IsomiR Utility in Amyotrophic Lateral Sclerosis Prognostication

Yahel Cohen[1,2#], Ilan Sinai[1,2#], Iddo Magen[1,2#], Yehuda Matan Danino[1,2], Joanne Wuu[3], Andrea Malaspina*[4], Michael Benatar*[3], Eran Hornstein[1,2]*[†]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel,
[2]Department of Molecular Neuroscience, Weizmann Institute of Science, Rehovot, Israel
[3]Department of Neurology and ALS Center, University of Miami Miller School of Medicine, Miami, FL, USA
[4]UCL Queen Square Motor Neuron Disease Center, UCL Queen Square Institute of Neurology, University College London, Queen Square, London, UK

[#] These authors contributed equally
[*] Corresponding authors
[†] Lead corresponding author

Correspondence:
Andrea Malaspina (a.malaspina@ucl.ac.uk) and Michael Benatar (mbenatar@med.miami.edu) are corresponding authors for the ABC and PGB cohorts, respectively. Eran Hornstein (eran.hornstein@weizmann.ac.il) is the corresponding author for all other facets of the work.

Summary

Background: Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by progressive motor neuron loss. IsomiRs are microRNA isoforms that arise from alternative processing or editing events during miRNA biogenesis. While isomiRs may carry distinct biological and clinical relevance, their potential as cell-free biomarkers in neurodegeneration remains largely unexplored.
Method: Here we investigated the prognostic utility of plasma isomiRs in ALS, using next-generation sequencing and two orthogonal statistical approaches.
Findings: We profiled cell-free isomiRs in 154 ALS patients from a British cohort and identified higher levels of one isomiR, let-7g-5p.t, to be associated with longer survival. This finding was independently validated in an international ALS cohort of 200 patients in both approaches. let-7g-5p.t prognostic utility was comparable to that of neurofilament light chain (NfL) or miR-181.
Conclusion: These results establish isomiRs as a novel class of blood-based biomarkers in ALS with potential to refine prognostication in clinical trials for neurodegenerative diseases.

Funding: Target ALS; Israel Science Foundation, CReATe Consortium (ISF 3497/21, 424/22). All additional funding can be found in the Acknowledgements section of this paper.

Keywords

Prognostic biomarkers, neurodegenerative diseases, Amyotrophic Lateral Sclerosis, miRNA, isomiRs, survival analysis, translational neuroscience, machine learning in biomedicine, small RNA biomarkers.

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease of the human motor neuron system. Heterogeneity of patient progression rate[1] challenges the detection of therapeutic effects in clinical trials[2]. Biofluid-based biomarkers have garnered increasing attention as tools to monitor progression[3,4]. These include urinary p75[5], troponin T2[6] and microRNAs (miRNAs)[7]. Additionally, neurofilament light chain protein (NfL)[8] is a strong predictor of the future rate of disease progression, conveying meaningful prognostic information. Accordingly, we have recently demonstrated the utility of NfL in reducing the sample size required for clinical trials in a trial-like population[9]. miRNAs are small endogenous RNAs that contribute to post-transcriptional silencing of mRNA[10]. miRNAs are critical for motor neuron viability[11] and their biogenesis is known to be dysregulated in ALS[12-16]. Previous studies have shown that cell-free miRNAs and in particular miR-181 might have prognostic value[7,17]. miRNA biogenesis is a multistep process that starts with the transcription of a primary miRNA, which is cleaved by the nuclear microprocessor complex to yield an intermediate double-stranded hairpin precursor (pre-miRNA)[18]. After export of the pre-miRNA from nucleus to cytoplasm, it is cut by Dicer, creating a single strand 22nt guide that is loaded on to the RNA-induced silencing complex (RISC)[19]. IsomiRs display sequence variations that distinguish them from the reference sequence miRNAs (reference miRNA), including 3' or 5' sequence changes, or sequence polymorphism[20-22]. IsomiR production is often the result of aberrant miRNA biogenesis and isomiR levels in tissues or biofluids are often low and heterogeneous[23].

Mutations in the genes encoding for a handful of RNA binding proteins were found to be linked to increased risk of developing ALS, including Fused in Sarcoma (FUS)[24,25], and TAR DNA-binding protein 43 (TDP-43)[26]. Cytoplasmic mislocalization of TDP-43 into inclusions and loss of its nuclear function are reported in approximately 97% of patients[27]. Accordingly, cryptic exon splicing impacts dozens of mRNAs[28-30], some of which may have value as novel biomarkers relevant to ALS therapy development[31,32]. Interestingly, loss of TDP-43 or FUS also leads to dysregulation in isomiR profiles or miRNA biogenesis[33,34]. Furthermore, miRNA biogenesis proteins, Drosha and Dicer, were suggested to be involved in the disease[11,12,35]. Taken together, changes in isomiRs may be directly or indirectly related to ALS. The potential of cell-free isomiRs as biomarkers has been suggested in cancer[36-40] and in a recent work of Loher et al.[41]. In this work the authors provided profiles and performed analysis of transfer RNA-derived fragments (tRFs), ribosomal RNA-derived fragments (rRFs), Y RNA-derived fragments, and isomiR alterations in ALS plasma of a single cohort. Loher et al. found dozens of small non coding RNAs with diagnostic and prognostic potential. Here, we test the

potential value of isomiRs for ALS prognostication. We profiled cell-free isomiRs in a British cohort of 154 patients with ALS. Using Cox proportional hazard regression[42-44] and orthogonally machine learning (ML) survival prediction[42,44-47], we identified 18 isomiR predictors, which were successfully replicated in an independent international cohort of 200 patients. We further demonstrate that isomiR-let-7g-5p.t particularly carries prognostic value, with potential to complement established clinical features and previously described biomarkers, such as miR-181 and NfL. These findings establish isomiR as a novel family of biomarkers for neurodegeneration and highlight their potential as reliable predictors of ALS survival.

Results

Study population and baseline characteristics

We conducted our study in two independent cohorts of patients with ALS. Clinical data and blood samples were obtained for 154 patients from the ALS biomarker cohort (UK)[48], and for 200 patients from the Phenotype Genotype Biomarker (PGB) study[49]. Clinical and demographic cohort characteristics are summarized in **Table 1.** We found the discovery and replication cohorts to be clinically comparable. DeltaFRS (Mann-Whitney P-value = 0.18, **Figure 1A**), baseline ALSFRS-R score (Mann-Whitney P-value = 0.24, **Figure 1B**), site of symptom onset (χ^2 goodness of fit P-value = 0.08, **Figure 1C**), Riluzole treatment (χ^2 goodness of fit P-value = 0.34m, **Figure 1D**), and C9orf72 repeat expansion status (χ^2 goodness of fit P-value = 0.6). Imbalances in patients' sex distribution (χ^2 goodness of fit P-value = 0.009, **Figure 1E**) and age were noticed (t-test P-value ≤ 0.001, **Figure 1F**). The number of censored events were also statistically different between the cohorts (9% vs 54% in UK and PGB, respectively, χ^2 goodness of fit P-value ≤ 0.001, **Figure 1G**).

Our study explores isomiRs, which are sequence isoforms of miRNAs that arise from differences in processing, editing, or modifications of miRNA precursors (**Table S1** for nomenclature; **Figure 2A**), as ALS prognostic biomarkers. Here we use the term "reference miRNA" to refer to the most abundantly expressed (predominant) sequence variant among all isomiRs derived from a given miRNA gene. Noteworthy, the literature refers to a "miRNA" as the collective set of all dozens of annotated isomiRs[20].

Sequencing and data preprocessing were conducted in parallel across 154 and 200 patient samples from the UK and PGB cohorts, respectively. Following noise filtration, 1,480 isomiRs from

287 miRNA genes were co-identified in both the discovery and replication cohorts. The discovery isomiRs study was performed on the UK cohort. The PGB was held out as an unseen, independent replication cohort.

isomiR-let-7g-5p.t levels differentiate between short and long survival

To identify potential isomiRs associated with patient survival, we dichotomized the 1480 isomiRs that were shared between cohorts, based on their median abundance in the UK discovery cohort. A univariate Cox regression—used to assess the association between isomiR expression and time to death—revealed 290 statistically significant isomiRs (P-value ≤ 0.05), of which 16 remained significant after correction for multiple comparisons (FDR ≤ 0.05, **Figure 2B**). Levels above median of 15 isomiRs were associated with decreased rate of mortality in patients (univariate Cox, Hazard ratios (HR) < 1, 95% confidence interval (CI) = 0.34–0.77, FDR ≤ 0.05; **Figure 2C**), and levels above median of one isomiR were associated with increased rate of mortality (univariate Cox, HR = 1.8, 95% (CI) = 1.32–2.6, FDR ≤ 0.05; **Figure 2C**). Next, we tested these associations in the PGB replication cohort by dichotomizing isomiR levels by the threshold obtained in the UK discovery cohort. Only levels above median (≥246 unique molecular identifier (UMI) counts) of isomiR-let-7g-5p.t were found to be associated with decreased rate of mortality in the PGB replication cohort (univariate Cox, HR = 0.6, 95% CI = 0.39–0.91, P-value = 0.015; **Figure 2D**). In addition, levels above median of isomiR-let-7g-5p.t were significantly associated with longer survival in the UK discovery cohort (log rank, χ^2 = 18.7, P-value ≤ 0.001, median survival difference = 8.2 months; **Figure 2E**), and in the PGB replication cohort (log rank, χ^2 = 5.99, P-value ≤ 0.01, median survival difference = 9.1 months; **Figure 2F**). Mean isomiR-let-7g-5p.t levels were significantly different in the UK discovery cohort sub cohorts, defined by plasma isomiR-let-7g-5p.t levels that are either higher or lower than the median isomiR-let-7g-5p.t plasma value (351 vs. 176 UMIs in the above- and below-median, Mann-Whitney P-value ≤ 0.001,**Figure S1A**). We found a similar difference in the PGB replication cohort sub cohorts (324 vs. 198 UMIs, Mann-Whitney P-value ≤ 0.001, **Figure S1A**).

Post division by median levels, the sub-cohorts were clinically comparable within each cohort. Age at symptom onset (Mann-Whitney P-value ≥ 0.1, **Figure S1B**), Age at enrollment (Mann-Whitney P-value ≥ 0.08, **Figure S1C**), baseline ALSFRS-R score (Mann-Whitney P-value ≥ 0.27, **Figure S1D**), DeltaFRS (Mann-Whitney P-value ≥ 0.13, **Figure S1E**), sex distribution (χ2 goodness of fit P-value ≥ 0.1, **Figure S1F**), site of symptom onset (χ2 goodness of fit P-value ≥ 0.2, **Figure S1G**), Riluzole treatment (χ2 goodness of fit P-value ≥ 0.15, **Figure S1H**), and C9orf72 repeat expansion

status (7 vs 5 and 8 vs 12 carriers in above/below median groups, respectively, in each cohort, $\chi 2$ goodness of fit P-value ≥ 0.6). Therefore, isomiR-let-7g-5p.t levels stratify patients independently of key clinical variables.

We sought to compare the mortality rate associated with isomiRs and the respective miRNAs, which reflect the aggregate of dozens of expressed isomiRs assigned to a miRNA gene. While isomiR-let-7g-5p.t levels were associated with decreased mortality rate, no cognate miRNA was significantly associated with mortality rate in both cohorts, including let-7g-5p (univariate Cox, HR = 0.87, 1.1; 95% CI = 0.63–1.2, 0.65-1.8; P-value > 0.05 in both, respectively, **Table S2**). Accordingly, plasma levels of cognate miRNA let-7g-5p did not differentiate between short and long survival in the UK discovery or the PGB replication cohorts (log rank χ^2 = 0.63, 0.09; P-value > 0.05 in both, median survival difference = 0.3, 3.8 months, respectively; **Figure 2G, H**).

IsomiR-let-7g-5p.t is the best prognostication isomiR in two independent ALS cohorts

Although isomiR-let-7g-5p.t emerged as a promising candidate for ALS prognostication, the high dimensionality of our dataset motivated us to explore the predictive performance of isomiR combinations and their potential non-linear interactions using ML, as employed in our previous studies[50]. We sampled at random subsets of 20 isomiRs, with repetition, using bootstrapping, so each isomiR within the pool of 1480 would be included in at least 50 distinct subsets. To mitigate the risk of false positive discoveries inherent in large-scale analyses, the 154 patients of the UK discovery cohort were divided randomly into 10 bins, and a Sequential forward selection (SFS) with 10-fold cross-validation (CV) technique was performed for each of the isomiR subset[51,52]. 84 isomiRs were selected within 90% of the subsets in which they were included after SFS (≥45 out of 50, with Cox regression[42], random survival forest (RSF)[44], gradient survival boosting (XGB)[43]**, Figure S2A**). Of these three orthogonal survival models, XGB demonstrated the highest mean concordance index score (C-index) across the 84 isomiRs (**Figure S2B**).

Next, we modeled survival of the 84 isomiRs by XGB. IsomiRs were included if their subsets displayed a C-index higher than a threshold that was defined empirically by the elbow method[53] (> C-index of 0.74). Reassuringly, isomiR-let-7g-5p.t was detected, along with 17 other isomiRs: isomiR-let-7e-5p.G9C, isomiR-let-7f-5p.T16C.t, isomiR-let-7i-5p.G5A, isomiR-130b-5p.T, isomiR-142-3p.a, isomiR-451a.T.A, isomiR-144-3p, isomiR-146a-5p.T22A, isomiR-151a-3p.A9G, isomiR-25-3p.A22T,

isomiR-30d-5p.A20G.CT, isomiR-425-5p.as, isomiR-432-5p, isomiR-486-5p.C, isomiR-486-5p.T9G.T, isomiR-92a-3p.T1C, isomiR-92a-3p.T22C (**Figure S2C**).

After hyperparameter tuning with 5-fold CV (Supplementary Methods), a model based on these 18 isomiRs demonstrated strong prognostic capacity across patient survival duration of 84 months (time-dependent area under the curve (AUC(t)) values of 0.7 to 0.95; average 0.78; C-index = 0.7; survival duration 2 - 85 months; **Figure 3A**), but survival prediction was less effective in the PGB replication cohort (AUC(t) 0.49- 0.9, average 0.62; C-index = 0.55; survival duration 3–54 months; **Figure 3A**). Notably, prediction was better for patients with longer survival times. Additionally, miRNA and reference miRNA displayed poor performance over time (**Figure 3B, C**).

We demonstrate higher performance with the model trained on the 18 selected isomiRs relative to models trained on the collective set of all annotated isomiRs (the ”miRNA”, average AUC(t) = 0.64/0.53, C-index= 0.59/0.52 in the UK discovery and PGB replication cohorts, respectively) or the most abundant isomiR (“reference miRNAs” average AUC(t)= 0.68/0.46, and C-index = 0.63/0.47) **Figure 3D, E**).

To identify the isomiRs most valuable for survival prediction, we calculated Shapley Additive Explanations (SHAP) for the model average risk function. High levels of isomiR-let-7g-5p.t were associated with reduced risk of death and longer survival (**Figure 3F**). In contrast, high levels of isomiR-92a-3p.T22C, isomiR-92a-3p.T1C, isomiR-let-7f-5p.T16C.t, and isomiR-let-7i-5p.G5A associate with higher risk and shorter survival. A similar SHAP analysis of the PGB replication cohort was highly correlated, highlighting the model's robustness (**Figure 3G**). We then derived the HRs from SHAP values, as previously described[54]. IsomiR-92a-3p.T22C and isomiR-let-7g-5p.t levels were associated the most with increased / decreased mortality, respectively (**Figure 3H**).

