

Science Computing With Science Friends

James Stevenson

2022/04/27

OUTLINE

- Outline
- Friends
- Git
- Workflow
- Style
- Analysis

FRIENDS?

FRIENDS

HOW MANY OF US HAVE THEM

Our computational artifacts should be *accessible* to *all of them*.

KINDS OF COMPUTATIONAL ARTIFACTS:

- Tools
- Analyses

SCIENCE DEMANDS OPENNESS

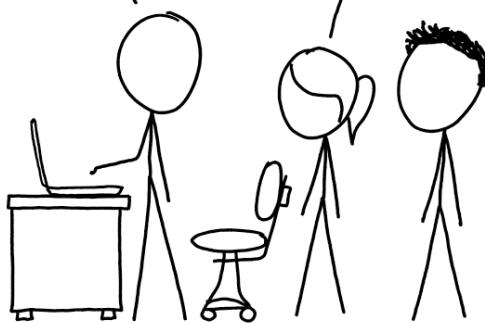
- Show your work (fight the replication crisis)
- If people find you, they can cite you (h-index ^^^^^^)
- Uncle Sam pays your salary, sort of
- It's a commitment (<https://www.big-data-biology.org/software/commitments/>)

GIT/GITHUB

THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.



WHAT'S IT FOR?

- Understand code
- Fix bugs
- Manage tasks
- Share work
- Collaborate with new friends
- Backup in case laptop destroyed by ninjas
- ~other things~

COMMIT HISTORY

Remove trust level where mentioned on DGldb.org #312

Merged

ahwagner merged 4 commits into `master` from `remove_trustlevel`  on Jan 2, 2018

Conversation 0

Commits 4

Checks 0

Files changed 4

Commits on Oct 23, 2017

remove source trust level from FAQ

 kcotto committed on Oct 23, 2017 

remove mention of trust level from news and getting started pages

 kcotto committed on Oct 23, 2017 

Commits on Nov 22, 2017

test html fix

 kcotto committed on Nov 22, 2017 

Commits on Nov 27, 2017

Merge branch 'error_test' into remove_trustlevel

 kcotto committed on Nov 27, 2017 

COMMIT HISTORY

- o  Add drugsatfda progress
- o  Add download method
- o  Stashing progress
- o  update DL and add meta methods
- o  Add analysis nb
- o  Add progress
- o  Add drugs@fda transform progress
- o  Add second class of drugsatfda products
- o  Move to shorter test filenames
- o  Update progress

COMMAND LINE WOWWWW COLORS

```
* 0ee02e0 - (6 days ago) Merge pull request #278 from cancervariants/staging - Kori Kuzma (tag: v0.4.0a2, origin/main)
  |
  * 316b796 - (6 days ago) Merge pull request #282 from cancervariants/hgvs-dup-del-mode-update - Kori Kuzma (HEAD -> staging)
    |
    * d08302e - (6 days ago) Fix typos - korikuzma
    * 4670ec0 - (7 days ago) Add tests for hgvs dup del mode in normalize - korikuzma
    * 8c4a17e - (7 days ago) Update docs + add tests for to_canonical_variation with hgvs_dup_del_mode - korikuzma
    * db5f167 - (8 days ago) Update default characteristics for hgvs dup del mode - korikuzma
    * 4e93432 - (11 days ago) Update docs for hgvs dup del mode enum - korikuzma (origin/issue-279)
    * 89c6c0f - (11 days ago) Add initial work for adding relative_cnv to hgvs dup del mode - korikuzma
  |
  * e7b4b14 - (12 days ago) Merge pull request #276 from cancervariants/issue-269-liftover - Kori Kuzma
    |
    * 7cf399e - (12 days ago) Consistent tests - korikuzma
    * 48a5c9f - (13 days ago) Cleanup adding liftover data + use tlr spdi - korikuzma
    * 5a2eaa3 - (13 days ago) Add more tests for genomic liftover for to_canonical - korikuzma
    * 0e14194 - (13 days ago) Add liftover genomic data to snv base for genomic sub/silent mutation - korikuzma
    * 9787747 - (13 days ago) Increment version - korikuzma
    * d03c8de - (13 days ago) Add initial work for hgvs/spdi liftover for genomic sub/del - korikuzma
  |
  * a348e3f - (2 weeks ago) Merge pull request #275 from cancervariants/staging - Kori Kuzma (tag: v0.4.0a1)
  |
  * 7e20ef5 - (2 weeks ago) Merge pull request #274 from cancervariants/issue-269 - Kori Kuzma
    |
    * cdf3e5a - (2 weeks ago) Return variation in hgvs_to_canonical_variation - korikuzma
    * 6324b8a - (2 weeks ago) Add TO_CANONICAL tag to endpoint - korikuzma
    * 0a50f2e - (2 weeks ago) Merge issue-272 - korikuzma
```

