



This is a repository copy of *Evaluation of a biomarker for amyotrophic lateral sclerosis derived from a hypomethylated DNA signature of human motor neurons.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/221987/>

Version: Published Version

---

**Article:**

Harvey, C., Nowak, A., Zhang, S. et al. (13 more authors) (2025) Evaluation of a biomarker for amyotrophic lateral sclerosis derived from a hypomethylated DNA signature of human motor neurons. BMC Medical Genomics, 18. 10. ISSN 1755-8794

<https://doi.org/10.1186/s12920-025-02084-w>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

RESEARCH

Open Access



# Evaluation of a biomarker for amyotrophic lateral sclerosis derived from a hypomethylated DNA signature of human motor neurons

Calum Harvey<sup>1</sup>, Alicja Nowak<sup>1</sup>, Sai Zhang<sup>2</sup>, Tobias Moll<sup>1</sup>, Annika K Weimer<sup>5</sup>, Aina Mogas Barcons<sup>1</sup>, Cleide Dos Santos Souza<sup>1</sup>, Laura Ferraiuolo<sup>1</sup>, Kevin Kenna<sup>3</sup>, Noah Zaitlen<sup>4</sup>, Christa Caggiano<sup>4</sup>, Pamela J Shaw<sup>1</sup>, Michael P Snyder<sup>5</sup>, Jonathan Mill<sup>6</sup>, Eilis Hannon<sup>6</sup> and Johnathan Cooper-Knock<sup>1\*</sup>

## Abstract

Amyotrophic lateral sclerosis (ALS) lacks a specific biomarker, but is defined by relatively selective toxicity to motor neurons (MN). As others have highlighted, this offers an opportunity to develop a sensitive and specific biomarker based on detection of DNA released from dying MN within accessible biofluids. Here we have performed whole genome bisulfite sequencing (WGBS) of iPSC-derived MN from neurologically normal individuals. By comparing MN methylation with an atlas of tissue methylation we have derived a MN-specific signature of hypomethylated genomic regions, which accords with genes important for MN function. Through simulation we have optimised the selection of regions for biomarker detection in plasma and CSF cell-free DNA (cfDNA). However, we show that MN-derived DNA is not detectable via WGBS in plasma cfDNA. In support of our experimental finding, we show theoretically that the relative sparsity of lower MN sets a limit on the proportion of plasma cfDNA derived from MN which is below the threshold for detection via WGBS. Our findings are important for the ongoing development of ALS biomarkers. The MN-specific hypomethylated genomic regions we have derived could be usefully combined with more sensitive detection methods and perhaps with study of CSF instead of plasma. Indeed we demonstrate that neuronal-derived DNA is detectable in CSF. Our work is relevant for all diseases featuring death of rare cell-types.

**Keywords** Amyotrophic lateral sclerosis (ALS), Biomarker, Cell-free DNA, DNA methylation, Whole-genome bisulfite sequencing, iPSC-derived motor neuron

\*Correspondence:

Johnathan Cooper-Knock  
j.cooper-knock@sheffield.ac.uk

<sup>1</sup>Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK

<sup>2</sup>Department of Epidemiology, University of Florida, Gainesville, FL, USA

<sup>3</sup>Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>4</sup>Departments of Computational Medicine and Neurology, UCLA, Los Angeles, CA, USA

<sup>5</sup>Center for Genomics and Personalized Medicine, Stanford University School of Medicine, Stanford, CA, USA

<sup>6</sup>University of Exeter Medical School, University of Exeter, Exeter, UK



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease where death results from motor neuron (MN) loss leading to respiratory failure. The design and development of novel therapeutics has been held back because of the lack of a specific biomarker. Currently, neurofilament proteins measured in plasma provide a non-specific readout of neuronal death [1]. Neurofilament proteins form important structural components of the large myelinated axons which are found in MN. MN death triggers the release of neurofilaments from the cytoplasm into the extracellular space [2]; as a result the level of detectable neurofilament is a function of the rate of MN death, and thus neurofilament measurement can be used as a biomarker of disease progression [1]. However, neurofilaments are not specific to MN and it is notable that serum neurofilament light chain (NfL) [3] is elevated in other neurological diseases. Indeed, for diagnosis of ALS, serum NfL is of limited value [4] even if it is useful for measuring the rate of progression. It follows that detection of a different marker which is released *only* from dying MN may outperform neurofilaments as a biomarker for ALS.

DNA methylation is fundamental to the control of gene expression and by inference, genomic methylation should be relatively cell specific. Cell-specific DNA methylation signals are stable between individuals, as was confirmed by a recent atlas of DNA methylation [5]. Moreover, DNA methylation is relatively stable over time [6]. Cell-free DNA (cfDNA) found in peripheral blood is the product of release from dying cells [7] and has been extensively proposed as a source of biomarkers in the cancer field [8]; methylated cfDNA is now the basis of FDA-approved applications e.g [9]. We hypothesised that a DNA methylation signature which is specific to MN, and is detectable within cfDNA, might be both sensitive and specific as a biomarker of the rate of MN death due to ALS.

We present whole genome bisulfite sequencing (WGBS) data from iPSC-derived MN from controls. These data complement our previously published epigenetic profiling from the same neurons [10]. It is practically difficult to obtain MN in sufficient quantity from post-mortem material to perform WGBS [11] and therefore we chose to focus on iPSC-derived MN which are a gold-standard model of ALS [12]. We note previous work demonstrating that DNA methylation changes detectable in iPSC-derived MN have correlates in ALS patient CNS tissue and in peripheral biofluids [13]. We have published WGBS of cfDNA from ALS patients and controls [14] but previously we lacked a MN signature for comparison. Here we show, using simulation and measurement, that MN-specific DNA methylation is not detectable within cfDNA in plasma by WGBS. Future work will evaluate

our MN DNA methylation signature by other means and in other biofluids. Our approach is summarised in Fig. 1.

## Results

### Cell-specific DNA methylation within control iPSC-derived MN is similar to human adult CNS neurons

WGBS was performed at high depth to profile DNA methylation within iPSC-derived MN from two independent differentiations from three neurologically normal individuals (Supplementary Table 1, Methods). A first question was whether the methylation signature of these neurons, which are derived *in vitro*, is consistent with CNS neurons obtained from human tissue.

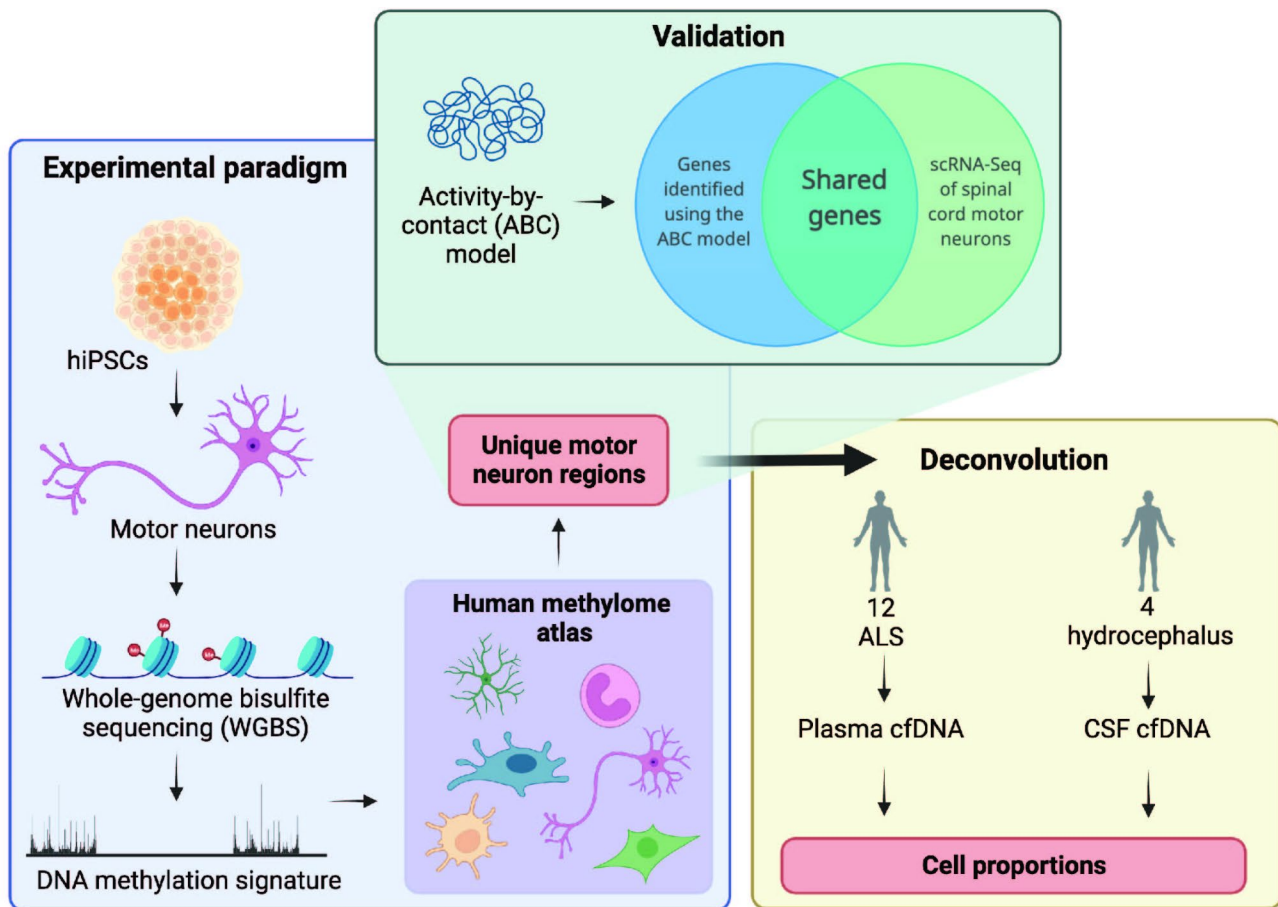
WGBS sequencing data were processed and quality control (QC) was performed according to the ENCODE 4 standards [16]. Methylation profiles of 205 samples covering 39 cell-types from an available methylation atlas [5] were combined with our samples, then used to segment the genome into blocks of co-methylated CpGs (Methods). Hierarchical unsupervised clustering was used to examine the relationships between samples (Methods, Fig. 2A). As expected, genome methylation within iPSC-derived MN clustered closely with CNS neuronal subtypes (Fig. 2B). On this basis we proceeded to use our data to identify MN-specific methylation (Methods).

