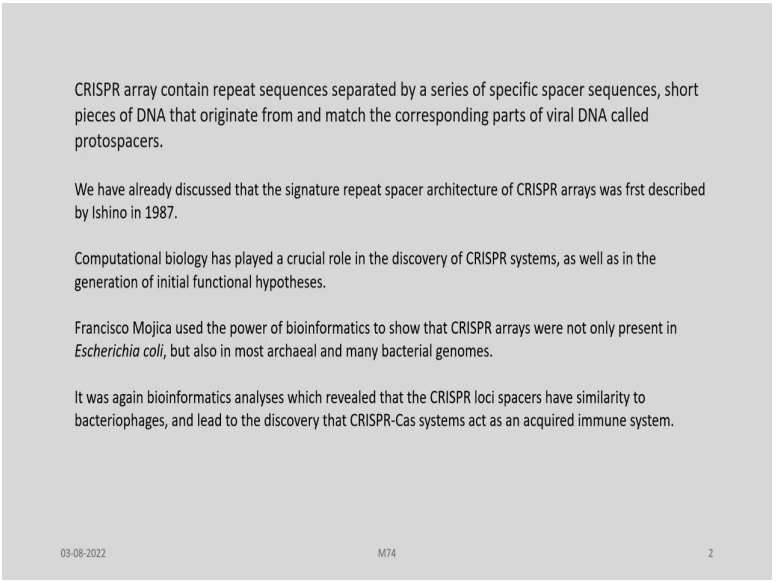
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 07
Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology
Lecture - 26
Computational Resources for CRISPR/Cas - Part A

Welcome to my course on Genome Editing and Engineering. We are discussing about CRISPR Cas 9 Technology. In today's lecture, we are going to study about the various computational resources for CRISPR Cas technology platform.

(Refer Slide Time: 00:49)



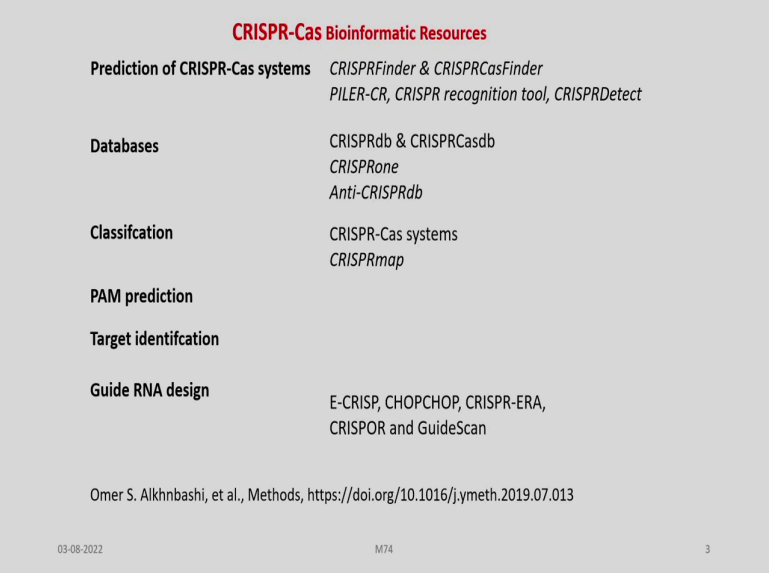
You know that CRISPR array contains repeat sequences which are separated by a series of specific spacer sequences, which are short pieces of DNA that originate from and match the corresponding parts of viral DNA called protospacers.

You also know that the signature repeat spacer architecture of CRISPR arrays was first described by Ishino et al in 1987. And then, Francisco Mojica also reported about them. And since the beginning, computational biology has played a crucial role in the discovery of CRISPR systems as well as in the generation of initial functional hypothesis.

Mojica used the power of bioinformatics to show that CRISPR arrays were not only present in *E. coli*, but also in most archaeal and many bacterial genomes. It was the power of bioinformatics analysis again which revealed that the CRISPR loci spaces have similarity to bacteriophages, and this led to the discovery that CRISPR-Cas systems act as an acquired immune system.

And this background note is being discussed. Just to give an idea that bioinformatics is a very important tool in the discovery as well as progress of CRISPR Cas technology.

(Refer Slide Time: 02:12)



So, we have many CRISPR Cas bioinformatics resources developed and each one of them serve different purposes. For example, in case of CRISPR Cas system discovery or prediction, we have CRISPRFinder and CRISPR Cas Finder, PILER-CR, CRISPR recognition tool or CRISPRDetect and so on. We have also many databases which have been built-up by various researchers and which help us further in CRISPR and Cas research and development of tools.

So, some of these databases are CRISPRdb, CRISPR Cas db and CRISPRone. And then there are other tools bioinformatic tools which helps us in the classification of CRISPR Cas system. And one such system is CRISPR Cas system itself and CRISPRmap.

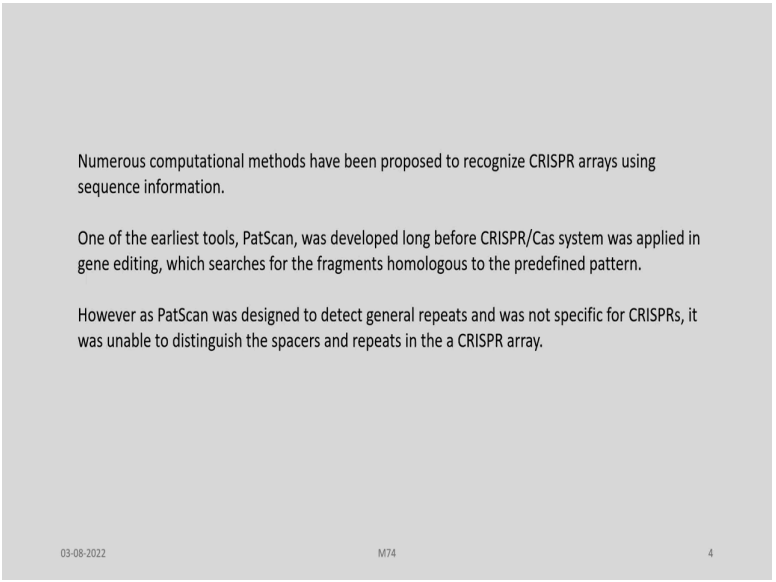
Then, there are workers who have developed tools which help us in the prediction of PAM as well as in the identification of targets. One of the important thing in CRISPR Cas9

technology is the development of the guide RNA. And we have discussed in brief how the various features of the guide RNA may help us in mobilizing the Cas9 to the target side.

So, guide RNA design is a very important step in this CRISPR Cas9 technology. And various researchers have developed numerous bioinformatics tools like E-CRISP, CHOPCHOP, CRISPR-ERA, CRISPOR and GuideScan, to help in the CRISPR RNA as regard RNA guide design.

So, if you are interested to know more about these various tools and softwares and online platforms, you can refer to this article by Omar in Methods, which has many important details.

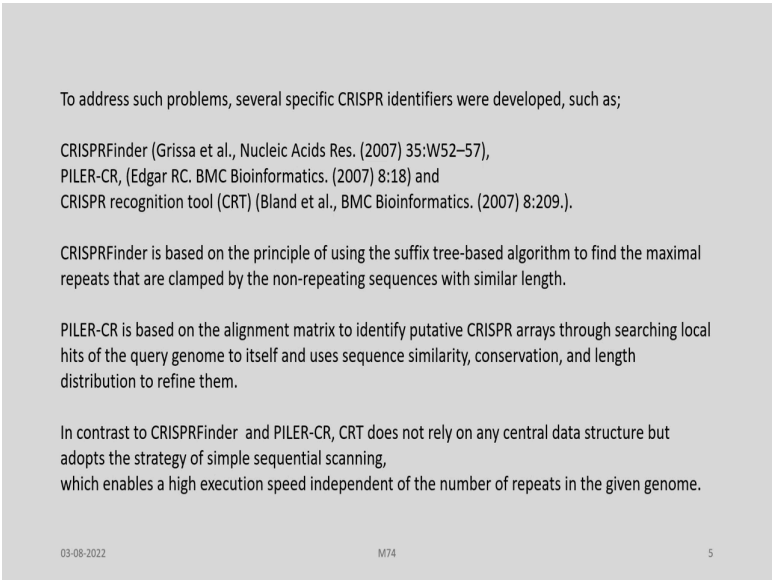
(Refer Slide Time: 04:34)



So, we now know that various computational tools have been proposed to recognize the CRISPR arrays using sequence information. One of the earliest tools that was used for recognizing CRISPR array is PatScan, but this was developed much before CRISPR Cas system was applied in gene editing. And these PatScan searches for the fragments homologous to the predefined pattern.

And it was not designed to detect specific CRISPR repeats, it was designed to detect only general repeats. And therefore, it was unable to distinguish the spacers and repeats in a CRISPR array.

