

## Preface

This user guide is intended for use by researchers with interest in single-cell omics data. The user guide will explain step-by-step three exemplary use-cases of the single-cell data integration platform.

## Start page

Figure (1) shows the start page of the Single-cell data integration platform ([www.single-cell.clst.riken.jp](http://www.single-cell.clst.riken.jp)). At first, the user selects the desired type of dataset to work with. Currently the platform distinguishes two types of datasets based on the source of the dataset:

1. Single-cell dataset generated at Division of Genomics Technologies (DGT), CLST RIKEN.
2. Dataset(s) published by other groups in peer-reviewed journals.

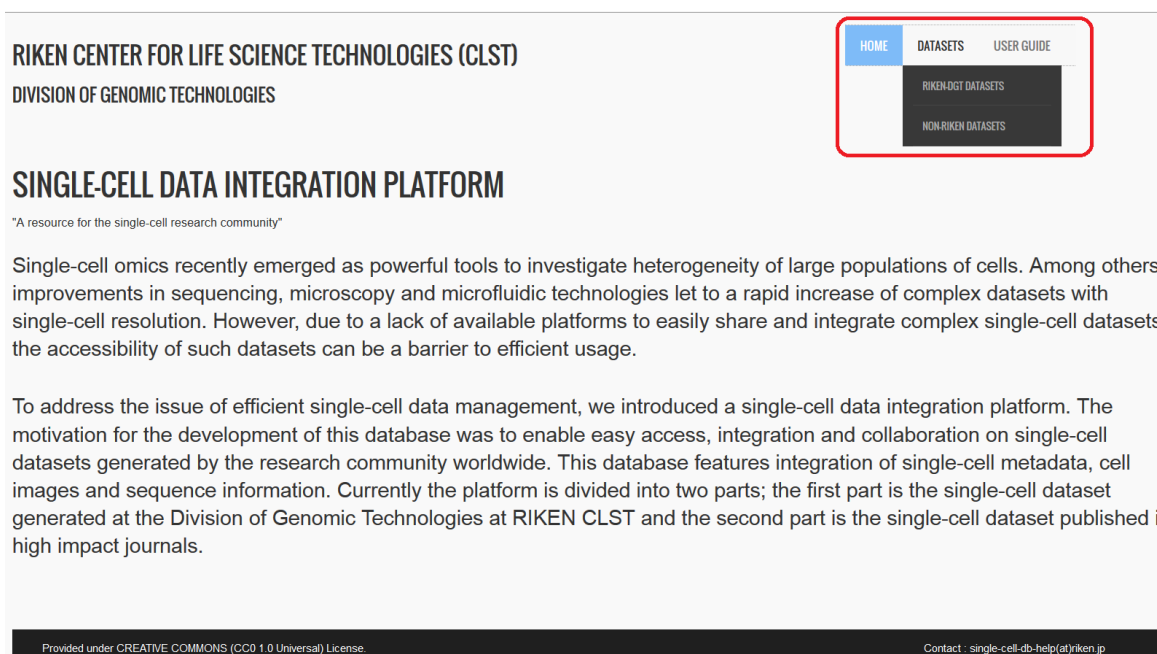


Figure 1: The start screen of the single-cell data integration platform

## Use-cases

The dataset from RIKEN CLST DGT features cell images, experimental metadata, sequence information, and gene expression data on the single-cell level. All these data can be used for different purposes. Here we introduce three potential use-cases:

- 1- Cell images enable quality checks on individual cells: users start from querying the metadata (e.g. cell fluorescence, cDNA concentration, read counts, etc.) and then select the cell/ or subset of cells of interest
- 2- Explore cell images and associated expression profiles
- 3- Search for genes of interest and find its associated expression values [TPM] and track a single cell based on a gene expression value.

The system is not limited to the above examples, it provides wide range of features that could be explored by the user to work with the available datasets.

### Use-case I: quality check of the individual cells

The cell with the cell\_id “1772-062-248\_A01” is displayed in the 3<sup>rd</sup> row of the first table of the [Fucci metadata](#) see Figure (2). The associated cell images (bright field, green fluorescence and red fluorescence mode) and sequencing information are linked to the cell “1772-062-248\_A01” click on the link [cell images](#) linked to cell images, clicking on “[Sequencing](#)” will redirect the user to [FASTQ files](#) (two files per cell) and links for BAM files , expression table and genome coordinate visualizations.

