GREGoR R02 data summary

GREGoR DCC

2024-11-26

Contents

List of Figures

Overview

This report provides data summaries for the second external release (R02) of the GREGoR Dataset. This dataset is available on AnVIL (https://anvilproject.org/). Researchers may apply for access via dbGaP (study ID: phs003047). Graphical and tabular summaries of participant, family, experiment, and phenotype information are generated from information provided by member Research Centers (RCs) and uploaded using data tables from the GREGoR data model (https://github.com/UW-GAC/gregor_data_models).

Table 1: Overview of the GREGoR R02 dataset

	Probands	Total
participants	2986	7394
families	2966	3059
short-read DNA (WES)	849	2242
short-read DNA (WGS)	2178	5182
short-read RNA	242	539
long-read DNA (Nanopore)	53	137
long-read DNA (PacBio)	44	77

Solve status summary

Table 2: Summary of solve status for probands in the GREGoR R02 Dataset

	No. of probands	%
Partially solved	13	0
Probably solved	75	0.03
Solved	360	0.12
Unaffected	2	0
Unsolved	2536	0.85

Participant and family summaries

Table 3: The number of participants and families in the GREGoR Dataset

None	Participants	Families
GRU	5274	2223
HMB	2120	836
Total	7394	3059

GRU = General research use and clinical care ; HMB = Health/medical/biomedical research and clinical care
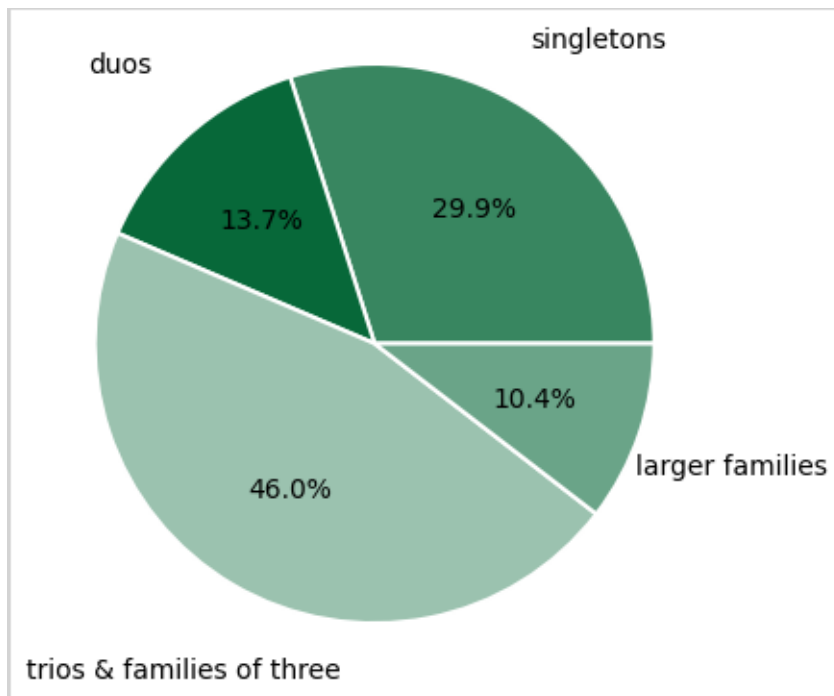

Figure 1: Pie chart summary of family structure in the GREGoR Dataset

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

Family Structure	No. of Families
Singletons	914
Duos	418
Trios & families of three	1408
Larger families	319
Total	3059

Phenotype Summaries

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

	No. of participants	%
Affected	3555	0.481
Possibly affected	8	0.001
Unaffected	3591	0.486
Unknown	240	0.032

