**).

When training the model, we used a normal model learning strategy. We applied cross-entropy loss with class-imbalance weight as the label prediction loss, the ADAM optimizer to train Pathformer, and the cosine annealing learning rate method to optimized learning rate. For hyperparameter optimization, we used grid search with 5-fold cross-validation in the discovery set with the macro-averaged F1 score as the selection criterion. For early stopping,

we divided the discovery set into the training set (75%) and the test set (25%) hierarchically, and used the macro-averaged F1 score of the test set as the criterion for stopping training. When testing the model, we used the best model trained with optimal hyperparametric combination in the validation set. More details of model training and test are described in **Supplementary Notes**.

## Model interpretability

To better understand Pathformer's decisions, we increased the interpretability of Pathformer by calculating contributions of different modalities, important pathways and their key genes, and hub module of the updated pathway crosstalk network.

***Contribution of each modality***

In Pathformer, row-attention is used to facilitate information interaction between different modalities, that is, row-attention map can represent the importance of each modality. According to the trained model, we obtained row-attention maps of 8 heads in 3 blocks for each sample. For the contribution of each modality, we first integrated all matrices of row-attention maps into one matrix by element-wise average. Then, we averaged this average row-attention matrix along with columns as the attention weights of modalities, i.e., the contribution of modalities. The calculation is as follows:

$$\boldsymbol{A}_{aver} = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{BL}\sum_{b=1}^{BL}\frac{1}{H}\sum_{h=1}^{H} softmax([[\boldsymbol{A}_2^{(h)}]^{(b)}]^{(n)})$$

$$attention\ weight_i = \frac{1}{D_p}\sum_{j=1}^{D_p} a_{ij},\ a_{ij} \text{ is the } i\text{th row and the } j\text{th columns of } \boldsymbol{A}_{aver}$$

, where $N$ is the number of samples, $BL$ is the number of blocks, $H$ is the number of heads, softmax is a normalized exponential function, and $attention\ weight_i$ is the attention weight of dimension $i$ of pathway embedding.

***Important pathways and their key genes***

SHapley Additive exPlanations[21] (SHAP) is an additive explanation model inspired by coalitional game theory, which regards all features as "contributors". SHAP value is the value assigned to each feature, which explains the relationship between pathways, genes and classification, implemented by "SHAP" package of *Python* v3.6.9.

Specifically, we calculated SHAP values of the gene embedding and the pathway embedding encoded by Transformer module corresponding to each sample and each category, denoted as $\boldsymbol{S}_{gn}^{(j)} \in \mathbb{R}^{D_p}$ and $\boldsymbol{S}_{pn}^{(j)} \in \mathbb{R}^{D_p}$ respectively. The SHAP values of genes and pathways are calculated as follows:

$$\text{SHAP}_g = \sum_{j=1}^{d_{out}}\sum_{e=1}^{D_p}\frac{1}{N}\sum_{n=1}^{N}\left|s_{gne}^{(j)}\right|, s_{gie}^{(j)} \in \boldsymbol{S}_{gi}^{(j)}$$

$$\mathrm{SHAP}_p = \sum_{j=1}^{d_{out}} \sum_{e=1}^{D_p} \frac{1}{N} \sum_{n=1}^{N} \left| s_{pne}^{(j)} \right|, s_{pie}^{(j)} \in \boldsymbol{S}_{pi}^{(j)}$$

, where $g = 1,2,\cdots,N_g$ is the $g$th gene, $g = 1,2,\cdots,N_p$ is the $p$th pathway, $n = 1,2,\cdots,N$ is the $n$th sample, $e = 1,2,\cdots,D_p$ is dimension $e$ of pathway embedding, and $j = 1,2,\cdots,d_{out}$ is the $j$th category of sample.

In addition, we calculated SHAP values of pathways and genes in different modalities, described as follows:

$$\mathrm{SHAP}_{gi} = \sum_{j=1}^{d_{out}} \sum_{e=e_1+\cdots+e_{i-1}}^{e_i} \frac{1}{N} \sum_{n=1}^{N} \left| s_{gne}^{(j)} \right|, s_{gie}^{(j)} \in \boldsymbol{S}_{gi}^{(j)}$$

$$\mathrm{SHAP}_{pi} = \sum_{j=1}^{d_{out}} \sum_{e=e_1+\cdots+e_{i-1}}^{e_i} \frac{1}{N} \sum_{n=1}^{N} \left| s_{pne}^{(j)} \right|, s_{pie}^{(j)} \in \boldsymbol{S}_{pi}^{(j)}$$

, where $i = 1,\cdots,m$ is the $i$th modality, $e_i$ is the length of gene embedding and pathway embedding for modality $i$.

Finally, pathways with the top 15 SHAP values in the classification task are considered as important pathways. For each pathway, genes with top 5 SHAP values are considered as the key genes of the pathway. The core modality on which one gene depends indicates that the SHAP value of that gene ranks higher on this modality than on the others.

***Hub module of the updated pathway crosstalk network***

In Pathformer, pathway crosstalk network matrix is used to guide the direction of information flow, and updated according to encoded pathway embedding in each Transformer block. Therefore, the updated pathway crosstalk network contains not only prior information but also multi-modal data information, which represents the specific regulatory mechanism in each classification task. We defined the sub-network score through SHAP value of each pathway in sub-network, so as to find foremost sub-network for prediction, that is, hub module of the updated pathway crosstalk network. The calculation of the sub-network score can be divided into four steps: average pathway crosstalk network matrix calculation, network pruning, sub-network boundary determination, and score calculation. More details of sub-network score calculations are described in **Supplementary Notes**.

# Declarations

**Data availability**

All datasets used in this study are publicly available for academic research usages. The details of usage are also fully illustrated in Methods and Supplementary Notes.

**Code availability**

Source code for data preprocessing and model training is freely available at Github (https://github.com/lulab/Pathformer) with detailed instructions. Source code for comparing the other methods is also included.

**Consent for publication**

All authors have approved the manuscript and agree with the publication.

**Competing interests**

The authors declare that they have no competing interests.

**Funding and Acknowledgements**

This work is supported by National Natural Science Foundation of China (32170671), National Key Research Program of China (2021YFA1301603), Tsinghua University Guoqiang Institute Grant (2021GQG1020), Tsinghua University Initiative Scientific Research Program of Precision Medicine (2022ZLA003), Bioinformatics Platform of National Center for Protein Sciences (Beijing) (2021-NCPSB-005). This study was also supported by Bayer Micro-funding, Beijing Advanced Innovation Center for Structural Biology, Bio-Computing Platform of Tsinghua University Branch of China National Center for Protein Sciences.

