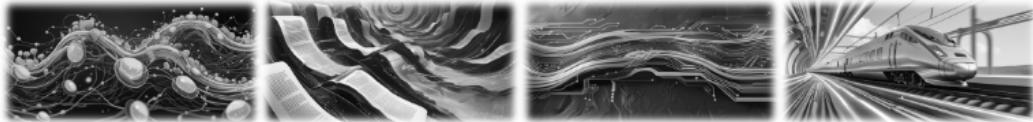


Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Hello! I'm **cellAtria**, your assistant for analyzing single-cell RNA-seq (scRNA-seq) datasets.

Here's what I can help you with:

1. Extract structured metadata from scientific articles (PDF or URL).
2. Store and organize metadata in structured project directories.
3. Access public databases (currently support GEO) and fetch associated sample metadata.
4. Download and organize scRNA-seq datasets.
5. Trigger CellExpress standardized single-cell data processing.

To see a list of all available actions, type `'help'`.

Let's see how far we can fly together.

To get started, let's set up your working directory. I can show you the current one or help create a new path. You can change your working directory anytime you wish.

Current Working Directory: /mnt/work/projects/cellatria/agent

Your Prompt b

Give me a task...

.pdf c

↑
 Drop File Here
 - or -
 Click to Upload

Submit Prompt/PDF

Live Logs d

Agent Backend e ▶

Terminal Panel f ▶

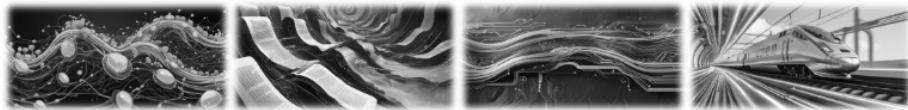
File Browser g ▶

Export Chat / Model h ▶

Figure 1. User interface of the CellAtria agentic system. **a)** Persistent chatbot window that mediates user-agent interaction. **b)** Input panel for free-text prompts. **c)** PDF upload widget enabling ingestion of documents and submit button that triggers the unified execution handler, dispatching tasks based on input context. **d)** Live log viewer providing real-time telemetry on agent execution status. **e)** The agent backend panel displays a live, step-by-step stream of the agent's internal reasoning, tool calls, and backend responses. **f)** Embedded terminal interface for issuing UNIX-style shell commands within the agent context. **g)** Interactive file browser for navigating directory structures, inspecting files. **h)** Export utility providing two downloads: a machine-readable user-agent conversation transcript and a structured metadata file that records the backend language model configuration. Note: the system-level panels (backend, terminal, browser, and export) are designed as collapsible accordion sections.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Name the list of tools you have been equipped with in compact view. Add a short description for each. Keep the in line spaces to minimum.

Here's a compact list of tools with short descriptions:

Metadata Analysis

1. `fetch_article_metadata_url`: Extract metadata from a scientific article URL.
2. `fetch_article_metadata_pdf`: Extract metadata from uploaded PDFs.
3. `store_article_metadata_file`: Save article metadata as JSON or TXT.
4. `refine_article_metadata`: Register article metadata for review.
5. `fetch_geo_metadata`: Fetch GEO dataset metadata.
6. `store_geo_metadata_file`: Save GEO metadata as JSON or CSV.
7. `refine_geo_metadata`: Register GEO metadata for review.

Data Downloads & Organization

8. `download_geo`: Download GEO datasets including supplementary files.
9. `download_gsm`: Download GSM-level supplementary files.
10. `download_file`: Download files from direct URLs.
11. `make_directory`: Create directories to organize files.
12. `list_directory`: Show directory contents in a tree view.
13. `inspect_file`: Preview CSV, JSON, or TXT files.
14. `get_file_size`: Check the size of a file in MB.
15. `move_file`: Move files to a new directory.
16. `rename_file`: Rename files or move them with new names.
17. `remove_file_or_dir`: Delete files or directories.
18. `set_working_directory`: Set or change the working directory.
19. `get_working_directory`: Display the current working directory.

File Preparation

20. `fix_10x_file_format`: Standardize file naming for 10X Genomics data.
21. `create_custom_csv`: Create metadata as CSV from tabular data.
22. `create_custom_json`: Save structured data as JSON metadata.
23. `create_custom_txt`: Generate plain text metadata files.

CellExpress Pipeline

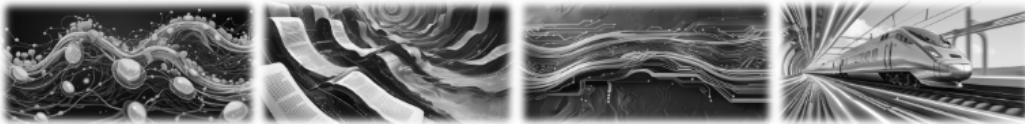
24. `get_cellexpress_info`: View details of the CellExpress pipeline.
25. `configure_cellexpress`: Set parameters for the pipeline.
26. `preview_cellexpress_config`: Preview and validate pipeline settings.
27. `reset_cellexpress_config`: Clear all pipeline configurations.
28. `validate_cellexpress_config`: Check if pipeline settings are valid.
29. `run_cellexpress`: Launch the CellExpress pipeline.
30. `review_cellexpress_log`: Read logs from an active pipeline session.
31. `terminate_cellexpress_job`: Stop a running pipeline job.
32. `check_cellexpress_status`: Check if the pipeline job is running or completed.

Let me know which tool you'd like to explore further!

Figure 2. Overview of the CellAtria toolchain in response to a prompt querying its functional capabilities. The agent enumerates its internal toolkit, which includes modules for article metadata extraction, dataset access, file system management, metadata structuring, and CellExpress configuration. Each tool is exposed as a callable function with defined scope and I/O specifications. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



👋 Hello! I'm **cellAtria**, your assistant for analyzing single-cell RNA-seq (scRNA-seq) datasets.

Here's what I can help you with:

1. Extract structured metadata from scientific articles (PDF or URL).
2. Store and organize metadata in structured project directories.
3. Access public databases (currently support GEO) and fetch associated sample metadata.
4. Download and organize scRNA-seq datasets.
5. Trigger CellExpress standardized single-cell data processing.

To see a list of all available actions, type '`help`'.

Let's see how far we can fly together. 🦅

To get started, let's set up your working directory. I can show you the current one or help create a new path. You can change your working directory anytime you wish.

📁 Current Working Directory: `/mnt/work/projects/cellatria/agent`

Set my working directory to: `/mnt/work/projects/cellatria/data/my_projects/infants_vaccine`

Your working directory has been successfully set to:

`/mnt/work/projects/cellatria/data/my_projects/infants_vaccine`.

What would you like to do next? Here are some options to consider:

1. Upload a scientific article (PDF or URL) for metadata extraction.
2. Provide a GEO accession ID to fetch metadata and download datasets.
3. Set up and configure a CellExpress pipeline job.
4. Organize your project files in the working directory.

Let me know how you'd like to proceed!

Your Prompt

Give me a task...

📁 .pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

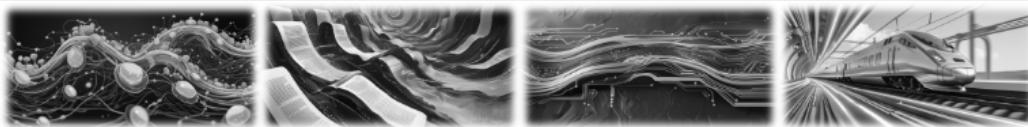
Live Logs

```
● New interaction started.  
✖ User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_pro  
jects/infants_vaccine  
🕒 Invoking agent...  
✓ Agent response received.  
🕒 Interaction complete.  
---
```

Figure 3. Interactive initialization of a working directory within the CellAtria agentic interface. In response to a user prompt, CellAtria updates the working directory and offers context-aware follow-up actions. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Extract metadata from the article: <https://www.nature.com/articles/s41467-023-43758-2>

The metadata extracted from the article has been structured as follows:

Field	Value
Project	Human blood Not disease-specific
Title	Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following initial immunizations
Authors	Nima Nouri; Raquel Giacomelli Cao; Eleonora Bunsow; Djamel Nehar-Belaid; Radu Marches; Zhaohui Xu; Bennett Smith; Santtu Heinonen; Sara Mertz; Amy Leber; Gaby Smits; Fiona van der Klis; Asunción Mejías; Jacques Banchereau; Virginia Pascual; Octavio Ramilo
Tissue	blood
Species	Human
Disease	Not disease-specific
Data Modality	scRNA-seq
Data Availability	GEO:GSE204716 and dbGAP:phs002926.v1.p1
Publication Date	02 December 2023
Publisher	Nature Communications
Conflict	

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```

New interaction started.
User input: Set my working directory to:
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine
Invoking agent...
Agent response received.
Interaction complete.