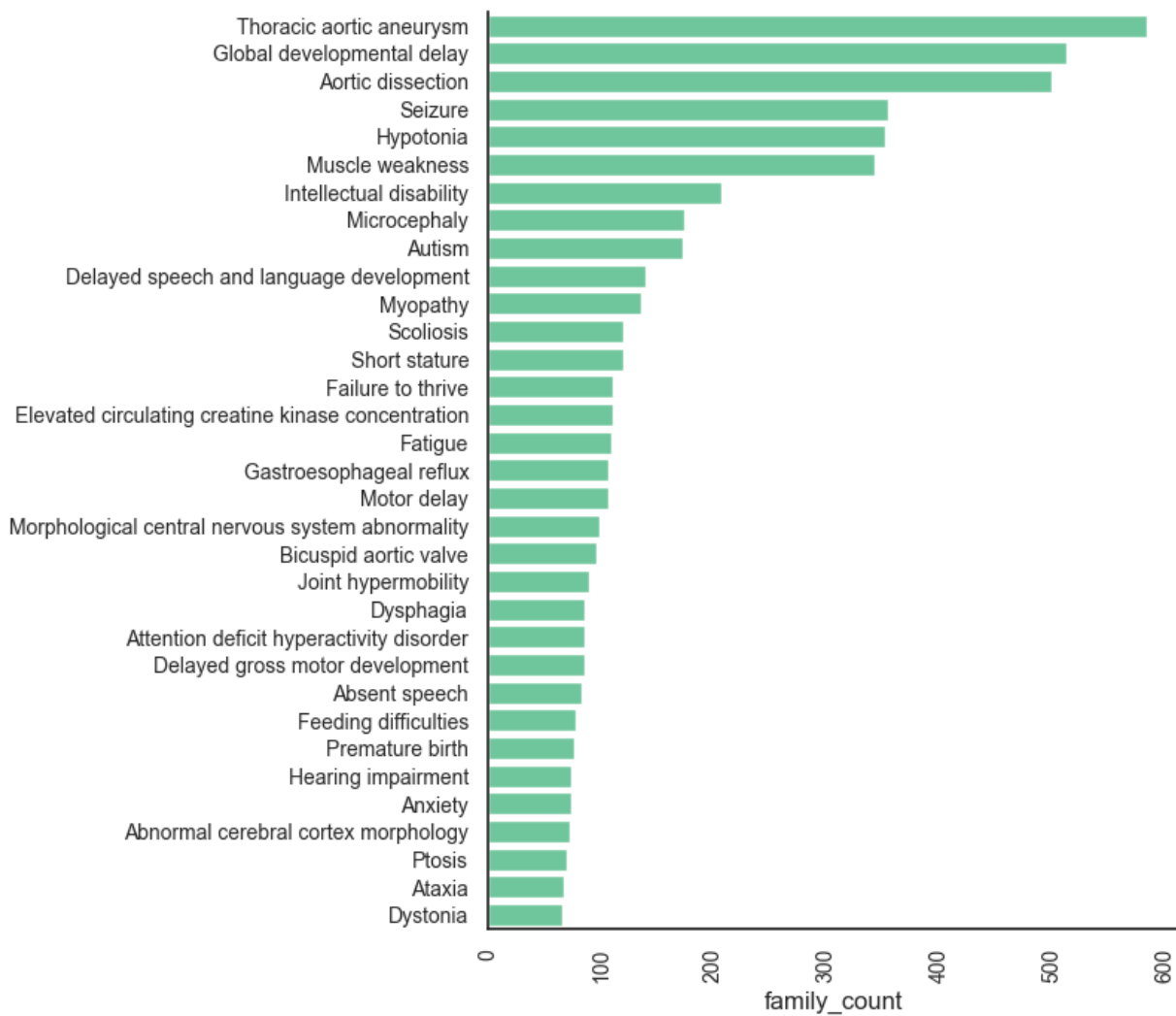Figure 2: Common phenotypes (HPO) in the GREGoR Dataset. Phenotypes (HPO names) are on the y-axis, in descending order and shown if family count > 65 (x-axis).

Experiment Summaries

Short-read DNA

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

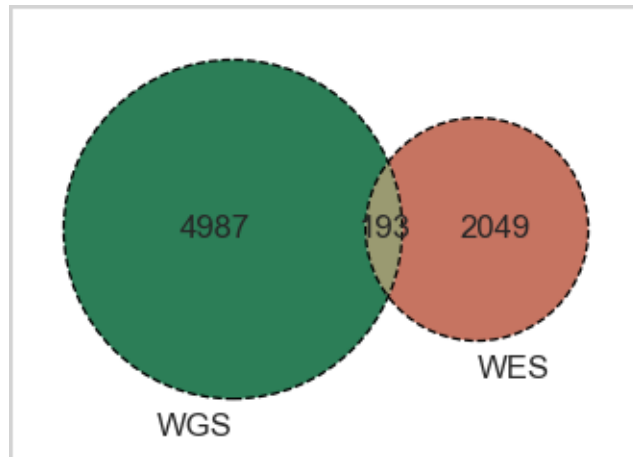|	No. of participants
Exome	2242
Genome	5180
Total	7422

4

Figure 3: 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 short-read RNA experiments by *experiment type* in the GREGoR Dataset.

	No. of participants
paired-end	171
paired-end & untargeted	358
untargeted	1
Total	530

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

Primary_biosample	No. of experiments
UBERON:0000178 (blood)	483
UBERON:0002385 (muscle tissue)	25
CL:0000057 (fibroblast)	19
UBERON:0019306 (nose epithelium)	7
CL:0000542 (lymphocyte)	2
UBERON:0000479 (tissue)	2
CL:0000034 (stem cell)	1

Long_read DNA

Table 9: The number of long-read whole genome experiments in the GREGoR Dataset.

	No. of participants
Nanopore	137
PacBio	77

Figure 4: Venn diagram showing participants with multi-omic data in the GREGoR Dataset.

Note: long-read DNA includes Nanopore and PacBio sequencing.

Summary of genetic findings

Table 10: The number of participants with genetic findings by GREGoR variant classification.

	No. of participants
Curation in progress	177
Likely pathogenic	29
Pathogenic	35
Uncertain significance	45
Uncertain significance - high	21
Uncertain significance - moderate	7
Well-established P/LP	32
NaN	273
Total	619

Table 11: Method of discovery for genetic finding entries.

	No. of participants
SR-GS	388
SR-ES	214
SR-ES & SR-GS	15
SR-GS & LR-GS	1
NaN	1

6

	No. of participants
Total	619

Table 12: Top 20 candidate genes in the genetic findings data table.

	No. of participants
TTN	15
RYR1	12
NEB	10
MSTO1	9
SYNE1	7
ABCA7	6
FBN1	6
DMD	4
ENG	4
COL3A1	4
CIAO1	4
PLA2G6	4
PKHD1	4
PKD1	4
RNU4-2	4
AFAP1L1	4
ZFHX2	3
FA2H	3
CACNA1A	3
LPL	3