

```
1 import Pkg, Revise; Pkg.activate(Base.current_project())
```



```
Activating project at `/DATA/cossio/SAM/2024/SamApp2024.jl`
```



```
1 import CairoMakie
```

```
1 import CSV
```

```
1 import FASTX
```

```
1 import HDF5
```

```
1 import Infernal
```

```
1 import KernelDensity
```

```
1 import Makie
```

```
1 import RestrictedBoltzmannMachines as RBMs
```

```
1 import Rfam
```

```
1 import SamApp2024
```

```
1 import StatsBase
```

```
1 using BioSequences: LongRNA
```

```
1 using DataFrames: DataFrame
```

```
1 using Distributions: Gamma
```

```
1 using Distributions: logpdf
```

```
1 using Distributions: pdf
```

```
1 using Distributions: Poisson
```

```
1 using LinearAlgebra: Diagonal
```

```
1 using LinearAlgebra: eigen
```

```
1 using Makie: @L_str
```

```
1 using NaNStatistics: nansum
```

```
1 using NaNStatistics: nanmean
```

```
1 using NaNStatistics: nanstd
```

```
1 using Random: bitrand
```

```
1 using RestrictedBoltzmannMachines: free_energy
```

```
1 using Statistics: cor
```

```
1 using Statistics: mean
```

```
1 using StatsBase: countmap
```

```
1 # load SHAPE data
2 shape_data_045 = SamApp2024.load_shapemapper_data_pierre_demux_20230920(; demux=true);
```

```
1 # split rep0 from rep4+5
2 shape_data_rep0 = SamApp2024.select_conditions_20231002(shape_data_045,
  filter(endswith("_rep0"), shape_data_045.conditions));
```

```
1 # split rep0 from rep4+5
2 shape_data_rep45 = SamApp2024.select_conditions_20231002(shape_data_045,
  filter(endswith("_rep45"), shape_data_045.conditions));
```

```
1 conds_sam_rep0 = identity.(indexin(["SAMAP_1M7_0-1SAM_5Mg_T30C_rep0", "SAMAP_1M7_0-
  5SAM_5Mg_T30C_rep0", "SAMAP_1M7_1SAM_5Mg_T30C_rep0"], shape_data_rep0.conditions));
```

```
1 conds_mg_rep0 = identity.(indexin(["SAMAP_1M7_noSAM_5Mg_T30C_rep0"],
  shape_data_rep0.conditions));
```

```
1 conds_30C_rep0 = identity.(indexin(["SAMAP_1M7_noSAM_noMg_T30C_rep0"],
  shape_data_rep0.conditions));
```

```
1 conds_sam_rep45 = identity.(indexin(["SAMAP_1M7_0-1SAM_5Mg_T30C_rep45",
  "SAMAP_1M7_1SAM_5Mg_T30C_rep45"], shape_data_rep45.conditions));
```

```
1 conds_mg_rep45 = identity.(indexin(["SAMAP_1M7_noSAM_5Mg_T30C_rep45"],
  shape_data_rep45.conditions));
```

```
1 conds_30C_rep45 = identity.(indexin(["SAMAP_1M7_noSAM_noMg_T30C_rep45"],
  shape_data_rep45.conditions));
```

```
1 @show conds_sam_rep0 conds_mg_rep0 conds_30C_rep0;
```

```
> conds_sam_rep0 = [1, 2, 3]
   conds_mg_rep0 = [4]
