TRANSPOSITION-MEDIATED IDENTIFICATION OF SPECIFIC BINDING OR FUNCTIONAL PROTEINS

Inventor: Ulf Grawunder

BACKGROUND OF THE INVENTION

Field of the Invention

(a) Technologies for the identification of specific functional and binding proteins

[0001] The discovery of target-specific proteins, including antibodies and fragments thereof, is of significant commercial interest, because the selection of highly selective functional proteins or binding proteins, including antibodies and fragments thereof, has a high potential for the development of new biological entities (NBEs) with novel therapeutic properties that very specifically integrate, or interfere with biological processes, and therefore are predicted to display lower side-effect profiles than conventional new chemical entities (NCEs). In that respect, particularly the development of highly target-specific, therapeutic antibodies, and antibody-based therapeutics, have paved the way to completely novel therapies with improved efficacy. As a consequence, therapeutic monoclonal antibodies represent the fastest growing segment in the development of new drugs over the last decade, and presently generate about USD 50 billion global revenues, which accounts for a significant share of the total global market of pharmaceutical drugs.

[0002] Therefore, efficient and innovative technologies, that allow the discovery of highly potent, but also well tolerated therapeutic proteins, in particular antibody-based therapeutics, are in high demand.

[0003] In order to identify a protein with a desired functionality or a specific binding property, as is the case for antibodies, it is required to generate, to functionally express and to screen large, diverse collections, or libraries of proteins, including antibodies and fragments thereof, for desired functional properties or target binding specificity. A number of technologies have been developed over the past twenty years, which allow expression of diverse protein libraries either in host cells, or on viral and phage particles and methods for their high-throughput screening and/or panning toward a desired functional property, or binding phenotype.

[0004] Standard, state-of-the-art technologies to achieve identification of target-specific binders or proteins with desired functional properties include, e.g. phage-display, retroviral display, bacterial display, yeast display and various mammalian cell display technologies, in combination with solid surface binding (panning) and/or other enrichment techniques. All of these technologies are covered by various patents and pending patent applications.

[0005] While phage and prokaryotic display systems have been established and are widely adopted in the biotech industry and in academia for the identification of target-specific binders, including antibody fragments (Hoogenboom, *Nature Biotechnology 23*, 1105-1116 (2005)), they suffer from a variety of limitations, including the inability to express full-length versions of larger proteins, including full-length antibodies, the lack of proper post-translational modification, the lack of proper folding by vertebrate chaperones, and, in the case of antibodies, an artificially enforced heavy and light chain combination. Therefore, in case of antibody discovery by these methods, "reformatting" into full-length antibodies and mammalian cell expression is required. Due to the above-mentioned limitations this frequently results in antibodies with unfavorable biophysical properties (e.g. low stability, tendency to aggregate, diminished affinity), limiting the therapeutic and diagnostic potential of such proteins. This, on one hand, leads to significant attrition rates in the development of lead molecules generated by these methods, and, on the other hand, requires significant effort to correct the biophysical and molecular liabilities in these proteins for further downstream drug development.

[0006] Therefore, protein and antibody discovery technologies have been developed using lower eukaryotic (e.g. yeast) and, more recently, also mammalian cell expression systems for the identification of proteins with desired properties, as these technologies allow (i) expression of larger, full-length proteins, including full-length antibodies, (ii) better or normal post-translational modification, and, (iii) in case of antibodies, proper heavy-light chain pairing (Beerli & Rader, *mAbs 2*, 365-378 (2010)). This, in aggregate, selects for proteins with favorable biophysical properties that have a higher potential in drug development and therapeutic use.

[0007] Although expression and screening of proteins in vertebrate cells would be most desirable, because vertebrate cells (e.g. hamster CHO, human HEK-293, or chicken DT40 cells) are preferred expression systems for the production of larger therapeutic proteins, such as antibodies, these technologies are currently also associated with a number of limitations, which has lead to a slow adoption of these technologies in academia and industry.

[0008] First, vertebrate cells are not as efficiently and stably genetically modified, as, e.g. prokaryotic or lower eukaryotic cells like yeast. Therefore, its remains a challenge to generate diverse (complex) enough vertebrate cell based proteins libraries, from which candidates with desired properties or highest binding affinities can be identified. Second, in order to efficiently isolate proteins with desired properties, usually iterative rounds of cell enrichment are required. Vertebrate expression either by transient transfection of plasmids (Higuchi *et al J. Immunol. Methods 202*, 193-204(1997)), or transient viral expression systems, like sindbis or vaccinia virus (Beerli *et al. PNAS 105*, 14336-14341 (2008), and WO02102885) do not allow multiple

rounds of cell selection required to efficiently enrich highly specific proteins, and these methods are therefore either restricted to screening of small, pre-enriched libraries of proteins, or they do require tedious virus isolation/cell re-infection cycles.

[0009] In order to achieve stable expression of binding proteins and antibodies in vertebrate cells, that do allow multiple rounds of selections based on stable genotype-phenotype coupling, technologies have been developed, utilizing specific recombinases (flp/frt recombinase system, Zhou *et al. mAbs 5*, 508-518 (2010)), or retroviral vectors (WO2009109368). However, the flp/frt recombination is a low-efficient system for stable integration of genes into vertebrate host cell genomes and therefore, again, only applicable to small, pre-selected libraries, or the optimization of selected protein or antibody candidates.

[0010] In comparison to the flp/frt recombinase system, retroviral vectors allow more efficient stable genetic modification of vertebrate host cells and the generation of more complex cellular libraries. However, (i) they are restricted to only selected permissible cell lines, (ii) they represent a biosafety risk, when human cells are utilized, (iii) retroviral expression vectors are subject to unwanted mutagenesis of the library sequences due to low-fidelity reverse transcription, (iv) retroviral vectors do not allow integration of genomic expression cassettes with intact intron/exon structure, due to splicing of the retroviral genome prior to packaging of the vector into retroviral particles, (v) retroviruses are subject to uncontrollable and unfavorable homologous recombination of library sequences during packaging of the viral genomes, (vi) are subject to retroviral silencing, and (vii) require a tedious two-step packaging-cell transfection / host-cell infection procedure. All these limitations represent significant challenges and linitations, and introduce significant complexities for the utility of retroviral vector based approaches in generating high-quality/high complexity vertebrate cell libraries for efficient target-specific protein, or antibody discovery.

[0011] Therefore, clearly a need exists for a more efficient, more controllable and straightforward technology that allows the generation of high-quality and highly complex vertebrate cell based libraries expressing diverse libraries of proteins including antibodies and fragments thereof from which proteins with highly specific function and/or binding properties and high affinities can be isolated.

(b) Transposases/Transposition:

[0012] Transposons, or transposable elements (TEs), are genetic elements with the capability to stably integrate into host cell genomes, a process that is called transposition (Ivics *et al. Mobile DNA* 1, 25 (2010)) (incorporated herein by reference in its entirety). TEs were already postulated in the 1950s by Barbara McClintock in genetic studies with maize, but the first functional models for transposition have been described for bacterial TEs at the end of the

1970s (Shapiro, *PNAS 76*, 1933-1937 (1979)) (incorporated herein by reference in its entirety).

[0013] Meanwhile it is clear that TEs are present in the genome of every organism, and genomic sequencing has revealed that approximately 45% of the human genome is transposon derived (International Human Genome Sequencing Consortium *Nature 409*: 860-921 (2001)) (incorporated herein by reference in its entirety). However, as opposed to invertebrates, where functional (or autonomous) TEs have been identified (Fig. 1a), humans and most higher vertebrates do not contain functional TEs. It has been hypothesized that evolutionary selective pressure against the mutagenic potential of TEs lead to their functional inactivation millions of years ago during evolution.

[0014] Autonomous TEs comprise DNA that encodes a transposase enzyme located in between two inverted terminal repeat sequences (ITRs), which are recognized by the transposase enzyme encoded in between the ITRs and which can catalyze the transposition of the TE into any double stranded DNA sequence (FIG. 1a). There are two different classes of transposons: class I, or retrotransposons, that mobilize via an RNA intermediate and a "copy-and-paste" mechanism (FIG. 2b), and class II, or DNA transposons, that mobilize via excision-integration, or a "cut-and-paste" mechanism (FIG. 2a) (Ivics *et al. Nat. Methods 6*, 415-422(2009)) (incorporated herein by reference in its entirety).

[0015] Bacterial, lower eukaryotic (e.g. yeast) and invertebrate transposons appear to be largely species specific, and cannot be used for efficient transposition of DNA in vertebrate cells. Only, after a first active transposon had been artificially reconstructed by sequence shuffling of inactive TEs from fish, which was therefore called "*Sleeping Beauty*" (Ivics *et al. Cell 91*, 501-510 (1997)) (incorporated herein by reference in its entirety), did it become possible to successfully achieve DNA integration by transposition into vertebrate cells, including human cells. *Sleeping Beauty* is a class II DNA transposon belonging to the Tc1/mariner family of transposons (Ni *et al. Briefings Funct. Genomics Proteomics 7*, 444-453 (2008)) (incorporated herein by reference in its entirety). In the meantime, additional functional transposons have been identified or reconstructed from different species, including Drosophila, frog and even human genomes, that all have been shown to allow DNA transposition into vertebrate and also human host cell genomes (FIG. 3). Each of these transposons, have advantages and disadvantages that are related to transposition efficiency, stability of expression, genetic payload capacity, etc.

[0016] To date, transposon-mediated technologies for the expression of diverse libraries of proteins, including antibodies and fragments thereof, in vertebrate host cells for the isolation of target

specific, functional binding proteins, including antibodies and fragments thereof, have not been disclosed in the prior art.

BRIEF SUMMARY OF THE INVENTION

[0017] The method disclosed herein describes a novel technology offering unparalleled efficiency, flexibility, utility and speed for the discovery and optimization of polypeptides having a desired binding specificity and/or functionality, including antigen-binding molecules such as antibodies and fragments thereof, for desired functional and/or binding phenotypes. The novel method is based on transposable constructs and diverse DNA libraries cloned into transposable vectors and their transfection into host cells by concomitant transient expression of a functional transposase enzyme. This ensures an efficient, stable introduction of the transposon-based expression vectors into vertebrate host cells in one step, which can then be screened for a desired functional or binding phenotype of the expressed proteins, after which the relevant coding sequences for the expressed proteins, including antibodies and fragments thereof, can be identified by standard cloning and DNA sequencing techniques.

[0018] In one embodiment, the invention is broadly directed to a method for identifying a polypeptide having a desired binding specificity or functionality, comprising:

[0019] (i) generating a diverse collection of polynucleotides encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence coding for a polypeptide disposed between first and second inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme;

[0020] (ii) introducing the diverse collection of polynucleotides of (i) into host cells;

[0021] (iii) expressing at least one transposase enzyme functional with said inverted terminal repeat sequences in said host cells so that said diverse collection of polynucleotides is integrated into the host cell genome to provide a host cell population that expresses said diverse collection of polynucleotides encoding polypeptides having different binding specificities or functionalities;

[0022] (iv) screening said host cells to identify a host cell expressing a polypeptide having a desired binding specificity or functionality; and

[0023] (v) isolating the polynucleotide sequence encoding said polypeptide from said host cell.

In a preferred embodiment, the polynucleotides are DNA molecules. In one embodiment, the diverse collection of polynucleotides comprises a ligand-binding sequence of a receptor, or a target binding sequence of a binding molecule. In a preferred embodiment, the polynucleotides comprise a sequence encoding an antigen-binding molecule, such as an antibody VH or VL domain, or an antigen-binding fragment thereof, or antibody heavy or

light chains that are full-length (i.e., which include the constant region). In certain embodiments, the polynucleotides may comprise a sequence encoding both a V_H and V_L region, or both antibody heavy and light chains. In another embodiment, the polynucleotides comprise a sequence encoding a single-chain Fv or a Fab domain.

In one embodiment, the diverse collection of polynucleotides is generated by subjecting V region gene sequences to PCR under mutagenizing conditions, for example, by PCR amplification of V region repertoires from vertebrate B cells. In another embodiment, the diverse collection of polynucleotides is generated by gene synthesis (e.g., by randomization of sequences encoding a polypeptide having known binding specificity and/or functionality). In one useful embodiment, the diverse collection of polynucleotides comprises plasmid vectors. In another useful embodiment, the diverse collection of polynucleotides comprises double-stranded DNA PCR amplicons. The plasmid vectors may comprise a sequence encoding a marker gene, such as a fluorescent marker, a cell surface marker, or a selectable marker. The marker gene sequence may be upstream or downstream of the sequence encoding the polypeptide having a binding specificity or functionality, but between the inverted terminal repeat sequences. Alternatively, the marker gene sequence may be downstream of said sequence encoding a polypeptide having binding specificity or functionality and separated by an internal ribosomal entry site.

[0024] In some embodiments, the diverse collection of polynucleotides encode a plurality of antigen-binding molecules of a vertebrate, such as a mammal, e.g., a human.

[0025] In one embodiment, step (ii) of the method comprises introducing into host cells polynucleotides comprising sequences encoding immunoglobulin V_H or V_L regions, or antigen-binding fragments thereof, and wherein said V_H and V_L region sequences are encoded on separate vectors. In another embodiment, step (ii) of the method of the invention comprises introducing into host cells polynucleotides comprising sequences encoding full-length immunoglobulin heavy or light chains, or antigen-binding fragments thereof, wherein said full-length heavy and light chain sequences are on separate vectors. The vectors may be introduced into the host cells simultaneously or sequentially. In another embodiment, sequences encoding V_H and V_L regions or full-length heavy and light chains are introduced into host cells on the same vector. In the event that the V_H and V_L sequences or the full-length antibody heavy and light chain sequences are introduced into the host cells on different vectors, it is useful for the inverted terminal repeat sequences on each vector to be recognized by and functional with different transposase enzymes.

[0026] The host cells are preferably vertebrate cells, and preferably mammalian cells, such as rodent or human cells. Lymphoid cells, e.g, B cells, are particularly useful. B cells may be

progenitor B cells or precursor B cells such as, for example, Abelson-Murine Leukemia virus transformed progenitor B cells or precursor B cells and early, immunoglobulin-null EBV transformed human proB and preB cells. Other useful host cells include B cell lines such as Sp2/0 cells, NS0 cells, X63 cells, and Ag8653 cells, or common mammalian cell lines such as CHO cells, Per.C6 cells, BHK cells, and 293 cells.

[0027] In one embodiment of the method of the invention, the expressing step (iii) comprises introducing into said host cells an expression vector encoding a transposase enzyme that recognizes and is functional with at least one inverted terminal repeat sequence in the polynucleotides. The vector encoding the transposase enzyme may be introduced into the host cells concurrently with or prior or subsequent to the diverse collection of polynucleotides. In one embodiment, the transposase enzyme is transiently expressed in said host cell. Alternatively, the expressing step (iii) may comprise inducing an inducible expression system that is stably integrated into the host cell genome, such as, for example, a tetracycline-inducible or tamoxifen-inducible system. In a preferred embodiment, step (iii) comprises expressing in the host cell(s) a vector comprising a functional *Sleeping Beauty* transposase or a functional *PiggyBac* transposase. In one useful embodiment, step (iii) comprises expressing in said host cell a vector comprising SEQ ID NO:11. In another useful embodiment, the vector encodes SEQ ID NO:12, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

[0028] In another useful embodiment, step (iii) comprises expressing in said host cell a vector comprising SEQ ID NO:17. In another useful embodiment, the vector encodes SEQ ID NO:18, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

[0029] In one embodiment of the method of the invention, the screening step (iv) comprises magnetic activated cell sorting (MACS), fluorescence activated cell sorting (FACS), panning against molecules immobilized on a solid surface panning, selection for binding to cell-membrane associated molecules incorporated into a cellular, natural or artificially reconstituted lipid bilayer membrane, or high-throughput screening of individual cell clones in multi-well format for a desired functional or binding phenotype. In one embodiment, the screening step (iv) comprises screening to identify polypeptides having a desired target-binding specificity or functionality. In a preferred embodiment, the screening step (iv) comprises screening to identify antigen-binding molecules having a desired antigen specificity. In one useful embodiment, the screening step further comprises screening to identify antigen-binding molecules having one or more desired functional properties. The screening step (iv) may

comprise multiple cell enrichment cycles with host cell expansion between individual cell enrichment cycles.

[0030] In one embodiment of the method of the invention, the step (v) of isolating the polynucleotide sequence encoding the polypeptide having a desired binding specificity or functionality comprises genomic or RT-PCR amplification or next-generation deep sequencing. In one useful embodiment, the polynucleotide sequence isolated in step (v) is subjected to affinity optimization. This can be done by subjecting the isolated polynucleotide sequence to PCR or RT-PCR under mutagenizing conditions. In another useful embodiment, the mutagenized sequence is then further subjected to steps (i)-(v) of the method of the invention. In a preferred embodiment, the polynucleotide sequence obtained in (v) comprises a sequence encoding a V_H or V_L region of an antibody, or an antigen-binding fragment thereof, and wherein said antibody optimization comprises introducing one or more mutations into a complementarity determining region or framework region of said V_H or V_L.

[0031] In one useful embodiment, the inverted terminal repeat sequences are from the *PiggyBac* transposon system and are recognized by and functional with the *PiggyBac* transposase. In one embodiment, the sequence encoding the upstream *PiggyBac* inverted terminal repeat sequence comprises SEQ ID NO:1. In another embodiment, the sequence encoding the downstream *PiggyBac* inverted terminal repeat sequence comprises SEQ ID NO:2.

[0032] In another useful embodiment, the inverted terminal repeat sequences are from the *Sleeping Beauty* transposon system and are recognized by and functional with the *Sleeping Beauty* transposase. In one embodiment, the sequence encoding the upstream *Sleeping Beauty* inverted terminal repeat sequence comprises SEQ ID NO:14. In another embodiment, the sequence encoding the downstream *Sleeping Beauty* inverted terminal repeat sequence comprises SEQ ID NO:15.

[0033] In one embodiment of the invention, the polynucleotides comprise V_H or V_L region sequences encoding a sequence derived from a human anti-TNF alpha antibody. In one embodiment, the human anti-TNF alpha antibody is D2E7.

[0034] In a useful embodiment, step (iii) comprises introducing into said host cell a vector comprising a sequence encoding a functional *PiggyBac* transposase. In one embodiment the vector comprises SEQ ID NO:11. In another embodiment, the vector encodes SEQ ID NO:12, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

[0035] In another useful embodiment, step (iii) comprises expressing in said host cell a vector comprising SEQ ID NO:17. In another useful embodiment, the vector encodes SEQ ID

NO:18, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

[0036] In preferred embodiments, the inverted terminal repeat sequences are recognized by and functional with at least one transposase selected from the group consisting of: *PiggyBac*, *Sleeping Beauty*, *Frog Prince*, *Himar1*, *Passport*, *Minos*, *hAT*, *Tol1*, *Tol2*, *Ac/Ds*, *PIF*, *Harbinger*, *Harbinger3-DR*, and *Hsmar1*.

