Using the Genomics England National Genomic Research Library (NGRL) and UK Biobank to investigate the genetic, phenotypic and clinical landscape of thymidine kinase 2 deficiency (TK2d)

Genomics of Rare Disease; Hinxton, Cambridge, UK; 9–11 April 2025

Miruna Carmen Barbu[1]*, Kaja Zarakowska[2], Martin Armstrong[3], Kate Witkowska[1], Olga Giannakopoulou[4], James Staley[4], Loukas Moutsianas[1], Ella Davyson[1,5], Nour Elkhateeb[1], Ana Lisa Taylor Tavares[1], Chris Odhams[1], Robert McFarland[6,7]

[1]Genomics England, London, UK; [2]UCB, Bulle, Switzerland; [3]UCB, Braine-l'Alleud, Belgium; [4]UCB, Slough, UK; [5]School of Informatics, The University of Edinburgh, Edinburgh, UK; [6]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK; [7]NHS Rare Mitochondrial Disorders Service, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

***Presenting author:** Miruna Carmen Barbu (MirunaCarmen.Barbu@genomicsengland.co.uk)

Introduction

- Thymidine kinase 2 deficiency (TK2d) is an ultra-rare, autosomal recessive, mitochondrial myopathy associated with progressive proximal muscle weakness, respiratory insufficiency and premature death[1–3]
- More than 60 pathogenic variants in the thymidine kinase 2 gene (*TK2*) have been identified[4,5]
  - Many further *TK2* variants of uncertain significance exist,[4,5] with insufficient scientific evidence currently available to reclassify them
- Diagnosis of TK2d is confounded by:
  - a lack of awareness among healthcare practitioners[6]
  - the inherently heterogeneous disease presentation, which also often overlaps with other neuromuscular and mitochondrial myopathies[2]
  - the limited understanding of genotype–phenotype relationships for this disease[2]
- Recognition and early diagnosis of TK2d is important to facilitate appropriate disease management as well as access to emerging treatments[2,7]
  - Although there are no approved therapies for TK2d,[8] an oral pyrimidine nucleoside therapy for TK2d is in clinical development[9]
- Large sequencing datasets such as the National Genomic Research Library (NGRL) and UK Biobank could provide deeper insights into the genetic, phenotypic and clinical landscape of TK2d, thereby facilitating improvements in the diagnostic yield[7]
- The NGRL, managed by Genomics England, is a secure database of de-identified genomic and health data, enriched for rare diseases.[10] These data are from participants enrolled from the following sources
  - The 100,000 Genomes Project (100kGP)[11]
  - The UK National Health Service Genomic Medicine Service (NHS GMS)[12]
- UK Biobank is a large-scale prospective population study of ~500,000 adults from across the UK[13]

Objective

- To use the NGRL to identify and characterize small and structural *TK2* variants; to characterize participant phenotypes with variation in *TK2*, using UK Biobank as an external cohort to compare findings; and to perform segregation-based filtering of the identified *TK2* variants

Methods

NGRL cohorts

- Whole-genome sequencing data were included from two de-identified participant cohorts in the NGRL, namely:
  - the 100kGP, data release version 17 (March 2023)
  - the NHS GMS, data release version 3 (March 2024)

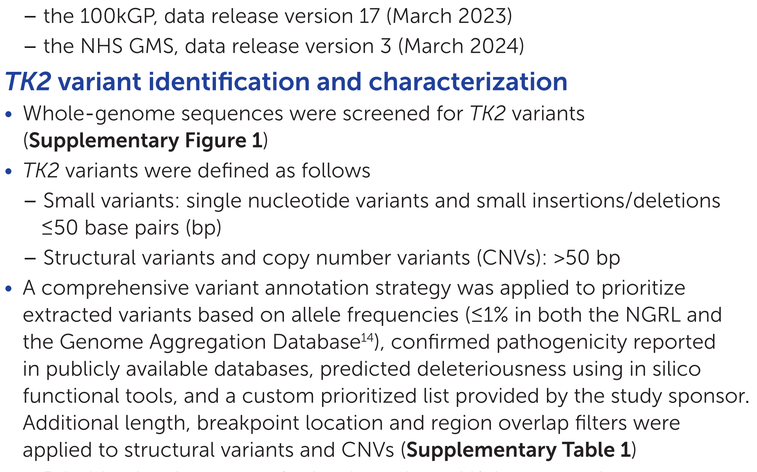
TK2 variant identification and characterization

