



# Deep learning enhancing guide RNA design for CRISPR/Cas12a-based diagnostics

Baicheng Huang<sup>1#</sup>, Ling Guo<sup>1#</sup>, Hang Yin<sup>1</sup>, Yue Wu<sup>1</sup>, Zihan Zeng<sup>1</sup>, Sujie Xu<sup>1</sup>,  
Yufeng Lou<sup>2</sup>, Zhimin Ai<sup>1</sup>, Weiqiang Zhang<sup>1</sup>, Xingchi Kan<sup>1</sup>, Qian Yu<sup>1</sup>, Shimin Du<sup>1</sup>, Chao Li<sup>3</sup>,  
Lina Wu<sup>4</sup>, Xingxu Huang<sup>1</sup>, Shengqi Wang<sup>5\*</sup>, Xinjie Wang<sup>6\*</sup>

<sup>1</sup>Zhejiang Laboratory, Hangzhou, China

<sup>2</sup>Department of Laboratory Medicine, the First Affiliated Hospital, Zhejiang University School of Medicine; Key Laboratory of Clinical In Vitro Diagnostic Techniques of Zhejiang Province; Institute of Laboratory Medicine, Zhejiang University, Hangzhou, China

<sup>3</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK.;  
School of Medicine, School of Science and Engineering, University of Dundee, Dundee, UK

<sup>4</sup>School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing, China

<sup>5</sup>Bioinformatics Center of AMMS, Beijing, China

<sup>6</sup>Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China



Baicheng Huang, Ling Guo, Hang Yin, Yue Wu, Zihan Zeng, Sujie Xu, Yufeng Lou, Zhimin Ai, Weiqiang Zhang, Xingchi Kan, Qian Yu, Shimin Du, Chao Li, Lina Wu, Xingxu Huang, Shengqi Wang, Xinjie Wang. 2024. Deep learning enhancing guide RNA design for CRISPR/Cas12a-based diagnostics. *iMeta* 3: e214. <https://doi.org/10.1002/imt2.214>



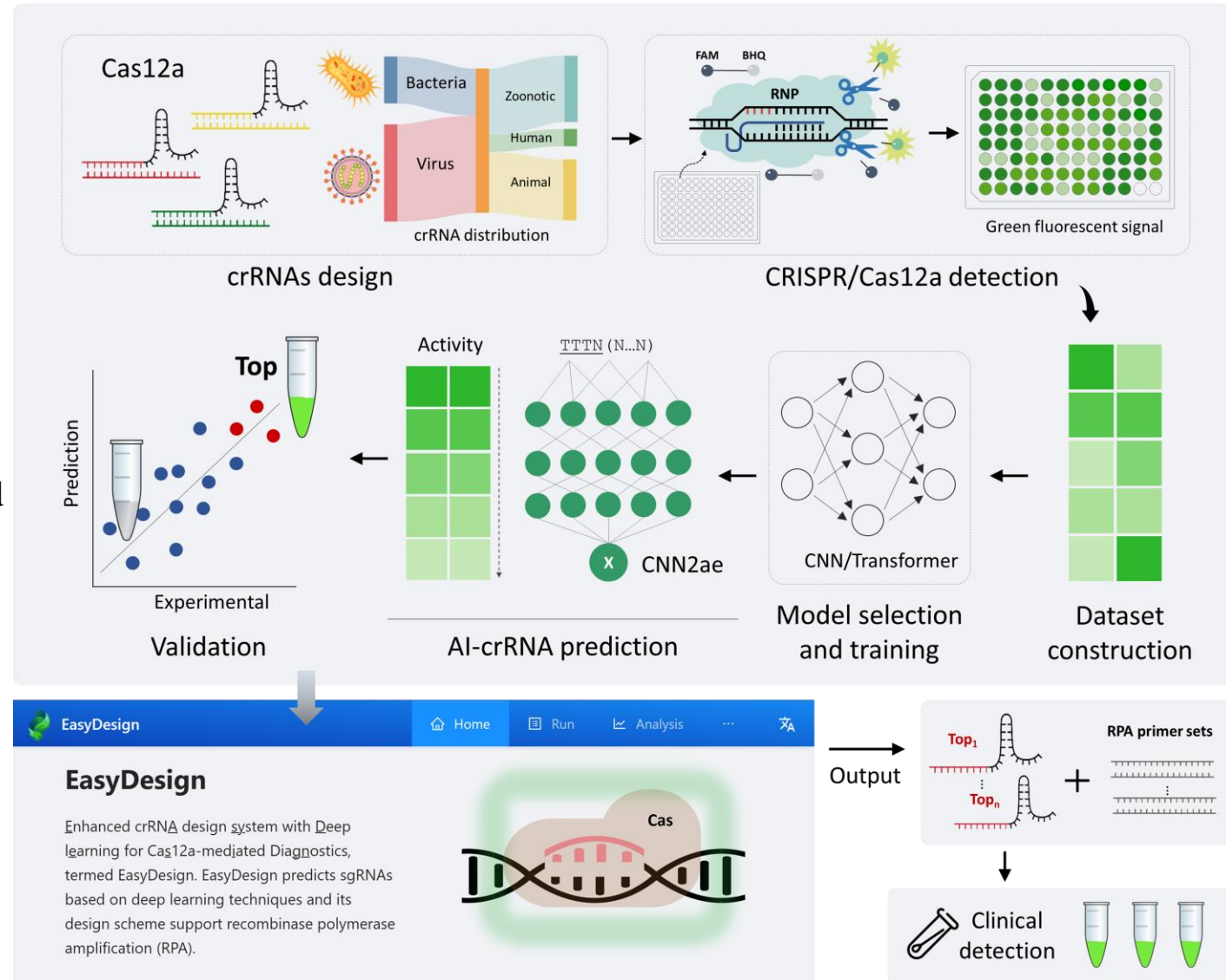
# Introduction

- Rapid and accurate diagnostic tests are fundamental for improving patient outcomes and combating infectious diseases, such as the outbreaks of COVID-19 and monkeypox.
- The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas12a-based detection system has emerged as a promising solution for on-site nucleic acid testing.
- The effective design of CRISPR RNA (crRNA) for Cas12a-based detection remains challenging and time-consuming.
- Deep learning has emerged as a powerful tool for designing crRNAs for CRISPR-based genome editing systems, such as convolutional neural networks (CNN) and transformer.



# Overview of the EasyDesign development

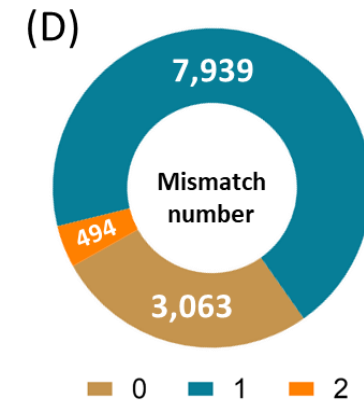
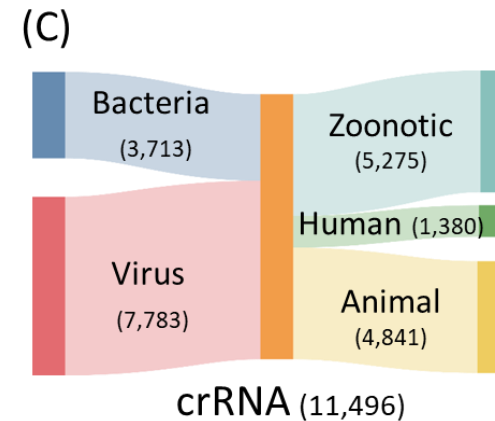
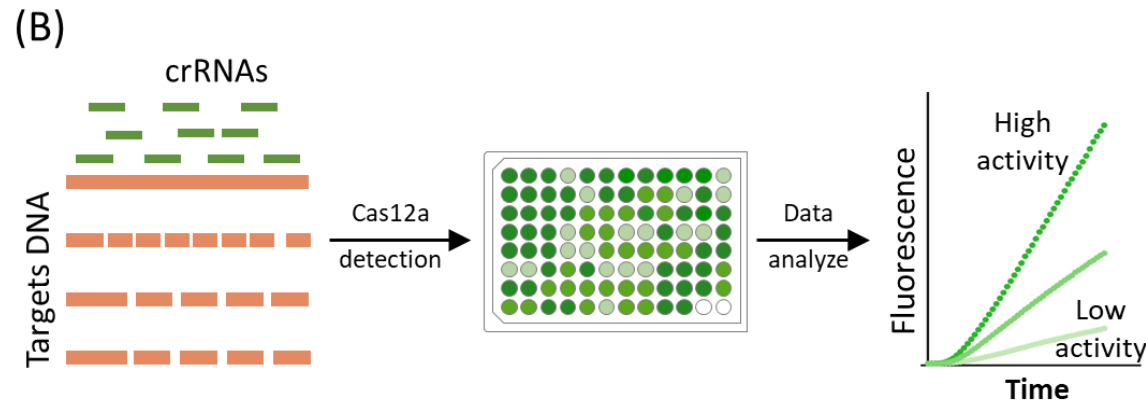
- An optimized CNN model, CNN12ae, developed by training on a comprehensive dataset of 11,496 experimentally validated Cas12a-based detection cases.
- The performance of CNN12ae in crRNA design for four pathogens (Monkeypox Virus, Enterovirus 71, Coxsackievirus A16, and *Listeria monocytogenes*) demonstrated superior prediction performance compared to the traditional experimental screening.
- A web server (<https://crispr.zhejianglab.com/>) that integrates EasyDesign with recombinase polymerase amplification primer design, which successfully designed optimal Cas12a crRNAs for six human papillomavirus subtypes.





