A two-component vector library system for rapid assembly and diversification of full-length T-cell receptor open reading frames

Field of the invention

The heterodimeric T-cell receptor (TCR) is central to adaptive immunity. Unique TCRs are continually generated during T-cell genesis whereby an array of gene segments are recombined into a single contiguous TCR open reading frame (ORF). Due to the large degree of diversity generated in this process of recombination, it is challenging to capture sequence data from TCR pairs that are expressed in single cells. Moreover, this diversity also makes it a challenging to provide these TCR open reading frames (ORFs) within genetic constructs on a high-throughput basis for testing and manipulation of TCR function. The present invention in the first aspect provides a pre-assembled two-component vector library system consisting of *Variable-Constant entry* vectors (V-C entry) and *Joining donor* (J donor) vectors comprising portions of TCR gene segments. The two component system is designed in such a way that when a V-C entry vector selected from the V-C entry vector library is combined with a J donor vector selected from the J donor vector library, along with a synthetic DNA *oligonucleotide duplex encoding TCR complementarity determining region 3* (odeCDR3) in a restriction enzyme digestion / ligase cycle reaction, a single vector is created reconstituting the full-length TCR ORF. Such a vector library system enables PCR-independent methods for rapid and cost effective generation of TCR ORFs in selected vector contexts. In addition, this system permits novel workflows for generating synthetic TCR sequences for affinity and/or functional maturation workflows. This TCR ORF Reconstitution and Engineering System (TORES) is thus a strong tool for TCR functional analysis and engineering. In the second aspect, a TORES with modified V-C entry vectors is provided along with a third component bidirectional terminator donor vector (BiT donor) as to enable a second step to adjoin two TCR ORF chain pairs, in an antiparallel coding sense, into a single product vector. This bidirectional TORES^2 system enables alternative workflows to deliver paired TCR ORF constructs for TCR manipulation and characterisation.

Background of the invention

Immune surveillance by T lymphocytes (T-cells) is a central function in the adaptive immunity of all jawed vertebrates. Immune surveillance by T-cells is achieved through a rich functional diversity across T-cell subtypes, which serve to eliminate pathogen-infected and neoplastic cells and orchestrate adaptive immune responses to invading

pathogens, commensal microorganisms, commensal non-self factors such as molecular components of foodstuffs, and even maintain immune tolerance of self. In order to respond to various foreign and self factors, T-cells must be able to specifically detect molecular constituents of these foreign and self factors. Thus T-cells must be able to detect a large cross-section of the self and non-self molecules that an individual encounters, with sufficient specificity to mount efficient responses against pathogenic organisms and diseased self, while avoiding the mounting of such responses against health self. The highly complex nature of this task becomes clear when considering the practically unlimited diversity of both foreign and self molecules, and that pathogenic organisms are under evolutionary pressure to evade detection by T-cells.

The T-cell Receptor (TCR)

T-cells are primarily defined by the expression of a T-cell receptor (TCR). The TCR is the component of the T-cell that is responsible for interacting with and sensing the targets of T-cell adaptive immunity. In general terms, the TCR is comprised of a heterodimeric protein complex presented on the cell surface. Each of the two TCR chains are composed of two extracellular domains, being the variable (V)-region and the constant (C)-region, both of the immunoglobulin superfamily (IgSF) domain, forming antiparallel β-sheets. These are anchored in the cell membrane by a type-I transmembrane domain, which adjoins a short cytoplasmic tail. The quality of the T-cells to adapt and detect diverse molecular constituents arises from variation in the TCR chains that is generated during T-cell genesis. This variation is generated by somatic recombination in a similar manner to antibody genesis in B-cells.

TCR chain diversity

The T cell pool consists of several functionally and phenotypically heterogeneous subpopulations. However, T cells may be broadly classified as αβ or γδ according to the somatically rearranged TCR form they express at their surface. There exist two TCR chain pair forms; TCR alpha (TRA) and TCR beta (TRB) pairs; and TRC gamma (TRG) and TCR delta (TRD) pairs. T-cells expressing TRA:TRB pairs are referred to as αβ T-cells, while T-cells expressing TRG:TRD pairs are often referred to as γδ T-cells.

TCRs of both αβ and γδ forms are responsible for recognition of diverse ligands, or 'antigens', and each T-cell generates αβ or γδ receptor chains de novo during T-cell maturation. These de novo TCR chain pairs achieve diversity of recognition through generation of receptor sequence diversity in a process called somatic V(D)J recombination af-

ter which each T-cell expresses copies of a single distinctly rearranged TCR. At the TRA and TRG loci, a number of discrete variable (V) and functional (J) gene segments are available for recombination and juxtaposed to a constant (C) gene segments, thus referred to as VJ recombination. Recombination at the TRB and TRD loci additionally includes a diversity (D) gene segment, and is referred to as VDJ recombination.

Each recombined TCR possess potential for unique ligand specificity, determined by the structure of the ligand-binding site formed by the α and β chains in the case of αβ T-cells or γ and δ chains in the case of γδ T-cells. The structural diversity of TCRs is largely confined to three short hairpin loops on each chain, called complementarity-determining regions (CDR). Three CDRs are contributed from each chain of the receptor chain pair, and collectively these six CDR loops sit at the membrane-distal end of the TCR extracellular domain to form the antigen-binding site.

Sequence diversity in each TCR chain is achieved in two modes. First, the random selection of gene segments for recombination provides basal sequence diversity. For example, TRB recombination occurs between 47 unique V, 2 unique D and 13 unique J germline gene segments. In general, the V gene segment contributes both the CDR1 and CDR2 loops, and are thus germline encoded. The second mode to generate sequence diversity occurs within the hypervariable CDR3 loops, which are generated by random deletion of template nucleotides and addition of non-template nucleotides, at the junctions between recombining V, (D) and J gene segments.

TCR:CD3 Complex

Mature αβ and γδ TCR chain pairs are presented at the cell surface in a complex with a number of accessory CD3 subunits, denoted ε, γ, δ and ζ. These subunits associate with αβ or γδ TCRs as three dimers (εγ, εδ, ζζ). This TCR:CD3 complex forms the unit for initiation of cellular signalling responses upon engagement of a αβ or γδ TCR with cognate antigen. The CD3 accessories associated as a TCR:CD3 complex contribute signalling motifs called immunoreceptor tyrosine-based activation motifs (ITAMs). CD3ε, CD3γ and CD3δ each contribute a single ITAM while the CD3ζ homodimer contains 3 ITAMs. The three CD3 dimers (εγ, εδ, ζζ) that assemble with the TCR thus contribute 10 ITAMs. Upon TCR ligation with cognate antigen, phosphorylation of the tandem tyrosine residues creates paired docking sites for proteins that contain Src homology 2 (SH2) domains, such as the critical ζ-chain-associated protein of 70 kDa (ZAP-70). Recruitment of such proteins initiate the formation of TCR:CD3 signalling com-

plexes that are ultimately responsible for T-cell activation and differentiation.

αβ T-cells

αβ T-cells are generally more abundant in humans than their γδ T-cell counterparts. A majority of αβ T-cells interact with peptide antigens that are presented by HLA complexes on the cell surface. These peptide-HLA (pHLA)-recognising T-cells were the first to be described and are by far the best characterised. More rare forms of αβ T-cells have also been described. Mucosal-associated invariant T (MAIT) cells appear to have a relatively limited α and β chain diversity, and recognise bacterial metabolites rather than protein fragments. The invariant natural killer T-cells (iNK T-cells) and germline-encoded mycolyl-reactive T-cells (GEM T-cells) are restricted to recognition of glycolipids that are cross-presented by non-HLA molecules. iNK T-cells are largely considered to interact with CD1d-presented glycolipids, whereas GEM T-cells interact with CD1b-presented glycolipids. Further forms of T-cells are thought to interact with glycolipids in the context of CD1a and CD1c, however, such cells are yet to be characterised in significant detail.

Conventional αβ T-cells

The key feature of most αβ T-cells is the recognition of peptide antigens in the context of HLA molecules. These are often referred to as 'conventional' αβ T-cells. Within an individual, self-HLA molecules present peptides from self and foreign proteins to T-cells, providing the essential basis for adaptive immunity against malignancies and foreign pathogens, adaptive tolerance towards commensal organisms, foodstuffs and self. The HLA locus that encodes HLA proteins is the most gene-dense and polymorphic region of the human genome, and there are in excess of 12,000 alleles described in humans. The high degree of polymorphism in the HLA locus ensures a diversity of peptide antigen presentation between individuals, which is important for immunity at the population level.

HLA class I and II

There are two forms of classical HLA complexes: HLA class I (HLAI) and HLA class II (HLAII). There are three classical HLAI genes: *HLA-A, HLA-B, HLA-C.* These genes encode a membrane-spanning α-chain, which associates with an invariant β2-microglobulin (β2M) chain. The HLAI α-chain is composed of three domains with an immunoglobulin fold: α1, α2 and α3. The α3 domain is membrane-proximal and largely invariant, while the α1 and α2 domains together form the polymorphic membrane-distal

antigen-binding cleft. There are six classical HLAII genes: *HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA*, and *HLA-DRB1*. These genes encode paired DP, DQ and DR heterodimeric HLA complexes comprising a α-chain and a β-chain. Each chain has two major structural domains with an immunoglobulin fold, where the α2 and β2 domain comprise membrane-proximal and largely invariant modules similar to that of HLAI α3 domain. The HLAII α2 and β2 domains together form the membrane-distal antigen-binding cleft and are regions of high polymorphism.

The antigen-binding cleft of HLAI and HLAII comprises two anti-parallel α-helices on a platform of eight anti-parallel β-sheets. In this cleft the peptide antigen is bound and presented in an extended conformation. The peptide-contacting residues in HLAI and HLAII are the location of most of the sequence polymorphism, which constitutes the molecular basis of the diverse peptide repertoires presented by different HLA alleles. The peptide makes extensive contacts with the antigen-binding cleft and as a result each HLA allele imposes distinct sequence constraints and preferences on the presented peptides. A given peptide will thus only bind a limited number of HLAs, and reciprocally each allele only accommodates a particular fraction of the peptide collection from a given protein. The set of HLAI and HLAII alleles that is present in each individual is called the HLA haplotype. The polymorphism of HLAI and HLAII genes and the co-dominant expression of inherited alleles drives very large diversity of HLA haplotype across the human population, which when coupled to the enormous sequence diversity of αβ TCRs, presents high obstacles to standardisation of analysis of these HLA-antigen-TCR interactions.

αβ TCR engagement of HLAI and HLAII

αβ TCRs recognize peptides as part of a mixed pHLA binding interface formed by residues of both the HLA and the peptide antigen (altered self). HLAI complexes are presented on the surface of nearly all nucleated cells and are generally considered to present peptides derived from endogenous proteins. T-cells can thus interrogate the endogenous cellular proteome of an HLAI-presenting cell by sampling pHLAI complexes of an interacting cell. Engagement of HLAI requires the expression of the TCR co-receptor CD8 by the interacting T-cell, thus HLAI sampling is restricted to $CD8^+$ αβ T-cells. In contrast, the surface presentation of HLAII complexes is largely restricted to professional APC, and are generally considered to present peptides derived from proteins exogenous to the presenting cell. An interacting T-cell can therefore interrogate the proteome of the extracellular microenvironment in which the presenting cell resides.

The engagement of HLAII requires the expression of the TCR co-receptor CD4 by the interacting T-cell, thus HLAII sampling is restricted to $CD4^+$ αβ T-cells.

Thymic selection of αβ TCRs

The role of αβ TCRs as described above is the detection of pHLA complexes, such that the TCR-presenting T-cell can raise responses germane to the role of that T-cell in establishing immunity. It should be considered that the αβ TCR repertoire generated within an individual must account for the immense and unforeseen diversity of all foreign antigens likely to be encountered in the context of a specific haplotype and prior to their actual occurrence. This outcome is achieved on a background where extremely diverse and numerous αβ TCRs are generated in a quasi-randomised manner with the potential to recognise unspecified pHLA complexes while only being specifically instructed to avoid strong interactions with self pHLA. This is carefully orchestrated during T-cell maturation in a process call thymic selection.

During the first step of T-cell maturation in the thymus, T-cells bearing αβ TCRs that are incapable of interacting with self-pHLA complexes with sufficient affinity, are deprived of a survival signal and eliminated. This step called positive selection assures that the surviving T-cells carry a TCR repertoire that is at least potentially capable of recognizing foreign or altered peptides presented in the right HLA context. Subsequently, αβ TCR that strongly interact with self-pHLA and thus have the potential to drive autoimmunity are actively removed through a process of negative selection. This combination of positive and negative selection results in only T-cells bearing αβ TCRs with low affinity for self-pHLA populating the periphery. This establishes an αβ T-cell repertoire that is self-restricted but not self-reactive. This highly individualised nature of T-cell genesis against HLA haplotype underscores the challenges in standardised analysis αβ TCR-antigen-HLA interactions. Moreover, it forms the basis of both graft rejection and graft versus host disease and the general principle that αβ TCRs identified in one individual may have completely different effect in a second individual, which has clear implications for TCR-based and T-cell based therapeutic and diagnostic strategies emerging in clinical practice.

Unconventional αβ T-cells

The non-HLA-restricted, or 'unconventional', forms of αβ T-cells have very different molecular antigen targets. These unconventional αβ T-cells do not engage classical HLA complexes, but rather engage conserved HLA-like proteins such as the CD1 family or

MR1. The CD1 family comprises four forms involved in antigen cross-presentation (CD1a,b,c and d). These cell surface complexes have an α-chain resembling HLAI, which forms heterodimers with β2-M. A small hydrophobic pocket presented at the membrane distal surface of the α-chain forms a binding site for pathogen-derived lipid-based antigens. Innate like NK T-cells (iNK T-cells) form the best-understood example of lipid/CD1 family recognition with GEM T-cells representing another prominent example. 'Type I' iNK T-cells are known to interact strongly with the lipid α-GalCer in the context of CD1d. These iNK T-cells display very limited TCR diversity with a fixed TCR α-chain (Vα10/Jα18) and a limited number of β-chains (with restricted vβ usage) and they have been likened to innate pathogen-associated molecular patterns (PAMPS) recognition receptors such as Toll-like and Nod-like receptors. In contrast, 'type II' NK T-cells present a more diverse TCR repertoire, and appear to have a more diverse mode of CD1d-lipid complex engagement. GEM T-cells recognize mycobacteria-derived glycolipids presented by CD1b, however, the molecular details of antigen presentation by CD1a, b and c as well as their T-cell recognition are only beginning to be understood.

MAIT cells largely express an invariant TCR α-chain (TRAV1-2 ligated to TRAJ33, TRAJ20, or TRAJ12), which is capable of pairing with an array of TCR β-chains. Instead of peptides or lipids MAIT TCRs can bind pathogen-derived folate- and riboflavin-based metabolites presented by the HLAI-like molecule, MR1. The limited but significant diversity in the TCRs observed on MAIT TCRs appear to enable the recognition of diverse but related metabolites in the context of the conserved MR1.

It is not well-understood how non-classical HLA-restricted αβ T-cell TCRs are selected in the thymus during maturation. However, it appears likely that the fundamental process of negative and positive selection outlined above still applies and some evidence suggests that this occurs in specialized niches within the thymus.

γδ T-cells

In contrast to the detailed mechanistic understanding of αβ TCR genesis and pHLA engagement, relatively little is known about the antigen targets and context of their γδ T-cell counterparts. This is in part due to their relatively low abundance in the circulating T-cell compartment. However, it is broadly considered that γδ T-cells are not strictly HLA restricted and appear to recognize surface antigen more freely not unlike antibodies. Additionally, more recently it has become appreciated that γδ T-cells can dominate the resident T-cell compartment of epithelial tissues, the main interaction site of the

immune system with foreign antigen. In addition, various mechanisms for γδ T-cell tumour immunuosurveillance and surveillance of other forms of dysregulated-self are beginning to emerge in the literature. The specific antigen targets of both innate-like and adaptive γδ T-cells remain poorly defined but the tissue distribution and fast recognition of PAMPs suggests a fundamental role for γδ T-cells both early in responses to foreign antigens as well as early in life when the adaptive immune system is still maturing.

The diverse functions of γδ T-cells appear to be based on different Vγ Vδ gene segment usage and can be broadly understood in two main categories in which γδ T-cells with largely invariant TCRs mediate innate-like recognition of PAMPs very early during infection. Beyond PAMPs these type of γδ T-cells are furthermore believed to recognize self-molecules, including phosphoantigens that could provide very early signatures of cellular stress, infection and potentially neoplastic development. Recognition of PAMPs and such so-called danger associated molecular patterns (DAMPS) as well as the large numbers of tissue-restricted innate-like γδ T-cells strongly suggests that these cells are suited to respond rapidly to antigenic challenge without the need for prior activation, homing and clonal expansion.

A second form of γδ T-cells are considered to be more adaptive in nature, with a highly diverse γδ TCR repertoire and the ability to peripherally circulate and access lymphoid tissues directly. Such antigen-specific γδ T-cells to common human pathogens such as CMV have been described and they appear to form a memory response. However, it has also been observed that γδ T-cells show only relatively limited clonal proliferation after activation and little data is available on the extent of TCR diversity and specific responses of γδ T-cells in peripheral circulation, or in tissues. Furthermore, while it is generally considered that γδ TCRs do not interact with pHLA complexes, and thus do not engage with peptide antigens in this context only few antigen targets of γδ T-cells have been characterised and the underlying molecular framework is only poorly understood.

The low frequency of peripheral γδ T-cells and the difficulty to study tissue-resident T-cells in humans has limited our knowledge of how this important and diverse type of T-cells participate in adaptive immune responses. This emerging area of research would require more reliable technologies with which to capture and characterise rare γδ T-cells, isolate their TCR pairs, and to identify their cognate antigens.

Antigens and Antigen-presenting cells

In the context of T-cells and TCRs, antigens may be defined as any molecule that may be engaged by a TCR and resulting in a signal being transduced within the T-cell. The most well characterised T-cell antigens are peptides presented in an HLAI and HLAII complex, and which are engaged by conventional αβ T-cells. However, in recent years it has become apparent that non-conventional αβ T-cells and γδ T-cells are able to engage a wide range of biomolecules as antigens, including lipids, lipopeptides, glycopeptides, glycolipds and a range of metabolites and catabolites. In addition, it has emerged that γδ T-cells may be able to engage fully folded proteins directly in an antibody-like fashion. Therefore, the view of T-cell antigens being largely restricted to HLA-presented peptides has expanded over the past two decades to include almost any biomolecule. With this concept in mind, it is relevant to define what may be considered an antigen-presenting cell (APC).

As defined in the above sections, HLAI and HLAII have a disparate expression profiles across cell types. It is widely accepted that nearly all nucleated cells present HLAI complexes on the cell surface, and are thus competent to present peptide antigens for T-cell sampling. In contrast, HLAII has a restricted expression profile, and at least in steady state conditions is only expressed on the surface of cells that have a specialist role in antigen presentation, including dendritic cells (DC), macrophage and B-cells. These specialist cell types are often referred to as professional APC. For the purposes of this document, the term APC is used to describe any nucleated cell that is capable of presenting an antigen for sampling by αβ or γδ T-cells. Such antigens are not restricted to those presented as 'cargo' in specific antigen-presenting complexes such as HLA and HLA-like molecules, but may also include any cell-surface presented moiety that is able to engage a αβ or γδ TCR-bearing cell.

Therapeutic use of TCRs

Adoptive transfer of primary T-cells was first trialled in a clinical setting in the early 1990s, starting with ex vivo expanded T-cells polarised towards viral antigens to confer viral immunity in immunocompromised patients. Similar approaches using primary T-cells expanded ex vivo against specific cancer antigens were soon after trialled in treatment of malignancies. One limitation in these early approaches that continues to be a challenge today is a lack of understanding of the nature and diversity of T-cells clashing with the need to finely-optimize their composition in the therapeutic product. At present, the use of ex vivo expanded primary T-cells has largely been abandoned by

the pharmaceutical industry with the exception of a handful of initiatives using primary T-cells with specificity for viral antigens.

In recent years the ability to reliably introduce genetic material into primary human cells has seen a variety of experimental genetically modified T-cell therapeutics arise. Such therapeutic cell products aim to harness the power of T-cell responses and redirect T-cell specificity towards a disease-associated antigen target, for example, an antigen uniquely expressed by malignant cells. These have largely relied on the transfer of a chimeric antigen receptor (CAR) into recipient T-cells, rather than actual TCR chain pairs. A CAR represents a targeting moiety (most often a single-chain antibody element targeting a surface expressed protein of malignant cells) grafted to signal receptor elements such as the ζ-chain of the CD3 complex, to produce a synthetic chimeric receptor that mimics CD3-TCR function. These so-called CAR T-cell (CAR-T) products have met mixed success in clinical trials to date and despite their potential are not easy to translate beyond tumours with inherent unique molecular targets such as B-cell malignancies. Alternatively, the transfer of full-length TCR chain pair ORFs into T-cells is of emerging interest. Such TCR-engineered T-cell therapeutics are at present limited by challenging manufacturing processes, and like the CAR-T products, a dearth of validated antigen targets and targeting constructs. To date this has been focused on the use of αβ TCRs for recognition of peptide antigens presented by HLAI on malignant cells and a fundamental challenge of this approach is the need for antigens that are specific to malignant cells.

It has been considered that since the TCR-pHLA interaction is of relatively low-affinity, native TCRs are likely to be suboptimal for TCR-engineered T-cell therapies. Several approaches have been devised to affinity-mature TCRs in vitro, in much the same manner as single-chain antibody affinity maturation. These TCR affinity maturation approaches generally also utilise a single-chain formats, wherein the V-region of one chain is fused to V-region of another chain to make a single polypeptide construct. Such single polypeptides may then be used in phage- or yeast- display systems adapted from antibody engineering workflows, and passed through rounds of selection based on target binding. Two inherent limitations exist in such a single-chain TCR approach in terms of yielding functional TCR chain pairs. Firstly, the selection is based on binding affinity to the target. However, it has been well documented that TCR affinity does not always correlate to the strength or competency of TCR signalling output. Secondly, the selection of single-chain constructs based on affinity does not always trans-

late to equivalent affinities once they are reconstituted as full-length receptors.

In a therapeutic context, there exists an additional limitation in affinity-matured TCR pairs. That is, considering their sequences have been altered, the resulting constructs by definition have no longer been subject to thymic selection, wherein TCRs that react strongly to self-antigens are deleted from the repertoire. Therefore, these modified TCRs carry an inherent risk of being auto-reactive, which is very difficult to rule out in vitro using current methods. For the same reason, any selected or engineered TCR for therapeutic application needs to be individualised. If TCRs are artificially engineered or native TCRs applied across individuals, cross-reactivities have to be ruled out on the basis of the HLA haplotype and presented peptide repertoire of each specific individual in order to avoid potentially catastrophic autoimmunity. This is due to the fact that thymic selection is conducted on a background of all available HLA molecules specific only to that given individual. The likelihood of such cross-reactivity is unclear. However, the ability of our TCR repertoire to recognize pHLA complexes of other individuals of the same species as foreign is a fundamental property of adaptive immunity and underpins graft rejection and graft versus host disease. Recent clinical trials using a matured TCR chain pair against the cancer-specific melanoma associated antigen (MAGE) highlighted the potential problem of bypassing thymic selection. When autologous T-cells harbouring the matured TCRs were infused back to two cancer patients, these patients rapidly developed a fatal heart disease. Subsequent studies determined that the MAGE-specific matured TCRs were cross-reactive with an HLAI-presented peptide from the heart protein titin. This strongly suggests that cross-reactivity is a distinct possibility in therapeutic use of TCRs.

A distinct avenue of utilising TCRs for therapeutic purposes is in their use as affinity reagents in much the same manner as antibody therapeutic substances. Single-chain TCR molecules have been trialled for delivery of conjugated drug substances to specific HLA-antigen expressing cell populations. Such an approach is generally considered safer than CAR-T or TCR engineered T-cell therapeutics, as administration of the drug substance may simply be withdrawn. However, the potential for cross-reactivity and off target effects that are difficult to predict remains a potential limitation in this setting.

TCR repertoire detection in clinical diagnostics

In a related aspect, there is an emerging interest in using the detection of the abundance of specific TCR sequences for clinical diagnostic purposes. With the rise of

deep-sequencing methods in particular, it is possible to capture the full TCR diversity within an individual globally and for matched αβ pairs in specific contexts. This potentially represents a means to diagnose specific conditions and disease states simply by detecting the abundance of expanded T-cell clones, as proxy readout for established immune response against a disease-associated antigen in the patient. However, such global approaches are currently limited to very strong immune responses with established clinical time-points and suffer from the underlying difficulty in identifying the specific antigen target of any particular TCR identified via sequencing.

Therapeutic and diagnostic use of T-cell antigens

The fundamental strength of harnessing adaptive immune responses translates into a central technical challenge in that the exquisite specificity of the TCR-antigen interaction requires detailed knowledge of the antigens specifically associated with each pathogen, cancer cell or autoimmune disease. Furthermore, each antigen may be presented by a specific antigen presenting complex, or allele thereof, such that antigen discovery has be performed for each relevant HLA gene and allele. For several infectious diseases like HIV, influenza and CMV that are associated with strong adaptive immune responses and generally display conserved epitope response hierarchies, the most important epitopes have been mapped in context of some common HLA. Similarly, the fields of cancer, allergy and autoimmunity have seen increased and systematic efforts to map the associated T-cell antigens. However, these are challenging procedures and the efforts to systematically describe T-cell antigens associated with different clinical contexts are hindered by the absence of efficient, robust, fast and scalable protocols.

Specifically, cancer cells represent a challenging and important aspect as most of the peptides presented on the surface of malignant cells are self antigens or very similar to self antigens. Therefore, thymic selection will have deleted TCRs that could strongly recognize these peptides, while at the same time the tumour has evolved to evade immune recognition. This means that potent immune responses against established tumours are relatively rare and targets difficult to predict or discover. However, these responses do exist and, importantly, are generally associated with better outcome. The target of such responses, tumour-associated-antigens (TAA), will in most cases have distinguishing characteristics from self and be derived from proteins that are overexpressed during cancer development, otherwise absent from the cell type at this stage of development or specifically altered through genetic mutation or post-translational modifications such as phosphorylation.

When available, the knowledge of such epitopes makes it possible to interrogate the associated T-cell response for fundamental discovery, diagnostic purposes and for example as a test of vaccine efficacy. Importantly, they also provide highly specific targets for T-cell tolerization in allergy and autoimmunity and, crucially, point towards valuable targets for specific immunotherapy and against malignant cells. Malignancies represent a particularly valuable target as the promise of cellular immunotherapies and the progress in the T-cell manipulations are slowed by a lack of validated target TAAs that go beyond the few cases where specific markers for the type of cancer happen to be available.

In the light of the potential of cellular therapy and lack of validated targets the identification of promising TCR antigens remains one of the most pressing bottlenecks of TCR-based immunotherapy, particularly in the effort to treat cancer.

Obtaining full-length TCR ORFs

Primarily due to the diversity of usage in V gene segments in naturally occurring TCRs, it is challenging to capture sequence data of paired TCR chains from pools of T-cells. In general, in order to reliably capture sequence with high reliability it is necessary to amplify TCR using reverse transcription and/or PCR methods, using a pool of primers that cover the full V region diversity of the target pool. This process generates ORF fragments from which sequence data may be obtained, however, make unsuitable starting material for readily cloning full-length TCR ORFs. Therefore, high-efficiency strategies for mining paired sequence data from pools of T-cells is generally incompatible with cloning of full-length TCR ORFs for downstream applications. Time consuming and expensive synthesis of TCR ORFs is thus required for any functional testing or application of the sequenced TCR pairs.

Diversification of full-length TCR ORFs

The emerging therapeutic use of TCRs has seen a rise in interest to diversify TCR sequences within affinity and/or functional maturation workflows to derive synthetic TCR pairs of defined specificity and function. Generally, this has been achieved through linking TCR chain pairs into a single-chain construct for phage-display methodologies, and introduction of sequence diversity into these single chain constructs. There is currently a lack of technologies that can rapidly diversify single TCR ORFs in a systematic manner, and where these full-length TCR ORFs are applicable to immediate testing and/or

selection in a context of surface expression on viable mammalian cells. Such a technology would be of great value in the maturation of TCR pairs, with highly defined specificity and signalling capacity, for therapeutic use.

Technological aspects of TCR and T-cell antigen analyses

Overall, the development of TCR-based therapies is still in its early stages, and success has been limited. Diagnostic approaches, while of immense potential, have seldom been deployed into controlled clinical studies that aim to assess patient disease states or response to therapy. Underdeveloped techniques for the reliable capture of native TCR chain pairs, and the systematic analysis of TCR-antigen interactions at high-throughput and in a functional context of cell-cell communication, have been the main hurdles to the development of TCR-based therapies and diagnostics.

Deep sequencing approaches have led to an improved understanding of T-cell receptor diversity in heath and disease. However, these approaches have generally focused on short stretches spanning the CDR3 regions, mainly of the TCR β-chain. Most studies have ignored the contribution of the TCR α-chain, and few have sought to analyse paired αβ chains as well as the antigen specificity of TCRs determined to be of interest. Recent workflows using single cell encapsulation and genetic barcoding has enabled the pairing of native TCR αβ or γδ chain pairs and analysis of full-length sequences, however, such workflows remain experimental.

Isolated TCR chain pairs may be analysed in terms of antigen specificity in either biophysical or functional modes. Biophysical analysis requires the recombinant production of both the TCR as well as the analyte antigen in soluble form. In the case of HLA-restricted TCRs this would thus require the generation of all individual TCRs as well as the cognate pHLA complexes. This is technically highly challenging, slow and very low-throughput. Furthermore, such analysis would only provide interaction affinities, which are not well-correlated with functional characteristics in predictable ways.

Until recently, the detailed functional analysis of isolated TCR sequences in a cellular context has been limited to laborious protocols of transfection of analyte TCR chain pairs into primary T-cells or immortal T-cell lines, and detection of cellular responses by traditional flow cytometric analysis of cell activation, or detection of secreted factors from the transfected cells upon antigen challenge. In a recent publication by Guo et al, rapid cloning, expression, and functional characterization of paired TCR chains from

single-cells was reported (Molecular Therapy – Methods and clinical development (2016) 3:15054). In this study, analyte human αβ TCR pairs were expressed in a reporter cell line that lacked αβ TCR expression, and which contained a green fluorescent protein (GFP) reporter system linked to the Nur77 promoter that is activated upon TCR stimulation. This system remains inefficient due to the lack of standardised TCR integration into the reporter cell line genome, and does not provide a systematic manner for cell-bound antigen challenge by an APC element.

Similar to workflows for identification of TCRs against known T-cell antigens, the de novo discovery of novel T-cell antigens in health and disease remains highly challenging. Most approaches remain biophysical in nature, and aim to produce candidate antigens that may be tested in immunisation protocols, or through identifying cognate TCRs as addressed above. Little or no standardisation exists in the field of T-cell antigen discovery, and the field is largely restricted to academic study.

With the accumulating interest in TCRs and their cognate antigens in both therapeutic and diagnostic use, and the emergence of means to capture significant numbers of native TCR αβ and γδ chain pairs, there remains a lack of reliable high-throughput and standardised technologies for the systematic analysis of TCR-antigen interactions. Importantly, there is a lack of standardised systems for rapid reconstitution and/or systematic diversification of full-length TCR ORFs, such that these ORFs may be directly applied to functional analysis of TCR chain pairs in the native context of an analyte TCR being presented on the surface of a viable cell in a native context. Such capability is important for achieving high-throughput analyses of native TCR chain pairs, but also affinity and/or functional maturation of TCR chain pairs, for therapeutic and diagnostic uses.

There is a clear need for rapid and systematic methods for TCR chain reconstitution, and their systematic diversification, in high-throughput methods that will enable the use of TCR diagnostics on an informatics and reagent basis, and also personalised TCR-based immunotherapies.

Detailed description of the invention

The present invention addresses the above-mentioned needs. The present invention provides in a first aspect, a two-component vector system comprising pre-assembled

libraries consisting of vectors harbouring variable (V), joining (J) and constant (C) sequences for TCR chains. The first component of such a system comprises a V-C entry vector containing V and C sequences. The second component of the system comprises a J donor vector containing J sequence. The two-component vector system is pre-assembled into libraries of V-C entry vectors and J donor vectors with all desirable V-C sequence combinations and J sequences, respectively. The two-component vector system is designed in such a manner that a single V-C entry vector and a single J donor vector with desired sequences can be combined with a short DNA oligonucleotide duplex encoding CDR3 (odeCDR3) sequence to reconstitute a full-length TCR ORF in vitro, in a single-tube reaction in a restriction enzyme and ligase dependent and PCR independent manner. In addition, the modular two-component system is ideally suited to rapidly generate large libraries of synthetic or mutant full-length TCR ORFs for affinity or functional maturation workflows. In a second aspect, modified V-C entry vectors for a reciprocal pair of TCR chains (i.e. TRA/TRB or TRD/TRG) are compiled into a 5-component vector system to provide a system in which reconstituted TCR chain pairs may be adjoined within a single vector. This second system utilises both modified V-C entry vectors, the same J donor vectors as the first system, and also a Bidirectional Terminator Donor Vector (BiT Donor) as to achieve adjoined TCR chain pairs encoded in antiparallel sense orientation in a final construct, with each TCR chain interposing a 'bidirectional terminator' element.

TCR ORF Reconstitution and engineering System (TORES)

The present invention first provides a two-component vector system with unique characteristics suitable for the above-mentioned uses. This TCR ORF reconstitution and engineering system (TORES) is used in conjunction with a third component, an oligonucleotide duplex encoding CDR3 (odeCDR3), to de novo assemble full-length TCR ORFs within a defined vector context and/or generate formulaic sequence diversity within a given TCR ORF.

The present invention is summarised in Figure 1A. A selected V-C entry vector containing V and C TCR gene segments required for a target full-length TCR ORF is combined with a J donor vector that contains the required J TCR gene segment. The full-length TCR ORF is completed by the addition of an oligonucleotide duplex encoding CDR3 (odeCDR3), which accounts for unfixed non-germline sequence generated during V(D)J recombination and interposed by fixed germline encoded V and J sequence encoded by the V-C entry vector and J donor vector, respectively. The two-component

vector system, and the third odeCDR3 component, is designed such that when combined into a restriction enzyme and ligase cycle reaction, the full V-CDR3-J-C TCR ORF is reconstituted. This is achieved via a Type IIS restriction enzyme(s) that are used to perform 'scarless' assembly of the genetic elements in a standardised manner. The two-component vector system is assembled into a library containing all required V, C and J gene segments for reconstitution of target full-length TCR ORFs (Figure 1B). For instance, a library can be constructed to contain all gene segments encoding native protein sequences of the human TRA repertoire as described in Examples 1 and 2, and the human TRB repertoire as described in Example 3.

