



Research article

Panomicon: A web-based environment for interactive, visual analysis of multi-omics data

Rodolfo S. Allendes Osorio^{a,*},¹ Johan T. Nyström-Persson^{b,**}, Yosui Nojima^a, Yuji Kosugi^b, Kenji Mizuguchi^{a,c}, Yayoi Natsume-Kitatani^{a,d,***}^a Laboratory of Bioinformatics, Artificial Intelligence Center for Health and Biomedical Research (ArCHER), National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki-shi, Osaka, 567-0085, Japan^b Lifematics Ltd., Sanshin Hatchobori bldg. 5F, 2-25-10 Hatchobori, Chuo-ku, Tokyo-to, 104-0032, Japan^c Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita-shi, Osaka, 565-0871, Japan^d Laboratory of In-silico Drug Design, Center of Drug Design Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki-shi, Osaka, 567-0085, Japan

ARTICLE INFO

Keywords:

Bioinformatics
Computational biology
Genetics
Gene networks
Multi-omics
microRNA
Visualization
Interactive analysis
Gene interactions
Lung cancer

ABSTRACT

Multi-omics analyses, combining transcriptomics, genomics, proteomics, and so on, have led to important insights in many areas of biology and medicine. To support these analyses, software that can handle the difficulties associated with multi-omics datasets is crucial. Here, we describe *Panomicon*, a web-based, interactive analysis environment for multi-omics data. Building on Toxygates, a tool previously created to study single-omics data that features interactive clustering, heatmaps, and user data uploads, Panomicon introduces improvements for the storage and handling of additional omics types, as well as tools for the generation and visualization of interaction networks between different types of omics data. Panomicon is a new type of environment for the collaborative study of multi-omics data, both for users uploading data to our server and for groups wishing to host their own deployment of Panomicon. We demonstrate Panomicon's capabilities by revisiting a microRNA-mRNA interaction networks study in a non-small cell lung cancer dataset.

1. Introduction

Increasingly widespread availability of multi-omics datasets is a driving force in the study of diverse biological phenomena. The joint analysis of messenger RNA (mRNA), micro-RNA (miRNA), proteomics, and other kinds of omics data has become central to many areas in biomedical and biological research. However, it is increasingly recognised that working with multi-omics data poses various challenges, among which data hosting, reproducibility of analyses, and availability of resources for integrative data visualization and exploration are at the top of the list (Conesa and Beck, 2019; Subramanian et al., 2020). There are various efforts being pushed in the development of storage, analysis and visualization capabilities that specifically target multi-omics data (Subramanian et al., 2020).

Omics datasets are stored on both public and private repositories—often large, multi-database repositories such as the ones provided by the National Center for Biotechnological Information (NCBI) (Sayes et al., 2020), the European Bioinformatics Institute (EBI) (Cook et al., 2020) or the National Genomics Data Center (NGDC) (Zhang et al., 2020). However, the majority of these repositories are limited to a single type of omics data.

As for analysis and visualization tools, a number of powerful tools exist, generally falling into one of two categories.

On the one hand, there are integrated databases and tools that focus on a specific task or biological theme: functional enrichment tools, such as ToppGene (Chen et al., 2009), pathway databases, such as KEGG (Kanehisa et al., 2019) and Reactome (Jassal et al., 2020), and miRNA-mRNA interaction databases, such as miRTarBase (Huang et al.,

* Corresponding author.

** Corresponding author.

*** Corresponding author.

E-mail addresses: rallendes@nibiohn.go.jp (R.S. Allendes Osorio), jtnystrom@gmail.com (J.T. Nyström-Persson), natsume@nibiohn.go.jp (Y. Natsume-Kitatani).¹ Lead Contact.

2020). Generally speaking, these are available on the web and serve a single specific purpose.

On the other hand, there are powerful general-purpose analysis tools, such as Cytoscape (Shannon et al., 2003), or R (R Core Team, 2018) (with suitable third-party libraries) which allow for standalone analysis, including network construction and editing, once the software environment has been properly configured and auxiliary tools and databases have been connected. Such general purpose tools are powerful, but they generally are not pre-packaged with integrations with other tools and databases—users must build or at least configure these themselves—and they require users to learn a wide array of general-purpose functionality, determine which of those functions suit their needs, and use them to construct a suitable workflow. In the case of R, programming skills are also needed. Moreover, even these general purpose tools often do not provide the full set of functionality necessary to analyse multi-omics datasets, requiring users to work with multiple such tools to conduct these analyses.

Thus, despite the growing number of existing tools, and following the conclusions of (Subramanian et al., 2020) there is a need for a single, integrated application in which a complete multi-omics analysis can be performed. To best support multi-omics research, such an application needs to facilitate the following: (1) storing and providing efficient access to heterogeneous data; (2) analysing data and identifying new features both within individual datasets and in the relationships between multiple datasets; and (3) reviewing the analysis results, all in order to facilitate knowledge discovery and dissemination.

Here, we introduce *Panomicon*, an actively developed application that aims to fulfill this need by integrating various data sources and analytics and combining these with visualizations and open/ended data import and export, all in the web browser. We believe that this is the best way to leverage existing resources to enable powerful collaborative data analysis.

Panomicon implements the following functionality:

1. In order to support storage and efficient data access, it implements a data model that supports the integration of multiple data types (demonstrated here through the simultaneous storage of mRNA and microRNA, although the concepts generalize to other types as well);
2. For the analysis and identification of new features in data, it provides fast, interactive, dynamic generation and regeneration of networks, based on mRNA-miRNA interaction databases that have been pre-loaded into the application; and
3. Finally, to facilitate the review of the analysis results, it allows the presentation and visualization, both in dual tables and through conventional 2-dimensional display of nodes and edges graphs, of the generated networks.

