Small non-coding vault RNA1-2 modulate expression of cell membrane proteins through nascent RNA silencing

Jana Terenova[1,3], Adele Alagia[1,3], Ruth F. Ketley[1], Irina Chelysheva[2] and Monika Gullerova[1,*]

[1] Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

[2] Oxford Vaccine Group, Department of Paediatrics, University of Oxford, and the NIHR Oxford Biomedical Research Centre, Oxford, UK

[3] these authors contributed equally

*Correspondence to: monika.gullerova@path.ox.ac.uk

Running title: svtRNA1-2 gene silencing

Key words: small RNA, Dicer, Argonaute, nascent RNA, membrane proteins, ERBB4

Abstract

Gene expression can be suppressed by post-transcriptional or by transcriptional gene silencing. Recently, we demonstrated that tRNA-derived small RNA (tsRNA) drive nascent RNA silencing of many genes. RNA polymerase III also transcribes the evolutionary conserved vault RNA (vtRNA), a crucial component of a large ribonucleoprotein complex Vault. Despite, only 5% of vault RNA are associated with the Vault complex, little is known about their Vault-independent cellular roles. Here we show that vtRNA are processed by Dicer into functional small vault RNA (svtRNA) fragments that associate with Argonaute2. Nascent RNA sequencing revealed the highest number of predicted svtRNA1-2 target genes among upregulated genes, in AGO2 and Dicer knockdown, in comparison to other svtRNA. SvtRNA1-2 target sites were detected predominantly in the intronic regions. Genomic deletion of vtRNA1-2 results in the severe inhibition of proliferation and upregulation of genes significantly associated with cell membrane physiology, including ERBB4, and cell adhesion. Finally, we show that svtRNA1-2 hybridises to target genes in a double-seeded manner with a loop protrusion required for intron targeting. Our data revealed a novel role for svtRNA1-2 in gene expression regulation, possessing unique molecular features for targeting genes in introns leading to nascent RNA silencing.

Background

Gene expression regulation is the hallmark of every living cell. In mammalian cells, genes can be silenced by two well-studied mechanisms. RNA interference (RNAi) facilitates post-transcriptional gene silencing (PTGS) by inhibition of mRNA translation[1] or by mRNA cleavage and degradation in the cytoplasm[2]. Another mechanism involves heterochromatin formation at target loci in the nucleus and leads to transcriptional gene silencing (TGS)[3, 4]. Furthermore, there are other alternative pathways regulating gene expression, for example, through tRNA-derived fragments[5] or various RNA modifications[6].

Various classes of small non-coding RNA (sncRNA) play canonical roles in many cellular processes, such as protein translation and splicing, but accumulating evidence suggests the existence of smaller fragments derived from non-canonical sncRNA (tRNA-derived small RNA (tsRNA)[7-9], rRNA-derived fragments (rRFs)[10] or Y-RNA fragments[11]) and their roles in gene expression regulation are similar to RNAi.

Recently, we have identified a group of tRNAs that are cleaved by Dicer into tsRNA and loaded onto AGO2 to target intronic sequences on many genes in the nucleus, leading to their nascent RNA silencing (NRS)[12]. In NRS, transcription of target genes remains unaffected, unlike in TGS the nascent RNA is degraded in the nucleus, in contrast to PTGS. Although we described the fundamental NRS mechanism, many of its molecular details remain enigmatic.

Similar to tRNA, vault RNA, another class of sncRNA is also transcribed by RNA polymerase III (RNAPIII). In the human genome, vault RNA are located on chromosome 5q31 within two different loci; the VAULT-1 locus consists of three genes (VTRNA1-1, VTRNA1-2 and VTRNA1-3), and the VAULT-2 locus encodes for VAULT2-1[13]. Both loci are under the control of an RNA Polymerase III type 2 promoter that contains box A and box B motifs typically[14] found within tRNA genes. Vault RNA molecules ranging from 80 to 140 nucleotides in length, depending on species[15, 16], therefore they are classified as small

ncRNAs. Vault RNA has been identified as a component of the large ribonucleoprotein parti known as Vaults that is involved in several cellular processes, such as nuclear transport[17, 18] , immune response[19], and drug resistance[20-22]. However, it has been shown that most of the vtRNA molecules (up to 95%) were not associated with the vault complex[23], suggesting that vtRNA may have additional roles within the cell. Indeed, some vtRNA have been shown to be processed into smaller fragments called svtRNA (small vault RNA). For example, svtRNA1-1 is associated with AGO proteins and is involved in the silencing of several target genes including CYP3A4 and CACNG7/8, via sequence complementarity[24, 25]. The svtRNA1-1 target genes are involved in pathways such as drug metabolism and encode voltage-gated calcium channels important for cell differentiations[26]. The unprocessed/intact vtRNA1-1 has also been shown to act as a riboregulator of autophagy[27].

Interestingly svtRNA2-1 were also shown to be AGO-associated, Drosha-independent and Dicer-dependent RNA molecules[24, 25, 28], which can lead to the silencing of target genes *via* sequence complementarity. Moreover, the biogenesis of some svtRNA (svtRNA1-1 and svtRNA1-3) have been described to be regulated by the NSUN2, RNA cytosine C (5)-methyltransferase, and serine/arginine-rich splicing factor 2 (SRSF2), similarly to tRNA[25, 26]. In fact, AGO association, Dicer processing, Drosha independence, RNAPIII transcription and RNA modifications are all common features between tRNA and vault RNA.

Decades of research on the mechanisms of ncRNA-target recognition have described rules governing ncRNA-mediated gene regulation. Factors such as the level of sequence complementarity and target sequence context have been demonstrated to influence the silencing outcome. Systematic analysis of the sequence determinants of RNA-induced silencing complex (RISC) association has integrated well-established rules, including seed region complementarity and 3’supplementary pairing, with previously undiscovered features

such as a central loop that allows the efficient pairing of the ncRNA to target nucleotides distant from each other[29].

Here we show that Dicer can process vtRNA into small fragments in human embryonic kidney 293T (HEK293T) cells. In particular, we identified a novel vtRNA1-2-derived fragment that associates with AGO2 in both the cytoplasm and the nucleus. Furthermore, we predicted 227 svtRNA1-2 target genes with dominant intronic complementarity. Genetic deletion of the vtRNA1-2 locus resulted in a severe proliferation defect. Analysis of chromatin-associated RNA-seq (ChrRNA-seq) derived from the vtRNA1-2 knockout (KO) cells, confirmed the involvement of svtRNA1-2 in nascent RNA silencing of the genes associated with plasma membrane and involved in cell signalling and cell adhesion regulation. Finally, we show that svtRNA1-2 hybridises to its target genes in a dual-seeded manner, which is required for intronic targeting. This study is the first report describing the mechanism of svtRNA1-2 biogenesis and its regulatory role in nascent RNA silencing.

Results

vtRNA1-2 is processed by Dicer into small RNA fragments

In order to characterise fragments derived from vtRNA and to examine the potential role of Dicer in the processing of vtRNA, we performed RNA sequencing of small RNA (sRNA) isolated from HEK293T Wild-Type (WT) and Dicer knockdown (KD) (HEK293 cells with integrated inducible shRNA) cells[12]. A principal component analysis (PCA) outlined clear differences between the two conditions and tight clustering of the replicates (**Fig. 1A**). The size distribution of mapped reads showed a distinct Dicer-dependent peak between 22-25nt long, including reads corresponding to mature miRNA (21-23nt) (**Fig. 1B**). The small difference in this peak between WT and Dicer KD is most likely affected by the presence of induced shRNA in Dicer KD sample (used to induce Dicer KD), which might mask the absence of proportion of mature miRNA. Among the identified sRNAs the most abundant species were miRNA, making up 47%, followed by tRNA with 23%. We also identified the four species of vault RNA (**Fig. 1C** and Supplementary Table 1), with >22nt long fragments. Interestingly, fragments derived from vtRNA1-2 were mostly 22-24nt long, suggesting that they might be Dicer products (Supplementary Fig. 1A). Indeed, moving average (MA) analysis showed that fragments derived from vtRNA1-2 and vtRNA2-1 were significantly down-regulated in Dicer KD (p-adj<0.05), suggesting Dicer is important for their biogenesis (**Fig. 1D** and Supplementary Fig. 1B). Interestingly, we identified more svtRNA1-1 reads upon Dicer KD when compared to WT, whilst the expression of vtRNA1-3 was not significantly altered between these samples (**Fig. 1D**), suggesting that Dicer is only involved in the processing of vtRNA1-2 and 2-1. These data were in agreement with additional analysis of small RNA-seq using a previously published dataset in WT and Dicer KD conditions (GSE55333)[30] (Supplementary Fig. 1C and Supplementary Table 4).

Next, we calculated the per-base read coverage of differentially expressed vtRNA in both WT and Dicer KD conditions. We identified regions along the vtRNA gene sequence with high read coverage, located

at the 3'end of vtRNA1-2 and the 5' end of vtRNA2-1, which were decreased in Dicer KD, indicating that vtRNA1-2 and vtRNA2-1 can be processed into small-vault RNA fragments in a Dicer-dependent manner (**Fig. 1E** and Supplementary Fig. 1C). Next, we analysed the reactivity[31] scores from (Qing-Jun Luo et al.) for each nucleotide on the vtRNA1-2 sequence and identified a cluster of highly reactive nucleotides between nucleotides 20 and 50, suggesting that this might be the region of vtRNA1-2 interaction with processing factors (Supplementary Fig. 1D). The newly identified svtRNA1-2 fragment was visualised on the vtRNA predicted secondary and tertiary structure (**Fig. 1F** and Supplementary Fig. 1D). Interestingly, detected reactive nucleotides on vtRNA1-1 were separated into two smaller clusters (around nucleotide 40 and nucleotide 60), which might suggest different processing of vtRNA1-1 to vtRNA1-2. The reactivity data on vtRNA2-1 were not included (Supplementary Fig. 1D). All four vtRNA sequences are highly similar in 5' and 3' regions but vary in the central domain (between nucleotides 30-60), which could be relevant to mentioned differences in reactive nucleotides distribution (Supplementary Fig. 2A). Interestingly, when we analysed the vault RNA1-2 sequence conservation we confirmed that 5' and 3' regions of vtRNA are highly conserved across vertebrates whilst the middle part is not (Supplementary Fig. 2B).

Altogether, these data reveal novel Dicer-dependent svtRNA1-2.

svtRNA1-2 associates with Argonaute proteins

Next, we hypothesised whether svtRNA1-2 is functional. We have previously identified that group of tsRNA-mediated NRS[12]. As vtRNA share many similarities with tRNA, we wished to assess the potential role of svtRNA1-2 in gene silencing. First, we investigated whether svtRNA1-2 fragments associate with Argonaute proteins by analysing PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation) publicly available datasets of AGO1-4 (GSE21918) and additionally AGO2/3 RIP-Seq (GSE55333) (Supplementary Table 4). Interestingly, we detected a

significant enrichment of Dicer-dependent svtRNA1-2 in the AGO2 PAR-CLIP. Fragments derived from vtRNA1-1 were present in the AGO2 PAR-CLIP data only at low levels, and svtRNA1-3 and 2-1 were not detected. We used microRNA-21 (miR-21) as a positive control (**Fig. 2A** and Supplementary Fig. 3A). We also observed svtRNA1-2 in the other Argonaute PAR-CLIP data (AGO1, AGO3 and AGO4) as well as in the Dicer PAR-CLIP dataset (Supplementary Table 4 and Supplementary Fig. 3A).

It has been shown that AGO2 can carry out its function in both the nucleus and the cytoplasm[32-34]. Given that we have shown that tsRNA function in the nucleus[12], we tested whether svtRNA1-2 might be present in a nuclear functional complex by analysing RNA bound to cytoplasmic and nuclear AGO2 in MCF-7 cells (GSE66665) (Supplementary Table 4). Surprisingly, the svtRNA1-2 fragment was enriched in both nuclear as well as in cytoplasmic AGO2 fractions (**Fig. 2B**). We used miR-21[35] as a positive control, and as expected it was detected in the nuclear fraction of AGO2 (**Fig. 2B**). Next, we transfected HEK239T cells with fluorescently labelled single- and double-stranded svtRNA1-2 and followed its cellular localisation at 4, 6, and 24 hours post-transfection. We observed mostly nuclear localisation of single-stranded and double-stranded svtRNA1-2 at all time points tested (**Fig. 2C** and Supplementary Fig. 3B). Taken together, this data shows that svtRNA1-2 fragments bind to AGO2 in both cellular compartments and can be detected in the nucleus, suggesting that it might play a role in nuclear gene silencing.

svtRNA1-2 targets selected genes on nascent RNA level

The Dicer dependency and association of svtRNA1-2 with AGO2, along with its nuclear localisation, prompted us to investigate whether svtRNA1-2 could regulate the gene expression on a nascent RNA level, similar to tsRNA. We employed ChrRNA-seq to characterise the nascent transcriptome of HEK293T cells in WT, AGO2 KD and Dicer KD conditions[12] (**Fig. 3A** and Supplementary Table 4). Differential expression analysis revealed that AGO2 and Dicer KD had a broad effect on the

transcriptome, with 1346 genes down-regulated and 2568 up-regulated in AGO2 KD, and 1052 genes down-regulated and 2391 up-regulated in Dicer KD (log2FC > 1/ log2FC < -1 and p-adj < 0.001). The up-regulated genes in AGO2 and Dicer KD conditions included mostly protein-coding genes and long non-coding RNA (**Fig. 3B and C** and Supplementary Table 2). In order to identify the role of svtRNA1-2 in gene silencing, we considered only the up-regulated protein-coding genes in both AGO2 and Dicer KD conditions (1272 genes), as we showed that svtRNA1-2 is Dicer-dependent and binds to AGO2 (**Fig. 3D**). We employed well-known and widely used MiRanda software[36] for *in silico* prediction of svtRNA1-2 gene targets through sequence complementarity, similar to miRNA binding sites, with restricted seed length and binding energy (see Methods). Interestingly, we found that the highest number of predicted gene targets are unique for svtRNA1-2 (227 genes, Supplementary Table 2), when compared to svtRNA1-1 (78 genes) or svtRNA1-3 (83 genes) (**Fig. 3D**). We used microRNA-4487 (miR-5587) as a negative control, as it has a sequence similar to svtRNA (**Fig. 3D** and Supplementary Fig. 4). Gene ontology enrichment analysis of the predicted svtRNA1-2 target genes revealed pathways involved in cell migration, motility and regulation of proliferation and growth (**Fig. 3E**). Indeed, transmembrane receptor tyrosine-protein kinase ERBB4 was among the top 5 predicted target genes with the highest log2 fold change (Supplementary Table 2). To investigate this further, we mapped the predicted binding sites along the transcript lengths and found a distinct density pattern for svtRNA1-2 distributed along the body of a gene. In contrast, the other svtRNA fragments were mapped mostly to the beginning of the transcripts (**Fig. 3F**). This could potentially suggest different roles for svtRNA1-2 and 1-1/1-3. Furthermore, we observed that svtRNA1-2 mapped to intronic regions with the highest occurrence suggesting a potential role of svtRNA1-2 in NRS (**Fig. 3G**). In summary, we identified genes that are up-regulated upon Dicer and AGO2 KD on a nascent level and are *in silico* predicted to be targeted by svtRNA1-2 mostly within introns.

svtRNA1-2 knockout leads to defects in cellular proliferation

To investigate svtRNA1-2 molecular function further, we applied an RNA-guided CRISPR Cas12a system targeting vtRNA1-2 with single guide RNAs (sgRNA) in HEK239T cells (**Fig. 4A** and Supplementary Fig. 5). We successfully generated individual vtRNA1-2 knock-out (KO) clones and confirmed partial deletion of vtRNA1-2 by sequencing and PCR (KO results in a shorter genomic fragment). As vault loci show relatively high sequence similarity (Supplementary Fig. 2A), we tested whether the other vtRNA loci remain intact in vtRNA1-2 KO. Indeed, vtRNA1-1, 1-3 and 2-1 genomic loci were not affected by vtRNA1-2 deletion (**Fig. 4B**). Next, we wished to validate the svtRNA1-2 predicted targets. We performed a qRT-PCR using RNA isolated from WT and stable vtRNA1-2 KO cell lines to validate the expression levels of ERBB4, a unique target for svtRNA1-2 (**Fig. 4C**) and PARP12 (a unique target for svtRNA1-1, used as negative control). We show that the ERBB4 expression was significantly increased at both the nascent and steady-state RNA level in vtRNA1-2 KO cells, whilst the expression of PARP12 remained unaffected (**Fig. 4D**).

