Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3072536 A1 2019/02/21

(21) 3 072 536

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2018/08/16
(87) Date publication PCT/PCT Publication Date: 2019/02/21
(85) Entrée phase nationale/National Entry: 2020/02/10
(86) N° demande PCT/PCT Application No.: GB 2018/052335
(87) N° publication PCT/PCT Publication No.: 2019/034887
(30) Priorité/Priority: 2017/08/16 (GB1713163.2)

(51) Cl.Int./Int.Cl. *A61K 39/12* (2006.01),
A61P 31/20 (2006.01)

(71) Demandeur/Applicant:
OXFORD UNIVERSITY INNOVATION LTD., GB

(72) Inventeurs/Inventors:
DORRELL, LUCY, GB;
BLIGHT, JOSHUA, GB;
REYES-SANDOVAL, ARTURO, GB

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : VACCIN ANTI-VPH
(54) Title: HPV VACCINE

Figure 1

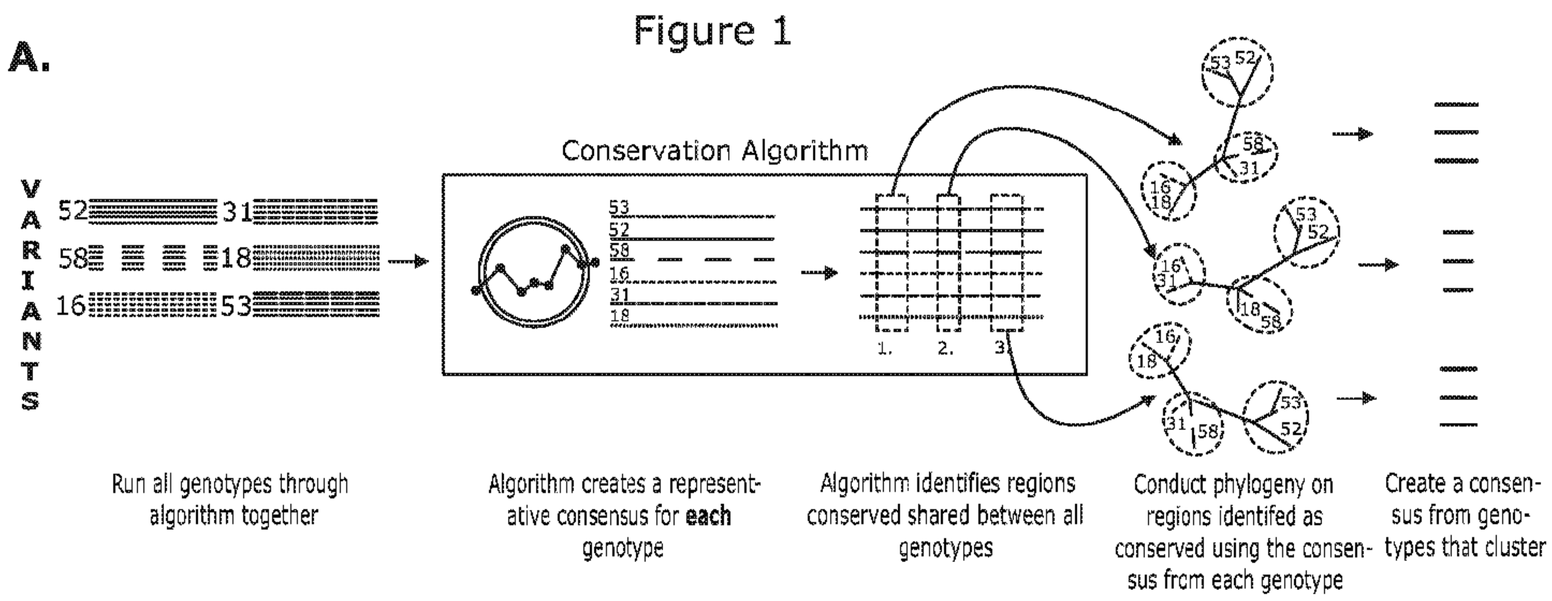
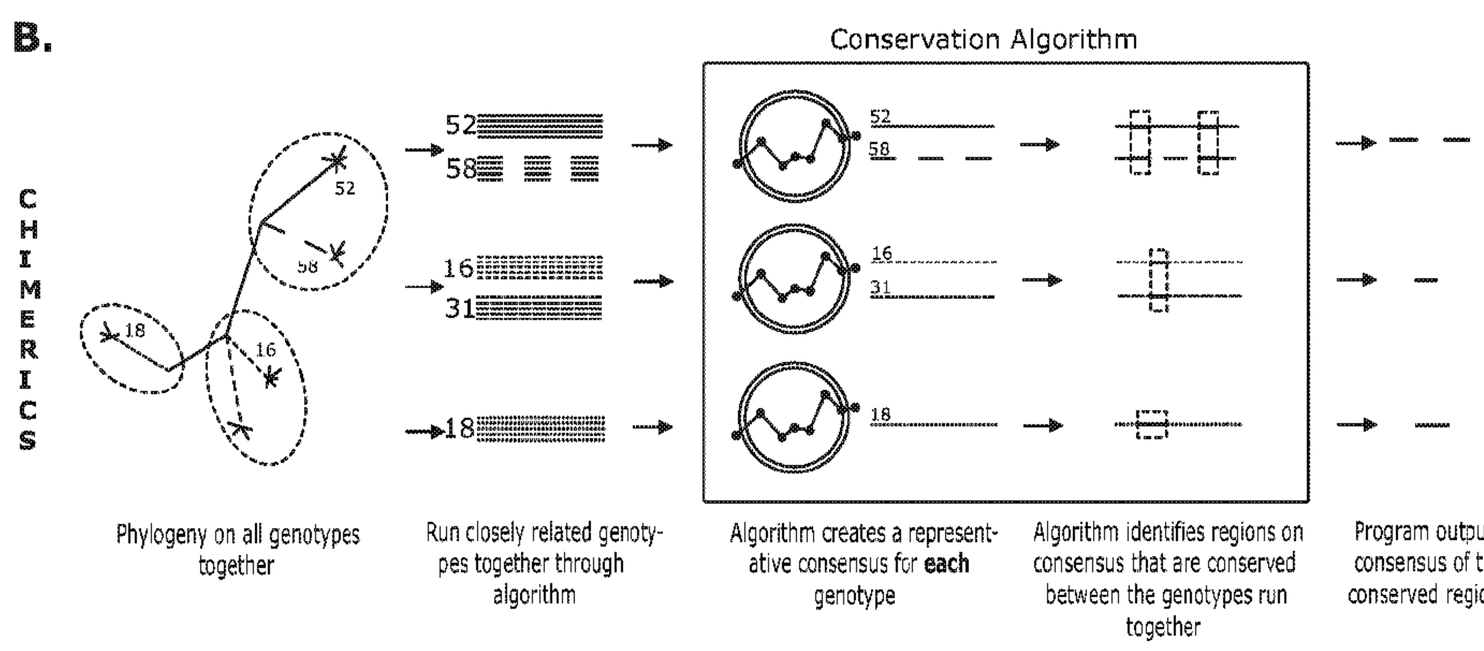


Figure 1 continued



(57) Abrégé/Abstract:
The invention relates to a nucleic acid encoding a polypeptide comprising a plurality of conserved peptide sequences, or variants thereof, wherein the conserved sequences are conserved across one or more HPV genotypes 16, 18, 31, 52, 53, and 58; and wherein the polypeptide comprises a conserved peptide sequence of each of the HPV proteins E1, E2, E4, E5, E6, and E7; and associated vaccines, viral vectors, treatment and prophylaxis.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 February 2019 (21.02.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/034887 A1

(51) International Patent Classification:
A61K 39/12 (2006.01) *A61P 31/20* (2006.01)

(21) International Application Number:
PCT/GB2018/052335

(22) International Filing Date:
16 August 2018 (16.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1713163.2 16 August 2017 (16.08.2017) GB

(71) Applicant: **OXFORD UNIVERSITY INNOVATION LTD.** [GB/GB]; Buxton Court, 3 West Way, Oxford Oxfordshire OX2 0JB (GB).

(72) Inventors: **DORRELL, Lucy**; c/o Nuffield Department of Medicine, University of Oxford, Old Road Campus, Headington, Oxford Oxfordshire OX3 7BN (GB). **BLIGHT, Joshua**; c/o Nuffield Department of Medicine, University of Oxford, Old Road Campus, Headington, Oxford Oxfordshire OX3 7BN (GB). **REYES-SANDOVAL, Arturo**; c/o Nuffield Department of Medicine, University of Oxford, Old Road Campus, Headington, Oxford Oxfordshire OX3 7BN (GB).

(74) Agent: **BARKER BRETTELL LLP**; Medina Chambers, Town Quay, Southampton Hampshire SO14 2AQ (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: HPV VACCINE

Figure 1

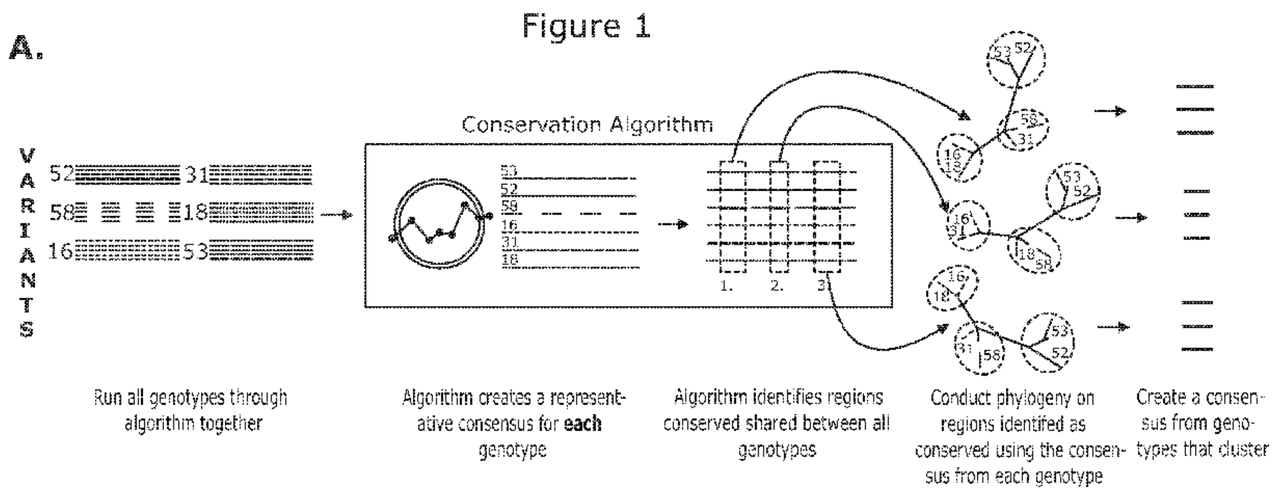
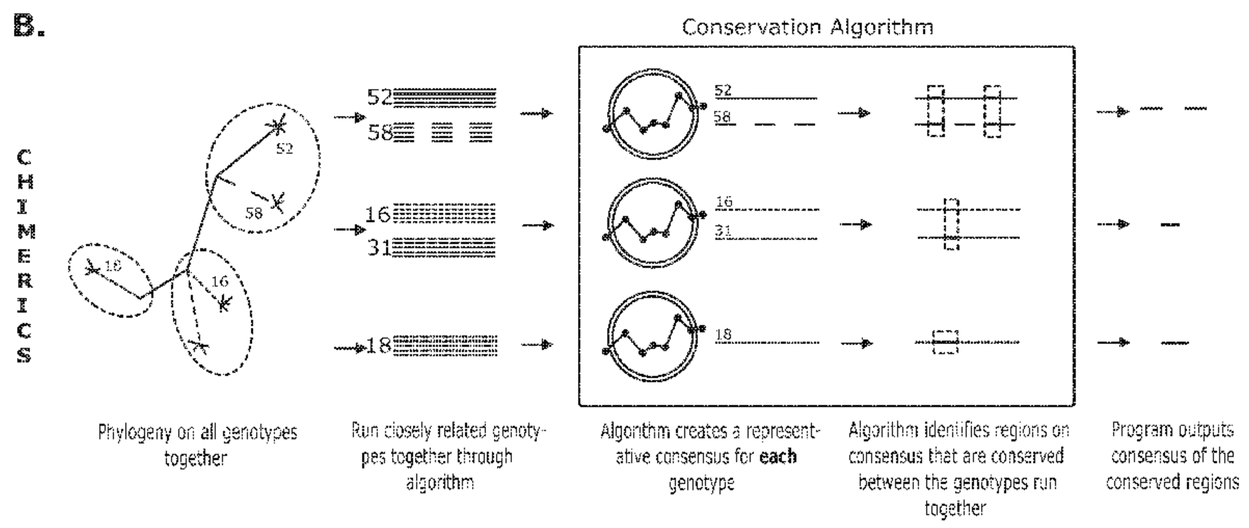


Figure 1 continued



(57) Abstract: The invention relates to a nucleic acid encoding a polypeptide comprising a plurality of conserved peptide sequences, or variants thereof, wherein the conserved sequences are conserved across one or more HPV genotypes 16, 18, 31, 52, 53, and 58; and wherein the polypeptide comprises a conserved peptide sequence of each of the HPV proteins E1, E2, E4, E5, E6, and E7; and associated vaccines, viral vectors, treatment and prophylaxis.

[Continued on next page]

WO 2019/034887 A1

WO 2019/034887 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

HPV VACCINE

This invention relates to viral-vectored vaccines for use in a vaccination against HPV infection.

Human papillomavirus infection is an infection by human papillomavirus (HPV). Most HPV infections cause no symptoms and resolve spontaneously. However, in some cases they persist and this can result in the development of warts or precancerous lesions. The precancerous lesions increase the risk of cancer of the cervix, vulva, vagina, penis, anus, mouth, or throat.

There are approximately 0.5 million cases of HPV-attributable cervical cancer that occur annually worldwide, and over half of these are fatal. About 85% of cases occur in low/middle income countries where there is limited or no treatment available. Women who have not received a prophylactic vaccine require 3-yearly screening to identify and treat cervical intra-epithelial neoplasia (CIN). Screening costs the UK National Health Service approximately £175 million annually.

Current therapy for CIN is ablation of abnormal cervical tissue by electrocautery or surgery. There is no current therapy available that eliminates HPV viral infection. Women require multiple follow-up visits after treatment to ensure that there is no recurrence. Therapy is also associated with increased risk of pre-term birth.

HPV vaccines that have been developed for therapy of existing HPV infection include Inovio - VGX-3100 (DNA encoding E6, E7) similarly Genexine (GX-188E); Janssen - Ad26/Ad35 +/-MVA encoding E2, E6, E7 fusion protein; Synthetic long peptides (E6, E7) and similar eg. PepCan, GTL001; Advaxis - ADXS-HPV; and *L. Monocytogenes* encoding E7. However, such developed vaccines have been targeted to HPV16 and 18 only, have safety concerns, and/or are low efficiency. For example, the net efficacy of VGX-3100 was 18% in a phase IIb randomised controlled trial (48% in vaccine arm vs. 30% in placebo arm).

What is needed is a vaccine that is safe, easy to deliver and to have greater efficacy than the therapeutic vaccine candidates tested to date. Therefore, an aim of the present invention is to provide an improved vaccine for HPV infection.

According to a first aspect of the invention, there is provided a nucleic acid encoding a polypeptide comprising a plurality of conserved peptide sequences, or variants thereof,

wherein the conserved sequences are conserved across one or more HPV genotypes 16, 18, 31, 52, 53, and 58; and

wherein the polypeptide comprises a conserved peptide sequence of each of the HPV proteins E1, E2, E4, E5, E6, and E7.

The invention advantageously provides a novel alternative and safer approach to vaccination whereby T cells can be induced to the relatively conserved antigens of the virion. The use of specially designed conserved viral segments from the non-structural proteins can provide protection against multiple important genotypes.

In one embodiment, the polypeptide is a fusion polypeptide. The polypeptide may not be a wild-type polypeptide. The polypeptide may be synthetic/artificial, for example, the polypeptide may not exist in nature. In one embodiment, the polypeptide may not comprise a complete gene sequence. The polypeptide may consist essentially of conserved peptide sequences. In another embodiment, the polypeptide may consist essentially of conserved peptide sequences and a peptide adjuvant sequence. In another embodiment, the polypeptide may consist essentially of conserved peptide sequences and one or more linkers therebetween. In another embodiment, the polypeptide may consist essentially of conserved peptide sequences, a peptide adjuvant sequence and one or more linkers therebetween. In one embodiment, the polypeptide is a recombinant polypeptide, such as a recombinant fusion polypeptide.

The term “fusion polypeptide” used herein is understood to mean a polypeptide comprising a combination of sequences derived from different gene products (for example different HPV proteins) or combinations of sequences from the same gene product (for example a single HPV protein), wherein the sequences are from distinct/separate regions of the wild-type gene product. For example the fusion polypeptide may comprise combinations of sequences which are normally separated by other sequence segments in wild-type, and the separating sequence(s) have been removed.

The term "conserved peptide sequence" or "conserved segment" used herein is defined as a sequence that is conserved in one or more genotypes, as defined below. Prior to assessment of conservation all available full-length sequences for HPV proteins E1, E2, E4, E5, E6 and E7 from genotypes 16, 18, 31, 52, 53 and 58 were collected from the NCBI Protein database (accessed 2014) and used as input for the approach of the invention. All available sequences were used to ensure the selected conserved peptide sequences would equally represent the whole environmental population (See Table 1). Conserved peptide sequences were identified using the 'variant' approach (Figure 1A); all genotypes were aligned and sequences within each genotype weighted prior to conservation assessment to ensure equal representation of genotype diversity and thus ensure the vaccine candidates were representative of the whole environmental population. Conservation within genotypes (intra-genotype conservation) was then assessed using a 15 amino acid sliding window, whereby for each window a conservation value was determined based on combining the amino acid prevalence within the window and weighting value of each sequence to identify fragments conserved within each genotype, and a normalised intra-genotype consensus created for each window. 'Normalised consensus' meaning an amino acid sequence that represented the weighted set of genotype sequence, not the most common amino acid at each position. To be classed as conserved the window must have a conservation value within the first quartile of all window conservation values for the protein. Subsequently, conserved intra-genotype windows at the same position across all genotypes were identified independent of the percentage identity of shared intra-genotype normalised consensus between genotypes (inter-genotype conservation). A phylogeny was then created of the resultant regions and tree ingroup sequences combined to create an inter-genotype consensus with a high level of shared consensus identity. In this case 'inter-genotype consensus' refers to a consensus created using the normalised consensus created from each genotype. In some scenarios, a 'modified variant' was created where conserved intra-genotype windows at the same position across all proteins were identified which shared greater than 60% shared intra-genotype normalised consensus percentage identity between genotypes.

If the identified inter-serotype fragments from tree ingroups had a percentage identity less than 60%, the sequences were classed as "highly divergent", in contrast to "less divergent" sequences used in the variant approach. For proteins highly divergent between genotypes a 'chimeric' approach was used to identify conserved peptide

sequences (Figure 1B); a phylogeny was created and only genotype ingroups were aligned together and intra-genotype conserved assessed. Therefore, inter-genotype conservation was only assessed between ingroups and intra-serotype conserved windows with greater than 60% shared consensus identity across genotypes selected. In some cases genotypes were run as 'chimerics' but inter-genotype conservation was not assessed ('chimeric-variants').

The plurality of conserved peptide sequences may comprise 10 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 15 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 20 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 25 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 30 or more conserved peptide sequences. The plurality of conserved sequences may comprise 35 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 40 or more conserved peptide sequences. The plurality of conserved peptide sequences may comprise 45 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 50 or more conserved peptide sequences. In another embodiment, the plurality of conserved peptide sequences may comprise 55 or more conserved peptide sequences. In one embodiment, the plurality of conserved peptide sequences consists of about 56 or more conserved peptide sequences. In one embodiment, the plurality of conserved peptide sequences consists of about 57 or more conserved peptide sequences. In one embodiment, the plurality of conserved peptide sequences consists of about 58 or more conserved peptide sequences. In one embodiment, the plurality of conserved peptide sequences consists of about 59 or more conserved peptide sequences. In one embodiment, the plurality of conserved peptide sequences consists of about 59 conserved peptide sequences.

The encoded polypeptide may comprise at least 3 different conserved peptide sequences of each of the HPV proteins E1, E2, E4, E5, E6, and E7. The encoded polypeptide may comprise at least 3 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 3 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least

3 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 3 different conserved peptide sequences of HPV protein E5. The encoded polypeptide may comprise at least 3 different conserved peptide sequences of HPV protein E6. The encoded polypeptide may comprise at least 3 different conserved peptide sequences of HPV protein E7.

The encoded polypeptide may comprise at least 4 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 4 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 4 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 4 different conserved peptide sequences of HPV protein E6. The encoded polypeptide may comprise at least 4 different conserved peptide sequences of HPV protein E7.

The encoded polypeptide may comprise at least 5 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 5 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 5 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 5 different conserved peptide sequences of HPV protein E6.

The encoded polypeptide may comprise at least 6 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 6 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 6 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 6 different conserved peptide sequences of HPV protein E6.

The encoded polypeptide may comprise at least 7 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 7 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 7 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 7 different conserved peptide sequences of HPV protein E6.

The encoded polypeptide may comprise at least 8 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 8 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 8 different conserved peptide sequences of HPV protein E4. The encoded polypeptide may comprise at least 8 different conserved peptide sequences of HPV protein E6.

The encoded polypeptide may comprise at least 9 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 9 different conserved peptide sequences of HPV protein E2. The encoded polypeptide may comprise at least 9 different conserved peptide sequences of HPV protein E4.

The encoded polypeptide may comprise at least 10 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 10 different conserved peptide sequences of HPV protein E2.

The encoded polypeptide may comprise at least 11 different conserved peptide sequences of HPV protein E1. The encoded polypeptide may comprise at least 11 different conserved peptide sequences of HPV protein E2.

The encoded polypeptide may comprise at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 different conserved peptide sequences of HPV protein E2.

The encoded polypeptide may comprise or consist of 11 or more different conserved peptide sequences of HPV protein E1, 24 or more different conserved peptide sequences of HPV protein E2, 9 or more different conserved peptide sequences of HPV protein E4, 3 or more different conserved peptide sequences of HPV protein E5, 8 or more different conserved peptide sequences of HPV protein E6, and 4 or more different conserved peptide sequences of HPV protein E7.

The plurality of conserved peptide sequences may be derived from distinct regions of sequence relative to each other (i.e. not-naturally concurrent). For example, reference to "different conserved peptide sequences" may comprise sequences that are derived from distinct regions of wild-type sequence relative to each other (i.e. not-naturally concurrent). For example, in the wild-type genotype the conserved sequences may be

separated in the wild-type genotypes by variable/non-conserved sequences. The plurality of conserved peptide sequences may not, or may not significantly, overlap with each other. Two or more, or all, of the plurality of conserved peptide sequences may be directly joined together in the polypeptide, for example not comprising any non-conserved/variable residues therebetween. The polypeptide sequence may not be found in nature. The polypeptide may not comprise non-conserved sequences or residues. The conserved peptide sequences may not be distanced apart by more than 1, 2, 3, 4, or 5 residues in the polypeptide sequence, for example in embodiments where there are linker/junction residues between the conserved peptide sequences. Alternatively, the conserved peptide sequences may not be distanced apart by more than 6, 7, 8, 9, or 10 residues in the polypeptide sequence, for example in embodiments where there are linker/junction residues between the conserved peptide sequences. The polypeptide may not comprise non-conserved sequences longer than 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids.

In one embodiment, linker residues may be provided between one or more, or all, conserved peptide sequences (e.g. providing junctions between the conserved peptide sequences in the polypeptide). The linker residues may comprise random amino acid sequences, or amino acids that have been selected to be non-immunogenic based on epitope prediction computer programs or experiments in animal models. For example, a linker may not be considered if it is predicted or known to be an epitope (i.e. in order to avoid an immune response to epitopes, e.g. artificial epitopes, not found in HPV. The linker may be flexible. The linker may comprise or consist of K, G, P, A or S amino acid residues, or combinations thereof. In one embodiment, the linker may comprise or consist of G and/or P amino acid residues. In one embodiment, the linker may comprise or consist of one or more alanine (A) amino acid residues. The linker residues may be between 1 and 10 amino acids in length. In another embodiment, the linker residues may be between 2 and 8 residues in length. In another embodiment, the linker residues may be between 1 and 6 residues in length. The conserved peptide sequences may be distanced apart by between 1 and 10 residues in the polypeptide sequence, for example in embodiments where there are linker/junction residues between the conserved peptide sequences.

In one embodiment, the polypeptide may consist essentially of conserved peptide sequences and one or more linkers, optionally wherein the one or more linkers are disposed between adjacent conserved peptide sequence.

The conserved peptide sequences may be selected from any of the group comprising SEQ ID NOs: 1 to 59; variants thereof or combinations thereof. In another embodiment, the conserved peptide sequences may be selected from any of the group comprising SEQ ID NOs: 1 to 59; variants thereof or combinations thereof, in any order. In one embodiment, the conserved peptide sequences may consist of the group comprising SEQ ID NOs: 1 to 59.

The polypeptide may comprise one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11; one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35; one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44; one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47; one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55; and one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59.

The polypeptide may comprise two or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11; two or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35; two or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44; two or more conserved E5 sequence(s) selected from any of SEQ ID NOs: 45-47; two or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55; and two or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59.

The polypeptide may comprise three or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11; three or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35; three or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44; three or more conserved E5 sequence(s) selected from any of SEQ ID NOs: 45-47; three or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55; and three or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59.

The conserved sequences are conserved across one or more of HPV genotypes 16, 18, 31, 52, and 58. The conserved sequences are conserved across all of HPV genotypes 16, 18, 31, 52, and 58.

The polypeptide may comprise:

one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E1 sequence;

one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E2 sequence;

one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E4 sequence;

one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E5 sequence;

one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E6 sequence; and

one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E7 sequence.

The polypeptide may comprise:

one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E1 sequence;

one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E2 sequence;

one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E4 sequence;

one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E5 sequence;

one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E6 sequence; and

one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented by at least one conserved E7 sequence.

The polypeptide may comprise:

two or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E1 sequences;

two or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E2 sequences;

two or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E4 sequences;

two or more conserved E5 sequence(s) selected from any of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E5 sequences;

two or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E6 sequences; and

two or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E7 sequences.

The polypeptide may comprise:

two or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E1 sequences;

two or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E2 sequences;

two or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E4 sequences;

two or more conserved E5 sequence(s) selected from any of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E5 sequences;

two or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E6 sequences; and

two or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E7 sequences.

The polypeptide may comprise:

three or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E1 sequences;

three or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E2 sequences;

three or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E4 sequences;

three conserved E5 sequence(s) selected from SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E5 sequences;

three or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E6 sequences; and

three or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented in the group of conserved E7 sequences.

The polypeptide may comprise:

three or more conserved E1 sequence(s) selected from any of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E1 sequences;

three or more conserved E2 sequence(s) selected from any of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E2 sequences;

three or more conserved E4 sequence(s) selected from any of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E4 sequences;

three conserved E5 sequence(s) selected from SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E5 sequences;

three or more conserved E6 sequence(s) selected from any of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E6 sequences; and

three or more conserved E7 sequence(s) selected from any of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, and 58 are represented in the group of conserved E7 sequences.

Reference to "each of the genotypes 16, 18, 31, 52, 53, and 58 are represented" or "each of the genotypes 16, 18, 31, 52, and 58 are represented" is intended to mean that each of the identified genotypes has been used to define at least one consensus sequence of a conserved peptide sequence. Therefore, a given group may comprise a conserved peptide from each genotype, or a conserved peptide may be derived from a consensus of two or more genotypes. If sequence identities are sufficiently similar, all the genotypes 16, 18, 31, 52, 53, and 58 or 16, 18, 31, 52, and 58 could be represented by a single conserved peptide sequence, which may be a consensus of all the genotypes 16, 18, 31, 52, 53, and 58 or 16, 18, 31, 52, and 58 respectively. However, due to differences in sequence identities, a single conserved peptide may not be able to represent a consensus sequence from all genotypes 16, 18, 31, 52, 53, and 58 or 16, 18, 31, 52, and 58 and instead two or more conserved peptide sequences are required to cover/represent all the genotypes 16, 18, 31, 52, 53, and 58 or 16, 18, 31, 52, and 58. For example (for illustrative purposes only), one conserved E6 peptide sequence may represent E6 genotypes 16 and 18, another may represent E6 genotype 52, and a

third may represent E6 genotypes 53 and 58, such that all three conserved E6 peptide sequences in a group represent all E6 genotypes 16, 18, 31, 52, 53, and 58 or 16, 18, 31, 52, and 58.

The nucleic acid may comprise or consist of the sequence of SEQ ID NO: 60, or variants thereof. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 60, or variants thereof, and without encoding the TPA lead sequence. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 60, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 65, or variants thereof.

Variants of the nucleic acid may comprise or consist of a sequence having at least 80% identity with SEQ ID NO: 60 or 65. Alternatively, variants of the nucleic acid may comprise or consist of a sequence having at least 85% identity with SEQ ID NO: 60 or 65. Variants of the nucleic acid may comprise or consist of a sequence having at least 90% identity with SEQ ID NO: 60 or 65. Variants of the nucleic acid may comprise or consist of a sequence having at least 95% identity with SEQ ID NO: 60 or 65. Variants of the nucleic acid may comprise or consist of a sequence having at least 98% identity with SEQ ID NO: 60 or 65. Variants of the nucleic acid may comprise or consist of a sequence having at least 99% identity with SEQ ID NO: 60 or 65. The skilled person will understand that a variant of the nucleic acid may include redundant codon variants that encode the same peptide as SEQ ID NO: 60 or 65.

The nucleic acid may comprise or consist of the sequence of SEQ ID NO: 62, or variants thereof. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 62, or variants thereof, and without encoding the TPA lead sequence. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 62, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence.

Variants of the nucleic acid may comprise or consist of a sequence having at least 80% identity with SEQ ID NO: 62. Alternatively, variants of the nucleic acid may comprise or consist of a sequence having at least 85% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least

90% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least 95% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least 98% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least 99% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least 99.5% identity with SEQ ID NO: 62. Variants of the nucleic acid may comprise or consist of a sequence having at least 99.9% identity with SEQ ID NO: 62. The skilled person will understand that a variant of the nucleic acid may include redundant codon variants that encode the same viral vector and/or peptide as SEQ ID NO: 62.

The nucleic acid may comprise or consist of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof, and without encoding the TPA lead sequence. In another embodiment, the nucleic acid may comprise or consist of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence.

Variants of the nucleic acid may comprise or consist of a sequence having at least 80% identity with SEQ ID NO: 71, 73 or 75. Alternatively, variants of the nucleic acid may comprise or consist of a sequence having at least 85% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 90% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 95% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 98% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 99% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 99.5% identity with SEQ ID NO: 71, 73 or 75. Variants of the nucleic acid may comprise or consist of a sequence having at least 99.9% identity with SEQ ID NO: 71, 73 or 75. The skilled person will understand that a variant of the nucleic acid may include redundant codon variants that encode the same viral vector and/or peptide as SEQ ID NO: 71, 73 or 75.

The polypeptide may comprise or consist of the sequence of SEQ ID NO: 61, or variants thereof. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 61, or variants thereof, and without the TPA lead sequence. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 61, or variants thereof, with a different/alternative peptide adjuvant than the TPA lead sequence. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 66, or variants thereof.

The polypeptide may comprise or consist of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof, and without the TPA lead sequence. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof, with a different/alternative peptide adjuvant than the TPA lead sequence. In another embodiment, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof.

In one embodiment, the polypeptide may consist essentially of conserved peptide sequences and a peptide adjuvant. In one embodiment, the polypeptide may consist essentially of conserved peptide sequences, one or more linkers, and a peptide adjuvant. The one or more linkers may be disposed between adjacent conserved peptide sequence. The peptide adjuvant may be N-terminal.

Variants of the polypeptide may comprise or consist of a sequence having at least 80% identity with SEQ ID NO: 61, 66, 72, 74 or 76. Alternatively, variants of the polypeptide may comprise or consist of a sequence having at least 85% identity with SEQ ID NO: 61, 66, 72, 74 or 76. Variants of the polypeptide may comprise or consist of a sequence having at least 90% identity with SEQ ID NO: 61, 66, 72, 74 or 76. Variants of the polypeptide may comprise or consist of a sequence having at least 95% identity with SEQ ID NO: 61, 66, 72, 74 or 76. Variants of the polypeptide may comprise or consist of a sequence having at least 98% identity with SEQ ID NO: 61, 66, 72, 74 or 76. Variants of the polypeptide may comprise or consist of a sequence having at least 99% identity with SEQ ID NO: 61, 66, 72, 74 or 76.

Variants of conserved peptide sequences may comprise or consist of a truncated sequence of the conserved peptide sequences. For example, any one or more of the sequences of SEQ ID NOs: 1 to 59, herein may be truncated and still provide immunogenicity in the polypeptide. The truncated sequence may comprise a sufficient number of amino acids to form a recognisable epitope (e.g. at least the minimum number of residues for specific T cell recognition) from a sequence within any one of the sequences of SEQ ID NOs: 1 to 59. The truncated sequence may comprise at least 7 amino acids of the sequences of SEQ ID NOs: 1 to 59. Alternatively, the truncated sequence may comprise at least 8 amino acids of the sequences of SEQ ID NOs: 1 to 59. Alternatively, the truncated sequence may comprise at least 9, 10, 11 or 12 amino acids of the sequences of SEQ ID NOs: 1 to 59. Multiple truncated sequences may be provided within one of the conserved peptide sequences of SEQ ID NOs: 1 to 59.

In one embodiment, any one of the conserved peptide sequences of SEQ ID NOs: 1 to 59 may be varied, for example by residue substitution, addition or deletion. In another embodiment, some or all of the conserved peptide sequences of SEQ ID NOs: 1 to 59 may be varied, for example by residue substitution, addition or deletion. The variant conserved peptide sequences may still function to provide recognisable HPV epitopes. The skilled person will understand that natural variation exists in any given population and that these variants may have some sequence variation with the consensus sequence, or example patient sequences provided in SEQ ID NOs: 1 to 59. Therefore, a variant conserved peptide sequence may have at least 70% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 74% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 75% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 79% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 80% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 82% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 83% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 85% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide

sequence may have at least 88% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 90% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 92% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 95% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 98% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 99% sequence identity with any one of SEQ ID NOs: 1 to 59. In another embodiment, a variant conserved peptide sequence may have at least 99.5% sequence identity with any one of SEQ ID NOs: 1 to 59.

Reference to sequence "identity" used herein may refer to the percentage identity between two aligned sequences using standard NCBI BLASTp parameters (http://blast.ncbi.nlm.nih.gov).

The conserved peptide sequences may vary in length, with the minimum length being defined as the minimum number of residues required to form a recognisable epitope. Therefore, the conserved peptide sequence may be from about 7 to 250 amino acids in length, or more. For example, at least one conserved peptide sequence may be at least about 7 amino acids in length. In another embodiment, at least one conserved peptide sequence may be at least about 8 amino acids in length. In another embodiment, at least one conserved peptide sequence may be at least about 10 amino acids in length. In another embodiment, at least one conserved peptide sequence may be at least about 15 amino acids in length. In another embodiment, at least one conserved peptide sequence may be at least about 20 amino acids in length. In another embodiment, at least one conserved peptide sequence may be at least about 30 amino acids in length. In one embodiment, at least one conserved peptide sequence may be between about 20 and about 220 amino acids in length. In one embodiment, at least one conserved peptide sequence may be no more than about 300 amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 250 amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 200 amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 150

amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 100 amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 55 amino acids in length. In another embodiment, at least one conserved peptide sequence may be no more than about 54 amino acids in length.

The conserved peptide sequences may be an average length of between about 15 and about 50 amino acids in a population of conserved peptide sequences.

In some embodiments of the invention, the polypeptide may further comprise a peptide adjuvant, such as a TPA (tissue plasminogen activator) sequence, or functional variants thereof. The TPA may comprise or consist of the sequence: MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRR (SEQ ID NO: 63), or a functional variant thereof. In one embodiment, the peptide adjuvant may comprise a Shark invariant chain, for example of the sequence SLLWGGVTVLAAMLIAGQVASSVVFLV (SEQ ID NO: 64), or a functional variant thereof. The peptide adjuvant may be N-terminal on the polypeptide of the invention. A functional variant of a peptide adjuvant may be a truncated or mutated peptide variant, which can still function as an adjuvant, for example a truncated or mutated variant of the TPA or shark invariant chain, which still function as an adjuvant. The skilled person will appreciate that 1, 2, 3, 4, 5 or more amino acid residues may be substituted, added or removed without affecting function. For example, conservative substitutions may be considered. In embodiments, where a peptide adjuvant is provided (or encoded as appropriate), there may additionally be provided a linker sequence provided (or encoded) between the peptide adjuvant and the first conserved peptide sequence. In embodiments without the peptide adjuvant, the first linker sequence may not be provided.

Combinations of nucleic acids may encode different polypeptides according to the invention may be provided as a vaccine. For example, a prime and/or boost vaccine formulation may comprise nucleic acid or viral vector encoding two or more polypeptides of the invention, which may be different relative to each other.

