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

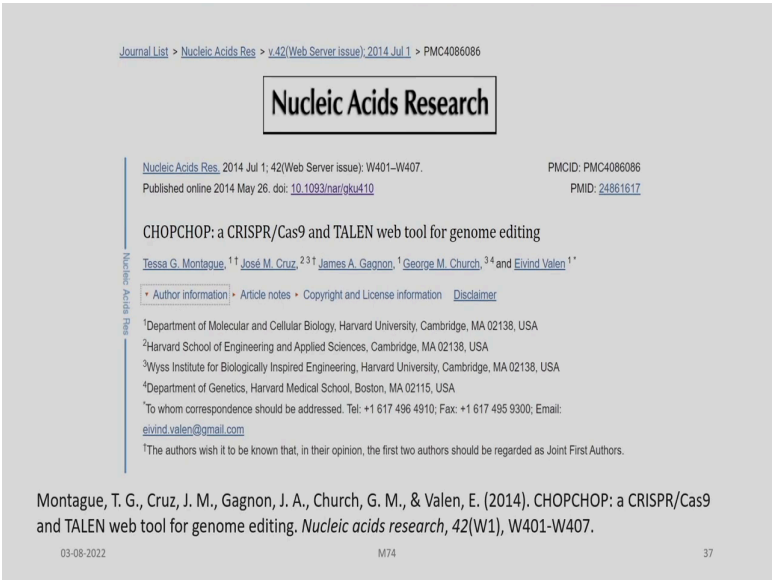
Module - 07
Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology
Lecture - 04
Computational Resources for CRISPR/Cas - Part B

Welcome back to my course on Genome Editing and Engineering. We are discussing about CRISPR Cas9 Computational Resources.

(Refer Slide Time: 00:41)

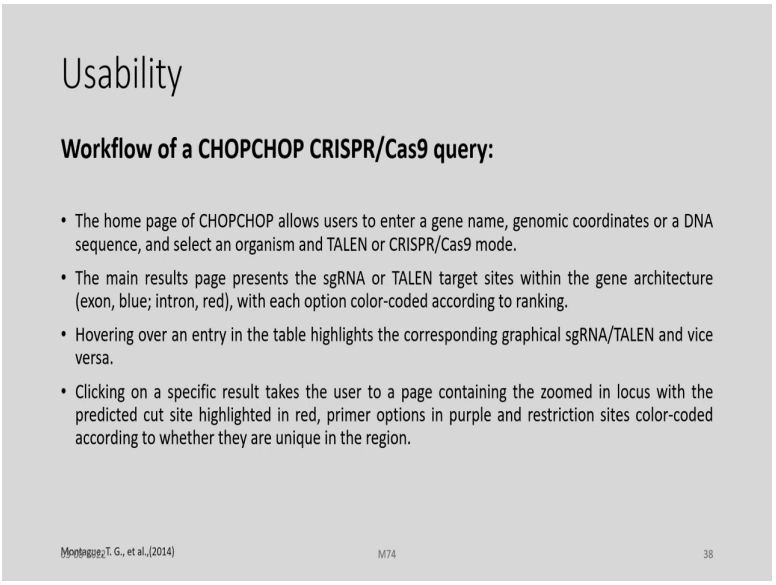
So, let us look into some of the Tools available for Design of sgRNA.

(Refer Slide Time: 00:45)



And you can see one of the tools developed by Montague and his colleagues are CHOPCHOP, a CRISPR, Cas9 and TALEN web tool for genome editing. So, this tool can be helpful in case of both CRISPR and TALEN.

(Refer Slide Time: 01:07)

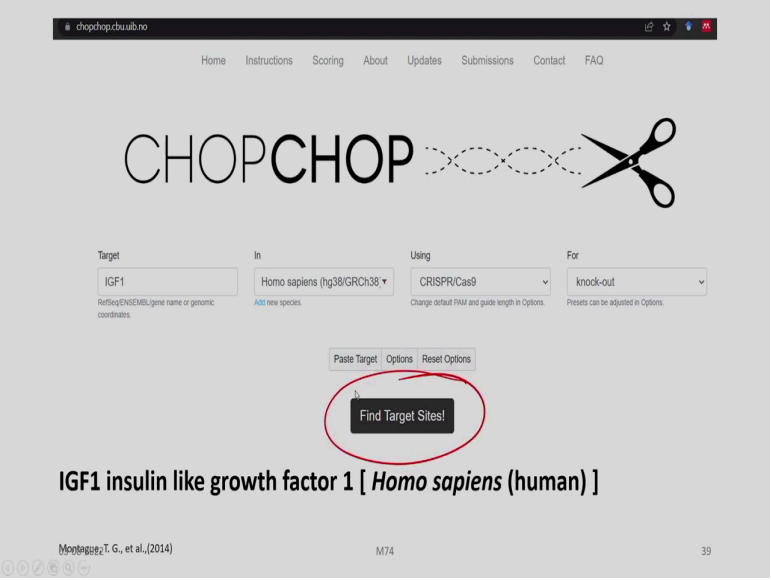


So, the usability of these particular CHOPCHOP is that you have a query which we will show in one of the slides later, how it can be used for design of sgRNA. The home page of CHOPCHOP allows users to enter a gene name, genomic coordinates, or a DNA sequence, and select an organism and TALEN or CRISPR Cas mode. So, if you want to develop a tool

for TALEN, we have to select TALEN. And if you want to develop for CRISPR, Cas9 we have to select CRISPR Cas9 mode.

The main results page presents the single guide RNA or TALEN target sites within the gene architecture which we have given as the input, with each option color coded according to ranking. Hovering over an entry in the table highlights the corresponding graphical single guide RNA or TALEN and vice-versa. Clicking on a specific result takes the user to a page containing the zoomed in locus with the predicted cut site highlighted in red, primer options in purple and restriction sites color coded according to whether they are unique in the region.

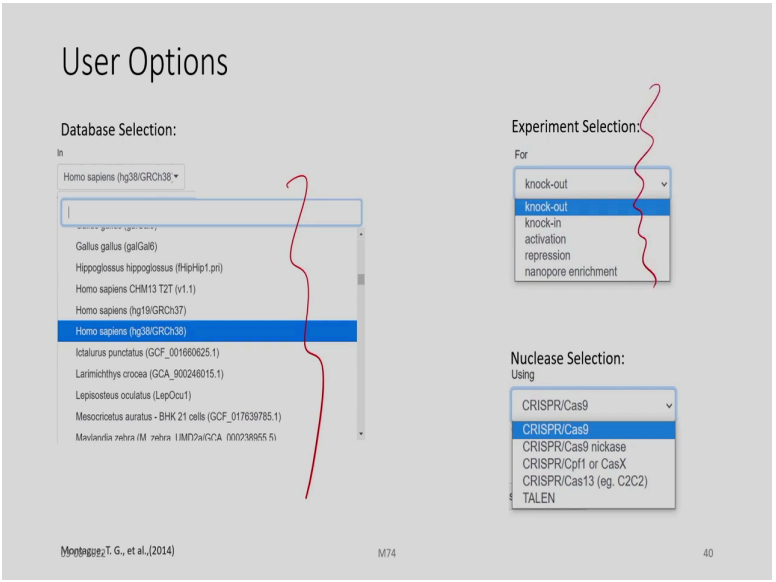
(Refer Slide Time: 02:32)



So, this is the web page of this CHOPCHOP. So, for example, here we have given a target IGF 1, which is insulin like growth factor. And this is we have selected in the *Homo sapiens*. So, here it is a pull-down menu. And here we are using the option CRISPR Cas9.

