A human RNA ligase that operates via auto- and RNA-AMPylation

Authors: Yizhi Yuan[1,2], Florian M. Stumpf[1,2], Lisa A. Schlor[1,2], Olivia P. Schmidt[1], Luisa B. Huber[1,2], Matthias Frese[1,2], Martin Scheffner[2,3], Florian Stengel[2,3], Kay Diederichs[2,3], Andreas Marx[1,2]*

Affiliations:

[1]Department of Chemistry, University of Konstanz, Konstanz, Germany

[2]Konstanz Research School Chemical Biology, University of Konstanz, Konstanz, Germany

[3]Department of Biology, University of Konstanz, Konstanz, Germany

*Correspondence: andreas.marx@uni-konstanz.de

Abstract:

Different forms of life are known to express RNA ligases that catalyze the condensation of a 3'-hydroxy group and a 5'-terminal phosphate of RNA. No such RNA ligases have yet been identified in vertebrates. Here, we report that the hitherto uncharacterized human protein chromosome 12 open reading frame 29 (C12orf29) is a 5'-3' RNA ligase. C12orf29 catalyzes RNA ligation via auto-AMPylation of a critical lysine residue by using ATP as a cosubstrate and subsequent AMP transfer to the 5'-phosphate of an RNA substrate followed by phosphodiester bond formation. Studies at the cellular level reveal the involvement of C12orf29 in maintaining RNA integrity upon cellular stress induced by reactive oxygen species. These findings highlight the importance of RNA ligation for cellular fitness.

One Sentence Summary: C12orf29, a protein of hitherto unknown function, is the first example of a human 5'-3' RNA ligase that is involved in RNA maintenance and cellular resilience.

DNA ligases seal nicks between a 3'-hydroxy group and a 5'-terminal phosphate in DNA (*1-3*). They are indispensable for maintaining genome integrity due to their involvement in DNA replication, recombination and repair throughout all domains of life. In several forms of life, but not in higher eukaryotes such as vertebrates, RNA ligases that catalyze the corresponding reaction with RNA have been described and linked to RNA repair (*4, 5*). These enzymes, termed 5'-3' RNA ligases, operate via a three-steps mechanism alike DNA ligases (Fig. 1A) (*4,6*). First, the ligase auto-AMPylates the catalytically essential lysine residue by the use of ATP resulting in the ligase-(lysyl-*N*)-AMP species. Then, the AMP is transferred onto the 5'-phosphate end of RNA (pRNA) to yield the RNA-adenylate intermediate (AppRNA). Finally, the two RNA ends are ligated by phosphodiester bond formation upon nucleophilic attack of the 3'-OH to the AppRNA to liberate AMP. Surprisingly, proteins with 5'-3' RNA ligase activity have not yet been identified in higher eukaryotes, such as humans, although cell culture experiments conducted already in 1983 suggested their presence (*7*). Here, we report the first human enzyme of that kind by showing that the hitherto uncharacterized human protein chromosome 12 open reading frame 29 (C12orf29) is a 5'-3' RNA ligase.

C12orf29 is an RNA ligase operating via sequential auto- and RNA AMPylation

C12orf29 is a 37 kDa human protein consisting of 325 amino acids. Sequence analysis shows that it is highly conserved among higher eukaryotes like vertebrates (fig. S1) but absent in lower eukaryotes species such as yeast. To obtain insight into its potential function, we initially performed structure prediction of C12orf29 using bioinformatic tools. Intriguingly, a putative structure of a truncated form of C12orf29 was obtained by Phyre^2 (*8*) based on the identity of 22 out of 71 residues (195-265) with those of *Naegleria gruberi* RNA ligase (*Ngr*Rnl) (*9, 10*). This motivated us to investigate whether C12orf29 also possesses RNA ligase activity. Recently, AlphaFold (*11*) structure predictions for almost all proteins of the human genome were made available at https://alphafold.ebi.ac.uk (*12*). We submitted the AlphaFold structure prediction for C12orf29, which was expected to be of high quality (average pLDDT = 91.4), to the Dali server (*13*), and compared it exhaustively to a representative subset of the Protein Data Bank. The results of structural superposition by Dali are highly significant and

display RNA ligases as most similar structures (top hit, Z-score = 7.4 for *Ngr*Rnl), confirming our initial sequence-based suggestion for the functional assignment of C12orf29. In turn, the structure of C12orf29 predicted by AlphaFold was superimposed on *Ngr*Rnl (Fig. 1B, C, fig. S2). While an N-terminal oligonucleotide-binding (OB) domain as in *Ngr*Rnl was not observed for C12orf29, several secondary structure patterns in the predicted C12orf29 structure precisely overlapped with the C-terminal nucleotidyltransferase (NT) domain of *Ngr*Rnl (Fig. 1B, C). Of note, C12orf29 contains a lysine at position 57 within the conserved sequence motif KX(D/H/N)G that defines the superfamily of nucleotidyltransferases (fig. S2) (*6*).