- Whole-genome sequences were screened for *TK2* variants (**Supplementary Figure 1**)
- *TK2* variants were defined as follows
  - Small variants: single nucleotide variants and small insertions/deletions ≤50 base pairs (bp)
  - Structural variants and copy number variants (CNVs): >50 bp
- A comprehensive variant annotation strategy was applied to prioritize extracted variants based on allele frequencies (≤1% in both the NGRL and the Genome Aggregation Database[14]), confirmed pathogenicity reported in publicly available databases, predicted deleteriousness using in silico functional tools, and a custom prioritized list provided by the study sponsor. Additional length, breakpoint location and region overlap filters were applied to structural variants and CNVs (**Supplementary Table 1**)
  - Prioritized variants were further investigated if they were alternate homozygous or compound heterozygous in at least one participant

Participant prioritization and segregation analysis

- An expert clinical review was conducted to systematically screen participants carrying prioritized variants for phenotypes relevant to TK2d, utilizing various ontologies (recruited rare disease, Human Phenotype Ontology [HPO] and International Statistical Classification of Diseases and Related Health Problems 10th Revision [ICD-10]; **Supplementary Table 2**)
  - Secondary phenotypic data for some participants in the NHS GMS cohort were later linked after the initial analysis had been conducted
- Family members of participants carrying prioritized variants were also screened for phenotypes relevant to TK2d to assess penetrance within families
- Participants carrying prioritized variants were categorized into three groups based on zygosity and phenotypic information (**Figure 1**)
- Families were constructed from participants in groups 1 and 2. Variant segregation patterns within families and variant penetrance within families and across the entire cohort were assessed

Figure 1. Participant grouping



Penetrance was assessed by screening family members of participants carrying prioritized variants for TK2d-related phenotypes
TK2d, thymidine kinase 2 deficiency

Cross-referencing variants with an external cohort

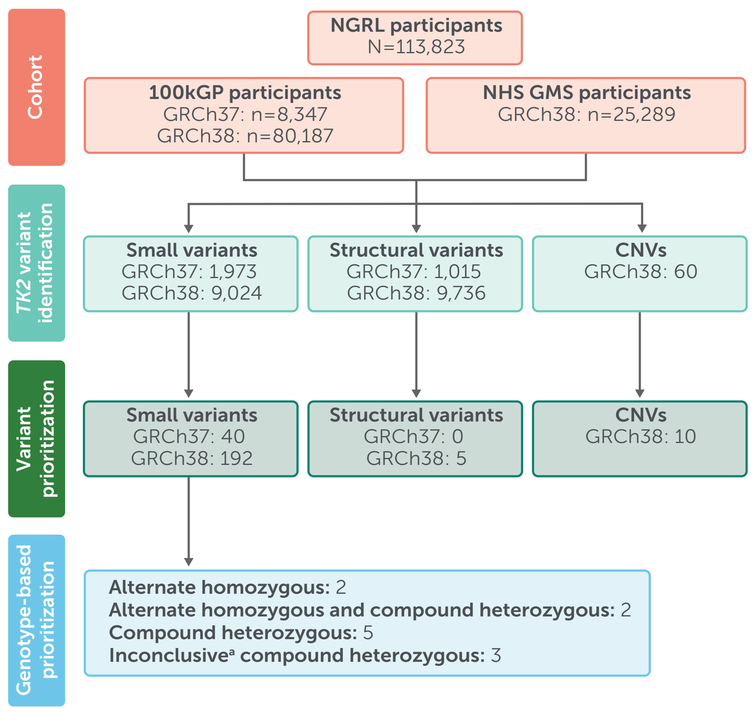
- To further investigate their pathogenic significance, prioritized alternate homozygous small variants identified in the NGRL were screened for in UK Biobank
  - Genotype analysis used whole-exome sequencing data aligned to Genome Reference Consortium human build 38 (GRCh38)
  - Phenotypic investigations included hospital admission records as of April 2024 and expert clinical screening of disease terms

Results

Participants and variant prioritization

- In total, 113,823 participants were included from the NGRL (100kGP, n=88,534; NHS GMS, n=25,289; **Figure 2**)
  - Prioritized for further investigation were 12 small variants, 5 structural variants and 10 CNVs (**Figure 2**; **Supplementary Table 3**)
  - Prioritization criteria were not met by 98% of small variants, 99.8% of structural variants and 88.5% of CNVs

Figure 2. Process flowchart of variant prioritization



Most participants had genomes aligned to GRCh38; approximately 9% of participants in the 100kGP cohort had genomes aligned to GRCh37
[a]Inconclusive compound heterozygous variants are those that could not be confirmed through family structure or phasing information
100kGP, 100,000 Genomes Project; CNV, copy number variant; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; NGRL, National Genomic Research Library; NHS GMS, National Health Service Genomic Medicine Service; *TK2*, thymidine kinase 2 gene