# High-quality datasets of crRNA mediate Cas12a detection

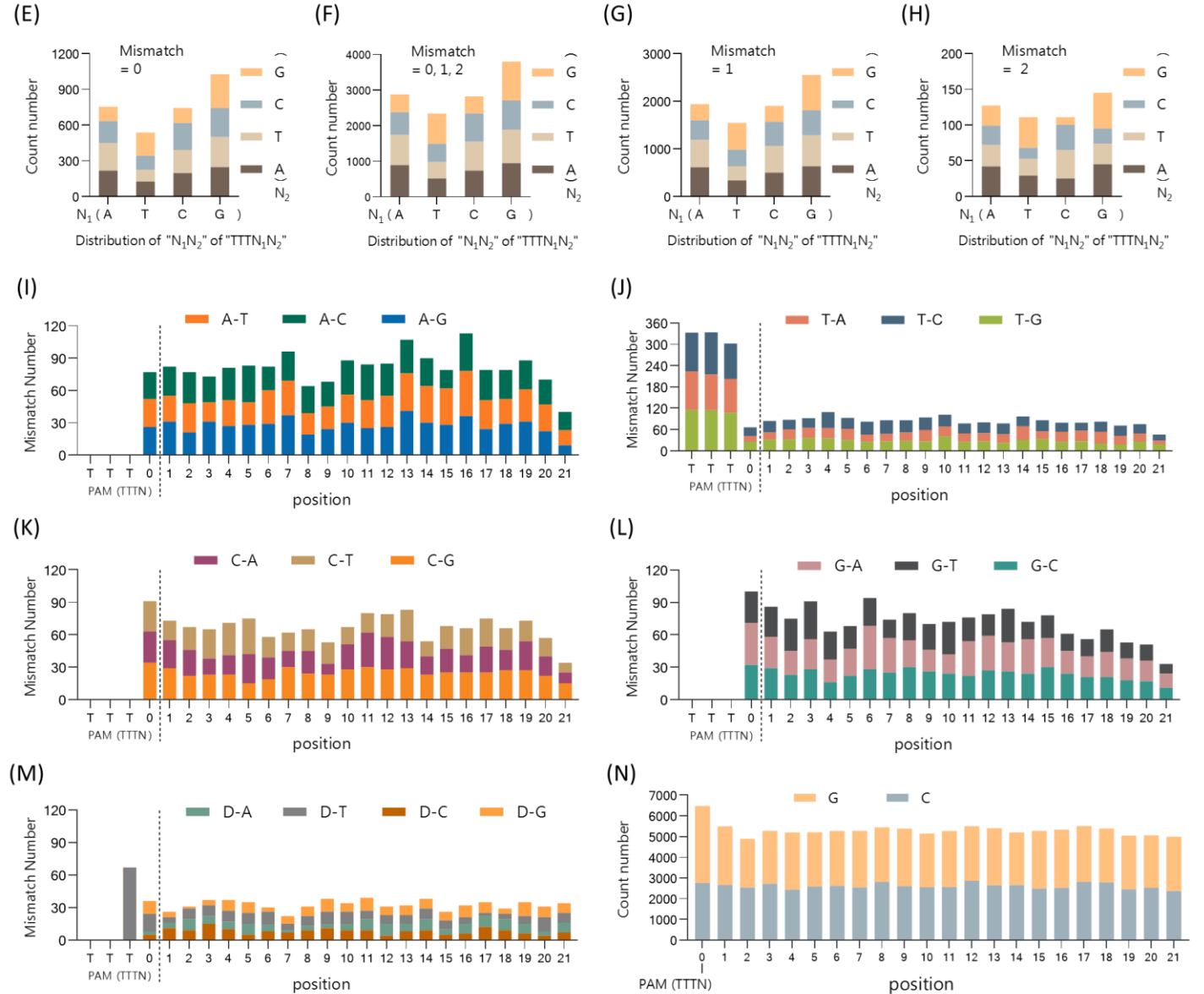
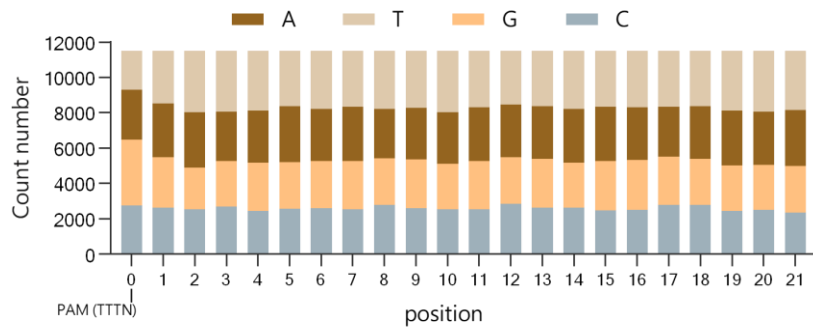
- Using 11,496 fluorescence readouts as the training dataset, include 7,783 pairs from viral sources and 3,713 pairs from bacterial sources.
- 1,533 crRNAs targeting 34 natural bacterial and viral pathogens (in 198 DNA templates).
- Of these, 5,275 pairs are zoonotic pathogens, 1,380 pairs are human pathogens, and 4,841 pairs are animal pathogens.
- 3,063 pairs with no mismatches; 7,939 pairs with one-base mismatches; 494 pairs with two-base mismatches.





# High-quality datasets of crRNA mediate Cas12a detection

- A uniform distribution of base types.
- No significant bias observed in mutation types in all guide-to-target pairs containing mismatches.
- The distribution of base types at the 21st positions downstream of the PAM remained relatively consistent.