To elucidate potential enzymatic activities of C12orf29, we prepared the recombinant full-length wild-type (WT) protein. Surprisingly, LC-MS measurement of the protein purified from *Escherichia coli* (*E. coli*) revealed a molecular mass that is 329 Da higher than the calculated mass of 37,771 Da suggesting that C12orf29 is AMPylated (Fig. 1D). Indeed, AMPylation of C12orf29 was validated by Western blot (WB) analysis using a monoclonal antibody against AMPylation (*14*) providing the first evidence that C12orf29 may act as an RNA ligase (Fig. 1D). Meanwhile, we also developed an expression and purification protocol that allowed the preparation of C12orf29 without AMPylation (Fig. 1D ad SI). Interestingly, the AMPylated and the non-modified form of C12orf29 showed different migration behaviors in SDS-PAGE. Moreover, in line with our expectation, mutation of the critical lysine at position 57 to alanine (K57A) abolished AMPylation. With C12orf29^{WT} and C12orf29^{WT}-AMP in hand, we further proved that C12orf29 was auto-AMPylated using ATP as a cosubstrate in the presence of Mg^{2+} as the most proficient cofactor (Fig. 1E).

Encouraged by the auto-AMPylation activity, we next investigated whether C12orf29 has 5'-3' RNA ligase activity. We initially studied nick sealing activity as described for *Ngr*Rnl and used substrates that consists of an oligonucleotide splint bringing two oligonucleotides in a complex to form a nick to be ligated (*9*) but did not detect any activity (fig. S3). In our further search for suitable substrates, we were guided by the following hypothesis: Compared to double-stranded RNA regions, the increased flexibility of single-stranded RNA (ssRNA) allows in-line attack of the 2'-OH group at the bridging phosphodiester bond (*15*). This favors transesterification and in consequence RNA strand cleavage. We further reasoned that enzymes that repair and ultimately ligate such ssRNA strand breaks should be beneficial to a cell (*4,5,16*). Therefore, we focused on ssRNA as substrates and studied oligonucleotide constructs that were composed of two strands and folded into an “open hairpin”. The 5'-terminus of one strand (17 nt) at the putative ligation site was labeled with ^{32}P-phosphate while the other strand (10 nt) was unlabeled (Fig. 2A, fig. S4). Successful ligation would lead

to 27 nt oligonucleotide with significantly altered migration detectable by denaturing PAGE analysis. Indeed, we found that C12orf29 was proficient in ligating the ssRNA overhangs to give a product that migrated at the expected length (Fig. 2A). Moreover, formation of cyclized RNA as well as AppRNA was also observed. Interestingly, the enzyme exhibited profound selectivity for RNA: RNA constructs were efficiently ligated within short time while the respective DNA constructs were not detectably converted under the very same conditions. We also investigated whether nucleotides other than ATP represented proficient cosubstrates and found that other adenosine nucleotides such as diadenosine tri- and tetraphosphate (Ap_3A, Ap_4A), and to a lesser extent even dATP are used as cosubstrates, while NAD^+ is not used for promoting RNA ligation (Fig. 2B). Among the other nucleotides investigated, only GTP can promote the RNA ligation reaction (Fig. 2B). Additional kinetic investigations, however, revealed that the apparent catalytic efficiency (k_{cat}/K_M) of C12orf29 is about 29-fold higher with ATP than with GTP. This difference is mainly due to differences in K_M while the k_{cat} is similar for both nucleotides (Fig. 2C, fig. S5).

Next, we investigated if 3'-termini bearing a 2',3'-cyclic phosphate (cPO_4) or a 2'-phosphate end can be used by C12orf29 in the ligation reaction. We found that both constructs were not processed by the protein (fig. S6). However, when the RNA substrate with a 2',3'-cPO_4 at the 3'-terminus was incubated in the presence of N-terminus-truncated ANGEL2 (ANGEL2-ΔN), an enzyme recently reported to cleave this phosphate from the 3'-terminus resulting in non-phosphorylated 3'-termini (*17*), we observed C12orf29-dependent ligation to occur (Fig. 2D). Conclusively, these results show that non-phosphorylated 3'-RNA termini are essential for the ligase reaction.

In order to investigate the structural requirements for C12orf29-catalyzed ligation within the single-stranded regions, we designed and studied single RNA oligonucleotides modified with ^{32}P-phosphate at the 5' ends, which fold into structures bearing two single-stranded regions at the 5' and 3' ends (Fig. 3A, B, fig. S4). When varying the nucleotide composition at the ligation site we found that RNA bearing purines at the ligation site are most efficiently processed (Fig. 3A). Moreover, longer 5'-overhangs are also ligated more efficiently (Fig. 3B).

