RESEARCH ARTICLE

Check for updates

Multi-omics data integration for topology-based pathway activation assessment and personalized drug ranking

Nicolas Borisov,[ab] Yaroslav Ilnytsky,[cd] Boseon Byeon,[e] Olga Kovalchuk*[cd] and Igor Kovalchuk [ID] *[cd]

Although multi-omics analysis is popular for revealing diverse physiological effects and biomarkers in many branches of state-of-the-art molecular and cell biology and bioinformatics, there is still no consensus on a gold standard protocol for the integration of various multi-omics profiles into a uniformly shaped system bioinformatics platform. In the current study, we performed the integration of data on DNA methylation, and the expression of coding RNA (mRNA), micro-RNA (miRNA), and long non-coding RNA into a joint platform for calculation of signaling pathway impact analysis (SPIA) and drug efficiency index (DEI). We found that the mirrored and balanced DEI values fitted the DNA methylome data better than the original DEI. Additionally, the protein-coding mRNA-based values correlated more strongly with antisense lncRNA-based values than with miRNA-based values. The whole correlation between the mRNA-based and antisense lncRNA-based values was generally positive. This platform allowed integrative analysis of several levels of gene expression regulation of protein-coding genes and their regulators, including methylation and noncoding RNAs.

Introduction

Diversity of multi-omics data

Multi-omics data integration has been extensively used to study normal and pathological conditions by assessing molecular pathway activation. A PubMed query with the keywords "multi-omics" and "pathway" retrieves 7449 items as of end of June 2025. Each type of omics data—genomics, transcriptomics, epigenomics, proteomics, metabolomics, lipidomics, glycomics, and microbiomics—provides unique insights into different aspects of biological systems.

Consideration of multi-omics events at the integrated level is important because it provides a comprehensive understanding of biological systems by combining data from various omics layers used.[1,2] Its advantages include examining multiple molecular levels simultaneously, offering a more complete picture than single-omics approaches.[3–5] The multi-omics approach also helps cross-validate the findings from different omics layers, increases the reliability and accuracy of the results, and, second, improves the identification of robust biomarkers for disease diagnosis, prognosis, and treatment monitoring by considering multiple types of molecular data.[6,7]

Multi-omics data aggregation for pathway activation assessment

Several approaches have been developed to integrate diverse multi-omics data into molecular pathway analysis, each with its advantages and challenges. These algorithms may utilize different approaches for data processing and produce output data in different formats. Considering the mathematical approach to data processing, we propose the following approaches: statistical and enrichment approaches, machine learning approaches, and network-based approaches.

Statistical and enrichment approaches include simple enrichment analysis and quantitative statistical analysis. At the current moment, the mostly qualitative approach based on Gene Ontology classification has largely gone out of favor.[8,9] In contrast, quantitative statistical analysis using tools such as Integrated Molecular Pathway-Level Analysis (IMPaLA),[10] Pathway Multiomics,[11] MultiGSEA,[12] PaintOmics[13] and ActivePathways[14] allows for integration of multiple omics layers to compute

[a] *Armenian Bioinformatics Institute, 7 Ezras Hasratyan str., 0014, Yerevan, Armenia. E-mail: Nicolas.borissoff@abi.am*

[b] *Vivan Therapeutics, (My Personal Therapeutics Ltd.), The Westworks, White City Place 195 Wood Lane, London, W12 7FQ, England, UK. E-mail: nikolay@mypersonaltherapeutics.com*

[c] *Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada. E-mail: igor.kovalchuk@uleth.ca, olga.kovalchuk@uleth.ca, slava.ilyntskyy@uleth.ca, bbyeon@gmail.com*

[d] *Pathway Rx., 2 Fortress Rise SW, Alberta, T3H 4Z2, Canada*

[e] *Biomedical and Health Informatics, Computer Science Department, State University of New York, 2 S Clinton St, Syracuse, NY 13202, USA*

pathway enrichment scores, which provide statistical significance and visual representations of pathway activities.

Machine learning approaches involve supervised and unsupervised learning. Supervised learning techniques, such as DIABLO,[15] or OmicsAnalyst,[16] which apply the LASSO regression,[17] use annotated (phenotype groups are used as class labels) datasets to predict pathway activities based on integrated multi-omics data, enhancing predictive performance and accuracy. Unsupervised learning methods, like clustering,[16,18] principal component analysis (PCA),[18] and tensor decomposition,[18] discover latent features and patterns in multi-omics data without predefined labels.

Network-based approaches construct interaction networks from multi-omics data, identifying key regulatory nodes and pathways. A realistic picture of pathway activation can only be revealed by topological network-based methods that consider the biological reality of pathways by incorporating data on the type and direction of protein interactions.[19] Not surprisingly, topology-based methods have outperformed their counterparts in benchmarking tests.[19] Different researchers suggested a wide repertoire of algorithms and toolkits for quantitative pathway topology-based assessment of pathway activation levels (PALs), like Oncobox,[20] topology analysis of pathway phenotype association (TAPPA),[21] topology-based score (TBScore),[22] pathway-express (PE),[23] signaling pathway impact analysis (SPIA),[24] *in silico* pathway activation network decomposition analysis (iPANDA),[25] Drug Efficiency Index (DEI),[26,27] *etc.* Such pathway activation level calculations utilize high-throughput gene expression or mutation profiles. Diverse methods, algorithms, and software for automated curation of pathway topology databases and uniformly shaped annotations of their content have also been developed.[28,29]

It may seem like it is sufficient to obtain the data from whole transcriptome sequencing (WTS, RNA-seq), as it allows evaluating the level of activation/inactivation of various pathways. However, it is known that ncRNAs, especially miRNAs, are able to regulate mRNA expression negatively through translational inhibition, and thus mRNA sequencing does not fully represent changes in the pathways. Different ncRNAs interfere with the gene expression process at different stages and with different affinities for distinct mRNAs. For example, small interfering RNAs (siRNAs) are RNA duplexes with typically 21–23 nucleotides that bind to a strictly specific mRNA molecule and prevent their movement from the nucleus to the cytoplasm; thus, mRNA is quickly cleaved in the nucleus as well as the cytoplasm.[30,31] Although micro-RNAs (miRNAs) have almost the same length (19–25 nucleotides), they are not so gene-specific[32] and bind to the target mRNA molecules in the cytoplasm, preventing translation and accelerating mRNA degradation by RNAases.[30]

In contrast to miRNAs, most antisense RNAs (asRNAs) are longer than 200 nucleotides, although shorter asRNAs also exist.[33] Like siRNAs, asRNAs are gene-specific; like miRNAs, they bind to mRNA molecules in the cytoplasm and prevent translation. The influence of asRNAs on the abundance of mRNAs is controversial: although asRNA may stimulate mRNA cleavage, the complexing of asRNA with mRNA can protect mRNA from RNAase and inhibit its degradation.[34–36]

Multi-omics integration study design

- Methylation profile integration
  Diseases: renal cell carcinoma, chronic myelomonicytic leukemia, myelodysplastic syndrome, type 2 diabetes, multiple sclerosis
  Data aggregation level: distinct genes, pathway activation levels, drug efficiency scores

- Non-coding RNA (micro-RNA and antisense lncRNA) profile integration
  Diseases: colon cancer, lung cancer, B cells, polycystic ovary syndrome, heart failure, temporal lobe epilepsy
  Data aggregation level: distinct genes, pathway activation levels

Fig. 1 Research pipeline of the current study.

Also, asRNAs can bind to the DNA template strand, preventing the transcription machinery from producing mRNA. Therefore, asRNAs can affect the splicing of pre-mRNA, leading to different mRNA isoforms.

There are no examples for the incorporation of the results of DNA methylome, siRNAs, dsRNAs, asRNAs or miRNAs profiling into the analysis of dysregulated pathways. Integration of mRNA-seq data with siRNA-seq data may help better understand the transcription and translation events. In the current study (see Fig. 1), we report on the systemic multi-omics integration of protein-coding mRNA expression profiles, and non-coding RNA expression profiles, including micro-RNA and long non-coding RNA/anti-sense RNA (antisense lncRNA/asRNA) profiles, into the SPIA/DEI-based computational platform[26,27] for pathway activation assessment and drug efficiency scoring.

