COMPOSITIONS AND METHODS FOR T-CELL RECEPTOR GENE ASSEMBLY

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/829,813, filed April 5, 2019, U.S. Provisional Patent Application No. 62/838,465, filed April 25, 2019, U.S. Provisional Patent Application No. 62/898,053, filed September 10, 2019, and U.S. Provisional Patent Application No. 62/972,231, filed February 10, 2020, each of which is entirely incorporated herein by reference.

SEQUENCE LISTING

[0001.1] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on May 15, 2020, is named 53563-705_601_SL.txt and is 69,187 bytes in size.

BACKGROUND OF THE INVENTION

[0002] The T-cell receptor (TCR) can be responsible for the recognition of the antigen-major histocompatibility complex, leading to the initiation of an inflammatory response. Many T cell subsets exist, including cytotoxic T cells and helper T cells. Cytotoxic T cells (also known as CD8+ T cells) kill abnormal cells, for example virus-infected or tumor cells. Helper T cells (also known as CD4+ T cells) aid in the activation and maturation of other immune cells. Both cytotoxic and helper T cells carry out their function subsequent to the recognition of specific target antigens which triggers their respective responses. The antigen specificity of a T cell can be defined by the TCR expressed on the surface of the T cell. T-cell receptors are heterodimer proteins composed of two polypeptide chains, most commonly an alpha chain and a beta chain, but a minority of T cells can express a gamma and delta chain. The specific amino acid sequence of the TCR and the resultant three-dimensional structure defines the TCR antigen specificity and affinity. The amino acid and coding DNA sequences of the TCR chains for any individual T cell are almost always unique or at very low abundance in an organism's entire TCR repertoire, since there are a vast number of possible TCR sequences. This large sequence diversity can be achieved during T cell development through a number of cellular mechanisms and may be a critical aspect of the immune system's ability to respond to a huge variety of potential antigens.

[0003] Analyzing the TCR repertoire may help to gain a better understanding of the immune system features and of the aetiology and progression of diseases, in particular those with

unknown antigenic triggers. The extreme diversity of the TCR repertoire and the bipartite nature of TCRs can represent a major analytical challenge. High-throughput sequencing can allow greater sequencing depth and significantly more accurate quantification of TCR clonotype abundance, albeit at a greater expense than spectratyping.

SUMMARY OF THE DISCLOSURE

[0004] Provided herein are compositions and methods to assemble nucleic acid sequences encoding natively paired T-cell receptors (TCRs) (or cognate TCR pairs). For example, a TCR can comprise a TCR alpha chain and a TCR beta chain or a TCR can comprise a TCR gamma chain and a TCR delta chain. Sequences encoding natively paired TCRs can be identified using various methods, including but not limited to using single cell barcoding and sequencing technologies. After obtaining the sequences encoding natively paired TCRs, compositions and methods described herein can be used to construct or assemble one or more nucleic acid sequences to express the natively paired TCRs in any given host cell(s) in a quick, high-throughput and cost-effective manner. The one or more nucleic acid sequences can comprise greater than or equal to about 1, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences encoding different TCRs.

[0004a] In one aspect, the present disclosure provides a method for generating a plurality of nucleic acid molecules encoding a plurality of different T-cell receptor (TCR) alpha chains or portion thereof in a same compartment, the method comprising:

(a) providing a first plurality of nucleic acid molecules, each nucleic acid molecule of the first plurality of nucleic acid molecules comprising a sequence encoding a CDR3 alpha (CDR3α) of a different TCR alpha chain;

(b) providing a second plurality of nucleic acid molecules, each nucleic acid molecule of the second plurality of nucleic acid molecules comprising a sequence from a TCR alpha variable (TRAV) gene, wherein the second plurality of nucleic acid molecules comprises at least two different sequences from at least two different TRAV genes; and

generating a third plurality of nucleic acid molecules, each nucleic acid molecule of the third plurality comprising a sequence from one of the sequences encoding a CDR3α from the first plurality of nucleic acid molecules of (a) and a sequence from one of the at least two different TRAV genes of the second plurality of nucleic acid molecules of (b), wherein generating comprises contacting the first plurality of nucleic acid molecules of (a) to the second plurality

of nucleic acid molecules of (b) in a same compartment, wherein a nucleic acid molecule of the first plurality of nucleic acid molecules of (a) comprises a connector sequence that is capable of specifically linking to an anti-connector sequence of a nucleic acid molecule of the second plurality of nucleic acid molecules of (b) that comprises a sequence from a single given TRAV gene of the at least two different TRAV genes, thereby generating the plurality of nucleic acid molecules encoding the plurality of different TCR alpha chains or portion thereof, and wherein the single given TRAV gene is cognate to the CDR3α to which it is linked.

[0005] In an example, the present disclosure provides a method for generating a nucleic acid molecule encoding a T-cell receptor (TCR) chain or portion thereof, comprising: (a) providing at least one nucleic acid molecule comprising a sequence encoding a CDR3 of a TCR chain; (b) providing a plurality of nucleic acid molecules, each nucleic acid molecule of the plurality comprising a sequence derived from a TCR V gene, wherein the plurality of nucleic acid molecules comprises at least two different sequences derived from at least two different TCR V genes; and (c) contacting the at least one nucleic acid molecule of (a) to the plurality of nucleic acid molecules of (b) in a same compartment, wherein the at least one nucleic acid molecule of (a) is capable of linking to a nucleic acid molecule of the plurality of nucleic acid molecules to generate a third nucleic acid molecule comprising the sequence encoding the CDR3 and a sequence derived from one of the at least two different TCR V genes, thereby generating the nucleic acid molecule encoding the TCR chain or portion thereof. In some examples, the at least one nucleic acid molecule comprises at least about 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences. In some examples, the plurality of nucleic acid molecules, each nucleic acid molecule of the plurality comprising a sequence derived from a TCR V gene, comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80 or more different sequences derived from different TCR V genes.

[0006] In some examples, the at least one nucleic acid molecule comprises a first plurality of nucleic acid molecules, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a CDR3 of a TCR chain. In some examples, the at least one nucleic acid molecule of (a) is capable of specifically linking to a nucleic acid

molecule of the plurality of nucleic acid molecules that comprises a sequence derived from any single given TCR V gene of the at least two different TCR V genes. In some examples, the at least one nucleic acid molecule further comprises a J region of the TCR chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a J region of a TCR chain. In some examples, the at least two TCR V genes are human TCR V genes or mouse TCR V genes. In some examples, the at least two TCR V genes are selected from the group consisting of a human TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, and TRAV41. In some examples, the at least two TCR V genes are selected from the group consisting of a human TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, and TRBV30. In some examples, each sequence of the plurality of sequences derived from the at least two different TCR V genes comprises a sequence encoding L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, and/or FR3. In some examples, the TCR chain is a TCR alpha chain, a TCR beta chain, a TCR gamma chain, or a TCR delta chain. In some examples, the at least one nucleic acid molecule further comprises an additional sequence encoding an additional CDR3 of an additional TCR chain. In some examples, the at least one nucleic acid molecule comprises an additional J region of the additional TCR chain. In some examples, the sequence encoding the CDR3 and the additional sequence encoding the additional CDR3 are separated by at most 100 nucleotides. In some examples, the TCR chain and the additional TCR chain are a cognate pair of TCR chains. In some examples, the at least one nucleic acid molecule comprises a connector sequence, which connector sequence is capable of linking the at least one nucleic acid molecule to the nucleic acid molecule of the plurality of nucleic acid molecules to generate the third nucleic acid molecule. In some examples, the at least one nucleic acid molecule and the nucleic acid molecule of the plurality of nucleic acid molecules encodes a functional TCR chain or portion thereof. In some

