Non-Coding Genetic Analysis Implicates Interleukin 18 Receptor Accessory Protein 3′UTR in Amyotrophic Lateral Sclerosis

Authors:

Chen Eitan[1], Elad Barkan[2], Tsviya Olender[1], Kristel R. van Eijk[3], Matthieu Moisse[4,5], Sali M. K. Farhan[6,7], Aviad Siany[1], Shu-Ting Hung[8-10], Nancy Yacovzada[1,2], Johnathan Cooper-Knock[11], Kevin P. Kenna[3], Rick A. A. van der Spek[3], William Sproviero[12], Ahmad Al Khleifat[12], Alfredo Iacoangeli[12], Aleksey Shatunov[12], Ashley R. Jones[12], Elik Chapnik[1], Daphna Rothschild[2,13,14], Omer Weissbrod[2], Sebastian Werneburg[15], Dorothy P. Schafer[15], Robert H. Brown Jr[16], Pamela J. Shaw[11], Philip Van Damme[4,5,17], Leonard H. van den Berg[3], Hemali P. Phatnani[18], Eran Segal[2], Justin K. Ichida[8-10], Ammar Al-Chalabi[12,19], Jan H. Veldink[3], Project MinE ALS Sequencing Consortium[20], NYGC ALS Consortium[20] and Eran Hornstein[1]*

Affiliations:

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel.

[2]Department of Computer Science And Applied Math, Weizmann Institute of Science, Rehovot 7610001, Israel.

[3]Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, 3584 CG, The Netherlands.

[4]KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology, B-3000 Leuven, Belgium.

[5]VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium.

[6]Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA.

[7]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

[8]Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA.

[9]Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research at USC, Los Angeles, CA 90033, USA.

[10]Zilkha Neurogenetic Institute, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90033, USA.

[11]Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield S10 2HQ, UK.

[12]King's College London, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology & Neuroscience, De Crespigny Park, London, SE5 8AF, United Kingdom.

[13]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
[14]Department of Genetics, Stanford University, Stanford, CA 94305, USA

[15]Department of Neurobiology, Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA.

[16]Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01655, USA.

[17]University Hospitals Leuven, Department of Neurology, Leuven, Belgium.

[18]Center for Genomics of Neurodegenerative Disease, New York Genome Center.

[19]King's College Hospital, Denmark Hill, London, SE5 9RS, United Kingdom.

[20]A list of Consortiums PIs and affiliations appears in the Supplementary Information.

*Corresponding author. Tel: +972 89346215; Fax: +972 89342108; E-mail: Eran.hornstein@weizmann.ac.il

Abstract:

The non-coding genome is substantially larger than the protein-coding genome, but the lack of appropriate methodologies for identifying functional variants limits genetic association studies. Here, we developed analytical tools to identify rare variants in pre-miRNAs, miRNA recognition elements in 3′UTRs, and miRNA-target networks. Region-based burden analysis of >23,000 variants in 6,139 amyotrophic lateral sclerosis (ALS) whole-genomes and 70,403 non-ALS controls identified Interleukin-18 Receptor Accessory Protein (IL18RAP) 3′UTR variants significantly enriched in non-ALS genomes, replicate in an independent cohort and associate with a five-fold reduced risk of developing ALS. IL18RAP 3′UTR variants modify NF-κB signaling, provide survival advantage for cultured ALS motor neurons and ALS patients, and reveal direct genetic evidence and therapeutic targets for neuro-inflammation. This systematic analysis of the non-coding genome and specifically miRNA-networks will increase the power of genetic association studies and uncover mechanisms of neurodegeneration.

One Sentence Summary: Non-coding genetics demonstrate rare variants in IL-18 receptor 3'UTR that modifies ALS risk and progression.

[Main Text:]

Introduction

Genomic sequencing technologies facilitate identification of variants in open reading frames (ORFs). Although allelic variants in non-coding regions are expected to be numerous [1, 2] they are largely overlooked because current analytical approaches do not adequately prioritize variants that are likely to be functional, over neutral background variation.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative syndrome, primarily affecting the human motor neuron system with a strong genetic predisposing component [3, 4]. Thus, mutations in approximately 25 protein-coding genes have been associated with ALS [3, 5, 6] and a hexanucleotide repeat expansion in an intronic sequence of the *C9orf72* gene is the most common genetic cause of ALS [7-9]. However, non-coding nucleotide variants in ALS have yet to be systematically explored.

MicroRNAs (miRNAs) are endogenous posttranscriptional repressors that silence mRNA expression through sequence complementarity. miRNA dysregulation has been implicated in ALS pathogenesis, and ALS-associated RNA-binding proteins, *TARDBP*/TDP-43 and FUS, regulate miRNA biogenesis [10-21]. miRNA primarily act on 3′ untranslated regions (3′UTRs) [22], which are non-coding parts of messenger RNAs (mRNAs) and often regulate degradation and translation [23].

Burden analysis is a genetics approach that is based on the rationale that different rare variants in the same gene may have a cumulative contribution [24]. Therefore, burden analysis allows the identification of genes containing an excess of rare and presumably functional variation in cases relative to controls. Although de novo mutations in non-coding regions were recently shown in family-based autism studies [25], variants in non-coding regions are not routinely included in rare-variant burden association studies. The application of burden analysis to non-coding regulatory variation is constrained by the availability of whole genome sequencing (WGS) data, and the ability to recognize functional variants in non-coding regulatory regions, which is currently far less effective than for protein-coding genes.

To effectively apply burden analyses to regulatory variation, appropriate ways to determine what are the relevant (qualifying) variants should be determined. In the case of miRNAs and miRNA recognition elements such a framework can be achieved with relatively high confidence because of miRNA high

conservation and wealth of mechanistic insight about sequence impact on function. Therefore, miRNA regulatory networks make an excellent gene set to explore.

Here, we developed tools that identify, or call, qualifying variants in miRNAs and 3′UTR of mRNAs, and performed collapsed genetic analysis [26], to test if these regulatory RNAs are associated with ALS. We discovered an enrichment of rare variants in the IL18RAP 3′UTR, implicating the IL-18 pathway in ALS. Non-coding variants analysis in miRNA networks may impact research of human traits, increase the power of rare-variant association methods and encourage systematic exploration of non-coding regions, to uncover genetic mechanisms of disease.

Results

To test whether genetic variations in non-coding regulatory regions are associated with ALS, we analyzed regions of interest WGS data from the Project MinE ALS sequencing consortium [27] (Supplementary Fig. 1A,B and Supplemental Tables 1,2). The discovery cohort consisted of 3,955 ALS patients and 1,819 age- and sex-matched controls, for a total of 5,774 whole-genomes from the Netherlands, Belgium, Ireland, Spain, United Kingdom, United States and Turkey (Project MinE Datafreeze 1). We tested 295 genes, including candidates from sporadic ALS GWAS [28] or encoding RNA-binding proteins, and analyzed both their 3′UTRs and open reading frames (Supplementary Table 3). In addition, we tested all 1,750 human precursor miRNA genes (pre-miRNAs; miRBase v20) [29].

We devised a method for identifying rare genetic variants with minor allele frequencies (MAF) ≤0.01, in non-coding RNA regulation that: (1) abrogate miRNA recognition elements in 3′UTRs; or (2) result in a predicted *de novo* gain of miRNA binding; or (3) change the sequence of mature miRNAs and/or pre-miRNAs (miRBase v20[29]). We performed region-based burden test, in which variants within miRNA recognition elements in 3′UTRs and within pre-miRNA genes that satisfy above criteria (qualifying variants), were binned together to weight their contribution to disease.

We also identified rare variants in open reading frame of the 295 miRNA-relevant or ALS-relevant protein coding genes, that are predicted to cause frameshifting, alternative splicing, an abnormal stop codon, or a deleterious non-synonymous amino acid substitution that were detected in ≥ 3 of 7 independent dbNSFP prediction algorithms [30] (Fig. 1A and Supplementary Table 3). In total 28,211 rare qualifying variants were identified (Supplementary Table 4).

As positive control we performed an association analysis of rare variants in protein-coding sequences, using the Optimized Sequence Kernel Association Test (SKAT-O) [31]. SKAT-O identified a significant excess of deleterious minor alleles in the ALS genes *NEK1* (127 cases; 19 controls [3.21%; 1.04%]: $P = 7.04\times10^{-7}$; $P_{corrected} = 2.13\times10^{-4}$), comparable with a reported prevalence of 3% [32], and *SOD1* (36 cases [0.91%]; 0 controls: $P = 2.61\times10^{-4}$; $P_{corrected} = 3.76\times10^{-2}$) [33], which is below the reported 2% prevalence [5, 34] (Fig. 1B, Supplementary Fig. 2A and Data File S1). Other known ALS genes did not reach statistical significance (Supplementary Table 3), consistent with reported statistical power limitations of Project MinE WGS data in assessing the burden of rare variants [35]. Our analysis did not consider *C9orf72* hexanucleotide (GGGGCC) repeat expansion region.

We determined the burden of rare variants in all autosomal pre-miRNAs in the human genome (1,750 genes). We did not identify disease association for any pre-miRNAs, nor for any of the predicted genetic networks based on variants aggregated over specific mature miRNAs and their cognate downstream 3'UTR targets. This may be because the small size of miRNA genes makes genetic aggregation studies particularly challenging (Supplementary Fig. 2B,C).

Finally, we tested the burden of variants that are potentially either abrogating conserved miRNA binding sites or creating new miRNA binding sites in 3′UTRs. The strongest association was for the 3′UTR of *IL18RAP* (Fig. 1B, Supplementary Fig. 2D and Data File S1). This association was higher than expected at random (P = 3.34×10^{-5}, $P_{corrected} = 9.31 \times 10^{-3}$) and from the association gained for all protein-coding ALS genes in this cohort, with the exception of *NEK1*. Notably, the signal was more prevalent in controls [9/1819, 0.49%] relative to ALS patients [4/3955, 0.10%], indicating that these variants are protective against ALS.

Because the number of ALS genomes was ~2.17-fold larger than the number of controls, the data depict a 4.89-fold enrichment in the abundance of variants in controls over cases. *IL18RAP* 3′UTR protective variants reduced the disease odds ratio by five-fold (OR = 0.20; Fig. 2A), and was consistent across independent population strata (Fig. 2B), whereas *NEK1* and *SOD1* increased the disease odds ratio (OR = 3.14, 33.89, respectively; Fig. 2A).

IL18RAP 3′UTR also ranked as the top hit when we relaxed the analysis by including all 3′UTR variants, regardless of predicted miRNA recognition elements. Therefore, the robust association of *IL18RAP* 3′UTR is independent of the assumptions about specific miRNA binding (SKAT-O P = 1.88×10^{-5}, $P_{corrected} = 5.62 \times 10^{-3}$; variant in controls [12/1819, 0.66%], cases [6/3955, 0.15%], OR = 0.23; Fig. 2A,C, Fig. 3A, Supplementary Table 5 and Data File S1). Three other algorithms – the Sequence Kernel Association Test (SKAT, P = 1.73×10^{-5}; permutated P-value < 10^{-4}), the Combined Multivariate and Collapsing (CMC, P = 8.66×10^{-4}) or Variable Threshold (VT) with permutation analysis (permutated P-value = 2.10×10^{-3}) – all ranked the *IL18RAP* 3′UTR association above any other studied 3'UTR, suggesting that the association does not depend on a particular statistical genetics method (Supplementary Fig. 3). Together, rare variants in *IL18RAP* 3'UTR are depleted in ALS patients, suggesting that they are protective against ALS.

To determine if the rare *IL18RAP* 3′UTR variants are depleted in another ALS cohort, we performed independent replication studies. Similar results for rare *IL18RAP* 3'UTR variants were reproduced in the

New York Genome Center (NYGC) ALS Consortium cohort (2,184 ALS genomes), which was studied against: (i) 263 non-neurological controls from the NYGC; (ii) 62,784 non-ALS genomes from NHLBI's Trans-Omics for Precision Medicine (TOPMed); and (iii) 5,537 non-ALS genomes from gnomAD. This replication effort yielded a joint analysis P-value = 9.58×10^{-4} ($\chi 2$ with Yate's correction; OR=0.32; 95% CI: 0.16 – 0.64; Fig. 2C and Supplementary Table 6). Combining this cohort with our discovery cohort from Project MinE, yielded a superior joint P-value < 1.00×10^{-5} ($\chi 2$ with Yate's correction; OR=0.20; 95% CI: 0.12 – 0.34; Fig. 2C). A meta-analysis of Project MinE datafreeze 1 and 2, which consisted of 5,185 ALS patients and 2,262 age- and sex-matched controls, reproduced the initial signal (p= 7.6×10^{-4}).