*Funding for open access charge: Tsinghua University Guoqiang Institute Grant (2021GQG1020).*

# References


1 Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome biology* **18**, 1-15 (2017).

2 Tarazona, S., Arzalluz-Luque, A. & Conesa, A. Undisclosed, unmet and neglected challenges in multi-omics studies. *Nature Computational Science* **1**, 395-402 (2021).

3 Shen, R., Olshen, A. B. & Ladanyi, M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* **25**, 2906-2912 (2009).

4 Lando, M. *et al.* Gene dosage, expression, and ontology analysis identifies driver genes in the carcinogenesis and chemoradioresistance of cervical cancer. *PLoS genetics* **5**, e1000719 (2009).

5 Cabassi, A. & Kirk, P. D. Multiple kernel learning for integrative consensus clustering of omic datasets. *Bioinformatics* **36**, 4789-4796 (2020).

6 Wang, T. *et al.* MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nature Communications* **12**, 1-13 (2021).

7 Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nature methods* **11**, 333-337 (2014).

8 Lock, E. F., Hoadley, K. A., Marron, J. S. & Nobel, A. B. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The annals of applied statistics* **7**, 523 (2013).

9 Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology* **13**, e1005752 (2017).

10 Wang, T. *et al.* MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. *Nat Commun* **12**, 3445 (2021). https://doi.org:10.1038/s41467-021-23774-w

11 Moon, S. & Lee, H. MOMA: A Multi-Task Attention Learning Algorithm for Multi-Omics Data Interpretation and Classification. *Bioinformatics* (2022). https://doi.org:10.1093/bioinformatics/btac080

12 Ma, J. *et al.* Using deep learning to model the hierarchical structure and function of a cell. *Nature methods* **15**, 290-298 (2018).

13 Hao, J., Kim, Y., Kim, T. K. & Kang, M. PASNet: pathway-associated sparse deep neural network for prognosis prediction from high-throughput data. *BMC Bioinformatics* **19**, 510 (2018). https://doi.org:10.1186/s12859-018-2500-z

14 Elmarakeby, H. A. *et al.* Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348-352 (2021).

15 Oh, J. H. *et al.* PathCNN: interpretable convolutional neural networks for survival prediction and pathway analysis applied to glioblastoma. *Bioinformatics* **37**, i443-i450 (2021). https://doi.org:10.1093/bioinformatics/btab285

16 Li, Y., Agarwal, P. & Rajagopalan, D. A global pathway crosstalk network. *Bioinformatics* **24**, 1442-1447 (2008).

17 Hu, R. & Singh, A. in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 1439-1449.

18 Cancer Genome Atlas Research Network, J. The cancer genome atlas pan-cancer analysis project. *Nat. Genet* **45**, 1113-1120 (2013).

19 Ogris, C., Guala, D., Helleday, T. & Sonnhammer, E. L. A novel method for crosstalk analysis of biological networks: improving accuracy of pathway annotation. *Nucleic acids research* **45**, e8-e8 (2017).

20 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).

21 Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30** (2017).
22 Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology* **27**, 1160 (2009).
23 Urra, F. A., Muñoz, F., Lovy, A. & Cárdenas, C. The mitochondrial complex (I) ty of cancer. *Frontiers in oncology* **7**, 118 (2017).
24 Kopinski, P. K., Singh, L. N., Zhang, S., Lott, M. T. & Wallace, D. C. Mitochondrial DNA variation and cancer. *Nature Reviews Cancer* **21**, 431-445 (2021).
25 Silwal-Pandit, L. *et al.* TP53 Mutation Spectrum in Breast Cancer Is Subtype Specific and Has Distinct Prognostic RelevanceTP53 in Breast Cancer. *Clinical Cancer Research* **20**, 3569-3580 (2014).
26 Sundvall, M. *et al.* Role of ErbB4 in breast cancer. *Journal of mammary gland biology and neoplasia* **13**, 259-268 (2008).
27 Ribeiro, E. *et al.* Triple negative breast cancers have a reduced expression of DNA repair genes. *PLoS One* **8**, e66243 (2013).
28 Srihari, S. *et al.* Understanding the functional impact of copy number alterations in breast cancer using a network modeling approach. *Molecular BioSystems* **12**, 963-972 (2016).
29 Oshi, M. *et al.* The E2F pathway score as a predictive biomarker of response to neoadjuvant therapy in ER+/HER2− breast cancer. *Cells* **9**, 1643 (2020).
30 Rocca, M. S. *et al.* E2F1 copy number variations in germline and breast cancer: a retrospective study of 222 Italian women. *Molecular Medicine* **27**, 1-7 (2021).
31 Harari, D. & Yarden, Y. Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene* **19**, 6102-6114 (2000).
32 Vasan, N., Toska, E. & Scaltriti, M. Overview of the relevance of PI3K pathway in HR-positive breast cancer. *Annals of oncology* **30**, x3-x11 (2019).
33 Davis, J. D. & Lin, S.-Y. DNA damage and breast cancer. *World journal of clinical oncology* **2**, 329 (2011).
34 Kotevic, I., Kirschner, K. M., Porzig, H. & Baltensperger, K. Constitutive interaction of the P2Y2 receptor with the hematopoietic cell-specific G protein Gα16 and evidence for receptor oligomers. *Cellular signalling* **17**, 869-880 (2005).
35 Rosson, D. & O'Brien, T. G. AP-1 activity affects the levels of induced erythroid and megakaryocytic differentiation of K562 cells. *Archives of biochemistry and biophysics* **352**, 298-305 (1998).
36 Schwarzenbach, H., Hoon, D. S. & Pantel, K. Cell-free nucleic acids as biomarkers in cancer patients. *Nature Reviews Cancer* **11**, 426-437 (2011).
37 6 7 Lehmann Kjong-Van 3 4 6 7 Liu Fenglin 10 Shiraishi Yuichi 11 Soulette Cameron M. 12 Urban Lara 2, P. T. C. G. C. C. D. N. R. D. D. F. N. A. H. Y. K. A. *et al.* Genomic basis for RNA alterations in cancer. *Nature* **578**, 129-136 (2020).
38 Vorperian, S. K. *et al.* Cell types of origin of the cell-free transcriptome. *Nature biotechnology* **40**, 855-861 (2022).
39 Ryu, S., Howland, A., Song, B., Youn, C. & Song, P. I. Scavenger receptor class A to E involved in various cancers. *Chonnam medical journal* **56**, 1-5 (2020).
40 Kzhyshkowska, J., Gratchev, A. & Goerdt, S. Stabilin‐1, a homeostatic scavenger receptor with multiple functions. *Journal of cellular and molecular medicine* **10**, 635-649 (2006).
41 Lee, I.-S. *et al.* A blood-based transcriptomic signature for noninvasive diagnosis of gastric cancer. *British Journal of Cancer* **125**, 846-853 (2021).
42 Zuehlke, A. D., Beebe, K., Neckers, L. & Prince, T. Regulation and function of the human HSP90AA1 gene. *Gene* **570**, 8-16 (2015).
43 Zhang, P. j. *et al.* Genes expression profiling of peripheral blood cells of patients with hepatocellular carcinoma. *Cell biology international* **36**, 803-809 (2012).