(Refer Slide Time: 05:17)



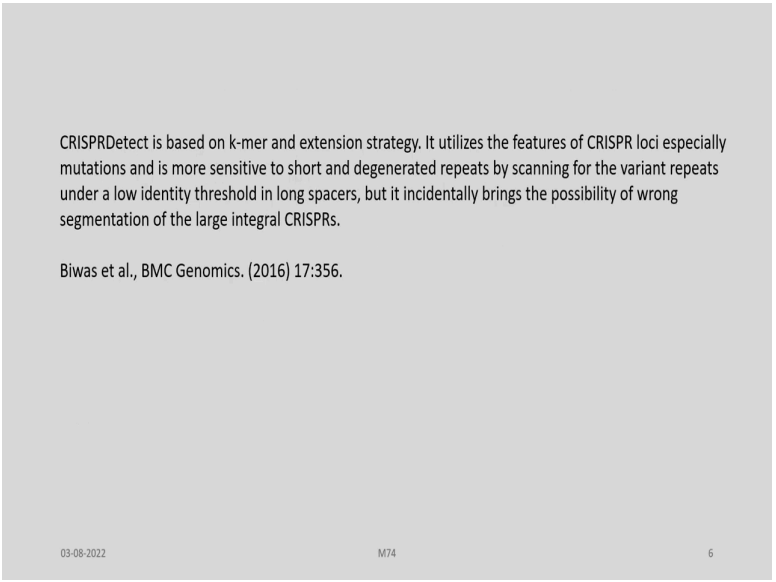
So, PatScan was not a very efficient tool as regards CRISPR array scanning. And to address such problems, several CRISPR specific tools were developed some of them are CRISPRFinder by Grissa et al. Then, we have PILER-CR by Edgar. Then CRISPR recognition tool or CRT by Bland.

The CRISPRFinder is based on the principle of using the suffix tree-based algorithm to find the maximal repeats that are claimed by the non-repeating sequences with a similar length.

While, PILER-CR is based on the alignment matrix to identify putative CRISPR arrays through searching local hits of the query genome to itself and uses sequence similarity, conservation, and the length distribution to refine them.

And in contrast to these CRISPRFinder and PILER-CR, a CRISPR recognition tool do not rely on any central data structure, but adopts the strategy of simple sequential scanning. And these enables a high execution speed, independent of the number of repeats in the given genome.

(Refer Slide Time: 06:35)



There is another tool called crisp CRISPRDetect which is based on k-mer and extension strategy. It utilizes the features of CRISPR loci, especially the mutations and is more sensitive to short and degenerated repeats by scanning for the variant repeats under a low identity threshold in long spacers, but it incidentally brings the possibility of wrong segmentation of the large integral CRISPRs.

So, as you will you can observe, various tools have various operating principles based on which they are designed. And some of them have certain advantages, and some of them have certain disadvantages.

(Refer Slide Time: 07:23)

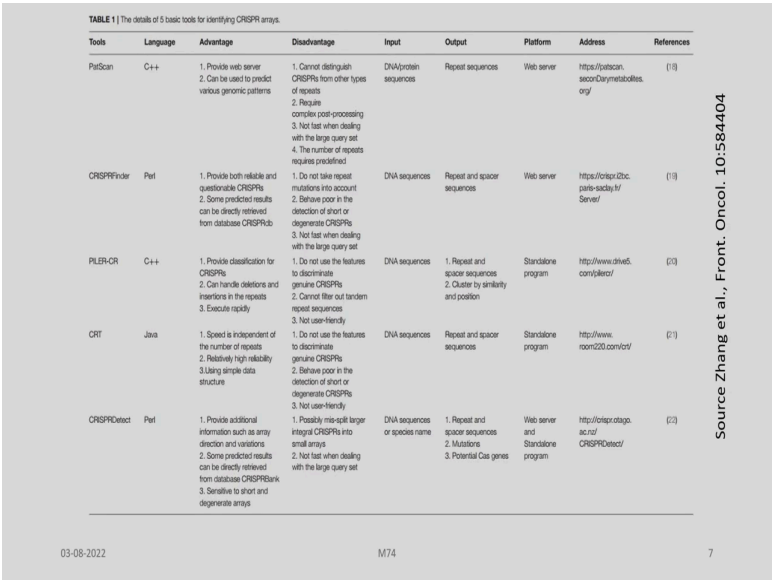


TABLE 1 | The details of 5 basic tools for identifying CRISPR arrays.

Tools	Language	Advantage	Disadvantage	Input	Output	Platform	Address	References
PatScan	C++	1. Provide web server 2. Can be used to predict various genomic patterns	1. Cannot distinguish CRISPRs from other types of repeats 2. Require complex post-processing 3. Not fast when dealing with the large query set 4. The number of repeats requires predefined	DNA/protein sequences	Repeat sequences	Web server	https://patscan.secondarymetabolites.org/	(18)
CRISPRFinder	Perl	1. Provide both reliable and questionable CRISPRs 2. Some predicted results can be directly retrieved from database CRISPRdb	1. Do not take repeat mutations into account 2. Behave poor in the detection of short or degenerate CRISPRs 3. Not fast when dealing with the large query set	DNA sequences	Repeat and spacer sequences	Web server	https://crispr.i2bc.paris-saclay.fr/Server/	(19)
PILER-CR	C++	1. Provide classification for CRISPRs 2. Can handle deletions and insertions in the repeats 3. Execute rapidly	1. Do not use the features to discriminate genuine CRISPRs 2. Cannot filter out tandem repeat sequences 3. Not user-friendly	DNA sequences	1. Repeat and spacer sequences 2. Cluster by similarity and position	Standalone program	http://www.drive5.com/pilercr/	(20)
CRT	Java	1. Speed is independent of the number of repeats 2. Relatively high reliability 3.Using simple data structure	1. Do not use the features to discriminate genuine CRISPRs 2. Behave poor in the detection of short or degenerate CRISPRs 3. Not user-friendly	DNA sequences	Repeat and spacer sequences	Standalone program	http://www.room220.com/crt/	(21)
CRISPRDetect	Perl	1. Provide additional information such as array direction and variations 2. Some predicted results can be directly retrieved from database CRISPRBank 3. Sensitive to short and degenerate arrays	1. Possibly mis-split larger integral CRISPRs into small arrays 2. Not fast when dealing with the large query set	DNA sequences or species name	1. Repeat and spacer sequences 2. Mutations 3. Potential Cas genes	Web server and Standalone program	http://crispr.otago.ac.nz/CRISPRDetect/	(22)

Source Zhang et al., Front. Oncol. 10:584404

You can see here the details of the 5 basic tools for identifying CRISPR array. Like, PatScan, CRISPRFinder, PILER-CR, CRT, CRISPRDetect which we discussed just now. And you can see the advantages listed against each of them and the disadvantages.

And one interesting thing is what is the input that is taken up or received by these bioinformatics tools also varies. In certain cases like, PatScan, it takes both DNA and protein sequences. But in other cases CRISPRFinder, PILER, CRT, it only accepts DNA sequences as input.

And the output is also varying. For example, in PatScan the output is repeat sequences. In CRISPRFinder we have both the repeat and spacer sequences. As already discussed PatScan was never intended or never designed to be used specifically for CRISPR. So, it is unable to predict or you know give the output of the special sequences. That is the reason why the other tools were developed.

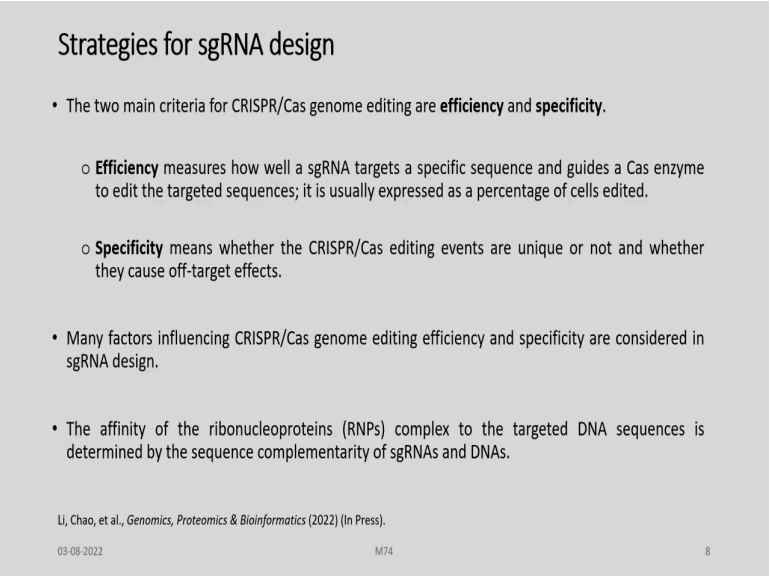
A PILER-CR, it gives repeat and spacer sequences as output and apart from that it also clusters by similarity and position, CRT gives repeat and spacers sequences. CRISPRDetect apart from giving the repeat spacers, it also offers information about the mutations and also the potential Cas genes.