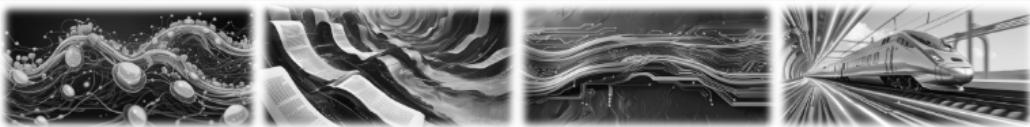
New interaction started.
User input: Extract metadata from the article:
https://www.nature.com/articles/s41467-023-43758-2
Invoking agent...
Agent response received.
Interaction complete.

```

Figure 4. Metadata extraction from a DOI-based URL using the CellAtria agentic interface. In response to a user-provided article URL, the agent retrieves and parses the full-text document, converting key study descriptors into a structured tabular format. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Disease	Not disease-specific
Data Modality	scRNA-seq
Data Availability	GEO:GSE204716 and dbGAP:phs002926.v1.p1
Publication Date	02 December 2023
Publisher	Nature Communications
Conflicts of Interest	J.B. served on boards and advisory roles linked to pharmaceutical companies. Several authors declared competing interests related to grants and stocks.

Would you like to store this metadata? If yes, please specify the format (JSON or TXT) and any filename preference.

Store the extracted metadata in JSON format in the working directory.

The extracted metadata has been successfully saved as a JSON file at:

/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/metadata_81b651af.json.

Would you like to:

1. Inspect the saved file?
2. Move on to fetch GEO dataset metadata (e.g., GSE204716)?
3. Begin processing this project with CellExpress?

Let me know how you'd like to proceed!

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

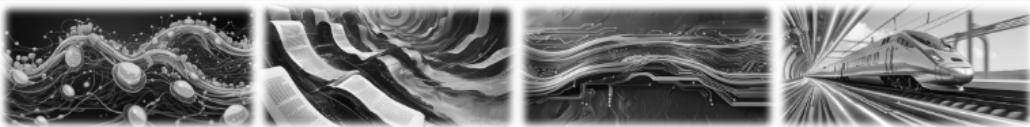
```

Agent response received.
Interaction complete.
-----
New interaction started.
User input: Extract metadata from the
article:
https://www.nature.com/articles/s41467-
023-43758-2
Invoking agent...
Agent response received.
Interaction complete.
-----
New interaction started.
User input: Store the extracted
metadata in JSON format in the working
directory.
Invoking agent...
Agent response received.
Interaction complete.
-----
```

Figure 5. Structured metadata export to file following article parsing in CellAtria. Upon user request, the metadata is saved as a JSON file in the active working directory. The interface confirms the file path and proposes context-aware next steps. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Conflicts of Interest

J.B. served on boards and advisory roles linked to pharmaceutical companies. Several authors declared competing interests related to grants and stocks.

Would you like to store this metadata? If yes, please specify the format (JSON or TXT) and any filename preference.

Store the extracted metadata in JSON format in the working directory.

The extracted metadata has been successfully saved as a JSON file at:

```
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/metadata_81b651af.json
```

Would you like to:

1. Inspect the saved file?
2. Move on to fetch GEO dataset metadata (e.g., GSE204716)?
3. Begin processing this project with CellExpress?

Let me know how you'd like to proceed!

Rename the generated file metadata_81b651af.json to article_metadata.json

The file has been successfully renamed to:

```
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/article_metadata.json
```

What would you like to do next? Here are some suggestions:

1. Inspect the renamed file for content verification.
2. Fetch GEO dataset metadata (e.g., GSE204716) for this project.
3. Start configuring a CellExpress pipeline.

Let me know how you'd like to proceed!

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

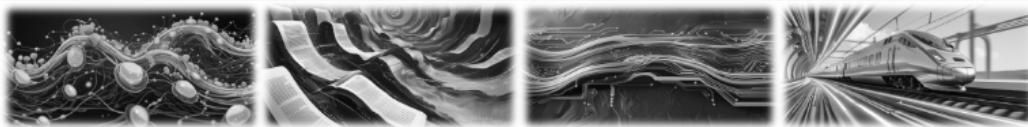
Live Logs

```
Invoking agent...
✓ Agent response received.
● Interaction complete.
—
● New interaction started.
● User input: Store the extracted
metadata in JSON format in the working
directory.
● Invoking agent...
✓ Agent response received.
● Interaction complete.
—
● New interaction started.
● User input: Rename the generated file
metadata_81b651af.json to
article_metadata.json
● Invoking agent...
✓ Agent response received.
● Interaction complete.
—
```

Figure 6. File management automation for structured metadata workflows. CellAtria autonomously renames the output file based on a user-specified name within the active working directory. The task includes confirmation prompts to validate the action. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Show the contents of article_metadata.json

The contents of the file article_metadata.json are as follows:

```
{  
  "Project": "Human blood Not disease-specific",  
  "Title": "Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following SARS-CoV-2 infection",  
  "Authors": "Nima Nouri; Raquel Giacomelli Cao; Eleonora Bunsow; Djamel Nehar-Belaid; Radu Marches; Zhaohui Xu; Bennett Smith; Sarit Sircar",  
  "Tissue": "blood",  
  "Species": "Human",  
  "Disease": "Not disease-specific",  
  "Data_Modality": "scRNA-seq",  
  "Data_Availability": "GEO:GSE204716 and dbGAP:phs002926.v1.p1",  
  "Publication_Date": "02 December 2023",  
  "Publisher": "Nature Communications",  
  "Conflicts_of_Interest": "J.B. served on boards and advisory roles linked to pharmaceutical companies. Several authors declared competing financial interests. See the article for further details."  
}
```

Let me know if you need any further actions, such as modifying, appending new data, or proceeding with dataset analysis!

List the current contents of my working directory.

The current contents of your working directory (`/mnt/work/projects/cellatria/data/my_projects/infants_vaccine`) are:

```
└── article_metadata.json
```

What would you like to do next?

1. Inspect or modify article_metadata.json.
2. Fetch GEO dataset metadata or download related datasets (e.g., GSE204716).
3. Configure and start a CellExpress pipeline for analysis.

Let me know how you'd like to proceed!

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

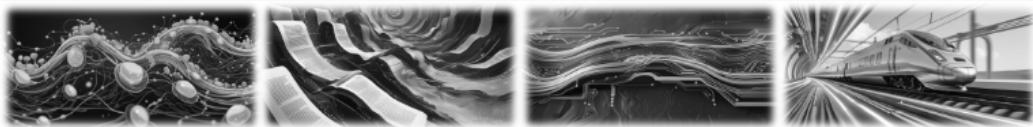
Live Logs

```
Received article_metadata.json to  
article_metadata.json  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—  
🕒 New interaction started.  
🕒 User input: Show the contents of  
article_metadata.json  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—  
🕒 New interaction started.  
🕒 User input: List the current contents  
of my working directory.  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—
```

Figure 7. Structured file inspection and metadata verification. Upon user request, CellAtria displays the contents of the stored metadata file and inspects the current working directory. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Fetch GEO metadata for accession GSE204716

The GEO metadata for accession **GSE204716** has been extracted and structured as follows:

Field	Value
Title	Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following initial immunizations
Organism	Homo sapiens
GEO Link	GSE204716

Samples Details:

Sample	Description	BioSample ID	SRA IDs
GSM6189249	Infant 1 - Before vaccination	N/A	Unavailable
GSM6189250	Infant 1 - 1 week after vaccination	N/A	Unavailable
GSM6189251	Infant 1 - 2 months after vaccination	N/A	Unavailable
GSM6189252	Infant 2 - Before vaccination	N/A	Unavailable
GSM6189253	Infant 2 - 1 week after vaccination	N/A	Unavailable
GSM6189254	Infant 2 - 2 months after vaccination	N/A	Unavailable
GSM6189255	Infant 3 - Before vaccination	N/A	Unavailable
GSM6189256	Infant 3 - 1 week after vaccination	N/A ↓	Unavailable

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
user input: show the contents of article_metadata.json
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.

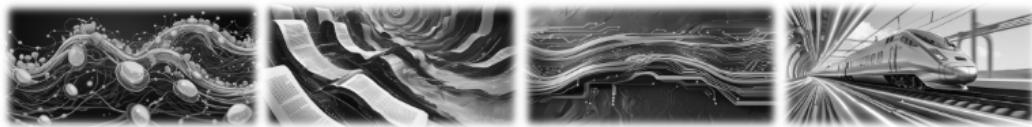
🕒 New interaction started.
🕒 User input: List the current contents of my working directory.
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.

🕒 New interaction started.
🕒 User input: Fetch GEO metadata for accession GSE204716
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.
```

Figure 8. Agent-driven GEO metadata extraction and sample summary. CellAtria retrieves and parses GEO study metadata, displaying structured sample-level information including descriptions and accession statuses. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



GSM6189265	Infant 6 - 1 week after vaccination	N/A	Unavailable
GSM6189266	Infant 6 - 2 months after vaccination	N/A	Unavailable

Would you like to store this metadata? If yes, specify the format (JSON or CSV) and any filename preference.