Characterization of participants with prioritized variants

- Prioritized variants carried by participants categorized into groups 1 and 2 are outlined in **Table 1**
  - Prior to this analysis, variant p.Thr108Met has previously been identified through the Genomics England Rare Disease Interpretation and Diagnostic Pipeline and reported as 'pathogenic' in relation to TK2d (ClinVar ID: 12710)
  - Variant p.Arg32Trp has previously been reported as 'benign/likely benign' in relation to TK2d (ClinVar ID: 215261)
  - The structural deletion has previously been reported as of 'uncertain significance' in VarSome but has not been reported in ClinVar
- All other prioritized variants were carried by participants without TK2d-related phenotypes and categorized into group 3, suggesting that they are not implicated in TK2d

Table 1. Variants carried by prioritized participants in groups 1 and 2

Group	Type	Variant ID	Variant and location	AF	Zygosity in prioritized participants and family members	n in prioritized participants and family members
1	Small missense variant	chr16:66531432_G_A	p.Thr108Met in exon 5	3.74×10^{-5}	Alternate homozygous	<5
2	Small missense variant	chr16:66583871_G_A (GRCh37)/ chr16:66549968_G_A (GRCh38)	p.Arg32Trp in exon 1	$\leq6.01\times10^{-3}$	Alternate homozygous	6
					Heterozygous	15
2	Structural deletion	chr16:66543128_ 66546738_C_<DEL>	Feature truncating intronic variant 3,611 bp in length spanning intron 2	1.25×10^{-5}	Heterozygous	<5

AF, allele frequency; bp, base pairs; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38

Penetrance of variants carried by prioritized participants in groups 1 and 2

- Following the application of variant-disease segregation filters, the penetrance of prioritized variants carried by participants in groups 1 and 2 was assessed (**Table 2**)
  - Small variant p.Thr108Met followed an assumed simple recessive segregation and exhibited complete penetrance
  - Small variant p.Arg32Trp followed both simple recessive and autosomal dominant segregations across families of participants; participants carrying this variant as alternate homozygous without TK2d-related phenotypes were observed, indicating reduced penetrance
  - The structural deletion of 3,611 bp spanning intron 2 followed an assumed autosomal dominant segregation and was observed in participants with age of first phenotypic TK2d presentation ≤2 years but also in older participants (currently aged >30 years) with no TK2d-related phenotypes, indicating reduced penetrance

Table 2. Variant-specific counts across family structures for prioritized participants in groups 1 and 2

			Exhibits TK2d-related phenotypes					Does not exhibit TK2d-related phenotypes				
Group	Variant ID	Genotype	Singleton	Sibling-pair	Duo	Trio	Quintet	Singleton	Sibling-pair	Duo	Trio	Quintet
1	chr16:66531432_G_A	Alternate homozygous	<5									
2	chr16:66583871_G_A (GRCh37)/ chr16:66549968_G_A (GRCh38)	Alternate homozygous or heterozygous		<5		<5	<5			<5	<5	
2	chr16:66543128_ 66546738_C_<DEL>	Heterozygous	<5					<5				

GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; TK2d, thymidine kinase 2 deficiency

UK Biobank analysis

- Data from 469,707 participants were included in the analysis
  - No UK Biobank participants were alternate homozygous carriers of p.Thr108Met
  - In total, 25 participants were alternate homozygous carriers of p.Arg32Trp; none of these participants had TK2d-related phenotypes, validating the benign classification of this variant

Conclusions and Outlook

This multidisciplinary study demonstrates how large-scale whole-genome and -exome sequencing datasets and deep phenotyping can be used to study ultra-rare diseases such as TK2d

- Variant p.Thr108Met followed an assumed simple recessive segregation, consistent with a biallelic mode of inheritance of TK2d, and exhibited complete penetrance, indicating that it is a key contributor to the disease
- Variants carried by participants in group 2 following an assumed autosomal dominant segregation and exhibiting reduced penetrance are likely benign in relation to TK2d
- Cross-referencing findings from the NGRL, which is enriched for rare diseases, with a nominally healthy external cohort such as UK Biobank provided further evidence for variant classification

The *TK2* variants identified in this study could be further explored by integrating multi-omics data and cross-referencing with TK2d-focused datasets

A deeper understanding of the genetic and clinical architecture of TK2d could potentially:

- facilitate improvements in diagnostic approaches and prevalence estimates
- support genetic counselling, helping to inform families and clinicians about carrier frequency and inheritance risks
- influence future policy recommendations regarding newborn screening or genetic testing protocols
- provide valuable data for clinical trials researching drug development and therapeutic advancements for TK2d