RIKEN Center for Life Science Technologies (CLST), Division of Genomic Technologies (DGT)  
Single-cell data integration platform - User guide

	Cell id	cDNA concentration	Error	Fluo QC	Ch2 corrected	Ch3 corrected	Low complexity	SPIKE 1	SPIKE 4	SPIKE 7	SPIKE 3	SPIKE 6	rRNA 18S	rRNA 28S	rRNA 5.8S	Nextera	HPV	HPV As	Control	Reads	HiSeq QC	Discard
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A01	3.74	0	FALSE	39.87	94.24	3,294.44	4,149.69	292.77	41.50	0.00	0.00	2,344.95	105.45	0.25	108,543.31	2,084.77	753.56	N4	3928008	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A02	0.17	1	TRUE	0.00	0.00	34,459.27	5,752.50	182.57	19.68	0.00	0.00	364.04	177.10	3.28	251,886.37	804.61	452.59	PC	914732	TRUE	TRUE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A03	2.83	0	FALSE	14.06	30.26	2,430.86	6,193.90	657.04	20.22	0.00	0.00	2,404.90	140.90	0.00	105,107.56	2,729.48	964.81	N4	4698330	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A04	3.74	0	FALSE	52.94	11.26	2,511.15	5,268.03	434.34	8.98	0.00	0.00	2,311.71	91.43	0.00	107,656.11	2,441.60	876.28	N4	4342234	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A05	4.20	3	TRUE	0.00	0.00	2,888.15	4,752.61	425.94	42.89	0.00	0.00	1,522.33	115.56	1.48	107,481.82	2,918.71	1,057.20	N4	4056928	TRUE	TRUE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A06	4.49	0	FALSE	19.30	50.07	3,480.43	3,927.83	323.39	42.97	0.00	0.00	1,385.26	75.63	0.74	103,527.78	2,340.10	818.17	N4	4072482	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A07	2.44	0	FALSE	14.01	66.43	2,525.47	7,935.73	850.45	145.06	0.00	0.00	2,484.46	145.06	0.95	119,995.43	1,744.66	624.54	N4	5280674	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A08	0.09	1	TRUE	0.00	0.00	5,164.95	530,263.16	27,778.87	3,795.19	0.00	0.00	43.49	9.82	0.00	194,424.03	16.83	11.22	N4	712880	TRUE	TRUE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A09	4.40	0	FALSE	21.20	95.07	3,517.27	3,657.90	258.71	36.30	0.00	0.00	1,744.36	114.90	0.35	112,681.92	2,246.98	769.44	N4	2837146	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A10	0.05	1	TRUE	0.00	0.00	9,221.41	526,919.67	36,647.38	3,305.86	0.00	0.00	22.34	0.00	0.00	189,241.81	18.61	11.17	NC	268614	TRUE	TRUE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A11	1.67	0	FALSE	13.97	23.41	1,782.11	12,399.56	1,290.19	205.92	0.00	0.00	3,033.69	216.68	0.00	131,754.42	371.95	126.96	N4	6041156	TRUE	FALSE
<input type="checkbox"/>	Sequencia (1) Cell Images (1) 1772-062-248_A12	4.78	0	FALSE	60.36	53.72	3,567.04	3,432.23	343.53	5.57	0.00	0.00	1,111.92	69.88	0.31	118,721.95	4,229.99	1,542.03	N4	3234052	TRUE	FALSE

Figure 2: List of 12 cells (one batch) in the Fucci metadata view

Use-case II: Visually explore cell images and its expression profile in the ZENBU (<http://fantom.gsc.riken.jp/zenbu/>) browser

Here we explain how cell images can be easily explored using the query interface. The query enables the filtering of the cells according to specific data values (quantitate and binary attributes). For example, one can filter a cell according to the distribution of fluorescence values in the cell. In the *Fucci* metadata the cells could be filtered according to a cut-off value in a fluorescence mode (e.g. Ch2\_corrected column in the metadata table) in which the green fluorescence value is recorded (Figure 3-1).

Fucci Metadata - Advanced search

Criteria:  All conditions  Any condition

NOT

Cell id  Contains

cDNA concentration  Equals

Error  Equals

Fluo QC  Contains

Ch2 corrected  More than

Figure 3-1: Search for a cell with green fluorescence concentration of more than 120 TPM

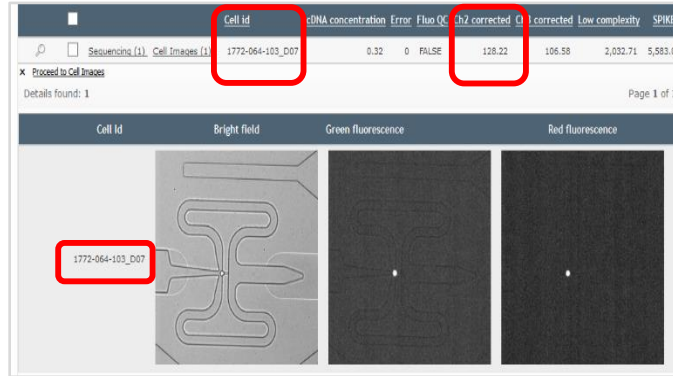


Figure 3.2: Explore cell image and expression profile.

