Universidade de Lisboa

Faculdade de Farmácia

SELECTION OF IgG SYNTHETIC LIBRARIES BY CRISPR INDUCED VARIABILITY

Jéssica Alexandra de Carvalho Nereu

Dissertação orientada pelo Professor Doutor João Gonçalves.

Mestrado em Ciências Biofarmacêuticas

2021

Universidade de Lisboa

Faculdade de Farmácia

SELECTION OF IgG SYNTHETIC LIBRARIES BY CRISPR INDUCED VARIABILITY

Jéssica Alexandra de Carvalho Nereu

Dissertação orientada pelo Professor Doutor João Gonçalves.

Mestrado em Ciências Biofarmacêuticas

2021

AGRADECIMENTOS

I would like to acknowledge several people that helped me in many different ways during the course of this project. The following acknowledgments will be addressed in Portuguese:

Para que este trabalho se pudesse desenvolver é importante agradecer a várias pessoas que não só o tornaram possível, bem como o tornaram num importante marco da minha vida.

Gostaria assim de agradecer em primeiro lugar ao Dr. João Gonçalves pela simpatia, por acreditar sempre nas minhas capacidades e por ter sido um excelente orientador, sabendo me guiar e me conceder, na mesma medida, o espaço para que pudesse desenvolver um trabalho autónomo.

Aos meus colegas do iMed.ULisboa - Miguel, Inês, Margarida, Rita, Pedro, Tatiana, Vera e Luciana pela disposição que sempre demonstraram para me ajudar, pelo carinho e sobretudo por terem feito do tempo que passei no departamento uma experiência que recordarei sempre de forma agradável; porque é muito melhor fazer o que se gosta num sítio com pessoas das quais se gosta. À Paula pelo apoio, aprendizagem e disponibilidade sempre que necessário.

À Joana Ministro e Joana Oliveira dos laboratórios da Technophage pela ajuda prestada durante o meu trabalho.

A toda a minha família, mas em especial aos meus pais e irmãos, por me apoiarem incondicionalmente, tentando constantemente fazer com que os meus sonhos e ambições sejam possíveis e me darem as ferramentas para tal; por me incentivarem e mostrarem que se encararmos os desafios com otimismo seremos bem sucedidos.

A todos aqueles que sempre me apoiam, me ouvem e me animam quando é necessário: aos meus amigos, porque o bem-estar pessoal é fundamental para o sucesso académico/profissional.

A todos, um sincero obrigada!

LIST OF ABBREVIATIONS

293-F	FreeStyle 293-F
AA	Amino acid
alt-NHEJ	Alternative nonhomologous end-joining
BSA	Bovine Serum Albumin
C3orf17	Human chromosome 3 open reading frame 17
Cas9	CRISPR associated protein 9
CD20	Cluster of differentiation 20
CDK	cyclin-dependent kinase
CDR	Complementarity determining region
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSR	Class switch recombination
DMD	Duchenne Muscular Dystrophy
DMEM	Dulbecco Modified Eagle's Medium
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
DSB	DNA double strand break
dsDNA	Double-strand DNA
EFS	EF-1 alpha short promoter
EGFP	Enhanced green fluorescent protein
Fab	Fragment antigen binding
FACS	Fluorescence activated cell sorting
FBS	Fetal Bovine Serum
FC	Flow cytometry
Fc	Fragment crystallizable
FDA	Food and Drug Administration

Gag	Group specific antigen polyprotein
GFP	Green fluorescent protein
gRNA	Guide RNA
HA	Hemagglutinin tag
HC	Heavy chain
HEK293T	Human embryonic kidney 293T
HHS	Health and Human Service
HR	Homologous recombination
HRP	Horseradish peroxidase
HT1	Hereditary Tyrosinemia type I
hUBC	Human ubiquitin
HuCAL	Human combinatorial antibody library
HV	Hypervariable region
Ig	Immunoglobulin
IgG	Immunoglobulin G
Indels	Small insertions and/or deletions
IRES	Internal ribosome entry site
IVC	*In vitro* compartmentalization
JKT	Jurkat
LC	Light chain
mAb	Monoclonal antibody
MMEJ	Microhomology-mediated end joining
NGS	Next-Generation Sequencing
NHEJ	Non-homologous end joining
NIH	National Institute of Health
P2A	Porcine teschovirus 2A sequence
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PAR	poly-ADP-ribose

PBS	Phosphate-buffered saline solution
PCR	Polymerase chain reaction
PE	Phycoerythrin dye
PE-Cy7	Phycoerythrin and cyanine dye
REC	Recognition domain
Rev	Regulator of virus protein expression
RIPA	Radio-Immunoprecipitation Assay
RNA	Ribonucleic acid
RPMI	Roswell Park memorial institute medium
RSS	Recombination signal sequences
SCD	Sickle Cell Disease
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
SH	Somatic Hypermutation
ssDNA	Single-strand DNA
TALENs	Transcription activator-like effector nucleases
Tc	Cytotoxic T lymphocytes
Th	Helper T lymphocytes
TNFα	Tumor necrosis factor
tracrRNA	Trans-activating crRNA
V(D)J	Variable, Diversity and Joining recombination
VH	Variable heavy chain domain
VL	Variable light chain domain
VSV-G	Vesicular stomatitis virus, glycoprotein-G
WB	Western blot
ZFNs	Zinc finger nucleases

ABSTRACT

Antibody discovery has become progressively important in almost all areas of modern medicine, where antibody libraries have proven an invaluable resource for the isolation of diagnostic and potentially therapeutic antibodies. Antibody libraries must follow a selection step to collect the leads with the required affinity, specificity and stability. However, the existing display and screening techniques offer some technical obstacles and require an extended protocol being more time-consuming and expensive. Here, we introduce a distinctive mammalian cell system capable of producing antibody libraries that generate its diversity, linked to a display method for a high-throughput selection. Mimicking the V(D)J recombination process that occurs inside B cells to generate diversity, it was possible to generate *in situ* diversity by inducing double-strand breaks that would be inaccurately repaired. The diversity process is controlled using a well-known antibody backbone designed to include specific DNA sequences inside the CDRs, recognized by CRISPR/Cas9 nuclease complex. With this new model, we successfully introduced variability inside the CDRs and created a cell-based platform to combine this strategy with a selection mechanism where antigen and antibodies interact in their native form.

Overall, the results endorse the proof of concept of a distinct and successful approach capable of selecting antibody binders against streptavidin from a diverse library exposed to different selection mechanisms.

This platform increases the likelihood of a selected antibody being well-tolerated and highly effective when employed for the development of therapeutic solutions in humans.

Keywords: Antibody discovery, Antibody library; CDRs; Cell surface display; CRISPR/Cas9; Genome editing.

RESUMO

Os anticorpos há muito tempo são considerados componentes essenciais da resposta imunológica adaptativa e têm se tornado ferramentas importantes em diversas áreas, tais como a Biotecnologia, Biologia Molecular, Indústria Farmacêutica ou Medicina. A utilização de anticorpos para identificar e neutralizar elementos patogénicos foi proposta há mais de um século por Paul Ehrlich, na sua “Teoria da cadeia lateral”. Os seus conceitos proporcionaram um avanço no campo da medicina moderna, onde encontrar moléculas específicas que neutralizam antígenos infeciosos ou tumorais pode garantir o sucesso da terapêutica. Devido à sua excelente especificidade, afinidade e estabilidade, os anticorpos são ferramentas de alto valor terapêutico, de diagnóstico, biotecnológico e são o segmento de crescimento mais rápido do mercado de produtos biológicos.

Os anticorpos são moléculas estáveis e facilmente manipuláveis por engenharia genética, em que o seu tamanho pode ser alterado e a sua função corrigida mediante a fusão com outros anticorpos, proteínas ou toxinas. Tendo em conta as suas características excecionais, a procura por uma melhor produção e seleção de anticorpos mais eficazes tem vindo a aumentar drasticamente. Porém, as tecnologias disponíveis de seleção de anticorpos que permitem posteriormente aperfeiçoar a afinidade e a estabilidade expõem certas limitações.

A tecnologia do hibridoma foi um desenvolvimento pioneiro para a produção de anticorpos monoclonais, imunoglobulinas homogéneas que, por definição, reconhecem um único epítopo. Nesta técnica ocorre a fusão de uma célula do mieloma com um linfócito B. proveniente de um animal imunizado com o antigénio conhecido. No entanto, as diferenças entre o sistema imunológico humano e o do animal originam reações adversas nos doentes tratados com estes anticorpos, o que que conduz à sua rápida eliminação do organismo, à ocorrência de reações de hipersensibilidade e a uma redução na capacidade de atingirem o local alvo de ação e, portanto, uma ineficácia terapêutica.

A evolução das tecnologias recombinante e de *display* permitiram superar a limitação de imunogenicidade com a implementação de uma molécula completamente humana. Tais anticorpos podem ser produzidos por métodos *in vivo* ou *in vitro.* O método *in vivo* consiste em utilizar animais transgénicos, no qual, os genes de imunoglobulinas dos animais são substituídos pelos genes humanos. No caso de desenvolver anticorpos humanos por métodos *in vitro*, uma técnica bastante utilizada é

o *Phage Display* de anticorpos. Apesar de versátil, a tecnologia de *Phage display* não é realizada num ambiente de uma célula eucariota, comprometendo a conformação final, tanto das proteínas alvo como dos anticorpos. Consequentemente, durante a fase de seleção a probabilidade de sucesso em ensaios funcionais é diminuída pelo facto dos anticorpos e antigénios não serem apresentados nas suas conformações nativas. Ademais, poderá haver a necessidade de alguns anticorpos, já selecionados, necessitarem posteriormente de modificações para aperfeiçoar a sua função, afinidade ou estabilidade. Desta forma, todo o processo de produção e seleção de um anticorpo eficaz acaba por se tornar mais demorado. É então pertinente, desenvolver novas estratégias que permitam ultrapassar os atuais obstáculos e que possibilitem a produção de anticorpos com maior biodisponibilidade, afinidade e especificidade, com o objetivo de obter uma maior eficácia terapêutica.

Reconhecendo a importância desta necessidade, este projeto tem como finalidade desenvolver uma nova tecnologia que permita gerar bibliotecas de anticorpos com elevada variabilidade num sistema celular eucariota, que possibilite selecionar anticorpos funcionais com alta afinidade e contra a expressão de qualquer antígeno na sua forma nativa. A ideia principal desta teoria passa pela mimetização do mecanismo de recombinação V(D)J, que ocorre nos linfócitos B durante a formação dos anticorpos, que consiste na introdução de cortes de cadeia dupla (do inglês *Double-Strand Breaks*) em locais específicos do DNA e a sua posterior reparação, que é naturalmente tendenciosa à inserção de mutações no local de corte.

Desta forma, para desenvolver este projeto foi utilizado um anticorpo Humano conhecido modificado especificamente para conter sequências-alvo reconhecidas pelo complexo CRISPR/Cas9 (do inglês *Clustered Regularly Interspaced Short Palindromic Repeats associated protein 9 nuclease*) nas regiões determinantes de complementaridade (do inglês *Complementary Determening Regions*), para que apenas seja gerada variabilidade nos locais de ligação ao antigénio. O anticorpo modificado foi introduzido no genoma das células 293-F, de modo a gerar aleatoriamente uma biblioteca de anticorpos a partir de uma única sequência de anticorpo.

Numa primeira fase foi testada a eficácia de diferentes complexos CRISPR/Cas9 como agentes mutagénicos para a introdução de cortes de cadeia dupla, a fim de estudar a combinação mais propicia à criação de uma biblioteca de elevada variabilidade.

Posteriormente à construção da linha celular com o complexo mutagénico escolhido, todas as células mutadas aleatoriamente foram sujeitas a um mecanismo de seleção intracelular para identificar anticorpos com capacidade de ligação à streptavina. Usando

a Citometria de Fluxo foi possível selecionar a população positiva para a expressão do anticorpo anti-streptavidina à superfície da célula e isolá-la.

Com o trabalho desenvolvido ao longo da tese de mestrado, podemos comprovar que é possível criar um grande repertório de anticorpos com variabilidade intracelular a partir de um único anticorpo. Estas descobertas podem fornecer um grande impacto no desenvolvimento de uma plataforma distinta para a geração de bibliotecas de anticorpos. Ademais, o fato de ter sido utilizada uma IgG Humana para criar esta plataforma, dá-nos a capacidade de superar questões significativas, como a toxicidade ou a tolerância, associadas a imunizações de animais. Esta plataforma única representa um grande progresso para o desenvolvimento de bibliotecas altamente diversificadas, que podem resolver desafios convencionais significativos tornando todo este processo mais rápido e eficaz

Palavras-chave: Descoberta de anticorpos; Biblioteca de anticorpo; Edição genética; Mutagénese dirigida;

TABLE OF CONTENTS

1. GENERAL INTRODUCTION

1.1 IMMUNE SYSTEM AND IMMUNOGLOBULINS

The immune system has two defense mechanisms to obviate the presence of pathogens: natural immunity and adaptive immunity. The natural response, also called innate response, is the first defense against pathogens acting immediately upon the invasion process through phagocytic cells (neutrophils and macrophages), cells with cytotoxic properties (natural killer), or other mediators as cytokines that regulate and coordinate immunity activities[1]. In contrast to innate immunity, there is another immune response whose defensive ability intensifies with subsequent exposure to the same pathogen. Since this response adapts to the ongoing infection, it is named adaptive or acquired immunity, which can also be divide into two types: adaptive cellular response and adaptive humoral response, mediated by different components of the immune system[2]. In adaptive cellular immunity, the helper T (Th) lymphocytes take part in the recognition of antigenic peptides connected to proteins of host cells, releasing cytokines and triggering effector response where the invading pathogen will be destroyed. In addition to helper lymphocytes, cytotoxic T lymphocytes (Tc) are also part of the cell response which recognize and directly destroy infected cells[2]. Adaptive humoral immunity is mediated by B lymphocytes and the products secreted by them - antibodies. Antibody proteins, also known as immunoglobulin (Ig), are a key element of the immune system's defense against a variety of antigens, which is any molecule, macromolecule, virus particle, or cell that contains a structure recognized and bound by an immunoglobulin or T-cell. Upon activation by helper T lymphocytes, B lymphocytes proliferate and differentiate into plasma B cells, which then secrete large amounts of antibodies with distinct classes and functions. Initially, plasma B cells secrete IgM class antibodies. With prolonged exposure to the antigen and the action of helper T cells, plasma B cells are induced to produce antibodies of different classes (IgD, IgG, IgE, IgA) aiming to improve the humoral response[3,4].

Based on their structural differrences in the constant part of the antibody molecule, immunoglobulins are classified into 5 classes or isotypes: IgG, IgM, IgD, IgA and IgE. The structural differences confer distinctive effector functions on the isotopes, which means they each interact with a different subset of immune system proteins. Each isotype is differentiated by unique single amino acid sequences in the constant region of

the heavy chain, wherein all five classes have a percentage of similarity between the amino acid sequences. The most predominant in humans is IgG, up about 75% of the antibodies in the blood, which can be further divided into four subclasses - IgG1, IgG2, IgG3, and IgG4 - varying their internal and antigenic structure. The light chain has only two isotypes —kappa (κ) and lambda (λ), where its relative abundances vary with the species of animal. In humans, two-thirds of the antibodies hold κ chains and one-third have λ chains. There is no functional difference between antibodies carrying κ light or λ light chains. Each antibody, however, contains either κ or λ light chains, not both[2].

1.1.1 IgG Antibody

IgG antibody molecules are homodimer glycoproteins composed by four polypeptide chains: two identical heavy chains (HC), each about 250 amino acids long, and two identical light chains (LC), each about 150 amino acids long. A disulfide bond connects a cysteine at the *C*-termini of each light chain to a cysteine of a heavy chain. In turn, the heavy chains are associated with each other by two disulfide bonds in the hinge region, which covers the variable and constant domains of each heavy chain (Figure 1). This molecule supple hinge region can be cleaved with proteases in order to produce two types of antibody fragments which have complementary functions: Fragment antigen binding (Fab), where lies the site of antigen binding and confers specificity on the antibody, and Fragment crystallizable (Fc) which interacts with other immune system components and is far less variable in its amino acid sequence[5].

The Fab region comprises three Complementarity-Determining Regions (CDRs), also known as hypervariable regions (HV), of high sequence and length variation. The six CDRs consist of an antigen binding region that contains a specific amino acid sequence that recognizes and interacts with the antigen[2,3] knowing that the Fab region only binds to a specific antigen. With the purpose of having antibodies available that can bind to multiple different antigens, several distinct antibody molecules are required. The Fc fragment mediates the effector functions of the antibody molecule by binding to serum proteins and Fc receptors on the surface of cells, such as macrophages. This special structure of the antibody determines its classification, the immune system cells it will bind to and consequently its function. The Fc region has a limited disparity in amino acid sequence amid different antibodies, which is why is It known as the constant region[2,3,6].

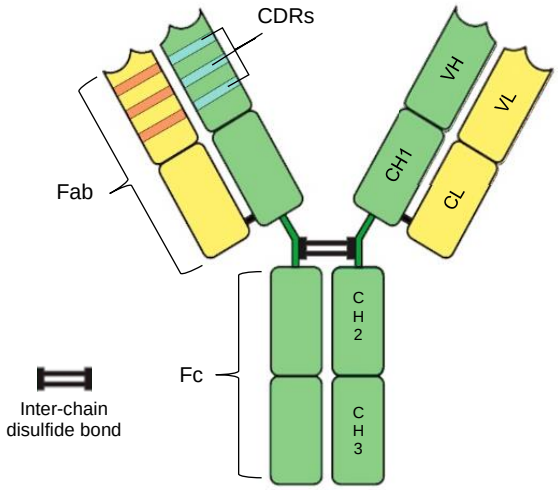


Figure 1| Common structural features of immunoglobulins.

Fab: Fragment antigen binding; Fc: Fragment crystallizable; CDRs: Complementarity-Determining Regions; CH1: constant heavy chain domain; CH2: CH3: VH: Variable heavy chain domain; VL: Variable light chain domain; CL: Constant light chain domain.

1.2 Antibody Diversity Generation

1.2.1 Diversity generation before encountering an antigen

Although an individual immunoglobulin binds to a limited and defined series of linkers, a population of immunoglobins can bind to an almost unlimited matrix of antigens sharing low or no similarity. This property, demonstrated by studies of several antibody molecules, is due to the difference of amino acids concentrated within the CDRs 1, 2, and 3 of heavy and light-chain variable domains. Before the encounter with antigen, in a process termed V(D)J recombination, antibody heavy and light-chains loci somatically assemble a family of gene segments, namely Variable (V), Diversity (D) and Joining (J), sequentially arrayed along the chromosome with each set of segments containing alternative versions of the antibody variable region (Figure 2).

Variances in the sequences of the V gene segments determine the diversity of CDR1 and CDR2, while in CDR3 is determined by differences in the D gene segments and the junctions it makes with the V and J gene segments[6,7]. The light-chain variable region is encoded by the V and J gene segments, whereas the heavy-chain locus includes an

additional D gene segment. This means that for a light chain a single recombination will occur, amid V_L and a J_L, while for a heavy chain will be necessary two recombinations, the first with D_H and J_H and the second will be the union between DJ and V_H. In either case, the particular V, D, and J gene segments that are joined together are selected at random[7] (Figure 2).

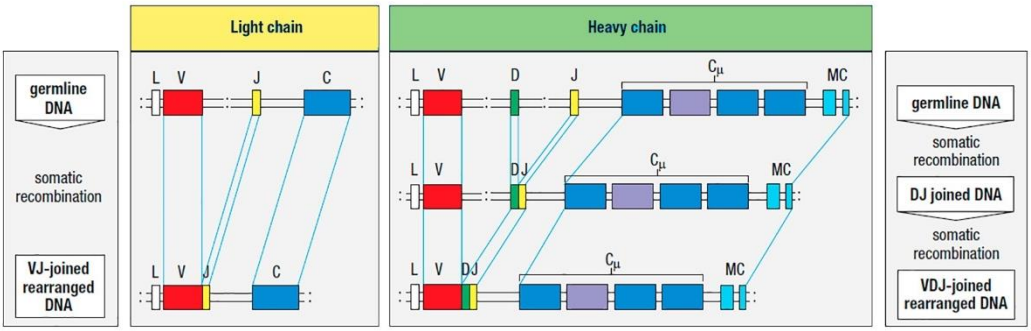


Figure 2| Random recombination of gene segments for diversity production in the antigen-binding sites of immunoglobulins.

Light-chain V-region genes are constructed from two segments, a variable (V) and a joining (J) gene segment in the genomic DNA which are joined to form a complete light-chain V-region (VL) exon. After rearrangement, the light-chain gene consists of three exons, encoding the leader (L) peptide, the V region, and the C (constant) region, which are separated by introns. Heavy-chain V regions are constructed from three gene segments. First the diversity (D) and J gene segments join, then the V gene segment joins to the combined DJ sequence, forming a complete heavy-chain V-region (VH) exon. Reprinted with permission from Parham P. The Immune System. Fourth Edi. Science G, editor. Vol. 90, The Quarterly Review of Biology. New York: Garland Science; 2015.

1.2.2 Diversity generation after B cells encountering an antigen

After B cells have been activated and proliferated, they are capable of enhancing their diversity during the maturation stage. Maturation involves two steps: Class Switch Recombination (CSR) in which a B cell can change the class of antibody it produces and Somatic Hypermutation (SH), in which the rearranged genes for the B cell receptor can mutate to increase the average affinity for their equivalent antigen[8].

The class of an antibody is determined by the constant Fc region of its heavy chain and it is the messenger RNA of this heavy chain that undergoes splicing to yield an M class constant region or a D class constant region. However, by default, when a

newly formed B cell is activated, it produces mostly IgM antibodies. As a B cell undergoes maturation, it has the ability to change the antibody class designed to one of the other antibody classes: IgG, IgE, or IgA. Located just next to the gene segment that encodes the constant region for IgM and IgD are the constant region segments for IgG, IgE, and IgA. Therefore, for the class switch recombination to take place certain special switching signals, located between the constant region segments, allow a B cell to cleave and delete the IgM and IgD constant region DNA and incorporate one of the other constant regions. The result is that the antibody gets a new Fc region, establishing how the antibody will function[8,9].

After the V, D, and J segments have been selected and usually, after class switching recombination has taken place, somatic hypermutation occurs. This event represents a high-rate mutation that intervenes in those regions that contain the V, D, and J gene segments. Somatic hypermutation alters the part of the rearranged antibody gene that encodes the antibody Fab region. These mutations can result in three outcomes: the antibody molecule affinity of its associate antigen may remain unchanged, it may be improved or it may be diminished[8,10]. The purpose of somatic hypermutation is to end up with many more B cells whose receptors present a higher affinity that enables them to bind tightly to their respective antigen. B cells whose receptors have mutated to higher affinity compete more successfully for T cell help. Consequently, they proliferate more frequently than do B cells with lower affinity receptors.

