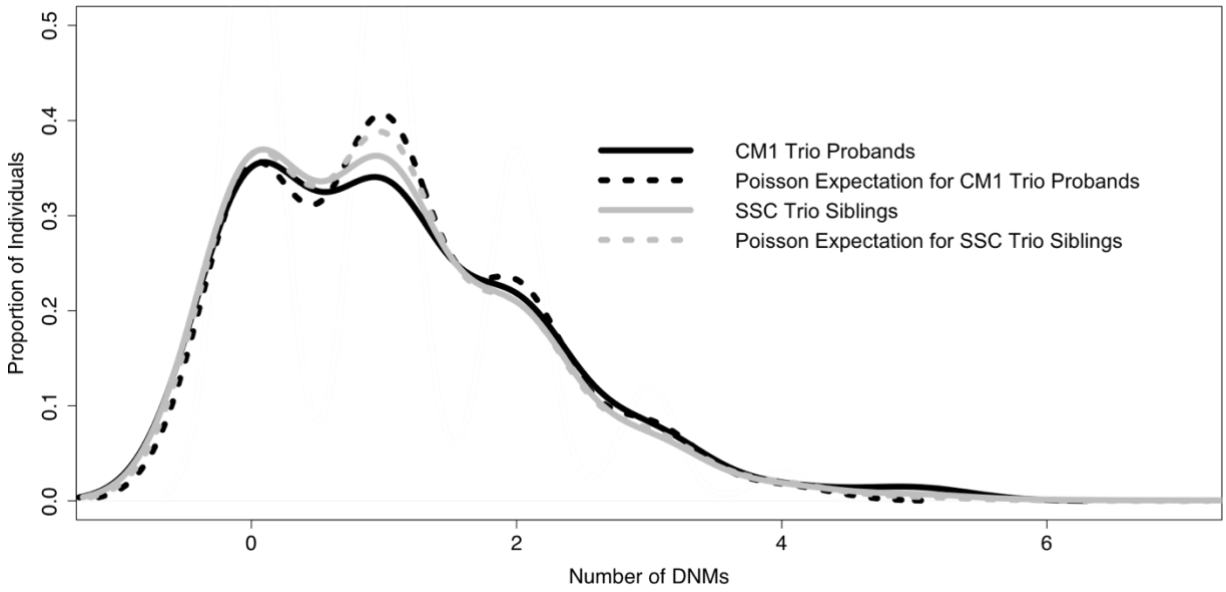
The American Journal of Human Genetics, Volume *108*

