

Chapter 3. Searching for perfectness

Each database has its own format for saving variants, though Harvard asked to unify it to LOVD, most of the tools work with similar to GnomAd variant format, it is especially useful, when dealing with VCF files (Variant Calling File). Because it can be split in to position, ref and alt according to VCF format.

Downloading of LOVD is pretty easy, data for any gene can be downloaded from static url. For example,

```
1 https://databases-lovd-nl.translate.google.com/shared/genes/download/all/gene/EYS
```

For another gene use different name instead of EYS gene.

However, Integration of GnomAd and Clinvar wasn't that successful. Link for downloading is dynamic. We developed turnaround for this one. We used Python Selenium for going to page and pressing download button. I didn't like that approach, it was literally crutch. GnomAd and Clinvar provide an API for retrieving data which looks like a much better solution. Right? RIGHT?

I did wrong decision and we wasted a lot of time to make it work. Clinvar was not implemented because it was giving so bad output that I don't even want to think about it. It was such a long time ago that it's hard for me to remember exactly, but I'm not even sure if downloading GnomAd function gives correct output, it looks not really understandable and lack some comments for sure.

Roughly same situation was with tests. Originally, we didn't plan software, that's why everything we were implementing in Google Colab notebook. Once we changed our mind, we had to move everything to GitHub and start development in IDE. I kept that file for a while as example for newcomers, it was explaining workflow much better. After some time I decided to move forward and replace that file with normal testing. So, for this movement I had to be sure that my team members know how to write unit test in Python and that they have enough time to do it. We had nothing. Testing in *pipeline.ipynb* was just easier and faster.



I swear to God! I personally deleted *pipeline.ipynb* file three times! Three! And it's still their, maybe even in two copies with different names...

After all, we implemented downloading for LOVD and GnomAd and implemented merging of them.

Not even once say you appreciate me, I deserve respect

I've done my best to give you nothin' less than perfectness

And I know that if I end this, I'll no longer have nothin' left

25 to life - Eminem