Materials and methods

Topological pathway activation assessment according to SPIA (signal pathway impact analysis)

The pathway-Express (PE)-score for a pathway K can be calculated as follows:[23]

$$PE_K = \log(1/p) + \frac{\sum_{g \in K} |PF(g)|}{|\Delta E| N_d(P)}$$

The first term here is the p-value for the probability to obtain the observed or a greater number N_d of differentially expressed genes (between the pools of case and normal samples) randomly, assuming a hypergeometrical distribution of N_d. The second term is a summation over the perturbation factors (PF) for all genes g of the pathway K,

$$PF(g) = \Delta E(g) + \sum_{\gamma \in U_g} \beta_{\gamma g} \frac{PF(\gamma)}{n_{down}(\gamma)}$$

Here $\Delta E(g)$ is the signed log-fold-change (LFC) of gene g expression in a given sample compared with the expected value for the pool of control samples. The latter term expresses the summation over all the genes γ that belong to the set U_g of the upstream genes for the gene g. The value of $n_{down}(\gamma)$ denotes the number of downstream genes for gene γ. The weight factor $\beta_{\gamma g}$ indicates the interaction type between γ and g: $\beta_{\gamma g} = 1$ if γ

activates g, and $\beta_{\gamma g} = -1$ when γ inhibits g. Although the value of PF may be positive or negative, the overall score of PE is obligatory positive. The search for upstream/downstream genes is performed according to the depth-first search method.

To obtain an estimator for pathway perturbation that is positive for an up-regulated pathway and negative for a down-regulated pathway, use the second term in the formula for the perturbation factor (PF) from the precious paragraph, resulting in the accuracy value,

$$\mathrm{Acc}(g) = \mathrm{PF}(g) - \Delta E(g).$$

It can be shown that this accuracy vector may be expressed as follows:[24]

$$\mathrm{Acc} = \mathbf{B} \cdot (\mathbf{I} - \mathbf{B})^{-1} \cdot \Delta \mathbf{E},$$

where

$$\mathbf{B} = \begin{pmatrix} \dfrac{\beta_{11}}{n_{\mathrm{down}}(g_1)} & \dfrac{\beta_{12}}{n_{\mathrm{down}}(g_2)} & \cdots & \dfrac{\beta_{1n}}{n_{\mathrm{down}}(g_n)} \\ \dfrac{\beta_{21}}{n_{\mathrm{down}}(g_1)} & \dfrac{\beta_{22}}{n_{\mathrm{down}}(g_2)} & \cdots & \dfrac{\beta_{2n}}{n_{\mathrm{down}}(g_n)} \\ \dfrac{\beta_{n1}}{n_{\mathrm{down}}(g_1)} & \dfrac{\beta_{n2}}{n_{\mathrm{down}}(g_2)} & \cdots & \dfrac{\beta_{nm}}{n_{\mathrm{down}}(g_n)} \end{pmatrix},$$

$\mathbf{I}$ is the identity matrix, and

$$\Delta \mathbf{E} = \begin{pmatrix} \Delta E(g_1) \\ \Delta E(g_2) \\ \cdots \\ \Delta E(g_n) \end{pmatrix}.$$

The resulting score for pathway perturbation is calculated as follows: $\mathrm{SPIA} = \sum\limits_{g} \mathrm{Acc}(g)$.

Curation of pathway databases

We used the Oncobox pathway databank, OncoboxPD[29] that accumulates 51 672 uniformly processed human molecular pathways extracted from different source databases. It is the largest knowledge base of human pathways with annotated gene functions, i.e. ready for the pathway activation calculations. Superposition of the enclosed pathways formed an interactome graph of protein–protein interactions and metabolic reactions totaling 361 654 interactions and 64 095 molecular participants. All pathways were functionally classified according to their main underlying biological processes using the Gene Ontology (GO) tree. Each pathway node was algorithmically functionally annotated by a specific activation/repressor role index. This enables direct calculation of pathway activation levels (PALs, i.e. using the SPIA method) using human RNA/protein expression profiles.

Using the Drug Efficiency Index, DEI, software,[26,27] the user can analyze custom expression data to evaluate SPIA scores in samples of interest against a built-in or custom set of controls and statistically evaluate differentially regulated pathways.

Integration of non-coding RNA profiles into SPIA calculations

For calculations of pathway-based values, such as signaling pathway impact analysis, SPIA[24,26,27] using the long noncoding/antisense RNA (lncRNA/asRNA) expression profiles, we considered the influence of long noncoding/antisense RNA in a manner similar to what has been done for microRNA.[37] Fig. 2 shows the effect of various pre-translation events that regulate gene expression and, subsequently, the pathway activation process.

Considering the fact that small RNAs typically direct the methylation of specific loci, and that both non-coding RNA (ncRNA) and DNA methylation downregulate gene expression (Fig. 2), we suggested calculating the methylation-based and ncRNA-based SPIA values with the negative sign compared to standard, transcriptome/mRNA-based values, using the same pathway topology graphs: $\mathrm{SPIA}_{\mathrm{methyl,ncRNA}} = -\mathrm{SPIA}_{\mathrm{mRNA}}$.

Drug efficiency index (DEI)

The method for assessment of personalized drug efficiency index (DEI)[26,27] consists of the following steps:

1. Calculate the pathway activation level (PAL) values for all molecular pathways (e.g. SPIA[24]).

2. Calculate the values of the pathway weight (w_{p}) factor as follows. For pathways with a positive mean PAL score of the case samples, $w_{\mathrm{p}} = $ ((number of case samples with a positive PAL score)/(total number of case samples)). For pathways with a negative mean PAL score of the case samples, $w_{\mathrm{p}} = $ ((number of case samples with a negative PAL score)/(total number of case samples)).

3. Adjust the mean PAL score of each pathway by the weight factor,

$$\mathrm{PAL}_{\mu} = \mathrm{mean}(\mathrm{PAL}) \cdot w_{\mathrm{p}}.$$

4. Perform the Student's t-test if the values of PAL_{μ} for the pool of case samples are different from 0 (for the pool of control samples, the values of PAL_{μ} are clearly equal to 0). During the Student's t-test, the following case classes are considered: (a) untreated case (U), e.g. the pathological state before drug application, should be far from the control (C); (b) treated case (T), e.g. the pathological state after drug application, should be close to the control.
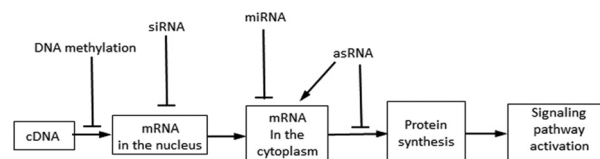
The following output values result from such calculations:

Fig. 2 Multi-omics chain of events that interfere with the gene expression process.

(a) $|t_U|$ = absolute t-value for the Student's t-test for U-$vs.$-C profiles; (b) $|t_T|$ = absolute t-value for the Student's test for T-$vs.$-C profiles.

5. In addition to the first-generation DEI metric[26] for individual drug activities, $\text{DEI} = 2 \cdot \left(\frac{|t_U|}{|t_U| + |t_T|} - \frac{1}{2} \right) = \frac{|t_U| - |t_T|}{|t_U| + |t_T|}$, which is equal to 1 when $t_T = 0$.

An alternative metric is called the mirrored DEI:[27]

$$\text{DEI}_M = 2 \cdot \left(\frac{2 \cdot |t_U|}{2 \cdot |t_U| + |t_T| + t_U} - \frac{1}{2} \right) = \frac{2 \cdot |t_U| - |t_T| + t_U}{2 \cdot |t_U| + |t_T| + t_U}.$$

The DEI_M metric is equal to 1 when $t_T = -t_U$; this is the maximum possible value of this metric.