Interestingly, we observed a severe growth phenotype in the vtRNA1-2 KO cells, when compared to WT cells, as shown by the clonogenic survival assay and MTS proliferation assay (**Fig. 4E** and **F**). The decreased cell viability and reduced proliferation of vtRNA1-2 KO cells, suggest that vtRNA1-2 plays important role in normal cell function.

Deletion of vtRNA1-2 alters the expression of genes associated with cellular membrane function

To further explore the role of svtRNA1-2 in nascent RNA gene silencing, we isolated nascent RNA from HEK293T WT and vtRNA1-2 KO cells and performed ChrRNA-seq (**Fig. 5A**). The purity of the fractions was confirmed by western blotting using H3 as a chromatin marker and Tubulin as a cytoplasmic marker (Supplementary Fig. 6A). A principal-component analysis outlined clear differences between the two conditions and tight clustering of the replicates (Supplementary Fig. 6B and C). The

differential expression analysis revealed 547 down-regulated and 944 up-regulated, mostly protein-coding genes in vtRNA1-2 KO cells (Supplementary Fig. 6D and E). Closer inspection showed that ERBB4 was among the upregulated genes in vtRNA1-2 KO, providing further validation for svtRNA1-2 mediated downregulation of ERBB4 (**Fig. 5B** and **C**).

Ontology enrichment analysis of genes upregulated in vtRNA1-2 KO cells revealed an association of protein-coding genes with biological processes such as regulation of membrane transport and potential (**Fig. 5D** and Supplementary Fig. 7A), while down-regulated genes were not enriched in these categories (Supplementary Fig. 7B). Interestingly, enrichment analysis that includes both up and down-regulated genes revealed a significant association of differentially expressed genes upon vtRNA1-2 KO with early and late estrogen response, most likely due to deregulated ERBB4 (**Fig. 5E**). Finally, the KEGG pathway confirmed altered ERBB4 signalling pathway (**Fig. 5F**).

These data propose that vtRNA1-2 is involved, directly and indirectly, in the regulation of membrane proteins through svtRNA1-2 silencing abilities.

SvtRNA1-2 target genes are associated with cellular membrane function

Next, we used MiRanda software to predict target genes for all potential svtRNA fragments in vtRNA1-2 KO up-regulated protein-coding genes. As expected, we identified most of the target genes for svtRNA1-2 (Supplementary Fig. 8A). Close inspection of MiRanda prediction revealed 139 target genes and 205 target sites for svtRNA1-2 (Supplementary Fig. 8B). To compare MiRanda with another target site prediction software PITA, we observed a good overlap of unique target genes and also more stringent MiRanda resulted in more predicted genes (Supplementary Fig. 8C and Supplementary Table 5 and 6). Therefore, for further analysis, we used targets identified by MiRanda. Metagene alignment showed an even distribution of svtRNA1-2 target sites, along the gene length, whilst svtRNA1-1 and 1-3 target sites were accumulated at the beginning of the genes, similar to what we observed in **Fig. 3F** (Supplementary

Fig. 8D). Finally, svtRNA1-2 target sites were predominantly mapped to the intronic regions. These results are consistent with the previous predictions (Supplementary Fig. 8E).

Focusing on unique svtRNA1-2 target genes, we analysed whether previously predicted svtRNA1-2 target genes that were upregulated upon Dicer and AGO2 knockdown and predicted to be targeted by svtRNA1-2 were also upregulated in vtRNA1-2 KO cells. We found 29 common genes including ERBB4 (**Fig. 6A**, Supplementary Fig. 8F and Supplementary Table 3). Interestingly, proteins encoded by these genes were mostly localised in the cellular membrane (**Fig. 6B**). Indeed, STRING[37] analysis identified 14 out of 29 to be involved in signalling and 17 out of 29 to be glycoproteins (**Fig. 6C**). Glycoproteins are membrane proteins that play a key role in cell migration, proliferation and adhesion[38]. Indeed, complex pathway analysis revealed that svtRNA1-2 target genes are involved in cell adhesion regulation (**Fig. 6D**). To validate the result, we seeded equal amount of WT and vtRNA1-2 KO cells and monitor their adhesion abilities. Interestingly, we observed elongated vtRNA1-2 KO cells with reduced cell to cell contact most likely as potential consequence of negatively regulated cell-to-cell adhesion (**Fig. 6E**). Overall, these data demonstrate that vtRNA1-2 is involved in the regulation of nascent RNA of protein-coding genes that are associated with cell membrane physiology.

svtRNA1-2 targets intronic regions through double-seeded complementarity and loop protrusion

Closer inspection of the MiRanda outputs for svtRNA1-2 predicted targets revealed that, unlike miRNA, svtRNA maps to its targets through a 5’ seed region and a 3’ supplementary seed region (Supplementary Table 5). In order to validate this observation, we compared the mapping of WT svtRNA1-2 to target genes with svtRNA1-2 5’ and 3’ seed regions swapped (svtRNA1-2 swapped), svtRNA1-2 with the 5’ seed region shuffled (shuffled 5’) and svtRNA1-2 with 3’ seed region shuffled (shuffled 3’) (**Fig. 7A**, 5’ seed sequence is highlighted in green, 3’ seed sequence is highlighted in yellow, shuffled bases are depicted in red; svtRNA1-1 was used as negative control). To prevent a biasing artefact, we kept the base

composition of all modified svtRNA sequences similar (**Fig. 7B**). All svtRNA sequences were run through MiRanda software against up-regulated genes in vtRNA1-2 KO to obtain their target predictions. Mapping of 5' shuffled svtRNA1-2 sequence resulted in complete loss of targets when compared to the svtRNA1-2 WT sequence, whilst 3' shuffled svtRNA1-2 mapped to only two targets. The swapped svtRNA1-2 mapped to 150 targets, but these were largely different from svtRNA1-2 WT. The observed partial overlap of targets between WT and swapped svtRNA1-2 was caused by swapped svtRNA1-2 mapping to svtRNA1-2 WT targets, but at different binding sites (**Fig. 7C** and **D**). Interestingly, swapped svtRNA1-2 was mapping mostly to the Exons, unlike svtRNA1-2 WT which was mapping predominantly to introns (**Fig. 7E**). Furthermore, we observed that svtRNA1-2 WT is mapping to its targets with a gap in between the 5' and 3' seed regions, creating a loop structure in the target nascent RNA sequence. This loop structure was lost in targets predicted for the swapped svtRNA1-2 (**Fig. 7F**). Finally, we tested whether the loop structure is associated with a certain number of targets for svtRNA1-2 WT and swapped. We compared the length of the gaps in all targets in svtRNA1-2 WT and the swapped svtRNA1-2. 23nt represents the original length of svtRNA1-2 with no gaps in the alignment. 28 nt contains a 5 nt gap across the alignment, which is the longest gap observed, creating a loop in the target sequence. We observed a striking difference between svtRNA1-2 WT and swapped svtRNA1-2 with the major feature for svtRNA1-2 WT is the presence of the loop in 179 targets (out of 205). Most of the swapped svtRNA1-2 targets corresponded to the length of the original sequence (23 nt). The distribution of 24-25-26-27-28 nt was distributed equally which indicates potential randomness in the alignments, which is not the case for svtRNA1-2 WT (**Fig. 7G**).

Altogether these analyses suggest that svtRNA1-2 uniquely maps to intronic regions of its targets through double-region complementarity, creating a loop structure in the target's nascent RNA.

Synthetic svtRNA1-2 downregulate ERBB4 expression in breast cancer cells

ERBB4 (HER4) activation stimulates several intracellular signalling pathways such as RAS/MEK/ERK, or PI3K/AKT leading to increased cell proliferation and migration. Increased expression or activation of ERBB proteins is frequently observed in various cancers[39]. Downregulation of ERBB4 has been shown to inhibit proliferation in invasive breast cancer cells MCF7[40]. To further assess the regulatory effect of svtRNA1-2 on ERBB4, we transfected human breast cancer MCF7 cells with synthetic single-stranded svtRNA1-2 WT, 5' shuffled svtRNA1-2 and 3' shuffled svtRNA1-2 and performed qRT-PCR. We observed that the nascent RNA levels of ERBB4 were significantly decreased in the presence of WT but not 5' or 3' shuffled svtRNA1-2 at 24 and 48 hours post-transfection (**Fig. 8A**). Next, we wished to see whether induced downregulation of ERBB4 by synthetic svtRNA1-2 WT might have any effect on the proliferation of MCF7 cells. Indeed, transfection of svtRNA1-2 WT led to severe inhibition of proliferation. We also observed a partial proliferation phenotype in cells transfected with 5' and 3' shuffled svtRNA1-2, although to a much lesser extent. This could be explained by either due to general toxicity or indirect effects of sRNA transfection (**Fig. 8B**). Finally, Ingenuity pathway analysis (Qiagen) illustrates how ERBB4 is involved in cell proliferation in breast cancer cells (**Fig. 8C**).

These data suggest that synthetic svtRNA1-2 can lead to specific downregulation of ERBB4 in MCF7 which can in part lead to inhibition of proliferation.

In summary, we show that Dicer-dependent svtRNA1-2 regulates a group of protein-coding genes that are associated with cell membrane physiology. SvtRNA1-2 is bound by AGO2 and promotes nuclear NRS by targeting intronic regions through canonical 5' and additional 3' sequence complementarity. Synthetic svtRNA1-2 can downregulate the expression of ERBB4 in breast cancer cells, which opens the possibility of using svtRNA fragments for therapeutic purposes.

Discussion

In this work, we have identified Dicer-dependent small RNA fragments derived from vtRNA1-2 and vtRNA2-1. We also found vtRNA1-1 to be independent of Dicer processing in HEK293T cells, which is in line with previous observations by Rybak-Wolf *et al.*[30]. However, it is worth noting that there are also contrasting results that show vtRNA1-1 as a Dicer substrate[24, 41]. This discrepancy could be potentially explained by the different sensitivity of the experimental techniques and also perhaps by differing efficiencies of Dicer KD. Furthermore, vtRNA are also differentially expressed in various tissues[16, 42-44], so vtRNA1-1 might be processed by Dicer in other cell lines. The analysis of RNA bound to Argonaute proteins using two different experimental methods identified svtRNA fragments derived from vtRNA1-1/2 to be bound to Argonaute proteins, whilst svtRNA2-1 fragment was not detected. In contrast, recent work showed the binding of AGO2 to the 5' vtRNA2-1 fragment by qRT-PCR in SH-SY5Y cells[28]. The potential explanation for this contrasting observation could be a differential expression of vault RNA in various tissues and pathological contexts. Genome-wide expression studies of small non-coding RNA, including svtRNA, in different cell types or tissues, would allow the identification of more detailed vtRNA expression patterns and could provide important insights into their biogenesis.

We found svtRNA1-2 to be associated with AGO2 not only in the cytoplasm but also in the nucleus. The nuclear localisation of svtRNA1-2 was confirmed by confocal microscopy using synthetic fluorescently labelled svtRNA1-2. As vtRNA share similar features with tRNA, we hypothesised that svtRNA1-2 could guide AGO2 in the nucleus to promote NRS[12]. Furthermore, AGO2 has previously been proposed to regulate gene expression in the nucleus[33].

In order to investigate the potential role of svtRNA1-2 in the NRS mechanism we generated chromatin-associated RNA datasets from Dicer and AGO2 KD cells and predicted svtRNA1-2 binding sites. We found that svtRNA1-2 mapped with the highest occupancy to the intronic regions in many genes that are

involved in cell membrane physiology, including ERBB4. Interestingly, genomic deletion of vtRNA1-2 resulted in inhibition of cell growth and proliferation. Such dramatic defects in cell growth and proliferation were not seen previously in vtRNA1-1 and vtRNA1-3 KO cells[45], suggesting that vtRNA1-2 might play a role in different cellular processes. Analysis of nascent RNA isolated from HEK293T WT and vtRNA1-2 KO further confirms the role of svtRNA1-2 in NRS.

Well-established rules for functional miRNA-target pairing, classify target sites into 3 main categories; canonical sites that pair at both 5' and 3'ends, seed sites, depending on minimal 5'end base pairing, and 3'compensatory sites characterised by weak 5'base pairing and strong compensatory pairing to the 3'end[46]. To assess the NRS requirements of the svtRNA1-2 molecule, we analysed svtRNA1-2 WT pairing conditions between svtRNA1-2 and predicted targets. A large set of predicted targets for svtRNA1-2 WT showed perfect 5'seed complementarity with extended wobble base pairs at 3'end. Nucleotide conservation analysis of the predicted targets (**Fig. 7F**) highlighted the presence of a GC-enriched segment at the 5'end that might have additional roles such as the 5' Terminal Oligo Guanine (TOG motif) described as a characteristic feature of some 5' tsRNA[47]. We cannot exclude that svtRNA1-2 might form intermolecular RNA G-quadruplex structures and gain further functions in translational inhibition, protein recognition, or cellular localization. In addition, svtRNA1-2 central (10-11-12) positions predominantly follow the base pairs scheme Wobble-WC-Wobble. It should be noted that most of the predicted targets have an unpaired loop of 5 nucleotides opposite to positions 11 and 12 of the svtRNA1-2, with a common sequence 5'-ACUAC or AUUAC. It has been demonstrated that AGO2 can accommodate a central loop present on the target gene, without loss of binding affinity[48]. Furthermore, we hypothesised that the target loop might be recognised as a consensus motif by some proteins. To further evaluate the minimal pairing requirements between svtRNA1-2 and the targets, we designed two svtRNA1-2 molecules with a shuffled sequence at either the 5'end or 3'supplementary region (shuffled svtRNA1-2 5' and shuffled svtRNA1-2 3', respectively). Interestingly, the svtRNA1-2

shuffled 5' designed to follow the rules of thermodynamic asymmetry, present in large numbers of miRNA molecules, did not have any common target genes with svtRNA1-2 WT (**Fig. 7C** and **D**). Loss of common targets could be connected to higher free energy values of shuffled svtRNA1-2 5' with respect to the WT svtRNA1-2. Shuffled svtRNA1-2 at the 3'end showed similar results to the svtRNA1-2 shuffled at the 5'end, resulting in no common targets with svtRNA1-2 WT (**Fig. 7C**). This data demonstrates that complementarity at both the 5' and 3'ends of the svtRNA1-2 is crucial for the identification of target genes. Since we reasoned that the complete loss of common targets between WT and 5' and 3' shuffled svtRNA1-2 might be partially due to the miRanda free energy cut-off, we designed a svtRNA1-2 molecule with a swapped 5' and 3' seed sequence. In this way, we preserved not only the nucleotide composition but also the sequence identity, while following the rules for the design of efficient miRNA molecules. From the analysis of the target predictions, we observed some increase in the putative targets and a dramatic change in the target sequence complementarity localisation. The swapped svtRNA1-2 hybridises within the Exons of the predicted genes, in contrast to the svtRNA1-2 WT which hybridises within introns via loop protrusion. This data sheds light on the importance of the G-rich region at the 5'end for the svtRNA1-2-mediated NRS.

Conclusions

In conclusion, in this work, we uncovered a Dicer-dependent small RNA fragment derived from vault RNA1-2. We show that svtRNA1-2 mediates nuclear gene silencing by targeting the introns of protein-coding genes that are associated with cell membrane function. Deletion of vtRNA1-2 results in a deregulated proliferation phenotype, which underpins the biological significance of vtRNA1-2. Finally, we show that svtRNA1-2 employs double complementarity with the formation of a loop protrusion within its target sequences, which is required for intronic binding. Although we cannot yet fully appreciate how the loop protrusion might be recognised and used for nascent RNA degradation, we

demonstrate a novel biological role for vtRNA1-2 and its potential for future RNA-based cancer therapies.

Methods

Small RNA analysis

FASTQ files were trimmed to remove the 3' adapter sequence (GATCGGAAGAGCACACGTCTGA ACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG) and filtered based on the length of the read (15 to 30) using Cutadapt (v.1.8.3)[49]. Following data pre-processing reads were then aligned with STAR aligner (v.2.5.3a)[50] to the human genome (hg38) downloaded from the GENCODE (v.34) with basic human genome annotation. Custom small non-coding RNAs annotation file was built by merging annotation of miRNA from miRbase (v.22.1)[51], tRNA from GENCODE (v.34) and snoRNA, snRNA, miscRNA, scaRNA, rRNA and vault RNA from GENCODE (v34) (hg38) basic genome annotation. Reads overlapping the custom annotation were quantified using featureCounts (v.1.6.2)[52]. DEGs were identified using the DESeq2[53] with an adjusted p-value < 0.05. The vault RNA fragments were identified using bedtools (v.2.29.2)[54] genomecov with the following settings –dz and –scale, which calculates the read coverage along the genome.

