Abstract: According to a first aspect of the invention, a method for the generation of a cell line is provided, comprising the steps of (a) providing a plurality of mammalian B cells, wherein each of the plurality of B cells comprises a transgenic genomic DNA sequence encoding a marker protein inserted into an endogenous immunoglobulin locus comprised in said B cell, and wherein the transgenic genomic DNA sequence is amenable to cleavage by a site directed nuclease, particularly Cas9; (b) replacing the transgenic genomic DNA sequence encoding a marker protein with a second transgenic DNA sequence encoding a protein of interest; (c) sorting B cells based on the presence or absence of the marker protein; and (d) collecting B cells in which the marker protein is absent.
Mammalian Cell Line for Protein Production and Library Generation

The invention relates to an engineered cell line and to its use for rapid generation of stable cells for protein expression and generation of protein libraries for directed evolution and protein engineering.

5 Background

Currently available methods for protein production from mammalian cells rely primarily on random transgene integration in three cell lines: Human embryonic kidney 293 (HEK293) cells, mouse myeloma cells (Sp20 and NSO), and Chinese hamster ovary (CHO) cells. These methods are both expensive and time consuming. The biotechnology industry standard for therapeutic protein production requires the generation of stable recombinant protein producing cell lines, capable of producing protein near indefinitely, when renewed from frozen cell stocks.

Engineering proteins by directed evolution requires the generation of DNA libraries by in vitro mutagenesis methods (e.g. error prone PGR, degenerate primer PGR mutagenesis, DNA shuffling), followed by cloning into expression hosts and high-throughput screening. The size of libraries that can be generated in mammalian cells is limited due to poor transfection efficiencies. Therefore, high-throughput screening of protein libraries often employs orthogonal surface display platforms, largely based on in vitro or microbial expression (e.g. ribosome, phage, yeast and bacterial display). Compared to mammalian cells, these hosts can have a large impact on protein folding, glycosylation patterns, and expression. Critically though, the standard systems are unable to express and screen some complex proteins effectively (e.g. full-length antibodies). Additionally, in the case of therapeutic proteins, following library screening in in vitro or microbial systems, there is often the need for genetic sub-cloning into mammalian expression systems for proper characterization.

The problem underlying the present invention is to provide a fast, reliable and inexpensive method for the generation of stable mammalian cell lines that can be used for recombinant protein expression and for the generation and screening of large protein libraries for protein engineering and directed evolution applications. This problem is solved by the subject matter of the independent claims.

30 Definitions

In the context of the present specifications the terms sequence identity and percentage of sequence identity refer to the values determined by comparing two aligned sequences. Methods for alignment of sequences for comparison are well-known in the art. Alignment of

One example for comparison of amino acid sequences is the BLASTP algorithm that uses the default settings: Expect threshold: 10; Word size: 3; Max matches in a query range: 0; Matrix: BLOSUM62; Gap Costs: Existence 11, Extension 1; Compositional adjustments: Conditional compositional score matrix adjustment. One such example for comparison of nucleic acid sequences is the BLASTN algorithm that uses the default settings: Expect threshold: 10; Word size: 28; Max matches in a query range: 0; Match/Mismatch Scores: 1.; Gap costs: Linear. Unless otherwise stated, sequence identity values provided herein refer to the value obtained using the BLAST suite of programs (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) using the above identified default parameters for protein and nucleic acid comparison, respectively.

In the context of the present specification, the term "B cell" refers to a cell of lymphoid lineage that has completed genomic rearrangement of the immunoglobulin heavy chain and light chain gene loci by V(D)J recombination.

In the context of the present specification, the term "IgH locus" refers to the immunoglobulin heavy chain gene locus.

In the context of the present specification, the term "site-directed endonuclease" refers to an endonuclease selected from the group comprising a CRISPR-associated endonuclease, a zinc finger nuclease (ZFN), a transcription activator-like effector-based nuclease (TALEN) and a meganuclease.

In the context of the present specification, the term "CRISPR-associated endonuclease (Cas9)" refers to the Cas9 endonuclease of Streptococcus pyogenes (SpyCas9), to orthologues of SpyCas9 or to engineered protein variants of SpyCas9 or its orthologues.

In the context of the present specification, the term "orthologue" refers to a gene and its corresponding polypeptide that evolved by vertical descent from a single ancestral gene. In other words, orthologues genes/polypeptides share a common ancestor and were divided when a species diverged into two separate species. The copies of a single gene in the two resulting species are then referred to as orthologues. To ascertain that two genes are orthologues a person skilled in the art can carry out a phylogenetic analysis of the gene
lineage by comparing the aligned nucleotide or amino acid sequences of genes or polypeptides.

In the context of the present specification, the term *antibody* is used in its meaning known in the art of cell biology and immunology; it refers to whole antibodies including but not limited to immunoglobulin type G (IgG), type A (IgA), type D (IgD), type E (IgE) or type M (IgM), any antigen binding fragment or single chains thereof and related or derived constructs. A whole antibody is a glycoprotein comprising at least two heavy (H) chains and two light (L) chains inter-connected by disulfide bonds. Each heavy chain is comprised of a heavy chain variable region \( (V_H) \) and a heavy chain constant region \( (C_H) \). The heavy chain constant region is comprised of three domains, \( C_{H1} \), \( C_{H2} \) and \( C_{H3} \). Each light chain is comprised of a light chain variable region (abbreviated herein as \( V_L \) ) and a light chain constant region \( (C_L) \). The light chain constant region is comprised of one domain, \( C_L \). The variable regions of the heavy and light chains contain a binding domain that interacts with an antigen. The constant regions of the antibodies may mediate the binding of the immunoglobulin to host tissues or factors, including various cells of the immune system (e.g., effector cells) and the first component of the classical complement system.

The term *antibody-like molecule* in the context of the present specification refers to a molecule capable of specific binding to another molecule or target with high affinity, a \( K_d \leq 10^{-8} \text{ mol/l} \). An antibody-like molecule binds to its target similarly to the specific binding of an antibody. The term *antibody-like molecule* encompasses a repeat protein, such as a designed ankyrin repeat protein (Molecular Partners, Zurich), a polypeptide derived from armadillo repeat proteins, a polypeptide derived from leucine-rich repeat proteins and a polypeptide derived from tetratricopeptide repeat proteins.

The term *antibody-like molecule* further encompasses a polypeptide derived from protein A domains (a protein A domain derived polypeptide), a polypeptide derived from fibronectin domain FN3, a polypeptide derived from consensus fibronectin domains, a polypeptide derived from lipocalins, a polypeptide derived from Zinc fingers, a polypeptide derived from Src homology domain 2 (SH2), a polypeptide derived from Src homology domain 3 (SH3), a polypeptide derived from PDZ domains, a polypeptide derived from gamma-crystallin, a polypeptide derived from ubiquitin, a polypeptide derived from a cysteine knot polypeptide and a polypeptide derived from a knottin.

The term *protein A domains derived polypeptide* refers to a molecule that is a derivative of protein A and is capable of specifically binding the Fc region and the Fab region of immunoglobulins.
The term *armadillo repeat protein* refers to a polypeptide comprising at least one armadillo repeat, wherein a armadillo repeat is characterized by a pair of alpha helices that form a hairpin structure.

The term *humanized camelid antibody* in the context of the present specification refers to an antibody consisting of only the heavy chain or the variable domain of the heavy chain (VHH domain) and whose amino acid sequence has been modified to increase their similarity to antibodies naturally produced in humans and, thus show a reduced immunogenicity when administered to a human being.


In the context of the present specification, the term *fragment crystallizable (Fc) region* is used in its meaning known in the art of cell biology and immunology; it refers to a fraction of an antibody comprising two identical heavy chain fragments comprised of a C_{H2} and a C_{H3} domain, covalently linked by disulfide bonds.

Further definitions of terms used herein are given throughout the document where appropriate.

**Description of the invention**

According to a first aspect of the invention, a method for the generation of a cell line is provided, comprising the steps of

a. providing a plurality of mammalian B cells, wherein each of the plurality of B cells comprises a transgenic genomic DNA sequence encoding a marker protein, wherein the transgenic genomic DNA sequence is inserted into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, and wherein said transgenic genomic DNA sequence is amenable to cleavage by a site-directed nuclease, particularly CRISPR-associated endonuclease (Cas9);

b. replacing said transgenic genomic DNA sequence encoding said marker protein with a second transgenic DNA sequence encoding a protein of interest;

c. sorting said B cells based on the presence or absence of said marker protein; and

d. collecting B cells in which said marker protein is absent.

In certain embodiments, the marker protein is a fluorescent protein. In certain embodiments, the marker protein is a protein conferring resistance to an antibiotic drug.
In certain embodiments, said marker protein is a fluorescent protein and said sorting is done
by flow cytometry.

In the context of the present specification, the expression “replacing said transgenic genomic
DNA sequence encoding said marker protein with a second transgenic DNA sequence
encoding a protein of interest” relates to both actual replacement of the DNA sequence and
functional replacement, in the way that the marker protein is no longer expressed and the
protein of interest is expressed instead.

The skilled person is aware that the expression “said transgenic genomic DNA sequence
[encoding a marker protein] is amenable to cleavage by a site directed nuclease” comprises
both cleavage within the DNA sequence encoding the marker protein and cleavage within the
immediate 3’ and/or 5’ flanking regions.

In certain embodiments, said transgenic genomic sequence is flanked by a 5’ flanking
sequence tract and a 3’ flanking sequence tract, wherein said 5’ flanking sequence tract
and/or said 3’ flanking sequence tract is amenable to cleavage by a site directed nuclease.

In certain embodiments, said flanking sequence tracts comprise 0 to 1500 nucleotides,
particularly 0 nucleotides or 1 to 700 nucleotides, more particularly 0 nucleotides or 1 to 100
nucleotides.

The sites amenable to cleavage by the site-directed nuclease, particularly Cas9, are typically
within 500 nucleotides of the sequence encoding the marker protein or within the sequence
encoding the marker protein.

In certain embodiments, the protein of interest is an antibody, an antibody-like molecule, a
humanized camelide antibody or an immunoglobulin antigen-binding fragment.

In those instances where the later inserted gene of interest is an antibody, the endogenous
immunoglobulin V\textsubscript{H} gene is replaced or disrupted by the transgenic genomic DNA sequence
encoding a marker protein.

In the context of the present specification, the term “V\textsubscript{H} gene” refers to the DNA sequence
encoding the variable region of an immunoglobulin heavy chain.

In the context of the present specification, the term “V\textsubscript{L} gene” refers to the DNA sequence
encoding the variable region of an immunoglobulin light chain.

In certain embodiments, the plurality of B cells is selected from the group comprising primary
B cells, immortalized B cells, hybridoma cells, myeloma cells, plasmacytoma cells, and
lymphoma cells.

In certain embodiments, the marker protein and/or the protein of interest is expressed under
control of an endogenous immunoglobulin promoter, particularly the V\textsubscript{H} promoter.
In the context of the present specification, the term "V<sub>H</sub> promoter" refers to the promoter of the "V<sub>H</sub> gene".