Similar to the previous DEI metric, $\text{DEI}_M = 0$ when $t_T = t_U$, and $\text{DEI}_M = -1$ when $|t_T| \gg |t_U|$.

The third metric,[27] balanced DEI, $\text{DEI}_B = \frac{\text{DEI} + \text{DEI}_M}{2}$ is the mean value of the DEI and mirrored DEI. In our previous work,[27] we validated the DEI, DEI_M, and DEB_B, methods, including their ability to distinguish clinically effective and ineffective treatments, the pathological and healthy samples, and treated and untreated patients.

Datasets for DNA methylome integration into SPIA/DEI analysis

As the database of miRNA targets, we used the current version of MiRTarBase.[39] MiRTarBase provides information on 15 641 distinct genes affected by the miRNAs.

The use of external normal references, or even synthetic controls, may require special cross-platform normalization methods, like those in our work,[38] for a direct comparison. Although we made a lot of efforts in developing these normalization methods, they may, however, significantly impact the case-to normal log-fold-changes (LFC).[38] This may introduce some unforeseen artifacts; therefore, we decided to utilize samples from the same cohort as a reference in the current study.

To make the first test for the methylation module, we used the data on the antiproliferative activity of the DNA hypomethylating agent 5-aza-2′-deoxycytidine (DAC)[40] (see GEO dataset GSE198673). Li $et\ al.$ (2023) tested whether DAC can inhibit the growth of clear cell renal cell carcinoma (ccRCC), both for the wild-type (WT) and knock-out (KO, SETD2−/−) variants, since the SETD2 (Su(var)3-9, Enhancer of Zeste, and Trithorax Domain Containing 2) gene is one of the major histone methyltransferases.[40] Both WT and KO ccRCC cells were treated with 300 nM of DAC. Then the cell growth rate and DNA methylation profile were monitored for 40 days; DNA methylation and the expression of protein-coding mRNA were profiled on days 0, 5, 15, and 40 after the DAC treatment.

Additionally, we curated four other recently published cohorts of DNA methylation profiles, collected for myelodysplastic syndrome (MDS) (GSE119617), type 2 diabetes (GSE145746),[41] multiple sclerosis (MS) (GSE151017),[42] and chronic myelomonocytic leukemia (CMML) (GSE221269) – see Table 1.

Table 1 Overview of curated cohorts with DNA methylation profiles

Paper reference	—	(Bansal et al., 2020)[41]	(Bansal et al., 2020)[42]	(Bansal et al., 2020)[42]	(Bansal et al., 2020)[42]	(Ringh et al., 2021)[42]	(Ringh et al., 2021)[42]	(Ringh et al., 2021)[42]	(Ringh et al., 2021)[42]	(Ringh et al., 2021)[42]
GSE ID	GSE119617	GSE145745	GSE145745	GSE145745	GSE145745	GSE151017	GSE151017	GSE151017	GSE151017	GSE221269
Disease	Myelo-dysplastic syndrome	Type 2 diabetes	Type 2 diabetes	Type 2 diabetes	Type 2 diabetes	Multiple sclerosis	Multiple sclerosis	Multiple sclerosis	Multiple sclerosis	Chronic myelo-monocytic leukemia
Drug	5-Azacitidine	TGFB1 24 h	TGFB1 72 h	TGFB1 96 h	IFNb; relapse	IFNb; remission	Tysabr; remission	Other drugs; remission	Azathio-prine (AZA)
Untreated cases	8	4	4	4	2	4	6	6	10
Untreated controls	5	4	4	4	44	44	44	44	5
Treated cases	8	4	2	2	4	10	6	6	10
Treated controls	5	4	2	2	44	44	44	44	5
Methylation sites	34 669	825 425	825 425	825 425	734 078	734 078	734 078	734 078	719 859
Affected genes	1488	23 039	23 039	23 039	23 314	23 314	23 314	23 314	22 390

Table 2 Overview of curated cohorts for antisense lncRNA/miR vs. protein-coding mRNA profiling

Paper reference	(Ma et al., 2015; Ma and Hu, 2023)[43,44]	(Yu et al., 2021)[45]	(Zhao et al., 2021)[46]	(He et al., 2024)[47]	(Liao et al., 2023)[48]	(Wang et al., 2022)[49]
GSE ID	GSE127905	GSE164595	GSE168404	GSE194299	GSE197671	GSE205661
Profiling platform	Illumina HiSeq X Ten	Illumina HiSeq 4000	Illumina HiSeq 2500	Illumina HiSeq 2000	Illumina NovaSeq 6000	Agilent-046064 miRNA microarray Agilent-052909 antisense lncRNA/mRNA mircoarray
Disease/sample type	Colon cancer/HCT116 cells with P14AS over-expression or/and AUF1 knockdown vs. controls	B cells treated and untreated with methylation inhibitors	Polycystic ovary syndrome/granulosa cells	Lung cancer	Heart failure/cardiomyocytes	Temporal lobe epilepsy with hippocampal sclerosis/brain tissue
mRNA profiles	Protein-coding mRNA, antisense lncRNA	Protein-coding mRNA, antisense lncRNA	Protein-coding mRNA, antisense lncRNA, miRNA	Protein-coding mRNA, antisense lncRNA	Protein-coding mRNA, antisense lncRNA	Protein-coding mRNA, miRNA
Disease cases	4	8	5	3	8	6
Controls	4	12	5	3	8	9
Number of gene targets of ncRNA	1141	1079	5775	1381	1381	10 227

Datasets for non-coding RNA integration into SPIA analysis

We compared the protein-coding-based SPIAs vs. non-coding-based ones for the following six multi-omics datasets obtained from the Gene Expression Omnibus (GEO) portal (Table 2). We included in our analysis only those multi-omics profiles that contain at least 1000 distinct genes and their targets.

We then performed Gene Ontology (GO) enrichment analysis of antisense lncRNA molecular targets from these cohorts using the enrichGO software tool[50] and Metascape online service.[51]

Results

Integration of methylome profiles into SPIA/DEI pathway activation and drug efficiency assessment

We confirmed that DAC decreased the overall methylation rate compared with DAC-untreated (T = 0 days) samples for clear human renal cell carcinoma (ccRCC)[40] (GEO reference GSE198673). DNA methylation was inhibited in both WT and SETD2−/− KO samples (Fig. 3). To calculate the case-to-control LFC, we used methylated cite/gene reads as cases and corresponding unmethylated sites/gene reads as controls (Fig. 3).

The DEI calculations require three types of profiles:

(a) Control samples (C), used as a reference for LFC computations for every gene expression.

(b) Untreated case samples (U) for the U-vs.-C comparison.

(c) Treated case samples (T) for the T-vs.-C comparison.

The study by Li et al. (2023) did not include any normal or healthy samples as a control.[40] That is why we used the sample, which showed the slowest proliferation rate (SETD2−/−, exposed with 300 nM of DAC, five days after treatment), as a quasi-normal control reference (Table 3).