To reconstitute a full-length TCR ORF, from sequence information that is sufficient to define V, J and C gene segment usage, along with unfixed CDR3 sequence interposed by fixed V and J segments, the V-C entry vector and J donor vector that correspond to the V/C and J usage of the target TCR ORF are first selected. An odeCDR3 corresponding the unfixed CDR3 sequence that is needed to complete the full-length TCR ORF is also generated. These three components are combined with a Type IIS restriction enzyme and DNA ligase enzyme in a cycle reaction to generate the target full-length TCR ORF as described in Figure 4 and Example 7. The resulting reconstituted full-length TCR is contained within the V-C entry vector backbone, thus contains all vector features contained within this parent construct.

The action of the Type IIS restriction enzyme of the three combined components (Figure 4 a, b, c) within a restriction enzyme / ligase cycle reaction, results in two reaction by-products and two reaction intermediates. The V-C entry vector derived reaction by-product is the excised native selection marker and Type IIS binding sites (Figure 4d). The J donor vector backbone from which the J segment part has been excised represents a second reaction by-product (Figure 4e). The excised J segment part from the J donor vector represents a reaction intermediate, and contains both a J segment part, a small C part from the C segment, and single stranded overhangs required for ligation (Figure 4f). The second reaction intermediate is the parental V-C entry backbone containing the V and C segments, and single stranded overhangs required for ligation (Figure 4g). The final product of reaction represents a full-length TCR ORF reconstituted within the parental V-C entry vector backbone, comprised of ligation of the odeCDR3 (Figure 4c), the excised J segment part (Figure 4f) and the V-C entry backbone carrying the V and C gene segments (Figure 4g).

The V-C entry vector and J donor vector components

In the present context, a combined two-component system includes one or more V-C entry vector/s containing

a. origin of replication,
b. a first positive selection marker,
c. 5' genetic element, or elements,
d. Kozak Sequence,
e. TCR variable gene segment,
f. a first Type IIS sequence, for site specific recognition and cleavage by a Type IIS restriction enzyme,
g. a negative selection marker,
h. a second Type IIS sequence,
i. TCR constant gene segment, and
j. 3' genetic element, or elements

wherein, a) and b) are used for propagation and selection of both parental V-C entry vector and the reconstituted TCR-containing vector in a bacterial host; c) and j) are used to define the downstream application of the reconstituted full-length TCR ORF; d) ensures efficient initiation of translation in eukaryotic cells, which could alternatively represent a Shine-Dalgarno sequence for transitional regulation in prokaryotes and archaea; e) represents the variable (V) gene segment from the start codon to a motif at the 5' edge of the CDR3 region conserved across all V segments in a given two-component vector system; f) represents a Type IIS recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang at the 3' end of the V gene segment; g) represents a negative selection marker to eliminate parental V-C entry vector during operation of the system to reconstitute a full-length TCR ORF; h) represents a Type IIS recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang at the 5' end of the C gene segment; i) represents the constant (C) gene segment from a motif at the 5' end of the C gene segment conserved across all C segments in a given two-component vector system, and which defines the boundary with the J segment (see Figures 2 and 4).

A V-C entry vector that is used for genetic reconstitution of the full-length TCR ORF without the need for downstream biological application, for example, to be used as a template for molecular biology workflows, the minimal V-C entry vector would comprise elements a), b), e), f), h) and i), lacking regulatory elements.

A V-C entry vector can also contain one or more transcriptional units for the expression of additional ORFs suitable for downstream applications, for example, a mammalian antibiotic resistance gene or reporter construct.

The combined two-component system includes one or more J-donor vector containing

a. origin of replication,
b. a second positive selection marker,
c. a third Type IIS sequence,
d. TCR Joining gene segment,
e. a C part, corresponding to a small 5' portion of a constant gene segment, and
f. a fourth Type IIS sequence.

wherein, a) and b) are used for propagation and selection of the J donor vector; c) represents a Type IIS recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang at the 5' end of the J gene segment; d) represents the Joining (J) gene segment starting from a 5' from the motif defining the 3' edge of the CDR3 region conserved across all J segments in a given two-component vector system, to a 3' sequence that incorporates C part, representing a 5' portion of the C segment encoded by V-C entry vector(s) contained within the two-component system; represents a Type IIS recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang at the 3' end of the J gene segment, and contained within the C part portion of the sequence (see figures 3 and 4).

A J-donor vector does not strictly need to carry a C part sequence, encoding a small 5' portion of the C gene segment. This C part is used to optimise and standardise overhangs for the reconstitution reaction during operation of a TORES. This is because of the higher sequence variation found at the 3' end of J gene segments, such that inclusion of a C part allows standardisation by generation of overhangs within the less diverse C gene segment. In the instance of constructing a TORES for a TCR loci from other organism that does not have 3' J segment diversity, or using synthetic J gene segments, this C-part may be omitted in favour of standardisation of overhangs within said J segments. This would reduce the complexity of the J donor library construction.

Each of first, second, third and fourth Type IIS sequences may be the same or differ-

ent. Preferably, they are the same. This ensures that each of the restriction sites within the two-component vector system is compatible with the same Type IIS enzyme, and only a single enzyme is needed for the restriction enzyme / ligase cycle reaction during reconstitution of full-length TCR ORF using the system. Type IIS enzymes do not cut within their recognition sequence, and thus the single-stranded overhangs are generated extrinsic to the recognition sequence. Therefore, the nature of the overhang generated upon Type IIS restriction enzyme action is dependent on both the orientation of the recognition sequence, and indeed the adjacent sequence (see examples 1 to 4).

Alternatively, each of the Type IIS restriction sequences may be different from one another. However, with the addition of each unique recognition sequence, an additional Type IIS enzyme must be incorporated into the restriction enzyme / ligase cycle reaction. This would increase the complexity and cost of a reconstitution reaction for assembling a full-length TCR ORF.

The first and second positive selection markers within the V-C entry vector and J donor vector, respectively, are normally different. This is to ensure that the V-C entry vector, which provides the backbone of the final full-length TCR ORF product, can be selected for independently of the J donor vector, and thus eliminate transformants that carry undigested or re-circularised J donor vectors that would otherwise contribute background to the reconstitution reaction (see Figures 2 and 3 and Example 7).

The positive selection markers can be selected from

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

wherein the choice, formatting and application of such positive selection markers are well known to those skilled in the art.

The 5' genetic element incorporated into a V-C entry vector comprises one or more elements selected from

a. gene cis/acting element,
b. heterospecific recognition site for recombinase enzymes,
c. a 5' homologous recombination arm for a genomic site of interest'
d. a mRNA splice acceptor site,

e. an internal ribosomal entry site, and
f. epigenetic insulator sequence

wherein, a) drives expression of the transcript encoded by the full-length TCR ORF product reconstituted within the V-C entry vector backbone; b) represents a sequence that directs site-directed recombination in the presence of recombinase enzymes to insert the full-length TCR ORF product reconstitute within the V-C entry vector backbone into a specific genetic context; c) represents a sequence that directs site-directed homologous recombination to insert the full-length TCR ORF product reconstituted within the V-C entry vector backbone into a specific genetic context; d) permits engineered domain-fusion approaches to manipulate the form of the protein expressed from the full-length TCR ORF reconstituted in the V-C entry vector backbone e) permits cap-independent initiation of translation of the mRNA expressed from the full-length TCR ORF reconstituted in the V-C entry vector backbone f) permits insulation of transcriptional activity otherwise affected by enhancer elements in a genomic context of where the full-length TCR ORF reconstituted in the V-C entry vector backbone may be inserted.

A cis/acting element may be used to drive transient expression of TCRs reconstituted into a V-C entry vector backbone provided in Examples 1 and 3, when said vector containing a reconstituted TCR ORF is transfected into mammalian cells.

A heterospecific recognition site for recombinase enzymes may be used to permit recombinase mediated cassette exchange of TCRs reconstituted into a V-C entry vector backbone provided in Example 4, when said vector containing a reconstituted TCR ORF is transfected into mammalian cells in the presence of appropriate recombinase enzyme.

A first Type IIS recognition sequence that is included in the V-C entry vector is oriented to cleave 5' of said recognition sequence and within the TCR variable gene segment (Figure 4a) to produce a single-stranded DNA overhang at the 3' end of the variable gene segment (Figure 4g) that is complementary to that at the 5' end of the synthesised odeCDR3 (Figure 4c) For details on how this first Type IIS recognition sequence is designed, see Examples 1, 3, 5 and 6.

A V-C entry vector contains a negative selection marker between the first Type IIS recognition sequence, and the second Type IIS recognition sequence (infra vide, Fig-

ure 2). This negative selection marker is selected from

a. a restriction enzyme recognition site not contained elsewhere in the first component or within the TCR joining gene segment,
b. a bacterial suicide gene, and
c. a reporter element.

wherein, the negative selection marker is used to eliminate host cells transformed with parental V-C entry vector, and thus reduce the background of a reconstitution reaction when using the first positive selection marker to select for transformants containing the target TCR ORF within the V-C entry vector backbone (see Example 7).

With the exception of the negative selection marker itself, all other sequences in the two-part system must be devoid of said sequence as to not confer undue negative selection on the basis of the inclusion of this sequence elsewhere in the system.

In the present context, a second Type IIS recognition sequence that is included in the V-C entry vector is orientated to cleave 3' of said recognition sequence and within the TCR constant gene segment (Figure 4a) to produce a single-stranded DNA overhang at the 5' end of the constant gene segment (Figure 4g) that is complementary to that at the 3' end of the J donor fragment reaction intermediate (Figure 4f). For details on how this second Type IIS recognition sequence is designed, see Examples 1, 2, 3 and 5.

The 3' genetic element incorporated into a V-C entry vector comprises one or more elements selected from

a. a terminator element,
b. heterospecific recognition site for recombinase enzymes,
c. a 3' homologous recombination arm for a genomic site of interest,
d. a mRNA splice donor site,
e. an internal ribosomal entry site, and
f. epigenetic insulator sequence.

wherein a) represents a sequence that directs transcriptional termination for effective mRNA production of the TCR ORF in situ and may encode a poly-A signal; b) represents a sequence that directs site-directed homologous recombination to insert the full-length TCR ORF product reconstituted within the V-C entry vector backbone into a specific genetic context; c) permits the fusion of a TCR ORF to a transcriptional unit after integration into a genomic locus encoding an downstream mRNA splice acceptor

site to manipulate the strength of TCR expression levels or form of the protein expressed from the full-length TCR ORF reconstituted in the V-C entry vector backbone e) permits cap-independent initiation of translation of the mRNA expressed from the full-length TCR ORF reconstituted in the V-C entry vector backbone f) prevent inappropriate interaction between adjacent chromatin domains, thus insulating the full-length TCR ORF from adjacent transcriptional regulation or spread of heterochromatin in a genomic context of where the reconstituted TCR ORF in the V-C entry vector backbone may be inserted

A terminator element is used to ensure transcriptional termination during expression of TCRs reconstituted into a V-C entry vector backbone provided in Examples 1 and 3, when said vector containing a reconstituted TCR ORF is a transfected into mammalian cells.

A heterospecific recognition site for recombinase enzymes is used to permit recombinase mediated cassette exchange of TCRs reconstituted into V-C entry vector backbones provided in Examples 4, when said vector containing a reconstituted TCR ORF is transfected into mammalian cells in the presence of appropriate recombinase enzyme.

A J donor vector contains a J gene segment with a C-part sequence, representing a 5' fragment of the C gene segment, to the 3' of the J gene segment (Figure 3).

The C-part sequence is designed to standardise the single stranded overhangs generated by Type IIS enzyme action within the at the 3' end of the J donor vector-derived J fragment reaction intermediate (Figure 4f), and that at the 5' end of the C gene segment within the Type IIS digested open V-C entry vector reaction intermediate (Figure 4g).

A third Type IIS recognition sequence that is included in the J donor vector is oriented to cleave 3' of said recognition sequence and within the TCR joining gene segment (Figure 4b) to produce a single-stranded DNA overhang at the 5' end of the joining gene segment (Figure 4f) that is complementary to that at the 5' end of the synthesised odeCDR3 (Figure 4c) For details on how this third Type IIS recognition sequence is designed, see Examples 2, 3, 5 and 6.

A fourth Type IIS recognition sequence that is included in the J donor vector is orientated to cleave 5' of said recognition sequence and within the TCR C-part (Figure 4b) to produce a single-stranded DNA overhang at the 3' end of the C-part (Figure 4f) that is complementary to that at the 5' Type IIS digested open V-C entry vector reaction intermediate (Figure 4g). For details on how this third Type IIS recognition sequence is designed, see Examples 1, 2, 3, 5 and 6.

The two-part vector system, all encoded TCR gene segments and parts should not contain Type IIS recognition sequences that are used for operation, or assembly, of the V-C entry vector or J donor vector. Inclusion of such sequences would result in Type IIS restriction enzyme action within the encoded gene segments or parts, and result in disruption of the TCR reconstitution process. Similarly, the Type IIS recognition sequences should not be included in the vector backbones, or in any 5' and 3' genetic elements within these vectors, nor the cloning fragments used to assemble the two-part vector system, nor the odeCDR3 representing a third system component (*infra vide*).

A two-component vector system of the TORES may be constructed for any collection of TCR chains. In examples 1 to 4 below, two-component vector systems are constructed for the human TRA and TRB loci, encoding the human TCR alpha and beta chains, respectively. The construction of such a TORES is equally applicable in the context of the TRD and TRG loci, encoding the TCR delta and gamma chain pair, respectively, or indeed for any TRA/TRB, TRD/TRG or variant TCR chain pair system found in jawed vertebrates.

The third odeCDR3 component

To reconstitute a full-length TCR ORF using any given TORES, a small ORF fragment not encoded by the two-component V-C entry vector and J donor vector system is required as a third component. This third component takes the form of an oligonucleotide duplex encoding CDR3 (odeCDR3).

Such a third component, odeCDR3, comprises

a. a first single strand overhang sequence complimentary to first Type IIS restriction enzyme recognition and cleavage site orientated to cleave 5' of the recognition sequence and within the TCR variable gene segment of the V-C entry vector,
b. a double strand segment encoding a TCR CDR3 region and de-

void of negative selection element, which negative selection element is as defined in item 10, and also devoid of any Type IIS restriction sequences of the first or second part, and

c. a second single strand overhang sequence complimentary to the third Type IIS restriction enzyme recognition and cleavage site orientated to cleave 3' of the recognition sequence and within the TCR joining gene segment of the J donor vector.

,

Alternatively, the odeCDR3 can be comprised of a dsDNA molecule and/or plasmid DNA encoding the CDR3 flanked by two Type IIS enzymes consistent with the first (V-C entry vector) or second (J donor vector) component, oriented such that when digested a product comprising of a, b and c described previously is generated, and two by-products encoding short dsDNA fragments flanked by the Type IIS sites. This alternative dsDNA odeCDR3 is compatible the restriction enzyme / ligase reaction, not necessarily requiring prior digestion or processing.

As an alternative to the use of V-C entry vector and J donor vector configuration in a TORES, J-C entry vector and V donor vector configuration may also be used by applying the same conceptual framework.

Methods to use a TORES to reconstitute full-length TCR ORFs

A TORES can be used to reconstitute a full-length TCR ORF in a genetic vector context, from sequence information, as is presented for a human TRA/TRB chain pair in Example 7.

To operate a TORES to reconstitute a full-length TCR ORF from sequence information, given the resource of a two-component vector system for a given TCR chain, the method comprises

a. selecting a V-C entry vector,
b. selecting a J donor vector,
c. selecting an odeCDR3,
d. combining a, b and c to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme and iii) subjecting the combined mix to a thermocycling reaction,

e. transforming the reaction product obtained from step d to a selectable host organism competent for DNA vector propagation, and

f. performing a selection of host organism to obtain full length reconstituted TCR open reading frame in the V-C entry vector backbone.

wherein, a) and b) are selected on the basis of the selected vector encoding the V,J and C gene segments in the target full-length TCR ORF; c) is selected on the basis of completing the full-length TCR ORF sequence not encoded by the V-C entry or J donor vectors selected in a) and b), and bounded by the Variable and Joining segments encoded therein; d) combining the three selected components into a reaction mixture along with a restriction enzyme that will cut the first, second, third and fourth Type IIS restriction enzyme recognition sequences within the V-C entry and J donor vectors; e) generally represents transformation-competent bacteria; f) selection of host is on the basis of the first positive selection marker provided by the V-C entry vector backbone.

Generally, a workflow to select and define the genetic elements of a full-length TCR ORF for reconstitution entails de novo sequencing of TCR chains from target organism tissues. Example 8 below presents the *de novo* identification of a set of TRA/TRB chain pairs specific for a HCMV antigen in a HLA-B*07:02 restricted context. The workflow described in Figure 13, incorporates reverse transcription and PCR based amplification of TCR chain pairs from sorted single cells with subsequent Sanger sequencing. There exists a requirement for high-quality sequence information spanning V, CDR3, J and C segments of the TCR ORF, which dictates the specific sequencing approach(es) taken.

A method for selecting and reconstituting a TCR open reading frame thus comprises

a. Obtaining a TCR open reading frame sequence wherein said sequence information is sufficient to identify i) variable gene segment usage ii) constant gene segment usage iii) joining gene segment usage iv) a full CDR3 sequence spanning the variable gene segment border to the joining gene segment border, and

b. selecting a V-C entry vector corresponding to the variable and constant gene segments identified in step a. i) and a. ii), respectively, and

c. selecting a J donor vector corresponding to the joining gene segment identified in step a, iii), and

d. generating an odeCDR3 corresponding to CDR3 sequence identified in step a. iv), and
e. combining b, c and d to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme, iii) subjecting the combined mix to a thermocycling reaction, and
f. transforming the reaction product obtained from step e. to a selectable host organism competent for plasmid replication, and
g. performing a selection of host organism to obtain full length reconstituted TCR open reading frame in the V-C entry vector backbone.

wherein, a) is conducted by sequencing methods well known to one skilled in the art, able to obtain sufficient sequence length and quality to identify all four required genetic elements; b) and c) are selected from a TORES library containing required vectors; d) is synthesised de novo or selected from an odeCDR3 library; e) is conducted in a single reaction vessel.

In order to select the appropriate V-C entry vector, J donor vector and odeCDR3, target TCR sequences were aligned against a library of V, C and J gene segments for their corresponding TCR chains to determine the V, C and J segment usage of the target chain. This sequence alignment and analysis step must also permit the definition of the CDR3 coding sequence, and thus the definition of odeCDR3 sequence. Thus, overall such sequence analysis permits the selection of V-C entry vectors and J donor vectors for TCR chain reconstitution. The analysis also permits the synthesis of odeCDR3 for each chain reconstitution reaction. This process is well described in Example 8 and summarised as part of Figure 13.

It is desirable to conduct the Type IIS digestion and DNA ligase-dependent ligation (step e) in a single cycle reaction. This minimises processing steps and is made possible by the design of the system, with Type IIS restriction enzymes cutting outside their recognitions sequences, such that a number of unique overhangs may be generated with a single enzyme, thus maintaining efficient directional cloning of the J donor vector reaction intermediate and odeCDR3 into the V-C entry vector backbone.

Alternatively, the Type IIS restriction digest and DNA ligation may be performed in se-

quential procedures.

In Example 8, the application of the TORES is exemplified in the context of single-cell fluorescence-activated cell sorting (FACS) of antigen-specific CD8 T-cells from human peripheral blood for reverse transcription and PCR based amplification of TRA/TRB TCR chain pairs, followed by Sanger sequencing. This is a generally applicable workflow, wherein any tissue may be the source of T-cells from any jawed vertebrate, and cells may be sorted based on any phenotypic characteristic. Importantly, the single-sorted cells need not be stained for antigen specificity using HLA-multimer reagents.

The TCR sequencing approach used is not restricted to any particular method or technology, provided sufficient high-quality sequence information is obtained such that the above-defined genetic characteristics of the TCR ORF(s) can be defined based on said sequence information.

The use of FACS for partitioning single cells such that native TCR chain pairs may be sequenced and identified is a powerful method due to the accurate and rich phenotypic information that may be collected with multi-specificity antibody panels. However, other methods exist to partition cells, including; emulsion PCR; digital PCR approaches using microfluidic cell encapsulation, digital PCR using physical partitioning substrates.

It is generally desirable to obtain native TCR pairs from a source material, as both chains of a TCR pair contribute to HLA-antigen engagement and recognition. However, there are instances where recovery of just a single chain may be desirable, such as high-throughput screening of a single chain against a set specificity. In such a case, TCRs may be amplified and/or sequenced from non-partitioned cells.

Methods to use a TORES to generate full-length TCR ORFs with diversified sequence

A TORES system is ideally suited to generate diversified full-length TCR ORFs in several systematic modes. Such systematic diversification may be applied to affinity and/or functional maturation workflows for TCR chains. Such diversification of target TCR chain sequences is well described in Examples 9 and 10.

Such TCR ORF sequence diversification methods follow the same general scheme as for a reconstitution reaction. Diversification can be conducted in multiple parallel reconstitution reactions, whereby a single variant TCR ORF is generated per reaction. How-

ever, in most scenarios it is desirable to generate a pool of variant TCR ORFs in a single reaction. Each of these approaches is achieved by providing multiple variants of one or more of each genetic component(s) (V-C entry vector, J donor vector, odeCDR3) to a reconstitution reaction.

As described in Example 9, a TCR ORF can be systematically diversified at the CDR3 region by adding a pool of odeCDR3 with defined positional sequence diversity.

A method for selecting and reconstituting a TCR open reading frame to achieve TCR ORF diversity in the CDR3 region, thus comprises

a. Obtaining a TCR open reading frame sequence wherein said sequence information is sufficient to identify i) variable gene segment usage ii) constant gene segment usage iii) joining gene segment usage iv) a full CDR3 sequence spanning the variable gene segment border to the joining gene segment border, and
b. selecting a V-C entry vector corresponding to the variable and constant gene segments identified in step a. i) and a. ii), respectively, and
c. selecting a J donor vector corresponding to the joining gene segment identified in step a, iii), and
d. generating two or more odeCDR3 corresponding to CDR3 sequence identified in step a. iv), with variant sequence composition, and
e. combining b, c and d to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme, iii) subjecting the combined mix to a thermocycling reaction, and
f. transforming the reaction product obtained from step e. to a selectable host organism competent for plasmid replication, and
g. performing a selection of host organism to obtain full length reconstituted TCR open reading frame in the V-C entry vector backbone,

wherein, a) is conducted by sequencing methods well known to one skilled in the art, able to obtain sufficient sequence length and quality to identify all four required genetic elements; b) and c) are selected from a TORES library containing required vectors; d)

is synthesised de novo, or selected from an odeCDR3 library; e) is conducted in a single reaction vessel.

Such a method can be achieved by pooling all odeCDR3 variants to a single reaction to generate a pool of sequence-diversified but may be equally achieved by proving each odeCDR3 variant to a parallel reaction.

Variant odeCDR3 can be generated via a variety of methods well known to those skilled in the art. The selection of position and extent of odeCDR3 degeneracy/diversity can range from a single residue change at a single position, to completely degenerate sequence to the length of the odeCDR3.

As described in Example 10, a TCR ORF can be systematically diversified by maintaining the CDR3 region via provision of odeCDR3, but diversifying V, C and J segment usage by providing two or more of the V-C entry vector and/or J donor vector to the reconstitution reaction.

A method for selecting and reconstituting a TCR open reading frame with diversified V, C and/or J segment usage, thus comprises

a. Obtaining a TCR open reading frame sequence wherein said sequence information is sufficient to identify i) variable gene segment usage ii) constant gene segment usage iii) joining gene segment usage iv) a full CDR3 sequence spanning the variable gene segment border to the joining gene segment border, and
b. selecting two or more V-C entry vectors not corresponding to the variable and constant gene segments identified in step a. i) and a. ii), respectively, and
c. selecting two or more J donor vectors not corresponding to the joining gene segment identified in step a, iii), and
d. generating an odeCDR3 corresponding to CDR3 sequence identified in step a. iv), and
e. combining b, c and d to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme, iii) subjecting the combined mix to a thermocycling reaction, and

f. transforming the reaction product obtained from step e. to a selectable host organism competent for plasmid replication, and
g. performing a selection of host organism to obtain full length reconstituted TCR open reading frame in the V-C entry vector backbone.

wherein, a) is conducted by sequencing methods well known to one skilled in the art, able to obtain sufficient sequence length and quality to identify all four required genetic elements; b) and c) are selected from a TORES library containing required vectors; d) is synthesised de novo, or selected from an odeCDR3 library; e) is conducted in a single reaction vessel.

Such a method can be achieved by pooling all V-C entry vectors and/or J donor vevtor variants to a single reaction to generate a pool of sequence-diversified but may be equally achieved by proving each vector variant to a parallel reaction.

Each V-C entry and J donor vector from a given library could be selected to provide full coverage of V, C and J gene segments.

Any combination of CDR3 and V, C and J diversification describe above could be used to generate pools or libraries of diversified TCR ORFs.

This system can be used to generate entirely synthetic libraries of TCRs ORFs with full coverage of native V, C and J gene segment usage, and defined CDR3 characteristics.

Features of a TORES with regard to Reconstitution/Diversification Methods

As mentioned above, it is desirable to conduct the assembly cycle reaction with a single Type IIS restriction enzyme. This economises the use of restriction enzyme and is made possible by the nature of Type IIS action, and the design of unique single stranded overhangs in the two-component vector system and odeCDR3.

Alternatively, up to four Type IIS restriction enzyme recognition sequences across the four Type IIS recognition sites of the V-C entry vector and J donor vector.

For efficient cloning of TCR ORF products, at least one step of negative selection is performed during the assembly of a full-length TCR ORF using the TORES, selected from

a. performing restriction enzyme digest of reaction product to eliminate parental V-C entry vector
b. performing a suicide gene selection to eliminate competent hosts transformed with parental V-C entry vector, and/or
c. performing selection of host cells transformed with parental V-C entry vector by way of reporter identification.

wherein, the negative selection is used to eliminate parental V-C entry vector that have remained undigested by the Type IIS enzyme(s), or have re-ligated to the parental form after digestion.

Elimination of parental V-C entry vector is critical, considering that the V-C entry vector backbone, and thus the positive selection marker carried in this backbone, is used for positive selection of the vector containing the full-length TCR ORF reaction product.

In the present context, negative selection is performed using a restriction enzyme site has been designed within the reaction by-product excised from the V-C entry vector (Figure 4d). This negative selection procedure is described in examples 7 and 8.

Any one, or a combination of the above-mentioned negative selection methods can be employed to eliminate parental V-C entry vector from the final cloned products. Such a negative selection procedure may be omitted if the cloning efficiency is deemed high enough for efficient recovery of cloned reaction products.

The selection of the cloned full-length TCR ORF containing vectors in transformed host organism is required to obtain the final cloned product. Such selections are well-known to those skilled in the art.

A host organism represents a transformation-competent bacterium, and the selection of transformants containing the full-length TCR ORF contained in a V-C entry vector backbone comprises antibiotic selection. This entails adding antibiotic to the culture system in which the transformed cells are placed, and resistance to this antibiotic is encoded by the gene represented as the first positive selection marker in the V-C entry vector backbone.

Alternatively, removal of auxotrophic factors of the culture system in which transformats are placed can be a form of positive selection, wherein auxotrophic complementation is

conferred by a gene product encoded in the V-C entry vector backbone. A combination of the above-described positive selections may be employed.

V-C entry vector and J donor vector libraries comprising a TORES

For the efficient operation of a TORES to perform reconstitution or diversification of selected TCR ORFs, the pre-construction of a V-C entry vector and J donor vector library is required. It is from this library, which is specific for each TCR chain form that selections are made to fulfil the V/J/C usage of the target TCR ORF sequence, when complemented with the odeCDR3 sequence.

V-C entry and J donor vector libraries may be constructed to contain all germline TCR variable, constant and joining gene segments of an organism having such TCRs. Such a library may also include all V-C combinations in the V-C entry vector, as for the TRB locus specific TORES presented in Example 3, wherein the library is replicated with both Constant gene segments against each Variable segment.

A library of V-C entry and J donor vectors may contain V/J/C gene segments, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segments.

Such a library permits change in the underlying nucleic acid sequence as to generate a library otherwise devoid of unwanted Type IIS recognition sequences, or positive and negative selection elements. Changes in the underlying nucleic acid sequence can also be used for codon optimisation, for expression reconstituted TCR chains in cells from different host organisms.

Alternatively, a library of V-C entry and J donor vectors may contain V/J/C gene segments, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segments.

Such a library may be used to construct TCRs with characteristics that are optimised for different diagnostic or therapeutic uses. Changes in framework residues or regions within the V/J/C gene segments could be used to increase expression or stability in various scenarios, such as expression of TCRs as soluble reagents. Similarly, alterations in framework regions that are not involved in direct HLA-antigen contacts may be used to alter the signalling capacity of reconstituted TCRs produced by the TORES.

Affinity tags or immunogenic sequences may also be encoded within framework regions as to aid in purification and/or detection of the reconstituted TCRs in downstream applications.

V-C entry and J donor vector libraries may be assembled into kit comprising a combination of

a. one or more V-C entry vectors encoding combinations of Variable and Constant gene segments, and
b. one or more J donor vectors encoding J gene segments, and optionally
c. one or more standardised odeCDR3 with single stranded overhangs matched to V-C entry vector and J donor vector single strand overhangs as positive control odeCDR3, and optionally
d. A pre-assembled full-length TCR ORF as a reference

wherein, a) and b) cover the required genetic diversity of gene segments from a target organism, with unmodified or modified amino acid sequence relevant for the intended application; c) is used as a positive control in reconstitution reactions d) is used as a positive control in downstream applications of full-length TCR ORFs reconstituted with the V-C entry vector and J donor vector libraries provided in said kit.

Method to construct V-C entry vectors

The assembly of V-C entry and J donor vector libraries may be achieved by a variety of molecular biology methods well known to those skilled in the art, including direct DNA synthesis of the required vectors. However, a rapid and cost-effective combinatorial approach using small gene segment-containing fragments is desirable. Such a method permits rapid cycling of V-C entry and J donor vector forms that are important for TCR engineering workflows. Similarly, a rapid expansion of a given V-C entry vector and/or J donor vector library can be used to account for single nucleotide polymorphism and other allelic differences of TCR gene segments between individuals of a given population, which may have functional significance or impact the immunogenicity of a TCR sequence in a pseudo-allogeneic therapeutic context. Systematic methods for assembling V-C entry vector and J donor vector libraries are well described in Examples 1, 2 and 3.

A method to construct a V-C entry comprises combining three DNA components selected from

a. a Variable gene segment cloning fragment
b. a Constant gene segment cloning fragment
c. a V-C entry vector backbone

wherein, a) contains the Variable gene segment; b) contains the constant gene segment; c) represent the V-C entry vector backbone into which the Variable and Constant gene fragments are assembled.

In the present context, a Variable gene segment cloning fragment comprises

a. a 5' primer bind sequence for polymerase chain reaction dependent propagation of the fragment
b. a fifth Type IIS sequence orientated to cut in the 3' direction
c. a first overhang sequence that encodes a defined single stranded overhang upon Type IIs enzyme action on the fifth Type IIs sequence in b.
d. a Kozak sequence
e. a TCR variable gene segment
f. a first Type IIS sequence
g. a 5' sequence segment of a negative selection marker
h. a sixth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated within the 5' sequence segment of the negative selection marker in g.
i. a 3' primer bind sequence for polymerase chain reaction dependent propagation of the fragment

wherein, b) and h) encode Type IIS sequences used in the assemble of the V-C entry vector; f) encodes a Type IIS sequences used in the operation of a reconstitution reaction; g) represents a fragment of the negative selection marker sequence that is completed by a complementary fragment provided in the Constant gene segment cloning fragment.

A schematic representation of the Variable gene segment cloning fragment is presented in Figure 5. Examples 1 and 3 describe the format and use of these cloning fragments to assemble V-C entry vectors for the human TRA and TRB loci, respectively.

These examples define Human TRA Variable gene segment cloning fragments as SEQ0001 to SEQ0046, and Human TRB Variable gene segment cloning fragments as SEQ0435 to SEQ0481.

To assemble a V-C entry vector, a Variable gene segment cloning fragment must be combined with a Constant gene segment cloning fragment.

A Constant gene segment cloning fragment comprises

a. a 5' primer bind sequence for polymerase chain reaction dependent propagation of the fragment
b. a seventh Type IIS sequence orientated to cut in the 3' direction such that a single stranded overhang is generated within the 3' sequence segment of the negative selection marker in c.
c. a 3' sequence segment of a negative selection marker
d. a second Type IIS sequence
e. a TCR constant gene segment
f. a second overhang sequence that encodes a defined single stranded overhang upon Type IIs enzyme action on the eighth Type IIs sequence in g.
g. an eighth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated in overhang sequence of f.
h. a 3' primer bind sequence for polymerase chain reaction dependent propagation of the fragment

wherein, b) and g) encode Type IIS sequences used in the assemble of the V-C entry vector; g) encodes a Type IIS sequences used in the operation of a reconstitution reaction; c) represents a fragment of the negative selection marker sequence that is completed by a complementary fragment provided in the Variable gene segment cloning fragment.

A schematic representation of the Variable gene segment cloning fragment is presented in Figure 6. Examples 1 and 3 describe the format and use of these cloning fragments to assemble V-C entry vectors for the human TRA and TRB loci, respectively.

These examples define Human TRB Constant gene segment cloning fragment as SEQ0047, and Human TRB Constant gene segment cloning fragments as SEQ0482 and SEQ0483.

The Variable and Constant gene segment cloning fragments are combined into a V-C

entry vector backbone to assemble a V-C entry vector,

A V-C entry vector backbone comprises

a. an origin or replication
b. a first positive selection marker
c. a 5' genetic element
d. a first restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to the overhang created within the Variable gene segment cloning fragment by Type IIS action during the assemble reaction
e. a second restriction enzyme recognition sequence permitting digestion of the backbone to create the Constant gene segment cloning fragment by Type IIS action during the assemble reaction
f. a 3' genetic element.

wherein, c) and f) represent genetic elements used for application of reconstituted TCR ORFs in differing biological systems as mentioned above.