Most of these are available as webserver programs, but PILER-CR and CRT, these are available only as stand-alone programs. And you can see the web address for downloading all these tools. So, you can download some of these tools and start working on them.

I will not go into the details of these advantages and disadvantages. You can refer to these paper source by Zhang et al in Frontier Oncology, and there you can study about the various advantages and disadvantages offered by each and every program.

Based on these advantages and disadvantages you can take a decision, which tool would be suitable for your type of application. As well as your availability of web connectivity. If you do not have web connectivity, you can download the standalone program and use it as such.

(Refer Slide Time: 10:14)

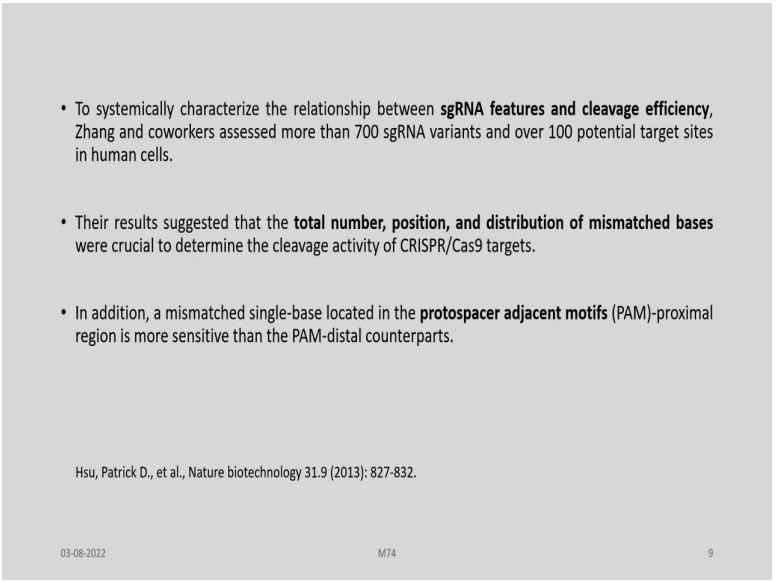


There are certain important strategies which are adopted for single guide RNA design and the two main criteria for CRISPR Cas genome editing are efficiency and specificity. Efficiency measures how well a single guide RNA targets a specific sequence and guides a Cas enzyme to edit the targeted sequences. It is usually expressed as a percentage of cells edited.

Specificity means whether the CRISPR Cas editing events are unique or not, and whether they cause off-target effects. Many factors influencing CRISPR Cas genome editing efficiency and specificity are considered in single guide RNA design.

The affinity of the ribonuclearproteins complex to the targeted DNA sequences is determined by the sequence complementary of sgRNAs and DNAs.

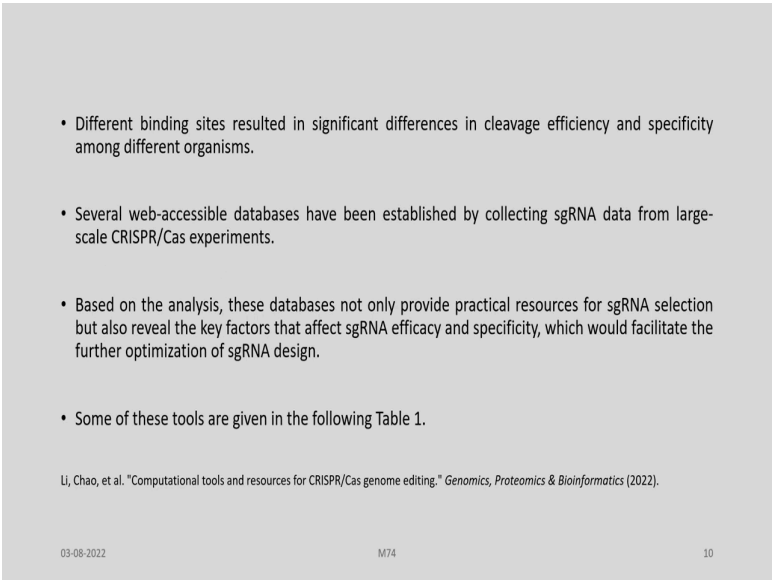
(Refer Slide Time: 11:11)



To systematically characterize the relationship between sgRNA features and cleavage efficiency, Zhang and co-workers assessed more than 700 single guide RNA variance and over 100 potential target sites in human cells. And from the analysis they suggested that the total number, position, and distribution of mismatched bases crucial to determine the cleavage activity of CRISPR Cas and targets.

They also told that, a mismatch single base located in the photosphere motifs proximal region is more sensitive than the PAM-distal counterparts.

(Refer Slide Time: 11:52)



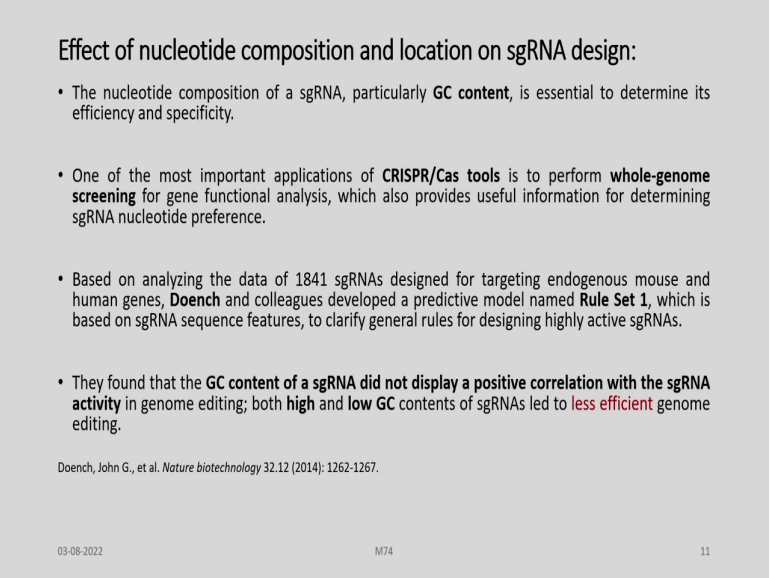
Different binding sites resulted in significant differences in cleavage efficiency and specificity among different organisms. Several web accessible databases have been established by collecting sgRNA data from the large scale CRISPR Cas experiments.

Based on the analysis these databases not only provide practical resources for sgRNA selection, but also reveal the key factors that affect sgRNA efficacy and specificity, which would facilitate the further optimization of sgRNA design.

Different binding sites resulted in significant differences in cleavage efficacy and specificity among different organisms. Several web accessible databases have been established by collecting sgRNA data from large scale CRISPR Cas experiments.

Based on the analysis these databases not only provide practical resources for sgRNA selection, but also reveal the key factors that affect sgRNA efficacy and specificity, which would facilitate further optimization of sgRNA design.

(Refer Slide Time: 12:57)



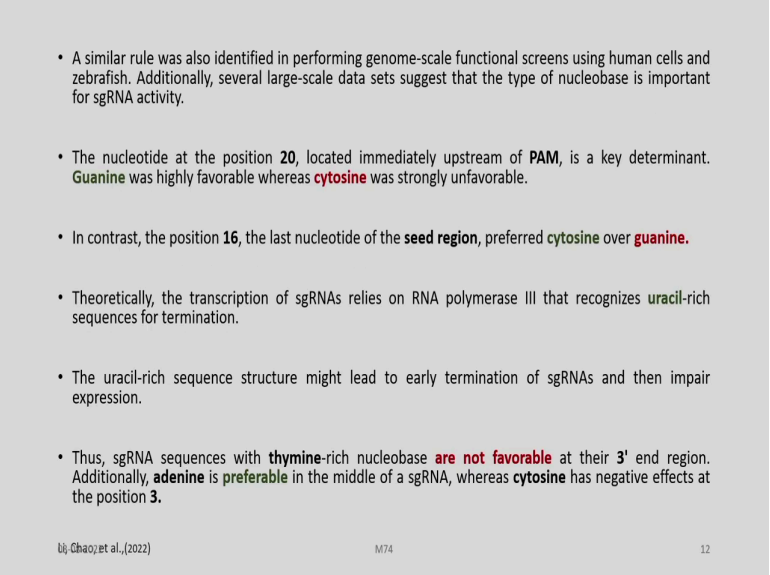
Effect of nucleotide composition and location on sgRNA design. The nucleotide composition of a sgRNA, particularly GC content is essential to determine its efficiency and specificity.

One of the most important applications of CRISPR Cas tools is to perform whole-genome screening of gene function analysis, which also provides useful information for determining sgRNA nucleotide preferences.

Based on analyzing the data of around 1900 sgRNAs designed for targeting endogenous mouse and human genes, Doench and colleagues developed a predictive model named Rule Set 1, which is based on sgRNA sequence features, to clarify general rules for designing highly active sgRNAs.

