SUPPLEMENTAL INFORMATION TO THE MANUSCRIPT:

*μ*IVC-Useq: a microfluidic-assisted high-throughput functionnal screening in tandem with next generation sequencing and artificial neural network to rapidly characterize RNA molecules

Roger Cubi[a,c], Farah Bouhedda[a,c], Mayeul Collot[b], Andrey S. Klymchenko[b] and Michael Ryckelynck[a,*]

a. Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France
b. Université de Strasbourg, CNRS, Laboratoire de Bioimagerie et Pathologies, UMR 7021, 67401 Illkirch, France
c. RC and FB equally contributed to this work
*: m.ryckelynck@unistra.fr

Table of content

Table S1…… p 2
Table S2…… p 3
Supplemental Fig. 1……p 4
Supplemental Fig. 2 …….. p 5
Supplemental Fig. 3 …….. p 6
Supplemental Fig. 4 …….. p 7
Supplemental Fig. 5 …….. p 8
Supplemental Fig. 6 …….. p 9
Supplemental Fig. 7…..…….. p 10
Supplemental Fig. 8…..…….. p 11

TABLE S1: μIVC experimental parameters

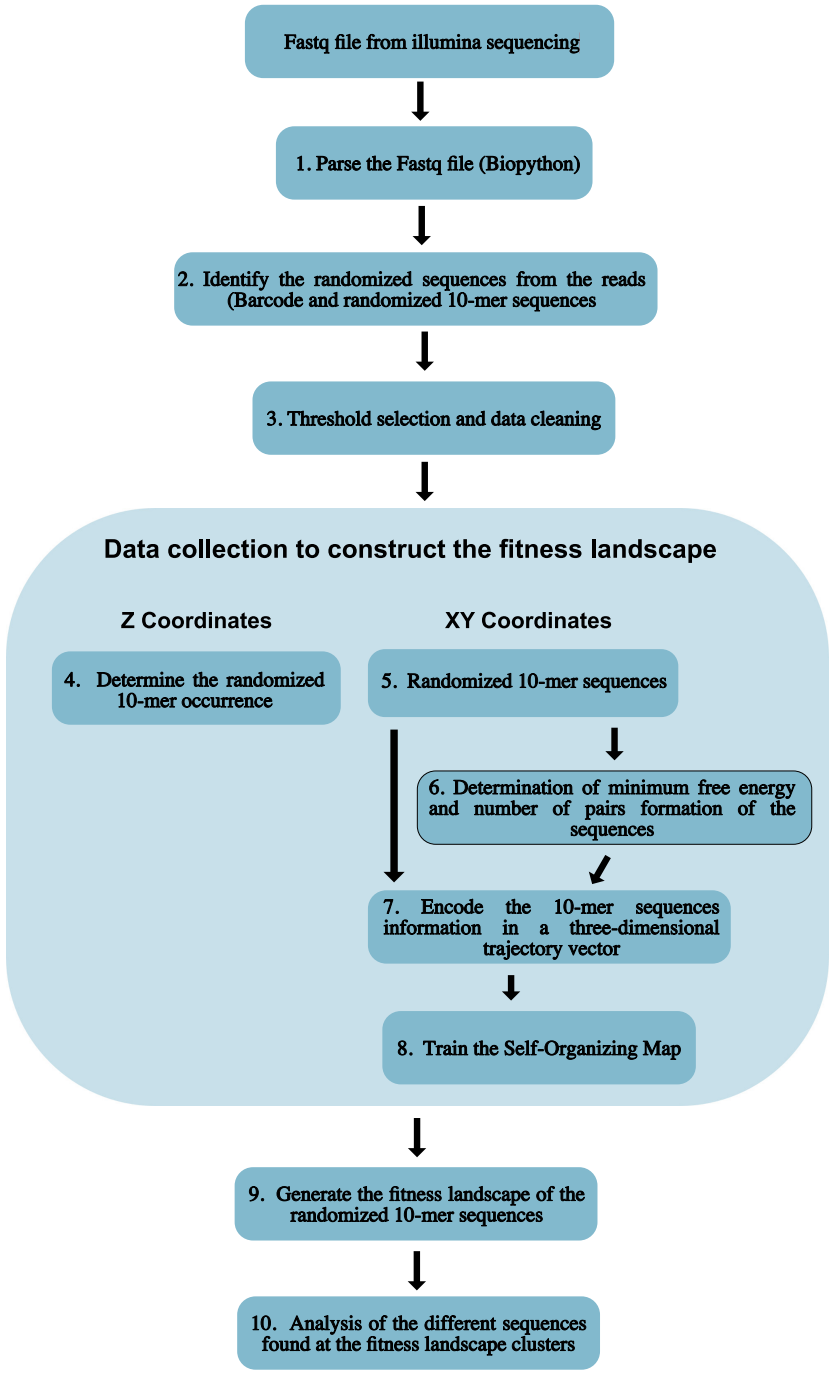
Round	PCR droplets occupancy (%)	λ value[a]	Fusion efficiency (%)	Number of analyzed droplets	Number of sorted droplets
R1	86	2	95	3 875 501	168 805
R2	18	0.2	86	1 906 236	7 120
R3A	7.7	0.08	92	785 958	4 085
R3B	7.7	0.08	92	561 338	1 422