AGO1-4 PAR-CLIP analysis

The AGO1-4 PAR-CLIP datasets adapter sequences were trimmed using Cutadapt (v.1.8.3)[49]. Then the reads were mapped to the reference human genome (hg38) by Bowtie (v.1.1.2)[55], allowing for two alignment errors (mutation, insertion or deletion). For each read, only the best mapping was reported out of a maximum of 10 genomic matches. Any tag with over 10 genomic matches was discarded. After the conversion subtraction, reads that mapped to only one genomic location were retained for further analysis.

The clusters of reads (AGO binding sites) were identified based on T-to-C conversions using PARalyzer (v.1.5)[56] detailed settings described in (Supplementary Table 4). Then, these files were visualised using Integrative Genome Viewer (IGV)[57, 58].

AGO2/3 RIP-seq

The AGO2/3 RIP-seq datasets consisted of 1 replicate per AGO2/3 sample (HEK293T cell type). Next, the 3' adapter sequence (TCTCGTATCGTATGCCGTCTTCTGCTTG) was trimmed using fastx-clipper from fastx toolkit (v.0.0.13.2) from Hannon Lab with a quality score of 33 and read length longer than 15 nt.

The index of the human genome was built with bowtie-build (v.1.1.2)[55] and reads mapped to the reference human genome (hg38) using Bowtie (v.1.1.2)[55]. The Stranded-Coverage tool from GitHub (https://github.com/pmenzel/stranded-coverage), was used to calculate the read coverage for vtRNA in the genome with the following settings -f, -x, -n (normalisation of the coverage (Reads Per Million)) -s 1. Output files were visualised using IGV.

AGO2 RIP-seq Nuclear/Cytoplasmic

The reads were trimmed using cutadapt with the following settings -e 0.05 -q 23 -m 15 -M 30 -a "A[39]" and aligned to hg38 using STAR aligner allowing maximum of 3 mismatches and retaining only uniquely mapped reads. Next bamCoverage function was used from deeptools[59] to calculate coverage files using –binSize 1 and –normalizeUsing RPKM into BigWig files which were next visualised by IGV.

Dicer PAR-CLIP analysis

FASTQ files (three biological replicates) were trimmed to remove the 3' adapter sequence (TCTCACGTCGTATGCCGTCTTCTGCTTG,TCTCCATTCGTATGCCGTCTTCTGCTTG,TCTCC

CATCGTATGCCGTCTTCTGCTTG) and select reads longer than 15 bases in size using Cutadapt (v.1.8.3). The alignment was done using Bowtie (v.1.2.3) with the following settings –v 3 –a –m1 –best –strata. The clusters of reads (Dicer binding sites) were identified based on T-to-C conversions using PARalyzer (v.1.5) detailed settings described in (Supplementary Table 4). Then, these files were visualised using Integrative Genome Viewer (IGV).

ChrRNA-seq Dicer and AGO2 KD

The adapter sequences (-A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT and -a AGATCGGAAGAGCACACGTCT) were trimming from datasets (2 replicates per condition) by Cutadapt (v.1.8.3), discarding processed reads shorter than 10 nt. The processed reads were then mapped to the human reference genome (hg38) using STAR aligner (v.2.5.3a). The basic genome annotation was downloaded from the GENCODE website. The genes were quantified using featureCounts (v.1.6.2). Any gene with less than 5 read counts per replicate in at least two conditions were removed, and the counts' matrix was then normalised using DESeq2.

Generation of svt1.2 knockout cell line

Cas12a-gRNA ribonucleoprotein complexes containing one crRNA targeting the vault1-2 gene were transfected in HEK293T using the Alt-R CRISPR-Cas12a system (IDT). Transfected cells were incubated for 48h, thereafter single cells were sorted into 96-well plates. After 2 weeks, viable clones were genotyped using primers flanking the vault1-2 gene. Cells bearing homozygous edited genes were expanded and deletion of vtRNA1-2 was confirmed by Sanger sequencing.

MTS Proliferation assay

WT and vtRNA1-2 KO cells were seeded into 96 well plates at a low density and left to proliferate for 7 days. The MTS assay was carried out using the CellTiter 96® AQueous One Solution Cell Proliferation Assay (MTS) (Promega) according to the manufacturer's instructions, and absorbance was measured at 490nm.

Colony Formation Assay

To assess the impact of knockout of vtRNA1-2 on cell survival, colony formation assays were carried out. WT and vtRNA1-2 KO cells were seeded into 12 well plates at a density of 1000 cells per well and left for 10 days until visible colonies formed. Plates were then washed in PBS and stained in Crystal Violet staining solution (0.5% crystal violet, 20% ethanol) for 30 minutes. Plates were washed in water, left to dry, and visualised and quantified using the ColonyArea plugin on ImageJ.

Confocal and bright light microscopy

HEK293T cells were grown on a poly-L-lys coverslip and transfected with or single-stranded (100nM final concentration) or double-stranded (40nM final concentration) 5'-Cy3-labelled svtRNA1.2 24h after seeding. Cells were fixed at specific time points (4, 6, 24h) by paraformaldehyde (4% in PBS) and subsequently, the sample was imaged on an FV1200 Olympus laser scanning confocal microscope equipped with an argon laser and a He/Ne 633 nm laser, using 60x/NA1.4 oil immersion objective. Comparative immunofluorescence analyses were performed maintaining identical acquisition parameters. Nuclear foci were analysed with Cellprofile software and statistical analysis was performed using GraphPad Prism.

Bright light microscopy was performed using 40x lens.

svtRNA1-2 transfection

MCF-7 cells were reverse-transfected in 0.01% poly-L-Lys coated 6-well plates at a density of 1.2 × 106 cells per well with 100 nM of single-stranded vtRNA1.2 or 60 nM of double-stranded vtRNA1-2, using 9ul of Lipofectamine RNAiMAX reagent and following the manufacturer's instructions 24h after transfection, cells were split, and parallel cultures were maintained for 24h without any further treatments.

Chromatin associated RNA-sequencing sample preparation

Chromatin RNA samples for sequencing were prepared according to previous publications[12, 60] with some minor modifications. Briefly, WT and vtRNA1-2 KO cells were harvested from 15 cm dishes and lysed in HLB+N buffer (10mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40), and underlaid with HLB+NS buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl2, 0.5% NP-40, 10% sucrose), before centrifugation at 420g for 5 minutes. Nuclear pellets were then resuspended in NUN1 buffer (20 mM Tris-HCl (pH 7.9), 75 mM NaCl, 0.5 mM EDTA, 50% glycerol). Chromatin was extracted by addition of NUN2 buffer (20 mM HEPES-KOH (pH 7.6), 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl2, 1% NP-40, 1M urea) on ice for 15 minutes with vortexing. Samples were centrifuged at 17,000g for 10 minutes and chromatin pellets collected. Chromatin was digested by the addition of Proteinase K and Turbo DNase (Thermo Fisher). RNA was extracted from the chromatin samples using TRIzol (Invitrogen) and chloroform extraction, followed by the Monarch® Total RNA Miniprep Kit (NEB) according to the manufacturer's instructions and eluted in water.

Western blotting

4x Laemmli buffer (0.2 M Tris–HCl, 8% (w/v) SDS, 40% glycerol, 20% (v/v) β-mercaptoethanol, 0.005% bromophenol blue) was used to treat the whole cell and chromatin extracts following 95ºC

incubation for 5 min and sonication. Samples were separated on mini-PROTEAN® TGX™ gels (Bio-Rad Laboratories) followed by transfer onto nitrocellulose membrane (Protran, GE Healthcare) and probed with Tubulin (55kDa) and H3 (15kDa) antibodies.

qRT-PCR

Total RNA for gene expression analysis was extracted from cultured cells using Trizol and then phenol/chloroform. The RNA was precipitated and then stained by glycoblue. cDNA was prepared by reverse transcription using SuperScript III (1ug of RNA). The RT-PCR setup was divided into 2 steps. First 70°C for 5 mins with following the second step of 55°C for 1 hour and 70°C for 15mins.

cDNA samples were analysed using SYBR green. Cycling conditions were 95 °C for 10 minutes, and 50 cycles of 95 °C for 15 s, 63 °C for 15s and 72°C for 20s. GAPDH mRNA was used to normalise relative levels of mRNA for tested genes using the $2^{-\Delta\Delta Ct}$ method. All qRT-PCR experiments were performed using three biological replicates.

Primer sequences

GAPDH_REV_EX 5’-GACGAACATGGGGGCATCAG-3'

GAPDH_FWD 5’-AATTCCATGGCACCGTCAAGG-3'

GAPDH_INT 5’-CGCCAGACCCTGCACTTTTTAAG-3'

ERBB4_REV_EX 5’-CTGACTGAGAATCGCTGGGC-3'

ERBB4_FWD 5’-GCGACAGGACTTTGGGTCTG-3'

ERBB4_INT 5’-AGGCTCCCAAACAAAACGCT-3'

PARP12_REV_EX 5’-AAAGCCAGGGGTCGTTCTGA-3'

PARP12_FWD 5’-AACTCATGGCGTTGACCACC-3'

PARP12_INT 5’-GTGGAACTGGGTTGTTGGCA-3'

ChrRNA-seq WT and KO

The samples were prepared by isolating chromatin-associated RNA from HEK293T cells in Wild-Type and svtRNA1-2 knock-out conditions in triplicates. The reads were trimmed by Cutadapt (v.1.8.3), discarding processed reads shorter than 10 nt. The processed reads were then mapped to the human reference genome (hg38) using STAR aligner (v.2.5.3a). The basic genome annotation was downloaded from the GENCODE website. The genes were quantified using featureCounts (v.1.6.2) and the counts' matrix was then normalised using DESeq2. The Gene Ontology enrichment analysis was performed using ShinyGO 0.76[61].

Miranda target predictions

The sequences of the identified significantly up-regulated genes in Dicer KD, AGO2 KD conditions were downloaded from Biomart[62]. The identified svtRNA1-2 target genes were predicted by running miRanda with the following settings -sc 150 -en -30 -strict (v.3.3a)[36]. The information such as gene target name mapped svtRNA name, its coordinates and the scores were obtained from the miRanda output file. The annotation of the target genes was obtained from Biomart. The custom gene feature file (Intron, Exon, UTR5, UTR3, UTR5/Exon meaning UTR Exon junction same for UTR3/Exon) in BED format was constructed from GENCODE basic annotation file. The canonical transcript file was constructed from the GENCODE basic annotation file taking one transcript per gene with the highest number of relevant tags that represent transcript confidence. Next, we used bedtools (v.2.27.10) intersect function with selected options: -wa, -wb, -loj to identify the features of the mapped vault RNA fragment. By the intersection of predicting mapping coordinates and the genomic features coordinates file. Then we used Python 3 scripts to visualise the result. The calculated percentage of predicted mappings in a particular feature was done by normalising the number of predicted mappings to the feature's length.

PhastCons Conservation Analysis

We used phastCons scores resulting from the multiple alignment of 100 vertebrate genomes to the human genome. The files were downloaded in wigFix format from the UCSC database (hg38.100way.phastCons). We extracted the phastCons scores corresponding to vtRNA1-2 and calculated the average score. The density of the normalised phastCons scores was visualised using python3.

Secondary and tertiary structures of vtRNA

We used RNA immunoprecipitation pull-down selective 2′-hydroxyl acylation followed by Primer Extension (icSHAPE-MaP)[31] dataset to refine vtRNA secondary and tertiary structure predictions. RNAfold[63] command line version with vault RNA sequences in FASTA format and additional --shape=vault RNA icSHAPE-MaP data were used to predict secondary structures.

>vtRNA1-1

GGGCUGGCUUUAGCUCAGCGGUUACUUCGACAGUUCUUUAAUUGAAACAAGCAACCUG
UCUGGGUUGUUCGAGACCCGCGGGCGCUCUCCAGUCCUUUU

(((((((.....((((.((((((...((((...(((.....))).....(((((((....)))))))))))))).))))))))).....))))))).... (-58.45)

>vtRNA1-2

GGGCUGGCUUUAGCUCAGCGGUUACUUCGAGUACAUUGUAACCACCUCUCUGGGUGGU
UCGAGACCCGCGGGUGCUUUCCAGCUCUUUU

(((((((....(((((.((((....((((.....((((((((....)))))))))))))..)))).)).)))..))))))).... (-53.30)

>vtRNA1-3

GGGCUGGCUUUAGCUCAGCGGUUACUUCGCGUGUCAUCAAACCACCUCUCUGGGUUGU
UCGAGACCCGCGGGCGCUCUCCAGCCCUCUU

(((((((.....((((.((((.....(((.((.(((.(((.....))).))).)))))))))))))).....))))))).... (-58.21)

There were no icSHAPE-MaP data for vault RNA2-1.

Next, we used RNAComposer Automated RNA Structure 3D Modeling Server[64] to model 3D structures of the vtRNA. The output structure was then downloaded in .pdb format and imported into Jmol 9 (http://www.jmol.org/) for visualisation and further modifications.

Acknowledgements

We thank to all members of the Gullerova group. We acknowledge the ENCODE Consortium and the ENCODE production laboratory(s) for generating the dataset(s). This work was supported by the Senior Research Fellowship by Cancer Research UK [grant number BVR01170] awarded to M. G., EPA Trust Fund [BVR01670] awarded to M.G. and Lee Placito Fund awarded to M.G.

The ERBB4 target explorer was generated through the use of QIAGEN Ingenuity Target Explorer (QIAGEN, Inc., https://targetexplorer.ingenuity.com/).

The experimental illustrations were created with BioRender.com

Author contributions

J.T. conducted all bioinformatics analyses under I.C. supervision. J.T. performed qRT-PCR experiments, Image J colonies quantification, western blotting and vtRNA1-2 KO colony screening and wrote the first draft of the manuscript. A.A., R.F.K. and M.G. performed the experiments and analysed the data. M.G. conceived, designed and supervised the study and wrote the final draft of the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

Information about all published data sets used in this study is summarised in Supplementary Table 4. Raw data are deposited in GEO accession number GSE198298 and accessible with token: gtczyqgcvlmntkz (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198298). ChrRNA-seq data from WT and vtRNA1-2 KO cells can be viewed here: https://genome-euro.ucsc.edu/s/Janca9/vtRNA1%2D2_KO_chrRNA%2Dseq

FIGURE LEGENDS

Fig. 1 small RNA-seq reveals Dicer processing of vtRNA1-2 into smaller RNA fragments.

A) Principal-component analysis (PCA) plot shows three biological replicates corresponding to the samples from HEK293T cells in WT and Dicer KD conditions. The RNA composition of cells in the WT condition differed significantly from those in the Dicer KD condition. **B)** Size distribution of mapped reads to the human genome in Dicer KD and WT conditions. **C)** Pie chart show**s the** relative abundance of different RNA species with differentially altered expression. **D)** MA plot of Dicer KD vs WT. Points in red indicate sRNAs with an adjusted p-value < 0.05 and the points encircled in blue indicate vault RNA. **E)** Per base read coverage of vtRNA. The x-axis shows the length of transcripts, and the y-axis shows the normalised number of mapped reads RPM (Reads Per Million). miR-21 acts as a positive control. **F)** Predicted tertiary structure of vtRNA1-2. The identified small vault RNA1-2 fragment is coloured in red.

Fig. 2 svtRNA1-2 is bound to Argonaute proteins.

A) Integrative genomics viewer **(**IGV) tracks of AGO1-4 PAR-CLIP and AGO2/3 RIP-Seq show the association of AGO proteins with svtRNA1-2 (data values are shown in the top left corner). miR-21 was used as a positive control. The bottom black line indicates the length of the transcript and the sequence below represents the interacting fragment. Note: no reads were detected for vtRNA2-1. **B)** IGV tracks showing reads mapping to vtRNA in nuclear and cytoplasmic AGO2 PAR-CLIP data. The bottom black line indicates the length of the transcript, and the sequence below represents the interacting fragment. The top chromosome ideogram shows the position of each gene. miR-21 acts as a positive control for AGO2 binding. **C)** Confocal microscopy images show nuclear localisation of single-stranded (ss) svtRNA1-2 (green) at different time points. DAPI was used to stain the nuclei of the cells. Images were

quantified using Cell profiler software. Significance was calculated using t-test (****p< 0.0001; ***p< 0.001).