A schematic representation of the V-C entry vector backbone is presented in Figure 7. Examples 1, 3 and 4 describe the format and use of different V-C entry vector backbone forms to assemble V-C entry vectors for TRA and TRB loci, respectively.

A V-C entry vector backbone used for downstream transient expression of reconstituted TCR ORFs in mammalian cells is defined as SEQ0048, whereas a pair of V-C entry vector backbones used for recombinase mediated cassette exchange of reconstituted TCR ORFs are defined as SEQ0688 and SEQ0689,

Rapid cycling of Variable and Constant gene cloning fragments into different V-C entry vector backbones is a cost-effective approach for altering the characteristics of the vector context of reconstituted TCR ORFs from a given V/J/C combination. This permits rapid replication of re-tasking of native or synthetic TCR gene segment libraries into different biological applications.

Within the examples of Variable and Constant gene segment cloning fragments and V-C entry vector backbones cited herein, the Type IIS enzyme used for vector assembly is BbsI, whereas the Type IIS enzyme used for TCR ORF reconstitution is BsaI. The V-

C entry vector backbone contains restriction enzyme recognition sites for Acc65I and XbaI, to create compatible overhangs with the 5' overhang and 3' overhang generated in the Variable and Constant gene cloning fragments, respectively, by BbsI action.

Any other combination of restriction enzymes could be used for assembly and reconstitution reactions, provided they satisfy the above-described criteria.

Preferably, the Type IIS sequences used for assembly of the V-C entry vector, designated the fifth, sixth, seventh and eighth Type IIS sequences above, are the same.

Alternatively, up to four different Type IIS recognition sequences could be used for this procedure.

The Type IIS sequences used for assembly of the V-C entry vector, designated the fifth, sixth, seventh and eighth Type IIS sequences above, must be different from those used for the reconstitution reaction; designated first, second third and fourth Type IIS sequences above.

The method for combining of the Variable and Constant gene segment cloning fragments with the V-C entry vector backbone to assemble a V-C entry vector is well described in examples 1 and 3.

A method for assembly of a V-C entry vector comprises

a. Digestion of the V-C entry vector backbone with the two restriction enzymes specific for the recognition sequences contained within the V-C entry vector backbone
b. Combining the digested V-C entry vector backbone with the V cloning fragment and the C cloning fragment, along with DNA ligase enzyme and one or more Type IIS restriction enzyme(s) recognising the fifth, sixth, seventh and eighth Type IIS sequences, and subjecting the combined mix to a thermocycling reaction, and
c. Transformation of the resulting reaction product into competent host organism and positive selection using said first positive selection marker to obtain complete V-C entry vector

wherein; a) creates single stranded overhangs complimentary with those generated in

Variable and Constant cloning fragments by Type IIS enzyme action in b); b) represents the digestion of Variable and Constant fragments to generate overhangs that will ligate to the overhangs generated in a), and permit ligation of the complementary overhangs generated within the negative selection marker fragments to ligate the Variable and Constant fragments; c) represents selection and propagation of the V-C entry vector product.

Generic features of V-C entry vectors and construction elements

The above descriptions use the human HLA TRA and TRB loci as templates for definition of a TORES system, as also outlined in Examples 1 to 4. However, a gene segment family from any TCR loci can be assembled into a TORES. In the above description, guidelines is given on where there is flexibility in the design of a TORES system, both for the human TRA/TRB loci, but also applicable to any other loci, from any organism. In the current section, generic features of a system are described using the TRA/TRB design presented in Examples 1 to 4 as a template.

To achieve a TORES for any given TCR loci, four sequence elements are required specific for a TCR chain encoded by said TCR loci:

X – a variable (V) gene segment fragment

Y – a constant (C) genes segment fragment

Y'– a constant (C) gene segment part

Z – a joining (J) gene segment fragment

According to the above description and examples below, these four forms of TCR sequence element can be assembled into various vector contexts to construct and deploy a TORES for any given V-J-C combination for any given TCR chain. For example, systems for the native human TRG and TRD locus, variant and/or synthetic TCR chain forms, or native TCR chain forms of an organism other than humans.

The V and C gene segment fragments are those assembled into a V-C entry vector via a V cloning fragment and C cloning fragment, respectively. This process is well described in examples 1 and 3.

A generic Variable gene segment cloning fragment, wherein the first Type IIS sequence encodes recognition for the enzyme BsaI, and the fifth and sixth Type IIS sequences encode recognition for the enzyme BbsI, and the Variable gene segment cloning fragment is thus represented by the sequence SEQ0690. The encoded Variable

gene segment fragment is denoted as XN_n, wherein X is the designation for the Variable gene segment, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

A generic Constant gene segment cloning fragment, wherein the first Type IIS sequence encodes recognition for the enzyme BsaI, and the seventh and eighth Type IIS sequences encode for recognition for the enzyme BbsI, and the Constant gene segment cloning fragment is thus represented by the sequence SEQ0691. The encoded Constant gene segment fragment is denoted as YN_n, wherein Y is the designation for the Constant gene segment, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

A V-C entry vector backbone represented by sequence SEQ0048, is used to construct V-C entry vectors suitable for transient expression or reconstituted full-length TCR open reading frames in a mammalian host cell, wherein the first and second overhangs are generated by Acc65I and XbaI enzyme action of the backbone, respectively, and the 5' and 3' genetic elements are represented by constitutive promoter and polyadenylation signal, respectively.

A V-C entry vector backbone represented by sequence SEQ0688, is used to construct V-C entry vectors suitable for recombinase mediated cassette exchange with matched genetic targets with suitable heterospecific recombinase sequences, wherein the first and second overhangs are generated by Acc65I and XbaI enzyme action of the backbone, respectively, and the 5' and 3' genetic elements are represented by F14 and F15 heterospecific recombinase sequences directing flippase activity, respectively.

A V-C entry vector backbone represented by sequence SEQ0689, is used to construct V-C entry vectors suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences, wherein the first and second overhangs are generated by Acc65I and XbaI enzyme action of the backbone, respectively, and the 5' and 3' genetic elements are represented by FRT and F3 heterospecific recombinase sequences directing flippase activity, respectively.

By combining the above-described generic V cloning fragment, C cloning fragment with a selected V-C entry vector backbone, Variant V-C entry vectors can be constructed for different downstream application of TCR ORFs reconstituted within these varying vec-

tor contexts.

A generic V-C entry vector constructed with the use of the V-C entry vector backbone having sequence SEQ0048, is represented by sequence SEQ0692. This resulting V-C entry vector is suitable for transient expression or reconstituted full-length TCR open reading frames in a mammalian host cell, wherein the first and second Type IIS sequences encode recognition for the enzyme BsaI, and the the Variable gene segment fragment is denoted XN_n, and the Constant gene segment fragment is denoted YN_n, wherein X and Y are designations for said sequences, N represents any nucleotide, and n represents the number of nucleotides in each sequence.

A generic V-C entry vector constructed with the use of the V-C entry vector backbone having sequence SEQ0688, is represented by sequence SEQ0693, containing F14 and F15 sequences suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences, wherein the first and second Type IIS sequences encode recognition for the enzyme BsaI, and the the Variable gene segment fragment is denoted XN_n, and the Constant gene segment fragment is denoted YN_n, wherein X and Y are designations for said sequences, N represents any nucleotide, and n represents the number of nucleotides in each sequence.

A generic V-C entry vector constructed with the use of the V-C entry vector backbone having sequence SEQ0689, is represented by sequence SEQ0694, containing FRT and F3 sequences suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences, wherein the first and second Type IIS sequences encode recognition for the enzyme BsaI, and the the Variable gene segment fragment is denoted XN_n, and the Constant gene segment fragment is denoted YN_n, wherein X and Y are designations for said sequences, N represents any nucleotide, and n represents the number of nucleotides in each sequence.

The use of pairs of V-C entry vectors with differing heterospecific recombinase sites may be used for each chain of a TCR chain pair, as presented in Example 4 for the human TRA and TRB chain pair. This means that in downstream application of TCR chains reconstituted in this paired TORES system, can be delivered into a genetic context with dual heterospecific recombinase receiver sites. For example, a cell line containing such dual heterospecific recombinase receiver sites for genomic integration of the TCR chain pair.

The above description of V-C entry vector, and the components from which they are assembled, is based on the use of Type IIS enzymes BbsI and BsaI for construction and operation, respectively. Based on the guidance given above, one or more alternative Type IIS enzymes may be used for each of these tasks.

Method to construct J donor vectors

As for the above-described combinatorial method for construction of V-C entry vectors, a combinatorial method may be used to construct J donor vectors. In the present method, a J donor vector is constructed in a two-step process involving the construction of an intermediate J receiving cassette vector in a first step, into which J segment parts are inserted in a second step to form the J donor vector. In this context, a J receiving cassette vector, as the J donor vector derived from it, contains a small fragment of a Constant gene segment, termed a C part. Thus a rapid combinatorial method or construction is desirable to iterate different J donor vector forms that require differential usage of Constant gene segments. This method is well described practically in Examples 2 and 3.

A method to construct J donor vector comprises combining four DNA components selected from

a. J receiving cassette fragment
b. J donor vector backbone
c. J receiving cassette vector
d. J segment part

wherein, a) contains the above-mentioned C-part and four distinct Type IIS cloning sites for vector assemble and reconstitution reaction operation; b) is the vector backbone into which one is inserted to create c); d) is the J gene segment part that is combined with c) to create a J donor vector.

A J receiving cassette fragment comprises

a. a first single stranded overhang at the 5' end complimentary to overhang sequence generated in the J donor vector backbone
b. a third Type IIS sequence orientated to cut in the 3' direction, joined with a sequence that forms a single stranded overhang when acted upon by the enzyme directed by a ninth Type IIS sequence mentioned in c.

c. a ninth Type IIs sequence orientated to cut in the 5' direction, and to create a single stranded overhang mentioned in b.
d. a negative selection marker
e. a tenth Type IIS sequence orientated to cut in the 3' direction and create a single stranded overhang at the 5' of the sequence described in f.
f. a C-part representing a 5' portion of the Constant gene fragment, with an overhang sequence at the 5' end generated by enzyme action directed by the tenth Type IIS sequence, and an overhang sequence at the 3' end generated by enzyme action directed by the fourth Type IIS sequence mentioned in g.
g. a fourth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated within the 5' sequence containing the C-part mentioned in g
h. a second single stranded overhang at the 3' end commentary to overhang sequence generated in the J donor vector backbone

wherein; a) and h) are used for directional cloning into the J donor vector backbone; b) and g) encode Type IIS sequences used in the operation of a reconstitution reaction; c) and e) encode Type IIS sequences used in the assemble of the J donor vector; d) represents a negative selection marker for elimination of parental J receiving cassette vector during assembly of a J donor vector; f) represents the fragment of the Constant gene segment carried by the J donor vector.

A schematic representation of the J receiving cassette fragment is presented in Figure 8. Examples 2 and 3 describe the format and use of J receiving cassette fragments to assemble J receiving cassette vectors for the human TRA and TRB loci, respectively.

In the present context, the J receiving cassette fragments are formed by annealing partially complimentary single stranded oligonucleotides resulting in a DNA duplex with single stranded overhangs at each termini.

A J receiving cassette fragment for the TRA locus are described as SEQ0098 and SEQ0099.

A J receiving cassette fragments for the TRB locus are described as SEQ0578 and SEQ0581, wherein two forms exist to account for the two Constant gene segments uti-

lised at the human TRB locus.

In the provided examples, the J receiving cassette fragment, the Type IIS enzyme used for J donor vector assembly is BbsI, whereas the Type IIS enzyme used for TCR ORF reconstitution is BsaI.

A J donor vector backbone comprises

a. an origin or replication
b. a second positive selection marker
c. a first restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to an overhang in the J receiving cassette fragment
d. a second restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to an overhang in the J receiving cassette fragment

wherein c) and d) permit the directional cloning of the J receiving cassette fragments through the use of unique complimentary overhang sequences.

A J donor vector backbone is represented schematically in Figure 9 and described in detail in examples 2 and 3. A J donor vector backbone sequences is represented as SEQ0097.

The J donor vector backbone contains restriction enzyme recognition sites for EcoRI and XhoI, to create compatible overhangs with the 5' overhang and 3' overhangs provided in the J receiving cassette gene cloning fragments.

Any other combination of restriction enzymes could be used for assembly and reconstitution reactions, provided they satisfy the above-described criteria.

The method for generating a J receiving cassette vector by combining a J receiving cassette fragment and J donor vector backbone is well described in in examples 2 and 3.

A J receiving cassette vector is constructed by combining the TRAJ receiving cassette fragment and the TRAJ J donor vector backbone, wherein the method comprises

a. Digestion of the J donor vector backbone with the two restriction enzymes specific for the recognition sequences contained within the J donor vector backbone
b. Combining the digested J donor vector backbone with the J receiving cassette fragment with DNA ligase enzyme
c. Transformation of the resulting reaction product into competent host organism and positive selection using said the second positive selection marker to obtain complete J receiving cassette vector

wherein; a) creates single stranded overhangs complimentary with those contained within the J receiving cassette fragments; b) represents ligation of complimentary overhangs to form the J receiving cassette vector; c) represents selection and propagation of the J receiving cassette vector product.

A schematic representation of a J receiving cassette vector is presented in Figure 10. Examples 2 and 3 describe the construction of J receiving cassette vectors for the human TRA and TRB loci, respectively.

A J receiving cassette vector for the TRA locus is described as SEQ0098, whereas those for the TRB locus are described as SEQ0582 and SEQ0583. The TRB locus utilises two Constant gene segments, thus the replica TRB J receiving cassette vectors contain C part sequences specific for each Constant gene segment.

A J receiving cassette vector is combined with a J segment part to create a J donor vector. The J segment part encodes the bulk of the J gene segment.

Said J segment part comprises

a. A first single stranded overhang sequence at the 5' end that is complementary to the overhang generated in the J receiving cassette vector, when acted on by the enzyme directed by the ninth Type IIS sequence
b. A Joining gene segment part
c. A second single stranded overhang sequence at the 3' end that is complementary to the overhang generated in the J receiving cassette vector, when acted on by the enzyme directed by the tenth Type IIS sequence

wherein, a) and c) direct directional cloning of the J segment part into the J receiving cassette vector that is digested by the Type IIS enzyme(s) provided in the assembly reaction; b) encodes the joining gene segment that is to be carried by the J donor vector product.

A schematic representation of a J segment part is presented in Figure 11. Examples 2 and 3 describe J segment parts used to construct J donor vectors for the human TRA and TRB loci, respectively.

Short J segment parts spanning the J gene segment usage of the human TRA locus are described as SEQ0099 to SEQ0210, wherein the J segment part is generated by annealing partially complimentary single stranded oligonucleotides resulting in a DNA duplex with single stranded overhangs at each termini.

A short J segment part can be used to generate J donor vectors with standardised J segment parts of minimal length, as is presented in Examples 2 and 3.

A long J segment part can be used to generate J donor vectors with extended coverage of the J gene segment to minimise the size of the odeCDR3 used in TCR reconstitution reactions. Shortening of the odeCDR3 element economises on synthesis cost for said element, and also minimises the mutational load within these oligonucleotide duplexes.

Long J segment parts spanning the J gene segment usage of the human TRA locus are described as SEQ0211 to SEQ0322, wherein the J segment part is generated by annealing partially complimentary single stranded oligonucleotides resulting in a DNA duplex with single stranded overhangs at each termini.

Short J segment parts spanning the J gene segment usage of the human TRB locus are described as SEQ0584 to SEQ0609, wherein the J segment part is generated by annealing partially complimentary single stranded oligonucleotides resulting in a DNA duplex with single stranded overhangs at each termini.

Long J segment parts spanning the J gene segment usage of the human TRB locus are described as SEQ0610 to SEQ0635, wherein the J segment part is generated by annealing partially complimentary single stranded oligonucleotides resulting in a DNA

duplex with single stranded overhangs at each termini.

In the present context, the Type IIS sequences used for assembly of the J donor vector, designated the ninth and tenth Type IIS sequences above, are the same.

Alternatively, two different Type IIS recognition sequences could be used for this procedure.

The Type IIS sequences used for assembly of the J donor vector, designated the ninth, and tenth Type IIS sequences above, must be different from those used for the reconstitution reaction; designated first, second, third and fourth Type IIS sequences above.

The Type IIS sequences used for assembly of the J donor vector, designated the ninth, and tenth Type IIS sequences above, are the same as those used to assemble the V-C entry vector; designated fifth, sixth, seventh and eighth Type IIS sequences above.

These Type IIS sequences used to assemble the V-C entry and J donor vectors need not be the same or different, as they are treated independently.

A J segment part is combined with a J receiving cassette vector that contains a matched C part to generate a J donor vector.

A J donor vector is constructed by combining a J donor vector backbone with a J segment part, wherein the method comprises

a. Combining J receiving cassette vector with the J segment part, along with DNA ligase enzyme and one or more Type IIS restriction enzyme(s) recognising the ninth and tenth Type IIS sequences, and subjecting the combined mix to a thermocycling reaction, a

b. Transformation of the resulting reaction product into competent host organism and positive selection using said second positive selection marker to obtain complete J donor vector

wherein, the Type IIS enzyme action on the ninth and tenth Type IIS sequences in step a) creates single stranded overhangs within the J receiving cassette vector, through excision of the negative selection marker sequence.

A schematic representation of a resulting J donor vector is presented in Figure 3. Examples 2 and 3 describe the use of J segment parts and a J receiving cassette vectors to construct J donor vectors for the human TRA and TRB loci, respectively.

J donor vectors containing short J segments for the human TRA locus are described as SEQ0323 to SEQ0378, whereas J donor vectors containing long J segments for the human TRA locus are described as SEQ0379 to SEQ0434.

J donor vectors containing short J segments for the human TRB locus paired with C1 constant gene segment are described as SEQ0636 to SEQ0648, whereas J donor vectors containing long J segments for the human TRA locus are described as SEQ0662 to SEQ0674.

J donor vectors containing short J segments for the human TRB locus paired with C2 constant gene segment are described as SEQ0649 to SEQ0661, whereas J donor vectors containing long J segments for the human TRA locus are described as SEQ0675 to SEQ0687.

In the provided examples, the Type IIS enzyme used to assemble a J donor vector is BbsI, whereas the Type IIS enzyme used within the reconstitution of a TCR is BsaI. The negative selection marker is a NotI restriction enzyme sequence.

The method to construct a J donor vector also entails a negative selection step to eliminate parental J receiving cassette vector prior to transformation and selection.

In the provided example, this negative selection entails a NotI digestion to eliminate parental J receiving cassette vector prior to transformation and selection.

Generic features of J donor vectors and construction elements

As described above and in Example 5 below, several features of a V-C entry and J donor vector library may be generic, inasmuch that TCR gene segment elements are constructed into a generic context to achieve a TORES system.

To achieve a TORES for any given TCR loci, four sequence elements are required specific for a TCR chain encoded by said TCR loci:

X – a variable (V) gene segment fragment

Y – a constant (C) genes segment fragment

Y'– a constant (C) gene segment part

Z – a joining (J) gene segment fragment

According to the above description and examples below, these four forms of TCR sequence element can be assembled into various vector contexts to construct and deploy a TORES for any given V-J-C combination for any given TCR chain. For example, systems for the native human TRG and TRD locus, variant and/or synthetic TCR chain forms, or native TCR chain forms of an organism other than humans.

The C gene segment part and J gene segment fragment are those assembled into J donor vector via a J receiving cassette fragment and a J cloning fragment, respectively. This process is well described in Examples 2 and 3.

A generic J receiving cassette fragment represented by the sequences SEQ0695 and SEQ0696 wherein the third and fourth Type IIS sequences encode recognition for the enzyme BsaI, and the ninth and tenth Type IIS sequences encode recognition for the enzyme BbsI, and the negative selection marker represents a NotI restriction enzyme recognition sequence. Within this sequence, the encoded C-part is denoted as $Y'N_n$, wherein Y' is the designation for the C-part, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

The J receiving cassette fragments is generated by annealing the single stranded oligonucleotides represented by the pair of sequences with partially complementary sequence.

A J donor vector backbone to construct a J receiving cassette vector with the above generic J receiving cassette fragment is represented by SEQ0097, wherein the restriction enzymes EcoRI and XhoI are used to generate single stranded DNA overhangs required for insertion of the J receiving cassette fragment sequence.

A J receiving cassette vector constructed by the combination of the J receiving cassette fragment generated from the J receiving cassette fragments represented as sequences SEQ0695 and SEQ0696, and a J donor vector backbone represented by sequence SEQ0097, is thus represented by SEQ0697. Within this resulting J receiving

cassette vector, the third and fourth Type IIS sequences encode recognition for the enzyme BsaI, and the ninth and tenth Type IIS sequences encode recognition for the enzyme BbsI, and the negative selection marker represents a NotI restriction enzyme recognition sequence. As derived from the J receiving cassette fragment, the encoded C-part is denoted as $Y'N_n$, wherein Y' is the designation for the C-part, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

The J receiving cassette vector has a J segment part inserted to obtain a J donor vector.

A generic J segment part used to insert into the generic J receiving cassette vector represented by SEQ0697, is thus represented by complementary sequences SEQ0698 and SEQ0699, wherein the J gene segment part is denoted as ZN_n, wherein Z is the designation for the J gene segment, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

A J segment part is generated by annealing the two single stranded oligonucleotides represented by SEQ0698 and SEQ0699.

A generic J donor vector assembled from the combination of the J receiving cassette vector represented by SEQ0697 and the J segment part represented by complementary sequences SEQ0698 and SEQ0699 is thus represented by sequence SEQ0700. This resulting J donor vector thus contains two annotated sequence inserts of ZN_n and $Y'N_n$, wherein Z is the designation for the J gene segment, and Y' is the designation of the C segment part, N represents any nucleotide, and n represents the number of nucleotides in said sequences.

The above description of V-C entry vector, and the components from which they are assembled, is based on the use of Type IIS enzymes BbsI and BsaI for construction and operation, respectively. Based on the guidance given above, one or more alternative Type IIS enzymes may be used for each of these tasks.

Use of Full-length TCR ORFs in defined vector contexts for diagnostics and therapeutics

A key challenge in harnessing T-cell immunity for treatment of disease is the inter-individual diversity of the HLA and TCR systems, along with the massive intra-

individual content of the TCR repertoire. This means that medicaments and therapeutic strategies that rely on the assessment and/or provision of TCRs require robust assessment of TCR function in a bona fide biological context.

To achieve accurate assessment of native TCR chain pairs, a reliable and cost-effective high-throughput method for delivering captured TCR ORFs in defined vector contexts is needed. A TORES for any given chain pair is a means to deliver such TCR ORFs.

A defined vector context for TCR ORFs can be, for example, a transient expression vector, for instance, whereby the TCRs can be rapidly characterised on the surface of human cells as described in Example 8. In this example, TCRs could be sequenced and re-expressed to confirm expected specificity. This permits the use of validated TCR sequences for following TCR clonotype abundance by sequence or amplification- or probed- based assays for diagnostic procedures during immunotherapeutic interventions on a personalised basis. Similarly, rapid capture and validation of TCR chain pairs mean that said TCR pairs could be provided in personalised medicine – such as the delivery of soluble TCR constructs as a medicinal compound, or provision of the TCR in an effector cell as a cellular therapeutic.

In a similar approach, a TORES system is ideally suited for engineering TCR ORFs to change specificity and/or function, as outlined in Examples 9 and 10. This engineering can be used to enhance or reduce signalling strength and/or change or redirect the specificity of TCR chain pairs towards specific disease antigens. Such engineered TCR sequences could be provide in personalised immunotherapeutic strategies in place of native TCR chain pairs.

The bidirectional TCR ORF Reconstitution and engineering System (TORES^2)

The above-described TORES system treats each TCR chain within independent two-part vector library systems. Thus, the products of TORES operation are discrete vectors encoding reconstituted TCR ORFs. In a separate aspect, the present invention provides an alternative system to adjoin reciprocal TCR chain pairs into a single product vector using a two-step method. This is achieved in a five-component vector library system, wherein modified V-C entry vectors are combined with the J-donor vectors of the TORES system and odeCDR3 to achieve a reconstitution of a TCR chain pair in discrete reactions. The modified V-C entry vectors then permit the reconstitution TCR

ORFs to be adjoined in a second step by addition of a fifth vector component, the Bidirectional Terminator donor vector (BiT donor), to achieve a vector encoding the two TCR OFR pairs in antiparallel coding sense.

This bidirectional TORES (TORES^2) represents a combined five-component vector system as modifications to V-C entry vectors encoding elements of the reciprocal chain pairs are non-symmetrical. Meaning that the V-C entry vectors comprise distinct arrangements. That is, the reconstituted ORF of one chain is excised from the product vector of the first step and ligated in an antisense orientation into the reciprocal product vector of the first step. Therefore the pairing of the V-C entry vectors is critical, as one of the original V-C entry vectors of the TORES^2 system represents the final product backbone, and thus encodes the desired 3' and 5' genetic elements for downstream application of the reconstituted and adjoined TCR pair. For clarity, the description and examples presented below fix the TRA chain as the final product backbone (e.g. V-Cα entry vector) and the TRB as the chain integrated to this final product backbone (e.g. V-Cβ entry vector). The reciprocal arrangement is equally as valid, as is any other combination of reciprocal TCR chain pairs.

V-C entry vector components of TORES^2

As mentioned above, the five-component vector library system comprising TORES^2 differs from the TORES system in the provision of modified V-C entry vector contexts. These modifications incorporate distinct Type IIS sites (labelled as Type IIS #2) and negative selection elements (labelled as –ve selection #2) from those utilised for TCR ORF reconstitution in a first step, to direct the adjoining of those reconstituted ORFs into a single vector in a second step. In effect, the TORES^2 system incorporates the TORES system within a new V-C entry vector context. The description below treats the TRA chain as the backbone of the final product, and the TRB chain as that ligated into the reconstituted TRA chain backbone in an antisense orientation. The same framework applies for any pair of TCR chains, and there is no particular reasons as to why either of the TCR chains need to be in one of the described vector arrangements, or the other.

The V-Cα entry vector contains,

a. origin of replication,
b. a first positive selection marker,
c. 5' genetic element, or elements,

d. Kozak Sequence,
e. TCR alpha variable gene segment,
f. a first Type IIS sequence, for site specific recognition and cleavage by a Type IIS restriction enzyme (for Type IIS enzyme #1),
g. a first negative selection marker,
h. a second Type IIS sequence (for Type IIS enzyme #1),
i. TCR constant gene segment,
j. A third Type IIS sequence (for Type IIS restriction enzyme #2),
k. A second negative selection marker,
l. A fourth Type IIS sequence (for Type IIS restriction enzyme #2), and
m. 3' genetic element, or elements

wherein, a) and b) are used for propagation and selection of both parental V-C entry vector and the reconstituted TCR-containing vector in a bacterial host; c) and m) are used to define the downstream application of the reconstituted and adjoined full-length TCR ORFs, and may include elements as described above for the TORES; d) ensures efficient initiation of translation in eukaryotic cells, which could alternatively represent a Shine-Dalgarno sequence for transitional regulation in prokaryotes and archaea; e) represents the variable (V) alpha gene segment from the start codon to a motif at the 5' edge of the CDR3 region conserved across all V segments in a given two-component vector system; f) represents a Type IIS #1 recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang at the 3' end of the V gene segment; g) represents a negative selection marker #1 to eliminate parental V-C entry vector during operation of the system to reconstitute a full-length TCR ORF; h) represents a Type IIS #1 recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang at the 5' end of the C gene segment; i) represents the constant (C) gene segment from a motif at the 5' end of the C gene segment conserved across all C segments in a given two-component vector system, and which defines the boundary with the J segment (see Figures 2 and 4); j) represents a Type IIS #2 recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang at the 3' end after the C gene segment, this Type IIS enzyme is different to the first and second Type IIS #1 en-

zymes; k) represents a negative selection marker #2 to eliminate parental TRA ORF vector during the operation of creating a bidirectional vector construct, it is distinct from the first negative selection marker; l) represents a Type IIS #2 recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang before the 3' genetic element, this Type IIS enzyme is different to the first and second Type IIS #1 enzymes; m) represents the 3' genetic element, or elements.

The arrangement of the above-described V-Cα entry vector is depicted in Figure 20 a.

Generally, the first and second Type IIS enzymes (designated as Type IIS #1) are the same but may also be different. Similarly, the third and fourth Type IIS enzymes (designated as Type IIS #2) are generally the same but may also be different. Critically, the Type IIS enzyme(s) used for the first and second Type IIS (Type IIS #1) sites must be different from the enzyme(s) used for the third and fourth Type IIS (Type IIS #2) sites.

A V-Cα entry vector backbone is represented by SEQ0756, as presented in Example 11.

Example 11 below presents a V-Cα entry vector of the TORES[2] for the human TRA locus (SEQ0756), wherein some of the V/C sequences have been modified relative to those incorporated into the TORES system as to be devoid of the Type IIS enzyme used in the third and fourth sites (SEQ0757 to SEQ0763)

The V-Cβ entry vector contains,

a. origin of replication,
b. a first positive selection marker,
c. 5' genetic element, or elements,
d. a first Type IIS sequence, for site specific recognition and cleavage by a Type IIS #2 restriction enzyme
e. Kozak Sequence,
f. TCR beta variable gene segment,
g. a second Type IIS sequence, for site specific recognition and cleavage by a Type IIS restriction enzyme (Type IIS #1),
h. a first negative selection marker,

i. a third Type IIS sequence (Type IIS #1),
j. TCR constant gene segment,
k. a fourth Type IIS sequence (Type IIS #2),
l. a second negative selection marker, and
m. 3' genetic element, or elements

wherein, a) and b) are used for propagation and selection of both parental V-C entry vector and the reconstituted TCR-containing vector in a bacterial host; c) and m) optionally define the downstream application of the reconstituted full-length TCR ORF as described above for the TORES system; d) represents a Type IIS #2 recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang after the 5' genetic element and before the Kozak sequence, this Type IIS enzyme is different to the second and third Type IIS #1 enzymes; e) ensures efficient initiation of translation in eukaryotic cells, which could alternatively represent a Shine-Dalgarno sequence for transitional regulation in prokaryotes and archaea; f) represents the variable (V) gene segment from the start codon to a motif at the 5' edge of the CDR3 region conserved across all V segments in a given two-component vector system; g) represents a Type IIS #1 recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang at the 3' end of the V gene segment; h) represents a negative selection marker to eliminate parental V-C entry vector during operation of the system to reconstitute a full-length TCR ORF; i) represents a Type IIS #1 recognition sequence that directs a Type IIS restriction enzyme to cut in the 3' direction as to create a standardised single stranded overhang at the 5' end of the C gene segment; j) represents the constant (C) gene segment from a motif at the 5' end of the C gene segment conserved across all C segments in a given two-component vector system, and which defines the boundary with the J segment (see Figures 2 and 4); k) represents a Type IIS #2 recognition sequence that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang after the C segment but before the 3' genetic element, this Type IIS enzyme is different to the second and third Type IIS #1 enzymes; m) represents a second negative selection marker, which is different to the first -negative selection marker, l) represents 3' genetic element or elements.

The arrangement of the above-described V-Cβ entry vector is depicted in Figure 20b.

The 3' and 5' genetic elements are optional in the V-Cβ entry vector context, as the ORF reconstituted within this vector is later excised and ligated into the V-Cα entry vector backbone context. That is, the V-Cα entry vector backbone.

Generally, the second and third Type IIS enzymes (designated Type IIS #1) are the same but may also be different. Similarly, the first and fourth Type IIS enzymes (designated Type IIS #2) are generally the same but may also be different. Critically, the Type IIS enzyme(s) used for the second and third Type IIS sites (Type IIS #1) must be different from the enzyme(s) used for the first and fourth Type IIS sites (Type IIS #2).

Generally, the first and second Type IIS sites (Type IIS #1) of the V-Cα entry vector are the same as the second and third of the V-Cβ entry vector, but need not be as these reactions may be operated independently. The reconstitution of TRA chains and TRB chains in the described system could be conducted in a single reaction if the first and second Type IIS sites of the V-Cα entry vector were distinct from those of the second and third Type IIS sites of the V-Cβ entry vectors.

Generally, the third and fourth Type IIS sites (Type IIS #2) of the V-Cα entry vectors are the same as the first and fourth of the V-Cβ entry vectors as to require only the addition of a single Type IIS enzyme (Type IIS #2) to the second reaction adjoining reaction.

Generally, the second negative selection marker (-ve selection #2) of each V-Cα and V-Cβ entry vectors are the same as to facilitate efficient negative selection of parental vectors in the adjoining step of the TORES^2 operation, but may also be different.

A V-Cβ entry vector backbone is represented by SEQ0764, as presented in Example 11.

Example 11 below presents a V-Cβ entry vector of the TORES^2 for the human TRB locus (SEQ0764), wherein some of the V/C sequences have been modified relative to those incorporated into the TORES system as to be devoid of the Type IIS enzyme used in the first and fourth sites (SEQ0765 to SEQ0776)

J donor and odeCDR3 components of the TORES^2

The J donor vector and the oligonucleotide duplex encoding the CDR3 are the same as

already described for the conventional TORES above. Both are used within the first step reaction(s) to reconstitute the reciprocal TCR chain pairs within the designated V-C entry vector backbone contexts in a procedure identical to that described above for the TORES.

Bidirectional terminator donor vector (BiT donor) of the TORES²

The first four vector library components of the TORES² represent the V-Cα and V-Cβ entry vectors, along with the J donor vectors for each of the TRA and TRB chain in the TRA/TRB focused description. Again, it should be noted that any TCR chains may be incorporated into this system following the described design principles, and the specific TRA and TRB configuration is fixed for clarity.

The fifth component represents Bidirectional Terminator Donor vector (BiT donor), which provides a bidirectional terminator element that is interposed by the antiparallel sense TRA and TRB chains reconstituted in the first step reaction, and adjoined by this bidirectional terminator element in the second step reaction.

The bidirectional terminator is generally provided as a vector and is also surrounded by two Type IIS restriction enzymes (Type IIS #2). It is equally as relevant to provide the bidirectional terminator element as a linear dsDNA construct with these Type IIS sites, or with target single strand DNA overhangs without Type IIS sites.

