TDP-43 pathology induces CD8$^+$ T cell activation through cryptic epitope recognition

Shahab Chizari[1,*], Matteo Zanovello[2,*], Steven Kong[1], Vidur Saigal[1], Anna-Leigh Brown[2], Valentina Turchetti[2], Luca Zampedri[2], Iwona Skorupinska[2], Giacomo Maria Minicuci[3], Francesca Paron[4], Paola Tonin[5], Giulia Marchetto[5], Ziyi Li[6], Jennifer M. Colón-Mercado[7], Dario Dattilo[2], Simone Barattucci[2], Ariana Gatt[2], Andy Qi[6], Michael Hanna[2], Michael Ward[6,7], Leonard Petrucelli[8], Maurizio Romano[9], Gaetano Vattemi[5], Emanuele Buratti[4], Andrea Malapsina[2], Ashirwad Merve[2], Pedro M Machado[2,10], Gianni Soraru[3], Pietro Fratta[2,11,✉], Ning Jiang[1,12,13,14,15,16,✉]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, USA; [2]Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, UK; [3]Padua Neuroscience Centre, University of Padua, Italy; [4]International Centre for Genetic Engineering and Biotechnology; [5]Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Italy; [6]Center for Alzheimer's and Related Dementias, NIH; [7]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD, USA; [8]Mayo Clinic College of Medicine and Science, Mayo Clinic, USA; [9]Department of Life Sciences, University of Trieste, Italy; [10]NIHR University College London Hospitals Biomedical Research Centre; [11]The Francis Crick Institute, London, UK; [12]Penn Institute for Immunology and Immune Health, University of Pennsylvania, Philadelphia; [13]Center for Cellular Immunotherapies, University of Pennsylvania, Philadelphia; [14]Abramson Cancer Center, University of Pennsylvania, Philadelphia; [15]Institute for RNA Innovation, University of Pennsylvania, Philadelphia; [16]Penn Center for Precision Engineering for Health, University of Pennsylvania, Philadelphia, USA

*: These authors contributed equally. ✉: Corresponding authors.

Abstract

Aggregation and nuclear depletion of the RNA binding protein TDP-43 are the crucial pathological features of amyotrophic lateral sclerosis (ALS) and inclusion body myositis (IBM), two degenerative diseases of the CNS and muscle. The loss of TDP-43 nuclear function results in the aberrant inclusion of cryptic exons in mRNA transcripts, leading to the expression of *de novo* proteins. Clonally expanded and highly differentiated $CD8^+$ T cells have been observed in individuals with TDP-43 proteinopathies and therapeutics modulating the T cell response have recently been found to extend survival. However, the target antigens mediating T cell activation have remained elusive. Here, we investigate whether the *de novo* proteins induced by aberrant cryptic splicing due to TDP-43 nuclear loss can act as neo-antigens. We detect the HDGFL2 cryptic peptide and multiple other TDP-43 cryptic exons in IBM skeletal muscle, where their presence correlates with enrichment of T cells and class I antigen presentation pathways. Furthermore, we identify epitopes deriving from HDGFL2 and IGLON5 cryptic peptides which are recognized by clonally expanded and functionally differentiated populations of $CD8^+$ T cells in ALS and IBM Patients. Finally, we demonstrate that T cells engineered to express the identified TCRs can bind and activate in response to the cryptic peptide derived epitopes (cryptic epitopes) and are able to kill TDP-43 deficient astrocytes. This work identifies for the first time specific T cell antigens in ALS and IBM, directly linking adaptive immune response to TDP-43 pathology.

Introduction

Aggregation and nuclear depletion of the RNA-binding protein TDP-43 is a defining pathological feature of amyotrophic lateral sclerosis (ALS), inclusion body myositis (IBM), and several other neurodegenerative conditions[1,2]. TDP-43 is a ubiquitously expressed protein that resides predominantly in the nucleus, where it has been shown to be a repressor of cryptic exons, intronic sequences that are normally excluded from mRNA[3]. Upon TDP-43 mislocalisation to the cytoplasm, these cryptic exons become derepressed and incorporated into transcripts. Although most of these events generate frameshifts and premature stop codons leading to RNA degradation[4], we and others have shown that some cryptic exons can be translated to form cryptic peptides in the brain of patients with ALS[5,6]. Similarly, loss of TDP-43 mediated splicing repression also generates cryptic exons and cryptic peptides in IBM skeletal muscle[7,8].

Highly differentiated and clonally expanded $CD8^+$ T cell populations have recently been described to be enriched in ALS[9,10] and IBM[11]. These cell populations exist both in the peripheral blood mononuclear cells (PBMCs)[9] and in disease relevant tissue such as cerebrospinal fluid (CSF)[10] and skeletal muscle[11]. T cell clonal expansion in addition to the increased expression of proteins associated with activation and cytotoxicity supports an antigen-specific T cell response[12] however, no clear antigenic targets have been described, and attempts at finding them have largely been unsuccessful[13–15]. $CD8^+$ T cells can recognize novel peptides derived from non-canonical reading frame translation in cancer cells[16,17], raising the crucial question of whether $CD8^+$ T cells can mount a response against cells expressing cryptic epitopes in TDP-43 proteinopathies.

Here, we examined IBM skeletal muscle using immunohistochemistry, proteomics, and RNA sequencing to further evaluate the cryptic exon expression pattern and signature for immune activation (Fig. 1a, left). We detected the HDGFL2 cryptic peptide co-localized with TDP-43 loss along with widespread cryptic exon expression which correlated with T cell infiltration and class I antigen presentation pathways in the diseased tissue. We then utilized a high-throughput and multidimensional single T cell profiling approach, TetTCR-SeqHD[18,19], to show that cryptic epitope specific $CD8^+$ T cells isolated from ALS and IBM PBMCs are polyclonal, highly clonally expanded, highly differentiated and functional, and capable of binding multiple cryptic epitopes, including those derived from HDGFL2 and IgLON5 (Fig. 1a, right). Additionally, we showed that cells engineered to express HDGFL2 and IgLON5 cryptic epitope specific T cell receptors

(cryptic-specific TCRs) are capable of not only binding the cryptic epitopes, but also mediating T cell specific activation and effector responses. Furthermore, we demonstrated that TDP-43 knockdown (KD) in an astrocyte cell line led to HDGFL2 cryptic exon expression as well as activation and cytotoxicity of primary $CD8^+$ T cells transduced to express an HDGFL2 cryptic-specific TCR. This work highlights for the first time T cell antigenic targets in ALS and IBM, linking them to the crucial pathological hallmark of these disorders.

Cryptic peptides are present in IBM skeletal muscle and correlate with TCR and MHC-I gene expression

TDP-43 aggregation and T cell infiltration in skeletal muscle are hallmarks of IBM[11], and recent work has suggested that degenerative features such as rimmed vacuoles, TDP-43 loss of function, and cryptic exons expression are present in IBM tissue[7]. Recently, two different groups have developed antibodies that bind to the HDGFL2 cryptic peptide[6,20]. Using these antibodies, they showed that the HDGFL2 cryptic peptide can be detected in CSF and blood of presymptomatic and early-stage of ALS/FTD individuals[6] and that it is significantly increased in brain regions with TDP-43 pathology in frontotemporal lobar degeneration with TDP-43 pathology (FTLD-TDP)[20] and Alzheimer's disease with TDP-43 pathology (AD-TDP)[20], as well as in muscle biopsies from IBM cases affected by TDP-43 pathology[8], compared to non-TDP-43 controls.

To further investigate HDGFL2 cryptic peptide in IBM patient muscle, we performed immunohistochemical staining (IHC) for HDGFL2 cryptic peptide using a newly generated antibody[20], TDP-43, and p62 (an autophagy component that colocalizes with protein aggregates), in 10 IBM patients and four controls (Supplementary Table 1). We were able to visualize the specific accumulation of TDP-43 in p62-positive aggregates, with co-occurrence of TDP-43 nuclear loss in some fibers (Fig. 1b,c). Importantly, we were able to detect HDGFL2 cryptic peptide expression in 9/10 IBM cases and 0/4 controls (Fig. 1b,c), and we found immune infiltrates in samples with higher levels of TDP-43 aggregation (Fig. 1b,c). Intriguingly, HDGFL2 cryptic peptide showed a similar localisation to TDP-43 and p62, suggesting it may also aggregate in affected IBM myofibers.

We then used proteomics data to evaluate the relationship between the presence of the HDGFL2 cryptic peptide with antigen processing/presentation and T cell infiltration. Strikingly, CD3ε and

TCRβ proteins were virtually absent in controls and expressed only in a subset of IBM cases with higher expression level of the HDGFL2 cryptic peptide (Fig. 1d). Proteins in the MHC-I peptide processing and presentation pathway and HLA proteins were also significantly higher in these cases (Fig. 1d and Extended Data Fig. 1a). Differential expression and GO enrichment analysis showed a reduction of muscle-specific terms and increased antigen processing and presentation terms in the IBM cases compared to controls (Extended Data Fig. 1b,c).

As TDP-43 depletion leads to the expression of many cryptic peptides (Fig. 1e), we aimed to assess the broader relationship between TDP-43 loss of function and T cell responses in IBM. We performed RNA sequencing in TDP-43 deficient iPSC-derived myocytes to identify high-confidence muscle CE events predicted to give rise to cryptic peptides. We then investigated their presence in patient tissue using RNA sequencing data from 25 IBM muscle tissues and 15 controls (Extended Data Fig 1d,e and Extended Data Fig. 2b)[7]. We detected multiple cryptic exons in the muscles of IBM patients, while being almost undetectable in controls (Fig. 1f). *HDGFL2* cryptic exon showed the highest specificity (0/15 healthy controls) and sensitivity (17/25 IBM cases). A number of cryptic exons such as *ACTL6B* and *IgLON5*, which are found in brain tissue from people affected from ALS (Supplementary Data 2), were not identified in the IBM cohort, possibly due to the low expression of these genes in muscles (Extended Data Fig. 2c)[21]. Intriguingly, we detected the *STMN2* cryptic exon, which has neuronal specificity[22], in some IBM cases (Extended Data Fig. 2d). To address whether TDP-43 loss of function and the T cell response are correlated in diseased tissue, we used the cumulative detection of cryptic exons as a measure of TDP-43 loss of function burden, and tested its relation to the expression of antigen presentation (MHC-I) and T cell infiltrate (TCR and CD3 transcripts). We found an elevated expression of TCR and HLA genes in the IBM cases (Extended Data Fig. 2e), and we observed a strong correlation between cryptic exon burden and the expression of both TCR and HLA genes indicating a link between these events in disease (Fig. 1g).

Cryptic epitope Tetramer$^+$ T cells are enriched and clonally expanded in the PBMCs of ALS and IBM

To address whether CD8$^+$ T cells are involved in the response to cryptic peptide in ALS and IBM, we used our previously established DNA-barcoded pMHC tetramer-based high-throughput multiomic single cell profiling technology, TetTCR-SeqHD[18,19] (Fig. 1a, right). From our RNA

sequencing of IBM muscle and TDP-43 deficient iPSC-derived muscles and neurons along with previously validated cryptic exons from neuronal cells[5], we defined a complete list of cryptic peptide sequences as T cell epitope candidates (Fig. 1e). We employed a suite of pMHC binding algorithms[23] to computationally predicted a panel of 379 cryptic epitopes for CD8+ T cells, covering 18 cryptic peptides and 9 HLA-alleles, to generate DNA barcoded pMHC tetramers (Tetramer).

$CD8^+$ T cells enriched from the PBMCs from ALS (N=9) and IBM (N=4) patients, and healthy controls (N=7) were stained with a donor-specific cryptic epitope Tetramer panel, a control viral Tetramer panel, along with a panel of fluorophore labeled and DNA-barcoded phenotyping/sample-tagging antibodies. $Tetramer^+$ cells were sorted for high-throughput amplification of T cell-bound Tetramer DNA-barcodes TCR genes, as well as 558 genes and 17 surface proteins indicative of T cell activation and differentiation. There were significantly higher cryptic epitope $Tetramer^+$ (CE-$Tetramer^+$) cells in ALS and IBM patients compared to healthy controls (Fig. 2a,b). Importantly, in ALS and IBM cases, CE-$Tetramer^+$ $CD8^+$ T cells were significantly more clonally expanded (Fig. 2c), with merely 3.7% of CE-$Tetramer^+$ T cells expanded in healthy donors (>2 cells in a clonotype), compared to the 31% and 37.8% in ALS and IBM donors respectively. Furthermore, "hyperexpanded" (>15 cells with the same TCR sequences) and "large" (>10 cells with the same TCR sequences) CE-$Tetramer^+$ T cell clones were only identified in IBM and ALS donors (Fig. 2d).

CE-$Tetramer^+$ T cells are associated with activation and cytotoxicity phenotypes

To gain more information on the subtype and status of the identified CE-$Tetramer^+$ T cells, we performed weighted-nearest neighbor (WNN) analysis and subsequent UMAP visualization of our single cell gene and surface protein expression data (Fig. 2e). We observed five clusters across three major phenotype groups including: Naive-like ($CD27^+$, $CD28^+$, $CCR7^+$, $CD62L^+$, $CD45RA^+$), $GZMK^{HI}$ T effector memory (T_{EM}) ($CD27^+$, $CD28^+$, $CCR7^-$, $CD45RO^+$, $GZMK^+$), and $GNLY^{HI}$ T effector memory RA (T_{EMRA}) ($CD27^-$, $CD28^-$, $CCR7^-$, $CD45RA^+$, $GNLY^+$). CE-$Tetramer^+$ $CD8^+$ T cells from ALS and IBM patients were enriched for the $GZMK^{HI}$ T_{EM} and $GNLY^{HI}$ T_{EMRA} populations, whereas CE-$Tetramer^+$ $CD8^+$ T cells from healthy control were largely Naive-like (Fig. 2e,f). Previous single cell RNA sequencing of CSF T cells isolated from

AD and Parkinson's disease (PD) patients show a similarly high expression of related cytolytic and granzyme genes associated with a T_{EMRA} phenotype[24,25].

$CD8^+$ T cells are capable of mounting polyclonal responses to HDGLF2 and IGLON5 cryptic epitopes

We combined single cell Tetramer sequencing data with TCR sequencing data to assign epitope specificities to T cell clones. To generate high confidence assignments, T cell clones (>3 cells in a clone) were given a particular cryptic epitope specificity only when at least 60% of all cells within the TCR clonotype share the same maximum tetramer barcode signal. Under these conditions, 11 T cell clones were assigned specificities, 10/11 of which recognized cryptic epitopes derived from either HDGFL2 or IGLON5 (Fig 2g-j). These clones were found across the $GZMK^{HI}$ T_{EM} and $GNLY^{HI}$ T_{EMRA} clusters (Fig 2g).

To experimentally validate the specificity and activation ability of the above-identified cryptic epitope specific (cryptic-specific) TCRs, we engineered J76-CD8 T cells, a human $CD4^+$ T cell line engineered with a human CD8 molecule, to express the clonally expanded TCRs that we found bind to the epitopes derived from HDGFL2 and IgLON5 through our high-throughput screen.

We found that two TCRs were able to robustly bind their cognate cryptic epitope and become activated, one originally identified from an IBM patient, TCR-4, and the other from an ALS patient, TCR-11 (Fig. 3a). TCR-4, derived from donor IBM-UCL-2, was able to bind a cryptic epitope sequence from HDGFL2 cryptic peptide when presented on HLA-B*08:01. TCR-11, derived from donor ALS-UP-2, was able to bind a cryptic epitope sequence from IgLON5 cryptic peptide when presented on HLA-A*03:01. We then tested the ability of these TCRs to induce T cell activation by co-incubating our engineered J76-CD8 cryptic-specific TCR cell lines with their respective tetramer coated on a plate and measuring CD69 upregulation after 18 hours. Both cryptic-specific TCRs were able to robustly upregulate CD69, whereas stimulation with an irrelevant antigen or mismatched TCR resulted in background level expression (Fig. 3b).