Acknowledgements: This analysis was funded by UCB. This research was made possible through access to data and findings in the National Genomic Research Library via the Genomics England Research Environment. This research has been conducted using UK Biobank under Application Number 44411. The authors acknowledge Tamsin Chambers PhD of PharmaGenesis Cambridge, Cambridge, UK for writing and editorial assistance, which was funded by UCB, in accordance with Good Publication Practice 2022 (GPP 2022) guidelines (https://www.ismpp.org/gpp-2022). The authors would like to thank Margarita Lens MSci CMPP of UCB for publication coordination. The authors thank the patients and their caregivers, in addition to the investigators and their teams who contributed to this study.
References: 1. Garone C, *et al. J Med Genet* 2018;55:515–21. 2. Berardo A, *et al. J Neuromuscul Dis* 2022;9:225–35. 3. Wang J, *et al. Mol Genet Metabol* 2018;124:124–30. 4. National Library of Medicine. ClinVar [Internet]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/?term=TK2%5BGene%5D (Accessed 5 March 2025). 5. VarSome. Gene: TK2. [Internet]. Available from: https://varsome.com/gene/hg38/TK2 (Accessed 5 March 2025). 6. Grier J, *et al. Neurol Genet* 2018;4:e230. 7. Stenton SL, Prokisch H. *EBioMedicine* 2020;56:102784. 8. de Barcelos IP, *et al. Curr Opin Neurol* 2019;32:715–21. 9. ClinicalTrials.gov [Internet]. NCT03845712. Available from: https://classic.clinicaltrials.gov/ct2/show/NCT03845712 (Accessed 5 March 2025). 10. Genomics England. The National Genomic Research Library v5.1. 2020. Available from: https://figshare.com/articles/dataset/GenomicEnglandProtocol_pdf/4530893?file=22714349 (Accessed 26 March 2025). 11. Smedley D, *et al. N Engl J Med* 2021;385:1868–80. 12. Snape K, *et al. Clin Med (Lond)* 2019;19:273–7. 13. Bycroft C, *et al. Nature* 2018;562:203–9. 14. gnomAD browser [Internet]. Available from: https://gnomad.broadinstitute.org/ (Accessed 5 March 2025).
Disclosures: **Miruna Carmen Barbu**, **Kate Witkowska**, **Loukas Moutsianas**, **Nour Elkhateeb**, **Ana Lisa Taylor Tavares** and **Chris Odhams** are employees of Genomics England. **Kaja Zarakowska**, **Martin Armstrong**, **Olga Giannakopoulou** and **James Staley** are employees of and stockholders in UCB. **Ella Davyson** received funding from Genomics England as part of a 6-month internship and receives funding from the UK Research and Innovation Centre for Doctoral Training in Biomedical Artificial Intelligence (grant EP/S02431X/1) at The University of Edinburgh School of Informatics. **Robert McFarland** receives funding from Action Medical Research, the Leigh Syndrome International Consortium, The Lily Foundation and Wellcome, and has consultancy agreements with Abliva, Pretzel Therapeutics and UCB.

Please use this QR code to access the supplementary data and download a PDF of this poster.

Using the Genomics England National Genomic Research Library (NGRL) and UK Biobank to investigate the genetic, phenotypic and clinical landscape of thymidine kinase 2 deficiency (TK2d)

