Why is the computational method section so important

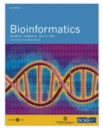
2021.03.02

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

Bioinformatics

ISCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Issues | Advance articles | Submit ▼ | Purchase | Alerts | About ▼ | All Bioinformatics | Advanced Search



Volume 25, Issue 14
15 July 2009

Article Contents

< Previous Next >

Fast and accurate short read alignment with Burrows–Wheeler transform

Heng Li, Richard Durbin Author Notes

Bioinformatics, Volume 25, Issue 14, 15 July 2009, Pages 1754–1760,
https://doi.org/10.1093/bioinformatics/btp324
Published: 18 May 2009 **Article history ▼**

PDF | Split View | Cite | Permissions | Share ▼

Abstract

Motivation: The enormous amount of short reads generated by the new DNA sequencing technologies call for the development of fast and accurate read alignment programs. A first generation of hash table-based methods has been developed, including MAQ, which is accurate, feature rich and fast enough to align short reads from a single individual. However, MAQ does not support gapped alignment for single-end reads, which makes it unsuitable for alignment of longer reads where indels may occur frequently. The speed of MAQ is also a concern when the alignment is scaled up to the resequencing of hundreds of individuals.

Results: We implemented Burrows-Wheeler Alignment tool (BWA), a new read alignment package that is based on backward search with Burrows–Wheeler Transform (BWT), to efficiently align short sequencing reads against a large reference sequence such as the human genome, allowing mismatches and gaps. BWA supports both base space reads, e.g. from Illumina sequencing machines, and color space reads from AB SOLiD machines. Evaluations on both simulated

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 <u>concise</u> 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

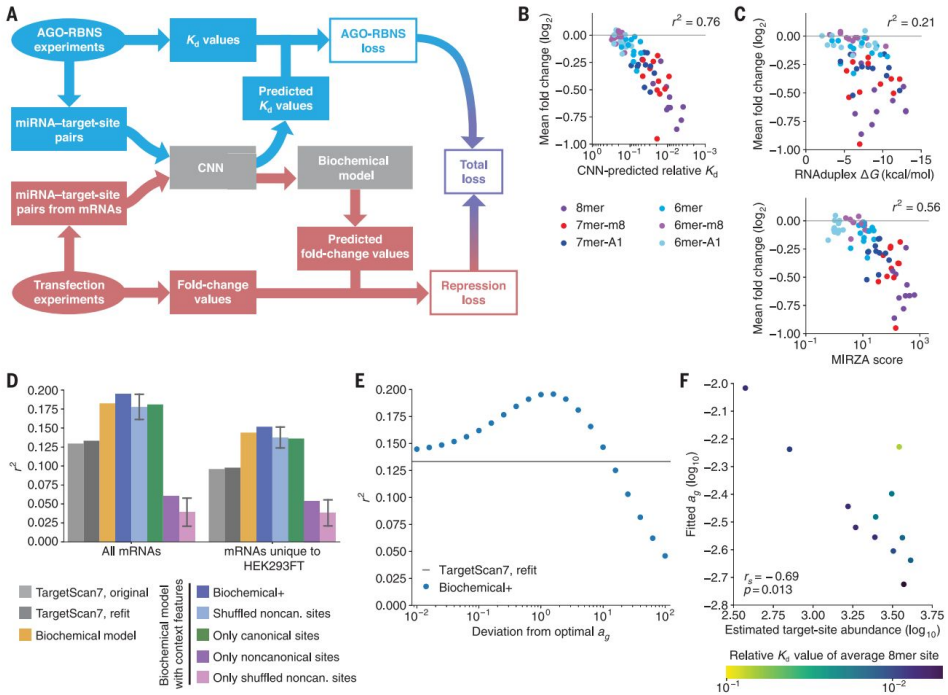
Format for research papers in *Bioinformatics*

- 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

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 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 (*22*). 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 targeting. In the framework of our model, the fitted value of 1.8 observed for the parameter b suggested that a typical mRNA bound to an average of one silencing complex will experience a near tripling of its decay rate, which would lead to a ~60% reduction in its abundance. In the concentration regimes of our transfection experiments, this occupancy can be achieved with two to three median 7mer-m8 sites. In addition, our fitted value for the ORF-site penalty suggested that the translation machinery reduces site affinity by 5.5-fold.

Another parameter was a_0, that is, the intracellular concentration of AGO loaded with the transfected miRNA and not bound to a target site. Whereas values of the other parameters could be fit globally in HeLa cells and 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 allowed to deviate from the fitted values, the biochemical+ model still outperformed TargetScan in predicting test-set repression over a 100-fold range of values (Fig. 6K), which indicated that even with rough estimates of miRNA abundances, our modeling framework had an advantage over other predictive methods in new contexts. Information that might be used to more accurately estimate a_0 values should come with the determination of these values for more miRNAs in more cellular contexts, together with the observation that, as expected (*29*, *40*), fitted a_0 values are higher for miRNAs with lower predicted target abundance and lower general affinity for their targets (Fig. 6F).

Our work replaced the correlative models of targeting efficacy with a principled biochemical model that explains and predicts about half of the variability attributable to the direct 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 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, leveraging substantially expanded repression datasets as well as site-affinity values. In this way, the complete regulatory landscape, as specified by AGO within this essential biological pathway, might ultimately be computationally reconstructed.

Methods summary

AGO2-miRNA complexes were generated by adding synthetic miRNA duplexes to lysate from cells that overexpressed recombinant AGO2, and then these complexes were purified on the basis of affinity to the miRNA seed. RNA libraries were generated by in vitro transcription of synthetic DNA templates. For AGO-RBNS, purified AGO2-miRNA complex was incubated with a large excess of library molecules, and after reaching binding equilibrium, library molecules bound to AGO2-miRNA complex were isolated 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 simultaneously determined by maximum likelihood estimation, fitting to AGO-RBNS results obtained over a 100-fold range in AGO2-miRNA concentration.