Fig. 3 svtRNA1-2 targets selected genes at the nascent RNA level.

A) Schematic representation of ChrRNA-seq method. **B)** Volcano plots of differentially expressed genes in AGO2 KD conditions. The red points show up-regulated genes and blue points the down-regulated genes with log2FC > 1/log2FC < -1 and p-adj < 0.001 respectively. Pie-charts show different gene types of up (red) or down (blue) regulated genes. **C**) As in B using Dicer KD samples. **D)** Venn diagram shows an overlap of up-regulated protein-coding genes from Dicer and AGO2 KD conditions (left). Venn diagram shows the number of predicted target genes for svtRNA1-1/2/3 and miR-5587 (right). svtRNA1-1/3 and miR-5587 were used as controls as their sequence similarity has the highest score with svtRNA1-2. **E)** GO analysis of vtRNA1-2 predicted targets **F)** Distribution of predicted svtRNA binding sites mapped along normalised target transcripts length and their corresponding gene features. **G)** svtRNA binding site counts normalised to the length of a feature displayed in overall percentage.

Fig. 4 vtRNA1-2 Knock-Out leads to severe cellular phenotype.

A) Schematic representation of CRISPR/Cas12a strategy. **B)** PCR analysis from purified genomic DNA from WT and vtRNA1-2 KO showing each vault RNA **C)** Schematic representation of ERBB4 gene with predicted binding sites for svtRNA1-2. **D**) qRT-PCR quantification of ERBB4 as a predicted target of svtRNA1-2 and PARP12 as a predicted target of svtRNA1-1 (act as control) in svtRNA1-2 KO cells normalised to GAPDH. The data represent the mean fold changes from three independent experiments (****p< 0.0001; ***p< 0.001; n.s not significant). **E)** Clonogenic assay of vtRNA1-2 KO and WT cells.

The cells were stained and counted after 10 days of growing (*** P <0.001). **F)** MTS assay showing proliferation of WT and vtRNA1-2 KO cells after 7 days (****P < 0.0001).

Fig. 5 Deletion of vtRNA1-2 alters the expression of genes associated with cellular membrane function.

A) Schematic representation of ChrRNA-seq method. B) IGV tracks of ChrRNA-seq from WT and vtRNA1-2 KO cells displaying the ERBB4 reads. The bottom blue line indicates the composition of ERBB4 gene. **C)** Volcano plots of differentially expressed genes in vtRNA1-2 KO and WT conditions. The red points show up-regulated genes and green points the down-regulated genes with log2FC > 1/log2FC < -1 and p-adj < 0.001 respectively. The yellow point shows ERBB4. D) GO analysis of up-regulated genes in vtRNA1-2 KO when compared to WT.

E) Gene Set Enrichment Analysis (GSEA) with gene set from vtRNA1-2 KO cells using Hallmark gene sets in MSigDB. Genes are ranked by Wald statistic values in a ChrRNA-seq experiment (KO vs. WT). A positive enrichment score indicates higher expression after vtRNA1-2 KO. F) GSEA with gene set from vtRNA1-2 KO cells using KEGG pathways.

Fig. 6 svtRNA1-2 target genes are associated with regulation of cell adhesion.

A) Venn diagram showing the overlap between up-regulated protein-coding genes from vtRNA1-2 KO (compared to WT), vtRNA1-2 KO svtRNA1-2 predicted targets and svtRNA1-2 predicted targets from up-regulated genes in AGO2 and Dicer KD cells (top). Tables showing names of 29 svtRNA1-2 target genes. **B)** Pie chart showing cellular localisation of 29 target genes. **C)** STRING analysis showing interactions among 29 svtRNA1-2 target genes. Proteins involved in signalling are marked in blue, and glycoproteins are marked in green. **D)** Pathways analysis showing association of svtRNA1-2 target genes

with negative regulation of cell adhesion. **E)** Representative examples of bright field microscopy showing differences in cell adhesion phenotype in WT and vtRNA1-2 KO cells.

Fig. 7 svtRNA1-2 target nascent RNA with double seeded complementarity

A) Sequence representation of svtRNA1-1, svtRNA1-2 WT, svtRNA1-2 swapped, svtRNA1-2 shuffled 5' and svtRNA1-2 shuffled 3'. svtRNA1-1 acts as control as its sequence has high sequence similarity to svtRNA1-2. **B)** Bar plot represents a base composition of each sequence displayed in percentage. **C)** Venn diagram showing an overlap of target genes predicted for each sequence. **D)** Table reporting number of predicted targets and number of predicted binding sites per sequence. **E)** Stacked bar plot showing predicted svtRNA binding sites mapped to different gene features with normalised counts to the length of a feature. Displayed in overall percentage. **F)** Sequence logo of significantly enriched bases in svtRNA1-2 predicted target sites (top). Line plot representing scores for predicted paired bases of svtRNA1-2 to their target sites (middle). Line plot representing scores for predicted paired bases of swapped svtRNA1-2 to their target sites (bottom). **G)** Table showing correlation analysis of the loop/no loop and number of target sites. 23 nt represents alignment with no loop, 28 nt contains a 5 nt long gap across the alignment, loop (the longest loop observed).

Fig. 8 svtRNA1-2 induced downregulation of ERBB4 leads to cell death in breast cancer cells.

A) qRT-PCR quantification of nascent ERBB4 in MCF7 cells after svtRNA1-2 WT, svtRNA1-2 shuffled 3' and svtRNA shuffled 5' sequence transfection. Samples were isolated after 24 and 48 hours. The values were normalised to the GAPDH. The data represent the mean fold changes from three/two independent experiments (***$P < 0.001$; **$P < 0.01$; n.s not significant). **B)** Clonogenic assay of MCF7 WT and cells transfected with 3' end and 5' end shuffled svtRNA1-2 sequence. The cells were stained and counted after 10 days of growing (****$P < 0.0001$, *** $P < 0.001$, **$P < 0.01$). **C)** Ingenuity pathway

analysis (IPA) of differentially expressed genes showing ERBB4 in the regulation of proliferation in breast cancer cells. **D)** The model of svtRNA1-2 nascent RNA silencing.

Supplementary Fig. 1

A) Size distribution of mapped reads to the vault RNA in Dicer KD (dashed line) and WT (full line) conditions. The shaded area around the line represents the standard deviation of the mean. **B)** Heatmap showing small RNA-seq data upon knock-down of Dicer in HEK293T cells. The colour scale indicates normalized intensities (z-score). The heatmap contains all genes that were significantly altered (adjusted $p < 0.05$) upon Dicer KD. Zoomed heatmap shows significantly expressed vtRNA genes. **C)** Per base read coverage of vtRNA. The x-axis shows the length of transcripts, and the y-axis shows the normalised number of mapped reads in RPM (Reads Per Million). miR-21 acts as a positive control. **D)** Predicted secondary structures of vtRNA1-2, 1-1 and 2-1. Reactivity score is shown for vtRNA1-2, 1-1 (left for each).

Supplementary Fig. 2

A) Vault RNA sequences alignment showing high sequence similarity. **B)** Conservation analysis of vtRNA1-2. The x-axis shows the sequence of vtRNA1-2, and the y-axis shows the conservation scores calculated from the phastCons multiple alignments of 100 vertebrate species.

Supplementary Fig. 3

a) IGV tracks of AGO1-4 PAR-CLIP, AGO2/3 RIP-Seq, Dicer PAR-CLIP and small RNA-seq in Dicer KD and WT conditions. Data values are shown in the top left corner. miR-21 as a positive control. The bottom line indicates the length of the transcript. **B)** Confocal microscopy images show nuclear localisation of double-stranded (ds) svtRNA1-2 (green) at different time points. DAPI was used to stain

nuclei of the cells. Images were quantified using Cell profiler software. Significance was calculated using t-test (****p< 0.0001; ***p< 0.001).

Supplementary Fig. 4

svtRNA and miR-5587 sequence alignment show bases with high sequence similarity.

Supplementary Fig. 5

Sanger sequencing results confirm vtRNA1-2 knock-out. Retained and removed parts of vtRNA1-2 gene are highlighted.

Supplementary Fig. 6

A) Western blot of chromatin and whole cell lysate fractions components isolated for ChrRNA sequencing. **B)** Principal-component analysis (PCA) plot shows three biological replicates corresponding to the samples from HEK293T cells in WT and vtRNA1-2 KO conditions. The RNA composition of cells in the WT condition differed significantly from those in the vtRNA1-2 KO condition. **C)** Box plot showing log2-transformed read counts in 3 biological replicates per condition. **D)** Pie chart showing gene type distribution of down-regulated genes in vtRNA1-2 KO cells. **E)** as in **D**, for up-regulated genes

Supplementary Fig. 7

A) GO analysis showing up-regulated biological processes in vtRNA1-2-KO when compared to WT cells. Bars represent the −log10 (P-value) and the frequency polygon (black line) denotes the number of genes. **B)** As in **A**, for down-regulated genes.

Supplementary Fig. 8

A) Venn diagram showing an overlap of target genes predicted for each svtRNA and miR-5587. **B)** Bar chart showing predicted unique target genes (left) and binding sites (right) for each svtRNA and miR-5587. **C)** Pie chart comparing a number of predicted target genes identified by PITA and MiRanda software for each svtRNA and miR-5587. **D)** Distribution of predicted svtRNA binding sites mapped along normalised target transcripts length and their corresponding gene features. **E)** svtRNA binding site counts normalised to the length of a feature displayed in overall percentage. **F)** Heatmap showing levels of significantly overexpressed 29 svtRNA1-2 target genes.

References

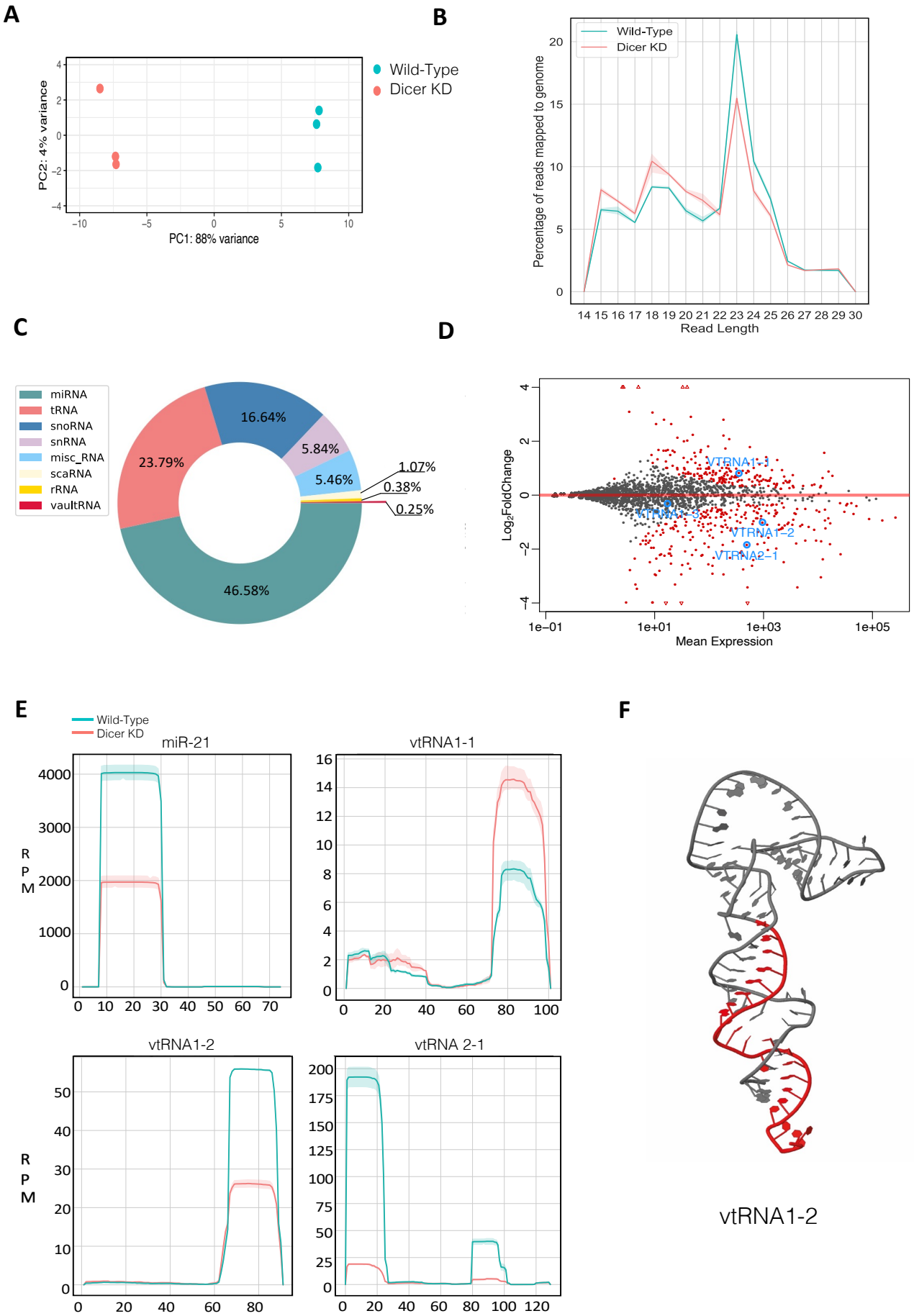
1. Jonas S, Izaurralde E: **Towards a molecular understanding of microRNA-mediated gene silencing.** *Nat Rev Genet* 2015, **16:**421-433.
2. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ: **Argonaute2 is the catalytic engine of mammalian RNAi.** *Science* 2004, **305:**1437-1441.
3. Castel SE, Martienssen RA: **RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond.** *Nat Rev Genet* 2013, **14:**100-112.
4. Martienssen R, Moazed D: **RNAi and heterochromatin assembly.** *Cold Spring Harb Perspect Biol* 2015, **7:**a019323.
5. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P: **Angiogenin-induced tRNA fragments inhibit translation initiation.** *Mol Cell* 2011, **43:**613-623.
6. Fu Y, Dominissini D, Rechavi G, He C: **Gene expression regulation mediated through reversible m(6)A RNA methylation.** *Nat Rev Genet* 2014, **15:**293-306.
7. Soares AR, Fernandes N, Reverendo M, Araujo HR, Oliveira JL, Moura GM, Santos MA: **Conserved and highly expressed tRNA derived fragments in zebrafish.** *BMC Mol Biol* 2015, **16:**22.
8. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, Green PJ, Barton GJ, Hutvagner G: **Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs.** *RNA* 2009, **15:**2147-2160.
9. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: **Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs.** *Genes Dev* 2008, **22:**2773-2785.
10. Chen Z, Sun Y, Yang X, Wu Z, Guo K, Niu X, Wang Q, Ruan J, Bu W, Gao S: **Two featured series of rRNA-derived RNA fragments (rRFs) constitute a novel class of small RNAs.** *PLoS One* 2017, **12:**e0176458.
11. Dhahbi JM, Spindler SR, Atamna H, Boffelli D, Mote P, Martin DI: **5'-YRNA fragments derived by processing of transcripts from specific YRNA genes and pseudogenes are abundant in human serum and plasma.** *Physiol Genomics* 2013, **45:**990-998.
12. Di Fazio A, Schlackow M, Pong SK, Alagia A, Gullerova M: **Dicer dependent tRNA derived small RNAs promote nascent RNA silencing.** *Nucleic Acids Res* 2022, **50:**1734-1752.