Genomics of Rare Disease; Hinxton, Cambridge, UK; 9–11 April 2025

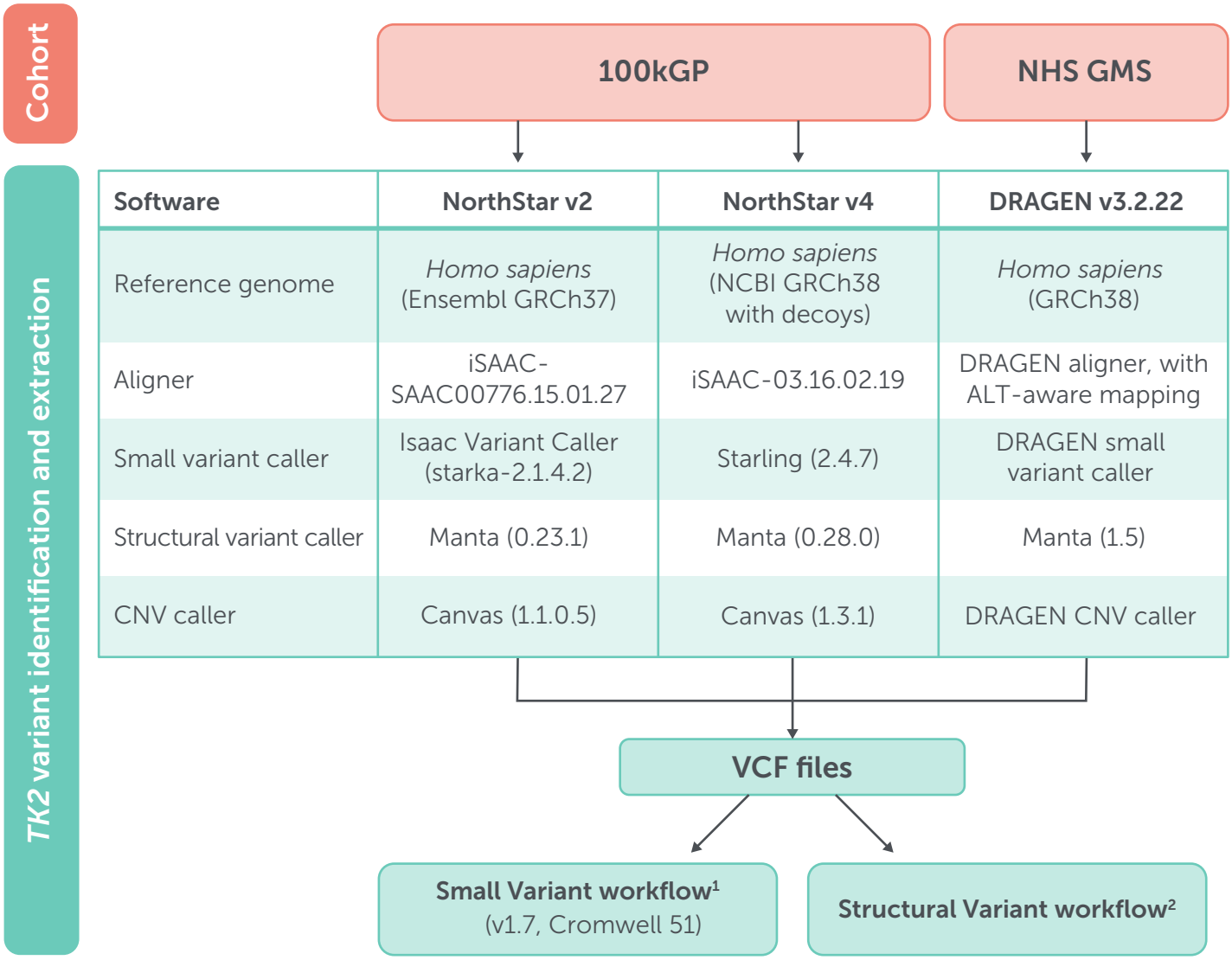
Miruna Carmen Barbu[1]*, Kaja Zarakowska[2], Martin Armstrong[3], Kate Witkowska[1], Olga Giannakopoulou[4], James Staley[4], Loukas Moutsianas[1], Ella Davyson[1,5], Nour Elkhateeb[1], Ana Lisa Taylor Tavares[1], Chris Odhams[1], Robert McFarland[6,7]

[1]Genomics England, London, UK; [2]UCB, Bulle, Switzerland; [3]UCB, Braine-l'Alleud, Belgium; [4]UCB, Slough, UK; [5]School of Informatics, The University of Edinburgh, Edinburgh, UK; [6]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK; [7]NHS Rare Mitochondrial Disorders Service, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

***Presenting author:** Miruna Carmen Barbu (MirunaCarmen.Barbu@genomicsengland.co.uk)

Supplementary materials: Supplementary Figure 1 and Supplementary Tables 1, 2 and 3

Supplementary Figure 1. Workflow for identifying *TK2* variants within whole-genome sequencing data



Approximately 9% of participants in the 100kGP cohort had genomes aligned to GRCh37

The canonical transcript for both genome builds was investigated (GRCh37 transcript: ENST00000451102.2; GRCh38 transcript: ENST00000544898.6)

100kGP, 100,000 Genome Project; ALT, alternate haplotype; CNV, copy number variant; DRAGEN, Dynamic Read Analysis for GENomics; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; NCBI, US National Center for Biotechnology Information; NHS GMS, National Health Service Genomic Medicine Service; *TK2*, thymidine kinase 2 gene; v, version; VCF, variant call format