Panomicon positions itself as a whole workflow support platform that allows end-to-end integration and analysis of multi-omics data.

Panomicon provides a rich, web-based integrated environment for multi-omics data analysis with data upload and management capabilities. We think that this would be especially helpful for groups of people wishing to analyse the same data, whether in a single laboratory or across multiple institutions. In such a setting, one person could act as the data uploader and manager, while others could explore and analyse the data that has been uploaded. Administrators can also create their own managed versions of *Panomicon*, given that it is open source, rather than use the server that we provide. In this way, *Panomicon* offers a solution to collaborative data analysis and management needs, while at the same time providing everything needed for individuals to establish their own multi-omics databases, even at large scale. All tools in *Panomicon* can handle large datasets and have been tested on more than 20,000 samples and 800 million data points in a single database.

We emphasise particularly the efficiency of *Panomicon*'s network manipulation engine. A network consisting of all mRNA and miRNA probes, as well as interactions between them, can be manipulated in-

memory on the server, while users can inspect and visualise a partial view of this network. Because only a small amount of data needs to be sent to the user, even for large networks consisting of over 100,000 nodes, updating, sorting, and filtering takes 1–2 s. To the best of our knowledge, this is unique among web-based multi-omics analysis tools.

Furthermore, the tools in *Panomicon* are specifically designed for omics analysis, particularly mRNA and miRNA data, and comparative analysis of different sample conditions, such as treated and control groups. It contains a pre-designed data environment where many widely used databases and third party data sources, such as mRNA-miRNA interaction databases, pathways, and gene ontology (GO) terms, are already available. Since tools and data access in *Panomicon* have all been developed together with this domain in mind, many valuable studies can be performed with very little configuration and without external software, which would be required when conducting similar analyses with more general purpose tools. Further details can be found in the *Panomicon* user guide (available at <https://toxygates.nibiohn.go.jp/panomicon/toxygatesManual.pdf>) and in our paper about *Panomicon*'s precursor (Nyström-Persson et al., 2017).

Panomicon is freely available to the public at <https://toxygates.nibiohn.go.jp/panomicon>. The source code of the web application is also available, under a GNU General Public License v2.0, at <https://github.com/Toxygates/Panomicon>.

2. Results

Three kinds of functionality make up the core of *Panomicon*: (1) handling of multi-omics data; (2) dynamic generation of networks; and (3) interactive visualization of networks. In this section, we discuss each of them in detail.

2.1. Multi-omics data handling

To provide fundamental support for this kind of data analysis, *Panomicon* accommodates multiple types of omics data. The system is divided into multiple screens that support various workflows. On the sample group definition screen, where an analysis can be started by selecting experimental samples to include in a sample group to be analysed, users can select an omics data type such as mRNA or miRNA in addition to sample conditions. Sample groups allow expression values for an omics type to be averaged across the samples contained in the group. To perform analyses involving multiple data types, users can create one or more groups of each type. Whenever multiple groups with different data types are activated simultaneously in this way, the “dual table” mode, described below, will be activated.

2.2. Linked expression data tables

On the data screen, where users can view expression values and other data for their selected sample groups, we have introduced a split table (“dual table”) view that simultaneously displays data from two different omics types (If only a single type of sample has been chosen, then a simpler single table view is displayed instead). In this view, two tables are active simultaneously, each corresponding to one data type. For example, the left (primary) table might display mRNA data, in which case the right (secondary) table would display miRNA data. Users can switch the positions of the two tables by clicking a button that “flips” the tables.

In the dual table view, the left-hand table is always considered to be the main table. It drives the display of the right-hand table, which is considered to be secondary. The main table can be controlled through paging, sorting, filtering, and choosing particular gene sets of interest, which indirectly also controls the right table. For example, if the main table, after user operations such as filtering and sorting, displays a set of 50 mRNAs, then the right table will immediately display those miRNAs that are associated with these 50 mRNAs. mRNA-miRNA association sources are configured by the user to construct the dual table view.

Currently we include, for human, rat and mouse data, miRTarBase (Huang et al., 2020) and miRDB (Chen and Wang, 2020); and for human data, prediction results from miRAW (Pla et al., 2018). These can be filtered by score and support type cutoffs.

2.3. Visual network analysis

By using the cytoscape.js library (Franz et al., 2016), we have added interactive network visualization capabilities to Panomicon. These capabilities have been deeply integrated with previously existing functions for gene set management as far as possible, as we will show below.

To construct a network, it is first necessary to load a dual table display of two omics types, as described above. When the user activates network visualization, the application will construct a network by taking probes from the main table (for example, mRNA probes), based on the currently active sort order and expression value filters (for example, with respect to *p*-values and log-2 fold values), then adding probes from the right table, together with their interactions (regulatory relationships, in the case of mRNAs being regulated by miRNAs). The interaction network will represent probes from the main and secondary tables as nodes, and interactions between probes as edges.

Such a network is fully defined by the following parameters:

- The mRNA and miRNA sample groups for which expression values are being displayed in the two tables
- The chosen miRNA-mRNA association sources and any cutoff limits applied to them
- The probes visible in the main table (for example, those associated with a chosen gene set)
- Any filtering that the user may have applied to the primary and secondary tables.