### Identification of cell-specific hypomethylated genomic regions

Next we derived DNA methylation changes specific to MN via comparison with the methylation profiles of 205 samples covering 39 cell-types from an available methylation atlas [5]. Blocks of co-methylated CpGs that exhibited hyper- or hypomethylation specifically in MN were identified (Methods) and taken forward for further analysis. In total 8,729 regions were specifically hypomethylated in MN (Supplementary Table 2); hypomethylation indicates increased genomic accessibility suggestive of MN-specific function. A similar analysis identified 5,690 blocks which were specifically hypomethylated in the total set of human CNS neurons compared to other cell-types. The number of regions identified per cell-type varied dramatically from 61,693 for gallbladder to 436 for colon fibroblasts.

### MN-specific DNA methylation is linked to MN function but not to genetic risk for ALS

Cell-specific DNA methylation is typically hypomethylated [5], which should be coincident with increased accessibility of underlying DNA over regulatory regions including enhancers [17]. As a validation of the regions we have identified, we examined the overlap of MN-specific hypomethylated enhancers and their target genes, with independent measurements of MN gene expression and ALS heritability (Fig. 3A).



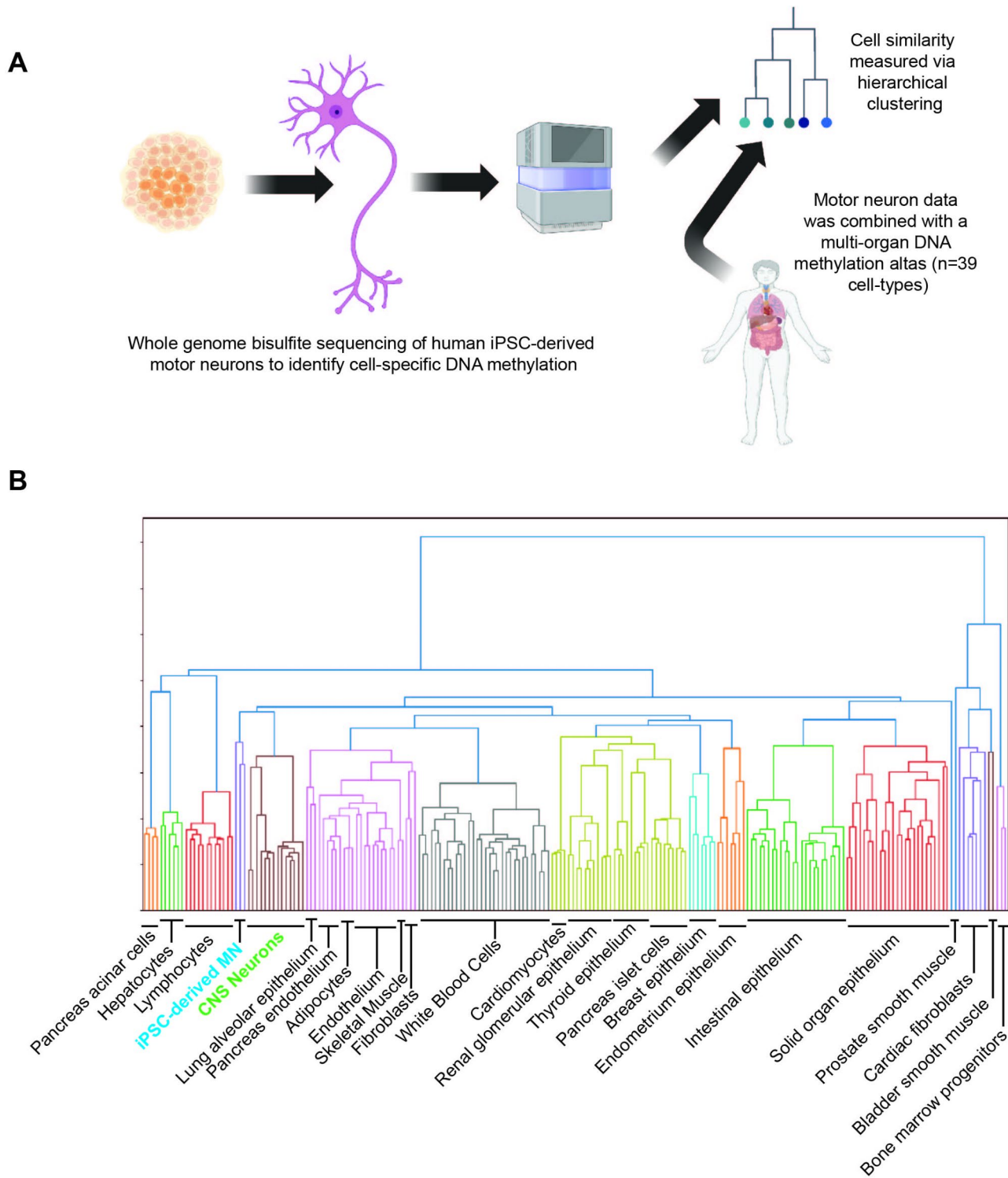
**Fig. 1** Derivation and biomarker evaluation of a hypomethylated DNA signature from whole genome bisulfite sequencing (WGBS) of human motor neurons. MN-specific DNA hypomethylation was used to assess the proportion of MN DNA within cfDNA in plasma from ALS patients ( $n=12$ ) and CSF from controls ( $n=4$ ). We sort to verify the validity of MN-specific DNA hypomethylated regions by linking regions to target genes and cross-checking those genes with independent observations of MN gene expression; we hypothesised that correctly identified hypomethylated regions should indicate regions of open, active and transcribed chromatin which should be statistically enriched in measures of MN-specific gene expression. We linked regions to target genes using the activity-by-contact (ABC) model [15]

To derive associated genes from MN-specific hypomethylated DNA blocks, we applied the activity-by-contact (ABC) model [15] to link regulatory regions to expressed genes within iPSC-derived MN (Methods). We found the total list of hypomethylated regions is associated with 2,046 expressed genes. We then tested this gene list for enrichment with gene expression specific to human cell types and tissues included in ARCHS4 [18] using Enrichr [19], and found they were most significantly enriched for genes expressed specifically in spinal motor neurons isolated from post-mortem tissue [20] (Fisher's exact test,  $p=4.22e-19$ , OR=1.79, using the ARCHS4 database [18], Fig. 3B). This demonstrates that the methylation profiles of the iPSC derived motor neurons are congruent with transcriptional profiles of human motor neurons.

