eRNAkit: Expanding the Functional Atlas of human Enhancer RNAs Beyond the Nucleus

Natalia Benova[1*], Rene Kuklinkova[1*], Mahmoud K. Eldahshoury[1], Chinedu A. Anene[1 #]

1. Centre for Biomedical Science Research, School of Health, Leeds Beckett University, Leeds, LS1 3HE, United Kingdom
* These authors contributed equally.
# Corresponding author. Centre for Biomedical Science Research, School of Health, Leeds Beckett University, Leeds LS1 3HE, United Kingdom. E-mail: c.a.anene@leedsbeckett.ac.uk

Abstract

Enhancer RNAs (eRNAs) are a class of non-coding RNAs transcribed from active enhancers that regulate various aspects of transcription. Although traditionally viewed as nuclear-localised and unstable transcripts, large transcriptomic studies reveal they vary widely in localisation and biochemical properties, including detectable accumulation in cytoplasmic compartments. These observations suggest potential non-nuclear functions for eRNAs, yet existing databases remain focused on nuclear, cis-regulatory roles, limiting systematic exploration of their broader regulatory repertoire. Here we present **eRNAkit**, a comprehensive and accessible resource designed to address this gap by integrating subcellular localisation, RNA-RNA interaction and expression data for annotated eRNAs. Leveraging fractionation-based RNA-Seq datasets, eRNAkit profiles eRNA distribution across nuclear and cytoplasmic compartments. It incorporates gene expression data spanning major human organs and primary cell types, enabling tissue-specific analysis of eRNA function. Crucially, eRNAkit includes experimentally derived RNA-RNA interaction data from RIC-Seq, PARIS, and KARR-Seq, supporting exploration of trans-acting and cytoplasmic roles for eRNAs grounded in physical interaction evidence. eRNAkit expands current eRNA resources beyond the enhancer-promoter paradigm, offering a robust platform for dissecting non-canonical functions of eRNAs and advancing our understanding of their full regulatory potential in human biology. The eRNAkit resource is available to download at https://github.com/AneneLab/eRNAkit.

Introduction

eRNAs are a class of non-coding RNAs transcribed from active enhancer regions [1]. They regulate gene expression by activating transcription, bridging enhancer-promoter interactions, and remodelling chromatin [2]. Although initially considered nuclear, unstable and non-polyadenylated, transcriptomic data — including those from the ENCODE consortium, reveal that eRNAs vary widely in both biochemical properties and cellular localisation [3,4]. While most studies report nuclear enrichment based on relative abundance, they also detect eRNAs in cytoplasmic fractions, raising the question of what functions these cytoplasmic eRNAs might perform — and which molecular targets they may regulate outside the nucleus.

Existing eRNA databases predominantly emphasise *cis*-regulatory mechanisms, reflecting the prevailing model in which enhancers influence the expression of nearby genes through

spatial proximity. Resources such as eRNAbase, HeRA, Anima-eRNAdb, eRic and GPIeR [5–9] systematically catalogue eRNA-target relationships based on genomic co-localisation, chromatin conformation, or correlative expression patterns. While these approaches have illuminated many aspects of enhancer biology, their design inherently favours local interactions and overlook more distal, cytoplasmic effects. Furthermore, these databases largely ignore the localisation of eRNAs, an essential factor in understanding their functional potential. Nevertheless, mounting evidence suggests that certain eRNAs can function beyond their sites of origin, modulating the expression of distant genes or engaging in post-transcriptional regulation. Some studies have uncovered eRNAs that act in *trans*, regulating gene expression far from their sites of origin. For example, a distal regulatory region eRNA transcribed from the MYOD1 locus mediates cohesion recruitment and Myogenin gene expression in trans to control myogenic differentiation [10]. Similarly, an eRNA transcribed from the KLK3 locus enhance androgen receptor-dependent gene expression in human prostate cancer [11]. In both examples, the eRNAs were found to be polyadenylated, which is contrary to the standard model of eRNA processing. In a different study, knockdown of two p53-induced enhancer RNAs, p53BER2 and p53BER4, impaired the induction of distant target genes in response to p53 activation [12]. This regulation occurred despite the absence of detectable p53 binding at the promoters of these target genes, suggesting that the eRNAs act independently of local enhancer–promoter contact. Moreover, given that siRNA knockdowns used to validate most of these observations are more effective for cytoplasm localised RNAs [13], these observations raise the possibility that some of these effects is through post-transcriptional mechanisms, such as mRNA stabilisation or miRNA sponging, rather than classical nuclear transcriptional activation.

Therefore, these findings may support a broader model in which cytoplasmic functions may contribute significantly to gene regulation and function. However, a comprehensive understanding of their subcellular distribution remains incomplete, and the functional significance of eRNAs beyond the nuclear compartment, particularly in the cytoplasm, has yet to be systematically explored. A centralised and accessible resource cataloguing eRNA localisation across subcellular compartments and mapping potential eRNA-eRNA interactions is currently lacking. This gap hampers efforts to distinguish nuclear-restricted enhancer activity from putative post-transcriptional functions and limits our ability to interrogate the full functional repertoire of eRNAs in gene expression control.

To address this gap, we developed **eRNAkit**, a comprehensive and user-friendly resource for exploring eRNA subcellular localisation, expression, and cytoplasmic functions. eRNAkit integrates publicly available fractionation-based transcriptomic datasets to systematically profile eRNA distribution across nuclear and cytoplasmic compartments. It also incorporates expression data across major human organs and primary cell types, allowing users to assess the tissue and cell-type specificity of individual eRNAs. Additionally, eRNAkit includes experimentally derived RNA-RNA interaction data from technologies such as RIC-Seq, PARIS, and KARR-Seq, enabling users to explore putative eRNA-eRNA interactions grounded in physical proximity and hybridisation evidence. By combining localisation, expression, and interaction data, eRNAkit offers a powerful framework to interrogate the full regulatory potential of eRNAs beyond the classical enhancer–promoter paradigm.

Results

eRNAkit Overview

eRNAkit represents the first comprehensive resource specifically designed to investigate the cytoplasmic functions of eRNAs (Figure 1). It integrates a suite of analytical workflows for processing and analysing eRNA-related data, facilitating tasks such as data manipulation, statistical analysis, and visualisation. To assist users, eRNAkit provides a transcriptome-wide landscape of eRNA subcellular localisation, eRNA-mRNA interactions and eRNA-ribosome associations, thereby elucidating their function beyond the nucleus. Compared to existing resources, eRNAkit offers the most extensive collection of cytoplasmic eRNA-specific functional information and analytical methods (Table 1), enabling a comprehensive interrogation of eRNA functions. To facilitate exploration of the database, it includes a user-friendly Rshiny interface.

Table 1: Leading eRNA resources for comparison to eRNAkit

	Subcellular localisation	**mRNA interaction**	**Ribosome interaction**	**Target assignment**
eRNAkit	Nucleus, Cytosol, Nucleolus, Chromatin, Nucleoplasm, Membrane, Insoluble cytoplasmic fraction	KARR-Seq, RIC-Seq, PARIS	Ribo-Seq, Polysome-Seq	Physical eRNA-mRNA (TAD used to differentiate nuclear from cytoplasmic)
eRNAbase	None	RIC-Seq	None	Genomic distance (multiple thresholds)
HeRA	None	None	None	Genomic distance (1MB) and Correlation (Rho > 0.3)
Anima — eRNAdb	None	None	None	Genomic distance (1MB) and Correlation (Rho > 0.3)
eRic	None	None	None	Genomic distance (1MB)
GPIeR	None	None	None	Genomic distance (1MB)

eRNAkitApp

Extracting relevant information for a few eRNAs from the transcriptome-wide database can be tedious due to the complexity of navigating multiple interlinked tables. Thus, we build an interactive Rshiny UI called eRNAkitApp. The interface includes two core components: search panel and interactive visualisation modules. The search panel allows users to query the database by eRNA ID, genomic coordinates, or Ensemble gene ID, returning all relevant records. The visualisation modules include subcellular localisation (Supplementary Figure S1b), tissue/cell expression and specificity (Supplementary Figure S1c), eRNA-mRNA interactions (Supplementary Figure S1a) and eRNA-ribosome associations. Finally, the application offers straightforward results download option, allowing users to export search results tables. Additionally, the full database is also available for download for integration into other workflows. Collectively, this tool serves as a powerful platform for researchers to interactively explore and analyse eRNA-related data.

Data Summary

We obtained 46,650 consensus eRNAs by integrating enhancer annotations from five large databases (Methods, Figure 1a). The median length of these eRNAs was 992 bp, with an interquartile range of 238–3,241 bp, and the longest extending up to 3,198,781 bp. We found 11 extremely large regions (> 300, 000 bp), which are derived from overlapping super-enhancers in the source databases. Across tissues and primary cells, 81.19% of the eRNAs were detected at a >1 CPM threshold (Supplemental Figure 1a-b). The number of detectable eRNAs ranged from 5,209 in the liver to 13,444 in the spinal cord for tissues, and from 1,957 in astrocytes to 20,688 in bladder smooth muscle cells (SMCs) for primary cells. Notably, SMCs and fibroblast populations accounted for the majority of the top 10 cell types. All tissues and cell types had at least one specific eRNA, with the vagina showing the highest number among organs (2,383) and mononuclear cells having the most among cell types (1,202).