Store the metadata as a CSV file in my working directory.

The GEO metadata has been successfully saved as a CSV file at:

/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/metadata_b2642794.csv.

Would you like to:

1. Inspect the saved file?
2. Proceed to download the GEO dataset files?
3. Start working with CellExpress analysis setup?

Let me know how you'd like to proceed!

Rename the generated file mmetadata_b2642794.csv to metadata.csv

The file has been successfully renamed to:

/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/metadata.csv.

Would you like to:

1. Inspect the renamed file for verification.
2. Proceed to download the GEO dataset.
3. Begin setting up a CellExpress pipeline.

Let me know what you'd like to do next!

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```

User input: Given task accepted for
accession GSE204716
Invoking agent...
Agent response received.
Interaction complete.

New interaction started.
User input: Store the metadata as a
CSV file in my working directory.
Invoking agent...
Agent response received.
Interaction complete.

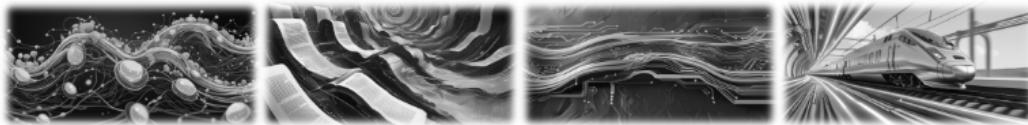
New interaction started.
User input: Rename the generated file
mmetadata_b2642794.csv to metadata.csv
Invoking agent...
Agent response received.
Interaction complete.

```

Figure 9. Structured storage and renaming of extracted GEO metadata. CellAtria stores the parsed metadata as a CSV file in the working directory and facilitates user-guided file renaming. The interface confirms path updates and suggests next actions. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Download dataset GSE204716 into the working directory.

The dataset **GSE204716** has been successfully downloaded into your working directory:

```
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine.
```

Would you like to:

1. Explore and verify the contents of the downloaded data.
2. Organize or configure the dataset for analysis.
3. Set up a CellExpress pipeline for processing.

Let me know how you'd like to proceed!

List the current contents of my working directory.

The current contents of your working directory (`/mnt/work/projects/cellatria/data/my_projects/infants_vaccine`) are:

```
└── GSE204716_family.soft.gz
    ├── article_metadata.json
    ├── metadata.csv
    └── GSM6189249
        ├── GSM6189249_JB19050_barcodes.tsv.gz
        ├── GSM6189249_JB19050_features.tsv.gz
        └── GSM6189249_JB19050_matrix.mtx.gz
    └── GSM6189250
        ├── GSM6189250_JB19051_barcodes.tsv.gz
        ├── GSM6189250_JB19051_features.tsv.gz
        └── GSM6189250_JB19051_matrix.mtx.gz
    └── GSM6189251
        ├── GSM6189251_JB19052_barcodes.tsv.gz
        ├── GSM6189251_JB19052_features.tsv.gz
        └── GSM6189251_JB19052_matrix.mtx.gz
    └── GSM6189252
        ├── GSM6189252_JB19053_barcodes.tsv.gz
        ├── GSM6189252_JB19053_features.tsv.gz
        └── GSM6189252_JB19053_matrix.mtx.gz
    └── GSM6189253
```

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

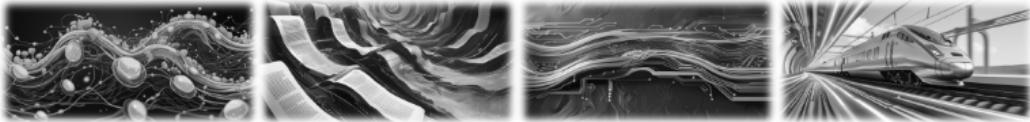
Live Logs

```
🕒 User input: Rename the generated file nmetadata.b2642794.csv to metadata.csv
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.
—
🕒 New interaction started.
🕒 User input: Download dataset GSE204716 into the working directory.
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.
—
🕒 New interaction started.
🕒 User input: List the current contents of my working directory.
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.
—
```

Figure 10. Automated dataset retrieval. Following a user request, CellAtria downloads all supplementary files for the specified GSE dataset and stores them in the working directory. The agent then provides a complete file listing upon user request. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Standardize the 10X Genomics file names in the working directory.

The 10X Genomics file names have been standardized successfully in the directory:

```
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine.
```

Would you like to:

1. Inspect or verify the contents of any specific directory.
2. Begin setting up a CellExpress pipeline for analysis.
3. Perform any additional organization or file operations.

Let me know how you'd like to proceed!

List the current contents of my working directory.

The current contents of your working directory (`/mnt/work/projects/cellatria/data/my_projects/infants_vaccine`) are as follows:

```
└── GSE204716_family.soft.gz
    ├── article_metadata.json
    ├── metadata.csv
    └── GSM6189249
        ├── barcodes.tsv.gz
        ├── features.tsv.gz
        └── matrix.mtx.gz
    └── GSM6189250
        ├── barcodes.tsv.gz
        ├── features.tsv.gz
        └── matrix.mtx.gz
    └── GSM6189251
        ├── barcodes.tsv.gz
        ├── features.tsv.gz
        └── matrix.mtx.gz
    └── GSM6189252
        ├── barcodes.tsv.gz
        ├── features.tsv.gz
        └── matrix.mtx.gz
    └── GSM6189253
```

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
of my working directory.
🕒 Invoking agent...
✓ Agent response received.
🕒 Interaction complete.

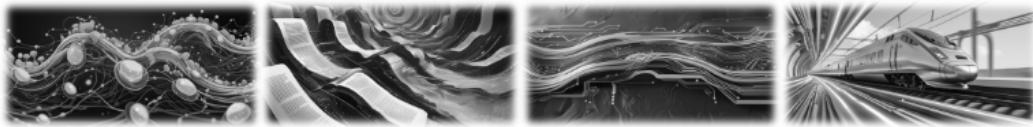
—
🕒 New interaction started.
🕒 User input: Standardize the 10X Genomics file names in the working directory.
🕒 Invoking agent...
✓ Agent response received.
🕒 Interaction complete.

—
🕒 New interaction started.
🕒 User input: List the current contents of my working directory.
🕒 Invoking agent...
✓ Agent response received.
🕒 Interaction complete.
```

Figure 11. Standardized file renaming for downstream compatibility. CellAtria programmatically renames all files to conform with 10X Genomics conventions, ensuring compatibility with downstream pipelines. A directory listing is presented upon request. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



👋 Hello! I'm **cellAtria**, your assistant for analyzing single-cell RNA-seq (scRNA-seq) datasets.

Here's what I can help you with:

1. Extract structured metadata from scientific articles (PDF or URL).
2. Store and organize metadata in structured project directories.
3. Access public databases (currently support GEO) and fetch associated sample metadata.
4. Download and organize scRNA-seq datasets.
5. Trigger CellExpress standardized single-cell data processing.

To see a list of all available actions, type `'help'`.

Let's see how far we can fly together. 🦅

To get started, let's set up your working directory. I can show you the current one or help create a new path. You can change your working directory anytime you wish.

📁 Current Working Directory: `/mnt/work/projects/cellatria/agent`

Set my working directory at `/mnt/work/projects/cellatria/data/my_projects/tumor_specific_T_cells`

Your working directory has been successfully set to:

`/mnt/work/projects/cellatria/data/my_projects/tumor_specific_T_cells`

How would you like to proceed?

1. Extract metadata from a scientific article (PDF or URL).
2. Retrieve GEO dataset metadata or download files.
3. Organize files and folders.
4. Explore analysis pipelines using **CellExpress**.

Simply let me know what you'd like to do next!

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

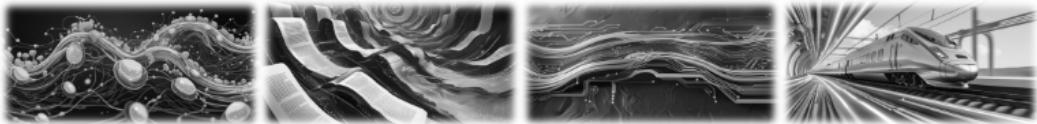
Live Logs

```
🕒 New interaction started.  
👤 User input: Set my working directory  
at  
/mnt/work/projects/cellatria/data/my_pro  
jects/tumor_specific_T_cells  
🕒 Invoking agent...  
✅ Agent response received.  
⭕ Interaction complete.  
---
```

Figure 12. Interactive initialization of a working directory within the CellAtria agentic interface. In response to a user prompt, CellAtria updates the working directory and offers context-aware follow-up actions. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Extract metadata from the article: [https://www.cell.com/cancer-cell/fulltext/S1535-6108\(23\)00082-X](https://www.cell.com/cancer-cell/fulltext/S1535-6108(23)00082-X)

I couldn't extract metadata directly from the provided URL because the site restricts automated access. Could you please upload the PDF version of the article instead? That way, I can proceed with the metadata extraction.