In certain embodiments, the endogenous V<sub>H</sub> gene and the endogenous V<sub>L</sub> gene are disrupted in each of said plurality of B cells. This means that the B cells can neither express an immunoglobulin heavy chain nor an immunoglobulin light chain.

In certain embodiments, the plurality of B cells is genetically modified to constitutively express CRISPR-associated endonuclease.

In certain embodiments, the plurality of B cells is genetically modified to express an activation-induced cytidine deaminase (AID) in an inducible and titratable manner.

In the context of the present specification, the term "activation-induced cytidine deaminase" refers to an enzyme able to deaminate cytosine bases within genomic DNA, turning them into uracil (EC 3.5.4.38).

In certain embodiments, the plurality of B cells comprises a safe harbor locus, and a first expression cassette comprising a DNA sequence encoding the CRISPR-associated endonuclease is inserted into said safe harbor locus.

In the context of the present specification, the term "safe harbor locus" refers to a chromosomal location amenable for integration of transgenes. Transgenes integrated in a safe harbor locus are stably expressed and do not perturb endogenous gene activity. Examples of safe harbor loci are the murine Rosa26 locus or the AAVS1 locus.

In certain embodiments, the CRISPR-associated endonuclease inserted into said safe harbor locus is under the control of a constitutively active promoter, particularly the CAG promoter, CBh promoter, or the CMV promoter.

The CAG promoter is a hybrid construct consisting of the cytomegalovirus enhancer fused to the chicken beta-actin promoter (Jun-ichi et al., Gene 79(2):269-277). CBh promoter is a hybrid form of the CBA (chicken beta-actin) promoter (Gray et al., Hum Gene Ther., 2011, 22(9):143-153). The term "CMV promoter" refers to the human cytomegalovirus immediate early enhancer and promoter sequence of a human herpesvirus such as human herpesvirus 5 strain Toledo (GenBank GU937742.2). Exemplary CMV sequences are deposited in GenBank under the references X03922.1, M64940.1, M64941.1, M64942.1, M64943.1, M64944.1 and K03104.1.

In certain embodiments, the plurality of B cells comprises a safe harbor locus, and a second expression cassette comprising a DNA sequence encoding said activation-induced cytidine deaminase (AID) is inserted into said safe harbor locus.

In certain embodiments, the second expression cassette comprises regulatory sequences
amenable to inducible activation by an activator molecule.

In certain embodiments, the second expression cassette comprises an inducible expression system, particularly the Tet-On inducible expression system. The Tet-One inducible expression system comprises a transactivator protein (Tet-On) and an inducible promoter (PTRE3GS). In the presence of the antibiotic doxycycline, Tet-On binds to tetO sequences in PTRE3GS and activates a high level of transcription of the gene downstream of the promoter, i.e. of the gene encoding activation-induced cytidine deaminase (AID). If doxycycline concentration is reduced, expression is reduced, thus generating a titratable system of AID expression and somatic hypermutation.

In the context of the present specification, the term "somatic hypermutation (SHM)" refers to a cellular mechanism by which multiple genomic mutations are generated. SHM involves the deamination of cytosine to uracil in DNA by the enzyme activation-induced cytidine deaminase (AID). The resulting basepair mismatch (uracil-guanosine, U:G) can bring about a genomic mutation, e.g. via excision of the uracil base and filling of the gap by an error-prone DNA polymerase, or via DNA replication during which the uracil is treated as a thymididine.

In the context of the present specification, the term "inducible synthetic somatic hypermutation (iSSHM)" refers to SHM that can be induced in said plurality of B cells by activation of AID expression from an inducible expression system.

In certain embodiments, said replacing of said transgenic genomic DNA sequence is mediated by Cas9-mediated site-directed DNA cleavage and subsequent integration of said second transgenic DNA sequence by homology directed repair (HDR) or non-homologous end joining (NHEJ). This method comprises providing a guide RNA and a replacement DNA.

In the context of the present specification, a guide RNA or gRNA is a short synthetic RNA composed of a sequence necessary for Cas9-binding and a user-defined "targeting sequence" of approximately 23 nucleotides which defines the genomic target to be modified (see table 2).

Such guide RNAs can be provided by the transfection of in vitro transcribed RNA or commercially synthesized oligonucleotide RNA. Alternatively, the guide RNA can be provided by transfection or viral transduction of a DNA sequence into the cell, wherein such DNA sequence encodes (and expresses in the cell) the guide RNA. Multiple guide RNAs can be provided for simultaneous cleavage of multiple genomic sites in the IgH locus (this improves efficiency of transgene insertion).

The replacement DNA comprises said second transgenic DNA sequence encoding a protein of interest. Replacement DNA can be provided by transfection or viral transduction of either circular or linear double-stranded DNA (dsDNA) or single-stranded DNA (ssDNA) (e.g.,
In certain embodiments of this aspect of the invention, said protein of interest is a full-length antibody. The full-length antibody can be based on a synthetic antigen binding fragment (sFAB) construct. To reintroduce a new full-length antibody, while avoiding targeting of both the immunoglobulin heavy chain locus and the immunoglobulin light chain locus, both full-length light and heavy chains can be expressed from the native $V_\text{H}$ promoter as a single transcript (Fig. 2B).

In certain embodiments, the plurality of B cells are mouse hybridoma cells.

In certain embodiments, the plurality of B cells are mouse hybridoma cells and the murine $C_\text{H}$ region is replaced with a $C_\text{H}$ region of a different species, particularly with a human $C_\text{H}$ region. This allows the generation of a human antibody by a mouse hybridoma cell.

In the context of the present specification, the term "$C_\text{H}$ region" refers to the DNA sequence encoding the constant region of an immunoglobulin heavy chain.

In certain embodiments of this aspect of the invention, said second transgenic nucleic acid sequence comprises more than one gene of interest and additional promoters. This way, a cell line that stably expresses multiple genes or an entire synthetic genetic network can be generated by the inventive method.

Current state of the art methods for recombinant protein expression from stable mammalian cells take at least 8-10 weeks to develop, with commercial entities charging upwards of 10,000 USD per protein. The generation of stable cell lines for industrial therapeutic purposes takes up to one year.

Usually, analysis of multiple clones is required, due to varying efficiency of protein production. Factors affecting clonal productivity are number of integrations and integration site, as gene silencing is known to occur over time at some integration sites.

According to a second aspect of the invention, a method for the generation of a library of protein variants is provided. This method comprises a method according to the first aspect of the invention or any of its embodiments mentioned above, followed by the additional step of modifying regions of the transgenic DNA with randomized nucleic acid sequences (through randomized regions on donor dsDNA or ssDNA). Thus, genomic mutations within the protein of interest by site-directed mutagenesis are generated.

According to an alternative of this aspect of the invention, a method for the generation of a library of protein variants comprises the method according to the first aspect of the invention or any of its embodiments mentioned above, followed by the additional step(s) of

  a. inducing expression of activation-induced cytidine deaminase (AID), thus generating
multiple genomic mutations within the protein of interest by inducible synthetic somatic hypermutation (iSSHM), or

b. modifying regions of said transgenic DNA with randomized nucleic acid sequences, thus generating genomic mutations within the protein of interest by site-directed mutagenesis.

According to a third aspect of the invention, a method for the generation of a library of protein variants is provided. The method comprises the method according to the first aspect of the invention, followed by the additional step(s) of inducing expression of activation-induced cytidine deaminase (AID), thus generating multiple genomic mutations within the protein of interest by inducible synthetic somatic hypermutation (iSSHM).

AID is especially efficient in generating mutations within the immunoglobulin (IgH or IgK or IgL) locus. It is therefore advantageous for the iSSHM that the DNA sequence encoding the protein of interest is inserted into the immunoglobulin locus, such as the IgH locus.

In certain embodiments, said modifying of said regions of the transgenic genomic DNA sequence is mediated by Cas9-mediated site-directed DNA cleavage and either

a. subsequent integration of said randomized transgenic DNA sequences by homology directed repair (HDR) or non-homologous end joining (NHEJ), or

b. subsequent insertion or deletion of nucleotides through the repair of the transgenic DNA sequence by NHEJ.

This method comprises providing a guide RNA and replacement DNA in an analogous manner to the first aspect of the invention. Guide RNA and randomized nucleic acid sequences can be provided by the transfection or viral transduction of dsDNA or ssDNA, comprising degenerate nucleotides or trinucleotide codons. To improve the HDR efficiency of ssDNA, phosphorothioate bonds are introduced at the 5’ and 3’ ends, as well as throughout the ssDNA oligonucleotide. Phosphorothioate bonds are an exchange of the non-bridging oxygen of the phosphate backbone with a sulfur atom. The sulfur atom in the phosphate backbone increases the ssDNA’s resistance to nuclease degradation.

In the context of the present specification, the term “degenerate nucleotide” refers to the position of a DNA sequence encoding any mixed nucleotide composition.

In the context of the present specification, the term “trinucleotide”, or trimer phosphoramidite, refers to three contiguous nucleotides of a DNA sequence encoding any mixed amino acid composition.

The skilled person is aware that the generation of a library of protein variants according to the second aspect of the invention comprises the generation of a library of cell lines.
In those instances, where the protein of interest is an antibody, the mutated antibodies are screened for improved or novel antigen binding. Fluorescently labelled antigen is added to the cells and the cells are screened by FACS. Alternatively, an initial screening step can be carried out by magnetic-associated cell sorting (MACS) using magnetic beads conjugated to antigen.

Several rounds of screening and iSSHM or site-directed mutagenesis can be performed to continue the engineering of an antibody or protein.

In certain embodiments, the marker protein is a fluorescent protein.

By way of non-limiting example, such fluorescent protein may be selected from green fluorescent protein (GFP) from *Aequorea victoria* and derivatives thereof, such as

- enhanced blue fluorescent protein (EBFP), enhanced blue fluorescent protein 2 (EBFP2), azurite, mKalamal, sirius
- enhanced green fluorescent protein (EGFP), emerald, superfolder avGFP, T-sapphire
- yellow fluorescent protein (YFP), enhanced yellow fluorescent protein (EYFP), citrine, venus, YPet, topaz, SYFP, mAmetrine
- enhanced cyan fluorescent protein (ECFP), mTurquoise, mTurquoise2, cerulean, CyPet, SCFP.

A fluorescent protein for practicing the invention may also be selected from the group comprising fluorescent protein from *Discosoma striata* and derivatives thereof:

- mTagBFP,
- TagCFP, AmCyan, Midoriishi Cyan, mTFP1
- Azami Green, mWasabi, ZsGreen, TagGFP, TagGFP2, TurboGFP, CopCFP, AceGFP
- TagYFP, TurboYFP, ZsYellow, PhiYFP
- Kusabira Orange, Kusabira Orange2, mOrange, mOrange2, dTomato, dTomato-Tandem, DsRed, DsRed2, DsRed-Express (T1), DsRed-Express2, DsRed-Max, DsRed-Monomer, TurboRFP, TagRFP, TagRFP-T
- mRuby, mApple, mStrawberry, AsRed2, mRFP1, JRed, mCherry, eqFP611, tdRFP611, HcRed1, mRaspberry
- tdRFP639, mKate, mKate2, katushka, tdKatushka, HcRed-Tandem, mPlum, AQ143.