[0037] The present invention is further directed to a library of polynucleotide molecules encoding polypeptides having different binding specificities or functionalities, comprising a plurality of polynucleotide molecules, wherein said polynucleotide molecules comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme. Preferably the polynucleotides are DNA.molecules and comprise a ligand-binding sequence of a receptor or a target-binding sequence of a binding molecule. In a particularly preferred embodiment, the library comprises polynucleotides, wherein each polynucleotide comprises a sequence encoding an antigen-binding sequence of an antibody. In one embodiment, the library comprises polynucleotides encoding a V_H or V_L region of an antibody or an antigen-binding fragment thereof. Alternatively, the polynucleotides may encode a V_H region and a V_L region. In a preferred embodiment, the polynucleotides of the library comprise a sequence encoding a full-length antibody heavy or light chain (i.e., including the constant region) or an antigen-binding fragment thereof. Alternatively, the polynucleotides may encode both a full-length immunoglobulin heavy and light chain. In other embodiments, the polynucleotides of the library comprise a sequence encoding a single-chain Fv or a Fab domain. In preferred embodiments, the polynucleotides of the library are in the form of plasmids or double stranded DNA PCR amplicons. In certain embodiments, the plasmids of the library comprise a marker gene. In another embodiment, the plasmids comprise a sequence encoding a transposase enzyme that recognizes and is functional with the inverted terminal repeat sequences. In one embodiment, the library of the invention comprises polynucleotides that encode the full-length immunoglobulin heavy chain including the natural intron/exon structure of an antibody heavy chain. The full-length immunoglobulin heavy chain may comprise the endogenous membrane anchor domain.

[0038] The present invention is also directed to a method for generating a library of transposable polynucleotides encoding polypeptides having different binding specificities or functionality, comprising (i) generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence encoding polypeptide having a binding specificity or

functionality disposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme.

[0039] The present invention is also directed to a vector comprising a sequence encoding a V_H or V_L region of an antibody, or antigen-binding portion thereof, disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme. In certain embodiments, the vector encodes a full-length heavy or light chain of an immunoglobulin. Preferably, the sequence encoding the V_H or V_L or the heavy or light chain is a randomized sequence generated by, for example, PCR amplification under mutagenizing conditions or gene synthesis. In one embodiment, the vector comprises inverted terminal repeat sequences that are recognized by and functional with the *PiggyBac* transposase. In another embodiment, the inverted terminal repeat sequences are recognized by and functional with the *Sleeping Beauty* transposase. In one embodiment, the vector comprises a VH or VL region sequence derived from an anti-TNF alpha antibody such as, for example, D2E7. In certain embodiments, the vector comprises at least one sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:19, SEQ ID NO: 20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, and SEQ ID NO:41.

[0040] The present invention is also directed to a host cell comprising a vector of the invention as described above. In a preferred embodiment, the host cell further comprises an expression vector comprising a sequence encoding a transposase that recognizes and is functional with at least one inverted terminal repeat sequence in the vector encoding said V_H or V_L region sequence.

[0041] The present invention is still further directed to antigen-binding molecules, e.g., antibodies, produced by a method comprising claim 1.

[0042] The present invention is also directed to a method for generating a population of host cells capable of expressing polypeptides having different binding specificities or functionalities, comprising:

[0043] (i) generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme; and

[0044] (ii) introducing said diverse collection of polynucleotides into host cells.

BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0045] FIG. 1: a.) This drawing depicts the configuration of an autonomous transposable element (TE), which can transpose or "jump" into any target DNA sequence. The key components of a TE are an active transposase enzyme that recognizes the inverted terminal repeats (ITRs) flanking the transposase enzyme itself up- and downstream of its sequence. TEs catalyze either the copying or the excision of the TE, and the integration in unrelated target DNA sequences. b.) This drawing depicts the configuration of a transposon vector system, in which the expression of an active transposase enzyme is effected by an expression vector that is not coupled to the TE itself. Instead, the TE may contain any sequence(s), or gene(s) of interest that is/are cloned in between the up- and downstream ITRs. Integration of the TE containing any sequence(s), or gene(s) of interest (e.g. a DNA library encoding a library of proteins) may integrate into unrelated target DNA sequences, if the transposase enzyme expression is provided in *trans*, e.g. by a separate transposase expression construct, as depicted here.

[0046] FIG. 2: a) This drawing depicts the two different ways how TEs can "jump" or transpose into unrelated target DNA. For group II transposons, the transposase enzyme in a first step recognizes the ITRs of the transposable element and catalyzes the excision of the TE from DNA. In a second step, the excised TE is inserted into unrelated target DNA sequence, which is also catalyzed by the transposase enzyme. This results in a "cut-and-paste" mechanism of transposition. For group I transposons (shown in b.) the coding information of the TE is first replicated (e.g. transcribed and reverse transcribed, in the case of retrotransposons) and the replicated TE then integrates into unrelated target DNA sequence, which is catalyzed by the transposase enzyme. This results in a "copy-and-paste" mechanism of transposition.

[0047] FIG. 3: This figure provides an overview of active transposase enzymes that have been identified and/or reconstructed from dormant, inactive TEs, and that have been shown to be able to confer transposition in various vertebrate and also human cells, as provided in the table. The table has been adapted from Table I of publication Ni *et al. Briefings Functional Genomics Proteomics* 7, 444-453 (2008) (incorporated herein by reference in its entirety)

.

[0048] FIG. 4: This figure outlines the principle of the method disclosed herein, for the isolation of coding information for proteins, including antibodies and fragments thereof, with a desired function, e.g. the binding to a target of interest, as depicted here. The gene(s) of interest, e.g. a diverse transposable DNA library encoding proteins, including antibody polypeptide chains or fragments thereof, that is cloned in between inverted terminal repeats (ITRs) of a transposable

construct is introduced into a vertebrate host cell together with an expression vector for an active transposase enzyme (see top of the drawing). The expression of the transposon enzyme in said host cells *in trans* and the presence of the gene(s) of interest cloned in between ITRs that can be recognized by the transposase enzyme allows the stable integration of the ITR-flanked gene(s) of interest into the genome of the host cells, which can then stably express the protein(s) of interest encoded by the genes of interest. The cellular library expressing the protein(s) of interest can then be screened for a desired functionality of the expressed proteins, e.g., but not limited to the binding to a target protein of interest, as depicted here. By means of cell separation technqiues known in the art, e.g. MACS or FACS, the cells expressing the protein(s) of interest with the desired phenotype and which therefore contain the corresponding genotype, can be isolated and the coding information for the gene(s) of interest can be retrieve from the isolated cells by cloning techniques known in the art, e.g. but not limited to genomic PCR cloning, as depicted here.

[0049] FIGs. 5a) and 5b): This drawing outlines the cloning strategy for the generation of a transposable human immunoglobulin (Ig) kappa light chain (LC) expression vector, as described in Example 1. FIG. 5 a.) depicts the cloning strategy for the insertion of 5'- and 3'-ITRs from the *PiggyBac* transposon into the mammalian expression vector pIRES-EGFP (Invitrogen, Carlsbad, CA, USA), which already contains the strong mammalian cell promoter element pCMV(IE) (immediate early promoter of CMV), and intron/polyA signals for strong mammalian host cell expression. In addition, downstream of the ClaI, EcoRV, NotI, EcoRI containing multiple cloning site, into which gene(s) of interest can be cloned, pIRES-EGFP contains an internal ribosomal entry site (IRES) with a downstream ORF of enhanced green fluorescent protein (EGFP), which effects the coupling of expression of gene(s) of interest cloned upstream of the IRES. Bacterial functional elements (ampicillin resistance gene, amp^R) and a bacterial origin of replication (Col E1) for amplification and selection of the plasmid in E. coli are depicted as well. The resulting *PiggyBac* ITRs containing plasmid is designated pIRES-EGFP-T1T2. FIG 5b) then depicts the insertion of a gene synthesized human Ig kappa LC into the unique EcoRV restriction enzyme site of pIRES-EGFP-T1-T2, which positions the human Ig kappa LC upstream of the IRES-EGFP cassette, and thereby couples the expression of the human Ig kappa LC to EGFP marker gene expression. The insertion of the human Ig kappa LC results in transposable human Ig kappa LC expression vector pIRES-EGFP-T1T2-IgL. The drawings show selected unique restriction enzyme sites in the plasmids, as well as selected duplicated sites resulting from cloning steps.

[0050] FIG. 6: This drawing outlines the cloning of a transposable human immunoglobulin (Ig) gamma 1 heavy chain (HC) expression vector, which can be generated by exchange of the human Ig kappa LC open reading frame (ORF) against the ORF for a human Ig gamma 1 HC ORF. The design of the final Ig gamma 1 HC ORF is similar, also with regard to the engineering of a unique Eco47III restriction enzyme site separating the variable (V) from the constant (C) coding regions, which allows the exchange of a single antibody V coding region against a diverse library of antibody V coding regions, as described in Example 3.

[0051] Fig. 7: This drawing depicts the cloning of a mammalian PiggyBac transposase enzyme expression vector, as described in the Example 4, using pCDNA3.1(+) hygro as the backbone of the mammalian expression vector, into which the gene synthesized ORF from PiggyBac transposase is cloned into the unique EcoRV restriction enzyme site of pCDNA3.1(+) hygro, resulting in PiggyBac transposon expressin vector pCDNA3.1(+) hygro-PB.expression vector pCDNA3.1(+) hygro-PB. Also in this drawing the relative position of other mammalian functional elements (CMV-IE promoter, BGH-polyA signal, SV40-polyA segment, hygromycinB ORF) and bacterial functional elements (ampicillin resistance gene, ampR, origin of replication, ColE1), as well as selected relevant restriction enzyme recognition sites are shown.

[0052] Fig. 8: This drawing depicts the cloning of a Sleeping Beauty transposable human immunoglobulin kappa light chain (Ig-kappa LC) expression vector, as described in Example 5. The cloning can be performed by sequentially replacing the PiggyBac 5' and 3' ITRs with Sleeping Beauty 5' and 3'ITRs in construct pIRES-EGFP-T1T2-IgL. Also in this drawing the relative position of other mammalian functional elements (CMV-IE promoter, BGH-polyA signal, SV40-polyA segment, hygromycinB ORF) and bacterial functional elements (ampicillin resistance gene, ampR, origin of replication, ColE1), as well as selected relevant restriction enzyme recognition sites are shown.

[0053] Fig. 9: This drawing depicts the cloning of a mammalian Sleeping Beauty transposase enzyme expression vector, as described in the Example 6, using pCDNA3.1(+) hygro as the backbone of the mammalian expression vector, into which the gene synthesized ORF from Sleeping Beauty transposase is cloned into the unique EcoRV restriction enzyme site of pCDNA3.1(+) hygro, resulting in Sleeping Beauty transposon expression vector pCDNA3.1(+) hygro-SB. Also in this drawing the relative position of other mammalian functional elements (CMV-IE promoter, BGH-polyA signal, SV40-polyA segment, hygromycinB ORF) and bacterial

functional elements (ampicillin resistance gene, ampR, origin of replication, ColE1), as well as selected relevant restriction enzyme recognition sites are shown.

[0054] Fig. 10: This drawing shows the arrangement of functional elements and position of selected unique restriction enzyme sites within the gene-synthesized DNA fragments 1.) and 2.) that were utilized in Example 4, in order to clone both "empty" IgH chain expression vectors allowing transposition utilizing either *PiggyBac* or *Sleeping Beauty* transposase. The origin of the functional elements is disclosed in detail in the description of the Example.

[0055] Fig. 11: This drawing shows the final design and plasmid map of the transposable expression vectors for human, membrane bound Ig-gamma1 heavy chains (left) and human Ig kappa light chains (right). For the IgH expression vector, the V_H-coding region may be replaced by V_H coding regions of any other monoclonal antibody, or by a V_H-gene library, using unique restriction enzyme sites NotI and NheI, flanking the V_H coding region in this vector. For the IgL expression vector, the V_L-coding region may be replaced by V_L coding regions of any other monoclonal antibody, or by a V_L-gene library, using unique restriction enzyme sites NotI and BsiWI, flanking the V_L coding region in this vector. The 8 vector constructions for *PiggyBac* and Sleeping Beauty transposable IgH and IgL vectors, disclosed in detail in Example 4 all share this general design. The two vector maps displayed here correspond to the vector maps of pPB-EGFP-HC-Ac10 (left) and pPB-EGFP-LC-Ac10 (right), and the additional vectors for hBU12 heavy chain (HC) or light chain (LC), containing either *PiggyBac* or *Sleeping Beauty* ITRs, are provided in the tables below. Sequences of all vectors in this figure are provided in Example 4.

[0056] Fig. 12: This figure shows two dimensional FACS dot-plots, in which the surface expression of human IgG from transfected and transposed IgHC and IgLC expression vectors is detected on the surface of 63-12 A-MuLV transformed murine proB cells derived from RAG-2-deficient mice. d2 post TF means that the FACS analysis was performed 2 days after transfection of vector constructs into 63-12 cells. The FACS plots in the left-hand column represents negative and positive controls for the transfection. NC=mock electroporation of cells without plasmid DNA. pEGFP-N3=transfection control with pEGFP-N3 control vector, which controls for the transfection efficiency by rendering transfected cells green. The second column from the left shows FACS plots from 63-12 cells co-transfected with either *PiggyBac*-transposase vector, pPB-EGFP-HC-Ac10, pPB-EGFP-LC-Ac10 vectors (top row), or *PiggyBac*-transposase vector, pPB-EGFP-HC-hBU12, pPB-EGFP-LC-hBU12 vectors (middle row), or with *Sleeping Beauty*-transposase vector, pSB-EGFP-HC-Ac10, pSB-EGFP-LC-Ac10 vectors (bottom row). The second-left column labeled "d2 post TF" shows the analysis

for cells expressing IgG on the cell surface (Y-axis) and EGFP expression (X-axis) two days post co-transfection of the vectors as mentioned above. Surface IgG and EGFP double positive cells were FACS sorted as indicated by the rectangular gate. The second-right column labeled "d9 1x sorted" shows the analysis of surface IgG and EGFP expression in the cell population that was sorted at day 2 after transfection, analyzed in the same way. Sorting gates for the second FACS sort are also provided as rectangular gates. The rightmost column labeled "d16 2x sorted" shows the analysis of surface IgG and EGFP expression of the cell populations that had been re-sorted at day 9 after transfection, and analyzed in the same way for surface IgG and EGFP expression as in the previous experiments.

[0057] Fig. 13: This figure depicts the demonstration that proB cells expressing CD30-specific IgG on the surface of 63-12 cells can specifically be stained and detected by CD30 antigen, and that the CD30-specific cells be detected and re-isolated from a large population of cells expressing surface IgG of unrelated specificity (here CD19), in which the CD30-specific cells have been spiked in with decreasing frequency. The FACS dot-plot on top shows the detection of IgG (via anti-kappaLC staining) and CD30 binding (via CD30-antigen staining) on the surface of the positive control cells, which are 63-12 cells stably transposed and 2x sorted for expressing human anti-CD30 IgG, clone Ac10 on the cell surface. As expected, a quite pure population (97.3%) of IgG-positive/CD30-reactive was detectable in the upper right quadrant of the FACs-dot-plot. The numbers on top of each FACS-plot indicates the number of live cells based on FSC/SSC gating that were acquired in each experiment. The middle row shows the FACS analysis for IgG-positive/CD30 reactive cells detectable in a background of IgGpositive/CD19 specific cells. The number above the number of events indicates the dilution factor of anti-CD30 specific IgG positive cells that were used for the generation of the "spiked-in" population of anti-CD30 mAb IgG positive cells in a background of anti-CD19 mAb IgG positive cells. The sorting gates are indicated that were used to specifically isolate IgG-positive/CD30 antigen reactive cells from the spiked-in populations. Larger numbers of events needed to be acquired in order to allow detection and isolation of the IgG-positive/CD30 cells at higher dilutions. The lower row of FACS plots then shows the re-analysis of sorted cells after the cells had been expanded for 12 days for the same parameters (IgG-expression & CD30 antigen specificity).

[0058] Fig. 14: This figure shows the cloning of a transposable vector for a human Ig-gamma1 heavy chain (HC) in genomic configuration. The linear fragment on top represents the human gamma1 exon and introns for membrane-bound Ig-gamma1-HC, with flanking NheI and BstBI restriction sites added to allow ligation into Ig-gamma1 HC cDNA vector pPB-EGFP-

HC-Ac10. H-designates the Hinge-region exon, M1 and M2 represent the exons encoding the trans-membrane region of surface expressed Ig heavy chain. With a simple one-step ligation the cDNA C-gamma1 region of the transposable human heavy chain vector is replaced by its genomic counterpart as indicated in the figure. Using this strategy, the V_H coding region will be ligated in-frame to the C_H1 coding exon of human C-gamma1.

[0059] Fig. 15: This figure shows the sequence and overall design of the kappa light chain library. CDR3 coding region is underlined. Useful restriction sites are indicated.

[0060] Fig. 16: This figure shows the sequence and overall design of the gamma heavy chain library, showing as an example the library fragment randomized using the NNK4 randomization strategy. The gamma heavy chain library fragments randomized using the NNK6, NNK8 and NNK10 randomization strategies differ only in the number of randomized amino acid residues in the HCDR3 region. HCDR3 coding region is underlined. The ARG codon encodes Lysine and Arginine. Useful restriction sites are indicated.

[0061] Fig. 17: This figure shows the digestion of PCR templates prior to amplification with primers. (A) Digestion of pUC57_Jkappa2-Ckappa with the restriction endonuclease ScaI produces a blunt-ended DNA fragment ideal for priming with the primer LCDR3-NNK6-F. (B) Digestion of pUC57_J_H4 with the restricion endonuclease DrdI produces a DNA fragment ideal for priming with the primers HCDR3-NNK4-F, HCDR3-NNK6-F, HCDR3-NNK8-F, and HCDR3-NNK10-F

[0062] Fig. 18: This figure shows the electropherograms spanning the randomized LCDR3 and HCDR3 region of the PCR amplicons generated to diversify the LCDR3 region by the NNK-6 approach for Vkappa (A), and the HCDR3 region by the NNK4-approach for V_H, as disclosed in Examples 12 and 13, respectively.

DETAILED DESCRIPTION OF THE INVENTION

[0063] *Definitions*

[0064] As used herein, "diverse collection" means a plurality of variants or mutants of particular functional or binding proteins exhibiting differences in the encoding nucleotide sequences or in the primary amino acid sequences, which define different functionalities or binding properties.

[0065] As used herein, "library" means a plurality of polynucleotides encoding polypeptides having different binding specificities and/or functionalities. In certain embodiments, the library may comprise polynucleotides encoding at least 10^2, at least 10^3, at least 10^4, at least 10^5, at least 10^6, at least 10^7, at least 10^8, or at least 10^9 unique polypeptides, such as, for example, full-length antibody heavy or light chains or VH or VL domains.