DIFFS

```
% git diff 3b768ea 9ba82f8
```

[github.com/ga4gh/vrs-python/compare
/9ba82f8cc5a..4e9742b](https://github.com/ga4gh/vrs-python/compare/9ba82f8cc5a..4e9742b)

GIT BLAME

WIP: 2-10-22 team call	3 months ago	159	version
		160	type
		161	desc
WIP: CV Descriptor	3 months ago	162	category
add categorical variation descriptor	3 months ago	163	\$ref
		164	desc
WIP: 2-10-22 team call	3 months ago	165	The
add categorical variation descriptor	3 months ago	166	category
		167	\$ref
		168	desc
WIP: 2-10-22 team call	3 months ago	169	The
WIP: 2-10-22 call	3 months ago	170	member
		171	desc
		172	Var
		173	of
		174	type

GIT BLAME

Merged Feat: Initial pass at groupers #84 Changes from all commits ▾ File filter ▾ Conversations ▾ ⚙ ▾

22 server/app/models/drug_claim.rb

```
32      46
33      47      def cleaned_names
34 -      @cleaned_names ||= names.map { |element| element.gsub(/[^\\w_]+/, '') }.to_set
35 +      @cleaned_names ||= names.map { |element| element.gsub(/[^\\w_]+/, '') }.to_set
36      49      end
37      50
38      51      def original_data_source_url
39          @@ -54,6 +68,8 @@ def original_data_source_url
40              base_url
41              when 'GuideToPharmacology'
42                  "https://www.guidetopharmacology.org/GRAC/LigandTextSearchForward?searchString=#"
```

ZENHUB

New Issues	Backlog	Sprint Backlog	In Progress	Review
23 Issues	5 Issues	1 Issue	7 Issues	4 Issues
dgidb-v5 #98 Health check endpoint	dgidb-v5 #24 Standardize workflow and values for drug approval attributes	dgidb-v5 #46 Check and update source citation data	dgidb-v5 #38 Spin up staging box on AWS	dgidb-v5 #21 Move displays to GraphQL endpoints
dgidb-v5 #97 Directionality returns as "0" for both inhibiting and activating	dgidb-v5 #54 JAX-CKB Importer: Fix API breaks		dgidb-v5 #27 Drug attributes	Filter by epic issues
dgidb-v5 #96 Interaction charts	requirement		dgidb-v5 #3 Design a Front-end UI mock-up	epic
dgidb-v5 #95 Add type declarations	dgidb-v5 #52 Refactor DrugBank importer		dgidb-v5 #51 Retrieve NCI data	dgidb-v5 #42 Prototype drug & gene record pages
dgidb-v5 #93 Use env variables for API URI	requirement		dgidb-v5 #69 Results page	frontend requirement
dgidb-v5 #91 Source grouper-provided data to original sources, not VICC normalizer	discussion			dgidb-v5 #34 Search page autocomplete
dgidb-v5 #90 Add post-group routines: source counters, normalize drug types, source trust levels, interaction types	dgidb-v5 #6 Find additional sources for anti-neoplastic interactions and cancer relevant genes			frontend epic
dgidb-v5 #89 Handle non-ASCII characters in grouper routines	discussion		dgidb-v5 #22 Determine filters for searches	dgidb-v5 #41 Prototype results page
dgidb-v5 #88 Query interactions by drug			Move displays to GraphQL endpoints	frontend
dgidb-v5 #86 Theme configuration				
dgidb-v5 #82 Drug indications				

