KSBi-BIML 2023

Bioinformatics & Machine Learning(BIML)
Workshop for Life Scientists, Data Scientists,
and Bioinformatians

생물정보학&머신러닝 워크샵(온라인)

Bioinformatics and AI for microRNA

백대현 _ 서울대학교

SBi 한국생명정보학회
Korean Society for Bioinformatics

본 강의 자료는 한국생명정보학회가 주관하는 BIML 2023 워크샵 온라인 수업을 목적으로 제작된 것으로 해당 목적 이외의 다른 용도로 사용할 수 없음을 분명하게 알립니다.

이를 다른 사람과 공유하거나 복제, 배포, 전송할 수 없으며 만약 이러한 사항을 위반할 경우 발생하는 **모든 법적 책임은 전적으로 불법 행위자 본인에게 있음을 경고**합니다.

KSBi-BIML 2023

Bioinformatics & Machine Learning (BIML)
Workshop for Life Scientists, Data Scientists, and Bioinformatians

안녕하십니까?

한국생명정보학회가 개최하는 동계 교육 워크샵인 BIML-2023에 여러분을 초대합니다. 생명정보학 분야의 연구자들에게 최신 동향의 데이터 분석기술을 이론과 실습을 겸비해 전달하고자 도입한 전문 교육 프로그램인 BIML 워크샵은 2015년에 시작하여 올해로 9차를 맞이하게 되었습니다. 지난 2년간은 심각한 코로나 대유행으로 인해 아쉽게도 모든 강의가 온라인으로 진행되어 현장 강의에서만 가능한 강의자와 수강생 사이에 다양한 소통의 기회가 없음에 대한 아쉬움이 있었습니다. 다행히도 최근 사회적 거리두기 완화로 현장 강의가 가능해져 올해는 현장 강의를 재개함으로써 온라인과 현장 강의의 장점을 모두 갖춘 프로그램을 구성할 수 있게 되었습니다.

BIML 워크샵은 전통적으로 크게 인공지능과 생명정보분석 두 개의 분야로 구성되었습니다. 올해 AI 분야에서는 최근 생명정보 분석에서도 응용이 확대되고 있는 다양한 심층학습(Deep learning) 기법들에 대한 현장 강의가 진행될 예정이며, 관련하여 심층학습을 이용한 단백질구조예측, 유전체 분석, 신약개발에 대한 이론과 실습 강의가 함께 제공될 예정입니다. 또한 싱글셀오믹스 분석과 메타유전체분석 현장 강의는 많은 연구자의 연구 수월성 확보에 큰 도움을 줄 것으로 기대하고 있습니다. 이외에 다양한 생명정보학 분야에 대하여 30개 이상의 온라인 강좌가 개설되어 제공되며 온라인 강의의 한계를 극복하기 위해서 실시간 Q&A 세션 또한 마련했습니다. 특히 BIML은 각 분야 국내 최고 전문가들의 강의로 구성되어 해당 분야의 기초부터 최신 연구 동향까지 포함하는 수준 높은 내용의 강의가 될 것입니다.

이번 BIML-2023을 준비하기까지 너무나 많은 수고를 해주신 BIML-2023 운영위원회의 남진우, 우현구, 백대현, 정성원, 정인경, 장혜식, 박종은 교수님과 KOBIC 이병욱 박사님께 커다란 감사를 드립니다. 마지막으로 부족한 시간에도 불구하고 강의 부탁을 흔쾌히 허락하시고 훌륭한 현장 강의와 온라인 강의를 준비하시는데 노고를 아끼지 않으신 모든 연사분께 깊은 감사를 드립니다.

2023년 2월

한국생명정보학회장 이 인 석

강의개요

Bioinformatics and AI for microRNA

본 강의에서는 인간 microRNA의 생성 및 타겟팅에 대한 최신 생물정보학 연구 내용을 소개한다. 또한, 인공 지능(AI)을 활용한 microRNA 타겟팅 연구를 소개하고, AI가 어떻게 microRNA 타겟 발굴의 예측 정확도를 높일 수 있는지 고찰한다. 끝으로, 코로나 19를 일으킨 원인 바이러스인 SARS-CoV-2의 microRNA가 어떻게 host immune을 회피하는 지에 대한 최근 연구 결과에 대해 논의한다.

본 강의는 다음의 내용을 포함한다:

- 인간 microRNA 생성 기작
- 인간 microRNA 타겟팅 기작에 대한 최근 연구
- 인간 microRNA 타겟팅 연구를 위한 AI 기법
- SARS-CoV-2 microRNA

* 참고강의교재:

  강의자료에 첨부된 논문 2편(The regulatory impact of RNA-binding proteins on microRNA targeting, A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome)

* 교육생준비물:

  노트북 (메모리 8GB 이상, 디스크 여유공간 30GB 이상)

* 강의 난이도: 초급

* 강의: 백대현 교수 (서울대학교 생명과학부)

Curriculum Vitae

Speaker Name: Daehyun Baek, Ph.D.

▶ **Personal Info**

Name	Daehyun Baek
Title	Associate Professor
Affiliation	Seoul National University

▶ **Contact Information**

Address	Rm 423, Bldg 504, Seoul National University Seoul, South Korea, 08826
Email	baek@snu.ac.kr
Phone Number	010-7737-0810

Research Interest

Artificial Intelligence (Deep Learning) for Biology and Medicine, Computational Biology and Bioinformatics, Noncoding Genome, Cancer Genomics

Educational Experience

1999	B.S. in Electrical Engineering at KAIST (Minor in Biological Sciences)
2007	Ph.D. in Bioengineering at University of Washington (Advisor: Phil Green)

Professional Experience

2007-2010	Postdoctoral Fellow at Whitehead Institute / MIT / HHMI (Advisor: David Bartel)
2010-Present	Assistant & Associate Professor of School of Biological Sciences at SNU

Selected Publications (5 maximum)

1. D. Kim*, S. Kim*, J. Park*, H. R. Chang*, J. Chang*, J. Ahn*, ..., M.-S. Park#, Y. K. Kim#, and **D. Baek#**, A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome, ***Nature Communications***, 2021 (**IF=14.92**)
2. S. Kim*, S. Kim*, H. R. Chang*, D. Kim*, ..., C. Shin#, and **D. Baek#**, The regulatory impact of RNA-binding proteins on microRNA targeting, ***Nature Communications***, 2021 (**IF=14.92**)
3. D. Kim*, Y. M. Sung*, J. Park*, ..., and **D. Baek**, General Rules for Functional MicroRNA Targeting, ***Nature Genetics***, 2016 (**cited 106 times, IF=38.33**)
4. D. Garcia*, **D. Baek***#, ..., and D. Bartel#, Weak Seed-Pairing Stability and High Target-Site Abundance Decrease the Proficiency of lsy-6 and Other miRNAs, ***Nature Structural and Molecular Biology***, 2011. (**cited 920 times, IF=15.37**)
5. **D. Baek***, J. Villen*, C. Shin*, ..., and D. Bartel, The Impact of MicroRNAs on Protein Output, ***Nature***, 2008. (**cited 4,005 times, IF=49.96**)

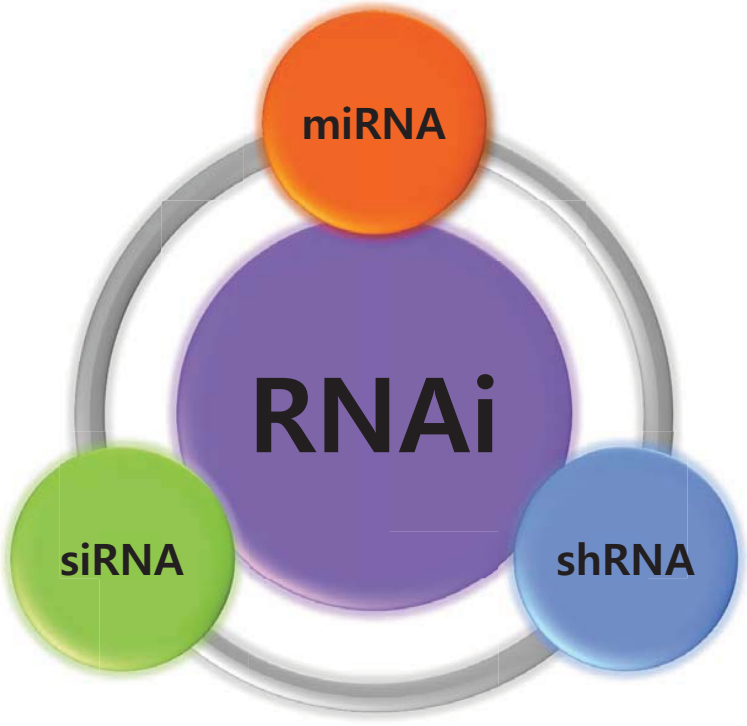
(*co-first authors, #co-corresponding authors)

Bioinformatics and AI for MicroRNA

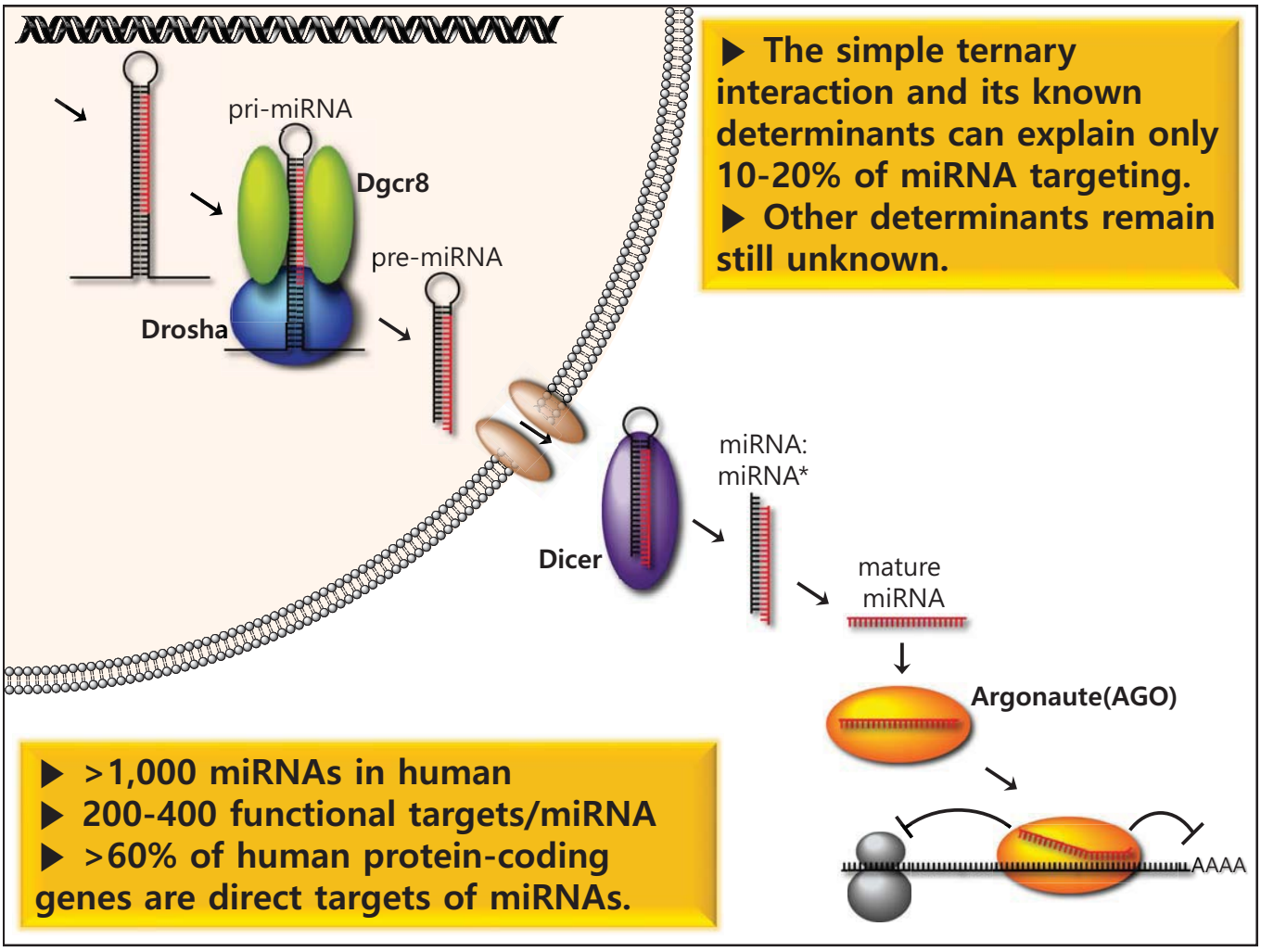
Daehyun Baek

School of Biological Sciences
Seoul National University

RNA Interference



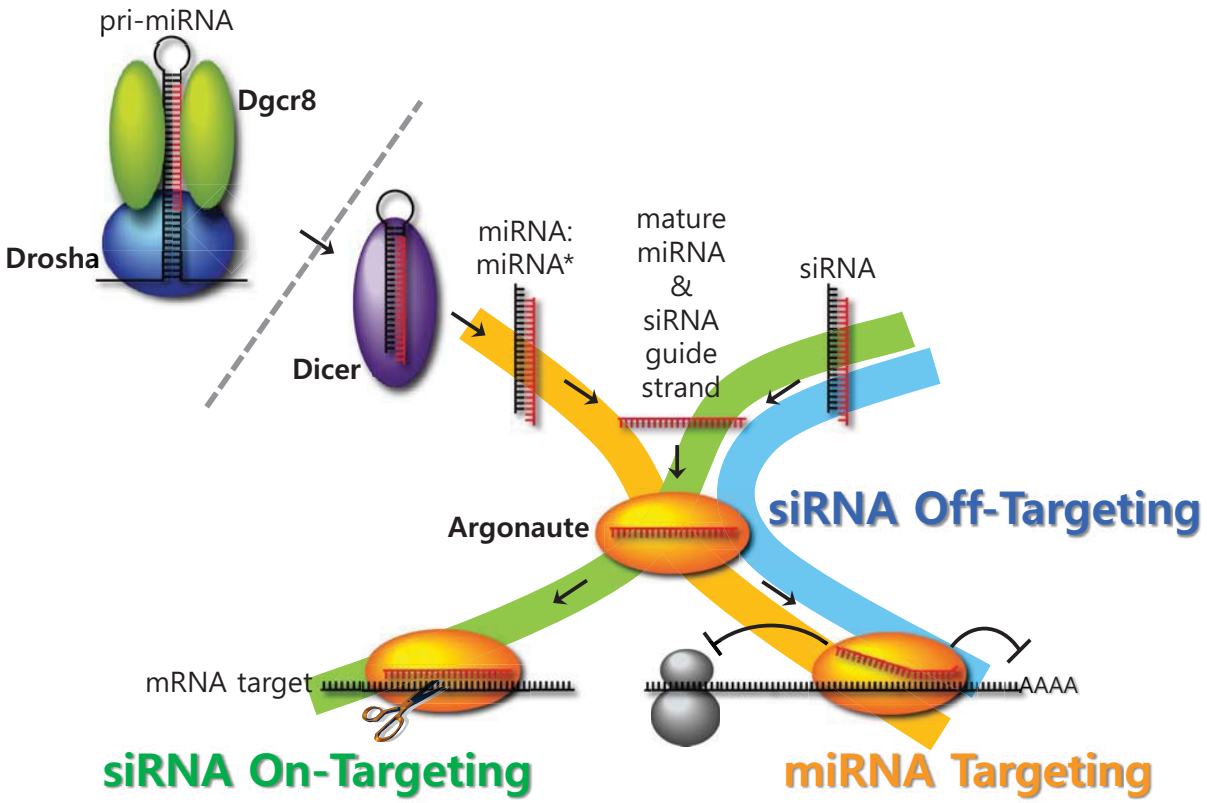
생성(biogenesis) 및 유전자 제어(targeting) 기작 규명 필요



▶ The simple ternary interaction and its known determinants can explain only 10-20% of miRNA targeting.
▶ Other determinants remain still unknown.

▶ >1,000 miRNAs in human
▶ 200-400 functional targets/miRNA
▶ >60% of human protein-coding genes are direct targets of miRNAs.

RNAi Biogenesis and Targeting

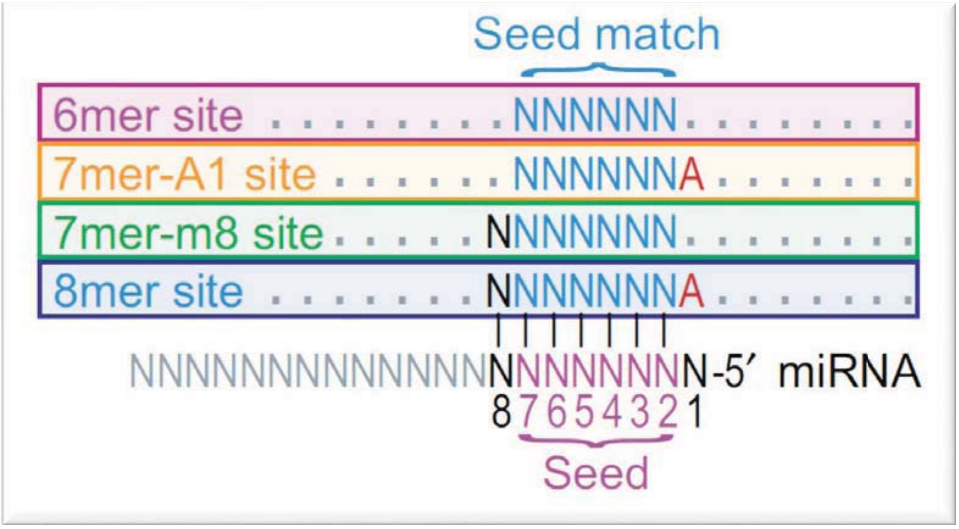


General Rules for Functional MicroRNA Targeting

Daehyun Baek

School of Biological Sciences
Seoul National University

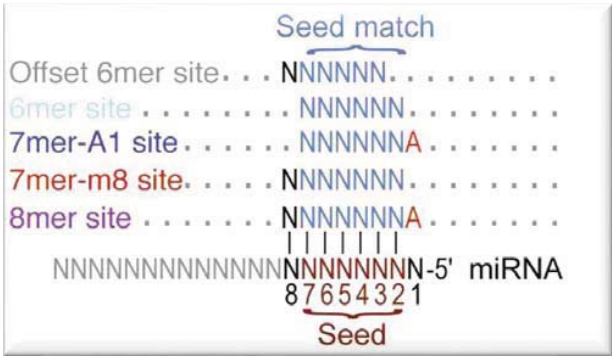
Canonical Site Types(CSTs) of miRNA Targeting



(Baek *et al.*, Nature, 2008)

Noncanonical Site Types(NSTs) of miRNA Targeting

Offset 6mer



Centered Site

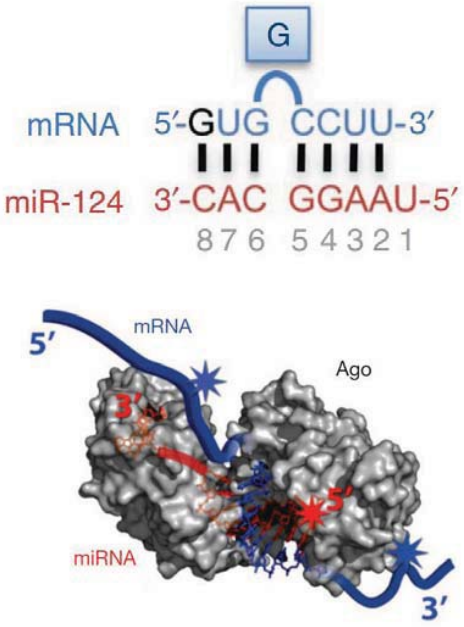
```
5′-nnnnnnnACAUGAAACUACnnn-3′  mRNA
          ||||||||||||
3′-GGGUUGUUGUACUUUGAUGGAU-5′  miRNA
```

(Friedman *et al.*, Genome Research, 2009)

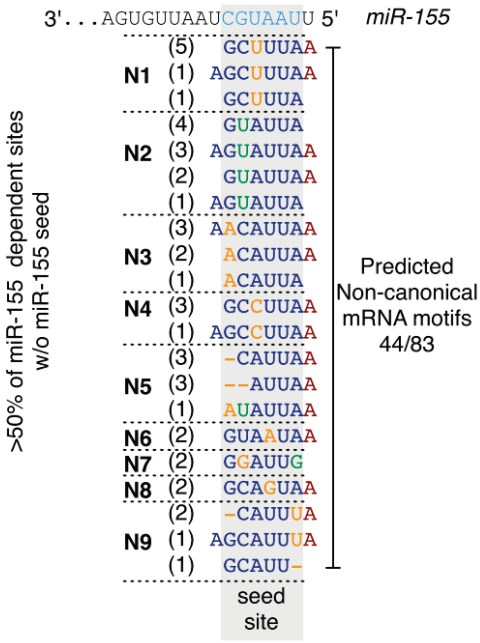
(Shin *et al.*, Molecular Cell, 2010)

Noncanonical Site Types(NSTs) of miRNA Targeting

Pivot Pairing



Single Mismatch Sites



(Chi *et al.*, Nature, 2009)

(Chi *et al.*, NSMB, 2011)

(Loeb *et al.*, Molecular Cell, 2012)

Limitations and Solutions

Incomplete Searches

- Canonical target sites
- Offset 6-mer site
- Centered site

Indirect Evidence

- AGO CLIPSeq-based analysis
- Based on limited number of miRNAs
- Pivot pairing site
- MIRZA sites

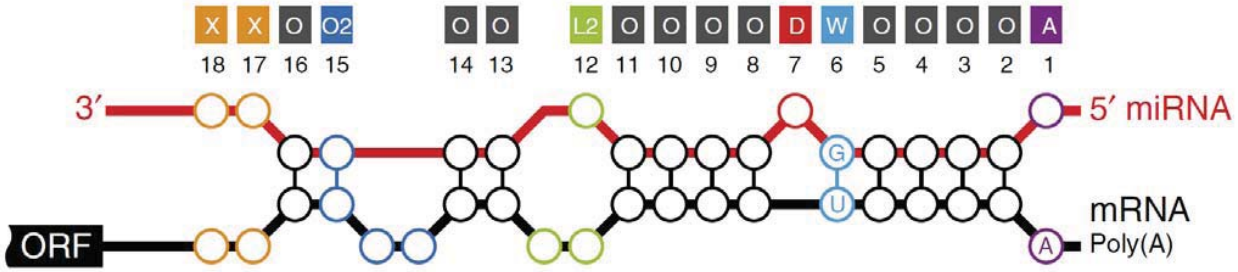
Limitations

- No one has performed the comprehensive and systematic search for functional miRNA targeting rules.

Solution: Extensive Bioinformatics Analysis

- Massive-scale search for functional, meaning those targets that elicit detectable mRNA repression, miRNA targets.
- **Goal: The most comprehensive discovery of miRNA targeting principles**

Challenge: Complexity of miRNA-mRNA Interactions



O Match X Mismatch D Bulge on miRNA A Adenine at 1st mRNA position
B Other at 1st mRNA position W Wobble pairing O1–O9 Bulge on mRNA
L1–L9 Internal loop ø All but match . Does not matter

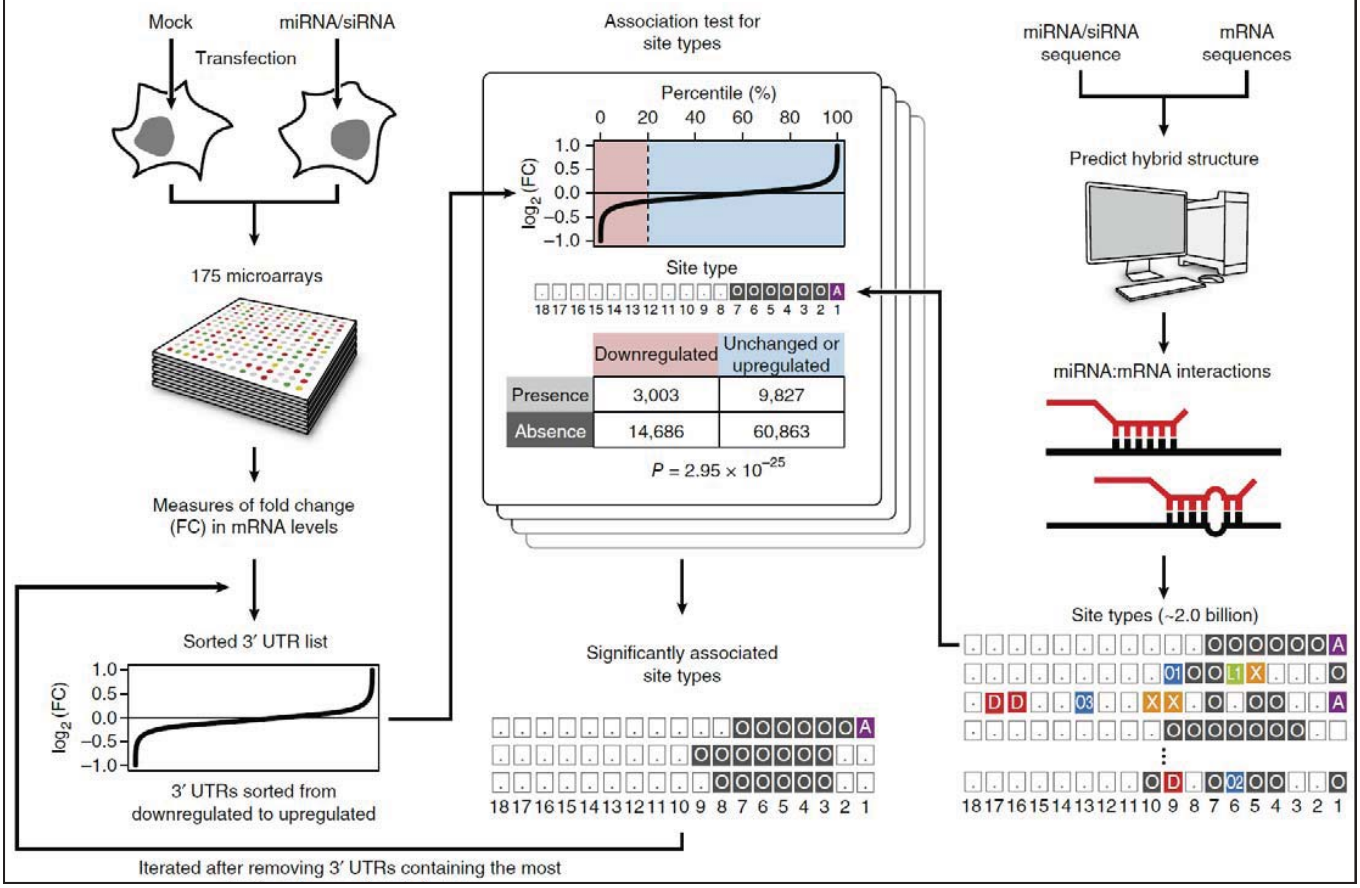
The number of site types(STs) that can occur between human miRNAs and mRNAs with >8 targets: ~2 Billion

Comparison of Search Spaces

Previously evaluated <1 million STs

>99.9% of the STs have remained unexamined.