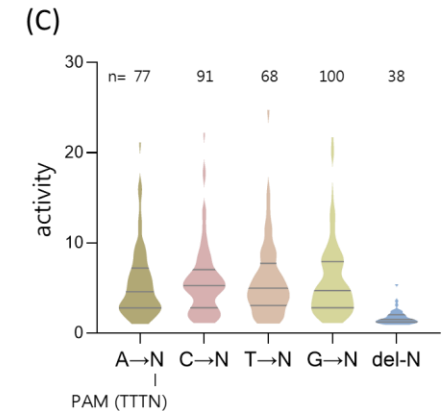
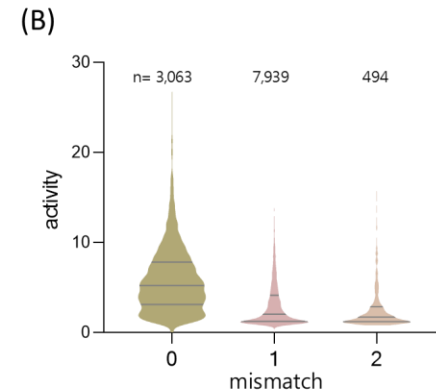
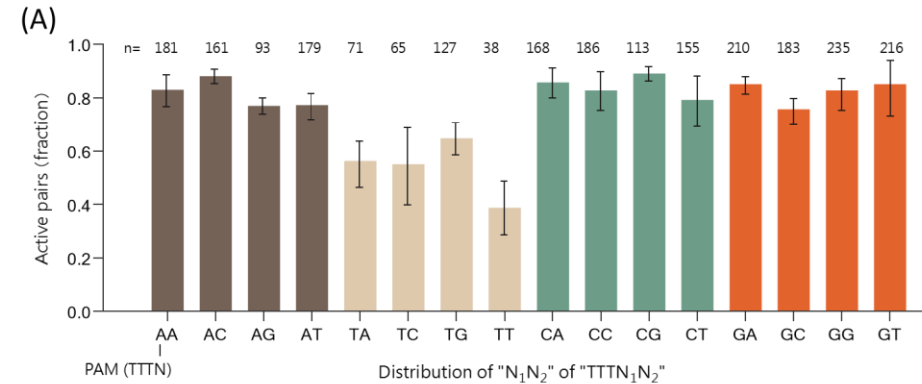
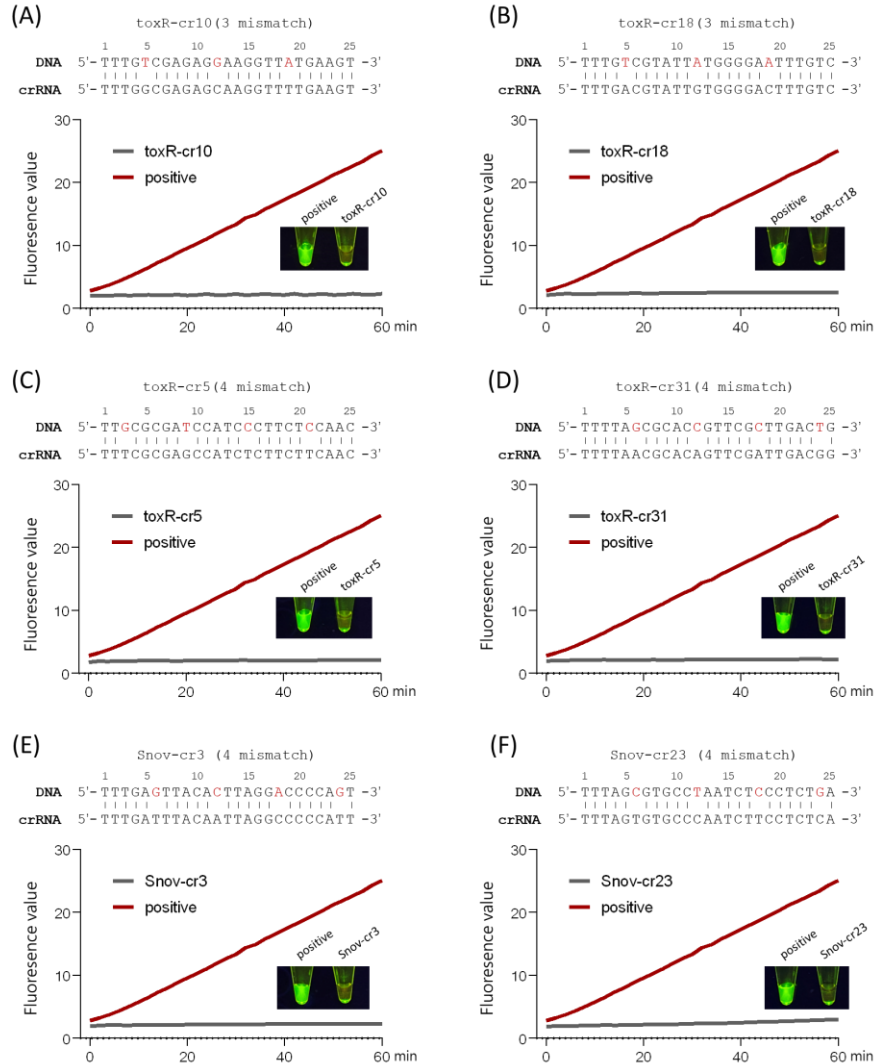




# High-quality datasets of crRNA mediate Cas12a detection

■ A high degree of mismatch leads to reduced activity.

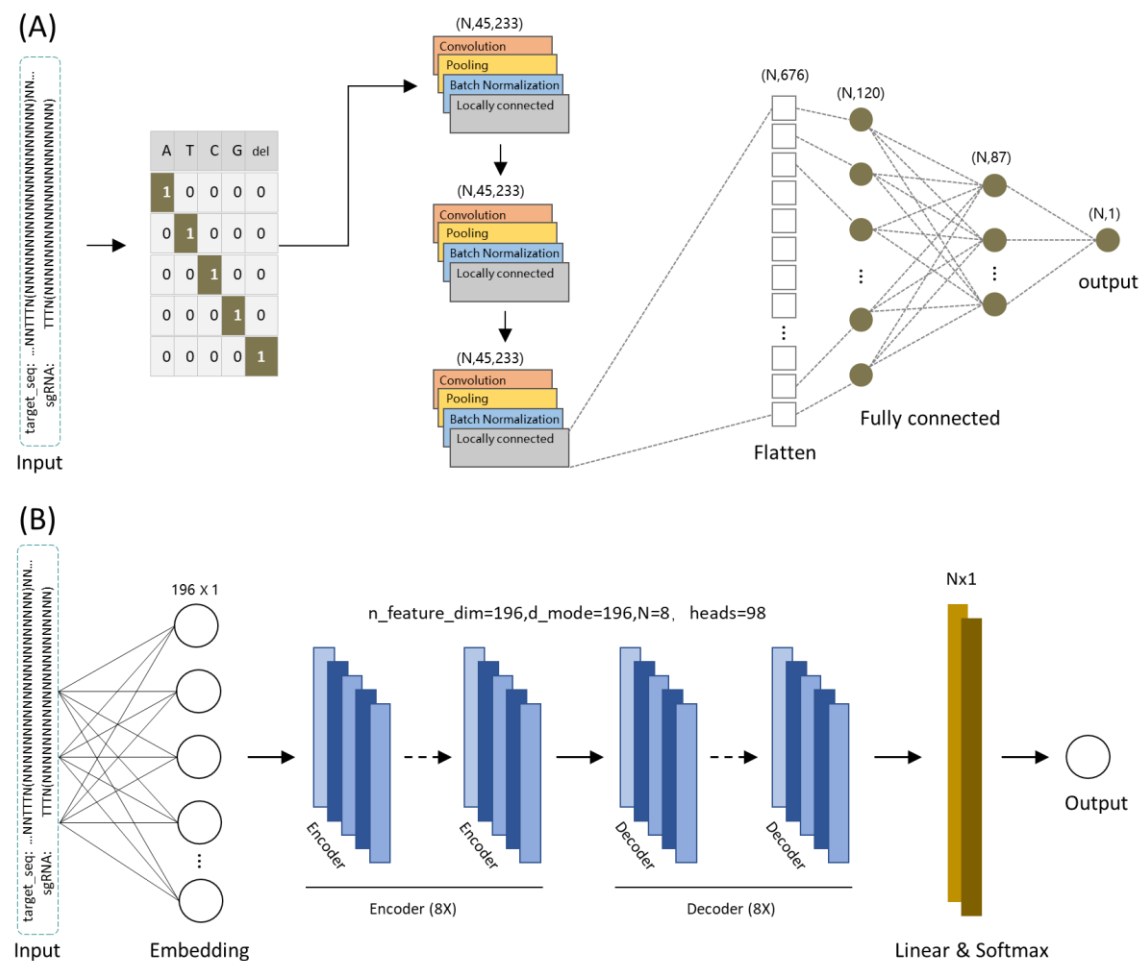
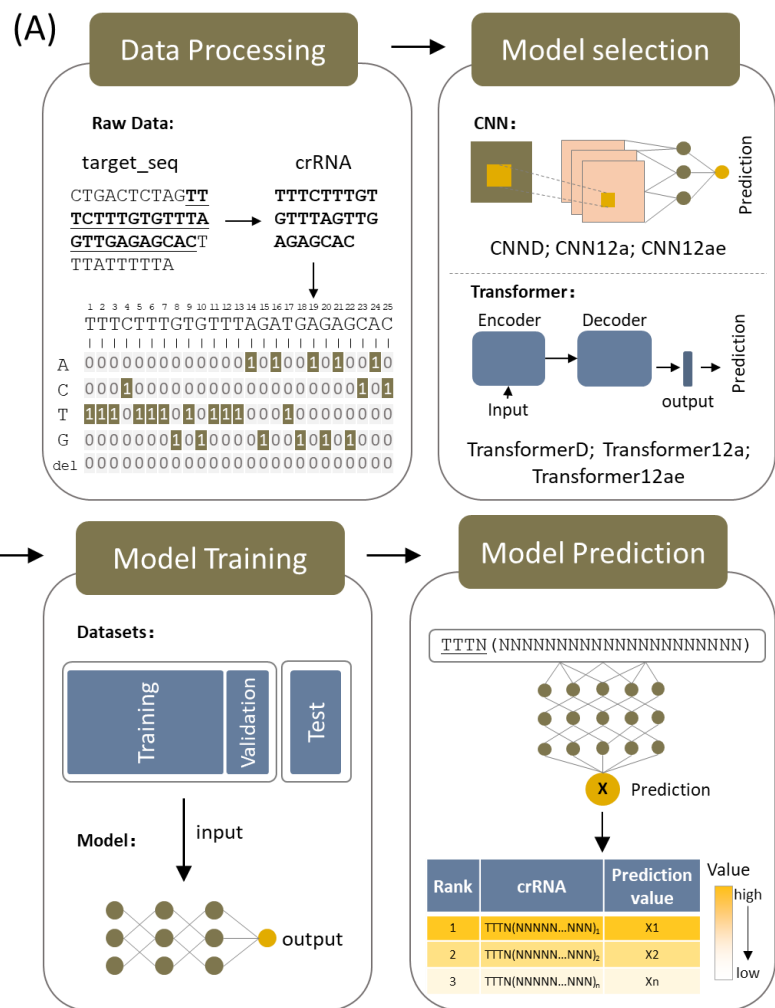
■ ‘TT’ exhibits lower activity compared to TA, TC, and TG; pairs without mismatches exhibit higher activity than those with mismatches; no significant effect on activity from the mutation of ‘N’ in the PAM sequence (TTTN).





