

Rare Variant Burden in Known Dystonia Genes in Population Controls and Sporadic Dystonia Patients

Elizabeth T. Cirulli, Ph.D.^{1,2}, Patrick Hickey, D.O.³, Samantha Tracy³, Julia Johnson, M.D.³, Zachary Caffall, M.S.³, Kaylin Tsukayama³, Burton Scott, M.D., Ph.D.³, Mark Stacy, M.D.³ and Nicole Calakos, M.D., Ph.D.^{3,4*}

Affiliations:

¹Department of Molecular Genetics and Microbiology

²Center for Applied Genomics and Precision Medicine

³Department of Neurology, Movement Disorders section

⁴Department of Neurobiology

Duke University Medical Center, Durham, North Carolina, USA

*Corresponding Author: Nicole Calakos, M.D., Ph.D., Box 2900 DUMC, Durham, NC 27715
USA, (919) 684-2423, nicole.calakos@duke.edu

Word Count: 3287

Running Title: Rare variants in dystonia genes

Keywords: dystonia, genetics, exome, exonic, rare variant

Financial Disclosures/Conflict of Interest concerning the research related to the manuscript: NC, PH, ZC, KT, MS, BS, ETC, ST, JJ have nothing to disclose.

Funding: Funding for collection of dystonia subjects and analysis was generously provided by Duke University's Institute of Genome Sciences and Policy (N.C.). The collection of control samples and data was funded in part by: the Duke Chancellor's Discovery Program Research Fund 2014; Gilead Sciences, Inc.; Bryan ADRC NIA P30 AG028377; B57 SAIC-Fredrick Inc M11-074; The Ellison Medical Foundation New Scholar award AG-NS-0441-08; National Institute of Mental Health (K01MH098126, R01MH097993, RC2MH089915); National Institute of Allergy and Infectious Diseases (Division of Intramural Research); National Human Genome Research Institute (U01HG007672); National Institute of Neurological Disorders and Stroke (U01NS077303, U01NS053998, U01NS077274, U01NS077276, U01NS077364, U01NS077367, and U01NS077275); National Institute of Allergy and Infectious Diseases Center (U19-AI067854, UM1-AI100645); and the Bill and Melinda Gates Foundation.

Abstract

Background: Rare mutations in genes associated with Mendelian forms of disease are a potential mechanism for sporadic disease. The need to assess the clinical significance of such variants is increasing as personalized medicine and genome sequencing increases.

Objective: To evaluate the rate of rare, functional variants in dystonia genes in the general population to improve interpretation of the clinical relevance of potentially pathogenic variants in dystonia cases.

Methods: We performed an “aggregated” collapsing analysis of exome sequence that considered rare coding variants in genes previously associated with dystonia, a rare neurological movement disorder, on 2,372 population controls of European ethnicity. We then performed a pilot study in sporadic dystonia to assess whether there was a substantially greater incidence of individuals with rare variation in dystonia genes.

Results: Nearly half of population controls had a rare coding variant when 148 genes associated with a dystonia phenotype were considered. When the subset of genes causing isolated dystonia (14 genes) was evaluated, 3-4% of controls harbored rare qualifying variants. Our pilot study of case exomes was powered to identify a five-fold higher or greater rate of qualifying variants in isolated dystonia genes in sporadic dystonia cases compared to population controls; we did not find such an enrichment.

Conclusions: We provide the first systematic analysis of rare variation in dystonia genes considered collectively. Our findings emphasize the need to consider the overall frequency of variants in rare disease-related genes in the general population when considering their potential role in clinical presentations.

1

INTRODUCTION

2

3 Human geneticists are increasingly appreciative of the role of rare variants in disease [1-4]. A
4 role for novel mutations in genes associated with familial, Mendelian forms of diseases has been

1 identified as a cause for sporadic presentations in a number of diseases, such as *de novo*
2 mutations in Mendelian mental retardation genes [5]. Moreover, exome sequencing studies of
3 sporadic cases in diseases such as amyotrophic lateral sclerosis and myocardial infarction have
4 found support for contributions by both novel risk genes and known familial disease genes [6, 7],
5 indicating heterogeneous genetic contributions.

6

7 Dystonia is an involuntary, centrally driven movement disorder involving slow twisting
8 movements and sustained abnormal postures [8]. Dystonia can present alongside other
9 neurological symptoms, as in the setting of neurodegenerative disease, metabolic disorders,
10 trauma or stroke, or occur in isolation, as in the setting of drug side effects, rare inherited
11 diseases, and the most common of isolated presentations, sporadic disease. Although a genetic
12 component is clearly demonstrated in rarer forms of dystonia [9], much less is known about the
13 etiology of adult-onset, sporadic disease, which includes presentations such as blepharospasm,
14 cervical dystonia, and spasmodic dysphonia. Environmental, epigenetic, and genetic factors are
15 all potential contributors.

16

17 In dystonia, rare coding variants in known dystonia genes have been reported in isolated sporadic
18 cases [10-20]. In such isolated cases, a clinician may be tempted to attribute the rare variant to
19 the patient's clinical presentation. However, this approach suffers from ascertainment bias, since
20 the individual's frequency is "100%" *a priori*. Secondly, while large population databases are
21 available to look up frequencies of rare variants, typically, either only the incidence of that
22 specific variant or the overall incidence of variants in that specific gene are considered. As the
23 finding of a rare missense variant in any gene associated with the patient's clinical phenotype

1 would have prompted making an attribution to their presentation, it is important to first consider
2 the overall frequency of individuals in the general population that have similarly qualifying
3 missense variants in disease-associated genes *considered collectively*. For example, to date,
4 when the burden of rare variants has been considered, it has typically been assessed in single
5 specific dystonia genes [11, 19, 21-23]. Thus, the overall likelihood of finding a rare variant in
6 any dystonia-associated gene in the general population is not readily accessible to clinicians as
7 yet because this has not been systematically evaluated.

8

9 In this study, we performed exome sequencing in 2,372 population controls and a small pilot
10 cohort of 20 sporadic dystonia cases to determine the overall incidence of individuals with
11 qualifying variants in dystonia genes. In this analysis, the presence of 1 or more rare missense
12 variants in any dystonia-associated gene qualified an individual to be counted. Thus, rather than
13 a gene-by-gene assessment, this analysis assesses the overall contribution of dystonia gene
14 variation to population controls and sporadic cases. Importantly, this analysis does not consider
15 the specific role of any given variant or gene in its association with dystonia cases.

16

17 This approach involves “collapsing” all variants in a group of genes for the purpose of assessing
18 the overall population frequency of dystonia gene rare variation (reviewed in [24, 25]). This
19 analysis also enables an assessment of whether rare variation in dystonia-associated genes occurs
20 more commonly in cases than controls. Here, we provide a first assessment of this possibility
21 using a small pilot cohort of twenty sporadic cases evaluated in our Movement Disorders
22 specialty clinic.

23

1

2 **METHODS**

3

4 *Samples*

5 The study was approved by the Duke University Medical Center IRB, and all subjects gave
6 written informed consent. Control samples were sequenced as part of other studies at Duke
7 University Medical Center and were not enriched for (but not specifically screened for) dystonia
8 or other neurological disorders (Table 2). Cases were recruited at the Movement Disorders
9 Center at Duke University Medical Center, Durham, NC. We collected 20 unrelated patients
10 diagnosed with adult onset, sporadic dystonia (3 men and 17 women) with a mean age of
11 dystonia onset of 47.4 ± 9.59 years. A diagnostic workup was conducted by a movement
12 disorders specialist to confirm the symptoms of dystonia with muscle involvement classified as
13 focal, segmental, multifocal, or generalized. Only presumptive primary cases were recruited. As
14 is typical for our sporadic dystonia patients, no prior genetic testing had been performed for the
15 majority (1 case had negative Dyt1 testing). Secondary dystonias associated with conditions such
16 as Parkinson's disease or other neurodegenerative disease were excluded. Cases suggestive of
17 Mendelian inheritance were also excluded. A complete family and medical history was collected
18 including common toxic exposures and medical comorbidities.