   conds_30C_rep0 = [5]
```

```
1 @show conds_sam_rep45 conds_mg_rep45 conds_30C_rep45;
```

```
> conds_sam_rep45 = [1, 2]
   conds_mg_rep45 = [4]
   conds_30C_rep45 = [6]
```

```
► (bps = [1, 2, 3, 4, 5, 6, 7, 8, 13, ... more ,108], nps = [9, 10, 11, 12, 18, 19, 20, 24, 32, ... more
```

```
1 (; bps, nps, pks) = SamApp2024.RF00162_sites_paired()
```

```
rbm_seqs =
```

```
► [207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 22
```

```
1 rbm_seqs = findall(shape\_data\_045.aptamer_origin .== "RF00162_syn_rbm")
```

```
inf_seqs = ► [258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 291]
```

```
1 inf_seqs = findall(shape\_data\_045.aptamer_origin .== "RF00162_syn_inf")
```

```
full_seqs =
```

```
► [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ... more ,46, 47, 48, 49, 50,
```

```
1 full_seqs = findall(shape\_data\_045.aptamer_origin .== "RF00162_full30")
```

```
seed_seqs =
```

```
► [56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ... more ,197, 198,
```

```
1 seed_seqs = findall(shape\_data\_045.aptamer_origin .== "RF00162_seed70")
```

```
1 nat_seqs = full\_seqs ∪ seed\_seqs;
```

```
1 aptamer_rbm_energies = [  
2     ismissing(seq) ? missing :  
3     free_energy(SamApp2024.rbm2022(), SamApp2024.onehot(LongRNA{4}(seq)))  
4     for seq in shape\_data\_045.aligned_sequences  
5 ];
```

```
1 _rbmlo = rbm\_seqs ∩ findall((!ismissing).(aptamer\_rbm\_energies) .&&  
(aptamer\_rbm\_energies .< -300));
```

```
1 _rbmhi = rbm\_seqs ∩ findall((!ismissing).(aptamer\_rbm\_energies) .&&  
(aptamer\_rbm\_energies .> -300));
```

```
1 ΔR_sam = (  
2     nanmean(shape\_data\_rep0.shape_reactivities[:, :, conds\_sam\_rep0]; dim=3) .-  
3     shape\_data\_rep0.shape_reactivities[:, :, only(conds\_mg\_rep0)]  
4 );
```

```
ΔR_sam_avg_seed =
```

```
► [NaN, NaN, -0.142272, -0.248336, -0.0404359, -0.282604, 0.110346, -0.555953, -0.194717, -1.3511
```

```
1 ΔR_sam_avg_seed = nanmean(ΔR_sam[:, seed\_seqs]; dim=2)
```

```
1 ΔR_sam_std_seed = nanstd(ΔR_sam[:, seed\_seqs]; dim=2);
```

```
ΔR_sam_avg_full =
```

```
► [NaN, NaN, -0.0681035, 0.00266552, -0.476075, -0.335042, 0.0248142, -0.869353, -0.472555, -1.71
```

```
1 ΔR_sam_avg_full = nanmean(ΔR_sam[:, full\_seqs]; dim=2)
```

```
1 ΔR_sam_std_full = nanstd(ΔR_sam[:, full\_seqs]; dim=2);
```

```
 $\Delta R_{sam\_avg\_rbmlo} =$ 
```

```
► [NaN, NaN, 0.0653607, -0.464177, 0.448063, -0.0677384, 0.407376, -0.0844439, -0.513369, -0.8908
```

```