44 Campbell, K. S. & Colonna, M. DAP12: a key accessory protein for relaying signals by natural killer cell receptors. *The international journal of biochemistry & cell biology* **31**, 631-636 (1999).
45 Placke, T., Kopp, H.-G. & Salih, H. R. Modulation of natural killer cell anti-tumor reactivity by platelets. *Journal of innate immunity* **3**, 374-382 (2011).
46 Cooper, E. & Plesner, T. Beta‐2‐microglobulin review: Its relevance in clinical oncology. *Medical and Pediatric Oncology* **8**, 323-334 (1980).
47 Zeestraten, E. *et al.* Combined analysis of HLA class I, HLA-E and HLA-G predicts prognosis in colon cancer patients. *British journal of cancer* **110**, 459-468 (2014).
48 Liu, L. *et al.* A three-platelet mRNA set: MAX, MTURN and HLA-B as biomarker for lung cancer. *Journal of Cancer Research and Clinical Oncology* **145**, 2713-2723 (2019).
49 Qi, P., Zhou, X.-y. & Du, X. Circulating long non-coding RNAs in cancer: current status and future perspectives. *Molecular cancer* **15**, 1-11 (2016).
50 Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature biotechnology* **37**, 1452-1457 (2019).
51 Liu, L. *et al.* Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nature communications* **10**, 470 (2019).
52 Chen, S. *et al.* Cancer type classification using plasma cell-free RNAs derived from human and microbes. *eLife* **11**, e75181 (2022).
53 Tao, Y. *et al.* Cell-free multi-omics analysis reveals tumor status-informative signatures in gastrointestinal cancer patients' plasma. *bioRxiv*, 2023.2001. 2031.526431 (2023).
54 Li, S. *et al.* exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic acids research* **46**, D106-D112 (2018).
55 Yu, S. *et al.* Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for the detection of pancreatic ductal adenocarcinoma. *Gut* **69**, 540-550 (2020).
56 Best, M. G. *et al.* RNA-Seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer cell* **28**, 666-676 (2015).
57 Best, M. G. *et al.* Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer cell* **32**, 238-252. e239 (2017).
58 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).
59 Schaefer, C. F. *et al.* PID: the pathway interaction database. *Nucleic acids research* **37**, D674-D679 (2009).
60 Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research* **39**, D691-D697 (2010).
61 Nishimura, D. BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient* **2**, 117-120 (2001).

# Figure Legends

**Figure 1. Overview of Pathformer.**

**a.** Schematic of the Pathformer framework and its applications. Pathformer integrates biological multi-modal data based on pathway crosstalk information, which can be applied to different biological classification problem and can identity key genes, pathways, modalities, and regulatory processes for better biological interpretability. **b.** Schematic of biological multi-modal embedding. Biological multi-modal data has different levels, which can be different omics or different variant features of the same omics. Pathformer supports any type and quantity of multi-modal data as input, including single modal data as well. $F_E$, statistical indicators in the gene embedding; circles, neurons in the neural network; arrows, represent the direction of information flow; G, gene; P, pathway; W, weight of pathway-based sparse neural network. The weights of the pathway-based sparse neural network represent the importance of different genes in different pathways. **c**. A block of Transformer module with pathway crosstalk network bias (3 blocks used in **a**). The pathway embedding matrix is used as input and the pathway crosstalk network matrix is used as bias. $N_p$, number of pathways; $D_p$, dimensionality of pathway embedding; $h$, number of attention heads; $d$, attention dimension; $\boldsymbol{V}_1$, $\boldsymbol{K}_1$, $\boldsymbol{Q}_1$, $\boldsymbol{A}_1$: vale, key, query and attention map of col-attention; $\boldsymbol{V}_2$, $\boldsymbol{K}_2$, $\boldsymbol{Q}_2$, $\boldsymbol{A}_2$: vale, key, query and attention map of row-attention; $+$, element-wise addition; $\times$, matrix multiplication; $\circ$, matrix dot product; $\beta$, constant coefficient for row-attention.

**Figure 2. Performance comparison between Pathformer and other multi-modal integration methods**

Bar charts show the macro-averaged F1 score of different multi-modal integration methods in different classification tasks of TCGA datasets. Error bars are from 2 times 5-fold cross-validation, representing 95% confidence intervals. XGBoost refers to the early integration methods based on gradient boosted tree, while XGBoost (late) refers to the late integration methods based on gradient boosted tree.

**Figure 3. Ablation analysis of Pathformer for the classification of early- and late-stage cancer patients.**

**a.** Different types of data (modalities) were used as input for TCGA cancer early- and late-stage classification. **b.** Ablation analysis of different modules in Pathformer. Error bars are from 2 times 5-fold cross-validation across 8 datasets, representing 95% confidence intervals. CC-attention, Pathformer without pathway crosstalk network bias; Transformer, Pathformer without either pathway crosstalk network bias or row-attention; PSNN, Pathformer without Transformer module; NN, classification module only.

682

683 **Figure 4. Breast cancer subtype related modalities, pathways and genes revealed by Pathformer.**

684 **a.** Contributions of different modalities for breast cancer (BRCA) subtype classification calculated by attention
685 weights (averaging attention maps of row-attention). **b.** Important pathways and their key genes with top SHapley
686 Additive exPlanations (SHAP) values for BRCA subtype classification. Among the key genes, different colors
687 represent different pillar modalities of the genes. **c**. Heatmap for the pathway embedding processed by Transformer
688 module on BRCA subtype. Pathways are composed of top 10 pathways with the highest SHAP value for each
689 subtype.

