Dystonia-specific mutations in *THAP1* alter transcription of genes associated with neurodevelopment and myelin

Aloysius Domingo,[1,2,3,4] Rachita Yadav,[1,2,3,4] Shivangi Shah,[3] William T. Hendriks,[2,3] Serkan Erdin,[1,4] Dadi Gao,[1,2,3,4] Kathryn O'Keefe,[1,4] Benjamin Currall,[1,2,4] James F. Gusella,[1,2,4,5,6] Nutan Sharma,[2,3] Laurie J. Ozelius,[2,3] Michelle E. Ehrlich,[7] Michael E. Talkowski,[1,2,3,4,5,*] and D. Cristopher Bragg[2,3,*]

Summary

Dystonia is a neurologic disorder associated with an increasingly large number of genetic variants in many genes, resulting in characteristic disturbances in volitional movement. Dissecting the relationships between these mutations and their functional outcomes is critical in understanding the pathways that drive dystonia pathogenesis. Here we established a pipeline for characterizing an allelic series of dystonia-specific mutations. We used this strategy to investigate the molecular consequences of genetic variation in *THAP1*, which encodes a transcription factor linked to neural differentiation. Multiple pathogenic mutations associated with dystonia cluster within distinct THAP1 functional domains and are predicted to alter DNA-binding properties and/or protein interactions differently, yet the relative impact of these varied changes on molecular signatures and neural deficits is unclear. To determine the effects of these mutations on THAP1 transcriptional activity, we engineered an allelic series of eight alterations in a common induced pluripotent stem cell background and differentiated these lines into a panel of near-isogenic neural stem cells (n = 94 lines). Transcriptome profiling followed by joint analysis of the most robust signatures across mutations identified a convergent pattern of dysregulated genes functionally related to neurodevelopment, lysosomal lipid metabolism, and myelin. On the basis of these observations, we examined mice bearing *Thap1*-disruptive alleles and detected significant changes in myelin gene expression and reduction of myelin structural integrity relative to control mice. These results suggest that deficits in neurodevelopment and myelination are common consequences of dystonia-associated *THAP1* mutations and highlight the potential role of neuron-glial interactions in the pathogenesis of dystonia.

Introduction

The dystonias are a group of complex neurologic disorders characterized by sustained involuntary muscle contractions and twisted postures.[1] A multitude of pathogenic mechanisms has been proposed, most of which relate broadly to signaling cascades within neurons.[2–4] A small number of genes have been unambiguously associated with specific phenotypes,[5,6] but the existence of shared molecular pathways disrupted across dystonic syndromes remains uncertain. Moreover, in several genes associated with dystonia, the functional implications of diverse genetic perturbations (e.g., loss-of-function and missense point mutations or copy number variation) remain largely unexplored. Establishing the neurobiological basis for dystonia across the diverse genes and mutations underlying its genetic architecture can provide insights into other neurodegenerative and/or neuropsychiatric disorders given the many anatomic, neurophysiologic, and even phenotypic overlaps in these diseases, raising the potential for broadly applicable therapeutic intervention.

The integration of induced pluripotent stem cell (iPSC) modeling and genome editing technologies offers an emerging opportunity to study mutational mechanisms in cells specifically relevant to disorders of the nervous system.[7] Combining these cellular modeling tools with transcriptomic profiling can be a powerful approach to identify shared molecular signatures that converge on common affected pathways, as well as molecular changes associated with distinct mutations.[8–10] However, there are considerable technical and biological barriers to such an approach, and there have been few examples of CRISPR modeling of disease-associated mutational mechanisms in near-isogenic neuronal models of disease.[11] Here, we applied this strategy to explore the functional impact of genetic variation in *THAP1* (Thanatos-associated [THAP]-domain containing apoptosis-associated protein 1 [MIM: 609520]), which causes familial and sporadic cases of dystonia (DYT6 [MIM: 602609]).[12]

The THAP1 protein is a zinc finger transcription factor consisting of an N-terminal DNA-binding domain, a central proline-rich region, and a C-terminal coiled-coil domain bearing a leucine zipper motif and nuclear

[1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; [2]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; [3]The Collaborative Center for X-Linked Dystonia-Parkinsonism, Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA; [4]Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; [5]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA; [6]Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA; [7]Departments of Neurology, Pediatrics, and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
*Correspondence: bragg@helix.mgh.harvard.edu (D.C.B.), mtalkowski@mgh.harvard.edu (M.E.T.)
https://doi.org/10.1016/j.ajhg.2021.09.017.

localization signal (NLS).[13–15] The cellular functions regulated by THAP1 are not fully defined, but studies in mice show links to the reciprocal repression of the pluripotency program and the promotion of neural differentiation via activation of progenitor-specific genes in embryonic stem cells.[16] Missense and truncating variants in the DNA-binding domain represent the majority of known pathogenic variation, but biophysical studies suggest that they may affect the protein in different ways by (1) altering affinity for the THAP1 DNA-binding sequence, (2) changing the specificity for this element, or (3) decreasing protein stability.[17,18] Less is known about variation within C-terminal domains, although ones that alter key residues within the NLS and leucine zipper have been suggested to interfere with nuclear translocation and homodimerization, respectively.[15,19,20] Most of these genetic variants occur exclusively in dystonia probands, although some are found in large control databases, suggesting that they represent either dystonia-associated mutations that display reduced penetrance or benign variation.

Here we engineered an allelic series of mutations in *THAP1* and performed transcriptome analyses to specifically interrogate varied mutational changes and consequent functional alterations in neural stem cells (NSCs). The NSCs represent multipotent cells within the germinal zone that precede specification into neuronal and glial lineages to give rise to neurons, oligodendrocytes, and astrocytes within the adult brain.[21,22] Transcriptional profiling identified a core set of dysregulated genes in NSCs across diverse *THAP1* mutations that converge on signatures related to neurodevelopment and myelination. We further assessed myelin status directly in mice with *Thap1* disruption and confirmed that these perturbed pathways lead to dysregulated myelin gene expression and reduced myelin structural integrity in adult animals. Our results underscore previously unappreciated neurodevelopmental aspects of *THAP1*-dystonia and highlight the potential contribution of glial dysfunction to mechanisms of neurologic diseases.

Material and methods

CRISPR engineering of THAP1 variants and NSC differentiation

We designed single-guide RNAs targeting *THAP1* mutations by using the known genomic sequence of *THAP1* (GRCh37, ENSG0000013
1931) and the Benchling CRISPR tool. Guide RNAs were synthesized by Synthego. To generate the genomic deletion, we used double-guide RNAs targeting sequences in the 5′ and 3′ untranslated regions. A karyotypically normal human male iPSC line from an individual without known disease (33362D, passage < 30) was seeded at 100,000 cells per well and maintained on mTESR1 media (StemCell Technologies,). On day 3 after plating and upon reaching 50% confluence, colonies were pretreated with nocodazole 25 ng/μL media overnight and then transfected the next day with Lipofectamine Stem (Thermo Fisher), the appropriate guide-RNA, Cas9 protein (Invitrogen), a single-stranded 100-bp oligo donor nucleotide (ssODN) template, and GFP-mRNA (TriLink Biotechnologies). Fluorescence was used as a proxy of transfection efficiency and for sorting positive cells at day 3 after transfection. GFP-positive, sorted cells were divided into "sibs," seeded into 16 wells of a 96-well plate, and maintained until confluent.[23] At this point, colonies were still a mixture of edited and unedited iPSCs. Editing efficiency within each sib was assessed via digital droplet PCR (ddPCR) and Genome Edit Detection assays (Bio-Rad). Two sibs with highest editing efficiencies were selected, divided into 16 sibs, replated, and regrown. This sib-selection method was repeated until edit efficiency within a sib reached $>5\%$ or within three rounds of assaying. Eventually, a selected sib was expanded and then sorted as single cells into wells of a 96-well plate. DNA was extracted and Sanger sequencing was performed for identification of clonal populations bearing desired mutations in *THAP1*. Independent mutation-positive lines and unedited controls were expanded (average = four per mutation). Prior to cryopreservation for future differentiation, we performed another round of Sanger sequencing to verify the mutation and clonality and subjected up to four random clones from each group to routine karyotyping (WiCell), which did not reveal any abnormalities. To identify and verify *THAP1*-deletion cell lines, we used a combination of PCR and fragment length analysis on agarose gels, Sanger sequencing, and ddPCR-based copy number assays (Bio-Rad). Guide RNAs, oligonucleotide templates, and ddPCR/Sanger sequencing primers are available from the authors upon request.

Cryopreserved iPSCs were thawed and grown on mTESR1 media for three passages prior to induction of neural differentiation. Differentiation into NSCs was performed following standard manufacturer's protocol (GIBCO). Briefly, this involves a 7-day neural induction stage until passage zero, followed by expansion. All inductions were performed with supplements from the same batch/lot, and all NSCs used for further experiments were from passage three. Differentiation experiments were performed by batch within genotypes, i.e., CRISPR editing-matched clonal populations of mutation-positive/edited and unedited lines were included in each differentiation batch. This resulted in eight differentiation batches for the study wherein each batch of samples contained up to 16 independently differentiated and clonal iPSC/NSC lines grown in parallel, obtained from balanced numbers of mutant and control samples from two different CRISPR experiments each (Table S1). To assess efficiency of neural differentiation, we subjected up to four randomly selected samples per differentiation batch to immunocytochemical staining with known markers of NSCs (Nestin, Msi1, Pax6) (Figure S1). The pluripotency marker Oct4 was negative in all samples tested (data not shown).

Library preparation for RNA sequencing and quality control

Once all differentiations were complete, RNA was extracted from all samples via Trizol lysis and a standard RNA isolation kit (QIAGEN). 1 μg of each RNA sample was taken for library preparation via the Illumina TruSeq protocol and adapters. 150 bp paired-end reads were generated with a NovaSeq S4 (Broad Institute Genomics Platform). Up to 65 million reads (mean: 38.4 million, median: 37.8 million) were generated per sample, resulting in a median of 73.2× mean per base coverage per sample. Sequence adapters were trimmed with Trimmomatic and then aligned to

the *Homo sapiens* GRCh37.75 reference sequence via the STAR Aligner (version 2.5.3).[24] As part of post-alignment quality control, two samples with >7% intergenic rate were excluded from downstream analysis. Further, mutation loci were verified by visualization with the Integrative Genomics Viewer (IGV version 2.3.94). From this, one sample was relabeled from p.Asn12Lys_5 (mutant) to p.Asn12wt_6 (control). Other quality control metrics are found in Table S1. Counts were generated from the STAR Aligner. We performed allele-specific expression by using the Genome Analysis ToolKit (GATK version 3.5). We applied the ASEReadCounter tool and a binomial test to determine significant deviation from the expected equal expression of reference and variant alleles (p value < 0.05). Off-target effects secondary to CRISPR editing were investigated with CasOFF-Finder.[25] Exonic off-target hits with up to three mismatches from the on-target site were queried via samtools for the presence of alternative alleles. Among sites investigated, a missense variant in Chr3: 119,886,818 (*GPR1*) was seen in two mutant (p.Pro26Arg_4 and p.Pro26Arg_6) and two control lines (p.Pro26wt_3 and p.Pro26wt_4); otherwise, no other relevant exonic off-target effects were identified in other groups (Table S2).