B cells can change their constant Fc region by CSR and their antigen-binding Fab region by SH, and these two modifications produce B cells that are improved to deal with pathogens. Usually, the assistance of helper T cells is required for B cells to make either of these maturation upgrades. Thus, B cells that are not activated with T cell help generally do not undergo either CSR or SH[10].

1.3 Antibody Libraries

Acquired immunity is mediated through numerous genetic and cellular processes that generate favorable somatic variants of antigen-binding receptors under evolutionary selection pressure by pathogens and other factors. Antibody libraries mirror an acquired immune system from which desirable antigen-binding antibodies can be selected[11]. Over the years three main library types were established: immune, naïve and synthetic. These

libraries vary on the origin of the antibody's variable fragments, sizes, compatible display platforms, and practical applications.

1.3.1 Immune Libraries

Immune libraries, which were the initial display libraries, derive from immune donors and are therefore predisposed to recognize specific antigens. These libraries are created from the antigen-sensitized IgG repertoire of the host's available B lymphocytes. Immune libraries represent the expression of antibody cDNA from the available B-cell lymphocyte repertoire, wherever its origin, of a donor that has been immunized, infected or exposed to an antigen[12].

Besides having a high quality and affinity, the antibodies selected from immune libraries have the drawback of requiring immunization for each antigen.

1.3.2 Naïve libraries

Naïve repertoires consist of a collection of variable genes from resting B cells (IgM) of non-immunized donors. These libraries are constructed from IgM bearing cells prior to exposure to antigen. This means that the donors of antibody expressing cells used to generate this library did not suffer immunization against particular antigens. The antibody expressing cells can be isolated from the peripheral blood of lymphocytes, bone marrow, or spleen cells[12,13].

In contrast to immune, naïve libraries have the advantage of being applied in an unlimited range of antibodies, in other words, a single naïve library can be used in the creation of antibodies facing a diverse range of antigens including toxins and self-antigens.

1.3.3 Synthetic Antibody Libraries

The first libraries were created from antibody genes isolated from natural sources, however, evolution brought a growing interest in the construction of rationally designed synthetic antibodies, in which individual library members incorporate structural features that are beneficial for practical applications. These libraries are constructed using

designed synthetic DNA that allows the use of highly optimized human frameworks and enables the introduction of defined chemical diversity at CDRs[14].

Synthetic libraries develop an adaptive immune system by antigen stimulation through somatic mutations, receptor editing, and gene rearrangements to ensure a result of mutated antibody genes. Moreover, high affinity antibodies can be generated against most antigens by introducing diversity into only a subset of positions within the CDRs. Some studies even have demonstrated, in extreme cases, that a binary code of tyrosine and serine is sufficient for generating antigen binding sites capable of recognizing diverse proteins. The ability to precisely define the diversity of the final library simplifies the process of isolating, characterizing, and optimizing an antibody lead[15].

Frameworks for synthetic antibodies can be chosen for particular properties, including high stability and expression. Similarly, for therapeutic applications, optimized human frameworks can be used to minimize the risk of immunogenicity, thus obviating the need for humanization. Lastly, design features can be incorporated to allow control over parameters like the subgroups of the light and heavy chain to enable affinity maturation. In general, the size of the library is proportional to the possibilities for selecting a good clone. Diverging from immune or naïve, synthetic libraries should be very large to yield high-affinity binders[15,16].

Immune libraries can be custom-designed to select specific classes of antibodies, however, a single library from synthetic repertoire can be used for the selection of antibodies against any antigen in theory, which offers a potential solution to the problem of identifying antigen binders in cases where it is impossible to immunize, such as lack of antigen, toxic or not immunogenic antigen. Therefore, we obtain a synthetic diversity that bypasses natural predispositions and allows a more thorough sampling of binding-site architecture than one allowed by natural antibodies. This is the main reason why synthetic libraries are noted for providing a wealth of antibodies against virtually any imaginable target.

Given the vast array of potential antibody applications, techniques for generating recombinant antibodies with desired specificities and affinities have an enormous potential for impact on biological research. Well-designed synthetic antibody libraries have several distinct advantages, which include high levels of expression, good solubility and stability, and the facility of optimization[17]. This allows the development of libraries from which molecules with a new function or unique molecules can be selected.

1.4 ANTIBODY LIBRARIES SELECTION TECHNOLOGIES

The numerous recombinant antibody display technologies available nowadays can be roughly divided into *in vitro* technologies, including phage display, ribosome/ mRNA display, antibody arrays or *in vitro* compartmentalization and into *in vivo* display platforms, such as bacterial, yeast and mammalian cell surface display. All of these techniques are based on an expression system in which the antibody gene (genotype) and the antibody molecule (phenotype) are linked from a vector[17].

Phage display is by far the most popular method, where is possible to express functional antibody fragments on the coat of filamentous bacteriophage and quickly isolate recombinant antibodies from libraries on the basis of antibody-antigen binding. In ribosome display, the system generates stable antibody–ribosome–mRNA complexes using *in vitro* transcription and translation. *In vitro* compartmentalization selects within compartmentalized aqueous droplets, which enable the screening of extremely large numbers of protein variants at each step. Cell display provides a platform with a direct interaction between antigens and specific antibodies that are expressed on the surface of a bacteria, yeast or mammalian cell[18].

Choosing the appropriate selection platform for an antibody library is of extreme importance since it is well documented that by varying the platform, different antibodies are isolated from identical gene repertoires[19].

1.4.1 Phage Display

In this approach, the antibody fragment is displayed on the bacteriophage coat through the fusion of its gene to a gene that encodes for a bacteriophage coat protein and is carried on a vector that can be packaged into bacteriophage particles. Although other phage genes exist, it is most common to use the phage gene III or VIII, which codes for phage surface protein pIII or pVIII - proteins present at the end of the phage involved in the extrusion process that leads to phage assembly[20]. Therefore, we have production of phage particles with the protein or peptide displayed on the surface of the phage as a fusion to the outer coat protein. To display antibody fragment libraries on phage, the library is first cloned into a special type of plasmid called phagemid. Phagemids contain a phage origin of replication to enable the packing of the phagemid DNA into phage particles and also a selectable marker to ensure the selection of the cells containing the

phagemid. The phagemid DNA is then transformed into the host *E. coli* (Figure 3). Therefore, the size of the resulting phage library is limited by the transformation inefficiencies of *E. coli*[18,20]. A culture of the cells transformed with phagemid is infected with M13 helper phage that provides all of the genes required for replication and packaging of the phagemid into the phage virion. The helper phage contains a defective origin of replication or packaging signal resulting in the preferential packing of the phagemid genome over the helper phage genome. After infection, the expression of the antibody-pIII fusions from the phagemid is induced. The wild-type pIII, encoded by the helper phage genome and the antibody-pIII fusion, encoded by the phagemid, compete for incorporation into the virion. As a result, the antibody is not displayed on every one of the five pIII of each particle[18].

To extract those phages from a library, the selection is carried by several rounds of panning. Panning selections consist of several rounds of adsorption of the phage-library and desorption of bound and washed phage particles. The library population is enriched from round to round by increasing the stringency of binding to select for clones exhibiting enhanced affinity. Weak interactions between phages expressing proteins and the target are disrupted by successive washes, while those phages that expose molecules with high affinity for the target are recovered by elution[18].

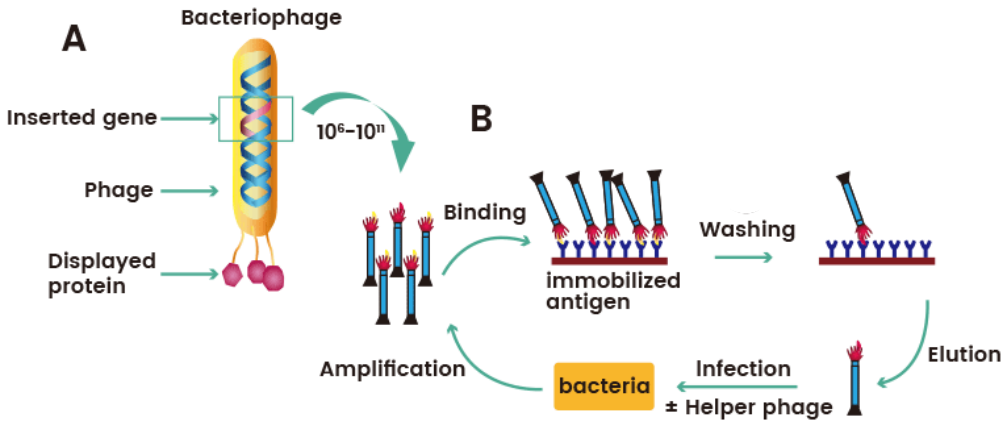


Figure 3| Steps involved in phage display antibody selection.

A: The gene of interest (pink) is cloned into the gene of phage DNA, which results in the display of the protein product on the surface of the phage as a polypeptide fusion. **B:** A phage library is incubated with immobilized antigen, where unbound phages are removed by washing and bound phages are eluted. *E. coli* are infected with eluted phage with or without helper phage to amplify eluted candidates. Process is reiterated 2–3 times resulting in enriched population of antibody/peptide fragments for the antigen of interest. Reprinted with permission from: Inc. SB. Phage Display Overview. 2017[21].

1.4.2 Ribosome and mRNA display

This display method use a DNA cassette containing a promoter and an ORF, encoding a library of the antibodies that is transcribed *in vitro*. The resulting mRNA does not contain a stop codon, which causes a “stalling” translation of the ribosome at the end of mRNA. The translation can run to the physical end of this mRNA, forming complexes consisting of the antibody, the ribosome and mRNA. This complex is selected through panning by binding to an immobilized target antigen. From these bound complexes, the mRNAs are isolated, reverse transcribed and PCR amplified to serve as the input of another round. After 3–5 rounds, the resulting DNA fragments are ligated into an expression vector and *E. coli* is transformed. The different proteins made by individual *E. coli* clones can then be further evaluated[22,23].

The main advantage of this technique is the PCR step, since it allows not only maturation of the antibody affinity, but also a production of large libraries ranging from 10^12 to 10^14 individual clones, larger than phage display. On the other hand, reverse transcription and PCR steps appear to be the most frequent focus of troubleshooting when designing a new system from scratch, which could compromise translation efficiency. Fortunately, these issues are easily and rapidly evaluated. Other drawbacks are the conditions required to stabilize the antibody-ribosome-mRNA complex, which is not always ideal for antibody-antigen interactions. Moreover, the large-scale production of this complex is extremely expensive[23]. Nonetheless, *in vitro* methods give the user full control over where mutations should occur in the sequence, which residue types are to be introduced, or how many random mutations should occur on average.

1.4.3 *In vitro* compartmentalization

In vitro compartmentalization (IVC) is a powerful tool for antibody engineering, allowing the construction of highly complex antibody libraries, compared to previously described methods, and thereby allows a significantly greater number of protein sequence variants at every stage of specificity redesign[11].

IVC technology uses a water-in-oil emulsion to physically compartmentalize the genetic material, translational factors and substrate. Within each droplet, is employed a mechanism that enables *in vitro* transcription and translation to combine an enzyme function to the gene. When the emulsion is broken, functional genes can be recovered

with each emulsion containing a single mutant sequence. These functional genes are amplified afterward by PCR and subjected to further rounds of selection and mutagenesis until the required characteristics have been achieved. Unlike techniques that utilize organisms, this approach is not limited by transfection efficiencies and libraries of up to 10^{10} per mL of an emulsion can be easily screened[24,25].

1.4.4 Cell Surface display

Cell surface display allows the expression of not only fragments, but also of full-length antibodies on the surface of bacteria, yeast, or mammalian cells as anchoring motifs. These systems have been used for many different biotechnical and biomedical applications, proving their effectiveness in protein engineering studies, such as vaccine development, bioabsorbants, biocatalysts, biosensors and so on[26]. Two important properties of cell display for protein engineering purposes are the large particle size and the multivalent surface expression of recombinant proteins. Together, these properties enable detection and subsequent analysis of antigen binding properties in a flow cytometer. Consequently, in contrast to phage and ribosome display where selections from libraries are generally based on a capture and elution procedure, selections from cell-displayed libraries are typically performed using fluorescence-activated cell sorting (FACS). In FACS, the antigen is labeled with a fluorophore and the cell-displayed protein library is incubated with the fluorescent antigen in solution[27]. Compared with other display strategies, cell surface display systems exhibit distinct advantages. First, the molecule presented at the cell surface is freely accessible for any kind of binding or activity studies without the need for a substrate or binding partner to cross a membrane barrier. Second, when connected to a matrix, proteins have proven to be more stable than free molecules. Third, the need for preparation or purification of molecules for many applications is unnecessary as whole cells displaying the molecule of interest can be applied to reactions or analytical assays and can be removed afterward by a simple centrifugation step[28].

- Bacterial surface autodisplay

Bacterial surface autodisplay uses a novel autotransporter pathway in which the autotransporters are synthesized as precursor proteins containing all structural requirements for the transport to the cell surface. IgA1 protease *from Neisseria*

gonorrhoeae was the first member of this protein family that was discovered and characterized by Meyer and co-workers in the late 1980s. It was soon realized that this secretion mechanism could be exploited for the transport of a recombinant protein in *E. coli*, by replacing the coding region for the natural passenger - the IgA1 protease - with the coding region for the recombinant protein of interest and subsequent expression and surface display in *E. coli*[29].

- Yeast surface display

Yeast display is emerging as an effective technology for isolating and engineering antibodies for therapeutics development and a variety of biomedical applications. In this system, the antibody is displayed on the yeast surface by fusing to the yeast agglutinin protein Aga2p, which attaches to the yeast cell wall through the binding with Aga1p by two disulfide bonds. Expression of the antibody-Aga2p and Aga1p are under the control of galactose inducible *GAL1* promoter on the yeast display plasmid which is maintained in yeast episomally[30]. One of the main advantages this technology offers is its eukaryotic system providing very sophisticated protein folding and chaperones machinery, which enables the construction of yeast display library with large size up to 10^{10} individual clones. This makes yeast display comparable to phage display system in terms of library size and thus further simplifies the initial antibody or protein isolation process[30,31].

- Mammalian cells surface display

Antibody expression systems within mammalian cells might be divided into two large categories, transient expression systems in which introduced genes do not integrate into the cell's genome, thus the new gene will not be replicated and stable expression systems in which heterogeneous DNA integration is selected using a marker gene resulting in long-lasting antibody expression. Both systems have been exploited for antibody display with their own advantages[26,32].

In transient expression systems, special mRNAs encoding fragments of antibodies could be transcribed from the transformed vector directly and then translated and assembled into complete antibody molecules within mammalian cells. Different approaches had been described to improve the quantity of mammalian cell display libraries using transient expression systems. The conventional one is to use a two-vector system with each vector carrying heavy and light chain, respectively. However, it is sometimes difficult to verify the amount of the transmitted vectors, leading to unequal

molar production of a heavy and light chains. One of the newest approaches supports the insertion of both heavy and light chain genes in one diachronic vector, under a sole promoter, with an additional internal ribosomal entry site located between them[33]. A single mRNA molecule encoding both heavy and light chains can then be transcribed from this diachronic vector.

Stable expression systems enable the introduced genes to integrate into the host cell genome using a standby virus expression system where these integrated genes could then transcribe and translate within cell metabolism. Those expression systems could provide long last and stable antibody expression on the surface of mammalian cells[26]. Since these host cells are relatively large, flow cytometry (FC) can be used to screen these libraries. It is possible to use this method for affinity selections, where the hosts displaying an antibody library are incubated with a fluorescently-labeled antigen. FC is able to examine individual cells quantitatively at rates of up to 10^9 cells/hour. The light scattering of each cell triggers the flow cytometer to measure the fluorescence intensity of the cell, which is a direct measure of the amount of bound fluorescent antigen. The number of bound antigen molecules per cell depends on several factors: affinity of the expressed antibody for the antigen, antibody expression level and the protein proper folding[33]. Enrichment between rounds of selection can be assayed without any additional experiments by comparing the fluorescence of each round's total population of cells. The fluorescence corresponds to the antigen binding potential of the cell population from each round of selection. The real-time quantitative analysis of parameters such as the kinetics and equilibrium coupled with high-throughput screening makes FC and cell display a powerful combination when it comes to screening antibody libraries. Besides, this platform is the most similar to a natural environment[34].

1.5 DNA Double-Strand Breaks Genome Engineering

Genome engineering is a field in which specific chromosomal loci are modified at precisely defined sites, generating cells and organisms with heritable DNA sequence alterations[35]. Several tools have been developed for this purpose, including meanucleases[36], zinc finger nucleases (ZFNs)[37], transcription activator-like effector nucleases (TALENs) [38] and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) endonucleases[39]. Meganucleases are naturally occurring, compact DNA

cleavage enzymes that recognize long (~20 base pairs) DNA targets[40]. ZFNs can be purchased or engineered using publicly available resources but often display measurable off-target activity[37]. ZFN proteins are more compact in size than TALENs and CRISPR associated protein 9 (Cas9). In contrast, TALENs and CRISPRs can be more easily reprogrammed to a wider range of DNA sequences, but their lengthy reading frames may complicate packaging and delivery in certain contexts and applications[41].

1.5.1 Meganucleases

Meganucleases are sequence-specific endonucleases that recognize large (>12 base pairs) sequence DNA targets[40] and can therefore cleave their equivalent site without affecting global genome integrity[42]. Meganucleases have been shown to hold a variety of properties, including high cleavage specificities and minor monomeric folds, that are suitable for genome editing. These proteins are ideal tools for genome engineering as they are specific enough to bind and cleave only one site in a designated genome. However, they tolerate different sequence variations at these sites[42].

Knowing that meganucleases natural available repertoire is limited, there is a need to increase this library to reach a greater number of targets. Meganuclease family members can be employed as templates to engineer tools that cleave DNA sequences other than their original wild-type targets, promoting double-strand breaks (DBS) repaired by homologous recombination (HR) of defective genes with a very low level of toxicity[36,43] Consequently, previous knowledge of the structure/function relationship of these enzymes with their DNA targets is crucial to design and construct custom enzymes that target a DNA sequence of interest while preventing harmful side effects in cells.

1.5.2 Zinc-finger nucleases

Zinc-finger proteins are one of the amplest groups of proteins and have a wide range of molecular functions. Given the wide variety of zinc finger domains, ZFNs are able to interact with DNA, RNA, PAR (poly-ADP-ribose) and other proteins[37]. ZFN is an artificial protein consisting of a C2H2 (Cys_2-His_2) zinc-finger DNA binding domain and the nuclease domain of the FokI restriction enzyme. When a pair of ZFNs binds to DNA in a tail-to-tail configuration, the two FokI nuclease domains dimerize and induce a DNA

double-stranded break[44]. The C2H2 zinc finger domain is a peptide comprised of ~30 amino acids (aa). It contains two β-strands and one α-helix that recognizes three to four bases of DNA. These domains can be fused to achieve a recognition of 9 to 18 base pairs DNA sequences with high specificity and affinity[45].

The applications of genome editing using ZFNs are based on the introduction of a site-specific DNA DSB into the locus of interest. All eukaryotic cells efficiently restore DSBs via homologous recombination or non-homologous end joining (NHEJ) pathways[31–33]. These highly conserved pathways can be exploited to create defined genetic outcomes across a wide range of cell types and species[46].

Due to its relative simplicity, ZFN-mediated gene disruption is the first ZFN-based approach that has been taken to the clinic. Specifically, for the treatment of glioblastoma[47] and HIV[48]. The inherent risk of a therapeutic approach using a ZFN-induced DSB is the potential for low-frequency off-target cleavage events at undesired locations in the genome. Validating ZFNs for clinical application thus requires more sensitive methods, like ultradeep sequencing or bioinformatics.

1.5.3 Transcription activator-like effector nuclease - TALENs

Transcription activator-like effector nuclease (TALENs) are proteins from the pathogenic *Xanthomonas* bacteria, whose main well-established function is to induce expression of specific host plant genes that enhance virulence[49]. TALENs contain a nuclear localization signal, a transcriptional activation domain, and a DNA binding domain composed of a series of 33–35 amino acid repeat domains where each monomer recognizes a single bp in the DNA target. The order and number of repeats in the DNA binding domain varies and establish the binding specificity for the DNA sequence[41.]

Generating TALENs to specifically induce DSBs, involves fusion between TALENs and the endonuclease domain of FokI. These TALENs, in the form of functional protein pairs, bind and cleave their DNA targets, resulting in double-stranded breaks that trigger the host cell's DNA repair systems[50]. While the TALEN-DNA binding repeats single base recognition affords larger design flexibility than zinc finger proteins, the cloning of repeat TALEN arrays poses an elevated technical challenge due to extensive matching repeat sequences.

TALENs are useful and advantageous in view of their minimal cytotoxicity and off-target editing[51]. TALENs have high flexibility in design for specific target sites and several

examples of high editing efficiency. However, they take longer to build, are larger, hence more challenging to deliver and are prone to generate an immune response. There is also a targeting limitation, consensus in the literature, which implies that TALENs binding site should start with a T base[52].

1.5.4 Clustered Regularly Interspaced Short Palindromic repeats - CRISPR

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system was originally discovered as part of the adaptive immune system in bacteria and archaea against invasion by viruses or bacteriophages and since then it has been exploited for the genetic manipulation of eukaryotic cells[53]. The CRISPR locus in microbes typically consists of a set of noncoding mature CRISPR RNAs (crRNAs) that act as guide RNA (gRNA) to recognize the target foreign DNA and cleave it with the help of Cas proteins. The sequence guide within gRNA has a length of ~20 bp and is the precise complementary sequence of the target site within the genome, also referred to as “protospacer”. The Cas9 cleavage site is located within the protospacer and positioned at around 3 bp upstream the 5′ end of the protospacer adjacent motif (PAM) “NGG” trinucleotide motif[54].

CRISPR/Cas technology can be divided into six types: types I, III and IV are associated with multi-protein effector complexes (e.g. Cascade, Cmr, Csf1) and types II, V and VI are recognized by single-subunit effector (e.g. Cas9, Cpf1, Cas13)[54]. The type II CRISPR/Cas9 system is the best characterized and most used technique among the systems. CRISPR/Cas9 detects the target sequence through base pairing and afterward, Cas9 induces double-strand breaks that stimulate natural cellular repair machinery.

Cas9 protein consists of two major domains: a large recognition domain (REC) and a small nuclear domain which cleave the complementary and non-complementary DNA target strand. To recognize a target-specific site, Cas9 requires a guide RNA which consists of a trans-activating crRNA (tracrRNA), that facilitates the maturation of crRNA and it’s loading onto Cas9, and the short-conserved sequence (5’-NGG-3’) PAM. The RNA-guided nuclease Cas9 assumes an inactive or active conformation depending on whether is free in the cellular environment or connected to gRNA or target DNA, respectively[55]. As soon as gRNA binds to Cas9, occurs an alteration in Cas9 structural conformation which enables the formation of RNA–DNA hetero-duplex. Then, the

Cas9/gRNA complex drifts through the DNA in order to identify the 5'-NGG-3' PAM site. After PAM site recognition, the Cas9 and gRNA aggregate begin to separate the DNA strand and search for the adjacent DNA complementarity to the gRNA[56]. The formation of Cas/gRNA and target DNA hetero-duplex generates the creation of an R-loop structure which afterward activates the Cas9 nuclease activity, creating a DSB at the DNA target site[57].

