

# Why is the computational method section so important

2021.03.02

# The most cited ( $n = 27,425$ ) paper in *Bioinformatics*

The screenshot shows the Bioinformatics journal website. At the top, there's a navigation bar with links for 'Issues', 'Advance articles', 'Submit', 'Purchase', 'Alerts', and 'About'. On the right of the navigation bar is a search bar with 'All Bioinformatics' dropdown and a magnifying glass icon. To the right of the search bar is the ISCB logo (International Society for Computational Biology).

The main content area features a thumbnail for 'Volume 25, Issue 14, 15 July 2009' with a DNA helix graphic. Below the thumbnail, the title of the article is displayed: 'Fast and accurate short read alignment with Burrows–Wheeler transform' by Heng Li and Richard Durbin. The publication details show it was published in Bioinformatics, Volume 25, Issue 14, 15 July 2009, Pages 1754–1760, with the DOI <https://doi.org/10.1093/bioinformatics/btp324>. The article was published on 18 May 2009.

Below the article title, there are download options: PDF, Split View, Cite, Permissions, and Share. The abstract section starts with the heading 'Abstract' and describes the motivation for developing a fast and accurate read alignment tool, mentioning the need for handling short reads from a single individual and the challenge of aligning longer reads with indels. It also discusses the speed of the tool and its scalability.

The sidebar on the left contains a table of contents for the issue, including sections like 'Introduction', 'Methods', 'Results', 'Discussion', 'Acknowledgements', 'References', and 'Author notes'. Navigation arrows at the bottom of the sidebar allow users to 'Previous' or 'Next' through the issue.

On the right side of the page, there's a 'View Metrics' button with a circular icon containing the number '49'. Below this are sections for 'Email alerts' (with links to 'Article activity alert', 'Advance article alerts', and 'New issue alert') and a promotional box for 'Exclusive offers and updates from Oxford Academic'.

At the bottom of the page, there's a section titled 'Related articles' with links to other papers: 'SHOGUN: a modular, accurate and scalable framework for microbiome quantification' by Benjamin Hillmann et al., 'An improved encoding of genetic variation in a Burrows–Wheeler transform' by Thomas Büchler et al., and 'LordFAST: sensitive and Fast Alignment Search Tool for LOnG noisy Read sequencing Data' by Ehsan Haghshenas et al.

# Format for research papers in *Bioinformatics*

- Title
- Authors and affiliations
- Abstract

The very first (and maybe only) part readers will read and so must be concise and catchy

# Format for research papers in *Bioinformatics*

- Title
- Authors and affiliations
- Abstract
- Introductions

Sets the stage for your work and thus should provide the minimal amount of background related to your work

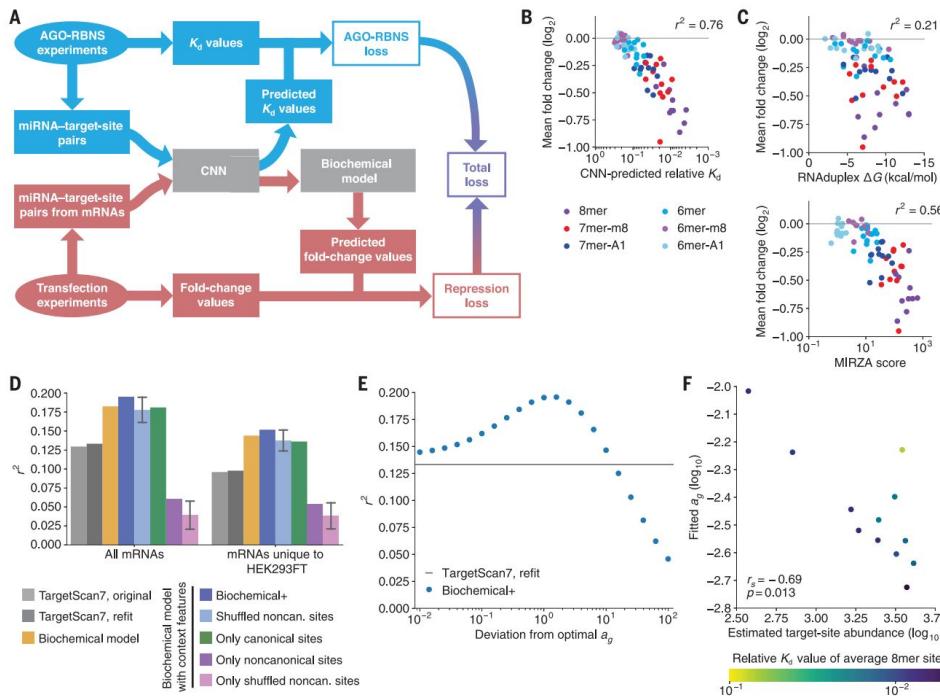
# Format for research papers in *Bioinformatics*

- Title
- Authors and affiliations
- Abstract
- Introductions
- Methods
- Results
- Discussion
- Acknowledgements
- References

# Methods before Results in *Bioinformatics*

- Title
- Authors and affiliations
- Abstract
- Introductions
- **Methods**
- **Results**
- Discussion
- Acknowledgements
- References

# Recent computational paper in *Science*



McGeary and Lin et al. 2020 from Bartel Lab

BiS800: Methods in functional genomics and computational molecular biology

# Method section is at the end in *Science*

## RESEARCH | RESEARCH ARTICLE

Although some target prediction algorithms (such as TargetScan) do not reward pairing to these motifs, most algorithms assume that such pairing enhances site affinity. Likewise, our model suggests that the reason why it is better that pairing to position 9 reduces site affinity (6), another reports that it increases *affinity* (22). We found that extending pairing to more sites in the miRNA 3' region diminished affinity in the context of seed-matched sites (Fig. 4), whereas extending pairing to nucleotide 9 or 10 enhanced affinity in the context of seed-unmatched sites (Fig. 4D). These results support the idea that extensive pairing to the miRNA 3' region unlocks productive pairing to nucleotides 9 to 12, which are typically paired to the 5' region.

The biochemical parameters fit by our model provided additional insights into miRNA targetting. In the first framework of our model, the fitted  $K_d$  values of silencing by the 3' region of the miRNA 3' region were higher than those of the 5' region. This suggested that a typical miRNA bound to an average of one silencing complex will experience a rate of tripling its decay rate, which would lead to a rapid increase in silencing abundance. In the congruence regions of our transcript experiments, this occupancy can be increased to as many as 100-fold by the 3' region. In addition, our final value for the OBP penalty suggested that the transylation machinery reduces site affinity by 5.5-fold.

After fitting the model to the cellular concentration of AGO loaded with the transfected miRNA and not bound to a target site. Whereas values of the other parameters were fixed, the OBP parameter was then used for testing;  $a_0$  was fit separately for each miRNA and passenger strand of each transfection experiment. Nonetheless, when  $a_0$  values were set to zero, the model predicted the same, the biochemical<sup>+</sup> model still outputted perfect TargetScan in predicting test-set repression over a 100-fold range of values (Fig. 6E), which is consistent with the observed range of miRNA abundances, our modeling framework had an advantage over other predictive methods in new contexts.

Our work extends the correlative models of targeting efficacy with a principle biochemical model that explains and predicts how miRNAs affect gene expression and the effects of miRNAs on their targets, raising the question of how the understanding and prediction of miRNA-mediated repression might be further improved. Acquiring site-affinity profiles for additional miRNAs with diverse

sequences will improve the CNN-predicted miRNA-mRNA affinity landscape and further flesh out the two major sources of targeting variability revealed by our study, that is, the sequence context of the target and the sequence context of the miRNA. Our model was also applied to different miRNAs and the substantial influence of local (12-n) site context. We suspect additional improvement will come from accounting for the fact that the main cause of targeting variability, which is the variability imparted by miRNA features more distant from the site. This variability is likely to be captured by the OBP parameter added to the biochemical model to generate the biochemical<sup>+</sup> model. Perhaps the most promising strategy for accounting for these features is to use a machine learning approach, such as a machine-learning approach that uses entire mRNA sequences to predict repression, leveraging substantially expanded repression datasets as well as sequence context to reflect the complete regulatory landscape, as specified by AGO within this essential biological pathway, might ultimately be computationally tractable.

### Methods summary

AGO-miRNA complexes were generated by adding synthetic miRNA duplexes to lysate from cells that overexpressed recombinant AGO2, and then these complexes were purified. The total cellular concentration of miRNA libraries was generated by *in vitro* transcription of synthetic DNA templates. For AGO-miRNAs, purified AGO-miRNA complex was transfected with the corresponding library molecules, and after reaching binding equilibrium, library molecules bound to AGO2 were isolated and separated and prepared for high-throughput sequencing. Examination of k-mers enriched within the bound library sequences identified miRNA target sites, and relative  $K_d$  values for each of these sites were calculated by using maximum likelihood estimation, fitting to AGO-miRNA results obtained over a 100-fold range in miRNA abundance.

Interspecies miRNA-mediated repression was measured by performing RNA-seq on HeLa cells that had been transfected with a synthetic miRNA duplex. In other words, the observation that, as expected (Fig. 6A), repression was higher for miRNAs with lower predicted gain abundance and lower general affinity for their targets (Fig. 6B).

Our work extends the correlative models of targeting efficacy with a principle biochemical model that explains and predicts how miRNAs affect gene expression and the effects of miRNAs on their targets, raising the question of how the understanding and prediction of miRNA-mediated repression might be further improved. Acquiring site-affinity profiles for additional miRNAs with diverse

**sequences will improve the CNN-predicted miRNA-mRNA affinity landscape and further flesh out the two major sources of targeting variability revealed by our study, that is, the sequence context of the target and the sequence context of the miRNA. Our model was also applied to different miRNAs and the substantial influence of local (12-n) site context. We suspect additional improvement will come from accounting for the fact that the main cause of targeting variability, which is the variability imparted by miRNA features more distant from the site. This variability is likely to be captured by the OBP parameter added to the biochemical model to generate the biochemical<sup>+</sup> model. Perhaps the most promising strategy for accounting for these features is to use a machine learning approach, such as a machine-learning approach that uses entire mRNA sequences to predict repression, leveraging substantially expanded repression datasets as well as sequence context to reflect the complete regulatory landscape, as specified by AGO within this essential biological pathway, might ultimately be computationally tractable.**

