# 1 Supplementary Text

## 1.1 Theoretical P-value calculation by the "hurdle model"

Suppose, ASSA produces the $SumEnergy = x$ when it is applied to a transcript pair with the $LASTAL$ score threshold set to *score*. To estimate the statistical significance of the observed value $x$, a background distribution is needed. This distribution can be defined as follows. Consider two random sequences S1 and S2 that have the lengths (L1 and L2) and GC contents (GC1 and GC2) same as the original transcripts, such that $L1 \leq L2$. Define a random variable $X = -1 \times SumEnergy_{rand}$ where the $SumEnergy_{rand}$ is the interaction energy predicted by ASSA for sequences S1 and S2 with the $LASTAL$ score threshold set to *score*. We simulate the distribution of $X$ by combining the Bernoulli and Gamma distributions into the hurdle model. Given the parameter of the Bernoulli distribution $P = P(X \neq 0)$ and the parameters of the Gamma distribution shape = $\alpha$ and rate = $\beta$, the CDF of $X \in [0, \infty)$ is defined as:

$$CDF_X(x) = P(X \leq x) = P(X = 0) + P(X \neq 0) \times CDF_{Gamma}(x|\alpha, \beta) \quad (1)$$

Taking into account that $P(X = 0) = 1 - P(X \neq 0)$ and $P(X \geq x) = 1 - CDF_X(x)$, it can be shown that for any $x > 0$ the P-value can be computed as:

$$\text{P-value}(x) = P(X \geq x) = P(X \neq 0) \times P(Gamma(\alpha, \beta) \geq x) \quad (2)$$

Thus, the distribution of $X$ is parameterized by $\alpha$, $\beta$ and $P = P(X \neq 0)$. It should be noted that the parameters of Gamma distribution ($\alpha$ and $\beta$) can be obtained from the mean and variance using the method of moments:

$$\alpha = \frac{mean^2}{var}; \quad \beta = \frac{mean}{var} \quad (3)$$

Importantly, the three distribution parameters ($P$, mean and variance) depend on the properties of the random sequences (L1, L2, GC1, GC2) as well as the $LASTAL$ score threshold used for ASSA run. Thus, our goal was to determine the coefficients of the linear regression formulas to predict the hurdle model parameters based on the following five features: *log(L1)* (the logarithm of the length of the shorter random sequence in a pair), *log(L2)* (the logarithm of the longer sequence length), *GCaver* (the average GC content of the sequences = $(GC1 + GC2)/2$), *GCdiff* (the absolute difference between GC contents = $|GC1 - GC2|$) and *score* (the $LASTAL$ score threshold). Thus, the corresponding regression equations to compute the log(mean), log(var) and logit(P) target variables are as follows:

$$\log(mean) = \theta_0^m + \theta_1^m \cdot \log(L1) + \theta_2^m \cdot \log(L2) + \theta_3^m \cdot \text{GCaver} + \theta_4^m \cdot \text{GCdiff} + \theta_5^m \cdot \text{score}$$

$$\log(var) = \theta_0^v + \theta_1^v \cdot \log(L1) + \theta_2^v \cdot \log(L2) + \theta_3^v \cdot \text{GCaver} + \theta_4^v \cdot \text{GCdiff} + \theta_5^v \cdot \text{score}$$

$$logit(P) = \theta_0^P + \theta_1^P \cdot \log(L1) + \theta_2^P \cdot \log(L2) + \theta_3^P \cdot \text{GCaver} + \theta_4^P \cdot \text{GCdiff} + \theta_5^P \cdot \text{score}$$

We used the log's of sequence lengths because there is a linear dependence between the $\log(mean)$, $\log(var)$ or $logit(P)$ on the $\log(L1)$ or $\log(L2)$ (see Supplementary Figure 11).

To determine the liner regression coefficients the 20532 sets of *SumEnergies* were used as a training data. To have enough statistics for reliable mean and variance calculation we only used the 4524 sets that had 20 or more sequence pairs with non-zero *SumEnergies*, each computed from at least 3 putative duplexes. To train the $logit(P)$ model the 17121 sets where P was not equal to either 0 or 1 (the logit function is not defined for these values) were used. All the regression coefficients were determined by the `lm()` function from R (see Supplementary Table 2). The three obtained models had good fit to the training data – the R-squared values for the $log(mean)$, $log(var)$ and $logit(P)$ models were 0.94, 0.926 and 0.831, respectively. It should be noted that we did not perform the regularization of the coefficients because overfitting is unlikely for the first order polynomial regressions.

The regression equations above can be simplified using linear algebra notations. By defining a column-vector $F$ containing all the feature values and 1 for the intercept term, i.e. $F = (1, \log(L1), \log(L2), GC1, GC2, score)^T$, and the three vectors with coefficients $\Theta^m$, $\Theta^v$ and $\Theta^P$, the formulas to compute the target variables can be written as: $\log(mean) = F^T \times \Theta^m$, $\log(var) = F^T \times \Theta^v$ and $logit(P) = F^T \times \Theta^P$. Thus, the values for the $mean$, $var$ and $P$ can be computed as follows:

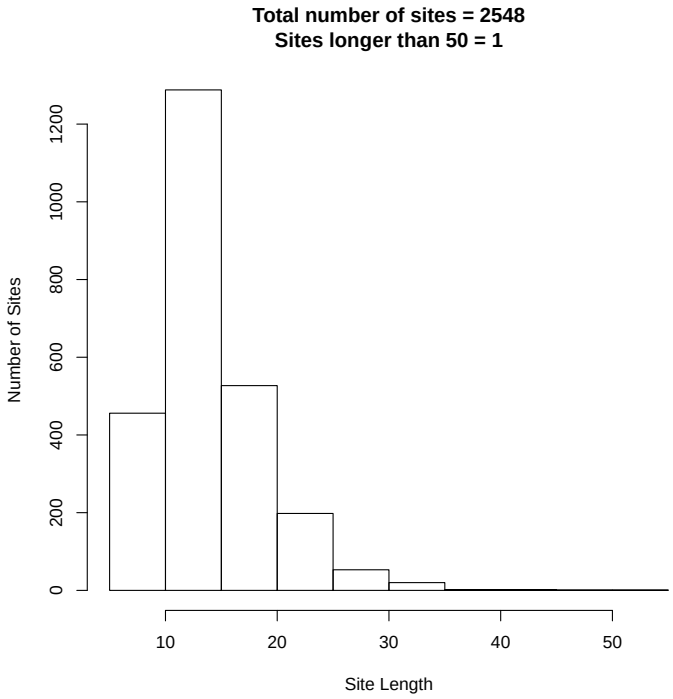
$$mean = \exp(F^T \times \Theta^m) \tag{4}$$

$$var = \exp(F^T \times \Theta^v) \tag{5}$$

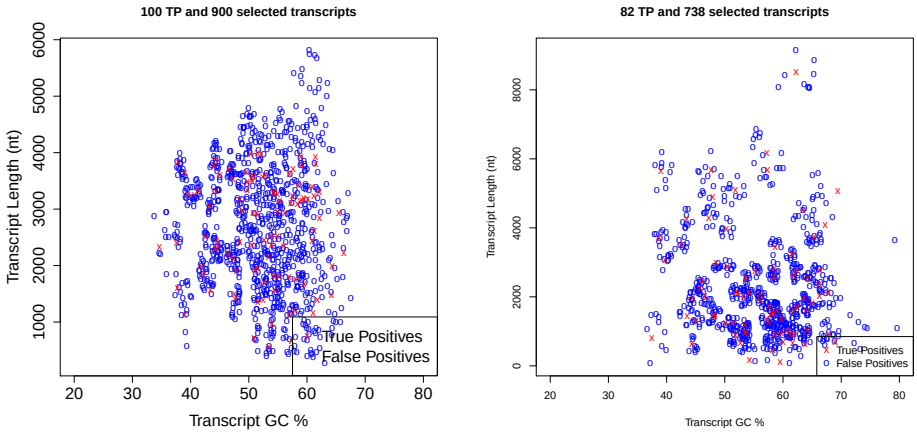
$$P = \frac{1}{1 + \exp(-F^T \times \Theta^P)} \tag{6}$$

Finally, the parameters of the Gamma distribution are obtained using the equations (3) and the Theoretical P-value of the observed $SumEnergy = x$ is computed with respect to the predicted background distribution using (2).

# 2 Supplementary Figures



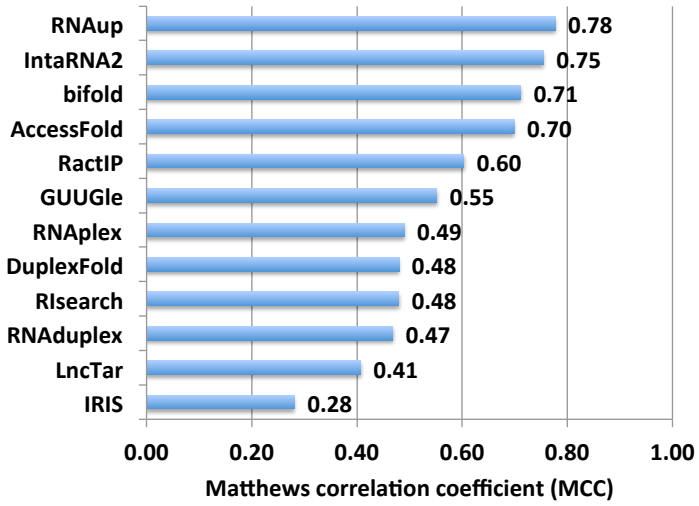

Supplementary Figure 1: The distribution of the antisense site lengths identified for 100 randomly selected human lncRNA-mRNA pairs (the LASTAL score threshold = 30).



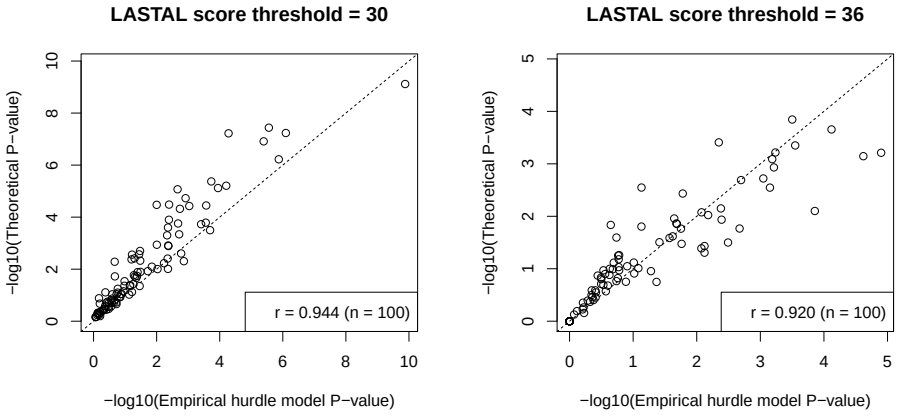

(a) TINCR: mix with human transcripts

(b) ACTB: mix with human transcripts

Supplementary Figure 2: Features (length and GC content) of the human transcripts included in the (a) TINCR and (b) ACTB "mix with human transcripts" test sets. The two test sets consist of 100 and 82 true targets (red crosses) and 900 and 738 other human transcripts with similar length and GC content (blue circles), respectively.