Closed ~25 issues / 33 PRs

ACTIONS

🔒 [jsstevenson / respected_wizard](#) = Private

Code Issues Pull requests ZenHub Actions Projects Wiki

fix: correct type annotations (#12) github-actions #8

Summary Triggered via push 2 days ago Status Success

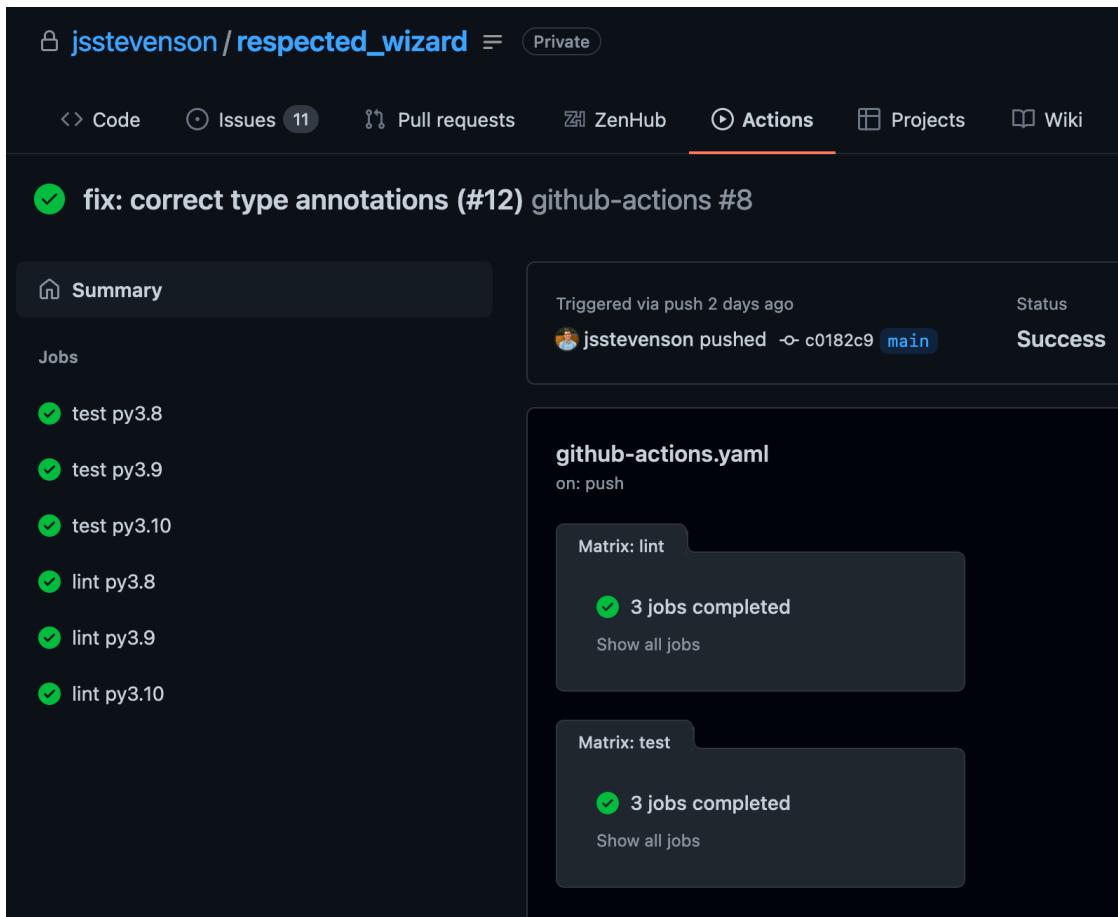
Jobs jsstevenson pushed -o- c0182c9 main

test py3.8 test py3.9 test py3.10 lint py3.8 lint py3.9 lint py3.10

github-actions.yaml on: push

Matrix: lint 3 jobs completed Show all jobs

Matrix: test 3 jobs completed Show all jobs



WHAT DO WE NEED FOR ALL OF THAT ^ ?

