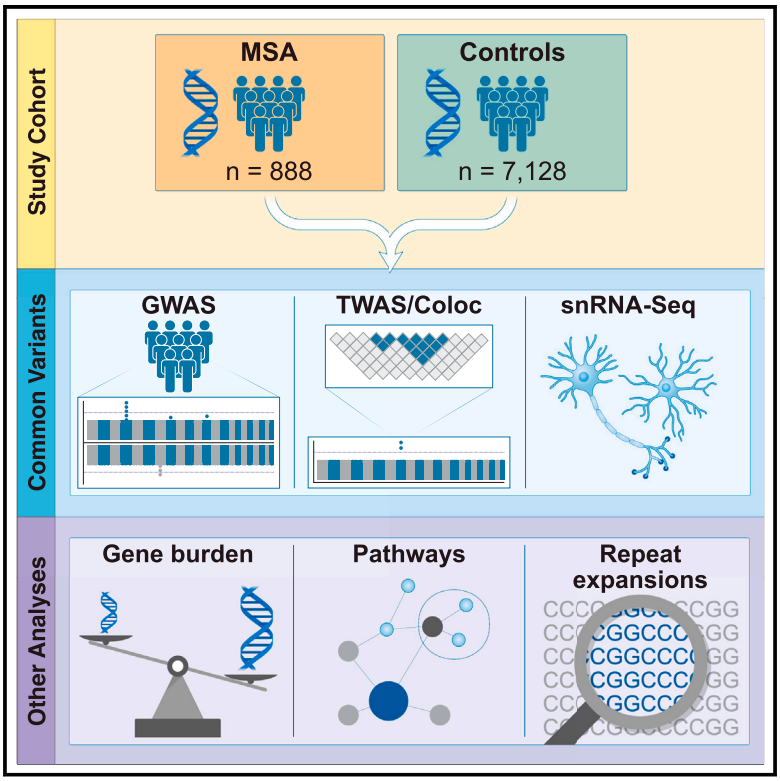
Neuron

Article

Genome sequence analyses identify novel risk loci for multiple system atrophy

Graphical abstract



Highlights

- Generation of a foundational genomic resource in multiple system atrophy
- GWAS identifies novel risk loci at *GAB1*, *lnc-LRRC49-3*, *TENM2*, and *RABGEF1*
- Functional genomics implicates *USP38-DT*, *KCTD7*, and *lnc-KCTD7-2* within these loci
- Gene-burden analysis identifies nominal enrichment of rare missense mutations in *KCTD7*

Authors

Ruth Chia, Anindita Ray, Zalak Shah, ..., Owen A. Ross, Henry Houlden, Sonja W. Scholz

Correspondence

sonja.scholz@nih.gov

In brief

Chia et al. comprehensively analyzed genome sequence data from patients with multiple system atrophy (MSA) and controls. The study identified four novel risk loci associated with MSA and prioritized significantly associated genes (*USP38-DT*, *KCTD7*, and *lnc-KCTD7-2*) within these loci. This initiative's data constitute a valuable resource for the research community.

Chia et al., 2024, Neuron *112*, 1–15
July 3, 2024 Published by Elsevier Inc.
https://doi.org/10.1016/j.neuron.2024.04.002

CellPress

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Neuron

CellPress
OPEN ACCESS

Article

Genome sequence analyses identify novel risk loci for multiple system atrophy

Ruth Chia,[1,82] Anindita Ray,[2,82] Zalak Shah,[2] Jinhui Ding,[3] Paola Ruffo,[1,4] Masashi Fujita,[5] Vilas Menon,[5] Sara Saez-Atienzar,[1] Paolo Reho,[2] Karri Kaivola,[2] Ronald L. Walton,[6] Regina H. Reynolds,[7,8,9] Ramita Karra,[1] Shaimaa Sait,[2] Fulya Akcimen,[1] Monica Diez-Fairen,[10] Ignacio Alvarez,[10] Alessandra Fanciulli,[11] Nadia Stefanova,[11] Klaus Seppi,[11] Susanne Duerr,[11] Fabian Leys,[11] Florian Krismer,[11] Victoria Sidoroff,[11] Alexander Zimprich,[12] Walter Pirker,[13] Olivier Rascol,[14] Alexandra Foubert-Samier,[15] Wassilios G. Meissner,[15,16,17] François Tison,[15,16] Anne Pavy-Le Traon,[18] Maria Teresa Pellecchia,[19] Paolo Barone,[19] Maria Claudia Russillo,[19] Juan Marín-Lahoz,[20,21,22] Jaime Kulisevsky,[20,21] Soraya Torres,[21] Pablo Mir,[23,24,25] Maria Teresa Periñán,[23,26] Christos Proukakis,[27] Viorica Chelban,[28,29] Lesley Wu,[28] Yee Y. Goh,[28] Laura Parkkinen,[30] Michele T. Hu,[31] Christopher Kobylecki,[32] Jennifer A. Saxon,[33,34] Sara Rollinson,[34] Emily Garland,[35] Italo Biaggioni,[35] Irene Litvan,[36]

(Author list continued on next page)

[1]Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[2]Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA
[3]Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[4]Medical Genetics Laboratory, Department of Pharmacy, Health and Nutritional Sciences, University of Calabria, Rende, Italy
[5]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain, New York, NY, USA
[6]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
[7]NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK
[8]Great Ormond Street Institute of Child Health, Genetics and Genomic Medicine, University College London, London, UK
[9]Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK
[10]Memory and Movement Disorders Units, Department of Neurology, University Hospital Mutua de Terrassa, Barcelona, Spain
[11]Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[12]Department of Neurology, Medical University of Vienna, Vienna, Austria
[13]Department of Neurology, Klinik Ottakring - Wilhelminenspital, Vienna, Austria
[14]MSA French Reference Center and CIC-1436, Department of Clinical Pharmacology and Neurosciences, University of Toulouse, Toulouse, France

(Affiliations continued on next page)

SUMMARY

Multiple system atrophy (MSA) is an adult-onset, sporadic synucleinopathy characterized by parkinsonism, cerebellar ataxia, and dysautonomia. The genetic architecture of MSA is poorly understood, and treatments are limited to supportive measures. Here, we performed a comprehensive analysis of whole genome sequence data from 888 European-ancestry MSA cases and 7,128 controls to systematically investigate the genetic underpinnings of this understudied neurodegenerative disease. We identified four significantly associated risk loci using a genome-wide association study approach. Transcriptome-wide association analyses prioritized *USP38-DT*, *KCTD7*, and *lnc-KCTD7-2* as novel susceptibility genes for MSA within these loci, and single-nucleus RNA sequence analysis found that the associated variants acted as *cis*-expression quantitative trait loci for multiple genes across neuronal and glial cell types. In conclusion, this study highlights the role of genetic determinants in the pathogenesis of MSA, and the publicly available data from this study represent a valuable resource for investigating synucleinopathies.

INTRODUCTION

The three main synucleinopathies—neurological conditions characterized by abnormal α-synuclein protein aggregates—are Parkinson's disease, Lewy body dementia, and multiple system atrophy (MSA).[1] Considerable progress has been made in unraveling the genetic architecture of Parkinson's disease and Lewy body dementia.[2,3] By contrast, the molecular causes of

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Ileana Rubio,[36] Roy N. Alcalay,[37,38] Kimberly T. Kwei,[37] Steven J. Lubbe,[39] Qinwen Mao,[40,41] Margaret E. Flanagan,[40,42,43] Rudolph J. Castellani,[40] Vikram Khurana,[44,45,46] Alain Ndayisaba,[11,44] Andrea Calvo,[47] Gabriele Mora,[48] Antonio Canosa,[47] Gianluca Floris,[49] Ryan C. Bohannan,[50] Anni Moore,[3] Lucy Norcliffe-Kaufmann,[51] Jose-Alberto Palma,[51] Horacio Kaufmann,[51] Changyoun Kim,[52] Michiyo Iba,[52] Eliezer Masliah,[52] Ted M. Dawson,[53,54,55,56] Liana S. Rosenthal,[53] Alexander Pantelyat,[53] Marilyn S. Albert,[53] Olga Pletnikova,[57,58] Juan C. Troncoso,[57] Jon Infante,[59] Carmen Lage,[59] Pascual Sánchez-Juan,[59,60] Geidy E. Serrano,[61] Thomas G. Beach,[61] Pau Pastor,[62,63] Huw R. Morris,[64]

(Author list continued on next page)

[15]Service de Neurologie des Maladies Neurodégénératives, French Reference Center for MSA, NS-Park/FCRIN Network, CHU Bordeaux, Bordeaux, France
[16]University of Bordeaux, CNRS, IMN, UMR 5293, Bordeaux, France
[17]Department of Medicine, University of Otago, and the New Zealand Brain Research Institute, Christchurch, New Zealand
[18]French Reference Center for MSA, Department of Neurosciences, Centre d'Investigation Clinique de Toulouse CIC1436, UMR 1048, Institute of Cardiovascular and Metabolic Diseases (I2MC), University Hospital of Toulouse, INSERM, Toulouse, France
[19]Neuroscience Section, Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", University of Salerno, Salerno, Italy
[20]Movement Disorders Unit, Neurology Department, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
[21]Institut d'Investigacions Biomèdiques Sant Pau (IIB-Sant Pau), Centro de Investigación en Red Enfermedades Neurodegenerativas (CIBERNED), Universitat Autònoma de Barcelona, Barcelona, Spain
[22]Servicio de Neurología, Hospital Universitario Miguel Servet, Zaragoza, Spain
[23]Unidad de Trastornos del Movimiento Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla Hospital Universitario Virgen del Rocío, Universidad de Sevilla, Seville, Spain
[24]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas, Madrid, Spain
[25]Departamento de Medicina Facultad de Medicina, Universidad de Sevilla, Seville, Spain
[26]Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University, London, UK
[27]Department of Clinical and Movement Neurosciences, University College London Queen Square Institute of Neurology, London, UK
[28]Department of Neuromuscular Diseases, University College London Queen Square Institute of Neurology, London, UK
[29]The National Hospital for Neurology and Neurosurgery, London, UK
[30]Nuffield Department of Clinical Neurosciences, Oxford Parkinson's Disease Centre, University of Oxford, Oxford, UK
[31]Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, UK
[32]Department of Neurology, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Sciences Centre, The University of Manchester, Manchester, UK
[33]Cerebral Function Unit, Manchester Centre for Clinical Neurosciences, Salfort, UK
[34]Division of Neuroscience and Experimental Psychology, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK
[35]Autonomic Dysfunction Center, Vanderbilt University Medical Center, Nashville, TN, USA
[36]Department of Neurosciences, University of California, San Diego, San Diego, CA, USA
[37]Department of Neurology, Columbia University Irving Medical Center, New York, NY, USA
[38]Neurological Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
[39]Ken and Ruth Davee Department of Neurology and Simpson Querrey Center for Neurogenetics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[40]Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[41]Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, USA
[42]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, UT Health San Antonio, San Antonio, TX, USA
[43]Department of Pathology, UT Health San Antonio, San Antonio, TX, USA
[44]Ann Romney Center for Neurologic Disease, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[45]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[46]Harvard Stem Cell Institute, Cambridge, MA, USA
[47]"Rita Levi Montalcini" Department of Neuroscience, University of Turin, Turin, Italy
[48]Istituti Clinici Scientifici Maugeri, IRCCS, Milan, Italy
[49]Department of Neurology, University Hospital of Cagliari, Cagliari, Italy
[50]Department of Neurobiology and Behavior, University of California, Irvine, Irvine, CA, USA
[51]Department of Neurology, New York University School of Medicine, New York, NY, USA
[52]Molecular Neuropathology Section, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA
[53]Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD, USA
[54]Neuroregeneration and Stem Cell Programs, Institute of Cell Engineering, Johns Hopkins University Medical Center, Baltimore, MD, USA
[55]Department of Pharmacology and Molecular Science, Johns Hopkins University Medical Center, Baltimore, MD, USA
[56]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University Medical Center, Baltimore, MD, USA
[57]Department of Pathology (Neuropathology), Johns Hopkins University Medical Center, Baltimore, MD, USA

(Affiliations continued on next page)

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

Diego Albani,[65] Jordi Clarimon,[66,67] Gregor K. Wenning,[68,83] John A. Hardy,[9,69,70,71,72] Mina Ryten,[7,8] Eric Topol,[73] Ali Torkamani,[73] Adriano Chiò,[47,74,75] David A. Bennett,[76] Philip L. De Jager,[5] Philip A. Low,[77] Wolfgang Singer,[77] William P. Cheshire,[78] Zbigniew K. Wszolek,[78] Dennis W. Dickson,[6,84] Bryan J. Traynor,[1,53,79,84] J. Raphael Gibbs,[3,84] Clifton L. Dalgard,[80,84] Owen A. Ross,[6,81,84] Henry Houlden,[28,29,84] and Sonja W. Scholz[2,53,84,85,*]

[58]Department of Pathology and Anatomical Sciences, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY, USA
[59]Neurology Service, University Hospital Marqués de Valdecilla-IDIVAL-UC-CIBERNED, Santander, Spain
[60]Alzheimer's Centre Reina Sofia-CIEN Foundation-ISCIII, Madrid, Spain
[61]Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Sun City, AZ, USA
[62]Genomics and Transcriptomics of Synucleinopathies, Neurosciences, The Germans Trias i Pujol Research Institute (IGTP), Badalona, Barcelona, Spain
[63]Unit of Neurodegenerative Diseases, Department of Neurology, University Hospital Germans Trias i Pujol, Badalona, Barcelona, Spain
[64]Department of Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, University College London, London, UK
[65]Department of Neuroscience, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy
[66]Sant Pau Biomedical Research Institute, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain
[67]The Network Center for Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain
[68]Autonomic Unit – Division of Neurobiology, Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria
[69]UK Dementia Research Institute of UCL, UCL Institute of Neurology, University College London, London, UK
[70]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, University College London, London, UK
[71]UCL Movement Disorders Centre, University College London, London, UK
[72]Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[73]Scripps Research Translational Institute, Scripps Research, La Jolla, CA, USA
[74]Institute of Cognitive Sciences and Technologies, C.N.R., Rome, Italy
[75]Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Turin, Italy
[76]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[77]Department of Neurology, Mayo Clinic, Rochester, MN, USA
[78]Department of Neurology, Mayo Clinic, Jacksonville, FL, USA
[79]RNA Therapeutics Laboratory, Therapeutics Development Branch, National Center for Advancing Translational Sciences, Rockville, MD, USA
[80]Department of Anatomy, Physiology and Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA
[81]Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL, USA
[82]These authors contributed equally
[83]Author deceased
[84]Senior author
[85]Lead contact
*Correspondence: sonja.scholz@nih.gov
https://doi.org/10.1016/j.neuron.2024.04.002

MSA are poorly understood due to its rarity in the community (~15,000 patients in the United States), its sporadic nature, the heterogeneous clinical manifestations, and the possibility of mimic syndromes.[4] Consequently, MSA remains the least understood member within the synucleinopathy triad.