Fluorescent proteins also comprise proteins derived from alpha-allophycocyanin from the cyanobacterium *Trichodesmium erythraeum* such as small ultra red fluorescent protein (smURFP).

Fluorescent proteins also comprise proteins derived from alpha-allophycocyanin from the cyanobacterium *Trichodesmium erythraeum* such as small ultra red fluorescent protein (smURFP).
The skilled person is aware that the method according to the first aspect of the invention may also work with other fluorescent proteins not included in the list above.

Another aspect of the invention provides a human B cell line comprising an expressed transgenic genomic DNA sequence inserted into an endogenous immunoglobulin locus comprised in said B cell obtained by a method according to the first, second or third aspect of the invention.

According to yet another aspect of the invention, a protein obtained by a method according to the first, second or third aspect of the invention is provided.

According to yet another aspect of the invention, a library of protein variants obtained by the method according to the second or third aspect of the invention is provided.

According to yet another aspect of the invention, a mammalian B cell is provided, wherein

- a transgenic genomic DNA sequence encoding a marker protein is inserted into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, wherein the marker gene encodes a fluorescent protein, and wherein

- said transgenic genomic sequence is amenable to cleavage by a site directed nuclease, particularly CRISPR-associated endonuclease (Cas9).

In certain embodiments of this aspect of the invention, the endogenous V_H gene and the endogenous V_L gene of said B cell are disrupted.

In certain embodiments, the mammalian B cell is a human cell.

Yet another aspect of the invention relates to a plurality of mammalian B cells, each of which encodes a variant of a transgene protein. The plurality in its entirety constitutes a library of such variants. Each member of the plurality of B cells contained in the plurality comprises a transgenic genomic DNA sequence encoding a variant of a protein of interest. The transgenic genomic DNA sequence is inserted into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, and each variant encoded by a member of said plurality is different from any other variant encoded by another member of said plurality.

The skilled person understands that each member of the plurality is likely to be represented by more than one individual cell, i.e. several cells of a clone may constitute a member. The important aspect here is that the variant is the same for one member, and that a large number (in certain embodiments, equal or more than ≥ 100, ≥ 1000, ≥10,000 or even ≥ 100,000 different variants may be present in the plurality.

In certain embodiments of this aspect of the invention, each variant encoded in the plurality of B cells is different from another variant in one to five positions of its amino acid sequence.

In certain embodiments of this aspect of the invention, each variant is at least 80%, 85%,
90%, 95%, 97%, 98% or 99% identical to any another variant encoded by a member of said plurality. In other words, the variants may share a significant part of their sequence between individual members, and yet represent a very broad spectrum of variants.

In certain embodiments of this aspect of the invention, the plurality of B cells is selected from the group comprising primary B cells, immortalized B cells, hybridoma cells, myeloma cells, plasmacytoma cells, and lymphoma cells.

Wherever alternatives for single separable features are laid out herein as "embodiments", it is to be understood that such alternatives may be combined freely to form discrete embodiments of the invention disclosed herein.

**Examples**

The invention is further illustrated by the following examples and figures, from which further embodiments and advantages can be drawn. These examples are meant to illustrate the invention but not to limit its scope.

To realize the invention, the inventors have generated a plug-and-(dis)play (PnP) mammalian cell line. The inventors used mouse B-lymphocytes (hybridoma cells), which function as production factories for antibody proteins. The PnP cell line consists of the following components:

1) The endogenous $V_H$ gene in the IgH locus was replaced with a fluorescent protein (mRuby, originally referred to as mRuby2, originated from the Addgene.org plasmid #: 40260 (Lam et al., Nat Methods 2012, 9:1005-1012)). This was accomplished by transfecting WT hybridoma cells with a CRISPR-Cas9 plasmid (pX458) (Addgene.org plasmid #: 48138) (Cong et al., Science 2013, 339:819-823) with a guide RNA (gRNA) targeting the intron between $V_H$ and IgG CH1 genes. Also co-transfected was a donor DNA construct consisting of the mRuby gene and homology arms corresponding to the IgH locus of WT hybridoma cells. The CRISPR-Cas9 system introduced a targeted double-strand break (DSB) in the DNA of the WT cells, which promoted DNA repair mechanisms of homology directed repair (HDR) or non-homologous end joining (NHEJ) resulting in site specific integration of the mRuby gene in place of the $V_H$ gene. The native IgH promoter is used for expression of mRuby (Fig. 1 a-d).

2) The endogenous $V_L$ gene in the IgK locus was deleted to generate a light chain knockout cell line. This was accomplished by transfecting the hybridoma cells (described in 1) with pX458 with gRNAs targeting two sites flanking the $V_L$ gene in the IgK locus. This results in deletion of the $V_L$ gene and knockout of the endogenous light chain expression in hybridoma cells (Fig. 1e-g).
The resulting cells are referred to as PnP-mRuby cells.

3) PnP-mRuby cells were converted into cells that express a new antibody, these are referred to as PnP-IgG cells. This was accomplished by transfecting PnP-mRuby cells with pX458 with gRNA targeting the mRuby gene (now integrated in the IgH locus). Also co-transfected was a donor DNA construct of a synthetic antibody fragment (sFAb) with homology arms corresponding to the IgH locus. Similar to 1), CRISPR-Cas9 promotes a DSB and HDR or NHEJ, which results in targeted integration of the sFAb into the IgH locus. The outcome of this is that full-length antibody (IgK and IgH) is expressed from the IgH locus as a single RNA transcript (Fig. 2).

PnP-mRuby cells can be converted into a stable mammalian cell line expressing recombinant protein (e.g., PnP-IgG) with a single transfection and selection step (Fig. 3, 4).

4) PnP-mRuby cells were engineered for constitutive expression of Cas9 (PnP-mRuby-Cas9 cells). This was accomplished by transfecting PnP-mRuby cells with pX458 with gRNA targeting a site in the safe harbor locus of Rosa26 and a donor DNA construct consisting of Cas9-2A-puromycin, under the control of a constitutively active promoter (e.g. CAG promoter or CMV promoter) (Addgene.org plasmid # 48139) (Piatt et al., CELL 2014, 159:440-455; Ran et al., Nat Protoc 2013, 8:2281-2308). An advantage of these cells is that they eliminate the need to transfect with CRISPR-Cas9 plasmid (pX458), therefore to convert PnP-mRuby cells to a cell line expressing a recombinant protein (e.g., PnP-IgG) only gRNA (in vitro transcribed or commercially synthesized) and donor construct (replacement DNA, e.g., sFAb) need to be transfected, which improves efficiency (Fig. 5a).

In vitro transcription of gRNAs is performed by using template DNA consisting of a T7 promoter, a customized spacer region encoding the gRNA and the trans-activating region in a manner according to the pX458 design. This construct serves as a template for the MEGAscript® T7 transcription kit (Thermo, AM1334), thus in vitro transcription results in a chimeric single gRNA. The protocol is adapted from: https://www.protocols.io/view/In-vitro-transcription-of-guide-RNAs-d4w8xd?step=3

5) PnP-IgG cells expressing antigen specific antibodies for HEL (PnP-HEL23) were further modified in the complementary determining region 3 of the V\textsubscript{H} gene (CDR-H3) to generate a large library of protein variants by site-directed mutagenesis, which can then be screened for increased antigen affinity or novel antigen binding. This was accomplished in the following manner:

The sFAb gene of PnP-HEL23 cells was modified to knock out antibody expression. By
transfecting these cells with the pX458 vector containing guide RNA targeting the CDR-H3, antibody expression is knocked out by the insertion or deletion of nucleotides through repair via NHEJ. The insertion or deletion of nucleotides cause frameshift mutations and subsequently alters all downstream amino acids remaining in the gene. Single cell clones negative for antibody expression can be isolated by flow cytometry and expanded. A suitable clone is then selected (PnP-HEL23-lgH') by phenotypic and genotypic characterization. A library of protein variants can then be derived from said cell clone through integration of randomized nucleic acids.

Affinity maturation of the antibody can be accomplished by providing a guide RNA targeting the CDR-H3 and replacement DNA by transfection or viral transduction, wherein the replacement DNA contains a region of three randomized nucleic acids corresponding to a single amino acid at each position found within the original CDR-H3 (Fig. 11).

A protein library for the discovery of novel, antigen specific antibodies may also be generated by complete replacement of the CDR-H3 in an analogous manner to the method described for antibody affinity maturation. Guide RNA targeting the CDR-H3 and replacement DNA are provided by transfection or viral transduction, wherein the replacement DNA contains a region comprising three to 69 randomized nucleic acids corresponding to variable lengths of the CDR-H3. (Fig. 11).

In both instances, cells producing functional antibodies against the target antigen are screened and sorted by flow cytometry.

6) This plug-and-(dis)play (PnP) mammalian cell line was further engineered so that they are able to generate a large library of protein variants by inducible synthetic somatic hypermutation iSSHM, which can then be used for directed evolution and high-throughput screening (PnP-iAID cells). This was accomplished in the following manner:

PnP-mRuby2 cells were transfected with pX458 vector containing sgRNA targeting the safe harbor locus ROSA26 and a donor construct (replacement DNA). The donor construct comprises the Tet-One System (Clontech) (Heinz et al., Human Gene Therapy 2011, 22:166-176), composed of the following components (Fig. 5b):

i. In the forward direction, a human phosphoglycerate kinase 1 (hPGK) promoter that provides constitutive expression of the Tet-On 3G protein;

ii. in the forward direction, coding sequence of Tet-On 3G Transactivator Protein, a fusion protein of rTetR linked to the VP16 activation domain (rTA);

iii. in the reverse direction, PTRE3GS inducible promoter, a modified version of PTRE3G, which consists of a modified tet responsive element (TRE) and contains 7 direct
repeats of the tet operator joined to a minimal CMV promoter;

iv. a gene of interest (GOI), whose expression is driven by the P<sub>PTRE3GS</sub> promoter, and which comprises DNA sequences encoding:

i. a fluorescent reporter protein (e.g., GFP or blue fluorescent protein (BFP));

ii. a "self-cleaving" 2A peptide;

iii. activation-induced cytidine deaminase (AID);

v. homology arms (>500 bp) corresponding to the Rosa 26 locus.

Alternatively, a donor construct without homology arms can be used. In these instances, the donor construct can be integrated via NHEJ.

In an alternative approach, PnP-mRuby-Cas9 cells were used as a starting platform for the creation of the PnP-iAID cell line. In this instance, PnP-mRuby-Cas9 cells were transfected with in vitro transcribed gRNA targeting an orthogonal site in the safe harbor locus of Rosa26 and a donor construct (replacement DNA). The donor comprises the same components described above in 4).