The vector containing the bidirectional terminator contains,

a. origin of replication,
b. a second positive selection marker,
c. first Type IIS sequence (Type IIS #2),
d. bidirectional terminator, and
e. second Type IIS sequence (Type IIS #2)

wherein, a) and b) are used for propagation and selection of the vector carrying the bi-directional terminator in a bacterial host, c) represents a Type IIS recognition sequence that directs a Type IIS restriction enzyme (Type IIS #2) to cut in the 3' direction as to create a standardised single stranded overhang, d) represents the bidirectional termi-nator which is used to ensure transcriptional termination on both sense and antisense DNA strands during expression of TCRs reconstituted into the TRA backbone, when said vector containing a reconstituted TCR ORF is transfected into mammalian cells; e)

represents a Type IIS recognition sequence (Type IIS #2) that directs a Type IIS restriction enzyme to cut in the 5' direction as to create a standardised single stranded overhang.

The arrangement of the above-described BiT donor vector is depicted in Figure 21.

Generally, the positive selection marker contained within the BiT donor is distinct from that of positive selection markers carried by the V-Cα and V-Cβ entry vectors, as to minimise carryover of parental BiT donor in the second step adjoining reaction.

Generally, the first and second Type IIS sites (Type IIS #2) in the BiT donor are the same as both the third and fourth Type IIS sites of the V-Cα entry vectors and the first and fourth of the V-Cβ entry vectors, as to require only the addition of a single Type IIS enzyme (Type IIS #2) to the second adjoining reaction.

The bidirectional terminator element sequence is represented by SEQ0777.

Operation of TORES^2 to reconstitute and adjoin full length TCR ORF pairs

The operation of TORES^2 is a two-step process, incorporating a first TCR ORF reconstitution step, and second TCR ORF adjoining step. The method for reconstitution of TCR ORFs is identical to that described above for the TORES (see Figures 1 and 4). The products of the first step are therefore analogous to the products of TORES, comprising vectors encoding full-length TCR ORFs containing the V-CDR3-J-C element arrangement. However, the TORES^2 system provides these TCR ORFs in a distinct backbone context from the TORES system, wherein the TORES^2 backbone contains additional cloning sites to facilitate the second adjoining step reaction. This second step restriction enzyme and ligase cycle reaction entails the adjoining of the reconstituted TCR ORFs from the first step, in an antiparallel sense and interposing a bidirectional terminator element provided by the BiT donor. The overall process is depicted in Figure 22, which depicts the above-described TRA/TRB TORES^2 arrangement.

The reconstituted TRA encoding vector (Figure 22a), TRB encoding vector (Figure 22b) and the BiT donor (Figure 22c) are reacted in a single tube with Type IIS enzyme (Type IIS #2) and ligase and subjected to a restriction enzyme and ligase cycle reaction. The Type IIS restriction enzyme digests the three provided vectors into three reaction by-products, and three reaction intermediates with designed overhangs for direc-

tional ligation into the reaction product.

The three reaction by-products are represented by the excised –ve selection element #2 and Type IIS #2 sites from the TRA vector (figure 22d), the opened TRB encoding vector backbone from which the reconstituted TRB ORF has been excised (Figure 22e) and the opened BiT donor backbone from which the bidirectional terminator element has been excised (Figure 22f).

The first of the reaction intermediates is the open TRA vector, encoding the reconstituted TRA ORF in the V-Cα entry vector backbone context (Figure 22g). The third Type IIS #2 site in the original V-Cα entry vector as described above directs the enzyme to cut in the 5' direction to create a standardised single stranded overhang at the 3' end after the C gene segment (Figure 22 g, overhang ‡ 1-5'), while the fourth Type IIS #2 site directs enzyme cleavage in the 3' direction to create a standardised single stranded overhang before the 3' genetic element (Figure 22 e overhang ‡ 3-3').

The second reaction intermediate is the TRB ORF encoding fragment excised from the V-Cβ entry vector context. The first Type IIS #2 site in the V-Cβ entry vector described above directs the enzyme to cut in the 3' direction to create an overhang prior to the Kozak sequence (Figure 22h overhang ‡ 3-5'). The fourth Type IIS #2 site directs the enzyme to cut in 5' direction, creating an overhang after the C-segment (Figure 22h overhang ‡ 2-3'). It should be noted that Figure 22h depicts this intermediate in the antisense orientation in which it will be ligated into the product vector context.

The third reaction intermediate is the bidirectional terminator element excised from the BiT donor, where the first Type IIS #2 site directs the enzyme to cut in the 5' direction to yield a standardised single stranded overhang (Figure 22i overhang ‡ 1-3'). The second Type IIS #2 site directs the enzyme to cut in the 3' direction to create a standardised single stranded overhang (Figure 22i overhang ‡ 2-5').

Within the restriction enzyme and ligase cycle reaction, the single stranded overhangs drive ligase-mediated directional ligation into the product vector, adjoining both TRA and TRB chains within the original V-Cα entry vector backbone context (Figure 22j).

A TORES^2 may be constructed for any collection of TCR chains. For clarity, the TRA and TRB loci are used as a case study for description. The construction of such a

TORES^2 is equally applicable in the context of the TRD and TRG loci, encoding the TCR delta and gamma chain pair, respectively, or indeed for any TRA/TRB, TRD/TRG or variant TCR chain pair system found in jawed vertebrates.

In the examples below, a TORES^2 system is provided for the human TRA/TRB loci, and applied to the reconstitution of a model human TCR ORF, and subsequently integrated into a matched engineered cell line harbouring RMCE sites matched with the 5' and 3' genetic elements of the originating V-Cα entry vector backbone context. This demonstrates that TORES^2 is readily used in combination with matched engineered cell lines optimised for integration and presentation of reconstituted TCR ORFs in a two-part device configuration similar to that described in application WO 2018/083318 A1.

In the following is given a table showing the sequences mentioned herein.

SEQ ID	Name	Reference example	Description
0001-0046	TRA V cloning fragments	Example 1	Full DNA sequences of the TRA V fragment
0047	TRA C constant cloning fragment	Example 1	Full DNA sequence of the TRA C fragment
0048	V-C entry vector backbone transient	Example 1	DNA sequence of the vector backbone from the 5′ genetic element encoding the CMV constitutive promoter to the 3′ genetic element encoding the SV40pA polyadenylation signal
0049-0094	TRA V-C entry vector library sequence	Example 1	DNA sequences of the cloned V-C fragments that make up the TRA V-C entry vector library
0095-0096	TRA J receiving cassette fragments	Example 2	Full DNA sequence of the TRA J receiving cassette fragment oligonucleotides
0097	J donor backbone	Example 2	J donor vector backbone is used to insert the TRA J receiving cassette fragment to create the TRA J receiving cassette vector
0098	TRA J receiving cassette vector	Example 2	See above

0099-0210	TRA J Short segment part	Example 2	Encodes all amino acids from the start of the CDR3-J border Phe codon
0211-0322	TRA J Long segment part	Example 2	Encodes more amino acids N-terminal of the CDR3 border amino acids
0323-0378	TRA J Short donor vector	Example 2	TRA short J donor library
0379-0434	TRA J Long donor vector	Example 2	TRA long J donor library
0435-0481	TRB V cloning fragment	Example 3	Full DNA sequences for the TRB V cloning fragments
0482-0483	TRB C constant cloning fragments	Example 3	Full DNA sequences of the TRB C cloning fragments
0484-0577	TRB V-C entry vector library sequence	Example 3	Sequences of the cloned V-C fragments that make up the TRA V-C entry vector library
0578-0581	TRB J receiving cassette fragments	Example 3	TRB J receiving cassette fragments are constructed and inserted into a J donor vector backbone to create a TRB J receiving cassette vector
0582-0583	TRB J receiving cassette vectors	Example 3	See above
0584-0609	TRB J Short segment part	Example 3	DNA sequences of the short TRB J segment parts
0610-0635	TRB J Long segment part	Example 3	DNA sequences of the long TRB J segment parts
0636-0648	TRB C1 J Short donor vector	Example 3	TRB C1 short J donor library
0649-0661	TRB C2 J Short donor vector	Example 3	TRB C2 short J donor library
0662-0674	TRB C1 J Long donor vector	Example 3	TRB C1 long J donor library
0675-0687	TRB C2 J Long donor vector	Example 3	TRB C2 long J donor library
0688	V-C entry vector backbone F14-F15	Example 4	F14/F15 V-C entry vector backbone sequence used to construct TRA V-C entry library
0689	V-C entry vector backbone FRT-F3	Example 4	FRT/F3 V-C entry vector backbone sequence used to construct TRB V-C entry library

0690	Generic V cloning fragment XN_n (left part)	Example 5	Contain the required V gene segment fragment in the context of appropriate cloning sites for final V-C entry vector assembly (full sequence: 0690-XN_n-0745)
0691	Generic C cloning fragment YN_n (left part)	Example 5	full sequence: 0691-YN_n-0746
0692	Generic Transient V-C Entry vector XN_n-YN_n (left part)	Example 5	full sequence: 0692-XN_n-0747-YN_n-0748
0693	Generic F14-F15 V-C Entry vector XN_n-YN_n (left part)	Example 5	full sequence: 0693-XN^n-0749-YN_n-0750
0694	Generic FRT-F3 V-C Entry vector XN_n-YN_n (left part)	Example 5	full sequence: 0694-XN_n-0751-YN_n-0752
0695-0696	Generic J receiving cassette oligonucleotides $Y'N_n$ (left part)	Example 5	full sequence of 0695: 0695-$Y'N_n$-TGAGACCC; full sequence of 0696: 0696-Y'Nn-0753
0697	Generic J receiving vector $Y'N_n$ (left part)	Example 5	full sequence: 0697-$Y'N_n$-0754
0698-0699	Generic J segment oligonucleotides ZN_n	Example 5	full sequence of 0698: ctcgZN_n; full sequence of 0699: Y'N(4)ZN_n; for both 0698 and 0699, the dummy sequence “aaaaaaaaaa” was included in the sequence listing due to the 10 nucleotide minimum requirement of the ST.25
0700	Generic J Donor vector ZN_n-$Y'N_n$ (left part)	Example 5	full sequence: 0700-ZN_n-$Y'N_n$-0755
0701-0702	JG9 TRA and TRB full sequences copy	Example 6	DNA sequences of the TRA and TRB chains
0703-0706	JG9 odeCDR3 sequences	Example 6	odeCDR3 synthesised for the TRA and TRB chains
0707-0718	Sequence results from Example 8	Example 8	
0719-0742	odeCDR3 collection for reconstitution Example 8	Example 8	
0743-0744	degenerate TRA odeCDR3s Example 9	Example 9	odeCDR3 oligos

0745	Generic V cloning fragment XN_n (right part)	Example 5	Contain the required V gene segment fragment in the context of appropriate cloning sites for final V-C entry vector assembly (full sequence: 0690-XN_n-0745)
0746	Generic C cloning fragment YN_n (right part)	Example 5	full sequence: 0691-YN_n-0746
0747	Generic Transient V-C Entry vector XN_n-YN_n (middle part)	Example 5	full sequence: 0692-XN_n-0747-YN_n-0748
0748	Generic Transient V-C Entry vector XN_n-YN_n (right part)	Example 5	full sequence: 0692-XN_n-0747-YN_n-0748
0749	Generic F14-F15 V-C Entry vector XN_n-YN_n (middle part)	Example 5	full sequence: 0693-XN_n-0749-YN_n-0750
0750	Generic F14-F15 V-C Entry vector XN_n-YN_n (right part)	Example 5	full sequence: 0693-XN_n-0749-YN_n-0750
0751	Generic FRT-F3 V-C Entry vector XN_n-YN_n (middle part)	Example 5	full sequence: 0694-XN_n-0751-YN_n-0752
0752	Generic FRT-F3 V-C Entry vector XN_n-YN_n (right part)	Example 5	ull sequence: 0694-XN_n-0751-YN_n-0752
0753	Generic J receiving cassette oligonucleotides $Y'N_n$ (right part)	Example 5	full sequence: 0696-$Y'N_n$-0753
0754	Generic J receiving vector $Y'N_n$ (right part)	Example 5	full sequence: 0697-$Y'N_n$-0754
0755	Generic J Donor vector ZN_n-$Y'N_n$ (right part)	Example 5	full sequence: 0700-ZN_n-$Y'N_n$-0755
0756	V-Cα entry vector backbone FRT/F3	Example 11	V-Cα entry vector backbone for TORES2
0757-0763	Modified TRA V-C sequences	Example 11	Full DNA sequences of the TRA V fragment, modified for TORES2
0764	V-Cβ entry vector backbone FRT/F3	Example 11	V-Cβ entry vector backbone for TORES2
0765-0776	Modified TRB V-C sequences	Example 11	Full DNA sequences of the TRB V fragment, modified for TORES2
0777	Bidirectional terminator element	Example 11	Bidirectional terminator element (Esp3I to Esp3I site)

0778-0779	Model TRA/TRB pair	Example 11	known specificity for a HLA-A*02:01-restricted antigen.

List of abbreviations

APC	Antigen-presenting cell
BiT	Bidirectional Terminator donor
CAR-T	CAR T-cell
CAR	Chimeric antigen receptor
CDR	Complementarity-determining regions
C segment	Constant segment (also C region)
CMV	Cytomegalovirus
DAMPS	Danger associated molecular patterns
DC	Dendritic cells
DNA	Deoxyribonucleic acid
dsDNA	Double stranded Deoxyribonucleic acid molecule
D segment	Diversity segment (also D region)
eAPC	Engineered antigen-presenting cell
FACS	Fluorescence-activated cell sorting
FRT	Flippase recognition target
GEM T-cells	Germ line-encoded mycolyl-reactive T-cells
GFP	Green fluorescent protein
HLAI	HLA class I
HLAII	HLA class II
HDR	Homology directed recombination
HLA	Human leukocyte antigen
IgSF	Immunoglobulin superfamily
IRES	Internal ribosome entry site
ITAM	Immunoreceptor tyrosine-based activation motif
iNK T-cells	Invariant natural killer T-cells
J segment	Joining segment (also J region)
MACS	Magnetic-activated cell sorting
MAGE	Melanoma associated antigen
MAIT	Mucosal-associated invariant T

odeCDR3	Oligonucleotide duplex encoding CDR3
ORF	Open reading frame
PAMPS	Pathogen-associated molecular patterns
PCR	Polymerase chain reaction
RMCE	Recombinase mediated cassette exchange
RFP	Red fluorescent protein
RT	Reverse Transcription
DNA	Ribonucleic acid
SH2	Src homology 2
T-cells	T lymphocytes
TCR	T-cell Receptor
TRA	TCR alpha
TRB	TCR beta
TRD	TCR delta
TRG	TCR gamma
TAA	Tumour-associated-antigens
TORES	TCR ORF Reconstitution and Engineering System
TORES2	Bidirectional TCR ORF Reconstitution and Engineering System
V segment	Variable segment (also V region)
β2M	β2-microglobulin
ZAP-70	ζ-chain-associated protein of 70 kDa

List of definitions

A pair of complementary TCR chains: two TCR chains wherein the translated proteins are capable of forming a TCRsp on the surface of a TCR presenting cell

Affinity: Kinetic or equilibrium parameter of an interaction between two or more molecules or proteins

Allele: Variant form of a given gene

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

Analyte: an entity that is of interest to be identified and/or measured and/or queried

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.

APC: Antigen-presenting cell. A cell capable of presenting antigen on its cell surface, generally in the context of an HLA.

C (-segment): Constant segment. Also Constant region. One of the gene segments that is used to assemble the T-cell receptor. The c-region is a distinct segment that rather than driving diversity of the TCR, defines its general function in the immune system.

C cloning fragment: Constant Cloning fragment. Also referred to as a C gene segment cloning fragment. A construct carrying a portion of a C gene segment used to construct a V-C entry vector.

C-part: Constant part. A small portion of Constant gene segment sequence carried by a J receiving cassette fragment, J receiving cassette and J donor vector to standardise overhang sequences for operation of the TORES to reconstitute TCR ORFs.

CDR: complementarity-determining regions. Short sequences on the antigen-facing end of TCRs and antibodies that perform most of the target binding function. Each antibody and TCR contains six CDRs and they are generally the most variable part of the molecules allowing detection of a large number of diverse target molecules.

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

D (-segment): Diversity segment. Also D region. One of the gene segments that is used to assemble the T-cell receptor. Each individual has a large number of different variations of these regions making it possible for each individual to arm T-cells with a very large variety of different TCR.

Diversification: A process where a sequence is diversified.

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

Endogenous: Substance that originated from within a cell

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 an 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*.

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

HLA I: Human Leukocyte Antigen class I. A gene that is expressed in humans in all nucleated cells and exported to the cell surface where it presents as cargo short fragments, peptides, of internal proteins to T-cell receptors. As such it presents fragments of potential ongoing infections along with intrinsic proteins. The HLA I can additionally present as cargo peptides that are added to the culture medium, generated from proteins expressed form introduced genetic elements or generated from proteins that are taken up by the cell. HLA class I genes are polymorphic meaning that different individuals are likely to have variation in the same gene leading to a variation in presentation. Related to HLA class II.

HLA II: Human Leukocyte Antigen Class II. A gene that is expressed in humans in specific cells that are coordinating and helping the adaptive immune response for example dendritic cells. Related to HLA class I. HLA class II proteins are exported to the cell surface where they present as cargo short fragments, peptides, of external proteins

to T-cell receptors. As such it presents fragments of potential ongoing infections along with intrinsic proteins. The HLA II can additionally present as cargo peptides that are added to the culture medium, generated from proteins expressed form introduced genetic elements or generated from proteins that are taken up by the cell. HLA class II genes are polymorphic meaning that different individuals are likely to have variation in the same gene leading to a variation in presentation.

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.

Immune surveillance: Process in which the immune system detects and becomes activated by infections, malignancies or other potentially pathogenic alterations.

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

J (-segment): Joining segment. Also J region. One of the gene segments that is used to assemble the T-cell receptor. Each individual has a large number of different variations of these regions making it possible for each individual to arm T-cells with a very large variety of different TCR.

J-C entry vector: The vector of the two-component vector system that carries the J and C TCR segments, and which receives sequences from the V donor vectors and odeCDR3 during reconstitution of a full-length TCR ORF.

J donor backbone: Joining donor backbone. The vector backbone into which a J receiving cassette fragment is inserted to create a J receiving cassette vector.

J donor vector: The vector of the two-component vector system that carries the J TCR segment, and donates this segment to the V-C entry vector during reconstitution of a full-length TCR ORF.

J receiving cassette fragment: Joining receiving cassette fragment. A cloning fragment that carries a C-part used to construct a J receiving cassette vector.

J receiving cassette vector: Joining receiving cassette vector. The vector, carrying a C-part, into which a J segment part is inserted to create a J donor vector.

J segment part: Joining segment part. A DNA construct carrying a portion of a J gene segment that is inserted into a J receiving cassette vector to generate a J donor vector.

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)

Major HLA class I: a Family of APX that comprise of the genes HLA-A, HLA-B and HLA-C

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

Meganuclease recognition site: A DNA sequence that is recognized by a endodeoxyribonuclease, commonly referred to as a meganuclease

Mobile genetic element: A DNA sequence that can permit the integration of DNA with the activity of transposase enzymes

N is a nucleotide code indicatin4g0 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

odeCDR3: oligonucleotide duplex encoding complementarity-determining regions. A synthetic construct carrying CDR3 genetic sequence with terminal overhangs, used in conjunction with the two-component vector system to reconstitute a full-length TCR ORF.

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 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 a protozoan.

Polymorphic: Present in different forms in individuals of the same species through the presence of different alleles of the same gene.

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

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.

Reconstitution: In the present context, reconstitution describes the operation of the TORES and TORES^2 to assemble a TCR ORF. A 'reconstitution reaction' is the reaction mix and thermocycling reaction that this operation entails.

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.

Selectable marker: A DNA sequence that confers a trait suitable for artificial selection methods.

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.

T-cell: T lymphocyte. White blood cell that expresses a T-cell receptor on its surface. Selected by the immune system to not react with the own body but have the potential to recognize infections and malignancies as well as reject grafts from most members of the same species.

TCR: T-cell Receptor. Affinity molecule expressed by a subgroup of lymphocytes called T-lymphocytes. In humans the TCR recognizes cargo presented by APX CM or APX AM, including fragments from virus or bacterial infections or cancerous cells. Therefore, the TCR recognition is an integral part of the adaptive immune system. The TCR consists of two chains that are paired on the cell surface. The TCR expressed on the surface of each cells is assembled at random from a large pool of varied genes (the v,d,j and c segments) and thus each individual has a pool of T-cells expressing a very large and diverse repertoire of different TCRs.

TRA: TCR alpha encoding locus. One of the four different locus encoding genes that can form a VDJ recombined TCR chain. Translated TCR alpha chain proteins typically pair with translated TCR beta chain proteins to form alpha/beta TCRsp.

TRB: TCR beta encoding locus. One of the four different locus encoding genes that can form a VDJ recombined TCR chain. Translated TCR beta chain proteins typically

pair with TCR alpha chain proteins to form alpha/beta TCRsp.

TRD: TCR delta encoding locus. One of the four different locus encoding genes that can form a VDJ recombined TCR chain. Translated TCR delta chain proteins typically pair with translated TCR gamma chain proteins to form gamma/delta TCRsp.

TRG: TCR gamma encoding locus. One of the four different locus encoding genes that can form a VDJ recombined TCR chain. Translated TCR gamma chain proteins typically pair with translate TCR delta chain proteins to form gamma/delta TCRsp.

Type IIS Restriction Enzyme: restriction enzymes that recognize asymmetric DNA sequences and cleave outside of their recognition sequence.

V (-segment): Variable region. Also V region. One of the gene segments that is used to assemble the T-cell receptor. Each individual has a large number of different variations of these regions making it possible for each individual to arm T-cells with a very large variety of different TCR.

V-C entry vector: The vector of the two-component vector system that carries the V and C TCR segments, and which receives sequences from the J donor vectors and odeCDR3 during reconstitution of a full-length TCR ORF.

V cloning fragment: Variable Cloning fragment. Also referred to as a V gene segment cloning fragment. A construct carrying a portion of a V gene segment used to construct a V-C entry vector.

V donor vector: The vector of the two-component vector system that carries the V TCR segment and donates this segment to the J-C entry vector during reconstitution of a full-length TCR ORF.

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

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

Legends to figures

Figure 1. Two-component vector system for rapid full-length TCR ORF reconstitution.

A)

Depicted are the core features of the two-component vector system (Box i), the required oligo duplex input (Box ii) and the resulting full-length TCR ORF vector (Box iii). The first component represents a V-C entry vector, incorporating a selected TCR V (variable) gene segment and a TCR C (constant) gene segment. This entry vector represents the backbone in which the final full-length TCR ORF will be carried. Thus, any desirable features of this vector backbone are tailored to the intended downstream application. Usually, this will be represented by at least a 5' and 3' genetic element. The required V gene segment and C gene segment of the target full-length TCR ORF is selected from a library of V-C entry vectors to construct a TCR ORF of any desired V/C combination (Box i). The second component represents a J donor vector, containing a selected J (joining) gene segment. This donor vector acts to donate the J gene segment, and thus the backbone of this vector is a reaction by-product. The desired J gene segment of the target TCR ORF is selected from a library J donor vectors. To complete the sequence required for the full-length TCR ORF, a short oligo duplex is synthesized corresponding largely to the CDR3 region of the mature TCR ORF (Box ii). This oligo duplex is synthesized with single strand overhangs compatible with unique overhangs generated by restriction enzyme digestion of V-C entry and J donor vectors. When combined into a single-tube reaction along with appropriate restriction and ligase enzymes, the full-length TCR ORF is reconstituted from the sequences provided by the V-C entry and J donor vectors selected from a pre-existing library, and the synthesized oligo duplex (Box iii).

B)

Function of the overall library feature of the TCR ORF reconstitution system is illustrated. A single V-C entry vector is selected from a library of V-C entry vectors with varying V-C combinations (Box i). This selection is based on the required V-C combination sequences for a selected TCR chain. A single J entry vector is selected from a library of J donor vectors with varying J gene segments encoded (Box ii). This selection is based on the required J combination sequences for the same selected TCR chain as that of the V-C entry vector. Finally, an *oligomeric duplex encoding CDR3* (odeCDR3) is selected as to complete the full-length ORF of the target TCR chain (Box iii). These three components (Box i, ii and iii) are combined into a single reaction along with appropriate restriction and ligase enzymes. The reaction cycle produces a reconstituted full-length TCR ORF in a single step in the V-C entry vector backbone context (Box iv).

Figure 2. Plasmid schematic of the genetic features of a generic V-C entry vector. Depicted is a circularized plasmid schematic of a V-C entry vector with minimally required genetic elements depicted as labelled boxes. **Kozak,** refers to consensus sequence that plays a role in the efficient initiation of translation. **V-segment,** refers to a selected sequence encoding a proportion of a TCR variable germline ORF, or mutant/synthetic ORF. **Type IIS ←**, refers to a Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 5' direction. **Type IIS →**, refers to a Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 3' direction. **-ve selection,** refers to a negative selection element designed to be detrimental to a plasmid harbouring the sequence during the full-length TCR reconstruction reaction, or subsequent selection steps. **C-segment,** refers to a selected sequence encoding a proportion of a TCR constant germline ORF, or mutant/synthetic ORF. **+ve selection #1,** refers to a the first positive selection marker of the two-component system used to convey a selective advantage to the organism harbouring the vector, and which is different to the positive selection marker of the second vector component (see figure 3). **Ori,** refers to an origin of replication used for the propagation of plasmid within a compatible host. **5′ genetic element,** refers to any desired genetic element that provides attributes required for downstream application of the reconstructed full-length TCR, and should be situated 5' of the reconstructed full-length TCR , for example, a cis-acting element. **3′ genetic element, refers to a** to any desired genetic element that provides attributes required for downstream application of the reconstructed full-length TCR, and should be situated 3' of the full-length TCR ORF, for example, a transcriptional terminator element.

Figure 3. Plasmid schematic of the genetic features of a generic J donor vector. Depicted is a circularized plasmid schematic of a J donor vector with minimally required genetic features depicted as labelled boxes. **J segment part**, refers to a DNA sequence encoding a proportion of a TCR joining germline ORF, or mutant/synthetic J gene segment. **C part**, refers to a small 5′ portion of the TCR Constant gene segment. **Type IIS ←**, refers to a Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 5' direction. **Type IIS →**, refers to a Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 3' direction. **+ve selection #2,** refers to the first positive selection marker of the two-component system used to convey a selective advantage to the organism harbouring the vector, and which is different to the positive selection marker of the first vector component (see figure 2). **Ori,** refers

to an origin of replication used for the propagation of plasmid within a compatible host.

Figure 4. Generic description of genetic input, by-products, intermediates and product of the two-component vector system to reconstruct a full-length TCR ORF.

Depicted are the two components of the vector system (a and b), the oligonucleotide duplex (c), when these three components are combined into a single reaction with Type IIS restriction enzyme and ligase, two reaction by-products (d and e), two reaction intermediates (f and g) and one reaction product (h) is generated. Input vectors and product of the two-component system are depicted as circularized plasmid schematics with genetic elements depicted as labelled boxes; open plasmid vectors that represent by-product or intermediate are non-circularized plasmid schematics with genetic elements depicted as labelled boxes; and linear DNA are depicted as series of labelled boxes describing genetic elements.

a) A generic V-C entry plasmid as depicted in figure 2.

b) A generic J donor plasmid as depicted in figure 3.

c) A DNA oligo duplex containing CDR3 sequence flanked by two single stranded DNA overhangs, overhang ‡1-5' and overhang ‡1-3'. Overhang ‡1-5' is compatible with the overhang ‡1-3' in the open V-C entry vector intermediate (g). Overhang ‡2-3' is compatible with the overhang ‡2-5' in the donor fragment intermediate (f).

d) Digestion of the V-C entry vector (a) by the Type IIS restriction enzyme results in a linear DNA V-C entry vector reaction by-product containing the –ve selection element and the **Type IIS ←** and **Type IIS →** elements.

e) Digestion of the J donor vector (b) by the Type IIS restriction enzyme results in a non-circularised plasmid by-product containing all genetic elements of the parental plasmid except those carried in the excised J donor fragment intermediate (f).

f) Digestion of the J donor vector (b) by the Type IIS restriction enzyme results in a linear DNA fragment containing the J segment part and C part flanked by single strand DNA overhangs, overhang ‡2-5' and overhang ‡3-3'. Overhang ‡2-5' is compatible with the overhang ‡2-3' in CDR3 DNA oligonucleotide duplex (c). Overhang ‡3-3' is

compatible with the overhang ‡3-5' in the open V-C entry vector intermediate (g).

g) Digestion of the V-C entry vector (a) by the Type IIS restriction enzyme results in a non-circularized plasmid intermediate containing all genetic elements of the parental plasmid except those carried in the excised linear DNA V-C entry vector reaction by-product (d). Digestion has additionally created two single stranded DNA overhangs, overhang ‡1-3' and overhang ‡3-5'. Overhang ‡1-3' is compatible with the overhang ‡1-5' in the CDR3 DNA oligonucleotide duplex (c). Overhang ‡3-5' is compatible with the overhang ‡3-3' in the J donor fragment intermediate (f).

h) Ligation of all three compatible single stranded DNA overhangs results in the full-length TCR ORF vector as circularized plasmid. This plasmid contains all genetic elements of the parental V-C entry vector (a) with the exception of the excised V-C entry vector reaction by-product (d). In addition, the full-length TCR ORF vector incorporates the CDR3 sequence from the CDR3 DNA oligonucleotide duplex (c) and J segment part and C part from the J donor fragment reaction intermediate. Arrows indicate the approximate points of ligation between compatible single stranded DNA overhangs ‡1, ‡2 and ‡3. Ligation point ‡1 is comprised of the ‡1-3' and ‡1-5' elements donated by the V-C entry vector reaction intermediate (g) and CDR3 DNA oligonucleotide duplex (c), respectively. Ligation point ‡2 is comprised ‡2-3' and ‡2-5' elements donated by the CDR3 DNA oligonucleotide duplex (c) and the J donor fragment reaction intermediate (f), respectively. Ligation point ‡3 is comprised ‡3-3' and ‡3-5' elements donated by the J donor fragment reaction intermediate (f) and the V-C entry vector reaction intermediate (g), respectively.

Figure 5 Arrangement of V cloning fragments for the construction of V-C entry vectors

Depicted is a representation of the V cloning fragments used to assemble V-C entry vectors of a TORES for human TRA and TRB TCR chains as described in Examples 1 to 4.

The V cloning fragment is flanked by unique primer bind sequences at 5' and 3' end to facilitate PCR-mediated amplification of the cloning fragments. BbsI sites represent a specific Type IIS restriction enzyme binding sites used in the assembly of the V-C entry vector, where ➔ indicates that the recognition site is orientated to cut in the 3' direction of the site, and ← indicates that the site is orientated to cut in the 5' direction. The BbsI

→ site cuts 5' of the encoded Kozak sequence to create overhang*1. The BsaI ← site cuts to create the 5' NotI overhang within the NotI 5' fragment. Overhang*1 and the 5' NotI overhang ultimately ligate with overhang*1' of digested V-C entry vector backbone, and the 3' NotI overhang of the digested C cloning fragment, respectively, in assembly of the V-C entry vector. The NotI 5' fragment represents a 6 nucleotide 5' fragment of the NotI recognition sequence, wherein NotI acts as the negative selection marker to eliminate parental V-C entry vector in operation of the TORES. The complete NotI recognition site is reconstituted with the 3' NotI fragment, provided by the C cloning fragment. The V-segment represents the TCR V gene segment that is to be encoded by the final V-C entry vector and encodes from the ATG start codon of the give V segment to the last Cys codon of the V segment that defines the border of the CDR3 region. The BsaI ← site is the Type IIS restriction enzyme recognition sequence used during operation of the TORES system to reconstitute a full-length TCR ORF. Action of the BsaI enzyme, wherein the site is orientated to cut in the 5' direction, results in the creation of overhang **‡1** at the 3' end of the V segment that encompasses the three nucleotides of the last Cys codon of each V segment, and the third nucleotide of the codon preceding that Cys codon. This overhang is standardized among all V segments in a given TORES set. Ultimately, the overhang **‡1** at the 3' of the V segment ligates with overhang **‡1** at the 5' odeCDR3 in operation of the TORES system to reconstitution of a full-length TCR ORF. All **sp** denote the addition of one or more nucleotides to create the correct spacing between the Type IIS recognition sequences and the target overhang sequences, or to space the NotI recognition and cut site for efficient action.

Figure 6 Arrangement of C cloning fragments for the construction of V-C entry vectors

Depicted is a representation of the C cloning fragments used to assemble V-C entry vectors of a TORES for human TRA and TRB TCR chains as described in Examples 1 to 4.

The C cloning fragment is flanked by unique primer bind sequences at 5' and 3' end to facilitate PCR-mediated amplification of the cloning fragments. BbsI sites represent a specific Type IIS restriction enzyme binding sites used in the assembly of the V-C entry vector, where → indicates that the recognition site is orientated to cut in the 3' direction of the site, and ← indicates that the site is orientated to cut in the 5' direction. The BbsI → site cuts to create the 3' NotI overhang within the NotI 3' fragment. The BsaI ← site cuts 3' of the stop codon of the C segment to create **overhang*2** at the 3' end of the C

segment. **Overhang*2** and the 3' NotI overhang ultimately ligate with **Overhang*2'** of the digested V-C entry vector backbone, and the 5' NotI overhang of the digested V cloning fragment, respectively, in assembly of the V-C entry vector. The NotI 3' fragment represents a 6 nucleotide 3' fragment of the NotI recognition sequence, wherein NotI acts as the negative selection marker to eliminate parental V-C entry vector in operation of the TORES. The complete NotI recognition site is reconstituted with the 5' NotI fragment, provided by the V cloning fragment.

The C-segment represents the TCR C gene segment that is to be encoded by the final V-C entry vector and encodes from the cytosine residue 5' of the first Glu codon of the C gene segment to the stop codon. The BsaI ➔ site is the Type IIS restriction enzyme recognition sequence used during operation of the TORES system to reconstitute a full-length TCR ORF. Action of the BsaI enzyme, wherein the site is orientated to cut in the 3' direction, results in the creation of overhang **‡3** at the 5' end of the C segment. This overhang is standardized among all C segments in a given TORES set. Ultimately, the overhang **‡3** at the 5' of the C segment ligates with overhang **‡3** at the 3' C part of the J donor vector in operation of the TORES system to reconstitution of a full-length TCR ORF. All **sp** denote the addition of one or more nucleotides to create the correct spacing between the Type IIS recognition sequences and the target overhang sequences, or to space the NotI recognition and cut site for efficient action.