Together, *IL18RAP* 3'UTR sequence variants contribute to a lower risk of suffering from ALS, that is approximately one fifth of the general population, although it did not reach conventional exome-wide multiplicity-adjusted significance threshold ($\alpha \approx 2.6 \times 10^{-6}$, ref. [24]) in our study.

Finally, genome-wide analysis of all known human 3'UTRs (RefSeq, [36]) identified *IL18RAP* 3'UTR as the most significant 3'UTR associated with ALS, in the Project MinE cohort (Fig. 3B), followed by the *GPATCH8* 3'UTR (SKAT-O P = 1.92×10^{-5}, $P_{corrected}$ =0.16; variant in controls [63/1819, 3.46%], cases [68/3955, 1.72%], OR = 0.49), the *CDC14B* 3'UTR (SKAT-O P = 3.64×10^{-5}, $P_{corrected}$ =0.19; variant in controls [112/1819, 6.16%], cases [163/3955, 4.12%], OR = 0.66), and the *RAB3GAP2* 3'UTR (SKAT-O P = 4.51×10^{-5}, $P_{corrected}$ =0.19, variant in controls [147/1819, 8.08%], cases [485/3955, 12.26%], OR = 1.59). *GPATCH8* is involved in hyperuricemia pathophysiology, *CDC14B* is a dual-specificity phosphatase involved in the DNA damage response, and *RAB3GAP2* is involved in neurotransmitter and hormone exocytosis and highly expressed in the brain, however, their potential role in neurodegeneration is unknown.

To investigate the source of the signal in the *IL18RAP* 3′UTR in a post-hoc analysis, we divided the 11 variant nucleotides into two synthetic sets, of either nine singleton variants (9 variants / 3 controls / 6 patients) or two variants that were identified solely in controls (2 variants / 9 controls / 0 patients). While the signal of the nine singleton variants was not statistically significant, analysis of the two control variants, which were identified in multiple samples, derived an improved significance compared to the original signal (SKAT-O P = 4.36×10^{-6}). Thus, these two rare variants (V1, Chr2:103068691 C>T; V3, Chr2:103068718 G>A) are likely central in generating the genetic association signal in *IL18RAP* 3′UTR.

IL18RAP is a receptor subunit which dimerizes with IL18R1 upon binding of the interleukin IL-18. IL18 receptor is expressed in T-cells, neurons, astrocytes, and microglia [37] and induces NF-κB signaling. To determine the functional impact of the *IL18RAP* variants we analyzed *IL18RAP* expression in lymphoblastoid cell lines (LCLs) from the UK MNDA DNA Bank [38] that were originally derived from two different individuals, one carrying the putative *IL18RAP* protective variant (V3, Chr2:103068718 G>A) and the other carrying the canonical *IL18RAP* 3'UTR (control). LCLs harboring the *IL18RAP* 3'UTR protective variant significantly down-regulated IL18RAP protein expression (Fig. 4A and Data File S2), with p-NF-κB protein levels also being significantly reduced (Fig. 4B and Data File S3). Therefore, a variant form of *IL18RAP* 3'UTR attenuates its endogenous expression and downstream NF-κB signaling.

To further establish the functional relevance of the *IL18RAP* 3'UTR variants, we subcloned wild-type *IL18RAP* 3'UTR (WT) and the two most prevalent 3'UTR variants, (Chr2:103068691 C>T (V1) and Chr2:103068718 G>A (V3)), downstream of a Renilla luciferase reporter (hRluc). Variants V3 and V1 reduced luciferase activity by ~33% and ~30%, respectively, relative to the WT *IL18RAP* 3'UTR (Fig. 4C,D). Thus, the protective *IL18RAP* variants regulate *IL18RAP* mRNA expression.

To determine the ability of the *IL18RAP* variants V3 and V1 to induce NF-κB activity, we co-transfected U2OS cells with different *IL18RAP* coding region (CDS) and 3'UTR constructs, along with an NF-κB activity reporter that drives luciferase (Luc2P) transcription via five copies of the NF-κB response element. NF-κB signaling was induced by adding a human recombinant IL-18 to the medium. Variants V3 and V1 of the *IL18RAP* 3'UTR reduced NF-κB activity by ~10% and ~21%, respectively, relative to the WT *IL18RAP* 3'UTR (Fig. 4E,F). GFP vector and a dominant negative coding mutant E210A-Y212A-Y214A CDS + WT 3'UTR (3CDS) [39], served as controls. We conclude that V3 and V1 3'UTR variants affect IL18RAP capacity to induce NF-κB signaling, although physiological relevance cannot be elucidated from a reporter assay.

Levels of the cytokine IL-18 are elevated in ALS patient tissues and biofluids [40-42], but the expression in motor neurons is not characterized. To study *IL-18* and *IL18RAP* expression levels specifically in human motor neurons of patients with ALS, we mined human NGS data. *IL18RAP* and *IL-18* mRNA expression are higher in laser capture microdissection– enriched surviving motor neurons from lumbar spinal cords of patients with sALS with rostral onset and caudal progression, relative to non-neurodegeneration controls (Supplementary Fig. 4A,B; data from ref. [43]). Consistently, higher *IL-18* mRNA levels were also

in induced human motor neurons of patients with ALS (Supplementary Fig. 4C; data from ref. [44])). Thus, *IL-18* and *IL18RAP* receptor subunit are abnormally high in human ALS motor neurons.

To study the impact of *IL18RAP* 3'UTR mutations in a model of human ALS motor neurons, we performed a survival analyses using induced pluripotent stem cells (iPSCs) derived from human ALS patients that harbor a *C9orf72* hexanucleotide repeat expansion [45]. iPSCs from patients and from healthy controls were differentiated to Hb9::RFP+ human motor neurons (iMNs) [45] and time-lapse microscopy was used to quantify their subsequent survival after withdrawal of neurotrophic factors and in the presence of the cytokine IL-18 (Fig. 5A). As expected, degeneration (cellular death) was significantly more severe in *C9orf72* patient iMNs than in iMNs derived from healthy controls (n=2 patients, n=2 controls, iMNs from each genotype combined into a single trace for clarity; Fig. 5B). Notably, this was partially rescued by transducing *C9orf72* iPSCs with the *IL18RAP* 3'UTR variant V3 (including CDS) expression vector, which ameliorated motor neuron toxicity, relative to wild-type *IL18RAP* 3'UTR. This was similar to the effect of a *IL18RAP* dominant negative coding mutant, 3CDS (E210A-Y212A-Y214A; one-sided log-rank test, $P<0.05$, n=2 patients, n=2 controls, Fig. 5B-D). The V1 variant had a more limited effect, which might be related to activity of V1 in cells other than motor neurons. Additionally, *C9orf72* iMNs harboring the V3 variant displayed a lower hazard ratio (cellular death propensity) than *C9orf72* iMNs harboring the wild-type *IL18RAP* 3'UTR, (n=2 patients, n=2 controls (Fig. 5E). Thus, both IL18RAP and its ligand IL-18 are upregulated in human ALS motor neurons and a rare variant of *IL18RAP* 3'UTR confers protection to human *C9orf72* motor neurons in tissue culture.

To determine whether the mutant *IL18RAP* 3'UTR is also protective in human patients with ALS, we tested the association between age of diagnosis and age of death in ALS patients harboring wild-type or variants of the *IL18RAP* 3'UTR. Of 4216 patients for whom data on age of diagnosis was available (Project MinE and NYGC cohorts), 8 harbored *IL18RAP* 3'UTR variants. Of 4263 patients for whom age of death was available, 9 harbored *IL18RAP* 3'UTR variants. *IL18RAP* variants are expected to be depleted in ALS genomes, nonetheless in those extremely rare patients harboring *IL18RAP* 3'UTR variants, these were associated with an older age of death, on average 6.1 years after the average for patients with canonical *Il18RAP* 3'UTR (one-sided Mann-Whitney test $P = 0.037$; Fig. 5F), and an older age of diagnosis, on average 6.2 years after the average for patients with canonical *Il18RAP* 3'UTR (one-sided Mann-Whitney test $P = 0.06$; Fig. 5G). Thus, variants in *IL18RAP* 3'UTR are protective against ALS and provide survival advantage for patients suffering from the disease.

Discussion

Data from the ALS consortia used in this study, Project MinE and NYGC, provide unprecedented opportunities for investigating the role of the non-coding genome in ALS, and will drive the development of computational methodologies for weighting the effect of variants outside of protein open reading frames. By identifying qualifying variants and performing rare variant aggregation analysis in 1,750 pre-miRNA genes, and 295 protein coding ORFs and their 3′UTRs, we demonstrated that variants in the 3′UTR of *IL18RAP* are enriched in non-ALS genomes, indicating that these are relatively depleted in ALS. *IL18RAP* 3′UTR variants reduced the chance of the disease five-fold, increased the survival of human *C9orf72* motor neurons and delayed onset and therefore age of death in people with ALS. The discovery of functional, disease-modifying *IL18RAP* 3'UTR variants underscores the need to explore the role of additional non-coding genomic regions in ALS.

Protective protein-coding variants have been identified in Alzheimer's disease [46-49] and implicated in ALS as well [50, 51] and deleterious variants were suggested in VEGF promoter/5'UTR [52]. However, the 3′UTR of *IL18RAP* is the first protective non-coding allele associated with a neurodegenerative disease.

Neuro-inflammation is prevalent in neurodegeneration, including in ALS [53], and is often characterized by the activation of microglia, astrocytes, and the accumulation of infiltrating T-cells at sites of neurodegeneration [54-57]. The soluble ligand, IL-18, is part of this neuro-inflammatory milieu, promoting receptor subunit (IL18RAP, IL18R1) dimerization on the membrane of T-cells, neurons, astrocytes, and microglia [37], and activating intracellular signaling cascades, including NF-κB. Polymorphic forms of IL18RAP are genetically associated with autoimmune / inflammatory diseases [58-62], suggesting that perhaps changes to IL18RAP via its 3'UTR, alter ALS risk or severity in a dose or expression-dependent manner.

Our study directly links dysregulation of IL18RAP signaling to ALS. The protective effect of IL18RAP 3'UTR variants causes downregulation of IL-18 signaling, which might be hyperactive in ALS motor neurons and is supported by previous observations showing that ALS patients have elevated levels of the cytokine IL-18 in tissues and biofluids [40-42] and that IL-18 secretion is triggered from microglia in a model of TDP-43 proteinopathy [63]. Our data from human motor neurons further suggest that IL18RAP may be acting directly on motor neurons. Therefore, variants in *IL18RAP* 3′UTR may modify IL-18 signaling in

the central nervous system of ALS patients. However, the regulatory changes affected by the *IL18RAP* 3′UTR variants remains to be elucidated.

A limitation of our study is that only 295 candidate genes were initially tested. However, the key findings were reproduced in a genome-wide study of all human 3'UTRs. While *IL18RAP* 3'UTR signal is comparable to that of protein-coding ALS-causing genes, such as *SOD1* and *NEK1*, limitations in the statistical power may be overcome with larger ALS and control cohorts, which are not currently available. Furthermore, the genetic involvement of *IL18RAP* 3'UTR in other neurodegenerative diseases remains to be explored. Finally, the mechanism underlying *IL18RAP* dose sensitivity is not fully understood. While we provide evidence that *IL18RAP* 3'UTR endows neuroprotection to human motor neurons and is associated with survival advantage in humans with ALS, additional studies should explore the mechanism by which IL18RAP protects motor neurons and the degree to which other cell types, such has microglia, are involved.