examples, the nucleic acid molecule of the plurality of nucleic acid molecules comprises an anti-connector sequence, which anti-connector sequence is complementary to the connector sequence of the at least one nucleic acid molecule of (a). In some examples, the method further comprises linking the at least one nucleic acid molecule of (a) and the nucleic acid molecule of the plurality of nucleic acid molecules of (b). In some examples, linking comprises hybridizing the at least one nucleic acid molecule of (a) and the nucleic acid molecule of the plurality of nucleic acid molecules of (b). In some examples, hybridizing comprises hybridizing the connector sequence of the at least one nucleic acid molecule of (a) with the anti-connector sequence of the nucleic acid molecule of the plurality of nucleic acid molecules of (b). In some examples, the method further comprises (i) extending a free 3' end of the nucleic acid molecule of the plurality of nucleic acid molecules using the at least one nucleic acid molecule of (a) as a template, and/or (ii) extending a free 3' end of the at least one nucleic acid molecule of (a) using the nucleic acid molecule of the plurality of nucleic acid molecules as a template, to generate the third nucleic acid molecule. In some examples, the method further comprises ligating the at least one nucleic acid molecule of (a) and the nucleic acid molecule of the plurality of nucleic acid molecules (b). In some examples, the method further comprises contacting the third nucleic acid molecule with a restriction enzyme to generate a sticky end. In some examples, the method further comprises contacting the third nucleic acid molecule with an additional nucleic acid molecule. In some examples, the additional nucleic acid molecule encodes a constant region or portion thereof of a TCR chain. In some examples, the method further comprises ligating the third nucleic acid molecule and the additional nucleic acid molecule. In some examples, a plurality of nucleic acid molecules, each encoding a different TCR chain or portion thereof, are generated in the same compartment. In some examples, at least five different nucleic acid molecules of the plurality of nucleic acid molecules are generated in the same compartment. In some examples, at least ten different nucleic acid molecules of the plurality of nucleic acid molecules are generated in the same compartment. In some examples, at least 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different nucleic acid molecules of the plurality of nucleic acid molecules are generated in the same compartment. In some examples, the same compartment is a well, a tube, or a droplet. In some examples, the at least one nucleic acid molecule comprises a unique barcode. In some examples, the unique barcode is a primer binding site. In some examples, the connector

sequence comprises a unique barcode. In some examples, the unique barcode is a primer binding site.

[0007] In another example, the present disclosure provides a composition comprising (a) a plurality of nucleic acid molecules, wherein each nucleic acid molecule of the plurality of nucleic acid molecules comprises a sequence derived from a T-cell receptor (TCR) V gene and does not comprise a CDR3 sequence, wherein a first nucleic acid molecule of the plurality comprises a first anti-connector sequence and a second nucleic acid molecule of the plurality comprises a second anti-connector sequence, wherein the first anti-connector sequence is different from the second anti-connector sequence, and wherein the sequence derived from a TCR V gene of the first nucleic acid molecule and the second nucleic acid molecule are derived from a different TCR V gene; and (b) at least one nucleic acid molecule comprising a sequence encoding a CDR3 of a TCR chain, wherein the at least one nucleic acid molecule further comprises a first connector sequence complementary to the first anti-connector sequence.

[0008] In some examples, the composition is a liquid composition. In some examples, the plurality of nucleic acid molecules of (a) and the at least one nucleic acid molecule of (b) are in a same compartment. In some examples, the sequence derived from the TCR V gene comprises at least ten nucleotides of the TCR V gene. In some examples, the TCR V gene is a TRAV gene, a TRBV gene, a TRGV gene, or a TRDV gene. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, and/or FR3. In some examples, the at least one nucleic acid molecule further comprises a J region of the TCR chain. In some examples, the at least one nucleic acid molecule further comprises an additional sequence encoding an additional CDR3 of an additional TCR chain. In some examples, the at least one nucleic acid molecule further comprises an additional J region of the additional TCR chain. In some examples, the sequence encoding the CDR3 and the additional sequence encoding the CDR3 are separated by at most 100 nucleotides. In some examples, the TCR chain and the additional TCR chain are a cognate pair of TCR chains. In some examples, the at least one nucleic acid molecule of (b) comprises a first plurality of nucleic acid molecules, and wherein each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a CDR3 of a TCR chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules encodes a different CDR3 of a different TCR chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a different connector sequence, which different connector sequence is capable of

specifically linking to a nucleic acid molecule of the plurality of nucleic acid molecules that comprises a sequence derived from any single given TCR V gene. In some examples, the first anti-connector sequence or the second anti-connector sequence comprises a TCR V gene sequence. In some examples, the TCR V gene sequence comprises at least three nucleotides of the TCR V gene adjacent to a sequence encoding a CDR3 in a rearranged gene. In some examples, the first anti-connector sequence or the second anti-connector sequence comprises a pre-determined sequence. In some examples, the first connector sequence hybridizes to the first anti-connector sequence. In some examples, the at least one nucleic acid molecule of (b) comprises a unique barcode. In some examples, the unique barcode is a primer binding site. In some examples, the first connector sequence of the at least one nucleic acid molecule comprises a unique barcode. In some examples, the unique barcode is a primer binding site.

[0009] In another example, the present disclosure provides a method for generating a plurality of nucleic acid molecules, comprising: providing a first plurality of nucleic acid molecules, wherein a nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a first CDR3 of a first T-cell receptor (TCR) chain and a second CDR3 of a second TCR chain, wherein the first CDR3 and the second CDR3 are from a cognate pair of TCR chains; providing a second plurality of nucleic acid molecules, wherein a nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene, wherein the nucleic acid molecule does not comprise a sequence encoding a constant domain; and contacting the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules, wherein the nucleic acid molecule of the first plurality of nucleic acid molecules links with the nucleic acid molecule of the second plurality of nucleic acid molecules to form a nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence derived from the TCR V gene, wherein the sequence encoding the first CDR3 and the second CDR3 and the TCR V gene are derived from the cognate pair of TCR chains.

