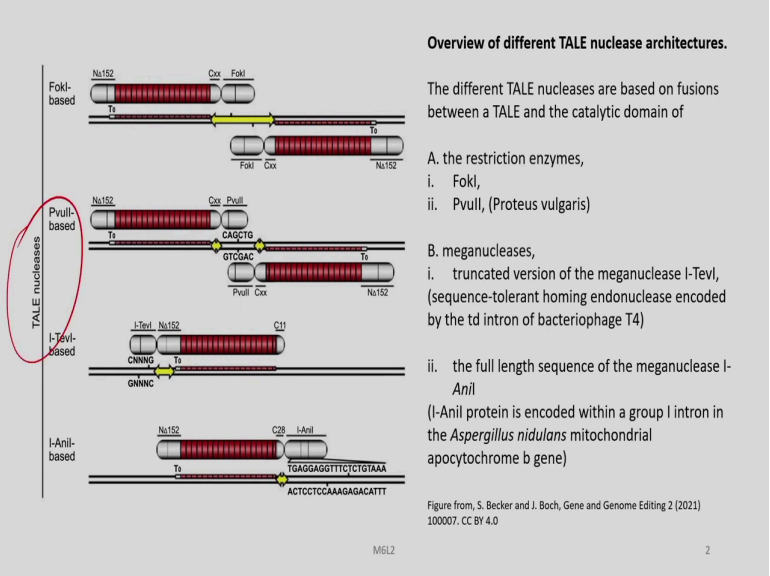
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 06
Transcription activator-like effector nuclease (TALEN) Technology
Lecture - 02
Design of TALEN for genome editing-PART A

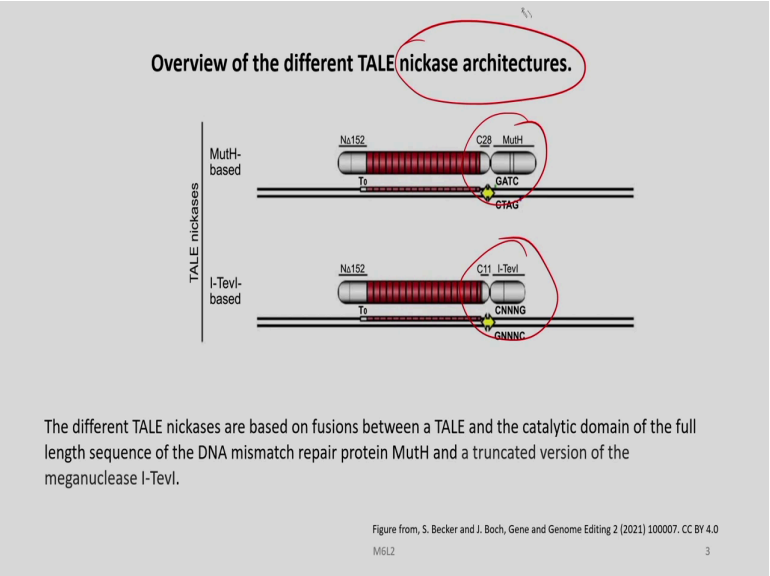
Welcome to my course on Genome Editing and Engineering; we were discussing about basics of TALEN in the last lecture. In this lecture under module 6, we will be discussing about the Design of TALEN for genome engineering, where we use some of the concepts we have learned in the last lecture for the design of TALEN.

(Refer Slide Time: 00:49)



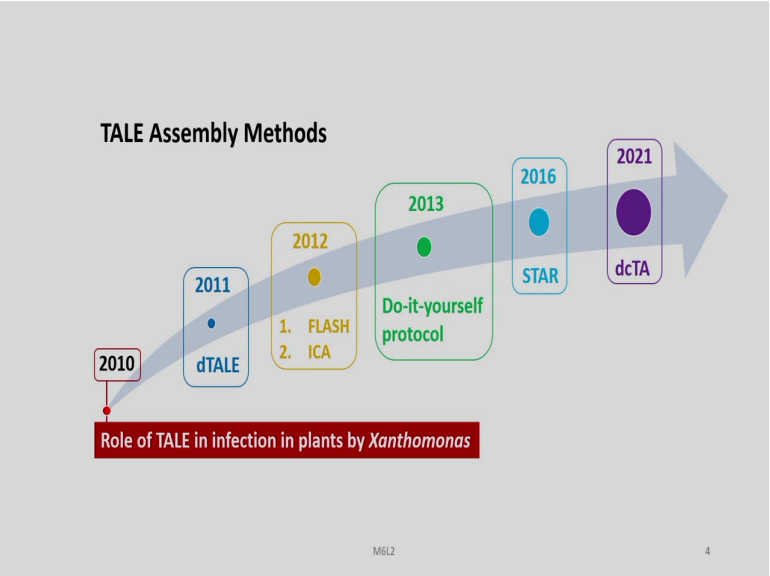
This is a simple overview of different TALE nuclease architectures and some of these are already known to you. In the last lecture we discussed how the TALE domains can be fused with various functional proteins and we can thereby design synthetic protein molecules having DNA specificities to carry out diverse functions. So, this is just to briefly show you about some of those capabilities once more.

(Refer Slide Time: 01:33)



So, whereby, we can create various nucleases, as well as nickases by fusing certain functionalities to the C terminus of the TALE proteins.

(Refer Slide Time: 01:44)

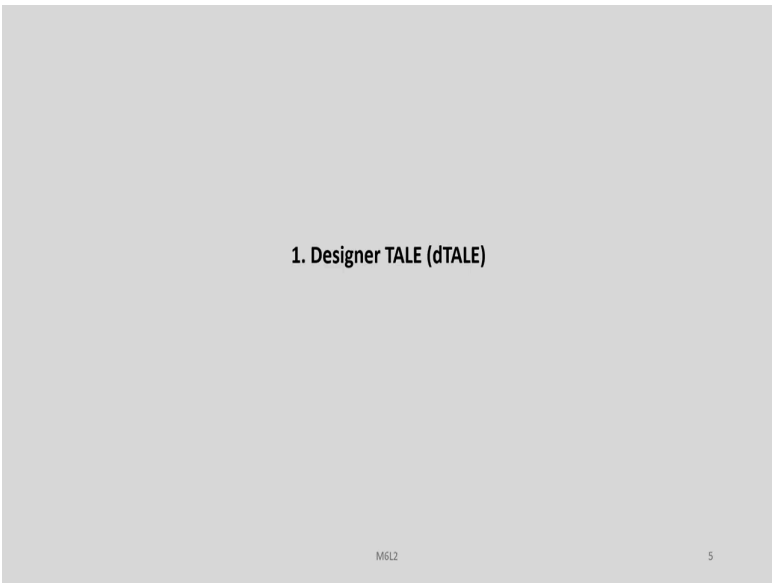


So, to obtain all those diverse functionalities, various scientists have developed numerous TALE fusion proteins over time. Since the first discovery of the role of TALE in infection in plants by *Xanthomonas* and the discovery of the principles by which the TALE recognizes DNA sequences, many developments have been done in this regard of engineering TALEs into functional proteins, which are nucleases as well as DNA modifying functions.