[0066] As used herein, "inverted terminal repeat sequence" or "ITR" means a sequence identified at the 5' or 3' termini of transposable elements that are recognized by transposases and which mediate the transposition of the ITRs including intervening coding information from one DNA construct or locus to another DNA construct or locus.

[0067] As used herein, "transposase" means an enzyme that has the capacity to recognize and to bind to ITRs and to mediate the mobilization of a transposable element from one target DNA sequence to another target DNA sequence.

[0068] As used herein, "antigen binding molecule" refers in its broadest sense to a molecule that specifically binds an antigenic determinant. A non-limiting example of an antigen binding molecule is an antibody or fragment thereof that retains antigen-specific binding. By "specifically binds" is meant that the binding is selective for the antigen and can be discriminated from unwanted or nonspecific interactions.

[0069] As used herein, the term "*antibody*" is intended to include whole antibody molecules, including monoclonal, polyclonal and multispecific (*e.g.*, bispecific) antibodies, as well as antibody fragments having an Fc region and retaining binding specificity, and fusion proteins that include a region equivalent to the Fc region of an immunoglobulin and that retain binding specificity. Also encompassed are antibody fragments that retain binding specificity including, but not limited to, VH fragments, VL fragments, Fab fragments, $F(ab')_2$ fragments, scFv fragments, Fv fragments, minibodies, diabodies, triabodies, and tetrabodies (*see, e.g.*, Hudson and Souriau, *Nature Med.* 9: 129-134 (2003)) (incorporated by reference in its entirety).

[0070] An embodiment of the invention disclosed herein is a method for the identification of specific functional or binding polypeptides, including, but not limited to antibody chains or fragments thereof (Fig. 4), which comprises:

i. cloning of diverse transposable DNA libraries encoding proteins, including antibody polypeptide chains or fragments thereof, in between inverted terminal repeats (ITRs) derived from transposable elements and recognizable by and functional with at least one transposase enzyme,

ii. introduction of one or more diverse transposable DNA libraries of step (i) into vertebrate host cells by standard methods known in the art,

iii. providing temporary expression of at least one functional transposase enzyme in said vertebrate host cells in trans, such that said one or more diverse transposable DNA libraries are stably integrated into the vertebrate host cell genomes, thereby providing a vertebrate host cell population that then stably expresses diverse libraries of proteins, including antibody chains or fragments thereof,

iv. screening of said diverse cellular libraries, stably expressing proteins, including antibodies or fragments thereof, for a desired functional or binding phenotype by methods known in the art,

v. optionally, including iterative enrichment cycles with the stably genetically modified vertebrate host cells for a desired binding or functional phenotype, and

vi. isolation of the corresponding genes from the enriched host cells encoding the desired binding or functional phenotype by standard cloning methods, known in the art, for instance, but not limited to, PCR (polymerase chain reaction), using primers specific for the sequences contained in the one or more transposed DNA library constructs.

[0071] A preferred embodiment of step (i) is to generate diverse transposable DNA libraries either by gene synthesis, or by polymerase chain reaction (PCR) using appropriate primers for the amplification of diverse protein coding regions, and DNA templates comprising a diversity of binding proteins, including antibodies, or fragments thereof, by methods known in the art.

[0072] For the generation of diverse antibody libraries, a diverse collection of antibody heavy and light chain sequences may be generated by standard gene synthesis in which the V region coding sequences may be randomized at certain positions, e.g. but not limited to, any or all of the complementarity determining regions (CDRs) of the antibody heavy or light chain V-regions. The diversity can be restricted to individual CDRs of the V-regions, or to a particular or several framework positions, and/or to particular positions in one or more of the CDR regions. The V regions with designed variations, as described above, can be synthesized as a fragment encoding entire antibody heavy or light chains that are flanked by inverted terminal repeats functional for at least one desired transposase enzyme. Preferably, the DNA library containing diverse variable domains encoding V regions for antibody heavy or light chains is generated, and flanked by appropriate cloning sites, including but not limited to restriction enzyme recognition sites, that are compatible with cloning sites in antibody heavy or light chain expression vectors. Useful transposon expression systems for use in the methods of the invention include, for example, the *PiggyBac* transposon system as described, for example, in US Pat. Nos. 6,218,185; 6,551,825; 6,962,810; 7,105,343; and 7,932,088 (the entire contents

of each of which are hereby incorporated by reference) and the *Sleeping Beauty* transposon system as described in US Pat. Nos. 6,489,458; 7,148,203; 7,160,682; US 2011 117072; US 2004 077572; and US 2006 252140 (the entire contents of each of which are hereby incorporated by reference.)

[0073] Diverse antibody heavy and light chain libraries may also be obtained from B cell populations isolated from desired vertebrate species, preferably humans, and preferably from cellular compartments containing B cells, e.g., but not limited to peripheral or cord blood, and lymphoid organs like bone marrow, spleen, tonsils and lymph-node tissues. In this case, diverse antibody V region sequences for antibody heavy and light chains can be isolated by RT-PCR or by genomic PCR using antibody heavy and light chain specific degenerate PCR primer pairs, that can amplify the majority of V-region families by providing upstream primers that bind to homologous sequences upstream of, or within leader sequences, upstream of or within V-region frameworks, and by providing downstream primers that bind in regions of homology within or downstream of the J joining gene segment of variable domain coding regions, or within or downstream of the coding regions of the constant regions of antibody heavy or light chains.

[0074] The PCR primer sets utilized for the amplification of diverse variable coding regions may be flanked by appropriate cloning sites, e.g. but not limited to restriction enzyme recognition sites, that are compatible with cloning sites in antibody heavy or light chain expression vectors.

[0075] The transposable DNA libraries of step (i) encoding diverse proteins, including antibodies and antibody fragments thereof, can be provided in the form of plasmid libraries, in which the gene-synthesized or the PCR amplified transposable DNA libraries are cloned using appropriate cloning sites, as mentioned above. Alternatively, the transposable DNA libraries encoding diverse libraries of binding proteins, such as antibodies and fragments thereof, can be provided in form of linear, double-stranded DNA constructs, directly as a result of DNA synthesis, or as a result of PCR amplification. The latter approach of providing the transposable DNA libraries as linear double-stranded DNA PCR amplicons, that have not been cloned into expression vectors or plasmids (in comparison to all other vertebrate cell expression systems) has the advantage that the maximum molecular complexity of the transposable DNA libraries is maintained and not compromised by a limited cloning or ligation efficiency into an expression vector. In contrast, cloning by ligation, or otherwise, into plasmid expression or shuttle vectors is a necessary intermediate for all other plasmid-based or viral vector based vertebrate cell expression systems.

[0076] However, the use of plasmid-based transposon expression vectors containing the diverse transposable DNA libraries encoding diverse binding proteins, including antibodies and antibody fragments thereof, has the advantage that these expression vectors can be engineered to contain additional functional elements, that allow the screening, or, alternatively, the selection for stably transposed vertebrate host cells for the stable integration of the transposon expression vector in transposed vertebrate host cells.

[0077] This is achieved by providing in operable linkage to the diverse transposable DNA libraries, i.e. cloned into the transposon expression vectors in *cis*, expression cassettes for marker genes including., but not limited to, fluorescent marker proteins (e.g. green, yellow, red, or blue fluorescent proteins, and enhanced versions thereof, as known in the art), or expression cassettes for cell surface markers including, but not limited to, CD markers, against which specific diagnostic antibodies or other diagnostic tools are available.

[0078] Alternatively, expression cassettes for selectable markers, that allow selection of transposed vertebrate host cells for antibiotic resistance, including, but not limited to, puromycin, hygromycinB, bleomycin, neomycin resistance, can be provided in operable linkage to the diverse transposable DNA libraries, i.e. cloned into the transposon expression vectors in *cis*.

[0079] The operable linkage can be achieved by cloning of said expression cassettes for marker genes or antibiotic resistance markers, either up- or downstream of the coding regions comprising said diverse transposable DNA libraries, but within the inverted terminal repeats of the transposon vector.

[0080] Alternatively, the operable linkage can be achieved by cloning of the coding regions for said marker or antibiotic resistance genes downstream of the coding regions comprising said diverse transposable DNA libraries, but separated by internal ribosomal entry site (IRES) sequences, that ensure transcriptional coupling of the expression of said diverse transposable DNA libraries with said marker or antibiotic resistance genes, and thereby allowing the screening for or selection of stably transposed vertebrate host cells.

[0081] In step (ii) of the method disclosed herein, said diverse transposable DNA libraries encoding diverse libraries of proteins, including antibodies and fragments thereof, are introduced into desired vertebrate host cells by methods known in the art to efficiently transfer DNA across vertebrate cell membranes, including., but not limited to, DNA-transfection using liposomes, Calcium phosphate, DEAE-dextran, polyethyleneimide (PEI) magnetic particles, or by protoplast fusion, mechanical transfection, including physical, or ballistic methods (gene gun), or by nucleofection. Any of the above-mentioned methods and other appropriate methods to transfer DNA into vertebrate host cells may be used individually, or in combination for step (ii) of the method disclosed herein.

[0082] In the case of dimeric proteins, including, but not limited to, antibodies and fragments thereof, it is a useful embodiment of the method disclosed herein to introduce diverse transposable DNA libraries and/or transposon vectors for antibody heavy or light chains contained in separate transposable vectors, which can independently be introduced into the vertebrate host cells. This either allows the sequential introduction of diverse transposable DNA libraries for antibody heavy or light chains into said cells, or their simultaneous introduction of diverse transposable DNA libraries for antibody heavy or light chains, which, in either case, allows the random shuffling of any antibody heavy with any antibody light chain encoded by the at least two separate diverse transposable DNA libraries.

[0083] Another useful embodiment of the previous embodiment is to utilize separate transposon vectors and/or diverse DNA transposable libraries for antibody heavy and light chains, where said constructs or libraries are contained on transposable vectors recognized by different transposase enzymes (Fig. 3). This allows the independent transposition of antibody heavy and antibody light chain constructs without interference between the two different transposase enzymes, as one transposable vector is only recognized and transposed by its specific transposase enzyme. In case of sequential transposition of transposable vectors or DNA libraries encoding antibody heavy or light chains, the advantage of utilizing different transposase enzymes with different ITR sequences is, that upon the second transposition event, the first already stably transposed construct is not again mobilized for further transposition.

[0084] This embodiment also allows the discovery of antibodies by the method of guided selection (Guo-Qiang *et al. Methods Mol. Biol. 562*, 133-142 (2009)) (incorporated herein by reference in its entirety). Guided selection can e.g. be used for the conversion of any non-human antibody specific for a desired target/epitope specificity and with a desired functionality into a fully human antibody, where the same target/epitope specificity and functionality is preserved. The principle of guided selection entails the expression of a single antibody chain (heavy or light chain) of a reference (the "guiding") antibody, in combination with a diverse library of the complementary antibody chains (i.e. light, or heavy chain, respectively), and screening of these heavy-light chain combinations for the desired functional or binding phenotype. This way, the first antibody chain, "guides" the selection of one or more complementary antibody chains from the diverse library for the desired functional or binding phenotype. Once the one or more novel complementary antibody chains are isolated, they can be cloned in expression vectors and again be used to "guide" the selection of the second, complementary antibody chain from a diverse antibody chain library. The end-result of this two-step process is that both original antibody heavy and light chains of a reference antibody are replaced by

unrelated and novel antibody chain sequences from the diverse libraries, but where the novel antibody heavy-light chain combination exhibits the same, or similar functional or binding properties of the original reference antibody. Therefore, this method requires the ability to independently express antibody heavy and light chain constructs or libraries in the vertebrate host cells, which can be achieved by the preferred embodiment to provide antibody heavy and light chain expression cassettes in different transposable vector systems, recognized by different transposon enzymes.

[0085] However, diverse transposable DNA libraries can also be constructed in a way, that the coding regions for multimeric proteins, including antibodies and fragments thereof, are contained in the same transposon vector, i.e. where the expression of the at least two different subunits of a multimeric protein, for example V_H and V_L regions or full-length heavy and light chains, isoperably linked by cloning of the respective expression cassettes or coding regions into the same transposable vector.

[0086] Useful vertebrate host cells for the introduction of transposable constructs and/or transposable DNA libraries of step (ii) are cells from vertebrate species that can be or that are immortalized and that can be cultured in appropriate cell culture media and under conditions known in the art. These include, but are not limited to, cells from e.g. frogs, fish, avians, but preferably from mammalian species, including, but not limited to, cells from rodents, ruminants, non-human primate species and humans, with cells from rodent or human origin being preferred.

[0087] Useful cell types from the above-mentioned species include, but are not limited to cells of the lymphoid lineage, which can be cultured in suspension and at high densities, with B-lineage derived cells being preferred, as they endogenously express all the required proteins, factors, chaperones, and post-translational enzymes for optimal expression of many proteins, in particular of antibodies, or antibody-based proteins. Of B-lineage derived vertebrate cells, those are preferred that represent early differentiation stages, and are known as progenitor (pro) or precursor (pre) B cells, because said pro- or preB cells in most cases do not express endogenous antibody chains that could interfere with exogenous or heterologous antibody chain expression that are part of the method disclosed herein.

[0088] Useful pro- and pre- B lineage cells from rodent origin are Abelson-Murine Leukemia virus (A-MuLV) transformed proB and preB cells (Alt *et al. Cell 27*, 381-390(1981) (incorporated herein by reference in its entirety)) that express all necessary components for antibody expression and also for their proper surface deposition, including the B cell receptor components Ig-alpha (CD79a, or mb-1), and Ig-beta (CD79b, or B-29) (Hombach *et al. Nature 343*, 760-762 (1990)) (incorporated herein by reference in its entirety), but as mentioned above, mostly lack the expression of endogenous antibody or immunoglobulin

chains. Here, A-MuLV transformed pro- and preB cells are preferred that are derived from mouse mutants, including, but not limited to, mouse mutants defective in recombination activating gene-1 (RAG-1), or recombination activating gene-2 (RAG-2), or animals carrying other mutations in genes required for V(D)J recombination, e.g. XRCC4, DNA-ligase IV, Ku70, or Ku80, Artemis, DNA-dependent protein kinase, catalytic subunit (DNA-PK_{cs}), and thus lack the ability to normally express of endogenous antibody polypeptides.

[0089] Additional useful types of progenitor (pro) and precursor (pre) B lineage cells are early, immunoglobulin-null (Ig-null) EBV transformed human proB and preB cells (Kubagawa *et al. PNAS 85*, 875-879(1988)) (incorporated herein by reference in its entirety) that also express all the required factors for expression, post-translational modification and surface expression of exogenous antibodies (including CD79a and CD79b).

[0090] Other host cells of the B lineage can be used, that represent plasma cell differentiation stages of the B cell lineage, preferably, but not limited to Ig-null myeloma cell lines, like Sp2/0, NSO, X63, Ag8653, and other myeloma and plasmacytoma cells, known in the art. Optionally, these cell lines may be stably transfected or stably genetically modified by other means than transfection, in order to over-express B cell receptor components Ig-alpha (CD79a, or mb-1), and Ig-beta (CD79b, or B-29), in case optimal surface deposition of exogenously expressed antibodies is desired.

[0091] Other, non-lymphoid mammalian cells lines, including but not limited to, industry-standard antibody expression host cells, including, but not limited to, CHO cells, Per.C6 cells, BHK cells and 293 cells may be used as host cells for the method disclosed herein, and each of these cells may optionally also be stably transfected or stably genetically modified to over-express B cell receptor components Ig-alpha (CD79a, or mb-1), and Ig-beta (CD79b, or B-29), in case optimal surface deposition of exogenously expressed antibodies is desired.

[0092] Essentially, any vertebrate host cell, which is transfectable, can be used for the method disclosed herein, which represents a major advantage in comparison to any viral expression systems, such as., but not limited to vaccinia virus, retroviral, adenoviral, or sindbis virus expression systems, because the method disclosed herein exhibits no host cell restriction due to virus tropism for certain species or cell types, and furthermore can be used with all vertebrate cells, including human cells, at the lowest biosafety level, adding to its general utility.

[0093] Step (iii) of the method disclosed herein results in the stable genetic modification of desired vertebrate host cells with the transfected transposable constructs of step (ii) by temporary, or transient expression of a functional transposase enzyme, such that a stable population of

vertebrate host cells is generated that expresses diverse libraries of proteins encoded by said constructs.

[0094] A useful embodiment of step (iii) is to transiently introduce into the host cells, preferably by co-transfection, as described above, a vertebrate expression vector encoding a functional transposase enzyme together with said at least one diverse transposable DNA library. It is to be understood that transient co-transfection or co-integration of a transposase expression vector can either be performed simultaneously, or shortly before or after the transfer of the transposable constructs and/or diverse transposable DNA libraries into the vertebrate host cells, such that the transiently expressed transposase can optimally use the transiently introduced transposable vectors of step (ii) for the integration of the transposable DNA library into the vertebrate host cell genome.

[0095] Another useful embodiment of step (iii) is to effect the stable integration of the introduced transposable vectors and/or transposable DNA libraries of step (ii) by transiently expressing a functional transposase enzyme by means of an inducible expression system known in the art, that is already stably integrated into the vertebrate host cell genome. Such inducible and transient expression of a functional transposase may be achieved by e.g., tetracycline inducible (tet-on/tet-off) or tamoxifen-inducible promoter systems known in the art. In this case, only the one or more transposable vector or DNA library needs to be introduced into the host cell genome, and the stable transposition of the constructs and the stable expression of the proteins encoded by the one or more transposable vector or DNA library is effected by the transiently switched on expression of the functional transposase enzyme in the host cells.

[0096] Step (iv) of the method disclosed herein effects the isolation of transposed vertebrate host cells expressing proteins with a desired functionality or binding phenotype.

[0097] A preferred embodiment of step (iv) is to screen for and to isolate the transposed host cells of step (iii) expressing desired proteins, including antibodies and fragments thereof, with target-binding assays and by means of standard cell separation techniques, like magnetic activated cell sorting (MACS) or high-speed fluorescence activated cell sorting (FACS) known in the art. Especially, in a first enrichment step of a specific population of transposed vertebrate host cells, where large number of cells need to be processed, it is preferred to isolate target specific cells from a large number of non-specific cells by MACS-based techniques.

[0098] Particularly, for additional and iterative cell enrichment cycles, FACS enrichment is preferred, as potentially fewer numbers of cells need to be processed, and because multi channel flow cytometry allows the simultaneous enrichment of functionalities, including., but not limited to, binding to a specific target of more than one species, or the specific screening for particular epitopes using epitope-specific competing antibodies in the FACS screen.

[0099] If proteins, including antibodies and fragments thereof, are to be discovered that interact with soluble binding partners, these binding partners are preferably labeled with specific labels or tags, such as but not limited to biotin, myc, or HA-tags known in the art, that can be detected by secondary reagents, e.g. but not limited to, streptavidin or antibodies, that themselves are labeled magnetically (for MACS based cell enrichment) or with fluorochromes (for FACS based cell enrichment), so that the cell separation techniques can be applied.