Figure 7 Arrangement of V-C entry vector backbone for the construction of V-C entry vectors

Depicted is a representation of the V-C entry vector backbone used to assemble V-C entry vectors of a TORES for human TRA and TRB TCR chains as described in Examples 1 to 4.

The circular plasmid DNA contains an origin of replication (Ori) and a positive selection marker #1. This selection marker is used for selection of transformed hosts when isolating clones of V-C entry vector backbone and V-C entry vectors during the assembly, and also for the selection of vectors containing full-length TCR ORFs during operation of the TORES. 5' and 3' genetic elements encode the target elements that flank the final TCR ORF after generation of full-length TCR ORF after its generation by TORES operation. A 5' genetic element might represent a mammalian promoter element to drive the expression of TCR transcripts, and a 3' genetic element might represent a transcriptional terminator sequence. The ACC65I site represents a restriction enzyme

recognition sequence, wherein action of the Acc65I enzyme results in the creation of Overhang*1'. This Overhang*1' ligates with Overhang*1 in the digested V cloning fragment during assembly of the V-C entry vector. The XbaI site represents a restriction enzyme recognition sequence, wherein action of the XbaI enzyme results in the creation of Overhang*2'. This Overhang*2' ligates with Overhang*2 in the digested C cloning fragment during assembly of the V-C entry vector. Sp denotes the addition of nucleotides to space the Acc65I and XbaI recognition sites for efficient action of both enzymes.

Figure 8 Arrangement of the J receiving cassette fragment

Depicted is a representation of a J receiving cassette fragment used in the assembly of J donor vectors of a TORES for human TRA and TRB TCR chains as described in Examples 2 and 3. A J receiving cassette fragment is inserted into a J donor backbone to generate a J receiving cassette vector.

A J receiving cassette fragment is generated by annealing two complimentary oligonucleotides to create a linear double stranded DNA construct with 4-nucleotide single stranded overhangs at the 5' and 3' ends that are used for insertion of the fragment to the J donor vector backbone. Overhang*3 at the 5' end of the J receiving cassette fragment ligates with Overhang*3' of the digested J donor vector backbone, whereas Overhang*4 at the 3' end ligates with Overhang*4' of the digested J donor vector backbone.

The BsaI sites represent the Type IIS restriction recognition sites used in the operation of the TORES to assemble a full-length TCR ORF. BsaI ← site is orientated to cut in the 5' direction, and acts upon the C part sequence to generate Overhang **‡3** at the 3' C part. BsaI → site ultimate acts on the J segment part of the J donor vector to create Overhang **‡2** at the 5'end of the J segment part. BsaI → element also contains Overhang*5, which is generated by action of the BbsI on the BbsI ← site during assembly of the J donor vector.

The BbsI sites represent the Type IIS restriction recognition sites used to assemble the J donor vector. The BbsI ← site cuts the BsaI → element to generate Overhang*5, whereas the BbsI → site cuts the 5' end of the C part to generate Overhang*6. Overhang*5 and Overhang*6 ultimately ligate with Overhang*5' and Overhang*6' of the J segment part, respectively.

The C part represents a small portion of the target C gene segment to permit standardized generation of non-palindromic overhangs during operation of the TORES. This C part is ultimate carried at the 3' end of the J segment part, and forms part of the sequence that ligates with the C segment carried by the digested V-C entry vector in operation of the TORES to generate a full-length TCR ORF.

The NotI site represents a negative selection marker used to eliminate the parental J receiving cassette vector during generation of the J donor vector.

All **sp** denote the addition of one or more nucleotides to create the correct spacing between the Type IIS recognition sequences and the target overhang sequences, or to space the NotI recognition and cut site for efficient action.

Figure 9 Arrangement of the J donor backbone

Depicted is a representation of J a donor vector backbone used in the assembly of J donor vectors of a TORES for human TRA and TRB TCR chains as described in Examples 2 and 3. A J receiving cassette fragment is inserted into a J donor backbone to generate a J receiving cassette vector.

The circular plasmid DNA contains an origin of replication (Ori) and a positive selection marker #2. This selection marker is used for selection of transformed hosts when isolating clones of J donor vector backbone and J donor vectors during the assembly. Importantly, this positive selection marker is distinct from positive selection marker #1 within the V-C entry vectors, such that parental J donor vectors are eliminated under positive selection on #1 during operation of the TORES to generate full-length TCR ORFs in the context of the V-C entry vector backbone.

The EcoRI site represents a restriction enzyme recognition sequence, wherein action of the EcoRI enzyme results in the creation of Overhang*3'. This Overhang*3' ligates with Overhang*3 in the annealed J receiving cassette fragment during assembly of the J receiving cassette vector. The XbaI site represents a restriction enzyme recognition sequence, wherein action of the XbaI enzyme results in the creation of Overhang*4'. This Overhang*4' ligates with Overhang*4 in the annealed J receiving cassette fragment during assembly of the J receiving cassette vector. Sp denotes the addition of nucleotides to space the Acc65I and XbaI recognition sites for efficient action of both

enzymes.

Figure 10 Arrangement of the J receiving cassette vector

Depicted is a representation of a J donor vector backbone used in the assembly of J donor vectors of a TORES for human TRA and TRB TCR chains as described in Examples 2 and 3. A J receiving cassette vector is created by insertion of a J receiving cassette fragment into a J donor backbone.

The circular plasmid DNA contains an origin of replication (Ori) and a positive selection marker #2. This selection marker is used for selection of transformed hosts when isolating clones of J donor vector backbone and J donor vectors during the assembly. Importantly, this positive selection marker is distinct from positive selection marker #1 within the V-C entry vectors, such that parental J donor vectors are eliminated under positive selection on #1 during operation of the TORES to generate full-length TCR ORFs in the context of the V-C entry vector backbone.

The BsaI sites represent the Type IIS restriction recognition sites used in the operation of the TORES to assemble a full-length TCR ORF. BsaI ← site is orientated to cut in the 5' direction, and acts upon the C part sequence to generate Overhang **‡3** at the 3' C part. BsaI → site ultimate acts on the J segment part of the J donor vector to create Overhang **‡2** at the 5'end of the J segment part. BsaI → element also contains Overhang*5, which is generated by action of the BbsI on the BbsI ← site during assembly of the J donor vector.

The BbsI sites represent the Type IIS restriction recognition sites used to assemble the J donor vector. The BbsI ← site cuts the BsaI → element to generate Overhang*5, whereas the BbsI → site cuts the 5' end of the C part to generate Overhang*6. Overhang*5 and Overhang*6 ultimately ligate with Overhang*5' and Overhang*6' of the J segment part, respectively.

The C part represents a small portion of the target C gene segment to permit standardized generation of non-palindromic overhangs during operation of the TORES. This C part is ultimate carried at the 3' end of the J segment part, and forms part of the sequence that ligates with the C segment carried by the digested V-C entry vector in operation of the TORES to generate a full-length TCR ORF.

The NotI site represents a negative selection marker used to eliminate the parental J receiving cassette vector during generation of the J donor vector.

All **sp** denote the addition of one or more nucleotides to create the correct spacing between the Type IIS recognition sequences and the target overhang sequences, or to space the NotI recognition and cut site for efficient action.

Figure 11 Arrangement of a J segment part

Depicted is a representation of a J segment part that is used in the assembly of J donor vectors of a TORES for human TRA and TRB TCR chains as described in Examples 2 and 3. A J segment part is inserted into a J receiving cassette vector to create a J donor vector.

Annealing complimentary single stranded oligonucleotides to form a linear double stranded DNA construct with single stranded overhangs at either terminus generates a J segment part. Overhang*5' at the 5' terminus anneals with Overhang*5 generated within the J receiving cassette vector digested with BbsI. Overhang*6' at the 3' terminus anneals with Overhang*6 generated within the J receiving cassette vector digested with BbsI.

The J segment part represents the target J gene segment sequence. Depending on the style of the J donor vector being constructed (i.e. short or long) the 5' border of the J segment part is defined differently. For short J donor vectors, the 5' border of the J segment part is defined as the Phe-Ala/Gly or Trp-Gly motifs that are used to define the canonical border between the J and CDR3 portions of a full-length TCR ORF. For long J donor vectors, the 5' border of the J segment part is extended ten to twelve nucleotides 5' of the Phe-Ala/Gly or Trp-Gly motif. This extends the portion of the overall TCR ORF encoded by the J donor vector, and conversely shortens the length of the odeCDR3 required to construct a full-length TCR ORF in operation of the TORES. At the 3' end of the J segment part is encoded a single Adenine residue (A), which represent the first nucleotide of the C fragment. This adenine is excluded from the J receiving cassette vector.

Figure 12 Validation of specificity of reconstituted model TCR TRA/TRB pair

A model TRA/TRB TCR chain pair with specificity for a HLA-A2*01- restricted HCMV pp65 derived antigen was reconstituted with the TORES system as described in exam-

ple 7. The two TRA and TRB plasmid products were transiently transfected into a human cell line constitutively expressing CD3 components, but lacking endogenous expression of TCR TRA and TRB chains. 48 hours after the transfection, cells were with antibodies against CD3, TCRalpha/beta and specific HLA-A2*01- NLVPMVATV tetramer and analysed by flow cytometry. The model TCR pair presented on the cell surface as indicated by positive staining with CD3 and TCRalpha/beta antibodies (left, top panel). The model TCR pair also displayed positive staining for the HLA-A2*01–NLVPMVATV tetramer reagent (left, bottom panel), indicating expected specificity of binding. An irrelevant pair of TRA and TRB plasmid constructs was also transfected in parallel (centre panels), as was empty vector (rightmost panels). The irrelevant TCR was presented on the cell surface as indicated by positive staining for CD3 and TCRalpha/beta antibodies (centre, top panel), but displayed no binding for the HLA-A2*01 - NLVPMVATV tetramer reagent (centre, bottom panel). Empty vector transfected cells displayed no staining for either CD3 or TCRalpha/beta antibodies, or for the HLA-A2*01- NLVPMVATV tetramer reagent.

Figure 13 Workflow for the identification and reconstitution of a set of TCR TRA/TRB chain pairs from human peripheral blood

To demonstrate the use of the TORES system for the reconstitution of native TCR TRA/TRB chain pairs captured from a human specimen, a set of TCRs specific for the HLA-B*07:02 restricted HCMV pp65 antigen, TPRVTGGGAM, was generated via capture of sequence information from single-cell sorting of tetramer-positive cells from a human PBMC fraction. This process is described in Example 8.

i) A PBMC fraction from a donor with HLA-B*07:02 allele is stained with a HLA-B*07:02-TPRV tetramer reagent. Single $CD8^+$ T-cells showing positive staining for tetramer a deposited into single PCR tubes by way of FACS.

ii) Single cell containing tubes were subjected to a reverse transcription reaction with primer sets anchored in all expressed variable and constant gene segments for both TRA and TRB chains.

iii) Nested PCR reactions are performed on the reverse transcription product with sets of internal primers, still anchored in all expressed V and C gene segments, independently for the TRA and TRB chains.

iv) The nested PCR product is subjected to Sanger sequencing

v) PCR product sequences are aligned against a library of human germline TCR gene segments to determine the V and J (and C) gene segment usage for each amplified TCR chain. The CDR3 sequence from the 5' Cys codon to the 3' Phe-Ala/Gly or Trp-Gly motifs are determined.

vi) A V-C entry vector is selected from the TORES library corresponding to the determined V and C gene segment usage of each TRA and TRB chain to be reconstituted.

vii) A J donor vector is selected from the TORES library corresponding to the determined J gene segment usage of each TRA and TRB chain to be reconstituted.

viii) An odeCDR3 is synthesized corresponding the CDR3 region determined from each TRA and TRB chain to be reconstituted.

ix) An independent restriction enzyme (RE) / Ligase cycle reaction is assembled for each TRA and TRB TCR to be reconstituted, containing the selected V-C entry vector, J donor vector and synthesized odeCDR3.

x) The RE/Ligase cycle reaction product is transformed into bacteria and placed under antibiotic selection for the first positive selection marker (i.e. V-C entry vector backbone).

xi) Clones are propagated and confirmed by Sanger sequencing of the final full-length TCR ORF.

Figure 14 Validation of specificity for a set of TCR TRA/TRB chain pairs captured from human peripheral blood and reconstituted with TORES

A set of six TCR TRA/TRB chain pairs were captured from human peripheral blood and reconstituted via the workflow outlined in Figure 13 and described in Example 8. The six TCR pairs had expected specificity for the B*07:02-TPRV tetramer reagent used to isolate single $CD8^+$ T-cells. Each of the six captured pairs, an irrelevant TCR chain pair, and an empty vector control, were transfected into human cells expressing CD3 components, but lacking endogenous expression of TCR TRA and TRB chains. 48 hours after the transfection, flow cytometric analysis was used to stain the cells with

antibodies against CD3, TCR alpha and beta chains and the specificity of the reconstituted and expressed TCR was confirmed by specific HLA-B*07:02-TPRV tetramer reagent. The data are presented as contour plots. Each panel displays CD3 signal on the X-axis and HLA-B*07:02-TPRV tetramer on the Y-axis. Inset at the bottom right of each panel is the ratio of HLA-B*07:02-TPRV tetramer signal of the CD3 positive cells over the CD3 negative cells as an arbitrary unit of the degree of tetramer staining. All six TCR chain pairs presented display some degree of specific tetramer staining compared to the irrelevant TCR pair control.

Figure 15 Operation of the TORES to generate CDR3-diversified TCR chains

Depicted is a schematic representation of the TORES when used to generate full-length TCR chains with diversified CDR3 inserts. A parental TCR is defined with V-J-C usage, and defined CDR3 region sequence. The corresponding single V-C entry vector (box i) and single J donor vector (box ii) are placed in the reaction tube. A pool of odeCDR3 with defined positional nucleotide degeneracy and/or point mutagenesis that changes the coded amino acid sequence is synthesized (Box iii). Such a CDR3 pool could include completely randomized CDR3 sequences within the bounds of the defined odeCDR3 framework, as to create a 'synthetic' full-length TCR ORFs CDR3 with germline V-J-C usage. These three components (Box i, ii and iii) are combined into a single reaction along with appropriate restriction and ligase enzymes. The reaction cycle produces a number of variant reconstituted full-length TCR ORFs, proportional to the number of variant odeCDR3 included, in a single step in the V-C entry vector backbone context (Box iv).

Figure 16 A Single round of TCR chain diversification using limited CDR3 degeneracy generates a TCR set with a large spectrum of target affinities

The TRA chain of a model TRA/TRB TCR chain pair with specificity for a HCMV pp65 derived antigen as described in example 7, was subject to a CDR3-diversification cycle according to the scheme presented in Figure 15 and as described in Example 9. The TRA chain odeCDR3 was synthesized with 4-fold nucleotide degeneracy at 3 separate positions, altering the codon usage at those three positions. The resulting reaction product contained 64 full-length TRA chain variants, including the parental sequence. Each of the 64 TRA chains was cloned and sequence confirmed.

A) Each of the 64 TRA chain plasmids was transfected along with the parental TRB plasmid into a human cell line constitutively expressing CD3 components, but lacking

endogenous expression of TCR TRA and TRB chains. 48 hours after transfection, the cells were stained with antibodies against CD3 and with a HLA-A2*01 - NLVPMVATV tetramer reagent and analysed by flow cytometry. The ratio of the mean fluorescence intensity (MFI) of HLA-A2*01 - NLVPMVATV tetramer signal for the CD3+ population over the CD3- population was plotted, as an indication of the binding strength of each TCR chain pair variant. The arrow indicates the data point corresponding to the parental TRA. Ovals indicate high-binder and non-binder variants at either end of the spectrum.

B) Each TRA chain variant is presented in a table with the MFI ratio of HLA-A2*01 - NLVPMVATV tetramer signal for the CD3+ population over the CD3- population, and the amino acid present at each of the three diversified positions (Pos1, 2 and 3), in one letter code. The arrow indicates the data point corresponding to the parental TRA. Brackets indicate high-binder and non-binder variants at either end of the spectrum.

Figure 17 Operation of the TORES to generate V-segment diversified TCR chains
Depicted is a schematic representation of the TORES when used to generate full-length TCR chains with diversified V-segment usage. A parental TCR is defined with V-J-C usage, and defined CDR3 region sequence. The corresponding single J donor vector (box ii) is placed in the reaction tube, as is the single odeCDR3 synthesized to correspond with parental CDR3 region sequence (Box iii). A selection of V-C entry vectors is also added to the reaction tube, corresponding to the V- and C- segments desired in the product V-segment diversified full length TCR ORF product (Box i). These three components (Box i, ii and iii) are combined into a single reaction along with appropriate restriction and ligase enzymes. The reaction cycle produces a number of variant reconstituted full-length TCR ORFs, proportional to the number of variant V-C entry vectors included, in a single step in the V-C entry vector backbone context (Box iv).

Figure 18 Operation of the TORES to generate J-segment diversified TCR chains
Depicted is a schematic representation of the TORES when used to generate full-length TCR chains with diversified J-segment usage. A parental TCR is defined with V-J-C usage, and defined CDR3 region sequence. The corresponding single V-C entry vector (box i) is placed in the reaction tube, as is the single odeCDR3 synthesized to correspond with parental CDR3 region sequence (Box iii). A selection of J donor is also added to the reaction tube, corresponding to the J segments desired in the product J-segment diversified full length TCR ORF product (Box ii). These three components

(Box i, ii and iii) are combined into a single reaction along with appropriate restriction and ligase enzymes. The reaction cycle produces a number of variant reconstituted full-length TCR ORFs, proportional to the number of variant J donor vectors included, in a single step in the V-C entry vector backbone context (Box iv).

Figure 19 Operation of the TORES to generate V/J-segment diversified TCR chains

Depicted is a schematic representation of the TORES when used to generate full-length TCR chains with diversified V- and J- segment usage. A parental TCR is defined with V-J-C usage, and defined CDR3 region sequence. The corresponding single odeCDR3 synthesized to correspond with parental CDR3 region sequence (Box iii). A selection of V-C entry vectors and J donor vectors are added to the reaction tube, corresponding to the combination of V- (C-) and J- segments desired in the product V/J-segment diversified full length TCR ORF product (Box ii). These three components (Box i, ii and iii) are combined into a single reaction along with appropriate restriction and ligase enzymes. The reaction cycle produces a number of variant reconstituted full-length TCR ORFs, proportional to the number of V-C and J donor vector combinations possible from those included, in a single step in the V-C entry vector backbone context (Box iv).

Figure 20. Schematic of the genetic features of TORES^2 V-C□ and V-C□ entry vectors.

Depicted is a circularized plasmid schematic of a V-C□ (a) and V-C□ (b) entry vectors with minimally required genetic elements depicted as labelled boxes. **Kozak,** refers to consensus sequence that plays a role in the efficient initiation of translation. **V-segment,** refers to a selected sequence encoding a proportion of a TCR variable germline ORF, or mutant/synthetic ORF. **Type IIS #1←**, refers to a first Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 5' direction. **Type IIS #1 →**, refers to a first Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 3' direction. **-ve selection #1,** refers to a negative selection element designed to be detrimental to a plasmid harbouring the sequence during the full-length TCR reconstruction reaction, or subsequent selection steps. **C-segment,** refers to a selected sequence encoding a proportion of a TCR constant germline ORF, or mutant/synthetic ORF. **Type IIS #2 ←**, refers to a second Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 5' direction. The enzyme is different than the first Type IIS. **Type IIS #2 →**, refers to a second Type IIS restriction

enzyme binding site orientated such the enzyme cleaves in the 3' direction. The enzyme is different than the first Type IIS. **-ve selection #2,** refers to a negative selection element designed to be detrimental to a plasmid harbouring the sequence during assembly of the full-length bidirectional TCR ORF vector. **-ve selection #2** is different to the first -ve selection. **+ve selection #1,** refers to the first positive selection marker of the two-component system used to convey a selective advantage to the organism harbouring the vector, and which is different to the positive selection marker a second (J donor) vector component (see figure 3). **Ori,** refers to an origin of replication used for the propagation of plasmid within a compatible host. **5′ genetic element,** refers to any desired genetic element that provides attributes required for downstream application of the reconstituted full-length TCR, and should be situated 5' of the reconstituted full-length TCR, for example, an RMCE element. **3′ genetic element, refers to** any desired genetic element that provides attributes required for downstream application of the reconstructed full-length TCR, and should be situated 3' of the full-length TCR ORF, for example, a second RMCE element.

Figure 21. Plasmid schematic of the genetic features of the bidirectional terminator donor (BiT donor) vector.

Depicted is a circularized plasmid schematic of a bidirectional terminator donor vector with minimally required genetic features depicted as labelled boxes. **The bidirectional terminator element (T-T)**, refers to a DNA sequence encoding a bidirectional terminator. **Type IIS #2 →**, refers to a second Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 3' direction. **Type IIS #2 ←**, refers to a second Type IIS restriction enzyme binding site orientated such the enzyme cleaves in the 5' direction. **+ve selection #2,** refers to the first positive selection marker of the two-component system used to convey a selective advantage to the organism harbouring the vector, and which is different to the positive selection marker of the first vector component (see figure 20). **Ori,** refers to an origin of replication used for the propagation of plasmid within a compatible host.

Figure 22. Generic description of genetic input, by-products, intermediates and product of the bidirectional two-component vector system to reconstruct a paired TRA and TRB vector.

Depicted are the TRA (a), TRB (b) and the bidirectional terminator donor vector (c), when these three components are combined into a single reaction with second Type IIS restriction enzyme and ligase, three reaction by-products (d, e and f), three reaction

intermediates (g, h and i) and one reaction product (j) are generated. Input vectors and product of the two-component system are depicted as circularized plasmid schematics with genetic elements depicted as labelled boxes; open plasmid vectors that represent by-product or intermediate are non-circularized plasmid schematics with genetic elements depicted as labelled boxes; and linear DNA are depicted as series of labelled boxes describing genetic elements.

a) A reconstituted TRA ORF in V-C□ entry vector context as depicted in Figure 20a.

b) A reconstituted TRA ORF in V-C□ entry vector context as depicted in Figure 20b

c) A BiT donor vector as depicted in Figure 21

d) Digestion of the TRA vector (a) with Type IIS #2 restriction enzyme results in a linear DNA TRA vector reaction by-product containing the –ve selection element #2 and the **Type IIS #2 ←** and **Type IIS #2 →** binding site.

e) Digestion of the TRB vector (b) by the Type IIS #2 restriction enzyme results in a linear DNA TRB vector reaction by-product containing the 5' and 3' genetic elements, the **Type IIS #2 →** and **Type IIS #2 ←** elements, the –ve selection element #2, the +ve selection element #1 and origin of replication element.

f) Digestion of the bidirectional terminator donor vector (c) by the Type IIS #2 restriction enzyme results in a non-circularised plasmid by-product containing all vector elements of the parental plasmid except the excised bidirectional terminator fragment intermediate (i).

g) Digestion of the TRA vector (a) by the Type IIS #2 restriction enzyme results in a non-circularized plasmid intermediate containing all genetic elements of the parental plasmid except those carried in the excised linear DNA TRA vector reaction by-product (d). Digestion has additionally created two single stranded DNA overhangs, overhang ‡1-5' and overhang ‡3-3'. Overhang ‡1-5' is compatible with the overhang ‡1-3' in the bidirectional terminator intermediate (i). Overhang ‡3-3' is compatible with the overhang ‡3-5' in the TRB fragment intermediate (h).

h) Digestion of the TRB vector (b) by the Type IIS #2 restriction enzyme results in a lin-

ear DNA fragment containing the C segment, C part, J segment, CDR3, V□ segment and Kozak sequence flanked by single strand DNA overhangs, overhang ‡2-3' and overhang ‡3-5'. Overhang ‡2-3' is compatible with the overhang ‡2-5' in the bidirectional terminator intermediate (i). Overhang ‡3-5' is compatible with the overhang ‡3-3' in the open TRA vector intermediate (g).

i) Digestion of the bidirectional terminator donor vector (c) by the Type IIS #2 restriction enzyme results in a linear DNA fragment intermediate containing the bidirectional terminator. Digestion has additionally created two single stranded DNA overhangs, overhang ‡1-3' and overhang ‡2-5'. Overhang ‡1-3' is compatible with the overhang ‡1-5' in the open TRA vector intermediate (g). Overhang ‡2-5' is compatible with the overhang ‡2-3' in the TRB fragment intermediate (h).

j) Ligation of all three compatible single stranded DNA overhangs results in the bidirectional paired TRA and TRB vector as circularized plasmid. This plasmid contains all genetic elements of the parental TRA vector (a) with the exception of the excised TRA vector reaction by-product (d). In addition, the full-length TCR ORF vector incorporates the TRB fragment intermediate (h) and the bidirectional terminator intermediate (i). Arrows indicate the approximate points of ligation between compatible single stranded DNA overhangs ‡1, ‡2 and ‡3. Ligation point ‡1 is comprised of the ‡1-5' and ‡1-3' elements donated by the TRA vector reaction intermediate (a) and the bidirectional terminator intermediate (i), respectively. Ligation point ‡2 is comprised ‡2-3' and ‡2-5' elements donated by the TRB fragment intermediate (h) and the bidirectional terminator fragment (i), respectively. Ligation point ‡3 is comprised ‡3-3' and ‡3-5' elements donated by the TRA vector reaction intermediate (a) and the TRB fragment reaction intermediate (h), respectively.

Figure 23 Validation of specificity of a model TCR TRA/TRB pair reconstituted using TORES^2

A model TRA/TRB TCR chain pair with specificity for a HLA-A*02:01-restricted antigen was reconstituted with the TORES^2 system as described in example 11. The bidirectional TRA/TRB donor vector was used to stably integrate the TCR ORFs into a human cell line constitutively expressing CD3 components but lacking endogenous expression of TRA and TRB chains, and further encoding a genomic receiver site compatible with the RMCE sites contained within the product donor vector that have been contributed by the original V-Cα entry vector. The functionality of the reconstituted full length paired

TRA and TRB donor vector generated by the TORES^2 was determined by surface staining of the CD3 complex, and expected specificity of the reconstituted and expressed TCR was confirmed by specific HLA-A*02:01- SLLMWITQV tetramer staining.

a) The model TCR pair was integrated into the genome of the engineered TCR presenting cell line containing genomic receiver sites matched to the donor vector. This genomic receiver site contains promoter elements driving two separate fluorescent markers (red fluorescent protein, RFP, and blue fluorescent protein, BFP) in an anti-parallel arrangement, interposed by a bidirectional terminator element similar to the final TRA/TRB donor vector product of the TRA/TRB TORES^2 described in Example 11. Upon transfection and outgrowth of the receiver cells, flow cytometric analysis was conducted to observe the persistence of these florescent reporters. The absence of the genomic receiver sites (RFP-BFP-) therefore indicated that the cells have successfully integrated the TRA/TRB donor vector. While the cells that did not integrate the TRA/TRB donor vector demonstrated presence of the genomic receiver sites (RFP+BFP+). Majority of the cells have successfully integrated the TRA/TRB donor vector sequences into genomic receiver site.

b) Each of the RFP-BFP- and RFP+BFP+ populations of a) were gated and redisplayed in a second overlay histogram plot. The cells that integrated the bidirectional construct also showed positive staining with CD3 antibody, demonstrating that TCR was expressed on the cell surface (light histogram). The cells that did not integrate the bidirectional construct failed to stain with the CD3 antibody (dark histogram).

c) and d) The RFP-BFP- cells as analysed in a) and b) were gated and redisplayed as contour plots. Parallel samples of the cells had been stained with HLA-A*02:01-SLLMWITQV tetramer reagent c), or not d), wherein tetramer/CD3 double positivity in c) indicates the expected specificity of the reconstituted and integrated TCR ORF pair.

Materials and Methods

DNA Sequencing

All sequencing referred to within the presented examples was conducted by the Sanger method, and conducted by *GATC Biotec AB*, Sweden.

DNA Synthesis

All DNA synthesis referred to within the presented examples was conducted by *Inte-*

grated DNA technologies BVBA, Belgium.

DNA Fragments >125 bp were synthesised as linear double stranded DNA molecules as a *'gBlock Gene Fragments'* product.

DNA Fragments 15-60 nt were synthesised as single stranded DNA molecules as a *'Custom Oligonucleotide Fragment'* product.

DNA Fragments 61-124 nt were synthesised as single stranded DNA molecules as a *'Ultramer DNA oligonucleotide Fragment'* product.

Vector library assembly and cloning

The construction of vectors described in the examples comprises a variety of methods well known to those skilled in the art, and specific reaction compositions are outlined in detail in Examples 1 to 3. The following key materials were used in the described procedures:

Product	**Supplier**	**Supplier Number**
Acc651	New England BioLabs	R0599L
BbsI HF	New England BioLabs	R3539L
DH5alpha competent cells	Thermo Fisher Scientific	18265017
DNA clean and concentrator kit	Zymo Research	D4030
EcoR1	New England BioLabs	R3101S
NotI	New England BioLabs	R3189L
QIAamp DNA Mini kit	Qiagen	51306
QIAquick Gel Extraction kit	Qiagen	28704
Qiagen Plasmid Plus Midi kit	Qiagen	12945
T4 ligase	New England BioLabs	M0202L
T4 ligase buffer 10x	New England BioLabs	B0202S
XbaI	New England BioLabs	R0145S
XhoI	New England BioLabs	R0146S

Oligonucleotide duplex encoding CDR3 (odeCDR3) assembly

odeCDR3 were routinely assembled by annealing partially complementary single

stranded oligonucleotides. A detailed description of reaction compision and conditions is provide in Example 6. The following key materials were used in the described procedures:

Product	Supplier	Supplier Number
T4 ligase buffer 10 x	New England BioLabs	B0202S
T4 PNK	New England BioLabs	M0201L

TCR reconstitution

Operation of a TORES to reconstitute full-length TCR ORFs is described in detail in Examples 7 to 9. The following key materials were used in the described procedures:

Product	Supplier	Supplier Number
BsaI-HF	New England BioLabs	R3535L
CutSmart buffer 10 x	New England BioLabs	B7204S
DH5alpha competent cells	Thermo Fisher Scientific	18265017
NotI-HF	New England BioLabs	R3189L
QIAamp DNA Mini kit	Qiagen	51306
T4 Ligase	New England BioLabs	M0202L
T4 Ligase buffer 10 x	New England BioLabs	B0202S

Fluorescence activated cell sorting (FACS)

Single HLA-multimer positive $CD8^+$ T-cells were sorted by FACS for amplification and sequencing of TCR chains. This was achieved through standard cell sorting methodologies using a *BDInflux* instrument. Briefly, cells were stained with HLA-multimer reagent on ice for 10 mins, then with CD3 and CD8 antibodies as markers of $CD8^+$ T-cells. Cells with CD3+CD8+Multimer+ signal were sorted to PCR plate pre-loaded with 5 µL of nuclease-free water. Specimens were snap-frozen until subsequent processing. The following key materials were used in the described procedures:

Product	Supplier	Supplier Number
CMV-B.07:02-TPRVT-pp65 (PE)	Immudex	WH2136

Anti-CD8 (clone RPA-T8) (BV510)	BD	563256
Anti-CD3 (clone SK7) (APC-H7)	BD	560176

Sequencing of TCR alpha and beta chains from single T-cells

Singly FACS-sorted T-cells were subjected to a two-step amplification process that entails a V-region specific primer collection for each TRA and TRB, followed by paired nested PCR reactions that create TRA and TRB amplicons for sequence analysis as described in Example 8. This procedure is described previously (Han et. al. Nat Biotechnol. 2014 32(7): 684-692). The following materials were used in the described procedures:

Product	**Supplier**	**Supplier Number**
2x Reaction Mix	Thermo Scientific	12574035
5X Phusion HF Buffer	Thermo Fisher Scientific	F-549S
dNTPs	Thermo Fisher Scientific	10297018
Nuclease free water	Qiagen	129114
Phusion Hot Start II DNA Polymerase	Thermo Fisher Scientific	F-549S
SuperScript® III One- Step RT-PCR System with Platinum® Taq High Fidelity DNA Polymerase	Thermo Scientific	12574035

Example 1

Design and assembly of TRA V-C entry vector library for native human TRA repertoire

A TORES consists of a V-C entry vector library and J donor vector library for a given TCR chain. When combined with a target odeCDR3 sequence to be inserted into a selected V-J-C context, a full-length TCR ORF can be reconstituted. Through varying odeCDR3 sequence features and/or V/J/C selection, this reconstitution step may also represent a sequence diversification step in TCR ORF engineering workflows. In the

present example, the design and assembly of a TRA V-C entry vector library that contains the native human TRA V-C sequence repertoire. This is an example of the generic V-C entry vector type depicted in Figure 2.

The DNA components required for a TRA V-C vector library are:

I. A TRA V cloning fragment for each functional TRA V gene segment encoded in the human genome
II. Single TRA C cloning fragment
III. A V-C entry vector backbone

In the present example, the TRA V and TRA C cloning fragments were synthesized and used to assemble into a target V-C entry vector backbone in a single restriction enzyme and ligase reaction. In the present example, the target V-C entry backbone was designed to permit transient expression of reconstituted TRA ORFs within mammalian cells.

In the present example, the TRA V-C entry vector library is constructed using Type IIS restriction enzyme BbsI. The Type IIS restriction enzyme used in functioning of the complete TORES to reconstitute full-length TRA ORFs is BsaI.

Design of synthetic TRA V cloning fragments

The arrangement of genetic elements of the TRA V cloning fragments in the present example is depicted in Figure 5.

Each end of the TRA V cloning fragment encodes a standardized 5' and 3' **primer bind** DNA sequence of 20 nucleotides for propagation of the overall fragment by PCR.

Proximal to the **5′ primer bind** a BbsI Type IIS restriction enzyme binding site is encoded, wherein the direction of the BbsI binding site guides the BbsI enzyme to cut the DNA 3′ to its recognition sequence. Overhangs generated by BbsI enzymatic activity are encoded by **Overhang *1.** This overhang is designed to permit directed ligase-dependent cloning with an arm of the V-C entry vector backbone.

A consensus **kozak** sequence is encoded 5′ of the ATG start codon within the TRA V gene segment for efficient initiation of translation of the final reconstituted and ex-

pressed TRA mRNA. In the present example, each **TRA V segment** encodes all amino acids from the start methionine residue until its last cysteine (Cys) of the TRA V segment. This Cys residue is generally recognised as a border of the TRA variable gene segments, the deletion of which is rare in naturally occurring recombined and functional TRA chains. Where necessary, native human TRA V consensus sequences have been edited to remove recognition sequences for any restriction enzymes used within assembly or reconstitution operations with the TORES, and also any enzymes used in downstream applications.