Clinically, MSA is classified as a sporadic, adult-onset neurodegenerative disease that presents with variable combinations of parkinsonism, cerebellar ataxia, pyramidal signs, and dysautonomia.[5] The mean age at disease onset is 55 years, and most patients die within 6–10 years.[6] Pathologically, MSA is defined by widespread neuronal loss and gliosis, with the deposition of fibrillar α-synuclein in oligodendroglial cells that spreads throughout the brain using prion-like mechanisms.[7] As for most neurodegenerative diseases, no disease-modifying therapies are available, and treatments are directed toward managing the patient's symptoms.

Identifying genetic risk loci is at the heart of efforts to understand the pathogenesis of MSA and inform translational efforts. Candidate gene studies have implicated risk variants within the *SNCA*, *GBA1*, *MAPT*, and *COQ2* genes, though replicating these loci has been challenging.[8–16] Here, we attempted to fill this critical knowledge gap by generating a whole genome sequencing dataset and analyzing it to discover risk variants driving the risk of MSA. Our genome-wide association studies (GWASs) and transcriptomic analyses identified novel risk genes. We further examined the genetic architecture of MSA by mapping pathogenic repeat expansions and investigating rare, damaging variants. Importantly, we provide a valuable resource to stimulate and advance research in this understudied, fatal neurodegenerative disease.

RESULTS

Genome sequence data from 888 patients diagnosed with MSA and 7,128 control subjects were included in the analysis after quality control (see Figure S1 for a workflow diagram; see Table 1 for demographic characteristics).

GWAS identifies novel MSA risk loci on chromosomes 4q31.21 and 15q23

We conducted a GWAS of ~9.2 million variants with a minor allele frequency (MAF) of 1% or higher to identify genetic loci associated with MSA. In Figure 1A, we show the Manhattan plot for the genome-wide analysis using the additive model,

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Table 1. Characteristics of study participants

Variable	Cases	Controls	
		Neurologically healthy controls	TOPMed controls
Number	888	3,018	4,110
Female sex, *n* (%)	416 (47%)	1,612 (53%)	2,129 (52%)
Mean age (years, range)	64 (38–91)[a]	77 (16–110)	62 (20–93)
Diagnostic category			
Clinically probable MSA, *n* (%)	420 (47%)	–	–
Definite MSA, *n* (%)	468 (53%)	–	–
Clinical subtype			
MSA-P, *n* (%)	202 (23%)	–	–
MSA-C, *n* (%)	127 (14%)	–	–
Not available, *n* (%)	559 (63%)	–	–

[a]Age information was not available for 116 MSA cases.

and Table 2 provides details of the most-associated variants. Under this model, we identified a significantly associated locus on chromosome 4q31.21 (rs55894236, odds ratio [OR] = 1.37, 95% confidence interval [CI] = 1.23–1.52, $p = 1.43 \times 10^{-8}$). The index variant was within the first intron of the *GAB1* gene (see Figure S2 for a regional association plot). We discovered a second locus on chromosome 15q23 near the RNA gene *lnc-LRRC49-3* (rs142721461, OR = 2.30, 95% CI = 1.71–3.08, $p = 3.40 \times 10^{-8}$). The quantile-quantile (QQ) plot showed no notable population stratification (additive model $\lambda_{1,000} = 1.045$; Figure S3), and a conditional analysis based on the index variants of these two loci found no additional signals (Figure S4).

A recessive GWAS identifies two additional loci at chromosomes 5q34 and 7q11.21

We also performed a GWAS under a recessive model, as prior research suggested a recessive inheritance pattern within MSA families.[8,17] This analysis identified a locus on chromosome 5q34 within the *TENM2* gene (rs77075949, OR = 7.40, 95% CI = 3.74–14.67, $p = 9.73 \times 10^{-9}$; Figure 1B; Table 2). A second associated locus was located on the long arm of chromosome 7, within *RABGEF1* and near the *KCTD7* gene, and significantly associated with MSA (rs11766262, OR = 1.95, 95% CI = 1.55–2.46, $p = 1.65 \times 10^{-8}$; see Figure 1B for the Manhattan plot; Table 2). Regional association and QQ plots for the recessive model ($\lambda_{1,000} = 0.97$) are shown in Figures S2 and S3. A conditional analysis based on the index variants found no additional signals (Figure S4).

Rare variant analyses identify enrichment of missense mutations in *KCTD7*

Next, we examined our whole genome sequence data for evidence of damaging mutations in candidate genes that included *GBA1, SNCA, MAPT, COQ2, KCTD7, GAB1*, and *TENM2*. To do this, we used gene-level sequence kernel association-optimized (SKAT-O) tests of missense and loss-of-function mutations with a MAF threshold of $\leq 1\%$ and a minor allele count ≥ 2.[18] Although genome-wide significance was not achieved due to the relatively small cohort size, we found that rare missense mutations in *KCTD7* were nominally associated with the risk of developing MSA ($p = 7.9 \times 10^{-3}$, Table 3) when the analysis was limited to the previously implicated genes. The variants identified in canonical and non-canonical transcripts of the *KCTD7* gene are listed in Table S1, and their distribution relative to the gene and protein structure is shown in Figure S5. No statistically significant associations were identified in other tested candidate genes or in a genome-wide gene-based analysis (Figure S6).

Robustness of the GWAS signals

We tested the associations' robustness by performing leave-one-out analyses in the overall cohort. These tests showed the same directions of effect at the four GWAS loci (*GAB1*, *lnc-LRRC49-3*, *TENM2*, and *RABGEF1*), demonstrating the strength of the association signals (Figure S7). We also performed sensitivity analyses by performing GWAS using only the pathologically confirmed subset of the MSA cases ($n = 468$ cases versus 7,128 controls). Although none of the loci achieved genome-wide significance in this investigation due to the smaller cohort size, the sensitivity evaluation demonstrated the same directions of effect and overall robustness of our findings (*GAB1*: rs55894236:C, OR = 1.42, 95% CI = 1.23–1.64, $p = 1.23 \times 10^{-6}$; *lnc-LRRC49-3*: rs142721461:A, OR = 2.12, 95% CI = 1.44–3.11, $p = 1.21 \times 10^{-4}$; *TENM2*: rs77075949:T, OR = 7.70, 95% CI = 3.35–17.71, $p = 1.56 \times 10^{-6}$; *RABGEF1*: rs11766262:C, OR = 1.89, 95% CI = 1.39–2.56, $p = 4.31 \times 10^{-6}$).

Transcriptome-wide association analyses implicate additional RNA transcripts

We performed a transcriptome-wide association study (TWAS) using bulk RNA sequence data from the Genotype-Tissue Expression (GTEx) portal (https://www.gtexportal.org). We evaluated a broad range of CNS regions ($n = 13$) due to the multisystem nature of the disease and the widespread distribution of neuropathological changes observed in MSA brains at autopsy.[19] This analysis identified transcripts associated with MSA (see Figures 1A and 1B, lower). At the 4q31.21 locus, the TWAS prioritized *USP38-DT*, a long non-coding RNA

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

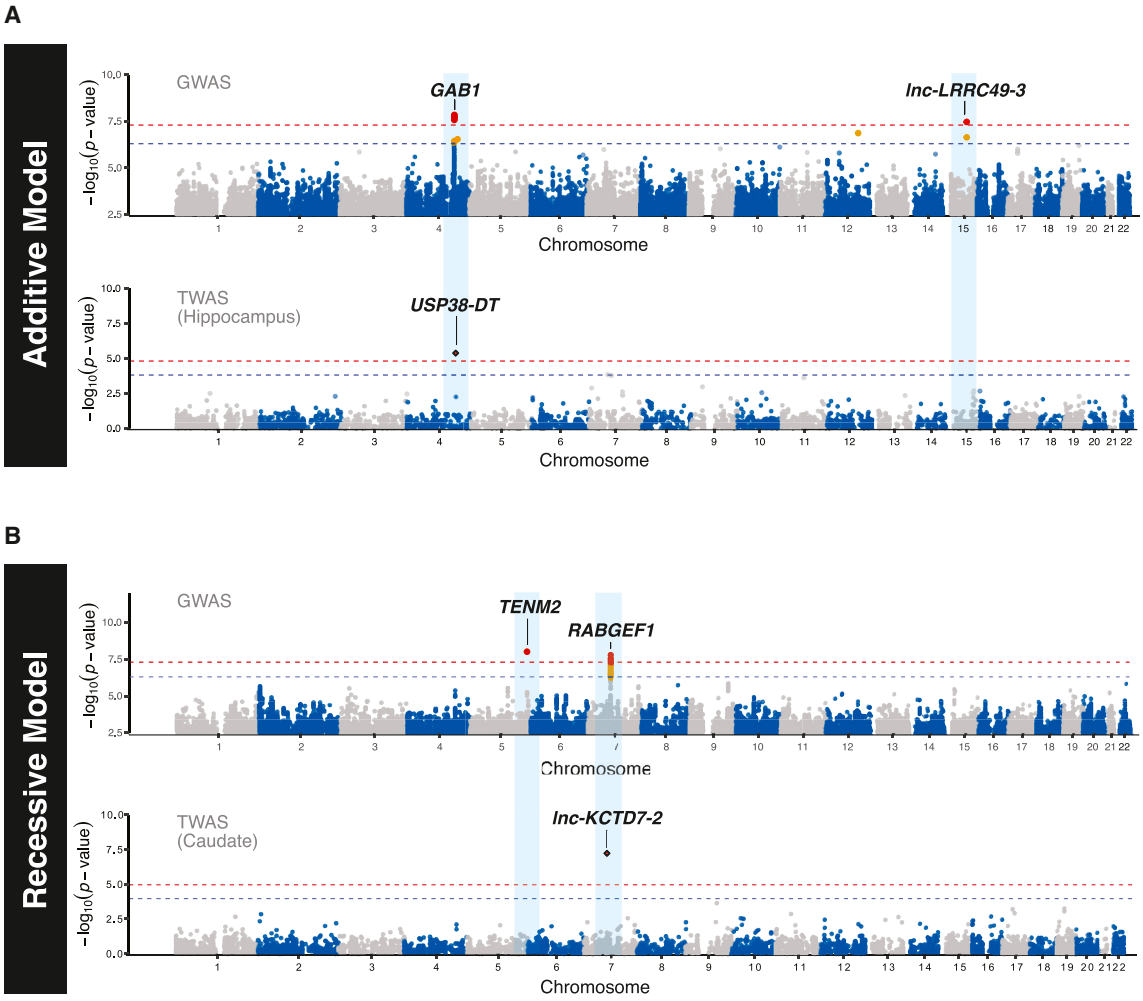


Figure 1. Genome-wide and transcriptome-wide association study results in MSA

(A and B) (A) Composite figure showing the additive MSA GWAS model (upper) with the corresponding TWAS results (lower) in 888 cases and 7,128 controls (MAF > 1%). The x axis denotes the chromosomal position in hg38 and the y axis shows the association *p* values on a negative-$\log_{10}$ scale. Each dot represents a variant (GWAS) or a transcript (TWAS). The TWAS results were generated using GTEx gene expression data (version 8) for the hippocampus (additive model) and the caudate (recessive model). Red dots indicate genome-wide significant variants, while orange dots are sub-significant signals. A red dashed line indicates the Bonferroni threshold for genome-wide significance (5.0×10^{-8} for the GWAS, 1.38×10^{-5} for the TWAS in the hippocampus, and 6.77×10^{-6} for the TWAS in the caudate). The blue dashed line denotes the threshold for declaring variants to be sub-significant (5.0×10^{-7}). The gene(s) closest to the index variant at each locus in the GWAS is listed. In the TWAS plot, the dot with a black diamond outline indicates a colocalization posterior prior probability hypothesis H4 > 0.80. (B) shows the recessive MSA GWAS model with the corresponding TWAS results in the caudate.