Differential expression analyses and meta-analysis

We performed differential expression analysis by using the DESeq2 package for R (version 1.26.0).[26] We performed analysis for each mutation individually by comparing profiles in mutation-positive samples versus control samples within the same NSC differentiation batches (n = 8). Our batching protocol allowed us to exploit controls from parallel genome editing experiments to increase sample size in controls while keeping analyses within batches. Thus, there were four to six mutant profiles and 11 to 13 internal control profiles per comparison in each optimized differential gene expression experiment (Table S1). Prior to performing DESeq2 analysis, data were filtered by inclusion of genes with >0.5 counts per million (cpm) in ≥60% of samples within an analysis group, resulting in an average of 16,493 "expressed genes" analyzed per group. Normalization was performed via DESEq2's median of ratios method.[26] Exploratory principal-component analysis (PCA) of uncorrected expression profiles revealed that differentiation batch effects dominated the variance. Visual comparison of principal components revealed that profiles appropriately clustered according to genotype after we applied surrogate variable (SV) analysis by using the sva package (version 3.34.0)[27] and corrected for the SVs in the regression models (Figure S2). We thus obtained differentially expressed genes (DEGs) from SV-corrected counts and by using a predefined cutoff of false discovery rate (FDR)-adjusted p value < 0.05. We used the estimated fold change (log2FC) obtained from DESeq2 and a cutoff of log2FC > 0.59 or log2FC < -0.59 to determine significant DEGs. To express log2FCs as fold-change (FC), we used $x = 2^{\text{log2FC}}$. We used Annotation DB (version 1.48.0) to obtain gene names from Ensembl IDs. All DESeq2 results can be found in Table S1. To generate models for overlapping DEGs, we estimated background expectation rates via the SuperExact R package.[28] Fisher's exact test p values were then adjusted via the Bonferroni method according to the number of overall set intersection probabilities.

To compare expression profiles in different comparisons beyond overlapping DEGs, rank scores per gene (q values) were derived from signed −(log10) transformed p values in each comparison. We then obtained Spearman's pairwise rank correlation for every comparison group by using different thresholds to subset q values and obtain correlation patterns. Subsequently, we applied a meta-analytic approach to obtain a core set of DEGs across multiple mutations in the DNA-binding domain and NLS. First, per-gene two-sample means *Z* score testing was performed between mutant and control SV-corrected counts. These scores were then weighted according to the total number of samples within that comparison.[29] Meta-analysis *Z* scores per gene were obtained via weighted *Z* scores from the following mutational groups: p.Asn12Lys, p.Ser21Thr:het, p.Pro26Arg, p.Cys54Tyr, and p.Arg146Aspfs*9. From these, $_{\text{JOINT}}$DEGs were determined with a cutoff of Benjamini-Hochberg (BH)-adjusted meta-analysis p value < 0.01 and a sign test, which required all individual components to be in the same direction of expression change. All meta-analysis results can be found in Table S3.

Functional enrichment analyses and weighted gene coexpression network analysis

We performed Gene Ontology (GO) term enrichment analysis of $_{\text{JOINT}}$DEGs and $_{\text{DEL}}$DEGs (FDR < 0.01) by using the companion R package for g:Profiler2 (version 0.2.0).[30] All GO terms enriched at FDR < 0.05 are reported in Table S4. To reduce these lists for visualization, we applied semantic similarity by using the SimRel measure in REVIGO to generate a small list (allowed similarity 0.5) of related significantly enriched GO terms.[31,32] We additionally manually curated reduced lists (from $_{\text{DEL}}$DEGs) to show common enriched GO terms in up- and downregulated DEGs. Consensus THAP1 targets were obtained from MSigDB: M30208, which derived results from ChIP-seq experiments uploaded into the Gene Transcription Regulation Database.[33] We applied a hypergeometric test to determine significant overlap with $_{\text{JOINT}}$DEGs and $_{\text{DEL}}$DEGs.

We performed coexpression analysis by using the Weighted Gene Coexpression Network Analysis (WGCNA) R package (version 1.69).[34] We subsetted expression count matrices (n = 87 samples after removal of outliers) to include only protein-coding genes (total genes = 12,661 genes) and then used them to construct signed networks from adjacency and topological overlap matrices at soft power threshold of 15, which we calculated to reduce mean connectivity < 100, as suggested by package authors. This resulted in 59 modules of correlated genes, which was then reduced to 47 after merging with the mergeCloseModules function. An additional custom filtering step was then applied wherein only genes with module membership p value of <0.10 were considered to belong to their assigned modules; otherwise, genes were assigned to M0 ("gray"/unclassified module, n = 702 genes) (Table S5). Modules of interest were identified by obtaining the biweight mid-correlation between the module eigengene (ME) and *THAP1* expression[35] and by testing for significant enrichment of $_{\text{JOINT}}$DEGs and $_{\text{DEL}}$DEGs in each module via a Fisher's test (p value < 0.10). To derive functional insights from these modules of genes, we performed GO term enrichment analysis as above (Table S5). MSigDB was likewise queried for significant overlap with hallmark non-cancer-related gene sets (version 7.2, build date September 2020). Additionally, significant overrepresentation of module genes in custom lists was tested. These lists were assembled on the basis of hypotheses derived from known functions of *THAP1* as well as enriched GO terms in modules of interest and included a list of genes related to lysosomal regulatory pathways,[36] a list of human myelin-related genes,[37] polycomb genes related to developmental regulators in human stem cell populations,[38] and modules of genes in different periods

of brain development derived from functional genomic analyses (PsychEncode dataset).[39] Overrepresentation in these and other brain-related gene lists was investigated via Fisher's tests, applied through the WGCNA and anRichment R packages (version 1.18), and adjusted for multiple testing. For all enrichment testing (GO/custom lists/DEGs), custom background sets corresponding to genes assayed ("expressed genes," mean 16,303 genes across all comparisons) or to protein-coding genes (12,661 genes, for enrichment testing in WGCNA-derived results) were used.

Protein-protein interaction (PPI) networks were constructed with the stringApp application for Cytoscape (version 3.8.2) and from (1) protein-coding $_{JOINT}$DEGs and $_{DEL}$DEGs and (2) *THAP1* coexpressed genes (p value < 0.01, n = 458 genes) from the WGCNA results. For both networks, only high-confidence interactions were included (stringDB interaction score > 0.7) and singleton nodes were omitted, as well as subnetworks with less than four nodes. Nodes/genes intersecting with $_{JOINT}$DEGs and $_{DEL}$DEGs were annotated manually. We performed functional enrichment within the application and by using only nodes/genes from the main/largest network.

Genetic mouse models for myelin gene expression assays and staining

Experiments in mice were in compliance with the United States Public Health Service's Policy on Humane Care and Use of Experimental Animals and were approved by the Institutional Animal Care and Use Committee at Icahn School of Medicine at Mount Sinai. Genetic mouse models included heterozygote *Thap1*$^{p.Cys54Tyr/+}$ (p.Cys54Tyr-KI) and *Thap1*$^{-/+}$ (ex2-KO) mice, backcrossed over 10 times onto a C57Bl6/J background. Controls were wild-type (WT) littermates of each mutant genotype. For targeted gene expression assays of myelin-specific genes, tissue from the cortex, cerebellum, and striatum of postnatal day 21 (P21, n = 5 p.Cys54Tyr-KI, n = 7 WT littermates; n = 10 ex2-KO, n = 12 WT littermates) and 2-month-old mice (adult, n = 4 per group) were lysed and homogenized in QIAzol Lysis Reagent (QIAGEN). Total RNA purification was performed with the miRNeasy Micro Kit (QIAGEN) according to manufacturer's instructions. 500 ng of RNA were reverse-transcribed with the High Capacity RNA-to-cDNA Kit (Applied Biosystems). Real-time quantitative PCR was then performed on resulting cDNAs in a Step-One Plus system (Applied Biosystems) with PerfeCTa SYBR Green FastMix ROX (Quanta BioSciences). Quantitative PCR for each target consisted of 40 cycles, each with 15 s at 95°C and 30 s at 60°C, followed by dissociation curve analysis. Primer oligonucleotide sequences are available upon request. Gene expression levels were obtained via the ΔΔCt method (normalized to *Gapdh* expression) and then expressed as mean FC relative to the mean expression in the control group.

For immunocytochemical staining of myelin-specific genes, mice (adult, n = 4; p.Cys54Tyr-KI, n = 5; ex2-KO, n = 7 WT) were anesthetized in a CO_2 chamber and transcardially perfused with ice-cold phosphate buffer solution (PBS). Brains were post-fixed in 4% polymeric formaldehyde and sliced into 30 μm sections with a vibratome (Leica) for histological analyses. Sections were permeabilized with 0.1% Triton X-100 in PBS, blocked with PBS/5% goat serum, and then incubated with primary antibodies: Mobp (1/1,000, rabbit, Abgent) and Cnp (1/1,000, mouse, Abcam) prior to exposure to fluorescent Alexa secondary antibodies. Immunoreactive proteins were visualized with Enhanced Chemiluminescence on a Fujifilm LAS4000 imaging device, and optical densities of protein of interest were measured and analyzed with Fiji software (ImageJ). For non-fluorescent myelin immunostaining, we used luxol fast blue (LFB) stain to measure myelin structural integrity in the corpus callosum of adult mice (n = 4 p.Cys54Tyr-KI, n = 5 WT littermates; n = 8 ex2-KO, n = 8 WT littermates). Briefly, sections were dried for 30 min before being immersed in a de-fat solution (1:1 alcohol/chloroform for 2 h), after which, sections were transferred to 0.1% LFB overnight at 56°C. After several washes in 95% ethyl alcohol, followed by distilled water, the sections were first differentiated in a lithium carbonate solution for 30 s, followed by a second differentiation in 70% ethyl alcohol for 30 s. Sections were checked microscopically for verification of the sharp definition of gray versus white matter before being dehydrated in graded alcohol baths, cleared in xylene, and mounted with resinous medium. Images were obtained with an Olympus BX61 microscope and analyzed with Fiji software (ImageJ). Quantification was performed and absolute values were normalized to mean WT quantification, which was set to 100%. We used one-sided Student's t tests to determine significance of percent stain relative to WT in mutant and WT groups.

