

<https://doi.org/10.1038/s41531-024-00723-0>

Profiling complex repeat expansions in *RFC1* in Parkinson's disease

Check for updates

Pilar Alvarez Jerez^{1,2,3}, Kensuke Daida^{1,2}, Abigail Miano-Burkhardt^{1,2}, Hirotaka Iwaki^{1,2,4}, Laksh Malik², Guillaume Cogan^{1,5}, Mary B. Makarios^{1,6}, Roisin Sullivan⁷, Jana Vandrovcova⁷, Jinhui Ding¹, J. Raphael Gibbs¹, Androo Markham⁸, Mike A. Nalls^{1,2,4}, Rupesh K. Kesharwani⁹, Fritz J. Sedlazeck^{1,9,10,11}, Bradford Casey¹², John Hardy³, Henry Houlden^{1,6,7}, Cornelis Blauwendraat^{1,2}, Andrew B. Singleton^{1,2} & Kimberley J. Billingsley^{1,2} ✉

A biallelic (AAGGG) expansion in the poly(A) tail of an AluSx3 transposable element within the gene *RFC1* is a frequent cause of cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS), and more recently, has been reported as a rare cause of Parkinson's disease (PD) in the Finnish population. Here, we investigate the prevalence of *RFC1* (AAGGG) expansions in PD patients of non-Finnish European ancestry in 1609 individuals from the Parkinson's Progression Markers Initiative study. We identified four PD patients carrying the biallelic *RFC1* (AAGGG) expansion and did not identify any carriers in controls.

A biallelic pentanucleotide repeat expansion (AAGGG) in the poly(A) tail of an AluSx3 transposable element in the replication factor C subunit 1 (*RFC1*) gene is a frequent cause of cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS)¹. Further, the length of the biallelic "AAGGG" expansion is disease-modifying, as an inverse correlation was observed between the size of expansions and age at neurological onset, age at onset of dysarthria and/or dysphagia, and age at the use of one stick².

More recent genetic studies have broadened the phenotypic spectrum of *RFC1* expansions. Several groups have investigated the prevalence of *RFC1* expansions in Parkinsonian disorders including multiple system atrophy with conflicting findings^{3,4}. In terms of its association with Parkinson's disease (PD) specifically, Kyotovuori et al. identified that three out of 569 patients with PD were carriers for the biallelic *RFC1* (AAGGG) expansion, suggesting that this expansion may be a rare cause of PD in the Finnish population⁵.

In this study, we aimed to profile the biallelic *RFC1* "AAGGG" repeat expansion in PD patients from non-Finnish European ancestry in 903 cases and 706 controls from the PPMI cohort. Due to the complexity of the *RFC1* repeat, short-read whole genome sequencing (WGS) data can yield false

positives, hence experimental validation is required. From the short-read analysis, five patients were predicted to carry the biallelic expansion. However, through the Oxford Nanopore Technologies (ONT) long-read DNA WGS, one predicted carrier was identified as a false positive, leaving four validated carriers. The four remaining carriers were PD patients resulting in an estimated frequency of 0.43% in PD. No controls carried the "AAGGG" *RFC1* repeat expansion. From the ONT long-read analysis, the biallelic "AAGGG" expansion repeat units varied from 333 to 1183 in the four carriers, which is slightly larger than the 144–820 reported in PD patients in the Finnish population⁵, but a smaller range than what was observed in CANVAS patients from European ancestry, which ranged from 400 to 2000 repeats¹ (Supplementary Table 2).

For the four "AAGGG" *RFC1* carriers, some variation was observed in the clinical phenotypic description (Table 1). However, overall in agreement with previous observations of the repeat expansion in PD patients, the clinical phenotype was that neither the presentation nor disease course differed from those in other PD patients. Patient 1 developed PD at the age of 57. She presented tremors, rigidity, and bradykinesia as motor symptoms (MDS-UPDRS 24 pts, Hoehn and Yahr (H&Y) stage 2), depression, mild

¹Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA. ²Center for Alzheimer's and Related Dementias, National Institute on Aging, Bethesda, MD, USA. ³Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, UK. ⁴Data-Tecnica LLC, Washington, DC, USA. ⁵Sorbonne Université, Institut du Cerveau-Paris Brain Institute-ICM, Institut National de la Recherche Médicale-U1127, Centre National de la Recherche Scientifique, Paris, France. ⁶UCL Movement Disorders Centre, University College London, London, UK. ⁷Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London, UK. ⁸Oxford Nanopore Technologies, Oxford, UK. ⁹Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. ¹⁰Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ¹¹Department of Computer Science, Rice University, Houston, TX, USA. ¹²The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA. ✉e-mail: kimberley.billingsley@nih.gov