[0010] In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a different first CDR3 of a first TCR chain and/or a different CDR3 of a second TCR chain. In some examples, the first plurality of nucleic acid molecules comprises at least about 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences. In some examples, each nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a different TCR V gene. In some examples, the second

plurality of nucleic acid molecules comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80 or more different TCR V genes. In some examples, the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules are contacted in a same compartment. In some examples, the nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a connector sequence, wherein the connector sequence links the nucleic acid molecule of the first plurality of nucleic acid molecules and the nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, the nucleic acid molecule of the second plurality of nucleic acid molecules further comprises an anti-connector sequence, which anti-connector sequence is complementary to the connector sequence. In some examples, the connector sequence hybridizes to the anti-connector sequence to link the nucleic acid molecule of the first plurality of nucleic acid molecules and the nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, the connector sequence is codon-diversified such that the connector sequence of the nucleic acid molecule of the first plurality of nucleic acid molecules is different from other connector sequences of other nucleic acid molecules of the first plurality of nucleic acid molecules. In some examples, the nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a first J region of the first TCR chain and/or a second J region of the second TCR chain. In some examples, (i) the first TCR chain is a TCR alpha chain and the second TCR chain is a TCR beta chain or (ii) the first TCR chain is a TCR gamma chain and the second TCR chain is a TCR delta chain. In some examples, the TCR V gene is a TRAV gene, a TRBV gene, a TRGV gene, or a TRDV gene. In some examples, the nucleic acid molecule of the second plurality of nucleic acid molecules is a double-stranded nucleic acid molecule. In some examples, the nucleic acid molecule of the second plurality of nucleic acid molecules further comprises a sequence encoding a portion of a self-cleaving peptide. In some examples, the anti-connector sequence is an overhang of the nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, the connector sequence or the anti-connector sequence is at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more nucleotides in length. In some examples, the method further comprises (i) extending a 3' end of the nucleic acid molecule of the first plurality of nucleic acid molecules hybridized thereto with the nucleic acid molecule of the second plurality of nucleic acid

molecules and/or (ii) extending a 3' end of the nucleic acid molecule of the second plurality of nucleic acid molecules hybridized thereto with the nucleic acid molecule of the first plurality of nucleic acid molecules. In some examples, the method further comprises ligating the nucleic acid molecule of the first plurality of nucleic acid molecules with the nucleic acid molecule of the second plurality of nucleic acid molecule.

[0011] In some examples, the method further comprises contacting the nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence derived from the TCR V gene with a restriction enzyme to generate a sticky end. In some examples, the method further comprises contacting the nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence derived from the TCR V gene with an additional nucleic acid molecule comprising a sequence encoding a constant region or portion thereof. In some examples, the method further comprises ligating the nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence derived from the TCR V gene with the additional nucleic acid molecule through the sticky end. In some examples, the sequence encoding the first CDR3 and the second encoding the second CDR3 are separated by at most about 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, or 5 nucleotides. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding FR1, CDR1, FR2, CDR2, and FR3. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, and FR3.

[0012] In another example, the present disclosure provides a composition comprising: a first plurality of nucleic acid molecules, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a first CDR3 of a first T-cell receptor (TCR) chain and a second CDR3 of a second TCR chain, wherein the first CDR3 and the second CDR3 are from a cognate pair of TCR chains; and a second plurality of nucleic acid molecules, wherein each nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene, and wherein each nucleic acid molecule of the second plurality of nucleic acid molecules does not comprise a sequence encoding the first CDR3 and the second CDR3; wherein (i) each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a different first CDR3 and/or second CDR3, and/or (ii) each nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a different TCR V gene. In some examples, the first plurality of nucleic acid molecules comprises at least about 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000,

7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences. In some examples, the second plurality of nucleic acid molecules comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80 or more different TCR V genes.

[0013] In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a connector sequence, wherein a given connector sequence is usable to link a given nucleic acid molecule of the first plurality of nucleic acid molecules and a given nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, each nucleic acid molecule of the second plurality of nucleic acid molecules further comprises an anti-connector sequence, which anti-connector sequence is complementary to the connector sequence. In some examples, the connector sequence is codon-diversified such that the given connector sequence of the given nucleic acid molecule of the first plurality of nucleic acid molecules is different from other connector sequences of other nucleic acid molecules of the first plurality of nucleic acid molecules. In some examples, the connector sequence encodes an amino acid sequence. In some examples, the connector sequence is in frame with the sequence encoding the first CDR3 of the first TCR chain and the second CDR3 of the second TCR chain. In some examples, the connector sequence comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more nucleotides. In some examples, the connector sequence comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more nucleotides of the TCR V gene adjacent to the sequence encoding the first CDR3 of the first TCR chain or the second CDR3 of the second TCR chain. In some examples, a given amino acid sequence encoded by the given connector sequence is the same or substantially the same as at least one other amino acid sequence encoded by at least one other connector sequence. In some examples, a given amino acid sequence encoded by the given connector sequence is different from other amino acid sequences encoded by other connector sequences. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a first J region of the first TCR chain and/or a second J region of the second TCR chain. In some examples, the composition is a liquid composition. In some examples, the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules are

within a same compartment. In some examples, the given nucleic acid molecule of the first plurality of nucleic acid molecules is linked to the given nucleic acid molecule of the second plurality of nucleic acid molecules through the given connector sequence. In some examples, the given nucleic acid molecule of the first plurality of nucleic acid molecules hybridizes to the given nucleic acid molecule of the second plurality of nucleic acid molecules through the given connector sequence hybridized to a given anti-connector sequence. In some examples, the sequence encoding the first CDR3 and the sequence encoding the second CDR3 are separated by at most 100 nucleotides. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding FR1, CDR1, FR2, CDR2, and FR3. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, and FR3. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules or the second plurality of molecules is chemically synthesized. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules is at most about 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, or 50 nucleotides long.

[0014] In another example, the present disclosure provides a composition comprising a plurality of nucleic acid molecules, each nucleic acid molecule of the plurality of nucleic acid molecules comprising a sequence derived from a T-cell receptor (TCR) V gene sequence, wherein the plurality of nucleic acid molecules comprises a first nucleic acid molecule having a first connector sequence and a second nucleic acid molecule having a second connector sequence, wherein the first connector sequence is different from the second connector sequence.

[0015] In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules comprises a sequence derived from a different TCR V gene. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules comprises a different connector sequence. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules does not comprise a sequence encoding a CDR3 of a TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules does not comprise a sequence encoding a constant domain of a TCR chain. In some examples, the sequence derived from the TCR V gene comprises at least 10, 20, 30, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more nucleotides of the TCR V gene. In some examples, the TCR V gene is a TRAV gene, a TRBV gene, a TRGV gene, or a TRDV gene.

[0016] In another example, the present disclosure provides a composition comprising a plurality of nucleic acid molecules, each nucleic acid molecule of the plurality of nucleic acid molecules encoding a CDR3 or a portion thereof of a T-cell receptor (TCR) chain, wherein the plurality of nucleic acid molecules comprises a first nucleic acid molecule having a first connector sequence and a second nucleic acid molecule having a second connector sequence, wherein the first connector sequence is different from the second connector sequence.
[0017] In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules further comprises a J region of a TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules encodes a first CDR3 or a portion thereof of a first TCR chain and a second CDR3 or a portion thereof of a second TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules further comprises a first J region of a first TCR chain and a second J region of a second TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules encodes a different CDR3 or a portion thereof of a different TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules comprises a different connector sequence. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules does not comprise greater than 200, 150, 100, 80, 50, 40, 30, 20, or 10 nucleotides TCR V gene. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules does not comprise a sequence encoding a constant domain of a TCR chain. In some examples, the first connector sequence or the second connector sequence comprises a sequence derived from a TCR V gene. In some examples, the sequence derived from the TCR V gene comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more nucleotides of the TCR V gene adjacent to a sequence encoding a CDR3 in a rearranged gene. In some examples, the first connector sequence or the second connector sequence comprises a pre-determined sequence. In some examples, the first connector sequence or the second connector sequence comprises a sequence complementary to a TCR V gene sequence. In some examples, the composition further comprises a second plurality of nucleic acid molecules, each nucleic acid molecule of the second plurality of nucleic acid molecules comprising a sequence derived from a TCR V gene. In some examples, a first nucleic acid molecule of the second plurality comprises a first anti-connector sequence, which first anti-connector sequence is complementary to the first connector sequence. In some examples, a second nucleic acid molecule of the second plurality comprises a second anti-connector sequence, which second anti-connector sequence

is complementary to the second connector sequence. In some examples, the first anti-connector sequence of the first nucleic acid molecule of the second plurality is linked to the first connector sequence of the first nucleic acid molecule of the first plurality. In some examples, the second anti-connector sequence of the second nucleic acid molecule of the second plurality is linked to the second connector sequence of the second nucleic acid molecule of the first plurality.