19

20 *DNA sequencing*

21 Genomic DNA samples from 2,372 population controls and twenty subjects with dystonia were
22 exome sequenced at Duke University. Control samples were either exome sequenced using the
23 Agilent All Exon (37MB, 50MB or 65MB) or the Nimblegen SeqCap EZ V2.0 or 3.0 Exome

1 Enrichment kit or whole-genome sequenced, and case samples were sequenced using the Agilent
2 All Exon 37MB or 50MB kit, using Illumina GAIIX or HiSeq 2000 or 2500 sequencers
3 according to standard protocols. All samples were processed using the same methods, as follows.
4 The Illumina lane-level fastq files were aligned to the Human Reference Genome (NCBI Build
5 37) using the Burrows-Wheeler Alignment Tool (BWA)[26]. We then used Picard software
6 (<http://picard.sourceforge.net>) to remove duplicate reads and process these lane-level SAM files,
7 resulting in a sample-level BAM file that is used for variant calling. We used GATK to
8 recalibrate base quality scores, realign around indels, and call variants [27]. Variants were
9 required to have a quality score (QUAL) of at least 20, a genotype quality (GQ) score of at least
10 20, at least 10x coverage, a quality by depth (QD) score of at least 2 and a mapping quality (MQ)
11 score of at least 40. Indels were required to have a maximum strand bias (FS) of 200 and a
12 minimum read position rank sum (RPRS) of -20. SNVs were restricted according to VQSR
13 tranche (calculated using the known SNV sites from HapMap v3.3, dbSNP, and the Omni chip
14 array from the 1000 Genomes Project): the cutoff was a tranche of 99.9%. Variants were
15 excluded if marked by EVS as being failures [28]. Variants were annotated to Ensembl 73 using
16 SnpEff [29].

17

18 Only genetically European ethnicity samples were included in the analysis. Samples were
19 screened with KING [30] to remove second-degree or higher relatives; samples with incorrect
20 sexes according to X:Y coverage ratios were removed, as were contaminated samples according
21 to VerifyBamID [31].

22

23 *Statistical and informatic analysis*

1 Our study used a gene-based collapsing methodology as previously described [6]. For each gene,
2 each sample was indicated as carrying or not carrying a qualifying variant. Qualifying variants
3 were defined for a dominant model requiring at least one qualifying variant per gene with a
4 minor allele frequency (MAF) cutoff of 0.1% internally and 0.01% in each population of the
5 ExAC database, which has been available for longer than the gnomad database[32]. These allele
6 frequency thresholds used a leave-one-out method for the combined sample of cases and controls
7 (where the MAF of each variant was calculated using all samples except for the sample in
8 question). We performed analyses of CCDS genes using two methods to identify qualifying
9 variants: 1) all non-synonymous and canonical splice variants (coding model), and 2) all
10 canonical splice and non-synonymous coding variants except those predicted by PolyPhen-2
11 HumDiv [33] to be benign (likely gene disrupting model). Qualifying variants were identified
12 using Analysis Tools for Annotated Variants v.6.0
13 (<http://redmine.igm.cumc.columbia.edu/projects/atav/wiki>). This is an in-house software package
14 maintained by the Institute for Genomic Medicine (IGM) at Columbia University that compiles
15 genotype and quality information on all cases and controls sequenced at the IGM as well as
16 annotation data from sources like Ensembl, Polyphen, and ExAC in an easy to use format. We
17 also downloaded qualifying variants from the ExAC database [32] using the same criteria for the
18 coding and likely gene disrupting models, restricting to variants that pass ExAC QC and
19 requiring MAFs to be below 0.01% in each population of the ExAC database. While individual-
20 level data are not available from ExAC, making it impossible to determine exactly how many
21 people had a variant in each gene, we used a conservative estimate by adding up the number of
22 carriers of each rare variant in each gene, and we calculated the proportion of people with a
23 qualifying variant in each gene by assuming that no one had more than one variant in the gene.

1

2 In addition to comparing the numbers of cases and controls with qualifying variants for each
3 model in each gene, we also performed an aggregate collapsing analysis that compared the
4 number of cases and controls with a qualifying variant in any of the known dystonia genes,
5 considered collectively.

6

7 Candidate dystonia genes were identified as OMIM genes containing the word “dystonia” in the
8 clinical synopsis plus *FTL* (*NBIA3*), *PANK2* (*NBIA1*), *PLA2G6* (*NBIA2*), *ATN1*, *HTT*, *PARK2*,
9 *TAF1* (*DYT3*), *TOR1A*, *ARSG*, and *CIZ1* (n =148 genes). Isolated dystonia genes were those
10 covered in a recent dystonia review (n =14 genes)[9]. Overall, 94.9% (88.7%) of the coding
11 bases had at least 10x coverage in our control (case) samples, indicating that the sequence data
12 would be able to pick up most causal variants in these genes if they existed. There were 7 genes
13 that had <50% of their bases with at least 10x coverage (<50% coverage in cases and controls:
14 *ARX*, *GJC2*, and *SDHAF1*; poor coverage in cases only: *TSFM*, *CACNA1B*, *CACNA1A*, and
15 *PDX1*). For the aggregate collapsing analysis across all candidate genes, a Fisher’s exact test was
16 performed using <http://www.langsrud.com/fisher.htm>. Power was calculated using the Genetic
17 Power Calculator [34], with a dystonia prevalence of 1 in 10,000, 80% power, a D-prime of 1,
18 marker allele frequency of 3% (to reflect the frequency of qualifying coding mutations occurring
19 in any of the isolated dystonia genes in the general population) and equal to risk allele frequency,
20 Aa relative risk equal to AA relative risk (dominant model), 20 cases, a control:case ratio of
21 118.6 (for 2,372 controls), and an alpha of 0.05 for the one test of all isolated dystonia genes in
22 aggregate.

23

1

2

3

RESULTS

4

5 To characterize the prevalence of individuals with genetic variants that could potentially be
6 described as pathogenic in genes associated with dystonia, we first sequenced the exomes of
7 2,372 population controls of European ancestry. We additionally sequenced the exomes of 20
8 patients with sporadic dystonia and of European ancestry, whose clinical characteristics are
9 summarized in Table 1.

10

11 Two sets of “dystonia-associated” genes were curated for the analysis. The first is a more
12 stringent set of genes (14 in total) that have been associated with dystonia in isolation from other
13 phenotypes, “isolated dystonia” (reviewed in [9]). The second set of genes (n=148) more broadly
14 considered those genes in which dystonia had been described as part of the clinical phenotype
15 (see Methods for further details). We first performed a gene-based collapsing analysis where
16 each sample was coded to indicate whether they had a qualifying variant in each sequenced gene.
17 “Qualifying” was defined based on two different genetic models: “coding” and “likely gene
18 disrupting”. We then identified the total number of subjects with a qualifying variant in any of
19 the dystonia genes analysed to arrive at the overall frequency.

20

21 Among the group of isolated dystonia genes, we found that 4.2% of controls had qualifying
22 variants in the coding model and 3.0% in the likely gene disrupting model (Table 2). At this rate,
23 we would have expected fewer than 1 of the 20 dystonia cases to have qualifying variants in at

1 least one of these genes by chance, and we found that 0 had such variants. Though we only had
2 20 cases, our aggregate analysis across all known isolated dystonia genes had 80% power to
3 identify a signal where at least 15% of sporadic dystonia cases were due to pathogenic or
4 possibly pathogenic coding variants in these genes.