As shown in Fig. 1C, figs S1, 2, several conserved residues are proposed to be critical for the RNA ligase activity of C12orf29 since they make up the catalytic center and are proposed to be involved e.g., in substrate and cofactor coordination. To examine the importance of these residues for RNA ligation as well as some residues that were reported to be mutated in cancers, a series of C12orf29 variants bearing mutations at the respective residues were prepared and tested. We found that several of these

mutations abolish RNA ligation (Fig. 3C). Noteworthy, among the C12orf29 variants studied, the mutations D59N, R77L, E123D, and K263N were found in patients suffering from esophageal squamous cell carcinoma (*18*), glioblastoma (*19*), chronic lymphocytic leukemia (*20*), and ovarian serous cystadenocarcinoma (*21*), respectively. While E123D has no influence on the RNA ligase activity, D59N, R77L, and K263N impair RNA ligation.

Knockout (KO) of C12orf29 increases cellular vulnerability towards reactive oxygen species

To gain insights into the function of C12orf29 in a cellular context, we generated human embryonic kidney (HEK293) cells, in which the *C12ORF29* gene (fig. S7) was knocked-out by CRISPR/Cas, and compared the properties of these cells with those of parental HEK293 cells expressing the WT enzyme. When cells were grown under physiological conditions, WT and KO cells have similar phenotypes with respect to cell growth, adherence, and appearance (Fig. 4A). However, when cells were treated with menadione, a molecule known to generate reactive oxygen species (ROS)-based cellular stress (*22, 23*), significant differences were observed (Fig. 4A). In comparison to WT cells, KO cells start to round up and to detach at lower concentrations of menadione. Moreover, KO cells are much more vulnerable to menadione that cell viability is affected at lower concentration of menadione in comparison to WT cells (Fig. 4B). Next, we investigated whether this difference stems from a varying ROS formation in those cell lines. We found that KO cells die at much lower ROS levels than their WT counterpart (Fig. 4C). Therefore, we conclude that the observed increase in vulnerability towards ROS of the KO cells is not caused by difference of ROS formation in the cell lines.

Next, we isolated and analyzed the total RNA from the two cell lines treated with various concentrations of menadione (Fig. 5). Very prominent in this analysis are the signals of 28S and 18S rRNA. We found that the amount of RNA degraded depends on the concentration of menadione applied. Interestingly, while the signal for 18S rRNA remains almost constant even at higher menadione concentrations, 28S rRNA appears to be more vulnerable. More importantly, in KO cells 28S rRNA is already degraded at a much lower concentration of menadione than in WT cells. (Fig. 5, fig. S8). Since the intracellular menadione-triggered ROS formation does not vary between both cell lines (Fig. 4C), we conclude that the observed increase in RNA decay in the KO cells is due to the lack of C12orf29. Therefore, we conclude that C12orf29 is important in maintaining RNA integrity under stress conditions.

Discussion

This study reports on the discovery of human C12orf29 as the first 5'-3' RNA ligase in higher eukaryotes. Our results elucidate the substrate spectrum of C12orf29 and show that the enzyme ligates RNA via successive protein and RNA AMPylation. Our studies show that HEK293 cells are more vulnerable to menadione-triggered ROS when *C12ORF29* is knocked out. We found decrease of resilience towards ROS in the KO cells with a simultaneous increase in RNA decay indicating that RNA ligation has an important role in stress-adaptive RNA repair.

Indeed, enzymes with 2',3'-cyclic phosphatase (ANGEL2) (*17*) and 5'-kinase (*Hs*CLP1) (*24*) activities have been identified in humans and may work in concert with C12orf29 in the repair of RNA strands that result from transesterification. The data obtained indicates the importance of maintaining a functional RNA pool (*ribonucleostasis*) for cellular fitness. Since malfunction of RNA repair in humans is connected to the onset of several diseases such as neurodegeneration and cancer (*5,16*) our findings may have implications for the development of new therapeutics. Our results demonstrate that the human protein C12orf29 is an RNA ligase and therefore we propose to name the protein *Homo sapiens* RNA ligase (*Hs*Rnl).

References

1. I. R. Lehman, *Science* **186**, 790-797 (1974).
2. T. Ellenberger, A. E. Tomkinson, *Annu. Rev. Biochem.* **77**, 313-338 (2008).
3. S. Shuman, *J. Biol. Chem.* **284**, 17365-17369 (2009).
4. J. Popow, A. Schleiffer, J. Martinez, *Cell. Mol. Life Sci.* **69**, 2657-2670 (2012).
5. A. M. Burroughs, L. Aravid, *Nucleic Acids Res.* **44**, 8525-8555 (2016).
6. S. Shuman, C. D. Lima, *Curr. Opin. Struc. Biol.* **14**, 757-764 (2004).
7. W. Filipowicz, M. Konarska, H. J. Gross, A. J. Shatkin, *Nucleic Acids Res.* **11**, 1405-1418 (1983).
8. L. Kelley, S. Mezulis, C. Yates, M. N. Wass, M. J. Sternberg *Nat. Protoc.* **10,** 845–858 (2015).
9. M.-C. Unciuleac, S. Shuman, *RNA* **21**, 824–832 (2015).