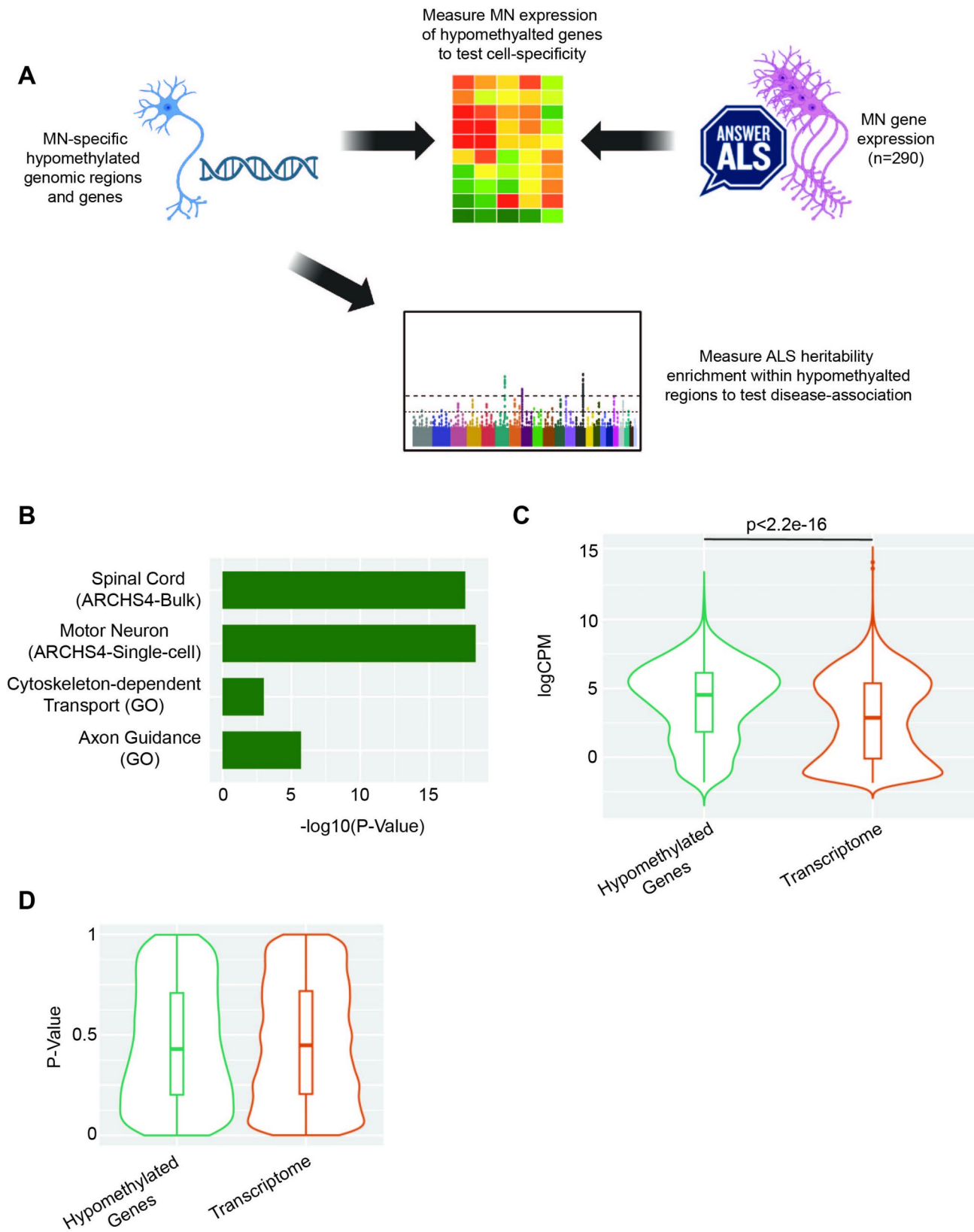
To further characterise the function of MN-specific hypomethylated genes we examined RNA-sequencing from iPSC-derived motor neurons obtained from

245 ALS patients and 45 controls ([www.answerals.org](http://www.answerals.org)) (Methods). Genes linked to hypomethylated regions in MN were highly expressed within iPSC-derived MN compared to the background transcriptome (Wilcoxon rank sum test,  $p < 2.2e-16$ , Fig. 3C) which is consistent with an important role in MN function. Four genes were reported as differentially expressed (FDR < 0.05, negative binomial test) between ALS patients and controls in this data, but genes linked to hypomethylated regions in MN were not enriched within ALS-associated differentially expressed genes (Wilcoxon rank sum test,  $p=0.25$ , Fig. 3D).

Finally, we performed linkage disequilibrium score regression (LDSC) [21] using a recent GWAS study of ALS [22] to examine disease-specific heritability enrichment within MN-specific hypomethylated regions. Heritability for ALS was enriched within hypomethylated regions but this was not statistically significant (OR=25.2, se=26.05,  $p=0.38$ , LDSC, Methods). We conclude that MN-specific DNA hypomethylation is



**Fig. 2** iPSC-derived MN maintain a DNA methylation signature consistent with human adult neurons. **(A)** Whole genome bisulfite sequencing (WGBS) of genomic DNA derived from human iPSC-derived MN was used to derive a profile of genomic methylation within MN for comparison with methylation profiles of 205 samples covering 39 cell-types [5]. **(B)** Unsupervised clustering was used to assess cell-similarity and revealed that iPSC-derived MN (blue text) cluster together with human CNS neurons (green text)



**Fig. 3** MN specific DNA methylation is linked to MN function but not to genetic risk for ALS. **(A)** We used independent measurements of MN gene expression and ALS heritability to verify the biological validity of identified MN-specific hypomethylated genomic regions. MN-specific hypomethylated genes are enriched with genes expressed in human MN **(B)** and in human iPSC-derived MN **(C)**. MN-specific hypomethylated genes are not differentially expressed in ALS iPSC-derived MN compared to control MN **(D)**

associated with gene expression linked to MN function, but we find no conclusive evidence that there is a specific association with genes dysregulated in MN in a disease context.

#### **An optimum set of hypomethylated DNA regions for ALS biomarker design**

An important use of cell-type-specific methylation profiles is for the deconvolution of complex mixes of DNA to identify the proportions of contributing cell types. This has the potential to lead to a novel biomarker of ALS: Cell-free DNA (cfDNA) found within plasma is released from dying cells and thus, the quantity of DNA sourced from CNS neurons, and MN in particular, should be proportional to the rate of MN death. Neuronal DNA is not normally seen in the plasma [5], which may be due to a low rate of neuron death or to the blood brain barrier, but brain-derived DNA has been detected in plasma under pathological conditions [23, 24] demonstrating its potential to serve as a biomarker.

To deconvolute plasma cfDNA we optimised the UXM algorithm [5] for the low coverage (~10x) typical of methylation studies of cfDNA; in particular we optimised the choice and configuration of MN-specific methylation blocks. The UXM algorithm was chosen as it makes use of read level methylation data, and has achieved accurate deconvolution of cell types present at proportions as low as 0.1% [5]. Optimisation was performed using synthetic data generated by spiking WGBS data derived from plasma cfDNA of healthy individuals, with sequencing reads derived from human MN at a known proportion between 0.01 and 10% (Methods, Fig. 4A). We simulated relatively low coverage (10x) to match coverage in the actual ALS cfDNA samples. We observed a linear correlation between the actual and predicted percentage of spike-in MN DNA with an adjusted  $r^2 < 0.9$  in all marker sets (Fig. 4B). A configuration of UXM using 500 MN-specific blocks with a minimum of 3 CpGs produced the highest detection probability at 1% spike-in, but 500 blocks with a minimum of 4 CpGs performed better at both 0.5% and 0.1% spike-in (difference in detection probability between 0.1 and 0.2 at each % spike-in, Fig. 4C). However, we note that at spike-ins of  $\leq 0.5\%$ , AUC was poor for all sets of MN marker blocks. The greatest AUC (0.69) at 1% spike-in was achieved with 500 blocks with a minimum of 3 CpGs, in keeping with its higher probability of detection (Supplementary Fig. 1A); this was the configuration taken forward to analyse ALS patient samples.

As seen in [5, 25], deconvolution frequently identified false-positive cell-types within the synthetic mixture (Supplementary Fig. 2B). We used a linear model to examine the effect of coverage and number of marker regions the total number of cell types identified in a

sample. Both coverage ( $p=0.04$ ) and number of markers ( $p=3.7e-4$ ) were significantly negatively correlated with the number of cell types identified, suggesting that increased coverage and using more marker regions per cell-type will reduce the number of cell types falsely identified within a mixture.