In summary, we have identified the *IL18RAP* 3′UTR as a non-coding genetic disease modifier by burden analysis of WGS data using ALS case-control cohorts. We show that IL-18 signaling modifies ALS susceptibility and progression, delineating a neuro-protective pathway and identifying potential therapeutic targets for ALS. Whereas the 3′UTR of *IL18RAP* is the first protective non-coding allele associated with a neurodegenerative disease, the increasing wealth of WGS data in Project MinE, NYGC and elsewhere, indicates that the exploration of non-coding regulatory genomic regions should reveal further disease-relevant genetic mechanisms.

Methods

Human genetic cohorts

All participants contributed DNA after signing informed consent at the submitting sites. Human materials were studied under approval of the Weizmann Institute of Science Institutional Review Board (Weizmann IRB: 1039-1).

Discovery cohort: Project MinE ALS sequencing consortium Datafreeze 1 includes 3,955 ALS patients and 1,819 age- and sex-matched controls, free of any neurodegenerative disease, for a total of 5,774 quality control (QC) passing whole-genomes, from the Netherlands, Belgium, Ireland, Spain, United Kingdom, United States and Turkey. Rare variant association in cases versus controls was evaluated for regions of interest, when we could identify ≥2 variants per region, by SKAT-O, SKAT, CMC and VT in RVTESTS environment [64], with sex and the top 10 principal components (PCs) as covariates. To construct the PCs of the population structure, an independent set of ~450,000 SNPs was sampled from WGS, (MAF ≥0.5%) followed by LD-pruning.

Replication cohorts: Utilized for testing rare variant alleles (MAF < 0.01) in human *IL18RAP* 3'UTR (GRCh37/hg19 chr2:103068641-103069025 or GRCh38 chr2:102452181-102452565) from Project MinE datafreeze 2: ~1300 European heritage ALS genomes without middle eastern (Turkish and Israelis) genomes. The New York Genome Center (NYGC) ALS Consortium (2,184 ALS Spectrum MND and 263 non-neurological control genomes from European/Americas ancestries), NHLBI's Trans-Omics for Precision Medicine (TOPMed; 62,784 non-ALS genomes) and gnomAD (5,537 non-ALS genomes; Europeans, non-Finnish, non-TOPMed). Joint analysis in replication cohort, was performed by Chi square test with Yate's correction. Meta-analysis was not possible because TOPMed and gnomAD covariate information is not available.

Quality control procedures in Project MinE genomics

Sample selection, data merging and sample- and variant level quality control procedures for Project MinE ALS sequencing consortium genomes are described in full previously [35]. Briefly, 6,579 Project MinE ALS sequencing consortium whole genomes sequenced on Illumina HiSeq2000 or HiSeqX platforms. Reads were aligned to human genome build hg19 and sequence variants called with Isaac Genome Alignment

Software and variant caller [65]. Individual genomic variant call format files (GVCFs) were merged with 'agg' tool: a utility for aggregating Illumina-style GVCFs. Following completion of the raw data merge, multiple QC filtering steps were performed: (i) setting genotypes with GQ<10 to missing; (ii) removing low-quality sites (QUAL< 30 and QUAL< 20 for SNPs and indels, respectively); (iii) removing sites with missingness > 10%. (iv) Samples excluded if deviated from mean by more than 6SD for total numbers of SNPs, singletons and indels, Ti/Tv ratio, het/hom-non-ref ratio and inbreeding (by cohort). (v) missingness > 5%, (vi) genotyping-sequence concordance (made possible by genotyping data generated on the Illumina Omni 2.5M SNP array for all samples; 96% concordance), (vii) depth of coverage, (viii) a gender check (to identify mismatches), (ix) relatedness (drop samples with >100 relatedness pairs). (x) Related individuals were further excluded until no pair of samples had a kinship coefficient > 0.05. (xi) missing phenotype information. Following QC, 312 samples with expended/inconsistent *C9orf72* status were omitted from further analysis. A total of 5,774 samples (3,955 ALS patients and 1,819 healthy controls) passed all QC and were included in downstream analysis. Per-nucleotide site QC was performed on QC-passing samples only, for Biallelic sites: variants were excluded from analysis based on depth (total DP < 10,000 or > 226,000), missingness > 5%, passing rate in the whole dataset < 70%, sites out of Hardy–Weinberg equilibrium (HWE; by cohort, controls only, $p < 1x10^{-6}$) and sites with extreme differential missingness between cases and control samples (Overall and by cohort, $p < 1x10^{-6}$). Non-autosomal chromosomes and multiallelic variants were excluded from analysis.

Selection of regions of interest

Discontinuous regions of interest approximating in total ~5Mb, include coding sequences and 3′ untranslated regions (3′UTRs) of 295 genes (Supplementary Table 3) encoding for proteins that were: (i) previously reported to be associated with ALS, (ii) RNA-binding proteins including miRNA biogenesis or activity factors [UCSC gene annotation; [66]]. In addition to (iii) all 1,750 human pre-miRNA genes [miRBase v20; [29]]. In addition, genome-wide analysis of all known human 3'UTRs (RefSeq [36]). Variants in regions of interest were extracted from Project MinE ALS sequencing consortium genomes using vcftools [67] according to BED file containing genomic coordinates of interest (hg19) ±300 bp that ensures covering splice junctions and sequence (Supplementary Table 7).

Annotation and burden analysis

After quality control and extraction of regions of interest we performed functional annotation of all variants. Indels were left-aligned and normalized using bcftools and multiallelic sites were removed. For variant annotation we developed a pipeline that calculates the impact of genetic variation in coding regions as well as in 3'UTR and miRNA regions, using ANNOVAR [68]. The frequency of the variants in the general population was assessed by screening the 1000 Genomes Project, the Exome Aggregation Consortium (ExAC) and NHLBI Exome Sequencing Project (ESP). For protein coding ORFs, association analysis of deleterious rare variants was performed, i.e., frameshift variants, deviation from canonical splice variant, stop gain/loss variants or a non-synonymous substitution, as predicted by at least three prediction programs (SIFT, Polyphen2 HVAR, LRT, MutationTaster, MutationAssessor, FATHMM, MetaLR) in dbNSFP environment [v2.0; [30]].

Non-coding sequence burden analysis included (i) 3′UTRs, (ii) variants in miRNA recognition elements in 3′UTRs (Supplementary Table 3): Variants that impaired conserved-miRNA binding sites in 3′UTRs (predicted loss of function) were called by TargetScan [v7.0; [69]]. Newly created miRNA binding sites in 3′UTRs (predicted gain of function) were called by textual comparison of all possible mutated seeds around a variant to all known miRNA seed sequences in the genome, (iii) all human pre-miRNAs (mirBase v20 [29]) and (iv) miRNAs:target gene networks: mature miRNA sequences (mirBase v20 [29]) and cognate targets within the 3'UTRs (Supplementary Table 3).

Mammalian Cell Cultures

Lymphoblastoid cell lines (LCLs) from the UK MNDA DNA Bank [38] were originally derived from an individual carrying the suggested *IL18RAP* protective variant (V3, Chr2:103068718 G>A) and another individual with the canonical *IL18RAP* 3'UTR (Weizmann IRB: 537-1). LCLs were cultured in RPMI-1640 (Gibco, 21875091) with 20% inactivated fetal bovine serum (FBS, Biological Industries, 04-001-1A), 1% L-glutamine and 1% penicillin-streptomycin (Biological Industries, 03-0311B) at 37°C, 5% CO2. Human Bone Osteosarcoma Epithelial Cells (U2OS), were maintained in Dulbecco's Modified Eagle Medium (DMEM, Biological Industries, 01-050-1A) supplemented with 10% FBS, 1% penicillin-streptomycin at 37°C, 5% CO2.

Cloning

Full *IL18RAP* coding sequence (CDS) and 3'UTR sequence (2223bp) in pMX vector was purchased from GeneArt (Invitrogen, Supplementary Table 8) and subcloned with V5 epitope into pcDNA3. Different mutants, including: WT *IL18RAP* CDS + mutant 3'UTR (V1 or V3), and a dominant negative coding mutant E210A-Y212A-Y214A CDS + WT 3'UTR (3CDS) [39] created by Transfer-PCR mutagenesis[70]. Next, WT and mutants full *IL18RAP* were subcloned into pUltra vector (a gift from Malcolm Moore, Addgene plasmid #24130, for which mCherry was replaced with EGFP), downstream of the human Ubiquitin C promoter and EGFP-P2A. Human *IL18RAP* 3'UTR sequences (422bp) in pMX vector were purchased from GeneArt (Invitrogen, Supplementary Table 8) and subcloned into hRluc reporter in psiCHECK-2 vector (Promega). All cloning procedures were done via restriction free cloning [71]. List of primers used for cloning and Transfer-PCR mutagenesis described in Supplementary Table 9.

Transfection and luciferase assays

Transfection to U2OS Cells at 1.9 cm^2 Corning plates was performed at 70–80% confluence, 24 h post plating in antibiotic- free media, using Lipofectamine 2000, 0.5 µL per well (Thermo Fisher Scientific, Cat# 11668027). Each well was considered as a single replicate. miRNA sensor: U2OS cells were harvested 72 h post-transfection with human *IL18RAP* 3'UTR downstream to hRluc reporter (psiCHECK-2 vector 500 ng / 1.9 cm^2 plate), for Dual luciferase reporter assay (Promega). NF-κB reporter assay: U2OS cells were induced with recombinant IL-18 (5ng/ml) 72 h post-transfection with full coding sequence of *IL18RAP* coding region + 3'UTRs (pUltra vector 500 ng / 1.9 cm^2 plate), luc2P/NF-κB-RE (pGL4.32 100 ng) luciferase and Renilla luciferase (hRluc 10 ng). 6 h post later cells were harvested and luminescence quantified.

Cell lysis and Western blot

LCLs were washed in PBSx1, centrifuged at 800 × *g* for 5 min at 4°C, pelleted and lysed in ice-cold RIPA buffer supplemented with cOmplete™ Protease Inhibitor Cocktail (Roche, 4693116001) and PhosSTOP™ (Roche, 4906837001). The lysates were cleared by centrifugation at 15,000 × g for 10 min at 4°C. Plasma Membrane Protein Extraction Kit (abcam, ab65400) was used for extraction of membrane-bound IL18RAP.

Protein concentrations quantified with Protein Assay Dye Reagent (Bio-Rad, 500-0006), resolved at 50µg of total protein/well by 10% polyacrylamide / SDS gel electrophoresis at 100-120 V for 70 min. After gel electrophoresis and transferred to nitrocellulose membrane (Whatmann, 10401383) at 250mA for 70 min. Membranes were stained with Ponceau (Sigma, P7170), blocked for 1 hour at RT with 3% Bovine albumin fraction V (MPBio 160069) or 5% milk protein in PBST (PBS containing 0.05% TWEEN-20) and then incubated with primary antibodies [Rabbit anti IL-18R Beta antibody (Bioss, catalog# BS-2616R, 1:500), mouse anti GAPDH (Thermo Fisher, catalog# AM4300, 1:5000), mouse anti p-NFκB p65 antibody (Santa Cruz, catalog# sc-135769, 1:200), Mouse Anti-beta Actin antibody [AC-15] (abcam, catalog# ab6276, 1:9,000)] O.N. at 4°C with rocking in antibody-Solution [5% albumin, 0.02% sodium azide, 5 drops of phenol red in 0.05% PBST]. Following primary antibody incubation, membranes were washed 3 times for 5 min at RT with 0.05% PBST then incubated for 1 hour at RT with horseradish peroxidase (HRP)-conjugated species-specific secondary antibodies, washed 3 x 5 min in 0.05% PBST at RT and visualized using EZ-ECL Chemiluminescence (Biological Industries, 20500-120) by ImageQuant™ LAS 4000 (GE Healthcare Life Sciences). Densitometric analysis performed using ImageJ (NIH).