**REFERENCES AND NOTES**

- 1. D. H. Kim, *Nature Methods*, **17**, 20–23 (2020).
- 2. E. Lewis, *E. Lewis*, Towards a molecular understanding of miRNA function. *Cell* **141**, 956–966 (2010). doi:10.1016/j.cell.2010.07.023
- 3. P. Bartel, *C. Cell* **128**, 21–33 (2007). doi:10.1016/j.cell.2007.01.022
- 4. J. A. Giraldo, C. M. Lai, S. R. George, D. R. Bartel, Most miRNA molecules are recovered targets of miRNA expression profiles. *Nat. Commun.* **10**, 2000 (2019). doi:10.1038/s41564-019-04648
- 5. N. H. Lee, *C. Cell* **128**, 34–45 (2007). doi:10.1016/j.cell.2007.01.023
- 6. W. L. Samore, S. M. Joly, M. J. Moore, P. D. Zarnies, *Cell* **128**, 46–57 (2007). doi:10.1016/j.cell.2007.01.024
- 7. P. D. Zarnies, *Cell* **128**, 58–69 (2007). doi:10.1016/j.cell.2007.01.025
- 8. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *J. Cell* **128**, 70–81 (2007). doi:10.1016/j.cell.2007.01.026
- 9. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12534–12539 (2007). doi:10.1073/pnas.0703558104
- 10. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12540–12545 (2007). doi:10.1073/pnas.0703559104
- 11. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12546–12551 (2007). doi:10.1073/pnas.0703560104
- 12. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12552–12557 (2007). doi:10.1073/pnas.0703561104
- 13. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12558–12563 (2007). doi:10.1073/pnas.0703562104
- 14. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12564–12569 (2007). doi:10.1073/pnas.0703563104
- 15. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12570–12575 (2007). doi:10.1073/pnas.0703564104
- 16. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12576–12581 (2007). doi:10.1073/pnas.0703565104
- 17. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12582–12587 (2007). doi:10.1073/pnas.0703566104
- 18. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12588–12593 (2007). doi:10.1073/pnas.0703567104
- 19. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12594–12599 (2007). doi:10.1073/pnas.0703568104
- 20. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12600–12605 (2007). doi:10.1073/pnas.0703569104
- 21. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12606–12611 (2007). doi:10.1073/pnas.0703570104
- 22. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12612–12617 (2007). doi:10.1073/pnas.0703571104
- 23. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12618–12623 (2007). doi:10.1073/pnas.0703572104
- 24. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12624–12629 (2007). doi:10.1073/pnas.0703573104
- 25. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12630–12635 (2007). doi:10.1073/pnas.0703574104
- 26. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12636–12641 (2007). doi:10.1073/pnas.0703575104
- 27. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12642–12647 (2007). doi:10.1073/pnas.0703576104
- 28. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12648–12653 (2007). doi:10.1073/pnas.0703577104
- 29. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12654–12659 (2007). doi:10.1073/pnas.0703578104
- 30. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12660–12665 (2007). doi:10.1073/pnas.0703579104
- 31. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12666–12671 (2007). doi:10.1073/pnas.0703580104
- 32. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12672–12677 (2007). doi:10.1073/pnas.0703581104
- 33. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12678–12683 (2007). doi:10.1073/pnas.0703582104
- 34. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12684–12689 (2007). doi:10.1073/pnas.0703583104
- 35. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12690–12695 (2007). doi:10.1073/pnas.0703584104
- 36. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12696–12701 (2007). doi:10.1073/pnas.0703585104
- 37. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12702–12707 (2007). doi:10.1073/pnas.0703586104
- 38. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12708–12713 (2007). doi:10.1073/pnas.0703587104
- 39. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12714–12719 (2007). doi:10.1073/pnas.0703588104
- 40. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12720–12725 (2007). doi:10.1073/pnas.0703589104
- 41. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12726–12731 (2007). doi:10.1073/pnas.0703590104
- 42. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12732–12737 (2007). doi:10.1073/pnas.0703591104
- 43. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12738–12743 (2007). doi:10.1073/pnas.0703592104
- 44. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12744–12749 (2007). doi:10.1073/pnas.0703593104
- 45. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12750–12755 (2007). doi:10.1073/pnas.0703594104
- 46. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12756–12761 (2007). doi:10.1073/pnas.0703595104
- 47. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12762–12767 (2007). doi:10.1073/pnas.0703596104
- 48. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12768–12773 (2007). doi:10.1073/pnas.0703597104
- 49. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12774–12779 (2007). doi:10.1073/pnas.0703598104
- 50. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12780–12785 (2007). doi:10.1073/pnas.0703599104
- 51. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12786–12791 (2007). doi:10.1073/pnas.0703600104
- 52. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12792–12797 (2007). doi:10.1073/pnas.0703601104
- 53. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12798–12803 (2007). doi:10.1073/pnas.0703602104
- 54. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12804–12809 (2007). doi:10.1073/pnas.0703603104
- 55. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12810–12815 (2007). doi:10.1073/pnas.0703604104
- 56. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12816–12821 (2007). doi:10.1073/pnas.0703605104
- 57. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12822–12827 (2007). doi:10.1073/pnas.0703606104
- 58. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12828–12833 (2007). doi:10.1073/pnas.0703607104
- 59. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12834–12839 (2007). doi:10.1073/pnas.0703608104
- 60. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12840–12845 (2007). doi:10.1073/pnas.0703609104
- 61. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12846–12851 (2007). doi:10.1073/pnas.0703610104
- 62. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12852–12857 (2007). doi:10.1073/pnas.0703611104
- 63. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12858–12863 (2007). doi:10.1073/pnas.0703612104
- 64. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12864–12869 (2007). doi:10.1073/pnas.0703613104
- 65. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12870–12875 (2007). doi:10.1073/pnas.0703614104
- 66. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12876–12881 (2007). doi:10.1073/pnas.0703615104
- 67. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12882–12887 (2007). doi:10.1073/pnas.0703616104
- 68. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12888–12893 (2007). doi:10.1073/pnas.0703617104
- 69. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12894–12899 (2007). doi:10.1073/pnas.0703618104
- 70. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12900–12905 (2007). doi:10.1073/pnas.0703619104
- 71. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12906–12911 (2007). doi:10.1073/pnas.0703620104
- 72. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12912–12917 (2007). doi:10.1073/pnas.0703621104
- 73. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12918–12923 (2007). doi:10.1073/pnas.0703622104
- 74. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12924–12929 (2007). doi:10.1073/pnas.0703623104
- 75. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12930–12935 (2007). doi:10.1073/pnas.0703624104
- 76. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12936–12941 (2007). doi:10.1073/pnas.0703625104
- 77. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12942–12947 (2007). doi:10.1073/pnas.0703626104
- 78. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12948–12953 (2007). doi:10.1073/pnas.0703627104
- 79. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12954–12959 (2007). doi:10.1073/pnas.0703628104
- 80. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12960–12965 (2007). doi:10.1073/pnas.0703629104
- 81. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12966–12971 (2007). doi:10.1073/pnas.0703630104
- 82. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12972–12977 (2007). doi:10.1073/pnas.0703631104
- 83. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12978–12983 (2007). doi:10.1073/pnas.0703632104
- 84. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12984–12989 (2007). doi:10.1073/pnas.0703633104
- 85. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12990–12995 (2007). doi:10.1073/pnas.0703634104
- 86. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12996–12997 (2007). doi:10.1073/pnas.0703635104
- 87. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12998–12999 (2007). doi:10.1073/pnas.0703636104
- 88. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13000–13001 (2007). doi:10.1073/pnas.0703637104
- 89. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13002–13003 (2007). doi:10.1073/pnas.0703638104
- 90. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13004–13005 (2007). doi:10.1073/pnas.0703639104
- 91. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13006–13007 (2007). doi:10.1073/pnas.0703640104
- 92. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13008–13009 (2007). doi:10.1073/pnas.0703641104
- 93. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13010–13011 (2007). doi:10.1073/pnas.0703642104
- 94. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13012–13013 (2007). doi:10.1073/pnas.0703643104
- 95. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13014–13015 (2007). doi:10.1073/pnas.0703644104
- 96. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13016–13017 (2007). doi:10.1073/pnas.0703645104
- 97. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13018–13019 (2007). doi:10.1073/pnas.0703646104
- 98. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13020–13021 (2007). doi:10.1073/pnas.0703647104
- 99. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13022–13023 (2007). doi:10.1073/pnas.0703648104
- 100. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13024–13025 (2007). doi:10.1073/pnas.0703649104
- 101. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13026–13027 (2007). doi:10.1073/pnas.0703650104
- 102. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13028–13029 (2007). doi:10.1073/pnas.0703651104
- 103. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13030–13031 (2007). doi:10.1073/pnas.0703652104
- 104. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13032–13033 (2007). doi:10.1073/pnas.0703653104
- 105. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13034–13035 (2007). doi:10.1073/pnas.0703654104
- 106. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13036–13037 (2007). doi:10.1073/pnas.0703655104
- 107. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13038–13039 (2007). doi:10.1073/pnas.0703656104
- 108. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13040–13041 (2007). doi:10.1073/pnas.0703657104
- 109. J. T. Stine, J. Gao, Gohdharan, I. T. Hwang, *Proc.*

# Or supplementary materials ...

RESEARCH | RESEARCH ARTICLE

Although some target-prediction algorithms (such as TargetScan) do not reward pairing to these nucleotides, most algorithms assume that such pairing enhances site affinity. Likewise, although one biochemical study reports that pairing to position 9 reduces site affinity (6), another reports that it increases affinity (12). We found that extending pairing to nucleotide 9 or 10 neither enhanced nor diminished affinity in the context of seed-matched sites (Fig. 4), whereas extending pairing to nucleotide 9 or 10 enhanced affinity in the context of 3'-only sites (Fig. 2, C and D). These results support the idea that extensive pairing to the miRNA 3' region unlocks productive pairing to nucleotides 9 to 12, which is otherwise inaccessible (1).

The biochemical parameters fit by our model provided additional insights into miRNA tar-

sequences will improve the CNN-predicted miRNA-mRNA affinity landscape and further flesh out the two major sources of targeting variability revealed by our study, that is, the widespread differences in site preferences observed for different miRNAs and the substantial influence of local (12-nt) site context. We suspect additional improvement will come with increased ability to predict the other major cause of targeting variability, which is the variability imparted by mRNA features more distant from the site. This variability is captured only partially by the three features added to the biochemical model to generate the biochemical+ model. Perhaps the most promising strategy for accounting for these more distal features will be an unbiased machine-learning approach that uses entire mRNA sequences to predict repression. lever-

sured  $K_d$  values and the repression observed in the HeLa transfection experiments, and tested on the repression of endogenous mRNAs observed after transfecting miRNAs into HEK293T cells. Results were also tested on external datasets examining either intracellular binding of miRNAs by CLIP-seq or repression of endogenous mRNAs after miRNAs had been transfected, knocked down, or knocked out. **The details of each of these methods are described in the supplementary materials.**

## REFERENCES AND NOTES

1. D. P. Bartel. Metazoan microRNAs. *Cell* **173**, 20–51 (2018). doi: [10.1016/j.cell.2018.03.006](https://doi.org/10.1016/j.cell.2018.03.006); pmid: [29570994](https://pubmed.ncbi.nlm.nih.gov/29570994/)
2. S. Jonas, E. Izaurralde. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16**, 421–433 (2015). doi: [10.1038/nrg3965](https://doi.org/10.1038/nrg3965); pmid: [26077373](https://pubmed.ncbi.nlm.nih.gov/26077373/)
3. D. P. Bartel. MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215–233 (2009). doi: [10.1016/j.cell.2009.01.002](https://doi.org/10.1016/j.cell.2009.01.002); pmid: [19167326](https://pubmed.ncbi.nlm.nih.gov/19167326/)

# 7 + 21 = 28 pages of supplementary materials

incubating at 37°C for 30 min. The samples were then re-extracted with phenol–chloroform, EtOH-precipitated, and resuspended in water to their original volumes. Reverse transcription, PCR, and formamide gel purification to generate amplicons for RNA-seq were performed as described (section 5) with the following modifications: 1) an RT primer was designed to reverse transcribe the variants at 3' ends of the replicative library, and added homology to the 3' PCR primers; 2) for each RNA-seq library preparation (D or S), the volume of the PCR reactions were scaled up, using 1 μL of SuperScript II in 30 μL of total reaction per 5 μg of total RNA, 3) after base-cleaning up, using 1 μL of SuperScript II in 30 μL of total reaction per 5 μg of total RNA, 3) after base-cleaning, and neutralization with HEPES, each RT reaction was EtOH-precipitated and resuspended in 60 μL of water before the P30 step, and 4) after performing a pilot PCR using 4 μL of the cDNA in a 50 μL reaction to determine the minimal number of cycles to achieve amplification, the remaining 56 μL of cDNA was amplified in seven 100 μL PCR reactions. These seven reactions were combined, and DNA was precipitated and resuspended for formamide-gel purification. These modifications, which scaled up the input and amplification volume, were designed to increase the number of distinct library mRNAs contributing to the measured expression of each variant. All seven components (the six miRNA duplexes, the sequencing library, and water) were combined in 1 μL of water. The seven samples were sequenced with multiplexing on the lanes of an Illumina HiSeq 2500 run in rapid mode with 100-nt single-end reads. For analysis, reads were first subjected to quality-control filtering (section 9, steps 1–5). Reads passing these criteria were then assigned to one of the 29,992 sequences designed for the library, requiring a perfect match to the sequence. For each sequence, counts were normalized to the total number of perfectly matching counts to obtain counts per million (cpm).

## Computational and mathematical methods

### 9 RBNS read quality control

Each RBNS sequencing dataset was used if it satisfied the following criteria: 1) it passed the Illumina quality filter, as indicated by the presence of the number 1 rather than 0 in the final position of the fastq header line; 2) it did not contain any “N” base calls; 3) it did not contain any positions with a Phred quality score ( $\mathcal{Q}$ ) of B or lower; 4) the sequenced fast sample-multiplexing barcode associated with the read was identical to one of the barcodes used when generating the sequencing library; 5) it did not match either strand of the phi-X genome; 6) it did not nearly match (allowing up to two single-nucleotide-substitutions/insertions/deletions) the standards added to the samples during library workup, and 7) it contained either a TCG at positions 38–40 in the library of the first AGO2–miR-1 experiment or a TGT at these positions for all other experiments.