#### **MN-derived DNA is not detectable within plasma cfDNA**

When we applied our optimised deconvolution utilising 500 MN-specific methylation blocks with a minimum of 3 CpGs to plasma cfDNA WGBS from  $n=12$  ALS patients we did not identify MN-derived DNA in any sample (Fig. 4D) suggesting that if MN DNA is present it is below the detectable limit of ~1% of plasma cfDNA (Fig. 4B-C).

#### **Neuronal-derived DNA is detectable in CSF cfDNA**

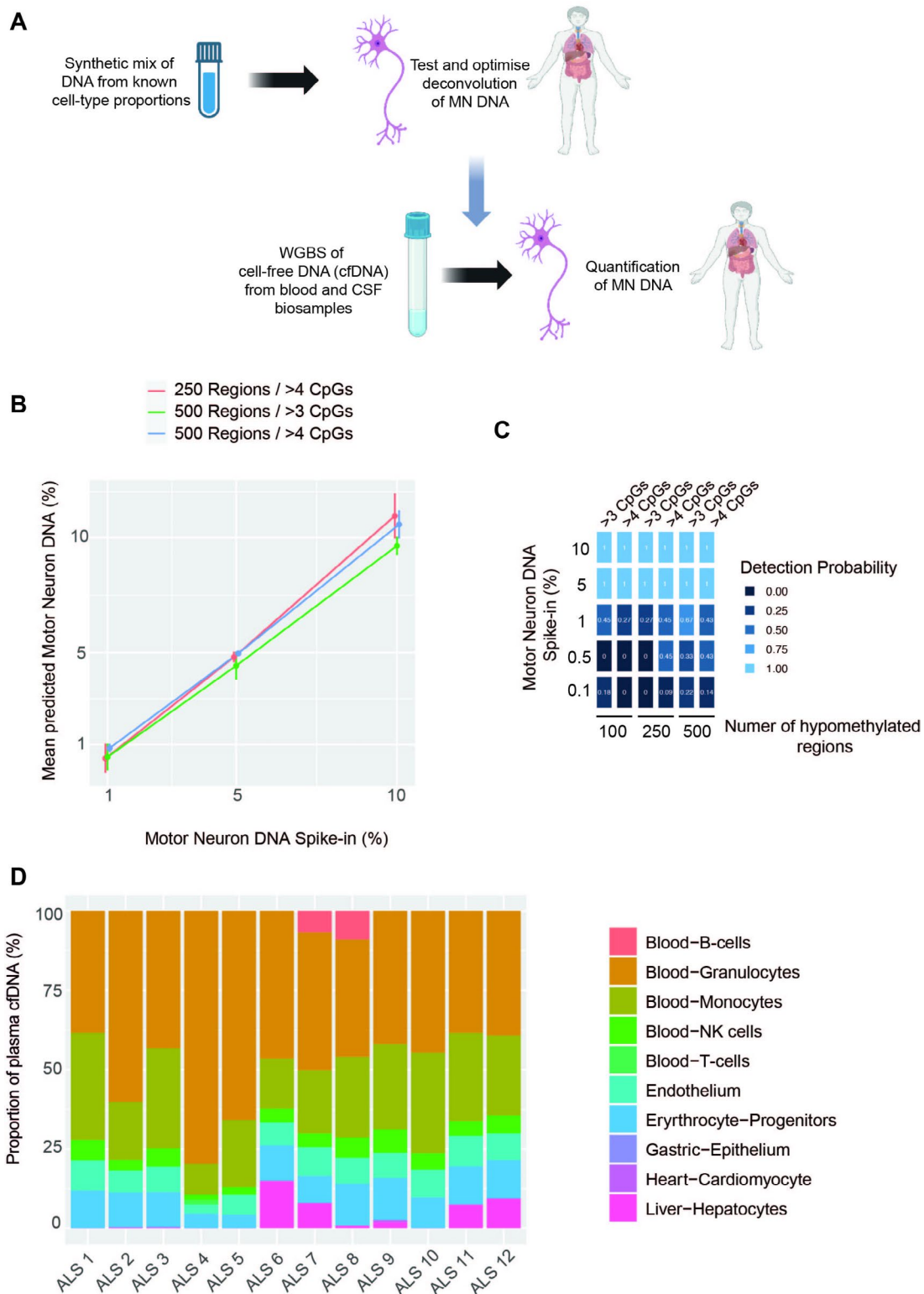
The cerebrospinal fluid (CSF) surrounds the spinal cord and brain, and is encapsulated by the inner cerebrospinal fluid-brain barrier. It might be expected that CSF cfDNA is enriched in neuronal DNA compared to plasma and so we attempted to fully characterise the contributing cell types within CSF cfDNA (Methods).

No WGBS data was available from ALS patient CSF cfDNA. We analysed WGBS of CSF cfDNA derived from four hydrocephalus patients [26]. Coverage was very low (0.12-0.45x, Supplementary Table 3) due to the low concentration of cfDNA within the CSF, so samples were merged to improve deconvolution accuracy. We discovered that neuronal and oligodendrocyte DNA comprised 13% and 14% of the total cfDNA with the remainder largely composed of a mix of blood, epithelial, and adipocyte cell types (Supplementary Fig. 2); MN-derived DNA was not detectable in any sample. The contribution of adipocytes may in part reflect the lumbar puncture procedure used to collect CSF as DNA. The lack of a number of CNS-specific cell-types such as microglia within the reference leads to a possible assignment error which is impossible to quantify, and is likely responsible for the small proportion of epithelial and pancreatic cell types identified.

#### **The theoretical maximum proportion of MN-derived DNA within plasma cfDNA is very low**

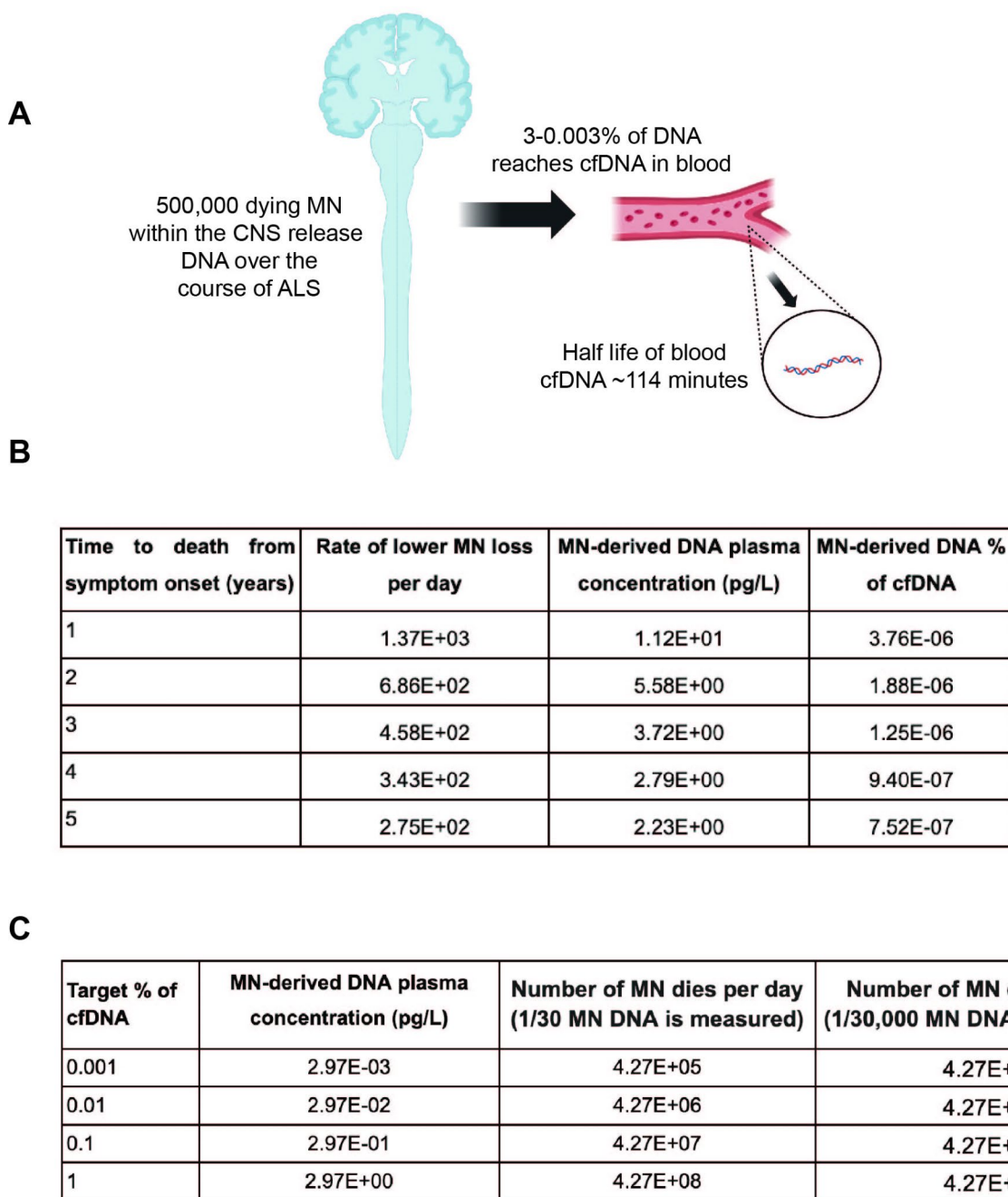
We did not detect MN DNA in any ALS patient sample suggesting that if MN DNA is present it is below ~1% of plasma cfDNA. We questioned if this was a detection deficiency or whether there might be insufficient MN DNA for detection. To address this we modelled the theoretical maximum proportion of MN DNA that might be expected within plasma cfDNA (Fig. 5A).