In the network visualization, a size limit cutoff may apply, with probes being selected according to the sort order of mRNAs or miRNAs applied by the user. The main reason for limiting the size of networks (currently networks are limited to a maximum of 100 nodes of the primary type, with no limit for the secondary type) is to speed up the network layout process and user interactions, since many network layout algorithms become slow when the network is large.

In designing the network analysis functions our principle is to allow multiple analysis functions to work together in an open-ended, interactive environment supporting many use cases. Networks can be constructed from gene sets, and conversely the nodes in networks can be saved as gene sets. As described above, filtering and sorting of gene sets is reflected in network display. It is also possible to interactively tune network parameters and inspect the results visually, until one is satisfied with the overall outcome. For example, filters on expression values in data columns as well as the choice of miRNA-mRNA associations may be edited directly from the network display, without closing the network view. Such changes are immediately reflected in a new, updated network. This allows users to gradually adjust filters until the desired outcome is obtained.

It is also possible to comparatively analyse two networks side by side and highlight common nodes. For example, networks based on two different experimental conditions or on two different filtering conditions can easily be compared. In such an analysis it is easy to find the location of common nodes in the networks, as selecting a node in one of them—either in the left or in the right network—will highlight the node corresponding to the same probe in the other network, if one exists. Such side-by-side networks can at any time be merged together and automatically laid out again in a way that intuitively highlights features shared by the two networks. Merging two networks in this way places common nodes in a visually separate area for easy identification.

In summary, Panomicon supports simultaneous analysis of mRNA and miRNA so that networks can be constructed by choosing a gene set of either of the two types as a starting point. While inspecting and manipulating such networks, whether in dual table mode or in network visualization mode, the user can construct new gene sets of either type. Thus, it is possible to draw immediate conclusions from networks as well as downstream conclusions from further analysing derived gene sets. All of these operations are sufficiently general to apply to any pair of omics types, as long as quantitative probe data and interaction pair data is available. For example, as a next step in Panomicon development, we are considering co-analysis of proteomics and mRNA data. In such a case, by designating mRNA data as a common denominator it is theoretically possible to compare mRNA, miRNA, and proteomics data simultaneously by inspecting mRNA-protein and mRNA-miRNA networks.

2.4. Case study

In an effort to identify biomarkers for Non-Small Cell Lung Cancer (NSCLC), Ma and colleagues (Ma et al., 2011) used microarrays to measure the expression levels of miRNA and mRNA simultaneously in paired NSCLC and adjacent normal tissues for 6 patients. Using TargetScan (Agarwal et al., 2015) they identified the down-regulated mRNA targets for hsa-miR-96 and then performed Gene Ontology enrichment for Biological Process using WebGestalt (Liao et al., 2019). Their conclusion was that hsa-miR-96 has potential as a biomarker for NSCLC.

To demonstrate the capabilities of Panomicon, we will replicate the whole of the analysis undertaken by Ma et al. (2011). We will show that it is possible to store multi-omics datasets on Panomicon and use the integrated functionalities available to find relationships between different data types, analyse them, and review the biology behind the generated results. All these operations require little to no interaction with other repositories or tools.

Figure 1 summarizes the flow of activities that can be performed using Panomicon. In the following, we give details on how these can be used to replicate the study of NSCLC data.

The original data used by Ma and colleagues is publicly available on the Gene Expression Omnibus. In addition, for the purpose of this exercise, this data has been uploaded and made publicly available as part of the GEO dataset on Panomicon. A detailed description of the upload process can be found in the supplementary file “Panomicon - a case study” provided together with this manuscript.

Although the example studied here is relatively small for the sake of simplicity, Panomicon has been tested with large databases of over 20,000 samples. Panomicon contains separate tools for the efficient management of such large datasets. Thus, the methods we demonstrate below can also be used in the same fashion to analyse large scale data, which makes Panomicon a suitable technology choice for groups that want to create their own large multi-omics databases.

Data selection in Panomicon is achieved through the definition of *Sample Groups*, i.e., lists of experimental samples to be analyzed. Our study deals with the relationship between mRNA and miRNA expression values, so two sample groups need to be defined: one for mRNA and one for miRNA. The NSCLC mRNA sample group is defined using all Adenocarcinoma and Squamous samples available for the mRNA/human/in vivo/lung/simple combination. Similarly, the NSCLC miRNA sample group is defined using all samples available when selecting the miRNA option in the first drop-down box of the data source selection controls.

Once two sample groups have been defined, it is possible to generate a network. For the generation of mRNA-miRNA networks, a number of sources for mRNA-miRNA interactions are available for use on Panomicon. Here we choose miRTarBase (Huang et al., 2020), with a “Functional MTT” support type as cutoff, so as to include only experimentally verified relations.

