A cell surface tag exchange (CSTE) system for tracing and manipulation of cells during recombinase mediated cassette exchange integration of nucleic acid sequences to engineered receiver cells

Field of Invention

The present invention relates to the fields of cell engineering and recombinant DNA technology. In particular the invention relates to systems and methods for generating, purifying, tracing and manipulating engineered cells during recombinase mediated cassette exchange (RMCE) based integration of a gene of interest.

Background

Within the field of genetic manipulation and immunology the Recombinase-mediated cassette exchange (RMCE) has become a useful tool for targeted genetic modifications, see e.g. Turan et al. Gene (2013):515:1-27. Generally, post-RMCE methods to confirm integration of the gene of interest (GOI) in the correct location require cellular disruption and sample processing. A recently described approach overcomes such need for cellular disruption and demonstrates that recombination events can be monitored and isolated by following expression of fluorescent marker proteins (Phan et al. Sci Rep. (2017):7(1):17771). Although the approach by Phan et al avoids the need for cellular disruption, as it relies on intracellular protein fluorescence, the technique is not suited for high-content screening of integrated constructs due to a limited set of unique spectral properties provided by the fluorescent proteins; as a result only a limited number of genes may be labelled and detected simultaneously using the described strategy. Furthermore, intracellular fluorescent proteins are not amenable to physical manipulation to partition cells via substrate affinity methods such as magnetic activated cell sorting (MACS), which provides potential for high throughput and parallelisation during cell engineering workflows, or potential for increasing the efficiency of high-content cell library engineering.

Currently, with the RMCE technique at hand there is a need for tools for rapid and robust cell engineering in both high-throughput and high-content applications, particularly in cases where cells have been simultaneously transfected with a pool of vectors containing a pool of individual genes of interest, or in cases where selection of cells by substrate affinity methods is desirable. There is thus a clear need for providing effective and improved methods.

Summary

The present invention relates to the provision of a two-component cell surface tag exchange (CSTE) system for tracing and manipulation of engineered cells during recombinase mediated cassette exchange (RMCE) based integration of a gene of interest (GOI). The first part of the CSTE system represents a tag-exchange donor vector (TEDV) encoding the GOI for integration, along with a first cell surface tag (CST) flanked by a 3' fragment of an intron sequence including the splice acceptor site. The TEDV construct is flanked by RMCE elements. The second part of the CSTE represents a tag-exchange receiver site (TERS) contained within the genome of an engineered target cell, encoding a selection gene, and a second CST with an in-frame transmembrane domain, adjoined by a complete intron sequence that also encodes a RMCE element, and a second RMCE element flanking the GOI coding sequence. The CST encoded by the TERS is essentially encoded as two exons adjoined by a single intron. Paired RMCE elements included in both the TEDV and TERS are designed such that execution of RMCE between the two constructs results in exchange of CST and GOI from the TEDV into the TERS, with reciprocal loss off the TERS CST and selection gene encoded by the TERS. The CST delivered by the TEDV utilises the transmembrane domain encoded by the TERS, wherein execution of RMCE results in exchange of the unique CST exon sequences. Promoter elements that drive CST and selection gene/GOI expression from the TERS are extrinsic to the RMCE exchanged constructs, thus sequences delivered by the TEDV will generally only be expressed upon faithful execution of RMCE. This CSTE system enables rapid and robust RMCE that is conditionally reported by detection of the affinity epitopes presented at the cell surface by the CST, and which also enables physical partitioning of cells with substrate-immobilised affinity methods. This CSTE system also enables multiplexing and lineage tracing of GOI integration through use of multivalent affinity epitopes as CST elements. The CSTE system represents a powerful tool for rapid and robust cell engineering in both high-throughput and high-content applications.

In a first aspect the present invention provides a combined system comprising two separate components, wherein the first component is a tag-exchange donor vector (TEDV) encoding a first cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation, and the second component is an engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a second CST exon adjoined by a full intron sequence to an exon encoding a transmembrane domain, and also encoding a reporter gene in the antisense

orientation, wherein paired recombinase mediated cassette exchange (RMCE) elements are included in the TEDV and TERS such that execution of RMCE between the TEDV and TERS results in exchange of the reporter element for the GOI encoded by the TEDV, and exchange of the first CST exon for the second CST exon, such that the derivative engineered cell now expresses the first CST and GOI, in place of the second CST and the reporter gene.

In one embodiment the first component is a TEDV comprising

a. a first RMCE element – 5'RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment
b. a 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site
c. an exon comprising the TEDV-encoded CST in the 5' to 3' direction
d. a first transcriptional terminator sequence for encoded CST in the 5' to 3' direction
e. a second transcriptional terminator for the 3' to 5' encoded GOI
f. a sequence encoding GOI in the 3' to 5' direction
g. a Kozak sequence
h. a 5' RMCE element

wherein, the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

In another embodiment the second component is a TERS comprising

a. a transcriptional promoter element
b. a Kozak sequence
c. a Type 2 membrane protein transmembrane domain exon
d. a 5' intron splice donor site
e. a 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment, as equivalent, and paired with, the 5' RMCE element of the TEDV
f. functional sequences of the 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site
g. an exon comprising the TERS-encoded CST in the 5' to 3' direction (that is different to the TEDV-encoded CST)
h. a transcriptional terminator sequence for encoded CST in the 5' to 3' direction

i. a transcriptional terminator sequence for the 3' to 5' direction
j. a sequence encoding the selection gene in the 3' to 5' direction
k. a Kozak sequence for efficient translational initiation of the selection gene transcript
l. a 3' RMCE element
m. a 3' genomic element responsible for regulating expression and tracking of the CST transcript,

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

In a second aspect the present invention provides a method for generating derivative engineered cells expressing a TEDV-encoded GOI from the TERS locus, said method comprising,

a. generating a TEDV encoding a GOI
b. delivering said TEDV to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein
c. contacting cells with two or more affinity reagents specific for both the TEDV-encoded CST and TERS-encoded CST
d. selecting derivative engineered cells on the basis of diminished expression of the reporter gene and TERS-encoded CST, and increased expression of the TEDV-encoded CST, as a proxy for selection of cells with integrated GOI.

In a third aspect the present invention provides a tag-exchange donor vector (TEDV) encoding a cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation.

In a fourth aspect the present invention provides an engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a cell surface tag (CST) exon adjoined by a full intron sequence to an exon encoding a transmembrane domain, and also encoding a reporter gene in the antisense orientation, wherein a recombinase mediated cassette exchange (RMCE) element is included in the TERS such that

execution of RMCE between TERS and a tag-exchange donor vector (TEDV) results in exchange of the reporter element for a gene of interest (GOI) encoded by the TEDV.

The present invention provides a system to integrate a GOI into an engineered cell line by means of RMCE, wherein reporting of said integration is conditionally detected by analysis of a co-integrated cell surface tag (CST). Moreover, upon integration of the GOI, a concomitant loss of a distinct CST from the engineered cell line permits dual positive/negative selection of GOI-expressing derivative engineered cells. This 'tag-exchange' is highly robust and the basis for the stringent positive/negative selection enabled by the CSTE system. A second selection gene is also lost from the TERS site, enabling a highly robust double negative selection and single positive selection during generation of derivative engineered cells expressing the GOI. Importantly, the nature of the CST system means that not only are these markers conditional, requiring selective addition of cognate affinity reagents for detection, but may also be leveraged to physically partition target cell populations by use of substrate-immobilised affinity reagent methodologies. This enables efficient high-throughput methods for derivative engineered cell generation through the partitioning of cells using methods such as magnetic activated cell sorting (MACS), as opposed to more time-consuming and difficult to parallelise methodologies such as florescence-activated cell sorting (FACS). Finally, the CSTE system, in replacing the intracellular expression of fluorescence proteins and antibiotic resistance genes as markers of integration, provides a much larger selectable marker space. In particular, the generation of a large number of unique CST epitopes for conditional selection is readily achieved, as opposed to florescent proteins with broad spectral overlap, or the highly limited set of antibiotic resistance genes. This broader access to selectable markers enables different forms of high-throughput derivative cell production with increased efficiency, and further enables high-content methods for both cell library creation and cell lineage tracking in cellular engineering, pathway engineering and experimental workflows.

Detailed description

The overall architecture and operation of the CSTE system is depicted in Figure 1. The CSTE system operates as a donor/receiver pair, wherein the tag-exchange donor vector (TEDV) acts to deliver a GOI sequence, and a cell-surface tag (CST) exon to a paired tag-exchange receiver site (TERS), which is generally contained within the genome of an engineered cell line. The TERS encodes a distinct CST exon and a

selection gene, which are exchanged with the TEDV-encoded CST and GOI, respectively.

The TEDV-encoded CST exon is encoded in the sense (5' to 3') direction and in frame with a 3' intron fragment, which includes splice acceptor site, polypyrimidine tract and branch point sequences. To the 5' of these functional intronic sequences, within the 'non-functional' intronic sequence, is encoded a first RMCE element. The GOI is encoded in the antisense (3' to 5') direction. To the 3' end of the construct adjacent to the GOI transcriptional start site a second RMCE element is encoded.

The paired TERS construct carried by an engineered cell line encodes a CST exon in frame with a complete intron sequence, with a first RMCE element within the 'non-functional' intronic sequence, paired with, and in equivalent context of the first RMCE element of the TEDV. The complete intron sequence in which this first RMCE site is encoded includes a 5' splice donor site, itself encoded in-frame with a transmembrane domain (TD) exon. This TD contains the transcriptional start site and promoter sequences to drive transcription to the 5' of the TERS construct. Thus, the TERS-encoded CST is expressed at the cell surface once translated from the spliced transcript. Similarly, once RMCE is executed between TEDV and TERS, the TEDV-encoded in the CST exon is exchanged with the TERS-encoded CST exon. This results in the execution of a 'tag-exchange', wherein the derivative engineered cell line now expresses the TEDV-encoded CST at the cell surface, and not the TERS-encoded CST. In this process, the TEDV-encoded CST exon 'acquires' the 5' splice donor sites, TD domain, transcriptional start site and promoter sequences from the TERS construct to enable the expression of the TEDV-encoded CST, providing a stringent site-specific control over integration of TEDV-encoded sequences.

As mentioned above, the TERS encodes a selection gene in the antisense (3' to 5') direction, with a promoter sequence able to drive the transcription of this selection gene to the 3' end of the TERS construct. Between the transcriptional start site of the selection gene and the 3' promoter sequence of the TERS, a second RMCE element is encoded, paired to the second 5' RMCE element encoded in the TEDV. Therefore, when RMCE is executed between the TEDV and TERS, the TEDV-encoded GOI is exchanged with the TERS-encoded selection gene.

It is important to note that the TEDV contains no promoter sequences, and thus both the TEDV-encoded CST exon and GOI must be complimented with promoter sequences to enable their transcription. Importantly, the inclusion of the intron and TD exon complementation for the CST exon, as outlined above, provides an even more stringent site selectivity for integration of TEDV-encoded sequences. The use of the intron element in the CSTE provides an additional level of assurance that random integration of the construct will not provide an expressible TEDV-encoded CST on the cell surface. In contrast to a donor vector that would deliver the full CST ORF, including the TD, the CSTE system provides only the CST epitope exon and a splice acceptor site. Thus, a random aberrant integration of the donor construct would not only need to integrate adjacent to an active promoter, but would also require acquisition of a Type II transmembrane domain, or equivalent, with an appropriate 5' splice donor sequence. This results in the occurrence of a randomly integrated construct expressing a TEDV-encoded CST being extremely unlikely, minimising the selection of such random and undesired events during generation of derivative engineered cells expressing the delivered GOI.

The specific architecture of the CSTE system componentry, comprising the TEDV and the TERS, is detailed below.

TERS Architecture

The TERS, contained within an engineered cell line, constitutively expresses a TERS-encoded CST and selection gene, which upon execution of RMCE with a paired TEDV are exchanged with a TEDV-encoded CST and a GOI, respectively. The generic architecture of a TERS is depicted in Figure 3.