[0018] In another example, the present disclosure provides a composition comprising a plurality of nucleic acid molecules, each comprising a sequence encoding at least ten amino acids (e.g., in some cases, encoding at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or more amino acids) of a T-cell receptor (TCR) chain, wherein the plurality of nucleic acid molecules comprises a first nucleic acid molecule having a first connector sequence and a second nucleic acid molecule having a second connector sequence, wherein the first connector sequence is different from the second connector sequence, wherein the first connector sequence or the second connector sequence encodes a portion of a TCR chain and wherein the first connector sequence or the second connector sequence is in frame with the sequence encoding at least ten (e.g., in some cases, encoding at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or more amino acids) amino acids of a TCR chain.

[0019] In some examples, the first connector sequence or the second connector sequence comprises at least 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more contiguous nucleotides of a TCR chain gene and is in frame with the sequence encoding at least ten amino acids of a TCR chain. In some examples, the first connector sequence and the second connector sequence encodes at least two contiguous amino acids of a TCR chain. In some examples, the TCR chain of the portion of the TCR chain and the TCR chain encoded by the sequence encoding at least ten amino acids is the same. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules encodes a CDR3 or portion thereof of the TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules further comprises a J region of the TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules encodes a first CDR3 or portion thereof of a first TCR chain and a second CDR3 or portion thereof of a second TCR chain. In some examples, each nucleic acid molecule of the plurality of nucleic acid molecules further comprises a first J region of a first TCR chain and a second J region of a second TCR chain. In some examples,

a sequence encoding the first CDR3 or portion thereof and a sequence encoding the second CDR3 or portion thereof are separated by at most 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, or 5 nucleotides. In some examples, the first connector sequence or the second connector sequence comprises a sequence derived from a TCR V gene. In some examples, the first connector sequence or the second connector sequence comprises a pre-determined sequence. In some examples, the first connector sequence comprises at least one nucleotide that is different from a nucleotide of the second connector sequence. In some examples, the first connector sequence encodes a same amino acid sequence as the second connector sequence. In some examples, the first connector sequence encodes a different amino acid sequence from the second connector sequence.

[0020] In another example, the present disclosure provides a method for generating a plurality of nucleic acid molecules, each nucleic acid molecule of the plurality encoding a T-cell receptor (TCR) chain or region thereof, the method comprising: contacting a first plurality of nucleic acid molecules and a second plurality of nucleic acid molecules to generate a third plurality of nucleic acid molecules comprising at least two (e.g., at least about 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more) different nucleic acid molecules, wherein each of the at least two different nucleic acid molecules has a different sequence encoding a different TCR chain or region thereof, and wherein the at least two different nucleic acid molecules are generated in a same compartment.

[0021] In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a CDR3 of the TCR chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules comprises a J region of the TCR chain. In some examples, each nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene of the TCR chain. In some examples, the TCR V gene is a human TCR V gene. In some examples, the TCR V gene is a human TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, or TRAV41. In some examples, the TCR V gene is a human TRBV2, TRBV3-1, TRBV4-1, TRBV4-2,

TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, or TRBV30. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding FR1, CDR1, FR2, CDR2, and FR3. In some examples, the sequence derived from the TCR V gene comprises a sequence encoding L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, and FR3. In some examples, the TCR chain is a TCR alpha chain, a TCR beta chain, a TCR gamma chain, or a TCR delta chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules further comprises an additional sequence encoding an additional CDR3 of an additional TCR chain. In some examples, each nucleic acid molecule of the first plurality of nucleic acid molecules comprises an additional J region of the additional TCR chain. In some examples, the TCR chain and the additional TCR chain are a cognate pair of TCR chains. In some examples, a nucleic acid molecule of the plurality of nucleic acid molecules encodes a different TCR or portion thereof. In some examples, a given nucleic acid molecule of the first plurality of nucleic acid molecules comprises a connector sequence, which connector sequence is usable for linking the given nucleic acid molecule of the first plurality of nucleic acid molecules to a given nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, the given nucleic acid molecule of the first plurality of nucleic acid molecules and the given nucleic acid molecule of the second plurality of nucleic acid molecules encodes a functional TCR chain or portion thereof. In some examples, the given nucleic acid molecule of the second plurality of nucleic acid molecules comprises an anti-connector sequence, which anti-connector sequence is complementary to the connector sequence of the given nucleic acid molecule of the first plurality of nucleic acid molecules. In some examples, the method further comprises linking the given nucleic acid molecule of the first plurality of nucleic acid molecules and the given nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, linking comprises hybridizing the given nucleic acid molecule of the first plurality of nucleic acid molecules and the given nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, hybridizing comprises hybridizing the connector sequence of the given nucleic acid molecule of the first plurality of nucleic acid molecules with the anti-connector sequence of the given nucleic acid molecule of the second plurality of nucleic acid

molecules. In some examples, the method further comprises (i) extending a free 3' end of the given nucleic acid molecule of the second plurality of nucleic acid molecules using the given nucleic acid molecule of the first plurality of nucleic acid molecules as a template, and/or (ii) extending a free 3' end of the nucleic acid molecule of the first plurality of nucleic acid molecules using the given nucleic acid molecule of the second plurality of nucleic acid molecules as a template, to generate a nucleic acid molecule of the third plurality of nucleic acid molecules. In some examples, the method further comprises ligating the given nucleic acid molecule of the first plurality of nucleic acid molecules and the given nucleic acid molecule of the second plurality of nucleic acid molecules. In some examples, the method further comprises contacting the nucleic acid molecule of the third plurality of nucleic acid molecules with a restriction enzyme to generate a sticky end. In some examples, the method further comprises contacting the nucleic acid molecule of the third plurality of nucleic acid molecules with an additional nucleic acid molecule. In some examples, the additional nucleic acid molecule encodes a constant region or a portion thereof of a TCR chain. In some examples, the method further comprises ligating the nucleic acid molecule of the third plurality of nucleic acid molecules and the additional nucleic acid molecule. In some examples, at least five (e.g., in some cases, at least about 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1,000, 2,000, 3,000, 4,000, 5,000, 10,000, 20,000, 30,000, 40,000, or more) different nucleic acid molecules of the third plurality of nucleic acid molecules are generated in the same compartment. In some examples, at least ten different nucleic acid molecules of the third plurality of nucleic acid molecules are generated in the same compartment. In some examples, the same compartment is a well, a tube, or a droplet.