Intracellular miRNA-mediated repression was measured by performing RNA-seq on HeLa cells that had been transfected with a synthetic miRNA duplex. For sites that were sufficiently abundant in endogenous 3′UTRs, efficacy was measured on the basis of their influence on levels of endogenous mRNAs of HeLa cells. Site efficacy was also evaluated using massively parallel reporter assays, which provided information for the rare sites as well as the more abundant ones. The biochemical and biochemical+ models of miRNA-mediated repression were constructed and fit using the measured K_d values, and the repression of endogenous mRNAs was observed after transfecting miRNAs into HeLa cells. The CNN was built using TensorFlow, trained using the measured 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

McGeary *et al.*, *Science* **366**, eaav1741 (2019) 20 December 2019 12 of 13

McGeary and Lin et al. 2020 from Bartel Lab

BiS800: Methods in functional genomics and computational molecular biology

Or supplementary materials ...

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 target 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, levers 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; pmid: 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; pmid: 26077373
3. D. P. Bartel, MicroRNAs: Target recognition and regulatory functions. *Cell* **136**, 215–233 (2009). doi: 10.1016/j.cell.2009.01.002; pmid: 19167326

McGeary and Lin et al. 2020 from Bartel Lab

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) the RT primer was designed to reverse transcribe the variable 3′ UTR region of the reporter library and add homology to the 3′ PCR primer used for small RNA-seq library preparation (Data S1), 2) the volumes of the RT 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 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 conditions (the six miRNA duplex transfections and the mock transfection) were performed in duplicate, and the fourteen samples were sequenced with multiplexing on two 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 read was used if it satisfied the following criteria: 1) it passed the Illumina chastity 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 (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/insertion/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 purifications

7

McGeary and Lin et al. 2020 from Bartel Lab

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) the RT primer was designed to reverse transcribe the variable 3′ UTR region of the reporter library and add homology to the 3′ PCR primer used for small RNA-seq library preparation (Data S1), 2) the volumes of the RT 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 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 conditions (the six miRNA duplex transfections and the mock transfection) were performed in duplicate, and the fourteen samples were sequenced with multiplexing on two 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 read was used if it satisfied the following criteria: 1) it passed the Illumina chastity 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 (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/insertion/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 purifications

7

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

$$f_{grad}(\boldsymbol{\theta}) = \nabla f_{cost}(\boldsymbol{x}(\boldsymbol{\theta})). \tag{11.1.2}$$

The optimization routine was performed with the *optim* function in R (*53*) using the L-BFGS-B method, supplying both $f_{cost}(\boldsymbol{x})$ and $f_{grad}(\boldsymbol{x})$ 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_d values per AGO-RBNS experiment.

11.2 Derivation of $\boldsymbol{x}(\boldsymbol{\theta})$

The function $\boldsymbol{x}(\boldsymbol{\theta})$ 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 mth 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 $\boldsymbol{l} = (l_1, \ldots, l_m)$, the total concentration of AGO–miRNA complex (hereafter referred to as "AGO") in each binding reaction $\boldsymbol{a} = (a_1, \ldots, a_n)$, the K_d value describing the binding between AGO and each site type $\boldsymbol{K} = (K_1, \ldots, K_m)$, 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 $\boldsymbol{l}$ is estimated using

$$\boldsymbol{l} = \frac{\boldsymbol{y}^i}{\sum_{i=1}^{m} y_j^i} \times 100 \text{ nM}, \tag{11.2.1}$$

where $\boldsymbol{y}^i$ 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 $\boldsymbol{a}$ is calculated from the experimentally determined dilution series

$$\begin{aligned} \boldsymbol{a} &= a \times \boldsymbol{s} \\ &= a \times (0.4\%,\ 1.27\%,\ 4\%,\ 12.7\%,\ 40\%), \end{aligned} \tag{11.2.2}$$

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

$$(K_1, K_2, \ldots, K_m, a, b). \tag{11.2.3}$$

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

10

McGeary and Lin et al. 2020 from Bartel Lab

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) the RT primer was designed to reverse transcribe the variable 3' UTR region of the reporter library and add homology to the 3' PCR primer used for small RNA-seq library preparation (Data S1), 2) the volumes of the RT 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 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 conditions (the six miRNA duplex transfections and the mock transfection) were performed in duplicate, and the fourteen samples were sequenced with multiplexing on two 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 read was used if it satisfied the following criteria: 1) it passed the Illumina chastity 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 (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/insertion/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 purifications

7

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

$$f_{\text{grad}}(\boldsymbol{\theta}) = \nabla f_{\text{cost}}(\boldsymbol{x}(\boldsymbol{\theta})). \tag{11.1.2}$$

The optimization routine was performed with the *optim* function in R (*53*) using the L-BFGS-B method, supplying both $f_{\text{cost}}(\boldsymbol{x})$ and $f_{\text{grad}}(\boldsymbol{x})$ 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_d values per AGO-RBNS experiment.