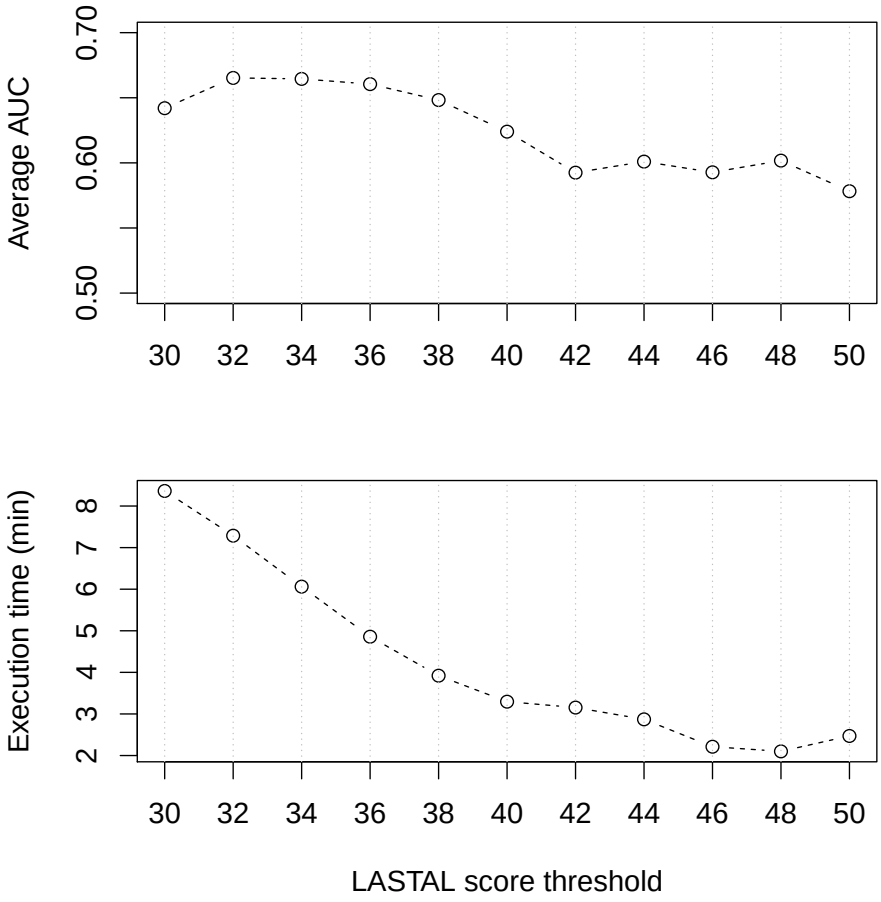



Supplementary Figure 3: Performance of 12 thermodynamics-based tools on a test set consisting of simulated duplexes

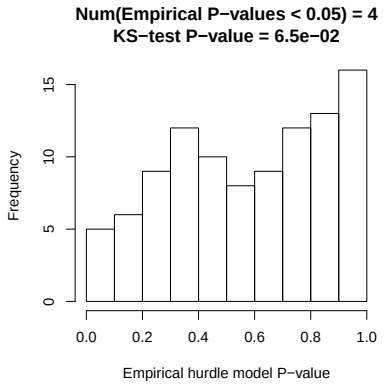



(a) (b)

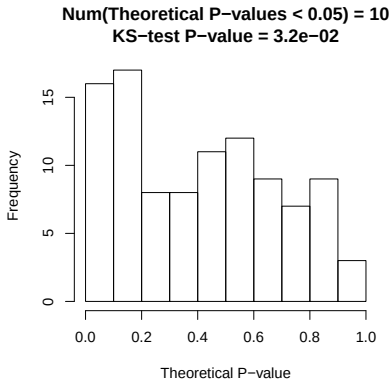
Supplementary Figure 4: The correlation between the Theoretical and the Empirical hurdle model P-values computed for 100 randomly selected lncRNA-mRNA pairs with the *LASTAL* score threshold equal to (a) 30 or (b) 36. Each transcript pair was used to generate 900 random sequence pairs which provided the SumEnergies for the MLE-estimates of the empirical hurdle model parameters.