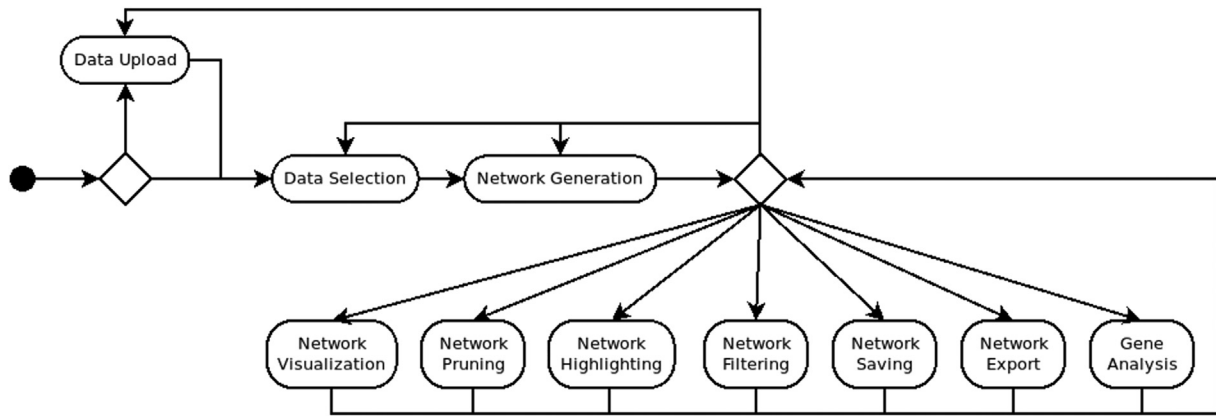


Figure 1. Sample analysis workflow for Panomicon. Panomicon's flexible implementation gives users a large degree of freedom in analysing data, starting from the optional Data Upload stage.

Next, the network must be pruned, in order to focus on the subgraph consisting only of miRNA hsa-miR-96-5p and its target genes. To do so, the network construction must be driven through a selection of miRNA probes, and a *gene set* consisting only of hsa-miR-96-5p must be enabled.

By convention, the probes in the left table (in the *View Data* tab) drive the search for targets, which are displayed in the right table. Thus, to analyze a network driven by the specification of miRNA elements, we first need to exchange the position of the two omics data types. This can be done by clicking on the *Flip mRNA-miRNA* button.

Once miRNA expression values are displayed on the left table we use the *Gene Sets* menu to search for hsa-miR-96-5p. Applying a *gene set* restricts a table view to only the probes or genes contained in the set. In our case, once the *gene set* is saved (and thus automatically applied), hsa-miR-96-5p is shown in the miRNA table on the left, and only mRNA targeted by hsa-miR-96-5p are shown in the right table. This list consists of 26 target genes (in 114 individual probes).

To visualize this network using a traditional node and edge representation, the user simply needs to select the corresponding option from the *Network* menu of the *View Data* tab.

Since the original authors' interest lies in the identification of targets for the treatment of NSCLC, we focus on the genes that depict an up-down regulation relationship to miR-96, i.e., those that are down-regulated due to upregulation of miR-96. This highlighting of the network can be achieved via the application of a color scale.

Furthermore, filters can also be used to prune elements from the visualization. In the visualization screen, we define a filter to be applied to mRNA elements whose log2 fold change is less than or equal to 0. The displayed network reduces to only 10 genes, in 11 individual probes.

Figure 2 shows the pruned network, that includes only hsa-miR-96-5p and its down-regulated targets. Notice that, in accordance with the applied color scale, all mRNA are displayed using different shades of red, indicating down-regulation, whilst miR-96 is shown in blue due to its up-regulation.

In the final step in demonstrating Panomicon's capacity to handle a complete analysis pipeline, we perform functional enrichment analysis in order to investigate the biological context behind the generated network and explore its relation to biological processes.

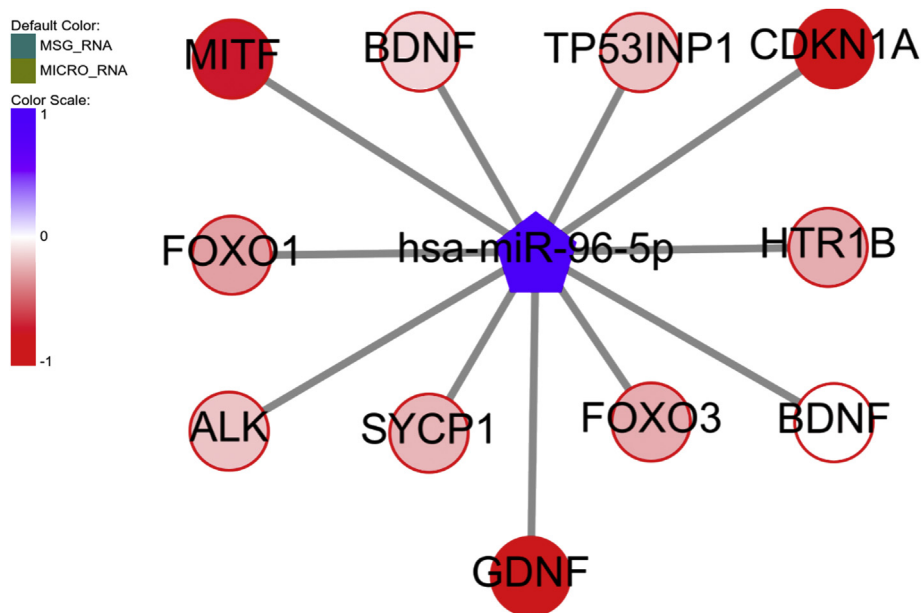


Figure 2. Network of Down-regulated targets for hsa-miR-96-5p. Notice that although only 10 individual target genes are found, two probes associated with BDNF are shown separately.

Functional enrichment analysis in Panomicon can be performed thanks to its integration with TargetMine (Chen et al., 2019), a data warehouse for drug discovery. The functional enrichment analysis is performed using the list of genes currently driving the network generation (left-hand side of the table view), so in order to perform it, we must select the targets from the network.

From the network visualization, we first save the network, then select the *Save probes as gene set* option of the *Network* menu to extract a Gene set that includes only mRNA probes. Finally, we apply the gene set, making sure that mRNA samples are currently being displayed on the left-side table. The functional enrichment analysis is then performed from the *Tools* menu.