Induced neuron survival assay

Survival assay of *Hb9::*RFP+ iMNs was conducted as described previously [45], with the following modifications: (i) iMNs were infected on day 2 with lentiviruses expressing the full pUltra-IL18RAP constructs, and (ii) longitudinal microscopic tracking was performed every 48 hours, following neurotrophic factor withdrawal and IL-18 treatment (10ng/mL), starting on day 15 for 18 additional days. iMN survival assays were performed using three individual replicates / line / condition. iMNs were from two independent donors for each genotype (CTRL/C9-ALS) were combined into one survival trace in the Kaplan-Meier plots for clarity. Ichida lab human lymphocytes from healthy subjects and ALS patients were obtained from the National Institute of Neurological Disorders and Stroke (NINDS) Biorepository at the Coriell Institute for Medical Research and reprogrammed into iPSCs as previously described [45]. The NINDS Biorepository requires informed consent from patients. The experiment involved mouse glial isolation, performed at University of Southern California (USC) and was done in compliance with ethical regulations approved by the USC IACUC committee (Los Angeles, USA).

Statistical analysis

Statistics performed with Prism Origin (GraphPad). Shapiro-Wilk test was used to assess normality of the data. Pairwise comparisons passing normality test were analyzed with Student's *t*-test, whereas the Mann-

Whitney test was used for pairwise comparison of nonparametric data. Multiple group comparisons were analyzed using ANOVA with post hoc tests. For iMN survival experiments, statistical analysis was performed using a one-sided log-rank test to account for events that did not occur (i.e. iMNs that did not degenerate before the end of the experiment). For each line, the number of iMNs that were analyzed to generate the survival curve is indicted in the figure. Statistical P values <0.05 were considered significant. Data are shown as box plots, or as noted in the text.

Supplementary Materials

Fig. S1. Study design.

Fig. S2. Region-based rare-variant association analyses.

Fig. S3. 3′UTR-based rare-variant association analysis, using different algorithms.

Fig. S4. Evaluation of IL18RAP and IL-18 mRNA expression in motor neurons of patients with ALS.

Table S1. Total number of samples before and after quality control procedures, stratified by country.

Table S2. Samples Quality Control Procedures.

Table S3. Candidate genes list.

Table S4. Number of rare qualifying genetics variants identified.

Table S5. Identified *IL18RAP* 3′UTR variants in Project MinE discovery cohort.

Table S6. Identified *IL18RAP* 3′UTR variants in discovery and replication cohorts.

Table S7. BED file containing genomic coordinates of regions of interest.

Table S8. Synthetic *IL18RAP* sequences used for cloning into pMX vectors.

Table S9. List of primers used for cloning and Transfer-PCR mutagenesis.

Data File S1. Detailed description of variants in protein coding sequences of *NEK1* and *SOD1* and the *IL18RAP* 3′UTR, in Project MinE discovery cohort.

Data File S2. Source data for IL18RAP western blot studies.

Data File S3. Source data for p-NF-κB western blot studies.

Project MinE ALS Sequencing Consortium PI List

NYGC ALS Consortium PI List

Acknowledgments:

We gratefully acknowledge the contributions of all participants and the investigators who provided biological samples and data for Project Mine ALS sequencing consortium, the New York Genome Center (NYGC) ALS Consortium, the Genome Aggregation Database (gnomAD) and Trans-Omics for Precision Medicine (TOPMed) of the National Heart, Lung, and Blood Institute (NHLBI, https://www.nhlbiwgs.org/topmed-banner-authorship). Samples used in this research were in part obtained from the UK National DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust. We acknowledge sample management undertaken by Biobanking Solutions funded by the Medical Research Council at the Centre for Integrated Genomic Medical Research, University of Manchester. The authors would like to thank the NINDS Biorepository at Coriell Institute for providing the cell lines used for this study at J.K.I. lab. We thank LSE for language and scientific editing. Hornstein lab is supported by friends of Dr. Sydney Brenner. EH is Head of Nella and Leon Benoziyo Center for Neurological Diseases and incumbent of Ira & Gail Mondry Professorial chair. **Funding:** The work is funded by Legacy Heritage Fund, Bruno and Ilse Frick Foundation for Research on ALS, Teva Pharmaceutical Industries Ltd as part of the Israeli National Network of Excellence in Neuroscience (NNE) and Minna-James-Heineman Stiftung through Minerva. The research leading to these results has received funding to E.H. from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 617351. Israel Science Foundation, the ALS-Therapy Alliance, AFM Telethon (20576 to E.H.), Motor Neuron Disease Association (UK), The Thierry Latran Foundation for ALS research, ERA-Net for Research Programmes on Rare Diseases (FP7), A. Alfred Taubman through IsrALS, Yeda-Sela, Yeda-CEO, Israel Ministry of Trade and Industry, Y. Leon Benoziyo Institute for Molecular Medicine, Kekst Family Institute for Medical Genetics, David and Fela Shapell Family Center for Genetic Disorders Research, Crown Human Genome Center, Nathan, Shirley, Philip and Charlene Vener New Scientist Fund, Julius and Ray Charlestein Foundation, Fraida Foundation, Wolfson Family Charitable Trust, Adelis Foundation, MERCK (UK), Maria Halphen, Estates of Fannie Sherr, Lola Asseof, Lilly Fulop. To A.A.C. from Neurodegenerative Disease Research (JPND), Medical Research Council (MR/L501529/1; STRENGTH, MR/R024804/1; BRAIN-MEND), Economic and Social Research Council (ES/L008238/1; ALS-CarE)), MND Association. National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. To P.V.D.: Project MinE Belgium was supported by a grant from IWT (n° 140935), the ALS Liga België, the National Lottery of Belgium and the KU Leuven Opening the

Future Fund. P.V.D. holds a senior clinical investigatorship of FWO-Vlaanderen and is supported by E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the ALS Liga België and the KU Leuven funds "Een Hart voor ALS", "Laeversfonds voor ALS Onderzoek" and the "Valéry Perrier Race against ALS Fund". Several authors of this publication are members of the European Reference Network for Rare Neuromuscular Diseases (ERN-NMD). To P.J.S: from the Medical Research Council, MND Association, NIHR Senior Investigator Award, National Institute for Health Research (NIHR) Sheffield Biomedical Research Centre and NIHR Sheffield Clinical Research Facility. To P.M.A.: Knut and Alice Wallenberg Foundation, the Swedish Brain Foundation, the Swedish Science Council, the Ulla-Carin Lindquist Foundation. H.P.P. and sequencing activities at NYGC were supported by the ALS Association (ALSA) and The Tow Foundation. C.E. was supported by scholarship from Teva Pharmaceutical Industries Ltd as part of the Israeli National Network of Excellence in Neuroscience (NNE). S.M.K.F. is supported by the ALS Canada Tim E. Noël Postdoctoral Fellowship. R. H. Brown Jr. was funded by ALS Association, ALS Finding a Cure, Angel Fund, ALS-One, Cellucci Fund and NIH grants (R01 NS104022, R01 NS073873 and NS111990-01 to R.H.B.J.). J.K.I. is a New York Stem Cell Foundation-Robertson Investigator. Work at J.K.I. lab was supported by NIH grants R01NS097850, U.S. Department of Defense grant W81XWH-19-PRARP-CSRA, and grants from the Tau Consortium, the New York Stem Cell Foundation, the ALS Association, and the John Douglas French Alzheimer's Foundation. To R.L.McL.: Science Foundation Ireland (17/CDA/4737). To A.N.B.: Suna and Inan Kirac Foundation. To J.E.L.: National Institute of Health/NINDS (R01 NS073873). **Author contributions:** C.E. led the project; C.E. contributed to research conception, design and interpretations and wrote the manuscript with E.H.; C.E., E.B., T.O., K.R.V.E., S.L.P., M.M., S.M.K.F., N.Y., J.C.-K., K.P.K., R.A.A.V.D.S., W.S., A.A.K., A.I., A.S., A.R.J., E.C., D.R., O.W., R.H.B.J., P.J.S., P.V.D., L.H.V.D.B., H.P.P., E.S., A.A.-C. and J.H.V. collected samples, were involved in the sequence analysis pipeline, phenotyping, variant calling, provided expertise or were involved in the genetic association analysis of rare non-coding variants in human patients with ALS; S.-T.H. and J.K.I. performed iMNs experiments and interpreted data; A.S. and C.E. performed molecular biology studies in LCLs and U2OS cell lines, including reporter assays and protein quantification by western blots; S.W. and D.P.S. helped performing research; E.H. conceived and supervised the study and wrote the manuscript with C.E. All co-authors provided approval of the manuscript. **Competing interests:** J.K.I. is a co-founder of AcuraStem Incorporated. J.K.I. declares that he is bound by confidentiality agreements that prevent him from disclosing details of his financial interests in this work. E.H. is inventor on pending patent family

PCT/IL2016/050328 entitled "Methods of treating motor neuron diseases". All other authors declare that they have no competing interests. **Data availability:** Human genetics data is publically available from the sequencing consortia: Project Mine ALS sequencing consortium, the New York Genome Center (NYGC) ALS Consortium, the Genome Aggregation Database (gnomAD) and NHLBI's Trans-Omics for Precision Medicine (TOPMed). All Other data used for this manuscript are available in the manuscript. **Code availability:** Variant annotation scripts are available at GitHub: https://github.com/TsviyaOlender/Non-coding-Variants-in-ALS-genes-.

References

1. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature reviews. Genetics* **10**, 184-194 (2009).
2. Knight, J.C. Regulatory polymorphisms underlying complex disease traits. *Journal of molecular medicine* **83**, 97-109 (2005).
3. Brown, R.H. & Al-Chalabi, A. Amyotrophic Lateral Sclerosis. *The New England journal of medicine* **377**, 162-172 (2017).
4. Taylor, J.P., Brown, R.H., Jr. & Cleveland, D.W. Decoding ALS: from genes to mechanism. *Nature* **539**, 197-206 (2016).
5. Renton, A.E., Chio, A. & Traynor, B.J. State of play in amyotrophic lateral sclerosis genetics. *Nature neuroscience* **17**, 17-23 (2014).
6. Al-Chalabi, A., van den Berg, L.H. & Veldink, J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat Rev Neurol* **13**, 96-104 (2017).
7. DeJesus-Hernandez, M., *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256 (2011).
8. Renton, A.E., *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268 (2011).
9. La Spada, A.R. & Taylor, J.P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature reviews. Genetics* **11**, 247-258 (2010).
10. Haramati, S., *et al.* miRNA malfunction causes spinal motor neuron disease. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13111-13116 (2010).
11. Emde, A., *et al.* Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: a new mechanism for ALS. *The EMBO journal* **34**, 2633-2651 (2015).
12. Eitan, C. & Hornstein, E. Vulnerability of microRNA biogenesis in FTD-ALS. *Brain research* (2016).
13. Campos-Melo, D., Droppelmann, C.A., He, Z., Volkening, K. & Strong, M.J. Altered microRNA expression profile in Amyotrophic Lateral Sclerosis: a role in the regulation of NFL mRNA levels. *Molecular brain* **6**, 26 (2013).
14. Buratti, E., *et al.* Nuclear factor TDP-43 can affect selected microRNA levels. *The FEBS journal* **277**, 2268-2281 (2010).

15. Kawahara, Y. & Mieda-Sato, A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3347-3352 (2012).

16. Morlando, M., *et al.* FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *The EMBO journal* **31**, 4502-4510 (2012).

17. Hoye, M.L., *et al.* MicroRNA Profiling Reveals Marker of Motor Neuron Disease in ALS Models. *J Neurosci* **37**, 5574-5586 (2017).

18. Rotem, N., *et al.* ALS Along the Axons - Expression of Coding and Noncoding RNA Differs in Axons of ALS models. *Sci Rep* **7**, 44500 (2017).

19. Butovsky, O., *et al.* Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS. *J Clin Invest* **122**, 3063-3087 (2012).

20. Figueroa-Romero, C., *et al.* Expression of microRNAs in human post-mortem amyotrophic lateral sclerosis spinal cords provides insight into disease mechanisms. *Mol Cell Neurosci* **71**, 34-45 (2016).

21. Williams, A.H., *et al.* MicroRNA-206 delays ALS progression and promotes regeneration of neuromuscular synapses in mice. *Science* **326**, 1549-1554 (2009).

22. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233 (2009).

23. Mayr, C. Regulation by 3'-Untranslated Regions. *Annual review of genetics* **51**, 171-194 (2017).

24. Povysil, G., *et al.* Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nature reviews. Genetics* **20**, 747-759 (2019).

