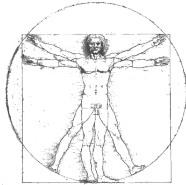


Structure-based Whole Genome Realignment Reveals Many Novel Non-coding RNAs

Sebastian Will, Michael Yu, Bonnie Berger

Previous Genome-wide Screens for Structural ncRNA



RNAz Washietl, *Nat Biotech*, 2005

Rose, *BMC Genomics*, 2007

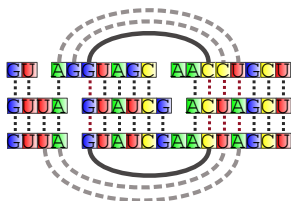
EvoFold Pedersen, *PLoS CB*, 2006

Bradley, *PLoS One*, 2006

- predict thousands of putative structural ncRNAs
- energy model + comparative information:
 WGAs, *conserved stable structure*
- for efficiency: only *correctly aligned* structure

Challenges for ncRNA Screens

- WGA tools regularly misalign structural RNA



↓ RNAz

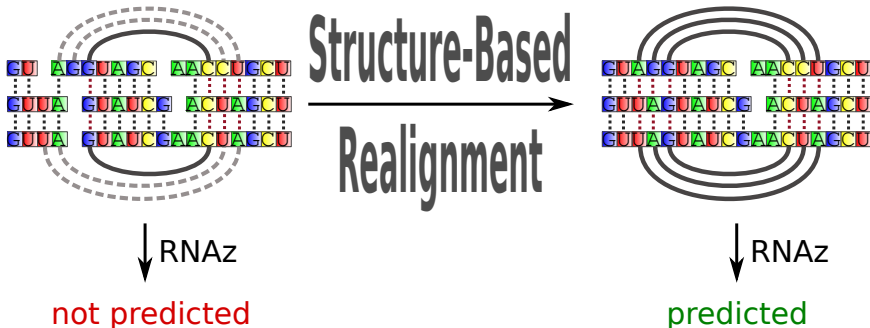
not predicted

Conjecture:

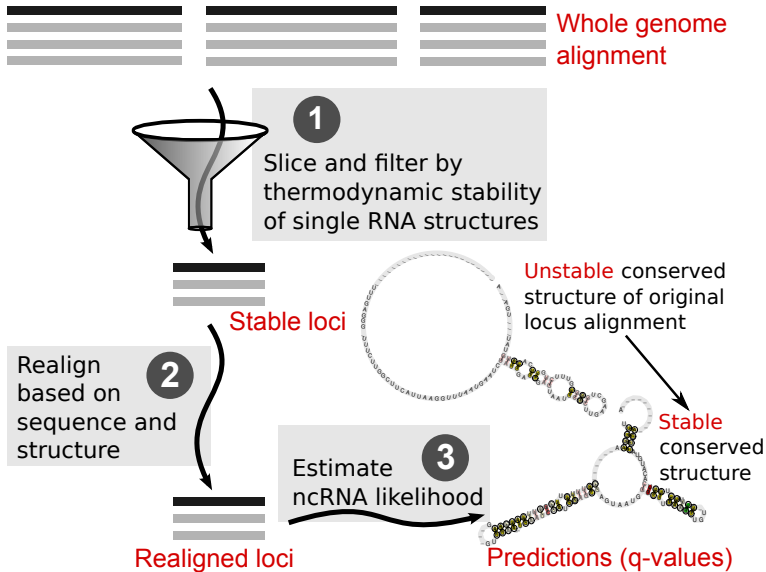
- underestimate true number of structural ncRNAs*
- biased for 'easy targets'*

Revealing Hidden ncRNA

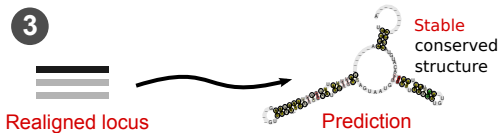
Idea: patch mistakes by structure-based realignment



The REAPR Pipeline



Estimate ncRNA Likelihood



Apply RNAz

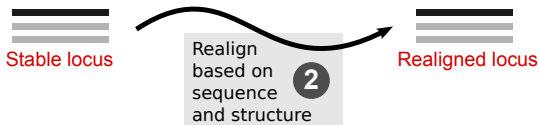


recall: RNAz sees only *correctly aligned* conserved structure

 Washietl *et al.* PNAS, 2005.