So, let us start with some of such functionalities; for example just one year after the discovery of the role of TALE in infection by *Xanthomonas* in plants in 2011, the development of detail took place and immediately next year 2 more technologies FLASH and ICA were developed. As well as in 2013, the Do it yourself protocol was developed, whereby the laboratories can develop these TALE fusion proteins on their own.

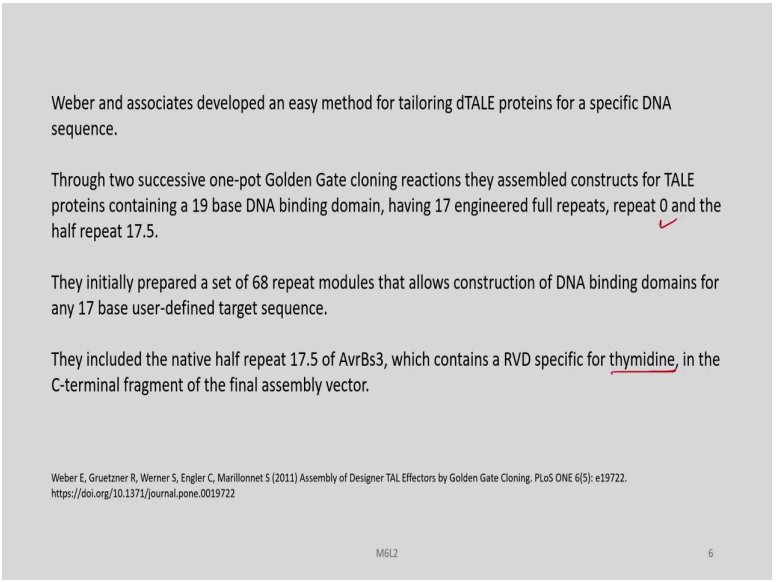
And in 2016, there was the development of STAR and just recently the development of dcTA. Besides these platforms, the numerous other platforms which has not been included here; we will be discussing only a few of these in today's lecture due to the paucity of time.

(Refer Slide Time: 03:44)



So, let us start with the first technology, the designer TALE or dTALE.

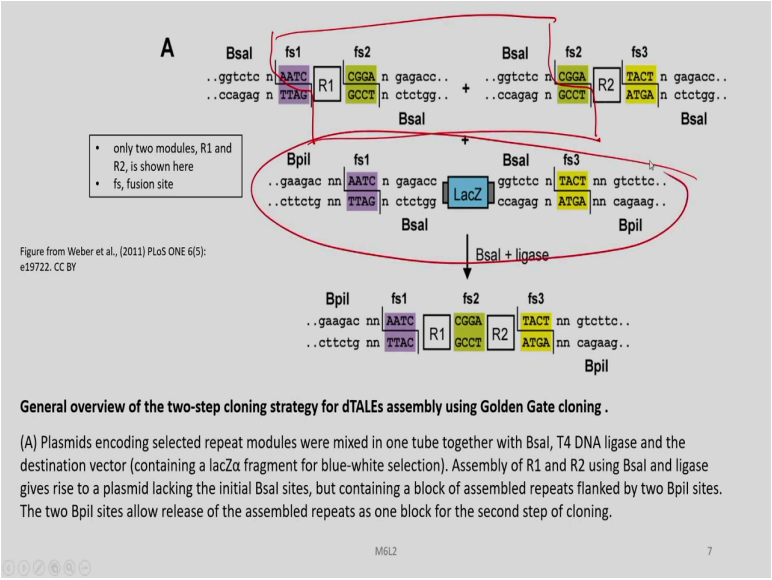
(Refer Slide Time: 03:45)



So, this was developed by Weber and associates and it was an easy method, which can be used for tailoring dTALE proteins for a specific DNA sequence. So, through 2 successive one-pot Golden Gate cloning reactions, they assembled constructs for TALE proteins containing a 19 base DNA binding domain, having 17 engineered full repeats and then repeat 0 and a half repeat 17.5; so, 17, 18 and total 19 repeats.

And initially they prepared a set of 68 repeat modules that allows construction of DNA binding domains for any 17 base user defined target sequences. They included the native half repeat of 17.5 of AvrBs3, which contains an RVD specific for thymidine, in the C terminal fragment of the final assembly vector.

(Refer Slide Time: 04:54)



So, let us see this figure a little bit closely. So, here only 2 modules, R1 and R2 are shown, but we can go for many modules at the same time. These fs represents fusion sites, ok. So, these concepts are very very important in understanding the development of this technology platform.

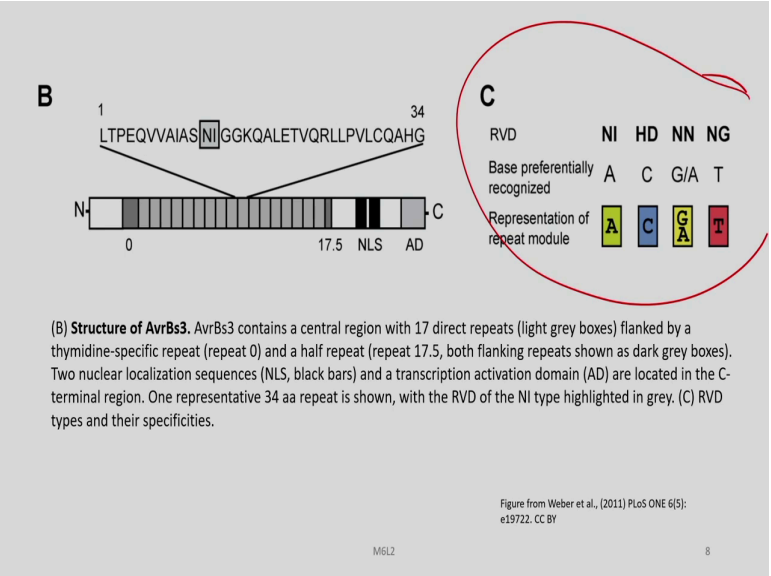
So, plasmids encoding selected repeat modules R1 and R2 were mixed in one tube together with BsaI, T4 DNA ligase and destination vector containing a lacZ fragment for blue-white selection. So, this is the one-pot reaction everything is put together into a single reaction vessel.

So, since these endonuclease Bsa1 sites are there; the Bsa will cleave it and the T4 DNA which is present in the reaction vessel, it will ligate the cut DNA. And this will be ligated into a destination vector, which contains a lacZ fragment that helps in the selection.

So, here we are just showing the example of 2 modules getting assembled and this assembly of the R1 and R2 using BsaI and ligase gives rise to a plasmid lacking the initial BsaI sites, but containing a block of assembled repeats flanked by 2 BpiI site. The 2 BpiI site allows the release of the assembled repeats as one block for the second step, step of cloning.

So, if you can see over here due to the BsaI restriction digestion. So, there is a cut over here and there is also a cut over here and this part is eliminated from the inclusion in the destination vector finally.