Method Overview



Solution

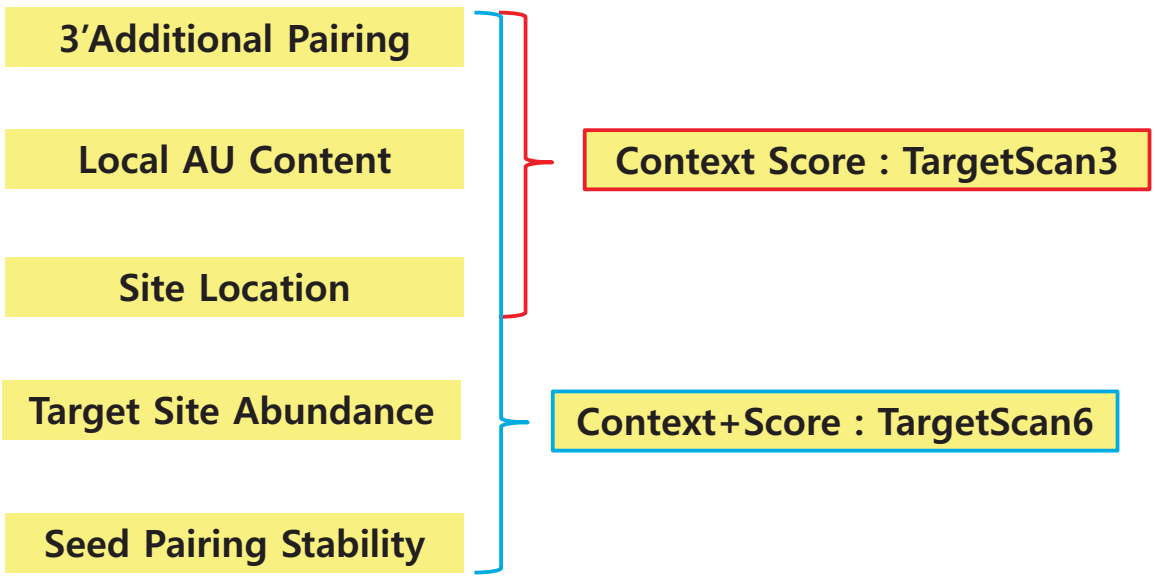
High-Performance Hardware

- High-performance server system
- >1,500 CPU cores
- >1.2 PB of storage

Optimized Software

- Implemented in C/C++
- Massive use of bit-operation
- Hash table based optimization
- Fast compression algorithm for data transfer

Known Local Context of CSTs

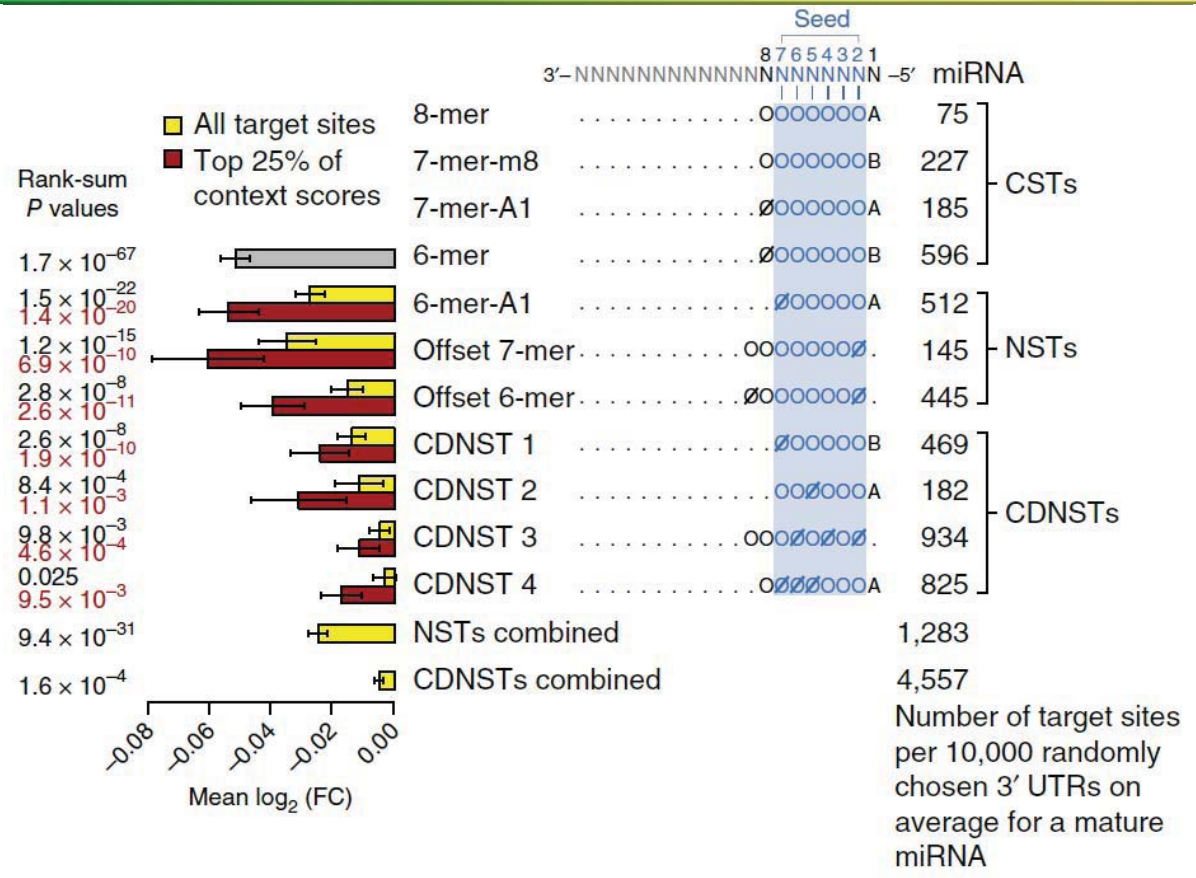


5 known determinants for the overall proficiency of CSTs

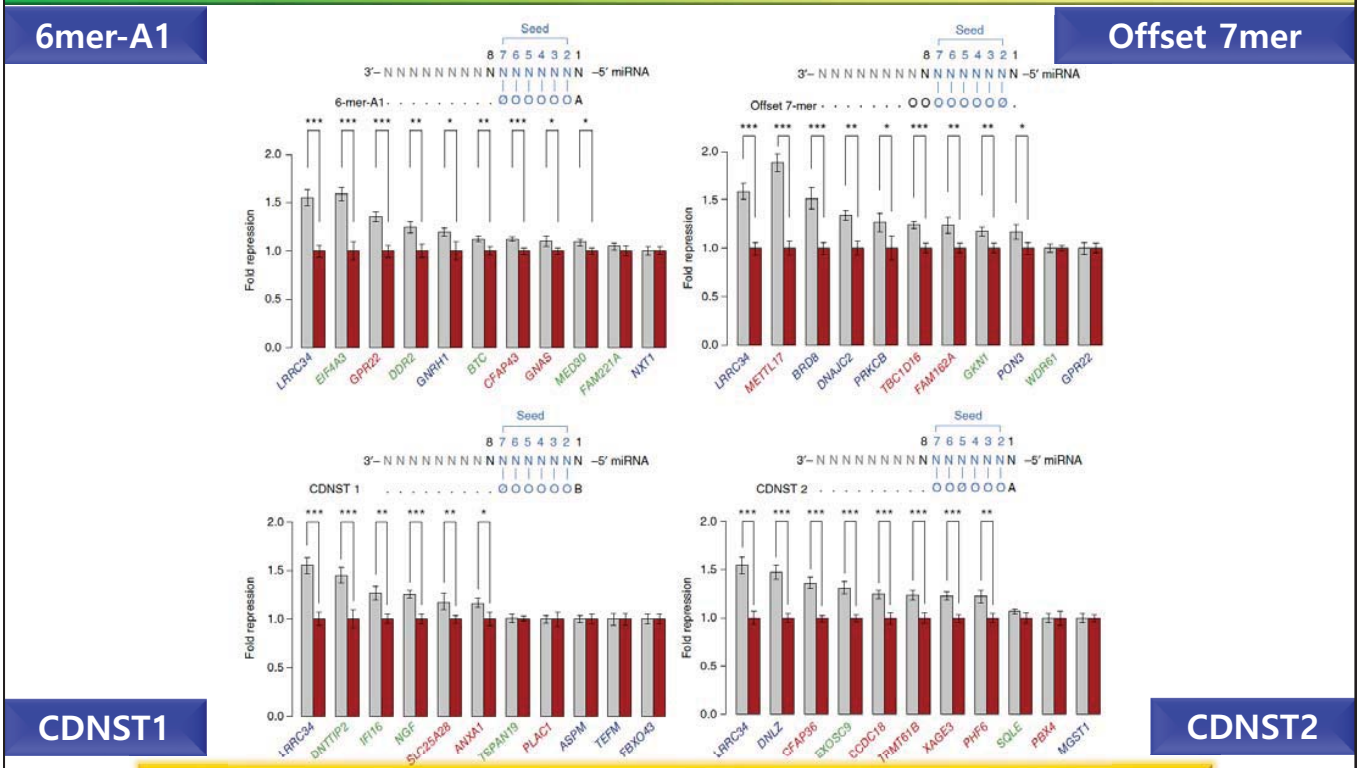
(Grimson *et al.*, Molecular Cell, 2007)

(Garcia and Baek *et al.*, NSMB, 2011)

Discovery of Functional CDNSTs

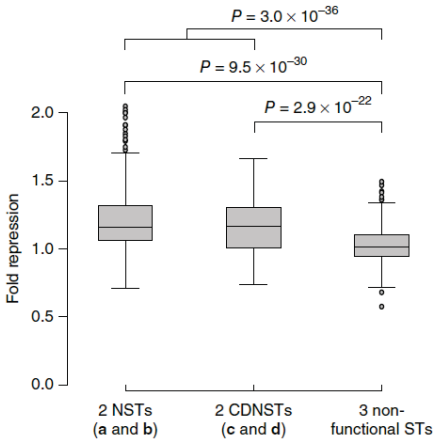


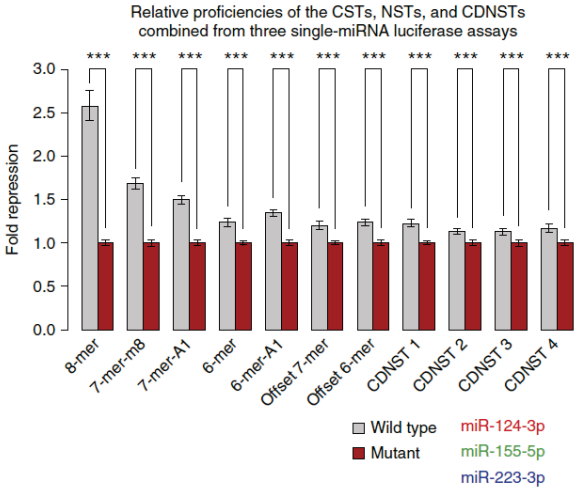
Validation of NSTs and CDNSTs: Luciferase Assays



70% of NST and CDNST targets were validated by luciferase assays.

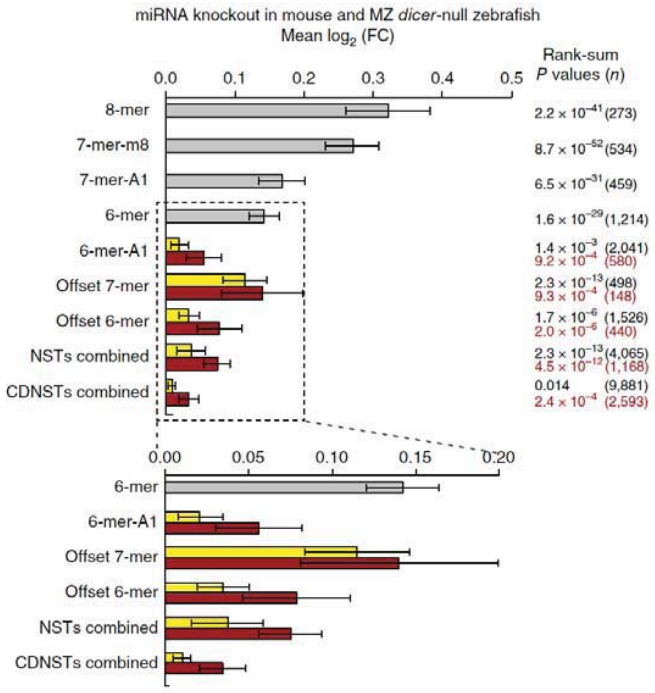
Validation of NSTs and CDNSTs: Luciferase Assays

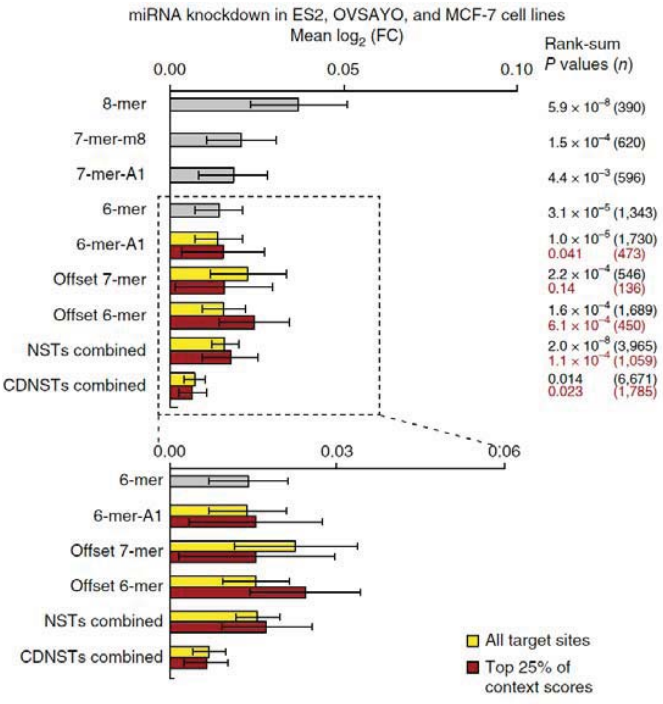




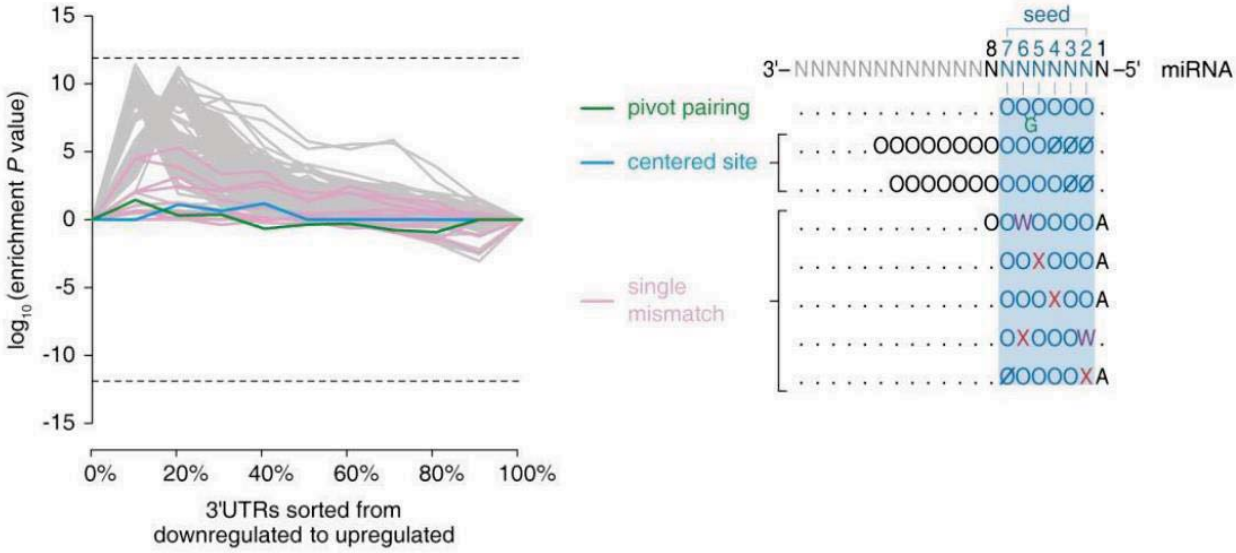
70% of NST and CDNST targets were validated by luciferase assays.

Validation of NSTs and CDNSTs: miRNA KO/KD Data

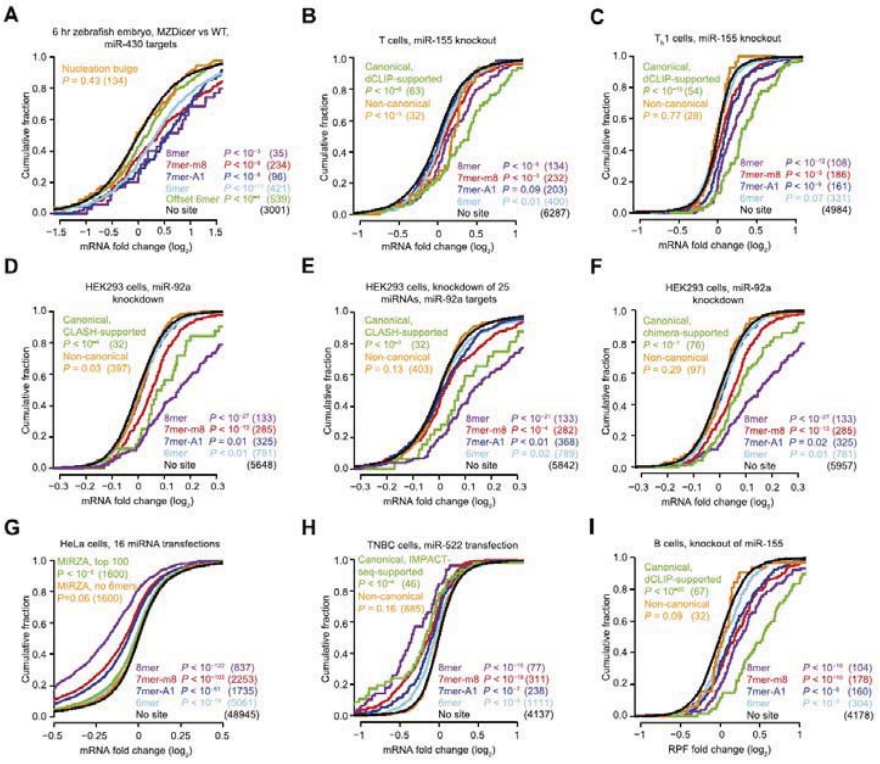




Evaluation of Previously Reported NSTs



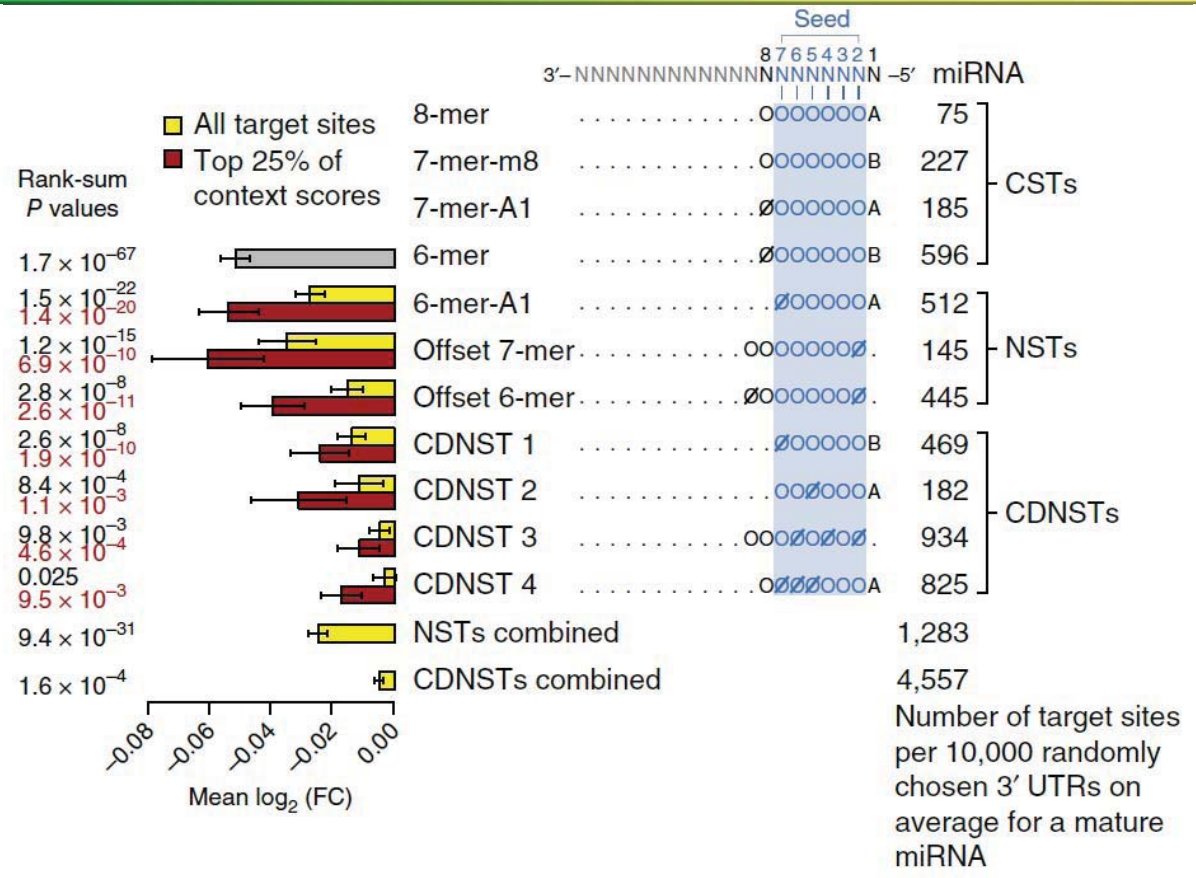
Evaluation of Previously Reported NSTs



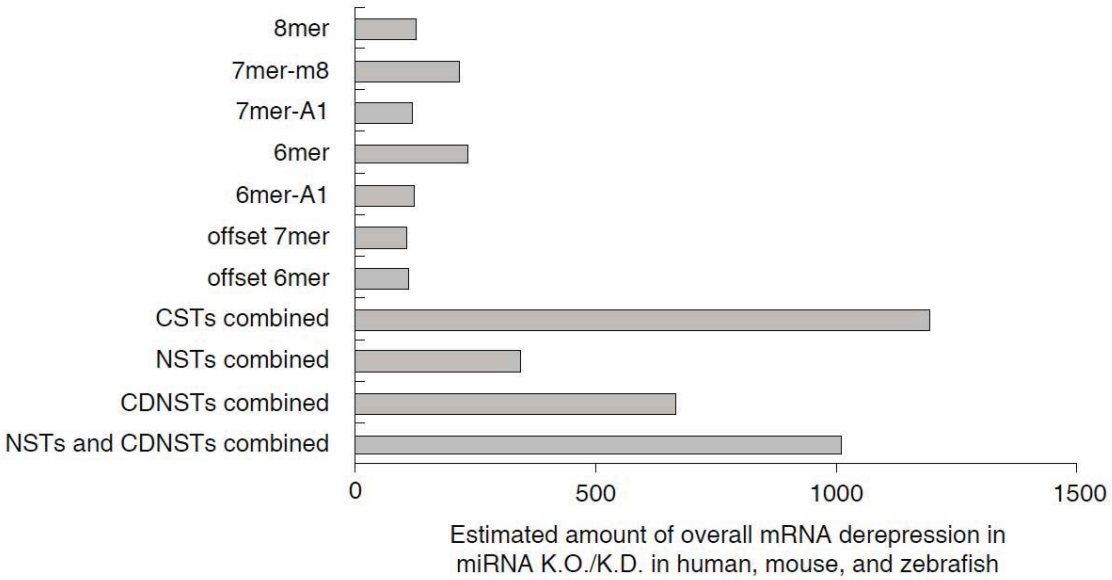
Consistent with a recent work by the Bartel lab.

