Genomic Database Task

۱۴۰۲/۱۰/۱۹

Here we have designed a challenging yet rewarding task that will not only test your technical abilities but also enhance your skills in genomic data handling.

Genomic Database Construction

Your primary objective is to construct a genomic variant database using a set of prepared VCF (Variant Call Format) files, along with the VCF file of ClinVar provided to you. The database should be organized and structured to facilitate efficient retrieval and analysis of genomic variants.

Requirements

1. Research about existing solutions for genomic variant databases. Select two tools of your choice for the construction of the genomic database. Give us a report on why you chose those tools. Consider factors such as ease of use, speed, accuracy, scalability, and any specific features relevant to genomic data integration. (Do step 2 and 3 for both databases and step 4 for just one of them)
2. Integrate the provided VCF files into a unified genomic database.
3. Incorporate the ClinVar VCF file into the database.
4. Provide a user-friendly UI to query for specific variants or regions. (It can be command based; Graphical UI is a plus)

Submission:

1. Database:

- Submit the constructed genomic databases along with a brief documentation outlining its structure and how to query them.

2. Tool Comparison:

- Provide a comparative analysis of the selected tools, highlighting their strengths and weaknesses.
- Include any challenges encountered during the database construction process and how each tool addressed them.

Important Notes:

- You are encouraged to document your workflow and decisions made during the task.
- Please ensure that your submission is well-organized and easily understandable.
- Utilizing LLM assistance is permissible, provided that it is reviewed and acknowledged.

Best of luck!