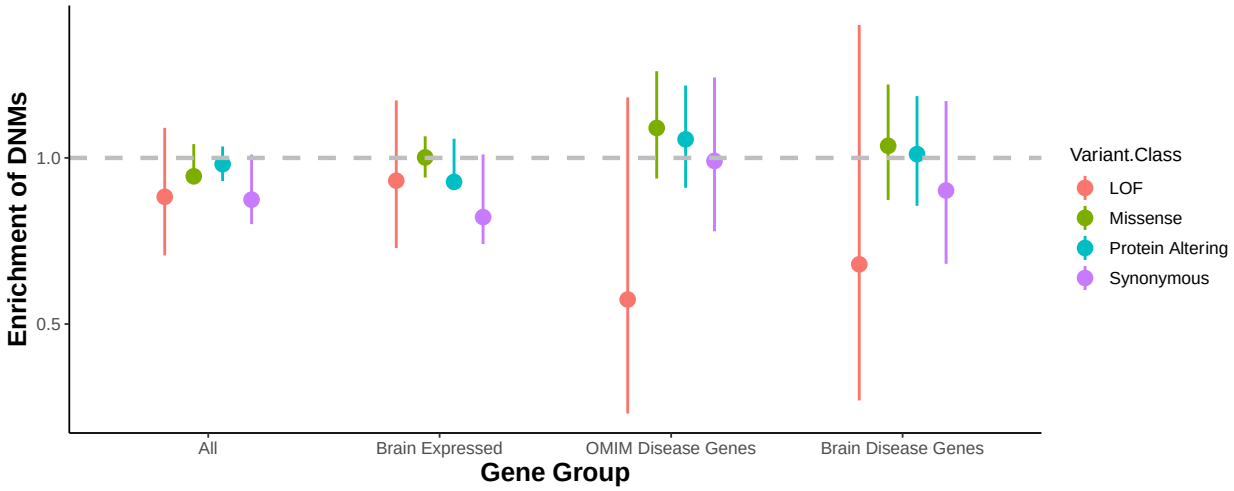
Supplemental Data

Rare and *de novo* coding variants in chromodomain genes in Chiari I malformation

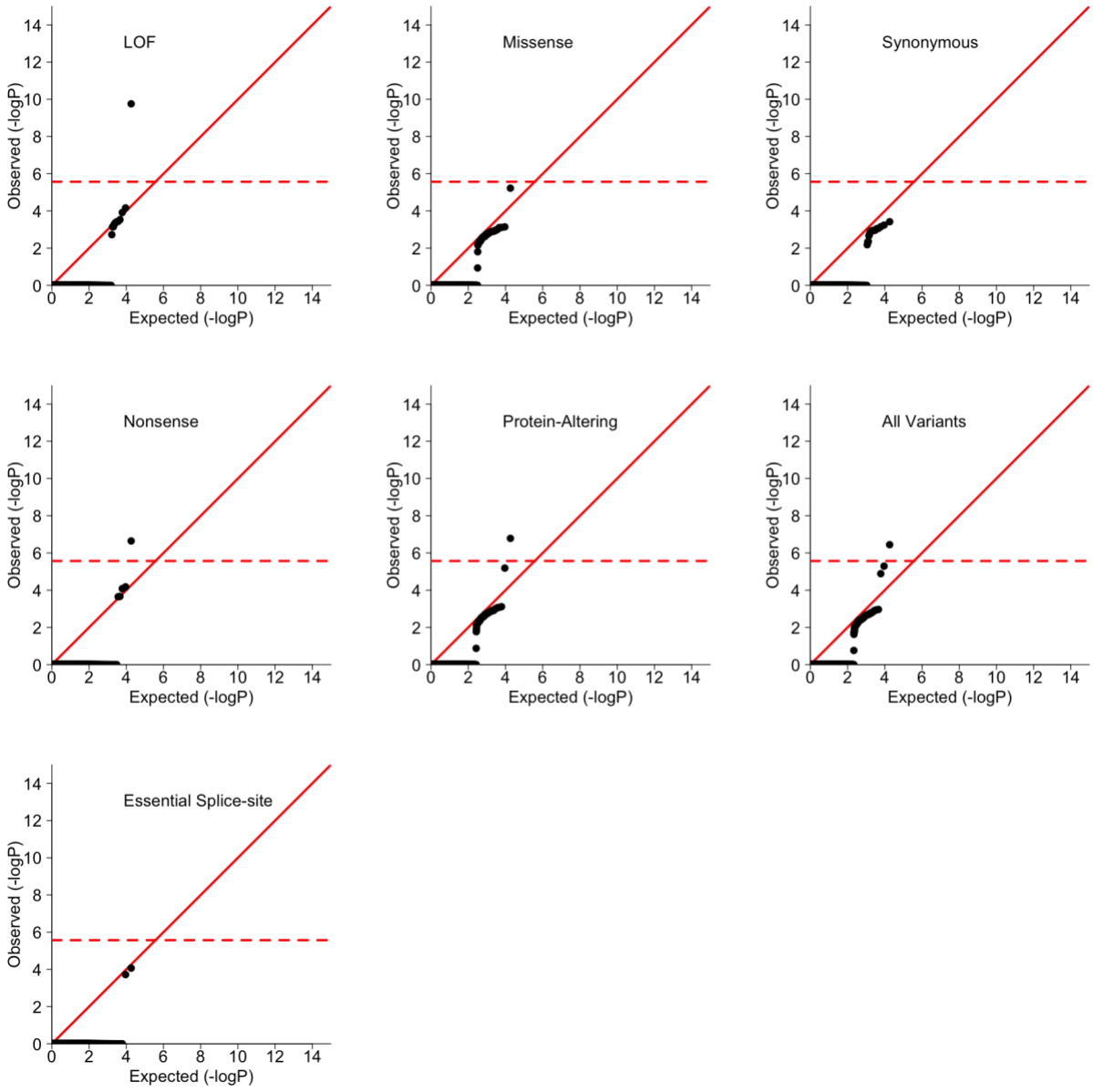
Brooke Sadler, Jackson Wilborn, Lilian Antunes, Timothy Kuensting, Andrew T. Hale, Stephen R. Gannon, Kevin McCall, Carlos Cruchaga, Matthew Harms, Norine Voisin, Alexandre Reymond, Gerarda Cappuccio, Nicola Brunetti-Pierri, Marco Tartaglia, Marcello Niceta, Chiara Leoni, Giuseppe Zampino, Allison Ashley-Koch, Aintzane Urbizu, Melanie E. Garrett, Karen Soldano, Alfons Macaya, Donald Conrad, Jennifer Strahle, Matthew B. Dobbs, Tychele N. Turner, Chevis N. Shannon, Douglas Brockmeyer, David D. Limbrick, Christina A. Gurnett, and Gabe Haller