We perform two different types of functional enrichment analysis. First, we perform GO Term enrichment analysis using the following configuration: Biological-process filtering, 0.1 *p*-value cutoff (no results were found under a 0.05 threshold value), and Benjamini-Hochberg correction. The complete list of enriched terms is shown in Table 1. Second, we perform functional enrichment analysis against the KEGG database, using a *p*-value cutoff of 0.05 and a Benjamini-Hochberg correction. The complete list of enriched pathways is shown in Table 2. In the following, we provide some biological interpretations of these results.

Among the GO Term results, we highlight the enrichment of “Apoptotic process”. It is widely known that apoptosis is suppressed in cancer cells due to the mutation of p53. Wild-type p53 is a potent inducer of apoptosis, but the expression of p53 is down-regulated in lung cancer tumor tissue compared with non-tumor tissue (Muller and Vousden, 2014; Su et al., 2017). Accordingly, Panomicon shows that gene expression of Tumor Protein P53 Inducible Nuclear Protein 1 (TP53INP1) is down-regulated by miR-96-5p (see Figure 2), which phosphorylates p53 and is involved in inducing apoptosis (Okamura et al., 2001; Shah-bazi et al., 2013).

As for the enriched pathways, worth noticing are pathways entitled “Non-small cell lung cancer” and “FoxO signaling pathway”, where the matched genes for the first pathway are ALK, CDKN1A, (also called p21) and FOXO3 whilst those for the second pathway are p21, FOXO1, and FOXO3 (see Table 2).

It has been shown that p21, together with p53, is involved in cell cycle arrest when down-regulated in tumor tissue (Normand et al., 2005; Zhu et al., 2019). FOXO1 and FOXO3 are involved in inducing apoptosis and both have been found to be related to Non-small cell lung cancer progression when down-regulated (Gao et al., 2018; Meng et al., 2010;

Mikse et al., 2010; Shi et al., 2018). Thus, the results found with Panomicon are consistent with what one would expect biologically.

There is one aspect of our results which are contrary to previous findings: ALK has been found to be up-regulated in lung cancer tumor tissue (Wang et al., 2011). However, ALK is widely known as a fusion gene: the C-terminus of ALK protein fuse with several proteins and the expression of ALK gene in lung cancer cells is different in each exon (Furugaki et al., 2019; Scheijen and Griffin, 2002), which in turn could explain the difference between our results and previous reports.

3. Discussion

The growing availability of multi-omics datasets has become a driving force in the study of diverse biological phenomena. However, together with the opportunities these new datasets provide, a series of new challenges, including data hosting, reproducibility of analyses, and facilitation of integrative data exploration are also introduced (Conesa and Beck, 2019).

We address all of these issues by introducing Panomicon, a web-based integrated environment suitable for the storage, analysis, and visualization of multi-omics datasets, with a focus on the analysis of biological networks.

To achieve this goal, Panomicon introduces the following core functionalities:

1. Data handling. Extension of the underlying data structure model to accommodate the storage and handling of multiple types of omics datasets.
2. Network generation. Built-in interaction databases that allow elements from different omics types to be linked together, allowing the generation of biological networks
3. Network visualization and analysis. Integration of the Cytoscape. js library to allow the visual display of the generated networks and provision of functionality for the analysis, persistence and sharing of networks.

Panomicon adheres to open source development methods, and provides integrated access to other biological resources, such as TargetMine.

The design philosophy of Panomicon emphasizes an interactive environment with multiple interacting tools that can be combined freely, as well as a hybrid data model, to support these tools. The hybrid data model also allows for two kinds of scalability: quantitative scaling of the

Table 1. Gene Ontology enrichment. List of all GO terms enriched (biological process) for the set of down-regulated genes found using Panomicon's integration to TargetMine. Holm-Bonferroni correction, *p*-value < 0.1.

ID	Description	<i>p</i> -value	Matches
GO:0051402	Neuron apoptotic process	0.0600	3
GO:0010604	Positive regulation of macromolecule metabolic process	0.0619	8
GO:0012501	Programmed cell death	0.0664	7
GO:0009267	Cellular response to starvation	0.0673	3
GO:0006915	Apoptotic process	0.0676	7
GO:0043523	Regulation of neuron apoptotic process	0.0705	3
GO:0031325	Positive regulation of cellular metabolic process	0.0714	8
GO:0008219	Cell death	0.0730	7
GO:0010508	Positive regulation of autophagy	0.0739	3
GO:0042594	Response to starvation	0.0780	3
GO:0010941	Regulation of cell death	0.0787	6
GO:0051173	Positive regulation of nitrogen compound metabolic process	0.0837	8
GO:0043067	Regulation of programmed cell death	0.0870	6
GO:0009893	Positive regulation of metabolic process	0.0951	8
GO:0042981	Regulation of apoptotic process	0.0992	6

Table 2. KEGG Pathway Enrichment results. List of KEGG enriched pathways found for the set of down-regulated genes using Panomicon's integration with TargetMine. Benjamini Hochberg correction, p-value < 0.05.

ID	Description	p-value	Matches
hsa05223	Non-small cell lung cancer	0.00614	ALK CDKN1A FOXO3
hsa04068	FoxO signaling pathway	0.0240	CDKN1A FOXO1 FOXO3
hsa04218	Cellular senescence	0.0289	CDKN1A FOXO1 FOXO3
hsa05202	Transcriptional misregulation in cancer	0.0338	CDKN1A FOXO1 MITF
hsa05200	Pathways in cancer	0.0382	ALK CDKN1A FOXO1 MITF

data (where large volumes of data are supported) and increasing complexity of data (where unforeseen or unexpected data structures are supported). Panomicon has built on these concepts. Different tools in Panomicon communicate by passing gene sets and networks to each other as well as by generating gene sets from networks and vice versa, as we have shown in the case study above. Dual table networks can scale to tens of thousands of probes (genes) on either side of the table and respond instantly to sorting and filtering.