(lncRNA) located 293 kilobases (kb) upstream of *GAB1*. Lower expression of *USP38-DT* was predicted to increase the risk for MSA (rs300925, *p* value within hippocampus = 4.38×10^{-6}, Z = −4.59, Table 4). The risk allele of the lead GWAS SNP (rs55894236-C) in this locus was similarly associated with decreased expression of *USP38-DT* in the GTEx bulk brain samples (Figure 2A).

At the 7q11.21 locus, our TWAS analysis identified an lncRNA (*lnc-KCTD7-2*) as being associated with MSA in ten brain regions (Table 4); decreased expression of *lnc-KCTD7-2* was predicted to increase disease risk, with the most prominent association being observed in the caudate (rs6958520, $p = 5.96 \times 10^{-8}$, Z = −5.42). We also found evidence of association for the *KCTD7* transcript in three brain regions, including the amygdala, the hippocampus, and the cervical spinal cord. Similar to *lnc-KCTD7-2*, decreased expression of *KCTD7* was predicted to increase MSA risk across these regions (for example, rs10215516 in the amygdala, $p = 3.23 \times 10^{-6}$, Z = −4.66; Table 4). The risk allele of the lead GWAS SNP (rs11766262-C) in this locus was similarly associated with decreased expression of *KCTD7* in the GTEx bulk brain samples (Figure 2B).

Colocalization analysis nominates genes in the pathogenesis of MSA

Next, we determined which genes within each locus might be driving susceptibility to MSA. To do this, we used *cis*-expression

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Table 2. Top association signals in the MSA GWAS using additive or recessive models

Chr.	Position (SNP ID)	Nearest gene	EA/OA	EAF cases/controls	OR	95% CI	*p* value
Additive GWAS model							
4	143,370,884 (rs55894236)	*GAB1*	C/T	0.44/0.39	1.37	1.23–1.52	1.43×10^{-8}
15	70,380,309 (rs142721461)	*lnc-LRRC49-3*	A/C	0.04/0.02	2.30	1.71–3.08	3.40×10^{-8}
Recessive GWAS model							
5	168,050,511 (rs77075949)	*TENM2*	T/TC	0.08/0.07	7.40	3.74–14.67	9.73×10^{-9}
7	66,699,548 (rs11766262)	*RABGEF1*	C/T	0.63/0.59	1.95	1.55–2.46	1.65×10^{-8}

The gene in closest proximity to the index variant at each locus is represented. The chromosomal position is shown according to hg38. The genome-wide significance was defined as a $p < 5.00 \times 10^{-8}$. The effect allele was defined as the allele associated with an increased risk of disease (i.e., odds ratio > 1.0). Chr., chromosome; EA, effect allele; OA, other allele; EAF, effect allele frequency; OR, odds ratio; 95% CI, 95% confidence interval.

quantitative trait locus (eQTL) data generated using bulk brain expression data in the GTEx project (version 8). We performed co-localization tests on the SNPs with a $p < 1.00 \times 10^{-4}$ within each GWAS locus to estimate the probability that a given variant is associated with both disease risk and gene expression.[20] This approach prioritized *USP38-DT* as the likely causal gene in the 4q31.21 locus, as it had a high posterior probability in the hippocampus (rs300925: posterior probability of hypothesis four [PPH4] = 0.84), a brain region known to be affected in MSA (Figure 2C). It also prioritized *KCTD7* and *lnc-KCTD7-2* as the likely causal genes in the 7q11.21 locus, based on high posterior probability values across multiple brain regions (Figure 2C; Table 4). No genes were prioritized by the colocalization analysis in the 5q34 and 15q23 loci. A summary of the GWAS, TWAS, colocalization, and rare variant analyses is provided for each locus in Figure 2D.

Cell-type-specific expression

We also explored the cell type expression of these genes using single-nucleus RNA sequencing (snRNA-seq) expression data generated from sorted CNS cell types for the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP) brain autopsy collection (n = 424 healthy subjects).[21] In the 4q31.21 locus, the lead GWAS SNP (rs55894236:C) was associated with decreased expression of *GAB1* in excitatory neurons ($\beta = -0.26$, $p = 4.7 \times 10^{-28}$; Figure 3A) and inhibitory neurons ($\beta = -0.079$, $p = 9.1 \times 10^{-3}$). Cell-type-specific data were not available for the *USP38-DT* lncRNA transcript. However, we noted a transcript within this locus, *INPP4B*, that is regulated by *USP38-DT*. *INPP4B*, which has been previously implicated in Alzheimer's disease.[22] We found that rs55894236:C increases the expression of *INPP4B* in astrocytes ($\beta = 0.28$, $p = 1.0 \times 10^{-7}$), oligodendrocytes ($\beta = 0.37$, $p = 4.7 \times 10^{-13}$), and oligodendrocyte precursor cells ($\beta = 0.42$, $p = 5.9 \times 10^{-13}$), suggesting that it may also play a role in increasing susceptibility for MSA (Figure 3B).

In the 7q11.21 locus, the lead GWAS SNP (rs11766262:C) was associated with markedly decreased expression of *KCTD7* in oligodendrocytes ($\beta = -0.64$, $p = 4.7 \times 10^{-29}$) and, to a lesser extent, in microglia ($\beta = -0.19$, $p = 1.7 \times 10^{-5}$). By contrast, the same SNP was associated with mildly increased expression of *KCTD7* in excitatory neurons ($\beta = 0.1$, $p = 2.3 \times 10^{-3}$; Figure 3C). Cell-type-specific data were not available for the *lnc-KCTD7-2* transcript.

Table 3. Rare variant association analysis

Gene	Mutation type	SKAT-O *p* value
GBA1	loss of function	–
	missense	0.69
SNCA	loss of function	–
	missense	–
MAPT	loss of function	–
	missense	0.91
COQ2	loss of function	–
	missense	1.00
RABGEF1	loss of function	–
	missense	–
KCTD7	loss of function	–
	missense	7.9×10^{-3}
GAB1	loss of function	–
	missense	0.47
TENM2	loss of function	0.25
	missense	0.68

SKAT-O test results for rare missense and loss-of-function variants in the gene set analysis (MAF < 0.01, minor allele count $\geq$ 2, minor transcript count $\geq$ 2). SKAT-O, sequence kernel association-optimized.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

Table 4. Transcriptome-wide association analysis results in MSA

GTEx tissue	Gene	eQTL ID	EA	Locus	*Z* scores eQTL	GWAS	TWAS	*p* value	PPH4
Hippocampus	*USP38-DT*	rs300925	C	4q31.21	−4.20	3.89	−4.59	4.38×10^{-6}	0.84
Caudate	*lnc-KCTD7*	rs6958520	C	7q11.21	−10.32	5.42	−5.42	5.96×10^{-8}	0.99
Cortex	*lnc-KCTD7*	rs6958520	C	7q11.21	−10.81	5.42	−5.11	3.26×10^{-7}	0.99
Cerebellar hemisphere	*lnc-KCTD7*	rs10215516	A	7q11.21	−9.87	5.45	−5.08	3.70×10^{-7}	0.99
Hippocampus	*lnc-KCTD7*	rs10215516	A	7q11.21	−10.07	5.45	−5.05	4.44×10^{-7}	0.99
Frontal cortex (BA9)	*lnc-KCTD7*	rs10215516	A	7q11.21	−10.17	5.45	−4.99	6.04×10^{-7}	0.99
Hippocampus	*KCTD7*	rs10215516	A	7q11.21	−6.31	5.45	−4.95	7.47×10^{-7}	0.99
Amygdala	*lnc-KCTD7*	rs17566701	C	7q11.21	−8.79	5.44	−4.93	8.05×10^{-7}	0.99
Anterior cingulate	*lnc-KCTD7*	rs6958520	C	7q11.21	−9.72	5.42	−4.90	9.58×10^{-7}	0.99
Putamen	*lnc-KCTD7*	rs6958520	C	7q11.21	−9.37	5.42	−4.89	1.01×10^{-6}	0.99
Cervical spinal cord	*KCTD7*	rs6958520	C	7q11.21	−8.24	5.42	−4.87	1.11×10^{-6}	0.99
Substantia nigra	*lnc-KCTD7*	rs6958520	C	7q11.21	−7.81	5.42	−4.83	1.38×10^{-6}	0.99
Amygdala	*KCTD7*	rs10215516	A	7q11.21	−4.97	5.45	−4.66	3.23×10^{-6}	0.98
Nucleus accumbens	*lnc-KCTD7*	rs6958520	C	7q11.21	−10.67	5.42	−4.54	5.69×10^{-6}	0.99

Significant eQTLs with coloc PPH4 values ≥ 0.8 and permutation $p < 0.05$ are listed. The eQTL ID refers to the best eQTL in a given locus. EA, effect allele; GTEx, Genotype-Tissue Expression portal (https://gtexportal.org); GWAS, genome-wide association study; PPH4, posterior probability of hypothesis 4; TWAS, transcriptome-wide association study.

Repeat expansion mapping in MSA identifies rare pathogenic allele carriers

We used the ExpansionHunter Targeted tool (version 5) to map repeat elements in ten genes known to carry pathogenic repeat expansions (*AR*, *ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *C9orf72*, *DMPK*, *FMR1*, *FXN*, and *HTT*). This analysis identified 8 (0.9%) MSA cases with pathogenic expansions in the genes *ATXN1*, *ATXN3*, *HTT*, and *AR* (Table S2). Remarkably, seven of these cases had pathologically confirmed MSA, arguing against mimic syndromes as an alternative cause of their neurological syndrome. The observation of the repeat expansions in our MSA cases may reflect the relatively high prevalence of these alleles in the general population, an observation corroborated by their frequent occurrence among our control cohort (n = 20 controls [0.66%] carried a pathogenic expansion; Table S2).[23,24]

Previously nominated genetic loci were not associated with MSA

We investigated our GWAS data for common variation in loci previously reported to be associated with MSA, including the *COQ2*,[8] *MAPT*,[13] *SNCA*,[12,25] *ZIC1-ZIC4*,[26] and *PLA2G4C* loci.[27] None of the tested SNPs in these loci surpassed the