[00100] If proteins, including antibodies and fragments thereof are to be discovered against membrane bound proteins, which cannot easily be expressed as soluble proteins, like e.g. but not limited to, tetraspannins, 7-transmembranbe spanners (like G-coupled protein coupled receptors), or ion-channels, these may be expressed in viral particles, or overexpressed in specific cell lines, which are then used for labeling or panning methods known in the art, which can enrich the vertebrate host cells expressing the proteins from the transposed constructs, including antibodies and fragments thereof.

[00101] Due to the stable genotype-phenotype coupling in the stably transposed vertebrate host cell population, a useful embodiment of step (v) is to repeat cell enrichment cycles for a desired functional or binding phenotype, until a distinct population of cells is obtained that is associated with a desired functional or binding phenotype. Optionally, individual cell clones can be isolated e.g., but not limited to, by single-cell sorting using flow cytometry technology, or by limiting dilution, in order to recover the transposed DNA information from individual cell clones that are coupled to a particular, desired functional or binding phenotype.

[00102] For the identification of functional target-specific antibodies it is often favorable to not only screen and to select for a particular binding phenotype, but to additionally screen for additional functional properties of target specific antibodies, in particular antagonistic or agonistic effects in biological assay.

[00103] Therefore, it is desirable to be able to efficiently "switch" cell membrane bound antibody expression to secreted antibody expression in the vertebrate host cells with sufficient yields, in order to produce enough quantity of a particular antibody clone for functional assays.

[00104] In natural B lineage cells the switch from membrane bound to secreted antibody expression occurs via a mechanism of alternative splicing, in which in preB and B cells an alternative splice donor near the 3'end of the last heavy chain constant region exon is preferentially spliced to a splice acceptor of a membrane anchor exon downstream of the heavy chain constant regions exons. This way, an antibody heavy chain is produced in B cells with an extended C-terminal, membrane spanning domain, that anchors the heavy chain and thereby the entire heavy-light chain containing antibody in the cell membrane. The C-terminal, membrane spanning domain also interacts non-covalently with the membrane spanning

components Ig-alpha (CD79a or mb-1) and Ig-beta (CD79b or B29), which likely results in better membrane anchoring and higher surface immunoglobulin expression in B lineage cells.

[00105] Once, a B cell differentiates further to the plasma cell stage, the alternative splicing does not occur any more and the alternative splice donor near the 3'end of the last heavy chain constant region is no longer recognized or utilized, and the mRNA template is terminated downstream of the heavy chain constant region stop codon, and a heavy chain of a secreted antibody is translated.

[00106] In order to exploit this natural mechanism of alternative splicing and "switching" from membrane bound to secreted expression of expressed antibodies, it is a useful embodiment of the method disclosed herein to construct the transposable vectors and diverse DNA libraries encoding proteins, including antibodies or fragments thereof, in such a way that the natural intron/exon structure of a constant antibody heavy chain, including the exons encoding the membrane spanning domains is maintained. This embodiment represents a clear advantage against retroviral expression systems, as the retroviral vector genome is already spliced before it is packaged into a retroviral particle and stably transduced into the host cell genome.

[00107] Other viral vector systems may be restricted in the length of the DNA insert that can be incorporated into the vectors, thereby precluding the cloning of larger genomic regions into such expression vectors and thereby preventing the exploitation of the natural "switching" from membrane-bound to secreted antibody expression by alternative splicing. Certain transposons (e.g. *Tol2*, see Fig. 3), have been characterized to be able to efficiently transpose more than 10 kb DNA fragments into vertebrate host cells without any loss in transposition efficiency (Kawakami *Genome Biol. 8, Suppl I*, S7 (2007)) (incorporated herein by reference in its entirety). Therefore, it is a useful embodiment of the method disclosed herein to construct transposable expression vectors comprising genomic exon/intron structures for better and proper expression and for the natural regulation switching from membrane bound to secreted antibody expression. The methods of the invention are useful to transpose DNA fragments at least 5kb, at least 6kb, at least 7kb, at least 8kb, at least 9 kb, at least 10 kb in size into host cell genomes.

[00108] The differentiation of earlier B lineage differentiation stage that favors membrane bound antibody expression, to a later, plasma cell stage, that favors secreted antibody expression can be induced by B cell differentiation factors, such as, but not limited to, CD40/IL4 triggering, or stimulation by mitogens, such as, but not limited to, lipopolysaccharide (LPS), or other polyclonal activators, Staph. aureus Cowan (SAC) strain activators, and CpG nucleotides, or any combination thereof.

[00109] Preferrably, this differentiation is effected in transformed cells, in which the proliferation can artificially be inhibited, such that proper B cell differentiation can again occur, as it has been described for A-MuLV transformed murine preB cells, in which the Abelson tyrosine kinase is specifically inhibited by the tyrosine inhibitor Gleevec (Muljo *et al. Nat. Immunol 4*, 31-37 (2003)) (incorporated herein by reference in its entirety). Therefore, it is a preferred embodiment to utilize Ig-null A-MuLV transformed murine preB cells for the method, which by treatment with Gleevec, can again differentiate to more mature B cell stages, including plasma cells, which then secrete sufficient amounts of secreted antibody for additional functional testing on the basis of alternative splicing of genomic heavy chain expression constructs. It is a preferred embodiment of the method disclosed herein, to further improve such B-lineage cell differentiation by stable overexpression of anti-apoptotic factors, known in the art, including, but not limited to, bcl-2 or bcl-x_L.

[00110] After step (iv), the enrichment of transposed vertebrate host cells as described above has been performed, optionally, additional cell enrichments according to the above-mentioned methods may be performed (step (v)), until cell populations, or individual cells are isolated expressing proteins, including antibodies and fragments thereof, with desired functional and/or binding properties.

[00111] Step (vi) of the method disclosed herein is then performed in order to isolate the relevant coding information contained in the transposed vertebrate host cells, isolated for a desired functional and/or binding property.

[00112] A useful embodiment of step (vi) for the isolation cloning and sequencing of the relevant coding information for a desired functional or binding protein, including an antibody or antibody fragment thereof, contained in the isolated cells, is to utilize genomic or RT-PCR amplification with specific primer pairs for the relevant coding information comprised in the transposed DNA constructs, and to sequence the genomic or RT-PCR amplicons either directly, or after sub-cloning into sequencing vectors, known in the art, e.g., but not limited into TA- or Gateway-cloning vectors.

[00113] Cloning and sequencing of the relevant coding information for a desired functional or binding protein, including an antibody or antibody fragment thereof, as described in the previous paragraph by genomic or RT-PCR amplification can also be performed with transposable IgH and IgL expression vector, such that the binding protein coding region cannot only be identified, but at the same time also be expressed upon stable transposition into mammalian host cells as disclosed herein. For the expression of secreted antibodies this would only require the use of transposable Ig heavy chain expression vectors lacking the IgH transmembrane spanning coding region.

[00114] Another useful embodiment of step (vi) is to subject the enriched cell populations of steps (iv) or (v), which exhibit a desired functional or binding phenotype to next-generation ("deep") sequencing (Reddy *et al. Nat. Biotech. 28*, 965-969 (2010)) (incorporated herein by reference in its entirety), in order to retrieve directly and in one step a representative set of several thousands of sequences for the coding information contained in the transposed DNA constructs. Based on a bioinformatics analysis of the relative frequency of sequences identified from the enriched cell populations, it allows a prediction about which sequences encoded a functional or binding protein, including an antibody or fragment thereof (Reddy *et al. Nat. Biotech. 28*, 965-969 (2010)) (incorporated herein by reference in its entirety). Statistically overrepresented sequences are then resynthesized and cloned into expression vector for expression as recombinant proteins, antibodies or fragments thereof, in order to characterize them functionally and for their binding properties. This method can significantly accelerate the identification of relevant sequences within a functionally and phenotypically enriched cell population, that expresses proteins with functional or target specific properties.

[00115] Yet another useful embodiment of the method disclosed herein is to utilize transposition-mediated vertebrate cell expression of proteins, including antibodies or fragments thereof, for the mutagenesis and optimization of desired proteins, including the affinity optimization of antibodies and fragments thereof.

[00116] This can be achieved by isolating the genes encoding the proteins, including antibody chains or fragments thereof, from transposed vertebrate cell populations enriched for a desired binding or functional phenotype according to the methods disclosed in step (iv), such as but not limited to, by genomic PCR or RT-PCR amplification under mutagenizing conditions, know in the art. The mutagenized sequences can then be re-cloned into transposition vectors and then again be transposed into vertebrate host cells, in order to subject them to screening according to the methods disclosed herein, for improved functional or binding properties.

[00117] In one useful embodiment of this approach, specific primers are used that allow the PCR amplification under mutagenizing conditions of complete transposed constructs, including the flanking ITRs.

[00118] By this method a mutagenized PCR amplicon containing a defined average frequency of random mutations is generated from the functionally or phenotypically selected transposed cells. Said PCR amplicon with controlled mutations (variations) of the original templates can now directly be re-transposed into new vertebrate host cells, according to preferred embodiments disclosed in the methods applicable in step (ii).

[00119] The main advantage of this method over other approaches of genetically modifying vertebrate cells is, that with this technology no time-consuming re-cloning of the mutagenized PCR

amplicons and time consuming quality control of the mutagenized sequences into expression vectors is required, which is a mandatory requirement in all other plasmid-based or viral expression systems, if a mutagenized sequence shall be subjected to another round of screening.

[00120] Because transposition of DNA only requires the presence of ITRs flanking the coding region of genes of interest, PCR-amplified mutagenized PCR amplicons can directly be re-introduced and re-transposed into novel vertebrate host cells for expression and screening for improved properties and/or affinity matured mutants.

[00121] Taken together, the methods disclosed herein, of utilizing TEs for the stable genetic modification of vertebrate host cells with transposable constructs and/or diverse transposable DNA libraries encoding proteins, including antibodies and fragments thereof, offers unparalleled efficiency, flexibility, utility and speed for the discovery and optimization of said proteins for optimal desired functional or binding phenotypes.

Examples:

Example 1: Instruction for cloning of basic *PiggyBac* transposable light chain expression vector for human antibody kappa light chains compatible with the *PiggyBac* transposase enzyme

[00122] A basic transposable expression vector for human kappa light chains can be generated by cloning of the ITRs from the *PiggyBac* transposon up and downstream of a human immunoglobulin kappa light chain expression cassette.

[00123] For this, as a first step, the minimal sequences for the up- and downstream ITRs of the *PiggyBac* transposon can be derived from pXLBacII (published in US 7,105,343) (incorporated herein by reference in its entirety) and can be gene synthesized with flanking restriction enzyme sites for cloning into the mammalian expression vector pIRES-EGFP (PT3157-5, order #6064-1, Invitrogen-Life Technologies, Carlsbad, CA, USA)

[00124] The upstream *PiggyBac* ITR sequence with the 5' terminal repeat has to be gene synthesized with flanking MunI restriction enzyme sequence, compatible with a unique MunI restriction enzyme site in pIRES-EGFP, and additional four random nucleotides (in lowercase letters) allowing proper restriction enzyme digestion. This sequence is provided in Seq-ID1.

[00125] The downstream *PiggyBac* ITR sequence with the 3' terminal repeat has to be gene synthesized with flanking XhoI restriction enzyme sequence compatible with a unique XhoI restriction enzyme site in pIRES-EGFP, and additional four random nucleotides allowing

proper restriction enzyme digestion. This sequence is provided in Seq-ID2.Upon MunI restriction enzyme digestion of the gene synthesized Seq-ID1, the DNA fragment can be ligated into MunI linearized pIRES-EGFP, generating pIRES-EGFP-TR1 according to standard methods, known in the art. The proper orientation of the insert can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned construct (Fig. 5a).

[00126] In a next step gene synthesized and XhoI digested DNA fragment Seq-ID2, can be ligated into XhoI linearized pIRES-EGFP-T1 (Fig. 5a) by standard methods known in the art, in order to generate pIRES-EGFP-T1T2, containing both *PiggyBac* ITRs up and downstream of the IRES-EGFP expression cassette (Fig. 5b). The proper orientation of the insert can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned construct (Fig. 5b).

[00127] For the cloning of a human immunoglobulin kappa light chain into the vector pIRES-EGFP-T1T2, the human Ig kappa light chain from human anti-TNF-alpha specific antibody D2E7 can be synthesized, which can be retrieved from European patent application EP 1 285 930 A2 (incorporated herein by reference in its entirety).

[00128] The coding region for human Ig kappa light chain of human anti-TNF-alpha specific antibody D2E7, in which the V region of D2E7 is fused in frame to a Vk1-27 leader sequence (Genbank entry: X63398.1, which is the closest germ-line gene V-kappa family member V-kappa of D2E7), and to the human kappa constant region (Genbank entry: J00241) has the following nucleotide sequence, which is provided in Seq-ID3.

[00129] The nucleotide sequence of Seq-ID3 translates in the amino acid sequence Seq-ID4.The DNA fragment Seq-ID3 encoding the D2E7 Ig kappa light chain can be gene synthesized and directly ligated by blunt-ended ligation into the unique EcoRV restriction enzyme site (which is also a blunt cutter), by methods know in the art, resulting in construct pIRES-EGFP-T1T2-IgL (Fig. 5b)

[00130] Seq-ID3 has been engineered to contain a unique Eco47III restriction enzyme site in between the V-kappa and the C-kappa coding regions (highlighted in boldface and underlined), which allows the replacement of V-kappa regions in this construct against other V-kappa regions or V-kappa libraries, using a unique restriction enzyme upstream of V-kappa coding region in the construct, together with Eco47III. The proper orientation of the kappa light chain insert can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned construct (Fig. 5b).

[00131] The entire sequence for the transposable human antibody kappa light chain vector pIRES-EGFP-T1T2-IgL is provided in sequence Seq-ID5.

[00132] Sequences referred to in this Example 1:

[00133] Seq-ID1 (327 bp long *PiggyBac* 5'-ITR sequence. The MunI restriction enzyme sites at each end are underlined and typed in boldface print, the random nucleotide additions at the termini are printed in lowercase)

[00134] Seq- ID2 (264 bp long *PiggyBac* 3'-ITR sequence. The XhoI restriction enzyme sites at each end are underlined and typed in boldface print, the random nucleotide additions at the termini are printed in lowercase)

[00135] Seq-ID3 (711 bp long Ig-kappaLC coding region of anti-TNF-alpha-specific mAb D2E7)

[00136] Seq-ID4 (236 amino acids long sequence of anti-TNF-alpha-specific mAb D2E7)

[00137] Seq-ID5 (6436bp long DNA sequence of *PiggyBac* transposable Ig-kappa-LC expression vector pIRES-EGFP-T1T2-IgL)

Example 2: Instruction for cloning of a basic *PiggyBac* transposable heavy chain expression vector for membrane spanning human antibody gamma1 heavy chains

[00138] In order to clone a transposable Ig heavy chain expression vector, the kappa light chain ORF from pIRES-EGFP- T1T2-IgL (Seq-ID5) needs to be exchanged with an ORF encoding a fully human IgG1 heavy chain coding region.

[00139] For the replacement of the human kappa light chain in vector pIRES-EGFP- T1T2-IgL by a human immunoglobulin gamma-1 heavy chain, the V_H region of antibody D2E7, which is specific for human TNF-alpha (see: EP 1 285 930 A2) (incorporated herein by reference in its entirety) can be synthesized. For this, a leader sequence of a close germ-line V_H3-region family member is fused in frame to the the V_H region of antibody D2E7, which then is fused in frame to the coding region of a human gamma1 constant region (Genbank: J00228) including the membrane spanning exons (Genbank: X52847). In order to be able to replace the human Ig kappa light chain from pIRES-EGFP- T1T2-IgL, unique ClaI and NotI restriction enzyme sites need to be present at the 5' and the 3' end of the sequence (underlined), respectively. Additionally, four nucleotides flanking the restriction enzyme sites (highlighted in lowercase letters at the ends of the sequence) allow proper restriction enzyme digestion of the gene-synthesized DNA fragment and ligation into the ClaI-NotI linearized pIRES-EGFP-T1T2-IgL backbone, according to standard methods. The sequence that needs to be gene synthesized is provided in Seq-ID6.

[00140] From the start codon in position 11 of Seq-ID6 (highlighted in boldface print), this nucleotide sequence translates to the human IgG1 heavy chain of anti-TNF-alpha specific clone D2E7 (see: EP 1 285 930 A2) (incorporated herein by reference in its entirety), but including the human gamma1 transmembrane exons M1 and M2. The protein translation of Seq-ID6 is provided in Seq-ID7.The DNA fragment Seq-ID6 encoding the D2E7 Ig gamma-1 heavy chain can then be double-digested by ClaI and NotI restriction enzymes and directionally ligated into ClaI and NotI linearized pIRES-EGFP- T1T2-IgL, resulting in construct pIRES-EGFP- T1T2-IgH (Fig. 6)

[00141] Seq-ID6 has also been engineered to contain a unique Eco47III restriction enzyme site in between the V-heavy variable and the C-gamma1 constant coding regions (highlighted in boldface and underlined in Seq-ID7), which allows the replacement of V-heavy regions in this construct against other V-heavy regions or V-heavy libraries, using a unique restriction enzyme upstream of V-heavy coding region in the construct, together with Eco47III. The correct ligation of the insert can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned construct (Fig. 6).

[00142] The entire sequence for the transposable human antibody gamma-1 heavy chain vector pIRES-EGFP- T1T2-IgH is provided in sequence Seq-ID8

[00143] Examples 1 and 2 provide instructions for the cloning of basic *PiggyBac* transposable expression vectors for human antibody kappa light and human gamma-1 heavy chains (membrane bound form) and therefore for full-length, membrane bound human IgG1, that can be utilized for the reduction to practice of the invention.

[00144] Sequences referred to in this Example 2:

[00145] Seq-ID6 (1642 bp long DNA fragment containing the coding region for membrane bound Ig-gamma1-HC of anti-TNF-alpha-specific mAb D2E7)

[00146] Seq-ID7 (539 amino acids long sequence of membrane bound Ig-gamma1-HC of anti-TNFalpha antibody

[00147] Seq-ID8 (7341bp long DNA sequence of *PiggyBac* transposable human Ig-gamma1-membrane-HC expression vector pIRES-EGFP-T1T2-IgH)

Example 3: Instructions for cloning of basic transposable light chain expression vector for human antibody kappa light chains compatible with the *Sleeping Beauty* transposase enzyme

[00148] In order to transpose human immunoglobulin heavy and light chain expression vectors contained in a transposable vector independently into host cells, a transposable immunoglobulin light chain construct with different inverted terminal repeat (ITR) sequences can be constructed that are recognized by the *Sleeping Beauty* transposase.