In an alternative approach, PnP-mRuby-Cas9 cells were transfected with in vitro transcribed gRNA targeting the native AID genomic locus of mice. In this case the donor construct consisted of the same elements described above in 4) but with exception that in iv) only the first intron of the gene encoding activation-induced cytidine deaminase (AID) is present (Fig. 5c).

In the presence of the antibiotic doxycycline (Dox), Tet-On binds to tetO sequences in PTRE3GS and activates a high level of transcription. However, in the presence of decreasing amounts of Dox, expression is reduced, thus generating a titratable system of AID expression and SHM. PnP-iAID cells can be cultured for long periods of time with Dox in order to generate large libraries of protein variants in the IgH locus. Directed evolution and engineering of proteins can then be accomplished by high-throughput screening via flow cytometry (Fig. 5d).

**Description of the figures**

Fig. 1 shows the generation of PnP-mRuby cells. (A) Schematic shows wildtype (WT) hybridoma cells expressing antibody will be converted into PnP-mRuby. (B) Shown is the targeting of WT IgH genomic locus with the following annotations: leader sequence (L), mRNA splice sites (SS), VH and IgG constant heavy region (CH1). The CRISPR-Cas9 gRNA target site (black) is in the intron between VH and CH1. The donor construct consists of
mRuby gene with a stop codon flanked by two homology arms of 732 and 711 bp. The PnP-mRuby IgH locus is generated by transfection of WT cells with CRISPR-Cas9 plasmid (pX458) and donor construct, which will result in HDR-based exchange of the V_H region with mRuby. (C) Flow cytometry dot plot shows WT cells are exclusively IgH-negative, where PnP-mRuby cells are exclusively mRuby-positive and IgH-negative. (D) PGR was performed on genomic DNA from WT and PnP-mRuby cells using a forward primer in 5' HA and reverse primer that is external of the 3' HA. Agarose gel shows the expected size of bands. The band from PnP-mRuby cells was extracted and Sanger sequencing confirmed mRuby exchange of the V_H region. (E) Shown is the targeting of WT hybridoma IgK locus with the following annotations: V_L and IgK constant light region (CK), other annotations same as in shown in (A). Two gRNA target sites are utilized in order to delete the V_L region. (F) Flow cytometry dot plot shows WT cells are strongly IgH- and IgK-positive, where PnP-mRuby cells are exclusively IgH- and IgK-negative. (E) PGR was performed on genomic DNA from WT and PnP-mRuby cells using a forward primer 5' of the gRNA-F site and reverse primer 3' of gRNA-H site. Agarose gel shows the expected size of band for WT cells and nearly no amplification product for the PnP-mRuby cell line. Throughout this figure, WT cells correspond to clone WEN1.3 and PnP-mRuby cells correspond to clone 1E9.C3 (see table 1).

Fig. 2 shows the generation of PnP-mRuby hybridomas reprogrammed to surface express and secrete a new antibody. (A) Schematic shows PnP-mRuby cells expressing mRuby will be converted back into hybridomas expressing a new antibody via sFAb. (B) Shown is simplified design of the sFAb donor construct (for complete design details, see fig. S8). (C) Shown is the PnP-mRuby IgH locus where gRNA-J target site is in the mRuby gene. PnP-mRuby cells transfected with pX458.2 and sFAb donor will result in HDR-driven genomic replacement of mRuby. The new antibody will then be expressed on a single mRNA transcript. (D) Flow cytometry dot plot shows the different populations that emerge following transfection of PnP-mRuby with pX458 and sFAb donor. Cells which were positive for IgH expression were sorted. (E) Flow cytometry dot plot shows initial population of PnP-mRuby cells and resulting cells (PnP-HEL23) from sorted IgH-positive population in (D), which are now strongly positive for IgH and IgK expression. (F) Graph shows sandwich ELISA results (capture anti-IgK, primary detection anti-IgH) on hybridoma culture supernatant, PnP-HEL23 show IgG secretion levels similar to WT. (G) PGR was performed on the genomic DNA of WT, PnP-mRuby, PnP-HEL23 cells using primers shown in (C). Agarose gel from genomic PGR shows the predicted band size in PnP-HEL23. (H) RT-PCR from mRNA results in a visible band present of the correct size present only in PnP-HEL23. The bands from PnP-HEL23 were extracted and Sanger sequencing confirmed correct integration of the PnP-sFAb construct. Throughout this figure, WT cells correspond to clone WEN1.3, PnP-mRuby
cells correspond to clone 1E9.C3, and PnP-HEL23 correspond to clone Y (see table 1).

Fig. 3 shows PnP-HEL23 cells that surface express and secrete an antigen-specific antibody. (A) Flow cytometry histogram shows PnP-HEL23 cells surface express antibody specific for cognate antigen HEL. (B) ELISA data shows that PnP-HEL23 cells secrete antibody specific for HEL. WT cells correspond to clone WEN1.3, PnP-mRuby cells correspond to clone 1E9.C3, PnP-IgG cells correspond to clone Y (see table 1).

Fig. 4 shows the rapid and reproducible generation of PnP-antibody producing cells. (A) Flow cytometry dot plot shows PnP-mRuby following transfection with pX458.2 and different PnP-sFAb donor constructs. Cells were sorted for IgH expression. (B) Flow cytometry dot plots show that all three PnP cell lines express IgH and IgK following sorting in (A). (C) As in Fig. 2G and 2H, PGR and RT-PCR was performed on genomic DNA and mRNA, respectively. (D, E) Similar to (A) and (B) are flow cytometry dot plots for PnP-HEL23.2 cells, with the exception that Cas9 sorting step was omitted. PnP-HyHEL10 corresponds to clone U, PnP-EBV-2G4 corresponds to clone AA, and PnP-EBV-4G7 corresponds to clone AB, PnP-HEL23 corresponds to clone AC (see table 1).

Fig. 5 shows the generation of PnP hybridoma cell that express Cas9 and AID. (a) Shown is the CRISPR-Cas9 targeting of PnP-mRuby Rosa 26 locus. A donor construct will provide Cas9 gene with the 2A peptide and puromycin resistance gene. Integration into this locus will provide constitutive expression of Cas9 in all PnP-mRuby cells. (b) and (c) Shown is the integration of an inducible AID locus into PnP-mRuby-Cas9 cells, either by integration into the Rosa26 locus or into the native AID locus. (d) Schematic shows that large libraries will be generated by inducible synthetic somatic hypermutation via expression of AID. These libraries can then be used for directed evolution and high-throughput screening by flow cytometry.

Fig. 6 shows the validation of CRISPR-Cas9 targeting of immunoglobulin loci of hybridoma cells. (A) Flow cytometry dotplot shows expression of Cas9-2A-GFP in WEN1.3 cells following transfection with with pX458. (B) Surveyor results validate CRISPR-Cas9 targeting in IgH locus (agarose gels of all gRNA sites tested). (C) Surveyor results validate CRISPR-Cas9 targeting in IgK locus (agarose gels of all gRNA sites tested).

Fig. 7 shows targeting of the ROSA26 locus by CRISPR-Cas9. (A) ROSA26 locus (mouse chromosome 6) with CRISPR targets identified; displayed are also the primers used for fragment amplification for cleavage analysis. (B) DNA gels from Surveyor assay showing that all 4 tested guides induce successful CRISPR cleavage. The fragments match the expected size, shown in (A).
Fig. 8 shows design and validation of PnP-Cas9 cell lines. (A) The ROSA26 locus is targeted for CRISPR-Cas9 induced HDR integration of the constitutive Cas9 cassette. Contained in the cassette are two separate genes. The SpCas9-2A-Puromycin gene with bovine growth hormone (bCGh), the GFP gene is under transcriptional control of the murine pPGK promoter. The 5’ and 3’ homology arms are also present in the construct. 1: SV40 pA; 2: AID; 3: F2A; 4: GFP; 5: TRE3Gs promoter; 6: hPGK promoter; 7: Tet-On 3G; 8: SV40 pA; 9: homology arm; 10: pCAG promoter; 11: SpCas9; 12: T2A; 13: Puromycin; 14: bGH pA; 15: mPGK promoter; 16: GFP; 17: homology arm. (B) A close up of the Cas9 donor construct, shown are guide RNA target sites within GFP or pPGK, which are used to subsequently inactivate GFP by Cas9-induced NHEJ. (C) Sanger sequencing results before or after introduction of guide RNA in PnPCas9. (D) The T7E1 assay confirms that PnP-Cas9 cells in the presence of guide RNA lead to Cas9-induced NHEJ of GFP cells. (E) Flow cytometry plots show that in PnP-mRuby-Cas9 cells, the addition of gRNA targeting mRuby leads to knockout of mRuby expression in nearly all cells.

Fig. 9 shows generation and selection of PnP-iAID-mRuby cell lines. (A) Shown is the integration of the iSSHM donor cassette (GFP-2A-AID construct under a Doxinducible promoter system) is integrated into the Rosa26 locus of hybridoma cells by Cas9-induced HDR. Hybridoma cells express mRuby in their reprogrammed IgH locus. (B) Cell expressing Cas9 (2A-BFP) are sorted, then in the presence of Dox (or absence for negative controls), GFP-positive cells are single cell sorted and expanded. (C) Characterization of GFP expression with or without Dox in the single-cell sorted colonies from B.

Fig. 10 shows generation and selection of PnP-iAID-lgG cell lines. (A) Shown is the integration of the iSSHM donor cassette (GFP-2A-AID construct under a Doxinducible promoter system) is integrated into the Rosa26 locus of hybridoma cells by Cas9-induced HDR. Hybridoma cells express lgG through sFAb in their reprogrammed IgH locus. (B) Cell expressing Cas9 (2A-BFP) are sorted, then in the presence of Dox (or absence for negative controls), GFP-positive cells are single cell sorted and expanded. (C) Characterization of GFP expression with or without Dox in the single-cell sorted colonies from B.

Fig. 11 shows the general workflow for optimizing ssODN induced HDR and constructing a synthetic antibody (sAb) library with CDRH3 genetic diversity. (A) A.I: PnP-HEL23 sAb construct in the heavy chain gene locus with a CDRH3 sgRNA site to direct cleavage by Cas9 following transfection with the Cas9 vector, pX458. The Cas9 induced double stranded break introduces insertions/deletions (InDels) near the cut site through NHEJ causing a frameshift mutation and dysfunctional protein expression. The sequence shown in exploded view is SEQ ID No 17. A.IIa: Antibody expression can then be restored through HDR promoted by donor ssODNs with codon rearrangements for the CDRH3. A.lib: Additional
genetic diversity into the CDRH3 of the sAb cassette through HDR incorporation of
degenerate ssODNs (NNK randomization). Sorting results: upper left: A.1A upper panel; upper
right: A.1A lower panel ( △ A.lla/b upper panel); lower left: A.lla lower panel; A.1lb lower panel.
(B) HDR percentages are estimated by flow cytometric analysis though labeling with a
fluorescently tagged antigen. Percentages displayed are cells that had regained antigen
specific antibody expression towards HEL. Data presented is representative of n = 2
experiments after screening for transfection positive (GFP+) cells. 1: Cas9 plasmid; 2: Cas9
RNP; 3: Cas9 RNP + Modified ssODN; 4: Cas9 Cell + Modified ssODN.

