Single-cell data integration platform workflow and milestones

Large scale data managing unit

RIKEN CLST

Characteristics of single-cell biology

1- VERY LARGE sample size

2- Various data types and measurements

3- Tracking of raw and processed data

4- Collaboration, publishing and disseminations

To meet the above requirements, a single-cell database mgt system created with the following functionalities:

- Store, process and integrate all types of data
- Automate data curation process
- Image processing and size reduction
- Enable data sharing and publishing

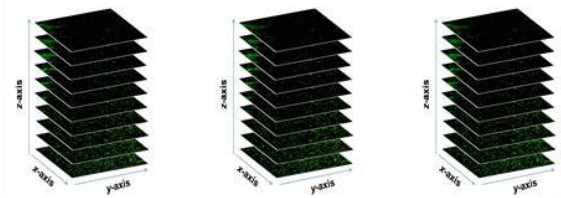
Data types from c1 workflow

Numerical

- Fluorescence measurements
- cDNA concentration
- Library inforamtion
- SPIKE
- Raw and processed seq data

Cellomics™ C0 cell images

InCell Analyzer 6000™ images



IN Cell Analyzer 6000™ image processing

IN Cell Analyzer 6000™
11 z-stack high resolution cell image / field
3168 (cell images)

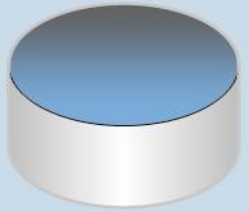
Convert to 96 well coordinates system
A - 02(fld 048 wv Green - dsRed z