To investigate whether TCR-4 and TCR-11 can bind to proteomically processed variants of the originally identified cryptic epitopes, we generated peptides that range from 7-12 aa, each of which contain some portion of the discovered epitope (Fig. 3c,d). Tetramer binding and activation were quite restrictive for TCR-4, and was largely limited to the discovered core epitope, FGKGHSGM.

Extending the peptide at the C-terminus maintained the binding, but significantly diminished CD69 expression. Losing either the phenylalanine or methionine completely abrogated binding and activation (Fig. 3c). On the other hand, TCR-11 proved to tolerate changes to the epitope. The original 11 aa epitope, SSLSAWCQLHR could be reduced to the 8 aa, SAWCQLHR, and maintain the same level of cell activation, with a moderate reduction in binding. Additionally, binding could be enhanced by extending the epitope by one amino acid at the C-terminus, though leading to slightly reduced activation (Fig. 3d).

To further validate specificity and functional capacity of cryptic epitope mediated T cell responses, we transduced an astrocytic-like cell line CCF-STTG1 that is HLA-B*08:01^{+} to constitutively express tagGFP2 (CCF-STTG1-GFP) in order to assess cell survival by monitoring the GFP signal overtime when co-culturing with human primary CD8^{+} T cells transduced to express TCR-4. After co-culturing the CCF-STTG1-GFP engineered to express either the HDGFL2.17 cryptic epitope or a scramble control with TCR-4 transduced primary CD8^{+} T cells, we observed TCR-specific target cell death as measured by GFP loss over time only in the condition where TCR-4 and HDGFL2.17 epitopes were matched (Fig 3e). These data demonstrate that if present, the HDGFL2.17 epitope can be loaded by the cell onto MHC and presented to T cells where they can become targets for killing through TCR recognition.

Cryptic epitope specific TCR engineered CD8^{+} T cells can kill TDP-43 deficient astrocytes

Given the success of TCR antigen validation using defined epitopes, we wanted to further evaluate whether primary T cells transduced with our TCRs can have a functional response in a TDP-43 proteinopathy relevant biological model. We continued to use the CCF-STTG1-GFP astrocytic-like cell line, as antigen target bearing cells, as glial dysfunction has been found to be relevant in disease pathology[26–30], and this cell line carries the HLA-B*08:01 allele which is what TCR-4 is restricted to. We used siRNA to knock down TDP-43 in CCF-STTG1-GFP and confirmed that *HDGFL2* cryptic exon is induced by TDP-43 loss in this line (Fig. 4a). This confirms that *HDGFL2* cryptic exon is consistent across cell lines and tissues upon TDP-43 loss.

The astrocytic-like cell line CCF-STTG1-GFP treatment with control siRNA resulted in no or limited cell death over time, regardless of the T cell co-culture condition (Fig. 4b, black lines), while TDP-43 siRNA treatment resulted in slightly reduced health comparatively (Fig. 4b, purple lines). Stimulated but TCR untransduced CD8^{+} T cells were able to induce some cell death (Fig.

4b, dark blue line). Strikingly, stimulated $CD8^+$ cells transduced to express TCR-4 resulted in a significantly higher amount of cell death (Fig. 4b,c). Using a shRNA targeting β2M to downregulate the expression of MHC class I molecule in CCF-STTG1-GFP line significantly reduced cell death (Fig. 4b, right panel) confirming this TCR-mediated effect and highlighting the TCR specificity toward the cryptic peptide generated by TDP-43 downregulation.

Overall, our result confirms that TDP-43 cryptic peptides could be processed, loaded to MHC class I molecules, and presented on the surface of neural cells, which is in turn recognized by cryptic-epitope specific TCRs. It also signifies that neural cells are susceptible to $CD8^+$ T cell mediated killing, and further implicates $CD8^+$ T cells in the pathogenesis of TDP-43 proteinopathies.

Discussion

ALS and IBM are both disorders where TDP-43 proteinopathy represents the predominant pathology[2]. Recently, highly differentiated and clonal $CD8^+$ T cells have been observed in ALS and IBM[9,11]. However, their antigen targets and their potential role in ALS and IBM have remained elusive. TDP-43 loss of function leads to the generation of cryptic peptides, which are foreign to the human immune system as the encoding cryptic exons do not exist in the human thymus, raising the possibility that T cells recognizing these peptides are retained. In this study, we highlight significant disease specific expression of HDGFL2 cryptic peptide in IBM skeletal muscle tissue co-localized with TDP-43 nuclear loss using IHC in addition to an increase of T cell infiltrates and MHC class I antigen presentation pathways through RNA sequencing and proteomics. We further identify clonally expanded and highly differentiated $CD8^+$ T cell clones which recognize epitopes derived from these cryptic peptides using a high-throughput and multi-dimensional integrated single T cell profiling platform we previously developed[18,19]. Finally, we show that activated $CD8^+$ T cells have the capacity to kill astrocytes with TDP-43 loss through cryptic-specific TCR recognition.

Recent work has highlighted clonal and activated T cell populations infiltrating disease-relevant tissue sites across various neurodegenerative disorders[9,10,14,15,24,25,31–37] and auto-immune diseases of the nervous system[38–42]. However, in many cases, the antigens driving the T cell responses in neurodegenerative diseases are unknown, with one of the main exceptions being PD. It has been shown that aggregates of α-synuclein, the primary pathological feature in PD, can be processed

and presented to $CD4^+$ and $CD8^+$ T cells as self-antigens[34], that the α-synuclein TCR repertoires of PD patients are highly clonal and diverse[36] and are present in early stages of disease[35,37], and associated T cells are in distinctive functional subsets[25]. However, not all aggregation-prone neuronal proteins seem to have the capacity to induce disease-specific T cell activation. T cell responses to amyloid precursor protein, amyloid beta, tau, α-synuclein, and TDP-43 were measured in AD patients, but no difference was observed compared to healthy controls[14]. Similarly, T cell responses to TDP-43 peptides in ALS patients proved to be weak, with healthy controls having moderately higher levels of activation[13], leaving a knowledge gap on whether TDP-43 pathology and T cell activation are linked.

Differently from other proteins that accumulate in neurodegenerative disease, TDP-43 mislocalisation induces the expression of multiple novel cryptic peptides which are never present in the normal state. We hypothesized these peptides can serve as neo-antigens and our data identify a T cell response against these cryptic peptides, particularly those deriving from IgLON5 and HDGFL2.

IgLON5 is a cell adhesion molecule belonging to the Ig superfamily that is primarily expressed in the brain and testis with poorly understood function[43]. The IgLON5 protein has been recently implicated as a target in an autoimmune and neurodegenerative condition that is characterized by autoantibodies against the IgLON5 protein that is concurrent with neuronal tau deposits[43]. Additionally, a case of a patient with co-existence of neuronal tau pathology along with microglia and neuronal TDP-43 pathology with autoantibodies against IgLON5 has been reported[44].

HDGFL2 is a ubiquitously expressed histone-binding protein. Cryptic exons derived from HDGFL2 are reproducibly detected via transcript analysis and proteomics in various cell lines including human iPSC-derived neurons, HeLa cells, in addition to CSF samples from patients with ALS/FTD and AD[5,6,20]. Furthermore, two groups have developed antibodies that bind to the cryptic peptide within HDGFL2[6,20] and showed that HDGFL2 cryptic peptide can be detected in the presymptomatic ALS/FTD[6] and is significantly increased in brain regions with TDP-43 pathology in FTLD-TDP and AD-TDP, compared to non-TDP-43 controls[20]. In accordance with previous work[7], we find a florid expression of cryptic exons in IBM muscle tissue, and further we show that HDGFL2 cryptic peptide is also identified in IBM tissue using one of the HDGFL2 cryptic peptide specific antibodies[20]. Intriguingly, we find the HDGFL2 cryptic peptide to co-aggregate in IBM

diseased muscle fibers raising the possibility it could further contribute to the aggregation pathogenic cascade. The consistent and widespread detection of HDGFL2 cryptic peptide in CSF, plasma, brain, and muscle supports its role as a potentially strong target for an adaptive immune response. Our findings show that it can induce T-cell mediated cell death and support a potential active role in the degenerative process.

From our single cell RNA and surface protein phenotyping data of the cryptic Tetramer^{+} T cells, we observed an enrichment for a CD8^{+} T_{EMRA} subset in both ALS and IBM conditions, which has previously been reported in ALS-4[9] and AD[24]. Within this subset, we noted several distinct subpopulations, including a GNLY^{Hi} subset expressing *GZMH*, *GZMB*, and *PRF1*. To investigate whether T cells isolated from other TDP-43 related conditions had shared features with the cryptic-epitope specific cells identified in our study, we identified publicly available single cell TCR sequencing data from the CSF and/or PBMC of AD, PD, ALS, and healthy donors across two distinct datasets[9,24] (Extended Data Fig. 10a). We identified two TCRs from two different datasets that have nearly identical TCR sequences, one isolated from the PBMCs of an individual with ALS-4 and another from the CSF of an individual with MCI (Extended Data Fig. 10b), compared to two TCRs we identified in this study. This supports the fact that cryptic-epitope specific T cells may have the potential to traffic into the CSF, and that conserved clonotypes can be identified between different individuals. Furthermore, this highlights that our dataset can be utilized as a resource to evaluate potential cryptic epitope specific T cells in already existing or future TCR sequencing datasets.

Modulating adaptive immunity may have the potential to improve clinical outcomes in TDP-43 proteinopathies. It has been shown that individuals with ALS which have a high frequency of activated regulatory T cells (T_{regs}) have better survival[10]. Similarly, a recent phase 2b clinical trial utilizing low dose IL-2 (MIROCALS, NCT03039673) has reported that for individuals with low to moderate levels of CSF phosphorylated neurofilament heavy chain (CSF-pNFH, biomarker of disease stage), there was a significant decrease in the risk of death of 48%[45]. Our findings raise the possibility that the effectiveness of this approach is mediated by dampening the response to cryptic peptides and subsequent trials will allow to specifically address this.

Other immunomodulatory clinical trials are ongoing in both IBM and ALS. A phase II/III clinical trial is evaluating the use of an anti-KLRG1 antibody (Ulviprubart) to deplete cytotoxic T cells in

IBM (NCT05721573). A Phase 2A trial, the use of tegoprubart targeting CD40-CD40L interaction has shown a decrease in inflammatory markers in people with ALS[46]. Recent trials are also evaluating T_{reg} based cell therapy products. The RAPA-501 (NCT06169176) phase II/III trial is evaluating the use of autologous hybrid T_{reg}/TH2 cells manufactured *ex vivo* in high-risk ALS patients. The CK0803 trial (NCT05695521) is evaluating the use of allogeneic, umbilical cord blood derived T_{regs} that express neurotropic homing markers. These approaches, which hope to broadly control the hyperinflammation that is associated with ALS and IBM, could be further refined with an antigen-specific approach in the future. A recent study has shown the neuroprotective value of antigen-specific T cells in a mouse model of CNS injury[47]. This result underscores the value of engineering cells toward a known antigen target. Our work highlights cryptic epitopes as critical T cell antigens and demonstrates the feasibility of engineering T cells to recognize and respond to these antigens.

Altogether, our data demonstrate a previously unknown pathological arm in TDP-43 proteinopathies where the adaptive immune system can be activated through MHC class I antigen presentation pathway, initiating from TDP-43 loss and support the rationale for immune-targeted therapies for TDP-43 proteinopathies.

Acknowledgements: We thank members of Jiang lab and Fratta lab for stimulating discussions. We thank P. Parikh and S. Salmaso for help with international shipping. We thank NIH Tetramer Core Facility at Emory University for providing MHC monomers. We thank M. Heemskerk lab from Leiden University Medical Center in the Netherlands for the J76 TCR^- Jurkat cell line and Paul Thomas lab at St. Jude Children's Research Hospital for providing human CD8 chain constructs. We are particularly grateful to Penn Center for Cellular Immunotherapies and X. Huang at Drexel University for Incucyte access. We appreciate the statistical analysis guidance

provided to us by J. Jin at the Department of Biostatistics, Epidemiology and Informatics of Penn Perelman School of Medicine. A special thank you to R. Balderas at BD Biosciences for providing BD Rhapsody scanner and some of the reagents. This study was supported by researchers at the National Institute for Health and Care Research University College London Hospitals Biomedical Research Centre.

Funding: This research was supported by

The Chan Zuckerberg Initiative Neurodegeneration Challenge Network Ben Barres Early Career Acceleration Awards 2021-005904 (NJ)

Penn Institute for Immunology and Immune Health pilot award 2022 (NJ)

The Chan Zuckerberg Initiative Collaborative Pairs Pilot Project Awards 2024-338587 (PF, NJ)

National Institute of Allergy and Infectious Diseases predoctoral training grant T32AI055428 (SC)

MZ is supported by Neuro Research Trust

UK Medical Research Council Senior Clinical Fellowship and MNDA (MR/M008606/1 and

MR/S006508/1) (PF)

NIH U54NS123743 (PF)

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services; Project number ZIAAG000535.

Author contributions:

Conceptualization: NJ, PF
Data curation: SC, MZ, NJ, PF
Formal analysis: SC, MZ, SK
Funding acquisition: NJ, PF

Investigation: SC, MZ, SK, VS, ALB, VT, LZ, IS, GMM, JMCM, DD, SB, AG, FP, PT, GM, ZL, AQ, LP, MR, GV, EB, MW, AMa, PMM, GS, PF, NJ

Methodology: SC, MZ, NJ, PF

Project administration: SC, MZ, NJ, PF

Resources: AQ, LP, MH, MR, GV, EB, MW, AMa, PMM, GS, PF, NJ

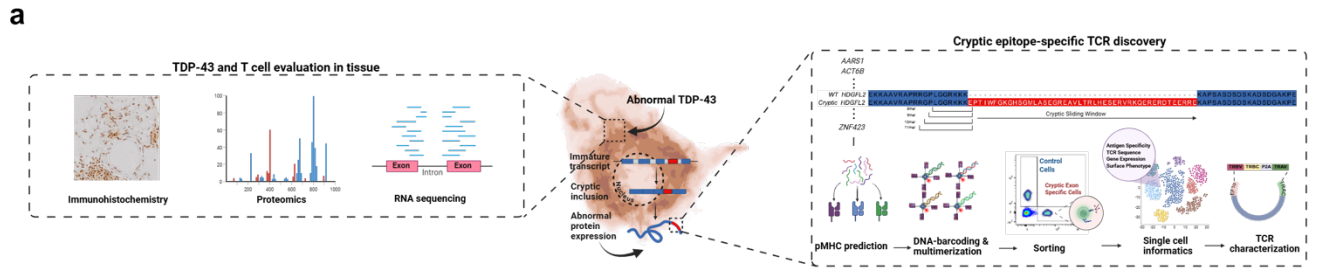
Software: SC, MZ, SK

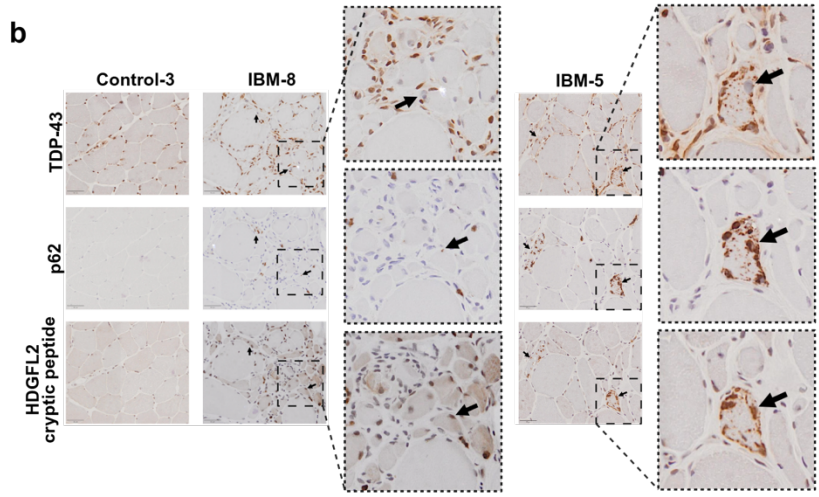
Supervision: NJ, PF

Writing: SC, MZ, NJ, PF

Writing, review and editing: all authors

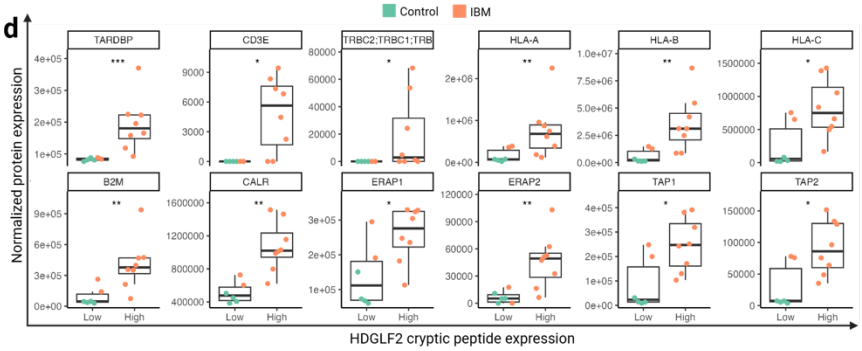
Competing interests: The University of Pennsylvania has applied for patents on the antigens and TCRs.