To the 3' end of the **TRA V segment** a BsaI Type IIS restriction enzyme binding site is encoded, **BsaI ←**. The direction of the BsaI binding site guides the BsaI enzyme to cut the DNA 5′ to its recognition sequence. The resulting overhang sequence is designed to encompass the last cysteine codon of the V segment element and the 3rd nucleotide for amino acid codon preceding the cysteine. Thus the action of BsaI on the designed sequence creates a **TRA V Cys-overhang ‡1** at the 3' end of the **TRA V segment**. In the present example, this **Cys-overhang ‡1** is standardized among all included TRA V segments to simplify and unify the cloning strategy. Where necessary the nucleotides encoding the TRA V genetic element were changed to encode this standardised overhang but not change the translated amino acid sequence. This **BsaI ←** site is utilized during the full length TRA reconstitution reaction.

In this present example, the V-C entry vector **negative selection marker** is a NotI restriction enzyme binding site. To construct a NotI binding site, two halves of the site are combined when the **TRA V cloning fragment** and **TRA C cloning fragment** are ligated together. The TRA V cloning fragment encodes the **NotI 5′** segment of six nucleotides.

To the 5' end of the **3′ primer bind** sequence encodes a second BbsI restriction site, that directs BbsI enzyme to cut the DNA 5′ to its recognition sequence, **BbsI ←**. The action of BbsI on the designed sequence thus creates an overhang of 4 nucleotides, **NotI 5′ overhang**, which is designed to be complementary to the overhang generated on the TRA C DNA fragment and reconstitute a NotI binding site upon ligation.

Sp denote nucleotide additions to specific points of the TRA V cloning fragment to achieve the correct spacing of Type IIS restriction enzyme binding site and the cut site,

when adjacent to such sites. **Sp** blocks flanking the NotI restriction enzyme binding site sequence have been used to space the NotI binding and cut site appropriately for efficient action. The selection of nucleotides considered the potential impact of DAM methylation of the BsaI binding site.

Full DNA sequences for the **TRA V cloning fragments** in the present example of native human TRA chains are provided as SEQ0001 to SEQ0046. These sequences includes the 5' primer bind and 3' primer bind sequences.

Design of synthetic TRA C cloning fragment

The arrangement of genetic elements of the TRA C cloning fragments in the present example is depicted in Figure 6.

Each end of the TRA C cloning fragment encodes a standardized 5' and 3' **primer bind** DNA sequence of 20 nucleotides for propagation of the overall fragment by PCR.

Proximal to the **5′ primer bind** sequence a BbsI restriction enzyme recognition site is encoded, such that BbsI enzyme will cut the DNA 3′ to its recognition sequence, **BbsI ➔**.

The TRA C cloning fragment encodes the **NotI 3′** segment of six nucleotides, which completes a NotI recognition site that will make up the V-C entry vector **negative selection marker.** The adjacent **BbsI ➔** restriction site acts upon the **NotI 3′** element to create the **NotI 3′ overhang** of four necleotides. This overhang is designed to be complementary to the NotI 5' overhang generated on the TRA V DNA fragment and reconstitute a full NotI binding site upon assembly of V-C entry vectors.

To the 3' end of the **NotI 3′** element, the TRA C cloning fragment encodes a BsaI restriction enzyme binding site, **BsaI ➔**. The direction of the BsaI binding site guides the BsaI enzyme to cut the DNA 5′ to its recognition sequence. The resulting overhang sequence is designed to start from the first cytosine of the TRA C genetic fragment, **TRA C overhang ‡3**. This **BsaI ➔** site is utilized during the full length TRA reconstitution reaction. The **BsaI ➔** enzyme acts upon the **TRA C segment** encoded in the V-C entry vector to create the necessary **TRA C overhang ‡3** during reconstitution reactions. A consensus TRA C sequence from the cytosine residue 5′ of the first glutamine codon

until the stop codon is included in the TRA C cloning fragment in the present example

To the 5' of the **3' primer bind** encodes a BbsI restriction enzyme recognition sequence, **BbsI ←**. The direction of the BbsI binding site guides the BbsI enzyme to cut the DNA 5' to its recognition sequence. Overhangs generated by BbsI enzymatic activity are encoded by **Overhang *2**. The design of this overhang permits directed ligase-dependent cloning with an arm of the V-C entry vector backbone during assembly.

Sp denote nucleotide additions to specific points of the TRA C cloning fragment to achieve the correct spacing of Type IIS restriction enzyme binding site and the cut site, when adjacent to such sites. **Sp** blocks flanking the NotI restriction enzyme binding site sequence have been used to space the NotI binding and cut site appropriately for efficient action. The selection of nucleotides considered the potential impact of DAM methylation of the BsaI binding site

The full DNA sequence for the **TRA C cloning fragment** in the present example of native human TRA chains are presented as, SEQ0047. This sequence includes the 5' primer bind and 3' primer bind sequences.

Design of V-C entry vector backbone for transient expression of reconstituted TRA ORF in mammalian cells

In the present example, the V-C entry vector backbone is derived from the pMA plasmid. It encodes a Col E1 origin of replication, **ori,** along with antibiotic resistance beta-lactamase gene, **positive selection #1.** Beta-lactamase confers resistance to the penicillin group of beta-lactam antibiotics such as ampicillin and carbenicillin.

The vector backbone, as depicted in Figure 7, encodes the required genetic elements that confer the appropriate functionality for downstream applications of the fully reconstituted TRA ORF. In this present example, the **5' genetic element** encodes the CMV constitutive mammalian promoter and the **3' genetic element** encodes the SV40pA polyadenylation signal to permit transient expression of the fully reconstituted TRA ORF in a mammalian cell.

In the present example, the vector backbone encodes **Acc65I** and **XbaI** restriction enzyme binding sites that generate **overhang *1'** and **overhang *2'**, respectively. **Overhang *1'** is complementary to **overhang *1** within the TRA V cloning fragment (figure

5). **Overhang *2'** is complementary to **overhang *2** within the TRA C cloning fragment (figure 6). These complementary overhangs permit directed cloning of the TRA V and TRA C cloning fragments into the V-C entry vector backbone.

Sp feature denotes nucleotides added between the **Acc65I** and **XbaI** restriction enzyme recognition sites required for distancing the two sites for efficient action.

The sequence of the vector backbone from the **5′ genetic element** encoding the CMV constitutive promoter, to the **3′ genetic element** encoding the SV40pA polyadenylation signal is presented as SEQ0048.

Method to assemble TRA V-C entry vector library

This method utilizes standard molecular biology techniques to assemble selected **TRA V cloning fragment** (Figure 5) and **TRA C cloning fragment** (Figure 6) into a given **V-C entry vector backbone** (Figure 7) to create a **TRA V-C entry vector** (Figure 2). In this present example, the method performs the restriction enzyme digestion and ligation reaction in a single reaction.

RE digestion and ligation reaction

100 ng of linear vector backbone (linearised by ACC65I and XbaI digestion)
10 ng of TRA V genetic fragment
20 ng of TRA C genetic fragment
2 µl 10x NEB ligase buffer
0.5 µl of BbsI
1 µl of T4 DNA ligase
Up to 20 µl of H_2O

Reaction conditions

Step 1; 2 min at 37˚C
Step 2; 3 min at 16˚C
Repeat step 1 and 2, 20 times
5 min at 50˚C
5 min at 80˚C
Return to room temperature

Resulting product is transformed into competent E.coli cells that are selected for car-

benicillin-resistant colonies. Plasmids isolated from selected colonies are sequenced to determine correctly assembled constructs. The procedure is repeated for each independent V segment cloning fragment. The resulting constructs make up the TRA V-C entry vector library for use in reconstitution of full-length TRA ORFs for later use in transient expression of said reconstituted TRA in mammalian cells. The sequence of the cloned V-C fragments that make up the TRA V-C entry vector library is presented as SEQ0049 to SEQ0094. The presented sequences include all the Kozac sequence preceding the start codon of the variable segment, to the stop codon of the C segment.

Example 2

Design and assembly of TRA J Donor vector library for native human TRA repertoire

A TORES consists of a V-C entry vector library and J donor vector library for a given TCR chain. When combined with a target odeCDR3 sequence to be inserted into a selected V-J-C context, a full-length TCR ORF can be reconstituted. Through varying odeCDR3 sequence features and/or V/J/C selection, this reconstitution step may also represent a sequence diversification step in TCR ORF engineering workflows

In the present example, the design and assembly of a TRA J Donor vector library that contains the native human TRA J sequence repertoire. This is an example of the generic J Donor vector type depicted in Figure 3.

In the present example, a **TRA J receiving cassette fragments** are constructed and inserted to a J donor vector backbone to create a **J receiving cassette vector**. Subsequently, a synthetic **TRA J segment parts** may be assembled into a **TRA J receiving cassette vector** to create the **J Donor vector** library. This flexible multistep assembly method allows rapid and cost effective engineering of J donor segment features, such as variations in J segment length.

The DNA components required for a **TRA J donor vector** library are:

I. TRA J receiving cassette fragment
II. J donor vector backbone
III. TRA J receiving cassette vector
IV. TRA J segment part

Design of synthetic TRA J receiving cassette fragment

The annealing of two single stranded DNA oligonucleotides is used to generate the receiving site cassette fragment that by design contains 4-nucleotide single-strand overhangs at each end of the DNA fragment; **Overhang *3** and **Overhang *4**. The 4-nucleotide overhangs to permit directed ligase-dependent cloning into a J donor vector backbone to create the **TRA J receiving cassette vector**, depicted in Figure 8.

The pair of Type IIS restriction sites, **BsaI ←** and **BsaI →** are positioned at the 5′ and 3′ end of the receiving site cassette DNA fragment. The direction of the BsaI recognition site is to guide BsaI enzyme to cut the DNA towards the centre of the construct. These sites are used during TRA ORF reconstitution protocol by generating **overhang ‡2-5′** and **overhang ‡3-3′. Overhang ‡3** is a component of the TRA C part encoded in the receiving cassette fragment, while **overhang ‡2** is defined after the TRA J segment part is cloned (*infra vide*).

The BbsI pair of Type IIS recognition sites **BbsI ←** and **BbsI →** are encoded near the middle of the cassette and used for assembly of the TRA J donor vector, in creating complementary overhangs included in synthesized **TRA J segment parts** (*infra vide*). The 5' BbsI site, **BbsI ←**, cuts into the BsaI site to create **overhang *5** at the 3' end of this feature. The 3 BbsI site, **BbsI →**, cuts into the TRA C part element, to create **overhang *6** at the 5' end of this element. These overhangs are encoded within the BsaI and TRA C part features of this construct as to avoid addition of non-native nucleotides that would be incorporated into the final reconstituted TRA ORF.

The region between **BbsI →** enzyme generated overhang and the **BsaI ←** enzyme generated overhang encodes a proportion of the TRA C region starting from the second nucleotide of the TRA C genetic fragment, **TRA C part**. The motivation for starting from the second nucleotide of the TRA C genetic fragment is because in the present example of a human TRA locus TORES, the resulting overhang is TATC and not a palindromic overhang, which would be the case if the beginning of the TRA C genetic fragment were including (resulting overhang ATAT). A palindromic overhang should be avoided, as it would permit two vector ends joining without the required TRA J segment part insert. The orientation of the **BbsI →** site permits the in-frame ligase dependent cloning of all TRA J fragments 3′ end to the 5′ beginning of the TRA C region in the receiving site cassette. The orientation of the **BsaI ←** site permits the in-frame ligase-dependent cloning of the beginning of the TRA C region with the remaining TRA C fragments in the final step of the TRA full length ORF reconstitution protocol using a

complete TORES.

Between the two BbsI binding sites is an 8 nucleotide recognition sequence for the enzyme **NotI**. This restriction site is utilized as a **negative selection** marker to reduce the background of the parental plasmid colonies. This is achieved when NotI enzyme is added after the TRA J gene fragment insertion has been performed. Therefore plasmids correctly cloning a TRA J gene fragment would remain circular in the presence of NotI enzyme but parental plasmids that did not exchange its NotI site for a TRA J gene fragment will be linearized, in turn biasing the bacterial transformation to propagate a complete circular TRA J fragment-containing plasmid.

Sp denote nucleotide additions to specific points of the TRA J receiving cassette fragment to achieve the correct spacing of Type IIS restriction enzyme binding site and the cut site, when adjacent to such sites. **Sp** blocks flanking the NotI restriction enzyme binding site sequence have been used to space the NotI binding and cut site appropriately for efficient action. Additional nucleotides have been included to maintain correct reading frame within the final reconstituted full-length TRA. The selection of nucleotides considered the potential impact of DAM methylation of the BsaI binding site.

The full DNA sequence for the **TRA J receiving cassette fragment** oligonucleotides in the present example of native human TRA chains are presented as, SEQ0095 and SEQ0096. Both forward (F1) and reverse (R1) oligonucleotide sequences are listed.

Design of the J donor vector backbone

The **J donor vector backbone** is used to insert the **TRA J receiving cassette fragment** to create the **TRA J receiving cassette vector**. The backbone is thus carried through to the J Donor vector library. In the final reaction to create TRA full-length ORFs, this backbone is a reaction byproduct (Figure 4e), and thus carries minimal features as depicted in Figure 9.

In the present example, the **J donor vector backbone** encodes a Col E1 origin of replication, **ori**. The antibiotic resistance is the aminoglycoside 3'-phosphotransferase gene, **positive selection selection #2.** Aminoglycoside 3'-phosphotransferase confers resistance to antibiotic substrates such as kanamycin, streptomycin, neomycin, and gentamicin. This alternate positive selection is used to ensure J donor vectors are not selected for after full-length TCR ORF reconstitution, which are selected on **positive**

selection #1.

In the present example the vector **EcoRI** and **XhoI** restriction enzyme binding sites that generates complementary overhang, **overhang *3'** and **overhang *4'**, respectively. **Overhang *3'** is complementary with **Overhang *3** contained within the **TRA J receiving cassette fragment. Overhang *4'** is complementary with **Overhang *4** contained within the **TRA J receiving cassette fragment.** These overhangs permits directed cloning of the TRA J receiving cassette fragment.

Sp block denotes nucleotides added between the **EcoRI** and **XhoI** restriction enzyme binding sites for distancing the two sites to ensure efficient action.

In the present example, the J donor backbone is presented as SEQ0097.

Method to assemble the TRA J receiving cassette vector

This method utilizes standard molecular biology techniques to assemble the given **TRA J receiving cassette fragments** (Figure 8) into a given **J donor vector backbone** (Figure 9) to create a **TRA J receiving cassette vector** (Figure 10). The resulting **TRA J receiving cassette vector** is used to insert **TRA J segment parts** (Figure 11) to construct **TRA J Donor vectors** (Figure 3).

First, the two oligonucleotides to form the TRA J receiving cassette DNA fragment must be phosphorylated and annealed.

<u>Reaction mix</u>

Oligonucleotide (sense strand) (100 µM)	1 µl
Oligonucleotide (anti-sense strand) (100 µM)	1 µl
T4 ligase buffer 10x	1 µl
T4 PNK	1 µl
H_2O	6 µl

<u>Reaction conditions</u>

Incubate for 37˚C for 1 hour

Denature at 95˚C for 5 min

Anneal sense and anti-sense oligonucleotides by slowly cooling the reaction down to 25˚C at 3˚C per min

Assembly ligation of TRA J receiving cassette fragments and J donor vector backbone.

Reaction mix

Linear vector backbone	100 ng
Receiving site cassette DNA fragment (0.5 µM)	2 µl
T4 ligase buffer 10x	2 µl
T4 ligase	0.5 µl
H_2O	up to 20 µl

Reaction conditions

Incubate for 1 hour at 25°C

Heat inactivate at 65°C for 10 min

Resulting product is transformed into competent E.coli cells and selected for Kanamycin resistant colonies. Resistant colonies are selected to determine correctly assembled constructs. The resulting plasmid is the **TRA J receiving cassette vector**. In the present example, the TRA J receiving cassette vector is presented as SEQ0098 and depicted in Figure 10.

Design of synthetic TRA J segment parts

Having generated the **TRA J receiving cassette vector** synthetic TRA J segment parts must be generated to insert into this vector. Each TRA J sequence is inserted into an independent **TRA J receiving cassette vector** context to generate the **TRA J donor vector library as** part of the human TRA TORES.

The **TRA J donor vector library** comes in two different forms, comprised of a long or short J segment part. The **short TRA J segment part** encodes all amino acids from the start of the CDR3 border codon. However, considering that the majority of TRA J segments are trimmed back by less than 10 nucleotides during TCR rearrangement, a TRA J donor library containing a longer TRA J germline segment is designed, **long TRA J segment part**. The motivation for a longer TRA J gene fragment library is that a shorter oligonucleotide duplex encoding CDR3 (odeCDR3) would be required for the full length TRA reconstitution, than if the **short TRA J fragment** would be used. Since highly variable sequences are provided as short oligonucleotide duplexes, **odeCDR3**, a shorter CDR3 oligonucleotide synthesis is less likely to contain truncated or mutated

oligonucleotide contaminants and therefore reduce the likelihood of oligonucleotide duplex with sequence errors being cloned during full length TRA reconstruction. Furthermore, shorter **odeCDR3** syntheses are cost-saving.

The **TRA J segment parts** are constructed by annealing two single-stranded DNA oligonucleotides designed to contain 4-nucleotide single-strand overhangs at each end of the DNA fragment. The resulting TRA J segment part is depicted in Figure 11.

The 5' overhang designated **Overhang *5'** is complementary to the **Overhang *5** generated within **J donor receiving cassette vector** by BbsI action. The 3' overhang designated **Overhang *6'** is complementary to the **Overhang *6** generated within **J donor receiving cassette vector** by BbsI action. This pair of complementary overhangs permits directional cloning of the **TRA J segment parts** into the **TRA J receiving cassette vector**.

The **Short TRA J segment part** encodes all amino acids from the start of the CDR3-J border Phe codon. The CDR3 is defined as the sequence flanked by the C-terminal-conserved Cys of the V region, and Phe of the J region which is part of the Phe-Gly/Ala conserved motif. This conserved Phe-Gly/Ala motif is utilized to standardize the 5′ overhangs of the TRA J fragments to TTTG for downstream TRA reconstitution. The exceptions to this standardization in the present example are human TRAJ33 and TRAJ38 that border the CDR3 region with Trp and Gly. The 5′ overhangs are TGGG for both TRAJ33 and TRAJ38 in the present example.

The **long TRA J segment part** is designed to encode more amino acids N-terminal of the CDR3 border amino acids. The start point of each long gene fragment is at the first nucleotide of an amino acid codon positioned 10-12 nt from the 5′ end of the germline encoded TCR joining element. The 5' end of each **long TRA J segment part** remains identical to that of the **short TRA J segment part.**

To both **short and long TRA J segment parts** an adenine, represented as the A block in figure 11, is added to the 3′ end of each TRA J segment part. This adenine represents the first nucleotide of the **TRA C fragment** that is excluded from the **TRA J receiving cassette**.

The sequences of the short TRA J segment parts of the present example of native hu-

man J segments are presented as SEQ0099 to SEQ0210 and the long TRA J segment parts SEQ0211 to SEQ0322. In both cases, both forward (F1) and reverse (R1) oligonucleotide sequences are listed.

Method to assemble the Short or Long J-Donor vector library

This method utilizes standard molecular biology techniques to clone the **Short TRA J segment** or **Long TRA J segment part part** (Figure 11) into the **TRA J receiving cassette vector** (Figure 10) to create **TRA J donor vectors** (Figure 3) containing the **short** or **long TRA J segments**. In this present example, the method performs the restriction enzyme digestion and ligation reaction in a single reaction.

The DNA components required for a J donor vector library is as follows:

I. Short TRA J segment part or Long TRA J segment part
II. J donor receiving cassette vector

Phosphorylation and Annealing two oligonucleotides to form the TRA J segment part DNA fragment

Reaction mix

Oligonucleotide (sense strand) (100 µM)	1 µl
Oligonucleotide (anti-sense strand) (100 µM)	1 µl
T4 ligase buffer 10x	1 µl
T4 PNK	1 µl
H_2O	6 µl

Reaction conditions

Incubate for 37˚C for 1 hour

Denature at 95˚C for 5 min

Anneal sense and anti-sense oligonucleotides by slowly cooling the reaction down to 25˚C at 3˚C per min

RE digestion and ligation reaction

TRA J receiving cassette backbone	100 ng
TRA J DNA fragment (0.5 µM)	2 µl
10x NEB T4 ligase buffer	2 µl

BbsI	0.5 µl
T4 DNA ligase	0.5 µl
H_2O	up to 20 µl

Reaction conditions

Step 1; 2 min at 37°C

Step 2; 3 min at 16°C

Repeat step 1 and 2, 20 times

5 min at 50°C

5 min at 80°C

Return to room temperature

Add 0.5 µl of NotI enzyme and incubate for 30 min at 37°C to linearize parental vector.

Reaction product is transformed into competent E.coli cells and selected for Kanamycin resistance. Selected resistant colonies are sequenced to determine correctly assembled constructs. The resulting constructs make up the TRA J donor vector library, encoding either a long or a short TRA J gene fragment.

The sequence of the resulting libraries, excluding backbone sequence outside of the BsaI recognition sites, are presented as SEQ0323 to SEQ0378 for the **TRA short J donor library** and SEQ0379 to SEQ0434 for the **TRA long J donor library**.

Example 3

Design and assembly of TRB V-C entry vector and TRB J Donor vector libraries for native human TRB repertoire

In the above examples, the design and assembly of V-C entry vector and J donor vector libraries for the native human TRA repertoire was described in detail. The overall design and assembly of such vector libraries encoding sequences of the TRB repertoire is essentially the same. In the present example, the design and assembly of the **TRB V-C entry vector** and **TRB J Donor vector libraries** will be briefly outlined in order to construct a TORE or the native human TRB TCR locus.

Design and assembly of TRB V-C entry vector library for native human TRB repertoire

In the present example, the design and assembly of a TRB V-C entry vector library that contains the native human TRB V-C sequence repertoire. This is an example of the

generic V-C entry vector type depicted in Figure 2.

The DNA components required for a TRB V-C vector library are:

I. A TRB V cloning fragment for each functional TRB V gene segment encoded in the human genome
II. TRB C1 or TRB C2 cloning fragment
III. A V-C entry vector backbone

In contrast to the human TRA locus, the human TRB locus encodes two distinct constant segments, TRB C1 and C2. Thus, to capture both constant regions, two V-C entry vector sets are constructed to pair each of the V segments with each C1 and C2 segments.

In the present example, the TRB V and TRB C cloning fragments were synthesized and used to assemble into a target V-C entry vector backbone in a single restriction enzyme and ligase reaction. In the present example, the target V-C entry backbone was designed to permit transient expression of reconstituted TRB ORFs within mammalian cells.

In the present example, the TRB V-C entry vector library is constructed using Type IIS restriction enzyme BbsI. The Type IIS restriction enzyme used in functioning of the library to reconstitute full-length TRB ORFs is BsaI.

Design of synthetic TRB V cloning fragments

The arrangement of genetic elements of the TRB V cloning fragments is identical to those of the TRA V cloning fragments described in Example 1, as depicted in Figure 5.

Full DNA sequences for the **TRB V cloning fragments** in the present example of native human TRB chains are presented as SEQ0435 to SEQ481.

Design of synthetic TRB C cloning fragment

The arrangement of genetic elements of the TRB C cloning fragments is identical to those of the TRA C cloning fragments described in Example 1, as depicted in Figure 6.

The TRB locus encodes two distinct C segments, and both are included in the design

of the **TRB V-C entry vector** library.

The full DNA sequence for the **TRB C cloning fragments** in the present example of native human TRB chains are presented as SEQ0482 and SEQ0483.

Method to assemble TRB V-C entry vector library

The method to assemble the given **TRB V and TRB C cloning fragments** into a given **V-C entry vector backbone** to create a **TRB V-C entry vector** is identical to that described in Example 1. The V-C entry vector backbone used, designed for transient expression of full-length TRA or TRB ORFs in mammalian cells is used in the present example, and in example 1 (SEQ0048).

The sequence of the cloned V-C fragments that make up the TRA V-C entry vector library is presented as SEQ0484 to SEQ0577.

Design and assembly of TRB J Donor vector library for native human TRB repertoire

In the present example, the design and assembly of a TRB J Donor vector library that contains the native human TRB J sequence repertoire. This is an example of the generic J Donor vector type depicted in Figure 3.

In the present example, a **TRB J receiving cassette fragments** are constructed and inserted to a J donor vector backbone to create a **TRB J receiving cassette vector**. Subsequently, a synthetic **TRB J segment part** may be assembled into a **TRB J receiving cassette vector** to create the **TRB J Donor vector** library. This flexible multi-step assembly method allows rapid and cost effective engineering of J donor segment features, such as variations in J segment length.

This procedure follows the same pattern as the TRA J donor vector assembly described in Example 2. However, it should be noted that since the J receiving cassette fragments contain parts of the C segment, the TRA J and TRB J receiving cassette fragments differ with regard to the C part sequence, that must correspond to the respective C gene segments. Moreover, in contrast to TRA J scenario that only requires a single J receiving cassette fragments, the TRB J requires two distinct J receiving cassette fragments to account for the use of alternate C1 and C2 segments.

The DNA components required for a **TRB J donor vector** library are:

I. TRB J C1 or TRB J C2 receiving cassette fragment
II. J donor vector backbone
III. TRB J C1 or TRB J C2 receiving cassette vector
IV. TRB J segment part

Design of synthetic TRA J receiving cassette fragment

The annealing of two single stranded DNA oligonucleotides is used to generate the **receiving cassette fragments**, which contain 4-nucleotide single-strand overhangs at each end of the DNA fragment, depicted in Figure 8. The 4-nucleotide overhangs permit directed ligase-dependent cloning into a J donor vector backbone to create the **TRB J receiving cassette vector**,

The two receiving cassette fragments required for alternate use of C1 and C2 segments are presented as SEQ0578 and SEQ0581. For each fragment, the forward (F1) and reverse (R1) oligonucleotide sequences are provided.

Method to assemble the TRB J receiving cassette vectors

The method for assembly of the **TRB J receiving cassette vectors** is identical to that of the method for assembly of **TRA J receiving cassette vectors** described in Example 2. The same **J donor vector backbone** (SEQ0097) is used to generate two **TRB J receiving cassette vectors**, each containing one C1 or C2 part corresponding to the alternate C segments for the TRB locus.

The resulting two **TRB J receiving cassette vector** is used to insert **TRB J segment parts** to construct **TRB J Donor vectors**.

The resulting **TRB J receiving cassette vectors** are presented as SEQ0582 and SEQ0583.

Design of synthetic TRB J segment parts

The **TRB J segment parts** are constructed by annealing two single-stranded DNA oligonucleotides designed to contains 4-nucleotide single-strand overhangs at each end of the DNA fragment. The arrangement of this part and method of assembly are identical to that of the TRA J segment parts, and depicted in Figure 11.

In the case of the **Short TRB J segment part** encodes all amino acids from the start of the CDR3-J border Phe codon. The CDR3 is defined as the sequence flanked by the C-terminal-conserved Cys of the V region, and Phe of the J region, which is part of the Phe-Gly motif conserved across all human TRB J segments. This conserved Phe-Gly motif is utilized to standardize the 5′ overhangs of the TRA J fragments to TTTG for downstream TRB reconstitution. Unlike the TRA J segments, there are no exceptions to this standardized overhang in the TRA J segment parts in the present example.

To both **short and long TRB J segment parts** an adenine, represented as the A block in figure 11, is added to the 3′ end of each TRB J segment part. This adenine represents the first nucleotide of the **TRB C fragment** that is excluded from the **TRB J receiving cassettes**.

The sequences of the short TRB J segment parts of the present example of native human J segments are presented as SEQ0584 to SEQ0609, and the long TRB J segment parts SEQ0610 to SEQ0635. In both cases, both forward (F1) and reverse (R1) oligonucleotide sequences are listed.

Method to assemble TRB Short or Long J Donor vector library

The procedure to assemble the TRB J donor libraries is identical to that of the TRA libraries described in Example 2. However, in the case of the TRB libraries, there are four libraries to generate, in contrast to the short and long libraries for the TRA locus segments.

In the case of TRB libraries, each short and long libraries can be constructed to carry each of the alternate C1 and C2 C segments, resulting in four subsets within the TRB J donor library.

The DNA components required for a J donor vector library is as follows:

I. Short TRB J segment part or Long TRB J segment part
II. TRB J C1 or TRB J C2 receiving cassette vector

Following the same procedure as described in example 2, the four resulting subsets within the TRB J donor library are generated. The sequence of the resulting libraries, excluding backbone sequence outside of the BsaI recognition sites is presented.

TRB C1 short J donor library presented as SEQ0636 to SEQ0648

TRB C2 short J donor library presented as SEQ0649 to SEQ0661

TRB C1 long J donor library presented as SEQ0662 to SEQ0674

TRB C2 long J donor library presented as SEQ0675 to SEQ0687

Example 4

Design of TRA and TRB V-C entry vectors for recombinase mediated cassette exchange application

The design of the TRA and TRB V-C entry vectors in examples 1 and 3, respectively, are such that the full-length TCRs reconstituted with the use of these TORES for each TCR chain may be directly applicable to transient expression in mammalian cells by transfection. The overall design of the TORES allows the V-C entry vector backbone to be exchanged for any backbone suitable for downstream applications, and the necessary V-C entry vector libraries to be rapidly assembled from V and C cloning fragments.

In the present example, a pair of *heterospecific FRT* V-C entry vector backbones are used to assemble TRA and TRB V-C entry vector libraries. Each TRA and TRB V-C entry vector libraries are constructed with vector backbones containing distinct *flippase* recognition target (FRT) sequences. One downstream application of such vectors, is the submission of final reconstituted TRA and TRB pairs to rapid genomic integration into mammalian cells harboring relevant FRT sites.

In the present example, TRA V-C entry vector of the design depicted in Figure 7, contains F14 and F15 FRT sequences as the 5' and 3' genetic elements, respectively. This F14/F15 V-C entry vector backbone sequence is presented as SEQ0688.

This backbone was used to construct TRA V-C entry library of the design outlined in Figure 2, with the use of TRA V cloning fragments (SEQ0001 to SEQ0046) and TRA C cloning fragment (SEQ0047) using methods outlined in Example 1. This construction yielded a TRA V-C entry library with sequences identical to that of the transient expression TRA V-C entry vector library (SEQ0049 to SEQ0094), but flanked by the FRT sequences in the context of the F14/F15 V-C entry vector backbone.

In the present example, TRB V-C entry vector of the design depicted in Figure 7, contains FRT and F3 FRT sequences as the 5' and 3' genetic elements, respectively. This FRT/F3 V-C entry vector backbone sequence is presented as SEQ0689.

This backbone was used to construct TRB V-C entry library of the design outlined in Figure 2, with the use of TRA V cloning fragments (SEQ0435 to SEQ0481) and TRB C cloning fragments (SEQ0482 and SEQ0483) using methods outlined in Example 3. This construction yielded a TRB V-C entry library with sequences identical to that of the transient expression TRB V-C entry vector library (SEQ0484 to SEQ0577), but flanked by the FRT sequences in the context of the FRT/F3 V-C entry vector backbone.

Overall, the TRA and TRB V-C entry vectors in the present example may be used interchangeably in with V-C entry vectors in above examples. In each case, the J donor vector remains unchanged, as does design and provision of odeCDR3 (see Example 6). In effect, different V-C entry vector backbones provide full-length TCR ORFs in desired vector contexts for downstream applications as a direct product of the TCR reconstitution reaction.

Example 5

Design of V-C entry vectors and J donor vectors suitable for provision of any V-J-C combinations

In the above examples of V-C entry vectors and J donor vectors, native human TRA and TRC TCR chains are used as a practical demonstration of the TORES. However, it is clear that any native or non-native TCR chain can be employed in the TORES format upon insertion of required TCR chain sequences into the specified vector formats.

To achieve a TORES for any given TCR chain, four sequence elements are required specific for said TCR chain:

X – a variable (V) gene segment fragment
Y – a constant (C) genes segment fragment
Y'– a constant (C) gene segment part
Z – a joining (J) gene segment fragment

According to the above examples, these four forms of sequence element can be assembled into various vector contexts to construct and deploy a TORES for any given V-

J-C combination for any given TCR chain. For example, systems for the native human TRG and TRD locus, variant synthetic human TCR chain forms, or native TCR chain forms of an organism other than humans.

V-C entry vector assembly and final construct

In order to achieve assemble of the V-C entry vector, two intermediates can be constructed that contain the relevant V and C gene segment fragments.

A V cloning fragment, according to Figure 5, is designed to contain required V gene segment fragment in the context of appropriate cloning sites for final V-C entry vector assembly. In the present example, as is also presented in examples 1, 3 and 4, the assembly of V-C entry vectors was achieved through use of BbsI restriction enzyme sites create target overhangs at the 5' and 3' end as to match complementary overhangs in the V-C entry vector backbone and C cloning fragment, respectively. The exemplified strategy also incorporates a NotI restriction enzyme site as a negative selection marker.

The generic sequence of such a cloning approach is presented as SEQ-0690, wherein the V gene segment fragment is bounded by BbsI sites at the 5' and 3' ends, and the V gene segment fragment is denoted as XN_n, wherein X is the designation for the V gene segment, N represents any nucleotide, and n represents the number of nucleotides in said sequence. The V gene segment fragment is bounded by Kozac and spacer nucleotides at the 5' and 3' ends, respectively, and include the translation start codon.

The second intermediate that can constructed to assemble a V-C entry vector libraries for specific TCR chain sequences is the C cloning fragment as depicted in Figure 6. In the present example, as is also presented in examples 1, 3 and 4, the assembly of V-C entry vectors was achieved through use of BbsI restriction enzyme sites create target overhangs at the 5' and 3' end as to match complementary overhangs in the V cloning fragment and backbone the V-C entry vector, respectively. The exemplified strategy also incorporates a NotI restriction enzyme site as a negative selection marker.

The generic sequence of such a cloning approach is presented as SEQ-0691, wherein the C gene segment fragment is bounded by BbsI sites at the 5' and 3' ends, and the C gene segment fragment is denoted as YN_n, wherein Y is the designation for the C gene segment, N represents any nucleotide, and n represents the number of nucleotides in

said sequence. The C gene segment fragment is bounded by the cytosine 5' of the first Glu codon of the C gene segment, or equivalent conserved site, and spacer nucleotide(s) at the 5' and 3' ends, respectively, and would include the translation stop codon.