A TERS typically comprises:

a) a Kozak sequence

b) an exon encoding a Type 2 membrane protein transmembrane domain.

c) a 5' intron splice donor site.

d) a 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment, as equivalent, and paired with, the 5' RMCE element of the TEDV.

e) the functional sequences of the 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site.

f) an exon comprising the TERS-encoded CST in the 5' to 3' direction.

g) a transcriptional terminator sequence for encoded CST in the 5' to 3' direction.

h) a transcriptional terminator sequence for the 3' to 5' encoded selection gene.

i) a sequence encoding the selection gene in the 3' to 5' direction

j) a Kozak sequence for efficient translational initiation of the selection gene transcript

k) a 3' RMCE element

x) a 5' genomic element responsible for regulating expression and tracking of the CST transcript. This genomic element minimally includes a promoter sequence that drives CST transcription.

y) a 3' genomic element responsible for regulating expression and tracking of the selection gene transcript. This genomic element minimally includes a promoter sequence that drives selection gene transcription;

wherein; a) is provided to ensure efficient translational initiation from the adjoined TD / CST transcript driven from the TERS, both for the TERS-encoded CST prior to execution to RMCE, and the TEDV-encoded CST subsequent to execution of RMCE; b) represents the TD that is adjoined to the CST exon by transcript splicing, and mediates the inserts of the translated protein construct in the cell plasma membrane and presents the CST to the extracellular space; c) is the RMCE element that is situated within the intron sequence of the TERS construct, and is equivalent to the 5' RMCE element of the TEDV, mediating CST exon exchange upon execution of RMCE within the CSTE system; e) is the splice acceptor site and associated functional sequence of the intron, enabling splicing of the intron, containing the RMCE element, from the TD / CST transcript driven from the TERS; f) is the exon encoding the CST epitope(s) within the TERS site, such that the expressed TD / CST protein presents

said epitope to the extracellular space, and which can be conditionally detected with cognate affinity reagent; g) represents a transcriptional terminator for the TD / CST open reading frame encoded in the 5' to 3' direction; h) represents a transcriptional terminator for the selection gene open reading frame encoded in the 3' to 5' direction within the TERS; i) represents an open reading frame encoding a selection gene expressed from the TERS in the engineered cell line; j) is a Kozak sequence provided to ensure efficient translational initiation from the selection gene transcript driven from the TERS prior to execution of RMCE; k) is the RMCE element that is situated between the transcriptional start site of the selection gene and the promoter sequences driving transcription of this open reading frame, and is equivalent to the 3' RMCE element of the TEDV, mediating exchange of selection gene with the TEDV-encoded GOI upon execution by RMCE within the CSTE system; x) minimally represents promoter sequences that drive the transcription of the TD / CST open reading frame encoded by the TERS prior to RMCE, and the TEDV-encoded TD / CST subsequent to execution of RMCE; y) minimally represents promoter sequences that drive the transcription of the selection gene open reading frame encoded by the TERS prior to RMCE, and the TEDV-encoded GOI subsequent to execution of RMCE.

A schematic representation of the TERS cloning fragment is presented in Figure 3.

The TERS essentially encodes a fully functional intron sequence that contains the 5' RMCE site. Such a sequence may represent any intron that enables efficient splicing of the TD/CST transcript in the context of the host engineered cell. The splice acceptor site of the TEDV is required to be functional in concert with the splice donor site of the TERS. Thus, the TEDV and TERS encoded splice acceptor sequence should be the same or equivalent.

The non-coding and 'non-functional' intron sequence, wherein the RMCE site is encoded, may also encode further genetic elements such as transcriptional enhancers, transcriptional insulators, distinct open reading frames encoding separate desirable components for function or reporting of the TERS construct within the engineered cell, or other unique sequences used for transcript and/or construct tracing and quantification.

In the context of the TERS, both the encoded TD/CST and the selection gene may contain 5' and 3' untranslated regions (UTR), which encode unique sequences used for

transcript and/or construct tracing and quantification, or impart regulation of transcript stability, for example.

The selection gene can be selected from

a. an antibiotic resistance gene,
b. a reporter gene
c. an auxotroph complementing gene,
d. an inducible suicide gene

wherein the choice, formatting and application of such positive selection markers are well known to those skilled in the art. The use of an inducible suicide gene may be used to eliminate parental engineered cells after execution of RMCE in the generation of derivative engineered cells expressing the GOI. A reporter gene may represent a unique TD/CST construct for conditional or constitutive reporting at the cell surface. Multiple selection genes may be included in this portion of the TERS, to enable positive selection during engineered cell line generation, and negative selection during derivative engineered cell line generation.

The 5' and 3' genomic elements minimally encode promoter sequences that mediate the transcription of the TD/CST and selection genes. Such promoters may represent constitutive of inducible promoters. These genomic elements may also encode further genetic elements such as transcriptional enhancers, transcriptional insulators, distinct open reading frames encoding separate desirable components for function or reporting of the TERS construct within the engineered cell, or other unique sequences used for construct tracing and quantification.

It is important to note, that in the present description, the TD/CST is encoded in the sense (5' to 3') direction, and the selection gene in the antisense (3' to 5') direction (as depicted in Figure 3). This is for clarity of description. There is no practical reason why these orientations could not be reversed in the TERS/TEDV context, provided that the TD/CST construct can incorporate a functional intron sequence that also encodes a RMCE site paired with an equivalent RMCE site within the 3' intron fragment of the TEDV.

The inclusion of a selection gene is not absolutely required for the operation of the CSTE system, however, is desirable as to aid in the triple selection for the expression of the GOI interest upon execution of RMCE. Moreover, the generation and

maintenance of the engineered cell line with the use of a reporter in both sense and antisense direction ensures that promoter sequences contained within the 5' and 3' genomic elements are fully functional prior to execution of RMCE to generate the derivative engineered cell expressing the TEDV-encoded GOI.

Cell Surface Tag (CST) architecture

One of the key functional features of the CSTE system is the use of multiple unique CST constructs to conditionally report the presence of the initial TERS construct, and the exchanged construct incorporating TEDV-encoded sequences upon execution of RMCE.

To achieve the TD complementation arrangement within the CSTE system, only certain types of TD exons are suitable. Namely, those that may be spliced with the CST exon situated to the 3', while also permitting the CST protein fragment to be exposed extracellularly.

With the objective to simplify the expressed CST epitope-presenting protein product, and to minimise the size and complexity of the TD exon, TDs from Type II membrane proteins are the most suitable. Type II TDs are single pass, and are situated at the N-terminal end of a protein, which faces the cytoplasmic side of the membrane, allowing a C-terminal CST epitope to be exposed extracellularly. Other TDs may be used, including multi-pass TD, with the proviso that the spliced CST transcript exposes the CST epitope to the extracellular space.

The CST exon, which is adjoined to the TD in the spliced transcript, has the function of simply presenting unique epitopes extracellularly that may be conditionally detected by use of cognate affinity reagents. Generally, such epitopes comprise synthetic sequences, sequences encoded by distinct organisms from that of the host engineered cell, or sequences from the same organism as the host engineered cell but are not expressed extracellularly.

It is generally desirable to include a linker domain between the TD and the CST epitope. Since CST epitopes could comprise of as little as a few amino acids, such a linker region ensures the availability of the CST at the cell surface for affinity reagent engagement. Such linker regions may include flexible 'unstructured' regions, or structural fully folded protein domains.

The nature of the epitopes encoded by the CST exons only reflect uniqueness in the context of epitopes otherwise expressed at the cell surface of the engineered cell containing the TERS, and thus may represent any protein sequence to which a specific affinity reagent may be raised.

In general, the CST construct as a whole should be functionally neutral with regard to cell function. Transmembrane domains simply fused, often via a linker sequence; to inert epitope structures without extraneous functional domains are preferred.

In the present context, an affinity reagent is defined as any antibody, peptide, nucleic acid, or other small molecule that specifically binds to a larger target molecule in order to identify, track, capture, or otherwise influence the activity of the CST epitope. Often such affinity reagents will be labelled with fluorescent, colorimetric, radiometric or other detectable labels for tracking of cells expressing a cognate CST. Alternatively, such affinity reagents may be functionalized to substrates to enable substrate affinity enrichment methods to partition CST-expressing engineered cells.

Engineered cell line containing TERS

The cell line that contains the TERS must be engineered, in as much that the synthetic TERS construct has been inserted. In general, this would mean the integration of the TERS construct to the genome of the engineered cell, but may also include other methods for nuclear maintenance of the TERS construct, such as methods for episomal maintenance of genetic constructs.

Methods to integrate constructs to the genome of target mammalian cells are well known to those skilled in the art, and may be achieved via homology directed recombination (HDR) and/or random integration methods, wherein HDR may be promoted by targeted mutation of the genomic loci at which HDR is to occur, and can be achieved via different means, including but not limited to site directed mutagenesis via

i. zinc-finger nucleases
ii. CRISPR/Cas9 mediated targeting
iii. Synthetic transcription activator-like effector nucleases (TALEN)

wherein said site-directed nucleases induce site-specific DNA-repair by HDR at target loci. After such events, a proportion of cells will have incorporated HDR vector, and can be selected and/or determined via any combination of the following,

iv. Non-destructive phenotypical expression analysis

v. Destructive phenotypical expression analysis

vi. Genetic analysis

Wherein iv and vi are the preferred methods for selection and determination of successful genomic integration events.

Alternatively, viral vectors could be used to deliver the required components in a site-directed or undirected manner.

The methodology of TERS construct integration is not central to the operation of the TERS, provided there is detectable TERS-encoded CST at the cell surface, and presence of an expressed selection gene, indicating that the construct is functional with regard to transcription of contained coding sequences. Indeed, both the TERS-encoded CST and selection genes are convenient selectable markers for the generation of the engineered cell line containing a TERS.

One important aspect to note is the requirement for control of copy number of integrated TERS constructs is required for some applications, especially those leveraging the ability to generate cell-based arrays of derivative engineered cell lines expressing single GOI from a library of TEDV constructs, and similar 'multiplex' methods. Noting this, episomal maintenance of the TERS construct would largely be incompatible with such multiplex methods for derivative engineered cell line generation. Methods for the elucidation of construct copy number within a genome of the engineered cell are well known to those skilled in the art.

TEDV Architecture

The TEDV encodes the CST exon and GOI to be exchanged with the CST exon and selection gene encoded by the TERS upon execution of RMCE, respectively, where the TERS is contained within an engineered cell line. The generic architecture of a TEDV is depicted in Figure 2.

A TEDV typically comprises

1) a 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment.

2) a 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site.

3) an exon comprising the TEDV-encoded CST in the 5' to 3' direction.

4) a transcriptional terminator sequence for encoded CST in the 5' to 3' direction.

5) a transcriptional terminator sequence for the 3' to 5' encoded GOI.

6) a sequence encoding the GOI in the 3' to 5' direction.

7) a Kozak sequence.

8) a 3' RMCE element.

wherein; 1) represents the RMCE element that is situated within the intron non-coding and non-functional sequence of the 3' intron fragment of the TEDV construct, and is equivalent to the 5' RMCE element of the TERS, mediating CST exon exchange upon execution of RMCE within the CSTE system; 2) a 3' intron fragment that provides a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site for execution of efficient splicing of the TD/CST transcript of the integrated TEDV-encoded CST exon after execution of RMCE between TEDV and TERS to generate derivative engineered cell expressing the GOI; 3) is the exon encoding the CST epitope(s) within the TEDV construct, such that the expressed TD / CST protein complemented, subsequent to execution of RMCE between TEDV and TERS, presents said epitope to the extracellular space, and which can be conditionally detected with cognate affinity reagent; 4) represents a transcriptional terminator for the CST exon encoded in the 5' to 3' direction, acting as a transcriptional terminator of the complimented TD/CST with the TEDV-encoded CST upon execution of RMCE; 5) represents a transcriptional terminator for the GOI open reading frame encoded in the 5' to 3' direction within the TEDV, acting as transcriptional termination of the encoded GOI once integrated to the TERS site upon execution of RMCE; 6) represents an open reading frame encoding a GOI to be integrated to the engineered cell to generate a derivative engineered cell

expressing the GOI; 7) is provided to ensure efficient translational initiation from the GOI transcript driven from the TERS subsequent to execution of RMCE; 8) is the RMCE element at the 3' end of the TEDV construct, and equivalent to the 3' RMCE site of the TERS situated between the transcriptional start site of the selection gene and the promoter sequences driving transcription of this open reading frame.

A schematic representation of the TEDV cloning fragment is presented in Figure 2.

The TEDV may represent a synthesised DNA construct, or a cloned construct as produced by methods well known to those skilled in the art. To aid in cloning methodologies, a TEDV may include restriction endonuclease sequences for insertion of varying CST exon and/or GOI sequences into the construct. The inclusion and use of such cloning sites is well known to those skilled in the art.

In the present context, a TEDV will generally be a plasmidic construct propagated in bacteria, and thus may also comprise an origin of replication and a selection gene.

The TEDV may also represent a construct with sequence motifs that enable packaging of said construct into a specific delivery vectors, such as viral vectors well known to those skilled in the art. The use of viral vector transduction of engineered cells containing the TERS would be beneficial for use of the CSTE system in cells that are otherwise difficult to transfect.

The TEDV may represent an RNA construct, which may be reverse transcribed with the provision of appropriate reverse transcriptase.

In the present context, a GOI is defined by any nucleic coding or non-coding sequence of interest. This may include any protein- or polypeptide- open reading frames, non-protein-coding RNA such as microRNA, short hairpin RNA, tRNA or rRNA.

Importantly, a GOI may represent a library of variants of a single open reading frame, and the TEDV-encoding such a GOI may be complied into a pooled library of TEDV. Due to the precise control over copy number enabled by the CSTE system, and the reliable multi-factor and conditional reporting of GOI integration, high-content libraries of derivative engineered cells may be produced expressing just a single GOI in each derivative engineered cell in the pool of target cells. This may be leveraged for a range

of cellular, pathway, protein and specific GOI engineering in a viable cell context (see below).