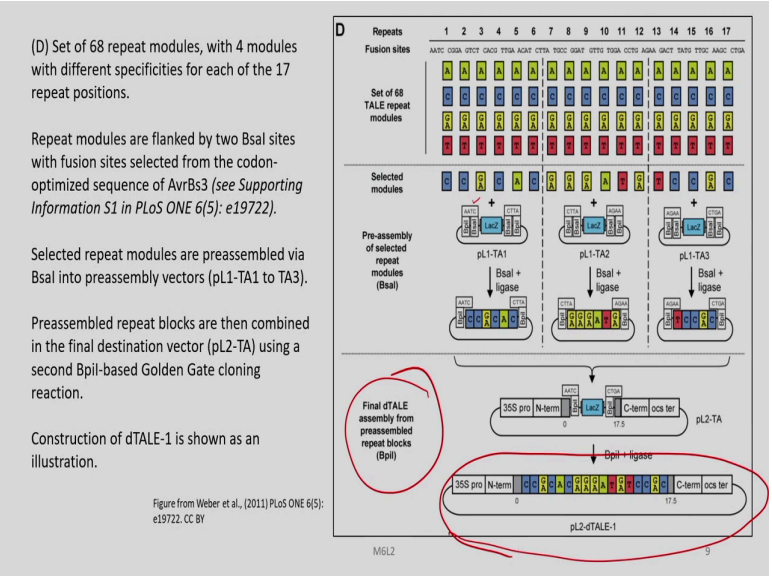
(Refer Slide Time: 07:21)



Now, here you can see the structure of AvrBs3, which contains a central region with 17 direct repeats. These are the grey boxes flanked by a thymidine specific repeat, which is repeat 0 and a half repeat, which is repeat 17.5; both flanking repeats shown as dark grey boxes. Then there are 2 nuclear localization sequences in a transcription activation domain about which we discussed yesterday located in the C terminal region.

Then one representative 34 amino acid repeat is shown here with the RVD of the NI type which is highlighted in grey and this is the codon binding preferences or base preferences as you can see over here.

(Refer Slide Time: 08:27)



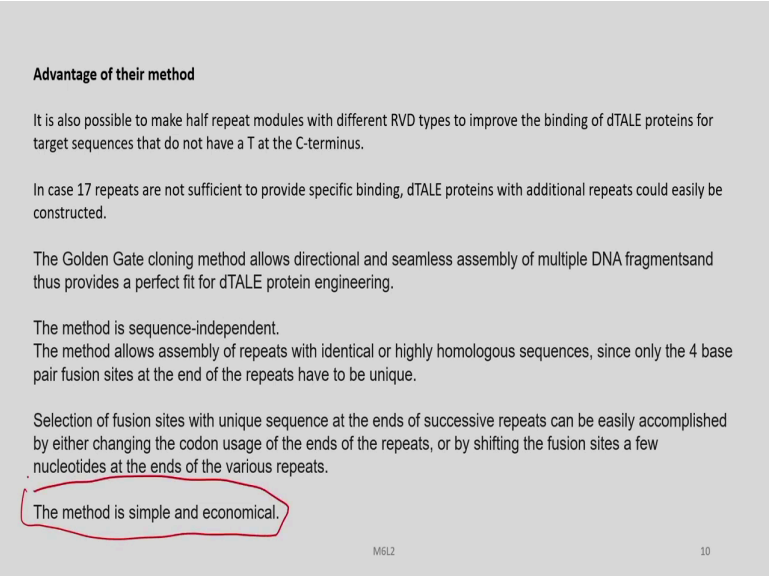
So, a set of 68 repeat modules, with 4 modules with different specificity for each of the 17 repeat positions were generated. These repeat modules are flanked by 2 Bsa1 sites with fusion sites selected from the codon optimized sequence of AvrBs3 and for more information you can visit these publications.

The selected repeat modules are preassembled via BsaI into preassembly vectors that we discussed just recently. So, these are the selected modules and then they are getting preassembled here and you can see here the BpiI site and BsaI site flanking both the regions and this is the LacZ, ok. So, this is the preassembly vector pL1 - TA1 and we also use pL1-TA2, pL1-TA3. So, we add BsaI plus ligase into the reaction vessel as we have told.

And as a result of putting all these together into one single reaction vessel, we have these repeat modules cloned into the destination vector. And here you can see there is no presence of any BsaI sequence anymore. So, the preassembled repeat blocks are then combined in the final destination vector pL2-TA using a second BpiI based golden cloning reaction.

So, we do the preassembly into the pL1 vectors initially and then we put all these together for the final dTALE assembly, and then in this step, we also use this endonuclease BpiI and ligase; here we use BsaI and ligase in the first reaction and then this finally, will give us the vector which is loaded with the TALE protein molecule sequences.

(Refer Slide Time: 11:06)

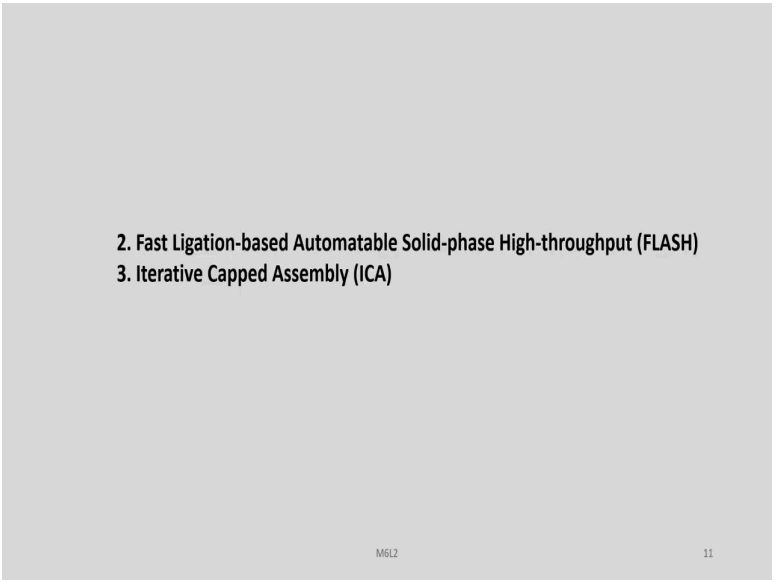


What is the advantage of this method? With this method, it is possible to make half-repeat modules with different RVD types to improve the binding of dTALE proteins for target sequences that do not have a T at the C terminus. In case 17 repeats are not sufficient to provide specific binding, dTALE proteins with additional repeats could easily be constructed.

So, you can add more and more of the repeat modules. The Golden Gate cloning method allows directional and seamless assembly of multiple DNA fragments and thus provides a perfect fit for dTALE protein engineering. The method is sequence-independent. The method allows assembly of repeats with identical or highly-homologous sequences since only the 4 base pair fusion sites at the end of the repeats have to be unique.

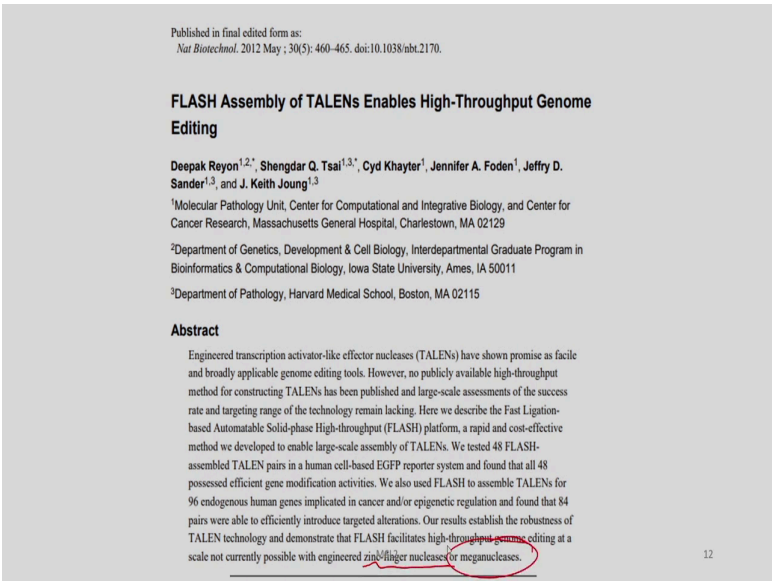
Selection of fusion sites with unique sequences at the ends of successive repeats can be easily accomplished by either changing the codon usage of the ends of the repeats or by shifting the fusion sites a few nucleotides at the ends of the various repeats. Overall the method is simple and economical as claimed by the developers.