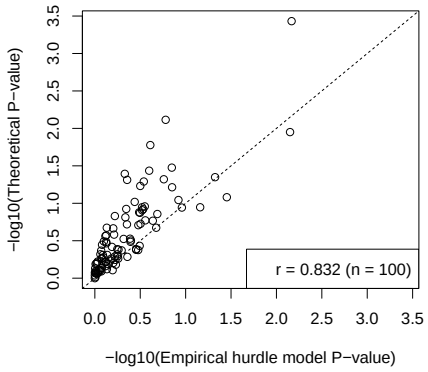

Supplementary Figure 5: The average AUC values and the average execution times demonstrated by ASSA on four validation sets when different LASTAL score thresholds were used




(a) Empirical P-values

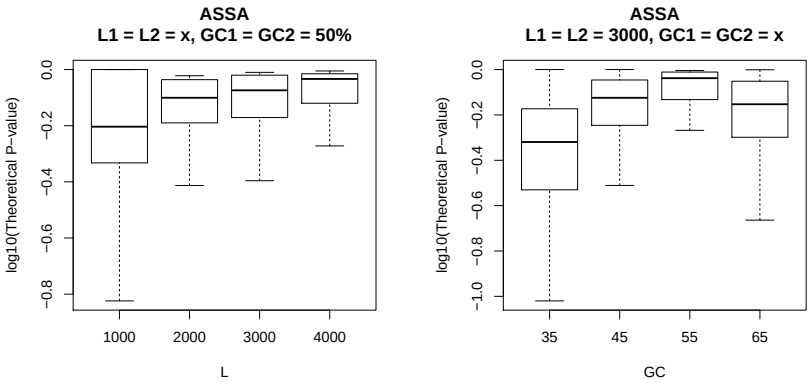



(b) Theoretical P-values




(c) Correlation between log's

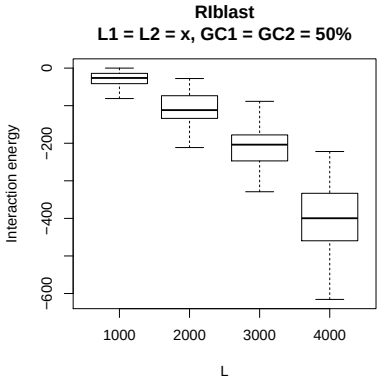
Supplementary Figure 6: Comparison of the Theoretical and the Empirical hurdle model P-values (LASTAL score threshold = 30) computed for 100 pairs of random sequence pairs with various lengths and GC contents.



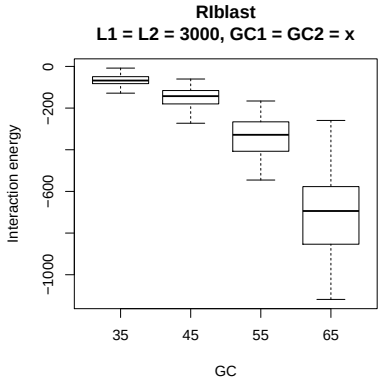

(a) ASSA Pvalue ∼ Length




(b) ASSA Pvalue ∼ GC content

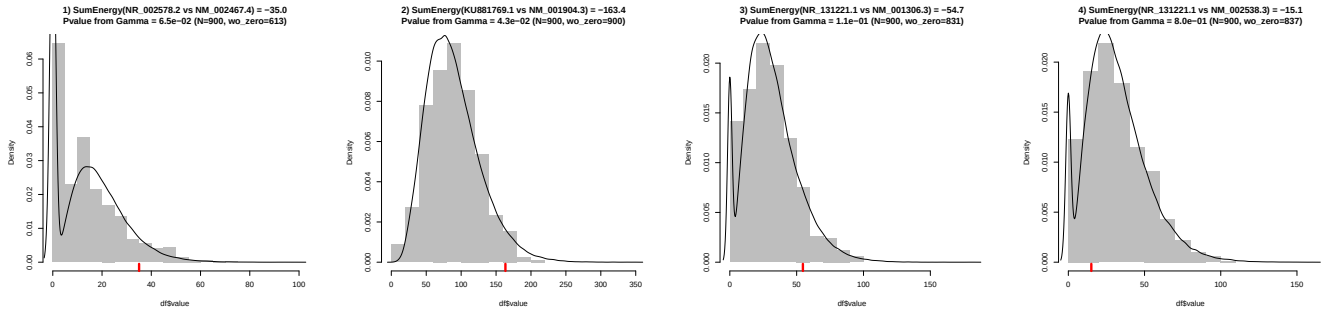



(c) RIblast Energy ∼ Length




(d) RIblast Energy ∼ GC content

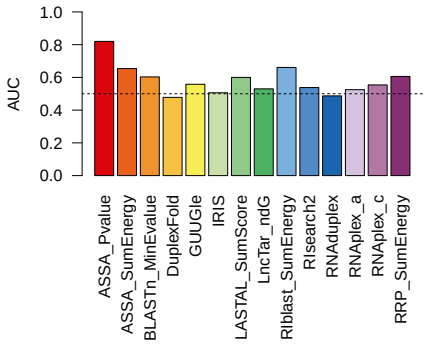
Supplementary Figure 7: The dependence of the (a,b) log10(ASSA P-value) with the LASTAL score threshold = 36 and (c,d) RIblast SumEnergy on the length and GC content of the random sequences. Each boxplot represents distribution of 100 values obtained for random sequence pairs with the identical lengths and GC contents (see the chart titles). The values of the parameter indicted as "x" in the title are on the X-axis of the corresponding chart.