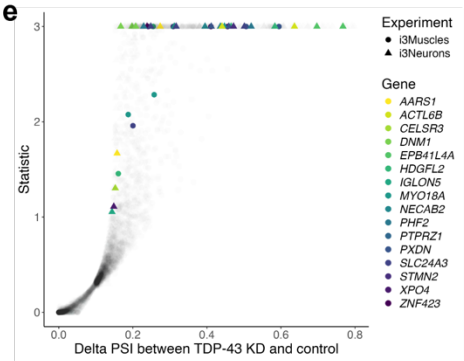


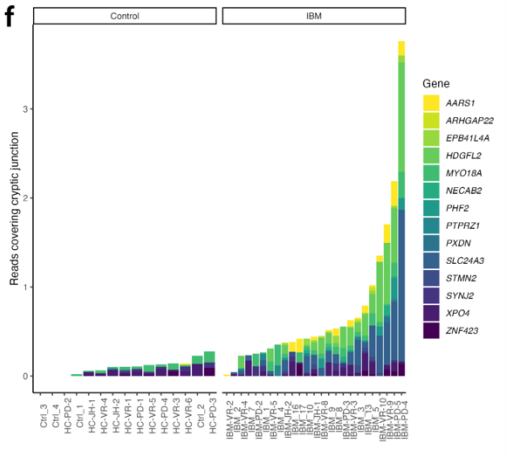


c

Case	Immune infiltrate	TDP-43 loss	p62	HDGFL2-CE
Control-1	-	NA	-	-
Control-2	-	-	-	-
Control-3	-	-	-	-
Control-4	-	-	-	-
IBM-17	-	-	-	-
IBM-16	+	+	+	+
IBM-2	+	++	++	++
IBM-5	+	++	++	++
IBM-7	++	++	++	++
IBM-9	++	NA	++	++
IBM-10	++	+++	++	++
IBM-8	++	+++	++	+++
IBM-3	++	+++	+++	+++
IBM-13	++	+++	+++	+++







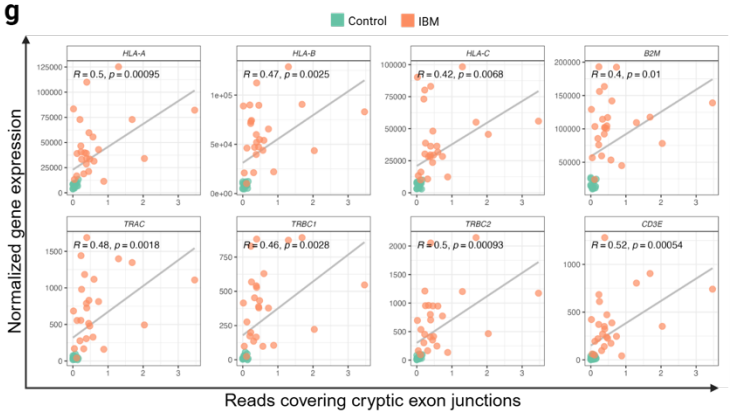


Fig. 1. Cryptic peptides and MHC I / TCR pathways are found in IBM tissues.

(a) Schematic of the experimental workflow. IHC, proteomics, and RNA sequencing were used to analyze the expression of cryptic peptide, TCR, and MHC expression and T cell infiltration in IBM muscle tissues (left). Validated cryptic peptides were used to informatically predict epitopes binding to patient specific HLA molecules. Then, fluorescence labeled and DNA-barcoded pMHC tetramers were generated and used to sort cryptic epitope Tetramer$^+$ CD8$^+$ T cells for downstream single cell analysis and *in vitro* characterization (right). **(b)** Visualization of seriate immunohistochemistry staining for TDP-43, p62, and HDGFL2 cryptic peptide in muscle from one control and two IBM cases. For each case, arrows show colocalisation of HDGFL2 cryptic peptide in fibers with p62 and TDP-43 nuclear depletion and/or aggregation. **(c)** Semi-quantitative scores for immune infiltrates, TDP-43 loss, p62 and HDGFL2 cryptic peptide in control and IBM cases. **(d)** Normalized protein expression as quantified by mass spectrometry. The same group of four controls and 10 IBM cases were grouped based on the expression level of HDGFL2 cryptic peptide quantified by antibody in (c). Box plots show median and all points are shown. **(e)** Volcano plot showing upregulated splicing junctions upon TDP-43 knockdown in iPSC-derived skeletal muscles (circles) and iPSC-derived glutamatergic cortical neurons (triangles). **(f)** Normalized counts of RNA-seq reads covering cryptic exon junctions in controls (n = 15) and IBM cases (n = 25). **(g)** Correlation between RNA-seq normalized read counts for TCR α and β constant chain genes (*TRAC*, *TRBC1*, and *TRBC2*), *CD3E*, *HLA-A*, *HLA-B*, *HLA-C*, and β2-microglobulin (*B2M*) and cryptic exon normalized counts in controls (n = 15) and IBM cases (n = 25).

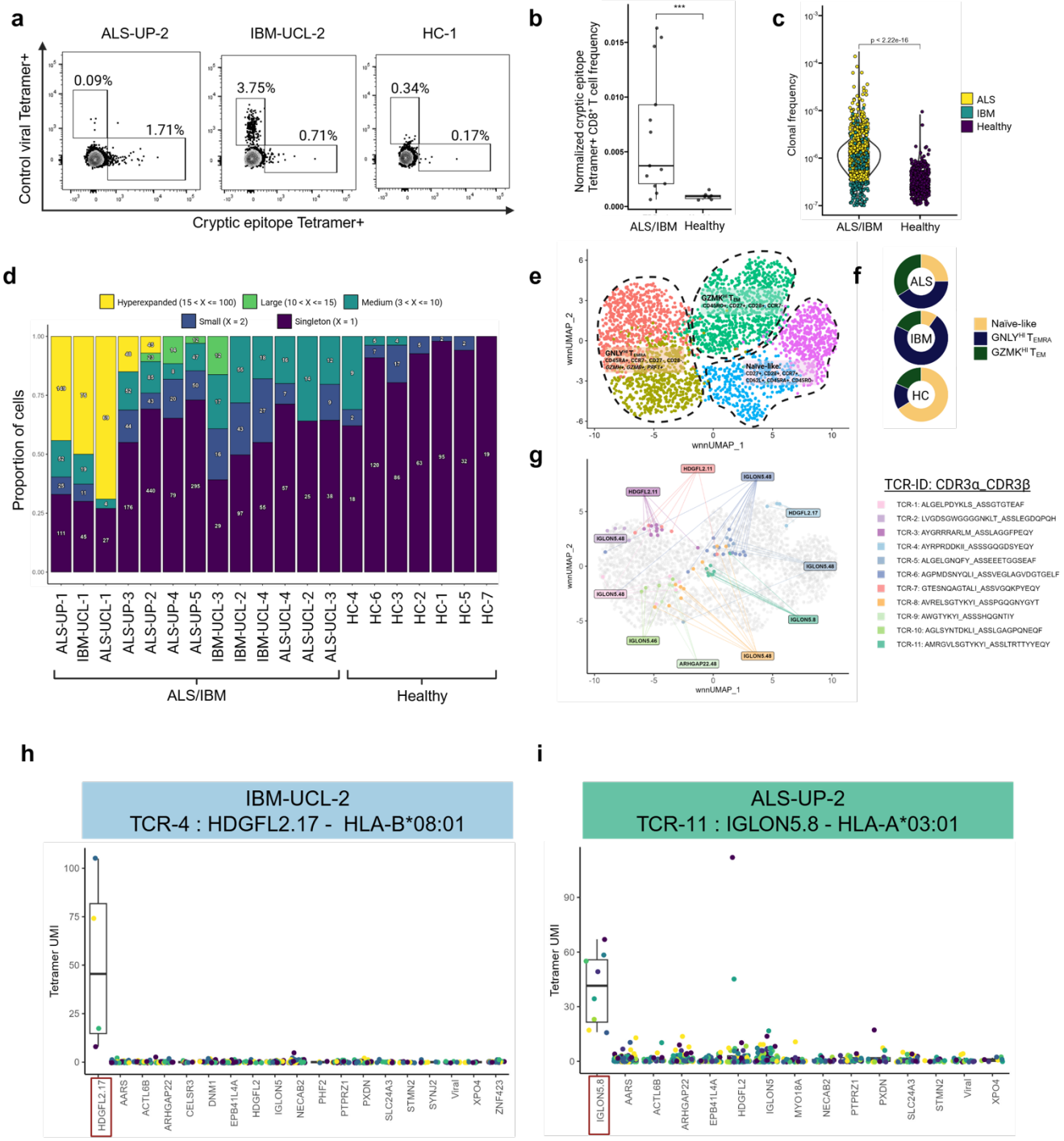

a
ALS-UP-2
IBM-UCL-2
HC-1
Control viral Tetramer+
Cryptic epitope Tetramer+
0.09%
1.71%
3.75%
0.71%
0.34%
0.17%
b
Normalized cryptic epitope Tetramer+ CD8+ T cell frequency
***
ALS/IBM
Healthy
c
Clonal frequency
p < 2.22e-16
ALS
IBM
Healthy
d
Hyperexpanded (15 < X <= 100)
Large (10 < X <= 15)
Medium (3 < X <= 10)
Small (X = 2)
Singleton (X = 1)
Proportion of cells
ALS-UP-1
IBM-UCL-1
ALS-UCL-1
ALS-UP-3
ALS-UP-2
ALS-UP-4
ALS-UP-5
IBM-UCL-3
IBM-UCL-2
IBM-UCL-4
ALS-UCL-4
ALS-UCL-2
ALS-UCL-3
HC-4
HC-6
HC-3
HC-2
HC-1
HC-5
HC-7
ALS/IBM
Healthy
e
wnnUMAP_1
wnnUMAP_2
f
ALS
IBM
HC
Naïve-like
GNLY^Hi T_EMRA
GZMK^Hi T_EM
g
TCR-ID: CDR3α_CDR3β
TCR-1: ALGELPDYKLS_ASSGTGTEAF
TCR-2: LVGDSGWGGGGNKLT_ASSLEGDQPQH
TCR-3: AYGRRRARLM_ASSLAGGFPEQY
TCR-4: AYRPRDDKII_ASSSGQGDSYEQY
TCR-5: ALGELGNQFY_ASSEEETGGSEAF
TCR-6: AGPMDSNYQLI_ASSVEGLAGVDGTGELF
TCR-7: GTESNQAGTALI_ASSVGQKPYEQY
TCR-8: AVRELSGTYKYI_ASSPGQGNYGYT
TCR-9: AWGTYKYI_ASSSHQGNTIY
TCR-10: AGLSYNTDKLI_ASSLGAGPQNEQF
TCR-11: AMRGVLSGTYKYI_ASSLTRTTYYEQY
h
IBM-UCL-2
TCR-4 : HDGFL2.17 - HLA-B*08:01
Tetramer UMI
HDGFL2.17
AARS
ACTL6B
ARHGAP22
CELSR3
DNM1
EPB41L4A
HDGFL2
IGLON5
NECAB2
PHF2
PTPRZ1
PXDN
SLC24A3
STMN2
SYNJ2
Viral
XPO4
ZNF423
i
ALS-UP-2
TCR-11 : IGLON5.8 - HLA-A*03:01
Tetramer UMI
IGLON5.8
AARS
ACTL6B
ARHGAP22
EPB41L4A
HDGFL2
IGLON5
MYO18A
NECAB2
PTPRZ1
PXDN
SLC24A3
STMN2
Viral
XPO4
j
EKKAAVRAPRRGPLGGRKKKEPTIWFGKGHSGMLASEGREAVLTRLHESERVRKQERERDTEERREKAPSASDSDSKADSDGAKPE
HDGFL2.17
TCR-4
HDGFL2.11
TCR-3
TCR-7
ARLRLLAAAALAGLAVISRGQPRAGPSHSRSPRSSSRPSSLSAWCQLHRLRMPSPHSLALVQPAGHGLQRDDCVGVTEPACCAEVCEWAWAHTCKHALANLSTQCACVDVCVCKGLLSQSLEFNSPADNYTVCEG
IGLON5.8
TCR-11
IGLON5.46
TCR-10
TCR-6
TCR-8
TCR-1
TCR-5
TCR-2
IGLON5.48
Clonal Response
≥ 20
10 - 20
5 - 10
< 5

Fig. 2. TDP-43 dysfunction leads to a heterogenous and polyclonal CD8$^+$ T cell response to cryptic epitopes.

(a) Representative FACS plot depicting cryptic epitope Tetramer$^+$ cells (x-axis) and control viral Tetramer$^+$ cells (y-axis) across three donors. **(b)** Summary of the complete data represented in (A). Cryptic epitope Tetramer$^+$ CD8$^+$ T cell frequency was normalized by the number of total pMHC Tetramers for each individual donor. Wilcoxon Rank Sum test was used to evaluate significance. **(c)** Violin plot of clone size normalized by starting number of CD8$^+$ T cells across ALS, IBM, and healthy donors. The Wilcoxon Rank Sum test was used to evaluate significance. **(d)** T cell clonality per donor ordered by greatest to least clonality. Clones were stratified by their CDR3β according to the following bins: Hyperexpanded ($15 < X <= 100$), Large ($10 < X <= 15$), Medium ($3 < X <= 10$), Small ($X = 2$), Singleton ($X = 1$). **(e)** UMAP representation of single cells by a weighted nearest neighbor calculation of gene and surface protein expression. Clusters were manually annotated based on their differential gene and protein expression. Genes are italicized. **(f)** Donut graph of phenotypic composition across ALS, IBM, and healthy controls. **(g)** Tetramer specificity assignments on T cell clones overlaid on UMAP. Clonal specificities were assigned such that at least 60% of the cells in a given clonotype must share the same epitope assignment based on tetramer barcode sequencing. Cells that do not meet that threshold are colored grey. **(h/i)** Box plot visualization of tetramer barcode UMI counts for TCR-4 and TCR-11 with assigned specificities to HDGFL2.17 (left) and IgLON5.8 (right) epitopes. Each dot depicts an individual cell's signal for each tetramer barcode (x-axis) within the specified clone. All tetramer barcodes were grouped into their parent cryptic protein except for the top hit (represented by the red box). All points are shown and center line represents median **(j)** Depiction of identified cryptic epitopes across HDGFL2 and IgLON5. Epitopes found to be targeted are marked by a bar which spans the epitope sequence. The tick marks within each bar represent different T cell clones targeting the same epitope and the distance within each tick represents the relative clone size. The color intensity represents the sum of all T cell clones targeting the particular epitope.