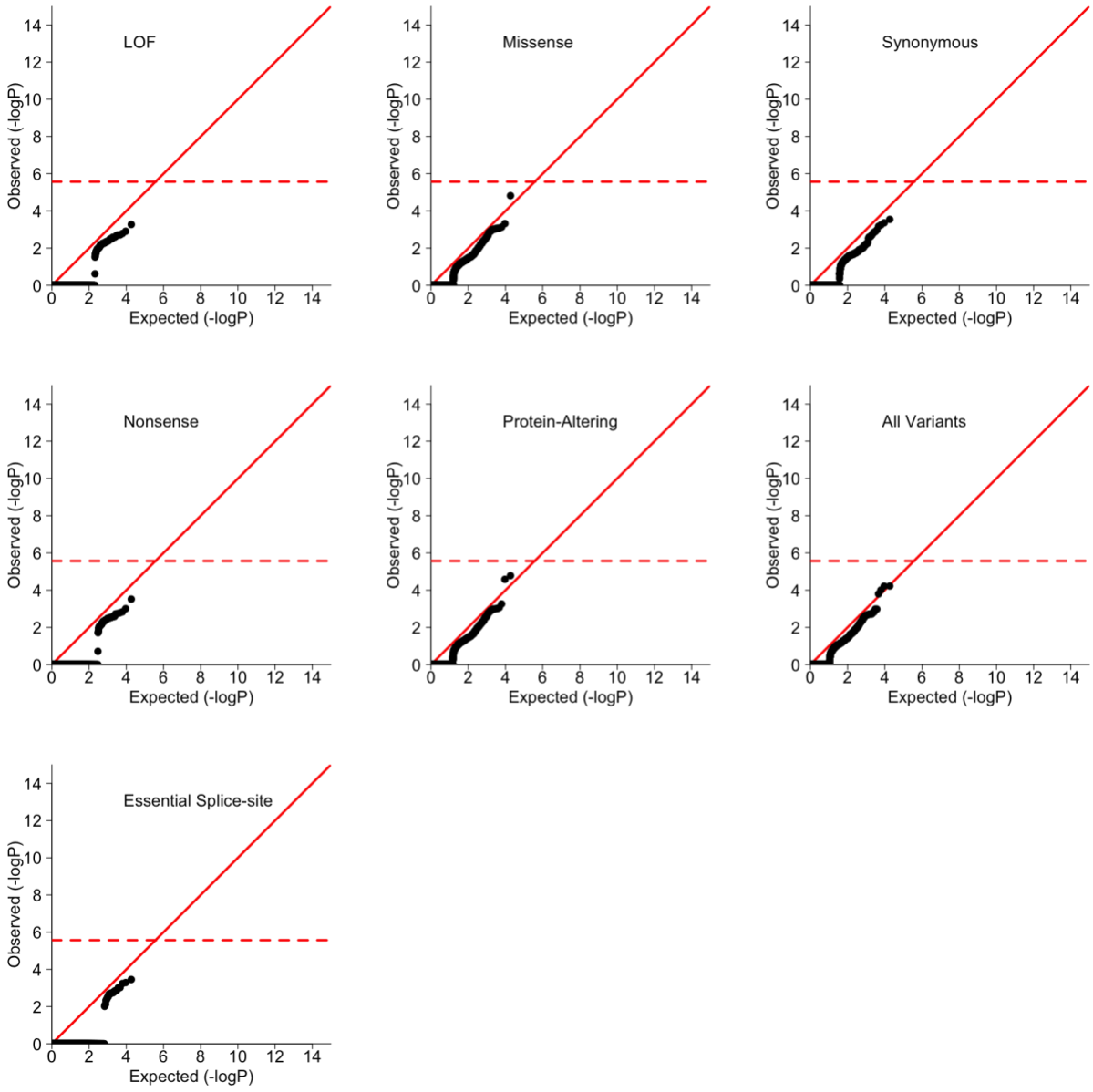
Supplementary Figure 1. Distribution and Poisson Expected Distributions for the number of De Novo Mutations (DNMs) observed per person in CM1 trios (N=67) and Control trios (N=1911). No difference was seen between the expected distribution of DNMs per person in CM1 trios (p=0.87) or in Control trios (p=0.74).