We calculated the DEI values for the following combinations of untreated (U) and treated (T) case samples (Table 3). For the (A) experiment, the treatment procedure was DAC addition, and the T-vs.-U comparison implied juxtaposition of samples which had received DAC, and those which had not. For the (B) experiment, the treatment procedure was the knockout (KO)
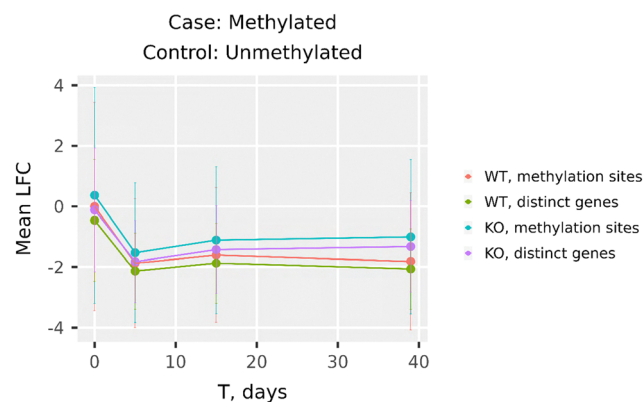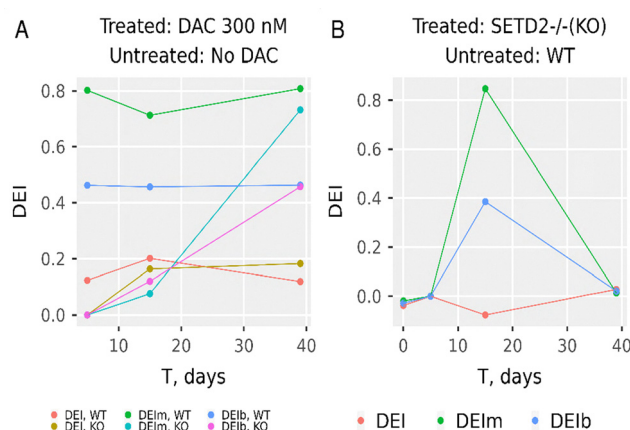

Fig. 3 Methylated-vs.-unmethylated log 2-fold change (LFC) in DNA reads for WT and SETD2−/− KO ccRCC samples[40] (GEO reference GSE198673).

Table 3 Control (C), untreated (U), and treated (T) samples for DEI calculations for the GSE198673 ccRCC dataset

Panel of Fig. 3	Control samples (C)	Untreated case samples (U)	Treated case samples (T)
(A)	KO (SETD2−/−), treated with 300 nM of DAC, five days after treatment (the sample, which showed the slowest proliferation rate)	WT and KO, with no DAC 5, 15, and 40 days after DAC treatment of DAC-treated samples	WT and KO, with 300 nM of DAC 5, 15, and 40 days after DAC treatment
(B)		WT, with 300 nM of DAC 0, 5, 15, and 40 days after DAC addition	KO, with 300 nM of DAC 0, 5, 15, and 40 days after DAC addition

Fig. 4 Drug efficiency index (DEI), mirrored DEI (DEIm), and balanced DEI (DEIb) for two comparisons of ccRCC methylome profiles (Table 3, GSE198673). Panel (A): DAC-treated *vs.* untreated samples. Panel (B): SETD2−/− KO *vs.* WT samples.

of SETD2, and the T-*vs.*-U comparison used juxtaposition of SETD2−/− (KO) and WT samples.

Based on our group comparisons, we demonstrated the beneficial role of both DAC (Fig. 4) and SETD2 knockout (Fig. 4(B)) for inhibition of cell proliferation. We observed this effect in terms of the drug efficiency index (DEI), as well as of the mirrored (DEIm) and balanced (DEIb) modifications of DEI.[26]

Other four DNA methylation case-*vs.*-control cohorts (Tables 4 and 5, totaling 82 case samples and 76 control samples) confirm the more adequate role of DEIm and DEIb values (compared to the old DEI metric) for the assessment of drug activity in such different diseases as MDS, type 2 diabetes, MS, and CMML. In particular, the DEIm and DEIb metrics were always positive for all these cohorts except in two cases: (1)

Table 4 Disease *vs.* control cohorts curated in the current work

Paper reference	—	(Bansal *et al.*, 2020)[41]	(Ringh *et al.*, 2021)[42]	—
GSE ID	GSE119617	GSE145745	GSE151017	GSE221269
Disease	MDS	Type 2 diabetes	MS	CMML
Cases	16	12	34	20
Controls	10	12	44	10
Methylation sites	34 669	825 425	734 078	719 859
Affected genes	1488	23 039	23 314	22 390

relapsed MS and (2) type 2 diabetes at the longest time after drug administration (Table 5).

Comparison of miRNA and antisense lncRNA as regulatory molecules for protein-coding mRNA pathway activation

Although the overall role of miRNA in gene expression is inhibitory, it is not easy to obtain a stable and robust negative correlation between the mRNA and miRNA values, both at the level of distinct genes and at the pathway activation levels.[52]

We analyzed the correlation between mRNA-based *vs.* miRNA-based, as well as between mRNA-based and antisense lncRNA-based values, at different levels of data aggregation (case-to-control LFC for each gene, and SPIA for pathways) in the six multi-omics cohorts listed in Table 2. For these cohorts, the correlation between the antisense lncRNA s-based and protein-coding mRNA-based values was generally higher than between the miRNA-based and protein-coding mRNA-based values (Fig. 4). Note that no false discovery rate (FDR) correction is required for *p*-values shown in Fig. 5. These correction methods, like the Benjamini–Hochberg one, provide more reliable marker sets for high-throughput profiles when multiple features, like distinct genes, are tested. However, in Fig. 5, we compare correlation coefficients between gene expression/pathway activation profiles, rather than the expression/pathway activation levels distinctly. Consequently, for single-value statistical tests, the BH correction is trivial: p_adj = p_raw, and no adjustment is needed. We added the corresponding explanation to the paper text, preventing the question about the BH adjustment from the readers.

Although it may seem counter-intuitive, the overall correlation between antisense lncRNA- and mRNA-based case-to-control LFCs was positive (Fig. 5(A)). Indeed, many authors found that antisense lncRNAs may increase the abundance of sequestered (and, therefore, inactivated) mRNA in the cytoplasm, not only in bacteria but also in mammals.[34–36]

Hence, we found that antisense lncRNA values correlated better with mRNA values than miRNA values correlated with mRNA values. Therefore, antisense lncRNAs may be more informative than miRNAs in the analysis of interference in signaling pathway activation caused by the non-coding transcriptome.

To reveal the gene expression modulating effect of antisense lncRNAs and miRNAs involved in the current study, we applied gene ontology enrichment analysis according to the enrichGO method to the targets of antisense lncRNAs and miRNAs in six

Table 5 DEI values for case-*vs.*-control methylome profiling cohorts

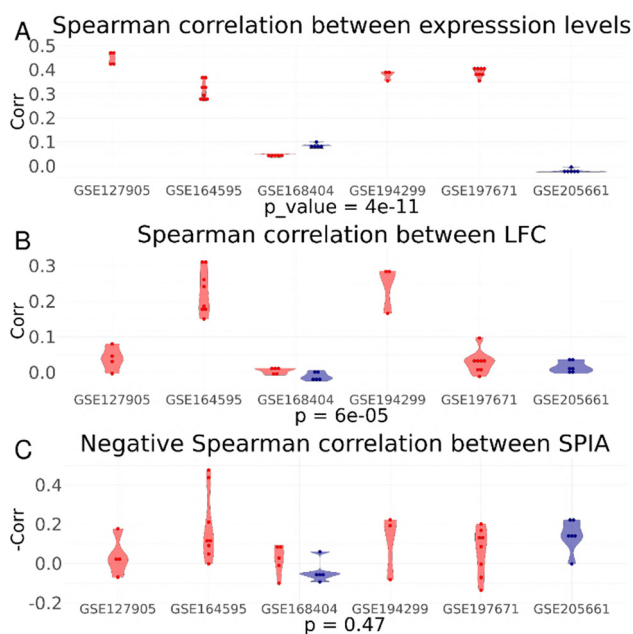| Reference	—	(Bansal *et al.*, 2020)[41]	(Bansal *et al.*, 2020)[41]	(Bansal *et al.*, 2020)[41]	(Ringh *et al.*, 2021)[42]	(Ringh *et al.*, 2021)[42]	(Ringh *et al.*, 2021)[42]	(Ringh *et al.*, 2021)[42]	—
GSE ID	GSE119617	GSE145745	GSE145745	GSE145745	GSE151017	GSE151017	GSE151017	GSE151017	GSE221269
Disease	Myelo-dysplastic syndrome	Type 2 diabetes	Type 2 diabetes	Type 2 diabetes	Multiple sclerosis	Multiple sclerosis	Multiple sclerosis	Multiple sclerosis	Chronic myelo-monocytic leukemia
Drug	5-Azaci-tidine	TGFB1 24 h	TGFB1 72 h	TGFB1 96 h	IFNb; relapse	IFNb; remission	Tysabr; remission	Other drugs; remission	Azathio-prine (AZA)
tU	2.35	3.05	3.05	3.05	−0.57	−2.19	−2.85	−2.85	2.91
tT	−3.80	2.74	−5.22	5.59	−1.63	2.93	−0.62	3.66	−5.52
DEI	−0.24	0.05	−0.26	−0.29	−0.48	−0.15	0.64	−0.13	−0.31
DEIm	0.53	0.03	0.48	−0.17	−0.32	0.71	0.24	0.75	0.38
DEIb	0.14	0.04	0.11	−0.23	−0.40	0.28	0.44	0.31	0.04

