

Hackathon Brainstorming

Alán Muñoz

April 25, 2024

Outline

Overview

Current JUMP toolkit

How to organise a hackathon (the fun parts)

Brainstorm session

Overview

Goals

Showcase the tools we have built to approach JUMP data

CytoData Hackathon: Resources, Challenges and Limitations

Collect feedback and ideas on both tools and plans

Current JUMP toolkit

Is it really that hard to access/process JUMP data?

Is it really that hard to access/process JUMP data?

Yes.

Is it really that hard to access/process JUMP data?

- Loads of data
 - ~115k chemical perturbations
 - ~15k genes (KO'd, overexpressed or both).

Is it really that hard to access/process JUMP data?

- Scattered metadata
 - How do we visualise the images from which a profile was produced?
 - Are these two perturbations from the same batch/plate?

Is it really that hard to access/process JUMP data?

- Diverse backgrounds
 - How can biologists with limited coding experience benefit?
 - How can developers build their own tools?

Some user stories of biologists

I have -omics data for X modified gene or chemical compound.

- How do cells affected by X look?

Some user stories of biologists

I have -omics data for X modified gene or chemical compound.

- What else produces similar morphologies?

Some user stories of biologists

I have -omics data for X modified gene or chemical compound.

- What are the distinctive features of X-perturbed cells?

Some user stories of biologists

I have -omics data for X modified gene or chemical compound.

- Can my X be found under a different name?

We preprocess data and publish it via Datasette

jump_rr (JUMP Round-Robin) is our Python tool to perform pairwise computations on profiles or features using the GPU. It also processes metadata to build whole image paths.

Datasette is a Python library to visualise, query and edit databases; it is deployable via WebAssembly (runs on the browser).

How do cells affected by X look?

gallery: Explore all the images with an associated profile.

12 rows where Gene/Compound in ["MYT1", "PROZ"] sorted by rowid

Gene/Compound:	In	-	MYT1, PROZ
- column -		*	*
<input type="button" value="Apply"/>			

[View and edit SQL](#)

This data as [JSON](#), [CSV \(advanced\)](#)

Link	rowid	index	Gene/Compound	External resources	Metadata_JCP2022	Foci 0	Foci 1	Foci 2	Foci 3	Foci 4	Foci 5	Foci 6	Foci 7	Foci 8	
	6	6	5	PROZ	External	JCP2022_805564									
	1055	1055	1054	MYT1	External	JCP2022_804400									
	9671	9671	9670	PROZ	External	JCP2022_805564									
	10704	10704	10703	MYT1	External	JCP2022_804400									
	19344	19344	19343	PROZ	External	JCP2022_805564									
	20390	20390	20389	MYT1	External	JCP2022_804400									
	29042	29042	29041	MYT1	External	JCP2022_804400									
	29521	29521	29520	PROZ	External	JCP2022_805564									

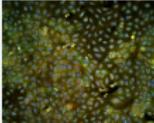
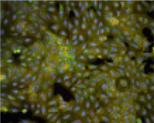
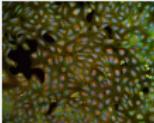
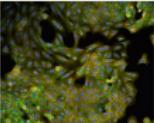
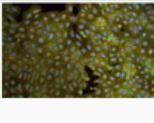
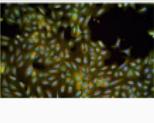
What else produces similar morphologies?

simile: Find the most similar (and dissimilar) perturbations.

home / data

content

50 rows where Gene/Compound = "MYT1" sorted by Similarity descending

Gene/Compound	=	MYT1								
- column -	=									
<input type="button" value="Apply"/>										
View and edit SQL										
This data as json , CSV (advanced)										
Link	rowid	index	Gene/Compound	Match	Gene/Compound Example	Match Example	Similarity ▲	Match resources	JCP2022	
	47551	47551	47550	MYT1	SOS1			0.7195221781730652	External	JCP2022_804400
	47552	47552	47551	MYT1	IL13RA2			0.6855948567390442	External	JCP2022_804400
	47553	47553	47552	MYT1	NLRP12			0.678930401802063	External	JCP2022_804400

[format](#)

What are the distinctive features of X-perturbed cells?

feature: Find the most distinctive features of a perturbation, or amongst all.