CRISPR outstands other methods, like ZFN or TALENs, due to relatively low noise, lower off-target effects and better consistency. Employing the inexpensive generation of single gRNA lentiviral libraries leads to large-scale screens using CRISPR which helps the study of a much broader phenotypic array than was previously possible[58]. Combined with fast and easy DNA and RNA synthesis, CRISPR allows more trackable editing on a genome-wide scale and can instantaneously correct genetic mutations or modify regulatory patterns in patients. Research has highlighted the possibility that CRISPR may also have therapeutic uses in various diseases, such as cancer[59], hemoglobinopathies[60], muscular dystrophy[61], nervous system diseases[62], or liver diseases[63].

This technology has broken barriers in research and human therapeutics. In 2015, for the first time, Chinese scientists edited the human embryo with CRISPR/Cas9 technology to replace the HBB gene, responsible for beta-thalassaemia[64]. Most recently, were presented by the scientist He Jiankui, the first 'CRISPR' baby which harbors multiple edited versions of the *ccr5* gene[64]. There are however some drawbacks to this technology that could affect the application. For instance, the off-target mutation or the editing efficiency. In order to overcome these obstacles, a reduction of the gRNA length to 17/18 nt has been applied which enhances the specificity of Cas9 system[65]. Off-target mutation can also be controlled with the replacement of positively charged amino acids with neutral ones in the nontarget DNA which would diminish the interaction amid Cas9 and the non-target strand[66].

The emergence of ZFNs, TALENs and CRISPR as a programmable genome editing tool for a diverse range of purposes have triggered a revolution in the field of human genome engineering. These methods have the ability to revolutionize biological research and influence medicine. Certainly, these developing technologies have truly expanded the ability to modify and study model organisms and also demonstrate a possibility to repair the genetic causes behind many diseases. However, to attain the complete potential of these technologies, many significant questions and challenges

must be addressed. Being the most pertinent, the specificity of each nuclease model and accurate delivery to cells or tissue *in situ*.

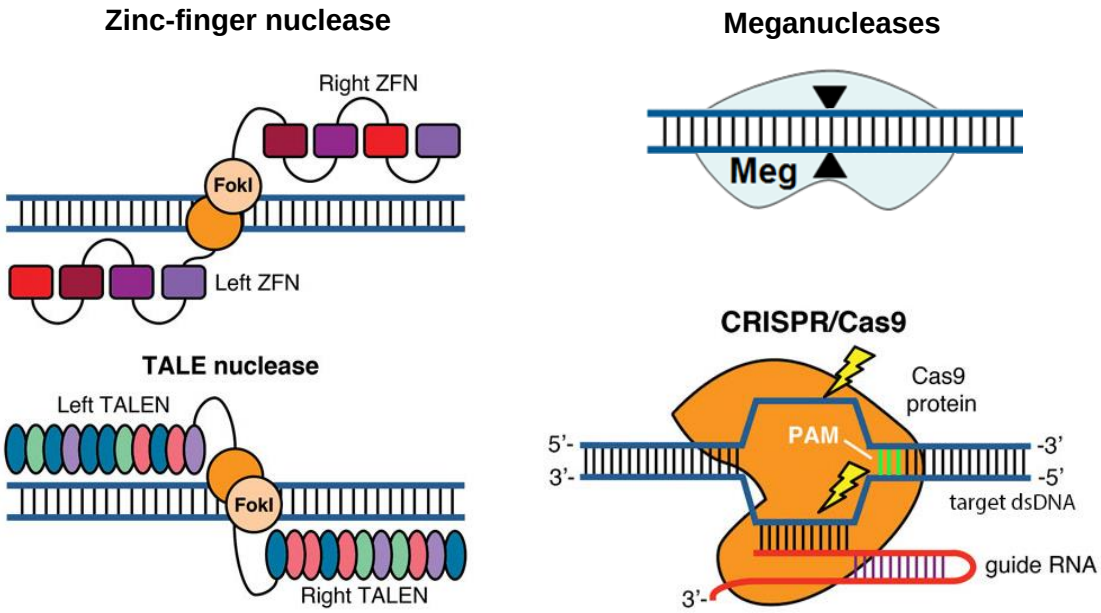


Figure 4| The structure and mode of DNA recognition of meganucleases, zinc finger nucleases, TALE nucleases, and CRISPR/Cas9. Adapted from Gersbach CA. 2014[67].

1.6 DNA Double Strand Break Repair

For the purpose of maintaining the genome integrity between cellular divisions, eukaryotic cells have evolved to repair several types of DNA damage. Amongst types of damage, DNA double-strand breaks are especially adverse as they can result in insertions, deletions or chromosomal translocations which are the main transforming step in numerous human cancers[68]. All eukaryotic cells efficiently repair DSBs via the homologous recombination or non-homologous end joining pathways. Homologous recombination allows high fidelity DSB repair and healthy cell growth. Whereas, a non-homologous procedure has greater potential to contribute to mutations[46,68] (Figure 5).

Pathological DSBs can emerge from both exogenous (radiation) or endogenous (DNA replication errors) sources. Both pathological and physiological DSBs entail

efficient procedures for repair that result in minimal to no change of the broken chromosome. However, not all DSBs are pathological and there are several cases when DSBs are introduced deliberately, which serves specific physiological purposes. As an example, is the V(D)J recombination where cells use DSB repair pathways to generate diversity. V(D)J recombination is assisted by non-homologous end-joining elements and does not entail homologous recombination[69]. Disruption of NHEJ elements results in serious immune disorders in humans and mice, suggesting their crucial role in V(D)J recombination[70]. Such as V(D)J recombination, class switch recombination is also dependent on end-joining mechanisms. Different from V(D)J or CSR, somatic hypermutation obtains antibody variation through mutagenesis in the V regions of light and heavy chains. Most of these mutations are single base substitutions with a small fraction of short insertions or deletions. This indicates that the development of a DSB is not a require step in SHM and in fact NHEJ-deficient cells do not reveal significant defects in SHM[71]. Additionally, it has been demonstrated that DSBs during SHM can be repaired by homologous recombination[71,72]. Clarifying the specifics and the regulation of these DSB systems will be an exciting direction of future investigation. Knowing that misregulation of these DSB repair pathways results in genome rearrangements that are typical in many cancer types, understanding these processes is extremely relevant for human health. To date, many of the key repair factors for each pathway have been recognized, however, the reason why a cell determines which pathway to use for DSB repair is still poorly understood.

1.6.1 Homologous recombination

Homologous recombination is a process that requires a homologous sequence as a template for repair. More precisely, is when a targeting construct containing the desired genomic modifications, flanked by sites homologous to the gene of interest, is inserted into the desired locus in the genome[73]. This allows the recombination apparatus to re-establish any missing genetic information in the proximity of the break site, and as a result, HR is largely accurate. Compared to non-homologous end-joining, HR is systematically more complex, involves a larger number of enzymes and is thus comparatively slower but more precise[73,74].

In HR, the first procedure is the DNA restriction at the break site which exposes extended sequences of ssDNA (single strand). Then, this ssDNA is used in the search for a homologous dsDNA (double-strand) sequence that will be used as a template for the

DSB repair by the recombination pathway[75]. The DNA end resection initiation is strictly controlled through the activation of key resection factors by cyclin-dependent kinase (CDK)-catalysed phosphorylation. This control mechanism allows HR to start only when a repair template is available and thus limits the possibilities for illegitimate recombination[76,77]. Additionally, unsuccessful repair and prolonged cell cycle arrest leading to the activation of apoptosis in higher eukaryotes[46].

1.6.2 Non-homologous End Joining

The end-joining pathways can be divided into canonical non-homologous end-joining (NHEJ) and alternative non-homologous end-joining (alt-NHEJ), also called microhomology-mediated end-joining (MMEJ). NHEJ and MMEJ involve the direct ligation of two DSB ends with slight or no sequence homology required[74]. Consequently, a repair template is not necessary, so it can take place during any phase of the cell cycle.

Both end-joining pathways normally lead to a partial loss of genetic information which results in short deletions at the DSB site. In addition, it can be exploited to introduce small insertions and/or deletions at the break site, an outcome that can be exploited to disrupt a target gene. Also, since NHEJ and MMEJ are template independent, binding of inaccurate ends may produce large deletions or chromosomal rearrangements if multiple DSBs are present[74].

NHEJ requires little processing of the broken DNA ends. The primary action of NHEJ implies the binding between the DNA ends and the Ku70-80 heterodimer, which creates a ring that surrounds the duplex DNA[78]. This action will protect the DNA ends from degradation and recruits additional NHEJ elements. In the case of DNA ends that comprise DNA overhangs, gaps, or blocking chemical groups, additional DNA end processing may be required. This involves removal by the human Artemis nuclease of nucleotides or chemical groups overhangs, which cleaves at the junctions of ssDNA and dsDNA, and is activated by DNA-dependent protein kinase catalytic subunit (DNA-PKcs)[74,79,80]. Alternatively, the filling of DNA gaps at breaks is implemented by DNA polymerases μ and λ[81].

Several shreds of evidence indicate that NHEJ frequently acts first to repair DSBs[82–86]. A study[85] with fluorescent reporter constructs incorporated into the chromosomes of human cell lines revealed that NHEJ is faster than homologous recombination with a duration of 30 minutes versus several hours for HR and is associated with approximately 75% of

repair events. According to these assessments, HR might be considered a pathway that works in particular contexts, when NHEJ is inactive or unsuccessful.

Regarding the end-joining systems, NHEJ substantially diverges from MMEJ. Whereas NHEJ does not require or requires the slightest homology (<4 nt) between the broken DNA molecules, MMEJ was established as a DNA end-joining method that occurs apart of the key NHEJ factors and typically involves small sections of microhomology (2–20 nt) between the two broken DNA ends to mediate repair [87]. Therefore, in the absence of the key canonical NHEJ factors cells can still repair DSBs through MMEJ. The use of microhomology for repair indicates that MMEJ and recombination-based mechanisms may share elements of DNA end resection. This is also supported by observations that Ku70-80 inhibits MMEJ[88].

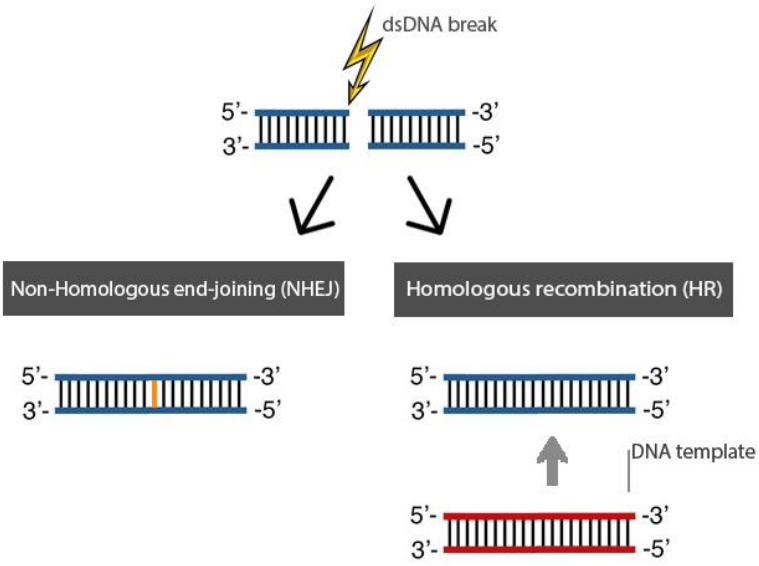


Figure 5| Alternative pathways of DNA double-strand break repair.

1.7. IMPACT OF THERAPEUTIC ANTIBODIES

Since its discovery, antibodies have been used as therapeutic agents for the treatment of infectious diseases, especially bacterial ones, as initially observed by Von Behring and Kitasato in their studies with diphtheria and tetanus. The study demonstrated a serotherapy using polyclonal antibodies which represent a heterogeneous mixture that recognizes several epitopes at the same time.[89] These

findings aroused the attention not only of the medical community but also the public in general. However, despite the efficiency in marking and eliminating specific targets, antibodies were not economically attractive to the pharmaceutical industry, mainly due to the high cost of production, their difficulty in being produced on a large scale, their undesirable effects provided by constant use and by the multiple specificities of different isotopes in the serum. A key strategy for overcoming some of these obstacles was the development of hybridoma technology to produce monoclonal antibodies (mAbs). Using these homogeneous immunoglobulins, it is possible to recognize a single epitope and have higher specific activity than polyclonal antibodies. In general, monoclonal antibodies are superior to polyclonal antibodies in terms of homogeneity, specificity and safety. Since 2008, 48 new mAbs have been approved, contributing to a total global market of 79 mAbs (58 in Europe) in clinical use at the end of 2019, according to the US FDA and EMA.[90]

This has been possible with the creation of several techniques that increase immunogenic potential and efficacy while making possible the therapeutic use of antibodies for an extended duration. Humanization of antibodies by the complementary determining region grafting technique, the creation of fully human antibody libraries or the development of bispecific antibodies notably accelerate the approval of therapeutic mAbs directed against diseases that require long-term treatment, such as cancer and autoimmune diseases[91]. Based on the success of humanized mAbs in the clinic, the first fully human therapeutic antibody, adalimumab (Humira), an anti-tumor necrosis factor α (TNFα) human antibody, was approved in 2002 by the US FDA for rheumatoid arthritis and in 2018 was the world's best-selling drug[92]. Despite mAbs being usually employed in cellular biology, biochemistry or medical research, possibly its most beneficial application is their use as therapeutic drugs for the treatment of human diseases, such as breast cancer, asthma, rheumatoid arthritis, psoriasis, Crohn's disease, transplant rejection or infectious diseases[93].

The mAb market enjoys a healthy course and is expected to grow at an improving rate. Human, humanized, chimeric, and murine antibodies respectively account for 51%, 34.7%, 12.5% and 2.8% of all mAbs in clinical use, making human and humanized mAbs the dominant modalities in the field of therapeutic antibodies. However, their stigma is yet intense due to an idea of huge spending in general health and pharmaceutics[94]. In a global drug market that represents just over 720 billion euros, which is currently dominated by seven companies: Genentech (30.8%), Abbvie (20.0%), Johnson & Johnson (13.6%), Bristol-Myers Squibb (6.5%), Merck Sharp & Dohme (5.6%), Novartis (5.5%), Amgen (4.9%), organic medicine produce about 45%. Half of this sum

corresponds to immunotherapy, whose market is growing strongly, with an average annual growth rate of 13.5% between 2016 and 2019[94]. Pharmaceutics are encouraged to select medicines that are on the list of standard therapies over new immunotherapies since the market will easily cover the production and sale of standard drugs. In Portugal, according to the National Program for Oncological Diseases, 200 million euros are annually spent on chemotherapy products alone. Products that are mass-produced and give millions to pharmaceutical companies. Given the major therapeutic benefits that antibodies bring in an increasingly significant number of severe pathologies, their stigma seems unfounded. Besides, it would be appropriate to deduct from the cost of therapeutic antibodies their beneficial impact on the overall consumption of healthcare and medical goods.

1.7.1. Challenges of CRISPR/Cas9 technology as a therapeutic tool

Gene therapy is undoubtedly a promising treatment option for human diseases, including inherited disorders, some types of cancer and certain viral infections, because it has the capacity of determining unique molecular characteristics of individual patients. CRISPR/Cas9 is amongst the available tools for precision-based gene therapy, allowing for simple, economic and time-saving genome editing. Since 2013, when validated in human cells, the RNA-guided CRISPR/Cas9 system has been applied for DNA and RNA manipulation in numerous cell lines and organisms[95]. Considering its extensive repertoire of applications, CRISPR/Cas9 is a promising system to optimize biomedical research and innovate treatment approaches for gene-associated diseases. Start-up companies and Universities are leading the innovations of medical treatment in humans with CRISPR-Cas technology. Submissions for patents increased between 2012 and 2017 with 75 patents already issued[96]. To date, only studies on cells and small animals have provided support for the therapeutic effectiveness of CRISPR/Cas9 gene editing technology with the purpose of repairing pathological mutations that cause genetic diseases. Examples with a positive outcome include gene repair in Duchenne Muscular Dystrophy (DMD)[97], Sickle Cell Disease (SCD)[98], β-thalassemia[99] and Hereditary Tyrosinemia type I (HT1)[100].

CRISPR-Cas genome editing has been also considered a powerful tool in cancer research. Due to its efficient and direct editing of the target gene along with adaptation for diverse delivery strategies. CRISPR/Cas9 technique has been evidenced to knockout

the Trp53, Pten and Nf1 genes accountable for glioblastoma in the mouse brain and Ptch1 gene responsible for medulloblastoma[101]. Disruption of the drug resistance-related gene is also another possibility of action to avoid one major limitation in cancer treatment, the development of resistance to chemotherapy, as evidenced in Ha et al., 2016[102]. Clinical trials in humans have been approved for investigations of Sickle Cell disorder, β-thalassemia, multiple myeloma and sarcoma and Leber Congenital Amaurosis, the most common type of inherited blindness in children. In the majority of these trials, the extracted autologous T cells from patients were edited by CRISPR/Cas9 system and then injected back into patients[103]. Regardless CRISPR/Cas9 system being a highly efficient genome editing tool, both for genetic improvement of plants and animals and for the clinical investigation of human genetic diseases, it also expresses potential risks in certain cases. In this context, there are three key concerns for this technology: off-target editing, immunogenicity of Cas9 nuclease and illegal or irresponsible experimentation by the scientific community.

- Off-target editing:

Off-target editing presents a potential risk in therapeutics as it may cause loss-of-function mutations in proper functional genes or incorrect repairing of disease-inducing genes, due to binding and breaking at sites other than the target DNA sequence. Off-target editing may also lead to chromosomal rearrangements and other types of mutations, including the integration of DNA mismatches into the PAM-distal position of the sgRNA sequence[104].

Efforts to reduce off-target editing have led researchers to improve gRNA design, generate new versions of the Cas9 nuclease and optimized delivery vehicles. Creating unique gRNA for a specific gene using bioinformatics tools like CRISPRfinder, RGENs, E-CRISP, CRISPR Design, ChopChop and ZiFiT can design exclusive gRNAs for target sequences that avoid binding to any off-target sites in the genome by at least 2–3 nucleotides variation in gRNA[105]. Another suggestion is the use of different versions of the Cas9 nuclease like mutated Cas9 D10A and H840A, which have one active site and produce single-strand DNA only, creating dual nicks at two different recognition domains of DNA to generate two single-strand breaks at two sites[106]. Another approach to minimize the late induction of off-target mutation is the delivery of Cas9-gRNA via electroporation, lipofection or protein transduction, as it induces on-target mutation immediately after insertion and degrades soon[107].

Nevertheless, recent studies by Feng et al.[108] and Lee et al.[109] found that off-target mutations by CRISPR/Cas systems are very rare or undetectable. Application of CRISPR/Cas9 in plant genome editing offers evidence of a few low frequency off-target mutations[110] and lower genetic variation than radiation-mutagenesis[111].

- Immunogenicity of Cas9 nuclease:

The immunogenicity of Cas9 nuclease is another concern that should be reflected during the clinical translation of CRISPR/Cas9 technology. Approximately 40% of the human population is colonized by *Staphylococcus aureus* and 12% of the children under 18 have asymptomatic colonization with *Streptococcus pyogenes*. About 80% of healthy individuals have been detected with anti-saCas9 (*S. aureus*) or anti-SpCas9 (*S. pyogenes*) antibodies. However, most of these reactions are against secreted proteins and proteins found on the surface membrane of the bacteria, which are easily accessible to the immune system[112,113].

An option to minimize the probability of an adverse immune response after CRISPR/Cas9 gene therapy would be to screen patients for anti Cas9 antibodies and T-cells before any intervention. Though, the substantial amount of the population to which this applies prompts its reassessment. To circumvent this problem, alternative strategies can be employed, like the induction of immune tolerance or immune suppression, the Cas9 structural modification to mask immunogenic epitopes, the use of Cas9 orthologs from non-pathogenic bacteria, or provide a microenvironment actively compliant of a foreign protein such as the immune privileged organs[114].

- Illegal or irresponsible experimentation by the scientific community:

The burst of CRISPR/Cas9 applications highlighted its potential but also the ethical apprehensions associated with the possible creation of permanent and inheritable modifications in the human genome. That is the reason discussions about ethical guidelines within international multidisciplinary groups are imperative to regulate and diminish the potential risks associated with this powerful tool. Working with human cells requires safety and ethical consent for any gene variation by the National Institute of Health (NIH). Currently, NIH does not fund studies using CRISPR on human embryos and also opposes the practice of CRISPR on germline cells since any such changes would be permanent and heritable[115]. Chinese scientist He Jiankui's violated these rules when engineering human twin IVF embryonic genomes, where these modifications might

be passed on to future generations. Chinese court recently found the researcher guilty of illegal medical practice and has sentenced him to three years in prison[116].

Despite the existence of some ethical concerns, CRISPR/Cas9 technology has made major progress to rectify pathogenic gene mutations in human embryos. This gene editing tool has effectively helped rectify MYBPC3 mutations associated with hypertrophic cardiomyopathy in human embryos, with a targeting efficiency of 72.2%, and the resulting embryos did not display any kind of mosaicism, off-target gene editing or other abnormalities. Moreover, embryos favored the gene copies from the healthy parent to the exogenous DNA donor as a repair template[117]. Such results may reduce ethical worries to a certain level, although further studies are essential to improve the targeting efficiency and determine the long-term safety of CRISPR/Cas9 technology.

With CRISPR tools becoming widely available and applicable, it is possible for the technology to be misused by an individual or organization. CRISPR expressively decreases the cost and expertise barriers of previous gene editing approaches. Methods such as ZFN genome editing, offering a lower percentage of mutated cells than CRISPR, can be quite expensive and are currently only available through Sigma Aldrich. ZFNs were used broadly for years, but the tool's effectiveness frequently required adding different enzymes or complicated procedures, such as cold culture conditions. This diverges from the current CRISPR kits, which may offer *E. coli* as an original organism and entail fewer supplements[118].

The ability to rapidly modify a genome at a relatively low cost compared to previous methods could make CRISPR systems attractive for nefarious actors at all levels, from individuals through nation states. Presently, living the epidemic from COVID-19, CRISPR poses a considerable possible biosecurity threat that may be misused to create increased virulence pathogens, neurotoxins and even *de novo* organisms. Creating an entirely novel organism is theoretically conceivable, but it is likely to require extensive training, funding and time for research and development, which is less possible for some types of individuals[119]. Still, there are solutions that may help to detect the misemployment of this technology: guidelines for gene synthesis companies to screen orders of possible misuse, like the ones established by the US Department of Health and Human Services (HHS); support scientists ability to self-govern, to give them control and authority to develop rules which may eventually lead to regulations; provide biosecurity training at the institutional level, a knowledge that they will carry throughout their work in academia or industry; make CRISPR Biosafety guidelines public to encourage bioresearch safety since this technology is widely available through traditional providers

such as ThermoFisher Scientific and nontraditional providers such as Odin Technologies[107].