Results

Generation of an allelic series of *THAP1* mutations in neural stem cells

To define the transcriptional changes of *THAP1*-associated mutational mechanisms, we focused primarily on mutations within the DNA-binding domain given that (1) 70% of pathogenic sequence variation associated with *THAP1*-dystonia are found in this module[40] and (2) mutations in this region have been proposed to directly alter the protein's transcriptional activity[14,18] (Figure 1A). We chose genetic variants on the basis of their distribution within the three-dimensional structure of this domain, which is composed of an anti-parallel β sheet and a helix-loop-helix motif.[18,41] This allelic series consisted of four missense mutations: c.36C>A (p.Asn12Lys) and c.61T>A (p.Ser21Thr), which are localized within the first loop (L1) of this domain; c.77C>G (p.Pro26Arg) within the second loop (L2); and c.161G>A (p.Cys54Tyr) within the β sheet.[17] All four mutations are (1) documented in dystonia probands;[42–44] (2) absent from population-based genetic references in the Genome Aggregation Database (gnomAD);[45] (3) predicted to be damaging across multiple algorithms (Combined Annotation Dependent Depletion [CADD], Mendelian Clinically Applicable Pathogenicity [M-CAP], Rare Exome Variant Ensemble Learner [REVEL]);[46–48] and (4) classified as pathogenic on the basis of existing criteria[49] (Figure 1A). For comparison, we also engineered iPSCs bearing two genotypes to model *THAP1* loss of function and haploinsufficiency, respectively: (1) a pathogenic dystonia-specific frameshift mutation, c.436_443del (p.Arg146Aspfs*9),[50] that results in a prematurely terminated protein and prevents nuclear translocation by abolishing the NLS and (2) a deletion allele (DEL) in which the entire coding sequence was excised via a dual guide-RNA approach. To date there has been one report of a family affected with a severe and atypical form of dystonia

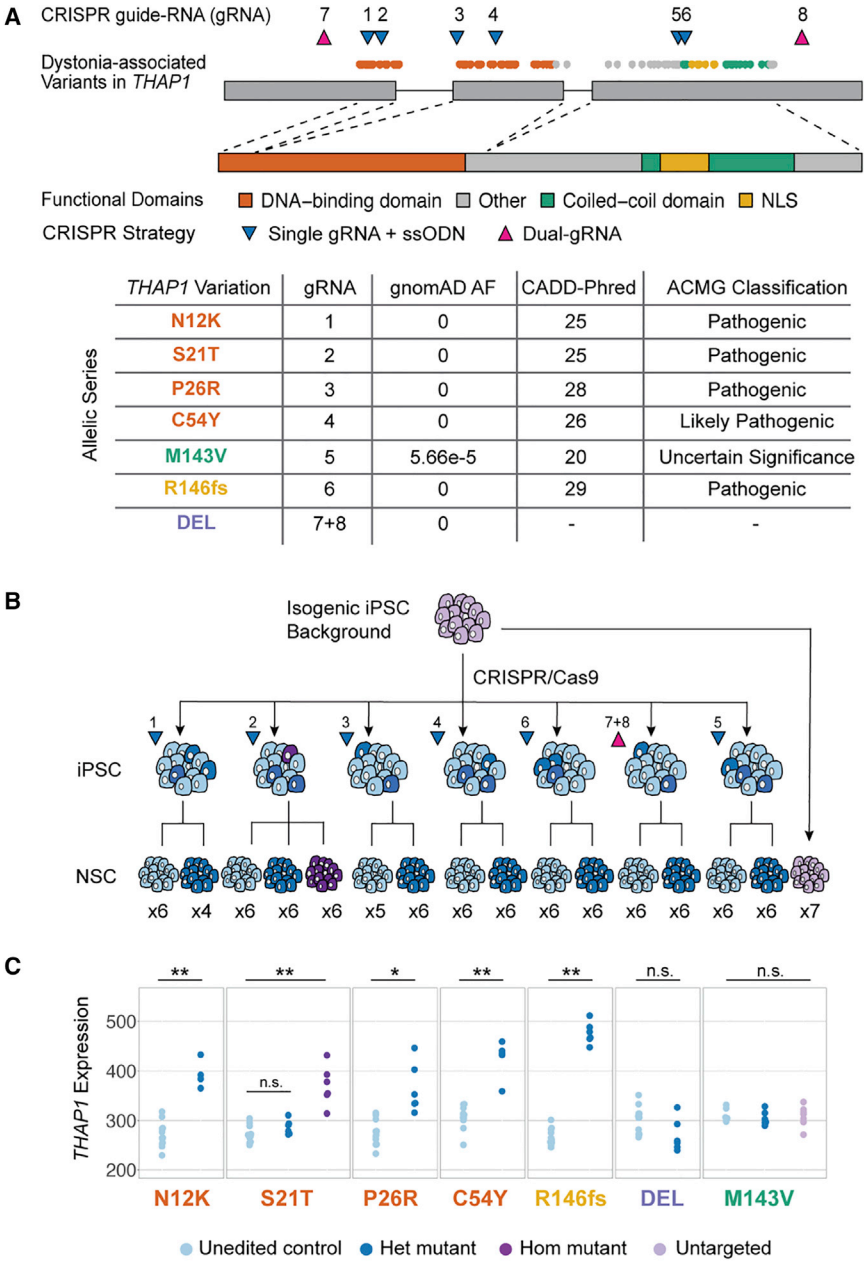

Figure 1. Generation of an allelic series of *THAP1* mutations in neural stem cells

(A) Dystonia-associated genetic variation in *THAP1* localize within known functional domains of the protein, with 70% of disease-specific, pathogenic mutations corresponding to residues in the DNA-binding domain. To create an allelic series, we introduced four single-nucleotide variants in this domain (c.36C>A [p.Asn12Lys, N12K], c.61T>A [p.Ser21Thr, S21T], c.77C>G [p.Pro26Arg, P26R], c.161G>A [p.Cys54Tyr, C54Y]) and one in the coiled-coil domain (c.427A>G [p.Met143Val, M143V]) via CRISPR-Cas9-based genome editing. We used single-guide RNAs (gRNAs) in combination with a single-stranded oligo donor nucleotide (ssODN) template to generate the desired mutations. An 8 base pair small deletion resulting in a frameshift (c.436_443del [p.Arg146Aspfs*9, R146fs]) in the NLS was generated with the same approach. Among these, mutations in the DNA-binding domain and the frameshift in the NLS are considered pathogenic on the basis of absence in gnomAD, a large population-based database of exome sequences that is depleted of individuals with neurologic disorders, bioinformatic predictive methods, including the CADD Phred-score, and classification with criteria from the American College of Medical Genetics and Genomics (ACMG). The p.Met143Val is classified as a variant of unknown significance via these criteria. Additionally, we used gRNAs placed in untranslated regions of the gene to engineer a haploinsufficiency model via deletion of the entire coding sequence of *THAP1* (DEL). The AF column denotes allele frequency in gnomAD v.2.1.1.

(B) Mutations were generated from a single human induced pluripotent stem cell (iPSC) line ("background"), making the models isogenic beyond the mutation introduced in *THAP1* and effectively isolating mutational effects in subsequent experiments. Multiple independently edited and matched unedited cells from each CRISPR experiment were grown as clonal iPSC lines and then differentiated in parallel into NSCs. For each mutation, up to six edited NSC lines and up to six unedited control lines were used. Additionally, six NSC lines were generated from an iPSC line bearing a homozygous c.61T>A (p.Ser21Thr) mutation. Seven NSC lines that were directly differentiated from the background line were also used as controls (total n = 94 NSC lines). See also Figure S1.

(C) *THAP1* expression was assessed in each mutation versus corresponding control comparison. The results recapitulate previously described upregulation of *THAP1* in lines carrying pathogenic genetic variation, i.e., heterozygous (Het) p.Asn12Lys, p.Pro26Arg, p.Cys54Tyr, and p.Arg146Aspfs*9 and homozygous (Hom) p.Ser21Thr variations. The DEL model induced minimal reduction of *THAP1* expression, while gene expression was unchanged in NSCs with the p.Met143Val variation. *THAP1* expression plotted as normalized counts. Wald test statistic from differential expression analysis: *p value < 0.05; **FDR < 0.05; n.s., not significant.

carrying a segregating ~500-kilobase genomic deletion of the 8p11.21 locus that included all exons of *THAP1*,[51] and there are no *THAP1* gene deletions in gnomAD-SV v.2.1.[52] Finally, we generated iPSCs bearing a missense mutation, c.427A>G (p.Met143Val), within the C-terminal coiled-coil domain. Although this genetic variant has been described recurrently in dystonia families,[53–55] its frequency is relatively high in gnomAD, i.e., above the maximum tolerated allele frequency in a reference database for *THAP1* and for an estimated disease penetrance of 60%.[56] Thus, we hypothesized that this variation may be benign and would behave as a control.

The CRISPR-targeting method across seven genome-editing sites produced individual iPSC clones that were heterozygous for each mutation. For the c.61T>A locus, the editing strategy generated clones with heterozygous (p.Ser21Thr:het) and homozygous (p.Ser21Thr:hom) alterations in the same experiment, and we included the latter in

transcriptional profiling to investigate potential dosage effects. Multiple independently edited iPSC clones and matched unedited controls that were exposed to the same guide RNAs, as well as controls that were directly isolated from the background, were each independently differentiated into NSCs in parallel, grouped in predefined differentiation batches (four to seven NSC lines per genotype per mutation, total n = 94 NSC lines) (Figure 1B). There were no obvious growth or morphologic differences observed between mutant and control iPSC/NSC lines, and equivalent expression of neural markers was seen across differentiation batches (Figure S1). Based on previous reports of *THAP1* autoregulation,[16,57] an interesting aspect of these analyses was the opportunity to directly quantify the expression of *THAP1* across all mutations in the allelic series. From these analyses, we were able to note (1) significant upregulation of *THAP1* when disrupted by a dystonia-specific pathogenic variant (mean FC = 1.4 across lines with a heterozygous p.Asn12Lys, p.Ser21Thr, p.Pro26Arg, p.Cys54Tyr, or p.Arg146Aspfs*9 alteration); (2) minimal impact of complete deletion of the *THAP1*-coding sequence on expression, suggesting strong autoregulation of *THAP1* (FC = 0.89, p value 0.08); and (3) no change in *THAP1* expression induced by the p.Met143Val variation (Figure 1C). While allele-specific expression was not significantly different between reference and variant alleles in the p.Cys54Tyr, p.Pro26Arg, p.Ser21Thr:het, and p.Met143Val groups, expression of the reference allele was moderately increased in the p.Asn12Lys lines (mean fraction of reference allele/total = 0.72; p value 0.05, binomial test).

Pathogenic variation in the *THAP1* DNA-binding domain and nuclear localization signal induce overlapping transcriptome-wide alterations

The mutations we engineered localize to different THAP1 structural and functional domains. We therefore asked whether they produced distinct patterns of transcriptional dysregulation in differential expression analyses within experimental groups and by matching mutations to unedited controls within differentiation batches. After controlling for variance in global expression profiles associated with batch effects from editing and differentiation, PCAs revealed that all profiles clustered according to genotype (Figure S2). For each comparison, we identified a set of DEGs (FDR < 0.05, designated hereafter as significant if FC > 1.5). Full gene deletion resulted in the most profound transcriptome-wide perturbation (2,622 $_{DEL}$DEGs, 386 significant), followed by introduction of a homozygous c.61T>A missense mutation (533 $_{p.Ser21Thr:hom}$DEGs, 130 significant) (Figure 2A and Table S1).