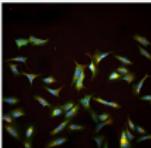
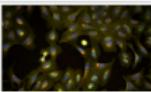
67,700 rows

- column - =

Apply

% [View and edit SQL](#)

This data as [JSON](#), [CSV \(advanced\)](#)

Channel	11	Mask	4							
• RNA	12,350	• Nuclei	19,100							
• Mito	12,050	• Cells	17,500							
• ER	11,750	• Cytoplasm	17,050							
• DNA	11,450	• Image	14,050							
• AGP	11,150									
• ±	5,100									
• OrigDNA	800									
• OrigRNA	800									
• OrigAGP	750									
• OrigER	750									
• OrigMito	750									
Link	rowid	index	Mask	Feature	Channel	Statistic	Gene/Compound	Metadata.image	Median	JCP2022
1	1	0	Cells	Correlation_K_Mito	AGP	8.349928508374835e-29	AVPR2		-0.21309228714749423	JCP2022_905491
2	2	1	Cells	Correlation_K_Mito	AGP	3.325815194945794e-19	EEF1AKMT2		0.13034929387328936	JCP2022_912740

forrest

Does X have an alternative identity?

broad_babel: Translates IDs (JUMP, Symbol/InChiKey, Entrez id, Broad) and which are controls.

home / babel

babel

140,843 rows

- column - =

Apply

[View and edit SQL](#)

This data as [json](#), [CSV \(advanced\)](#)

Link	rowid	JCP2022	NCBI_Gene_ID	standard_key	broad_sample	pert_type
1	1	JCP2022_900002	9	NAT1	ccsbBroad304_00001	trt
2	2	JCP2022_900003	15	AANAT	ccsbBroad304_00002	trt
3	3	JCP2022_900004	18	ABAT	ccsbBroad304_00003	trt
4	4	JCP2022_900005	37	ACADVL	ccsbBroad304_00007	trt
5	5	JCP2022_900006	41	ASIC1	BRDN0001485349	trt
6	6	JCP2022_900006	41	ASIC1	BRDN0001480309	trt
7	7	JCP2022_900008	41	ASIC1	ccsbBroad304_00008	trt
8	8	JCP2022_900007	47	ACLY	ccsbBroad304_00009	trt
9	9	JCP2022_900008	48	ACO1	ccsbBroad304_00010	trt
10	10	JCP2022_900009	52	ACP1	ccsbBroad304_00011	trt
11	11	JCP2022_900010	54	ACP5	ccsbBroad304_00012	trt
12	12	JCP2022_900011	56	ACRV1	ccsbBroad304_00013	trt
13	13	JCP2022_900012	59	ACTA2	ccsbBroad304_00014	trt
14	14	JCP2022_900013	70	ACTC1	ccsbBroad304_00015	trt
15	15	JCP2022_900014	81	ACTN4	ccsbBroad304_00016	trt
16	16	JCP2022_900015	88	ACTN2	ccsbBroad304_00017	trt

format

The tools can be organised based on question and dataset...