Furthermore, newly reported anti-CRISPR proteins may be harnessed as effective regulators of CRISPR systems in gene-editing applications for biotechnology and medicine. It also could help as biosecurity countermeasures capable of handle in some genome edited bioweapon. Bondy-Denomy together with Davidson, microbiologist Karen Maxwell and fellow graduate student April Pawluk, have discovered these new anti-CRISPRs. These proteins feature the capability of blocking the cut-and-paste action of CRISPR systems, providing scientists with a toolkit for keeping gene editing controlled[120].

For Type II CRISPR, which uses a single enzyme as Cas9, researchers have discovered more than 50 anti-CRISPR (Acr) proteins that switch off DNA-editing activity in different ways. Two commonly observed mechanisms inhibit CRISPR complexes from binding target DNA (AcrIIA4 and AcrIIC1) or specifically block the DNA-cleavage (AcrIE1)[121]. Even with a growing number of projected applications and proof-of-concept experiments, investigators have yet to pin down the therapeutic potential of this anti-CRISPR system.

2 OBJECTIVES

The general aim of this thesis is to develop a unique cell-based platform to create a Semi-synthetic antibody library capable of creating its diversity, integrating fully human antibodies.

To achieve this purpose, three major goals were proposed:

- Selection of a suitable target CRISPR/Cas9 nuclease;
- Intracellular antibody library generation via targeted mutagenesis;
- Projection of an intracellular selection mechanism.

The plan was to design a mammalian cell system capable of mimicking the V(D)J recombination mechanism that occurs in B cells for antibody generation. For that reason, a Human antibody gene was engineered to have no affinity for any target but to include specific recognition sites in the CDRs for the targeted nuclease CRISPR/Cas9. It was predicted that this system formed double-strand breaks in the target DNA, more precisely inside the CDRs, that would be repaired by the non-homologous end joining mechanism able to promote the introduction of *indels* in the cleavage site.

To reach a single method for the generation and selection of antibodies, the antibody library was engineered to interact with a target antigen in its native conformation. Through flow cytometry analysis, the positive population for the expression of the membrane protein at the cell surface can be sorted, the genomic DNA extracted and the antibodies isolated.

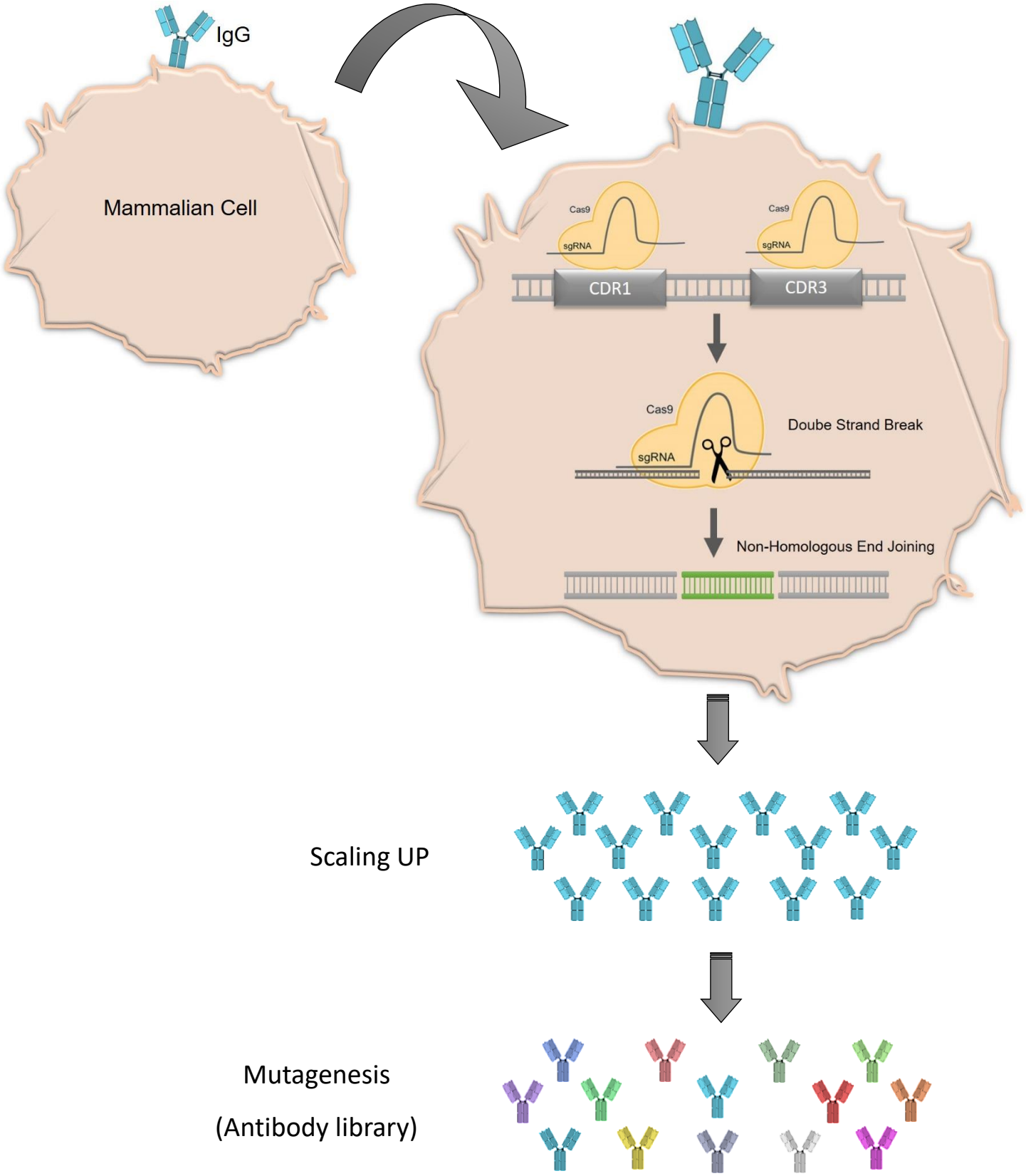


Figure 6| Display of the intracellular technology used to create an antibody library. Representation of an antibody expressed by a mammalian cell system with hotspots for CRISPR/Cas9 in the CDRs which induce double-strand breaks in the DNA. NHEJ repair system produces indels at the break sites, generating different antibody sequences that represent the antibody library. CDR: complementary-determining region; NHEJ: non-homologous end joining.

3 MATERIAL AND METHODS

3.1- BACTERIAL STRAINS

With the view to propagate lentiviral clone's DNA, it was used the NEB® Stable (C3040H) New England Biolabs, Inc.) and Stellar™ competent cells (Clontech Laboratories Inc.). These bacterial strains, suitable for high efficiency transformation, were grown in LB medium (10 g tryptone; 5 g yeast extract and 10 g NaCl in 1 L ddH2O) supplemented with plasmid's selectable marker Ampicillin (100 µg/mL). Incubation was performed at 30º C to reduce recombination activity. Bacterial cells were transformed according to the manufacturer's protocol. Genotypes are shown in Annexes.

3.2- PLASMID CONSTRUCTIONS

All antibody genes were cloned in the lentiviral vector FugW, under control of the human ubiquitin C (*hUBC*) constitutive promoter, fused to a leader sequence for the secretory pathway, derived from the murine Igκ chain, to the P2A self-cleavage peptide and to the hemagglutinin (HA) tag. Heavy and light chain domain antibody backbone were fused to the fluorescent protein GFP or to CD20 protein marker, respectively.

To assess CRISPR/Cas9 nuclease activity, two variable light domain (V_L and V_L*) variants (in FugW) were designed to contain CRISPR/Cas9 recognition sequences in the place of CDR1 or CDR3. The recognition sequences are shown in Table 1.

CRISPR/Cas9 vectors were purchased from Addgene. All these vectors have the same lentiviral backbone pXPR_001, which contains the human codon-optimized Cas9 protein with a FLAG tag upstream and a puromycin resistance gene, under the control of EF-1 Alpha Short promoter (EFS). The specific targeting synthetic single-guided RNA (sgRNA) elements are under control of the U6 promoter.

After analysis of the antibody variants, separate vectors for light and heavy chain both containing a hotspot for lentiCRISPR-EGFP sgRNA 1 in CDR1 and a hotspot for the lentiCRISPR-EGFP sgRNA 4 in CDR3 were designed (in FugW) to proceed for cell lines construction.

As a negative control was used pNeuLite plasmid which has the promoter region of the human gene Her-2/neu cloned in the plasmid pGL2-basic, fused to a Luciferase tag. Expresses the human epidermal growth factor receptor 2 (HER2/neu) oncogene, member of the human epidermal growth factor receptor family and comprises an ampicillin resistance. pNeuLite was a gift from Mien-Chie Hung[122].

pMDLg/pRRE, pCMV-VSV-G and pRSV-REV obtained from Addgene repository, encoding the packaging proteins Gag-Pol, VSV-G and Rev respectively, were used to produce lentiviral particles for cell transduction.

Table 1| sgRNA CRISPR/Cas9 sequences.

Name and Recognition Sequence	Source
lentiCRISPR - EGFP sgRNA 1 [GGGCGAGGAGCTGTTCACCG]	Gift from Feng Zhang. Expresses human codon-optimized Cas9 protein and puromycin resistance from *EFS* promoter and an EGFP targeting synthetic single-guide RNA (sgRNA) element from U6 promoter. Lentiviral backbone.
lentiCRISPR - EGFP sgRNA 2 [GAGCTGGACGGCGACGTAAA]	Gift from Feng Zhang. Expresses human codon-optimized Cas9 protein and puromycin resistance from *EFS* promoter and an EGFP targeting synthetic single-guide RNA (sgRNA) element from U6 promoter. Lentiviral backbone.
lentiCRISPR - EGFP sgRNA 4 [GGAGCGCACCATCTTCTTCA]	
lentiCRISPR - EGFP sgRNA 5 [GAAGTTCGAGGGCGACACCC]	
lentiCRISPR - EGFP sgRNA 6 [GGTGAACCGCATCGAGCTGA]	

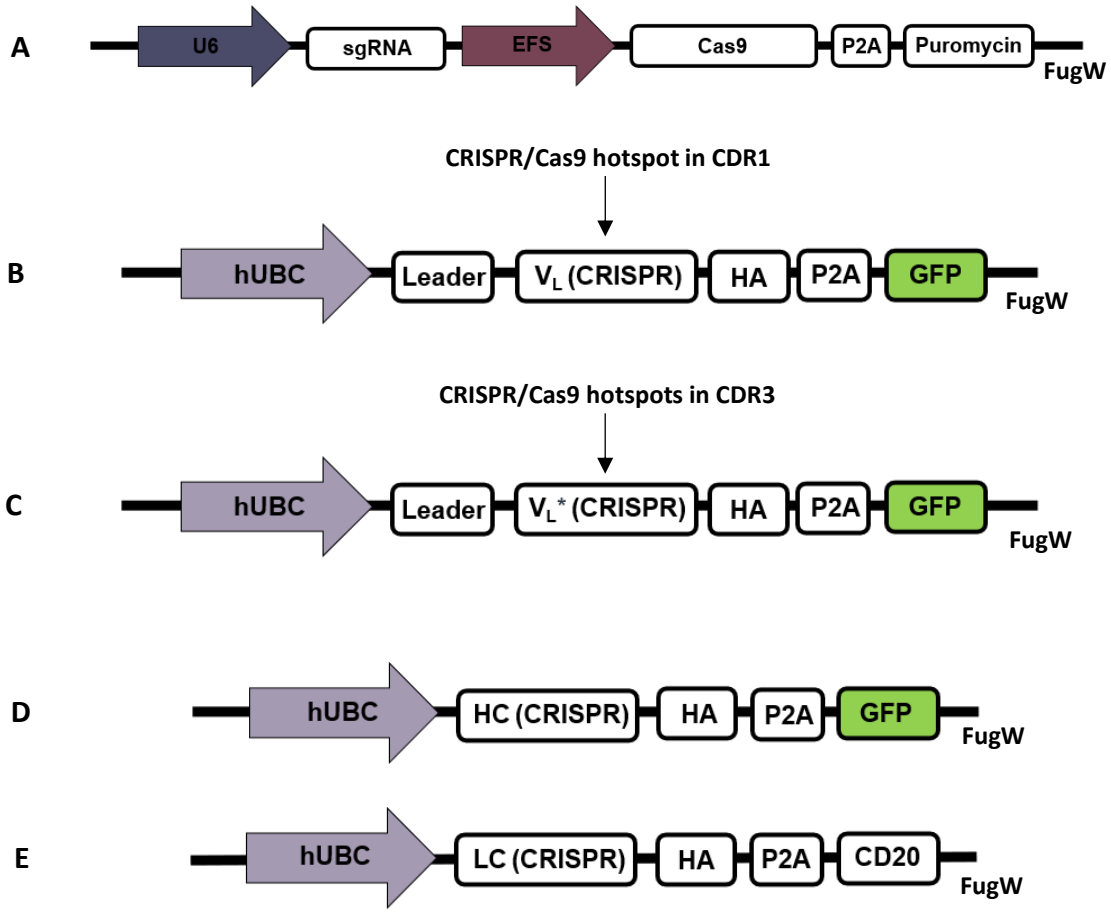


Figure 7| Representation of DNA constructs.

A: Lentiviral expression vector for Cas9 and sgRNA (lentiCRISPR) with puromicyn resistance gene. **B:** antibody construct VL with a hotspot in CDR1 for CRIPR/Cas9 sgRNA 1. **C:** antibody construct VL* with a hotspot in CDR3 for CRIPR/Cas9 sgRNA 2,4,5,6. **D:** Heavy chain vector with a hotspot for the lentiCRISPR-EGFP sgRNA 1 in CDR1 and the lentiCRISPR-EGFP sgRNA 4 in CDR3. **E:** Light chain vector with a hotspot for the lentiCRISPR-EGFP sgRNA 1 in CDR1 and the lentiCRISPR-EGFP sgRNA 4 in CDR3.

U6: U6 promoter; sgRNA: single-guided RNA; EFS: EF-1 alpha short promoter; P2A: porcine 2A self-cleaving peptide; CDR: Complementary Determining Region; hUBC: Leader: IgG k-chain leader sequence; V_L: Light Chain Variable region; HA: hemagglutinin tag; GFP: green fluorescent protein; HC: heavy chain; LC: light chain; CD20: cell receptor cluster of differentiation 20.

Table 2| Plasmids used to produce lentiviral particles.

Name	Source
pMDLg/pRRE	3rd generation lentiviral packaging plasmid; Contains Gag and Pol pMDLg/pRRE was a gift from Didier Trono.
pCMV-VSV-G	Envelope protein for producing lentiviral and MuLV retroviral particles. pCMV-VSV-G was a gift from Bob Weinberg.
pRSV-REV	3rd generation lentiviral packaging plasmid; pRSV-Rev was a gift from Didier Trono.

Plasmid DNA extraction was performed using the NucleoBond Xtra Midi kit (Macherey – Nagel).

3.3- SYNTHETIC OLIGONUCLEOTIDES

Table 2| Oligonucleotides used for PCR amplification in Surveyor assay.

Primer	5' - 3' Sequence
Heavy Chain Forward	CTTGTTGCTCCACGCCGC
Heavy Chain Reverse	GGGGAAGTAGTCCTTGACCAGGC
Light Chain Forward	CCACTGGCCTTGTTGCTCCAC
Light Chain Reverse	TCTCTGGGATAGAAGTTATTCAGCAGGCAC

3.4- Cell Culture

During the master thesis, were used three cell lines: Human embryonic kidney cells 293-T (HEK 293T) (ATCC®, USA), Jurkat E6.1 cells (ATCC®) and FreeStyle™ 293-F cells (Invitrogen, R79007). 293T cell line are cells isolated from human embryonic kidneys transformed with large T antigen. HEK 293T cells have fast reproduction, easy maintenance, the ability to transfection using a wide variety of methods and high efficiency of transfection and protein production. This cell line was used for lentiviral

production and also to study the CRISPR's DNA constructs expression. HEK 293T were cultured in Dulbecco Modified Eagle's Medium (DMEM) (Lonza, Basel, Switzerland) with 10% (v/v) heat-inactivated Fetal Bovine Serum (FBS) (Biowest, France), 2 mM L-Glutamine (Lonza) and antibiotic/antimycotic PSA (100 U Penicillin, 100 µg Streptomycin, 0.25 µg Amphotericin B) (Lonza).

Jurkat E6.1 cells are an immortalized line of human leukemic CD4+ T lymphocyte cells. This cell line was employed for lentiviral transduction and consequent construction of stable cell lines. Jurkat E6.1 cells (ATCC®) were cultured in RPMI-1640 (Lonza) with 10% (v/v) heat-inactivated FBS (Biowest), 2 mM L-Glutamine (Lonza), PSA (100 U Penicillin, 100 µg Streptomycin, 0.25 µg Amphotericin B) (Lonza) and 2.09g/L Sodium bicarbonate.

FreeStyle™ 293-F cells are a variant cell line from HEK 293 cells covering the same benefits as HEK 293T with the difference of enabling transfection of cells at large volumes, allowing these cells to produce high levels of protein. FreeStyle™ 293-F cells were cultured in FreeStyle™ 293 Expression Medium (Gibco™) serum-free with GlutaMAX™-I.

HEK 293T cells are adherent to polarized plastic surfaces, whereas Jurkat E6-1 and 293-F cells are suspension cultures. Since FreeStyle™ 293-F cell line is derived from the 293 cell line primary embryonal human kidney it can also be cultured as adherent. However, its preferable state is in suspension. These three cell lines effectively multiply in a humidified atmosphere incubator at 37° C and 8% CO_2.

3.5- LENTIVIRAL PRODUCTION VIA 293T CO-TRANSFECTION

Lentiviral production was carried by two different transfection reagents: Lipofectamine™ 3000 (Invitrogen) and jetPRIME® (Polyplus Transfection®) in HEK 293T cells, following each corresponding manufacturer's protocol. Lipofectamine 3000 is a lipid nanoparticle technology that provides superior transfection performance in the widest variety of biological cell types. jetPRIME® is a transfection agent based on a polymer formulation that ensures effective and reproducible DNA transfection into mammalian cells. These reagents require low amounts of nucleic acid per transfection, hence resulting in very low cytotoxicity.

Using Lipofectamine™ 3000, $1x10^6$ HEK 293T cells were seeded in 6-well plate (Orange Scientific) or $2x10^5$ in 24-well plate (Sarstedt) with DMEM+/+ to achieve an 80% confluency on the day after. According to the manufacturer's protocol, 3 µg (6-well plate) or 1 µg (24-well plate) of DNA per well were transfected. The total amount of DNA includes: pMDLg/pRRE, pCMV-VSV-G, pRSV-REV and the transfer plasmid in a proportion of 1:2:1:1, respectively. After 6 hours of incubation, the medium was replaced with fresh DMEM+/+ and then incubated overnight. At 24- and 48-hours post-transfection the cells, the supernatant was recovered, aliquoted and placed at -80° C.

With jetPRIME® reagent, 24h prior transfection, $2.5x10^5$ HEK 293T cells were seeded in 6-well plate (Orange Scientific) or $7x10^4$ in 24-well plate (Sarstedt) with DMEM+/+. According to the manufacturer's protocol, 2 µg (6-well plate) or 0.5 µg (24-well plate) of DNA per well were transfected. The total amount of DNA includes: the transfer vector of interest, pMDLg/pRRE, pCMV-VSV-G , pRSV-REV and the transfer plasmid in a proportion of 1:1:1:1. After 4 hours of incubation, the medium was replaced with fresh pre-warmed DMEM+/+ and then incubated overnight. At 24- hours post-transfection the cells, the supernatant was recovered, aliquoted and placed at 4° C (if used in the same day) or -80° C.

3.6- CONSTRUCTION OF JURKAT CELL LINES

Using a 24-well plate, $5x10^5$ Jurkat cells were plated in 500 µL of complete RPMI in each well. Transduction was performed using 250µL of lentiviral particles, produced in HEK293T with 8 µg/ml Polybrene (Hexadimethrine bromide – Sigma-Aldrich, MO, USA) to facilitate adsorption. After 6-hour incubation at 37^o C, change the medium for fresh complete RPMI and incubate overnight. 24 hours past-transduction, cells were centrifuged 5 times and the pellet resuspended with fresh pre-warmed complete medium, maintaining the initial concentration of $5x10^5$ Jurkat cells per 500 µL in a 24-well plate. To ensure that the transduction of lentiviral particles, produced by Lipofectamine™ 3000 and jetPRIME® was efficient, cells were analyzed by direct and indirect flow cytometry.

3.7- CONSTRUCTION OF 293-F CELL LINES

Using a 24-well plate, $5x10^5$ 293-F cells were plated in 500 µL of complete FreeStyle medium in each well. Transduction was performed using 250µL of lentiviral particles, produced in HEK293T with 8 µg/ml Polybrene (Hexadimethrine bromide – Sigma-Aldrich, MO, USA) to facilitate adsorption. After 6-hour incubation at 37º C, change the medium for fresh complete FreeStyle and incubate overnight. 24 hours past-transduction, cells were centrifuged 5 times and the pellet resuspended with fresh pre-warmed complete medium, maintaining the initial concentration of $5x10^5$ 293-F cells per 500 µL in a 24-well plate. When transducing the lentiCRISPR constructs, cells were selected by adding puromycin (0.5 µg/ml) (Invitrogen) to the cell culture medium 48 hours after transduction so that only cells expressing Cas9 were preserved.

To ensure that the transduction of lentiviral particles, produced by Lipofectamine™ 3000 and jetPRIME® was efficient, cells were analyzed by direct and indirect flow cytometry.

3.8- WESTERN BLOT

- Cell dissociation and lysis

Before starting Western Blot technique, it is required to extract the desired protein from our samples, in this case, adherent or suspension cells. In the presence of adherent cells, there is a primary requirement to dissociate the cells from the culture vessels. For this purpose, we used a dissociation protocol centered on the reagent Cell Dissociation Buffer, enzyme-free, PBS (Gibco™), which is a membrane-filtered, isotonic and enzyme-free solution of salts, chelating agents, and cell-conditioning agents in calcium-free and magnesium-free PBS. Following, it was used a lysis buffer – RIPA buffer (Sigma-Aldrich) combined with a protease inhibitor to attain the lysate from cultured mammalian cells in suspension. The mixture was incubated on ice for 30 minutes for immediate use or frozen at -80 °C until needed. After centrifugation at 16 000 g for 30 minutes at 4° C, the supernatant containing proteins was quantified by Bradford Protein assay (AppliChem) and 50 µg of each sample was used for analysis.