The major strength of Panomicon is its suitability for supporting entire biological workflows, from data storage to analysis, visualization, and biological interpretation. This is demonstrated by our case study, where the need for multiple software tools is obviated by the different functionalities integrated into Panomicon. A step-by-step guide for the replication of our case study is provided as supplementary information. Moreover, with the improvements described above, Panomicon now fundamentally supports a wide range of multi-omics datasets.

The use of Panomicon is not limited by the data and functionality currently implemented in the publicly available version. Thanks to its efficient data model and open source philosophy, Panomicon can be deployed by institutions to create new large scale multi-omics databases with integrated capabilities for collaborative data analysis.

3.1. Future directions

We see the potential for Panomicon to be used even more widely by researchers as well as application developers working with multi-omics data. To that end, in the future we plan to pursue the following improvements:

1. Further generalizing the user interface and data model, as well as importing more data sets of general interest;
2. Developing system-wide support for more omics types and platforms, to facilitate import and work with different kinds of data, and
3. Providing well-defined and documented configurability and extensibility which, together with the open-source spirit that informs the development of Panomicon, would allow users to develop their own derived applications to meet the needs of their own research.

4. Experimental procedures

4.1. Resource availability

4.1.1. Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Rodolfo S. Allendes Osorio (ralles@nibiohn.go.jp).

4.1.2. Materials availability

This study did not generate new unique reagents. There are no restrictions for use of the software presented.

4.1.3. Data and code availability

The dataset analyzed in this article, GSE29250 (Ma et al., 2011), is publicly available from the Gene Expression Omnibus (GEO) at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29250>.

Panomicon is an open source project, released under a GNU General Public License v2.0. The main application URL is <https://toxygates.nibiohn.go.jp/panomicon/>. The source code for Panomicon can be found at: <https://github.com/toxygates/panomicon>. Documentation can be found at <https://toxygates.nibiohn.go.jp/panomicon/toxygatesManual.pdf>.

4.2. Method details

To help construct miRNA-mRNA networks in Panomicon, we have added the miRBase v. 21 miRNA set (Kozomara et al., 2019) and several miRNA-mRNA association sources. MiRDB v5.0 contains predicted targets for human, mouse and rat (Chen and Wang, 2020). miRTarBase contains experimentally verified targets for human, mouse and rat (Huang et al., 2020). Finally, the MiRAW algorithm was used to generate human predicted targets, using a novel method that applies deep learning to the entire transcript, and not only the predicted binding sites (Pla et al., 2018). The parameters used to generate the data in Panomicon were 6_1_10_AE10_NLL (candidate binding site contains at least 6 base pairs between nucleotides 1 and 10, accessible energy filter -10 kcal/mol, negative log-likelihood).

The dataset GSE29250 was imported from the Gene Expression Omnibus (GEO) and preloaded to Panomicon. MicroRNA data, measured using the platform GPL8179 (Illumina Human v2 MicroRNA expression beadchip) was converted into miRBase identifiers using an approximate method. Messenger RNA data, originally measured using GPL10558 (Illumina HumanHT-V2 4.0 expression beadchip) was converted into RefSeq transcript identifiers using official annotations from Illumina. Ensembl release 99 (Cunningham et al., 2019) was used to obtain further RefSeq annotations. The imported data can be found under the label 'GEO' in the list of Data Sources available to use in Panomicon. Detailed instructions for the uploading of this dataset are included in the Case Study supplementary guide, provided together with this manuscript.

The case study was performed on the publicly hosted version of Panomicon, available at <https://toxygates.nibiohn.go.jp/panomicon/>. This public version is maintained at the AI Center for Health and Biomedical Research, part of the National Institutes of Biomedical Innovation, Health and Nutrition, in Osaka, Japan.

Declarations

Author contribution statement

Rodolfo S. Allendes Osorio: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Johan T. Nyström-Persson Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Yosui Nojima: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Yuji Kosugi: Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Kenji Mizuguchi, Yayoi Natsume-Kitatani: Conceived and designed the experiments; Wrote the paper.

Funding statement

Kenji Mizuguchi was supported by Japan Society for the Promotion of Science (Grants-in-aid for Scientific Research 25430186 and 25293079). Yayoi Natsume-Kitatani was supported by Ministry of Health, Labour and Welfare and Japan Agency for Medical Research and Development (16ak0101010h0005).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2020.e04618>.

Acknowledgements

The authors wish to thank Chen Yi-An (NIBIOHN) for help, feedback and suggestions during the Toxygates development process.