5

6 We then broadened our scope to all genes previously reported as associated with dystonia in
7 OMIM and literature reviews as compared to controls, 148 genes in total. We found that 50.6%
8 of controls had qualifying coding variants in at least one of these genes. The frequency among
9 cases was remarkably similar, 45.0%. For the likely gene disrupting model, qualifying variants
10 were found in 37.1% of controls and again, a similar frequency of 35.0% in cases.

11

12 Although the primary outcome for this analysis was the aggregated contribution of rare variation
13 in dystonia genes when considered as a group, in Table 3 we also indicate the proportion of cases
14 and controls with qualifying variants for each gene analysed. The dystonia-associated gene that
15 was most often mutated in population controls was *DST*, with 4.1% of controls having variants in
16 the coding model and 1.6% in the likely gene disrupting model. Additionally, all of the primary
17 dystonia genes had qualifying variants in at least one of our controls under the coding model,
18 while 4.7% of the genes in the extended candidate dystonia gene list were never mutated in this
19 control set.

20

21 Table 3 and S1 Table also present an estimate for the proportion of people in the general
22 population from several ethnicities who have qualifying variants in each of these genes. The data
23 for this section come from the ExAC database, a collection of exome sequence data from

1 >60,000 people at the time of access [32]. Unlike the data presented for our locally sequenced
2 controls, the proportion of carriers of qualifying variants in each ExAC population is an estimate
3 because individual-level data are not available from ExAC; that is, if the same person has two
4 qualifying variants in the same gene, then you will accidentally count them twice. In our locally
5 sequenced controls, we found that 0.5% and 0.9% of controls had more than one qualifying
6 variant in a single candidate dystonia gene under the coding and likely gene disrupting models,
7 respectively. This result indicates that it is uncommon for a person to have multiple qualifying
8 variants in a single candidate dystonia gene, and so the frequency of qualifying variants per gene
9 from ExAC is fairly accurate, which is also supported by the strong similarity between the
10 frequencies of qualifying variants for each gene in our locally sequenced controls and the
11 European ExAC population (S1 Fig and S2 Fig). However, it is important to note that one cannot
12 use ExAC data to estimate how often a person would have a qualifying mutation in any of the
13 genes, i.e., considered in aggregate. In our controls, 15.5% of individuals had rare qualifying
14 variants in more than one dystonia gene (of the 148 dystonia genes) for the coding model, and
15 8.5% for the likely gene disrupting model.

16

17

18 DISCUSSION

19

20 Here, we have performed exome sequencing and evaluated the frequency of individuals with rare
21 variation in previously reported dystonia genes in 2,372 population controls and a pilot sample of
22 20 sporadic dystonia patients. We found that it is fairly common for population controls to have
23 likely gene disrupting variants in these genes. Depending on the variant-calling model used, we

1 found that between 37 and 51% of the control population had qualifying variants in dystonia
2 genes. Among the subset of 14 genes associated with causing only dystonia and not being part of
3 a larger syndrome (“isolated”), we found that 3-4% of the population harbored qualifying
4 variants. These frequencies are noteworthy since dystonia is a rare disease, estimated to affect
5 significantly fewer than 0.1% of the population [35]. Thus, the results of our study provide an
6 important point-of-reference for providers in the clinical setting that are considering the
7 significance of a rare variant found in an individual with dystonia.

8

9 For the purpose of counseling a patient, knowing that there is not a strong enrichment of variants
10 in dystonia genes in sporadic patients promotes using caution when assigning causality to a
11 variant. Based on our analysis, if all individuals with a qualifying rare variant in a dystonia gene
12 were to be clinically considered for the presence of dystonia, at most 1 individual in 2000
13 subjects with rare variants in isolated dystonia genes would be expected to have dystonia based
14 on an estimated clinical incidence of 1 in 10,000 and a frequency of such variants in the general
15 population of 3% (and <15% of sporadic dystonia cases). Conversely, for an individual with
16 clinically confirmed dystonia, the chance that their rare variant is incidental is 1 in 25-33 for
17 isolated dystonia genes and 1 in 2-3 for any dystonia-associated gene.

18

19

20 A number of isolated cases of sporadic disease have been reported that possess rare, likely gene-
21 disrupting mutations in previously described dystonia genes [10-13, 15, 17-19], and more
22 recently, the collective burden of rare variation has been assessed in single dystonia genes of
23 interest [11, 19, 21, 36]. The present study is the first to formally describe the overall frequency

1 of rare dystonia gene variants. We focused the study on population controls, but also
2 systematically searched for evidence of enrichment in sporadic dystonia cases. Though our case
3 sample size was small, the collapsing nature of this analysis, in which all dystonia genes are
4 considered as a group, was sufficiently powered to ascertain whether a substantial fraction of
5 sporadic dystonia cases (>15% or 3 subjects) could be explained by rare missense variation in
6 isolated dystonia genes. We found no evidence for an enrichment of rare functional variants in
7 dystonia genes among sporadic dystonia patients. Of course, specific variants and/or individual
8 dystonia-associated genes may still play a role in some cases and larger scale studies are
9 necessary to determine their significance. Our study subjects had predominantly adult-onset focal
10 dystonias, while a recent report indicates a potentially larger contribution of rare variants in
11 dystonia genes to early-onset dystonia [37]. Larger scale studies are necessary to determine these
12 possibilities. Our results indicate however, that this genetic mechanism, on its own, is not
13 sufficient to explain the bulk of sporadic cases and emphasize the need for a better understanding
14 of other potential genetic and non-genetic factors that give rise to dystonia in the population with
15 late-onset sporadic presentation.

16
17 We also note that our conclusions specifically apply to rare coding variants and not other sources
18 of genetic variation. For example, exome sequencing does miss many insertions and deletions as
19 well as structural variants and non-coding variants. Causal variants in some of the known
20 dystonia genes, such as the intronic retrotransposon insertion in *TAF1* (DYT3)[38] and some of
21 the larger deletions in *SGCE* [39, 40], cannot therefore be accurately screened in our study.

22

1 Finally, we present the approximate rate at which controls in the ExAC database have rare
2 coding variants in dystonia genes (Table 3 and S1 Table). These numbers are approximate
3 because individual-level data are not available from databases like ExAC and gnomad, and thus
4 proportions will be inflated as you cannot distinguish a situation where one person has two rare
5 functional variants from a situation where two people each have one rare functional variant. This
6 is especially true if one attempts to calculate what proportion of people will have a mutation in
7 any one of a set of genes, in aggregate. For example, adding up all the control carrier frequencies
8 for the 148 genes in Table 3 would indicate that 71.2% of people have qualifying variants in
9 these genes under the coding model, but because 15.5% have mutations in two or more genes, it
10 is actually only 50.5% who have a mutation in at least one of these genes. As a final caveat, all
11 sequencing data sets have their own particular parameters in terms of, for example, the kit used,
12 the sequencing quality, the processing software used, and the quality and annotation cutoffs used
13 to identify variants of interest. It is therefore imprecise to compare aggregate data across multiple
14 variants or genes from datasets such as ExAC or gnomad to one's own data, and the gold
15 standard will always be to compare sequenced cases to a large number of controls that have been
16 sequenced by the same group using identical methods. It is especially pertinent to be aware of
17 the coverage statistics, as some exons or sites may have poor or no coverage in some datasets.
18
19 In conclusion, here we present the first systematic analysis of the aggregated contribution of rare
20 variation in dystonia genes. Our analysis provides a statistical framework that can be used as the
21 field continues to develop more advanced and accurate abilities to ascertain the significance of
22 rare genetic variants in rare diseases. Such ongoing efforts include *in vitro* functional
23 phenotyping and the assembly of genotype-phenotype clinical databases (MDSgene) [41, 42].