Unlike protein-coding genes, the functional output of eRNAs is the RNA molecule itself. As such, their regulatory potential is closely linked to their subcellular localisation and physical interactions with other RNAs. Defining these spatial and molecular contexts is essential for understanding the mechanisms by which eRNAs exert their functions. As expected, more eRNAs are detectable in the nucleus, regardless of polyadenylation status (Figure 2b). However, polyadenylation influences the number of eRNAs detectable in the cytoplasm, potentially through controlling eRNA stability and transport. Analysis of the cytoplasm-nucleus REI distribution across cell types and RNA classes reveals a trimodal pattern, with subsets of eRNAs enriched in the nucleus, in the cytoplasm, or showing no strong bias (Figure 2c). While nuclear enrichment is more common, a substantial proportion of eRNAs are cytoplasm enriched. In the two cell lines with subcytoplasmic fractionation data, we observe widespread localisation of eRNAs in both the insoluble (organelle-associated) and membrane fractions (Figure 2d).

Our curated dataset of eRNA-mRNA interactions identified 82,587 pairs present in at least two samples, encompassing 25.28% (11,793 out of 46,650) of the annotated consensus eRNAs (Figure 3a). Given the conservative nature of our filtering, this number is expected to increase with broader dataset inclusion. Surprisingly, correlation analysis between eRNA length and the number of interactions revealed a significant weak negative correlation (*rho* = −0.207, $p < 2.2 \times 10^{-16}$; $S = 3.30 \times 10$, Figure 2a). This suggests that longer eRNAs may be slightly less likely to engage in RNA-RNA interactions. However, the low correlation coefficient indicates that factors beyond transcript size-such as sequence features, localisation, or binding affinity-likely contribute to their interaction potential. Indeed, stratified analysis of these interactions across diverse cellular perturbations reveals that some are dependent on active translation or RNA-binding proteins (Figure 3b). Notably, similar perturbations-such as translation inhibition by cycloheximide or harringtonine, and viral infection by RSV or VSV-lead to consistent alterations in the RNA-RNA interaction landscape (Figure 3b). Additionally, only 28.7% (23, 703 out of 82, 587) of the mRNA were located within the same TAD or within 5kb of their corresponding eRNA pair (Methods). Interestingly, 30% of all the interactions pairs occur between eRNAs and mRNAs located on different chromosomes. These observations suggest that most eRNA-mRNA interactions do not adhere to the canonical *cis*-regulatory function of eRNAs.

Case study

Potential miRNA sponging by eRNAs

An established cytoplasmic function of several classes of non-coding RNAs is the ability to act as miRNA sponges — sequestering miRNAs and preventing them from repressing their target mRNAs. This mechanism has been widely described for circular RNAs and long non-coding RNAs [14,15]. Given this context, it is plausible that eRNAs, especially those with stable cytoplasmic presence — may also function as miRNA sponges. This possibility has not been systematically explored, but the pervasive localisation of eRNAs hints at such roles. To illustrate this, we examined the eRNA targets of miR130b and miR96, both of which have been implicated in diverse cellular processes, including metabolism and tissue development. miR-130B has been shown to regulate adipogenesis [16], while miR-96 is involved in neuronal cell differentiation [17]. Using eRNAkitApp, we first input their Ensemble IDs "ENSG00000283871, ENSG00000199158" and selected **TARGET** (Figure S1a). The results showed two distinct clusters of target eRNAs, indicating potential differences in function. Then we selected the **LOCALISATION** tab, which revealed that most of the associated eRNAs are detectable in the cytoplasm and sub-cytoplasmic fractions (Figure S1b). Given the strong cytoplasmic localisation of these eRNAs in this context, and the precedence of miRNA sponging among other cytoplasmic non-coding RNAs, we proposed that miRNA sequestration may be a functional mechanism for these eRNAs. To investigate this possibility, we examined changes in eRNA expression in response to miR-130 overexpression in mesenchymal stem cells (GSE202135) and miR-96 knockdown in prostate epithelial cells (GSE255227). Following miR-130 overexpression (by mimics), 60% (12 out of 20) of the detectable (CPM > 1 in > 2 samples) eRNA targets were downregulated (log2FC < -0.5), a proportion markedly higher than expected by chance (empirical p = 0.0, Figure 3c). By contrast, all detectable eRNA targets of miR-96 (6 out of 6) were upregulated (log2FC > 0.5) upon miR-96 knockdown (by siRNA), also surpassing random expectation (empirical p = 0.0019, Figure 3d). These findings suggest that, like other cytoplasmic ncRNAs, eRNAs may contribute to post-transcriptional regulation by modulating miRNA availability and illustrate the biological value of the eRNAkit.

Inter-organelle communication through eRNAs

Bi-directional communication between the nucleus and mitochondria is critical for maintaining cellular function [18]. Emerging research suggests that RNA-mediated nuclear signal can influence mitochondrial biology [19]. We identified an eRNA, "en135763" located on chromosome 5, which is altered under various cellular stress conditions (data not shown), and searched eRNAkitApp using the eRNA ID option (Figure S3a). In the **LOCALISATION** tab, results showed that this eRNA exhibits comparable expression levels in the cytoplasm and nucleus. It is also enriched in the organelle-associated fraction in K562 cells and has eRNA-RNA interactions with mitochondrial encoded genes (Figure S3b), indicating that eRNAkit could support research into RNA mediated inter-organelle communication. The **EXPRESSION** tab revealed that this eRNA is widely expressed across organs and primary cell types, with notably higher levels in T-cell populations, consistent with the involvement of immune cells in the stress response. These findings highlight the underappreciated value of eRNA subcellular localisation and physical interactions.

Discussion

eRNAs have long been considered nuclear-localised and unstable byproducts of enhancer activation, with functions largely confined to cis-regulatory roles at nearby promoters. However, emerging data challenge this restricted view, revealing that eRNAs are more biochemically and functionally diverse than previously appreciated. Unlike protein-coding genes, the final product of eRNAs is RNA itself. Consequently, eRNA function relies on the physical properties and interactions of the RNA molecule, which are inherently dependent on spatial proximity. Here, we present eRNAkit, a comprehensive resource that consolidates subcellular localisation, expression specificity, and RNA-RNA interaction data to systematically map the full regulatory repertoire of eRNAs, including potential trans-acting and post-transcriptional functions. Compared to existing databases, eRNAkit offers several advances. Most notably, it is the first to integrate larger scale subcellular localisation data alongside physical RNA-RNA interactions, thereby facilitating functional hypotheses that extend beyond proximity-based enhancer-promoter models.

A key component of eRNAkit is the transcriptome-wide landscape of eRNA localisation across subcellular compartments. Our REI-based analyses demonstrate that while nuclear enrichment is prevalent, a considerable proportion of eRNAs are detectable — and in some cases, enriched in cytoplasmic and subcytoplasmic fractions. This observation supports prior reports suggesting the presence of cytoplasmic eRNAs [3] and offers a transcriptome-wide perspective that challenges the long-standing assumption of nuclear restriction. Notably, the influence of polyadenylation on cytoplasmic localisation suggests that RNA processing features may be critical determinants of eRNA stability and trafficking, potentially mirroring the behaviour of other non-coding RNAs with post-transcriptional roles, such as lncRNAs and circRNAs. eRNAkit also highlights strong cell- and tissue-specific expression patterns of eRNAs, consistent with the known context-dependence of enhancer activity [1,20]. The presence of highly specific eRNAs in reproductive organs and in immune cells highlight their potential roles in tissue-specialised gene regulatory programs. Such specificity may be leveraged in future studies to prioritise candidate eRNAs for functional validation or biomarker development.

The inclusion of tens of thousands of robust eRNA-mRNA interactions further expands the functional landscape of eRNAs. These interactions, derived from multiple RNA-RNA interactome technologies (RIC-Seq, PARIS, KARR-Seq), provide direct evidence of physical proximity between eRNAs and potential target transcripts. Importantly, a substantial fraction of these interactions is translation or RBP-dependent, suggesting that eRNAs may engage in post-transcriptional regulation through mechanisms such as mRNA stabilisation, translational control, or miRNA competition. The weak negative correlation between eRNA length and interaction count also points to a nuanced interaction landscape that may be shaped more by structure, binding motifs, or localisation than by transcript size alone.

Despite its utility, several limitations should be noted. First, while the use of multiple datasets enhances confidence in reported interactions and expression patterns, coverage across all cell types and compartments remains incomplete. Second, although physical interactions provide strong evidence for potential function, mechanistic validation is essential to determine whether these interactions are regulatory, structural, or stochastic. Lastly, the dynamic regulation of eRNA expression and localisation — particularly under stress, development, or disease contexts — remains to be fully characterised and could

257 provide further insight into the functional adaptability of eRNAs. To address these
258 limitations, we plan to update eRNAkit regularly by incorporating additional subcellular
259 localisation data and functional eRNA interaction profiles as more high-throughput datasets
260 become available. Due to our work in cellular senescence, we are also building an extension
261 to characterise the expression profiles of these eRNAs in cellular stress response. We are
262 confident that eRNAkit provides a critical first step toward uncovering the full regulatory
263 potential of eRNAs and will serve as a foundational tool for advancing our understanding of
264 their diverse roles in human health and disease.