As isomiR expression can vary across cohorts, we assessed abundance and inter-cohort correlation to evaluate biological reproducibility. Among the 18 selected isomiRs, mean expression ranged from 4–9 UMIs (UK cohort), and 5–14 UMIs (PGB cohort) and the detection rates were comparable between cohorts (**Figure S3A, B**). IsomiR-142-3p.a had the highest expression (UK = 974 ± 833 UMIs, PGB = 1984 ± 945 UMIs), followed by isomiR-let-7g-5p.t (UK discovery cohort: 264 UMIs ± 129; PGB replication cohort: 259 UMIs ± 94). IsomiR abundance was well correlated between cohorts ($r = 0.74$, $P = 8.1e{-}28$, **Figure S3C, D**) but no correlation was noted between the abundance of isomiRs and their “cognate miRNA”, nor with the “reference miRNA” (**Table S3**).

Lower isomiR-let-7g-5p.t levels in non-neurodegeneration control blood relative to ALS

Having found that isomiR-let-7g-5p.t holds prognostic value for the disease in two orthogonal approaches, we compared its levels between ALS patients and controls. To do so, we reanalyzed plasma samples from two independent control cohorts from Queen Mary hospital (QMUL - CTL) and University College London (UCL - CTL), respectively. isomiR-let-7g-5p.t levels were comparable across the two ALS cohorts within the same sub cohort (mean difference ≤ 27 UMI, Tukey HSD P-value ≥ 0.99, **Figure S4**), and although variable across control cohorts (mean difference = 324, Tukey HSD P-value ≤ 0.001), we observed significantly lower levels in the below-median ALS sub cohorts compared to both controls and the above-median ALS sub cohorts (mean difference ≥ 369 UMI and mean difference ≥ 522 UMI, respectively, Tukey HSD P-value ≤ 0.001, **Figure S4**). Therefore, isomiR-let-7g-5p.t is also of potential diagnostic value.

Lastly, after examining isomiR-let-7g-5p.t biomarker potential, we tested if there was a change in the isomiR's predicted targets compared to its reference miRNA. We retrieved the predicted targets of miRNA let-7g-5p from miRCarta[55], an online repository, and subjected them to a miRNA-mRNA hybridization prediction algorithm[56], compared to the reference miRNA let-7g-5p and isomiR-let-7g-5p.t. Out of potential ~4000 mRNA targets, 3189 had at least one significant hybridization site with either the reference miRNA let-7g-5p or isomiR-let-7g-5p.t. However, 358 out of 3189 mRNA targets were predicted to lose at least one hybridization site, and 43 were predicted to gain at least one such isomiR-let-7g-5p.t site (Wilcoxon p-value = 2.4e-47, rank-biserial correlation effect size = 0.67). Enrichment for targets associated with ALS was not noticed.

Prognostic Value of Plasma isomiR-let-7g-5p.t in ALS Is Comparable to NfL

As we previously reported[17,57], high (above median) blood NfL levels were significantly associated with shorter survival in the UK discovery cohort (log-rank, χ^2 = 9.4, P-value = 2e-3, median survival difference = 7.1 months; **Figure 4A**), and this was consistent in the PGB replication cohort, when NfL dichotomized according to the median plasma value in the UK cohort (log rank, χ^2 = 17.7, P-value = 2.7e-5, median survival difference = 13.7 months; **Figure 4B**). Above-median NfL levels were associated with an increased mortality rate in both cohorts (UK discovery cohort: univariate Cox, HR = 1.7, 95% CI = 1.2-2.3, P-value = 2.4e-4, C-index = 0.56 ; PGB replication cohort

Cox HR = 2.4, 95% CI 1.6-3.7, P-value = 4.5e-6, C-index = 0.63; **Figure 4C, D**). In contrast, higher levels of seven (out of eighteen selected by ML, **Figure S3C**) isomiRs were associated with a decreased mortality rate (univariate Cox, HR < 0.72, 95% CI = 0.37-0.99, P-value ≤ 0.05, C-index > 0.54; **Figure 4C**). However, only isomiR-let-7g-5p.t retained significant HR in univariate COX regression in the PGB replication cohort (UK discovery cohort: HR = 0.58, 95% CI = 0.41-0.81, P-value = 1.4e-3, C-index = 0.57; PGB replication cohort: univariate Cox, HR = 0.6, 95% CI = 0.4-0.9, P-value = 0.015, C-index = 0.55; **Figure 4C, D**). Accordingly, isomiR-let-7g-5p.t was the only isomiR that significantly distinguishes short and long survival times in both cohorts using the log-rank test (**Figures S5A, B**).

In multivariate Cox analysis, NfL remained significantly associated with increased mortality rate in both cohorts (UK discovery HR = 1.7, 95% CI = 1.2-2, P-value = 6e-4; PGB replication HR = 3.3, 95% CI = 2.4-5.3, P-value = 2e-7), whereas isomiR-let-7g-5p.t was significant only in the PGB cohort (multivariate Cox, HR = 0.52, 95% CI = 0.32-0.86, P-value = 0.01; **Figure S6A, B**). Noteworthy, isomiRs modestly improved the prognostic capacity, compared to a model which only includes NfL, in both cohorts (C-index = 0.67/0.72 in UK discovery or PGB replication, respectively, compared to 0.56/0.63 for NfL alone).

When considering only isomiR-let-7g-5p.t and NfL together, isomiR-let-7g-5p.t showed a stronger association with risk of mortality than NfL in the UK discovery cohort (HR of NfL in a multivariate Cox model = 1.64, 95% CI = 1.2-2.3, P-value = 3.9e-4; HR of isomiR-let-7g-5p.t in a multivariate Cox model = 0.49, 95% CI = 0.35-0.69, P-value = 3.8e-6; **Figure S6C**). Together, NfL and isomiR-let-7g-5p.t yield better prognostication than either marker alone (multivariate Cox C-index = 0.63 vs. C-index = 0.56 NfL alone or C-index = 0.59 isomiR-let-7g-5p.t alone). In the PGB cohort, NfL had a stronger association with mortality rate (NfL multivariate Cox, HR = 3, 95% CI = 1.9-4.6, P-value = 1e-7; isomiR-let-7g-5p.t multivariate Cox, HR = 0.46, 95% CI = 0.29-0.7, P-value = 4.8e-5; **Figure S6D**). But again, both markers together improved survival prediction compared to either one individually (multivariate Cox C-index = 0.67, NfL alone C-index = 0.63, isomiR-let-7g-5p.t alone C-index = 0.56). Moreover, when combined with relevant clinical features (baseline ALSFRS-R, site of symptom onset, age at onset), only isomiR-let-7g-5p.t (UK discovery multivariate Cox, HR = 0.53, 95% CI = 0.38-0.75, P-value = 3e-5; PGB replication multivariate Cox, HR = 0.52, 95% CI = 0.33-0.83, P-value =6.5e-3) and baseline ALSFRS-R (UK discovery multivariate Cox, HR = 0.97, 95% CI = 0.94-0.99, P-value = 2.6e-3; PGB replication multivariate Cox, HR = 0.96, 95% CI = 0.92-0.99, P-value = 0.013) were significantly associated with mortality risk in both cohorts (**Figure S6E, F**). In

contrast, NfL levels association with mortality were only approaching significance in the UK discovery cohort (multivariate Cox, HR = 1.4, 95% CI = 0.97-1.94, P-value = 0.08; **Figure S6E**) while being significantly associated with mortality in the PGB cohort, with an effect size larger than that of isomiR-let-7g-5p.t (multivariate Cox, HR = 2.5, 95% CI = 1.5-4, P-value ≤ 0.001; **Figure S6F**).

Since above-median levels of isomiR-let-7g-5p.t and NfL were associated with opposing effects on mortality risk, we next assessed their joint prognostic capacity by calculating the ratio between isomiR-let-7g-5p.t and NfL levels. High values of this ratio slightly improved distinction between short and long survival compared to NfL alone (UK discovery log-rank, χ^2 = 11.8, P-value = 5.9e-4, median survival difference = 8 months; PGB replication log-rank, χ^2 = 19.8, P-value = 8.8e-6, median survival difference = 14 months; **Figures 4E, F**). Notably, the ratio showed a stronger inverse association with mortality risk compared to either marker alone, and improved overall prognosis (UK discovery univariate Cox, HR = 0.56, 95% CI = 0.4-0.78, P-value = 6.2e-4, C-index = 0.58; PGB replication univariate Cox, HR = 0.4, 95% CI = 0.26-0.6, P-value = 1.8e-5, C-index =0.64 ;**Figure S6G**). Taken together, these results suggest that isomiR-let-7g-5p.t performs comparably to NfL, and that combining isomiR-let-7g-5p.t with NfL improved predictive performance, at least as reflected by the C-index. These findings highlight the value of isomiR-let-7g-5p.t as a potential prognostic biomarker in ALS.

Prognostic Value of Plasma isomiR-let-7g-5p.t in ALS Is Comparable to miR-181

Our UK cohort was a subset of the cohort where we originally discovered miR-181, a multiplication of sister miRNAs miR-181a-5p and miR-181b-5p, as a prognostic marker[17]. A univariate Cox analysis revealed that high miR-181 levels were associated with a >2.1-fold increased risk of mortality, again demonstrating that isomiR-let-7g-5p.t has an equivalent prognostic capacity to previously published markers (UK discovery cohort HR = 2.1, 95% CI = 1.4–3.2, P-value ≤ 0.001, C-index = 0.58; PGB replication cohort HR = 3.0, 95% CI = 1.7–5.4, P-value ≤ 0.001, C-index = 0.55; **Figure 4C, D**).

miR-181 levels distinguished between short and long survivors in the UK cohort (log rank on dichotomized miR-181 levels: χ^2 = 13.7, P-value = 2.1e-4, median survival difference = 7.2 months; **Figure 5A**), and also in the PGB cohort (log rank: χ^2 = 15.1, P-value = 1e-4, median survival difference = 9.7 months; **Figure 5B**). Notably, isomiR-let-7g-5p.t was the only isomiR to remain statistically significant in both cohorts, in a combined analysis of miR-181 and the 18 isomiR chosen

by ML (UK discovery multivariate Cox, HR = 0.63, 95% CI = 0.4-0.98, P-value = 0.04; PGB replication multivariate Cox, HR = 0.57, 95% CI = 0.36-0.9, P-value = 0.017; **Figure 5C, D**). To assess joint prognostic capacity of isomiR-let-7g-5p along with miR-181, we tested isomiR-let-7g-5p / miR-181 ratio capacity to distinguish between short and long survival. We found a significant association of the ratio levels to patient survival in both cohorts (UK discovery log rank: χ^2 = 20.6, P-value = 5.6e-6, median survival difference = 8 months; PGB replication log rank: χ^2 = 7, P-value = 8e-3, median survival difference = 6 months; **Figures 5E, F**). Moreover, the HR of isomiR-let-7g-5p / miR-181 ratio was 0.38 in the discovery and 0.44 in the replication cohort (UK discovery univariate Cox, 95% CI = 0.19–0.77, C-index = 0.58, P-value = 8e-6; PGB replication univariate Cox, 95% CI = 0.18-0.8, C-index = 0.53, P-value = 0.01, **Figure S6G**). Taken together, these results highlight the potential of isomiR-let-7g-5p.t as a robust biomarker for ALS survival.

Discussion

IsomiRs display small variations of either sequence insertion, deletion, or single nucleotide polymorphism that distinguish them from the reference miRNA form. While miRNAs are extensively studied in ALS[12,14-17,58-66], we are first to demonstrate the value of isomiRs as biomarkers for neurodegeneration in two large, independent cohorts. Among 1480 studied isomiRs, isomiR-let-7g-5p.t emerges as the most promising prognostic biomarker. High plasma levels of isomiR-let-7g-5p.t are significantly associated with lower mortality rate and longer survival. This effect was not explained by clinical characteristics, suggesting that the information encompassed in isomiR-let-7g-5p.t is not directly associated with a single clinical feature. IsomiR-let-7g-5p.t levels were reduced in two ALS cohorts relative to two control cohorts, suggesting diagnostic potential. Higher plasma levels of isomiR-let-7g-5p.t are comparable but reciprocal to miR-181 and NfL, which we previously associated with ALS prognosis[7,66,67]. Additionally, the incorporation of isomiR-let-7g-5p.t to either miR-181 or NfL enhances their prognostic value.

Dysregulation of miRNAs in ALS has previously been demonstrated in mice and humans[11,15,16]. Numerous studies have proposed miRNA utility as potential biomarkers for prognosis for ALS [7,59,63]. However, isomiRs have been overlooked in part due to their sequence heterogeneity and their low abundance[23]. Nevertheless, our work, the work of Loher et al.[41], and notable dysregulation of RNA-binding proteins such as TDP-43, DICER, DROSHA, or FUS in ALS[11,12,34,62,68-70] should prompt additional studies of isomiRs in ALS. In addition, it is possible that isomiRs play a role

in other neurodegenerative disorders, with implications ranging from basic research to clinical applications.

Our analysis found that isomiR-let-7g-5p.t accounts for approximately 40% of the reads of let-7g-5p in the plasma and perhaps predominates the let-7g-5p repertoire. In this context, microRNA let-7g-5p has previously been reported to be lowered in abundance in the blood of patients with sporadic ALS (sALS)[63], and in patients with frontotemporal dementia (FTD) due to a *C9orf72* repeat expansion, compared to healthy controls[69]. Loher et al[41] reported hundreds of differentially abundant isomiRs in ALS patient plasma compared to controls, with roughly equal numbers up- and down-regulated. 53 out of 56 Let-7 family isomiRs were elevated in ALS. The association of reduced plasma isomiR-let-7g-5p.t levels to severe disease, observed in our study, suggests a distinct behavior not captured by Loher et al[41]. Let-7g-5p blood levels are also decreased in individuals with Alzheimer's disease (AD) or mild cognitive impairment (MCI)[71-78], and correlate with AD severity[71]. In addition, let-7g-5p was found to be downregulated in brain samples of Huntington patients compared to controls[79]. This could reflect a specific decrease that is mediated, at least in part, via isomiR-let-7g-5p.t.

Members of the let-7 miRNA family have been repeatedly linked to neurodegenerative diseases. In ALS, let-7d and let-7f are reduced in patient plasma, while let-7b is elevated in leukocytes. In AD, several let-7 members, including let-7b, let-7d, let-7e, and let-7g, are upregulated in patient biofluids or neuronal vesicles, with let-7b also correlating with disease progression. In Parkinson's disease, let-7a and let-7f are decreased in plasma, whereas let-7g-3p is increased in CSF-derived vesicles[65,80-90]. Together, these findings underscore the relevance of the let-7 family in neurodegeneration and reinforce the significance of isomiR-let-7g-5p.t as a prognostic marker in ALS.

Seventeen other isomiRs displayed potential relevance in our analysis, though to a weaker extent. Some of the miRNAs associated with these isomiRs (miR-451, miR-144-3p, and miR-486-5p) are linked to oxidative stress and inflammation[91-99], which potentially play a role in neurodegeneration. However, because we analyze the isomiRs in plasma, the tissue expression origin is unclear.

The mechanisms underlying isomiR levels, including isomiR-let-7g-5p.t, and their functional role in disease progression is yet to be fully elucidated. The loss of a single nucleotide at the 3′ end occurs outside the seed region and has limited impact on target specificity[100-103]. Accordingly, using

RNA22 v2[56], we found that a subset of predicted targets interacts differently with isomiR-let-7g-5p.t relative to let-7g-5p, but no enrichment was observed for ALS-related pathways.