1772-123-289_A01_BF_z01

Raw images conversion and compression
8,193 > 566 KB

Generate z-stack image movie

Select representative zstack from 11
Assign cell meta information to images



Single-cell database

IN CELL ANALYZER 6000™ IMAGE PROCESSING

Convert to 96 well coordinates system

02(fld 048 wv Green - dsRed z

772-123-289_A

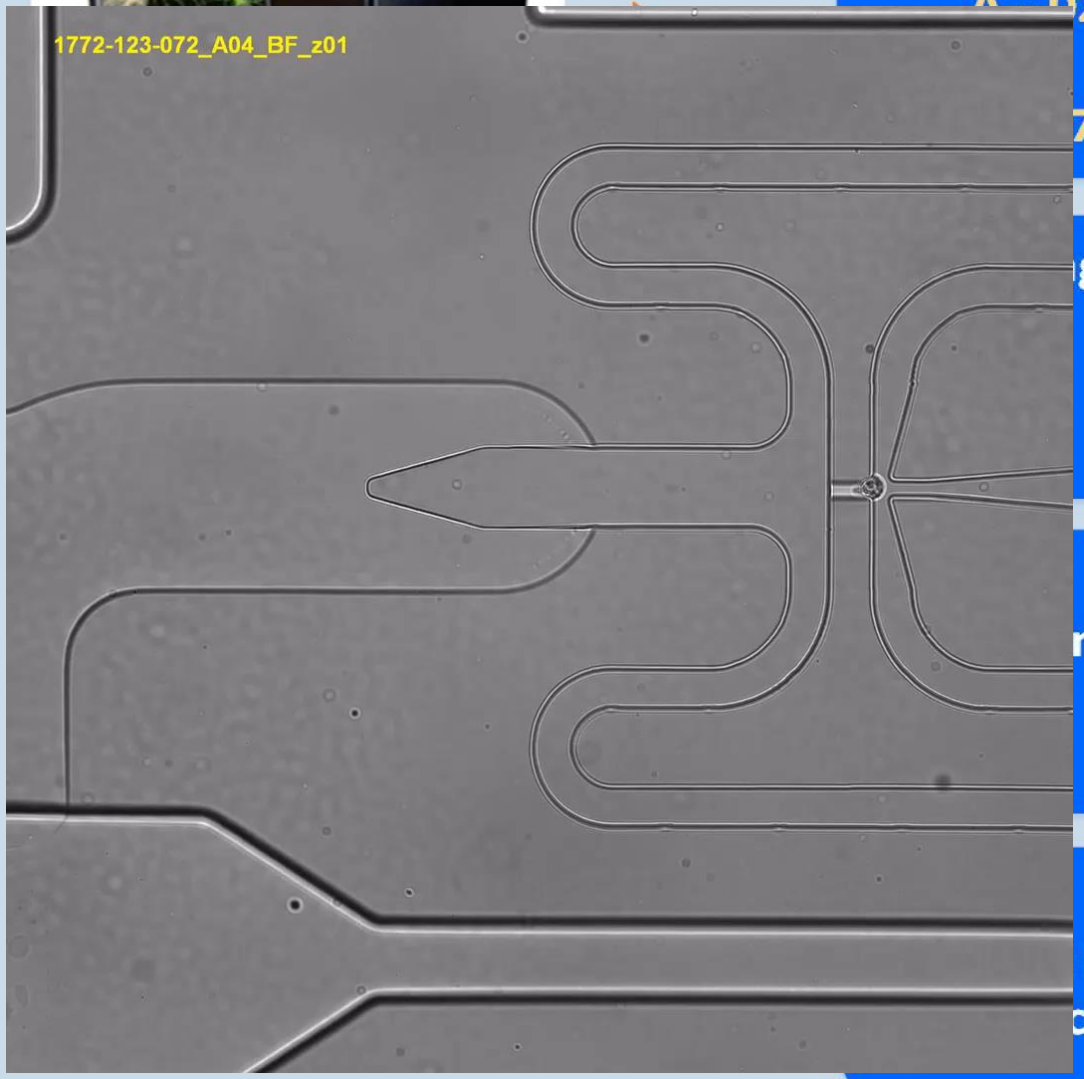
ges conversion

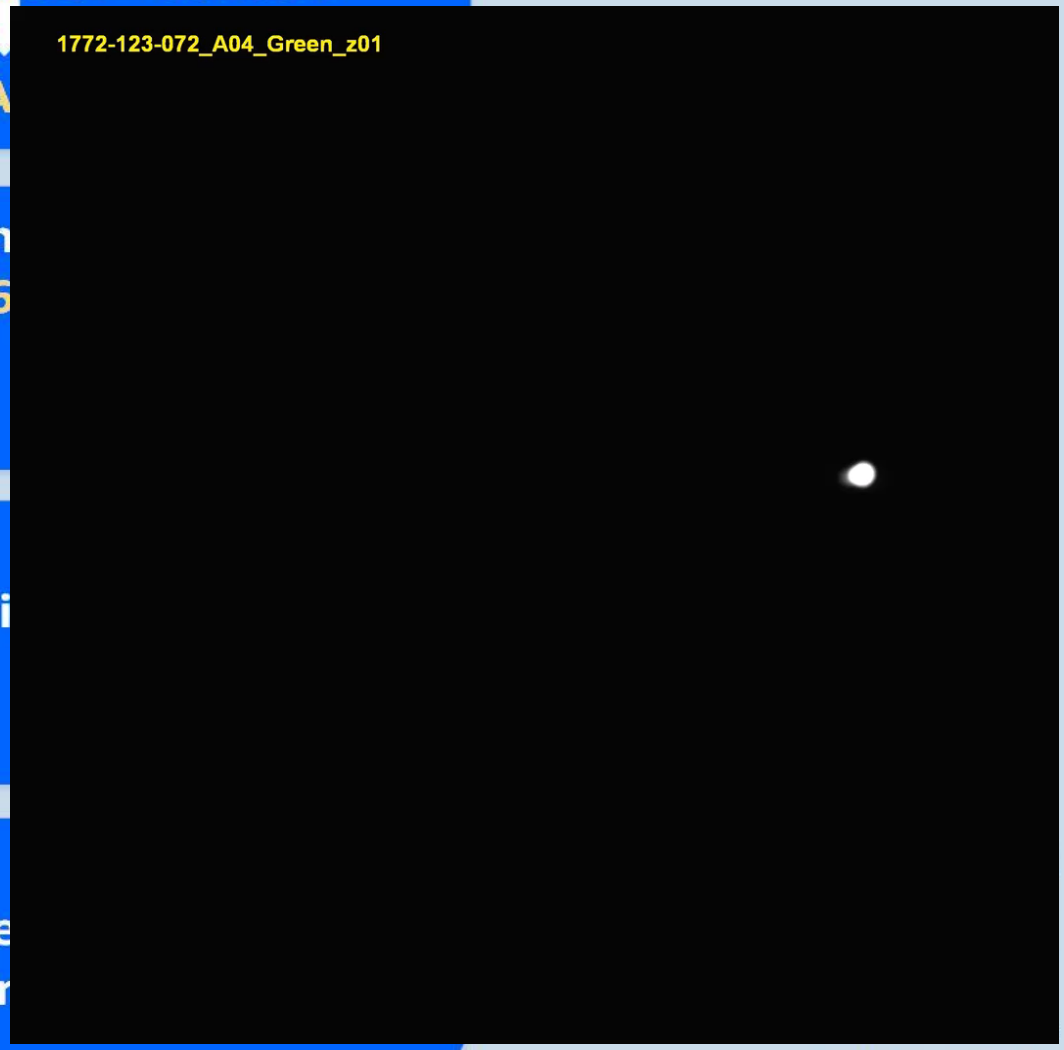
8,193 > 56

nerate z-stack i

representative

cell meta inforr





Data integration

Fluorescence measurements

cDNA concentration

Library inforamtion

SPIKE

Raw and processed seq data

Numerical

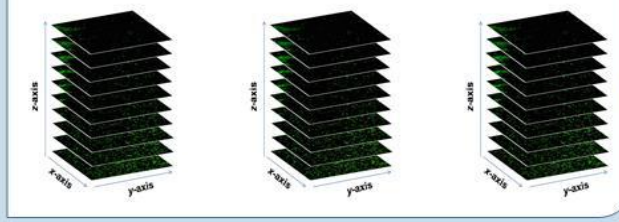
C1 chip dataset for single-run

Integration

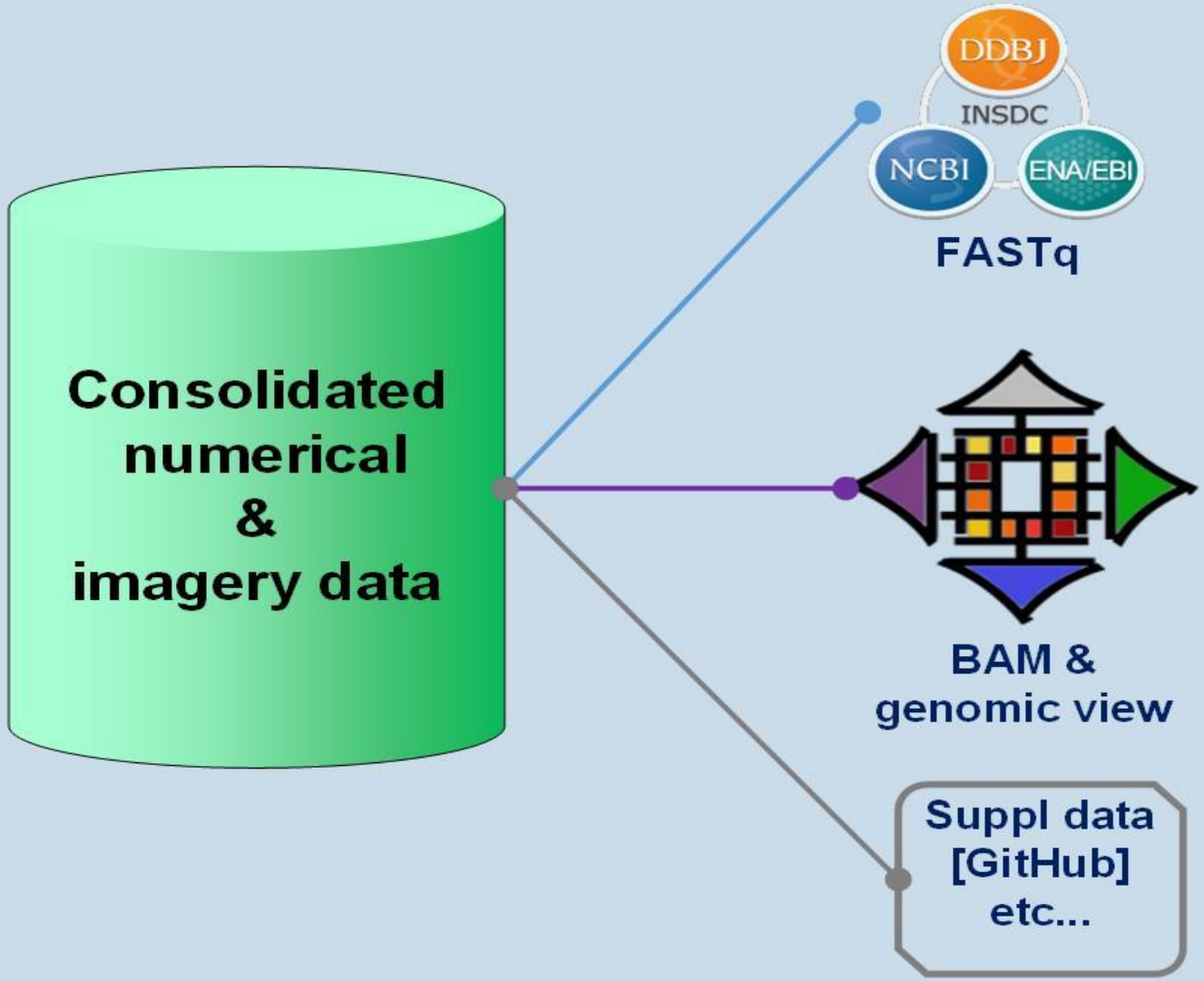
Consolidated numerical & Imagery data

Cellomics™ C0 cell images

InCell Analyzer 6000™ images



Integrate with external databases



DATA ACQUISITION, PROCESSING AND ANALYSIS OF PUBLISHED SINGLE-CELL [HUMAN & MOUSE]

1- Single-cell data curation and ontology annotation

2- Single-cell transcriptomic & genomic analysis

3- Analysis of single-cell genomic contaminations

4- Functional annotation [Goterm, KEGG, Pathways]

5 - Database system creation

DATA ACQUISITION, PROCESSING AND ANALYSIS OF PUBLISHED SINGLE-CELL [HUMAN & MOUSE]

1- Single-cell data curation and ontology annotation

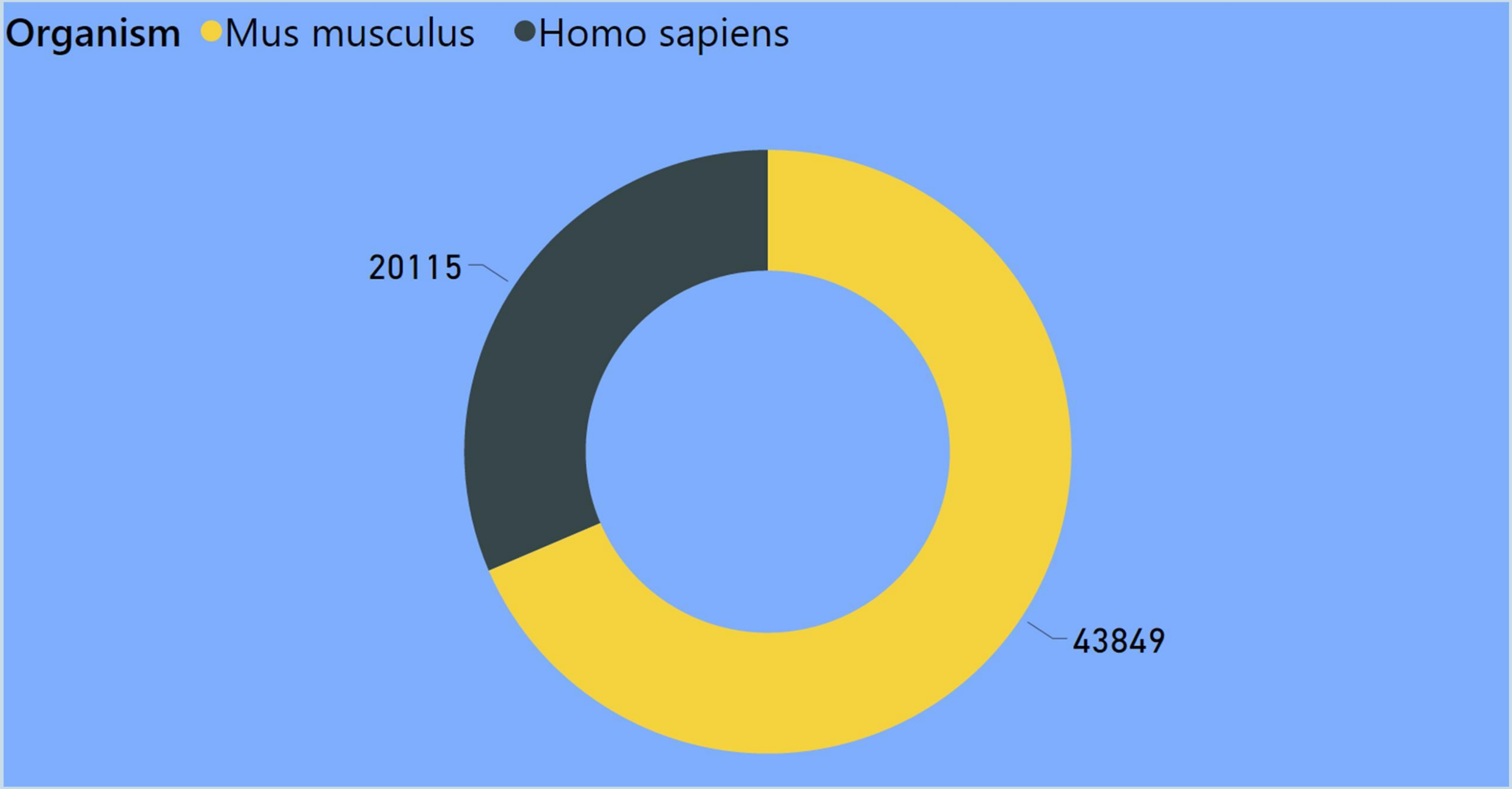
2- Single-cell transcriptomic & genomic analysis