(Agarwal *et al.*, eLife, 2015)

Comprehensive View on Functional miRNA Targeting



The Impact of Functional miRNA Targeting



The Impact of Functional miRNA Targeting

Site type name	Site type	Median PhyloP score[a] Site	Median PhyloP score[a] Control	Rank-sum *P* value
8-mer	.....OOOOOOOA	0.521	0.218	$<1.0 \times 10^{-320}$[b]
7-mer-m8	.....OOOOOOOB	0.241	0.172	$<1.0 \times 10^{-320}$[b]
7-mer-A1	.....ØOOOOOOA	0.291	0.221	$<1.0 \times 10^{-320}$[b]
6-mer	.....ØOOOOOOB	0.193	0.175	6.3×10^{-48}
6-mer-A1	.....ØOOOOOA	0.281	0.265	1.9×10^{-39}
Offset 7-mer	.....OOOOOOOØ.	0.235	0.184	2.8×10^{-257}
Offset 6-mer	.....ØOOOOOOØ.	0.210	0.177	7.2×10^{-120}
CDNST 1[c]	.....ØOOOOOB	0.263	0.252	1.6×10^{-12}
CDNST 2[c]	.....OOØOOOA	0.387	0.361	3.5×10^{-29}
CDNST 3[c]	.....OOOØOØOØ.	0.189	0.172	9.0×10^{-19}
CDNST 4[c]	.....OØØØOOOA	0.275	0.291	1.0

Conclusions

◈ **We have constructed a massive-scale bioinformatics pipeline that aims to systematically and comprehensively evaluate miRNA-target interactions.**

◈ **We discovered 7 NSTs and CDNSTs, many of which have not been reported previously.**

◈ **Luciferase assays and independent data analyses suggest that most of the newly discovered NSTs and CDNSTs may be functional.**

◈ **The miRNA-target interactions and their gene regulatory networks may be substantially more complex than currently perceived.**

nature genetics

General rules for functional microRNA targeting

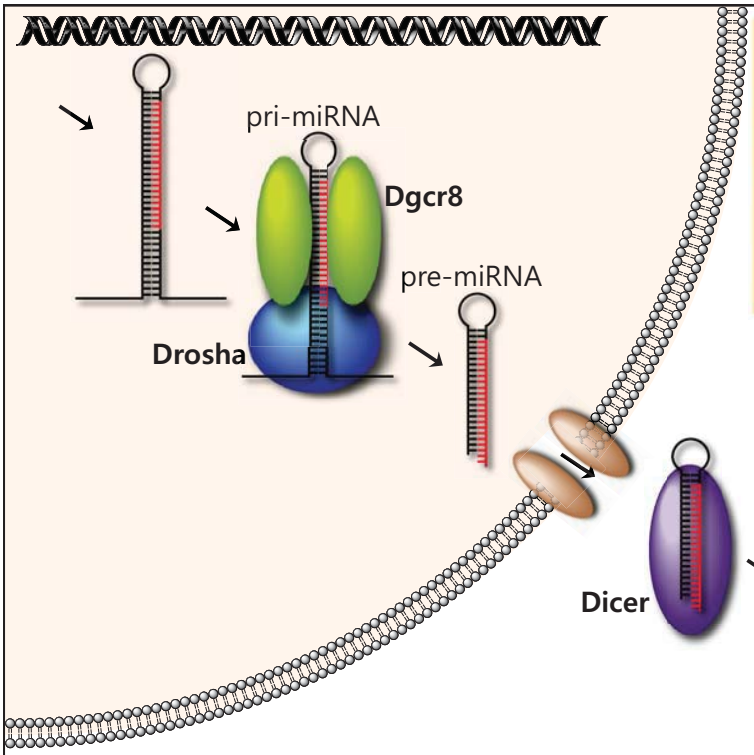
Doyeon Kim[1,2,4], You Me Sung[2,4], Jinman Park[1,2,4], Sukjun Kim[1,2], Jongkyu Kim[1,2], Junhee Park[2], Haeok Ha[2], Jung Yoon Bae[2], SoHui Kim[1,2] & Daehyun Baek[1–3]

The functional rules for microRNA (miRNA) targeting remain controversial despite their biological importance because only a small fraction of distinct interactions, called site types, have been examined among an astronomical number of site types that can occur between miRNAs and their target mRNAs. To systematically discover functional site types and to evaluate the contradicting rules reported previously, we used large-scale transcriptome data and statistically examined whether each of approximately 2 billion site types is enriched in differentially downregulated mRNAs responding to overexpressed miRNAs. Accordingly, we identified seven non-canonical functional site types, most of which are novel, in addition to four canonical site types, while also removing numerous false positives reported by previous studies. Extensive experimental validation and significantly elevated 3′ UTR sequence conservation indicate that these non-canonical site types may have biologically relevant roles. Our expanded catalog of functional site types suggests that the gene regulatory network controlled by miRNAs may be far more complex than currently understood.

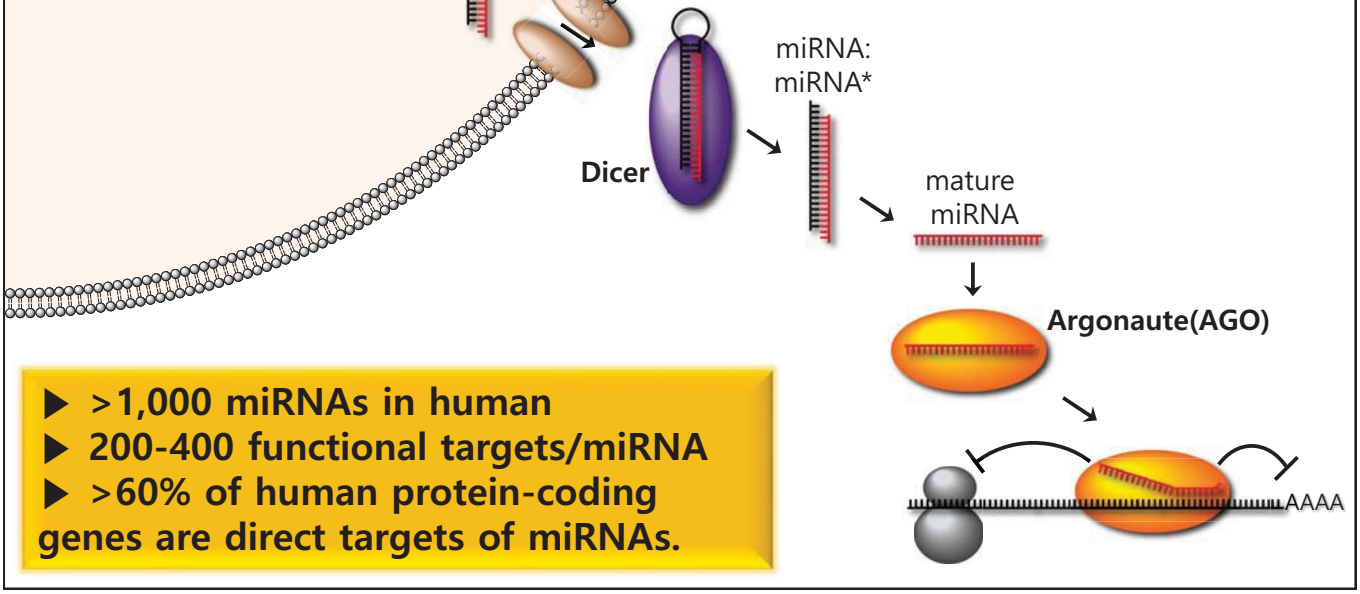
Widespread Impact of RNA-Binding Proteins on MicroRNA Targeting

Daehyun Baek

School of Biological Sciences
Seoul National University

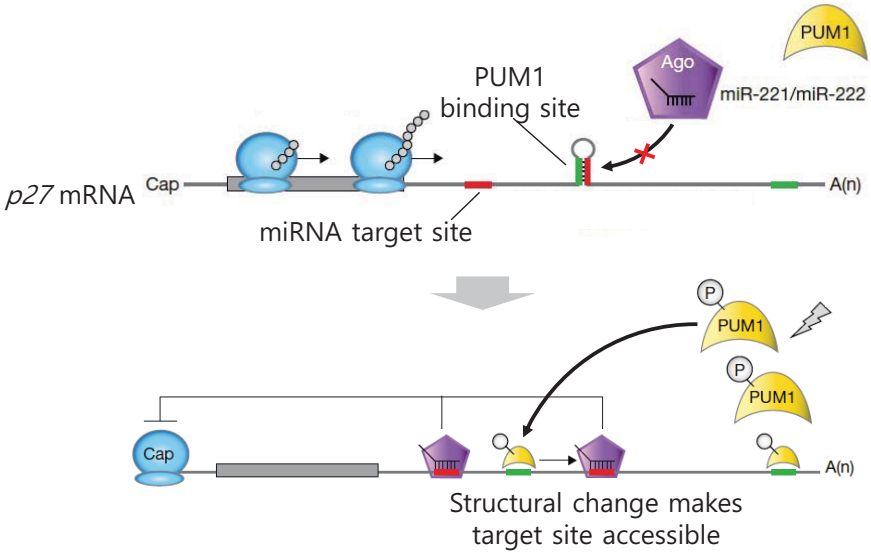


▶ The simple ternary interaction and its known determinants can explain only 10-20% of miRNA targeting.
▶ Other determinants remain still unknown.



▶ >1,000 miRNAs in human
▶ 200-400 functional targets/miRNA
▶ >60% of human protein-coding genes are direct targets of miRNAs.

miRNA Targeting(MT)에 영향을 주는 RNA-결합 단백질(RBP)

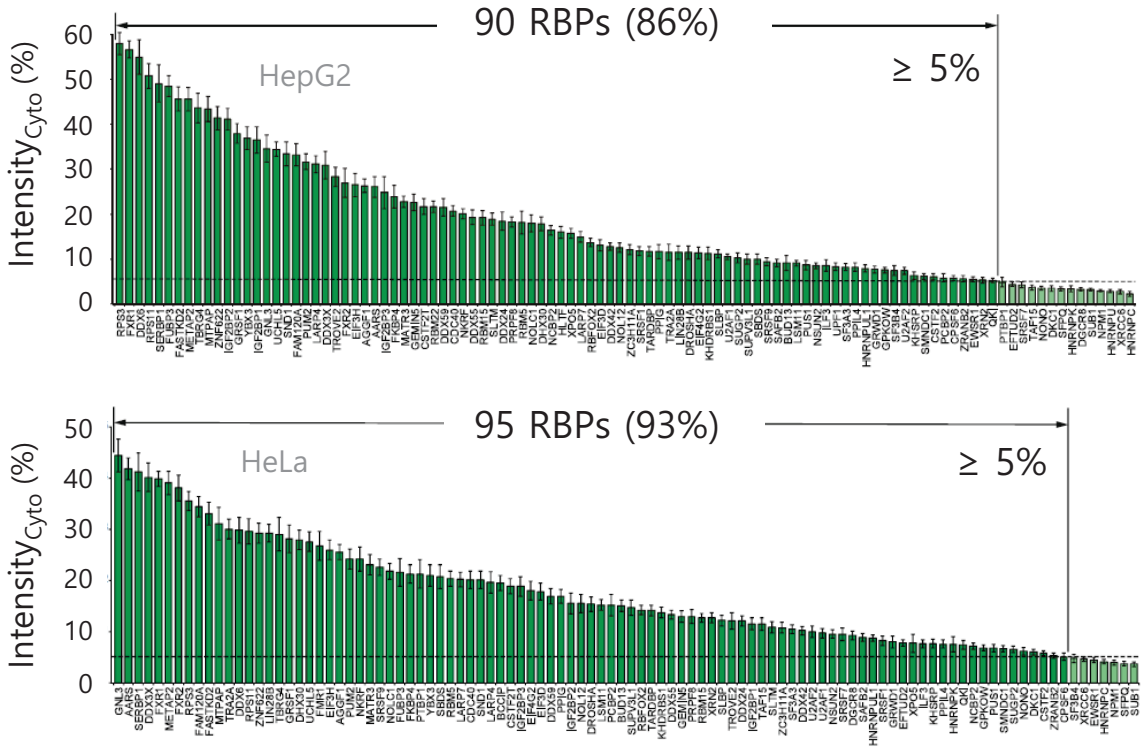


- MT 인헨서: Pumilio, PCBP2, FUS, and PTBP1
- MT 서프레서: Dnd1, RBM38, HuR, IGF2BP1, and PTBP1

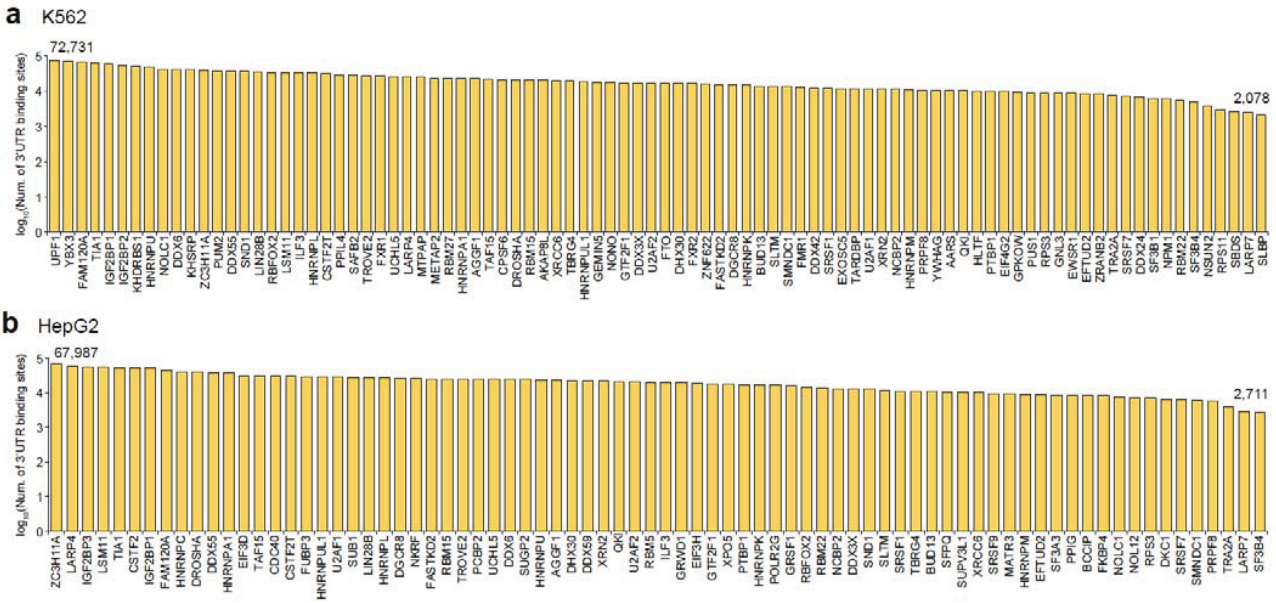
>800 RBPs x 22,000 3'UTR RBP-결합 사이트 =
~17 million 개에 이르는 결합 중 극히 일부만 연구됨

(Triboulet *et al.*, 2010, Nature Cell Biology)

Cytoplasmic Fraction of RBPs

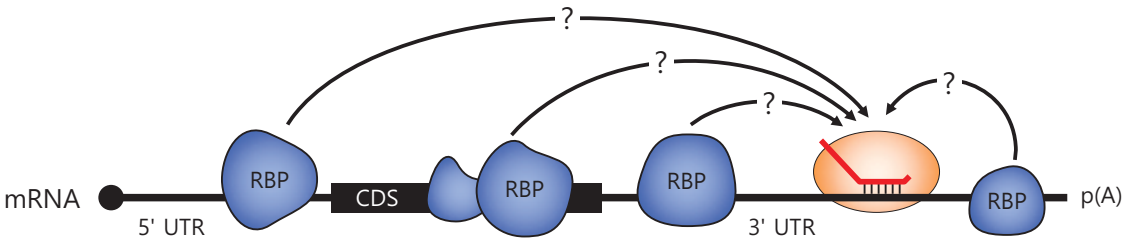


Num of 3'UTR Binding Sites of Individual RBPs



>2,000 Binding Sites of Each RBP Located in the 3' UTR

핵심 가설: RBP가 MT 조절에 중요한 역할 수행

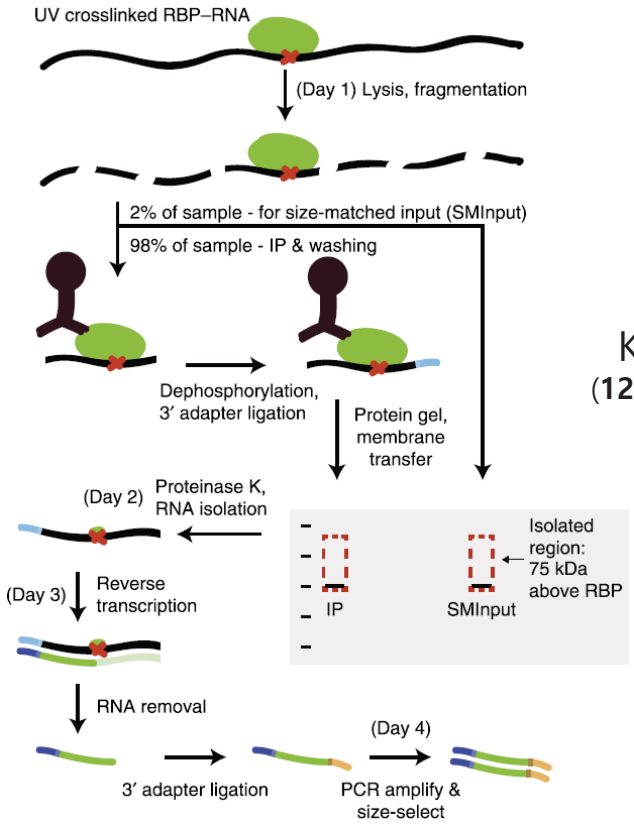


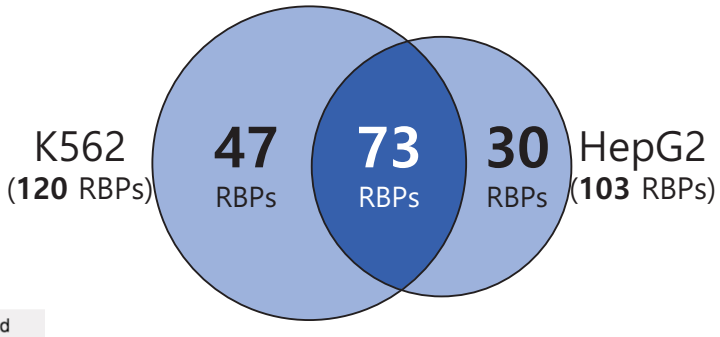
What features of RBPs affect miRNA targeting (MT) efficacy?

- The **distance** to the **5' end** or **3' end** of mRNA?
- The **distance** to the **CDS start** or **CDS end**?
- The **distance** to the **miRNA target site**?
- The **number** of RBP binding sites (RBSs)?
- The **density** of RBS?
- The **intensity** of RBP binding?

⋮

ENCODE eCLIP-Seq 데이터(mRNA 상의 RBP-결합 위치 정보)

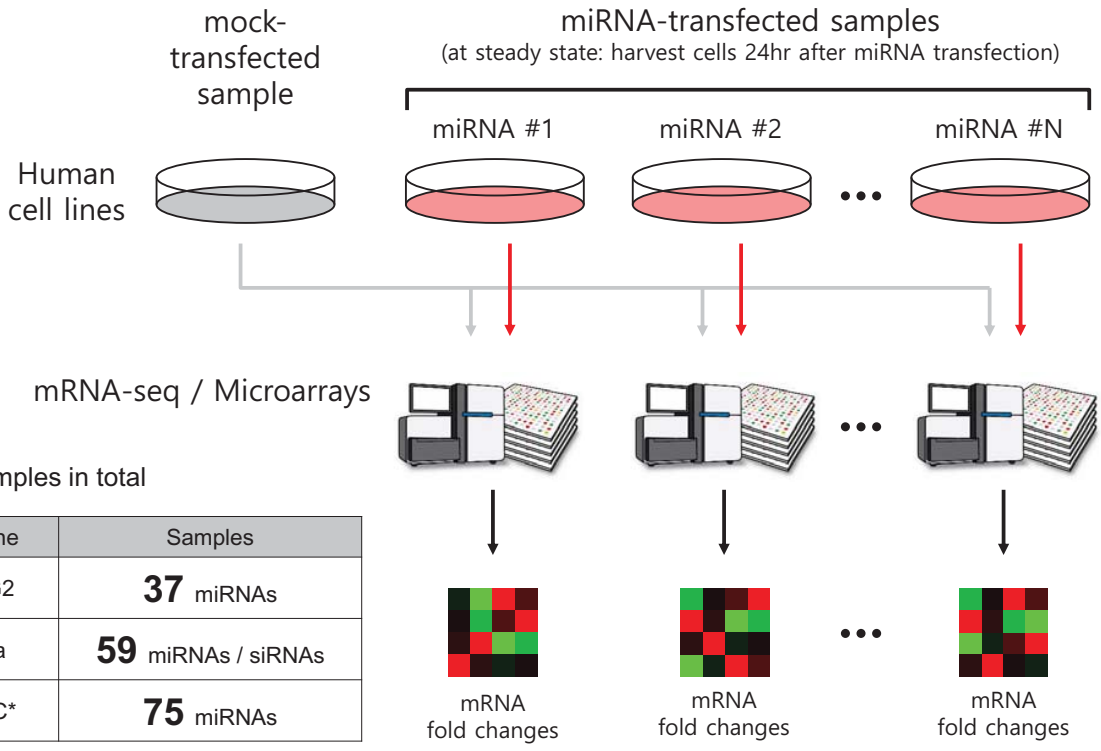




150 RBPs

(Nostrand *et al.*, 2016, Nature Methods)

대규모 전사체 데이터를 통한 MT 효율 측정

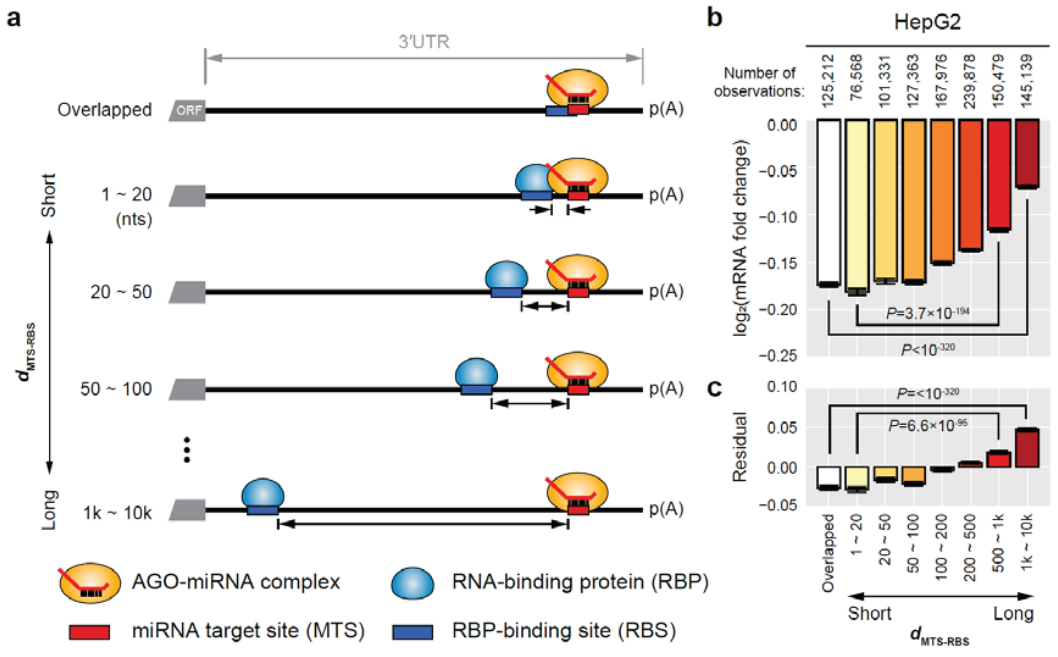


171 samples in total

Cell line	Samples
HepG2	**37** miRNAs
HeLa	**59** miRNAs / siRNAs
OHCC*	**75** miRNAs

