# GREGoR Data Summary

### Release R05

## Table of contents

### Overview

This report summarizes participant, family, phenotype, experiment, and research-center-level pipeline information from the GREGoR Dataset for Release **R05**.

**Consent codes**

- **GRU** — General research use and clinical care
- **HMB** — Health/medical/biomedical research and clinical care

Table 1: Number of participants, families, and probands in the GREGoR Dataset by consent.

| release | consent_code | participants | families | probands |
|---|---|---|---|---|
| R05 | GRU | 9219 | 3847 | 3741 |
| R05 | HMB | 3073 | 1182 | 1172 |

| release | consent_code | participants | families | probands |
|---|---|---|---|---|
| R05 | Total | 12292 | 5029 | 4913 |

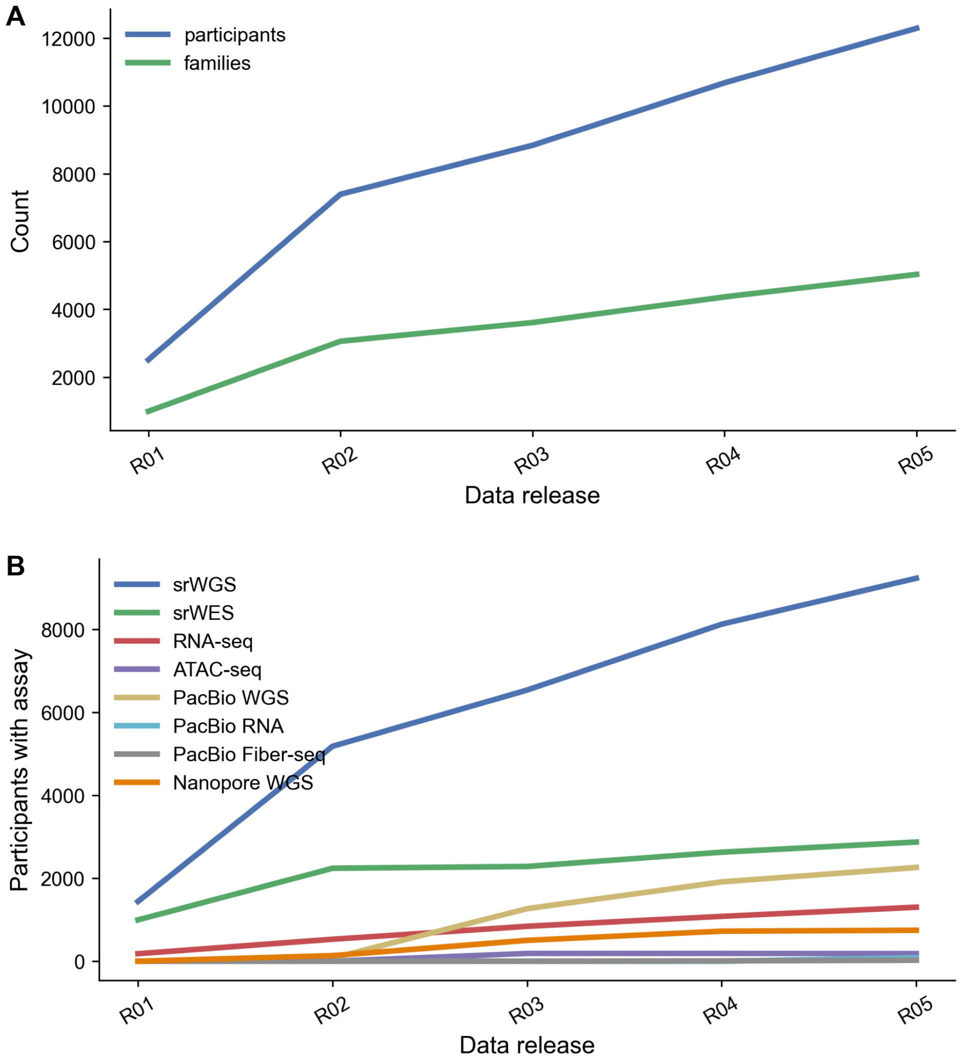

Figure 1: Growth of the GREGoR Dataset across releases. (A) Cohort size including participants and families. (B) Growth of participants with -omics data by assay type. Participant, family, and assay counts are aggregated across all consent groups.