PDF is attached.

Received PDF: /tmp/gradio/bacce16c51e0c7fdcadad5e8b4d09c8b3212a76e76d2790bf69abc0614d1fe9b/PIIS153561082300082X.pdf.
I can extract metadata from it!

Extract metadata

The metadata structure from your provided PDF has been processed successfully. Below is the extracted information:

Field	Extracted Data
Project	Human Non-small Cell Lung Cancer (NSCLC) Tumor-specific T cells
Title	Lineage tracing reveals clonal progenitors and long-term persistence of tumor-specific T cells during immune checkpoint blockade
Publication Date	April 10, 2023
Publisher	Elsevier (Cancer Cell)
Authors	Joy A. Pai, Matthew D. Hellmann, Jennifer L. Sauter, Charles M. Rudin, Andrew Chow, Ansuman T. Satpathy, et al.

↓

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
NEW INTERACTION STARTED.
User input: Extract metadata from the
article: https://www.cell.com/cancer-
cell/fulltext/S1535-6108(23)00082-X
Invoking agent...
Agent response received.
Interaction complete.

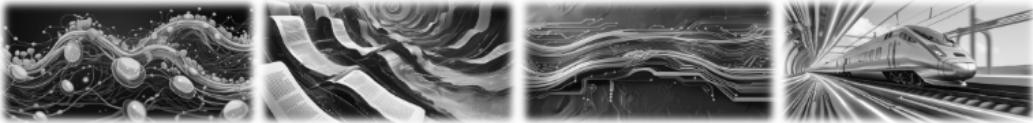
NEW INTERACTION STARTED.
User input: PDF is attached.
Invoking agent...
Agent response received.
Interaction complete.

NEW INTERACTION STARTED.
User input: Extract metadata
Invoking agent...
Agent response received.
Interaction complete.
```

Figure 13. PDF-driven metadata extraction workflow in CellAtria. When automated access to an article URL is restricted, the user uploads a locally stored PDF. The agent parses the document and extracts structured metadata which is then presented in a tabulated format. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Fetch GEO metadata for accession GSE185206

Here is the extracted metadata for GEO accession GSE185206:

General Information:

- **Title:** Systematic lineage tracing reveals clonal progenitors and long-term persistence of tumor-specific T cells during immune checkpoint blockade
- **Organism:** Homo sapiens (Human)
- **GEO Link:** [GSE185206](#)

Samples Summary:

GSM ID	Description	BioSample ID	SRA ID
GSM5608032	MSK1263_A1_RNA	SAMN22021642	SRX12457939
GSM5608033	MSK1263_A3_RNA	SAMN22021643	SRX12457940
GSM5608034	MSK1263_A4_RNA	SAMN22021644	SRX12457941
GSM5608035	MSK1263_LN_RNA	SAMN22021645	SRX12457942
GSM5608036	MSK1263_Normal_RNA	SAMN22021646	SRX12457943
...

The study includes a total of **199 samples**, categorized into RNA, TCR sequencing, and multimer data pertaining to NSCLC samples.

Would you like to download the supplementary files for this GEO entry or export this metadata as [JSON](#) or [CSV](#)? Let me know how you'd like to proceed!

Your Prompt

Give me a task...

.pdf

Drop File Here
- OR -
Click to Upload

Submit Prompt/PDF

Live Logs

- New interaction started.
- User input: PDF is attached.
- Invoking agent...
- Agent response received.
- Interaction complete.

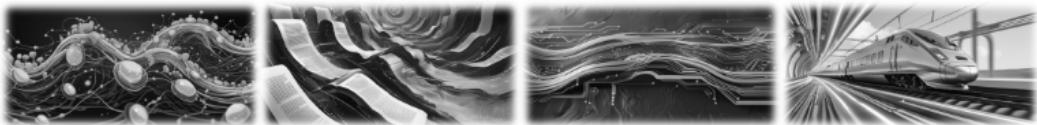
- New interaction started.
- User input: Extract metadata
- Invoking agent...
- Agent response received.
- Interaction complete.

- New interaction started.
- User input: Fetch GEO metadata for accession GSE185206
- Invoking agent...
- Agent response received.
- Interaction complete.

Figure 14. GEO metadata extraction in CellAtria following PDF-based article parsing. Upon user request, the agent retrieves and displays study metadata associated with the referenced GEO accession. The interface summarizes sample-level information and prompts the user to proceed with file download or metadata export. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Download dataset from GSM5608032, GSM5608033, and GSM5608034



The datasets for the following GSM accessions have been successfully downloaded to your working directory

(/mnt/work/projects/cellatria/data/my_projects/tumor_specific_T_cells):

- **GSM5608032:** MSK1263_A1_RNA
- **GSM5608033:** MSK1263_A3_RNA
- **GSM5608034:** MSK1263_A4_RNA

Would you like to proceed with additional actions? For example:

1. Organize the downloaded datasets.
2. Start analyzing the data using **CellExpress**.
3. Fetch additional metadata or datasets.

Let me know how you'd like to continue!

List the current contents of my working directory



Here are the current contents of your working directory (/mnt/work/projects/cellatria/data/my_projects/tumor_specific_T_cells):

```
.dominokeep
└── GSM5608032
    ├── GSM5608032_MSK1263_A1_RNA_barcodes.tsv.gz
    ├── GSM5608032_MSK1263_A1_RNA_features.tsv.gz
    └── GSM5608032_MSK1263_A1_RNA_matrix.mtx.gz
└── GSM5608033
    ├── GSM5608033_MSK1263_A3_RNA_barcodes.tsv.gz
    ├── GSM5608033_MSK1263_A3_RNA_features.tsv.gz
    └── GSM5608033_MSK1263_A3_RNA_matrix.mtx.gz
└── GSM5608034
    ├── GSM5608034_MSK1263_A4_RNA_barcodes.tsv.gz
    ├── GSM5608034_MSK1263_A4_RNA_features.tsv.gz
    └── GSM5608034_MSK1263_A4_RNA_matrix.mtx.gz
```



Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
⌚ User input: fetch GSE metadata for accession GSE185206
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.

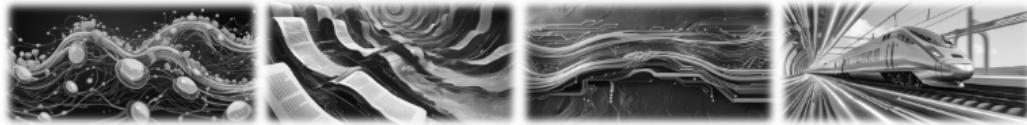
⌚ New interaction started.
⌚ User input: Download dataset from GSM5608032, GSM5608033, and GSM5608034
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.

⌚ New interaction started.
⌚ User input: List the current contents of my working directory.
🕒 Invoking agent...
✅ Agent response received.
🕒 Interaction complete.
```

Figure 15. GSM-based sample-level dataset acquisition in CellAtria. Upon user request, the agent downloads the specified GSM accessions and confirms successful retrieval. The interface presents a structured view of the file contents for each sample. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Set my working directory at /mnt/work/projects/cellatria/data/my_projects/cancer_data

Your working directory is now set to:

📁 /mnt/work/projects/cellatria/data/my_projects/cancer_data

Would you like to:

1. Extract metadata from a scientific article (URL or PDF)?
2. Fetch or download a dataset (e.g., from GEO)?
3. Organize or process an existing dataset (e.g., for the CellExpress pipeline)?

Let me know how you'd like to proceed!

Download <https://datasets.cellxgene.cziscience.com/76c942bd-45c3-47ec-b290-1e695ec9c177.h5ad> and <https://datasets.cellxgene.cziscience.com/f6e35982-3bef-47fe-b14a-60d2e8965f20.h5ad> to my current working directory

Both files have been successfully downloaded to your current working directory:

📁 /mnt/work/projects/cellatria/data/my_projects/cancer_data

1. 76c942bd-45c3-47ec-b290-1e695ec9c177.h5ad
2. f6e35982-3bef-47fe-b14a-60d2e8965f20.h5ad

What would you like to do next?

- o Process these files with the **CellExpress** pipeline?
- o Organize them further?
- o Perform a quick inspection of the content?

Your Prompt

Give me a task...

📤 .pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
at /mnt/work/projects/cellatria/data/my_projects/cancer_data
  📁 Invoking agent...
  ✓ Agent response received.
  ✨ Interaction complete.