690

691 **Figure 5. Pathformer integrates single-cell multi-omics data for cell annotation.**

692 **a.** Clustering of different cell types based on pathway embedding processed by Transformer module and original
693 gene embedding. Cells are colored by the predicted values of Pathformer, where red represents cells that were
694 wrongly assigned. **b.** Performance comparison between Pathformer and neural networks based on gene embedding.
695 Here, AUC is measured as the macro-averaged AUC of the one-vs-rest scheme. **c.** Important pathways and their
696 key genes with top SHAP values for single cell annotation. Among the key genes, different colors represent different
697 pillar modalities of the genes. **d.** A hub module of the updated pathway crosstalk network for the cell annotation.
698 Color depth and size of node represent degree of node. Line thickness represents weight of edge. All links are
699 predicted by Pathformer, where the known links were reported previously and the new links are new predictions. **e.**
700 Performance of Pathformer for K562 cells' prediction on an independent validation dataset.

701

702 **Figure 6. Pathformer integrates multi-modal liquid biopsy data for non-invasive cancer diagnosis.**

703 **a**. Contributions of different input features and their statistical indicators when classifying cancer patients from
704 healthy controls using three liquid biopsy datasets (cell free RNA-seq). All mean represents the sum of mean,
705 weighted mean and window weighted mean. Each type of RNA splicing is the sum of all statistical indicators in
706 this type. **b**. Classification performance of different input combinations on three liquid biopsy datasets. Each value
707 is the mean of 5-fold cross-validation.

708

709 **Figure 7. Interpretation of cancer patients' plasma data using Pathformer.**

710 **a**. Important pathways and their key genes revealed by Pathformer in the plasma cell free RNA-seq data when
711 classifying cancer patients from healthy controls. The pathways and their key genes were selected with top SHAP

values. Among the key genes, different colors represent different pillar modalities (e.g., RNA expression, RNA editing, etc) of the genes. **b**. Hub modules of pathway crosstalk network are shown for plasma cell free RNA-seq data. Color depth and size of node represent the degree of node. Line thickness represents the weight of edge. All links are predicted by Pathformer, where known links are reported by the initial crosstalk network and new links are new predictions.