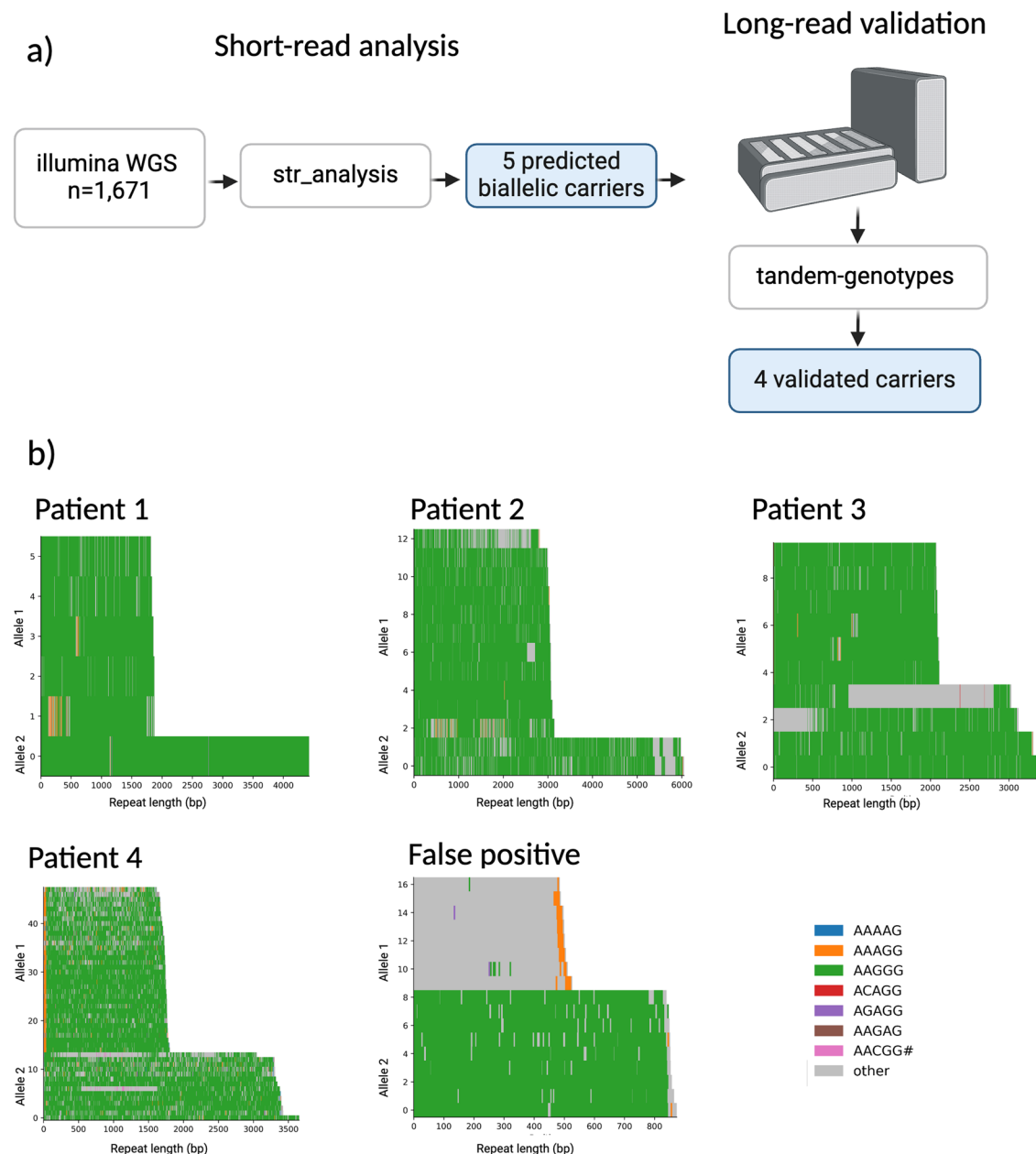


Fig. 1 | Investigation of the RFC1 biallelic expansion in PD patients from European ancestry. a Overview of the study design and rationale behind the analysis included in the work. **b** Waterfall plots of the ONT long-read sequencing data showing the four predicted carriers and one false positive. Created with [BioRender.com](#).

cognitive decline, constipation, and insomnia as non-motor symptoms. Her symptoms did not show much progression until the latest PPMI visit (one and a half years after onset) since she had not taken any medications. Her dopamine transporter (DaT) imaging was normal at the initial diagnosis and she showed a negative reaction in alpha-synuclein (aSyn) SAA.