[a] average number of template DNA molecules per droplet.

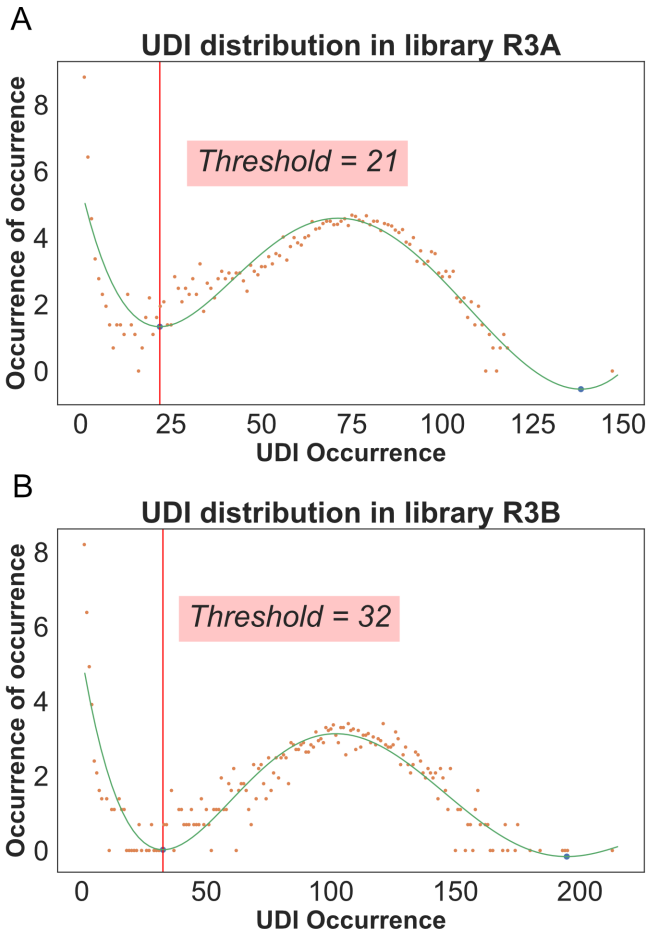
TABLE S2 : Templates and primers

Name	Sequence (5'-3')
Final construct	*TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG*aaNNNNNNNNNNNNNNNNNNaaTATcTAATACGACTCACTATA**GGGAGACAGCTAGAGTAC**GGAACCTCGCTTCGGCGATGATGGAGNNNNNCAAGGTTAACNNNNNCAGGTTCC**GACACGAGCACAGTGTAC***CTGTCTCTTATACACATCTCCGAGCCCACGAGAC*
SRB-2 P3N10	GGAACCTCGCTTCGGCGATGATGGAGNNNNNCAAGGTTAACNNNNNCAGGTTCC
SRB-2 P3N10-ext	**GGGAGACAGCTAGAGTAC**GGAACCTCGCTTCGGCGATGATGGAGNNNNNCAAGGTTAACNNNNNCAGGTTCC**GACACGAGCACAGTGTAC**
Primer 1.A	CTTTAATACGACTCACTATA**GGGAGACAGCTAGAGTAC**
Primer 1.B	**GTACACTGTGCTCGTGTC**
Primer 2.A	*TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG*aaNNNNNNNNNNNNNNNNNNaaTATcTAATACGACTCACTATA**GGGAGACAGCTAGAGTAC**
Primer 3.A	*TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG*
Primer 2.B	*GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG***GTACACTGTGCTCGTGTC**