Doench and his colleagues found that the GC content of sgRNA did not display a positive correlation with the sgRNA activity in genome editing; both high and low GC contents of sgRNA to less efficient genome editing.

(Refer Slide Time: 13:58)



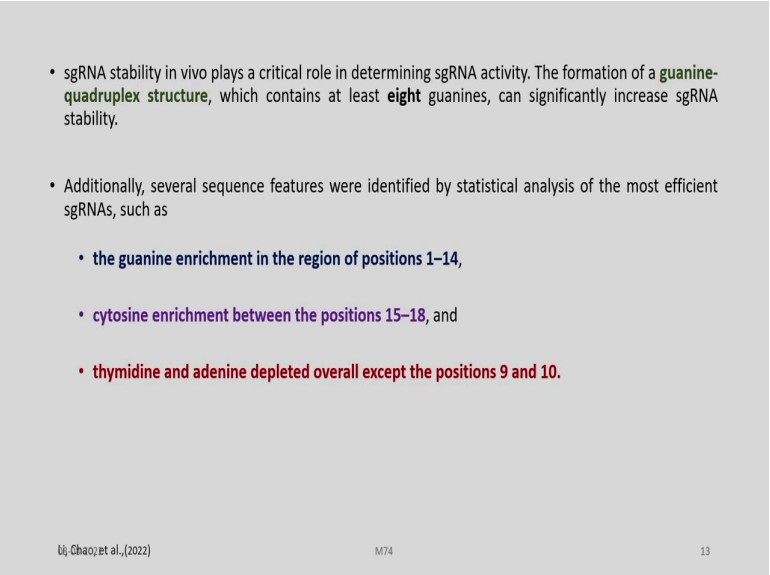
A similar rule was also identified in performing genome-scale functional screens using human cells and zebrafish. Additionally, several large-scale data sets suggest that the type of nucleobase is important for sgRNA activity. The nucleeotide at the position 20, located immediately upstream of PAM, is a key determinant. Guanine was highly favourable whereas, cytosine was strongly unfavourable.

In contrast, the position 16, the last nucleotide of the seed region, a preferred cytosine over guanine. Theoretically, the transcription of sgRNA relies on RNA polymerase 3, that recognizes uracil-rich sequences for termination.

The uracil-rich sequence structure might lead to early termination of sgRNAs and then impair expression. Thus, sgRNA sequences with thymine-rich nucleobases are not favourable at

their 3’ end region. Additionally, adenine is preferable in a middle of sgRNA whereas, cytosine has negative effects at position 3.

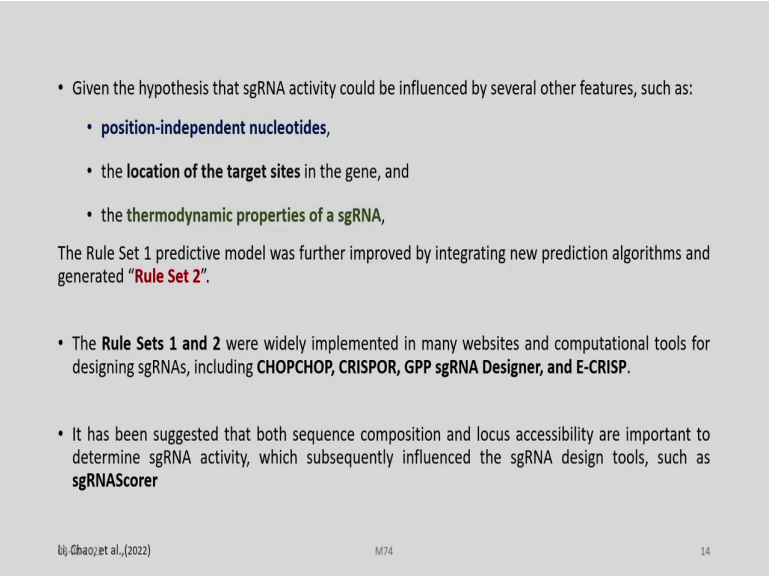
(Refer Slide Time: 15:05)



sgRNA stability in vivo plays a critical role in determining sgRNA activity. The formation of a guanine quadruplex structure, which contains at least 8 guanines, can significantly increase sgRNA stability.

Additionally, several sequence features were identified by statistical analysis of the most efficient sgRNA, such as, the guanine enrichment in the region of positions 1 to 14. Cytosine enrichment between the positions 15 to 18, and thymidine and adenine depleted overall except the positions 9 and 10.

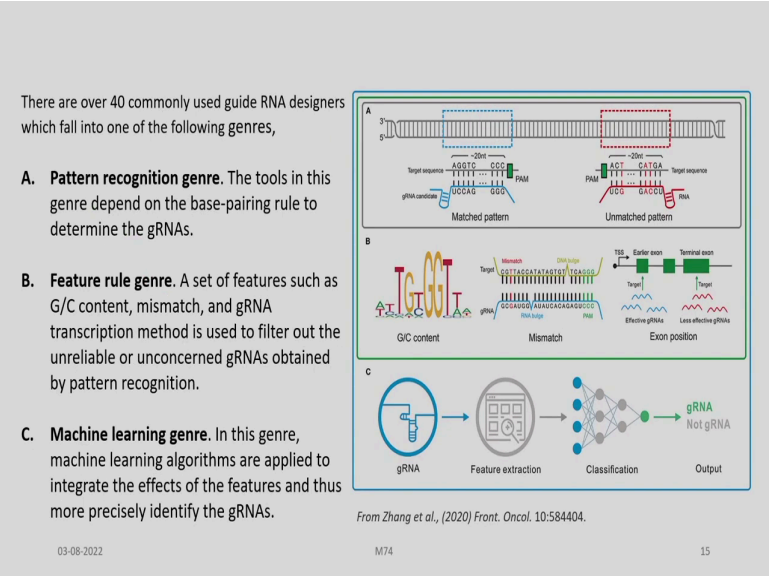
(Refer Slide Time: 15:40)



Given the hypothesis that sgRNA activity could be influenced by several other features, such as, position independent nucleotides, the location of the target sites in the gene, and the thermodynamic properties of a sgRNA. The Rule Set 1 predictive model was further improved by integrating new prediction algorithms and generated Rule Set 2.

The Rule Sets 1 and 2 were widely implemented in many web based tools and computational tools for designing sgRNAs, including CHOPCHOP, CRISPOR, GPP sgRNA designer, and E-CRISP. It has been suggested that both sequence composition and locus accessibility are important to determine sgRNA activity, which subsequently influence the sgRNA design tools, such as in the case of sgRNAScorer.

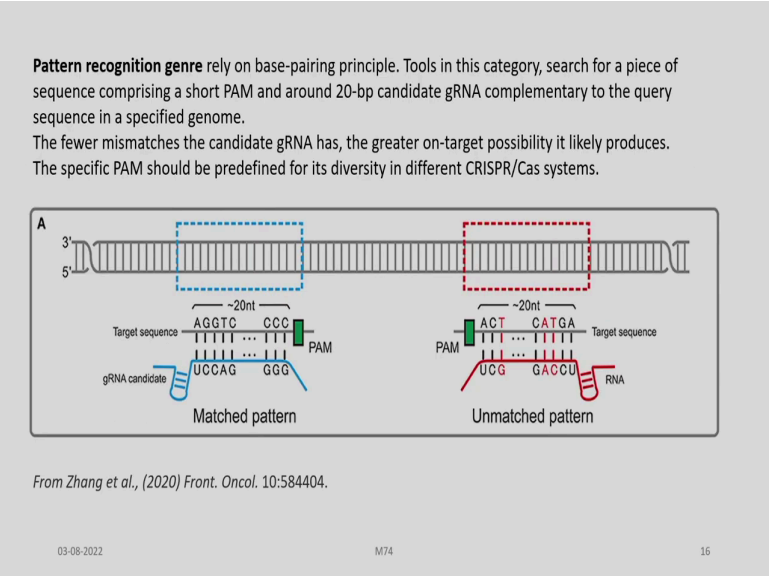
(Refer Slide Time: 16:31)



There are over 40 commonly used guide RNA designers which fall into one of the following genres. The first genre is the pattern recognition genre. The tools in these genre depend on the base-pairing rule to determine the gRNAs.

The second is the feature rule genre. Here a set of features such as GC content, mismatch, and gRNA transcription method is used to filter out the unreliable or unconcerned gRNAs obtained by pattern recognition. The third genre is the machine learning genre, where machine learning algorithms are applied to integrate the effects of the features and thus more precisely identify the gRNAs.