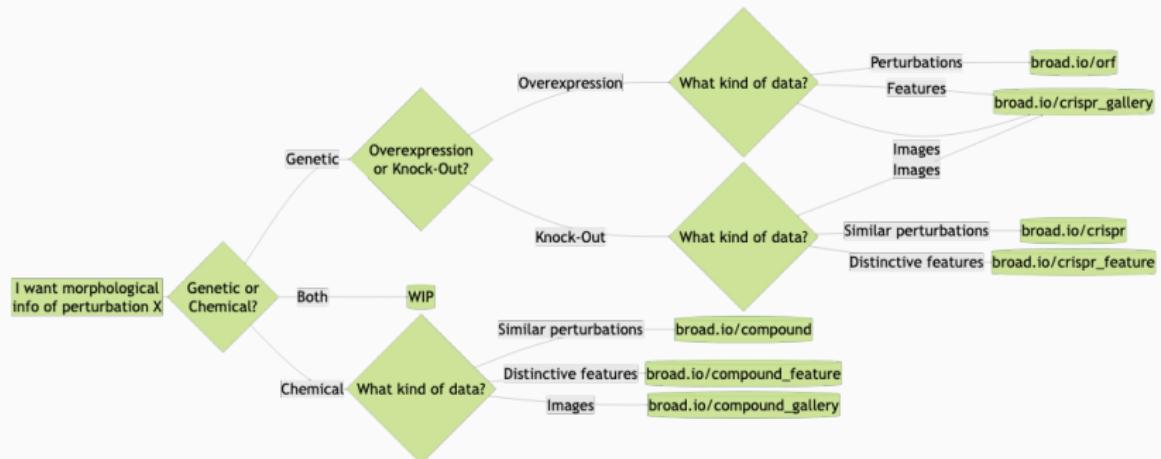
Perturbation	Similar perturbations	Distinctive features	Images
Overexpression	orf	orf_feature	orf_gallery
Knocked-Out	crispr	crispr_feature	crispr_gallery
Compound	compound	compound_feature	compound_gallery

For example:

- broad.io/orf
- broad.io/orf_feature
- broad.io/orf_gallery

(Broad shortlinks are an underused asset. Keep it low-key.)

...or as a decision flowchart



Ok, that's "fun" and all, but what about developer experience?

pandas! pandas! pandas!

The following tools are available via pip.

broad_babel: Three SQL wrappers in a trenchcoat.

Tool to translate its and provides essential information.

- ID goes in, ID(s) come out
- ID goes in, control metadata comes out

It doubles as a central accessible source of ground truth for identities and metadata.

`jump_portrait`: Spice-up your workflow with cell images

Fetch a subset of images associated to a perturbation.
Optionally, include their respective negative controls.

- Foci-level images are accessible, not single-cell.
- Negative controls is essential to account for batch effects.
- Useful to train Deep Learning models on images on-the-fly(?).

Other nice JUMP-adjacent tools

- jump-dti: Fetch data from drug-target interaction (@John)
- cpg-data: Fetch images for any CellPainting data (@Ank)
- SMILES without borders: Cheminformatics black magic (@Srijit)

Aren't you just procrastinating the dive into biology?

Aren't you just procrastinating the dive into biology?

Perhaps.

So, we need biology vignettes

Or at least tutorials/documentation for others to make their own

So, we need biology vignettes

Or at least tutorials/documentation for others to make their own

¿Por qué no los dos?

As a side-note, it would be neat not to suffer an overdose of .ipynb

JUMP central concentrates knowledge, including (battle-proven) code to keep the evergrowing info in a leash.

How-To Guides > Basic JUMP data access

Other Formats
Jupyter

JUMP documentation and examples

Home

How-To Guides

Basic JUMP data access

Explanations

Analyses

Available resources

Search

Basic JUMP data access

This is a tutorial on how to access profiles from the [JUMP Cell Painting datasets](#). We will use polars to fetch the data frames lazily, with the help of `s3fs` and `pyarrow`. We prefer lazy loading because the data can be too big to be handled in memory.

▼ Code

```
import polars as pl
from pyarrow.dataset import dataset
from s3fs import S3FileSystem
```

The shapes of the available datasets are:

- a. `cpg0016-jump[crispr]`: CRISPR knockouts genetic perturbations.
- b. `cpg0016-jump[orf]`: Overexpression genetic perturbations.
- c. `cpg0016-jump[compound]`: Chemical perturbations.

Their explicit location is determined by the transformations that produce the datasets. The aws paths of the dataframes are built from a prefix below:

► Code

We use a S3FileSystem to avoid the need of credentials.