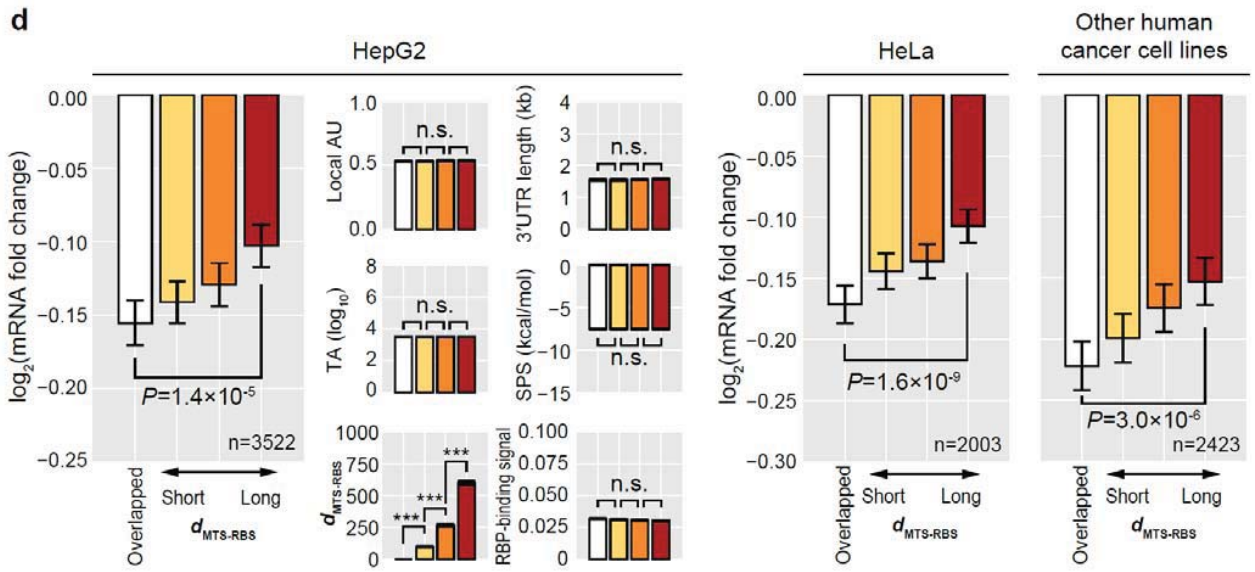
* OHCC: Other human cancer cell lines

RBP-결합 위치에 따른 MT 효율의 영향



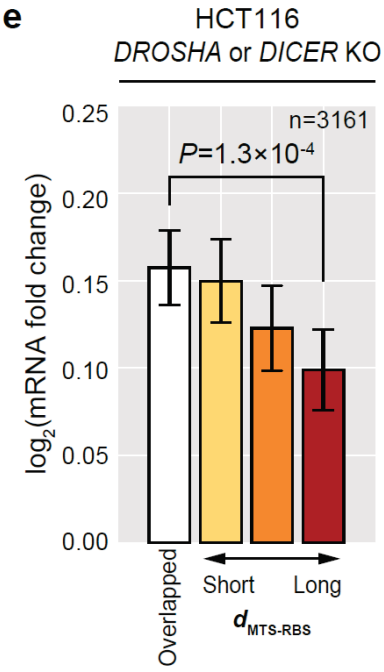
RBP-결합 위치가 miRNA 타겟 사이트에 가까울수록 MT 효율이 강하게 증가함 ➔ RBP가 MT 인헨서로 동작

RBP는 강력한 MT 인헨서



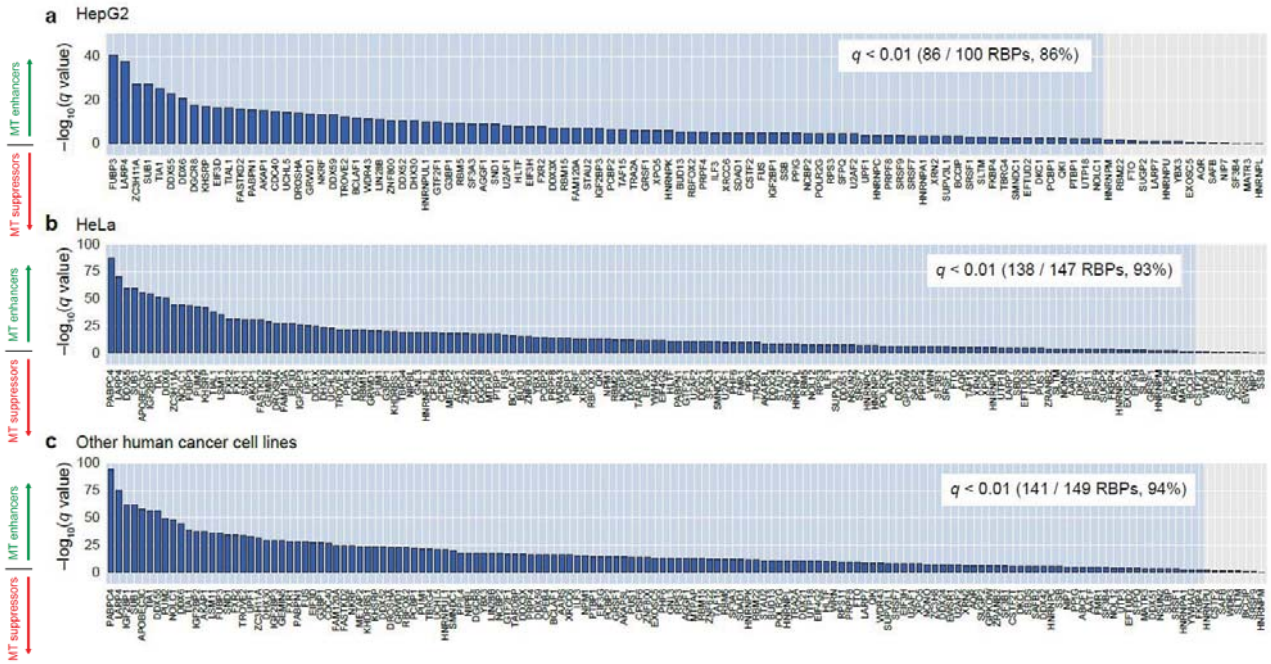
서로 다른 여러가지 세포 조건에서 동일 현상 관찰됨

Endogenous 조건에서도 RBP는 강력한 MT 인헨서로 동작



miRNA를 과발현시킨 조건뿐만 아니라
miRNA KO시킨 endogenous 조건에서도 동일 현상 관찰됨

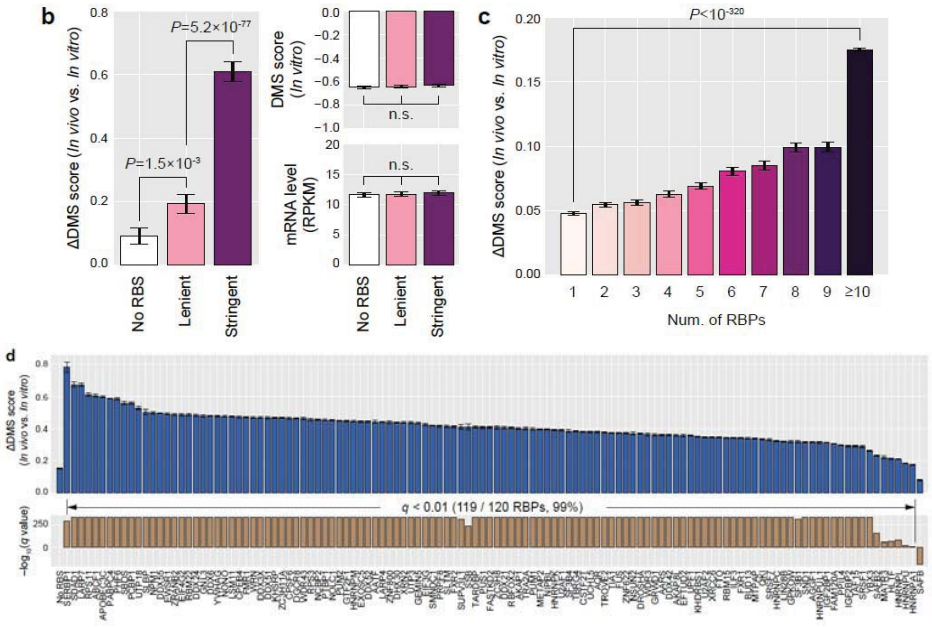
거의 모든 RBP들이 MT 인헨서로 동작



RBP 개별 분석: ≥86% RBP들이 MT 인헨서로 동작하고
MT 서프레서로는 단 하나도 동작하지 않음

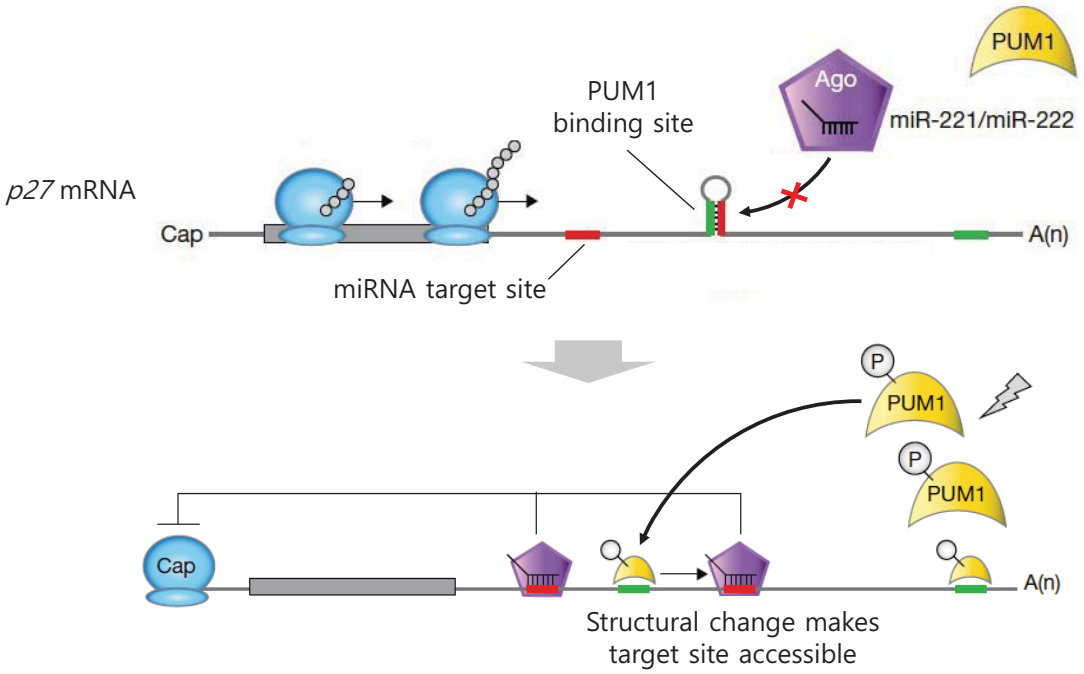
기작: RBP-결합이 mRNA의 Secondary Structure를 Open

- DMS-seq detects unpaired nucleotides *in vivo*.



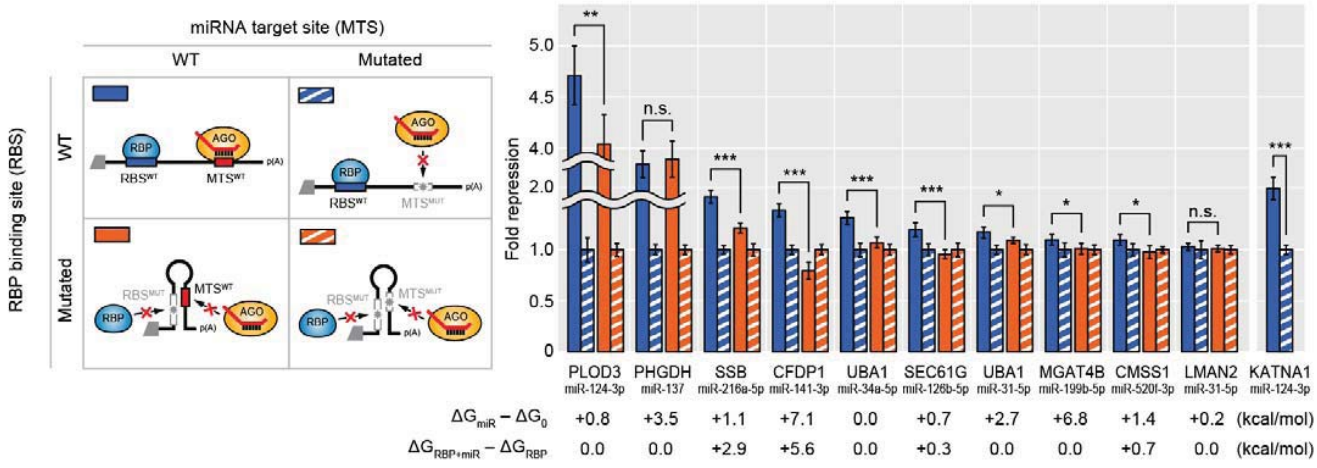
PUM1에서처럼, RBP-결합이 mRNA의
secondary structure를 open하여 MT 효율을 증가

MT에 영향을 주는 RBP: PUM1



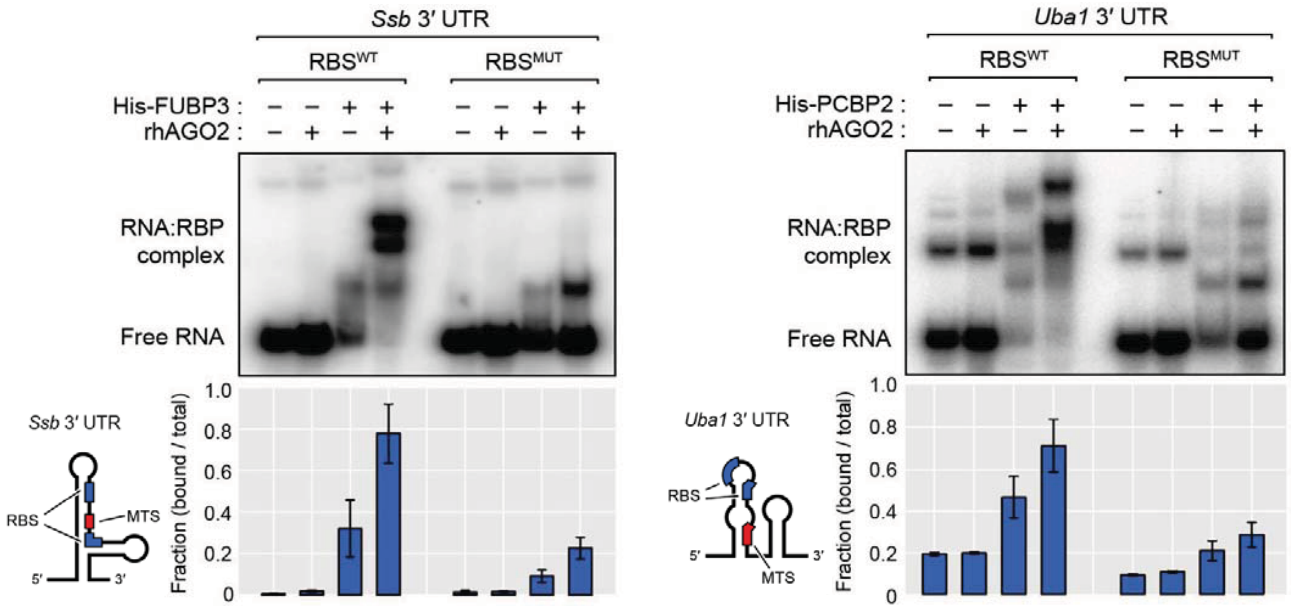
(Triboulet *et al.*, 2010, Nature Cell Biology)

실험적 검증: Luciferase Reporter Assay



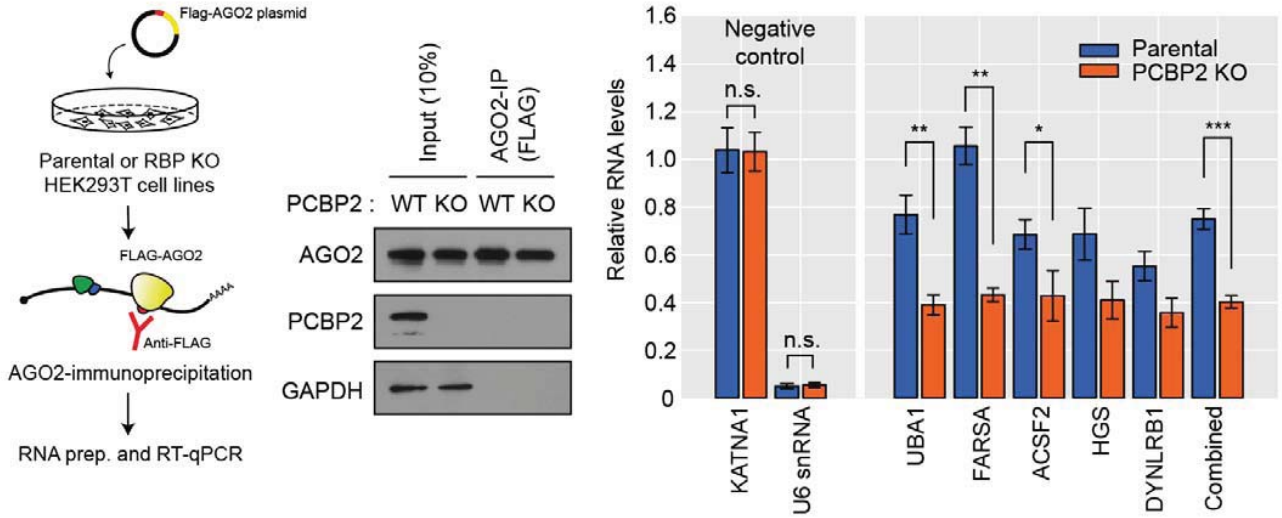
80%의 3'UTR에서 RBP가 MT 효율을 직접적으로 제어

AGO Binding Changes *in vitro* - Gel Shift Assay (EMSA)



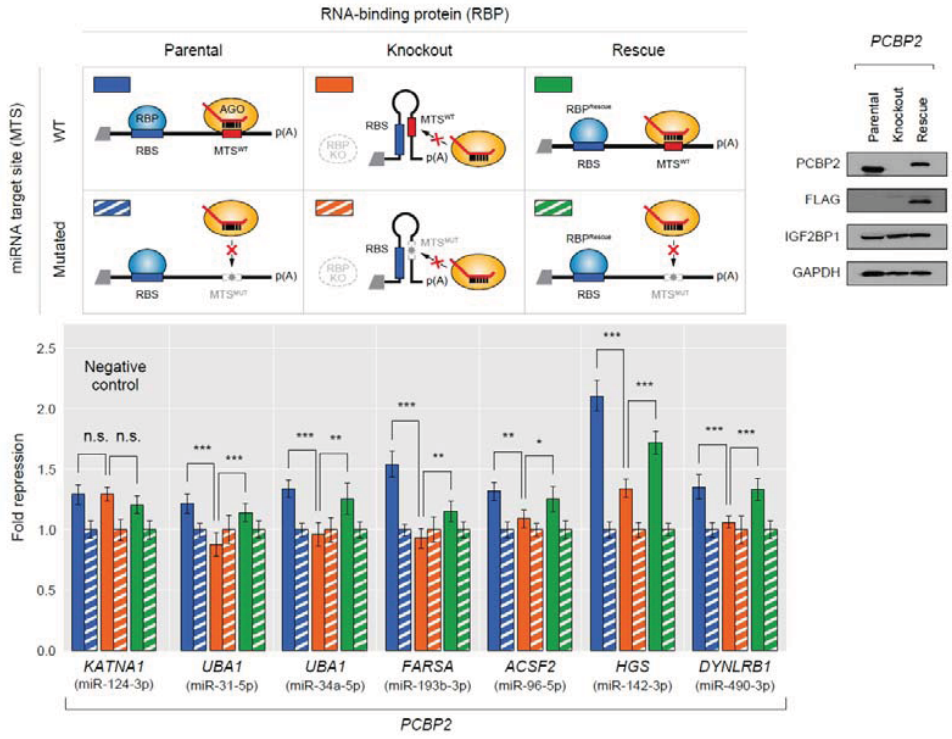
The Disrupted RBS Reduces AGO Binding to the MTS.

AGO Binding Changes *in vivo* - RNA IP Experiment



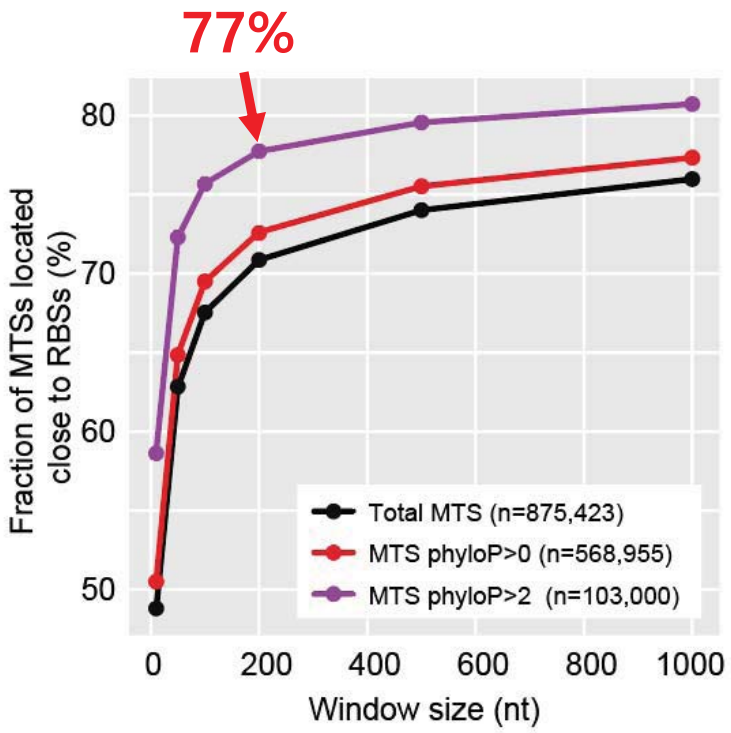
The Absence of an RBP Reduces AGO Binding to the MTS *in vivo*.

Endogenous 조건에서의 실험적 검증: PCBP2 KO



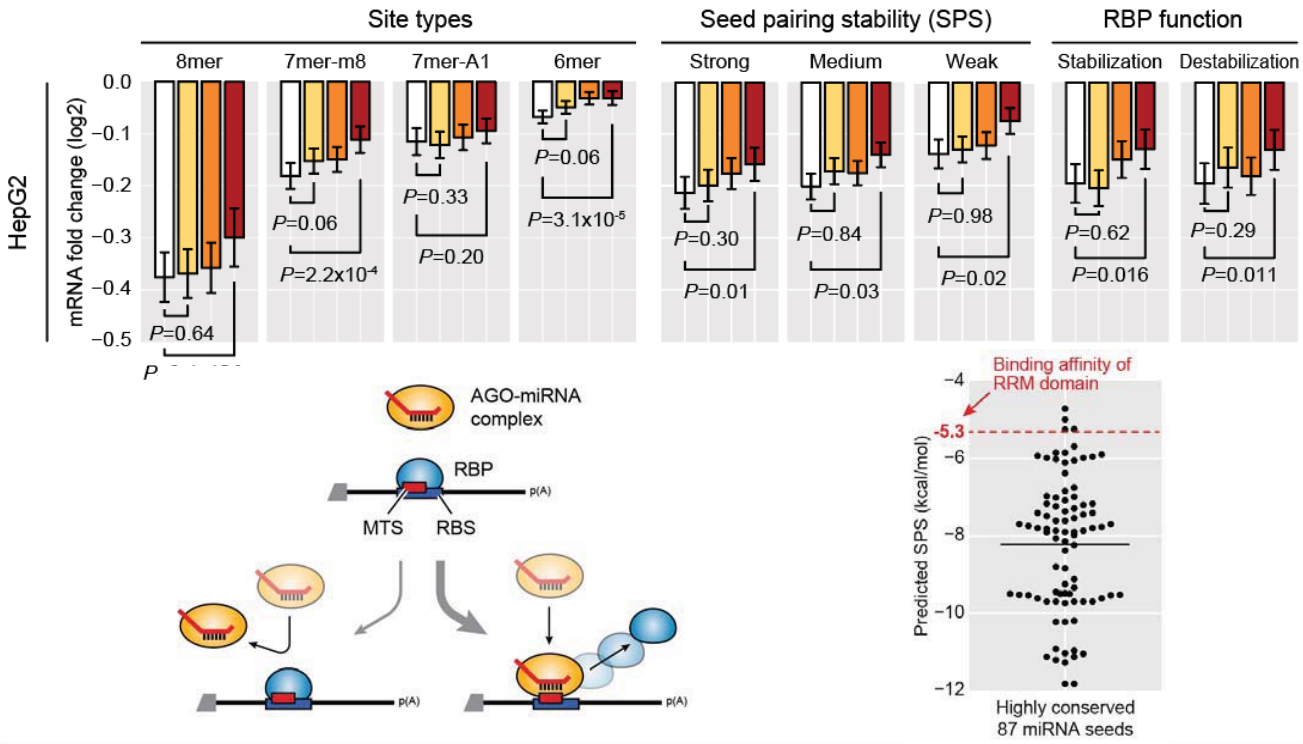
RBP를 제거하자 MT 효율이 유의미하게 감소

Evolutionary Insight – Widespread Regulatory Impact



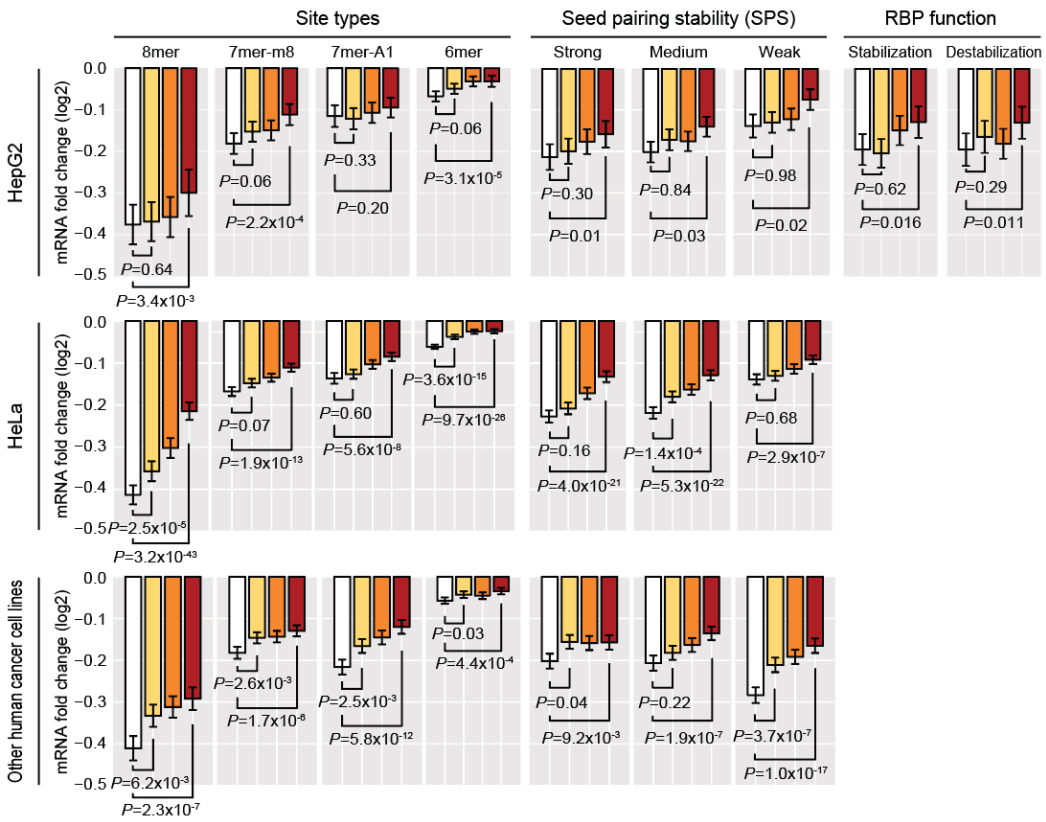
>77% of Conserved Target Sites Contain 1≥ RBSs in Their Vicinities