- Protein electrophoresis in polyacrylamide gel (SDS Page)

The samples, together with the loading buffer (NewEngland BioLabs), as the PageRuler™ Prestained Protein Ladder (Thermo Scientific™) were applied in a denaturing polyacrylamide gel (12 or 15%) and run in an electrophorator using a low voltage (120 V) for separating gel and higher voltage (180 V) for stacking gel (5%). The gels were subsequently subjected to Western blotting technique.

- Electrotransfer

After electrophoresis in polyacrylamide gel, proteins were electrotransferred to a nitrocellulose membrane (Amersham HybondTM-C, GE Healthcare). The transfer was performed for 60 min at 90 V. To ensure that the transfer was complete, the membrane was stained with Ponceau S 0.1% (Panreac Applichem) in 1% acetic acid (Merck). Following the confirmation, the membrane was blocked for 1 hour at room temperature or overnight at 4° C, under gentle agitation with 5% whole milk in PBS / 0.1% Tween 20 (VWR Chemicals BDH prolabo). The primary antibody, horseradish peroxidase (HRP)-conjugated anti-HA Monoclonal Antibody (Roche), in PBS-Tween 20 0.1% was added and incubated for 90 min at room temperature or overnight at 4° C with gentle agitation. Posterior incubation, three washes were performed with PBS/ 0.1% Tween 20 and later revelation with Amersham ECL HRP-Conjugated Antibodies (Cytiva).

3.9- DIRECT AND INDIRECT LABELING FOR FLOW CYTOMETRY

For flow cytometry was used direct and indirect labeling, which indicates the use of a single antibody directed against the target of interest (direct) or the use of two antibodies, where the primary antibody is unconjugated and a fluorophore-conjugated secondary antibody is directed against the primary antibody for detection (indirect)[33].

Cells were harvested and washed with ice cold 3%BSA/ PBS and then were incubated, at least for 30 min, with a primary antibody (1 µg/1x106 cells) at room temperature or 4º C in the dark. Following incubation, cells were washed three times by centrifugation at 400 g for 5 min with resuspension in ice cold PBS.
When proceeding with indirect labeling, there is subsequent incubation with a secondary antibody in 3% BSA/PBS (1 µg/1x106 cells) for at least 20-30 minutes at room

temperature or 4o C in the dark. Cells are washed again by centrifugation as before and analyzed by flow cytometry as soon as possible. All flow cytometry data were analyzed with FlowJo software (TreeStar).

3.10- SURVEYOR NUCLEASE ASSAY

The GeneArt® Genomic Cleavage Detection Kit provides a simple, reliable, and rapid method for the detection of locus specific cleavage of genomic DNA. The assay uses genomic DNA extracted from cells transfected with constructs expressing engineered nucleases such as TALEN, CRISPR/Cas9, or Zinc-finger nuclease.

Following cleavage, genomic insertions or deletions are created by the cellular repair mechanisms. Loci where the gene-specific double-strand breaks occur are amplified by PCR. The PCR product is denatured and reannealed so that mismatches are generated as strands with an indel re-annealed to strands with no indel or a different indel. The mismatches are subsequently detected and cleaved by Detection Enzyme and then the resultant bands are analyzed by gel electrophoresis and band densitometry.

The use of Surveyor nuclease to detect mutations involves four steps: (i) PCR to amplify target DNA from both mutant and wild-type reference DNA; (ii) hybridization to form heteroduplexes between mutant and wild-type reference DNA; (iii) treatment of both heteroduplex and reference homoduplex DNA with Surveyor nuclease to cleave heteroduplexes; and (iv) analysis of DNA products using any suitable separation platform[123]. 1x10^6 cells from each sample were collected for genomic DNA extraction. Following the manufacturer's protocol of GeneArt Genomic Cleavage detection Kit (Invitrogen™) it was possible to extract the genomic DNA and quantify the performance of the genetic tool used – CRISPR/Cas9.

4 RESULTS

4.1- VERIFY CRISPR/CAS9 EFFICIENCY FOR THE GENERATION OF DOUBLE-STRAND BREAKS

The first part of this master thesis was the evaluation of CRISPR/Cas9 system to verify their nuclease efficiency to create DNA double-strand breaks in the antibody gene. To accomplish this purpose, two antibody variants were constructed to hold five different sgRNAs targeting enhanced green fluorescent protein (EGFP) in CDR1 or CDR3[124]. One of the V_L constructs contains a hotspot targeted by sgRNA EGFP 1 in the place of CDR1, yet the other V_L* construct comprises a hotspot in the place of CDR3 targeted by sgRNA EGFP 2,4,5 and 6 (targeting a different coding sequence). Each antibody variant was co-transfected in HEK293T cells with the corresponding sgRNA in a proportion of 1:2 (antibody vector: nuclease vector). To obtain a negative control the antibody variant was co-transfected in HEK293T with pNeuLite and for the specificity control the 293T was co-transfected with V_L* plus sgRNA EGFP 1.

The results displayed in figure 9 were analyzed 48 hours after transfection by Western blot to assess changes in antibody construct size, caused by target nucleases, with anti-HA-HRP monoclonal antibody. The positive control sgRNA EGFP 1, demonstrated in the previous reference[124], expressed positive guidance to Cas9 resulting in the DNA's cut since it is possible to see a faded band of 15 kDa in lane 1. This 15 kDa represents the protein size of the V_L construct which after an efficient cut by the Cas9 cannot be translated into a normal protein. sgRNA EGFP 4 also shows a probable positive effect as Cas9 guide due to the fading of the V_L* band. In contrast, sgRNA EGFP 2, 5 and 6 did not validate their efficiency to guide Cas9 to perform DNA's cut, as the WB outcome reveals the presence of a 15 kDa band similar to the negative control using pNeulite. Cas9 precision and specificity are likewise proven when co-transfecting sgRNA EGFP 1 with the V_L* construct, displaying a WB perfect band at 15 kDa.

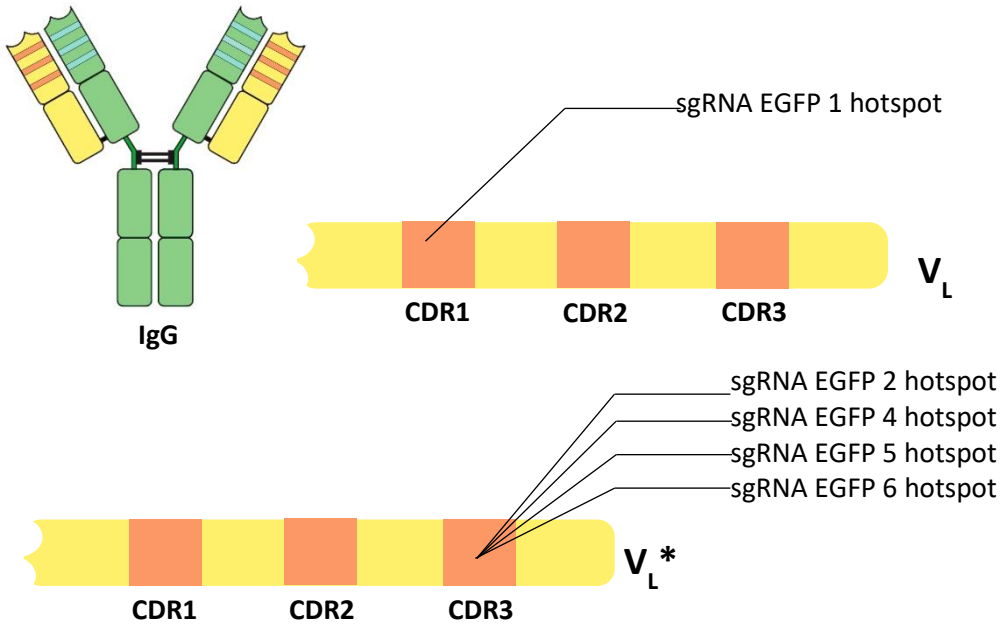


Figure 8| Schematic representation of hotspots localization in V_L constructs.

V_L only has the hotspot for sgRNA EGFP 1 inserted in CDR1, V_L* has the hotspot for sgRNA 2, 4, 5, 6 inserted in CDR 3. IgG: Immunoglobulin G; CDR 1 and 3: Complementarity Determining Region 1 and 3; V_L: Light Chain Variable region; sgRNA: Single-guide RNA; EGFP: enhanced green fluorescent protein.

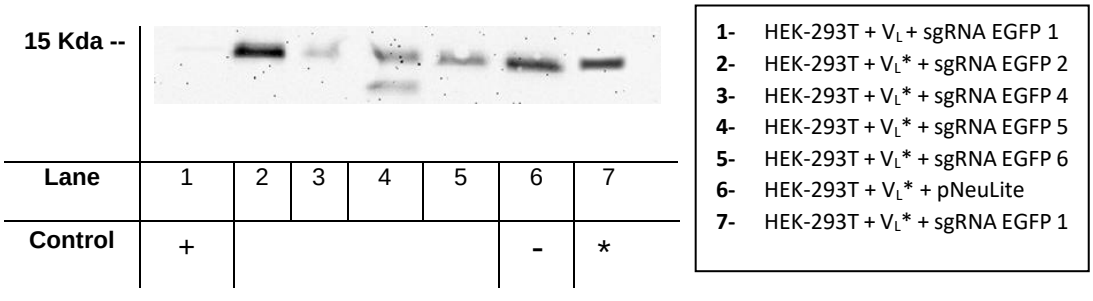


Figure 9| Western blot analysis of VL antibody constructs present in the protein lysates of the HEK-293 T cell line to evaluate CRISPR/Cas9 efficiency for the generation of double-strand breaks.

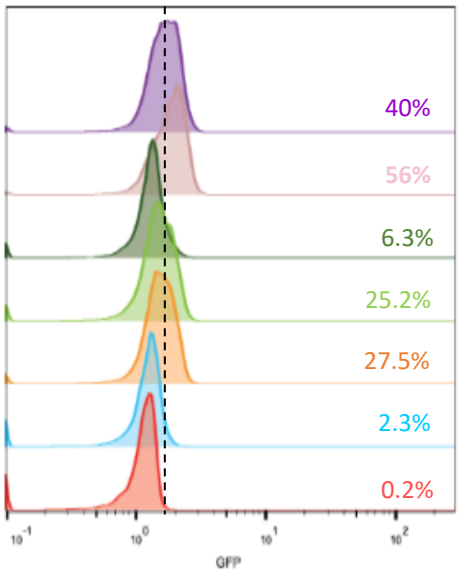
HEK293T cells were co-transfected with the desired V_L, V_L* and the CRIPR/Cas9 sgRNA 1,2,4,5 or 6. Negative control for gene modification was prepared by transfecting HEK293T with pNeuLite. HEK-293T: Human Embryonic Kidney 293 cells; V_L: Light Chain Variable region with a hotspot in CDR1 for CRIPR/Cas9 sgRNA 1; V_L*: Light Chain Variable region with a hotspot in CDR3 for CRIPR/Cas9 sgRNA 2,4,5,6; sgRNA: Single-guide RNA; EGFP: enhanced green fluorescent protein; pNeuLite: plasmid with the promoter region of human gene Her-2/neu.

4.2. TRANSDUCTION OPTIMIZATION IN JURKAT E6.1 CELL LINE

For the purpose of constructing a stable cell line, Jurkat E6.1 was used to define the perfect transduction conditions to be used in the course of this project. The human T lymphocytes cell line was engineered to express a full antibody at the cell surface using FugW lentiviral vectors fused with Heavy or Light chain backbone, produced trough lipofectamine™ 3000 transfection kit in HEK 293T. Both constructs were attached to a fluorescent protein - EGFP or the Human T cell receptor cluster of differentiation 20 - CD20 to control their presence and enable selection. The transduction optimization was based in two parameters: transductions hours (time past transduction) and Lentivirus concentration.

Results were analyzed by Flow Cytometry, either to detect directly the EGFP expression (Figure 10) or indirectly immunostained with antibodies that target CD20 and emit Alexa Fluor 405 fluorescence, (Figure 10).

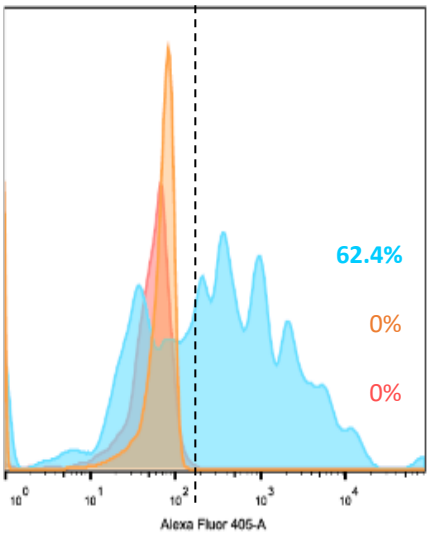
Exhibited in Figure 10 are de data from the lentiviral transduction of Heavy Chain in Jurkat E6.1, where is possible to identify the desirable conditions as being 24 hours past-transduction with lentivirus concentration of 100µL per 5x10^5 cells, since it achieved a higher EGFP expression. As control there is non-transduced JKT cells.



Cell Line	Transduction Hours	Lentivirus concentration (µL per 5x10^5 cells)
JKT + HC:EGFP	24	200
JKT + HC:EGFP	**24**	**100**
JKT + HC:EGFP	24	10
JKT + HC:EGFP	6	200
JKT + HC:EGFP	6	100
JKT + HC:EGFP	6	10
WT JKT	/	/

Figure 10| Flow Cytometry analysis of EGFP expression for transduction optimization in Jurkat E6.1. Lentiviral particles transduced in JKT by lipofectamine™ 3000 with different lentiviral concentration per 5x10^5 cells and distinct post-transfection times. Finest transduction condition was assessed over EGFP expression. JKT: Jurkat E6.1; WT: wild type; HC: Heavy Chain; EGFP: enhanced green fluorescent protein.

Once recognized the ideal transduction characteristics, JKT_HC:EGFP was subsequently transduced with FugW lentiviral vector fused with Light Chain backbone. Figure 11 expose Flow Cytometry results for the Alexa Fluor 405 fluorescence expression, revealing even higher transduction efficiency than JKT_HC:EGFP.



Cell Line	Transduction Hours	Lentivirus concentration (µL per 5x10^5 cells)
JKT + HC:EGFP + LC:CD20	24	100
WT JKT^	---	---
WT JKT	---	---

Figure 11| Flow Cytometry analysis of Alexa Fluor 405 expression for transduction optimization in JKT_HC:EGFP_LC:CD20. Light Chain:CD20 lentiviral particles transduced in JKT_HC:EGFP by lipofectamine™ 3000. After 24 hours cells were immunostained for the CD20 tag and followed with the analysis by flow cytometry to detect Alexa Fluor 405 expression. JKT + HC:EGFP + LC:CD20 were marked with 1st and 2nd antibody and WT JKT^ were marked with 2nd antibody. JKT: Jurkat E6.1; WT: wild type; HC: Heavy Chain; EGFP: enhanced green fluorescent protein; LC: Light Chain; CD20: human T cell receptor cluster of differentiation 20.

4.3. CONSTRUCTION OF 293-F_HC CELL LINE

After validating the ideal transduction conditions using JKT_HC:EGFP and JKT_HC:EGFP:LC:CD20 cell lines, the next step was to replicate the process in FreeStyle™ 293-F cells. The lentiviral having HC:EGFP antibody construct was integrated in 293-F cells and after 24 hours EGFP expression was studied by flow cytometry. Figure 12 shows a high efficiency transduction of 73,3%.

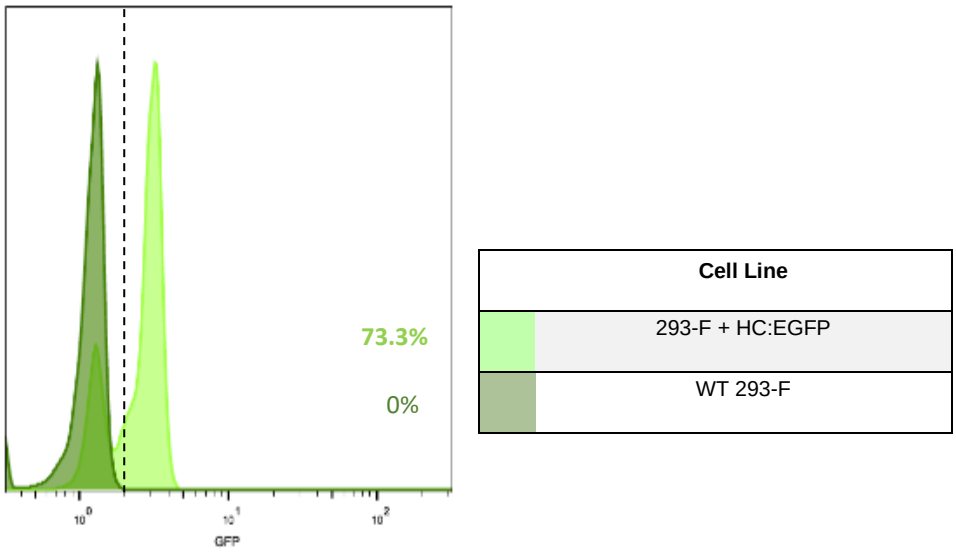


Figure 12| Flow Cytometry analysis of EGFP expression in 293-F_HC:EGFP. 100µg HC:EGFP lentiviral particles transduced in 5x10^5 293-F cells by lipofectamine™ 3000. After 24 hours cells were collected and examined by flow cytometry for EGFP expression. 293-F: FreeStyle™ 293-F cells; WT: wild type; HC: Heavy Chain; EGFP: enhanced green fluorescent protein.

4.4. CONSTRUCTION OF 293-F_HC_LC CELL LINE

Following the production of 293-F_HC:EGFP, there was a second phase transduction to complete the IgG formation by integrating the Light Chain in the same cell line. After lentiviral production by transfection in HEK293T, the LC:CD20 lentiviral particles were also incorporated in 293-F_HC:EGFP cell line by lipofectamine™ 3000. 24 hours later cells were harvested and indirectly stained with an antibody expressing the fluorescence of Alexa Fluor 405 expression. Figure 13 shows the flow cytometry results with a transduction efficacy of 83,3%.

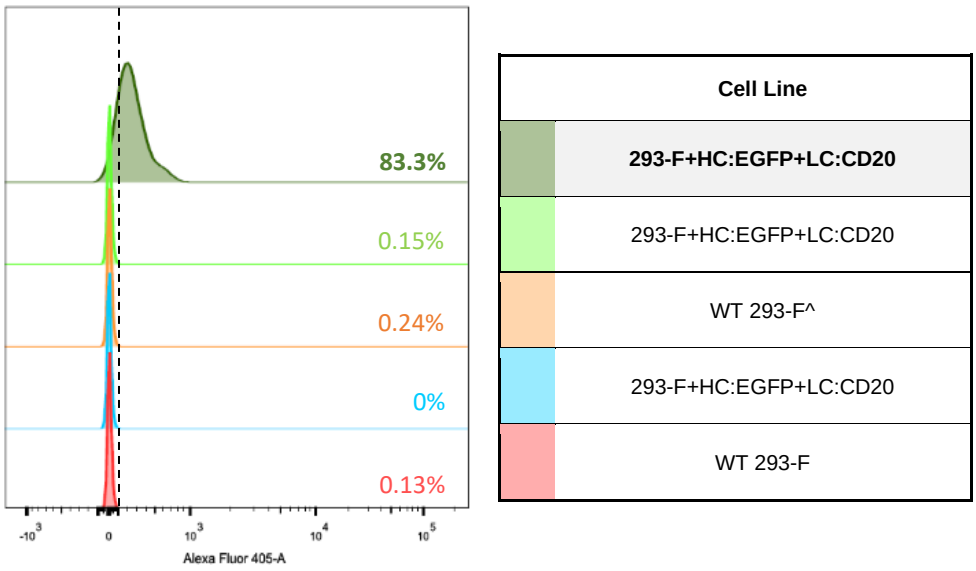


Figure 13| Flow Cytometry analysis of Alexa Fluor 405 expression in 293-F_HC:EGFP_LC:CD20. LC:CD20 lentiviral particles transduced in 293-F_HC:EGFP cell line by lipofectamine™ 3000. After 48 hours cells were collected, indirectly immunostained and examined by flow cytometry for Alexa Fluor 405 expression. 293-F+HC:EGFP+LC:CD20 were marked 1st and 2nd antibody, 293-F+HC:EGFP+LC:CD20 and WT 293-F^ were marked with 2nd antibody.

293-F: FreeStyle™ 293-F cells; WT: wild type; HC: Heavy Chain; EGFP: enhanced green fluorescent protein; LC: Light Chain; CD20: human T cell receptor cluster of differentiation 20.

4.5. CONSTRUCTION OF 293-F_HC_LC_CRISPR/CAS9 CELL LINE

Having the antibody construct integrated in 293-F cell line, the next step to construct the antibody library is to incorporate in 293-F_HC_LC cell line the CRISPR/Cas9 system correspondent to sgRNA 1 and 4, previously selected as ideal in the first step. Lentiviral nuclease constructs were transfected into HEK293T and then transduced orderly in 293-F_HC_LC, first CRISPR/Cas9_sgRNA 1 (C1) and then CRISPR/Cas9_sgRNA 4 (C4).

Prior to selection, there is a need to verify if the cell lines constructed are expressing the full antibody. To access this question, 48 hours past-transduction the cells were prepared and stained with an antibody anti-human IgG (Fab) - goat which targets the presence of IgG, emitting a fluorescent Dylight 650 signal. Figure 13 represents the flow cytometry results for the expression of Dylight 650, where 293-F_HC_LC_C1 (95,5%), 293-F_HC_LC_C4 (75,7%) and 293-F_HC_LC_C1_C4 (62,1%) cell lines obtained very positive expression values as expected.