Good issues

Good commits

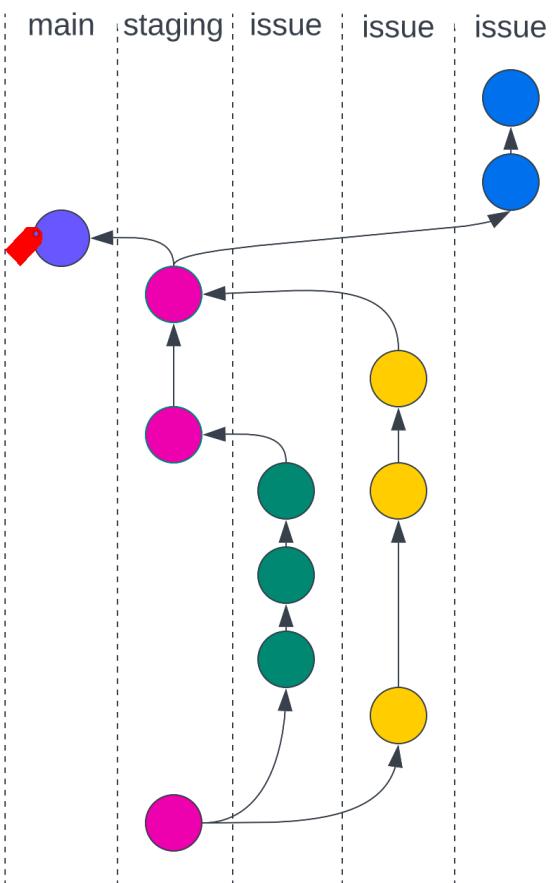
A ban on procrastination

BUT THAT'S HARD!



The Wagner Flow™

HIGH LEVEL



CONSTRAINTS (LIMITS ARE OUR FRIENDS)



Issues create structure

Commits create sub-structure

1) ISSUES

- When in doubt, create them
- Errors, features, stray thoughts
- Either edit or add comments with notes
- Break into new issues where appropriate
- They are extremely googleable

2) CLAIMING AN ISSUE

- Self-assign on GitHub
- Create a branch locally, eg

```
git checkout staging
```

```
git checkout -b issue-42
```

ISSUES: ASK YOURSELF

- Does this issue require tests?
 - Write them first maybe! (TDD)
- Does this issue require documentation?
 - Write it! (probably last)

3) COMMITS

- "should" encompass complete, discrete changes
- brief messages in imperative mood, completes the sentence "This commit will..."
- Can use body/footer for additional details (

```
git commit
```

w/ no

```
-m "message"
```

)

PRE-COMMIT

In projects that use it:

```
pre-commit install
```

Style checks, other security-y things before commits go through

4) PULL REQUESTS

- Your branch → staging
- Most repos have review policies: assign review to maintainers
- Explicate any outstanding questions
- Pay attention to test results in Actions
- When it's ready and approved, merge it yourself
- Manually close issue

LINK PRS TO ISSUES

Add service meta to response #132

Merged

korikuzma merged 1 commit into [staging](#) from [issue-131](#)  on Dec 14, 2021

 Conversation 0

 Commits 1

 Checks 0

 Files changed 4



korikuzma commented on Dec 14, 2021

Closes [#131](#)

Add service meta to response has no dependencies

CI: ACTIONS

Powered by ZenHub

✗ f1713e6

- o  Use uta-tools
- o  korikuzma self-assigned this 25 days ago
- o  korikuzma added 5 commits 23 days ago
 - o  Use uta-tools pypi package 427423f
 - o  Fix positions when going from gnomad_vcf to protein alt ✓ 53d6a52
 - o  Clean up init variables ✓ 015d2da
 - o  Add more tests for gnomad_vcf_to_protein sub ✓ 0b540bd
 - o  await coroutines in main ✓ 0336cb5