10. M.-C. Unciuleac, Y. Goldur, S. Shuman, *Proc. Natl. Acad. Sci. U.S.A.***112**, 13868-13873 (2015).
11. J. Jumper, R. Evans, A. Pritzel, A. et al. *Nature* **596**, 583–589 (2021).
12. K. Tunyasuvunakool, J. Adler, Z. Wu, Z. et al. *Nature* **596**, 590–596 (2021).
13. L. Holm, *Nucleic Acids Res.* **50** (W1), W210–W215 (2022).
14. D. Höpfner, J. Fauser, M. S. Kaspers, C. Pett, C. Hedberg, A. Itzen, *iScience* **23**, 101800 (2020).
15. G. A. Soukup, R. D. Breaker, *RNA* **5**, 1308-1325 (1999).
16. L. L. Yan, H. S. Zahner, *J. Biol. Chem.* **294**, 15158-15171 (2019).
17. P. H. Pinto, A. Kroupova, A. Schleiffer, K. Mechtler, M. Jinek, S. Weitzler, J. Martinez, *Science* **369**, 524-530 (2020).
18. Y.-B. Gao, *et al.*, *Nat. Genet.* **46**, 1097-1102 (2014).
19. V. Frattini, *et al.*, *Nat. Genet.* **45**, 1141-1149 (2013).
20. Sanger Institute. COSMIC, The Catalogue of Somatic Mutations in Cancer. https://cancer.sanger.ac.uk/cosmic (accessed May 25, 2022).
21. Cancer Genome Atlas Research Network, *Nature* **474**, 609-615 (2011).
22. H. P. Misra, I. Fridovich, *J. Biol. Chem.* **247**, 188-192 (1972).
23. S. Goffart, P. Tikkanen, C. Michell, T. Wilson, J. L. O. Pohjoismäki *Cells* **10**, 1075 (2021).
24. S. Weitzer, J. Martinez *Nature* **447**, 222–226 (2007).
25. P. Emsley, K. Cowtan, *Acta Crystallogr.* **D60**, 2126-2132 (2004).

Acknowledgments

We thank A. Marquardt and A. Sladewska-Marquardt (Proteomics Center, University of Konstanz) for assistance in mass spectrometry. Y.Y., F.M.S., L.A.S., L.B.H., and M.F. thank the Konstanz Research School Chemical Biology for support. We thank D. Höpfner and A. Itzen, UKE Hamburg, for the provision of the anti-AMPylation antibody.

Funding

A.M. acknowledges the European Research Council for funding of this project (ERC AdG AMP-Alarm), F.S. acknowledges the Deutsche Forschungsgemeinschaft for funding (STE 2517/5-1), and O.P.S. acknowledges support by the Alexander von Humboldt-Foundation.

Author contributions

Y.Y., F.M.S., L.A.S., O.P.S. and A.M. conceived the study and experimental approach; Y.Y., F.M.S., L.A.S., O.P.S. generated C12orf29 and its mutants, performed AMPylation and RNA ligation experiments, Y.Y. conducted the modelling and MS analysis, F.M.S. conducted cellular experiments. L.B.H. and M.F. provide expert expertise in PAGE analysis and AMPylation reactions, respectively. M.S., F.S., K.D. provided expert expertise on cellular studies, mass spectrometry and modelling, respectively. Y.Y., F.M.S., L.A.S., O.P.S. and A.M. analyzed the data, A.M. wrote the manuscript with input from all authors.

Competing interests: Authors declare that they have no competing interests.