[00149] For this, the human Ig-kappa light chain expression vector pIRES-EGFP-T1T2-IgL (Seq-ID-5) of example 1 can be used to replace the 5' and 3' ITRs of the *PiggyBac* transposon system, contained in this vector, with the 5' and 3' ITRs of the *Sleeping Beauty* transposon system. The sequences for the *Sleeping Beauty* 5'ITR and 3'ITR, recognized and functional with the *Sleeping Beauty* transposase, can be retrieved from patent document US7160682B1/US2003154500A1.

[00150] The upstream *Sleeping beauty* ITR sequence with the 5' terminal repeat has to be gene synthesized with flanking MunI restriction enzyme sequences, allowing the replacement of the MunI flanked *PiggyBac* 5'ITR in construct pIRES-EGFP-T1T2-IgL (Seq-ID-5) of example 1 by the *Sleeping Beauty* 5'ITR sequence. This sequence is provided as Seq-ID14 below, at the end of this Example.

[00151] The downstream *Sleeping beauty* ITR sequence with the 3' terminal repeat (also published in US7160682B1/US2003154500A1) has to be gene synthesized with flanking XhoI restriction enzyme sequences, allowing the replacement of the XhoI flanked *PiggyBac* 3'ITR in construct pIRES-EGFP-T1T2-IgL (Seq-ID-5) of example 1 by the *Sleeping beauty* 3'ITR sequence. This sequence is as provided in Seq-ID15 below, at the end of this Example (XhoI restriction enzyme sites are highlighted in boldface print and 4 additional flanking random nucleotides, allowing proper restriction enzyme digestion of the gene synthesized fragment, are indicated in lowercase letters):

[00152] In a first step, the MunI-flanked *PiggyBac* 5'ITR of construct pIRES-EGFP-T1T2-IgL (Seq-ID-5) has to be replaced by the *Sleeping Beauty* 5'ITR by digesting pIRES-EGFP-T1T2-IgL (Seq-ID-5) with MunI restriction enzyme and by ligating the MunI digested gene-synthesized fragment from Seq-ID-14 into the MunI linearized vector backbone of pIRES-EGFP-T1T2-IgL (Seq-ID-5). The correct orientation of *Sleeping Beauty* 5'ITR can be checked by diagnostic restriction enzyme digestions and/or DNA sequencing. The resulting plasmid is called pIRES-EGFP-sbT1-pbT2-IgL (Fig. 8).

[00153] In a second step, the XhoI-flanked *PiggyBac* 3'ITR of construct still contained in pIRES-EGFP-sbT1-pbT2-IgL has to be replaced by the *Sleeping Beauty* 3'ITR by digesting pIRES-EGFP-sbT1-pbT2-IgL with XhoI restriction enzyme and by ligating the XhoI digested gene-synthesized fragment from Seq-ID-15 into the XhoI linearized vector backbone of pIRES-EGFP-sbT1-pbT2-IgL. The correct orientation of *Sleeping Beauty* 3'ITR can be checked by diagnostic restriction enzyme digestions and/or DNA sequencing. The resulting plasmid is called pIRES-EGFP-sbT1T2-IgL (Fig. 8).

[00154] The entire sequence of the Ig-kappa LC expression vector pIRES-EGFP-sbT1T2-IgL transposable by the *Sleeping Beauty* transposase is provided in Seq-ID-16 below, at the end of this Example.

[00155] Sequences referred to in this Example 3:

[00156] Seq-ID14 (246 bp long DNA fragment containing the 5'ITR of the *Sleeping Beauty* transposon system. Flanking MunI restriction enzyme sites are printed in boldface and underlined)

[00157] Seq-ID15 (248 bp long DNA fragment containing the 3'ITR of the *Sleeping Beauty* transposon system)

[00158] Seq-ID16 (6339bp long DNA sequence of *Sleeping Beauty* transposable Ig-kappa-LC expression vector pIRES-EGFP-sbT1T2-IgL)

Example 4: Cloning of *PiggyBac* and *Sleeping Beauty* transposable vectors for membrane bound human IgG_1

[00159] In addition to the cloning instructions for basic *PiggyBac* transposable IgH and IgL expression vectors provided in Examples 1 and 2, and the construction of a basic *Sleeping Beauty* transposable IgL expression vector provided in Example 3, additionally cloning of improved *PiggyBac* and *Sleeping Beauty* transposable IgH and IgL expresison vectors for a chimeric anti-humanCD30 mAb and for a humanized anti-humanCD19 mAb has been performed, in order to reduce the invention to practice.

[00160] For this, in a first step, the following two gene fragments have been synthesized (commissioned to Genscript, Piscataway, NJ, USA):

[00161] 1.) A 4975 bp DNA fragment containing an expression cassette, in which the expression of a human membrane bound IgG_1 heavy chain is driven by the EF1-alpha promoter (basepairs 1-1335 of Clontech expression vector pEF1-alpha-IRES, Cat-No. #631970), and in which the

expression of Ig chains is linked to EGFP expression via an internal ribosomal entry site (IRES). The DNA sequences for the IRES and EGFP regions were derived from pIRES-EGFP (basepairs 1299-1884 and 1905-2621, respectively, of Clontech expression vector pIRES-EGFP (Cat.-No. #6064-1, Life Technologies). In addition, the synthesized DNA fragment contained a chimeric intron positioned in between the Ig constant coding region and the IRES sequence, whose sequence was derived from pCI mammalian expression vector (basepairs 857-989. od Promega, Cat.-No. #E1731). At the 3' end of the expression cassette the synthesized fragment contained a bovine growth hormone polyadenylation signal (BGH-polyA), whose sequence was derived from pCDNA3.1-hygro(+) expression vector (basepairs 1021-1235 of Invitrogen-Life Technologies, Cat.-No. #V870-20). The expression cassette was flanked up-and downstream by *PiggyBac* transposon ITRs already disclosed in Seq-ID1 and Seq-ID2 further above.

[00162] A map of the elements and their arrangement in the gene-synthesized DNA fragment is provided in Fig. 10, including additionally added unique restriction enzyme sites that can be used to excise or to replace any of the functional elements of the expression cassette.

[00163] The sequence of the 4975 bp long gene-synthesized fragment is provided as Seq-ID20 below, at the end of this Example.

[00164] It shall be noted here that the gene synthesized expression cassette for human IgH chains provided in Seq-ID20, on purpose, did not yet contain the coding region for a V_H domain, such that the construct can be used for the insertion of any desired V_H coding region and/or V_H coding gene library using unique restriction enzyme sites NotI and NheI. This construct therefore is designated "empty" Ig-gamma1-HC expression cassette.

[00165] 2.) In order to provide a plasmid backbone for the transposable expression cassette of Seq-ID20, a 2774 bp long DNA fragment had been gene synthesized (performed by Genscript, Piscataway, NJ, USA) that contained a bacterial ColE1 ori and an ampicillin resistance gene. The sequence information for these plasmid backbone components were derived from the plasmid backbone of expression vector pCI (Promega, Cat.-No. #E1731). The synthetic gene fragment additionally contained 5' and 3' ITRs of the *Sleeping Beauty* transposon, already disclosed in Seq-ID14 and Seq-ID15, respectively. This fragment needed to be circularized and could be propagated in *E.coli* as an autonomous plasmid, due to the presence of the ColE1 ori and the ampicillin resistence gene.

[00166] A map of the elements and their arrangement in the gene-synthesized DNA fragment is provided in Fig. 10, including position of additionally added unique restriction enzyme sites that can be used to excise or to replace any of the functional elements of the expression vector.

[00167] The sequence of the 2774 bp long gene-synthesized fragment is provided as Seq-ID21 below at the end of this Example:

[00168] These two gene fragments allowed the construction of both *PiggyBac* and *Sleeping Beauty* transposable vectors by ligating fragments from these vectors, upon digestion with different restriction enzymes, followed by ligation, as follows:

[00169] The *PiggyBac* transposable vector was cloned by ligating EcoRI-ClaI fragments from Seq-ID20 and Seq-ID21, such that the resulting construct contains the entire *PiggyBac* ITR-flanked expression cassette of Seq-ID20, and the ColE1-amp containing backbone without the *Sleeping Beauty* ITRs of Sq-ID21. Conversely, the ligation of XbaI-MluI fragments from Seq-ID20 and Seq-ID21 resulted in the ligation of the expression cassette without the *PiggyBac* ITRs into the linearized plasmid backbone of Seq-ID21 still containing the *Sleeping Beauty* ITRs. Miniprep plasmids resulting from the two ligations were analyzed by diagnostic restiction enzyme digestions using a mixture of XhoI-NheI-BamHI and in addition with PvuI restriction enzymes, in order to identify correctly ligated plasmids. One selected DNA clone of each ligation was retransformed into *E.coli* to generate a DNA maxiprep, which was verified by DNA sequencing using sequencing primers allowing sequencing of the entire plasmid sequence.

[00170] The entire sequences of *PiggyBac* and *Sleeping beauty* transposable vectors (containing the "empty" human gamma1-HC expression cassette) generated as described above and verified by DNA sequencing is provided as Seq-ID22 (*PiggyBac* transposable vector) and Seq-ID23 (*Sleeping Beauty* transposable vector) below, at the end of this Example.

[00171] V_H and V_L coding regions of chimeric anti-human CD30 antibody brentuximab (clone Ac10) could be retrieved from sequences 1 and 9 of patent application US2008213289A1, and are provided below as Seq-ID24 and Seq-ID25, respectively.

[00172] V_H and V_L coding regions of humanized anti-human CD19 antibody hBU12 were retrieved from patent document US 8,242,252 B2 as sequence variants HF and LG, respectively, and are provided in Seq-ID26 and Seq-ID27 further below, at the end of the Example.

[00173] In order to allow construction of final *PiggyBac* and *Sleeping Beauty* transposable anti-CD30 and anti-CD19 IgHC expression vectors, the DNA fragments for the V_H domains were designed to have flanking NheI and NotI restriction enzyme sites. The nucleotide sequence encoding the V_H of anti-CD30 antibody brentuximab (clone Ac10) has additionally been modified to also contain a leader sequence for mammalian cell expression. The DNA sequences of the NotI-NheI fragments encoding the V_H of anti-CD30 and anti-CD19 mAbs are provided in Seq-ID28 and Seq-ID29 at the end of this Example. The DNA fragments had

been gene synthesized by GeneArt, Regensburg, Germany (NotI and NheI sites are underlined).

[00174] In order to generate anti-CD30 and anti-CD19 IgH chain expression vectors that are transposable with either *PiggyBac* or *Sleeping Beauty* transposase, the NotI-NheI digested fragments of Seq-ID28 Seq-ID29 were ligated into NotI-NheI linearized vectors disclosed in Seq-ID22 or SeqID23, respectively. This resulted in the generation of four vectors containing a fully functional heavy chain (HC) of anti-CD30 mAb brentuximab (clone Ac10) and of anti-CD19 mAb hBU12 and the constructs were designated: pPB-EGFP-HC-Ac10, pPB-EGFP-HC-hBU12, pSB-EGFP-HC-Ac10, and pSB-EGFP-HC-hBU12 and their vector maps are provided in Fig. 11. These vectors have specifically been designed to allow surface expression of the heavy chains, and, upon co-exprssion of light chains, surface IgG expression. However, simple omission of the coding region of for the membrane spanning region of teh Ig heavy chains would result in transposable expression vectors for secreted IgG.

[00175] In order to generate anti-CD30 and anti-CD19 IgL chain expression vectors that are transposable with either *PiggyBac* or *Sleeping Beauty* transposase, the IgH constant region genes from the vectors disclosed in Seq-ID22 and Seq-ID23 needed to be replaced with IgL chain coding regions of anti-CD30 and anti-CD19 antibodies. This was achieved by gene synthesizing gene fragments containing the V_L coding regions as disclosed in Seq-ID25 and Seq-ID27 fused in-frame to a human constant kappa light chain coding region, with a leader sequence at the 5' end and flanked by NotI-BstBI cloning sites that allow the ligation of the NotI-BstBI digested fragment into NotI-BstBI linearized vectors disclosed in Seq-ID22 and Seq-ID23, thereby replacing the IgH constant coding region of Seq-ID22 and Seq-ID23 with the IgL coding regions of anti-CD30 mAb Ac10 and anti-CD19 mAb hBU12.

[00176] The gene-fragments containing the IgL coding regions of anti-CD30 mAb Ac10 and anti-CD19 mAb hBU12, with leader sequence and flanked by NotI-BstBI cloning sites is disclosed in Seq-ID30 and Seq-ID31 below, at the end of the Example. The gene synthesis of these DNA fragments was performed by Genscript (Piscataway, NJ, USA).

[00177] In order to generate anti-CD30 and anti-CD19 IgL chain expression vectors that are transposable with either *PiggyBac* or *Sleeping Beauty* transposase, NotI-BstBI digested fragments of Seq-ID30 and Seq-ID31 had been ligated into NotI-BstBI linearized vectors disclosed in Seq-ID22 or SeqID23. The resulting four vectors were called: pPB-EGFP-LC-Ac10, pPB-EGFP-LC-hBU12, pSB-EGFP-LC-Ac10, and pSB-EGFP-LC-hBU12 and their vector maps are provided in Fig. 11.

[00178] Complete sequences of the *PiggyBac* and *Sleeping beauty* anti-CD30 and anti-CD19 IgH and IgL constructs (eight combinations) are provided in Seq-ID-32 (pPB-EGFP-HC-Ac10), Seq-

ID-33 (pPB-EGFP-HC-hBU12), Seq-ID-34 (pSB-EGFP-HC-Ac10), Seq-ID-35 (pSB-EGFP-HC-hBU12), and in Seq-ID-36 (pPB-EGFP-LC-Ac10), Seq-ID-37 (pPB-EGFP-LC-hBU12), Seq-ID-38 pSB-EGFP-LC-Ac10), and Seq-ID-39 (pSB-EGFP-LC-hBU12) below, at the end of this Example

[00179] Sequences referred to in this Example:

[00180] Seq-ID20 (4975 bp long DNA sequence containing a *PiggyBac* ITR-flanked expression cassette for membrane spanning human Ig-gamma1 heavy chains)

[00181] Seq-ID21 (2774 bp long DNA sequence containing vector backbone components ColE1 and ampicillin resistance flanked by 5' and 3' ITRs of *Sleeping Beauty*)

[00182] Seq-ID22 (7242 bp long sequence of *PiggyBac* transposable "empty" human gamma1-HC vector)

[00183] Seq-ID23 (7146 bp long sequence of *Sleeping Beauty* transposable "empty" human gamma1-HC vector)

[00184] SeqID-24 (351 bp long V_H coding region of anti-human CD30 antibody brentuximab)

[00185] Seq-ID25 (333 bp long V_L coding region of anti-human CD30 antibody brentuximab)

[00186] Seq-ID26 (417 bp long V_H coding region of anti-human CD19 mAb huB12, including leader)

[00187] Seq-ID27 (375 bp long V_L coding region of anti-human CD19 mAb huB12, including leader)

[00188] Seq-ID28 (423 bp long DNA fragment, containing NotI-NheI-flanked V_H coding region of the V_H domain of anti-human CD30 mAb brentuximab)

[00189] Seq-ID29 (432 bp long DNA fragment, containing NotI-NheI-flanked V_H coding region of the V_H domain of anti-human CD19 mAb hBU12)

[00190] Seq-ID30 (733 bp long DNA fragment containing IgL coding region of anti-CD30 mAb Ac10 and flanked by NotI and BstBI restriction enzyme sites)

[00191] Seq-ID31 (718 bp long DNA fragment containing IgL coding region of anti-CD19 mAb hBU12 and flanked by NotI and BstBI restriction enzyme sites)

[00192] Seq-ID32 (7645 bp sequence of pPB-EGFP-HC-Ac10

[00193] Seq-ID33 (7654 bp sequence of pPB-EGFP-HC-hBU12)

[00194] Seq-ID34 (7549 bp sequence of pSB-EGFP-HC-Ac10)

[00195] Seq-ID35 (7558 bp sequence of pSB-EGFP-HC-hBU12)

[00196] Seq-ID36 (6742 bp long sequence of pPB-EGFP-LC-Ac10)

[00197] Seq-ID37 (6727 bp long sequence of pPB-EGFP-LC-hBU12)

[00198] Seq-ID38 (6646 bp long sequence of pSB-EGFP-LC-Ac10)

[00199] Seq-ID39 (6631 bp long sequence of pSB-EGFP-LC-hBU12)

Example 5: Instructions for cloning of a *PiggyBac* transposase expression vector

[00200] The ORF of functional *PiggyBac* transposase enzyme can be retrieved from US Patent US 7,105,343 B1(incorporated herein by reference in its entirety) and is provided in Seq-ID11 below, at the end of this Example. The DNA sequence of Seq-ID11 translates into the amino acid Seq-ID12 also provided at the end of this Example.

[00201] In order to generate a vertebrate cell expression vector for the *PiggyBac* transposase enzyme, this ORF can be gene synthesized and cloned as a blunt ended DNA into the unique, blunt-cutting restriction enzyme site EcoRV in the standard vertebrate cell expression vector pCDNA3.1-hygro(+) (catalogue # V870-20, Invitrogen, Carlsbad, CA, USA), by methods know in the art. The correct ligation of the *PiggyBac* ORF, relative to the pCDNA3 promoter can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned *PiggyBac* expression construct pCDNA3.1-hygro(+)-PB (Fig. 7). The construction of a *PiggyBac* expresison vector was performed as described herein and the vector design was verified by diagnostic restriction enzyme digestion, and DNA sequencing.

[00202] The sequence of the *PiggyBac* expression construct pCDNA3.1-hygro(+)-PB is provided as Seq-ID 13, below at the end of this Example

[00203] Sequences referred to in this Example 5:

[00204] Seq-ID11 (ORF of *PiggyBac* transposase)

[00205] Seq-ID12 (amino acid sequence of *PiggyBac* transposase)

[00206] Seq-ID13 (pCDNA3.1-hygro(+)-*PiggyBac* expression vector)

Example 6: Instructions for cloning of a *Sleeping Beauty* transposase expression vector

[00207] The open reading frame (ORF) of the *Sleeping Beauty* transposase enzyme can be found in patent reference US7160682B1/US2003154500A1. The sequence is provided in Seq-ID17, below at the end of this Example. This DNA sequence of Seq-ID17 translates into the amino acid sequence of Seq-ID18, also provided at the end of this Example, further below.

[00208] In order to generate a vertebrate cell expression vector for the *Sleeping Beauty* transposase enzyme, this ORF can be gene synthesized and cloned as a blunt ended DNA into the unique, blunt-cutting restriction enzyme site EcoRV in the standard vertebrate cell expression vector

pCDNA3.1-hygro(+) (catalogue # V870-20, Invitrogen, Carlsbad, CA, USA), by methods know in the art. The correct ligation of the *Sleeping Beauty* ORF, relative to the pCDNA3 promoter can be verified by diagnostic restriction enzyme digestion, and/or by DNA sequencing of the cloned *Sleeping Beauty* expression construct pCDNA3.1-hygro(+)-SB (Fig. 9).

[00209] The sequence of the *Sleeping Beauty* expression construct pCDNA3.1-hygro(+)-SB is provided in Seq-ID19, below, at the end of this Example.