The nucleic acid may be used in a vaccine in combination with another therapeutically or prophylactically active ingredient. The nucleic acid may be used in a vaccine in combination with an adjuvant.

According to another aspect of the invention there is provided a composition comprising a plurality of different nucleic acids according to the invention, optionally wherein the composition is a pharmaceutically acceptable composition.

According to another aspect of the invention there is provided a polypeptide encoded by the nucleic acid according to the invention herein.

In one embodiment the polypeptide is an isolated polypeptide. The polypeptide, nucleic acid encoding the polypeptide, or associated viral particle may be provided in a pharmaceutically acceptable carrier.

The nucleic acid may be a plasmid vector for vaccination. The nucleic acid may comprise viral vector sequences.

According to another aspect of the invention there is provided a viral vector comprising the nucleic acid according to the invention herein.

The viral vector may comprise a virus. The viral vector may comprise an adenovirus, such as a human or simian adenovirus. The viral vector may comprise an adenovirus when used in a prime vaccine of a prime boost regime. The viral vector may comprise ChAdOx1 (a group E simian adenovirus, like the AdCh63 vector used safely in malaria trials) or ChAdOx2 (as described in Morris *et al* 2016. Future Virol 11(9), pp. 649-659). The ChAdOx2 sequence may comprise or consist of the sequence described herein (e.g. SEQ ID NOs: 67+68). The viral vector may comprise AdCh63. The viral vector may comprise AdC3 or AdH6. The viral vector may be a human serotype. The viral vector may comprise Modified Vaccinia Ankara (MVA). The viral vector may comprise F11 MVA (e.g. MVA with the nucleic acid construct insert at the F11 locus). The nucleic acid of the invention (the HPV vaccine construct insert) may be inserted at the TK locus of parental MVA virus under the control of the p7.5 promoter, for example through recombination with the p7.5 MVA shuttle plasmid (SEQ ID NO: 158). The nucleic acid may comprise the sequence of SEQ ID NO: 158 with the

nucleic acid vaccine construct insert as provided in SEQ ID NO: 158 (underlined), or with an alternative nucleic acid vaccine construct in accordance with the invention herein. In another embodiment, the nucleic acid of the invention (the HPV vaccine construct insert) may be inserted at the F11 locus of parental MVA virus under the control of the F11 promoter, for example through recombination with the F11 shuttle plasmid (SEQ ID NO: 159). The nucleic acid may comprise the sequence of SEQ ID NO: 159 with the nucleic acid vaccine construct insert as provided in SEQ ID NO: 159 (underlined), or with an alternative nucleic acid vaccine construct in accordance with the invention herein. The MVA sequence may comprise or consist of the sequence described herein (e.g. SEQ ID NOs: 69+70). The viral vector may comprise MVA when used as a vaccine boost in a prime boost regime. The viral vector may comprise Adeno-associated virus (AAV) or lentivirus. The viral vector may be an attenuated viral vector. The polypeptide sequence of the invention may be cloned into any suitable viral vector that is known to elicit good immune response. Suitable viral vectors have been described in Dicks et al (Vaccine. 2015 Feb 25;33(9):1121-8. doi: 10.1016/j.vaccine.2015.01.042. Epub 2015 Jan 25), Antrobus et al (Mol Ther. 2014 Mar;22(3):668-74. doi: 10.1038/mt.2013.284. Epub 2013 Dec 30.), and (Warimwe et al. (Virol J. 2013 Dec 5;10:349. doi: 10.1186/1743-422X-10-349), which are incorporated herein by reference.

According to another aspect of the invention there is provided a composition comprising one or more of:

-the polypeptide according to the invention;

-the nucleic acid according to the invention; and

-the viral vector according to the invention.

The composition may be immunogenic, for example in a mammal, such as a human. The composition may comprise a pharmaceutically acceptable carrier. The composition may be a pharmaceutical composition comprising a pharmaceutically acceptable carrier. The composition may be for use in the prophylaxis or treatment of HPV infection.

According to another aspect of the invention there is provided a method of treatment or prophylaxis of HPV infection comprising the administration of:

-the polypeptide according to the invention;

-the nucleic acid according to the invention;
-the composition according to the invention or
-the viral vector according to the invention.

The method of treatment or prophylaxis of HPV infection may be a method of vaccination.

According to another aspect of the invention there is provided an agent for use in the prophylaxis or treatment of HPV infection, the agent comprising or consisting of:

-the polypeptide according to the invention;
-the composition according to the invention;
-the nucleic acid according to the invention; or
-the viral vector according to the invention.

In one embodiment, the treatment or prophylaxis of HPV infection comprises the treatment or prophylaxis of an anogenital HPV-driven lesion, such as anal, vulval, vaginal, or penile intraepithelial neoplasia. Additionally or alternatively, the treatment or prophylaxis of HPV infection comprises the treatment or prophylaxis of an oropharyngeal lesion that is caused by HPV.

According to another aspect of the invention there is provided the polypeptide according to the invention; the composition according to the invention; the nucleic acid according to the invention; or the viral vector according to the invention; for use in, or as, a vaccine.

According to another aspect of the invention there is provided a vaccine comprising the nucleic acid of the invention; the polypeptide according to the invention; the composition according to the invention; or the viral vector according to the invention.

The vaccine may be a prime vaccine. The vaccine may be a boost vaccine. Where a boost vaccine is provided following a prime vaccine, the polypeptide may be different. For example, the polypeptide may comprise a re-ordered sequence of conserved peptide sequences. The conserved peptide sequences may be identical, but the order in which they are provided in the polypeptide may be changed. Therefore, the invention herein provides any of the sequences/embodiments of the invention wherein the order

in which conserved peptide sequences are provided may be changed. Such embodiments may also include re-ordered or differed linker/junction sequences.

Advantageously, the re-ordering of the conserved peptide sequences of the polypeptide between prime and boost vaccines can avoid the provision of "false" epitopes formed across junctions of one conserved peptide sequence with another conserved peptide sequence. i.e. the same junction may not occur in the re-ordered polypeptide.

According to another aspect of the invention, there is provided a nucleic acid or polypeptide according to the invention for use in, or as, a vaccine.

According to another aspect of the invention, there is provided a prime boost vaccination kit comprising

-a prime vaccination according to the invention;

-a boost vaccination according to the invention.

The prime and boost vaccinations may be different. The prime and boost vaccination may differ in the polypeptide sequence. The prime and boost vaccination may comprise different viral vectors (i.e. from different virus families such as MVA vs adenovirus).

According to another aspect of the invention, there is provided a composition comprising a nucleic acid according to the invention herein, and a pharmaceutically acceptable carrier.

The composition may not comprise wild-type HPV. The composition may not comprise full length HPV protein sequence. The viral vector or nucleic acid may not encode non-conserved protein/peptide sequence of HPV.

The use may be with a pharmaceutically acceptable carrier. Additionally or alternatively, the use may be with an adjuvant.

The term "immunogenic", when applied to the nucleic acid, polypeptide or composition of the present invention means capable of eliciting an immune response in a human or animal body. The immune response may be protective.

The term "protective" means prevention of a disease, a reduced risk of disease infection, transmission and/or progression, reduced severity of disease, a cure of a condition or disease, an alleviation of symptoms, or a reduction in severity of a disease or disease symptoms.

The term “prophylaxis” means prevention of or protective treatment for a disease. The prophylaxis may include a reduced risk of disease infection, transmission and/or progression, or reduced severity of disease.

The term “treatment”, means a cure of a condition or disease, an alleviation of symptoms, or a reduction in severity of a disease or disease symptoms.

The skilled person will understand that optional features of one embodiment or aspect of the invention may be applicable, where appropriate, to other embodiments or aspects of the invention.

Embodiments of the invention will now be described in more detail, by way of example only, with reference to the accompanying drawings.

Figure 1 - Chimeric and Variant methods used to create HPV candidates.

Figure 2 - Conserved regions with resultant variants below.

Figure 3 - A) Regions identified as conserved in the two genotypes used to form chimerics. B) Conservation plot of Modified variant.

Figure 4 - A) Regions identified as conserved in the two genotypes used to form chimerics. B) Conservation plot of variants

Figure 5 - Regions identified as conserved in the two genotypes used to form chimerics

Figure 6 - Regions identified as conserved in the two genotypes used to form chimeric-variants

Figure 7 - Regions identified as conserved in the two genotypes used to form chimeric

Figure 8 - High frequencies of vaccine-specific T cells are induced following prime boost vaccination. IFNγ Elispot performed on PBMCs from C57BL/6 mice (six/group) primed intramuscularly with DNA-HPV, MVA-HPV or ChAdOx1-5GHPV3 and then boosted intramuscularly with a heterologous or homologous vaccine two weeks later. PBMCs were collected by tail vein bleed two weeks post prime and one and two weeks post boost.

Figure 9 – Prime boost vaccination is capable of inducing responses to all antigens encoded in the immunogen. PBMCs were collected at two weeks boost and used in an IFNγ Elispot with peptides spanning the entire immunogen sequence, pooled according to protein source. Peptides spanning the E2 region of the immunogen were split into two pools because of the large number of peptides and peptides for regions spanning E4 and E5 were combined into one pool.

Figure 10 - Vaccine-specific CD8+ T cells can be detected at high frequencies six weeks after ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost. 10a IFNγ Elispot using PBMCs collected by tail vein bleed two, three, four and six weeks post boost (DM; DNA-5GHPV3 prime MVA-5GHPV3 boost, CM; ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost). 10b ICS performed on PBMCs obtained one week and six weeks post ChAdOx1-5GHPV3 prime, MVA-5GHPV3 boost. PBMCs stimulated with E6 and E7 peptide pools. Measured IFN-γ, CD107, TNF-α and IL2.

Figure 11 - HPV E6- and E7-specific CD8+ T cell responses are polyfunctional and have cytotoxic potential. PBMCs from a tail vein bleed collected one week post ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost were stimulated with immunodominant peptide pools E6 and E7 and sub pool 22

which is the dominant sub pool within E6. Responding CD8+ T cells predominantly express three functions (CD107, IFNγ and TNFα).

Figure 12 - ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost also primes HPV E6- and E7-specific CD4+ T cell responses that are still detectable six weeks post boost. ICS performed on PBMCs obtained one week and six weeks post ChAdOx1-5GHPV3 prime, MVA-5GHPV3 boost. PBMCs stimulated with E6 and E7 peptide pools. Measured IFN-γ, CD107, TNF-α and IL2.

Figure 13 - Most E6 and E7-specific CD4+ T cells express two functions. PBMCs from a tail vein bleed collected one week post ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost were stimulated with immunodominant peptide pools E6 and E7 and sub pool 22 which is the dominant sub pool within E6. Responding CD4+ T cells predominantly express two functions (Discounting CD107+ monofunctional cells).

Figure 14 - HPV E6- and E7-specific CD8+ T cells can be detected in the cervix. ICS was performed on cervicovaginal lymphocytes isolated from mice two weeks post ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost and stimulated with immunodominant peptide pools E6 and E7 and sub pool 22. Measured IFN-γ, CD107, TNF-α and IL2.

Figure 15 – Cervicovaginal HPV E6- and E7-specific CD8+ T cell responses are polyfunctional. Cervicovaginal lymphocytes collected one week post ChAdOx1-5GHPV3 prime MVA-5GHPV3 boost were stimulated with immunodominant peptide pools E6 and E7 and sub pool 22 which is the dominant sub pool within E6. Responding CD8+ T cells predominantly express three functions (CD107, IFNγ and TNFα).

Figure 16 - Vaccine-induced E6 and E7-specific CD8+ and CD4+ T cells in the cervix are almost exclusively of effector phenotype. Naïve: CD44-CD62L+. Antigen-experienced: Central memory - CD62L+, CD127+; Effector memory - CD62L-, CD127+; Effector - CD62L-, CD127-. Cervical lymphocytes from six mice pooled into three pairs, due to low lymphocyte numbers

Figure 17 - In contrast to cervix, vaccine-induced E6 and E7-specific CD8+ and CD4+ T cells in the spleen comprise effector and effector memory populations.

Figure 18 - Mice vaccinated with ChAdOx1-5GHPV3 prime, MVA-5GHPV3 boost show increased survival over control mice. Mice were inoculated with $5x10^4$ TC-1 cells on day 0 and then primed on day 3 and boosted on day 17. Tumours were measured with digital callipers every two days and mice culled when tumours reached 10mm in any one direction.

Figure 19 – Vaccination of outbred CD1 mice induces high frequency T cell responses. IFNγ Elispot performed on PBMCs from CD1 mice (ten/group) primed intramuscularly with DNA-5GHPV3, MVA-5GHPV3 or ChAdOx1-5GHPV3 and then boosted intramuscularly with a heterologous or homologous vaccine two weeks later. PBMCs were collected by tail vein bleed two weeks post prime and two and three weeks post boost.

Figure 20 - T cell responses in vaccinated CD1 mice are directed across the entire immunogen. PBMCs were collected at two weeks boost and used in an IFNγ Elispot with peptides spanning the entire immunogen sequence, pooled according to protein source. Peptides spanning the E2 region of the immunogen were split into two pools because of the large number of peptides and peptides for regions spanning E4 and E5 were combined into one pool.

Figure 21 - Top panel: Peripheral blood mononuclear cells (PBMC) from 76 women aged 16-24 years were tested for recognition of peptides based on early proteins from high risk HPV (hrHPV) in ex vivo IFN-γ Elispot assays. 'Insert' is a pool of 15-mer peptides overlapping by 11 amino acid spanning the hrHPV transgene. 'Reference' peptides were pools of peptides based on early proteins from HPV16 and HPV52, which were combined as follows: E1/E2, E4/E5 and E6/E7. 'FEC' (flu, EBV and CMV) peptides and PHA (phytohaemaglutinnin) were used as positive controls. The data shown are the spot-forming units (SFU) obtained from peptide-stimulated wells after subtraction of negative control values (mock-stimulated cells). The cut-off for a positive response was set at 25 SFU/million PBMC (derived from the mean of mock-

stimulated values from all donors + 2 standard deviations). Women were tested concurrently for hrHPV DNA on vaginal sampling: 26% tested positive. The data show that 9/76 women with current hrHPV infection or prior exposure recognised HPV sequences encoded in the transgene. Bottom panel: Responses to the insert pool were interrogated further in one responding donor by testing PBMC with sub-pools of the insert pool (left, SFU per million PBMC), followed by individual peptides within the pools (right, SFU per well), thus confirming the presence of a true HPV-specific response. SFU – spot-forming units reported as either per well, which contains 200,000 PBMC or per million PBMC).

Figure 22 - Subpool mapping in C57BL/6 mice (top panel) and CD1 mice (bottom panel) following ChAdOX1-5G-HPV3 prime MVA-5GHPV3 boost. Mice were culled two weeks post boost and splenocytes isolated. Splenocytes used in an IFNγ Elispot assay using subpools that cover the immunogen sequence. Subpools 21, 25 and 26 (for example) contain no HPV53 sequences and still get high magnitude responses, thus providing evidence that a sequence without the HPV53 segments would still be immunogenic.

HPV Immunogen Design

The HPV immunogen is composed of amino acid fragments conserved and geographically representative of the global HPV population at a protein level. Each fragment is created using a conservation algorithm which has been utilised to create either Chimeric or Variant based fragments. The choice being dependent on characteristics of each HPV protein used.

The core conservation algorithm uses a sliding window approach in combination with a normalisation method that accounts for collection bias to identify windows which are conserved both within (intra-genotype) and between (inter-genotype) genotypes. Intra-genotype conserved windows are classed as windows with a conservation value less than one quartile of the sum of all window conservation values for the whole sequence. A normalised representative consensus sequence is created for each genotype during this process. Subsequently inter-genotype windows are identified as windows at the same position in each genotype which are conserved and have a shared consensus identity of > 60%.

A key challenge in creating sequences which represent all the selected genotypes for a region of each protein was that many genotypes showed unique phylogenetic clustering. There is a critical balance between the identification of regions conserved across genotypes and the level of shared % consensus identity between these regions. In many cases the inter-genotype diversity was so extreme that you sacrifice shared % consensus identity. Meaning some genotypes were so distinct from one another it was impossible to create a representative sequence. To solve this, raw sequences were inputted into the core algorithm in one of two ways (Fig 1). (i) Variants; Sequences for all genotypes are input into the algorithm and regions of inter-genotype conservation are identified independent of shared % consensus identity. A phylogeny was created from the resultant regions and ingroup sequences combined to create a consensus with a high level of shared consensus identity (Fig1b). (ii) Chimerics; in some instances the distance between genotypes is so significant that genotypes do not align suitably. Therefore, only genotype ingroups are inputted together into the algorithm, and regions conserved between ingroups with > 60% shared consensus identity are identified (Fig1a).

HPV Candidates

Candidate sequences were created for six HPV proteins (E1-2, E4-7) using input sequences collected from the NCBI protein database, aligned and manually audited to remove incomplete and poor quality sequences (Table 1).

	Genotype					
Protein	*16*	*18*	*31*	*52*	*53*	*58*
E1	126	49	24	27	15	53
E2	195	56	26	32	18	54
E4	161	48	24	24	14	52
E5	197	48	24	24	—	60
E6	1205	78	90	218	131	185
E7	566	70	85	193	114	199

Table 1 | Number of sequences used as input post audit.

E1 Candidate

Approach:	*Variants*	
Fragments:	1	E1_V1_52+58: DEDETAYDSGTDLIDFIDDS (SEQ ID NO: 1) E1_V1_31+16+18: DENENDSDTGEDMVDFIDN (SEQ ID NO: 2) E1_V1_53: DETDEESTESDLDGFIDNS (SEQ ID NO: 3)
	2	Excluded
	3	E1_V3_31+53: AQLADSDSNACAFLK (SEQ ID NO: 4) E1_V3_52+58+18+16: AQLADVNSNAAAFLK (SEQ ID NO: 5)
	4	E1_V4_16+31: NCILLYGAANTGKSLFGMSL (SEQ ID NO: 6) E1_V4_18+52+58: NCLVLCGPANTGKSYFGMSL (SEQ ID NO: 7) E1_V4_53: NCLVIYGPPNTGKSCFAMSL (SEQ ID NO: 8)
	5	E1_V5_16+31+52: WPYLHSRLVVFTFPNPF (SEQ ID NO: 9) E1_V5_18+58: WPYLESRITVFEFPNAF (SEQ ID NO: 10) E1_V5_53: LRYLHSRIHVLQFLNPF (SEQ ID NO: 11)

Identified 5 fragments within the E1 protein with windows conserved at the same position within their respective genotypes. Cladistics identified the most suitable genotype combinations providing high level shared consensus identity (Fig 2).

E2 Candidate

Approach:	*Chimerics*

Fragments:	1 *(16+31)*	E2_C1-1_16+31	NVCQDKILEHYENDSKD (SEQ ID NO: 12)
		E2_C1-2_16+31	ILEHYENDSKDLCDHI (SEQ ID NO: 13)
		E2_C1-3_16+31 (SEQ ID NO: 14)	CDHIDYWKHIRLECAIMYKAR
		E2_C1-4_16+31	IRLECAIMYKAREMGFH (SEQ ID NO: 15)
		E2_C1-5_16+31	QFDGDICNTMHYTNW (SEQ ID NO: 16)
		E2_C1-6_16+31	IYICEDAQCTVVEGQVD (SEQ ID NO: 17)
		E2_C1-7_16+31	KKWEVHAGGQVILCPES (SEQ ID NO: 18)
		E2_C1-8_16+31	GQRRIKRPRSE (SEQ ID NO: 19)
		E2_C1-9_16+31	NCHPNKLL (SEQ ID NO: 20)
		E2_C1-10_16+31	ILKCLRYRFKKHCKL (SEQ ID NO: 21)
		E2_C1-11_16+31	SSTWHWTCHDGKHK (SEQ ID NO: 22)
		E2_C1-12_16+31	WHWTCHDGKHKNAIVTLTY (SEQ ID NO: 23)
	2 *(52+58)*	E2_C1-1_52+58 (SEQ ID NO: 24)	YEADKNDLNAQIEHWKLIRMECAIFYKAKELGIS
		E2_C1-2_52+58 (SEQ ID NO: 25)	ICHQVVPPLAASKAKACQAIELQLALEALNASPY
		E2_C1-3_52+58 NO: 26)	DEWTLQQTSLEMWLAEPQ (SEQ ID
		E2_C1-4_52+58 ID NO: 27)	FKKHGITITVQYDNDKANTMDYTNWKEIY (SEQ
		E2_C1-5_52+58	VIVCPASIPSDEISTEEA (SEQ ID

		NO: 28)
	3 *(53+18)*	E2_C1-1_53+18 DHIDYWKAIRQENAIFFAAR (SEQ ID NO: 29) E2_C1-2_53+18 HQVVPALNICKAKACKAIE (SEQ ID NO: 30) E2_C1-3_53+18 WNTEPKHCFKKGGQHIEVWFD (SEQ ID NO: 31) E2_C1-4_53+18 YVAWDSVYYCGDDGWCKT (SEQ ID NO: 32) E2_C1-5_53+18 EAEKYGCKGTWEVHFG (SEQ ID NO: 33) E2_C1-6_53+18 NSIDCNDSMCSTFDDNVSATELVK (SEQ ID NO: 34)
Approach:	*Modified Variant*	
Fragments:	1	E2_FC1_All DHIDYWKLIRLECAIFYKAR (SEQ ID NO: 35)

Due to alignment inconsistencies three chimerics were created based on phylogeny (16 & 31, 52 & 58, 53 & 18) (Fig 3a). Additionally all genotypes were inputted into the algorithm in a similar fashion to creation of variants but the programs filter for only selecting conserved windows from each genotype with a shared % consensus identity of greater than 60% was not disabled. This identified one fragment, referred to as 'modified variant' (Fig 3b).

E4 Candidate

Approach:	*Chimerics*	
Fragments:	1 *(16+31)*	E4_C1-1_16+31 RRLSSDQDQSQ (SEQ ID NO: 36)
	2 *(52+58)*	E4_C1-1_52+58 LVTKYPLLKLLS (SEQ ID NO: 37)
	3 *(53)*	E4_C1-1_53

		RPPNMGVKAHGKCIWENKVFIVPTLCPVPLDPTYP LLKLLT (SEQ ID NO: 38) E4_C1-2_53 TQTTTPENTSLVELRVTTPKSTVVIRLHL (SEQ ID NO: 39)
	4 *(18)*	E4_C1-1_18 TTRYPLLSLLNSYSTPPHRIPAPCPWAPQRP (SEQ ID NO: 40)
Approach:	*Variants*	
Fragments:	1	E4_V1_16+31 PIPKPSPWAP (SEQ ID NO: 41) E4_V1_18 RIPAPCPWAP (SEQ ID NO: 42) E4_V1_52 PRPPHCPWVP (SEQ ID NO: 43) E4_V1_53 PPPPPRPWAP (SEQ ID NO: 44) E4_V1_58 Excluded

Four chimerics were created, two of the fragments are genotype specific (Fig 4a). Additionally, variant analysis identified one region (Fig 4b).

E5 Candidate

Approach:	*Chimerics*	
Fragments:	1 *(16+31)*	E5_C1_16_31 CFLLCFCVLLCVCLLIRPLLLSVSTY (SEQ ID NO: 45)
	2 *(52+58)*	E5_C1_52+58 LRPLLLSISVYAQVLVLVLLLWVSIGS (SEQ ID NO: 46)
	3 *(18)*	E5_C1_18 LLPSVCMCAYAWVLVFVYIVVITSPATA (SEQ ID NO: 47)

Three chimerics were created (Fig 5).

E6 Candidate

Approach:	*Chimeric-Variants*

Fragments:	1 *(16+18)*	E6_CV1_16 IVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQY NKPLCDLLIRCIN (SEQ ID NO: 48) E6_CV1_18 VVYRDSIPHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLTN TGLYNLLIRCLR (SEQ ID NO: 49)
	2 *(53+31)*	E6_CV2-1_53 VFCKKALTASEVYNFAYTDLRVVYRD (SEQ ID NO: 50) E6_CV2-2_53 SKVRKLRYYNCSVYGASL (SEQ ID NO: 51) E6_CV2-1_31 VYCKGQLTETEVLDFAFTDLTIVYRD (SEQ ID NO: 52) E6_CV2-2_31 SKVSEFRWYRYSVYGTTL (SEQ ID NO: 53)
	3 *(52_58)*	E6_CV3-1_52+58 CVECKKTLQRSEVYD (SEQ ID NO: 54) E6_CV3-2_52+58 CQRPLCPQEKKRHVDLNKRFH (SEQ ID NO: 55)

The E6 protein showed very limited conservation across genotypes. Instead ingroups were processed to produce chimerics without the shared % consensus identity filter, but the conserved windows were not combined to produce a consensus, except for genotypes 52 and 58. (Fig 6)

E7 Candidate

Approach:	*Chimerics*	
Fragments:	1 *(16+31)*	E7_C1_16+31 TLHEYMLDLQPETTDLYCYEQ (SEQ ID NO: 56)
	2 *(52+58)*	E7_C1_58_52 PETTDLHCYEQLGDSSDEEDTGGLDG (SEQ ID NO: 57)
	3 *(53+18)*	Excluded
Approach:	*Chimeric-Variants*	
Fragments	1	E7_V1_53

:		DEDEDEVDHLQEQPQQARRDEQHPCYLIETQCCR CESLV (SEQ ID NO: 58) E7_V1_18 EENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCE ARI (SEQ ID NO: 59)

Chimerics were created for Genotypes 16 & 31 and 52 & 58 (Fig 7). Genotypes 53 and 18 are chimeric variants.

Summary of Vaccine Fragments with SEQ ID NOs.

SEQ ID NO:	Protein	Fragment
1	E1_V1_52+58	DEDETAYDSGTDLIDFIDDS
2	E1_V1_31+16+18	DENENDSDTGEDMVDFIDN
3	E1_V1_53	DETDEESTESDLDGFIDNS
4	E1_V3_31+53	AQLADSDSNACAFLK
5	E1_V3_52+58+18 +16	AQLADVNSNAAAFLK
6	E1_V4_16+31	NCILLYGAANTGKSLFGMSL
7	E1_V4_18+52+58	NCLVLCGPANTGKSYFGMSL
8	E1_V4_53	NCLVIYGPPNTGKSCFAMSL
9	E1_V5_16+31+52	WPYLHSRLVVFTFPNPF
10	E1_V5_18+58	WPYLESRITVFEFPNAF
11	E1_V5_53	LRYLHSRIHVLQFLNPF
12	E2_C1-1_16+31	NVCQDKILEHYENDSKD
13	E2_C1-2_16+31	ILEHYENDSKDLCDHI
14	E2_C1-3_16+31	CDHIDYWKHIRLECAIMYKAR
15	E2_C1-4_16+31	IRLECAIMYKAREMGFH
16	E2_C1-5_16+31	QFDGDICNTMHYTNW
17	E2_C1-6_16+31	IYICEDAQCTVVEGQVD
18	E2_C1-7_16+31	KKWEVHAGGQVILCPES
19	E2_C1-8_16+31	GQRRIKRPRSE
20	E2_C1-9_16+31	NCHPNKLL

21	E2_C1-10_16+31	ILKCLRYRFKKHCKL
22	E2_C1-11_16+31	SSTWHWTCHDGKHK
23	E2_C1-12_16+31	WHWTCHDGKHKNAIVTLTY
24	E2_C1-1_52+58	YEADKNDLNAQIEHWKLIRMECAIFYKAKELGIS
25	E2_C1-2_52+58	ICHQVVPPLAASKAKACQAIELQLALEALNASPY
26	E2_C1-3_52+58	DEWTLQQTSLEMWLAEPQ
27	E2_C1-4_52+58	FKKHGITITVQYDNDKANTMDYTNWKEIY
28	E2_C1-5_52+58	VIVCPASIPSDEISTEEA
29	E2_C1-1_53+18	DHIDYWKAIRQENAIFFAAR
30	E2_C1-2_53+18	HQVVPALNICKAKACKAIE
31	E2_C1-3_53+18	WNTEPKHCFKKGGQHIEVWFD
32	E2_C1-4_53+18	YVAWDSVYYCGDDGWCKT
33	E2_C1-5_53+18	EAEKYGCKGTWEVHFG
34	E2_C1-6_53+18	NSIDCNDSMCSTFDDNVSATELVK
35	E2_FC1_All	DHIDYWKLIRLECAIFYKAR
36	E4_C1-1_16+31	RRLSSDQDQSQ
37	E4_C1-1_52+58	LVTKYPLLKLLS
38	E4_C1-1_53	RPPNMGVKAHGKCIWENKVFIVPTLCPVPLDPTYPLLKLLT
39	E4_C1-2_53	TQTTTPENTSLVELRVTTPKSTVVIRLHL
40	E4_C1-1_18	TTRYPLLSLLNSYSTPPHRIPAPCPWAPQRP
41	E4_V1_16+31	PIPKPSPWAP
42	E4_V1_18	RIPAPCPWAP
43	E4_V1_52	PRPPHCPWVP
44	E4_V1_53	PPPPPRPWAP
45	E5_C1_16_31	CFLLCFCVLLCVCLLIRPLLLSVSTY
46	E5_C1_52+58	LRPLLLSISVYAQVLVLVLLLWVSIGS
47	E5_C1_18	LLPSVCMCAYAWVLVFVYIVVITSPATA
48	E6_CV1_16	IVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIR CIN
49	E6_CV1_18	VVYRDSIPHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLTNTGLYNLLIR CLR
50	E6_CV2-1_53	VFCKKALTASEVYNFAYTDLRVVYRD
51	E6_CV2-2_53	SKVRKLRYYNCSVYGASL
52	E6_CV2-1_31	VYCKGQLTETEVLDFAFTDLTIVYRD
53	E6_CV2-2_31	SKVSEFRWYRYSVYGTTL

54	E6_CV3-1_52+58	CVECKKTLQRSEVYD
55	E6_CV3-2_52+58	CQRPLCPQEKKRHVDLNKRFH
56	E7_C1_16+31	TLHEYMLDLQPETTDLYCYEQ
57	E7_C1_58_52	PETTDLHCYEQLGDSSDEEDTGGLDG
58	E7_V1_53	DEDEDEVDHLQEQPQQARRDEQHPCYLIETQCCRCESLV
59	E7_V1_18	EENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCEARI

HPV3 Nucleotide sequence (SEQ ID NO: 60)

From Start codon, starting with the TPA leading sequence encoded with an additional linker (TPA and linker are bold and underlined).

ATGGATGCTATGAAGAGGGGCCTGTGCTGCGTGCTGCTGCTGTGTGGCGC
CGTGTTTGTGTCCCCCAGCCAGGAAATCCACGCCCGGTTCAGAAGAGGCA
GCAAGCTGGCCGACGAGGACGAGACAGCCTACGACAGCGGCACCGACCTG
ATCGACTTCATCGACGACAGCGACGAGAATGAGAACGACTCCGACACCGG
CGAGGACATGGTGGATTTCATCGACAACGACGAAACCGACGAAGAGAGCA
CCGAGAGCGACCTGGACGGCTTTATCGACAACTCCGCCCAGCTGGCTGAC
AGCGACAGCAATGCCTGCGCCTTCCTGAAGGCTCAGCTGGCAGACGTGAA
CAGCAACGCCGCTGCTTTTCTGAAGAACTGCATCCTGCTGTACGGCGCTG
CCAACACCGGCAAGAGCCTGTTCGGCATGAGCCTGAACTGCCTGGTGCTG
TGCGGCCCAGCCAATACCGGAAAGTCCTACTTCGGCATGTCCCTGAATTG
TCTCGTGATCTACGGCCCACCTAACACAGGCAAGTCCTGCTTTGCCATGT
CTCTGTGGCCCTACCTGCACAGCAGACTGGTGGTGTTTACCTTCCCCAAC
CCCTTCTGGCCTTACCTGGAAAGCCGGATCACCGTGTTCGAGTTCCCCAA
TGCCTTTCTGAGATACCTGCACTCCCGGATCCACGTGCTGCAGTTTCTGA
ACCCCTTCAACGTGTGCCAGGACAAGATCCTGGAACACTACGAGAACGAC
AGCAAGGACATTCTGGAACATTATGAGAATGATTCCAAGGACCTGTGCGA
CCACATCTGCGATCACATCGACTACTGGAAGCACATCCGGCTGGAATGCG
CCATCATGTACAAGGCCCGGATCAGACTGGAATGTGCTATTATGTATAAG
GCTCGCGAGATGGGCTTCCACCAGTTCGACGGCGACATCTGCAACACCAT
GCACTACACCAACTGGATCTATATCTGCGAGGACGCCCAGTGCACCGTGG
TGGAAGGCCAGGTGGACAAGAAATGGGAGGTGCACGCTGGCGGCCAAGTG
ATCCTGTGTCCTGAGAGCGGCCAGCGGCGGATCAAGAGGCCCAGAAGCGA
GAACTGCCACCCCAACAAGCTGCTGATCCTGAAGTGCCTGCGGTACAGAT
TCAAGAAGCACTGCAAGCTGAGCAGCACCTGGCACTGGACCTGCCACGAC
GGCAAGCACAAGTGGCATTGGACATGTCACGATGGGAAACACAAGAACGC
CATTGTGACCCTGACCTACTACGAGGCCGACAAGAACGACCTGAACGCCC
AGATCGAGCACTGGAAACTGATCCGGATGGAATGTGCAATCTTCTATAAG
GCCAAAGAGCTGGGCATCAGCATCTGCCACCAGGTGGTGCCTCCACTGGC
CGCCTCTAAAGCCAAAGCCTGCCAGGCCATCGAACTGCAGCTGGCCCTGG
AAGCCCTGAATGCCAGCCCCTACGATGAGTGGACCCTGCAGCAGACCAGC
CTGGAAATGTGGCTGGCCGAGCCCCAGTTTAAGAAGCACGGCATCACCAT
CACCGTGCAGTACGACAATGACAAGGCCAATACCATGGATTACACAAATT
GGAAAGAAATCTACGTGATCGTGTGCCCCGCCAGCATCCCCTCCGATGAG
ATCAGCACCGAGGAAGCCGACCACATTGATTATTGGAAAGCCATCAGGCA
GGAAAACGCCATCTTCTTCGCCGCCAGACACCAGGTGGTGCCCGCCCTGA
ATATCTGCAAGGCCAAGGCCTGTAAAGCCATCGAGTGGAACACCGAGCCC
AAGCACTGCTTCAAGAAGGGCGGCCAGCACATCGAAGTGTGGTTCGACTA
CGTGGCCTGGGACAGCGTGTACTACTGCGGCGACGATGGCTGGTGCAAGA
CCGAGGCCGAGAAGTACGGCTGCAAGGGCACCTGGGAAGTGCATTTCGGC

AACAGCATCGACTGCAACGACTCCATGTGCAGCACCTTCGACGACAACGT
GTCCGCCACCGAGCTCGTGAAGGACCATATCGACTATTGGAAGCTGATTC
GCCTGGAATGTGCCATTTTTTACAAGGCCAGACGGCGGCTGTCCAGCGAC
CAGGATCAGTCTCAGCTCGTGACCAAGTACCCCCTGCTGAAGCTGCTGTC
CAGACCCCCCAACATGGGCGTGAAGGCCCACGGCAAGTGCATCTGGGAGA
ACAAGGTGTTCATCGTGCCCACCCTGTGCCCCGTGCCTCTGGATCCAACA
TATCCTCTGCTGAAACTGCTGACCACCCAGACCACCACCCCCGAGAATAC
CTCCCTGGTGGAACTGAGAGTGACCACCCCCAAGAGCACAGTCGTGATCA
GGCTGCACCTGACCACCAGATACCCACTGCTGTCACTGCTGAACAGCTAC
AGCACCCCCCCTCACCGGATCCCTGCTCCATGTCCTTGGGCTCCTCAGAG
GCCCCCCATCCCTAAGCCTTCTCCATGGGCCCCTAGAATCCCTGCCCCTT
GCCCCTGGGCACCTCCTAGACCTCCACACTGTCCATGGGTGCCCCCTCCA
CCTCCTCCAAGACCTTGGGCCCCTTGCTTCCTGCTGTGCTTTTGTGTGCT
GCTGTGCGTGTGCCTGCTGATCAGACCCCTGCTGCTGAGTGTGTCCACCT
ACCTGAGGCCTCTGCTGCTGTCTATCAGCGTGTACGCTCAGGTGCTGGTG
CTGGTGCTGCTGCTGTGGGTGTCCATCGGAAGCCTGCTGCCCAGCGTGTG
CATGTGTGCCTATGCCTGGGTGCTGGTGTTCGTGTACATCGTCGTGATTA
CCAGCCCCGCCACCGCCATCGTGTACCGGGATGGCAATCCTTACGCCGTG
TGCGACAAGTGCCTGAAGTTCTACAGCAAGATCAGCGAGTACCGGCACTA
CTGCTACAGCCTGTACGGCACCACCCTGGAACAGCAGTACAACAAGCCCC
TGTGCGATCTGCTGATTCGGTGCATCAACGTGGTGTACAGAGACTCCATC
CCCCACGCCGCCTGCCACAAGTGTATCGACTTCTACTCCAGAATCAGAGA
GCTGCGGCACTACAGCGACTCCGTGTACGGCGATACCCTGGAAAAGCTGA
CCAACACTGGCCTGTACAACCTGCTGATTAGATGCCTGCGGGTGTTCTGC
AAGAAGGCCCTGACAGCCAGCGAGGTGTACAACTTCGCCTACACCGATCT
GCGGGTGGTGTATCGGGACAGCAAAGTGCGGAAGCTGAGGTACTACAACT
GCTCTGTGTATGGCGCCAGCCTGGTGTATTGCAAGGGACAGCTGACCGAG
ACAGAGGTGCTGGATTTCGCCTTCACAGACCTGACAATCGTGTATCGCGA
CTCCAAGGTGTCCGAGTTCCGGTGGTACAGATATTCCGTGTATGGCACCA
CACTGTGCGTGGAATGCAAGAAAACCCTGCAGAGATCTGAGGTGTACGAC
TGCCAGCGGCCACTGTGTCCGCAGGAAAAGAAAAGACACGTGGACCTGAA
CAAGCGGTTCCACACCCTGCACGAGTACATGCTGGATCTGCAGCCCGAGA
CAACCGACCTGTACTGCTACGAGCAGCCTGAAACCACTGATCTGCACTGT
TATGAGCAGCTGGGAGACAGCTCCGATGAAGAGGACACTGGCGGCCTGGA
TGGGGACGAGGATGAGGACGAAGTGGACCATCTGCAGGAACAGCCCCAGC
AGGCTAGACGGGACGAACAGCACCCTTGCTATCTGATCGAGACACAGTGC
TGCAGATGCGAATCTCTGGTGGAAGAGAACGACGAGATCGACGGCGTGAA
CCACCAGCATCTGCCCGCTAGAAGGGCCGAGCCTCAGAGACACACCATGC
TGTGTATGTGCTGCAAGTGCGAGGCCAGAATCGCCGGCTCTGGACCTGGC
GCCTCTGGCAAGCCTATCCCCAATCCACTGCTGGGCCTGGACTCCACCCG
GACCTGATAA