Supplementary Materials

Materials and Methods

Figs. S1 to S11

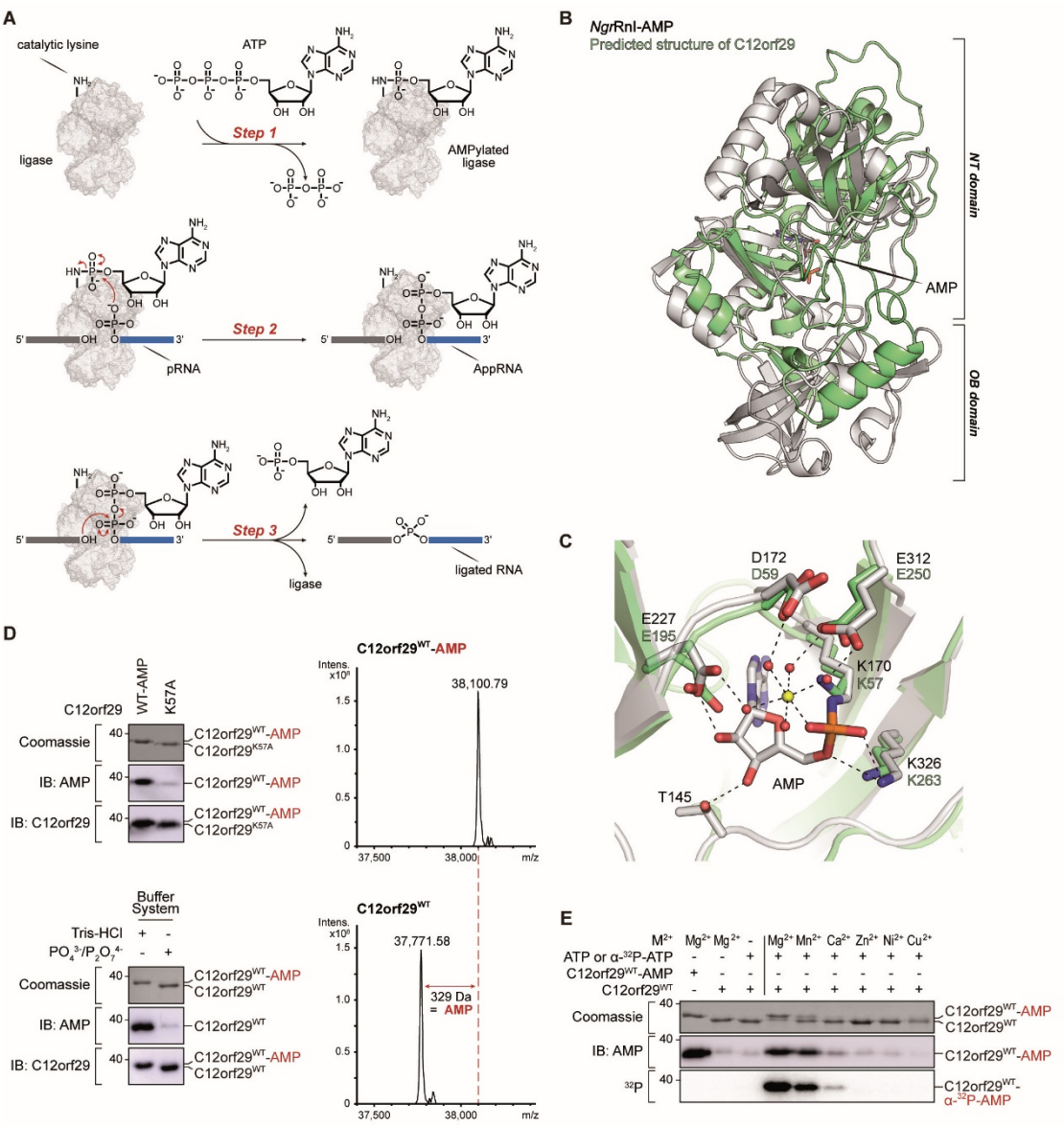

Fig. 1. RNA ligase mechanism, structural prediction and auto-AMPylation activity of C12orf29. (**A**) Schematic display of the three-step mechanism of RNA ligation by a 5'-3' RNA ligase. In step 1, the ligase is auto-AMPylated on the catalytic lysine using ATP as the cosubstrate. In step 2, the AMP is transferred from the catalytic lysine to the 5'-PO_4 end of RNA (pRNA), giving the RNA-adenylate intermediate (AppRNA). In step 3, the ligated RNA is obtained upon the attack of 3'-OH to the AppRNA in the presence of the ligase along with the liberation of AMP. (**B**) Superimposition of the structure of C12orf29 predicted by AlphaFold (green) (*11, 12*) on the structure of *Ngr*Rnl (PDB ID: 5COT, grey). OB, oligonucleotide-binding. NT, nucleotidyltransferase. Structures were superposed in Coot (*25*) using structural equivalent residues identified by the DALI webserver (*13*). (**C**) Enlarged view of the catalytic pocket of *Ngr*Rnl (PDB ID: 5COT) and predicted C12orf29. In the *Ngr*Rnl structure, AMP is covalently attached to the side chain of K170. D172, E227, and E312 bind Mn^{2+} via water-mediated contacts and K326 contacts the phosphate moiety of AMP. Corresponding residues in the putatively catalytic pocket of C12orf29 are indicated. Mn^{2+} and water molecules were depicted as yellow and red spheres, respectively. Atomic contacts were depicted as dashed lines. (**D**) Top: Immunoblotting of C12orf29^{WT}-AMP and C12orf29^{K57A} and MS analysis of C12orf29^{WT}-AMP. C12orf29^{K57A} is deficient in auto-AMPylation activity. Mass spectra indicating a 329

Da increase in mass upon auto-AMPylation of C12orf29^{WT}. C12orf29^{WT}-AMP: calc. 38,100.34 Da, found 38,100.79 Da. C12orf29^{WT}: calc. 37,771.29 Da, found 37,771.58 Da. Bottom: Preparation of C12orf29^{WT} and C12orf29^{WT}-AMP with different buffer systems. (**E**) Divalent metal ion dependency of auto-AMPylation activity of C12orf29^{WT}.