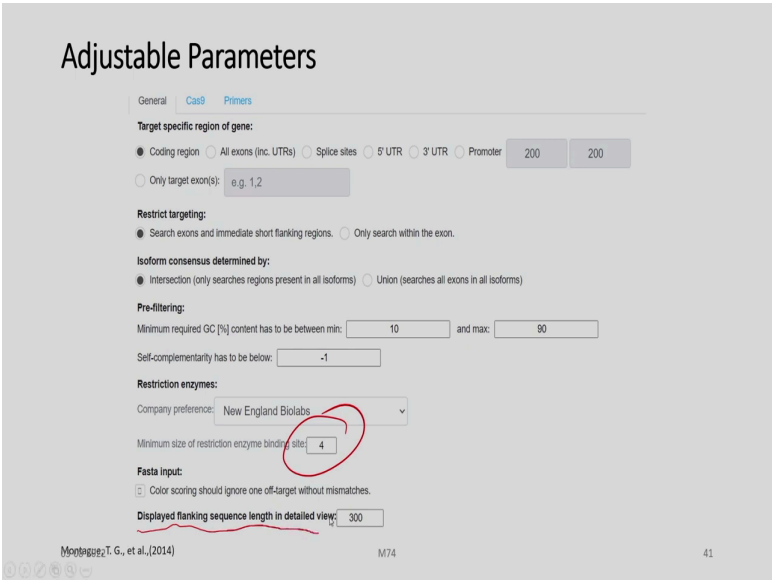
Again, we are defining it for what? For knock-out of this particular gene target so, we have various options. Now, if you land up in this web page, chopchop.cbu.uib.no, you can try typing in a gene of your interest and selecting an organism from the pull-down menu over here. And whether you are using for CRISPR Cas9 or TALEN, you can choose, and the various options whether knock-out, knock-in and so on, can be selected from over here.

(Refer Slide Time: 03:45)



And then, we can find the target sites. So, there is a facility for the user to give certain options as I have told you in the earlier case. The various organisms, the various experimental sections knock-out, knock-in, activation, repression. And, there is a facility for selection of the nuclease that we are going to use whether CRISPR Cas9 or nickase or CRISPR, Cas13 or even TALEN.

(Refer Slide Time: 04:21)

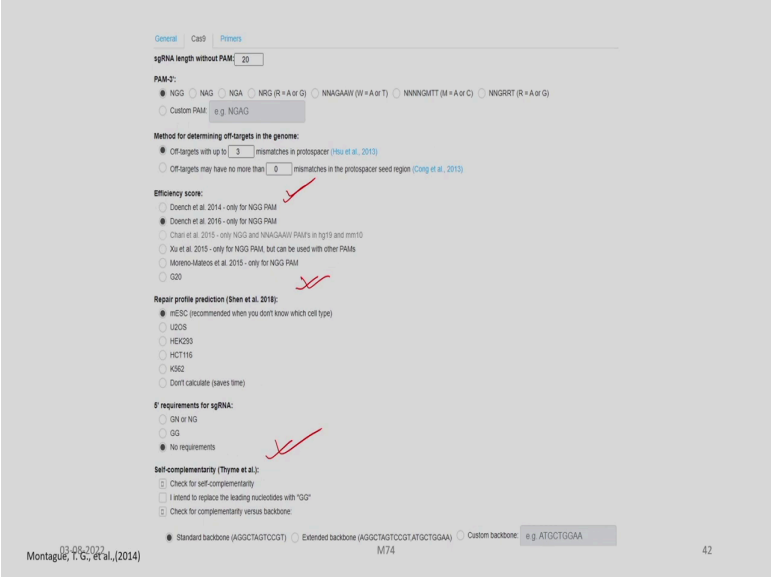


And we can adjust certain parameters. We can target specific regions of the gene, the cooling region or the exons or splice sites in UTR of 5' or 3' and so on. And then, we can also restrict

the targeting. We can just search within the exon. Then, we can have an option for intersection, then pre-filtering, we can minimum required GC percentage and the maximum here. You can choose and so self-complementary, t also we can choose.

And restriction enzymes from which commercial company it is available because once we plan the experiment, we need to find out the source of the enzyme and the supplier. So, you can plan quite in advance from where you are going to buy and whether these will be available with this companies or not. The minimum size of the restriction enzyme binding site and then faster input, and we also, have the option for display flanking sequence length in detailed view and so on.

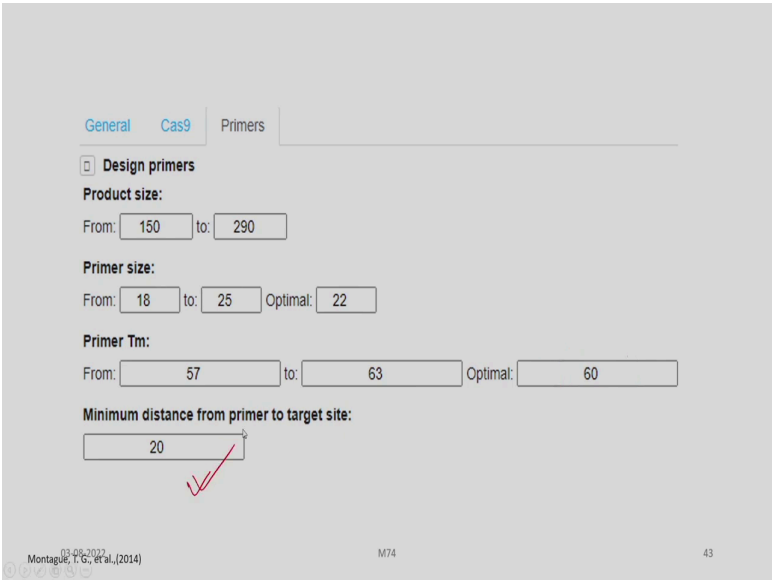
(Refer Slide Time: 05:40)



So, for this is the general options and then, for a Cas9 also we have various parameters which we can opt, PAM 3' can be NGG or NGA and method for determining off targets of the genome. We can feed a value here after gets with up to 3 mismatches and as per the protocol developed by Hsu et al or by Cong et al, you can make an choice.

Then efficiency score according to the Doench et al, 2014 method or 2016 method. Chari et al, Xu et al and so on and then, repair the profile prediction by Shen et al, 2018 and then 5' requirement of the sgRNA. We may simply choose as no requirement as well. And also, self-complementarity check as defined by Thyme et al.