HPV3 Nucleotide sequence without encoding a peptide adjuvant/TPA (SEQ ID NO: 65)

GACGAGGACGAGACAGCCTACGACAGCGGCACCGACCTG
ATCGACTTCATCGACGACAGCGACGAGAATGAGAACGACTCCGACACCGG
CGAGGACATGGTGGATTTCATCGACAACGACGAAACCGACGAAGAGAGCA
CCGAGAGCGACCTGGACGGCTTTATCGACAACTCCGCCCAGCTGGCTGAC
AGCGACAGCAATGCCTGCGCCTTCCTGAAGGCTCAGCTGGCAGACGTGAA
CAGCAACGCCGCTGCTTTTCTGAAGAACTGCATCCTGCTGTACGGCGCTG
CCAACACCGGCAAGAGCCTGTTCGGCATGAGCCTGAACTGCCTGGTGCTG
TGCGGCCCAGCCAATACCGGAAAGTCCTACTTCGGCATGTCCCTGAATTG
TCTCGTGATCTACGGCCCACCTAACACAGGCAAGTCCTGCTTTGCCATGT
CTCTGTGGCCCTACCTGCACAGCAGACTGGTGGTGTTTACCTTCCCCAAC
CCCTTCTGGCCTTACCTGGAAAGCCGGATCACCGTGTTCGAGTTCCCCAA
TGCCTTTCTGAGATACCTGCACTCCCGGATCCACGTGCTGCAGTTTCTGA

ACCCCTTCAACGTGTGCCAGGACAAGATCCTGGAACACTACGAGAACGAC
AGCAAGGACATTCTGGAACATTATGAGAATGATTCCAAGGACCTGTGCGA
CCACATCTGCGATCACATCGACTACTGGAAGCACATCCGGCTGGAATGCG
CCATCATGTACAAGGCCCGGATCAGACTGGAATGTGCTATTATGTATAAG
GCTCGCGAGATGGGCTTCCACCAGTTCGACGGCGACATCTGCAACACCAT
GCACTACACCAACTGGATCTATATCTGCGAGGACGCCCAGTGCACCGTGG
TGGAAGGCCAGGTGGACAAGAAATGGGAGGTGCACGCTGGCGGCCAAGTG
ATCCTGTGTCCTGAGAGCGGCCAGCGGCGGATCAAGAGGCCCAGAAGCGA
GAACTGCCACCCCAACAAGCTGCTGATCCTGAAGTGCCTGCGGTACAGAT
TCAAGAAGCACTGCAAGCTGAGCAGCACCTGGCACTGGACCTGCCACGAC
GGCAAGCACAAGTGGCATTGGACATGTCACGATGGGAAACACAAGAACGC
CATTGTGACCCTGACCTACTACGAGGCCGACAAGAACGACCTGAACGCCC
AGATCGAGCACTGGAAACTGATCCGGATGGAATGTGCAATCTTCTATAAG
GCCAAAGAGCTGGGCATCAGCATCTGCCACCAGGTGGTGCCTCCACTGGC
CGCCTCTAAAGCCAAAGCCTGCCAGGCCATCGAACTGCAGCTGGCCCTGG
AAGCCCTGAATGCCAGCCCCTACGATGAGTGGACCCTGCAGCAGACCAGC
CTGGAAATGTGGCTGGCCGAGCCCCAGTTTAAGAAGCACGGCATCACCAT
CACCGTGCAGTACGACAATGACAAGGCCAATACCATGGATTACACAAATT
GGAAAGAAATCTACGTGATCGTGTGCCCCGCCAGCATCCCCTCCGATGAG
ATCAGCACCGAGGAAGCCGACCACATTGATTATTGGAAAGCCATCAGGCA
GGAAAACGCCATCTTCTTCGCCGCCAGACACCAGGTGGTGCCCGCCCTGA
ATATCTGCAAGGCCAAGGCCTGTAAAGCCATCGAGTGGAACACCGAGCCC
AAGCACTGCTTCAAGAAGGGCGGCCAGCACATCGAAGTGTGGTTCGACTA
CGTGGCCTGGGACAGCGTGTACTACTGCGGCGACGATGGCTGGTGCAAGA
CCGAGGCCGAGAAGTACGGCTGCAAGGGCACCTGGGAAGTGCATTTCGGC
AACAGCATCGACTGCAACGACTCCATGTGCAGCACCTTCGACGACAACGT
GTCCGCCACCGAGCTCGTGAAGGACCATATCGACTATTGGAAGCTGATTC
GCCTGGAATGTGCCATTTTTTACAAGGCCAGACGGCGGCTGTCCAGCGAC
CAGGATCAGTCTCAGCTCGTGACCAAGTACCCCCTGCTGAAGCTGCTGTC
CAGACCCCCCAACATGGGCGTGAAGGCCCACGGCAAGTGCATCTGGGAGA
ACAAGGTGTTCATCGTGCCCACCCTGTGCCCCGTGCCTCTGGATCCAACA
TATCCTCTGCTGAAACTGCTGACCACCCAGACCACCACCCCCGAGAATAC
CTCCCTGGTGGAACTGAGAGTGACCACCCCCAAGAGCACAGTCGTGATCA
GGCTGCACCTGACCACCAGATACCCACTGCTGTCACTGCTGAACAGCTAC
AGCACCCCCCCTCACCGGATCCCTGCTCCATGTCCTTGGGCTCCTCAGAG
GCCCCCCATCCCTAAGCCTTCTCCATGGGCCCCTAGAATCCCTGCCCCTT
GCCCCTGGGCACCTCCTAGACCTCCACACTGTCCATGGGTGCCCCCTCCA
CCTCCTCCAAGACCTTGGGCCCCTTGCTTCCTGCTGTGCTTTTGTGTGCT
GCTGTGCGTGTGCCTGCTGATCAGACCCCTGCTGCTGAGTGTGTCCACCT
ACCTGAGGCCTCTGCTGCTGTCTATCAGCGTGTACGCTCAGGTGCTGGTG
CTGGTGCTGCTGCTGTGGGTGTCCATCGGAAGCCTGCTGCCCAGCGTGTG
CATGTGTGCCTATGCCTGGGTGCTGGTGTTCGTGTACATCGTCGTGATTA
CCAGCCCCGCCACCGCCATCGTGTACCGGGATGGCAATCCTTACGCCGTG
TGCGACAAGTGCCTGAAGTTCTACAGCAAGATCAGCGAGTACCGGCACTA
CTGCTACAGCCTGTACGGCACCACCCTGGAACAGCAGTACAACAAGCCCC
TGTGCGATCTGCTGATTCGGTGCATCAACGTGGTGTACAGAGACTCCATC
CCCCACGCCGCCTGCCACAAGTGTATCGACTTCTACTCCAGAATCAGAGA
GCTGCGGCACTACAGCGACTCCGTGTACGGCGATACCCTGGAAAAGCTGA
CCAACACTGGCCTGTACAACCTGCTGATTAGATGCCTGCGGGTGTTCTGC
AAGAAGGCCCTGACAGCCAGCGAGGTGTACAACTTCGCCTACACCGATCT
GCGGGTGGTGTATCGGGACAGCAAAGTGCGGAAGCTGAGGTACTACAACT
GCTCTGTGTATGGCGCCAGCCTGGTGTATTGCAAGGGACAGCTGACCGAG
ACAGAGGTGCTGGATTTCGCCTTCACAGACCTGACAATCGTGTATCGCGA
CTCCAAGGTGTCCGAGTTCCGGTGGTACAGATATTCCGTGTATGGCACCA
CACTGTGCGTGGAATGCAAGAAAACCCTGCAGAGATCTGAGGTGTACGAC
TGCCAGCGGCCACTGTGTCCGCAGGAAAAGAAAAGACACGTGGACCTGAA
CAAGCGGTTCCACACCCTGCACGAGTACATGCTGGATCTGCAGCCCGAGA
CAACCGACCTGTACTGCTACGAGCAGCCTGAAACCACTGATCTGCACTGT

TATGAGCAGCTGGGAGACAGCTCCGATGAAGAGGACACTGGCGGCCTGGA
TGGGGACGAGGATGAGGACGAAGTGGACCATCTGCAGGAACAGCCCCAGC
AGGCTAGACGGGACGAACAGCACCCTTGCTATCTGATCGAGACACAGTGC
TGCAGATGCGAATCTCTGGTGGAAGAGAACGACGAGATCGACGGCGTGAA
CCACCAGCATCTGCCCGCTAGAAGGGCCGAGCCTCAGAGACACACCATGC
TGTGTATGTGCTGCAAGTGCGAGGCCAGAATCGCCGGCTCTGGACCTGGC
GCCTCTGGCAAGCCTATCCCCAATCCACTGCTGGGCCTGGACTCCACCCG
GACCTGATAA

HPV3 polypeptide sequence (SEQ ID NO: 61)

From Start codon, starting with the TPA leading sequence.

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGSKLADEDETAYDSGTDL
IDFIDDSDENENDSDTGEDMVDFIDNDETDEESTESDLDGFIDNSAQLAD
SDSNACAFLKAQLADVNSNAAAFLKNCILLYGAANTGKSLFGMSLNCLVL
CGPANTGKSYFGMSLNCLVIYGPPNTGKSCFAMSLWPYLHSRLVVFTFPN
PFWPYLESRITVFEFPNAFLRYLHSRIHVLQFLNPFNVCQDKILEHYEND
SKDILEHYENDSKDLCDHICDHIDYWKHIRLECAIMYKARIRLECAIMYK
AREMGFHQFDGDICNTMHYTNWIYICEDAQCTVVEGQVDKKWEVHAGGQV
ILCPESGQRRIKRPRSENCHPNKLLILKCLRYRFKKHCKLSSTWHWTCHD
GKHKWHWTCHDGKHKNAIVTLTYYEADKNDLNAQIEHWKLIRMECAIFYK
AKELGISICHQVVPPLAASKAKACQAIELQLALEALNASPYDEWTLQQTS
LEMWLAEPQFKKHGITITVQYDNDKANTMDYTNWKEIYVIVCPASIPSDE
ISTEEADHIDYWKAIRQENAIFFAARHQVVPALNICKAKACKAIEWNTEP
KHCFKKGGQHIEVWFDYVAWDSVYYCGDDGWCKTEAEKYGCKGTWEVHFG
NSIDCNDSMCSTFDDNVSATELVKDHIDYWKLIRLECAIFYKARRRLSSD
QDQSQLVTKYPLLKLLSRPPNMGVKAHGKCIWENKVFIVPTLCPVPLDPT
YPLLKLLTTQTTTPENTSLVELRVTTPKSTVVIRLHLTTRYPLLSLLNSY
STPPHRIPAPCPWAPQRPPIPKPSPWAPRIPAPCPWAPPRPPHCPWVPPP
PPPRPWAPCFLLCFCVLLCVCLLIRPLLLSVSTYLRPLLLSISVYAQVLV
LVLLLWVSIGSLLPSVCMCAYAWVLVFVYIVVITSPATAIVYRDGNPYAV
CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINVVYRDSI
PHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLTNTGLYNLLIRCLRVFC
KKALTASEVYNFAYTDLRVVYRDSKVRKLRYYNCSVYGASLVYCKGQLTE
TEVLDFAFTDLTIVYRDSKVSEFRWYRYSVYGTTLCVECKKTLQRSEVYD
CQRPLCPQEKKRHVDLNKRFHTLHEYMLDLQPETTDLYCYEQPETTDLHC
YEQLGDSSDEEDTGGLDGDEDEDEVDHLQEQPQQARRDEQHPCYLIETQC
CRCESLVEENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCEARIAGSGPG
ASGKPIPNPLLGLDSTRT**

HPV3 polypeptide sequence without the TPA/peptide adjuvant sequence (SEQ ID NO: 66)

DEDETAYDSGTDL
IDFIDDSDENENDSDTGEDMVDFIDNDETDEESTESDLDGFIDNSAQLAD
SDSNACAFLKAQLADVNSNAAAFLKNCILLYGAANTGKSLFGMSLNCLVL
CGPANTGKSYFGMSLNCLVIYGPPNTGKSCFAMSLWPYLHSRLVVFTFPN
PFWPYLESRITVFEFPNAFLRYLHSRIHVLQFLNPFNVCQDKILEHYEND
SKDILEHYENDSKDLCDHICDHIDYWKHIRLECAIMYKARIRLECAIMYK
AREMGFHQFDGDICNTMHYTNWIYICEDAQCTVVEGQVDKKWEVHAGGQV
ILCPESGQRRIKRPRSENCHPNKLLILKCLRYRFKKHCKLSSTWHWTCHD
GKHKWHWTCHDGKHKNAIVTLTYYEADKNDLNAQIEHWKLIRMECAIFYK
AKELGISICHQVVPPLAASKAKACQAIELQLALEALNASPYDEWTLQQTS
LEMWLAEPQFKKHGITITVQYDNDKANTMDYTNWKEIYVIVCPASIPSDE
ISTEEADHIDYWKAIRQENAIFFAARHQVVPALNICKAKACKAIEWNTEP

KHCFKKGGQHIEVWFDYVAWDSVYYCGDDGWCKTEAEKYGCKGTWEVHFG
NSIDCNDSMCSTFDDNVSATELVKDHIDYWKLIRLECAIFYKARRRLSSD
QDQSQLVTKYPLLKLLSRPPNMGVKAHGKCIWENKVFIVPTLCPVPLDPT
YPLLKLLTTQTTTPENTSLVELRVTTPKSTVVIRLHLTTRYPLLSLLNSY
STPPHRIPAPCPWAPQRPPIPKPSPWAPRIPAPCPWAPPRPPHCPWVPPP
PPPRPWAPCFLLCFCVLLCVCLLIRPLLLSVSTYLRPLLLSISVYAQVLV
LVLLLWVSIGSLLPSVCMCAYAWVLVFVYIVVITSPATAIVYRDGNPYAV
CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINVVYRDSI
PHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLTNTGLYNLLIRCLRVFC
KKALTASEVYNFAYTDLRVVYRDSKVRKLRYYNCSVYGASLVYCKGQLTE
TEVLDFAFTDLTIVYRDSKVSEFRWYRYSVYGTTLCVECKKTLQRSEVYD
CQRPLCPQEKKRHVDLNKRFHTLHEYMLDLQPETTDLYCYEQPETTDLHC
YEQLGDSSDEEDTGGLDGDEDEDEVDHLQEQPQQARRDEQHPCYLIETQC
CRCESLVEENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCEARIAGSGPG
ASGKPIPNPLLGLDSTRT**

Viral vector sequence ChAdOx1 with immunogen coding sequence insert.

Start and end codons of the immunogen coding sequence insert are underlined. Lead TPA sequence and linker is in bold.

(SEQ ID NO: 62)

GTTTAAACGCGGCCGCCAGGCCTACCCACTAGTCAATTCGGGAGGATCGAAACGGCAGATCGCAA
AAAACAGTACATACAGAAGGAGACATGAACATGAACATCAAAAAAATTGTAAAACAAGCCACAGT
TCTGACTTTTACGACTGCACTTCTGGCAGGAGGAGCGACTCAAGCCTTCGCGAAAGAAAATAACC
AAAAAGCATACAAAGAAACGTACGGCGTCTCTCATATTACACGCCATGATATGCTGCAGATCCCT
AAACAGCAGCAAAACGaAAAATACCAAGTGCCTCAATTCGATCAATCAACGATTAAAAATATTGA
GTCTGCAAAAGGACTTGATGTGTGGGACAGCTGGCCGCTGCAAAACGCTGACGGAACAGTAGCTG
AATACAACGGCTATCACGTTGTGTTTGCTCTTGCGGGAAGCCCGAAAGACGCTGATGACACATCA
ATCTACATGTTTTATCAAAAGGTCGGCGACAACTCAATCGACAGCTGGAAAAACGCGGGCCGTGT
CTTTAAAGACAGCGATAAGTTCGACGCCAACGATCCGATCCTGAAAGATCAGACGCAAGAATGGT
CCGGTTCTGCAACCTTTACATCTGACGGAAAAATCCGTTTATTCTACACTGACTATTCCGGTAAA
CATTACGGCAAACAAAGCCTGACAACAGCGCAGGTAAATGTGTCAAAATCTGATGACACACTCAA
AATCAACGGAGTGGAAGATCACAAAACGATTTTTGACGGAGACGGAAAAACATATCAGAACGTTC
AGCAGTTTATCGATGAAGGCAATTATACATCCGGCGACAACCATACGCTGAGAGACCCTCACTAC
GTTGAAGACAAAGGCCATAAATACCTTGTATTCGAAGCCAACACGGGAACAGAAAACGGATACCA
AGGCGAAGAATCTTTATTTAACAAAGCGTACTACGGCGGCGGCACGAACTTCTTCCGTAAAGAAA
GCCAGAAGCTTCAGCAGAGCGCTAAAAAACGCGATGCTGAGTTAGCGAACGGCGCCCTCGGTATC
ATAGAGTTAAATAATGATTACACATTGAAAAAAGTAATGAAGCCGCTGATCACTTCAAACACGGT
AACTGATGAAATCGAGCGCGCGAATGTTTTCAAAATGAACGGCAAATGGTACTTGTTCACTGATT
CACGCGGTTCAAAAATGACGATCGATGGTATTAACTCAAACGATATTTACATGCTTGGTTATGTA
TCAAACTCTTTAACCGGCCCTTACAAGCCGCTGAACAAAACAGGGCTTGTGCTGCAAATGGGTCT
TGATCCAAACGATGTGACATTCACTTACTCTCACTTCGCAGTGCCGCAAGCCAAAGGCAACAATG
TGGTTATCACAAGCTACATGACAAACAGAGGCTTCTTCGAGGATAAAAAGGCAACATTTGCGCCA

AGCTTCTTAATGAACATCAAAGGCAATAAAACATCCGTTGTCAAAAACAGCATCCTGGAGCAAGG
ACAGCTGACAGTCAACTAATAACAGCAAAAAGaAAATGCCGATACTTCATTGGCATTTTCTTtTA
TTTCTCAACAAGATGGTGAATTGACTAGTGGGTAGATCCACAGGACGGGTGTGGTCGCCATGATC
GCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGA
CAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCAT
AGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACC
AAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTT
CATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGA
TGATAAGCTGTCAAACATGAGAATTGATCCGGAACCCTTAATATAACTTCGTATAATGTATGCTA
TACGAAGTTATTAGGTCCCTCGACTATAGGGTCACCGTCGACAGCGACACACTTGCATCGGATGC
AGCCCGGTTAACGTGCCGGCACGGCCTGGGTAACCAGGTATTTTGTCCACATAACCGTGCGCAAA
ATGTTGTGGATAAGCAGGACACAGCAGCAATCCACAGCAGGCATACAACCGCACACCGAGGTTAC
TCCGTTCTACAGGTTACGACGACATGTCAATACTTGCCCTTGACAGGCATTGATGGAATCGTAGT
CTCACGCTGATAGTCTGATCGACAATACAAGTGGGACCGTGGTCCCAGACCGATAATCAGACCGA
CRAYACGAGTGGGAYCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTC
CCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTYCCAGWCTRATWATCAGACCGAC
GATACRAGTGGRACMGTGGKCCCAGASAKAATAWTCAGRCCgAGWTAYGcWKTCKGGCCTGTAAC
AAAGGACATTAAGTAAAGACAGATAMRMGTgRGACtaaaaCGTGGTCCCAGTCTGATTATCAGAC
CGACGATACGAGTGGGACCGTGGTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTG
GTCCCAGACTAATAATCAGACCGACGATACGAGTGGGACCGTGGTCCCAGTCTGATTATCAGACC
GACGATACAAGTGGAACAGTGGGCCCAGAGAGAATATTCAGGCCAGTTATGCTTTCTGGCCTGTA
ACAAAGGACATTAAGTAAAGACAGATAAACGTAGACTAAAACGTGGTCGCATCAGGGTGCTGGCT
TTTCAAGTTCCTTAAGAATGGCCTCAATTTTCTCTATACACTCAGTTGGAACACGAGACCTGTCC
AGGTTAAGCACCATTTTATCGCCCTTATACAATACTGTCGCTCCAGGAGCAAACTGATGTCGTGA
GCTTAAACTAGTTCTTGATGCAGATGACGTTTTAAGCACAGAAGTTAAAAGAGTGATAACTTCTT
CAGCTTCAAATATCACCCCAGCTTTTTTCTGCTCATGAAGGTTAGATGCCTGCTGCTTAAGTAAT
TCCTCTTTATCTGTAAAGGCTTTTTGAAGTGCATCACCTGACCGGGCAGATAGTTCACCGGGGTG
AGAAAAAAGAGCAACAACTGATTTAGGCAATTTGGCGGTGTTGATACAGCGGGTAATAATCTTAC
GTGAAATATTTTCCGCATCAGCCAGCGCAGAAATATTTCCAGCAAATTCATTCTGCAATCGGCTT
GCATAACGCTGACCACGTTCATAAGCACTTGTTGGGCGATAATCGTTACCCAATCTGGATAATGC
AGCCATCTGCTCATCATCCAGCTCGCCAACCAGAACACGATAATCACTTTCGGTAAGTGCAGCAG
CTTTACGACGGCGACTCCCATCGGCAATTTCTATGACACCAGATACTCTTCGACCGAACGCCGGT
GTCTGTTGACCAGTCAGTAGAAAAGAAGGGATGAGATCATCCAGTGCGTCCTCAGTAAGCAGCTC
CTGGTCACGTTCATTACCTGACCATACCCGAGAGGTCTTCTCAACACTATCACCCCGGAGCACTT
CAAGAGTAAACTTCACATCCCGACCACATACAGGCAAAGTAATGGCATTACCGCGAGCCATTACT
CCTACGCGCGCAATTAACGAATCCACCATCGGGGCAGCTGGTGTCGATAACGAAGTATCTTCAAC
CGGTTGAGTATTGAGCGTATGTTTTGGAATAACAGGCGCACGCTTCATTATCTAATCTCCCAGCG
TGGTTTAATCAGACGATCGAAAATTTCATTGCAGACAGGTTCCCAAATAGAAAGAGCATTTCTCC
AGGCACCAGTTGAAGAGCGTTGATCAATGGCCTGTTCAAAAACAGTTCTCATCCGGATCTGACCT

TTACCAACTTCATCCGTTTCACGTACAACATTTTTTAGAACCATGCTTCCCCAGGCATCCCGAAT
TTGCTCCTCCATCCACGGGGACTGAGAGCCATTACTATTGCTGTATTTGGTAAGCAAAATACGTA
CATCAGGCTCGAACCCTTTAAGATCAACGTTCTTGAGCAGATCACGAAGCATATCGAAAAACTGC
AGTGCGGAGGTGTAGTCAAACAACTCAGCAGGCGTGGGAACAATCAGCACATCAGCAGCACATAC
GACATTAATCGTGCCGATACCCAGGTTAGGCGCGCTGTCAATAACTATGACATCATAGTCATGAG
CAACAGTTTCAATGGCCAGTCGGAGCATCAGGTGTGGATCGGTGGGCAGTTTACCTTCATCAAAT
TTGCCCATTAACTCAGTTTCAATACGGTGCAGAGCCAGACAGGAAGGAATAATGTCAAGCCCCGG
CCAGCAAGTGGGCTTTATTGCATAAGTGACATCGTCCTTTTCCCCAAGATAGAAAGGCAGGAGAG
TGTCTTCTGCATGAATATGAAGATCTGGTACCCATCCGTGATACATTGAGGCTGTTCCCTGGGGG
TCGTTACCTTCCACGAGCAAAACACGTAGCCCCTTCAGAGCCAGATCCTGAGCAAGATGAACAGA
AACTGAGGTTTTGTAAACGCCACCTTTATGGGCAGCAACCCCGATCACCGGTGGAAATACGTCTT
CAGCACGTCGCAATCGCGTACCAAACACATCACGCATATGATTAATTTGTTCAATTGTATAACCA
ACACGTTGCTCAACCCGTCCTCGAATTTCCATATCCGGGTGCGGTAGTCGCCCTGCTTTCTCGGC
ATCTCTGATAGCCTGAGAAGAAACCCCAACTAAATCCGCTGCTTCACCTATTCTCCAGCGCCGGG
TTATTTTCCTCGCTTCCGGGCTGTCATCATTAAACTGTGCAATGGCGATAGCCTTCGTCATTTCA
TGACCAGCGTTTATGCACTGGTTAAGTGTTTCCATGAGTTTCATTCTGAACATCCTTTAATCATT
GCTTTGCGTTTTTTTATTAAATCTTGCAATTTACTGCAAAGCAACAACAAAATCGCAAAGTCATC
AAAAAACCGCAAAGTTGTTTAAAATAAGAGCAACACTACAAAAGGAGATAAGAAGAGCACATACC
TCAGTCACTTATTATCACTAGCGCTCGCCGCAGCCGTGTAACCGAGCATAGCGAGCGAACTGGCG
AGGAAGCAAAGAAGAACTGTTCTGTCAGATAGCTCTTACGCTCAGCGCAAGAAGAAATATCCACC
GTGGGAAAAACTCCAGGTAGAGGTACACACGCGGATAGCCAATTCAGAGTAATAAACTGTGATAA
TCAACCCTCATCAATGATGACGAACTAACCCCCGATATCAGGTCACATGACGAAGGGAAAGAGAA
GGAAATCAACTGTGACAAACTGCCCTCAAATTTGGCTTCCTTAAAAATTACAGTTCAAAAAGTAT
GAGAAAATCCATGCAGGCTGAAGGAAACAGCAAAACTGTGACAAATTACCCTCAGTAGGTCAGAA
CAAATGTGACGAACCACCCTCAAATCTGTGACAGATAACCCTCAGACTATCCTGTCGTCATGGAA
GTGATATCGCGGAAGGAAAATACGATATGAGTCGTCTGGCGGCCTTTCTTTTTCTCAATGTATGA
GAGGCGCATTGGAGTTCTGCTGTTGATCTCATTAACACAGACCTGCAGGAAGCGGCGGCGGAAGT
CAGGCATACGCTGGTAACTTTGAGGCAGCTGGTAACGCTCTATGATCCAGTCGATTTTCAGAGAG
ACGATGCCTGAGCCATCCGGCTTACGATACTGACACAGGGATTCGTATAAACGCATGGCATACGG
ATTGGTGATTTCTTTTGTTTCACTAAGCCGAAACTGCGTAAACCGGTTCTGTAACCCGATAAAGA
AGGGAATGAGATATGGGTTGATATGTACACTGTAAAGCCCTCTGGATGGACTGTGCGCACGTTTG
ATAAACCAAGGAAAAGATTCATAGCCTTTTTCATCGCCGGCATCCTCTTCAGGGCGATaAAAAAC
CACTTCCTTCCCCGCGAAACTCTTCAATGCCTGCCGTATATCCTTACTGGCTTCCGCAGAGGTCA
ATCCGAATATTTCAGCATATTTAGCAACATGGATCTCGCAGATACCGTCATGTTCCTGTAGGGTG
CCATCAGATTTTCTGATCTGGTCAACGAACAGATACAGCATACGTTTTTGATCCCGGGAGAGACT
ATATGCCGCCTCAGTGAGGTCGTTTGACTGGACGATTCGCGGGCTATTTTTACGTTTCTTGTGAT
TGATAACCGCTGTTTCCGCCATGACAGATCCATGTGAAGTGTGACAAGTTTTTAGATTGTCACAC
TAAATAAAAAAAGAGTCAATAAGCAGGGATAACTTTGTGAAAAAACAGCTTCTTCTGAGGGCAATT
TGTCACAGGGTTAAGGGCAATTTGTCACAGACAGGACTGTCATTTGAGGGTGATTTGTCACACTG

AAAGGGCAATTTGTCACAACACCTTCTCTAGAACCAGCATGGATAAAGGCCTACAAGGCGCTCTA
AAAAAGAAGATCTAAAAACTATaAAAAAAATAATTATAAAAATATCCCCGTGGATAAGTGGATAA
CCCCAAGGGAAGTTTTTtCAGGCATCGTGTGTAAGCAGAATATATAAGTGCTGTTCCCTGGTGCT
TCCTCGCTCACTCGAGGGCTTCGCCCTGTCGCTCAACTGCGGCGAGCACTACTGGCTGTAAAAGG
ACAGACCACATCATGGTTCTGTGTTCATTAGGTTGTTCTGTCCATTGCTGACATAATCCGCTCCA
CTTCAACGTAACACCGCACGAAGATTTCTATTGTTCCTGAAGGCATATTCAAATCGTTTTCGTTA
CCGCTTGCAGGCATCATGACAGAACACTACTTCCTATAAACGCTACACAGGCTCCTGAGATTAAT
AATGCGGATCTCTACGATAATGGGAGATTTTCCCGACTGTTTCGTTCGCTTCTCAGTGGATAACA
GCCAGCTTCTCTGTTTAACAGACAAAAACAGCATATCCACTCAGTTCCACATTTCCATATAAAGG
CCAAGGCATTTATTCTCAGGATAATTGTTTCAGCATCGCAACCGCATCAGACTCCGGCATCGCAA
ACTGCACCCGGTGCCGGGCAGCCACATCCAGCGCAAAAACCTTCGTGTAGACTTCCGTTGAACTG
ATGGACTTATGTCCCATCAGGCTTTGCAGAACTTTCAGCGGTATACCGGCATACAGCATGTGCAT
CGCATAGGAATGGCGGAACGTATGTGGTGTGACCGGAACAGAGAACGTCACACCGTCAGCAGCAG
CGGCGGCAACCGCCTCCCCAATCCAGGTCCTGACCGTTCTGTCCGTCACTTCCCAGATCCGCGCT
TTCTCTGTCCTTCCTGTGCGACGGTTACGCCGCTCCATGAGCTTATCGCGAATAAATACCTGTGA
CGGAAGATCACTTCGCAGAATAAATAAATCCTGGTGTCCCTGTTGATACCGGGAAGCCCTGGGCC
AACTTTTGGCGAAAATGAGACGTTGATCGGCACGTAAGAGGTTCCAACTTTCACCATAATGAAAT
AAGATCACTACCGGGCGTATTTTTTGAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGG
AGAAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAG
GCATTTCAGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTT
AAAGACCGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTTGCCCGCCTGA
TGAATGCTCATCCGGAGTTCCGTATGGCAATGAAAGACGGTGAGCTGGTGATATGGGATAGTGTT
CACCCTTGTTACACCGTTTTCCATGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCA
CGACGATTTCCGGCAGTTTCTACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGG
CCTATTTCCCTAAAGGGTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTC
ACCAGTTTTGATTTAAACGTGGCCAATATGGACAACTTCTTCGCCCCCGTTTTCACCATGGGCAA
ATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCAGGTTCATCATGCCGTTTGTG
ATGGCTTCCATGTCGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGG
GCGTAAtTTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTGAATAAGTGA
TAATAAGCGGATGAATGGCAGAAATTCGATGATAAGCTGTCAAACATGAGAATTGGTCGACGGCG
CGCCAAAGCTTGCATGCCTGCAGCCGCGTAACCTGGCAAAATCGGTTACGGTTGAGTAATAAATG
GATGCCCTGCGTAAGCGGGGCACATTTCATTACCTCTTTCTCCGCACCCGACATAGATAATAACT
TCGTATAGTATACATTATACGAAGTTATCTAGTAGACTTAATCGCGTTTAAACCCATCATCAATA
ATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTGGGAAGGGAGGA
AGGTGATTGGCCGAGAGAAGGGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGACC
GCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTGAAC
ACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTAGGCGGATGCAAGT
GAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTAATTTCGCGTTTA
TGACAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGGGGTTTCGATTA

CCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTCAGC

TGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAA

GAGTTTTCTCCTCCGCGCGCGAGTCAGATCTACACTTTGAAAGGCGATCGCTAGCGACATCGATC

ACAAGTTTGTACAAAAAAGCAGGCTCCACCATGGGAACCAATTCAgTCGAGCCTTTCACTCATTA

GATGCATGTCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCC

CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA

TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC

GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT

GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT

TGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCAT

TGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACT

CCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTC

CCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCG

TCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCA

GCCTCCGGTTAAGCTcGgtacCGCTAGCCGCGCCGCCACC**ATGGATGCTATGAAGAGGGGCCTGT**

GCTGCGTGCTGCTGCTGTGTGGCGCCGTGTTTGTGTCCCCCAGCCAGGAAATCCACGCCCGGTTC

AGAAGAGGCAGCAAGCTGGCCGACGAGGACGAGACAGCCTACGACAGCGGCACCGACCTGATCGA

CTTCATCGACGACAGCGACGAGAATGAGAACGACTCCGACACCGGCGAGGACATGGTGGATTTCA

TCGACAACGACGAAACCGACGAAGAGAGCACCGAGAGCGACCTGGACGGCTTTATCGACAACTCC

GCCCAGCTGGCTGACAGCGACAGCAATGCCTGCGCCTTCCTGAAGGCTCAGCTGGCAGACGTGAA

CAGCAACGCCGCTGCTTTTCTGAAGAACTGCATCCTGCTGTACGGCGCTGCCAACACCGGCAAGA

GCCTGTTCGGCATGAGCCTGAACTGCCTGGTGCTGTGCGGCCCAGCCAATACCGGAAAGTCCTAC

TTCGGCATGTCCCTGAATTGTCTCGTGATCTACGGCCCACCTAACACAGGCAAGTCCTGCTTTGC

CATGTCTCTGTGGCCCTACCTGCACAGCAGACTGGTGGTGTTTACCTTCCCCAACCCCTTCTGGC

CTTACCTGGAAAGCCGGATCACCGTGTTCGAGTTCCCCAATGCCTTTCTGAGATACCTGCACTCC

CGGATCCACGTGCTGCAGTTTCTGAACCCCTTCAACGTGTGCCAGGACAAGATCCTGGAACACTA

CGAGAACGACAGCAAGGACATTCTGGAACATTATGAGAATGATTCCAAGGACCTGTGCGACCACA

TCTGCGATCACATCGACTACTGGAAGCACATCCGGCTGGAATGCGCCATCATGTACAAGGCCCGG

ATCAGACTGGAATGTGCTATTATGTATAAGGCTCGCGAGATGGGCTTCCACCAGTTCGACGGCGA

CATCTGCAACACCATGCACTACACCAACTGGATCTATATCTGCGAGGACGCCCAGTGCACCGTGG

TGGAAGGCCAGGTGGACAAGAAATGGGAGGTGCACGCTGGCGGCCAAGTGATCCTGTGTCCTGAG

AGCGGCCAGCGGCGGATCAAGAGGCCCAGAAGCGAGAACTGCCACCCCAACAAGCTGCTGATCCT

GAAGTGCCTGCGGTACAGATTCAAGAAGCACTGCAAGCTGAGCAGCACCTGGCACTGGACCTGCC

ACGACGGCAAGCACAAGTGGCATTGGACATGTCACGATGGGAAACACAAGAACGCCATTGTGACC

CTGACCTACTACGAGGCCGACAAGAACGACCTGAACGCCCAGATCGAGCACTGGAAACTGATCCG

GATGGAATGTGCAATCTTCTATAAGGCCAAAGAGCTGGGCATCAGCATCTGCCACCAGGTGGTGC

CTCCACTGGCCGCCTCTAAAGCCAAAGCCTGCCAGGCCATCGAACTGCAGCTGGCCCTGGAAGCC

CTGAATGCCAGCCCCTACGATGAGTGGACCCTGCAGCAGACCAGCCTGGAAATGTGGCTGGCCGA

GCCCCAGTTTAAGAAGCACGGCATCACCATCACCGTGCAGTACGACAATGACAAGGCCAATACCA

TGGATTACACAAATTGGAAAGAAATCTACGTGATCGTGTGCCCCGCCAGCATCCCCTCCGATGAG
ATCAGCACCGAGGAAGCCGACCACATTGATTATTGGAAAGCCATCAGGCAGGAAAACGCCATCTT
CTTCGCCGCCAGACACCAGGTGGTGCCCGCCCTGAATATCTGCAAGGCCAAGGCCTGTAAAGCCA
TCGAGTGGAACACCGAGCCCAAGCACTGCTTCAAGAAGGGCGGCCAGCACATCGAAGTGTGGTTC
GACTACGTGGCCTGGGACAGCGTGTACTACTGCGGCGACGATGGCTGGTGCAAGACCGAGGCCGA
GAAGTACGGCTGCAAGGGCACCTGGGAAGTGCATTTCGGCAACAGCATCGACTGCAACGACTCCA
TGTGCAGCACCTTCGACGACAACGTGTCCGCCACCGAGCTCGTGAAGGACCATATCGACTATTGG
AAGCTGATTCGCCTGGAATGTGCCATTTTTTACAAGGCCAGACGGCGGCTGTCCAGCGACCAGGA
TCAGTCTCAGCTCGTGACCAAGTACCCCCTGCTGAAGCTGCTGTCCAGACCCCCCAACATGGGCG
TGAAGGCCCACGGCAAGTGCATCTGGGAGAACAAGGTGTTCATCGTGCCCACCCTGTGCCCCGTG
CCTCTGGATCCAACATATCCTCTGCTGAAACTGCTGACCACCCAGACCACCACCCCCGAGAATAC
CTCCCTGGTGGAACTGAGAGTGACCACCCCCAAGAGCACAGTCGTGATCAGGCTGCACCTGACCA
CCAGATACCCACTGCTGTCACTGCTGAACAGCTACAGCACCCCCCCTCACCGGATCCCTGCTCCA
TGTCCTTGGGCTCCTCAGAGGCCCCCCATCCCTAAGCCTTCTCCATGGGCCCCTAGAATCCCTGC
CCCTTGCCCCCTGGGCACCTCCTAGACCTCCACACTGTCCATGGGTGCCCCCTCCACCTCCTCCAA
GACCTTGGGCCCCTTGCTTCCTGCTGTGCTTTTGTGTGCTGCTGTGCGTGTGCCTGCTGATCAGA
CCCCTGCTGCTGAGTGTGTCCACCTACCTGAGGCCTCTGCTGCTGTCTATCAGCGTGTACGCTCA
GGTGCTGGTGCTGGTGCTGCTGCTGTGGGTGTCCATCGGAAGCCTGCTGCCCAGCGTGTGCATGT
GTGCCTATGCCTGGGTGCTGGTGTTCGTGTACATCGTCGTGATTACCAGCCCCGCCACCGCCATC
GTGTACCGGGATGGCAATCCTTACGCCGTGTGCGACAAGTGCCTGAAGTTCTACAGCAAGATCAG
CGAGTACCGGCACTACTGCTACAGCCTGTACGGCACCACCCTGGAACAGCAGTACAACAAGCCCC
TGTGCGATCTGCTGATTCGGTGCATCAACGTGGTGTACAGAGACTCCATCCCCCACGCCGCCTGC
CACAAGTGTATCGACTTCTACTCCAGAATCAGAGAGCTGCGGCACTACAGCGACTCCGTGTACGG
CGATACCCTGGAAAAGCTGACCAACACTGGCCTGTACAACCTGCTGATTAGATGCCTGCGGGTGT
TCTGCAAGAAGGCCCTGACAGCCAGCGAGGTGTACAACTTCGCCTACACCGATCTGCGGGTGGTG
TATCGGGACAGCAAAGTGCGGAAGCTGAGGTACTACAACTGCTCTGTGTATGGCGCCAGCCTGGT
GTATTGCAAGGGACAGCTGACCGAGACAGAGGTGCTGGATTTCGCCTTCACAGACCTGACAATCG
TGTATCGCGACTCCAAGGTGTCCGAGTTCCGGTGGTACAGATATTCCGTGTATGGCACCACACTG
TGCGTGGAATGCAAGAAAACCCTGCAGAGATCTGAGGTGTACGACTGCCAGCGGCCACTGTGTCC
GCAGGAAAAGAAAAGACACGTGGACCTGAACAAGCGGTTCCACACCCTGCACGAGTACATGCTGG
ATCTGCAGCCCGAGACAACCGACCTGTACTGCTACGAGCAGCCTGAAACCACTGATCTGCACTGT
TATGAGCAGCTGGGAGACAGCTCCGATGAAGAGGACACTGGCGGCCTGGATGGGGACGAGGATGA
GGACGAAGTGGACCATCTGCAGGAACAGCCCCAGCAGGCTAGACGGGACGAACAGCACCCTTGCT
ATCTGATCGAGACACAGTGCTGCAGATGCGAATCTCTGGTGGAAGAGAACGACGAGATCGACGGC
GTGAACCACCAGCATCTGCCCGCTAGAAGGGCCGAGCCTCAGAGACACACCATGCTGTGTATGTG
CTGCAAGTGCGAGGCCAGAATCGCCGGCTCTGGACCTGGCGCCTCTGGCAAGCCTATCCCCAATC
CACTGCTGGGCCTGGACTCCACCCGGACCTGATAAGCggccgctcgagcatgcatctagagggcc
ctattctatagtgtcacctaaatgctagagctcgctgatcagcctcgactgtgccttctagttgc
cagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgt

cctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggg
gtggggtggggcaggacagcaaggggaggattgggaagacaatagcaggcatgctggggatgcg
gtgggctctatggcttctgaggcggaaagaaccagctggggctcgaggggggatcgatcccgtcG
AGATATCTAGACCCAGCTTTCTTGTACAAAGTGGTGATCGATTCGACAGATCGCGATCGCAGTGA
GTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATGACTGAAATCTGTGCTTTTCTG
TGTGTTGCAGCATCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACG
GGGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCC
CGTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGGTGGACGCAG
CTGCCGCCGCAGCTGCTGCATCCGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTAC
TACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCT
GCTGCTGCTGATGGCCCAGCTTGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGG
CTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAA
TAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCG
CGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTA
GAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGGTGGAGGTAGCTCCATT
GCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCGTGG
TGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTAC
AAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGA
GATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTG
TATCCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGAC
GCCCTTGTGTCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCAATGGGCCCGTGGGCGG
CGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCA
TAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCC
GGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCAGAGGGGGGGATCATGT
CCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAG
TTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAAATGACCCCGATGACCGGCTGCAG
GTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCATCT
CGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGAGATAGGAGC
TCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGT
CTGTTGCAAGAGTTCCAAGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCA
GACCTCCTCGTTTCGCGGGTTGGGACGACTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCG
CAGCCAGGGTCCGGTCCTTCCAGGGCCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAG
GGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCG
CTCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCT
CGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGG
GACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGAATCGGGGGCGTAGGCGTCCGCGCCGCA
GTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCA
GTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGG
GTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCC

GCGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGA
AGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACTTTTTCCAGGGTATGC
AAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACC
GGGGGTCCCGGCCGGGGGGGGTATAAAAGGGGGCGGGCCCCTGCTCGTCCTCACTGTCTTCCGGAT
CGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCA
CTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCAGCGGAGATGCCTTT
CAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGAtTTTTTTGTTGTCGAGCTTGGTGGCGAAGG
AGCCGTAGAGGGCGTTGGAAAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTtCCTtGTCG
GCGCGCTCCTTGGCCGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAA
GACGGTGGTCATCTCGTCGGGCACGATTCTGACCTGCCAACCTCGATTATGCAGGGTGATGAGGT
CCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCGTTGGTCCAGCAGAGGCGGCCGCCCTTGCGC
GAGCAGAAGGGGGGCAGAGGGTCCAGCATGACCTCGTCGGGGGGGGTCGGCATCGATGGTGAAGAT
GCCGGGCAGGAGATCGGGGTCGAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGAAGCTTGCC
ATTCGCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGGTGGGTG
AGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTA
GGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGCG
CGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGAAAG
ATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGGAGGCC
GACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGTTTGGCGACGAGCTCGGCGGTGACGA
GGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGGCCCTTTTGT
TTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCC
GTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGC
CCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAG
GTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGAAATCGATATCGTCGCAGcCCCCCTGCTC
CCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGA
AAAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGCTGGGGCACCTCGGCC
CGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAACCGTTGATGTTGTGGCCCACGAT
GTAGAGTTCCACGAATCGCGGGCGGCCCTTGACGTGGGGCAGCTTCTTGAGCTCCTCGTAGGTGA
GCTCGTCGGGGTCGCTGAGACCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGG
AGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTG
CTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGGT
CCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGGCGGTCGTCCCCTGAGAGTTTC
ATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACATC
GTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCCTGCC
ACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACACTCG
TGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAG
CTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGT
GCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGC
CCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAG

GCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGT
TGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCG
TTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTT
CTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACG
CGCGCCGGGCGGCAGAGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGC
GCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACG
TCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTC
GACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGT
TGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCCTGAAGGTCTCCGCGACCG
GCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCAT
GCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCGGGCGCGCATGACCA
CCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGG
TAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCAT
CTCGCTGACGTCGCCCAGCGCCTCCAAGCGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGA
AAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATG
GTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCCTCTTCTTCCTCCTC
CACTAACATCTCTTCTACTTCCTCCTCAGGCGGTGGTGGTGGCGGGGGAGGGGGCCTGCGTCGCC
GGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTC
TCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTG
GCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCGTAG
GGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGTTGAACGAAGGCT
TCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGCCGGGTCATGTTGGGGAGC
GGGGCGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGA
GGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGG
TCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTC
GCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGG
CGACGACGCGCTCGGCGAGGATGGCCTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCGTCAAAG
TCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTT
GACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGA
AGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGG
CGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTG
GTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACT
CGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCC
GTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTC
CGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAG
GCTGGAGCCGCAGCTAACGTGGTACTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCA
GGATACGGAGGCGGGTCGTTTTGCAACTTTTTTtGGAGGCCGGAAATGAAACTAGTAAGCGCGGA
AAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGT
GTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCA

AGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTT
GCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCT
CCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAGCAGCAACTTCCAGCCACGACCGCCGC
GGCCGCCGTGAGCGGGGCTGGACAGACTTCTCAGTATGATCACCTGGCCTTGGAAGAGGGCGAGG
GGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCT
CGCGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGAT
GCGCGCGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGA
GGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCG
GCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAA
CAACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACC
TGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTG
CAGCATAGTCGGGACAACGAGGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCG
CTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGT
CCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTCTGGGCAAGTACTACGCTAGGAAGATC
TACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGAC
CCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGCGCGG
TGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCACAGCCTGCAGCGGGCCCTG
ACCGGGGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAG
CCGCCGGGCCTTGGAGGCGGCAGGCGGTCCCCCCTACATAGAAGAGGTGGACGATGAGGTGGACG
AGGAGGGCGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAGCCAC
CTCCTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGAT
TGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCA
GCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGC
ACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCC
GGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAA
CCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGT
CCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGG
GGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAG
CGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGA
ACCTGAGCCAGGCGTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGC
GCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCAC
GGACAGCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCA
TCGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGC
CAGGACGACCCGGGCAATCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGAT
CCCGCCCCAGTACACGCTCAGCGCCGAGGAGGAGCGCATCCTGCGATACGTGCAGCAGAGCGTGG
GCCTGTTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATG
GAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGC
GGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGG
GGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGAC

AGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGA
CCGACGCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCA
GTCCTTTCCCGAGCTTGCCCTTCTCGCTGAACAGTATTCGCAGCAGCGAGCTGGGCAGGATCACG
CGCCCGCGCTTGCTGGGCGAGGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAA
GAACTTCCCCAATAACGGGATAGAGAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGC
AGGAGCACAGGGACGATCCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGG
TGGCACGACAGGCAGCGGGGACTGATGTGGGACGATGAGGATTCCGCCGACGACAGCAGCGTGTT
GGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGCATCGGGCGCATGATGTAAG
AGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTG
TTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTG
ATGCAGCAGGCGATGGCGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCC
GCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACG
ATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAAC
GACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCAC
CCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGTCAGCTGAAAACCATCATGCACACCA
ACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGC
AAGACCCCCAACGGGGTGACAGTGACAGATGGTAGTCAGGATATCTTGGAGTATGAATGGGTGGA
GTTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCA
TCGACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTCCTGGAGAGCGATATCGGCGTGAAGTTC
GACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACAC
CAACGAGGCCTTCCACCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCC
GCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAGGGCTTCCAGATCATGTAC
GAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGA
GGAGAGCGCCGCCGCGGCGACTGCAGCTGTAGCCACCGCCTCTACCGAGGTCAGGGGCGATAATT
TTGCCAGCCCTGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAG
CCGGTGGAGAAGGATAGCAAGGACAGGAGCTACAACGTGCTGCCGGACAAGATAAACACCGCCTA
CCGCAGCTGGTACCTGGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGC
TCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAA
GACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCT
GCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCA
CCTCGCTCACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCC
ACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAG
CAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCCCTACGTCT
ACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAAAAATGTCCATT
CTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGC
TCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCC
TCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCG
CGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGC
CGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGCA

CCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATG
CTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCAC
GGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCG
ACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGcCCCCCTCGCACTTGAAGATGTTCACTT
CGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTC
CAGGTCATCGCGCCTGAGATCTACGGCCCCGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAAT
CAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGATGACGATCTGGTGGAGTTTGTGCGCGAGTTCG
CCCCCcGGCGGCGCGTGCAGTGGCGCGGGCGGAAAGTGCACCCGGTGCTGAGACCCGGCACCACC
GTGGTCTTCACGCCCGGCGAGCGCTCCGGCAGCGCTTCCAAGCGCTCCTACGACGAGGTGTACGG
GGACGAGGACATCCTCGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCC
GCCCCGCCCTGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCGAGCCTC
AAGCCCGTGACCCTGCAGCAGGTGCTGCCGAGCGCAGCGCCGCGCCGGGGGTTCAAGCGCGAGGG
CGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGG
AGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCC
CCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATCCCCACGGAGCCCATGGAAACGCAGACCGA
GCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTC
CTAGCCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCAT
CCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCGGTCATACAACCAGCCG
CCGCCGCAAGACCACCACCCGCCGCCGCCGTCGCCGCACAGCCGCTGCATCTACCCCTGCCGCCC
TGGTGCGGAGAGTGTACCGCCGCGGCCGCGCGCCTCTGACCCTACCGCGCGCGCGCTACCACCCG
AGCATCGCCATTTAAACTTTCGCCTGCTTTGCAGATGGCCCTCACATGCCGCCTCCGCGTTCCCA
TTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCAC
CACCATCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGgAGGCTTCCTGCCCGCGCTGATCCC
CATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGC
GCCACTGAGACACTTGGAAAACATCTTGTAATAAACCAATGGACTCTGACGCTCCTGGTCCTGTG
ATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGCG
GCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGA
GCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGG
AACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGT
CGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGA
TCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAG
CTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGACGCGGAGGAGACGCTGCTGAC
GCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCA
TCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAGTAATAAGCCCGCGACCCTGGACTTGCCT
CCTCCCGCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCCGCGCGCG
ACCCGGGGGCTCCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGG
GAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTG
TGTGTGTATGTATTATGTCGCCGCTGTCCGCCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTT
GCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCT

TCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTCGCCCGCGCCACAGACACCTACTTCAGTCTGGG
GAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGACCACCGACCGCAGCCAGCGGC
TGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACCTACTCGTACAAAGTGCGCTACACG
CTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACATCCGCGGCGTGCT
GGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGCCTGGCTCCCAAGGGAG
CGCCCAATTCCAGCCAGTGGGAGCaAAAAAAGGCAGGCAATGGTGACACTATGGAAACACACACA
TTTGGTGTGGCCCCAATGGGCGGTGAGAATATTACAATCGACGGATTACAAATTGGAACTGACGC
TACAGCTGATCAGGATAAACCAATTTATGCTGACAAAACATTCCAGCCTGAACCTCAAGTAGGAG
AAGAAAATTGGCAAGAAACTGAAAGCTTTTATGGCGGTAGGGCTCTTAAAAAAGACACAAGCATG
AAACCTTGCTATGGCTCCTATGCTAGACCCACCAATGTAAAGGGAGGTCAAGCTAAACTTAAAGT
TGGAGCTGATGGAGTTCCTACCAAAGAATTTGACATAGACCTGGCTTTCTTTGATACTCCCGGTG
GCACAGTGAATGGACAAGATGAGTATAAAGCAGACATTGTCATGTATACCGAAAACACGTATCTG
GAAACTCCAGACACGCATGTGGTATACAAACCAGGCAAGGATGATGCAAGTTCTGAAATTAACCT
GGTTCAGCAGTCCATGCCCAATAGACCCAACTATATTGGGTTCAGAGACAACTTTATTGGGCTCA
TGTATTACAACAGTACTGGCAATATGGGGGTGCTGGCTGGTCAGGCCTCACAGCTGAATGCTGTG
GTCGACTTGCAAGACAGAAACACCGAGCTGTCATACCAGCTCTTGCTTGACTCTTTGGGTGACAG
AACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGTTATGATCCTGATGTGCGCATTATTG
AAAACCATGGTGTGGAAGACGAACTTCCCAACTATTGCTTCCCCCTGGATGGGTCTGGCACTAAT
GCCGCTTACCAAGGTGTGAAAGTAAAAAATGGTAACGATGGTGATGTTGAGAGCGAATGGGAAAA
TGATGATACTGTCGCAGCTCGAAATCAATTATGCAAGGGCAACATTTTTGCCATGGAAATTAACC
TCCAAGCCAACCTGTGGAGAAGTTTCCTCTACTCGAACGTGGCCCTGTACCTGCCCGACTCTTAC
AAGTACACGCCAGCCAACATCACCCTGCCCACCAACACCAACACTTATGATTACATGAACGGGAG
AGTGGTGCCTCCCTCGCTGGTGGACGCCTACATCAACATCGGGGCGCGCTGGTCGCTGGACCCCA
TGGACAACGTCAATCCCTTCAACCACCACCGCAACGCGGGCCTGCGCTACCGCTCCATGCTCCTG
GGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCT
CCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGC
AGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTC
TACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGA
CACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCA
ACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTC
ACGCGCCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTC
GGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCA
CCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATC
AAGCGCACCGTCGACGGCGAGGGATACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCT
GGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGG
ACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAAC
TACAAGGACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCT
CGCGCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCGTACCCGCTCATCGGCAAGA
GCGCCGTCACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTC

TCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTC
CGCCCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTG
TCTTCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTAC
CTGCGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAATTGCTACTTGCATGATGGCTGAGCC
CACAGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCC
TGGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTC
AACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAA
CACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCG
AGTACGAGGGCCTGCTGCGCCGTAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAG
TCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCA
CGCCTTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGG
TGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTC
TACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCAC
CGCCTTCGACCGCATGAACAATCAAGACATGTAAACCGTGTGTGTATGTTTAAAATATCTTTTAA
TAAACAGCACTTTAATGTTACACATGCATCTGAGATGATTTTATTTTAGAAATCGAAAGGGTTCT
GCCGGGTCTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTG
AACTCGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGT
CAGCTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCT
GCGCGCGAGAGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACG
CTCGCCAGCACCGCCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAA
GGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGC
AGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATG
AAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGA
CTTGCTAGAGAACTGGTTGGTGGCACAGCCGGCATCGTGCACGCAGCAGCGCGCGTCGTTGTTGG
CCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTC
AGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGT
GGTCCCGTGCAGGCACCGCAGTTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACC
CGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCGCGTGCACGAACCCTTGCAGGAAGCGG
CCCATCATGGTCGTCAGGGTCTTGTTGCTAGTGAAGGTCAACGGGATGCCGCGGTGCTCCTCGTT
GATGTACAGGTGGCAGATGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGTTGGAAGTTGGCTT
TCAGGTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATGCCCTTCTCCCAG
GCCGAGACGATGGGCAGGCTCATAGGGTTCTTCACCATCATCTTAGCACTAGCAGCCGCGGCCAG
GGGGTCGCTCTCATCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGG
GGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTG
ACGTCCTGCATGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGTGGCGGCGGAGA
TGCTTGTGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCG
AGGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCG
CGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGA
CTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGC

CATCGCCAACCTCGCCATCTGCCCCCACCGCCGGCGACGAGAAGCAGCAGCAGCAGAATGAAAGC
TTAACCGCCCCGCCGCCCAGCCCCGCCTCCGACGCAGCCGCGGTCCCAGACATGCAAGAGATGGA
GGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCATGAGGAGGAGCTGGCAGTGC
GCTTTCAATCGTCAAGCCAGGAAGATAAAGAACAGCCAGAGCAGGAAGCAGAGAACGAGCAGAGT
CAGGCTGGGCTCGAGCATGGCGACTACCTCCACCTGAGCGGGGAGGAGGACGCGCTCATCAAGCA
TCTGGCCCGGCAGGCCACCATCGTCAAGGACGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCG
TGGAGGAGCTCAGCCGCGCCTACGAGCTCAACCTCTTCTCGCCGCGCGTGcCCCCCAAGCGCCAG
CCCAACGGCACCTGCGAGCCCAACCCCCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGC
CCTGGCCACCTACCACATCTTTTtCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCA
CCCGCGCCGACGCCCTCTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCTCCTTGGAA
GAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGG
AGAAGGAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGC
TGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGcCCCCG
AAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGA
GGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGG
GTCCTAATGCTACCCCTCAAAGTTTGGAAGAGCGGCGCAAGCTCATGATGGCCGTGGTCCTGGTG
ACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGA
GAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGC
TGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCAC
ACCACCCTGCGCGGGGAGGCCCGCCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCA
CACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCT
GCAAGCTCCTGCAAAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGGACCACCGCCTCG
GACCTGGCCGACCTCATCTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTT
TATGAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCG
CCACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTG
TGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGA
CGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCC
TGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCC
AGCGAGGGCGAGGGAGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTT
GCGCAAGTTCGTGCCCGAGGATTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCAGC
CGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCC
ATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGAC
CGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAG
CTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGATGGAGGA
AGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGG
AGGCAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCA
AGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGGCCCCACAGTAGATGGGACGA
GACCGGGCGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCT
GGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACC

CGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTCCCCCCGCAACATCTTGCATTACTACCGTCA
CCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGaAAAAGACCAGAAAACCAGCT
AGAAAATCCACAGCGGCGGCAGCGGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAGAC
CCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGG
AGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAG
AGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCT
CACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGC
CCTTCGCCCTAGCCGCCTCCACCCAGCACCGCCATGAGCAAAGAGATTCCCACGCCTTACATGTG
GAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGCGCCGCCCAGGACTACTCCACCCGCATGAATT
GGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATA
CTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGC
CGCCCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCG
AAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCC
GCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAG
CTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCA
CGCCTCGTCAGGCGGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGCGGCATCGGC
ACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGG
CCACTACCCGGACGAGTTCATCCCGAACTTTGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
GATTAATTAATCAACTAACCCCTTACCCCTTTACCCTCCAGTAAAAATAAAGATTAAAAATGATT
GAATTGATCAATAAAGAATCACTTACTTGAAATCTGAAACCAGGTCTCTGTCCATGTTTTCTGTC
AGCAGCACTTCACTCCCCTCTTCCCAACTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCCT
CCACACTCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTTATCTTCTATCAGA
TGTCCAAAAAGCGCGCGCGGGTGGATGATGGCTTCGACCCCGTGTACCCCTACGATGCAGACAAC
GCACCGACTGTGCCCTTCATCAACCCTCCCTTCGTCTCTTCAGATGGATTCCAAGAAAAGCCCCT
GGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAATGGGGCTGTCACCCTCAAGC
TGGGGGAGGGGGTGGACCTCGACGACTCGGGAAAACTCATCTCCAAAAATGCCACCAAGGCCACT
GCCCCTCTCAGTATTTCCAACGGCACCATTTCCCTTAACATGGCTGCCCCTTTTTACAACAACAA
TGGAACGTTAAGTCTCAATGTTTCTACACCATTAGCAGTATTTCCCACTTTTAACACTTTAGGTA
TCAGTCTTGGAAACGGTCTTCAAACTTCTAATAAGTTGCTGACTGTACAGTTAACTCATCCTCTT
ACATTCAGCTCAAATAGCATCACAGTAAAAACAGACAAAGGACTCTATATTAATTCTAGTGGAAA
CAGAGGGCTTGAGGCTAACATAAGCCTAAAAAGAGGACTGATTTTTGATGGTAATGCTATTGCAA
CATACCTTGGAAGTGGTTTAGACTATGGATCCTATGATAGCGATGGGAAAACAAGACCCATCATC
ACCAAAATTGGAGCAGGTTTGAATTTTGATGCTAATAATGCCATGGCTGTGAAGCTAGGCACAGG
TTTAAGTTTTGACTCTGCCGGTGCCTTAACAGCTGGAAACAAAGAGGATGACAAGCTAACACTTT
GGACTACACCTGACCCAAGCCCTAATTGTCAATTACTTTCAGACAGAGATGCCAAATTTACCCTA
TGTCTTACAAAATGCGGTAGTCAAATACTAGGCACTGTTGCAGTAGCTGCTGTTACTGTAGGTTC
AGCACTAAATCCAATTAATGACACAGTAAAAAGCGCCATAGTATTCCTTAGATTTGACTCTGACG
GTGTGCTCATGTCAAACTCATCAATGGTAGGTGATTACTGGAACTTTAGGGAAGGACAGACCACC
CAAAGTGTGGCCTATACAAATGCTGTGGGATTCATGCCCAATCTAGGTGCATATCCTAAAACCCA

AAGCAAAACACCAAAAAATAGTATAGTAAGTCAGGTATATTTAAATGGAGAAACTACTATGCCAA
TGACACTGACAATAACTTTCAATGGCACTGATGAAAAAGACACAACACCTGTGAGCACTTACTCC
ATGACTTTTACATGGCAGTGGACTGGAGACTATAAGGACAAGAATATTACCTTTGCTACCAACTC
CTTTACTTTCTCCTACATGGCCCAAGAATAAACCCTGCATGCCAACCCCATTGTTCCCACCACTA
TGGAAAACTCTGAAGCAGaAAAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTCtcaca
gaaccctagtattcaacctgccacctccctcccaacacacagagtacacagtcctttctccccgg
ctggccttaaaaagcatcatatcatgggtaacagacatattcttaggtgttatattccacacggt
ttcctgtcgagccaaacgctcatcagtgatattaataaactccccgggcagctcacttaagttca
tgtcgctgtccagctgctgagccacaggctgctgtccaacttgcggttgcttaacgggcggcgaa
ggagaagtccacgcctacatgggggtagagtcataatcgtgcatcaggatagggcggtggtgctg
cagcagcgcgcgaataaactgctgccgccgccgctccgtcctgcaggaatacaacatggcagtgg
tctcctcagcgatgattcgcaccgcccgcagcataaggcgccttgtcctccgggcacagcagcgc
accctgatctcacttaaatcagcacagtaactgcagcacagcaccacaatattgttcaaaatccc
acagtgcaaggcgctgtatccaaagctcatggcggggaccacagaacccacgtggccatcatacc
acaagcgcaggtagattaagtggcgacccctcataaacacgctggacataaacattacctctttt
ggcatgttgtaattcaccacctcccggtaccatataaacctctgattaaacatggcgccatccac
caccatcctaaaccagctggccaaaacctgcccgccggctatacactgcagggaaccgggactgg
aacaatgacagtggagagcccaggactcgtaaccatggatcatcatgctcgtcatgatatcaatg
ttggcacaacacaggcacacgtgcatacacttcctcaggattacaagctcctcccgcgttagaac
catatcccagggaacaacccattcctgaatcagcgtaaatcccacactgcagggaagacctcgca
cgtaactcacgttgtgcattgtcaaagtgttacattcgggcagcagcggatgatcctccagtatg
gtagcgcgggtttctgtctcaaaaggaggtagacgatccctactgtacggagtgcgccgagacaa
ccgagatcgtgttggtcgtagtgtcatgccaaatggaacgccggacgtagtcatatttcctgaag
caaaaccaggtgcgggcgtgacaaacagatctgcgtctccggtctcgccgcttagatcgctctgt
gtagtagttgtagtatatccactctctcaaagcatccaggcgccccctggcttcgggttctatgt
aaactccttcatgcgccgctgccctgataacatccaccaccgcagaataagccacacccagccaa
cctacacattcgttctgcgagtcacacacgggaggagcgggaagagctggaagaaccatGATTAA
CTTTATTCCAAACGGTCTCGGAGCACTTCAAAATGCAGGTCCCGGAGGTGGCACCTCTCGCCCCC
ACTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGACACGGTTCTCGAGATGTTCCACGGTGG
CTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGAGGACAGCGAAAGCGGGAGCGTTTTCT
AATTCCTCAATCATCATATTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTG
AATGATTCGTATTAGTTCCTGAGGTAAATCCAAGCCAGCCATGATAAAAAGCTCGCGCAGAGCGC
CCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGAGATTCTGCTCCTGGTTCACCTGCA
GCAGATTAACAATGGGAATATCAAAATCTCTGCCGCGATCCCTAAGCTCCTCCCTCAACAATAAC
TGTATGTAATCTTTCATATCATCTCCGAAATTTTTAGCCATAGGGCCGCCAGGAATAAGAGCAGG
GCAAGCCACATTACAGATAAAGCGAAGTCCTCCCCAGTGWGCATTGCCAAATGTAAGATTGAAAT
AAGCATGCTGGCTAGACCCTGTGATATCTTCCAGATAACTGGACAGAAAATCAGGCAAGCAATTT
TTAAGAAAATCAACAAAAGAAAAGTCGTCCAGGTGCAGGTTTAGAGCCTCAGGAACAACGATGGA
ATAAGTGCAAGGAGTGCGTTCCAGCATGGTTAGTGtTTTTTTGGTGATCTGTAGAACAAAAAATA

AACATGCAATATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCACTCTTTCCAGCACCAG
GCAGGCTACGGGGTCTCCGGCGCGACCCTCGTAGAAGCTGTCGCCATGATTGAAAAGCATCACCG
AGAGACCTTCCCGGTGGCCGGCATGGATGATTCGAGAAGAAGCATACACTCCGGGAACATTGGCA
TCCGTGAGTGAAAAAAaGCGACCTATAAAGCCTCGGGGCACTACAATGCTCAATCTCAATTCCAG
CAAAGCCACCCCATGCGGATGGAGCACAAAATTGGCAGGTGCGTAAAAAATGTAATTACTCCCCT
CCTGCACAGGCAGCAAAGCCCCCGCTCCCTCCAGAAACACATACAAAGCCTCAGCGTCCATAGCT
TACCGAGCACGGCAGGCGCAAGAGTCAGAGAAAAGGCTGAGCTCTAACCTGACTGCCCGCTCCTG
TGCTCAATATATAGCCCTAACCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAAAT
GACACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGTGCGCTTCCTC
AAACGCCCAAACCGGCGTCATTTCCGGGTTCCCACGCTACGTCACCGCTCAGCGACTTTCAAATT
CCGTCGACCGTTAAAAACGTCACTCGCCCCGCCCCTAACGGTCGCCCTTCTCTCGGCCAATCACC
TTCCTCCCTTCCCAAATTCAAACGCCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATA
TATTTGAATGATG

ChAdOx2 sequence (SEQ ID NO: 67 + 68)

The ChAdOx2 sequence 5' to the immunogen cassette is provided as SEQ ID NO: 67 and the ChAdOx2 sequence 3' to the immunogen cassette is provided as SEQ ID NO: 68.

MVA sequence (SEQ ID NO: 69 + 70)

The MVA sequence 5' to the immunogen cassette is provided as SEQ ID NO: 69 and the MVA sequence 3' to the immunogen cassette is provided as SEQ ID NO: 70.

MVA vaccines have been made using two different shuttle plasmids:

1. P7.5 shuttle plasmid. HPV insert with upstream and downstream flanks that are homologous to regions of the TK locus in parental MVA virus. Insert under control of p7.5 promoter.

2. F11 shuttle plasmid. HPV insert with upstream and downstream flanks that are homologous to regions of the F11 locus in parental MVA virus. Insert under control of F11 promoter.

Chicken embryo fibroblast cells are then infected with MVA parental virus and transfected with either p7.5 or F11 MVA shuttle plasmids to allow homologous recombination with the MVA genome. So you get parental MVA with the gene of

interest inserted into the MVA genome at either the TK locus (p7.5 shuttle plasmid) or F11 locus (F11 shuttle plasmid). Resulting in two versions of the MVA vaccine.