3- Analysis of single-cell genomic contaminations

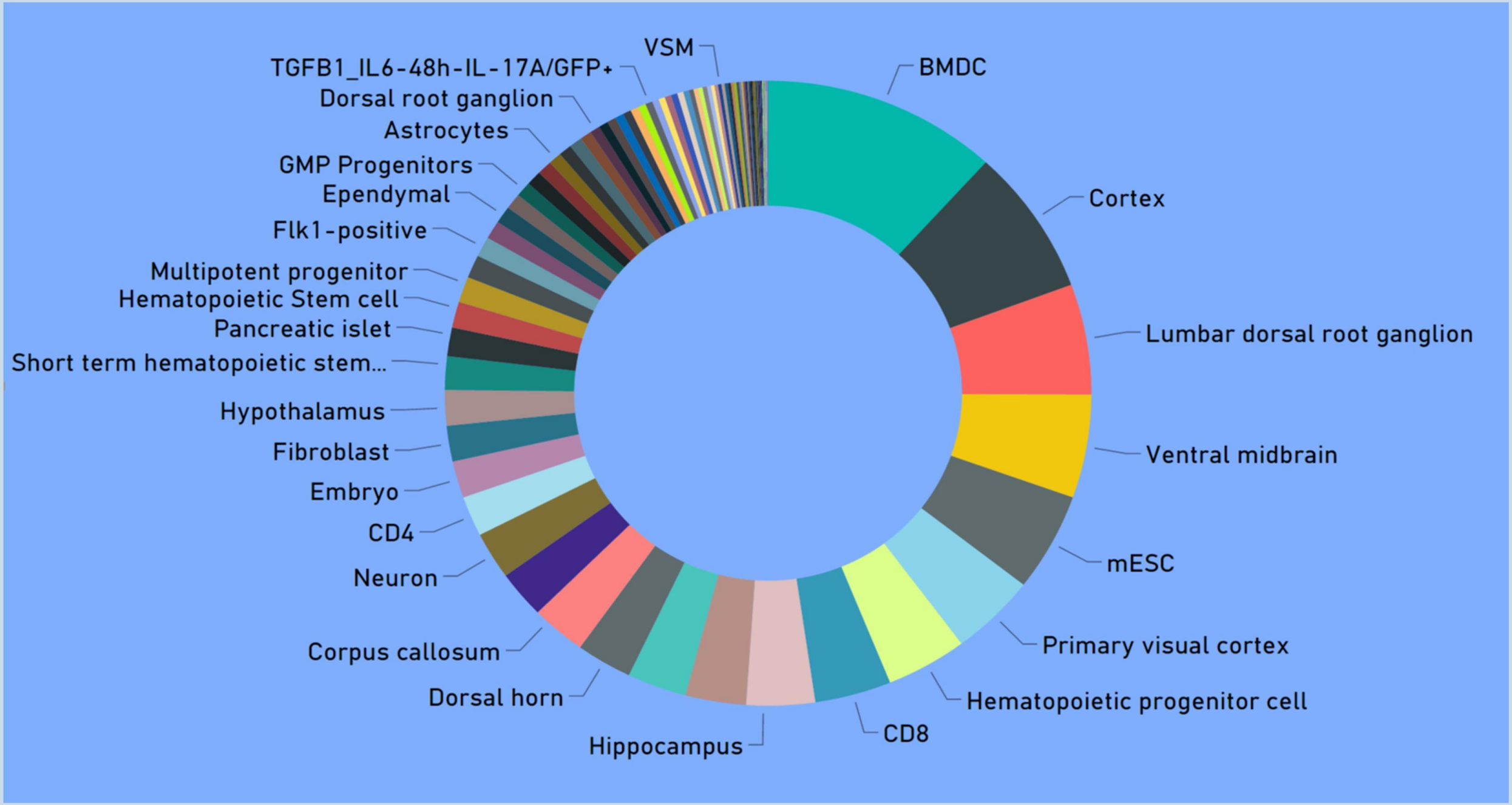
4- Functional annotation [Goterm, KEGG, Pathways]

5 – Database system creation

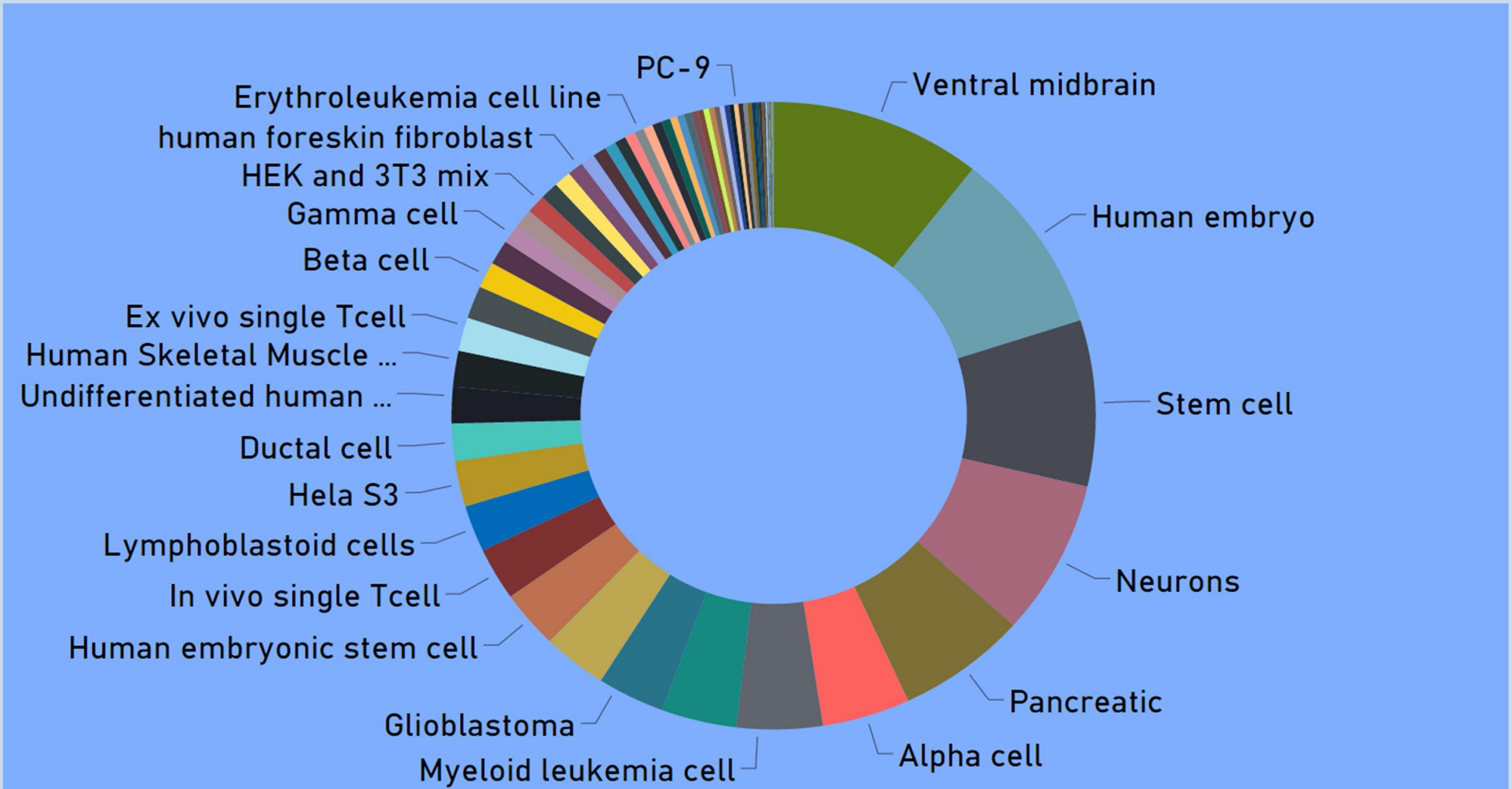
Total number of curated and processed single-cell