(Refer Slide Time: 06:51)

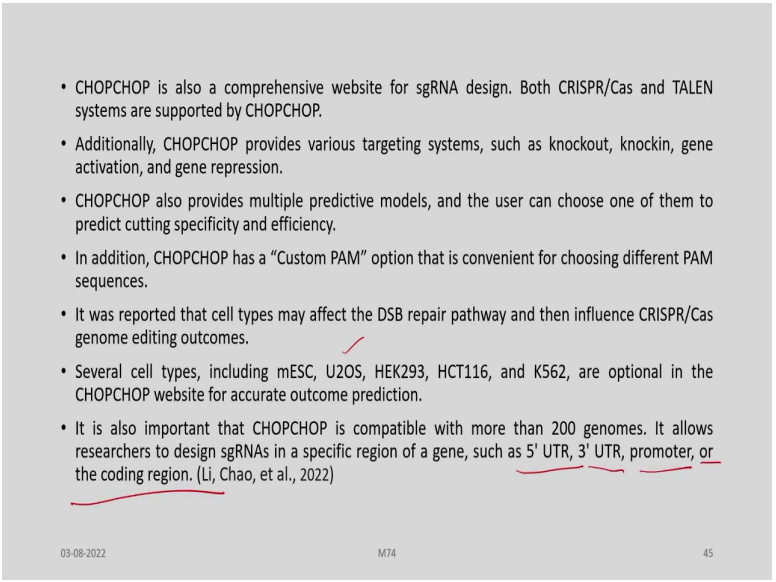


So, we can use all these options. And we can also have the options for the primers, product size from position say here 150 to 290 has been chosen. And the primer size from 18 to 25, and optimal we are giving around 22. Then, primer Tm or melting temperature from a minimum to maximum and optimal. And minimum distance from primer to target site also can be adopted here.

(Refer Slide Time: 07:30)

So, as already told, we will get a result. And if you hover around all these particular target sequences, we will get the further details.

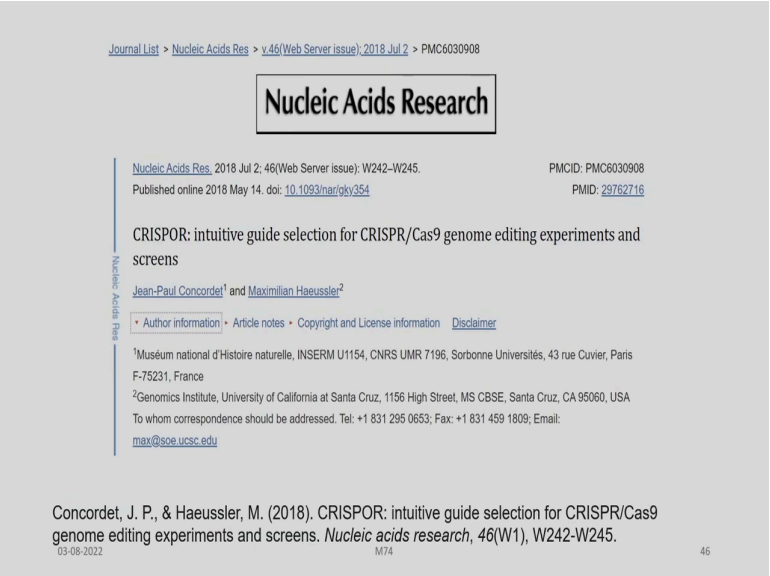
(Refer Slide Time: 07:46)



CHOPCHOP is a comprehensive website for sgRNA design, and we know that both CRISPR and TALEN systems are supported by these web resource. It provides various targeting systems, such as knock-out, knock-in that is now known to you. And also various predictive models can be chosen. So, the CHOPCHOP is a "Custom PAM" option that is convenient for choosing different PAM sequences.

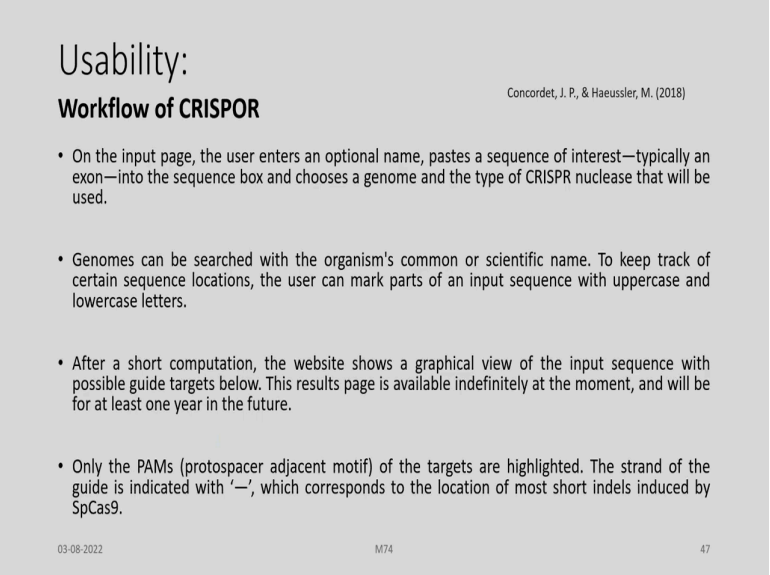
And it has been reported that cell types may affect the DSB repair pathway and then influence CRISPR, Cas genome editing outcomes. Several cell types, including, mESC, U20S, etcetera are optional in the CHOPCHOP website which we have just shown you for accurate outcome prediction. It is also important that CHOPCHOP is compatible with more than 200 genomes. It allows researchers to design sgRNAs in specific region of a gene, such as the 5' or 3' UTR, promoter, or the coding region.

(Refer Slide Time: 08:55)



Another tool which is available is CRISPR developed by Concordet and Haeussler.

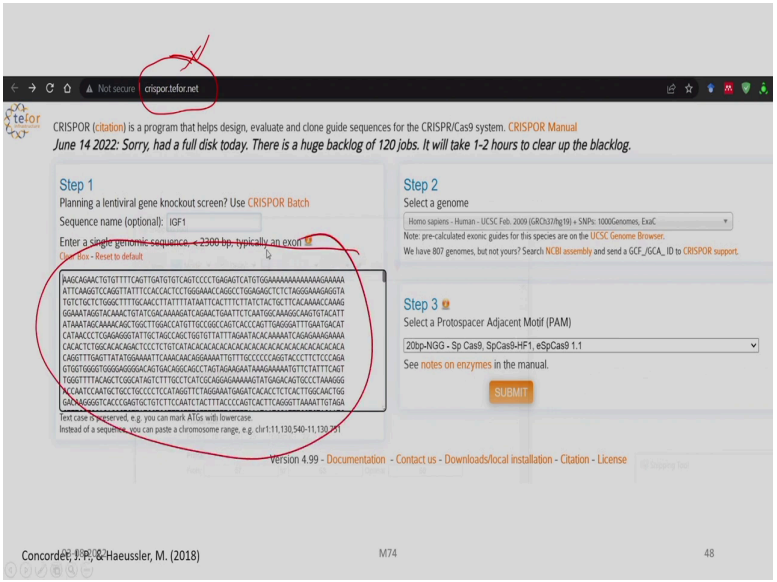
(Refer Slide Time: 09:05)



So, here, on the input page, the user enters an optional name, pastes a sequence of interest, typically an exon, into the sequence box and chooses a genome and the type of CRISPR nucleus that will be used. Genomes can be searched with the organism's common or scientific name. To keep track of certain sequence locations, the user can mark parts of an input sequence with upper case and lowercase letters.