# Figure 1. Overview of Pathformer

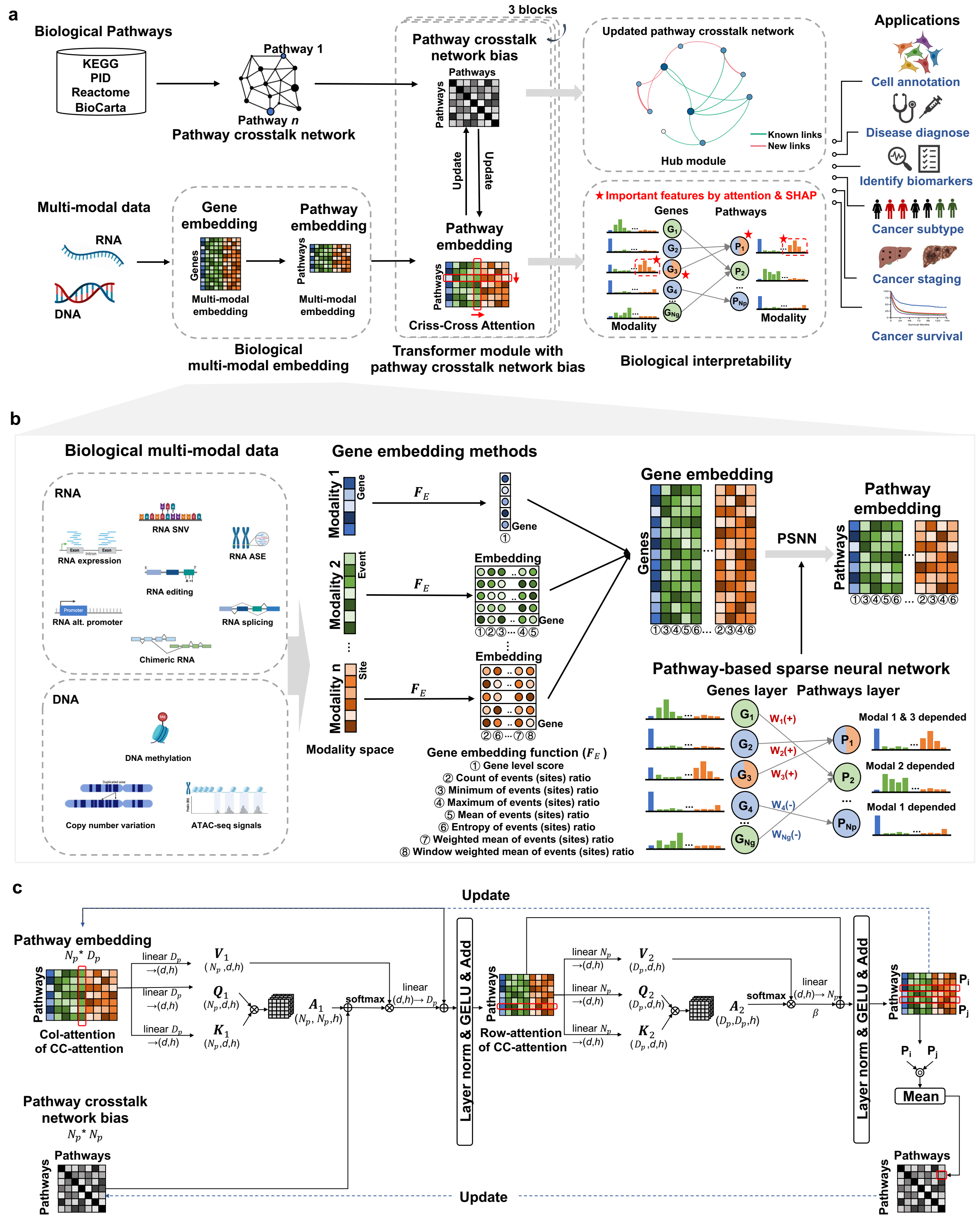


**Figure 2. Performance comparison between Pathformer and other multi-modal