Table 2: Number of participants with short-read and long-read experiments. Participant counts are aggregated across consent groups

| experiment | participants |
|---|---|
| Short-read DNA | 11358 |
| Short-read RNA | 1303 |
| Short-read ATAC | 186 |
| Long-read DNA (Nanopore) | 748 |
| Long-read DNA (PacBio) | 2261 |
| Long-read RNA (PacBio) | 85 |
| PacBio fiber-seq | 24 |

## GREGoR participant and family summaries

This section summarizes family structure and participant ascertainment across consent groups in the GREGoR Dataset for release **R05**.

Table 3: Summary of the sex field in the participant table.

| sex | n_participants | pct |
|---|---|---|
| Male | 6149 | 50.0% |
| Female | 6091 | 49.6% |
| Unknown | 52 | 0.4% |

Table 4: Summary of the reported_race field in the participant table.

| reported_race | n_probands | pct |
|---|---|---|
| White | 2827 | 57.5% |
|  | 1317 | 26.8% |
| Asian | 306 | 6.2% |
| Middle Eastern or North African | 245 | 5.0% |
| Black or African American | 171 | 3.5% |
| American Indian or Alaska Native | 14 | 0.3% |
| Asian\|White | 13 | 0.3% |
| Native Hawaiian or Other Pacific Islander | 9 | 0.2% |
| Middle Eastern or North African\|White | 3 | 0.1% |

| reported_race | n_probands | pct |
|---|---|---|
| Black or African American\|White | 3 | 0.1% |
| Native Hawaiian or Other Pacific Islander\|White | 2 | 0.0% |
| American Indian or Alaska Native\|White | 1 | 0.0% |
| American Indian or Alaska Native\|Asian\|Native Hawaiian or Other Pacific Islander\|White | 1 | 0.0% |
| Black or African American\|Native Hawaiian or Other Pacific Islander | 1 | 0.0% |

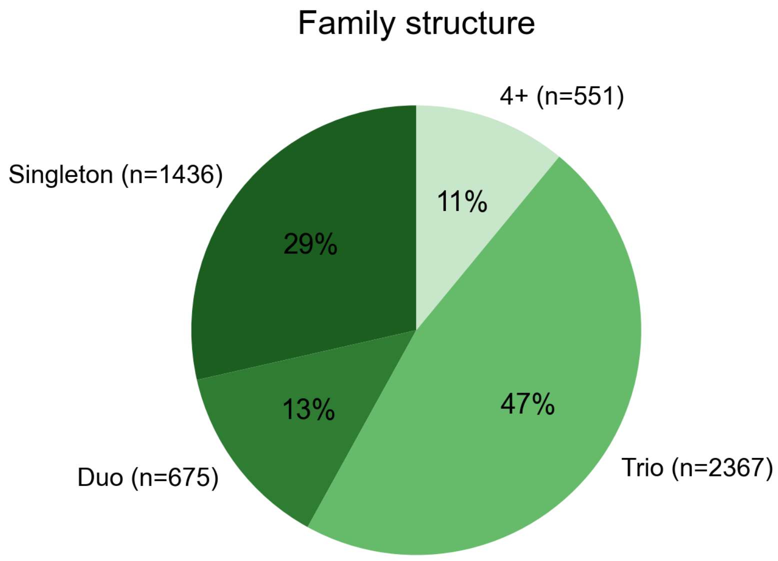

Figure 2: Pie chart summary of family structure in the GREGoR Dataset.

Table 5: Summary of family structure in the GREGoR Dataset.

| family_structure | n_families | pct_families |
|---|---|---|
| Singleton | 1436 | 28.5% |
| Duo | 675 | 13.4% |
| Trio | 2367 | 47.1% |
| 4+ | 551 | 11.0% |

## Solve status

This section summarizes solve_status for probands in the GREGoR Dataset for release **R05**.