265 **Methods**

266 **eRNAkit Design**

267 eRNAkit comprises three main layers: analytics, database and Rshiny application (Figure 1).
268 The analytics layer includes a suite of functions designed to characterise the cytoplasmic
269 functions of eRNAs. Key routines encompass rna2rna for identifying eRNA-mRNA
270 interactions from STAR aligner chimeric junctions, entropyTS for assessing tissue or cell
271 specificity and getREI for calculating the relative expression index. The database layer,
272 referred to as emi, contains four structured tables that store consensus eRNA annotations,
273 transcriptome-wide eRNA-mRNA interactions, subcellular localisation data of eRNA classes
274 and their association with ribosomes, as well as expression levels of eRNAs across major
275 human organs and primary cell types, including tissue and cell-type specificity metrics.
276 Additionally, the database integrates genome-wide enhancer-target gene information,
277 facilitating the differentiation between nuclear and cytoplasmic functions of eRNAs. The R
278 Shiny application layer provides an interactive interface for users to explore and visualise
279 the database, enabling dynamic querying and presentation of results.
280
281 **Consensus human eRNA annotation**
282 To build consensus eRNAs annotation, we collected previously published human enhancer
283 elements across multiple tissues and cell types from EnhancerAtlas 2.0 [21], Ensemble
284 regulatory build [22], eRNAbase [5], neuronal transcribed enhancers [23], and cancer associated
285 eRNAs [24]. We manually downloaded 2, 492, 442 coordinates in bed format and converted
286 those in hg19 genome build to hg38 coordinates using liftOver tool [25]. Then, overlapping
287 elements across the five datasets were merged to generate a consensus file using BEDTools
288 [26]. To ensure robust genomic coordinates, we validated all annotations against the hg38
289 chromosome sizes, excluding any entries with start or end positions outside of chromosome
290 bounds. To ensure that eRNA quantification is not confounded by normal gene expression,
291 we removed elements overlapping with the Ensemble gene annotation v112. We extended
292 the gene regions by 500 bp at both sides for this analysis to remove possible unknown
293 variants. Finally, we removed elements with a size < 10 bp, which are within the range of
294 largely unstable nuclear eRNAs.
295
296 **eRNA subcellular localisation**
297 Subcellular RNA-Seq data representing two RNA classes (polyA− and polyA+) across seven
298 cellular fractions (nucleus, cytosol, nucleolus, chromatin, nucleoplasm, membrane, insoluble
299 cytoplasmic fraction) were obtained from ENCODE [3]. BAM files aligned to the hg38 human
300 genome assembly for 13 cell lines (SK-N-SH, K562, A549, H1, HeLa-S3, MCF-7, HepG2, IMR-

90, GM12878, SK-N-DZ, SK-MEL-5, NCI-H460, and HT1080) were downloaded from the ENCODE Data portal (Supplemental Table S2). Read counts were quantified using HTSeq (v0.11.1)[27] on the consensus eRNA GTF file. Each RNA class and subcellular fraction had biological replicates, and read counts were summarised by the median count per million (CPM) across replicates.

To assess relative compartment expression of eRNAs across RNA classes and cell types, we calculated the relative expression index (REI) across predefined pairs of subcellular compartments for each RNA class (polyA−, polyA+ or total RNA) and cell line. REI was defined as:

$$REI = log_2\left(\frac{C_1}{C_2}\right) \qquad Eq.1$$

where C_1 and C_2 represent the median of a given eRNA in two compartments. A higher REI value indicates greater expression in C_1 relative to C_2.

Compartment pairs were defined such that finer cytoplasmic subfractions (e.g., insoluble cytoplasmic fraction, membrane) were compared to the cytosol, and nuclear subfractions (e.g., nucleoplasm, nucleolus, chromatin) were compared to the nucleus. The complete set of compartment comparisons were cytosol-nucleus, insoluble cytoplasmic fraction-cytosol, membrane-cytosol, nucleolus-nucleus, nucleoplasm-nucleus, chromatin-nucleus. For each RNA class and cell line, REIs were computed only when expression values were available for at least one compartments being compared. To investigate eRNA processing, a second REI was calculated between polyA− and polyA+ samples within each compartment and cell using the same filtering strategy.

Calculation of tissue and cell specificity for eRNAs

To assess tissue and cell specificity, we downloaded and processed total RNA-Seq BAM files (hg38) for 32 major tissues and 77 primary cell types from the ENCODE data portal. Biological replicates were summarised by the median CPM across replicates. eRNAs that were lowly expressed (<1 CPM) in all tissues were removed. We call an eRNA specific (tissue or cell) when its maximum median CPM was more than double the second largest median CPM and the difference between the logarithm of the total number of tissues/cells and the Shannon entropy of the median CPMs is greater that 1 [28].

eRNA-mRNA interaction

The eRNA-mRNA interactions in eRNAkit were identified using three high-throughput approaches, including KARR-Seq [29], RIC-Seq [30] and PARIS [31] (Table 2). These methods differ in their underlying chemistries, interaction specificity, and compartmental bias, together providing a more comprehensive and nuanced map of eRNA-mRNA contacts. By integrating datasets with differing sensitivities to nuclear versus cytoplasmic interactions, eRNAkit enables exploration of potential post-transcriptional regulatory roles for eRNAs that might be overlooked by any single approach.

Table 2: Comparison of RNA–RNA Interaction Mapping Methods

Feature	KARR-Seq	RIC-Seq	PARIS
Chemistry	Kethoxal-modified guanine bases (selective to ssRNA)	Formaldehyde (reversible; targets protein and RNA interactions)	Psoralen (intercalates in double-stranded RNA)
Interaction	Base-paired duplexes, especially transient or regulatory pairs	Both direct and indirect (via RNA-binding proteins) interactions	Base-paired duplexes with high confidence
Specificity	Prefers G-rich and dynamic regions	Protein-mediated proximity bias	Enriches for stable RNA duplexes
Compartment Bias (Whole Cell Lysate)	**Cytoplasmic-enriched** — favours mature, exported RNAs	**Balanced** — captures both nuclear and cytoplasmic interactions	**Nuclear-biased** — stronger detection of pre-mRNA and nuclear lncRNA
Advantages	Sensitive to dynamic, potentially regulatory interactions	Captures a broad spectrum of RNA interactions	Provides structural information; precise mapping of duplexes
Limitations	May miss highly structured or non-G-rich regions	Less resolution for direct base-pairing	Bias toward stable, long helices

We manually downloaded 78 raw sequencing datasets from 9 cells lines via the NCBI GEO database (Supplemental Table S1). To identify eRNA-mRNA interactions, reads were quality filtered (Q < 20) and adapter trimmed using Trimmomatic (v0.39) [32]. Cleaned reads were aligned to the human reference genome (hg38) using STAR [33] with chimeric alignment options enabled (--runMode alignReads --chimOutType Junctions --chimJunctionOverhangMin 8 --chimSegmentMin 15 --outFilterMultimapNmax 100 --alignIntronMin 1 --scoreGapNoncan -4 --scoreGapATAC -4 --alignSJoverhangMin 15 --alignSJDBoverhangMin 10 --alignSJstitchMismatchNmax 5 -1 5 5).

Chimeric junctions were then processed using a custom R-based workflow implemented in eRNAkit (rna2rna). Briefly, chimeric alignments were parsed into a pseudo-BEDPE format by splitting each read into two genomic intervals (n1 and n2) and extracting positions and strand information using the CIGAR string. These were intersected with Ensemble v111 gene annotations and the consensus eRNA annotations. Only junctions connecting an eRNA to a known gene were retained as putative eRNA-mRNA interactions. To enhance robustness, only eRNA-mRNA pairs detected in at least two independent samples were included in the final interaction set. This conservative filtering strategy prioritises reproducibility across biological contexts and offers a high-confidence, global view of eRNA interactions.

Some of the samples analysed by KARR-Seq included cellular perturbations such as translation inhibition, RNA-binding protein knockdowns and viral infections (Supplemental Table S1). These experimental contexts provided an opportunity to explore potential

mechanisms driving eRNA-mRNA interactions. To assess perturbation responsiveness, we focused on binary changes-specifically, the complete gain or loss of individual interactions in treated versus untreated conditions. Interactions were considered perturbation-sensitive if they were entirely absent in control samples but appeared in treated samples, or vice versa. This stringent approach highlights interactions potentially under regulatory control in response to environmental or cellular cues.

eRNA genomic targets

To identify potential genomic targets to facilitate differentiation of cytoplasm exclusive interactions, we used two strategies (1) distal targets, based on co-localisation within the same topologically associating domains (TADs), and (2) proximal targets, based on linear distance to the eRNA loci. For the distal approach, we obtained human TAD annotations across multiple cell types and tissues from the ENCODE database [34]. Using BEDTools (v2.31.1) [26] and human gene annotation (hg38, Ensemble v111), we identified genes residing in the same TAD as each eRNA locus. For the proximal approach, genes located within ±5 kb of an eRNA locus were annotated as their proximal targets. Note that both approaches mirror what is used by existing tools to call eRNA targets (Table 1).

RNA-Seq and Ribo-Seq analysis

Standard RNA-Seq and Ribo-Seq (Supplemental Table S3) analysis was performed using our existing pipeline [35]. Briefly, the raw reads were quality filtered (Q < 20) and adapter trimmed using Trimmomatic (v0.39) [32]. Cleaned reads were aligned to the human reference genome (hg38) using HISAT2 (v2.1.0) in default settings [36]. Read counts were quantified using HTSeq (v0.11.1)[27] on the consensus eRNA annotation. The expression levels were normalised by CPM.

Permutation test

To determine whether the number of differentially expressed targets associated with a given miRNA was significantly greater than expected by chance, we performed a non-parametric permutation test. The observed count was defined as the number of eRNA targets in a specific threshold (Log_2 fold change (Log2FC) > 1 or < -1) among the set of detected eRNA targets of miR-130b or miR-96. To generate the null distribution, we randomly sampled the same number of eRNA (x10,000) from all the detected eRNA and counted the number matching the threshold. The empirical p-value was computed as the proportion of permutations in which the number of random eRNAs in the specified threshold was greater than or equal to the observed count. This one-tailed test estimates the probability of obtaining the observed level of enrichment under the null hypothesis of random association.

Data availability

The research community can access eRNAkit freely from the GitHub repository https://github.com/AneneLab/eRNAkit.

Author contributions

Project was conceived of by C.A.A. Data curation and processing were done by N.B, R.K, M.E, and C.A.A. Software and database creation was done by C.A.A, N.B and R.K. C.A.A supervised the project and wrote the manuscript. All authors edited the final manuscript.