Overlapping Sites: MTSs May Outcompete RBSs



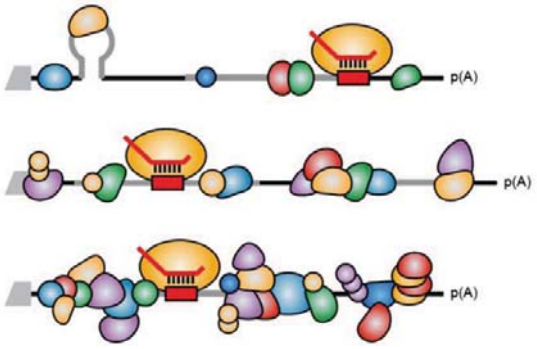
For Numerous Cases, MTSs Consistently Outcompete the Overlapping RBSs

Overlapping Sites: MTSs May Outcompete RBSs



Conclusions

- To gain a global insight into the regulatory impact of RBPs on MT, we have systematically evaluated the quantitative effect of 117 RBPs on MT efficacy.
- Most RBPs, if not all, significantly enhance MT, while no RBP detectably suppresses MT on a global scale.
- RBPs make the local secondary structure of the MTS more accessible to AGO and therefore enhance MT.
- MT should be understood in a context of hundreds of co-regulating RBPs rather than the currently accepted simplified model of a ternary interplay between AGO, miRNA, and mRNA target.
- Our study illuminates the previously unappreciated, widespread regulatory impact of RBPs on MT, unveiling the complex nature of the gene regulatory network governed by metazoan miRNAs.



nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-25078-5 OPEN

The regulatory impact of RNA-binding proteins on microRNA targeting

Sukjun Kim[1,11], Soyoung Kim[2,11], Hee Ryung Chang[1,11], Doyeon Kim[1,11], Junehee Park[1], Narae Son[1], Joori Park[3,4], Minhyuk Yoon[2], Gwangung Chae[2], Young-Kook Kim[5], V. Narry Kim[1,6], Yoon Ki Kim[3,4], Jin-Wu Nam[7], Chanseok Shin[2,8,9✉] & Daehyun Baek[1,10✉]

Argonaute is the primary mediator of metazoan miRNA targeting (MT). Among the currently identified >1,500 human RNA-binding proteins (RBPs), there are only a handful of RBPs known to enhance MT and several others reported to suppress MT, leaving the global impact of RBPs on MT elusive. In this study, we have systematically analyzed transcriptome-wide binding sites for 150 human RBPs and evaluated the quantitative effect of individual RBPs on MT efficacy. In contrast to previous studies, we show that most RBPs significantly affect MT and that all of those MT-regulating RBPs function as MT enhancers rather than suppressors, by making the local secondary structure of the target site accessible to Argonaute. Our findings illuminate the unappreciated regulatory impact of human RBPs on MT, and as these RBPs may play key roles in the gene regulatory network governed by metazoan miRNAs, MT should be understood in the context of co-regulating RBPs.

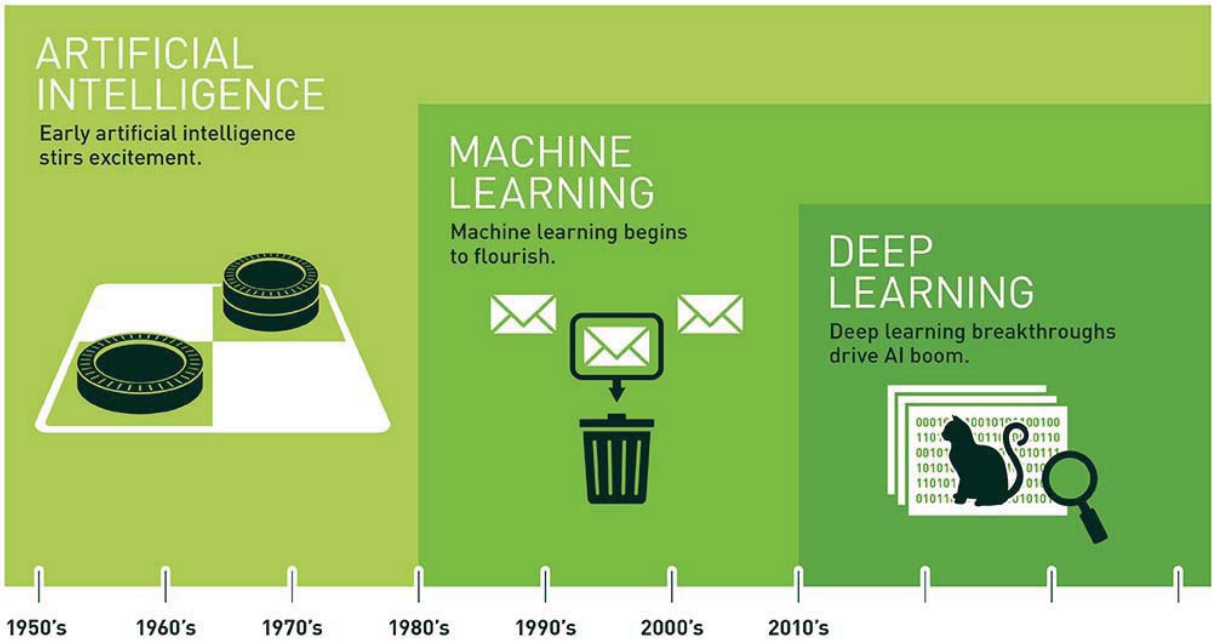
[1] School of Biological Sciences, Seoul National University, Seoul, Republic of Korea. [2] Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea. [3] Creative Research Initiatives Center for Molecular Biology of Translation, Korea University, Seoul, Republic of Korea. [4] Division of Life Sciences, Korea University, Seoul, Republic of Korea. [5] Department of Biochemistry, Chonnam National University Medical School, Hwasun, Jeollanam-do, Republic of Korea. [6] Center for RNA Research, Institute for Basic Science, Seoul, Republic of Korea. [7] Department of Life Science, College of Natural Sciences, Hanyang University, Seoul, Republic of Korea. [8] Research Institute of Agriculture and Life Sciences, and Plant Genomics and Breeding Institute, Seoul National University, Seoul, Republic of Korea. [9] Research Center for Plant Plasticity, Seoul National University, Seoul, Republic of Korea. [10] Bioinformatics Institute, Seoul National University, Seoul, Republic of Korea. [11] These authors contributed equally: Sukjun Kim, Soyoung Kim, Hee Ryung Chang, Doyeon Kim. ✉email: cshin@snu.ac.kr; baek@snu.ac.kr

AI Prediction for Functional MicroRNA Targeting

Daehyun Baek

School of Biological Sciences
Seoul National University

Artificial Intelligence vs. Deep Learning

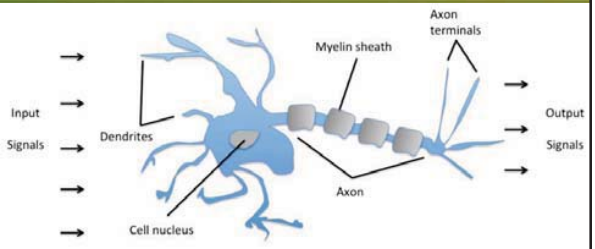


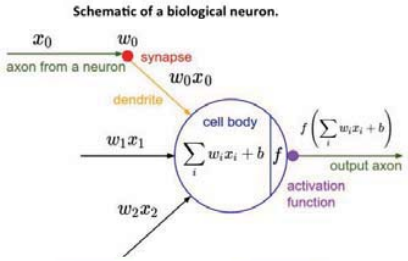
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

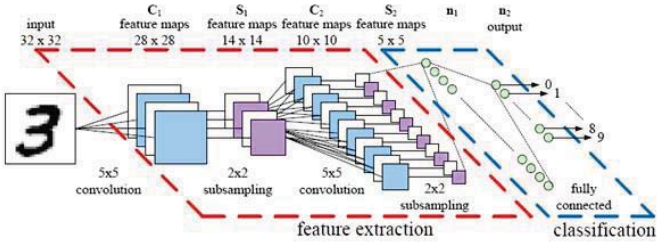
(blogs.nvidia.com)

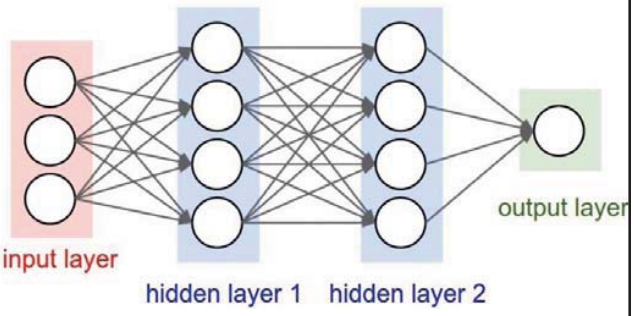
Deep Learning 기반의 miRNA 타겟 예측

- Artificial neural network with multiple layer of simple but non-linear functions
- Good at high-dimensional, big data
- Convolutional neural network (CNN): Addition of feature extraction step for image and reduction of the number of model parameters



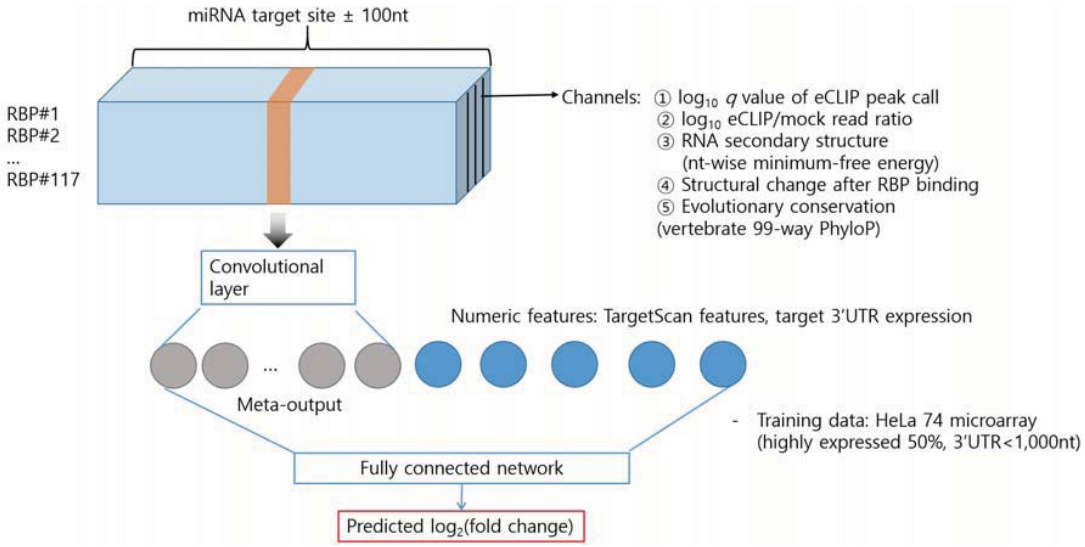


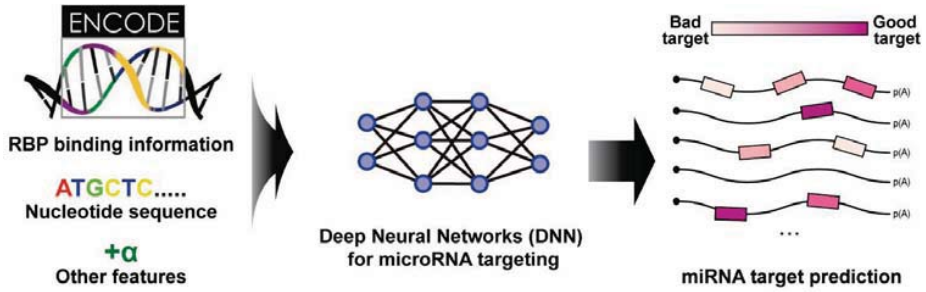




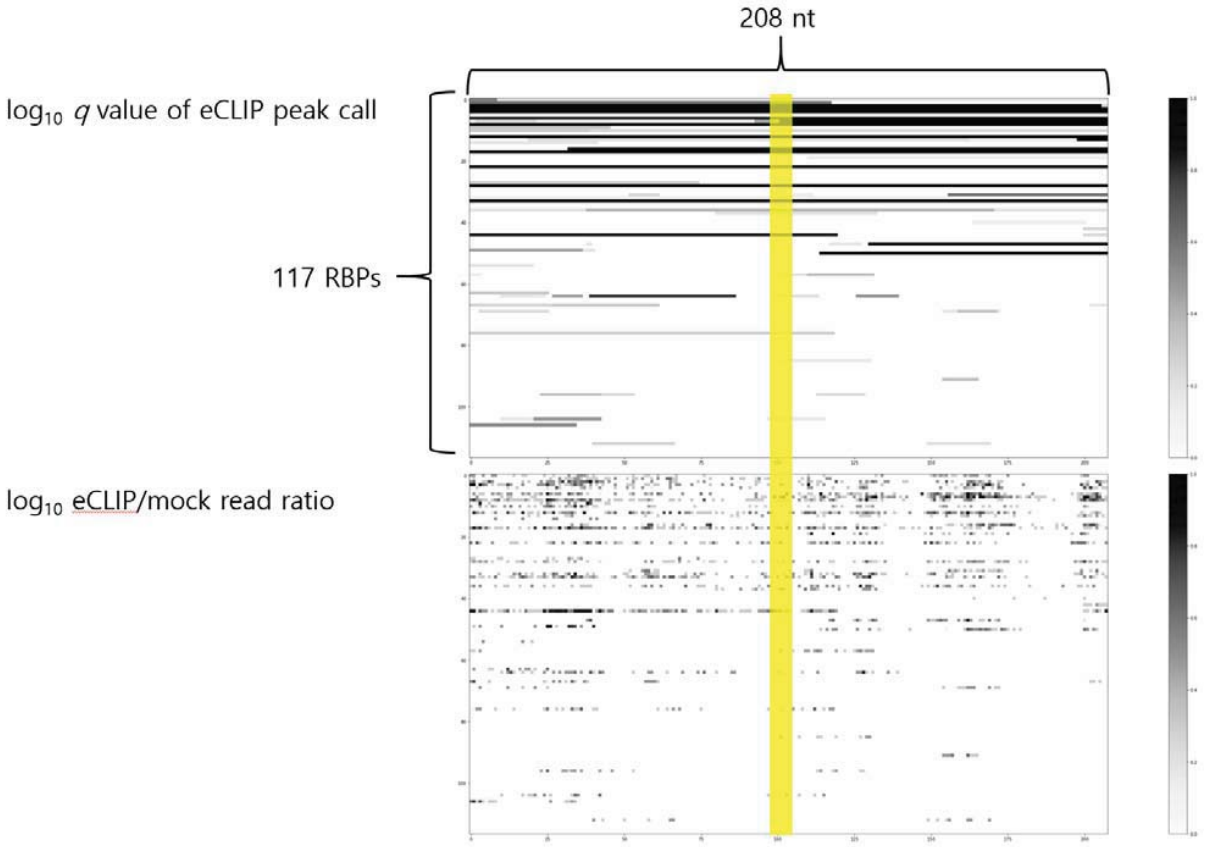
(www.cs231n.github.io)

Convolution Neural Network(CNN) 모델 for MT

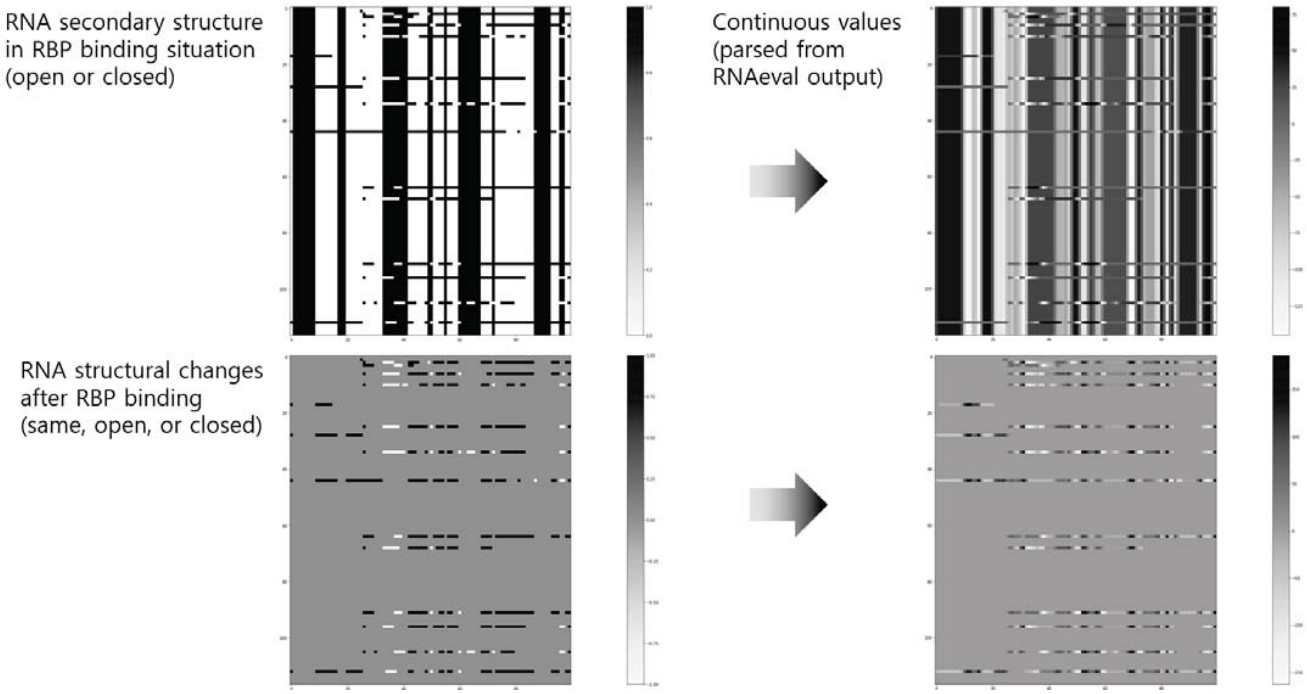




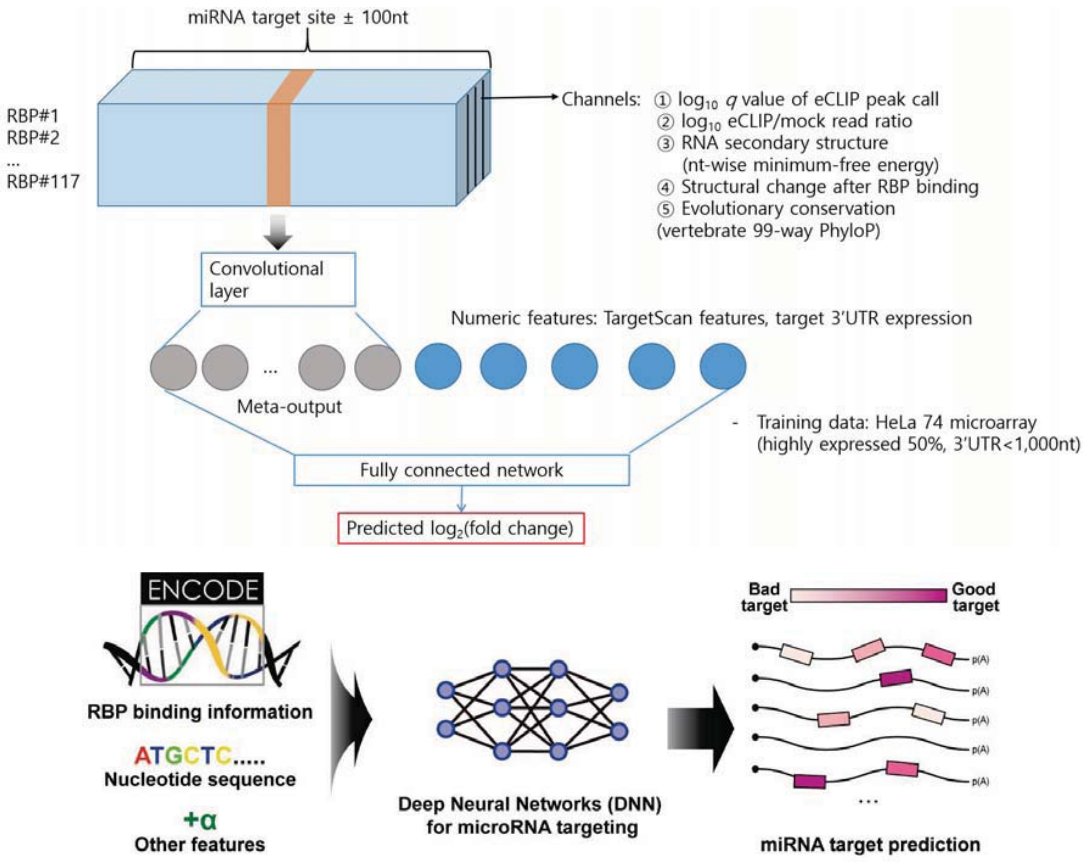
CNN Features: RBP Binding Information



CNN Features: RNA Secondary Structure

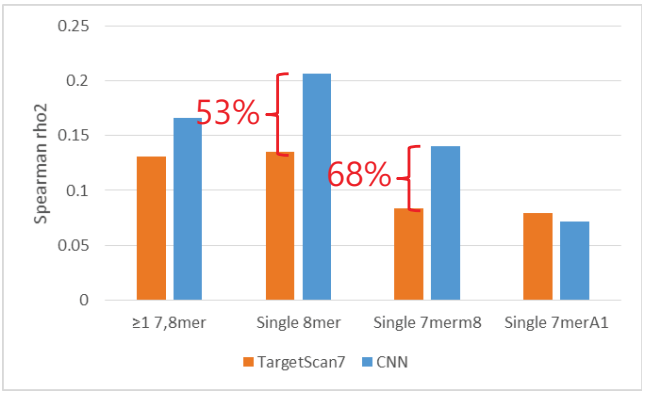


Convolution Neural Network(CNN) 모델 for MT

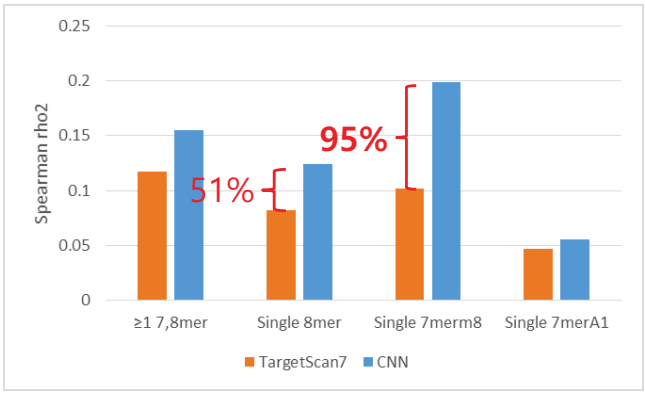


Deep Learning 기반의 miRNA 타겟 예측

HeLa



HCT116



TargetScanHuman
Prediction of microRNA targets
Release 7.2: March 2018

RBP-결합 정보를 활용하는 Deep Learning 적용 결과, miRNA 타겟 예측 정확도가 대폭 향상

A High-Resolution Temporal Atlas of the SARS-CoV-2 Translatome and Transcriptome

Daehyun Baek

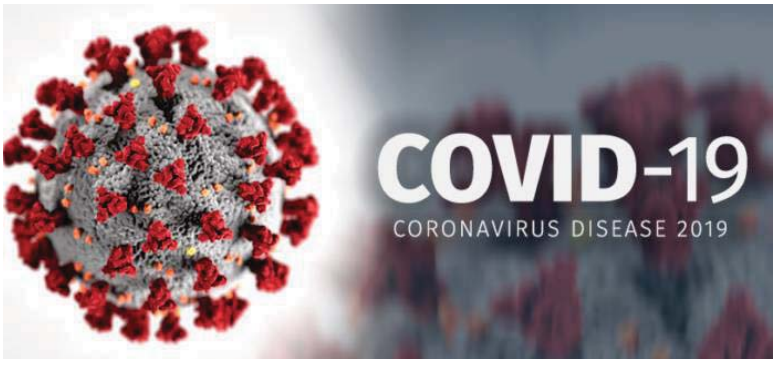
School of Biological Sciences
Seoul National University

COVID-19

- COVID-19 is caused by severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2), which infected >34 million people resulting in >1 million deaths.