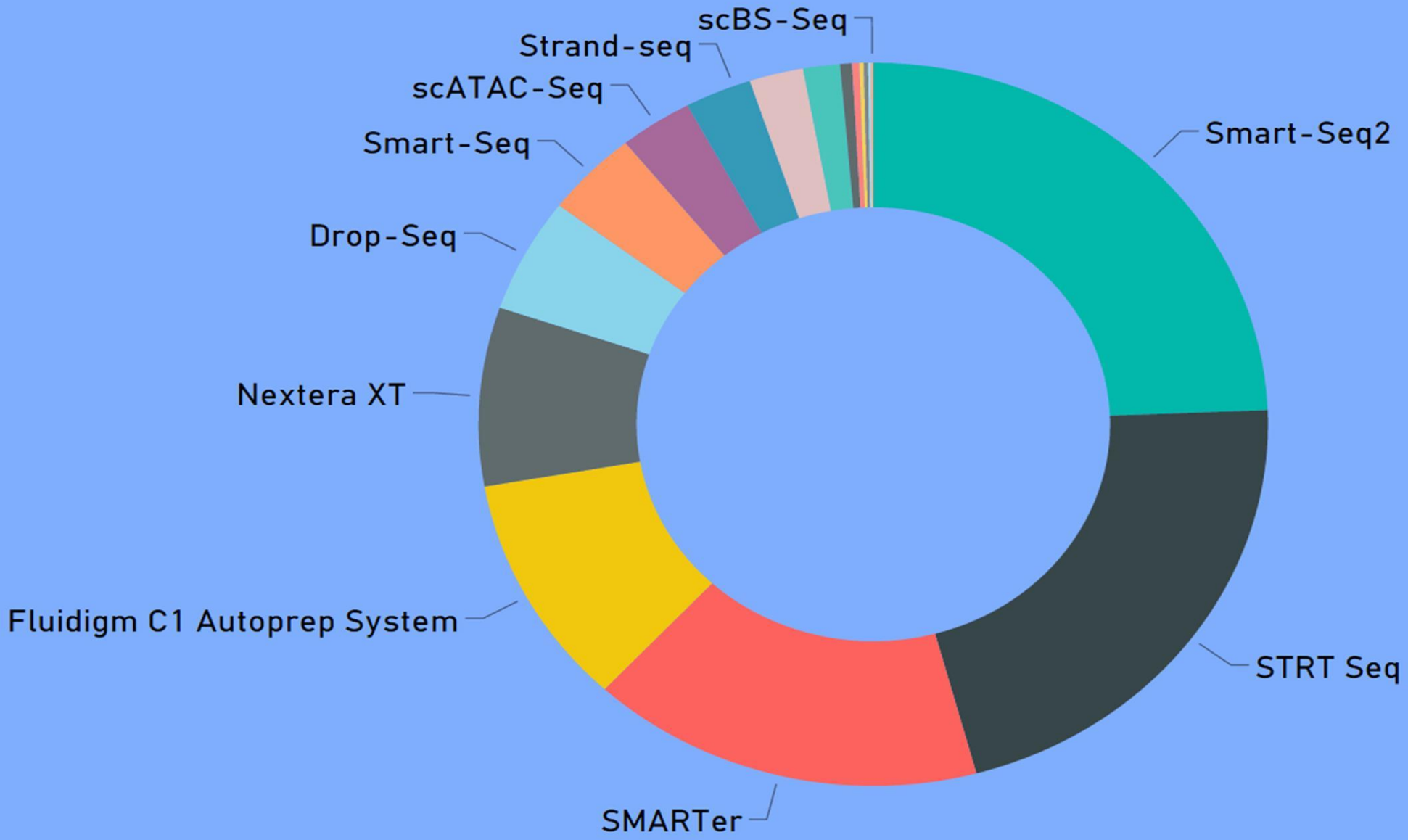
MOUSE SINGLE-CELL TYPES



Human single-cell types

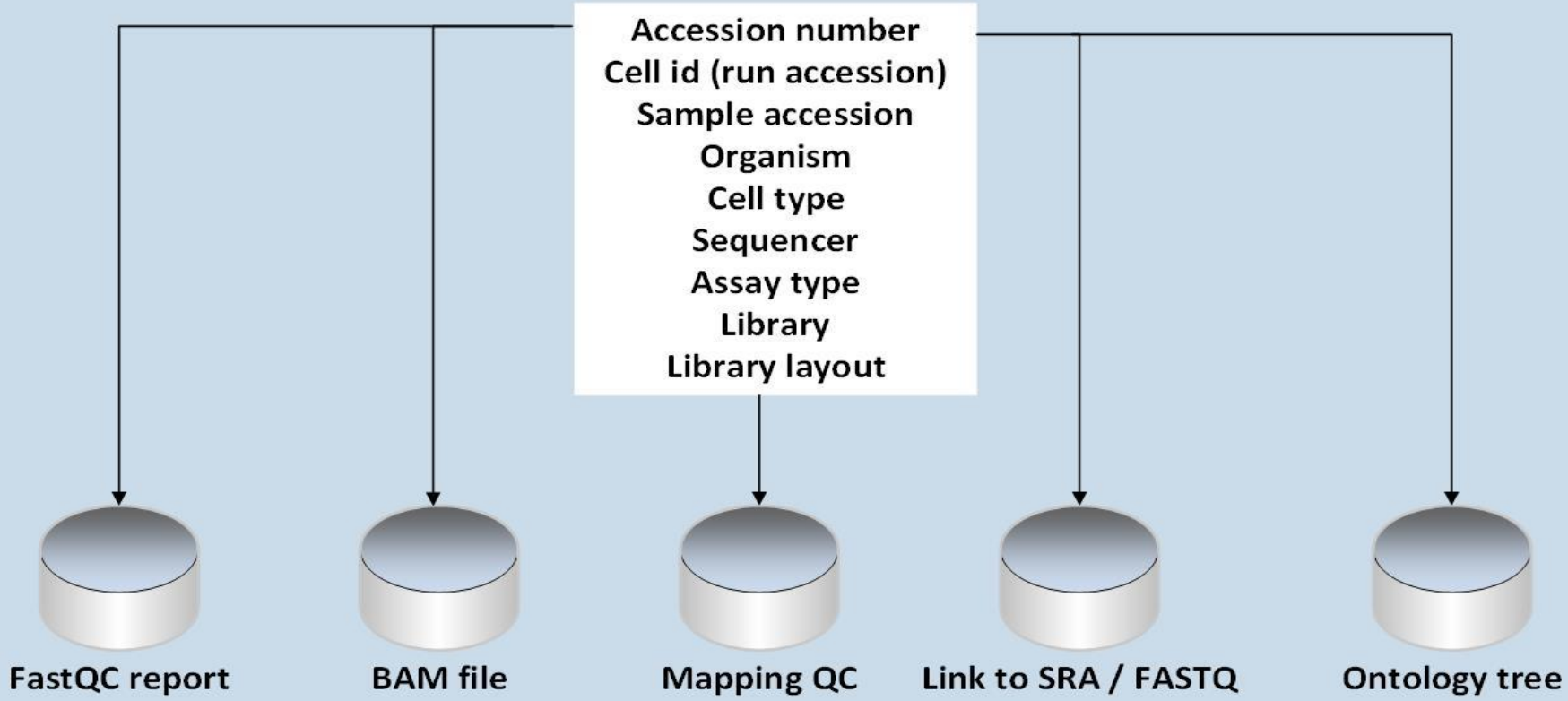


Single-cell library protocols



Curation and ontology annotation

1- Manually curate dataset metadata provided by GEO / ArrayExpress / DDBJ. Select the following items as minimum meta information



A DATASET VIEW

Study Meta Info

single-cell.clst.riken.jp/non_riken_data/study_meta_info_list.php

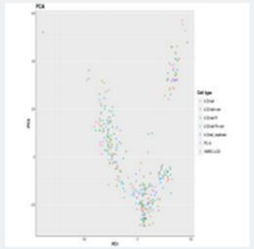
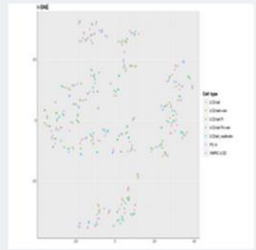
Most Visited Getting Started Small RNA lab - SciLif...