References

- Agarwal, V., Bell, G.W., Nam, J.-W., Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4, e05005.
- Chen, Y., Wang, X., 2020. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48, D127–D131.
- Chen, J., Bardes, E.E., Aronow, B.J., Jegga, A.G., 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* 37, W305–W311.
- Chen, Y.-A., Tripathi, L.P., Fujiwara, T., Kameyama, T., Itoh, M.N., Mizuguchi, K., 2019. The TargetMine data warehouse: enhancement and updates. *Front. Genet.* 10.
- Conesa, A., Beck, S., 2019. Making multi-omics data accessible to researchers. *Sci. Data* 6, 1–4.
- Cook, C.E., Stroe, O., Cochrane, G., Birney, E., Apweiler, R., 2020. The European Bioinformatics Institute in 2020: building a global infrastructure of interconnected data resources for the life sciences. *Nucleic Acids Res.* 48, D17–D23.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddut, S., Cummins, C., Davidson, C., Dodiya, K.J., Gall, A., Girón, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Marugán, J.C., Maurel, T., McMahon, A.C., Moore, B., Morales, J., Mudge, J.M., Nuhn, M., Ogegh, D., Parker, A., Parton, A., Patricio, M., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sparrow, H., Stapleton, E., Szuba, M., Taylor, K., Threadgold, G., Thormann, A., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Yates, A.D., Zerbino, D.R., Flicek, P., 2019. Ensembl 2019. *Nucleic Acids Res.* 47, D745–D751.
- Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O., Bader, G.D., 2016. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 32, 309–311.
- Furugaki, K., Mochizuki, M., Kohno, M., Shu, S., Harada, N., Yoshimura, Y., 2019. Expression of C-terminal ALK, RET, or ROS1 in lung cancer cells with or without fusion. *BMC Canc.* 19, 301.
- Gao, Z., Liu, R., Ye, N., Liu, C., Li, Xiuli, Guo, X., Zhang, Z., Li, Xiaoxi, Yao, Y., Jiang, X., 2018. FOXO1 inhibits tumor cell migration via regulating cell surface morphology in non-small cell lung cancer cells. *Chem. Pharm. Bull.* 48, 138–148.
- Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., Tang, Y., Chen, Y.-G., Jin, C.-N., Yu, Y., Xu, J.-T., Li, Y.-M., Cai, X.-X., Zhou, Z.-Y., Chen, X.-H., Pei, Y.-Y., Hu, L., Su, J.-J., Cui, S.-D., Wang, F., Xie, Y.-Y., Ding, S.-Y., Luo, M.-F., Chou, C.-H., Chang, N.-W., Chen, K.-W., Cheng, Y.-H., Wan, X.-H., Hsu, W.-L., Lee, T.-Y., Wei, F.-X., Huang, H.-D., 2020. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res.* 48, D148–D154.
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorsler, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D’Eustachio, P., 2020. The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., Tanabe, M., 2019. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595.
- Kozomara, A., Birgaonu, M., Griffiths-Jones, S., 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162.
- Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., Zhang, B., 2019. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47, W199–W205.
- Ma, L., Huang, Y., Zhu, W., Zhou, S., Zhou, J., Zeng, F., Liu, X., Zhang, Y., Yu, J., 2011. An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. *PLoS One* 6, e26502.
- Meng, J., Fang, B., Liao, Y., Chresta, C.M., Smith, P.D., Roth, J.A., 2010. Apoptosis induction by MEK inhibition in human lung cancer cells is mediated by bim. *PLoS One* 5, e13026.
- Mikse, O.R., Blake, D.C., Jones, N.R., Sun, Y.-W., Amin, S., Gallagher, C.J., Lazarus, P., Weisz, J., Herzog, C.R., 2010. FOXO3 encodes a carcinogen-activated transcription factor frequently deleted in early-stage lung adenocarcinoma. *Cancer Res.*
- Muller, P.A.J., Vousden, K.H., 2014. Mutant p53 in cancer: new functions and therapeutic opportunities. *Canc. Cell* 25, 304–317.
- Normand, G., Hemmati, P.G., Verdoot, B., von Haefen, C., Wendt, J., Güner, D., May, E., Dörken, B., Daniel, P.T., 2005. p14ARF induces G2 cell cycle arrest in p53- and p21-deficient cells by down-regulating p34cdc2 kinase activity. *J. Biol. Chem.* 280, 7118–7130.
- Nyström-Persson, Johan, Natsume-Kitatani, Yayoi, Igarashi, Yoshinobu, Satoh, Daisuke, Mizuguchi, Kenji, 2017. Interactive toxicogenomics: gene set discovery, clustering and analysis in toxycates. *Sci. Rep.* 7 (1).
- Okamura, S., Arakawa, H., Tanaka, T., Nakanishi, H., Ng, C.C., Taya, Y., Monden, M., Nakamura, Y., 2001. p53DINP1, a p53-Inducible Gene, Regulates p53-Dependent Apoptosis. *Mol. Cell* 8, 85–94.
- Pla, A., Zhong, X., Rayner, S., 2018. miRAW: a deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. *PLoS Comput. Biol.* 14, e1006185.
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Sayers, E.W., Beck, J., Brister, J.R., Bolton, E.E., Canese, K., Comeau, D.C., Funk, K., Ketter, A., Kim, S., Kimchi, A., Kitts, P.A., Kuznetsov, A., Lathrop, S., Lu, Z., McGarvey, K., Madden, T.L., Murphy, T.D., O’Leary, N., Phan, L., Schneider, V.A., Thibaud-Nissen, F., Trawick, B.W., Pruitt, K.D., Ostell, J., 2020. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 48, D9–D16.
- Scheijen, B., Griffin, J.D., 2002. Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene* 21, 3314–3333.
- Shahbazi, J., Lock, R., Liu, T., 2013. Tumor protein 53-induced nuclear protein 1 enhances p53 function and represses tumorigenesis. *Front. Genet.* 4.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Shi, F., Li, T., Liu, Z., Qu, K., Shi, C., Li, Y., Qin, Q., Cheng, L., Jin, X., Yu, T., Di, W., Que, J., Xia, H., She, J., 2018. FOXO1: another avenue for treating digestive malignancy? *Sem. Cancer Biol. FOXO Family Cancer* 50, 124–131.
- Su, P., Wang, F., Qi, B., Wang, T., Zhang, S., 2017. P53 regulation-association long non-coding RNA (lncRNA PRAL) inhibits cell proliferation by regulation of P53 in human lung cancer. *Med. Sci. Mon.* 23, 1751–1758.
- Subramanian, I., Verma, S., Kumar, S., Jere, A., Anamika, K., 2020. Multi-omics data integration, interpretation, and its application. *Bioinf. Biol. Insights* 14.