Recent work [27] has estimated the effect of cellular turnover on the proportion of DNA derived from different cell-types detectable within plasma cfDNA.



**Fig. 4** Optimised set of MN-specific hypomethylated genomic regions is not detectable in ALS patient plasma cfDNA. **(A)** We used a synthetic mix of WGBS reads from non-diseased plasma cfDNA together with spike-in reads from iPSC-derived MN to determine the optimum set of MN-specific regions for detection in ALS patient biosamples. **(B)** At spike-ins of 1–10% there is a linear relationship between spike-in and predicted MN DNA concentrations for all sets of MN-specific methylation blocks;  $p < 0.02$ , adjusted  $r^2 > 0.998$ , Pearson's product moment correlation coefficient. **(C)** At spike-ins  $\leq 1\%$  it is possible to detect reads derived from MN-specific regions but the detection probability is  $< 0.5$ . **(D)** MN-specific DNA is not detectable within ALS patient plasma





**Fig. 5** The theoretical maximum proportion of MN-derived DNA within plasma cfDNA is very low. **(A)** We can estimate the proportion of plasma cfDNA derived from MN based on the number of MN dying, the proportion of released DNA which reaches plasma cfDNA and the half-life of cfDNA. **(B)** For different disease durations between one and five years we estimate the proportion of plasma cfDNA derived from MN; and **(C)** we estimate the rate of MN-death necessary to achieve a given concentration within plasma cfDNA

The proportion of DNA released from dying cells that reaches cfDNA varies dramatically, from 3% of released DNA for megakaryocytes and endothelial cells, to 0.003% for erythrocyte progenitors. Although there are >86 billion neurons in the human CNS [28], lower MN are a rare subtype of neurons, and previous work has estimated that there may be <500,000 in total [11].

Assuming optimum availability then 3% of released MN DNA will be detectable within plasma cfDNA, equal to that of megakaryocytes. If we assume all lower MN die over the course of disease, we can estimate the theoretical maximum proportion of MN DNA as a part of total plasma cfDNA as a function of the rate of disease progression (Methods, Fig. 5B). From this we can calculate

that even for the fastest theoretical disease progression rate, the plasma concentration of MN DNA would be several orders of magnitude smaller than our threshold for detection, primarily because of the small number of MN relative to other cell types. We have assumed a half life for cfDNA of 114 min [29]. In our simulation experiments we achieved a detection probability greater than chance only when the proportion of cfDNA attributed to MN was >1% (Fig. 4B-C) which determined the threshold for theoretical detection.

We sought to estimate what rate of MN death would be required to produce a detectable concentration within cfDNA. Using the proportion of DNA from cellular turnover detectable as cfDNA in the plasma from endothelial cell and erythroblasts as maximum and minimum estimates, we show that even if all lower MN died within 24 h, their contribution to cfDNA would still be below the limit of detection for WGBS (Fig. 5C). We consider this estimate of wider use to the field as it predicts whether a detectable quantity of cfDNA will be present from a known rate of cell death.

## Methods

Our pipeline is summarised in Supplementary Fig. 1 and details of all software are provided in Supplementary Table 4.

### Tissue culture and development of pure iPSC-derived MN

iPSCs were cultured in mTesR Plus media (StemCell Technologies) in Matrigel-coated 6-well plates (Corning) and maintained at 37 °C, 5% CO<sub>2</sub>. Cells were passaged when ~80% confluent using ReLeSR (StemCell Technologies), according to the manufacturer's instructions. iPSCs were differentiated into neural progenitor cells (NPCs) using a modified version of the dual SMAD inhibition protocol [10, 30]. On the day after plating (day 1), after the cells had reached ~100% confluence, the cells were washed once with PBS and the medium was replaced with neutralization medium (50% of KnockOut DMEM/F-12, 50% of Neurobasal), 0.5× N2 supplement (ThermoFisher), 1x Gibco GlutaMAX Supplement (ThermoFisher), 0.5x B-27 supplement (ThermoFisher), 50 U ml<sup>-1</sup> penicillin and 50 mg ml<sup>-1</sup> streptomycin, supplemented with SMAD inhibitors (DMH-1 2 µm; SB431542 10 µm; and CHIR99021 3 µm). This medium was changed every day for 6 days, and on day 7 the medium was replaced for neural medium supplemented with DMH-1 2 µm, SB431542 10 µm and CHIR 1 µm, All-Trans Retinoic Acid 0.1 µm (RA), and Purmorphamine 0.5 µm (PMN), the cells were kept in this medium until day 12 when it was possible to observe a uniform neuroepithelial sheet. At this point the cells were split 1:6 with Accutase (Gibco), onto matrigel substrate in the presence of 10 µm of ROCK inhibitor (Y-27632 dihydrochloride,

Tocris), giving rise to a sheet of neural progenitor cells (NPC). After 24 h of incubation the medium was changed to neural medium supplemented with RA 0.5 µm and PMN 0.1 µm, and the medium was changed every day for 6 more days. On day 19 MN progenitors were split with accutase onto to matrigel-coated plates and the medium was replaced with neural medium supplemented with RA 0.5 µm, PMN 0.1 µm, compound E 0.1 µm (Cpd E, Tocris), BDNF 10ng/mL, CNTF 10ng/mL and IGF 10ng/mL until day 28. On day 29, the media was replaced with neuronal media (neurobasal media supplemented with 1% of B27, BDNF 10ng/mL, CNTF 10ng/mL and IGF 10ng/mL). The cells were then fed on alternate days with neuronal medium until day 40.

For immunostaining to confirm the purity of MN cultures, neurons were washed with phosphate-buffered saline (PBS) and fixed with 4% paraformaldehyde for 10 min at room temperature. After fixation samples were washed three times with PBS and permeabilized with 0.3% Triton X-100 diluted in PBS for 5 min. The cells were subsequently blocked in 5% Donkey serum (Millipore) for 1 h at room temperature. After blocking, cell cultures were incubated with the appropriate primary antibodies: guinea pig anti-MAP2 1:1000 [Synaptic Systems #188004]; chicken anti-TUJ1 1:1000 [Merck Millipore #AB9354]; or mouse anti-SMI32 1:1000 [Biolegend #801701]; diluted in PBS containing 5% of DS overnight. Cells were then washed with PBS three times. Fluorescent secondary antibodies (Alexa Fluor 488, 555, 594 or 647, diluted 1:400 with DS) (ThermoFisher Scientific #A-21202, #A-21432, A-21450, #A-32744, #A-21206) were subsequently added to the cells and incubated for 1 h. The samples were washed with PBS three more times and incubated with Hoechst 33,342 (Invitrogen) for nuclear staining for 5 min. All experiments included cultures where the primary antibodies were not added; non-specific staining was not observed in these negative controls. Images were obtained using an Opera Phenix™ High Content Screening System at 40x magnification. Images were analyzed using the Harmony™ Image analysis system. 405, 488, 594, and 647 nm lasers along with the appropriate excitation and emission filters were used. These settings were kept consistent while taking images from all separate differentiations. In all cases we observed >95% purity of MN cultures.

