GREGoR R03 Data Summary

GREGoR DCC

2025-06-11

Contents

Overview . . . 1
Summary of solve status for probands in the GREGoR Dataset . . . 2
GREGoR participant and family summaries . . . 3
Phenotype Summaries . . . 4
Experiment Summaries . . . 5
Summary of genetic findings in the GREGoR Dataset . . . 8

List of Figures

1 Overview of the GREGoR Dataset . . . 2
2 Pie chart of family structure. . . . 4
3 Common phenotypes in the GREGoR Combined Consortium Dataset . . . 5
4 Participants with WGS and WES data in the GREGoR Dataset . . . 6
5 Multi-omic data in the GREGoR Dataset . . . 7
6 Multi-omic data for probands in the GREGoR Dataset . . . 8

Overview

This report provides data summaries for the third release of the GREGoR Dataset(R03) which is available on AnVIL. Graphical and tabular summaries of participant, family, experiment, and phenotype information are generated from information provided by member Research Centers (RCs) and uploaded to AnVIL data tables using the GREGoR data model (https://github.com/UW-GAC/gregor_data_models).

Abbreviations:
RCs:
BCM = Baylor College of Medicine Research Center
BROAD = Broad Institute
UCI = University of California, Irvine
GSS = GREGoR Stanford Site
UW-CRDR = University of Washington Center for Rare Disease Research
Consent codes:
GRU = General research use and clinical care
HMB = Health/medical/biomedical research and clinical care

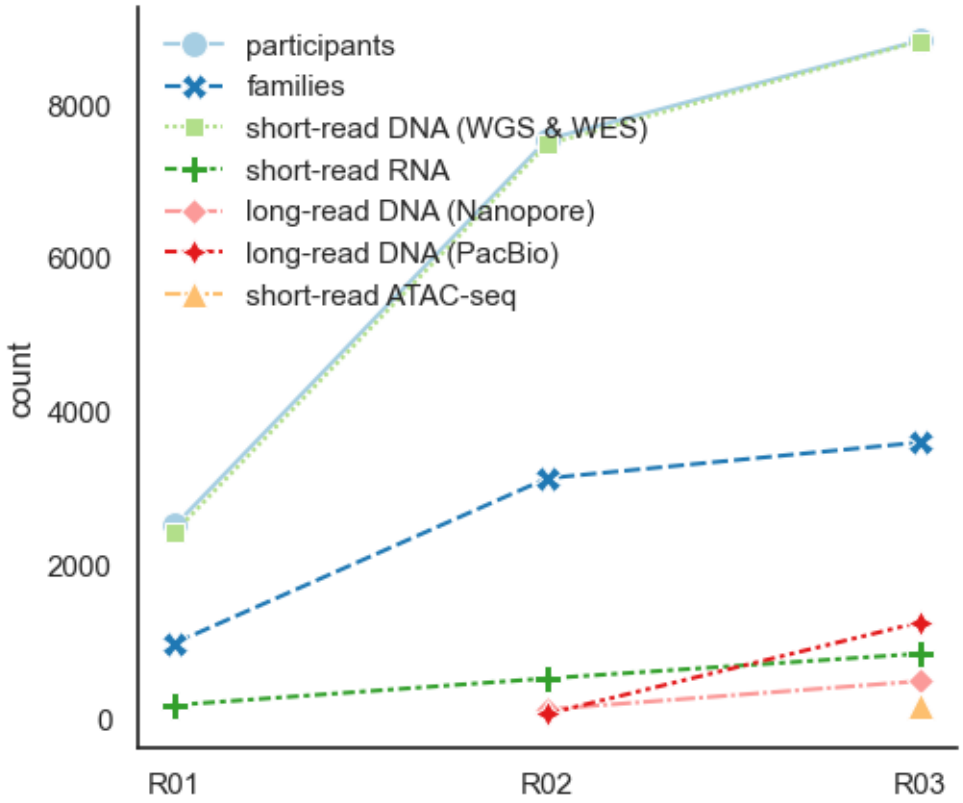

Figure 1: Overview of the GREGoR Dataset across data releases.

Table 1: The number of participants, families and experiments in the GREGoR Dataset

	Number of entries
Participants	8840
Families	3610
Short-read dna (wgs & wes)	8829
Short-read rna	860
Long-read dna (nanopore)	503
Long-read dna (pacbio)	1269
Short-read atac-seq	189

Summary of solve status for probands in the GREGoR Dataset

Table 2: Summary of solve status for probands in the GREGoR Dataset

	No. of probands	%
Partially solved	22	0.01
Probably solved	115	0.03

	No. of probands	%
Solved	472	0.13
Unaffected	1	0
Unsolved	2904	0.83

Solve status definitions:
Solved:

Dominant: Pathogenic/Likely Patthogenic (P/LP) variant with matching inheritance pattern in a gene that also matches the phenotype where there is at least evidence of moderate gene-disease validity (with at least 1 prior publication or preprint or submission to GenCC by any submitter, including GREGoR center)