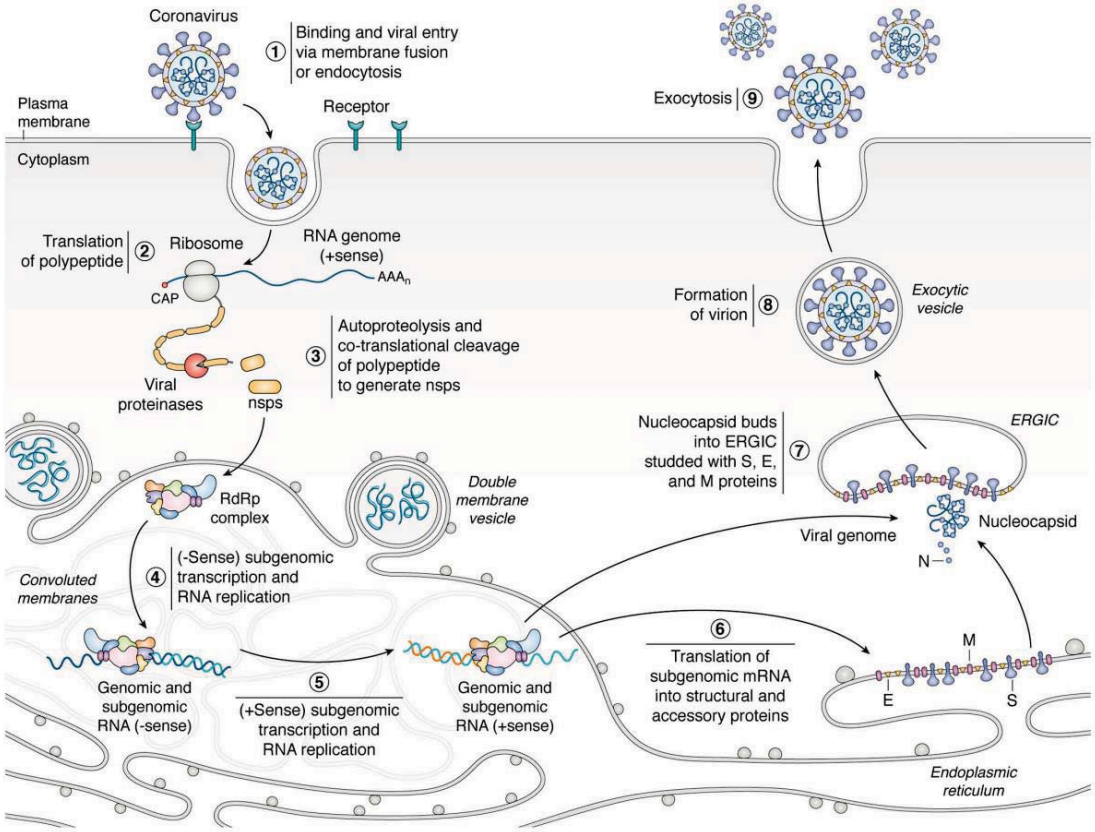
- As the United Nations has recently declared, COVID-19 is not only a pandemic but also a substantial crisis deeply affecting the societies and economics on a global scale

- Although the SARS-CoV-2 transcriptome has been recently reported (Kim et al., 2020), temporal landscape of the SARS-CoV-2 translatome and its impact on the human genome remain unexplored.



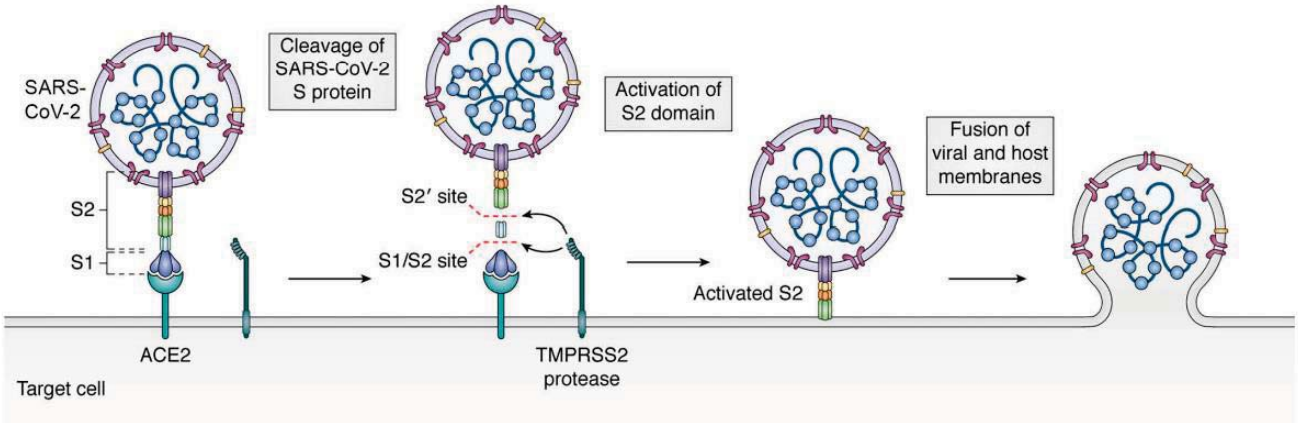
(www.fscluster.org)

The Viral Life Cycle of Coronavirus



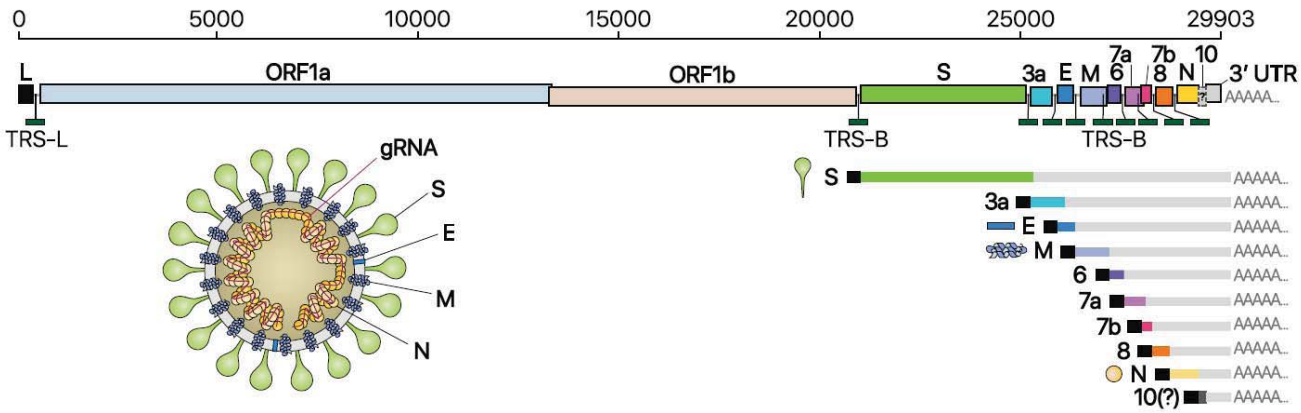
(Hartenian *et al.*, 2020)

The Mechanism of SARS-CoV-2 Viral Entry



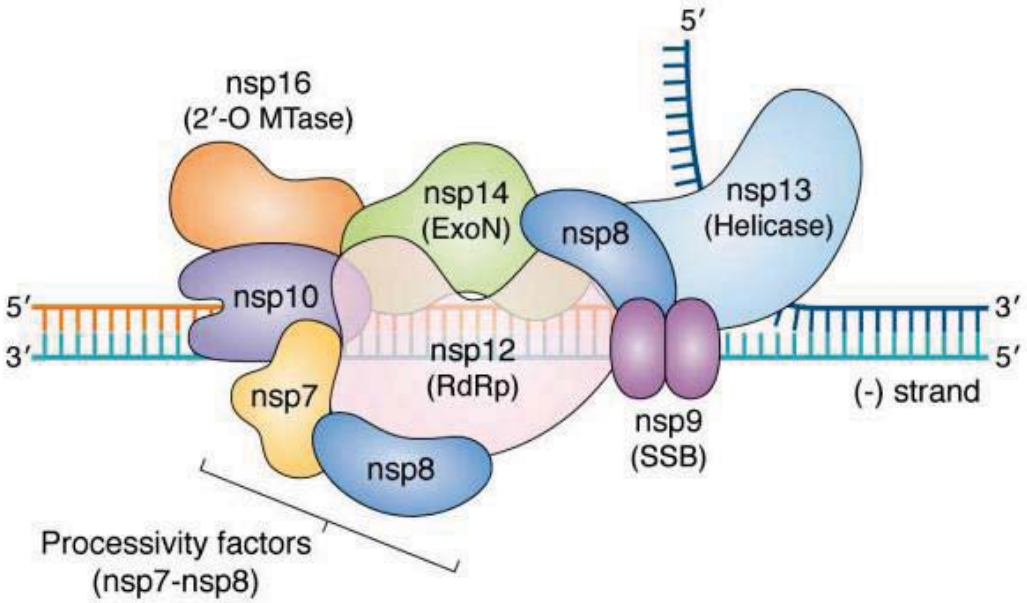
(Hartenian *et al.*, 2020)

Genome Organization of SARS-CoV-2



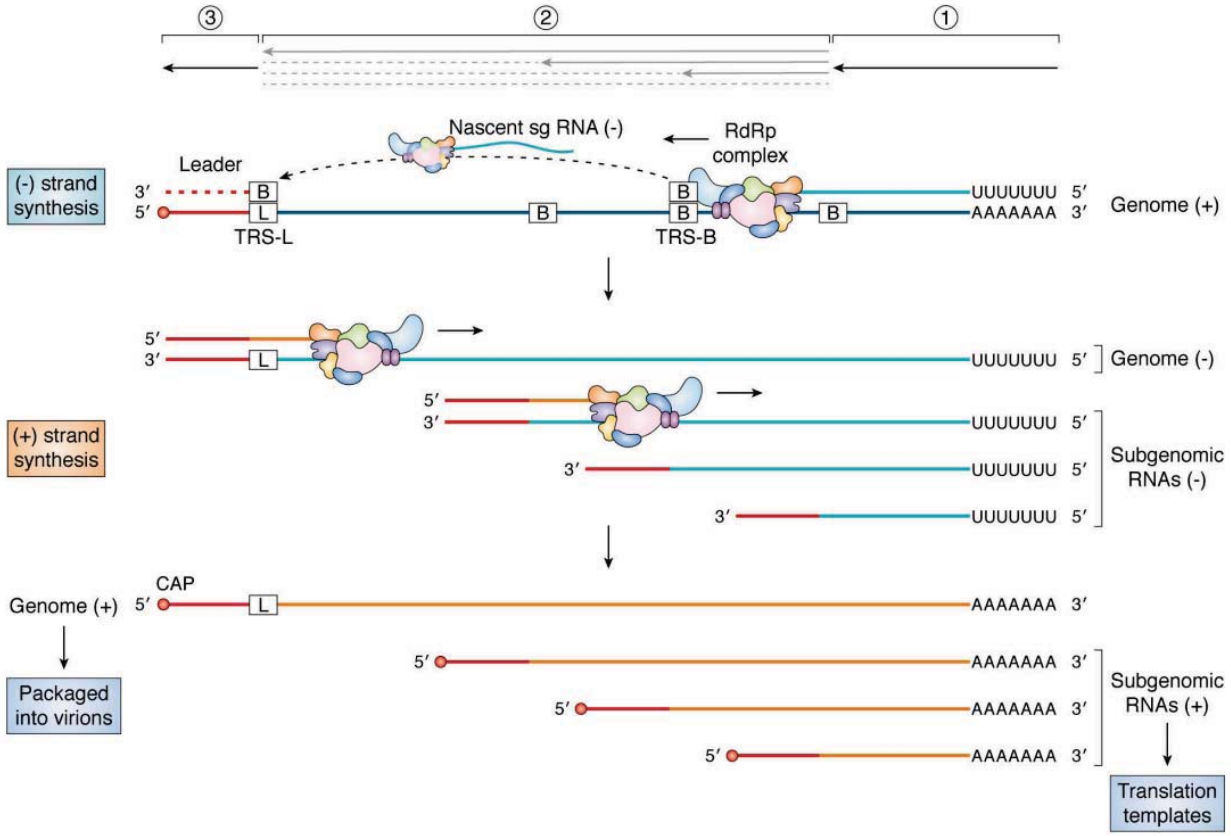
(Kim et al., Cell, 2020)

Model of Putative Coronavirus Replisome



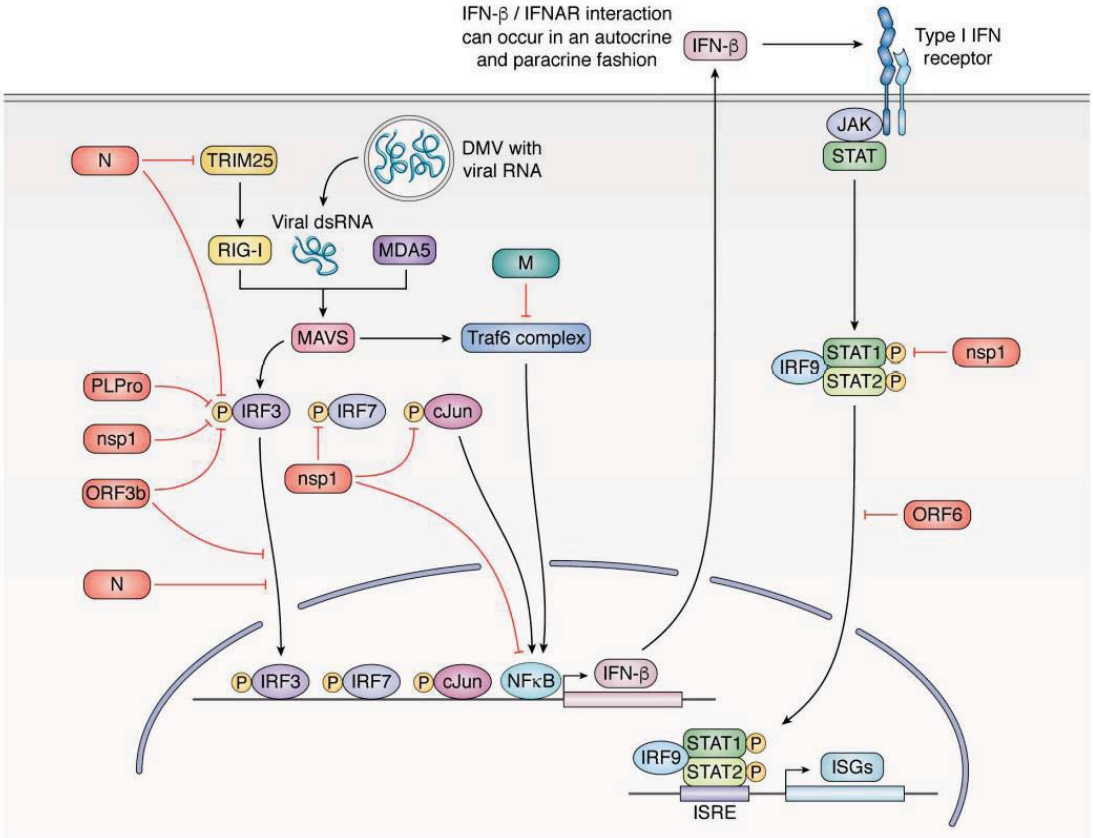
(Hartenian *et al.*, 2020)

Discontinuous Transcription



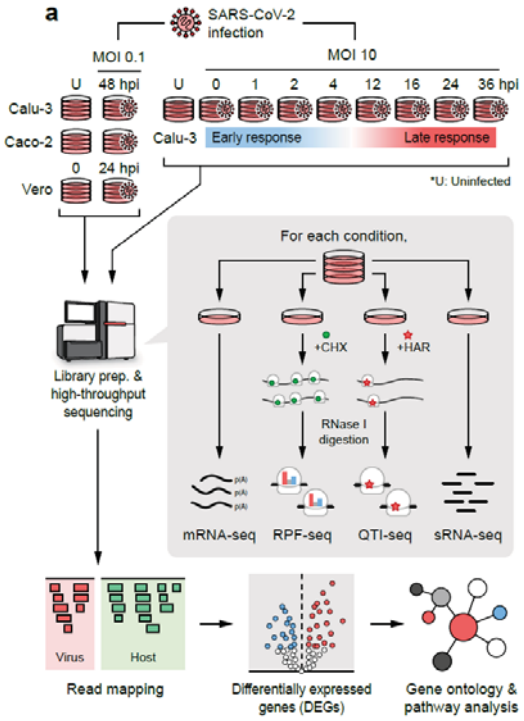
(Hartenian *et al.*, 2020)

Innate Immune Antagonism by SARS-CoV



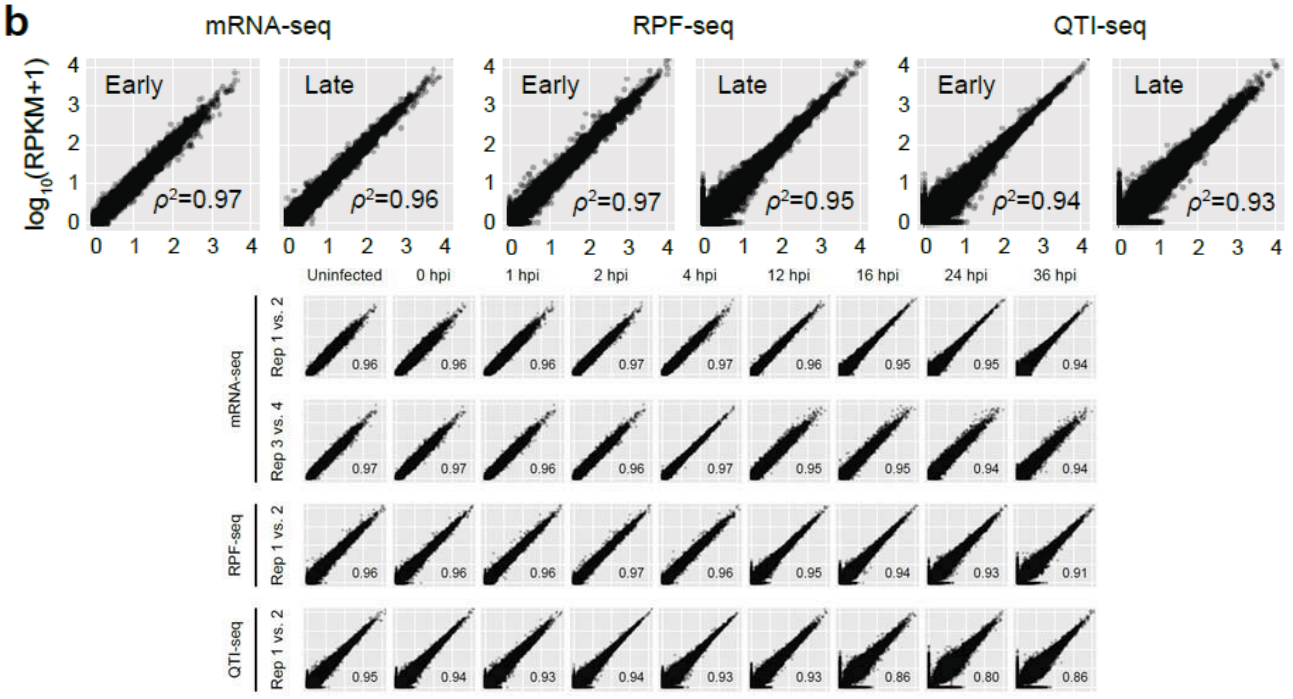
(Hartenian *et al.*, 2020)

Experimental Design



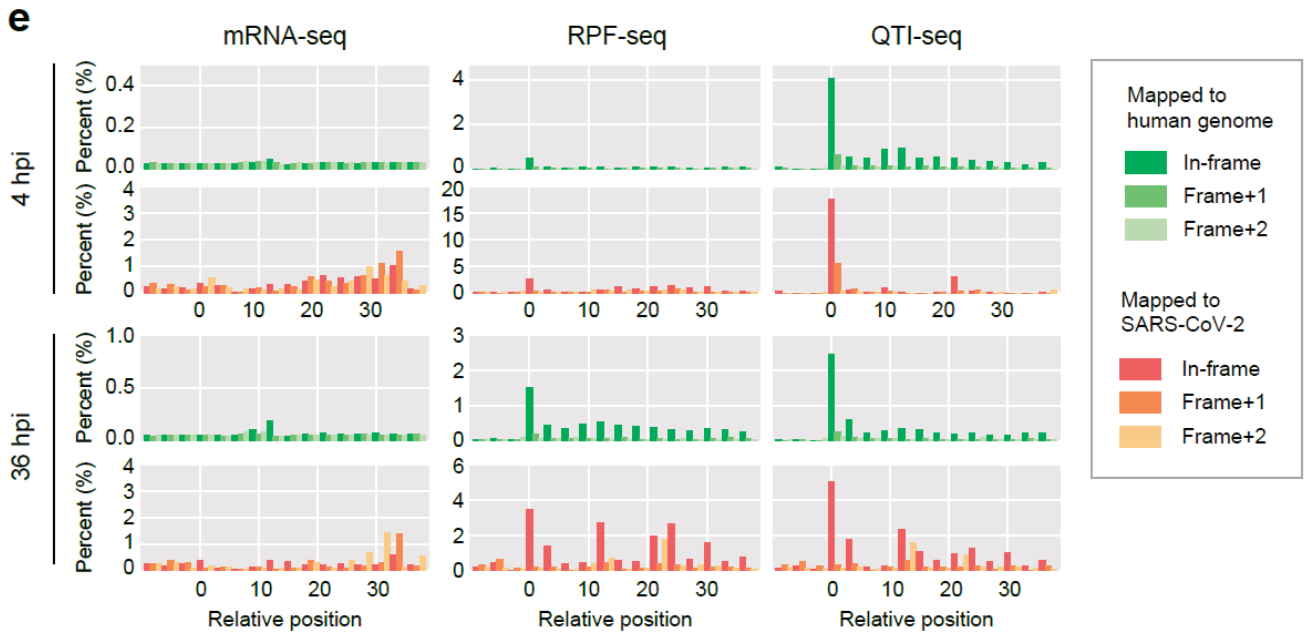
Generation of massive-scale datasets of the SARS-CoV-2 translatome and transcriptome

Highly Reliable Data Quality



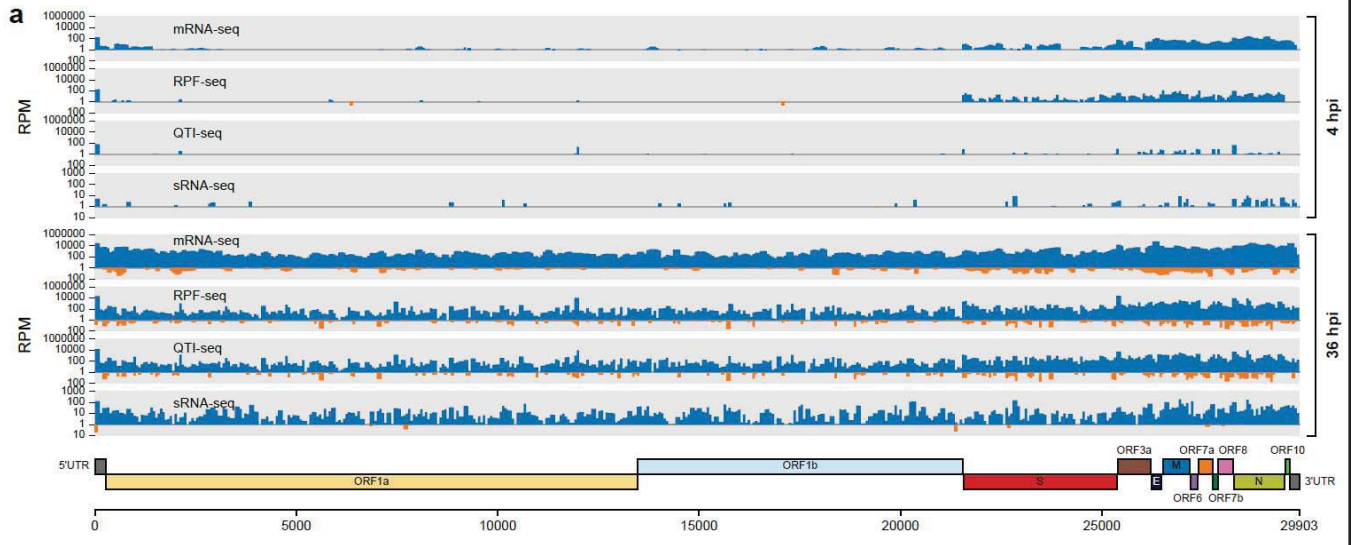
Strong correlation between replicates indicates that our datasets are highly reliable.