### 10 De novo site identification

To identify sites of an AGO–miRNA complex using RBNS results, we performed an analysis in which we 1) calculated the enrichment of all 10-nt  $k$ -mers in the library from the binding reaction with the greatest concentration of AGO–miRNA, 2) defined a site by computationally assisted manual curation of the ten most highly enriched 10-nt  $k$ -mers, as outlined below, and 3) removed all reads containing the identified site from both the input and the bound libraries corresponding to that AGO-RBNS experiment. This three-step process was repeated until no 10-nt  $k$ -mer with an enrichment  $>10$ -fold remained. For miR-1, miR-124, and miR-7, this process was performed with two separate AGO-RBNS experiments, each of which had used a separately purified AGO–miRNA complex (section 1). The AGO-RBNS experiments performed with second purification

# Some fancy mathematical equations

incubating at 37°C for 30 min. The samples were then re-extracted with phenol–chloroform, EtOH-precipitated, and resuspended in water to their original volumes. Reverse transcription, PCR, and formamide gel purification to generate amplicons for RNA-seq were performed as described (section 5) with the following modifications: 1) an RT primer was designed to reverse transcript the variant 3' ends of the input replicates library, and added homology to the 3' PCR primers; 2) no cDNA-seq library preparation (DpnI S1) was used; 3) the PCR reactions were scaled up, using 1  $\mu$ L of SuperScript II in 30  $\mu$ L of total reaction plus 5  $\mu$ g of total RNA, 3) after base-hydrolysis of the RT reactions, and neutralization with HEPES, each RT reaction was EtOH-precipitated and resuspended in 60  $\mu$ L of water before the P30 step, and 4) after performing a pilot PCR using 4  $\mu$ L of the cDNA in a 50  $\mu$ L reaction to determine the minimal number of cycles to achieve amplification, the remaining 56  $\mu$ L of cDNA was amplified in seven 100  $\mu$ L PCR reactions. These seven reactions were combined, and DNA was precipitated and resuspended for formamide-gel purification. These modifications, which scaled up the input and the amplification volume, were designed to increase the number of distinct library mRNAs contributing to the measured expression of each variant. All seven components (the six mRNA duplexes, the sequencing adapter, and water) were combined in equal amounts. The final samples were sequenced with multiplexing on the lanes of an Illumina HiSeq 2500 run in rapid mode with 100-nt single-end reads. For analysis, reads were first subjected to quality-control filtering (section 5, steps 1–5). Reads passing these criteria were then assigned to one of the 29,992 sequences designed for the library, requiring a perfect match to the sequence. For each sequence, counts were normalized to the total number of perfectly matching counts to obtain counts per million (cpm).

## Computational and mathematical methods

### 9 RBNS read quality control

Each RBNS sample was used if it satisfied the following criteria: 1) it passed the Illumina quality filter, as indicated by the presence of the number 1 rather than 0 in the final position of the fastq header line; 2) it did not contain any “N” base calls; 3) it did not contain any positions with a Phred quality score ( $\mathcal{Q}$ ) of B or lower; 4) the sequenced 6-nt sample-multiplexing barcode associated with the read was identical to one of the barcodes used when generating the sequencing library; 5) it did not match either strand of the phi-X genome; 6) it did not nearly match (allowing up to two single-nucleotide-substitutions/insertions/deletions) the standards added to the samples during library workup; and 7) it contained either a TCG at positions 38–40 in the library of the first AGO2–miR-1 experiment or a TGT at these positions for all other experiments.

### 10 De novo site identification

To identify sites of an AGO–miRNA complex using RBNS results, we performed an analysis in which we 1) calculated the enrichment of all 10-nt  $k$ -mers in the library from the binding reaction with the greatest concentration of AGO–miRNA, 2) defined a site by computationally assisted manual curation of the ten most highly enriched 10-nt  $k$ -mers, as outlined below, and 3) removed all reads containing the identified site from both the input and the bound libraries corresponding to that AGO-RBNS experiment. This three-step process was repeated until no 10-nt  $k$ -mer with an enrichment >10-fold remained. For miR-1, miR-124, and miR-7, this process was performed with two separate AGO-RBNS experiments, each of which had used a separately purified AGO–miRNA complex (section 1). The AGO-RBNS experiments performed with second purification

where  $p(y|x(\theta))$  is the probability of observing the sequencing counts  $y$  given the model-simulated abundances  $x(\theta)$  (itself a function of  $\theta$ ). We first describe the derivation of  $x(\theta)$  and then of  $f_{\text{cost}}(x)$ , a cost function scaling monotonically with  $\ln p(y|x(\theta))$  and therefore having a minimum value coincident with the MLE parameter estimates. We then derive the gradient of the cost function:

$$f_{\text{grad}}(\theta) = \nabla f_{\text{cost}}(x(\theta)) \quad (11.1.2)$$

The optimization routine was performed with the `optim` function in R (53) using the L-BFGS-B method, supplying both  $f_{\text{cost}}(x)$  and  $f_{\text{grad}}(x)$  to the optimizing function as compiled C scripts through the `.C` interface. This enabled efficient, simultaneous estimation of a large set (>50,000) of  $K_a$  values per AGO-RBNS experiment.

### 11.2 Derivation of $x(\theta)$

The function  $x(\theta)$  produces an  $m \times n$  matrix where each element  $x_{ij}$  specifies a model estimate of the concentration of library RNA molecules of site type  $j$  recovered from binding reaction  $i$  for a particular AGO-RBNS experiment. The dimensions  $m$  and  $n$  are therefore determined by the number of distinct types of sites (where library RNA molecules that do not contain a site constitute the  $m$ th site type) and the total number of binding reactions comprising that AGO-RBNS experiment, respectively. In practice,  $n=5$  for all experiments other than that with AGO2–miR-7, for which  $n=4$  because the 4% dilution sample was discarded for technical reasons. This calculation requires as input the total concentration of each site type  $I = (I_1, \dots, I_n)$ , the total concentration of AGO–miRNA complex (hereafter referred to as “AGO”) in each binding reaction  $a = (a_1, \dots, a_n)$ , the  $K_a$  value describing the binding between AGO and each site type  $K = (K_1, \dots, K_n)$ , and the concentration of library RNA recovered due to nonspecific binding to the nitrocellulose filter  $b$ , which is assumed to be constant across all five samples and therefore given by a single parameter. The vector  $I$  is estimated using

$$I = \frac{y'}{\sum_{i=1}^m y'_i} \times 100 \text{ nM}, \quad (11.2.1)$$

where  $y'$  is the vector of read counts corresponding to each site type as measured in the sequencing of the input library. Each element  $a_j$  of  $a$  is calculated from the experimentally determined dilution series

$$a = \alpha \times s$$

where  $s$

$$= \alpha \times (0.4\%, 1.27\%, 4\%, 12.7\%, 40\%),$$

where  $\alpha$  is the stock (pre-dilution) concentration of AGO, and so only the parameter  $\alpha$  is included in  $\theta$ . The set of parameters to be optimized is therefore

$$(K_1, K_2, \dots, K_n, a, b) \quad (11.2.3)$$

Because these parameters represent either binding affinities or concentrations, for which negative values are physically meaningless,  $x(\theta)$  performs an exponential transformation on  $\theta$ :

# And more fancy mathematical equations

incubating at 37°C for 30 min. The samples were then re-extracted with phenol–chloroform, EtOH-precipitated, and resuspended in water to their original volumes. Reverse transcription, PCR, and formamide gel purification to generate amplicons for RNA-seq were performed as described (section 5) with the following modifications: 1) an RT primer was designed to reverse transcript the variant 3' ends of the input representation library; and 2) analogously the 3' PCR primers for small RNA sequencing preparation (DyNab S3) were redesigned so that the reactions were scaled up, using 1 μL of SuperScript II in 30 μL of total reaction per 5 μg of total RNA, 3) after base-hydrolysis, using 1 μL of the RT reactions, and neutralization with HEPES, each RT reaction was EtOH-precipitated and resuspended in 60 μL of water before the P30 step, and 4) after performing a pilot PCR using 4 μL of the cDNA in a 50 μL reaction to determine the minimal number of cycles to achieve amplification, the remaining 56 μL of cDNA was amplified in seven 100 μL PCR reactions. These seven reactions were combined, and DNA was precipitated and resuspended for formamide-gel purification. These modifications, which scaled up the input and the amplification volume, were designed to increase the number of distinct library mRNAs contributing to the measured expression of each variant. All seven components (the six mRNA duplexes, the sequencing adapter, and the workflow) were used to generate the sequencing samples were sequenced with multiplexing on the lanes of an Illumina HiSeq 2500 run in rapid mode with 100-nt single-end reads. For analysis, reads were first subjected to quality-control filtering (section 5, steps 1–5). Reads passing these criteria were then assigned to one of the 29,992 sequences designed for the library, requiring a perfect match to the sequence. For each sequence, counts were normalized to the total number of perfectly matching counts to obtain counts per million (cpm).

## Computational and mathematical methods

### 9 RBNs read quality control

Each RBNs sequencing process used if it satisfied the following criteria: 1) it passed the Illumina quality filter, as indicated by the presence of the number 1 rather than 0 in the final position of the fastq header line; 2) it did not contain any “N” base calls; 3) it did not contain any positions with a Phred quality score ( $\mathcal{Q}$ ) of B or lower; 4) the sequenced fastq sample-multiplexing barcode associated with the read was identical to one of the barcodes used when generating the sequencing library; 5) it did not match either strand of the ph-X genome; 6) it did not nearly match (allowing up to two single-nucleotide-substitutions/insertions/deletions) the standards added to the samples during library workup; and 7) it contained either a TCG at positions 38–40 in the library of the first AGO2–miR-1 experiment or a TGT at these positions for all other experiments.

### 10 De novo site identification

To identify sites of an AGO–miRNA complex using RBNs results, we performed an analysis in which we 1) calculated the enrichment of all 10-nt  $k$ -mers in the library from the binding reaction with the greatest concentration of AGO–miRNA, 2) defined a site by computationally assisted manual curation of the ten most highly enriched 10-nt  $k$ -mers, as outlined below, and 3) removed all reads containing the identified site from both the input and the bound libraries corresponding to that AGO–RBNs experiment. This three-step process was repeated until no 10-nt  $k$ -mer with an enrichment  $>10$ -fold remained. For miR-1, miR-124, and miR-7, this process was performed with two separate AGO–RBNs experiments, each of which had used a separately purified AGO–miRNA complex (section 1). The AGO–RBNs experiments performed with second purification

where  $p(y|x(\theta))$  is the probability of observing the sequencing counts  $y$  given the model-simulated abundances  $x(\theta)$  (itself a function of  $\theta$ ). We first describe the derivation of  $x(\theta)$  and then of  $f_{\text{cost}}(x)$ , a cost function scaling monotonically with  $\ln p(y|x(\theta))$  and therefore having a minimum value coincident with the MLE parameter estimates. We then derive the gradient of the cost function:

$$f_{\text{grad}}(\theta) = \nabla f_{\text{cost}}(x(\theta)) \quad (11.1.2)$$

The optimization routine was performed with the `optim` function in R (53) using the L-BFGS-B method, supplying both  $f_{\text{cost}}(x)$  and  $f_{\text{grad}}(x)$  to the optimizing function as compiled C scripts through the `C` interface. This enabled efficient, simultaneous estimation of a large set ( $>50,000$ ) of  $K_a$  values per AGO–RBNs experiment.