Table 1 . Summary of hybridoma clones

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>WEN1.3</td>
<td>Wen 1.3 cells are derived from a mouse infected with LCMV. They express IgG2c and are specific for LCMV GP-1 antigen.</td>
</tr>
<tr>
<td>Cell Line</td>
<td>Selection</td>
<td>Description</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PnP-mRuby</td>
<td>1E9.C3</td>
<td>WEN 1.3 cells were transfected with pX458 with gRNA-E and mRuby donor construct and sorted for Cas9 positive expression (2A-GFP). This was followed by a first round of sorting for mRuby-positive cells, followed by a second single cell sort for mRuby. A single cell clone was selected and then transfected with pX458 with gRNA-F and H and sorted for Cas9 positive expression (2A-GFP). Cells were then sorted for IgK negative expression. A second round of single cell sorting was performed followed by genomic PGR to identify a clone with VL deletion. This final clone represents 1E9.C3</td>
</tr>
<tr>
<td>PnP-mRuby-pA</td>
<td>D2</td>
<td>PnP-mRuby-pA cells include a polyadenylation signal following the mRuby gene's stop codon to increase cell fluorescence. PnP-mRuby-pA cells were generated in an identical manner to PnP-mRuby cells, but with a donor construct including the polyadenylation signal.</td>
</tr>
<tr>
<td>PnP-mRuby-Cas9</td>
<td>1AD</td>
<td>Clone D2 was transfected with pX458 with gRNA-P and Cas9-2A-Puro-GFP HDR donor linearized. Cells were sorted for Cas9 positive expression (2A-BFP) and expanded. Cells were sorted for GFP positive expression and expanded. Cells were then cultured in growth medium supplemented with 2.5 μg/ml puromycin for up to one week. Single cells were sorted for GFP positive expression and expanded. A suitable clone was selected based on genotypic and phenotypic characterization.</td>
</tr>
<tr>
<td>PnP-HEL23</td>
<td>Y</td>
<td>Clone 1E9.C3 was transfected with pX458 with gRNA-J and HEL23-2A HDR donor linearized. Cells were sorted for Cas9 positive expression (2A-BFP) and expanded. Cells were then sorted for surface IgH expression and expanded, and finally characterized for IgH and IgK expression.</td>
</tr>
<tr>
<td>PnP-HEL23-IgH&lt;sup&gt;-&lt;/sup&gt;</td>
<td>IgH&lt;sup&gt;-&lt;/sup&gt;</td>
<td>Clone Y was transfected with pX458 with gRNA-Q. Cells were sorted for Cas9 positive expression (2A-GFP) and expanded. A single cell sort for cells lacking surface IgH expression was performed followed by genomic PGR and Sanger sequencing to genotypically characterize the individual clones. A suitable clone was selected based on genotypic characterization.</td>
</tr>
<tr>
<td>PnP-HyHEL10</td>
<td>U</td>
<td>Clone 1E9.C3 was transfected with pX458 with gRNA-J and HyHELIO- 2A HDR donor linearized. Cells were sorted for Cas9 positive expression (2ABFP) and expanded. Cells were then sorted for surface IgH expression and expanded, and then they underwent a</td>
</tr>
</tbody>
</table>
second IgH sort. They were finally characterized for IgH and IgK expression.

<table>
<thead>
<tr>
<th>Clone</th>
<th>HDR</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PnP-EBV-2G4</td>
<td>AA</td>
<td>Clone 1E9.C3 was transfected with pX458 with gRNA-J and 2G4-2A HDR donor linearized. Cells were sorted for Cas9 positive expression (2A-BFP) and expanded. Cells were then sorted for surface IgH expression and expanded, and then they underwent a third sort for IgH and IgK expression. They were finally characterized for IgH and IgK expression.</td>
</tr>
<tr>
<td>PnP-EBV-4G7</td>
<td>AB</td>
<td>Clone 1E9.C3 was transfected with pX458 with gRNA-J and 4G7-2A HDR donor linearized. Cells were sorted for surface IgH expression and expanded, and then they underwent a third sort for IgH and IgK expression. They were finally characterized for IgH and IgK expression.</td>
</tr>
<tr>
<td>PnP-HEL23-2.0</td>
<td>AC</td>
<td>Clone 1E9.C3 was transfected with pX458 with gRNA-J and HEL23-2A HDR donor linearized. Cells were NOT sorted for Cas9 positive expression. They were sorted for surface IgH expression and expanded. They were finally characterized for IgH and IgK expression. [In order to achieve a higher purity, cells were eventually sorted a second time for IgH and IgK expression]</td>
</tr>
<tr>
<td>PnP-mRuby-pA-</td>
<td>(winner</td>
<td>PnP-mRuby-pA cells were transfected with px458-BFP with gRNA-0 and sorted for Cas9 expression (2A-BFP). Cells were then induced by 1 µg/ml Doxycycline, single-cell sorted for GFP expression, and characterized by further induction cycles (GFP+, mRuby knock-out activity), genotyping and transcript analysis.</td>
</tr>
<tr>
<td>AID</td>
<td>selection in progress)</td>
<td></td>
</tr>
<tr>
<td>(winner selection in progress)</td>
<td>PnP-HEL23-lgH cells were transfected with px458-BFP with gRNA-0 and sorted for Cas9 expression (2A-BFP). Cells were then induced by 1 µg/ml Doxycycline, single-cell sorted for GFP expression, and characterized by further induction cycles, genotyping and transcript analysis.</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. List of gRNAs

<table>
<thead>
<tr>
<th>Target region</th>
<th>Targeting sequence (5′-3′ Sequence + PAM)</th>
<th>Resident plasmid</th>
</tr>
</thead>
</table>

21
<table>
<thead>
<tr>
<th>gRNA</th>
<th>Leader/Downstream</th>
<th>SEQ ID NO</th>
<th>Overhang</th>
<th>Booster</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Wenl.3 leader-VH</td>
<td>01:</td>
<td>GCTGTCGGGAGAAAGAAATTGTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>B</td>
<td>leader-λH</td>
<td>02:</td>
<td>GCCCTATCTCTCTCTGATTTG</td>
<td>pX458</td>
</tr>
<tr>
<td>C</td>
<td>Wenl.3 leader-VH</td>
<td>03:</td>
<td>GTTCCAATCTGAAGAAGGATAGG</td>
<td>pX458</td>
</tr>
<tr>
<td>D</td>
<td>Wenl.3 JH downstream</td>
<td>04:</td>
<td>GGAGCATGACGAGACTACTTCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>E</td>
<td>Wenl.3 JH downstream</td>
<td>05:</td>
<td>GTTGGTTTAGCGGAGTCCCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>F</td>
<td>Wenl.3 VK leader</td>
<td>06:</td>
<td>GGAGAACGAGACCACATAGCAGG</td>
<td>pX458</td>
</tr>
<tr>
<td>G</td>
<td>Wenl.3 VK leader</td>
<td>07:</td>
<td>GGCTATGGTCCTGCTTCTCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>H</td>
<td>Wenl.3 JH downstream</td>
<td>08:</td>
<td>GGGATCTTCTATGCTACAGG</td>
<td>pX458</td>
</tr>
<tr>
<td>I</td>
<td>Wenl.3 JH downstream</td>
<td>09:</td>
<td>GTGGCTAAATGAGCCATTCCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>J</td>
<td>mRuby2</td>
<td>10:</td>
<td>GTCATGGAAGGTTGGGATTGCAAGG</td>
<td>pX458.2 (BFP)</td>
</tr>
<tr>
<td>K</td>
<td>mRuby2</td>
<td>11:</td>
<td>GCATGCCGTGATCANCGCCTGG</td>
<td>pX458.2 (BFP)</td>
</tr>
<tr>
<td>L</td>
<td>ROSA26</td>
<td>12:</td>
<td>GAGACCTCCATCGGCACCTCCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>M</td>
<td>ROSA26</td>
<td>13:</td>
<td>GCAGACCTCCATCGGCACCTCCTGG</td>
<td>pX458</td>
</tr>
<tr>
<td>N</td>
<td>ROSA26</td>
<td>14:</td>
<td>GCCCTCGATGAAAATGACTCCGAGG</td>
<td>pX458</td>
</tr>
<tr>
<td>O</td>
<td>ROSA26</td>
<td>15:</td>
<td>GCGATGGAAAAATGACTCCGAGG</td>
<td>pX458 (BFP)</td>
</tr>
</tbody>
</table>
The sequences listed in this table (SEQ ID NO 01 - SEQ ID NO 18) refer to the DNA sequences encoding the targeting sequences of the respective gRNAs.

Methods

5 Hybridoma cell culture conditions

The WT hybridoma cell line (Wen1.3) was obtained as a gift from Prof. Annette Oxenius (ETH Zurich). All hybridoma cell lines were cultivated in high-glucose Dulbecco’s Modified Eagle Medium [(DMEM), Thermo Fisher Scientific (Thermo), 11960-044] supplemented with 10% heat inactivated fetal bovine serum [(FBS), Thermo, 10082-147], 100 U/ml Penicillin/Streptomycin (Thermo, 15140-122), 2 mM Glutamine (Sigma-Aldrich, G7513), 10 mRS M HEPES buffer (Thermo, 15630-056) and 50 µM 2-mercaptoethanol (Sigma-Aldrich, M3148). All hybridoma cells were maintained in incubators at a temperature of 37 °C and 5% CO2. Hybridomas were typically maintained in 10 ml of culture in T-25 flasks (Thermo, NC-156367), and split every 48/72 hours.