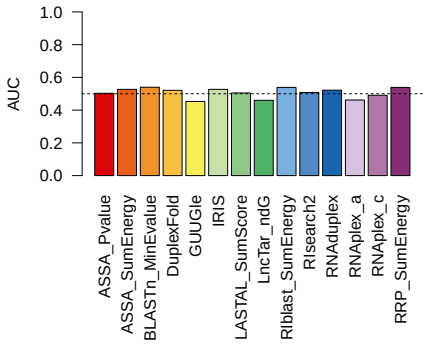

(a) GAS5::MYC (ASSA) (b) LincRNA-p21::CTNNB1 (c) SPRY4-IT1::CLDN3 (d) SPRY4-IT1::OCLN

(e) GAS5::MYC (RIblast) (f) LincRNA-p21::CTNNB1 (g) SPRY4-IT1::CLDN3 (h) SPRY4-IT1::OCLN

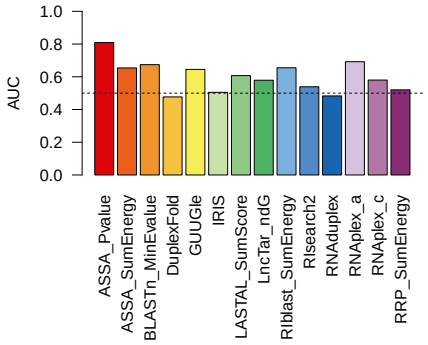
Supplementary Figure 8: The hurdle model empirical P-values computed from (a-c) ASSA (LASTAL score threshold = 36) and (d-f) RIblast SumEnergies for the functional short-trans interactions with Theoretical P-values $> 0.05$. Every background distribution consists of 900 SumEnergies computed for mono-nucleotide shuffled sequences.



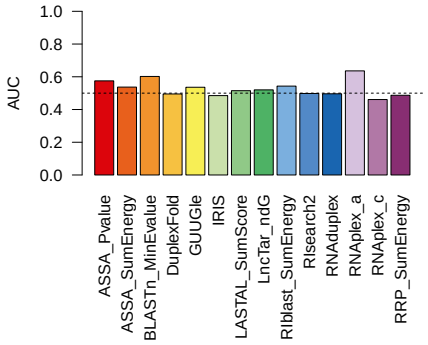

(a) TINCR: mix with shuffled sequences




(b) TINCR: mix with human transcripts

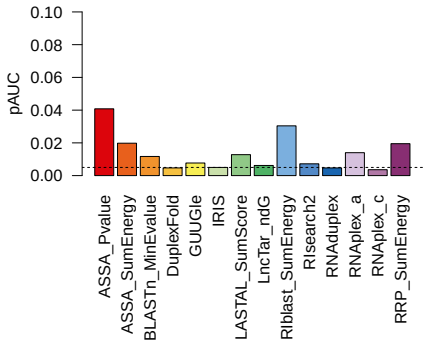



(c) ACTB: mix with shuffled sequences

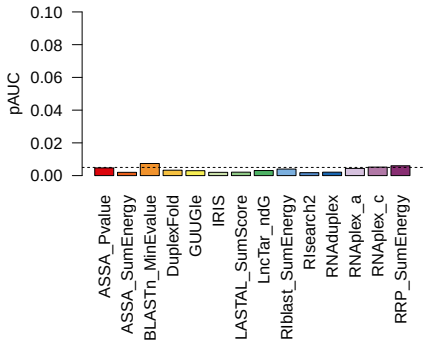



(d) ACTB: mix with human transcripts

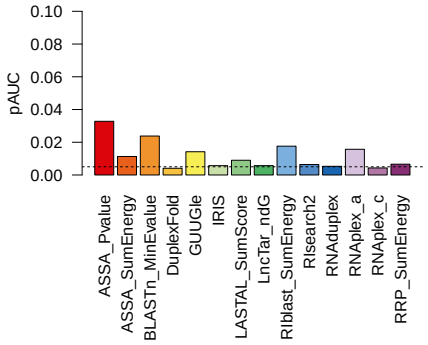
Supplementary Figure 9: AUC values obtained by different RNA-RNA prediction approaches on four test sets.