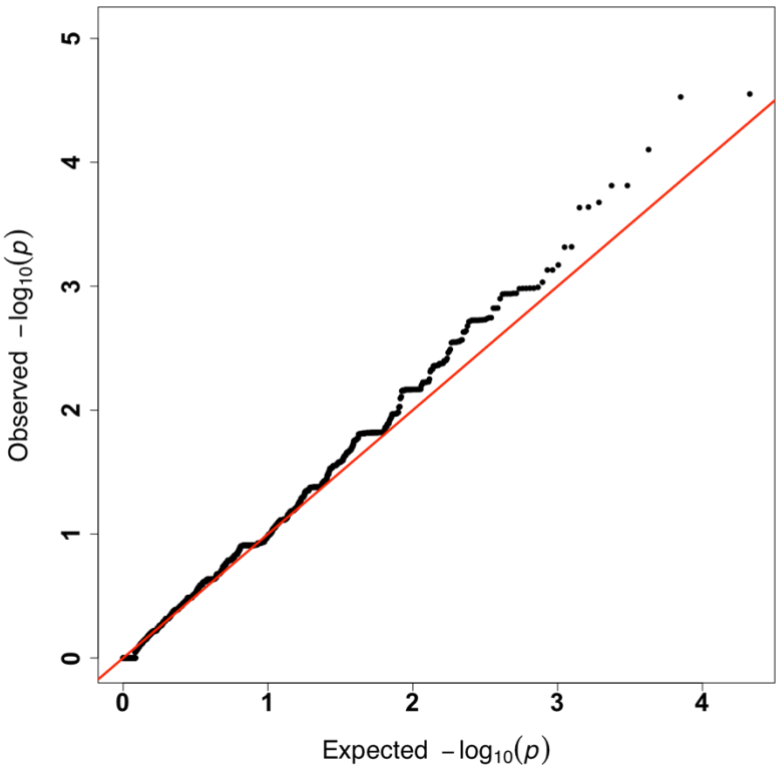
Supplementary Figure 2. Enrichment of DNMs in SSC Controls (N=1911) within gene classes. All enrichment tests presented in this figure were not significant in a two-sided Poisson test. Error bars are 95% CI.



