NHGRI Data Access Committee dbGaP Activity Report 2019-11-24-2020-12-24

Hoyin Chu, Christopher Steven Marcum

09 December, 2020

The NHGRI Data Access Committee (DAC) currently manages 16492 data access requests (DARs) for access to 5103 projects in dbGaP.

# 1 Data Access Requests

Between 2019-11-24 and 2020-12-24 NHGRI reviewed 7182 DARs. Of these, 5712 were accepted, 631 were downloaded, 2383 had a previous version downloaded, and 1023 were rejected. The average amount of time from when the Principle Investigator (PI) submited a DAR to the final decision by the DAC was 9.3 days. The average time to an accepted decision was 9.4 days, while the average time to a rejected decision was 9.3 days. Figure 1.1 is a barplot comparing the NHGRI DAC to time to final decision to the average across all NIH DACs during the same time interval.

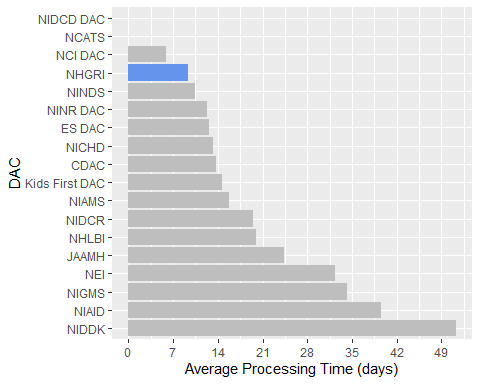


Figure 1.1: Comparison of DAR Processing Time among all DACs

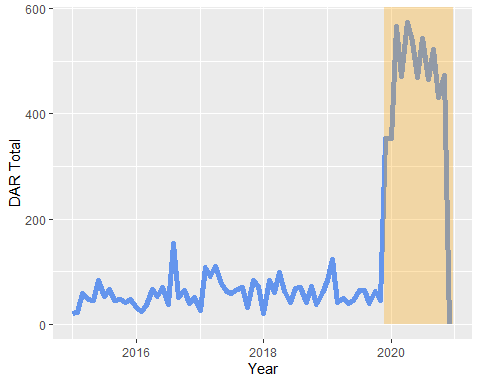


Figure 1.2: Data Access Requests Submitted to NHGRI Per Month Since 2015

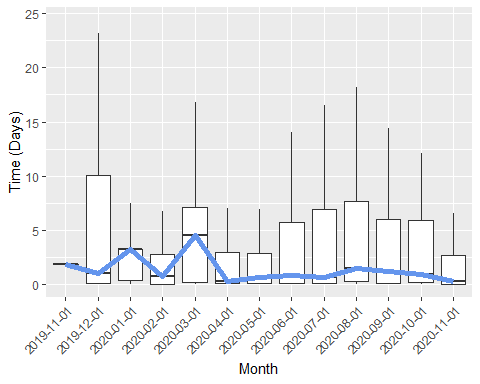


Figure 1.3: DAR Processing Time: From PI Submission to SO Approval

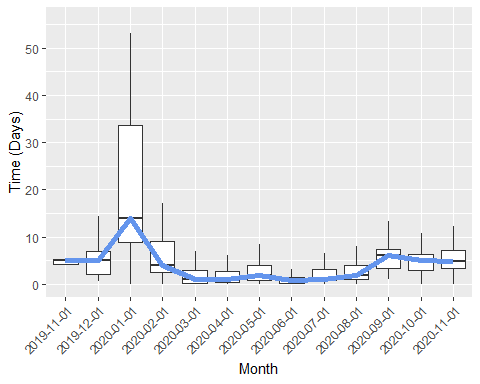


Figure 1.4: DAR Processing Time: From SO Approval to DAC Approval

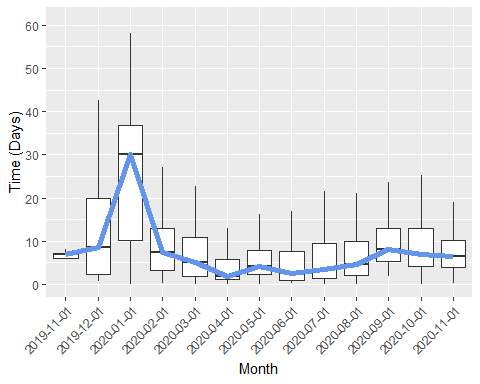


Figure 1.5: DAR Processing Time: From PI Submission to DAC Approval

Between 2019-11-24 and 2020-12-24, 1715 PIs have submitted DAR to studies released by NHGRI. Among these PIs, 1323 have also submitted DAR to studies released by other DACs. PIs who submitted DARs to NHGRI also submits, on average, 19.16 DARs to all DACs for 2.11 projects.