ACTIONS

James Stevenson <notifications@github.com>
To: cancervariants/gene-normalization; Cc: Ci activity

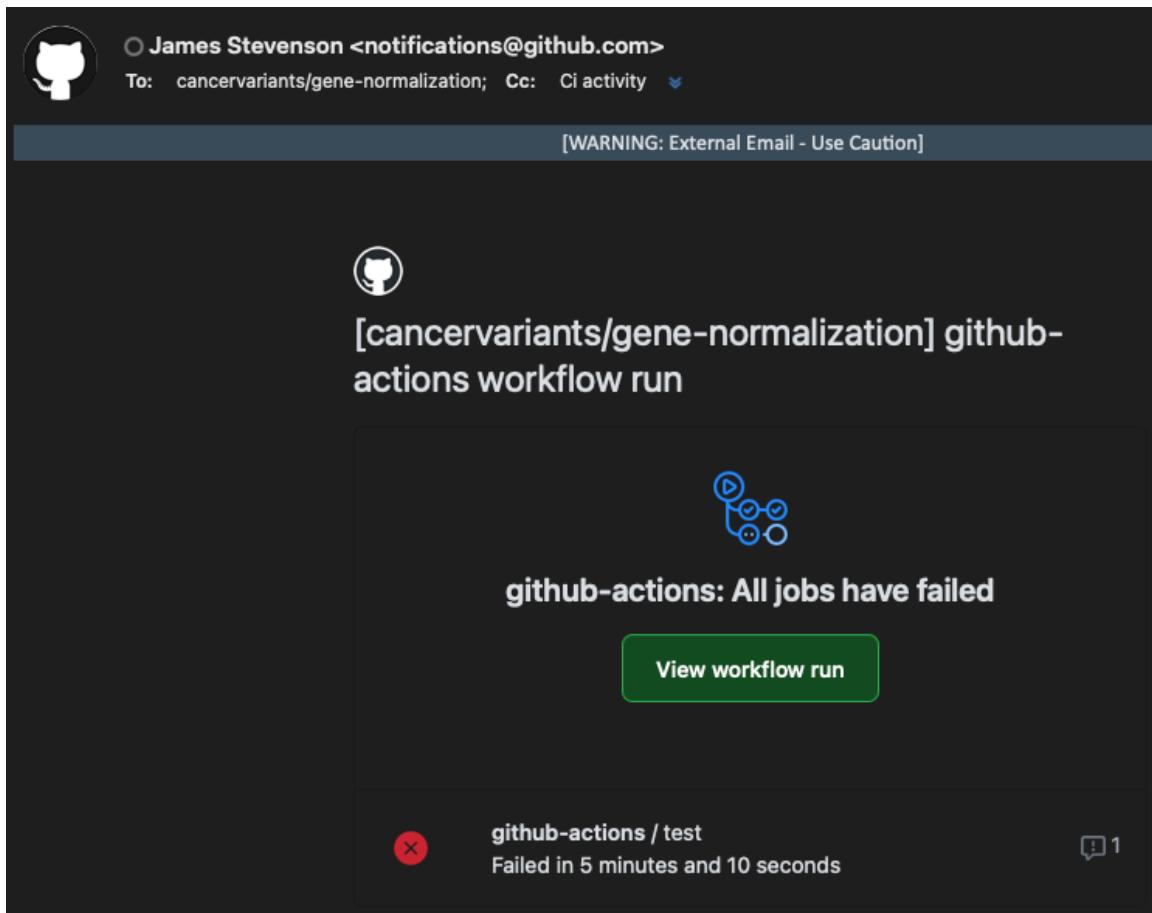
[WARNING: External Email - Use Caution]

[cancervariants/gene-normalization] github-actions workflow run

github-actions: All jobs have failed

[View workflow run](#)