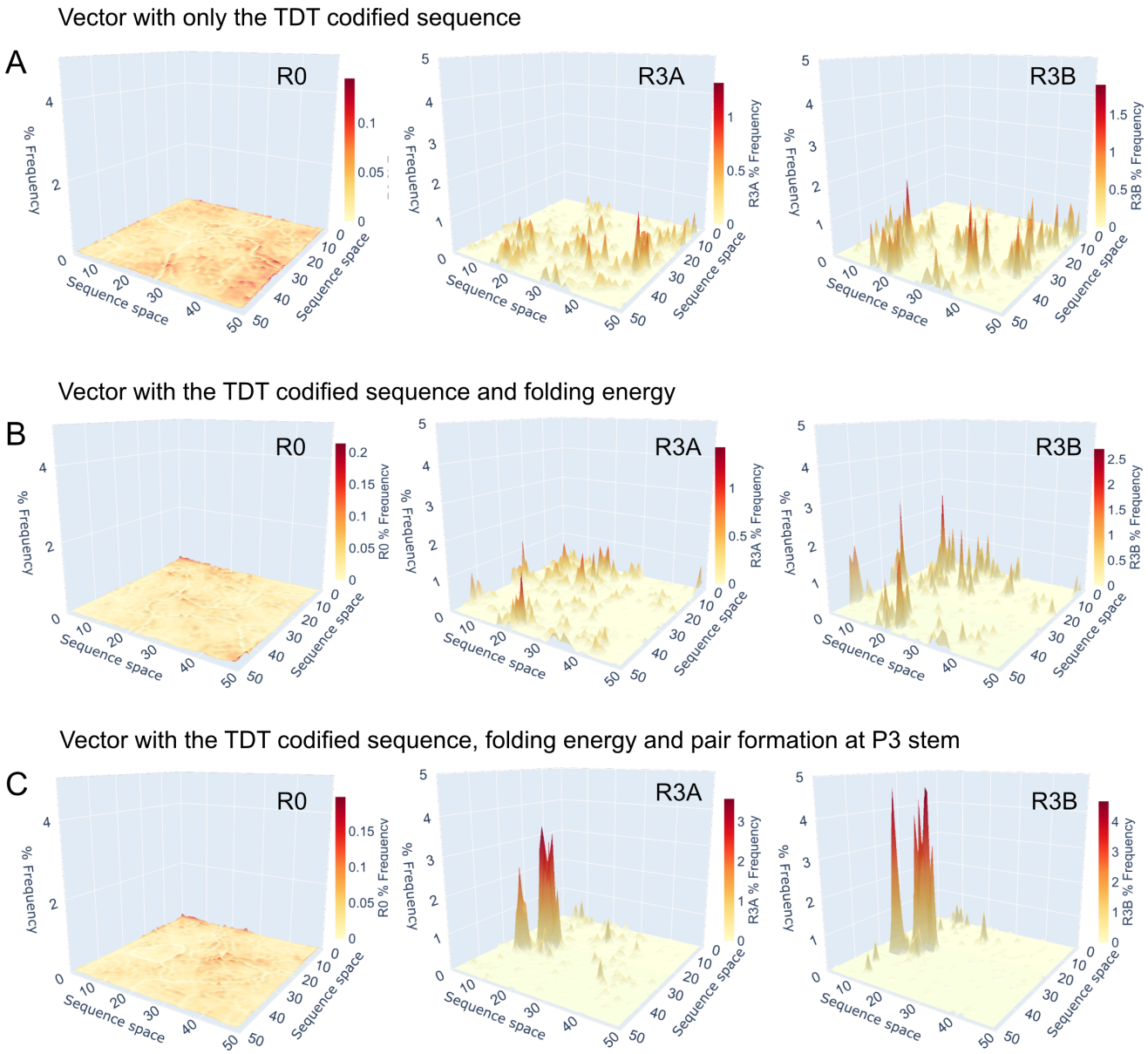
Italicized sequences correspond to Read1 (in 3') or Read2 (in 5') Illumina adaptor sequences
$[N]_{20}$ corresponds to a randomized sequence used as Unique Droplet Identifier (UDI) barcode
The two $[N]_5$ corresponds to the randomized SRB-2 P3 stem
Underlined sequence corresponds to the T7 RNA polymerase promoter
Bolded sequences correspond to PCR primer annealing sites and have a 55° < Tm < 65°