1  $\Delta R_{sam\_avg\_rbmlo} = \text{nanmean}(\Delta R_{sam}[:, \text{\_rbmlo}]; \text{dim}=2)$ 
```

```
1  $\Delta R_{sam\_std\_rbmlo} = \text{nanstd}(\Delta R_{sam}[:, \text{\_rbmlo}]; \text{dim}=2);$ 
```

```
 $\Delta R_{sam\_avg\_rbmhi} =$ 
```

```
► [NaN, NaN, 0.242838, -0.0336922, 0.347019, 0.224733, 0.502098, -0.378962, 0.147099, -0.510053, 0.102
```

```
1  $\Delta R_{sam\_avg\_rbmhi} = \text{nanmean}(\Delta R_{sam}[:, \text{\_rbmhi}]; \text{dim}=2)$ 
```

```
1  $\Delta R_{sam\_std\_rbmhi} = \text{nanstd}(\Delta R_{sam}[:, \text{\_rbmhi}]; \text{dim}=2);$ 
```

```
 $\Delta R_{sam\_avg\_inf} =$ 
```

```
► [NaN, NaN, -0.148521, -0.0592776, 0.344908, -0.0223373, -0.0172388, 0.155964, 0.0645069, -0.102
```

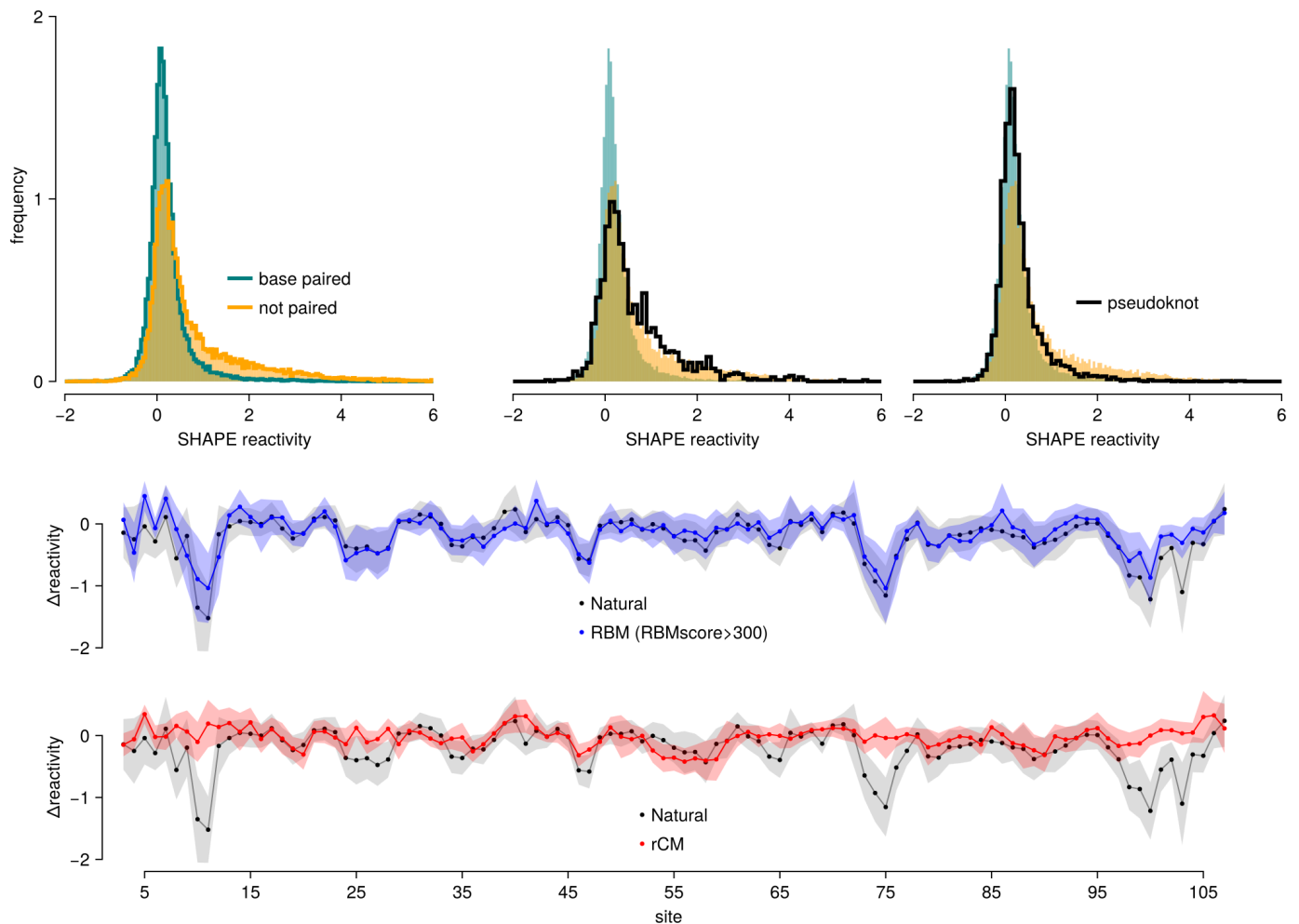
```
1  $\Delta R_{sam\_avg\_inf} = \text{nanmean}(\Delta R_{sam}[:, \text{inf\_seqs}]; \text{dim}=2)$ 
```

```
1  $\Delta R_{sam\_std\_inf} = \text{nanstd}(\Delta R_{sam}[:, \text{inf\_seqs}]; \text{dim}=2);$ 
```

```
1  $\text{bps\_reactivities\_rep0} = \text{shape\_data\_rep0.shape\_reactivities}[\text{bps}, \text{seed\_seqs},$   
 $\text{conds\_sam\_rep0}];$ 
```

```
1  $\text{nps\_reactivities\_rep0} = \text{shape\_data\_rep0.shape\_reactivities}[\text{nps}, \text{seed\_seqs},$   
 $\text{conds\_sam\_rep0}];$ 
```

```
1  $\text{all\_reactivities\_rep0} = \text{shape\_data\_rep0.shape\_reactivities}[:, \text{nat\_seqs}, \text{conds\_sam\_rep0}];$ 
```



```

1 let fig = Makie.Figure()
2 ax = Makie.Axis(
3     fig[1,1], width=300, height=300, xlabel="SHAPE reactivity", ylabel="frequency",
4     xgridvisible=false, ygridvisible=false, xticks=-2:2:6, yticks=0:2,
5     xtrimspine=true, ytrimspine=true
6 )
7 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[bps,
8     nat_seqs, conds_sam_rep0])), normalization=:pdf, bins=-2:0.05:6, color=(teal, 0.5),
9     gap=-0.01)
10 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[nps,
11     nat_seqs, conds_sam_rep0])), normalization=:pdf, bins=-2:0.05:6, color=(orange,
12     0.5), gap=-0.01)
13 Makie.stepphist!(ax, filter(x -> -2 < x < 6,
14     vec(shape_data_rep0.shape_reactivities[bps, nat_seqs, conds_sam_rep0])), label="base
15     paired", normalization=:pdf, bins=-2:0.05:6, linewidth=3, color=:teal)
16 Makie.stepphist!(ax, filter(x -> -2 < x < 6,
17     vec(shape_data_rep0.shape_reactivities[nps, nat_seqs, conds_sam_rep0])), label="not
18     paired", normalization=:pdf, bins=-2:0.05:6, linewidth=3, color=:orange)
19 Makie.xlims!(-2.2, 6)