### 11.2 Derivation of $x(\theta)$

The function  $x(\theta)$  produces an  $m \times n$  matrix where each element  $x_{ij}$  specifies a model estimate of the concentration of library RNA molecules of site type  $i$  recovered from binding reaction  $j$  for a particular AGO–RBNs experiment. The dimensions  $m$  and  $n$  are therefore determined by the number of distinct types of sites (where library RNA molecules that do not contain a site constitute the  $m$ th site type) and the total number of binding reactions comprising that AGO–RBNs experiment, respectively. In practice,  $n=5$  for all experiments other than that with AGO2–miR-7, for which  $n=4$  because the 4% dilution sample was discarded for technical reasons. This calculation requires as input the total concentration of each site type  $I = (I_1, \dots, I_n)$ , the total concentration of AGO–miRNA complex (hereafter referred to as “AGO”) in each binding reaction  $a = (a_1, \dots, a_n)$ , the  $K_a$  value describing the binding between AGO and each site type  $K = (K_1, \dots, K_n)$ , and the concentration of library RNA recovered to nonspecific binding to the nitrocellulose filter  $b$ , which is assumed to be constant across all five samples and therefore given by a single parameter. The vector  $I$  is estimated using

$$I = \frac{y'}{\sum_{i=1}^n y'_i} \times 100 \text{ nM}, \quad (11.2.1)$$

where  $y'$  is the vector of read counts corresponding to each site type as measured in the sequencing of the input library. Each element  $a_j$  of  $a$  is calculated from the experimentally determined dilution series

$$a = \alpha \times s \quad (11.2.2)$$

where  $s$  is the stock (pre-dilution) concentration of AGO, and so only the parameter  $\alpha$  is included in  $\theta$ . The set of parameters to be optimized is therefore

$$(K_1, K_2, \dots, K_n, a, b) \quad (11.2.3)$$

Because these parameters represent either binding affinities or concentrations, for which negative values are physically meaningless,  $x(\theta)$  performs an exponential transformation on  $\theta$ :

(11.2.14) cannot be used directly; it requires a value for the concentration of unbound AGO in sample  $j$ ,  $a_j^f$ . This value is obtained by invoking the conservation of mass for AGO in sample  $j$ :

$$a_j = a_j^f + \sum_{i=1}^n c_{ij}. \quad (11.2.15)$$

Because each  $c_{ij}$  value is itself a function of  $I$ ,  $K$ , and  $a$  according to equation (11.1.2), equation (11.2.15) specifies a single value of  $a_j^f$ . However, this equation cannot be rearranged to an explicit expression for  $a_j^f$ . Therefore, each time  $x$  is calculated during the optimization routine requires that  $a_j^f$  first be numerically approximated by finding the root of

$$f(a_j^f) = ax_j - a_j^f - \sum_{i=1}^n \frac{I_i a_i^f}{a_j^f + K_i} \quad (11.2.16)$$

within the interval  $0 < a_j^f < as$ . This was performed using compiled C code modified from the `zeroin`/Fortran root-finding subroutine.

### 11.3 Derivation of $f_{\text{cost}}(x)$

The cost function  $f_{\text{cost}}(x)$  is derived from the product of the negative log multinomial probability mass function for each column /

$$f_{\text{cost}}(x) = -\ln \prod_{j=1}^n f_{\text{usb}}(Y_j, \pi_j) \\ = -\ln \prod_{j=1}^n \frac{Y_j!}{\prod_{i=1}^m v_{ij}!} \prod_{i=1}^m \pi_i^{v_{ij}}, \quad (11.3.1)$$

where  $\pi_j$  is the expected frequency of each site type  $i$  in sample  $j$  according to the model values  $y_g$ , and  $Y_j = \sum_{i=1}^m v_{ij}$ . Each expected frequency vector  $\pi_j$  is trivially given by  $x_j / X_j$  (where  $X_j = \sum_{i=1}^m v_{ij}$ ), thereby providing the link between the model simulation and subsequent likelihood estimation. Substituting  $\pi_j$  and distributing the natural log yields

$$f_{\text{cost}}(x) = \sum_{j=1}^n \left[ Y_j \ln X_j - \sum_{i=1}^m Y_j \ln y_g + \sum_{i=1}^m \ln y_g - \ln Y_j \right]. \quad (11.3.2)$$

After discarding the third and fourth terms in equation (11.3.2) because they do not contain any terms of  $x_j$ , and are therefore not related to the MLE estimation of  $\theta$ , the final cost function is given by

# And more fancy mathematical equations

incubating at 37°C for 30 min. The samples were then re-extracted with phenol-chloroform, EtOH-precipitated, and resuspended in water to their original volumes. Reverse transcription, PCR, and formamide gel purification to generate amplicons for RNA-seq were performed as described (section 5) with the following modifications: 1) no RT primer was designed to reverse transcribe the variable 3'UTR region of the reporter library, and add homology to the 3'UTR primers; 2) 100 ng total RNA was used for each reaction; 3) 100 pmol of each RT reaction were scaled up, using 1 μL of SuperScript II in 30 μL of total reaction per 5 μg of total RNA; 3) after base-hydrolysis of the RT reactions, and neutralization with HEPES, each reaction was EtOH-precipitated and resuspended in 60 μL of water before the P30 step; and 4) after performing a pilot PCR using 4 μL of the cDNA in a 50 μL reaction to determine the minimal number of cycles to achieve amplification, the remaining 50 μL of cDNA was amplified in seven 100 μL PCR reactions. These seven reactions were combined, and DNA was precipitated and resuspended for formamide-gel purification. These modifications, which scaled up the input and the amplification volume, were designed to increase the number of distinct library mRNAs contributing to the measured expression of each variant. All seven conditions (the six mRNA duplex modifications and the control transcript) were run in duplicate, and the fourteen samples were processed with barcoding on the basis of an Agilent HiSeq 2500 run in rapid mode with 100-mt single-end reads. For analysis, reads were first subjected to quality filtering (section 9, steps 1–5). Reads passing these criteria were then assigned to 29,992 sequences designed for the library, requiring a perfect match to the sequence; counts were normalized to the total number of perfectly matching counts per million (cpm).

## Computational and mathematical methods

### 9 RBNS read quality control

Each RBNS sequencing read was used if it satisfied the following criteria: 1) it passed Illumina quality filter, as indicated by the presence of the number 1 rather than 0 position of the fastq file line, 2) it did not contain any “N” base calls, 3) it did not contain any positions with a Phred-quality score (Q) of 0 or less, 4) the expected 6 nt sample-multiplexing barcode associated with the read was identical to one of the barcodes used when generating the sequencing library, 5) it did not match either strand of the phi-X genome, 6) it did not nearly match (allowing up to two single-nucleotide-substitutions/insertions/deletions) the standards added to the samples during library workflow, and 7) it contained either a TCG at positions 38–40 of the phi-X of the first AGO2–miR-1 experiment or a TGT at these positions for all other experiments.

### 10 De novo site identification

To identify enriched AGO-RBNS complex using RBNS results, we performed an analysis in which we 1) calculated the enrichment of all 10-nt *k*-mers in the library from the binding reaction with the greatest concentration of AGO–miRNA, 2) defined a site by computationally assisted manual curation of the ten most highly enriched 10-nt *k*-mers, as outlined below, and 3) removed all reads containing the identified site from both the input and the bound libraries corresponding to that AGO-RBNS experiment. This three-step process was repeated until no 10-nt *k*-mer remained for that AGO-RBNS experiment. For miR-1, miR-124, and miR-7, this process was performed with two separate AGO-RBNS experiments, each of which had used a separately purified AGO–miRNA complex (section 1). The AGO-RBNS experiments performed with second purification

I'm pretty sure there's only one person who fully understands the computational methods of this paper.

where  $p(y|x(\theta))$  is the probability of observing the sequencing counts  $y$  given the model-simulated abundances  $x(\theta)$  (itself a function of  $\theta$ ). We first describe the derivation of  $x(\theta)$  and then of  $f_{\text{cost}}(x)$ , a cost function scaling monotonically with  $\ln p(y|x(\theta))$  and therefore having a minimum value coincident with the MLE parameter estimates. We then derive the gradient of the cost function

$$f_{\text{prob}}(\theta) = \nabla f_{\text{cost}}(x(\theta)). \quad (11.1.2)$$

The optimization routine was performed with the *optim* function in R (53) using the L-BFGS-B method, supplying both  $f_{\text{cost}}(x)$  and  $f_{\text{prob}}(x)$  to the optimizing function as compiled C scripts through the *C* interface. This enabled efficient, simultaneous estimation of a large set ( $>50,000$ ) of  $K_4$  values per AGO-RBNS experiment.

### 11.2 Derivation of $x(\theta)$

The function  $x(\theta)$  produces an  $m \times n$  matrix where each element  $x_{ij}$  specifies a model estimate of the concentration of library RNA molecules of site type  $j$  recovered from binding reaction  $i$  for a particular AGO-RBNS experiment. The dimensions  $m$  and  $n$  are therefore determined by the

five samples and therefore given by a single parameter. The vector  $I$  is estimated using

$$I = \frac{y'}{\sum_{i=1}^m y'_i} \times 100 \text{ nM}, \quad (11.2.1)$$

where  $y'$  is the vector of read counts corresponding to each site type as measured in the sequencing of the input library. Each element  $a_j$  of  $a$  is calculated from the experimentally determined dilution series

$$a = \alpha \times (0.4\% \text{, } 1.27\% \text{, } 4\% \text{, } 12.7\% \text{, } 40\%), \quad (11.2.2)$$

where  $\alpha$  is the stock (pre-dilution) concentration of AGO, and so only the parameter  $\alpha$  is included in  $\theta$ . The set of parameters to be optimized is therefore  $(K_1, K_2, \dots, K_m, a, b)$  (11.2.3)

Because these parameters represent either binding affinities or concentrations, for which negative values are physically meaningless,  $x(\theta)$  performs an exponential transformation on  $\theta$ :

(11.2.14) cannot be used directly; it requires a value for the concentration of unbound AGO in sample  $j$ ,  $a_j^T$ . This value is obtained by invoking the conservation of mass for AGO in sample  $j$ :

$$a_j = a_j^T + \sum_{i=1}^m c_{ij}. \quad (11.2.15)$$

Because each  $c_{ij}$  value is itself a function of  $I$ ,  $K_i$ , and  $a$  according to equation (11.1.12), equation (11.2.15) specifies a single value of  $a_j^T$ . However, this equation cannot be rearranged to an explicit expression for  $a_j^T$ . Therefore, each time  $x$  is calculated during the optimization routine requires that  $a_j^T$  first be numerically approximated by finding the root of

$$f(a_j^T) = a_j - a_j^T - \sum_{i=1}^m \frac{I_i a_i^T}{a_i^T + K_i} \quad (11.2.16)$$

within the interval  $0 < a_j^T < a_j$ . This was performed using compiled C code modified from the

e product of the negative log multinomial

$$\begin{aligned} & -\ln \prod_{j=1}^m f_{\text{mb}}(Y_j; \pi_j) \\ & = -\ln \prod_{j=1}^m \frac{Y_j!}{\prod_{i=1}^n Y_{ij}!} \prod_{i=1}^n \pi_{ij}^{Y_{ij}}, \end{aligned} \quad (11.3.1)$$

where  $\pi_j$  is the expected frequency of each site type  $i$  in sample  $j$  according to the model values  $y_g$ , and  $Y_j = \sum_{i=1}^n Y_{ij}$ . Each expected frequency vector  $\pi_j$  is trivially given by  $x_j / X_j$  (where  $X_j = \sum_{i=1}^n Y_{ij}$ ), thereby providing the link between the model simulation and subsequent likelihood estimation. Substituting  $\pi_j$  and distributing the natural log yields

$$f_{\text{cost}}(x) = \sum_{j=1}^m \left[ Y_j \ln X_j - \sum_{i=1}^n Y_{ij} \ln y_g + \sum_{i=1}^n \ln Y_{ij}! - \ln Y_j! \right]. \quad (11.3.2)$$

After discarding the third and fourth terms in equation (11.3.2) because they do not contain any terms of  $x_j$ , and are therefore not related to the MLE estimation of  $\theta$ , the final cost function is given by

# Method section is seemingly the most unimportant section

- Title
- Authors and affiliations
- Abstract
- Introductions
- Results
- Discussion
- Acknowledgements
- References
- Methods

“Plus de détails, plus de détails, disait-il-à  
son fils, il n'y a de originalité et de vérité que  
dans les détails.”

From *Lucien Leuwen* by Stendhal

“More details, more details,” he said to his son, “there is originality and truth only in the details.”

By Google Translate

“More details, more details,” he said to his son, “there is **originality** and **truth** only in the details.”

By Google Translate

# Choices that you'll make as a scientist or engineer

Researcher vs. Journalist

# Choices that you'll make as a scientist or engineer

Researcher vs. Journalist

Professional expert vs. Intramural research

# Choices that you'll make as a scientist or engineer

Researcher vs. Journalist

Professional expert vs. Intramural research

Deep vs. shallow learning (not referring to machine learning)

# Choices that you'll make as a scientist or engineer

Researcher vs. Journalist

Professional expert vs. Intramural research

Deep vs. shallow  The attention to detail is what makes you a professional in the field, and that detail, if available, is always in the method section.