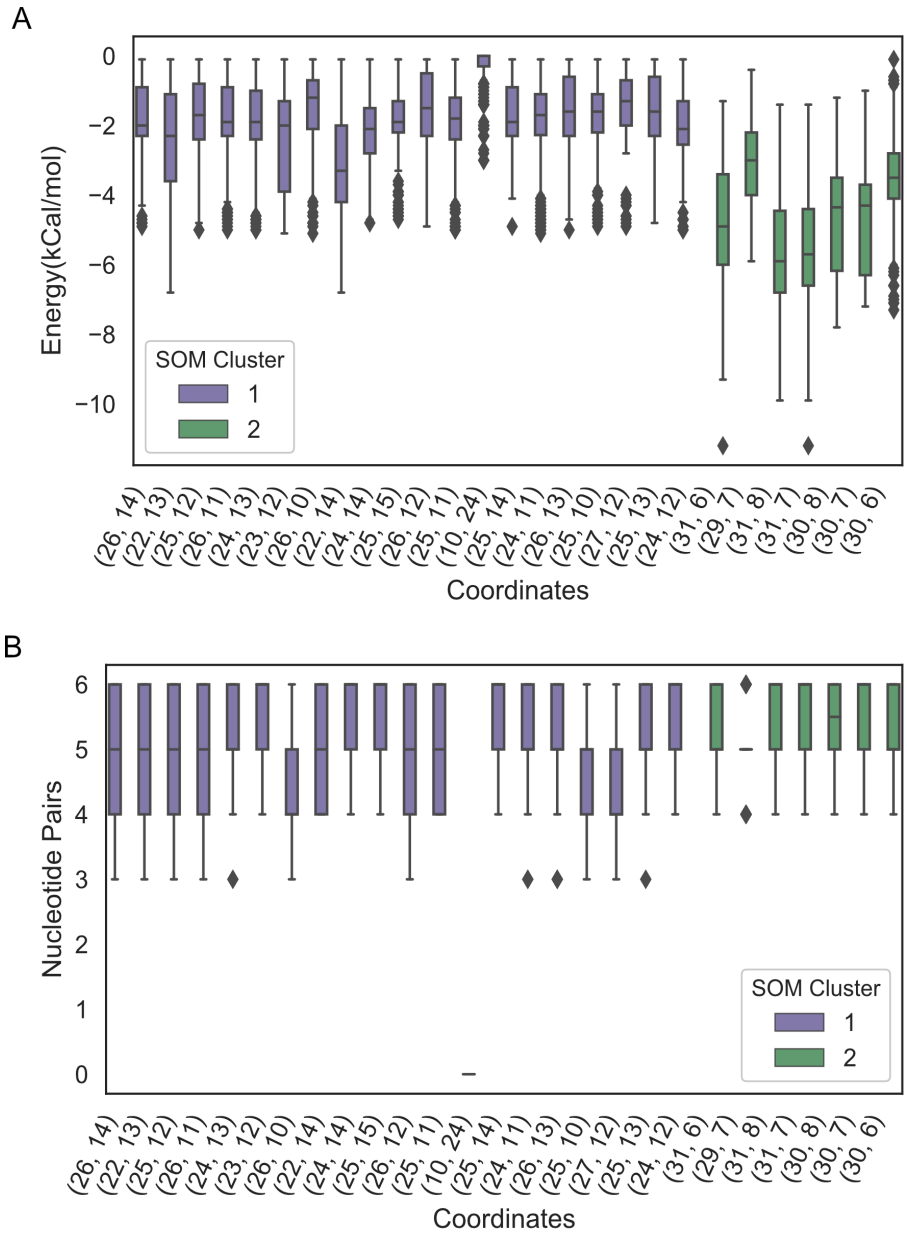
SUPPLEMENTAL FIG. 1: Bioinformatic pipeline used in the study.