25. An, J.Y., *et al.* Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362** (2018).

26. Lee, S., Abecasis, G.R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *American journal of human genetics* **95**, 5-23 (2014).

27. Project MinE Consortium, Van Rheenen, W. & et al. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. (2017).

28. Dunckley, T., *et al.* Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *The New England journal of medicine* **357**, 775-788 (2007).

29. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140-144 (2006).

30. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Human mutation* **34**, E2393-2402 (2013).

31. Lee, S., *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics* **91**, 224-237 (2012).

32. Kenna, K.P., *et al.* NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nature genetics* **48**, 1037-1042 (2016).

33. Rosen, D.R., *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59-62 (1993).

34. Chio, A., *et al.* Prevalence of SOD1 mutations in the Italian ALS population. *Neurology* **70**, 533-537 (2008).

35. van der Spek, R.A.A., *et al.* The project MinE databrowser: bringing large-scale whole-genome sequencing in ALS to researchers and the public. *Amyotrophic lateral sclerosis & frontotemporal degeneration* **20**, 432-440 (2019).

36. O'Leary, N.A., *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic acids research* **44**, D733-745 (2016).

37. Alboni, S., Cervia, D., Sugama, S. & Conti, B. Interleukin 18 in the CNS. *Journal of neuroinflammation* **7**, 9 (2010).

38. Smith, L., *et al.* Establishing the UK DNA Bank for motor neuron disease (MND). *BMC Genet* **16**, 84-84 (2015).

39. Tsutsumi, N., *et al.* The structural basis for receptor recognition of human interleukin-18. *Nature communications* **5**, 1-13 (2014).

40. Kadhim, H., Deltenre, P., Martin, J.J. & Sebire, G. In-situ expression of Interleukin-18 and associated mediators in the human brain of sALS patients: Hypothesis for a role for immune-inflammatory mechanisms. *Medical hypotheses* **86**, 14-17 (2016).

41. Johann, S., *et al.* NLRP3 inflammasome is expressed by astrocytes in the SOD1 mouse model of ALS and in human sporadic ALS patients. *Glia* **63**, 2260-2273 (2015).

42. Italiani, P., *et al.* Evaluating the levels of interleukin-1 family cytokines in sporadic amyotrophic lateral sclerosis. *Journal of neuroinflammation* **11**, 94 (2014).

43. Krach, F., *et al.* Transcriptome–pathology correlation identifies interplay between TDP-43 and the expression of its kinase CK1E in sporadic ALS. *Acta neuropathologica* **136**, 405-423 (2018).

44. Thompson, L. iMN (Exp 2)—ALS. *SMA and control (unaffected) iMN cell lines differentiated from iPS cell lines using a long differentiation protocol—RNA-seq., LINCS (collection)* (2017).

45. Shi, Y., *et al.* Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nature medicine* **24**, 313 (2018).

46. Ayers, K.L., *et al.* A loss of function variant in CASP7 protects against Alzheimer's disease in homozygous APOE epsilon4 allele carriers. *BMC genomics* **17 Suppl 2**, 445 (2016).

47. Benitez, B.A., *et al.* Missense variant in TREML2 protects against Alzheimer's disease. *Neurobiology of aging* **35**, 1510 e1519-1526 (2014).

48. Jonsson, T., *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488**, 96-99 (2012).

49. Sims, R., *et al.* Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature genetics* (2017).

50. Landers, J.E., *et al.* Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9004-9009 (2009).

51. Farhan, S.M.K., *et al.* Exome sequencing in amyotrophic lateral sclerosis implicates a novel gene, DNAJC7, encoding a heat-shock protein. *Nature neuroscience* **22**, 1966-1974 (2019).

52. Lambrechts, D., *et al.* VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nature genetics* **34**, 383-394 (2003).

53. Beers, D.R. & Appel, S.H. Immune dysregulation in amyotrophic lateral sclerosis: Mechanisms and emerging therapies. *The Lancet Neurology* **18**, 211-220 (2019).

54. Engelhardt, J.I. & Appel, S.H. IgG reactivity in the spinal cord and motor cortex in amyotrophic lateral sclerosis. *Archives of neurology* **47**, 1210-1216 (1990).

55. Engelhardt, J.I., Tajti, J. & Appel, S.H. Lymphocytic infiltrates in the spinal cord in amyotrophic lateral sclerosis. *Archives of neurology* **50**, 30-36 (1993).

56. Henkel, J.S., *et al.* Presence of dendritic cells, MCP-1, and activated microglia/macrophages in amyotrophic lateral sclerosis spinal cord tissue. *Annals of neurology* **55**, 221-235 (2004).

57. Philips, T. & Robberecht, W. Neuroinflammation in amyotrophic lateral sclerosis: role of glial activation in motor neuron disease. *The Lancet. Neurology* **10**, 253-263 (2011).

58. Zhernakova, A., *et al.* Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. *American journal of human genetics* **82**, 1202-1210 (2008).

59. Hirota, T., *et al.* Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nature genetics* **44**, 1222-1226 (2012).

60. Liu, H., *et al.* Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. *American journal of human genetics* **91**, 935-941 (2012).

61. Hunt, K.A., *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nature genetics* **40**, 395-402 (2008).

62. Smyth, D.J., *et al.* Shared and distinct genetic variants in type 1 diabetes and celiac disease. *The New England journal of medicine* **359**, 2767-2777 (2008).

63. Leal-Lasarte, M.M., Franco, J.M., Labrador-Garrido, A., Pozo, D. & Roodveldt, C. Extracellular TDP-43 aggregates target MAPK/MAK/MRK overlapping kinase (MOK) and trigger caspase-3/IL-18 signaling in microglia. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **31**, 2797-2816 (2017).

64. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-1426 (2016).

65. Raczy, C., *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041-2043 (2013).

66. Tyner, C., *et al.* The UCSC Genome Browser database: 2017 update. *Nucleic acids research* **45**, D626-D634 (2017).

67. Danecek, P., *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).

68. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **38**, e164 (2010).

69. Agarwal, V., Bell, G.W., Nam, J.W. & Bartel, D.P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4** (2015).

70. Erijman, A., Dantes, A., Bernheim, R., Shifman, J.M. & Peleg, Y. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. *Journal of structural biology* **175**, 171-177 (2011).

71. Peleg, Y. & Unger, T. Application of the Restriction-Free (RF) cloning for multicomponents assembly. *Methods in molecular biology* **1116**, 73-87 (2014).

FIGURES

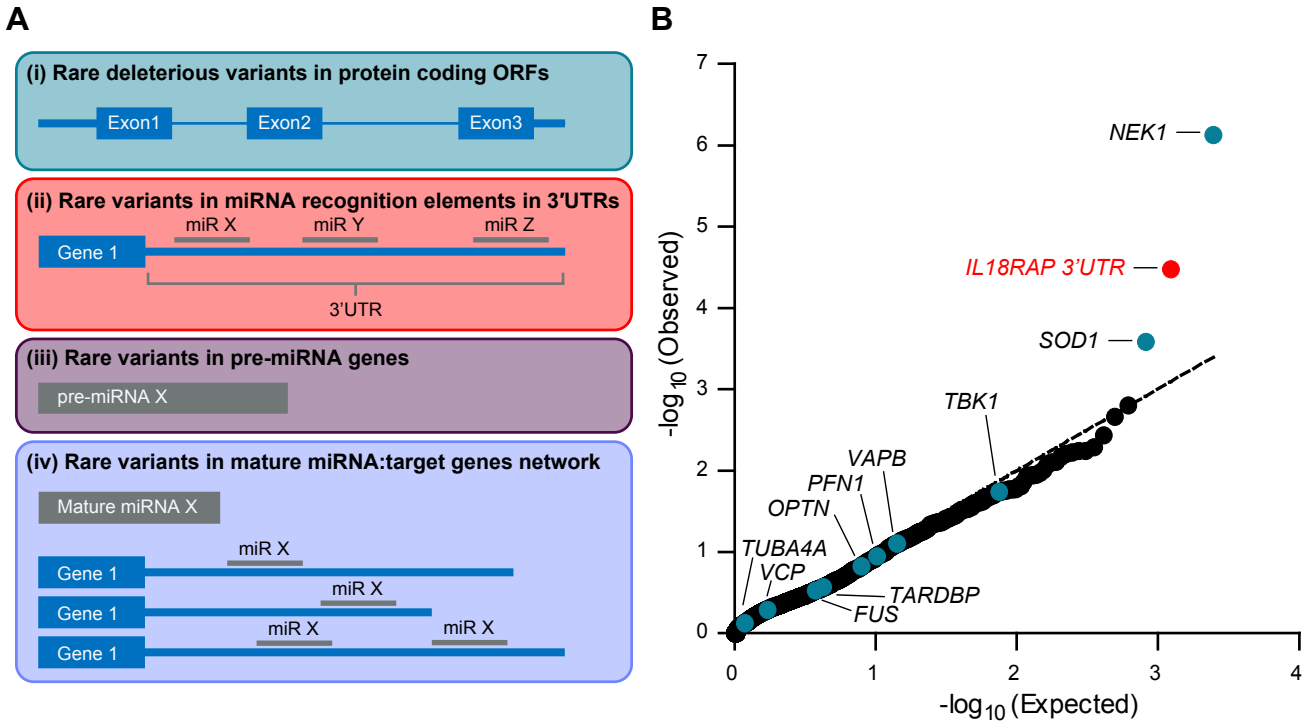


Fig.1 - Eitan et al. (Hornstein)

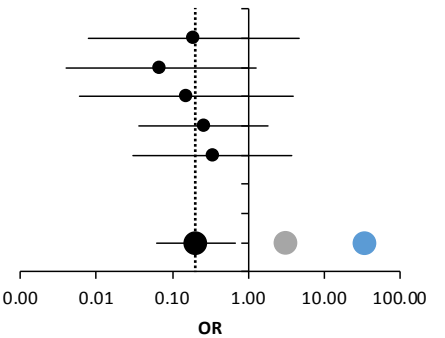
Fig. 1. Region-based rare-variant association analysis. (A) Diagram for region-based rare-variant association studies. Collapsed region-based association analysis was performed on rare (MAF ≤0.01) qualifying variants in: (i) 295 candidate protein-coding genes (Supplementary Table 3) encoding for ALS-relevant proteins or proteins associated with miRNA biogenesis/activity. Variants were included if predicted to cause frameshifting, alternative splicing, an abnormal stop codon, or a deleterious non-synonymous amino acid substitution, in ≥ 3 of 7 independent dbNSFP prediction algorithms; (ii) variants abrogating or gaining miRNA recognition elements in 3′UTRs of the 295 genes (Supplementary Table 3); (iii) all known pre-miRNA genes in the human genome; and (iv) predicted networks, comprised of aggregated variants detected in a specific mature miRNA sequence and its cognate down-stream 3'UTR targets. **(B)** QQ plot of obtained and expected P-values for the burden of rare variants (log scale) gained by collapsed region-based association analysis of all genomic regions described in (A). Data were obtained from 3,955 ALS cases and 1,819 controls (Project MinE). Features positioned on the diagonal line represent results obtained under the null hypothesis. Open-reading frames of 10 known ALS genes (blue). *IL18RAP* 3′UTR miRNA recognition elements (red). Genomic inflation $\lambda = 1.21$.