407
408 **Disclosure and declaration**
409 We declare no conflict of interest.
410
411 **Figure Legends**
412 **Figure 1: Overview of the main components of eRNAkit.**
413 eRNAkit consists of three layers: **Analytics**, **Database**, and **eRNAkitApp**. The **Analytics** layer
414 provides workflows for processing eRNA-related data and supports the design of validation
415 experiments. The **Database** contains data on eRNA expression across major tissues and
416 primary cells, subcellular localisation, and physical eRNA–mRNA interactions. The
417 **eRNAkitApp** offers a user-friendly interface for querying the database and includes modules
418 for exploring eRNA function.
419
420 **Figure 2: Summary statistics from the database layer.**
421 **(a)** Distribution of consensus eRNAs across chromosomes.
422 **(b)** Number of detected eRNAs (CPM > 1) in both the nucleus and cytoplasm across multiple
423 cell types. polyA+ indicates selection of polyadenylated RNAs prior to sequencing, while
424 polyA− indicates depletion of polyadenylated RNAs.
425 **(c)** Distribution of the relative expression index (REI) between cytoplasm (C1) and nucleus
426 (C2) across RNA classes (see Methods). Positive values indicate enrichment in the
427 cytoplasm, and negative values indicate enrichment in the nucleus.
428 **(d)** Number of detected eRNAs (CPM > 1) in cytoplasmic sub-fractions across RNA classes in
429 HepG2 and K562 cells.
430
431 **Figure 3: eRNA–mRNA interactions reveal potential cytoplasmic functions.**
432 **(a)** Distribution of eRNA–mRNA interactions across mRNA biotypes.
433 **(b)** Similarity of eRNA–mRNA interactions (gain or loss) under different cellular
434 perturbations. SRSF1 and DDX3X represent knockdowns of RNA-binding proteins.
435 Harringtonine and cycloheximide indicate translational inhibition via drug treatment. RSV
436 and VSV represent infection with respiratory syncytial virus or vesicular stomatitis virus,
437 respectively. Arsenite denotes induction of cellular stress.
438 **(c–d)** Scatter plots showing the median expression of eRNAs in control cells versus the $\log_2$
439 fold change in expression upon perturbation with miR-130 **(c)** or miR-96 **(d)**.
440
441 **Figure S1: Main interface and usage of eRNAkitApp.**
442 **(a)** Navigation bar with the TARGET module activated.
443 **(b)** View of the LOCALISATION module.
444 **(c)** View of the EXPRESSION module.
445 **Note:** Graphs shown in S1 relate to searching for "ENSG00000283871, ENSG00000199158"
446 (MIR130 and MIR96).
447
448 **Figure S2:** eRNAs are widely expressed across major tissues and primary cells.
449 **(a–b)** Number of detected eRNAs (CPM > 1) across tissues **(a)** and primary cells **(b)**.
450
451 **Figure S3: Searching eRNAkitApp using eRNA ID.**
452 **(a)** Navigation bar with the LOCALISATION module activated.
453 **(b)** View of the TARGET module.

454 **(c)** View of the EXPRESSION module.
455 **Note:** Graphs shown in S3 relate to searching for "en135763".
456
457

458 **References**

459 1. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues.
460 *Nature* **507**, 455–461 (2014).
461 2. Li, Q. *et al.* Enhancer RNAs: mechanisms in transcriptional regulation and functions in
462 diseases. *Cell Communication and Signaling* **21**, 191 (2023).
463 3. Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
464 4. Sartorelli, V. & Lauberth, S. M. Enhancer RNAs are an important regulatory layer of the
465 epigenome. *Nature structural & molecular biology* **27**, 521–528 (2020).
466 5. Song, C. *et al.* eRNAbase: a comprehensive database for decoding the regulatory eRNAs in
467 human and mouse. *Nucleic Acids Research* **52**, D81–D91 (2024).
468 6. Zhang, Z. *et al.* HeRA: an atlas of enhancer RNAs across human tissues. *Nucleic acids*
469 *research* **49**, D932–D938 (2021).
470 7. Jin, W. *et al.* Animal-eRNAdb: a comprehensive animal enhancer RNA database. *Nucleic*
471 *acids research* **50**, D46–D53 (2022).
472 8. Zhang, Z. *et al.* Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-
473 targeted therapy in cancer. *Nature communications* **10**, 4562 (2019).
474 9. Zhang, Z. *et al.* Genetic, pharmacogenomic, and immune landscapes of enhancer RNAs
475 across human cancers. *Cancer Research* **82**, 785–790 (2022).
476 10. Tsai, P.-F. *et al.* A muscle-specific enhancer RNA mediates cohesin recruitment and
477 regulates transcription in trans. *Molecular cell* **71**, 129–141 (2018).

478 11. Hsieh, C.-L. *et al.* Enhancer RNAs participate in androgen receptor-driven looping
479 that selectively enhances gene activation. *Proceedings of the National Academy of*
480 *Sciences* **111**, 7319–7324 (2014).
481 12. Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene
482 transcription. *Molecular cell* **49**, 524–535 (2013).
483 13. Lennox, K. A. & Behlke, M. A. Cellular localization of long non-coding RNAs affects
484 silencing by RNAi more than by antisense oligonucleotides. *Nucleic acids research* **44**,
485 863–877 (2016).
486 14. Singh, S., Sinha, T. & Panda, A. C. Regulation of microRNA by circular RNA. *Wiley*
487 *Interdisciplinary Reviews: RNA* **15**, e1820 (2024).
488 15. Paci, P., Colombo, T. & Farina, L. Computational analysis identifies a sponge
489 interaction network between long non-coding RNAs and messenger RNAs in human
490 breast cancer. *BMC systems biology* **8**, 1–15 (2014).
491 16. Wang, Y. *et al.* Circulating miR-130b mediates metabolic crosstalk between fat and
492 muscle in overweight/obesity. *Diabetologia* **56**, 2275–2285 (2013).
493 17. Kuhn, S. *et al.* miR-96 regulates the progression of differentiation in mammalian
494 cochlear inner and outer hair cells. *Proceedings of the National Academy of Sciences* **108**,
495 2355–2360 (2011).
496 18. Soledad, R. B., Charles, S. & Samarjit, D. The secret messages between mitochondria
497 and nucleus in muscle cell biology. *Archives of biochemistry and biophysics* **666**, 52–62
498 (2019).
499 19. Vendramin, R., Marine, J. & Leucci, E. Non-coding RNA s: the dark side of nuclear–
500 mitochondrial communication. *The EMBO journal* **36**, 1123–1133 (2017).

501 20. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal
502 signatures of enhancer activity during embryonic development. *Nature genetics* **44**, 148–
503 156 (2012).
504 21. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation
505 in 586 tissue/cell types across nine species. *Nucleic acids research* **48**, D58–D64 (2020).
506 22. Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T. & Flicek, P. R. The ensembl
507 regulatory build. *Genome biology* **16**, 1–8 (2015).
508 23. Dong, P. *et al.* Population-level variation in enhancer expression identifies disease
509 mechanisms in the human brain. *Nature genetics* **54**, 1493–1503 (2022).
510 24. Chen, H. & Liang, H. A high-resolution map of human enhancer RNA loci
511 characterizes super-enhancer activities in cancer. *Cancer cell* **38**, 701–715 (2020).
512 25. Hinrichs, A. S. *et al.* The UCSC genome browser database: update 2006. *Nucleic acids*
513 *research* **34**, D590–D598 (2006).
514 26. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
515 genomic features. *Bioinformatics* **26**, 841–842 (2010).
516 27. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
517 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
518 28. Li, S. *et al.* exoRBase: a database of circRNA, lncRNA and mRNA in human blood
519 exosomes. *Nucleic acids research* **46**, D106–D112 (2017).
520 29. Wu, T. *et al.* KARR-seq reveals cellular higher-order RNA structures and RNA–RNA
521 interactions. *Nature Biotechnology* **42**, 1909–1920 (2024).
522 30. Cai, Z. *et al.* RIC-seq for global in situ profiling of RNA–RNA spatial interactions.
523 *Nature* **582**, 432–437 (2020).

524 31. Lu, Z. *et al.* RNA duplex map in living cells reveals higher-order transcriptome
525 structure. *Cell* **165**, 1267–1279 (2016).
526 32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
527 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
528 33. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
529 (2013).
530 34. Hammal, F., De Langen, P., Bergon, A., Lopez, F. & Ballester, B. ReMap 2022: a
531 database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an
532 integrative analysis of DNA-binding sequencing experiments. *Nucleic acids research* **50**,
533 D316–D325 (2022).
534 35. Cogan, J. A., Benova, N., Kuklinkova, R., Boyne, J. R. & Anene, C. A. Meta-analysis of
535 RNA interaction profiles of RNA-binding protein using the RBPInper tool. *Bioinformatics*
536 *Advances* **4**, vbae127 (2024).
537 36. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory
538 requirements. *Nature methods* **12**, 357 (2015).