Patient 2 developed PD at the age of 65. She presented tremors, rigidity, bradykinesia, and postural instability at the diagnosis (H&Y stage 1, MDS-UPDRS part 3, 18 points). Approximately 2 years after the onset, her symptoms progressed (H&Y stage 3, MDS-UPDRS part 3, 35 points) with 900 mg of levodopa equivalent dose (LEDD) and she showed a negative reaction in aSyn SAA.

Patient 3 developed PD at the age of 54, presenting with tremors, bradykinesia, and hyposmia. At the age of 61, 9 years from the onset, her H&Y stage was 2 with 900 mg of LEDD. She showed a positive reaction in aSyn SAA. DaT imaging showed decreased binding in the putamen.

Patient 4 developed PD at the age of 76. A year after the onset, his H&Y stage was 2, MDS-UPDRS part 3 was 16, accompanied by constipation and

insomnia. Clinical data of follow-up visits were not available and aSyn SAA was not performed for this patient. Genetic testing revealed that he was a carrier of the known damaging *LRRK2* p.G2019S variant. DaT imaging results were not available.

In this study, we leveraged short-read WGS data from the PPMI cohort and the computational tool str-analysis to genetically screen 1609 individuals for the biallelic “AAGGG” repeat expansion and identified four PD patient carriers and no control carriers giving a frequency of 0.44% in PD. To note, when we excluded carriers of known pathogenic variants in *LRRK2*, *GBA1*, and *SNCA*, and those with scans without evidence of dopaminergic deficits (SWEDD), the estimated frequency in PD is higher (0.84%), which is slightly higher than what was previously reported in the Finnish population, who report a frequency of 0.53% in PD patients⁵.

Interestingly, the reported clinical phenotype of these patients is in line with typical PD symptoms and no clear red flags in the clinical data were observed that the diagnosis was incorrect. However, it is worth noting that no specific ataxia phenotype data is collected and therefore we cannot

exclude misdiagnosis. Actually, only one out of three patients that had data available showed a positive reaction in aSyn SAA. Notably, SAA positivity is generally very high in PPMI PD cases and is influenced by genetic status. 67.5% of LRRK2 cases are SAA positive, whereas typical non-LRRK2 cases show a remarkably high SAA positivity rate of 93.3%.

As demonstrated in this study, long-read DNA sequencing is a powerful tool and a required step to validate potential pathogenic repeat expansion carriers. Short-read sequencing methods are notorious for over or underestimating repeat expansion lengths and the RFC1 locus is further complicated by its variable motif sequence. Therefore, although the allele frequency reported in this present study is inline if not slightly higher than what was identified in the Finnish study using PCRs for large (XL-PCR) amplicons and repeat primed PCR (see ref. 5), given the limitations, short-read sequencing can still lead to false negatives. As such, generating population-scale long-read DNA sequencing datasets to capture repeat expansions that are currently hidden using traditional methods is an essential step towards solving the architecture of complex genetic disorders⁶. For PD specifically, the Global Parkinson's Genetics Program (GP2 www.gp2.org) is leading a large-scale initiative to long-read DNA sequence ~1000 case-control samples (Fig. 1)

Methods

Cohort information

Samples were obtained from the Parkinson's Progression Markers Initiative (PPMI; <https://www.ppmi-info.org/>). Clinical and demographic characteristics of the PPMI cohort are shown in (Supplementary Table 1). Participants included PD cases clinically diagnosed by experienced neurologists and control individuals. All PD cases met the criteria defined by the UK PD Society Brain Bank⁷. All individuals were of European descent and were not age or gender-matched. This included a total of 903 cases and 706 neurologically healthy controls. PD cases ranged from 33 years to 90 years of age at diagnosis (mean 61.7 ± 11.05 , median 63.0) and included 62 individuals who showed SWEDD and 368 individuals who carry known genetic mutations associated with PD (within *LRRK2*, *GBA1*, and *SNCA*). Control subjects ranged from 19 years to 86 years of age (mean 58.3 ± 11.54 , median 60.0) and included 503 individuals who carry known genetic mutations associated with PD.x.