20 Makie.ylims!(-0.07, 2)
21 #Makie.axislegend(ax, framevisible=false, patchlabelgap=3, position=(-0.02, 1))
22 Makie.axislegend(ax, position=(0.7, 0.2), framevisible=false)
23 Makie.hidespines!(ax, :t, :r)
24
25 _dummy_ax = Makie.Axis(fig[1,2], width=20, xgridvisible=false, ygridvisible=false)
26 Makie.hidespines!(_dummy_ax, :t, :b, :r, :l)
27 Makie.hidexdecorations!(_dummy_ax)
28 Makie.hideydecorations!(_dummy_ax)
29
30 ax = Makie.Axis(fig[1,3], width=300, height=300, xlabel="SHAPE reactivity",
31     ylabel="frequency", xgridvisible=false, ygridvisible=false, xticks=-2:2:6,

```

```

yticks=0:2, xtrimspine=true, ytrimspine=true)
21 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[bps,
nat_seqs, conds_sam_rep0])), label="b.p.", normalization=:pdf, bins=-2:0.05:6, color=
(:teal, 0.5), gap=-0.01)
22 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[nps,
nat_seqs, conds_sam_rep0])), label="n.p.", normalization=:pdf, bins=-2:0.05:6, color=
(:orange, 0.5), gap=-0.01)
23 Makie.stephist!(ax, filter(x -> -2 < x < 6,
vec(shape_data_rep0.shape_reactivities[pks, nat_seqs, conds_mg_rep0])),
label="p.k.", normalization=:pdf, bins=-2:0.1:6, linewidth=3, color=:black)
24 Makie.xlims!(-2.2, 6)
25 Makie.ylims!(-0.07, 2)
26 Makie.hidespines!(ax, :t, :r, :l)
27 Makie.hideydecorations!(ax)
28
29 ax = Makie.Axis(fig[1,4], width=300, height=300, xlabel="SHAPE reactivity",
xgridvisible=false, ygridvisible=false, xticks=-2:2:6, yticks=0:2, xtrimspine=true,
ytrimspine=true)
30 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[bps,
nat_seqs, conds_sam_rep0])), normalization=:pdf, bins=-2:0.05:6, color=(teal, 0.5),
gap=-0.01)
31 Makie.hist!(ax, filter(x -> -2 < x < 6, vec(shape_data_rep0.shape_reactivities[nps,
nat_seqs, conds_sam_rep0])), normalization=:pdf, bins=-2:0.05:6, color=(orange,
0.5), gap=-0.01)
32 Makie.stephist!(ax, filter(x -> -2 < x < 6,
vec(shape_data_rep0.shape_reactivities[pks, nat_seqs, conds_sam_rep0])),
label="pseudoknot", normalization=:pdf, bins=-2:0.1:6, linewidth=3, color=:black)
33 Makie.xlims!(-2.2, 6)
34 Makie.ylims!(-0.07, 2)
35 Makie.axislegend(ax, position=(0.7, 0.2), framevisible=false)
36 Makie.hidespines!(ax, :t, :r, :l)
37 Makie.hideydecorations!(ax)
38
39 _xs = 3:107
40
41 ax = Makie.Axis(fig[2,:], width=900, height=150, xticks=5:10:108, yticks=-2:1:1,
xgridvisible=false, ygridvisible=false, ylabel="Δreactivity", xtrimspine=true,
ytrimspine=true)
42
43 Makie.band!(ax, _xs, (ΔR_sam_avg_seed - ΔR_sam_std_seed/2)[_xs], (ΔR_sam_avg_seed +