1 Our findings provide an easily accessible resource (Table 3) and highlight the necessity of
2 considering general population frequencies when evaluating the significance of rare variants to
3 rare diseases. The approach, in principle, is also applicable to other rare diseases as genome
4 sequence data increasingly enter the equation of care for an individual patient.

5

6

7

8

9 **Acknowledgements:** We are grateful to the patients and healthy volunteers who participated in
10 this research. We thank Changrui Xiao for technical assistance and David Goldstein for advice.
11 We would like to acknowledge the following individuals or groups for the contributions of
12 control samples: K. Welsh-Bomer, C. Hulette; W. Lowe; D. Marchuk; S. Schuman, E. Nading;
13 J. Burke; S. Palmer; J. Milner; P. Lugar; C. Moylan; A. M. Diehl; M. Abdelmalek; DUHS (Duke
14 University Health System) Nonalcoholic Fatty Liver Disease Research; M. Winn, R.
15 Gbadegesin; A. Holden; D. Levy; E. Behr; D. Daskalakis; R Buckley; E. Holtzman; M. Hauser;
16 J. Hoover-Fong, N. L. Sobreira and D. Valle; A. Poduri; T. Young and K. Whisenhunt; Z. Farfel,
17 D. Lancet, and E. Pras; G. Cavalleri; N. Delanty; G. Nestadt; J. Samuels, Y. Wang; V. Shashi;
18 M. Carrington; S. Kerns, H. Oster; The Murdock Study Community Registry and Biorepository;
19 C. Woods, K. Schmader, S. MacDonald, M. Yanamadala, H. White, and Crosdaile Retirement
20 Communities; National Institute of Allergy and Infectious Diseases Center for HIV/AIDS
21 Vaccine Immunology (CHAVI) (U19-AI067854), National Institute of Allergy and Infectious
22 Diseases Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery (UM1-
23 AI100645); and the Epi4K Consortium and Epilepsy Phenome/Genome Project.

1
2 **Authors' Roles:** E.T.C. (1B,C; 2A,B,C; 3A,B) P.H. (1C; 3B), J.J. (1B,C; 3B), K.T. (1C; 3B),
3 S.T. (1C; 3B), Z.C. (1C; 3B), B.S. (1C, 3B), M.S. (1C, 3B), N.C. (1A,B,C; 2C; 3A,B)
4
5 **Financial Disclosures of all authors (for the preceding 12 months):** N.C. receives royalty
6 payment from Circuit Therapeutics (Redwood City, California), serves on the Tourette
7 Syndrome Association scientific advisory board, receives honoraria solely from non-profit
8 academic organizations and has received grants from NINDS, NIMH, Tyler's Hope for a
9 Dystonia Cure, Bachmann Strauss Dystonia Parkinson Foundation, Dystonia Medical Research
10 Foundation, McKnight Foundation, and the Harrington Discovery Institute. E.T.C. receives
11 support from the National Institute of Mental Health of the National Institutes of Health under
12 award number K01MH098126 and has received grant support from Biogen, Idec. P.H. has
13 consulted for and received grant support from Medtronic, Allergan, Merz, and GE Medical. P.H.
14 serves as Fellowship director of the Duke Movement Disorders Center, which has received
15 funding from Medtronic, Allergan, and Boston Scientific. M.S. reports grant support from the
16 Michael J. Fox Foundation, Parkinson Study Group and has received compensation as a
17 consultant for Eli Lilly, Merz, Osmotica, ProStrakan and SK Life Sciences. He serves on
18 protocol steering committees for Allergan and Biotie, and receives royalties for the "Handbook
19 of Dystonia". B.S. received grants for clinical trials from Merz, Acadia, Auspex, and CHDI. No
20 other disclosures are reported.
21
22
23

References

1. Do R, Kathiresan S, Abecasis GR. Exome sequencing and complex disease: practical aspects of rare variant association studies. *Hum Mol Genet*. 2012;21(R1):R1-9. doi: 10.1093/hmg/ddz387. PubMed PMID: 22983955; PubMed Central PMCID: PMC3459641.
2. Hu H, Roach JC, Coon H, Guthery SL, Voelkerding KV, Margraf RL, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nat Biotechnol*. 2014;32(7):663-9. doi: 10.1038/nbt.2895. PubMed PMID: 24837662; PubMed Central PMCID: PMC4157619.
3. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet*. 2013;93(1):42-53. doi: 10.1016/j.ajhg.2013.05.010. PubMed PMID: 23768515; PubMed Central PMCID: PMC3710762.
4. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*. 2010;11(6):415-25. doi: 10.1038/nrg2779. PubMed PMID: 20479773.
5. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. *Nat Genet*. 2010;42(12):1109-12. doi: 10.1038/ng.712. PubMed PMID: 21076407.
6. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science*. 2015. doi: 10.1126/science.aaa3650. PubMed PMID: 25700176.
7. Do R, Stitzel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare *LDLR* and *APOA5* alleles conferring risk for myocardial infarction. *Nature*. 2015;518(7537):102-6. doi: 10.1038/nature13917. PubMed PMID: 25487149; PubMed Central PMCID: PMC4319990.
8. Albanese A, Bhatia K, Bressman SB, Delong MR, Fahn S, Fung VS, et al. Phenomenology and classification of dystonia: a consensus update. *Mov Disord*. 2013;28(7):863-73. doi: 10.1002/mds.25475. PubMed PMID: 23649720; PubMed Central PMCID: PMC3729880.
9. Lohmann K, Klein C. Genetics of dystonia: what's known? What's new? What's next? *Mov Disord*. 2013;28(7):899-905. doi: 10.1002/mds.25536. PubMed PMID: 23893446.
10. Calakos N, Patel VD, Gottron M, Wang G, Tran-Viet KN, Brewington D, et al. Functional evidence implicating a novel *TOR1A* mutation in idiopathic, late-onset focal dystonia. *Journal of medical genetics*. 2010;47(9):646-50. doi: 10.1136/jmg.2009.072082. PubMed PMID: 19955557; PubMed Central PMCID: PMC2891583.
11. Dufke C, Sturm M, Schroeder C, Moll S, Ott T, Riess O, et al. Screening of mutations in *GNAL* in sporadic dystonia patients. *Mov Disord*. 2014;29(9):1193-6. doi: 10.1002/mds.25794. PubMed PMID: 24408567.
12. Groen JL, Andrade A, Ritz K, Jalalzadeh H, Haagmans M, Bradley TE, et al. *CACNA1B* mutation is linked to unique myoclonus-dystonia syndrome. *Hum Mol Genet*. 2015;24(4):987-93. doi: 10.1093/hmg/ddu513. PubMed PMID: 25296916.
13. Hettich J, Ryan SD, de Souza ON, Saraiva Macedo Timmers LF, Tsai S, Atai NA, et al. Biochemical and cellular analysis of human variants of the *DYT1* dystonia protein,

TorsinA/TOR1A. *Hum Mutat.* 2014;35(9):1101-13. doi: 10.1002/humu.22602. PubMed PMID: 24930953; PubMed Central PMCID: PMC4134760.

