

Release Notes

Release 2 (R02):

To see what's new in a 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 2 at <https://gregorconsortium.org/data/release-notes/r02-errata>

Release Date: November, 2024

AnVIL Workspaces:

- [AnVIL_GREGoR_R02_GRU](#)
- [AnVIL_GREGoR_R02_HMB](#)

Summary of changes since previous release:

1. R02 adds data for an additional 4882 participants as well as long-read data (Nanopore and PacBio) to the GREGoR Dataset.
2. The R02 version of GREGoR Data Model includes data tables that support long-read data (Nanopore and PacBio) and genetic findings reported by GREGoR Research Centers.
3. A subset of aligned DNA short-read files were uniformly processed by the GREGoR Data Coordinating Center (DCC). The R02 dataset contains these reprocessed, aligned DNA short-read files instead of the aligned DNA short-read files separately processed and uploaded by GREGoR Research Centers. The `aligned_dna_short_read_id` for each reprocessed file begins with "GREGoR_DCC_A1".
4. Multi-sample VCFs (split by chromosome) and single-sample genomic VCFs (gVCFs) derived from DCC reprocessed aligned DNA short-read files are available in the R02 release. These VCFs contain genotype data for single-nucleotide variants and short-indels and can be found in the `called_variants_dna_short_read` table.
5. The following participants were present in R01 and removed in the R02 dataset due to sex and ped check failures:
 - BCM_Subject_BH15330_2
 - BCM_Subject_BH16155_1
 - BCM_Subject_BH16330_4
 - GSS795357