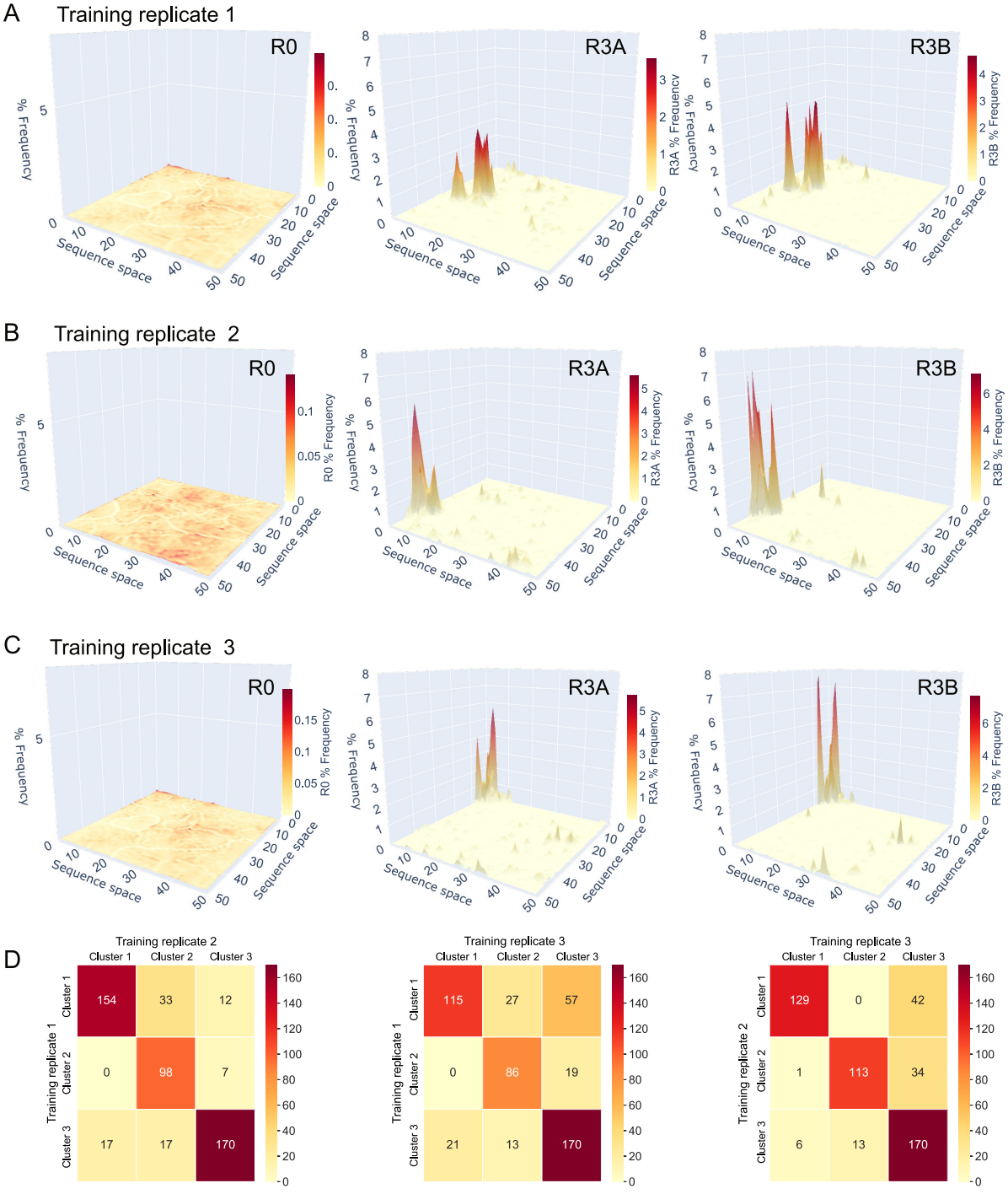
SUPPLEMENTAL FIG. 2: Occurrence distribution of the sequences in R3 rounds. The occurrence of each sequence was measured for each round and a polynomial curve was fit to the data. The first minimum of the curve was used to define an occurrence threshold (red line) used as a cut-off value to filter-out undesired minor sequences likely to result from PCR mutations or sequencing errors.



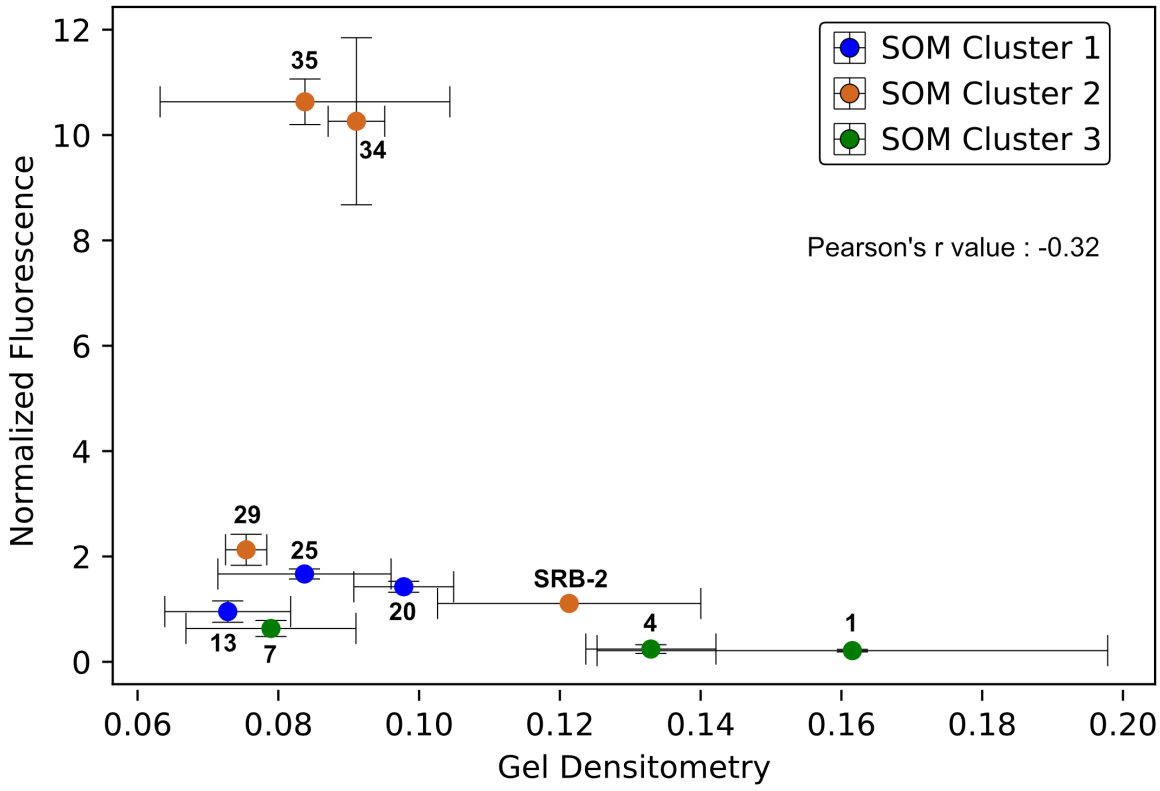
SUPPLEMENTAL FIG. 3: Fitness landscape representations obtained including different information levels to the sequence codifying vector used to generate the SOM map. (*A*) SOM map obtained using the TDT vector representing the sequence. (*B*) SOM map obtained using the TDT vector representing the sequence including its folding energy. (*C*) SOM map obtained using the TDT vector representing the sequence including its folding energy and the number of pairs formed at the P3 stem.