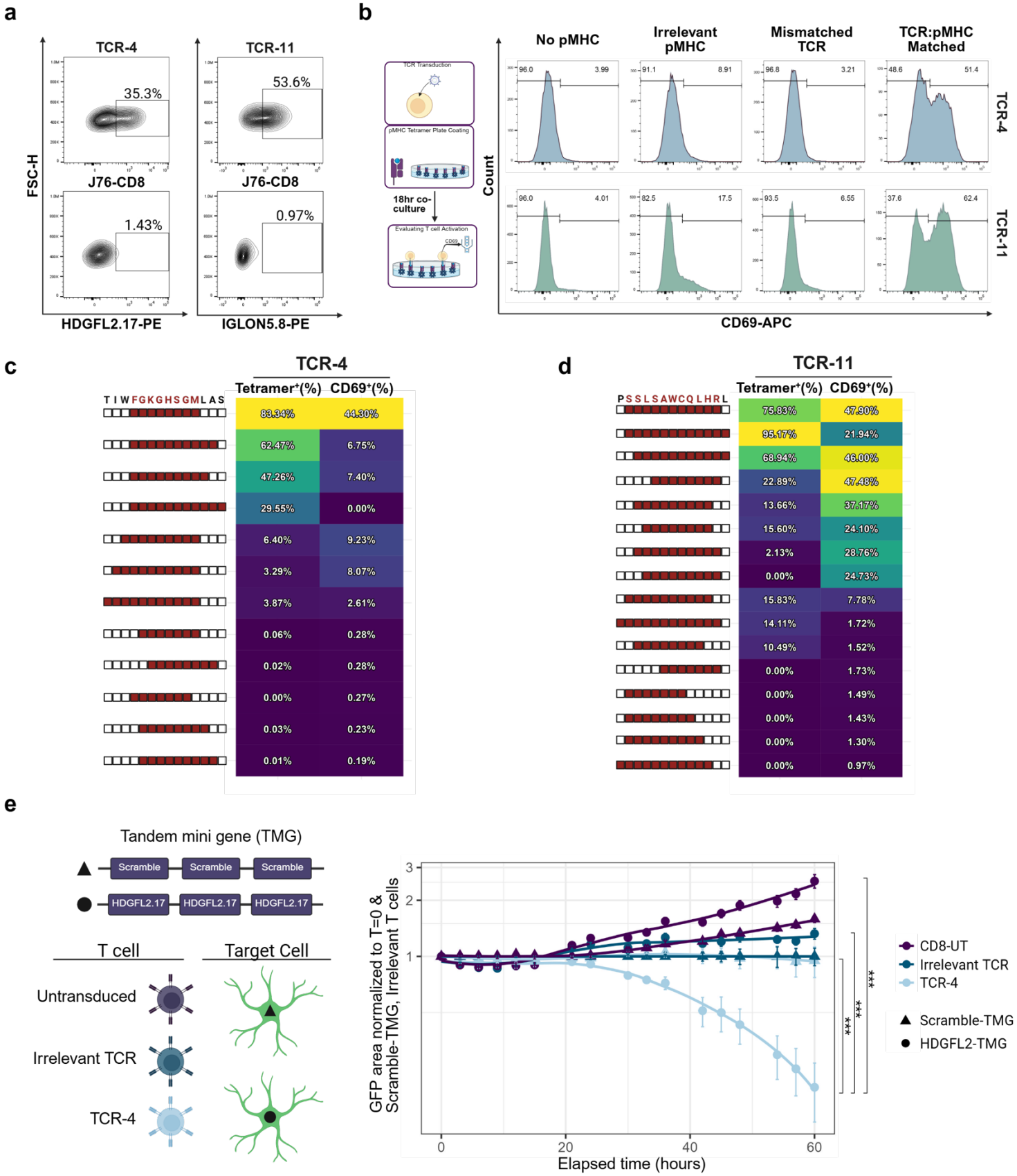

a
TCR-4
35.3%
TCR-11
53.6%
J76-CD8
1.43%
J76-CD8
0.97%
FSC-H
HDGFL2.17-PE
IGLON5.8-PE
b
TCR Transduction
pMHC Tetramer Plate Coating
18hr co-culture
Evaluating T cell Activation
CD69
No pMHC
Irrelevant pMHC
Mismatched TCR
TCR:pMHC Matched
96.0
3.99
91.1
8.91
96.8
3.21
48.6
51.4
TCR-4
96.0
4.01
82.5
17.5
93.5
6.55
37.6
62.4
TCR-11
Count
CD69-APC
c
TCR-4
Tetramer+(%) CD69+(%)
T I W F G K G H S G M L A S
83.34% 44.30%
62.47% 6.75%
47.26% 7.40%
29.55% 0.00%
6.40% 9.23%
3.29% 8.07%
3.87% 2.61%
0.06% 0.28%
0.02% 0.28%
0.00% 0.27%
0.03% 0.23%
0.01% 0.19%
d
TCR-11
Tetramer+(%) CD69+(%)
P S S L S A W C Q L H R L
75.83% 47.90%
95.17% 21.94%
68.94% 46.00%
22.89% 47.48%
13.66% 37.17%
15.60% 24.10%
2.13% 28.76%
0.00% 24.73%
15.83% 7.78%
14.11% 1.72%
10.49% 1.52%
0.00% 1.73%
0.00% 1.49%
0.00% 1.43%
0.00% 1.30%
0.00% 0.97%
e
Tandem mini gene (TMG)
Scramble
Scramble
Scramble
HDGFL2.17
HDGFL2.17
HDGFL2.17
T cell
Target Cell
Untransduced
Irrelevant TCR
TCR-4
GFP area normalized to T=0 & Scramble-TMG, Irrelevant T cells
Elapsed time (hours)
CD8-UT
Irrelevant TCR
TCR-4
Scramble-TMG
HDGFL2-TMG
***
***
***

Fig. 3. T cells with engineered cryptic epitope specific TCRs can bind and activate in response to cryptic epitopes.

(a) pMHC tetramer binding of J76-CD8 T cell expressing TCR-4 targeting the HDGFL2 epitope FGKGHSGM, and TCR-11 targeting the IgLON5 epitope SSLSAWCQLHR. **(b)** Flow cytometry histogram plots of CD69 upregulation following 18 hour incubation in a cryptic epitope tetramer coated plate across the following conditions: no pMHC coated on the plate, CMV pp65 (irrelevant) pMHC coated on the plate, irrelevant TCR co-cultured with the matched cryptic epitope pMHC, matched TCR co-cultured with the matched cryptic epitope pMHC. Results for TCR-4 (top row) and TCR-11 (bottom row) are shown. **(c/d)** Evaluation of pMHC tetramer binding and activation of J76-CD8 T cell expressing TCR-4 (c) and TCR-11 (d) across various possible epitope variations of identified cognate cryptic epitopes. The values shown on the heatmaps are background subtracted based on an irrelevant CMV pp65-specific TCR. **(e)** Constructs carrying HDGFL2.17 epitope or scramble epitope were transduced into CCF-STTG1-GFP astrocytes and co-incubated with primary $CD8^+$ T cells transduced to express TCR-4 or irrelevant TCR or untransduced primary $CD8^+$ T cells. GFP area over time was detected as a measure of cell death over time. GFP signals were normalized to the scramble construct with the irrelevant T cell condition. The Kolmogorov–Smirnov (two-sided) test was performed on possible differences in the means of the curves. The standard error for each normalized value is shown across four technical replicates.

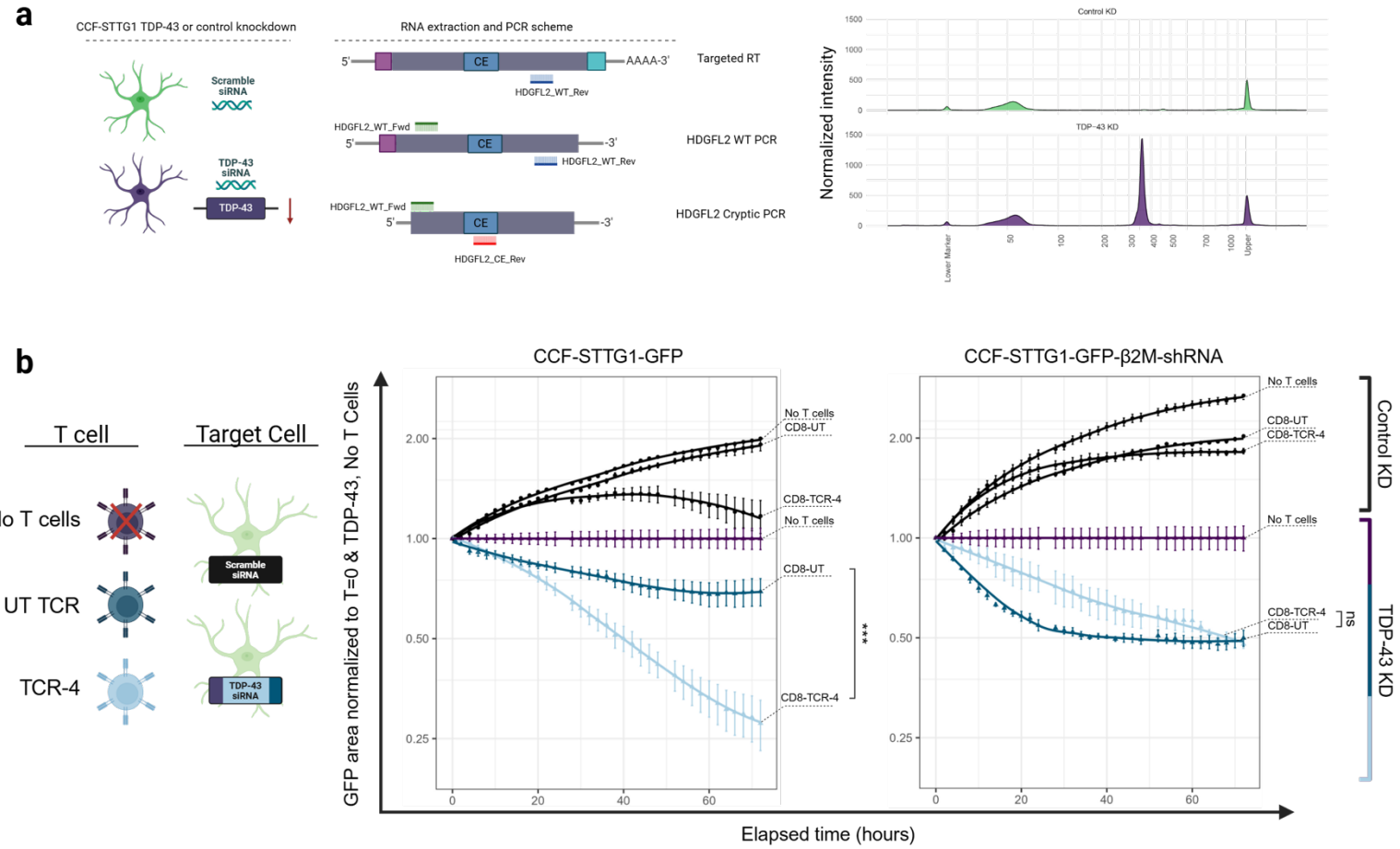
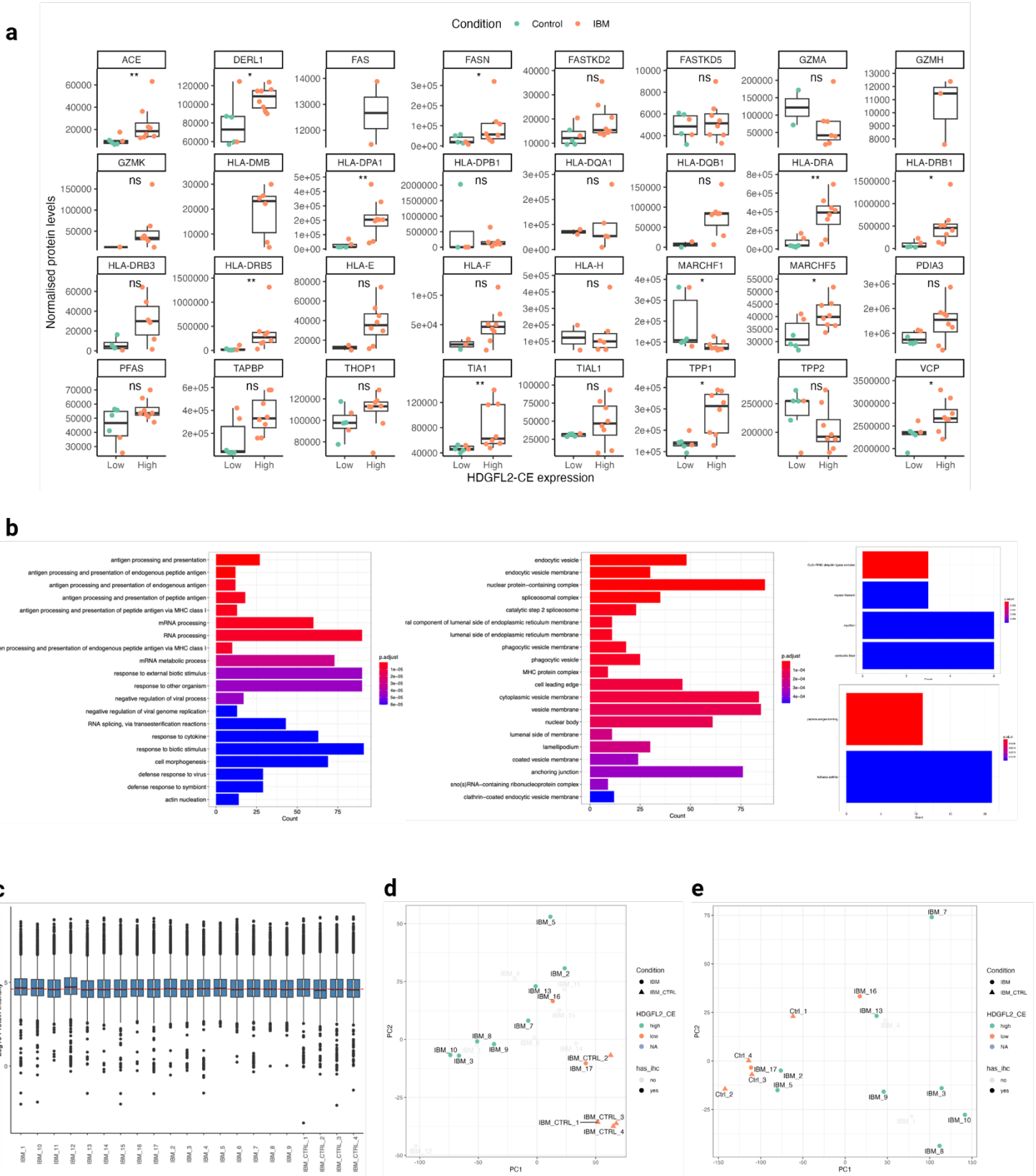

a
CCF-STTG1 TDP-43 or control knockdown
Scramble siRNA
TDP-43 siRNA
TDP-43
RNA extraction and PCR scheme
5'
CE
AAAA-3'
Targeted RT
HDGFL2_WT_Rev
HDGFL2_WT_Fwd
-3'
HDGFL2 WT PCR
HDGFL2 Cryptic PCR
HDGFL2_CE_Rev
Control KD
TDP-43 KD
Normalized intensity
Lower Marker
Upper
b
T cell
Target Cell
No T cells
UT TCR
TCR-4
Scramble siRNA
TDP-43 siRNA
CCF-STTG1-GFP
CCF-STTG1-GFP-β2M-shRNA
GFP area normalized to T=0 & TDP-43, No T Cells
Elapsed time (hours)
No T cells
CD8-UT
CD8-TCR-4
***
ns
Control KD
TDP-43 KD
c
TDP-43 KD
CD8-TCR-4
CTRL KD
CD8-TCR-11
0d 0h 0m
0d 3h 0m
0d 6h 0m
0d 15h 0m
0d 18h 0m
1d 21h 0m