Together, our results highlight the potential of isomiR-let-7g-5p.t as a prognostic biomarker in ALS, potentially offering comparable value to clinical features, miR-181, and NfL. Moreover, isomiR-let-7g-5p.t enhances effective prognostication in conjunction with these markers. These findings warrant further exploration of the utility of isomiR-let-7g-5p.t and other isomiRs in neurodegenerative diseases.

Limitations of Study

Despite the promising prognostic potential of isomiRs in ALS, several limitations must be noted: some uncertainty remains about the reference miRNA due to high sequence similarity within the broad family of Let-7 miRNAs. Future advances in sequencing technologies and alignment algorithms may improve isomiR resolution and assignment. Larger and more diverse population cohorts would strengthen the generalizability of the findings and help assess isomiRs' robustness across different ALS subtypes and disease stages. Last, given the availability and strong prognostic performance of NfL, the practical utility of isomiRs as standalone biomarkers may be limited until an effective way to measure them is developed. These limitations highlight the need for continued research on isomiRs in ALS.

Acknowledgments

Eran Hornstein is the Mondry Family Professorial Chair and Head of the Nella and Leon Benoziyo Centre for Neurological Diseases and of the Andi and Larry Wolfe Centre for Neuroimmunology and Neuromodulation. We acknowledge patients for their contributions. We thank ALS biomarker and the Phenotype-Genotype biomarker studies co-workers and their contributions to the biobanking project, the Crown Genomics Team of Nancy and Stephen Grand Israel National Center for Personalized Medicine for small RNA sequencing. The authors thank Dr. Isidore Rigoutsos and Phillipe Loher from Thomas Jefferson University for their assistance with the revised manuscript. The authors thank participants in the CReATe Consortium's Phenotype-Genotype-Biomarker (PGB1) study; research staff at each clinical site; and the CReATe Consortium project management and data management teams, genomics sub-core, and biorepository. This study was funded by several programs. Eran Hornstein lab is funded by the Andi and Larry Wolfe Centre for Neuroimmunology and Neuromodulation; US Department of Defense (DoD) Congressionally Directed Medical Research

Programs (CDMRP) grant AL240063; The Binational Science Foundation (BSF); Association Française Contre les Myopathies (AFM) grants 24882, 28680; Muscular Dystrophy Association (MDA) grant 1280000; Target ALS; Israel Science Foundation (ISF 3497/21, 424/22); ALS Canada; Minna-James-Heineman Stiftung through, Minerva Foundation, with funding from the Federal German Ministry for Education and Research; Robert Packard Center for ALS Research at Johns Hopkins; McGill University; EU - ERA-Net; Radala Foundation for ALS Research; Additional support generously provided by the Kekst Family Institute for Medical Genetics. Weizmann SABRA - Yeda-Sela - WRC Program, the Estate of Emile Mimran, and The Maurice and Vivienne Wohl Biology Endowment. Nella and Leon Benoziyo Center for Neurological Diseases, Goldhirsh-Yellin Foundation. Dr. Sydney Brenner and friends, Weizmann - Center for Research on Neurodegeneration. Redhill Foundation – Sam and Jean Rothberg Charitable Trust Dr. Dvora and Haim Teitelbaum Endowment Fund and Weizmann – Institute for Artificial Intelligence - Acceleration Grant. The CReATe Consortium (U54NS092091) is part of the NIH Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), NCATS. This Consortium is funded through a collaboration between NCATS and the NINDS. This work was also supported by a Clinical Trial Readiness grant (U01NS107027) from NINDS and by a grant from the ALS Association to support the CReATe Biorepository (grant ID 16-TACL-242).

Author contributions

E.H., Y.C. and I.M. oversaw the conceptualization of this study, M.B., A.M. and J.W. recruited participants and collected clinical data and biofluid samples. E.H., Y.C., I.M. & I.S. designed the study and had unrestricted access to all data. Y.C. and I.M. oversaw molecular data collection. Y.C., I.M. and I.S. performed, and replicated the statistical analyses and E.H. supervised them. Y.C., I.M, Y.M.D, M.B, A.M. and E.H participated in drafting the initial manuscript draft and E.H. Y.C, I.M, Y.M.D reviewed and revised the draft. All authors agreed to submit the manuscript, read and approved the final draft, and take full responsibility for its content, including the accuracy of the data and the fidelity of the trial to the registered protocol and its statistical analysis.

Declaration of interests

The authors declare no competing interests

Figures titles and legends

Figure 1 – Comparison of the UK discovery and PGB replication cohort metadata. Boxplot of **(A)** DeltaFRS, measured as the reduction in the ALSFRS-R score per month or **(B)** baseline ALSFRS-R score at study enrollment. Pie charts depicting **(C)** SSO - site of symptom onset **(D)** Riluzole treatment, or **(E)** sex for the cohorts. **(F)** Boxplots of age at study enrollment and age at symptom onset. **(G)** Pie charts showing the number of censored and deceased events for the UK discovery (left) and the PGB replication (right) cohorts. *** P-value ≤ 0.001, two-sided Student's t-test.

Figure 2 – High plasma isomiR-let-7g-5p.t levels predict longer survival in patients with ALS. (A) miRNA biogenesis yields several forms of miRNA isoforms (isomiRs). Following transcription, the primary miRNA is cleaved by the nuclear microprocessor, exported as a precursor to the cytoplasm, and further processed by Dicer into a ~22-nt guide RNA loaded onto RISC for target mRNA binding and silencing. Variability in biogenesis, such as misplacement of processing factors, leads to the generation of isomiRs. **(B)** Volcano plot of 1480 plasma isomiRs univariate Cox Hazard ratios (HR, x axis) and their significance (y axis). Red line denotes significance threshold (P-value = 0.05). Only 16 isomiRs remained significant post multiple comparison correction (black line denotes FDR = 5.2e-4). Forest plots of univariate Cox proportional hazard ratios of these 16 significant isomiRs in UK discovery **(C)** or PGB replication **(D)** cohorts (HR ≤ 0.6 or HR = 1.8, P-value ≤ 0.015). Only isomiR-let-7g-5p.t remained significant in both cohorts (denoted by green text). The Cumulative survival (Kaplan–Meier) curves for isomiR-let-7g-5p.t in the **(E)** UK discovery or **(F)** PGB replication cohorts. (UK: 77/77 patients in sub- vs. supra-threshold cohorts. PGB: 104/96 patients in sub- vs. supra-threshold cohorts). Kaplan–Meier curves for the mature miRNA counterpart, let-7g-5p, in the **(G)** UK discovery or **(H)** PGB replication cohorts. (UK: 77/77 patients in sub- vs. supra-threshold sub cohorts. PGB: 40/160 patients in sub- vs. supra-threshold sub cohorts subthreshold using the same threshold for dichotomization.)

Figure 3 – IsomiR based survival prediction in two independent cohorts. Plots of AUC (y axis) in the UK discovery (blue) or PGB replication (purple) cohorts for patient survival, as a function of duration from enrolment (x axis) for isomiR **(A)**, cognate miRNA **(B)**, or reference miRNA **(C)**. Bar graphs of **(D)** average AUC across time (47/42 time points for patients from the UK discovery or PGB replication, respectively) **(E)** or C-index for a model based on 18 selected isomiRs. SHAP analysis the contribution of 18 isomiRs to risk prediction in the **(F)** UK discovery or **(G)** PGB replication cohorts.

Range of isomiR levels (low–blue to high–red). Positive or negative SHAP values reveal higher or lower effect size of isomiR on the mortality risk (x axis), respectively. **(H)** Plot of HR, on the basis of XGB SHAP values calculated according to Sundrani & Lu[64].

Figure 4 – isomiR-let-7g-5p.t-based prognostication compared to NfL. **(A)** Cumulative survival (Kaplan–Meier) curves for NfL plasma levels (76/76 subthreshold (blue) /suprathreshold (red) patients, respectively, from enrollment in the UK discovery cohorts and **(B)** similarly, NfL serum levels (140/60 subthreshold (blue) /suprathreshold (red) patients, respectively) in PGB replication cohort. Forest plots of univariate Cox proportional hazard ratios of 18 isomiRs, NfL and miR-181 in **(C)** the UK discovery cohort or, **(D)** PGB replication cohort. NfL, miR-181 and isomiR-let-7g-5p.t are highlighted in green. **(E)** Cumulative survival (Kaplan–Meier) curves for isomiR-let-7g-5p.t/NfL ratio (76/76 subthreshold (blue) / suprathreshold (red) patients, respectively) from enrollment in the UK discovery cohorts and **(F)** PGB replication (53/147 subthreshold (blue) / suprathreshold (red) patients, respectively)

Figure 5 – isomiR-let-7g-5p.t-based prognostication compared to miR-181. Cumulative survival (Kaplan–Meier) for miR-181 (122/32 subthreshold (blue) / suprathreshold (red) patients, respectively) in the UK discovery **(A)** or PGB replication (181 /19 patients subthreshold (blue) / suprathreshold (red) patients, respectively) **(B)** cohorts. Forest plots of multivariate Cox proportional hazard ratios of 18 isomiRs and miR-181 in the UK discovery **(C)** or PGB replication **(D)** cohorts. isomiR-let-7g-5p.t and miR-181 are highlighted in green. **(E)** Cumulative survival (Kaplan–Meier) for isomiR-let-7g-5p.t/miR-181 ratio (24/130 subthreshold (blue) / suprathreshold (red) patients, respectively) from enrollment in the UK discovery cohorts and **(F)** PGB replication (11/189 subthreshold (blue) / suprathreshold (red) patients, respectively)

Tables with titles and legends

	UK – discovery (N = 154)	PGB – replication (N = 200)	Overall (N = 354)
Sex:			
Males	91 (59%)	90 (45%)	181 (51%)
Females	63 (41%)	110 (55%)	173 (49%)
Age at symptom onset (years) (Mean ± SD)	62.3 (± 11)	57 (± 12)	59.3 (± 12)
Age at enrollment (years) (Mean ± SD)	63.9 (± 11)	58.4 (± 12)	60.8 (± 12)
Disease duration at enrollment (months) (Mean ± SD)	19.3 (± 10)	16.5 (± 8)	17.7 (± 9)
Baseline ALSFRS-R (AU) (Median [Q1, Q3])	38 [33, 43]	39 [35, 41]	39 [34, 42]
DeltaFRS (-ALSFRS-R unit / month) (Median [Q1, Q3])	0.6 [0.3, 0.9]	0.6 [0.4, 0.9]	0.6 [0.3, 0.9]
Site of symptom onset			
Bulbar	52 (34%)	55 (28%)	107 (30%)
Non-bulbar	102 (66%)	145 (72%)	247 (70%)
C9orf72 status			
Carrier	12 (8%)	20 (10%)	32 (9%)
Non-carrier	142 (92%)	180 (90%)	322 (91%)
Riluzole treatment			
Yes	109 (71%)	134 (67%)	243 (69%)
No	45 (29%)	66 (33%)	111 (31%)
Survival from symptom onset (months) (Median [Q1, Q3])	33 [26.7, 46.7]	33.4 [22.5, 43.9]	33 [23.7, 44.5]
Survival from enrollment (months) (Median [Q1, Q3])	13.9 [8.9, 23.9]	14.7 [8.5, 24.1]	14.3 [8.7, 24.1]
Censored event			
Yes	9 (6%)	108 (54%)	117 (33%)
No	145 (94%)	92 (46%)	237 (67%)

Table 1. Clinical and demographic description of the study cohorts. Abbreviations: Baseline ALSFRS-R - ALSFRS-R score at baseline assessment. DeltaFRS – change of ALSFRS-R score from symptom onset to baseline assessment divided by months passed. Age at enrollment – age at baseline assessment. Survival from symptom onset / enrollment – time from symptom onset or baseline assessment to death/event censored. Censored event – number and percentage of right-censored cases. SD standard deviation. Q1 - 25th percentile, Q3 - 75th percentile.

STAR★METHODS

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Eran Hornstein (eran.hornstein@weizmann.ac.il)

Materials availability

This study did not generate new unique reagents.

Data and code availability

Newly generated transcriptomic small RNA-seq datasets have been uploaded to GEO: GSE298532 and GSE307456 and the programming code with analyzed data have been uploaded to Mendeley, doi: 10.17632/2xxw2f28p6.2. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Experimental model and subject details

Human samples

Two ALS cohorts were included in this study. The ALS biomarker study (REC Reference: 09/H0703/27 IRAS ID: 47352) cohort which consists of 154 British patients (London, UK) matched in their inclusion criteria to a “trial-like” subset of the PGB study (registered at clinicaltrials.gov NCT02327845) consisting of 200 patients from 12 U.S. sites and one South African site, associated to the Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) Consortium[9,49]. Patients were diagnosed with ALS by experienced ALS neurologists. Inclusion criteria

were as follows: time from symptom onset to first blood draw shorter than 36 months, a baseline slow vital capacity (SVC) equal to or above 50%, and at least two study visits with measurement of the revised amyotrophic lateral sclerosis functional rating scale (ALSFRS-R)[104]. Cognitive impairment (ALSci) or behavioral impairment (ALSbi) were determined by applying the revised Strong criteria[105] to Edinburgh Cognitive and Behavioral ALS Screen (ECAS[106]) data. Patients with either neuroinflammatory or neurodegenerative disorders, recent injuries, systemic or organ-specific autoimmune disorders, and recent treatment with steroids, immunosuppressants, or immunoglobulins were excluded from the British cohort. Control cohorts included 99 adult individuals from the ALS biomarker study in Queen Mary Hospital (ethics approval 09/H0703/27) and 56 adult individuals enrolled in the longitudinal FTD cohort studies at University College London (ethics approval 15/0805). All participants provided written consent (or verbal permission for a carer to sign on their behalf) to be enrolled in the ALS biomarker study if they met inclusion criteria. Ethics approval was obtained from East London and the City Research Ethics Committee 1 or the University of Miami central institutional review board (IRB), which serves as the IRB of record for the CReATe Consortium (Internal Protocol ID #20160603). The University of Cape Town Health Sciences Human Research Ethics Committee approved the study in South Africa (REF number 165/2017). Participants information on sex, and age was self-reported. Information on gender, race, ethnicity and socioeconomic status was not collected. Patients were de-identified. Disease severity was assessed with the ALSFRS-R. The progression rate prior to enrollment was estimated using the DeltaFRS, calculated by subtracting the ALSFRS-R score at baseline assessment from a maximum score of 48, and dividing by the number of months passing from symptom onset to baseline assessment. Any use of riluzole was recorded. Index date (follow-up “time zero”) was defined as the date of enrollment and survival time was measured from the index date to events for either mortality, tracheostomy or permanent assisted ventilation (PAV). The censoring date for survival analysis was documented in a case where an event was not observed before the study ended. Experimenters were blinded during the steps of the molecular analysis. Whole blood was collected by venipuncture in ethylenediaminetetraacetic acid (EDTA) tubes, and plasma was recovered by centrifugation at room temperature (RT) for 10 min at 1300g- 1750g. Plasma was stored at < −70 °C prior to use. Clinical information for patients and healthy controls is shown in Table 1.