11.2 Derivation of $\boldsymbol{x}(\boldsymbol{\theta})$

The function $\boldsymbol{x}(\boldsymbol{\theta})$ 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 mth 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 $\boldsymbol{l} = (l_1, \ldots, l_m)$, the total concentration of AGO–miRNA complex (hereafter referred to as "AGO") in each binding reaction $\boldsymbol{a} = (a_1, \ldots, a_n)$, the K_d value describing the binding between AGO and each site type $\boldsymbol{K} = (K_1, \ldots, K_m)$, 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 $\boldsymbol{l}$ is estimated using

$$\boldsymbol{l} = \frac{\boldsymbol{y}^i}{\sum_{i=1}^{m} y_i^j} \times 100 \text{ nM}, \tag{11.2.1}$$

where $\boldsymbol{y}^i$ 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 $\boldsymbol{a}$ is calculated from the experimentally determined dilution series

$$\begin{aligned} \boldsymbol{a} &= a \times \boldsymbol{s} \\ &= a \times (0.4\%,\ 1.27\%,\ 4\%,\ 12.7\%,\ 40\%), \end{aligned} \tag{11.2.2}$$

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

$$(K_1,\ K_2,\ \ldots,\ K_m,\ a,\ b). \tag{11.2.3}$$

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

10

(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:

$$a_j = a_j^f + \sum_{i=1}^{m} c_{ij}. \tag{11.2.15}$$

Because each c_{ij} value is itself a function of $\boldsymbol{l}$, $\boldsymbol{K}$, and a according to equation (11.2.12), 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 $\boldsymbol{x}$ is calculated during the optimization routine requires that a_j^f first be numerically approximated by finding the root of

$$f(a_j^f) = as_j - a_j^f - \sum_{i=1}^{m} \frac{l_i a_j^f}{a_j^f + K_i} \tag{11.2.16}$$

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

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

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

$$\begin{aligned} f_{\text{cost}}(\boldsymbol{x}) &= -\ln \prod_{j=1}^{n} f_{\text{mnm}}(\boldsymbol{y}_j, \boldsymbol{\pi}_j) \\ &= -\ln \prod_{j=1}^{n} \frac{Y_j! \prod_{i=1}^{m} \pi_{ij}^{y_{ij}}}{\prod_{i=1}^{m} y_{ij}!}, \end{aligned} \tag{11.3.1}$$

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

$$f_{\text{cost}}(\boldsymbol{x}) = \sum_{j=1}^{n} \left(Y_j \ln X_j - \sum_{i=1}^{m} y_{ij} \ln x_{ij} + \sum_{i=1}^{m} \ln y_{ij}! - \ln Y_j! \right). \tag{11.3.2}$$

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

13

McGeary and Lin et al. 2020 from Bartel Lab

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) the RT primer was designed to reverse transcribe the variable 3′ UTR region of the reporter library and add homology to the 3′ PCR primer used for small RNA-seq library preparation (Data S1), 2) the volumes of the RT 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 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 conditions (the six miRNA duplex transfections and the mock transfection) were performed in duplicate, and the fourteen samples were sequenced with multiplexing on two lanes of an Illumina HiSeq 2500 run in rapid mode with 100-nt single-end reads. For analysis, reads were first subjected to qua 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 sequen sequence, counts were normalized to the total number of perfectly matching coun 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 pa Illumina chastity filter, as indicated by the presence of the number 1 rather than 0 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 (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/insertion/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 purifications

7

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

$$f_{\mathrm{grad}}(\boldsymbol{\theta}) = \nabla f_{\mathrm{cost}}(\mathbf{x}(\boldsymbol{\theta})). \tag{11.1.2}$$

The optimization routine was performed with the *optim* function in R (5.5.5) using the L-BFGS-B method, supplying both $f_{\mathrm{cost}}(\mathbf{x})$ and $f_{\mathrm{grad}}(\mathbf{x})$ 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_d values per AGO-RBNS experiment.

11.2 Derivation of $\mathbf{x}(\boldsymbol{\theta})$

The function $\mathbf{x}(\boldsymbol{\theta})$ 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

five samples and therefore given by a single parameter. The vector $\boldsymbol{l}$ is estimated using

$$l = \frac{\mathbf{y}^i}{\sum_{i=1}^{n} y_i^j} \times 100\ \mathrm{nM}, \tag{11.2.1}$$

where $\mathbf{y}^i$ 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 $\boldsymbol{a}$ is calculated from the experimentally determined dilution series

$$\begin{aligned} \boldsymbol{a} &= a \times \mathbf{s} \\ &= a \times (0.4\%,\ 1.27\%,\ 4\%,\ 12.7\%,\ 40\%), \end{aligned} \tag{11.2.2}$$

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

$$(K_1,\ K_2,\ \ldots,\ K_m,\ a,\ b). \tag{11.2.3}$$

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

10

(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:

$$a_j = a_j^f + \sum_{i=1}^{m} c_{ij}. \tag{11.2.15}$$

Because each c_{ij} value is itself a function of $\boldsymbol{l}$, $\boldsymbol{K}$, and a according to equation (11.2.12), 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 $\mathbf{x}$ is calculated during the optimization routine requires that a_j^f first be numerically approximated by finding the root of

$$f(a_j^f) = as_j - a_j^f - \sum_{i=1}^{m} \frac{l_i a_j^f}{a_j^f + K_i} \tag{11.2.16}$$

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

e product of the negative log multinomial

$$\begin{aligned} &-\ln \prod_{j=1}^{n} f_{\mathrm{mult}}(\mathbf{y}_j, \boldsymbol{\pi}_j) \\ &= -\ln \prod_{j=1}^{n} \frac{Y_j!\prod_{i=1}^{m} \pi_{ij}^{y_{ij}}}{\prod_{i=1}^{m} y_{ij}!}, \end{aligned} \tag{11.3.1}$$

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

$$f_{\mathrm{cost}}(\mathbf{x}) = \sum_{j=1}^{n} \left(Y_j \ln X_j - \sum_{i=1}^{m} y_{ij} \ln x_{ij} + \sum_{i=1}^{m} \ln y_{ij}! - \ln Y_j! \right). \tag{11.3.2}$$

After discarding the third and fourth terms in equation (11.3.2) because they do not contain any terms of $\mathbf{x}_{j,i}$ and are therefore not related to the MLE estimation of $\boldsymbol{\theta}$, the final cost function is given by

13

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

McGeary and Lin et al. 2020 from Bartel Lab

Method section is seemingly the most <u>unimportant</u> 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 le

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 <u>read/write</u> the method section of papers from a specific field, then you're a <u>professional</u>.