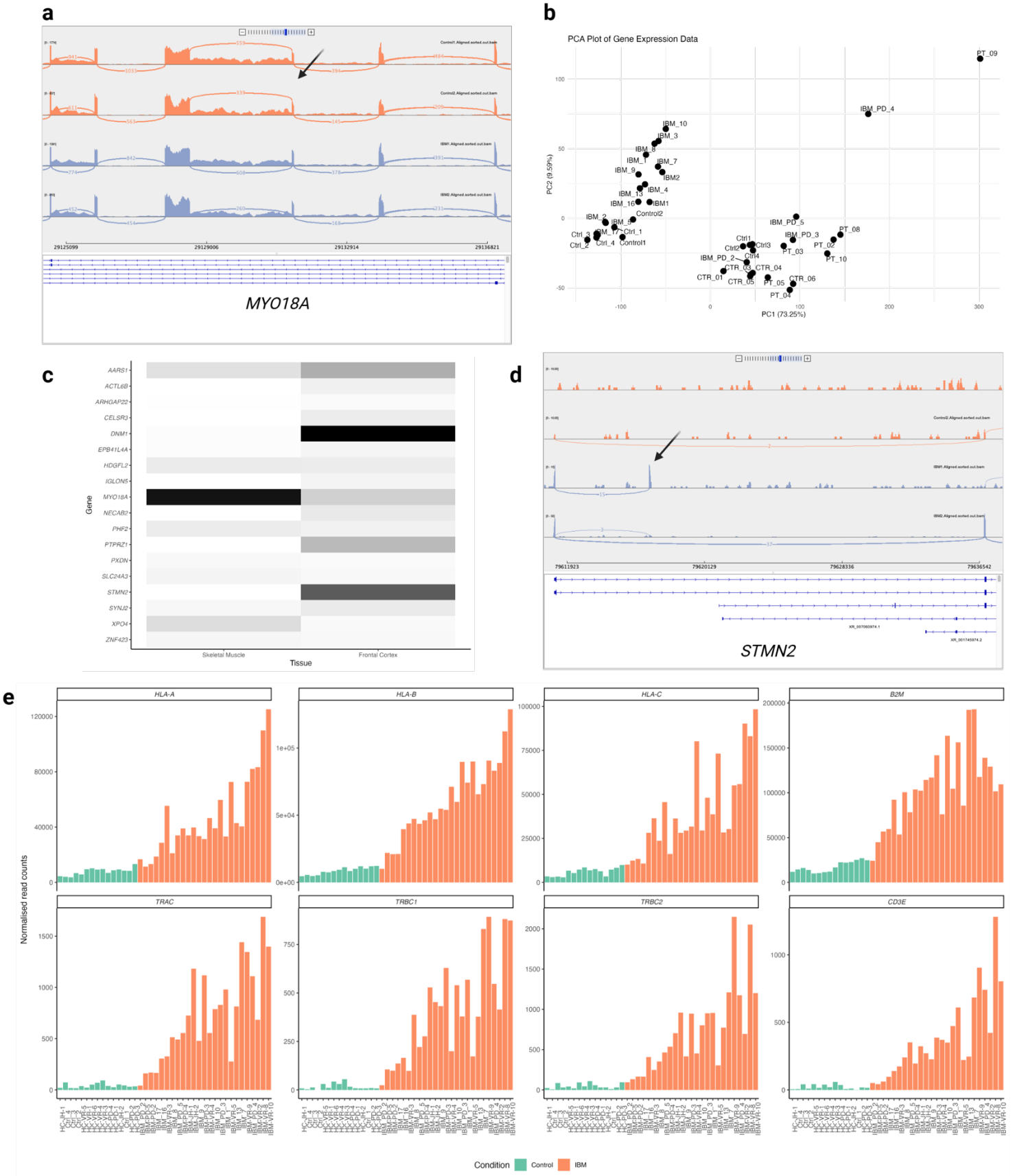
Fig. 4. HDGFL2 cryptic epitope specific CD8$^+$ T cells can efficiently kill TDP-43 deficient astrocytes.

(a) siRNA knockdown and HDGFL2 cryptic exon PCR strategy diagram (left) and Tapestation trace of final HDGFL2 cryptic exon PCR product across control and TDP-43 knockdown (right). **(b)** Cytotoxicity of HDGFL2 cryptic epitope specific CD8$^+$ T cells toward astrocyte-like cell line, CCF-STTG1-GFP, and its HLA-negative version, CCF-STTG1-GFP-β2M-shRNA, across time as measured by GFP expression. Three T cell conditions across control and siRNA TDP-43 knockdown were measured. CD8-TCR-4: human primary activated CD8$^+$ T cells transduced to express TCR-4 and enriched via flow cytometry to be 100% tetramer positive using HDGFL2.17 pMHC tetramer; CD8-UT: primary activated and TCR untransduced CD8$^+$ T cells from matching donor; No T cell: no T cells were added to the co-culture. GFP signals were normalized to the No T cell condition. The Kolmogorov–Smirnov (two-sided) test was performed on possible differences in the means of the curves between CD8-TCR-4 and CD8-UT in both panels. ***p = 8.305e-07. ns, p = 0.5041. **(c)** representative images of the CCF-STTG1-GFP:CD8-TCR-4 co-culture across TDP-43 (top) and control (bottom) siRNA knockdown. The red mask marks cells undergoing cell death as indicated by caspase 3/7 dye. The standard error for each normalized value is shown across three technical replicates.


a
Condition
Control
IBM
ACE
DERL1
FAS
FASN
FASTKD2
FASTKD5
GZMA
GZMH
GZMK
HLA-DMB
HLA-DPA1
HLA-DPB1
HLA-DQA1
HLA-DQB1
HLA-DRA
HLA-DRB1
HLA-DRB3
HLA-DRB5
HLA-E
HLA-F
HLA-H
MARCHF1
MARCHF5
PDIA3
PFAS
TAPBP
THOP1
TIA1
TIAL1
TPP1
TPP2
VCP
Normalised protein levels
HDGFL2-CE expression
Low
High
b
antigen processing and presentation
antigen processing and presentation of endogenous peptide antigen
antigen processing and presentation of endogenous antigen
antigen processing and presentation of peptide antigen
antigen processing and presentation of peptide antigen via MHC class I
mRNA processing
RNA processing
mRNA metabolic process
response to external biotic stimulus
response to other organism
negative regulation of viral process
negative regulation of viral genome replication
RNA splicing, via transesterification reactions
response to cytokine
response to biotic stimulus
cell morphogenesis
defense response to virus
defense response to symbiont
actin nucleation
Count
p.adjust
endocytic vesicle
endocytic vesicle membrane
nuclear protein−containing complex
spliceosomal complex
catalytic step 2 spliceosome
lumenal side of endoplasmic reticulum membrane
phagocytic vesicle membrane
phagocytic vesicle
MHC protein complex
cell leading edge
cytoplasmic vesicle membrane
vesicle membrane
nuclear body
lumenal side of membrane
lamellipodium
coated vesicle membrane
anchoring junction
sno(s)RNA−containing ribonucleoprotein complex
clathrin−coated endocytic vesicle membrane
c
Log10 Protein Intensity
d
PC1
PC2
Condition
IBM
IBM_CTRL
HDGFL2_CE
high
low
NA
has_ihc
no
yes
e

Extended Data Fig 1. Additional data from the tissue proteomics.

(a) MHC-I and -II antigen processing pathway expression in controls and IBM cases, divided by condition and coloured based on the upper or lower 50% by TDP43 levels. **(b)** Normalized count distribution for the proteomics cohort. **(c)** GO analysis summary from the proteomics analysis. **(d)** PCA plot from normalised protein expression from the proteomics experiment. **(e)** PCA for the NHNN RNA-seq cohort. Note how the PCA from proteomics and that from RNA are fairly overlapping.


a
MYO18A
b
PCA Plot of Gene Expression Data
PC2 (9.59%)
PC1 (73.25%)
c
Gene
AARS1
ACTL6B
ARHGAP22
CELSR3
DNM1
EPB41L4A
HDGFL2
IGLON5
MYO18A
NECAB2
PHF2
PTPRZ1
PXDN
SLC24A3
STMN2
SYNJ2
XPO4
ZNF423
Skeletal Muscle
Frontal Cortex
Tissue
d
STMN2
e
HLA-A
HLA-B
HLA-C
B2M
TRAC
TRBC1
TRBC2
CD3E
Normalised read counts
Condition
Control
IBM

Extended Data Fig 2. Additional data from RNA-seq analysis.

(a). IGV tracks for MYO18A cryptic exon in a subset of controls and IBM muscle biopsy RNA-seq. **(b)** PCA for the full RNA-seq cohort. The clear split between two groups in PC1 is likely driven by different library preparation methods. **(c)** Heatmap showing gene expression of genes with cryptic peptides in skeletal muscle and frontal cortex. Expression data obtained from ASCOT[21] **(d)** IGV tracks for STMN2 cryptic exon in a subset of controls and IBM muscle biopsy RNA-seq. **(e)** RNA-seq normalized read counts for TCR constant chains (*TRAC*, *TRBC1*, and *TRBC2*), *CD3E*, *HLA-A*, *HLA-B*, *HLA-C*, and beta2-microglobulin (*B2M*) in control and IBM cases.

Supplementary Table 1. Tissue analysis cohort.

Sample Name	Experiment	Disease	Cohort	Alias
IBM-2	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-2
IBM-3	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-3
IBM-5	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-5
IBM-7	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-7
IBM-8	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-8
IBM-9	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-9
IBM-10	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-10
IBM-13	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-13
IBM-16	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-16
IBM-17	Proteomics/IHC/RNA-seq	IBM	UCL_proteomics	IBM-17
Ctrl-1	Proteomics/IHC/RNA-seq	HC	UCL_proteomics	Ctrl-1
Ctrl-2	Proteomics/IHC/RNA-seq	HC	UCL_proteomics	Ctrl-2
Ctrl-3	Proteomics/IHC/RNA-seq	HC	UCL_proteomics	Ctrl-3
Ctrl-4	Proteomics/IHC/RNA-seq	HC	UCL_proteomics	Ctrl-4
IBM-1	Proteomics/RNA-seq	IBM	UCL_proteomics	IBM-1
IBM-4	Proteomics/RNA-seq	IBM	UCL_proteomics	IBM-4
IBM-6	Proteomics	IBM	UCL_proteomics	IBM-6
IBM-11	Proteomics	IBM	UCL_proteomics	IBM-11
IBM-12	Proteomics	IBM	UCL_proteomics	IBM-12
IBM-14	Proteomics	IBM	UCL_proteomics	IBM-14
IBM-15	Proteomics	IBM	UCL_proteomics	IBM-15
Control1	RNA-seq	HC	JH_RNA	HC-JH-1

Control2	RNA-seq	HC	JH_RNA	HC-JH-2
IBM1	RNA-seq	IBM	JH_RNA	IBM-JH-1
IBM2	RNA-seq	IBM	JH_RNA	IBM-JH-2
Ctrl1	RNA-seq	HC	UP_RNA	HC-UP-1
Ctrl2	RNA-seq	HC	UP_RNA	HC-UP-2
Ctrl3	RNA-seq	HC	UP_RNA	HC-UP-3
Ctrl4	RNA-seq	HC	UP_RNA	HC-UP-4
IBM_PD_2	RNA-seq	IBM	UP_RNA	IBM-UP-2
IBM_PD_3	RNA-seq	IBM	UP_RNA	IBM-UP-3
IBM_PD_4	RNA-seq	IBM	UP_RNA	IBM-UP-4
IBM_PD_5	RNA-seq	IBM	UP_RNA	IBM-UP-5
CTR_01	RNA-seq	HC	TS_RNA	HC-TS-1
CTR_03	RNA-seq	HC	TS_RNA	HC-TS-3
CTR_04	RNA-seq	HC	TS_RNA	HC-TS-4
CTR_05	RNA-seq	HC	TS_RNA	HC-TS-5
CTR_06	RNA-seq	HC	TS_RNA	HC-TS-6
PT_02	RNA-seq	IBM	TS_RNA	IBM-TS-2
PT_03	RNA-seq	IBM	TS_RNA	IBM-TS-3
PT_04	RNA-seq	IBM	TS_RNA	IBM-TS-4
PT_05	RNA-seq	IBM	TS_RNA	IBM-TS-5
PT_08	RNA-seq	IBM	TS_RNA	IBM-TS-8
PT_09	RNA-seq	IBM	TS_RNA	IBM-TS-9
PT_10	RNA-seq	IBM	TS_RNA	IBM-TS-10

Methods

Ethics for patient sample collection, MTA, clinical-demographics

For the TCR analysis, we collected PBMCs from n = 4 ALS cases from UCL, n = 5 ALS cases from University of Padua, n = 4 IBM cases from UCL, and n = 7 healthy controls from the human immunology core at the University of Pennsylvania. All cases gave written informed consent in accordance with the principles outlined in the Declaration of Helsinki.

iMuscle CRISPRi cell line development from human iPSCs

The WTC11 human induced pluripotent stem cell (iPSC) line was used as the parent cell line. The cultures were kept at 37.5°C, 5% CO2 atmosphere, and approximately 90-95% humidity. Cultures were grown using Essential-8™ media with supplement (Life Technologies; #A1517001) and rock 2 inhibitor Chroman-1 (MedChem Express; #HY-15392) was used to aid in single cell survival during passage. Single cells were seeded at 0.033 M/cm^2 into 5% Matrigel-coated plates (Corning® Matrigel hESC-Qualified, LDEV free; #354277) A 5% Matrigel coating solution was prepared using cold KnockOut™ DMEM/F-12 (Gibco; #12660012) and chilled pipette tips[49]. The coated vessels were incubated at 37.5°C, 5% CO2 for 1 hour or overnight. Cultures were passed before colony overlap, or when approximately 75% confluent. Routine morphology assessments were performed. Stable cell lines with transcription factors under the conditional Tet-On 3G promoter were engineered to rapidly differentiate iPSCs into disease relevant cell types. The Zim3 dCas9 CRISPR inhibition system was delivered via PiggyBac flanked between 5' and 3' ITRs co-delivered with the EF1-α Transposase at 2:1 ratio (3 µg total DNA per 1.5 M iPSCs and 5 µLs of Lipofectamine STEM (Thermofisher; STEM00001).The dCas9 protein was cloned under the chicken beta-actin promoter with co-expression of BFP-NLS which was observed after overnight incubation. The polyclonal population was expanded and sorted using the SH800S Cell Sorter (Sony Biotechnology; Model LE-SH800SZFCPL; San Jose, CA) with a sterile sorting chip and a 100 µM nozzle size. A single cell suspension was obtained followed by dilution, cell counting, and resuspension on Essential-8™ media with 50 nM Chroman-1 and Penicillin-Streptomycin (Gibco; #15070063), to prevent contamination after sorting. The iPSCs were kept at 4°C while in suspension and during sorting. Positive (mApple or BFP) and negative controls (non-transfected iPSCs) were used to set the forward (FSC) and side scatter (SSC) gates, followed by single cell gating, and fluorescent marker gating to identify positive populations for each marker. The top

10% double positive (mApple + BFP) population was gated and sorted into 12 well plates coated with Matrigel and Essential-8™ with Chroman-1, selecting cells within the 104-105 intensity. The pre-sorted population was enriched from 5.2% to 96.9% double positive. After sorting, iPSCs were returned to the incubator and allowed to rest for 3-4 hours or overnight before refreshing media. The rock inhibitor was kept until the colonies reached 4-6 cells, passaged 2-3 times before removing the antibiotics, and tested negative for mycoplasma before moving out of quarantine. To rapidly generate skeletal muscle, we subcloned shRNA targeting the miRNA sequence for Oct4 (GTCCGAGTGTGGTTCTGTA) downstream of MyoD1, as described previously[50]. To enable CRISPR inhibition, the Zim3-dCas9 machinery was delivered via PiggyBac followed by hygromycin selection (1.50 mg/mL for 4-7 days).

Modeling loss of function of TDP-43 via CRISPR interference

The loss of function system was engineered by combining the Zim3 repressor, dCas9 system, and dual sgRNA delivered via a lentiviral transduction. The dual guides were subcloned by creating inserts containing sgRNAs between a tRNA spacer[51] to deliver 2 non-targeting guides (control) or 2 guides targeting TDP-43. The sequences were either ordered as a GBlock from IDT or amplified by PCR, using primers with sequences related to the desired sgRNA. For the insert, non-targeting sgRNA-1 (gaggcaaccgctctagcgcg) or non-targeting sgRNA-2 (ggttgaaccgcccccgccga) were subcloned under the mouse U6 (mU6) promoter or the human U6 (hU6) promoter respectively. For efficient knockdown of TDP-43, sgRNA-1 (gggaagtcagccgtgagacc) and sgRNA-2 (gaggactaggcccggtctca) were subcloned downstream of the hU6 promoter or the mU6 promoter, respectively. The hygromycin selection cassette followed by eGFP-NLS under constitutive expression (EF1-α) was used as a selection marker, when appropriate. The lentiviral backbone was modified to increase viral titer by subcloning a WPRE upstream of the 3' LTR.