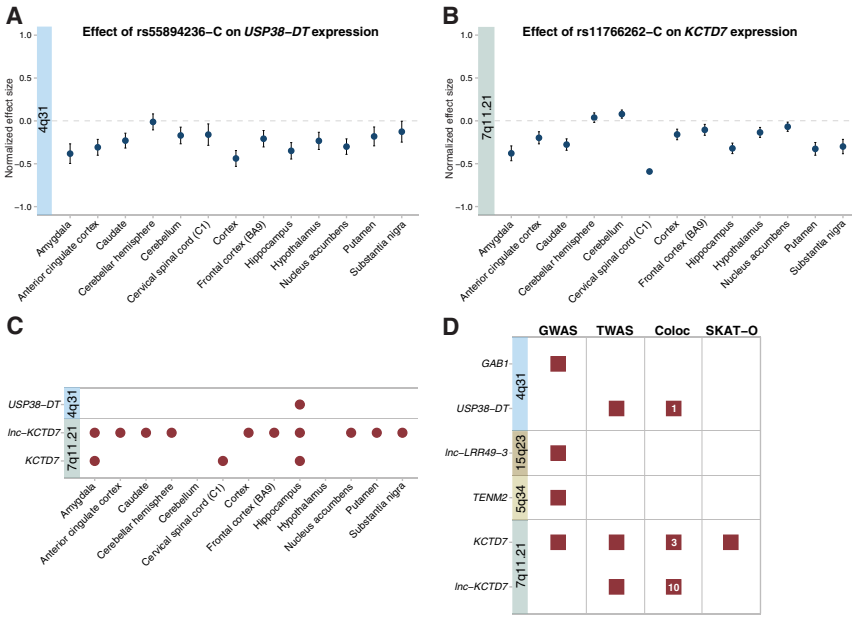


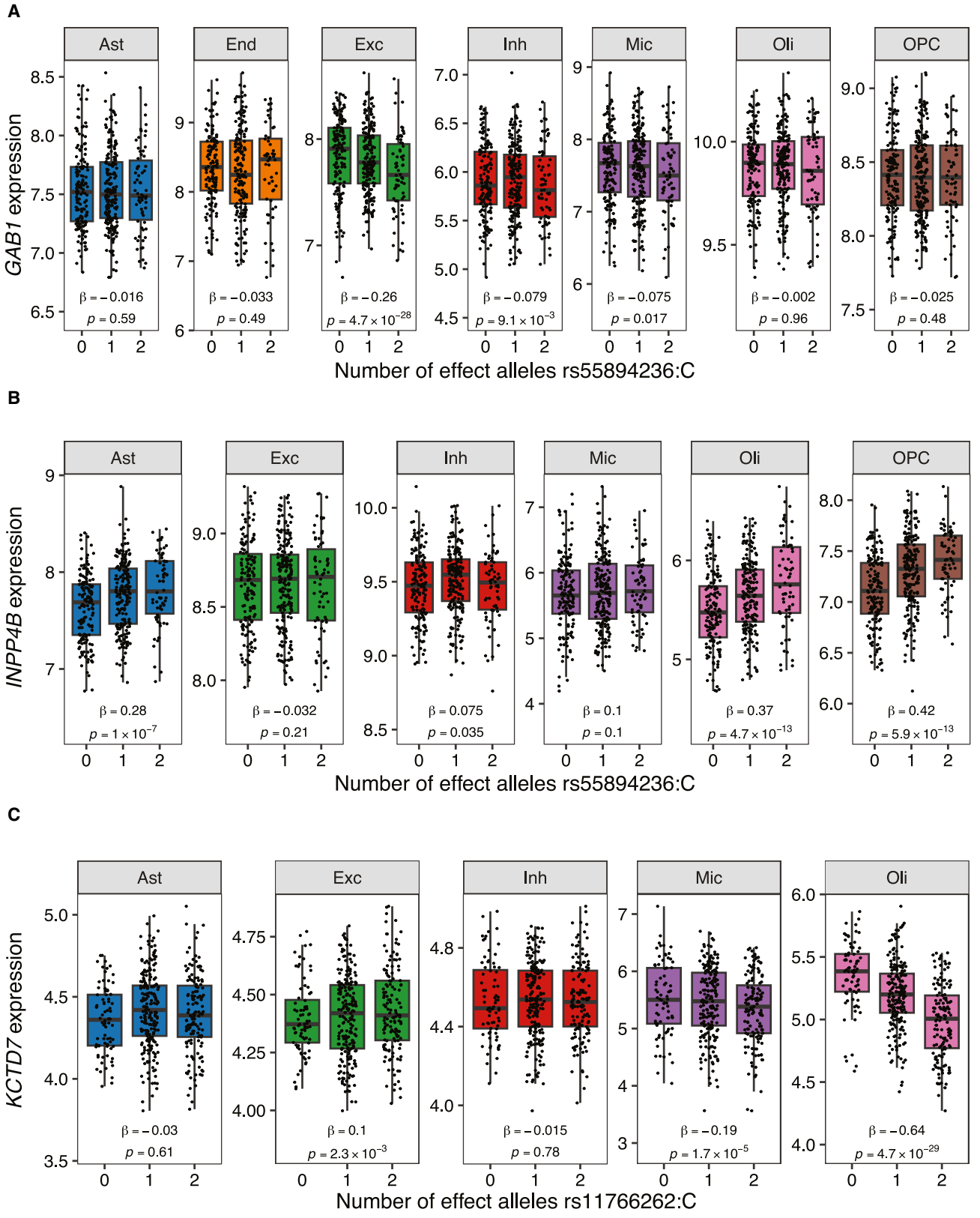
Figure 2. Bulk RNA sequencing and colocalization analyses of the MSA loci

(A) The effect of the 4q31.21 locus index variant, rs55894236-C allele, on *USP38-DT* expression in brain tissues from the GTEx consortium is shown. Error bars indicate the standard error.

(B) The effect of the 7q11.21 locus index variant, rs11766262-C allele, on *KCTD7* expression is depicted.

(C) Summary of significant colocalization signals (PPH4 > 0.80) for transcripts at the 4q31.21 and 7q11.21 loci across the GTEx brain tissues.

(D) Schematic summary of the GWAS, TWAS, colocalization, and SKAT-O results at the four MSA risk loci. The red squares depict a significant analysis result for the listed gene. The numbers in the red squares show the number of tissues that had a significant colocalization PPH4 signal. GWAS, genome-wide association study; TWAS, transcriptome-wide association study; coloc, colocalization analysis; PPH4, posterior probability of hypothesis four; SKAT-O, sequence kernel association-optimized.



(legend on next page)

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Bonferroni threshold for multiple testing (Figure S8). The most associated SNP was rs242557, located within the first intron of the *MAPT* gene on the long arm of chromosome 17, with a *p* value of 0.049, which is far from genome-wide or even regional Bonferroni-adjusted significance (Table S3). We, therefore, lack evidence that common genetic variation in these loci plays a major role in MSA risk in the European/Northern American population. We also used the Manta algorithm to examine our whole genome sequence data for evidence of *SNCA* duplications or triplications; none of the cases ($n = 888$) carried a specific structural variant of this gene.

Gene set enrichment analysis identifies a pathway associated with MSA

Pathway enrichment analysis was performed based on the GWAS summary statistics. We discovered a significant pathway under the recessive model, namely 3′–5′ DNA helicase activity, also known as ATP-dependent DNA helicase activity (GO:0003678, number of genes = 15, $p = 3.05 \times 10^{-5}$, FDR-adjusted $p = 0.0495$). None of the other pathways achieved significance under the additive or recessive models after correction for multiple testing.

DISCUSSION

Our analyses of whole genome sequence data illustrate the impact of common and rare variants in MSA, a fatal neurodegenerative disease. Specifically, our GWAS identified several novel loci associated with MSA risk, and gene-burden tests implicated *KCTD7* in this synucleinopathy. Functional mapping using TWAS and colocalization analyses also revealed changes in the *USP38-DT*, *KCTD7*, and *lnc-KCTD7-2* transcripts at these risk loci, connecting their expression to disease risk. Our pathway analysis implicated 3′–5′ DNA helicase activity in the MSA pathogenesis, which has not been previously implicated in the disease. Our investigations highlight the value of unbiased, data-driven evaluations, which may open new avenues for future exploration. These analytical approaches were chosen as these databases already exist, and, indeed, they complemented each other. Finally, we showed that MSA is characterized, at least in part, by a primary molecular deficit localized within oligodendrocytes, corroborating the converging evidence from preclinical models and *post-mortem* studies indicating that MSA is a primary oligodendrogliopathy.[4,28]

We identified a new locus on the long arm of chromosome 4 using an additive GWAS model (Figure 1A). Functional mapping using TWAS and colocalization analysis narrowed the candidate genes in this region to *USP38-DT* (Figures 1A and 3). *USP38-DT* is a ubiquitously expressed, validated lncRNA, a transcript class that can upregulate or decrease the expression of genes via *cis*- or *trans*-mechanisms.[29] Interestingly, this lncRNA regulates the expression of the nearby *INPP4B* (https://lncipedia.org), an Alzheimer's disease-related gene involved in lysosomal homeostasis and the autophagic clearance of protein aggregates.[30] Additionally, in cellular studies, overexpression of INPP4B was necessary for α-synuclein-mediated endocytosis,[31] likely due to its activity in phosphorylating the phosphoinositol membrane lipids for micropinocytosis.[32]

Our transcriptome-wide analysis of MSA identified a significant association of *USP38-DT* at the chromosome 4q31.21 locus in the hippocampus rather than the cerebellum, brainstem, and basal ganglia, which are the more prominently affected brain regions in autopsies. Nevertheless, the hippocampus is known to be affected in MSA, reflecting the widespread neuropathology associated with the disease. Indeed, prominent pathological hippocampal changes were observed in a large autopsy series of MSA,[33] and cognitive impairment in patients correlated with this hippocampal involvement.[34,35] Further work will be needed to examine the phenotype correlation of these transcripts. However, our findings already highlight the power of genomics to uncover novel results when applied in a broad empirical manner.

Besides *USP38-DT*, the location of the GWAS association signal within the 4q31.21 locus also points to the *GAB1* gene as a potential candidate underlying the molecular pathomechanism. The scaffolding protein encoded by *GAB1* regulates oligodendrocyte development,[36,37] a cell type that is particularly affected in MSA.[28] Conditional deletions of *Gab1* in a murine model have been shown to impair myelination by affecting oligodendrocyte progenitor cell differentiation.[37] Interestingly, patients with MSA show early myelin dysfunction and relocation of myelin proteins,[38] making *GAB1* a plausible risk gene. *GAB1* has also been shown to be involved in Parkinson's disease,[39] and a key paralog of this gene, *GAB2*, is the principal activator of phosphatidylinositol-3 kinase, which has been implicated in Alzheimer's disease.[40] Our genomic data prompt us to speculate how disruption of this locus may increase the risk for disease; a preliminary model suggests that complex interactions involving multiple genes within the chromosome 4q31.21 locus may lead to MSA.[41] Despite this complexity, these observations could have therapeutic implications, as antisense oligonucleotides targeting *USP38-DT*, *INPP4B*, or *GAB1* in the CNS are worth exploring as a means to slowing symptom progression in patients with MSA.

A second GWAS risk locus, located on chromosome 15q23, was detected under the additive model. This association signal was located within a distal enhancer-like signature (ENCODE accession #: EH38E1774799) and downstream of the RNA gene *lnc-LRRC49-3*. However, the TWAS analysis found no significant expression changes at this locus (Figure 1). Further work is therefore needed to pinpoint a possible molecular mechanism associated with this risk locus.

Figure 3. Single-nucleus RNA sequence analyses of common variants at the 4q31.21 and 7q11.21 loci

(A) Single-nucleus RNA sequence analyses identified *cis*-eQTLs for rs55894236 for *GAB1* in excitatory neurons, inhibitory neurons, and microglia (*p* value threshold < 0.05).

(B) This SNP was also a *cis*-eQTL for *INPP4B* in astrocytes, oligodendrocytes, oligodendrocyte precursor cells, and inhibitory cells.

(C) Additionally, we identified a *cis*-eQTL for rs11766262 regulating *KCTD7* expression in oligodendroglia, microglia, and excitatory neurons. Ast, astrocytes; End, endocytes; Exc, excitatory neurons; Inh, inhibitory neurons; Mic, microglia; Oli, oligodendroglia; OPC, oligodendroglia precursor cell.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Under a recessive GWAS model, we detected a significant MSA risk locus within the *TENM2* gene on chromosome 5q34 (Figure 1). *TENM2* encodes the teneurin transmembrane protein 2, which is involved in neuronal migration,[42] calcium-mediated signaling,[43] cell-cell adhesion,[44] and retrograde *trans*-synaptic signaling.[45] However, no TWAS signal was identified at this locus, and more work is needed to understand the molecular mechanism by which genetic variation is associated with susceptibility for MSA.

We detected a second risk locus under the recessive model, located on chromosome 7q11.21 within *RABGEF1* and downstream of *KCTD7* (Figure 1). Our TWAS and colocalization analyses prioritized both *KCTD7* and *lnc-KCTD7-2* as the genes within this region most likely contributing to the pathogenesis of MSA. Gene-burden analysis also found a nominally significant enrichment of missense mutations in *KCTD7* in MSA cases (Table 3). Although this association was not significant at the genome-wide level, it does provide additional evidence supporting its role in MSA pathogenesis. Like other neurodegenerative disease loci, such as *GBA*,[2] *GRN*,[46] and *LRRK2*,[47] where common and rare variants have been implicated, *KCTD7* may be pleomorphic in MSA. It also remains possible that the expression of other transcripts in this locus is affected and contributes to the pathogenesis of MSA.

KCTD7 is a member of the potassium channel tetramerization domain-containing protein family that is highly expressed in the cerebellum and modulates neuron excitability.[34] Mutations in this gene have already been linked to a severe neurodegenerative disease called progressive myoclonic epilepsy, type 3 (OMIM: 611726); this syndrome manifests with intractable myoclonic seizures before the age of two, developmental regression, and truncal ataxia, a clinical feature that is also frequently observed among patients with MSA.[48]

Intriguingly, KCTD7 was recently found to regulate calpains, a group of non-lysosomal cysteine proteases, by inducing ubiquitination.[49] Loss of this KCTD7-induced ubiquitination leads to calpain hyperactivation, aberrant cleavage of downstream targets, and caspase-3 activation.[49] CRISPR-Cas9-mediated knockout of *Kctd7* in mice phenotypically recapitulated human KCTD7 deficiency and resulted in calpain hyperactivation, behavioral impairments, and neurodegeneration; these phenotypes were largely prevented by pharmacological inhibition of calpains.[49] Overall, our genomic data implicate a novel molecular mechanism in the pathogenesis of MSA. Therapeutic strategies targeting malfunctions of calpains are also under development[50] and, based on our work, may be appropriate for therapeutic development in MSA.

We used a recessive inheritance model to identify the association of *KCTD7* in MSA; intriguingly, this locus was not detected under the additive model. The importance of evaluating non-additive inheritance models is increasingly recognized for complex traits,[51] such as obesity,[52] type 2 diabetes,[53] and autoimmune diseases.[54] Examining the recessive model may be particularly beneficial for age-related illnesses like MSA, where recessive loci with reduced penetrance may contribute.[17] Such traits appear to occur sporadically within the population as they rarely recur within families. Despite this, recessive alleles are easier to map from fewer affected individuals, provided the appropriate model is deployed.[55] Overall, our data reinforce the relevance of adopting the recessive model in GWAS studies and highlight the contribution of recessive variants to late-onset neurological diseases.

We found eight MSA patients with pathogenic repeat expansions in disease-related genes, making up less than 1% of the cohort. This rate is similar to that in the general population,[23,24,56] suggesting that the patients carried the genetic risk variants for two neurological diseases coincidentally. Alternatively, these pathogenic repeat expansions may produce phenotypic syndromes indistinguishable from MSA. There is a growing awareness that mutations in one gene can lead to different neuro-psychiatric syndromes.[23] Indeed, the eight patients' diagnoses were verified by medical record review, and seven had classical MSA features on *post-mortem* evaluation, ruling out mimic syndromes as an alternative diagnosis. Regardless, screening for these mutations should be considered part of the initial MSA evaluation, especially because new treatments targeting these loci are emerging.[57–60]

Our study had limitations. First, although our cohort constituted a large genome sequence dataset generated for MSA, the sample size was relatively small by genomic standards, limiting our power to detect common genetic variants of small effect size. Second, our study only included individuals of European ancestry, as this was the population in which large cohorts of MSA cases and matching control data were readily available; thus, our findings may not be generalizable to non-European populations.[61] Third, the clinical diagnosis of MSA can be challenging, and some of the clinically diagnosed cases could have been mimic syndromes arising from other diseases. To minimize this possibility, we only included patients who fulfilled consensus criteria for clinically probable disease[5] and prioritized whole genome sequencing of pathologically confirmed cases.