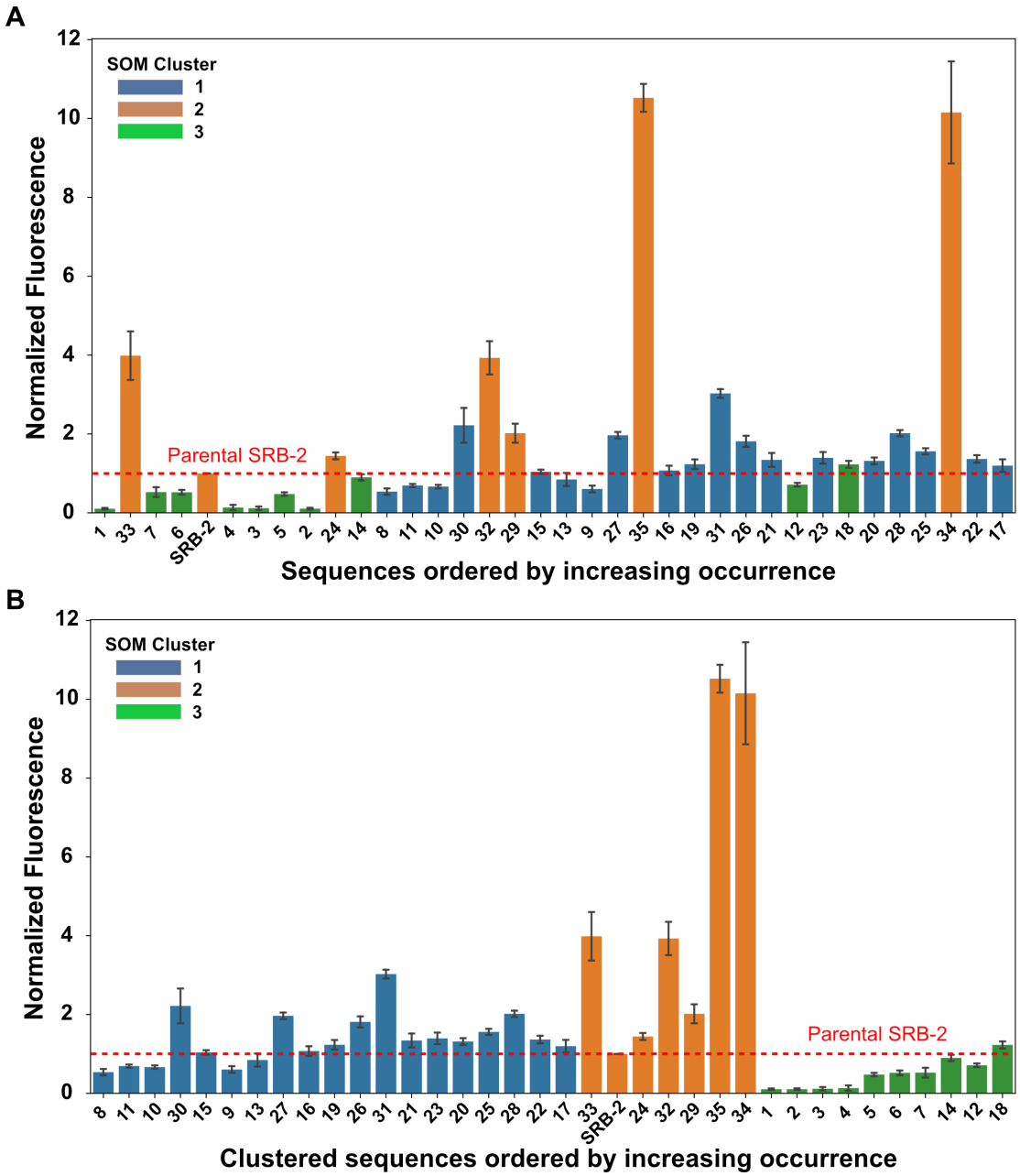
SUPPLEMENTAL FIG. 4: Homogeneity within clusters. (*A*) Folding energy of the sequences found at the different coordinates of SOM clusters 1 and 2 computed using RNAfold program from the ViennaRNA Package. (*B*) Base-pairs formation in the P3-L3 stem loop of the sequences found at the different coordinates of SOM clusters 1 and 2. The number of base pairs formed was extracted from the optimal secondary structure in dot-bracket notation generated by the RNAfold program of the ViennaRNA Package.