Lentiviral Packaging and sgRNA Delivery via Lentiviral Transduction

The sgRNAs for TDP-43 knockdown and non-targeting controls were packaged into lentiviral particles using in-house lentiviral vectors transfected into HEK cells as described previously[52]. After transfection, viral particles were concentrated using Lenti-X™ Concentrator (TakaraBio; #631232) at a 1:10 ratio in dPBS and aliquoted for storage at -80°C until the start of the experiments. The control and TDP-43 sgRNAs were qualitatively evaluated for viral titer by confocal imaging. Briefly, Zim3-dCas9 iPSCs were seeded in Matrigel-coated 96 well plates at

0.04 M/cm^2 in half medium volume plus Chroman-1 and spun down at 300 g for even coating. The plate was returned to the incubator for 1-3 hours to allow the iPSCs to attach to the surface. A lentiviral aliquot was thawed out on ice, and serially diluted 10 times. After 24 hours of incubation, the media was refreshed, the rock inhibitor was removed and returned to the incubator. The plate was imaged at 72 hours post transduction and the viral titer was estimated based on the percentage of eGFP positive cells. The transduction schedule was designed to maximize guide delivery and disease modeling per cell type. To deliver sgRNAs to target cells, lentiviral particles containing control or TDP-43-targeting sgRNAs were directly added to the iPSCs while in suspension. The iPSCs were singularized and spun down at 300 g for 5 minutes. The cell pellet was resuspended in Essential-8™ with 50 nM Chroman-1 followed by cell counting. Cell populations with a viability above 90% or more were used for the experiments. Based on the viral titer, spin down the desired number of iPSCs, remove the supernatant, and resuspend the cell pellet directly on the lentiviral particles. Allow to rest at room temperature in the biosafety cabinet for 10 minutes. After incubation, resuspend the iMuscle iPSCs on Day 0 media. On Day -2, the cultures were observed, and fresh media with hygromycin was supplied, as needed.

Direct differentiation of iPSCs into MyoD-1 iMuscle

After puromycin selection, two stable cell lines harboring MyOD-1-shRNA-OCT4 with Zim3-dCas9 machinery were obtained. To generate human skeletal muscle from iPSCs, a transcription factor cassette containing MyoD-1 and shRNA against Oct-4 was cloned under the Tet-On promoter, as described previously[50]. Briefly, on Day 0, iPSCs were singularized and seeded in Myogenic Induction Media (MIM) composed of DMEM/F12, 1X Sodium pyruvate (Gibco; #11360070), 1X NEAA, 110 µM 2-mercaptoethanol (Gibco; #21985023), 10 µg/mL Insulin (R&D; 11376497001), 50 nM Chroman-1, and 2 µg/mL Doxycycline. On Day 1, fresh warm media with a rock inhibitor was provided. The rock inhibitor was removed on Day 2 and daily media changes were performed until Day 6. At Day 7 the MIM was replenished and supplemented with 3 µM CHIR 99021 (Tocris; #4423) for 48 hours. On Day 10, a complete media change was performed using Maturation Media composed of Neurobasal A (Gibco; #10888022), 1X B27 plus, 1X NEAA, 1X GlutaMAX, 1X CultureONE, 10µg/mL NT-3, 50 ng/mL Sonic Hedgehog (Peprotech; #100-45), 100 ng /mL Agrin (R&D Systems ; #6624-AG-050), 50 ng/mL Laminin, 200 µM L-Ascorbic Acid (Thomas Scientific; #C988E92), 10 ng/mL human IGF-1(Peprotech;

#100-11), 2 µg/mL Doxycycline. Half media changes were performed three times per week from Day 12 until harvest on Day 14.

Sample Harvest, RNA Purification, and RNA sequencing

The iPS-derived cells were harvested on Day 14 for RNA analysis. The media was aspirated, and 300 µL of TRI Reagent® were added directly to the culture and harvested by scraping. Samples in TRI Reagent®were stored on -80°C for automated high throughput processing using Direct-zol™-96 MagBead RNA (Zymo Research # R2101) and Kingfisher® Apex Automated Purification Instrument with 96 DW head (Thermofisher; #5400930). Briefly, the RNA extraction was performed following the manufacturer's protocol, using a magnetic bead purification system, DNAse I treatment, and a final elution volume of 55 µL. The RNA concentration was determined Qubit™ Flex Fluorometer (ThermoFisher; #Q33327), and the RNA HS Assay Kit (Invitrogen #Q32855). Samples were normalized to 1 ng/µL and RNA quality was assessed by bioanalyzer. The New York Genome Center prepared libraries using the KAPA Hyper Stranded RNA kit at a sequencing length of 350 bp and depth of 40 M reads. Differential splicing analysis was performed as described below.

Publicly available data

Cross-linking and immunoprecipitation (CLIP) data, used for the analysis of the binding of TDP-43 and other RNA-binding proteins around and within CEs, was downloaded from the POSTAR database[48] and visualised by custom scripts in R. Information about gene expression in different tissues was obtained from ASCOT.[53] Genes of interest were selected and displayed using the tidyverse package in R.

Cryptic Exon Epitope Prediction and Synthesis

The pVACbind feature from pvactools[23] (v3.1.2) was utilized to evaluate epitope binding across the following HLA: HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-B*07:02, HLA-B*08:01, HLA-B*35:01, HLA-B*40:01, HLA-C*07:02. For each cryptic exon amino acid peptide sequence, 8 amino acids from the left and right flanking exon were included and 8 – 11mer sliding window epitopes were considered using the following MHC binding algorithms: MHCflurry, NetMHC, NetMHCpan, PickPocket, SMM, and SMMPMBEC. An epitope was considered a binder if any of the 6 algorithms called it a either a top 0.5% binder (HLA-B and HLA-C), or a top 1% binder (HLA-A). This strategy resulted in 379 distinct

peptide:MHC pairs across a total of 331 epitopes. All predicted epitopes were synthesized by Genscript.

High-throughput screening of peptide-MHC candidates for non-specific binding and noise.

Our lab has data suggesting that a few short 8 – 11mer amino acid peptides have the capacity to non-specifically bind to streptavidin and subsequently bind to the cell surface, independent of MHC interaction (data unpublished). This phenomenon greatly increases noise and reduces the specificity of Tetramer:TCR binding which is critical, especially in high-throughput screening. Therefore, following the synthesis of all peptides of interest and prior to UV-mediated loading of peptides onto HLA molecules, the peptides were screened for non-specific interactions with APC and PE Streptavidin. In this method, APC or PE Streptavidin is co-incubated with each peptide individually at a final concentration of 30nM and 150uM respectively in the absence of any HLA monomer. The streptavidin-peptide mixture is then co-incubated with either J76 or primary T cells for 30 minutes, then washed and evaluated using flow cytometry. Any peptide exhibiting more than 2% APC or PE signal was flagged and removed from the final tetramer panel.

High-dimensional, tetramer-associated T cell antigen receptor (TCR) sequencing (TetTCR-SeqHD)

TetTCR-SeqHD[18,19] is a previously described single cell sequencing method developed in our lab capable of profiling T cell cognate antigen specificities, T cell receptor sequences, gene expression, and surface protein expression. In brief, a large library of fluorescently labeled and DNA barcoded pMHC tetramers are generated through peptide synthesis and ultraviolet light-mediated peptide exchange. Next, pMHC tetramers are pooled and used to stain T cells along with barcoded antibodies (Abseq) for surface phenotyping and sample multiplexing. Antigen-bound T cells are sorted and are loaded onto a BD Rhapsody single cell capture chip. Following several PCR processing steps, the libraries are sequenced with next generation sequencing, and computationally processed to generate cell-UMI matrices across gene expression, surface phenotype, TCR expression, and tetramer-specificity.

Informatic Processing of Gene Expression, Abseq, Sampletag, TCR, and Tetramer

Raw single cell sequencing files were processed using a combination of BD Rhapsody's single cell pipeline in addition to a custom pipeline which generates tetramer cell-UMI matrices. Seurat (v4.0.2) was used to generate a merged Seurat object consisting of all experimental runs. RNA and

Abseq counts were normalized using the SCTransform function. A weighted nearest neighbor graph was generated using the FindMultiModalNeighbors function utilizing dimensionally reduced RNA and Abseq modalities, and a UMAP was constructed. Clusters were manually annotated using a combination of the differentially expressed genes and Abseq per cluster. Tetramer specificities per cell were assigned using a modified version of the Seurat HTODemux function. Specificities were refined by assigning *clonal specificities* such that at least 60% of the cells in a given clonotype as defined by their CDR3b amino acid sequence must share the same epitope assignment. TCR clonality was visualized with the scRepertoire (v.1.7.2). The frequency of clonotypes was binned by patient with the following parameters: Singleton = 1, Small = 3, Medium = 10, Large = 15, Hyperexpanded = 100, and clones were determined by the CDR3b amino acid sequence.

TCR Cloning, Transfection, and Transduction

The alpha and beta CDR3 sequences identified through the BD Rhapsody Sequence Analysis pipeline were aligned to the IMGT TCR reference to generate full length TRBV and TRAV sequences. The sequences were codon-optimized and IDT eblocks were ordered for each fragment. Fragments were assembled using HIFI assembly with an optimized backbone compatible with lentiviral transfection. The assemblies were transformed into NEB Stable Competent E.coli (High Efficiency) and a standard cloning protocol was used to generate plasmids. Sanger sequencing validated TCR plasmid were transfected into HEK293T cells in combination with a lentiviral packaging mix to generate virus. Virus was collected and concentrated for subsequent TCR transduction.

TCR Binding and Activation Validation

The binding specificity of the TCR-transduced cell lines was validated using flow cytometry with pMHC tetramers. To create pMHC complexes, 6 µL of 200 µM peptide and 2 µL of 0.2 mg/mL monomer were combined and UV-exchanged for an hour. After incubating at 4°C overnight, the pMHC complexes were first evaluated for stability using streptavidin-coated beads (6-8 µm) and incubated at 37°C for 1 hour. Following washing, PE-labeled anti-human β2m monoclonal antibody was used to stain for 30 minutes at 4°C and then analyzed by flow cytometry. For TCR binding validation, the stable pMHC complexes were tetramerized with 4 µL of 0.045 mg/mL fluorescently-labeled PE-streptavidin. 50,000 transduced cells were used for each condition and

stained with 1.5 µL of tetramer in 50 µL of FACS buffer (PBS + 2% FBS + 5mM EDTA) for 1 hour at 4°C before washing and running on flow cytometry. Untransduced J76-CD8 cell lines and irrelevant peptides were used as negative controls. To evaluate the activation of the TCR-transduced cell lines by the cognate peptides, we first coated 0.1 µg of streptavidin in an ELISA 96-well plate and incubated at room temperature overnight. After washing with 0.05% Tween20 in PBS, blocking buffer (2% BSA in PBS) was added and incubated at room temperature for 30 minutes. The pMHC complexes were diluted to 450 ng with blocking buffer for each reaction. CMV-specific pp65 pMHC with the corresponding transduced cell lines were used as controls for the activation assay. 100 µL of diluted pMHC monomer or 500 µM biotin solution was added to each sample or control wells, respectively, followed by incubation at 37°C for 1 hour. 125,000 TCR-transduced cells were added and incubated at 37°C for 18 hours. After washing with FACS buffer, the cells were stained with APC-CD69 antibody for 30 minutes and evaluated using flow cytometry.

CCF-STTG1 TDP-43 Knockdown: RNA-seq and HDGFL2 cryptic exon PCR Analysis

CCF-STTG1 cells were seeded in twelve-well plates at a density of 150,000 cells / well in 800 µL of RPMI 1640 +10% FBS without antibiotics. The following day, cells were transfected with either MISSION siRNA targeting human TARDBP (cat. EHU109221) or with MISSION siRNA Universal Negative Control #1 (cat. EHU109221). siRNA-Lipofectamine complexes were generated by combining 40 pmol of siRNA in 100 µL of OptiMem with 2 µL of Lipofectamine 2000 in 100 µL of OptiMem. After a 20 minute incubation, the combined mixture was added to the cells. Following a 48 hour incubation, RNA was isolated from each well by following Zymo's Quick-RNA Miniprep Kit (cat. R1054). RNAseq libraries were prepared using NEBNext Ultra II Directional with Poly-A selection and sequenced using Novaseq X 2x150, with 40M PE reads per sample.

To evaluate the HDGFL2 cryptic exon expression, a modified strategy described by Irwin et al.[6] was implemented. Briefly, reverse transcription was performed with a reverse primer targeting the exon region which follows the cryptic exon inclusion of HDGFL2: 5'-CTCGGAGGCTTCTTCACAGACA-3'. PCR was then performed using a primer pair targeting the exons flanking the cryptic exon inclusion: Fwd-primer 5'-AAGACGCCTGCGCTAAAGAT-3', Rev-primer 5'-CTCGGAGGCTTCTTCACAGACA-3'. Finally, HDGFL2 cryptic exon expression was evaluated with the following primer pair: Forward primer 5'-

AAGACGCCTGCGCTAAAGAT-3' and HDGFL2 CE Reverse primer 5'-GCTTCCCTCCCTTCTGATGC-3'.

CCF-STTG1:T cell Co-culture

The CCF-STTG1-GFP cell line was first generated by virally transducing CCF-STTG1 cells to constitutively express tagGFP2 using the viral transduction strategy described above. These cells were then plated at a seeding density of 2,000 cells/well in a flat 96-well plate in 50 µL of RPMI+10% FBS without antibiotics. The following day, TDP-43 or control siRNA was complexed with Lipofectamine 2000 as described above, and 1 pmol of siRNA was added to the cell culture. 24 hours following the siRNA transfection, cells were thoroughly washed with media and then co-incubated with the following primary T cell conditions: CD4/CD8 transduced, CD4/CD8 untransduced, and CD4/CD8 naive. In each condition, $CD4^+$ or $CD8^+$ T cells were enriched using EasySep Human $CD4^+$ or $CD8^+$ negative enrichment kit. The transduced and untransduced cells were both initially stimulated with ImmunoCult Human CD3/CD28 T Cell Activator. The naive cell conditions were unstimulated and CD45RA+CCR7+ cells were FACS sorted. A portion of the T cells were lysed and RNA was collected for RNAseq prior to co-incubation. The T cells were resuspended in RPMI + 10% FBS + 1x Caspase-3/7 Red and 5,000 T cells / well were co-incubated with the siRNA-treated CCF-STTG1-GFP cells. Each condition was performed with 3-4 replicates. Cells were monitored using the IncuCyte system where scans were taken every 3 hours.

CCF-STTG1 B2M-deficient shRNA generation

To generate stable B2M deficient shRNA cell lines, the pLKO.1 shRNA backbone was used to target the following sequences: 5'-

AGTTAAGCGTGCATAAGTTAACTCGAGTTAACTTATGCACGCTTAACT-3' and 5'-TTCAATCTCTTGCACTCAAAGCTCGAGCTTTGAGTGCAAGAGATTGAA-3'. The shRNA assemblies were transformed into NEB Stable Competent E. coli (High Efficiency), and standard molecular cloning protocols were used to generate plasmids, which were validated by full plasmid sequencing. The validated plasmids were transfected with a lentiviral packaging mix into HEK293T cells to produce lentiviral particles, which were collected, filtered, and concentrated. The virus was subsequently used to transduce CCF-STTG1-GFP astrocyte cells to establish stable shRNA-expressing cell lines.