Small RNA next-generation sequencing

The molecular experimenters were blinded to the identity of the samples. Total RNA was extracted from 500µl plasma using miRNeasy Micro Kit following the manufacturer's protocol except

adding 2.5ml of Qiazol to each plasma sample (Qiagen, #217084). RNA was quantified with a Qubit fluorometer using the RNA Broad Range Assay Kit (Thermo Fisher Scientific, #Q10211). Small RNA next-generation sequencing libraries were prepared using the QIAseq miRNA Library Kit (Qiagen, #331502) from 7.5 ng of total RNA. If RNA concentration was below the limit of detection, 5µl of undiluted RNA were used. Precise linear quantification of miRNAs was achieved by 12-nucleotide long UMIs[107]. Samples were randomly assigned to the preparation of libraries indexed with QIAseq miRNA NGS 48 Index IL (Qiagen, #331595) and sequenced in batches of 48 or fewer samples. Library concentration was determined with a Qubit fluorometer (dsDNA High Sensitivity Assay Kit, Thermo Fisher Scientific, #Q32854) and library size with TapeStation D1000 (Agilent). Libraries with different indices were multiplexed and sequenced on a NextSeq 500/550 v2 (Illumina, #20024906) or a NovaSeq 6000 flow cell (Illumina, #20028401), with a 75-bp single read or paired-end reads with 6-bp index. FASTQ files were de-multiplexed using the user-friendly transcriptome analysis pipeline[108]. Human miRNAs and isomiRs, as defined by miRBase - V22[109], were mapped using CLC genomic workbench 22.0 (Qiagen) only to the positive strand. Reads were trimmed in two consecutive fashions; Phred score[110], based error ($p_{error} = 10^{-\frac{q}{10}}$) was calculated for every nucleotide, high values reflect poor read quality. A trimming score defined by the subtraction of the error from a threshold of 0.05 was set for each nucleotide. Then, a running sum of the nucleotide errors was calculated per read. The region between the first positive value of the running sum and the highest value of the running sum was kept with non-positive score reads being removed completely. Consecutively, reads' ends were trimmed to keep the maximum length region containing 2 or fewer ambiguous nucleotides. This was followed by the library's 3'-adaptors (TGGAATTCTCGGGTGCCAAGG) trimming, with reads that did not contain any adaptors being excluded from downstream analyses. Reads were filtered based on lengths of 15-55 bps. IsomiRs were considered as 2bps changes (insertion, deletion, shift, and mismatch) up/downstream from miRBase reference sequence with a maximum of 2 mismatches allowed in alignment. The algorithm prioritizes perfect matches to reference miRNAs, followed by sequences of increasing mismatch nucleotides. When a sequence aligns equally well to multiple isomiRs, counts are assigned uniformly across all matching forms.

IsomiR definitions

Reference miRNA refers to the most abundantly expressed sequence variant among all isomiRs derived from a given miRNA gene. The collective set of all isomiRs, often comprising around 50 distinct variants annotated for a specific miRNA gene is defined as the miRNA.

QC and pre-processing of sequencing data

Sequencing data were pre-processed separately for each cohort (British n=154 / PGB n=200) and yielded ~280K/~228K isomiRs, respectively but 99% of identified isomiRs were extremely rare and sparsely detected across samples. Therefore, isomiRs were included in the analysis only if their average expression level across cohort's samples was above 5 UMIs, yielding a total of 2046 and 2003 isomiRs in the British and PGB cohorts, respectively. As geometric mean cannot be calculated with missing values, the remaining isomiR missing reads were imputed using multiple imputation with a denoising autoencoder algorithm (parameters: 40 forward passes, three layers with 512 nodes each, probability of corruption of 0.95, learning rate of $1e^{-6}$). Missing values were replaced by the mean isomiR levels across 50 imputed datasets. Clinical features used for estimation of similarity between patients: age, sex, site of symptom onset, and Riluzole treatment. Python-V3.9.2 and R-V4.0.4 package *rMIDAS*-V.0.3[111] were used for the imputation. Using *DESeq2*[112], libraries were corrected for size by geometric mean, under the assumption that miRNA counts followed negative binomial distribution[112]. Then, imputed values of original missing values were transformed back to zero to avoid bias in batch correction. Finally, data were corrected for the library preparation batch to reduce its potential bias by the R Combat function in *SVA*-V3.38 package[113]. Then, levels of 1480 isomiR identities shared across the two cohorts were standardized and isomiRs were considered for downstream analyses.

Machine learning-based isomiR selection for disease prognostication

Biomarker discovery analysis (training) was performed using the UK cohort that we have recently characterized[17] while the PGB cohort was reserved for external validation. The 1,480 isomiR candidates were randomly sampled with repetitions to generate 3,718 bootstrap subsets, each containing 20 isomiRs. The sampling process ensured that each isomiR was included in at least 50 distinct subsets. Importantly, any given subset contained one specific isomiR of interest alongside 19 other randomly selected isomiRs, resulting in at least 50 different subset compositions for each individual isomiR. For each bootstrap set survival analysis was conducted using three survival models: Cox regression analysis[42], RSF[44], and XGB[43]. The analyses were performed using the scikit-survival library (version 0.23.0) in Python. SFS[114], adds iteratively isomiRs in different combinations to optimize the C-index[115] (*mlxtend* library in Python).

The process explored approximately 1,000,000 possible combinations of isomiRs for each subset $(\sum_{k=0}^{20} \binom{20}{k} = 2^{20} = 1,048,576)$, resulting in roughly 3.9 billion combinations across all 3,718 isomiRs subsets. Each isomiR combination was evaluated using 10-fold CV on subsets of the UK discovery cohort within each of the three orthogonal survival models. A selection threshold of ≥90% in the SFS process yielded 84 isomiRs. Subsequently, additional iterations of feature selection were performed on the 84 isomiRs, generating 222 random subsets for further analysis. A C-index-based threshold was determined using the elbow method[53] and applied across all random subsets. This process identified two final panels, together consisting of 18 unique isomiRs, which were selected for downstream analyses.

isomiRs panel prediction validation

After feature selection, an XGB model was trained on the UK discovery cohort to predict patient survival time from enrollment. Hyperparameter tuning was conducted using 5-fold CV over three rounds of *Optuna* optimization[116] (1000 trials per round, with a random sampler and seed = 42), followed by manual fine-tuning. The optimized model was trained on the selected isomiR levels from the UK cohort and validated using the corresponding isomiR levels from the PGB replication cohort. SHAP[117] were used to identify feature importance within the model and were calculated on the individual predicted hazard function. HRs, based on SHAP values, were calculated as previously described[54]. Individual SHAP values were exponentiated across features. Values were averaged separately for all patients below feature mean level and those above it. The average value of above mean levels was divided by the average value of below mean level to yield the hazard ratio. The performance of the chosen isomiR-based model was compared against alternative models trained on: reference miRNA levels only, miRNAs levels (defined as the aggregate of all isomiRs within a miRNA annotation).

NfL assay

NfL levels were quantified using single molecule array (Simoa) technology on the Simoa HD-1 Analyzer (Quanterix) in plasma/serum samples from 152 and 200 patients in the UK discovery and PGB replication cohorts, respectively. The assay utilized the Simoa NfL Advantage Kit (Quanterix, #102258), with standards, primary and secondary antibodies, and detection ranges specified by the manufacturer, including defined lower and upper detection limits. Univariate or multivariate Cox regression, with or without clinical features, and associations between these and mortality were analyzed as described below.

Feature engineering

For univariate/multivariate Cox regressions isomiR levels and NfL levels were dichotomized into

binary variables based on the median levels in the UK discovery cohort. miR-181 levels were dichotomized using a previously published threshold[17]. Combinations of isomiR-let-7g-5p.t with NfL or miR-181 were evaluated by their ratios, followed by a dichotomization based on the median levels in the UK discovery cohort for NfL combination, or by an in-house algorithm that evaluated dichotomization cutpoint by optimizing univariate Cox mortality HR on the UK discovery cohort.

miRNA target prediction

mRNA targets of the reference miRNA were retrieved from miRCarta[55], retaining those listed in at least one of miRTarBase, micro-T-CDS, or TargetScan. Target canonical transcripts sequences were obtained from Ensembl[118] using *BioPython* V-1.83[119]. miRNA and mRNA sequences were analyzed using RNA22-v2[56] with sensitivity-optimized parameters (see Supplementary Methods). Significant hybridization sites were counted and compared between the reference miRNA and the isomiR. Targets with differing site counts were analyzed for enrichment using Enrichr[120-122] via *GSEApy V-1.1.9*[123] and tested against an ALS gene list obtained from Eitan et al[124].

Statistical analysis

The Student's t-test was applied to assess significant differences in the age at symptom onset or study enrollment between groups. The Mann–Whitney U test was used to evaluate differences in ALSFRS-R, DeltaFRS scores, and isomiR-let-7g-5p.t levels between the groups. Differences in the proportions of censored events, sex prevalence, or Riluloze treatment were tested using a chi-squared goodness-of-fit test. One-way ANOVA with Tukey-HSD post-hoc analysis assesses differences in isomiR-let-7g-5p.t levels across ALS sub cohorts and control. The Wilcoxon signed-rank test was used to evaluate enrichment of miRNA targets for ALS genes. Pearson correlation coefficients were calculated to examine the relationships between isomiR levels within/between the two cohorts. Kaplan–Meier survival estimators were analyzed using the log-rank Mantel–Cox test, with survival data censored at a fixed date. Univariate and multivariate Cox proportional hazard models were employed to estimate mortality hazard ratios from the time of enrollment, and model fit was assessed using the C-index. The performance of the XGB model was evaluated using the dynamic AUC(t)[125] across the range of survival periods, the average AUC(t), and the C-index. All statistical analyses were conducted, and visualizations were generated using Python (version 3.9.2), R (version 4.0.4), and BioRender.

Supplemental Information titles and legends

Table S1 - IsomiR Nomenclature

Mutations are separated by a dot. Examples of mutations are highlighted in bold and red.

Base exchanges are represented by the original nucleotide, followed by its position in the reference sequence, and then the new nucleotide (e.g., A12G).

Reference miRNAs do not have any suffixes. Deletions are indicated by lowercase letter suffixes. Insertions are indicated by uppercase letter suffixes.

5′ indels are prefixed with the letter “s”; otherwise, indels are assumed to be at the 3′ end. The position of base exchange in indels remains relative to the reference sequence.

Table S2 – miRNAs, a term reflecting on the integration of all isomiRs associated to a miRNA gene, display poor prognostication capacity. Cox univariate results of the miRNA associated with isomiRs from Figure 2. Results are shown for UK – discovery cohort (blue) and PGB – replication (purple). P – p-value, HR – hazard ratio, C-index – concordance index, CI – confidence interval.

Table S3 – plasma levels of reference miRNA associated to the 18 isomiRs studied. Levels of reference miRNAs and miRNAs (integrated of all isomiRs associated to a defined miRNA gene),

across patients in the UK discovery and the PGB replication cohorts. Units are UMIs, unique molecular identifiers. Mean plasma levels and standard deviation.

Figure S1 – Above/below median levels of isomiR-let-7g-5p.t are not correlated to clinical features. Boxplots of **(A)** isomiR-let-7g-5p.t levels, **(B)** Age at disease symptom onset, **(C)** Age at study enrollment**, (D)** baseline ALSFRS-R score or **(E)** DeltaFRS, measured as the reduction in the ALSFRS-R score per month in above/below median groups across cohorts. Pie charts of **(F)** Sex, **(G)** SSO - site of symptom onset or **(H)** Riluzole treatment above/below median groups across cohorts. *** P-value ≤ 0.001, two-sided Mann-Whitney U test. $N_{UK\ above}$ = 77, $N_{UK\ below}$ = 77, $N_{PGB\ above}$ = 160, $N_{PGB\ below}$ = 40.

Figure S2 - Feature selection process yielded 18 candidate isomiRs as optimal predictors for survival. (A) The workflow outlines the multi-step feature selection strategy applied to 1,480 isomiRs profiled in a UK discovery cohort of 154 ALS patients. Initially, three different survival models — Cox proportional hazards (COX), Random Survival Forest (RSF), and gradient survival boosting (XGB) — were each applied across repeated data splits, composed of 20 isomiRs. For each model, sequential feature selection (SFS) was performed to identify isomiRs contributing most to predictive performance. IsomiRs selected in at least 90% of the SFS runs were retained, yielding a set of 84 highly stable candidates. This refined set underwent a second round of XGB modeling. The top-scoring model from this phase, based on concordance index (C-index), identified a panel of 18 isomiRs that supports survival outcomes. **(B)** Scatter plot of the mean C-index (N = 50) across 3,718 isomiR subsets (y axis) and percentage of times isomiRs were selected in sequential forward selection (SFS) in three survival analyses: Cox (green), RSF (blue), or XGB (purple). **(C)** A plot of C-index (y axis) for sampling of 20 isomiRs from a total of 84, with repetition, using XGB (N = 222). 18 isomiRs within two subsets, above the red dashed line, displayed the best C-index.

Figure S3 – isomiR characteristics across the discovery and replication cohorts. (A) Boxplot of corrected isomiR levels (y-axis) in the UK discovery cohort (blue, N = 154) and the PGB replication cohort (purple, N = 200). The 1st and 99th percentiles of isomiR levels are denoted by whiskers. Extreme values are marked by circles **(B)** Bar plot of missing isomiR values in patients from the UK discovery (blue) and PGB replication (purple) cohorts. Heatmaps of the correlation between plasma isomiR levels in **(C)** the UK discovery cohort and **(D)** the PGB replication cohort clustered by hierarchical clustering according to Euclidean distance and average linkage by UK discovery data. Pearson's correlation r from -1.0 (complete negative correlation, blue) to 1.0 (complete positive correlation, red).

Figure S4 – isomiR-let-7g-5p.t plasma levels are reduced in ALS patients compared to controls. Boxplot of isomiR-let-7g-5p.t levels in plasma of controls from two independent cohorts (green , N = 99; yellow, N = 56) and ALS sub cohorts above, (light blue, n = 77; light purple, N = 160) or below (dark blue, N = 77; dark purple, N = 40) in the UK discovery or PGB - replication cohorts, respectively. * P-value ≤ 0.05, ** P-value ≤ 0.01 , *** P-value ≤ 0.001, Tukey's HSD post-hoc test.

Figure S5 - Cumulative survival (Kaplan–Meier) curves for ML selected isomiRs. Data dichotomized by UK discovery isomiR median values. Curves represent values from the **(A)** UK - discovery or **(B)** PGB – replication cohorts, respectively. Subthreshold or suprathreshold groups are denoted in blue / red, respectively. Group median survival and Logrank statistics are presented.

Figure S6 - Prognostic performance of joint isomiR-let-7g-5p.t and NfL. Forest plots of multivariate Cox proportional hazard ratios of 18 isomiRs and NfL in **(A)** the UK discovery or **(B)** PGB replication cohorts. Multivariate and univariate Cox hazard ratios of NfL and isomiR-let-7g-5p.t in the UK discovery **(C)** or PGB replication **(D)** cohorts Multivariate Cox hazard ratios of NfL, isomiR-let-7g-5p.t and clinical features in the UK discovery **(E)** or PGB replication **(F)**. Univariate Cox proportional hazard ratios of isomiR-let-7g-5p.t, NfL, miR-181 and the ratios of isomiR-let-7g-5p.t,/NfL or isomiR-let-7g-5p.t/miR-181, in both cohorts **(G)**.