Table 6: Summary of solve status for probands in the GREGoR Dataset.

| solve_status | n_probands | pct |
|---|---|---|
| Unsolved | 4130 | 84.1% |
| Solved | 612 | 12.5% |
| Probably solved | 142 | 2.9% |
| Partially solved | 28 | 0.6% |
| Unaffected | 1 | 0.0% |

## Genetic Findings

This section summarizes genetic findings across consent groups in the GREGoR Dataset for release **R05**.

Table 7: Overview of the genetic findings table in GREGoR Dataset.

| total_participants | n_genetic_findings | participants_with_findings | participants_without_findings |
|---|---|---|---|
| 12292 | 2265 | 1615 | 10677 |

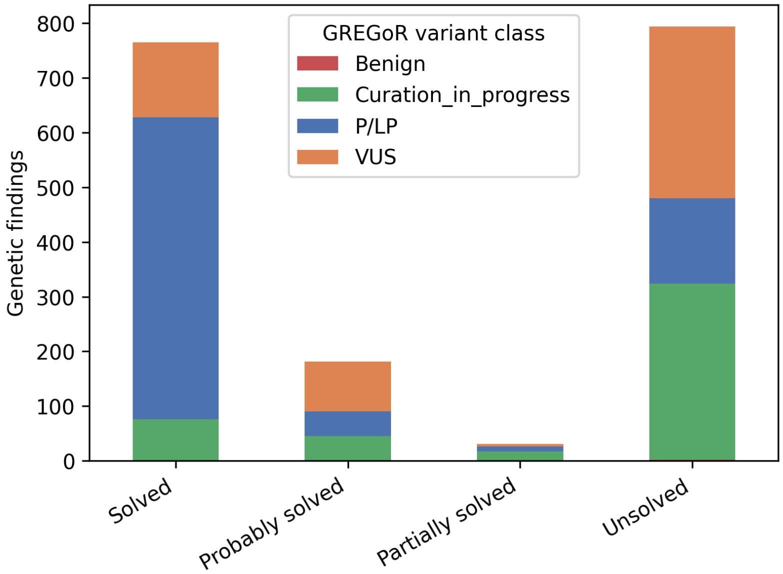

Figure 3: Overview of genetic findings by solve status and GREGoR variant classification. VUS high, moderate, and low categories are collapsed into a single VUS category.

Table 8: Summary of variant_type in genetic findings table.

| n_findings | count | percent |
|---|---|---|
| SNV | 1731 | 76.4% |
| INDEL | 353 | 15.6% |
| SV | 133 | 5.9% |
| RE | 23 | 1.0% |
| CNV | 23 | 1.0% |
| STR | 1 | 0.0% |
| MEI | 1 | 0.0% |

Table 9: Summary of gene_for_known_phenotype category in genetic findings table.

| n_findings | count | percent |
|---|---|---|
| Known | 1238 | 54.7% |
| Candidate | 1027 | 45.3% |

## Phenotype summaries

This section summarizes phenotype-related attributes represented in the current release.

Table 10: Summary of affected status in the GREGoR Dataset.

| category | n_participants | pct |
|---|---|---|
| Proband | 4913 | 40.0% |
| Affected family member | 819 | 6.7% |
| Possibly affected family member | 47 | 0.4% |
| Unknown family member | 379 | 3.1% |
| Unaffected family member | 6134 | 49.9% |