- Wang, Y.-W., Tu, P.-H., Lin, K.-T., Lin, S.-C., Ko, J.-Y., Jou, Y.-S., 2011. Identification of oncogenic point mutations and hyperphosphorylation of anaplastic lymphoma kinase in lung cancer. *Neoplasia* 13, 704–724.
- Zhang, Zhang, Zhao, W., Xiao, J., Bao, Y., He, S., Zhang, G., Li, Yixue, Zhao, G., Chen, R., Gao, Y., Zhang, C., Yuan, L., Zhang, G., Xu, S., Zhang, C., Gao, Y., Ning, Z., Lu, Y., Xu, S., Zeng, J., Yuan, N., Zhu, Junwei, Pan, M., Zhang, H., Wang, Q., Shi, S., Jiang, M., Lu, M., Qian, Q., Gao, Q., Shang, Y., Wang, Jinyue, Du, Z., Xiao, J., Tian, D., Wang, P., Tang, B., Li, Cuiping, Teng, Xufei, Liu, X., Zou, D., Song, S., Xiong, Z., Li, Mengwei, Yang, F., Ma, Y., Sang, J., Li, Zhaohua, Li, R., Wang, Zhonghuang, Zhu, Q., Zhu, Junwei, Li, Xin, Zhang, S., Tian, D., Kang, Hailong, Li, Cuiping, Dong, L., Ying, C., Duan, G., Song, S., Li, Menghua, Zhao, W., Zhi, X., Ling, Y., Cao, R., Jiang, Z., Zhou, H., Lv, D., Liu, W., Klenk, H.-P., Zhao, G., Zhang, G., Zhang, Yadong, Zhang, Zhewen, Zhang, H., Xiao, J., Chen, T., Zhang, S., Chen, Xu, Zhu, Junwei, Wang, Zhonghuang, Kang, Hailong, Dong, L., Wang, Y., Ma, Y., Wu, S., Li, Zhaohua, Gong, Z., Chen, Meili, Li, Cuiping, Tian, D., Teng, Xufei, Wang, P., Tang, B., Liu, X., Zou, D., Song, S., Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, Xiyuan, Teng, Xueyi, Sun, X., Sun, L., Chen, R., Zhao, Y., Wang, Jiajia, Zhang, P., Li, Yanyan, Zheng, Y., Chen, R., He, S., Teng, Xueyi, Chen, Xiaomin, Xue, H., Teng, Y., Zhang, P., Kang, Q., Hao, Y., Zhao, Y., Chen, R., He, S., Cao, J., Liu, L., Li, Zhao, Li, Q., Zou, D., Du, Q., Abbasi, A.A., Shireen, H., Pervaiz, N., Batool, F., Raza, R.Z., Ma, L., Niu, G., Zhang, Yuansheng, Zou, D., Zhu, T., Sang, J., Li, Mengwei, Hao, L., Zou, D., Wang, G., Li, Mengwei, Li, R., Li, Mengwei, Li, R., Bao, Y., Yan, J., Sang, J., Zou, D., Li, Chen, Wang, Zhennan, Zhang, Yuansheng, Zhu, T., Song, S., Wang, X., Hao, L., Li, Zhonghai, Zhang, Yang, Zou, D., Zhao, Y., Wang, H., Zhang, Yi, Xia, X., Guo, H., Zhang, Zhang, Zou, D., Ma, L., Dong, L., Tang, B., Zhu, Junwen, Zhou, Q., Wang, Zhonghuang, Kang, Hongen, Chen, Xu, Lan, L., Bao, Y., Zhao, W., Zou, D., Zhu, Junwei, Tang, B., Bao, Y., Lan, L., Zhang, Xin, Ma, Y., Xue, Yongbiao, Sun, Y., Zhai, S., Yu, L., Sun, M., Chen, H., Zhang, Zhang, Zhao, W., Xiao, J., Bao, Y., Hao, L., Hu, H., Guo, A.-Y., Lin, S., Xue, Yu, Wang, C., Xue, Yu, Ning, W., Xue, Yu, Zhang, Xinxin, Xiao, Y., Li, Xia, Tu, Y., Xue, Yu, Wu, W., Ji, P., Zhao, F., Luo, H., Gao, F., Guo, Y., Xue, Yu, Yuan, H., Zhang, Y.E., Zhang, Q., Guo, A., Zhou, J., Xue, Yu, Huang, Z., Cui, Q., Miao, Y.-R., Guo, A.-Y., Ruan, C., Xue, Yu, Yuan, C., Chen, Ming, Jin, J.-P., Tian, F., Gao, G., Shi, Y., Xue, Yu, Yao, L., Xue, Yu, Cui, Q., Li, Xiangshang, Li, C.-Y., Tang, Q., Guo, A.-Y., Peng, D., Xue, Yu, 2020. Database resources of the national genomics data center in 2020. *Nucleic Acids Res.* 48, D24–D33.
- Zhu, H., Ji, Y., Li, W., Wu, M., 2019. Identification of key pathways and genes in colorectal cancer to predict the prognosis based on mRNA interaction network. *Oncol. Lett.* 18, 3778–3786.