539

540
541
542

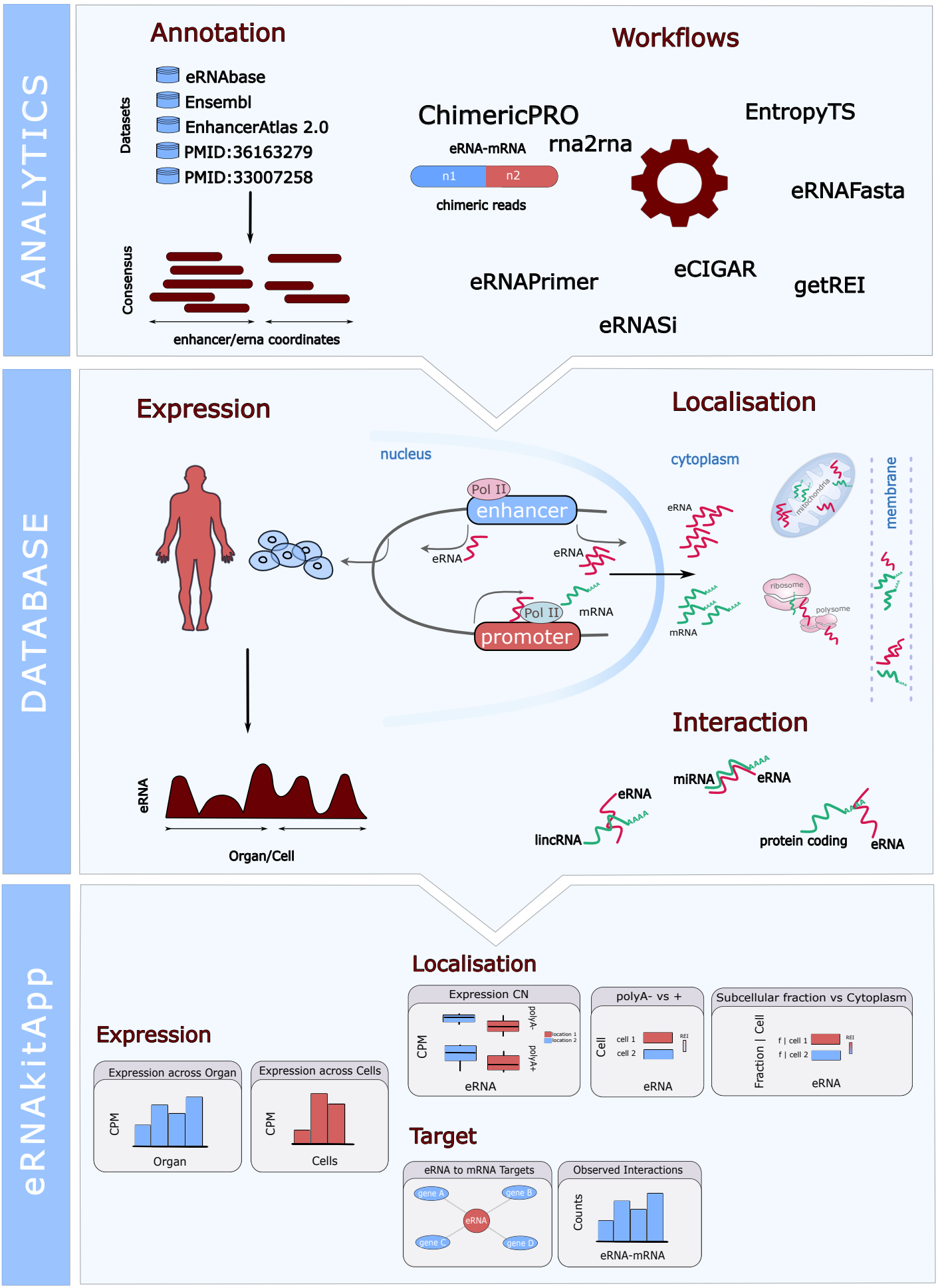

ANALYTICS
Annotation
Datasets
eRNAbase
Ensembl
EnhancerAtlas 2.0
PMID:36163279
PMID:33007258
Consensus
enhancer/erna coordinates
Workflows
ChimericPRO
eRNA-mRNA
n1
n2
chimeric reads
rna2rna
EntropyTS
eRNAFasta
eRNAPrimer
eCIGAR
getREI
eRNASi
DATABASE
Expression
nucleus
Pol II
enhancer
eRNA
eRNA
Pol II
mRNA
promoter
Localisation
cytoplasm
eRNA
mitochondria
membrane
ribosome
polysome
mRNA
eRNA
Organ/Cell
Interaction
miRNA
eRNA
eRNA
lincRNA
protein coding
eRNA
eRNAkitApp
Expression
Expression across Organ
CPM
Organ
Expression across Cells
CPM
Cells
Localisation
Expression CN
CPM
polyA-
polyA+
location 1
location 2
eRNA
polyA- vs +
Cell
cell 1
cell 2
REI
eRNA
Subcellular fraction vs Cytoplasm
Fraction | Cell
f | cell 1
f | cell 2
REI
eRNA
Target
eRNA to mRNA Targets
gene A
gene B
eRNA
gene C
gene D
Observed Interactions
Counts
eRNA-mRNA

Figure 1

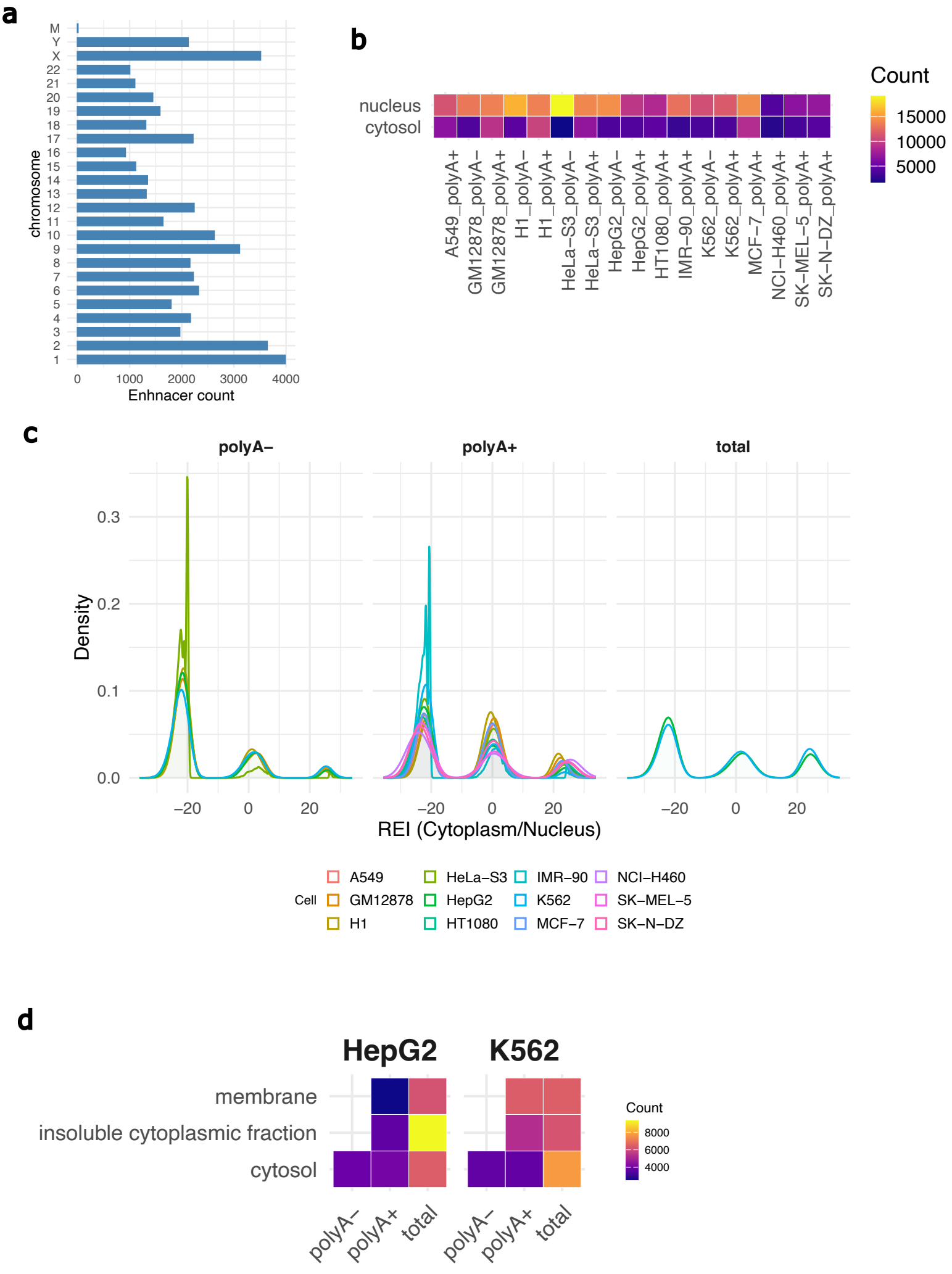

a
chromosome
M
Y
X
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
1000
2000
3000
4000
Enhnacer count
b
Count
15000
10000
5000
nucleus
cytosol
A549_polyA+
GM12878_polyA−
GM12878_polyA+
H1_polyA−
H1_polyA+
HeLa−S3_polyA−
HeLa−S3_polyA+
HepG2_polyA−
HepG2_polyA+
HT1080_polyA+
IMR−90_polyA+
K562_polyA−
K562_polyA+
MCF−7_polyA+
NCI−H460_polyA+
SK−MEL−5_polyA+
SK−N−DZ_polyA+
c
polyA−
polyA+
total
Density
0.3
0.2
0.1
0.0
−20
0
20
REI (Cytoplasm/Nucleus)
Cell
A549
GM12878
H1
HeLa−S3
HepG2
HT1080
IMR−90
K562
MCF−7
NCI−H460
SK−MEL−5
SK−N−DZ
d
HepG2
K562
membrane
insoluble cytoplasmic fraction
cytosol
polyA−
polyA+
total
Count
8000
6000
4000