# Development of a deep learning model for crRNA design

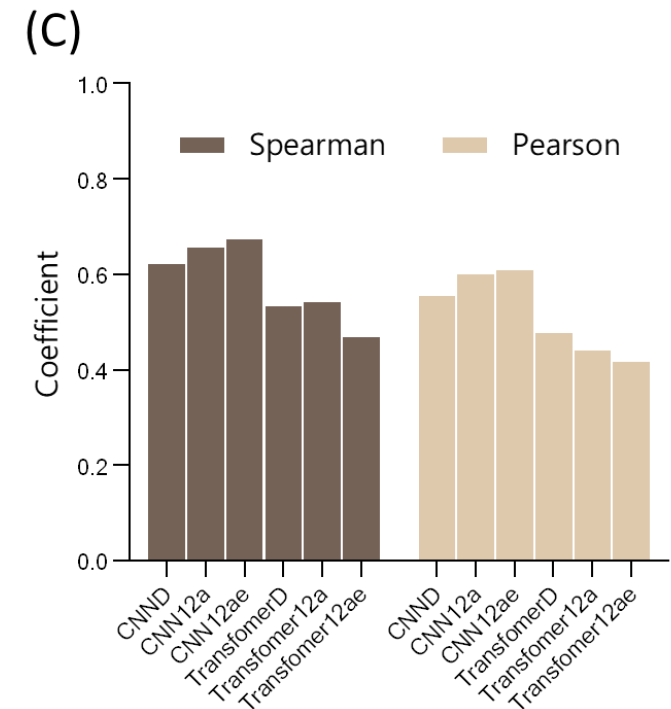
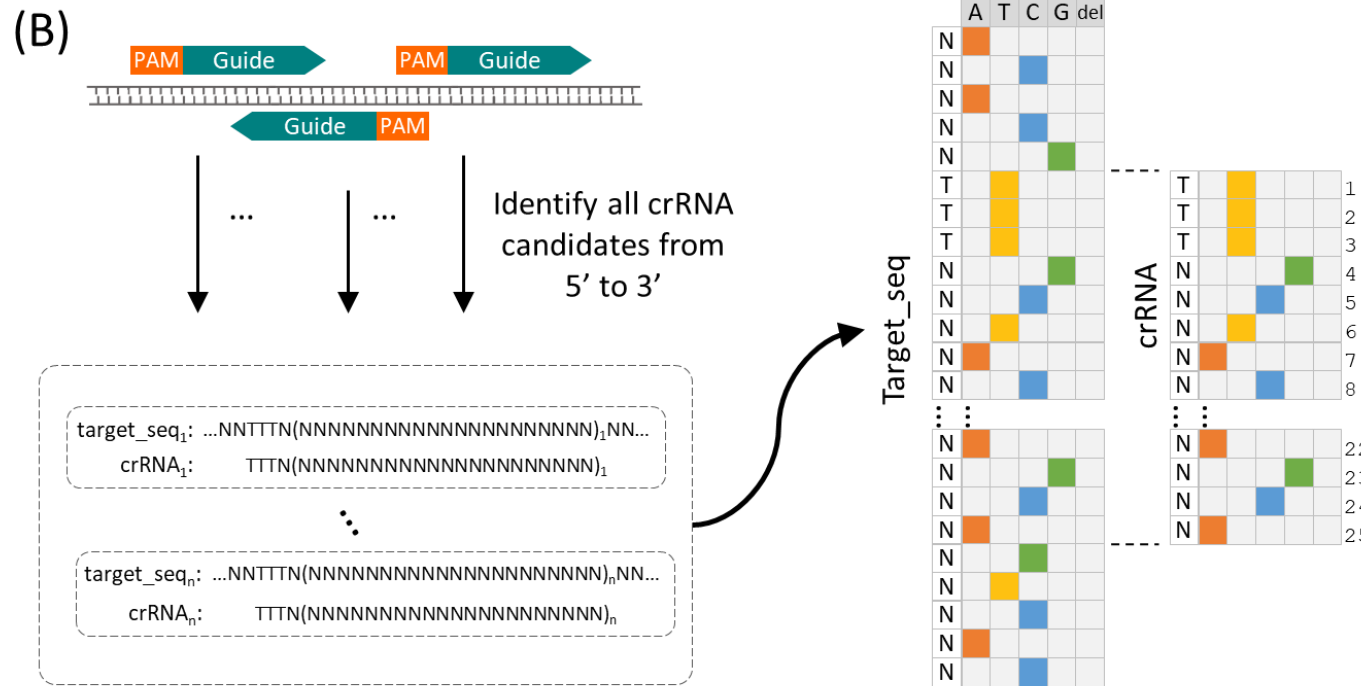
- A four-step process for training the deep learning models.
- CNN12a and Transformer12a, by adapting CNND and TransformerD models to the Cas12a diagnostic scenarios.





