



## Release Notes

### Release 3 (R03):

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 3 at <https://gregorconsortium.org/data/release-notes/r03-errata>

**Release Date:** July 2025

#### AnVIL Workspaces:

- [AnVIL\\_GREGoR\\_R03\\_GRU](#)
- [AnVIL\\_GREGoR\\_R03\\_HMB](#)

#### Summary of changes since previous release:

- R03 adds an additional 1,446 participants as well as short-read ATAC-seq data to the GREGoR Dataset. In total, R03 includes a total of 8,840 participants.
- The R03 version of GREGoR Data Model includes data tables that support short-read ATAC-seq data.
- A subset of aligned DNA short-read files were uniformly processed by the GREGoR Data Coordinating Center (DCC). The R03 dataset contains these reprocessed, aligned DNA short-read files instead of the aligned DNA short-read files separately processed by GREGoR Research Centers. The `aligned_dna_short_read_id` for each reprocessed file begins with "GREGoR\_DCC\_A1".
- Jointly called, multisample VCFs (split by chromosome) and single-sample genomic VCFs (gVCFs) are available for DCC reprocessed aligned DNA short-read files. These VCFs contain genotype data for single-nucleotide variants and short indels and can be found in the `called_variants_dna_short_read` table.
- R03 fixes the following errors identified in R02:
  - The subject UW\_CRDR-sub-1443008 has been removed from the R03 dataset as it was inappropriately included in the R02 GRU release consent group.
  - R03 corrects metadata for family PMGRC-1067.
  - A small subset of participants were present in the dataset R02 with multiple, unique participant\_ids. These were entered erroneously because the participant has multiple tissue (analyte) or experiment types. We identified this issue in R02 release for the following families:
    - Broad\_BON\_B09\_24

- Broad\_BON\_B11-09
- Broad\_BON\_B16-87
- Broad\_BON\_B17-01
- Broad\_BON\_B17-54
- Broad\_BON\_B17-77
- Broad\_BON\_B17-97
- Broad\_BON\_B17\_15
- Broad\_BON\_B18-37
- Broad\_BON\_B18-97
- Broad\_BON\_B19-52
- Broad\_BON\_B20\_32
- Broad\_BON\_UC35
- Broad\_BON\_UC355
- Broad\_BON\_UC522
- Broad\_HK079
- Broad\_HK080
- Broad\_HK081
- Broad\_HK085
- UW\_CRDR-fam-1590041
- UW\_CRDR-fam-1590067

R03 corrects this issue by retaining one unique participant\_id for each individual and mapping both analyte\_ids to this participant\_id. The file [R03\\_remapped\\_participant\\_ids.tsv](#) has a table with the mapping of participant\_ids retired in R03 to new participant ids.