P7.5 shuttle plasmid (SEQ ID NO: 158) (insert underlined)

agcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactgga
aagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggct
cgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcttgcatgcA
TCTGGAAACGGGCATCTCCATTTAAGACTAGAtGCCACGGGGTTTAAAATACT
AATCATGACATTTTGTAGAGCGTAATTACTTAGTAAATCCGCCGTACTAGGTT
CATTTCCTCCTCGTTTGGATCTCACATCAGAAATTAAAATAATCTTAGAAGGA
TGCAGTTGTTTTTTGATGGATCGTAGATATTCCTCATCAACGAACCGAGTCAC
TAGAGTCACATCACGCAATCCATTTAAAATAGGATCATGATGGCGGCCGTCA
ATTAGCATCCATTTGATGATCACTCCTAAATTATAGAAATGATCTCTCAAATA
ACGTATATGTGTACCGGGAGCAGATCCTATATACACTACGGTGGCACCATCTA
ATATACCGTGTCGCTGTAACTTACTAAGAAAAAATAATTCTCCTAGTAATAGT
TTTAACTGTCCTTGATACGGtAGTTTTTTTGCGACCTCATTTGCACTTTCTGGTT
CGTAATCTAACTCATTATCAATTTCCTCAAAATACATAAACGGTTTATCTAAC
GACACAACATCCATTTTTAAGTATTATATTAAAATTTAATCAATGTTTATTTTT
AGTTTTTTAGATAAAAAATATAATATTATGAGTCGATGTAACACTTTCTACAC
ACCGATTGATACATATCATTACCTCCTATTATcTCTATCTCGGTTTCCTCACCC
AATCGTTTAGAAAAGGAAGCCTCCTTAAAGCATTTCATACACACAGCAGTTAG
TTTTACCACCATTTCAGATAATGGAATAAGATTCAAAATATTATTAAACGGTT
TACGTTGAAATGTCCCATCGAGTGCGGCTACTATAACTATTTTTCCTTCGTTTG
CCaTACAGATCCTACGTACTCGAGCGGCCGCTTATCAGGTCCGGGTGGAGTCC
AGGCCCAGCAGTGGATTGGGGATAGGCTTGCCAGAGGCGCCAGGTCCAGAGC
CGGCGATTCTGGCCTCGCACTTGCAGCACATACACAGCATGGTGTGTCTCTGA
GGCTCGGCCCTTCTAGCGGGCAGATGCTGGTGGTTCACGCCGTCGATCTCGTC
GTTCTCTTCCACCAGAGATTCGCATCTGCAGCACTGTGTCTCGATCAGATAGC
AAGGGTGCTGTTCGTCCCGTCTAGCCTGCTGGGGCTGTTCCTGCAGATGGTCC
ACTTCGTCCTCATCCTCGTCCCCATCCAGGCCGCCAGTGTCCTCTTCATCGGAG
CTGTCTCCCAGCTGCTCATAACAGTGCAGATCAGTGGTTTCAGGCTGCTCGTA
GCAGTACAGGTCGGTTGTCTCGGGCTGCAGATCCAGCATGTACTCGTGCAGGG
TGTGGAACCGCTTGTTCAGGTCCACGTGTCTTTTCTTTTCCTGCGGACACAGTG
GCCGCTGGCAGTCGTACACCTCAGATCTCTGCAGGGTTTTCTTGCATTCCACG

CACAGTGTGGTGCCATACACGGAATATCTGTACCACCGGAACTCGGACACCTT
GGAGTCGCGATACACGATTGTCAGGTCTGTGAAGGCGAAATCCAGCACCTCT
GTCTCGGTCAGCTGTCCCTTGCAATACACCAGGCTGGCGCCATACACAGAGCA
GTTGTAGTACCTCAGCTTCCGCACTTTGCTGTCCCGATACACCACCCGCAGAT
CGGTGTAGGCGAAGTTGTACACCTCGCTGGCTGTCAGGGCCTTCTTGCAGAAC
ACCCGCAGGCATCTAATCAGCAGGTTGTACAGGCCAGTGTTGGTCAGCTTTTC
CAGGGTATCGCCGTACACGGAGTCGCTGTAGTGCCGCAGCTCTCTGATTCTGG
AGTAGAAGTCGATACACTTGTGGCAGGCGGCGTGGGGGATGGAGTCTCTGTA
CACCACGTTGATGCACCGAATCAGCAGATCGCACAGGGGCTTGTTGTACTGCT
GTTCCAGGGTGGTGCCGTACAGGCTGTAGCAGTAGTGCCGGTACTCGCTGATC
TTGCTGTAGAACTTCAGGCACTTGTCGCACACGGCGTAAGGATTGCCATCCCG
GTACACGATGGCGGTGGCGGGGCTGGTAATCACGACGATGTACACGAACACC
AGCACCCAGGCATAGGCACACATGCACACGCTGGGCAGCAGGCTTCCGATGG
ACACCCACAGCAGCAGCACCAGCACCAGCACCTGAGCGTACACGCTGATAGA
CAGCAGCAGAGGCCTCAGGTAGGTGGACACACTCAGCAGCAGGGGTCTGATC
AGCAGGCACACGCACAGCAGCACACAAAAGCACAGCAGGAAGCAAGGGGCC
CAAGGTCTTGGAGGAGGTGGAGGGGGCACCCATGGACAGTGTGGAGGTCTAG
GAGGTGCCCAGGGGCAAGGGGCAGGGATTCTAGGGGCCCATGGAGAAGGCTT
AGGGATGGGGGGCCTCTGAGGAGCCCAAGGACATGGAGCAGGGATCCGGTG
AGGGGGGGTGCTGTAGCTGTTCAGCAGTGACAGCAGTGGGTATCTGGTGGTC
AGGTGCAGCCTGATCACGACTGTGCTCTTGGGGGTGGTCACTCTCAGTTCCAC
CAGGGAGGTATTCTCGGGGGTGGTGGTCTGGGTGGTCAGCAGTTTCAGCAGA
GGATATGTTGGATCCAGAGGCACGGGGCACAGGGTGGGCACGATGAACACCT
TGTTCTCCCAGATGCACTTGCCGTGGGCCTTCACGCCCATGTTGGGGGGTCTG
GACAGCAGCTTCAGCAGGGGGTACTTGGTCACGAGCTGAGACTGATCCTGGT
CGCTGGACAGCCGCCGTCTGGCCTTGTAAAAAATGGCACATTCCAGGCGAAT
CAGCTTCCAATAGTCGATATGGTCCTTCACGAGCTCGGTGGCGGACACGTTGT
CGTCGAAGGTGCTGCACATGGAGTCGTTGCAGTCGATGCTGTTGCCGAAATGC
ACTTCCCAGGTGCCCTTGCAGCCGTACTTCTCGGCCTCGGTCTTGCACCAGCC
ATCGTCGCCGCAGTAGTACACGCTGTCCCAGGCCACGTAGTCGAACCACACTT
CGATGTGCTGGCCGCCCTTCTTGAAGCAGTGCTTGGGCTCGGTGTTCCACTCG
ATGGCTTTACAGGCCTTGGCCTTGCAGATATTCAGGGCGGGCACCACCTGGTG
TCTGGCGGCGAAGAAGATGGCGTTTTCCTGCCTGATGGCTTTCCAATAATCAA
TGTGGTCGGCTTCCTCGGTGCTGATCTCATCGGAGGGGATGCTGGCGGGGCAC
ACGATCACGTAGATTTCTTTCCAATTTGTGTAATCCATGGTATTGGCCTTGTCA

TTGTCGTACTGCACGGTGATGGTGATGCCGTGCTTCTTAAACTGGGGCTCGGC
CAGCCACATTTCCAGGCTGGTCTGCTGCAGGGTCCACTCATCGTAGGGGCTGG
CATTCAGGGCTTCCAGGGCCAGCTGCAGTTCGATGGCCTGGCAGGCTTTGGCT
TTAGAGGCGGCCAGTGGAGGCACCACCTGGTGGCAGATGCTGATGCCCAGCT
CTTTGGCCTTATAGAAGATTGCACATTCCATCCGGATCAGTTTCCAGTGCTCG
ATCTGGGCGTTCAGGTCGTTCTTGTCGGCCTCGTAGTAGGTCAGGGTCACAAT
GGCGTTCTTGTGTTTCCCATCGTGACATGTCCAATGCCACTTGTGCTTGCCGTC
GTGGCAGGTCCAGTGCCAGGTGCTGCTCAGCTTGCAGTGCTTCTTGAATCTGT
ACCGCAGGCACTTCAGGATCAGCAGCTTGTTGGGGTGGCAGTTCTCGCTTCTG
GGCCTCTTGATCCGCCGCTGGCCGCTCTCAGGACACAGGATCACTTGGCCGCC
AGCGTGCACCTCCCATTTCTTGTCCACCTGGCCTTCCACCACGGTGCACTGGG
CGTCCTCGCAGATATAGATCCAGTTGGTGTAGTGCATGGTGTTGCAGATGTCG
CCGTCGAACTGGTGGAAGCCCATCTCGCGAGCCTTATACATAATAGCACATTC
CAGTCTGATCCGGGCCTTGTACATGATGGCGCATTCCAGCCGGATGTGCTTCC
AGTAGTCGATGTGATCGCAGATGTGGTCGCACAGGTCCTTGGAATCATTCTCA
TAATGTTCCAGAATGTCCTTGCTGTCGTTCTCGTAGTGTTCCAGGATCTTGTCC
TGGCACACGTTGAAGGGGTTCAGAAACTGCAGCACGTGGATCCGGGAGTGCA
GGTATCTCAGAAAGGCATTGGGGAACTCGAACACGGTGATCCGGCTTTCCAG
GTAAGGCCAGAAGGGGTTGGGGAAGGTAAACACCACCAGTCTGCTGTGCAGG
TAGGGCCACAGAGACATGGCAAAGCAGGACTTGCCTGTGTTAGGTGGGCCGT
AGATCACGAGACAATTCAGGGACATGCCGAAGTAGGACTTTCCGGTATTGGC
TGGGCCGCACAGCACCAGGCAGTTCAGGCTCATGCCGAACAGGCTCTTGCCG
GTGTTGGCAGCGCCGTACAGCAGGATGCAGTTCTTCAGAAAAGCAGCGGCGT
TGCTGTTCACGTCTGCCAGCTGAGCCTTCAGGAAGGCGCAGGCATTGCTGTCG
CTGTCAGCCAGCTGGGCGGAGTTGTCGATAAAGCCGTCCAGGTCGCTCTCGGT
GCTCTCTTCGTCGGTTTCGTCGTTGTCGATGAAATCCACCATGTCCTCGCCGGT
GTCGGAGTCGTTCTCATTCTCGTCGCTGTCGTCGATGAAGTCGATCAGGTCGG
TGCCGCTGTCGTAGGCTGTCTCGTCCTCGTCGGCCAGCTTGCTGCCTCTTCTGA
ACCGGGGCGTGGATTTCCTGGCTGGGGGGACACAAACACGGCGCCACACAGCAG
CAGCACGCAGCACAGGCCCCTCTTCATAGCATCCATGGTGGCGGCGCGGCTA
GCGGTACCGgatctagatGGGGATCCGTCACtGTTCTTTATGATTCTACTTCCTTACC
GTGCAATAAATTAGAATATATTTTCTACTTTTACGAGAAATTAATTATTGTATT
TATTATTTATGGGTGAAAAACTTACTATAAAAAGCGGGTGGGTTTGGAATTAG
TGATCAGTTTATGTATATCGCAACTACCGGgCATATGGCTATCGACATCGAGA
ACATTACCCACATGATAAGAGATTGTATCAGTTTCGTAGTCTTGAGTATTGGT

ATTACTATATAGTATATagatGTCGACCTGcaggtcGACGAAGTTCCTATACTTTCT
AGAGAATAGGAACTTCGCAGCCAAGCTGGAATTcaTCCACTTTGGATAAGAAA
TCTGCATGATAAATATATTGATATCCTACCACCTATTAAAGTACCATTATCTA
ATAGCAATAAGATAGATAAACAAATGTTTTTTGATGAAGTTATTACGTGGATA
AATATATATCTTCAGGAAAAGGGTATTATGTTACCAGATGATATAAGAGAACT
CAGAGATGCTATTATTCCTTAACTAGTTACGTCTCTTTAGGTACTTATTTTGAT
ACGTTACAAGTAAAAAACTATCAAATATAAATGGAATCTGATTCTAATATAGC
GATTGAAGAggaTCCACCGGTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA
CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
CGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA
GCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT
TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAG
CCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAC
TTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCA
CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGAT
CACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGA
CGAGCTGTACAAGTAAAGCGGccgcgaagttcctatactttctagagaataggaacTTCAACAAT
GTCTGGAAAGAACTGTCCTTCATCGATACCTATCACGGAGAAATCTGTAATTG
ATTCCAAGAcATCACATAGTTTAGTTGCTTCCAATGCTTCAAAATTATTCTTAT
CATGCGTCCATAGTCCCGTTCCGTATCTATTATCGTTAGAATATTTTATAGTCA
CGCATTTATATTGAGCTATTTGATAACGTCTAACTCGTCTAATTAATTCTGTAC
TTTTACCTGAAAACATGGGGCCGATTATCAACTGAATATGTCCGCCGTTCATG
ATGACAATAAAGAATTAATTATTGTTCACTTTATTCGACTTTAATATATCCATC
ACGTTAGAAAATGCGATATcGCGACGAGGATCTATGTATCTAACAGGATCTAT
TGCGGTGGTAGCTAGAGctGATTCTTTTTTGAATCGCATCAAACTAATCACAAA
GTCGAACAAATATCCTTTATTAAGTTTGACCCTTCCATCTGTAACAATAGGGA
CCTTGTTAAACAGTTTTTTAAAATCTTGAgAGTCTGTGAATTTTGTCAATTGTC
TGTATTCCTCTGAAAGAGATTCATAACAATGACCCACGGCTTCTAATTTATTTT
TTGATTGGATCAATAATAATAACAGAAAGTCTAGATATTGAGTGATTTGCAAT
ATATCAGATAATGAAGATTCATCATCTTGACTAGCCAAATACTTAAAAAATGA

ATCATCATCTGCGAAGAACATCGTTAAGAGATACTGGTTGTGATCCATTTATga
gctcgcgaaagcttggcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgc
agcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaa
tggcgaatggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatc
tgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccg
gcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcga
gacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcgg
ggaaatgtgcgcggaaccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaat
gcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttttgcggcattttgccttcct
gtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactgg
atctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtgg
cgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtact
caccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacac
tgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactc
gccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggca
acaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataa
agttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctc
gcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactat
ggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatat
actttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaac
gtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgc
tgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactcttttccgaaggtaact
ggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccaccacttcaagaactctgtagcacc
gcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaag
acgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacc
tacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc
ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcg
ggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgc
ggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtatta
ccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaag
a

F11-HPV shuttle plasmid (SEQ ID NO: 159) (vaccine construct insert is underlined)

AGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGAC
AGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTA
GGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAAC
AATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTATTTAGGTGACACTATAGAATA
CTCAAGCTATGCATCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGG
AATTCGCCCTTgtaatctattcgatataccgttgctaacagtatactggcccaataactgtggat
ggaaaatctataataatacattaatatcatccgatggtgctagggttatttggatggatgcgtat
aaattttcttgcggtttatctttacaagactattgttatcattggggtagcaaaccagagagccg
accattcgatttaataaaaaaatcagatgctaaacgcaattctaaatcgttggtcaaagaatcta
tggcatccttgaaatccttgtacgaggcattcgagacacaatcaggagcgttagaagttttaatg
agtccatgtaggatgttttcgttttctagaatagaagacatgttcttaactagtgtcattaatag
agtatccgagaatactggaatggggatgtattatcctaccaacgatataccttctctatttatcg
aatcatctatctgtctagattatattatagtaaataatcaggaatccaacaaatatcgtatcaaa
tctgttctcgatatcatttcttcaaaacaataccctgcaggacgtcccaactacgttaaaaatgg
tacaaaaggaaagttatatatcgcgttgtgtaaagttaccgtacctactaacgaccatattccag
tagtttatcacgatgatgacaatactaccacctttattacagtattgacgtccgtcgatattgaa
actgctatcagagcaggatattcgatagtcgaattaggggctttacaatgggataataatattcc
agaacttaaaaacggtttactggatagtatcaagatgatttatgacttgaacgcagttacaacaa
ataatttattggaacagctcatagaaaatattaactttaacaactctagtataatttcgttgttt
tatacatttgccattagttattgccgagcattcatttactcaattatggaaaccatagatccggt
gtatatatctcagttcagttataaagaattatacgttagtagctcttataaagatattaatgaat
ccatgagtcagatggtaaaattataaaaagtgaaaaacaatattatttttatcgttggttgttac
actATGGATGCTATGAAGAGGGGCCTGTGCTGCGTGCTGCTGCTGTGTGGCGCCGTGTTTGTGTC
CCCCAGCCAGGAAATCCACGCCCGGTTCAGAAGAGGCAGCAAGCTGGCCGACGAGGACGAGACAG
CCTACGACAGCGGCACCGACCTGATCGACTTCATCGACGACAGCGACGAGAATGAGAACGACTCC
GACACCGGCGAGGACATGGTGGATTTCATCGACAACGACGAAACCGACGAAGAGAGCACCGAGAG
CGACCTGGACGGCTTTATCGACAACTCCGCCCAGCTGGCTGACAGCGACAGCAATGCCTGCGCCT
TCCTGAAGGCTCAGCTGGCAGACGTGAACAGCAACGCCGCTGCTTTTCTGAAGAACTGCATCCTG
CTGTACGGCGCTGCCAACACCGGCAAGAGCCTGTTCGGCATGAGCCTGAACTGCCTGGTGCTGTG
CGGCCCAGCCAATACCGGAAAGTCCTACTTCGGCATGTCCCTGAATTGTCTCGTGATCTACGGCC
CACCTAACACAGGCAAGTCCTGCTTTGCCATGTCTCTGTGGCCCTACCTGCACAGCAGACTGGTG
GTGTTTACCTTCCCCAACCCCTTCTGGCCTTACCTGGAAAGCCGGATCACCGTGTTCGAGTTCCC
CAATGCCTTTCTGAGATACCTGCACTCCCGGATCCACGTGCTGCAGTTTCTGAACCCCTTCAACG
TGTGCCAGGACAAGATCCTGGAACACTACGAGAACGACAGCAAGGACATTCTGGAACATTATGAG
AATGATTCCAAGGACCTGTGCGACCACATCTGCGATCACATCGACTACTGGAAGCACATCCGGCT
GGAATGCGCCATCATGTACAAGGCCCGGATCAGACTGGAATGTGCTATTATGTATAAGGCTCGCG
AGATGGGCTTCCACCAGTTCGACGGCGACATCTGCAACACCATGCACTACACCAACTGGATCTAT
ATCTGCGAGGACGCCCAGTGCACCGTGGTGGAAGGCCAGGTGGACAAGAAATGGGAGGTGCACGC
TGGCGGCCAAGTGATCCTGTGTCCTGAGAGCGGCCAGCGGCGGATCAAGAGGCCCAGAAGCGAGA

ACTGCCACCCCAACAAGCTGCTGATCCTGAAGTGCCTGCGGTACAGATTCAAGAAGCACTGCAAG
CTGAGCAGCACCTGGCACTGGACCTGCCACGACGGCAAGCACAAGTGGCATTGGACATGTCACGA
TGGGAAACACAAGAACGCCATTGTGACCCTGACCTACTACGAGGCCGACAAGAACGACCTGAACG
CCCAGATCGAGCACTGGAAACTGATCCGGATGGAATGTGCAATCTTCTATAAGGCCAAAGAGCTG
GGCATCAGCATCTGCCACCAGGTGGTGCCTCCACTGGCCGCCTCTAAAGCCAAAGCCTGCCAGGC
CATCGAACTGCAGCTGGCCCTGGAAGCCCTGAATGCCAGCCCCTACGATGAGTGGACCCTGCAGC
AGACCAGCCTGGAAATGTGGCTGGCCGAGCCCCAGTTTAAGAAGCACGGCATCACCATCACCGTG
CAGTACGACAATGACAAGGCCAATACCATGGATTACACAAATTGGAAAGAAATCTACGTGATCGT
GTGCCCCGCCAGCATCCCCTCCGATGAGATCAGCACCGAGGAAGCCGACCACATTGATTATTGGA
AAGCCATCAGGCAGGAAAACGCCATCTTCTTCGCCGCCAGACACCAGGTGGTGCCCGCCCTGAAT
ATCTGCAAGGCCAAGGCCTGTAAAGCCATCGAGTGGAACACCGAGCCCAAGCACTGCTTCAAGAA
GGGCGGCCAGCACATCGAAGTGTGGTTCGACTACGTGGCCTGGGACAGCGTGTACTACTGCGGCG
ACGATGGCTGGTGCAAGACCGAGGCCGAGAAGTACGGCTGCAAGGGCACCTGGGAAGTGCATTTC
GGCAACAGCATCGACTGCAACGACTCCATGTGCAGCACCTTCGACGACAACGTGTCCGCCACCGA
GCTCGTGAAGGACCATATCGACTATTGGAAGCTGATTCGCCTGGAATGTGCCATTTTTTACAAGG
CCAGACGGCGGCTGTCCAGCGACCAGGATCAGTCTCAGCTCGTGACCAAGTACCCCCTGCTGAAG
CTGCTGTCCAGACCCCCCAACATGGGCGTGAAGGCCCACGGCAAGTGCATCTGGGAGAACAAGGT
GTTCATCGTGCCCACCCTGTGCCCCGTGCCTCTGGATCCAACATATCCTCTGCTGAAACTGCTGA
CCACCCAGACCACCACCCCCGAGAATACCTCCCTGGTGGAACTGAGAGTGACCACCCCCAAGAGC
ACAGTCGTGATCAGGCTGCACCTGACCACCAGATACCCACTGCTGTCACTGCTGAACAGCTACAG
CACCCCCCCTCACCGGATCCCTGCTCCATGTCCTTGGGCTCCTCAGAGGCCCCCCATCCCTAAGC
CTTCTCCATGGGCCCCTAGAATCCCTGCCCCTTGCCCCTGGGCACCTCCTAGACCTCCACACTGT
CCATGGGTGCCCCCTCCACCTCCTCCAAGACCTTGGGCCCCTTGCTTCCTGCTGTGCTTTTGTGT
GCTGCTGTGCGTGTGCCTGCTGATCAGACCCCTGCTGCTGAGTGTGTCCACCTACCTGAGGCCTC
TGCTGCTGTCTATCAGCGTGTACGCTCAGGTGCTGGTGCTGGTGCTGCTGCTGTGGGTGTCCATC
GGAAGCCTGCTGCCCAGCGTGTGCATGTGTGCCTATGCCTGGGTGCTGGTGTTCGTGTACATCGT
CGTGATTACCAGCCCCGCCACCGCCATCGTGTACCGGGATGGCAATCCTTACGCCGTGTGCGACA
AGTGCCTGAAGTTCTACAGCAAGATCAGCGAGTACCGGCACTACTGCTACAGCCTGTACGGCACC
ACCCTGGAACAGCAGTACAACAAGCCCCTGTGCGATCTGCTGATTCGGTGCATCAACGTGGTGTA
CAGAGACTCCATCCCCCACGCCGCCTGCCACAAGTGTATCGACTTCTACTCCAGAATCAGAGAGC
TGCGGCACTACAGCGACTCCGTGTACGGCGATACCCTGGAAAAGCTGACCAACACTGGCCTGTAC
AACCTGCTGATTAGATGCCTGCGGGTGTTCTGCAAGAAGGCCCTGACAGCCAGCGAGGTGTACAA
CTTCGCCTACACCGATCTGCGGGTGGTGTATCGGGACAGCAAAGTGCGGAAGCTGAGGTACTACA
ACTGCTCTGTGTATGGCGCCAGCCTGGTGTATTGCAAGGGACAGCTGACCGAGACAGAGGTGCTG
GATTTCGCCTTCACAGACCTGACAATCGTGTATCGCGACTCCAAGGTGTCCGAGTTCCGGTGGTA
CAGATATTCCGTGTATGGCACCACACTGTGCGTGGAATGCAAGAAAACCCTGCAGAGATCTGAGG
TGTACGACTGCCAGCGGCCACTGTGTCCGCAGGAAAAGAAAAGACACGTGGACCTGAACAAGCGG
TTCCACACCCTGCACGAGTACATGCTGGATCTGCAGCCCGAGACAACCGACCTGTACTGCTACGA
GCAGCCTGAAACCACTGATCTGCACTGTTATGAGCAGCTGGGAGACAGCTCCGATGAAGAGGACA

CTGGCGGCCTGGATGGGGACGAGGATGAGGACGAAGTGGACCATCTGCAGGAACAGCCCCAGCAG
GCTAGACGGGACGAACAGCACCCTTGCTATCTGATCGAGACACAGTGCTGCAGATGCGAATCTCT
GGTGGAAGAGAACGACGAGATCGACGGCGTGAACCACCAGCATCTGCCCGCTAGAAGGGCCGAGC
CTCAGAGACACACCATGCTGTGTATGTGCTGCAAGTGCGAGGCCAGAATCGCCGGCtaattttta
taaccgagtttctgcattattgtaattcgtatgctggcaccatcaaagaatcacttctaaaagat
atcaatatcacacatacaaatattactaccctattgaatgagacagccaaggttatcaagttagt
aaaatctctggtagataaagaagatactgatattgtgaataatttcattaccaaagaaattaaaa
acagagacaaaatagttaatagtttgtctctatcaaacctggactttcgtttgtaaattggggct
Ttttgtacaataaatgggtgttgccaatgattcatcccctgaatatcaatggatgtctccccata
gattatcagatactgttatattaggagactgtttgtattttaacaatataatgtcccaattagat
ttacaccaaaattgggctccatcagttagattgttaaattattttaagaattttaataaggaaac
actactaaagatagaagagaatgattacattaattcatccttttccaacaaaaggataaacgat
tttatcctataaacgacgatttttatcacatatctacaggaggatatggtatagtctttaagata
gataactatgtagtaaaatttgtattcgaggccacaaaattatatagtcccatggaaactacggc
ggagttcacagtacccaaatttctatacaacaatctaaagggagatgaaaaaaaattaatcgtgt
gtgcgtgggccatgggattaaactataaattaacattttacatactctgtataaacgtgttctt
catatgttgctattattgatacaaactatggatggtcaggaactatcattgagatattcttctaa
agtttttttaaaggcgtttaacgagagaaaggacagtatcaaattcgtgaaattactatcccact
tttatccggcagttattaacagtaatattaatgttataaactattttaaccgcatgtttcacttt
ttcgaacatgaaaagagaactaactacgaatacgaaagaggaaatattataatttttccccctagc
actgtattcggcagataaagtagataccgagctagctatcaaattaggatttaaatctttggtac
aatacataaagtttatcttttacagatggctctgttatacattaaaatttacgaactaccatgc
tgcgacaacttttacacgcagatcttaaacccgataatatcttacttttgattccaatgaacc
aataataattcatctaaaggataaaaagtttgtttttaatgaacgtattaaatcggcattaaacg
actttgacttttcccaagAAGGGCGAATTCTGCAGATATCCATCACACTGGCggccgcTTACTTG
TACAGCTCGTCCATGCCGAGAGTGATCCCGGCGGCGGTCACGAACTCCAGCAGGACCATGTGATC
GCGCTTCTCGTTGGGGTCTTTGCTCAGGGCGGACTGGGTGCTCAGGTAGTGGTTGTCGGGCAGCA
GCACGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCGAGCTGCACGCTGCCGTCC
TCGATGTTGTGGCGGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATATA
GACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTGCCCCAGGATGTTGCCGTCCTCCTTGAAGT
CGATGCCCTTCAGCTCGATGCGGTTCACCAGGGTaTCGCCCTCGAACTTCACCTCGGCGCGGGTC
TTGTAGTTGCCGTCGTCCTTGAAGAAGATGGTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGA
CTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTCA
GGGTGGTCACGAGGGTcGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTC
AGCTTGCCGTAGGTGGCATCGCCCTCGCCCTCGCCGGACACGCTGAACTTGTGGCCGTTTACGTC
GCCGTCCAGCTCGACCAGGATGGGCACCACCCCGGTaAACAGCTCCTCGCCCTTGCTCACCATgt
ttaaacTTTATATTCCAAAAAAAAAAAATAAAATTTCAATTTTTgtttaaacgttGTACGGCAGT
TTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATT
ATTGACACGCCGGGGCGACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAGT

CTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATA
TGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGAC
ATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCATGAGATTATCAAAAAG
GATCTTCACCTAGATCCTTTTCACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGA
ATGTCAGCTACTGGGCTATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGC
AGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACCGGAATTGCC
AGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTCGCCGC
CAAGGATCTGATGGCGCAGGGGATCAAGCTCTGATCAAGAGACAGGATGAGGATCGTTTCGCATG
ATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGA
CTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCC
CGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGG
CTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGG
AAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTG
CCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGC
CCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGT
CGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCA
AGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATC
ATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTA
TCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCT
TCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGAC
GAGTTCTTCTGAATTATTAACGCTTACAATTTCCTGATGCGGTATTTTCTCCTTACGCATCTGTG
CGGTATTTCACACCGCATACAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTT
ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAAT
AATAGCACGTGAGGAGGGCCACCATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGC
GACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGGCTCGGGTTCTCCCGGGACTTCGTGGAGGA
CGACTTCGCCGGTGTGGTCCGGGACGACGTGACCCTGTTCATCAGCGCGGTCCAGGACCAGGTGG
TGCCGGACAACACCCTGGCCTGGGTGTGGGTGCGCGGCCTGGACGAGCTGTACGCCGAGTGGTCG
GAGGTCGTGTCCACGAACTTCCGGGACGCCTCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCC
GTGGGGGCGGGAGTTCGCCCTGCGCGACCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGC
AGGACTGACACGTGCTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGAT
AATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAA
GATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAAC
CACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT
CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGG
TCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAG
ATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATC
CGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTAT

CTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGG
GGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGGCTTTTGCTGGC
CTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTT
GAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGC
GGAAG

HPV2-randomised nucleotide sequence (segments are in a randomisd order) (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO: 71)

ATGGATGCTATGAAGAGGGGCCTGTGCTGCGTGCTGCTGCTGTGTGGCGCCGTGTTTGTGTCCCC
CAGCCAGGAAATCCACGCCCGGTTCAGAAGAGGCAGCAAGCTGGCTGCTCAGCTGGCCGACAGCG
ACAGCAATGCCTGCGCCTTCCTGAAGCTGAGATACCTGCACAGCCGGATCCACGTGCTGCAGTTT
CTGAACCCCTTCACCCTGCACGAGTACATGCTGGACCTGCAGCCCGAGACAACCGACCTGTACTG
CTACGAGCAGGACGAGGATGAGGACGAGGTGGACCATCTGCAGGAACAGCCCCAGCAGGCCAGAA
GGGATGAGCAGCACCCCTGCTACCTGATCGAGACACAGTGCTGCAGATGCGAGAGCCTGGTGGCC
CAGCTGGCTGACGTGAACTCTAACGCCGCAGCCTTTCTGAAGAACAGCATCGACTGCAACGACAG
CATGTGCAGCACCTTCGACGACAACGTGTCCGCCACCGAGCTCGTGAAGAGAATCCCTGCCCCTT
GCCCCTGGGCCCCTGAGGAAAATGACGAGATCGACGGCGTGAACCACCAGCATCTGCCCGCTAGA
AGGGCCGAGCCTCAGAGACACACCATGCTGTGCATGTGCTGCAAGTGCGAGGCCCGGATCTGCTT
CCTGCTGTGCTTTTGTGTGCTGCTGTGCGTGTGCCTGCTGATCAGACCCCTGCTGCTGAGTGTGT
CCACCTACTGCCAGCGGCCGCTGTGTCCTCAGGAAAAGAAACGGCACGTGGACCTGAACAAGCGG
TTCCACATCTACATCTGCGAGGACGCCCAGTGCACCGTGGTGGAAGGCCAGGTGGACAGCAAGGT
GTCCGAGTTCCGGTGGTACAGATACAGCGTGTACGGCACCACCCTGGGCCAGCGGAGAATCAAGA
GGCCCAGATCCGAGGTGTACTGCAAGGGACAGCTGACCGAGACAGAGGTGCTGGACTTCGCCTTC
ACCGACCTGACCATCGTGTACCGGGACATCCGGCTGGAATGCGCCATCATGTACAAGGCCAGAGA
GATGGGCTTCCACGACCACATCGACTACTGGAAGCTGATTAGACTGGAATGTGCTATCTTCTACA
AAGCCCGGATCCTGAAGTGCCTGCGGTACAGATTCAAGAAGCACTGCAAGCTGTACGTGGCCTGG
GACTCCGTGTACTACTGCGGCGACGATGGCTGGTGCAAGACCAGCTCTACCTGGCACTGGACATG
CCACGACGGCAAGCACAAGAACGTGTGCCAGGACAAGATCCTGGAACACTACGAGAACGACTCCA
AGGACGACGAGGACGAGACAGCCTACGACAGCGGCACCGATCTGATCGACTTCATCGACGATAGC
AACTGCCACCCCAACAAGCTGCTGCGGAGACTGAGCAGCGACCAGGACCAGTCTCAGAGGCCCCC
CAACATGGGAGTGAAGGCCCACGGCAAGTGCATCTGGGAGAACAAGGTGTTCATCGTGCCCACCC
TGTGCCCCGTGCCTCTGGATCCAACATACCCCCTGCTGAAGCTGCTGACCCCCGAAACCACAGAT
CTGCACTGTTATGAGCAGCTGGGCGACTCCTCCGACGAAGAGGATACAGGCGGCCTGGATGGCTA
CGAGGCCGACAAGAACGACCTGAACGCCCAGATCGAGCACTGGAAACTGATCCGGATGGAATGTG
CAATTTTCTATAAGGCCAAAGAGCTGGGGATCAGCGACGAGAATGAGAACGACAGCGATACCGGC
GAGGACATGGTGGATTTCATCGACAATGAGGCCGAGAAGTACGGCTGCAAGGGCACCTGGGAGGT

GCACTTCGGCTTTAAGAAGCACGGCATCACCATCACCGTGCAGTACGACAACGACAAGGCCAACA
CCATGGACTACACCAACTGGAAAGAGATCTACCCCCCTCCCCCCCCCACGGCCTTGGGCTCCTCCA
ATTCCTAAGCCCTCTCCATGGGCCCCTCAGTTCGACGGCGACATCTGCAATACCATGCACTATAC
CAATTGGGTGGTGTACAGAGACAGCATCCCCCACGCCGCCTGCCACAAGTGTATCGACTTCTACA
GCAGAATCAGAGAGCTGCGGCACTACAGCGACTCTGTGTACGGCGATACCCTGGAAAAGCTGACC
AACACCGGCCTGTACAATCTGCTGATCCGGTGCCTGAGGCTCGTGACCAAGTATCCTCTGCTGAA
ACTGCTGTCCAACTGCATCCTGTACGGCGCTGCCAATACCGGCAAGAGCCTGTTCGGCATGAGCC
TGAGCAAAGTGCGGAAGCTGAGGTACTACAACTGCTCCGTGTATGGGGCCAGCCTGTGCGTGGAA
TGCAAGAAAACCCTGCAGCGGAGCGAAGTGTACGACGACGAAACCGACGAGGAAAGCACCGAGAG
CGACCTGGACGGCTTCATCGATAACAGCGTGATCGTGTGCCCCGCCTCCATCCCCTCCGATGAGA
TCTCTACCGAGGAAGCCCCCAGACCCCCTCACTGTCCTTGGGTGCCAGTGTTCTGCAAGAAGGCC
CTGACCGCCTCTGAGGTGTACAATTTTGCCTATACCGACCTGCGCGTGGTGTATAGGGACATTCT
GGAACATTATGAGAATGATAGCAAGGACCTGTGCGATCACATCAACTGCCTCGTGATCTACGGCC
CTCCTAACACCGGCAAGTCCTGCTTCGCCATGTCCCTGTGGAACACCGAGCCCAAGCACTGCTTC
AAGAAGGGCGGCCAGCACATCGAAGTGTGGTTCGATATTGTGTACAGGGACGGCAACCCTTACGC
CGTGTGCGACAAGTGCCTGAAGTTCTACTCCAAGATCAGCGAGTACCGCCACTACTGCTACTCCC
TGTATGGCACAACACTGGAACAGCAGTACAACAAGCCCCTGTGCGACCTGCTGATTCGCTGCATC
AACACCACCAGATACCCTCTGCTGTCCCTGCTGAACAGCTACAGCACCCCCCCTCATCGGATTCC
CGCCCCATGTCCATGGGCTCCACAGAGGCCTACCCAGACCACCACCCCCGAGAATACCTCCCTGG
TGGAACTGAGAGTGACCACCCCCAAGAGCACAGTCGTGATCAGGCTGCACCTGTGGCCCTACCTG
CACTCCAGACTGGTGGTGTTCACCTTCCCCAACCCCTTTCACCAGGTGGTGCCCGCCCTGAATAT
CTGCAAGGCCAAGGCCTGCAAAGCCATCGAGAAGAAATGGGAAGTGCACGCTGGCGGCCAAGTGA
TCCTGTGTCCTGAGAGCCTGCGGCCTCTGCTGCTGTCCATTAGCGTGTACGCCCAGGTGCTGGTG
CTGGTGCTGCTGCTGTGGGTGTCCATCGGCAGCAACTGTCTGGTGCTGTGCGGCCCTGCCAACAC
AGGGAAGAGTTACTTCGGCATGTCTCTGATCTGCCATCAGGTGGTGCCTCCACTGGCCGCCTCTA
AGGCTAAAGCCTGTCAGGCCATCGAACTGCAGCTGGCCCTGGAAGCCCTGAATGCCAGCCCCTAT
GATCACATTGATTACTGGAAAGCCATCCGGCAGGAAAATGCCATCTTCTTCGCCGCCAGATGGCA
TTGGACCTGTCACGATGGAAAACACAAGAATGCCATTGTGACCCTGACCTACCTGCTGCCCAGCG
TGTGTATGTGCGCCTACGCTTGGGTGCTGGTGTTCGTGTACATCGTCGTGATTACCAGCCCCGCC
ACCGCCGATGAGTGGACACTGCAGCAGACAAGCCTGGAAATGTGGCTGGCCGAGCCCCAGTGTGA
CCATATCGATTATTGGAAACACATCCGCCTGGAATGTGCTATTATGTATAAGGCCCGGTGGCCTT
ACCTGGAAAGCAGAACCGTGTTCGAGTTCCCCAATGCCTTCGCCGGCTCTGGACCTGGCGCCTCT
GGAAAACCCATCCCCAATCCACTGCTGGGCCTGGACTCCACCCGGACC

HPV2-randomised sequence polypeptide (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO: 72)

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGSKLAAQLADSDSNACAFLKLRYLHSRIHVLQF
LNPFTLHEYMLDLQPETTDLYCYEQDEDEDEVDHLQEQPQQARRDEQHPCYLIETQCCRCESLVA
QLADVNSNAAAFLKNSIDCNDSMCSTFDDNVSATELVKRIPAPCPWAPEENDEIDGVNHQHLPAR
RAEPQRHTMLCMCCKCEARICFLLCFCVLLCVCLLIRPLLLSVSTYCQRPLCPQEKKRHVDLNKR
FHIYICEDAQCTVVEGQVDSKVSEFRWYRYSVYGTTLGQRRIKRPRSEVYCKGQLTETEVLDFAF
TDLTIVYRDIRLECAIMYKAREMGFHDHIDYWKLIRLECAIFYKARILKCLRYRFKKHCKLYVAW
DSVYYCGDDGWCKTSSTWHWTCHDGKHKNVCQDKILEHYENDSKDDEDETAYDSGTDLIDFIDDS
NCHPNKLLRRLSSDQDQSQRPPNMGVKAHGKCIWENKVFIVPTLCPVPLDPTYPLLKLLTPETTD
LHCYEQLGDSSDEEDTGGLDGYEADKNDLNAQIEHWKLIRMECAIFYKAKELGISDENENDSDTG
EDMVDFIDNEAEKYGCKGTWEVHFGFKKHGITITVQYDNDKANTMDYTNWKEIYPPPPPRPWAPP
IPKPSPWAPQFDGDICNTMHYTNWVVYRDSIPHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLT
NTGLYNLLIRCLRLVTKYPLLKLLSNCILYGAANTGKSLFGMSLSKVRKLRYYNCSVYGASLCVE
CKKTLQRSEVYDDETDEESTESDLDGFIDNSVIVCPASIPSDEISTEEAPRPPHCPWVPVFCKKA
LTASEVYNFAYTDLRVVYRDILEHYENDSKDLCDHINCLVIYGPPNTGKSCFAMSLWNTEPKHCF
KKGGQHIEVWFDIVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI
NTTRYPLLSLLNSYSTPPHRIPAPCPWAPQRPTQTTTPENTSLVELRVTTPKSTVVIRLHLWPYL
HSRLVVFTFPNPFHQVVPALNICKAKACKAIEKKWEVHAGGQVILCPESLRPLLLSISVYAQVLV
LVLLLWVSIGSNCLVLCGPANTGKSYFGMSLICHQVVPPLAASKAKACQAIELQLALEALNASPY
DHIDYWKAIRQENAIFFAARWHWTCHDGKHKNAIVTLTYLLPSVCMCAYAWVLVFVYIVVITSPA
TADEWTLQQTSLEMWLAEPQCDHIDYWKHIRLECAIMYKARWPYLESRTVFEFPNAFAGSGPGAS
GKPIPNPLLGLDSTRT**