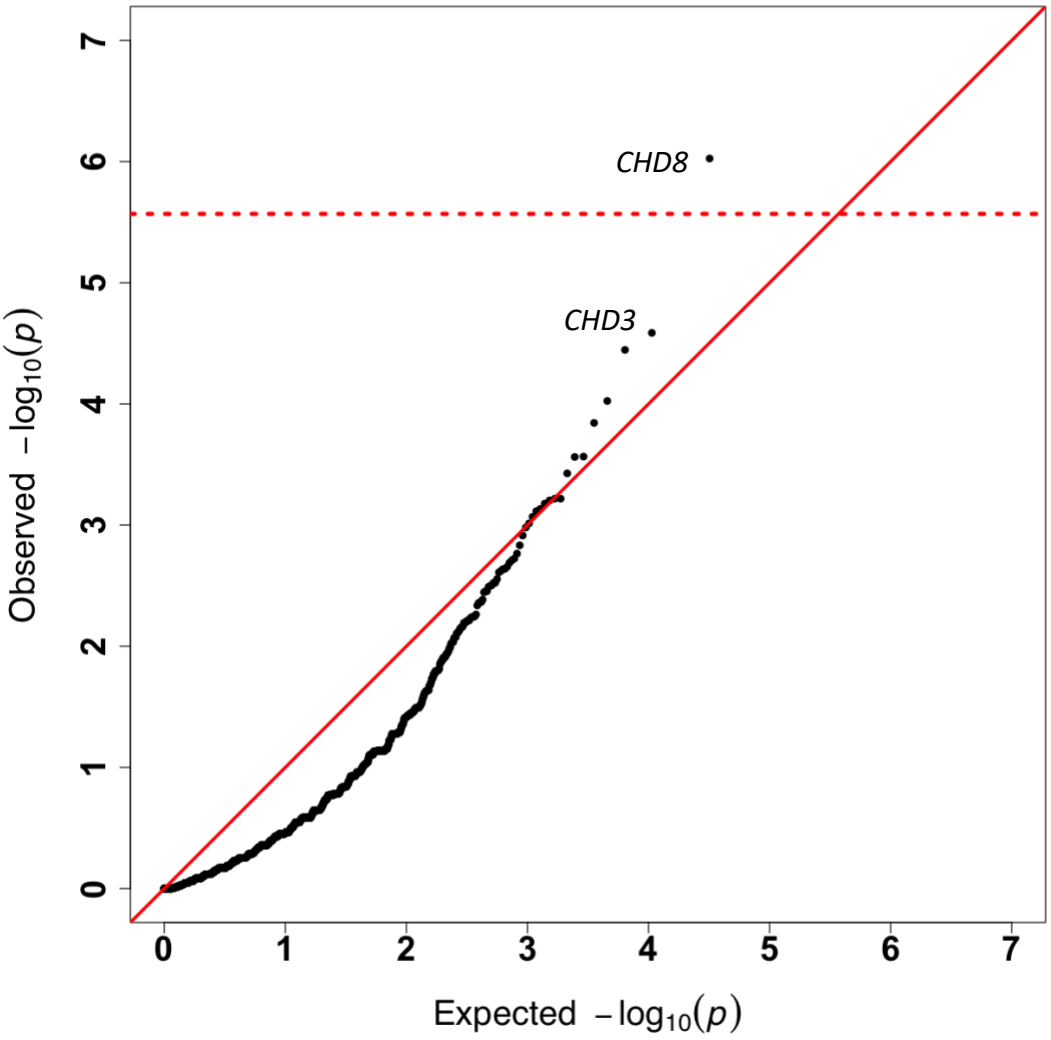
Supplementary Figure 3. QQ plots for denovolyzeR association of per gene *de novo* mutations (DNMs) in CM1 trios (N=67). See Supplementary Table 7 and Supplementary Table 8 for results for all data presented in this figure.



Supplementary Figure 4. QQ plots for denovolyzeR association of per gene *de novo* mutations (DNMs) in SSC Control trios (N=1911).