► Code

We will lazy-load the dataframes and print the number of rows and columns

► Code

shape: (3, 5)

dataset	#rows	#cols	#Metadata cols	Size (MB)
str	i64	i64	i64	i64

format

How to organise a hackathon (the fun parts)

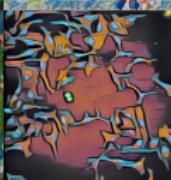
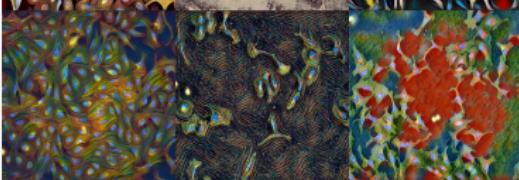
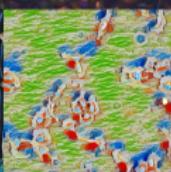
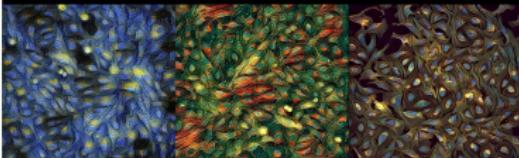
We have a visual!



"JUMP into Cell Painting data" Hackathon for discoveries from microscopy images

SAVE THE DATE! September 17th, 2024

Broad Institute of Harvard and MIT, Cambridge, MA



WikiHow to the rescue

wikiHow to do anything...

wikiHow is where trusted research and expert knowledge come together. Learn why people trust wikiHow

HOSPITALITY AND TOURISM BUSINESSES • EVENT MANAGEMENT

How to Organize a Hackathon

Co-authored by wikiHow Staff
Last Updated: February 26, 2024 [References](#)

A hackathon is typically a weekend-long event, during which computer programmers and code writers come together to solve a coding or programming problem. Hackathons can also occur outside of the programming industry. The word has recently taken on a broader meaning and now includes any group of professionals who come together to address a common problem within their field.^[1] To organize a hackathon, first put together clear and achievable goals, then begin inviting participants, finding sponsors, and arranging a venue, food, and transportation.

Part 1 Determining the Focus and Major Details [Download Article](#)



1 Determining the Focus and Major Details

2 Planning for Comfort and Convenience

3 Facilitating the Hackathon

OTHER SECTIONS

• Tips and Warnings

• Related Articles

• References

Co-authored by:  wikiHow Staff [Edit](#)

Co-authors: \$ Updated: February 26, 2024 Views: 28,655

100% of readers found this article helpful.

Click a star to add your vote!

Quizzes

Am I a Nervousist, or an Empath Quiz [Take Quiz](#)

Does He Like Me Quiz [Take Quiz](#)

What Is My Aesthetic Quiz [Take Quiz](#)

Am I Gay Quiz [Take Quiz](#)

Why Can't I Sleep Quiz [Take Quiz](#)

Zodiac Compatibility Quiz [Take Quiz](#)

You Might Also Like

How to Be a Good Master of [How to Organise a](#)

What are the goals?

Part 1 Determining the Focus and Major Details [Download Article](#)

1 Articulate a clear goal for the hackathon. In order for your hackathon to be a success, you and all of the participants will need to have a clear goal of what's actually being accomplished. Hackathons are a valuable networking resource, but most attempt to solve a specific problem or create new technology. For example, the goal of your hackathon could be:^[2]

- Solving a medical issue related to breathing-related deaths in infants.
- Addressing the need for more food startup companies.



wiki How to Organize a Hackathon



Featured Articles



Trending Articles



What are the goals?

- Promote JUMP (and Imaging) as an additional tool in the scientists' data arsenal
- Pair biologists with questions and computer scientists with who can tackle those questions
- Encourage folks to apply their own methods, knowledge and skills to find out new/better approaches to use morphological profiles

An open-ended hackathon

The data and tools will be available beforehand

Part 3 **Facilitating the Hackathon** [PDF Download Article](#)

The most successful team often receives a monetary reward or a tech item like a new computer or camera. Rules often include time limits: for example, participants may be working in multiple 60-minute time intervals. Also specify to participants what types of technology they are or are not allowed to use.^[12]

- On the other hand, if your event is meant to be collaborative rather than competitive, an elaborate structure of rules may not be necessary.