An additional limitation of our study was the need for a replication cohort. MSA is rare in the general population, making it challenging to collect large numbers of cases. This is an implicit obstacle to identifying the genetic causes of any rare disease. We hope that future studies involving larger cohorts can help us further understand MSA's genetic etiology; indeed, we have made the summary statistics publicly available with that goal in mind. In the meantime, researchers in the rare disease space must rely on orthogonal evidence to confirm the validity of their findings.[62] In our example, we discovered that genes linked to MSA were expressed in oligodendroglia, and *KCTD7* mutations were already known to cause a juvenile neurodegenerative disorder. We also performed leave-one-out analyses and a sensitivity analysis using only pathologically confirmed cases, demonstrating the robustness of the detected signals.

Conclusions

Our genomic analyses identified four novel risk loci for MSA, a rare and fatal adult-onset neurodegenerative disease. Our discoveries begin to unravel the missing genetic etiology of this understudied member of the synucleinopathy triad. We created a foundational genomic resource that can be systematically investigated to unravel the architecture of MSA. In this way, our study advances the understanding of MSA's pathogenesis and paves the way for modeling the disease and developing targeted treatments.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study cohorts
  - Data generation and preprocessing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Genome-wide association analyses (GWAS)
  - Transcriptome-wide association study (TWAS)
  - Colocalization analyses and gene prioritization
  - Cell type-specific expression analysis
  - Gene-based, rare variant association analyses
  - Repeat expansion analysis in short-read genomes from MSA cases and controls
  - Structural variant evaluation in the MSA cohort
  - Candidate gene analyses
  - Pathway analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2024.04.002.

ACKNOWLEDGMENTS

We thank the contributors who collected the samples and data used in this study and the patients and their families whose help and participation made this work possible. We thank the members of the Laboratory of Neurogenetics and the Neurodegenerative Diseases Research Unit (NIH) for their collegial support and technical assistance. This study used DNA samples and clinical data from the NINDS Repository at Coriell (www.coriell.org). We are grateful to the NIH NeuroBioBank for the provision of tissue samples. The ROS/MAP study was supported by the National Institute on Aging (RF1 AG057474 and U01 AG061356). The study used tissue samples and data from the Johns Hopkins Morris K. Udall Center of Excellence for Parkinson's Disease Research (NIH P50 NS38377). We thank the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona, for providing human biological materials. The Brain and Body Donation Program has been supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026, National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 and P30 AG072980, Arizona Alzheimer's Disease Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson's Disease Consortium), and the Michael J. Fox Foundation for Parkinson's Research. The Columbia Parkinsonism Brain Bank is funded by the Parkinson's Foundation. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIA, NIMH, and NINDS. Biospecimens used in this article were obtained from the Northwestern Movement Disorders Center (MDC) Biorepository. As such, the investigators within the MDC Biorepository contributed to the design and implementation of the MDC Biorepository and/or provided data and collected biospecimens but did not participate in the analysis or writing of this report. MDS Biorepository investigators include Rizwan Akhtar, MD, PhD; Tanya Simuni, MD; Dimitri Krainc, MD, PhD; Puneet Opal, MD, PhD; Steven Lubbe, PhD; Niccolo Mencacci, MD, PhD; Joanna Blackburn, MD; and Lisa Kinsley, MS, CGC. For up-to-date information on the study, visit https://www.feinberg.northwestern.edu/research/cores/units/parkinsons.html. This work was supported by the NUgene Project at Northwestern University. We thank the contributors who collected the samples used in this study and the patients whose help and participation made NUgene and this work possible. Several authors of this publication are members of the European Reference Network for Rare Neurological Diseases—project ID no. 739510. We acknowledge the Oxford Brain Bank, supported by Brains for Dementia Research (BDR) (Alzheimer Society and Alzheimer Research UK) and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). V.K. and A.N. acknowledge funding from the MSA Coalition, the Barbara Bloom Ranson Fund for MSA Research at Brigham and Women's Hospital, the Brigham Research Institute's Director's Transformative Award, and NIH 1R01NS109209-01A1. Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLB). Genome sequencing for NHLBI TOPMed: Lung Tissue Research Consortium (LTRC) was performed at Broad Genomics and Northwest Genomics Center. Core support, including centralized genomic read mapping and genotype calling along with variant quality metrics and filtering, was provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1, contract HHSN268201800002I). Core support, including phenotype harmonization, data management, sample identity QC, and general program coordination, was provided by the TOPMed Data Coordinating Center (R01HL-120393 and U01HL-120393, contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. This study utilized data provided by the LTRC supported by the NHLBI. This research was supported in part by the Intramural Research Program of the National Institutes of Health (the National Institute on Aging and the National Institute of Neurological Disorders and Stroke, project numbers 1ZIAAG000935 and 1ZIANS003154). This study used the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).

AUTHOR CONTRIBUTIONS

Conceptualization, H.H., O.A.R., C.L.D., J.R.G., and S.W.S.; data curation, C.L.D., J.R.G., J.D., R.C., A.R., Z.S., and S.W.S.; formal analysis, R.C., A.R., Z.S., P. Reho, P. Ruffo, M.F., V.M., S.S.-A., K.K., F.A., R.K., S.S., J.D., J.R.G., and S.W.S.; funding acquisition, B.J.T. and S.W.S.; investigation, R.C., A.R., Z.S., J.D., P. Ruffo, M.F., V.M., S.S.-A., P. Reho, K.K., R.L.W., R.H.R., R.K., F.A., M.D.-F., I.A., A.F., A.Z., W.G.M., F.T., A.P.-L.T., M.T. Pellecchia, P.M., V.C., L.W., L.P., M.T.H., I.B., I.L., I.R., R.N.A., S.J.L., V.K., A.M., E.M., T.M.D., L.S.R., A.P., M.S.A., O.P., J.C.T., J.I., G.E.S., T.G.B., P.P., J.C., M.R., P.A.L., W.S., W.P.C., Z.K.W., D.W.D., B.J.T., J.R.G., C.L.D., O.A.R., H.H., and S.W.S.; project administration, S.W.S.; resources, I.A., A.F., N.S., K.S., S.D., F.L., F.K., V.S., A.Z., W.P., O.R., A.F.-S., W.G.M., F.T., A.P.-L.T., M.T. Pellecchia, P.B., M.C.R., J.M.-L., J.K., S.T., P.M., M.T. Periñán, C.P., V.C., L.W., Y.Y.G., L.P., M.T.H., C.K., J.A.S., S.R., E.G., I.B., I.L., I.R., R.N.A., K.T.K., S.J.L., Q.M., M.E.F., R.J.C., V.K., A.N., A. Calvo, G.M., A. Canosa, G.F., R.C.B., A.M., L.N.-K., J.-A.P., H.K., E.M., T.M.D., L.S.R., A.P., M.S.A., O.P., J.C.T., J.I., C.L., P.S.-J., G.E.S., T.G.B., P.P., H.R.M., D.A., J.C., G.K.W., J.A.H., M.R., E.T., A.T., A. Chiò, D.A.B., P.L.D.J., P.A.L., W.S., W.P.C., Z.K.W., D.W.D., B.J.T., J.R.G., C.L.D., O.A.R., H.H., and S.W.S.; software, R.C., A.R., Z.S., J.D., P. Ruffo, S.S.-A., M.F., and J.R.G.; supervision, D.A.B., P.L.D.J., R.C., B.J.T., H.H., O.A.R., C.L.D., J.R.G., and S.W.S.; validation, R.C., A.R., Z.S., and P. Ruffo; visualization, R.C., A.R., Z.S., P. Reho, P. Ruffo, M.F., S.S.-A., S.S., and S.W.S.; writing – original draft, R.C., A.R., B.J.T., and S.W.S.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

T.G.B. is a consultant for Aprinoia Therapeutics, Vivid Genomics, and Avid Radiopharmaceutical and is a scientific advisory board member for Vivid Genomics. J.A.H., H.R.M., B.J.T., and H.R.M. hold US, EU, and Canadian patents on the clinical testing and therapeutic intervention for the hexanucleotide repeat expansion of *C9orf72*. B.J.T. and S.W.S. receive research support from Cerevel Therapeutics. B.J.T. is an editorial and advisory board member for *Brain*, *eClinicalMedicine*, *Journal of Neurology, Neurosurgery, and Psychiatry*, and *Neurobiology of Aging*. H.R.M. reports paid consultancy from Biogen, Biohaven, Lundbeck, UCB, and Denali as well as lecture fees and honoraria from the Wellcome Trust and the Movement Disorders Society. H.R.M.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

received research grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, CBD Solutions, the Drake Foundation, and the Medical Research Council. H.K. is editor-in-chief of *Clinical Autonomic Research*, serves as principal investigator (PI) of a clinical trial sponsored by Biogen MA Inc. (TRACK MSA, S19-01846), and received consultancy fees from Lilly USA LLC, Biohaven Pharmaceuticals Inc., Takeda Pharmaceutical Company Ltd., Ono Pharma UK Ltd., Lundbeck LLC, and Theravance Biopharma US Inc. A.F. reports royalties from Springer Verlag; speaker fees and honoraria from Theravance Biopharma, GE Health Care, Broadview Ventures, Austrian Autonomic Society, Stopp-HSP, and Elsevier; and research grants from the FWF-Austrian Science Fund, Medical University of Innsbruck, US MSA Coalition, Dr. Johannes and Hertha Tuba Foundation, and Austrian Exchange Program, outside of the present work. J.-A.P. is an editorial board member of *Movement Disorders*, *Parkinsonism & Related Disorders*, *BMC Neurology*, and *Clinical Autonomic Research*. I.B. received consultancy fees from Theravance Biopharma US Inc., Amenal Pharmaceutics, Regeneron Pharmaceuticals, Takeda Pharmaceuticals, and Neurawell Therapeutics. S.W.S. serves on the scientific advisory board of the Lewy Body Dementia Association and the Multiple System Atrophy Coalition. S.W.S. is an editorial board member for the *Journal of Parkinson's Disease* and *JAMA Neurology*. A.P. serves on the board of directors for CurePSP, has received research grants from the National Institutes of Health and the Michael J. Fox Foundation, and has received consultancy fees from AbbVie Inc., Biogen Inc., SciNeuro Pharmaceuticals, Ono Pharma, and Ferrer Internacional, S.A. A.T. serves on the scientific advisory board for Vivid Genomics. R.H.R. is currently employed by CoSyne Therapeutics; all work performed for this publication was performed on her own time and not as a part of her duties as an employee. Z.K.W. is partially supported by the NIH/NIA and NIH/NINDS (1U19AG063911, FAIN: U19AG063911), the Mayo Clinic Center for Regenerative Medicine, the gifts from the Donald G. and Jodi P. Heeringa Family, the Haworth Family Professorship in Neurodegenerative Diseases fund, and the Albertson Parkinson's Research Foundation. He serves as PI or co-PI on Biohaven Pharmaceuticals Inc. (BHV4157-206) and Vigil Neuroscience Inc. (VGL101-01.002, VGL101-01.201, PET tracer development protocol, Csf1r biomarker and repository project, and ultra-high field MRI in the diagnosis and management of CSF1R-related adult-onset leukoencephalopathy with axonal spheroids and pigmented glia) projects/grants. He serves as co-PI of the Mayo Clinic APDA Center for Advanced Research, as an external advisory board member for Vigil Neuroscience Inc., and as a consultant on neurodegenerative medical research for Eli Lilli & Company. F.K. received personal fees from Institut de Recherches Internationales Servier, Takeda Pharmaceuticals, Sanofi, Teva, Vial, and the Austrian Society of Neurology in the past 12 months, and he has ongoing grant support from the Austrian Science Fund (FWF) and the National Institutes of Health outside of the submitted work. W.G.M. has received fees for editorial activities with Elsevier and has served as an advisor for Lundbeck, Biohaven, Roche, Alterity, Servier, Inhibikase, Takeda, and Teva.