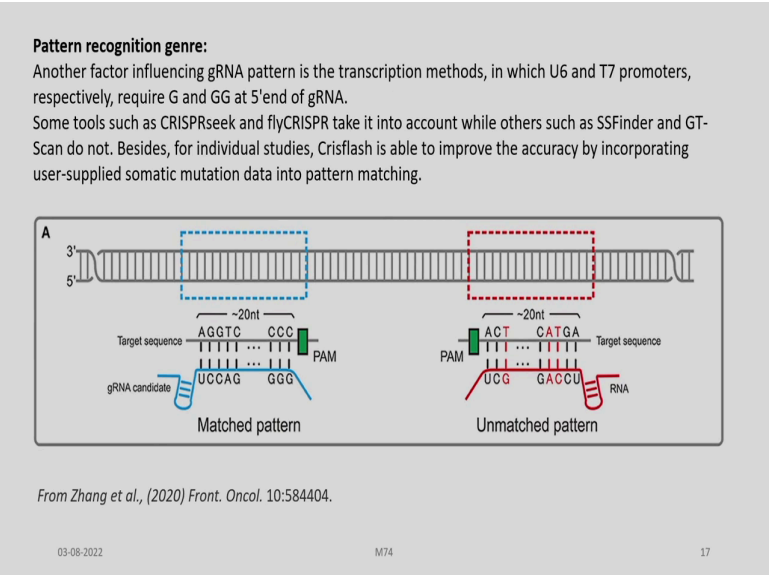
(Refer Slide Time: 17:27)



Let us see how the pattern recognition genre works. It rely on base-pairing principle. Tools in this category, search for a piece of sequence comprising a short PAM and around 20 base pair candidate guide RNA which is complementary to the query sequence in a specified genome.

The fewer mismatches the candidate gRNA has, the greater on-target possibility it likely produces. The specific PAM should be predefined for its diversity in different CRISPR Cas system.

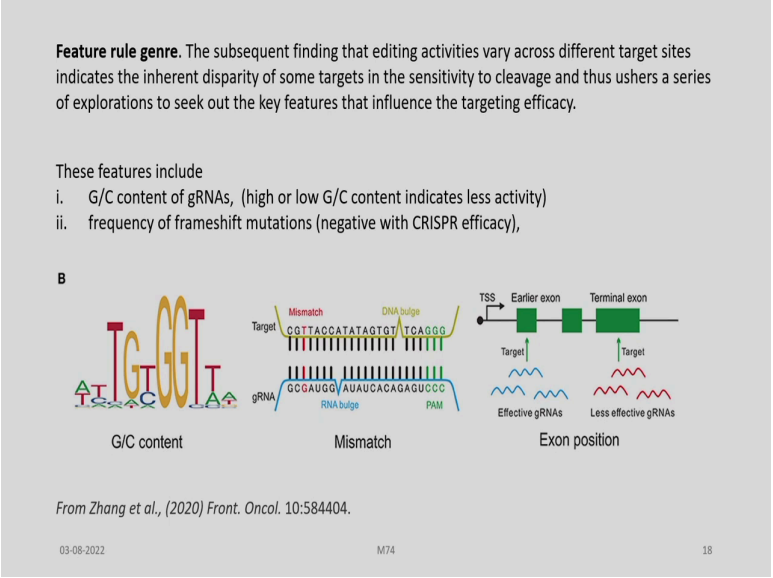
(Refer Slide Time: 18:07)



Another factor which influence gRNA pattern is the transcription method, based on which the pattern recognition genre functions. Here, U6 and T7 promoters, respectively, require G and GG at 5' end of gRNA. Some tools such as CRISPRseek and flyCRISPR take it into account while others such as SSFinder and GT Scan do not.

Besides, for individual studies, Crisflash is able to improve the accuracy by incorporating user-supplied somatic mutation data into pattern matching.

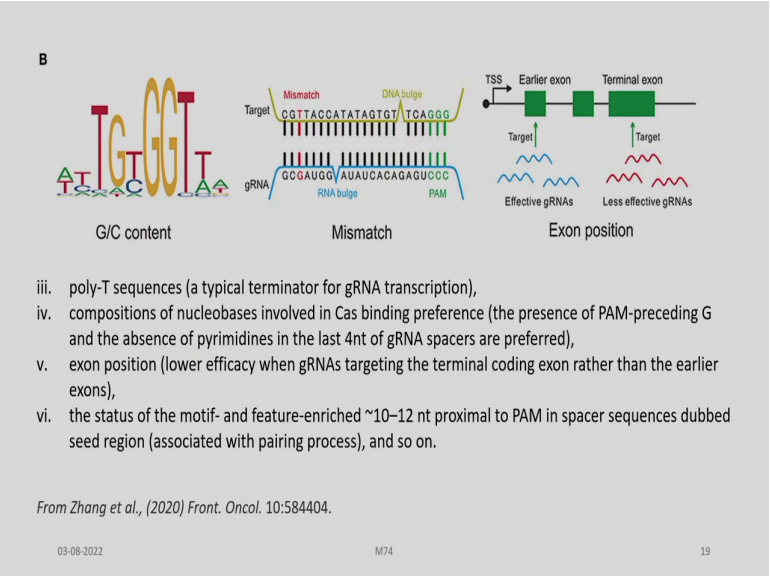
(Refer Slide Time: 18:56)



Let us discuss about the feature rule genre. The subsequent finding that editing activities vary across different target sites indicates the inherent disparity of some targets in the sensitivity to cleavage and thus assess a series of explorations to seek out the key features that influence the targeting efficacy.

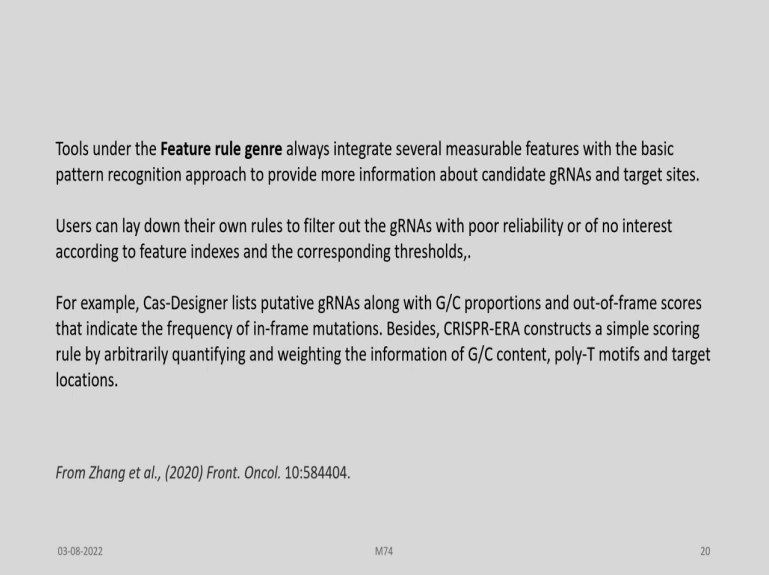
And these features include, number 1, GC content of gRNAs. Frequency of frameshift mutations.

(Refer Slide Time: 19:27)



Then, third is the poly-T sequences which is a typical terminator for gRNA transcription. Then, fourth is the compositions of nucleobases involved in Cas binding preference. Fifth is the axon position. And sixth is the status of the motif and feature-enriched 10 to 12 nucleotide proximal to PAM in spacer sequences dubbed the seed region.

(Refer Slide Time: 19:53)

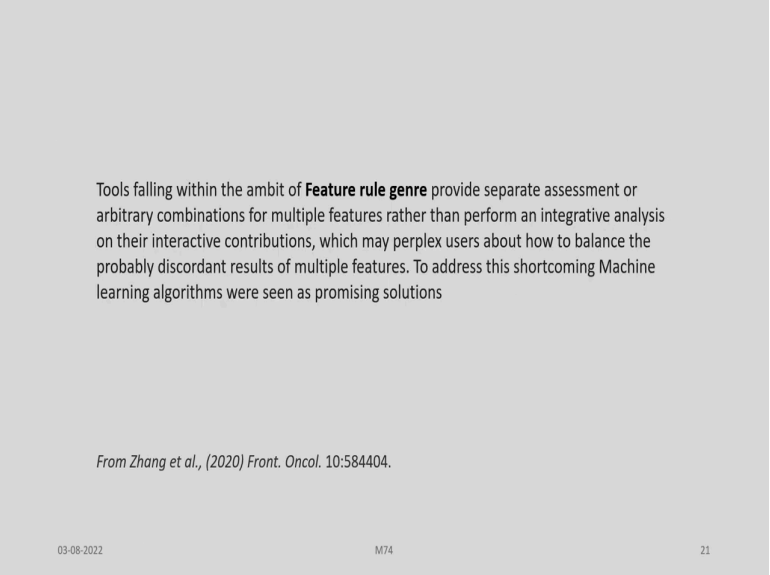


Tools under the Feature rule genre always integrate several measurable features with the basic pattern recognition approach to provide more information about candidate gRNA and

target sites. Users can lay down their own rules to filter out the gRNA with poor reliability or of no interest according to feature indexes and the corresponding thresholds.