Recessive: Biallelic P/LP variants with good phenotype and inheritance mode match in gene with at least moderate evidence of gene-disease validity. Can include cases where phase is unknown if the phenotype match is strong (otherwise downgrade to probably solved)

Dual diagnosis/blended phenotype: Can include cases where some components of the phenotype are not explained, particularly phenotypes that may be non-Mendelian or familial

Partially (phenotype) solved: P/LP variant(s) with matching inheritance pattern in a gene with at least moderate gene-disease validity that only accounts for part of the phenotype (i.e. a P variant to explain hearing loss in a patient with hearing loss and intellectual disability); If multiple partial solves are discovered that together explain the majority of the phenotype, the case would be considered solved

Probably solved: i.e. high chance that causal variants have been identified but need more support to reach P/LP
Unsolved: includes cases with a low or moderate candidate listed in the genetics findings table; these are cases where full analysis effort should still be put forth
Unaffected: Any unaffected participant

GREGoR participant and family summaries

Table 3: The number of participants and families in the GREGoR Dataset by GREGoR Research Center and Consent group

GREGoR Center	Participants	Families
GRU	6628	2738
HMB	2212	872
Total	8840	3610

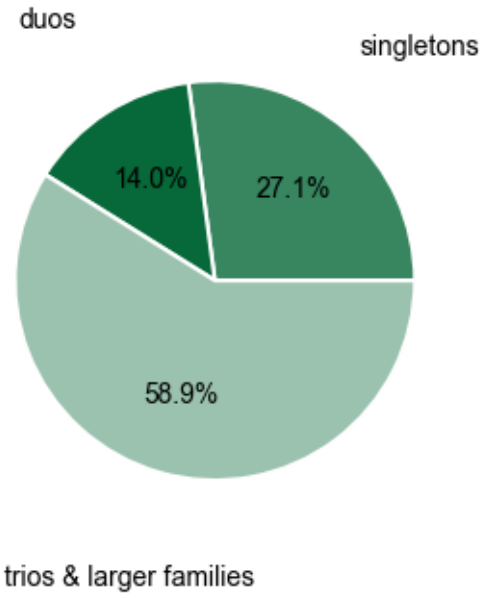

Figure 2: Pie chart summary of family structure in the GREGoR Dataset

Table 4: Table summary of family structure in the GREGoR Dataset

Family Structure	No. of Families
Singletons	979
Duos	505
Trios & larger families	2126
Total	3610

Phenotype Summaries

Table 5: Summary of affected status in the GREGoR Dataset.

	No. of participants	%
Affected	4127	0.467
Possibly affected	12	0.001
Unaffected	4435	0.502
Unknown	266	0.03