Fig. 5 Spearman correlation between protein-coding mRNA values and ncRNA values (red – antisense lncRNA, blue – miRNA) for six GEO cohorts (see Table 2). (A) Gene expression levels; (B) case-to-control log-FC (LFC); (C) SPIA. The *p*-value is shown for two-sided Student's test between red and blue groups of correlation coefficients.

multi-omics cohorts listed in Table 2. For the GO analysis, we embraced different sets of genes: (A) all antisense lncRNA; (B) those genes, which have high correlation (the top 25% quantile) in the expression level between the corresponding protein-coding mRNA and miRNA values; and (C) those genes, which have high correlation (the top 25% quantile) in the expression level between the corresponding protein-coding mRNA and antisense lncRNA values (Fig. 6). We showed that the overall GO terms for the options A, B, and C overlap significantly (Fig. 6(D)). Note also that the most explicitly manifested GO terms for all these options are related to the developmental processes.

Table S1 contains results for GO annotation of target genes for antisense lncRNA from these six multi-omics cohorts

(Table 2), which we obtained using the Metascape online service for GO analysis.[51] We provided (see Table S2, Fig. 7) the target gene statistics for antisense lncRNA and KEGG pathways that we curated in our pathway database. This analysis reveals the high enrichment levels for cancer-related pathways, which comprise 11 out of 20 top enriched signalling cascades (Fig. 7).

Discussion

Multi-omics data integration is a rapidly evolving field that seeks to combine different types of omics data to provide a comprehensive view of biological systems. The sequence of events that foreruns translational processes and governs gene expression has been extensively studied,[3] and attention has been sometimes shifted from miRNA to long non-coding RNAs, which some researchers consider more relevant for controlling the abundance of protein-coding mRNA.[53,54]

The integration of these diverse data types is crucial for understanding complex biological processes and identifying the molecular pathways involved in various diseases. By combining omics profiles, researchers can gain a comprehensive understanding of pathway activations and the complex molecular mechanisms underlying various diseases.[7] This holistic approach enhances the accuracy and robustness of pathway activation assessments, providing critical insights for personalized medicine and therapeutic development.[55,56]

In the current work, we investigated the integration of multi-omics profiles into the calculation of pathway activation levels according to the SPIA method.[24] Specifically, we used our multi-omics SPIA platform[26,27] to integrate mRNA, ncRNA and antisense lncRNA data. All these additional (beyond the standard protein-coding mRNA) regulatory processes *i.e.*, DNA methylome, miRNAs, and antisense lncRNAs, theoretically inhibit gene expression, that is why we calculated the SPIA values for these additional profiles with the sign opposite to SPIA values based on protein-coding mRNA data.

For the multi-omics-based SPIA assessments, we obtained the following results. First, similar to the post-traumatic stress disorder (PTSD) protein-coding mRNA-based data,[26] the mirrored and balanced DEI values proved more adequate than the
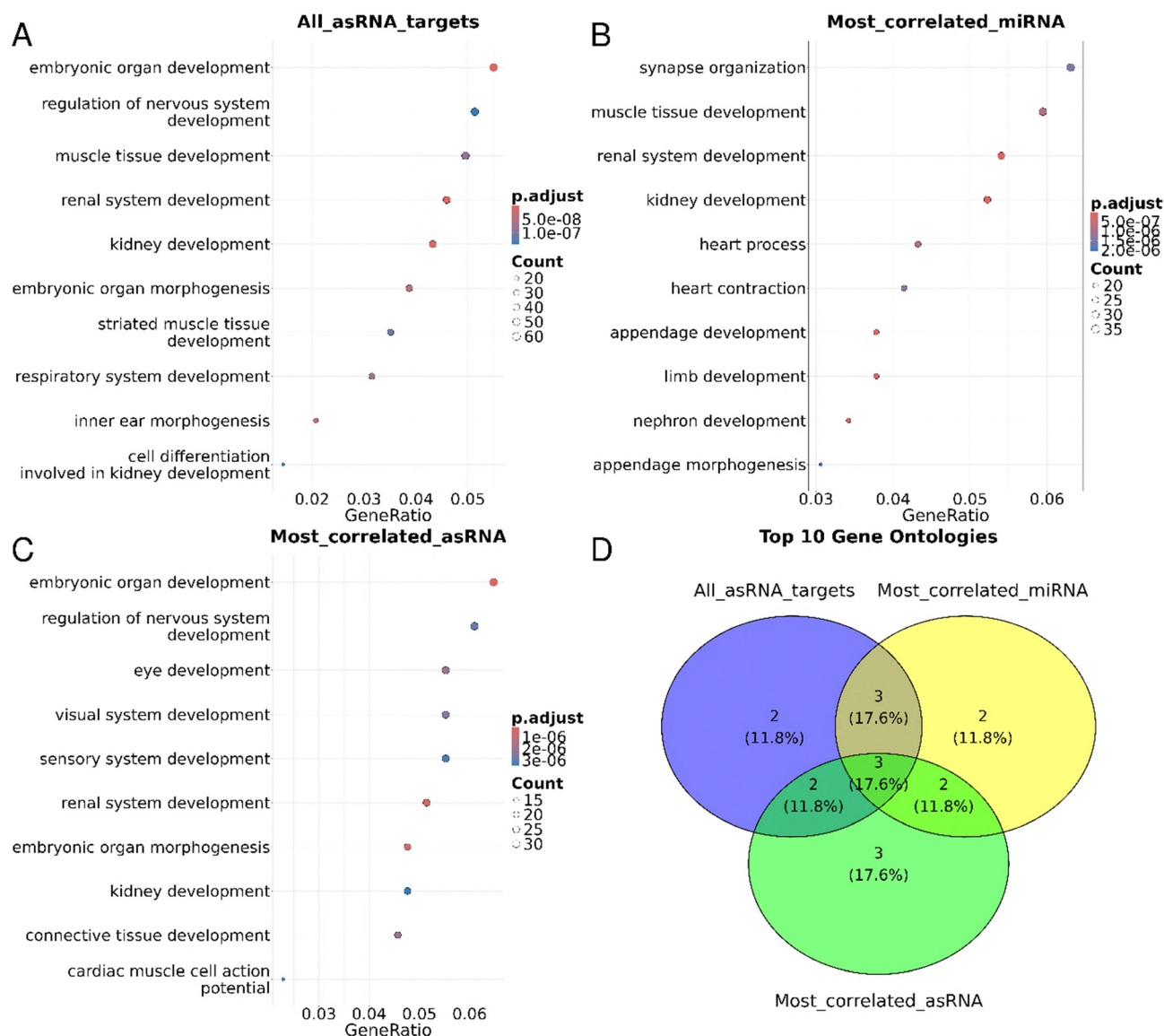
Fig. 6 GO enrichment analysis of antisense lncRNA molecular targets combined from six multi-omics cohorts (Table 2). (A) All antisense lncRNA targets; (B) top quartile of positively correlated genes between protein-coding mRNA and miRNA; (C) top quartile of positively correlated genes between protein-coding mRNA and antisense lncRNA; (D) intersection of GO terms shown in panels A–C.