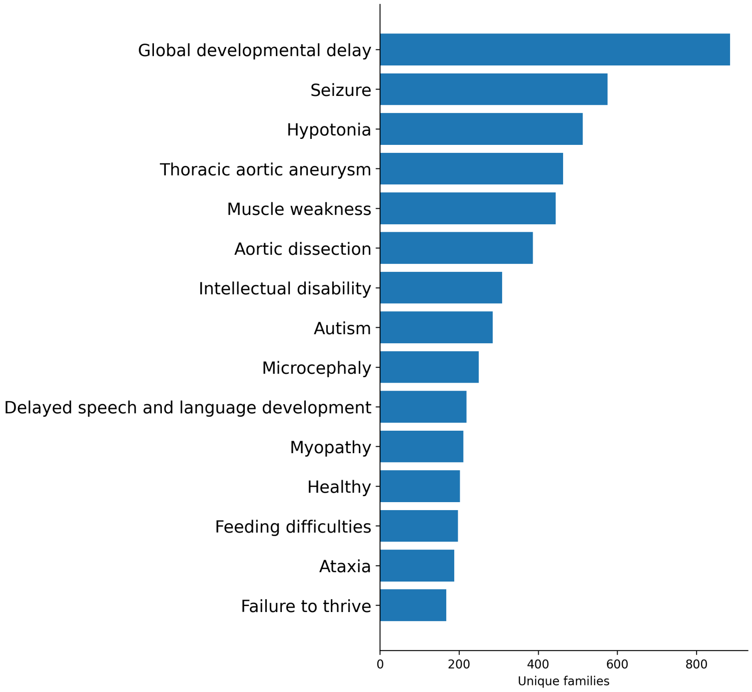

Figure 4: Most common HPO terms in the GREGoR Dataset.

## Experiment summaries by GREGoR Research Center

This section summarizes available -omics data by consent and, where applicable these are separated by experiment type.

### Short-read DNA

Table 11: Overall summary of short-read DNA sequencing experiments by experiment type.

| experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|
| genome | 9271 | 9259 | 9230 |
| exome | 2872 | 2872 | 2872 |
| targeted | 18 | 18 | 18 |

Table 12: Summary of short-read DNA sequencing experiments by consent and experiment type.

| consent_group | experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|---|
| GRU | Total | 9083 | 9071 | 9042 |
| GRU | exome | 1907 | 1907 | 1907 |
| GRU | genome | 7159 | 7147 | 7118 |
| GRU | targeted | 17 | 17 | 17 |
| HMB | Total | 3078 | 3078 | 3078 |
| HMB | exome | 965 | 965 | 965 |
| HMB | genome | 2112 | 2112 | 2112 |
| HMB | targeted | 1 | 1 | 1 |

n_records: total number of experiment records (rows)
n_analytes: number of unique analytes
n_participants: number of unique participants

### Short-read RNA

Table 13: Summary of short-read RNA sequencing experiments by consent and experiment type.

| consent_group | experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|---|
| GRU | Total | 1171 | 1164 | 1150 |
| GRU | paired-end | 552 | 552 | 538 |
| GRU | paired-end\|untargeted | 619 | 612 | 612 |
| HMB | Total | 154 | 153 | 153 |
| HMB | paired-end | 55 | 54 | 54 |
| HMB | paired-end\|untargeted | 98 | 98 | 98 |
| HMB | untargeted | 1 | 1 | 1 |

### Short-read ATAC-seq

Table 14: Summary of short-read ATAC-seq experiments by consent and experiment type.

| consent_group | experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|---|
| GRU | Total | 186 | 186 | 186 |
| GRU | genome | 186 | 186 | 186 |

### Long-read PacBio

Table 15: Summary of long-read PacBio sequencing experiments by consent and experiment type.

| consent_group | experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|---|
| GRU | Total | 2736 | 2376 | 2356 |
| GRU | fiberseq-genome | 19 | 19 | 19 |
| GRU | genome | 2619 | 2259 | 2258 |
| GRU | isoseq | 38 | 38 | 19 |
| GRU | masseq | 60 | 60 | 60 |
| HMB | Total | 20 | 20 | 14 |
| HMB | fiberseq-genome | 5 | 5 | 5 |
| HMB | genome | 3 | 3 | 3 |
| HMB | isoseq | 12 | 12 | 6 |

### Long-read Nanopore

Table 16: Summary of long-read Nanopore sequencing experiments by consent and experiment type.

| consent_group | experiment_type | n_records | n_analytes | n_participants |
|---|---|---|---|---|
| GRU | Total | 729 | 729 | 729 |
| GRU | genome | 729 | 729 | 729 |
| HMB | Total | 19 | 19 | 19 |
| HMB | genome | 19 | 19 | 19 |

## Overlapping assays

This section summarizes participant overlap across assay modalities.