# Development of a deep learning model for crRNA design

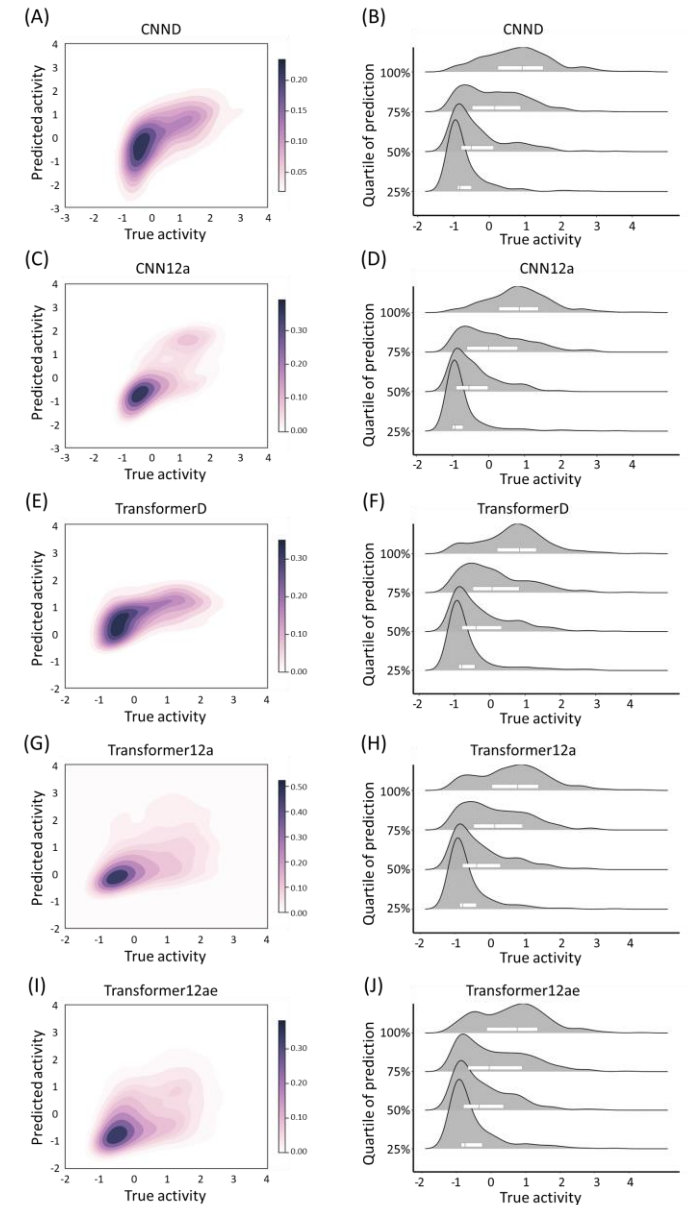
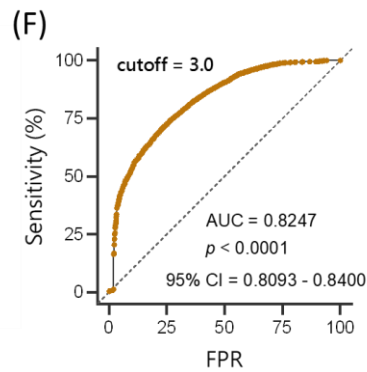
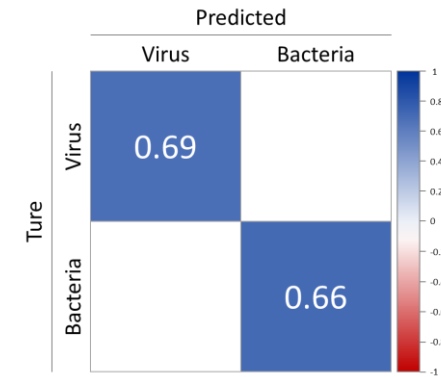
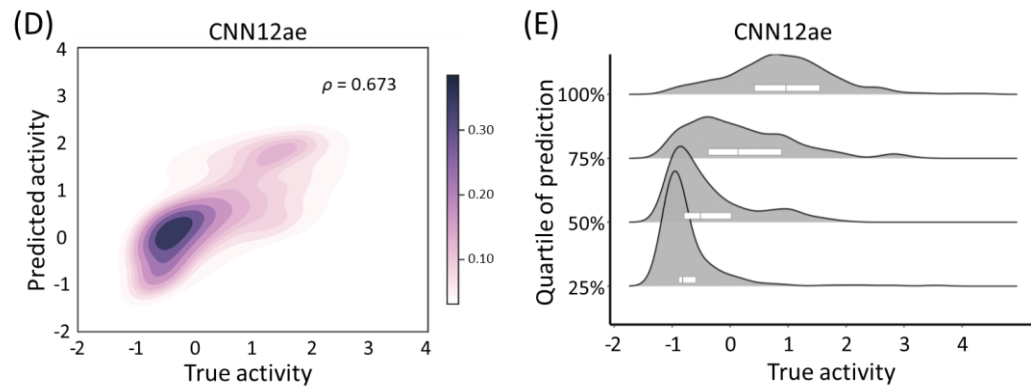
- The crRNA was extended by an additional 10 nucleotides at both the 5' and 3' terminals of the target, resulting in a 45-nucleotide target sequence.
- Upon evaluation using the test set, CNN12ae achieves improved performance over the CNN12a model (Spearman  $\rho = 0.673$ ), and also outperforms Transformer12ae ( $\rho = 0.467$ ).





# Development of a deep learning model for crRNA design

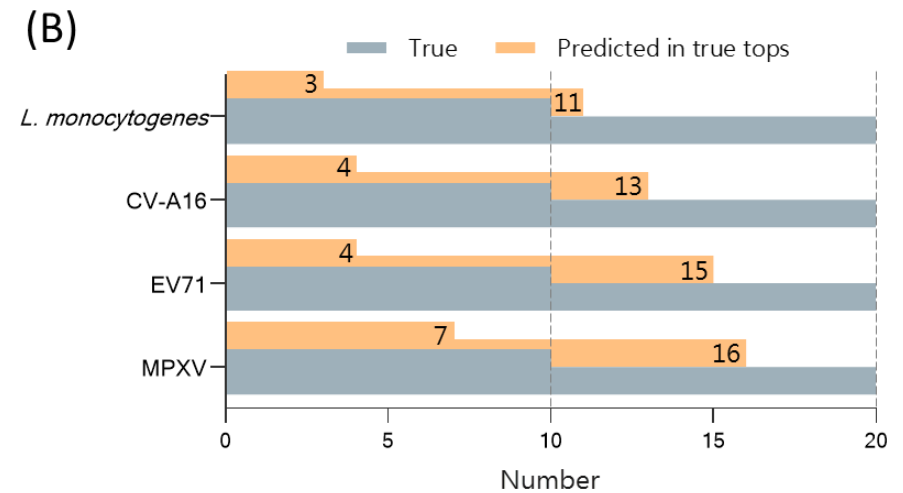
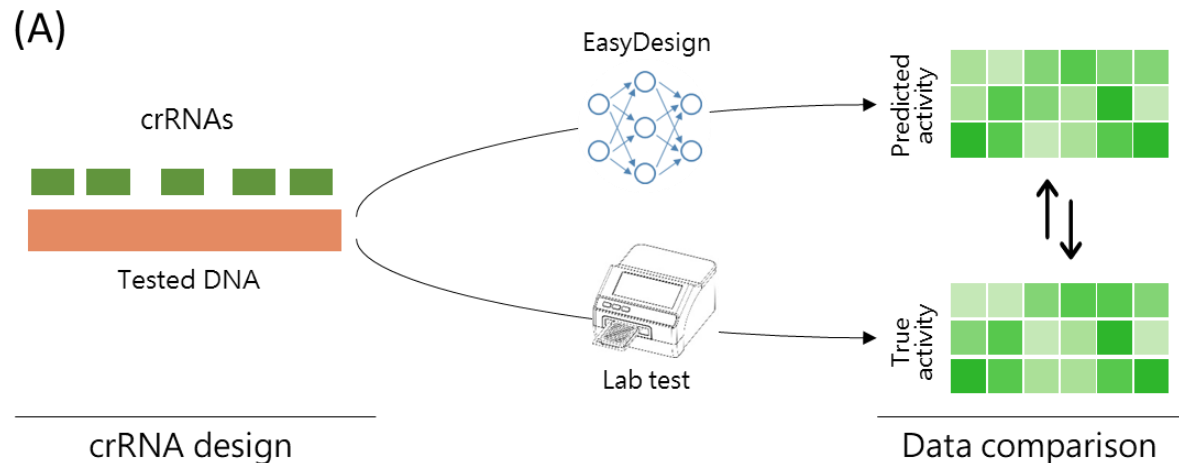
- CNN12ae shows distinct regions characterized by different levels of activity compared to other models.
- The classification performance of the CNN12ae model, which yielded an AUC value of 0.8247 with a p-value  $< 0.0001$ , indicating strong classification capabilities.
- The predictive performance of CNN12ae was consistent for viral (0.66) and bacterial (0.69) pairs.