14. Kock N, Naismith TV, Boston HE, Ozelius LJ, Corey DP, Breakefield XO, et al. Effects of genetic variations in the dystonia protein torsinA: identification of polymorphism at residue 216 as protein modifier. *Human molecular genetics.* 2006;15(8):1355-64. doi: 10.1093/hmg/ddl055. PubMed PMID: 16537570.

15. Lohmann K, Uflacker N, Erogullari A, Lohnau T, Winkler S, Dendorfer A, et al. Identification and functional analysis of novel THAP1 mutations. *Eur J Hum Genet.* 2012;20(2):171-5. doi: 10.1038/ejhg.2011.159. PubMed PMID: 21847143; PubMed Central PMCID: PMC3260936.

16. Mencacci NE, R'Bibo L, Bandres-Ciga S, Carecchio M, Zorzi G, Nardocci N, et al. The CACNA1B R1389H variant is not associated with myoclonus-dystonia in a large European multicentric cohort. *Hum Mol Genet.* 2015. doi: 10.1093/hmg/ddv255. PubMed PMID: 26157024.

17. Nibbeling E, Schaake S, Tijssen MA, Weissbach A, Groen JL, Altenmuller E, et al. Accumulation of rare variants in the arylsulfatase G (ARSG) gene in task-specific dystonia. *J Neurol.* 2015;262(5):1340-3. doi: 10.1007/s00415-015-7718-3. PubMed PMID: 25825126.

18. Vemula SR, Xiao J, Zhao Y, Bastian RW, Perlmutter JS, Racette BA, et al. A rare sequence variant in intron 1 of THAP1 is associated with primary dystonia. *Mol Genet Genomic Med.* 2014;2(3):261-72. doi: 10.1002/mgg3.67. PubMed PMID: 24936516; PubMed Central PMCID: PMC4049367.

19. Zech M, Gross N, Jochim A, Castrop F, Kaffe M, Dresel C, et al. Rare sequence variants in ANO3 and GNAL in a primary torsion dystonia series and controls. *Mov Disord.* 2014;29(1):143-7. doi: 10.1002/mds.25715. PubMed PMID: 24151159.

20. Takeshita E, Saito Y, Nakagawa E, Komaki H, Sugai K, Sasaki M, et al. Late-onset mental deterioration and fluctuating dystonia in a female patient with a truncating MECP2 mutation. *Journal of the neurological sciences.* 2011;308(1-2):168-72. doi: 10.1016/j.jns.2011.06.008. PubMed PMID: 21722922.

21. Groen JL, Ritz K, Warner TT, Baas F, Tijssen MA. DRD1 rare variants associated with tardive-like dystonia: a pilot pathway sequencing study in dystonia. *Parkinsonism & related disorders.* 2014;20(7):782-5. doi: 10.1016/j.parkreldis.2014.04.002. PubMed PMID: 24768614.

22. Zech M, Jochim A, Boesch S, Weber S, Meindl T, Peters A, et al. Systematic TOR1A non-c.907_909delGAG variant analysis in isolated dystonia and controls. *Parkinsonism & related disorders.* 2016;31:119-23. doi: 10.1016/j.parkreldis.2016.07.013. PubMed PMID: 27477622.

23. Long Y, Chen Y, Qian Y, Wang J, Luo L, Huang X, et al. A rare variant in TOR1A exon 5 associated with isolated dystonia in southwestern Chinese. *Mov Disord.* 2017;32(7):1083-7. doi: 10.1002/mds.27016. PubMed PMID: 28432771.

24. Basu S, Pan W. Comparison of statistical tests for disease association with rare variants. *Genetic epidemiology.* 2011;35(7):606-19. doi: 10.1002/gepi.20609. PubMed PMID: 21769936; PubMed Central PMCID: PMC3197766.

25. Cirulli ET. The Increasing Importance of Gene-Based Analyses. *PLoS genetics.* 2016;12(4):e1005852. doi: 10.1371/journal.pgen.1005852. PubMed PMID: 27055023; PubMed Central PMCID: PMC4824358.

26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303. doi: 10.1101/gr.107524.110. PubMed PMID: 20644199; PubMed Central PMCID: PMC2928508.
28. ExomeVariantServer. NHLBI GO Exome Sequencing Project (ESP) NHLBI GO Exome Sequencing Project (ESP): Seattle, WA; URL: <http://evs.gs.washington.edu/EVS/>. Available from: URL: <http://evs.gs.washington.edu/EVS/>.
29. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92. doi: 10.4161/fly.19695. PubMed PMID: 22728672; PubMed Central PMCID: PMC3679285.
30. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26(22):2867-73. doi: 10.1093/bioinformatics/btq559. PubMed PMID: 20926424; PubMed Central PMCID: PMC3025716.
31. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *American journal of human genetics*. 2012;91(5):839-48. doi: 10.1016/j.ajhg.2012.09.004. PubMed PMID: 23103226; PubMed Central PMCID: PMC3487130.
32. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-91. doi: 10.1038/nature19057. PubMed PMID: 27535533; PubMed Central PMCID: PMCPMC5018207.
33. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7 20. doi: 10.1002/0471142905.hg0720s76. PubMed PMID: 23315928.
34. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England)*. 2003;19(1):149-50. PubMed PMID: 12499305.
35. Defazio G. The epidemiology of primary dystonia: current evidence and perspectives. *Eur J Neurol*. 2010;17 Suppl 1:9-14. doi: 10.1111/j.1468-1331.2010.03053.x. PubMed PMID: 20590802.
36. Zech M, Jech R, Havrankova P, Fecikova A, Berutti R, Urgosik D, et al. KMT2B rare missense variants in generalized dystonia. *Mov Disord*. 2017;32(7):1087-91. doi: 10.1002/mds.27026. PubMed PMID: 28520167.
37. Zech M, Boesch S, Jochim A, Weber S, Meindl T, Schormair B, et al. Clinical exome sequencing in early-onset generalized dystonia and large-scale resequencing follow-up. *Mov Disord*. 2016. doi: 10.1002/mds.26808. PubMed PMID: 27666935.
38. Makino S, Kaji R, Ando S, Tomizawa M, Yasuno K, Goto S, et al. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet*. 2007;80(3):393-406. doi: 10.1086/512129. PubMed PMID: 17273961; PubMed Central PMCID: PMCPMC1821114.

39. Asmus F, Salih F, Hjermind LE, Ostergaard K, Munz M, Kuhn AA, et al. Myoclonus-dystonia due to genomic deletions in the epsilon-sarcoglycan gene. *Ann Neurol.* 2005;58(5):792-7. doi: 10.1002/ana.20661. PubMed PMID: 16240355.
40. Zimprich A, Grabowski M, Asmus F, Naumann M, Berg D, Bertram M, et al. Mutations in the gene encoding epsilon-sarcoglycan cause myoclonus-dystonia syndrome. *Nat Genet.* 2001;29(1):66-9. doi: 10.1038/ng709. PubMed PMID: 11528394.
41. Lill CM, Mashychev A, Hartmann C, Lohmann K, Marras C, Lang AE, et al. Launching the movement disorders society genetic mutation database (MDSGene). *Mov Disord.* 2016;31(5):607-9. doi: 10.1002/mds.26651. PubMed PMID: 27156390.
42. Capetian P, Pauly MG, Azmitia LM, Klein C. Striatal cholinergic interneurons in isolated generalized dystonia-rationale and perspectives for stem cell-derived cellular models. *Front Cell Neurosci.* 2014;8:205. doi: 10.3389/fncel.2014.00205. PubMed PMID: 25120431; PubMed Central PMCID: PMCPMC4112996.