Corollary: if you can read/write the method section of papers from a specific field, then you're a professional.

# Case study: Dr. Young-suk Lee

BSc Computer Science and BSc Mathematics

PhD Computer Science

# Case study: Dr. Young-suk Lee publishes in SARS2??

BSc Computer Science and BSc Mathematics

PhD Computer Science

Manuscript [Click here to view linked References ▾](#)

**The SARS-CoV-2 RNA interactome**

**Authors**  
Sungyul Lee<sup>1,2,4</sup>, Young-suk Lee<sup>1,2,4</sup>, Yeon Choi<sup>1,2</sup>, Ahyeon Son<sup>1,2</sup>, Youngran Park<sup>1,2</sup>, Kyung-Min Lee<sup>3</sup>, Jeesoo Kim<sup>1,2</sup>, Jong-Seo Kim<sup>1,2</sup>, and V. Narry Kim<sup>1,2,\*</sup>

**Affiliations**  
<sup>1</sup>Center for RNA Research, Institute for Basic Science, Seoul 08826, Republic of Korea.  
<sup>2</sup>School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea.  
<sup>3</sup>International Vaccine Institute, Seoul 08826, Republic of Korea.  
<sup>4</sup>These authors contributed equally to this work.  
\*Corresponding author. Email: narrykim@snu.ac.kr

**Abstract**  
SARS-CoV-2 is an RNA virus whose success as a pathogen relies heavily on its ability to repurpose host RNA-binding proteins to form its own RNA interactome. To uncover the SARS-CoV-2 RNA interactome, we developed and applied a highly sensitive ribonucleoprotein capture method called MORIP (Massive Oligo-hybridized RNA Interactome Profiling) which led to the identification of 109 host factors. Applying MORIP on HCoV-OC43 revealed evolutionarily conserved interactions between viral RNAs and host proteins. Leveraging published data, we delineated antiviral factors stimulated by JAK-STAT signaling and proviral factors responsible for hijacking multiple steps of mRNA life cycle. Knockdown experiments indicated that LARP1, PARP12, FUBP3, ZC3HAV1 (ZAP), TRIM25, SND1, RPS9, RPS3, EIF4H, FAM120A/C, CELF1, RALY, HDLBP (Vigilin), and CNBP function as antiviral factors while EIF3A, EIF3D, and CSDE1 are required for viral replication. Overall, this study provides a comprehensive list of RBPs regulating coronaviral replication and opens new avenues for therapeutic interventions.

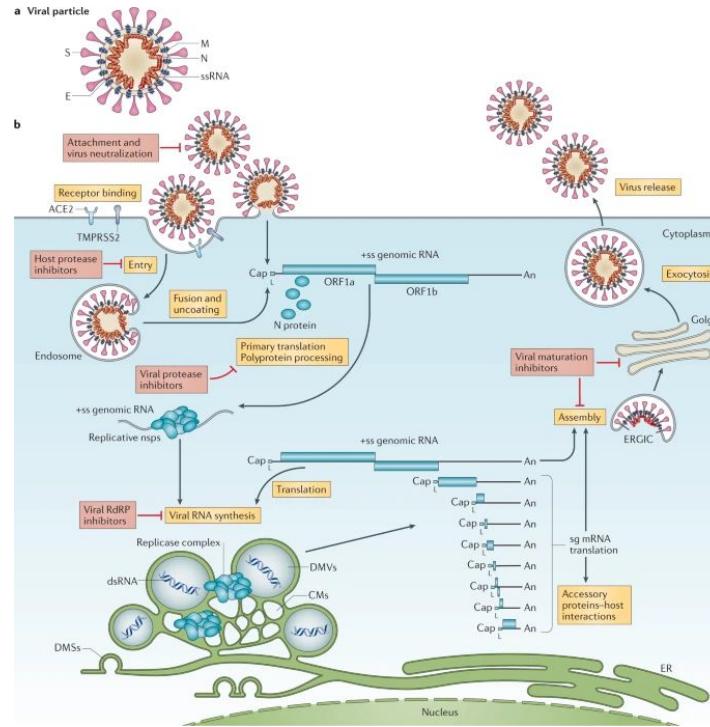
**Keywords:** SARS-CoV-2; Coronavirus; COVID-19; Host-Pathogen Interactions; RNA Interactome Capture; RNA-Binding Proteins; Ribonucleoproteins; HCoV-OC43; RNA Virus; Mass Spectrometry

# Case study: Dr. Young-suk Lee publishes in SARS2??

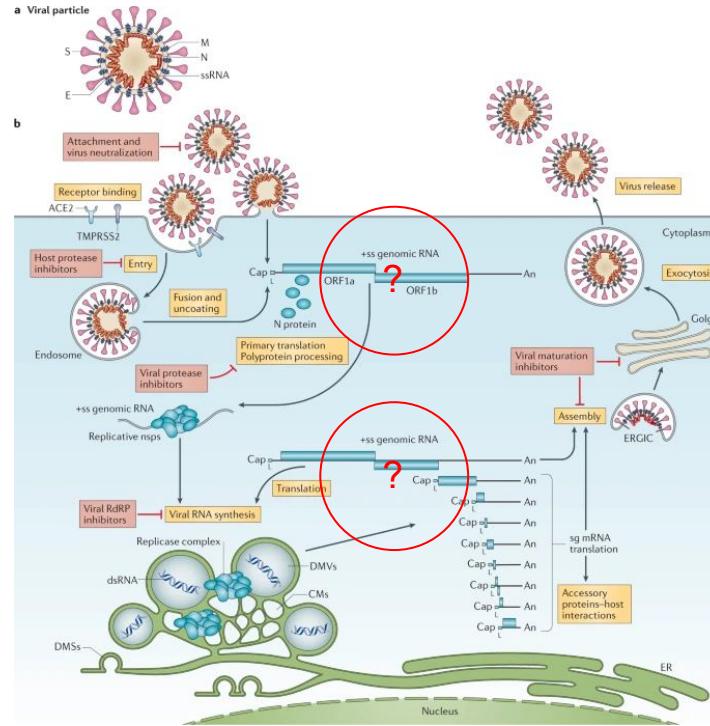
Hypothesis:

1. Solely computation work of SARS-CoV-2 data
2. Passenger author and no major contribution to the field
3. Corresponding author is a long-standing SARS-CoV-2 expert

# Case study: The life cycle of SARS-CoV-2



# Case study: The life cycle of SARS-CoV-2 RNAs?



Philip V'kovski et al. (2021) Review

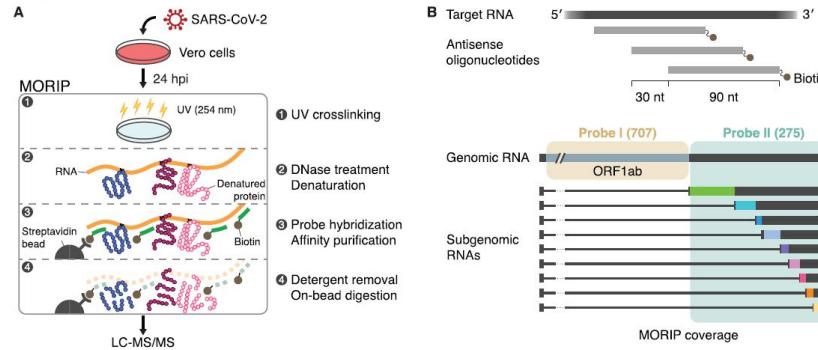
BiS800: Methods in functional genomics and computational molecular biology