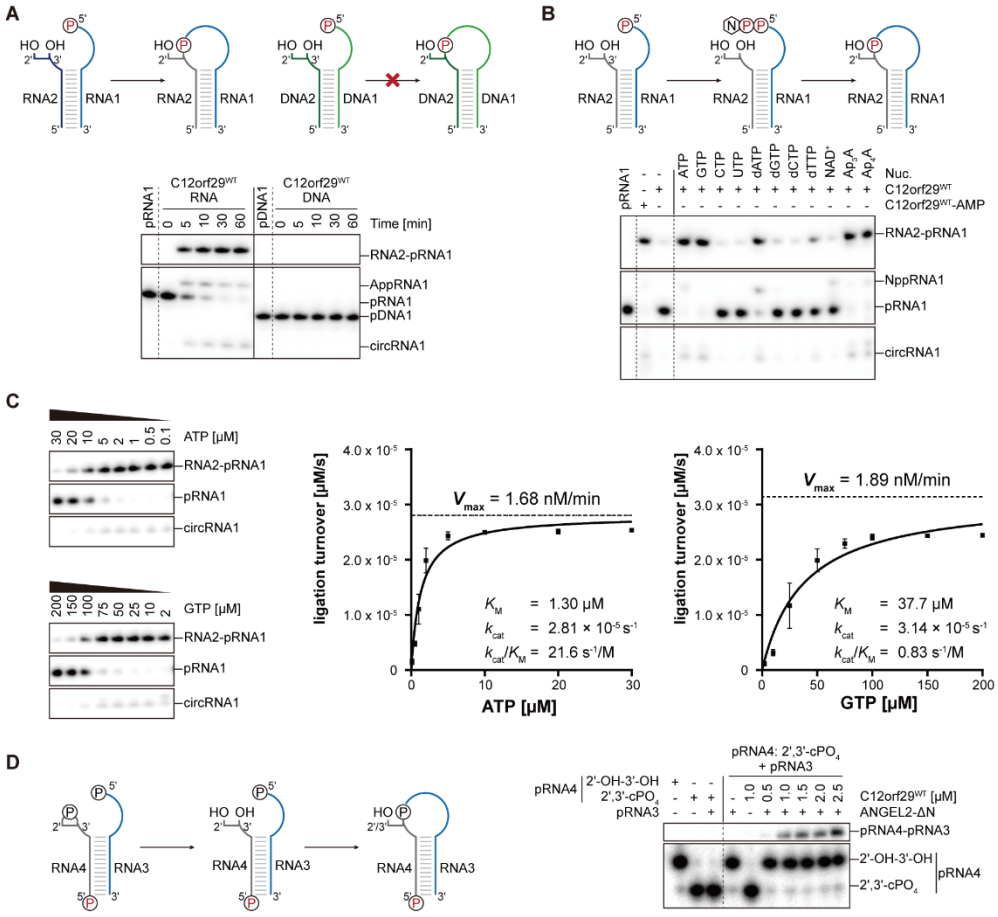

Fig. 2. RNA ligase activity of C12orf29. RNA/DNA oligonucleotides are schematically depicted in blue/green, respectively. The ^{32}P-labeled 5'-ends were depicted in red. The radioactive oligonucleotides were resolved by denaturing PAGE and analyzed by phosphorimaging. (**A**) (Top) Schematic display of two nucleic acid substrates used. (Bottom) PAGE analysis of reaction products after incubation with C12orf29 for various time points as indicated. Ligation was observed for 5'-phosphorylated RNA. (**B**) Reaction of C12orf29 with indicated RNA constructs using various nucleotides. Besides ATP, C12orf29 is able to efficiently catalyze ligation by processing GTP, dATP, Ap_3A and Ap_4A. (**C**) (Left) C12orf29 promoted ligation reaction at various concentrations of ATP and GTP. (Right) Michaelis-Menten fits to initial rates providing K_M, k_{cat} and k_{cat}/K_M for ATP and GTP. Plotted data represent the mean value ± SD for three biological replicates. (**D**) (Left) Proposed ligation scheme starting from 2',3'-cyclic phosphorylated RNA by the sequential action of ANGEL2-ΔN and C12orf29^{WT}. (Right) PAGE analysis of the depicted reaction. All oligonucleotide sequences are provided in fig. S4.