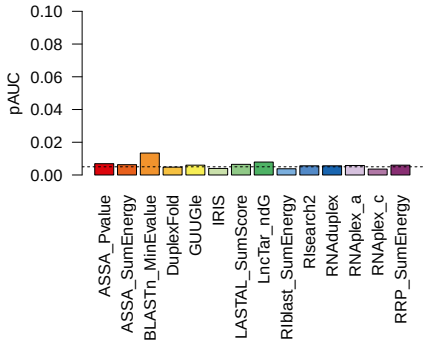

(a) TINCR: mix with shuffled sequences




(b) TINCR: mix with human transcripts

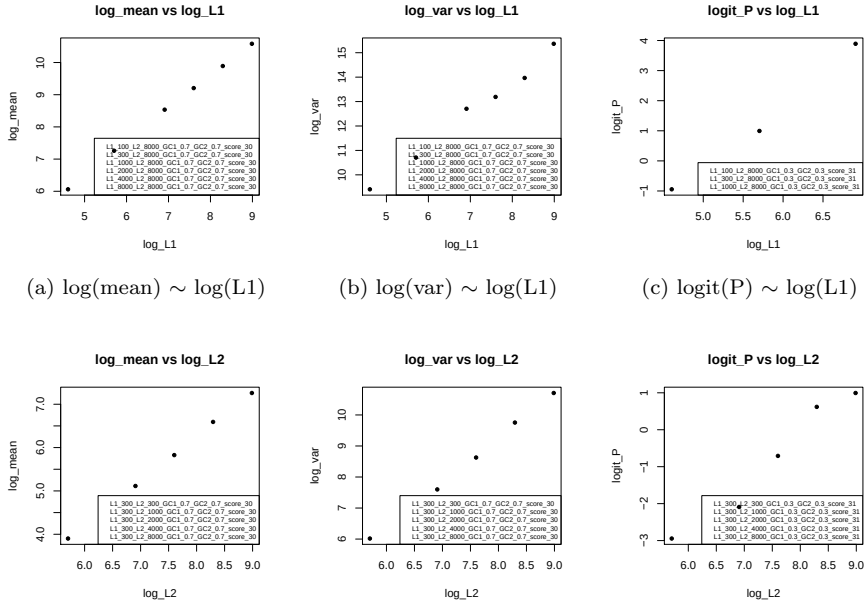



(c) ACTB: mix with shuffled sequences




(d) ACTB: mix with human transcripts

Supplementary Figure 10: Partial AUC values (false positive rate range between 0 and 0.1) obtained by different RNA-RNA prediction approaches




(a) log(mean) ∼ log(L1) (b) log(var) ∼ log(L1) (c) logit(P) ∼ log(L1)

(d) log(mean) ∼ log(L2) (e) log(var) ∼ log(L2) (f) logit(P) ∼ log(L2)

Supplementary Figure 11: The observed linear dependence between the log(mean)/log(var)/logit(P) and log(L1)/log(L2). Six different values were used for L1 (100, 300, 1k, 2k, 4k, 8k) and five different values – for L2 (300, 1k, 2k, 4k, 8k). Other properties of the random sequences and the LASTAL score thresholds are specified in the legend boxes. It should be noted that the mean and variance are not computed if the number of obtained SumEnergy values is less than 20; logit function is not defined if P = 1 or P = 0.

# 3 Supplementary Tables

Supplementary Table 1: Custom substitution matrix used for LASTAL run in ASSA. Since the LASTAL searches the reverse complement of the target sequences, the matrix diagonal contains weights for A-T and G-C base pairings and the G-A and T-C matrix weights correspond to the G-U wobble base pairing.

| | A | C | G | T |
|---|---|---|---|---|
| A | 2 | -6 | -6 | -6 |
| C | -6 | 4 | -6 | -6 |
| G | 1 | -6 | 4 | -6 |
| T | -6 | 1 | -6 | 2 |

Supplementary Table 2: Linear regression coefficients used in ASSA to predict the three parameters of the hurdle model.

| | | log(mean) | log(var) | logit(P) |
|---|---|---|---|---|
| $\theta_0$ | (Intercept) | -7.52533270114322 | -8.26126584589412 | -11.7030901792339 |
| $\theta_1$ | log(L1) | 0.959757241083511 | 1.09978076320836 | 1.21409710977686 |
| $\theta_2$ | log(L2) | 0.970083906725601 | 1.21962418536611 | 1.07599802655411 |
| $\theta_3$ | GCaver | 0.132465609760186 | 0.164629880507987 | 0.187136520807308 |
| $\theta_4$ | GCdiff | -0.0306943559106818 | -0.0300034089811152 | -0.0462828560516273 |
| $\theta_5$ | score | -0.260337628983731 | -0.298713533663101 | -0.340412558547227 |

Supplementary Table 3: ASSA predictions (LASTAL score threshold = 36) for the published cases of functional RNA-RNA interactions

| # | Gene 1 | Gene 2 | Ref | Type | ASSA P-value |
|---|---|---|---|---|---|
| 1 | BACE1-AS | BACE1 | [1] | *cis* | 2.86E-07 |
| 2 | DHPS | WDR83 | [2] | *cis* | 7.45E-07 |
| 3 | HIF1A-AS2 | HIF1A | [3] | *cis* | 1.86E-82 |
| 4 | IFNA1 AS | IFNA1 | [4] | *cis* | 7.96E-84 |
| 5 | NR1D1 | THRA | [5] | *cis* | 8.70E-07 |
| 6 | PCNA-AS1 | PCNA | [6] | *cis* | 3.16E-30 |
| 7 | PU.1-AS | SPI1 | [7] | *cis* | 2.97E-06 |
| 8 | Uchl1os | UCHL1 | [8] | *cis* | 7.84E-13 |
| 9 | UXT-AS1 | UXT | [8] | *cis* | 5.28E-29 |
| 10 | WRAP53 | TP53 | [9] | *cis* | 7.09E-08 |
| 11 | ZEB2-AS1 | ZEB2 | [10] | *cis* | 3.94E-22 |
| 12 | Arhgap20os | Kifc1 | [11] | *pseudo-cis* | 1.60E-13 |
| 13 | E330011O21Rik | Hdac1 | [11] | *pseudo-cis* | 4.62E-35 |
| 14 | Gm29811 | Hsp90ab1 | [11] | *pseudo-cis* | 2.97E-33 |
| 15 | Khdc1a | Oog4 | [11] | *pseudo-cis* | 9.44E-21 |
| 16 | 1/2-sbsRNA1 | SERPINE1 | [12] | *Alu*-based | 5.58E-05 |
| 17 | 1/2-sbsRNA2 | CDCP1 | [12] | *Alu*-based | 1.03E-08 |
| 18 | 1/2-sbsRNA3 | MTAP | [12] | *Alu*-based | 2.75E-12 |
| 19 | 1/2-sbsRNA4 | MTAP | [12] | *Alu*-based | 9.20E-16 |
| 20 | RAB11FIP1 | TUFT1 | [13] | *Alu*-based | 6.14E-22 |
| 21 | RAB11FIP1 | CDCP1 | [13] | *Alu*-based | 3.97E-11 |
| 22 | RAB11FIP1 | PPID | [13] | *Alu*-based | 3.79E-25 |
| 23 | SOWAHC | CDCP1 | [13] | *Alu*-based | 5.31E-11 |
| 24 | SOWAHC | TUFT1 | [13] | *Alu*-based | 1.95E-10 |
| 25 | SOWAHC | PPID | [13] | *Alu*-based | 2.91E-15 |
| 26 | 7SL | TP53 | [14] | *short-trans* | 2.53E-07 |
| 27 | GAS5 | MYC | [15] | *short-trans* | 1.23E-01 |
| 28 | LincRNA-p21 | CTNNB1 | [16] | *short-trans* | 9.18E-02 |
| 29 | LincRNA-p21 | JUNB | [16] | *short-trans* | 3.50E-03 |
| 30 | lncRNA-ATB | IL11 | [17] | *short-trans* | 4.02E-08 |
| 31 | SPRY4-IT1 | F11R | [18] | *short-trans* | 6.74E-06 |
| 32 | SPRY4-IT1 | CLDN3 | [18] | *short-trans* | 1.71E-01 |
| 33 | SPRY4-IT1 | CLDN1 | [18] | *short-trans* | 2.04E-02 |
| 34 | SPRY4-IT1 | OCLN | [18] | *short-trans* | 6.36E-01 |