### Whole genome bisulfite sequencing (WGBS) of DNA derived from iPSC-derived MN

We generated WGBS libraries following the Whole-Genome Bisulfite Sequencing Data Standards and Processing Pipeline (<https://www.encodeproject.org/data-standards/wgbs/>). In brief, genomic DNA was extracted from ~50,000 cells per technical replicate before shearing and bisulfite treatment. Libraries were amplified by

PCR and purified. Library concentrations were measured (Qubit). WGBS libraries were paired-end sequenced on a NovaSeq 6000 system (Illumina) with target 30X coverage. Raw data were processed with the ENCODE 4 pipeline for WGBS according to ENCODE 4 standards. Files are available at [encodeproject.org](https://encodeproject.org) and with the following accession numbers: ENCSR734EFX, ENCSR509LMK, ENCSR978LOX; data are also available at the Gene Expression Omnibus (GEO) with the following accession numbers: GSE215710, GSE215617 and GSE215648.

Paired-end FASTQ files were mapped to the human (hg38), lambda, pUC19 and viral genomes using *bwa-meth* (v.0.2.0) then converted to BAM files using *SAMtools* (v.1.9)52. Duplicated reads were marked by *Sambamba* (v.0.6.5) with parameters ‘-l 1 -t 16 --sort-buffer-size 16000 --overflow-list-size 10000000’ [31]. Reads with low mapping quality, duplicated or not mapped in a proper pair were excluded using *SAMtools* view with parameters ‘-F 1796 -q 10’. Reads were stripped from nonCpG nucleotides and converted to PAT files using *wgbstools* (v.0.2.0, downloaded from Github [github.com/nloyfer/wgbstools](https://github.com/nloyfer/wgbstools) in September 2022), command *wgbstools bam2pat --genome hg38*. Methylation across the MN samples was examined using a PCA plot, and technical replicates were found to have low heterogeneity. Technical replicates were then merged to allow inclusion in the *wgbstools* pipeline.

#### Genome segmentation into methylation blocks

Using all three of our samples and 205 samples from a methylation atlas we segmented the genome into 1,630,133 blocks of 4 or more CpGs using the *wgbstools* command ‘*wgbstools segment --min\_cpg 4 --max\_bp 5000*’. PAT and BETA files for all 207 available samples mapped to GRCh38 were downloaded from GEO (accession number GSE186458) [5] on the 20th of September 2022. As per the original publication we excluded two cardiomyocyte samples due to low coverage. We also segmented the genome into 1,938,130 blocks of 3 CpGs were identified using the *wgbstools* command *wgbstools segment --min\_cpg 3 --max\_bp 5000*; these blocks of 3 CpGs were used only for marker selection.

#### Unsupervised clustering of DNA methylation profiles

Average methylation per block (of at least 4 CpGs in size) for each sample was extracted using the *wgbstools* command ‘*beta\_to\_table*’, replacing blocks with less than 10x coverage in a sample with ‘NA’. We then selected the top 1% of blocks by variance, excluding blocks with any ‘NA’ values across all samples, and used these for clustering. Unsupervised clustering was performed using Python version 3.10.8, Dask version 2023.9.2, SciPy 1.9.1, options *method='average'*, *metric='cityblock'*, *optimal\_ordering = True*.

#### Derivation of MN-specific hypomethylated genomic regions

We applied the *wgbstools* command ‘*find\_markers*’ together with all 205 samples used for segmentation. Default parameters were used to remove low coverage regions, samples with a read depth of less than 5 in a segment had the value set to NA, and segments with greater than 1 in 3 NA values in either the target or background cell type were removed. Regions were considered MN-specific if there was a difference of at least 0.3 between the mean motor neuron methylation and mean of all other samples’ methylation within that block, and the p value of a t-test was equal to or below 0.05.

#### Identification of genes linked to MN-specific hypomethylated genomic regions

We implemented the ABC model [15] following the guidelines provided at <https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction> using ATAC-seq data from the same iPSC-derived MN as described in [32]. Firstly we called peaks for the ATAC-seq profiling using MACS2, and then identified the candidate enhancer elements using ‘*makeCandidateRegions.py*’ with parameters *peakExtendFromSummit = 250* and *nStrongestPeaks = 150,000*. The black-listed regions generated by the ENCODE 4 (<https://www.encodeproject.org/>) were used for removing enhancers overlapping regions with anomalous sequencing reads. Second, we applied ‘*run.neighborhoods.py*’ to quantify the enhancer activities by counting ATAC-seq and H3K27ac ChIP-seq reads in candidate enhancer regions. RNA-seq profiling of iPSC-derived MNs was also provided to inform expressed genes. Quantile normalisation was applied using K562 epigenetic data as the reference. At last, using ‘*predict.py*’ we computed the ABC scores by combining the enhancer activities (calculated by the second step) with the Hi-C profiling. Hi-C data was fit to the power-law model. The default threshold 0.02 was used to define valid enhancer-promoter links. Finally we overlapped enhancers with hypomethylated regions and thereby inferred a link to a target gene for that enhancer.

#### Transcriptome analysis

For AnswerALS data [33], gene expression profiling of iPSC-derived MNs and phenotype data were obtained for 245 ALS patients and 45 neurologically normal controls (<https://www.answerals.org/>). Gene expression was normalised by the trimmed mean of M-values normalisation method (TMM). We used a negative binomial test to determine genes differentially expressed between ALS patients and controls. Significance testing was performed for all genes expressed in MN ( $n = 22,976$ ) defined as count above zero in more than half of samples; in addition we excluded the bottom 25% of genes based on mean

count across all samples. All analyses were carried out in EdgeR [34] v4.0 implemented in R 4.4.1.

#### LD score regression (LDSC)

We used LDSC [35] to determine whether there was a significant enrichment of SNP-based heritability for ALS within MN-specific hypomethylated regions. Partitioned heritability was calculated following guidelines at <https://github.com/bulik/ldsc/wiki/Partitioned-Heritability> [21]. Briefly, for all SNPs found within the total set of hypomethylated regions we examined the proportion of total SNP-based heritability. Enrichment was calculated by comparing the ratio of partitioned heritability to the quantity of genetic material.

#### Generation of synthetics mixes of MN-derived DNA together with plasma cfDNA

WGBS of plasma cfDNA samples produced by Caggiano C. et al. [14] were downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164600> in February 2023, including 12 ALS patients. Raw FastQ files were trimmed with Trim Galore version 6.7 using the options ‘trim\_galore --paired -clip\_R1 4 --clip\_R2 4 --three\_prime\_clip\_R1 12 --three\_prime\_clip\_R2 12’ and then aligned to GRCh38 using the bowtie 2 aligner in Bismark version 22.3. Duplicate reads were removed with Bismark and Samtools version 1.16.1 was used to remove reads with a MAPQ score below 10. BAM files were then converted to PAT and BETA files using wgbstools.

Using wgbstools command ‘mix\_pat’, synthetic mixes of MN sample PGP\_M\_55\_iPSC (Supplementary Table 1) or cerebral neuron sample Cortex-Neuron-Z0000042F [5] and the either the 12 plasma cfDNA samples from healthy volunteers, or the 4 CSF cfDNA samples from hydrocephalus patients were created. By down- or up-sampling the cfDNA and neuronal reads, spike-ins were made at 0–10%, and coverage was varied from 2.5–30x.

#### Deconvolution of plasma cfDNA and optimisation of a deconvolution algorithm

We derived uniquely hypomethylated regions for each cell-type to use for deconvolution. In this process we excluded the two samples used for spike-in to prevent overfitting. Segmentation was repeated as before to derive two sets of regions, one with a minimum length of 3 CpGs and one with a minimum length of 4 CpGs. For both sets of regions cell type specific marker regions were found using wgbstools ‘find\_markers’ with a minimum difference between target and background means of 0.3 and a t-test p-value equal to or below 0.05. To derive different numbers of marker regions, for each cell-type the marker regions were ordered by the difference between the 75th-centile in the target group and the 2.5th centile in the background and then 25, 50, 100, 250, 300, 400, or

500 marker regions were selected. Marker regions for all cell types were then used to create an atlas of the fragment based methylation for each region across all cell types using the UXM tool downloaded from [https://github.com/nloyfer/UXM\\_deconv](https://github.com/nloyfer/UXM_deconv) on the 31st of January 2023. We then used UXM to deconvolve the synthetic mixes, producing estimated cell type contributions for each mix. These were then analysed using R version 4.3.1 (2023-06-16). To optimise region selection we tested using smaller or larger regions, and more or less regions per cell-type in order to maximise the probability of detection of spiked-in DNA, and minimise the normalised root mean squared error (RMSE).