GLIPH2 TCR analysis

We performed GLIPH2[54] clustering using our experimentally identified TCRs along with reference datasets from literature[9,24]. We further filtered the output from GLIPH2 to identify TCRs with a higher probability of sharing specificities. The GLIPH2 output was filtered to include only patterns that contained samples from both our experimental dataset and our reference set from the literature. For each pattern, we calculated the minimum Levenshtein distances between the TCRβ CDR3 sequences of the two groups as well as between TCRα CDR3a sequences, excluding any empty or missing values. The TRBV allele was compared to identify matching V gene segments between the groups within each pattern. We included patterns which contained CDR3 Levenshtein distances of one or less along with a matching TRBV allele between our experimental dataset and the reference.

Muscle biopsies

The use of samples for the present study was approved by Brain UK. All cases gave written informed consent in accordance with the principles outlined in the Declaration of Helsinki. Anonymisation of study samples was performed by a member of the clinical care team who was not directly involved in the study. We selected cases based on the following inclusion and exclusion criteria:

1. IBM group: Any person diagnosed with IBM by a qualified clinician, according to internationally accepted diagnostic criteria[55]
2. Controls: Normal muscle tissue, from patients investigated for cramps or fatigue with normal examination and neuro-physiology tests and normal histology

Quadriceps muscles from 17 IBM and 4 control cases were biopsied, flash-frozen in liquid nitrogen, and stored at -80 °C in the Neuropathology department of the National Hospital for Neurology and Neurosurgery (NHNN) in London, UK.

Immunohistochemistry

For IHC, 8 μm sections from 10 IBM cases and 4 controls were cut with a cryostat. DAB-staining consisted of 10 minutes of fixation with 4% PFA, followed by blocking endogenous peroxidases with 3% H2O2 in methanol. Pre-blocking in 10% milk/TBS-T for 30 minutes was followed by a 2-hour incubation with the following primary antibodies:

- TDP-43, mouse monoclonal (Abnova, 1:800)

- phosphoTDP-43, mouse polyclonal (Cosmo, 1:4000)
- p62, mouse monoclonal (Abcam, 1:100)
- HDGFL2 cryptic peptide, rabbit polyclonal (1:500)[20]

This was followed by a 30-minute incubation with secondary antibodies. The visualization step consisted of 30 minutes of incubation with ABC complex and DAB staining for 3 minutes, followed by counterstaining in Mayer's hematoxylin and mounting overnight prior to visualization on the microscope and scanning. Quantification of immune infiltration via IHC was performed by visually assessing the extent of immune cells in the whole sections in a blind way, and confirmed by a neuropathologist. Quantification of TDP-43, p62 inclusions, and HDGFL2 cryptic peptide burden analyses were based on DAB staining signal and performed in a blind, semiquantitative fashion, and confirmed by a neuropathologist.

Proteomics

4 25 **µ**m thick flash-frozen muscle biopsies from each IBM and control case were collected and 20 mg of tissue were dissolved in 100 **µ**l of lysis buffer, consisting of 50 mM Tris-HCl pH 8, 50 mM NaCl, 1% v/v SDS, 1% v/v Triton X-100, 1% v/v NP-40, 1% v/v Tween 20, 1% v/v glycerol, 1% sodium deoxycholate (w/v), 5 mM EDTA pH 8, 5 mM dithiothreitol, 5 KU benzonase and protease inhibitors. Following homogenization and spinning at 20000g at 4 °C for 15 minutes, cells were snap-frozen.

The sample preparation for mass spectrometry (MS)-based proteomics was performed on our previously published fully automated pipeline[56]. Briefly, Cell lysates were incubated at 65 °C for 30 min, followed by addition of 10 mM iodoacetamide and 30 min incubation in the dark. The denatured proteins were measured and normalized using an automated colorimetric assay on a Bravo (Agilent Technologies) robot. The protein enrichment and tryptic on-beads digestion were carried out on a KingFisher robot. The final resulting tryptic peptides were separated on a nano column (75 µm × 500 mm, 2 µm C18 particle) using a 2-h efficient linear gradient (Phase B, 2-35% ACN) on an UltiMate 3000 nano-HPLC system. We used data independent acquisition (DIA) discovery proteomics on a hybrid Orbitrap Eclipse mass spectrometer. Specifically, MS1 resolution was set to 120K, and MS2 resolution was set to 30K. For DIA isolation, the precursor range was set to 400–1000 m/z, and the isolation window was 8 m/z, resulting in 75 windows for each scan cycle (3 s). High collision dissociation was used for fragmentation with 30% collision

energy. The AGC target was set to 800% for MS2 scan (improve MS2 spectra quality). The MS2 scan range was defined as 145–1450 m/z, which covers most fragment ions. The database search was performed at the protein level by Spectronaut software 16.0. We used a direct DIA, which does not require a preconstructed high-quality project-specific spectral library but only uses a proteome sequence file. Specifically, the MS/MS spectra were searched against a proteome reference database containing 17,125 proteins (reviewed) mapped to the human reference genome (UP000005640) obtained via the UniProt Consortium. All spectra were allowed ''trypsin and lysC'' enzyme specificity, with up to two missed cleavages. Allowed fixed modifications included carbamidomethylation of cysteine and the allowed variable modifications for whole proteome datasets were acetylation of protein N-termini, oxidized methionine. The false discovery rates of peptide and protein were all set as 1%. The post-analysis and data visualization were carried out using Protpipe informatics suite[57]. Custom scripts written in R software were used to analyze and visualize the normalised count data.

RNA-seq analysis

RNA-seq data for n = 2 controls and n = 2 IBM were publicly available and were obtained from the authors (Johns Hopkins - JH cohort)[7]. Novel data included a cohort from Padua (PD), including n = 4 controls and n = 4 IBM, and a cohort from Verona (VR), including n = 5 controls and n = 7 IBM. Moreover, we sequenced 12 IBM and 4 control from our skeletal muscle biopsy cohort (Supplementary Table 1). 4 25 **µ**m thick flash-frozen muscle biopsies from each IBM and control case were collected and 20 **µ**g of tissue were dissolved in QIAzol Lysis Reagent (QIAGEN) prior to RNA extraction using the miRNeasy mini kit (QIAGeN). For the PD cohort, sequencing libraries were prepared with polyA enrichment and sequenced at UCL Genomics with the following specifics: 2×150 bp, depth >40M/sample. For the VR cohort, sequencing libraries were prepared with polyA enrichment and sequenced by Novogene with the following specifics: 2x150 bp, depth >40M/sample. For the NHNN cohort, sequencing libraries were prepared with ribo-depletion and sequenced by Novogene with the following specifics: 2x150 bp, depth >60M/sample. Raw sequences in fastq format were trimmed using Fastp[58], and aligned using STAR (v2.7.0f)[59] to the GRCh38 reference genome with gene models from GENCODE v42[60], including the generation of splice junction counts. For visualisation purposes, aligned BAM files were sorted and indexed by the read coordinates location using samtools[61]. To generate normalised

gene counts, trimmed fastqc files were aligned to the transcriptome using Salmon (v1.5.1)[62], before running DeSEQ2[63] without covariates, using an index built from GENCODE v42[60]. The DESeq2 median of ratios, which controls for both sequencing depth and RNA composition, was used to normalise gene counts. We kept genes with at least 5 counts per million in more than 2 samples. The alignment pipeline is implemented in Snakemake (v5.5.4)[64], and run on the UCL high-performance cluster.

To assess the presence of cryptic exon encoding for cryptic peptides in the IBM RNA-seq cohort, we used the splice junction output from STAR[59]. A cryptic exon was considered detected in a sample if there was at least one uniquely mapped spliced read supporting either the acceptor or the donor junction, and for quantification purposes, only the higher between the two was considered. Cryptic splice junctions were exported as a BED file that was used as a query to quantify cryptic exon expression in the IBM RNA-seq data. This returned a table with coordinates, gene name, and how many spliced reads in each sample supported that junction. Lastly, we normalised read count by the total number of mapped read per sample. This pipeline uses bedtools[65] and bash parsing with awk and is implemented in Snakemake v5.5.4[64]. Data visualisation was run on R using custom scripts.

References

1. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q. & Lee, V. M.-Y. Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. *Science* **314**, 130–133 (2006).
2. de Boer, E. M. J., Orie, V. K., Williams, T., Baker, M. R., De Oliveira, H. M., Polvikoski, T., Silsby, M., Menon, P., van den Bos, M., Halliday, G. M., van den Berg, L. H., Van Den Bosch, L., van Damme, P., Kiernan, M., van Es, M. A. & Vucic, S. TDP-43 proteinopathies: a new wave of neurodegenerative diseases. *J. Neurol. Neurosurg. Psychiatry* **92**, 86–95 (2021).
3. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349**, 650–5 (2015).
4. Brown, A. L., Wilkins, O. G., Keuss, M. J., Hill, S. E., Zanovello, M., Lee, W. C., Bampton, A., Lee, F. C. Y., Masino, L., Qi, Y. A., Bryce-Smith, S., Gatt, A., Hallegger, M., Fagegaltier, D., Phatnani, H., Newcombe, J., Gustavsson, E. K., Seddighi, S., Reyes, J. F., Coon, S. L., Ramos, D., Schiavo, G., Fisher, E. M. C., Raj, T., Secrier, M., Lashley, T., Ule, J., Buratti, E., Humphrey, J., Ward, M. E., Fratta, P. & Consortium, N. A. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature* **603**, 131–137 (2022).
5. Seddighi, S., Qi, Y. A., Brown, A. L., Wilkins, O. G., Bereda, C., Belair, C., Zhang, Y. J., Prudencio, M., Keuss, M. J., Khandeshi, A., Pickles, S., Kargbo-Hill, S. E., Hawrot, J., Ramos, D. M., Yuan, H., Roberts, J., Sacramento, E. K., Shah, S. I., Nalls, M. A., Colon-Mercado, J. M., Reyes, J. F., Ryan, V. H., Nelson, M. P., Cook, C. N., Li, Z., Screven, L., Kwan, J. Y., Mehta, P. R., Zanovello, M., Hallegger, M., Shantaraman, A., Ping, L., Koike, Y., Oskarsson, B., Staff, N. P., Duong, D. M., Ahmed, A., Secrier, M., Ule, J., Jacobson, S., Reich, D. S., Rohrer, J. D.,

Malaspina, A., Dickson, D. W., Glass, J. D., Ori, A., Seyfried, N. T., Maragkakis, M., Petrucelli, L., Fratta, P. & Ward, M. E. Mis-spliced transcripts generate de novo proteins in TDP-43-related ALS/FTD. *Sci Transl Med* eadg7162 (2024) doi:10.1126/scitranslmed.adg7162.

6. Irwin, K. E., Jasin, P., Braunstein, K. E., Sinha, I. R., Garret, M. A., Bowden, K. D., Chang, K., Troncoso, J. C., Moghekar, A., Oh, E. S., Raitcheva, D., Bartlett, D., Miller, T., Berry, J. D., Traynor, B. J., Ling, J. P. & Wong, P. C. A fluid biomarker reveals loss of TDP-43 splicing repression in presymptomatic ALS-FTD. *Nat Med* (2024) doi:10.1038/s41591-023-02788-5.
7. Britson, K. A., Ling, J. P., Braunstein, K. E., Montagne, J. M., Kastenschmidt, J. M., Wilson, A., Ikenaga, C., Tsao, W., Pinal-Fernandez, I., Russell, K. A., Reed, N., Mozaffar, T., Wagner, K. R., Ostrow, L. W., Corse, A. M., Mammen, A. L., Villalta, S. A., Larman, H. B., Wong, P. C. & Lloyd, T. E. Loss of TDP-43 function and rimmed vacuoles persist after T cell depletion in a xenograft model of sporadic inclusion body myositis. *Sci. Transl. Med.* **14**, eabi9196.
8. Ikenaga, C., Wilson, A. B., Irwin, K. E., Peethambaran Mallika, A., Kilgore, C., Sinha, I. R., Michelle, E. H., Ling, J. P., Wong, P. C. & Lloyd, T. E. Loss of TDP-43 Splicing Repression Occurs in Myonuclei of Inclusion Body Myositis Patients. *Ann. Neurol.* **97**, 629–641 (2025).
9. Campisi, L., Chizari, S., Ho, J. S. Y., Gromova, A., Arnold, F. J., Mosca, L., Mei, X., Fstkchyan, Y., Torre, D., Beharry, C., Garcia-Forn, M., Jiménez-Alcázar, M., Korobeynikov, V. A., Prazich, J., Fayad, Z. A., Seldin, M. M., De Rubeis, S., Bennett, C. L., Ostrow, L. W., Lunetta, C., Squatrito, M., Byun, M., Shneider, N. A., Jiang, N., La Spada, A. R. & Marazzi, I. Clonally expanded CD8 T cells characterize amyotrophic lateral sclerosis-4. *Nature* **606**, 945–952 (2022).
10. Yazdani, S., Seitz, C., Cui, C., Lovik, A., Pan, L., Piehl, F., Pawitan, Y., Kläppe, U., Press, R., Samuelsson, K., Yin, L., Vu, T. N., Joly, A.-L., Westerberg, L. S., Evertsson, B., Ingre, C., Andersson, J. & Fang, F. T cell responses at diagnosis of amyotrophic lateral sclerosis predict disease progression. *Nat. Commun.* **13**, 6733 (2022).

11. Greenberg, S. A., Pinkus, J. L., Kong, S. W., Baecher-Allan, C., Amato, A. A. & Dorfman, D. M. Highly differentiated cytotoxic T cells in inclusion body myositis. *Brain* **142**, 2590–2604 (2019).
12. Hu, D., Xia, W. & Weiner, H. L. CD8+ T cells in neurodegeneration: friend or foe? *Mol. Neurodegener.* **17**, 59 (2022).
13. Ramachandran, S., Grozdanov, V., Leins, B., Kandler, K., Witzel, S., Mulaw, M., Ludolph, A. C., Weishaupt, J. H. & Danzer, K. M. Low T-cell reactivity to TDP-43 peptides in ALS. *Front. Immunol.* **14**, (2023).
14. Dhanwani, R., Pham, J., Premlal, A. L. R., Frazier, A., Kumar, A., Pero, M. E., Bartolini, F., Dutra, J. R., Marder, K. S., Peters, B., Sulzer, D., Sette, A. & Lindestam Arlehamn, C. S. T Cell Responses to Neural Autoantigens Are Similar in Alzheimer's Disease Patients and Age-Matched Healthy Controls. *Front. Neurosci.* **14**, (2020).
15. Jiang, N., Malone, M. & Chizari, S. Antigen-specific and cross-reactive T cells in protection and disease. *Immunol. Rev.* **316**, 120–135 (2023).
16. Laumont, C. M. & Perreault, C. Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy. *Cell Mol Life Sci* **75**, 607–621 (2018).
17. Ely, Z. A., Kulstad, Z. J., Gunaydin, G., Addepalli, S., Verzani, E. K., Casarrubios, M., Clauser, K. R., Wang, X., Lippincott, I. E., Louvet, C., Schmitt, T., Kapner, K. S., Agus, M. P., Hennessey, C. J., Cleary, J. M., Hadrup, S. R., Klaeger, S., Su, J., Jaeger, A. M., Wolpin, B. M., Raghavan, S., Smith, E. L., Greenberg, P. D., Aguirre, A. J., Abelin, J. G., Carr, S. A., Jacks, T. & Freed-Pastor, W. A. Pancreatic cancer–restricted cryptic antigens are targets for T cell recognition. *Science* **388**, eadk3487 (2025).
18. Ma, K. Y., Schonnesen, A. A., He, C., Xia, A. Y., Sun, E., Chen, E., Sebastian, K. R., Guo, Y. W., Balderas, R., Kulkarni-Date, M. & Jiang, N. High-throughput and high-dimensional single-cell analysis of antigen-specific CD8(+) T cells. *Nat Immunol* **22**, 1590–1598 (2021).