In the CRISPR editing and neural differentiation experimental design, mutants and controls were independently differentiated in batches defined by CRISPR-edited genotypes. This grouping facilitated differential expression analyses and subsequent enrichment tests for shared DEGs across mutations by comparison to null expectation models.[28] Intersections of DEGs induced by any two heterozygous mutations localized to either the DNA-binding domain or NLS were always significantly above expectation (minimum 9.6-fold enrichment of observed/expected overlapping DEGs, Bonferroni-corrected p value [FWER] 1.1×10^{-6} [$_{p.Asn12Lys}$DEGs-$_{p.Ser21Thr:het}$DEGs], Table S1). This observation was consistent even when considering multi-set intersections (i.e., overlaps across three to six DEG sets), indicating a significant overlap in the transcriptional consequences of mutations affecting these functional domains. The most consistently observed effect was downregulation of *GALC* [MIM: 606890], *ZNF785*, and *DPH6* [MIM: 618391] across all six mutational groups (3.7×10^{9}-fold enrichment over expected, FWER 1.6×10^{-26}) and dysregulation of dosage-sensitive DEGs induced by the p.Ser21Thr:het/hom variation (7.7-fold-enrichment of observed/expected overlapping DEGs, FWER 2.4×10^{-15}), where 78% of overlapping DEGs displayed greater FCs associated with the homozygous alteration (Figure S3 and Table S1). Importantly, while only 0.6% of $_{DEL}$DEGs intersected with DEGs induced by any mutation localized to either the DNA-binding domain or NLS, closer inspection of expression profiles indicated consistent expression changes in the DEL group at nominal thresholds (Figure 2B and Figure S3). These results suggest that pathogenic *THAP1* mutations affecting the DNA-binding domain and NLS exert highly concordant transcriptomic effects and that, although these molecular signatures were largely distinct from those observed in haploinsufficiency, there were potentially convergent signatures across all mutational mechanisms.

We further explored the common transcriptional signatures by using pairwise correlations of extreme rank orders of genes (i.e., upper and lower 2.5% of q values, corresponding to up to the 830 most differential genes in every set). In this subset of expression profiles, the rank orders in DNA-binding domain mutants correlated with each other and with those in the NLS group (mean Spearman r = 0.56; p value < 0.05; outlier: p.Cys54Tyr-p.Arg146Aspfs*9) but not with DEL or p.Met143Val q values (Figure 2C). Further, the correlation patterns were maintained at different thresholds (Figure S4), suggesting a robust intersecting signature induced by the p.Asn12Lys, p.Ser21Thr, p.Pro26Arg, p.Cys54Tyr, and p.Arg146Aspfs*9 variations that extend beyond the few overlapping significant DEGs already discovered. Meta-analysis with signed weighted *Z* scores further defined a molecular signature consisting of 871 genes (termed subsequently as $_{JOINT}$DEGs, representing meta-analysis FDR < 0.01) that were consistently altered in a direction-specific manner across these dystonia-specific pathogenic mutations in the DNA-binding domain and the NLS (Figure 2D and Table S3).

Functional analysis reveals convergent signatures of neurodevelopment and myelination across diverse *THAP1* mutations

To identify potential cellular consequences of *THAP1* transcriptional dysregulation, we performed functional

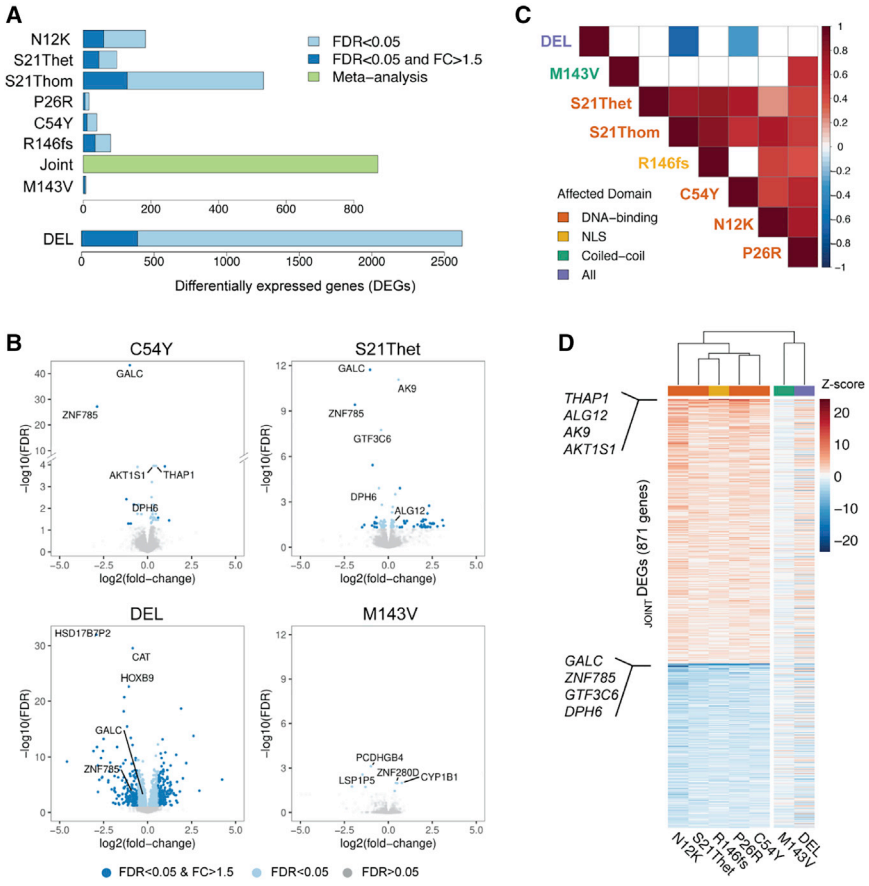

Figure 2. Pathogenic variation in the *THAP1* DNA-binding domain and nuclear localization signal induce overlapping transcriptome-wide alterations

(A) Differential expression analysis was performed within mutational groups independently, identifying a set of robust and significant DEGs for each comparison (FDR < 0.05, FC > 1.5). Full gene deletion resulted in the most profound transcriptome-wide perturbation (2,622 $_{DEL}$DEGs, 386 significant), followed by introduction of a homozygous c.61T>A missense mutation (533 $_{p.Ser21Thr:hom}$DEGs, 130 significant). $_{JOINT}$DEGs are genes identified via meta-analytic combination of results in p.Asn12Lys, p.Ser21Thr:het, p.Pro26Arg, p.Cys54Tyr, and p.Arg146Aspfs*9 groups.

(B) Volcano plots showing DEG discovery in selected mutational groups: p.Cys54Tyr and p.Ser21Thr:het (representing DNA-binding domain variation), DEL, and p.Met143Val, with notable overlapping DEGs induced by mutations localized to the DNA-binding domain labeled (i.e., p.Cys54Tyr and p.Ser21Thr:het; upregulation of *THAP1*, *ALG12*, *AK9*, and *AKT1S1* and downregulation of *GALC*, *ZNF785*, *GTF3C6*, and *DPH6*). These were also the most consistently dysregulated genes in the joint analysis. $_{DEL}$DEGs intersected with few DEGs induced by other pathogenic mutations (e.g., *GALC* and *ZNF785*). In contrast, there were only a few DEGs induced by the p.Met143Val variation. See also Figures S2 and S3.

(C) Pairwise correlations of extreme rank orders of genes revealed a striking overlap in the expression signatures induced by missense mutations in the DNA-binding domain and by the frameshift in the NLS (mean Spearman rank correlation r = 0.5, p values < 0.05; outlier: p.Cys54Tyr-p.Arg146Aspfs*9), consistent with the observation of overlapping DEGs discovered in these mutational groups. See also Figure S4.

(D) Meta-analytic combination of profiles discovered 871 genes ($_{JOINT}$DEGs, meta-analysis FDR < 0.01) that are consistently altered in a direction-specific manner by heterozygous mutations localized in the DNA-binding domain or NLS (i.e., p.Asn12Lys, p.Ser21Thr:het, p.Pro26Arg, p.Cys54Tyr, p.Arg146Aspfs*9 mutational groups). This signature was distinct from $_{DEL}$DEGs, and gene expression was not altered by the p.Met143Val variation, as shown in this heatmap of meta-analysis *Z* scores across different comparisons. Genes with the highest/lowest *Z* scores represent the most consistently dysregulated genes and include *THAP1*, *ALG12*, *AK9*, *AKT1S1*, *GALC*, *ZNF785*, *GTF3C6*, and *DPH6*.

enrichment analyses of the discovered DEGs: (1) genes identified via meta-analysis ($_{JOINT}$DEGs), representing the common signature of genetic variation in the DNA-binding domain and NLS, and (2) $_{DEL}$DEGs, representing haploinsufficiency. The analyses revealed significant enrichment of multiple GO terms primarily related to transcription regulation in $_{JOINT}$DEGs (FDR < 0.05, Fisher's test, Figure 3A and Table S4), consistent with the characterization of THAP1 as a transcription factor.[14,58] Further, $_{JOINT}$DEGs overlapped significantly with consensus THAP1 target genes[33] as expected (FDR 1.8×10^{-9}) given that the mutational groups that comprise this aggregation represent genetic variants that functionally impact THAP1 DNA binding or impair nuclear import.[14,17,19] Meanwhile, $_{DEL}$DEGs were strongly related to developmental processes including nervous system development (FDR 2.3×10^{-7}) (Figure 3B and Table S4), recapitulating previous observations in Thap1-haploinsufficient mouse embryonic stem cells.[16,59] Intriguingly, enrichment analysis in $_{DEL}$DEGs also revealed enrichment of terms related to lipid metabolism and synthesis (e.g., steroid biosynthetic process, FDR 3.4×10^{-5}).