To assemble a V-C entry vector using the above defined V and C cloning fragments, these fragments are combined with a V-C entry vector backbone as depicted in Figure 7 and described in examples 1 and 3. In the present and above examples, this backbone is defined with the **Acc65I** and **XbaI** restriction enzyme binding sites that generate **overhang *1'** and **overhang *2'**, respectively. **Overhang *1'** is complementary to **overhang *1** within the V cloning fragment (figure 5). **Overhang *2'** is complementary to **overhang *2** within the C cloning fragment (figure 6). These complementary overhangs permit directed cloning of the V and C cloning fragments into the V-C entry vector backbone.

Sp feature denotes nucleotides added between the **Acc65I** and **XbaI** restriction enzyme recognition sites required for distancing the two sites for efficient action.

As outlined in example 4, a V-C entry vector backbone (Figure 7), by way of providing the 5' and 3' genetic elements to the V-C entry vector (Figure 2), ultimately provides the vector context into which a full length TCR ORF is reconstituted (Figure 4h). Thus, a range of selections can be made for the V-C entry vector backbone / V-C entry vector to satisfy a range of downstream applications for the reconstituted TCR ORF.

In the present example, as in examples 1, 3 and 4 above, three different V-C entry vector backbones are provided.

The sequence of the V-C entry vector backbone with the **5′ genetic element** encoding the CMV constitutive promoter, and the **3′ genetic element** encoding the SV40pA polyadenylation signal is presented as SEQ0048. This vector context permits transient expression of reconstituted TCR ORFs in mammalian cells.

When this transient expression V-C entry vector backbone SEQ0048 is combined with the generic V cloning fragment (SEQ0690) and C cloning fragment (SEQ0691) in the present example, a generalized V-C entry vector can be defined by the sequence SEQ0692.

The sequence of the V-C entry vector backbone with the **5′ genetic element** encoding the FTR site, F14, sequence, and the **3′ genetic element** encoding the FRT site, F15, sequence is presented as SEQ0688. This vector context permits flippase mediate cassette exchange of reconstituted TCR ORFs into genetic contexts bounded by F14 and F15 heterospecific FRT sites. For example, such an approach can be employed for stable integration of TCR ORFs into mammalian cell lines harboring such heterospecific FRT receiver sites.

When this F14/F15 RCME V-C entry vector backbone SEQ0688 is combined with the generic V cloning fragment (SEQ0690) and C cloning fragment (SEQ0691) in the present example, a generalized V-C entry vector can be defined by the sequence SEQ0693.

The sequence of a second V-C entry vector backbone with the **5′ genetic element** encoding the FRT stie, FRT, sequence, and the **3′ genetic element** encoding the FRT site, F3, sequence is presented as SEQ0689. This vector context permits flippase mediate cassette exchange of reconstituted TCR ORFs into genetic contexts bounded by FRT and F3 heterospecific FRT sites. For example, such an approach can be employed for stable integration of TCR ORFs into mammalian cell lines harboring such heterospecific FRT receiver sites.

When this FRT/F3 RCME V-C entry vector backbone SEQ0689 is combined with the generic V cloning fragment (SEQ0691) and C cloning fragment (SEQ0692) in the present example, a generalized V-C entry vector can be defined by the sequence SEQ0694.

In the present example, one TCR chain can be reconstituted in one of the two RCME contexts, and a second chain in the other RCME context. For example, a TRA ORF can be reconstituted in the F14/F15 context, and a TRB ORF can be reconstituted in the FRT/F3 context. Thus, paired TRA/TRB constructs can be integrated into a mammalian cell line harboring F14/F15 and FRT/F3 sites by RCME for stable expression of the TRA/TRB pair in said cell line.

J donor vector assembly and final construct

Described above are generic constructs that may be used to assemble various V and C TCR gene segment combinations into V-C entry vectors as a component of a TORES.

In order to complete a TORES, appropriate J donor vectors are required. As described above in example 2, a J donor vector may be assembled in a multistep process to achieve a J donor system with flexibility to design features of J segments. One important aspect to note is that the J donor vector contributes a C part corresponding to the C segment of a matched V-C entry vector library.

To assemble a J donor vector library, four different constructs are required;

I. J receiving cassette fragment
II. J donor vector backbone
III. J receiving cassette vector
IV. J segment part

In the present example, as in example 2, the J donor vector is assembled using BbsI restriction sites, and the overall TORES operates with the use if BsaI restriction sites carried in both the resulting J donor vector and matched V-C entry vector.

The J receiving cassette fragment as depicted in Figure 8 importantly contains a C part corresponding to the C segment carried by the matched V-C donor vector library. The generic sequence of such a cloning approach is presented as SEQ0695 to SEQ0696, wherein the C part encoding a 5' portion on the C segment is bounded by a BbsI recognition site to the 5' and a BsaI site to the 3'. The two presented sequences represent the sense and antisense oligonucleotides annealed to generate the J receiving cassette fragment, wherein the Y' sequences provided are complementary. The C part in the presented sequence is designated Y'Nn, wherein Y' is the designation for the C part matching the Y C segment described above, N represents any nucleotide, and n represents the number of nucleotides in said sequence. The BbsI site is spaced and orientated to cut in the 3' direction, to create a cloning overhang to match a J segment part to be inserted. The BsaI site is spaced and cuts in the 5' direction to create the C overhang in the final assembly reaction for generating full-length TCR ORFs in operation of the TORES. A second BsaI site at the 5' terminus of the J receiving cassette fragment, cutting in the 3' direction, is also used for the final operation of the TORES. The internal BbsI site cutting in the 5' direction cerates the second cloning overhang for assembly of the J-donor vector. In the present example, a spaced NotI restriction site is encoded between the BbsI recognition sites to act as a negative selection marker to eliminate parental fragments from the J donor vector assembly process. The J receiv-

ing fragment is assembled by annealing two complimentary oligonucleotides designed to produce two 4-nucleodtide single-stranded overhangs, designated overhang*3 at the 5' terminus, and overhang*4 at the 3' terminus as described in Example 2.

To assemble the intermediate J receiving cassette vector with the above described J receiving cassette fragment, a J donor backbone must be created, as depicted in Figure 9. In the present example, the cloning features of this vector backbone are EcoRI and XhoI restriction sites spaced to allow efficient restriction enzyme action. Digestion of this backbone with EcoRI and XhoI generate overhang*3' and overhang*4' complementary with overhang*3 and overhang*4 in the assembled J receiving cassette fragment. This permits directional cloning of the J receiving cassette fragment into the J donor backbone to create the J receiving cassette vector as described in Example 2. An example of the J donor vector backbone with a positive selection marker different from that of the V-C entry vector described above is presented as SEQ0097.

The J receiving cassette vector obtained above is depicted in Figure 10, and essentially represents the J receiving cassette fragment in the context of the J donor vector backbone, and is presented as SEQ0697, wherein Y' is the designation for the C part matching the C segment that is designated Y and described above, N represents any nucleotide, and n represents the number of nucleotides in said sequence.

It is into the J receiving cassette vector that an array of J segments are inserted to construct a J donor vector library. The design of the J gene segment is depicted in figure 11, wherein a J gene segment is bounded by overhang*5' and overhang*6' to the 5' and 3' ends, respectively. The J segment part is spaced from the 3' overhang by a single adenine nucleotide in the present example, as to maintain enzyme recognition sequences and correct reading frame when inserting the J segment part into the J receiving cassette vector to create a J donor vector, as descried in Example 2. In the present example, a generic sequence for the J segment part is presented as SEQ0698 to SEQ0699 wherein the J gene segment part is denoted as ZN_n, wherein Z is the designation for the J gene segment, N represents any nucleotide, and n represents the number of nucleotides in said sequence. The two sequences presented represent sense and antisense oligonucleotides annealed to assemble the J segment part, wherein the Z sequences are complementary.

The generic final obtained J donor vector in the present example is presented as

SEQ0700, a representation of which is presented in Figure 3. This J donor vector in the present example represents an inserted J segment part and the C part from the J receiving cassette vector. The sequence is presented with two annotated sequence inserts of ZN_n and $Y'N_n$, wherein Z is the designation for the J gene segment, and Y' is the designation of the C segment part, N represents any nucleotide, and n represents the number of nucleotides in said sequences.

Within all provided sequences, the selection of nucleotides considered the potential impact of DAM methylation of the BsaI binding sites. All additional recognition sequences for assembly and selection enzymes within the vectors were removed. Any such recognition sequences should also be removed from germline or synthetic TCR V, J and C elements inserted into the described fragment and vector contexts.

Example 6

Design and generation of oligonucleotide duplex encoding CDR3 (odeCDR3)

In the above example, various formats of TORES are described in the design and generation of matched V-C entry vector and J donor vecror libraries for various applications. The utilization of these V-C entry vector and J donor vector libraries for one-step reconstitution of full-length TCR open reading frames requires an oligonucleotide duplex encoding CDR3 (odeCDR3) construct to be provided in order to complete the target full-length TCR chain sequence (Figure 4c). Once V-C entry vector and J donor vector libraries are generated, these vectors represent stock items that may be drawn upon indefinitely to select desired V-J-C combinations of target full-length TCR chains sequences. In contrast, the odeCDR3 represents a short unique sequence specific to the target full-length TCR ORF.

The present example describes the design and generation of odeCDR3 for use in the native human TRA and TRB vector platforms.

Design of the TRA odeCDR3

The annealing of two single stranded DNA oligonucleotides generates an **odeCDR3** that contains 4-nucleotide single-strand overhangs at each end of the DNA fragment, as depicted in Figure 4c. The 4-nucleotide overhangs are designed to permit directed ligase dependent cloning to the 3′ end of the TRA V segment encoded in the entry vector, (**Overhang ‡1-5′**) and the 5′ end of the TRA J fragment during TRA reconstitution (**Overhang ‡2-3′**). **Overhang ‡1-5′** is standardised to **CTGC**, complementary to the

standardized **Overhang ‡1-5** encoded in the V segment of the TRA V-C entry vector. In the case of **Overhang ‡2-3′**, there are two sequence forms that this can take, which is determined by sequence divergence among J segments from the human TRA locus. For native human TRA J segments TRAJ33 and TRAJ38, the **Overhang ‡2-3′** is standardized to **TGGG**, complementary to the **Overhang ‡2-3** encoded in the J donor vector of these two J segments. For all other human TRA J segments **Overhang ‡2-3′** is standardized to **TTTG**, complementary to the **Overhang ‡2-3** encoded in the J donor vector of these J segments (see Example 2).

Design of the TRB odeCDR3

As for the TRB odeCDR3, the annealing of two single stranded DNA oligonucleotides generates an **odeCDR3** that contains 4-nucleotide single-strand overhangs at each end of the DNA fragment, as depicted in Figure 4c. The 4-nucleotide overhangs are designed to permit directed ligase dependent cloning to the 3′ end of the TRB V segment encoded in the entry vector, **Overhang ‡1-5′,** and the 5′ end of the TRB J fragment during TRB reconstitution, **Overhang ‡2-3′**. The **Overhang ‡1-5′** is standardised to **TTGC**, complementary to the standardized **Overhang ‡1-5** encoded in the V segment of the TRB V-C entry vectors. In contrast to the TRA odeCDR3 where two alternative **Overhang ‡2-3** forms are required, for the TRB odeCDR3 **Overhang ‡2-3** is standardized to **TTTG**, complementary to the **Overhang ‡2-3** encoded in the J donor vector of all TRB J segments (see Example 3).

General odeCDR3 design

In general, an odeCDR3 design must be matched to the overhangs the 4-nucleotide overhangs are designed to permit directed ligase dependent cloning to the 3′ end of the V segment encoded in the entry vector (**Overhang ‡1-5′)** and the 5′ end of the J fragment during reconstitution, (**Overhang ‡2-3′)**.

Method to generate phosphorylated CDR3 DNA oligonucleotide duplex

Phosphorylation and Annealing two oligonucleotides to form the odeCDR3

Reaction mix

Oligonucleotide (sense strand) (100 µM)	1 µl
Oligonucleotide (anti-sense strand) (100 µM)	1 µl
T4 ligase buffer 10x	1 µl

T4 PNK	1 µl
H_2O	6 µl

Reaction conditions

Incubate for 37°C for 1 hour

Denature at 95°C for 5 min

Anneal sense and anti-sense oligonucleotides by slowly cooling the reaction down to 25°C at 3°C per min

Example 7

Reconstitution of TRA and TRB full-length ORFs from respective V-C entry and J donor vector libraries

JG9a/b example.

This example describes the steps used for defining the vector library components and odeCDR3 required to reconstitute TRA and TRB full length TCR ORFs given sequence information of the target TCRs. The present example also demonstrates the assembly process of full-length a model TRA and TRB TCR chain pair, and confirms its specificity via transient expression in a human cell model, and staining of surface-presented TCR with specific HLA-multimer reagent.

Selection of V-C entry vector, J donor vector and odeCDR3

The sequences of all possible germline fragments that are represented in the cloning library are aligned to a TRA or TRB sequences of interest. The genetic fragments with the highest identity to the TRA or TRB sequence determines which V, J and C genetic element will constitute the desired TRA or TRB clonotype sequences. For TRA, the appropriate V-C entry vector is selected based on the determination of the V usage of the desired TRA. For TRB, when sequence coverage is sufficient to determine the V and C usage, the appropriate V-C entry vector will be selected that corresponds to the V usage of the desired TRB clonotype, in addition to whether said clonotype uses TRBC1 or TRBC2.

In the case when both the short and long version of the specific TRAJ or TRB J genetic element align to the TRA and TRB sequence, respectively, the corresponding plasmids encoding the longer genetic elements will be used for the TRA reconstruction.

The odeCDR3 sequence required for the TRA to be synthesised is determined as the

region between the 3′ end of the TRA V aligned genetic fragment and the 5′ end of the aligned TRAJ genetic fragment. The oligonucleotide sense strand requires the additional 5′ 4-nucleotide overhang, **Overhang ‡1-5′,** CTGC that is universal to the overhang generated on the TRA V entry vector when digested with BsaI, **Overhang ‡1-3′**. The complementary oligonucleotide anti-senses strand requires the additional 5′ 4-nucleotide overhang, **Overhang ‡2-3′**, that is unique to the overhang specifically for the TRAJ vector added to the TRA reconstruction reaction, **Overhang ‡2-5′**.

The CDR3 sequence required for the TRB to be synthesised is determined as the region between the 3′ end of the TRB V aligned genetic fragment and the 5′ end of the aligned TRB J genetic fragment. The oligonucleotide sense strand requires the additional 5′ 4-nucleotide overhang, **Overhang ‡1-5′,** TTGC that is universal to the overhang generated on the TRB V entry vector when digested with BsaI, **Overhang ‡1-3′**. The complementary oligonucleotide anti-senses strand requires the additional 5′ 4-nucleotide overhang, **Overhang ‡2-3′,** that is unique to the overhang specifically for the TRB J vector added to the TCR reconstruction reaction **Overhang ‡2-5′**.

In the present example, a model TCR TRA/TRB pair is used with a known specificity for a HLA-A2*01- restricted antigen. The sequences of the TRA and TRB chains are presented as SEQ701 and SEQ702, respectively.

Based on this full-length sequence it was straightforward to select the appropriate V-C entry and J donor vectors from the TRA and TRB libraries. In the present example, V-C entry vector of the transient expression type were used, that is, with the backbone presented as SEQ0048.

In the present example the TRA V-C entry vector SEQ0088 (from list 0049 to 0094) and J donor vector SEQ0371 (from list 0323 to 0378) were selected.

In the present example the TRB V-C entry vector SEQ0563 (from list 0484 to 0577) and J donor vector SEQ0637 (from list 0636 to 0687) were selected.

The odeCDR3 synthesised for the TRA chain is presented in SEQ703 and SEQ704 as sense and antisense, respectively.

The odeCDR3 synthesised for the TRB chain is presented in SEQ705 and SEQ706 as sense and antisense, respectively.

Method for full-length reconstitution

For each of the TRA and TRB components selected above, restriction enzyme / ligase cycle reactions were performed as described below.

RE digestion and ligation reaction

V-C entry vector	100 ng
J donor vector	60 ng
odeCDR3 oligonucleotide duplex (0.5 µM)	2 µl
10x T4 ligase buffer	2 µl
BsaI	0.5 µl
T4 DNA ligase	0.5 µl
H_2O	up to 20 µl

Reaction conditions

Step 1; 2 min at 37˚C

Step 2; 3 min at 16˚C

Repeat step 1 and 2, 20 times

5 min at 50˚C

5 min at 80˚C

Return to room temperature

Add 0.5 µl of NotI enzyme and incubate for 30 min at 37˚C

The resulting reaction product was transformed into competent E.coli cells and plated on carbenicillin containing plates.

Screening and sequencing carbenicillin resistant colonies was conducted to determine correctly assembled constructs. Screening of colonies was performed by restriction enzyme diagnostic digest of isolated plasmid DNA, and the expected DNA fragment sizes were observed by DNA electrophoresis. The resulting constructs encode the full length TCR alpha and beta clone sequences.

Validation of reconstituted TRA and TRB vectors.

To verify the specificity of the reconstituted TCR TRA/TRB pair above, both constructs were transiently transfected into a cell line that expresses all CD3 components, but lacks TRA and TRB expression. This analysis is presented in Figure 12, wherein the reconstituted TCR is transfected along side an irrelevant TCR, and an empty vector control. The empty vector shows no surface staining with antibodies for either TCRalpha/beta or CD3, nor do these control cells stain for HLA-A2*01- NLVPMVATV tetramer reagent (Figure 12, rightmost panels). TCR expression is required for surface presentation of the CD3 complex. The irrelevant TCR stains positively for TCRalpha/beta and CD3, but shows no staining for the HLA-A2*01- NLVPMVATV tetramer reagent (Figure 12, centre panels). This indicates that a TCR is expressed on the cell surface, but is not specific for the target antigen loaded in the HLA-A2*01 tetrameric reagent. The target TCR stains positively for TCRalpha/beta and CD3, and also shows positive staining for the HLA-A2*01- NLVPMVATV tetramer reagent (leftmost panels). This indicates that the reconstituted TCR both presents on the cell surface, and is also specific for the target antigen loaded in the HLA-A2*01 tetrameric reagent, as expected.

Example 8

Reconstitution of a set of antigen-specific TCR TRA/TRB chain pairs identified from human peripheral blood

One of the key requirements of upcoming strategies in personalised TCR-based diagnostics and therapeutics is the ability to rapidly capture and characterise antigen-specific TCR chain pairs from individuals. In the present example, a workflow is conducted wherein TCR TRA/TRB chain pairs are amplified and sequenced from single T-cells isolated from peripheral human blood, followed by reconstitution and validation of these pairs via transient expression in mammalian cells.

Sequencing and reconstitution of TRA and TRB chain pairs

Figure 13 presents the overall workflow. Briefly, a fresh PBMC specimen was collected from a HLA-B*07:02 positive donor, and stained with an HLA-B*07:02-TPRV HLA-multimer reagent. This HLA-multimer is the HLA-B*07:02 allele loaded with a known immunodominant antigen from the HCMV pp65 gene, the full peptide sequence of which is; TPRVTGGGAM. Single HLA-B*07:02-TPRV multimer-positive cells were FACS sorted into 0.2mL PCR tubes containing 5µL of water. Samples were snap-frozen and stored at -80°C until further processing (Figure 13 step i).

Single-cell isolates were subjected to a two-step amplification process that entails a V-region specific primer collection for each TRA and TRB (Figure 13 step ii), followed by paired nested PCR reactions that create TRA and TRB amplicons (Figure 13 step iii) for sequence analysis (Figure 13 step iv). This procedure is described previously (Han et. al. Nat Biotechnol. 2014 32(7): 684-692).

In the present example, 6 paired clones are presented as an example of downstream processing. The obtained sequences are presented as SEQ0707 to SEQ0718.

These sequences were then aligned against the a library of V, C and J gene segment for their corresponding TRA and TRB chains to determine the V, C and J segment usage of each amplified chain. This sequence analysis step also permits the definition of the CDR3 coding sequence, and thus the definition of odeCDR3 sequence (Figure 13 step v). This sequence analysis permits the selection of V-C entry vectors (Figure 13 step vi) and J donor vectors (Figure 13 step vii) for TCR chain reconstitution. The analysis also permits the synthesis of odeCDR3 for each chain reconstitution reaction (Figure 13 step viii). The selections and sequences are summarised in Table 1.

The table below presents as summary of sequence IDs for identified TCR TRA/TRB pairs in de novo sequencing and reconstitution example

Chain	SEQ ID	V Segment Usage	V-C Entry Vector SEQ	J Segment Usage	J Donor Vector SEQ	odCDR3 Forward SEQ	odCDR3 Reverse SEQ
TRA	**SEQ0707**	TRAV26-1	SEQ0082	TRAJ49	SEQ0370	SEQ0719	SEQ0720
TRB	**SEQ0708**	TRBV27	SEQ0574	TRBJ2-7	SEQ0661	SEQ0721	SEQ0722
TRA	**SEQ0709**	TRAV14DV4	SEQ0071	TRAJ12	SEQ0334	SEQ0723	SEQ0724
TRB	**SEQ0710**	TRBV7-9	SEQ0554	TRBJ1-1	SEQ0636	SEQ0725	SEQ0726
TRA	**SEQ0711**	TRAV27	SEQ0084	TRAJ9	SEQ0331	SEQ0727	SEQ0728
TRB	**SEQ0712**	TRBV27	SEQ0574	TRBJ1-2	SEQ0637	SEQ0729	SEQ0730
TRA	**SEQ0713**	TRAV14DV4	SEQ0071	TRAJ9	SEQ0331	SEQ0731	SEQ0732

TRB	**SEQ0714**	TRBV27	SEQ0574	TRBJ1-2	SEQ0637	SEQ0733	SEQ0734
TRA	**SEQ0715**	TRAV3	SEQ0052	TRAJ12	SEQ0334	SEQ0735	SEQ0736
TRB	**SEQ0716**	TRBV7-9	SEQ0554	TRBJ2-1	SEQ0655	SEQ0737	SEQ0738
TRA	**SEQ0717**	TRAV17	SEQ0073	TRAJ12	SEQ0334	SEQ0739	SEQ0740
TRB	**SEQ0718**	TRBV7-9	SEQ0554	TRBJ2-5	SEQ0659	SEQ0741	SEQ0742

Selection of these components allows the assembly of independent restriction enzyme / ligase cycle reactions for each TRA and TRB chain to perform reconstitution of the full-length TCR ORFs (Figure 13 step ix). Each reaction is performed as described in Example 7 above, which includes transformation and antibiotic selection and (Figure 13 step x), then propagation and sequencing confirmation of the resulting reconstituted TCR ORFs (Figure 13 step xi).

Validation of TRA and TRB chain pairs

To verify the specificity of the set of reconstituted TCR TRA/TRB pairs above, paired constructs were transiently transfected into a cell line that expresses all CD3 components, but lacks TRA and TRB expression, as presented in example 7. This analysis is presented in Figure 14. Cells were transfected with each of the 6 TCR ORF construct pairs, in parallel with an irrelevant TRA/TRB clone pair, and an empty vector control. Cells were analysed by flow cytometry after staining with an antibody against CD3 and the HLA-B*07:02-TPRV tetramer reagent initially used to isolate the single T-cells from the PBMC specimen. The paired constructs in the present example all displayed binding to the target HLA-B*07:02-TPRV tetramer reagent, and quantitative differences are observed. The number inset into each box represents the MFI ratio of CD3 negative over CD3 positive cell events.

Example 9

Rapid TCR chain diversification via odeCDR3 degeneracy

The diversification and selection of TCR ORFs is desirable to engineer TCRs chain pairs with altered specificities, affinities and/or signalling capacity. The TORES system is suited to the rapid generation of collections of TCR chains that are systematically altered from the original target sequence. In the present example, an approach of diversi-

fying a model TCR chain pair by including an odeCDR3 to a reconstitution reaction with a defined and limited nucleotide degeneracy at selected codon positions is presented. This approach was used to diversify the TRA chain of the model TCR TRA/TRB pair presented in Example 7. This single-reaction diversification is shown to produce a TCR set with a wide range of affinities to a specific HLA-multimer reagent when presented on the surface of mammalian cells with its natural TRB chain pair. This approach is ideally suited for rapid TCR-engineering using full-length TCR ORFs that may be presented and selected in a functional context of viable mammalian cells.

Generation of diverse TRA chain collection

Figure 15 presents the overall strategy for generating a sequence-diversified collection of TCR chains in a single reaction by use of an odeCDR3 pool. A single C-V entry vector and J donor vector are selected to represent the target V,J and C gene segments in the final full-length TCR product (Figure 15, box i and box ii). An odeCDR3 pool is generated with selected diversity, such that there are a number of different CDR3 sequences represented in the odeCDR3 pool (Figure 15, Box iii). When all components are combined into a restriction enzyme / ligase cycle reaction, the resulting product are a collection of constructs containing full-length TCR chains of defined V,J C gene segment usage, and a defined diversity in the CDR3 region (Figure 15, Box iv). The number of diversified full-length TCR chains in the final product is directly proportional to the number of odeCDR3 variants in the initial odeCDR3 pool added to the reaction.

In the present example, a model TCR TRA/TRB pair is used with a known specificity for a Human cytomegalovirus (HCMV) antigen presented in HLA-A2*01 (The same pair as Example 7). This antigenic peptide is derived from the HCMV pp65 protein, and the full amino acid sequence of the peptide antigen that is presented in HLA-A2*01 is NLVPMVATV. The sequences of the TRA and TRB chains are presented as SEQ701 and SEQ702, respectively.

In the present example, the TRA chain was the target of sequence diversification, and this was achieved through synthesis of odeCDR3 sense and antisense oligos with nucleotide degeneracy at 3 distinct positions, each altering a separate codon to result in the possibility of 4 different amino acids at each of the three codons. The codons were selected for degeneracy were spaced across the CDR3 loop. The odeCDR3 oligos are presented as SEQ0743 and SEQ0744, wherein degenerate codons are denoted N.

The odeCDR3 oligos were annealed by the method outlined in Example 6, with the 4-fold amino acid degeneracy at 3 separate codon positions resulting in an odeCDR3 product pool with 64 unique sequences, including the original coding sequence (i.e.SEQ0701).

The odeCDR3 was used to assemble the full-length TRA ORFs by the method outlined in Example 7 to create 64 unique TRA ORFs with 4-fold amino acid degeneracy at 3 distinct codon positions. In the present example, the odeCDR3 was synthesised with degenerated nucleotide usage at the indicated positions, and thus reconstitution was performed in a single tube to generate all 64 chain variants. The approach is equally valid whereby a unique odeCDR3 is provided to discrete reactions in parallel. All of the expected clones were prepared and sequence confirmed from a single reaction. Each TRA chain was prepared as a separate plasmid stock for subsequent characterisation.

Characterisation of diversified TRA chains with TRB chain pair

To characterise the specificity each of the 64 TCR TRA chains derived above were co-transfected with the TRB chain (SEQ0702) into a cell line that expresses all CD3 components, but lacks TRA and TRB expression. These cells were then stained with an antibody against CD3, and a HLA-A2*01- NLVPMVATV multimer reagent and analysed by flow cytometry. This analysis is presented in Figure 16.

Each transfectant population is expressed as the ratio of the mean fluorescence intensity (MFI) of HLA-A2*01- NLVPMVATV multimer signal of CD3 positive over the CD3 negative populations, and plotted as a scatter plot ranked on ascending ratio (Figure 16a). It is clear that the diversified TRA chains create a range of affinities to the HLA-A2*01- NLVPMVATV multimer. The original alpha chain is indicated with an arrow. The majority of diversified clones showed poorer binding than the original. A significant number of non-binders are also observed. Strikingly, a number of clones showed significant increase in HLA-A2*01- NLVPMVATV multimer staining (high-binders), including clones that improve the relative signal by over 3-fold. The clones are also presented as a table of ascending rank of MFI ratio, with the amino acids presented and diversified positions 1, 2 and 3 listed (figure 16b). It can be observed that a Pro at position 2 is highly stabilising of HLA-A2*01- NLVPMVATV multimer binding, and to a lesser degree an Asp at position 3. Conversely, a His at position 3 completely abolishes binding, except in the presence of a Pro at position 2. This positional bias of amino acid substitutions towards increased and decreased binding of the HLA-A2*01- NLVPMVATV mul-

timer strongly suggests a bona fide alteration in binding affinity upon targeted amino acid degeneracy created in the TRA chain using the TORES.

Overall, this example demonstrates that minimal diversity in CDR3 loop introduced in a single step by odeCDR3 degeneracy within the TORES can create a collection of TCRs with diverse binding characteristics towards analyte HLA-antigen complex. This can be directly incorporated into TCR maturation workflows to generate synthetic TCRs with altered characteristics within various TCR engineering scenarios, while still maintaining the full-length native context of the TCR chains along with native V,J and C gene segment usage.

Example 10

Rapid TCR chain diversification via V and J segment shuffling

The diversification and selection of TCR ORFs is desirable to engineer TCRs chain pairs with altered specificities, affinities and/or signalling capacity. The TORES system is suited to the rapid generation of collections of TCR chains that are systematically altered from the original target sequence. In the present example, a TCR chain diversification approach is outlined, wherein a single odeCDR3 is provided to a full-length TCR chain reaction with multiple V-C entry vectors, J donor vectors or multiples of both V-C entry vectors and J donor vectors. Such an approach can be used to diversify a given TCR chain using native germline V, J and C segments, while maintaining the possibility of carrying through core CDR3 contacts with cognate HLA-antigen complex in de novo TCR chains. This approach is ideally suited for rapid TCR-engineering using full-length TCR ORFs that may be presented and selected in a functional context of viable mammalian cells.

An example of shuffling V gene segment usage is presented in Figure 17. This schematic example outlines an approach of maintaining a single odeCDR3 and J donor vector usage, with sequence matching the parental TCR chain, while providing a selection of V gene segments via provision of a number of V-C entry vectors to the reconstitution reaction. In the case of TCR chains that have the opportunity of multiple C gene segment usage (e.g. human TRB), this can also be factored into the diversification workflow. The product of a reaction containing single odeCRD3 and J donor vector, and a selection of V-C entry vectors, will be a number of full-length TCR chains containing parental CDR3/J sequence, and each of the provided V (and/or C) gene segments. Such an approach is equally valid to achieve in discrete reactions with a number of se-

lected V-C entry vectors, rather than the pooled reaction depicted in Figure 17.

An example of shuffling J gene segment usage is presented in Figure 18. This schematic example outlines an approach of maintaining a single odeCDR3 and V-C entry vector usage, with sequence matching the parental TCR chain, while providing a selection of J gene segments via provision of a selection of J donor vectors to the reconstitution reaction. The product of a reaction containing single odeCRD3 and V-C entry vector, and a selection of J donor vectors, will be a number of full-length TCR chains containing parental CDR3/V-C sequence, and each of the provided J donor gene segments. Such an approach is equally valid to achieve in discrete reactions with a number of selected J donor vectors, rather than the pooled reaction depicted in Figure 18.

A final example of shuffling both V and J gene segment usage is presented in Figure 19. This schematic example outlines an approach of maintaining a single odeCDR3, with sequence matching the parental TCR chain, while providing a selection of both V(-C) and J gene segments via provision of a selection of both V-C and J donor vectors to the reconstitution reaction. The product of a reaction containing single odeCRD3, and a selection of both V-C entry and J donor vectors, will be a number of full-length TCR chains containing parental CDR3 sequence, and each combination of V(-C) and J donor gene segments provided to the reaction. Such an approach is equally valid to achieve in discrete reactions with a number of selected V-C entry and J donor vectors, rather than the pooled reaction depicted in Figure 19.

Example 11

Reconstitution of full length TCR ORFs from TRA and TRB using TORES^2

This example describes a TORES^2 for the human TRA/TRB loci with RMCE sites as 5' and 3' genetic elements, and demonstrates the reconstitution of a model human TCR pair, and subsequent integration of the reconstituted ORFs into a engineered cell line harbouring RMCE sites matched with the 5' and 3' genetic elements of the originating V-Cα entry vector backbone context. The model TCR TRA/TRB pair has known specificity for a HLA-A*02:01-restricted antigen. The model TRA and TRB sequences are represented by SEQ0778 and SEQ0779, respectively. The antigen peptide has the amino acid sequence SLLMWITQV.

V-C entry vectors, J donor vector, odeCDR3 and Bidirectional terminator for TORES^2

The V-Cα, V-Cβ, J donor vector and the oligonucleotide duplex encoding the odeCDR3

are selected using the same selection criteria as outlined in example 7. As described above, the TORES^2 requires adaptation of the V-Cα and V-Cβ to incorporate distinct Type IIS sites and negative selection elements. The V-Cα and V-Cβ entry vector backbone sequences are represented by SEQ7056 and SEQ0764, respectively. Due to the addition of new restriction sites to these backbone sequences, some of the V-segment sequences require modification to underlying nucleotide sequences compared to those presented for the TORES system above, as to eliminate these new restriction sites from the TCR coding sequence. Modified TRA V-C sequences are represented by SEQ0757 to SEQ0763, and modified TRB V-C sequences by SEQ0765 to SEQ0776. All other TRA and TRB sequences used were as described for the TORES system above.

Bidirectional terminator donor component

The Bidirectional Terminator Donor vector (BiT donor), provides the bidirectional terminator element introduced in the second step reaction, and which ultimately adjoins the antiparallel TRA and TRB chains reconstituted in the first step reaction. In the present example, the bidirectional terminator element is represented by SEQ0777, and is provided in a suitable vector backbone context.

Reconstitution reactions

The first restriction enzyme / ligase cycle reaction reaction to reconstitute the target TRA and TRB ORFs is conducted as is described in Example 7. This results in reconstituted TRA and TRB, each within their respective V-C entry backbone contexts. In the second reaction, these two product vectors are combined with the BiT donor vector in a new enzyme ligase cycle reaction to generate the final product vector, encoding the reconstituted TRA and TRB ORFs in an antiparallel sense, and adjoined by the introduced bidirectional terminator element.