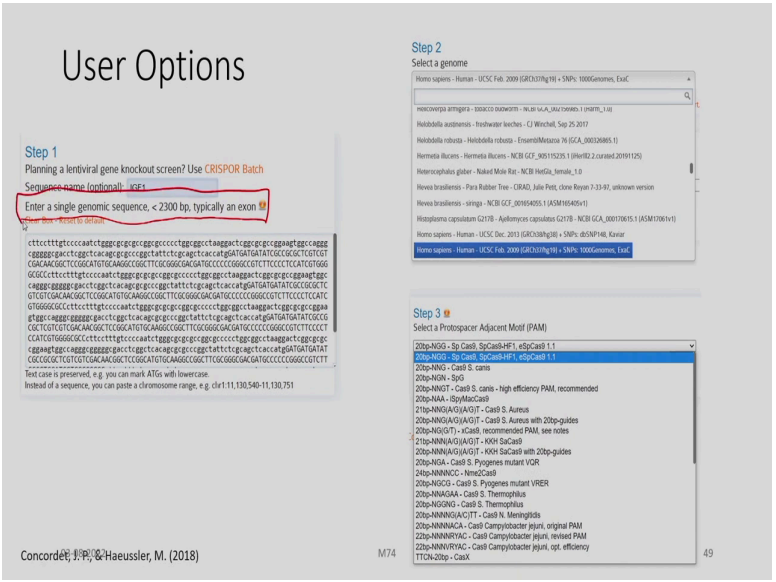
After a short computation, the website allows a graphical view of the input sequence with possible guide targets below. This result page is available indefinitely at the moment, and will be for at least one year in the future. So, only the PAMs of the target are highlighted. The strand of the guide is indicated with a dash, which corresponds to the location of the shortest indels induced by SpCas9.

(Refer Slide Time: 09:58)



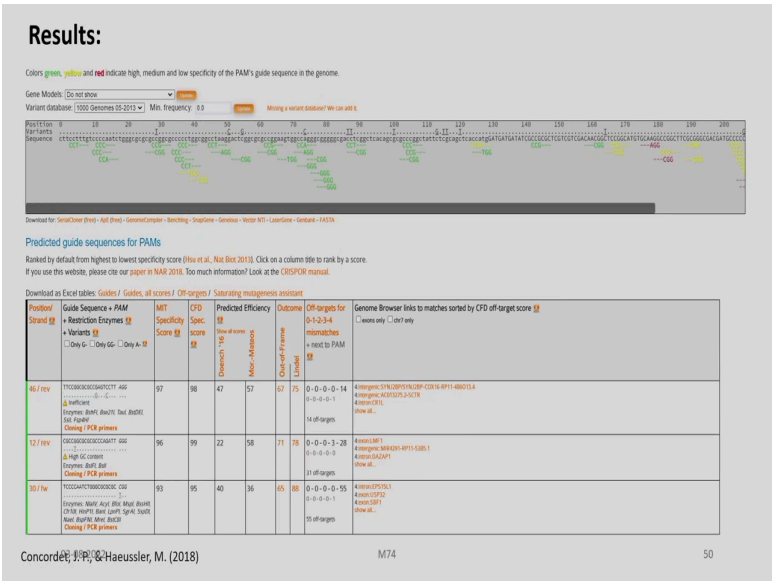
So, here you can see this website crispor.tefor.net, through which you can assess and you can enter a single genomic sequence here into this box, ok in step 1. And then, in step 2, you select a genome from this drop-down menu. And in step 3, you select a PAM motif, again from this drop-down menu.

(Refer Slide Time: 10:33)



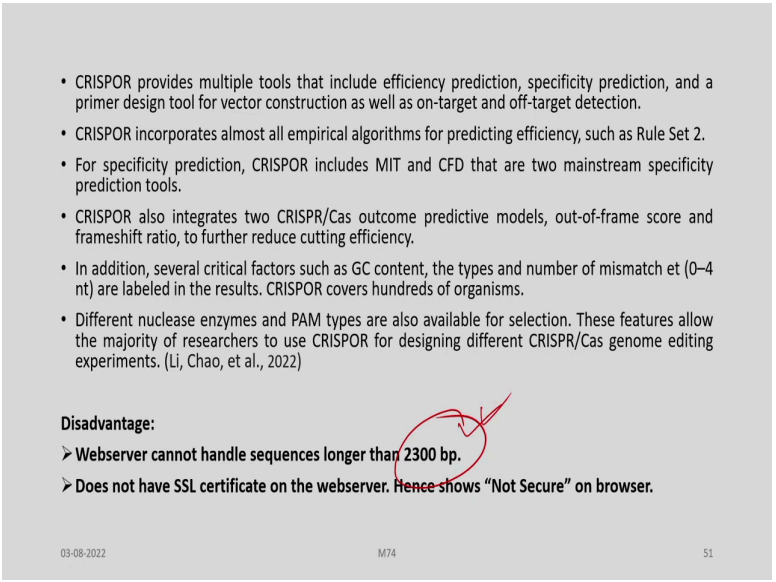
So, this already been discussed.

(Refer Slide Time: 10:38)



So, you have the results displayed in a page over here. So, colors green, yellow, and red indicate high, medium, and low specificity of the PAMs guide sequence in the genome. So, you have here the predicted guide sequences for PAMs. These are ranked by default from highest to lowest specificity scores. So, you can click on a column title to rank by a score. So, kindly visit this website and try to find out the sequences that you would desire for a genomic sequence of interest.

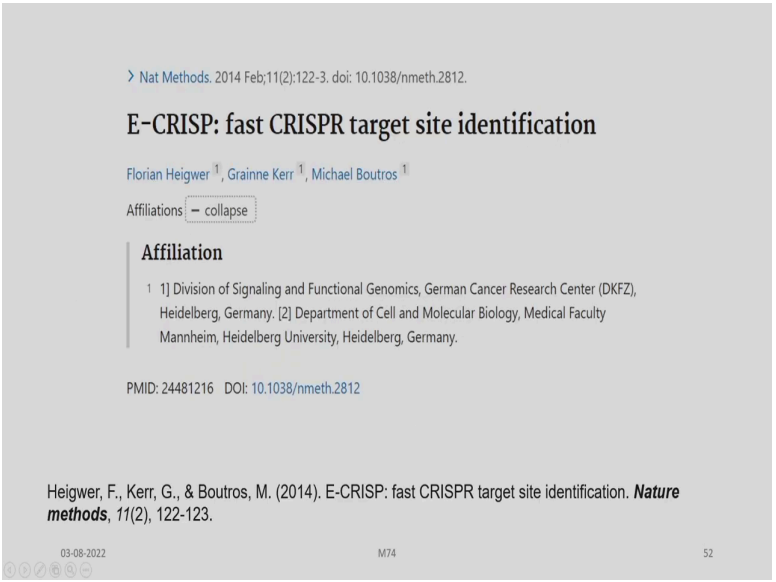
(Refer Slide Time: 11:28)



Overall, CRISPR provides multiple tools that include efficiency prediction, and specificity prediction, and a primer design tool for vector construction as well as on-target and off-target detection. It incorporates almost all empirical algorithms for predicting efficiency, such as Rule Set 2. For specificity prediction, CRISPOR includes MIT and CFD that are two mainstream specificity prediction tools. It integrates two CRISPR Cas outcome predictive models, out-of-frame score and frameshift ratio, to further reduce cutting efficiency.