Supplementary Table 1. Tools and databases used for variant prioritization

Category	Resource	Description	Variant filter
Frequency annotations	Genomics England Internal Allele Frequency (global)	AF calculated on Genomics England datasets (100kGP GRCh37; 100kGP GRCh38; NHS GMS GRCh38)	AF ≤0.01 (1%)
	gnomAD Allele Frequency (global)	AF in gnomAD dataset • Small variants: exomes for GRCh37 v2.1; genomes for GRCh38 v3.1.2 • Structural variants: GRCh37, v2.1; GRCh38, v4.0 • CNVs: v4.0	AF ≤0.01 (1%)
Custom	Custom variants	Custom variant list provided by the study sponsor	NA
Curated database	ClinVar (v01/19/2024)	Variant pathogenicity in relation to disease[3]	Pathogenic or likely pathogenic for TK2d
	VarSome (v02/21/2024)	Variant pathogenicity based on ACMG guidelines based on population frequency, aggregated pathogenicity predictions and links to HPO phenotypes[4]	Pathogenic or likely pathogenic
	Genomic Medicine Centre Exit Questionnaire (small variants only)	Genomics England internal dataset capturing clinical actionability of variants reported via ISO accredited pipeline (this information is only available for probands)	Pathogenic or likely pathogenic; whether a case is solved based on the variant(s), and if so, which disease was solved
In silico prediction	VEP (v110)	Predict the effect and consequence of variants	Moderate or high predicted consequence[a]
	Small variants: CADD (v1.6) Structural variants: CADD-SV (v1)	Deleteriousness of small variants or of deletions, insertions and duplications	CADD or CADD-SV score ≥15 (top ~3% most deleterious variants in the genome)
	REVEL (v1; small variants only)	Missense variant pathogenicity	Score ≥0.5
	LOFTEE (v1.0.4; small variants only)	Loss-of-function variation in stop-gained, splice-site disrupting and frameshift variants	High-confidence rare variants
	SpliceAI (v1.3; small variants only)	Predicted effect on splicing	Delta score ≥0.5
Structural and copy number variation filters (not applied to small variants)	Structural variation filter	Maximum structural variant length	≤500 kb in size
	Structural variation filter	Inversion breakpoint location	Inversions occurring within 5 Mb of the start or end point of *TK2* or within *TK2*
	CNV filter	Region overlap filter for copy number variation	>95% region overlap between CNV gains >95% region overlap between CNV losses
Allele segregation	Trio assessment via family structure and phasing data	Assessment of family structure for participants carrying more than one small variant to confirm compound heterozygosity	Parent-of-origin segregation in trio or phasing data available

Not all participants will form complete trios as singletons and duos are also recruited into the 100kGP and the NHS GMS cohorts

[a]VEP moderate and high classifications include protein-altering, missense, inframe-deletion, inframe-insertion, transcript-amplification, start-lost, stop-lost, frameshift, stop-gained, splice-donor, splice-acceptor and transcript-ablation variants

100kGP, 100,000 Genome Project; ACMG, American College of Medical Genetics and Genomics; AF, allele frequency; CADD, Combined Annotation Dependent Depletion; CADD-SV, CADD Structural Variant; CNV, copy number variant; gnomAD, Genome Aggregation Database; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; HPO, Human Phenotype Ontology; ISO, International Organization for Standardization; kb, kilobases; LOFTEE, Loss-Of-Function Transcript Effect Estimator; Mb, megabases; NA, not applicable; NHS GMS, National Health Service Genomic Medicine Service; REVEL, rare exome variant ensemble learner; *TK2*, thymidine kinase 2 gene; TK2d, thymidine kinase 2 deficiency; v, version; VEP, Variant Effect Predictor

Supplementary Table 2. Disease domains across TK2d phenotypes

Disease domain	Ontology	Inclusion phenotypes	Term code
Muscle weakness	Rare disease	Limb girdle muscular dystrophy; hypotonic infant; muscular dystrophy; myopathy	
	HPO term	Dysarthria; dysphagia; neck muscle weakness; ptosis; limb girdle muscular dystrophy; axial muscle weakness; distal lower limb muscle weakness; distal upper limb muscle weakness; muscle weakness; progressive muscle weakness; proximal lower limb amyotrophy; proximal muscle weakness in lower limbs; proximal muscle weakness in upper limbs; proximal upper limb amyotrophy; scapular muscle atrophy; scapular winging; skeletal muscle atrophy; abnormal skeletal muscle morphology; generalized hypotonia; bulbar signs; spinal rigidity	HP:0001260; HP:0002015; HP:0000467; HP:0000508; HP:0006785; HP:0003327; HP:0009053; HP:0008959; HP:0001324; HP:0003323; HP:0008956; HP:0008994; HP:0008997; HP:0008948; HP:0009060; HP:0003691; HP:0003202; HP:0011805; HP:0001290; HP:0002483; HP:0003306
	ICD-10 term	Myopathy, unspecified; dysphagia; ptosis of eyelid; progressive external ophthalmoplegia	G72.9; R13; H02.4; H49.4
Movement phenotypes	HPO term	Difficulty walking; progressive inability to walk; gait disturbance	HP:0002355; HP:0002505; HP:0001288
	ICD-10 term	Tendency to fall, not elsewhere classified	R29.6
Creatine kinase level	HPO term	Abnormal circulating creatine kinase circulation	HP:0040081
Anaemia	ICD-10 term	Anaemia, unspecified	D64.9
Development	Rare disease	Intellectual disability	
	HPO term	Delayed fine and/or gross motor development; global developmental delay; failure to thrive	HP:0010862; HP:0002194; HP:0001263; HP:0001508
Epilepsy	HPO term	Seizure	HP:0001250
	ICD-10 term	Localization-related (focal) (partial) symptomatic epilepsy and epileptic syndromes with simple partial seizures; localization-related (focal) (partial) symptomatic epilepsy and epileptic syndromes with simple complex seizures); grand mal seizures, unspecified (with or without petit mal)	G40.1; G40.2; G40.6
Hearing phenotypes	Rare disease	Congenital hearing impairment	
	HPO term	High-, mid- or low-frequency hearing loss; bilateral, congenital or profound sensorineural hearing impairment	HP:0005101; HP:0012781; HP:0008542; HP:0008619; HP:0008527; HP:0011476; HP:0000407
	ICD-10 term	Sensorineural hearing loss, bilateral; hearing loss, unspecified	H90.3; H91.9
Unrelated phenotypes		Phenotype unrelated to TK2d (all other disease terms not listed above)	