Short-read analysis. Short-read whole-genome sequencing data in bam format was downloaded through AMP-PD and has been reported in detail previously by Iwaki et al.⁸. For short-read data analysis, alignment was performed based on the GATK best practice pipeline, and the fastqs were aligned to the hg38 reference genome using BWA-mem. The STR detection tool str-analysis was used to screen for biallelic *RFC1* (AAGGG) expansion carriers (<https://github.com/broadinstitute/str-analysis>).

Long-read validation of expansion in carriers

To validate the five individuals predicted to carry pathogenic *RFC1* "AAGGG" biallelic repeat expansions, ONT whole-genome long-read DNA sequencing was performed. For all predicted carriers, a library was prepared from the DNA of the individuals with either the SQK-LSK110⁹ or the SQK-LSK114 ligation sequencing kit from ONT¹⁰. The samples were quantified using a Qubit fluorometer and were loaded onto a PromethION R9.4.1 (SQK-LSK110) or R.10.4 flow cell (SQK-LSK114) following ONT standard operating procedures and ran for a total of 72 h on a PromethION device (Supplementary Table 2).

Fast5 files containing the raw signal data were obtained from sequencing performed using MinKNOW (ONT). All fast5 files were used to perform super accuracy base calling on each sample with Guppy v6.0.1 (R.9) (ONT) or Dorado (v0.5.0). and sequencing statistics were obtained with seqkit v2.2.0 using fastq files that passed quality control filters in the super accuracy base calling. To accurately determine the length of *RFC1* repeat expansion from the ONT data, as required by tandem-genotypes, the fastqs were first mapped to the

Table 1 | Clinical characteristics of the four PD patients "AAGGG" biallelic *RFC1* expansion carriers

ID	Patient 1	Patient 2	Patient 3	Patient 4
Diagnosis	PD	PD	PD	PD
SAA	0	0	1	—
GBA1	—	—	—	—
LRRK2	—	—	—	G2019S
SEX	Female	Female	Female	0
FAMILY_HISTORY	1	0	0	1
Age at baseline	57.7	65.8	54.7	77.3
Age at onset	57.7	65.7	54.5	76
UPSIT	33	34	20	37
UPSIT percentile	20.5	50	1	NA
Hyposmia	0	0	1	NA
RBD	0	1	0	0
Depression	1	1	0	0
MCI	1	0	0	0
Constipation	1	1	0	1
Daytime sleepiness	0	0	0	0
Insomnia	1	1	0	1
Mean_caudate	3.72	2.93	2.43	NA
Mean_putamen	2.54	1.92	1.09	NA
Mean_striatum	3.13	2.42	1.76	NA
Follow-up years	1.51	1.91	7	—
Levodopa dosage	0	0	600	450
LEDD	0	0	900	900
Hoehn and Yahr stage	2	2	3	2
Tremor	1	1	1	NA
Rigidity	1	1	0	NA
Bradykinesia	1	1	1	NA
Postural instability	0	0	0	NA
MDS-UPDRS 1	9	10	24	19
MDS-UPDRS 2	5	3	17	27
MDS-UPDRS 3	24	22	18	35
MDS-UPDRS 4	0	0	0	0
MOCA	23	—	29	28

SAA seeding amplification assay, UPSIT the University of Pennsylvania Smell Identification Test, RBD rapid eye movement sleep behavior disorder, MCI mild cognitive impairment, LEDD levodopa equivalent daily dose, MDS-UPDRS Movement Disorder Society-sponsored revision of the unified Parkinson's disease rating scale, MOCAMo Montreal cognitive assessment.

hg38 reference using LAST as described in detail here (<https://github.com/mcfrith/last-rna/blob/master/last-long-reads.md>)¹¹. To size the expansion on each allele, tandem genotypes were then run using the mapped files¹².

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data used in the preparation of this article were obtained from the PPMI database (www.ppmi-info.org/access-data-specimens/download-data), RRID: SCR_006431. For up-to-date information on the study, visit www.ppmi-info.org. The PPMI cohort and the ONT raw data will be available at the LONI IDA.