original DEI values.[27] Second, the protein-coding mRNA-based values had better correlation with antisense lncRNA -based values than with miRNA-based values. Also, the correlation between the mRNA-based and antisense lncRNA -based values was mostly positive. This surprising effect may be caused by the antisense lncRNA-dependent sequestration of inactivated mRNA in the cytoplasm, which may artifactually inflate mRNA abundance estimates.[34–36]

Integration of DNA methylation and mRNA expression levels

Data on direct correlation between methylation and gene expression in the set of biological samples in mammals are abundant. In most cases, methylation at the promoter negatively correlates with gene expression. However, this is typically demonstrated for a subset of the genes, while

correlation at the whole genome level, that is, correlation between all DNA methylation changes and the expression of all genes is not that commonly established. For example, negative correlation between DNA methylation at the promoter region and expression of the subset of genes in the samples with acute myocardial infarction in mice heart was only established for 4183 genes (depending on the time point).[57] Also, in the ovarian cancer samples, negative correlation between DNA methylation and gene expression was established for a subset of 1118 genes.[58] Analysis of blood samples in people with coronary artery disease revealed correlation in 669 genes.[59] Analysis of DNA methylation and gene expression in the developing porcine placenta showed a total of 4774 genes whose DNA methylation levels on the promoter were negatively correlated with their expression levels ($R < -0.475$).[60]
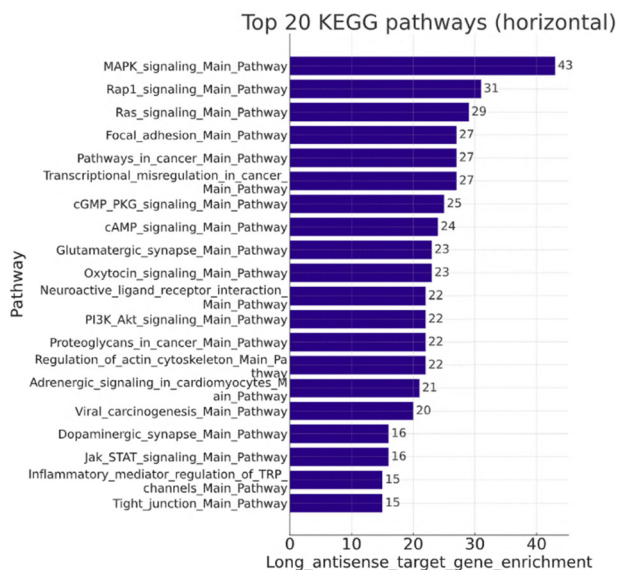
Top 20 KEGG pathways (horizontal)

Fig. 7 Top 20 KEGG pathways enriched with the antisense lncRNA targets.

In contrast, when correlation is attempted at the level of the whole genome, it is more difficult to demonstrate. Analysis of gene expressions and methylation pattern in horse sarcoids showed significant negative correlations between DNA methylation at the promoter regions and mRNA levels, with the R of ~ -0.23.[61] Similar correlation in DNA methylation in the introns showed a much weaker negative correlation with gene expression (~ -0.1), while no significant correlation was found between DNA methylation in exons and gene expression. The authors used MethGET (Methylation and Gene Expression Teller) software.[62] This software appeared to be superior to several other previously published tools for DNA methylation analysis such as COHCAP,[63] PiiL,[64] and ViewBS.[65]

Integration of DNA methylation and gene expression data is not a simple task and to date has been attempted with various degrees of success. Sajedi et al. (2023) developed the iNETgrate package that allows to integrate data from all genes, simultaneously building a comprehensive gene-level network. However, they do not rely on complex pathway topology graphs, preferring mostly statistical analysis such as correlations and principal components.[66] They utilized data from five independent human cohorts (cancer- and Alzheimer-related datasets) to understand the contribution of epigenome to the survival outcomes. When they analyzed the modalities individually based either on gene expression or DNA methylation, they achieved the p-value of 10^{-4}, while when they utilized both modalities (DNA methylation and gene expression), they were able to increase the significance to the p-value of 10^{-7}. This work demonstrated the power multi-omics data integration for the prognostic prediction capabilities of the survival model.

Several other multi-omics data integrations were proposed. Zachariou et al. (2018) developed a "super network", attempting integration of six different types of interactions to identify significant pathways related to a disease.[67] Their method allows

pathway analysis on top genes based on the quantity of shared information between gene pairs utilizing gene expression analysis. It was not demonstrated, however, whether this super network can integrate methylation data.

Ma et al. (2017) developed Edge-Based Module Detection Network (EMDN) for the analysis of differentially co-methylated and co-expressed networks.[68] After constructing multiple networks, the standard modules within these networks are defined as epigenetic modules. The authors compared the EMDN performance with Consensus clustering (CSC),[69] the multiple-modularity method (MolTi)[70] and spectral clustering (SPEC)[71] modules. EMDN outperformed other artificial networks, demonstrating higher accuracy.[68] EMDN, as many other similar algorithms, relies on the establishment of differentially methylated or expressed genes, and thus requires paired comparison, such as normal versus disease samples or case versus control, or untreated versus treated. Our approach, unless we need to calculate the DEI values, as well as iNETgrate approach does not have this limitation.

Another fairly advanced model, INTEND (Integration of Transcriptomic and Epigenomic Data), which addresses the integration of disjointed methylation and gene expression data, was recently published.[72] While INTEND integrates data from the same individual for multiple data sets, it does not use any information matching methylation and gene expression data to the same individual. Instead, INTEND learns a predictive model between the two by training on data sets having a large number of gene expression and methylome data sets from the same analyzed cohorts. At the first step, INTEND is trained to predict gene expression data based on methylation data located close to genes. Then, it compares predicted expression to the expression of the same set of genes stemming from transcriptome analysis. The authors evaluated INTEND performance on cancer datasets spanning 4329 patients by comparing it with four other integration methods: LIGER, Seurat v3, JLMA and MMD-MA, and demonstrated INTEND to be superior to all four.[72] However, the INTEND utility relies on ML procedures, such as LASSO regression rather than following the signal propagation along multiple highly branched pathways and networks.

Integration of protein-coding gene expression data and non-coding RNA expression data

Several software tools were developed to analyze inverse correlation between mRNA and miRNA expressions, including CORNA,[73] MMIA,[74] MAGIA[75] and miARma-Seq.[76]

miARma-Seq integrates the results of interaction between mRNA and miRNA based on the information stored in miRGate database;[77] it relies on established negative correlation. miRGate includes information on miRNA sequences from mRBase[78] and 3-UTR sequences from EnsEMBL,[79] as well as the information about experimentally validated targets stored in miRTarbase,[80] Tarbase[81] and OncomiDB.[82]

The authors profiled samples of colorectal cancer and identified 29 differently expressed miRNAs and 368 mRNA-encoding genes; they found that out of total possible 10 672

correlations, ~60% were statistically significant, with many of them having a positive correlation, rather than the expected negative correlation.[76]

More direct integration of mRNA and miRNA data could be possible if all associations between miRNA and mRNA are known. Since miRNA can target hundreds of mRNAs, and many different miRNAs can target the same mRNA, and also, since many miRNAs positively correlate with the expression of some genes (perhaps by targeting the miRNAs that inhibit those miRNAs), direct estimates of the effects of miRNAs on mRNA expression are hard to calculate.

While substantial effort was made to integrate data from miRNA and mRNA sequencing, almost no effort was made to do the same for lncRNAs and mRNA sequencing. This could be due to the fact that there is no clear relationship between lncRNAs and mRNA expression similar to miRNA/mRNA pairs. However, we found that protein-coding mRNA-based values are better correlated with the long antisense non-coding RNA-based ones rather than with micro-RNA-based. This effect was revealed because of the use of our bioinformatics integrated platform, which allows multi-omics analysis in terms of DEI and SPIA values.