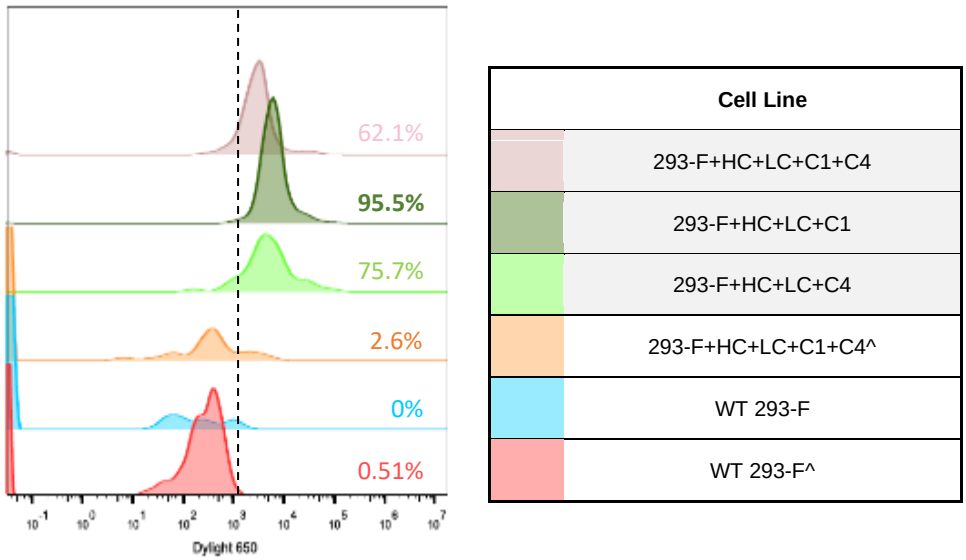


Figure 14| Flow Cytometry analysis of Dylight 650 expression in 293-F_HC_LC_CRISPR/Cas9. CRISPR/Cas9_sgRNA 1 and 4 lentiviral particles transduced in 293-F_HC_LC cell line by lipofectamine™ 3000. After 24 hours cells were collected, stained, and examined by flow cytometry for Dylight 650 expression. 293-F+HC+LC+C1+C4, 293-F+HC+LC+C1, 293-F+HC+LC+C4, WT 293-F were marked with Dylight 650. 293-F: FreeStyle™ 293-F cells; WT: wild type; HC: Heavy Chain; LC: Light Chain; C1: CRIPRS/Cas9 with a correspondent sgRNA 1: C4: CRIPRS/Cas9 with a correspondent sgRNA 4.

4.6. ANTI-STERPTAVIDIN LIBRARY DIVERSITY AND SELECTION

Established all the conditions to the creation of an antibody library, the following step is to identify if the system has the ability to generate antibodies with affinity against a desired target. Streptavidin protein was used for the objective, having attached the fluorescent dye PE-Cy7. In this early phase, the cell lines constructed 293-F_HC_LC_C1, 293-F_HC_LC_C4 and 293-F_HC_LC_C1_C4 were submitted to Streptavidin PE-Cy7 conjugate in a way of discovering any anti-streptavidin antibody generated. Figure 14 reveal interesting results of PE-Cy7 expression for two cell lines: 293-F_HC_LC_C1 (23%) and 293-F_HC_LC_C1_C4 (18,1%), however not so enticing results for the 293-F_HC_LC_C4 cell line (8%). Having a diversity generation system, it is normal that PE-Cy7 expression values were not so wide, also 293-F_HC_LC_C1 cell line showing a higher value than 293-F_HC_LC_C4 or 293-F_HC_LC_C1_C4 helps to prove the theory that it only takes one modification in one CDR to express variability.

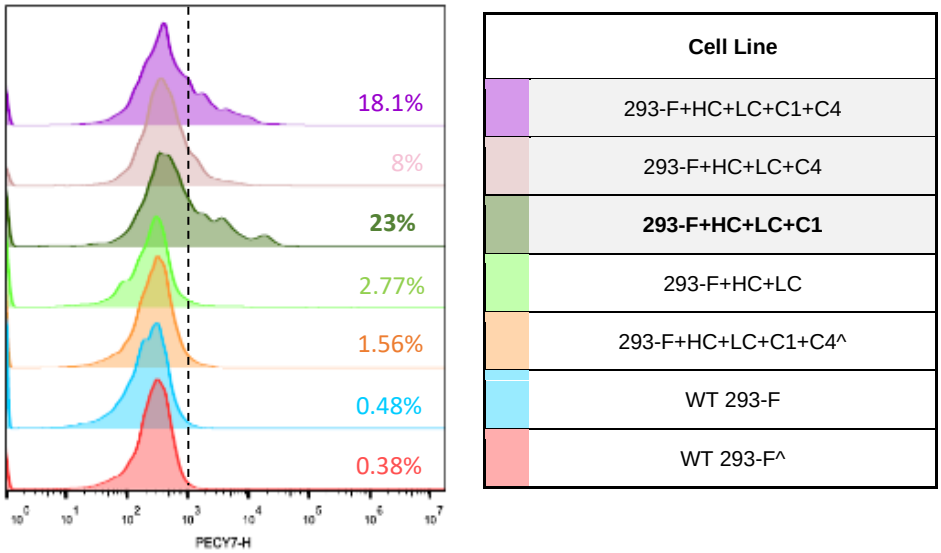


Figure 15| Flow Cytometry analysis of PE-Cy7 expression in 293-F_HC_LC_CRISPR/Cas9. After one week of culture, 293-F_CRISPR/Cas9 cell lines were collected and treated with Streptavidin PE-Cy7 conjugate to be examined by flow cytometry for PE-Cy7 expression. 293-F+HC+LC+C1+C4, 293-F+HC+LC+C4, **293-F+HC+LC+C1,** 293-F+HC+LC and WT 293-F were marked with Streptavidin PE-Cy7. 293-F: FreeStyle™ 293-F cells; WT: wild type; HC: Heavy Chain; LC: Light Chain; C1: CRIPRS/Cas9 with a correspondent sgRNA 1: C4: CRIPRS/Cas9 with a correspondent sgRNA 4.

Selection cell sorting was used to increase the population of interest, where the cell lines 293-F_HC_LC_C1_C4, 293-F_HC_LC_C1 and 293-F_HC_LC_C4 were incubated with Streptavidin PE-Cy7 conjugated and tested for the expression of anti-streptavidin antibodies at the cell surface. FACS Sorter achieved 93% of positive cells, which were cultured for two weeks and afterward submitted to protein purification by Dynabeads™ M-280 Streptavidin accordingly to the manufacturer's protocol.

Figure 16 unveils the flow cytometry results for PE-Cy7 expression using the cells after protein purification, with 293-F_HC_LC_C1_C4 cell line having the higher PE-Cy7 signal of 76%. 293-F_HC_LC_C1 (64,5%) and 293-F_HC_LC_C4 (61%) cell lines also presented great values of PE-Cy7 expression, showing again that an alteration in only one of the CDR's can generate diversity.

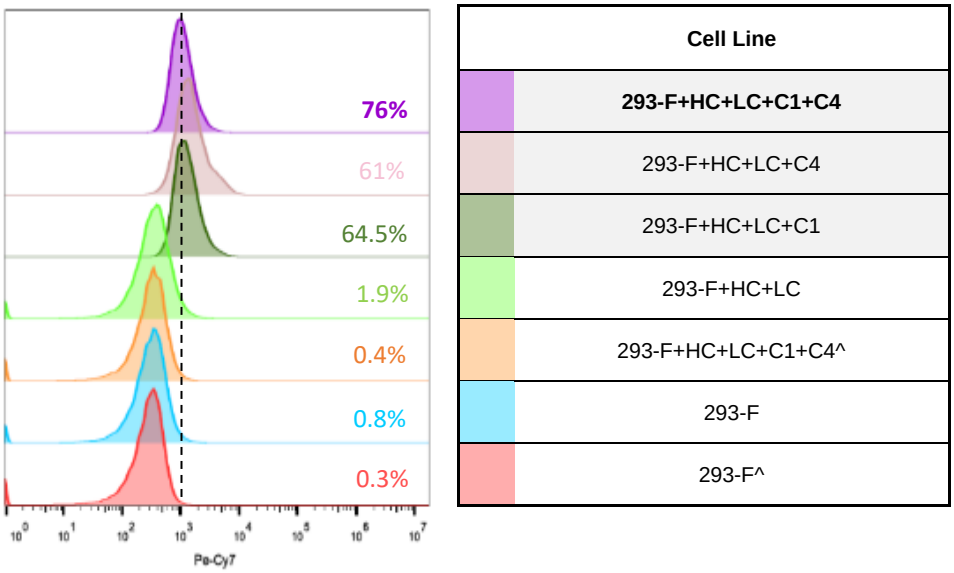


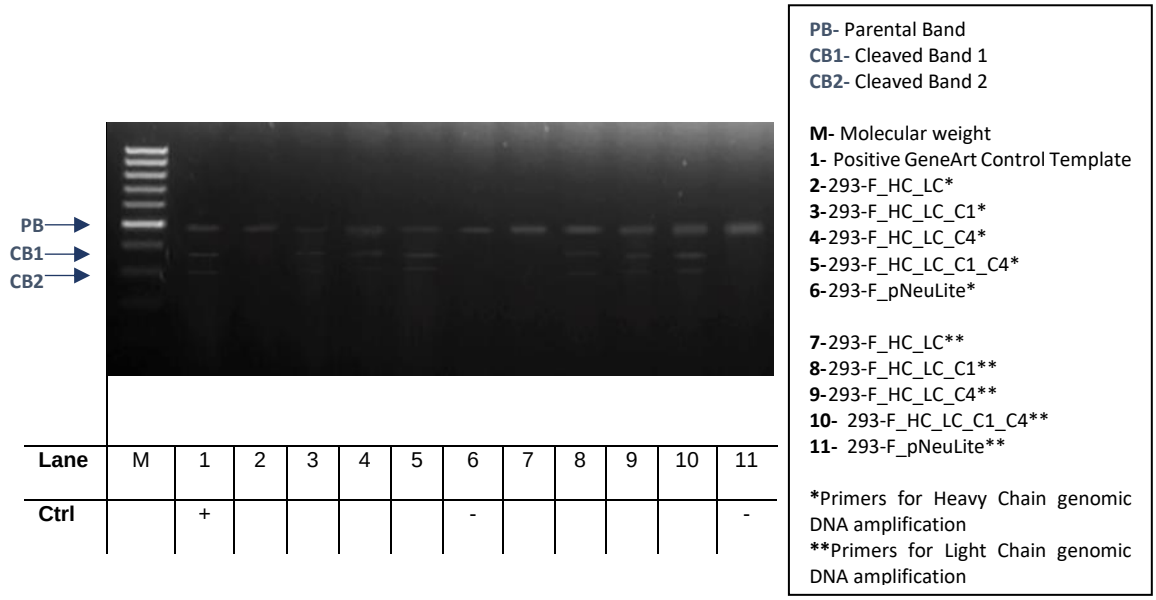
Figure 16| Flow Cytometry analysis of PE-Cy7 expression in 293-F_HC_LC_CRISPR/Cas9. After two weeks of culture, 293-F_CRISPR/Cas9 cell lines were submitted to protein purification by Dynabeads™ M-280 Streptavidin. Cells were collected and treated with Streptavidin PE-Cy7 conjugate to be examined by flow cytometry for PE-Cy7 expression. **293-F+HC+LC+C1+C4,** 293-F+HC+LC+C4, 293-F+HC+LC+C1, 293-F+HC+LC and 293-F were marked with PE-Cy7. 293-F: FreeStyle™ 293-F cells; WT: wild type; HC: Heavy Chain; LC: Light Chain; C1: CRIPRS/Cas9 with a correspondent sgRNA 1: C4: CRIPRS/Cas9 with a correspondent sgRNA 4.

4.7. ASSESSMENT OF CRISPR/CAS9 MUTAGENESIS

With previous results, it is possible to presume that antibody library diversity was achieved, and that was possible through CRISPR/Cas9 nuclease activity. To assess the veracity of these findings, a Genomic Cleavage Detection Kit (GeneArt®) was employed to identify any locus specific cleavage of genomic DNA from 293-F_HC_LC_CRISPR/Cas9 cell lines.

$1x10^5$ cells were harvested and lysed for genomic DNA extraction, used for the detection Kit. This kit amplifies by PCR the loci where the gene-specific double-strand break occurred, where the mismatches are detected and cleaved by its Detection Enzyme. Resultant bands are analyzed by gel electrophoresis (Figure 17), indicating the presence of genomic DNA cleavage in 293-F_HC_LC_CRISPR/Cas9 cell lines by portraying a similar lane profile as the Positive Kit control. Positive lanes profile results for confirmed gene modification efficiency consists of three bands: one Parental band (PB) that

represents the intact DNA and two other Cleaved bands (CB 1 and 2) that represent the cleavage product fraction.



Lane	M	1	2	3	4	5	6	7	8	9	10	11
Ctrl		+					-					-

Figure 17| Gel image of Genomic Cleavage Detection Assay using 293-F_HC_LC_CRISPR/Cas9 cells. 293-F_HC_LC cells transduced with CRISPR/Cas9 sgRNA 1 and(or) 4 by lipofectamine 3000™. Negative control for gene modification was prepared by transfecting 293-F with pNeuLite. The above samples were PCR amplified using the same set of primers flanking the region of interest. 293-F: FreeStyle™ 293-F cells; HC: Heavy Chain; LC: Light Chain; C1: CRIPRS/Cas9 with a correspondent sgRNA 1: C4: CRIPRS/Cas9 with a correspondent sgRNA 4. pNeuLite: plasmid with the promoter region of human gene Her-2/neu.

To ensure that CRISPR/Cas9 nuclease was present and active, 293-F_HC_LC_CRISPR/Cas9 cell lines were collected and submitted to western blot analysis. Figure 18 reveal a CRISPR/Cas9 positive activity for 293-F cell lines that contain the expressed antibody combined with Cas9 nuclease, which is shown by the different profile presented between a cell with and without Cas9. The negative control represented by 293-F_HC_LC without Cas9 has a lane with two perfect bands corresponding to the usual IgG profile: one heavy chain band at 50 kDa and one light chain band at 25 kDa. Whereas 293-F_HC_LC_CRISPR/Cas9 cell lines expose different results, not having any (293-F_HC_LC:C1; 293-F_HC_LC:C1:C4) or a slightest heavy/light chain bands (293-F_HC_LC:C4).

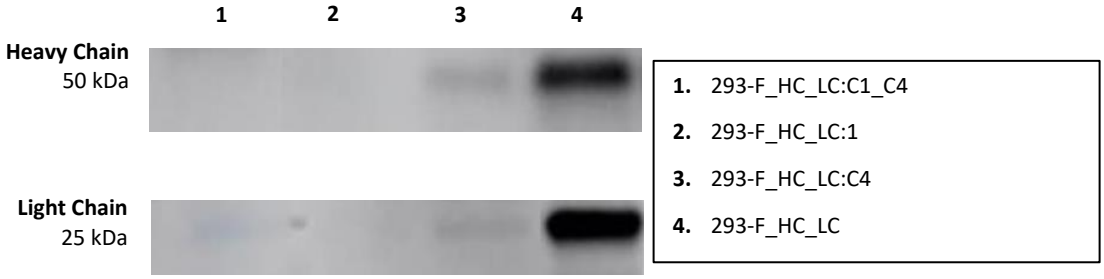


Figure 18| Western blot analysis of IgG antibody constructs present in the protein lysates of the 293-F cell lines to evaluate CRISPR/Cas9 efficiency for the generation of double-strand breaks. 293-F cells were transduced with Heavy and Light chain constructs plus the desired CRIPR/Cas9 sgRNA 1/4. Negative control for gene modification was prepared by transducing only Heavy and Light chains. 293-F: FreeStyle™ 293-F cells; HC: Heavy Chain; LC: Light Chain; C1: CRIPRS/Cas9 with a correspondent sgRNA 1: C4: CRIPRS/Cas9 with a correspondent sgRNA 4.

5 DISCUSSION

The humoral immune response recognizes a wide range of antigens and generates an antibody library with an immense number of distinct antibodies, either by the immune system or *in vitro* approaches. This variety is mediated by the presence of unique CDRs in heavy and light chain variable regions of IgG[6]. The *in vitro* antibody discovery technologies revolutionized the generation of target-specific antibodies that traditionally relied on the humoral response of immunized animals. Recently, *In vitro* technologies are highly affected to generate target-specific antibodies by easily isolating them from large antibody libraries, as the probability of finding a desirable antibody is proportional to the repertoire size.

The large sequence diversity in libraries originated from natural sources leads to large variability in physical properties, stability and expression levels in the pool of antibodies[19]. Meanwhile, a synthetic approach allows the construction of libraries built on a chosen limited set of sequences heading to more improved and consistent characteristics. Regardless of their lower structural variety, the possibility to select antibody fragments with high specificity and high affinity against any antigen from such libraries has been proven[125–127].

Framework's nature is another aspect that highly influences libraries diversity. Natural antibody libraries originate from the same source, being complex combinations of antibody clones that diverge from one another in their biophysical and biochemical properties. This results in unequal propagation and privileged enrichment of fast-growing clones and predispositions the selection output toward highly expressed clones rather than high affinity clones. Several synthetic antibody libraries use a single framework sequence, creating a simpler library design and construction, making clones more uniform in their properties. Sustaining this idea many reports are showing that large synthetic antibody libraries with a single or very limited number of framework sequences are capable of generating antibodies against diverse antigens and epitopes[125,128,129].

With this project, it was possible to create a synthetic antibody library from a Human antibody backbone with pre-defined CDRs and induced variability through CRISPR/Cas9 system. Developing this library CDR1 and CDR3 were engineered to be replaced by DNA sequences recognized by CRISPR/Cas9, CDR2 and the framework was maintained from the original antibody. Approaching the CDRs is a sharp tactic given that the majority of somatic mutations affecting affinity are mostly located in them rather

than in the frameworks. This decreases the probability of generating undesirably mutations that disturb antibody stability.

Primarily was evaluated the cutting efficacy of CRISPR/Cas9 nuclease, where we tested 5 different sgRNA's correspondent to different 20 base pair sequences from *gfp* gene. For this test, two antibody variants were designed with hotspots targeted by sgRNA EGFP 1 in the place of CDR1 (V_L) and by sgRNA EGFP 2,4,5,6 in the place of CDR3 (V_L*****), then the constructs were subjected to CRISPR/Cas9 nuclease system. Interpretation of western blot results reveals different cutting activities between the various CRISPR/Cas9 sgRNA's. sgRNA EGFP 2,5 and 6 did not reveal the ability to create double-strand breaks, on the other hand, sgRNA EGFP 1 showed high effectiveness in cutting the DNA and sgRNA EGFP 4 a slightly less cutting ability. Even though the results obtained showed varied cutting abilities for the five sgRNA's, the literature presented all of them as perfectly capable of creating double-strand breaks in the DNA. Is important to remember that all target sequences that exhibited different cutting activities were all introduced in the same CDR3, which indicates that the different reaction of Cas9/sgRNA's complexes is not justifiable by the accessibility to the target sequence. An explanation that could be counted for the diverse cutting activities is the exact location of PAM, read-out by the Cas9 protein to initiate base-pairing between sgRNA and DNA target. Correspondingly to sgRNA EGFP 1, the PAM sequence is immediately after the DNA target sequence, while on the other sgRNA's the PAM sequence is placed 3/4 nucleotides after the end of the DNA target sequence.

Although the results do not appear exactly as the literature, it was possible to fulfill the aim of finding two CRISPR/Cas9 systems that showed the highest ability to implement double-strand breaks in the DNA. Consequently, two separate vectors were designed for expression of light chain and heavy chain for IgG production both covering a hotspot for sgRNA EGFP 1 CRISPR/Cas9 in CDR1 and a hotspot for sgRNA EGFP 4 CRISPR/Cas9 in CDR3, to be submitted to the Cas9 activity. Such as the breaks required during V(D)J recombination and class switch recombination, the DSBs created by CRISPR/Cas9 system are very important to generate variability for the desired library. The repair mechanism by NHEJ, predominant in mammalian cells, results in small insertions and deletions (indels) relative to the original genomic template which contributes to the generation of antibody variability inside B cells[130].

Hereupon, the second goal was to create an *in vitro* platform for the generation of an antibody library capable of mimic the variability creation that happens inside B cells. Before the construction of the antibody library in 293-F cell line, there was a need to test

HEK 293T transfection protocol with different amounts and proportions of DNA and also test distinct conditions during the transduction procedure in Jurkat E6.1.

To achieve perfect transduction conditions, it was first engineered a T cell line to express the modified IgG with the hotspots for the chosen sgRNA CRISPR/Cas9 nuclease. Afterward, those ideal conditions were applied when developing the library in a 293-F cell line. Following the activation of Cas9 nuclease, the DSB created will consequently initiate a random repair mechanism that will generate the antibody library. The possibility of controlling the number of 293-F cells population subjected to Cas9 is important to regulate the library size.

293-F cell line was transduced first with Heavy Chain and Light Chain plasmids and lastly with CRISPR/Cas9. Heavy and Light Chains were cloned in two different vectors in favor of the CRISPR/Cas9 activity, since there is a need to induce CRISPR/Cas9 at the same time in the two antibody chains which is only possible having them expressed in different areas of the genome. Construction of the HC_LC_CRISPR/Cas9 293-F cell line was successful, actively demonstrating expressions of full IgG antibodies, with the peculiarity of having different values in the presence of sgRNA EGFP 1 (95.5%), sgRNA EGFP 4 (75.7%) or both (62.1%), and also showing a positive activity of Cas9 nuclease.

For proof of concept, Streptavidin protein, having attached the fluorescent dye PE-Cy7, was used to verify if the system created had the ability to generate antibodies with affinity against the desired target. Given that it was used two different sgRNA to target DNA, distinct expression efficiencies were also reflected (293-F_HC_LC_C1 = 23%; 293-F_HC_LC_C1_C4 = 18%; 293-F_HC_LC_C4 = 8%). After sorter, it was possible to select and maintain a population of cells that presented anti-streptavidin antibodies with higher expression PE-Cy7 values (293-F_HC_LC_C1 = 64,5%; 293-F_HC_LC_C1_C4 = 76%; 293-F_HC_LC_C4 = 61%), presuming that this sorted cells have been subjected to the DSBs and consequently to the process of NHEJ and random mutagenesis.

The final step in this process is a functional evaluation. Since NHEJ usually produces insertion and deletion mutations, to detect if the transfection was successful in incorporating the right CRISPR/Cas9 plasmid, it was used the SURVEYOR nuclease assay (GeneArt® Kit) to detect endogenous target cleavage. The SURVEYOR nuclease assay detects mismatch repairs and cleaves them, where the product can then be analyzed using gel electrophoresis. Results from the genomic cleavage assay proved that 293-F cells were correctly transfected with CRISPR/Cas9 nuclease plasmid and consequently this system has created DSB repaired by NHEJ mechanism. These

findings are proven by the similar lane profile between the positive control provided on the kit and the samples at test. To complement this result, a western blot assay verified that the prevailing Cas9 was still actively cutting the target DNA.
It is also important to further analyze the edited genomic sequence for potential frameshifts, truncations, alternate AUG codons, and splicing alterations. Next-Generation Sequencing (NGS) is a strong option tool to evaluate the library variability, screening a wide number of sequences present in the antibody library. With NGS we would be able to understand the exact efficiency of CRISPR/Cas9 and which percentage of the cells carried a different DNA sequence than the original antibody. More importantly, would be to comprehend if these mutations occurred specifically inside the CDR1 (hotspot for sgRNA EGF 1), or CDR3 (hotspot for sgRNA EGF 4).

Generally, we were able to prove that it is possible to create a large repertoire of antibodies with variability inside mammalian cells, in which they create their own diversity. These finds may provide a great impact in establishing a powerful platform for antibody library generation. The fact that we used Human IgG to create this platform, gives us the ability to surpass significant issues, like toxicity and self-tolerance, associated with animal immunizations, or complications regarding humanization and glycosylation.
This unique platform represents a major progress for the development of highly diverse libraries, which could settle significant conventional challenges.