References

1. Pfohl, S.R., Kim, R.B., Coan, G.S., and Mitchell, C.S. (2018). Unraveling the complexity of amyotrophic lateral sclerosis survival prediction. Front. Neuroinform. *12*. 10.3389/fninf.2018.00036.
2. Beghi, E., Chiò, A., Couratier, P., Esteban, J., Hardiman, O., Logroscino, G., Millul, A., Mitchell, D., Preux, P.-M., Pupillo, E., et al. (2011). The epidemiology and treatment of ALS: focus on the heterogeneity of the disease and critical appraisal of therapeutic trials. Amyotroph. Lateral Scler. *12*, 1-10. 10.3109/17482968.2010.502940.
3. Benatar, M., Boylan, K., Jeromin, A., Rutkove, S.B., Berry, J., Atassi, N., and Bruijn, L. (2016). ALS biomarkers for therapy development: State of the field and future directions. Muscle Nerve *53*, 169-182. 10.1002/mus.24979.
4. Turner, M.R., Bowser, R., Bruijn, L., Dupuis, L., Ludolph, A., McGrath, M., Manfredi, G., Maragakis, N., Miller, R.G., Pullman, S.L., et al. (2013). Mechanisms, models and biomarkers in amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Frontotemporal Degener. *14 Suppl 1*, 19-32. 10.3109/21678421.2013.778554.
5. Shepheard, S.R., Wuu, J., Cardoso, M., Wiklendt, L., Dinning, P.G., Chataway, T., Schultz, D., Benatar, M., and Rogers, M.-L. (2017). Urinary p75ECD: A prognostic, disease progression, and pharmacodynamic biomarker in ALS. Neurology *88*, 1137-1143. 10.1212/WNL.0000000000003741.
6. Dergai, O., Wuu, J., Koziczak-Holbro, M., Malaspina, A., Granit, V., Hernandez, J.P., Cooley, A., Sachdev, R., Yu, L., Bidinosti, M., et al. (2025). Skeletal muscle biomarkers of amyotrophic

lateral sclerosis: a large-scale, multi-cohort proteomic study. medRxiv. 10.1101/2025.04.23.25326161.

7. Dobrowolny, G., Martone, J., Lepore, E., Casola, I., Petrucci, A., Inghilleri, M., Morlando, M., Colantoni, A., Scicchitano, B.M., Calvo, A., et al. (2021). A longitudinal study defined circulating microRNAs as reliable biomarkers for disease prognosis and progression in ALS human patients. Cell Death Discov *7*, 4. 10.1038/s41420-020-00397-6.

8. Benatar, M., Wuu, J., and Turner, M.R. (2023). Neurofilament light chain in drug development for amyotrophic lateral sclerosis: a critical appraisal. Brain *146*, 2711-2716. 10.1093/brain/awac394.

9. Benatar, M., Macklin, E.A., Malaspina, A., Rogers, M.-L., Hornstein, E., Lombardi, V., Renfrey, D., Shepheard, S., Magen, I., Cohen, Y., et al. (2024). Prognostic clinical and biological markers for amyotrophic lateral sclerosis disease progression: validation and implications for clinical trial design and analysis. EBioMedicine *108*, 105323. 10.1016/j.ebiom.2024.105323.

10. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854. 10.1016/0092-8674(93)90529-y.

11. Haramati, S., Chapnik, E., Sztainberg, Y., Eilam, R., Zwang, R., Gershoni, N., McGlinn, E., Heiser, P.W., Wills, A.-M., Wirguin, I., et al. (2010). miRNA malfunction causes spinal motor neuron disease. Proc. Natl. Acad. Sci. U. S. A. *107*, 13111-13116. 10.1073/pnas.1006151107.

12. Emde, A., Eitan, C., Liou, L.-L., Libby, R.T., Rivkin, N., Magen, I., Reichenstein, I., Oppenheim, H., Eilam, R., Silvestroni, A., et al. (2015). Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: a new mechanism for ALS. EMBO J. *34*, 2633-2651. 10.15252/embj.201490493.

13. Katsu, M., Hama, Y., Utsumi, J., Takashina, K., Yasumatsu, H., Mori, F., Wakabayashi, K., Shoji, M., and Sasaki, H. (2019). MicroRNA expression profiles of neuron-derived extracellular vesicles in plasma from patients with amyotrophic lateral sclerosis. Neurosci. Lett. *708*, 134176. 10.1016/j.neulet.2019.03.048.

14. Reichenstein, I., Eitan, C., Diaz-Garcia, S., Haim, G., Magen, I., Siany, A., Hoye, M.L., Rivkin, N., Olender, T., Toth, B., et al. (2019). Human genetics and neuropathology suggest a link between miR-218 and amyotrophic lateral sclerosis pathophysiology. Sci Transl Med *11*. 10.1126/scitranslmed.aav5264.

15. Toivonen, J.M., Manzano, R., Oliván, S., Zaragoza, P., García-Redondo, A., and Osta, R. (2014). MicroRNA-206: a potential circulating biomarker candidate for amyotrophic lateral sclerosis. PLoS One *9*, e89065. 10.1371/journal.pone.0089065.

16. Wakabayashi, K., Mori, F., Kakita, A., Takahashi, H., Utsumi, J., and Sasaki, H. (2014). Analysis of microRNA from archived formalin-fixed paraffin-embedded specimens of amyotrophic lateral sclerosis. Acta Neuropathol Commun *2*, 173. 10.1186/s40478-014-0173-z.

17. Magen, I., Yacovzada, N.S., Yanowski, E., Coenen-Stass, A., Grosskreutz, J., Lu, C.-H., Greensmith, L., Malaspina, A., Fratta, P., and Hornstein, E. (2021). Circulating miR-181 is a prognostic biomarker for amyotrophic lateral sclerosis. Nat. Neurosci. *24*, 1534-1541. 10.1038/s41593-021-00936-z.

18. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419. 10.1038/nature01957.

19. Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366. 10.1038/35053110.

20. Guo, L., and Chen, F. (2014). A challenge for miRNA: multiple isomiRs in miRNAomics. Gene *544*, 1-7. 10.1016/j.gene.2014.04.039.

21. Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., et al. (2008). Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. *18*, 610-621. 10.1101/gr.7179508.
22. Newman, M.A., Mani, V., and Hammond, S.M. (2011). Deep sequencing of microRNA precursors reveals extensive 3' end modification. RNA *17*, 1795-1803. 10.1261/rna.2713611.
23. Neilsen, C.T., Goodall, G.J., and Bracken, C.P. (2012). IsomiRs--the overlooked repertoire in the dynamic microRNAome. Trends Genet. *28*, 544-549. 10.1016/j.tig.2012.07.005.
24. Kwiatkowski, T.J., Jr., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205-1208. 10.1126/science.1166066.
25. Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science *323*, 1208-1211. 10.1126/science.1165942.
26. Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., et al. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668-1672. 10.1126/science.1154584.
27. Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E.M., Logroscino, G., Robberecht, W., Shaw, P.J., Simmons, Z., and van den Berg, L.H. (2017). Amyotrophic lateral sclerosis. Nature Reviews Disease Primers *3*, 1-19. 10.1038/nrdp.2017.71.
28. Keuss, M.J., Harley, P., Ryadnov, E., Jackson, R.E., Zanovello, M., Wilkins, O.G., Barattucci, S., Mehta, P.R., Oliveira, M.G., Parkes, J.E., et al. (2024). Loss of TDP-43 induces synaptic dysfunction that is rescued by UNC13A splice-switching ASOs. bioRxivorg. 10.1101/2024.06.20.599684.
29. Ma, X.R., Prudencio, M., Koike, Y., Vatsavayai, S.C., Kim, G., Harbinski, F., Briner, A., Rodriguez, C.M., Guo, C., Akiyama, T., et al. (2022). TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. Nature *603*, 124-130. 10.1038/s41586-022-04424-7.
30. Melamed, Z.e., López-Erauskin, J., Baughn, M.W., Zhang, O., Drenner, K., Sun, Y., Freyermuth, F., McMahon, M.A., Beccari, M.S., Artates, J.W., et al. (2019). Premature polyadenylation-mediated loss of stathmin-2 is a hallmark of TDP-43-dependent neurodegeneration. Nature Neuroscience *22*, 180-190. 10.1038/s41593-018-0293-z.
31. 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., et al. (2024). HDGFL2 cryptic proteins report presence of TDP-43 pathology in neurodegenerative diseases. Molecular Neurodegeneration *19*, 1-5. 10.1186/s13024-024-00718-8.
32. 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., et al. (2024). A fluid biomarker reveals loss of TDP-43 splicing repression in presymptomatic ALS–FTD. Nature Medicine *30*, 382-393. 10.1038/s41591-023-02788-5.
33. Chen, X., Fan, Z., McGee, W., Chen, M., Kong, R., Wen, P., Xiao, T., Chen, X., Liu, J., Zhu, L., et al. (2018). TDP-43 regulates cancer-associated microRNAs. Protein Cell *9*, 848-866. 10.1007/s13238-017-0480-9.
34. Morlando, M., Dini Modigliani, S., Torrelli, G., Rosa, A., Di Carlo, V., Caffarelli, E., and Bozzoni, I. (2012). FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment: FUS participates in microRNA biogenesis. EMBO J. *31*, 4502-4510. 10.1038/emboj.2012.319.

35. Porta, S., Kwong, L.K., Trojanowski, J.Q., and Lee, V.M.Y. (2015). Drosha inclusions are new components of dipeptide-repeat protein aggregates in FTLD-TDP and ALS C9orf72 expansion cases. J. Neuropathol. Exp. Neurol. *74*, 380-387. 10.1097/NEN.0000000000000182.
36. Aiso, T., and Ueda, M. (2023). 5'isomiR is the most abundant sequence of miR1246, a candidate biomarker of lung cancer, in serum. Mol. Med. Rep. *27*. 10.3892/mmr.2023.12979.
37. Ibuki, Y., Nishiyama, Y., Tsutani, Y., Emi, M., Hamai, Y., Okada, M., and Tahara, H. (2020). Circulating microRNA/isomiRs as novel biomarkers of esophageal squamous cell carcinoma. PLoS One *15*, e0231116. 10.1371/journal.pone.0231116.
38. Koppers-Lalic, D., Hackenberg, M., de Menezes, R., Misovic, B., Wachalska, M., Geldof, A., Zini, N., de Reijke, T., Wurdinger, T., Vis, A., et al. (2016). Non invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. Oncotarget *7*, 22566-22578. 10.18632/oncotarget.8124.
39. Lan, C., Peng, H., McGowan, E.M., Hutvagner, G., and Li, J. (2018). An isomiR expression panel based novel breast cancer classification approach using improved mutual information. BMC Med. Genomics *11*, 118. 10.1186/s12920-018-0434-y.
40. Wang, S., Zheng, Z., Chen, P., and Wu, M. (2019). Tumor classification and biomarker discovery based on the 5'isomiR expression level. BMC Cancer *19*, 127. 10.1186/s12885-019-5340-y.
41. Loher, P., Londin, E., Ilieva, H., Pasinelli, P., and Rigoutsos, I. (2025). Re-Analyses of Samples From Amyotrophic Lateral Sclerosis Patients and Controls Identify Many Novel Small RNAs With Diagnostic And Prognostic Potential. Molecular Neurobiology *62*, 8135-8149. 10.1007/s12035-025-04747-2.
42. Cox, D.R. (1972). Regression Models and Life-Tables. J. R. Stat. Soc. Series B Stat. Methodol. *34*, 187-202. 10.1111/j.2517-6161.1972.tb00899.x.
43. Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., and Van Der Laan, M.J. (2005). Survival ensembles. Biostatistics *7*, 355-373. 10.1093/biostatistics/kxj011.
44. Ishwaran, H., Gerds, T.A., Kogalur, U.B., Moore, R.D., Gange, S.J., and Lau, B.M. (2014). Random survival forests for competing risks. Biostatistics *15*, 757-773. 10.1093/biostatistics/kxu010.
45. Breiman, L. (1996). Bagging predictors. Mach. Learn. *24*, 123-140. 10.1007/BF00058655.
46. Džeroski, S., and Ženko, B. (2004). Is Combining Classifiers with Stacking Better than Selecting the Best One? Mach. Learn. *54*, 255-273. 10.1023/B:MACH.0000015881.36452.6e.
47. Malviya, R., Ghinea, G., Dhanaraj, R.K., Balusamy, B., and Sundram, S. (2022). Deep Learning for Targeted Treatments: Transformation in Healthcare (John Wiley & Sons).
48. Faure, H., and Hrynaszkiewicz, I. (2011). The ISRCTN Register: achievements and challenges 8 years on: The ISRCTN Register. J. Evid. Based Med. *4*, 188-192. 10.1111/j.1756-5391.2011.01138.x.
49. ClinicalTrials.gov/study/NCT02327845. https://clinicaltrials.gov/study/NCT02327845.
50. Magen, I., Yacovzada, N.-S., Warren, J.D., Heller, C., Swift, I., Bobeva, Y., Malaspina, A., Rohrer, J.D., Fratta, P., and Hornstein, E. (2023). microRNA-based predictor for diagnosis of frontotemporal dementia. Neuropathol. Appl. Neurobiol., e12916. 10.1111/nan.12916.
51. Charilaou, P., and Battat, R. (2022). Machine learning models and over-fitting considerations. World J Gastroenterol *28*, 605-607. 10.3748/wjg.v28.i5.605.
52. Taylor, J.M.G., Ankerst, D.P., and Andridge, R.R. (2008). Validation of biomarker-based risk prediction models. Clin Cancer Res *14*, 5977-5983. 10.1158/1078-0432.CCR-07-4534.
53. Thorndike, R.L. (1953). Who belongs in the family? Psychometrika *18*, 267-276. 10.1007/BF02289263.
54. Sundrani, S., and Lu, J. (2021). Computing the hazard ratios associated with explanatory variables using machine learning models of survival data. JCO Clin. Cancer Inform. *5*, 364-378. 10.1200/CCI.20.00172.