# Development of biochemical and computational methods

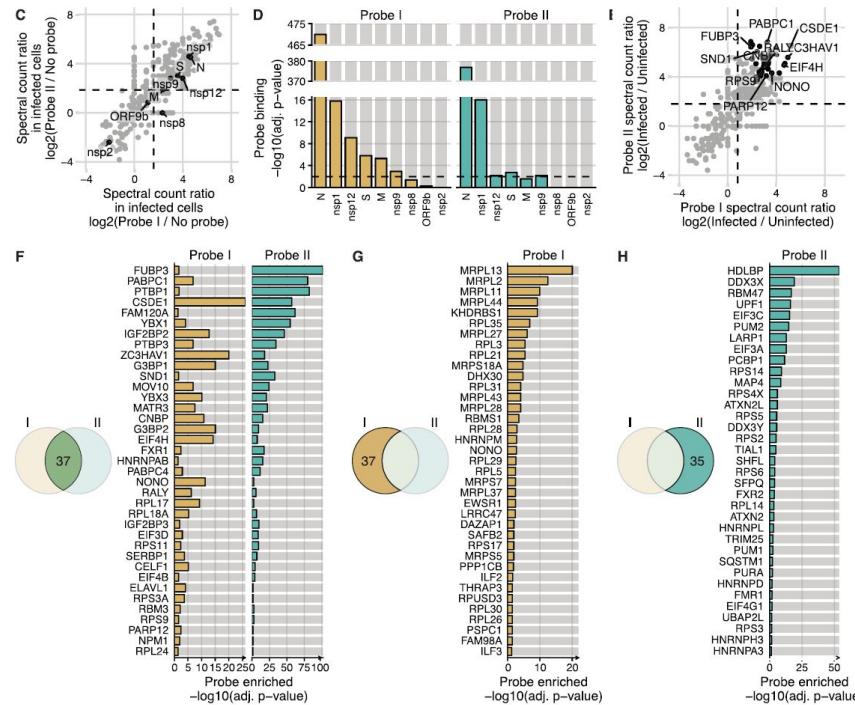
Figure 1

[Click here to access/download;Figure;Figure 1.pdf](#) ↗

Figure 1



# Identification of proteins that regulate SARS-CoV-2 RNA



# Details, details, and details

|    | A                   | B                   | C  | D                              | E            | F                | G   | H              |
|----|---------------------|---------------------|--|--------------------------------|--------------|------------------|---|----------------|
| 1  | Year of Publication | Journal             | Title  | Author                         | Article type | Topic            | Pubmed Link   | Misc           |
| 2  | 1990                | Journal of Virology | Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal.                  | Makino S, Yokomori K, Lal M    | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2243386">https://www.ncbi.nlm.nih.gov/pubmed/2243386</a>                           |                |
| 3  | 1991                | Journal of Virology | A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs.                              | van der Most RG1, Bredenbe     | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2033672">https://www.ncbi.nlm.nih.gov/pubmed/2033672</a>                           |                |
| 4  | 1992                | Journal of Virology | Identification and characterization of a coronavirus packaging signal.   | Fosmire JA1, Hwang K, Maki     | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/1316465">https://www.ncbi.nlm.nih.gov/pubmed/1316465</a>                           |                |
| 5  | 1994                | Virology            | Coronavirus translational regulation: leader affects mRNA efficiency.  | Tahara SM, Dietlin TA, Berg    | research     | Translation      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8030227">https://www.ncbi.nlm.nih.gov/pubmed/8030227</a>                           |                |
| 6  | 1996                | Virology            | The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus.  | Bos EC, Luytjes W, van der M   | research     | VLP              | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8615041">https://www.ncbi.nlm.nih.gov/pubmed/8615041</a>                           | reporter       |
| 7  | 1998                | Adv Exp Med Biol.   | Mouse hepatitis virus nucleocapsid protein as a translational effector of viral mRNAs.   | Tahara SM, Dietlin TA, Nelso   | research     | Translation      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/9782298">https://www.ncbi.nlm.nih.gov/pubmed/9782298</a>                           |                |
| 8  | 2000                | J Gen Virol         | High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA.                                  | Nelson GW1, Stohlman SA,       | research     | RNA-binding      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10640556">https://www.ncbi.nlm.nih.gov/pubmed/10640556</a>                         |                |
| 9  | 2000                | Journal of Virology | Identification of a bovine coronavirus packaging signal.   | Cologna R, Hogue BG,           | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10590153">https://www.ncbi.nlm.nih.gov/pubmed/10590153</a>                         |                |
| 10 | 2001                | Journal of Virology | Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging.  | Narayanan K1, Makino S,        | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/11533169">https://www.ncbi.nlm.nih.gov/pubmed/11533169</a>                         |                |
| 11 | 2003                | Acta Pharmacol Sin  | Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis.  | Qin L, Xiong B, Luo C, Guo Z   | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/12791173">https://www.ncbi.nlm.nih.gov/pubmed/12791173</a>                         |                |
| 12 | 2005                | Journal of Virology | Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent.                  | Hsieh PK1, Chang SC, Huan      | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16254430">https://www.ncbi.nlm.nih.gov/pubmed/16254430</a>                         |                |
| 13 | 2005                | Journal of Virology | Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis.                        | Sola I, Moreno JL, Zúñiga S,   | research     | template-switchi | <a href="https://www.ncbi.nlm.nih.gov/pubmed/15681451">https://www.ncbi.nlm.nih.gov/pubmed/15681451</a>                         |                |
| 14 | 2007                | Virology            | Coronavirus nucleocapsid protein is an RNA chaperone.  | Zúñiga S, Sola I, Moreno JL,   | research     | RNA chaperone    | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16979208">https://www.ncbi.nlm.nih.gov/pubmed/16979208</a>                         |                |
| 15 | 2007                | Journal of Virology | New structure model for the packaging signal in the genome of group Ia coronaviruses.  | Chen SC, van den Born E, v     | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/17426856">https://www.ncbi.nlm.nih.gov/pubmed/17426856</a>                         |                |
| 16 | 2009                | Journal of Virology | Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications       | I Chang CK, Hsu YL, Chang Y    | research     | RNA-binding      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19052082">https://www.ncbi.nlm.nih.gov/pubmed/19052082</a>                         |                |
| 17 | 2009                | J Mol Biol.         | Coronavirus N protein-N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA                | c Grosshoechme NE, Li T, Keane | research     | RNA-binding      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19782089">https://www.ncbi.nlm.nih.gov/pubmed/19782089</a>                         |                |
| 18 | 2010                | Journal of Virology | Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.   | Zúñiga S, Cruz JL, Sola I,     | research     | template switchi | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19955314">https://www.ncbi.nlm.nih.gov/pubmed/19955314</a>                         | in vitro templ |
| 19 | 2010                | Springer            | Molecular Biology of the SARS-CoV  | Lal K,                         | book         | SARS             | <a href="https://link.springer.com/book/10.1007/978-3-642-03683-5">https://link.springer.com/book/10.1007/978-3-642-03683-5</a> |                |
| 20 | 2011                | Journal of Virology | Cellular poly(c) binding protein 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1 $\beta$ and suppo | Beura L, Dinh PX, Osorio F     | research     | host factors     | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21976641">https://www.ncbi.nlm.nih.gov/pubmed/21976641</a>                         | immunoprecip   |
| 21 | 2011                | RNA Biol.           | RNA-RNA and RNA-protein interactions in coronavirus replication and transcription.   | Sola I, Mateos-Gomez PA, Al    | review       | template-switchi | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21378501">https://www.ncbi.nlm.nih.gov/pubmed/21378501</a>                         |                |
| 22 | 2011                | PLoS Pathogen       | SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Ind                    | Huang C, Lokugamage KG, I      | research     | nsp1             | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22174690">https://www.ncbi.nlm.nih.gov/pubmed/22174690</a>                         |                |
| 23 | 2011                | Journal of Virology | Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis.                       | Dufour D, Mateos-Gomez PA      | research     | template-switchi | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21388138">https://www.ncbi.nlm.nih.gov/pubmed/21388138</a>                         |                |
| 24 | 2011                | Journal of Virology | The poly(purine) tract-binding protein affects coronavirus RNA accumulation levels and relocates viral RNAs to novel cytoplasmic domai               | Sola I, Galán C, Mateos-Gó     | research     | host factor      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21415158">https://www.ncbi.nlm.nih.gov/pubmed/21415158</a>                         |                |
| 25 | 2012                | J Biol Chem.        | Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucle          | Keane SC, Liu P, Leibowitz J   | research     | RNA-binding      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22241479">https://www.ncbi.nlm.nih.gov/pubmed/22241479</a>                         |                |
| 26 | 2013                | Journal of Virology | The cellular interactome of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology.          | Emmott E, Munday D, Bicker     | research     | host factors     | <a href="https://www.ncbi.nlm.nih.gov/pubmed/23637411">https://www.ncbi.nlm.nih.gov/pubmed/23637411</a>                         | immunoprecip   |
| 27 | 2014                | Viruses             | The coronavirus nucleocapsid is a multifunctional protein  | McBride R, van Zyl M, Fielder  | review       | N protein        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25105276">https://www.ncbi.nlm.nih.gov/pubmed/25105276</a>                         |                |
| 28 | 2014                | Journal of Virology | Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid prot                 | Kuo L, Koetzner CA, Hurst K    | research     | packaging        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24501403">https://www.ncbi.nlm.nih.gov/pubmed/24501403</a>                         |                |
| 29 | 2014                | J Med Chem.         | Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target.                         | Lin SY, Liu CL, Chang YM, Zi   | research     | structure        | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24564608">https://www.ncbi.nlm.nih.gov/pubmed/24564608</a>                         |                |
| 30 | 2014                | Cell Host Microbe   | Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcript            | Wu CH, Chen PJ, Yeh SH,        | research     | PTM              | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25299332">https://www.ncbi.nlm.nih.gov/pubmed/25299332</a>                         |                |
| 31 | 2014                | Cell                | Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.  | Kutluay SB, Zang T, Blanco-I   | research     | CLIP             | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25416948">https://www.ncbi.nlm.nih.gov/pubmed/25416948</a>                         |                |
| 32 | 2015                | Journal of Virology | The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells.  | Cui L, Wang H, Ji Y, Yang J,   | research     | host factors     | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26085159">https://www.ncbi.nlm.nih.gov/pubmed/26085159</a>                         |                |
| 33 | 2015                | Annu Rev Virol.     | Continuous and Discontinuous RNA Synthesis in Coronaviruses.   | Sola I, Almazán F, Zúñiga S,   | review       | template-switchi | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26958916">https://www.ncbi.nlm.nih.gov/pubmed/26958916</a>                         |                |
| 34 | 2015                | Virology            | Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.  | Lloyd RE,                      | review       | host factors     | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25818028">https://www.ncbi.nlm.nih.gov/pubmed/25818028</a>                         |                |
| 35 | 2016                | PLoS Pathogen       | High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling   | Irigoyen N, Firth AE, Jones J  | research     | Translation      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26919232">https://www.ncbi.nlm.nih.gov/pubmed/26919232</a>                         |                |
| 36 | 2016                | Adv Virus Res.      | Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.   | Nakagawa K, Lokugamage K       | review       | Translation      | <a href="https://www.ncbi.nlm.nih.gov/pubmed/27712623">https://www.ncbi.nlm.nih.gov/pubmed/27712623</a>                         |                |

Explore

# Details, details, and details older than 30 years

|    | A                        | B  | C  | D   | E   | F              | G           | H    |
|----|--------------------------|--|--|---|---|----------------|-------------|------|
| 1  | Year of Publication      | Journal  | Title  | Author  | Article type  | Topic          | Pubmed Link | Misc |
| 3  | 1990 Journal of Virology | Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal.  | <b>Makino S</b> , Yokomori K, Lal M research   | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2243386">https://www.ncbi.nlm.nih.gov/pubmed/2243386</a>                           |                |             |      |
| 4  | 1991 Journal of Virology | A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs.  | van der Most RG1, Bredenbeek research          | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2033672">https://www.ncbi.nlm.nih.gov/pubmed/2033672</a>                           |                |             |      |
| 5  | 1992 Journal of Virology | Identification and characterization of a coronavirus packaging signal.   | Fosmire JA1, Hwang K, <b>Maki</b> research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/1316465">https://www.ncbi.nlm.nih.gov/pubmed/1316465</a>                           |                |             |      |
| 6  | 1994 Virology            | Coronavirus translational regulation: leader affects mRNA efficiency.  | Tahara SM, Dietlin TA, Berg research           | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8030227">https://www.ncbi.nlm.nih.gov/pubmed/8030227</a>                           |                |             |      |
| 7  | 1996 Virology            | The production of recombinant infectious Dl-particles of a murine coronavirus in the absence of helper virus.  | Bos EC, Luytjes W, van der M research          | VLP   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8615041">https://www.ncbi.nlm.nih.gov/pubmed/8615041</a>                           | reporter       |             |      |
| 8  | 1998 Adv Exp Med Biol.   | Mouse hepatitis virus nucleocapsid protein as a translational effector of viral mRNAs.   | Tahara SM, Dietlin TA, Nelso research          | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/9782298">https://www.ncbi.nlm.nih.gov/pubmed/9782298</a>                           |                |             |      |
| 9  | 2000 J Gen Virol         | High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA.  | Nelson GW1, Stohlmeyer SA, research            | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10640556">https://www.ncbi.nlm.nih.gov/pubmed/10640556</a>                         |                |             |      |
| 10 | 2000 Journal of Virology | Identification of a bovine coronavirus packaging signal.   | Cologna R, Hogue BG, research                  | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10590153">https://www.ncbi.nlm.nih.gov/pubmed/10590153</a>                         |                |             |      |
| 11 | 2001 Journal of Virology | Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging.  | Narayanan K1, <b>Makino S</b> , research       | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/11533169">https://www.ncbi.nlm.nih.gov/pubmed/11533169</a>                         |                |             |      |
| 12 | 2003 Acta Pharmacol Sin  | Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis.  | Qin L, Xiong B, Luo C, Guo Z research          | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/12791173">https://www.ncbi.nlm.nih.gov/pubmed/12791173</a>                         |                |             |      |
| 13 | 2005 Journal of Virology | Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent.  | Hsieh PK1, Chang SC, Huan research             | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16254430">https://www.ncbi.nlm.nih.gov/pubmed/16254430</a>                         |                |             |      |
| 14 | 2005 Journal of Virology | Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis.  | <b>Sola I</b> , Moreno JL, Zúñiga S, research  | template-switch   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/15681451">https://www.ncbi.nlm.nih.gov/pubmed/15681451</a>                         |                |             |      |
| 15 | 2007 Virology            | Coronavirus nucleocapsid protein is an RNA chaperone.  | Zúñiga S, <b>Sola I</b> , Moreno JL, research  | RNA chaperone   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16979208">https://www.ncbi.nlm.nih.gov/pubmed/16979208</a>                         |                |             |      |
| 16 | 2007 Journal of Virology | New structure model for the packaging signal in the genome of group Ia coronaviruses.  | Chen SC, van den Born E, vi research           | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/17426856">https://www.ncbi.nlm.nih.gov/pubmed/17426856</a>                         |                |             |      |
| 17 | 2009 J Mol Biol          | Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for efficient transcription.                  | I Chang CK, Hsu YL, Chang Y research           | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19052082">https://www.ncbi.nlm.nih.gov/pubmed/19052082</a>                         |                |             |      |
| 18 | 2010 Journal of Virology | Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.   | Zúñiga S, Cruz JL, <b>Sola I</b> , Me research | template switch   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19955314">https://www.ncbi.nlm.nih.gov/pubmed/19955314</a>                         | in vitro templ |             |      |
| 19 | 2010 Springer            | Molecular Biology of the SARS-CoVirus  | Sunil K, Lal book                              | SARS  | <a href="https://link.springer.com/book/10.1007/978-3-642-03683-5">https://link.springer.com/book/10.1007/978-3-642-03683-5</a> |                |             |      |
| 20 | 2011 Journal of Virology | Cellular poly(c) binding protein 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1 $\beta$ and suppose Beura LK, Dinh PX, Osorio F, research | host factors                                   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21976641">https://www.ncbi.nlm.nih.gov/pubmed/21976641</a> | immunoprecip  |                |             |      |
| 21 | 2011 RNA Biol.           | RNA-RNA and RNA-protein interactions in coronavirus replication and transcription.   | <b>Sola I</b> , Mateos-Gomez PA, Al review     | template-switch   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21378501">https://www.ncbi.nlm.nih.gov/pubmed/21378501</a>                         |                |             |      |
| 22 | 2011 PLoS Pathogen       | SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Ind  | Huang C, Lokugamage KG, I research             | nsp1  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22174690">https://www.ncbi.nlm.nih.gov/pubmed/22174690</a>                         |                |             |      |
| 23 | 2011 Journal of Virology | Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis.   | Dufour D, Mateos-Gomez PA research             | template-switch   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21389138">https://www.ncbi.nlm.nih.gov/pubmed/21389138</a>                         |                |             |      |
| 24 | 2011 Journal of Virology | The poly(pyrimidine tract)-binding protein affects coronavirus RNA accumulation levels and relocates viral RNAs to novel cytoplasmic domain  | <b>Sola I</b> , Galán C, Mateos-Gó research    | host factor   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21415158">https://www.ncbi.nlm.nih.gov/pubmed/21415158</a>                         |                |             |      |
| 25 | 2012 J Biol Chem.        | Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleic acid.  | Keane SC, Liu P, Leibowitz J research          | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22241479">https://www.ncbi.nlm.nih.gov/pubmed/22241479</a>                         |                |             |      |
| 26 | 2013 Journal of Virology | The cellular interaction of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology.  | Emmott E, Munday D, Bicker research            | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/23637411">https://www.ncbi.nlm.nih.gov/pubmed/23637411</a>                         | immunoprecip   |             |      |
| 27 | 2014 Viruses             | The coronavirus nucleocapsid is a multifunctional protein  | McBride R, van Zyl M, Fielder review           | N protein   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25105276">https://www.ncbi.nlm.nih.gov/pubmed/25105276</a>                         |                |             |      |
| 28 | 2014 Journal of Virology | Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein  | Kuo L, Koetzner CA, Hurst K research           | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24501403">https://www.ncbi.nlm.nih.gov/pubmed/24501403</a>                         |                |             |      |
| 29 | 2014 J Med Chem.         | Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target.   | Lin SY, Liu CL, Chang YM, Zi research          | structure   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24564608">https://www.ncbi.nlm.nih.gov/pubmed/24564608</a>                         |                |             |      |
| 30 | 2014 Cell Host Microbe   | Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcript  | Wu CH, Chen PJ, Yeh SH, research               | PTM   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25299332">https://www.ncbi.nlm.nih.gov/pubmed/25299332</a>                         |                |             |      |
| 31 | 2014 Cell                | Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.  | <b>Kutluay SB</b> , Zang T, Blanco-I research  | CLIP  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25416948">https://www.ncbi.nlm.nih.gov/pubmed/25416948</a>                         |                |             |      |
| 32 | 2015 Journal of Virology | The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells.  | Cui L, Wang H, Ji Y, Yang J, research          | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26085159">https://www.ncbi.nlm.nih.gov/pubmed/26085159</a>                         |                |             |      |
| 33 | 2015 Annu Rev Virol.     | Continuous and Discontinuous RNA Synthesis in Coronaviruses.   | <b>Sola I</b> , Almazán F, Zúñiga S, review    | template-switch   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26958916">https://www.ncbi.nlm.nih.gov/pubmed/26958916</a>                         |                |             |      |
| 34 | 2015 Virology            | Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.  | Lloyd RE, review                               | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25818028">https://www.ncbi.nlm.nih.gov/pubmed/25818028</a>                         |                |             |      |
| 35 | 2016 PLoS Pathogen       | High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling   | Irigoyen N, <b>Firth AE</b> , Jones J research | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26919232">https://www.ncbi.nlm.nih.gov/pubmed/26919232</a>                         |                |             |      |
| 36 | 2016 Adv Virus Res.      | Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.   | Nakagawa K, Lokugamage K review                | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/27712623">https://www.ncbi.nlm.nih.gov/pubmed/27712623</a>                         |                |             |      |