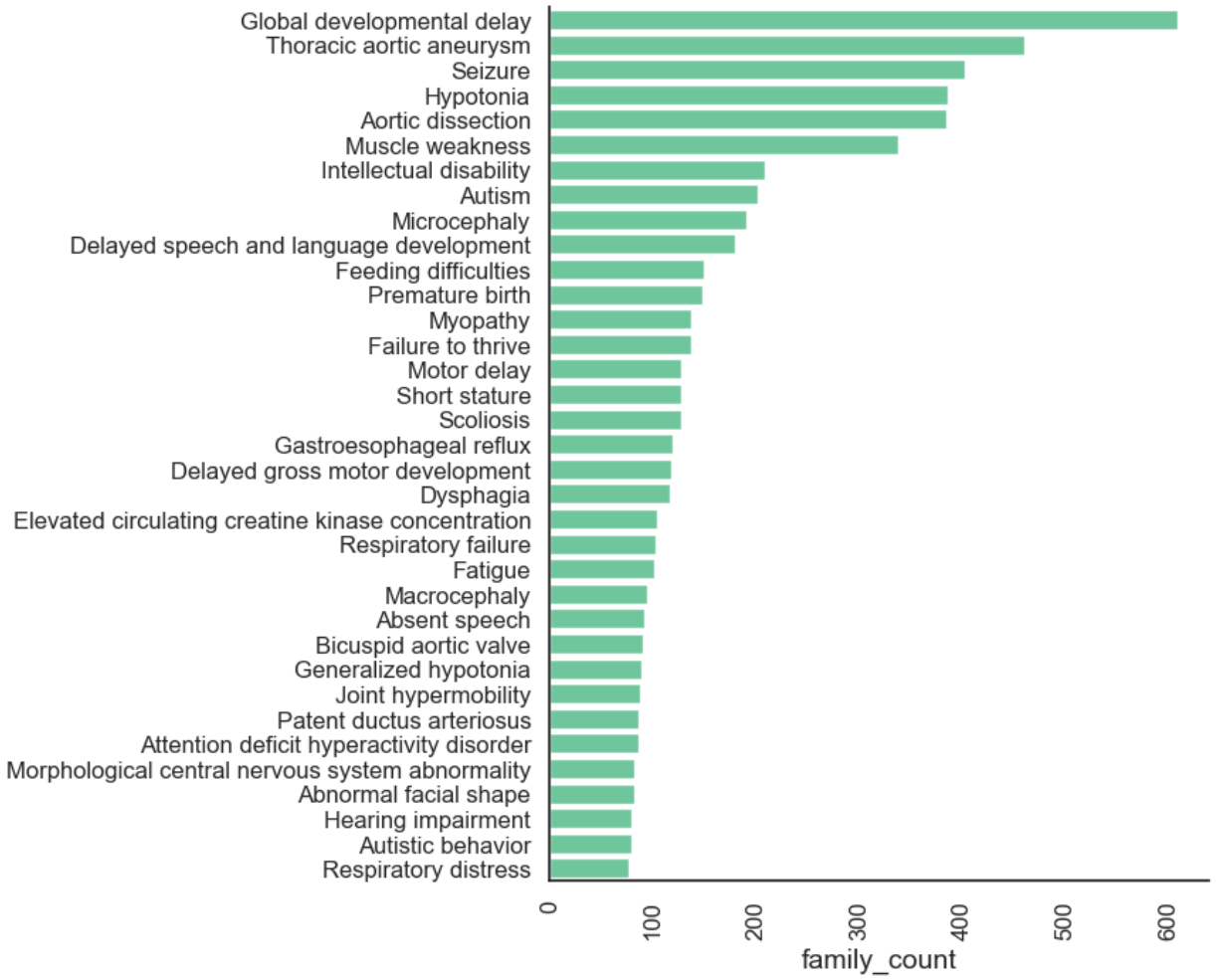

Figure 3: Common phenotypes (HPO) in the GREGoR Dataset. Phenotypes (HPO names) are on the y-axis, in descending order and shown if family count > 75 (x-axis).

Experiment Summaries

Short-read DNA

Table 6: The number of unique participants with short-read DNA sequencing experiments in the GREGoR Dataset.

GREGoR Center	Exome	Genome
GRU	1904	4736
HMB	380	1799
Total	2284	6535

Figure 4: Venn diagram showing participants with whole genome (WGS) and whole exome (WES) sequencing data in the GREGoR Dataset.

Short-read RNA

Table 7: The number of unique participants with short-read RNA in the GREGoR Dataset

GREGoR Center	paired-end	paired-end & untargeted	untargeted
GRU	347	446	0
HMB	49	0	1
Total	396	446	1

Table 8: Short-read RNA sequencing experiments by primary biosample

Primary_biosample	No. of experiments
UBERON:0000178 (blood)	674
CL:0000057 (fibroblast)	97
UBERON:0002385 (muscle tissue)	71
UBERON:0019306 (nose epithelium)	7
UBERON:0000479 (tissue)	6
CL:0000542 (lymphocyte)	3
CL:0000034 (stem cell)	1
UBERON:0001003 (skin epidermis)	1

Short-read ATAC-seq

Table 9: The number of unique participants with short-read ATAC-seq experiments.

GREGoR Center	No. of Participants
GRU	189
Total	189

GREGoR Center	No. of Participants

Table 10: The number short-read ATAC-seq experiments by primary biosample.

Primary_biosample	No. of experiments
CL: 0000576	189

Long-read DNA

Table 11: The number of unique participants with long-read whole genome experiments in the GREGoR Dataset.

GREGoR Center	Nanopore	PacBio
GRU	484	1268
HMB	19	0
Total	503	1268

Participants and probands with multiple data types

Figure 5: Venn diagram showing **participants** with multi-omic data in the GREGoR Dataset.

Figure 6: Venn diagram showing **probands** with multi-omic data in the GREGoR Dataset.

Summary of genetic findings in the GREGoR Dataset

Table 12: The number of participants with genetic findings by variant classification.

	No. of participants
Benign	1
Curation in progress	564
Likely benign	2
Likely pathogenic	215
Pathogenic	208
Uncertain significance	372
Uncertain significance - high	33
Uncertain significance - moderate	8
Well-established P/LP	107
nan	64
Total	1574

Table 13: Variant type(s) listed in the GREGoR genetic findings table.

	No. of entries
CNV	8
INDEL	146
RE	13
SNV	559
SNV/INDEL	775
SV	73
Total	1574

RE = repeat element; SNV/INDEL = single nucleotide variant OR insertion/deletions; SV = structural variant; CNV=copy number variant

Table 14: Method of discovery for genetic finding entries.

	No. of entries
SR-GS	1107
SR-ES	428
SR-ES & SR-GS	18
SR-GS-reanalysis	6
SR-GS & LR-GS	4
LR-GS	2
LR-GS & SR-GS	2
SR-ES-reanalysis	2
SR-GS & LR-GS	2
SR-ES & SR-GS & LR-GS	1
SR-GS & LR-GS & SNP array	1
nan	1
Total	1574

SR-GS = short-read genome; SR-ES = short-read exome; LR-GS = long-read genome