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[1,2,4], Young-suk Lee[1,2,4], Yeon Choi[1,2], Ahyeon Son[1,2], Youngran Park[1,2], Kyung-Min Lee[3], Jeesoo Kim[1,2], Jong-Seo Kim[1,2], and V. Narry Kim[1,2,*]

Affiliations

[1]Center for RNA Research, Institute for Basic Science, Seoul 08826, Republic of Korea.

[2]School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea.

[3]International Vaccine Institute, Seoul 08826, Republic of Korea.

[4]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

1

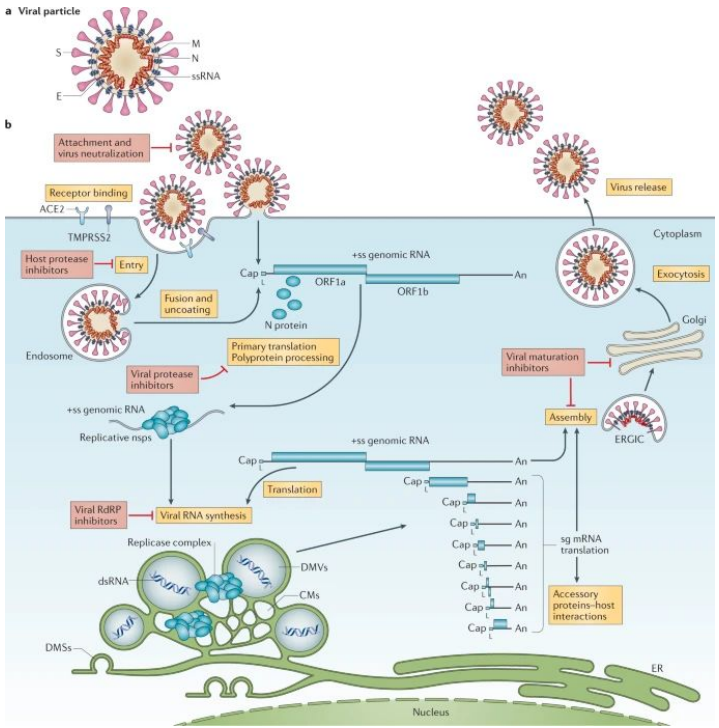
Lee and Lee et al. (2021) Under review

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

Hypothesis:

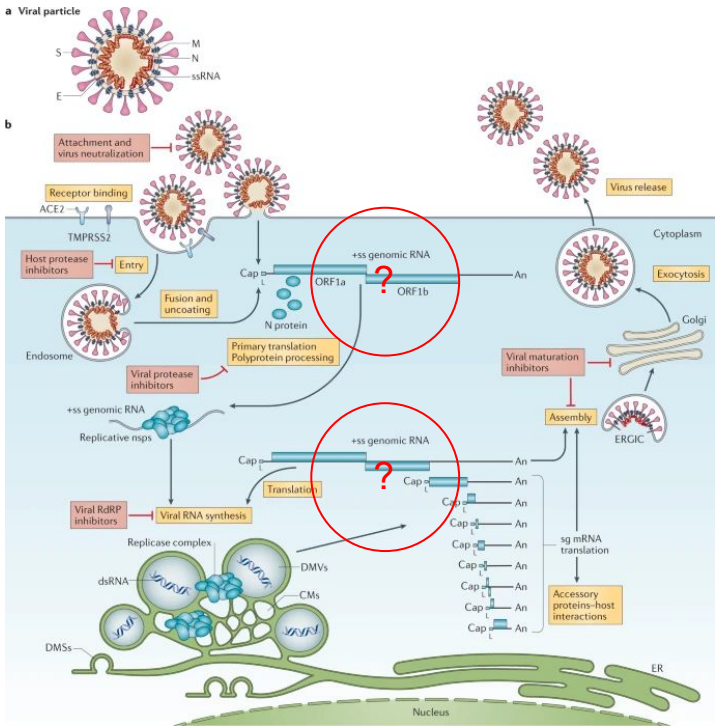
1. Solely computation work of SARS-CoV-2 data
2. Passenger author and no major contribution to the field
3. Corresponding author is a long-standing SARS-CoV-2 expert