Public single-cell dataset » **Dataset list** » | A user guide | Home page

Select all | With selected... | More... | Search for: | search

Details found: **68** [**1** 2 3 4] Page **1** of **4** Records Per Page: 20

Single-cell sample list (337)

Accession number	DRA001287
Dataset title	scRNA-Seq analysis of a series of lung adenocarcinoma cell lines. We analyzed a More ...
Article abstract	We analyze a total of 336 single-cell RNA-Seq libraries from seven cell lines. T More ...
Run metadata	Click to download study's detailed metadata (txt formatted)
Data repository	Link to to the public data repository [GEO - DDBJ - ArrayExpress]
PubMed record	Show PubMed record
Authors provided files	Click to download author's files (expression table, count info, etc.) .gz foramtted
PCA	 t-SNE 
Search gene expression (FPKM)	Search for a gene in FPKM exp. table

Curation and ontology annotation

1- Assign ontology term for each cell
   1-1 Cell from cell lines : Cell line ontology
   1-2 Primary cell: Cell ontology
   1-3 Cell from tissue: Uber Anatomy Ontology
2- Select the nearest term from the target ontology based on **is-a** or **part-of** relationship in the tree ontology
3- we utilized EBI Ontology Lookup web service

Hematopoietic

DATA ACQUISITION, PROCESSING AND ANALYSIS OF PUBLISHED SINGLE-CELL [HUMAN & MOUSE]

1- Single-cell data curation and ontology annotation

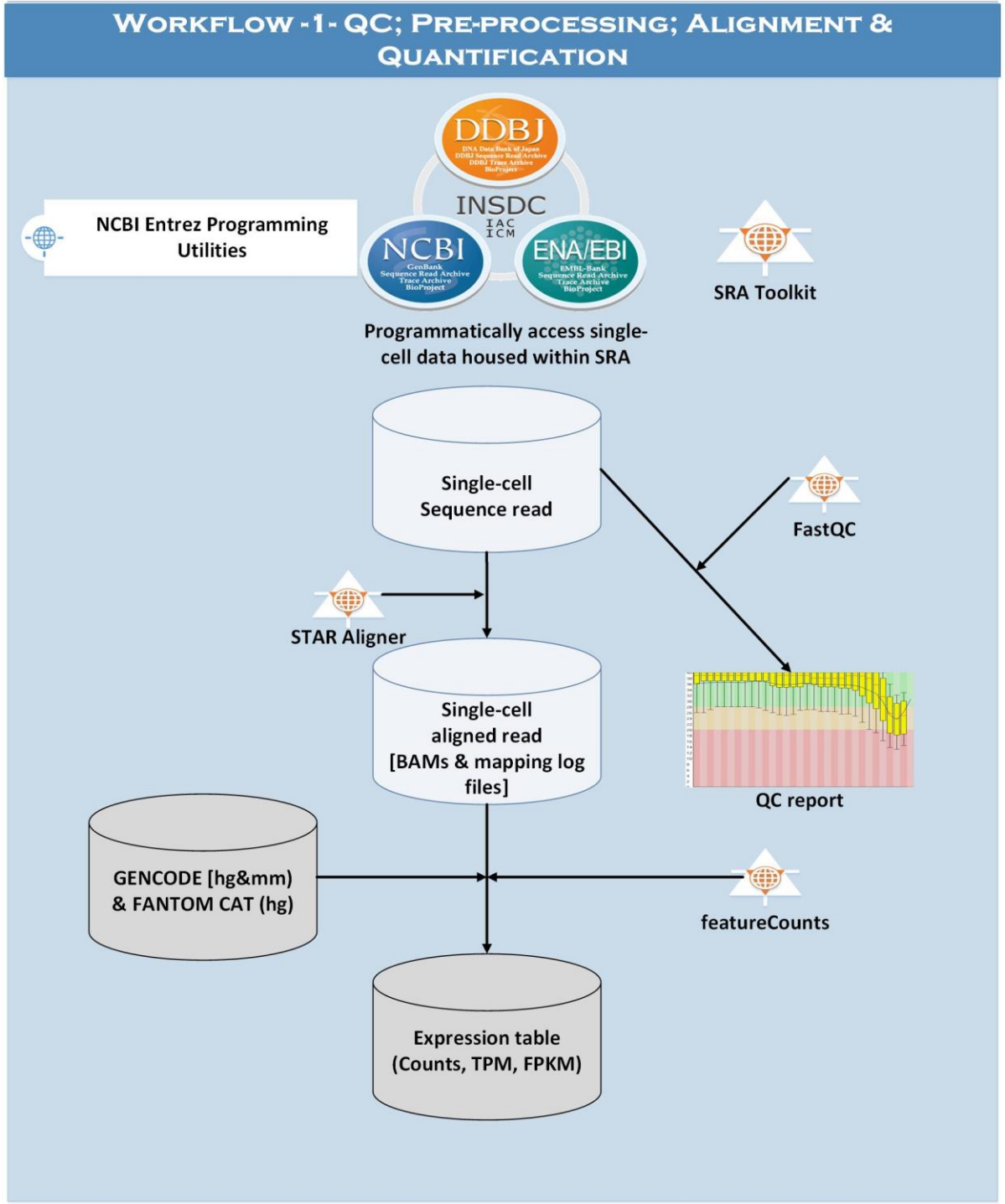
2- Single-cell transcriptomic & genomic analysis

3- Analysis of single-cell genomic contaminations

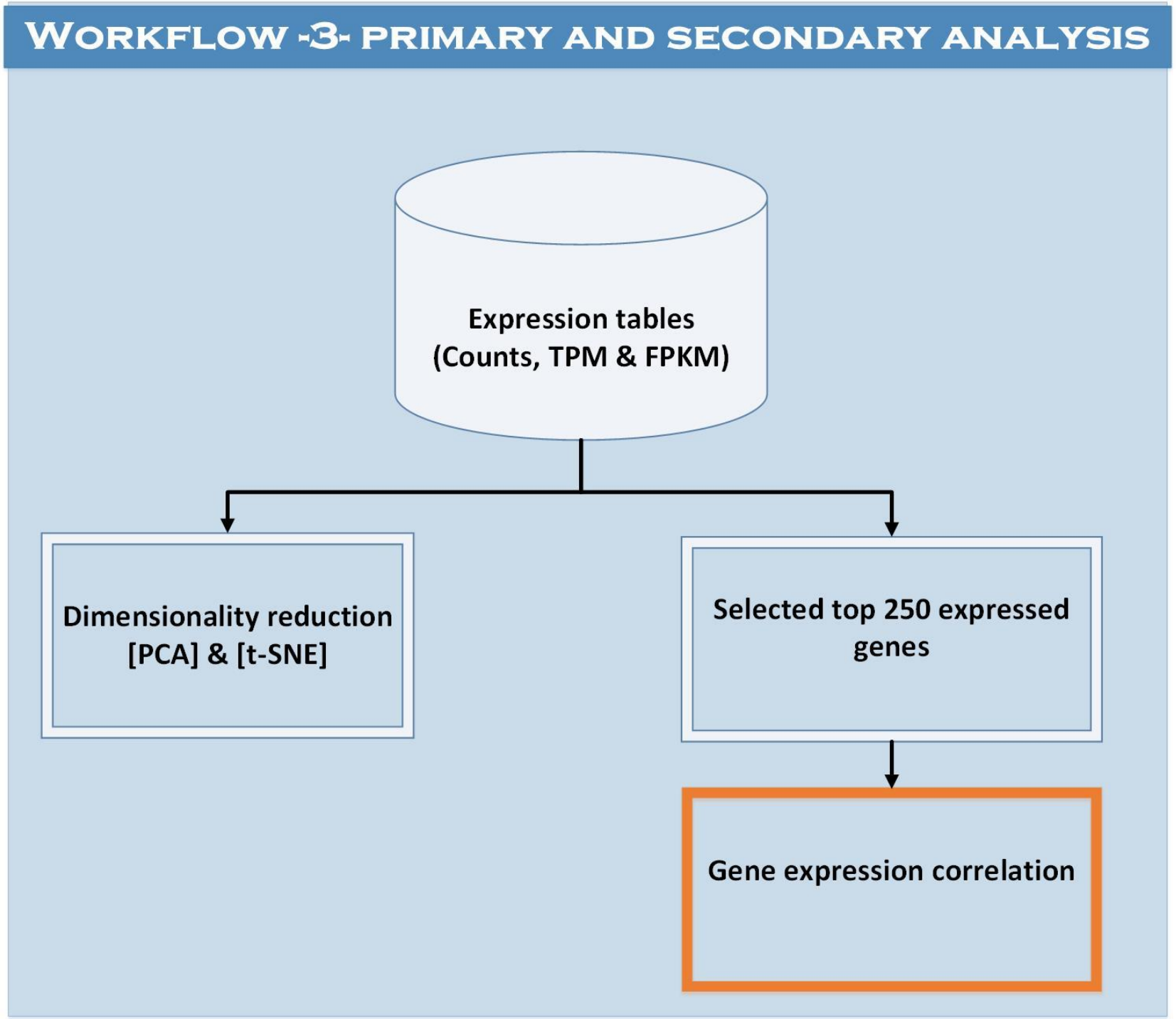
4- Functional annotation [Goterm, KEGG, Pathways]