(Refer Slide Time: 12:18)



Over time in around 2012 as I have already told 2 more technologies were developed; one is the FLASH, which stands for fast ligation based automatable solid phase high throughput and the 3rd one is the iterative capped assembly.

(Refer Slide Time: 12:40)


Published in final edited form as:
Nat Biotechnol. 2012 May ; 30(5): 460–465. doi:10.1038/nbt.2170.

FLASH Assembly of TALENs Enables High-Throughput Genome Editing

Deepak Reyon[1,2,*], **Shengdar Q. Tsai**[1,3,*], **Cyd Khayter**[1], **Jennifer A. Foden**[1], **Jeffry D. Sander**[1,3], and **J. Keith Joung**[1,3]

[1]Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129

[2]Department of Genetics, Development & Cell Biology, Interdepartmental Graduate Program in Bioinformatics & Computational Biology, Iowa State University, Ames, IA 50011

[3]Department of Pathology, Harvard Medical School, Boston, MA 02115

Abstract

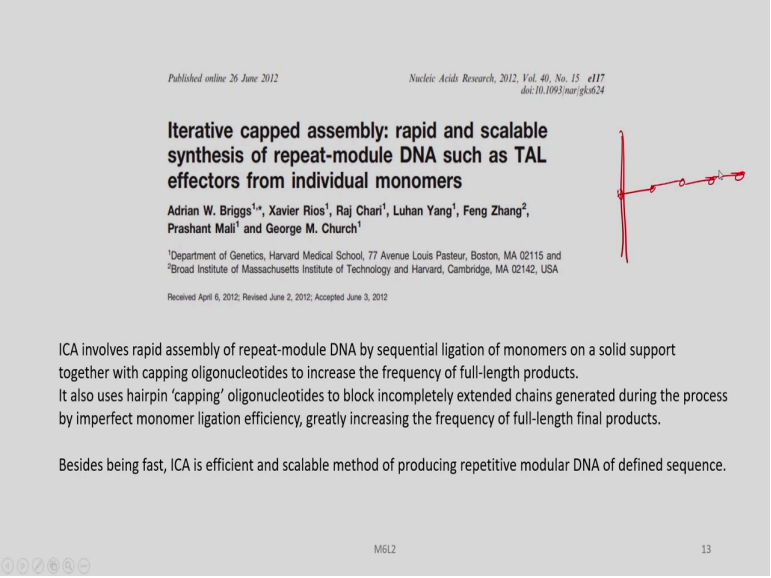
Engineered transcription activator-like effector nucleases (TALENs) have shown promise as facile and broadly applicable genome editing tools. However, no publicly available high-throughput method for constructing TALENs has been published and large-scale assessments of the success rate and targeting range of the technology remain lacking. Here we describe the Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) platform, a rapid and cost-effective method we developed to enable large-scale assembly of TALENs. We tested 48 FLASH-assembled TALEN pairs in a human cell-based EGFP reporter system and found that all 48 possessed efficient gene modification activities. We also used FLASH to assemble TALENs for 96 endogenous human genes implicated in cancer and/or epigenetic regulation and found that 84 pairs were able to efficiently introduce targeted alterations. Our results establish the robustness of TALEN technology and demonstrate that FLASH facilitates high-throughput genome editing at a scale not currently possible with engineered zinc-finger nucleases or meganucleases.

So, in the FLASH assembly it has been reported that it enables high throughput genome editing. So, they describe the fast ligation based automatable solid high throughput platform as a rapid and cost effective method, which was used to develop large scale assembly of TALENs. And with this technology, they tested 48 FLASH assembled TALEN pairs in a

human cell based EGFP reporter system and found that all 48 possessed efficient gene modification activities.

And they further used these FLASH to assemble TALENs for 96 endogenous human genes implicated in cancer and/or epigenetic regulation and found that 84 pairs were able to efficiently introduce targeted alterations. We claim that this method is robust and it facilitates high throughput genome editing at a scale of at that time which was not possible with other technologies that zinc finger nucleases and mega-nucleases.

(Refer Slide Time: 14:04)



In this course we are not discussing much about mega-nucleases, because it has become a little bit obsolete technology; but we are of course, referencing it in certain cases wherever required. The other technology developed by Briggs and his coworkers is the ICA or iterative capped assembly, it is also claimed to be very rapid and scalable.

And ICA as the name suggest is a rapid assembly of repeat module DNA by sequential ligation of monomers on a solid support. So, they will be using a solid support and on to this solid support, they will be sequentially ligating the repeat modules.

And while doing so, they do capping of the oligonucleotides to increase the frequency of full-length products. And they use hairpin capping oligonucleotides to block incompletely extended chains generated during the process by imperfect monomer ligation efficiency, greatly increasing the frequency of full-length final products. Besides being fast, ICA is also

claimed to be an efficient and scalable method of producing repetitive modular DNA of defined sequences.

(Refer Slide Time: 15:43)

Next let us discuss about one simple technology for TALEN or TALE based protein assembly.

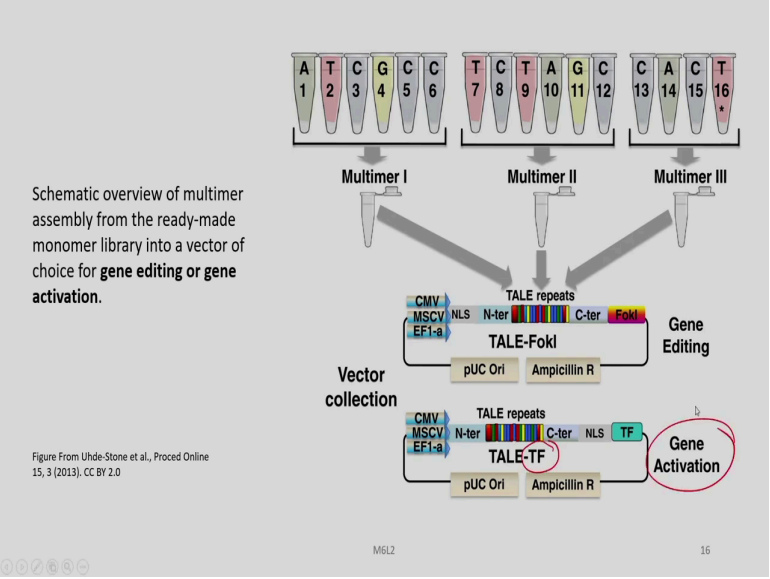
(Refer Slide Time: 16:01)

It is simply called as do it yourself protocol, which was developed by Biao Lu and his associates Uhde-Stone. So, this technology relies on a simple design of TALEs and coupled

with high binding predictability and specificity. This simple technology can be easily used by any standard molecular biology laboratory with only the limited resources.