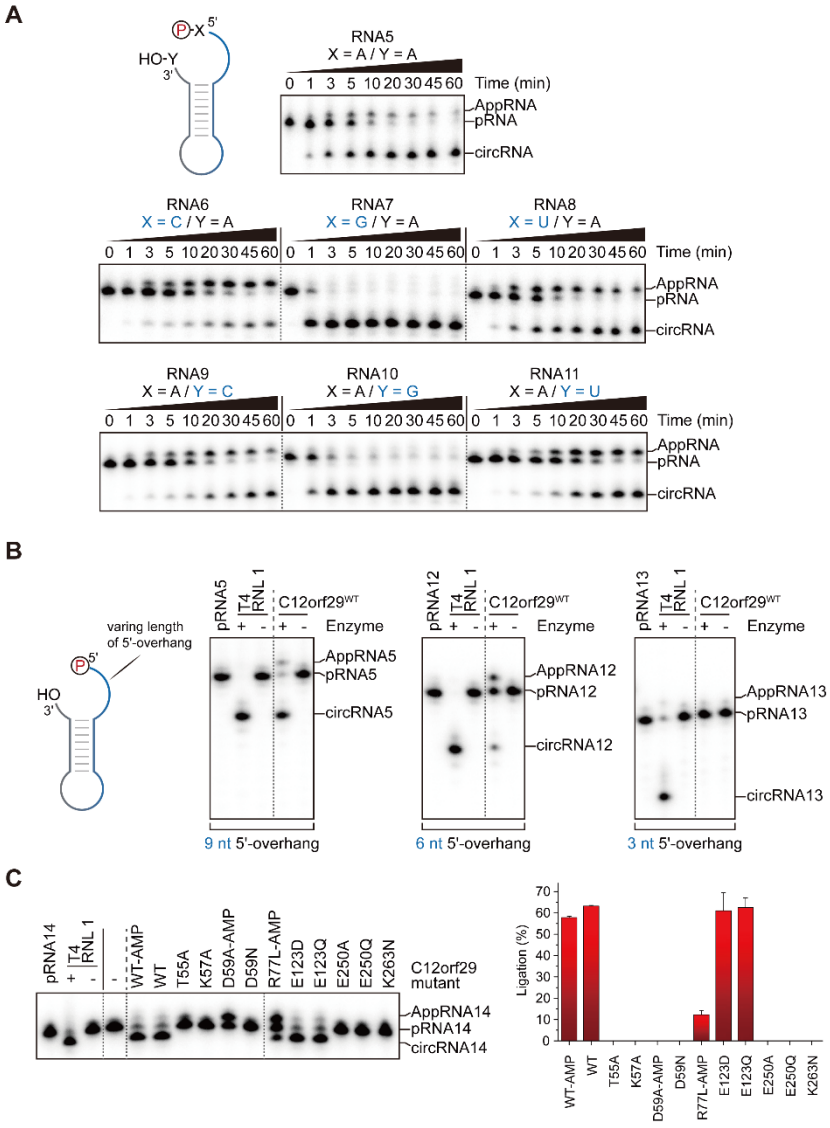

Fig. 3. Substrate scope of C12orf29. RNA oligonucleotides that were used for intramolecular ligation resulting in cyclization are schematically depicted. The ^{32}P-labeled 5'-ends were depicted in red. The radioactive oligonucleotides were resolved by denaturing PAGE and analyzed by phosphorimaging. (**A**) Investigation of the impact of the nucleobase composition at the 5'- and 3'- termini at the ligation site on the ligation efficiency of C12orf29. The depicted constructs were incubated under the same conditions for various time points. Most efficient ligation was observed when purines were at the ligation site. (**B**) Investigation of the impact of the length of the 5'-overhang on the ligation efficiency of C12orf29. (**C**) Impact of single site mutations on RNA ligase efficiency of C12orf29. All oligonucleotide sequences are provided in fig. S4.