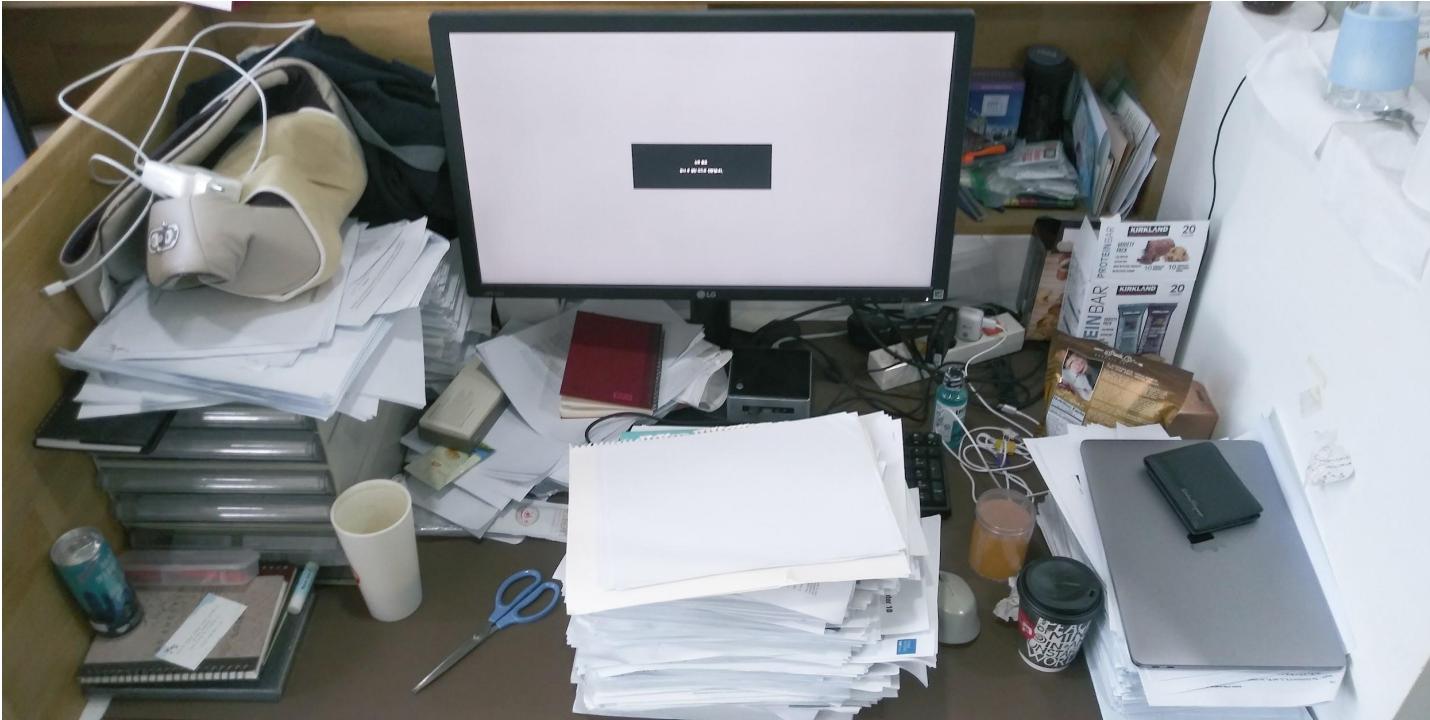
Explore

# Details, details, and details from field-specific journals

|    | A                   | B                   | C  | D                             | E            | F   | G   | H              |
|----|---------------------|---------------------|--|-------------------------------|--------------|---|---|----------------|
| 1  | Year of Publication | Journal             | Title  | Author                        | Article type | Topic   | Pubmed Link   | Misc           |
| 2  | 1990                | Journal of Virology | Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal.            | Makino S, Yokomori K, Lal M   | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2243386">https://www.ncbi.nlm.nih.gov/pubmed/2243386</a>                           |                |
| 3  | 1991                | Journal of Virology | A domain at the 3' end of the polymerase gene is essential for encapsidation of coronavirus defective interfering RNAs.                        | van der Most RG1, Bredenbe    | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/2033672">https://www.ncbi.nlm.nih.gov/pubmed/2033672</a>                           |                |
| 4  | 1992                | Journal of Virology | Identification and characterization of a coronavirus packaging signal.   | Fosmire JA1, Hwang K, Maki    | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/1316465">https://www.ncbi.nlm.nih.gov/pubmed/1316465</a>                           |                |
| 5  | 1994                | Virology            | Coronavirus translational regulation: leader affects mRNA efficiency.  | Tahara SM, Dietlin TA, Berg   | research     | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8030227">https://www.ncbi.nlm.nih.gov/pubmed/8030227</a>                           |                |
| 6  | 1996                | Virology            | The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus.                                  | Bos EC, Luytjes W, van der M  | research     | VLP   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/8615041">https://www.ncbi.nlm.nih.gov/pubmed/8615041</a>                           | reporter       |
| 7  | 1998                | Adv Exp Med Biol.   | Mouse hepatitis virus nucleocapsid protein as a translational effector of viral mRNAs.   | Tahara SM, Dietlin TA, Nelso  | research     | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/9782298">https://www.ncbi.nlm.nih.gov/pubmed/9782298</a>                           |                |
| 8  | 2000                | J Gen Virol.        | High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA.                            | Nelson GW1, Stohlmeyer SA,    | research     | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10640556">https://www.ncbi.nlm.nih.gov/pubmed/10640556</a>                         |                |
| 9  | 2000                | Journal of Virology | Identification of a bovine coronavirus packaging signal.   | Cologna R, Hogue BG,          | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/10590153">https://www.ncbi.nlm.nih.gov/pubmed/10590153</a>                         |                |
| 10 | 2001                | Journal of Virology | Cooperation of an RNA packaging signal and a viral envelope protein in coronavirus RNA packaging.  | Narayanan K1, Makino S,       | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/11533169">https://www.ncbi.nlm.nih.gov/pubmed/11533169</a>                         |                |
| 11 | 2003                | Acta Pharmacol Sin  | Identification of probable genomic packaging signal sequence from SARS-CoV genome by bioinformatics analysis.                                  | Qin L, Xiong B, Luo C, Guo Z  | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/12791173">https://www.ncbi.nlm.nih.gov/pubmed/12791173</a>                         |                |
| 12 | 2005                | Journal of Virology | Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent.            | Hsieh PK1, Chang SC, Huan     | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16254430">https://www.ncbi.nlm.nih.gov/pubmed/16254430</a>                         |                |
| 13 | 2005                | Journal of Virology | Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis.                  | Sola I, Moreno JL, Zúñiga S,  | research     | template-switchi  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/15681451">https://www.ncbi.nlm.nih.gov/pubmed/15681451</a>                         |                |
| 14 | 2007                | Virology            | Coronavirus nucleocapsid protein is an RNA chaperone.  | Zúñiga S, Sola I, Moreno JL,  | research     | RNA chaperone   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/16979208">https://www.ncbi.nlm.nih.gov/pubmed/16979208</a>                         |                |
| 15 | 2007                | Journal of Virology | New structure model for the packaging signal in the genome of group Ia coronaviruses.  | Chen SC, van den Born E, v    | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/17426856">https://www.ncbi.nlm.nih.gov/pubmed/17426856</a>                         |                |
| 16 | 2009                | Journal of Virology | Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications | I Chang CK, Hsu YL, Chang Y   | research     | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19052082">https://www.ncbi.nlm.nih.gov/pubmed/19052082</a>                         |                |
| 17 | 2009                | J Mol Biol.         | Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA c        | Grosoehme NE, Li T, Keane     | research     | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19782089">https://www.ncbi.nlm.nih.gov/pubmed/19782089</a>                         |                |
| 18 | 2010                | Journal of Virology | Grosoehme NE, Li T, Keane research   | Zúñiga S, Cruz JL, Sola I, M  | research     | template switchi  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/19955314">https://www.ncbi.nlm.nih.gov/pubmed/19955314</a>                         | in vitro templ |
| 19 | 2010                | Springer            | Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.                                   | Sunil K, Lal                  | book         | SARS  | <a href="https://link.springer.com/book/10.1007/978-3-642-03683-5">https://link.springer.com/book/10.1007/978-3-642-03683-5</a> |                |
| 20 | 2011                | Journal of Virology | Molecular Biology of the SARS-CoVirus  | Beura LK, Dinh PX, Osorio F,  | research     | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21976641">https://www.ncbi.nlm.nih.gov/pubmed/21976641</a>                         | immunoprecip   |
| 21 | 2011                | RNA Biol.           | Cellular poly(c) binding protein 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and supp   | Sola I, Mateos-Gómez PA,      | review       | template-switchi  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21378501">https://www.ncbi.nlm.nih.gov/pubmed/21378501</a>                         |                |
| 22 | 2011                | PLoS Pathogen       | RNA-RNA and RNA-protein interactions in coronavirus replication and transcription.   | AI review                     | nsp1         | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22174690">https://www.ncbi.nlm.nih.gov/pubmed/22174690</a> |   |                |
| 23 | 2011                | Journal of Virology | SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Ind              | Huang C, Lokugamage KG, I     | research     | template-switchi  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21389138">https://www.ncbi.nlm.nih.gov/pubmed/21389138</a>                         |                |
| 24 | 2011                | Journal of Virology | Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis.                 | Dufour D, Mateos-Gómez PA     | research     | host factor   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/21415158">https://www.ncbi.nlm.nih.gov/pubmed/21415158</a>                         |                |
| 25 | 2012                | J Biol Chem.        | The poly(pirimidine tract-binding protein affects coronavirus RNA accumulation levels and relocates viral RNAs to novel cytoplasmic domai      | Sola I, Galán C, Mateos-Gó    | research     | RNA-binding   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/22241479">https://www.ncbi.nlm.nih.gov/pubmed/22241479</a>                         |                |
| 26 | 2013                | Journal of Virology | Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucle    | Keane SC, Liu P, Leibowitz J  | research     | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/23637411">https://www.ncbi.nlm.nih.gov/pubmed/23637411</a>                         | immunoprecip   |
| 27 | 2014                | Viruses             | Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucle    | Emmott E, Munday D, Bicker    | research     | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/23637411">https://www.ncbi.nlm.nih.gov/pubmed/23637411</a>                         |                |
| 28 | 2014                | Journal of Virology | The coronavirus nucleocapsid is a multifunctional protein  | McBride R, van Zyl M, Fielder | review       | N protein   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25105276">https://www.ncbi.nlm.nih.gov/pubmed/25105276</a>                         |                |
| 29 | 2014                | J Med Chem.         | Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid prot           | Kuo L, Koetzner CA, Hurst K   | research     | packaging   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24501403">https://www.ncbi.nlm.nih.gov/pubmed/24501403</a>                         |                |
| 30 | 2014                | Cell Host Microbe   | Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target.                   | Liu SY, Liu CL, Chang YM, Zi  | research     | structure   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/24564608">https://www.ncbi.nlm.nih.gov/pubmed/24564608</a>                         |                |
| 31 | 2014                | Cell                | Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcrip       | Wu CH, Chen PJ, Yeh SH,       | research     | PTM   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25299332">https://www.ncbi.nlm.nih.gov/pubmed/25299332</a>                         |                |
| 32 | 2015                | Journal of Virology | Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.  | Kutluay SB, Zang T, Blanco-I  | research     | CLIP  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25416948">https://www.ncbi.nlm.nih.gov/pubmed/25416948</a>                         |                |
| 33 | 2015                | Annu Rev Virol.     | The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells.                                      | Cui L, Wang HJ, Yang J,       | research     | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26085159">https://www.ncbi.nlm.nih.gov/pubmed/26085159</a>                         |                |
| 34 | 2015                | Virology            | Continuous and Discontinuous RNA Synthesis in Coronaviruses.   | Sola I, Almazán F, Zúñiga S,  | review       | template-switchi  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26958916">https://www.ncbi.nlm.nih.gov/pubmed/26958916</a>                         |                |
| 35 | 2016                | PLoS Pathogen       | Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.  | Lloyd RE,                     | review       | host factors  | <a href="https://www.ncbi.nlm.nih.gov/pubmed/25818028">https://www.ncbi.nlm.nih.gov/pubmed/25818028</a>                         |                |
| 36 | 2016                | Adv Virus Res.      | High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling   | Irigoyen N, Firth AE, Jones J | research     | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/26919232">https://www.ncbi.nlm.nih.gov/pubmed/26919232</a>                         |                |
|    |                     |                     | Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.   | Nakagawa K, Lokugamage K      | review       | Translation   | <a href="https://www.ncbi.nlm.nih.gov/pubmed/27712623">https://www.ncbi.nlm.nih.gov/pubmed/27712623</a>                         |                |