13. Buscher M, Horos R, Hentze MW: **'High vault-age': non-coding RNA control of autophagy.** *Open Biol* 2020, **10:**190307.
14. van Zon A, Mossink MH, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EA: **Multiple human vault RNAs. Expression and association with the vault complex.** *J Biol Chem* 2001, **276:**37715-37721.
15. Ferro I, Gavini J, Gallo S, Bracher L, Landolfo M, Candinas D, Stroka DM, Polacek N: **The human vault RNA enhances tumorigenesis and chemoresistance through the lysosome in hepatocellular carcinoma.** *Autophagy* 2022, **18:**191-203.
16. Stadler PF, Chen JJ, Hackermuller J, Hoffmann S, Horn F, Khaitovich P, Kretzschmar AK, Mosig A, Prohaska SJ, Qi X, et al: **Evolution of vault RNAs.** *Mol Biol Evol* 2009, **26:**1975-1991.
17. Dickenson NE, Moore D, Suprenant KA, Dunn RC: **Vault ribonucleoprotein particles and the central mass of the nuclear pore complex.** *Photochem Photobiol* 2007, **83:**686-691.
18. Kedersha NL, Heuser JE, Chugani DC, Rome LH: **Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry.** *J Cell Biol* 1991, **112:**225-235.
19. Berger W, Steiner E, Grusch M, Elbling L, Micksche M: **Vaults and the major vault protein: novel roles in signal pathway regulation and immunity.** *Cell Mol Life Sci* 2009, **66:**43-61.
20. Scheffer GL, Wijngaard PL, Flens MJ, Izquierdo MA, Slovak ML, Pinedo HM, Meijer CJ, Clevers HC, Scheper RJ: **The drug resistance-related protein LRP is the human major vault protein.** *Nat Med* 1995, **1:**578-582.
21. Kickhoefer VA, Rajavel KS, Scheffer GL, Dalton WS, Scheper RJ, Rome LH: **Vaults are up-regulated in multidrug-resistant cancer cell lines.** *J Biol Chem* 1998, **273:**8971-8974.
22. Mossink MH, van Zon A, Scheper RJ, Sonneveld P, Wiemer EA: **Vaults: a ribonucleoprotein particle involved in drug resistance?** *Oncogene* 2003, **22:**7458-7467.
23. Kickhoefer VA, Poderycki MJ, Chan EK, Rome LH: **The La RNA-binding protein interacts with the vault RNA and is a vault-associated protein.** *J Biol Chem* 2002, **277:**41282-41286.
24. Persson H, Kvist A, Vallon-Christersson J, Medstrand P, Borg A, Rovira C: **The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs.** *Nat Cell Biol* 2009, **11:**1268-1271.
25. Hussain S, Sajini AA, Blanco S, Dietmann S, Lombard P, Sugimoto Y, Paramor M, Gleeson JG, Odom DT, Ule J, Frye M: **NSun2-mediated cytosine-5 methylation of vault noncoding RNA determines its processing into regulatory small RNAs.** *Cell Rep* 2013, **4:**255-261.
26. Sajini AA, Choudhury NR, Wagner RE, Bornelov S, Selmi T, Spanos C, Dietmann S, Rappsilber J, Michlewski G, Frye M: **Loss of 5-methylcytosine alters the biogenesis of vault-derived small RNAs to coordinate epidermal differentiation.** *Nat Commun* 2019, **10:**2550.
27. Horos R, Buscher M, Kleinendorst R, Alleaume AM, Tarafder AK, Schwarzl T, Dziuba D, Tischer C, Zielonka EM, Adak A, et al: **The Small Non-coding Vault RNA1-1 Acts as a Riboregulator of Autophagy.** *Cell* 2019, **176:**1054-1067 e1012.
28. Minones-Moyano E, Friedlander MR, Pallares J, Kagerbauer B, Porta S, Escaramis G, Ferrer I, Estivill X, Marti E: **Upregulation of a small vault RNA (svtRNA2-1a) is an early event in Parkinson disease and induces neuronal dysfunction.** *RNA Biol* 2013, **10:**1093-1106.
29. Filipowicz W, Bhattacharyya SN, Sonenberg N: **Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?** *Nat Rev Genet* 2008, **9:**102-114.
30. Rybak-Wolf A, Jens M, Murakawa Y, Herzog M, Landthaler M, Rajewsky N: **A variety of dicer substrates in human and C. elegans.** *Cell* 2014, **159:**1153-1167.

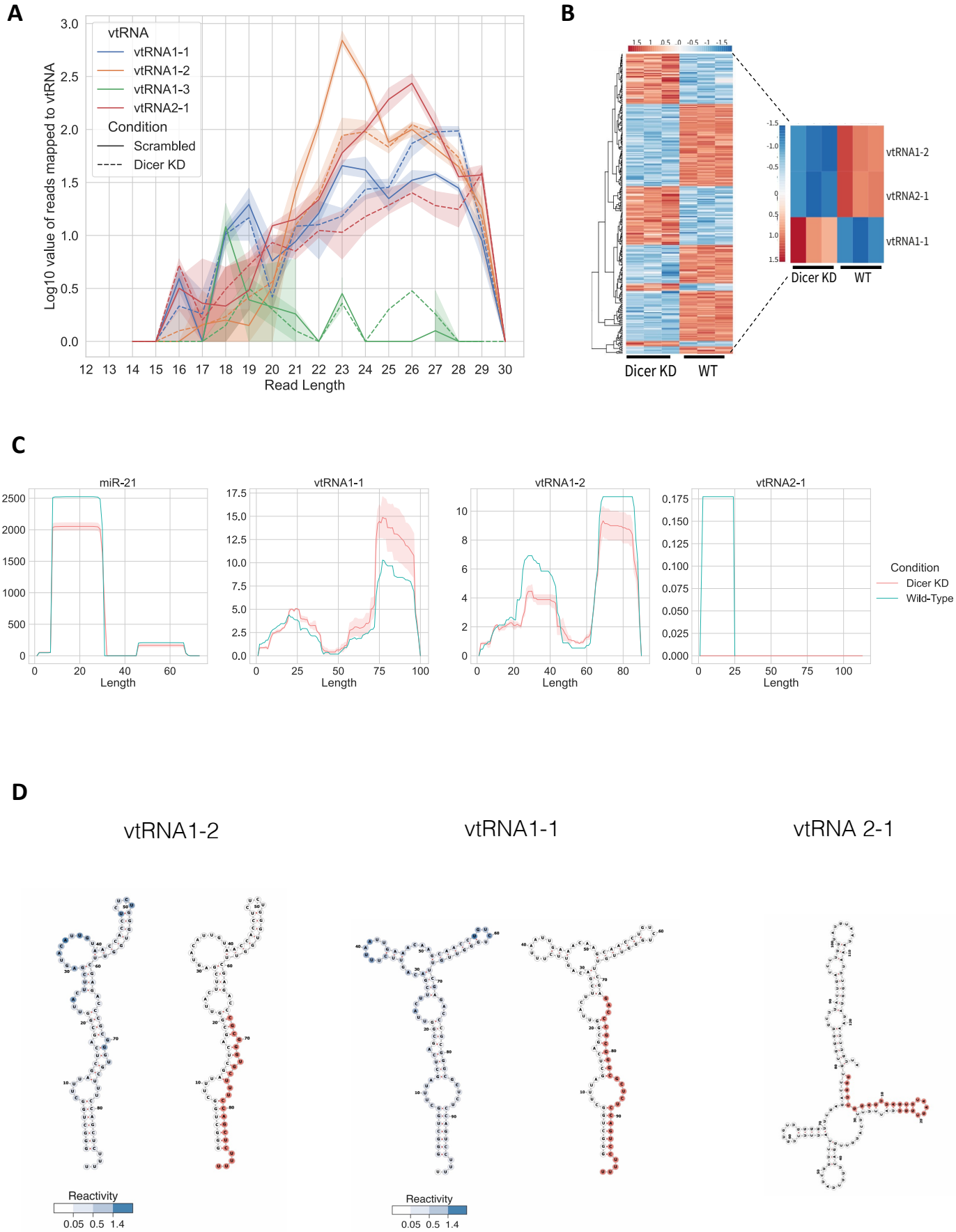
31. Luo QJ, Zhang J, Li P, Wang Q, Zhang Y, Roy-Chaudhuri B, Xu J, Kay MA, Zhang QC: **RNA structure probing reveals the structural basis of Dicer binding and cleavage.** *Nat Commun* 2021, **12:**3397.
32. Gong M, Zhang X, Wang Y, Mao G, Ou Y, Wei C, Hu X, Xiang S: **DDX21 interacts with nuclear AGO2 and regulates the alternative splicing of SMN2.** *Biosci Biotechnol Biochem* 2021, **85:**272-279.
33. Sarshad AA, Juan AH, Muler AIC, Anastasakis DG, Wang X, Genzor P, Feng X, Tsai PF, Sun HW, Haase AD, et al: **Argonaute-miRNA Complexes Silence Target mRNAs in the Nucleus of Mammalian Stem Cells.** *Mol Cell* 2018, **71:**1040-1050 e1048.
34. Kalantari R, Hicks JA, Li L, Gagnon KT, Sridhara V, Lemoff A, Mirzaei H, Corey DR: **Stable association of RNAi machinery is conserved between the cytoplasm and nucleus of human cells.** *RNA* 2016, **22:**1085-1098.
35. Kriegel AJ, Terhune SS, Greene AS, Noon KR, Pereckas MS, Liang M: **Isomer-specific effect of microRNA miR-29b on nuclear morphology.** *J Biol Chem* 2018, **293:**14080-14088.
36. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome Biol* 2003, **5:**R1.
37. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al: **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.** *Nucleic Acids Res* 2019, **47:**D607-D613.
38. Gu J, Isaji T, Xu Q, Kariya Y, Gu W, Fukuda T, Du Y: **Potential roles of N-glycosylation in cell adhesion.** *Glycoconj J* 2012, **29:**599-607.
39. Arteaga CL, Engelman JA: **ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics.** *Cancer Cell* 2014, **25:**282-303.
40. Okazaki S, Nakatani F, Masuko K, Tsuchihashi K, Ueda S, Masuko T, Saya H, Nagano O: **Development of an ErbB4 monoclonal antibody that blocks neuregulin-1-induced ErbB4 activation in cancer cells.** *Biochem Biophys Res Commun* 2016, **470:**239-244.
41. Langenberger D, Cakir MV, Hoffmann S, Stadler PF: **Dicer-processed small RNAs: rules and exceptions.** *J Exp Zool B Mol Dev Evol* 2013, **320:**35-46.
42. Hahne JC, Lampis A, Valeri N: **Vault RNAs: hidden gems in RNA and protein regulation.** *Cell Mol Life Sci* 2021, **78:**1487-1499.
43. Gallo S, Kong E, Ferro I, Polacek N: **Small but Powerful: The Human Vault RNAs as Multifaceted Modulators of Pro-Survival Characteristics and Tumorigenesis.** *Cancers (Basel)* 2022, **14**.
44. Fort RS, Duhagon MA: **Pan-cancer chromatin analysis of the human vtRNA genes uncovers their association with cancer biology.** *F1000Res* 2021, **10:**182.
45. Bracher L, Ferro I, Pulido-Quetglas C, Ruepp MD, Johnson R, Polacek N: **Human vtRNA1-1 Levels Modulate Signaling Pathways and Regulate Apoptosis in Human Cancer Cells.** *Biomolecules* 2020, **10**.
46. Alagia A, Eritja R: **siRNA and RNAi optimization.** *Wiley Interdiscip Rev RNA* 2016, **7:**316-329.
47. Shi J, Zhang Y, Zhou T, Chen Q: **tsRNAs: The Swiss Army Knife for Translational Regulation.** *Trends Biochem Sci* 2019, **44:**185-189.
48. Becker WR, Ober-Reynolds B, Jouravleva K, Jolly SM, Zamore PD, Greenleaf WJ: **High-Throughput Analysis Reveals Rules for Target RNA Binding and Cleavage by AGO2.** *Mol Cell* 2019, **75:**741-755 e711.
49. Marcel M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnetjournal* 2011, **17:**10--12.

50. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.
51. Kozomara A, Birgaoanu M, Griffiths-Jones S: **miRBase: from microRNA sequences to function.** *Nucleic Acids Research* 2019, **47:**D155-D162.
52. Liao Y, Smyth GK, Shi W: **featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.** *Bioinformatics* 2014, **30:**923-930.
53. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15:**550.
54. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841-842.
55. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
56. Corcoran DL, Georgiev S, Mukherjee N, Gottwein E, Skalsky RL, Keene JD, Ohler U: **PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data.** *Genome Biol* 2011, **12:**R79.
57. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer.** *Nat Biotechnol* 2011, **29:**24-26.
58. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** *Brief Bioinform* 2013, **14:**178-192.
59. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T: **deepTools: a flexible platform for exploring deep-sequencing data.** *Nucleic Acids Res* 2014, **42:**W187-191.
60. Nojima T, Gomes T, Carmo-Fonseca M, Proudfoot NJ: **Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide.** *Nat Protoc* 2016, **11:**413-428.
61. Ge SX, Jung D, Yao R: **ShinyGO: a graphical gene-set enrichment tool for animals and plants.** *Bioinformatics* 2020, **36:**2628-2629.
62. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al: **Ensembl BioMarts: a hub for data retrieval across taxonomic space.** *Database (Oxford)* 2011, **2011:**bar030.
63. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL: **ViennaRNA Package 2.0.** *Algorithms Mol Biol* 2011, **6:**26.
64. Antczak M, Popenda M, Zok T, Sarzynska J, Ratajczak T, Tomczyk K, Adamiak RW, Szachniuk M: **New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure.** *Acta Biochim Pol* 2016, **63:**737-744.

Figure 1: small RNA-seq reveals Dicer processing of vtRNA1-2 into smaller RNA fragments.