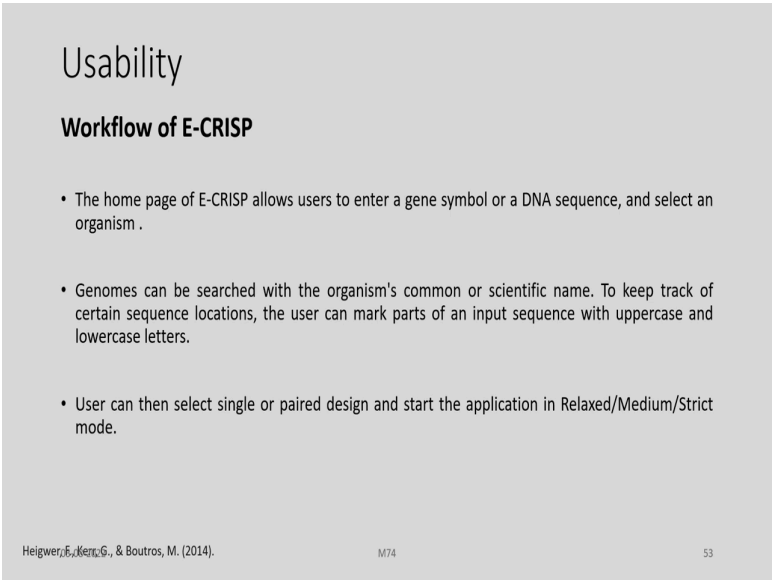
In addition, several critical factors such as GC content, the types and numbers of mismatches are labeled in the results. CRISPOR covers hundreds of organisms. Different nuclease enzymes and PAM types are also available for selection under CRISPOR. These features allow the majority of researchers to use CRISPOR for designing different CRISPR Cas genome editing experiments. However, there are certain disadvantages associated. The webserver cannot handle sequences longer than 2300 base pairs. And here it is highlighted, enter a single genomic sequences which should be lesser than 2300.

(Refer Slide Time: 12:53)



Let us now discuss about E-CRISP which have been developed by Heigwer and his colleagues which is used for first CRISPR target site identification.

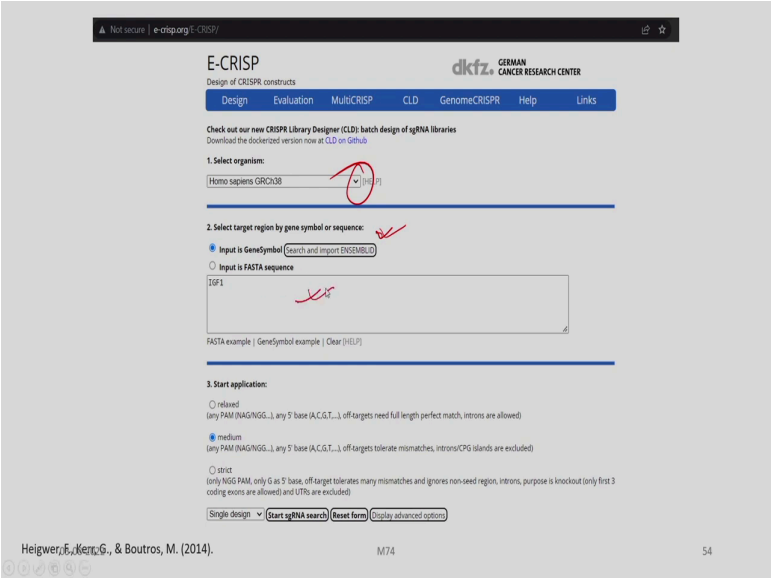
(Refer Slide Time: 13:06)



The home page of these E-CRISP allows users to enter a gene symbol or a DNA sequence and select an organism similar to the earlier web based software's that we have discussed. The genomes can be searched with the organism's common or scientific name. To keep track of certain sequence locations, the user can mark parts of an input sequence with uppercase

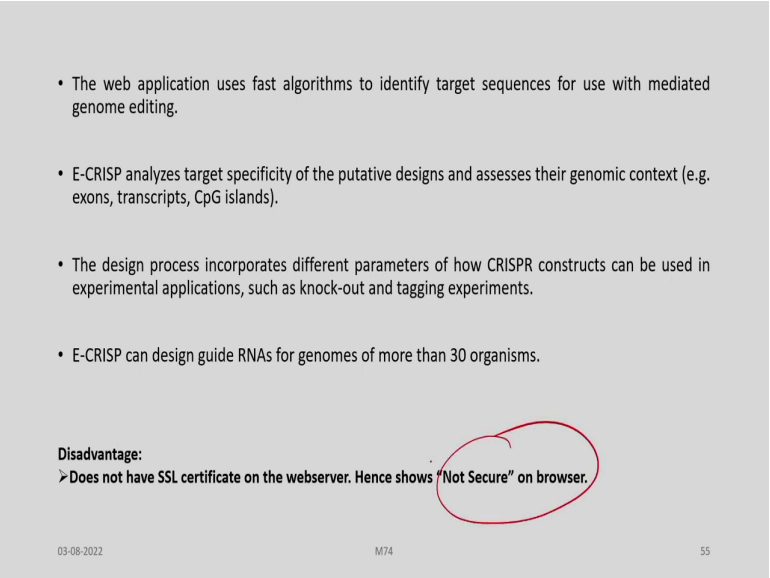
and lowercase letters. User can then select single or paired designs and start the application in Relaxed, Medium, Strict mode.

(Refer Slide Time: 13:41)



So, this is the website, e-crisp.org. And this is the German Cancer Research Center which facilitates this web page and this program. So, here this is the software which is used for design of CRISPR constructs. So, you can select the organism as already told, and input in the gene symbol, or you can also put as a fasta sequence by pasting it over here. And then you can choose the level of stringency whether relaxed, medium, or strict over here.