Highly Reliable Data Quality



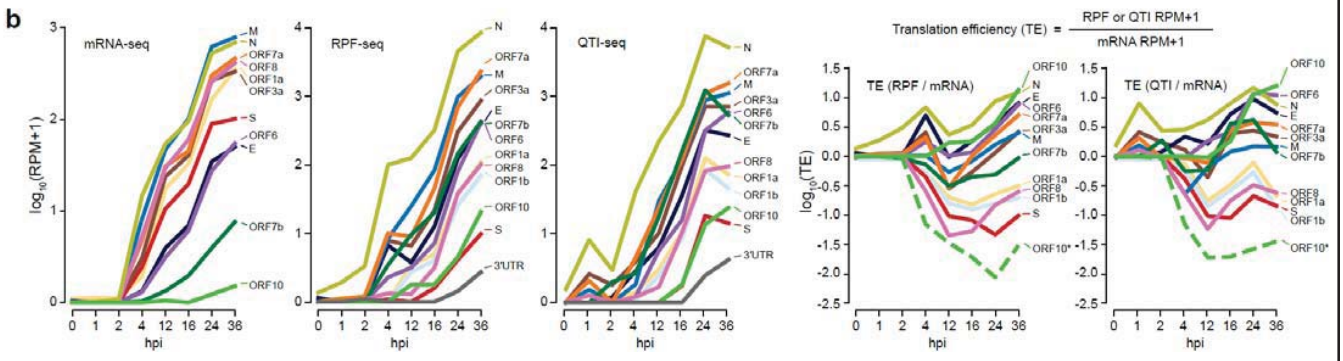
Strong correlation between replicates indicates that our datasets are highly reliable.

A High-Resolution Temporal Atlas of the SARS-CoV-2 Translatome



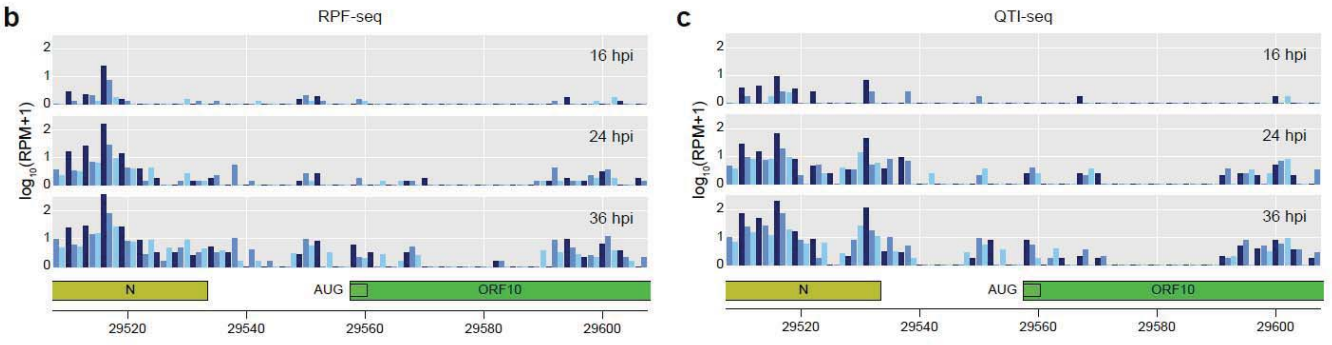
Employing RPF-seq, QTI-seq, mRNA-seq, and sRNA-seq, a temporal atlas of SARS-CoV-2 translatome and transcriptome was constructed.

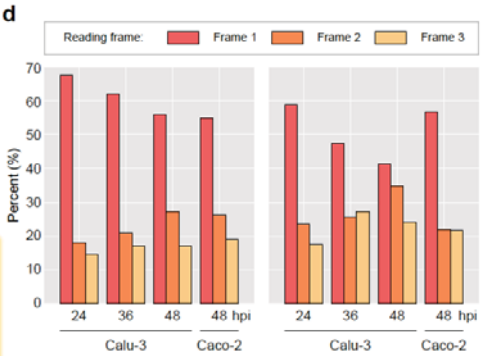
Temporal Expression of Individual SARS-CoV-2 Genes



The overall increment in expression level for all ORFs over time was observed on both mRNA and RPF levels.

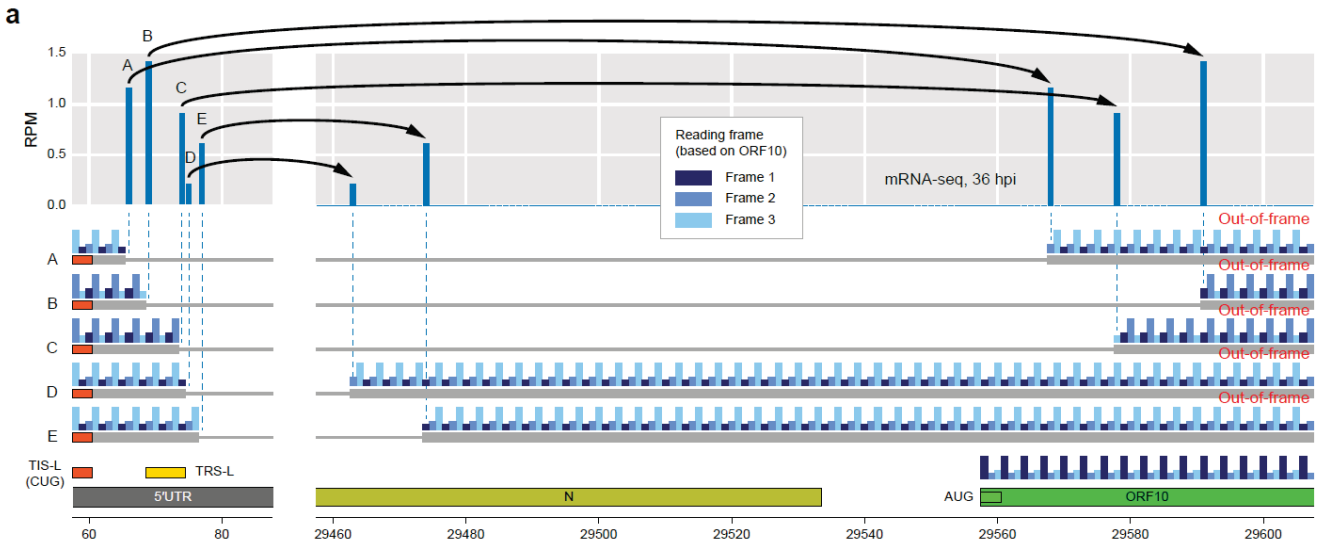
ORF 10 May Be Functional





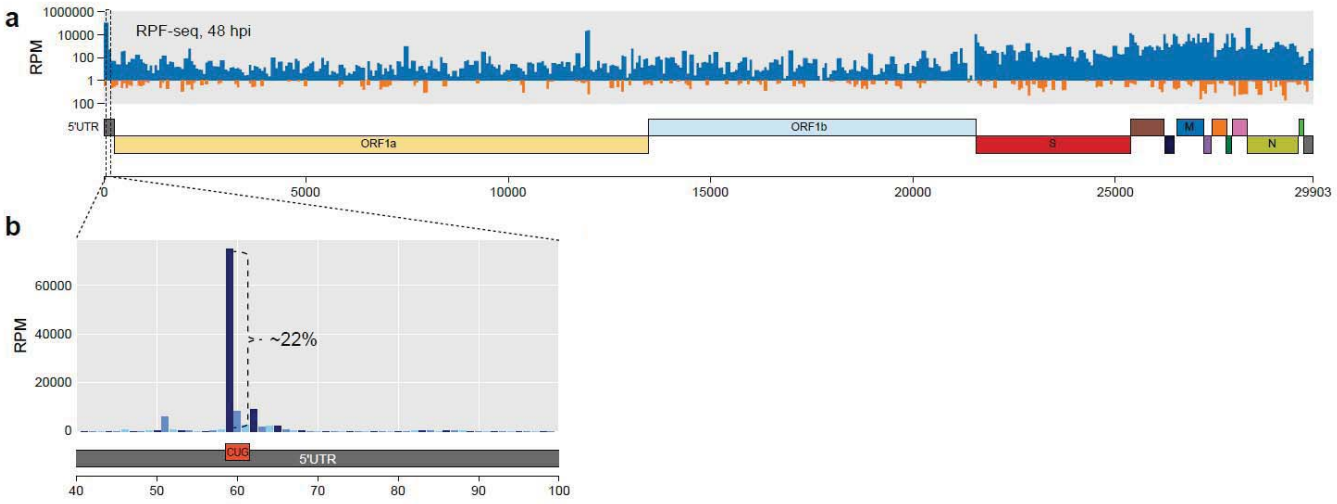
ORF 10 exhibited a very high TE, albeit to its modest RPF level, suggesting that ORF 10 might be functional.

ORF 10 May Be Functional



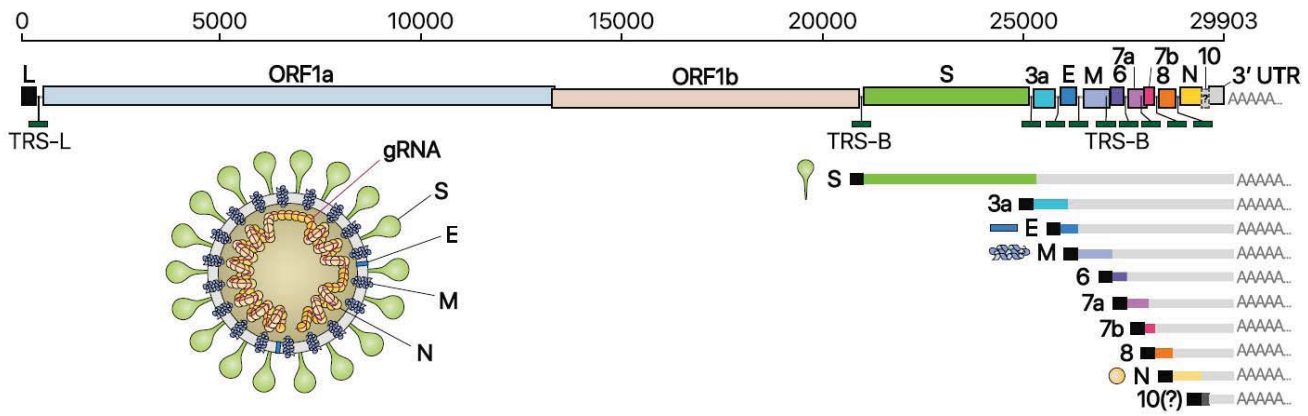
Leaky scanning of ribosome in N sgRNA might lead to the translation of ORF 10.

Translation Initiation Site Located in the Leader (TIS-L)



Substantial amount of the RPF-seq and QTI-seq reads were mapped on a CUG codon located in the leader sequence

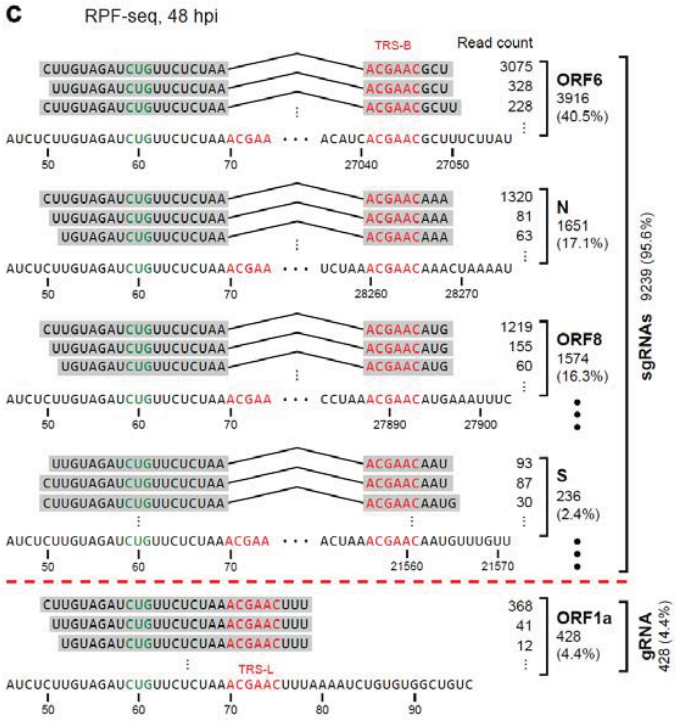
Leader Sequence in SARS-CoV-2



Leader sequence and TIS-L are included in all gRNAs and sgRNAs of SARS-CoV-2

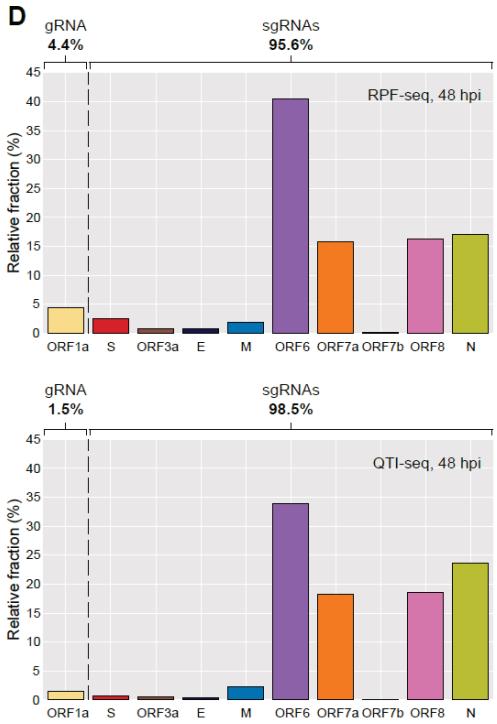
(Kim et al., Cell, 2020)

TIS-L Reads Uniquely Mapped to the SARS-CoV-2 Genome



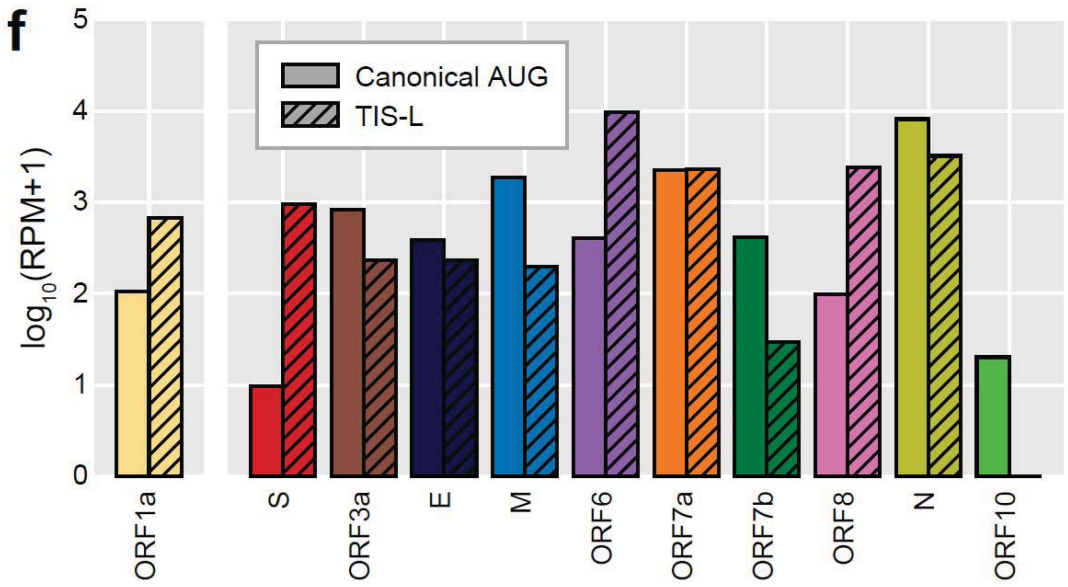
Most RPF-seq and QTI-seq reads (>95%) were mapped to sgRNAs, while <5% of the reads were mapped to gRNA

TIS-L Reads Uniquely Mapped to the SARS-CoV-2 Genome



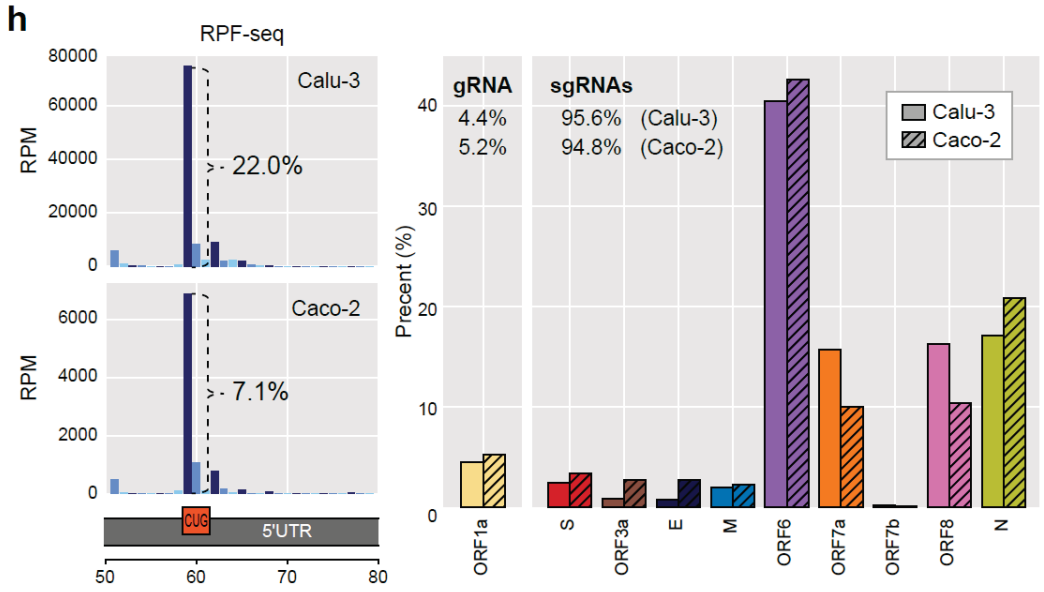
The largest number of TIS-L reads mapped to ORF 6, followed by ORFs N, 8, 7a, 1a, and S.

Translation Initiation at Annotated AUGs vs. TIS-L



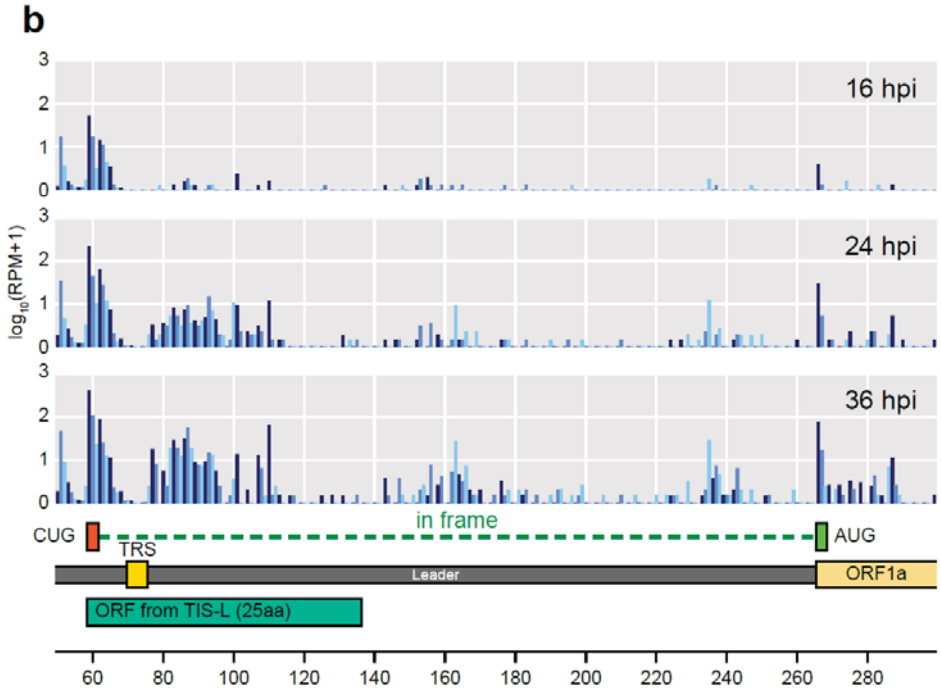
Translation initiation of TIS-L was even higher than that of the annotated AUG for several ORFs including ORF S

Translation Initiation at Annotated AUGs vs. TIS-L



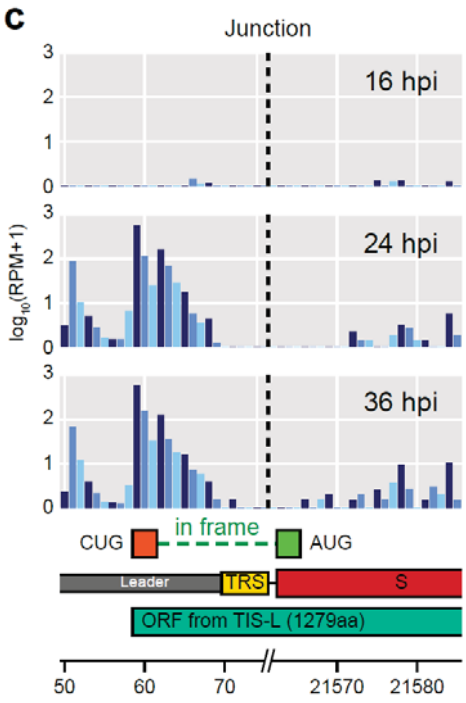
Consistent with Calu-3, a considerable amount of reads were mapped to TIS-L in Caco-2.

TIS-L for ORF 1a



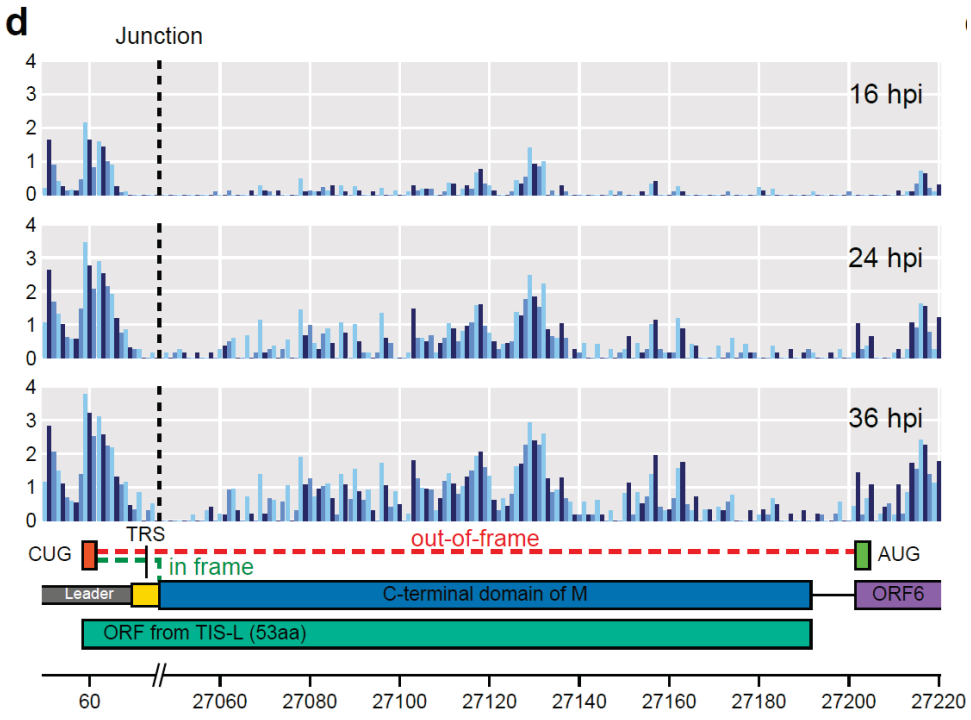
For ORFs 1a and N, TIS-L is expected to create a short uORF that is not overlapping with the annotated ORF

TIS-L for ORF S



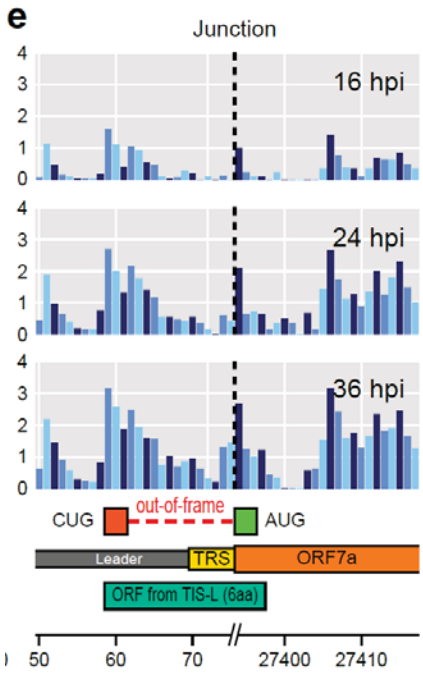
TIS-L is in-frame with ORF S and thus expected to yield an extended ORF or to function as a translation enhancer.