#### Processing and deconvolution of CSF cfDNA

WGBS of CSF cfDNA samples [26] were downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142241> in April 2023, including four hydrocephalus patients. Reads were trimmed with trim-galore version 6.7 using the paired option and default settings. Due to low mapping efficiency of the reads we followed the ‘Dirty Harry’ protocol described by the creators of the Bismark software. Reads were first aligned as paired end reads using the bowtie aligner within Bismark. Unmapped R1 reads were then aligned in directional mode, and R2 reads were then aligned in pbat mode before combining them into a single file. Duplicate reads were then removed with Bismark, then Samtools version 1.16.1 was used to remove reads with a MAPQ score below 10 before converting them into PAT and Beta files using wgbstools. Deconvolution of cfDNA WGBS was performed using the UXM tool as for plasma cfDNA.

#### Theoretical estimate of the maximum of MN-derived DNA within plasma cfDNA

The concentration of cfDNA produced from cell death is given by the standard pharmacokinetic equation for concentration produced by a drug infusion at a constant rate.

$$C = d(k_0 * t_{1/2}) / (\ln(2) * Vd).$$

Where C is the concentration in the plasma,  $k_0$  is the infusion rate,  $t_{1/2}$  is the half life, Vd is the volume of distribution, and d is the proportion of DNA from cell death present in the plasma. We were able to calculate the theoretical maximum concentration of MN DNA within plasma cfDNA as a function of the time period over which the DNA was released i.e. disease duration by making reasonable assumptions for each of these values. Using the values given for a 70 kg 20–25 year old man as has historically been used as standard, the volume of plasma is 3.0 L [36]. In the absence of a ground-truth for the proportion of DNA released from dying MN that reaches plasma cfDNA, we used observed maximum and minimum proportions for other cell-types: from 3% for megakaryocytes and endothelial cells to 0.003% for

erythrocyte progenitors [27]. Infusion rate is given by the rate of cell death, and converted to weight of DNA using the conversion 1 diploid genome = 6.46pg [37]. The total number of lower MN has been estimated at ~500,000 [11] and we estimate a constant rate of loss over the disease course based on the observation that neurofilament levels, a biomarker of neuronal death, rise prior to disease onset then reach a stable concentration that is proportional to speed of progression [38]. The half life of plasma cfDNA has been measured using a variety of means, including the decrease in foetal cfDNA following pregnancy, the decrease in tumour cfDNA following surgery, and the increase and decrease in cfDNA following exercise [39]. A key point is to distinguish between the distribution half life and steady state half life. As shown by experiments with radiolabeled double stranded DNA [40], following an infusion DNA is taken up by soft tissues causing its concentration in the plasma to decrease rapidly until an equilibrium is reached with equal movement of DNA between the soft tissues and plasma. Following this the concentration of DNA will reach a steady state where its concentration is determined by the infusion rate and the steady state half life. We use 114 min as our estimate for the steady state half life as this is based on the fall in circulating tumour DNA following complete resection of the tumour [41]. cfDNA from the tumour would have reached a steady state prior to the surgery and its decrease from the surgery would be in line with the steady state half life. When estimating the proportion of cfDNA we use the concentration of 297pg/ul as the expected concentration of plasma cfDNA as this was the average concentration in controls age and sex matched to ALS patients [14].

## Discussion

ALS is currently an incurable and invariably fatal neurodegenerative disease [42]. Biomarkers are crucial for translational medicine and the recent development of serum NfL as a biomarker for ALS [1] has been key to the development of new treatments [43]. However, a key deficiency of NfL measurement is that it is not specific to MN [3], the primary degenerating cell in ALS. We and others have hypothesised that detection of cell-specific methylation of DNA within plasma cfDNA might provide an alternative and more specific biomarker for ALS. Here we show theoretically and experimentally that this goal is potentially not achievable using WGBS of plasma cfDNA, at least under the experimental conditions we encountered. Alternative approaches are needed which may include alternative biofluids or detection methods.

We have developed a MN-specific set of hypomethylated genomic regions using WGBS in iPSC-derived MN from neurologically normal individuals, together with an atlas of tissue-specific methylation [5]. We demonstrate

that these regions are associated with genes which are key to MN function but not significantly enriched with ALS genetic risk. Our regions are likely to be useful for future works aiming to detect DNA derived from MN using different detection methods. The use of iPSC-derived MN is a potential limitation, despite that fact that these cells are a gold standard model of ALS [12] and recapitulate many features found in post mortem MN. It is possible that we have excluded certain MN-specific DNA methylation which is not found in their iPSC-derived counterparts; future work which includes isolation of post mortem in sufficient numbers [11] to perform WGBS, may revise our findings. Another potential limitation is the fact that iPSCs were derived from individuals of different ages, including one individual significantly below the age of onset of ALS (Supplementary Table 1). The reason for this inclusion was to try and isolate a MN-specific rather than an aging-specific signal but this could have impaired the detection of MN death in ALS patients in an unmeasurable manner. However, this does not affect the conclusions of our simulations and therefore our central conclusions are unchanged.

Our simulations and our measurements suggest that the sensitivity of WGBS is limited to 1% of plasma cfDNA which is significantly greater than the maximum proportion of plasma cfDNA derived from rapidly degenerating MN, which we estimate theoretically to be  $1.6 \times 10^{-5}\%$ . This is due to the relatively small number of MN compared to the ongoing turnover of other cell-types. It is not inconceivable that MN-derived DNA could be detected at this level but targeted amplification together with more sensitive detection will be necessary, perhaps using customised oligonucleotide probes for selected CpG sites.

An important limitation to our work, and the majority of deconvolution algorithms, is that they assume the sequenced DNA fragments are randomly distributed across the genome, which is not correct. It is known that the formation of cfDNA from genomic DNA leads to preferential preservation of nucleosome-bound DNA, so cfDNA from different cell types or tissues produces fragmentation patterns with greater depth at sites bound to nucleosomes [44]. Enrichment of MN-specific methylation blocks used for detection with nucleosome-bound genomic regions could potentially improve the performance of detection. Alternatively WGBS protocols including post-bisulfite adaptor tagging (PBAT) can aid detection of subnucleosomal fragments [45, 46].

It is possible that use of an alternative biofluid might enable detection of MN-specific DNA. CSF is the obvious choice given that, unlike blood, it is not separated from MN by the blood brain barrier (BBB). However, the extremely low concentration of cfDNA in CSF – 0.4ng/mL versus 7.7ng/mL in plasma [47] – may again be

prohibitive. Our preliminary analysis suggests that neuronal but not MN-derived DNA is detectable within CSF cfDNA via WGBS, but this did not include sequencing data from ALS patients.

Our study has contributed WGBS data from iPSC-derived MN (encodeproject.org, Methods) and the identification of MN-specific hypomethylated genomic regions. We have not achieved a new biomarker for ALS but we have delineated the challenge for this approach through both theoretical calculations and experimental measurements. We have shown that WGBS of cfDNA derived from plasma is not likely to lead to a new biomarker for ALS and that future research should focus on developing our MN-specific regions with a more sensitive detection method. Our approach is relevant to any disease defined by the progressive loss of a specific cell type.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-025-02084-w>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

### Acknowledgements

We are very grateful to the ALS patients and control subjects who generously donated biosamples. We acknowledge transcriptomic data provided by the AnswerALS Consortium. Figures were created using BioRender.com.

### Author contributions

CH, PJS, MPS, SZ, JM, EH and JCK conceived and designed the study. CH, AN, AMB and JCK performed statistical analyses. AKW, CDSS, LF and TM carried out experiments. CH, EH, JM, SZ, JCK, KK, CC, and NZ interpreted the data with assistance from all other authors. JCK, JM, PJS, and MPS supervised the work. CH, EH and JCK wrote the manuscript with feedback from all other authors.

### Funding

This work was supported by the National Institutes of Health (CEGS 5P50HG00773504, 1P50HL083800, 1R01HL101388, 1R01-HL122939, S10OD025212, P30DK116074, and UM1HG009442 to MPS), the Wellcome Trust (216596/Z/19/Z to JCK), and NIHR (NF-SI-0617-10077 to PJS). CH/JCK are supported by the MND (899–792). We also acknowledge support from a Kingsland fellowship (TM), and the NIHR Sheffield Biomedical Research Centre for Translational Neuroscience (IS-BRC-1215-20017) and the NIHR Sheffield Clinical Research Facility.