HPV53del nucleotide sequence (segments from HPV53 have been removed) (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO: 73)

ATGGATGCTATGAAGCGAGGACTGTGCTGCGTGCTGCTGCTGTGTGGCGCTGTGTTTGTGTCCCCTAGCC
AAGAGATCCACGCCAGATTCAGACGGGGCAGCAAACTGGCCGACGAGGATGAGACAGCCTACGACTCTGG
CACCGACCTGATCGACTTCATCGACGACAGCGACGAGAACGAGAATGACAGCGACACCGGCGAGGACATG
GTGGATTTCATCGACAATGCCCAGCTGGCCGACTCCGACTCTAATGCCTGTGCCTTTCTGAAGGCTCAGC
TGGCTGACGTGAACAGCAATGCCGCCGCTTTCCTGAAGAACTGCATCCTGCTGTACGGCGCTGCCAACAC
AGGCAAGAGCCTGTTTGGCATGAGCCTGAACTGCCTGGTGCTGTGCGGACCTGCCAATACCGGCAAAAGC
TACTTCGGCATGTCCCTGTGGCCTTACCTGCACAGCAGACTGGTGGTGTTTACATTCCCCAATCCTTTCT
GGCCCTACCTGGAAAGCCGGATCACCGTGTTCGAGTTCCCCAACGCCTTCAACGTGTGCCAGGACAAGAT
CCTGGAACACTATGAGAACGACAGCAAGGACATCCTTGAGCACTACGAAAACGACTCCAAGGACCTGTGC
GACCACATCTGCGATCACATCGACTACTGGAAGCACATCCGGCTGGAATGCGCCATCATGTACAAGGCCC
GGATCAGACTGGAATGTGCTATTATGTATAAGGCTCGCGAGATGGGCTTCCACCAGTTCGACGGCGACAT
CTGCAACACCATGCACTACACCAACTGGATCTATATCTGCGAGGACGCCCAGTGCACCGTGGTGGAAGGA
CAGGTGGACAAGAAATGGGAAGTGCACGCTGGCGGCCAAGTGATTCTGTGTCCTGAGAGCGGCCAGCGGA
GAATCAAGAGGCCCAGATCCGAGAACTGTCACCCCAACAAGCTGCTGATCCTGAAGTGCCTGCGGTACAG
ATTCAAGAAGCACTGCAAGCTGAGCAGCACCTGGCACTGGACATGCCACGATGGCAAGCACAAGTGGCAT
TGGACCTGTCACGACGGGAAACACAAGAACGCCATCGTGACCCTGACCTACTACGAGGCCGACAAGAACG
ACCTGAACGCCCAGATTGAGCACTGGAAACTGATCCGGATGGAATGTGCAATCTTCTATAAGGCCAAAGA
GCTGGGGATCAGCATCTGCCACCAGGTGGTGCCTCCACTGGCTGCCTCTAAAGCCAAAGCCTGTCAGGCC
ATCGAACTGCAGCTGGCCCTGGAAGCCCTGAACGCTAGCCCTTACGATGAGTGGACCCTGCAGCAGACCA
GCCTGGAAATGTGGCTGGCCGAGCCTCAGTTTAAGAAGCACGGCATCACCATCACCGTGCAGTACGACAA
CGACAAGGCCAATACCATGGATTACACGAATTGGAAAGAAATCTACGTGATCGTGTGCCCCGCCAGCATT

CCCTCCGATGAGATCTCTACCGAGGAAGCCGACCACATTGATTATTGGAAGGCCATCCGGCAAGAGAATG
CCATCTTCTTCGCCGCCAGACATCAGGTGGTCCCCGCTCTGAATATCTGCAAGGCCAAGGCCTGCAAAGC
CATCGAGTGGAACACCGAGCCTAAGCACTGCTTCAAGAAAGGCGGCCAGCACATCGAAGTTTGGTTCGAC
TACGTGGCCTGGGACAGCGTGTACTACTGCGGAGATGATGGCTGGTGCAAGACCGAGGCCGAGAAGTACG
GCTGTAAAGGCACCTGGGAAGTCCACTTCGGCAACAGCATCGACTGCAACGATAGCATGTGCAGCACCTT
CGACGACAACGTGTCCGCCACAGAGCTGGTCAAGGACCATATAGACTATTGGAAGCTGATCAGGCTTGAG
TGCGCCATTTTCTACAAGGCCAGACGGCGGCTGTCCAGCGACCAGGATCAATCTCAGCTCGTGACCAAGT
ATCCCCTGCTGAAGCTGCTGTCTACCCAGACCACCACACCTGAGAACACAAGCCTGGTGGAACTGAGAGT
GACCACACCTAAGAGCACCGTCGTGATTCGGCTGCACCTGACCACAAGATACCCTCTGCTGAGCCTGCTG
AACAGCTACAGCACCCCTCCACACAGGATCCCCGCTCCATGTCCTTGGGCTCCTCAGAGGCCTCCTATTC
CTAAGCCTTCTCCATGGGCTCCTAGAATCCCCGCACCTTGTCCATGGGCACCACCAAGACCTCCACATTG
CCCTTGGGTGCCCTGTTTCCTGCTGTGCTTTTGCGTGCTCCTGTGCGTGTGCCTGCTGATCAGACCTCTG
CTGCTGAGCGTGTCCACCTACCTTAGACCACTGCTCCTGTCCATCTCCGTGTACGCACAGGTGCTGGTGC
TGGTCCTGCTTCTGTGGGTGTCCATCGGAAGCCTGCTGCCTAGCGTGTGCATGTGTGCCTATGCTTGGGT
GCTCGTGTTCGTGTACATCGTGGTCATCACAAGCCCCGCCACAGCCATCGTGTACAGAGATGGCAATCCC
TACGCCGTGTGCGACAAGTGCCTGAAGTTCTACAGCAAGATCAGCGAGTACCGGCACTACTGCTACAGCC
TGTACGGCACCACACTGGAACAGCAGTACAACAAGCCCCTGTGCGATCTGCTGATTCGGTGCATCAACGT
GGTGTACCGGGACAGCATTCCTCACGCCGCCTGCCACAAGTGCATCGACTTCTACTCCAGAATCAGAGAG
CTGCGGCACTACAGCGACTCTGTGTACGGCGACACCCTGGAAAAGCTGACCAACACCGGCCTGTACAACC
TGCTGATTAGATGCCTGCGGGTGTACTGCAAGGGACAGCTGACAGAGACAGAGGTGCTGGACTTCGCCTT
CACCGATCTGACAATCGTGTATCGGGATAGCAAGGTGTCCGAGTTCCGGTGGTACAGATATAGCGTGTAC
GGAACAACCCTGTGCGTCGAGTGCAAGAAAACCCTGCAGAGAAGCGAGGTGTACGACTGCCAGAGGCCAC
TGTGCCCTCAAGAGAAGAAACGGCACGTGGACCTGAACAAGCGGTTTCACACCCTGCACGAGTACATGCT
GGACCTGCAGCCTGAGACAACCGACCTGTACTGCTACGAGCAGCCCGAAACCACAGATCTGCACTGTTAT
GAGCAGCTGGGCGACAGCAGCGACGAAGAGGATACAGGCGGACTGGACGGCGAGGAAAACGACGAAATTG
ACGGCGTGAACCACCAGCATCTCCCCGCCAGAAGGGCTGAACCTCAGAGACACACCATGCTGTGTATGTG
CTGCAAGTGCGAGGCCAGAATCGCCTGATGA

HPV53del polypeptide sequence (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO:74)

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGSKLADEDETAYDSGTDLIDFIDDSDENENDSDTGEDM
VDFIDNAQLADSDSNACAFLKAQLADVNSNAAAFLKNCILLYGAANTGKSLFGMSLNCLVLCGPANTGKS
YFGMSLWPYLHSRLVVFTFPNPFWPYLESRITVFEFPNAFNVCQDKILEHYENDSKDILEHYENDSKDLC
DHICDHIDYWKHIRLECAIMYKARIRLECAIMYKAREMGFHQFDGDICNTMHYTNWIYICEDAQCTVVEG
QVDKKWEVHAGGQVILCPESGQRRIKRPRSENCHPNKLLILKCLRYRFKKHCKLSSTWHWTCHDGKHKWH
WTCHDGKHKNAIVTLTYYEADKNDLNAQIEHWKLIRMECAIFYKAKELGISICHQVVPPLAASKAKACQA
IELQLALEALNASPYDEWTLQQTSLEMWLAEPQFKKHGITITVQYDNDKANTMDYTNWKEIYVIVCPASI
PSDEISTEEADHIDYWKAIRQENAIFFAARHQVVPALNICKAKACKAIEWNTEPKHCFKKGGQHIEVWFD
YVAWDSVYYCGDDGWCKTEAEKYGCKGTWEVHFGNSIDCNDSMCSTFDDNVSATELVKDHIDYWKLIRLE
CAIFYKARRRLSSDQDQSQLVTKYPLLKLLSTQTTTPENTSLVELRVTTPKSTVVIRLHLTTRYPLLSLL
NSYSTPPHRIPAPCPWAPQRPPIPKPSPWAPRIPAPCPWAPPRPPHCPWVPCFLLCFCVLLCVCLLIRPL
LLSVSTYLRPLLLSISVYAQVLVLVLLLWVSIGSLLPSVCMCAYAWVLVFVYIVVITSPATAIVYRDGNP
YAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINVVYRDSIPHAACHKCIDFYSRIRE
LRHYSDSVYGDTLEKLTNTGLYNLLIRCLRVYCKGQLTETEVLDFAFTDLTIVYRDSKVSEFRWYRYSVY
GTTLCVECKKTLQRSEVYDCQRPLCPQEKKRHVDLNKRFHTLHEYMLDLQPETTDLYCYEQPETTDLHCY
EQLGDSSDEEDTGGLDGEENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCEARIA**

HPV3-linkers nucleotide sequence (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO: 75)

ATGGATGCTATGAAGAGGGGCCTGTGCTGCGTGCTGCTGCTGTGTGGCGCCGTGTTTGTGTCCCCCAGCC
AGGAAATCCACGCCCGGTTCAGAAGAGGCAGCAAGCTGGCCGACGAGGACGAGACAGCCTACGACAGCGG
CACCGACCTGATCGACTTCATCGACGATAGCGCCGCTGCCGACGAGAATGAGAACGACAGCGATACCGGC
GAGGACATGGTGGATTTCATCGACAACGCTGCCGCCGACGAAACCGACGAAGAGAGCACCGAGAGCGACC

TGGACGGCTTTATCGACAACAGCGCAGCCGCCCAGCTGGCTGACAGCGACTCTAATGCCTGCGCCTTCCT
GAAGGCCGCTGCTCAGCTGGCAGACGTGAACAGCAATGCCGCCGCTTTTCTGAAGGCTGCCGCCAACTGC
ATCCTGCTGTACGGCGCTGCCAACACCGGCAAGAGCCTGTTCGGCATGTCTCTGGCCGCAGCCAACTGCC
TGGTGCTGTGCGGACCTGCCAATACTGGCAAAAGCTACTTCGGCATGAGCCTGGCAGCCGCCAATTGTCT
CGTGATCTACGGCCCTCCTAATACCGGCAAGTCCTGCTTTGCCATGAGTCTGGCCGCTGCCTGGCCCTAC
CTGCACTCTAGACTGGTGGTGTTCACCTTCCCCAACCCCTTCGCTGCCGCTTGGCCTTACCTGGAAAGCC
GGATCACCGTGTTCGAGTTCCCCAATGCCTTCGCCGCAGCCCTGAGATACCTGCACAGCAGAATCCACGT
GCTGCAGTTTCTGAACCCCTTTGCCGCCGCAAACGTGTGCCAGGACAAGATCCTGGAACACTACGAGAAC
GACTCCAAGGATGCCGCTGCCATTCTGGAACATTATGAGAATGATAGCAAGGACCTGTGCGACCACATTG
CTGCCGCCTGCGATCACATCGACTACTGGAAGCACATCCGGCTGGAATGCGCCATCATGTACAAGGCCAG
AGCCGCCGCTATCAGACTGGAATGTGCTATTATGTATAAGGCTCGCGAGATGGGCTTCCACGCTGCTGCC
CAGTTCGACGGCGACATCTGCAACACCATGCACTACACCAACTGGGCTGCCGCTATCTACATCTGCGAGG
ACGCCCAGTGCACCGTGGTGGAAGGACAGGTGGACGCCGCTGCTAAGAAATGGGAGGTGCACGCTGGCGG
CCAAGTGATCCTGTGTCCAGAGTCTGCTGCCGCAGGCCAGCGGAGAATCAAGAGGCCTAGAAGCGAGGCA
GCCGCTAACTGCCACCCCAACAAACTGCTGGCTGCTGCCATCCTGAAGTGCCTGCGGTACAGATTCAAGA
AGCACTGCAAACTGGCTGCAGCTAGCAGCACCTGGCACTGGACCTGTCACGACGGCAAGCACAAAGCCGC
CGCATGGCATTGGACATGCCACGATGGAAAACACAAGAACGCCATCGTGACCCTGACCTATGCAGCCGCC
TACGAGGCCGACAAGAACGACCTGAACGCCCAGATCGAGCACTGGAAGCTGATCAGGATGGAATGTGCAA
TCTTCTATAAGGCCAAAGAGCTGGGCATCAGCGCTGCCGCAATCTGCCACCAGGTGGTGCCTCCACTGGC
CGCCTCTAAAGCCAAAGCCTGCCAGGCCATCGAACTGCAGCTGGCCCTGGAAGCCCTGAATGCCAGCCCT
TATGCCGCAGCCGATGAGTGGACCCTGCAGCAGACCAGCCTGGAAATGTGGCTGGCCGAACCTCAGGCCG
CAGCTTTTAAGAAGCACGGCATCACCATCACCGTGCAGTACGACAACGACAAGGCCAATACCATGGATTA
CACCAATTGGAAAGAGATCTACGCCGCAGCTGTGATCGTGTGCCCCGCCAGCATCCCTAGCGACGAGATC
AGCACAGAGGAAGCAGCCGCCGACCACATCGATTATTGGAAAGCCATCAGACAGGAAAACGCCATCTTCT
TCGCCGCTAGAGCCGCTGCCCACCAGGTGGTGCCAGCCCTGAATATCTGCAAGGCCAAGGCCTGTAAAGC
CATCGAAGCCGCTGCTTGGAACACCGAGCCCAAGCACTGCTTCAAGAAGGGCGGCCAGCACATCGAAGTG
TGGTTCGACGCTGCAGCCTACGTGGCCTGGGACAGCGTGTACTACTGTGGCGACGACGGCTGGTGCAAGA
CCGCCGCTGCAGAGGCCGAGAAGTATGGCTGCAAGGGCACCTGGGAAGTGCATTTCGGCGCAGCTGCCAA
CTCCATCGACTGCAACGACAGCATGTGCAGCACCTTCGACGACAACGTGTCCGCCACCGAGCTCGTGAAA
GCTGCCGCTGACCATATTGATTACTGGAAACTGATTCGCCTGGAATGCGCTATTTTCTACAAAGCCAGGG
CCGCAGCACGGCGGCTGTCCTCAGATCAGGATCAGAGCCAGGCTGCTGCACTCGTGACCAAGTACCCCCT
GCTGAAGCTGCTGAGCGCCGCAGCAAGACCCCCCAACATGGGAGTGAAGGCCCACGGCAAGTGCATCTGG
GAGAACAAGGTGTTCATCGTGCCCACCCTGTGCCCCGTGCCTCTGGATCCAACATATCCTCTGCTGAAAC
TGCTGACCGCTGCCGCCACCCAGACCACCACACCTGAGAATACCTCCCTGGTGGAACTGAGAGTGACCAC
CCCCAAGAGCACAGTCGTGATCAGGCTGCACCTGGCTGCCGCAACCACCAGATACCCTCTGCTGTCCCTG
CTGAACAGCTACAGCACCCCCCCTCATCGGATCCCTGCCCCTTGTCCTTGGGCTCCTCAGAGGCCTGCCG
CTGCACCTATCCCTAAGCCTTCTCCATGGGCCCCTGCCGCAGCTAGAATCCCAGCTCCATGTCCATGGGC
ACCAGCTGCTGCTCCCAGACCTCCTCATTGCCCTTGGGTGCCAGCAGCCGCTCCTCCACCTCCTCCTAGA
CCTTGGGCCCCAGCCGCCGCTTGTTTCCTGCTGTGCTTCTGTGTGCTGCTGTGCGTGTGCCTGCTGATCA
GACCCCTGCTGCTGAGTGTGTCCACCTACGCAGCTGCTCTGCGGCCACTGCTGCTGTCCATCTCTGTGTA
CGCACAGGTGCTGGTGCTGGTGCTGCTGCTGTGGGTGTCCATCGGATCTGCCGCAGCACTGCTGCCCTCC
GTGTGCATGTGTGCCTATGCCTGGGTGCTGGTGTTCGTGTACATCGTCGTGATTACCAGCCCCGCCACCG
CAGCCGCAATCGTGTACAGGGACGGCAACCCTTACGCCGTGTGCGACAAGTGCCTGAAGTTCTACAGCAA
GATCAGCGAGTACCGCCACTACTGCTACAGCCTGTACGGCACCACCCTGGAACAGCAGTACAACAAGCCC
CTGTGCGATCTGCTGATCCGGTGCATCAACGCAGCCGCTGTGGTGTACAGAGACAGCATCCCACACGCCG
CCTGCCACAAGTGTATCGACTTCTACTCCCGGATCAGAGAGCTGAGACACTACTCCGACTCCGTGTACGG
CGATACCCTGGAAAAGCTGACCAATACCGGCCTGTACAACCTGCTGATTAGATGCCTGCGGGCAGCCGCA
GTGTTCTGCAAGAAAGCCCTGACCGCCAGCGAGGTGTACAACTTCGCCTACACCGATCTGCGGGTGGTGT
ACCGGGATGCTGCTGCCTCCAAAGTGCGGAAGCTGCGGTACTACAACTGCTCTGTGTATGGCGCCTCCCT
GGCAGCTGCCGTGTATTGCAAGGGACAGCTGACCGAGACAGAGGTGCTGGATTTCGCCTTCACAGACCTG
ACCATCGTGTATAGAGATGCAGCTGCTAGCAAGGTGTCCGAGTTCCGGTGGTACAGATATAGCGTGTACG
GAACAACACTGGCAGCAGCTTGCGTGGAATGCAAGAAAACACTGCAGCGGAGCGAAGTGTACGATGCTGC
AGCTTGCCAGAGGCCGCTGTGTCCTCAGGAAAAGAAAAGACACGTGGACCTGAACAAGCGGTTCCACGCA
GCAGCTACCCTGCACGAGTACATGCTGGACCTGCAGCCCGAGACAACCGACCTGTACTGCTACGAGCAGG
CAGCTGCACCCGAAACCACAGATCTGCACTGTTATGAGCAGCTGGGAGACAGCTCCGATGAAGAGGACAC
CGGCGGACTGGATGCTGCCGCTGGGGATGAGGACGAGGATGAGGTGGACCATCTGCAGGAACAGCCCCAG
CAGGCCAGAAGGGATGAGCAGCACCCCTGCTATCTGATCGAGACACAGTGCTGCAGATGCGAGAGCCTGG
TGGCCGCTGCTGAGGAAAACGACGAGATCGACGGCGTGAACCACCAGCATCTGCCCGCTAGAAGGGCCGA

GCCTCAGAGACACACCATGCTGTGTATGTGTTGCAAGTGCGAGGCCCGGATCGCCGGATCTGGACCTGGC
GCTAGCGGAAAGCCCATCCCCAATCCACTGCTGGGCCTGGACTCCACCCGGACCTGATAA

HPV3-linkers polypeptide sequence (includes tPA leading sequence and HindIII cloning linker, underlined) (SEQ ID NO: 76)

MDAMKRGLCCVLLLCGAVFVSPSQEIHARFRRGSKLADEDETAYDSGTDLIDFIDDSAAADENENDSDTG
EDMVDFIDNAAADETDEESTESDLDGFIDNSAAAQLADSDSNACAFLKAAAQLADVNSNAAAFLKAAANC
ILLYGAANTGKSLFGMSLAAANCLVLCGPANTGKSYFGMSLAAANCLVIYGPPNTGKSCFAMSLAAAWPY
LHSRLVVFTFPNPFAAAWPYLESRITVFEFPNAFAAALRYLHSRIHVLQFLNPFAAANVCQDKILEHYEN
DSKDAAAILEHYENDSKDLCDHIAAACDHIDYWKHIRLECAIMYKARAAAIRLECAIMYKAREMGFHAAA
QFDGDICNTMHYTNWAAAIYICEDAQCTVVEGQVDAAAKKWEVHAGGQVILCPESAAAGQRRIKRPRSEA
AANCHPNKLLAAAILKCLRYRFKKHCKLAAASSTWHWTCHDGKHKAAAWHWTCHDGKHKNAIVTLTYAAA
YEADKNDLNAQIEHWKLIRMECAIFYKAKELGISAAAICHQVVPPLAASKAKACQAIELQLALEALNASP
YAAADEWTLQQTSLEMWLAEPQAAAFKKHGITITVQYDNDKANTMDYTNWKEIYAAAVIVCPASIPSDEI
STEEAAADHIDYWKAIRQENAIFFAARAAAHQVVPALNICKAKACKAIEAAAWNTEPKHCFKKGGQHIEV
WFDAAAYVAWDSVYYCGDDGWCKTAAAEAEKYGCKGTWEVHFGAAANSIDCNDSMCSTFDDNVSATELVK
AAADHIDYWKLIRLECAIFYKARAAARRLSSDQDQSQAAALVTKYPLLKLLSAAARPPNMGVKAHGKCIW
ENKVFIVPTLCPVPLDPTYPLLKLLTAAATQTTTPENTSLVELRVTTPKSTVVIRLHLAAATTRYPLLSL
LNSYSTPPHRIPAPCPWAPQRPAAAPIPKPSPWAPAAARIPAPCPWAPAAAPRPPHCPWVPAAAPPPPPR
PWAPAAACFLLCFCVLLCVCLLIRPLLLSVSTYAAALRPLLLSISVYAQVLVLVLLLWVSIGSAAALLPS
VCMCAYAWVLVFVYIVVITSPATAAAIVYRDGNPYAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKP
LCDLLIRCINAAAVVYRDSIPHAACHKCIDFYSRIRELRHYSDSVYGDTLEKLTNTGLYNLLIRCLRAAA
VFCKKALTASEVYNFAYTDLRVVYRDAAASKVRKLRYYNCSVYGASLAAAVYCKGQLTETEVLDFAFTDL
TIVYRDAAASKVSEFRWYRYSVYGTTLAAACVECKKTLQRSEVYDAAACQRPLCPQEKKRHVDLNKRFHA
AATLHEYMLDLQPETTDLYCYEQAAAPETTDLHCYEQLGDSSDEEDTGGLDAAAGDEDEDEVDHLQEQPQ
QARRDEQHPCYLIETQCCRCESLVAAAEENDEIDGVNHQHLPARRAEPQRHTMLCMCCKCEARIAGSGPG
ASGKPIPNPLLGLDSTRT**

CLAIMS

1. A nucleic acid encoding a polypeptide comprising a plurality of conserved peptide sequences, or variants thereof,

wherein the conserved sequences are conserved across one or more HPV genotypes 16, 18, 31, 52, 53, and 58; and

wherein the polypeptide comprises a conserved peptide sequence of each of the HPV proteins E1, E2, E4, E5, E6, and E7.

2. The nucleic acid according to claim 1, wherein the polypeptide is a fusion polypeptide.

3. The nucleic acid according to claim 1 or claim 2, wherein the polypeptide does not comprise a complete gene sequence.

4. The nucleic acid according to any preceding claim, wherein the polypeptide consists essentially of conserved peptide sequences, a peptide adjuvant sequence and one or more linkers therebetween.

5. The nucleic acid according to any preceding claim, wherein the plurality of conserved peptide sequences comprises 10 or more conserved peptide sequences.

6. The nucleic acid according to any preceding claim, wherein the plurality of conserved peptide sequences comprises 40 or more conserved peptide sequences.

7. The nucleic acid according to any preceding claim, wherein the encoded polypeptide comprises at least 3 different conserved peptide sequences of each of the HPV proteins E1, E2, E4, E5, E6, and E7.

8. The nucleic acid according to any preceding claim, wherein the plurality of conserved peptide sequences are derived from distinct regions of sequence relative to each other and are not naturally concurrent in wild type.

9. The nucleic acid according to any preceding claim, wherein linker residues are provided between one or more, or all, conserved peptide sequences.

10. The nucleic acid according to claim 9, wherein the linker comprises or consists of K, G, P or S amino acid residues, or combinations thereof.

11. The nucleic acid according to any preceding claim, wherein the conserved peptide sequences are selected from any of the group comprising SEQ ID NOs: 1 to 59; variants thereof or combinations thereof, in any order.

12. The nucleic acid according to any preceding claim, wherein the polypeptide comprises one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11; one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35; one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44; one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47; one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55; and one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59.

13. The nucleic acid according to any preceding claim, wherein the polypeptide comprises:

one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E1 sequence;

one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E2 sequence;

one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E4 sequence;

one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E5 sequence;

one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E6 sequence; and

one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52, 53, and 58 are represented by at least one conserved E7 sequence.

14. The nucleic acid according to any preceding claim, wherein the polypeptide comprises:

one or more conserved E1 sequence(s) selected from any one of SEQ ID NOs: 1-11, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E1 sequence;

one or more conserved E2 sequence(s) selected from any one of SEQ ID NOs: 12-35, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E2 sequence;

one or more conserved E4 sequence(s) selected from any one of SEQ ID NOs: 36-44, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E4 sequence;

one or more conserved E5 sequence(s) selected from any one of SEQ ID NOs: 45-47, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E5 sequence;

one or more conserved E6 sequence(s) selected from any one of SEQ ID NOs: 48-55, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E6 sequence; and

one or more conserved E7 sequence(s) selected from any one of SEQ ID NOs: 56-59, wherein each of the genotypes 16, 18, 31, 52 and 58 are represented by at least one conserved E7 sequence.

15. The nucleic acid according to any preceding claim, wherein the nucleic acid comprises or consists of the sequence of SEQ ID NO: 60, or variants thereof; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 60, or variants thereof, and without encoding the TPA lead sequence; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 60, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 65, or variants thereof.

16. The nucleic acid according to any preceding claim, wherein the nucleic acid comprises or consists of the sequence of SEQ ID NO: 62, or variants thereof; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 62, or variants thereof, and without encoding the TPA lead sequence; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 62, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence.

17. The nucleic acid according to any preceding claim, wherein the nucleic acid comprises or consists of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof, and without encoding the TPA lead sequence; or

the nucleic acid comprises or consists of the sequence of SEQ ID NO: 71, 73 or 75, or variants thereof, with a different/alternative peptide adjuvant encoded than the TPA lead sequence.

71, 73 or 75,

18. The nucleic acid according to any preceding claim, wherein the polypeptide comprises or consists of the sequence of SEQ ID NO: 61, or variants thereof.

the polypeptide comprises or consists of the sequence of SEQ ID NO: 61, or variants thereof, and without the TPA lead sequence; or

the polypeptide comprises or consists of the sequence of SEQ ID NO: 61, or variants thereof, with a different/alternative peptide adjuvant than the TPA lead sequence; or

the polypeptide comprises or consists of the sequence of SEQ ID NO: 66, or variants thereof.

19. The nucleic acid according to any preceding claim, wherein the polypeptide comprises or consists of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof.

the polypeptide comprises or consists of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof, and without the TPA lead sequence; or

the polypeptide comprises or consists of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof, with a different/alternative peptide adjuvant than the TPA lead sequence; or

the polypeptide comprises or consists of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof.

20. The nucleic acid according to any preceding claim, wherein the conserved peptide sequences are from about 7 to 250 amino acids in length, or more.

21. The nucleic acid according to any preceding claim, wherein the polypeptide further comprises a peptide adjuvant.

22. The nucleic acid according to any preceding claim, wherein the nucleic acid is a plasmid vector for vaccination.

23. The nucleic acid according to any preceding claim, wherein the nucleic acid comprises viral vector sequence.

24. A vaccine comprising a nucleic acid according to any preceding claim.

25. The vaccine according to claim 24, comprising combinations of different nucleic acids encoding different polypeptides according to any of claims 1-20.

26. The vaccine according to claim 24 or 25, further comprising another therapeutically or prophylactically active ingredient.

27. The vaccine according to any one of claims 24-26, further comprising an adjuvant.

28. A composition comprising a plurality of different nucleic acids according to any one of claims 1-23, optionally wherein the composition is a pharmaceutically acceptable composition.

29. A polypeptide encoded by the nucleic acid according to any one of claims 1-23.

30. A viral vector comprising the nucleic acid according to any one of claims 1-23.

31. The viral vector according to claim 30, wherein the virus is adenovirus.

32. A composition comprising one or more of:

-the nucleic acid according to any one of claims 1-23

-the polypeptide according to claim 29; and/or

-the viral vector according claim 30 or 31.

33. The composition according to claim 32, wherein the composition is immunogenic in a mammal.

34. The composition according to claim 32 or 33, further comprising a pharmaceutically acceptable carrier.

35. The nucleic acid according to any one of claims 1-23; the vaccine according to any one of claims 24-27; the polypeptide according to claim 29; the viral vector according claim 30 or 31; or the composition according to any one of claims 28, or 32-34, for use in the prophylaxis or treatment of HPV infection.

36. A method of treatment or prophylaxis of HPV infection comprising the administration of the nucleic acid according to any one of claims 1-23; the vaccine according to any one of claims 24-27; the polypeptide according to claim 29; the viral vector according claim 30 or 31; or the composition according to any one of claims 28, or 32-34.

37. An agent for use in the prophylaxis or treatment of HPV infection, the agent comprising or consisting of the nucleic acid according to any one of claims 1-23; the polypeptide according to claim 29; the viral vector according claim 30 or 31.

38. A prime boost vaccination kit comprising:

-a prime vaccination comprising the nucleic acid according to any one of claims 1-23; the vaccine according to any one of claims 24-27; the polypeptide according to claim 29; the viral vector according claim 27 or 28; or the composition according to any one of claims 28, or 32-34; and

-a boost vaccination comprising a nucleic acid according to any one of claims 1-23; the vaccine according to any one of claims 24-27; the polypeptide according to claim 26; the viral vector according claim 30 or 31; or the composition according to any one of claims 28, or 32-34.

39. The prime boost vaccination kit according to claim 38, wherein the prime and boost vaccination are different.

40. The prime boost vaccination kit according to claim 39, wherein the prime and boost vaccination differ in the encoded polypeptide sequence and/or comprise different viral vectors.

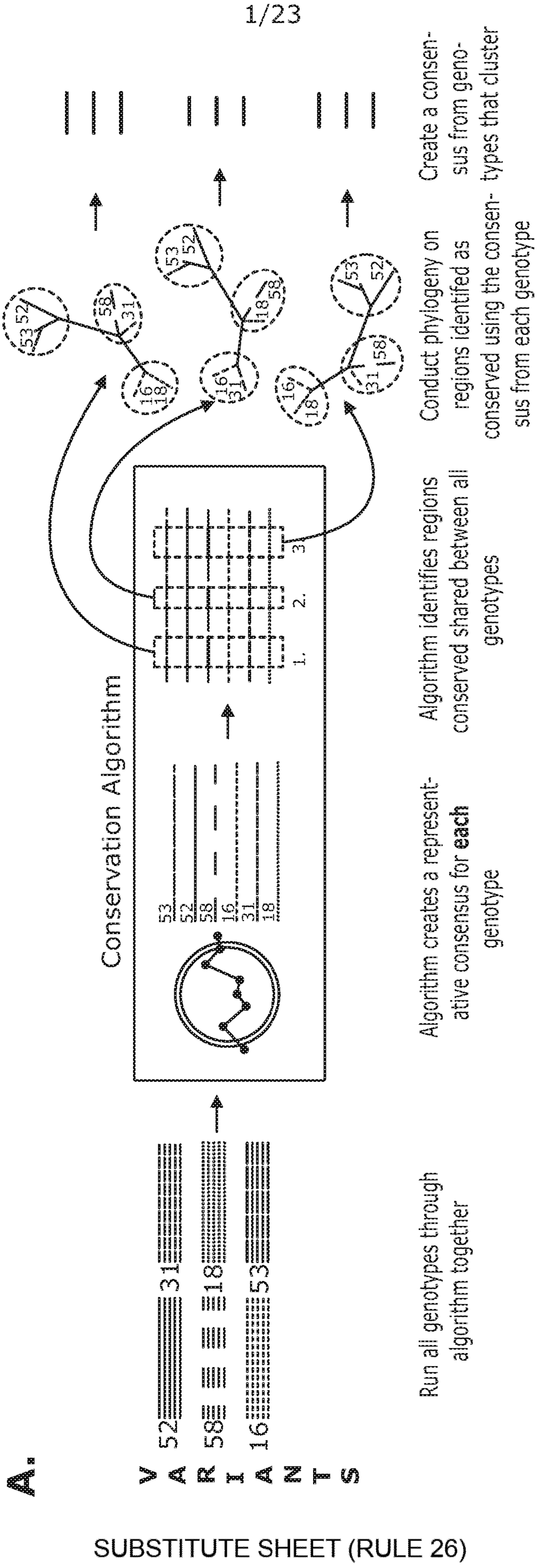

A.
VARIANTS
52
58
16
53
31
18
Run all genotypes through algorithm together
Algorithm creates a representative consensus for each genotype
Conservation Algorithm
1.
2.
3.
Algorithm identifies regions conserved shared between all genotypes
Conduct phylogeny on regions identifed as conserved using the consensus from each genotype
Create a consensus from genotypes that cluster