[0022] In another example, the present disclosure provides a method for generating a plurality of nucleic acid molecules, comprising: (a) providing a first plurality of nucleic acid molecules, wherein a nucleic acid molecule of the first plurality of nucleic acid molecules comprises a sequence encoding a first CDR3 of a first T-cell receptor (TCR) chain and a second CDR3 of a second TCR chain, wherein the first CDR3 and the second CDR3 are from a cognate pair of TCR chains; (b) providing a second plurality of nucleic acid molecules, wherein a nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene; and (c) contacting the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules, wherein the nucleic acid molecule of the first plurality of nucleic acid molecules links with the nucleic acid molecule of the second plurality of nucleic acid molecules to form a linear nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence

derived from the TCR V gene, wherein the sequence encoding the first CDR3 and the second CDR3 and the TCR V gene are derived from the cognate pair of TCR chains. In some examples, the first plurality of nucleic acid molecules comprises at least about 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences. In some examples, the second plurality of nucleic acid molecules comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80 or more different TCR V genes.

[0023] In another example, the present disclosure provides a method for generating a plurality of nucleic acid molecules, comprising: (a) providing a first plurality of nucleic acid molecules, wherein a nucleic acid molecule of the first plurality of nucleic acid molecules comprises (i) a synthetic sequence encoding a first CDR3 of a first T-cell receptor (TCR) chain and a second CDR3 of a second TCR chain and (ii) a synthetic sequence encoding a third CDR3 of a third T-cell receptor (TCR) chain and a fourth CDR3 of a fourth TCR chain , wherein the first CDR3 and the second CDR3 are from a first cognate pair of TCR chains and wherein the third CDR3 and the fourth CDR3 are from a second cognate pair of TCR chains; (b) providing a second plurality of nucleic acid molecules, wherein a nucleic acid molecule of the second plurality of nucleic acid molecules comprises a sequence derived from a TCR V gene; and (c) contacting the first plurality of nucleic acid molecules and the second plurality of nucleic acid molecules, wherein the nucleic acid molecule of the first plurality of nucleic acid molecules links with the nucleic acid molecule of the second plurality of nucleic acid molecules to form a nucleic acid molecule comprising the sequence encoding the first CDR3 and the second CDR3 and the sequence derived from the TCR V gene, wherein the sequence encoding the first CDR3 and the second CDR3 and the TCR V gene are derived from the cognate pair of TCR chains. In some examples, the first plurality of nucleic acid molecules comprises at least about 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, or more different sequences. In some examples, the second plurality of nucleic acid molecules comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60 61, 62,

63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80 or more different TCR V genes.

[0024] In another example, the present disclosure provides a method of identifying a sequence of a natively paired T-cell receptor (TCR) in a tissue sample from a subject, comprising: (a) identifying one or more paired sequences of one or more natively paired TCRs in a sample containing a plurality of peripheral T cells obtained from the subject, wherein each of the one or more paired sequences comprises a CDR3 sequence; and (b) identifying a tissue CDR3 sequence of a TCR chain of a TCR in the tissue sample for which the other TCR chain to which it is natively paired is unknown, wherein the tissue CDR3 sequence matches a CDR3 sequence of at least one paired sequence of the one or more paired sequences of the one or more natively paired TCRs, thereby identifying the at least one paired sequence as the sequence of the natively paired TCR in the tissue sample. In some examples, identifying in (a) comprises sequencing the one or more natively

paired TCRs in the sample containing the plurality of peripheral T cells. In some examples, the sequencing comprises single cell sequencing. In some examples, the single cell sequencing comprises partitioning the plurality of peripheral T cells into a plurality of compartments, each compartment comprising an individual peripheral T cell of the plurality of peripheral T cells. In some examples, the tissue sample is not a bodily fluid sample. In some examples, the tissue sample is a solid tumor sample. In some examples, the tissue sample is a fixed or frozen sample. In some examples, the sample containing the plurality of peripheral T cells is a peripheral blood mononuclear cell (PBMC) sample. In some examples, the method further comprises, prior to (a), obtaining a blood sample from the subject. In some examples, the method further comprises, prior to (a), isolating peripheral blood mononuclear cells from the blood sample. In some examples, the tissue sample comprises a tumor-infiltrating T cell.

[0025] In another example, the present disclosure provides a method of identifying a target-reactive T-cell receptor (TCR), comprising: (a) providing a cell comprising the TCR identified using the methods described herein; and (b) contacting the cell with a target antigen presented by an antigen-presenting cell (APC), wherein the cell binds to the target antigen presented by the APC via the TCR, thereby identifying the TCR as the target-reactive TCR. In some examples, the target antigen is a tumor antigen (e.g., tumor-associated antigens or tumor-specific antigens). In some examples, the method further comprises delivering a sequence encoding the target-reactive TCR into a host cell. In some examples, the method further comprises administering the host cell into the subject. In some examples, the host cell is a T cell. In some examples, the T cell is an autologous T cell. In some examples, the T cell is an allogeneic T cell. In some examples, the cell is a reporter cell line, which reporter cell line comprises a reporter gene that is expressed upon the cell binding to the target antigen presented by the APC.

[0026] Additional examples and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative examples of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different examples, and its several details are capable of modifications in various obvious respects, all without

departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0027] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative

amino acid sequence (e.g., the conserved portion of the TCR V gene adjacent to the CDR3). When the connector sequence encodes a conserved portion of a TCR V gene, the connector sequence can be codon-diversified such that the connector sequence can be used to link a nucleic acid molecule to another nucleic acid molecule specifically, resulting in a constructed nucleic acid molecule encoding a cognate pair of a TCR. In some embodiments, the connector sequence (or anti-connector sequence) comprises at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200, or more nucleotides of the TCR V gene adjacent to a sequence encoding a CDR3 in a rearranged gene. Because of the specificity of the connector sequence and the anti-connector sequence, a pool of nucleic acid molecules having different sequences which encode different TCRs can be constructed in a bulk reaction (e.g., in a same compartment). The connector sequence can prescribe which TCR V gene the sequence encoding a CDR3 should be linked to according to a reference sequence (e.g., a native sequence of a TCR chain determined by sequencing). The connector sequence (or the anti-connector sequence) can be an arbitrary (e.g., pre-determined) sequence which may not encode a portion of a TCR V gene. And in such cases, the arbitrary sequence can be removed after linking two nucleic acid fragments together.

[0094] FIG. 7 depicts an example to use arbitrary connector (701) and anti-connector (702) sequence to link a CDR3-J polynucleotide to a designated V gene germline polynucleotide (thin arrow). Here each V gene germline polynucleotide has as partially double-stranded structure. The top strand, with its 3' end to its right in this figure, has a single-stranded region at its 3' end. The connector and the anti-connector sequences may be single stranded and may hybridize to each other. The connector and anti-connector sequence only serves the purpose of specific hybridization and may not be related to TCR whatsoever, hence arbitrary. After the hybridization between the connector and the anti-connector, the 3' end of the top strand of the V gene germline polynucleotide may hybridize to the CDR3-J polynucleotide and may be extended by a DNA polymerase. The number of nucleotides on the 3' end of the top strand of the V gene germline polynucleotide that are hybridized to the CDR3-J polynucleotide may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15, or up to 20.

