Release Notes

Release 4 (R04):

To see what's new in this Release, please refer to the Release Notes below and GREGoR Dataset characteristics by release version. Additionally, the GREGoR Consortium Data Coordinating Center maintains a list of errata for Release 4 at https://gregorconsortium.org/data/release-notes/r04-errata

Release Date: October 2025

AnVIL Workspaces:

- AnVIL_GREGoR_R04_GRU
- AnVIL_GREGoR_R04_HMB

Summary of changes since previous release:

- R04 adds an additional 1852 participant IDs to the GREGoR Dataset, and retires 9. In total, R04 includes a total of 10,683 participants.
- A subset of aligned DNA short-read files were uniformly processed by the GREGoR Data Coordinating Center (DCC). The R04 dataset contains these harmonized, aligned DNA short-read files in addition to the aligned DNA short-read files separately processed by GREGoR Research Centers. The `aligned_dna_short_read_id` for each harmonized file begins with "GREGoR_DCC_A1" and can be found in the `aligned_dna_short_read` table.
- Jointly called multisample VCFs (split by chromosome) and single-sample genomic VCFs (gVCFs) are available for DCC harmonized data . These VCFs contain genotype data for single-nucleotide variants and short indels and can be found in the `called_variants_dna_short_read` table.
- R04 fixes the following errors identified in R03:
    - The families Broad_BON_B17-01 and Broad_BON_B19-52 have multiple participant_ids for the same individual. This was entered erroneously because participants in these families have multiple experiment types. This issue was corrected by retiring former participant IDs, as listed in this table.
    - 38 nanopore VCF files from R03 were truncated and should not be used; these files are not linked in R04.