

## Instruction to Access Controlled Data of the Paper

1. **Understand the data distributions and access privileges:** The raw data sets at individual level for 574 cases in the paper were separately registered under three different dbGaP accessions/projects due to the originations of patient sources shown in Fig. 1. Only having case IDs in dbGaP, the dbGaP serves as master of control access policy for users. The actual sequencing data sets are stored in corresponding study IDs at GDC where users can download after the access privileges are granted by dbGaP, and thereafter automatically synchronized with GDC.

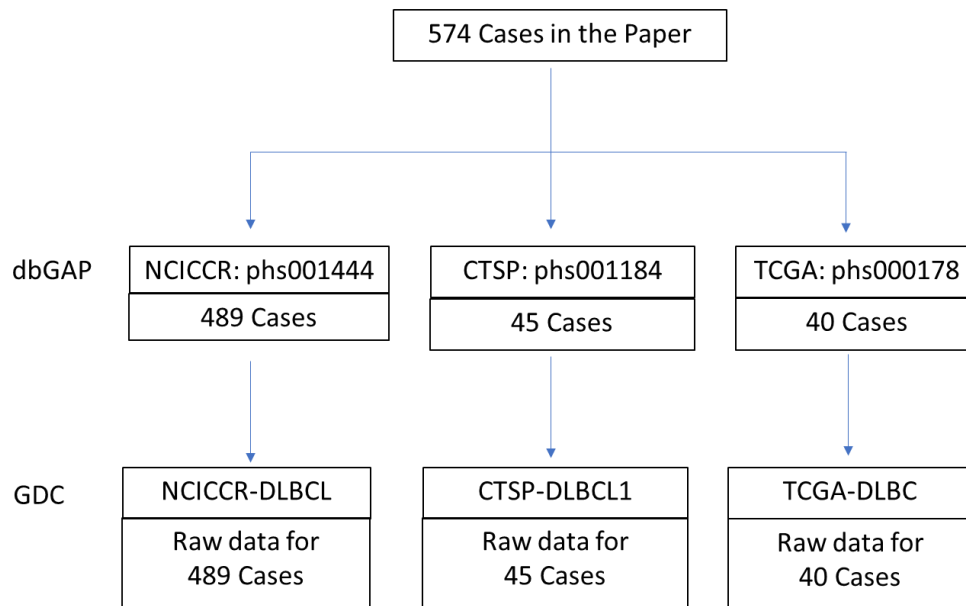


Fig. 1. The distribution of raw sequencing data sets of 574 cases.

2. **Access to the controlled data sets at GDC:** MAF result, Copy Number variation, as well as all raw sequencing data are considered as PHI, and subjected to control access via dbGaP policy.

Step 1: To request access permission from dbGaP for THREE accessions: phs001444, phs001184, phs000178. Importantly, you need three phs accessions in order to access all 574 cases.

Submit your access request by following dbGaP procedure:

<https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login>

Step 2. After permission granted by dbGaP, you can download SNV MAF files and Copy Number Variation files that were used in paper via GDC publication page. Those files are split into three sub files corresponding to the three phs accessions (as shown in Fig. 1) for the

reason of separate control access. You need download every sub files, and combine them into full set of 574 cases.

Step 3: After permission granted by dbGaP, you can download raw sequencing data by following GDC downloading protocol (<https://gdc.cancer.gov/access-data>). In brief:

3.1. Select files and generate manifest. On GDC exploration portal (<https://portal.gdc.cancer.gov/exploration>), select files you want to download. The raw data files of 574 cases are under three places of Program-Project (i.e. NCICCR-DLBCL, CTSP-DLBCL1 and TCGA-DLBC) as shown in Fig. 1. The 574 case names as well as corresponding phs accessions are listed in Table S9 of Supplementary Appendix.xlsx.

3.2. Download raw sequencing files to client computer via gdc-client software. You need firstly install gdc-client software in your computer, and then run gdc-client with the previous manifest file.

### 3. Need help?

- 3.1. For general questions how to use GDC website or gdc-client tools, please contact GDC help desk at <https://gdc.cancer.gov/support>
- 3.2. For general question about the procedure to request access permission from dbGaP, contact dbGaP help desk at <https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=email> or [https://www.ncbi.nlm.nih.gov/books/NBK99228/#DArequest.i have several questions re gar](https://www.ncbi.nlm.nih.gov/books/NBK99228/#DArequest.i_have_several_questions_regar)
- 3.3. For specific data format question about the 574 cases and associated data files, please contact the author of data scientist at [huangdawei@mail.nih.gov](mailto:huangdawei@mail.nih.gov)