[00210] The construction of a *Sleeping Beauty* expression vector was performed as described herein and the vector design was verified by diagnostic restriction enzyme digestion, and DNA sequencing. The coding regions for *PiggyBac* and *Sleeping Beauty* transposase enzymes had been gene synthesized by Genscript, Piscataway, NJ. With the eight different transposable IgH and IgL expression vectors for *PiggyBac* and *Sleeping Beauty* transposases, and the pCDNA3.1-hygro(+) expression vectors for *PiggyBac* and *Sleeping Beauty* transposase enzymes, all vectors have been generated that allow the expression of anti-CD30 and anti-CD19 antibodies on the cell surface of mammalian cells.

[00211] Sequences referred to in this Example 6:

[00212] Seq-ID17 (ORF of Sleeping Beauty transposase enzyme)

[00213] Seq-ID18 (amino acid sequence of Sleeping Beauty transposase)

[00214] Seq-ID19 (DNA sequence of Sleeping Beauty expression vector pCDNA3.1-hygro(+)-SB)

Example 7: Generation of murine preB cells stably expressing membrane bound human IgG from stably transposed expression vectors

[00215] In order to demonstrate stable expression of human IgG antibodies in mammalian cells, transposable human IgH and IgL expression constructs have been transfected into Abelson murine leukemia virus (A-MuLV) transformed proB cell line 63-12, originally derived from RAG-2 deficient mice (Shinkai et al. (1992) Cell 68, 855-867) and therefore unable to initiate V(D)J recombination. This host cell line represents a B cell lineage lymphocyte cell type that expresses all cellular components for optimal membrane bound antibody expression, including the B cell receptor co-factors Ig-alpha (CD79a or mb-1) and Ig-beta (CD79b or B29) that interact with the transmembrane spanning amino acids of membrane bound immunoglobulin. Therefore, these cells optimally anchor IgG molecules with a trans-membrane spanning region in the cell surface membrane. 63-12 cells were grown in in static culture in suspension using

IMDM medium supplemented with 2 % FCS, 0.03% Primatone™ RL-UF (Sheffield Bioscience), 2 mM L-glutamine, 50µM 2-mercaptoethanol, at 37°C in a humidified incubator and a 10% CO_2 atmosphere. For the co-transfection of the transposable IgH and IgL expression vectors (Example 4) with a transposon expression vector (Examples 5 or 6), the cells were passaged 24 hours prior to transfection and seeded at a density of $5x10^5$ cells/ml, in order to allow the cells to enter into log-phase growth until the time-point of transfection.

[00216] At the day of transfection, 63-12 cells were harvested by centrifugation and resuspended in RPMI 1640 medium without any supplements or serum at a density of $5x10^6$ cells/ml. 400µl of this cell suspension (corresponding to $2x10^6$ cells) were transferred into 0.4 cm electroporation cuvettes (BioRad order #165-2081) and mixed with 400µl of RPMI 1640 medium containing the desired plasmid DNA (or a mixture of plasmids). Cells were then transfected using a BioRad Gene Pulser II at 950µF/300V settings and incubated for 5 min at room temperature after a single electroporation pulse. After this, the cells were transferred into 5 ml IMDM-based growth medium and the cells were centrifuged once, in order to remove cell debris and DNA from the electroporation, before the cells were transferred into IMDM-based growth medium for recovery and expression of proteins from transfected plasmids.

[00217] The electroporation settings have been determined as the most optimal transfection conditions for A-MuLV transformed proB cell line 63-12 that routinely resulted in transient transfection efficiencies ranging between 30-40%. The result of such a transfection by electroporation is documented in the FACS analyses depicted in Fig. 12, where the transfection controls, two days post transfection, are depicted on the left column panels. The negative control (labeled NC), that was mock-electroporated without DNA, as expected, does not show any green fluorescent cells, whereas the transfection control that was transfected with 15µg pEGFP-N3 plasmid (Clontech, order #6080-1), showed that 38.8% of the cells were transiently transfected, as detected by cells expressing enhanced green fluorescent protein (see cell in lower right quadrant). As expected, the transfection controls do not show any Ig-kappa signal, because none of the transfection controls was transfected with an Ig-expression construct.

[00218] For transposition of IgH and IgL expresison vectors, 63-12 cells were also transfected by electroporation with a mix of 5µg each of a transposable IgH expression vector, 5µg of a transposable IgL expression vector, and 5µg of an expression vector allowing expression of a transposase mediating the transposition of the IgH and IgL expression vector. The result of this transfection is also shown in Fig. 12.

[00219] Expression of human IgG on the surface of the cells was detected by a biotinylated anti-human kappa light chain specific antibody (Affymetrix, ebioscience, order #13-9970-82) detected with streptavidin-allophycocyanin (strep-APC) (Affymetrix, ebioscience, order #17-

4317-82), and is shown on the Y-axis of the FACS dot-plots. As can be seen in Fig. 10, the measurements depicted in the second column from left show the analysis of cells transfected with IgH+IgL+transposase expression vectors two days after electroporation (labeled "d2 post TF"). The FACS analysis after two days of transfection showed that between 1.8% and 2.8% of the cells express human IgH+IgL on the cell surface, because IgL expression on the cell surface can only be detected, if IgH chains are co-expressed in the cells, such that a complete IgG can be expressed on the surface of the cells.

[00220] From this data it can be inferred that if ca. 38% of the cells are transiently transfected, ca. 5-7.5% of these cells have been co-transfected with both IgH and IgL expression vectors. From this experiment it is concluded that the transposable IgH and IgL expression constructs allow high-level expression of human IgG on the surface of murine A-MuLV transformed proB cells, which is comparable to surface IgG signals obtained by staining of human peripheral B lymphocytes with the same antibody staining reagents (data not shown).

[00221] As expected the cells showing IgG expression also displayed EGFP expression, because the EGFP expression was transcriptionally coupled to IgH or IgL expression via IRES sequences. However, the EGFP expression was significantly lower, as compared to the EGFP expression from the pEGFP-N3 control plasmid, which is expected, as the EGFP expression in pEGFP-N3 is directly driven by a strong constitutive promoter, whereas EGFP expression in the transposable IgH and IgL expression vectors is effected by transcriptional coupling to the IgH and IgL coding region using an internal ribosomal entry site (IRES). Nevertheless, as expected, cells displaying higher IgG expression also displayed higher EGFP signals (leading to a slightly diagonal Ig-kappa$^+$/EGFP$^+$ population), which clearly demonstrates, that both expression levels are coupled.

[00222] When cells were analyzed without cell sorting after one week of transfection, the EGFP signal in pEGFP-N3 control transfections was no longer detectable (data not shown), showing that the cells do not stably integrate expression constructs at any significant frequency. In contrast, a low ca. 1-2% IgG-EGFP double-positive population of cells was maintained in cells that have been co-transfected with transposable IgH&IgL vectors together with a transposase expression vector, already indicating that ca. 3-6% of the cell transiently transfected cells stably integrate simultaneously the transposable IgH and IgL expression vectors into their genome (data not shown).

[00223] In order to enrich for these stably transposed cells, Ig-kappa light chain and EGFP double positive cells have been FACS sorted at day 2 post transfection, as indicated by the sorting gates (black rectangles in the second left column FACS dot plots). Each 5'000 cells falling into this gate have been sorted from the *PiggyBac* transpositions with IgH&IgL of anti-CD30

mAb Ac10 (top row), and of anti-CD19 mAb hBU12 (middle row), and 3'000 cells have been sorted from the *Sleeping Beauty* transposition with IgH&IgL of anti-CD30 mAb Ac10 (bottom row), as indicated in Fig. 12..

[00224] The FACS-sorted cells were expanded for one week (representing day 9 post transfection), and were re-analyzed again for surface IgG expression by detection of IgG with an anti-kappa light chain antibody, as described above. As can be seen in Fig. 10 (second column from the right), over 30%, 50% and 5% of the one-time sorted cells stably expressed IgG on the cell surface, while these cells, as expected, were also EGFP-positive. This demonstrates that a significant percentage of transiently IgH & IgL co-transfected cells stably maintain IgG expression. The *PiggyBac* mediated transpositions in this experimental set-up appear to have occurred with about 6-10-fold higher efficiency than the *Sleeping Beauty* mediated transpositions.

[00225] A couple of additional conclusions can be drawn from this transposition experiment: First, from the IgG/EGFP double positive cells sorted on day two post transfection, about 75% of cells remained stably EGFP+ in the *PiggyBac* transpositions, as ca. 40% and 30% of the PB-Ac10 and PB-hBU12 transposition generate EGFP+ cells lacking IgG surface expression. These cells most likely have stably transposed only one of the two transposable IgH and IgL expression vectors, which does not allow for surface IgG expression, but sufficient to render the cells EGFP-positive. This also means that from the ca. 38% originally transiently transfected cells, at least 5% are stably transposed with at least one transposable Ig expression vector.

[00226] The numbers of stably transposed cells for the *Sleeping Beauty* transposition were lower, than those of the *PiggyBac* transposition, and after a first round of FACS sorting of IgH+IgL+transposase co-transfected cells, only about 5% of stably IgG expressing cells was obtained. However, if also the stably EGFP positive cells are considered, about 9% of stably transposed cells were obtained after the first FACS sorting cycle using Sleeping Beauty transposase.

[00227] When these stably IgG-positive and IgH/IgL transposed cells cells were FACS sorted again, over 99% stably IgG expressing cells were obtained (Fig. 12, rightmost column), and the stable expression phenotype was maintained for over four weeks, without any change in the percentage of IgG+ cells (data not shown). Therefore, it is concluded that the transposable expression vectors for human IgH and IgL chains, as disclosed in this invention, are functional and can stably be integrated into a mammalian host cell genome with high efficiency.

Example 8: Enrichment of stably IgG transposed and IgG expressing cells via specific antigen binding

[00228] In order to demonstrate that human IgG expressing proB cells, generated by transposition of IgH and IgL expression vectors, as disclosed in Example 7 above, can be used for the isolation of antigen-specific cells, decreasing numbers of the proB cell line 63-12, expressing anti-CD30 mAb (see d16, 2x sorted, 63-12+PB-Ac10, of Fig. 12) were mixed with proB cell line 63-12 expressing anti-CD19 mAb (see d16, 2x sorted, 63-12+PB-hBU12, of Fig. 12) at ratios 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6} (see Fig. 13). A total of 10^{7} cells were stained in 1 ml PBS, supplemented with 2%FCS, for 30 min on ice, with the following reagents:

- 0.1µg 6xHis-tagged, recombinant human CD30 (Sino Biological Inc., Bejing, China, order #10777-H08H), and
- 10µl mouse anti-human Ig-kappaLC-APC labeled antibody (Life Technologies, Invitrogen, order #MH10515).

[00229] After these primary reagents were removed from the cells, by centrifugation and washing in PBS, 2% FCS, a secondary staining was performed in 1 ml PBS, supplemented with 2%FCS, for 30 min on ice, with:

- 0.1 µg biotinylated anti-His-tag antibody (IBA Life Sciences, Göttingen, Germany, order #2-1590-001)
- -After this secondary reagent was removed from the cells, by centrifugation and washing in PBS, 2% FCS, the CD30-6xHis/anti-His-tag-bio combination was detected by staining in 1 ml PBS, 2% FCS, for 30 min on ice, with:
- 1/500 diluted streptavidin-Phycoerythrin (strep-PE) (Affymetrix ebioscience, order #12-4317-87) reagent.

[00230] After the final FACS staining, the cells were again washed twice in ice-cold PBS, 2%FCS, and then resuspended in 1.0 ml PBS, 2% FCS, after which the cells were subjected to FACS analysis and cell sorting of Ig-kappaLC/CD30 positive cells (see Fig. 13).

[00231] As can be seen from the results disclosed in Fig. 13, a specific population of IgG positive and anti-CD30 reactive cells is detectable in the upper-right quadrant of the FACS-plots of the positive control, and as expected the intensity of the FACS signal for surface IgG (detected via anti-Ig-kappa-APC) correlates with the FACS signal for anti-CD30 resulting in a diagonal staining pattern for this population.

[00232] In the mixtures of anti-CD30 mAb expressing cells with anti-CD19 mAb expressing cells, the level of dilution of the specific anti-CD30 mAb expressing cells is very well reflected by the frequency of CD30 antigen specific cells in the upper right quadrant and the stringently

defined FACS sorting gate (black square in the upper right quadrant). The very rare events corresponding to CD30-reactive/IgG positive cells upon increased dilution of the specific cells (1:10'000, 1:100'000, 1:1'000'000) are hardly visible on the printouts of the FACS-dot-plots, even, if increasing numbers of events were acquired, as indicated above the individual dot plots. However, the frequency of CD30 detectable cells correlated well with their frequency as expected from the dilution factor. From this result it is concluded that the display and antigen-specific detection of cells expressing an antigen-specific antibody by means of transposition mediated human IgG expression on the surface of mammalian and proB cells, as shown here, can reliably be performed.

[00233] The bottom row of FACS dot-plots in Fig.13 shows the re-analysis of the FACS sorted cells from the different spike-in dilution experiments. As can be seen from the results, the re-analysis of the cells FACS sorted from the 1:100, 1:1'000 and 1:10'000 dilution resulted in almost the same cell population being enriched, which showing ca. 90-95% antigen reactive cells. The FACS sorted cells from the 1:100'000 dilution contained a small, additional population that did not fall into the gate of IgG-positive/CD30 reactive cells, but also in this experiment ca. 85% of the FACS sorted cells were antigen-specific IgG-expressing cells. Surprisingly, the highest purity of cells, with regard to CD30-reactivity and IgG expression, resulted from the FACS sort, in which only 1 in 1'000'000 had been CD30 antigen specific, and where only ca. 14 cells had been sorted. This can only be explained that her almost clonal effects need to be considered such that the sort was not a mixture from IgG-positive-CD30 reactive cells, but rather a few clones that all represented IgG-positive-CD30 reactive cell clones.

[00234] Nevertheless, the results of the specific antigen-mediated staining and identification of antigen-specific antibody expressing cells and their successful enrichment by preparative FACS-mediated cell sorting clearly demonstrates the feasibility of the method disclosed herein for the isolation of cells expressing antibodies with a desired binding phenotype.

Example 9: Instruction for the generation and use of transposable IgH expression vectors that can be used to switch from membrane bound to secreted antibody expression

[00235] The transposable Ig expression vectors disclosed in Examples 1 to 4 only allow expression of human IgG on the surface of mammalian cells, such that the binding phenotype of antibodies can readily be identified and enriched for by antigen-binding to the cells, by means of FACS,

as exemplified in Example 8, or by cell-panning or batch enrichment methods (e.g. magnetic bead activated cell sorting, MACS). However, it is often desired to rapidly analyze the antigen-binding properties of a given antibody displayed by a cell also as a secreted antibody in solution. While it is possible to PCR-amplify the relevant V_H and V_L coding regions of an antigen-specific cell clone into expression vectors for secreted IgG expression, this approach is time consuming and labour intensive.

[00236] In the detailed description of the invention it is already disclosed that transposable IgH expression constructs can be employed that exploit the natural "switch" from membrane bound to secreted antibody expression, based on alternative splicing of genomic IgH chain constructs.

[00237] This switch from membrane bound to secreted antibody expression can be achieved as follows:

[00238] Instead of a cDNA-based expression cassette for human Ig-gamma1 heavy chains, the original genomic organization of human Ig-gamma1 gene locus needs to be cloned into the IgH expression vectors as disclosed before in Example 4. The sequence of the entire immunoglobulin gene locus in germline configuration can be retrieved from contig NT_010168 of the human genome project, which covers the human Ig heavy chain locus located on chromosome 14. The human Ig-gamma1 heavy chain gene locus starting from the first amino acid of the C_H1 domain at the 5' end to 500 bp downstream of the last stop codon of the second membrane-spanning exon gamma1-M2 at the 3' end has a length of 5807 base pairs and displays no internal NheI or BstBI sites. Therefore, this gene locus can be synthesized with flanking NheI and BstBI sites, that can be used for directional cloning. Such a gene synthesized fragment can then directly be used to replace the cDNA coding region of the membrane-bound gamma1-constant coding region in pPB-EGFP-HC-Ac10 (Seq-ID32)

[00239] The DNA sequence of a genomic human Ig-gamma1 fragment to be synthesized is provided in Seq-ID40 below, at the end of this Example (5'-NheI and 3'-BstBI sites are highlightes in boldface print).

[00240] The organization of the exon and introns of the human Ig-gamma1-heavy chain germline locus, including their membrane spanning exons M1 and M2 is depicted in Fig. 14. The coding and non-coding regions in this genomic gene fragment left in its original genomic configuration are supposed to contain all required *cis*-regulatory elements to allow alternative splicing of an Ig-gamma1 mRNA depending on the differentiation stage of the B-lineage cells, in which the mRNA is processed (Peterson et al. (2002) Mol. Cell. Biol. 22, 5606-5615). The cloning of fragment Seq-ID40 into a transposable Ig-gamma1 HC expression vector can be performed by replacing the C-gamma1 coding region in pPB-EGFP-HC-Ac10 by digesting

pPB-EGFP-HC-Ac10 with NheI and BstBI restriction enzymes and ligating the genomic fragment of Seq-ID40 as a NheI-BstBI digested fragment into the NheI-BstBI linearized vector fragment of pPB-EGFP-HC-Ac10.

[00241] The result of this ligation is shown schematically in Fig. 14, and the sequence of the construct is provided in Seq-ID41 below at the end of this Example.

[00242] A-MuLV transformed proB cells, like 63-12 cells, represent a suitable cell line to exploit the natural mechanism of alternative splicing of a genomic Ig-gamma1 HC construct, as it is possible to effect phenotypic differentiation of these cells to more mature B-lineage cells, if the transforming activity of the Abl-kinase is inhibited. This can specifically be achieved with the Abl-kinase inhibitor Gleevec (also known as Imatinib, or STI-571) (Muljo and Schlissel (2003) Nature Immunol. 4, 31-37). However, if A-MuLV transformed proB cells are treated with Gleevec, they not only initiate phenotypic differentiation to more mature B lineage stages, but this process is also associated with an induction of apoptosis (unpublished observation). This can be prevented by first establishing a 63-12 A-MuLV transformed cell line that is stably transfected with a bcl-2 expression vector.

[00243] The mouse bcl-2 mRNA sequence can be found in NCBI-Genbank entry NM_009741, and has the following sequence Seq-ID42, shown below at the end of this Example. This open reading frame translates into the following amino acid sequence Seq-ID43, also provided below, at the end of the Example.

[00244] In order to generate a mammalian bcl-2 expression vector, the murine bcl-2 coding region can be gene synthesized with flanking KpnI and XhoI restriction enzyme sites that are not present in the coding region of bcl-2 and a KpnI-XhoI double digested gene-synthesized DNA fragment can be ligated into pCDNA3.1-hygro(+) described further above in order to generate a mammalian expression vector for bcl-2 that can stably be transfected into 63-12 cells in order to select for stable bcl-2 transfectants.