5 – Database system creation

Workflow -1- QC; Pre-processing; Alignment & Quantification

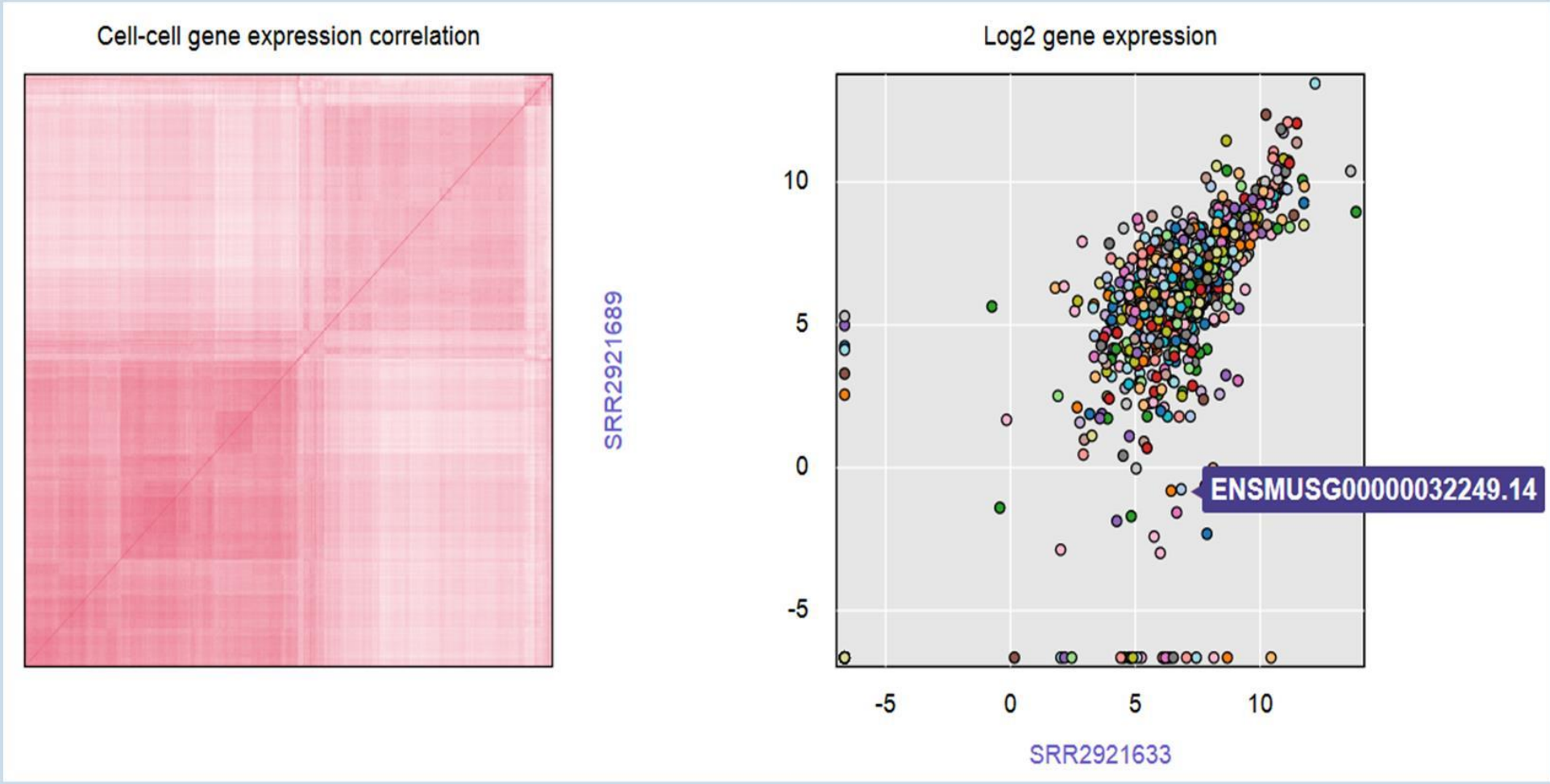


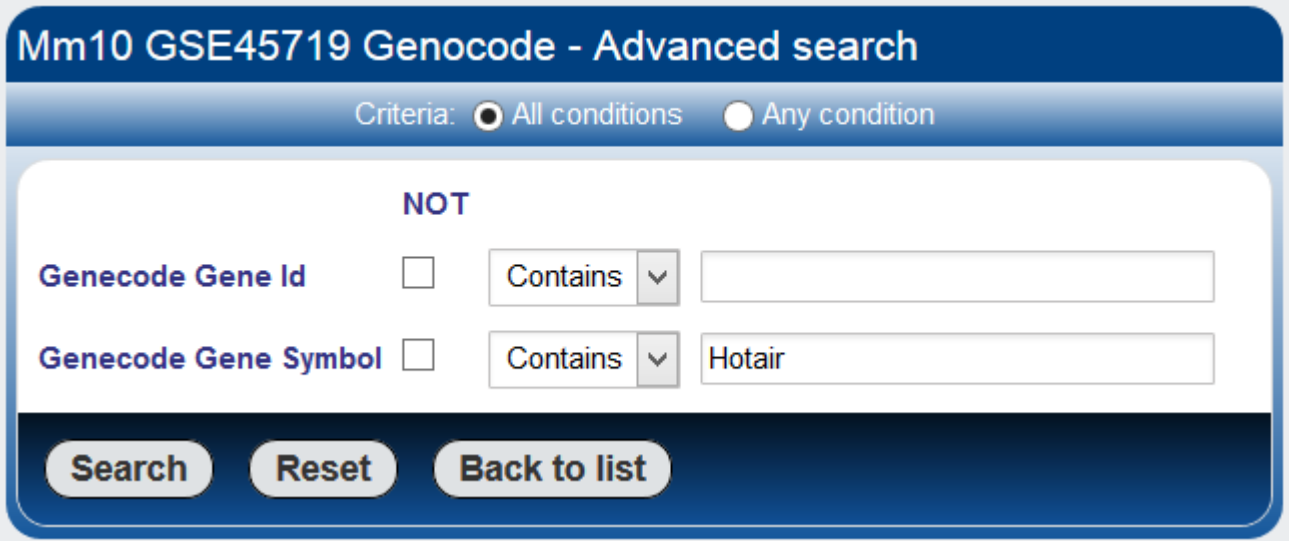
Workflow -3- primary and secondary analysis



Single-cell gene expression correlation

Identification of novel regulators of Th17 cell pathogenicity by single-cell genomics





	Genecode Gene Id	Genecode Gene Symbol
Mm10 GSE45719 Exp Table	ENSMUSG00000086903.4	Hotair
Mm10 GSE45719 Exp Table	ENSMUSG00000087658.2	Hotairm1

Details found: **317** [1 2 3 4 5 6 7 8 9 10] Next : Last Page **1** of **16**

	Gencode Gene Id	Cell Id	Expression
Single-cell sample list	ENSMUSG00000087658.2	SRR1041756	12.63
Single-cell sample list	ENSMUSG00000087658.2	SRR1041755	10.09
Single-cell sample list	ENSMUSG00000087658.2	SRR1041759	8.42
Single-cell sample list	ENSMUSG00000087658.2	SRR1041758	4.36
Single-cell sample list	ENSMUSG00000087658.2	SRR1041763	3.11
Single-cell sample list	ENSMUSG00000087658.2	SRR1041764	3.00
Single-cell sample list	ENSMUSG00000087658.2	SRR1041760	2.75
Single-cell sample list	ENSMUSG00000087658.2	SRR1041749	1.50
Single-cell sample list	ENSMUSG00000087658.2	SRR805386	1.45
Single-cell sample list	ENSMUSG00000087658.2	SRR805412	0.89
Single-cell sample list	ENSMUSG00000087658.2	SRR805445	0.74
Single-cell sample list	ENSMUSG00000087658.2	SRR805291	0.39
Single-cell sample list	ENSMUSG00000087658.2	SRR805288	0.27
Single-cell sample list	ENSMUSG00000087658.2	SRR805426	0.23
Single-cell sample list	ENSMUSG00000087658.2	SRR1041740	0.23
Single-cell sample list	ENSMUSG00000087658.2	SRR805292	0.22
Single-cell sample list	ENSMUSG00000087658.2	SRR805409	0.21
Single-cell sample list	ENSMUSG00000087658.2	SRR805250	0.14
Single-cell sample list	ENSMUSG00000087658.2	SRR805287	0.13
Single-cell sample list	ENSMUSG00000087658.2	SRR805306	0.10