[0095] The nucleic acid fragment used to construct nucleic acid molecule encoding a TCR or a portion thereof can comprise a self-cleaving peptide. The self-cleaving peptide can be a 2A peptide, an intein peptide, or a hedgehog peptide. Examples of 2A peptide include, but are not limited to, P2A (e.g., sequence: ATNFSLLKQAGDVEENPGP (SEQ ID NO: 284)), E2A

(e.g., sequence QCTNYALLKLAGDVESNPGP (SEQ ID NO: 285)), F2A (e.g., sequence VKQTLNFDLLKLAGDVESNPGP (SEQ ID NO: 286)), and T2A (e.g., sequence EGRGSLLTCGDVEENPGP (SEQ ID NO: 287)) peptide.

[0096] The nucleic acid fragment used to construct nucleic acid molecule encoding a TCR or a portion thereof can comprise a restriction enzyme recognition site. For example, the restriction enzyme recognition site can be a recognition site for Type IIS restriction enzyme. Examples of Type-IIS restriction enzymes which can be useful in the present disclosure include, but are not limited to, EarI, MnlI, PleI, AlwI, BbsI, BbvI, BcoDI, BsaI, BseRI, BsmAI, BsmBI, BspMI, Esp3I, HgaI, SapI, SfaNI, BbvI, BsmFI, BsrDI, BtsI, FokI, BseRI, HphI, MlyI and MboII. In some cases, two or more different restriction enzymes can be used during nucleic acid construction process. In some cases, a restriction enzyme that create a 4-bp 5' overhang (for example, BbsI, BbvI, BcoDI, BsaI, BsmBI, FokI, etc.) can be used. In some cases, a restriction enzyme that creates a blunt end or 3' overhang (for example, BseRI, BsrDI, BtsI, MlyI, etc.) can be used.

[0097] A nucleic acid fragment used to construct nucleic acid molecule encoding a TCR or a portion thereof can be circularized. For example, the nucleic acid fragment can be circularized by joining two ends of the nucleic acid fragment by ligation. The ligation can be blunt end ligation. The ligation can be performed after creating sticky ends using 5'-to-3' exonuclease (e.g, Gibson Assembly), 3'-to-5' exonuclease (e.g., sequence and ligase independent cloning or SLIC), or USER enzyme mix (e.g., USER friendly DNA recombination or USERec). Additional examples of circularization methods include, but are not limited to, circular polymerase extension cloning (CPEC) and seamless ligation cloning extract (SLiCE) assembly. Alternatively, these two ends can be joined by overlapping PCR. A variety of ligases can be used for ligation, for example, including but not limited to, T4 DNA ligase, T4 RNA ligase, E. coli DNA ligase.

[0098] The nucleic acid fragment used to construct the nucleic acid molecule encoding a TCR chain or portion thereof can be synthesized chemically. For example, the nucleic acid fragment can be pre-synthesized by chip-based synthesis. In some cases, the nucleic acid fragment synthesized can be equal to or greater than about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, or more nucleotides in length. In some cases, the nucleic acid fragment synthesized by can be equal to or less than about 500, 450, 400, 350, 300, 250, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, or 10 nucleotides in length.

[0099] The two nucleic acid sequences encoding two peptide chains of a TCR can be constructed in several orientations, for example, head-to-head, head-to-tail, and tail-to-tail. As described herein, "head" refers to "5' end" of a sense nucleic acid strand and "tail" refers

single-chain polypeptide. Each expressed chain may contain the full variable domain sequence including the rearranged V(D)J gene. Each expressed chain may contain the full variable domain sequence including CDR1, CDR2, and CDR3. Each expressed chain may contain the full variable domain sequence including FR1, CDR1, FR2, CDR2, FR3, and CDR3. In some cases, each expressed chain may further contain a constant domain sequence.

[00120] To create expression vectors, additional sequences may be added to the constructed nucleic acid molecules. These additional sequences include vector backbone (e.g., elements required for the vector's replication in target cell or in temporary host such as E. coli), promoters, IRES, sequence encoding the self-cleaving peptide, terminators, accessory genes (such as payloads), as well as partial sequences of the immunoreceptor polynucleotides (such as part of the sequences encoding the constant domains).

[00121] Protease cleavage sites include, but are not limited to, an enterokinase cleavage site: (Asp)4Lys (SEQ ID NO: 288); a factor Xa cleavage site: Ile-Glu-Gly-Arg (SEQ ID NO: 289); a thrombin cleavage site, e.g., Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO: 290); a renin cleavage site, e.g., His-Pro-Phe-His-Leu-Val-Ile-His (SEQ ID NO: 291); a collagenase cleavage site, e.g., X-Gly-Pro (where X is any amino acid); a trypsin cleavage site, e.g., Arg-Lys; a viral protease cleavage site, such as a viral 2A or 3C protease cleavage site, including, but not limited to, a protease 2A cleavage site from a picornavirus, a Hepatitis A virus 3C cleavage site, human rhinovirus 2A protease cleavage site, a picornavirus 3 protease cleavage site; and a caspase protease cleavage site, e.g., DEVD (SEQ ID NO: 292) recognized and cleaved by activated caspase-3, where cleavage occurs after the second aspartic acid residue. In some embodiments, the present disclosure provides an expression vector comprising a protease cleavage site, wherein the protease cleavage site comprises a cellular protease cleavage site or a viral protease cleavage site. In some embodiments, the first protein cleavage site comprises a site recognized by furin; VP4 of IPNV; tobacco etch virus (TEV) protease; 3C protease of rhinovirus; PC5/6 protease; PACE protease, LPC/PC7 protease; enterokinase; Factor Xa protease; thrombin; genenase I; MMP protease; Nuclear inclusion protein a(N1a) of turnip mosaic potyvirus; NS2B/NS3 of Dengue type 4 flaviviruses, NS3 protease of yellow fever virus; ORF V of cauliflower mosaic virus; KEX2 protease; CB2; or 2A. In some embodiments, the protein cleavage site is a viral internally cleavable signal peptide cleavage site. In some embodiments, the viral internally cleavable signal peptide cleavage site comprises a site from influenza C virus, hepatitis C virus, hantavirus, flavivirus, or rubella virus.

[00122] A suitable IRES element to include in the vector of the present disclosure can comprise an RNA sequence capable of engaging a eukaryotic ribosome. In some