RMCE elements and enzymes

The use of site-specific recombinases (SSRs) has proven to be a predictable tool to modify the genome of a cell. There are two major classes of SSRs, Ser integrases that comprise of but not limited to, ΦC31, γδ-res, ParA, Tn3, Gin, ΦBT1, R4, Bxb1, TP901-1, and Tyr recombinases, that comprise of but not limited to, Flp, Cre, R. Although both classes are capable of performing RMCE, Ser integrases have the limitation that depending on the initial crossover, thus introduction of either the desired TEDV-encoded CST and GOI or the TEDV vector backbone will occur. Although not all Tyr recombinases have been comprehensively experimentally evaluated, Cre and Flp have been extensively used to perform RMCE. With the objective of integrating a single copy of the TEDV-encoded CST and GOI genetic elements, the use of Flp in conjunction with well characterised heterospecific FRT sites is most suitable because of the lack of pseudo FRT sites encoded in the human genome. This is not the case with Cre which displays promiscuous activity for LoxP genomic pseudosites.

Methods for derivative engineered cell generation

The use of RMCE in mammalian cells is well known to those skilled in the art, where the donor construct may be delivered to the target cells by a range of methodologies, including but not limited to chemical transfection, electroporation or viral vector delivery. In addition to provision of the donor construct, the specific recombinase enzyme(s) must also be provided. This will generally be provided as a separate expression construct co-delivered to the target cells along with the donor construct. In some instances, it may be desirable to modify engineered cells expressing a TERS to conditionally express the required recombinase enzymes to increase the efficiency of the overall process.

In the context of the CSTE system, once the RMCE has been executed by delivery of the TEDV to the engineered cell containing the TERS, and followed by an outgrowth period, the target cell population may be analysed and or selected for generation of derivative engineered cells expressing the TEDV-encoded GOI. This may be conducted in three ways that are intrinsic to the CSTE system, and independent of detection of the GOI expression.

A derivative engineered cell line may be selected on the basis of a negative selection of the TERS-encoded CST, and/or a positive selection of the TEDV-encoded CST, as a reflection of a successful tag-exchange. These are conditionally reported, and require addition of cognate affinity reagents for each CST. Such selection methodologies may leverage FACS through addition of fluorescently labelled affinity reagents and selection of desired CST expression profiles. Due to the extracellular presentation of the CST epitopes, cell partitioning may also be achieved through substrate-based enrichment approaches, such as MACS.

A further negative selection may be achieved on the basis of the TERS-encoded selection gene, where said selection gene represents a reporter gene or an inducible suicide gene. The reporter gene may utilise FACS in the case of a fluorescent reporter gene, for instance, or FACS and/or MACS if the reporter gene itself represents a unique CST. An inducible suicide gene may be used to negatively select parental engineered cells from the culture; for instance, to enrich for derivative engineered cells expressing the GOI.

Overall the double positive selection and single negative selection provide a highly robust selection of derivative engineered cells, which may be reported conditionally, and may be used to partition the cells with substrate-immobilised affinity reagents. This selection may additionally include positive selection on TEDV-encoded GOI, if the nature of said GOI is amendable to non-destructive detection within the derivative engineered cell. The abovementioned selections may be used in series or in parallel to achieve high-throughput or precise high-content derivative engineered cell populations.

These above phenotypic analysis will generally be supported by genotypic confirmation of copy number of TEDV sequences in the derivative engineered cells, utilising methods well known to those skilled in the art. This may also be achieved phenotypically to an extent, considering the highly standardised context of the CST expression, wherein there very rare instances where aberrant integration of the TEDV constructs results in expression of a TEDV-encoded CST at the cell surface, could be negatively selected by exclusion of cells with high TEDV-encoded CST expression, particularly when utilising FACS selection methodologies. In the case of engineered cells that express multiple TERS sites (see below) a copy number control can similarly be achieved based on the degree of TEDV-encoded CST expressed at the cell surface.

Methods for substrate-immobilised affinity reagent partitioning of cells

A key advantage of the CSTE system is the cell surface exposure of the CST epitopes, and thus the ability to partition cells on the basis of substrate immobilised affinity methods. This is particularly useful for high-throughput methods, or stepwise selection in high-content derivative engineered cell library generation. Methodologies well-known to those skilled in the art are the ferromagnetic bead-based MACS approach, among a variety of other affinity reagent-functionalised substrates for capture and physical partitioning of cells based on specific expression of surface epitopes. Such rapid and conditional positive and/or negative selection based on cell surface CST expression is highly parallelisable and thus amenable to high-throughput methodologies, with far more speed and precision that present antibiotic resistance selection, and far more speed and reduced cost compared to FACS selection limited to intracellular fluorescent proteins. Moreover, the ability to positively and negatively select both parental engineered cell and derivative engineered cells leads to precise high-content library generation of derivative engineered cells expressing one or more GOI.

Methods for cell surface 'barcoding' with multiple and multivalent CST

An advantage of the CSTE system is the increased amount of selection marker space available. In contrast to standard fluorescent protein reporter and antibiotic resistance gene systems, which are limited to just a handful of parallelisable markers, many unique markers can be rapidly generated by using simple epitopes with cognate affinity reagents. One is not limited only by the number of CST epitopes and cognate reagents available, as multi-epitope CST domains may readily be constructed for all unique combinations of available CST epitope and cognate detection reagent. The expanded and parallelisable conditional reporter space enables a variety of high-throughput and high-content methodologies.

The use of multiple TEDV-encoded CST epitopes, each associated with a different GOI, or GOI variant sequence, can be used to increase efficiency in high-throughput cell engineering. For instance, a pool of such unique vectors can be integrated into a pool of engineered cells expressing a TERS. These can be bulk handled through negative selection as outlined above, and later partitioned on the basis of unique TEDV-encoded CST expression.

Similarly, the use of multiple TEDV-encoded CST epitopes enables the production of medium-content libraries for viable cell lineage tracking and selection. For instance, a library of GOI associated with unique CSTs can be integrated into a pool of derivative engineered cells, and persistence or function of those cell lineages expressing specific GOI can be traced across time. This can be applied to selections in culture systems for protein, pathway or cellular engineering or analysis, or may similarly be used in instances where cells are grafted to viable animals, and later recovered as viable cells for analysis.

High-content libraries and multi-TERS engineered cells

With the robust positive and negative selection of derivative engineered cells expressing a GOI, precise high-content libraries may be generated in multiple modes to support protein, pathway and cellular engineering workflows.

In the simplest conception, a library of TEDV encoding a single CST epitope, but multiple GOIs, may be integrated into a pool of cells. The application of a standard positive/negative CST selection with the TERS- and TEDV- encoded CST epitopes, for example, would derive a population of derivative engineered cells expressing a single GOI variant in each cell of the selected population. These high-content cell libraries may then be subjected to functional selection in workflows to engineer a protein, a cellular pathway or more generally a cellular function, based on the variant GOI incorporated.

With the addition of multiple TERS with unique CST epitopes to a single engineered cell, and multiple TEDV libraries with unique CST epitopes, multiple families of variant GOI may be incorporated into such engineering workflows, treating enzymes in a biosynthetic pathway as combinatorial variant GOI, for instance. Such a context, where an engineered cell containing multiple TERS, could be used to enhance the overall expression of a single GOI. Multi-TERS engineered cells would generally require the use of unique heterospecific recombinase sites for each TERS/TEDV pair to ensure efficiency and stability of GOI integration and selection.

Figure legends

Figure 1. Composition and operation of the Cell Surface Tag Exchange (CSTE) System

Schematic representation of the CSTE system. The top panel depicts system components, with the Tag Exchange Donor Vector (TEDV), right, and an engineered cell containing Tag Exchange Receiver Site (TERS), left.

The TEDV encodes RMCE elements at either terminus of the construct (open and closed triangles), which are paired with the RMCE sites contained within the TERS. The RMCE element at the 5' end of the TEDV construct (closed triangle) is encoded within a sequence that represents a 3' intron fragment, wherein immediately to the 3' of the RMCE element is are the 3' elements of an intron (open circle), including a branch point sequence, polypyrimidine tracts and splice acceptor site. Thus, the RMCE is contained within a 'non-functional' and non-coding intron sequence, and where the 3' intron fragment contained within the TEDV lacks a 5' splice donor site. Immediately 3' of the splice acceptor site, the TEDV encodes a 3' exon of a cell surface tag (CST), meaning the exon encodes the portion of the CST containing unique molecular binding motifs (grey rectangle). The CST sequence is encoded in the 5' to 3' direction, where the TEDV also encodes the gene of interest (GOI) to be integrated to the TERS encoded in the 3' to 5' direction (crosshatched rectangle).

The central part of the TERS, contained within an engineered cell, encodes elements distinct from those of the TEDV, though with the same architecture. That is, between the RMCE elements, paired with those of TEDV (open and closed triangles), the TERS encodes a CST exon distinct from that of the TEDV-encoded CST (chequered rectangle), with a splice acceptor site and associated 3' intronic sequence immediately to the 5' of this CST (open circle). Similarly, a selection gene (closed rectangle) is encoded in the antisense direction within the TERS, as is the GOI in the TEDV. At the 5' end of the construct, promoter sequences are included (rightward arrow), to drive transcription of the CST. To the 3' of this promoter sequence, a transmembrane domain (TD) exon is encoded (open rectangle), with a 5' intron sequence immediately to the 3' (closed circle). This means that the TD exon and CST exon, encoded in frame with an RMCE element containing intron, are produced as a contiguous transcript that is spliced to adjoin the exons into a single coding mRNA. The TERS-encoded TD-CST product (open/chequered dumbbell) is expressed on the cell surface. At the 3' terminus of the TERS construct is a separate promoter element that drives transcription of the selection gene (closed rectangle) in the 3' to 5' direction, resulting in the expressed selection gene (closed square).

The introduction of the TEDV to the engineered cell containing the TERS, along with an appropriate expression construct for recombinase specific for the paired RMCE elements in the TEDV/TERS, results in the execution of RMCE, and the generation of a derivative engineered cell expressing the GOI (bottom panel). The TERS-encoded elements have exchanged for the TEDV-encoded elements. The derivative cell line thus expresses the TEDV-encoded CST at the cell surface, as a TD-CST product with the originally TERS-encoded TD (open/grey dumbbell), and the GOI (crosshatched square).

Overall, execution of RMCE between the TEDV and TERS results in the generation of a derivative engineered cell that has lost expression of the selection gene and the TERS-encoded CST, and has gained expression of the TEDV-encoded CST and GOI.

Figure 2. The Tag-exchange delivery vector (TEDV) architecture

A schematic representation of the TEDV, depicted as a linear construct with each numbered box representing a key element of the construct architecture. The construct contains both the TEDV-encoded CST, and the gene of interest (GOI).

1) Represents the 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment.

2) Represents functional sequences of the 3' intron fragment, containing a branch point sequence, a polypirimidine tract and a 3' acceptor splice site.

3) Represents an exon encoding the TEDV-encoded CST in the 5' to 3' direction.

4) Represents a transcriptional terminator sequence for encoded CST in the 5' to 3' direction.

5) Represents the transcriptional terminator sequence for the 3' to 5' encoded GOI.

6) Represents the sequence encoding the GOI in the 3' to 5' direction.

7) Represents the Kozak sequence for efficient translational initiation of the GOI transcript.

8) Represents the 3' RMCE element

Figure 3. Tag-exchange receiver site (TERS) open architecture - before tag exchange

A schematic representation of the TERS, depicted as a linear construct with each lettered box representing a key element of the construct architecture. The construct contains both the TERS-encoded CST, and the gene of interest (GOI).

a) Represents a Kozak sequence for efficient translational initiation of the adjoined transmembrane domain (TD) / CST transcript.

b) Represents an exon encoding a Type 2 membrane scaffold protein domain.

c) Represents an 5' intron splice donor site.

d) Represents the 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment, as equivalent, and paired with, the 5' RMCE element of the TEDV.

e) Represents functional sequences of the 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site.

f) Represents an exon encoding the TERS-encoded CST in the 5' to 3' direction.

g) Represents a transcriptional terminator sequence for encoded CST in the 5' to 3' direction.

h) Represents the transcriptional terminator sequence for the 3' to 5' encoded selection gene.

i) Represents the sequence encoding the selection gene in the 3' to 5' direction

j) Represents the Kozak sequence for efficient translational initiation of the selection gene transcript

k) Represents the 3' RMCE element

x) Represents the 5' genomic element responsible for regulating expression and tracking of the CST transcript. This genomic element minimally includes a promoter sequence that drives CST expression.

y) Represents the 5' genomic element responsible for regulating expression and tracking of the CST transcript. This genomic element minimally includes a promoter sequence that drives CST expression.

Figure 4. Exchanged TERS of the derivative engineered cell

A schematic representation of TERS locus after RMCE-mediated exchange with TEDV-encoded elements. As depicted in Figure 1, the elements encoded between the flanking RMCE sites of the TEDV are exchanged with the sequences of equivalent architecture encoded between the flanking RMCE sites of the TERS. This results in an exchanged TERS within the derivative engineered cell line. The lettered and numbered boxes represent the key genetic elements as detailed in Figure 2 and Figure 3, respectively.