With the *Fucci* [metadata](#) the cells could be filtered according to the type of error that was assigned based on manually checking all cell image files.

Use-case III: Search for gene of interest, find its associated expression values and track a single cell based on its gene expression value.

This use-case demonstrates how to search for gene(s) of interest and retrieve the expression [TPM] and get all cells expressing a particular gene. The search interface accepts three parameters for search including the gene symbol, Encode gene id, and gene description.

The screenshot shows the 'FuCCI Metadata' section of the platform. A search for 'PER3' is performed, resulting in a list of genes. The first result, 'PER3', is selected. This leads to a detailed view of 'Cells and gene expression (TPM)' for PER3, showing 472 details. A table lists cell IDs and their corresponding expression values in TPM. A red box highlights the 'Expression (TPM)' column, and a callout box points to the 'FuCCI Metadata' column, indicating that clicking on a cell ID leads to its metadata.

Cell Id	Expression (TPM)
FuCCI Metadata, 1772-062-248_B06	99.999
FuCCI Metadata, 1772-064-103_B10	99.999
FuCCI Metadata, 1772-067-038_E12	99.999
FuCCI Metadata, 1772-064-103_E02	54.277
FuCCI Metadata, 1772-067-039_C06	47.610
FuCCI Metadata, 1772-062-248_B04	46.627
FuCCI Metadata, 1772-067-038_A03	44.274
FuCCI Metadata, 1772-067-039_C11	29.933
FuCCI Metadata, 1772-067-038_D04	28.247
FuCCI Metadata, 1772-062-249_G03	24.729
FuCCI Metadata, 1772-062-248_B02	23.661
FuCCI Metadata, 1772-062-249_B09	23.610
FuCCI Metadata, 1772-064-103_B01	22.919
FuCCI Metadata, 1772-062-249_H10	20.679
FuCCI Metadata, 1772-062-248_D09	15.376
FuCCI Metadata, 1772-062-249_G09	14.794
FuCCI Metadata, 1772-062-249_D01	14.201
FuCCI Metadata, 1772-067-039_H01	13.553
FuCCI Metadata, 1772-062-248_H07	11.452
FuCCI Metadata, 1772-062-249_F11	10.020
FuCCI Metadata, 1772-062-249_D09	9.036
FuCCI Metadata, 1772-062-249_A01	8.994
FuCCI Metadata, 1772-062-248_A12	8.186

Figure 4: Searching for gene of interest in all samples. The output is a list of single cells and the associated expression value [TPM] per cell.

## Bulk-download

All database data is provided for bulk download in two formats:

- 1- MySQL format: SQL files with the data definition and manipulation statement embedded all the tables are easy to import in any version of the MySQL database.
- 2- CSV format: We provide the data in CSV format. A common type widely used in Biology and Bioinformatics and easily accessible via text editor programs.

## GitHub repository

The dynamic documents for R scripts are available from the GitHub repository Population-Transcriptomics/Cell-Cycle-on-C1

## Data export

Selected data can be exported in different formats (e.g. CSV, XML). This flexibility in data format will allow users to obtain data in a way that is most suitable for their individual platform requirements.

## Non-RIKEN dataset

Under the Non-RIKEN section of the database, the user will find collected datasets that have been published elsewhere. For each dataset we provide an overview of the dataset, options for download, and link to the original journal publication (Figure 5).

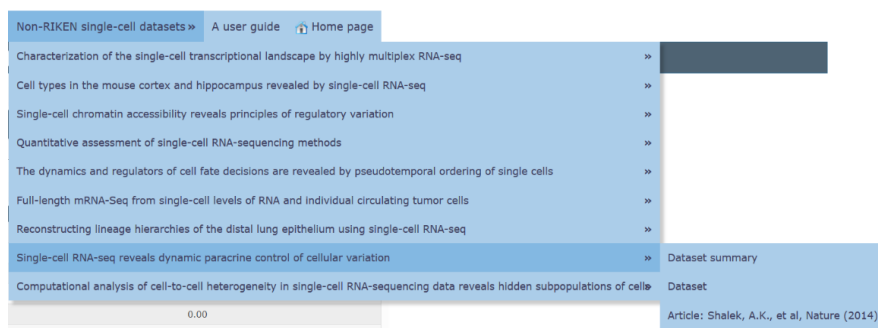


Figure 5: Non-RIKEN dataset layout.

## Contact information

Division of Genomic Technologies  
RIKEN Center for Life Science Technologies

RIKEN Center for Life Science Technologies (CLST), Division of Genomic Technologies (DGT)  
Single-cell data integration platform - User guide

1-7-22 Suehiro-chō , Tsurumi-ku, Yokohama, Kanagawa, 230-0045 JAPAN

Tel: +81-45-503-9245, Fax: +81-45-503-9216

email: **single-cell-db-help(at)riken.jp**