TIS-L for ORF 6



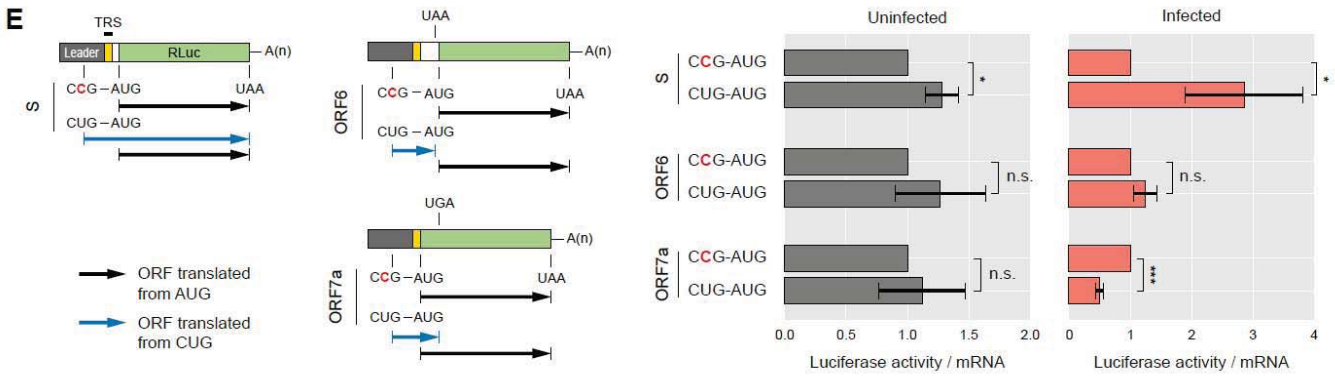
TRS-B of ORF 6 is embedded in the middle of ORF M producing an uORF in-frame with the C-terminal region of ORF M

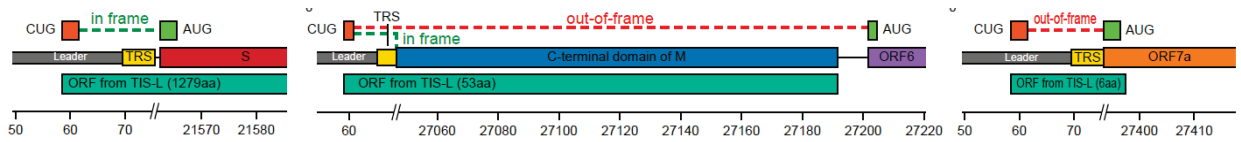
TIS-L for the Other ORFs (3a, E, M, 7a, 7b, and 8)



Most uORFs derived from TIS-L overlap with annotated ORFs and are out of frame with them, likely functioning as a translation suppressor.

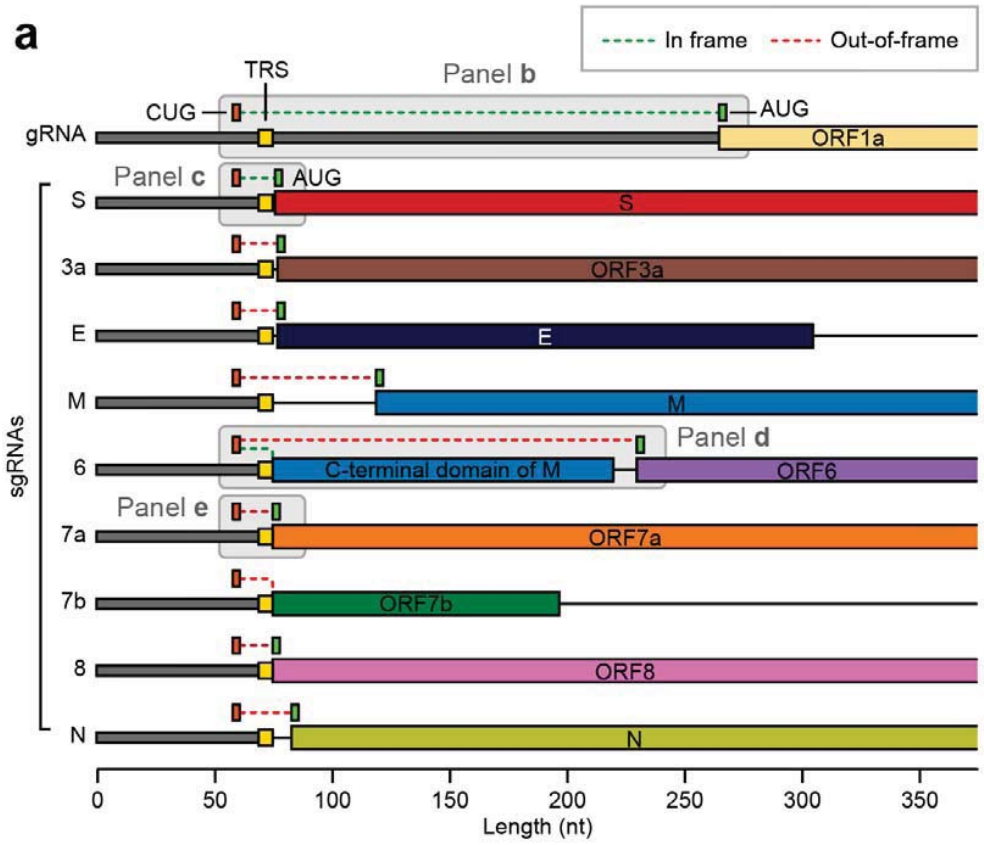
Experimental Validation by Luciferase Reporter Assay





These results demonstrate that TIS-L has a substantial regulatory impact on most SARS-CoV-2 ORFs either positively or negatively.

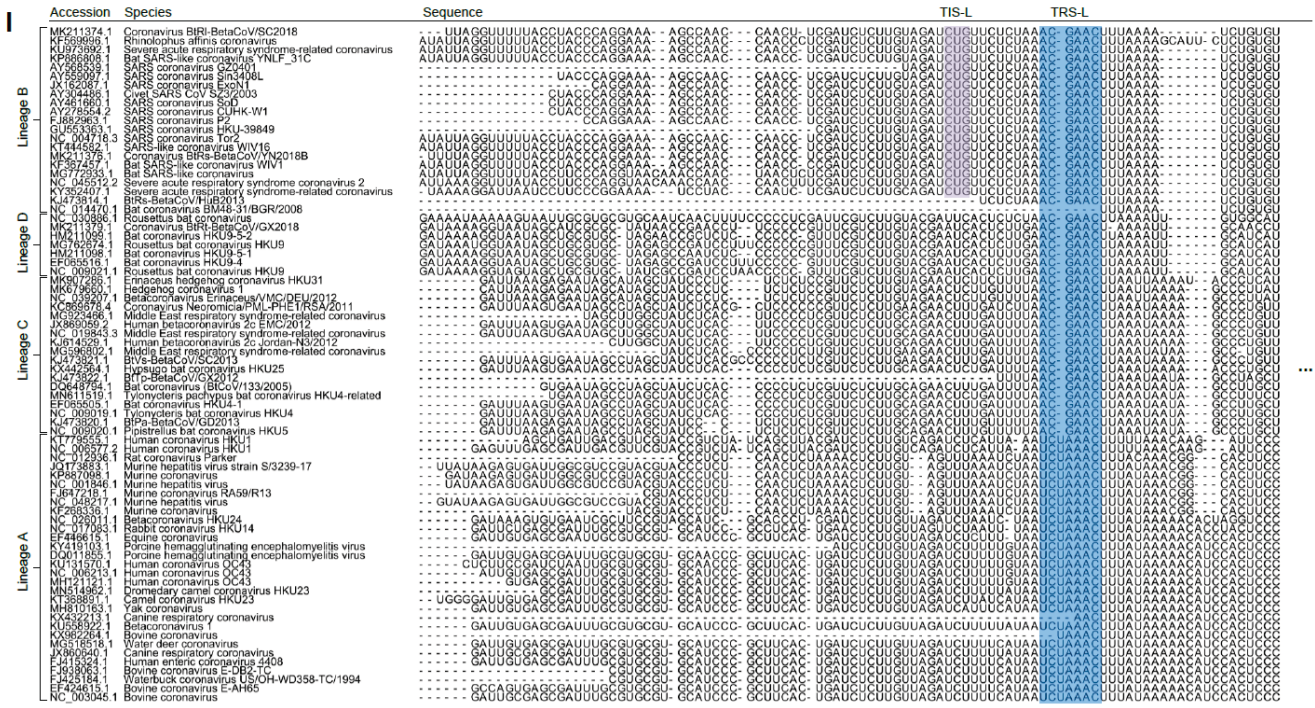
The Impact of TIS-L on the SARS-CoV-2 Translatome



Evolutionary Insight into TIS-L in Betacoronaviruses

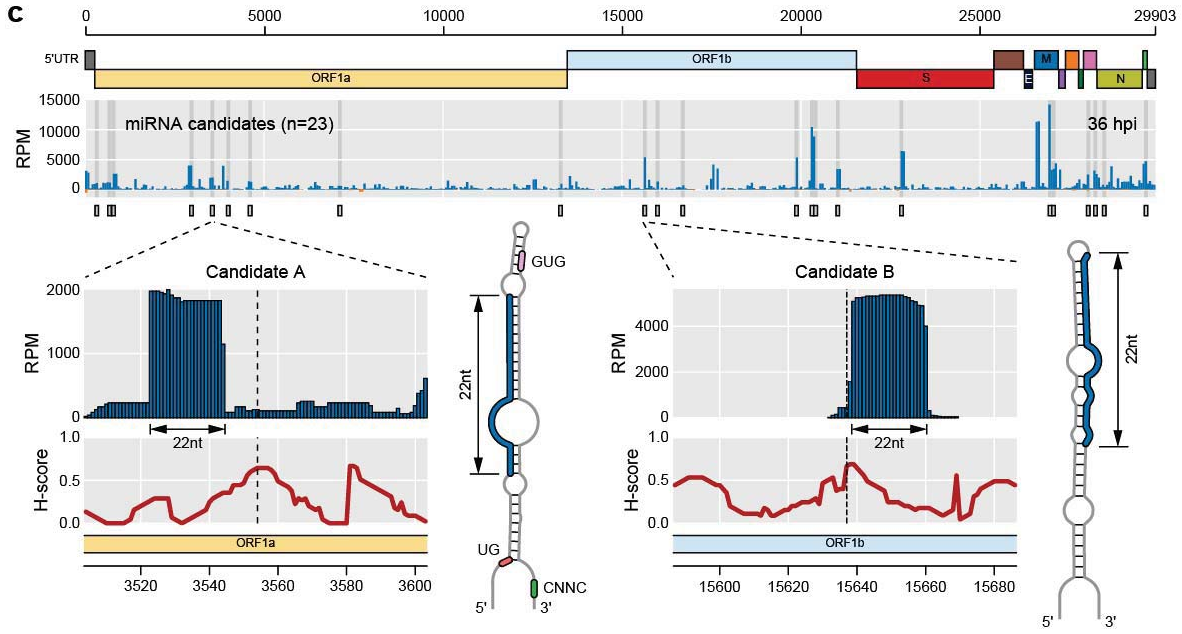
TIS-L may bypass the reduced global translation of the host cells in response to viral infection.

Evolutionary Insight into TIS-L in Betacoronaviruses



TIS-L is highly conserved in most of Lineage B viruses while absent in other lineages.

SARS-CoV-2 MicroRNAs

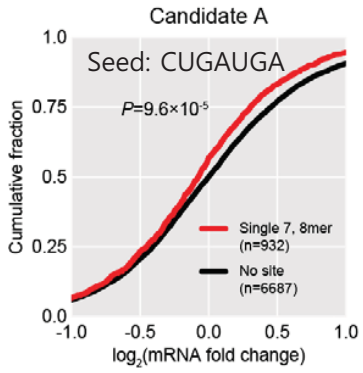


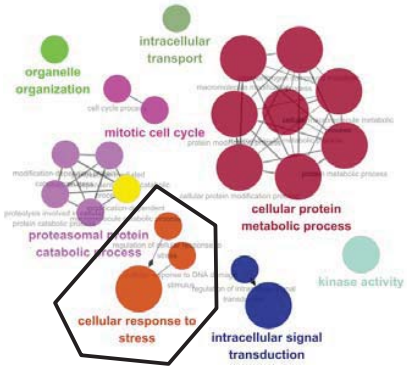
Two miRNA candidates detected by small RNA-seq

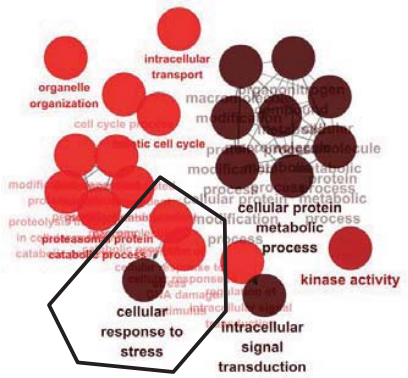
SARS-CoV-2 MicroRNAs

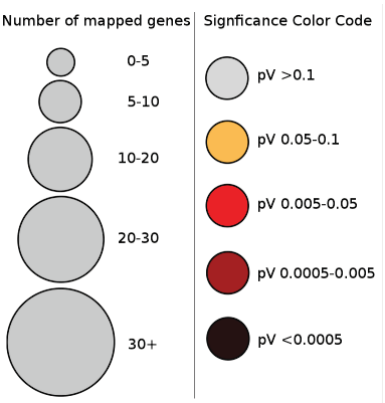
Color by groups

Color by significance









Target mRNAs with a single 7, 8mer site (n=932)
Associated GO terms (P < 0.05) were displayed.

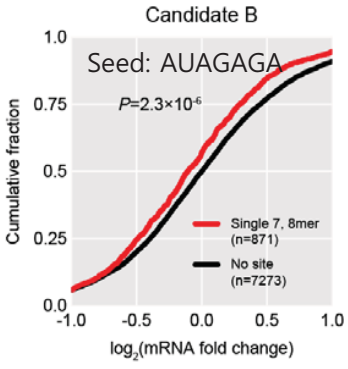
The most significant GO term:
Cellular response to stress (P=8.0x10^{-11})

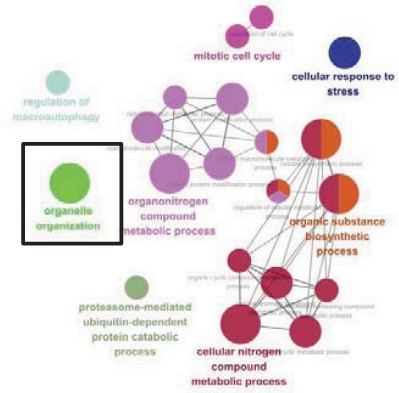
- 171 genes (7.7%) are associated

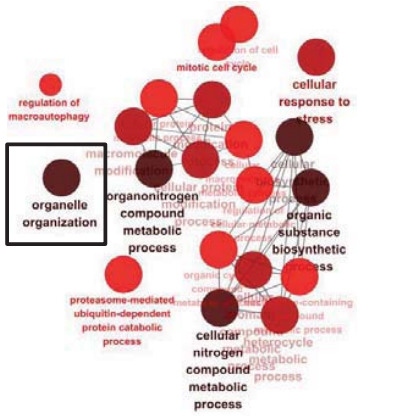
SARS-CoV-2 MicroRNAs

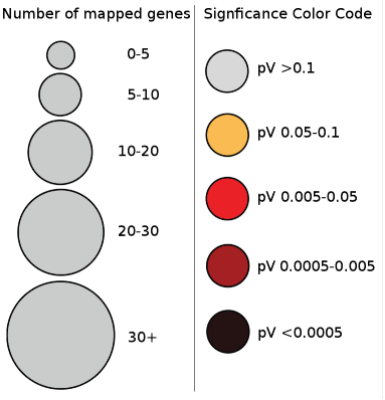
Color by groups

Color by significance







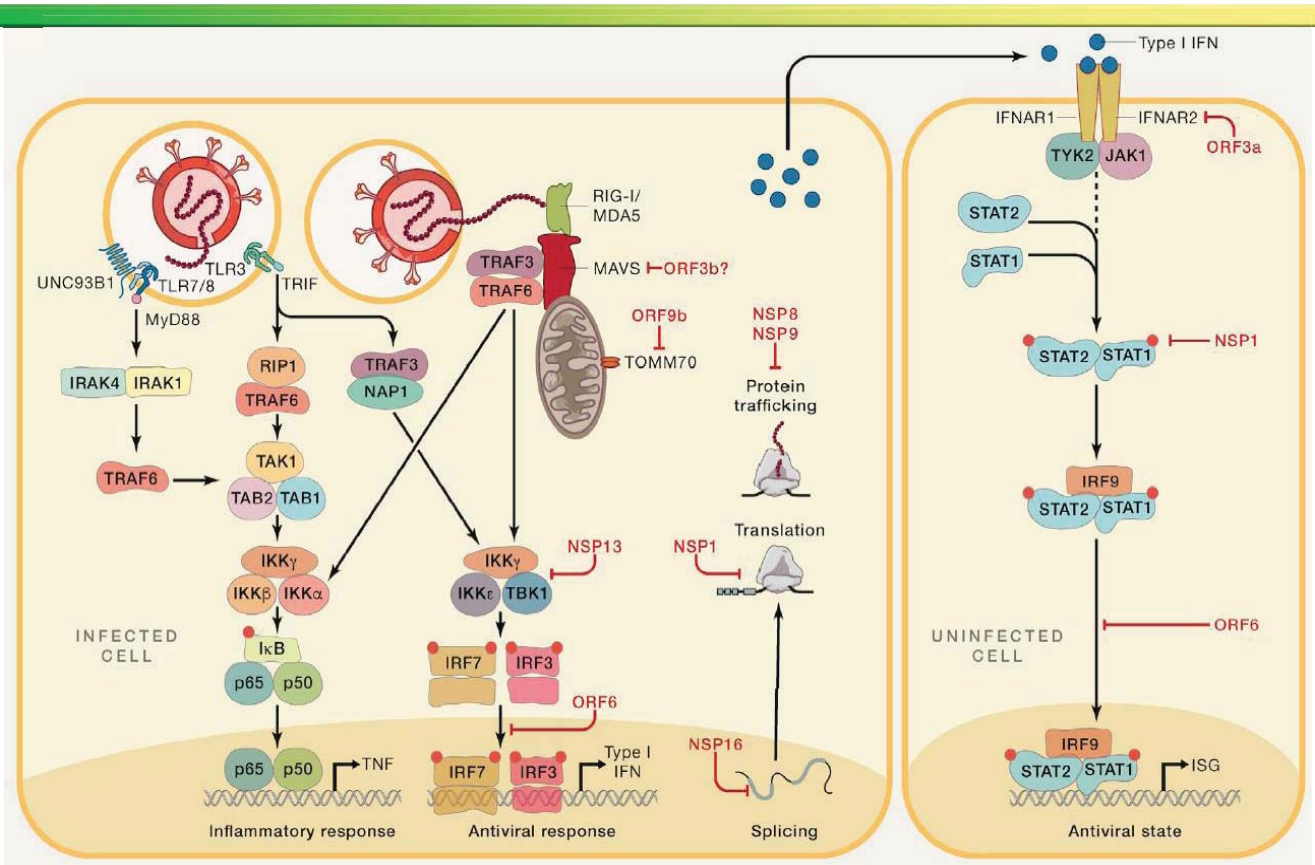


Target mRNAs with a single 7, 8mer site (n=871)
Associated GO terms (P < 0.05) were displayed.

The most significant GO term:
Organelle organization (P=4.9x10^{-8})

- 256 genes (6.1%) are associated

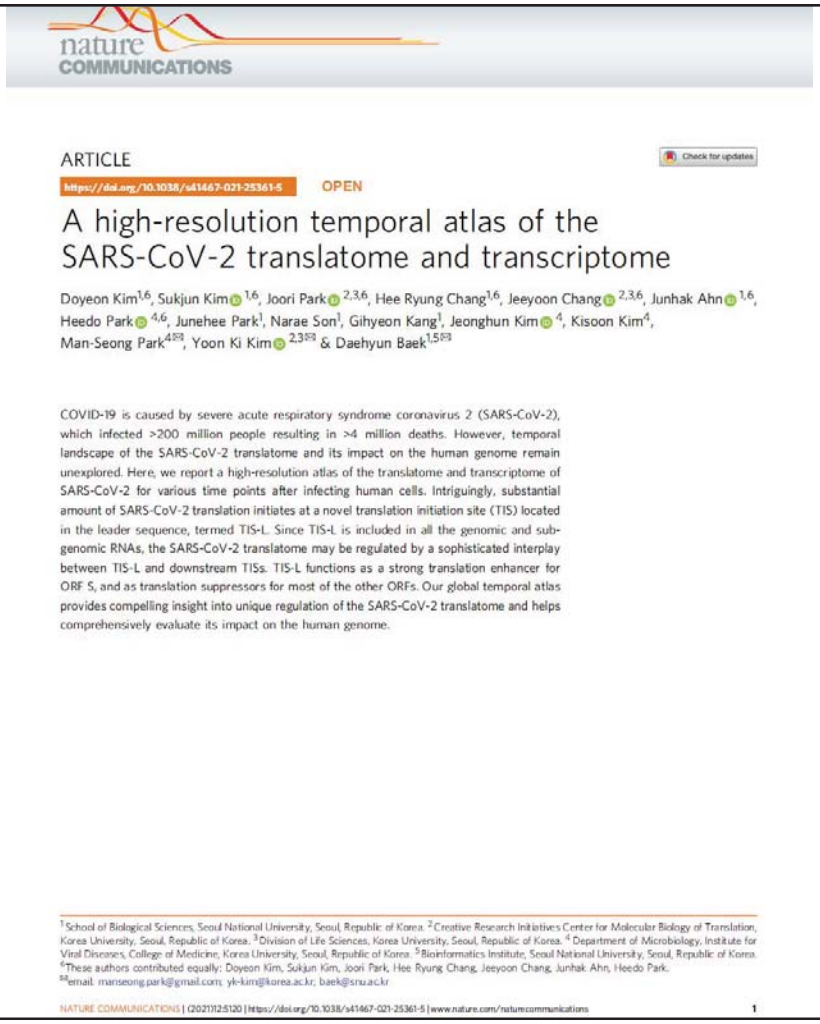
SARS-CoV-2 MicroRNAs May Help Evade Human Immune Response



(Schultze and Aschenbrenner, 2021)

Conclusions

- **We report the first high-resolution atlas of the translatome and transcriptome of SARS-CoV-2 for various time points after infecting human cells.**
- **Intriguingly, substantial amount of SARS-CoV-2 translation initiates at a novel translation initiation site (TIS) located in the leader sequence, that we termed TIS-L.**
- **Since TIS-L is included in all the genomic and subgenomic RNAs, the SARS-CoV-2 translatome may be regulated by a sophisticated interplay between TIS-L and downstream TISs.**
- **TIS-L functions as a strong translation enhancer for S protein, and as translation suppressors for most of the other proteins.**
- **Our global temporal atlas provides compelling insight into unique regulation of the SARS-CoV-2 translatome and helps comprehensively evaluate its impact on the human genome.**



nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-25361-5 OPEN

A high-resolution temporal atlas of the SARS-CoV-2 translatome and transcriptome

Doyeon Kim[1,6], Sukjun Kim[1,6], Joori Park[2,3,6], Hee Ryung Chang[1,6], Jeeyoon Chang[2,3,6], Junhak Ahn[1,6], Heedo Park[4,6], Junehee Park[1], Narae Son[1], Gihyeon Kang[1], Jeonghun Kim[4], Kisoon Kim[4], Man-Seong Park[4✉], Yoon Ki Kim[2,3✉] & Daehyun Baek[1,5✉]

COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which infected >200 million people resulting in >4 million deaths. However, temporal landscape of the SARS-CoV-2 translatome and its impact on the human genome remain unexplored. Here, we report a high-resolution atlas of the translatome and transcriptome of SARS-CoV-2 for various time points after infecting human cells. Intriguingly, substantial amount of SARS-CoV-2 translation initiates at a novel translation initiation site (TIS) located in the leader sequence, termed TIS-L. Since TIS-L is included in all the genomic and sub-genomic RNAs, the SARS-CoV-2 translatome may be regulated by a sophisticated interplay between TIS-L and downstream TISs. TIS-L functions as a strong translation enhancer for ORF S, and as translation suppressors for most of the other ORFs. Our global temporal atlas provides compelling insight into unique regulation of the SARS-CoV-2 translatome and helps comprehensively evaluate its impact on the human genome.

[1] School of Biological Sciences, Seoul National University, Seoul, Republic of Korea. [2] Creative Research Initiatives Center for Molecular Biology of Translation, Korea University, Seoul, Republic of Korea. [3] Division of Life Sciences, Korea University, Seoul, Republic of Korea. [4] Department of Microbiology, Institute for Viral Diseases, College of Medicine, Korea University, Seoul, Republic of Korea. [5] Bioinformatics Institute, Seoul National University, Seoul, Republic of Korea. [6] These authors contributed equally: Doyeon Kim, Sukjun Kim, Joori Park, Hee Ryung Chang, Jeeyoon Chang, Junhak Ahn, Heedo Park. ✉email: manseong.park@gmail.com; yk-kim@korea.ac.kr; baek@snu.ac.kr

Acknowledgements


Sukjun Kim


Hee Ryung Chang


Doyeon Kim


Chanseok Shin


Jinwu Nam


Yoon Ki Kim


Sael Lee

Research Funding


Ministry of Science and ICT


Ministry of Health and Welfare

Acknowledgements

Yoon Ki Kim

Man-Seong Park

Kisoon Kim

Kim lab: Joori Park and Jeeyoon Chang

Park lab: Heedo Park and Jeonghun Kim

Baek lab: Doyeon Kim, Sukjun Kim, Hee Ryung Chang, Junhak Ahn, Junehee Park, Narae Son, and Gihyeon Kang

Research Funding:

Ministry of Science and ICT

Ministry of Health and Welfare