An illustration showing several people in a workshop or office environment. In the foreground, a person with dark hair is seen from the side, looking at a computer screen. In the background, another person is standing and looking at a document, while others are seated at desks with computers. The scene depicts a collaborative and focused atmosphere typical of a hackathon.

Are more winners more fun?

3

Feature a few small competitions rather than 1 large contest. By multiplying the number of competitions featured during the hackathon, you'll also be multiplying the number of prizes offered and the number of chances to win. This will motivate hackathon participants and draw in participants that might not otherwise be interested.^[13]

- For example, offer a small prize for best innovation using a certain type of technology.
- Of course, the prizes for each contest will be smaller and less expensive than if the hackathon were to only have 1 overall prize.



We are set two independent tracks

Further a biological story

Build a tool that helps data processing

Prizes are more symbolic, we are aiming for collaborative
more than competitive.

How should we structure teams?

Part
3

Facilitating the Hackathon

 Download Article

- 1 Divide participants into effective groups. Most hackathons have many groups of 4-6 individuals working together to solve the problem posed by the hackathon.

Successful groups will blend subject-matter experts (e.g. medical professionals, professional coders, or restaurant managers) with relative newcomers (e.g. high school or college students, and interested amateurs).^[11]

- Avoid allowing groups comprised of only experts and only newcomers, as the newcomers will flounder and struggle to find ways to compete with the experts.



Structuring teams is hard

There are multiple options for forming groups

- Allow some teams to come prepared?
- On-site mixing of computer scientists and biologists
- Both (is it a fair competition to pitch these teams against each other?)

Sponsors are an option

4 Recruit 1 or more sponsors for your event. Ask around in your industry to find at least 1 commercial sponsor. Sponsors can help cover the costs of large hackathons in various ways: through providing a space to host the event, providing food, or simply providing cash. In return, sponsoring companies will expect something in return: perhaps time to address hackathon participants or a booth to sell their products.^[5]

- If the hackathon is a success, your sponsors will be likely to return year after year. If the hackathon continues to expand in size, you can recruit additional sponsors.



An absolute beginners' guide to applying makeup like a pro

How to make gummy bears

How to store cut ginger

How to draw a bunny step by step: easy and realistic methods

A 2x2 grid of article thumbnails. Top-left: A woman looking up, text overlay: 'An Absolute Beginners' Guide to Applying Makeup Like a Pro'. Top-right: A hand holding a glass of milk, text overlay: 'How to Make Gummy Bears'. Bottom-left: A hand holding a small container of ginger, text overlay: 'How to Store Cut Ginger'. Bottom-right: An illustration of a bunny, text overlay: 'How to Draw a Bunny Step by Step: Easy and Realistic Methods'.

Trending Articles

How to search for a word or phrase in a PDF document

What ‘Mid’ means and how to use it (with examples)

A 2x2 grid of article thumbnails. Top-left: A green bar with the text 'Trending Articles'. Top-right: A man looking at a computer screen, text overlay: 'How to Search for a Word or Phrase in a PDF Document'. Bottom-right: A man holding a book, text overlay: 'What ‘Mid’ Means and How to Use It (With Examples)'.

One sponsor has step up

- \$1000 dollars for prize + Ardigen-branded gadgets
- One data analyst and one biologist to help out with organisation

Limitations

- 40-60 maximum room capacity
- ~6.5 real hours for design/coding/analysis
- Wide range of technical and biological expertise

Brainstorm session

Suggestions from your own experiences?

- Things you liked about an event
- Specific considerations for Bio Hackathons (e.g., Allen Institute)

Current elements in the air

- Do we charge a registration fee?
- Should we give preference to registrants of Cytodata/SBI2?
- We have guaranteed \$1k for prizes. We may have up to 8 winners.
- What are attractive prizes that will be palatable to the sponsors and committee?

Any other suggestions?

Links and resources

- Slides:
github.com/afermg/2024_04_hackathon_brainstorm
- Monorepo of Carpenter-Singh Lab: broad.io/monorepo
- JUMP Info central: broad.io/jump