Although significant $_{JOINT}$DEGs and $_{DEL}$DEGs had minimal overlap, a PPI network constructed from these DEGs identified significant PPIs among dysregulated genes (PPI enrichment p value 1.0×10^{-16}, Figure S5). Functional analysis of this network revealed enrichment of GO terms related to development (FDR 1.0×10^{-16}) and neural differentiation (FDR 1.2×10^{-13}). Gene annotations of $_{JOINT}$DEGs were also enriched for terms related to neuron development (e.g., neural tube development FDR 0.02, Figure 3A) and overrepresented with genes exhibiting selective pressure and loss-of-function intolerance (FDR 0.01, genes with upper bound of the observed/expected confidence interval < 0.25 in gnomAD) that are enriched in neurodevelopmental disease.[45,60,61] Further, query for

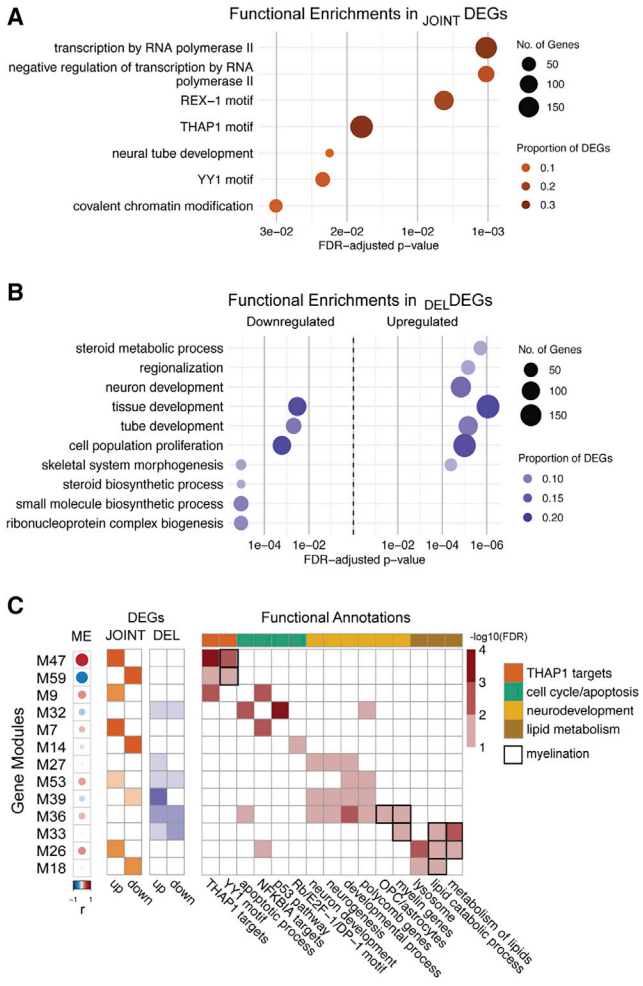

Figure 3. Functional analysis reveals convergent signatures of neurodevelopment and myelination across diverse *THAP1* mutations

(A) Functional analysis via GO term enrichment testing in DEGs identified transcription-related terms in $_{JOINT}$DEGs. Additionally, "neural tube development" was an enriched GO term. As expected, the THAP1-binding motif was overrepresented in these DEGs, given that the mutations that comprise the meta-analysis functionally alter THAP1 binding to its targets. YY1, which co-binds THAP1-bound sequences and is a known regulator of oligodendrocyte development, is also overrepresented, as well as REX-1, a transcription factor that has known activity in stem cells. All terms plotted have FDR < 0.05 in Fisher's tests for enrichment.

(B) $_{DEL}$DEGs were enriched for GO terms related to development, including specifically, "neuron development" (FDR 1.5×10^{-5}), as well as "cell population proliferation," consistent with known functions of THAP1. $_{DEL}$DEGs were not enriched for THAP1 targets. Additionally, $_{DEL}$DEGs were enriched for GO terms related to lipid metabolism or synthesis (in both up-/downregulated DEGs, FDR < 1.0×10^{-5} in both). All terms plotted have FDR < 0.01 in Fisher's tests for enrichment. In both of these enrichment plots (A and B), circle size represents the number of intersecting genes between each gene set (y axis) and DEGs, while color/shade represents the proportion of DEGs harboring a certain gene set term. For visualization, GO terms were reduced from the entire list of enriched terms via semantic similarity measures, as described in the material and methods.

(C) Modules of co-expressed genes where the module eigengene (ME) correlates/anticorrelates with *THAP1* expression (r, biweight mid-correlation) were probed for enrichment of $_{JOINT}$DEGs and $_{DEL}$DEGs (all shaded boxes corresponding to DEGs with p value < 0.01 in Fisher's test for enrichment). Modules were then annotated with GO terms and custom lists, revealing significant enrichment of terms that are related to known THAP1 functions in humans and mice (cell cycle control, apoptosis, and neural differentiation). This analysis also revealed modules that were associated with lipid metabolism and lysosomal function, as well as overlapping modules that are related to myelination. The module that is most correlated with *THAP1* expression (i.e., M47) includes *THAP1* itself and is enriched for the THAP1 motif (FDR 0.01) and for the YY1 motif (FDR 0.02), which include genes known to participate in oligodendrocyte development and differentiation, a critical pathway in myelin formation. Enrichment of annotations FDR < 0.10 when shaded in the heatmap with Fisher's test for enrichment; annotations/terms were curated for visualization, and each are defined in the material and methods.

overrepresented transcription-factor-binding motifs in $_{JOINT}$DEGs identified not only the THAP1 motif but also elements for YY1 and REX-1 (both FDR < 0.05, Figure 3A), both of which have roles in neurodevelopment.[62–65] Taken together, functional enrichment analyses suggest that convergent pathways and networks related to development characterized both $_{JOINT}$DEGs and $_{DEL}$DEGs, and imply neurodevelopmental consequences resulting from diverse classes of *THAP1* mutations and across different levels of *THAP1* dysregulation. This effect was more pronounced in haploinsufficiency, as previously shown.

To further probe the functional impact of *THAP1* perturbation, we performed gene coexpression network-based analyses to identify the modules of co-regulated genes representing active biological processes in our dataset.[66] We correlated module eigengenes (MEs) with *THAP1* expression, then layered with significant overlap with $_{JOINT}$DEGs and $_{DEL}$DEGs, to select the gene modules that were impacted by our mutational modeling (Table S5). The two modules with MEs most correlated with *THAP1* expression (module 47 [biweight midcorrelation r = 0.58; FDR 1.4×10^{-7}] and module 59 [r = −0.50; FDR 1.7×10^{-5}]) contained the highest number of $_{JOINT}$DEGs and were strongly enriched for THAP1 DNA targets, as expected. Overall, DEGs were distributed across 13 *THAP1*-associated modules that were not necessarily enriched with known THAP1 targets, suggesting that the transcriptional effects captured by our profiling and analyses were not exclusive to direct regulation by THAP1 but may also include downstream and secondary effects. Annotation of gene modules revealed functions that have previously been described for THAP1 in human and/or mouse cells, such as apoptosis and cell-cycle control via the p53 pathway and the pRB/E2F transcription factor, and neurodevelopmental processes (Figure 3C),[13,58,59,67] showing that our workflow and analyses are able to recover expected effects based on known THAP1 biology. Expanding and recapitulating the results of enrichment analyses performed on DEGs, we also discovered *THAP1*-associated gene modules that were functionally related to lipid metabolism, lysosomal function, glial development (i.e., oligodendrocyte progenitor cells [OPCs] and astrocytes),

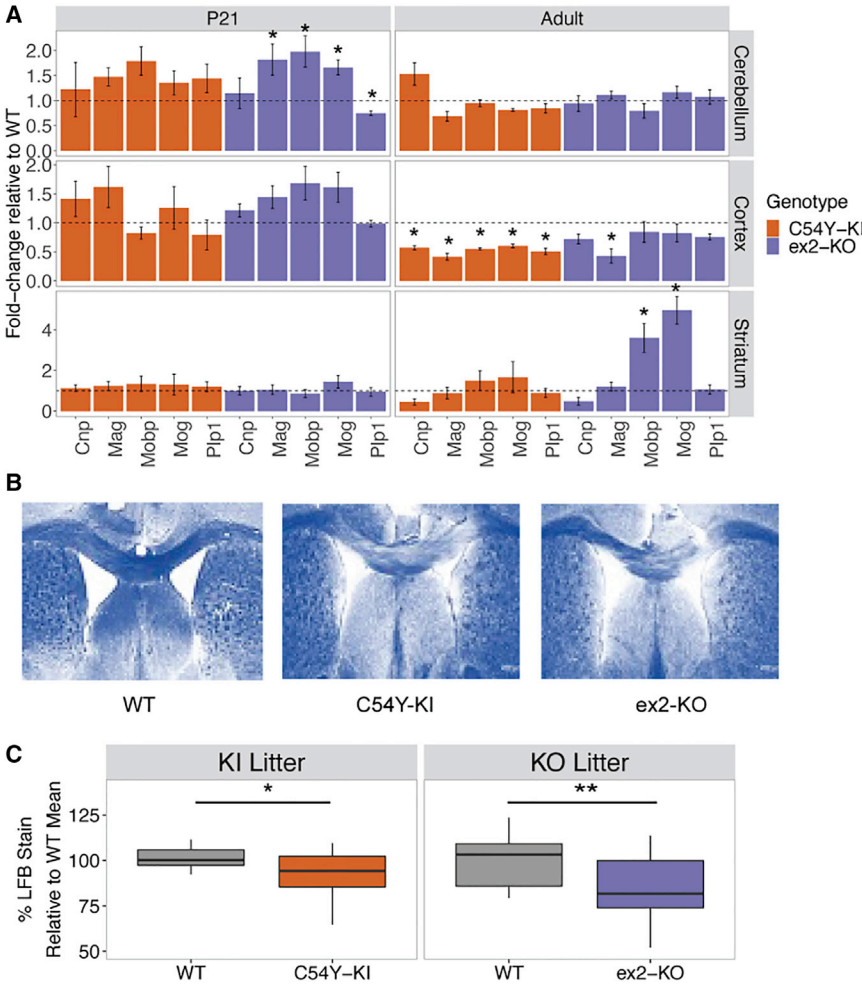

Figure 4. *Thap1* disruption is associated with defects in myelination in mouse models

(A) Quantitative PCR assays revealed significant decreases in the expression of myelin-related genes in the cortex of 2-month-old mice (adult) with heterozygous knockin of *Thap1* p.Cys54Tyr (p.Cys54Tyr-KI, $Thap1^{p.Cys54Tyr/+}$) and deletion of exon 2 (ex2-KO, $Thap1^{-/+}$) compared to WT littermates. These reductions were not seen in younger mice (postnatal day 21 [P21]). Further, significant increases in gene expression in the cerebellum of P21 mice as well as in the striatum of adult mice with ex2-KO was seen. For P21, n = 5 p.Cys54Tyr-KI, n = 7 WT littermates; n = 10 ex2-KO, n = 12 WT littermates. For adult, n = 4 mice per group. Quantification is shown as FC against mean expression in WT littermates. Error bars represent standard error of the mean, and significance was tested via Student's t tests on normalized expression values in mutant versus WT groups: *p value < 0.05. See also Figure S6.

(B) Representative LFB staining of myelin structural integrity in the corpus callosum of adult WT, p.Cys54Tyr-KI, and ex2-KO mice, showing myelin loss in adult mice with mutant *Thap1* alleles.