Table 1. Patient Characteristics

Men	3
Women	17
Mean age at onset	
(years)	47.4 +/- 9.59
Site of onset	
Arm	6
Leg	1
Cranial	3
Cervical	13
Site at examination	
Arm	6
Leg	1
Cranial	3
Cervical	13
Distribution	
Focal	6
Segmental	11
Multifocal	3
Family History	
Negative	16
Positive	2
Possible	2
Not Informed	0

Table 2. Characteristics of the 2,372 controls used in this study.

	n (%)
Phenotype	
Bone disease	3 (0.1%)
Cardiovascular disease	70 (3%)
Dementia caused by APOE 4/4 genotype	5 (0.2%)
Extreme birth weight	39 (1.6%)
Healthy family member of patient with bone, endocrine, hematological, metabolic, immune, or ophthalmic disease	36 (1.5%)
Healthy family member of patient with brain malformation	259 (10.9%)
Healthy family member of patient with epilepsy	548 (23.1%)
Healthy family member of patient with neuropsychiatric disease	590 (24.9%)
Healthy family member of patient with other neurodevelopmental or neurological disease	99 (4.2%)
Healthy volunteer	124 (5.2%)
Hematological disease	18 (0.8%)
Immune disorder	23 (1%)
Infectious disease	310 (13.1%)
Kidney, urological, or endocrine disease	32 (1.3%)
Liver disease	77 (3.2%)
Neurosensory disorder	3 (0.1%)
Ophthalmic disease	9 (0.4%)
Pulmonary disease	127 (5.4%)
Bone disease	3 (0.1%)
Sex	
Male	1279 (53.9%)
Female	1093 (46.1%)
Self-reported ethnicity*	
White	1920 (80.9%)
Middle Eastern	101 (4.3%)
Hispanic	13 (0.5%)
Other or unknown	338 (14.2%)
Sequencing method	
Whole genome	375 (15.8%)
37 MB Agilent All Exon	103 (4.3%)
50 MB Agilent All Exon	145 (6.1%)
65 MB Agilent All Exon	648 (27.3%)
Nimblegen SeqCap EZ V2.0 or 3.0 Exome Enrichment Kit	1101 (46.4%)

*All controls were genetically confirmed as being of white ethnicity. Ages of controls are not known.

Table 3. Percent (%) samples with qualifying variants in each of the previously reported dystonia genes for the coding (C) and likely gene disrupting (LGD) models.