Figure 5. Exchanged TEDV by-product

A schematic representation of TEDV by-product after RMCE-mediated exchange with TERS-encoded elements. As depicted in Figure 1, the elements encoded between the flanking RMCE sites of the TEDV are exchanged with the sequences of equivalent architecture encoded between the flanking RMCE sites of the TERS. This results in an unstable exchanged TEDV by-product during generation of the derivative engineered cell line. The lettered and numbered boxes represent the key genetic elements as detailed in Figure 2 and Figure 3, respectively.

Figure 6. Integration of TERS into engineered cells

a) An engineered cell line population, ACL-1163, was created by integration of a TERS cassette encoding a Myc epitope CST and RFP selection gene, into a parental cell line ACL-128 by way of homology directed recombination. 10 days after electroporation, cells were stained with anti-Myc antibody and analysed by flow cytometry for selection markers encoded by TERS, the Myc epitope CST and RFP. The plot displays live single cells as RFP versus Myc, showing that among the transfected cells, there is a population of transfectants that display high signal for both RFP and Myc (Q2 left panel) compared to the parental cells (Q2 right panel). b) The cells with the high RFP and Myc signals were selected and outgrown and a representative engineered monoclone ACL-1163 analysed by flow cytometry. Plots display RFP versus Myc parameters of gated live single cells. The monoclone ACL-1163 has high RFP and Myc signals as expected (Q2).

Figure 7. Execution of an RMCE with TEDV with tag-exchange and delivery of GOI to generate derivative engineered cells

A Flp recombinase mediated tag-exchange was performed in the ACL-1163

engineered cells. ACL-1163 cells harbouring a TERS encoding an RFP selection gene and the Myc epitope CST, were transfected together with TEDV; encoding a SBP epitope CST and a GOI, together with a construct encoding a flp recombinase. 7 days after electroporation cells were stained with anti-Myc and anti-SBP antibodies and analysed by flow cytometry for RFP, Myc and SBP signals. Cells displaying reduced RFP and Myc signal but high signal for SBP surface expression were sorted and expanded as monoclones. a and b) Contour plots showing a representative derivative engineered cell monoclone ACL-3426 (left) compared to parental cells (right). Derivative cells show, loss of the RFP and the Myc signals (Q4 top left panel) compared to the parental cells (Q4 top right panel). The derivative engineered cell monoclone ACL-3426 successfully expresses the TEDV-encoded SBP epitope CST as shown by the increased signal when stained with the anti-SBP antibody (Q5 bottom right panel). These results suggest a successfully performed Flp recombinase mediated tag-exchange, as ACL-3426 cells have lost the signal for the RFP and Myc markers and gained SBP surface expression, as expected due to tag-exchange of the CSTs between the TERS and the TEDV. c) Immunoblot showing three examples of expression of an integrated GOI with C-terminal FLAG-tag. Detection of the GOI (RSV-1 ORFs) expression was achieved by immunoblotting using an antibody against Flag-tag. The parental monoclone line ACL-1163 is included for control. CSTE was performed in three independent experiments whereby cells were transfected with a construct encoding a flp recombinase and either with a TEDV encoding; a SBP epitope CST and a RSV-1 P gene (resulting cell line monoclone is ACL-3374); or TEDV encoding a SBP epitope CST and a RSV-1 N gene (resulting cell line monoclone is ACL-3386); or TEDV encoding a SBP epitope CST and RSV-1 M2 long gene (resulting cell line monoclone is ACL- 3433). Protein was extracted from the monoclones, immunoblotting was performed using mouse anti-Flag antibody. The western blot results demonstrate that each RSV-1 ORF was expressed as evidenced by the presence of a single band in each corresponding cells corresponding to the expected molecular weight of each GOI, and absence of a signal in the parental cell line.

Figure 8. Magnetic affinity cell sorting (MACS) enrichment of a Tag2 (SBP) or depletion of Tag1 (Myc) from a mixed cell population after a successful tag exchange event

This figure demonstrates that the surface tag technology can be used for MACS enrichment of cells carrying Tag2 or depletion of cells carrying Tag1 after a successful tag exchange event from a mixed population of cells. A starting mixed population of

two engineered monoclone cell lines, APL-4535 and APL-3015, was used to either enrich cells expressing the Tag2 (SBP) or deplete Tag1 (Myc) after a successful tag exchange event by using MACS. The TERS within the APL-4535 cells encoded an SBP epitope CST and full-length FLAG tag GOI encoding an intracellular protein; while the TERS within the APL-3015 cells encoded a Myc epitope CST and RFP selection gene. The two cell populations were mixed together at a percentage of 90% (APL-3015) to 10% (APL-4535) cells and MACS was used to either enrich the SBP positive cells (a) or in separate experiment deplete the Myc positive cells (b). a) All cells were labelled with anti-SBP-Alexa 488 fluorophore and incubated with anti-mouse IgG iron beads. The samples were passed through MACS and fractions were collected at three experimental time points – pre-enrichment, flow through and bound. All three fractions were counter stained with anti-c-Myc-Alexa 405 and data was acquired on the BD Influx instrument. The graph displays the percentage of SBP and Myc positive cells in all three experimental fractions. The pre-MACS fraction demonstrates that the starting population of cells consisted of 90% Myc positive cells and 10% SBP positive cells. The successful enrichment of SBP positive cells was demonstrated by the lack of SBP signal in the flow-through fraction (>0.01% SBP positive cells), while following the SBP targeted enrichment 95% of the bound cells were SBP positive. b) In a separate experiment, all cells were labelled with anti-c-Myc-Alexa 405 fluorophore and incubated with anti-mouse IgG iron beads. The labeled cells were depleted using MACS and fractions were collected at three experimental time points as indicated above. All three fractions were counter stained with anti-SBP-Alexa 488 and data was acquired on the BD Influx instrument. The graph displays the percentage of SBP and Myc positive cells after Myc targeted depletion in all three experimental fractions. The pre-MACS fraction again demonstrated that the starting population of cells consisted of 90% Myc positive cells and 10% SBP positive cells. The successful depletion of Myc positive cells was demonstrated by reduced number of Myc positive cells (25%) and increased number of SBP positive cells (75%) in the flow-through fraction. Furthermore, the bound fraction contained > 90% Myc positive cells and < 5% SBP positive cells. c) A total of 546 SBP positive staining individual monoclones was assessed whether they encoded the SBP-linked GOI. The chart shows that 96.52% of the SBP positive cells did integrate the GOI, while 3.48% of the SBP positive cells did not encode the GOI.

The results demonstrate that MACS can be used to either enrich Tag2 (SBP) positive cells or deplete Tag1 (Myc) positive cells from a mixed population of cells. Furthermore, the presence of the Tag2 (SBP) can be used as an indicator of a successful tag exchange and GOI genomic integration.

Figure 9. Composition and operation of the Cell Surface Tag Exchange (CSTE) System using Barcodes

Schematic representation of the CSTE system using barcoding. a) Depicts components of the Tag Exchange Donor Vector (TEDV). Each TEDV encodes RMCE elements at either terminus of the construct (open and closed triangles), which are paired with the RMCE sites contained within the TERS. The RMCE element at the 5' end of the TEDV construct (closed triangle) is encoded within a sequence that represents a 3' intron fragment, wherein immediately to the 3' of the RMCE element are the 3' elements of an intron (open circle), including a branch point sequence, polypyrimidine tracts and splice acceptor site. Immediately 3' of the splice acceptor site, the TEDV encodes a 3' exon of a cell surface tag (CST), meaning the exon encodes the portion of the CST containing unique molecular binding motifs (grey rectangle). The CST sequence is encoded in the 5' to 3' direction, where the TEDV also encodes the gene of interest (GOI) to be integrated to the TERS encoded in the 3' to 5' direction (crosshatched rectangle).

Each CST in this example is composed of two different epitopes from a selection of 3 potential unique epitopes (A, B, C), which can be combined into 6 possible unique combinations. AB is effectively equivalent to BA. In this example, the AX combinations have been assigned to a GOI family, with three variants (GOI a-i, a-ii and a-iii), and the BX combinations have been assigned to a second GOI family with three variants (GOI b-i, b-ii, b-iii). The individual TEDV in this example are pooled. b) Depicts the components of the Tag Exchange Receiver Site (TERS). The TERS RMCE elements are paired with those of TEDV (open and closed triangles). The TERS encodes a CST exon distinct from that of the TEDV-encoded CST (chequered rectangle), with a splice acceptor site and associated 3' intronic sequence immediately to the 5' of this CST (open circle). Similarly, a selection gene (closed rectangle) is encoded in the antisense direction within the TERS, as is the GOI in the TEDV. At the 5' end of the construct, promoter sequences are included (rightward arrow), to drive transcription of the CST. To the 3' of this promoter sequence, a transmembrane domain (TD) exon is encoded (open rectangle), with a 5' intron sequence immediately to the 3' (closed circle). The TERS-encoded TD-CST product (open/chequered dumbbell) is expressed on the cell surface. At the 3' terminus of the TERS construct is a separate promoter element that drives transcription of the selection gene (closed rectangle) in the 3' to 5' direction, resulting in the expressed selection gene (closed square). c) Depicts the introduction of

the TEDV pool to the engineered cell population containing the TERS, along with an appropriate expression construct for recombinase specific for the paired RMCE elements in the TEDV/TERS, resulting in the execution of RMCE, and the generation of a derivative engineered cell population expressing a variety of GOIs. During RMCE, the TERS-encoded elements exchange for the TEDV-encoded elements. The derivative cell population thus expresses the TEDV-encoded CST at the cell surface, as a TD-CST product with the originally TERS-encoded TD (open/grey dumbbell), and the GOI (crosshatched square). d) Following RMCE, the pool of engineered cells expressing GOI can be further isolated into its individual members via FACS, based on the expression of the unique CST barcodes. This can be achieved either as bulk populations of a desired barcode, or individual cell isolation via single cell sorting methods. Alternatively, analysis of the pool can be conducted via FAC without sorting using the barcode as means to identify populations of interest within the digital datasets. e) As a proof of principle for using the CSTE system to barcode engineered cells, ACL-5 and ACL-1 cells were transfected, using chemical transfection (ACL-5) or electroporation (ACL-1) using standard methods known by those skilled in the art, with either a plasmid encoding a CST or a control plasmid without a CST. The CST comprised 3 unique epitopes, FLAG, MYC and HA (as represented by SEQ ID 20). 48 hours following the transfection, the cells were harvested and stained with the cognate antibodies conjugated with a fluorophore: anti-FLAG-PE, anti-MYC-AF647 and anti-HA-AF488. Cells were analysed by flow cytometry; live cells were gated by Forward Scatter (FSC) and Side Scatter (SSC). The Mean Fluorescence Intensity (MFI) of the live cells was determined for each of the 3 epitopes, and the percentage of live cells expressing the respective epitopes. All three epitopes were detected in high proportions and intensity in the cells transfected with the plasmid encoding the barcoded CST, compared to the empty vector, thereby demonstrating the ability of a CST to be composed of multiple epitopes.

The following list of non-limiting embodiments further illustrate the invention:

1. A combined system comprising two separate components, wherein the first component is a tag-exchange donor vector (TEDV) encoding a first cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation, and the second component is an engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a second CST exon adjoined by a full intron sequence to an exon encoding a transmembrane domain, and

also encoding a reporter gene in the antisense orientation, wherein paired recombinase mediated cassette exchange (RMCE) elements are included in the TEDV and TERS such that execution of RMCE between the TEDV and TERS results in exchange of the reporter element for the GOI encoded by the TEDV, and exchange of the first CST exon for the second CST exon, such that the derivative engineered cell now expresses the first CST and GOI, in place of the second CST and the reporter gene.

2. The combined system according to embodiment 1, wherein said first cell surface tag (CST) exon is different to said second CST.