 Gruber *et al.* PSB 2010. (*RNAz 2.0*)

Making ncRNAs Visible for De Novo Predictors

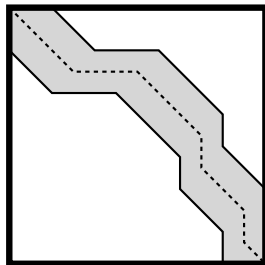
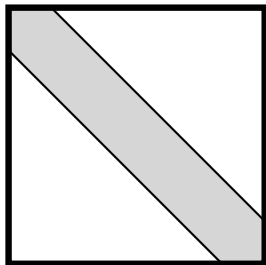
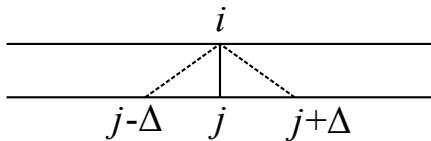


Apply “gold-standard” of structure-based RNA alignment: simultaneous alignment and folding

- use Sankoff-algorithm ? $O(n^6)$
- use fast and accurate Sankoff-variant ?
e.g., *LocARNA* [Will et al., 2007], $O(n^4)$

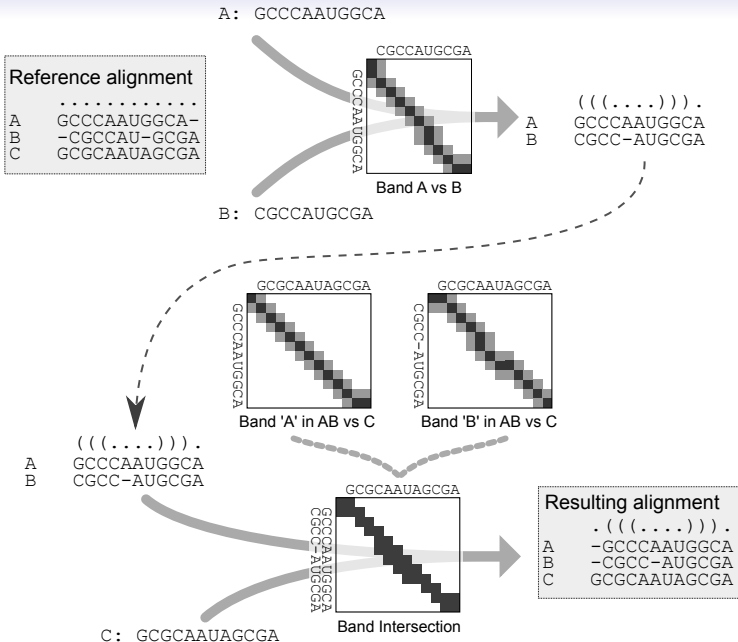
Efficient Realignment: Novel Banding Algorithm

novel **banding**: compute **realignment** *within limited deviation* Δ from a reference alignment, e.g., $\Delta = 20$ (± 20 nt)



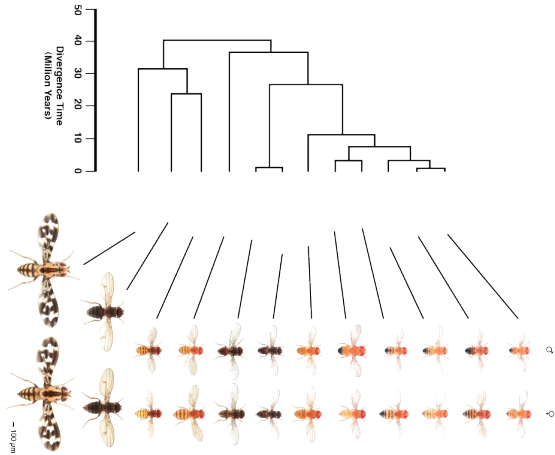
main technical challenge: banding for *multiple* alignment

Banding for Progressive Multiple Alignment



Applying REAPR: Reaping ncRNAs in Fly

- First computational screen of a eukaryotic genome that corrects structural misalignments
- control *w/o realignment*