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



Philip V'kovski et al. (2021) Review

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

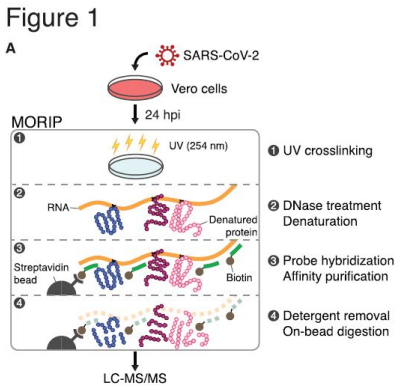


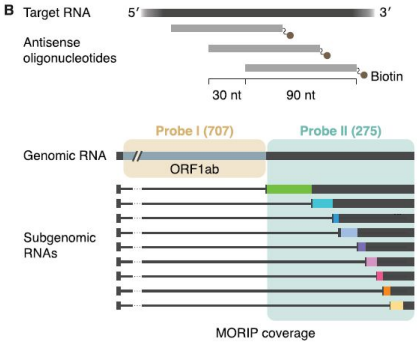
Philip V'kovski et al. (2021) Review

Development of biochemical and computational methods

Figure 1

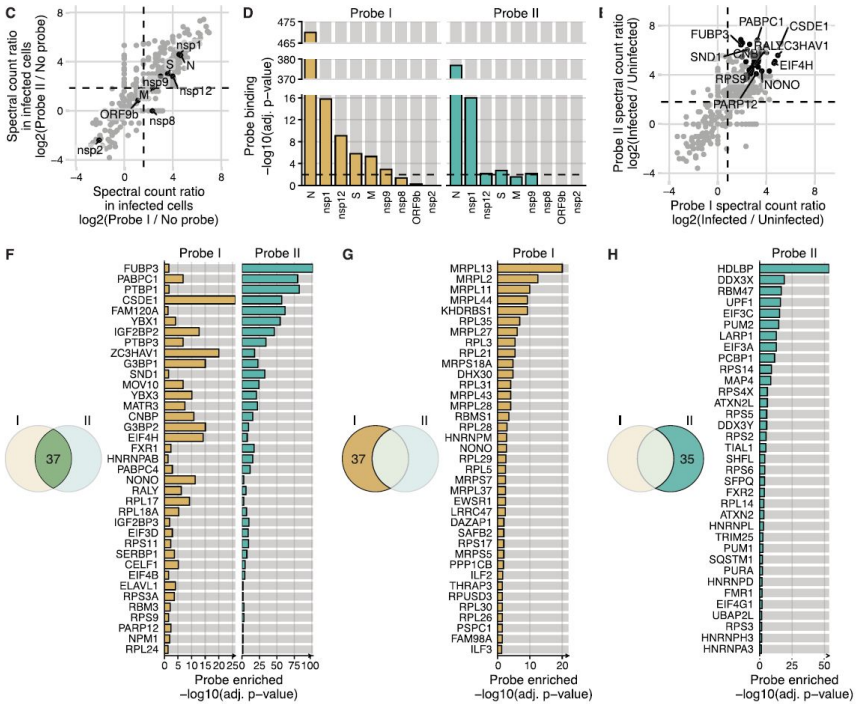
Click here to access/download;Figure;Figure 1.pdf





Lee and Lee et al. (2021) Under review

Identification of proteins that regulate SARS-CoV-2 RNA



Lee and Lee et al. (2021) Under review

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, Lai M	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/2243386	
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	https://www.ncbi.nlm.nih.gov/pubmed/2033672	
4	1992	Journal of Virology	Identification and characterization of a coronavirus packaging signal.	Fosmire JA1, Hwang K, Maki	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/1316465	
5	1994	Virology	Coronavirus translational regulation: leader affects mRNA efficiency.	Tahara SM, Dietlin TA, Bergn	research	Translation	https://www.ncbi.nlm.nih.gov/pubmed/8030227	
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	https://www.ncbi.nlm.nih.gov/pubmed/8615041	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	https://www.ncbi.nlm.nih.gov/pubmed/9782298	
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	https://www.ncbi.nlm.nih.gov/pubmed/10640556	
9	2000	Journal of Virology	Identification of a bovine coronavirus packaging signal.	Cologna R, Hogue BG.	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/10590153	
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	https://www.ncbi.nlm.nih.gov/pubmed/11533169	
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	https://www.ncbi.nlm.nih.gov/pubmed/12791173	
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	https://www.ncbi.nlm.nih.gov/pubmed/16254320	
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	https://www.ncbi.nlm.nih.gov/pubmed/15681451	
14	2007	Virology	Coronavirus nucleocapsid protein is an RNA chaperone.	Zúñiga S, Sola I, Moreno JL,	research	RNA chaperone	https://www.ncbi.nlm.nih.gov/pubmed/16979208	
15	2007	Journal of Virology	New structure model for the packaging signal in the genome of group IIa coronaviruses.	Chen SC, van den Born E, va	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/17428856	
16	2009	Journal of Virology	Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications f	Chang CK, Hsu YL, Chang Y	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19052082	
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	Grossoehme NE, Li L, Keane	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19782089	
18	2010	Journal of Virology	Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.	Zúñiga S, Cruz JL, Sola I, Ma	research	template switchi	https://www.ncbi.nlm.nih.gov/pubmed/19955314	in vitro templa
19	2010	Springer	Molecular Biology of the SARS-Coronavirus	Sunil K. Lal	book	SARS	https://link.springer.com/book/10.1007/978-3-642-03683-5	
20	2011	Journal of Virology	Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and suppo	Beura LK, Dinh PX, Osorio F	research	host factors	https://www.ncbi.nlm.nih.gov/pubmed/21976648	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	https://www.ncbi.nlm.nih.gov/pubmed/21378501	
22	2011	PLoS Pathogen	SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Indu	Huang C, Lokugamage KG, F	research	nsp1	https://www.ncbi.nlm.nih.gov/pubmed/22174690	
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	https://www.ncbi.nlm.nih.gov/pubmed/21389138	
24	2011	Journal of Virology	The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domain	Sola I, Galán C, Mateos-Gón	research	host factor	https://www.ncbi.nlm.nih.gov/pubmed/21411518	
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	https://www.ncbi.nlm.nih.gov/pubmed/22241479	
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	https://www.ncbi.nlm.nih.gov/pubmed/23637410	immunoprecip
27	2014	Viruses	The coronavirus nucleocapsid is a multifunctional protein	McBride R, van Zyl M, Fieldin	review	N protein	https://www.ncbi.nlm.nih.gov/pubmed/25105276	
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	https://www.ncbi.nlm.nih.gov/pubmed/24501403	
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, Z	research	structure	https://www.ncbi.nlm.nih.gov/pubmed/24564608	
30	2014	Cell Host Microbe	Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcrip	Wu CH, Chen PJ, Yeh SH.	research	PTM	https://www.ncbi.nlm.nih.gov/pubmed/25299332	
31	2014	Cell	Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.	Kutluay SB, Zang T, Blanco-M	research	CLIP	https://www.ncbi.nlm.nih.gov/pubmed/25416948	
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	https://www.ncbi.nlm.nih.gov/pubmed/26085159	
33	2015	Annu Rev Virol.	Continuous and Discontinuous RNA Synthesis in Coronaviruses.	Sola I, Almazán F, Zúñiga S,	review	template-switchi	https://www.ncbi.nlm.nih.gov/pubmed/26958916	
34	2015	Virology	Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.	Lloyd RE.	review	host factors	https://www.ncbi.nlm.nih.gov/pubmed/25818028	
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	https://www.ncbi.nlm.nih.gov/pubmed/26919232	
36	2016	Adv Virus Res.	Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.	Nakagawa K, Lokugamage K	review	Translation	https://www.ncbi.nlm.nih.gov/pubmed/27712623	