Supplementary Figure 5. Q-Q Plot for Gene-burden analysis of rare synonymous variants. See Supplementary Table 10 for summary statistics for all data presented in this figure. Lambda = 0.79, SE=0.00105.



Supplementary Figure 6. QQ plots for meta-analysis using per gene enrichment of *de novo* mutations (DNMs) in CM1 trios and gene burden analysis. For each, only protein-altering (missense, nonsense, splice-site, frameshift) variants were used. See Supplementary Table 7 for results for all data presented in this figure. Lambda = 0.31, SE=0.0018.

Table S1. Cohort Information, Quality Control, and Sequencing Metrics. Two exome library kits were used for the case/control cohorts with different target sizes and, therefore, varying coverage of RegSeq hg19 coding regions. To control for these factors, we determined the number of ‘‘callable’’ bp, or the number of bp that have coverage >10x. We then intersected these coordinates with RefSeq hg19 coding exons to determine the ‘‘total callable exome,’’ or the number of bp within RefSeq coding exons that had sufficient coveragefor de novo variant calling. Picard Tools (https://broadinstitute.github.io/picard/) generated capture, sequencing, alignment, and variant level quality metrics, and GATK DepthOfCoverage generated coverage metrics for the exome intervals. Where applicable, sequencing metrics include ±95% confidence intervals. For SSC Sibling Trios, data was not combined across capture methods so all captured bases were interogated and utilized when calculating the DNM probabilities.

	CM1 Cohort	CM1 Trios	In-house Unrelated Controls (IDT Subset)	In-house Unrelated Controls (Agilent Subset)	SSC Sibling Control Trios
Samples Sequenced (Families)	900 (668)	201 (67)	4437	527	5733 (1911)
Proband Male:Female (sex ratio)	257:411 (0.38)	35:32 (1.1)	2211:2226 (0.99)	240:287 (0.83)	900:1011
Paternal Age (95% CI)	-	32.4 (±1.78)	-	-	33.32 (±0.25)
Exome capture platform	IDT xGEN Exome Research Panel (v1.0)	IDT xGEN Exome Research Panel (v1.0)	IDT xGEN Exome Research Panel (v1.0)	Agilent SureSelect v5	Nimblegen EZ Exome V2
Size of Capture Region	39086460	39086460	39086460	50330201	44001748
RefSeq hg19 coding region covered	33418881	33418881	33418881	33048750	32586393
% Refseq hg19 coding region covered	97.81%	97.81%	97.81%	96.72%	96.33%
Intersection IDT/Agilent Capture Region	34236596	34236596	34236596	34236596	-
RefSeq hg19 coding region covered by Intersection	32622227	32622227	32622227	32622227	-
% Refseq hg19 coding region covered by Intersection[a]	95.47%	95.47%	95.47%	95.47%	-
Mean callable exome (million bp with >10x reads)	29.75 (±2.12)	33.81 (±3.34)	31.24 (±1.56)	28.97 (±2.56)	25.80 (±1.06)
Mean total reads per sample (million)	71.95 (±1.34)	89.9 (±3.56)	75.51 (± 1.22)	56.32 (± 2.48)	92.4 (±1.12)
Read length	76	76	76	76	76
Passing unique aligned reads (million)	65.41(±1.14)	81.6 (±3.24)	67.54 (± 1.22)	52.1 (± 2.34)	87.8 (±0.52)
% passing, unique reads aligned	99.23% (±0.23%)	99.33% (±0.38%)	99.31% (±0.06%)	99.21 (±0.31%)	99.78% (±0.38%)
% Duplicate reads	6.83% (±0.26%)	8.63% (±0.40%)	10.31% (±0.15%)	7.21% (±0.37%)	8.52% (±0.40%)
Mean coverage in target	71.25 (±1.89)	90.94 (±3.55)	78.72 (±0.87)	50.72 (±1.45)	90.47 (±0.54)
Median coverage in target	65.61 (±1.60)	88.34 (±3.42)	67.31 (±0.69)	41.72 (±1.23)	88.56(±0.56)