A

Gene	Region	ALS (3955)	Control (1819)	OR	OR 95% CI	P	$P_{corrected}$
NEK1	Coding	127	19	3.14	1.93-5.11	7.04×10^{-7}	2.13×10^{-4}
SOD1	Coding	36	0	33.89	2.08-552.47	2.61×10^{-4}	3.76×10^{-2}
IL18RAP	miRNA binding sites in 3'UTR	4	9	0.20	0.06-0.66	3.34×10^{-5}	9.31×10^{-3}
IL18RAP	3'UTR	*6*	*12*	0.23	0.09-0.61	1.88×10^{-5}	5.62×10^{-3}

B

Cohort	Cases	Control	OR	OR 95% CI
Ireland	0/239	1/136	0.19	0.008-4.662
Netherlands	0/1633	4/1004	0.07	0.004-1.265
Turkey	0/142	1/67	0.16	0.006-3.870
UnitedKingdom	2/1043	2/272	0.26	0.036-1.850
USA	2/398	1/68	0.34	0.030-3.784
Belgium	0/295	0/172	-	-
Spain	0/205	0/100	-	-
Total	4/3955	9/1819	0.20	0.063-0.662



C

Cohort	Cases	Control	OR	95% CI	P	
					χ2	SKAT-O
Discovery: Project MinE	6/3955	12/1819	0.23	0.086-0.611	3.00×10^{-3}	1.88×10^{-5}
Replication: NYGC, TOPMed & gnomAD	8/2184	786/68584	0.32	0.158-0.637	9.58×10^{-4}	-
Joint analysis: Discovery & replication	14/6139	798/70403	0.20	0.118-0.338	$< 1.00 \times 10^{-5}$	-

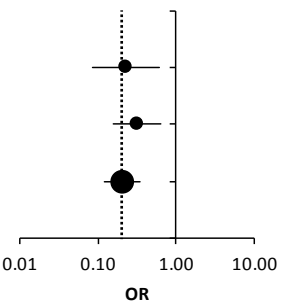
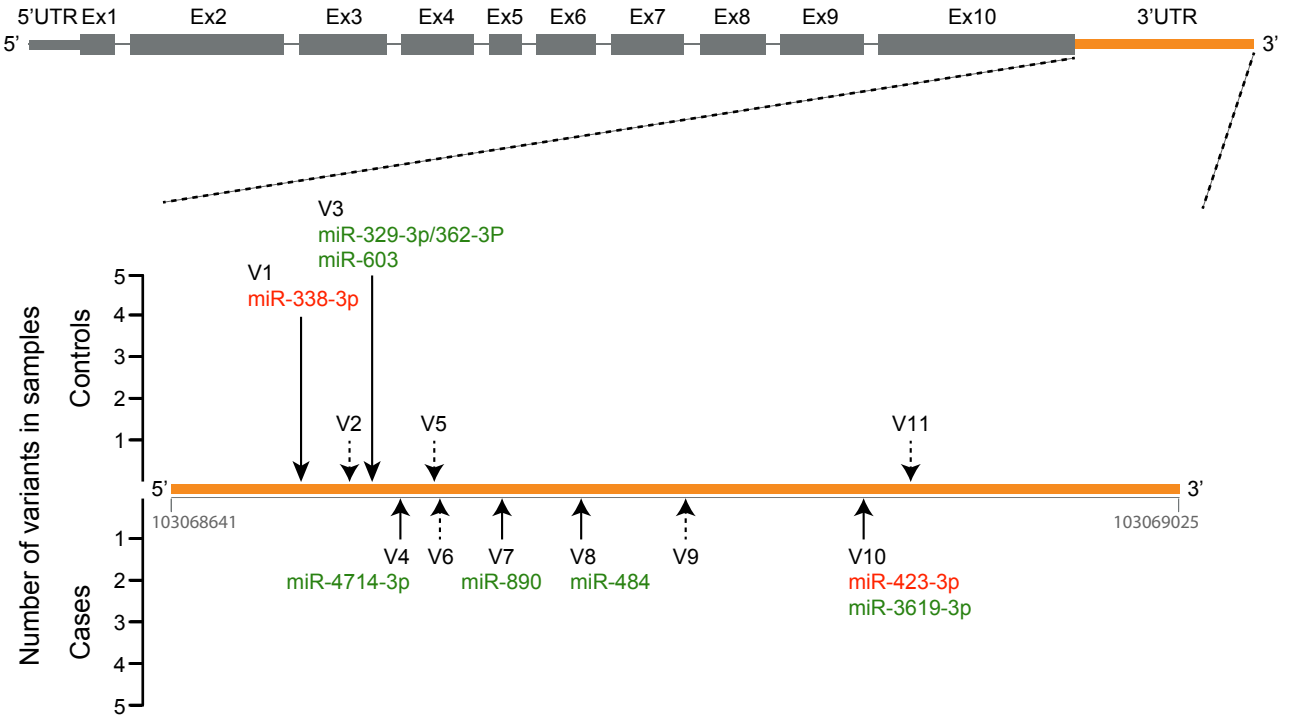


Fig.2 - Eitan et al. (Hornstein)

Fig 2. Odds of ALS is reduced with rare variants in the *IL18RAP* 3'UTR. (A) Odds ratio (OR) estimates with 95% confidence intervals (CI) for NEK1 (coding), SOD1 (coding), predicted miRNA recognition elements in the *IL18RAP* 3'UTR, and for all variants identified in the *IL18RAP* 3'UTR. P values corrected for false discovery rate (FDR). **(B)** Stratification of data pertaining to miRNA recognition elements in the *IL18RAP* 3'UTR in seven geographically-based sALS sub-cohorts and forest plot (Log scale). *NEK1* (grey) and *SOD1* (blue) signals are from combined data of all cohorts. Vertical dotted line denotes OR=0.2. **(C)** OR with 95% CI and forest plot (Log scale) across discovery and replication cohorts and joint analysis thereof. Vertical dotted line denotes OR=0.2. P-values, calculated with SKAT-O or Chi-squared test with Yate's correction.

A

Human IL18RAP
ENST00000264260 Chr2:103035149-103069025 [GRCh37/hg19]
3'UTR length: 384 Chr2:103068641-103069025 [GRCh37/hg19]



B

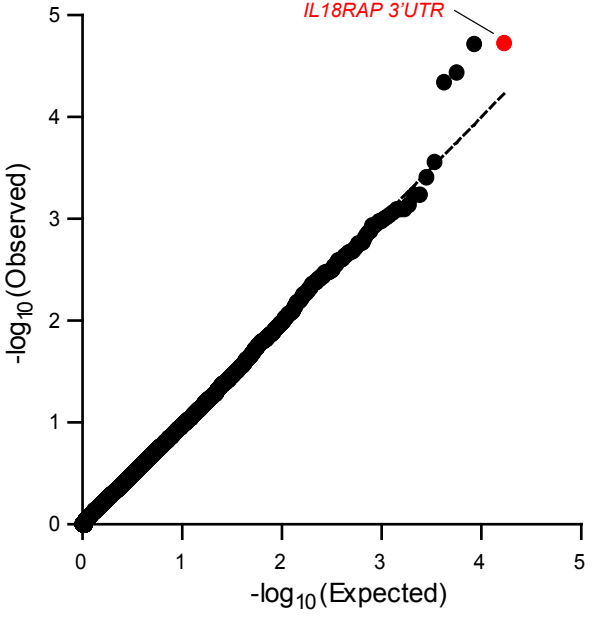


Fig. 3. Rare variants in the *IL18RAP* 3′UTR. (A) Schematic of the *IL18RAP* transcript and 3′UTR (5' to 3′) showing the number of control (upper) or ALS (lower) samples, in which miRNA recognition element variants (black arrow) or other variants (dashed arrow) were identified. Potentially lost (red) or created (green) miRNA recognition elements are marked (Supplementary Table 6). **(B)** QQ plot of obtained and expected P-values for the burden of rare variants (log scale) gained by collapsed region-based association analysis for all known human 3'UTRs (RefSeq), in Project MinE cohort (3,955 ALS cases and 1,819 non-ALS controls). Variants are not restricted to miRNA recognition elements. Features positioned on the diagonal line represent results obtained under the null hypothesis. The *IL18RAP* 3′UTR (red) is the most significant 3'UTR associated with ALS. Genomic inflation $\lambda = 0.97$.

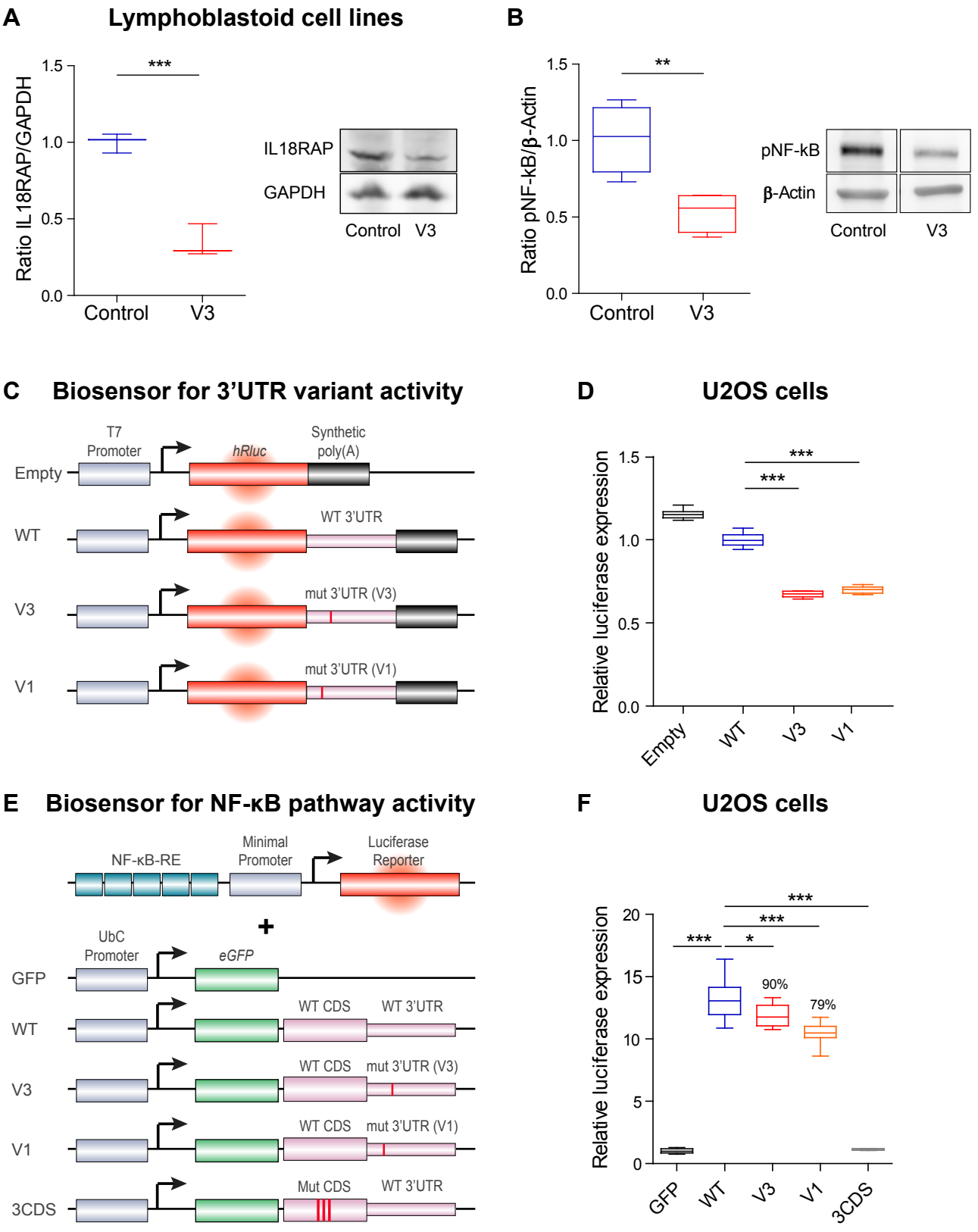

A
Lymphoblastoid cell lines
Ratio IL18RAP/GAPDH
1.5
1.0
0.5
0.0
***
Control
V3
IL18RAP
GAPDH
Control
V3
B
Ratio pNF-kB/β-Actin
1.5
1.0
0.5
0.0
**
Control
V3
pNF-kB
β-Actin
Control
V3
C
Biosensor for 3'UTR variant activity
T7 Promoter
hRluc
Synthetic poly(A)
Empty
WT
WT 3'UTR
V3
mut 3'UTR (V3)
V1
mut 3'UTR (V1)
D
U2OS cells
Relative luciferase expression
1.5
1.0
0.5
0.0
***
***
Empty
WT
V3
V1
E
Biosensor for NF-κB pathway activity
NF-κB-RE
Minimal Promoter
Luciferase Reporter
+
UbC Promoter
eGFP
GFP
WT
WT CDS
WT 3'UTR
V3
WT CDS
mut 3'UTR (V3)
V1
WT CDS
mut 3'UTR (V1)
3CDS
Mut CDS
WT 3'UTR
F
U2OS cells
Relative luciferase expression
20
15
10
5
0
***
***
***
*
90%
79%
GFP
WT
V3
V1
3CDS

Fig 4. *IL18RAP* 3'UTR variant attenuates IL-18 / NF-κB signaling. Quantification of protein expression for IL18RAP (n=3) **(A)** or p-NF-κB protein (n=4) **(B)** by Western blots of extracts from a lymphoblastoid cell line harboring an endogenous IL18RAP variant (V3, Chr2:103068718 G>A) relative to IL18RAP protein in a line with the canonical *IL18RAP* 3'UTR. Loading normalization with anti GAPDH or anti Beta-Actin. Two-sided student *t*-test. Diagram **(C)** and quantification **(D)** of hRluc Il18RAP 3'UTR reporter assays, in human U2OS cell line (Empty, WT, V3, V1; *n*=6). One-way ANOVA followed by Dunnett's multiple comparison test. Diagram **(E)** and quantification **(F)** of NF-κB reporter assay in human U2OS cell line (GFP, WT, V3, V1, n=9; 3CDS, *n*=4). One-way ANOVA followed by Dunnett's multiple comparison test was performed on square root-transformed data. Box plots depicting median, upper and lower quartiles, and extreme points. * P<0.05; ** P<0.01; *** P<0.001. Experiment repeated independently three times with similar results.