시트1 시트3 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
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, Lai M	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/2243386	
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	https://www.ncbi.nlm.nih.gov/pubmed/2033672	
4	1992	Journal of Virology	Identification and characterization of a coronavirus packaging signal.	Fosmire JA1, Hwang K, Maki	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/1316465	
5	1994	Virology	Coronavirus translational regulation: leader affects mRNA efficiency.	Tahara SM, Dietlin TA, Bergn	research	Translation	https://www.ncbi.nlm.nih.gov/pubmed/8030227	
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	https://www.ncbi.nlm.nih.gov/pubmed/8615041	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	https://www.ncbi.nlm.nih.gov/pubmed/9782298	
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	https://www.ncbi.nlm.nih.gov/pubmed/10640556	
9	2000	Journal of Virology	Identification of a bovine coronavirus packaging signal.	Cologna R, Hogue BG.	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/10590153	
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	https://www.ncbi.nlm.nih.gov/pubmed/11533169	
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	https://www.ncbi.nlm.nih.gov/pubmed/12791173	
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	https://www.ncbi.nlm.nih.gov/pubmed/16254320	
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	https://www.ncbi.nlm.nih.gov/pubmed/15681451	
14	2007	Virology	Coronavirus nucleocapsid protein is an RNA chaperone.	Zúñiga S, Sola I, Moreno JL,	research	RNA chaperone	https://www.ncbi.nlm.nih.gov/pubmed/16979208	
15	2007	Journal of Virology	New structure model for the packaging signal in the genome of group IIa coronaviruses.	Chen SC, van den Born E, va	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/17428856	
16	2009	Journal of Virology	Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications f	Chang CK, Hsu YL, Chang Y	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19052082	
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	Grossoehme NE, Li L, Keane	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19782089	
18	2010	Journal of Virology	Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.	Zúñiga S, Cruz JL, Sola I, Ma	research	template switchi	https://www.ncbi.nlm.nih.gov/pubmed/19955314	in vitro templa
19	2010	Springer	Molecular Biology of the SARS-Coronavirus	Sunil K. Lal	book	SARS	https://link.springer.com/book/10.1007/978-3-642-03683-5	
20	2011	Journal of Virology	Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and suppo	Beura LK, Dinh PX, Osorio F	research	host factors	https://www.ncbi.nlm.nih.gov/pubmed/21976648	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	https://www.ncbi.nlm.nih.gov/pubmed/21378501	
22	2011	PLoS Pathogen	SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Indu	Huang C, Lokugamage KG, F	research	nsp1	https://www.ncbi.nlm.nih.gov/pubmed/22174690	
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	https://www.ncbi.nlm.nih.gov/pubmed/21389138	
24	2011	Journal of Virology	The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domain	Sola I, Galán C, Mateos-Gón	research	host factor	https://www.ncbi.nlm.nih.gov/pubmed/21411518	
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	https://www.ncbi.nlm.nih.gov/pubmed/22241479	
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	https://www.ncbi.nlm.nih.gov/pubmed/23637410	immunoprecip
27	2014	Viruses	The coronavirus nucleocapsid is a multifunctional protein	McBride R, van Zyl M, Fieldin	review	N protein	https://www.ncbi.nlm.nih.gov/pubmed/25105276	
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	https://www.ncbi.nlm.nih.gov/pubmed/24501403	
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, Zh	research	structure	https://www.ncbi.nlm.nih.gov/pubmed/24564608	
30	2014	Cell Host Microbe	Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcrip	Wu CH, Chen PJ, Yeh SH.	research	PTM	https://www.ncbi.nlm.nih.gov/pubmed/25299332	
31	2014	Cell	Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.	Kutluay SB, Zang T, Blanco-M	research	CLIP	https://www.ncbi.nlm.nih.gov/pubmed/25416948	
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	https://www.ncbi.nlm.nih.gov/pubmed/26085159	
33	2015	Annu Rev Virol.	Continuous and Discontinuous RNA Synthesis in Coronaviruses.	Sola I, Almazán F, Zúñiga S,	review	template-switchi	https://www.ncbi.nlm.nih.gov/pubmed/26958916	
34	2015	Virology	Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.	Lloyd RE.	review	host factors	https://www.ncbi.nlm.nih.gov/pubmed/25818028	
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	https://www.ncbi.nlm.nih.gov/pubmed/26919232	
36	2016	Adv Virus Res.	Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.	Nakagawa K, Lokugamage K	review	Translation	https://www.ncbi.nlm.nih.gov/pubmed/27712623	

