MOVIS: A Multi-Omics Software Solution for Multi-modal Time-Series Clustering, Embedding, and Visualizing Tasks

Aleksandar Anžel^a, Dominik Heider^a, Georges Hattab^a

^aDepartment of Mathematics and Computer Science, University of Marburg, Hans-Meerwein-Strasse 6, Marburg, 35032, Hesse, Germany

Keywords: genomics, transcriptomics, metabolomics, proteomics, physico-chemical, omics, multi-omics, time-series, temporal, multimodal, modality, clustering, embedding, visualization

Supplementary material

The supplementary material for this paper consists of various screenshots of MOVIS. Each screenshot depicts one or more essential functionalities or showcases a different part of the tool that the user experiences.

Preprint submitted to Computational and Structural Biotechnology Journal, February 14, 2022

Email address: aleksandar.anzel@uni-marburg.de (Aleksandar Anžel)

MOVIS	×		Μ	OVIS			
MOVIS		Welcome to M time-series o temporal dat	MOVIS - a time-series multi-omics of mics data sets at once so that you a.	visualization tool. MOVIS allows you to explore multiple can easily pinpoint anomalies and/or patterns in your			
Aulti-Omics VISualization tool for ime-series data sets	GitHub	Currently, MOVIS supports genomics, proteomics, metabolomics, transcriptomics, and physico-chemical time- series data sets. You can see the examples of what kind of data sets MOVIS supports on <i>Example 1</i> and <i>Example 2</i> pages on the sidebar. Example 1 consists of genomics, proteomics, metabolomics, and physico-chemical time-series data.					
Avigation: Home Example 1 Example 2 Case study Upload		If you do not kn the sidebar). If y Changeld	ow how to use MOVIS, or want to s rou want to start exploring your da	ee how MOVIS works, check out our documentation (link or Ita, head out to the <i>Upload</i> page.			
		Date	Info				
ug report, <u>report here</u> sature request: propose here ocumentation: <u>see here</u>		03.09.2021.	MOVIS v1.0.0 released 🎘 🏷				

Figure 1: **The homepage.** The default page is a welcome page that briefly describes the tool and provides a changelog.



Figure 2: The upload functionality. Genomics and proteomics data are ready to be uploaded.

	×	Met	agenomics da	ita							
	_	What kind	l of data set do you want to see?								
		Raw F#	ASTA files								•
MOVIS		rmags	_filtered — 2011-04-11.fa — 2011-04-19.fa — 2011-04-19.fa — 2011-04-20.fa — 2011-04-21.fa — 2011-04-25.fa — 2011-04-26.fa								
MOVIS		7	7 files								
Multi-Omics VISualization tool for time-series data sets	GitHub	Calo	culate additional physico-chem al properties First 100 entries.	ical properti	es?						
			DateTime		Hydrogen	oond	Stacking	energy	Solvat	ion	
		0	2011-04-11T00:00:00+02:	00	-6.7835		-27.0692	87	-172.3	423	1
Navigation:		1	2011-04-18T00:00:00+02:	00	-7.3033		-27.1131		-171.6	519	1
		2	2011-04-19T00:00:00+02:	00	-7.4018		-27.1566		-171.6	411	
0		3	2011-04-20T00:00:00+02:	00	-7.0880		-27.1063		-172.0	904	
Home Frame 1		4	2011-04-21T00:00:00+02:	00	-7.5475		-27.1704		-171.3	783	
Example 1		5	2011-04-25T00:00:00+02:	00	-7.4765		-27.1512		-171.4	882	
Case study		6	2011-04-26T00:00:00+02:	00	-7.2997		-27.1522		-171.7	622	
O Upload		7	2011-04-27T00:00:00+02:	00	-7.3444		-27.1609		-171.7	337	
0.		8	2011-05-16T00:00:00+02:	00	-6.7082		-27.0149		-172.3	109	
		9	2011-05-17T00:00:00+02:	00	-7.5394		-27.0844		-171.1	767	
		Insert the	time series interval you want to kee	ep. Use ISO 860	1 format: YYYY-I	/M-DD					0
Bug report: <u>report here</u> Feature request: <u>propose here</u> Documentation: <u>see here</u>		Filte	er data set (remove rows and/o	r columns)							
		Choose yo	our visualization:								
		Choose	e an option								•
		Embedd	ed FASTA files First 50 entries a	ind first 8 fea	ures (column	s).					
		C	DateTime	dim 0	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	
		0 2	011-04-11T00:00:00+02:00	-0.4256	-2.3905	2.7292	1.9618	2.3907	-0.2310	-0.1565	1
		1 7	011 04 10700-00-00-00-00	0.0700	2 0114	1 4440	0.0501	1 35 43	3 4015	0.2010	

Figure 3: *Example 1* data. The screenshot shows additional physico-chemical properties calculated and displayed to the user. The first 50 entries of the embedded FASTA files are also displayed.