DATA ACQUISITION, PROCESSING AND ANALYSIS OF PUBLISHED SINGLE-CELL [HUMAN & MOUSE]

1- Single-cell data curation and ontology annotation

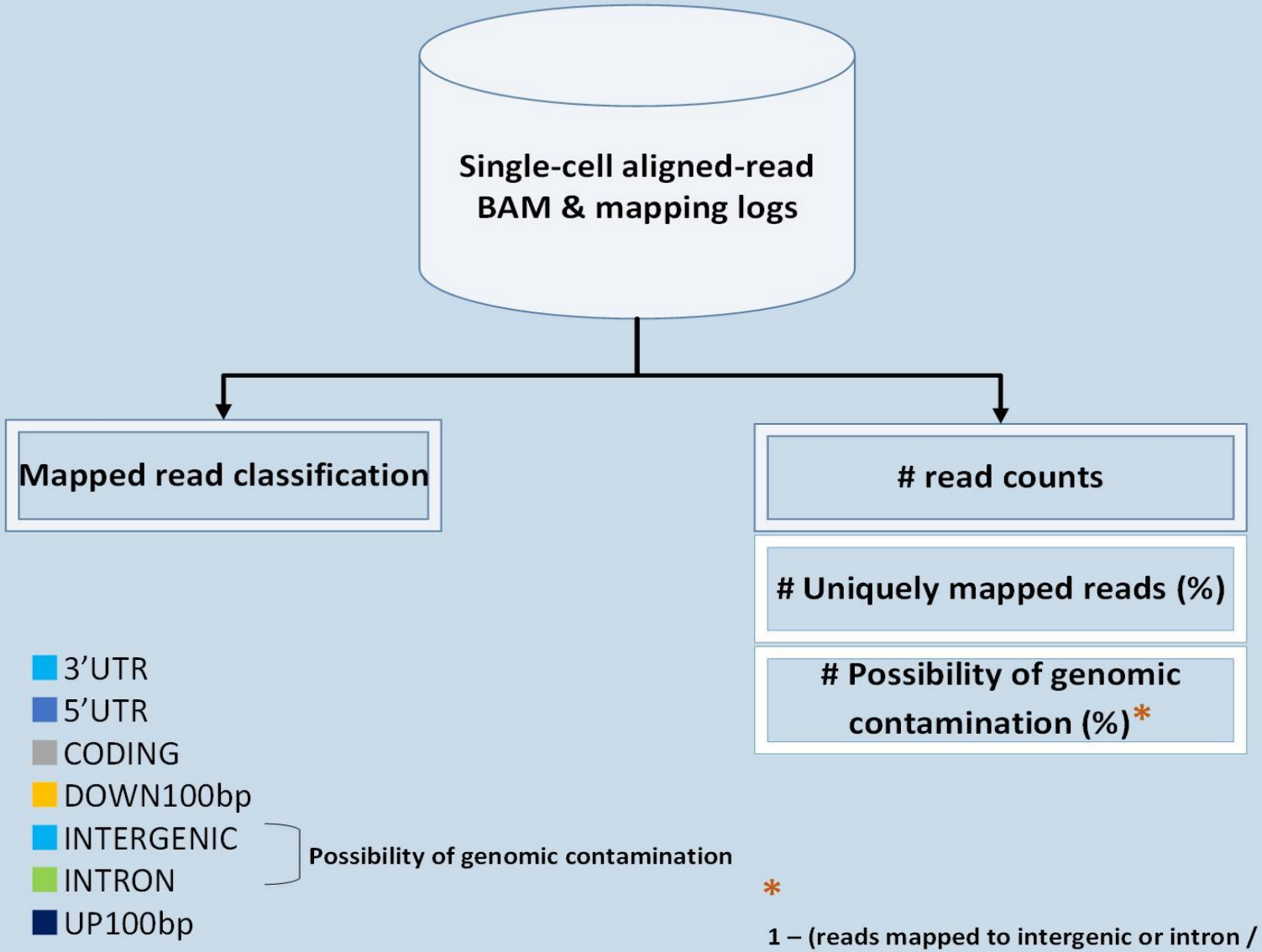
2- Single-cell transcriptomic & genomic analysis

3- Analysis of single-cell genomic contaminations

4- Functional annotation [Goterm, KEGG, Pathways]

5 – Database system creation

Workflow -2- Mapping QC and genomic contamination



DATA ACQUISITION, PROCESSING AND ANALYSIS OF PUBLISHED SINGLE-CELL [HUMAN & MOUSE]

1- Single-cell data curation and ontology annotation

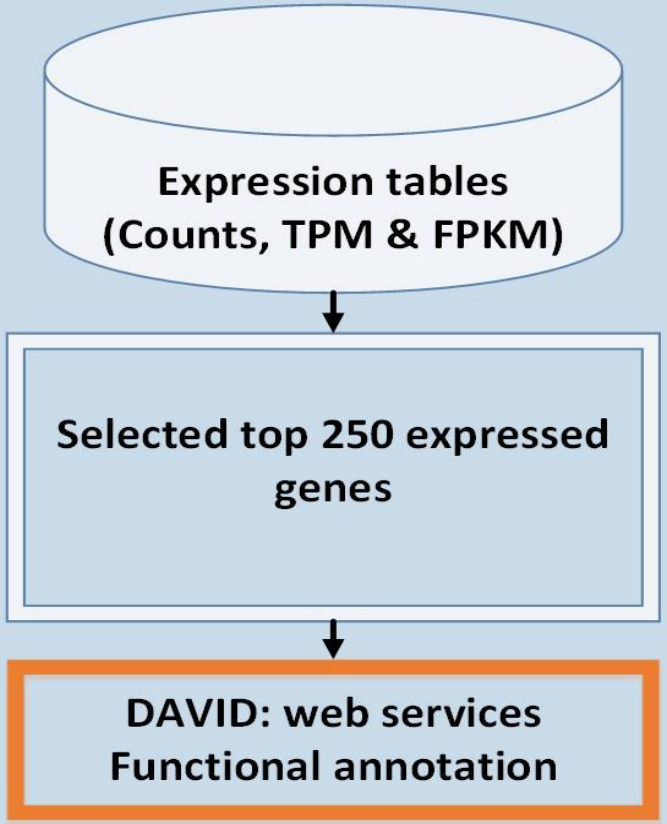
2- Single-cell transcriptomic & genomic analysis

3- Analysis of single-cell genomic contaminations

4- Functional annotation [Goterm, KEGG, Pathways]