HPO, Human Phenotype Ontology; ICD-10, International Classification of Diseases and Related Health Problems 10th Revision; TK2d, thymidine kinase 2 deficiency

Supplementary Table 3. Prioritized variants (page 1 of 2)

Variant type	Variant details	Variant ID	AF (cohort)	Zygosity	Prioritization annotation
Small variant	p.Thr108Met, exon 5, missense	chr16:66531432_G_A	3.74×10^{-5}	Alternate homozygous	CADD score: 25.5 REVEL score: 0.853 VEP: consequence with moderate impact ClinVar: pathogenic for mitochondrial DNA depletion syndrome and myopathic form VarSome: pathogenic GMC exit questionnaire: likely pathogenic variant for TK2d UCB priority list
	c.156+6T>G, intron 2, splice donor region and intronic variant	chr16:66548972_A_C	1.19×10^{-4} (NHS GMS)	Alternate homozygous	CADD score: 17.49 UCB priority list
	p.Arg32Trp, exon 1, missense	chr16:66583871_G_A (GRCh37) chr16:66549968_G_A (GRCh38)	5×10^{-3} (100kGP GRCh37) 6×10^{-3} (100kGP GRCh38) 6.01×10^{-3} (NHS GMS)	Alternate homozygous and compound heterozygous	CADD score: 15.82 (GRCh37), 19.09 (GRCh38) VEP: consequence with moderate impact
	p.Pro41His, exon 1, missense	chr16:66549940_G_T	7.48×10^{-4} (100kGP GRCh38) 1.0×10^{-3} (NHS GMS)	Alternate homozygous and compound heterozygous Inconclusive compound heterozygous	VEP: consequence with moderate impact
	p.Gly28Asp, exon 1, missense	chr16:66549979_C_T	$<1\times10^{-4}$ (100kGP GRCh38)	Compound heterozygous	VEP: consequence with moderate impact CADD score: 15.61
	p.Pro227Leu, exon 9, missense	chr16:66513750_G_A	3.06×10^{-4} (100kGP GRCh38) $<1\times10^{-4}$ (NHS GMS)	Compound heterozygous	VEP: consequence with moderate impact CADD score: 24.2
	Intron 2, intronic variant	chr16:66548066_C_A	3.74×10^{-5} (100kGP GRCh38)	Compound heterozygous	VarSome: likely pathogenic
	p.Glu207Val, exon 9, missense	chr16:66513810_T_A	1.62×10^{-4} (100kGP GRCh38)	Compound heterozygous	VEP: consequence with moderate impact CADD score: 32 REVEL score: 0.92
	Intron 4, intronic variant	chr16:66533657_A_G	$<1\times10^{-4}$ (100kGP GRCh38)	Compound heterozygous	CADD score: 15.95
	p.Arg12Gln, exon 1, missense	chr16:66550027_C_T	$<1\times10^{-4}$ (100kGP GRCh38, NHS GMS)	Inconclusive compound heterozygous	VEP: consequence with moderate impact CADD score: 15.12
	p.Val174Leu, exon 7, missense	chr16:66517807_C_G	6.24×10^{-5} (100kGP GRCh38) $<1\times10^{-4}$ (NHS GMS)	Inconclusive compound heterozygous	VEP: consequence with moderate impact CADD score: 15.69

