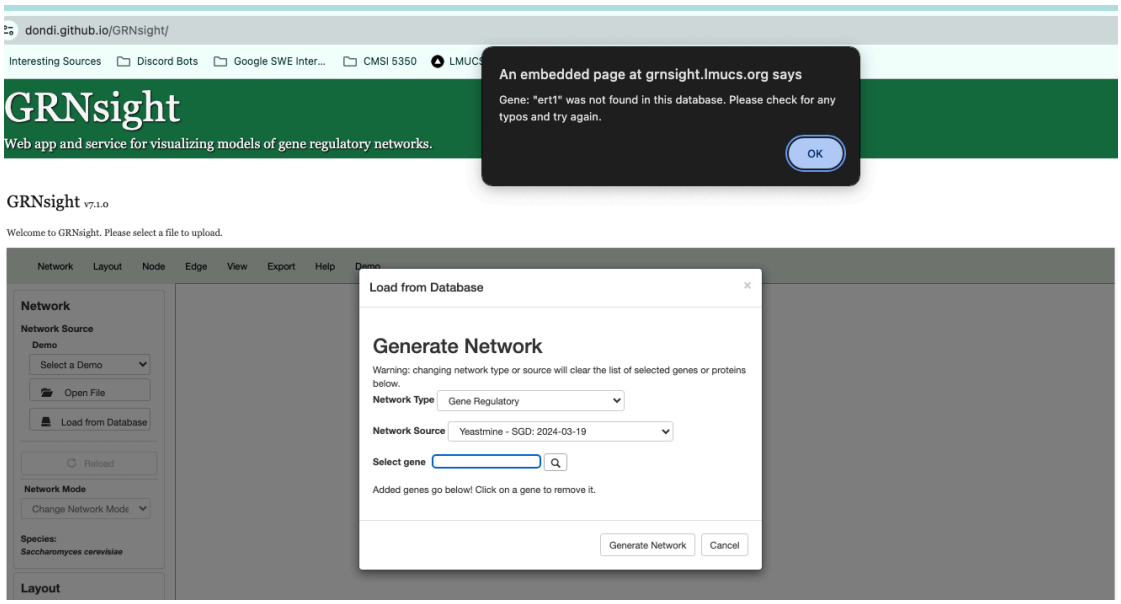
GRNsight Database

Background

GRNsight includes four primary databases: the expression database, grnsettings database, network database, and protein-protein database. This document will focus on the network database and the protein-protein database. Both of these databases contain a gene table, a source table, and a network table (referred to as physical interactions in the protein-protein database). Additionally, the protein-protein database features a protein table. All scripts for downloading and uploading data to these databases are located in the database folder.

Problems



ERT1 is not in the network database. After investigating the queries, I discovered that our original approach involved first querying the list of regulators, followed by querying the targets for each regulator. However, ERT1 was not identified as a regulator in this process. See here for the list of regulators.

In contrast, when I queried the list of regulations from AllianceMine, I found approximately 15 connections where ERT1 is recognized as a regulator. You can view the relevant data in Box.

Comparison of Databases: I conducted a comparison between the server network database and the AllianceMine network:

	Server Database	AllianceMine

Missing network	2961	17
Missing regulator genes	270	1
Missing target genes	44	6

For further details on the missing networks or genes, please refer to the provided link.

Migration Note: Also, YeastMine is no longer supported, and we need to transition to AllianceMine. The service link in the sample code is currently non-functional.

Proposed Solutions

ERT1 issue

Solution 1: Get the network by querying the list of regulations in one request.

As we can see from the AllianceMine network, we can see more regulatory networks from our current query. However, we don't know whether this is because of the query or because new data has been updated since Yeastmine is down.