Received: September 4, 2023
Revised: February 28, 2024
Accepted: April 2, 2024
Published: May 2, 2024

REFERENCES

1. McCann, H., Stevens, C.H., Cartwright, H., and Halliday, G.M. (2014). alpha-synucleinopathy phenotypes. Parkinsonism Relat. Disord. *20*, S62–S67. https://doi.org/10.1016/S1353-8020(13)70017-8.
2. Chia, R., Sabir, M.S., Bandres-Ciga, S., Saez-Atienzar, S., Reynolds, R.H., Gustavsson, E., Walton, R.L., Ahmed, S., Viollet, C., Ding, J., et al. (2021). Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. Nat. Genet. *53*, 294–303. https://doi.org/10.1038/s41588-021-00785-3.
3. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. *18*, 1091–1102. https://doi.org/10.1016/S1474-4422(19)30320-5.
4. Poewe, W., Stankovic, I., Halliday, G., Meissner, W.G., Wenning, G.K., Pellecchia, M.T., Seppi, K., Palma, J.A., and Kaufmann, H. (2022). Multiple system atrophy. Nat. Rev. Dis. Primers *8*, 56. https://doi.org/10.1038/s41572-022-00382-6.
5. Gilman, S., Wenning, G.K., Low, P.A., Brooks, D.J., Mathias, C.J., Trojanowski, J.Q., Wood, N.W., Colosimo, C., Dürr, A., Fowler, C.J., et al. (2008). Second consensus statement on the diagnosis of multiple system atrophy. Neurology *71*, 670–676. https://doi.org/10.1212/01.wnl.0000324625.00404.15.
6. Fanciulli, A., and Wenning, G.K. (2015). Multiple-system atrophy. N. Engl. J. Med. *372*, 1375–1376. https://doi.org/10.1056/NEJMc1501657.
7. Watts, J.C., Giles, K., Oehler, A., Middleton, L., Dexter, D.T., Gentleman, S.M., DeArmond, S.J., and Prusiner, S.B. (2013). Transmission of multiple system atrophy prions to transgenic mice. Proc. Natl. Acad. Sci. USA *110*, 19555–19560. https://doi.org/10.1073/pnas.1318268110.
8. Multiple-System Atrophy Research Collaboration (2013). Mutations in COQ2 in familial and sporadic multiple-system atrophy. N. Engl. J. Med. *369*, 233–244. https://doi.org/10.1056/NEJMoa1212115.
9. Ronchi, D., Di Biase, E., Franco, G., Melzi, V., Del Sorbo, F., Elia, A., Barzaghi, C., Garavaglia, B., Bergamini, C., Fato, R., et al. (2016). Mutational analysis of COQ2 in patients with MSA in Italy. Neurobiol. Aging *45*, 213.e1–213.e2. https://doi.org/10.1016/j.neurobiolaging.2016.05.022.
10. Ross, O.A., Vilariño-Güell, C., Wszolek, Z.K., Farrer, M.J., and Dickson, D.W. (2010). Reply to: SNCA variants are associated with increased risk of multiple system atrophy. Ann. Neurol. *67*, 414–415. https://doi.org/10.1002/ana.21786.
11. Sailer, A., Scholz, S.W., Nalls, M.A., Schulte, C., Federoff, M., Price, T.R., Lees, A., Ross, O.A., Dickson, D.W., Mok, K., et al. (2016). A genome-wide association study in multiple system atrophy. Neurology *87*, 1591–1598. https://doi.org/10.1212/WNL.0000000000003221.
12. Scholz, S.W., Houlden, H., Schulte, C., Sharma, M., Li, A., Berg, D., Melchers, A., Paudel, R., Gibbs, J.R., Simon-Sanchez, J., et al. (2009). SNCA variants are associated with increased risk for multiple system atrophy. Ann. Neurol. *65*, 610–614. https://doi.org/10.1002/ana.21685.
13. Vilariño-Güell, C., Soto-Ortolaza, A.I., Rajput, A., Mash, D.C., Papapetropoulos, S., Pahwa, R., Lyons, K.E., Uitti, R.J., Wszolek, Z.K., Dickson, D.W., et al. (2011). MAPT H1 haplotype is a risk factor for essential tremor and multiple system atrophy. Neurology *76*, 670–672. https://doi.org/10.1212/WNL.0b013e31820c30c1.
14. Wernick, A.I., Walton, R.L., Koga, S., Soto-Beasley, A.I., Heckman, M.G., Gan-Or, Z., Ren, Y., Rademakers, R., Uitti, R.J., Wszolek, Z.K., et al. (2020). GBA variation and susceptibility to multiple system atrophy. Parkinsonism Relat. Disord. *77*, 64–69. https://doi.org/10.1016/j.parkreldis.2020.06.007.
15. Yun, J.Y., Lee, W.W., Lee, J.Y., Kim, H.J., Park, S.S., and Jeon, B.S. (2010). SNCA variants and multiple system atrophy. Ann. Neurol. *67*, 554–555. https://doi.org/10.1002/ana.21889.
16. Zhao, Q., Yang, X., Tian, S., An, R., Zheng, J., and Xu, Y. (2016). Association of the COQ2 V393A variant with risk of multiple system atrophy in East Asians: a case-control study and meta-analysis of the literature. Neurol. Sci. *37*, 423–430. https://doi.org/10.1007/s10072-015-2414-8.
17. Hara, K., Momose, Y., Tokiguchi, S., Shimohata, M., Terajima, K., Onodera, O., Kakita, A., Yamada, M., Takahashi, H., Hirasawa, M., et al. (2007). Multiplex families with multiple system atrophy. Arch. Neurol. *64*, 545–551. https://doi.org/10.1001/archneur.64.4.545.
18. Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D.C., Wurfel, M.M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

small-sample case-control whole-exome sequencing studies. Am. J. Hum. Genet. *91*, 224–237. https://doi.org/10.1016/j.ajhg.2012.06.007.

19. Campese, N., Fanciulli, A., Stefanova, N., Haybaeck, J., Kiechl, S., and Wenning, G.K. (2021). Neuropathology of multiple system atrophy: Kurt Jellinger's legacy. J. Neural Transm. (Vienna) *128*, 1481–1494. https://doi.org/10.1007/s00702-021-02383-3.
20. Farrell, K., Humphrey, J., Chang, T., Zhao, Y., Leung, Y.Y., Kuksa, P.P., Patil, V., Lee, W.-P., Kuzma, A.B., Valladares, O., et al. (2023). Genetic, transcriptomic, histological, and biochemical analysis of progressive supranuclear palsy implicates glial activation and novel risk genes. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.09.565552.
21. Fujita, M., Gao, Z., Zeng, L., McCabe, C., White, C.C., Ng, B., Green, G.S., Rozenblatt-Rosen, O., Phillips, D., Amir-Zilberstein, L., et al. (2024). Cell-subtype specific effects of genetic variation in the the Alzheimer's disease brain. Nat. Genet. *56*, 605–614. https://doi.org/10.1038/s41588-024-01685-y.
22. Hadar, A., Milanesi, E., Squassina, A., Niola, P., Chillotti, C., Pasmanik-Chor, M., Yaron, O., Martásek, P., Rehavi, M., Weissglas-Volkov, D., et al. (2016). RGS2 expression predicts amyloid-beta sensitivity, MCI and Alzheimer's disease: genome-wide transcriptomic profiling and bioinformatics data mining. Transl. Psychiatry *6*, e909. https://doi.org/10.1038/tp.2016.179.
23. Dewan, R., Chia, R., Ding, J., Hickman, R.A., Stein, T.D., Abramzon, Y., Ahmed, S., Sabir, M.S., Portley, M.K., Tucci, A., et al. (2021). Pathogenic huntingtin repeat expansions in patients with frontotemporal dementia and amyotrophic lateral sclerosis. Neuron *109*, 448–460.e4. https://doi.org/10.1016/j.neuron.2020.11.005.
24. Zanovello, M., Ibáñez, K., Brown, A.L., Sivakumar, P., Bombaci, A., Santos, L., van Vugt, J.J.F.A., Narzisi, G., Karra, R., Scholz, S.W., et al. (2023). Unexpected frequency of the pathogenic AR CAG repeat expansion in the general population. Brain *146*, 2723–2729. https://doi.org/10.1093/brain/awad050.
25. Al-Chalabi, A., Dürr, A., Wood, N.W., Parkinson, M.H., Camuzat, A., Hulot, J.S., Morrison, K.E., Renton, A., Sussmuth, S.D., Landwehrmeyer, B.G., et al. (2009). Genetic variants of the alpha-synuclein gene SNCA are associated with multiple system atrophy. PLoS One *4*, e7114. https://doi.org/10.1371/journal.pone.0007114.
26. Hopfner, F., Tietz, A.K., Ruf, V.C., Ross, O.A., Koga, S., Dickson, D., Aguzzi, A., Attems, J., Beach, T., Beller, A., et al. (2022). Common variants near ZIC1 and ZIC4 in autopsy-confirmed multiple system atrophy. Mov. Disord. *37*, 2110–2121. https://doi.org/10.1002/mds.29164.
27. Nakahara, Y., Mitsui, J., Date, H., Porto, K.J., Hayashi, Y., Yamashita, A., Kusakabe, Y., Matsukawa, T., Ishiura, H., Yasuda, T., et al. (2023). Genome-wide association study identifies a new susceptibility locus in PLA2G4C for multiple system atrophy. Preprint at medRxiv. https://doi.org/10.1101/2023.05.02.23289328.
28. Wenning, G.K., Stefanova, N., Jellinger, K.A., Poewe, W., and Schlossmacher, M.G. (2008). Multiple system atrophy: a primary oligodendrogliopathy. Ann. Neurol. *64*, 239–246. https://doi.org/10.1002/ana.21465.
29. Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. *22*, 96–118. https://doi.org/10.1038/s41580-020-00315-9.
30. Rodgers, S.J., Jones, E.I., Arumugam, S., Hamila, S.A., Danne, J., Gurung, R., Eramo, M.J., Nanayakkara, R., Ramm, G., McGrath, M.J., and Mitchell, C.A. (2022). Endosome maturation links PI3Kalpha signaling to lysosome repopulation during basal autophagy. EMBO J. *41*, e110398. https://doi.org/10.15252/embj.2021110398.
31. Schechter, M., Atias, M., Abd Elhadi, S., Davidi, D., Gitler, D., and Sharon, R. (2020). Alpha-synuclein facilitates endocytosis by elevating the steady-state levels of phosphatidylinositol 4,5-bisphosphate. J. Biol. Chem. *295*, 18076–18090. https://doi.org/10.1074/jbc.RA120.015319.
32. Maekawa, M., Terasaka, S., Mochizuki, Y., Kawai, K., Ikeda, Y., Araki, N., Skolnik, E.Y., Taguchi, T., and Arai, H. (2014). Sequential breakdown of 3-phosphorylated phosphoinositides is essential for the completion of macropinocytosis. Proc. Natl. Acad. Sci. USA *111*, E978–E987. https://doi.org/10.1073/pnas.1311029111.
33. Ando, T., Riku, Y., Akagi, A., Miyahara, H., Hirano, M., Ikeda, T., Yabata, H., Koizumi, R., Oba, C., Morozumi, S., et al. (2022). Multiple system atrophy variant with severe hippocampal pathology. Brain Pathol. *32*, e13002. https://doi.org/10.1111/bpa.13002.
34. Koga, S., Parks, A., Uitti, R.J., van Gerpen, J.A., Cheshire, W.P., Wszolek, Z.K., and Dickson, D.W. (2017). Profile of cognitive impairment and underlying pathology in multiple system atrophy. Mov. Disord. *32*, 405–413. https://doi.org/10.1002/mds.26874.
35. Miki, Y., Foti, S.C., Hansen, D., Strand, K.M., Asi, Y.T., Tsushima, E., Jaunmuktane, Z., Lees, A.J., Warner, T.T., Quinn, N., et al. (2020). Hippocampal alpha-synuclein pathology correlates with memory impairment in multiple system atrophy. Brain *143*, 1798–1810. https://doi.org/10.1093/brain/awaa126.
36. Qian, X., Wang, H., Wang, Y., Chen, J., Guo, X., and Deng, H. (2020). Enhanced autophagy in GAB1-deficient vascular endothelial cells is responsible for atherosclerosis progression. Front. Physiol. *11*, 559396. https://doi.org/10.3389/fphys.2020.559396.
37. Zhou, L., Shao, C.Y., Xie, Y.J., Wang, N., Xu, S.M., Luo, B.Y., Wu, Z.Y., Ke, Y.H., Qiu, M., and Shen, Y. (2020). Gab1 mediates PDGF signaling and is essential to oligodendrocyte differentiation and CNS myelination. eLife *9*, e52056. https://doi.org/10.7554/eLife.52056.
38. Song, Y.J., Lundvig, D.M., Huang, Y., Gai, W.P., Blumbergs, P.C., Højrup, P., Otzen, D., Halliday, G.M., and Jensen, P.H. (2007). p25alpha relocalizes in oligodendroglia from myelin to cytoplasmic inclusions in multiple system atrophy. Am. J. Pathol. *171*, 1291–1303. https://doi.org/10.2353/ajpath.2007.070201.
39. Zhu, B., Park, J., Coffey, S., Hsu, I., Lam, T., Gopal, P., Ginsberg, S.D., Wang, J., Su, C., Zhao, H., et al. (2022). Single-cell transcriptomic and proteomic analysis of Parkinson's disease brains. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.14.480397.
40. Schjeide, B.M., Hooli, B., Parkinson, M., Hogan, M.F., DiVito, J., Mullin, K., Blacker, D., Tanzi, R.E., and Bertram, L. (2009). GAB2 as an Alzheimer disease susceptibility gene: follow-up of genomewide association results. Arch. Neurol. *66*, 250–254. https://doi.org/10.1001/archneurol.2008.552.
41. Leys, F., Eschlböck, S., Campese, N., Mahlknecht, P., Peball, M., Goebel, G., Sidoroff, V., Granata, R., Bonifati, V., Zschocke, J., et al. (2022). Family history for neurodegeneration in multiple system atrophy: does it indicate susceptibility? Mov. Disord. *37*, 2310–2312. https://doi.org/10.1002/mds.29202.
42. Del Toro, D., Carrasquero-Ordaz, M.A., Chu, A., Ruff, T., Shahin, M., Jackson, V.A., Chavent, M., Berbeira-Santana, M., Seyit-Bremer, G., Brignani, S., et al. (2020). Structural basis of teneurin-latrophilin interaction in repulsive guidance of migrating neurons. Cell *180*, 323–339.e19. https://doi.org/10.1016/j.cell.2019.12.014.
43. Silva, J.P., Lelianova, V.G., Ermolyuk, Y.S., Vysokov, N., Hitchen, P.G., Berninghausen, O., Rahman, M.A., Zangrandi, A., Fidalgo, S., Tonevitsky, A.G., et al. (2011). Latrophilin 1 and its endogenous ligand Lasso/teneurin-2 form a high-affinity transsynaptic receptor pair with signaling capabilities. Proc. Natl. Acad. Sci. USA *108*, 12113–12118. https://doi.org/10.1073/pnas.1019434108.
44. Beckmann, J., Schubert, R., Chiquet-Ehrismann, R., and Müller, D.J. (2013). Deciphering teneurin domains that facilitate cellular recognition, cell-cell adhesion, and neurite outgrowth using atomic force microscopy-based single-cell force spectroscopy. Nano Lett. *13*, 2937–2946. https://doi.org/10.1021/nl4013248.
45. Zhang, X., Lin, P.Y., Liakath-Ali, K., and Südhof, T.C. (2022). Teneurins assemble into presynaptic nanoclusters that promote synapse formation via postsynaptic non-teneurin ligands. Nat. Commun. *13*, 2297. https://doi.org/10.1038/s41467-022-29751-1.
46. Fernández, M.V., Kim, J.H., Budde, J.P., Black, K., Medvedeva, A., Saef, B., Deming, Y., Del-Aguila, J., Ibañez, L., Dube, U., et al. (2017). Analysis of

neurodegenerative Mendelian genes in clinically diagnosed Alzheimer disease. PLoS Genet. *13*, e1007045. https://doi.org/10.1371/journal.pgen.1007045.