# EasyDesign on pathogenic nucleic acid detection

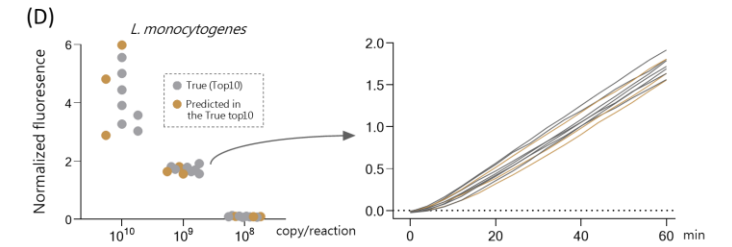
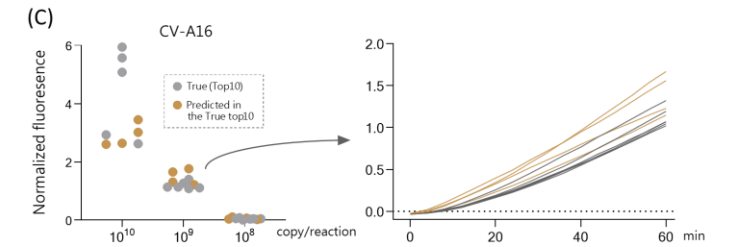
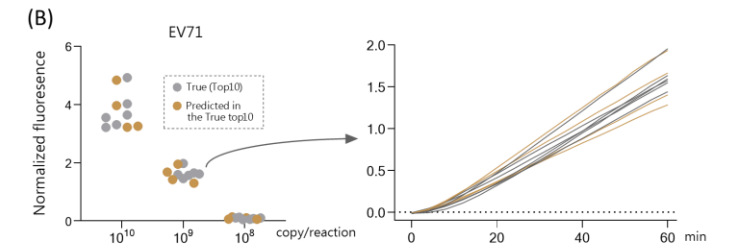
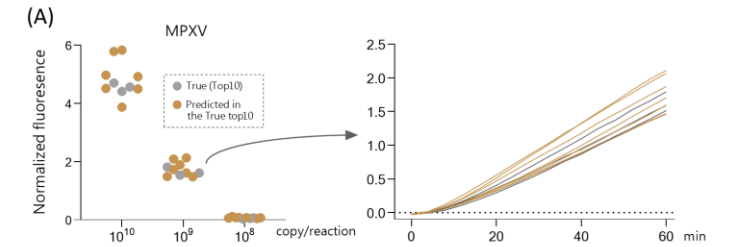
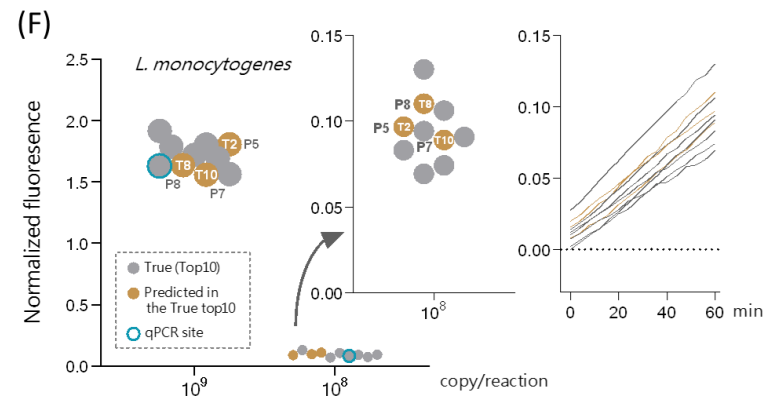
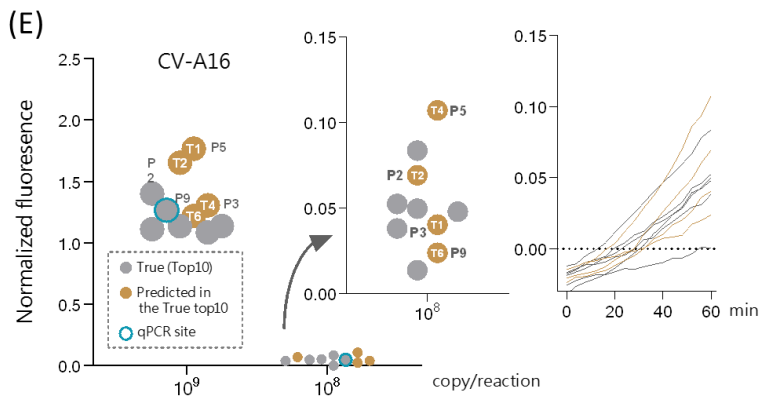
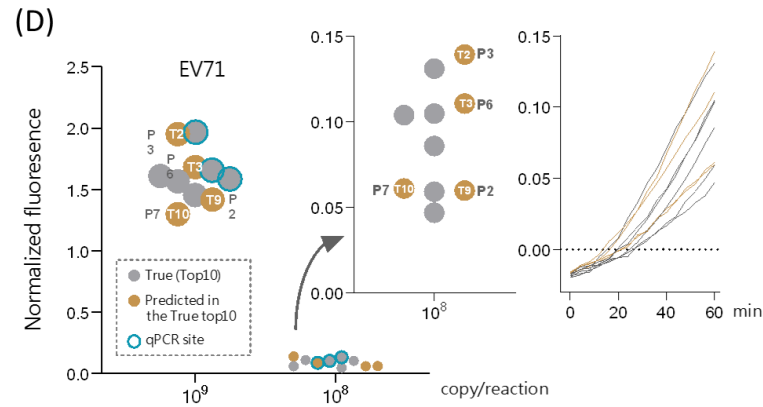
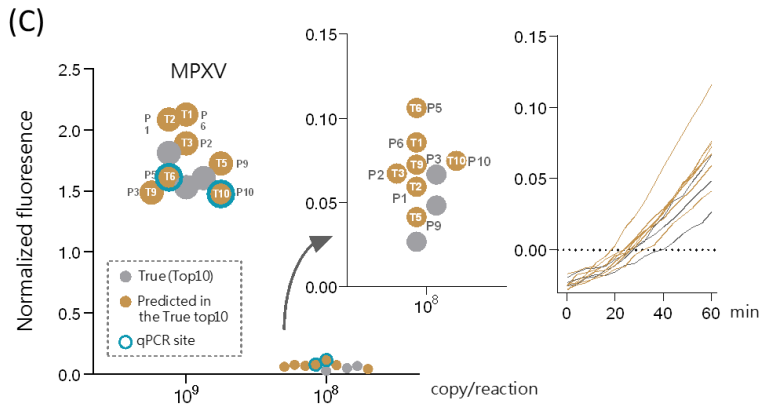
- The crRNA activity was experimental performed to validate the applicability of the developed model, and EasyDesign predicted crRNA ranking for four pathogenic nucleic acid sequences, 194 crRNA in total.
- Among four pathogens, the consistency between the predicted and experimentally verified crRNAs demonstrated in 7, 4, 4, and 3 cases from each pathogen for the top 10 crRNAs; whereas 16, 15, 13, and 11 cases for the top 20 crRNAs.