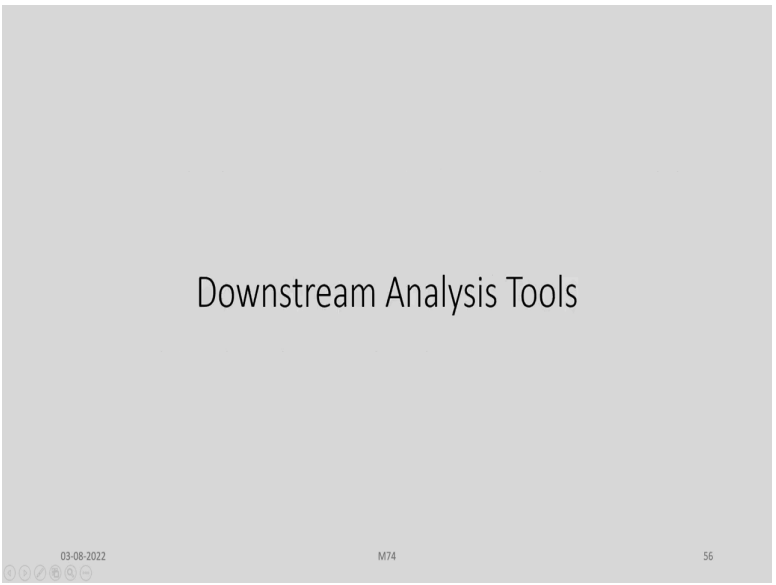
(Refer Slide Time: 14:31)



And this web application uses fast algorithms to identify target sequences for use with mediated genome editing. It analyzes target specificity of the putative designs and assesses their genomic context, example exons, transcripts, CpG islands and so on. The design process incorporates different parameters of how CRISPR constructs can be used in experimental applications, such as knock-out and tagging experiments.

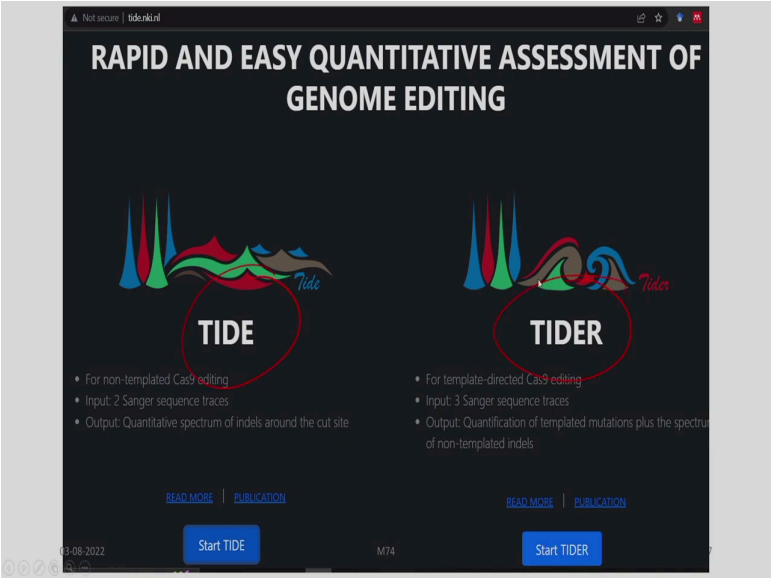
It can design guide RNAs for genomes of more than 30 organisms. The disadvantage of these particular program is that, is does not have SSL certificate on the webserver and therefore, it shows “Not Secure” on the browser.

(Refer Slide Time: 15:22)



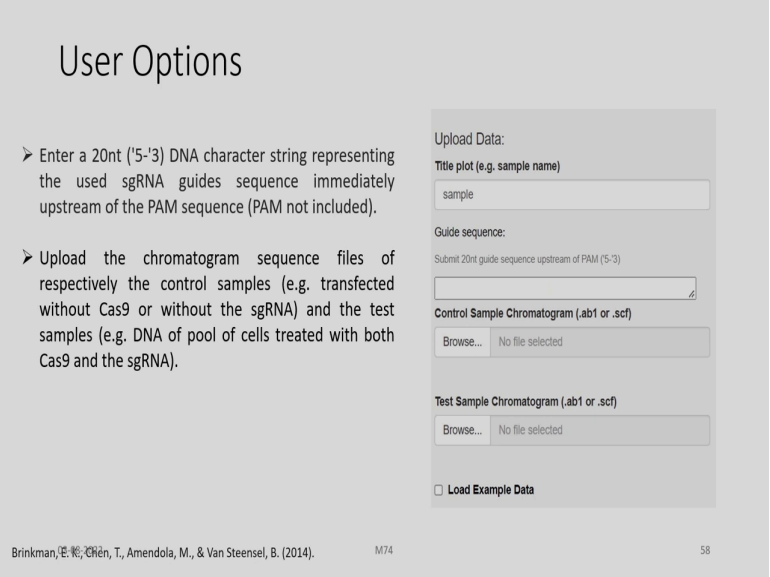
One of the important considerations in CRISPR Cas9 design tools is the Downstream Analysis Tools.

(Refer Slide Time: 15:32)



We have two such tools, one is the TIDE, another is the TIDER. So, this TIDE is for non-templated Cas9 editing and TIDER is for template directed Cas9 editing. In both cases, you have to provide the inputs either as Sanger sequence traces, here 2 and here 3. And then output is quantitative spectrum of indels around the cut site in the case of TIDE and in the case of TIDER the output is quantification of templated mutations plus the spectrum of non-templated indels. So, you can start either of these programs by visiting the particular program through these buttons which provides the option.

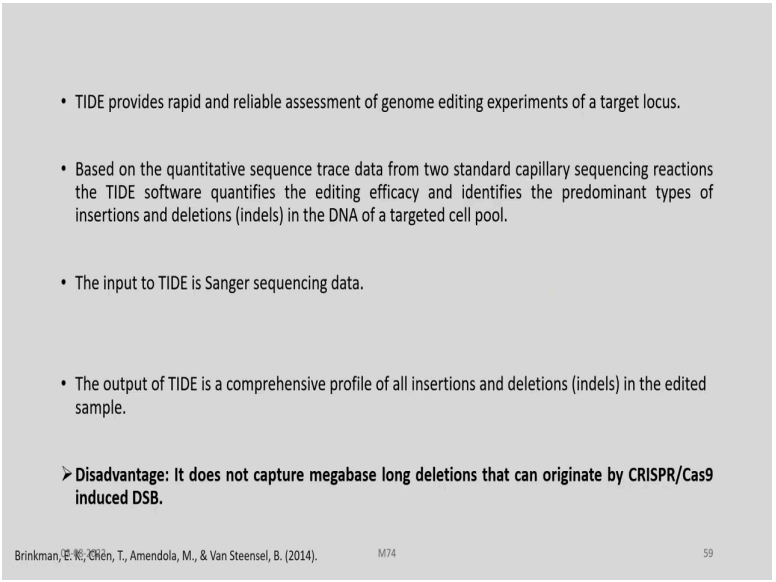
(Refer Slide Time: 16:30)



The users can give certain options. For example, he can enter a 20 nucleotide, 5' to 3' DNA character string representing the used single guide RNA sequence immediately upstream of the PAM sequence. PAM should not be included here. And then upload the chromatogram sequence files of respectively the control samples, example transfected without Cas9 or without the single guide RNA, and the test samples, example DNA of pool of cells are treated with both Cas9 and sgRNA.

