

SUPPLEMENTARY MATERIAL

MatrixDB 2024: an increased coverage of extracellular matrix interactions, a new Network Explorer and a new web interface

Kasun Wijesiriwardana Samarasinghe¹, Max Kotlyar², Sylvain D. Vallet³, Catherine Hayes¹, Alexandra Naba⁴, Igor Jurisica², Frédérique Lisacek^{1*}, and Sylvie Ricard-Blum^{5*}

SUPPLEMENTARY METHODS

Web portal

The Web portal application is mainly composed of a backend system and a single page frontend web application. The backend system interacts with the database and exposes the data in a meaningful manner such that the web interface can organize and visualize data. It consists of several sub-components as illustrated in Supplementary Figure S1. It exposes a REST API, as the standard interface for the frontend web application or to any other application, which intends to access MatrixDB data. It is implemented in Python v.3.11, along with Flask 3.0 API (<https://flask.palletsprojects.com/en/latest/api/>) as the API layer implementation.

The Web portal application is developed as a Single Page Application (SPA), which interacts with the MatrixDB API and retrieves the required data to produce the relevant pages. It is implemented with the React framework v.18.2 (<https://react.dev/>), in typescript. It uses the material UI, a standard reusable web component library to implement the pages and visual arrangement. The modular structure of the API, combined with a caching layer, optimizes resource usage and enables horizontal scaling, ensuring the system can efficiently support increasing demand over time.

High-confidence predicted interactions (PPIs)

Prediction studies defined their high confidence PPIs as follows:

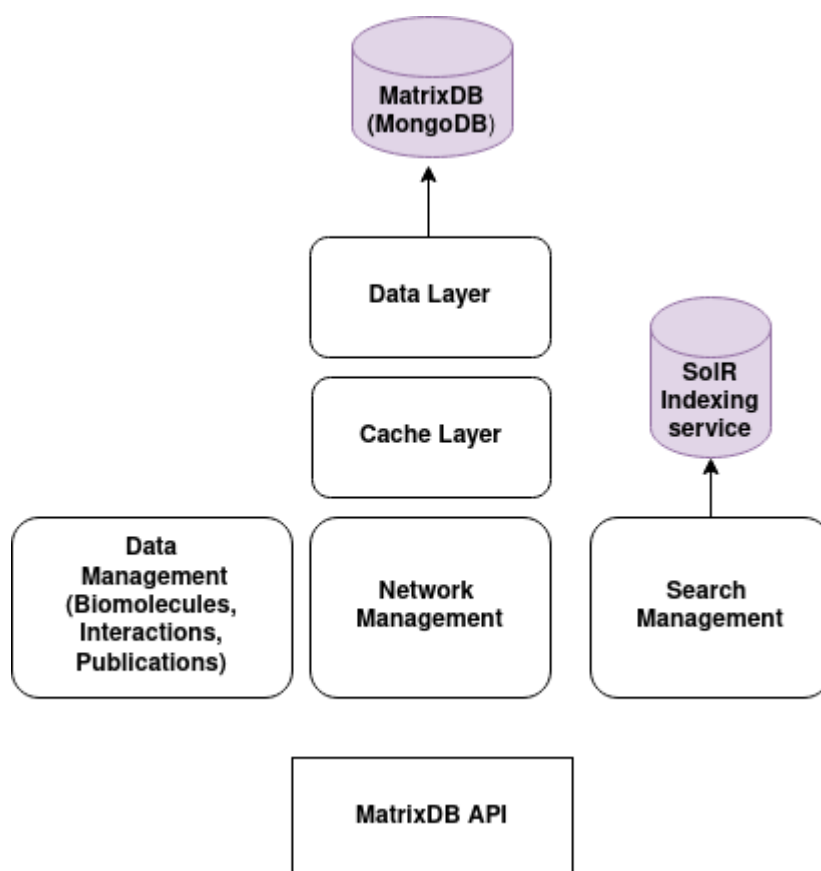
- Rhodes *et al.* (*Nat. Biotechnol.* 2005 23: 951-959) used a likelihood ratio cut-off of 381
- Lefebvre *et al.* (*In: Systems Biology and Computational Proteomics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 42–56*) used a probability greater than 0.5
- Elefsinioti *et al.* (*Mol. Cell. Proteomics* 2011 10 : M111.010629) used probabilities greater than 0.7
- Zhang *et al.* (*Nature* 2012 490 : 556-560) used likelihood ratio of at least 600
- Kotlyar *et al.* (*Nat. Methods* 2015 12: 79-84) used false discovery rate < 0.6