47. Cookson, M.R. (2015). LRRK2 pathways leading to neurodegeneration. Curr. Neurol. Neurosci. Rep. *15*, 42. https://doi.org/10.1007/s11910-015-0564-y.

48. Van Bogaert, P., Azizieh, R., Désir, J., Aeby, A., De Meirleir, L., Laes, J.F., Christiaens, F., and Abramowicz, M.J. (2007). Mutation of a potassium channel-related gene in progressive myoclonic epilepsy. Ann. Neurol. *61*, 579–586. https://doi.org/10.1002/ana.21121.

49. Sharma, J., Mulherkar, S., Chen, U.I., Xiong, Y., Bajaj, L., Cho, B.K., Goo, Y.A., Leung, H.E., Tolias, K.F., and Sardiello, M. (2023). Calpain activity is negatively regulated by a KCTD7-Cullin-3 complex via non-degradative ubiquitination. Cell Discov. *9*, 32. https://doi.org/10.1038/s41421-023-00533-3.

50. Ono, Y., Saido, T.C., and Sorimachi, H. (2016). Calpain research for drug discovery: challenges and potential. Nat. Rev. Drug Discov. *15*, 854–876. https://doi.org/10.1038/nrd.2016.212.

51. Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvoro, C., Miguel-Escalada, I., Carey, C.E., Cole, J.B., Rüeger, S., Atkinson, E., et al. (2021). The impact of non-additive genetic associations on age-related complex diseases. Nat. Commun. *12*, 2436. https://doi.org/10.1038/s41467-021-21952-4.

52. Wood, A.R., Tyrrell, J., Beaumont, R., Jones, S.E., Tuke, M.A., Ruth, K.S., GIANT Consortium, Yaghootkar, H., Freathy, R.M., Murray, A., et al. (2016). Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. Diabetologia *59*, 1214–1221. https://doi.org/10.1007/s00125-016-3908-5.

53. Moltke, I., Grarup, N., Jørgensen, M.E., Bjerregaard, P., Treebak, J.T., Fumagalli, M., Korneliussen, T.S., Andersen, M.A., Nielsen, T.S., Krarup, N.T., et al. (2014). A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature *512*, 190–193. https://doi.org/10.1038/nature13425.

54. Goyette, P., Boucher, G., Mallon, D., Ellinghaus, E., Jostins, L., Huang, H., Ripke, S., Gusareva, E.S., Annese, V., Hauser, S.L., et al. (2015). High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat. Genet. *47*, 172–179. https://doi.org/10.1038/ng.3176.

55. Mani, A., Meraji, S.M., Houshyar, R., Radhakrishnan, J., Mani, A., Ahangar, M., Rezaie, T.M., Taghavinejad, M.A., Broumand, B., Zhao, H., et al. (2002). Finding genetic contributions to sporadic disease: a recessive locus at 12q24 commonly contributes to patent ductus arteriosus. Proc. Natl. Acad. Sci. USA *99*, 15054–15059. https://doi.org/10.1073/pnas.192582999.

56. Gardiner, S.L., Boogaard, M.W., Trompet, S., de Mutsert, R., Rosendaal, F.R., Gussekloo, J., Jukema, J.W., Roos, R.A.C., and Aziz, N.A. (2019). Prevalence of carriers of intermediate and pathological polyglutamine disease-associated alleles among large population-based cohorts. JAMA Neurol. *76*, 650–656. https://doi.org/10.1001/jamaneurol.2019.0423.

57. Tabrizi, S.J., Leavitt, B.R., Landwehrmeyer, G.B., Wild, E.J., Saft, C., Barker, R.A., Blair, N.F., Craufurd, D., Priller, J., Rickards, H., et al. (2019). Targeting huntingtin expression in patients with Huntington's disease. N. Engl. J. Med. *380*, 2307–2316. https://doi.org/10.1056/NEJMoa1900907.

58. O'Callaghan, B., Hofstra, B., Handler, H.P., Kordasiewicz, H.B., Cole, T., Duvick, L., Friedrich, J., Rainwater, O., Yang, P., Benneyworth, M., et al. (2020). Antisense oligonucleotide therapeutic approach for suppression of ataxin-1 expression: A safety assessment. Mol. Ther. Nucleic Acids *21*, 1006–1016. https://doi.org/10.1016/j.omtn.2020.07.030.

59. Hauser, S., Helm, J., Kraft, M., Korneck, M., Hübener-Schmid, J., and Schöls, L. (2022). Allele-specific targeting of mutant ataxin-3 by antisense oligonucleotides in SCA3-iPSC-derived neurons. Mol. Ther. Nucleic Acids *27*, 99–108. https://doi.org/10.1016/j.omtn.2021.11.015.

60. Lopez, E.R., Borschel, W.F., and Traynor, B.J. (2022). New antisense oligonucleotide therapies reach first base in ALS. Nat. Med. *28*, 25–27. https://doi.org/10.1038/s41591-021-01629-7.

61. Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J. (2016). RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics *32*, 1423–1426. https://doi.org/10.1093/bioinformatics/btw079.

62. Cassa, C.A., Akle, S., Jordan, D.M., and Rosenfeld, J.A. (2017). When "N of 2" is not enough: integrating statistical and functional data in gene discovery. Cold Spring Harb. Mol. Case Stud. *3*, a001099. https://doi.org/10.1101/mcs.a001099.

63. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/s13742-015-0047-8.

64. Abraham, G., Qiu, Y., and Inouye, M. (2017). FlashPCA2: principal component analysis of biobank-scale genotype datasets. Bioinformatics *33*, 2776–2778. https://doi.org/10.1093/bioinformatics/btx299.

65. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

66. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

67. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252. https://doi.org/10.1038/ng.3506.

68. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

69. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

70. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

71. Dolzhenko, E., van Vugt, J.J.F.A., Shaw, R.J., Bekritsky, M.A., van Blitterswijk, M., Narzisi, G., Ajay, S.S., Rajan, V., Lajoie, B.R., Johnson, N.H., et al. (2017). Detection of long repeat expansions from PCR-free whole-genome sequence data. Genome Res. *27*, 1895–1903. https://doi.org/10.1101/gr.225672.117.

72. Dolzhenko, E., Weisburd, B., Ibañez, K., Rajan-Babu, I.-S., Anyansi, C., Bennett, M.F., Billingsley, K., Carroll, A., Clamons, S., Danzi, M.C., et al. (2022). REViewer: haplotype-resolved visualization of read alignments in and around tandem repeats. Genome Med. *14*, 84. https://doi.org/10.1186/s13073-022-01085-z.

73. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinform. *43*, 11.10.11–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43.

74. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Mägi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. *9*, 1192–1212. https://doi.org/10.1038/nprot.2014.071.

75. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

76. Dolzhenko, E., Weisburd, B., Ibañez, K., Rajan-Babu, I.S., Anyansi, C., Bennett, M.F., Billingsley, K., Carroll, A., Clamons, S., Danzi, M.C., et al. (2022). REViewer: haplotype-resolved visualization of read alignments in and around tandem repeats. Genome Med. *14*, 84. https://doi.org/10.1186/s13073-022-01085-z.

77. Ibañez, K., Polke, J., Hagelstrom, R.T., Dolzhenko, E., Pasko, D., Thomas, E.R.A., Daugherty, L.C., Kasperaviciute, D., Smith, K.R., et al.; WGS for Neurological Diseases Group (2022). Whole genome sequencing for the diagnosis of neurological repeat expansion disorders in the UK: a retrospective diagnostic accuracy and prospective clinical validation study. Lancet Neurol. *21*, 234–245. https://doi.org/10.1016/S1474-4422(21)00462-2.

78. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222. https://doi.org/10.1093/bioinformatics/btv710.

79. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human cerebellar brain tissue and/or whole blood	Comprehensive list of study sites where samples were collected is listed in the Table S1 of this paper	N/A
Critical commercial assays		
Maxwell RSC Tissue DNA Kit	Promega	Catalog # AS1610
PicoGreen dsDNA assay	Thermo Fisher	Catalog # P7589
TruSeq PCR-free Library Prep Kit	Illumina	Catalog # 20015963
HiSeq X Ten Reagent Kit (v.2.5 chemistry)	Illumina	Catalog # FC-502-2501
Deposited data		
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/genome/guide/human/; RRID:SCR_006553
Individual-level whole genome sequence data from neurologically healthy, aged controls	DementiaSeq	dbGAP (www.ncbi.nlm.nih.gov/gap/); Accession # phs001963 RRID: SCR_002709
Individual-level whole genome sequence data from neurologically healthy, aged controls	Wellderly cohort	Available upon request; Contact: Dr. Ali Torkamani (atorkama@scripps.edu)
TOPMed control genome data	TOPMed consortium	Available on dbGaP (www.ncbi.nlm.nih.gov/gap/); Accession # phs001662.v2.p1, phs00974.v5.p4, phs000951.v5.p5); RRID: SCR_002709
Gene expression data	GTEx (v.8)	https://gtexportal.org/home/; RRID: SCR_013042
Individual-level, whole genome sequencing data from MSA cases	This paper	dbGAP (www.ncbi.nlm.nih.gov/gap/); Accession #: phs001963; RRID: SCR_002709
MSA GWAS summary statistics	This paper	GWAS catalog: www.ebi.ac.uk/gwas/
Software and algorithms		
GATK	Broad Institute	https://gatk.broadinstitute.org/; RRID:SCR_001876
Pipeline Standardization	CCDG	https://github.com/CCDG/Pipeline-Standardization/blob/master/PipelineStandard.md
prod-wgs-germline-snps-indels	Broad Institute	https://github.com/gatk-workflows/broad-prod-wgs-germline-snps-indels
PLINK (v.2.0)	Chang et al.[63]	https://www.cog-genomics.org/plink/2.0/; RRID:SCR_001757
FlashPCA (v.2.0)	Abraham et al.[64]	https://github.com/gabraham/flashpca; RRID:SCR_021680
R (v.3.5.2)	R core team	https://www.r-project.org/; RRID: SCR_001905
SAMtools	Li et al.[65]	https://samtools.sourceforge.net/; RRID:SCR_002105
BCFtools (v.1.16.1)	Danecek et al.[66]	https://samtools.github.io/bcftools/bcftools.html; RRID:SCR_005227
FUSION	Gusev et al.[67]	https://github.com/gusevlab/fusion_twas

(*Continued on next page*)

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell Ranger (v.6.0.0)	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome; RRID:SCR_017344
Seurat (v.4.3.0)	Stuart et al.[68]	https://github.com/satijalab/seurat; RRID:SCR_016341
Matrix eQTL	Shabalin et al.[69]	http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/
Ensembl Variant Effect Predictor (VEP, v.101)	Ensembl	https://useast.ensembl.org/info/docs/tools/vep/index.html; RRID:SCR_007931
Loss-Of-Function Transcript Effect Estimator (LOFTEE)	Karczewski et al. (2020)	https://github.com/konradjk/loftee
RVTests (v.2.1.0)	Zhan et al.[61]	https://github.com/zhanxw/rvtests; RRID:SCR_007639
MAGMA (v.1.10)	de Leeuw et al.[70]	https://ctg.cncr.nl/software/magma
Gene set pathway analysis	Molecular Signatures Database (MSigDB), v7.5.1	https://www.gsea-msigdb.org/gsea/msigdb/human/genesets.jsp; RRID:SCR_016863
ExpansionHunter Targeted (v.5)	Dolzhenko et al.[71]	https://github.com/Illumina/ExpansionHunter
Repeat Expansion Viewer (REViewer, v.0.2.7)	Dolzhenko et al.[72]	https://github.com/Illumina/REViewer
MSA genomic analyses code	This paper	https://zenodo.org/records/10723069

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Sonja W. Scholz (sonja.scholz@nih.gov).

Materials availability

The study did not generate any new unique reagents.