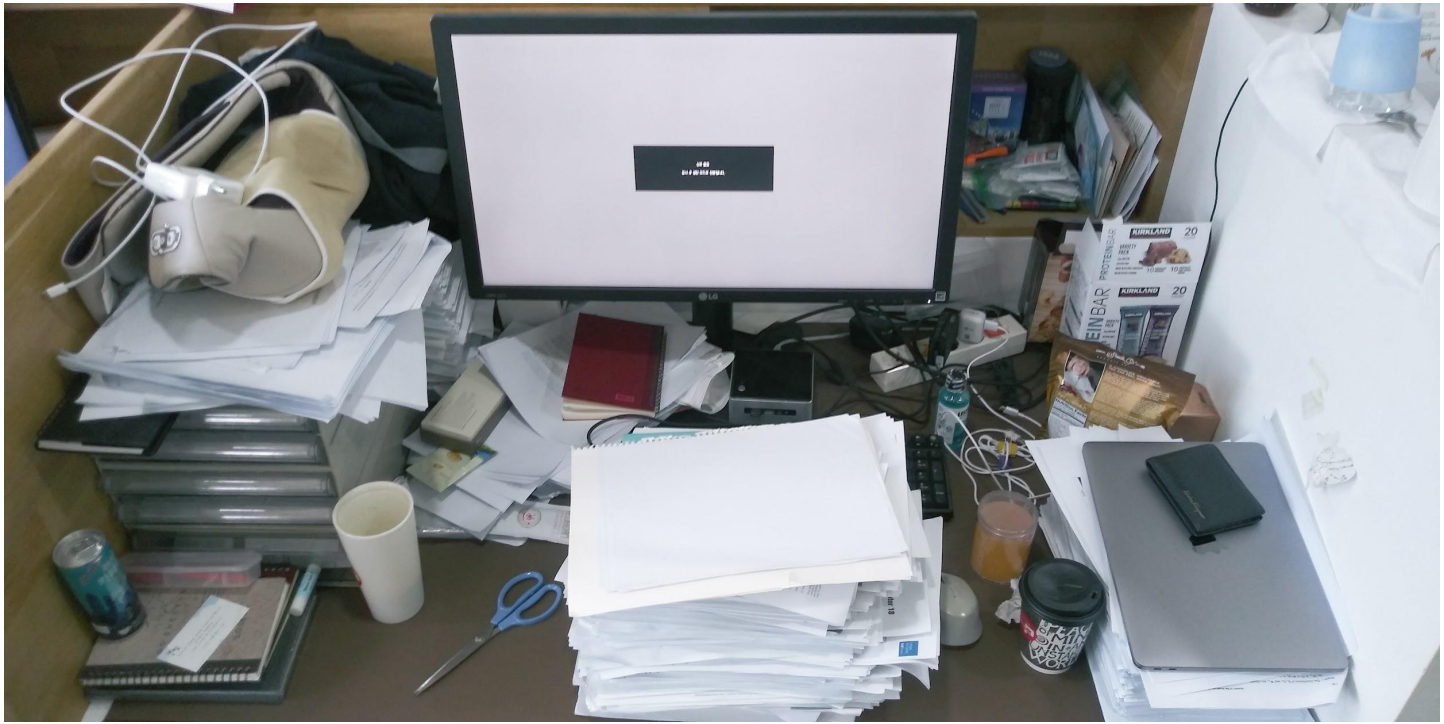
시트1 시트3 Explore

Details, details, and details from field-specific journals

	A		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, Lai M	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/2243386	
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	https://www.ncbi.nlm.nih.gov/pubmed/2033672	
4	1992	Journal of Virology	Identification and characterization of a coronavirus packaging signal.	Fosmire JA1, Hwang K, Maki	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/1316465	
5	1994	Virology	Coronavirus translational regulation: leader affects mRNA efficiency.	Tahara SM, Dietlin TA, Bergn	research	Translation	https://www.ncbi.nlm.nih.gov/pubmed/8030227	
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	https://www.ncbi.nlm.nih.gov/pubmed/8615041	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	https://www.ncbi.nlm.nih.gov/pubmed/9782298	
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	https://www.ncbi.nlm.nih.gov/pubmed/10640556	
9	2000	Journal of Virology	Identification of a bovine coronavirus packaging signal.	Cologna R, Hogue BG.	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/10590153	
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	https://www.ncbi.nlm.nih.gov/pubmed/11533169	
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	https://www.ncbi.nlm.nih.gov/pubmed/12791173	
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	https://www.ncbi.nlm.nih.gov/pubmed/16254320	
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	https://www.ncbi.nlm.nih.gov/pubmed/15681451	
14	2007	Virology	Coronavirus nucleocapsid protein is an RNA chaperone.	Zúñiga S, Sola I, Moreno JL,	research	RNA chaperone	https://www.ncbi.nlm.nih.gov/pubmed/16979208	
15	2007	Journal of Virology	New structure model for the packaging signal in the genome of group IIa coronaviruses.	Chen SC, van den Born E, va	research	packaging	https://www.ncbi.nlm.nih.gov/pubmed/17428856	
16	2009	Journal of Virology	Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications f	Chang CK, Hsu YL, Chang Y	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19052082	
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	Grossoehme NE, Li L, Keane	research	RNA-binding	https://www.ncbi.nlm.nih.gov/pubmed/19782089	
18	2010	Journal of Virology	Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription.	Zúñiga S, Cruz JL, Sola I, Ma	research	template switchi	https://www.ncbi.nlm.nih.gov/pubmed/19955314	in vitro templa
19	2010	Springer	Molecular Biology of the SARS-Coronavirus	Sunil K. Lal	book	SARS	https://link.springer.com/book/10.1007/978-3-642-03683-5	
20	2011	Journal of Virology	Cellular poly(c) binding proteins 1 and 2 interact with porcine reproductive and respiratory syndrome virus nonstructural protein 1β and suppo	Beura LK, Dinh PX, Osorio F	research	host factors	https://www.ncbi.nlm.nih.gov/pubmed/21976648	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	https://www.ncbi.nlm.nih.gov/pubmed/21378501	
22	2011	PLoS Pathogen	SARS Coronavirus nsp1 Protein Induces Template-Dependent Endonucleolytic Cleavage of mRNAs: Viral mRNAs Are Resistant to nsp1-Indu	Huang C, Lokugamage KG, F	research	nsp1	https://www.ncbi.nlm.nih.gov/pubmed/22174690	
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	https://www.ncbi.nlm.nih.gov/pubmed/21389138	
24	2011	Journal of Virology	The polypyrimidine tract-binding protein affects coronavirus RNA accumulation levels and relocalizes viral RNAs to novel cytoplasmic domain	Sola I, Galán C, Mateos-Gón	research	host factor	https://www.ncbi.nlm.nih.gov/pubmed/21411518	
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	https://www.ncbi.nlm.nih.gov/pubmed/22241479	
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	https://www.ncbi.nlm.nih.gov/pubmed/23637410	immunoprecip
27	2014	Viruses	The coronavirus nucleocapsid is a multifunctional protein	McBride R, van Zyl M, Fieldin	review	N protein	https://www.ncbi.nlm.nih.gov/pubmed/25105276	
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	https://www.ncbi.nlm.nih.gov/pubmed/24501403	
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, Z	research	structure	https://www.ncbi.nlm.nih.gov/pubmed/24564608	
30	2014	Cell Host Microbe	Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcrip	Wu CH, Chen PJ, Yeh SH.	research	PTM	https://www.ncbi.nlm.nih.gov/pubmed/25299332	
31	2014	Cell	Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis.	Kutluay SB, Zang T, Blanco-M	research	CLIP	https://www.ncbi.nlm.nih.gov/pubmed/25416948	
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	https://www.ncbi.nlm.nih.gov/pubmed/26085159	
33	2015	Annu Rev Virol.	Continuous and Discontinuous RNA Synthesis in Coronaviruses.	Sola I, Almazán F, Zúñiga S,	review	template-switchi	https://www.ncbi.nlm.nih.gov/pubmed/26958916	
34	2015	Virology	Nuclear proteins hijacked by mammalian cytoplasmic plus strand RNA viruses.	Lloyd RE.	review	host factors	https://www.ncbi.nlm.nih.gov/pubmed/25818028	
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	https://www.ncbi.nlm.nih.gov/pubmed/26919232	
36	2016	Adv Virus Res.	Viral and Cellular mRNA Translation in Coronavirus-Infected Cells.	Nakagawa K, Lokugamage K	review	Translation	https://www.ncbi.nlm.nih.gov/pubmed/27712623	