[00245] The entire sequence of the pCDNA3.1-hygro(+) expression vector containing the murine bcl-2 gene inserted into the KpnI and XhoI restriction sites of the multiple cloning site is provided as Seq-ID44 below, at the end of this Example.

[00246] In order to facilitate the generation of stable transfectants, this vector can e.g. be linearized outside of the expression cassettes for bcl-2 or hygromycinB using the enzyme FspI, that linearizes the vector in the bacterial ampicillin resistance gene. 20 µg of such a linearized vector can be transfected into $2x10^6$ 63-12 cells by electroporation at 950µF/300V exactly as disclosed further above for the transfection of transposable vectors. Following electroporation, the cells can then be diluted in 100ml growth medium and plated into five 96-well plates with each 200µl/well, which will result in the plating of ca. $4x10^3$ cells/well.

[00247] Stable transfectants can then selected by adding 800µg/ml hygromycinB 48 hours post transfection. Individual stably transfected cell clones, of which 20-100 can be expected from such an experiment, can then be obtained 2-3 weeks later. Stable bcl-2 transfected cell clones are best functionally tested for their ability to protect cells from apoptosis, by measuring the survival of individual clones upon exposure of 0.1 to 1 mM Gleevec (Imatinib, or STI-571). Once a 63-12 stable bcl-2 transfectant is identified that has high resistance to Gleevec (Imatinib, or STI-571), this clone can be utilized as a host cell for expression of human IgG from transposable genomic Ig-gamma1 HC and Ig-kappa LC expression vectors, e.g. utilizing vectors Seq-ID41 and Seq-ID36.

[00248] These vector can be co-transfected with PiggyBac expression vector (Seq-ID13) into the stably Bcl-2 transfected 63-12 cells, and stably transposed and IgG expressing cells can be established as described further above (equivalent to Example 7). Because proB cells represent a differentiation stage, where endogenous immunoglobulin is expressed as membrane bound immunoglobulin, it can be expected that also the Ig-HC expressed from a transposable Ig-gamma1 HC expression vector in genomic configuration will be expressed as membrane bound version.

[00249] However, if the cells are treated with 0.1 to 1 mM Gleevec (Imatinib, or STI-571), the Abl-kinase encoded by the A-MuLV is specifically inhibited, the cells are no longer transformed and continue their intrinsic differentiation program to more mature B cell differentiation stages. *In vitro*, this differentiation is independent of functionally expressed Ig proteins (Grawunder et al. (1995) Int. Immunol. 7, 1915-1925). It has even been shown that *in vitro* differentiation of non-transformed proB cells renders them responsive to T cell derived anti-IL4 and CD40 stimulation, upon which the cells even differentiate into plasma cell stage cells undergoing class-switch recombinationand where they can be fused with myeloma cells to generate hybridomas (Rolink et al. (1996) Immunity 5, 319-330).

[00250] This means that also A-MuLV transformed 63-12 cells, which are rendered resistant to apoptosis by stable expression of bcl-2 can be differentiated into cells of the plasma cell stage upon treatment with Gleevec and simultaneous incubation with 10µg/ml agonistic anti-CD40 antibody, and 100U/ml recombinant IL4, exactly as described in Rolink et al. (1996) Immunity 5, 319-330.

[00251] This treatment will induce a change in the cellular differentiation program, that will change the cellular alternative splicing program from membrane bound IgG expression to secreted IgG expression from an Ig HC expression construct in genomic organization. This will allow the production of secreted antibody from replica plated cell clones identified and isolated by surface display and antigen binding, without the need to re-clone V_H and V_L coding regions

from selected cell clones and without the need to to ligate them into expression vectors for secreted IgG antibodies. This is a functional feature of the vectors that cannot easily be incorporated in most mammalian cell expression system, particularly not into many virus-based expression systems, in which such extended genomic expression vectors cannot easily be inserted.

[00252] Sequences referred to in this Example 9:

[00253] Seq-ID40 (5812 bp long genomic human Ig-gamma1-heavy chain gene)

[00254] Seq-ID41 (transposable Ig-gamma1-HC expression vector in genomic configuration)

[00255] Seq-ID42 (murine bcl-2 coding region)

[00256] Seq-ID43 (amino acid sequence of murine Bcl-2 protein

[00257] Seq-ID44 (pCDNA3.1-hygro(+)-bcl2 mammalian expression vector)

Example 10: Instruction for the generation of vectors encoding basic human antibody heavy and light chain libraries as PiggyBac transposable vectors.

[00258] In order to generate simple transposable DNA libraries encoding human antibody heavy and light chain libraries, only the V_L and V_H regions from transposable vectors pIRES-EGFP-T1T2-IgL of Example 2 and pIRES-EGFP-T1T2-IgH of Example 3, respectively, need to be replaced. This can be done by gene synthesizing human V_H and V_L coding regions flanked by ClaI and Eco47III restriction enzyme sites, and by allowing nucleotide variations in certain HCDR and LCDR positions, as provided in Seq-ID9, which encodes libraries for variable heavy chain domains, and Seq-ID-10, which encodes libraries for variable light chain domains, and which are provided at the end of this Example. Both of these sequences contain a stretch of N-sequences in the HCDR3 (boldface), and LCDR3 (boldface), respectively. Both Seq-ID9 and Seq-ID10 sequences are flanked by ClaI and Eco47III restriction enzymes (underlined), respectively, including four nucleotides flanking the restriction enzyme sites (highlighted in lowercase letters at the ends of the sequence), allowing proper restriction enzyme digestion of the gene-synthesized DNA fragments and directed ligation into ClaI-Eco47III linearized pIRES-EGFP-T1T2-IgH and pIRES-EGFP-T1T2-IgL backbones, respectively.

[00259] This way, diverse transposable DNA libraries, encoding antibody heavy and light chains on separate vectors, in which the expression of the antibody chains are transcriptionally and therefore operably linked to a green fluorescent marker protein can be generated.

[00260] Seq-ID9 (VL domain coding region with variable N-sequence variation at positions encoding LCDR3)

[00261] Seq-ID10 (V_H domain coding region with variable N-sequence variation at positions encoding HCDR3)

Example 11: Instructions for the generation of a basic *Sleeping Beauty* transposable human Ig-kappa light chain expression library

[00262] In order to generate a diverse *Sleeping Beauty* transposable DNA library encoding human antibody light chain libraries, the V_L region of *Sleeping Beauty* transposable vector pIRES-EGFP-sbT1T2-IgL of Example 5 needs to be replaced with a diverse V_L gene repertoire. This can be done by gene synthesizing of human V_L coding regions flanked by ClaI and Eco47III restriction enzyme sites, and by allowing nucleotide variations in certain HCDR and LCDR positions, as already provided in Seq-ID-10 above. The Seq-ID10 sequence is flanked by ClaI and Eco47III restriction enzymes allowing directed ligation into ClaI-Eco47III linearized pIRES-EGFP-sbT1T2-IgL. This way a *Sleeping Beauty* transposable DNA library encoding diverse human antibody light chain can be generated.

[00263] This way, diverse transposable DNA libraries, encoding antibody heavy and light chains on separate vectors, in which the expression of the antibody chains are transcriptionally and therefore operably linked to a green fluorescent marker protein can be generated.

Example 12: Cloning of a transposable IgL chain expression library

[00264] A V-kappa light chain library with randomized LCDR3 region was constructed as described below. Six amino acid residues were randomized, *i.e.* encoded by the codon NNK (N = any nucleotide; K = T or G), which accomodates each of the 20 amino acids. The library was based on germline human Vkappa1-5 and Jkappa2 gene segments and was randomized between the conserved cysteine at the end of the framework 3 region and the Jkappa2-based framework 4 region as follows: Gln-Gln-$(NNK)_6$-Thr. The sequence and overall design of the kappa light chain library is shown in Figure 15.

[00265] A linear DNA molecule encoding the kappa light chain library was generated by PCR. For this, two templates were generated by total gene synthesis (performed by GenScript, Piscataway, NJ, USA). On one hand, a synthetic construct was generated comprising the Vkappa1-5 gene segment cloned into the EcoRV site of pUC57 (Genscript order #SD1176), pUC57_Vkappa1-5 (SEQ-ID45); on the other hand, a synthetic construct was generated comprising the Jkappa2 gene segment fused to the Ckappa coding region cloned into the EcoRV site of pUC57, pUC57_Jkappa2-Ckappa (SEQ-ID46).

[00266] A first linear DNA comprising the Vkappa1-5 gene segment was PCR amplified from pUC57_Vkappa1-5 using the primers pUC57-1 (5'-CGT TGT AAA ACG ACG GCC AG-3') and LCDR3-B (5'-CTG TTG GCA GTA ATA AGT TGC-3'). A second linear DNA comprising the randomized CDR3 region (Gln-Gln-$(NNK)_6$-Thr), the Jkappa2 gene segment and the Ckappa constant region was amplified from pUC57_Jkappa2-Ckappa using the primers LCDR3-NNK6-F (5'-GCA ACT TAT TAC TGC CAA CAG NNK NNK NNK NNK NNK NNK ACT TTT GGC CAG GGG ACC AAG-3') and pUC57-2 (5'-TCA CAC AGG AAA CAG CTA TG-3'). To prevent introduction of a sequence bias due to priming of the randomized region of the primer LCDR3-NNK6-F on pUC57_Jkappa2-Ckappa, the plasmid was first linearized by digestion with the restriction enzyme ScaI (Figure 17A).

[00267] The resulting DNA molecules (SEQ-ID47 and SEQ-ID48) displayed an overlap of 21bp and were assembled by PCR overlap extension using the primers pUC57-1 and pUC57-2, generating a DNA molecule comprising the kappa light chain library flanked by NotI and AsuII (=BstBI) restriction sites as shown in Figure 15. The PCR amplicon of the V-kappa light chain library was subjected to PCR-fragment sequencing, and the result shown in Fig. 18, demonstrate that indeed the expected diversity was introduced as designed in the positions of the LCDR3, as evidenced by overlapping electropherogram signals in the randomized positions. This PCR fragment was digested with the restriction endonucleases NotI and BstBI (an isoschizomer of AsuII) and cloned into the *PiggyBac*-transposable vector pPB-EGFP_HC-g1 (SEQ-ID049, resulting in a library consisting of 5.2×10^7 independent clones. The size of this library can easily be increased by a factor 10 by scaling up the ligation reaction.

[00268] Light chain libraries incorporating distinct randomization designs, or comprising Vkappa and Jkappa gene segments other than the ones used in this example, can be produced the same way. Likewise, the strategy described here can be employed for the production of Vlambda light chain libraries.

[00269] Sequences referred to in this Example 12:

[00270] SEQ-ID45 (pUC57_Vkappa1-5)

[00271] SEQ-ID46 (pUC57_Jkappa2-C-kappa)

[00272] SEQ-ID47 (Vkappa1-5 PCR product)

[00273] SEQ-ID48 (NNK6-Jkappa2-C-kappa PCR product

Example 13: Cloning of transposable IgH chain expression libraries with variable HCDR3 length

[00274] A human gamma1 heavy chain library with randomized HCDR3 region was constructed as described below. Several amino acid residues were randomized, *i.e.* encoded by the codon NNK (N = any nucleotide; K = T or G), which accomodates each of the 20 amino acids. The library was based on the V_H3-30 and J_H4 gene segments and was randomized between the conserved Cysteine residue at the end of the framework 3 region and the J_H4-based framework 4 region as follows: Ala-Lys/Arg-$(NNK)_n$-Asp-NNK. Various HCDR3 lengths were explored, with n = 4, 6, 8, or 10 (NNK4, NNK6, NNK8, and NNK10 randomization). The sequence and overall design of the gamma heavy chain library is shown in Figure 16.

[00275] A linear DNA molecule encoding the heavy chain variable region (V_H) library was generated by PCR. For this, two templates were generated by total gene synthesis (performed by GenScript, Piscataway, NJ, USA). On one hand, a synthetic construct was generated comprising the V_H3-30 gene segment cloned into the EcoRV site of pUC57, pUC57_V_H3-30 (SEQ-ID49); on the other hand, a synthetic construct was generated comprising the J_H4 gene segment cloned into the EcoRV site of pUC57, pUC57_J_H4 (SEQ-ID50).

[00276] A first linear DNA comprising the V_H3-30 gene segment was PCR amplified from pUC57_V_H3-30 using the primers V_H3-30-F (5'-GAT ATC CAA TGC GGC CGC ATG -3') and HCDR3-B (5'-CGC ACA GTA ATA CAC AGC CGT G -3'). Additional linear DNA molecules comprising the randomized HCDR3 regions Ala-Lys/Arg-$(NNK)_4$-Asp-NNK, Ala-Lys/Arg-$(NNK)_6$-Asp-NNK, Ala-Lys/Arg-$(NNK)_8$-Asp-NNK, or Ala-Lys/Arg-$(NNK)_{10}$-Asp-NNK fused to the J_H4 gene segment were amplified from pUC57_JH4 using, respectively, the primers HCDR3-NNK4-F (5'-CAC GGC TGT GTA TTA CTG TGC GAR GNN KNN KNN KNN KGA CNN KTG GGG CCA AGG AAC CCT GGT C-3'), HCDR3-NNK6-F (5'-CAC GGC TGT GTA TTA CTG TGC GAR GNN KNN KNN KNN KNN KNN KGA CNN KTG GGG CCA AGG AAC CCT GGT C-3'), HCDR3-NNK8-F (5'-CAC GGC TGT GTA TTA CTG TGC GAR GNN KNN KNN KNN KNN KNN KNN KNN KGA CNN KTG GGG CCA AGG AAC CCT GGT C-3'), or HDR3-NNK10-F (5'-CAC GGC TGT GTA TTA CTG TGC GAR GNN KNN KNN KNN KNN KNN KNN KNN KNN KNN KGA CNN KTG GGG

CCA AGG AAC CCT GGT C-3') in combination with the primer pUC57-3 (5'-CAG GTT TCC CGA CTG GAA AG-3'). To prevent introduction of a sequence bias due to priming of the randomized region of the primers HCDR3-NNK4-F, HCDR3-NNK6-F, HCDR3-NNK8-F and HCDR3-NNK10-F on pUC57_J_H4, the plasmid was first linearized by digestion with the restriction enzyme DrdI (Figure 17B).

[00277] The resulting V_H3-30 PCR product (SEQ-ID51) displayed an overlap of 22bp with the NNK4-J_H4, NNK6-J_H4, NNK8-J_H4 and NNK10-J_H4 PCR products (SEQ-ID52 to 55), and was assembled with each by PCR overlap extension in 4 separate reactions, using the primers V_H3-30-F and pUC57-3. The resulting DNA molecules comprised the V_H library flanked by NotI and NheI restriction sites as shown in Figure 16. All PCR amplicons obtained from the PCRs employing the NNK4-J_H4, NNK6-J_H4, NNK8-J_H4 and NNK10-J_H4 degenerate oligos were subjected to direct DNA sequencing, and it was confirmed that the designed randomization of the HCDR3 positions was obtained, as expected. This is shown by way of example in Fig. 18 (B), where the electropherogram of the region spanning the HCDR3 is provided. The randomized positions show expected sequence peak overlays demonstrating the nucleotide diversity in these positions (Fig. 18). The 4 different V_H library DNAs were mixed in equimolar ratio, digested with the restriction endonucleases NotI and NheI and cloned into the *PiggyBac*-transposable vector pPB-EGFP_HC-gamma1 (SEQ-ID22), upstream of the gamma1 heavy chain constant region, resulting in a library consisting of 3.7x 10^7 independent clones. The size of this library can easily be increased by a factor 10 by scaling up the ligation reaction.

[00278] Heavy chain libraries incorporating distinct randomization designs, or comprising V_H and J_H gene segments other than the ones used in this example, can be produced the same way.

[00279] DNA sequences referred to in this Example 13:

[00280] SEQ-ID49 (pUC57_VH3-30)

[00281] SEQ-ID50 (pUC57_J_H4)

[00282] SEQ-ID51 (V_H3-30 PCR product)

[00283] SEQ-ID52 (NNK4-J_H4 PCR product)

[00284] SEQ-ID53 (NNK6-J_H4 PCR product)

[00285] SEQ-ID54 (NNK8-J_H4 PCR product)

[00286] SEQ-ID55 (NNK10-J_H4 PCR product)

Example 14: Identification of variable light and heavy chain coding regions from antigen-reactive, enriched and stably transposed host cells

[00287] Due to the stable integration of the transposable expression vectors encoding antibody heavy and light chains in the host, the variable heavy and light chain coding regions can be re-isolated in a straightforward way by standard PCR amplification followed by direct sequencing of the PCR amplicons or, upon re-cloning, from re-cloned plasmid vectors. For this, isolated cells or cell clones, expressing antigen-specific antibodies are centrifuged for 5 minutes at 1200x g. Total RNA is isolated from these cells using TRIzol reagent (Sigma-Aldrich). First strand cDNA can be synthesized with PowerScript (Clontech-Life Technologies) using an oligio-dT primer. The light chain coding regions can then amplified by PCR using the primers SP-F (5'-GAG GAG GAG GCG GCC GCC ATG AAT TTT GGA C-3') and CK-rev (5'-GAG GAG GAG TTC GAA AGC GCT AAC ACT CTC-3'), which will result in a PCR amplicon of ca. 740 bp, depending on length of the V-kappa region contained in the PCR amplicon. If desired, this PCR amplicon can be digested with restriction endonucleases NotI and BstBI, and cloned into the vector pPB-EGFP_HC-g1 (Seq-ID22), in order to subclone individual clones of the PCR amplicon for further sequence identification. Individual V-kappa region clones can then be subjected to sequencing using the primer pPB-seq13 (5'-GGC CAG CTT GGC ACT TGA TG-3'), binding in the EF1-alpha promoter, upstream of the cloned V-coding region.

[00288] The heavy chain variable regions can be PCR amplified on cDNA, generated as above, using the primers SP-F (5'-GAG GAG GAG GCG GCC GCC ATG AAT TTT GGA C-3') and CG-revseq-1 (5'- GTT CGG GGA AGT AGT CCT TG-3') that will result in a PCR amplicon of ca. 530 bp expected size, depending in the length of the V_H-region contained in the PCR amplicon. If desired, this PCR amplicon can be digested with restriction endonucleases NotI and NheI, and cloned into the vector pPB-EGFP_HC-g1 (Seq-ID22). Individual clones are then subjected to sequencing of the V_H-region using the primer pPB-seq13 (5'-GGC CAG CTT GGC ACT TGA TG-3'), binding in the EF1-alpha promoter, upstream of the cloned V-coding region.