For example, CAS-Designer list putative gRNAs along with GC proportions and out-of-frame scores that indicate the frequency of in-frame mutations. Besides CRISPR-ERA constructs a simple scoring rule by arbitrarily quantifying and weighing the information of GC content, poly-T motifs and target locations.

(Refer Slide Time: 20:40)



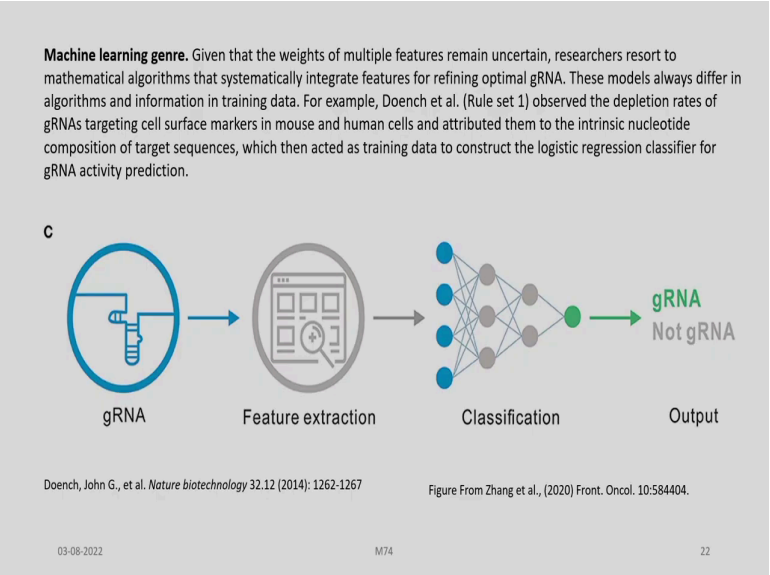
Tools falling within the ambit of Feature genre rule provides separate assessment or arbitrary combinations for multiple features rather than performing integrative analysis on their interactive contributions, which may perplexed users about how to balance the probably discordant results of multiple features. To address the shortcomings Machine Learning algorithms were seen as promising solutions.

Under the machine learning genre, given that the weights of multiple features remain uncertain, researchers resort to mathematical algorithms that systematically integrate features for refining optimal gRNA. These models always differ in algorithms and information in training data.

For example, we have discussed about dawn settle and the Rule Set 1. They observed the depletion rates of gRNA targeting cell surface markers in mouse and human cells and

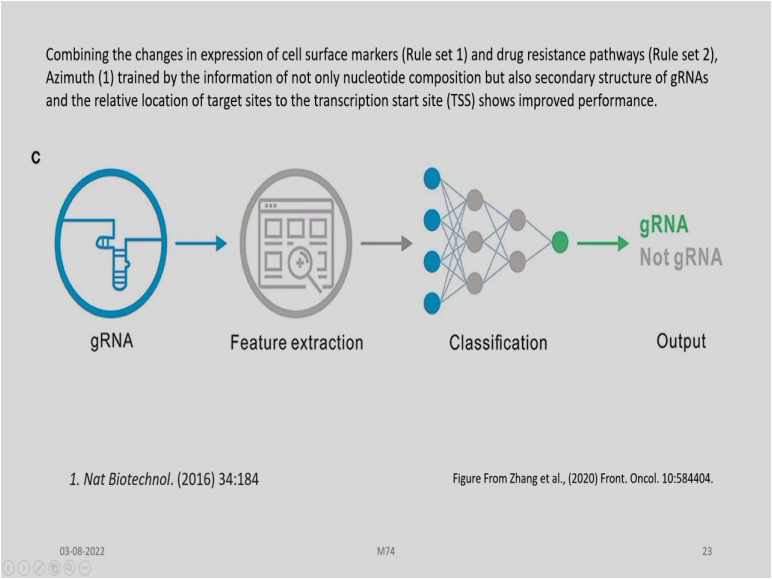
attributed them to the intrinsic nucleotide composition of target sequences, which then acted as training data to construct the logistic regression classifier for gRNA activity prediction.

(Refer Slide Time: 21:59)



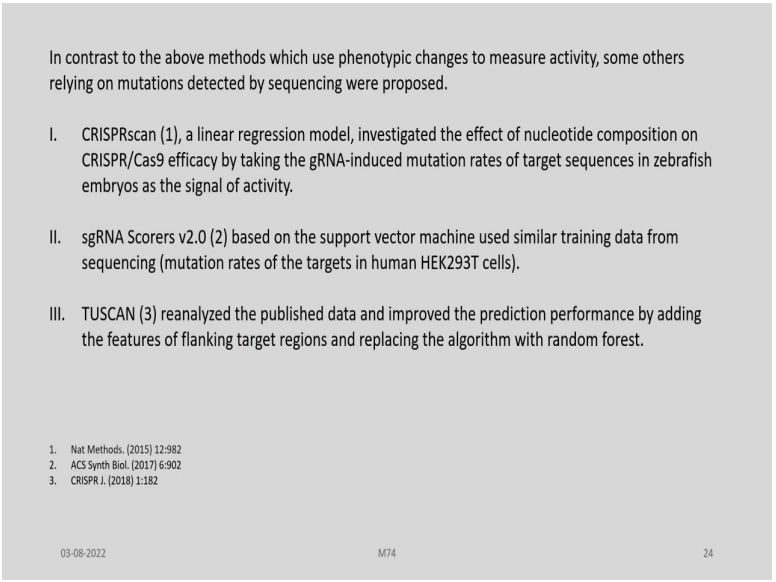
So, here you have in this figure the gRNAs and then the features are extracted. And based on these a classification takes place, and the probable candidates are selected as gRNA, while the remaining are left out as candidates which are not gRNAs. And these are the output.

(Refer Slide Time: 22:31)



Combining the changes in expression of cell surface markers and drug resistant pathways, which falls under the Rule Set 1 and Rule Set 2, respectively, they are trained by the information of not only nucleotide composition, but also secondary structure of gRNAs and the relative location of target sites to the transcription start site shows improved performances.

(Refer Slide Time: 22:57)

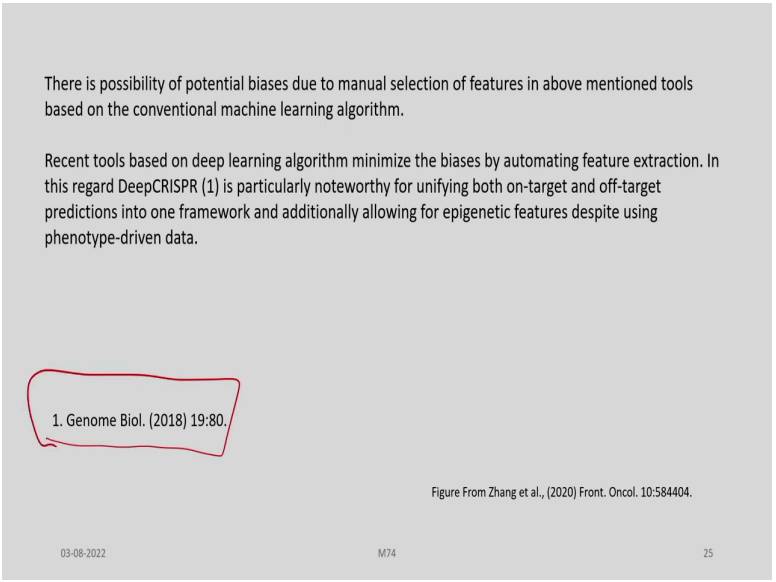


In contrast to the above methods which use phenotypic changes to measure activity, some others relying on mutations detected by sequencing were also proposed. For example, we have CRISPRscan, then we have sgRNA Scorers, then we have TUSCAN. For full details on these methods you can refer to these articles in Nature Methods, ACS Synthetic Biology and CRISPR Journal.

So, let us study in brief what are these CRISPRscan, sgRNA Scorers and TUSCAN. CRISPRscan here, a linear regression model investigated the effect of nucleotide composition on CRISPR Cas9 efficacy by taking the gRNA-induced mutation rates of target sequences in zebrafish embryos as the signal of activity. sgRNA Scorers based on the support factor machine used similar training data from sequencing mutation rates of the targets in human HEK293T cells.

TUSCAN reanalyzed the published data and improve the prediction performed by adding the features of flanking target regions and replacing the algorithm with random forest.