Data and code availability

The summary statistics from the additive and recessive GWAS models have been deposited in the GWAS catalog (https://www.ebi.ac.uk/gwas/). The individual-level sequence data for a subset of the MSA genomes (n = 683 cases) reported in this paper will be available upon publication in dbGaP (accession number: phs001963). Public data sharing was not feasible for the remaining 205 MSA genomes; access to these data will be granted to qualified researchers via appropriate collaboration agreements. The TOPMed control genome data are available in dbGaP (accession numbers: phs001662.v2.p1, phs00974.v5.p4, phs000951.v5.p5). The control genome data from 1,980 subjects from the DementiaSeq project are available in dbGaP (accession number: phs001963), and the remaining control genomes are available upon request from the Wellderly study team (contact: atorkama@scripps.edu). The programming code used in this paper is available at https://zenodo.org/records/10723069.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study cohorts

The study workflow is depicted in Figure S1. The cohort included 3,978 participants of European ancestry (958 MSA cases and 3,021 neurologically healthy controls). MSA cases were recruited across twenty North American and European sites and consortia to create a genomic resource for MSA research (see Table S4 for a list of the participating sites). MSA cases were diagnosed with clinically probable (n = 416 [47%]) or pathologically definite disease (n = 468 [53%]) according to the Gilman consensus criteria.[5] The control subjects were obtained from the DementiaSeq project (dbGaP accession number: phs001963) and selected based on a lack of evidence of cognitive decline in their clinical history and no neurological deficits on neurological examination.[2] The pathologically confirmed control individuals had no evidence of significant neurodegenerative disease on histopathological examination. We additionally obtained whole genome sequence data from 5,963 European-ancestry convenience controls from the TOPMed consortium (dbGaP accession numbers: phs001662.v2.p1, phs00974.v5.p4, phs000951.v5.p5). The demographic characteristics of the cohorts

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

are summarized in Table 1. The appropriate institutional review boards of participating institutions approved the study, and informed consent was obtained from all subjects or their surrogate decision-makers according to the Declaration of Helsinki.

Data generation and preprocessing

Whole genome sequencing

PCR-free libraries from genomic DNA samples were constructed using the Illumina TruSeq chemistry, according to the manufacturer's protocol. Whole genome sequencing was performed on an Illumina NovaSeq sequencer using 150 base pair, paired-end cycles (version 2.5 chemistry, Illumina). The control subjects were previously sequenced on an Illumina HiSeq X Ten using the same parameters, as described elsewhere.[2] The mean sequencing coverage of the samples was 35.85 (range, 18.34–70.75, see Figure S9).

Sequence alignment and variant calling

The whole genome sequence data were aligned to reference genome build hg38 and processed on the Google Cloud Platform, according to GATK (2016) Best Practices.[73] Variants were called by a combination of the publicly available GATK Best Practices and another workflow for joint discovery and Variant Quality Score Recalibration by the Broad Institute (https://github.com/gatk-workflows/broad-prod-wgs-germline-snps-indels). All genome sequence data were processed using a uniform pipeline. The convenience control genomes obtained from the TOPMed consortium were called separately and merged with the study data for quality control checks (Figure S1).

Quality control

A workflow diagram of the quality control steps is shown in Figure S1. The cohort consisted of a discovery dataset (n = 958 MSA cases and 3,021 controls) and a convenience control dataset from TOPMed (n = 5,963 controls). For sample-level quality control of the discovery dataset, we removed genomes based on the following criteria: (1) failed library preparations or sequencing, (2) abnormal heterozygosity (F-statistic outside of the -0.15 to 0.15 range), (3) low call rate ($\leq$ 95%), (4) sex check failure (i.e., a discrepancy between reported sex and genotypic sex), (5) non-European ancestry (based on principal component analysis when compared to Hapmap3 data; Figure S3), (6) duplicate samples (pi-hat statistic > 0.8), (7) related samples (pi-hat statistic > 0.125), and (8) cases in whom the final diagnosis was changed. For variant-level quality control of the discovery dataset, we excluded variants based on the following criteria: (1) spanning deletions, (2) minor allele frequencies (MAFs) significantly different in controls from reported frequencies in the NHLBI TransOmics TOPMed database (freeze 5b; www.nhlbiwgs.org) or gnomAD (version 3.1.2; https://gnomad.broadinstitute.org), (3) a significant departure from Hardy-Weinberg equilibrium in the control cohort (p-value $\leq 1 \times 10^{-6}$), (4) non-autosomal variants (X, Y, mitochondrial DNA), (5) non-random missingness between cases and controls (excluding variants with p-value $< 1 \times 10^{-4}$), (6) haplotype-based non-random missingness (excluding variants with p-value $\leq 1 \times 10^{-4}$), (7) variants with a high missingness rate (i.e., $\geq$ 5%), (8) variants mapping to variable, diversity, and joining (VDJ) recombination sites, and variants in the centromeric regions ± 10 kb (due to poor sequence alignment and incomplete resolution of the reference genome assembled at these sites), (9) variants failing gnomAD filters (version 3.1.2; https://gnomad.broadinstitute.org), and (10) variants with poor sequence alignment. A total of 888 MSA cases and 3,018 controls from the discovery dataset were included in downstream analyses. Of note, for rare variant analyses only this jointly called discovery dataset was used. Following the quality control steps of the discovery dataset, we merged the data with the convenience control genomes from TOPMed and applied the same sample- and variant-level quality control steps. Additionally, we excluded variants that had significantly different minor allele frequencies between the discovery control genomes and the TOPMed control genomes (excluding variants with per chromosome FDR-corrected p-value $\leq$ 0.05). The final dataset included 91,594,360 variants in 888 cases and 7,128 controls (Table 1), which were used for downstream common variant analyses (Figure S1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome-wide association analyses (GWAS)

The GWAS was performed in 888 MSA cases and 7,128 controls using the PLINK toolset (version 2.0).[63] We applied additive and recessive logistic regression models using an MAF threshold of > 1% (based on allele frequency estimates in the MSA cases). We determined the relevant genetic principal components in FlashPCA (version 2.0) and applied the *step* function in R Mass package (R version 3.5.2; https://www.R-project.org/) to calculate the number of principal components required for population substructure correction.[64] Based on this analysis, we included sex and seven principal components as covariates in our GWAS study. The Bonferroni threshold for genome-wide significance was 5×10^{-8}, and variants achieving a p-value less than or equal to 5.0×10^{-7} were considered subsignificant. The effect allele was defined as the allele associated with an increased risk of disease (i.e., odds ratio > 1.0).[74] All of the genes located within a 1 Mb upstream and downstream of each gene were included in transcriptomic analyses to ensure the detection of ancillary signals.

For conditional analyses on GWAS loci identified in the additive and recessive models, we additionally included the respective index variants in the covariates (Figure S4). To demonstrate the robustness of GWAS signals, we performed leave-one-out analyses by withholding samples based on their institutional source; there were twenty-four cohorts from the twenty different institutions at which samples were collected, meaning that twenty-four separate GWAS analyses were performed (Figure S7). A sensitivity analysis was

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

performed in a subset of pathologically confirmed MSA cases (n = 468) and healthy controls (n = 7,128) under both additive and recessive models, with sex and seven principal components included as covariates.

Transcriptome-wide association study (TWAS)

Tissue-specific expression was predicted based on the GWAS summary statistics from the additive and the recessive models by transcriptome-wide association analyses (TWAS). To do so, we obtained gene expression data from the Genotype-Tissue Expression portal (GTEx, version 8; https://gtexportal.org). To explore a gene's association with disease, a transcriptome-wide imputation was achieved using the FUSION pipeline,[67] where the precomputed gene expression weights obtained from the GTEx data for thirteen brain regions were considered. These regions included: (i) the amygdala, (ii) anterior cingulate cortex (BA24), (iii) caudate, (iv) cerebellar hemisphere, (v) cerebellum, (vi) cervical spinal cord (C-1), (vii) cortex, (viii) frontal cortex (BA9), (ix) hippocampus, (x) hypothalamus, (xi) nucleus accumbens, (xii) putamen, and (xiii) substantia nigra. The significant association threshold was defined as 0.05 divided by the number of genes in GTEx (version 8) in the thirteen types of brain regions. This threshold ranged from 2.17×10^{-5} in the substantia nigra to 6.77×10^{-6} in the cerebellum due to the variable number of genes expressed in each tissue. Variants achieving a *p*-value ten-fold higher than the significance threshold were considered subsignificant.

Colocalization analyses and gene prioritization

We used the COLOC function within the FUSION package[67] to test the hypothesis that an MSA risk variant colocalized with an eQTL variant in bulk RNA-seq data obtained from the GTEx project (version 8). For the four genome-wide significant loci in the GWAS (4q31.21, 5q34, 7q11.21, and 15q23), we extracted all SNPs with a *p*-value $< 1 \times 10^{-4}$ for colocalization analysis to evaluate the probability of the MSA loci and eQTL sharing a single causal variant for each region. In each eQTL dataset, we extracted the associations for the SNP-gene pairs within that range and tested for colocalization.[20] A colocalization posterior prior probability hypothesis 4 (PPH4) $\geq$ 80% and a permutation *p*-value < 0.05 was considered evidence for an eQTL-GWAS association that could substantially influence both the expression and the GWAS trait in that region for disease.

Cell type-specific expression analysis

We evaluated the expression of SNPs and nominated risk genes identified through GWAS and TWAS in a single-nucleus RNA-seq dataset generated using the Religious Orders Study/Memory and Aging Project (ROS/MAP) cohort.[21] The ROS/MAP data were derived from 424 dorsolateral prefrontal cortexes of individuals of advanced age using the 10x Genomics Single Cell 3′ kit, as described elsewhere.[21] Sequencing reads were processed, and the unique molecule identifier count matrix was generated using Cell Ranger software (version 6.0.0, 10x Genomics). The cell types were classified by clustering cells by gene expression using the R package Seurat (version 4).[68] The ''pseudobulk'' gene expression matrix was constructed by aggregating unique molecule identifier counts of the same cell type of the same donor and normalizing them to the $\log_2$ counts-per-million-reads-mapped values. Sample genotyping was performed by whole genome sequencing followed by GATK processing. The *cis*-eQTLs were mapped using Matrix-eQTL (version 2.3) for single nucleotide polymorphisms within 1 megabase of the transcription start sites.[69]

Gene-based, rare variant association analyses

A gene-based SKAT-O analyses of missense and loss-of-function mutations were conducted to determine the difference in the aggregate burden of rare coding variants in the MSA cases (n = 888) versus healthy controls (n = 3,018). All variants were annotated in Variant Effect Predictor (VEP; version 101),[75] with the 'LoFtee' plugin to annotate high-confidence loss-of-function variants using the default parameters. The variants were filtered using an MAF threshold of $\leq$ 1% and an MAC of $\geq$ 2. We then performed a SKAT-O analysis of filtered and annotated variants in RVTESTS (version 2.1.0),[61] including sex and five principal components as covariates. We used a genome-wide significance threshold of 3.03×10^{-6} (= 0.05/16,507 genes). From the genome-wide data, we extracted the values for eight genes (*COQ2*, *GBA1*, *MAPT*, *SNCA*, *GAB1*, *RABGEF1, KCTD7*, and *TENM2*) implicated in MSA and used a gene-wide significance threshold of 0.006 (= 0.05/8) to test for significant enrichment of coding mutations.

Repeat expansion analysis in short-read genomes from MSA cases and controls

As repeat expansion diseases can occasionally mimic the clinical features of MSA, we assessed the frequency of pathogenic repeat expansions in our MSA case-control whole genome sequence data (*n* = 888 cases and 3,018 controls) using the ExpansionHunter Targeted tool (version 5; Illumina).[71] This tool has been validated for measuring repeat expansions in ten known disease genes (*AR, ATN1, ATXN1, ATXN2, ATXN3, C9orf72, DMPK, FMR1, FXN*, and *HTT*).[23] Pathogenic repeat expansions were validated manually by visualization in the Repeat Expansion Viewer (REViewer; Illumina).[76,77]

Structural variant evaluation in the MSA cohort

We used the Manta algorithm to detect the structural variants (i.e., duplications) within the *SNCA* locus on chromosome 4q22.1 in the 888 MSA cases and the 3,018 neurologically healthy subjects.[78] This analysis used default settings and focused on the region defined by the *SNCA* gene [chr4:89,724,099-89,837,161]. The result files were merged with bcftools,[66] and missing genotypes were set to reference homozygotes.

Please cite this article in press as: Chia et al., Genome sequence analyses identify novel risk loci for multiple system atrophy, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.04.002

Candidate gene analyses

GWAS data were analyzed for evidence of association in six genes previously reported to be associated with MSA, including *COQ2*, *MAPT, SNCA, ZIC1, ZIC4*, and *PLA2G4C*.[8,12,13,25–27] The analysis was performed by subsetting gene regions from post-quality control variant files and testing for association using a generalized logistic regression model as described above.

Pathway analyses

Gene-set enrichment analyses were performed in MAGMA (version 1.10). The primary analysis used the binary PLINK files to annotate all the variants to genes if they were within the genic boundaries of 1.5 kb upstream and downstream. This was followed by a gene analysis, where the summary statistics from both the additive and recessive GWAS models were used to generate different gene-level metrics. Using the 'mean' test statistics (snp-wise = mean) across all the genes, pathway enrichment evaluation was performed using 13,159 gene sets from MSigDB (https://www.gsea-msigdb.org/gsea/msigdb; version 7.5.1) to identify potential pathways associated with MSA.[79]