3. The combined system according to any of embodiments 1-2, wherein the first component is a TEDV comprising

a. a first RMCE element
b. a 3' intron fragment
c. a CST exon
d. a first transcriptional terminator
e. a second transcriptional terminator
f. a GOI
g. a Kozak sequence
h. a second RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

4. The combined system according to any one of embodiments 1-3, wherein the first component is a TEDV comprising

a. a first RMCE element – 5'RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment
b. a 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site
c. an exon comprising the TEDV-encoded CST in the 5' to 3' direction
d. a first transcriptional terminator sequence for encoded CST in the 5' to 3' direction
e. a second transcriptional terminator for the 3' to 5' encoded GOI
f. a sequence encoding GOI in the 3' to 5' direction
g. a Kozak sequence

h. a 5' RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

5. The combined system according to any one of embodiments 1-3, wherein the first component is a TEDV containing

a. a first RMCE element
b. a 3' intron fragment
c. a CST exon
d. a first transcriptional terminator
e. a second transcriptional terminator
f. a GOI
g. a Kozak sequence
h. a second RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

6. The combined system according to any one of embodiments 1-5, wherein the second component is a TERS comprising

a. a transcriptional promoter element
b. a Kozak sequence
c. a transmembrane domain exon
d. an intron
e. a first RMCE element
f. a CST exon
g. a first transcriptional terminator
h. a second transcriptional terminator
i. a reporter gene
j. a Kozak sequence
k. a second RMCE element
l. a second transcriptional promoter element

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

7. The combined system according to any one of embodiments 1-6, wherein the second component is a TERS comprising

a. a transcriptional promoter element
b. a Kozak sequence
c. a Type 2 membrane protein transmembrane domain exon
d. a 5' intron splice donor site
e. a 5' RMCE element encoded in the non-coding and 'non-functional' 3' intron fragment, as equivalent, and paired with, the 5' RMCE element of the TEDV
f. functional sequences of the 3' intron fragment, containing a branch point sequence, a polypyrimidine tract and a 3' acceptor splice site
g. an exon comprising the TERS-encoded CST in the 5' to 3' direction (that is different to the TEDV-encoded CST)
h. a transcriptional terminator sequence for encoded CST in the 5' to 3' direction
i. a transcriptional terminator sequence for the 3' to 5' direction
j. a sequence encoding the selection gene in the 3' to 5' direction ~~reporter gene~~
k. a Kozak sequence for efficient translational initiation of the selection gene transcript
l. a 3' RMCE element
m. a 3' genomic element responsible for regulating expression and tracking of the CST transcript,

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

8. The combined system according to any one of embodiments 1-6, wherein the second component is a TERS containing

a. a transcriptional promoter element
b. a Kozak sequence

c. a transmembrane domain exon
d. an intron
e. a first RMCE element
f. a CST exon
g. a first transcriptional terminator
h. a second transcriptional terminator
i. a reporter gene
j. a Kozak sequence
k. a second RMCE element
l. a second transcriptional promoter element

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

9. The combined system according to any one of embodiments 1-8, wherein the first RMCE element of the TEDV is paired with the first RMCE element of the TERS, and the second RMCE element of the TEDV is paired with the second RMCE element of the TERS.

10. The combined system according to any one of embodiments 1-9, wherein each CST exon comprises a sequence encoding one or more molecular affinity tags and that the CST encoded by the TEDV and TERS are different.

11. The combined system according to any one of embodiments 1-10, wherein the engineered cell contains a single TERS in its genome.

12. A method for generating derivative engineered cells expressing a TEDV-encoded GOI from the TERS locus, said method comprising,

a. generating a TEDV encoding a GOI
b. delivering said TEDV to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein
c. contacting cells with two or more affinity reagents specific for both the TEDV-encoded CST and TERS-encoded CST

d. selecting derivative engineered cells on the basis of diminished expression of the reporter gene and TERS-encoded CST, and increased expression of the TEDV-encoded CST, as a proxy for selection of cells with integrated GOI.

13. The method according to embodiment 12, wherein the affinity reagents used in step c. are fluorescently labelled as to detect diminished expression of TERS-encoded CST and increased expression of TEDV-encoded CST, as to enable cell partitioning and selection based on said expression by way of florescence activated cell sorting.

14. The method according to embodiment 12, wherein the affinity reagents used in step c. are immobilised to a substrate such that cells expressing TERS-encoded CST may be depleted or that cells expressing TEDV-encoded CST may be enriched, in the target cell population using substrate affinity methods, such as magnetic activated cell sorting.

15. A method for generating multiple derivative engineered cells expressing a range of TEDV-encoded GOI, from a pool of TEDV, said method comprising,

a. generating a library of two or more TEDV, each encoding a unique GOI sequence, and each with a unique TEDV-encoded CST
b. delivering said library of TEDV as a pool, to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein
c. contacting cells with three or more affinity reagents specific for both the multiple TEDV-encoded CST and TERS-encoded CST
d. selecting derivative engineered cells on the basis of diminished expression of the reporter gene and TERS-encoded CST, and increased expression of each of the unique TEDV-encoded CST

16. A method for cell lineage tracing of derivative engineered cells expressing a range of TEDV-encoded GOI, within a pool of cells as generated by steps a. and b. of embodiment 15, said method comprising,

a. contacting cells with two or more affinity reagents specific for the multiple TEDV-encoded CSTs
b. analysing the content of derivative engineered cells on the basis of expression of each of the unique TEDV-encoded CSTs.

17. A method for cell lineage tracing of derivative engineered cells expressing a range of TEDV-encoded GOI, within a pool of cells as generated by the steps:

- generating a library of two or more TEDV, each encoding a unique GOI sequence, and each with a unique TEDV-encoded CST
- delivering said library of TEDV as a pool, to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein,

said method comprising,

a. contacting cells with two or more affinity reagents specific for the multiple TEDV-encoded CSTs
b. analysing the content of derivative engineered cells on the basis of expression of each of the unique TEDV-encoded CSTs.

18. A tag-exchange donor vector (TEDV) encoding a cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation.

19. The tag-exchange donor vector (TEDV) according to embodiment 18, comprising

a. a first RMCE element
b. a 3' intron fragment
c. a CST exon
d. a first transcriptional terminator
e. a second transcriptional terminator
f. a GOI
g. a Kozak sequence
h. a second RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

20. An engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a cell surface tag (CST) exon adjoined by a full intron sequence to an exon encoding a transmembrane domain, and also encoding a reporter gene in the antisense orientation, wherein recombinase mediated cassette exchange (RMCE) elements are included in the TERS such that execution of RMCE between TERS and a

tag-exchange donor vector (TEDV) results in exchange of the reporter element for a gene of interest (GOI) encoded by the TEDV.

21. The engineered cell according to embodiment 20 wherein said TERS comprises:

a. a transcriptional promoter element
b. a Kozak sequence
c. a transmembrane domain exon
d. an intron
e. a first RMCE element
f. a CST exon
g. a first transcriptional terminator
h. a second transcriptional terminator
i. a reporter gene
j. a Kozak sequence
k. a second RMCE element
l. a second transcriptional promoter element,

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

Materials and methods

Integration of TERS to an engineered cell line

Electroporation was used to deliver the required DNA constructs to generate engineered cells with a single TERS site integrated into the AAVS1 site by way of homology directed recombination.

Per reaction, $4x10^6$ cells were electroporated in 500 ul RPMI 1640 with Glutamax-I (Life Technologies) using the Gene Pulser Xcell™ (Bio-Rad) with the following setting Square Wave 285V, pulse length 12.5 ms and 2 pulses with 1s interval. The DNA concentrations used were 15 ug/ul for the TERS integration vector (V9.F.5), 10 ug/ml for the Cas9-P2A-GFP encoding plasmid (V1.A.8) and 7.5 ug/ml for the vector encoding the gRNA targeting the integration site AAVS1 site (V2.J.6) **(Table 3).**

After electroporation, cells were incubated in culture medium RPMI 1640 with Glutamax-I + 10% FBS (37°C, 5% CO_2) for two days, before analysis.

Sorting of polyclonal GFP-expressing transfectant cells

Cells electroporated with Cas9-P2A-GFP (V1.A.8) or with a plasmid encoding a GFP selection marker (V1.A.4) were sorted for transient GFP expression, using the FACSJAzz™ cell sorter (BD Biosciences). Cells were washed and resuspended in an adequate volume of DPBS before sorting in RPMI 1640 with Glutamax-I with 20% HI-FBS and Anti-Anti 100X (Life Technologies).

Sorting monoclonal cells with stable expression of TERS

FACS was used to obtain single cells constitutively expressing Myc epitope CST and RFP selection gene expressed from the TERS receiving cassette. To detect Myc CST, cells were stained with anti-Myc antibody (anti-c-Myc-Alexa 647, SantaCruz) prior to sorting. Cells were washed in DPBS before sorting in RPMI 1640 with Glutamax-I with 20% HI-FBS and Anti-Anti 100X (Life Technologies).

Table 3: Vectors

ID	Name
V1.A.8	SpCas9-2A-GFP
V9.F.5	AAVS-EF1aL-TxnCS-S-STv5a_G-in:FRT:MYC-RFP:F3-EF1a
V2.J.6	AAVSI_sg-sp-opti_3
V12.A.8	M2-gene-Long_RSV1-FLAG_Tx
V4.I.8	CMVpro_FLPo-sv40pA-V2

GFP and RFP fluorescence was detected on the Influx™ (BD Biosciences) FACS to the filter set listed in **Table 4.** Single cells expressing Myc and RFP were sorted into 96-well plates, containing 200 ul of growth medium to grow a collection of monoclones.

Table 4: FACSJazz and Influx filters

Protein	Fluorochrome	Excitation laser	Detection Filter
Cas9/GFP	GFP	488	530/40
RFP	RFP	561	585/29
Myc	Alexa647	640	670/30
Myc	PE	561	585/29
SBP	Alexa647	640	670/30
SBP	Alexa488	488	530/40
Myc	Alexa405	405	460/50
Flag	PE	561	585/29
HA	Alexa488	488	530/40

Phenotypic screening of monoclonal populations

A sample of 20,000 cells of the outgrown monoclone population was transferred into microtiter plates for analysis, cells were resuspended in 250 ul of DPBS 1X (Life Technologies) and analyzed on the LRSFortessa™ (BD Biosciences). The monoclonal population (ACL-1163) was screened for the presence of the Myc epitope CST and RFP. Myc expression was detected using the anti-Myc antibody labelled with the Alexa Fluor 647 fluorophore. Staining solution was prepared using the recommended antibody volume diluted in 100 ul of staining buffer (DPBS +2% FBS). Cells were incubated for 1 hour at 4°C and then washed twice with 500 ul of staining buffer, prior to analysis.

Genotypic screening of monoclones - confirmation of integration in correct genomic location

ACL-1163 cells were maintained in normal growth medium of RPMI 1640 with Glutamax-I + 10% HI-FBS. The confluence of cells was monitored every day, until they reached 10-12x10^6. DNA was extracted from 5x10^6 cells using the QIAamp DNA Minikit (Qiagen). The remaining cells were further expanded and cryopreserved at a density of 3x10^6 cells/ml, in 70% growth medium + 20% HI-FBS + 10% DMSO.

ACL-1163 monoclones were screened and assessed at a molecular level, this was done by PCR using Q5® Hot Start High-Fidelity DNA Polymerase (NEB), in 20 ul reactions, using the components and reaction conditions listed in tables 5 and 6, respectively. To determine whether TERS integration cassette integrated into the AACS1 locus, primers 15.F.9 and 19.E.7 were used **(Table 7)**, that target respectively, the region before the left homology arm and the transmembrane domain. Correct left homologous arm recombination was indicated by 2.1kb amplicons. Initially, a PCR Master Mix was prepared with all components (Q5® Reaction Buffer, dNTPs, Hot-Start Q5® DNA polymerase, primers Fwd and Rev, 100 ng of DNA template and H_20). PCR reactions were run using C1000 Touch™ Thermal Cycler (Bio-Rad). PCR products were run on a 1% Agarose gel in 1XTAE buffer, using the PowerPac Basic (Bio-Rad), stained with 10,000 dilution of sybersafe and analyzed with Fusion SL (Vilber Lourmat).

Table 5: PCR reagents for assessing integration of the TERS

Reaction Component	Volume per reaction
5xPhusion buffer	4 ul
DNTPs	0,2 ul
Phusion DNA polymerase	0,15 ul
15.F.9	0,5 ul
19.E.7	0,5 ul
H20	up to 20 ul
DNA (100ng)	1 ul (100 ng/ul)
DMSO 3%	0.6 ul

Table 6: PCR cycle conditions

Step	Temperature	Time
Initial Denaturation	98°C	30 sec
30 cycles	98°C	10 sec
	62°C	1:10 min
	72°C	15 sec
Final extension	72°C	10 min

Table 7. Primers

ID	Name	Sequence
1.I.7	turboRFP_GT_F1	GAGAGGCCATTCTCAGATGG
1.I.8	turboRFP_GT_R1	CGGGCATCTTCAGGTTCTTG
1.I.9	turboRFP_probe_FAM	CTACCTGCACTGCTCCTTCAAGACC
10.A.10	TRAC_TCRA-promoter_F1	CTGATCCTCTTGTCCCACAGATA
10.B.6	TRAC_probe (HEX)	ATCCAGAACCCTGACCCTGCCG
15.F.9	AAVS1_GT_F5	ACTCTGCCCTCTAACGCTG
19.E.7	AMPN-TMD_GT_R1	GCTGATGTAGAAGCCCTTGG
21.G.5	ORF-AM_GT_F2	TTCTGTAGCTCCATTGGCAG
21.G.8	ORF-AM_GT_R1	ATCCGTATGGTGACAAGACG

Identification of gene copy number

DNA of selected monoclones was evaluated for a number of TERS cassettes integrated within the cells genome. To achieve this, Droplet Digital PCR (ddPCR) was

performed using primer and probes specific to the TERS cassettes and a reference gene (TRAC) **(Table 8)**. The TERS specific probe was conjugated with FAM, and the reference gene specific probe conjugated with HEX. Integration copy number considered that the ACL-1163 cells are diploid for the reference gene (TRAC). Prior to ddPCR, DNA was digested with MfeI (NEB) to separate tandem integrations. The reaction setup and cycling conditions were followed according to the protocol for ddPCR™ Supermix for Probes (No dUTP) (Bio-Rad), using the QX200™ Droplet Reader and Droplet Generator and the C1000 Touch™ deep-well Thermal cycler (Bio-Rad). Data was acquired using the QuantaSoft™ Software, using Ch1 to detect FAM and Ch2 for HEX.