(C) Quantification of LFB stain intensity in 25 adult mice (n = 4 p.Cys54Tyr-KI, n = 5 WT; n = 8 ex2-KO, n = 8 WT) revealed significant myelin loss relative to control in the corpus callosum of p.Cys54Tyr-KI and ex2-KO mice. One-sided Student's t test on stain quantification: *p value < 0.05, **p value < 0.01.

and myelination (all FDR < 0.10). Module 47 in particular included *THAP1*, was overrepresented with $_{JOINT}$DEGs (38% of the genes in the module overlapped with $_{JOINT}$DEGs, Fisher's test p value < 2.2×10^{-16}), and was enriched for YY1 (MIM: 600013) targets (FDR 3.5×10^{-7}). This protein is a transcription factor with critical functions in oligodendrocyte differentiation[63] (Table S5).

Thap1 disruption is associated with defects in myelination in mouse models

The transcriptional profiling of NSCs thus indicated that *THAP1* disruption is associated with gene modules related to myelination. To directly assess the effects of *THAP1* perturbation on myelin gene expression and morphology, we analyzed two *Thap1* mouse models harboring (1) knockin of the p.Cys54Tyr variation (p.Cys54Tyr-KI, $Thap1^{p.Cys54Tyr/+}$) that we profiled in iPSCs and (2) a deletion allele based on knockout of *Thap1* exon 2 (ex2-KO, $Thap1^{-/+}$). Both mice display subtle motor abnormalities and impaired performance in behavioral assays compared to WT littermates.[68] We interrogated three brain regions that have been implicated in the pathogenesis of dystonia, including regions where microstructural white matter abnormalities have been documented in neuroimaging studies in humans,[69,70] and observed age-, region-, and genotype-specific variability in gene expression compared to control tissue. In the cerebellum and cortex, we detected increased expression of myelin-related genes (*Cnp*, *Mag*, *Mobp*, *Mog*, and *Plp1*) in postnatal day 21 (P21) mice, which was most significant in the ex2-KO model (n = 5 p.Cys54Tyr-KI, n = 7 WT littermates; n = 10 ex2-KO, n = 12 WT littermates). In 2-month-old (adult) mice, we detected a consistent and significant downregulation of these genes in the cortex (p value < 0.05, n = 4 p.Cys54Tyr-KI, n = 4 WT mice, Figure 4A). There was also a significant reduction in myelin structural integrity on the basis of LFB staining in the corpus callosum of p.Cys54Tyr-KI and ex2-KO mice relative to control (p value < 0.05; n = 4 p.Cys54Tyr-KI, n = 5 WT mice; n = 8 ex2-KO, n = 8 WT mice, Figure 4C). These data appear consistent with the transcriptional analyses and indicate that dysregulation of THAP1 activity impacts the formation and/or maintenance of myelin in the brain.

Discussion

In recent years, advances in human genetics and genomics have led to the discovery of an expanding list of genetic

variation associated with hereditary dystonia. This list currently includes genes such as *TOR1A* (MIM: 605204), in which a single mutation has been linked to dystonia in multiple populations (DYT1 [MIM: 128100]),[71] versus others such as *THAP1*, *KMT2B* (MIM: 606834, DYT28 [MIM: #617824]), *ANO3* (MIM: 610110, DYT24 [MIM: 615034]), *GNAL* (MIM: 139312, DYT25 [MIM: 615073]), *TAF1* (MIM: 313650, DYT3 [MIM: 314250]), and *ATP1A3* (MIM: 182350, DYT12 [MIM: 128235]), in which varied alterations have been associated with disease.[6,40] In some of these genes harboring multiple mutations (e.g., *THAP1*, *GNAL*, and *ANO3*), genetic variation results in a largely consistent neurologic phenotype, whereas in others, such as *ATP1A3* and *TAF1*, diverse mutations result in distinct clinical syndromes.[72–75] Dissecting the complex relationships between genetic variation and human disease requires a better understanding of their functional outcomes, the ways in which they differ, and in particular, the affected cellular pathways that they share.

Previous efforts to define such convergent mechanisms have gained insight from studies of (1) the impact of dystonia-specific genetic variation within structural domains of a target protein,[76] (2) functional inter-relationships among proteins implicated in dystonia,[4,77] and (3) overlapping brain expression patterns of genes associated with dystonia.[78] Here, we have outlined an additional approach that leverages an allelic series of dystonia-specific mutations in near-isogenic iPSC-based neuronal models to infer mutational mechanisms from their transcriptome-wide effects. By engineering sequence changes into an isogenic background, we isolated the effects of each mutation while capturing variation through the analysis of multiple clones per genotype. In the differential expression analyses, each mutational group was compared internally to its own set of controls, enabling systematic and repeated determinations of dysregulated genes. In this context, the overlap of differentially expressed genes across multiple sets (where each independent analysis met stringent criteria for statistical significance) was considerably beyond what could be expected by chance and thus defined a high-confidence signature of *THAP1* disruption. We used meta-analytic methods to expand this core set of dysregulated genes after obtaining proof of highly correlated mutational expression profiles. This experimental and analytical pipeline mitigates current challenges in discovery via functional genomics in iPSC-based models, i.e., noise introduced by individual genetic variation is resolved by the isogenic background, while false positive signals resulting from off-target genome-editing effects are reduced in the meta-analysis of results across an allelic series. Lastly, we performed validation in mice, which recapitulated the discovered transcriptomic signatures and implied physiologically relevant effects in a model organism and in different developmental time points.

The pathogenic missense mutants analyzed here have behaved differently in previous biochemical assays, raising the possibility that they might exert unique transcriptional effects. Unlike the p.Ser21Thr and p.Pro26Arg variants, the p.Asn12Lys alteration did not affect DNA-binding affinity but rather decreased thermostability.[14,17] The p.Cys54Tyr variation may also be unstable, as the substitution targets one of four residues that together bind a single zinc ion that may be required for structural integrity.[14,15] In our modeling, we found that the global transcriptional effects of these mutations are highly similar, implying a common perturbed pathway related to the regulatory effects of THAP1 that is independent of the mutation's impact on THAP1 stability. In comparing the signatures across the allelic series, we observed a stepwise increase in transcriptional effects induced by the heterozygous, homozygous, and haploinsufficiency models wherein the latter induced a number of dysregulated genes that was two orders of magnitude greater than that of the missense mutations. Although we hypothesized that the p.Arg146Aspfs*9 frameshift variation would be functionally equivalent to the deletion if the loss of residues in the NLS prevented nuclear import, its signature correlated with those of the missense mutants in the DNA-binding domain. These results imply two competing mutational mechanisms. One possibility is that genetic variation localized to the DNA-binding domain or NLS represent hypomorphic alleles, consistent with previous studies in mice reporting that exon deletion produced broader effects than heterozygous knockin of the p.Cys54Tyr variation[16,67] and that p.Cys54Tyr fails to rescue embryonic lethality of Thap1 null.[59,68] Alternatively, missense mutations localized in the DNA-binding domain may induce a novel gain of function in the cytoplasm, and that property is responsible for some of the shared expression changes with the p.Arg146Aspfs*9 variation that abolishes nuclear import of THAP1. To resolve these scenarios, future studies will be required to identify cytosolic functions of THAP1 and to determine whether the p.Arg146Aspfs*9 and similar variation in this domain retain some capacity for passive nuclear entry and DNA binding. As predicted, there was a paucity of transcriptional effects induced by the p.Met143Val variant, suggesting it could be benign. However, given the position of this alteration within the coiled-coil domain, we cannot exclude the possibility that it may affect critical interactions with partner proteins that do not directly disrupt transcriptional activity.

Although the dysregulated genes associated with missense versus haploinsufficiency models were largely exclusive, they appeared to converge on common cellular pathways associated with neurodevelopment, lipid metabolism, and myelin. We identified dysregulation of myelin-related transcripts associated with *THAP1/Thap1* disruption, including consistent downregulation of galactosylceramidase (*GALC*), which encodes a lysosomal enzyme that catabolizes the major lipid constituents of myelin. Mutations in this gene cause an autosomal-recessive neurodevelopmental disorder with leukoencephalopathy and demyelination, globoid cell leukodystrophy (GLD [MIM: 245200]),[79] and enzyme deficiency in mice results in reduced proliferative

potential in NSCs and functional impairment of neuronal and oligodendroglial progeny.[80] Further, we identified modules of genes that were correlated with *THAP1* perturbation and functionally associated with YY1 targets, lipid metabolism, and myelination. Previous studies in mice have linked Thap1 haploinsufficiency to impaired maturation of oligodendrocytes and loss of YY1 binding to a network of genes related to oligodendroglial differentiation.[59] Our results extend these observations further to human disease-specific variation and suggest that a complex interplay of dysregulated lysosomal sphingolipid metabolism, disrupted oligodendrocyte differentiation, and altered myelin formation and/or maintenance may contribute to the molecular and cellular mechanisms underlying *THAP1*-associated disease.[81–84] Although overt demyelination has not been specifically documented in dystonia, neuroimaging studies have reported white matter abnormalities in multiple forms of the disease, including (1) syndromes associated with genetic variation in *TOR1A*, *THAP1*, *SGCE* (MIM: 604149, DYT11 [MIM: 159900]), *TAF1*, *KMT2B*, and *COL6A3* (MIM: 120250, DYT27 [MIM: 616411])[20,69,85–89] and (2) idiopathic cases of cervical dystonia, writer's cramp, laryngeal dystonia, and blepharospasm.[90–93] The mechanisms underlying these disturbances in white matter tracts are not known, and further investigation is needed to determine whether they reflect common defects in any of the pathways perturbed by genetic variation in *THAP1*.

Identifying cellular deficits that may be shared among multiple forms of dystonia remains a high priority for translational research in these disorders. Monogenic dystonias are rare syndromes, and if convergent disease mechanisms exist, they may offer opportunities for broad therapeutic intervention. In this study, we outlined a systematic experimental and analytical variant-to-function platform of iPSC-based modeling and genome editing that we initially used to understand the cellular consequences of genetic variation in *THAP1*. These tools may similarly facilitate future studies that seek to understand how diverse mutations across multiple genes result in similar phenotypes while enabling the potential discovery of common disease targets.

Data and code availability

The accession number for the RNA sequencing data reported in this paper is dbGaP: phs001525. All other data are available on request.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.09.017.

Acknowledgments

Funding for these studies was provided by US National Institutes of Health grants NS102423 (M.E.T., D.C.B., and L.J.O.), NS087997 (D.C.B., L.J.O., and N.S.), and NS100529 (M.E.E.). A.D. was a recipient of fellowships from The Dystonia Medical Research Foundation and the MGH Collaborative Center for X-linked Dystonia Parkinsonism.

Declaration of interests

The authors declare no competing interests.