Conclusion

In this work, we attempted to correlate several omics data sets with the disease or/and treatment outcomes using the PAL/SPIA method. We demonstrated positive moderate correlation between antisense lncRNA expression and mRNA expression as well as negative correlation between DNA methylation and mRNA expression. In the future, it would be important to integrate several omics data sets (methylomics, transcriptomics and ncRNAomics, if such data are available from the sample) into the pathway analysis, assigning regulatory (modulatory) coefficients to each ncRNA.

Author contributions

Conceptualization: NB, OK, and IK; data curation: NB, BB, and YI; formal analysis: NB, BB, and YI; funding acquisition: OK and IK; methodology: NB, BB, and YI; project administration: OK and IK; resources: OK and IK; software: NB; supervision: OK and IK; validation: NB; visualization: NB; writing – original draft: NB; writing – review & editing: NB, OK, and IK.

Conflicts of interest

SI, IK and OK are employees of Pathway Rx; Pathway Rx provided funding for this research.

Data availability

No original/new data were used in this work. All data analyzed in this work are available in public databases.

Code availability

Can be found at GitHib: https://github.com/BorisovNM/SPIA_DEI.

Acknowledgements

Notes and references

1 Y. Tao, et al., Cell-free multi-omics analysis reveals potential biomarkers in gastrointestinal cancer patients' blood, Cell Rep. Med., 2023, 4, 101281.

2 A. Maimaiti, et al., DNA methylation regulator-mediated modification patterns and risk of intracranial aneurysm: a multi-omics and epigenome-wide association study integrating machine learning, Mendelian randomization, eQTL and mQTL data, J. Translat. Med., 2023, 21, 660.

3 S. Zhang, et al., Multi-omics analysis reveals Mn exposure affects ferroptosis pathway in zebrafish brain, Ecotoxicol. Environ. Saf., 2023, 253, 114616.

4 L. Ning, et al., Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts, Nat. Commun., 2023, 14, 7135.

5 J. Jeon, E. Y. Han and I. Jung, MOPA: An integrative multi-omics pathway analysis method for measuring omics activity, PloS One, 2023, 18, e0278272.

6 M. Du, et al., Integrated multi-omics approach to distinct molecular characterization and classification of early-onset colorectal cancer, Cell Rep. Med., 2023, 4, 100974.

7 L. Han, et al., A multi-omics integrative network map of maize, Nat. Genet., 2023, 55, 144–153.

8 N. T. Doncheva, et al., Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks, J. Proteome Res., 2023, 22, 637–646.

9 A. Majeed and S. Mukhtar, Protein–Protein Interaction Network Exploration Using Cytoscape, Methods Mol. Biol., 2023, 2690, 419–427.

10 A. Kamburov, et al., Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA, Bioinformatics, 2011, 27, 2917–2918.

11 G. J. Odom, et al., PathwayMultiomics: An R Package for Efficient Integrative Analysis of Multi-Omics Datasets With Matched or Un-matched Samples, Front. Genet., 2021, 12, 783713.

12 S. Canzler and J. Hackermüller, multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data, BMC Bioinf., 2020, 21, 561.

13 T. Liu, et al., PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple

pathway databases, *Nucleic Acids Res.*, 2022, **50**, W551–W559.

14 M. Paczkowska, *et al.*, Integrative pathway enrichment analysis of multivariate omics data, *Nat. Commun.*, 2020, **11**, 735.

15 A. Singh, *et al.*, DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays, *Bioinformatics*, 2019, **35**, 3055–3062.

16 G. Zhou, J. Ewald and J. Xia, OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data, *Nucleic Acids Res.*, 2021, **49**, W476–W482.

17 R. Tibshirani, The lasso method for variable selection in the Cox model, *Stat. Med.*, 1997, **16**, 385–395.

18 C. Wieder, *et al.*, PathIntegrate: Multivariate modelling approaches for pathway-based multi-omics data integration, *PLoS Comput. Biol.*, 2024, **20**, e1011814.

19 T.-M. Nguyen, *et al.*, Identifying significantly impacted pathways: a comprehensive review and assessment, *Genome Biol.*, 2019, **20**, 203.

20 N. Borisov, *et al.*, Quantitation of Molecular Pathway Activation Using RNA Sequencing Data, in *Nucleic Acid Detection and Structural Investigations*, ed. K. Astakhova and S. A. Bukhari, Springer, US, New York, NY, 2020, pp. 189–206. Available at: **https://link.springer.com/10.1007/978-1-0716-0138-9_15** [Accessed November 15, 2019].

21 S. Gao and X. Wang, TAPPA: topological analysis of pathway phenotype association, *Bioinformatics*, 2007, **23**, 3100–3102.

22 M. A.-H. Ibrahim, *et al.*, A topology-based score for pathway enrichment, *J. Comput. Biol.: J. Comput. Mol. Cell Biol.*, 2012, **19**, 563–573.

23 S. Draghici, *et al.*, A systems biology approach for pathway level analysis, *Genome Res.*, 2007, **17**, 1537–1545.

24 A. L. Tarca, *et al.*, A novel signaling pathway impact analysis, *Bioinformatics*, 2009, **25**, 75–82.

25 I. V. Ozerov, *et al.*, In silico Pathway Activation Network Decomposition Analysis (iPANDA) as a method for biomarker development, *Nat. Commun.*, 2016, **7**, 13427.

26 N. Borisov, *et al.*, System, Method and Software for Calculation of a Cannabis Drug Efficiency Index for the Reduction of Inflammation, *Int. J. Mol. Sci.*, 2020, **22**(1), 388.

27 N. Borisov, *et al.*, Application of Drug Efficiency Index Metric for Analysis of Post-Traumatic Stress Disorder and Treatment Resistant Depression Gene Expression Profiles, *Psychoactives*, 2023, **2**, 92–112.

28 M. Sorokin, *et al.*, Algorithmic Annotation of Functional Roles for Components of 3044 Human Molecular Pathways, *Front. Genet.*, 2021, **12**, 617059.

29 M. A. Zolotovskaia, *et al.*, OncoboxPD: human 51 672 molecular pathways database with tools for activity calculating and visualization, *Comput. Struct. Biotechnol. J.*, 2022, **20**, 2280–2291.

30 J. K. W. Lam, *et al.*, siRNA Versus miRNA as Therapeutics for Gene Silencing, *Mol. Ther.–Nucleic Acids*, 2015, **4**, e252.

31 I. Monga, *et al.*, ASPsiRNA: A Resource of ASP-siRNAs Having Therapeutic Potential for Human Genetic Disorders and Algorithm for Prediction of Their Inhibitory Efficacy, *G3 Genes*, 2017, **7**, 2931–2943.

32 H.-H. Huang, *et al.*, A novel meta-analysis based on data augmentation and elastic data shared lasso regularization for gene expression, *BMC Bioinf.*, 2022, **23**, 353.

33 C. Wahlestedt, Targeting long non-coding RNA to therapeutically upregulate gene expression, *Nat. Rev. Drug Discovery*, 2013, **12**, 433–446.

34 D. Stazic, *et al.*, Antisense RNA protects mRNA from RNase E degradation by RNA–RNA duplex formation during phage infection, *Nucleic Acids Res.*, 2011, **39**, 4890–4899.

35 M. Nishizawa, Post-transcriptional inducible gene regulation by natural antisense RNA, *Front. Biosci.*, 2015, **20**, 1–36.

36 T. Kimura, Non-coding Natural Antisense RNA: Mechanisms of Action in the Regulation of Target Gene Expression and Its Clinical Implications, *Yakugaku Zasshi*, 2020, **140**, 687–700.

37 M. A. Zolotovskaia, *et al.*, Pathway Based Analysis of Mutation Data Is Efficient for Scoring Target Cancer Drugs, *Front. Pharmacol.*, 2019, **10**, 1.

38 N. Borisov, *et al.*, Uniformly shaped harmonization combines human transcriptomic data from different platforms while retaining their biological properties and differential gene expression patterns, *Front. Mol. Biosci.*, 2023, **10**, 1237129.