15 Cloning and assembly of CRISPR-Cas9 targeting constructs

Unless otherwise noted, cloning of CRISPR-Cas9 plasmids and HDR donor constructs was done by Gibson assembly and cloning with the Gibson Assembly® Master Mix (NEB, E2611S) (Gibson et al., Nat Methods 2009, 6:343-345). When necessary, fragments for the Gibson assembly cloning were amplified with the KAPA HiFi HotStart Ready Mix [KAPA Biosystems (KAPA), KK2602]. All gRNAs were obtained from Integrated DNA Technologies (IDT) as single-stranded 5’-phosphorylated oligonucleotides purified by standard desalting. The basis for CRISPR-Cas9 experiments relied on the plasmid pSpCas9(BB)-2A-GFP (pX458), obtained as a gift from Feng Zhang (Addgene plasmid # 48138) (Ran et al., Nat Protoc 2013, 8:2281-2308). An alternate version of pX458 was generated by replacing the GFP (eGFP variant) with BPF (TagBFP variant) (pX458.2 or pSpCas9(BB)-2A-BFP). For cloning gRNAs, both versions of pX458 were digested with BbsI [New England BioLabs (NEB), R0539S], gRNA oligonucleotides were ligated into plasmids with DNA T4 ligase
(NEB, M0202S). The gene for mRuby (mRuby2 variant) was derived from the plasmid pcDNA3-mRuby2, a gift from Michael Lin (Addgene plasmid # 40260) (Lam et al., Nat Methods 2012, 9:1005-1012; Jinek et al., eLife 2013, 2:e00471-e00471). The HDR donors (mRuby and the antibody constructs) were cloned in the pUC57(Kan)-HDR plasmid, obtained from Genewiz. The vector was designed with homology arms according to the annotated mouse genomic sequence (GRCm38). The 2A antibody constructs were obtained as synthetic gene fragments (gBlocks, IDT). The HDR donor vectors were linearized by PGR with the KAPA HiFi HotStart ReadyMix (KAPA Biosystems, KK2602). All plasmid and linear versions of HDR donors, as well as pX458 and pX458-BFP, were ethanol precipitated as a final purification step.

**Hybridoma transfection with CRISPR-Cas9 constructs**

Hybridoma cells were transfected with the 4D-Nucleofector™ System (Lonza) using the SF Cell Line 4D-Nucleofector® X Kit L (Lonza, V4XC-2024) with the program CQ-104. Cells were prepared as follows: 10^6 cells were isolated and centrifuged at 90 xG for 5 minutes, washed with 1 ml of Opti-MEM® | Reduced Serum Medium (Thermo, 31985-062), and centrifuged again with the same parameters. The cells were finally re-suspended in 100 μl of total volume of nucleofection mix, containing the vector(s) diluted in SF buffer (per kit manufacturer guidelines). For the exchange of V<sub>H</sub> locus, 5 μg of pX458 (or pX458-BFP) with gRNA-E (targeting V<sub>H</sub>) or gRNA-J (targeting mRuby), and 5 μg of the circular or linear HDR donor constructs were nucleofected into cells. For V<sub>L</sub> deletion, 5 ug each of pX458 with gRNA-F and gRNA-H were co-transfected into cells. Following transfection, the cells were typically cultured in 1 ml of growth media in 24-well plates (Thermo, NC-142475). When a significant cell expansion was observed, cells were supplemented 24 hours later with 0.5-1.0 ml of fresh growth media. After sorting, typically 48 hours after transfection, cells were recovered in 24-well plates, and progressively moved into 6-well plates (Thermo, NC-140675) and T-25 flasks, following expansion. After replacing the V<sub>H</sub> with mRuby, cells were single-cell sorted in U-bottom 96-well plates (Sigma-Aldrich, M0812) in a recovery volume of 100 μl. The clones were eventually expanded in 24-well plates, 6-well plates and T-25 flasks.

**Genomic and transcript analysis of CRISPR-Cas9 targeting**

Genomic DNA of hybridoma cell lines were recovered from typically 106 cells, which were washed with PBS by centrifugation (250 xG, 5 minutes) and re-suspended in QuickExtract™ DNA Extraction Solution (Epicentre, QE09050). Cells were then incubated at 68 C for 15 minutes and 95 C for 8 minutes. For transcript analysis, total RNA was isolated from 106 - 5 x 106 cells. The cells were lysed with TRIzol® reagent (Thermo, 15596-026) and total RNA was extracted with the Direct-zol™ RNA MiniPrep kit (Zymo Research, R2052). Maxima Reverse Transcriptase (Thermo, EP0742) was used for cDNA synthesis
from total RNA (Taq DNA Polymerase with ThermoPol® Buffer, NEB, M0267S). Both genomic DNA and cDNA were used as templates for downstream PGR reactions.

The gRNAs targeting WT IgH and IgK loci and mRuby were initially tested for their activity by 30 induction of NHEJ. The targeted fragment was amplified by PGR with KAPA2G Fast ReadyMix (KAPA, KK5121) and the PGR product digested with the Surveyor nuclease for the detection of mismatches (Surveyor® Mutation Detection Kit, IDT, 706020). For HDR evaluation, PGR was performed on genomic and cDNA using primers binding inside and outside homology arms, followed by fragment size analysis on DNA agarose gels. Selected PGR products were subjected to Sanger sequencing.

10 Flow cytometry analysis and sorting of hybridomas

Flow cytometry-based analysis and cell isolation were performed using the BD LSR Fortessa™ and BD FACS Aria™ III (BD Biosciences), respectively. At 24 hours post-transfection, approximately 100 µl of cells were harvested, centrifuged at 250 xG for 5 minutes, resuspended in PBS and analyzed for expression of Cas9 (via 2A-GFP/-BFP). 48 hours post-transfection, all transfected cells were harvested and resuspended in Sorting Buffer (SB): PBS supplemented with 2 mM EDTA and 0.1% BSA). When labeling was required, cells were washed with PBS, incubated with the labeling antibody or antigen for 30 minutes on ice, protected from light, washed again with PBS and analyzed or sorted. The labeling reagents and working concentrations are described in table 3 below. For cell numbers different from 10^6, the antibody/antigen and incubation volume were adjusted proportionally.

Table 3. Flow cytometry labeling reagents with their working concentrations

<table>
<thead>
<tr>
<th>Target antigen</th>
<th>Working conc.</th>
<th>Dilution from stock</th>
<th>Incubation volume</th>
<th>Fluorophore</th>
<th>Product ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG2C</td>
<td>3.3 µg/ml</td>
<td>1:150</td>
<td>100 µl</td>
<td>Allophycocyanin (APC)</td>
<td>115-135-208 (Jackson ImmunoResearch)</td>
</tr>
<tr>
<td>IgG2C</td>
<td>1.6 mg/ml</td>
<td>1:100</td>
<td>100 µl</td>
<td>AlexaFluor® 488</td>
<td>115-547-188 (Jackson ImmunoResearch)</td>
</tr>
<tr>
<td>IgK</td>
<td>2.5 µg/ml</td>
<td>1:80</td>
<td>100 µl</td>
<td>Brilliant Violet 421™</td>
<td>409511 (BioLegend)</td>
</tr>
<tr>
<td>Hen egg lysozyme</td>
<td>0.1 µg/ml</td>
<td>1:62.5</td>
<td>100 µl</td>
<td>AlexaFluor® 647</td>
<td>62971-10G-F (Sigma- Aldrich)</td>
</tr>
</tbody>
</table>
Measurement of antibody secretion by ELISA

Sandwich ELISAs were used to measure the secretion of IgG from hybridoma cell lines. Plates were coated with capture polyclonal antibodies specific for Vk light chains (goat anti-mouse, Jackson ImmunoResearch, 115-005-174) concentrated at 4 µg/ml in PBS (Thermo, 10010-015). Plates were then blocked with PBS supplemented with 2% w/v milk (AppliChem, A0830) and 0.05% v/v Tween®-20 (AppliChem, A1389) (PBSMT). Supernatants from cell culture (10^6 cells/sample, volume normalized to least concentrated samples) were then serially diluted (at 1:3 ratio) in PBS supplemented with 2% w/v milk (PBSM). As a positive control, a purified mouse IgG2b, κ isotype control (BioLegend, 401202) was used at a starting concentration of 5 ng/µl (diluted in hybridoma growth media) and serially diluted as the supernatants. After blocking, supernatants and positive controls were incubated for 1 hour at RT or O/N at 4 °C, followed by 3 washing steps with PBS supplemented with Tween-20 0.05% v/v (PBST). A secondary HRP-conjugated antibody specific for mouse Fc region was used (goat anti-mouse, Sigma-Aldrich, A2554), concentrated at 1.7 µg/ml in PBSM, followed by 3 wash steps with PBST. ELISA detection was performed using a 1-StepTM Ultra TMB-ELISA Substrate Solution (Thermo, 34028) as the HRP substrate. Absorbance at 450 nm was read with Infinite® 200 PRO NanoQuant (Tecan). For antigen specificity measurements, plates were coated with purified hen egg lysozyme (Sigma-Aldrich, 62971-10G-F) concentrated at 4 µg/ml in PBS. Blocking, washing, and supernatant incubation steps were made analogously to the previously described procedure, with the exception of serial dilutions of supernatants at 1:5 ratios. A secondary HRP-conjugated antibody was used specific for Vk light chain (rat anti-mouse, Abeam, AB99617) concentrated at 0.7 µg/ml. ELISA detection by HRP substrate and absorbance reading was performed as previously stated.

Targeting of the ROSA26 locus

The mouse safe harbor locus ROSA26 was amplified and Sanger sequenced from Wen1.3 cells, and the sequence obtained was used to design DNA cassette homology arms. Guide RNA target sequences (gRNA-L to gRNA-P) were individually validated by Surveyor Assay (Fig. 7). This locus was targeted for the creation of the PnP-mRuby-AID, PnP-IgG-AID and the PnP-mRuby-Cas9 cell lines.

For the targeting of the ROSA26 in the creation of the PnP-mRuby-AID and the PnP-mRuby-Cas9 cell lines, gRNA-0 and gRNA-P were selected due to their high cleavage efficiency. Generation of the AID cell lines and induction by Doxycycline

Cloning of the donor cassette for the inducible AID (iSSHM) system was performed in three steps. (1) The Tet-One™ Inducible Expression System was purchased from Takara Clontech
(634301); GFP-2A-AID was obtained as synthetic gene fragment (gBlocks, IDT) and cloned into the pTetOne vector by Gibson assembly cloning. (2) Homology arms (829 and 821 bp) for the hybridoma's ROSA26 locus were obtained by genomic DNA PGR and cloned in a pUC57(Kan) plasmid. (3) Finally, the previously cloned - see point (1) - Tet-One-GFP-2A-AID construct (containing, in the forward orientation: the human phosphoglycerate kinase 1 promoter (hPGK), the Tet-On 3G transactivator protein, and the SV40 poly-A signal; in the reverse orientation: the P_{TRE3GS} Inducible promoter, the GFP-2A-AID construct and the SV40 poly-A signal) was inserted between the homology arms through Gibson assembly cloning. The HDR donor was linearized by PGR with restriction digestion with the enzyme Aful (Thermo, ER1951). gRNA-0 was obtained and cloned as previously described in pX458-BFP. As an alternative NHEJ insertion design, the TetOne-GFP-2A-AID construct is linearized without homology arms. The cell lines engineered for introduction of the TetOne-iSSHM system are: PnP-mRuby-pA (PnP-mRuby with a bGH poly-A tail); PnP-HEL23-lgH⁺ (PnP-HEL23 with a frameshift insertion in the HCDR3 knocking out antibody expression - see next sections). The workflow for these cells are shown in Fig. 9 and 10.

a. Cell lines generation

From the transfection stage, the cells were kept in Tet-free GM: regular growth media supplemented with Tet System Approved FBS, US-sourced (Takara Clontech, 631 105). PnP-HEL23-lgH⁺ and PnP-mRuby-pA cells were transfected with -2.5 pg px458-BFP with gRNA-0 and 2.5 pg linearized pTetOne-HDR donor (see previous section). 48 hours after transfection the cells were sorted for BFP and grown for recovery. Once recovered, induction experiments were performed to verify the system's functionality: Doxycycline (Takara Clontech, 631 311) was dissolved in nuclease-free H₂O at 1 mg/ml, sterile filtered and diluted in Tet-free GM at need directly before use. Concentrations in the range between 1 ng/ml and 2 pg/ml were tested, with 1 pg/ml proving to be the most efficient. Cells were induced by incubation at 37 C for 24 or 48 hours; however, 24 hour incubation gave the best results and was selected as main condition to check for positive integration and induction efficiency.