55. Backes, C., Fehlmann, T., Kern, F., Kehl, T., Lenhof, H.P., Meese, E., and Keller, A. (2018). miRCarta: a central repository for collecting miRNA candidates. Nucleic Acids Res *46*, D160-D167. 10.1093/nar/gkx851.
56. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell *126*, 1203-1217. 10.1016/j.cell.2006.07.031.
57. Lu, C.-H., Macdonald-Wallis, C., Gray, E., Pearce, N., Petzold, A., Norgren, N., Giovannoni, G., Fratta, P., Sidle, K., Fish, M., et al. (2015). Neurofilament light chain: A prognostic biomarker in amyotrophic lateral sclerosis. Neurology *84*, 2247-2257. 10.1212/WNL.0000000000001642.
58. Banack, S.A., Dunlop, R.A., Stommel, E.W., Mehta, P., and Cox, P.A. (2022). miRNA extracted from extracellular vesicles is a robust biomarker of amyotrophic lateral sclerosis. J Neurol Sci *442*, 120396. 10.1016/j.jns.2022.120396.
59. de Andrade, H.M.T., de Albuquerque, M., Avansini, S.H., de S Rocha, C., Dogini, D.B., Nucci, A., Carvalho, B., Lopes-Cendes, I., and França, M.C., Jr. (2016). MicroRNAs-424 and 206 are potential prognostic markers in spinal onset amyotrophic lateral sclerosis. J. Neurol. Sci. *368*, 19-24. 10.1016/j.jns.2016.06.046.
60. Figueroa-Romero, C., Hur, J., Lunn, J.S., Paez-Colasante, X., Bender, D.E., Yung, R., Sakowski, S.A., and Feldman, E.L. (2016). Expression of microRNAs in human post-mortem amyotrophic lateral sclerosis spinal cords provides insight into disease mechanisms. Mol Cell Neurosci *71*, 34-45. 10.1016/j.mcn.2015.12.008.
61. Hur, J., Paez-Colasante, X., Figueroa-Romero, C., Lo, T.W., Barmada, S.J., Paulsen, M.T., Ljungman, M., Alakwaa, F.M., Savelieff, M.G., Goutman, S.A., and Feldman, E.L. (2023). miRNA analysis reveals novel dysregulated pathways in amyotrophic lateral sclerosis. Hum Mol Genet *32*, 934-947. 10.1093/hmg/ddac250.
62. Kawahara, Y., and Mieda-Sato, A. (2012). TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc. Natl. Acad. Sci. U. S. A. *109*, 3347-3352. 10.1073/pnas.1112427109.
63. Liguori, M., Nuzziello, N., Introna, A., Consiglio, A., Licciulli, F., D'Errico, E., Scarafino, A., Distaso, E., and Simone, I.L. (2018). Dysregulation of MicroRNAs and Target Genes Networks in Peripheral Blood of Patients With Sporadic Amyotrophic Lateral Sclerosis. Front. Mol. Neurosci. *11*, 288. 10.3389/fnmol.2018.00288.
64. Maimon, R., Ionescu, A., Bonnie, A., Sweetat, S., Wald-Altman, S., Inbar, S., Gradus, T., Trotti, D., Weil, M., Behar, O., and Perlson, E. (2018). miR126-5p Downregulation Facilitates Axon Degeneration and NMJ Disruption via a Non-Cell-Autonomous Mechanism in ALS. J Neurosci *38*, 5478-5494. 10.1523/JNEUROSCI.3037-17.2018.
65. Raheja, R., Regev, K., Healy, B.C., Mazzola, M.A., Beynon, V., Von Glehn, F., Paul, A., Diaz-Cruz, C., Gholipour, T., Glanz, B.I., et al. (2018). Correlating serum micrornas and clinical parameters in amyotrophic lateral sclerosis. Muscle Nerve *58*, 261-269. 10.1002/mus.26106.
66. Saucier, D., Wajnberg, G., Roy, J., Beauregard, A.-P., Chacko, S., Crapoulet, N., Fournier, S., Ghosh, A., Lewis, S.M., Marrero, A., et al. (2019). Identification of a circulating miRNA signature in extracellular vesicles collected from amyotrophic lateral sclerosis patients. Brain Res. *1708*, 100-108. 10.1016/j.brainres.2018.12.016.
67. Koval, E.D., Shaner, C., Zhang, P., du Maine, X., Fischer, K., Tay, J., Chau, B.N., Wu, G.F., and Miller, T.M. (2013). Method for widespread microRNA-155 inhibition prolongs survival in ALS-model mice. Hum Mol Genet *22*, 4127-4135. 10.1093/hmg/ddt261.
68. Ma, H., Wu, Y., Niu, Q., Zhang, J., Jia, G., Manjunath, N., and Wu, H. (2016). A sliding-bulge structure at the Dicer processing site of pre-miRNAs regulates alternative Dicer processing to generate 5'-isomiRs. Heliyon *2*, e00148. 10.1016/j.heliyon.2016.e00148.

69. Kmetzsch, V., Latouche, M., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., French Research Network on, F.A., Le Ber, I., Colliot, O., and Becker, E. (2022). MicroRNA signatures in genetic frontotemporal dementia and amyotrophic lateral sclerosis. Ann. Clin. Transl. Neurol. *9*, 1778-1791. 10.1002/acn3.51674.
70. Kour, S., Fortuna, T., Anderson, E.N., Mawrie, D., Bilstein, J., Sivasubramanian, R., Ward, C., Roy, R., Rajasundaram, D., Sterneckert, J., and Pandey, U.B. (2023). Drosha-dependent microRNAs modulate FUS-mediated neurodegeneration in vivo. Nucleic Acids Res *51*, 11258-11276. 10.1093/nar/gkad774.
71. Aharon, A., Spector, P., Ahmad, R.S., Horrany, N., Sabbach, A., Brenner, B., and Aharon-Peretz, J. (2020). Extracellular vesicles of Alzheimer's disease patients as a biomarker for disease progression. Mol. Neurobiol. *57*, 4156-4169. 10.1007/s12035-020-02013-1.
72. Kumar, P., Dezso, Z., MacKenzie, C., Oestreicher, J., Agoulnik, S., Byrne, M., Bernier, F., Yanagimachi, M., Aoshima, K., and Oda, Y. (2013). Circulating miRNA biomarkers for Alzheimer's disease. PLoS One *8*, e69807. 10.1371/journal.pone.0069807.
73. Mendes-Silva, A.P., Pereira, K.S., Tolentino-Araujo, G.T., Nicolau, E.d.S., Silva-Ferreira, C.M., Teixeira, A.L., and Diniz, B.S. (2016). Shared biologic pathways between Alzheimer disease and major depression: A systematic review of MicroRNA expression studies. Am. J. Geriatr. Psychiatry *24*, 903-912. 10.1016/j.jagp.2016.07.017.
74. Satoh, J.-I., Kino, Y., and Niida, S. (2015). MicroRNA-Seq data analysis pipeline to identify blood biomarkers for Alzheimer's disease from public data. Biomark. Insights *10*, 21-31. 10.4137/BMI.S25132.
75. Tan, L., Yu, J.-T., Tan, M.-S., Liu, Q.-Y., Wang, H.-F., Zhang, W., Jiang, T., and Tan, L. (2014). Genome-wide serum microRNA expression profiling identifies serum biomarkers for Alzheimer's disease. J. Alzheimers. Dis. *40*, 1017-1027. 10.3233/JAD-132144.
76. Zhang, X., Wang, Y., Liu, W., Wang, T., Wang, L., Hao, L., Ju, M., and Xiao, R. (2021). Diet quality, gut microbiota, and microRNAs associated with mild cognitive impairment in middle-aged and elderly Chinese population. Am. J. Clin. Nutr. *114*, 429-440. 10.1093/ajcn/nqab078.
77. Zhang, X., Xi, Y., Yu, H., An, Y., Wang, Y., Tao, L., Wang, Y., Liu, W., Wang, T., and Xiao, R. (2019). 27-hydroxycholesterol promotes Aβ accumulation via altering Aβ metabolism in mild cognitive impairment patients and APP/PS1 mice: 27-hydroxycholesterol leads to cognitive dysfunction via Aβ. Brain Pathol. *29*, 558-573. 10.1111/bpa.12698.
78. Kafshdooz, T., Farajnia, S., Sharifi, R., and Najmi, S. (2023). Hsa-let-7g-5p, a circulating microRNA, as a biomarker for Alzheimer's disease. Inform. Med. Unlocked *38*, 101203. 10.1016/j.imu.2023.101203.
79. Martí, E., Pantano, L., Bañez-Coronel, M., Llorens, F., Miñones-Moyano, E., Porta, S., Sumoy, L., Ferrer, I., and Estivill, X. (2010). A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. Nucleic Acids Res. *38*, 7219-7235. 10.1093/nar/gkq575.
80. Awuson-David, B., Williams, A.C., Wright, B., Hill, L.J., and Di Pietro, V. (2023). Common microRNA regulated pathways in Alzheimer's and Parkinson's disease. Front Neurosci *17*, 1228927. 10.3389/fnins.2023.1228927.
81. Daneshafrooz, N., Joghataei, M.T., Mehdizadeh, M., Alavi, A., Barati, M., Panahi, B., Teimourian, S., and Zamani, B. (2022). Identification of let-7f and miR-338 as plasma-based biomarkers for sporadic amyotrophic lateral sclerosis using meta-analysis and empirical validation. Sci Rep *12*, 1373. 10.1038/s41598-022-05067-4.
82. Durur, D.Y., Tastan, B., Ugur Tufekci, K., Olcum, M., Uzuner, H., Karakulah, G., Yener, G., and Genc, S. (2022). Alteration of miRNAs in Small Neuron-Derived Extracellular Vesicles of Alzheimer's Disease Patients and the Effect of Extracellular Vesicles on Microglial Immune Responses. J Mol Neurosci *72*, 1182-1194. 10.1007/s12031-022-02012-y.

83. Gamez-Valero, A., Campdelacreu, J., Vilas, D., Ispierto, L., Rene, R., Alvarez, R., Armengol, M.P., Borras, F.E., and Beyer, K. (2019). Exploratory study on microRNA profiles from plasma-derived extracellular vesicles in Alzheimer's disease and dementia with Lewy bodies. Transl Neurodegener *8*, 31. 10.1186/s40035-019-0169-5.
84. Gui, Y., Liu, H., Zhang, L., Lv, W., and Hu, X. (2015). Altered microRNA profiles in cerebrospinal fluid exosome in Parkinson disease and Alzheimer disease. Oncotarget *6*, 37043-37053. 10.18632/oncotarget.6158.
85. Lehmann, S.M., Kruger, C., Park, B., Derkow, K., Rosenberger, K., Baumgart, J., Trimbuch, T., Eom, G., Hinz, M., Kaul, D., et al. (2012). An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration. Nat Neurosci *15*, 827-835. 10.1038/nn.3113.
86. Liu, Y., He, X., Li, Y., and Wang, T. (2018). Cerebrospinal fluid CD4+ T lymphocyte-derived miRNA-let-7b can enhances the diagnostic performance of Alzheimer's disease biomarkers. Biochem Biophys Res Commun *495*, 1144-1150. 10.1016/j.bbrc.2017.11.122.
87. Poursaei, E., Abolghasemi, M., Bornehdeli, S., Shanehbandi, D., Asadi, M., Sadeghzadeh, M., Rahmanpour, D., and Sadeh, R.N. (2022). Evaluation of hsa-let-7d-5p, hsa-let-7g-5p and hsa-miR-15b-5p plasma levels in patients with Alzheimer's disease. Psychiatr Genet *32*, 25-29. 10.1097/YPG.0000000000000303.
88. Samareh, A., Nematollahi, M.H., Pourghadamyari, H., Meimand, H.A.E., Shabani, M., and Asadikaram, G. (2024). Diagnostic Potential of NEAT1, hsa-let-7a-5p, and miR-506-3p in Early-stage Parkinson's Disease. Curr Med Chem. 10.2174/0109298673336756241016063552.
89. van Harten, A.C., Mulders, J., Scheltens, P., van der Flier, W.M., and Oudejans, C.B. (2015). Differential Expression of microRNA in Cerebrospinal Fluid as a Potential Novel Biomarker for Alzheimer's Disease. J Alzheimers Dis *47*, 243-252. 10.3233/JAD-140075.
90. Vrabec, K., Bostjancic, E., Koritnik, B., Leonardis, L., Dolenc Groselj, L., Zidar, J., Rogelj, B., Glavac, D., and Ravnik-Glavac, M. (2018). Differential Expression of Several miRNAs and the Host Genes AATK and DNM2 in Leukocytes of Sporadic ALS Patients. Front Mol Neurosci *11*, 106. 10.3389/fnmol.2018.00106.
91. Lin, Z., Ge, J., Wang, Z., Ren, J., Wang, X., Xiong, H., Gao, J., Zhang, Y., and Zhang, Q. (2017). Let-7e modulates the inflammatory response in vascular endothelial cells through ceRNA crosstalk. Scientific Reports *7*, 1-12. 10.1038/srep42498.
92. Liu, Q., Hu, Y., Zhang, M., Yan, Y., Yu, H., and Ge, L. (2018). microRNA-451 protects neurons against ischemia/reperfusion injury-induced cell death by targeting CELF2. Neuropsychiatr. Dis. Treat. *14*, 2773-2782. 10.2147/NDT.S173632.
93. Luo, Q., Zhu, J., Zhang, Q., Xie, J., Yi, C., and Li, T. (2020). MicroRNA-486-5p Promotes Acute Lung Injury via Inducing Inflammation and Apoptosis by Targeting OTUD7B. Inflammation *43*, 975-984. 10.1007/s10753-020-01183-3.
94. Ma, N., Li, S., Lin, C., Cheng, X., and Meng, Z. (2021). Mesenchymal stem cell conditioned medium attenuates oxidative stress injury in hepatocytes partly by regulating the miR-486-5p/PIM1 axis and the TGF-β/Smad pathway. Bioengineered *12*, 6434-6447. 10.1080/21655979.2021.1972196.
95. Mandolesi, G., De Vito, F., Musella, A., Gentile, A., Bullitta, S., Fresegna, D., Sepman, H., Di Sanza, C., Haji, N., Mori, F., et al. (2017). MiR-142-3p is a key regulator of IL-1β-dependent synaptopathy in neuroinflammation. J. Neurosci. *37*, 546-561. 10.1523/jneurosci.0851-16.2017.
96. Strycharz, J., Wróblewski, A., Zieleniak, A., Świderska, E., Matyjas, T., Rucińska, M., Pomorski, L., Czarny, P., Szemraj, J., Drzewoski, J., and Śliwińska, A. (2021). Visceral adipose tissue of prediabetic and diabetic females shares a set of similarly upregulated

microRNAs functionally annotated to inflammation, oxidative stress and insulin signaling. Antioxidants (Basel) *10*, 101. 10.3390/antiox10010101.

97. Tan, W., Gu, Z., Leng, J., Zou, X., Chen, H., Min, F., Zhou, W., Zhang, L., and Li, G. (2019). Let-7f-5p ameliorates inflammation by targeting NLRP3 in bone marrow-derived mesenchymal stem cells in patients with systemic lupus erythematosus. Biomed Pharmacother *118*, 109313. 10.1016/j.biopha.2019.109313.

98. Thomas, J.J., Harp, K.O., Bashi, A., Hood, J.L., Botchway, F., Wilson, M.D., Thompson, W.E., Stiles, J.K., and Driss, A. (2022). MiR-451a and let-7i-5p loaded extracellular vesicles attenuate heme-induced inflammation in hiPSC-derived endothelial cells. Front. Immunol. *13*, 1082414. 10.3389/fimmu.2022.1082414.

99. Gan, X., Zhao, H., Wei, Y., Jiang, Q., Wen, C., and Ying, Y. (2020). Role of miR-92a-3p, oxidative stress, and p38MAPK/NF-κB pathway in rats with central venous catheter related thrombosis. BMC Cardiovascular Disorders *20*, 1-8. 10.1186/s12872-020-01436-x.

100. Dika, E., Broseghini, E., Porcellini, E., Lambertini, M., Riefolo, M., Durante, G., Loher, P., Roncarati, R., Bassi, C., Misciali, C., et al. (2021). Unraveling the role of microRNA/isomiR network in multiple primary melanoma pathogenesis. Cell Death Dis *12*, 473. 10.1038/s41419-021-03764-y.

101. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798. 10.1016/s0092-8674(03)01018-3.

102. Wong, L.L., Fadzil, A.B., Chen, Q., Rademaker, M.T., Charles, C.J., Richards, A.M., and Wang, P. (2023). Interrogating the Role of miR-125b and Its 3'isomiRs in Protection against Hypoxia. Int J Mol Sci *24*. 10.3390/ijms242116015.

103. Yu, F., Pillman, K.A., Neilsen, C.T., Toubia, J., Lawrence, D.M., Tsykin, A., Gantier, M.P., Callen, D.F., Goodall, G.J., and Bracken, C.P. (2017). Naturally existing isoforms of miR-222 have distinct functions. Nucleic Acids Res *45*, 11371-11385. 10.1093/nar/gkx788.

104. Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., and Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J. Neurol. Sci. *169*, 13-21. 10.1016/s0022-510x(99)00210-5.