So, this is the option for uploading the data. So, here we give the title plot or the name of the sample. And the guide sequence which has been already discussed around 20 nucleotides. And we give a control sample chromatogram. So, we browse and upload the files. And also, the test sample chromatogram which is saved as a separate file and then upload it.

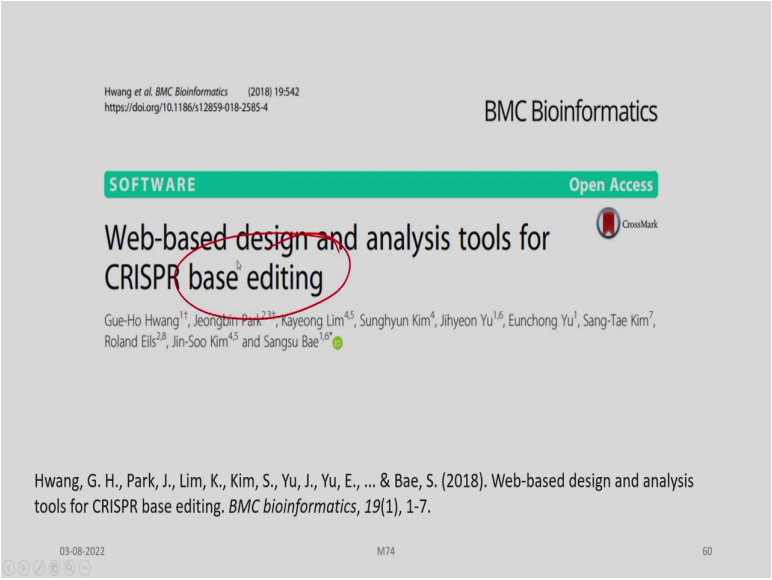
(Refer Slide Time: 17:41)



TIDE provides rapid and reliable assessment of genome editing experiments of a target locus. Based on the quantitative sequence trace data from two standard capillary sequencing reactions, the TIDE software quantifies the editing efficacy and identifies the predominant types of insertions and deletions in the DNA of a targeted cell pool.

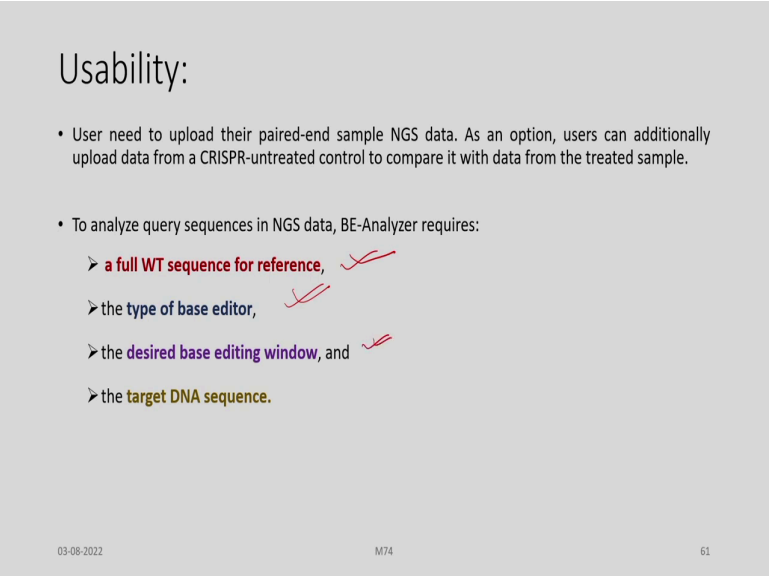
The input to TIDE is Sanger sequencing data as already told in the beginning. The output of TIDE is a comprehensive profile of all insertion and deletions, indels, in the edited sample. Disadvantage is that it does not capture megabase long deletions that can originate by CRISPR Cas9 induced double strand breaks. There are also tools which are developed for CRISPR based editing work.

(Refer Slide Time: 18:38)



So, this is one of the web-based design and analysis tools for CRISPR base editing developed by Hwang et al, and published in these journal BMC Bioinformatics in 2018.

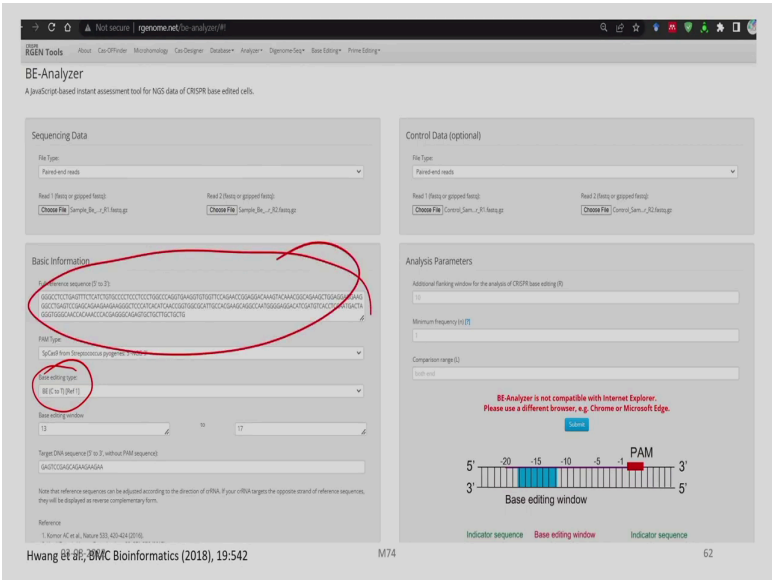
(Refer Slide Time: 18:47)



Here the user needs to upload their paired-end sample NGS data. As an option, users can additionally upload data from a CRISPR-untreated control to compare it with data from a treated sample. To analyze query sequences in NGS data, base editor analyzer requires the following: One, a full wild-type sequence for reference; two, the type of base editor to be

used or deployed; third, the desired base editing window and fourth, the target DNA sequence.

(Refer Slide Time: 19:30)



So, here you can visit this web base editor by visiting the site rgenome.net/be-analyzer. And here this is a JavaScript based instant assessment tool for NGS data of CRISPR base edited cells. So, in the options for sequencing data, you may select your file type from this drop-down menu. For example, here the type selected is paired-end reads.

You can also upload the files, the read one files in fastq format or the read two files and these are uploaded from pre-prepared files. Under the options for basic information, you can give full reference sequence 5' to 3' and you can choose the PAM type from the drop-down menu. In this case, the SpCas 9 from *Streptococcus pyogenes*, 5' NGG 3' is selected.