Supplementary Table 4: Comparison of the RNA-RNA interaction prediction tools on four test sets. Column notation: Time – total execution time on the corresponding test set (in minutes), AUC – area under the ROC curve, pAUC – partial AUC (false positive rate range between 0 and 0.1). The maximum AUC and pAUC values in each column are bold. The execution time of all the tools is measured on one CPU. *ASSA execution times on 24 CPUs were 57, 78, 20 and 24 minutes, respectively.

| | | **TINCR: mix with shuffled** | | | **TINCR: mix with human** | | | **ACTB: mix with shuffled** | | | **ACTB: mix with human** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Time** | **AUC** | **pAUC** | **Time** | **AUC** | **pAUC** | **Time** | **AUC** | **pAUC** | **Time** | **AUC** | **pAUC** |
| 1 | ASSA (P-value) | 1022* | **0.82** | **0.0408** | 1776* | 0.503 | 0.0047 | 358* | **0.809** | **0.0328** | 563* | 0.575 | 0.0069 |
| 2 | ASSA (SumEnergy) | 1022* | 0.654 | 0.0198 | 1776* | 0.527 | 0.002 | 358* | 0.654 | 0.0113 | 563* | 0.537 | 0.0063 |
| 3 | BLASTn (E-value) | 16 | 0.603 | 0.0117 | 23 | **0.54** | **0.0074** | 16 | 0.674 | 0.0238 | 12 | 0.602 | **0.0134** |
| 4 | DuplexFold | 743 | 0.478 | 0.0047 | 694 | 0.521 | 0.0033 | 173 | 0.477 | 0.0041 | 271 | 0.495 | 0.0048 |
| 5 | GUUGle | 2 | 0.558 | 0.0077 | 2 | 0.453 | 0.003 | 1 | 0.645 | 0.0142 | 1 | 0.536 | 0.006 |
| 6 | IRIS | 10236 | 0.506 | 0.005 | 8909 | 0.527 | 0.002 | 2376 | 0.505 | 0.0057 | 2395 | 0.485 | 0.004 |
| 7 | LASTAL (SumScore) | 22 | 0.6 | 0.0128 | 15 | 0.505 | 0.0021 | 5 | 0.607 | 0.009 | 5 | 0.515 | 0.0065 |
| 8 | LncTar (ndG) | 2783 | 0.53 | 0.0062 | 2882 | 0.46 | 0.0031 | 1012 | 0.579 | 0.0057 | 1012 | 0.52 | 0.0079 |
| 9 | RIblast (SumEnergy) | 4313 | 0.661 | 0.0304 | 4143 | 0.539 | 0.004 | 1481 | 0.655 | 0.0176 | 1460 | 0.543 | 0.0038 |
| 10 | RIsearch2 | 81 | 0.538 | 0.0072 | 61 | 0.508 | 0.0018 | 45 | 0.539 | 0.0064 | 23 | 0.498 | 0.0056 |
| 11 | RNAduplex | 1625 | 0.487 | 0.0047 | 1522 | 0.522 | 0.0021 | 475 | 0.483 | 0.0053 | 536 | 0.496 | 0.0056 |
| 12 | RNAplex-a | 508 | 0.525 | 0.014 | 449 | 0.462 | 0.0044 | 164 | 0.692 | 0.0157 | 192 | **0.636** | 0.0058 |
| 13 | RNAplex-c | 713 | 0.554 | 0.0036 | 458 | 0.491 | 0.0052 | 91 | 0.58 | 0.0043 | 128 | 0.461 | 0.0036 |
| 14 | RRP (SumEnergy) | 3182 | 0.606 | 0.0195 | 2982 | 0.539 | 0.006 | 1502 | 0.52 | 0.0066 | 1629 | 0.487 | 0.006 |