105. Strong, M.J., Abrahams, S., Goldstein, L.H., Woolley, S., McLaughlin, P., Snowden, J., Mioshi, E., Roberts-South, A., Benatar, M., HortobáGyi, T., et al. (2017). Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. Amyotroph Lateral Scler Frontotemporal Degener *18*, 153-174. 10.1080/21678421.2016.1267768.

106. Abrahams, S., Newton, J., Niven, E., Foley, J., and Bak, T.H. (2014). Screening for cognition and behaviour changes in ALS. Amyotroph. Lateral Scler. Frontotemporal Degener. *15*, 9-14. 10.3109/21678421.2013.805784.

107. Coenen-Stass, A.M.L., Magen, I., Brooks, T., Ben-Dov, I.Z., Greensmith, L., Hornstein, E., and Fratta, P. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. RNA Biol. *15*, 1133-1145. 10.1080/15476286.2018.1514236.

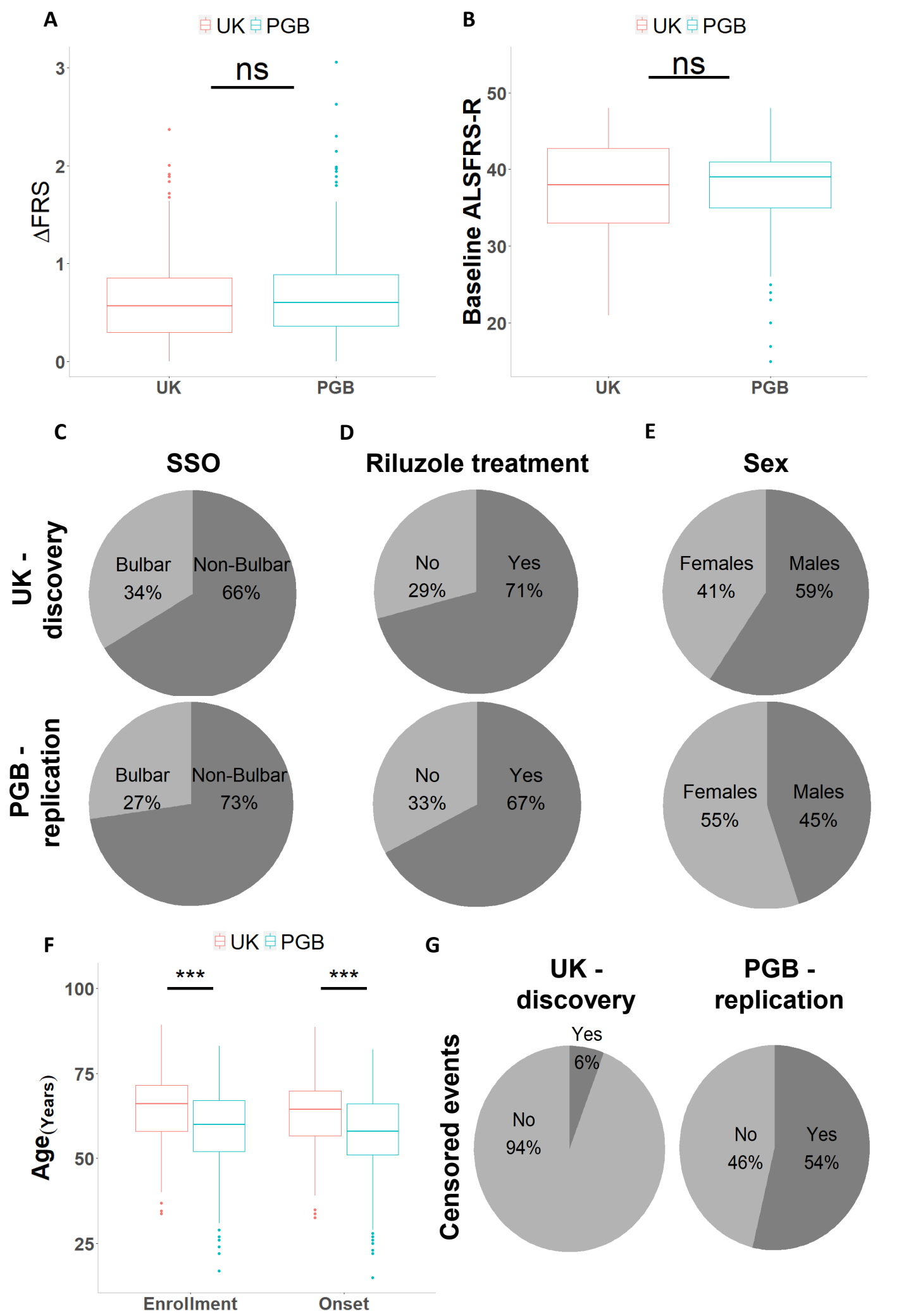
108. Kohen, R., Barlev, J., Hornung, G., Stelzer, G., Feldmesser, E., Kogan, K., Safran, M., and Leshkowitz, D. (2019). UTAP: User-friendly Transcriptome Analysis Pipeline. BMC Bioinformatics *20*, 154. 10.1186/s12859-019-2728-2.

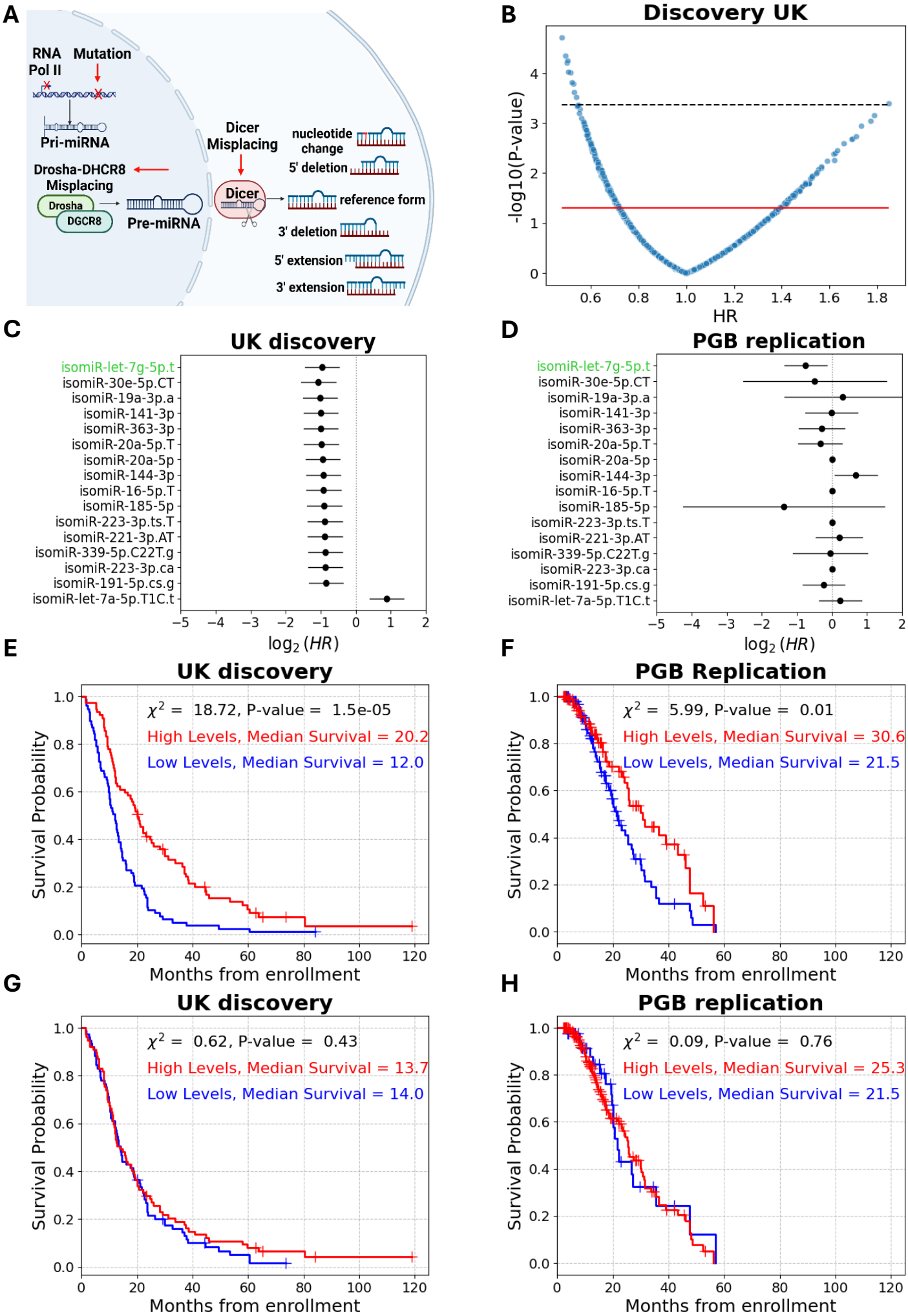
109. Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68-73. 10.1093/nar/gkt1181.

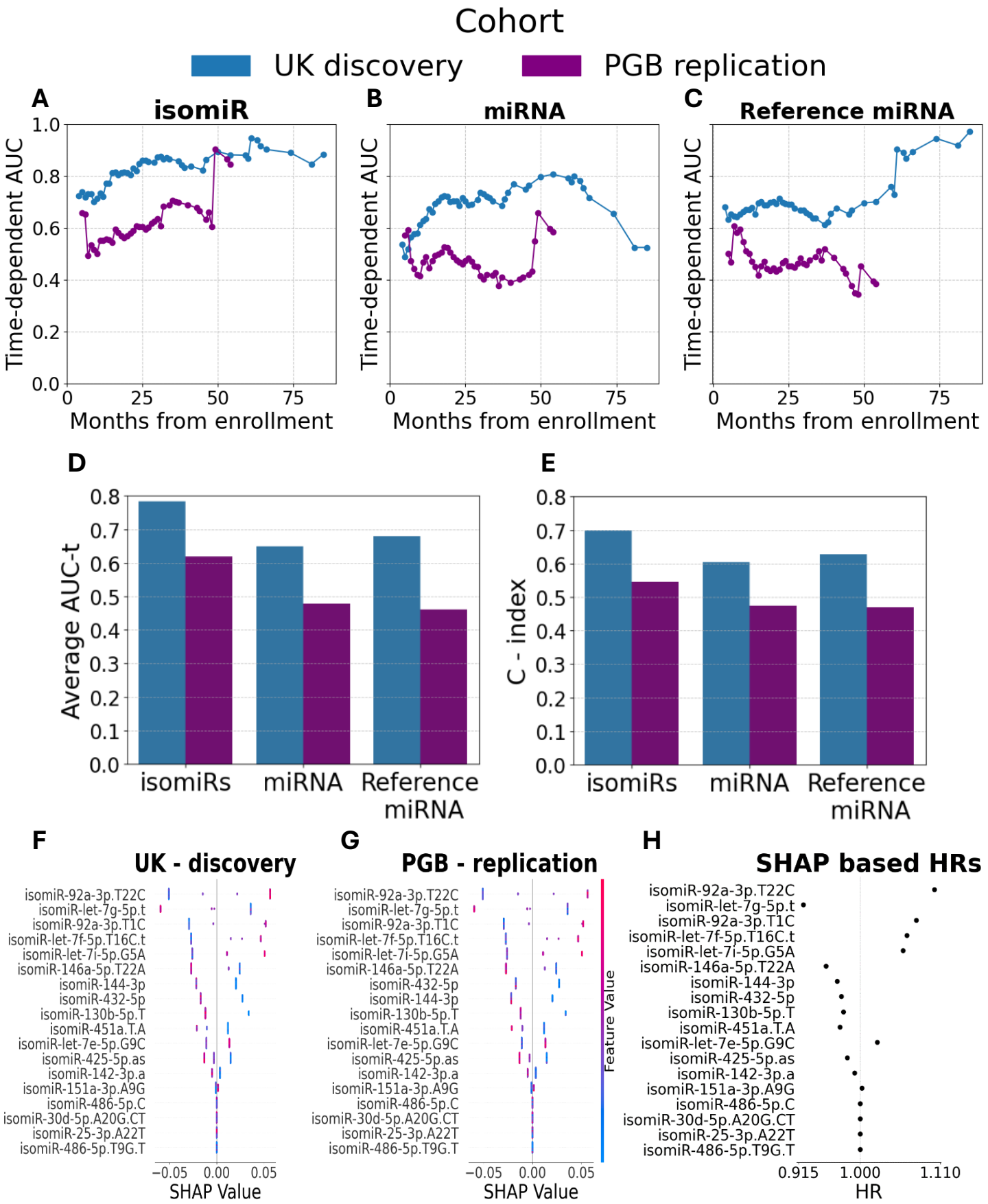
110. Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. *8*, 186-194. 10.1101/gr.8.3.186.

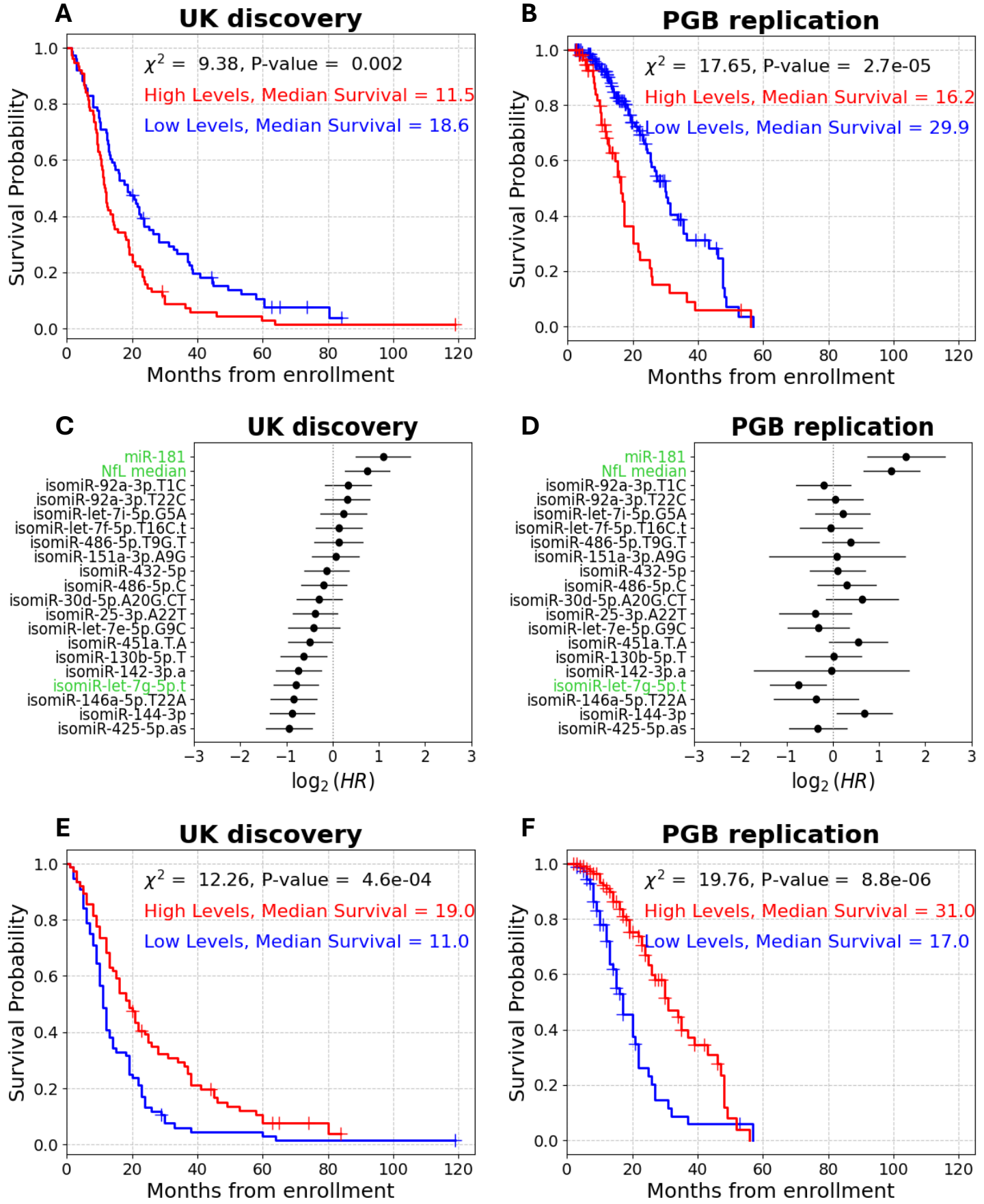
111. Lall, R., and Robinson, T. Efficient Multiple Imputation for Diverse Data in Python and R: MIDASpy and rMIDAS.