# EasyDesign on pathogenic nucleic acid detection

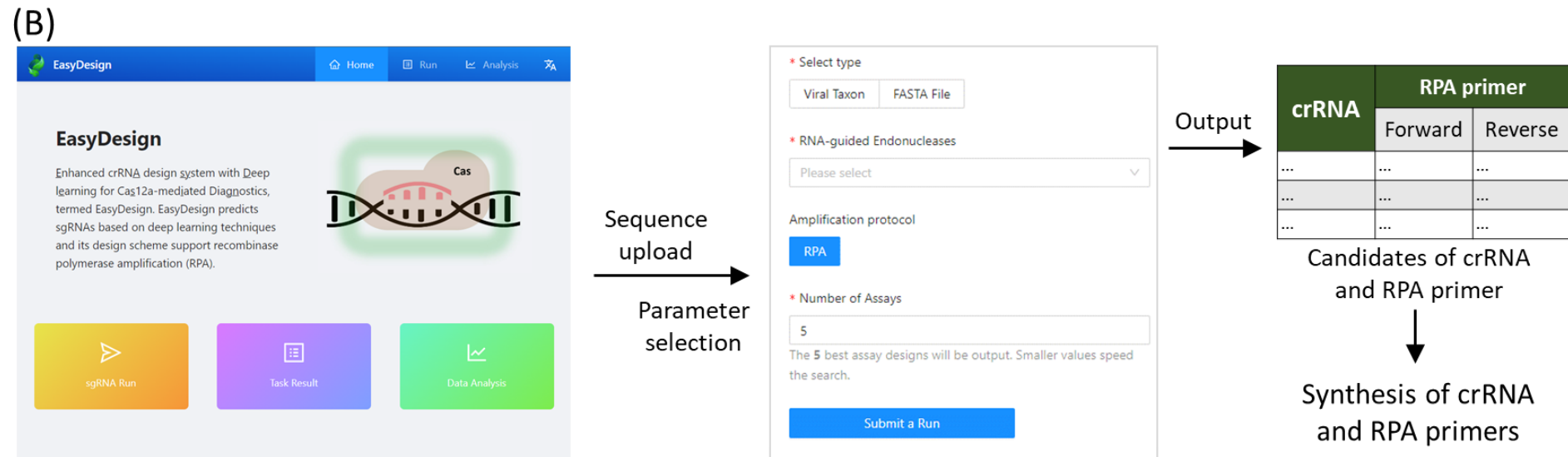
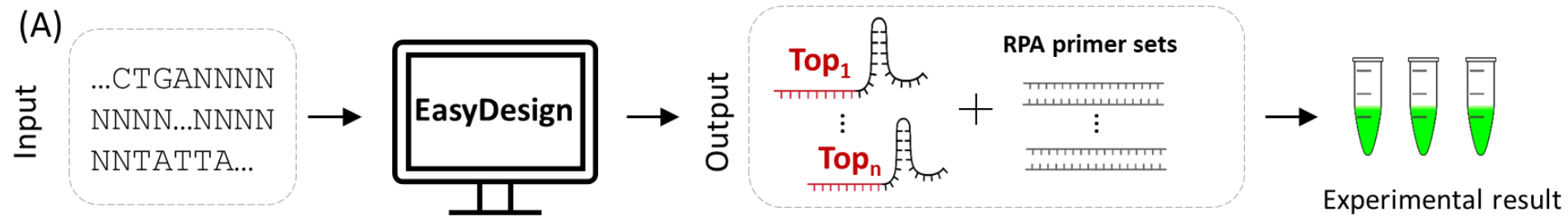
- The predicted crRNAs generated by CNN12ae model exhibits equal or higher reactivity than the reported qPCR sites, particularly at a target DNA concentration of  $10^9$  copies/reaction.





# Web tool for End-to-End Cas12a diagnostic design

- EasyDesign: accessed at <https://crispr.zhejianglab.com/>.
- This platform provides a comprehensive Cas12a-based diagnostic design experience that seamlessly integrates RPA primer design, facilitating the development of RPA-CRISPR assays with recommended crRNAs and RPA primers.

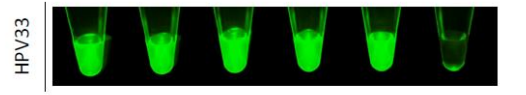
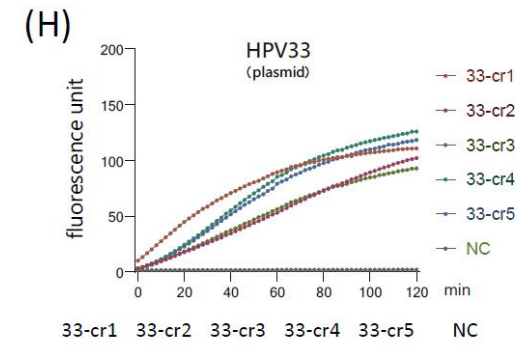
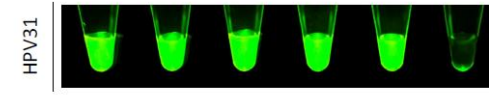
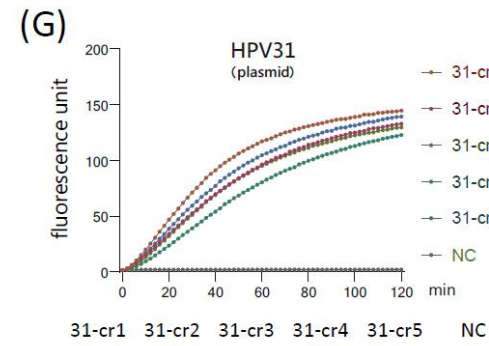
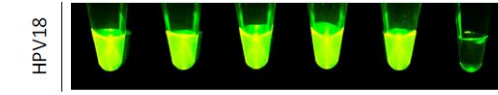
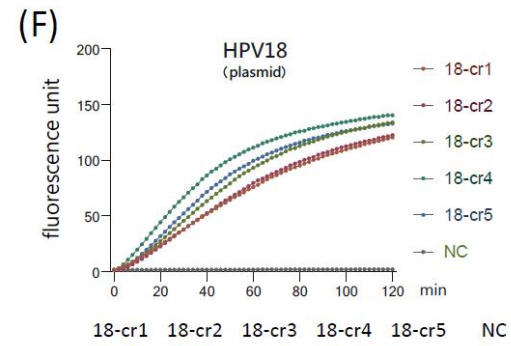
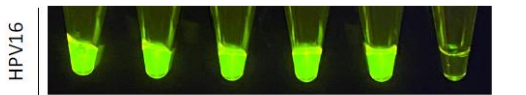
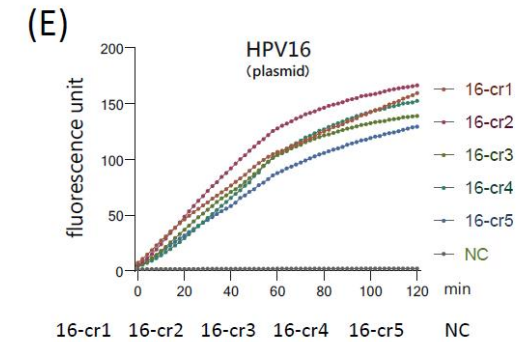
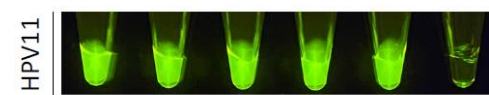
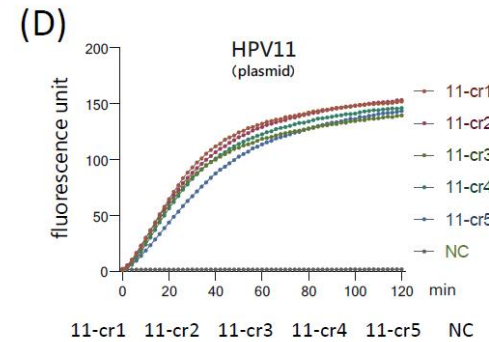
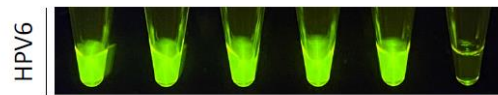
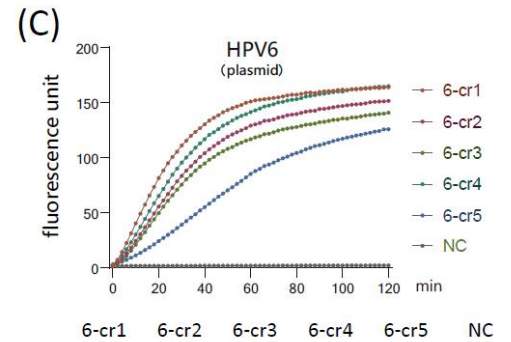




# EasyDesign facilitates HPV clinical sample diagnostic

- Using the input template DNA sequence and the combination of candidate crRNAs and RPA primers provided by online server <https://crispr.zhejianglab.com>, we achieved a robust fluorescent detection signal for synthetic DNA templates of HPV.