Received: June 21, 2021
Accepted: September 27, 2021
Published: October 20, 2021

Web resources

Molecular Signatures Database (MsigDB), https://www.gsea-msigdb.org/gsea/msigdb/

Online Mendelian Inheritance in Man (OMIM), https://www.omim.org/

References

1. Albanese, A., Bhatia, K., Bressman, S.B., Delong, M.R., Fahn, S., Fung, V.S.C., Hallett, M., Jankovic, J., Jinnah, H.A., Klein, C., et al. (2013). Phenomenology and classification of dystonia: a consensus update. Mov. Disord. *28*, 863–873.
2. Bragg, D.C., Armata, I.A., Nery, F.C., Breakefield, X.O., and Sharma, N. (2011). Molecular pathways in dystonia. Neurobiol. Dis. *42*, 136–147.
3. Rittiner, J.E., Caffall, Z.F., Hernández-Martinez, R., Sanderson, S.M., Pearson, J.L., Tsukayama, K.K., Liu, A.Y., Xiao, C., Tracy, S., Shipman, M.K., et al. (2016). Functional Genomic Analyses of Mendelian and Sporadic Disease Identify Impaired eIF2α Signaling as a Generalizable Mechanism for Dystonia. Neuron *92*, 1238–1251.
4. Gonzalez-Latapi, P., Marotta, N., and Mencacci, N.E. (2021). Emerging and converging molecular mechanisms in dystonia. J. Neural Transm. (Vienna) *128*, 483–498.
5. Lohmann, K., and Klein, C. (2017). Update on the Genetics of Dystonia. Curr. Neurol. Neurosci. Rep. *17*, 26.
6. Keller Sarmiento, I.J., and Mencacci, N.E. (2021). Genetic Dystonias: Update on Classification and New Genetic Discoveries. Curr. Neurol. Neurosci. Rep. *21*, 8.
7. Rehbach, K., Fernando, M.B., and Brennand, K.J. (2020). Integrating CRISPR Engineering and hiPSC-Derived 2D Disease Modeling Systems. J. Neurosci. *40*, 1176–1185.
8. Kiskinis, E., Sandoe, J., Williams, L.A., Boulting, G.L., Moccia, R., Wainger, B.J., Han, S., Peng, T., Thams, S., Mikkilineni, S., et al. (2014). Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell *14*, 781–795.
9. Aneichyk, T., Hendriks, W.T., Yadav, R., Shin, D., Gao, D., Vaine, C.A., Collins, R.L., Domingo, A., Currall, B., Stortchevoi, A., et al. (2018). Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. Cell *172*, 897–909.e21.
10. Zaslavsky, K., Zhang, W.-B., McCready, F.P., Rodrigues, D.C., Deneault, E., Loo, C., Zhao, M., Ross, P.J., El Hajjar, J., Romm, A., et al. (2019). SHANK2 mutations associated with autism spectrum disorder cause hyperconnectivity of human neurons. Nat. Neurosci. *22*, 556–564.

11. Kwart, D., Gregg, A., Scheckel, C., Murphy, E.A., Paquet, D., Duffield, M., Fak, J., Olsen, O., Darnell, R.B., and Tessier-Lavigne, M. (2019). A Large Panel of Isogenic APP and PSEN1 Mutant Human iPSC Neurons Reveals Shared Endosomal Abnormalities Mediated by APP β-CTFs, Not Aβ. Neuron *104*, 256–270.e5.
12. Fuchs, T., Gavarini, S., Saunders-Pullman, R., Raymond, D., Ehrlich, M.E., Bressman, S.B., and Ozelius, L.J. (2009). Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. Nat. Genet. *41*, 286–288.
13. Roussigne, M., Cayrol, C., Clouaire, T., Amalric, F., and Girard, J.-P. (2003). THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene *22*, 2432–2442.
14. Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F., and Girard, J.-P. (2005). The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc. Natl. Acad. Sci. USA *102*, 6907–6912.
15. Sengel, C., Gavarini, S., Sharma, N., Ozelius, L.J., and Bragg, D.C. (2011). Dimerization of the DYT6 dystonia protein, THAP1, requires residues within the coiled-coil domain. J. Neurochem. *118*, 1087–1100.
16. Aguilo, F., Zakirova, Z., Nolan, K., Wagner, R., Sharma, R., Hogan, M., Wei, C., Sun, Y., Walsh, M.J., Kelley, K., et al. (2017). THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation. Stem Cell Reports *9*, 92–107.
17. Campagne, S., Muller, I., Milon, A., and Gervais, V. (2012). Towards the classification of DYT6 dystonia mutants in the DNA-binding domain of THAP1. Nucleic Acids Res. *40*, 9927–9940.
18. Bessière, D., Lacroix, C., Campagne, S., Ecochard, V., Guillet, V., Mourey, L., Lopez, F., Czaplicki, J., Demange, P., Milon, A., et al. (2008). Structure-function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways. J. Biol. Chem. *283*, 4352–4363.
19. Osmanovic, A., Dendorfer, A., Erogullari, A., Uflacker, N., Braunholz, D., Rakovic, A., Vierke, G., Gil-Rodríguez, C., Münchau, A., Albrecht, M., et al. (2011). Truncating mutations in THAP1 define the nuclear localization signal. Mov. Disord. *26*, 1565–1567.
20. Cheng, F.B., Wan, X.H., Feng, J.C., Ma, L.Y., Hou, B., Feng, F., Wang, L., and Yang, Y.M. (2012). Subcellular distribution of THAP1 and alterations in the microstructure of brain white matter in DYT6 dystonia. Parkinsonism Relat. Disord. *18*, 978–982.
21. Tao, Y., and Zhang, S.-C. (2016). Neural Subtype Specification from Human Pluripotent Stem Cells. Cell Stem Cell *19*, 573–586.
22. Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron *103*, 785–801.e8.
23. Miyaoka, Y., Chan, A.H., and Conklin, B.R. (2016). Using Digital Polymerase Chain Reaction to Detect Single-Nucleotide Substitutions Induced by Genome Editing. Cold Spring Harb. Protoc. *2016*. https://doi.org/10.1101/pdb.prot086801.
24. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
25. Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475.
26. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
27. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.
28. Wang, M., Zhao, Y., and Zhang, B. (2015). Efficient Test and Visualization of Multi-Set Intersections. Sci. Rep. *5*, 16923.
29. Zaykin, D.V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. J. Evol. Biol. *24*, 1836–1841.
30. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47* (W1), W191–W198.
31. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE *6*, e21800.
32. Schlicker, A., Domingues, F.S., Rahnenführer, J., and Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. BMC Bioinformatics *7*, 302.
33. Kolmykov, S., Yevshin, I., Kulyashov, M., Sharipov, R., Kondrakhin, Y., Makeev, V.J., Kulakovskiy, I.V., Kel, A., and Kolpakov, F. (2021). GTRD: an integrated view of transcription regulation. Nucleic Acids Res. *49* (D1), D104–D111.
34. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.
35. Langfelder, P., and Horvath, S. (2012). Fast *R* Functions for Robust Correlations and Hierarchical Clustering. J. Stat. Soft. *46*, i11.
36. Jegga, A.G., Schneider, L., Ouyang, X., and Zhang, J. (2011). Systems biology of the autophagy-lysosomal pathway. Autophagy *7*, 477–489.
37. Stokowy, T., Polushina, T., Sønderby, I.E., Karlsson, R., Giddaluru, S., Le Hellard, S., Bergen, S.E., Sullivan, P.F., Andreassen, O.A., Djurovic, S., et al. (2018). Genetic variation in 117 myelination-related genes in schizophrenia: Replication of association to lipid biosynthesis genes. Sci. Rep. *8*, 6915.
38. Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.
39. Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*, eaat7615.
40. Domingo, A., Yadav, R., and Ozelius, L.J. (2021). Isolated dystonia: clinical and genetic updates. J. Neural Transm. (Vienna) *128*, 405–416.
41. Campagne, S., Saurel, O., Gervais, V., and Milon, A. (2010). Structural determinants of specific DNA-recognition by the THAP zinc finger. Nucleic Acids Res. *38*, 3466–3476.

