

Epilogue

Ok. From this point you should more or less understand status of KATH. I gave you all I got. Take my experience into account and try to avoid my mistakes.

Search for balance between *it works* and *it is perfect* and you will be fine. Last one thing I can left to you is list of tasks I find important for you at this moment.

1. Harvard can explain output from Clinvar API. However, I suggest you to take a look at this problem again and reconsider solution for problem called "download data from CADD". Download Clinvar and merge with LOVD.
2. Remove this cursed pipeline.ipynb file. I think the manual testing is the way to go, people can have it to test locally, but not on main branch for sure. Usage should be described either in Confluence or in Readme file.
3. Resolve imports issue in back_end folder, some files uses relative import and some uses absolute. Use relative and remove dynamic adding or exporting paths.
4. Have a standard for tools files like CADD or SpliceAI. I thought about file with single function that accepts path to csv file with data and some parameters. File can be executed as script or have import from it to use function somewhere else.
5. I think that it would be nice to hide manual merging. So, when user presses downloaded data, he specifies databases and they are automatically downloaded and merged.
6. Routing functions share same behavior, you can extract it and make decorator to save time.
7. Integrate REVEL.
8. From GnomAd you already can retrieve a lot of scores, use this info for merged output.
9. Finish with integration of alignment.
10. Fix Pylint issues.



Good luck!