Supplementary Table 5: Summary of the files in the Supplementary data

| # | Related to | Description |
|---|---|---|
| 01 | Supplementary Table 2 | The dataset used to optimize the linear regression coefficients for ASSA |
| 02 | Supplementary Figure 4 | The sequences of the randomly selected human lncRNAs and mRNAs |
| 03 | Supplementary Table 3 | The sequences of the published cases of the functional RNA-RNA interactions |
| 04 | Supplementary Table 4<br>Figure 6<br>Supplementary Figure 9<br>Supplementary Figure 10 | The four test sets used to compare different RNA-RNA prediction approaches |

# Supplementary References


[1] Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., Finch, C. E., St Laurent, G., r., Kenny, P. J., and Wahlestedt, C. (2008) Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med,* **14**(7), 723–30.

[2] Su, W. Y., Li, J. T., Cui, Y., Hong, J., Du, W., Wang, Y. C., Lin, Y. W., Xiong, H., Wang, J. L., Kong, X., Gao, Q. Y., Wei, L. P., and Fang, J. Y. (2012) Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. *Cell Res,* **22**(9), 1374–89.

[3] Uchida, T., Rossignol, F., Matthay, M. A., Mounier, R., Couette, S., Clottes, E., and Clerici, C. (2004) Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. *J Biol Chem,* **279**(15), 14871–8.

[4] Kimura, T., Jiang, S., Nishizawa, M., Yoshigai, E., Hashimoto, I., Nishikawa, M., Okumura, T., and Yamada, H. (2013) Stabilization of human interferon-alpha1 mRNA by its antisense RNA. *Cell Mol Life Sci,* **70**(8), 1451–67.

[5] Hastings, M. L., Ingle, H. A., Lazar, M. A., and Munroe, S. H. (2000) Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J Biol Chem,* **275**(15), 11507–13.

[6] Yuan, S. X., Tao, Q. F., Wang, J., Yang, F., Liu, L., Wang, L. L., Zhang, J., Yang, Y., Liu, H., Wang, F., Sun, S. H., and Zhou, W. P. (2014) Antisense long non-coding RNA PCNA-AS1 promotes tumor growth by regulating proliferating cell nuclear antigen in hepatocellular carcinoma. *Cancer Lett,* **349**(1), 87–94.

[7] Ebralidze, A. K., Guibal, F. C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M. A., Petkova, V., Rosenbauer, F., Huang, G., Dayaram, T., Klupp, J., O'Brien, K. B., Will, B., Hoogenkamp, M., Borden, K. L., Bonifer, C., and Tenen, D. G. (2008) PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. *Genes Dev,* **22**(15), 2085–92.

[8] Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A. R., Carninci, P., Biffo, S., Stupka, E., and Gustincich, S. (2012) Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature,* **491**(7424), 454–7.

[9] Mahmoudi, S., Henriksson, S., Corcoran, M., Mendez-Vidal, C., Wiman, K. G., and Farnebo, M. (2009) Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell,* **33**(4), 462–71.

[10] Beltran, M., Puig, I., Pena, C., Garcia, J. M., Alvarez, A. B., Pena, R., Bonilla, F., and de Herreros, A. G. (2008) A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev,* **22**(6), 756–69.

[11] Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R. M., and Hannon, G. J. (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature,* **453**(7194), 534–8.

[12] Gong, C. and Maquat, L. E. (2011) lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature,* **470**(7333), 284–8.

[13] Gong, C., Tang, Y., and Maquat, L. E. (2013) mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat Struct Mol Biol,* **20**(10), 1214–20.

[14] Abdelmohsen, K., Panda, A. C., Kang, M.-J., Guo, R., Kim, J., Grammatikakis, I., Yoon, J.-H., Dudekula, D. B., Noh, J. H., Yang, X., Martindale, J. L., and Gorospe, M. (Sep, 2014) 7SL RNA represses p53 translation by competing with HuR. *Nucleic Acids Res,* **42**(15), 10099–111.

[15] Hu, G., Lou, Z., and Gupta, M. (2014) The long non-coding RNA GAS5 cooperates with the eukaryotic translation initiation factor 4E to regulate c-Myc translation. *PLoS One,* **9**(9), e107016.

[16] Yoon, J. H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J. L., De, S., Huarte, M., Zhan, M., Becker, K. G., and Gorospe, M. (2012) LincRNA-p21 suppresses target mRNA translation. *Mol Cell,* **47**(4), 648–55.

[17] Yuan, J. H., Yang, F., Wang, F., Ma, J. Z., Guo, Y. J., Tao, Q. F., Liu, F., Pan, W., Wang, T. T., Zhou, C. C., Wang, S. B., Wang, Y. Z., Yang, Y., Yang, N., Zhou, W. P., Yang, G. S., and Sun, S. H. (2014) A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer Cell,* **25**(5), 666–81.

[18] Xiao, L., Rao, J. N., Cao, S., Liu, L., Chung, H. K., Zhang, Y., Zhang, J., Liu, Y., Gorospe, M., and Wang, J.-Y. (Feb, 2016) Long noncoding RNA SPRY4-IT1 regulates intestinal epithelial barrier function by modulating the expression levels of tight junction proteins. *Mol Biol Cell,* **27**(4), 617–26.