×	Met	agenomics data						
	What kind	What kind of data set do you want to see?						
	Raw F4	Raw FASTA files •						
	rmags	_filtered — 2011-04-11.fa — 2011-04-18.fa — 2011-04-20.fa — 2011-04-20.fa — 2011-04-21.fa — 2011-04-25.fa — 2011-04-26.fa						
MOVIS	7	7 files						
Multi-Omics VISualization tool for	Cale	culate additional physico-chemical prop	erties?					
time-series data sets <u>GitHub</u>	Additional properties First 100 entries.							
		DateTime	Hydrogen bond	Stacking energy	Solvation			
	0	2011-04-11T00:00:00+02:00	-6.7835	-27.0692	-172.3423			
Navigation:	1	2011-04-18T00:00:00+02:00	-7.3033	-27.1131	-171.6519			
	2	2011-04-19T00:00:00+02:00	-7.4018	-27.1566	-171.6411			
	3	2011-04-20T00:00:00+02:00	-7.0880	-27.1063	-172.0904			
Home Frample 1	4	2011-04-21T00:00:00+02:00	-7.5475	-27.1704	-171.3783			
Example 2	5	2011-04-25T00:00:00+02:00	-7.4765	-27.1512	-171.4882			
Case study	6	2011-04-26T00:00:00+02:00	-7.2997	-27.1522	-171.7622			
O Upload	7	2011-04-27T00:00:00+02:00	-7.3444	-27.1609	-171.7337			
	8	2011-05-16T00:00:00+02:00	-6.7082	-27.0149	-172.3109			
	9	2011-05-17T00:00:00+02:00	-7.5394	-27.0844	-171.1767			
Bug report: <u>report here</u>	Succe	ssfully kept 2011-04-11, 2011-04-27 time	e frame!					
Feature request: propose here	V Filte	er data set (remove rows and/or column	s)					
Documentation: see here	Select col	umns to remove						
	Choose	e an option			•			
	Succe	ssfully removed 0, 2, 3 rows!						
	Choose yo	our visualization:						
	Choose	e an option			•			

Figure 4: Genomics filtering on *Example 1* data. A filtering step is added to slice the data from 2011-04-11 to 2011-04-27. Additionally, certain row numbers are omitted.



Figure 5: **Embedding for** *Example 1* **data.** The embedding of the resulting genomics data is displayed (50 first entries). A clustering method is applied: OPTICS. A threshold or a minimum number of samples for the method is set using a slider. To evaluate the results in the high dimensional space, the t-SNE visualization is chosen.



Figure 6: Scatter plot matrix for selected features in physico-chemical data set. Four features are selected: *Inflow_pH*, *Outflow_pH*, *Inflow_conductivity*, and *Outflow_conductivity*; totalling 16 possible scatter plot combinations which are displayed as a matrix using the small multiple idiom. A tooltip and a selection are available to investigate certain data points and their characteristics.



Figure 7: Triangular heatmap of all features in the physico-chemical properties data set. A diverging color scheme palette is selected to showcase low and high correlations with less emphasis on the mid-range values. A tooltip is available to investigate specific heatmap cells.

×	Transcriptomics X	0 -						
	Concatenated transcriptomics data sets First 50 entries and first 8 features (columns).							
	Type DateTime thrL thrA thrB thrC yaa3 0 RENZ 1 2010-01-04700-30-00+01-00 447 9175 2886 3244 101	yaaA						
	1 BENZ 1 2019-01-04T12:00:00+01:00 1009 29923 9717 11596 338	424						
	2 BENZ 2 2019-01-04T00:30:00+01:00 444 10189 3333 3837 101	174						
	3 BENZ 2 2019-01-04T12:00:00+01:00 838 13186 4024 4357 98	267						
	4 CONTROL 1 2019-01-04T00:30:00+01:00 361 6778 2137 2425 67	167						
	5 CONTROL 1 2019-01-04T12:00:00+01:00 575 11758 3692 3871 114	161						
	6 CONTROL_2 2019-01-04T00:30:00+01:00 313 8218 2527 2916 77	176						
	7 CONTROL_2 2019-01-04T12:00:00+01:00 460 13114 4227 4834 131	200						
	8 CONTROL_3 2019-01-04T00:30:00+01:00 332 7077 2060 2361 61	140						
	9 CONTROL 3 2019-01-04T12:00:00+01:00 461 11648 3400 4092 95	192						
ie-series data sets <u>GitHub</u>	Filter data set (remove rows and/or columns)							
	Choose your visualization:							
avigation:	Feature through time 🛛 🗙	0 -						
) Home	Feature through time: select features to visualize							
Example 1 Example 2	thrL X thrC X	0 -						
Upload	Encode one nominal feature with color?							
,	Feature through time: select target feature							
	Туре	-						
sug report: <u>report here</u>	thrL through time	Turne						
eature request: <u>propose here</u> Documentation: <u>see here</u>		- BENZ_1 - BENZ_2 - BENZ_3 - CONTROL_1 - CONTROL_2 - CONTROL_3						
		PHE_1 PHE_2 PHE_3 POV_1						
	200-	- POV_2 - POV_3						
	0 F100 01AM 02AM 03AM 04AM 05AM 05AM 07AM 08AM 09AM 10AM 11AM 12PM DateTime							
	Download visualization							

Figure 8: **Example 2 transcriptomics data.** A time series of selected features is displayed as a line plot. Two features are selected: thrL and thrC. A categorical or nominal encoding is carried out for each of the 11 reported types. The *Tableau20* color palette is used.