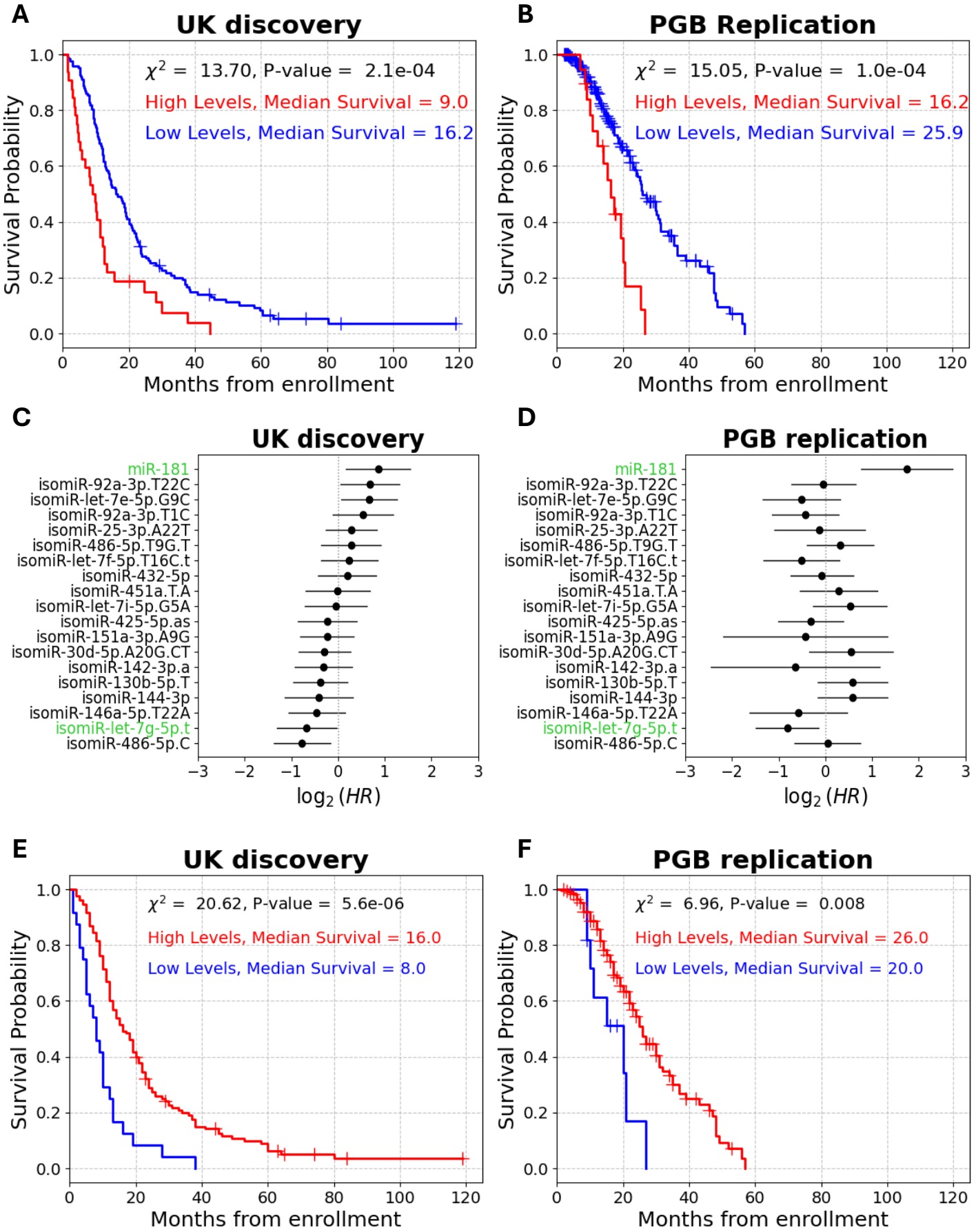
112. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.
113. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882-883. 10.1093/bioinformatics/bts034.
114. Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. Pattern Recognit. Lett. *15*, 1119-1125. 10.1016/0167-8655(94)90127-9.
115. Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. Journal of Open Source Software *3*, 638. 10.21105/joss.00638.
116. Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.
117. Lundberg, S., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv [cs.AI].
118. Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res *48*, D682-D688. 10.1093/nar/gkz966.
119. Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., and de Hoon, M.J. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics *25*, 1422-1423. 10.1093/bioinformatics/btp163.
120. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. 10.1186/1471-2105-14-128.
121. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.
122. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.
123. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*. 10.1093/bioinformatics/btac757.
124. Eitan, C., Siany, A., Barkan, E., Olender, T., van Eijk, K.R., Moisse, M., Farhan, S.M.K., Danino, Y.M., Yanowski, E., Marmor-Kollet, H., et al. (2022). Whole-genome sequencing reveals that variants in the Interleukin 18 Receptor Accessory Protein 3'UTR protect against ALS. Nat Neurosci *25*, 433-445. 10.1038/s41593-022-01040-6.
125. Lambert, J., and Chevret, S. (2016). Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. Stat Methods Med Res *25*, 2088-2102. 10.1177/0962280213515571.


A
UK PGB
ns
ΔFRS
3
2
1
0
UK
PGB
B
UK PGB
ns
Baseline ALSFRS-R
50
40
30
20
UK
PGB
C
SSO
UK - discovery
Bulbar 34%
Non-Bulbar 66%
PGB - replication
Bulbar 27%
Non-Bulbar 73%
D
Riluzole treatment
No 29%
Yes 71%
No 33%
Yes 67%
E
Sex
Females 41%
Males 59%
Females 55%
Males 45%
F
UK PGB
***
***
Age(Years)
100
75
50
25
Enrollment
Onset
G
UK - discovery
PGB - replication
Censored events
Yes 6%
No 94%
No 46%
Yes 54%


A
RNA Pol II
Mutation
Pri-miRNA
Drosha-DHCR8 Misplacing
Drosha
DGCR8
Pre-miRNA
Dicer Misplacing
Dicer
nucleotide change
5' deletion
reference form
3' deletion
5' extension
3' extension
B
Discovery UK
-log10(P-value)
HR
C
UK discovery
isomiR-let-7g-5p.t
isomiR-30e-5p.CT
isomiR-19a-3p.a
isomiR-141-3p
isomiR-363-3p
isomiR-20a-5p.T
isomiR-20a-5p
isomiR-144-3p
isomiR-16-5p.T
isomiR-185-5p
isomiR-223-3p.ts.T
isomiR-221-3p.AT
isomiR-339-5p.C22T.g
isomiR-223-3p.ca
isomiR-191-5p.cs.g
isomiR-let-7a-5p.T1C.t
log2 (HR)
D
PGB replication
E
UK discovery
χ2 = 18.72, P-value = 1.5e-05
High Levels, Median Survival = 20.2
Low Levels, Median Survival = 12.0
Survival Probability
Months from enrollment
F
PGB Replication
χ2 = 5.99, P-value = 0.01
High Levels, Median Survival = 30.6
Low Levels, Median Survival = 21.5
G
UK discovery
χ2 = 0.62, P-value = 0.43
High Levels, Median Survival = 13.7
Low Levels, Median Survival = 14.0
H
PGB replication
χ2 = 0.09, P-value = 0.76
High Levels, Median Survival = 25.3
Low Levels, Median Survival = 21.5


Cohort
UK discovery
PGB replication
A
isomiR
B
miRNA
C
Reference miRNA
Time-dependent AUC
Months from enrollment
D
Average AUC-t
isomiRs
miRNA
Reference miRNA
E
C - index
F
UK - discovery
G
PGB - replication
H
SHAP based HRs
isomiR-92a-3p.T22C
isomiR-let-7g-5p.t
isomiR-92a-3p.T1C
isomiR-let-7f-5p.T16C.t
isomiR-let-7i-5p.G5A
isomiR-146a-5p.T22A
isomiR-144-3p
isomiR-432-5p
isomiR-130b-5p.T
isomiR-451a.T.A
isomiR-let-7e-5p.G9C
isomiR-425-5p.as
isomiR-142-3p.a
isomiR-151a-3p.A9G
isomiR-486-5p.C
isomiR-30d-5p.A20G.CT
isomiR-25-3p.A22T
isomiR-486-5p.T9G.T
SHAP Value
Feature Value
HR


A
UK discovery
χ^2 = 9.38, P-value = 0.002
High Levels, Median Survival = 11.5
Low Levels, Median Survival = 18.6
Survival Probability
Months from enrollment
B
PGB replication
χ^2 = 17.65, P-value = 2.7e-05
High Levels, Median Survival = 16.2
Low Levels, Median Survival = 29.9
Survival Probability
Months from enrollment
C
UK discovery
miR-181
NfL median
isomiR-92a-3p.T1C
isomiR-92a-3p.T22C
isomiR-let-7i-5p.G5A
isomiR-let-7f-5p.T16C.t
isomiR-486-5p.T9G.T
isomiR-151a-3p.A9G
isomiR-432-5p
isomiR-486-5p.C
isomiR-30d-5p.A20G.CT
isomiR-25-3p.A22T
isomiR-let-7e-5p.G9C
isomiR-451a.T.A
isomiR-130b-5p.T
isomiR-142-3p.a
isomiR-let-7g-5p.t
isomiR-146a-5p.T22A
isomiR-144-3p
isomiR-425-5p.as
$\log_2(HR)$
D
PGB replication
miR-181
NfL median
isomiR-92a-3p.T1C
isomiR-92a-3p.T22C
isomiR-let-7i-5p.G5A
isomiR-let-7f-5p.T16C.t
isomiR-486-5p.T9G.T
isomiR-151a-3p.A9G
isomiR-432-5p
isomiR-486-5p.C
isomiR-30d-5p.A20G.CT
isomiR-25-3p.A22T
isomiR-let-7e-5p.G9C
isomiR-451a.T.A
isomiR-130b-5p.T
isomiR-142-3p.a
isomiR-let-7g-5p.t
isomiR-146a-5p.T22A
isomiR-144-3p
isomiR-425-5p.as
$\log_2(HR)$
E
UK discovery
χ^2 = 12.26, P-value = 4.6e-04
High Levels, Median Survival = 19.0
Low Levels, Median Survival = 11.0
Survival Probability
Months from enrollment
F
PGB replication
χ^2 = 19.76, P-value = 8.8e-06
High Levels, Median Survival = 31.0
Low Levels, Median Survival = 17.0
Survival Probability
Months from enrollment


A
UK discovery
χ^2 = 13.70, P-value = 2.1e-04
High Levels, Median Survival = 9.0
Low Levels, Median Survival = 16.2
Survival Probability
Months from enrollment
B
PGB Replication
χ^2 = 15.05, P-value = 1.0e-04
High Levels, Median Survival = 16.2
Low Levels, Median Survival = 25.9
Survival Probability
Months from enrollment
C
UK discovery
miR-181
isomiR-92a-3p.T22C
isomiR-let-7e-5p.G9C
isomiR-92a-3p.T1C
isomiR-25-3p.A22T
isomiR-486-5p.T9G.T
isomiR-let-7f-5p.T16C.t
isomiR-432-5p
isomiR-451a.T.A
isomiR-let-7i-5p.G5A
isomiR-425-5p.as
isomiR-151a-3p.A9G
isomiR-30d-5p.A20G.CT
isomiR-142-3p.a
isomiR-130b-5p.T
isomiR-144-3p
isomiR-146a-5p.T22A
isomiR-let-7g-5p.t
isomiR-486-5p.C
$\log_2(HR)$
D
PGB replication
miR-181
isomiR-92a-3p.T22C
isomiR-let-7e-5p.G9C
isomiR-92a-3p.T1C
isomiR-25-3p.A22T
isomiR-486-5p.T9G.T
isomiR-let-7f-5p.T16C.t
isomiR-432-5p
isomiR-451a.T.A
isomiR-let-7i-5p.G5A
isomiR-425-5p.as
isomiR-151a-3p.A9G
isomiR-30d-5p.A20G.CT
isomiR-142-3p.a
isomiR-130b-5p.T
isomiR-144-3p
isomiR-146a-5p.T22A
isomiR-let-7g-5p.t
isomiR-486-5p.C
$\log_2(HR)$
E
UK discovery
χ^2 = 20.62, P-value = 5.6e-06
High Levels, Median Survival = 16.0
Low Levels, Median Survival = 8.0
Survival Probability
Months from enrollment
F
PGB replication
χ^2 = 6.96, P-value = 0.008
High Levels, Median Survival = 26.0
Low Levels, Median Survival = 20.0
Survival Probability
Months from enrollment

	UK – discovery (N = 154)	PGB – replication (N = 200)	Overall (N = 354)
Sex:			
Males	91 (59%)	90 (45%)	181 (51%)
Females	63 (41%)	110 (55%)	173 (49%)
Age at symptom onset (years) (Mean ± SD)	62.3 (± 11)	57 (± 12)	59.3 (± 12)
Age at enrollment (years) (Mean ± SD)	63.9 (± 11)	58.4 (± 12)	60.8 (± 12)
Disease duration at enrollment (months) (Mean ± SD)	19.3 (± 10)	16.5 (± 8)	17.7 (± 9)
Baseline ALSFRS-R (AU) (Median [Q1, Q3])	38 [33, 43]	39 [35, 41]	39 [34, 42]
DeltaFRS (-ALSFRS-R unit / month) (Median [Q1, Q3])	0.6 [0.3, 0.9]	0.6 [0.4, 0.9]	0.6 [0.3, 0.9]
Site of symptom onset			
Bulbar	52 (34%)	55 (28%)	107 (30%)
Non-bulbar	102 (66%)	145 (72%)	247 (70%)
C9orf72 status			
Carrier	12 (8%)	20 (10%)	32 (9%)
Non-carrier	142 (92%)	180 (90%)	322 (91%)
Riluzole treatment			
Yes	109 (71%)	134 (67%)	243 (69%)
No	45 (29%)	66 (33%)	111 (31%)
Survival from symptom onset (months) (Median [Q1, Q3])	33 [26.7, 46.7]	33.4 [22.5, 43.9]	33 [23.7, 44.5]
Survival from enrollment (months) (Median [Q1, Q3])	13.9 [8.9, 23.9]	14.7 [8.5, 24.1]	14.3 [8.7, 24.1]
Censored event			
Yes	9 (6%)	108 (54%)	117 (33%)
No	145 (94%)	92 (46%)	237 (67%)