시트1 시트3 Explore

(Please do not copy or share)



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
Jan.13 9:00am	Jan.13 1:00am	Jan.12 7:00pm	Jan.12 4:00pm
Feb.17 9:00am	Feb.17 1:00am	Feb.16 7:00pm	Feb.16 4:00pm
Mar.10 5:00pm	Mar.10 9:00am	Mar.10 3:00am	Mar.10 0:00am
Jun.2 5:00pm	Jun.2 10:00am	Jun.2 4:00am	Jun.2 1:00am
Jul.14 9:00am	Jul.14 2:00am	Jul.13 8:00pm	Jul.13 5:00pm
Aug.11 9:00am	Aug.11 2:00am	Aug.10 8:00pm	Aug.10 5:00pm

WEBSITE
narrykim.org/en

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

SPEAKERS
Joan Steitz, 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
Noam Stern-Ginossar, Weizmann Institute of Science
Yong Sun Lee, National Cancer Center
Heeyoung Seok, Korea University
David Bartel, MIT
Chengqi Yi, Peking University
Olivia Rissland, University of Colorado
Jin-Hong Kim, IBS & Seoul National University

FREE
registration

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 <u>not</u> 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 <u>not</u> a programming course. Nevertheless, you are expected to read/write/present mathematical and computational details in your assignments.
3. This is <u>not</u> 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

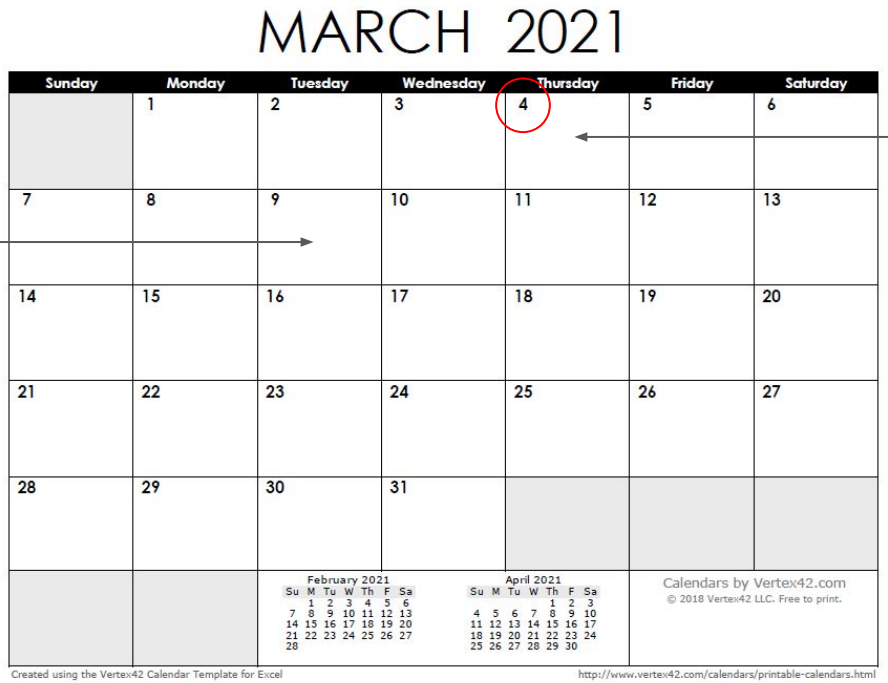
Written assignment will be anonymously graded (largely) based on three peer reviews under the Honor system.

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

https://en.wikipedia.org/wiki/Honor_system

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

Submit peer review of written response before the next class starts



Submit written response before class starts

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