39 H. Y. Huang, *et al.*, miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions, *Nucleic Acids Res.*, 2022, **50**(D1), D222–D230.

40 H. T. Li, *et al.*, RNA mis-splicing drives viral mimicry response after DNMTi therapy in SETD2-mutant kidney cancer, *Cell Rep.*, 2023, **42**(1), 112016.

41 A. Bansal, *et al.*, Integrative Omics Analyses Reveal Epigenetic Memory in Diabetic Renal Cells Regulating Genes Associated With Kidney Dysfunction, *Diabetes*, 2020, **69**(11), 2490–2502.

42 M. V. Ringh, *et al.*, Methylome and transcriptome signature of bronchoalveolar cells from multiple sclerosis patients in relation to smoking, *Mult. Sclerosis*, 2021, **27**(7), 1014–1026.

43 J. Ma, *et al.*, Statistical Methods for Establishing Personalized Treatment Rules in Oncology, *BioMed Res. Int.*, 2015, **2015**, 670691.

44 W. Ma and J. Hu, The linear ANRIL transcript P14AS regulates the NF-κB signaling to promote colon cancer progression, *Mol. Med.*, 2023, **29**, 162.

45 B. Yu, *et al.*, B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells, *Cell*, 2021, **184**, 1790–1803.

46 J. Zhao, *et al.*, Systems pharmacological study illustrates the immune regulation, anti-infection, anti-inflammation, and multi-organ protection mechanism of Qing-Fei-Pai-Du decoction in the treatment of COVID-19, *Phytomedicine*, 2021, **85**, 153315.

47 M. He, *et al.*, Systematic Analysis to Identify the MIR99AHG-has-miR-21-5p-EHD1 CeRNA Regulatory Network as Potential Biomarkers in Lung Cancer, *J. Cancer*, 2024, **15**, 2391–2402.

48 X. Liao, *et al.*, Effect of mechanical unloading on genome-wide DNA methylation profile of the failing human heart, *JCI Insight*, 2023, **8**, e161788.

49 J. Wang, *et al.*, Long noncoding RNA HOTAIR regulates the stemness of breast cancer cells via activation of the NF-κB signaling pathway, *J. Biol. Chem.*, 2022, **298**, 102630.

50 G. Yu, *et al.*, clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters, *OMICS: J. Integr. Biol.*, 2012, **16**, 284–287.

51 J. Zhou, *et al.*, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, *Nat. Commun.*, 2019, **10**, 1523.

52 A. A. Buzdin and N. M. Borisov, MiRImpact as a Methodological Tool for the Analysis of MicroRNA at the Level of Molecular Pathways, in *Handbook of Nutrition, Diet, and Epigenetics*, ed. V. B. Patel and V. R. Preedy, Springer International Publishing, 2019, pp. 2289–2308. Cham Available at: **https://link.springer.com/10.1007/978-3-319-55530-0_91** [Accessed May 29, 2021].

53 S. Panni, *et al.*, Non-coding RNA regulatory networks, *Biochim. Biophys. Acta, Gene Regul. Mech.*, 2020, **1863**(6), 194417.

54 P. Szafranski and P. Stankiewicz, Long Non-Coding RNA FENDRR: Gene Structure, Expression, and Biological Relevance, *Genes*, 2021, **12**(2), 177.

55 S. Sinha, *et al.*, PERCEPTION predicts patient response and resistance to treatment using single-cell transcriptomics of their tumors, *Nat. Cancer*, 2024, **5**, 938–952.

56 T. C. Freeman, *et al.*, Graphia: A platform for the graph-based visualisation and analysis of high dimensional data, *PLoS Comput. Biol.*, 2022, **18**, e1010310.

57 X. Luo, *et al.*, Integrative analysis of DNA methylation and gene expression reveals key molecular signatures in acute myocardial infarction, *Clin. Epigenet.*, 2022, **14**(1), 46.

58 G. Gong, T. Lin and Y. Yuan, Integrated analysis of gene expression and DNA methylation profiles in ovarian cancer, *J. Ovarian Res.*, 2020, **13**(1), 30.

59 L. Miao, *et al.*, Integrated DNA methylation and gene expression analysis in the pathogenesis of coronary artery disease, *Aging*, 2019, **11**(5), 1486–1500.

60 B. Tan, *et al.*, Integrated Analysis of DNA Methylation and Gene Expression in Porcine Placental Development, *Int. J. Mol. Sci.*, 2023, **24**(6), 5169.

61 E. Semik-Gurgul, A. Gurgul and T. Szmatoła, Transcriptome and methylome sequencing reveals altered long non-coding RNA genes expression and their aberrant DNA methylation in equine sarcoids, *Funct. Integr. Genomics*, 2023, **23**(3), 268.

62 C. S. Teng, *et al.*, MethGET: web-based bioinformatics software for correlating genome-wide DNA methylation and gene expression, *BMC Genomics*, 2020, **21**, 375.

63 C. D. Warden, *et al.*, COHCAP: an integrative genomic pipeline for single-nucleotide resolution DNA methylation analysis, *Nucleic Acids Res.*, 2013, **41**(11), e117.

64 B. T. Moghadam, *et al.*, PiiL: visualization of DNA methylation and gene expression data in gene pathways, *BMC Genomics*, 2017, **18**(1), 571.

65 X. Huang, *et al.*, ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data, *Bioinformatics*, 2018, **34**(4), 708–709.

66 S. Sajedi, *et al.*, Integrating DNA methylation and gene expression data in a single gene network using the iNETgrate package, *Sci. Rep.*, 2023, **13**, 21721.

67 M. Zachariou, *et al.*, Integrating multi-source information on a single network to detect disease-related clusters of molecular mechanisms, *J. Proteomics*, 2018, **188**, 15–29.

68 X. Ma, *et al.*, Multiple network algorithm for epigenetic modules via the integration of genome-wide DNA methylation and gene expression data, *BMC Bioinform.*, 2017, **18**(1), 1–13.

69 L. Cantini, *et al.*, Detection of gene communities in multi-networks reveals cancer drivers, *Sci. Rep.*, 2015, **5**, 17386.

70 G. Didier, C. Brun and A. Baudot, Identifying communities from multiplex biological networks, *Peer J.*, 2015, **3**, 1525.

71 M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2006, **74**, 036104.

72 Y. Itai, N. Rappoport and R. Shamir, Integration of gene expression and DNA methylation data across different experiments, *Nucleic Acids Res.*, 2023, **51**(15), 7762–7776.

73 X. Wu and M. Watson, CORNA: testing gene lists for regulation by microRNAs, *Bioinformatics*, 2009, **25**(6), 832–833.

74 S. Nam, *et al.*, microRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression, *Nucleic Acids Res.*, 2009, **37**, W356–W362.

75 G. Sales, *et al.*, MAGIA, a web-based tool for miRNA and genes integrated analysis, *Nucleic Acids Res.*, 2010, **38**, W356–W359.

76 E. Andrés-León and A. M. Rojas, miARma-Seq, a comprehensive pipeline for the simultaneous study and integration of miRNA and mRNA expression data, *Methods*, 2019, **152**, 31–40.

77 E. Andrés-León, *et al.*, miRGate: a curated database of human, mouseand rat miRNA-mRNA targets, *Database*, 2015, **2015**, bav035.

78 A. Kozomara and S. Griffiths-Jones, miRbase: annotating high confidence microRNAs using deep sequencing data, *Nucleic Acids Res.*, 2014, **42**, D68–D73.

79 A. Yales, *et al.*, Ensembl, *Nucleic Acids Res.*, 2016, **44**(D1), D710–D716.

80 C. H. Chou, *et al.*, miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database, *Nucleic Acids Res.*, 2016, **44**(D1), D239–D247.

81 T. Vergoulis, *et al.*, TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support, *Nucleic Acids Res.*, 2012, **40**, D222–D229.

82 D. Wang, *et al.*, OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs, *Bioinformatics*, 2014, **20**(15), 2237–2238.