Figure 2

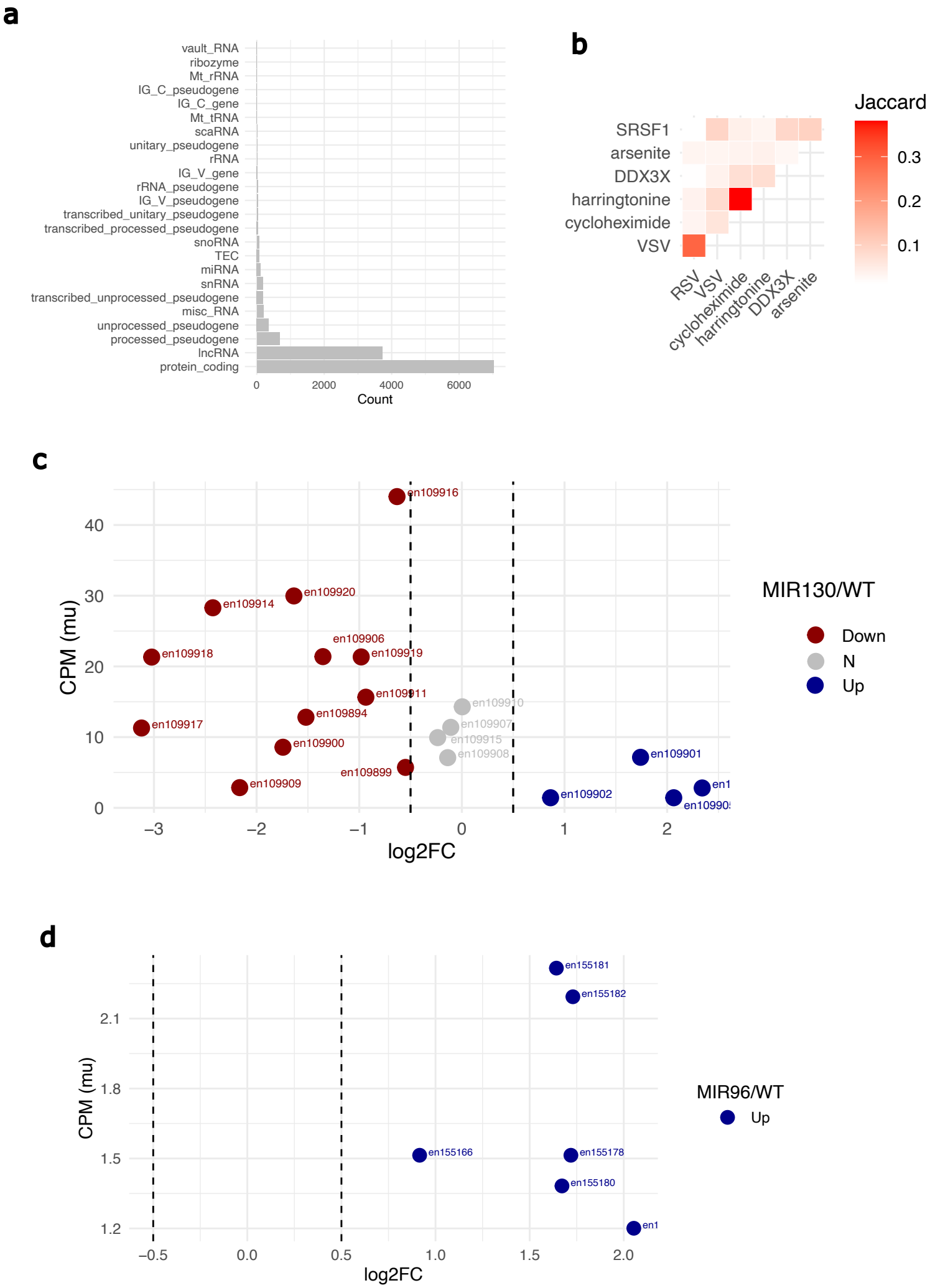
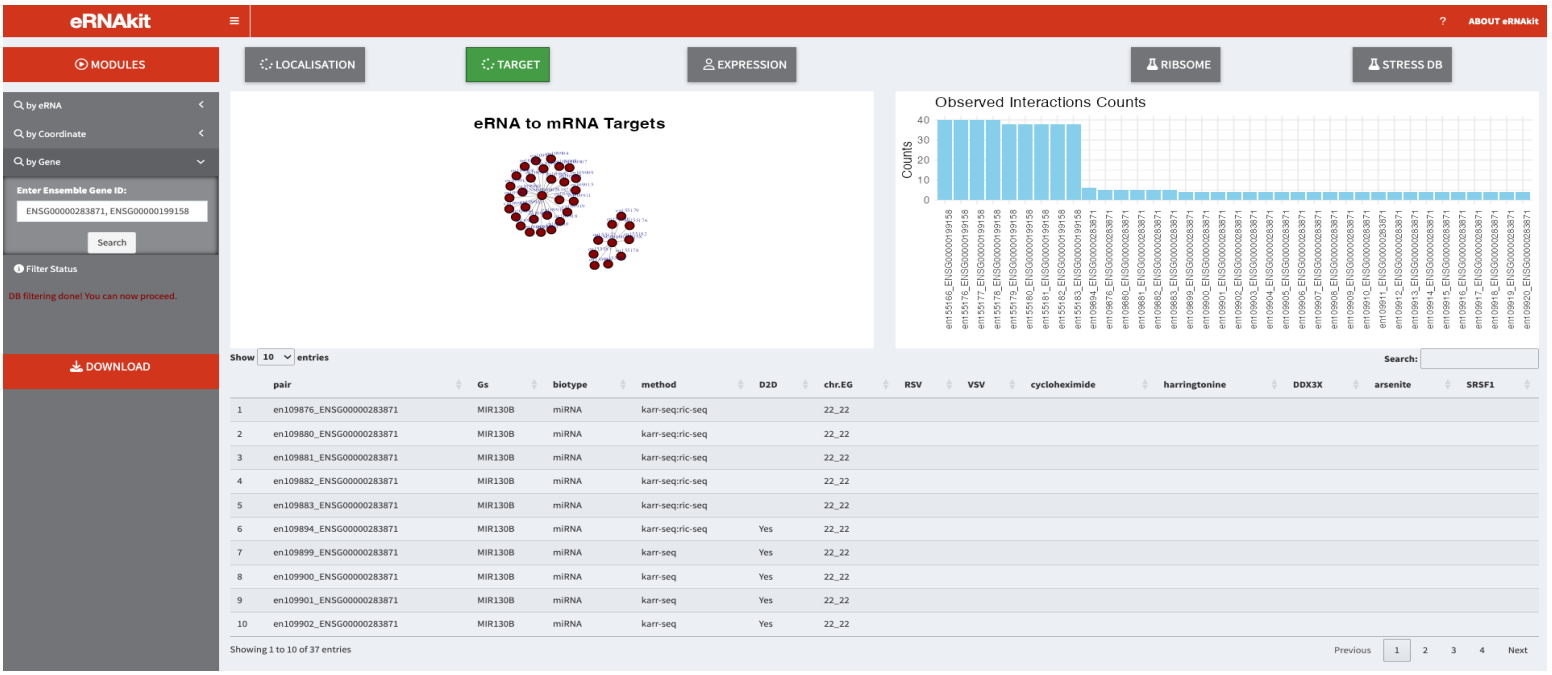
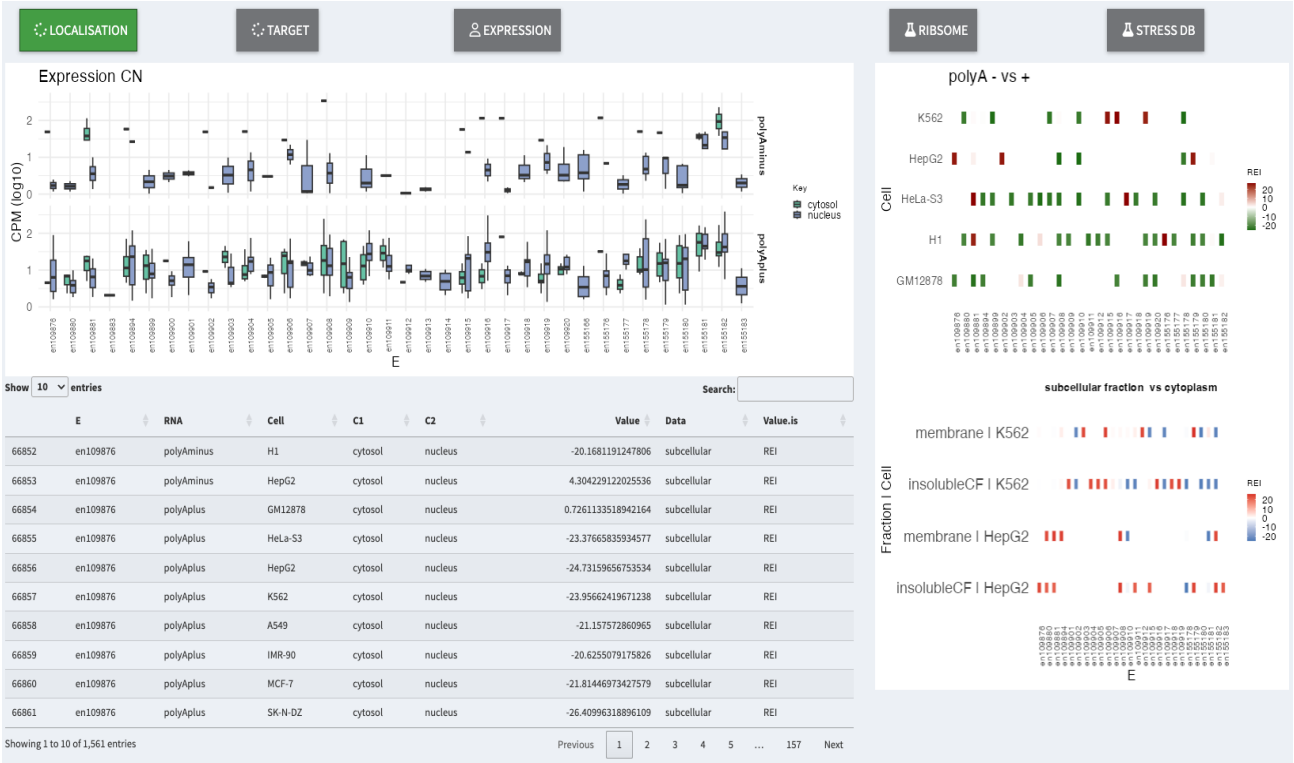

a
vault_RNA
ribozyme
Mt_rRNA
IG_C_pseudogene
IG_C_gene
Mt_tRNA
scaRNA
unitary_pseudogene
rRNA
IG_V_gene
rRNA_pseudogene
IG_V_pseudogene
transcribed_unitary_pseudogene
transcribed_processed_pseudogene
snoRNA
TEC
miRNA
snRNA
transcribed_unprocessed_pseudogene
misc_RNA
unprocessed_pseudogene
processed_pseudogene
lncRNA
protein_coding
0
2000
4000
6000
Count
b
Jaccard
0.3
0.2
0.1
SRSF1
arsenite
DDX3X
harringtonine
cycloheximide
VSV
RSV
VSV
cycloheximide
harringtonine
DDX3X
arsenite
c
CPM (mu)
log2FC
MIR130/WT
Down
N
Up
en109916
en109920
en109914
en109906
en109918
en109919
en109911
en109910
en109894
en109907
en109917
en109915
en109900
en109908
en109901
en109899
en109909
en109902
d
CPM (mu)
log2FC
MIR96/WT
Up
en155181
en155182
en155166
en155178
en155180