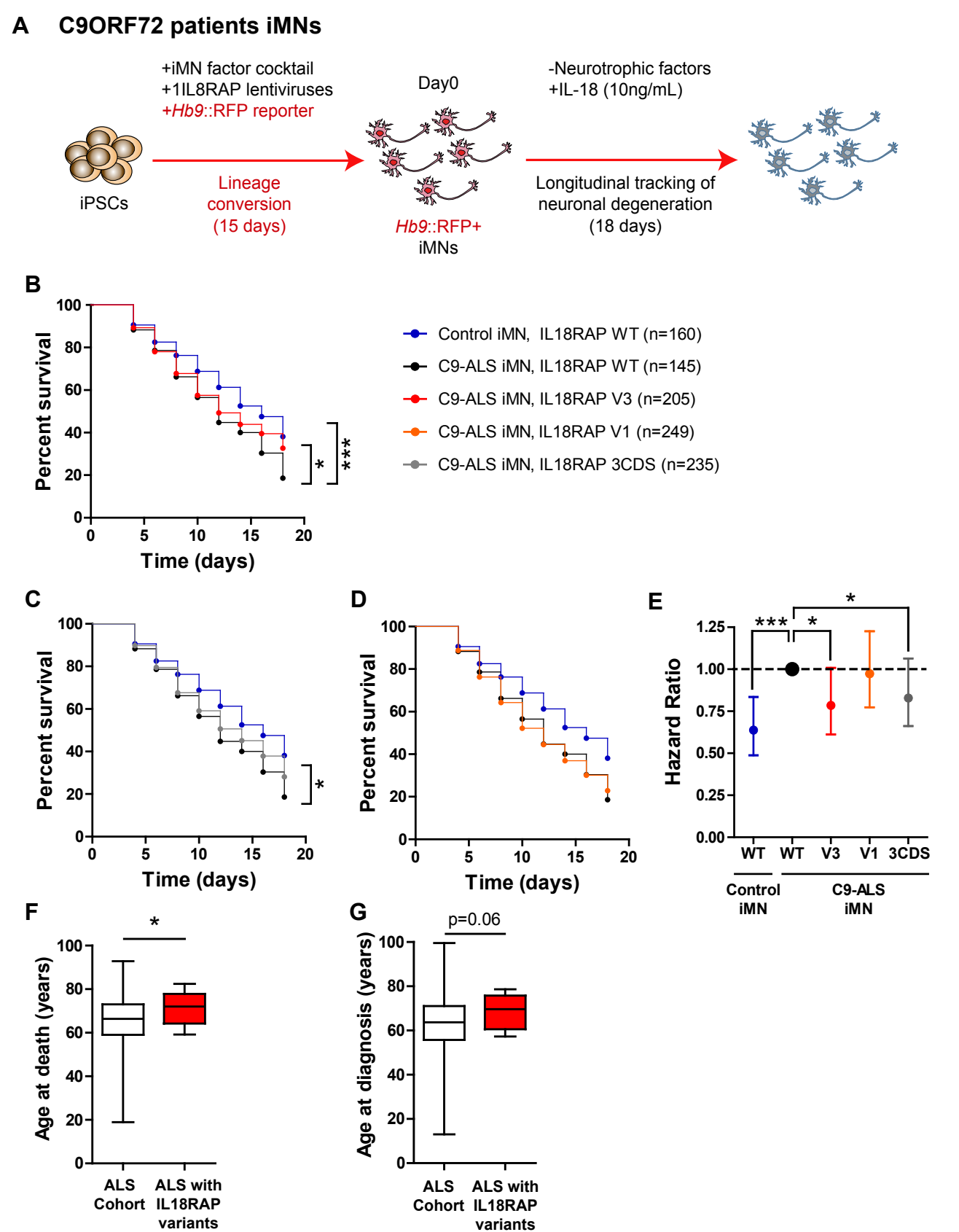
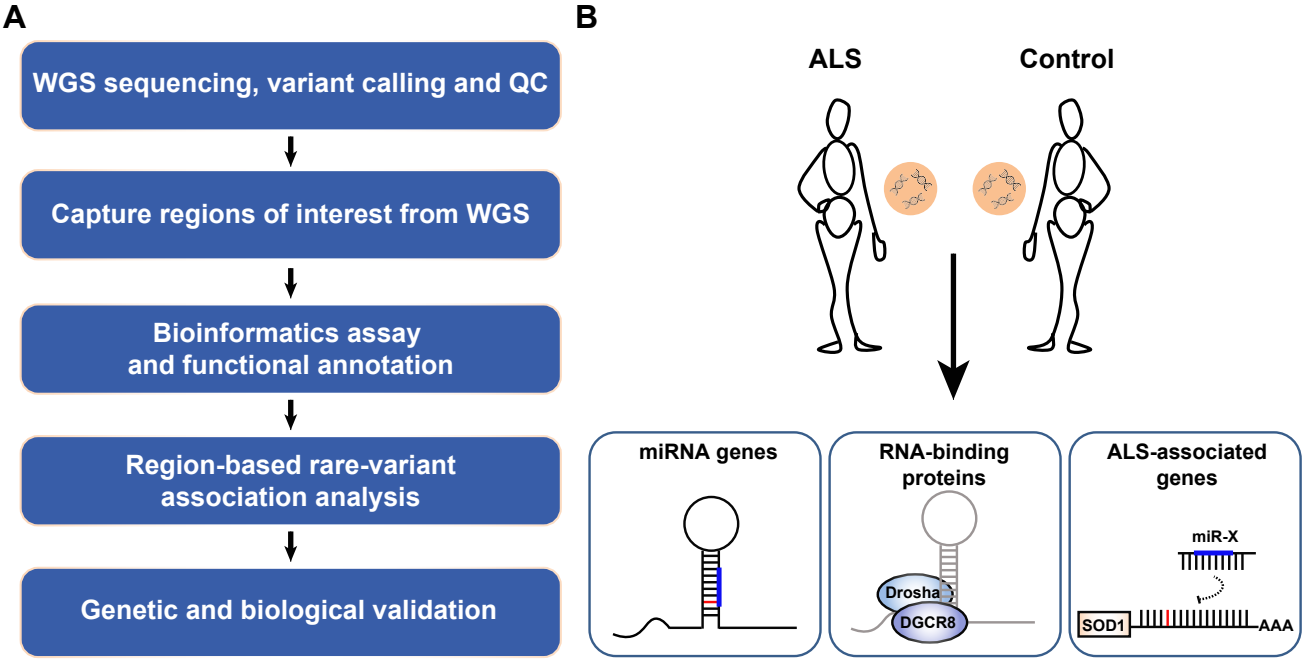
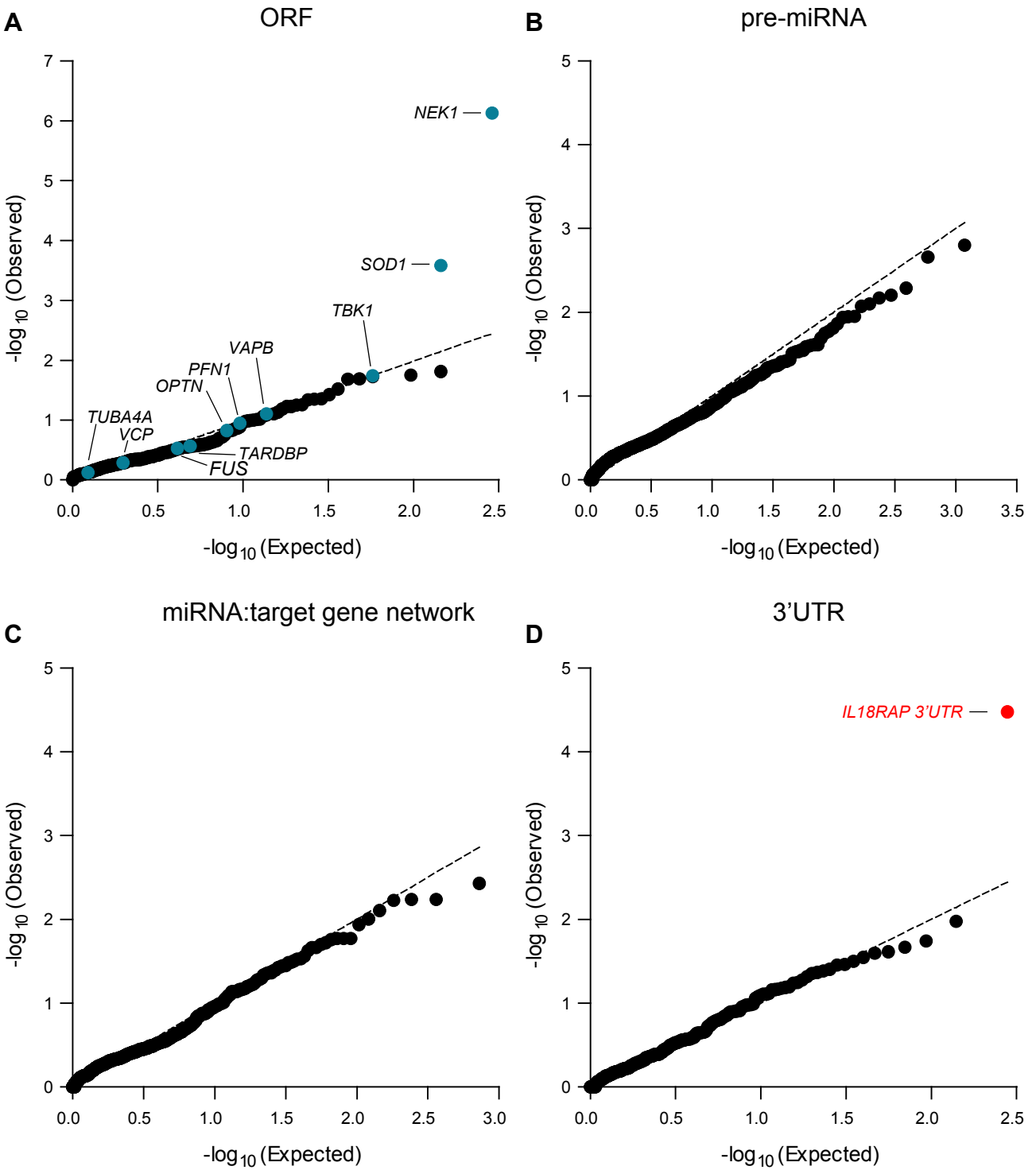