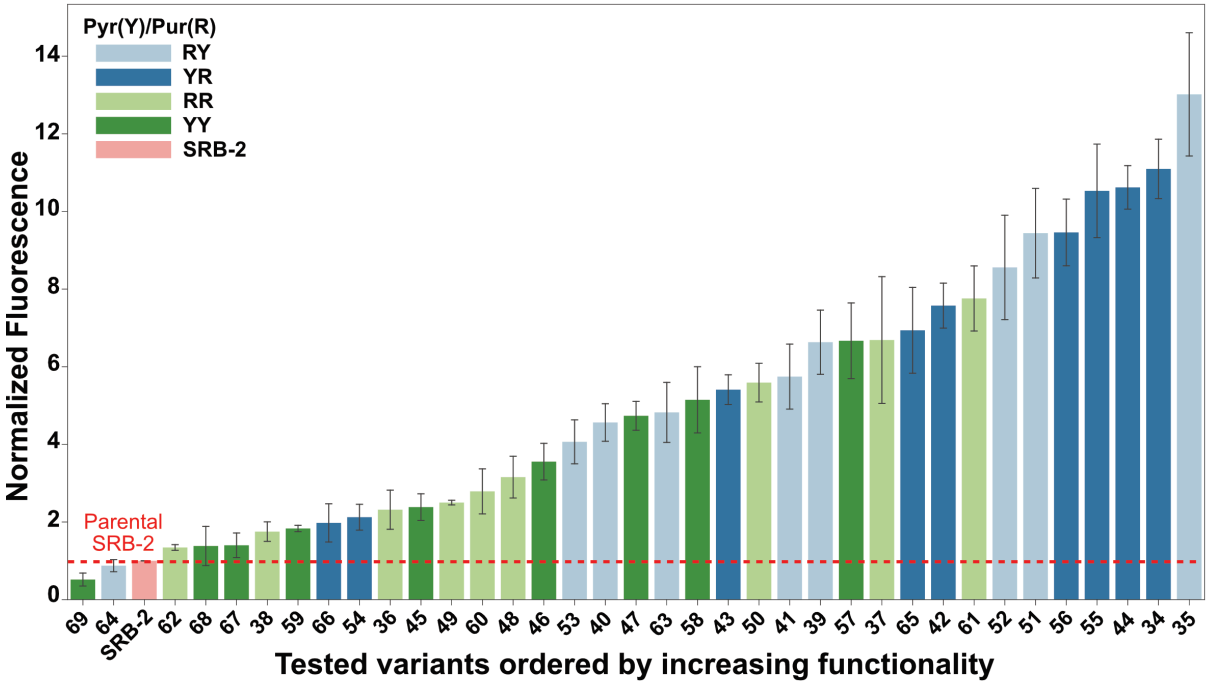
SUPPLEMENTAL FIG. 5: Robustness of the SOM map approach. (*A*, *B* and *C*) 3 SOM maps were generated independently from R3B dataset using the TDT vector representing the sequence including its folding energy and the number of pairs formed at the P3 stem at the different selection rounds. (D) Confusion matrix comparing the different models in relation of the sequences found at the R3B and present in the different identified clusters (Cluster 1, 2 and 3). The good correlation of the different replicates is given by the elevated values found in the diagonal.