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC Eur-LGD	Case % Bases Covered	Ctrl % Bases Covered
<i>TUBB4A</i>	0.00%	0.08%	0.12%	0.00%	0.04%	0.09%	95.64%	96.98%
<i>TOR1A</i>	0.00%	0.08%	0.16%	0.00%	0.08%	0.09%	82.99%	87.37%
<i>GCH1</i>	0.00%	0.13%	0.07%	0.00%	0.08%	0.01%	81.62%	83.53%
<i>GNAL</i>	0.00%	0.17%	0.05%	0.00%	0.13%	0.04%	75.46%	77.54%
<i>THAP1</i>	0.00%	0.17%	0.10%	0.00%	0.04%	0.03%	92.47%	99.06%
<i>ATP1A3</i>	0.00%	0.21%	0.17%	0.00%	0.04%	0.07%	96.78%	97.71%
<i>PRKRA</i>	0.00%	0.25%	0.12%	0.00%	0.21%	0.08%	93.28%	95.62%
<i>SGCE</i>	0.00%	0.30%	0.36%	0.00%	0.30%	0.26%	91.74%	93.99%
<i>SLC2A1</i>	0.00%	0.38%	0.21%	0.00%	0.25%	0.09%	97.31%	98.04%
<i>ANO3</i>	0.00%	0.42%	0.54%	0.00%	0.25%	0.29%	99.99%	99.85%
<i>TAF1</i>	0.00%	0.46%	0.25%	0.00%	0.30%	0.14%	99.63%	98.24%
<i>PRRT2</i>	0.00%	0.51%	0.25%	0.00%	0.38%	0.15%	94.02%	99.02%
<i>PNKD</i>	0.00%	0.55%	0.19%	0.00%	0.46%	0.13%	73.72%	89.43%
<i>CIZ1</i>	0.00%	0.55%	0.37%	0.00%	0.46%	0.25%	77.93%	92.23%
<i>GBA</i>	5.00%	0.21%	0.35%	5.00%	0.08%	0.19%	100.00%	99.81%
<i>B4GALNT1</i>	5.00%	0.63%	0.33%	0.00%	0.42%	0.23%	80.14%	89.61%
<i>TSEN2</i>	5.00%	0.67%	0.41%	5.00%	0.46%	0.24%	100.00%	99.90%
<i>AARS2</i>	5.00%	0.84%	0.72%	5.00%	0.63%	0.46%	97.34%	97.62%
<i>COL4A1</i>	5.00%	0.93%	0.79%	5.00%	0.89%	0.63%	96.74%	96.97%
<i>CTC1</i>	5.00%	0.93%	0.82%	5.00%	0.84%	0.51%	96.80%	99.26%
<i>IFIH1</i>	5.00%	1.18%	0.94%	0.00%	0.72%	0.66%	99.89%	99.85%
<i>ATN1</i>	5.00%	1.64%	0.70%	5.00%	1.43%	0.36%	85.77%	91.60%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC	Case % Bases Covered	Ctrl % Bases Covered
<i>VPS13A</i>	5.00%	2.23%	1.91%	5.00%	1.43%	1.13%	99.12%	99.14%
<i>ACTB</i>	0.00%	0.00%	0.02%	0.00%	0.00%	0.01%	99.20%	99.04%
<i>ARX</i>	0.00%	0.00%	0.02%	0.00%	0.00%	0.01%	25.16%	32.28%
<i>PLP1</i>	0.00%	0.00%	0.03%	0.00%	0.00%	0.01%	99.99%	98.67%
<i>ZC4H2</i>	0.00%	0.00%	0.03%	0.00%	0.00%	0.02%	99.56%	98.28%
<i>VAMP1</i>	0.00%	0.00%	0.06%	0.00%	0.00%	0.04%	95.36%	99.33%
<i>C19orf12</i>	0.00%	0.00%	0.08%	0.00%	0.00%	0.07%	94.27%	96.29%
<i>NDUFS8</i>	0.00%	0.00%	0.16%	0.00%	0.00%	0.08%	92.26%	97.54%
<i>SDHAF1</i>	0.00%	0.04%	0.01%	0.00%	0.00%	0.01%	1.04%	36.69%
<i>HPRT1</i>	0.00%	0.04%	0.02%	0.00%	0.00%	0.00%	95.53%	94.73%
<i>FOXG1</i>	0.00%	0.04%	0.07%	0.00%	0.04%	0.04%	66.57%	75.77%
<i>PDHA1</i>	0.00%	0.04%	0.07%	0.00%	0.04%	0.02%	91.01%	93.50%
<i>PTS</i>	0.00%	0.04%	0.09%	0.00%	0.04%	0.05%	81.17%	92.62%
<i>GNAO1</i>	0.00%	0.04%	0.14%	0.00%	0.04%	0.08%	93.99%	97.70%
<i>TTC19</i>	0.00%	0.04%	0.22%	0.00%	0.04%	0.13%	73.87%	75.94%
<i>TIMM8A</i>	0.00%	0.08%	0.01%	0.00%	0.00%	0.00%	100.00%	96.83%
<i>NPC2</i>	0.00%	0.08%	0.06%	0.00%	0.08%	0.04%	89.29%	96.89%
<i>SNCA</i>	0.00%	0.08%	0.07%	0.00%	0.08%	0.02%	100.00%	99.86%
<i>TPK1</i>	0.00%	0.08%	0.12%	0.00%	0.04%	0.10%	98.10%	99.59%
<i>VCP</i>	0.00%	0.08%	0.14%	0.00%	0.08%	0.05%	98.53%	99.16%
<i>MAT1A</i>	0.00%	0.08%	0.19%	0.00%	0.08%	0.11%	97.66%	99.31%
<i>NDUFA2</i>	0.00%	0.13%	0.08%	0.00%	0.00%	0.04%	98.51%	99.21%
<i>NDUFA12</i>	0.00%	0.13%	0.07%	0.00%	0.08%	0.03%	100.00%	99.94%
<i>WDR45</i>	0.00%	0.13%	0.09%	0.00%	0.04%	0.03%	99.84%	96.23%
<i>SLC6A8</i>	0.00%	0.13%	0.11%	0.00%	0.08%	0.04%	75.56%	84.90%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC Eur-LGD	Case % Bases Covered	Ctrl % Bases Covered
<i>CHMP2B</i>	0.00%	0.13%	0.10%	0.00%	0.08%	0.08%	100.00%	99.72%
<i>GJC2</i>	0.00%	0.13%	0.14%	0.00%	0.04%	0.10%	47.86%	44.49%
<i>SPR</i>	0.00%	0.13%	0.14%	0.00%	0.13%	0.09%	61.40%	65.91%
<i>MPV17</i>	0.00%	0.13%	0.15%	0.00%	0.08%	0.09%	100.00%	99.93%
<i>NADK2</i>	0.00%	0.13%	0.15%	0.00%	0.08%	0.13%	70.03%	78.30%
<i>PRKCG</i>	0.00%	0.13%	0.19%	0.00%	0.04%	0.07%	86.08%	95.71%
<i>VPS37A</i>	0.00%	0.13%	0.21%	0.00%	0.04%	0.09%	89.66%	97.26%
<i>DRD2</i>	0.00%	0.13%	0.23%	0.00%	0.04%	0.14%	98.85%	99.27%
<i>NDUFA9</i>	0.00%	0.13%	0.25%	0.00%	0.04%	0.15%	98.62%	99.37%
<i>UQCRCQ</i>	0.00%	0.17%	0.07%	0.00%	0.17%	0.05%	99.53%	99.15%
<i>BCAP31</i>	0.00%	0.17%	0.07%	0.00%	0.04%	0.04%	79.77%	94.31%
<i>FTL</i>	0.00%	0.17%	0.14%	0.00%	0.13%	0.08%	100.00%	99.78%
<i>NDUFS4</i>	0.00%	0.17%	0.15%	0.00%	0.04%	0.10%	99.90%	99.95%
<i>SUCLA2</i>	0.00%	0.17%	0.25%	0.00%	0.08%	0.13%	96.77%	98.94%
<i>TSFM</i>	0.00%	0.17%	0.18%	0.00%	0.17%	0.10%	28.48%	90.21%
<i>BCS1L</i>	0.00%	0.17%	0.24%	0.00%	0.17%	0.18%	100.00%	99.91%
<i>PPP2R2B</i>	0.00%	0.21%	0.18%	0.00%	0.17%	0.10%	94.04%	94.68%
<i>HIBCH</i>	0.00%	0.21%	0.26%	0.00%	0.13%	0.16%	98.85%	96.83%
<i>APTX</i>	0.00%	0.21%	0.31%	0.00%	0.04%	0.17%	100.00%	99.79%
<i>SLC6A3</i>	0.00%	0.21%	0.34%	0.00%	0.13%	0.20%	88.14%	98.17%
<i>NDUFS7</i>	0.00%	0.25%	0.10%	0.00%	0.25%	0.06%	55.04%	84.77%
<i>NDUFAF2</i>	0.00%	0.25%	0.10%	0.00%	0.13%	0.08%	99.80%	99.33%
<i>FA2H</i>	0.00%	0.25%	0.13%	0.00%	0.17%	0.09%	68.57%	77.90%
<i>SLC30A10</i>	0.00%	0.25%	0.14%	0.00%	0.21%	0.07%	74.01%	88.31%
<i>RNASET2</i>	0.00%	0.25%	0.18%	0.00%	0.21%	0.13%	86.46%	94.04%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC	Case % Bases Covered	Ctrl % Bases Covered
<i>DCAF17</i>	0.00%	0.25%	0.22%	0.00%	0.21%	0.14%	91.97%	93.21%
<i>SLC19A3</i>	0.00%	0.25%	0.30%	0.00%	0.17%	0.15%	100.00%	99.89%
<i>MCOLN1</i>	0.00%	0.25%	0.32%	0.00%	0.17%	0.19%	96.94%	96.04%
<i>AUH</i>	0.00%	0.30%	0.14%	0.00%	0.08%	0.08%	75.01%	87.71%
<i>COQ9</i>	0.00%	0.30%	0.23%	0.00%	0.21%	0.17%	92.25%	93.50%
<i>ACOX1</i>	0.00%	0.30%	0.38%	0.00%	0.17%	0.18%	99.66%	99.74%
<i>TH</i>	0.00%	0.30%	0.33%	0.00%	0.17%	0.20%	59.45%	81.84%
<i>PDX1</i>	0.00%	0.34%	0.05%	0.00%	0.25%	0.04%	48.06%	75.08%
<i>TPII</i>	0.00%	0.34%	0.13%	0.00%	0.17%	0.08%	86.09%	93.67%
<i>NDUFS3</i>	0.00%	0.34%	0.25%	0.00%	0.21%	0.20%	98.06%	99.46%
<i>DLD</i>	0.00%	0.34%	0.26%	0.00%	0.13%	0.14%	99.97%	99.91%
<i>KCNQ2</i>	0.00%	0.34%	0.31%	0.00%	0.21%	0.22%	71.78%	83.74%
<i>SUOX</i>	0.00%	0.34%	0.34%	0.00%	0.30%	0.21%	100.00%	99.94%
<i>SDHA</i>	0.00%	0.34%	0.34%	0.00%	0.34%	0.24%	96.85%	96.69%
<i>DDC</i>	0.00%	0.34%	0.34%	0.00%	0.21%	0.25%	95.05%	98.12%
<i>ARSG</i>	0.00%	0.34%	0.41%	0.00%	0.17%	0.24%	95.31%	99.59%
<i>PLA2G6</i>	0.00%	0.34%	0.50%	0.00%	0.30%	0.28%	80.30%	95.25%
<i>RNASEH2B</i>	0.00%	0.38%	0.17%	0.00%	0.17%	0.10%	89.21%	93.55%
<i>PSEN1</i>	0.00%	0.38%	0.18%	0.00%	0.21%	0.07%	100.00%	99.90%
<i>QDPR</i>	0.00%	0.38%	0.18%	0.00%	0.13%	0.11%	85.14%	90.09%
<i>NDUFAF6</i>	0.00%	0.38%	0.23%	0.00%	0.38%	0.14%	80.64%	82.89%
<i>SCP2</i>	0.00%	0.38%	0.28%	0.00%	0.38%	0.20%	96.39%	98.85%
<i>NDUFA10</i>	0.00%	0.42%	0.24%	0.00%	0.21%	0.13%	92.87%	93.74%
<i>SURF1</i>	0.00%	0.42%	0.25%	0.00%	0.30%	0.14%	87.63%	88.85%
<i>COX10</i>	0.00%	0.42%	0.32%	0.00%	0.38%	0.17%	99.81%	98.21%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC	Case % Bases Covered	Ctrl % Bases Covered
<i>SLC20A2</i>	0.00%	0.42%	0.37%	0.00%	0.34%	0.19%	99.43%	98.89%
<i>BSCL2</i>	0.00%	0.42%	0.38%	0.00%	0.21%	0.24%	99.55%	96.46%
<i>AFG3L2</i>	0.00%	0.42%	0.47%	0.00%	0.21%	0.26%	95.21%	95.50%
<i>GRIK2</i>	0.00%	0.42%	0.48%	0.00%	0.34%	0.29%	97.26%	98.51%
<i>DNAJC6</i>	0.00%	0.42%	0.54%	0.00%	0.30%	0.33%	94.77%	97.51%
<i>ARSA</i>	0.00%	0.46%	0.31%	0.00%	0.30%	0.25%	71.84%	92.18%
<i>ATP1A2</i>	0.00%	0.46%	0.39%	0.00%	0.25%	0.20%	99.01%	99.42%
<i>FBXL4</i>	0.00%	0.46%	0.37%	0.00%	0.25%	0.14%	100.00%	99.96%
<i>EARS2</i>	0.00%	0.46%	0.38%	0.00%	0.34%	0.21%	94.72%	97.48%
<i>PCCB</i>	0.00%	0.46%	0.40%	0.00%	0.34%	0.28%	84.23%	94.88%
<i>MECP2</i>	0.00%	0.46%	0.34%	0.00%	0.46%	0.20%	95.65%	86.99%
<i>GLB1</i>	0.00%	0.46%	0.43%	0.00%	0.25%	0.28%	95.54%	96.80%
<i>PARK2</i>	0.00%	0.46%	0.44%	0.00%	0.34%	0.33%	99.03%	99.49%
<i>MRE11A</i>	0.00%	0.46%	0.51%	0.00%	0.42%	0.32%	99.73%	99.33%
<i>AP4B1</i>	0.00%	0.46%	0.54%	0.00%	0.21%	0.36%	100.00%	99.95%
<i>SERAC1</i>	0.00%	0.51%	0.32%	0.00%	0.25%	0.19%	99.89%	99.80%
<i>MARS2</i>	0.00%	0.51%	0.40%	0.00%	0.34%	0.26%	99.88%	99.36%
<i>PNPT1</i>	0.00%	0.51%	0.43%	0.00%	0.34%	0.24%	99.75%	99.09%
<i>DRD5</i>	0.00%	0.51%	0.45%	0.00%	0.21%	0.31%	96.89%	98.55%
<i>PCCA</i>	0.00%	0.51%	0.51%	0.00%	0.25%	0.33%	92.52%	95.37%
<i>MTO1</i>	0.00%	0.51%	0.51%	0.00%	0.38%	0.35%	92.50%	97.02%
<i>TREX1</i>	0.00%	0.55%	0.33%	0.00%	0.30%	0.20%	99.84%	99.09%
<i>DLAT</i>	0.00%	0.55%	0.47%	0.00%	0.25%	0.26%	98.73%	99.30%
<i>PANK2</i>	0.00%	0.59%	0.34%	0.00%	0.25%	0.20%	68.77%	83.92%
<i>GCDH</i>	0.00%	0.59%	0.33%	0.00%	0.46%	0.21%	99.80%	99.18%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC	Case % Bases Covered	Ctrl % Bases Covered
<i>TBC1D24</i>	0.00%	0.59%	0.38%	0.00%	0.38%	0.24%	88.66%	90.61%
<i>COX15</i>	0.00%	0.63%	0.28%	0.00%	0.42%	0.20%	99.37%	93.77%
<i>NUP62</i>	0.00%	0.63%	0.43%	0.00%	0.51%	0.22%	98.24%	98.13%
<i>TBP</i>	0.00%	0.67%	0.10%	0.00%	0.63%	0.06%	99.77%	99.74%
<i>ATXN3</i>	0.00%	0.72%	0.24%	0.00%	0.51%	0.10%	99.26%	98.86%
<i>COASY</i>	0.00%	0.72%	0.46%	0.00%	0.42%	0.29%	96.39%	99.61%
<i>FBXO7</i>	0.00%	0.76%	0.31%	0.00%	0.55%	0.20%	89.89%	93.55%
<i>PINK1</i>	0.00%	0.76%	0.45%	0.00%	0.59%	0.33%	77.71%	79.54%
<i>FOXRED1</i>	0.00%	0.84%	0.35%	0.00%	0.51%	0.22%	98.25%	98.91%
<i>NPC1</i>	0.00%	0.89%	0.91%	0.00%	0.55%	0.46%	98.14%	98.53%
<i>MAPT</i>	0.00%	0.93%	0.47%	0.00%	0.55%	0.32%	92.67%	96.05%
<i>TRAPPC11</i>	0.00%	0.93%	0.66%	0.00%	0.72%	0.49%	100.00%	99.84%
<i>ATP13A2</i>	0.00%	0.93%	0.68%	0.00%	0.72%	0.38%	74.83%	92.07%
<i>ALS2</i>	0.00%	0.97%	0.83%	0.00%	0.67%	0.45%	98.54%	99.52%
<i>ZNF592</i>	0.00%	1.05%	0.73%	0.00%	0.72%	0.42%	99.69%	99.67%
<i>SYNJ1</i>	0.00%	1.05%	0.86%	0.00%	0.89%	0.44%	95.11%	97.11%
<i>JPH3</i>	0.00%	1.10%	0.56%	0.00%	0.67%	0.34%	71.92%	88.44%
<i>ADAR</i>	0.00%	1.10%	0.59%	0.00%	0.72%	0.37%	99.62%	99.71%
<i>CACNA1B</i>	0.00%	1.10%	0.98%	0.00%	0.63%	0.72%	28.61%	81.35%
<i>ATP7B</i>	0.00%	1.14%	1.13%	0.00%	0.89%	0.75%	99.35%	99.46%
<i>CACNA1A</i>	0.00%	1.56%	0.71%	0.00%	0.84%	0.39%	27.89%	81.14%
<i>ATM</i>	0.00%	1.56%	1.93%	0.00%	1.05%	1.12%	99.95%	99.18%
<i>SETX</i>	0.00%	2.02%	1.81%	0.00%	1.18%	0.95%	99.91%	99.94%
<i>DMXL2</i>	0.00%	2.23%	1.69%	0.00%	1.56%	1.12%	99.07%	99.63%
<i>HTT</i>	0.00%	2.36%	1.58%	0.00%	1.43%	1.02%	94.79%	96.91%