integration methods**

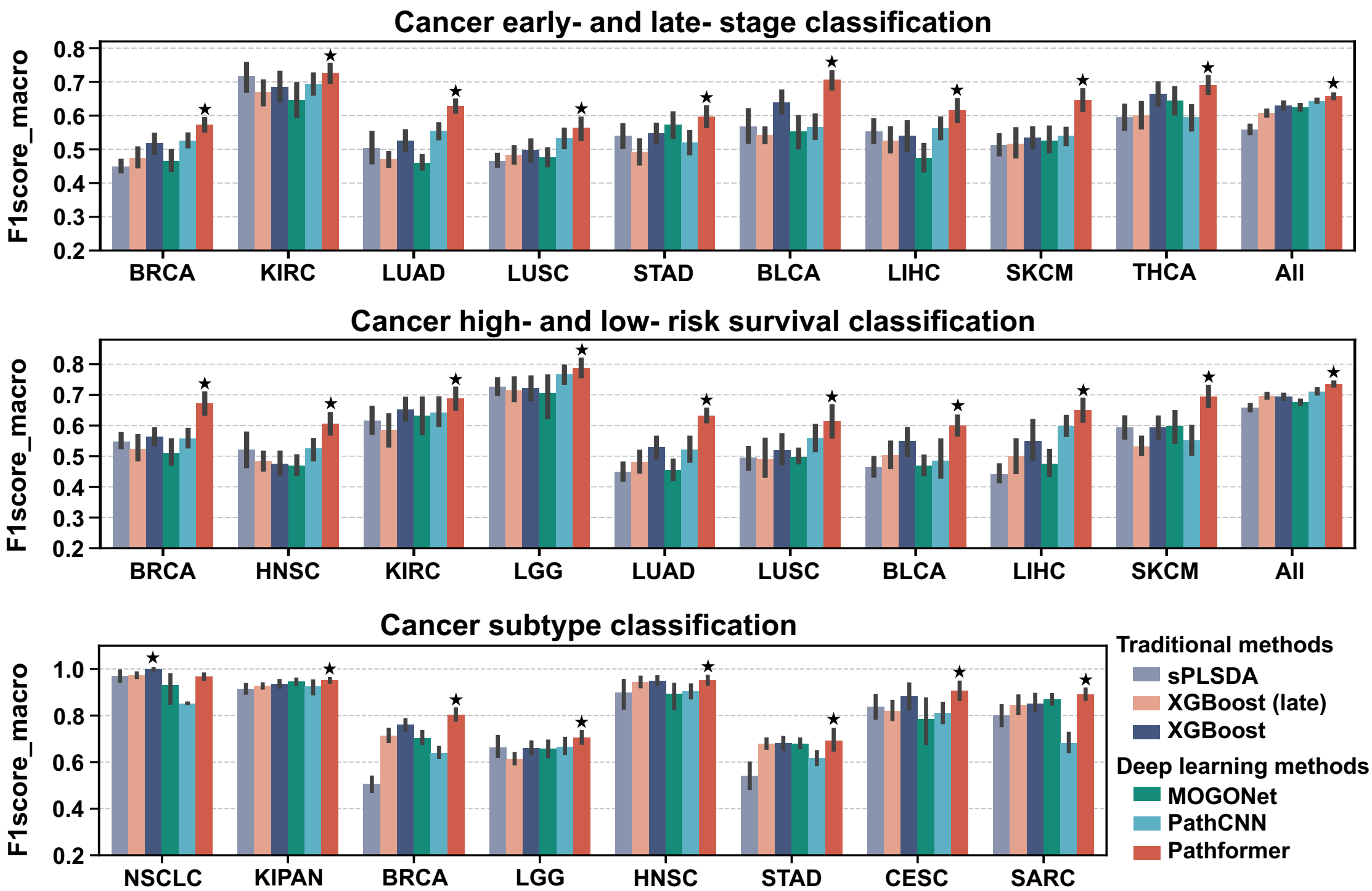


**Figure 3. Ablation analysis of Pathformer for the classification of early- and