<u>RE digestion and ligation reaction for TORES^2</u>

Reconstituted TRA ORF vector	25 ng
Reconstituted TRB ORF vector	25 ng
BiT donor vector	25 ng
10x T4 ligase buffer	2 µl
Esp3I	0.5 µl
T4 DNA ligase	0.5 µl
H_2O	up to 20 µl

Reaction conditions

Step 1; 2 min at 37°C

Step 2; 3 min at 16°C

Repeat step 1 and 2, 20 times

5 min at 50°C

5 min at 80°C

Return to room temperature

Add 0.5 µl of SalI and MluI enzymes and incubate for 30 min at 37°C

The resulting reaction product was transformed into competent E.coli cells and plated on carbenicillin containing plates.

Screening and sequencing carbenicillin resistant colonies was conducted to determine correctly assembled constructs. Screening of colonies was performed by restriction enzyme diagnostic digest of isolated plasmid DNA, and the expected DNA fragment sizes were observed by DNA electrophoresis. The resulting constructs encode the full length TCR alpha and beta clone sequences.

Validation of the reconstituted full length TCR ORF

To verify the functionality of the reconstituted bi-directional TRA/TRB donor vector generated by the TORES^2 system, the construct was delivered into a cell line that expresses all CD3 components but lacks TRA and TRB expression, and further encodes genomic receiver sites compatible with the RMCE sites contained within the product donor vector and which are contributed by the original V-Cα entry vector, and suitable antiparallel promoter elements at the 5' and 3' ends. The analysis is presented in Figure 23, wherein the reconstituted donor vector is transfected into cells. Cells that have lost the genomic receiver site markers (Figure 23 a, RFP- BFP-) are the cells that express the bidirectional TCR ORF, resulting in CD3 surface expression (Figure 23 b, light grey histogram). While the cells that did not lose the genomic receiver site markers are cells that failed to integrate the bidirectional TCR construct (Figure 23 a) RFP+BFP+) and do not stain for CD3 surface expression (Figure 23 b, dark grey histogram). TCR expression is required for surface presentation of the CD3 complex. All the RFP-BFP- cells that have lost the genomic receiver markers are positive for CD3 surface expression (Figure 23 b, light grey histogram) All the RFP+BFP+ cells that have

not lost the genomic receiver markers are negative for CD3 surface expression (Figure 23 b) dark grey histogram). These CD3+ cells expressing the reconstituted TCR also show positive staining for the HLA-A*02:01–SLLMWITQV tetramer reagent (Figure 23 c) compared to the cells that have not been stained with the said tetramer (Figure 23 d). Together the results indicate that the reconstituted TCR both presents on the cell surface and is specific for the target antigen loaded in the HLA-A*02:01 – SLLMWITQV tetramer reagent.

Specific embodiments

In the following is given more details regarding the invention and various aspects thereof.

1. A combined system comprising two separate components, wherein the first component is a vector carrying variable and constant (V-C) T-cell receptor (TCR) gene segments, and the second component is a vector carrying joining (J) TCR gene segments.

2. A combined system according to item 1 wherein the first component is a V-C entry vector containing
    a. origin of replication,
    b. a first positive selection marker
    c. 5' genetic element, or elements,
    d. Kozak Sequence,
    e. TCR variable gene segment,
    f. a first Type IIS sequence, for site specific recognition and cleavage by a Type IIS restriction enzyme,
    g. a negative selection marker,
    h. a second Type IIS sequence
    i. TCR constant gene segment, and
    j. 3' genetic element, or elements.

3. A combined system according to item 1 or 2 wherein the second component is a J donor vector containing
    a. origin of replication,
    b. a second positive selection marker,
    c. a third Type IIS sequence,

d. TCR Joining gene segment,
e. A C part, corresponding to a small 5' portion of a constant gene segment, and
f. a fourth Type IIS sequence.

4. A combined system according to item 2 or 3, wherein said first, second, third and fourth Type IIS sequences are the same or different.

5. A combined system according to item 2 or 3, wherein said first, second, third and fourth Type IIS sequences are the same.

6. A combined system according to any of items 2-5, wherein the first and second positive selection markers are different and are selected from an antibiotic resistance gene or auxotroph complementing gene.

7. A combined system according to any of items 2-6, wherein the 5' genetic element comprises one or more elements selected from
   a. gene cis/acting element,
   b. heterospecific recognition site for recombinase enzymes,
   c. a 5' homologous recombination arm for a genomic site of interest'
   d. a mRNA splice acceptor site,
   e. an internal ribosomal entry site, and
   f. epigenetic insulator sequence.

8. A combined system according to any of items 2-7, wherein the TCR variable gene segment does not contain a type IIS sequence contained in the first and second component.

9. A combined system according to any of items 2-8, wherein the first Type IIS sequence is orientated to cleave 5' of said recognition sequence and within the TCR variable gene segment.

10. A combined system according any of items 2-9, wherein the negative selection marker is selected from
    a. a restriction enzyme recognition site not contained elsewhere in the first component or within the TCR joining gene segment,

b. a bacterial suicide gene, and
c. a reporter element.

11. A combined system according items 2-10 wherein the second Type IIS sequence is orientated to cleave 3' of said recognition sequence and within the TCR constant gene segment.

12. A combined system according to any of items 2-11, wherein the TCR constant gene segment does not contain any of the type IIS sequences contained in the first and second component.

13. A combined system according to any of items 2-12, wherein the 3' genetic element comprises one or more elements selected from
a. a terminator element,
b. heterospecific recognition site for recombinase enzymes,
c. a 3' homologous recombination arm for a genomic site of interest,
d. a mRNA splice donor site,
e. an internal ribosomal entry site, and
f. epigenetic insulator sequence.

14. A combined system according to any of items 3-13, wherein the third Type IIS sequence is orientated to cleave 3‘ of said recognition sequence and within the TCR joining gene segment.

15. A combined system according to any of items 3-14, wherein the TCR joining gene segments do not contain any of the Type IIS sequences contained in the first and second component.

16. A combined system according to any of items 3-15, wherein the fourth Type IIS sequence is orientated to cleave 5‘ of said recognition sequence and within the TCR C part portion or the construct.

17. A combined system according to any of items 3-16, wherein the TCR C part does not contain any of the type IIS sequences contained in the first and second component.

18. A combined system according to any of the preceding items, wherein all contained sequences are devoid of any negative selection element apart from the negative selection marker as defined in item 10.

19. A combined system according to any of the preceding items further comprising a third component comprising an oligonucleotide duplex encoding CDR3 (odeCDR3).

20. A combined system according to item 19, wherein the odeCDR3 has
    a. a first single strand overhang sequence complimentary to first Type IIS restriction enzyme recognition and cleavage site orientated to cleave 5' of the recognition sequence and within the TCR variable gene segment,
    b. a double strand segment encoding a TCR CDR3 region and devoid of negative selection element, which negative selection element is as defined in item 10, and also devoid of any Type IIS restriction sequences of the first or second part, and
    c. a second single strand overhang sequence complimentary to the third Type IIS restriction enzyme recognition and cleavage site orientated to cleave 3' of the recognition sequence and within the TCR joining gene segment.

21. A library of V-C entry vectors as defined in any of items 1-18, wherein the library is a collection of one or more vectors representing all germline TCR variable and constant gene segments of an organism having such TCRs.

22. A library of J donor vectors as defined in any of items 1-18, wherein the library is a collection of one or more vectors representing germline TCR joining gene segments of an organism having such TCRs.

23. A library of V-C entry vectors as defined in any of items 1-18, wherein the library is a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segment.

24. A library of J donor vectors as defined in any of items 118, wherein the library is a collection of one or more vectors representing variant germline TCR joining gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segment.

25. A library of V-C entry vectors as defined in any of items 1-18, wherein the library is a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segment.

26. A library of J donor vectors as defined in any of items 1, 3-18, wherein the library is a collection of one or more vectors representing variant germline TCR joining gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segment.

27. A library comprising a combination of V-C entry vectors and J donor vectors as defined in any of items 1-26.

28. A kit comprising a combination of
    a. one or more V-C entry vectors encoding combinations of Variable and Constant gene segments, and
    b. one or more J donor vectors encoding J gene segments , and optionally
    c. one or more standardised odeCDR3 with single stranded overhangs matched to V-C entry vector and J donor vector single strand overhangs as positive control odeCDR3, and optionally
    d. A pre-assembled full-length TCR ORF as a reference

29. A method for in vitro reconstitution of a full length TCR open reading frame (ORF), said method comprising
    a. selecting a V-C entry vector,
    b. selecting a J donor vector,
    c. selecting an odeCDR3,

d. combining a, b and c to react with i) Type IIS restriction enzyme(s) to cleave all TypeIIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme and subjecting the combined mix to a thermocycling reaction,
e. transforming the reaction product obtained from step d. to a selectable host organism competent for DNA vector propagation, and
f. performing a selection of host organism to obtain full length reconstituted TCR open reading frame in the V-C entry vector backbone.

30. A method according to item 29, wherein the V-C entry vector is as defined in item 2, the J donor vector is as defined in item 3, and step f of item 29 is performed by using the first positive selection marker as defined in item 6 on the transformed host cells.

31. A method according to items 29 or 30, wherein a selected TCR open reading frame is reconstituted from a selected TCR sequence, said method comprising
a. Obtaining a TCR open reading frame sequence wherein said sequence information is sufficient to identify i) variable gene segment usage ii) constant gene segment usage iii) joining gene segment usage iv) a full CDR3 sequence spanning the variable gene segment border to the joining gene segment border, and
b. selecting a V-C entry vector corresponding to the variable and constant gene segments identified in step a. i) and a. ii), respectively, and
c. selecting a J donor vector corresponding to the joining gene segment identified in step a. iii), and
d. generating an odeCDR3 corresponding to CDR3 sequence identified in step a. iv, and
e. combining b, c and d to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme and subjecting the combined mix to a thermocycling reaction, and
f. transforming the reaction product obtained from step d. to a selectable host organism competent for vector replication, and
g. performing a selection of host organism to obtain full length reconstitut-

ed TCR open reading frame in the V-C entry vector backbone.

32. A method according to any of items 29-31, wherein step d. i) is performed using a single Type IIS restriction enzyme.

33. A method according to any of items 29-31, wherein step d. i) is performed using at least two Typ eIIS restriction enzymes.

34. A method according to any of items 29-33 further comprising at least one step of negative selection comprising
    a. performing restriction enzyme digest of reaction product to eliminate parental V-C entry vector prior to host transformation prior to step e. of item 29 or step f. of item 31, and/or
    b. performing a suicide gene selection to eliminate competent hosts transformed with parental V-C entry vector, and/or
    c. performing selection of host cells transformed with parental V-C entry vector by way of reporter identification.

35. A method according to any of items 29-34, wherein selection of host organisms comprises selecting a host organism having resistance to antibiotic present in the culture system or competition of auxotrophic factors not present in the culture system, and said antibiotic resistance or auxotrophic complementation being conferred by a gene product encoded in the V-C entry vector.

36. A method according to any of items 29-34, wherein one or more steps are performed in a culture system suitable for culture and selection of the host organism.

37. A method according to any of items 29 to 36 used to create a pool of two or more TCR ORFs with diversified CDR3 region, wherein two or more odeCDR3 forms are selected according to item 29 c, and combined to a single restriction enzyme / ligase cycle reaction according to item 29 d.

38. A method according to any of items 29 to 36 used to create a pool of two or more TCR ORFs with diversified V and/or C gene segment usage, wherein two or more V-C entry vectors are selected according to item 29 a, and combined to

a single restriction enzyme / ligase cycle reaction according to item 29 d.

39. A method according to any of items 29 to 36 used to create a pool of two or more TCR ORFs with diversified J gene segment usage, wherein two or more J donor vectors are selected according to item 29 b, and combined to a single restriction enzyme / ligase cycle reaction according to item 29 d.

40. A method according to any of items 29 to 36 used to create a pool of two or more TCR ORFs with diversified V and/or C and J gene segment usage, wherein two or more V-C entry and J donor vectors are selected according to item 29 a and b, and combined to a single restriction enzyme / ligase cycle reaction according to item 29 d.

41. A method according to any of items 29 to 36 used to create a pool of two or more TCR ORFs with diversified CDR3 in addition to diversified V and/or C and J gene segment usage, wherein two or more odeCDR3 are selected according to item 29 c, and two or more V-C entry and J donor vectors are selected according to item 29 a and b, and combined to a single restriction enzyme / ligase cycle reaction according to item 29 d.

42. A method to construct a V-C entry vector according to items 1 and 2, wherein the method comprises combining three DNA components selected from
    a. a Variable gene segment cloning fragment
    b. a Constant gene segment cloning fragment
    c. a V-C entry vector backbone

43. A method according to item 42 wherein said Variable gene segment cloning fragment comprises
    a. a 5' primer bind sequence for polymerase chain reaction dependent propagation of the fragment
    b. a fifth Type IIS sequence orientated to cut in the 3' direction
    c. a first overhang sequence that encodes a defined single stranded overhang upon Type IIs enzyme action on the fifth Type IIs sequence in b
    d. a Kozak sequence
    e. a TCR variable gene segment
    f. a first Type IIS sequence

g. a 5' sequence segment of a negative selection marker
h. a sixth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated within the 5' sequence segment of the negative selection marker in g
i. a 3' primer bind sequence for polymerase chain reaction dependent propagation of the fragment

44. A method according to item 42 or 43, wherein said Constant gene segment cloning fragment comprises
a. a 5' primer bind sequence for polymerase chain reaction dependent propagation of the fragment
b. a seventh Type IIS sequence orientated to cut in the 3' direction such that a single stranded overhang is generated within the 3' sequence segment of the negative selection marker in c
c. a 3' sequence segment of a negative selection marker
d. a second type IIS sequence
e. a TCR constant gene segment
f. a second overhang sequence that encodes a defined single stranded overhang upon Type IIs enzyme action on the eighth Type IIs sequence in g
g. an eighth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated in overhang sequence of f
h. a 3' primer bind sequence for polymerase chain reaction dependent propagation of the fragment

45. A method according to any of items 42-44 wherein said V-C entry vector backbone comprises
a. an origin or replication
b. a first positive selection marker
c. a 5' genetic element
d. a first restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to said first overhang mentioned in item 43 c
e. a second restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to said second overhang mentioned in item 44 f

f. a 3' genetic element.

46. A method according to any of items 42 to 45, wherein the fifth, sixth, seventh and eighth Type IIS sequences are the same or different.

47. A method according to any of items 42 to 45, wherein the fifth, sixth, seventh and eighth Type IIS sequences are the same.

48. A method according to any of items 42 to 45, wherein the fifth, sixth, seventh and eighth Type IIS sequences are different from the first, second third and fourth Type IIS sequences as defined in items 2 to 5.

49. A method according to any of items 42 to 48 comprising
    a. Digestion of the V-C entry vector backbone with said first restriction enzyme and said second restriction enzyme to create said first and second overhang, respectively
    b. Combining the digested V-C entry vector backbone with the V cloning fragment and the C cloning fragment, along with DNA ligase enzyme and one or more Type IIS restriction enzyme(s) recognising the fifth, sixth, seventh and eighth Type IIS sequences
    c. Transformation of the resulting reaction product into competent host organism and positive selection using said first positive selection marker to obtain complete V-C entry vector

50. A Variable gene segment cloning fragment represented by the sequence SEQ0690.

51. A Constant gene segment cloning fragment represented by the sequence SEQ0691.

52. A V-C entry vector represented by sequence SEQ0048, to construct V-C entry vectors suitable for transient expression or reconstituted full-length TCR open reading frames in a mammalian cells.

53. A V-C entry vector represented by sequence SEQ0688, to construct V-C entry vectors suitable for recombinase mediated cassette exchange with matched

genetic targets with suitable heterospecific recombinase sequences.

54. A V-C entry vector backbone represented by sequence SEQ0689, to construct V-C entry vectors suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences.

V-C entry vector generics

55. A V-C entry vector represented by sequence SEQ0692, suitable for transient expression or reconstituted full-length TCR open reading frames in a mammalian host cell.

56. A V-C entry represented by sequence SEQ0693 suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences.

57. A V-C entry vector represented by sequence SEQ0694, suitable for recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences.

58. A method to construct J donor vector according to items 1 or 3, wherein the method comprises combining four DNA components selected from
    a. J receiving cassette fragment
    b. J donor vector backbone
    c. J receiving cassette vector
    d. J segment part

59. A method according to item 58 wherein said J receiving cassette fragment comprises
    a. a first single stranded overhang at the 5' end commentary to overhang sequence mentioned in 60 c
    b. a third Type IIS sequence orientated to cut in the 3' direction, joined with a sequence that forms a single stranded overhang when acted upon by the enzyme directed by a ninth Type IIS sequence mentioned in c
    c. a ninth Type IIs sequence orientated to cut in the 5' direction, and to create a single stranded overhang mentioned in b

d. a negative selection marker
e. a tenth Type IIS sequence orientated to cut in the 3' direction and create a single stranded overhang at the 5' of the sequence described in f
f. a C-part representing a 5' portion of the Constant gene fragment, with an overhang sequence at the 5' end generated by enzyme action directed by the tenth Type IIS sequence, and an overhang sequence at the 3' end generated by enzyme action directed by the fourth Type IIS sequence mentioned in g
g. a fourth Type IIS sequence orientated to cut in the 5' direction such that a single stranded overhang is generated within the 5' sequence containing the C-part mentioned in f
h. a second single stranded overhang at the 3' end commentary to overhang sequence mentioned in 60 d

60. A method according to item 58 or 59 wherein said J donor vector backbone comprises
a. an origin or replication
b. a second positive selection marker
c. a first restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to said first overhang mentioned in item 59 a
d. a second restriction enzyme recognition sequence permitting digestion of the backbone to create a single stranded overhang complimentary to said second overhang mentioned in item 59 h

61. A method according to any of items 58 to 60 wherein said J receiving cassette vector in constructed by combining the TRAJ receiving cassette fragment described in item 53 and the TRAJ J donor vector backbone described in item 54, wherein the method comprises
a. Digestion of the J donor vector backbone with said first restriction enzyme and said second restriction enzyme to create said first and second overhang, respectively
b. Combining the digested J donor vector backbone with the J receiving cassette fragment
c. Transformation of the resulting reaction product into competent host organism and positive selection using said the second positive selection

marker to obtain complete J receiving cassette vector

62. A method according to any of items 58 to 61 wherein said J segment part comprises
    a. A first single stranded overhang sequence at the 5' end that is complementary to the overhang generated in the J receiving cassette fragment mentioned in item 59 b, when acted on by the enzyme directed by the ninth Type IIS sequence mentioned in item 59 b
    b. A Joining gene segment part
    c. A second single stranded overhang sequence at the 3' end that is complementary to the overhang generated in the J receiving cassette fragment mentioned in item 59 f, when acted on by the enzyme directed by the ninth Type IIS sequence mentioned in item 59 e

63. A method according to any of items 58 to 62, wherein the ninth and tenth Type IIS sequences are the same or different.

64. A method according to any of items 58 to 62, wherein the ninth and tenth Type IIS sequences are the same.

65. A method according to any of items 42 to 62, wherein the ninth and tenth Type IIS sequences are different from the first, second third and fourth Type IIS sequences as defined in items 2 to 5.

66. A method according to any of items 42 to 62, wherein the ninth and tenth Type IIS sequences are the same or different from the fifth, sixth, seventh and eighth Type IIS sequences as defined in items 2 to 5.

67. A method according to any of items 42 to 62, wherein the ninth and tenth Type IIS sequences are the same as the fifth, sixth, seventh and eighth Type IIS sequences as defined in items 2 to 5.

68. A method according to any of items 42 to 62 wherein said J donor vector is constructed be combining a J donor vector backbone with a J segment part, wherin the method comprises
    a. Combining J receiving cassette vector with the J segment part, along

with DNA ligase enzyme and one or more Type IIS restriction enzyme(s) recognising the ninth and tenth Type IIS sequences and subjecting the combined mix to a thermocycling reaction

b. Transformation of the resulting reaction product into competent host organism and positive selection using said second positive selection marker to obtain complete J donor vector

69. A J receiving cassette fragment represented by the sequences SEQ0695 and SEQ0696.

70. A J donor vector backbone represented by sequence SEQ0097.

71. A J receiving cassette vector represented by SEQ0697, constructed by the combination of the J receiving cassette fragment mentioned in item 69 and the J donor vector backbone mentioned in item 70.

72. A J segment part represented by sequences SEQ0698 and SEQ0699, to construct a J donor vector by insertion of the J segment part into a J receiving cassette vector mentioned in item 71.

73. A J donor vector represented by the sequence SEQ0700, assembled from J receiving cassette vector mentioned in item 71 and the J segment part mentioned in item 72.

74. A library of V-C entry vectors as defined in any items 2-19, containing Variable and Constant gene segments to recapitulate the gene segment usage of the human TRA locus, represented by sequences SEQ0049 to SEQ0094, wherein the vectors are suited for transient expression of a reconstituted full-length TCR open reading frame.

75. A library of J donor vectors as defined in any of items 3-19, containing Joining gene segments to recapitulate the gene segment usage of the human TRA locus, represented by sequences SEQ0323 to SEQ0378, wherein the vectors are used in conjunction with an odeCDR3 spanning the entire CDR3 region.

76. A library of J donor vectors as defined in any of items 3-, containing Joining

gene segments to recapitulate the gene segment usage of the human TRA locus, represented by sequences SEQ0379 to SEQ0434, wherein the vectors are used in conjunction with an odeCDR3 reduced in length by three to four codons.

77. A library of V-C entry vectors as defined in any of items 2-19 containing Variable and Constant gene segments to recapitulate the gene segment usage of the human TRB locus, represented by sequences SEQ0484 to SEQ0577, wherein the vectors are suited for transient expression of a reconstituted full-length TCR open reading frame.

78. A library of J donor vectors as defined in any of items 3-19 containing Joining gene segments to recapitulate the gene segment usage of the human TRB locus, represented by sequences SEQ0636 to SEQ0661, wherein the vectors are used in conjunction with an odeCDR3 spanning the entire CDR3 region.

79. A library of J donor vectors as defined in any of items 3-19 containing Joining gene segments to recapitulate the gene segment usage of the human TRB locus, represented by sequences SEQ0662 to SEQ0687, wherein the vectors are used in conjunction with an odeCDR3 reduced in length by three to four codons.

80. A combined system according to items 1 to 20, for use in providing native germline and/or diversified synthetic TCR open reading frames in defined vector contexts.

81. TCR open reading frames in defined vector contexts for use in diagnostic or therapeutic procedures.

Claims

1. A combined system comprising two separate components, wherein a first component is a vector carrying variable and constant (V-C) T-cell receptor (TCR) gene segments, and a second component is a vector carrying joining (J) TCR gene segments,

wherein the first component is a V-C entry vector containing

a. origin of replication,
b. a first positive selection marker
c. 5' genetic element, or elements,
d. Kozak Sequence,
e. TCR variable gene segment,
f. a first Type IIS sequence, for site specific recognition and cleavage by a first Type IIS restriction enzyme, wherein the sequence is orientated such that the enzyme cleaves 5' of said recognition sequence and at the 3' end of the TCR variable gene segment,
g. a negative selection marker,
h. a second Type IIS sequence for site specific recognition and cleavage by a second Type IIS restriction enzyme, wherein the sequence is orientated such that the enzyme cleaves to direct cleavage 3' of said recognition sequence and at the 5' end of the TCR constant gene segment,.
i. TCR constant gene segment, and
j. 3' genetic element, or elements,

wherein the second component is a J donor vector containing

a. origin of replication,
b. a second positive selection marker,
c. a third Type IIS sequence for site specific recognition and cleavage by a third Type IIS restriction enzyme, wherein the sequence is orientated such that the enzyme cleaves 3' of said recognition sequence and at the 5' end of the TCR joining gene segment,
d. TCR Joining gene segment,
e. A C-part, corresponding to a small 5' portion of a constant gene segment, and
f. a fourth Type IIS sequence for site specific recognition and cleavage by fourth Type IIS restriction enzyme, wherein the sequence is orientated such that the enzyme cleaves 5‘ of said recognition sequence and at the

3' end TCR C-part portion of the construct, such that the single strand overhang sequence generated is complimentary with that generated by said second Type IIS sequence.

2. A combined system according to claim 1, wherein the 5' genetic element comprises one or more elements selected from

a. gene cis/acting element,
b. heterospecific recognition site for recombinase enzymes,
c. a 5' homologous recombination arm for a genomic site of interest,
d. a mRNA splice acceptor site,
e. an internal ribosomal entry site, and
f. an epigenetic insulator sequence.

3. A combined system according any of claims 1-2, wherein the negative selection marker is selected from

a. a restriction enzyme recognition site not contained elsewhere in the first component or within the TCR joining gene segment,
b. a bacterial suicide gene, and
c. a reporter element.

4. A combined system according to any of claims 1-3, wherein the 3' genetic element comprises one or more elements selected from

a. a terminator element,
b. heterospecific recognition site for recombinase enzymes,
c. a 3' homologous recombination arm for a genomic site of interest,
d. a mRNA splice donor site,
e. an internal ribosomal entry site, and
f. an epigenetic insulator sequence.

5. A combined system according to any of claims 1-4, wherein the components do not contain any Type IIS sequences further than those defined above, and only contain said negative selection marker.

6. A combined system according to any of the preceding claims further comprising a third component comprising an oligonucleotide duplex encoding CDR3 (odeCDR3), wherein the odeCDR3 has

a. a first single strand overhang sequence complimentary to that generated by cleavage of the first Type IIS restriction site at the 3' end of TCR variable gene segment,
b. a double strand segment encoding a TCR CDR3 region and devoid of negative selection marker, which negative selection marker is as defined in claim 5, and said double strand element is also devoid of any Type IIS restriction sequences of the first or second component, and
c. a second single strand overhang sequence complimentary to that generated by cleavage of the third Type IIS restriction enzyme site at the 5' end of TCR joining gene segment.

7. A combined system according to any of claims 1-5 further comprising a third component comprising an oligonucleotide duplex encoding CDR3 (odeCDR3), wherein the odeCDR3 is a dsDNA molecule or plasmid DNA encoding the CDR3 flanked by two Type IIS recognition and cleavage sites, such that cleavage by said enzymes results in single strand overhangs complimentary to those generated by the first Type IIS cleavage in the V-C entry vector at the 5' end, and with the third Type IIS cleavage of the J donor vector at the 3' end, to obtain a digested product of said odeCDR3 comprising a, b and c as described in claim 6.

8. A combined system according to any of claims 1-7, wherein the first to fourth Type IIS sequences are the same or different.

9. A two-part vector library comprising a library of a V-C entry vectors as defined in claim 1 and a library of J donor vectors as defined in claim 1, wherein the library of V-C entry vectors is

a collection of one or more vectors representing all germline TCR variable and constant gene segments of an organism having such TCRs, or

a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segment, or

a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segment and wherein

the J donor vector library is a collection of one or more vectors representing germline TCR joining gene segments of an organism having such TCRs, or

a collection of one or more vectors representing variant germline TCR joining gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segment, or

a collection of one or more vectors representing variant germline TCR joining gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segment.

10. A method for in vitro reconstitution of a full length TCR open reading frame (ORF), said method comprising

a. selecting a V-C entry vector,
b. selecting a J donor vector,
c. selecting an odeCDR3,
d. combining a, b and c to react with i) Type IIS restriction enzyme(s) to cleave all Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vector and in the J donor vector and ii) DNA ligase enzyme and subjecting the combined mix to a thermocycling reaction,
e. transforming the reaction product obtained from step d. into a host organism competent for DNA vector propagation, and
f. propagating said host organism to obtain full length reconstituted TCR open reading frame in the resulting V-C entry vector backbone.

11. A V-C entry vector selected from
sequence SEQ0692,

sequence SEQ0693, and
sequence SEQ0694.

12. A J donor vector represented by the sequence SEQ0700.

13. A library of V-C entry vectors as defined in any of claims 1-8 for use in combination with a library of J donor vectors and an odeCDR3 to reconstitute a full-length TCR ORF, wherein the V-C entry vectors contain Variable and Constant gene segments to recapitulate the gene segment usage of the human TRA locus, represented by the sequences SEQ0049 to SEQ0094; and/or the V-C entry vectors contain Variable and Constant gene segments to recapitulate the gene segment usage of the human TRB locus, represented by the sequences SEQ0484 to SEQ0577.

14. A library of J donor vectors as defined in any of claims 1-8, containing Joining gene segments to recapitulate the gene segment usage of the human TRA locus, represented by sequences SEQ0323 to SEQ0378, wherein the vectors are used in conjunction with an odeCDR3 spanning the entire CDR3 region; and/or
containing Joining gene segments to recapitulate the gene segment usage of the human TRA locus represented by sequences SEQ0379 to SEQ0434, wherein the vectors are used in conjunction with an odeCDR3 reduced in length by three to four codons.

15. A library of J donor vectors as defined in any of claims 1-8, containing Joining gene segments to recapitulate the gene segment usage of the human TRB locus, represented by sequences SEQ0636 to SEQ0661, wherein the vectors are used in conjunction with an odeCDR3 spanning the entire CDR3 region, and/or
containing Joining gene segments to recapitulate the gene segment usage of the human TRB locus, represented by sequences SEQ0662 to SEQ0687, wherein the vectors are used in conjunction with an odeCDR3 reduced in length by three to four codons.

16. A library according to claim 14 or 15 for use in combination with a V-C entry vector and an odeCDR3 to reconstitute a full-length TCR ORF.

17. A combined system according to any of claims 1-8, wherein said first component is a V-C entry vector encoding a first TCR chain, the system further comprises a component comprising a Bidirectional Terminator (BiT) donor vector, and a component comprising a V-C entry vector encoding a second TCR chain complimentary to said first TCR chain.

18. A combined system according to claim 17, wherein said V-C entry vector of said first componentfurther comprises

a. a fifth Type IIS sequence for site specific recognition and cleavage by a fifth Type IIS restriction enzyme, wherein the sequence is oriented such that the enzyme cleaves 5' of said recognition sequence, and at the 3' end of the constant gene segment,
b. a negative selection marker located 3' of the fifth Type IIS sequence,
c. a sixth Type IIS sequence for site specific recognition and cleavage by a sixth Type IIS restriction enzyme located 3' of the negative selection marker of b., wherein the sequence is oriented such that the enzyme cleaves 3' of said recognition sequence, and to the 5' of the 3' genetic element(s).

19. A combined system according to claim 17 or 18, wherein said V-C entry vector encoding a second TCR chain complimentary to said first chain and further comprises

a. a seventh Type IIS sequence for site specific recognition and cleavage by a seventh Type IIS restriction enzyme, wherein the sequence is oriented such that the enzyme cleaves 3' of said recognition sequence and at the 5' of the Kozak sequence,
b. an eighth Type IIS sequence for site specific recognition and cleavage by an eighth Type IIS restriction enzyme, wherein the sequence is oriented such that the enzyme cleaves 5' of said recognition sequence, and at the 3' end of the constant gene segment,
c. a negative selection marker located at the 3' of the eighth Type IIS sequence.

20. A combined system according to any of claims 17-19, wherein said BiT donor vector contains

an origin of replication,

a. a positive selection marker,
b. a ninth Type IIS sequence for site specific recognition and cleavage by a ninth Type IIS restriction enzyme, wherein the sequence is oriented such that the enzyme cleaves 3' of said recognition sequence and the 5' end of a bidirectional terminator element, d,
c. a bidirectional terminator element encoding two transcriptional terminators in antisense arrangement, and
d. a tenth Type IIS sequence for site specific recognition and cleavage by a tenth Type IIS restriction enzyme, wherein the sequence is oriented such that the enzyme cleaves 5' of said recognition sequence and the 3' of the bidirectional terminator element, d.

21. A combined system according to any of claims 17-20, wherein the first to fourth Type IIS sequences are the same or different.

22. A combined system according to any one of claims 17-20, wherein the fifth to tenth Type IIS sequences are the same or different, and are different from those of the first to fourth Type IIS sequences.

23. A combined system according to any one of claims 17-22 wherein the negative selection marker is selected from

a. a restriction enzyme recognition site not contained elsewhere in the first component or within the TCR joining gene segment,
b. a bacterial suicide gene, and
c. a reporter element,

and wherein the negative selection markers defined in b. of claim 18 and in c. of claim 19 are different from the negative selection marker defined in g. of claim 1, and the negative selection marker defined in g. of claim 1 is different from those in the V-C entry vector of the first component and the V-C entry vector encoding saida second TCR chain complimentary to said first TCR chain.

24. A library of V-C entry vectors as defined in any of claims 17-23, wherein the library comprises a first and second V-C entry vector library, each containing

a collection of one or more vectors representing all germline TCR variable and constant gene segments of an organism having such TCRs, or

a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is unmodified in relation to the protein sequence encoded by the germline gene segment, or

a collection of one or more vectors representing a collection of variant germline TCR variable and constant gene segments of an organism having such TCRs, such that translated amino acid sequence of the encoded protein is modified in relation to the protein sequence encoded by the germline gene segment,

wherein the first and second V-C entry vector libraries encode reciprocal TCR chains.

25. A library of V-C entry vectors according to claim 24 for use to provide a pair of full length TCR open reading frames (ORFs), encoded in a antiparallel arrangement in a single product construct.

26. A method for in vitro reconstitution of a pair of full length TCR open reading frames (ORFs), encoded in an antiparallel arrangement in a single product construct, said method comprising

a. selecting a V-C entry vector for each TCR chain,
b. selecting a J donor vector for each TCR chain,
c. selecting an odeCDR3 for each TCR chain,
d. combining a, b and c for each TCR in an independent reaction to react with i) Type IIS restriction enzyme(s) to cleave the first to fourth Type IIS restriction enzyme recognition and cleavage sites present in the V-C entry vectors and in the J donor vectors and ii) DNA ligase enzyme and subjecting the combined mixes to thermocycling reactions to obtain a first and second reaction product,
e. combining both reaction products from step d. with a BiT donor vector as defined in claim 17 with i) Type IIS restriction enzyme(s) to cleave the fifth to tenth Type IIS restriction enzyme recognition and cleavage sites and ii) DNA ligase enzyme, and subjecting the combined reaction prod-

ucts to thermocycling reactions to obtain a third reaction product,

f. transforming the third reaction product from step e. into a host organism competent for DNA vector propagation, and

g. propagating said host organism to obtain full length reconstituted TCR open reading frame in the resulting V-C entry vector backbone.

27. A pair of V-C entry vectors represented by SEQ ID NO: 0756 and 0764.

28. A pair of V-C entry vectors according to claim 27, for use in recombinase mediated cassette exchange with genetic targets with matched heterospecific recombinase sequences.

29. A bidirectional terminator element, represented by the sequence SEQ0777.