[00157] Table 1. Connector sequences derived from mouse TRAV genes

SEQ ID NO:	Connector of mouse TRAV genes	Connector Sequence
1	> ConA of Trav1|ENSMUST00000103567.5	TCAAAGACTCTGCCTCATACCTCT
2	> ConA of Trav2|ENSMUST00000196939.1	CTGAGAGACGCAGCTGTGTATTACT
3	> ConA of Trav3-1|ENSMUST00000103569.2	GGGGACTCAGCCGTGTACTTCT
4	> ConA of Trav3-3|ENSMUST00000181768.2	GGTGACTCCGCAGCCTATTTCT
5	> ConA of Trav3-4|ENSMUST00000103670.3	TGGGGATAGCGCAGTCTATTTCT
6	> ConA of Trav3d-3|ENSMUST00000196023.1	CCGGAGACAGCGCAGTTTATTTTT
7	> ConA of Trav3n-3|ENSMUST00000197557.1	GGTGACAGCGCCGTCTATTTTT
8	> ConA of Trav4-2|ENSMUST00000103637.5	TGGAGGACTCAGGCACTTACTTCT
9	> ConA of Trav4-3|ENSMUST00000103655.2	TGGAGGACTCTGGGACATACTTTT
10	> ConA of Trav4-4-dv10|ENSMUST00000103663.5	TGGAGGACTCTGGCACCTATTTTT
11	> ConA of Trav4d-3|ENSMUST00000103592.1	ACTCGAGGATTCCGGTACTTATTTCT
12	> ConA of Trav4d-4|ENSMUST00000103600.2	GAAGACTCCGGGACCTACTTTT
13	> ConA of Trav4n-3|ENSMUST00000103618.1	GCTGGAGGATTCCGGAACCTATTTCT
14	> ConA of Trav4n-4|ENSMUST00000103627.2	CTCGAAGATAGCGGCACATATTTTT
15	> ConA of Trav5-1|ENSMUST00000103570.1	GCCTGGTGATAGCGCAATATACTTCT
16	> ConA of Trav5d-4|ENSMUST00000179701.1	TGGCGACTCTGCAATGTACTTCT
17	> ConA of Trav5n-4|ENSMUST00000179997.1	CCCGGAGACTCTGCTATGTATTTTT
18	> ConA of Trav6-1|ENSMUST00000103571.1	GGAATCCGATAGCGCAGTCTATTACT
19	> ConA of Trav6-2|ENSMUST00000198058.1	TCCGACAGCGCTGTCTACTACT
20	> ConA of Trav6-3|ENSMUST00000180549.2	AAGAGATTGATAGCGCTGTTTACTACT
21	> ConA of Trav6-4|ENSMUST00000184650.1	AGGAATCTGATTCCGCAGTCTATTTTT
22	> ConA of Trav6-5|ENSMUST00000181210.2	AGAATCTGATAGCGCCGTTTATTATT

96	> ConA of Trav14n-2|ENSMUST00000197614.1	CCCGGAGATTCTGCCACTTATTTCT
97	> ConA of Trav14n-3|ENSMUST00000103652.4	CTGGCGACAGCGCTACTTATTTCT
98	> ConA of Trav15-1-dv6-1|ENSMUST00000103653.2	AACCAGACGATTCGGGAAAGTATTTCT
99	> ConA of Trav15-2-dv6-2|ENSMUST00000103660.3	CAGAGGATTCAGGGACGTACTTCT
100	> ConA of Trav15d-1-dv6d-1|ENSMUST00000103616.4	CCAGACGACTCCGGAAAGTACTTTT
101	> ConA of Trav15d-2-dv6d-2|ENSMUST00000199800.1	CCGAGGACTCCGGTACATACTTCT
102	> ConA of Trav15n-1|ENSMUST00000103590.3	AACCCGATGACTCTGGTAAGTATTTTT
103	> ConA of Trav15n-2|ENSMUST00000199112.1	GCCAGAAGACTCCGGTACATATTTTT
104	> ConA of Trav16|ENSMUST00000103667.5	TCAAATTGAAGATTCTGCAGTCTACTTTT
105	> ConA of Trav16d-dv11|ENSMUST00000103606.1	GATTGAGGACTCGGCAGTATATTTCT
106	> ConA of Trav16n|ENSMUST00000199280.1	AAATCGAAGACTCTGCAGTTTACTTTT
107	> ConA of Trav17|ENSMUST00000103672.8	GAGCGACTCAGCCAAGTACTTCT
108	> ConA of Trav18|ENSMUST00000103673.10	AGGGGATGCTGGGATCTACTTTT
109	> ConA of Trav19|ENSMUST00000103674.5	CCCGAAGATACAGCTGTCTACCTGT
110	> ConA of Trav21-dv12|ENSMUST00000180938.2	AGGGACGCAGCAGTCTATCATT
111	> ConA of Trav23|ENSMUST00000199137.1	GCCACTCTGCCATCTACTTCTGT

[00158] Table 2. Connector sequences derived from mouse TRBV genes

SEQ ID NO:	**Connector of mouse TRBV genes**	**Connector sequence**
112	> ConB of Trbv1|ENSMUST00000103262.2	GGCGCACACTGTACTGCACAT
113	> ConB of Trbv2|ENSMUST00000103263.2	TGATGACTCGGCCACATACTTCT

specific hybridization are chosen. Next, hybridization yield of every ConX# to every ConX#* using the codon-diversified sequence set is then calculated to see if the codon diversification was successful (**FIG. 11A** and **FIG. 11B**). **FIG. 11A** shows hybridization yield of the codon-diversified connector sequences (ConA# to ConA#*). **FIG. 11B** shows hybridization yield of the codon-diversified connector sequences (ConB# to ConB#*).

Example 4. Connector sequences with arbitrary sequences

[00160] Table 3 provides arbitrary sequences that can be used as connector sequences to link CDR3-J polynucleotides and the designated V gene germline polynucleotides according to the scheme described in **FIG. 7**.

[00161] Table 3. Connector sequences with arbitrary sequences

SEQ ID NO:	Connector sequence	SEQ ID NO:	Connector sequence
134	CCGGGATTTTGTGACTCATC	209	CTTGTCCACTAAACGCAACG
135	GAGGATCGTATGTTTCGCAC	210	CGGGTATCACTGGGTAATGA
136	CTTGTGTGCACTTACCGTAC	211	GGAACAGAGACCAATCCAGT
137	TGATGCATCTCCAGTACAGG	212	GTGGGCATCCGAAATTTCAG
138	CTTCTGTGTGTACCTCGACA	213	CGCGACGACATTACCAATAG
139	CTTGCAATCCTTTACCGTGC	214	CGTCGGAATATGCTCTCAGA
140	GCAAGTGTGGAAAATGACCC	215	GATGCAGATCAATGAGTGGC
141	GAGTCTAGTCTCACAACCCA	216	ACTGCTTACAAGTGTCCACG
142	GAAATGTTGAGGACTCCACG	217	ACTGTATGCAAGCTAGTCCC
143	CCTAACAGATGCTACGTGGA	218	AGATCTCCCAAAAGTGTCCG
144	GTAGGTCCACACAGATTCCA	219	TTCCAGAACCATGTGATCCC
145	GCCAGTCACAGCAAATACAC	220	GCCTTGTCTTTCAACCTCTG
146	CCGCTACCAGTATGTACCTT	221	GATACGGATCTTCACATGCG
147	ACTGTGTTCCTTGTCTTCCG	222	CGCTCATCTAGGTTGGACTA
148	TGAATGCATCTACGGTACCG	223	CGCGTTCAGATTCCAAACAG
149	GCGCTTATCAATCTTGCTCG	224	GCCTGGTTACACATGCTATC
150	CGGTCAATTCAGTAGCCACT	225	GCAAAGGTCCTACAGGTTTC
151	GGACACATGTACACTAGCCA	226	GGCTTTCCATGTCTATGCTC
152	TGGGAGCTCTACGAAAATCC	227	GCGACATTAGCAGAGTAGGT
153	GTTCTCGAGATCGTCACACA	228	CTCGCCATACTATCTGCATG
154	CTTCTGCATTCGATCCTTGC	229	CTACTGAACACTTGGCAAGC
155	GCAGAGTTGTGTGATTGGAG	230	CTGTTCAATTCCTGTGCGAG
156	CCCATCAATTCGGAACCATC	231	CACTGAGATGGAATTTGGCG
157	ATCGTAACCCAAGTCTGTGG	232	GTCCATCACAACTTCCACTG
158	GGCGAAATGATCCCTGAATG	233	GGCTCAGTCTACTTTGCTTC
159	AGTGCTCAGAACTTTCAGGC	234	GTTAGCTTCCGACACAATGG
160	GATCGTTAACTCTTTGGGCG	235	CCGTGACACACTTTCATCTC
161	ACACGAGGATTGCTGTAGAG	236	CCGGGATGTCATTATGAGCA
162	TTCTACCACATTGTCTCGGG	237	GAGTGTCCTACGAGATCAGA
163	GAATGGCTAAACTGTGTCCC	238	TAACGTCTCTCTGAGTGTGG
164	CGGACTGTACGAGAAACTGA	239	ACCCTAGACAAGAGACACCT
165	CTTGCGACAAACTACTCCTG	240	TGCTCAGTACTCTTCATCGC