5 – Database system creation

WORKFLOW -3- PRIMARY AND SECONDARY ANALYSIS



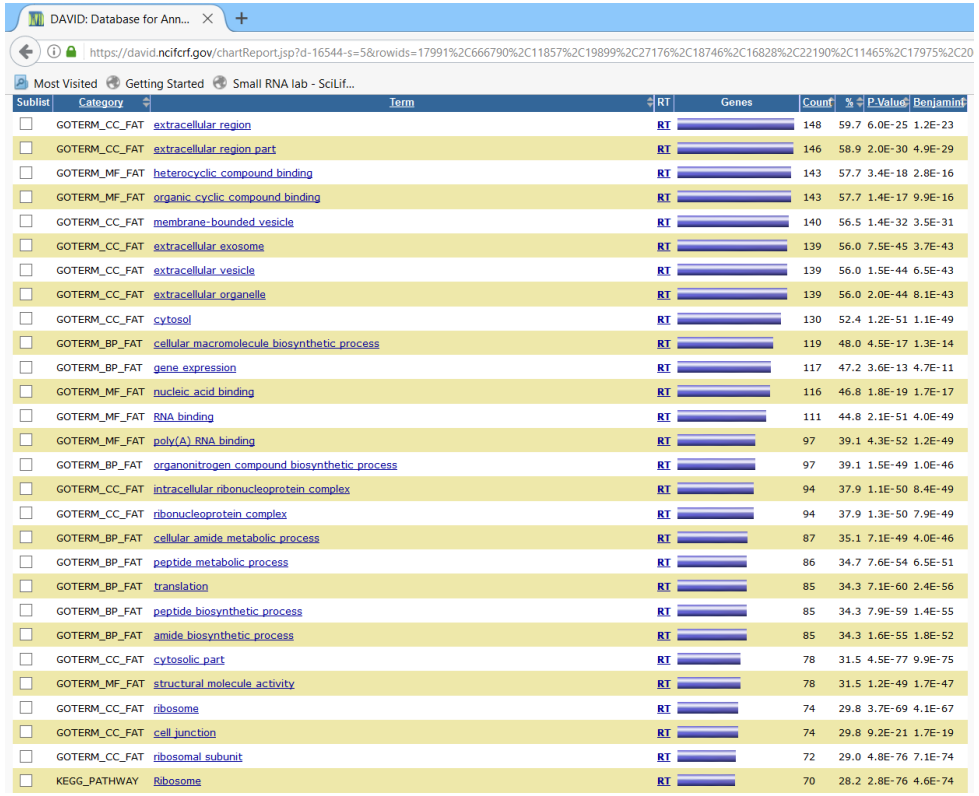
https://david.ncifcrf.gov/

KEGG pathways

Protein sequence analysis & classification

GOTERM

OMIM disease



DAVID: Database for Ann...

https://david.ncifcrf.gov/chartReport.jsp?d-16544-s=5&rowids=17991%2C666790%2C11857%2C19899%2C27176%2C18746%2C16828%2C22190%2C11465%2C17975%2C20

Most Visited Getting Started Small RNA lab - SciLif...

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	GOTERM_CC_FAT	extracellular region	RT		148	59.7	6.0E-25	1.2E-23
	GOTERM_CC_FAT	extracellular region part	RT		146	58.9	2.0E-30	4.9E-29
	GOTERM_MF_FAT	heterocyclic compound binding	RT		143	57.7	3.4E-18	2.8E-16
	GOTERM_MF_FAT	organic cyclic compound binding	RT		143	57.7	1.4E-17	9.9E-16
	GOTERM_CC_FAT	membrane-bounded vesicle	RT		140	56.5	1.4E-32	3.5E-31
	GOTERM_CC_FAT	extracellular exosome	RT		139	56.0	7.5E-45	3.7E-43
	GOTERM_CC_FAT	extracellular vesicle	RT		139	56.0	1.5E-44	6.5E-43
	GOTERM_CC_FAT	extracellular organelle	RT		139	56.0	2.0E-44	8.1E-43
	GOTERM_CC_FAT	cytosol	RT		130	52.4	1.2E-51	1.1E-49
	GOTERM_BP_FAT	cellular macromolecule biosynthetic process	RT		119	48.0	4.5E-17	1.3E-14
	GOTERM_BP_FAT	gene expression	RT		117	47.2	3.6E-13	4.7E-11
	GOTERM_MF_FAT	nucleic acid binding	RT		116	46.8	1.8E-19	1.7E-17
	GOTERM_MF_FAT	RNA binding	RT		111	44.8	2.1E-51	4.0E-49
	GOTERM_MF_FAT	poly(A) RNA binding	RT		97	39.1	4.3E-52	1.2E-49
	GOTERM_BP_FAT	organonitrogen compound biosynthetic process	RT		97	39.1	1.5E-49	1.0E-46
	GOTERM_CC_FAT	intracellular ribonucleoprotein complex	RT		94	37.9	1.1E-50	8.4E-49
	GOTERM_CC_FAT	ribonucleoprotein complex	RT		94	37.9	1.3E-50	7.9E-49
	GOTERM_BP_FAT	cellular amide metabolic process	RT		87	35.1	7.1E-49	4.0E-46
	GOTERM_BP_FAT	peptide metabolic process	RT		86	34.7	7.6E-54	6.5E-51
	GOTERM_BP_FAT	translation	RT		85	34.3	7.1E-60	2.4E-56
	GOTERM_BP_FAT	peptide biosynthetic process	RT		85	34.3	7.9E-59	1.4E-55
	GOTERM_BP_FAT	amide biosynthetic process	RT		85	34.3	1.6E-55	1.8E-52
	GOTERM_CC_FAT	cytosolic part	RT		78	31.5	4.5E-77	9.9E-75
	GOTERM_MF_FAT	structural molecule activity	RT		78	31.5	1.2E-49	1.7E-47
	GOTERM_CC_FAT	ribosome	RT		74	29.8	3.7E-69	4.1E-67
	GOTERM_CC_FAT	cell junction	RT		74	29.8	9.2E-21	1.7E-19
	GOTERM_CC_FAT	ribosomal subunit	RT		72	29.0	4.8E-76	7.1E-74
	KEGG_PATHWAY	Ribosome	RT		70	28.2	2.8E-76	4.6E-74

DAVID Bioinformatics Resources 6.8

National Institute of Allergy and Infectious Diseases (NIAID), NIH

*** Welcome to DAVID 6.8 with updated Knowledgebase (more info). ***

*** If you are looking for DAVID 6.7, please visit our development site. ***

Functional Annotation Table

Help and Manual

Current Gene List: List_1

Current Background: Homo sapiens

93 DAVID IDs

92 record(s)

Download File

4508	ATP synthase F0 subunit 6(ATP6)	Related Genes	Homo sapiens
GOTERM_BP_FAT	purine nucleotide metabolic process, purine nucleotide biosynthetic process, nucleoside phosphate metabolic process, ATP biosynthetic process, phosphorus metabolic process, phosphate-containing compound metabolic process, ion transport, cation transport, hydrogen transport, mitochondrial transport, aging, nucleoside metabolic process, nucleotide metabolic process, ribonucleoside metabolic process, nucleoside monophosphate metabolic process, nucleoside monophosphate biosynthetic process, purine nucleoside monophosphate metabolic process, purine nucleoside monophosphate biosynthetic process, nucleoside triphosphate metabolic process, nucleoside triphosphate biosynthetic process, purine nucleoside triphosphate metabolic process, purine nucleoside triphosphate biosynthetic process, purine ribonucleotide metabolic process, purine ribonucleotide biosynthetic process, ribonucleoside monophosphate biosynthetic process, ribonucleoside monophosphate metabolic process, nucleoside biosynthetic process, nucleotide biosynthetic process, purine ribonucleoside monophosphate metabolic process, purine ribonucleoside monophosphate biosynthetic process, ribonucleoside triphosphate metabolic process, ribonucleoside triphosphate biosynthetic process, purine ribonucleoside triphosphate metabolic process, purine ribonucleoside triphosphate biosynthetic process, ribonucleotide metabolic process, ribonucleotide biosynthetic process, response to abiotic stimulus, monovalent inorganic cation transport, energy coupled proton transport, down electrochemical gradient, ATP synthesis coupled proton transport, proton transport, heterocycle biosynthetic process, aromatic compound biosynthetic process, organophosphate metabolic process, ribose phosphate metabolic process, ion transmembrane transport, nucleobase-containing compound biosynthetic process, response to increased oxygen levels, purine nucleoside metabolic process, purine nucleoside biosynthetic process, ribonucleoside biosynthetic process, mitochondrial ATP synthesis coupled proton transport, ATP metabolic process, purine ribonucleoside metabolic process, purine ribonucleoside biosynthetic process, ribose phosphate biosynthetic process, intracellular transport, cellular localization, establishment of localization in cell, transmembrane transport, nucleobase-containing small molecule metabolic process, response to hyperoxia, response to oxygen levels, purine-containing compound metabolic process, purine-containing compound biosynthetic process, organophosphate biosynthetic process, cation transmembrane transport, inorganic ion transmembrane transport, inorganic cation transmembrane transport, carbohydrate derivative metabolic process, carbohydrate derivative biosynthetic process, nucleoside phosphate biosynthetic process, organonitrogen compound biosynthetic process, glycosyl compound metabolic process, glycosyl compound biosynthetic process, single-organism cellular localization, single-organism intracellular transport, hydrogen ion transmembrane transport, mitochondrial transmembrane transport,		
GOTERM_CC_FAT	extracellular region, mitochondrion, mitochondrial envelope, mitochondrial inner membrane, mitochondrial proton-transporting ATP synthase complex, proton-transporting two-sector ATPase complex, organelle inner membrane, mitochondrial membrane, organelle envelope, envelope, membrane-bounded vesicle, proton-transporting two-sector ATPase complex, proton-transporting domain, extracellular organelle, extracellular region part, mitochondrial part, mitochondrial membrane part, proton-transporting ATP synthase complex, proton-transporting		