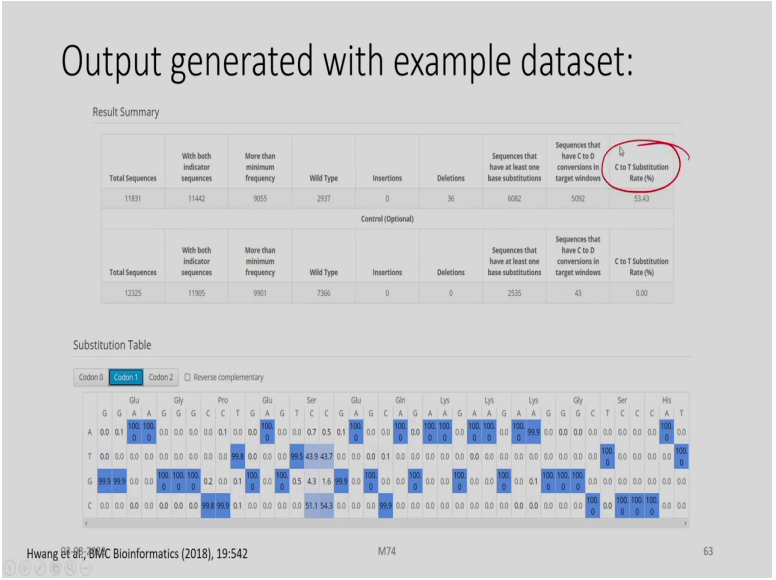
Then, the base editing type C to A is selected in this case. And the other type of base editing available can be explored from this dropdown menu. And the base editing window need to be defined from the minimum to maximum. And the target DNA sequence 5' to 3' is given here as a input without the PAM sequence.

You can also provide control data which is optional, and the file type can be chosen over here. See, as in the case of the sequencing data. And there are other parameters, like analysis parameters, where additional flanking window for the analysis of CRISPR base editing can

be defined in numerical terms. And the minimum frequency and the comparison range, whether it is from both end or single end can be given as a input.

We have to remember that BE-analyzer is not compatible with certain browsers like internet explorer. And it has to be used with either Chrome or Microsoft Edge. And you can see here the base editing window and the PAM sequence given here as a schematic. So, once all these parameters are fed into this webpage, you have to submit it for the webserver to compute. And give you the output.

(Refer Slide Time: 22:30)

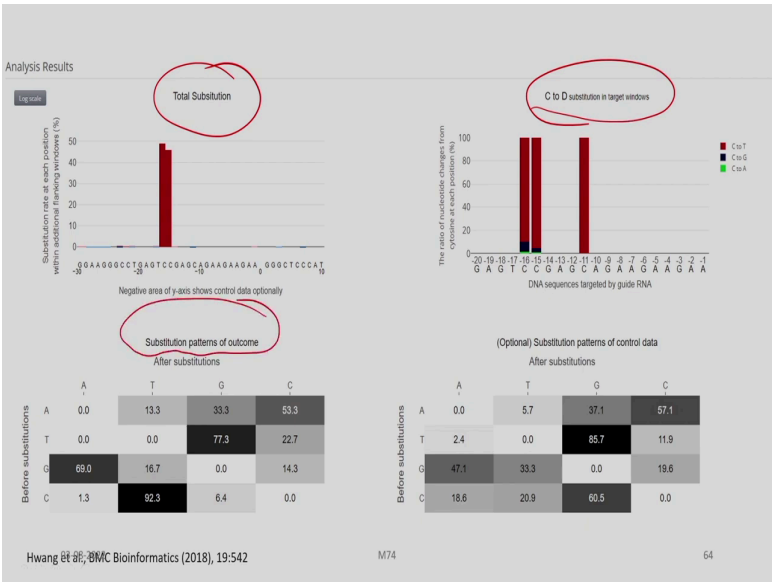


And the output generated with one of the examples that, for example, given in the earlier slide. So, you get a result summary which you need to use for planning your work. So, you are getting here around 11,800 total sequences. And with both indicator sequences which are roughly around 11400, and more than the minimum frequency. And you get the data about wild-type, and here the value of the insertions, 0 in this case.

And you are getting some idea about the number of deletions, 36. And sequences that have at least one base substitution, quite huge in this case around 6000 and sequences that have C to D conversions in the target window around 5000. And finally, you get a score in terms of percentage C to T substitution rate because we have chosen here the C to T base editing type.

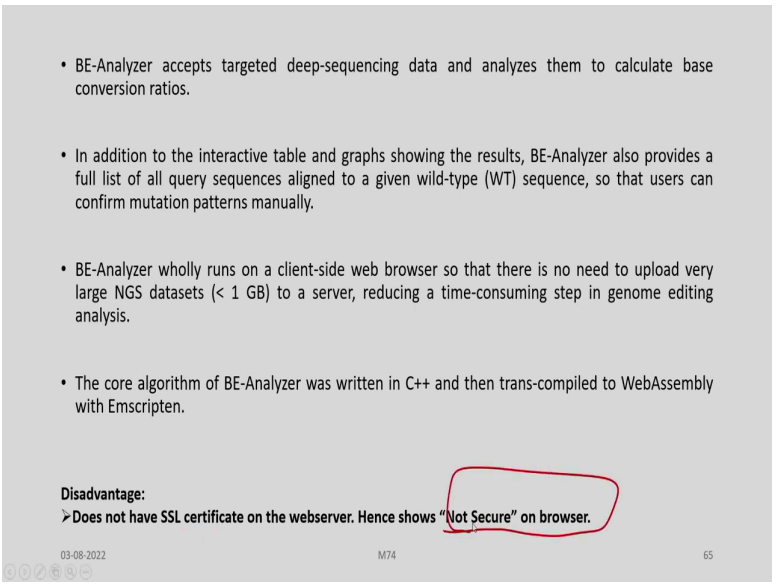
For example, here, we got around 53 percent. So, there is a substitution table also being displayed at the bottom of this output page. So, this output page looks like this when you give some input for a targeted based editing of a particular the sequence of your choice.

(Refer Slide Time: 24:15)



And you get certain analysis results in graphical form. So, you get here the total substitution, and then C to D substitution in the target window. And substitution patterns of outcome after the substitutions as well as the optional substitution patterns of control data after the substitutions.

(Refer Slide Time: 24:42)



Overall, BE-analyzer accepts targeted deep sequencing data and analyzes them to calculate base conversion ratios. In addition to the interactive table and graphs showing the results, BE-analyzer also provides a full list of all query sequences aligned to a given wild-type sequence, so that users can confirm mutation patterns manually.

BE-analyzer wholly runs on a client-side webs browser, so that there is no need to upload very large NGS datasets to a server, thereby this reduces a time-consuming step in genome editing analysis. The core algorithm of BE-analyzer was written in C++ and then trans-compiled to WebAssembly with Emscripten.

There is one disadvantage however, it does not have SSL certificate on the webserver. And hence when you open the web page, it will show it as ‘‘Not Secure’. So, with this we come to an end on the discussion on various bioinformatics and computational resources available for CRISPR Cas9 experiments.

We have discussed about the basic tools which I use for CRISPR Cas array scan as well as in detail about the various tools which are available for a single guide RNA design. We have also discussed about the various databases which are available. And also, on the web applications which are available for example, base editing.

Thank you for your patient hearing.