Table 8: ddPCR Primers/ probes

ID	Name	Sequence
1.I.7	turboRFP_GT_F1	GAGAGGCCATTCTCAGATGG
1.I.8	turboRFP_GT_R1	CGGGCATCTTCAGGTTCTTG
1.I.9	turboRFP_probe_FAM	CTACCTGCACTGCTCCTTCAAGACC
10.A.9	TRAC-TCRA-ex1-F1	CTGATCCTCTTGTCCCACAGATA
10.A.10	TRAC-TCRA-ex1-F1	GACTTGTCACTGGATTTAGAGTCTCT
10.B.6	TRAC-probe (HEX)	ATCCAGAACCCTGACCCTGCCG

Flp-mediated integration of GOI sequences in derivative engineered cell line

Electroporation was used to deliver the required DNA constructs to promote Flp recombinase mediated tag exchange. Per reaction, $4x10^6$ cells were electroporated in 500 ul RPMI 1640 with Glutamax-I (Life Technologies) using the Gene Pulser Xcell™ (Bio-Rad) with the following setting: Square Wave 285V, pulse length 12.5 ms and 2 pulses with 1s interval. The DNA concentrations used were 7.5 µg/ml for the TEDV vector (V9.F.5), 10 ug/ml for the FLPO encoding plasmid (V12.A.8) and 7.5 µg/ml for the vector encoding GFP for tracing DNA delivery (V1.A.4) **(Table 3).**

After electroporation, cells were incubated in culture medium RPMI 1640 with Glutamax-I + 10% FBS (37°C, 5% CO_2) for two days, before analysis and cell sorting

of GFP positive cells.

Phenotyping for tag exchange

To determine if Flp recombinase mediated tag exchange occurred, cells were stained for surface expression of Myc and SBP and measured for RFP fluorescence intensity. 7-10 days after electroporation, the cells were harvested, and surface stained for SBP and Myc using the following antibodies (anti-SBP-Alexa647and anti-c-Myc-AlexaPE, SantaCruz). GFP and RFP fluorescence was detected on the Influx™ (BD Biosciences) FACS to the filter set listed in **Table 4.**

Single cells expressing SBP but not Myc and RFP were sorted into 96-well plates, containing 200 ul of growth medium to grow a collection of monoclones.

Phenotyping of monoclones was performed 20-24 days post single cell sort. For flow cytometric analysis, the cells were transferred from the wells and 300 µl RPMI was added per tube. Cells were centrifuged for 3 min at 400 g, at 4°C, Supernatant was aspirated and cell pellets were resuspended in 25 µl stain mix or RPMI (unstained controls) (Stain mix: anti-SBP-Alexa647 and anti-c-Myc-AlexaPE) and incubated for 30 min at 4°C. Cells were washed twice with staining buffer (SB) (DBPS + 2% FBS) and centrifuged for 3 min at 400 g. Cells were resuspended in 200ul of SB and transferred to 96-well plates for data acquisition on LSRFortessa. Analysis was performed using FlowJo.

Confirmation of GOI expression

Cells were grown and following harvest the cells were lysed with 150 mM NaCl, 50 mM Tris pH 8, 1% CHAPS, 5mM Imidazole, 1mM PMSF, 1x protease and phosphatase inhibitors (Thermo) for 20 minutes on a rotor at 4˚C. Lysates were cleared by centrifugation at 17000g for 10 minutes at 4˚C and subjected to sodium Dodecyl Sulphate (SDS) gel electrophoresis on 10% acrylamide pre-cast gel (Biorad) at 140V for 1 hour. Gels were turbo blotted onto PVDF membranes (Biorad), which were then blocked for 15 minutes in Sea block 1x (Thermo) in Tris buffered saline/ tween 20 (TBST) 1x and incubated with mouse anti-flag antibody (Sigma) for 2 hours at room temperature. Membranes were washed from unbound primary antibody 3x for 5 minutes in TBST 1x and incubated with anti-mouse horseradish peroxidase (HRP)-conjugated goat antibody for 1 hour at room temperature. Membranes were finally washed from unbound secondary antibody 3x for 5 minutes in TBST 1x and finally, the

HRP signal was developed with ECL substrate (Biorad) and acquired with Fusion SL Vilber system.

Confirmation of GOI genomic integration

Monoclone cell lines expressing cell surface Tag 2 (SBP) were assessed at a molecular level for the cointegration of a GOI coding sequence. This was done by PCR using Q5® Hot Start High-Fidelity DNA Polymerase (NEB), in 30 ul reactions, using the components and reaction conditions listed in tables 9 and 10, respectively. To determine whether a GOI integrated into the genome, primers 12.G.5 and 21.G.8 were used **(Table 7)**, that target the 3´UTR region of the GOI. Initially, a PCR Master Mix was prepared with all components (Q5® Reaction Buffer, dNTPs, Hot-Start Q5® DNA polymerase, primers Fwd and Rev, 100 ng of DNA template and H_20). PCR reactions were run using C1000 Touch™ Thermal Cycler (Bio-Rad). PCR products were run on a 1% Agarose gel in 1XTAE buffer, using the PowerPac Basic (Bio-Rad), stained with 10,000 dilution of sybersafe and analyzed with Fusion SL (Vilber Lourmat). A band of correct size was confirmed to encode sequence of the GOI 3´UTR by sanger sequence.

Table 9: PCR mix

Reagent	Per/reaction (µL)	Master mix for 96 samples
5XPhusion HF buffer	6	660
DNTPs	0.3	33
Phusion Hot Star pol	0.3	33
21.G.5 ORF-AM_GT_F2 (100µM) stock	0.15	55
21.G.8 ORF-AM_GT-R1 (100µM) stock	0.15	55
H20	20.4	2244
Cell sample	2	NA

Table 10: PCR conditions

Step	Temprature	Time	Number of cycles
Initial Denaturation	98˚C	30 sec	1
Denaturation	98˚C	10 sec	35
Annealing	62˚C	20 sec	
Extension	72˚C	10 sec	
Finial Extension	72˚C	5 min	1

Hold	12˚C	hold	1

MACS enrichment/ depletion of Surface Tag expressing cells

The manufacturer's MACS protocol for enrichment / depletion was generally followed (Miltenyi Biotec, #130-047-101,IM0001377.PDF).

Sample preparation

Cells were harvested and washed once by centrifugation (300xg at 4°C for 3 min) in stain buffer-M (SB-M – cold Dulbecco's phosphate buffered saline (DPBS), 2% FBS, 2mM EDTA). SB-M resuspended cells were filtered through a 40-μm cell strainer to obtain single cell suspension. Cells were washed with 3ml cold SB-M and cell pellet was collected. 80 μl staining solution, containing appropriate antibodies or dead cell removal reagent, was used per 10 million cells and the sample was incubated at 4°C for 30 min. Cells were washed twice with 3ml SB-M buffer and pelleted.

Magnetic bead labelling

The pelleted cells were labelled with magnetic beads by resuspension of the cells in 160 μl SB-M and addition of anti-mouse IgG1 MACS MicroBeads (Miltenyi) (40 μl MACS MicroBeads were added per 10 million cells). After incubation at 4°C for 20 min, cells were washed with 3ml SB-M and finally resuspended in 500ul SB-M.

Magnetic separation

A LS column (Miltenyi) was placed in the magnetic field of a suitable MACS separator (Miltenyi) and rinsed wit 3ml SB-M. A pre MACS fraction was also collected. Cell suspension was added to the column and flow-through fraction containing unlabelled cells was collected into a 15 ml conical tube (this is referred to as the flow-through fraction). The column was removed from the separator and placed in a suitable collection tube. 5ml SB-M was added into the column containing the magnetically labelled cells. Using the supplied plunger, pressure was applied until the plunger reached the bottom of the column. The magnetically labelled cells were eluted from the column (this is referred to as the bound fraction) and the fraction was used in downstream applications.

Example 1 - Integration of TERS into an engineered cell line

This example describes the stable integration of a TERS into a cell line to produce an engineered cell line monoclone ACL-1163 containing a single TERS in the genome.

In this example, the TERS presented as SEQ ID 1, comprised of the following selected genetic elements that encode two genes. The first gene encoded in the sense direction, comprises of a EF1a promoter upstream of an ORF encoded across two exons. The first exon encodes a Transmembrane Type II protein domain (TD) and the second encodes the Myc epitope tag. The intron between the two exons was derived from the human GAPDH gene, and modified to encode the first heterospecific FRT site (FRT) between the 5' intron splice donor site and intron branch point sequence. 3' end of the ORF encodes a SV40 polyadenylation signal terminator. The second gene encoded in the antisense direction comprises of a EF1a promoter upstream of an ORF encoding the fluorescent reporter, RFP. The region between the Kozak sequence and the promoter encodes the second heterospecific FRT site (F3). 3' end of the RFP ORF encodes a bGHpA polyadenylation signal terminator.

To promote the stable genomic integration of the TERS into the genomic safe harbor locus, AAVS1, a plasmid was constructed, wherein; the DNA elements of the TERS were flanked with AAVS1 left and right homology arms. Each arm comprised of >500 bp of sequence homologous to the AAVS1 genomic locus. Stable integration of TERS was achieved through the process of homology directed recombination (HDR) at the genomic safe harbor locus, AAVS1.

The ACL-128 cell line was transfected with a plasmid that encoded the TERS genetic elements flanked by AAVS1 left and right homology arms, a plasmid that encoded an optimal gRNAs targeting the AAVS1 locus, and a plasmid that encoded Cas9-P2A-GFP. Cells positive for Cas9-P2A-GFP plasmid uptake were FACS sorted based on GFP fluorescence, 2 days after. The GFP sorted cells were further expanded for greater than 7 days. The TERS-transfected cells were stained with an anti-Myc antibody and analyzed by flow cytometry for presence of RFP and Myc epitope CST (Figure 6a). The cells that have integrated the TERS into their genome showed increased RFP and Myc signals. These cells were sorted and expanded to represent a collection of monoclones. A representative monoclone, ACL-1163 shown in Figure 6b, demonstrated constant and strong expression of RFP and Myc surface expression. To determine that the genomic receiver cassette integrated into the targeted AAVS1 site, genomic DNA was extracted from the selected ACL-1163 cell line and a PCR reaction was performed with a primer internal to the TERS receiving cassette and a primer specific to the AAVS1 locus (15.F.9 and 19.E.7, see Table 7). A PCR amplicon of expected size was detected (data not shown). Furthermore, ddPCR with the primers and probe (forward primer 1.I.7, reverse primer 1.I.8, probe 1.I.9, see Table 8) confirmed that only a single copy of the TERS receiving cassette was integrated (data not shown).

The resulting engineered cell line, ACL-1163, contained a single copy of the TERS designed for RMCE with suitable paired TEDV.

Example 2 - Execution of an RMCE with TEDV with tag-exchange and delivery of GOI to generate derivative engineered cells

This example demonstrates execution of the CSTE system that is the RMCE driven reaction between TEDV-encoded sequences and TERS results in a switch of CSTs, on the cell surface, thus reporting the exchanged construct incorporating TEDV-encoded sequences and the cargo GOI. In the present example, the above described ACL-1163 was used as the target engineered cell line.

The TEDV in the current example, encodes in the sense direction the first heterospecific FRT site (FRT), a 3' intron fragment, containing a branch point sequence, a polypyrimidine track and a 3' acceptor splice site; an exon encoding streptavidin binding peptide (SBP) and a SV40 polyadenylation signal terminator. Three separate GOI, originating from the respiratory syncytial virus (RSV), were encoded in the antisense direction in separate TEDV, each GOI ORF located between the second heterospecific FRT site (F3) and 3' bGHpA polyadenylation signal terminator. The sequences SEQ ID 3-SEQ ID 5 represent the three independent SBP epitope CST and GOI encoding TEDV constructs used.

In the present example, the engineered cell line, ACL-1163 constructed in example 1 was electroporated with a TEDV (selected from sequences SEQ ID 3 -SEQ ID 5) and with a vector encoding expression of the RMCE recombinase enzyme (FLPO, V4.1.8, Table 3, SEQ ID 2). Cells were incubated for 7-10 days to allow for the integration couple to occur and then were stained with anti-Myc and anti-SBP antibodies and analyzed by flow cytometry for RFP, Myc and SBP reporter signals. Cells displaying reduced RFP and Myc signals but increased SBP signal, indicating a 'tag exchange', were sorted and expanded to represent a collection of monoclones. Characterization of a representative monoclone, ACL-3426, is depicted in figure 7a and 7b. To confirm a successful surface tag exchange, the monoclone ACL-3426 was stained with an antibody against Myc and an antibody against SBP and analyzed by flow cytometry for the loss of RFP and Surface Myc expression (Figure 7a) and gain of SBP signal (Figure 7b). Indeed, the cells that performed the CSTE showed successful tag-exchange by displaying SBP and not Myc or RFP signals. Parental cells were analysed in parallel and showed persistent high signal for both RFP and Myc (Figure 7a right panel) and a low SBP signal (Figure 7b right panel).