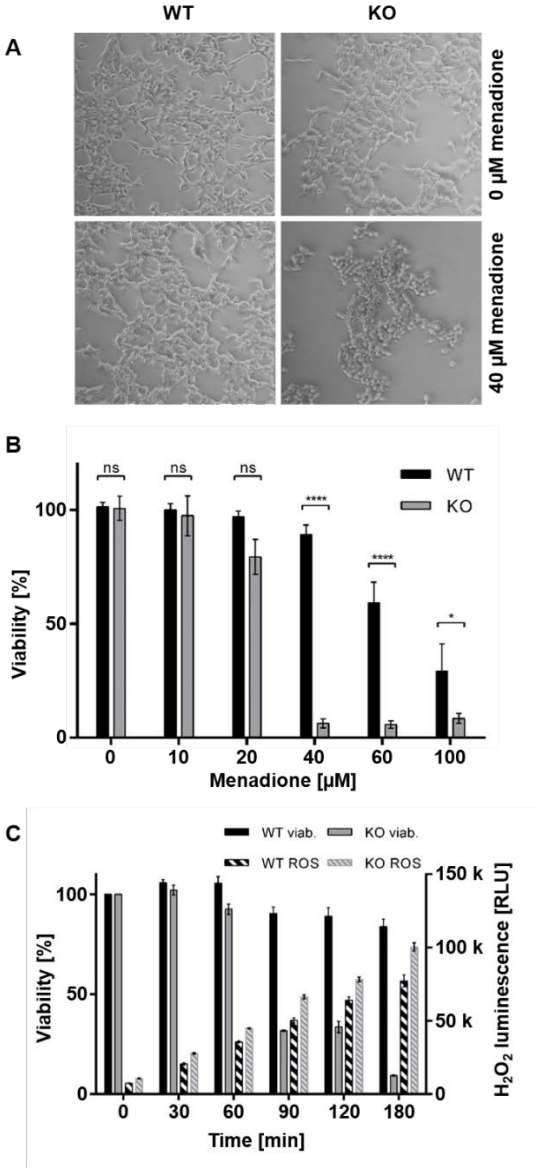

Fig.4. Effect of menadione-induced oxidative stress on WT and C12orf29-KO cells. (**A**) Light microscopy of HEK293 WT and *C12ORF29*-KO cells. The cells were treated with 40 µM menadione for 3 h or only with the carrier as control (0 µM menadione). (**B**) Cell viability of HEK293 WT and *C12ORF29*-KO cells after 3 h treatment with different menadione concentrations (n = 9, error bars represent the SD. Significance was calculated by two-way ANOVA with Sidak's multiple comparisons test: ns = P > 0.05; * = P ≤ 0.05; **** = P ≤ 0.0001). (**C**) Cell viability and corresponding H_2O_2 concentrations in HEK293 WT and *C12ORF29*-KO cells at different time points after treatment with 40 µM menadione (n = 3). Viab = viability.

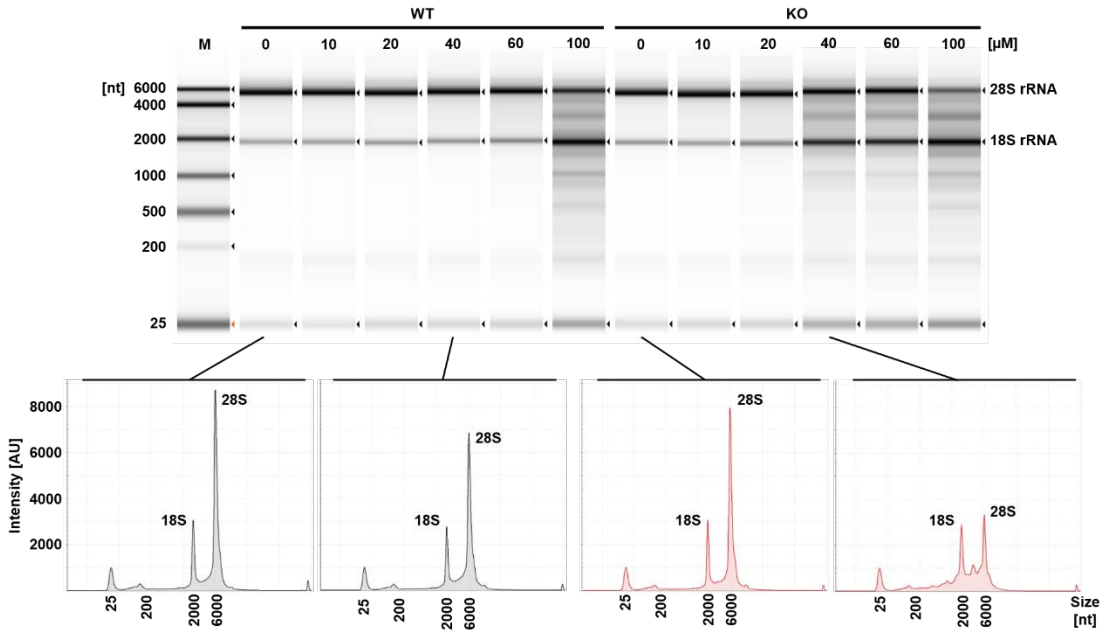

Fig.5. Analysis of total RNA from cell extracts of HEK293 WT and C12orf29-KO cells treated with various concentrations of menadione. (Top) Cells (left: HEK293 WT cells; right: C12orf29-KO HEK293 cells) were treated with different menadione concentrations for 3 h, than total RNA was isolated and subsequently analyzed by TapeStation. (Bottom) Electropherograms of the TapeStation analysis of total RNA from cell extracts of HEK293 WT (in black) and *C12ORF29*-KO (in red) cells.