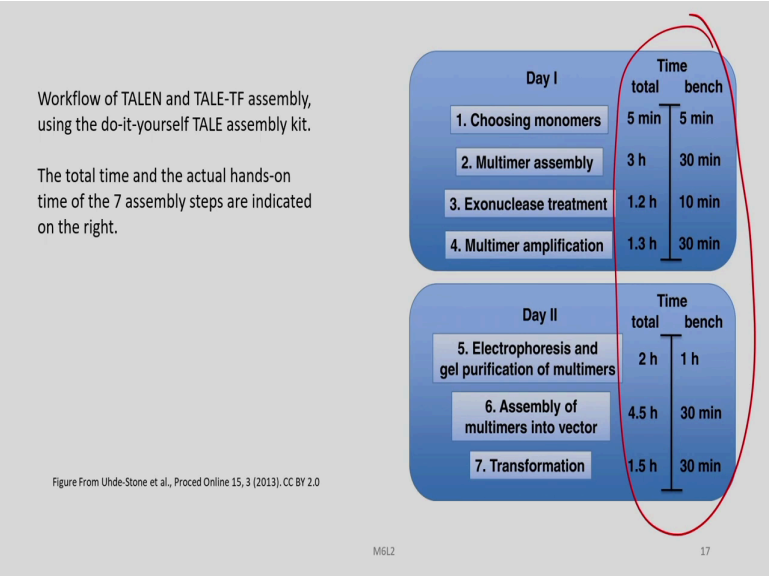
So, custom TALE assembly at that time was costly and it was only available to few research centers. So, this do it yourself approach, solved this kind of problems and made TALEN assembly a very simple method, which can be done in any laboratory with minimum instrumental setup.

(Refer Slide Time: 17:08)



So, this is a schematic overview of multimer assembly. So, you have, say for example, here 3 multimer, 1, 2, 3 from the readymade monomer library into a vector of choice. So, this is the vector of choice. So, in certain first cases we are interested to make a nucleases TALEN. So, we here we are attaching it to FokI. In the second case, we want to attach these tale modules to transcription factors for gene activation. So, there are many other such functionalities which can be added by this simple do-it-yourself method.

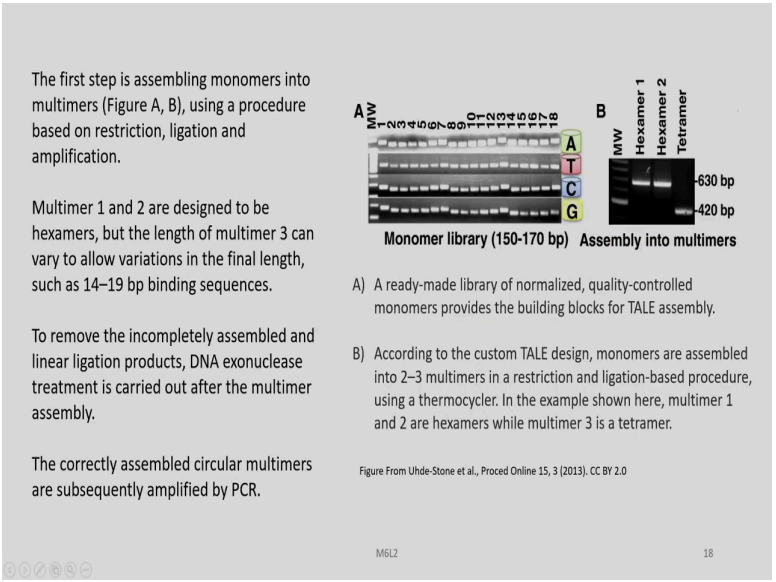
(Refer Slide Time: 18:06)



So, let us see how we can proceed in this particular method. So, this is the workflow of this do-it-yourself assembly comprising of 2 days and you can see the time taken for each step on the right side, ok. So, the first step is choosing the monomers, and then after the monomers are chosen; we go for the multimer assembly. Then there is exonuclease treatment and the fourth step is the multimer amplification. With these, the first day of work is all over.

So, next day we start with electrophoresis and gel purification of the multimers and these multimers we assemble into our targeted vector and then once that is done, we go for transformation.

(Refer Slide Time: 19:06)

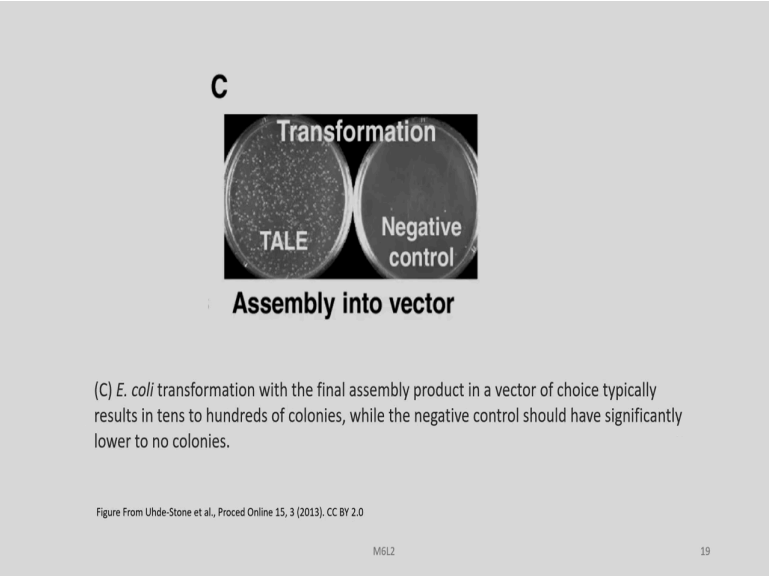


So, the first step is assembling monomers into multimers, ok. So, from monomer to multimer what is the procedure? So, you can see here in figure A and B, we are trying to show you this process and this process is based on the use of restriction, ligation, and amplification.

So, multimers 1 and 2 are designed to be hexamers; for example, here this is hexamer and the length of multimer 3 can vary to allow variations in the final length. So, here it is being taken as a tetramer. To remove the incompletely assembled and linear ligation products, DNA exonuclease treatment is carried out after the multimer assembly.

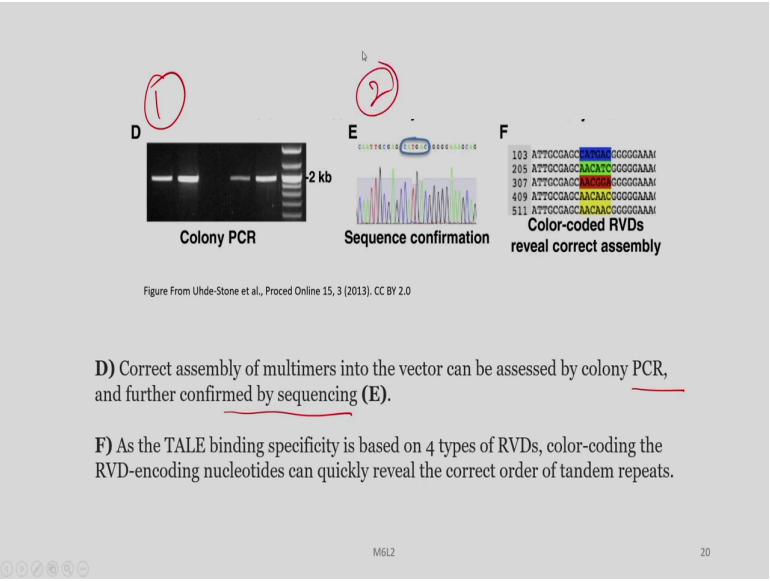
The correctly assembled circular multimers are subsequently amplified by PCR.