6 CONCLUDING REMARKS

Genome-editing tools, such as ZFN and TALENs, have been critical in the development of novel therapeutics over the past decades. However, the quick development and application of CRISPR/Cas9 have been reflected as a new promise to drug discovery. CRISPR/Cas9 technology has become the desirable choice for genome editing in the laboratory, which seems to increase the ability to perform systematic analyses of gene function, replicate animal models for human diseases phenotypes and use as a tool for gene therapy and screening of drug target candidate genes. Despite such potential, further improvements of this method are necessary to minimize off-target effects, improve its efficacy in primary cells and formulate safety guidelines. Ethical issues are a main concern and also need to be addressed rapidly for the implementation of this evolving technology.

For the current project, the idea was to mimic the V(D)J recombination process that occurs inside B cells to generate diversity. The recombinational process consists of a random selection of a pair of V, D, J segments, the introduction of double-strand breaks adjacent to each segment and repair of the intervening DNA by recombination or introducing small insertions and deletions, which contributes to immunoglobulin diversity in vertebrate immune systems. With the recreation of V(D)J mechanism in a mammalian cell system, it was possible to generate *in situ* diversity without the transformation of recombinant libraries. The location of the diversity process was controlled using a well-known antibody backbone designed to include specific DNA sequences inside the CDRs, recognized by CRISPR/Cas9 nuclease.

We opted for the generation of fully Human antibodies to overcome the need for a humanization and consequently immunogenicity and tolerance.

Overall, it may be said that the results attained in this project validate the proof of concept of a distinct and successful approach capable of creating an antibody library with its own variability. From a single antibody sequence, it was possible to create a diverse library that was exposed to selection mechanisms to discard nonfunctional antibody forms, and intelligently selecting only good binders. This platform increases the likelihood of a selected antibody being well-tolerated and highly effective when employed for the development of therapeutic solutions in humans.

7 REFERENCES

1. Charles A. Janeway Jr and RM. INNATE IMMUNE RECOGNITION. *Annual Review of Immunology*. 2002;20:197-216.

2. Sompayrac L. *How the Immune System Works*. Fifht Edi. Wiley Blackwell; 2015. doi:10.1016/S1081-1206(10)63314-5

3. Medzhitov R, Janeway C. Innate immunity. *The New England Journal of Medicine*. 2000;343(5):338-344. doi:10.1056/NEJM200008033430506

4. Maul RW, Gearhart PJ. Controlling somatic hypermutation in immunoglobulin variable and switch regions. *immunol Res*. 2010;47(1-3):113-122. doi:10.1007/s12026-009-8142-5.Controlling

5. Amzel M, Poljak RJ. Three-dimensional structure of immunoglobulins. *Annual Reviews*. 1979;(48):961-997.

6. Kasahara M, Suzuki T, Du Pasquier L. On the origins of the adaptive immune system: Novel insights from invertebrates and cold-blooded vertebrates. *Trends in Immunology*. 2004;25(2):105-111. doi:10.1016/j.it.2003.11.005

7. Weaver DT. *V(D)J Recombination and Double-Strand Break Repair*. Vol 58. Elsevier Masson SAS; 1995. doi:10.1016/S0065-2776(08)60619-7

8. Dudley DD, Chaudhuri J, Bassing CH, Alt FW. Mechanism and control of V(D)J recombination versus class switch recombination: Similarities and differences. *Advances in Immunology*. 2005;86:43-112. doi:10.1016/S0065-2776(04)86002-4

9. Maizels N. Recombination in the Immune System. *Encyclopedia of Genetics*. Published online 2001:1640-1642. doi:10.1016/B978-0-12-374984-0.01278-X

10. Martin A, Chahwan R, Parsa JY, Scharff MD. Somatic Hypermutation: The Molecular Mechanisms Underlying the Production of Effective High-Affinity Antibodies. In: Frederick W. Alt, Tasuku Honjo, Andreas Radbruch MR, ed. *Molecular Biology of B Cells*. 2nd Edi. Elsevier Ltd; 2015:363-388. doi:10.1016/B978-0-12-397933-9.00021-7

11. Benhar I. Design of synthetic antibody libraries. *Expert Opinion on Biological Therapy*. 2007;7(5):763-779. doi:10.1517/14712598.7.5.763

12. Ponsel D, Neugebauer J, Ladetzki-Baehs K, Tissot K. High affinity, developability and functional size: The holy grail of combinatorial antibody library generation. *Molecules*. 2011;16(5):3675-3700. doi:10.3390/molecules16053675

13. Hoogenboom HR. Selecting and screening recombinant antibody libraries. *Nature Biotechnology*. 2005;23(9):1105-1116. doi:10.1038/nbt1126

14. Mollova S, Retter I, Hust M, Dübel S, Müller W. Analysis of Single Chain Antibody Sequences Using the VBASE2 Fab Analysis Tool. In: Kontermann R, Dübel S, eds. *Antibody Engineering Vol. 2*. Second Edi. Springer; 2010:3-10.

15. Nelson B, S. Sidhu S. Synthetic Antibody Libraries. In: Voynov V, Caravela J, editors. Therapeutic Proteins: Methods and Protocols [Internet]. 2nd Edition. Springer Science; 2012. p. 27–41. Available from: http://www.springer.com/series/7651

16. Nemazee D, Hogquist KA. Antigen receptor selection by editing or downregulation of V(D)J recombination. *Current Opinion in Immunology*. 2003;15(2):182-189. doi:10.1016/S0952-7915(03)00008-6

17. George G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Stephen RQ. The promise and challenge of high-throughput sequencing of the antibody repertoire. 2007;32(2):158-168. doi:10.1038/nbt.2782.The

18. Pastan MH and I, Summary. Antibody phage display - methods and protocols 2nd Edition. In: Aitken R, ed. *Antibody Phage Display*. 2nd Edi. Human Press; 2011:99-114. doi:10.1146/annurev.physiol.66.032902.131835

19. Wark KL, Hudson PJ. Latest technologies for the enhancement of antibody affinity. *Advanced Drug Delivery Reviews*. 2006;58(5-6):657-670. doi:10.1016/j.addr.2006.01.025

20. Omidfar K, Daneshpour M. Advances in phage display technology for drug discovery. *Expert Opinion on Drug Discovery*. 2015;10(6):651-669. doi:10.1517/17460441.2015.1037738

21. Inc. SB. Phage Display Overview. Published online 2017.

22. He M, Khan F. Ribosome display: Next-generation display technologies for production of antibodies in vitro. *Expert Review of Proteomics*. 2005;2(3):421-430. doi:10.1586/14789450.2.3.421

23. Murray CJ, Baliga R. Cell-free translation of peptides and proteins: From high throughput screening to clinical production. *Current Opinion in Chemical Biology*. 2013;17(3):420-426. doi:10.1016/j.cbpa.2013.02.014

24. Magdassi S, Bernath K, Miller OJ, et al. Directed evolution by in vitro compartmentalization. *Nature Methods*. 2006;3(7):561-570. doi:10.1038/nmeth897

25. Seelig B. MRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nature Protocols*. 2011;6(4):540-552. doi:10.1038/nprot.2011.312

26. Qin CF, Li GC. Mammalian cell display technology coupling with AID induced SHM in vitro: An ideal approach to the production of therapeutic antibodies. *International Immunopharmacology*. 2014;23(2):380-386. doi:10.1016/j.intimp.2014.09.017

27. Daugherty PS, Iverson BL, Georgiou G. Flow cytometric screening of cell-based libraries. *Journal of Immunological Methods*. 2000;243(1-2):211-227. doi:10.1016/S0022-1759(00)00236-2

28. Jose J. Autodisplay: Efficient bacterial surface display of recombinant proteins. *Applied Microbiology and Biotechnology*. 2006;69(6):607-614. doi:10.1007/s00253-005-0227-z

29. Löfblom J. Bacterial display in combinatorial protein engineering. *Biotechnology Journal*. 2011;6(9):1115-1129. doi:10.1002/biot.201100129

30. Sazinsky SL, Hackel BJ, Lippow SM, Chao G, Lau WL, Wittrup KD. Isolating and engineering human antibodies using yeast surface display. *Nature Protocols*. 2006;1(2):755-768. doi:10.1038/nprot.2006.94

31. Thomas K, Johannes P, Thomas FM. The secretion pathway of IgA protease-type proteins in gram-negative bacteria. *BioEssays*. 1993;15(12):799-805.

32. Bostrom J, Lee CV, Haber L, Fuh G. Mammalian Cell Display for Antibody Engineering Mitchell. *Methods in molecular biology (Clifton, NJ)*. 2009;525(301):353. doi:10.1007/978-1-59745-554-1

33. Ito W, Iba Y, Kurosawa Y. Effects of substitutions of closely related amino acids at the contact surface in an antigen-antibody complex on thermodynamic parameters. *Journal of Biological Chemistry*. 1993;268(22):16639-16647.

34. Feldhaus MJ, Siegel RW, Opresko LK, et al. Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature Biotechnology*. 2003;21(2):163-170. doi:10.1038/nbt785

35. Segal DJ, Meckler JF. Genome Engineering at the Dawn of the Golden Age. *Annual Review of Genomics and Human Genetics*. 2013;14(1):135-158. doi:10.1146/annurev-genom-091212-153435

36. Gouble A, Duchateau P, Arnould S, et al. Engineered I-CreI Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells. *Journal of Molecular Biology*. 2007;371(1):49-65. doi:10.1016/j.jmb.2007.04.079

37. Lee HJ, Kweon J, Kim E, Kim S, Kim JS. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Cold Spring Harbor Laboratory Press*. 2012;22(3):539-548. doi:10.1101/gr.129635.111

38. Schmidt C, Voytas DF, Hummel A, et al. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*. 2010;186(2):757-761. doi:10.1534/genetics.110.120717

39. Cong L, Ran FA, Cox D, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. 2013;339(6121):819-823. doi:10.1126/science.1231143.Multiplex

40. Stoddard BL. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure*. 2012;19(1):7-15. doi:10.1016/j.str.2010.12.003.Homing

41. Mussolino C, Janssen JM, Maggio I, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*. 2012;41(5):e63-e63. doi:10.1093/nar/gks1446

42. Duchateau P, Galetto R, Montoya G, et al. Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Current Gene Therapy*. 2011;11(1):11-27. doi:10.2174/156652311794520111

43. Redel BK, Prather RS. Meganucleases Revolutionize the Production of Genetically Engineered Pigs for the Study of Human Diseases. *Toxicologic Pathology*. 2016;44(3):428-433. doi:10.1177/0192623315613160

44. Scott CT. The zinc finger nuclease monopoly. *Nature Biotechnology*. 2005;23(8):915-918. doi:10.1038/nbt0805-915

45. Wolfe SA, Nekludova L, Pabo CO. DNA Recognition by Cys2His2 Zinc Finger Proteins. *Annu Rev Biophys Biomol Struct*. 2000;29(3):183-212.

46. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. *Nature*. 2009;461(7267):1071-1078. doi:10.1038/nature08467

47. Pavletich NP, Pabo C 0. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 A. *The Poems of Sir Philip Sidney*. 1991;(May):809-817. doi:10.1093/oseo/instance.00007554

48. Perez EE, Wang J, Miller JC, et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol*. 2012;26(7):808-816.

49. Bogdanove AJ, Schornack S, Lahaye T. TAL effectors: Finding plant genes for disease and defense. *Current Opinion in Plant Biology*. 2010;13(4):394-401. doi:10.1016/j.pbi.2010.04.010

50. Baller JA, Doyle EL, Wang L, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*. 2011;39(12):e82-e82. doi:10.1093/nar/gkr218

51. Lütge F, Dannemann N, Mussolino C, Morbitzer R, Cathomen T, Lahaye T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research*. 2011;39(21):9283-9293. doi:10.1093/nar/gkr597

52. Miller JC, Zhang L, Xia DF, et al. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nature Methods*. 2015;12:465-471. doi:10.1038/nmeth.3330

53. Bolotin A, Quinquis B, Sorokin A, Dusko Ehrlich S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005;151(8):2551-2561. doi:10.1099/mic.0.28048-0

54. Wilen CB, Lee S, Hsieh L yang, et al. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. 2018;360(6385):204-208. doi:10.1126/science.aar3799.Tropism

55. Jinek M, Jinek M, Jiang F, et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*. 2014;343(6176):1-28. doi:10.1126/science.1247997

56. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2015;513(7519):569-573. doi:10.1038/nature13579.Structural

57. Nishimasu H, Ran FA, Hsu PD, et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*. 2014;156(5):935-949. doi:10.1002/cncr.27633.Percutaneous

58. Ferreira R, David F, Nielsen J. Advancing biotechnology with CRISPR/Cas9: recent applications and patent landscape. *Journal of Industrial Microbiology and Biotechnology*. 2018;45(7):467-480. doi:10.1007/s10295-017-2000-6

59. Weber J, Öllinger R, Friedrich M, et al. CRISPR/Cas9 somatic multiplex-mutagenesis for high-throughput functional cancer genomics in mice. *Proceedings of the National Academy of Sciences*. 2015;112(45):13982-13987. doi:10.1073/pnas.1512392112

60. Wang J, Exline CM, Declercq JJ, et al. Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nature Biotechnology*. 2015;33(12):1256-1263. doi:10.1038/nbt.3408

61. Nelson CE, Hakim CH, Ousterout DG, et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*. 2016;351(6271):403-407. doi:10.1126/science.aad5143.In

62. Suzuki K, Tsunekawa Y, Hernandez-Benitez R, et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration generated minicircle DNA vectors performed single-cell genotyping HHS Public Access. *Nature*. 2016;540(7631):144-149. doi:10.1038/nature20565

63. Yang Y, Wang L, Bell P, et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nature Biotechnology*. 2016;34(3):334-338. doi:10.1038/nbt.3469.A

64. Liang P, Xu Y, Zhang X, et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein and Cell*. 2015;6(5):363-372. doi:10.1007/s13238-015-0153-5

65. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs developed the updated version of ZiFiT software HHS Public Access Author manuscript. *Nat Biotechnol*. 2014;32(3):279-284. doi:10.1038/nbt.2808

66. Slaymaker IM, Yan WX, Gao L, Zhang F, Scott DA, Zetsche B. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2015;351(6268):84-88. doi:10.1126/science.aad5227

67. Gersbach CA. Comparing Genome Editing Technologies. *Vol4 No 5*. Published online 2014.

68. Moynahan ME, Jasin M. Mitotic homologous recombination maintains genomic stability. *Nat Rev Mol Cell Biol*. 2010;11(3):196-207. doi:10.1038/nrm2851

69. Shim H. Synthetic approach to the generation of antibody diversity. *BMB Reports*. 2015;48(9):489-494. doi:10.5483/BMBRep.2015.48.9.120

70. Walker LM, Pandita TK, Helmink BA, et al. MRN complex function in the repair of chromosomal Rag-mediated DNA double-strand breaks. *The Journal of Experimental Medicine*. 2009;206(3):669-679. doi:10.1084/jem.20081326

71. Papavasiliou FN, Schatz DG. Cell-cycle-regulated DNA double-strand breaks in somatic hypermutation of immunoglobulin genes. *Nature*. 2000;408(6809):216-221. doi:10.1038/35041599

72. Chen JE, Glover GH. AID-Dependent Generation of Resected Double-Strand DNA Breaks and Recruitment of Rad52/Rad51 in Somatic Hypermutation. *Immunity*. 2003;18(6):727-738. doi:10.1007/s11065-015-9294-9.Functional

73. Kowalczykowski SC. An overview of the molecular mechanismsof recombinational DNA repair. *Cold Spring Harbor Perspectives in Biology*. 2015;7(11):1-36. doi:10.1101/cshperspect.a016410

74. Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews Molecular Cell Biology*. 2017;18(8):495-506. doi:10.1038/nrm.2017.48

75. Symington LS, Gautier J. Double-Strand Break End Resection and Repair Pathway Choice. *Annual Review of Genetics*. 2011;45(1):247-271. doi:10.1146/annurev-genet-110410-132435

76. Ira G, Pellicioli A, Balijja A, et al. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature*. 2004;431(7011):1011-1017. doi:10.1038/nature02964.DNA

77. Huertas P, Cortés-ledesma F, Sartori AA, Aguilera A, Jackson SP. CDK targets Sae2 to control DNA-end resection and homologous recombination. *Cancer Research*. 2008;455(7213):689-692. doi:10.1038/nature07215.CDK

78. Gottlieb TM, Jackson P. The DNA-Dependent Protein Kinase for DNA Ends and Association with Ku Antigen. *Cell*. 1993;72(1):131-142. doi:doi:10.1016/0092-8674(93)90057-W

79. Ma Y, Pannicke U, Schwarz K, Lieber MR. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*. 2002;108(6):781-794. doi:10.1016/S0092-8674(02)00671-2

80. Löbrich M, Jeggo P. A Process of Resection-Dependent Nonhomologous End Joining Involving the Goddess Artemis. *Trends in Biochemical Sciences*. 2017;42(9):690-701. doi:10.1016/j.tibs.2017.06.011

81. Moon AF, Pryor JM, Ramsden DA, Kunkel TA, Bebenek K, Pedersen LC. Sustained active site rigidity during synthesis by human DNA polymerase μ. *Nat Struct Mol Biol*. 2014;21(3):253-260. doi:10.1158/0008-5472.CAN-10-4002.BONE

82. Shahar OD, Hareli S, Goldberg M, Meshorer E, Shimshoni E, Ram EVSR. Live imaging of induced and controlled DNA double-strand break formation reveals extremely low repair by homologous recombination in human cells. *Oncogene*. 2012;31(30):3495-3504. doi:10.1038/onc.2011.516

83. Beucher A, Birraux J, Tchouandong L, et al. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO Journal*. 2009;28(21):3413-3427. doi:10.1038/emboj.2009.276

84. Karanam K, Kafri R, Loewer A, Lahav G. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. *Molecular Cell*. 2012;47(2):320-329. doi:10.1016/j.molcel.2012.05.052

85. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)*. 2008;7(10):1765-1771. doi:10.1016/j.dnarep.2008.06.018.Comparison

86. Rothkamm K, Kru I, Thompson LH, Lo M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Molecular and Cellular Biology*. 2003;23(16):5706-5715. doi:10.1128/MCB.23.16.5706

87. Seol JH, Shim EY, Lee SE. Microhomology-mediated end joining: Good, bad and ugly. *Mutation Research*. 2018;809:81-87. doi:10.1016/j.mrfmmm.2017.07.002

88. Mimitou EP, Symington LS. Ku prevents Exo1 and Sgs1-dependent resection of DNA ends in the absence of a functional MRX complex or Sae2. *EMBO Journal*. 2010;29(19):3358-3369. doi:10.1038/emboj.2010.193

89. von Behring, E. & Kitasato S. The mechanism of diphtheria immunity and tetanus immunity in animals. 1890. *Mol Immunol*. 1991;28(1317):1319-1320. doi:https://doi.org/10.17192/eb2013.0164

90. Reichert JM. The Antibody Society. Antibody therapeutics approved or in regulatory review in the EU or US.

91. Watier H, Reichert JM. Evolution of Antibody Therapeutics. In: Vaughan T, Osbourn J, Jalla B, eds. *Protein Terapeutics*. Wiley-VCH Verlag GmbH & Co; 2017:25-49.

92. Kempeni J. Preliminary results of early clinical trials with the fully human anti- TNFα monoclonal antibody D2E7. *Annals of the Rheumatic Diseases*. 1999;58(SUPPL. 1):170-172. doi:10.1136/ard.58.2008.i70

93. Grand View Research Inc. *Monoclonal Antibodies Market Size, Share, Industry Growth Report 2024*.; 2016.

94. Co KK. *Consolidated Financial Summary (IFRS) Fiscal 2019*. Vol 4151.; 2019.

95. Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. *Cell*. 2018;172(6):1239-1259. doi:10.1016/j.cell.2017.11.032

96. Kwon D. A brief guide to the current CRISPR landscape. *The Scientist*. 2019.

97. Amoasii L, Long C, Li H, et al. Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Science Translational Medicine*. 2017;9(418):756-760. doi:10.1126/scitranslmed.aan8081

98. DeWitt MA, Magis W, Bray NL, et al. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Science Translational Medicine*. 2016;8(360). doi:10.1126/scitranslmed.aaf9336

99. Niu X, He W, Song B, et al. Combining single strand oligodeoxynucleotides and CRISPR/Cas9 to correct gene mutations in β-thalassemia-induced pluripotent stem sells. *Journal of Biological Chemistry*. 2016;291(32):16576-16585. doi:10.1074/jbc.M116.719237

100. Yin H, Xue W, Chen S, et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature Biotechnology*. 2014;32(6):551-553. doi:10.1038/nbt.2884

101. Zuckermann M, Hovestadt V, Knobbe-Thomsen CB, et al. Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nature Communications*. 2015;6(May):1-9. doi:10.1038/ncomms8391

102. Ha JS, Byun J, Ahn DR. Overcoming doxorubicin resistance of cancer cells by Cas9-mediated gene disruption. *Scientific Reports*. 2016;6(October 2015):1-7. doi:10.1038/srep22847

103. Saey TH. CRISPR enters its first human clinical trials. *Science News*. 2019.

104. Ghosh D, Venkataramani P, Nandi S, Bhattacharjee S. CRISPR-Cas9 a boon or bane: The bumpy road ahead to cancer therapeutics 06 Biological Sciences 0604 Genetics. *Cancer Cell International*. 2019;19(12):1-10. doi:10.1186/s12935-019-0726-0

105. Koo T, Lee J, Kim J soo. Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Molecules and Cells*. 2015;38(6):475-481.

106. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Cold Spring Harbor Laboratory Press*. 2014;24:132-141. doi:10.1101/gr.162339.113.Freely

107. You L, Tong R, Li M, Liu Y, Xue J, Lu Y. Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Molecular Therapy - Methods and Clinical Development*. 2019;13(June):359-370. doi:10.1016/j.omtm.2019.02.008

108. Feng C, Su H, Bai H, et al. High-efficiency genome editing using a dmc1 promoter-controlled CRISPR/Cas9 system in maize. *Plant Biotechnology Journal*. 2018;16(11):1848-1857. doi:10.1111/pbi.12920

109. Lee K, Zhang Y, Kleinstiver BP, et al. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnology Journal*. 2019;17(2):362-372. doi:10.1111/pbi.12982

110. Eş I, Gavahian M, Marti-Quijal FJ, et al. The application of the CRISPR-Cas9 genome editing machinery in food and agricultural science: Current status, future perspectives, and associated challenges. *Biotechnology Advances*. 2019;37(3):410-421. doi:10.1016/j.biotechadv.2019.02.006

111. Anderson JE, Michno JM, Kono TJY, et al. Genomic variation and DNA repair associated with soybean transgenesis: A comparison to cultivars and mutagenized plants. *BMC Biotechnology*. 2016;16(1):1-13. doi:10.1186/s12896-016-0271-z

112. Kolata JB, Kühbandner I, Link C, et al. The fall of a Dogma? Unexpected high T-cell memory response to staphylococcus aureus in humans. *Journal of Infectious Diseases*. 2015;212(5):830-838. doi:10.1093/infdis/jiv128

113. Mortensen R, Nissen TN, Blauenfeldt T, Christensen JP, Andersen P, Dietrich J. Adaptive Immunity against Streptococcus pyogenes in Adults Involves Increased IFN-γ and IgG3 Responses Compared with Children . *The Journal of Immunology*. 2015;195(4):1657-1664. doi:10.4049/jimmunol.1500804

114. Mehta A, Merkel OM. Immunogenicity of Cas9 Protein. *Journal of Pharmaceutical Sciences*. 2019;109(1):62-67. doi:10.1016/j.xphs.2019.10.003

115. Mirza Z, Karim S. Advancements in CRISPR/Cas9 technology—Focusing on cancer therapeutics and beyond. *Seminars in Cell and Developmental Biology*. 2019;96:1-9. doi:10.1016/j.semcdb.2019.05.026

116. Joseph A. CRISPR Babies Scientist Sentenced to 3 Years in Prison. *STAT*. Published online 2019.

117. Winblad N, Lanner F. At the heart of gene edits in human embryos. *Nature*. 2017;548(7668):398-400. doi:10.1038/nature23533

118. Thomas Gaj, Charles A. Gersbach CFB. ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Bone*. 2013;31(7):397-405. doi:10.1038/jid.2014.371

119. Gibson DG, Glass JI, Lartigue C, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*. 2010;329(5987):52-56. doi:10.1126/science.1190719

120. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013;493(7432):429-432. doi:10.1038/nature11723

121. Dolgin E. Finding the CRISPR off-switch. *Nature*. 2020;577(January):308-310.

122. Xing X, Wang SC, Xia W, et al. The ets protein PEA3 suppresses HER-2/neu overexpression and inhibits tumorigenesis. *Nature Medicine*. 2000;6(2):189-195. doi:10.1038/72294

123. Dmitry Y. Guschin, Adam J. Waite, George E. Katibah, Jeffrey C. Miller, Michael C. Holmes and EJR. A Rapid and General Assay for Monitoring Endogenous Gene Modificatio. *Methods in Molecular Biology*. 2010;649:247-256. doi:10.1007/978-1-60761-753-2_21

124. Ophir Shalem, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, Benjamin L. Ebert, David E. Root, John G. Doench FZ. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells Ophir. *Science*. 2014;343((6166)):84-88.

125. Silacci M, Brack S, Schirru G, et al. Design, construction, and characterization of a large synthetic human antibody phage display library. *Proteomics*. 2005;5(9):2340-2350. doi:10.1002/pmic.200401273

126. Philibert P, Stoessel A, Wang W, et al. A focused antibody library for selecting scFvs expressed at high levels in the cytoplasm. *BMC Biotechnology*. 2007;7:1-17. doi:10.1186/1472-6750-7-81

127. Knappik A, Ge L, Honegger A, et al. Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *Journal of Molecular Biology*. 2000;296(1):57-86. doi:10.1006/jmbi.1999.3444

128. Söderlind E, Strandberg L, Jirholt P, et al. Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. *Nature Biotechnology*. 2000;18(8):852-856. doi:10.1038/78458

129. Yang HY, Kang KJ, Chung JE, Shim H. Construction of a large synthetic human scFv library with six diversified CDRs and high functional diversity. *Molecules and Cells*. 2009;27(2):225-235. doi:10.1007/s10059-009-0028-9

130. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*. 2018;127(2):187-214. doi:10.1007/s00412-017-0658-1

8 ANNEXES

Bacterial genotypes

StellarTM competent cells: F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ (lacZYA - argF) U169, Δ (mrr - hsdRMS - mcrBC), ΔmcrA, λ–

NEB® Stable: F' proA+B+ lacIq ∆(lacZ)M15 zzf::Tn10 (TetR)/ ∆(ara-leu) 7697 araD139 fhuA ∆lacX74 galK16 galE15 e14- Φ80dlacZ∆M15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 ∆(mrr-hsdRMS-mcrBC)

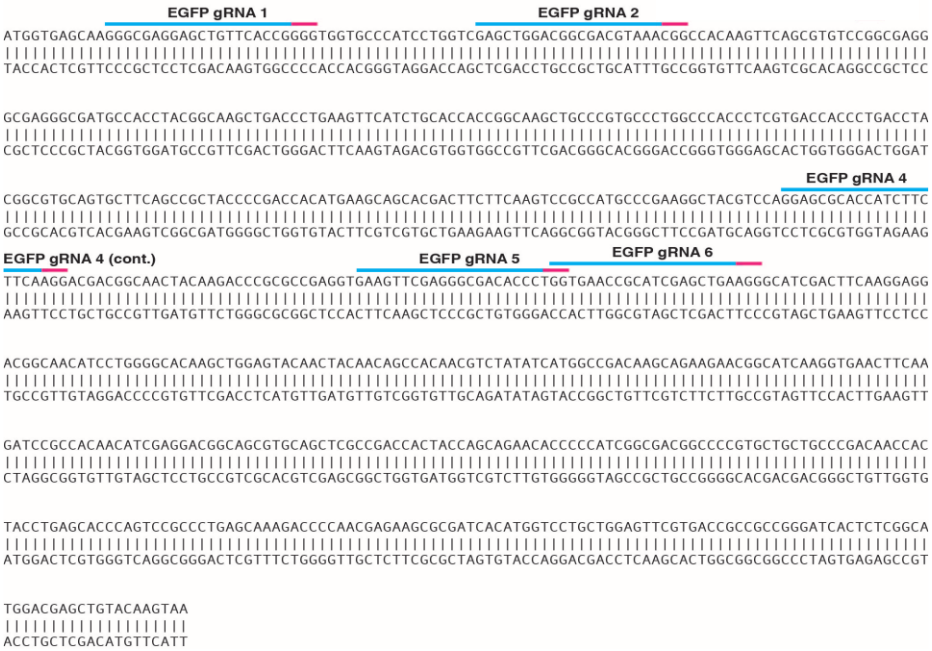


Figure A1| Design of sgRNAs to knock out EGFP. EGFP sequence is shown with five different sgRNAs that were designed to target the coding sequence.

Plasmid Maps

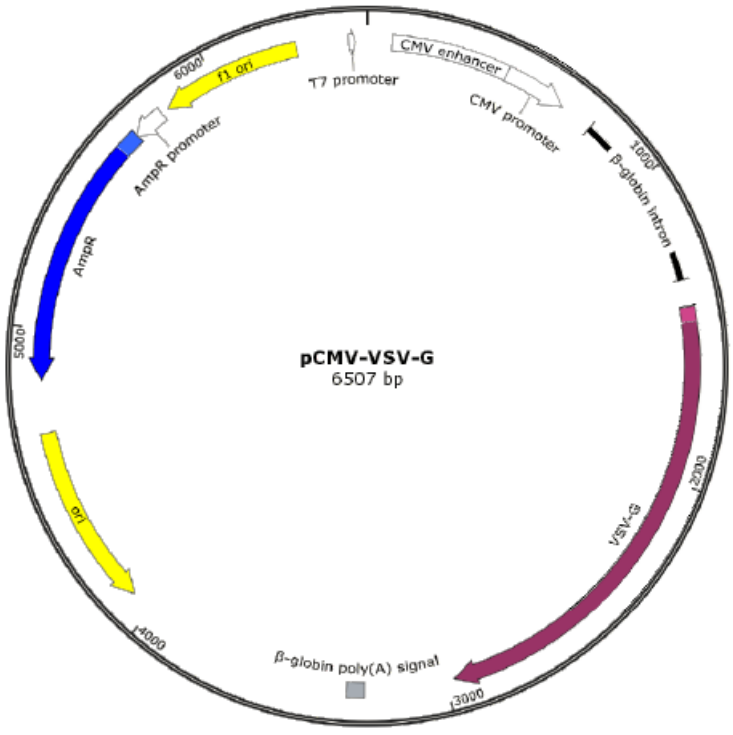


Figure A2| Genomic map of pCMV-VSV-G

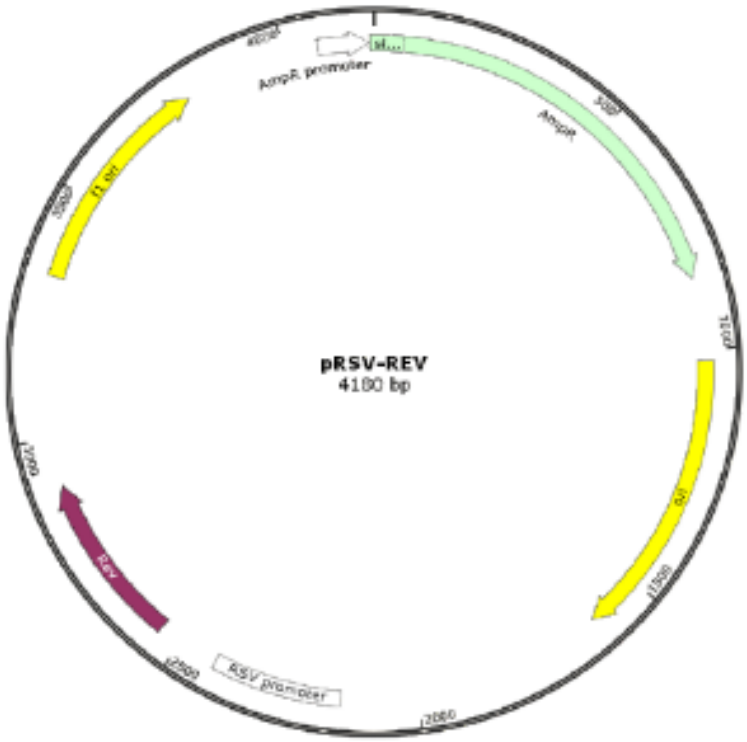


Figure A3| Genomic map of pRSV-REV

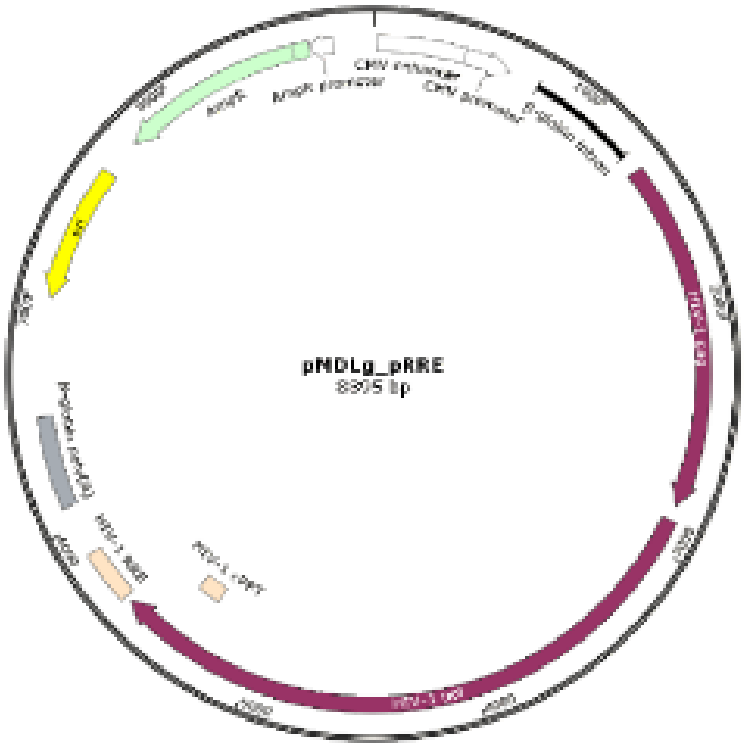


Figure A4| Genomic map of pNDLg/pRRE

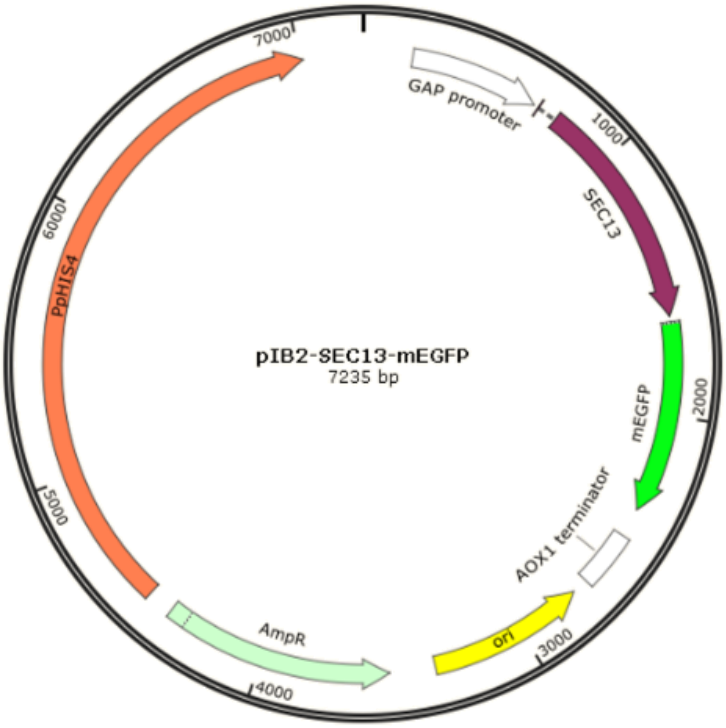


Figure A5| Genomic map of lentiCRISPR - C3orf17 sgRNA 1

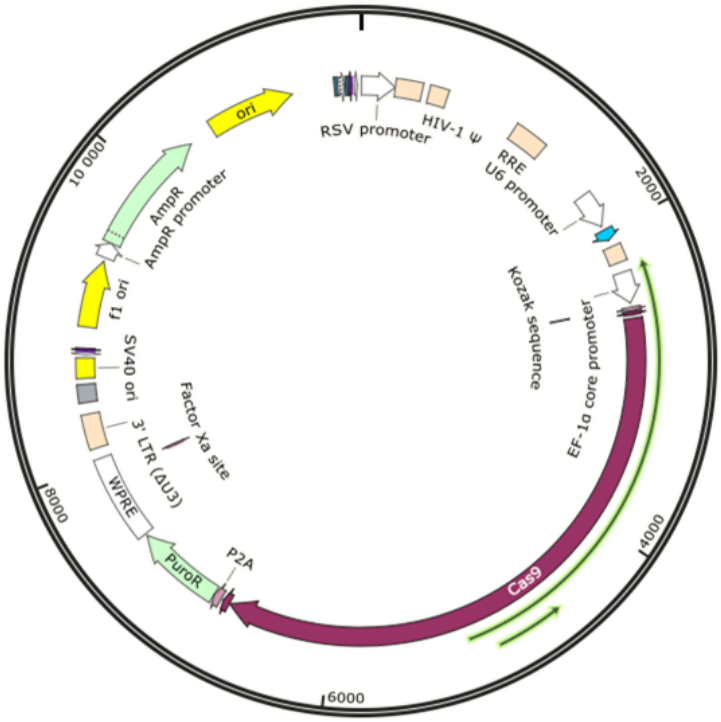


Figure A6| Genomic map of of lentiCRISPR EGFP sgRNA 2

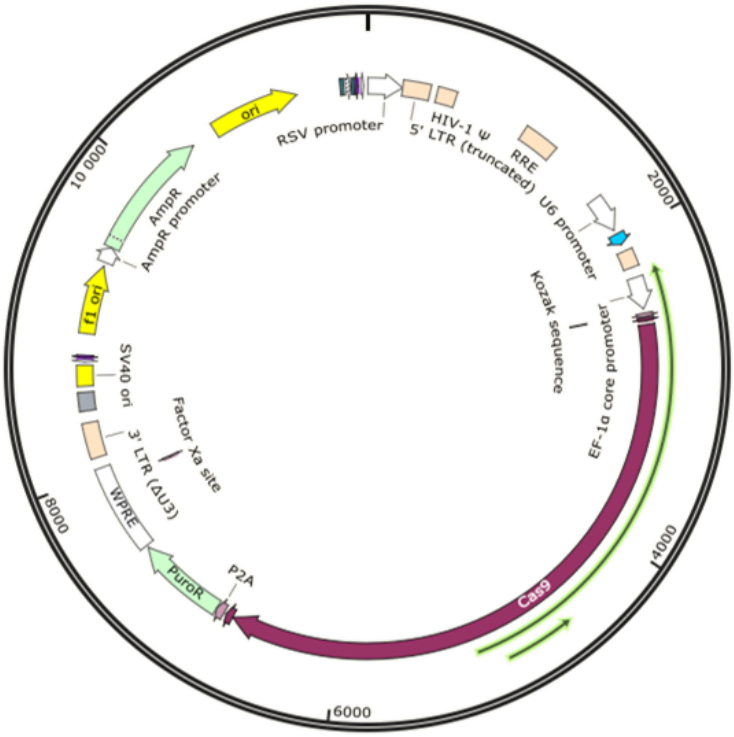


Figure A7| Genomic map of lentiCRISPR EGFP sgRNA 4

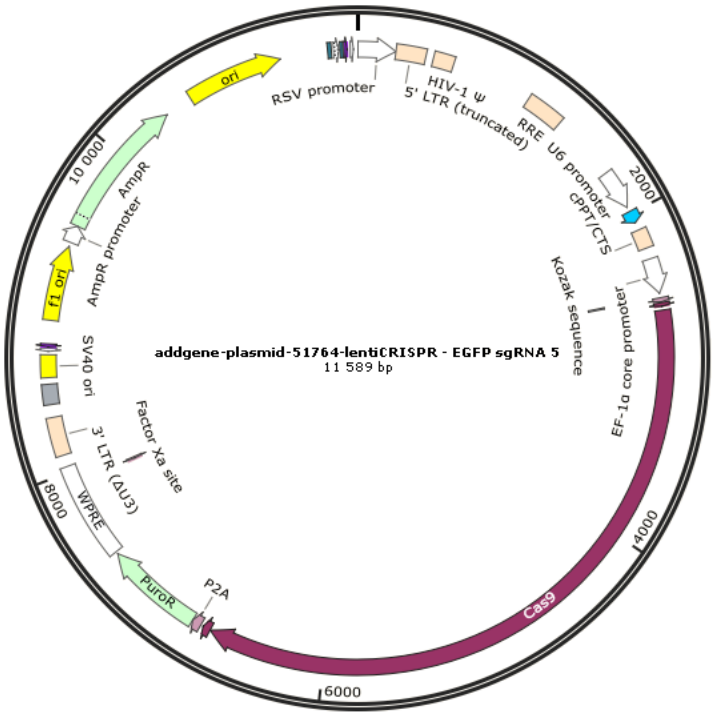


Figure A8| Genomic map of lentiCRISPR EGFP sgRNA 5

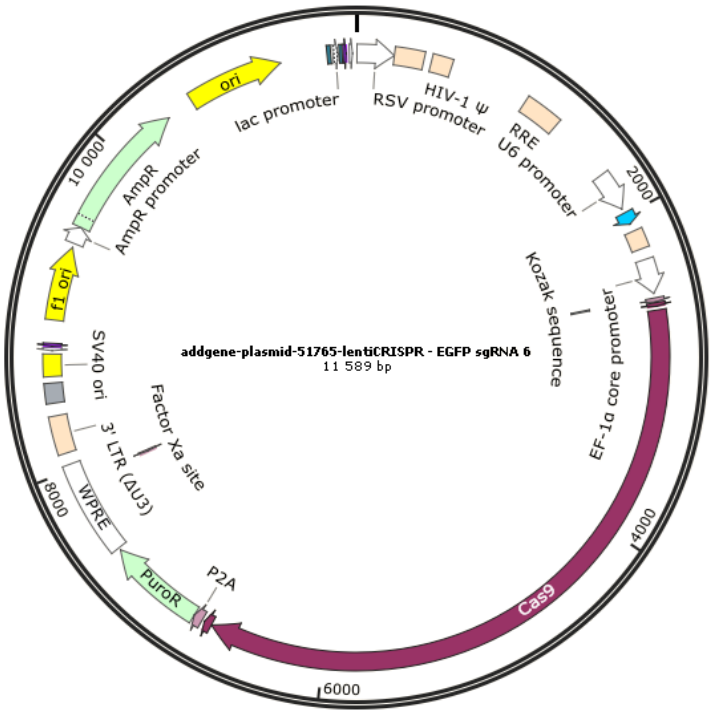


Figure A9| Genomic map of lentiCRISPR EGFP sgRNA 6

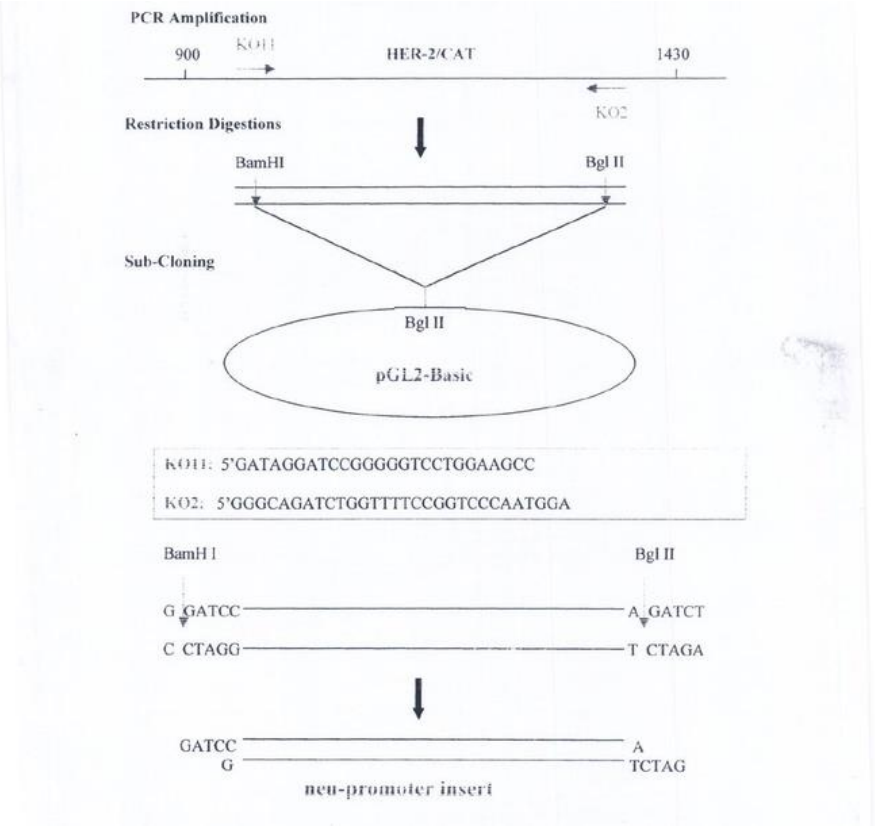


Figure A10| Plasmid pNeuLite Construction Scheme