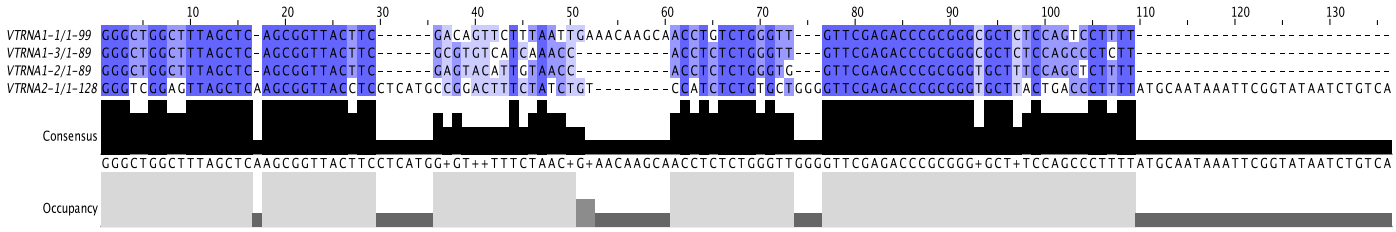


Supplementary Figure 1



Supplementary Figure 2

A



B

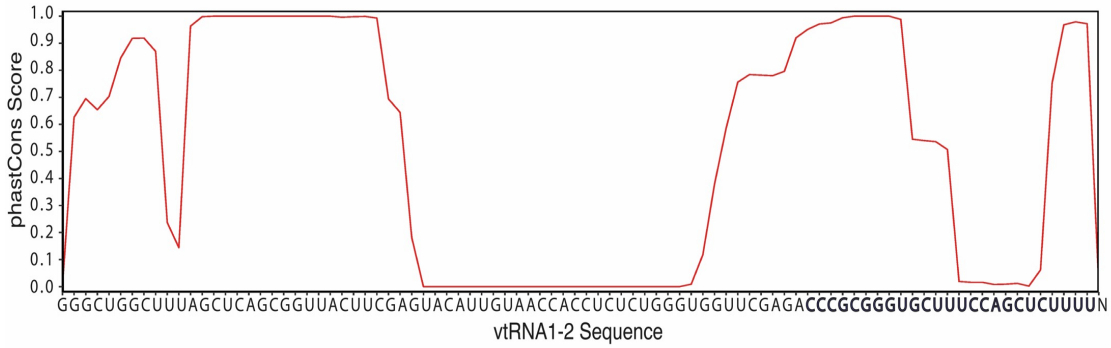
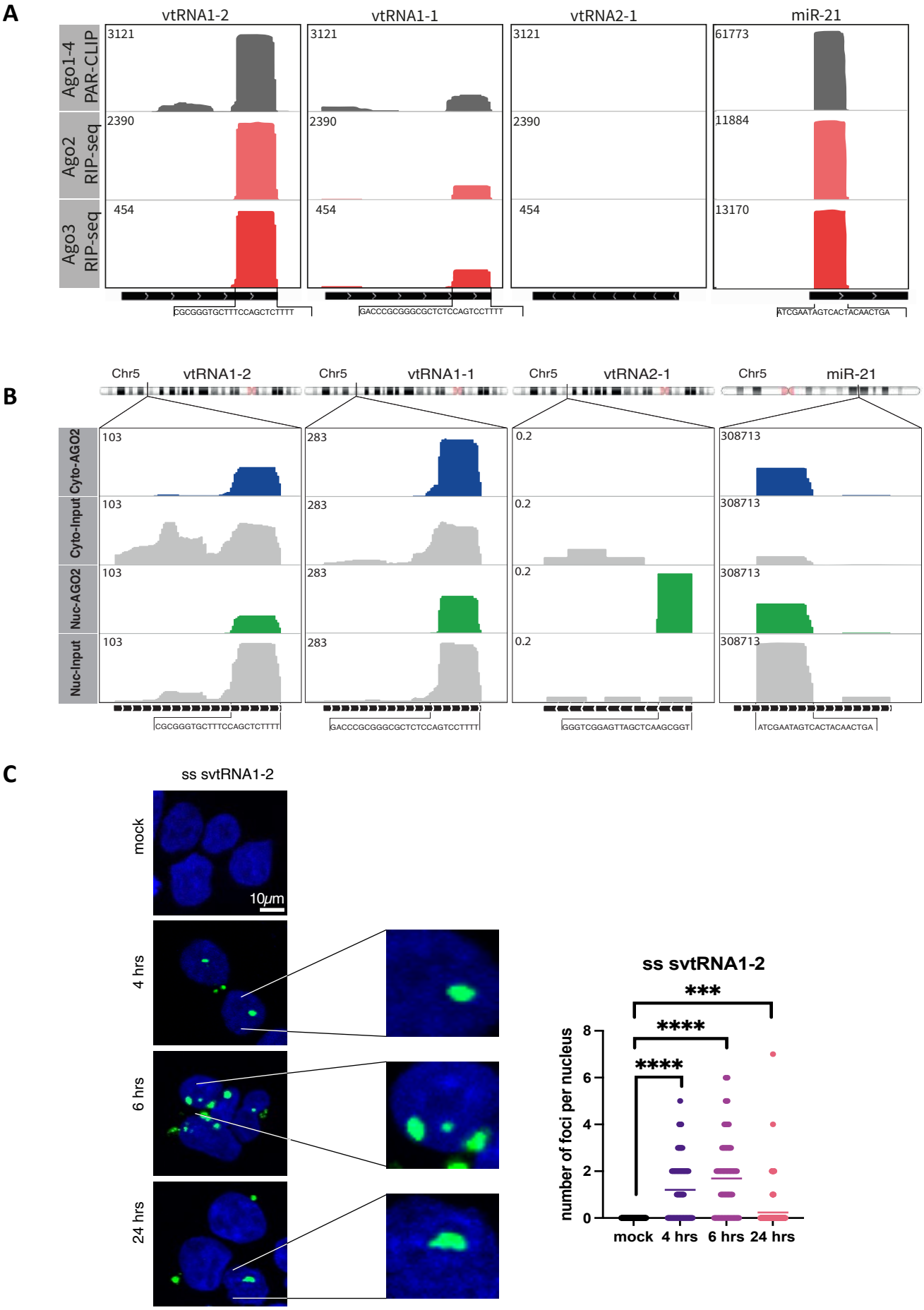
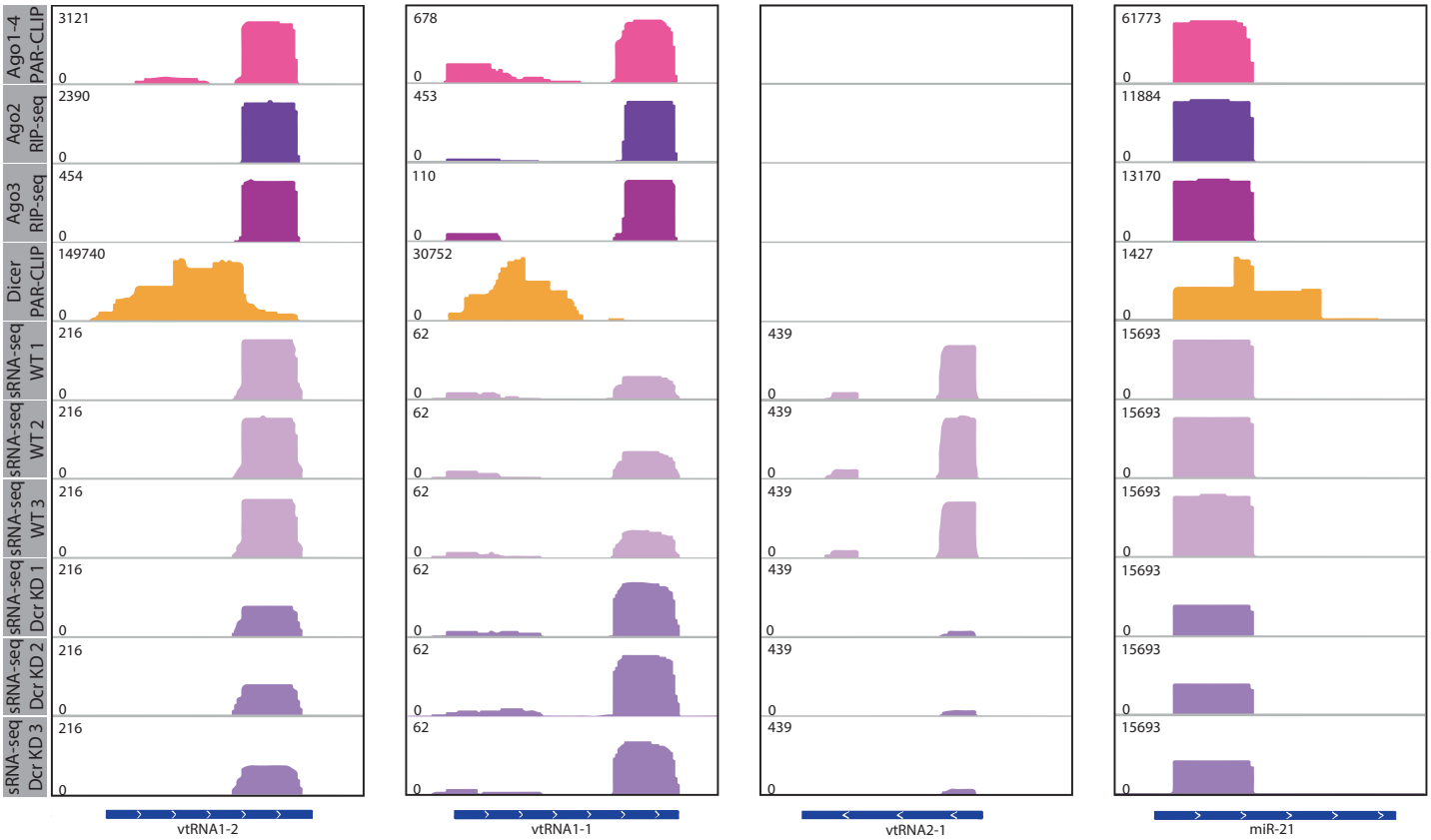


Figure 2: svtRNA1-2 is bound to Argonaute proteins.



Supplementary Figure 3

A



B

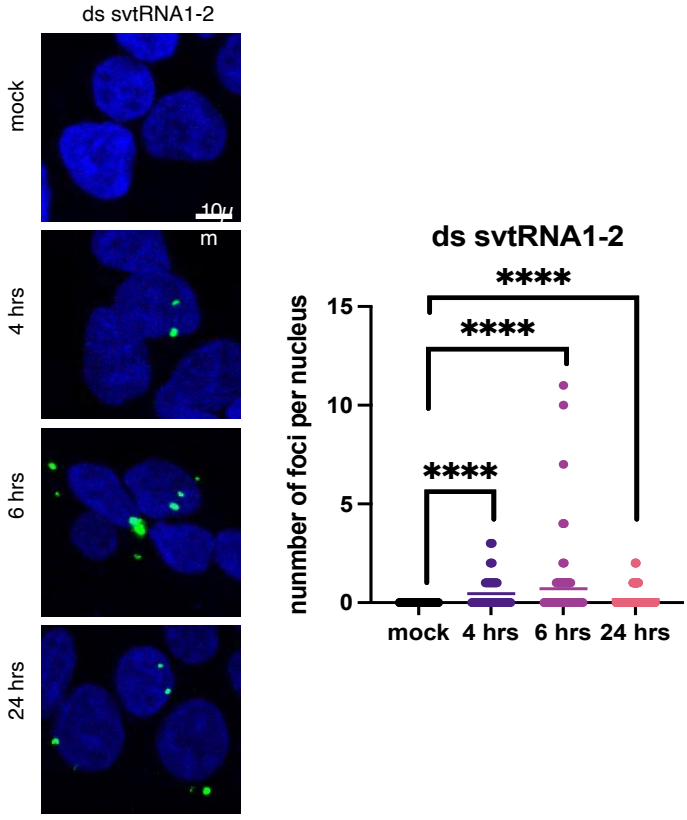
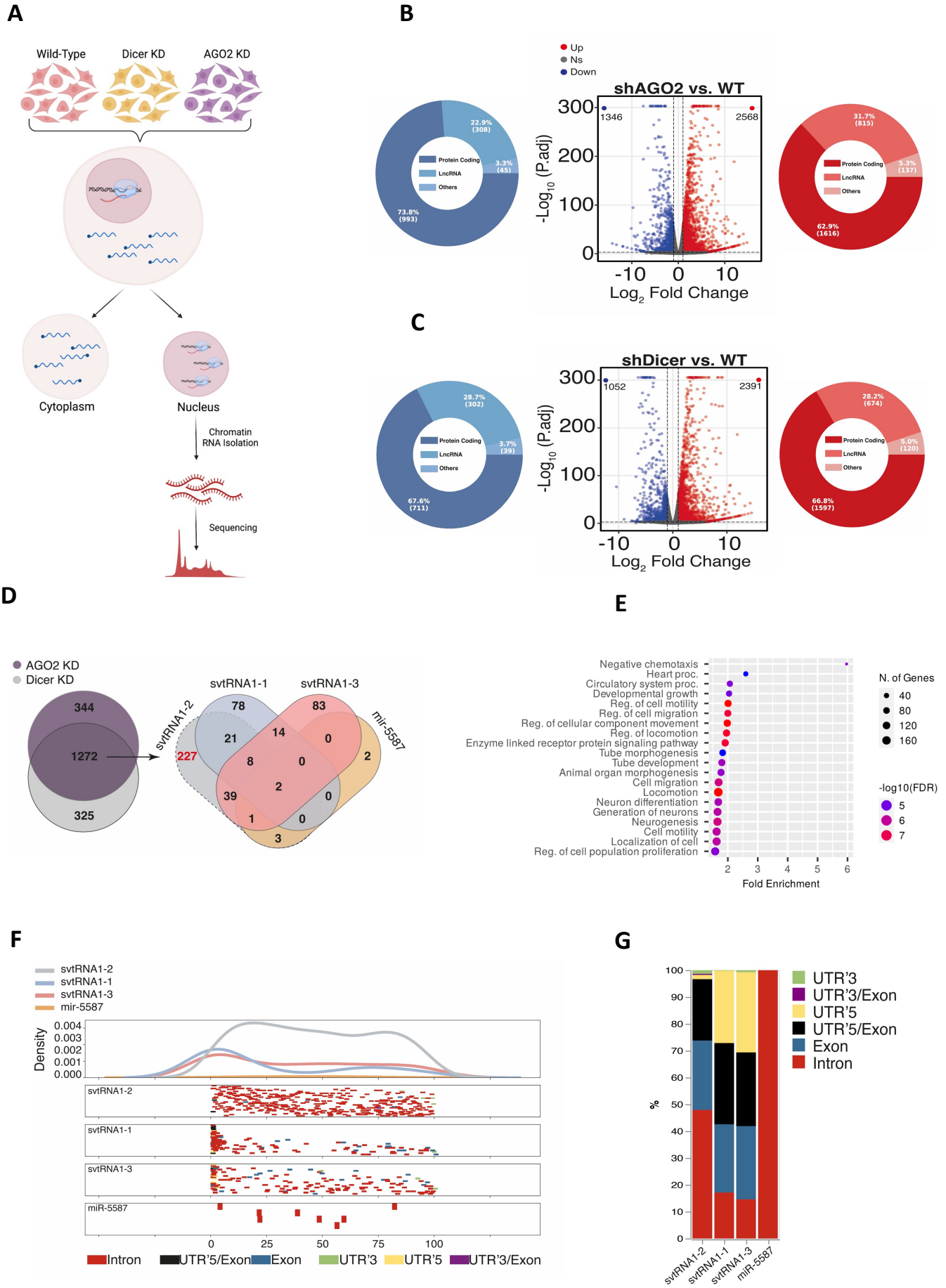


Figure 3: svtRNA1-2 targets selected genes on nascent RNA level.