(Refer Slide Time: 20:11)



And then these it is used for E coli transformation in a vector of choice, typically it would result in tens to hundreds of colonies and the negative controls will have a significantly lower or no colonies.

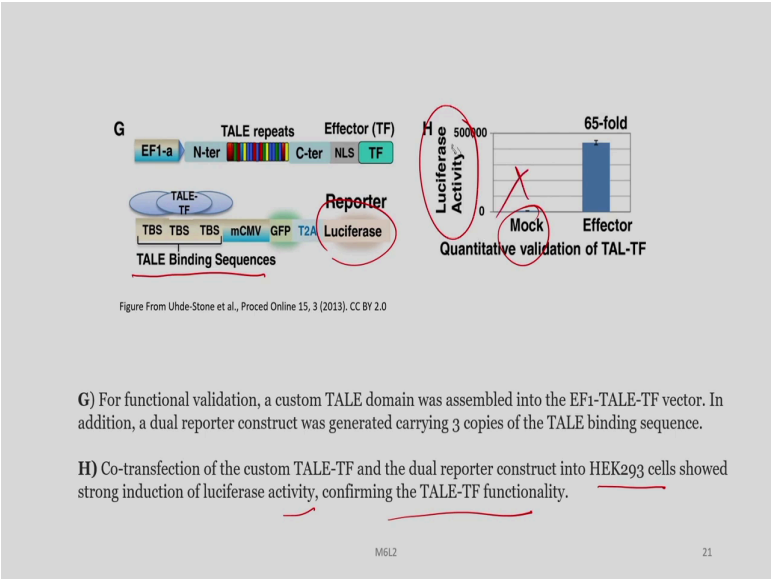
(Refer Slide Time: 20:28)



Then we go for assessing the correctness of the assembly; we go for colony PCR, we pick up the colony one by one and then we go for colony PCR, this is a quality assurance step.

So, the correct assembly of these multimers into the vectors are assessed by in the colony PCR in one instance and also it is confirmed by sequencing in another instance. And you can see in figure F as the TALE bending specificity is based on 4 types of RVDs, color coding the RVD encoding nucleotides can quickly reveal the correct order of the tandem repeats.

(Refer Slide Time: 21:19)

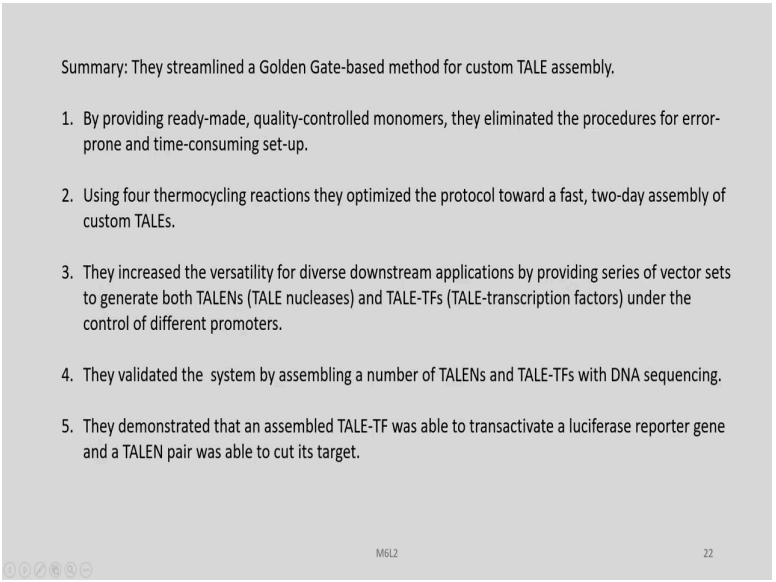


So, for functional validation, this is the validation only at the cloning level that the multimer assembly is proper, so we do the PCR as well as the sequencing. Now, for functional validation of these TALE constructs, a custom TALE domain is assembled into the EF1-TALE-TF vector. So, this is the EF1-TALE-TF vector and a dual reporter construct was generated carrying 3 copies of the TALE binding sequences, ok.

So, these are the TALE binding sequences in triplets. Then we go for co-transfection of these custom TALE-TF and dual reporter construct into HEK 293 cells, which showed strong induction of luciferase activity, confirming the TALE transcription functionality.

So, if you look into the mock or the control, there is no any luciferase activity; but in the transfected one, you have this luciferase activity, which is around 65-fold high. So, this confirms that the system is working fine.

(Refer Slide Time: 22:47)

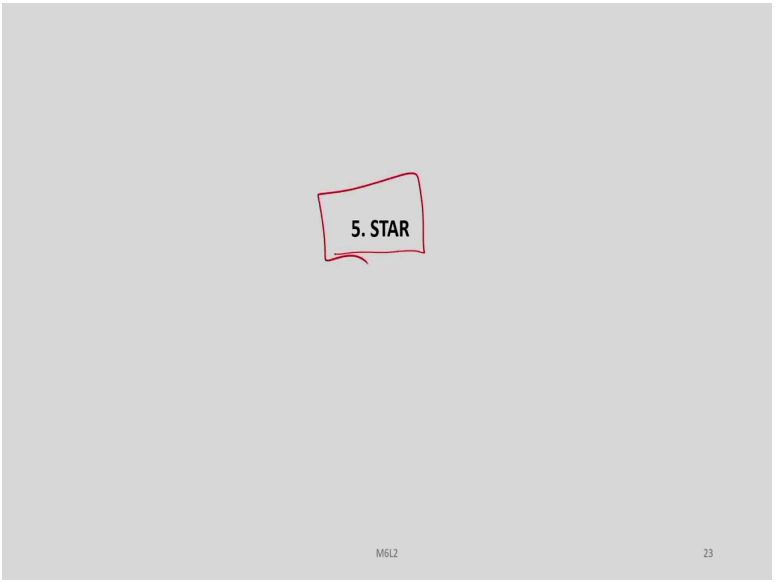


To sum up, in this procedure, they streamlined a Golden Gate-based method for custom TALE assembly.

By providing readymade quality controlled monomers, they eliminated the procedure for error-prone and time-consuming setup. Using 4 thermocycling reactions they optimized the protocol towards a fast, 2-day assembly of custom TALEs. And they increase the versatility for diverse downstream applications by providing a series of vector sets to generate both TALENs and TALE transcription factors under the control of different promoters.