 **github-actions / test**
Failed in 5 minutes and 10 seconds  1



ASIDE: CONVENTIONAL COMMITS

ci(freebsd): enable oldtests	2 months ago
ci: label all markdown edits as "documentation" #18120	2 days ago
chore: fix typos (#17755)	last month
ci: show failed message on Windows	2 months ago
fix(build): missing definitions for sizeof macros #16393	1 hour ago
chore(flake): bump flake (#17632)	2 months ago
docs(man): omit misleading mention of environment for -u NORC (#1...	5 months ago
fix(packaging): remove excess forward slash in Wix Patch (#18121)	2 days ago

type(scope): short description (#pr number)

CONVENTIONAL COMMITS TOOLS

INCLUDE:

- Autogenerate change logs
- Automatically bump consistent with semantic versioning
- (also good template for commits generally, guardrails are our friends)

RELEASES

Use semantic versioning: major.minor.patch

CD: CODE PIPELINE

igm-vicc-TherapyNormalization-dev

Notify ▾

Edit

Stop execution

Clone pipeline

Release change

⌚ Source Succeeded

Pipeline execution ID: [b4de5055-3c6a-4bb0-8788-8be181e11a05](#)

Source



[GitHub \(Version 2\)](#)

⌚ Succeeded - 1 day ago

[ab70af83](#)

[ab70af83](#) Source: feat: add unmerged normalize endpoint (#266) ...

Style



WHY DOES IT MATTER?

- Readability, accessibility
- Stunting
- Have some pride in your craft for god's sake

LINTING

Flake8: serene, minimal, difficult

Black: automated, deterministic, a little quirky

COMMENTS

"Code should be self-documenting"

Use comments to explain **why**, not **what**

PYTHON DOCSTRINGS

```
def add_location_id(self, fusion: Fusion) -> Fusion:  
    """Add `location_id` in fusion object.  
  
    :param Fusion fusion: A valid Fusion object.  
    :return: Updated fusion with `location_id` fields set  
    """
```

Here be monsters

TYPE ANNOTATIONS

The best kind of documentation

Enable static analysis

~required for wagner projects

DATES

Use ISO-8601.

- YYYY-MM-DD
- YYYYMMDD

Analysis

USE VIRTUAL ENVIRONMENTS

pip freeze → generate a lockfile

Sort of a foundation of reproducibility

Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten Simple Rules for Reproducible Computational Research. PLOS Computational Biology 9(10): e1003285.
<https://doi.org/10.1371/journal.pcbi.1003285>

FILE LOCATIONS

Don't hardcode!

- Relative paths from `__main__`
- Environment variables
- CLI arguments
- Software solutions: <https://github.com/cthoyt/pystow>

USE MARKDOWN CELLS

Gene

Summary: The Gene class wraps an external definition of a gene to be used as the subject of a abundance statement. For example, "PTEN del". Gene is not an identifiable class.

Genes are intended to be used to represent systemic variation, not molecular variation.