SUPPLEMENTAL FIG. 6: Correlation plot between RNA concentration and functionality. Upon functional testing by *in vitro* transcription monitoring in presence of G-561, produced RNAs were loaded on 8 % denaturing gel and their relative concentration was evaluated by gel densitometry measurements. Here, representant of cluster 1 (13, 20 & 25), cluster 2 (SRB-2, 29, 34 & 35) and cluster 3 (1, 4 & 7) were used to realize a correlation plot. The Pearson r value was computed (-0.32) and demonstrated no correlation between functionality and RNA concentration. For both x and y axis, the values are the mean of n = 3 independent experiments. The error bars correspond to ± 1 s.d.



SUPPLEMENTAL FIG. 7: Functional analysis of SRB-2 mutants organized by their occurrence with and without clustering. Different sequences were selected from the self-organizing map (SOM) shown on figure 5. Each construct was *in vitro* transcribed in the presence of 500 nM of Gemini-561 and the fluorescence monitored over time. The fluorescence apparition rate was computed and normalized to that of the parental SRB-2 aptamer. The values are the mean of n = 3 independent experiments. The error bars correspond to ± 1 s.d. The bars are color-coded with respect to the SOM cluster from which the sequence was originally selected from. (A) Sequences are ordered by their increasing occurrence at the end of R3B selection round. (B) Sequences were grouped according to the SOM cluster they originated from prior to being ordered by their increasing occurrence.



Label	P3 Sequence	Label	P3 Sequence	Label	P3 Sequence
SRB-2	5'-AGGCG-3' // 5'-CGCCT-3'	46	5'-GGCTC-3' // 5'-GAGCC-3'	59	5'-GGCTC-3' // 5'-GGGCC-3'
34	5'-GGTAC-3' // 5'-GTACC-3'	47	5'-GGTTC-3' // 5'-GAACC-3'	60	5'-GGGGC-3' // 5'-GCTCC-3'
35	5'-GGACC-3' // 5'-GGTCC-3'	48	5'-GGAGC-3' // 5'-GCTCC-3'	61	5'-GGGAC-3' // 5'-GTTCC-3'
36	5'-GGGGC-3' // 5'-GCCCC-3'	49	5'-GGGGC-3' // 5'-GTCCC-3'	62	5'-GGGGC-3' // 5'-GTTCC-3'
37	5'-GGGAC-3' // 5'-GTCCC-3'	50	5'-GGAAC-3' // 5'-GTTCC-3'	63	5'-GGGTC-3' // 5'-GGCCC-3'
38	5'-GGAGC-3' // 5'-GTTCC-3'	51	5'-GGGCC-3' // 5'-GGTCC-3'	64	5'-GGGTC-3' // 5'-GGTCC-3'
39	5'-GGGCC-3' // 5'-GGCCC-3'	52	5'-GGATC-3' // 5'-GATCC-3'	65	5'-GGTGC-3' // 5'-GCGCC-3'
40	5'-GGGTC-3' // 5'-GACCC-3'	53	5'-GGATC-3' // 5'-GGTCC-3'	66	5'-GGTGC-3' // 5'-GTGCC-3'
41	5'-GGGTC-3' // 5'-GATCC-3'	54	5'-GGCGC-3' // 5'-GTGCC-3'	67	5'-GGTTC-3' // 5'-GAGCC-3'
42	5'-GGCGC-3' // 5'-GCGCC-3'	55	5'-GGTGC-3' // 5'-GCACC-3'	68	5'-GGTTC-3' // 5'-GGACC-3'
43	5'-GGCAC-3' // 5'-GTGCC-3'	56	5'-GGTGC-3' // 5'-GTACC-3'	69	5'-GGTTC-3' // 5'-GGGCC-3'
44	5'-GGTAC-3' // 5'-GTGCC-3'	57	5'-GGTCC-3' // 5'-GGACC-3'		
45	5'-GGCCC-3' // 5'-GGGCC-3'	58	5'-GGTCC-3' // 5'-GGGCC-3'		

SUPPLEMENTAL FIG. 8: Functionality of $_{27}$GGNNC$_{31\text{-}42}$GNNCC$_{46}$ motif variants. The 36 possible sequence permutations allowing the formation of A-U, U-A, G-C, C-G, G.U and U.G base-pairs were individually transcribed in the presence of Gemini 561. The emitted fluorescence was recorded and normalized to that measured with of SRB-2. The nature (purine, R, or pyrimidine, Y) of the consecutive residues at positions 29 and 30 in the motif were color-coded. The measurements are the mean of n = 3 independent experiments and the error bars correspond to ± 1 s.d.