Supplementary Figure 4

A

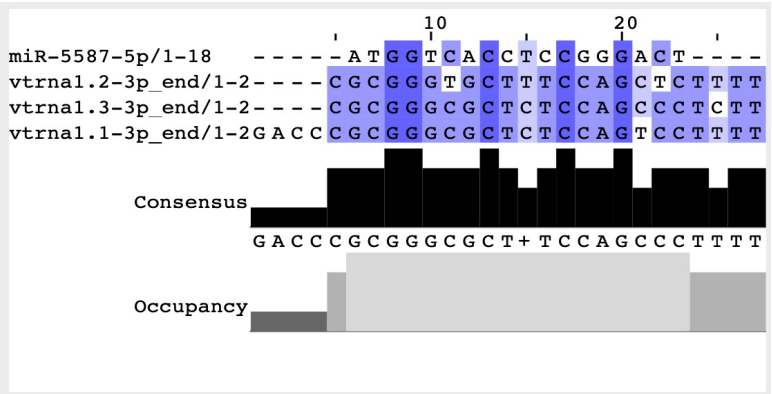
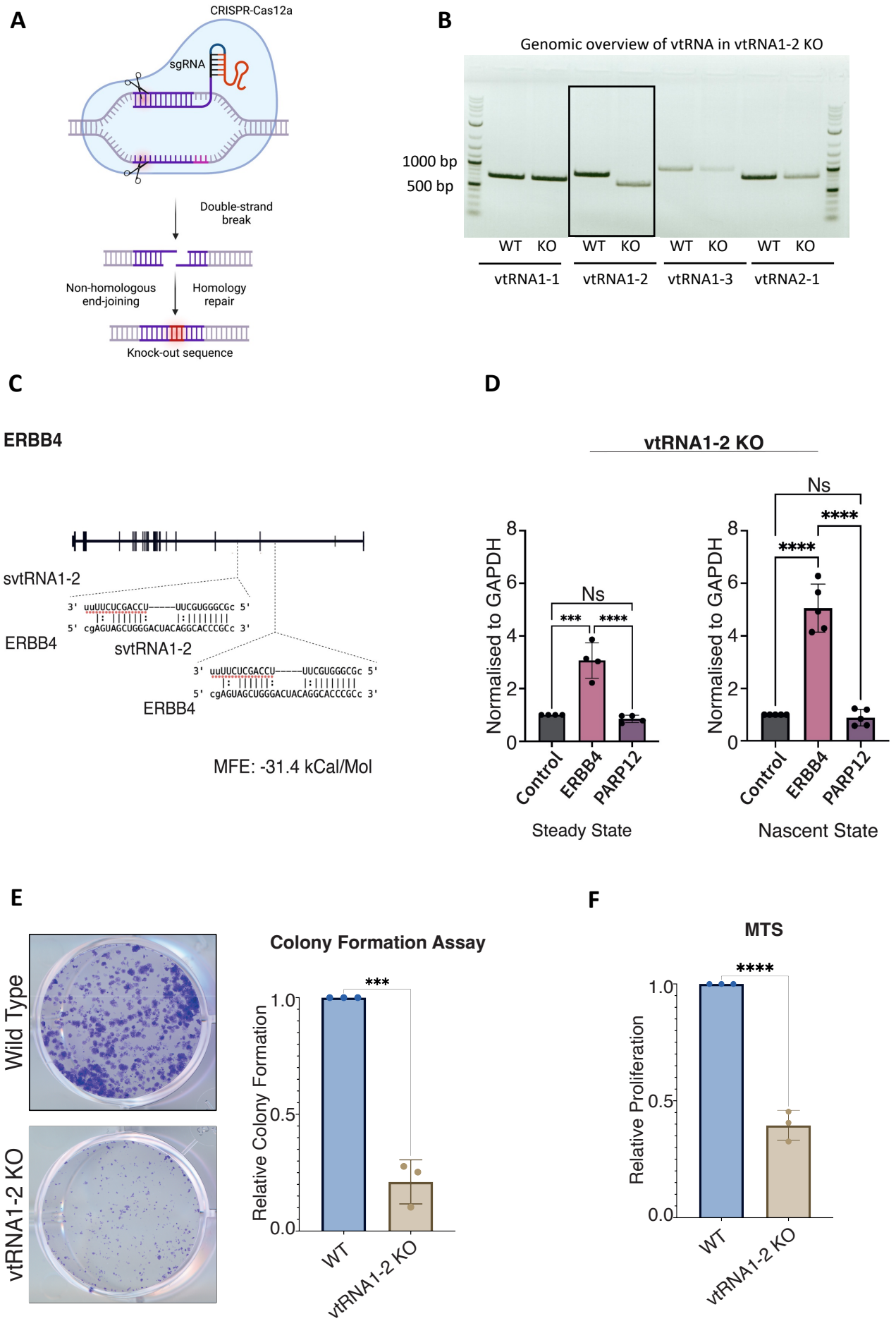
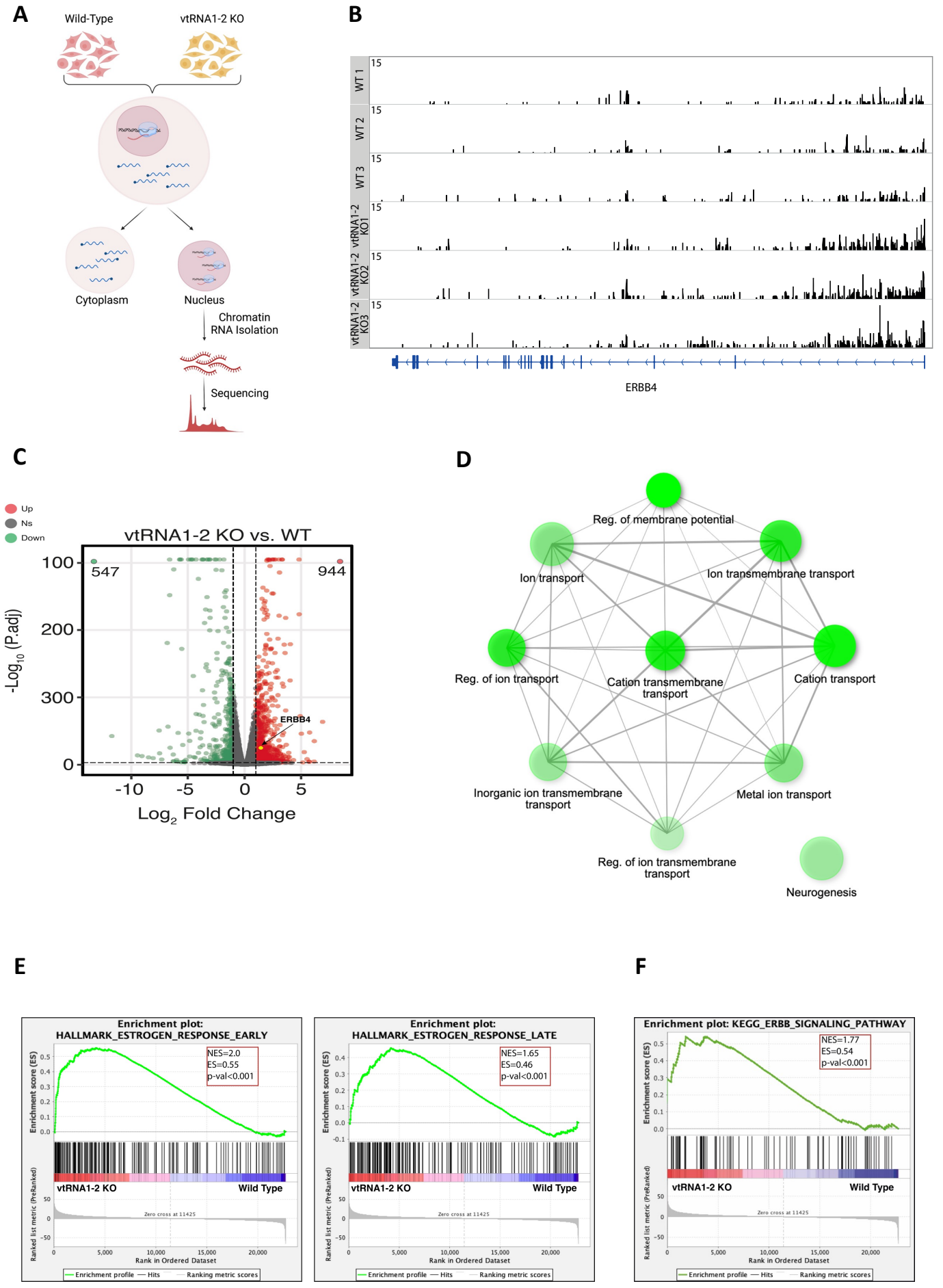


Figure 4: vtRNA1-2 knock-out leads to severe cellular phenotype.

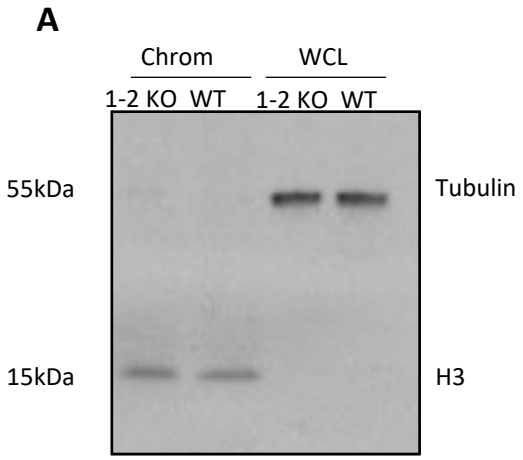


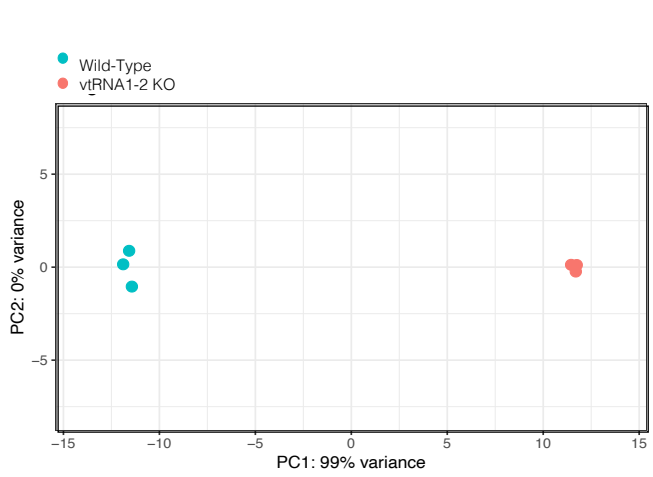
Supplementary Figure 5

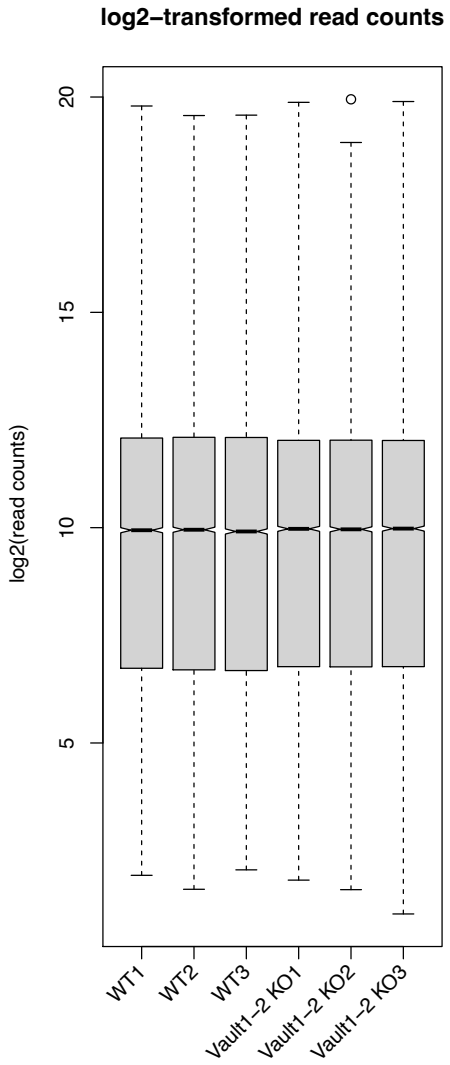
Figure 5: Deletion of vtRNA1-2 alters expression of genes associated with cellular membrane function .

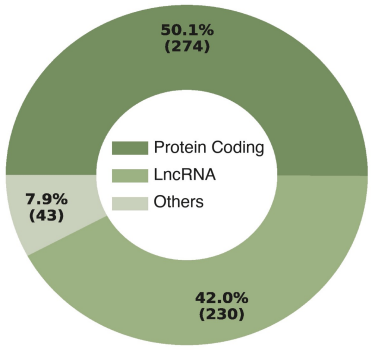


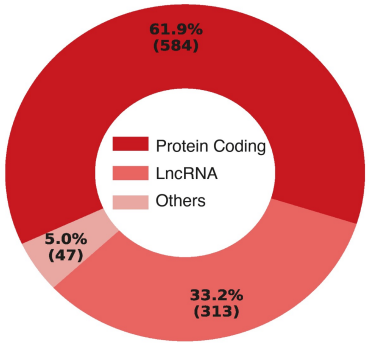
Supplementary Figure 6





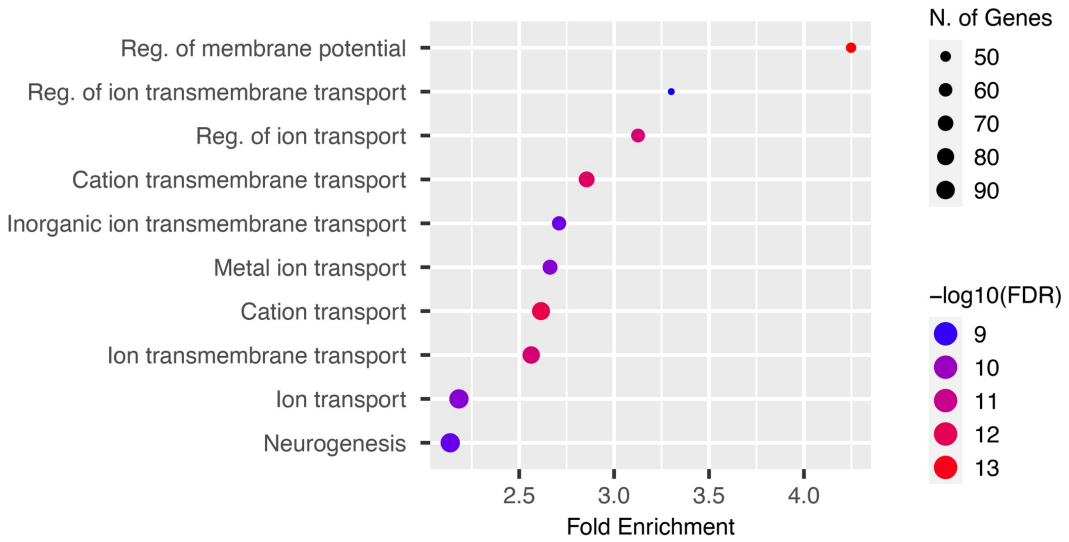






Supplementary Figure 7

A



B

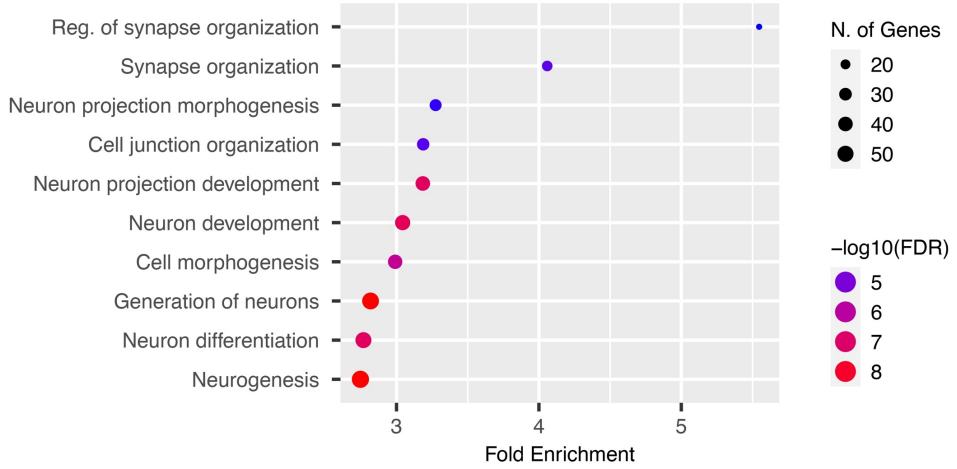
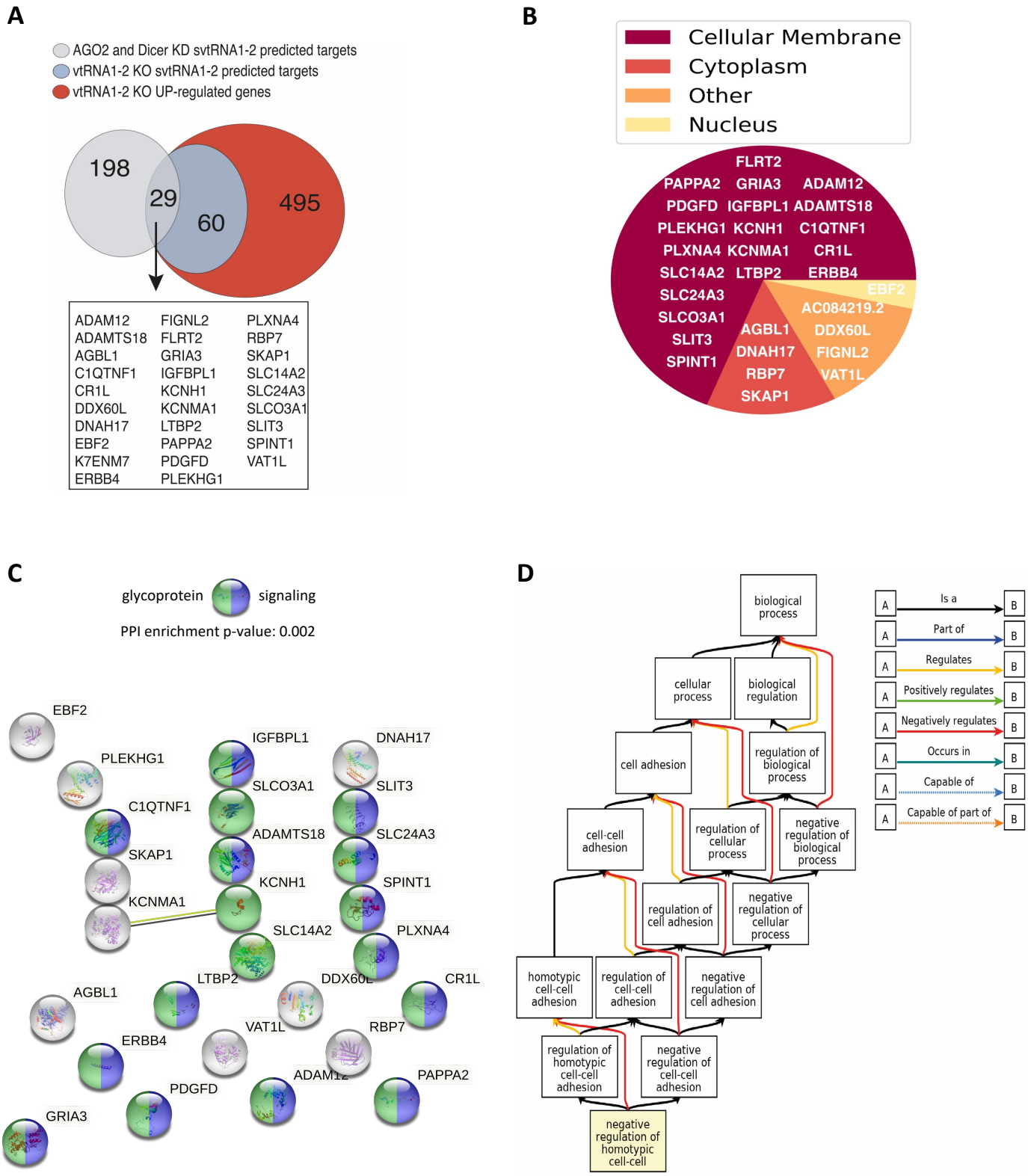


Figure 6: svtRNA1-2 target genes are associated with regulation of cell adhesion.