Figure 1

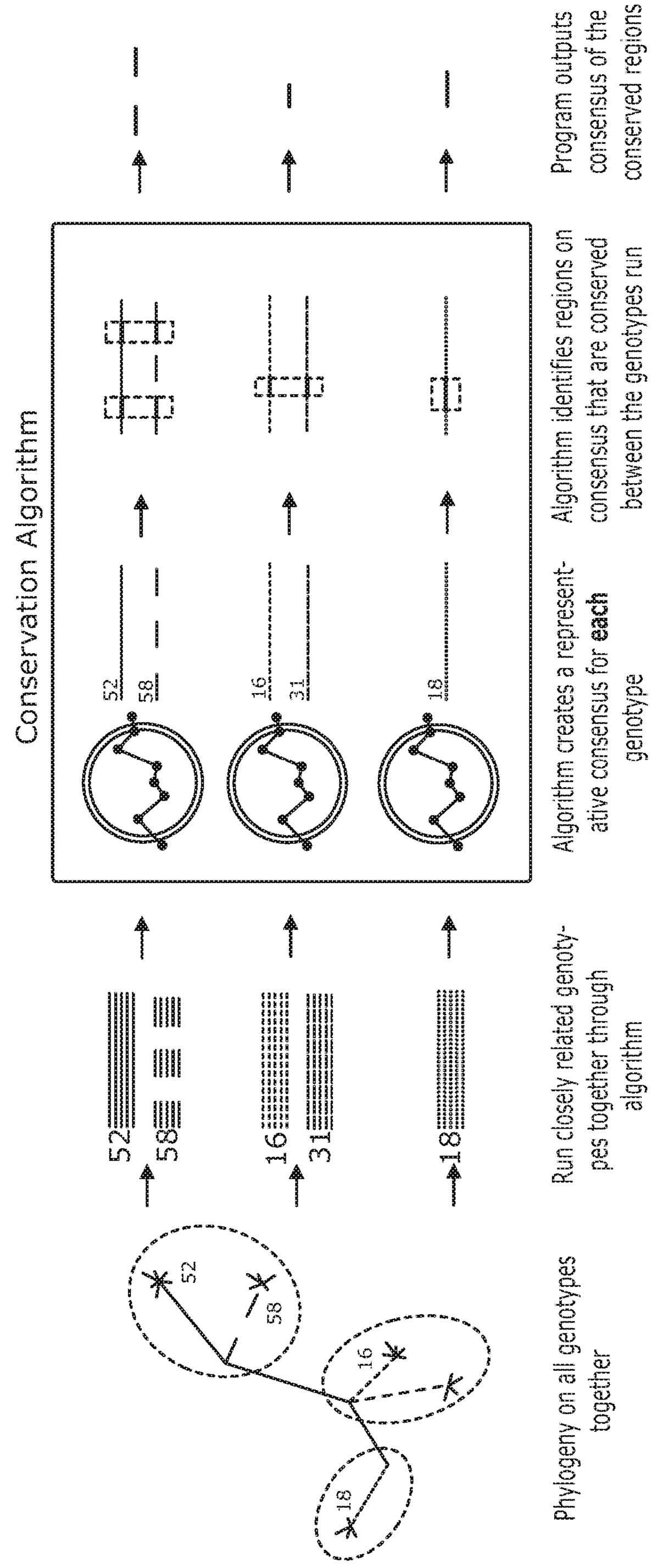


Figure 1 continued

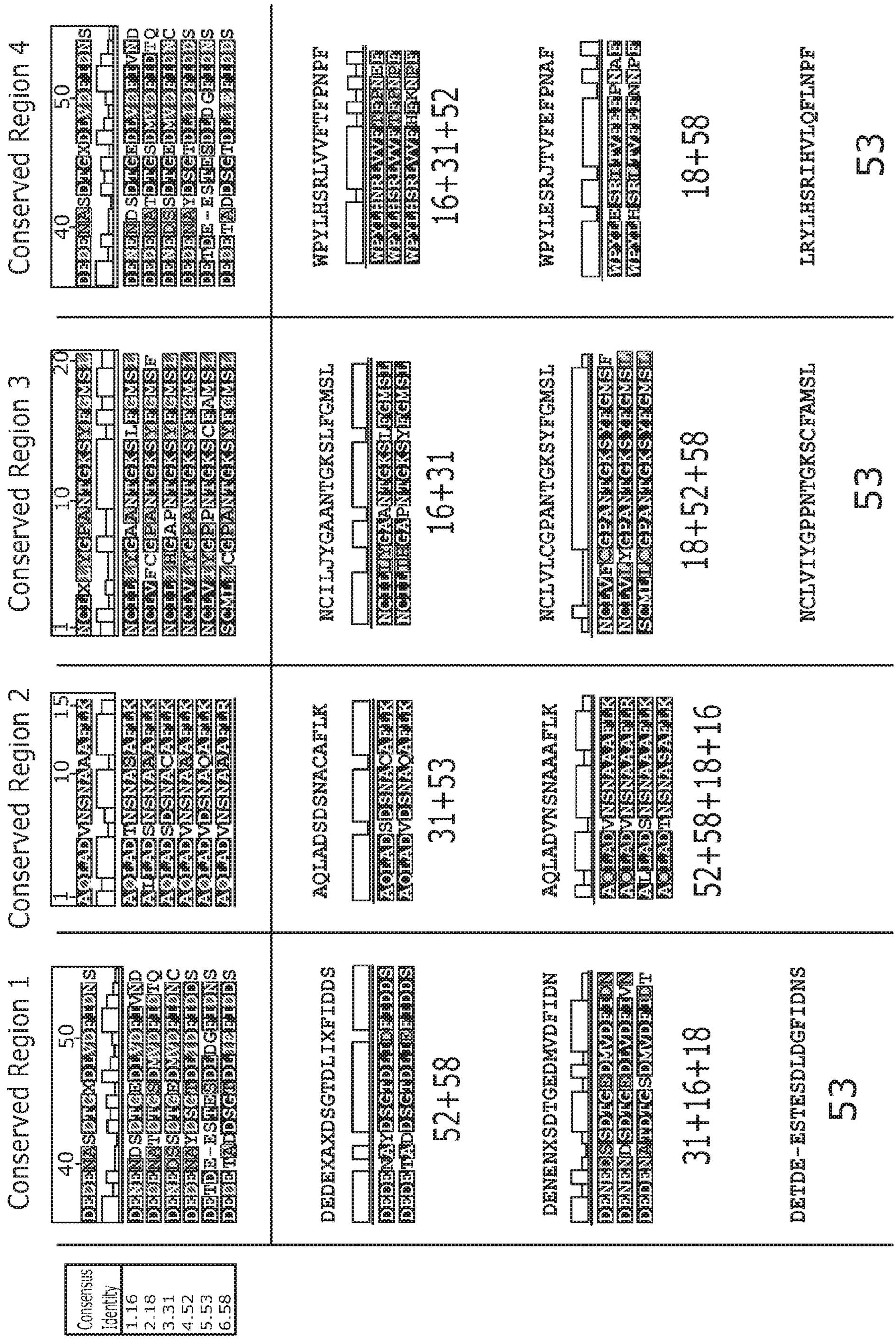

Conserved Region 4
Conserved Region 3
Conserved Region 2
Conserved Region 1
Consensus Identity
1.16
2.18
3.31
4.52
5.53
6.58
16+31+52
18+58
53
16+31
18+52+58
53
31+53
52+58+18+16
52+58
31+16+18
53

Figure 2

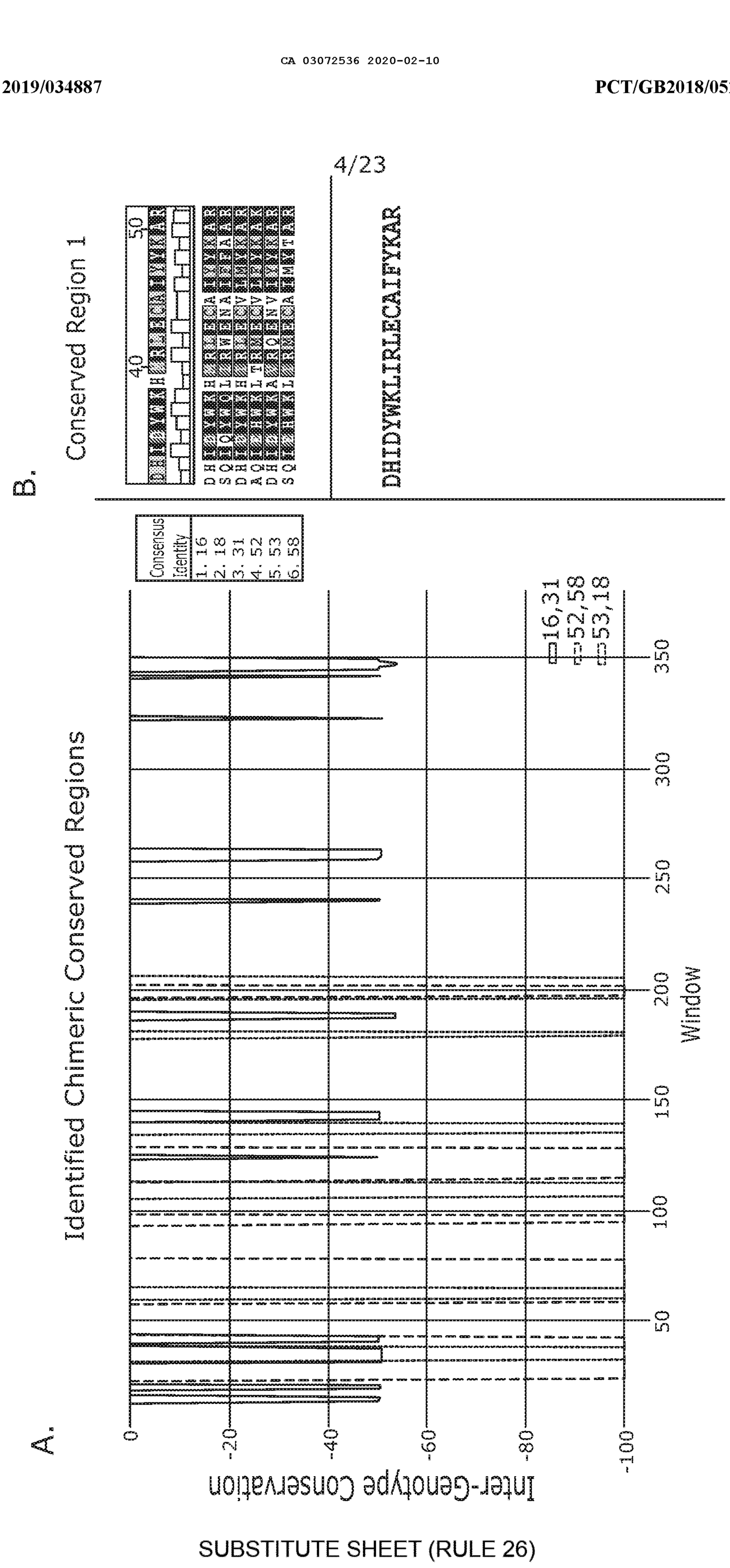

A.
Identified Chimeric Conserved Regions
Inter-Genotype Conservation
Window
0
-20
-40
-60
-80
-100
50
100
150
200
250
300
350
16,31
52,58
53,18
B.
Conserved Region 1
40
50
DHIDYWKLTRLECAIFYKAR
Consensus Identity
1. 16
2. 18
3. 31
4. 52
5. 53
6. 58

Figure 3

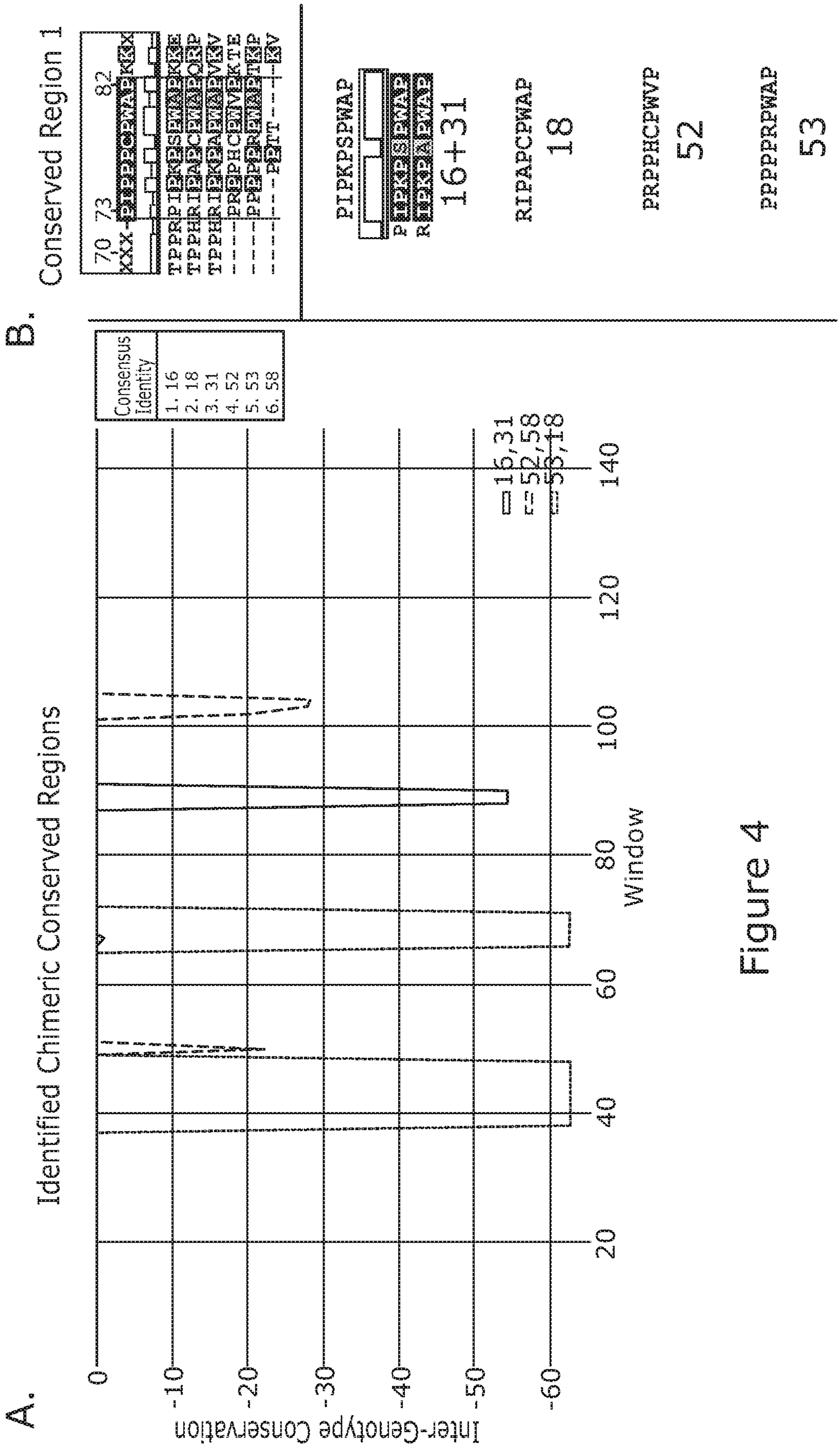

B.
Conserved Region 1
70 73 82
XXX+PPRPCPWAPKKX
TPPRPIEKESEMAEKKE
TPPHRIRAPCRMAPQRP
TPPHRIEKPAPMAPVKV
PRPPHCRMVPKTE
PPPPRPMAPTKP
PPT--KV
PIPKPSPWAP
PHPKPSPWAP
RHPKPRPWAP
16+31
RIPAPCPWAP
18
PRPPHCPWP
52
PPPPPRPWAP
53
Consensus Identity
1. 16
2. 18
3. 31
4. 52
5. 53
6. 58
=16,31
=52,58
=53,18
140
120
100
80
60
40
20
Window
Identified Chimeric Conserved Regions
A.
0
-10
-20
-30
-40
-50
-60
Inter-Genotype Conservation

Figure 4

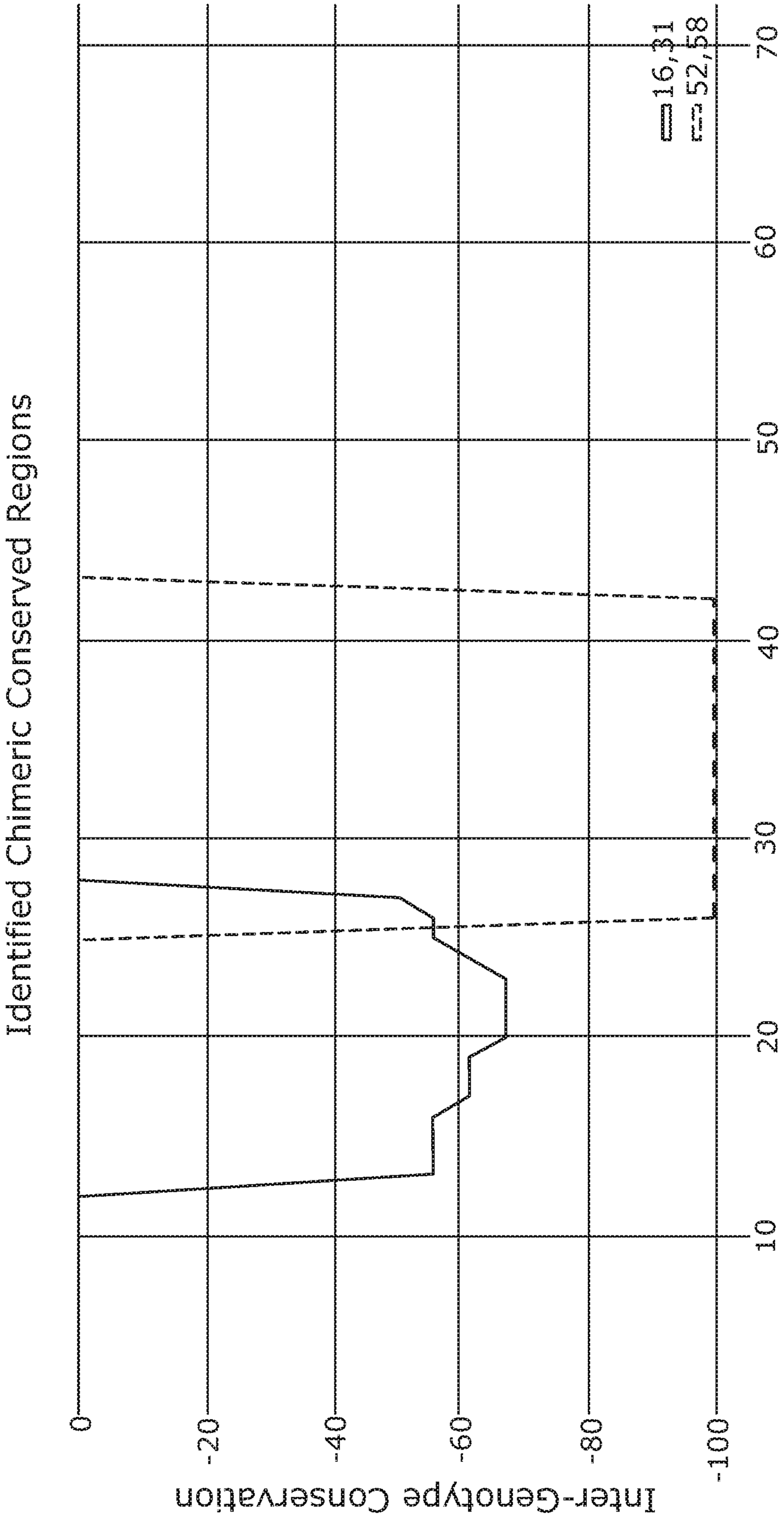

Identified Chimeric Conserved Regions
Inter-Genotype Conservation
Window
0
-20
-40
-60
-80
-100
10
20
30
40
50
60
70
16,31
52,58

Figure 5

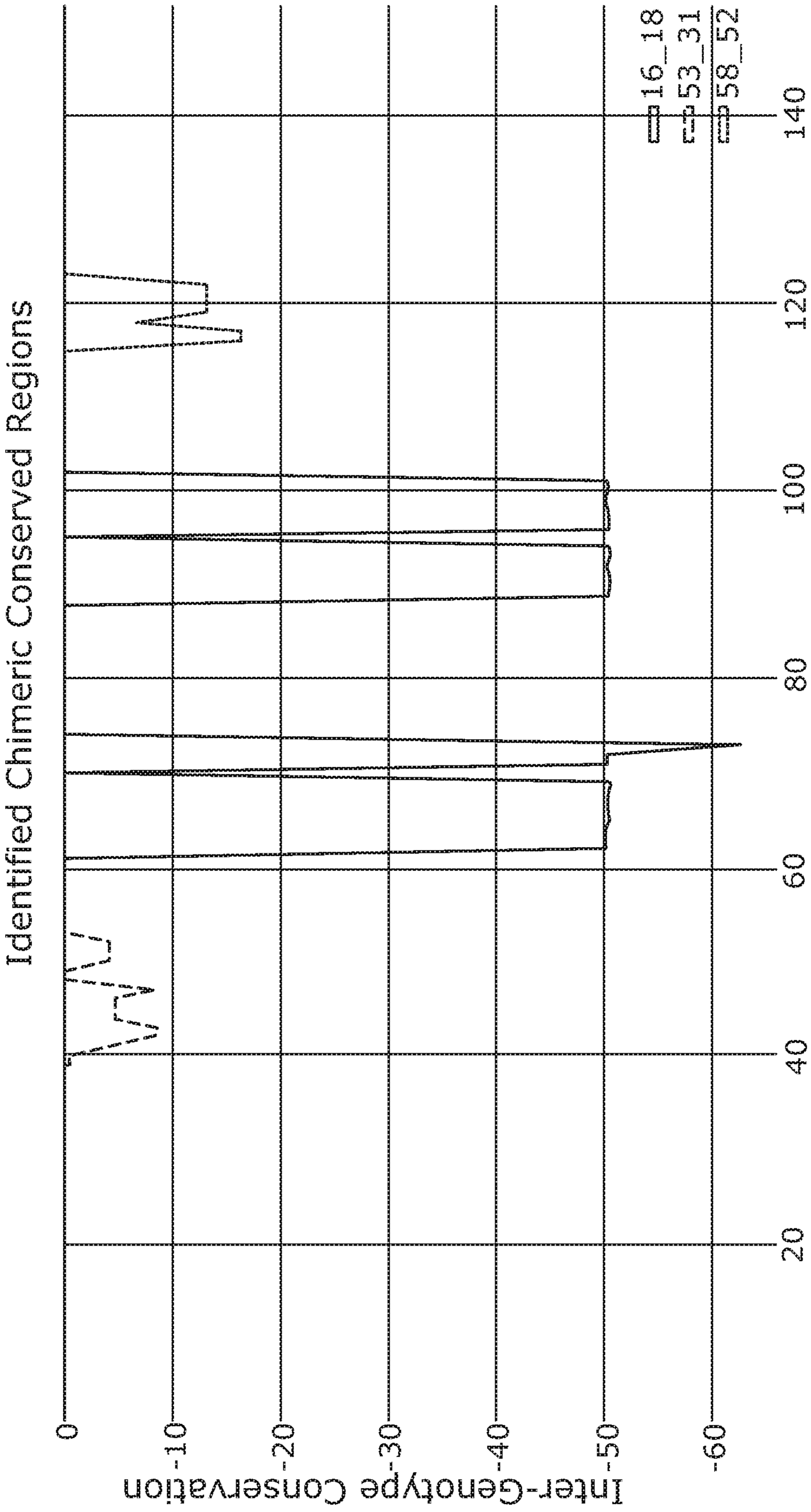

16_18
53_31
58_52
140
120
100
80
60
40
20
Window Low
Identified Chimeric Conserved Regions
0
-10
-20
-30
-40
-50
-60
Inter-Genotype Conservation

Figure 6

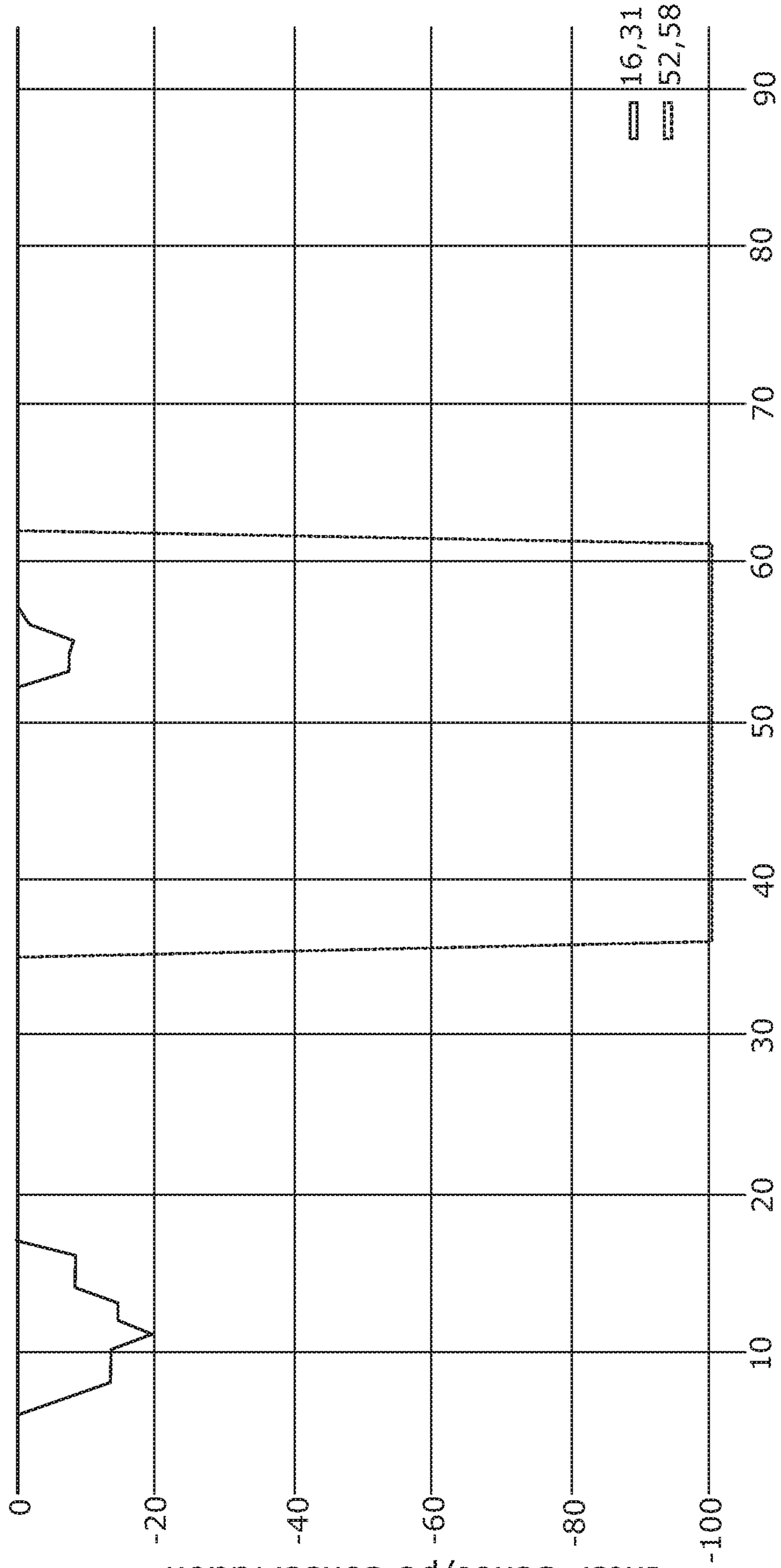

16,31
52,58
90
80
70
60
50
40
30
20
10
0
-20
-40
-60
-80
-100
Identified Chimeric Conserved Regions
Window
Inter-Genotype Conservation

Figure 7

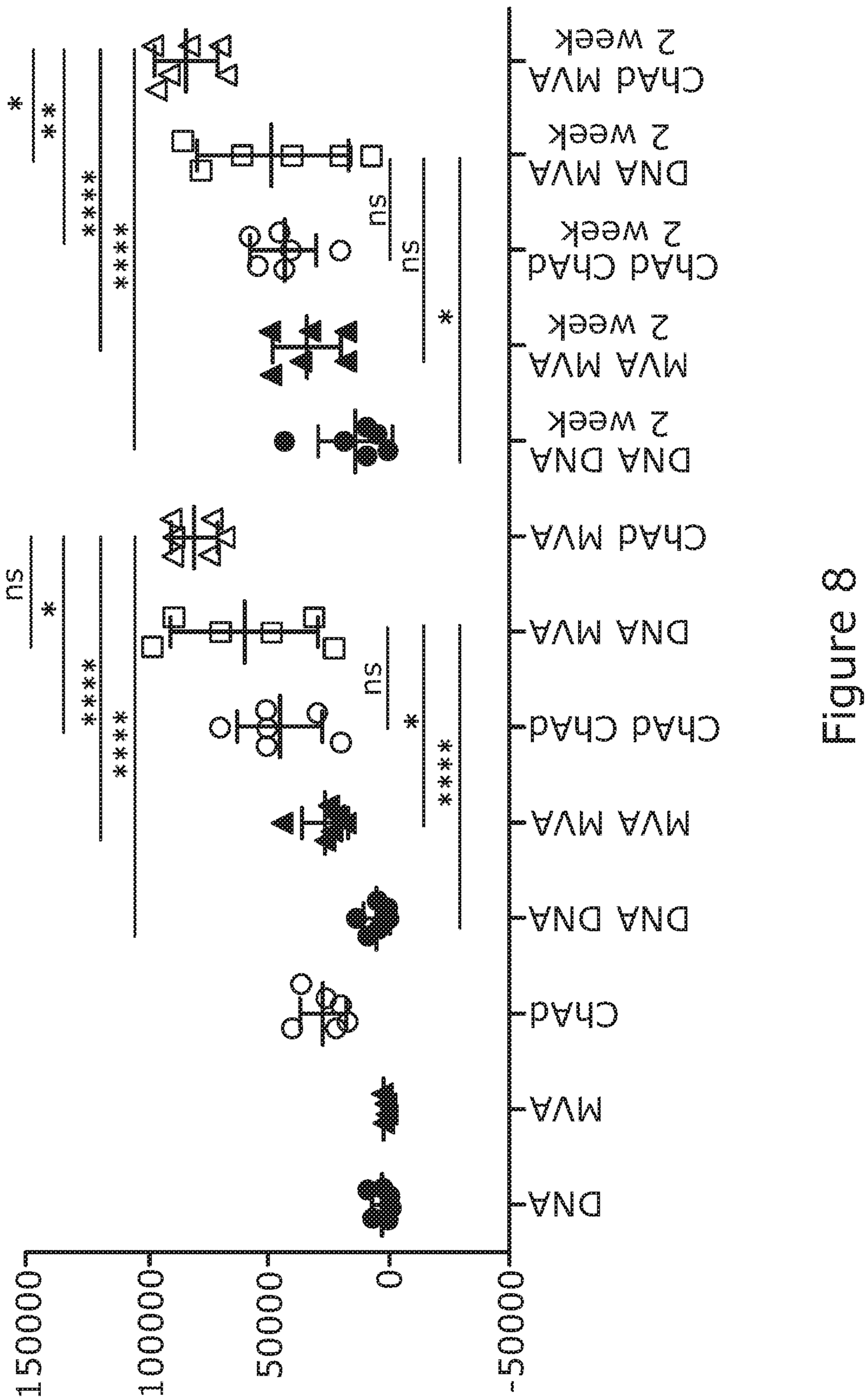

SFU/million PBMCs
150000
100000
50000
0
-50000
DNA
MVA
DNA MVA
Chad
MVA Chad
DNA Chad
MVA MVA
DNA MVA
Chad MVA
DNA DNA
MVA
Chad
DNA MVA
Chad Chad
MVA MVA
DNA MVA
Chad MVA
DNA DNA
MVA
Chad
DNA MVA
Chad Chad
MVA MVA
DNA MVA
Chad MVA
DNA DNA
2 week
2 week
2 week
2 week
2 week
2 week
2 week
2 week
ns
*
ns
*
****
****
**
*

Figure 8

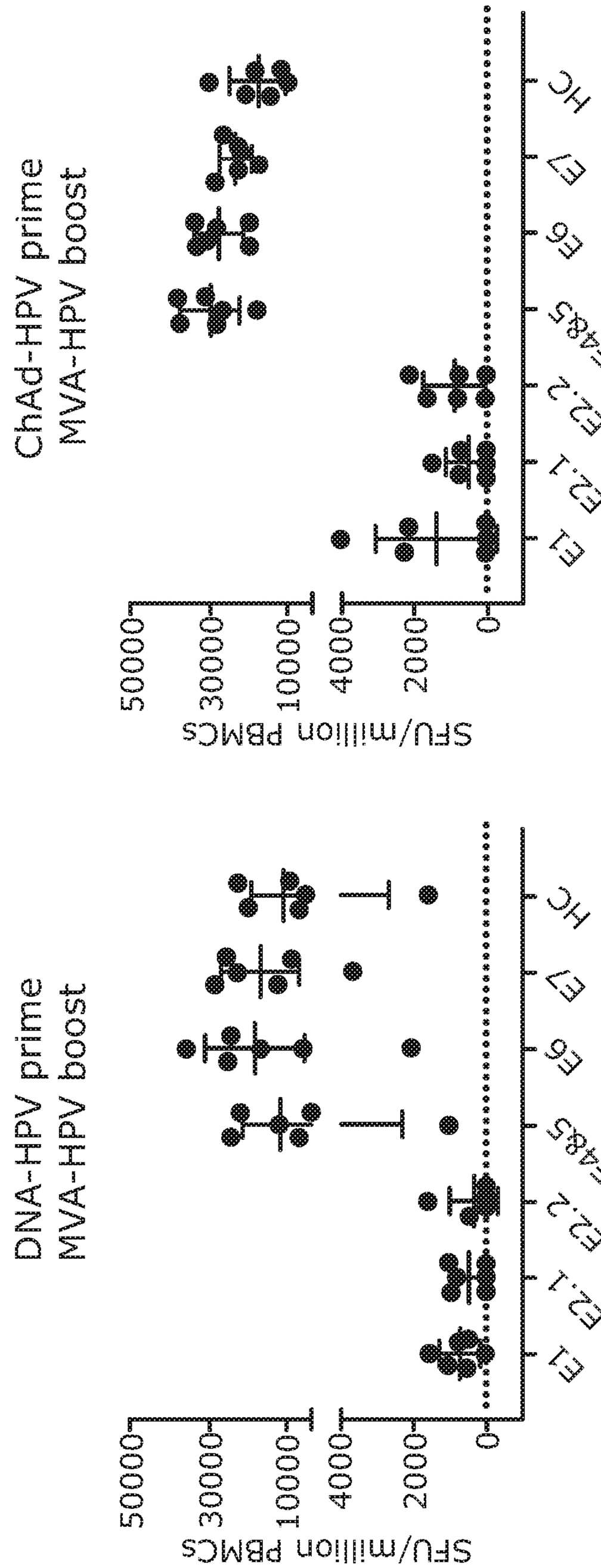

ChAd-HPV prime
MVA-HPV boost
SFU/million PBMCs
50000
30000
10000
4000
2000
0
HC
E7
E6
E4&5
E2.2
E2.1
E1
DNA-HPV prime
MVA-HPV boost
SFU/million PBMCs
50000
30000
10000
4000
2000
0
HC
E7
E6
E4&5
E2.2
E2.1
E1

Fig 9

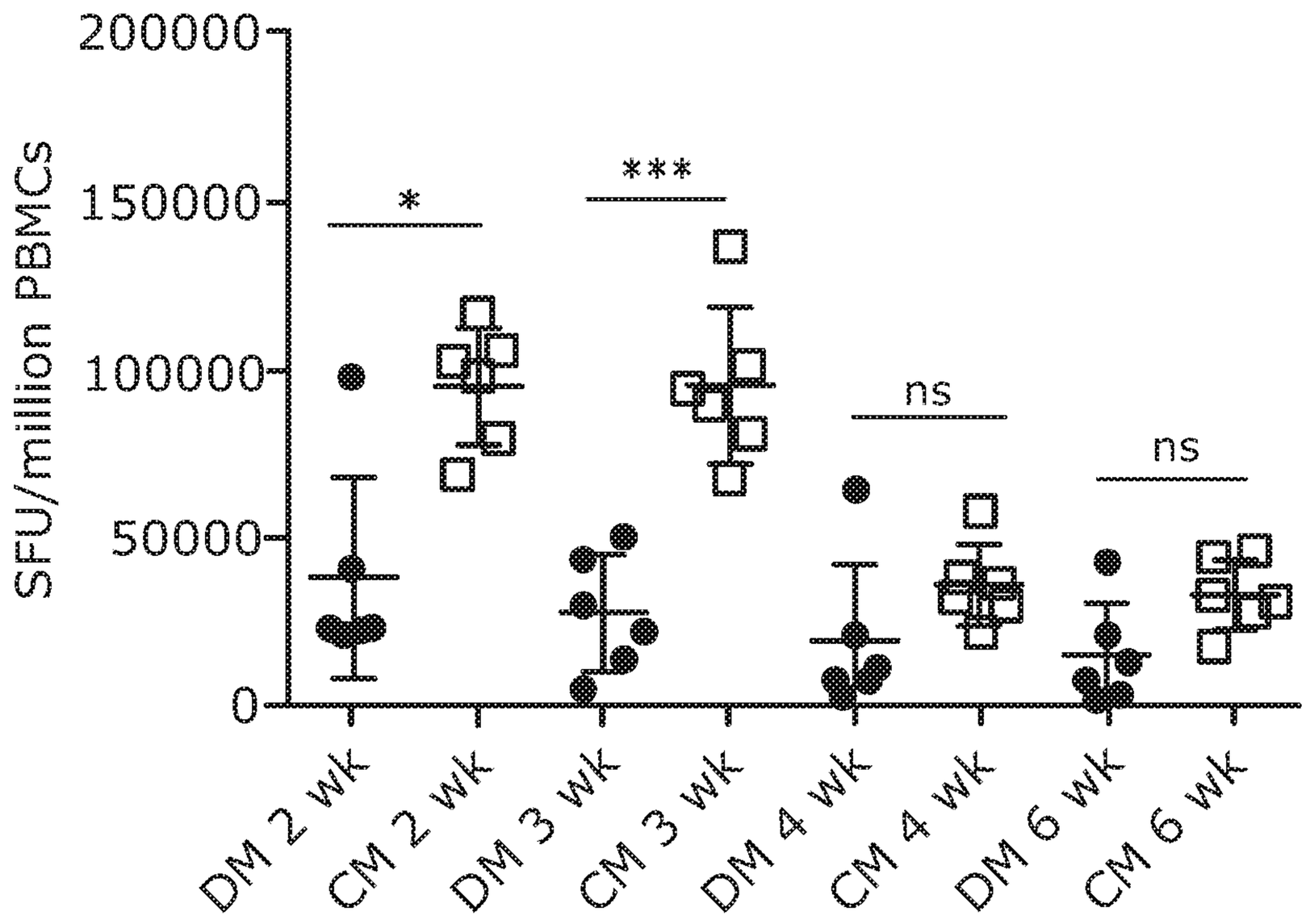

200000
150000
100000
50000
0
SFU/million PBMCs
*
***
ns
ns
DM 2 wk
CM 2 wk
DM 3 wk
CM 3 wk
DM 4 wk
CM 4 wk
DM 6 wk
CM 6 wk