The cells were sorted after 24 hours of induction: from the GFP positive population, single-cell clones were isolated, grown and characterized. An initial screening was performed to select the most positive clones after Dox induction: each sample was seeded at 1 pg Dox/10⁵ cells/1 ml culture and screened for GFP 24 hours later.

The best performing clones from the initial screening steps (9 for PnP-HEL23-lgH⁺, 4 for PnP-mRuby-pA) were used for genomic DNA extraction, locus-specific amplification and Sanger sequencing. According to the genomic sequence, one final clone was selected for each cell line.
b. Induction optimization

After selection of the best clone for each transfected cell line, a second and tighter titration was performed with concentrations in the range between 500 ng/ml and 1.5 μg/ml, and induction measured at different time points (ideally: 24 hrs; 48 hrs; 72 hrs; 96 hrs. Due to Doxycycline having a half-life of 24 hours, it was replaced in culture every 48 hours, as recommended by the manufacturer (Clontech).

For each time point, induction was assessed by:

- FACS analysis (GFP)
- RT-PCR (mRNA/cDNA)
- Western Blot

Amplification of AID from cDNA was performed with KAPA HiFi HotStart Ready Mix. For Western Blot, M-PER™ Mammalian Protein Extraction Reagent (Thermo, 78501), supplemented with Halt™ Protease Inhibitor Cocktail (Thermo, 78430) was used to obtain lysates from cultured hybridomas, typically from 10^6 cells. Anti-Human/Mouse Activation-Induced Cytidine Deaminase (AID) Purified (eBioscience, 14-5959-80) was used as primary antibody for AID detection via WB.

c. iSSHM (AID activity) assessment

For the PnP-mRuby-pA-AID cell line, hypermutation activity was first evaluated by FACS analysis and detection of decreasing mRuby fluorescence. For a more thorough evaluation, the mRuby gene was amplified from cDNA and analyzed by Sanger or next-generation sequencing (NGS) using the method of molecular amplification fingerprinting (Fig. 12).

For PnP-HEL23-lgH —AID cell line, restored antibody expression and/or antigen-specificity was evaluated by flow cytometry after labelling cells (see previous sections) using anti-lgG2C and anti-lgK (any positivity arisen through random mutations) and HEL-647 (re-gained HEL positivity).

For more definitive assessment and optimization of the system, the iSSHM workflow was repeated for a cell line (obtained by either of the two starting platforms) expressing a functional antibody. To obtain such a situation, PnP-mRuby-pA cells were transfected to exchange mRuby with a sAb donor; PnP-HEL23-lgH —AID cells were transfected to exchange the knocked-out HEL23 sAb with a functional one with the same or another specificity; in an alternative setting, the HEL23 frame was restored by HDR with a 120 ssODN containing a codon-mutated version of the original HCDR3. In case of no previously tested binders, antibodies with a known and testable antigen were typically chosen to evaluate affinity maturation.
Once obtained a PnP-sAb-AID cell line, AID was induced like previously described (ON). After a 48-96 hours induction, Doxycycline was retracted from the system (OFF). To assess affinity maturation, FACS labelling was performed as previously described, but with decreasing antigen concentrations (typically 1-0.001 μg/ml, decreasing ten-fold for each round of analysis/affinity maturation cycle). Positive cells were sorted and underwent a further iSSHM cycle; the cycle was repeated as needed. For each stage of affinity maturation, after Doxycycline retraction from the system, the effective switch OFF was evaluated by monitoring GFP fluorescence by FACS. Once in the OFF state, V_L and V_H regions were amplified from cDNA and iSSHM was evaluated by NGS using the method of molecular amplification fingerprinting.

**Generation of the PnP-mRuby2-Cas9 cell line**

Cloning of the donor cassette for constitutive expression of the Cas9 protein was performed in three steps. (1) The pSpCas9(BB)-2A-Puro vector (pX459) and MDH1-PGK-GFP_2.0 vector were obtained from Addgene (plasmid # 48139, # 11375, respectively). The Cas9-2A-Puro and GFP gene fragments were obtained from their respective vectors through PGR (KAPA HiFi HotStart ReadyMix) and assembled together with Gibson assembly cloning. (2) Homology arms (1,000 and 976 bp) for the hybridoma’s ROSA26 locus were obtained by genomic DNA PGR and assembled with the pUC57(Kan) plasmid backbone through Gibson assembly cloning. (3) Finally, the previously assembled fragments - see points (1 and 2) - were assembled through Gibson assembly cloning. The HDR donor was linearized by restriction digestion with the XhoI and MluI restriction endonucleases followed by gel electrophoresis purification. gRNA-P was cloned as previously described in pX458 (BFP). For the alternative NHEJ insertion design, the Cas9-2A-Puro-GFP construct was linearized without homology arms.

Following transfection with the Cas9-2A-Puro-GFP donor, GFP+ cells were sorted and expanded. Cells were then selected for stable integration of the donor construct by culturing in regular growth media supplemented with 2.5 μg/mL of Puromycin (Thermo, A113802) for up to one week before single-cell isolation, growth and PGR characterization. After identification of a single clone with correct integration of the Cas9-2A-Puro-GFP cassette, Cas9 activity within the cell was validated through transfection of a guide RNA targeting GFP. Cas9 cleavage activity was validated by T7E1 assay and Sanger sequencing of PGR amplicons (Fig. 8). GFP knock out effectiveness was confirmed by flow cytometry.

The cell lines engineered for constitutive Cas9 expression are: PnP-mRuby-pA; PnP-HEL23-IgH+. A more comprehensive cell line was designed to incorporate both the constitutive Cas9 and the inducible AID; due to the constructs bearing homology with two different regions of the ROSA26 locus, it was possible to incorporate them in tandem like shown in Fig. 8A. This
allowed us to merge the high HDR efficiency achieved by constitutively expressing Cas9 with the iSSHM workflow.

_Cas9 cells - In vitro transcription of guide RNA or synthetic oligonucleotides (IDT)_

The PnP-mRuby-Cas9 cell line, constitutively expressing Cas9, was transfected with the appropriate HDR donor and already transcribed guide RNAs. The latter were obtained as oligonucleotides from IDT or in vitro transcribed. In the case of in vitro transcription, the previously described guide-DNA oligodeoxynucleotides (see Cloning and assembly CRISPR-Cas9 targeting constructs section) served as templates for the MEGAscript® T7 transcription kit (Thermo, AM1334). An adapted protocol was as described previously (https://www.protocols.io/view/in-vitro-transcription-of-guide-RNAs-d4w8xd?step=3 accessed Feb. 21, 2017).
Claims

1. A method for the generation of a cell line, comprising the steps of

   a. providing a plurality of mammalian B cells, wherein each of said plurality of B cells comprises a transgenic genomic DNA sequence encoding a marker protein, wherein said transgenic genomic DNA sequence is inserted into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, and wherein said transgenic genomic DNA sequence is amenable to cleavage by a site-directed nuclease, particularly CRISPR-associated endonuclease (Cas9);

   b. replacing said transgenic genomic DNA sequence encoding said marker protein with a second transgenic DNA sequence encoding a protein of interest;

   c. sorting said B cells based on the presence or absence of said marker protein; and

   d. collecting B cells in which said marker protein is absent.

2. The method according to claim 1, wherein said plurality of B cells is selected from the group comprising primary B cells, immortalized B cells, hybridoma cells, myeloma cells, plasmacytoma cells, and lymphoma cells.

3. The method according to any one of the preceding claims, wherein said marker protein and/or said protein of interest is expressed under control of an endogenous immunoglobulin promoter, particularly the V\textsubscript{H} promoter.

4. The method according to any one of the preceding claims, wherein in each of said plurality of B cells the endogenous V\textsubscript{H} gene and the endogenous V\textsubscript{L} gene are disrupted.

5. The method according to any one of the preceding claims, wherein said plurality of B-cells is genetically modified to constitutionally express said CRISPR-associated endonuclease (Cas9).

6. The method according to any one of the preceding claims, wherein said plurality of B-cells is genetically modified to express an activation-induced cytidine deaminase (AID) in an inducible and titratable manner.
7. The method according to any one of the preceding claims, wherein said plurality of B cells comprises a safe harbor locus, and wherein a first expression cassette comprising a DNA sequence encoding said CRISPR-associated endonuclease is inserted into said safe harbor locus.

8. The method according to any one of the preceding claims, wherein said plurality of B cells comprises a safe harbor locus, and wherein a second expression cassette comprising a DNA sequence encoding said activation-induced cytidine deaminase (AID) is inserted into said safe harbor locus.

9. The method according to any one of the preceding claims, wherein said replacing of said transgenic genomic DNA sequence is mediated by Cas9-mediated site-directed DNA cleavage and subsequent integration of said second transgenic DNA sequence by homology directed repair (HDR) or non-homologous end joining (NHEJ).

10. The method according to any one of the preceding claims, wherein said protein of interest is a full-length antibody, in particular a full-length antibody based on a synthetic antigen binding fragment (sFab) construct.

11. A method for the generation of a library of protein variants, comprising a method according to any one of the preceding claims, followed by the additional step of modifying regions of said transgenic DNA with randomized nucleic acid sequences (through randomized regions on donor dsDNA or ssDNA).

12. A method for the generation of a library of protein variants, comprising a method according to any one of the preceding claims 6 to 11, followed by the additional step of inducing expression of said activation-induced cytidine deaminase (AID), thus generating multiple genomic mutations within said protein of interest by inducible synthetic somatic hypermutation (iSSHM).

13. The method according to any one of the preceding claims, wherein said marker protein is a fluorescent protein.

14. A human B cell line comprising an expressed transgenic genomic DNA sequence inserted into an endogenous immunoglobulin locus comprised in said B cell obtained by a method of any one of the preceding claims.

15. A mammalian B cell, wherein

a. a transgenic genomic DNA sequence encoding a marker protein is inserted
into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, wherein said marker gene encodes a fluorescent protein, and wherein

b. said transgenic genomic sequence is amenable to cleavage by a site-directed nuclease, particularly CRISPR-associated endonuclease (Cas9).