CLAIMS

What is claimed is:

1. A method for generating a plurality of nucleic acid molecules encoding a plurality of different T-cell receptor (TCR) alpha chains or portion thereof in a same compartment, the method comprising:
   (a) providing a first plurality of nucleic acid molecules, each nucleic acid molecule of the first plurality of nucleic acid molecules comprising a sequence encoding a CDR3 alpha (CDR3α) of a different TCR alpha chain;
   (b) providing a second plurality of nucleic acid molecules, each nucleic acid molecule of the second plurality of nucleic acid molecules comprising a sequence from a TCR alpha variable (TRAV) gene, wherein the second plurality of nucleic acid molecules comprises at least two different sequences from at least two different TRAV genes; and
   (c) generating a third plurality of nucleic acid molecules, each nucleic acid molecule of the third plurality comprising a sequence from one of the sequences encoding a CDR3α from the first plurality of nucleic acid molecules of (a) and a sequence from one of the at least two different TRAV genes of the second plurality of nucleic acid molecules of (b), wherein generating comprises contacting the first plurality of nucleic acid molecules of (a) to the second plurality of nucleic acid molecules of (b) in a same compartment, wherein a nucleic acid molecule of the first plurality of nucleic acid molecules of (a) comprises a connector sequence that is capable of specifically linking to an anti-connector sequence of a nucleic acid molecule of the second plurality of nucleic acid molecules of (b) that comprises a sequence from a single given TRAV gene of the at least two different TRAV genes, thereby generating the plurality of nucleic acid molecules encoding the plurality of different TCR alpha chains or portion thereof, and wherein the single given TRAV gene is cognate to the CDR3α to which it is linked.
2. The method of claim 1, wherein the sequence encoding the CDR3α and the at least two different TRAV genes are from a sample from a subject.
3. The method of claim 2, wherein the sequence from the single given TRAV gene and the sequence encoding the CDR3α to which it is specifically linked in (c) are from a same TCR alpha chain expressed in a T cell of the sample from the subject.
4. The method of any one of claims 1-3, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules further comprises a J region of a TCR alpha chain, and wherein the J region and the CDR3α are from a same TCR alpha chain.

5. The method of any one of claims 1-4, wherein each nucleic acid molecule of the second plurality of nucleic acid molecules further comprises a sequence encoding a first part of a leader peptide (L-PART1), a second part of the leader peptide (L-PART2), framework 1 (FR1), complementarity determining region 1 (CDR1), framework 2 (FR2), complementarity determining region 2 (CDR2), or framework 3 (FR3).
6. The method of any one of claims 1-5, wherein the plurality of nucleic acid molecules generated in (c) further encodes a plurality of different TCR beta chains or portion thereof, and wherein each nucleic acid molecule of the first plurality of nucleic acid molecules of (a) further comprises an additional sequence encoding a CDR3 beta (CDR3β) of a TCR beta chain.
7. The method of claim 6, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules of (a) further comprises an additional J region of the TCR beta chain, and wherein the additional J region and the CDR3β are from a same TCR beta chain.
8. The method of claim 6, wherein the sequence encoding the CDR3α and the additional sequence encoding the additional CDR3β of each nucleic acid molecule of the first plurality of nucleic acid molecules of (a) are separated by at most 100 nucleotides.
9. The method of claim 6, wherein the TCR alpha chain and the TCR beta chain is a cognate pair of TCR chains expressed in a sample from a subject.
10. The method of any one of claims 1-9, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules of (a) comprises a unique connector sequence.
11. The method of claim 10, wherein each nucleic acid molecule of the second plurality of nucleic acid molecules of (b) comprises a unique anti-connector sequence, and wherein the unique anti-connector sequence is complementary to the unique connector sequence.
12. The method of claim 11, wherein linking comprises hybridizing the unique connector sequence with the unique anti-connector sequence.
13. The method of claim 12, further comprising (i) extending a free 3’ end of the nucleic acid molecule of the second plurality of nucleic acid molecules of (b) using the nucleic acid molecule of the first plurality of nucleic acid molecules of (a) as a template, or (ii) extending a free 3’ end of the nucleic acid molecule of the first plurality of nucleic acid molecules of (a) using the nucleic acid molecule of the second plurality of nucleic acid molecules of (b) as a template, to generate the third nucleic acid molecule.
14. The method of any one of claims 1-13, further comprising ligating the nucleic acid molecule of the first plurality of nucleic acid molecules of (a) and the nucleic acid molecule of the second plurality of nucleic acid molecules of (b).

15. The method of any one of claims 1-14, further comprising contacting the third plurality of nucleic acid molecules of (c) with a restriction enzyme to generate a sticky end.
16. The method of claim 15, further comprising contacting the third plurality of nucleic acid molecules of (c) with a fourth plurality of nucleic acid molecules encoding a constant region or portion thereof of a TCR chain, wherein a nucleic acid molecule of the third plurality of nucleic acid molecules of (c) hybridizes with a nucleic acid molecule of the fourth plurality of nucleic acid molecules via the sticky end.
17. The method of claim 16, further comprising ligating a nucleic acid molecule of the third plurality of nucleic acid molecules of (c) and the nucleic acid molecule of the fourth plurality of nucleic acid molecules to generate a fifth plurality of nucleic acid molecules.
18. The method of claim 17, further comprising circularizing nucleic acid molecules of the fifth plurality of nucleic acid molecules, wherein circularizing comprises joining two ends of a nucleic acid molecule of the fifth plurality of nucleic acid molecules by ligation or overlapping polymerase chain reaction (PCR).
19. The method of any one of claims 1-18, wherein at least five nucleic acid molecules, each encoding a different TCR alpha chain or portion thereof, are generated in the same compartment.
20. The method of any one of claims 1-19, wherein each nucleic acid molecule of the first plurality of nucleic acid molecules of (a) further comprises a unique barcode.
21. The method of any one of claims 1-20, further comprising, prior to (a), sequencing a sample from a subject to obtain sequences of cognate pairs of TCRs expressed in the subject.
22. The method of any one of claims 1-21, further comprising expressing a sequence of at least one nucleic acid molecule of the plurality of nucleic acid molecules encoding the plurality of different TCR alpha chains or portion thereof generated in (c) in a cell.