late-stage cancer patients**

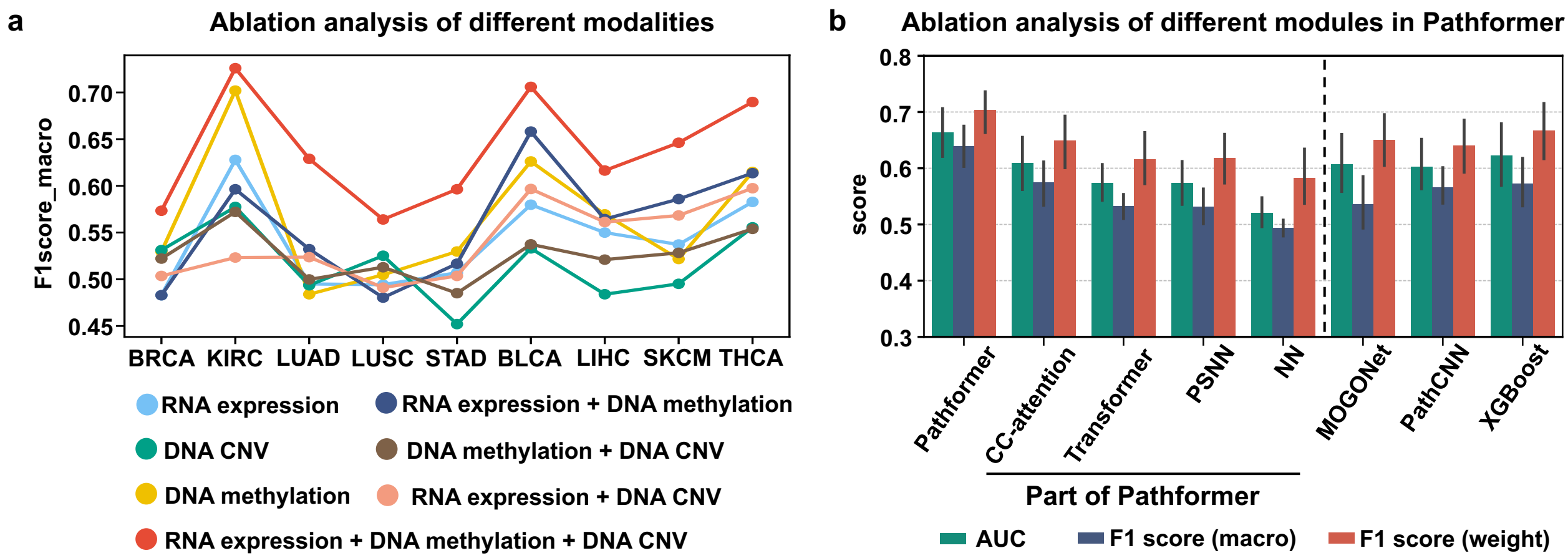
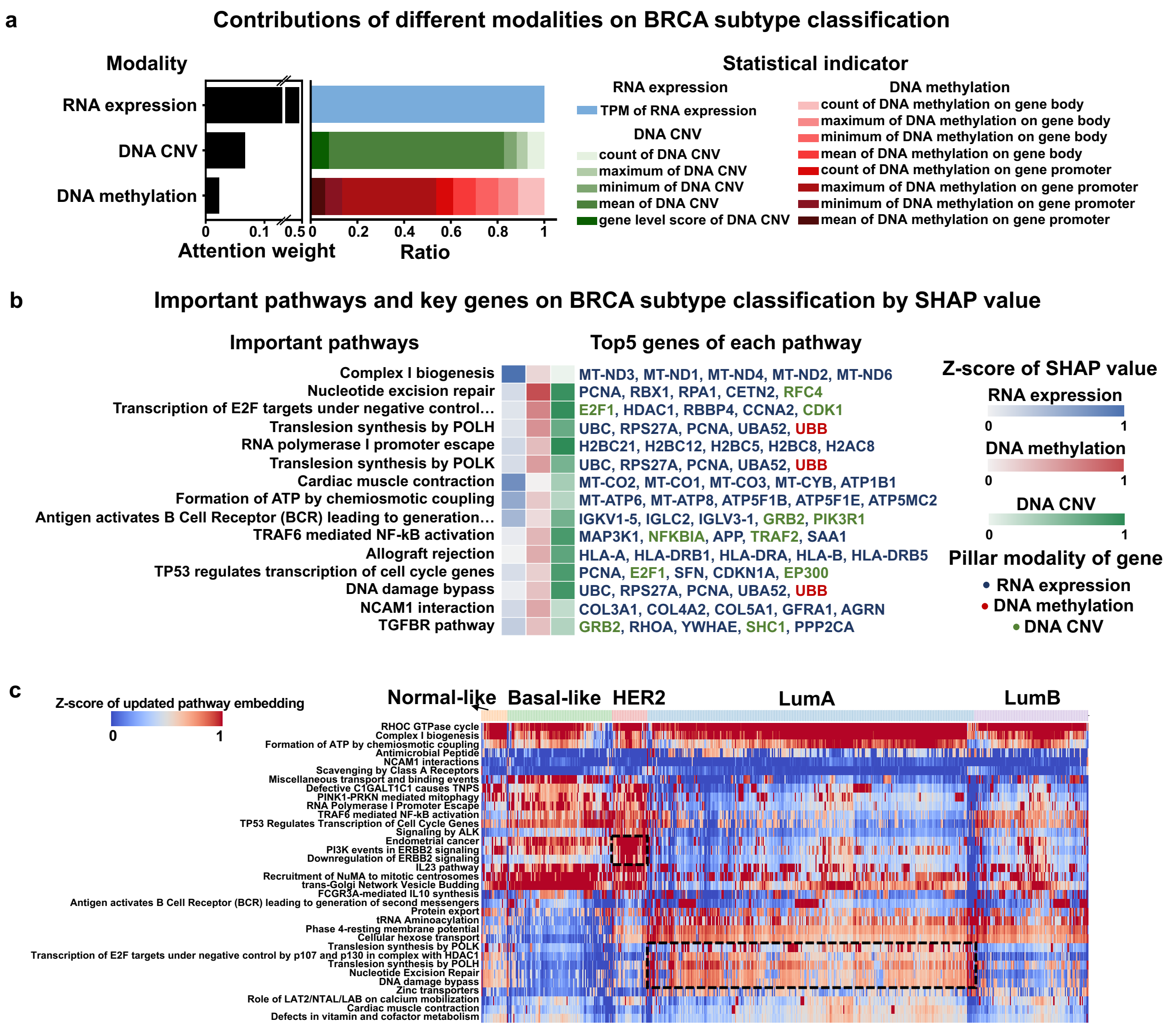
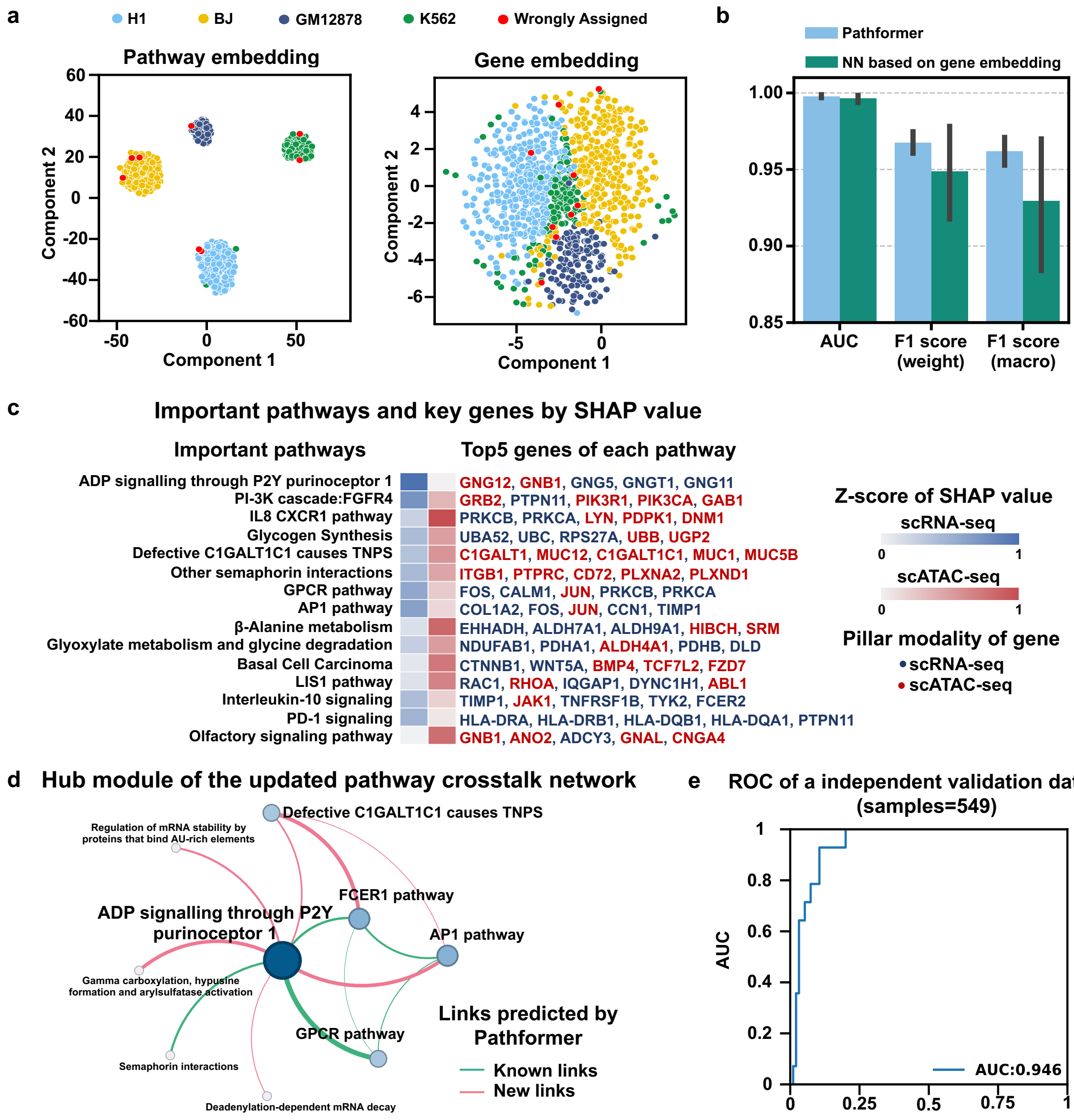