Pros	Cons
✅ Query all regulatory networks in one query. ✅ Capture more networks than querying all the regulators and find the targets for each query. ✅ Don't need to perform multiple manual queries if you can't request automatically by the service.	❌ Possibly missing some regulators

Solution 2: Combine the network database from the list of regulations query and the list of regulators and their corresponding target queries

To address the issue in solution 1, we can first query the list of regulations query. Then we query the list of regulators. If we don't see any gene in the regulators list in regulations, then we would query the target for each missing regulator.

Pros	Cons
✅ Capture all the regulatory networks	❌ Require a lot of queries if there are a lot of missing regulators (even worse if we can't automate the task since AllianceMine server is not working).

AllianceMine

Solution 1: Manually download data from AllianceMine, then process the file

Download the required data file using the AllianceMine query builder. Write a Python script to process the downloaded file, extracting relevant details such as genes and regulators (would be the same for protein-protein-interactions). Depending on the approach taken to resolve the ERT1 issue, we would either develop and process the data from the sheet or reuse the existing code.

Pros	Cons
✅ Simple, don't need to worry about web scraping ✅ Know the exact structure of the downloaded file ✅ Changes to the AllianceMine web interface or the service API won't affect the process (as long as the file download format remains the same)	❌ Required a manual download process, which might have errors and time-consuming.

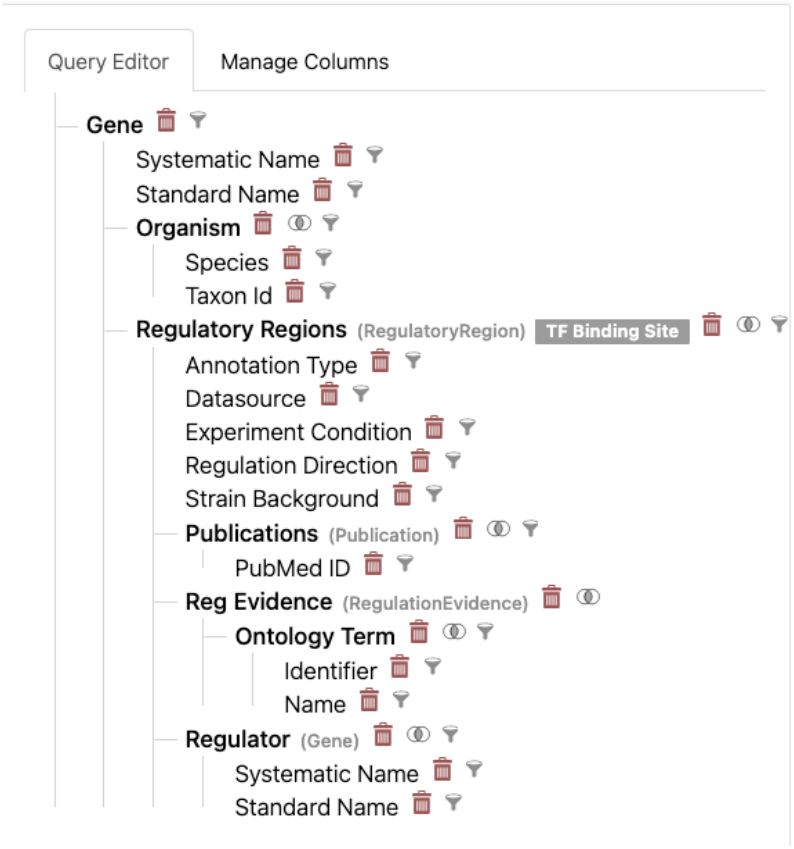
Solution 2: Automating Data Retrieval with Web Scraping

I'm not familiar with this process, and I require more investigation about this approach as how we can download the whole table.

Pros	Cons
✅ Prevent errors from manual work.	❌ Can break when there are changes to the website structure ❌ Violate the terms of service of the website (need to double check AllianceMIne policies) ❌ Requires frequent monitoring and maintenance to ensure the scraper works as expected.

Appendix

1. GRN Network AllianceMine Query Builder



2. Regulators Query