### Data availability

MN WGBS data generated during the current study are available at the Gene Expression Omnibus (GEO) repository with the following accession numbers: GSE215710, GSE215617 and GSE215648. ALS patient plasma cfDNA WGBS data generated as part of [14] and analysed during the current study are available at the GEO repository with the following accession numbers: GSM5014683, GSM5014684, GSM5014685, GSM5014686, GSM501469, GSM5014692, GSM5014693, GSM5014694, GSM5014695, GSM5014696, GSM5014697, GSM5014698. CSF cfDNA WGBS data generated as part of [26] and analysed during the current study are available at the GEO repository with the following accession numbers: GSM4223629, GSM4223630, GSM4223631, and GSM4223632. WGBS from 207 samples performed as part of a tissue atlas [5] and analysed during the current study are available at the GEO repository

under the accession number GSE186458. Gene expression profiling of iPSC-derived MNs for ALS patients and neurologically normal controls performed as part of [33] and analysed during the current study are available at data. answers.org.

### Declarations

#### Ethics approval and consent to participate

The study was approved by the South Sheffield Research Ethics Committee. Also, this study followed study protocols approved by Medical Ethical Committees for each of the participating institutions. Written informed consent was obtained from all participating individuals. All methods were performed in accordance with relevant national and international guidelines and regulations.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 5 November 2024 / Accepted: 6 January 2025

Published online: 14 January 2025

### References

- Lu C-H, Macdonald-Wallis C, Gray E, Pearce N, Petzold A, Norgren N, et al. Neurofilament light chain: a prognostic biomarker in amyotrophic lateral sclerosis. *Neurology*. 2015;84:2247–57.
- Yuan A, Rao MV, Veeranna, Nixon RA. Neurofilaments and Neurofilament Proteins in Health and Disease. *Cold Spring Harb Perspect Biol*. 2017;9.
- Verde F, Steinacker P, Weishaupt JH, Kassubek J, Oeckl P, Halbgebauer S, et al. Neurofilament light chain in serum for the diagnosis of amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*. 2019;90:157–64.
- Davies JC, Dharmadasa T, Thompson AG, Edmond EC, Yoganathan K, Gao J, et al. Limited value of serum neurofilament light chain in diagnosing amyotrophic lateral sclerosis. *Brain Commun*. 2023;5:fcad163.
- Loyfer N, Magenheimer J, Peretz A, Cann G, Bredno J, Klochendler A, et al. A DNA methylation atlas of normal human cell types. *Nature*. 2023;613:355–64.
- Li Y, Pan X, Roberts ML, Liu P, Kotchen TA, Cowley AW Jr, et al. Stability of global methylation profiles of whole blood and extracted DNA under different storage durations and conditions. *Epigenomics*. 2018;10:797–811.
- Kustanovich A, Schwartz R, Peretz T, Grinshpun A. Life and death of circulating cell-free DNA. *Cancer Biol Ther*. 2019;20:1057–67.
- Bronkhorst AJ, Ungerer V, Holdenrieder S. The emerging role of cell-free DNA as a molecular marker for cancer management. *Biomol Detect Quantif*. 2019;17:100087.
- Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, et al. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med*. 2011;9:133.
- Zhang S, Cooper-Knock J, Weimer AK, Shi M, Moll T, Marshall JNG, et al. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron*. 2022;110:992–e100811.
- Gautier O, Blum JA, Maksymetz J, Chen D, Schweingruber C, Mei I et al. Human motor neurons are rare and can be transcriptomically divided into known subtypes. *bioRxiv*. 2023;2023.04.05.535689.
- Sances S, Brujin Li, Chandran S, Eggan K, Ho R, Klim JR, et al. Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat Neurosci*. 2016;19:542–53.
- Catanese A, Rajkumar S, Sommer D, Masrori P, Hersmus N, Van Damme P, et al. Multiomics and machine-learning identify novel transcriptional and mutational signatures in amyotrophic lateral sclerosis. *Brain*. 2023;146:3770–82.
- Caggiano C, Celona B, Garton F, Mefford J, Black BL, Henderson R, et al. Comprehensive cell type decomposition of circulating cell-free DNA with CelfIE. *Nat Commun*. 2021;12:2717.
- Stamenova EK, Aiden EL, Lander ES, Engreitz JM. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature*. 2019.

16. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583:699–710.
17. Wiench M, John S, Baek S, Johnson TA, Sung M-H, Escobar T, et al. DNA methylation status predicts cell type-specific enhancer activity. *EMBO J*. 2011;30:3028–39.
18. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun*. 2018;9:1366.
19. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*. 2021;1:e90.
20. Nizzardo M, Taiana M, Rizzo F, Aguila Benitez J, Nijssen J, Allodi I, et al. Synaptotagmin 13 is neuroprotective across motor neuron diseases. *Acta Neuropathol*. 2020;139:837–53.
21. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228–35.
22. van Rheenen W, van der Spek RAA, Bakker MK, van Vugt JJFA, Hop PJ, Zwamborn RAJ, et al. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet*. 2021;53:1636–48.
23. Chatterton Z, Mendeleev N, Chen S, Carr W, Kamimori GH, Ge Y, et al. Bisulfite Amplicon sequencing can detect Glia and Neuron Cell-Free DNA in blood plasma. *Front Mol Neurosci*. 2021;14:672614.
24. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheimer J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A*. 2016;113:E1826–34.
25. Li S, Zeng W, Ni X, Liu Q, Li W, Stackpole ML, et al. Comprehensive tissue deconvolution of cell-free DNA by deep learning for disease diagnosis and monitoring. *Proc Natl Acad Sci U S A*. 2023;120:e2305236120.
26. Li J, Zhao S, Lee M, Yin Y, Li J, Zhou Y et al. Reliable tumor detection by whole-genome methylation sequencing of cell-free DNA in cerebrospinal fluid of pediatric medulloblastoma. *Sci Adv*. 2020;6.
27. Sender R, Noor E, Milo R, Dor Y. What fraction of cellular DNA turnover becomes cfDNA? *bioRxiv*. 2023.
28. Voytek B. Are there really as many neurons in the human brain as stars in the Milky Way. *Scitable, Nature Education*.
29. Chen K, Zhao H, Yang F, Hui B, Wang T, Wang LT, et al. Dynamic changes of circulating tumour DNA in surgical lung cancer patients: protocol for a prospective observational study. *BMJ Open*. 2018;8:e019012.
30. Du Z-W, Chen H, Liu H, Lu J, Qian K, Huang C-L, et al. Generation and expansion of highly pure motor neuron progenitors from human pluripotent stem cells. *Nat Commun*. 2015;6:6626.
31. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31:2032–4.
32. Zhang S, Moll T, Rubin-Sigler J, Tu S, Li S, Yuan E et al. Deep learning modeling of rare noncoding genetic variants in human motor neurons defines CCDC146 as a therapeutic target for ALS. *medRxiv*. 2024. <https://doi.org/10.1101/2024.03.30.24305115>
33. Baxi EG, Thompson T, Li J, Kaye JA, Lim RG, Wu J, et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. *Nat Neurosci*. 2022;25:226–37.
34. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *bioRxiv*. 2024.
35. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–5.
36. ICRP. ICRP publication 89: basic anatomical and physiological data for Use in Radiological Protection: reference values. SAGE Publications Limited; 2003.
37. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. *BMC Res Notes*. 2019;12:106.
38. Benatar M, Wu J, Andersen PM, Lombardi V, Malaspina A. Neurofilament light: a candidate biomarker of presymptomatic amyotrophic lateral sclerosis and phenoconversion. *Ann Neurol*. 2018;84:130–9.
39. Khier S, Lohan L. Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature. *Future Sci OA*. 2018;4:FSO295.
40. Emlen W, Mannik M. Effect of DNA size and strandedness on the in vivo clearance and organ localization of DNA. *Clin Exp Immunol*. 1984;56:185–92.
41. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med*. 2008;14:985–90.
42. Cooper-Knock J, Jenkins T, Shaw PJ. Clinical and molecular aspects of Motor Neuron Disease. *Colloquium Ser Genomic Mol Med*. 2013;2:1–60.
43. Miller TM, Cudkowicz ME, Genge A, Shaw PJ, Sobue G, Buccelli RC, et al. Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS. *N Engl J Med*. 2022;387:1099–110.
44. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its Tissues-Of-Origin. *Cell*. 2016;164:57–68.
45. Miura F, Ito T. Post-bisulfite adaptor tagging for PCR-free whole-genome bisulfite sequencing. *Methods Mol Biol*. 2018;1708:123–36.
46. Miura F, Shibata Y, Miura M, Sangatsuda Y, Hisano O, Araki H, et al. Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res*. 2019;47:e85.
47. Wu J, Liu Z, Huang T, Wang Y, Song MM, Song T, et al. Cerebrospinal fluid circulating tumor DNA depicts profiling of brain metastasis in NSCLC. *Mol Oncol*. 2023;17:810–24.

## Publisher's note

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