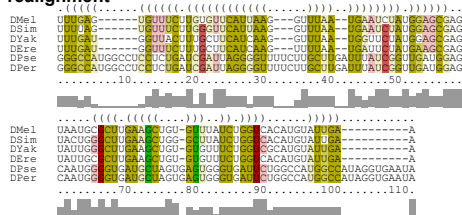


A Novel Putative Motif in the Long ncRNA roX1

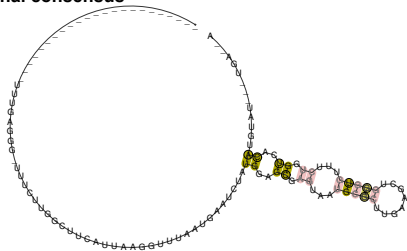
original alignment



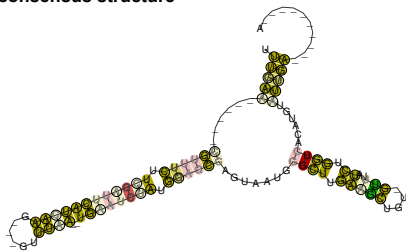
realignment



original consensus

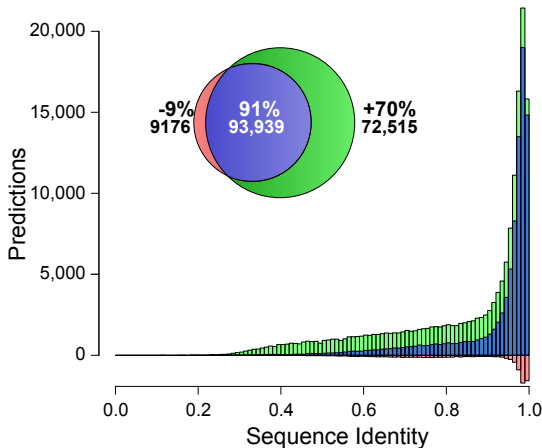


new consensus structure

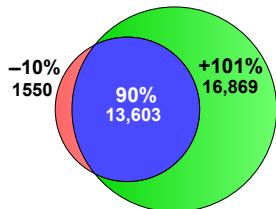


Novel Predictions due to Realignment

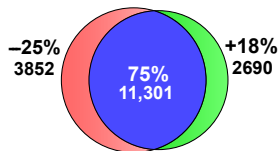
all predictions



D. melanogaster
predictions



REAPR with LocARNA

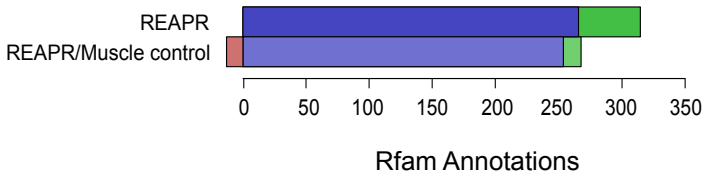


REAPR with Muscle

- = common predictions by REAPR and control
- = novel predictions by REAPR
- = predictions only by control

Sensitivity for Rfam ncRNAs in *D. melanogaster*

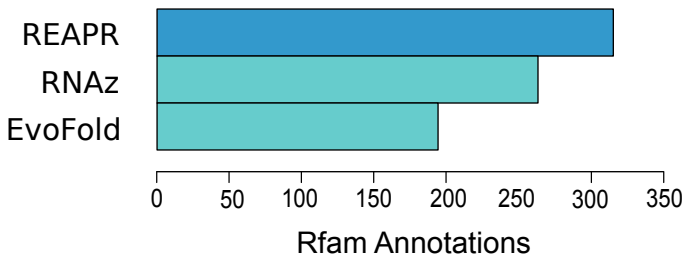
REAPR vs. Controls



- = common predictions of REAPR and control
- = novel predictions by REAPR
- = predictions only by control

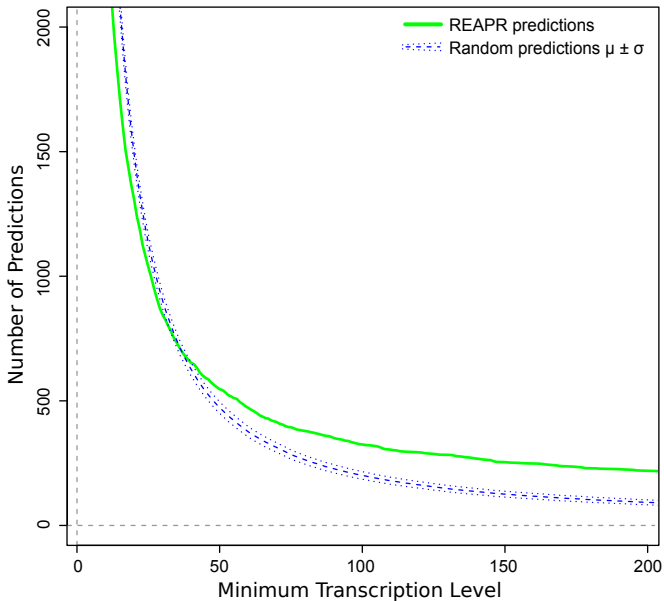
Sensitivity for Rfam ncRNAs in *D. melanogaster*