16. The mammalian B cell according to claim 15, wherein in said B cell the endogenous $V_H$ gene and the endogenous $V_L$ gene are disrupted.

17. The mammalian B cell according to any one of claims 15 or 16, wherein said cell is a human cell.

18. A plurality of mammalian B cells, wherein each member of said plurality of B cells comprises a transgenic genomic DNA sequence encoding a variant of a protein or interest, wherein said transgenic genomic DNA sequence is inserted into an endogenous immunoglobulin locus comprised in said B cell, particularly an IgH locus, and each variant encoded by a member of said plurality is different from any other variant encoded by another member of said plurality.

19. The plurality of mammalian B cells according to claim 18, wherein each variant is different from another variant in one to five positions of its amino acid sequence.

20. The plurality of mammalian B cells according to claim 18 or 19, wherein each variant is at least 80%, 85%, 90%, 95%, 97%, 98% or 99% identical to any another variant encoded by a member of said plurality.
Fig. 1

A

WT

PnP-mRuby

B

WT IgH locus

L SS V\text{H} gRNA-E IgG CH1

5' HA mRuby 3' HA

732 bp 711 bp

mRuby replaces V\text{H}

PnP-mRuby IgH locus
Fig. 4

A

PnP-HyHEL10

Q1 0.39%
96.7%
Q4

mRuby (561 nm)

PnP-2G4

Q1 0.18%
61.6%
Q4

PnP-4G7

Q1 0.33%
76.8%
Q4

B

PnP-HyHEL10

Q1 0.74%
8.46%
Q4

IgK (405 nm)

PnP-2G4

Q1 0.92%
1.18%
Q4

PnP-4G7

Q1 8.22%
7.01%
Q4

C

WT

PnP-HyHEL10

PnP-2G4

PnP-4G7

PnP-HEL23.2

2000 bp

1500 bp

1000 bp

D

PnP-HEL23.2

Q1 0.14%
38.7%
Q4

mRuby (561 nm)

Q2 0.033%
61.1%
Q3

E

PnP-HEL23.2

Q1 1.07%
8.97%
Q4

IgK (405 nm)

Q2 62.2%
27.8%
Q3
A Cas9 donor construct

Promoter  Cas9  Puro

HA1  HA2

PnP-mRuby Rosa26 locus  gRNA-X

+ CRISPR-Cas9 (pX458)

B PnP-mRuby-Cas9 Rosa26 locus

HA1  pTREG3G  Tet-on  BFP  2A  AID  HA2

gRNA-Y  + Cas9 (endogenous or via plasmid)

C PnP-mRuby AID locus

gRNA-Z  AID

+ endogenous Cas9
Fig. 5 (continued)
Fig. 7 (continued)

b

<table>
<thead>
<tr>
<th>CRISPR guides 1&amp;2</th>
<th>CRISPR guides 3&amp;4</th>
</tr>
</thead>
<tbody>
<tr>
<td>gRNA</td>
<td>+</td>
</tr>
<tr>
<td>Cas9</td>
<td>-</td>
</tr>
</tbody>
</table>

1000 bp
500 bp
200 bp
Sanger Sequencing Results

C

gRNA (-)  

GFP sgRNA Target  

eGFP

pPGK Fwd Primer

GFP Rev Primer

14:30

gRNA (+)  

GFP sgRNA Target  

eGFP

pPGK Fwd Primer

GFP Rev Primer
Fig. 8 (continued)

<table>
<thead>
<tr>
<th></th>
<th>T7E1 Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7E1</td>
<td>+ - + -</td>
</tr>
<tr>
<td>gRNA</td>
<td>+ + - -</td>
</tr>
</tbody>
</table>

1500 bp
1000 bp
750 bp
500 bp
250 bp
Fig. 8 (continued)

E

PnP-mRuby-Cas9

PnP-mRuby-Cas9 + gRNA targeting mRuby

mRuby
Fig. 9 (continued)
Fig. 9 (continued)

C

B9

E7

D4

E9

normalized to mode

GFP (488 nm)

+ Dox

- Dox

Negative ctrl
PnP-mRuby-pA
Fig. 10 (continued)

Normallized to mode

GFP (488 nm)

Negativectrl
PnP-HEL23-9PH

+ Dox
- Dox

3C10
1F4b
1F4
1D6
3C10b
3C5b
3D11b
1C9

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$

$10^0$ $10^2$ $10^4$ $10^6$
Fig. 11 (continued)
B

Functional Antibody Expression & HDR Editing Efficiencies

Antibody Expression  HEL⁺ (HDR)
a.

- PnP-mRuby-AID Red-high
- mRuby PCR (gDNA)
- Cloning in bacteria
- Sequencing single colonies

+ Dox
72 hrs

mRuby → mRuby

PnP-mRuby-AID Red-low
Fig. 12 (continued)

B

Dox-induced PnP-mRuby-AID

PnP-mRuby (Ctrl)

PnP-mRuby-AID Red-
low

PnP-mRuby-AID Red-
high

0.54% 39.2%

0.89% 89.1%

20.3% 79.7%

96.7% 1.27%

Count

mRuby (561 nm)

mRuby (561 nm)

GFP (488 nm)
Fig. 12 (continued)

iSSHM mutated mRuby - Mutations per kb

Mutated nucleotides/sequenced kb

0 10 20 30 40 50 60 70 80 90

(0.14) (0.26) (0) (84.39)

PnP-mRuby PnP-mRuby-AID PnP-mRuby-AID Red-high PnP-mRuby-AID Red-low
**INTERNATIONAL SEARCH REPORT**

**Inv. C12N5/0781 C12N15/64 C12N15/90**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, BIOSIS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>paragraphs [0144], [0147]; examples 4-6; table 31</td>
<td>1-13</td>
</tr>
<tr>
<td>X</td>
<td>TAEK-CHIN CHEONG ET AL: &quot;Editing of mouse and human immunoglobulin genes by CRISPR-Cas9 system&quot;, NATURE COMMUNICATIONS, vol. 7, 9 March 2016 (2016-03-09), page 10934, XP055272094, UK, ISSN: 2041-1723, DOI: 10.1038/ncomms10934</td>
<td>14-20</td>
</tr>
<tr>
<td>Y</td>
<td>page 2, last paragraph</td>
<td>1-13</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) one of which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

* document member of the same patent family

**Date of the actual completion of the international search**

2 May 2017

**Date of mailing of the international search report**

22/05/2017

**Name and mailing address of the ISA**

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

**Authorized officer**

Novak-Giese, Sabine
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>ANN RAN F ET AL: &quot;Genome engineering using the CRISPR-Cas9 system&quot;, NATURE PROTOCOLS, NATURE PUBLISHING GROUP, GB, vol. 8, no. 11, 1 November 2013 (2013-11-01), pages 2281-2308, XP002744123, ISSN: 1754-2189, DOI: 10.1038/NPROT.2013.143 [retrieved on 2013-10-24], figures 1, 4</td>
<td>1-13</td>
</tr>
<tr>
<td>Y</td>
<td>ANTOINE MARUANI ET AL: &quot;A plug-and-play approach to anti body-based therapeutics via a chemoselective dual click strategy&quot;, NATURE COMMUNICATIONS, vol. 6, 31 March 2015 (2015-03-31), page 6645, XP055272086, United Kingdom, ISSN: 2041-1723, DOI: 10.1038/ncomms7645, page 3, column 1, paragraph first</td>
<td>1-13</td>
</tr>
<tr>
<td>Y</td>
<td>Wo 2015/166272 A2 (INTAS LTD [GB]) 5 November 2015 (2015-11-05), claims 1, 22, 32; figure 21</td>
<td>1-13</td>
</tr>
<tr>
<td>Y</td>
<td>Wo 2013/079953 A1 (KYMAB LTD [GB]) 6 June 2013 (2013-06-06), claims 43, 46, 63</td>
<td>1-13</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>----------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>CA</td>
<td>18-06-2015</td>
<td></td>
</tr>
<tr>
<td>CN 105980568 A</td>
<td>28-09-2016</td>
<td></td>
</tr>
<tr>
<td>JP 2016198110 A</td>
<td>01-12-2016</td>
<td></td>
</tr>
<tr>
<td>JP 2016539655 A</td>
<td>22-12-2016</td>
<td></td>
</tr>
<tr>
<td>KR 20160095150 A</td>
<td>10-08-2016</td>
<td></td>
</tr>
<tr>
<td>US 2015159174 Al</td>
<td>11-06-2015</td>
<td></td>
</tr>
<tr>
<td>US 2015159175 Al</td>
<td>11-06-2015</td>
<td></td>
</tr>
<tr>
<td>US 2016060657 Al</td>
<td>03-03-2016</td>
<td></td>
</tr>
<tr>
<td>US 2017067078 Al</td>
<td>09-03-2017</td>
<td></td>
</tr>
<tr>
<td>W0 2015088643 Al</td>
<td>18-06-2015</td>
<td></td>
</tr>
<tr>
<td>WO 2015166272 A2</td>
<td>05-11-2015</td>
<td></td>
</tr>
<tr>
<td>AU 2015255013 Al</td>
<td>17-11-2016</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>05-11-2015</td>
<td></td>
</tr>
<tr>
<td>CN 106574264 A</td>
<td>19-04-2017</td>
<td></td>
</tr>
<tr>
<td>EP 3137602 A2</td>
<td>08-03-2017</td>
<td></td>
</tr>
<tr>
<td>SG 11201609123R A</td>
<td>29-11-2016</td>
<td></td>
</tr>
<tr>
<td>US 2017073664 Al</td>
<td>16-03-2017</td>
<td></td>
</tr>
<tr>
<td>W0 2015166272 A2</td>
<td>05-11-2015</td>
<td></td>
</tr>
<tr>
<td>WO 2013079953 A</td>
<td>06-06-2013</td>
<td></td>
</tr>
<tr>
<td>AU 2012343587 Al</td>
<td>15-05-2014</td>
<td></td>
</tr>
<tr>
<td>AU 2016101604 A4</td>
<td>06-10-2016</td>
<td></td>
</tr>
<tr>
<td>AU 2016244295 Al</td>
<td>03-11-2016</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td>06-06-2013</td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>104160031 A</td>
<td>19-11-2014</td>
</tr>
<tr>
<td>DE 202012013369 UI</td>
<td>23-08-2016</td>
<td></td>
</tr>
<tr>
<td>EP 2649184 Al</td>
<td>16-10-2013</td>
<td></td>
</tr>
<tr>
<td>EP</td>
<td>02-03-2016</td>
<td></td>
</tr>
<tr>
<td>HK 1185100 Al</td>
<td>26-08-2016</td>
<td></td>
</tr>
<tr>
<td>JP</td>
<td>22-01-2015</td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>19-09-2013</td>
<td></td>
</tr>
<tr>
<td>US 2016353719 Al</td>
<td>08-12-2016</td>
<td></td>
</tr>
<tr>
<td>W0</td>
<td>06-06-2013</td>
<td></td>
</tr>
</tbody>
</table>