% target at 10x	98.35% (±1.56%)	99.32% (±0.21%)	99.12% (±0.25%)	97.13% (±0.26%)	99.56% (±0.07%)
% target at 15x	88.35% (±1.31%)	96.06% (±0.34%)	92.27% (±0.51%)	87.52% (±0.53%)	96.67% (±0.14%)
% target at 20x	85.56% (±1.46%)	88.25% (±0.73%)	89.45% (±0.56%)	79.16% (±0.66%)	88.11% (±0.3%)
% target at 30x	72.34% (±1.69%)	79.92% (±1.05%)	75.23% (±0.91%)	73.12% (±0.21%)	72.14% (±0.63%)
Het SNP quality	11.36 (±0.07	10.97 (±0.06)	10.94 (±0.03)	10.84 (±0.08)	11.54 (±0.02)
Base pair error rate	0.0035 (±0.0001)	0.0042 (±0.0001)	0.0043 (±0.0001)	0.0041 (±0.0001)	0.002 (±0.0001)
Novel transition/transversion ratio	2.12 (±0.02)	2.16 (±0.02)	2.03 (±0.02)	2.21 (±0.03)	2.01 (±0.01)

Table S2. De Novo Variant Statistics

	CM1 Trios	SSC Sibling Control Trios
Samples Sequenced (Families)	201 (67)	5733 (1911)
Male:Female (sex ratio)	35:32 (1.1)	900:1011 (0.89)
Paternal Age (95% CI)	32.4 (±1.78)	33.32 (±0.25)
DNM Average SNP quality (95% CI)	97.85 (±1.62)	98.2 (±0.02)
DNM Ti/Tv	1.92	2.33
Average Read Depth at DNMs (95% CI)	80.23 (±13.6)	68.85 (±5.7)
DNM Per Person	1.20 (±0.27)	1.04 (±0.04)

Supplementary Acknowledgements

We thank the cases and their families for their role in this work. This study was funded in part by the generous support of Sam and Betsy Reeves and the Park-Reeves Syringomyelia Research Consortium. This research was supported by the University of Missouri Spinal Cord Injury Research Program and the Children's Discovery Institute of St Louis Children's Hospital and Washington University (CAG and GH). We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant #UL1RR024992 from the National Center for Research Resources (NCRR), a component of the NIH (National Institute of Health) and NIH Roadmap for Medical Research. Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially funded by NIH grants 1S10RR022984-01A1 and 1S10OD018091-01. We thank Carlos Cruchaga, Matthew Harms and Donald Conrad for their contribution of exomes for use as controls funded under grants R01-AG044546, R01-AG035083, the Alzheimer Association (NIRG-11-200110), K08-NS075094 R01-HD078641. Research reported in this publication was also supported by the Washington University Institute of Clinical and Translational Sciences grant UL1 TR000448 from the National Center for Advancing Translational Sciences (NCATS) of the NIH, the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number U54 HD087011 to the Intellectual and Developmental Disabilities Research Center at Washington University and grants from the Swiss National Science Foundation (31003A_182632) and the Jérôme Lejeune Foundation to AR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. A.T.H. is supported by the National Institutes of Health (F30HL143826) and Vanderbilt University Medical Scientist Training Program (T32GM007347). DNA samples obtained from Vanderbilt University Medical Center (BioVU) are supported by individual investigator-led projects including U01-HG004798, R01-NS032830, RC2-GM092618, P50-GM115305, U01-HG006378, U19-HL065962, and R01-HD074711. Additional funding sources for BioVU are listed at https://victr.vanderbilt.edu/pub/biovu/. Additionally, we thank Sheng Chih Jin, Ben Neale, Konrad Karczewski and Kaitlin Samocha for helpful discussion. We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet,

R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic and genetic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study (https://www.sfari.org/resource/simons-simplex-collection/) by applying at https://base.sfari.org.