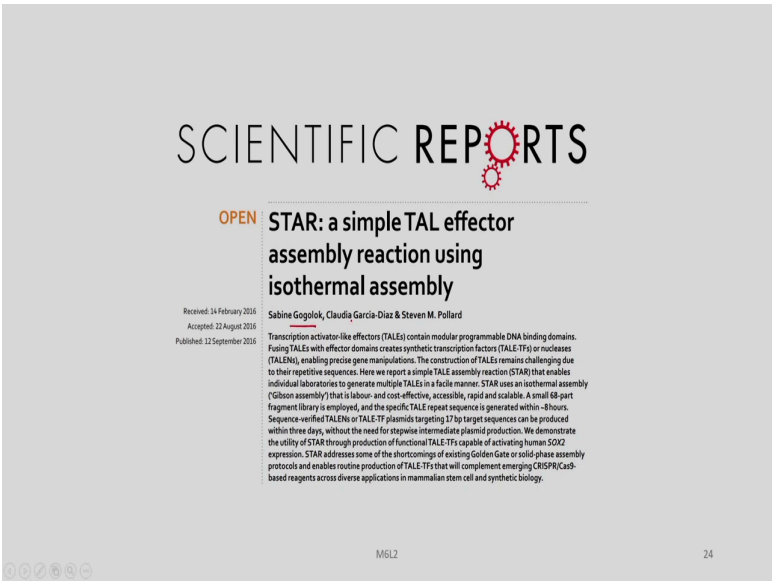
They validated the system by assembling a number of TALENs and TALE transcription factors with DNA sequencing. And demonstrated that an assembled TALE transcription factor was able to transactivate a luciferase reporter gene and a TALEN pair was able to cut at it is targets.

(Refer Slide Time: 23:46)



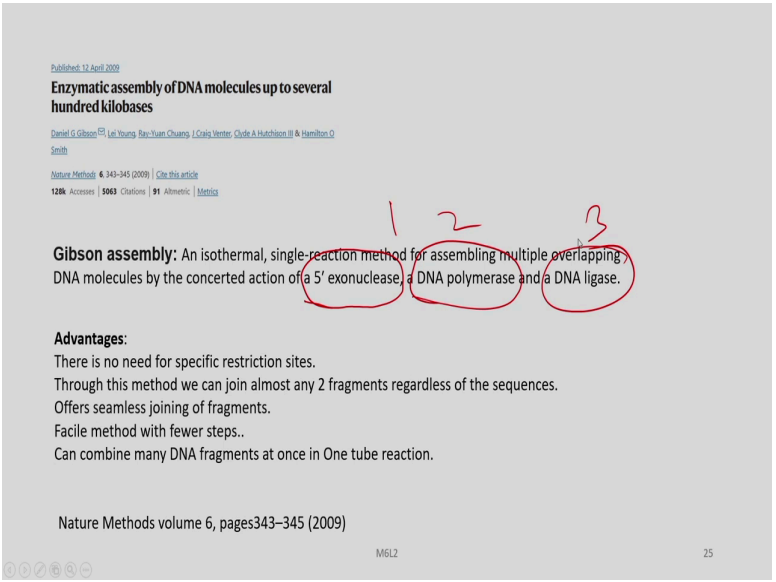
Let us now discuss about the 5th technology, the STAR assembly method.

(Refer Slide Time: 23:52)



A STAR stands for a simple TAL effector assembly reaction and this uses isothermal assembly and this was developed by Gogolok and his associates.

(Refer Slide Time: 24:13)



So, before we go into the details of the simple TAL effector assembly reaction; let us try to learn about the Gibson assembly, which we have referred to in the past in one or few of the assembly methods as well.

So, this paper published by Gibson and his team is about a method of enzymatic assembly of DNA molecules up to several hundred kilobases. So, we can have small small DNA fragments and then we can concatenate all these small DNA fragments into very large DNA fragments. So, it is believed that we can even go for constructing large chromosomes by this enzymatic assembly of the DNA molecules.

Gibson assembly is an isothermal single reaction method for assembling multiple overlapping DNA molecules by the concerted action of a 5 prime exonuclease number 1, a DNA polymerase number 2, and a DNA ligase. So, 3 enzymes are essential for carrying out this isothermal reaction; a 5 prime exonuclease, a DNA polymerase, and a DNA ligase.

Now, what are the advantages of the Gibson assembly? There is no need for specific restriction sites; through this method we can join almost any 2 fragments regardless of the sequences. It offers seamless joining of fragments. It is a facile method, I mean very simple with fewer steps and can combine many DNA fragments at once in one tube reaction, single step.

(Refer Slide Time: 26:10)

So, let us see how this is being done. So, these are say 2 DNA fragments to be joined; the green one and the black one, these are double-stranded. So, now, what we do? To one of these, we may do it to the green one or the black one. So, here we add an overlap to the black one by a PCR reaction and this overlap is with the green spaces of the DNA molecule.

So, this is the amount of overlap and this overlap is introduced with the help of a PCR reaction. Now, we add these to the Gibson assembly master matrix containing all the 3 enzymes that we discussed; here the exonuclease will chew back the 5 prime ends, ok.

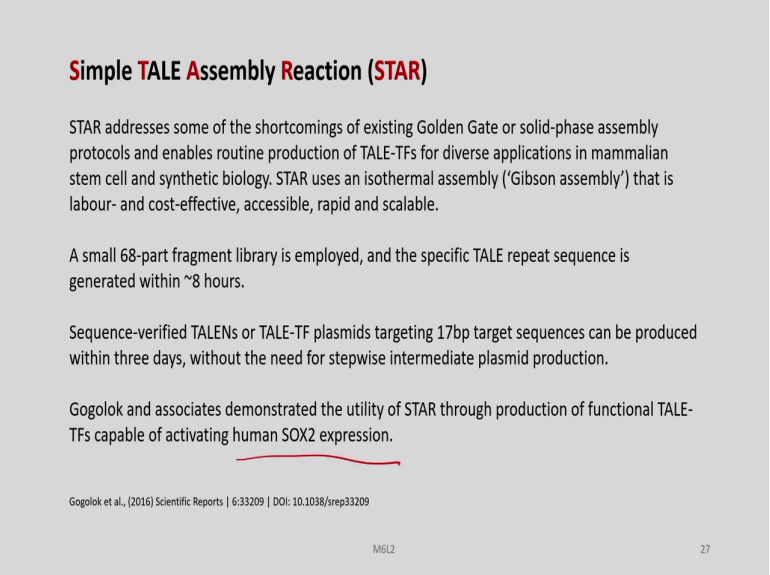
And then these fragments will anneal due to the complementarity at this point and once this is done, these gaps are filled up or closed by the action of DNA polymerase. And then finally, these small leaks which cannot be sealed by DNA polymerase are sealed by the action of DNA ligase. So, you require 3 enzymes here as already told to you; first one you need a 5 prime exonuclease, second you need a DNA polymerase, third one you need a DNA ligase and these are enzymes easily available.

And you need other enzymes for example, Taq polymerases to carry out the PCR reaction. So, for carrying out Gibson assembly, you need these many things: at least 4 enzymes and then DNA fragments and so on. And these are today possible to be done in any kind of simple molecular biology laboratory, which have a PCR and simple hot water bath and other incubation facilities.

So, this is just an example by which we have a seamless connection of 2 DNA molecules over here. We can use these to join not only just two but several such DNA molecules as shown here by some color coded fragments as a kind of a representation. So, technically speaking here we recess the DNA fragments yielding the single stranded DNA overhangs, that anneal the specifically and then these are joined covalently.

And Gibson assembly can be used to seamlessly construct synthetic and natural genes; genetic pathways and entire genomes and could be a useful molecular engineering tool.