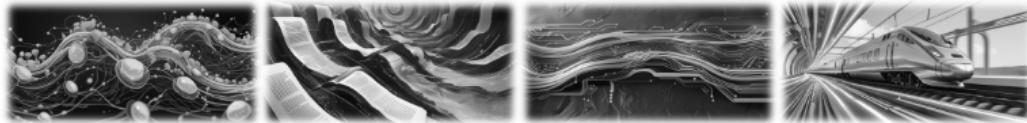
---  

  ● New interaction started.
  User input: Download
  https://datasets.cellxgene.cziscience.com/76c942bd-45c3-47ec-b290-1e695ec9c177.h5ad and
  https://datasets.cellxgene.cziscience.com/f6e35982-3bef-47fe-b14a-60d2e8965f20.h5ad to my current working
  directory
  📁 Invoking agent...
  ✓ Agent response received.
  ✨ Interaction complete.
```

Figure 16. Direct download and workspace integration of preprocessed datasets. Upon receiving a user prompt, the agent sets the working directory and confirms the change with contextual follow-up options. When provided with direct URLs, the agent retrieves and stores the datasets in the specified location. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



What would you like to do next?

- Process these files with the **CellExpress** pipeline?
- Organize them further?
- Perform a quick inspection of the content?

```
rename 76c942bd-45c3-47ec-b290-1e695ec9c177.h5ad to sample_1 and move it to  
/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_1
```

The file has been successfully renamed to **sample_1.h5ad** and moved to:

/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_1

Would you like to proceed with the second file, process these files using **CellExpress**, or perform any other tasks? Let me know!

```
rename f6e35982-3bef-47fe-b14a-60d2e8965f20.h5ad to sample_2 and move it to  
/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_2
```

The file has been successfully renamed to **sample_2.h5ad** and moved to:

/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_2

Both files are now organized in their respective directories. Would you like to:

1. Process these files with the **CellExpress** pipeline?
2. Perform any additional organization or inspection?
3. Start a new task?

Your Prompt

Give me a task...

.pdf



Drop File Here

- or -

Click to Upload

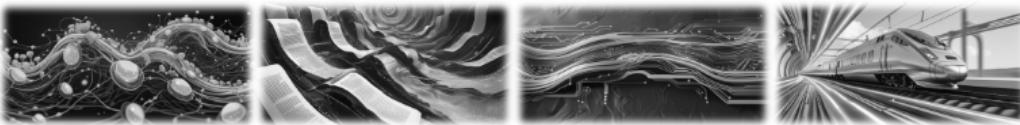
Submit Prompt/PDF

Live Logs

```
New interaction started.  
User input: rename 76c942bd-45c3-47ec-b290-1e695ec9c177.h5ad to sample_1  
and move it to  
/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_1  
Invoking agent...  
Agent response received.  
Interaction complete.  
  
New interaction started.  
User input: rename f6e35982-3bef-47fe-b14a-60d2e8965f20.h5ad to sample_2  
and move it to  
/mnt/work/projects/cellatria/data/my_projects/cancer_data/sample_2  
Invoking agent...  
Agent response received.  
Interaction complete.
```

Figure 17. Dataset renaming and organization within the working directory. Upon user request, the agent renames the datasets and organizes them into sample-specific subdirectories. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria



stroe a metadata.csv in my working directory with two columns, sample:sample_1,sample_2, data_type:paper_1, paper_2

A metadata file named **metadata.csv** has been successfully created in your working directory:

/mnt/work/projects/cellatria/data/my_projects/cancer_data

It contains the following data:

- o **Columns:** sample, data_type
 - o **Rows:** sample_1:paper_1, sample_2:paper_2

What would you like to do next?

- Inspect or edit the metadata file?
 - Proceed with processing the files using **CellExpress?**
 - Start a new task?

List the current contents of my working directory

Here are the current contents of your working directory:

/mnt/work/projects/cellatria/data/my_projects/cancer_data

```
└── metadata.csv  
└── sample_1  
    └── sample_1.h5ad  
└── sample_2  
    └── sample_2.h5ad
```

You now have:

1. The metadata file: **metadata.csv**.
 2. Two subdirectories: **sample_1** and **sample_2**, each containing their respective **.h5ad** files.

What would you like to do next?

- Inspect or edit any file?
 - Process these files with the **CellExpress** pipeline?
 - Start a new task?

Your Prompt

Give me a task...

.pdf

↑

Drop File

- or -

[Submit Prompt/PDE](#)

Live Logs

- ✓ Agent response received.
- Interaction complete.

-
- New interaction started.
- User input: stroe a metadata.csv in my working directory with two columns, sample:sample_1, sample_2,
- data_type:paper_1, paper_2
- Invoke agent...
- ✓ Agent response received.
- Interaction complete.

-
- New interaction started.
- User input: List the current contents of my working directory
- Invoke agent...
- ✓ Agent response received.
- Interaction complete.

Figure 18. Metadata file creation using CellAtria. Upon user request, the agent generates a structured metadata file based on the provided inputs. The live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

a

Agent Backend

Step: chatbot Model: gpt-4o-2024-11-20 **Tokens:** 33 completion, 9862 prompt, 9895 total

```
{"chatbot": {"messages": AIMessage(content=""), additional_kwargs={"tool_calls": [{"id": "call_pXXLCbuXlxDMVSLFYAYiLYFD", "function": {"arguments": {"url": "https://"}}, "status": "PENDING"}]}}, {"chatbot": {"messages": AIMessage(content="")}, "status": "PENDING"}
```

Step: tools Tool: fetch_article_metadata_url

```
{"tools": {"messages": [ToolMessage(content="Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection.", url="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1234567/abstract")]}}, {"chatbot": {"messages": AIMessage(content="The title of the article is:\n\n**Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection.**\n\nAbstract: Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection. These changes are likely due to the immature immune system and the unique physiological characteristics of the infant. The study found that the immune response in young infants is different from that in older children and adults, with higher levels of certain cytokines and lower levels of others. This may explain why young infants are more susceptible to certain infections and why they have a higher rate of complications than older children and adults. The study also found that the transcriptional changes in young infants are reversible, suggesting that the immune system can adapt to the changing environment and maintain homeostasis. The findings of this study provide new insights into the immune system of young infants and may help to develop better treatments for common childhood illnesses. The results of this study are important for understanding the biology of the immune system in young infants and for developing new treatments for common childhood illnesses. The results of this study are important for understanding the biology of the immune system in young infants and for developing new treatments for common childhood illnesses."}, "status": "PENDING"}}
```

Step: chatbot Model: gpt-4o-2024-11-20 **Tokens:** 31 completion, 35299 prompt, 35330 total

```
{"chatbot": {"messages": AIMessage(content="The title of the article is:\n\n**Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection.**\n\nAbstract: Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection. These changes are likely due to the immature immune system and the unique physiological characteristics of the infant. The study found that the immune response in young infants is different from that in older children and adults, with higher levels of certain cytokines and lower levels of others. This may explain why young infants are more susceptible to certain infections and why they have a higher rate of complications than older children and adults. The study also found that the transcriptional changes in young infants are reversible, suggesting that the immune system can adapt to the changing environment and maintain homeostasis. The findings of this study provide new insights into the immune system of young infants and may help to develop better treatments for common childhood illnesses. The results of this study are important for understanding the biology of the immune system in young infants and for developing new treatments for common childhood illnesses. The results of this study are important for understanding the biology of the immune system in young infants and for developing new treatments for common childhood illnesses."}, "status": "PENDING"}}
```

b

Terminal Panel

Command

Enter shell command (e.g., ls -la)

Terminal Output

```
$ cd /mnt/work/projects/cellatria/data/my_projects/cancer_data && pwd  
/mnt/work/projects/cellatria/data/my_projects/cancer_data  
--  
$ tree .  
. --  
|   metadata.csv  
|   sample_1  
|   |   sample_1.h5ad  
|   sample_2  
|       sample_2.h5ad  
2 directories, 3 files  
--
```

c

File Browser

Directory Path

/mnt/work/projects/cellatria/data/my_projects/cancer_data

Current Path: /mnt/work/projects/cellatria/data/my_projects/cancer_data

Folders: 2 | Files: 1

Subdirectories

- .. (Up)
- sample_1
- sample_2

Refresh Directory

Files in Directory

- metadata.csv

d

Export Chat / Model

Download Chat

Chat.json
chat_71b57ecc.json 1.3 KB ↓

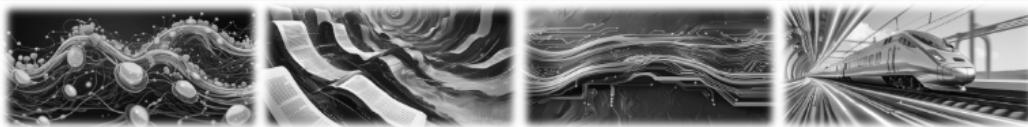
Download LLM Metadata

LLM.json
llm_ead1a923.json 206.0 B ↓

Figure 19. System-level panels of the CellAtria user interface. **a)** The “Agent Backend” panel displays step-by-step agent reasoning, tool calls, and backend responses. **b)** The “Terminal Panel” allows execution of shell commands within the agent runtime. **c)** The “File Browser” provides a navigable view of the directory structure. “Refresh Directory” updates the view when a new “Directory Path” is provided, and “Files in Directory” lists the files at that location. **d)** The “Export” panel allows users to retrieve session-level outputs via two buttons: one for downloading the full agent–user interaction log and another for exporting structured metadata associated with the active language model configuration. Note: The header and chatbot window were cropped to enhance panel visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