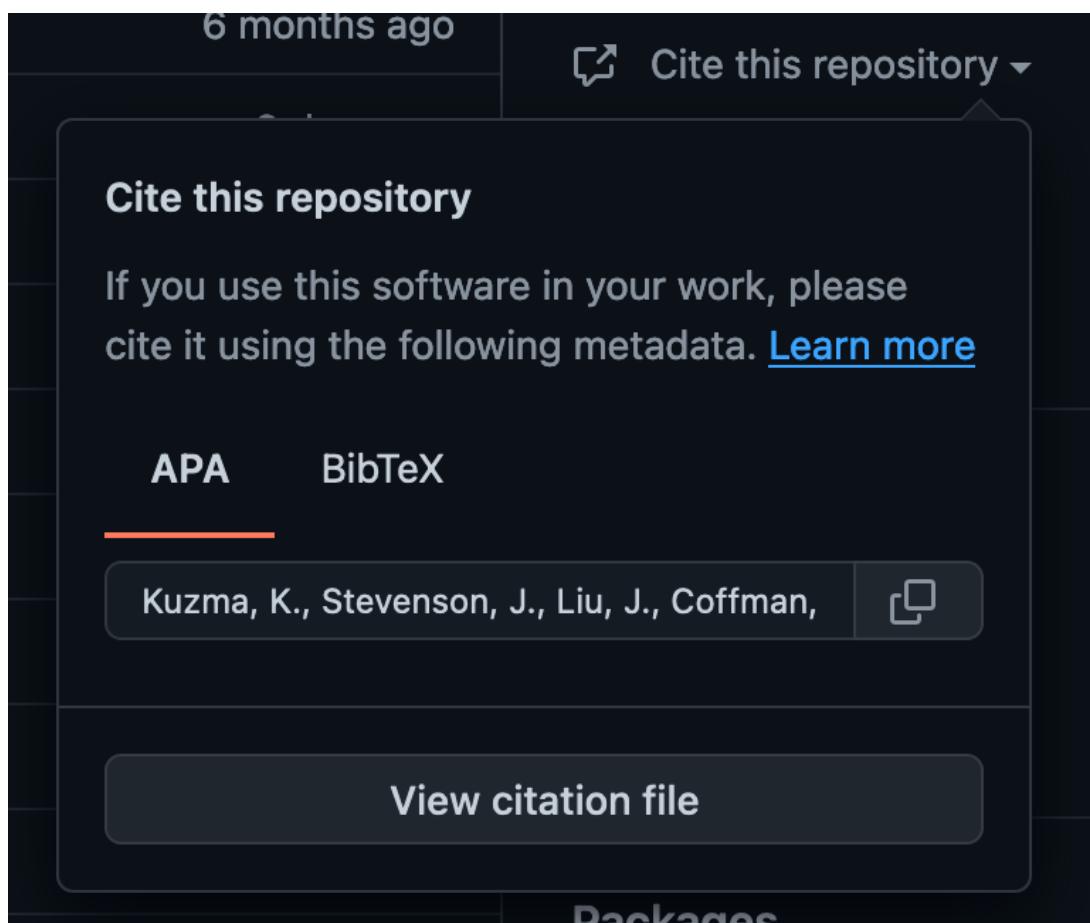
In [5]:

```
g = models.Gene(gene_id="ncbigene:1234")
ppo(g)
```

```
{
    "gene_id": "ncbigene:1234",
    "type": "Gene"
}
```

Maintain readable notes, like a lab notebook

ADD A CITATION TO YOUR REPO



I LIKED THIS EDITORIAL

Rule A, Birmingham A, Zuniga C, Altintas I, Huang SC,
et al. (2019) **Ten simple rules for writing and sharing
computational analyses in Jupyter Notebooks.** PLOS
Computational Biology 15(7): e1007007.
<https://doi.org/10.1371/journal.pcbi.1007007>

I ALSO LIKE THIS TALK

"I Like Notebooks" by Jeremy Howard

PEOPLE LOVE WRITING ABOUT THIS

Good enough practices in scientific computing

Greg Wilson  , Jennifer Bryan  , Karen Cranston  , Justin Kitzes  , Lex Nederbragt  , Tracy K. Teal  

Best Practices for Scientific Computing

Greg Wilson  , D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven H. D. Haddock, Kathryn D. Huff, Ian M. Mitchell, Mark D. Plumley, Ben Waugh, Ethan P. White, Paul Wilson

A Quick Guide to Organizing Computational Biology Projects

William Stafford Noble 

Ten Simple Rules for Effective Computational Research

James M. Osborne  , Miguel O. Bernabeu, Maria Bruna, Ben Calderhead, Jonathan Cooper, Neil Dalchau, Sara-Jane Dunn, Alexander G. Fletcher, Robin Freeman, Derek Groen, Bernhard Knapp, Greg J. McInerny, Gary R. Mirams, [...] , Charlotte Deane

Ten Simple Rules for Reproducible Computational Research

Geir Kjetil Sandve  , Anton Nekrutenko, James Taylor, Eivind Hovig

FURTHER READING

- Wilson et al, "Best Practices for Scientific Computing"
- MIT: "The Missing Semester"
- Luis Coelho lab, "Lab Software Tool Commitments"