VEP moderate and high classifications include protein-altering, missense, inframe-deletion, inframe-insertion, transcript-amplification, start-lost, stop-lost, frameshift, stop-gained, splice-donor, splice-acceptor and transcript-ablation variants

100kGP, 100,000 Genome Project; AF, allele frequency; bp, base pairs; CADD, Combined Annotation Dependent Depletion; CN, copy number; CNV, copy number variant; GMC, Genomic Medicine Centre; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; NA, not applicable; NHS GMS, National Health Service Genomic Medicine Service; REVEL, rare exome variant ensemble learner; *TK2*, thymidine kinase 2 gene; TK2d, thymidine kinase 2 deficiency; VEP, Variant Effect Predictor

Supplementary Table 3. Prioritized variants (page 2 of 2)

Variant type	Variant details	Variant ID	AF (cohort)	Zygosity	Prioritization annotation
Structural variant	Deletion, 4,844 bp, exon 4, introns 3–4, 10.9% *TK2* overlap	chr16:66532535_66537378	$<1\times10^{-4}$	Heterozygous	VEP: consequence with high impact UCB list: entire exon 4 deletion
	Deletion, 3,611 bp, intron 2, 8.1% *TK2* overlap	chr16:66543128_66546738	1.25×10^{-5}	Heterozygous	VEP: consequence with high impact
	Tandem duplication, 17,000 bp, exon 10, 0.5% *TK2* overlap	chr16:66491219_66508219	3.95×10^{-5}	Heterozygous	Exonic overlap: exon 10
	Tandem duplication, 5,168 bp, intron 6, 11.6% *TK2* overlap	chr16:66522744_66527912	1.98×10^{-5}	Heterozygous	VEP: consequence with high impact
	Deletion, 4,946 bp, exon 4, introns 3–4, 11.1% *TK2* overlap	chr16:66532520_66537466	3.95×10^{-5}	Heterozygous	VEP: consequence with high impact Exonic overlap: exon 4 UCB priority list
CNV	Gain, 3 CN, entire *TK2*, 100% *TK2* overlap	chr16:66461633:66598120	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, entire *TK2*, 100% *TK2* overlap	chr16:66475158:66601170 chr16:66475286:66598732	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, entire *TK2*, 100% *TK2* overlap	chr16:66481054:66600133 chr16:66480872:66599762 chr16:66481992:66599476	NA	NA	Spans *TK2* exonic region
	Gain, 4 CN, entire *TK2*, 100% *TK2* overlap	chr16:66494759:66604087	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, entire *TK2*, 100% *TK2* overlap	chr16:66502527:66564503	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, exons 1–4, introns 1–4, 40% *TK2* overlap	chr16:66534731:66642194	NA	NA	Spans *TK2* exonic region
	Gain, 4 CN, exons 1–2, introns 1–2, 18.6% *TK2* overlap	chr16:66544270:66583794	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, exons 1–5, introns 1–4, 47.3% *TK2* overlap	chr16:66531467:66637780	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, exons 1–3, introns 1–3, 31% *TK2* overlap	chr16:66540035:66637780, chr16:66538747:66637780, chr16:66538747:66639044	NA	NA	Spans *TK2* exonic region
	Gain, 3 CN, exons 1–3, introns 1–3, 28.1% *TK2* overlap	chr16:66540035:66644812	NA	NA	Spans *TK2* exonic region

VEP moderate and high classifications include protein-altering, missense, inframe-deletion, inframe-insertion, transcript-amplification, start-lost, stop-lost, frameshift, stop-gained, splice-donor, splice-acceptor and transcript-ablation variants

100kGP, 100,000 Genome Project; AF, allele frequency; bp, base pairs; CADD, Combined Annotation Dependent Depletion; CN, copy number; CNV, copy number variant; GMC, Genomic Medicine Centre; GRCh37, Genome Reference Consortium human build 37; GRCh38, Genome Reference Consortium human build 38; NA, not applicable; NHS GMS, National Health Service Genomic Medicine Service; REVEL, rare exome variant ensemble learner; *TK2*, thymidine kinase 2 gene; TK2d, thymidine kinase 2 deficiency; VEP, Variant Effect Predictor

References: 1. Genomics England Research Environment User Guide [Internet]. Small Variant workflow. Available from: https://re-docs.genomicsengland.co.uk/small_variant/ (Accessed 4 March 2025). 2. Genomics England Research Environment User Guide [Internet]. Structural Variant workflow. Available from: https://re-docs.genomicsengland.co.uk/structural_variant/ (Accessed 4 March 2025). 3. Landrum MJ, *et al*. *Nucleic Acids Res* 2020;48:D835–44.
4. Kopanos CV, *et al. Bioinformatics* 2019;35:1978–80.