```
GSM0107200  
└── barcodes.tsv.gz  
└── features.tsv.gz  
└── matrix.mtx.gz  
GSM6189266  
└── barcodes.tsv.gz  
└── features.tsv.gz  
└── matrix.mtx.gz
```

The file names have been standardized for 10X Genomics processing. What would you like to do next?

1. Set up a **CellExpress** pipeline for analysis.
2. Inspect or validate specific files in the directory.
3. Clean and organize your working directory further.

Let me know!

I need to initiate a CellExpress pipeline run

To initiate the **CellExpress pipeline**, you will need the following required parameters:

Parameters Required

1. `--input`: Provide the absolute path to the directory containing your sample folders and `metadata.csv`.
2. `--project`: A name for your project (e.g., `infants_vaccine`).
3. `--species`: Organism species, either `hs` (human) or `mm` (mouse).
4. `--tissue`: Tissue name or relevant UBERON ID (e.g., `blood`).
5. `--disease`: Disease name or MONDO ID (use MONDO:0000000 for healthy/control).

Pipeline Arguments

Additional optional arguments for quality control, normalization, clustering, etc., are available. Let me know if you'd like to include any.

Would you like me to help configure the pipeline by populating the required fields?

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

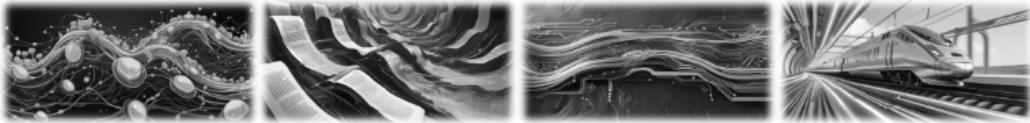
Live Logs

```
generated file named by the working directory.  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—  
🕒 New interaction started.  
🕒 User input: List the current contents of my working directory.  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—  
🕒 New interaction started.  
🕒 User input: I need to initiate a CellExpress pipeline run  
🕒 Invoking agent...  
✅ Agent response received.  
🕒 Interaction complete.  
—
```

Figure 20. Initialization of the CellExpress analysis pipeline. In response to user input, CellAtria displays the required parameters for initiating CellExpress and prompts the user to specify additional configurable options, enabling tailored pipeline execution. Note: the integrated live log viewer provides real-time feedback on execution status and confirms successful completion. The system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Configure pipeline arguments with the following configuration:

```
--input: current working directory  
--project: 3gz_test  
--species: hs  
--tissue: pbmc  
--disease: normal
```

Confirmed: The pipeline configuration is valid and ready to execute. Below are the details of your setup:

- **Input Directory:** /mnt/work/projects/cellatria/data/my_projects/infants_vaccine
- **Project Name:** 3gz_test
- **Species:** hs (Human)
- **Tissue:** pbmc
- **Disease:** normal
- **Defaults:** Quality control and analysis parameters are set to defaults (e.g., 750 min UMIs, Scrublet threshold of 0.25, etc.).

Would you like to proceed with running the pipeline using this configuration?

[Preview the current pipeline configuration.](#)

Here is your current **CellExpress** pipeline configuration:

```
--input /mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
--project 3gz_test  
--species hs  
--tissue pbmc  
--disease normal  
--min_umi_per_cell 750  
--min_genes_per_cell 250  
--min_cell 3  
--max_mt_percent 15.0
```

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

[Submit Prompt/PDF](#)

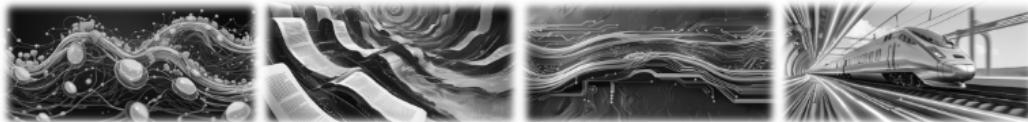
Live Logs

```
● New interaction started.  
>User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_pro  
jects/infants_vaccine  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
--input: current working directory
```

Figure 21. CellExpress pipeline configuration preview and validation. The agent validates user-provided parameters for the CellExpress pipeline and confirms readiness for execution. Default quality control thresholds are applied automatically. A YAML-formatted configuration summary is presented to the user for final review prior to pipeline initiation. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



I need to perform cell type annotation. What method does CellExpress use for cell type annotation?

CellExpress offers two powerful methods for cell type annotation, both of which can be used individually or combined based on your needs:

1. SCimilarity:

- A cross-tissue cell annotation framework.
- Matches scRNA-seq data with pre-trained datasets to achieve accurate annotations, even for rare or novel cell types.
- You need to specify the path to the SCimilarity model directory using `--sci_model_path`.

2. CellTypist:

- A highly detailed and accurate cell annotation tool optimized for tissue-specific classification.
- Requires access to pre-trained `.pkl` models provided by the CellTypist library.
- To use this method, configure:
 - The path to the CellTypist model directory (`--cty_model_path`).
 - The model name without `.pkl` extension (`--cty_model_name`).

Configuration Example:

You can set **annotation methods** using the `--annotation_method` argument.

For instance:

```
--annotation_method scimilarity,celltypist
```

Make sure to:

- Define the file paths for models when required (e.g., `--sci_model_path` for SCimilarity).
- Specify an appropriate model (e.g., `Immune_All_High` for CellTypist, using `--cty_model_name`).

Your Prompt

Give me a task...

.pdf

Drop File Here
- OR -
Click to Upload

Submit Prompt/PDF

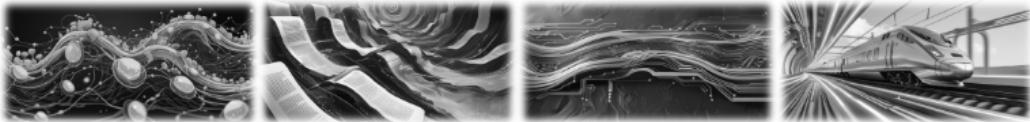
Live Logs

```
● New interaction started.  
>User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
—  
● New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
—  
● New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
--inout: current working directory
```

Figure 22. Cell type annotation strategies in CellExpress. Upon user request, the agent enumerates the automated cell type annotation methods supported by CellExpress. The provided configuration example illustrates how users specify annotation strategies and corresponding model paths through pipeline arguments. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Configuration Example:

You can set **annotation methods** using the `--annotation_method` argument.

For instance:

```
--annotation_method scimilarity,celltypist
```

Make sure to:

- Define the file paths for models when required (e.g., `--sci_model_path` for SCimilarity).
- Specify an appropriate model (e.g., `Immune_All_High` for CellTypist, using `--cty_model_name`).

Would you like help setting up these arguments for your pipeline?

Configure the pipeline to use SCimilarity for cell type annotation. specify `--scimilarity_model_path` as
`/mnt/work/projects/cellatria/data/scimilarity/model_v1.1`

Successfully configured the pipeline for cell type annotation using **SCimilarity**.

Updated Settings:

- Annotation Method:** scimilarity
- SCimilarity Model Path:** `/mnt/work/projects/cellatria/data/scimilarity/model_v1.1`

Would you like to:

1. Preview the full pipeline configuration?
2. Run the pipeline now?
3. Make additional adjustments to the configuration?

Let me know how you'd like to proceed!

Your Prompt

Give me a task...

.pdf

Drop File Here
- OR -
Click to Upload

Submit Prompt/PDF

Live Logs

```
New interaction started.  
User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
Invoking agent...  
Agent response received.  
Interaction complete.  
New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
Invoking agent...  
Agent response received.  
Interaction complete.  
New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
--input: current_working_directory
```

Figure 23. Configuration of SCimilarity-based cell type annotation in CellExpress. In response to user instruction, the agent configures the CellExpress pipeline to utilize SCimilarity for automated cell type annotation. The interface confirms the updated parameters and presents follow-up actionable options to proceed with execution. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Launch the CellExpress pipeline.

The CellExpress pipeline is now running successfully!

Job Details:

o **Process ID (PID):** 85399

o **Log File:** [View Log](#)

o **Command Used:**

```
python -m cellexpress.main --input /mnt/work/projects/cellatria/data/my_projects/infants_vaccine --project 3gz_test --species hs
```

Next Steps:

1. **Monitor the Log File for updates:** [/mnt/work/projects/cellatria/data/my_projects/infants_vaccine/cellexpress_8603a56f.log](#).

2. **Check Status** of the running process using the PID: 85399.

3. Once completed, review the outputs:

- o Quality-Controlled Data.
- o Cell Type Annotations.
- o Summary Reports.

Would you like to:

1. **Review the log file** for progress?

2. **Confirm the job is running** using its PID?

3. Anything else?

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

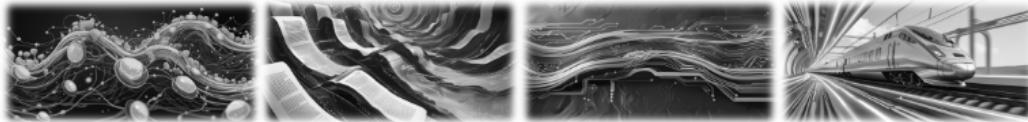
Live Logs