# 2 Study Released

During this reporting period, 23 new studies were released by NHGRI through dbGaP. A total of 748 DARs were made for these newly released studies. Study phs000452.v3.p1 (Melanoma Genome Sequencing Project) was the most requested dataset from the NHGRI DAC with 301 requests. The following word-cloud provides a graphical summary of the combined research use statements (RUS) from all DARs requesting this study. The larger the word appears in the wordcloud, the more frequent the word in the set of RUS.

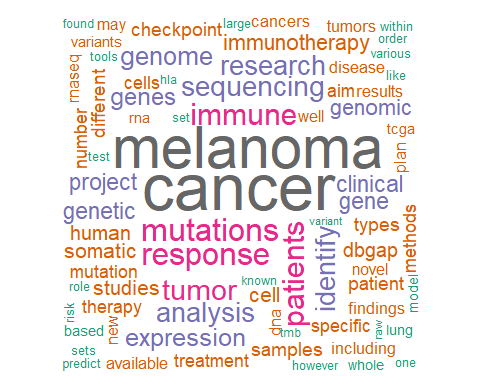


Figure 2.1: Word cloud of Research Statements for Study phs000452.v3.p1 (Melanoma Genome Sequencing Project)

According to Google Scholar, the most cited study during this time period is phs000452.v3.p1 (Melanoma Genome Sequencing Project) with 4 citations.

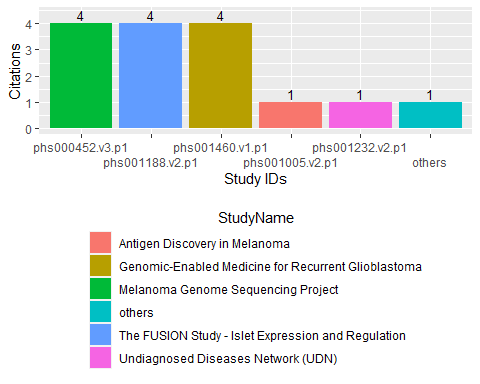


Figure 2.2: Top 5 Studies Managed by NHGRI with Most Google Scholar Citations Between 2019-11-24 and 2020-12-24

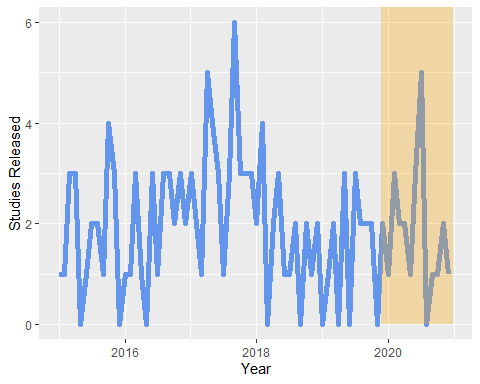


Figure 2.3: Number of Study Released by NHGRI Per Month Since 2015

# 3 Other Comparisons

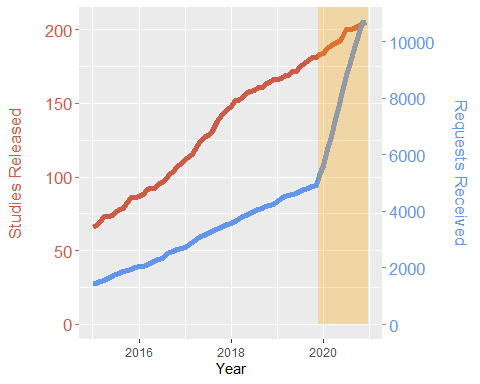


Figure 3.1: Comparison of Cummulative Requests Received and Studies Released by NHGRI

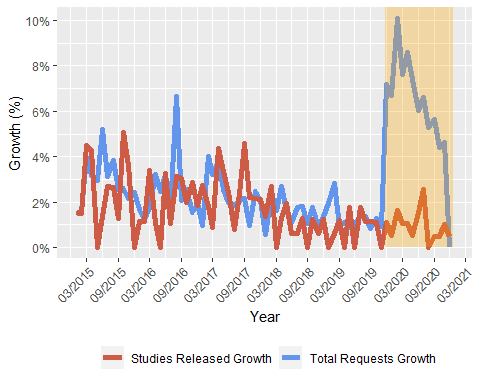


Figure 3.2: Comparison of DAR Growth and Studies Released Growth for NHGRI

This report was prepared using the DACReportingTool package for R, build 0.1.1 by Mr. Hoyin Chu and Dr. Christopher Steven Marcum.