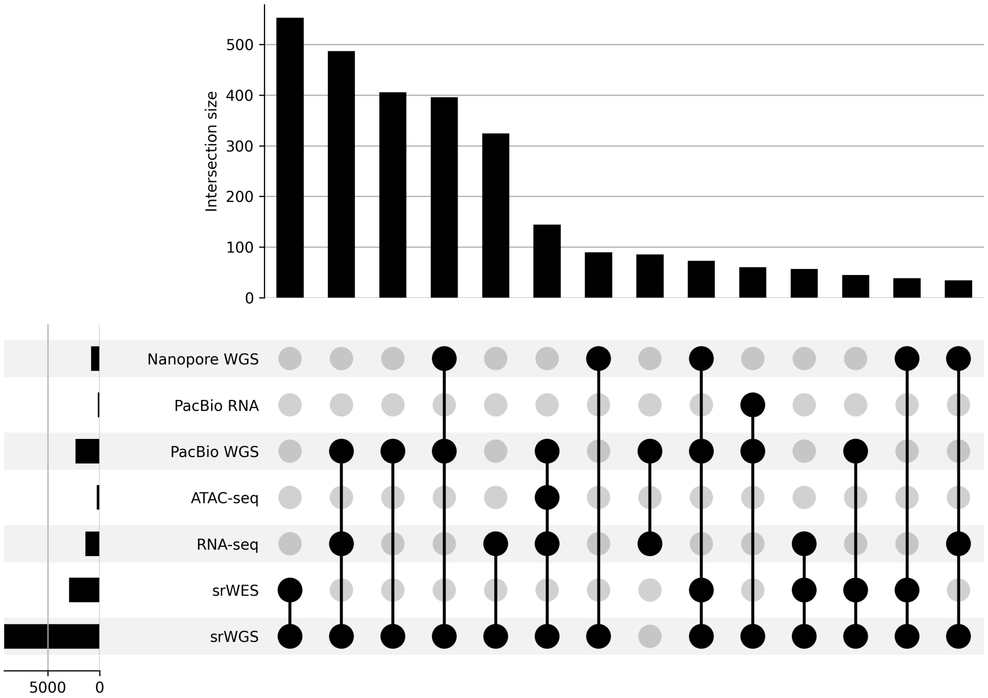

Figure 5: UpSet plot showing the number of participants with overlapping multi-omic assays.

Table 17: Number of participants with overlapping multi-omic assays.

| n_participants | n_assays | assay_combination |
|---|---|---|
| 551 | 2 | srWGS + srWES |
| 485 | 3 | srWGS + RNA-seq + PacBio WGS |
| 404 | 2 | srWGS + PacBio WGS |
| 394 | 3 | srWGS + PacBio WGS + Nanopore WGS |
| 323 | 2 | srWGS + RNA-seq |
| 143 | 4 | srWGS + RNA-seq + ATAC-seq + PacBio WGS |
| 88 | 2 | srWGS + Nanopore WGS |
| 84 | 2 | RNA-seq + PacBio WGS |
| 71 | 4 | srWGS + srWES + PacBio WGS + Nanopore WGS |
| 59 | 3 | srWGS + PacBio WGS + PacBio RNA |
| 55 | 3 | srWGS + srWES + RNA-seq |
| 43 | 3 | srWGS + srWES + PacBio WGS |

| n_participants | n_assays | assay_combination |
|---|---|---|
| 37 | 3 | srWGS + srWES + Nanopore WGS |
| 33 | 3 | srWGS + RNA-seq + Nanopore WGS |