To demonstrate that following CSTE, the GOI ORF integrated and is being expressed, three monoclones from independent experiments using each of the TEDV described above, encoding distinct RSV-1 GOIs were assessed by immunoblot (Figure 7c). The GOI ORFs each encoding a Flag-tag, donated by the TEDVs were 3 genes from the RSV-1; P, N genes and M2 gene (long). CSTE was performed whereby the cells were transfected with a construct encoding a flp recombinase and either with TEDV; a SBP epitope CST and a RSV-1 P gene (resulting cell line monoclone is ACL-3374); or TEDV encoding a SBP epitope CST and a RSV-1 N gene (resulting cell line monoclone is ACL-3386); or TEDV encoding a SBP epitope CSTand RSV-1 M2 long gene (resulting cell line monoclone is ACL- 3433). Protein was extracted from the monoclones, and the samples were immunoblotted using mouse anti-Flag primary antibody, followed by an incubation with an anti-mouse horseradish peroxidase (HRP)-conjugated goat antibody. The HRP signal was developed with ECL substrate. A successful CSTE was demonstrated by the presence of a single band positive for flag of expected molecular weight for each of the GOI (Figure 7c lanes 2-4). Parental cells were analyzed in parallel and showed absence of the Flag-tag signal (Figure 7c lane 5).

In summary, this example demonstrates that the cell surface tag-exchange can be used to conditionally report the presence of the initial TERS construct, and the exchanged construct incorporating TEDV-encoded sequences upon execution of RMCE, and report GOI integration and expression independently of detection of the GOI itself.

Example 3 – Magnetic affinity cell sorting (MACS) enrichment of a Tag2 (SBP) or a depletion of Tag1 (Myc) from a mixed engineered cell population after a successful tag exchange event

This example demonstrates that the surface tag technology can be used for MACS enrichment of a Tag 2 (SBP) or depletion of Tag1 (Myc). Additionally, the presence of Tag2 (SBP) can be used to monitor the incorporation of the gene of interest (GOI) after a successful tag exchange event. A starting mixed population of two engineered monoclone cell lines, APL–3015 and APL–4535, was used. The cell populations were mixed as 90% Myc positive APL-3015 cells and 10% SBP positive APL-4535 cells. The Tag exchange receiver site (TERS) within the APL-4535 cells encodes an SBP epitope CST and a full-length FLAG tag GOI encoding an intracellular protein, while the TERS within the APL-3015 cells encoded a Myc epitope CST and RFP selection gene. Magnetic affinity cell sorting (MACS) was used to enrich the SBP positive APL-4535 cells from the mixed population. In a separate experiment, MACS was used to deplete the Myc positive APL-3015 cells from the mixed population.

In the first instance, the mixed cell population was labeled with an anti-SBP-Alexa 488 fluorophore and subsequently incubated with anti-mouse IgG iron beads. The SBP labeled cells were enriched using MACS and counter stained with anti-c-Myc-Alexa 405 fluorophore. Figure 8a depicts a percentage of SBP and Myc positive cells in three experimental steps, 1) pre-MACS to ensure the starting ratio of SBP and Myc positive cells in the mixed cell population, 2) flow through to assess the percentage of SBP and Myc positive cells that were not captured by the MACS column when the column is placed in a magnetic field and 3) bound fraction is used to assess the percentage of SBP and Myc positive cells that were captured by the MACS column. All three fractions were counter stained with anti-c-Myc Alexa 405 fluorophore and data was acquired on the BD Influx instrument. Figure 8b demonstrates the successful enrichment of SBP positive cells by the lack of SBP positive cells in the flow through fraction (<0.01), while following the SBP targeted enrichment 95% of the bound cells were SBP positive.

To demonstrate the use of the surface tag technology to MACS-deplete base cell line expressing Tag1 (Myc) from a mixed cell population all cells were labelled with anti-c-Myc-Alexa 405 fluorophore and incubated with anti-mouse IgG iron beads. The labeled cells were depleted using MACS and fractions were collected as indicated. All three fractions were counter stained with anti-SBP-Alexa 488 and data was acquired on the BD Influx instrument. The successful depletion of Myc positive cells was demonstrated by reduced number of Myc positive cells (25%) and increased number of SBP positive cells (75%) in the flow-through fraction.

Additionally, to demonstrate that the presence of Tag2 (SBP) could be used as a reporter to monitor the incorporation of the GOI after a successful tag exchange event, a total of 546 SBP positive individual monoclones were assessed whether they encoded the SBP-linked GOI. The chart in Figure 8c shows that 96.52% of the SBP positive cells did encode the GOI, while 3.48% of the SBP positive cells did not express the GOI.

The results demonstrate that the surface tag technology is suitable for enrichment of monoclones expressing Tag2 (SBP) or depletion of unmodified base lines (monoclones expressing Tag1 (Myc)) from a mixed cell population after a successful tag exchange. Furthermore, the presence of Tag2 (SBP) correlated with the incorporation of the GOI after a successful tag exchange event and therefore presence of Tag2 (SBP) on the cell surface can be used as an indication that the GOI has been successfully delivered into the TERS within the engineered cell line.

Example 4 – Composition and operation of the Cell Surface Tag Exchange (CSTE) System using barcodes.

This example describes a schematic representation of the concept of using the CSTE system for barcoding engineered cells expressing a GOI.

Figure 9 is a schematic representation of using the CSTE system to barcode cells expressing the GOIs. Panels a and b depict the system components, with a pool of Tag Exchange Donor Vectors (TEDV), and an engineered cell containing the Tag Exchange Receiver Site (TERS), respectively.

Each TEDV encodes RMCE elements at the 5' and 3' termini of the construct, which are paired with the RMCE sites contained within the TERS. The RMCE element at the 5' end of the TEDV construct is encoded within a sequence that represents a 3' intron fragment, wherein immediately to the 3' of the RMCE element are the 3' elements of an intron, including a branch point sequence, polypyrimidine tracts and splice acceptor site. Thus, the RMCE is contained within a 'non-functional' and non-coding intron sequence, and where the 3' intron fragment contained within the TEDV lacks a 5' splice donor site. Immediately 3' of the splice acceptor site, the TEDV encodes a 3' exon of a cell surface tag (CST), meaning the exon encodes the portion of the CST containing unique molecular binding motifs. The CST sequence is encoded in the 5' to 3' direction, where the TEDV also encodes the gene of interest (GOI) to be integrated into the TERS encoded in the 3' to 5' direction.

Each CST in this example is composed of two different epitopes from a selection of 3 potential unique epitopes (A, B, C), which can be combined into 6 possible unique combinations. As the arrangement of each epitope is difficult to distinguish via current techniques, AB is effectively equivalent to BA. In this example, the AX combinations have been assigned to a GOI family, with three variants (GOI a-i, a-ii and a-iii), and the BX combinations have been assigned to a second GOI family with three variants (GOI b-i, b-ii, b-iii). The individual TEDVs in this example are pooled (Figure 9a).

The central part of the TERS, contained within an engineered cell, encodes elements distinct from those of the TEDV, though with the same architecture. That is, between the RMCE elements, paired with those of TEDV, the TERS encodes a CST exon distinct from that of the TEDV-encoded CST, with a splice acceptor site and associated 3' intronic sequence immediately to the 5' of this CST. Similarly, a selection gene is encoded in the antisense direction within the TERS, as is the GOI in the TEDV. At the 5' end of the construct, promoter sequences are included, to drive transcription of the CST. To the 3' of this promoter sequence, a transmembrane domain (TD) exon is encoded, with a 5' intron sequence immediately to the 3'. This means that the TD exon and CST exon, encoded in frame with an RMCE element containing intron, are produced as a contiguous transcript that is spliced to adjoin the exons into a single coding mRNA. The TERS-encoded TD-CST product is expressed on the cell surface. At the 3' terminus of the TERS construct is a separate promoter element that drives transcription of the selection gene in the 3' to 5' direction, resulting in the expressed selection gene (Figure 9b).

The introduction of the TEDV pool to the engineered cell population containing the TERS, along with an appropriate expression construct for recombinase specific for the paired RMCE elements in the TEDV/TERS, results in the execution of RMCE, and the generation of a pool of derivative engineered cells expressing the GOIs (Figure 9c). The TERS-encoded elements have exchanged for the TEDV-encoded elements. The derivative engineered cells thus express the TEDV-encoded CSTs at the cell surface, as a TD-CST product with the originally TERS-encoded TD, and the GOI.

Overall, execution of RMCE between the pool of TEDV and TERS results in the generation of a derivative pool of engineered cells that has lost expression of the selection gene and the original TERS-encoded CST, and each cell in the pool has gained expression of one of the members of the TEDV-encoded CST and GOI from the TEDV pool. The pool of engineered cells expressing GOI can be further analysed/ isolated into its individual members via FACS, for example, being isolated based on the expression of the unique CST barcode (Figure 9d). This can be achieved either as bulk populations of a desired barcode, or individual cell isolation via single cell sorting methods, for instance. Alternatively, analysis of the pool can be conducted via FAC without sorting using the barcode as a means to identify populations of interest within the digital datasets, with secondary analysis of cellular function, for example, to correlate variant GOI expression or cellular function.

To demonstrate the concept that the CST can be composed of multiple epitopes, ACL-1 and ACL-5 cells were transfected with either a plasmid encoding a CST or a control plasmid without a CST (Figure 9e). The CST was comprised of 3 unique epitopes, FLAG, MYC and HA and is represented by sequence SEQ ID 20. Cells were transfected using chemical transfection (ACL-5) or electroporation (ACL-1) using standard methods known by those skilled in the art. After 48 hours cells were harvested and stained with the cognate antibodies conjugated with a fluorophore: anti-FLAG-PE, anti-MYC-AF647 and anti-HA-AF488. Cells were analysed by flow cytometry, live cells were gated by Forward Scatter (FSC) and Side Scatter (SSC). The Mean Fluorescence Intensity (MFI) of live cells was determined for each of the 3 epitopes, and the percentage of live cells expressing the respective epitopes. All three epitopes were detected in high proportions and intensity for the cells in the sample transfected with the plasmid encoding the barcoded CST, compared to the empty vector, thereby demonstrating a CST composed of multiple epitopes.

The results demonstrate that the surface tag technology is suitable for barcoding cell lines as the CST can be composed of multiple epitopes.

List of abbreviations

AAVS1	Adeno-associated virus integration site 1
APC	Antigen-presenting cell
Cas9	CRISPR-associated gene 9
CMV	Cytomegalovirus
cre	Cre recombinase
CRISPR	Clustered regularly interspaced short palindromic repeats
CST	Cell surface tag
CSTE	Cell surface tag exchange
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate buffered saline
DSB	Double-strand break
dUTP	Deoxyuridine Thiphosphate
EDTA	Ethylenediaminetetraacetic acid
EF1 alpha	Elongation factor alpha (for eukaryotic translation)
FACS	Fluorescence-activated cell sorting
FAM	Fluorescein amidite
FBS	Fetal bovine serum
FLP	Flippase
FRT	Flippase recognition target
GFP	Green fluorescent protein
GOI	Gene of interest
gRNA	Guide ribonucleic acid
HDR	Homology directed recombination
HLA	Human leukocyte antigen
IRES	Internal ribosomal entry site
MACS	Magnetic-activated cell sorting
NEB	New England biolabs
NHEJ	Non-homologous end joining
ORF	Open reading frame

PCR	Polymerase chain reaction
RFP	Red fluorescent protein
RMCE	Recombinase mediated cassette exchange
RPMI	Roswell Park Memorial Institute
RSV	Respiratory syncytial virus
RT	Reverse Transcription
RNA	Ribonucleic acid
SBP	Streptavidin binding peptide
SSR	Site-specific recombinase
SV40	Simian virus 40
SV40pA	Simian virus 40 poly (A)
TAA	Tumour-associated-antigens
TALEN	Transcription-like effector nuclease
TAE	Tris-acetate-EDTA
T-cells	T lymphocytes
TCR	T-cell Receptor
TCS	Target coding sequence
TD	Transmembrane domain
TEDV	Tag-exchange donor vector
TERS	Tag-exchange receiver site
rRNA	Ribosomal RNA
tRNA	Transfer RNA
UTR	Untranslated region
ZNF	Zinc finger nuclease

List of definitions

Amplicon: a piece of DNA or RNA that is the source and/or product of artificial amplification using various methods including PCR.

Antibody: Affinity molecule that is expressed by specialized cells of the immune system called B-cells and that contains of two chains. B-cells express a very large and very diverse repertoire of antibodies that do generally not bind self proteins but can bind and neutralize pathogens or toxins that would threaten the host. Natural or artificially engineered antibodies are often used as affinity reagents.