Figure 3

a



b



c

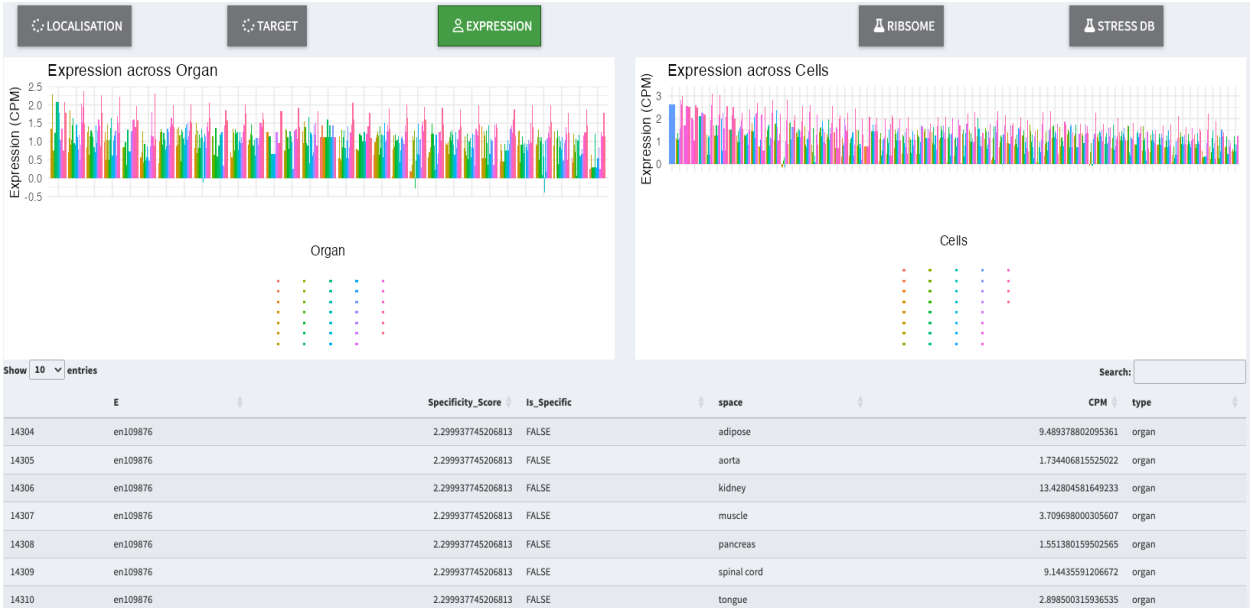
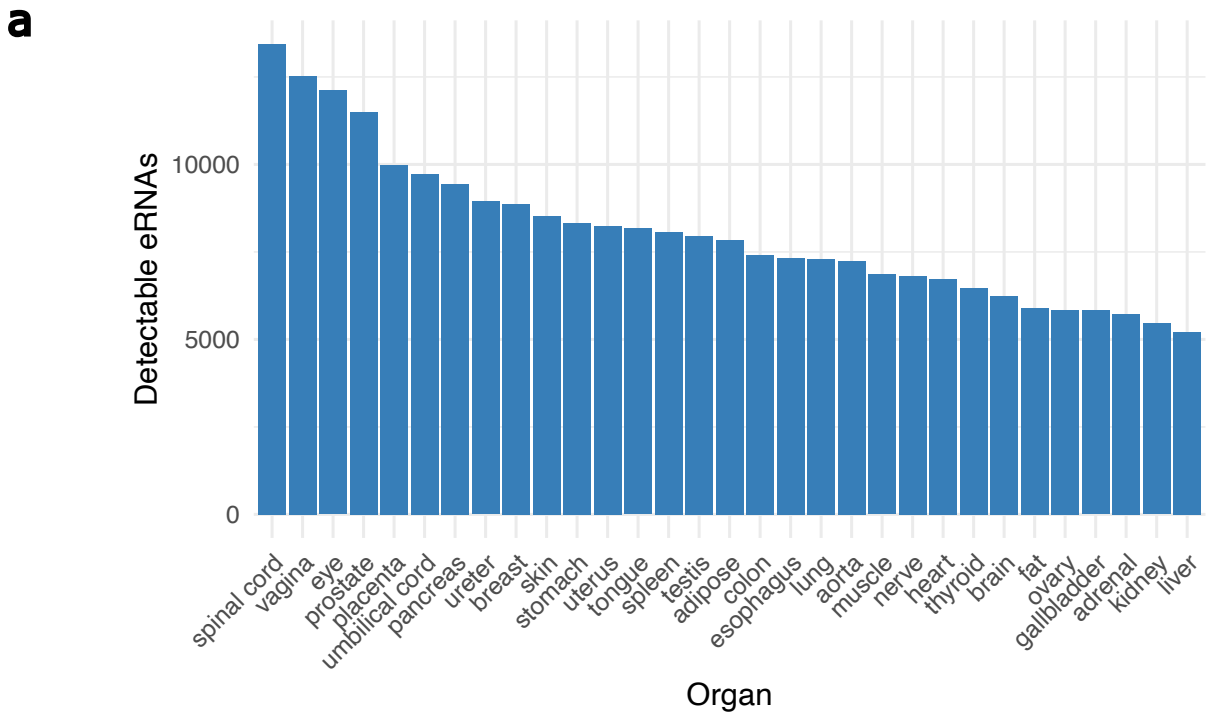


Figure S1


a
Detectable eRNAs
10000
5000
0
spinal cord
vagina
eye
prostate
placenta
umbilical cord
pancreas
ureter
breast
skin
stomach
uterus
tongue
spleen
testis
adipose
colon
esophagus
lung
aorta
muscle
nerve
heart
thyroid
brain
fat
ovary
gallbladder
adrenal
kidney
liver
Organ