19. Zhang, S.-Q., Ma, K.-Y., Schonnesen, A. A., Zhang, M., He, C., Sun, E., Williams, C. M., Jia, W. & Jiang, N. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat. Biotechnol.* **36**, 1156–1159 (2018).
20. Calliari, A., Daughrity, L. M., Albagli, E. A., Castellanos Otero, P., Yue, M., Jansen-West, K., Islam, N. N., Caulfield, T., Rawlinson, B., DeTure, M., Cook, C., Graff-Radford, N. R., Day, G. S., Boeve, B. F., Knopman, D. S., Petersen, R. C., Josephs, K. A., Oskarsson, B., Gitler, A. D., Dickson, D. W., Gendron, T. F., Prudencio, M., Ward, M. E., Zhang, Y. J. & Petrucelli, L. HDGFL2 cryptic proteins report presence of TDP-43 pathology in neurodegenerative diseases. *Mol Neurodegener* **19**, 29 (2024).
21. Ling, J. P., Wilks, C., Charles, R., Leavey, P. J., Ghosh, D., Jiang, L., Santiago, C. P., Pang, B., Venkataraman, A., Clark, B. S., Nellore, A., Langmead, B. & Blackshaw, S. ASCOT identifies key regulators of neuronal subtype-specific splicing. *Nat. Commun.* **11**, 137 (2020).
22. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., Fagerberg, L., Mulder, J., Luo, Y., Ponten, F., Uhlén, M. & Lindskog, C. A single–cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021).
23. Hundal, J., Kiwala, S., McMichael, J., Miller, C. A., Xia, H., Wollam, A. T., Liu, C. J., Zhao, S., Feng, Y. Y., Graubert, A. P., Wollam, A. Z., Neichin, J., Neveau, M., Walker, J., Gillanders, W. E., Mardis, E. R., Griffith, O. L. & Griffith, M. pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol Res* **8**, 409–420 (2020).
24. Gate, D., Saligrama, N., Leventhal, O., Yang, A. C., Unger, M. S., Middeldorp, J., Chen, K., Lehallier, B., Channappa, D., De Los Santos, M. B., McBride, A., Pluvinage, J., Elahi, F., Tam, G. K.-Y., Kim, Y., Greicius, M., Wagner, A. D., Aigner, L., Galasko, D. R., Davis, M. M. & Wyss-

Coray, T. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* **577**, 399–404 (2020).

25. Wang, P., Yao, L., Luo, M., Zhou, W., Jin, X., Xu, Z., Yan, S., Li, Y., Xu, C., Cheng, R., Huang, Y., Lin, X., Ma, K., Cao, H., Liu, H., Xue, G., Han, F., Nie, H. & Jiang, Q. Single-cell transcriptome and TCR profiling reveal activated and expanded T cell populations in Parkinson's disease. *Cell Discov.* **7**, 1–16 (2021).
26. Velebit, J., Horvat, A., Smolič, T., Prpar Mihevc, S., Rogelj, B., Zorec, R. & Vardjan, N. Astrocytes with TDP-43 inclusions exhibit reduced noradrenergic cAMP and Ca2+ signaling and dysregulated cell metabolism. *Sci. Rep.* **10**, 6003 (2020).
27. Peng, A. Y. T., Agrawal, I., Ho, W. Y., Yen, Y.-C., Pinter, A. J., Liu, J., Phua, Q. X. C., Koh, K. B., Chang, J.-C., Sanford, E., Man, J. H. K., Wong, P., Gutmann, D. H., Tucker-Kellogg, G. & Ling, S.-C. Loss of TDP-43 in astrocytes leads to motor deficits by triggering A1-like reactive phenotype and triglial dysfunction. *Proc. Natl. Acad. Sci.* **117**, 29101–29112 (2020).
28. Licht-Murava, A., Meadows, S. M., Palaguachi, F., Song, S. C., Jackvony, S., Bram, Y., Zhou, C., Schwartz, R. E., Froemke, R. C., Orr, A. L. & Orr, A. G. Astrocytic TDP-43 dysregulation impairs memory by modulating antiviral pathways and interferon-inducible chemokines. *Sci. Adv.* **9**, eade1282 (2023).
29. Yamanaka, K. & Komine, O. The multi-dimensional roles of astrocytes in ALS. *Neurosci. Res.* **126**, 31–38 (2018).
30. Walker, A. K., Daniels, C. M. L., Goldman, J. E., Trojanowski, J. Q., Lee, V. M.-Y. & Messing, A. Astrocytic TDP-43 Pathology in Alexander Disease. *J. Neurosci.* **34**, 6448–6458 (2014).
31. Xu, H. & Jia, J. Single-Cell RNA Sequencing of Peripheral Blood Reveals Immune Cell Signatures in Alzheimer's Disease. *Front. Immunol.* **12**, (2021).
32. Merlini, M., Kirabali, T., Kulic, L., Nitsch, R. M. & Ferretti, M. T. Extravascular CD3+ T Cells in

Brains of Alzheimer Disease Patients Correlate with Tau but Not with Amyloid Pathology: An Immunohistochemical Study. *Neurodegener. Dis.* **18**, 49–56 (2018).

33. Joshi, C., Sivaprakasam, K., Christley, S., Ireland, S., Rivas, J., Zhang, W., Sader, D., Logan, R., Lambracht-Washington, D., Rosenberg, R., Cullum, M., Hitt, B., Li, Q.-Z., Barber, R., Greenberg, B., Cowell, L., Zhang, R., Stowe, A., Huebinger, R., Kelley, B. & Monson, N. CSF-Derived CD4+ T-Cell Diversity Is Reduced in Patients With Alzheimer Clinical Syndrome. *Neurol. Neuroimmunol. Neuroinflammation* **9**, e1106 (2022).
34. Sulzer, D., Alcalay, R. N., Garretti, F., Cote, L., Kanter, E., Agin-Liebes, J., Liong, C., McMurtrey, C., Hildebrand, W. H., Mao, X., Dawson, V. L., Dawson, T. M., Oseroff, C., Pham, J., Sidney, J., Dillon, M. B., Carpenter, C., Weiskopf, D., Phillips, E., Mallal, S., Peters, B., Frazier, A., Lindestam Arlehamn, C. S. & Sette, A. T cells from patients with Parkinson's disease recognize α-synuclein peptides. *Nature* **546**, 656–661 (2017).
35. Lindestam Arlehamn, C. S., Dhanwani, R., Pham, J., Kuan, R., Frazier, A., Rezende Dutra, J., Phillips, E., Mallal, S., Roederer, M., Marder, K. S., Amara, A. W., Standaert, D. G., Goldman, J. G., Litvan, I., Peters, B., Sulzer, D. & Sette, A. α-Synuclein-specific T cell reactivity is associated with preclinical and early Parkinson's disease. *Nat. Commun.* **11**, 1875 (2020).
36. Singhania, A., Pham, J., Dhanwani, R., Frazier, A., Rezende Dutra, J., Marder, K. S., Phillips, E., Mallal, S., Amara, A. W., Standaert, D. G., Sulzer, D., Peters, B., Sette, A. & Lindestam Arlehamn, C. S. The TCR repertoire of α-synuclein-specific T cells in Parkinson's disease is surprisingly diverse. *Sci. Rep.* **11**, 302 (2021).
37. Galiano-Landeira, J., Torra, A., Vila, M. & Bové, J. CD8 T cell nigral infiltration precedes synucleinopathy in early stages of Parkinson's disease. *Brain* **143**, 3717–3733 (2020).
38. Varrin-Doyer, M., Shetty, A., Spencer, C. M., Schulze-Topphoff, U., Weber, M. S., Bernard, C. C. A., Forsthuber, T., Cree, B. A. C., Slavin, A. J. & Zamvil, S. S. MOG transmembrane and

cytoplasmic domains contain highly stimulatory T-cell epitopes in MS. *Neurol. Neuroimmunol. Neuroinflammation* **1**, e20 (2014).

39. Sabatino, J. J., Wilson, M. R., Calabresi, P. A., Hauser, S. L., Schneck, J. P. & Zamvil, S. S. Anti-CD20 therapy depletes activated myelin-specific CD8+ T cells in multiple sclerosis. *Proc. Natl. Acad. Sci.* **116**, 25800–25807 (2019).
40. Bronge, M., Ruhrmann, S., Carvalho-Queiroz, C., Nilsson, O. B., Kaiser, A., Holmgren, E., Macrini, C., Winklmeier, S., Meinl, E., Brundin, L., Khademi, M., Olsson, T., Gafvelin, G. & Grönlund, H. Myelin oligodendrocyte glycoprotein revisited—sensitive detection of MOG-specific T-cells in multiple sclerosis. *J. Autoimmun.* **102**, 38–49 (2019).
41. Bronge, M., Högelin, K. A., Thomas, O. G., Ruhrmann, S., Carvalho-Queiroz, C., Nilsson, O. B., Kaiser, A., Zeitelhofer, M., Holmgren, E., Linnerbauer, M., Adzemovic, M. Z., Hellström, C., Jelcic, I., Liu, H., Nilsson, P., Hillert, J., Brundin, L., Fink, K., Kockum, I., Tengvall, K., Martin, R., Tegel, H., Gräslund, T., Al Nimer, F., Guerreiro-Cacais, A. O., Khademi, M., Gafvelin, G., Olsson, T. & Grönlund, H. Identification of four novel T cell autoantigens and personal autoreactive profiles in multiple sclerosis. *Sci. Adv.* **8**, eabn1823 (2022).
42. Súkeníková, L., Mallone, A., Schreiner, B., Ripellino, P., Nilsson, J., Stoffel, M., Ulbrich, S. E., Sallusto, F. & Latorre, D. Autoreactive T cells target peripheral nerves in Guillain–Barré syndrome. *Nature* **626**, 160–168 (2024).
43. Zhang, Y. H., Ni, Y., Gao, Y. N., Shen, D. D., He, L., Yin, D., Meng, H. Y., Zhou, Q. M., Hu, J. & Chen, S. Anti-IgLON5 disease: a novel topic beyond neuroimmunology. *Neural Regen Res* **18**, 1017–1022 (2023).
44. Cagnin, A., Mariotto, S., Fiorini, M., Gaule, M., Bonetto, N., Tagliapietra, M., Buratti, E., Zanusso, G., Ferrari, S. & Monaco, S. Microglial and Neuronal TDP-43 Pathology in Anti-IgLON5-Related Tauopathy. *J Alzheimers Dis* **59**, 13–20 (2017).

45. Bensimon, G., Leigh, P. N., Tree, T., Malaspina, A., Payan, C. A., Pham, H.-P., Klaassen, P., Shaw, P. J., Khleifat, A. A., Amador, M. D. M., Attarian, S., Bell, S. M., Beltran, S., Bernard, E., Camu, W., Corcia, P., Corvol, J.-C., Couratier, P., Danel, V., Debs, R., Desnuelle, C., Dimitriou, A., Ealing, J., Esselin, F., Fleury, M.-C., Gorrie, G. H., Grapperon, A.-M., Hesters, A., Juntas-Morales, R., Kolev, I., Lautrette, G., Forestier, N. L., McDermott, C. J., Pageot, N., Salachas, F., Sharma, N., Soriani, M.-H., Sreedharan, J., Svahn, J., Verber, N., Verschueren, A., Yildiz, O., Suehs, C. M., Saker-Delye, S., Muller, C., Masseguin, C., Hajduchova, H., Kirby, J., Garlanda, C., Locati, M., Zetterberg, H., Asselain, B. & Al-Chalabi, A. Efficacy and safety of low-dose IL-2 as an add-on therapy to riluzole (MIROCALS): a phase 2b, double-blind, randomised, placebo-controlled trial. *The Lancet* **0**, (2025).
46. Perrin, S., Ladha, S., Maragakis, N., Rivner, M. H., Katz, J., Genge, A., Olney, N., Lange, D., Heitzman, D., Bodkin, C., Jawdat, O., Goyal, N. A., Bornstein, J. D., Mak, C., Appel, S. H. & Paganoni, S. Safety and tolerability of tegoprubart in patients with amyotrophic lateral sclerosis: A Phase 2A clinical trial. *PLOS Med*. **21**, e1004469 (2024).
47. Gao, W., Kim, M. W., Dykstra, T., Du, S., Boskovic, P., Lichti, C. F., Ruiz-Cardozo, M. A., Gu, X., Weizman Shapira, T., Rustenhoven, J., Molina, C., Smirnov, I., Merbl, Y., Ray, W. Z. & Kipnis, J. Engineered T cell therapy for central nervous system injury. *Nature* 1–9 (2024) doi:10.1038/s41586-024-07906-y.
48. Zhao, W., Zhang, S., Zhu, Y., Xi, X., Bao, P., Ma, Z., Kapral, T. H., Chen, S., Zagrovic, B., Yang, Y. T. & Lu, Z. J. POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Res*. **50**, D287–D294 (2022).
49. Fernandopulle, M. S., Prestil, R., Grunseich, C., Wang, C., Gan, L. & Ward, M. E. Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons: Rapid differentiation of iPSCs into neurons. *Curr. Protoc. Cell Biol*. **79**, e51 (2018).

50. Vila, O. F., Chavez, M., Ma, S. P., Yeager, K., Zholudeva, L. V., Colón-Mercado, J. M., Qu, Y., Nash, T. R., Lai, C., Feliciano, C. M., Carter, M., Kamm, R. D., Judge, L. M., Conklin, B. R., Ward, M. E., McDevitt, T. C. & Vunjak-Novakovic, G. Bioengineered optogenetic model of human neuromuscular junction. *Biomaterials* **276**, 121033 (2021).
51. Replogle, J. M., Bonnar, J. L., Pogson, A. N., Liem, C. R., Maier, N. K., Ding, Y., Russell, B. J., Wang, X., Leng, K., Guna, A., Norman, T. M., Pak, R. A., Ramos, D. M., Ward, M. E., Gilbert, L. A., Kampmann, M., Weissman, J. S. & Jost, M. Maximizing CRISPRi efficacy and accessibility with dual-sgRNA libraries and optimal effectors. *eLife* **11**, e81856 (2022).
52. Wu, W., Hill, S. E., Nathan, W. J., Paiano, J., Callen, E., Wang, D., Shinoda, K., Van Wietmarschen, N., Colón-Mercado, J. M., Zong, D., De Pace, R., Shih, H.-Y., Coon, S., Parsadanian, M., Pavani, R., Hanzlikova, H., Park, S., Jung, S. K., McHugh, P. J., Canela, A., Chen, C., Casellas, R., Caldecott, K. W., Ward, M. E. & Nussenzweig, A. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).
53. Ling, J. P., Wilks, C., Charles, R., Leavey, P. J., Ghosh, D., Jiang, L., Santiago, C. P., Pang, B., Venkataraman, A., Clark, B. S., Nellore, A., Langmead, B. & Blackshaw, S. ASCOT identifies key regulators of neuronal subtype-specific splicing. *Nat. Commun.* **11**, 137 (2020).
54. Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* **38**, 1194–1202 (2020).
55. Rose, M. R. 188th ENMC International Workshop: Inclusion Body Myositis, 2–4 December 2011, Naarden, The Netherlands. *Neuromuscul. Disord.* **23**, 1044–1055 (2013).
56. Reilly, L., Lara, E., Ramos, D., Li, Z., Pantazis, C. B., Stadler, J., Santiana, M., Roberts, J., Faghri, F., Hao, Y., Nalls, M. A., Narayan, P., Liu, Y., Singleton, A. B., Cookson, M. R., Ward, M. E. & Qi, Y. A. A fully automated FAIMS-DIA mass spectrometry-based proteomic pipeline. *Cell Rep.*

Methods **3**, 100593 (2023).

57. Li, Z., Weller, C. A., Shah, S., Johnson, N., Hao, Y., Roberts, J., Bereda, C., Klaisner, S., Machado, P., Fratta, P., Petrucelli, L., Prudencio, M., Oskarsson, B., Staff, N. P., Dickson, D. W., Cookson, M. R., Ward, M. E., Singleton, A. B., Nalls, M. A. & Qi, Y. A. ProtPipe: A Multifunctional Data Analysis Pipeline for Proteomics and Peptidomics. Preprint at https://doi.org/10.1101/2023.12.12.571327 (2023).
58. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
59. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J. M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., Gonzalez, J. M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O. G., Lagarde, J., Martin, F. J., Martínez, L., Mohanan, S., Muir, P., Navarro, F. C. P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B. M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J. S., Gerstein, M., Guigó, R., Hubbard, T. J. P., Kellis, M., Paten, B., Reymond, A., Tress, M. L. & Flicek, P. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
61. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
62. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

64. Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S. & Köster, J. Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).

65. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).