Received: 16 October 2023; Accepted: 10 May 2024;
Published online: 24 May 2024

References

1. Cortese, A. et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat. Genet.* **51**, 649–658 (2019).
2. Cortese, A. et al. Repeat expansion size predicts age of onset in RFC1 CANVAS and disease spectrum (S29.005). *Neurology* **98**, (2022).
3. Wan, L. et al. Biallelic intronic AAGGG expansion of RFC1 is related to multiple system atrophy. *Ann. Neurol.* **88**, (2020).
4. Sullivan, R. et al. Letter: RFC1-related ataxia is a mimic of early multiple system atrophy. *J. Neurol. Neurosurg. Psychiatry* **92**, 444 (2021).
5. Kytövuori, L. et al. Biallelic expansion in RFC1 as a rare cause of Parkinson's disease. *NPJ Parkinsons Dis.* **8**, 6 (2022).
6. De Coster, W., Weissensteiner, M. H. & Sedlazeck, F. J. Towards population-scale long-read sequencing. *Nat. Rev. Genet.* **22**, 572–587 (2021).
7. Gelb, D. J., Oliver, E. & Gilman, S. Diagnostic Criteria for Parkinson Disease. *Archives Neurol* **56**, 33 Preprint at <https://doi.org/10.1001/archneur.56.1.33> (1999).
8. Iwaki, H. et al. Accelerating medicines partnership: Parkinson's disease. Genetic resource. *Mov. Disord.* **36**, 1795–1804 (2021).
9. J Billingsley, K. Processing frozen human blood samples for population-scale Oxford nanopore long-read DNA sequencing SOP v1. <https://doi.org/10.17504/protocols.io.ewov1n93ygr2/v1> (2022).
10. Miano-Burkhardt, A. Processing frozen human blood samples for population-scale SQK-LSK114 Oxford nanopore long-read DNA sequencing SOP v1. <https://doi.org/10.17504/protocols.io.x54v9py8qg3e/v1> (2023).
11. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
12. Mitsuhashi, S. et al. Tandem-genotypes: robust detection of tandem repeat expansions from long DNA reads. *Genome Biol.* **20**, 58 (2019).

Acknowledgements

We would like to thank all of the participants who donated their time and biological samples to be a part of this study. This work was supported in part by the Intramural Research Programs of the National Institute on Aging (NIA) and the National Institute of Neurological Disorders and Stroke (NINDS), part of the National Institutes of Health, Department of Health and Human Services; project numbers AG000542, Z01-AG000949, 1Z1ANS003154. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). Short-read WGS data used in the preparation of this article was obtained from the Accelerating Medicine Partnership® (AMP®) Parkinson's Disease (AMP PD) knowledge platform. For up-to-date information on the study, visit <https://www.amp-pd.org>. The AMP® PD program is a public-private partnership managed by the Foundation for the National Institutes of Health and funded by the NINDS in partnership with the Aligning Science Across Parkinson's (ASAP) initiative; Celgene Corporation, a subsidiary of Bristol-Myers Squibb Company; GlaxoSmithKline plc (GSK); The Michael J. Fox Foundation for Parkinson's Research; Pfizer Inc.; Sanofi US Services Inc.; and Verily Life Sciences. ACCELERATING MEDICINES PARTNERSHIP and AMP are registered service marks of the U.S. Department of Health and Human Services. Clinical data and biosamples used in the preparation of this article were obtained from the MJFF Parkinson's Progression Marker Initiative (PPMI). PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, Abbvie, AcureX, Allergan, Amathus Therapeutics, ASAP, AskBio, Avid Radiopharmaceuticals, BIAL, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo

Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Janssen Neuroscience, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager Therapeutics, the Weston Family Foundation, and Yumanity Therapeutics. The PPMI Investigators have not participated in reviewing the data analysis or content of the manuscript. For up-to-date information on the study, visit www.ppmi-info.org. We would also like to thank the team at PPMI for sending frozen blood samples to complete the long-read DNA validation, specifically; Tatiana M. Foroud, Jan E. Hamer, Caitlin D. Schulz, Bradford Casey, and Mark Frasier. We would also like to thank Ben Weisburd for his guidance with the str-analysis tool. K. Daida was supported by the JSPS research fellowship for Japanese biomedical and behavioral researchers at NIH.

Author contributions

P.A.J., R.S., J.V., H.H., C.B., A.B.S., J.H., and K.J.B. designed, executed, reviewed, and critiqued the study. P.A.J., K.D., A.M.-B., L.M., J.D., J.R.B., A.M., M.A.N., R.K.K., F.J.S., B.C., and K.J.B. ran the analysis and generated the long-read data for validation. H.I., G.C., and M.B.M. reviewed and critiqued the analysis.

Funding

Open access funding provided by the National Institutes of Health.

Competing interests

ABS is an editor for npj Parkinson's Disease. ABS was not involved in the journal's review of, or decisions related to, this manuscript. All remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41531-024-00723-0>.

Correspondence and requests for materials should be addressed to Kimberley J. Billingsley.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024