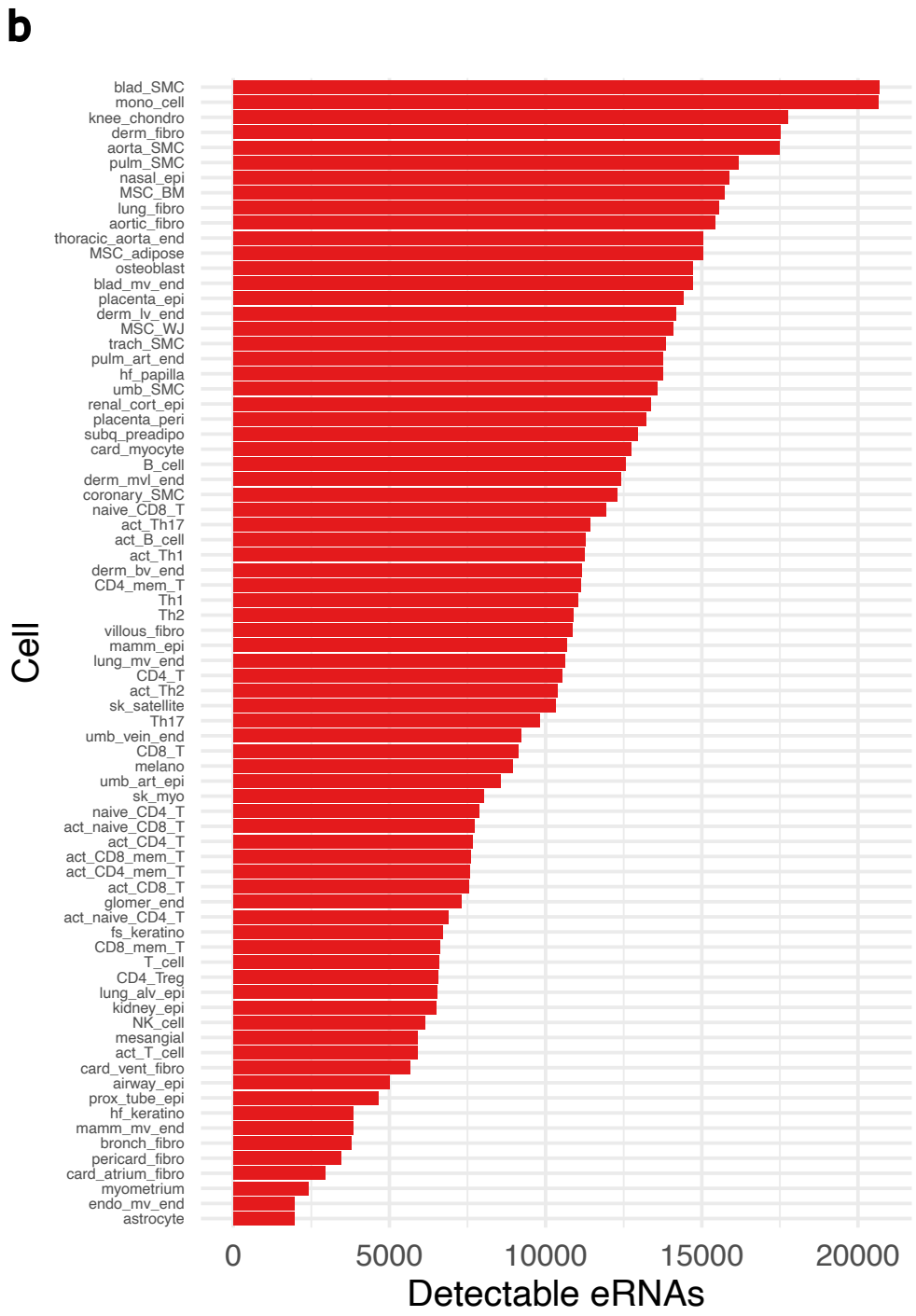

b
Cell
blad_SMC
mono_cell
knee_chondro
derm_fibro
aorta_SMC
pulm_SMC
nasal_epi
MSC_BM
lung_fibro
aortic_fibro
thoracic_aorta_end
MSC_adipose
osteoblast
blad_mv_end
placenta_epi
derm_lv_end
MSC_WJ
trach_SMC
pulm_art_end
hf_papilla
umb_SMC
renal_cort_epi
placenta_peri
subq_preadipo
card_myocyte
B_cell
derm_mvl_end
coronary_SMC
naive_CD8_T
act_Th17
act_B_cell
act_Th1
derm_bv_end
CD4_mem_T
Th1
Th2
villous_fibro
mamm_epi
lung_mv_end
CD4_T
act_Th2
sk_satellite
Th17
umb_vein_end
CD8_T
melano
umb_art_epi
sk_myo
naive_CD4_T
act_naive_CD8_T
act_CD4_T
act_CD8_mem_T
act_CD4_mem_T
act_CD8_T
glomer_end
act_naive_CD4_T
fs_keratino
CD8_mem_T
T_cell
CD4_Treg
lung_alv_epi
kidney_epi
NK_cell
mesangial
act_T_cell
card_vent_fibro
airway_epi
prox_tube_epi
hf_keratino
mamm_mv_end
bronch_fibro
pericard_fibro
card_atrium_fibro
myometrium
endo_mv_end
astrocyte
0
5000
10000
15000
20000
Detectable eRNAs

Figure S2

a

eRNAkit

ABOUT eRNAkit

MODULES | LOCALISATION | TARGET | EXPRESSION | RIBSOME | STRESS ERNA

by eRNA

Enter eRNA: en135763

Search

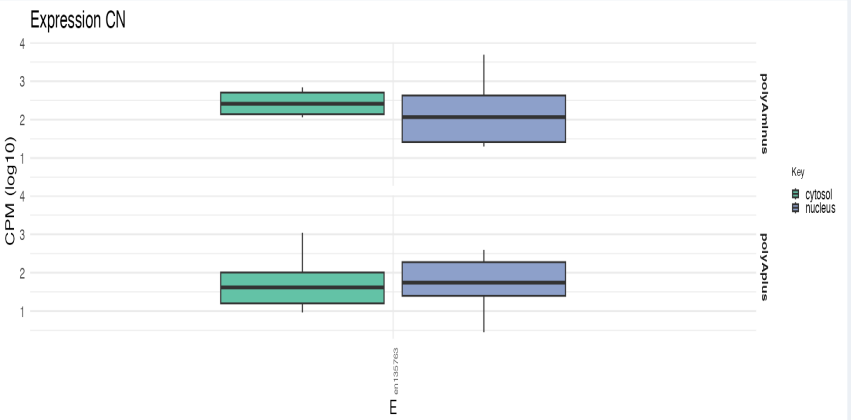
by Coordinate

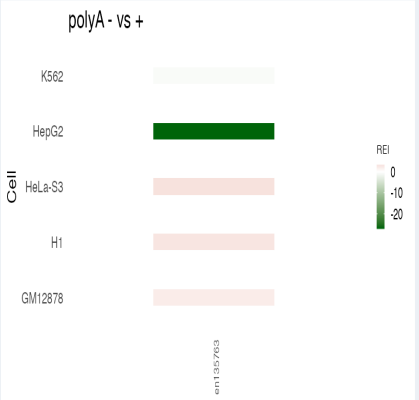
by Gene

Filter Status

DB filtering done! You can now proceed.

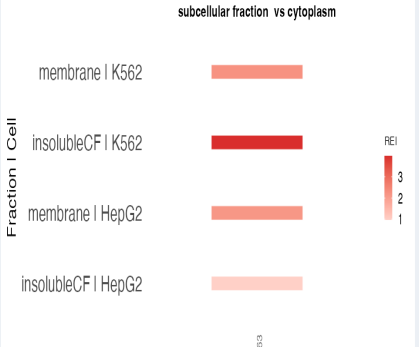
DOWNLOAD





Show 10 entries Search:

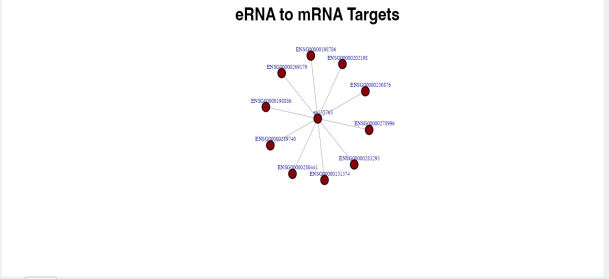
	E	RNA	Cell	C1	C2	Value	Data	Value.is
208689	en135763	polyAminus	GM12878	cytosol	nucleus	1.971989413367965	subcellular	REI
208690	en135763	polyAminus	H1	cytosol	nucleus	-1.531603088934351	subcellular	REI
208691	en135763	polyAminus	HeLa-S3	cytosol	nucleus	2.520848575599159	subcellular	REI
208692	en135763	polyAminus	HepG2	cytosol	nucleus	-24.63296582162621	subcellular	REI
208693	en135763	polyAminus	K562	cytosol	nucleus	-2.824517712029208	subcellular	REI
208694	en135763	polyAplus	GM12878	cytosol	nucleus	-0.2816103276069366	subcellular	REI
208695	en135763	polyAplus	H1	cytosol	nucleus	-4.390051565759349	subcellular	REI

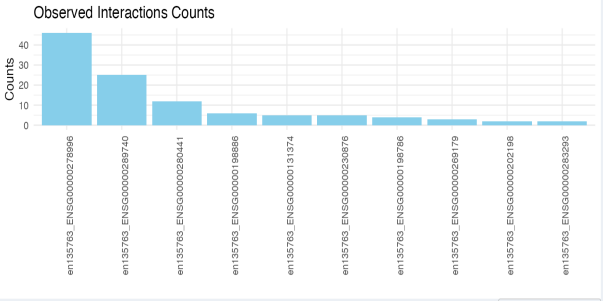


b

ABOUT eRNAkit

LOCALISATION | TARGET | EXPRESSION | RIBSOME | STRESS ERNA





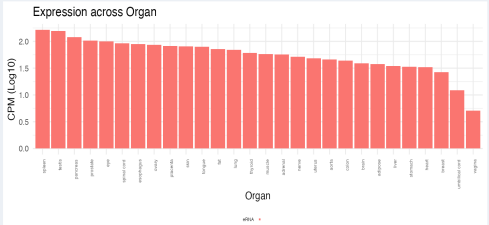
Show 10 entries Search:

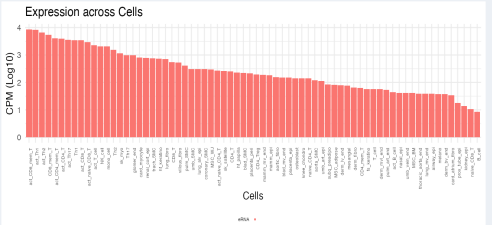
	pair	Gs	biotype	method	D2D	chr.EG	RSV	VSV	cycloheximide	harringtonine	DDX3X	arsenite	SRSF1
1	en135763_ENSG00000131374	TBC1D5	protein_coding	karr-seq:ric-seq		5_3							
2	en135763_ENSG00000198786	MT-ND5	protein_coding	karr-seq:ric-seq		5_MT						loss	
3	en135763_ENSG00000198886	MT-ND4	protein_coding	ric-seq:karr-seq		5_MT							loss
4	en135763_ENSG00000202198	7SK	misc_RNA	karr-seq		5_6						gain	
5	en135763_ENSG00000230876	LINC00486	lncRNA	ric-seq:paris		5_2							
6	en135763_ENSG00000269179		protein_coding	ric-seq:karr-seq		5_19							
7	en135763_ENSG00000278996		lncRNA	paris		5_21							

c

ABOUT eRNAkit

LOCALISATION | TARGET | EXPRESSION | RIBSOME | STRESS ERNA





Show 10 entries Search:

	E	Specificity_Score	Is_Specific	space	CPM	type
54426	en135763	0.3751372868217135	FALSE	adipose	37.51883175394445	organ
54427	en135763	0.3751372868217135	FALSE	adrenal	57.08033118362893	organ
54428	en135763	0.3751372868217135	FALSE	aorta	45.80370400852761	organ
54429	en135763	0.3751372868217135	FALSE	brain	38.83009245948874	organ
54430	en135763	0.3751372868217135	FALSE	breast	26.78111128221266	organ
54431	en135763	0.3751372868217135	FALSE	colon	43.53315049410126	organ
54432	en135763	0.3751372868217135	FALSE	esophagus	88.86255924170615	organ

Figure S3