ΔR_sam_std_seed/2)[_xs], markersize=5, color=:gray, 0.25))
44 Makie.lines!(ax, _xs, ΔR_sam_avg_seed[_xs], linewidth=1, color=:gray)
45 Makie.scatter!(ax, _xs, ΔR_sam_avg_seed[_xs], markersize=5, color=:black,
label="Natural")
46
47 Makie.band!(ax, _xs, (ΔR_sam_avg_rbmlo - ΔR_sam_std_rbmlo/2)[_xs], (ΔR_sam_avg_rbmlo
+ ΔR_sam_std_rbmlo/2)[_xs], markersize=5, color=:blue, 0.25))
48 Makie.lines!(ax, _xs, ΔR_sam_avg_rbmlo[_xs], linewidth=1, color=:blue)
49 Makie.scatter!(ax, _xs, ΔR_sam_avg_rbmlo[_xs], markersize=5, color=:blue, label="RBM
(RBMscore>300)")
50 Makie.axislegend(ax, position=(0.5, 0), framevisible=false, patchlabelgap=-3)
51 Makie.xlims!(1, 108)
52 Makie.hidespines!(ax, :t, :r, :b)
53 Makie.hidexdecorations!(ax)
54
55 ax = Makie.Axis(fig[3,:], width=900, height=150, xticks=5:10:108, yticks=-2:1:1,
xgridvisible=false, ygridvisible=false, xlabel="site", ylabel="Δreactivity",
xtrimspine=true, ytrimspine=true)
56
57

```

```

58 Makie.band!(ax, _xs, ( $\Delta R_{sam\_avg\_seed} - \Delta R_{sam\_std\_seed}/2$ )[_xs], ( $\Delta R_{sam\_avg\_seed} +$ 
59  $\Delta R_{sam\_std\_seed}/2$ )[_xs], markersize=5, color=:gray, 0.25))
60 Makie.lines!(ax, _xs,  $\Delta R_{sam\_avg\_seed}$ [_xs], linewidth=1, color=:gray)
61 Makie.scatter!(ax, _xs,  $\Delta R_{sam\_avg\_seed}$ [_xs], markersize=5, color=:black,
    label="Natural")

62 Makie.band!(ax, _xs, ( $\Delta R_{sam\_avg\_inf} - \Delta R_{sam\_std\_inf}/2$ )[_xs], ( $\Delta R_{sam\_avg\_inf} +$ 
63  $\Delta R_{sam\_std\_inf}/2$ )[_xs], markersize=5, color=:red, 0.25))
64 Makie.lines!(ax, _xs,  $\Delta R_{sam\_avg\_inf}$ [_xs], linewidth=1, color=:red)
65 Makie.scatter!(ax, _xs,  $\Delta R_{sam\_avg\_inf}$ [_xs], markersize=5, color=:red, label="rCM")
66 Makie.axislegend(ax, position=(0.5, 0), framevisible=false, patchlabelgap=-3)
67 Makie.hidespines!(ax, :t, :r)
68 Makie.xlims!(1, 108)

69
70 # Makie.Label(fig[1,1][1,1,Makie.TopLeft()], "A", font=:bold, padding=(0,0,10,10))
71 # Makie.Label(fig[1,2][1,1,Makie.TopLeft()], "B", font=:bold, padding=(0,0,10,10))
72 # Makie.Label(fig[1,3][1,1,Makie.TopLeft()], "C", font=:bold, padding=(0,0,10,10))
73 # Makie.Label(fig[2,:][1,1,Makie.TopLeft()], "D", font=:bold, padding=(0,0,0,0))
74 # Makie.Label(fig[3,:][1,1,Makie.TopLeft()], "E", font=:bold, padding=(0,0,0,0))
75
76 Makie.resize_to_layout!(fig)
77 #Makie.save("/workspaces/SamApp.jl/notebooks/2024-03-14 New paper
    figures/Figures/SHAPE reactivities.pdf", fig)
    fig

```