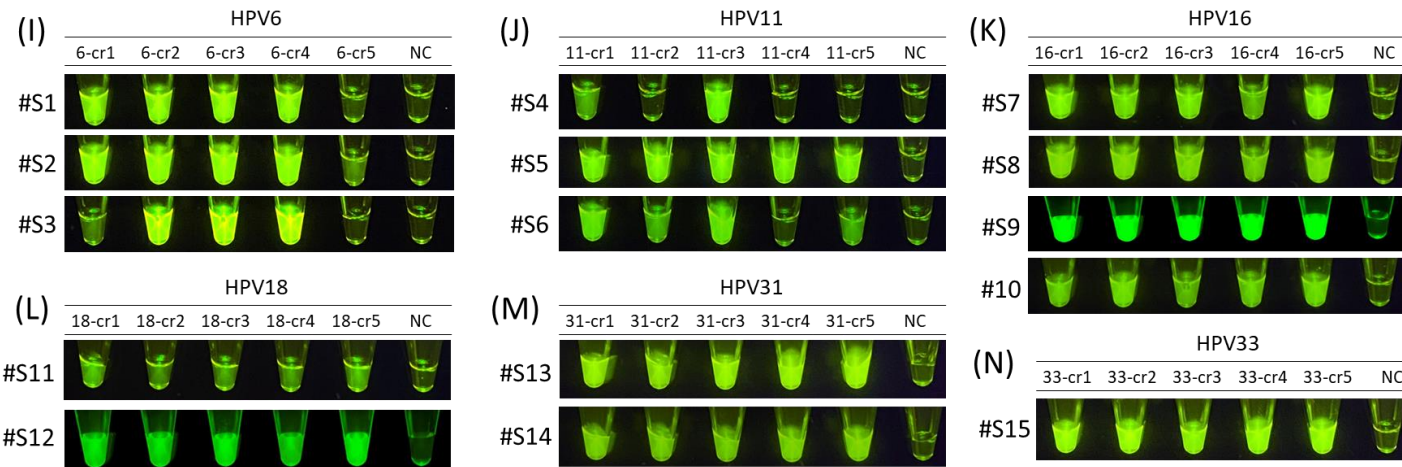
Fluorescence detection kinetic curves representing six human papillomavirus (HPV) subtypes (HPV6, HPV11, HPV16, HPV18, HPV31, and HPV33) using synthetic DNA templates. The optimal amplification primers and crRNAs were generated utilizing the EasyDesign web-based tool. Five optimal crRNAs and their corresponding primer pairs were generated for each template.



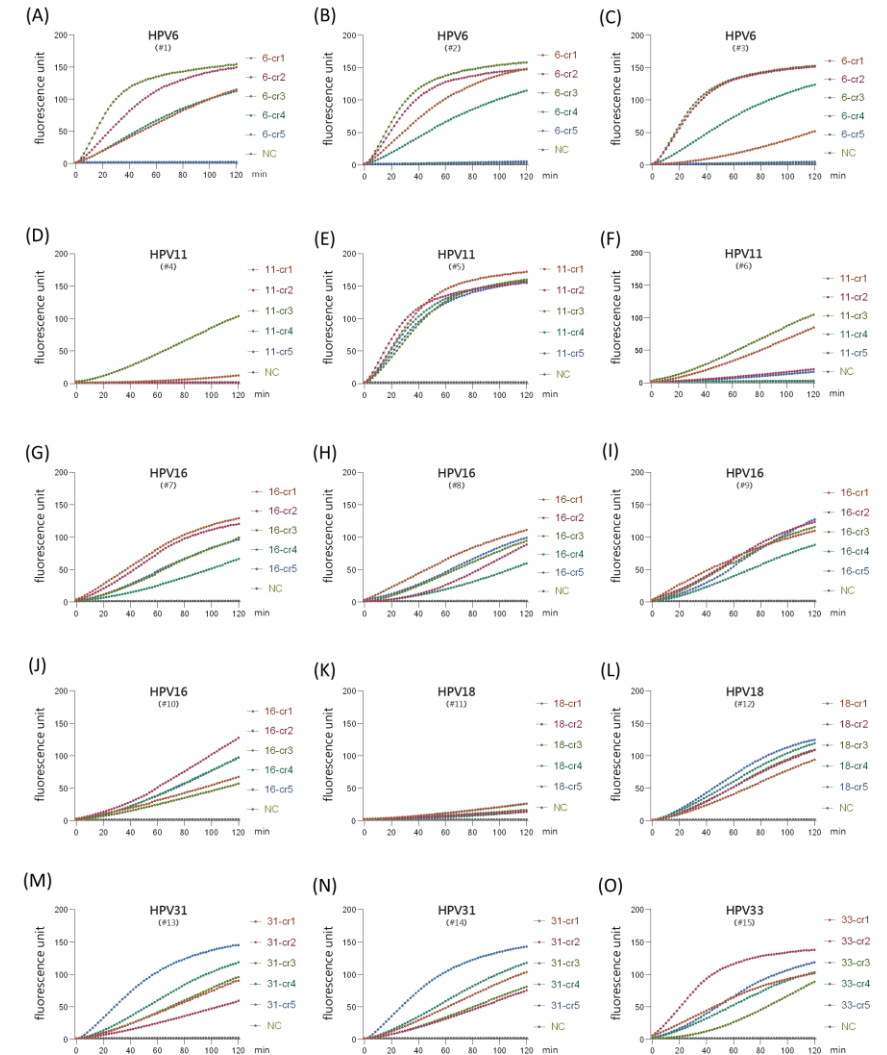


# EasyDesign facilitates HPV clinical sample diagnostic

- In clinical sample detection, all five candidate crRNAs of different HPV subtypes show significant fluorescence signals, demonstrating the efficacy of EasyDesign.



The positive clinical samples were identified as follows: #S1 to #S3 for HPV6, #S4 to #S6 for HPV11, #S7 to #S10 for HPV16, #S11 and #S12 for HPV18, #S13 and #S14 for HPV31, and #S15 for HPV33. NC, negative control.





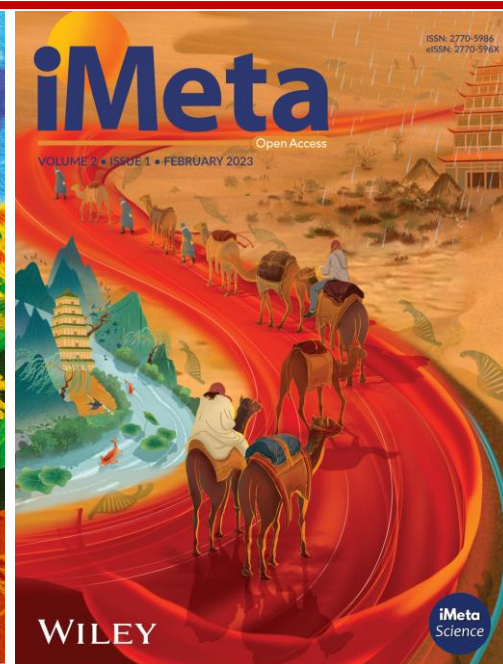
# Summary

- ❑ In this study, we developed an optimized CNN model, CNN12ae, for Cas12a-based detection design.
- ❑ The model performance of CNN12ae in crRNA design for four pathogens demonstrated superior prediction performance compared to the traditional experiment screening.
- ❑ A web server that integrates EasyDesign with recombinase polymerase amplification primer design was developed, which successfully designed optimal Cas12a crRNAs for six human papillomavirus subtypes.
- ❑ Website: <https://crispr.zhejianglab.com/>

Baicheng Huang, Ling Guo, Hang Yin, Yue Wu, Zihan Zeng, Sujie Xu, Yufeng Lou, Zhimin Ai, Weiqiang Zhang, Xingchi Kan, Qian Yu, Shimin Du, Chao Li, Lina Wu, Xingxu Huang, Shengqi Wang, Xinjie Wang. 2024. Deep learning enhancing guide RNA design for CRISPR/Cas12a-based diagnostics. *iMeta* 3: e214. <https://doi.org/10.1002/imt2.214>



**iMeta:** Integrated meta-omics to change the understanding of the biology and environment

**WILEY**



“**iMeta**” (IF **23.7**) is a Wiley partner journal launched by iMeta Science Society of scientists in bioinformatics and metagenomics in 2022. It aims to publish high-quality papers targeting a broad and diverse audiences. Its scope is similar to that of Nature Biotechnology, Nature Microbiology, Cell Host & Microbe. Its unique features include video submission, bilingual publishing, and social media dissemination with 500,000 followers. It has been published 190+ papers and been cited for 3400+ times since 2022, and indexed by [ESCI](#), [PubMed](#), [Google](#), and [Scopus](#).

“**iMetaomics**” is sister journal of “iMeta” launched in 2024, with target IF>10, and its scope is similar to Microbiome, ISME J, Nucleic Acids Research, Briefings in Bioinformatics, and Bioinformatics. All submissions are welcomed!

 Society: <http://www.imeta.science>  
Publisher: <https://wileyonlinelibrary.com/journal/imeta>  
 Submission: <https://wiley.atyponrex.com/journal/IMT2>  
<https://wiley.atyponrex.com/journal/IMO2>

 [office@imeta.science](mailto:office@imeta.science)

 [Promotion Video](#)

 [iMetaScience](#)

 [iMetaScience](#)