42. Bressman, S.B., Raymond, D., Fuchs, T., Heiman, G.A., Ozelius, L.J., and Saunders-Pullman, R. (2009). Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. Lancet Neurol. *8*, 441–446.
43. Houlden, H., Schneider, S.A., Paudel, R., Melchers, A., Schwingenschuh, P., Edwards, M., Hardy, J., and Bhatia, K.P. (2010). THAP1 mutations (DYT6) are an additional cause of early-onset dystonia. Neurology *74*, 846–850.
44. Panov, F., Tagliati, M., Ozelius, L.J., Fuchs, T., Gologorsky, Y., Cheung, T., Avshalumov, M., Bressman, S.B., Saunders-Pullman, R., Weisz, D., and Alterman, R.L. (2012). Pallidal deep brain stimulation for DYT6 dystonia. J. Neurol. Neurosurg. Psychiatry *83*, 182–187.
45. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
46. Jagadeesh, K.A., Wenger, A.M., Berger, M.J., Guturu, H., Stenson, P.D., Cooper, D.N., Bernstein, J.A., and Bejerano, G. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat. Genet. *48*, 1581–1586.
47. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885.
48. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
49. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
50. Clot, F., Grabli, D., Burbaud, P., Aya, M., Derkinderen, P., Defebvre, L., Damier, P., Krystkowiak, P., Pollak, P., Leguern, E., et al. (2011). Screening of the THAP1 gene in patients with early-onset dystonia: myoclonic jerks are part of the dystonia 6 phenotype. Neurogenetics *12*, 87–89.
51. Baker, M., Strongosky, A.J., Sanchez-Contreras, M.Y., Yang, S., Ferguson, W., Calne, D.B., Calne, S., Stoessl, A.J., Allanson, J.E., Broderick, D.F., et al. (2014). SLC20A2 and THAP1 deletion in familial basal ganglia calcification with dystonia. Neurogenetics *15*, 23–30.
52. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. Nature *581*, 444–451.
53. Söhn, A.S., Glöckle, N., Doetzer, A.D., Deuschl, G., Felbor, U., Topka, H.R., Schöls, L., Riess, O., Bauer, P., Müller, U., and Grundmann, K. (2010). Prevalence of THAP1 sequence variants in German patients with primary dystonia. Mov. Disord. *25*, 1982–1986.
54. LeDoux, M.S., Vemula, S.R., Xiao, J., Thompson, M.M., Perlmutter, J.S., Wright, L.J., Jinnah, H.A., Rosen, A.R., Hedera, P., Comella, C.L., et al. (2016). Clinical and genetic features of cervical dystonia in a large multicenter cohort. Neurol. Genet. *2*, e69.
55. Giri, S., Naiya, T., Equbal, Z., Sankhla, C.S., Das, S.K., Ray, K., and Ray, J. (2017). Genetic screening of THAP1 in primary dystonia patients of India. Neurosci. Lett. *637*, 31–37.
56. Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A.H., Karczewski, K., Ing, A.Y., Barton, P.J.R., Funke, B., Cook, S.A., MacArthur, D., and Ware, J.S. (2017). Using high-resolution variant frequencies to empower clinical genome interpretation. Genet. Med. *19*, 1151–1158.
57. Erogullari, A., Hollstein, R., Seibler, P., Braunholz, D., Koschmidder, E., Depping, R., Eckhold, J., Lohnau, T., Gillessen-Kaesbach, G., Grünewald, A., et al. (2014). THAP1, the gene mutated in DYT6 dystonia, autoregulates its own expression. Biochim. Biophys. Acta *1839*, 1196–1204.
58. Cayrol, C., Lacroix, C., Mathe, C., Ecochard, V., Ceribelli, M., Loreau, E., Lazar, V., Dessen, P., Mantovani, R., Aguilar, L., and Girard, J.P. (2007). The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. Blood *109*, 584–594.
59. Yellajoshyula, D., Liang, C.-C., Pappas, S.S., Penati, S., Yang, A., Mecano, R., Kumaran, R., Jou, S., Cookson, M.R., and Dauer, W.T. (2017). The DYT6 Dystonia Protein THAP1 Regulates Myelination within the Oligodendrocyte Lineage. Dev. Cell *42*, 52–67.e4.
60. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23.
61. Moyses-Oliveira, M., Yadav, R., Erdin, S., and Talkowski, M.E. (2020). New gene discoveries highlight functional convergence in autism and related neurodevelopmental disorders. Curr. Opin. Genet. Dev. *65*, 195–206.
62. Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. J. Biol. Chem. *281*, 23319–23325.
63. He, Y., Dupree, J., Wang, J., Sandoval, J., Li, J., Liu, H., Shi, Y., Nave, K.A., and Casaccia-Bonnefil, P. (2007). The transcription factor Yin Yang 1 is essential for oligodendrocyte progenitor differentiation. Neuron *55*, 217–230.
64. Garcia-Tuñon, I., Guallar, D., Alonso-Martin, S., Benito, A.A., Benítez-Lázaro, A., Pérez-Palacios, R., Muniesa, P., Climent, M., Sánchez, M., Vidal, M., and Schoorlemmer, J. (2011). Association of Rex-1 to target genes supports its interaction with Polycomb function. Stem Cell Res. (Amst.) *7*, 1–16.
65. Zurkirchen, L., Varum, S., Giger, S., Klug, A., Häusel, J., Bossart, R., Zemke, M., Cantù, C., Atak, Z.K., Zamboni, N., et al. (2019). Yin Yang 1 sustains biosynthetic demands during brain development in a stage-specific manner. Nat. Commun. *10*, 2192.
66. van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J.P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. Brief. Bioinform. *19*, 575–592.
67. Zakirova, Z., Fanutza, T., Bonet, J., Readhead, B., Zhang, W., Yi, Z., Beauvais, G., Zwaka, T.P., Ozelius, L.J., Blitzer, R.D., et al. (2018). Mutations in THAP1/DYT6 reveal that diverse dystonia genes disrupt similar neuronal pathways and functions. PLoS Genet. *14*, e1007169.
68. Ruiz, M., Perez-Garcia, G., Ortiz-Virumbrales, M., Méneret, A., Morant, A., Kottwitz, J., Fuchs, T., Bonet, J., Gonzalez-Alegre, P., Hof, P.R., et al. (2015). Abnormalities of motor function,

transcription and cerebellar structure in mouse models of *THAP1* dystonia. Hum. Mol. Genet. *24*, 7159–7170.
69. Carbon, M., Kingsley, P.B., Tang, C., Bressman, S., and Eidelberg, D. (2008). Microstructural white matter changes in primary torsion dystonia. Mov. Disord. *23*, 234–239.
70. Niethammer, M., Carbon, M., Argyelan, M., and Eidelberg, D. (2011). Hereditary dystonia as a neurodevelopmental circuit disorder: Evidence from neuroimaging. Neurobiol. Dis. *42*, 202–209.
71. Ozelius, L.J., Hewett, J.W., Page, C.E., Bressman, S.B., Kramer, P.L., Shalish, C., de Leon, D., Brin, M.F., Raymond, D., Corey, D.P., et al. (1997). The early-onset torsion dystonia gene (DYT1) encodes an ATP-binding protein. Nat. Genet. *17*, 40–48.
72. Balint, B., Mencacci, N.E., Valente, E.M., Pisani, A., Rothwell, J., Jankovic, J., Vidailhet, M., and Bhatia, K.P. (2018). Dystonia. Nat. Rev. Dis. Primers *4*, 25.
73. Bragg, D.C., Sharma, N., and Ozelius, L.J. (2019). X-Linked Dystonia-Parkinsonism: recent advances. Curr. Opin. Neurol. *32*, 604–609.
74. Weissbach, A., Saranza, G., and Domingo, A. (2021). Combined dystonias: clinical and genetic updates. J. Neural Transm. (Vienna) *128*, 417–429.
75. Cheng, H., Capponi, S., Wakeling, E., Marchi, E., Li, Q., Zhao, M., Weng, C., Stefan, P.G., Ahlfors, H., Kleyner, R., et al. (2019). Missense variants in TAF1 and developmental phenotypes: challenges of determining pathogenicity. Hum. Mutat. *41*, 449–464.
76. Sweadner, K.J., Arystarkhova, E., Penniston, J.T., Swoboda, K.J., Brashear, A., and Ozelius, L.J. (2019). Genotype-structure-phenotype relationships diverge in paralogs *ATP1A1*, *ATP1A2*, and *ATP1A3*. Neurol. Genet. *5*, e303.
77. Jinnah, H.A., and Sun, Y.V. (2019). Dystonia genes and their biological pathways. Neurobiol. Dis. *129*, 159–168.
78. Mencacci, N.E., Reynolds, R., Ruiz, S.G., Vandrovcova, J., Forabosco, P., Sánchez-Ferrer, A., Volpato, V., Weale, M.E., Bhatia, K.P., Webber, C., et al. (2020). Dystonia genes functionally converge in specific neurons and share neurobiology with psychiatric disorders. Brain *143*, 2771–2787.
79. Weinstock, N.I., Kreher, C., Favret, J., Nguyen, D., Bongarzone, E.R., Wrabetz, L., Feltri, M.L., and Shin, D. (2020). Brainstem development requires galactosylceramidase and is critical for pathogenesis in a model of Krabbe disease. Nat. Commun. *11*, 5356.
80. Santambrogio, S., Ricca, A., Maderna, C., Ieraci, A., Aureli, M., Sonnino, S., Kulik, W., Aimar, P., Bonfanti, L., Martino, S., and Gritti, A. (2012). The galactocerebrosidase enzyme contributes to maintain a functional neurogenic niche during early postnatal CNS development. Hum. Mol. Genet. *21*, 4732–4750.
81. Grillet, M., Dominguez Gonzalez, B., Sicart, A., Pöttler, M., Cascalho, A., Billion, K., Hernandez Diaz, S., Swerts, J., Naismith, T.V., Gounko, N.V., et al. (2016). Torsins Are Essential Regulators of Cellular Lipid Metabolism. Dev. Cell *38*, 235–247.
82. Cascalho, A., Foroozandeh, J., Hennebel, L., Swerts, J., Klein, C., Rous, S., Dominguez Gonzalez, B., Pisani, A., Meringolo, M., Gallego, S.F., et al. (2020). Excess Lipin enzyme activity contributes to TOR1A recessive disease and DYT-TOR1A dystonia. Brain *143*, 1746–1765.
83. Steel, D., Zech, M., Zhao, C., Barwick, K.E.S., Burke, D., Demailly, D., Kumar, K.R., Zorzi, G., Nardocci, N., Kaiyrzhanov, R., et al. (2020). Loss-of-Function Variants in HOPS Complex Genes *VPS16* and *VPS41* Cause Early Onset Dystonia Associated with Lysosomal Abnormalities. Ann. Neurol. *88*, 867–877.
84. Giovedì, S., Ravanelli, M.M., Parisi, B., Bettegazzi, B., and Guarnieri, F.C. (2020). Dysfunctional Autophagy and Endolysosomal System in Neurodegenerative Diseases: Relevance and Therapeutic Options. Front. Cell. Neurosci. *14*, 602116.
85. Carbon, M., Kingsley, P.B., Su, S., Smith, G.S., Spetsieris, P., Bressman, S., and Eidelberg, D. (2004). Microstructural white matter changes in carriers of the DYT1 gene mutation. Ann. Neurol. *56*, 283–286.
86. van der Meer, J.N., Beukers, R.J., van der Salm, S.M.A., Caan, M.W.A., Tijssen, M.A.J., and Nederveen, A.J. (2012). White matter abnormalities in gene-positive myoclonus-dystonia. Mov. Disord. *27*, 1666–1672.
87. Brüggemann, N., Heldmann, M., Klein, C., Domingo, A., Rasche, D., Tronnier, V., Rosales, R.L., Jamora, R.D.G., Lee, L.V., and Münte, T.F. (2016). Neuroanatomical changes extend beyond striatal atrophy in X-linked dystonia parkinsonism. Parkinsonism Relat. Disord. *31*, 91–97.
88. Jochim, A., Li, Y., Zech, M., Lam, D., Gross, N., Koch, K., Zimmer, C., Winkelmann, J., and Haslinger, B. (2018). Microstructural white matter abnormalities in patients with COL6A3 mutations (DYT27 dystonia). Parkinsonism Relat. Disord. *46*, 74–78.
89. Baizabal-Carvallo, J.F., and Alonso-Juarez, M. (2018). Generalized dystonia associated with mutation in the histone methyltransferase gene KMT2B (DYT28) and white matter abnormalities. Parkinsonism Relat. Disord. *49*, 116–117.
90. Bonilha, L., de Vries, P.M., Vincent, D.J., Rorden, C., Morgan, P.S., Hurd, M.W., Besenski, N., Bergmann, K.J., and Hinson, V.K. (2007). Structural white matter abnormalities in patients with idiopathic dystonia. Mov. Disord. *22*, 1110–1116.
91. Berndt, M., Li, Y., Gora-Stahlberg, G., Jochim, A., and Haslinger, B. (2018). Impaired white matter integrity between premotor cortex and basal ganglia in writer's cramp. Brain Behav. *8*, e01111.
92. Pontillo, G., Castagna, A., Vola, E.A., Macerollo, A., Peluso, S., Russo, C., Baglio, F., Manganelli, F., Brunetti, A., Cocozza, S., and Esposito, M. (2020). The cerebellum in idiopathic cervical dystonia: A specific pattern of structural abnormalities? Parkinsonism Relat. Disord. *80*, 152–157.
93. Tomić, A., Agosta, F., Sarasso, E., Svetel, M., Kresojević, N., Fontana, A., Canu, E., Petrović, I., Kostić, V.S., and Filippi, M. (2021). Brain Structural Changes in Focal Dystonia-What About Task Specificity? A Multimodal MRI Study. Mov. Disord. *36*, 196–205.