[00289] EMBODIMENTS OF THE INVENTION

1. A method for identifying a polypeptide having a desired binding specificity or functionality, comprising:

(i) generating a diverse collection of polynucleotides encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence coding for a polypeptide disposed between first and second inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme;

(ii) introducing the diverse collection of polynucleotides of (i) into host cells;

(iii) expressing at least one transposase enzyme functional with said inverted terminal repeat sequences in said host cells so that said diverse collection of polynucleotides is integrated into the host cell genome to provide a host cell population that expresses said diverse collection of polynucleotides encoding polypeptides having different binding specificities or functionalities;

(iv) screening said host cells to identify a host cell expressing a polypeptide having a desired binding specificity or functionality; and

(v) isolating the polynucleotide sequence encoding said polypeptide from said host cell.

2. A method according to 1, wherein said polynucleotides are DNA molecules.

3. A method according to 1, wherein said polynucleotides comprise a ligand-binding sequence of a receptor or a target-binding sequence of a binding molecule.

4. A method according to 1, wherein said polynucleotides comprise an antigen-binding sequence of an antibody.

5 A method according to 1, wherein said polynucleotides comprise a sequence encoding a VH or VL region of an antibody, or an antigen-binding fragment thereof.

6. A method according to 1, wherein said polynucleotides comprise a sequence encoding an antibody VH region and an antibody VL region.

7. A method according to 1, wherein said polynucleotides comprise a sequence encoding a full-length immunoglobulin heavy chain or light chain, or an antigen-binding fragment thereof.

8. A method according to 1, wherein said polynucleotides comprise a sequence encoding a single-chain Fv or a Fab domain.

9. A method according to 1, wherein generating said diverse collection of polynucleotides comprises subjecting V region gene sequences to PCR under mutagenizing conditions.

10. A method according to 1, wherein generating said diverse collection of polynucleotides comprises gene synthesis.

11. A method according to 1, wherein generating said diverse collection of polynucleotides comprises PCR amplification of V region repertoires from vertebrate B cells.

12. A method according to 1, wherein said diverse collection of polynucleotides comprises plasmid vectors.

13. A method according to 1, wherein said diverse collection of polynucleotides comprises double-stranded DNA PCR amplicons.

14. A method according to 4, wherein said antigen-binding sequence is of a vertebrate.

15. A method according to 4, wherein said antigen-binding sequence is mammalian.

16. A method according to 4, wherein said antigen-binding sequence is human.

17. A method according to 12, wherein said plasmid vectors further encode a marker gene.

18. A method according to 17, wherein said marker is selected from the group consisting of: a fluorescent marker, a cell surface marker and a selectable marker.

19. A method according to 17, wherein said marker gene sequence is upstream or downstream of the sequence encoding the polypeptide having a binding specificity or functionality, but between the inverted terminal repeat sequences.

20. A method according to 17, wherein said marker gene sequence is downstream of said sequence encoding a polypeptide having binding specificity or functionality and separated by an internal ribosomal entry site.

21. A method according to 1, wherein step (ii) comprises introducing into said host cells polynucleotides comprising sequences encoding immunoglobulin VH or VL regions, or antigen-

binding fragments thereof, and wherein said VH and VL region sequences are encoded on separate vectors.

22. A method according to 21, wherein step (ii) comprises introducing into said host cells polynucleotides comprising sequences encoding full-length immunoglobulin heavy or light chains, or antigen-binding fragments thereof, wherein said full-length heavy and light chain sequences are on separate vectors.

23. A method according to 21, wherein said vectors comprising VH sequences and said vectors comprising VL sequences are introduced into said host cells simultaneously.

24. A method according to 21, wherein said vectors comprising VH sequences and said vectors comprising VL sequences are introduced into said host cells sequentially.

25. A method according to 1, wherein step (ii) comprises introducing into said host cells a vector comprising sequences encoding antibody VH and VL chains.

26. A method according to 1, wherein step (ii) comprises introducing into said host cells a vector comprising sequences encoding a full-length immunoglobulin heavy chain and a full-length immunoglobulin light chain.

27. A method according to 21, wherein said vector comprising the VH sequence comprises inverted terminal repeat sequences that are recognized by a different transposase enzyme than the inverted terminal repeat sequences in the vector comprising the VL sequence.

28. A method according to 1, wherein the host cells of step (ii) are vertebrate cells.

29. A method according to 28, wherein said host cells are mammalian.

30. A method according to 29, wherein said host cells are human or rodent cells.

31. A method according to 28, wherein said vertebrate host cells are lymphoid cells.

32. A method according to 31, wherein said host cells are B cells.

33. A method according to 32, wherein said host cells are progenitor B cells or precursor B cells.

34. A method according to 33, wherein said host cells are selected from the group consisting of: Abelson-Murine Leukemia virus transformed progenitor B cells or precursor B cells and early, immunoglobulin-null EBV transformed human proB and preB cells.

35. A method according to 32, wherein said host cells are selected from the group consisting of: Sp2/0 cells, NS0 cells, X63 cells, and Ag8653 cells.

36. A method according to 29, wherein said host cells are selected from the group consisting of: CHO cells, Per.C6 cells, BHK cells, and 293 cells.

37. A method according to 1, wherein said expressing step (iii) comprises introducing into said host cells an expression vector encoding a transposase enzyme that recognizes and is functional with an least one inverted terminal repeat sequence.

38. A method according to 37, wherein said vector encoding said transposase enzyme is introduced into said host cells concurrently with or prior or subsequent to the diverse collection of polynucleotides.

39. A method according to 37, wherein said transposase enzyme is transiently expressed in said host cell.

40. A method according to 1, wherein said expressing step (iii) comprises inducing an inducible expression system that is stably integrated into the host cell genome.

41. A method according to 40, wherein said inducible expression system is tetracycline-inducible or tamoxifen-inducible.

42. A method according to 1, wherein said screening step (iv) comprises magnetic activated cell sorting (MACS), fluorescence activated cell sorting (FACS), panning against molecules immobilized on a solid surface panning, selection for binding to cell-membrane associated molecules incorporated into a cellular, natural or artificially reconstituted lipid bilayer membrane, or high-throughput screening of individual cell clones in multi-well format for a desired functional or binding phenotype.

43. A method according to 1, wherein said screening step (iv) comprises screening to identify polypeptides having a desired target-binding specificity or functionality.

44. A method according to 1, wherein said screening step (iv) comprises screening to identify antigen-binding molecules having a desired antigen specificity.

45. A method according to 44, wherein said screening step further comprises screening to identify antigen-binding molecules having one or more desired functional properties.

46. A method according to 1, wherein said screening step (iv) comprises multiple cell enrichment cycles with host cell expansion between individual cell enrichment cycles.

47. A method according to 1, wherein said step (v) of isolating the polynucleotide sequence encoding the polypeptide having a desired binding specificity or functionality comprises genomic or RT-PCR amplification or next-generation deep sequencing.

48. A method according to 1, further comprising (vi) affinity optimizing the polynucleotide sequence obtained in (v).

49. A method according to 48, wherein said affinity optimization comprises genomic PCR or RT-PCR under mutagenizing conditions.

50. A method according to 49, further comprising subjecting the mutagenized sequences to steps (i)-(v) of claim 1.

51. A method according to 1, wherein said inverted terminal repeat sequences are from the PiggyBac transposon system.

52. A method according to 51, wherein the sequence encoding the upstream PiggyBac inverted terminal repeat sequence comprises SEQ ID NO:1.

53. A method according to 51, wherein the sequence encoding the downstream PiggyBac inverted terminal repeat sequence comprises SEQ ID NO:2.

54. A method according to 5, wherein said VH or VL region sequences encode a sequence derived from a human anti-TNF alpha antibody.

55. A method according to 54, wherein said human anti-TNF alpha antibody is D2E7.

56. A method according to 55, wherein the VH and VL regions of D2E7 are encoded by separate transposable vectors.

57. A method according to 56, wherein said vector comprising said VL region sequence comprises SEQ ID NO:5.

58. A method according to 56, wherein said vector comprising said VH region sequence comprises SEQ ID NO:8.

59. A method according to 56, wherein said vector comprising said VH region sequence comprises a randomized sequence as set forth in SEQ ID NO:9.

60. A method according to 56, wherein said vector comprising said VL region sequence comprises a randomized sequence as set forth in SEQ ID NO:10.

61. A method according to 1, wherein step (iii) comprises introducing into said host cell a vector comprising a sequence encoding a functional PiggyBac transposase.

62. A method according to 61, wherein said vector comprises SEQ ID NO:11.

63. A method according to 61, wherein said vector encodes SEQ ID NO:12, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

64. A method according to claim1, wherein said inverted terminal repeat sequences are recognized by and functional with at least one transposase selected from the group consisting of: PiggyBac, Sleeping Beauty, Frog Prince, Himar1, Passport, Minos, hAT, Tol1, Tol2, Ac/Ds, PIF, Harbinger, Harbinger3-DR, and Hsmar1.

65. A library of polynucleotide molecules encoding polypeptides having different binding specificities or functionalities, comprising a plurality of polynucleotide molecules, wherein said polynucleotide molecules comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme.

66. A library according to 65, wherein said polynucleotides are DNA.molecules.

67. A library according to 65, wherein said polynucleotides comprise a ligand-binding sequence of a receptor or a target-binding sequence of a binding molecule.

68. A library according to 65, wherein said polynucleotides comprise at least one sequence encoding an antigen-binding sequence of an antibody.

69 A library according to 65, wherein said polynucleotides comprise a sequence encoding a VH or VL region of an antibody or an antigen-binding fragment thereof.

70. A library according to 65, wherein said polynucleotides comprise a sequence encoding an antibody VH region and an antibody VL region.

71. A library according to 65, wherein said polynucleotides comprise a sequence encoding a full-length immunoglobulin heavy chain or light chain, or an antigen-binding fragment thereof.

72. A library according to 65, wherein said polynucleotides comprise a sequence encoding a single-chain Fv or a Fab domain.

73. A library according to 65, wherein said polynucleotide molecules are plasmids.

74. A library according to 65, wherein said polynucleotide molecules are double stranded DNA PCR amplicons.

75. A library according to 73, wherein said plasmids further comprise a sequence encoding a marker gene.

76. A library according to 73, wherein said plasmids further comprise a sequence encoding a transposase enzyme that recognizes and is functional with the inverted terminal repeat sequences.

77. A method for generating a library of transposable polynucleotides encoding polypeptides having different binding specificities or functionality, comprising:

(i) generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence encoding polypeptide having a binding specificity or functionalitydisposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme.

78. A vector comprising a sequence encoding a VH or VL region of an antibody, or antigen-binding portion thereof, disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme.

79. A vector according to 78,comprising a sequence encoding a full-length heavy or light chain of an immunoglobulin.

80. A vector according to 78, wherein said VH or VL region sequence is randomized.

81. A vector according to 78, wherein said inverted terminal repeat sequences are recognized by and functional with the PiggyBac transposase.

82. A vector according to 78, wherein said VH or VL region sequence is derived from an anti-TNF alpha antibody.

83. A vector according to 82, wherein said antibody is D2E7.

84. A vector according to 78, comprising at least one sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:17, and SEQ ID NO:19.

85. A host cell comprising a vector according to any one of claims 78-84.

86. A host cell according to 85 further comprising an expression vector comprising a sequence encoding a transposase that recognizes and is functional with at least one inverted terminal repeat sequence in the vector encoding said VH or VL region sequence.

87. An antibody produced by a method comprising claim 1.

88. A method according to 1, wherein said inverted terminal repeat sequences are from the Sleeping Beauty transposon system.

89. A method according to 88, wherein the sequence encoding the upstream Sleeping Beauty inverted terminal repeat sequence comprises SEQ ID NO:14.

90. A method according to 88, wherein the sequence encoding the downstream Sleeping Beauty inverted terminal repeat sequence comprises SEQ ID NO:15.

91 A method according to 88, wherein step (iii) comprises expressing in said host cell a vector comprising a functional Sleeping Beauty transposase.

92. A method according to 48, wherein said polynucleotide sequence obtained in (v) comprises a sequence encoding a VH or VL region of an antibody, or an antigen-binding fragment thereof, and wherein said antibody optimization comprises introducing one or more mutations into a complementarity determining region or framework region of said VH or VL.

93. A library according to 71, wherein said full-length immunoglobulin heavy chain comprises the natural intron/exon structure of an antibody heavy chain.

94. A library according to 93, wherein said full-length immunoglobulin heavy chain comprises the endogenous membrane anchor domain.

95. A method for generating a population of host cells capable of expressing polypeptides having different binding specificities or functionalities, comprising:

(i) generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides

comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme; and

(ii) introducing said diverse collection of polynucleotides into host cells.

96. A vector according to 78, wherein said inverted terminal repeat sequences are recognized by and functional with the Sleeping Beauty transposase

97. A method according to 91, wherein step (iii) comprises expressing in said host cell a vector comprising SEQ ID NO:17.

98. A method according to 91, wherein said vector encodes SEQ ID NO:18, or a sequence with at least 95% amino acid sequence homology and having the same or similar inverted terminal repeat sequence specificity.

WHAT IS CLAIMED IS

1. A method for identifying a polypeptide having a desired binding specificity or functionality, comprising:

(i) generating a diverse collection of polynucleotides, preferably plasmid vectors or double stranded DNA PCR amplicons, encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence coding for a polypeptide disposed between first and second inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme;

(ii) introducing the diverse collection of polynucleotides of (i) into host cells;

(iii) expressing at least one transposase enzyme functional with said inverted terminal repeat sequences in said host cells so that said diverse collection of polynucleotides is integrated into the host cell genome to provide a host cell population that expresses said diverse collection of polynucleotides encoding polypeptides having different binding specificities or functionalities;

(iv) screening said host cells to identify a host cell expressing a polypeptide having a desired binding specificity or functionality; and

(v) isolating the polynucleotide sequence encoding said polypeptide from said host cell.

2. A method according to claim 1, wherein said polynucleotides comprise

a) a ligand-binding sequence of a receptor or a target-binding sequence of a binding molecule,

b) an antigen-binding sequence of an antibody,

c) a sequence encoding a V_H or V_L region of an antibody, or an antigen-binding fragment thereof.

d) a sequence encoding an antibody V_H region and an antibody V_L region.

e) a sequence encoding a full-length immunoglobulin heavy chain or light chain, or an antigen-binding fragment thereof, and/or

f) a sequence encoding a single-chain Fv or a Fab domain.

3 A method according to any of the aforementioned claims, wherein generating said diverse collection of polynucleotides comprises subjecting V region gene sequences to PCR under mutagenizing conditions.

4. A method according to any of the aforementioned claims, wherein step (ii) comprises introducing into said host cells polynucleotides comprising sequences encoding

a) immunoglobulin V_H or V_L regions, or antigen-binding fragments thereof, and wherein said V_H and V_L region sequences are encoded on separate vectors, and/or

b) full-length immunoglobulin heavy or light chains, or antigen-binding fragments thereof, wherein said full-length heavy and light chain sequences are on separate vectors.

c) a vector comprising sequences encoding antibody V_H and V_L chains.

d) a vector comprising sequences encoding a full-length immunoglobulin heavy chain and a full-length immunoglobulin light chain.

5. A method according to any of the aforementioned claims, wherein said expressing step (iii) comprises introducing into said host cells an expression vector encoding a transposase enzyme that recognizes and is functional with an least one inverted terminal repeat sequence, wherein said transposase enzyme is preferably transiently expressed in said host cell.

6. A method according to any of the aforementioned claims, wherein said screening step (iv) comprises magnetic activated cell sorting (MACS), fluorescence activated cell sorting (FACS), panning against molecules immobilized on a solid surface, selection for binding to cell-membrane associated molecules incorporated into a cellular, natural or artificially reconstituted lipid bilayer membrane, or high-throughput screening of individual cell clones in multi-well format for a desired functional or binding phenotype.

7. A method according to any of the aforementioned claims, wherein said step (v) of isolating the polynucleotide sequence encoding the polypeptide having a desired binding specificity or functionality comprises genomic or RT-PCR amplification or next-generation deep sequencing.

8. A method according to any of the aforementioned claims, wherein

a) said inverted terminal repeat sequences are from the *PiggyBac* transposon system or the *Sleeping Beauty* transposon system, and/or

b) step (iii) comprises introducing into said host cell a vector comprising a sequence encoding a functional *PiggyBac* transposase or *Sleeping Beauty* transposase.

9 A method according to any of the aforementioned claims, wherein said inverted terminal repeat sequences are recognized by and functional with at least one transposase selected from the group consisting of: *PiggyBac*, *Sleeping Beauty*, Frog Prince, Himar1, Passport, Minos, hAT, Tol1, Tol2, Ac/Ds, PIF, Harbinger, Harbinger3-DR, and Hsmar1.

10. A library of polynucleotide molecules encoding polypeptides having different binding specificities or functionalities, comprising a plurality of polynucleotide molecules, preferably plasmids or double stranded DNA PCR amplicons, wherein said polynucleotide molecules comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme.

11. A library according to claim 10, wherein said polynucleotides comprise

a) at least one sequence encoding an antigen-binding sequence of an antibody.

b) a sequence encoding a V_H or V_L region of an antibody or an antigen-binding fragment thereof.

c) a sequence encoding an antibody V_H region and an antibody V_L region.

d) a sequence encoding a full-length immunoglobulin heavy chain or light chain, or an antigen-binding fragment thereof.

e) a sequence encoding a single-chain Fv or a Fab domain.

12. A library according to claims 10 - 11, wherein said plasmids or double stranded DNA PCR amplicons further comprise a sequence encoding a transposase enzyme that recognizes and is functional with the inverted terminal repeat sequences.

13. A method for generating a library of transposable polynucleotides encoding polypeptides having different binding specificities or functionality, comprising generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence encoding polypeptide having a

binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme.

14. A vector comprising a sequence encoding a V_H or V_L region of an antibody, or antigen-binding portion thereof, disposed between inverted terminal repeat sequences that are recognized by and functional with at least one transposase enzyme.

15 A host cell comprising a vector according to claim 14.

16. A method for generating a population of host cells capable of expressing polypeptides having different binding specificities or functionalities, comprising:

(i) generating a diverse collection of polynucleotides comprising sequences encoding polypeptides having different binding specificities or functionalities, wherein said polynucleotides comprise a sequence encoding a polypeptide having a binding specificity or functionality disposed between inverted terminal repeat sequences that are recognized by and functional with a least one transposase enzyme; and

(ii) introducing said diverse collection of polynucleotides into host cells.

17. The method or vector according to any one of the aforementioned claims 2-9, 11-12 and 14-15, wherein the vector comprising the V_H sequence comprises inverted terminal repeat sequences that are recognized by a different transposase enzyme than the inverted terminal repeat sequences in the vector comprising the V_L sequence.

18. A method or host cell according to any one of the aforementioned claims 1-9 and 15-17, wherein the host cells are vertebrate cells, preferably mammalian cells, more preferably human or rodent cells, still more preferably lymphoid cells yet still more preferably B cell, even more preferably progenitor B cells or precursor B cells, particularly preferred Abelson-Murine Leukemia virus transformed progenitor B cells or precursor B cells and early, immunoglobulin-null EBV transformed human proB and preB cells.