Auxotroph: a mutant organism (especially a bacterium or fungus) that requires a particular additional nutrient which the normal strain does not.

Cis-acting element: regions of non-coding DNA that regulate the transcription of nearby ORFs.

CST: co-integrated cell surface tag that allows reporting of an integrated gene of interest

CSTE system: a system that operates as a donor/ receiver pair, wherein the tag-exchange donor vector acts to deliver a gene of interest sequence, and a cell-surface tag exon to a paired tag-exchange receiver site contained within the genome of an engineered cell line.

Derivative Engineered Cell: an engineered cell, that has been further genetically modified to exchange CST and integrate GOI

DNA: Desoxyribonucleic acid. Chemical name of the molecule that forms genetic material encoding genes and proteins.

Engineered Cell: A cell whereby the genome has been engineered through genetic modification.

Epitope: Region on an antibody target that is bound by an antibody or other affinity reagent.

Eukaryotic conditional regulatory element: A DNA sequence that can influence the activity of a promoter, which may be induced or repressed under defined conditions

Eukaryotic Promoter: A DNA sequence that encodes a RNA polymerase binding site and response elements The sequence of the promoter region controls the binding of the RNA polymerase and transcription factors, therefore promoters play a large role in determining where and when your gene of interest will be expressed.

Eukaryotic terminator/Signal terminator: A DNA sequence that are recognized by protein factors that are associated with the RNA polymerase II and which trigger the termination process of transcription. It also encodes the poly-A signal

FACS/Flow Cytometry: Fluorescence-activated cell sorting. Flow cytometry is a technique by which individual cells can be analyzed en masse for the expression of specific cell surface and intracellular markers. A variation of that technique, cell sorting, allows cells that carry a defined set of markers to be retrieved for further analysis.

Flippase: A recombinase (Flippase, Flp) derived from the 2 µm plasmid of baker's yeast *Saccharomyces cerevisiae*.

Fluorescent (protein) marker: Molecule that has specific extinction and emission characteristics and can be detected by Microscopy, FACS and related techniques.

Gene cis acting elements: are present on the same molecule of DNA as the gene they regulate whereas trans-regulatory elements can regulate genes distant from the gene from which they were transcribed. Cis-regulatory elements are often binding sites for one or more trans-acting factors.

Genetic barcoding: DNA barcoding is a taxonomic method that uses a short genetic marker in an organism's DNA to identify it as belonging to a particular species.

GOI: gene of interest defined as any nucleic acid coding or non-coding sequence of interest.

Heterospecific recombinase sites: A DNA sequence that is recognized by a recombinase enzyme to promote the crossover of two DNA molecules

Homologous arms: A stretch of DNA that has near identical sequence identity to a complement homologous arm and therefore promote the exchange of two DNA molecules by the cellular process, homology directed repair.

Insulator: A DNA sequence that prevents a gene from being influenced by the activation or repression of nearby genes. Insulators also prevent the spread of heterochromatin from a silenced gene to an actively transcribed gene.

Integration: The physical ligation of a DNA sequence into a chromosome of a cell

Internal ribosome entry site (IRES): A DNA sequence that once transcribed encodes a RNA element that allows the initiation of translation in a cap-independent manner

Intron: Noncoding sections of an RNA transcript, or the DNA encoding it, that are spliced out before the RNA molecule is translated into a protein

Intron branch point sequence: branch point nucleotide that initiates a nucleophilic attack on the 5′ donor splice site. The free end of the upstream intron then initiates a second nucleophilic attack on the 3′ acceptor splice site, releasing the intron as an RNA lariat and covalently combining the two exons

K is a nucleotide code indicating Keto (K = G or T)

Kozak Sequence: Short sequence required for the efficient initiation of translation

M is a nucleotide code indicating aMino (M = A or C)

MACS: Magnetic activated cell sorting: Cellular isolation technique in which cells are labelled with affinity molecules that contain magnetic particles for separation using magnetic fields.

Matched: When two components encode genetic elements that direct and restrict the interaction between the complemented components

Monoclone cell line: A defined group of cells produced from a single ancestral cell by repeated cellular replication. **N** is a nucleotide code indicating aNy nucleotide (N = A, T, C or G)

Native: an entity that is naturally occurring to the cell

Negative Selection Marker: A selectable marker that confers negative selection of a vector and/or of host organism carrying said marker-bearing vector

Non-coding gene: A non-protein coding DNA sequence that is transcribed into functional non-coding RNA molecules

Origin of replication: a particular sequence in a vector, plasmid or genome at which replication is initiated.

ORF: Open reading frame. Stretch of genetic material that encodes a translation frame for synthesis of a protein (polypeptide) by the ribosome

Overhang: A single stranded sequence at the terminus of a double stranded nucleic acid molecule. Often referred to as sticky or cohesive ends.

PCR: Polymerase chain reaction in which a specific target DNA molecule is exponentially amplified

Peptide: short string of amino acids, typically between 6 - 30 amino acids in length

Phenotypic analysis: Analysis of the observable characteristics of a cell.

Plasmid: A genetic construct can replicate independently of the chromosomes, typically a small circular DNA strand in the cytoplasm of a bacterium or protozoan.

Polypeptide: Protein consisting of a stretch of peptides, forming a three-dimensional structure.

Polypirimidine motif: (C_nT_n) motif high in pyrimidines and is present upstream of the CAG intron 3' end.

Positive Selection Marker: A selectable marker that confers positive selection of a vector and/or host organism carrying said marker-bearing vector

Primer: Short DNA sequence that allows specific recognition of a target DNA sequence for example during a PCR.

Promoter: Regulatory DNA element for the controlled initiation of gene expression

Recombinase: Enzymes that mediate genetic recombination, catalyses RMCE.

Reporter Element: A genetic element that mediates a reported signal in the organism or vector bearing said element. May be used as a positive or negative selection maker.

Restriction Enzyme Cleavage Sequence: The genetic sequence cleaved by a restriction enzyme, which can be intrinsic or intrinsic to the recognition sequence of said restriction enzyme

Restriction Enzyme Recognition Sequence: The genetic sequence recognised and engaged by a restriction enzyme

RMCE: Recombinase-mediated cassette exchange. Exchange of genetic material at the genomic receiver site catalysed by a recombinase.

Slice acceptor site: A DNA sequence at the 3' end of the intron AM, APX CM or affinity reagent for interaction with cells with TCRsp on the surface, or TCRsp based reagents

Slice donor site: A DNA sequence at the 5' end of the intron

Suicide gene: A gene that will mediate cell death within the host organism carrying said gene. May be used as a positive or negative selection marker.

Synthetic: an entity that is artificially generated

TEDV: tag-exchange donor vector paired to the tag-exchange receiver site contained within the genome of an engineered cell. It is used to deliver a gene of interest and a cell-surface tag exon

TERS: paired tag-exchange receiver site contained within the genome of an engineered cell line

Type II transmembrane domain: single non-cleavable transmembrane stretch of hydrophobic residues close to the N terminus which serves as a combined signal/ anchor sequence, with the N terminal portion on the interior of the membrane and the C terminal portion exposed on the cell exterior or in the ER lumen.

Vector: A vector is a genetic construct that carries genetic information. In the present context vector usually describes plasmidic DNA vectors. A vector can represent any such construct that can be propagated and selected in a host organism.

W is a nucleotide code indication Weak (W = A or T)

Claims

1. A combined system comprising two separate components, wherein the first component is a tag-exchange donor vector (TEDV) encoding a first cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation, and the second component is an engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a second CST exon, which is different to the first CST, adjoined by a full intron sequence to an exon encoding a transmembrane domain, and also encoding a reporter gene in the antisense orientation, wherein paired recombinase mediated cassette exchange (RMCE) elements are included in the TEDV and TERS such that execution of RMCE between the TEDV and TERS results in exchange of the reporter element for the GOI encoded by the TEDV, and exchange of the first CST exon for the second CST exon, such that the derivative engineered cell now expresses the first CST and GOI, in place of the second CST and the reporter gene.

2. The combined system according to claim 1 wherein the first component is a TEDV comprising

a. a first RMCE element
b. a 3' intron fragment
c. a CST exon
d. a first transcriptional terminator
e. a second transcriptional terminator
f. a GOI
g. a Kozak sequence
h. a second RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

3. The combined system according to claim 1 or 2, wherein the second component is a TERS comprising

a. a transcriptional promoter element
b. a Kozak sequence
c. a transmembrane domain exon
d. an intron

e. a first RMCE element encoded within the intron sequence d.
f. a CST exon
g. a first transcriptional terminator
h. a second transcriptional terminator
i. a reporter gene
j. a Kozak sequence
k. a second RMCE element
l. a second transcriptional promoter element

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.

4. The combined system according to any of claims 1 to 3, wherein the first RMCE element of the TEDV is paired with the first RMCE element of the TERS, and the second RMCE element of the TEDV is paired with the second RMCE element of the TERS.

5. The combined system according to any of claims 1 to 4, wherein each CST exon comprises a sequence encoding one or more molecular affinity tags and that the CST encoded by the TEDV and TERS are different.

6. The combined system according to any of claims 1 to 4, wherein the engineered cell contains a single TERS in its genome.

7. A method for generating derivative engineered cells expressing a TEDV-encoded GOI from the TERS locus, said method comprising,

a. generating a TEDV encoding a GOI
b. delivering said TEDV to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein
c. contacting cells with two or more affinity reagents specific for both the TEDV-encoded CST and TERS-encoded CST

d. selecting derivative engineered cells on the basis of diminished expression of the reporter gene and TERS-encoded CST, and increased expression of the TEDV-encoded CST, as a proxy for selection of cells with integrated GOI.

8. The method according to claim 7, wherein the affinity reagents used in step c. are fluorescently labelled as to detect diminished expression of TERS-encoded CST and increased expression of TEDV-encoded CST, as to enable cell partitioning and selection based on said expression by way of florescence activated cell sorting.

9. The method according to claim 7, wherein the affinity reagents used in step c. are immobilised to a substrate such that cells expressing TERS-encoded CST may be depleted or that cells expressing TEDV-encoded CST may be enriched, in the target cell population using substrate affinity methods, such as magnetic activated cell sorting.

10. A method for generating multiple derivative engineered cells expressing a range of TEDV-encoded GOI, from a pool of TEDV, said method comprising,

a. generating a library of two or more TEDV, each encoding a unique GOI sequence, and each with a unique TEDV-encoded CST

b. delivering said library of TEDV as a pool, to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein

c. contacting cells with three or more affinity reagents specific for both the multiple TEDV-encoded CST and TERS-encoded CST

d. selecting derivative engineered cells on the basis of diminished expression of the reporter gene and TERS-encoded CST, and increased expression of each of the unique TEDV-encoded CST

11. A method for cell lineage tracing of derivative engineered cells expressing a range of TEDV-encoded GOI, within a pool of cells as generated by the steps:

- generating a library of two or more TEDV, each encoding a unique GOI sequence, and each with a unique TEDV-encoded CST

- delivering said library of TEDV as a pool, to an engineered cell line containing a paired TERS, along with the recombinase enzyme matching the RMCE elements encoded therein,

said method comprising,

a. contacting cells with two or more affinity reagents specific for the multiple TEDV-encoded CSTs

b. analysing the content of derivative engineered cells on the basis of expression of each of the unique TEDV-encoded CSTs.

12. A tag-exchange donor vector (TEDV) encoding a cell surface tag (CST) exon flanked by a 3' intron fragment, and a gene of interest (GOI) in the antisense orientation.

13. The tag-exchange donor vector (TEDV) according to claim 12, comprising

a. a first RMCE element
b. a 3' intron fragment
c. a CST exon
d. a first transcriptional terminator
e. a second transcriptional terminator
f. a GOI
g. a Kozak sequence
h. a second RMCE element

wherein the CST exon and first transcriptional terminator are encoded in the antisense orientation from the GOI and associated transcriptional terminator and Kozak sequences.

14. An engineered cell containing within its genome a tag-exchange receiver site (TERS), encoding a cell surface tag (CST) exon adjoined by a full intron sequence to an exon encoding a transmembrane domain, and also encoding a reporter gene in the antisense orientation, wherein recombinase mediated cassette exchange (RMCE) elements are included in the TERS such that execution of RMCE between TERS and a tag-exchange donor vector (TEDV) results in exchange of the reporter element for a gene of interest (GOI) encoded by the TEDV.

15. The engineered cell according to claim 14, wherein said TERS comprises:

a. a transcriptional promoter element
b. a Kozak sequence
c. a transmembrane domain exon
d. an intron

e. a first RMCE element
f. a CST exon
g. a first transcriptional terminator
h. a second transcriptional terminator
i. a reporter gene
j. a Kozak sequence
k. a second RMCE element
l. a second transcriptional promoter element,

wherein the transmembrane domain exon and CST exon are encoded in the antisense orientation from the reporter gene, such that the first transcriptional promoter element drives transcription of the combined transmembrane domain and CST, and the second transcriptional promoter element drives transcription of the reporter gene.