A
C9ORF72 patients iMNs
+iMN factor cocktail
+1IL8RAP lentiviruses
+Hb9::RFP reporter
Day0
-Neurotrophic factors
+IL-18 (10ng/mL)
iPSCs
Lineage conversion (15 days)
Hb9::RFP+ iMNs
Longitudinal tracking of neuronal degeneration (18 days)
B
Percent survival
Time (days)
Control iMN, IL18RAP WT (n=160)
C9-ALS iMN, IL18RAP WT (n=145)
C9-ALS iMN, IL18RAP V3 (n=205)
C9-ALS iMN, IL18RAP V1 (n=249)
C9-ALS iMN, IL18RAP 3CDS (n=235)
C
D
E
Hazard Ratio
WT
WT
V3
V1
3CDS
Control iMN
C9-ALS iMN
F
Age at death (years)
ALS Cohort
ALS with IL18RAP variants
G
p=0.06
Age at diagnosis (years)

Fig 5. *IL18RAP* 3'UTR variant ameliorates disease in *C9orf72* patient iMNs and in patients with ALS. Experimental setup: Time-lapse survival tracking microscopy of *Hb9*::RFP+ iMNs, transduced with IL18RAP lentiviruses, after neurotrophic factors withdrawal and introduction of IL-18 cytokine **(A)**. Kaplan-Meier survival plots for control (CTRL) and *C9orf72* patient (C9-ALS) iMNs, with wild-type (WT) *IL18RAP*, *IL18RAP* harboring variants in the 3'UTR (V3, V1) or an *IL18RAP* dominant negative coding mutant (E210A-Y212A-Y214A) 3CDS. Traces of iMNs from 2 donors per genotype (control/ C9-ALS lines), quantified from 3 independent iMN differentiation experiments per line. Number of iMNs quantified per treatment denoted. One-sided log-rank test for the entire survival time course **(B-D)** and corresponding hazard ratio of cellular death, relative to C9-ALS iMNs with wild-type *IL18RAP* 3'UTR **(E)**. Association of age of death (9 patients with protective 3'UTR variants /4263 patients with available phenotypic data in Project MinE and NYGC cohorts **(F)** or diagnosis (8/4216 patients) **(G)**. *IL18RAP* variant is associated with delayed age of death (+6.1 years, * $P<0.05$) and age of diagnosis (+6.2 years, subthreshold significance of $P = 0.06$), relative to the mean age of all Project MinE and NYGC ALS patients. Box plots depicting median, upper and lower quartiles, and extreme points, one-sided Mann-Whitney test. * $P<0.05$; *** $P<0.001$.



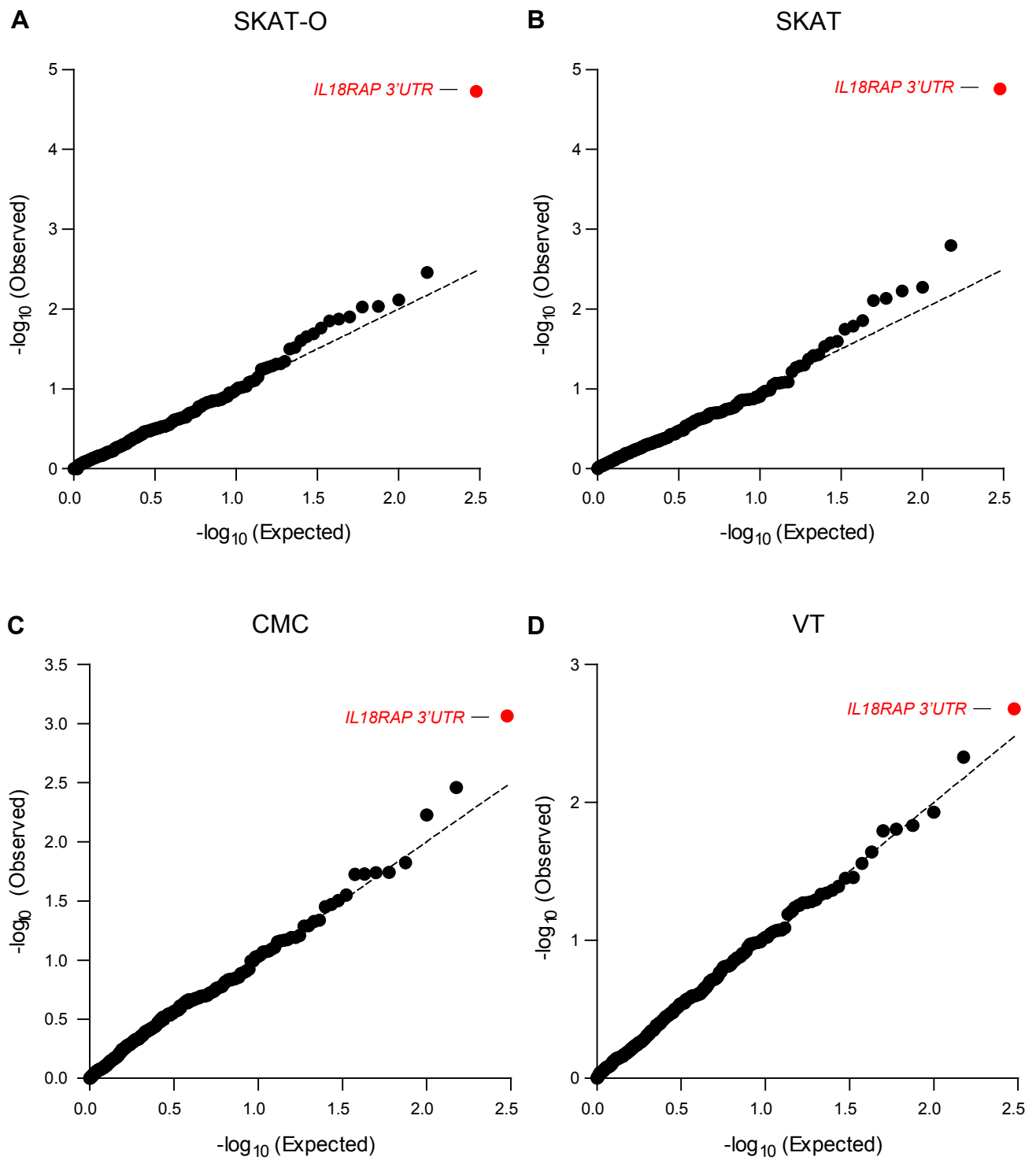
Supplementary Fig.1 - Eitan et al. (Hornstein)

Supplementary Fig. 1. Study design. (A) Flow chart of approach for discovery of region-based rare-variants in non-coding genomic regions via association studies and **(B)** diagram depicting regions of interest comprising of 1,750 human pre-miRNA genes, 295 open reading frames encoding for proteins of interest and 295 3′UTRs.



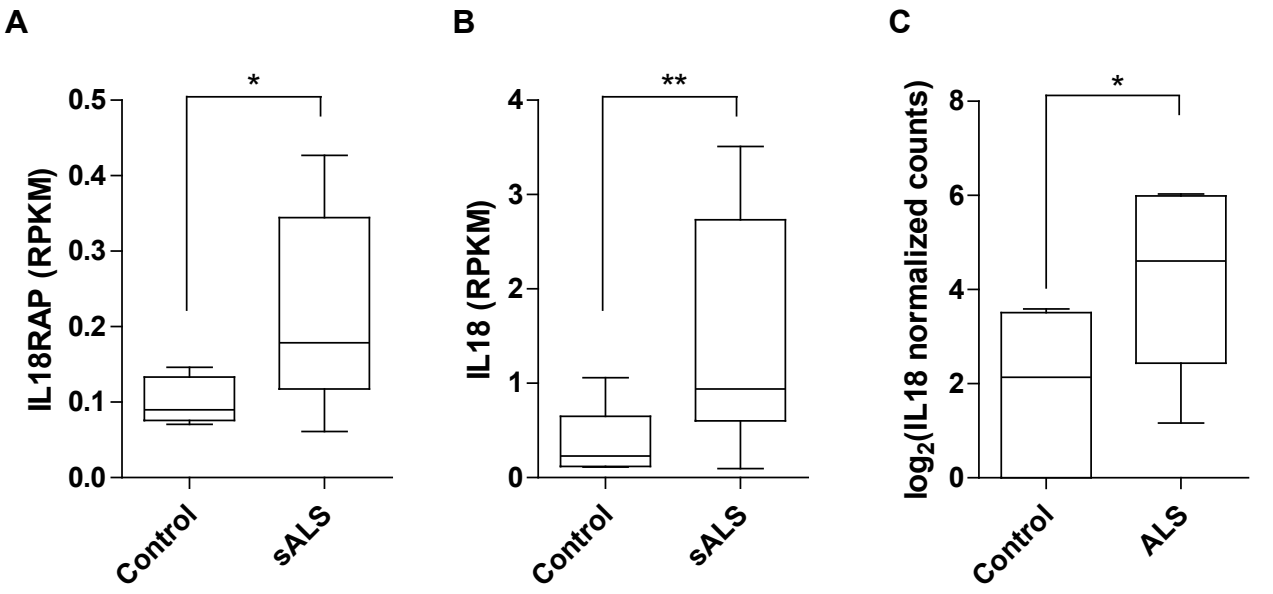
Supplementary Fig.2 - Eitan et al. (Hornstein)

Supplementary Fig. 2. Region-based rare-variant association analyses. (A-D) QQ plot of obtained and expected *P*-values for the burden of rare-variants (log scale) gained by collapsed region-based association analysis of different genomic regions, comprised of (A) 295 candidate protein-coding listed in Supplementary Table 3. These ORFs encode for ALS-relevant proteins or proteins that are associated with miRNA biogenesis or activity. Variants were depicted if predicted to cause frameshifting, alternative splicing, abnormal stop codon or a deleterious non-synonymous amino acid substitution, in ≥ 3 of 7 independent dbNSFP prediction algorithms (genomic inflation $\lambda = 0.97$), (B) All known pre-miRNA genes in the human genome (genomic inflation $\lambda = 1.30$), (C) predicted networks, comprised of aggregated variants detected on a specific mature miRNA sequence and its cognate downstream 3'UTR targets (genomic inflation $\lambda = 1.17$), and (D) variants abrogating or gaining miRNA recognition elements in 3′UTRs of same 295 genes listed in Supplementary Table 3 (genomic inflation $\lambda = 1.03$). Data was obtained from 3,955 ALS cases and 1,819 controls (Project MinE). Features positioned on the diagonal line represent results obtained under the null hypothesis. Open-reading frames of 10 known ALS genes (blue). *IL18RAP* 3′UTR miRNA recognition elements (red).



Supplementary Fig.3 - Eitan et al. (Hornstein)

Supplementary Fig. 3. 3′UTR-based rare-variant association analysis, using different algorithms. (A-D) QQ plot of obtained and expected P-values for the burden of rare variants (log scale) gained by collapsed region-based association analysis of genomic regions comprised of 295 3′UTRs listed in Supplementary Table 3, in Project MinE cohort (3,955 ALS cases and 1,819 non-ALS controls). Variants are not restricted to miRNA recognition elements. Features positioned on the diagonal line represent results obtained under the null hypothesis. *IL18RAP* 3′UTR (red) is the most significant 3'UTR associated with ALS, using different algorithms: (A) Optimized Sequence Kernel Association Test, SKAT-O (genomic inflation $\lambda = 0.98$), (B) Sequence Kernel Association Test, SKAT (genomic inflation $\lambda = 0.98$), (C) Combined Multivariate and Collapsing, CMC (genomic inflation $\lambda = 1.27$), (D) Variable Threshold with permutation analysis, VT (genomic inflation $\lambda = 1.07$).



Supplementary Fig.4 - Eitan et al. (Hornstein)

Supplementary Fig. 4. Evaluation of IL18RAP and IL-18 mRNA expression in motor neurons of patients with ALS. (A-B) mRNA expression of IL18RAP **(A)** and IL-18 **(B)**, as reads per kilobase million (RPKM), from NGS study of laser capture microdissection–enriched surviving motor neurons from lumbar spinal cords of patients with sALS with rostral onset and caudal progression (n = 12) and non-neurodegeneration controls (n = 9; [43] GSE76220). two-sided Student's t test with Welch's correction on log-transformed data. **(C)** IL-18 mRNA expression, as log2-normalized counts, from NGS study of induced ALS motor neurons (n = 4 different donors in duplicates) or non-neurodegeneration controls (n=3 different donors in duplicates; [44] DESeq analysis). Box plots depicting median, upper and lower quartiles, and extreme points. *P < 0.05; **P < 0.01.