(Refer Slide Time: 29:25)

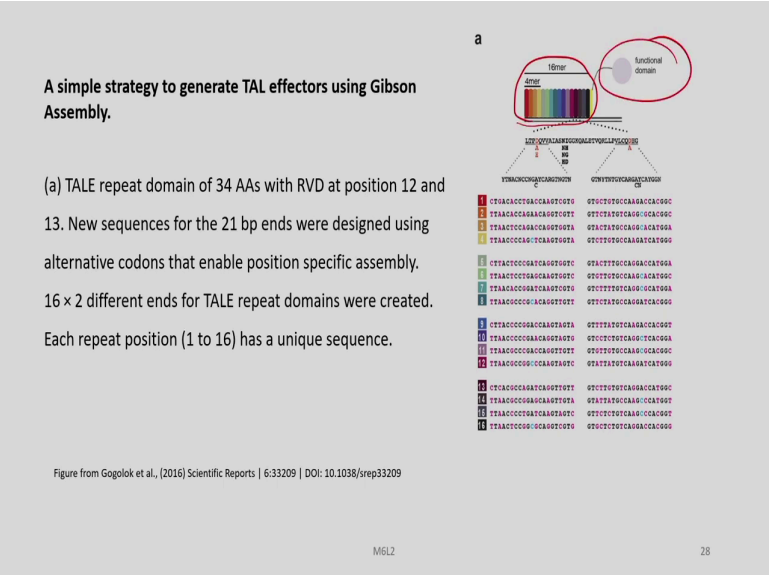


So, this method we are going to use in the STAR platform or the simple TALE assembly reaction. STAR addresses some of the shortcomings of the existing Golden Gate or solid phase assembly protocols and enables routine production of TALE transcription factors for diverse applications in mammalian stem cells and synthetic biology.

It uses an isothermal assembly, Gibson assembly which we just discussed; that is labour and cost effective, accessible, rapid, scalable, and very simple. A small 64-part fragment library is employed and the specific TALE repeat sequence is generated within 8 hours. Sequence-verified TALENs or TALE transcription factor plasmids targeting 17 base pair target sequences can be produced within 3 days, without the need for stepwise intermediate plasmid production.

Gogolok and associates demonstrated the utility of the STAR through production of functional TALE transcription factors capable of activating human SOX2 expression.

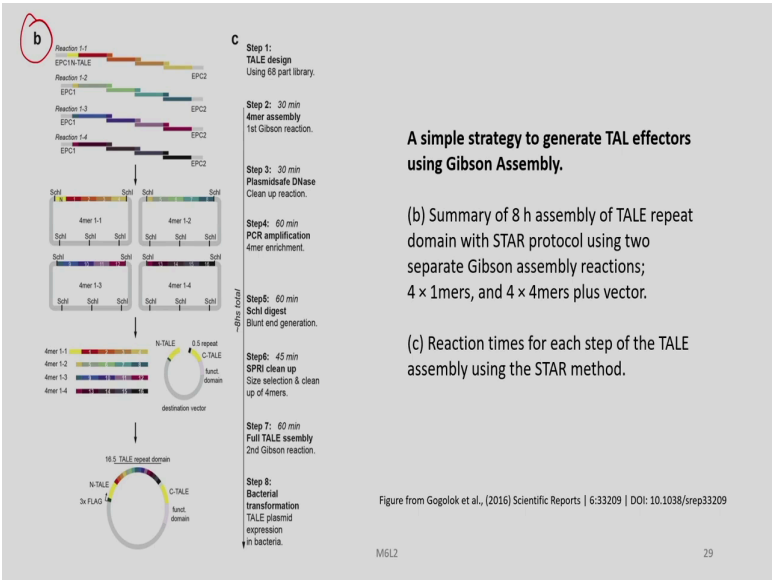
(Refer Slide Time: 30:32)



So, this is a simple strategy to generate TAL effectors using Gibson assembly. So, here you can see TALE repeat domains of 34 amino acids each with RVD positions at 12 and 13.

And new sequences for the 21 base pair ends were designed using alternative codons that enable position-specific assembly.16 into 2, 32 different ends of TALE repeat domains were created; each repeat position has a unique sequence. And this is the TALE domain and this is the functional domain and we we need to create lot of variations in this TALE domain, for which we are using this strategy.

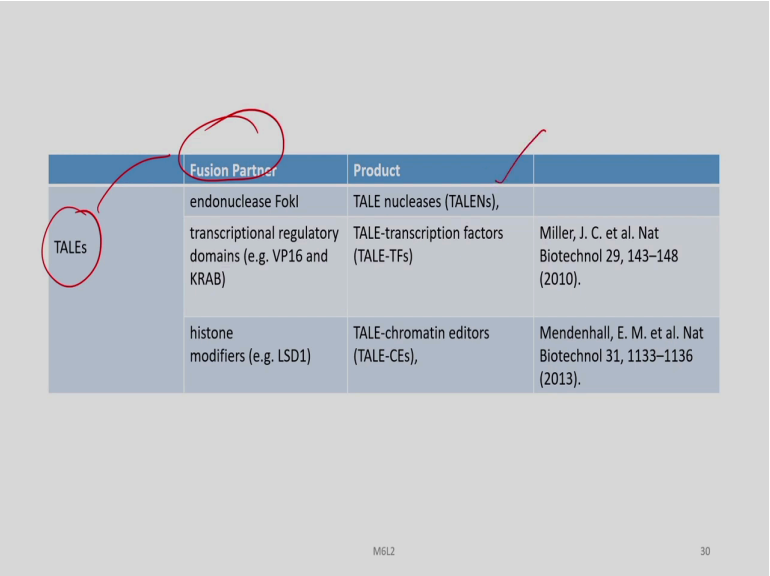
(Refer Slide Time: 31:32)



So, there are various steps in this entire process; the first step is TALE design using the 68-part library and then step 2 is a tetramer assembly, this is the first Gibson reaction. So, here you can see the first Gibson reaction taking place and the second step is a plasmid-safe DNase clean-up reaction and the next reaction is the PCR amplification followed by SchI digest blunt end generation.

And then next step is SPRI clean up, size selection and cleanup of the 4mers or tetramers. The full tale assembly, which involves the second Gibson reaction, and finally, the last step is the bacterial transformation, TALE plasmid expression in bacteria. So, in (b) you can see the summary of 8-hour assembly of TALE repeats domains with STAR protocol using 2 separate Gibson assembly reactions; 4 into 1mers, and 4 into 4mers plus the vector. And in (c), we have the description of the various steps.

(Refer Slide Time: 33:11)



So, using these technologies, here we have a functional domain and a TALE domain. So, depending on the variability of these functional domains, we may have different kind of products; this is a TALE domain and this is the fusion partner or the functional domain. So, if you use endonuclease FokI, it is known to you; it will be TALEN in the way ZF fused to nuclease is ZFN.

Then if we have transcriptional regulatory domains, you have we have TALE transcription factors or TALE-TFs. And then we may have also certain functional domains, which modify the DNA or modify proteins. So, we may have histone modifiers like LSD1. So, these will be TALE chromatin editors or TALE-CEs and these are some of the references, where you can find details of these kind of fusion products.

So, thank you for your patient hearing; we will be discussing more about the TALE design in the next part of this lecture.