Explore

(Please do not copy or share)



BiS800: Methods in functional genomics and  
computational molecular biology

You are what you read and write!

# 10 minute break

# Corrections and shameless advertisement

Office hours (until 6/17):

Tuesday 3:00 pm - 4:30 pm

Thursday 10:30 am - 12:00 pm



**2021 IBS-SNU  
MINI-SYMPOSIA ON  
RNA BIOLOGY  
& THERAPEUTICS**

The Center for RNA Research at the Institute for Basic Science (IBS) is hosting a series of 6 virtual mini-symposia titled 'RNA Biology and Therapeutics'. Each 2.5hr mini-symposium will consist of leading research presented by both senior and young scientists.

**TIME SCHEDULE \_2021**

| Seoul            | Berlin            | Boston           | LA               |
|------------------|-------------------|------------------|------------------|
| Jun.13<br>9:00am | Jun.13<br>10:00am | Jun.12<br>7:00pm | Jun.12<br>4:00pm |
| Feb.17<br>9:00am | Feb.17<br>10:00am | Feb.16<br>7:00pm | Feb.16<br>4:00pm |
| Mar.30<br>9:00pm | Mar.30<br>10:00pm | Mar.30<br>9:00pm | Mar.30<br>9:00pm |
| Jun.2<br>9:00pm  | Jun.2<br>10:00pm  | Jun.2<br>9:00pm  | Jun.2<br>10:00pm |
| Jul.14<br>9:00am | Jul.14<br>10:00am | Jul.13<br>8:00pm | Jul.13<br>5:00pm |
| Aug.11<br>9:00am | Aug.11<br>10:00am | Aug.10<br>8:00pm | Aug.10<br>5:00pm |

**WEBSITE**  
[narrykim.org/en](http://narrykim.org/en)

**ORGANIZERS**  
V. Narry Kim, Jin-Hong Kim, Yoosik Kim & Young-Suk Lee

**SPEAKERS**  
Joan Shitrit, Yale University  
Yoosik Kim, KAIST  
Sun Hur, Harvard Medical School  
Howard Y. Chang, Stanford University  
Joshua Mendell, UT Southwestern Medical Center  
Miao-Chih Tsai, Cell Press  
Ling-Ling Chen, CAS  
Naem Stern-Ginossar, Weizmann Institute of Science  
Yonsei Park, National Cancer Center  
Heeyoung Seo, Korea University  
David Bartel, MIT  
Chengqi Yu, Peking University  
Olivia Rissland, University of Colorado  
Jin-Hong Kim, IBS & Seoul National University

**FREE**

 Institute for Basic Science  SEOUL NATIONAL UNIVERSITY



# Ground rules

1. No plagiarism. If you discussed the assignments with someone, please mention that in your assignments
2. Course materials will be available on KLMS
3. All questions regarding the logistics should be directed to the TA
4. All questions regarding the content should be directed to the instructor
5. The Zoom call will be recorded but not distributed

# What this course is NOT

1. Participation is 25%, but this does not include attendance. This is a graduate course, so I expect everyone to be committed to this course based on their own schedule.
2. This is not a programming course. Nevertheless, you are expected to read/write/present mathematical and computational details in your assignments.
3. This is not an english course. While effective communication in english is required, the assignment will not be graded based on grammar, style, and other aspects important in literature.

# What this course IS

1. Participation is 25%. This is again a graduate course, so I expect everyone to actively participant in the discussions.
2. The R programming language is used to convey computational ideas in this course, especially concepts in data science.
3. This course includes a decent amount of reading and writing. While the emphasis of this course is in regards to the method section, you might end up reading the entire paper and also related work to complete the assignments.

Read and be prepared in advance!

# For those auditing (everything goes through TA)

1. 100% attendance and active participation is required
2. Written assignment to TA via email (Due 5/25)
3. Participate in the mock peer-review
4. Volunteer as a scribe for this course

Especially for those auditing, please do not hesitate to  
express your appreciation for his service

# Scribe notes

- Taking notes is important, but engaging into the discussion is more
- Write and summarize the materials and discussions for that lecture
- Scribe notes must be submitted **to TA** before the next lecture
- Scribe notes will be made available on KLMS
- Karma points!
- The goal is to assign a scribe for every lecture
  - Week 2: 3/9, 11
  - Week 3: 3/16, 18
  - Week 6: 4/6, 8
  - Week 7: 4/13, 15

# Written assignments (mock peer review)

Structured response based on the reading material which includes:

- One paragraph summary of reading material
- Description of what part of the method you appreciated and what you considered limitations
- Lastly, the impact and meaning of the computational work

Written assignment will be anonymously graded (largely) based on three peer reviews under the Honor system.

For each category above, assign a score from 1 to 3 and a short description of why you assigned that score.

# For example for this week on “T-Rex the chicken?”

Submit peer review of written response before the next class starts

| Sunday | Monday | Tuesday       | Wednesday | Thursday   | Friday | Saturday   |    |    |    |    |    |    |    |    |    |  |
|--------|--------|---------------|-----------|------------|--------|--|----|----|----|----|----|----|----|----|----|--|
|        | 1      | 2             | 3         | 4          | 5      | 6  |    |    |    |    |    |    |    |    |    |  |
| 7      | 8      | 9             | 10        | 11         | 12     | 13   |    |    |    |    |    |    |    |    |    |  |
| 14     | 15     | 16            | 17        | 18         | 19     | 20   |    |    |    |    |    |    |    |    |    |  |
| 21     | 22     | 23            | 24        | 25         | 26     | 27   |    |    |    |    |    |    |    |    |    |  |
| 28     | 29     | 30            | 31        |            |        |  |    |    |    |    |    |    |    |    |    |  |
|        |        | February 2021 |           | April 2021 |        | Calendars by Vertex42.com<br>© 2018 Vertex42 LLC. Free to print. |    |    |    |    |    |    |    |    |    |  |
|        |        | Su            | M         | Tu         | W      | Th   | F  | Sa | Su | M  | Tu | W  | Th | F  | Sa |  |
|        |        | 1             | 2         | 3          | 4      | 5  | 6  |    | 1  | 2  | 3  | 4  | 5  | 6  | 7  |  |
|        |        | 8             | 9         | 10         | 11     | 12   | 13 | 4  | 11 | 12 | 13 | 14 | 15 | 16 | 17 |  |
|        |        | 14            | 15        | 16         | 17     | 18   | 19 | 20 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |  |
|        |        | 21            | 22        | 23         | 24     | 25   | 26 | 27 | 25 | 26 | 27 | 28 | 29 | 30 |    |  |
|        |        | 28            |           |            |        |  |    |    |    |    |    |    |    |    |    |  |

Created using the Vertex42 Calendar Template for Excel

<http://www.vertex42.com/calendars/printable-calendars.html>

# First half: general principles

Week 1: Introduction and T-Rex the chicken?

Week 2: Data science

Week 3: Statistics

Week 4: Student data

Week 5: Student method

Week 6: Machine learning

Week 7: Applied machine learning

# First half: general principles, textbooks

Week 1: Introduction and T-Rex the chicken?

Week 2: Data science

Week 3: Statistics

Week 4: Student data

Week 5: Student method

Week 6: Machine learning

Week 7: Applied machine learning

# First half: general principles, your contributions

Week 1: Introduction and T-Rex the chicken?

Week 2: Data science

Week 3: Statistics

Week 4: Student data

Week 5: Student method

Week 6: Machine learning

Week 7: Applied machine learning

# Your introduction to the class

Week 4: Student data

- Biological question or subject matter

Week 5: Student method

- Computing techniques or mathematical concept

★ Basis of your final report ★

Come to office hours for help!

# Second half: oral presentations, published work

Week 9: High-dimensional data

Week 10: RNA-Seq

Week 11: Single-cell RNA-Seq

Week 12: Public data

Week 13: RNA biology (Response is extra credit)

Read all for that week, but pick one to write response

# Second half: oral presentations, (un)published work

Week 9: High-dimensional data

Week 10: RNA-Seq

Week 11: Single-cell RNA-Seq

Week 12: Public data

Week 13: RNA biology

Week 14: From future biologist

Week 15: From future engineers

# One final report is due May 27

(Re)write your own method section

Encourage to share unpublished work but introducing published work is completely acceptable

In general based on the data and method discussed in Week 4/5

But due before final presentations in Week 14/15

# Summary

Written assignments:

Due 4/27, 5/4, 5/11, 5/18, (5/25)

Oral presentations:

Due 3/23, 4/6, published (TBD), unpublished (TBD)

One final report:

Due 5/27