Most studies, including us, interpreted "confidence" by calculating a Bayesian probability of interaction: $P(\text{interaction} | \text{evidence}) \sim \text{likelihood} * \text{prior} \sim P(\text{evidence} | \text{interaction}) * P(\text{interaction})$. And $P(\text{interaction}) = |\text{estimated size of human interactome}| / |\text{number of human protein pairs}|$. For example, Zhang *et al.* assumed

that there are ~650K human PPIs, resulting in a prior of about 1/600, and then their high confidence threshold of likelihood=600 corresponded to $P(\text{interaction}|\text{evidence}) = 0.5$. While this may seem low, it is an improvement in probability from 1/600 to 1/2. Also, 650K PPIs is an under-estimation, so the real $P(\text{interaction}|\text{evidence})$ may be much higher than 0.5.

Ranking search results by relevance

Search results are collected and ranked using the indexing system of Solr so that free text search is performed based on a weighed scoring mechanism that can be defined in this software. Those pre-defined weights reflect biomolecule attributes resulting in a search score used to rank the results. Solr searching capabilities span fielded search, Boolean queries, phrase queries, fuzzy queries, spell check, wildcards, grouping, auto-complete and others across different types of data. The index of a search term is therefore linked with other indices and a hierarchy was defined to keep biomolecule names at the top and associated annotations below.

SUPPLEMENTARY FIGURES

Supplementary Figure S1: System architecture. There are three main management components (data, network and search), which handle different aspects of the Web portal.

The screenshot displays the MatrixDB search interface. At the top, there are tabs for 'BASIC SEARCH' and 'ADVANCED SEARCH'. A search bar contains the text 'name:collagen'. Below the search bar, there are two columns of related information:

- Biomolecule properties:**
 - id:GAG_1 name:Fibronectin
 - Gene name: [gene:LOX](#)
 - GO Terms, UniProt Keywords: [go:cytosol keywords:Metal-binding](#)
 - Reactome: [reactome:ECM proteoglycans](#)
- MatrixDB metadata:**
 - Matrixome division: [matrixome_division:matrixome-associated](#)
 - Matrixome category: [matrixome_category:ecm regulators](#)
 - GO term AND Species: [go:cytosol and species:homo sapiens](#)
 - Matrixome category AND Species: [matrixome_category:ecm regulators and species:homo sapiens](#)

Below this information is a table titled 'Biomolecules'. The table has two columns of results. A red box highlights three icons in the top right corner of the table: a magnifying glass (award icon), an eye (display all icon), and a dropdown arrow (filter icon).

Collagen-VI	20	Collagen-I	9
MULT_8_human CPX-1736		MULT_3_bovine CPX-3101	
Collagen-I	18	Collagen-V heterotrimer 2 different chains	8
MULT_3_human CPX-1650		MULT_5_VAR1_human CPX-1727	
Collagen-IV v1	16	Collagen-XI v1	7
MULT_4_VAR1_human CPX-1723		MULT_7_VAR1_bovine CPX-3108	
Collagen-III	14	Collagen-II	7
MULT_10_human CPX-1714		MULT_9_bovine CPX-3105	
Collagen-I	11	Collagen-IV v1	6
MULT_3_rat CPX-3104		MULT_4_VAR1_mouse CPX-2959	

Supplementary Figure S2: Querying MatrixDB database using basic and advanced search. The icons circled in red allow users to access the most relevant results (award icon), to display all the result pages (eye icon), and to select a biomolecule category and/or a species (filter icon).

SUPPLEMENTARY TABLES

Name of libraries	Use in MatrixDB	Reference
Cytoscape.js , a graph theory library for visualization and analysis	Network visualization with most of the standard Cytoscape functionalities and compatible formats for export and reuse	<i>Franz et al., 2016 Bioinformatics 32: 309-11</i>
Mol* , a 3D structure viewer	Visualization of 3D structures from the Protein Data Bank in Europe in Biomolecule pages to replace outdated LiteMol (<i>Sehnal et al., 2017, Nat. Methods 14:1121-1122</i>)	<i>Sehnal et al., 2021 Nucleic Acids Res. 49(W1):W431-W437</i>
D3 (or D3.js), JavaScript library for visualizing data	Graphs and charts for statistics display and other visualization-related components	https://d3js.org/

Supplementary Table S1: Third-party JavaScript libraries used to implement visual components in MatrixDB.