Supplementary Figure 8

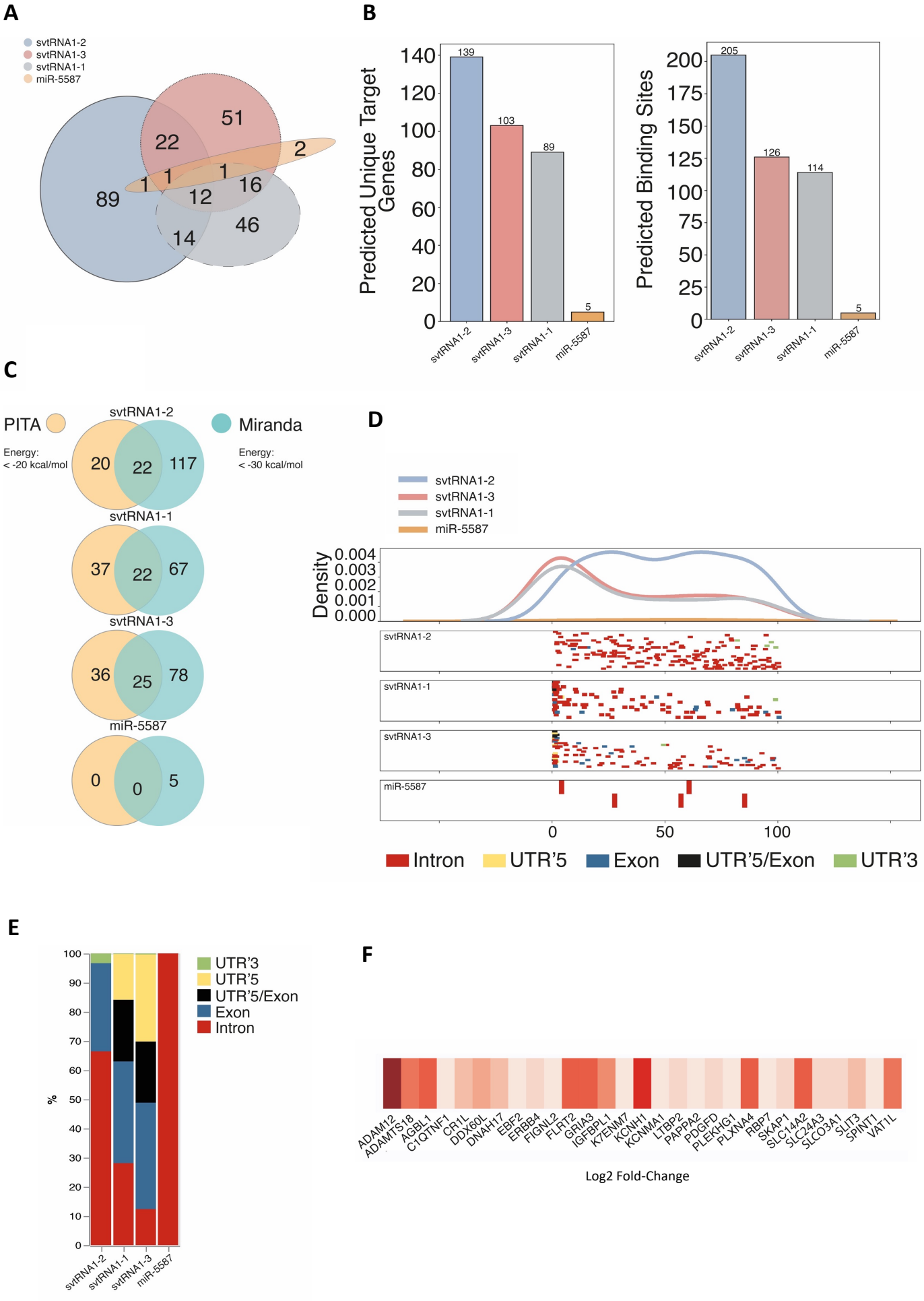


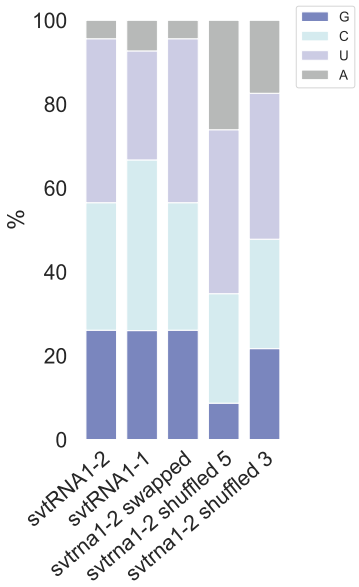
Figure 7: svtRNA1-2 targets nascent RNA with seed loop seed complementarity

A

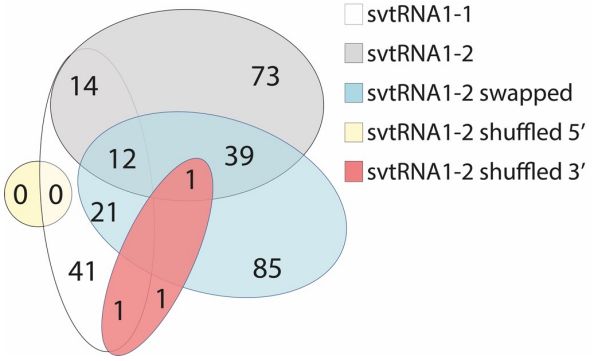
svtRNA1-1 wild-type 3' –UUUUCCUGACCUCUCGCGGGCGCCCAG-5'

svtRNA1-2 wild-type 3' –UUUUCUCGACCUUUCGUGGGCGC- 5'
svtRNA1-2 swapped 3' –UUUUCGUGGGCGUUCUCGACCUC-5'
svtRNA1-2 shuffled 5' 3' –UUUUCUCGACCUCAAAGUUUAAC-5'
svtRNA1-2 shuffled 3' 3' –UUUUCCAUCAAAUUCGUGGGCGC-5'

B



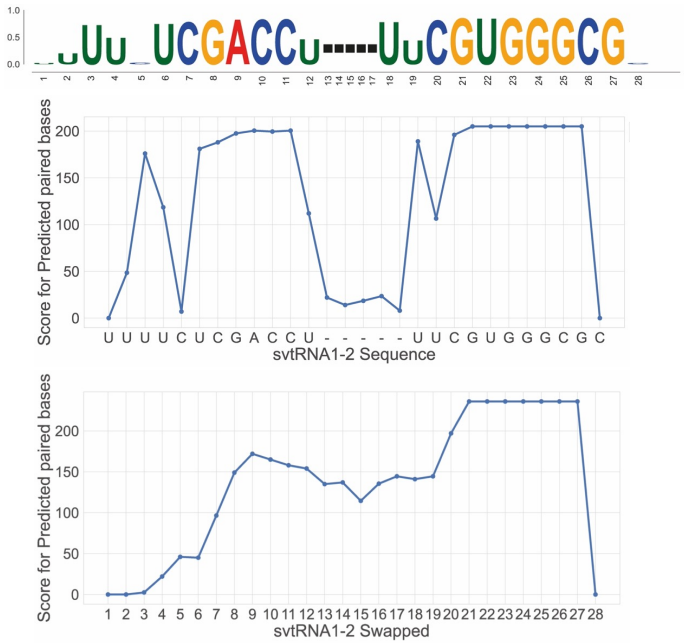
C



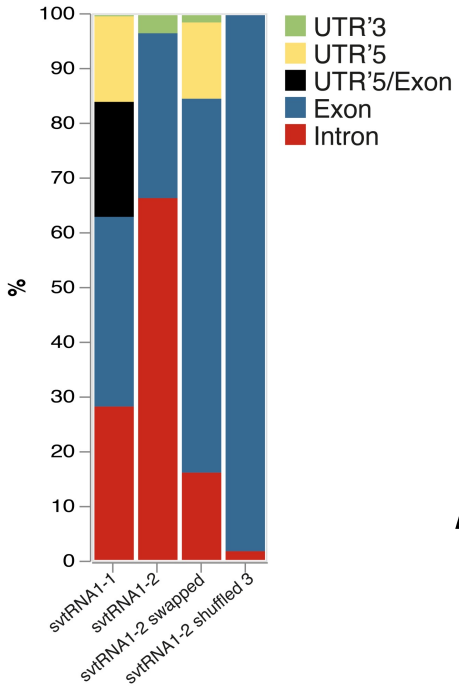
D

Sequence name	Number of Targets	Predicted Binding Sites
svtRNA1-1	89	114
svtRNA1-2	139	205
svtRNA1-2 swapped	158	236
svtRNA1-2 shuffled 5'	0	0
svtRNA1-2 shuffled 3'	3	3

F



E



G

loop

WT svtRNA1-2		Swapped svtRNA1-2	
length in nt (with gaps)	# of targets	length in nt (with gaps)	# of targets
23	10	23	**98**
24	5	24	49
25	5	25	29
26	2	26	18
27	4	27	34
28	**179**	28	8
total	205		236

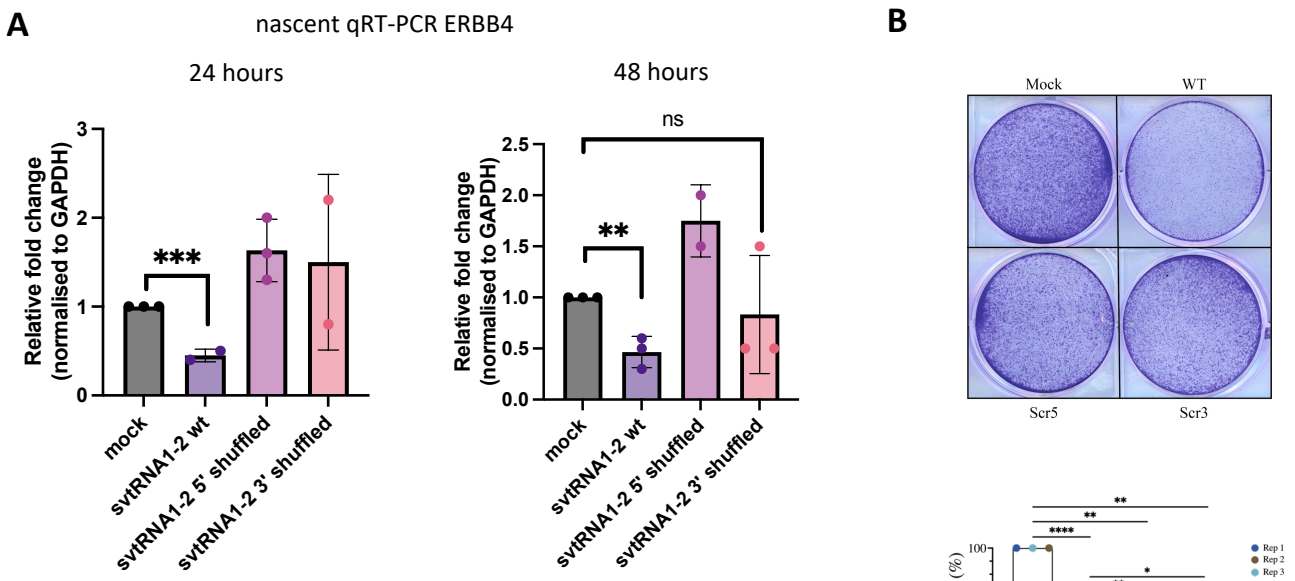
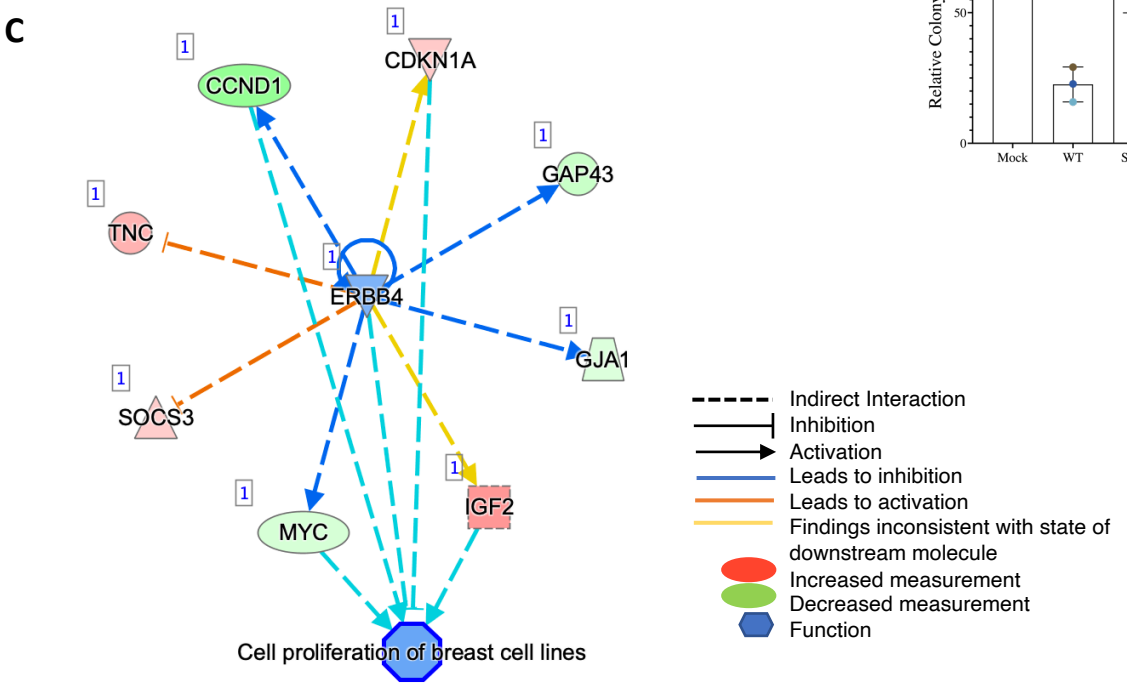
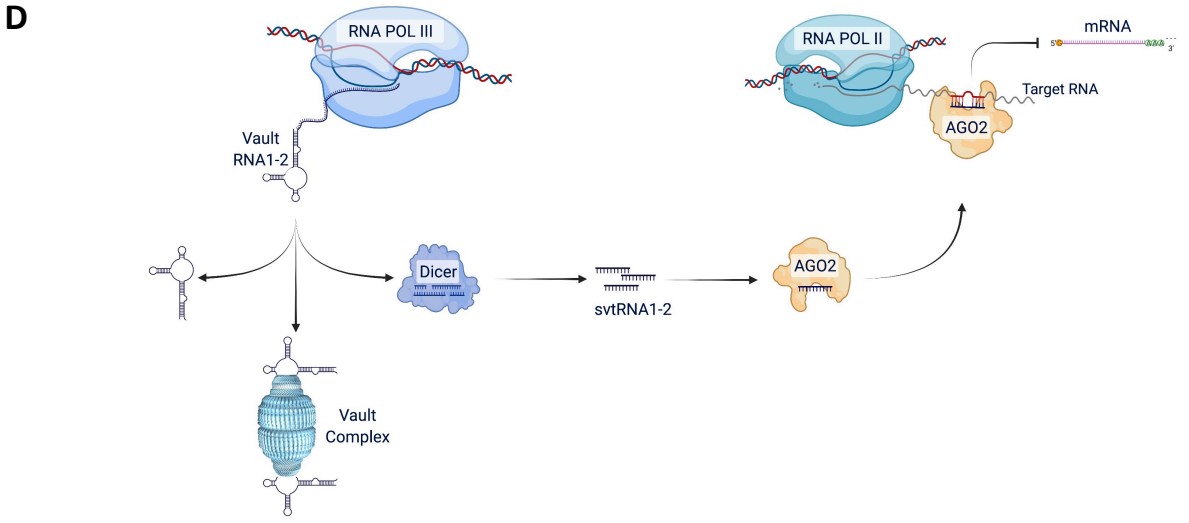



Figure 8: SvtRNA1-2 induced downregulation of ERBB4 leads to cell death in breast cancer cells.