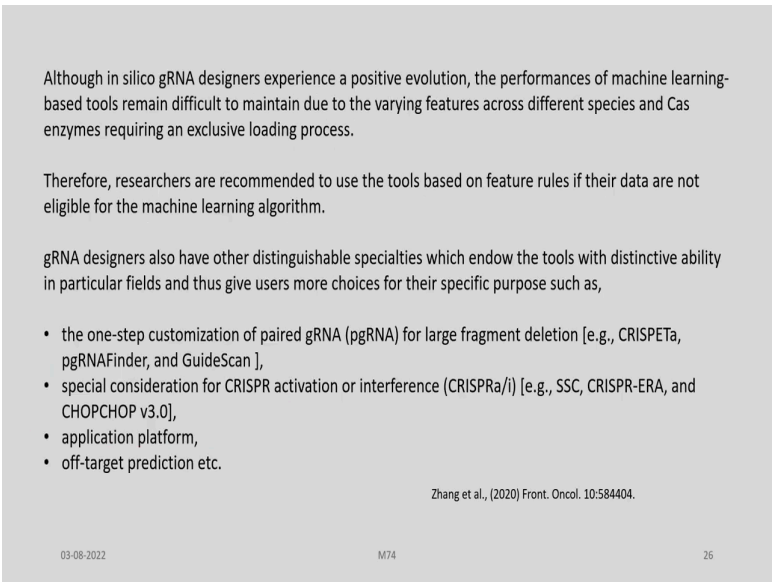
(Refer Slide Time: 24:17)



There is possibility of potential biases due to manual selection of features in above mentioned tools based on the conventional machine learning algorithm. Recent tools based on deep learning algorithm minimize the biases by automating feature extraction. In this regard DeepCRISPR is particularly noteworthy for unifying both on-target and off-target predictions into one framework and additionally allowing for epigenetic features despite using phenotype-driven data.

For further information, kindly consult this article in Genome Biology.

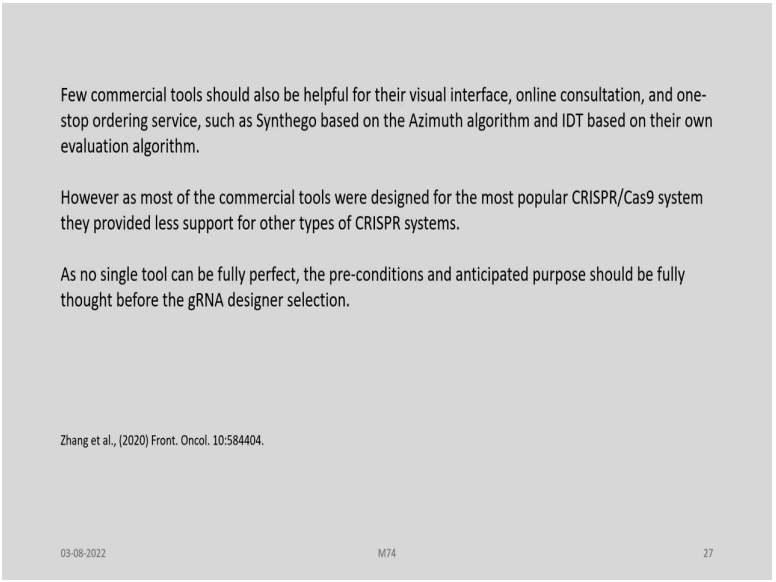
(Refer Slide Time: 24:58)



Although, in silico gRNA designers experience a positive evolution, the performances of machine learning based tools remain difficult to maintain due to the varying features across different species and Cas enzyme requiring an exclusive loading process. Therefore, researchers are recommended to use the tools based on feature rules if their data are not eligible for the machine learning algorithm.

gRNA designers also have other distinguishable specialities which endow the tools with distinctive ability in particular fields and thus give users more choices for their specific purposes. Such as, 1, the one-step customization of paired gRNA for large fragment deletion. Number 2, special consideration for CRISPR activation or interference. Number 3, the application platform. And number 4, off-target prediction.

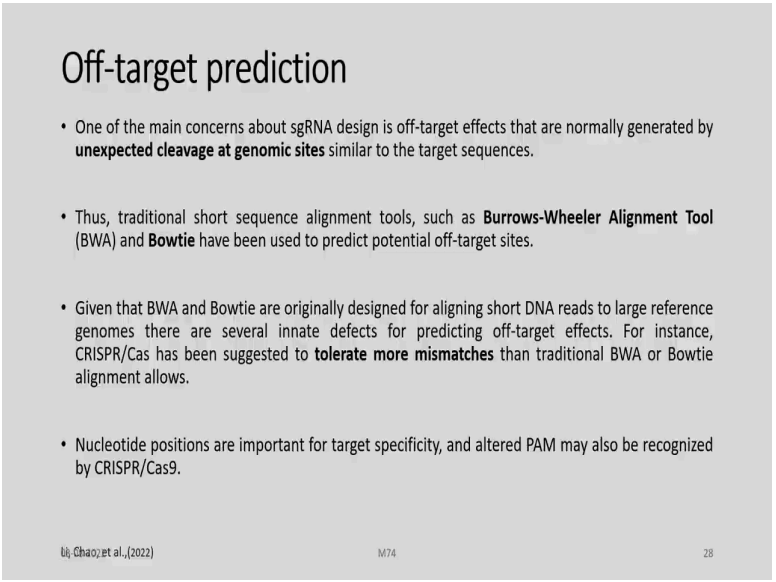
(Refer Slide Time: 25:57)



Few commercial tools are also helpful for their visual interface, online consultation, and one-stop ordering service, such as Synthego based on Azimuth algorithm and IDT based on their own evaluation algorithm.

However, as most of the commercial tools were designed for the most popular CRISPR Cas9 system, they provide less support for other types of CRISPR systems. So, we see that no any single tool can be fully perfect, the preconditions and anticipated purposes should be fully taught before the gRNA designer is selected for ones work.

(Refer Slide Time: 26:36)

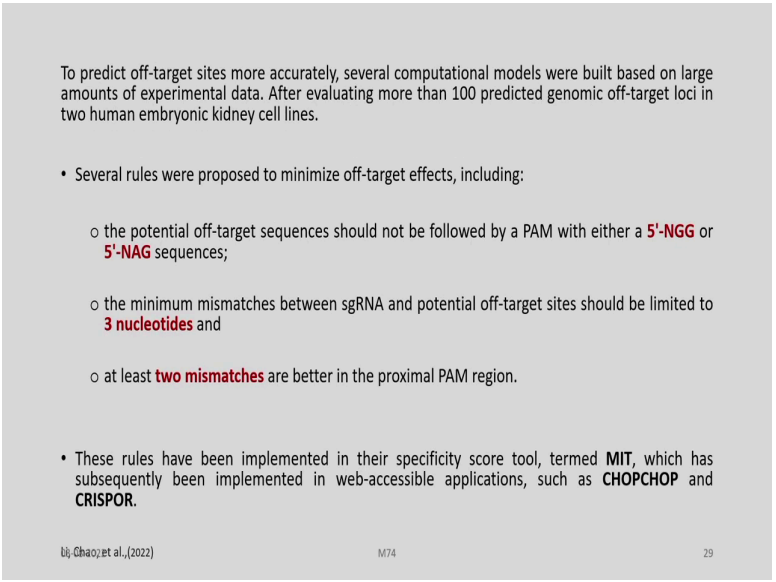


One of the important aspects of CRISPR Cas9 tools is the off-target reduction. So, therefore, the off-target prediction is very important. Traditional short sequence alignment tools, such as Barrows-Wheeler Alignment Tool and Bowtie have been used to predict potential off-target sites.

Given that BWA and bowtie are originally designed for aligning short DNA reads to large reference genomes. There are several innate defects for predicting off-target effects using these tools. For instance, CRISPR Cas has been suggested to tolerate more mismatches than traditional BWA or Bowtie alignments allows.

Nucleotide positions are important for target specificity, and altered PAM may also be recognized by CRISPR Cas 9.

(Refer Slide Time: 27:26)

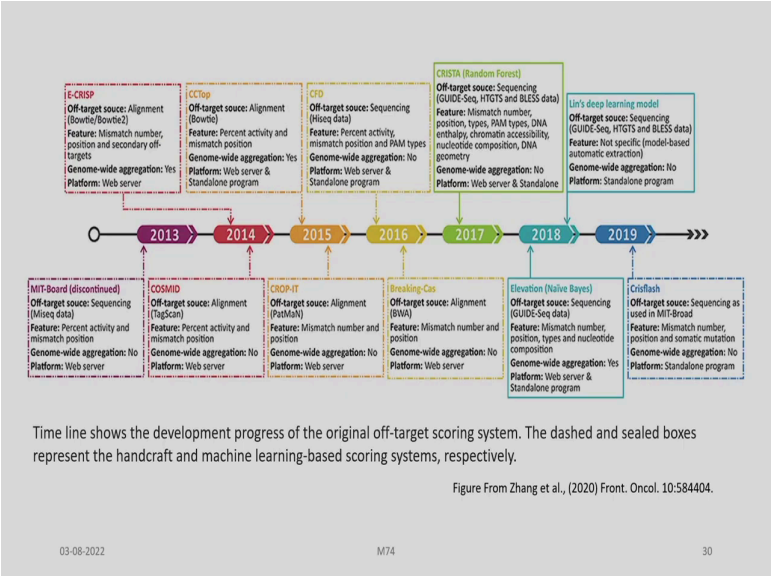


To predict off-target sites more accurately, several computational models have been built based on large amounts of experimental data. After evaluating more than 100 predicted genomic off-target loci in two human embryonic kidney cell lines.