```
● New interaction started.  
User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
--input: current_working_directory
```

Figure 24. Launch and background execution of the CellExpress pipeline. Upon receiving the execution request, the agent initiates the CellExpress pipeline in detached mode using the specified configuration. The process ID (PID) and log file path are displayed to facilitate runtime monitoring. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



check the status of process ID (PID): 85399

The pipeline job with PID: 85399 is still running smoothly.

Command Executed:

```
python -m cellexpress.main --input /mnt/work/projects/cellatria/data/my_projects/infants_vaccine --project 3gz_test --species hs
```

Would you like to:

1. Monitor progress by reviewing the log file?
2. Receive updates automatically once it's completed?
3. Do anything else while waiting (e.g., fetch other data or configure a new pipeline)?

Let me know! 😊

review the log file cellexpress_8603a56f.log

Log Summary:

The pipeline has successfully completed most of the standard analysis components, including quality control (QC), normalization, PCA, clustering, and embedding. Key outputs include:

1. QC Metrics:

- UMI counts, gene counts, and mitochondrial percentages were calculated and documented.
- Filters applied (e.g., min UMI: 750, min genes: 250).

2. Clustering:

↓

Your Prompt

Give me a task...

.pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

Live Logs

```
● New interaction started.  
User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
🕒 Invoking agent...  
✓ Agent response received.  
● Interaction complete.  
  
● New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
--input: current_working_directory
```

Figure 25. Real-time monitoring and log inspection of an active CellExpress pipeline. Given the process ID, the agent confirms the active execution status of the CellExpress pipeline, displays the full launch command, and provides access to the log file for real-time inspection. Upon user request, the agent parses the log to summarize completed stages of the pipeline. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



review the log file cellexpress_8603a56f.log

Log Summary:

The **CellExpress pipeline** has successfully completed its full execution, and all outputs have been generated and stored.

Key highlights from the log:

1. Interactive Report Generated:

- Path: /mnt/work/projects/cellatria/data/my_projects/infants_vaccine/outputs_cellexpress_v1-0-0_bk3hD2F/report_cellexpress_v1-0-0_bk3hD2F_2025-06-10.html.

2. Quality-Controlled Data Output:

- Data stored as an AnnData object:
 - Path: /mnt/work/projects/cellatria/data/my_projects/infants_vaccine/outputs_cellexpress_v1-0-0_bk3hD2F/counts_qced_cellexpress_v1-0-0_bk3hD2F_2025-06-10.h5ad.
 - Dimensions: **86,704 cells x 20,630 genes**.

3. Annotated Data File:

- Includes cell type annotations (`cellstate_scimilarity`, `celltype_scimilarity`).
- Path: /mnt/work/projects/cellatria/data/my_projects/infants_vaccine/outputs_cellexpress_v1-0-0_bk3hD2F/adata_cellexpress_v1-0-0_bk3hD2F_2025-06-10.h5ad.

4. Configuration Snapshot:

- Stored for reproducibility:
 - Path: /mnt/work/projects/cellatria/data/my_projects/infants_vaccine/outputs_cellexpress_v1-0-0_bk3hD2F/config_cellexpress_v1-0-0_bk3hD2F_2025-06-10.json.

5. Pipeline Status:

- Completed Successfully 🎉

Your Prompt

Give me a task...

↓ .pdf



Drop File Here

- OR -

Click to Upload

Submit Prompt/PDF

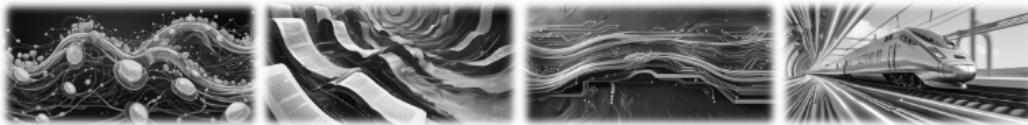
Live Logs

```
New interaction started.  
User input: Set my working directory  
to:  
/mnt/work/projects/cellatria/data/my_projects/infants_vaccine  
Invoking agent...  
Agent response received.  
Interaction complete.  
—  
New interaction started.  
User input: I need to initiate a  
CellExpress pipeline run  
Invoking agent...  
Agent response received.  
Interaction complete.  
—  
New interaction started.  
User input: Configure pipeline  
arguments with the following  
configuration:  
—input: current working directory
```

Figure 26. Completion of the CellExpress pipeline and structured output verification. Upon user request, the agent confirms successful completion of the CellExpress pipeline and enumerates the resulting output files. The integrated live log viewer provides real-time feedback on execution status and confirms successful completion. Note: the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

Welcome to cellAtria

Agentic Triage of Regulated single cell data Ingestion and Analysis



Set my working directory at `/mnt/work/projects/cellatria/data/my_projects/one_shot_exe`, and run cellexpress using single cell data from <https://www.nature.com/articles/s41467-023-43758-2>, no questions asked.

The CellExpress pipeline has been successfully launched! 🎉

Job Details:

- **PID:** 80067
- **Log File:** `/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/cellexpress_9910cf38.log`
- **Configuration:**
 - **Input Directory:** `/mnt/work/projects/cellatria/data/my_projects/one_shot_exe`
 - **Project Name:** Infant-Vaccine-Study
 - **Species:** *Homo sapiens* (hs)
 - **Tissue:** *blood*
 - **Disease:** Not disease-specific
- **Analysis Steps:** Default settings were used for quality control, clustering, and annotation.

You can monitor progress via the log file or let me know if you need updates on the pipeline's status.

[review the log file cellexpress_9910cf38.log](#)

Summary of the Log Output:

- **Pipeline Execution Completed Successfully!** 🎉

The CellExpress pipeline has finished all stages. A summary report and processed data have been generated.

◦ Key Outputs:

- **Report:**
`/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/outputs_cellexpress_v1-0-0_vNnf5sz/report_cellexpress_v1-0-0_vNnf5sz_2025-07-11.html`
- **Quality-Controlled Data (H5AD):**
`/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/outputs_cellexpress_v1-0-0_vNnf5sz/counts-qced_cellexpress_v1-0-0_vNnf5sz_2025-07-11.h5ad`
- **Annotated Data (H5AD):**
`/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/outputs_cellexpress_v1-0-0_vNnf5sz/adata_cellexpress_v1-0-0_vNnf5sz_2025-07-11.h5ad`
- **Configuration:**
`/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/outputs_cellexpress_v1-0-0_vNnf5sz/config_cellexpress_v1-0-0_vNnf5sz_2025-07-11.json`

◦ Processed Data Shape:

- **Cells x Genes:** `(86,704 × 20,630)`

◦ Analysis Highlights:

- Metadata columns for cells: `['sample_id', 'description', 'pct_counts_mito', 'leiden_cluster', ...]`.
- Multi-dimensional embeddings (e.g., PCA: 30 components, UMAP: 2D embedding) were saved.
- Differentially expressed genes (DEGs) were identified per Leiden cluster.

You can access the report for a comprehensive summary, or review the processed data for further analysis. Let me know if you need assistance visualizing or interpreting any of these results! 🎉

Figure 27. Autonomous execution of the CellExpress pipeline initiated by a single prompt. The CellAtria interface demonstrates complete orchestration of the document-to-analysis workflow following submission of a high-level user prompt - without iterative user-agent interactions. Note: the log viewer and the system-level panels (backend, terminal, file browser, and chat export) have been cropped for improved visibility.