Gene name	Case Freq-C	Ctrl Freq-C	ExAC Eur-C	Case Freq-LGD	Ctrl Freq-LGD	ExAC	Case % Bases Covered	Ctrl % Bases Covered
<i>DST</i>	0.00%	4.13%	2.49%	0.00%	2.49%	1.41%	80.95%	98.29%

* Shown are the frequency of qualifying mutation carriers in cases and controls. Primary dystonia genes are shown in bold. % bases covered columns indicate proportion of CCDS bases (including 2 bp into each intron) that had least 10x coverage. ExAC-Eur column indicates approximate frequencies of variant carriers in ExAC European non-Finnish population, with the caveat that samples with two variants in one gene cannot be isolated.

Supporting information

S1 Fig. Correlation between proportion of samples with samples with qualifying variants in each gene under the coding (C) model in our locally sequenced controls and in the European non-Finnish population in ExAC ($r^2=0.85$; $p<0.001$).

S2 Fig. Correlation between proportion of samples with samples with qualifying variants in each gene under the likely gene disrupting (LGD) model in our locally sequenced controls and in the European non-Finnish population in ExAC ($r^2=0.80$; $p<0.001$).

S1 Table. Approximate percent (%) ExAC samples with qualifying variants in each of the previously reported dystonia genes for the coding (C) and likely gene disrupting (LGD) models.