Several rules are proposed to minimize off-target effects. Like, number 1, the potential off-target sequences should not be followed by a PAM with either a 5’ NGG or 5’ NAG sequences. Number 2, the minimum mismatches between single guide RNA and potential off-target sites should be limited to 3 nucleotides. And number 3, at least two mismatches are better in the proximal PAM region.

These rules have been implemented in their specificity score tool, termed MIT, which has subsequently been implemented in web-accessible applications, such as CHOPCHOP and CRISPOR.

(Refer Slide Time: 28:32)



Time line shows the development progress of the original off-target scoring system. The dashed and sealed boxes represent the handcraft and machine learning-based scoring systems, respectively.

Figure From Zhang et al., (2020) Front. Oncol. 10:584404.

These shows the timeline of the development progress of the original off-target scoring system. The dashed line and the seal boxes represent the handcraft and machine learning based scoring systems, respectively. So, in 2013 you can see the MIT-Board which is now discontinued. Here the off-target source was the sequencing Miseq data. And the features are percent activity and mismatch position. And it do not have any genome-wide aggregation, and it is available in the web server.

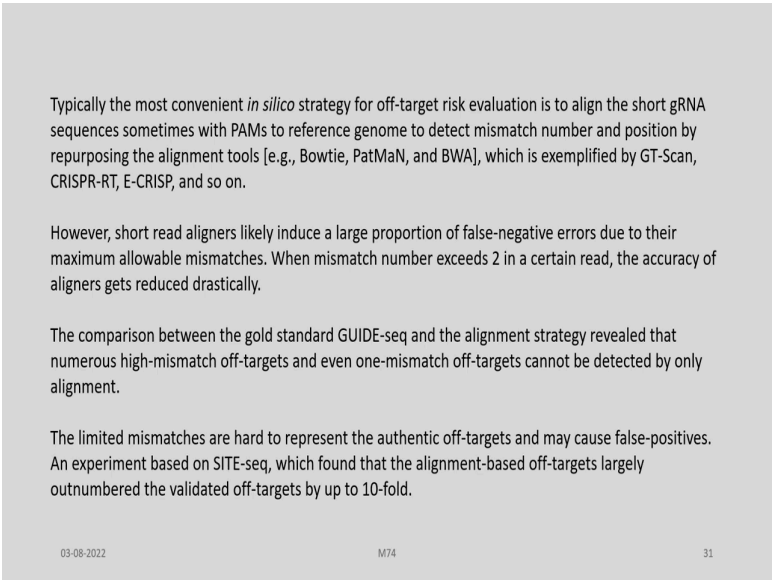
CCTop is not available in web server, it is a standalone program, so is CFD. As well as elevation of course, available both as a web server and a standalone program. Crisflash is only available as a standalone program, and so is Lin's deep learning model. While, CRISTA Random Forest is available both as a web server and standalone. The others are available over the web interface.

So, the whole idea of presenting these development of the various tools over time is that to, show the diversity and the continuous effort by researchers. So, you have so many of them continuously being developed over the years 13, 14, 15, 16, 17, 18, 19 and the latest of course, as shown in this figure is Crisflash.

So, here sequencing as used in MIT-Board. Feature is that mismatch number position and somatic mutations are displayed. And on Lin's deep learning model, here the sequencing is guide sequence is the GTS and BLESS data is used. A feature of course, these are not specific. So, in CRISTA you can see the mismatch number, position types, PAM types, DNA

enthalpy, chromatin accessibility, nucleotide composition, DNA geometry, so many things are being considered at the same time.

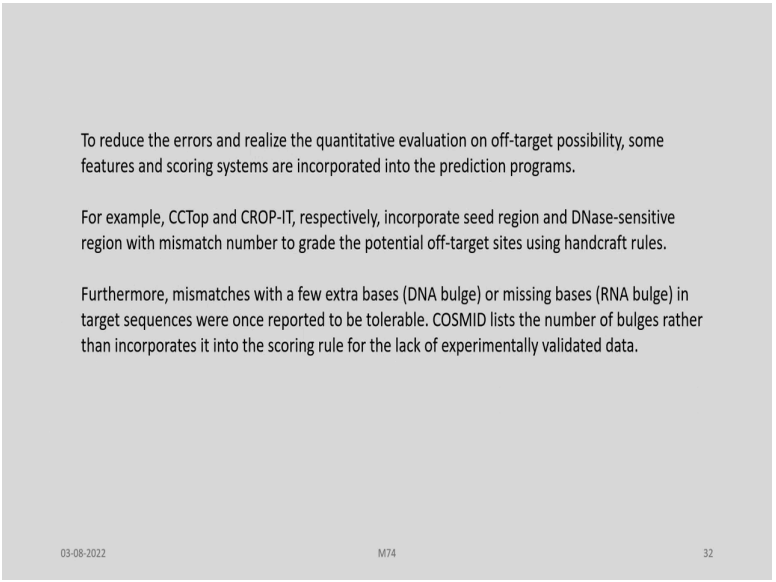
(Refer Slide Time: 31:00)



Typically, the most convenient in silico strategy for off-target evaluation is to align the short gRNA sequence sometime with PAMs to reference genome to detect mismatch number and position by repurposing the element alignment tools. However, short read aligners likely induce a large proportion of false-negative errors due to their minimum allowable mismatches. When mismatch number exceeds 2 in a certain read, the accuracy of aligners get reduced drastically.

The comparison between the gold standard, which is the GUIDE-seq, and the alignment strategy revealed that numerous high mismatch off-targets and even one-mismatch off-target cannot be detected by only alignment. The limited mismatches are hard to represent the authentic off-targets and may cause false-positives. An experiment based on SITE-seq, which found that the alignment based off-targets largely outnumbered the validated off-targets by up to 10-fold.

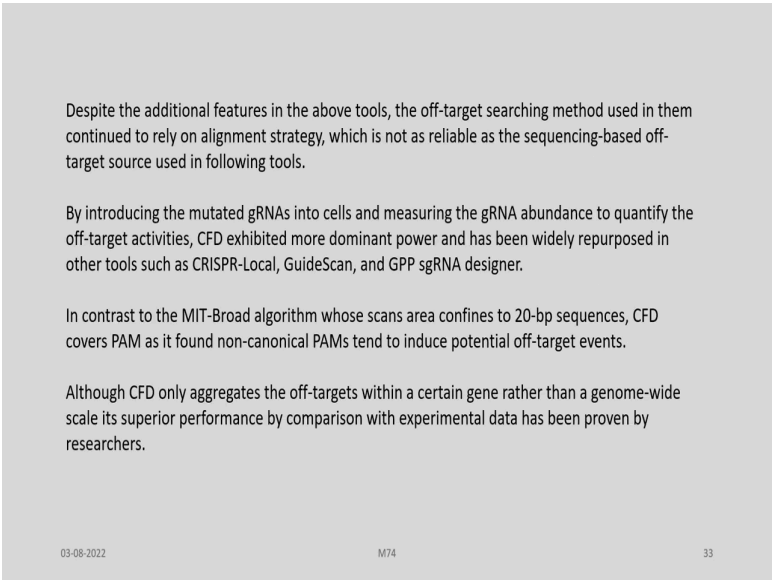
(Refer Slide Time: 32:02)



So, to reduce the errors and realize the quantitative evaluation on off-target possibility, some features and scoring systems are incorporated into the prediction programs. For example, in CCTop and CROP-IT, respectively, incorporate seed region and DNase-sensitive region with mismatch number to grade the potential off-target sites using handcraft rules.

Furthermore, mismatches with a few extra bases, where we have a DNA bulge or missing base, where we have a RNA bulge, in target sequences were once reported to be tolerable. COSMID lists the number of bulges rather than incorporates it into the scoring rule for the lack of experimentally validated data.

(Refer Slide Time: 32:49)



Despite the additional features in the above tools, the off-target searching method used in them continued to rely on alignment strategy, which is not as reliable as the sequencing-based off-target source used in following tools.

By introducing the mutated gRNAs into cells and measuring the gRNA abundance to quantify the off-target activities, CFD exhibited more dominant power and has been widely repurposed in other tools such as CRISPR-Local, GuideScan and GPP sgRNA designer.

In contrast to the MIT-Broad algorithm whose scans area confines to 20 base pair sequences, CFD covers PAM as it found non-canonical PAMs tend to induce potential off-target events. Although, CFD only aggregates the off-targets within a certain gene rather than a genome-wide scale its superior performance by comparison with experimental data has been proven by researchers.

The prediction of off-target is very important because of safety and toxicity concerns. So, to make CRISPR Cas 9, a better technology there is scope of developing better tools which can give us accurate off-target searching and predictions. Thank you we will continue our discussion about the various CRISPR Cas bioinformatics resources in the part b of this lecture.