```

Agent Backend
▼

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 32 completion, 10011 prompt, 10043 total
{"chatbot": {"messages": AIMessage(content='', additional_kwargs={"tool_calls": [{"id": "call_3W1pJvfHzNRCzP0Vg1vSideI", "function": {"arguments": {"path": "/mnt/work/projects/cellatria/data/my_projects/one_shot_exe"}}, "name": "set_working_directory"]}], "name": "chatbot", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: tools Tool: set_working_directory

```
{"tools": {"messages": [ToolMessage(content='✅ Working directory set to: /mnt/work/projects/cellatria/data/my_projects/one_shot_exe', name='set_working_directory')], "name": "tools", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 33 completion, 10074 prompt, 10107 total
{"chatbot": {"messages": AIMessage(content='', additional_kwargs={"tool_calls": [{"id": "call_Rrjj2DWxIvVsIFNKWkFpBuJz", "function": {"arguments": {"url": "https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10000000/"}}, "name": "fetch_article_metadata_url"]}], "name": "chatbot", "role": "agent"}, "name": "chatbot", "role": "model"}

Step: tools Tool: fetch_article_metadata_url

```
{"tools": {"messages": [ToolMessage(content='["Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection."]', name='fetch_article_metadata_url')], "name": "tools", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 103 completion, 35511 prompt, 35614 total
{"chatbot": {"messages": AIMessage(content='', additional_kwargs={"tool_calls": [{"id": "call_jdzAFbxSo0Pxf7FyOyzJJ4TQ", "function": {"arguments": {"Project": "Human", "path": "/mnt/work/projects/cellatria/data/my_projects/one_shot_exe"}}, "name": "refine_article_metadata"]}], "name": "chatbot", "role": "agent"}, "name": "chatbot", "role": "model"}

Step: tools Tool: refine_article_metadata

```
{"tools": {"messages": [ToolMessage(content='✅ Metadata prepared for review. Awaiting user confirmation to store.', name='refine_article_metadata', id='c997618c-4'), {"content": "The metadata has been successfully stored.", "name": "refine_article_metadata", "role": "agent"}], "name": "tools", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 66 completion, 35636 prompt, 35702 total
{"chatbot": {"messages": AIMessage(content='', additional_kwargs={"tool_calls": [{"id": "call_S42fzfC7F2Jtla6AoHpEpg3K", "function": {"arguments": {"path": "/mnt/work/projects/cellatria/data/my_projects/one_shot_exe"}}, "name": "make_directory"]}], "name": "chatbot", "role": "agent"}, "name": "chatbot", "role": "model"}

Step: tools Tool: make_directory Step: tools Tool: fetch_geo_metadata

```
{"tools": {"messages": [ToolMessage(content='✅ Created directory: /mnt/work/projects/cellatria/data/my_projects/one_shot_exe', name='make_directory', id='e5567ab'), ToolMessage(content='["Young infants display heterogeneous serological responses and extensive but reversible transcriptional changes following infection."]', name='fetch_geo_metadata', role='agent')], "name": "tools", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 740 completion, 36532 prompt, 37272 total
{"chatbot": {"messages": AIMessage(content='', additional_kwargs={"tool_calls": [{"id": "call_hp5icZ1M2w7D5GnjrD4JsqcR", "function": {"arguments": {"gse_id": "GSE204716", "path": "/mnt/work/projects/cellatria/data/my_projects/one_shot_exe"}}, "name": "download_geo"]}], "name": "chatbot", "role": "agent"}, "name": "chatbot", "role": "model"}

Step: tools Tool: download_geo Step: tools Tool: store_geo_metadata_file

```
{"tools": {"messages": [ToolMessage(content='X Download failed: [Errno 2] No such file or directory: \'/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/GSE204716\'', name='download_geo', role='agent'), ToolMessage(content='✅ Metadata saved to \'/mnt/work/projects/cellatria/data/my_projects/one_shot_exe/metadata_9fc6cd9f.csv\'', name='store_geo_metadata_file', role='agent')], "name": "tools", "role": "agent"}, "name": "chatbot", "role": "model"}
```

Step: chatbot Model: gpt-4o-2024-11-20 Tokens: 142 completion, 37378 prompt, 37520 total
{"chatbot": {"messages": AIMessage(content='The metadata for the GEO dataset \'GSE204716\' has been successfully saved as \'metadata_9fc6cd9f.csv\' in the working direc...', name: "chatbot", role: "agent"), "name": "chatbot", "role": "model"}}

Figure 28. Agentic backend reasoning trace for one-shot execution workflow. Internal execution trace showing stepwise agent behavior as it interprets user intent and autonomously orchestrates tool invocations without requiring follow-up prompts. Each step documents how the agent decomposed a single compound prompt into a series of reasoning and execution actions. "chatbot" entries reflect language model outputs with tool-calling instructions, while "tools" steps record the successful execution of those calls. Metadata such as model version, token usage, and generated text is included for transparency and reproducibility. Note: the header, chatbot window, log viewer, and system-level panels (including backend trace, terminal, file browser, and chat export) have been cropped to enhance visual clarity.

can you batch correct single cell data with tool named "magicharm"?

Currently, I cannot process batch correction using a tool named "**MAGICHARM**". However, **CellExpress** supports batch correction specifically using well-integrated tools like **Harmony** and **scVI** for harmonization of single-cell data. Here's a brief overview:

1. **Harmony:** Corrects batch effects while preserving biological variation in single-cell data. It is especially used for large integrated datasets.
2. **scVI (Single-Cell Variational Inference):** A deep generative model for robust normalization and batch effect correction by learning intrinsic biological representations.

If you're looking to utilize **MAGICHARM** and have further details, you might consider providing that information or using this external tool separately.

If you'd like to proceed with **Harmony** or **scVI** for batch correction within the **CellExpress** pipeline, let me know!

Your Prompt

Give me a task...

.pdf

Drop File Here
- or -
Click to Upload

Submit Prompt/PDF

Live Logs

```

● New interaction started.
>User input: can you batch correct
single cell data with tool named
"magicharm"?
👉 Invoking agent...
✅ Agent response received.
👉 Interaction complete.
---
```

Agent Backend

Terminal Panel

File Browser

Export Chat / Model

Figure 29. CellAtria response to an out-of-scope tool request. User prompts the agent to harmonize single-cell data using an arbitrary, non-existent tool (“magicharm”). CellAtria recognizes that the requested tool is not part of the integrated pipeline, issues a structured and informative warning, and redirects the user toward the supported alternative tool.

List all command-line arguments supported by CellExpress, including their names, expected input types, default values, and descriptions. No gap inline... make it compact, one line per each!

Here's the compact list of **CellExpress** command-line arguments:

- `--input str` (required): Path to input directory with sample folders and `metadata.csv`.
- `--project str` (required): Project name for outputs and reports.
- `--species str` (required): Organism species, `hs` (human) or `mm` (mouse).
- `--tissue str` (required): Tissue name.
- `--disease str` (required): Disease name.
- `--config str` (optional): Path to JSON config file.
- `--min_umi_per_cell int` (default: 750): Min total UMI counts per cell.
- `--max_umi_per_cell int` (default: None): Max total UMI counts per cell.
- `--min_genes_per_cell int` (default: 250): Min genes per cell.
- `--max_genes_per_cell int` (default: None): Max genes per cell.
- `--min_cell int` (default: 3): Genes expressed in at least N cells.
- `--max_mt_percent float` (default: 15): Max mitochondrial gene percentage per cell.
- `--doublet_method str` (default: None): Doublet detection method, e.g., `scrublet`.
- `--scrublet_cutoff float` (default: 0.25): Max allowed scrublet doublet score.
- `--norm_target_sum float` (default: 1e4): Target UMI per-cell total counts for normalization.
- `--n_top_genes int` (default: 2000): Number of highly variable genes to retain.
- `--regress_out str` (default: no): Regress out library size or MT%.
- `--scale_max_value float` (default: 10): Clip after scaling.
- `--n_pcs int` (default: 30): Number of retained PCA components.
- `--batch_correction str` (default: None): Batch correction method, e.g., `harmony`/`scvi`.
- `--batch_vars str` (default: None): Batch correction column names.
- `--n_neighbors int` (default: 15): k-NN neighbors for graph clustering.
- `--resolution float` (default: 0.6): Leiden clustering resolution.
- `--compute_tsne str` (default: no): Compute t-SNE embedding, options `yes`/`no`.
- `--annotation_method str` (default: None): Annotation, options: `scimilarity`, `celltypist`.
- `--sci_model_path str` (default: None): Path to SCimilarity model directory.
- `--cty_model_path str` (default: None): Path to CellTypist model directory.
- `--cty_model_name str` (default: None): CellTypist model name (no extension).
- `--pval_threshold float` (default: 0.05): Max FDR-adjusted p-value limit for marker genes.
- `--logfc_threshold float` (default: 0.25): Min log-fold change for marker gene filtering.
- `--dea_method str` (default: wilcoxon): Differential method: `wilcoxon`, `t-test`, etc.
- `--pts_threshold float` (default: 0.1): Min cells fraction where marker expressed.
- `--top_n_deg_leidn int` (default: 100): Top marker genes per Leiden group.
- `--top_n_deg_scim int` (default: 100): Top genes for SCimilarity annotation.
- `--top_n_deg_ctlpst int` (default: 100): Top genes for CellTypist annotation.
- `--doc_url str` (default: None): Documentation link.
- `--data_url str` (default: None): Dataset repository link.
- `--only_qc str` (default: no): Output QC-only summary.
- `--fix_gene_names str` (default: no): Replace Ensembl IDs with gene symbols.
- `--plot_alpha float` (default: 0.7): Overlay transparency (0.0–1.0) in projection plots.

Figure 30. Agentic query interface listing all CellExpress arguments. Example interaction showing how CellAtria responds to a user prompt requesting a comprehensive list of CellExpress command-line arguments. The output includes argument names, expected input types, default values, and concise descriptions.