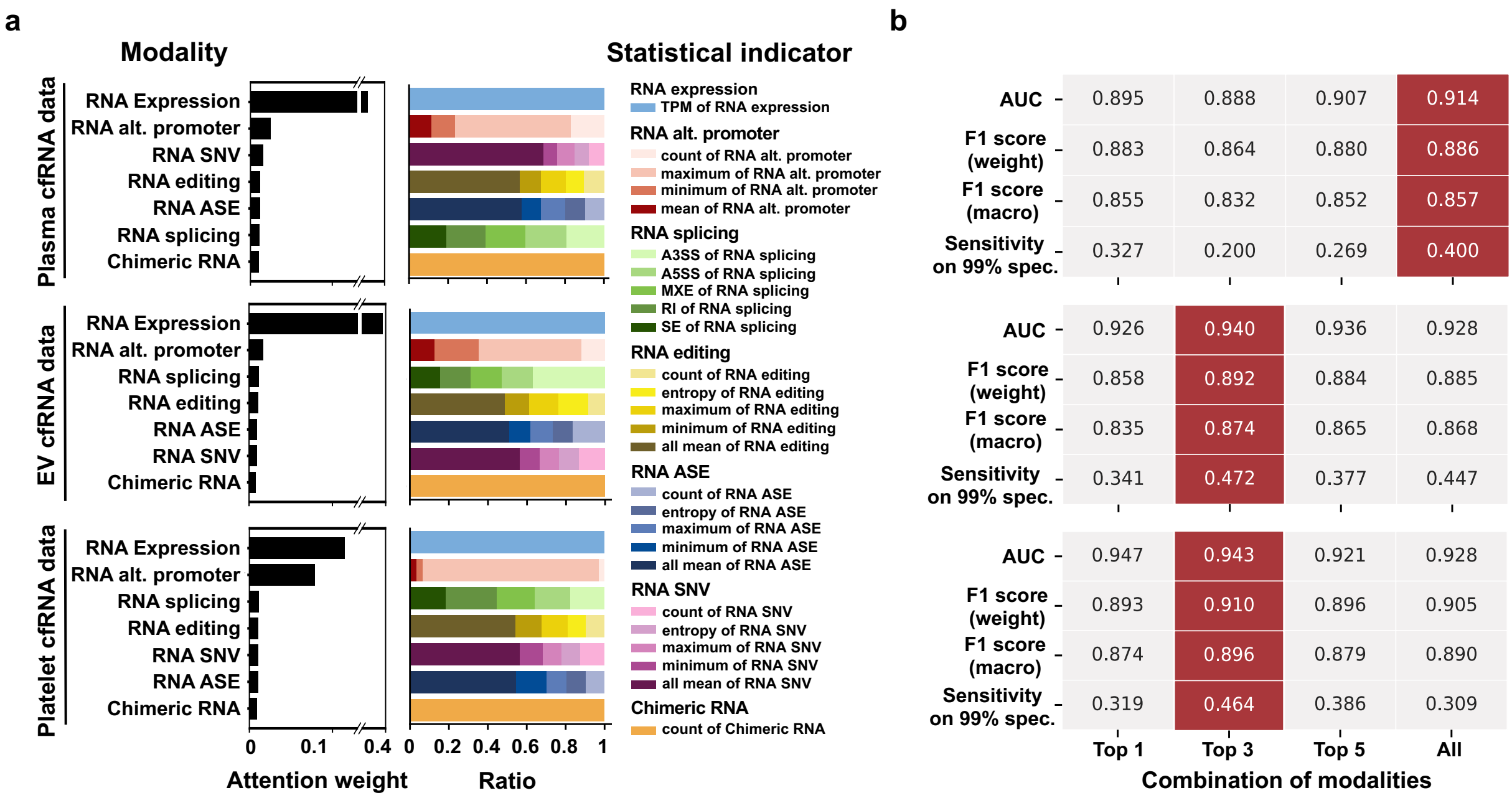
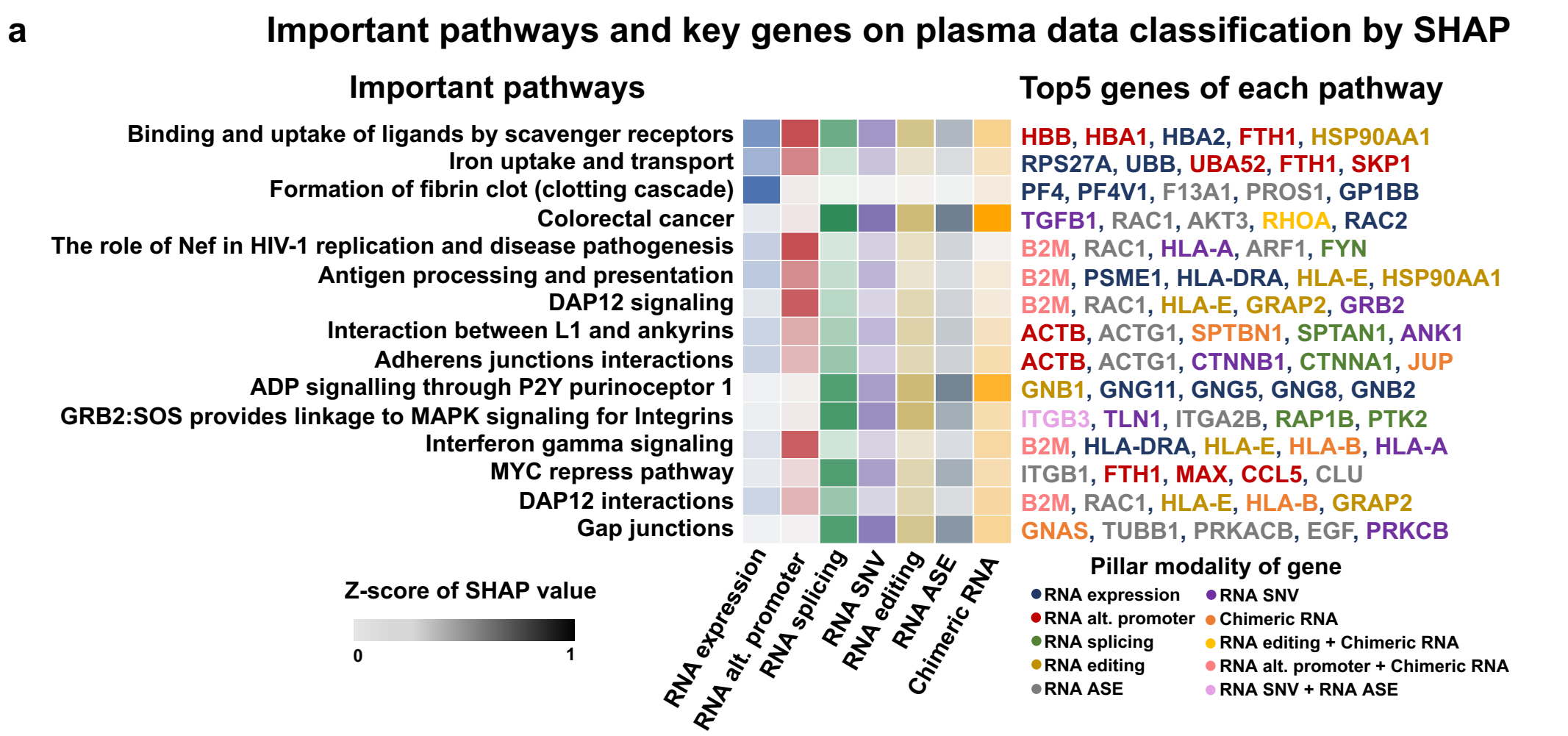
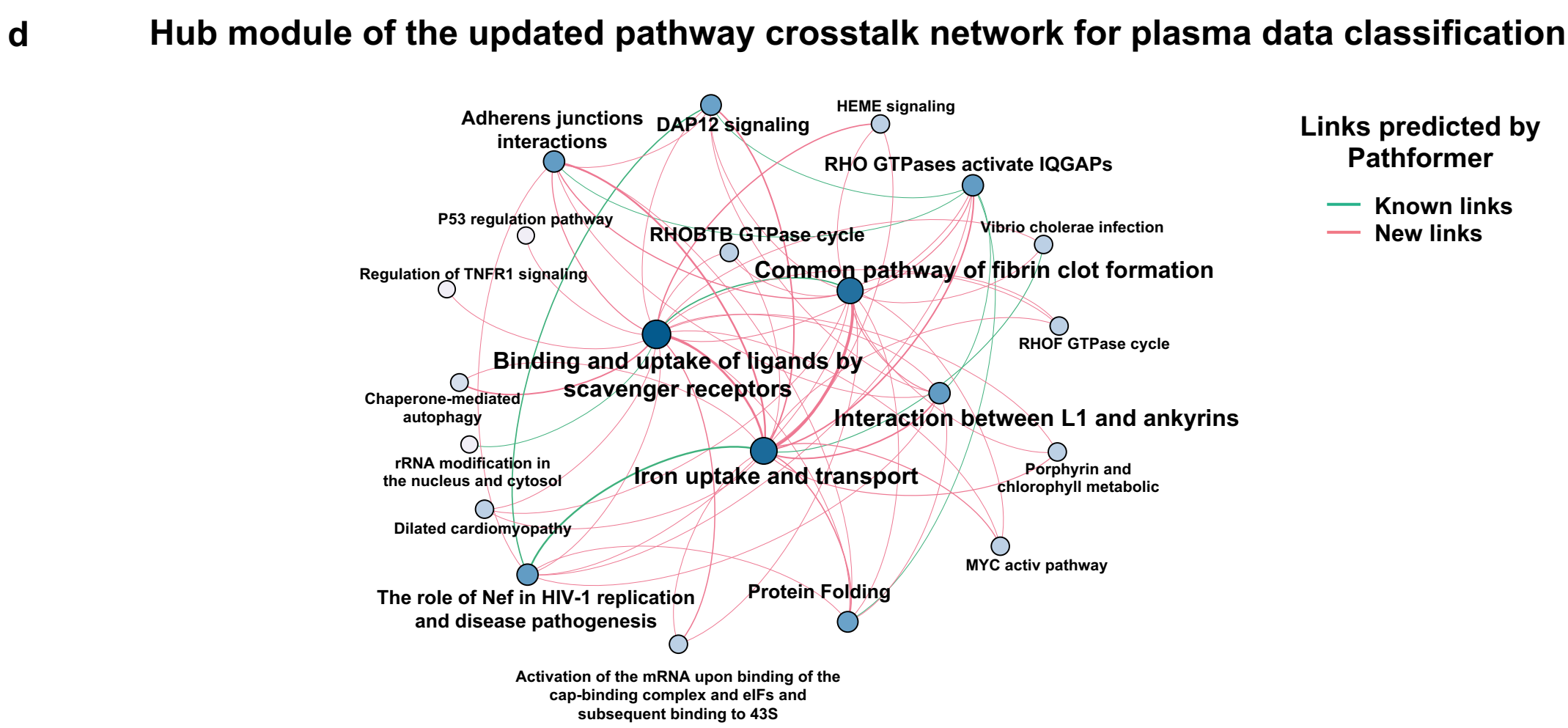
**Figure 4. Breast cancer subtype related modalities, pathways and genes revealed by Pathformer**



# Figure 5. Pathformer integrates single-cell multi-omics data for cell annotation



# Figure 6. Pathformer integrates multi-modal liquid biopsy data for non-invasive cancer diagnosis



# Figure 7. Interpretation of cancer patients' plasma data using Pathformer