REAPR vs. Previous Screens





RNA-seq (modENCODE) Confirms Predictions

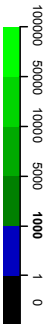




Combining Fly Embryo RNA-seq and REAPR



CH2R:18591419-18581570
 CH3:87474487-875006
 CH3:14607488-1607590
 CH3:16976868-16975088
 CH3:13016676-13016828
 CH2R:3072472-3072780
 CH3:3074339-3074497
 CH3:16400618-16400867
 CH3:16977850-16978043
 CH2R:4683224-16838114
 CH3:19987925-14858113
 CH2R:9760789-9738683
 CH2R:16841577-16841510
 CH3:19889267-19889002
 CH3:16874867-16878702
 CH3:16453519-16455648
 CHX:139078-1339216
 CH4:88830-87021
 CH3:2914276-2914416
 CH3R:19185518-19185701
 CH3:5688477-25688608
 CH3:1668815-2668164
 CHX:54878-548804
 CH3:1070745-1070315
 CH3:1091741-501489299
 CH2R:1091-87524
 CH3:6839179-16839292
 CH3:459619-454778
 CH3:1947374-17947524
 CH2:9106897-1122408
 CH2:2404788-2407294
 CH3:1676588-16765664
 CH3R:1122272-1122408
 CH3:5409897-15470039
 CH3:3018688-30186958
 CH3:4092845-21091456
 CH3:210689-1439886
 CH3:2454557-23484663
 CH3:1681880-16820460
 CH2R:16842771-16852966
 CH2R:2101099-2101299
 CH2R:103301-103415
 CH3:504999-5048090
 CH3:429029-2399881
 CH3:1092477-1092607
 CH3:1483634-14836740
 CH3R:859589-1859746
 CH3R:1047878-1048035
 CH3R:442018-4423173
 CH2R:8312898-18312807
 CH4:1088013-1088173



CH3R:7424276-27424475
 CH3R:106330-306910
 CH3R:431008-21361044
 CH2R:500989-2960980
 CH3R:3579174-5744934
 CH3R:311609-1331888
 CH3R:3484246-3484474
 CH3R:502471-1020926
 CH3R:1004479-5804622
 CH2R:410009-741012
 CH3R:4942577-14424204
 CH3R:1988853-704840709
 CH2R:1021028-1021248
 CH3R:1749077-17521092
 CH2R:4973820-9074020
 CH2R:3657839-3657931
 CH2:1894175-8894283
 CHX:619903-620074
 CH2R:0166644-0136796
 CHX:6376887-6376991
 CH3R:815446-815549
 CH2R:16979820
 CH2R:0188339-10388305
 CH2R:1608823-16088803
 CH2R:1968982-14988909
 CH2R:3215335-12315334
 CH2R:7715032-7715192
 CH2R:16979518-16979650
 CHX:16979184-16979334
 CH3R:15015907-15016009
 CH3R:3309517-3309757
 CH3R:8881298-18881726
 CH3:1680404-168178
 CH4:200-80589
 CH3R:3271800-1409890
 CH3R:1955589-19555964
 CH2R:21202607-21202984
 CH2R:1668100-21668302
 CH2R:9806683-19806827
 CH3R:603862-6648332
 CH3R:558923-6648394
 CH2R:323223-2372785
 CH3R:603340-1603385
 CH3R:197534-1076471
 CH3R:4024931-6033471
 CH3R:14236-142310
 CH3R:44748-147587
 CH3R:44509-245488
 CH3R:1884781-8488815
 CH2R:408020-7409173

Conclusions

- **REAPR**: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Conclusions

- *REAPR*: “Realignment for Prediction of ncRNAs”
- *First* genome-wide ncRNA screen that *corrects structural misalignments*
- *Twice as many D. melanogaster ncRNAs* while *less than doubling run-time*
- *Novel banding method* for fast multiple realignment implemented in the RNA alignment tool LocARNA
- Verified with *annotations and modEncode RNA-seq*
- REAPR + RNA-seq combined results in interesting structural ncRNAs
- *Future challenge*: detect non-locally misaligned RNAs

Acknowledgements

Bonnie Berger



Michael Yu



Rolf Backofen



German Research Foundation **DFG**