Figure 10A

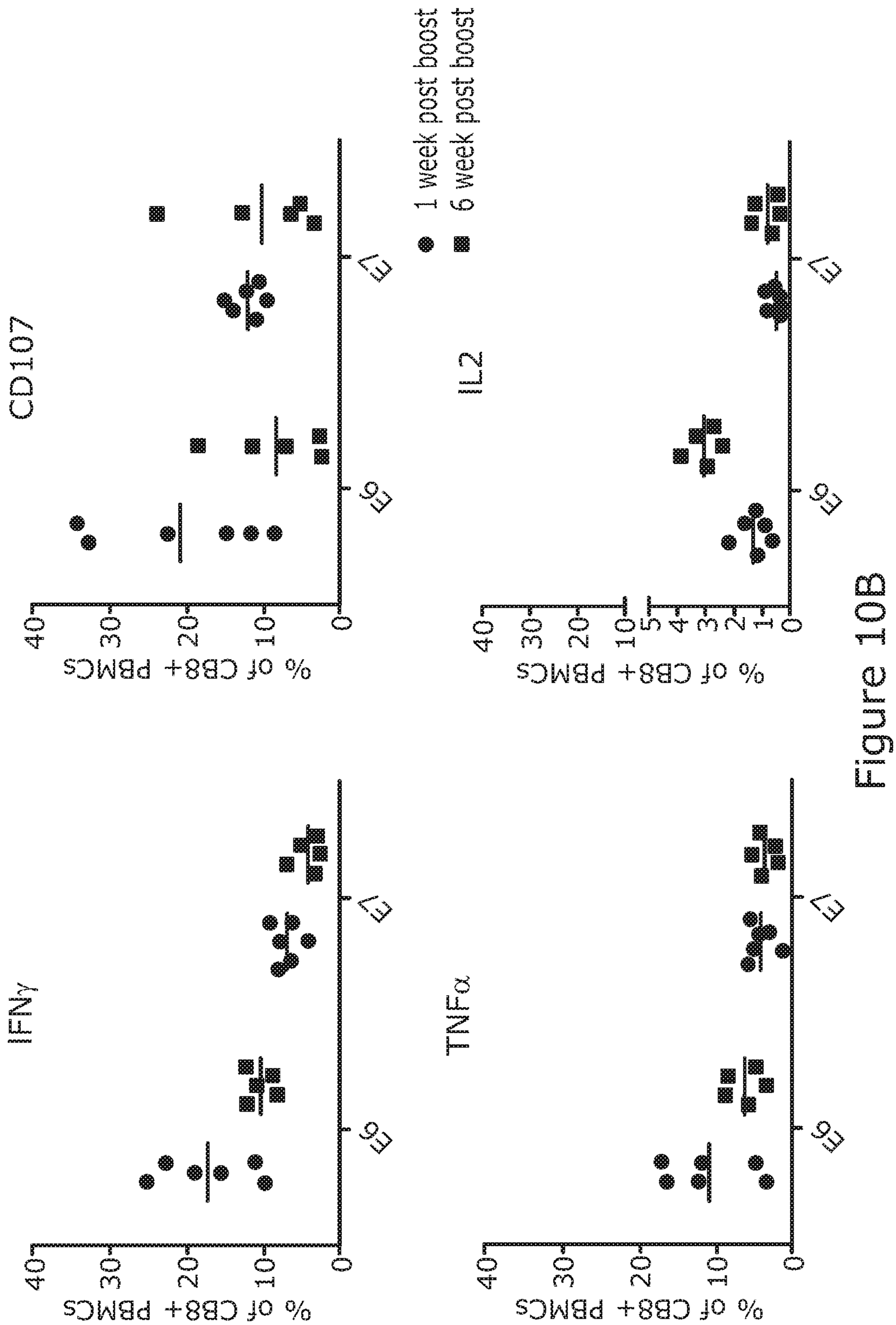

1 week post boost
6 weeks post boost
CD107
IL2
IFNγ
TNFα
% of CB8+ PBMCs
E9
E1
0
10
20
30
40
1
2
3
4
5

Figure 10B

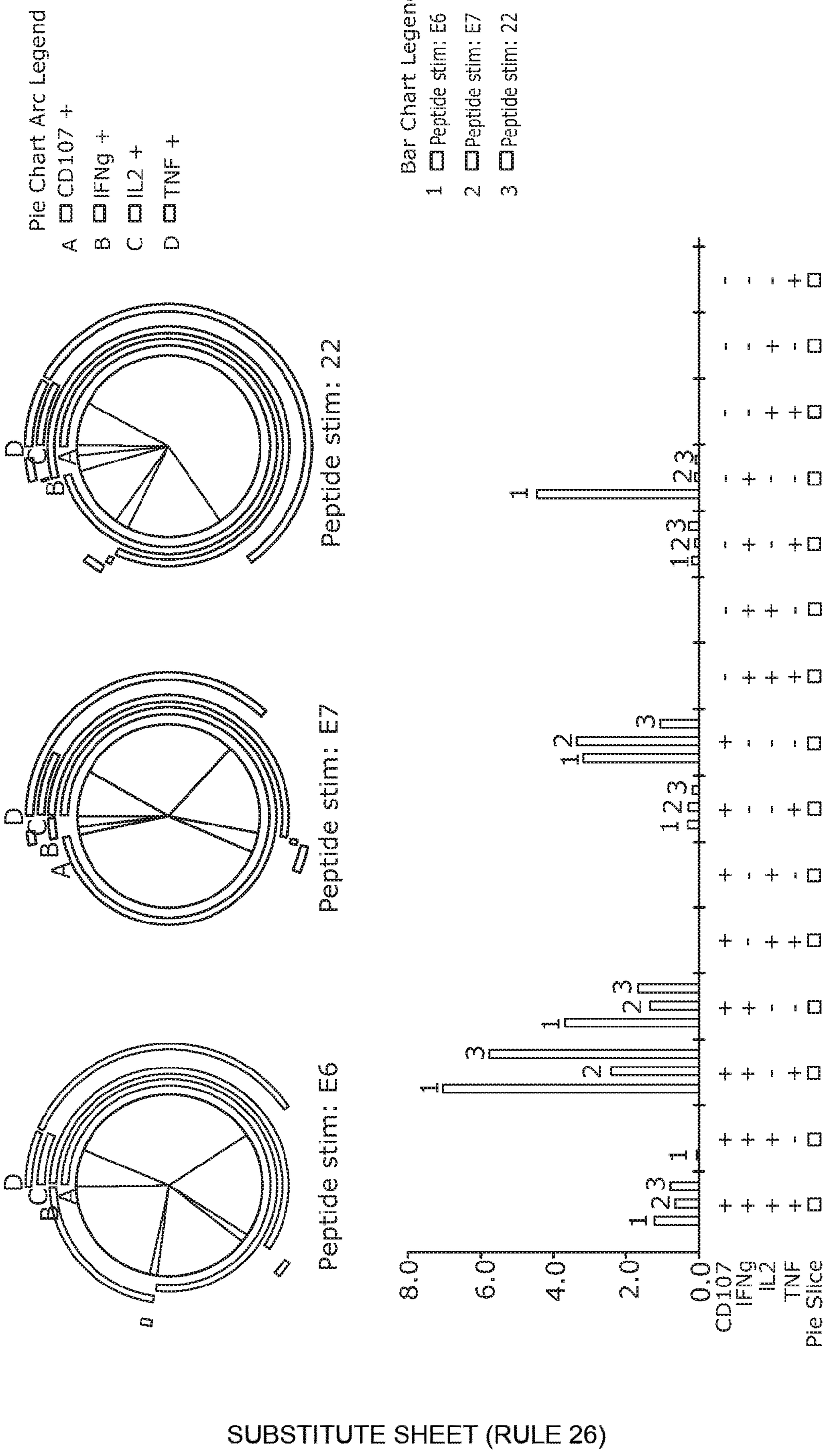

Pie Chart Arc Legend
A CD107 +
B IFNg +
C IL2 +
D TNF +
Peptide stim: 22
Peptide stim: E7
Peptide stim: E6
Bar Chart Legend
1 Peptide stim: E6
2 Peptide stim: E7
3 Peptide stim: 22
8.0
6.0
4.0
2.0
0.0
CD107
IFNg
IL2
TNF
Pie Slice

Figure 11

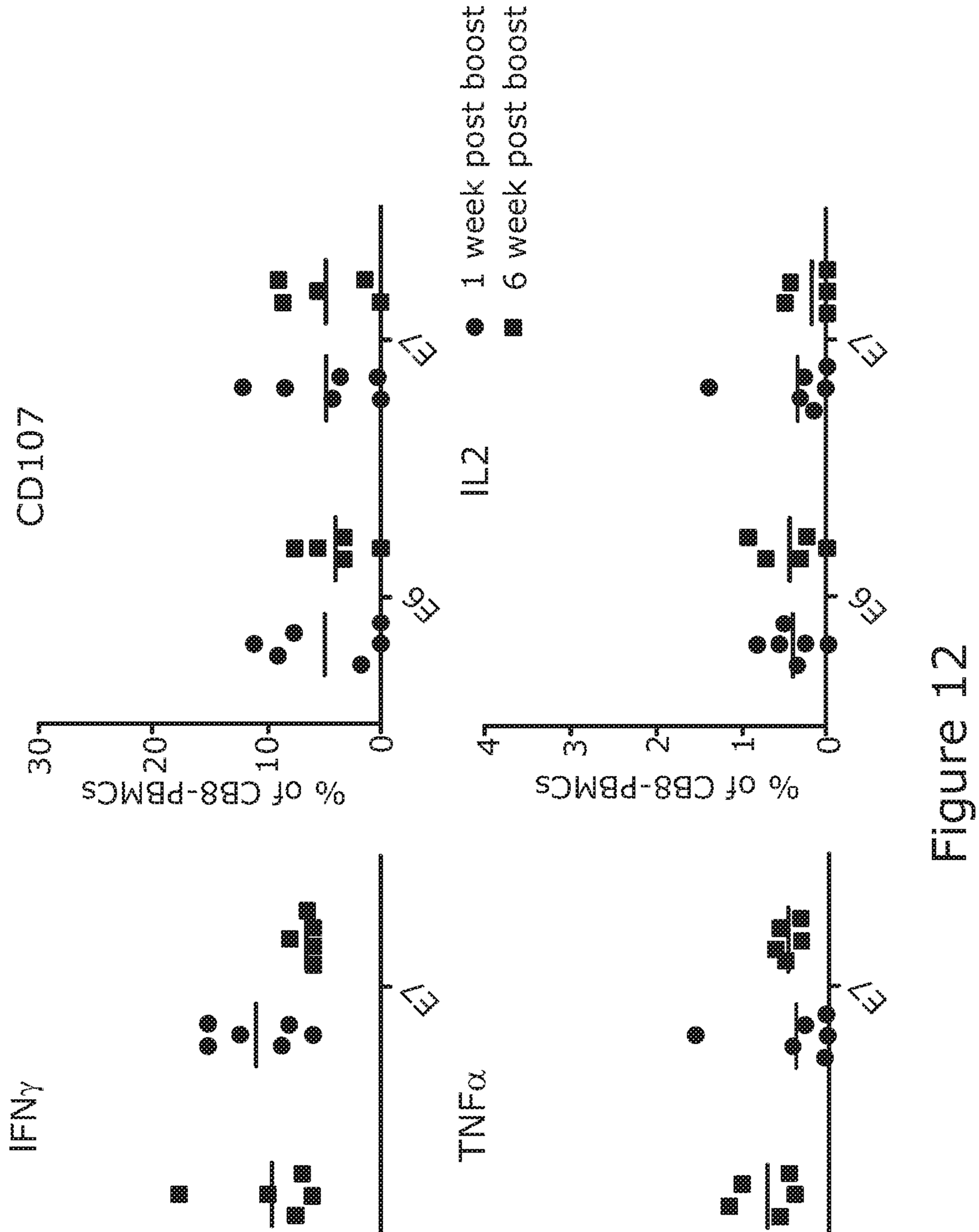

1 week post boost
6 week post boost
CD107
IL2
IFNγ
TNFα
% of CB8-PBMCs
E:T
0
10
20
30
1
2
3
4
6

Figure 12

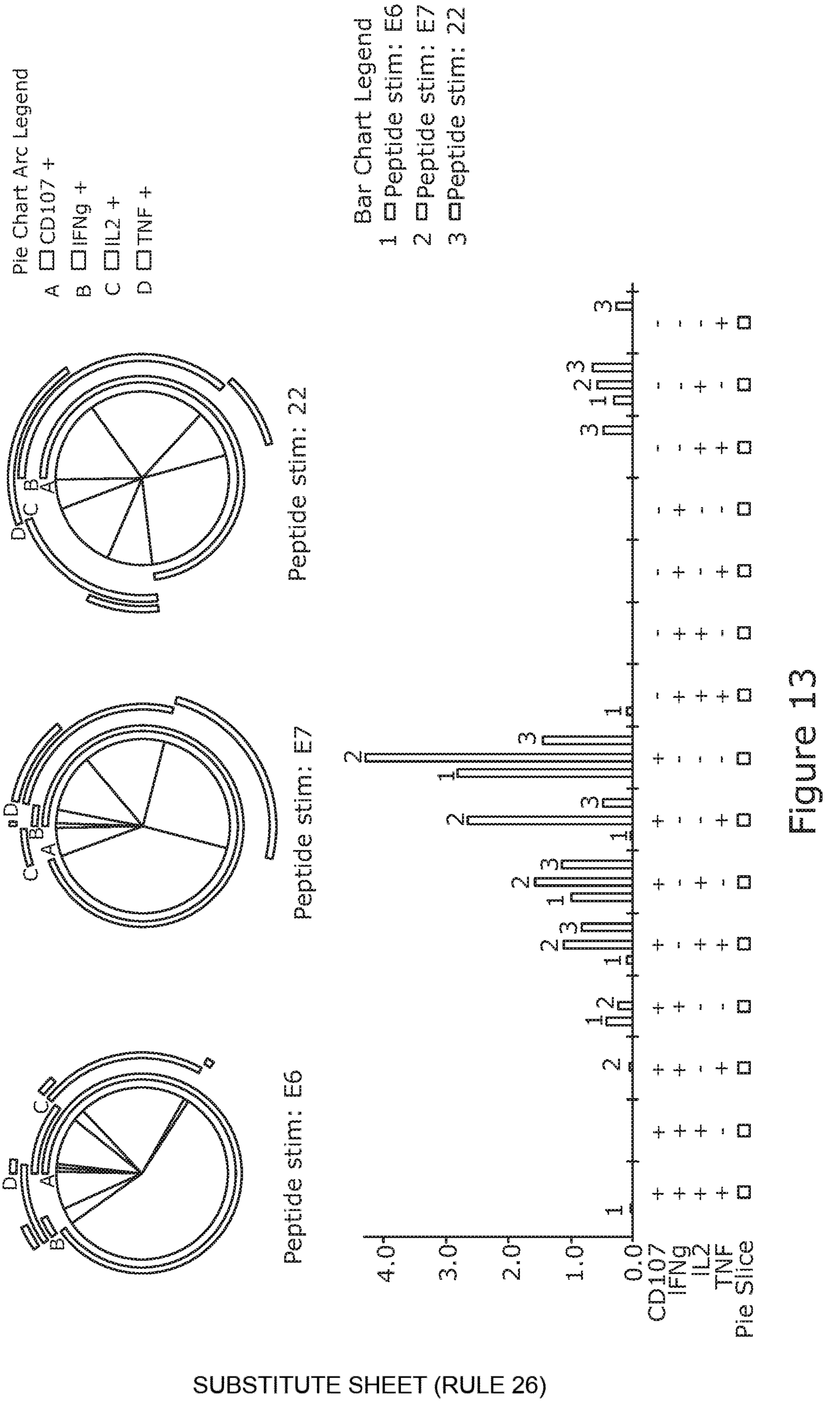

Pie Chart Arc Legend
A CD107 +
B IFNg +
C IL2 +
D TNF +
Bar Chart Legend
1 Peptide stim: E6
2 Peptide stim: E7
3 Peptide stim: 22
Peptide stim: 22
Peptide stim: E7
Peptide stim: E6
4.0
3.0
2.0
1.0
0.0
CD107
IFNg
IL2
TNF
Pie Slice

Figure 13

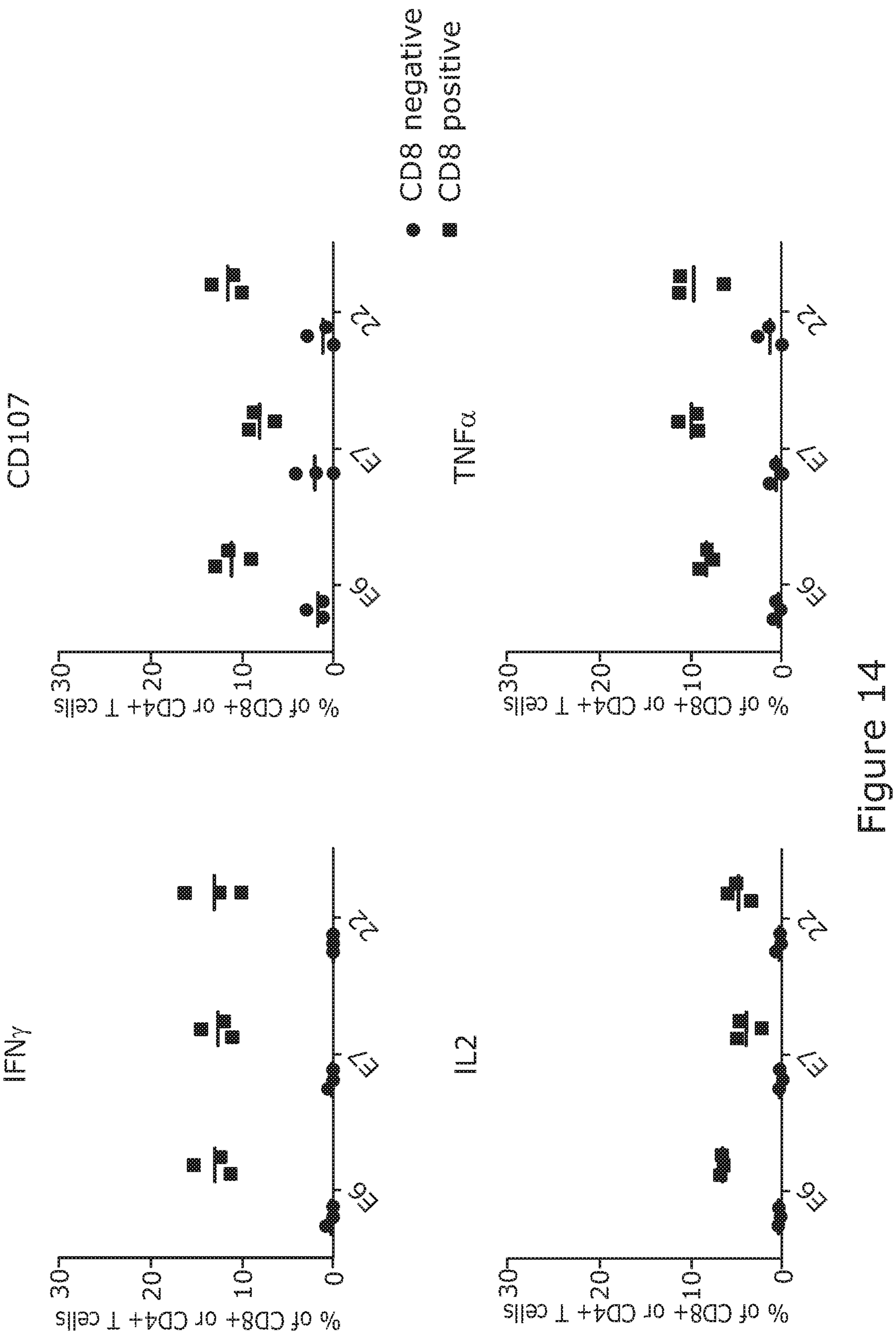

CD8 negative
CD8 positive
CD107
TNFα
IFNγ
IL2
% of CD8+ or CD4+ T cells
0
10
20
30
E1
E6
22

Figure 14

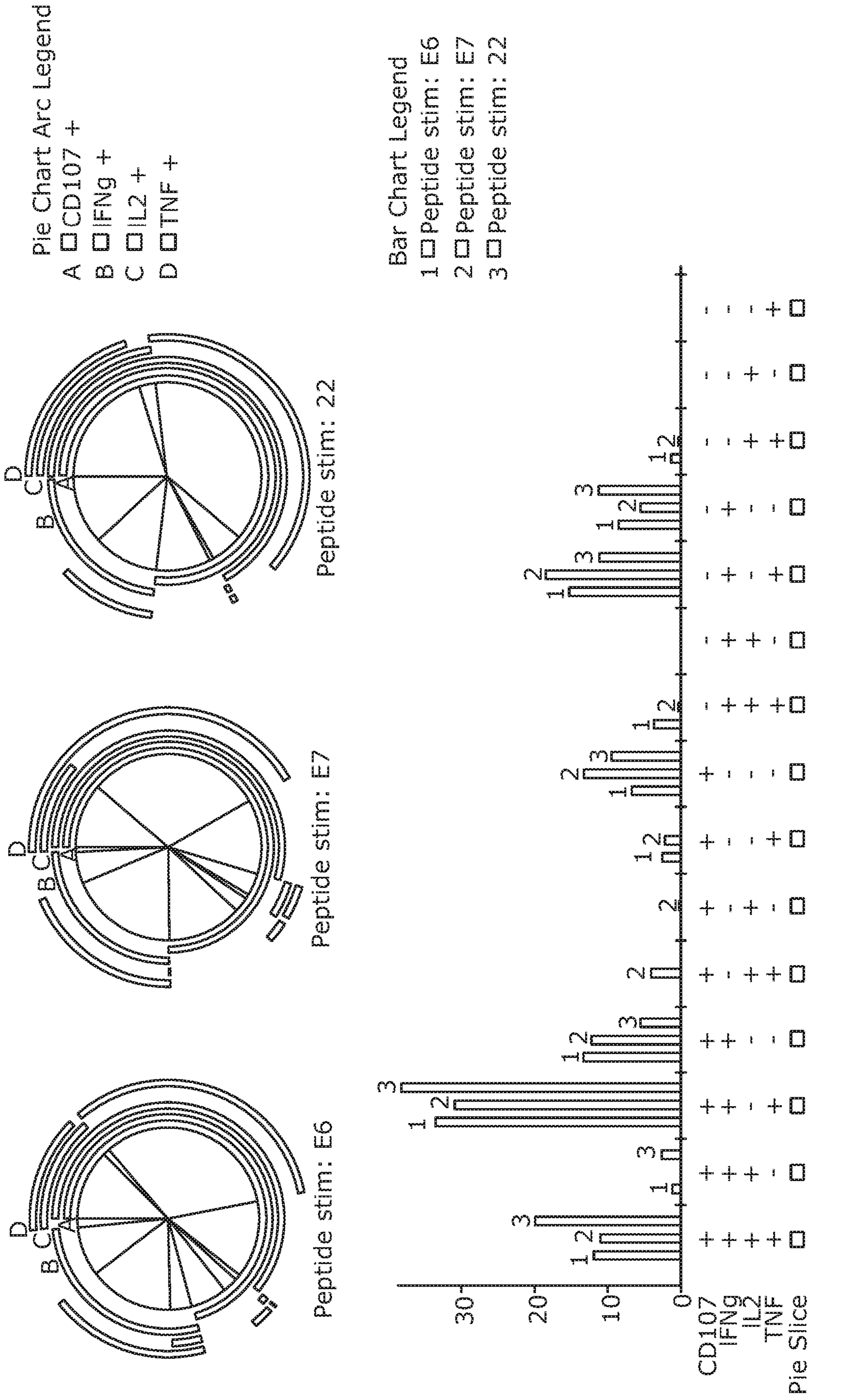

Pie Chart Arc Legend
A CD107 +
B IFNg +
C IL2 +
D TNF +
Bar Chart Legend
1 Peptide stim: E6
2 Peptide stim: E7
3 Peptide stim: 22
Peptide stim: 22
Peptide stim: E7
Peptide stim: E6
n
30
20
10
0
CD107
IFNg
IL2
TNF
Pie Slice

Figure 15

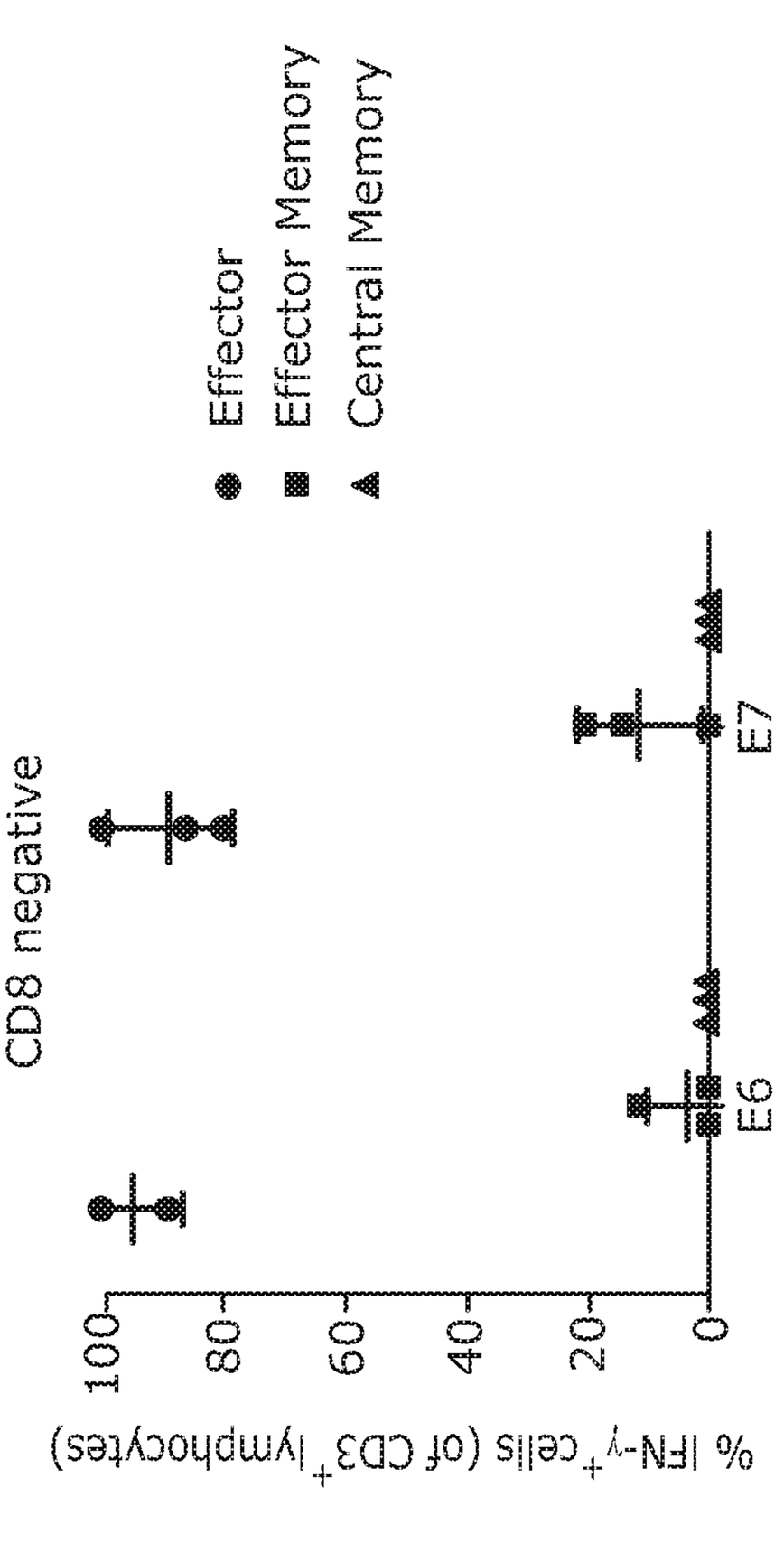


Figure 16

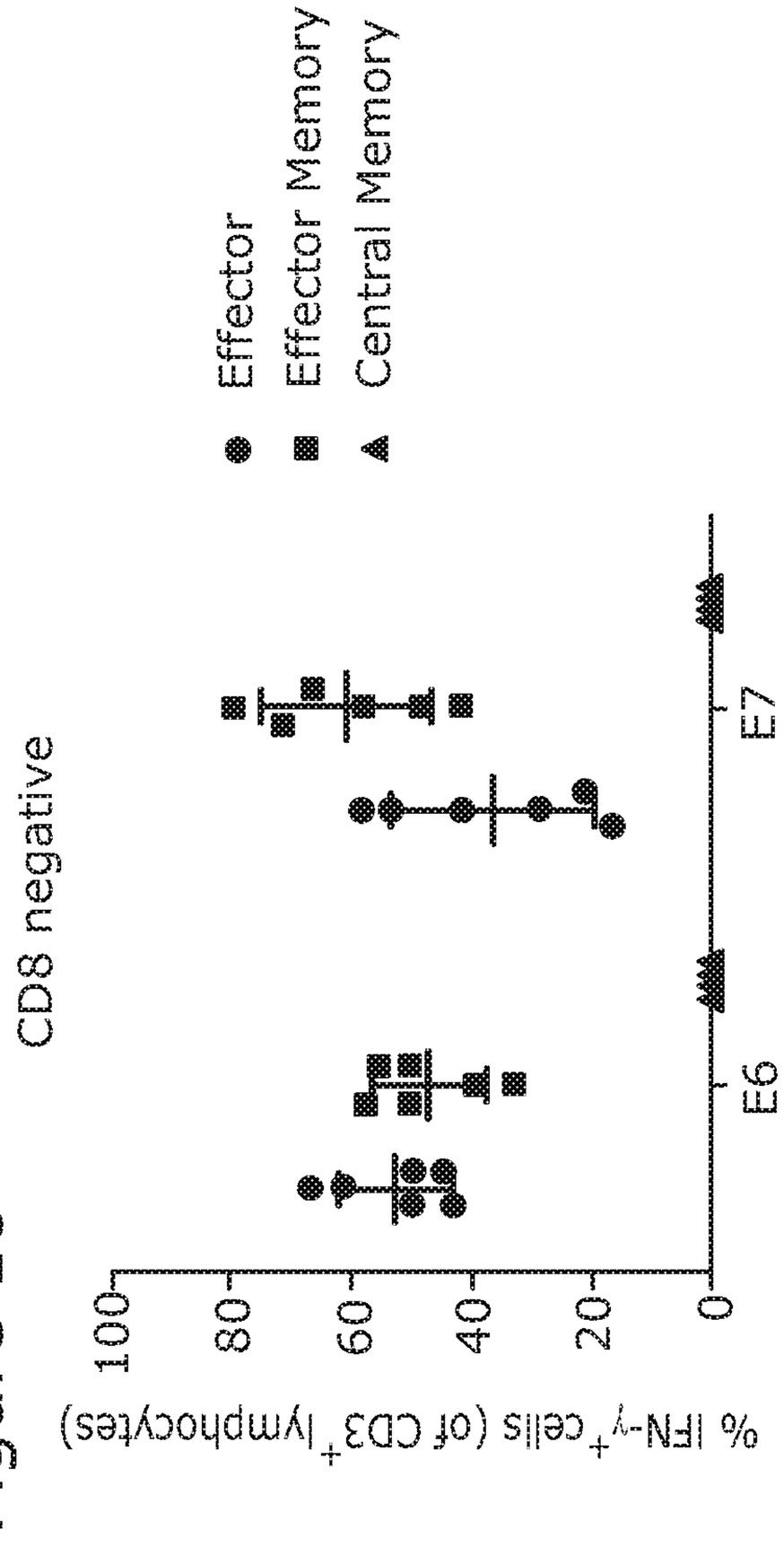
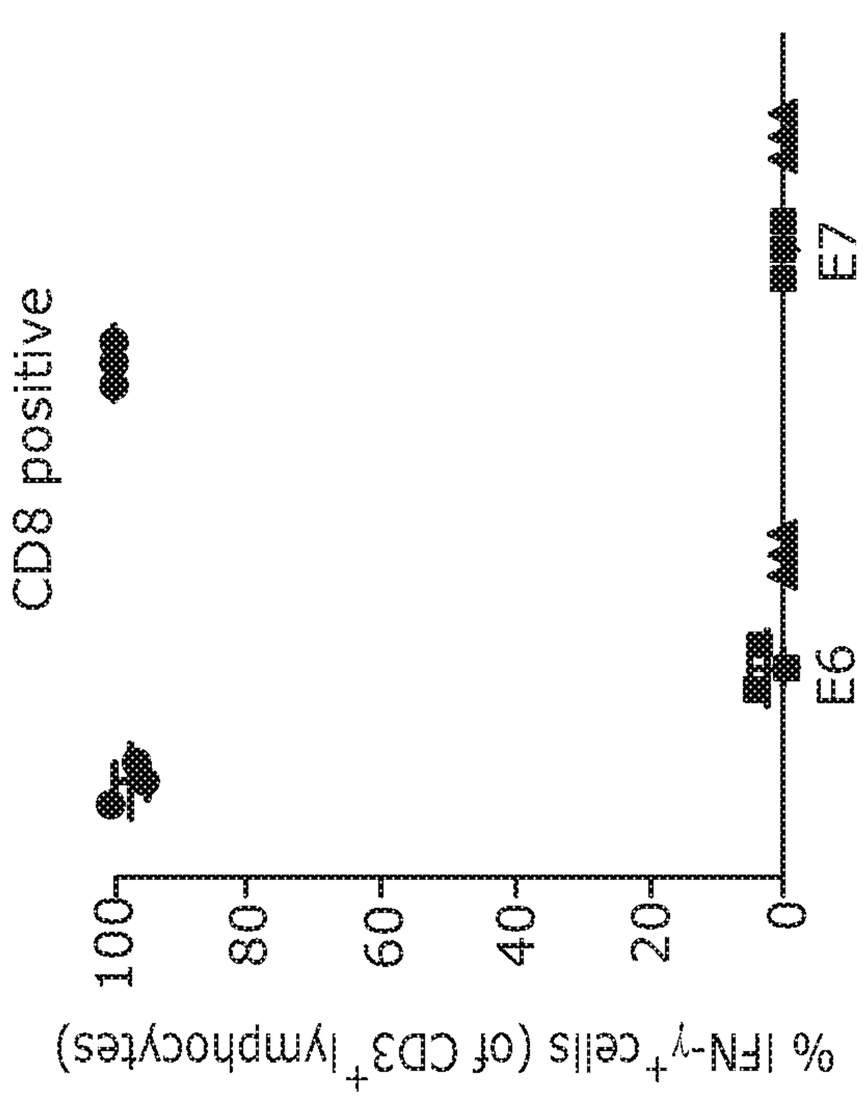
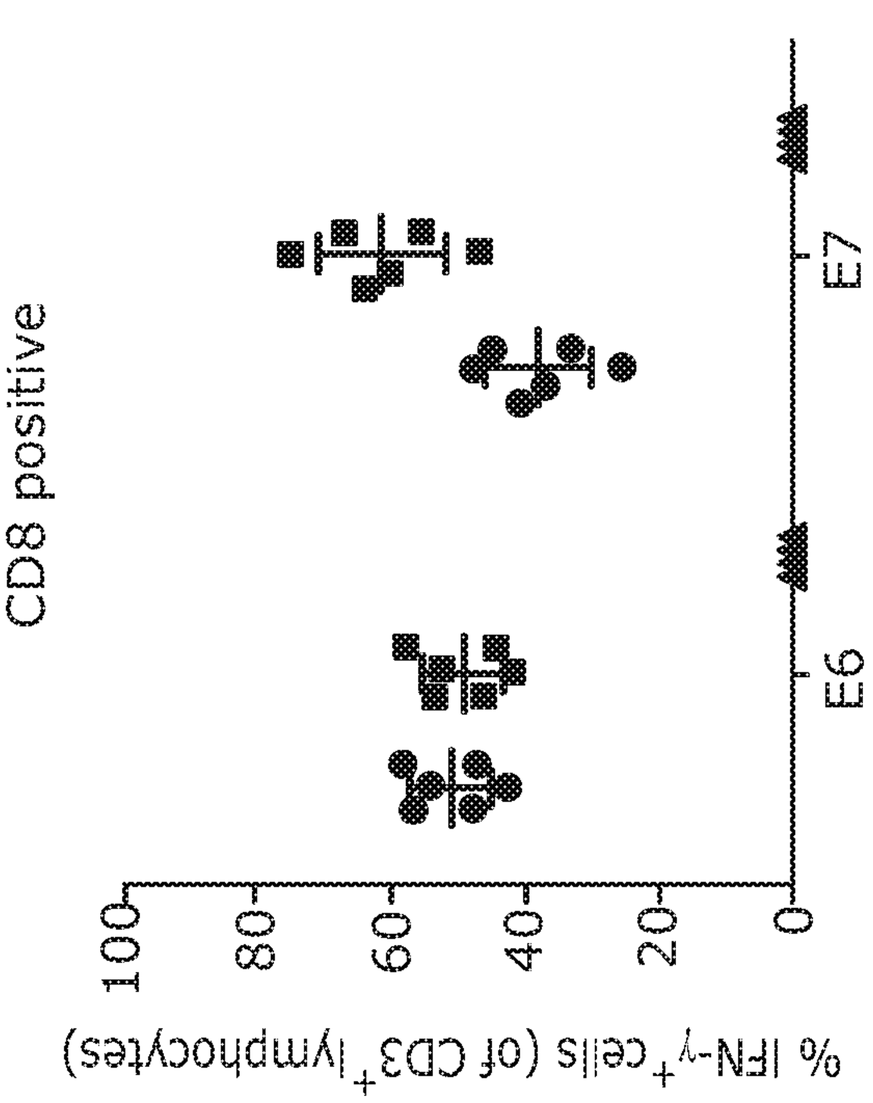


Figure 17





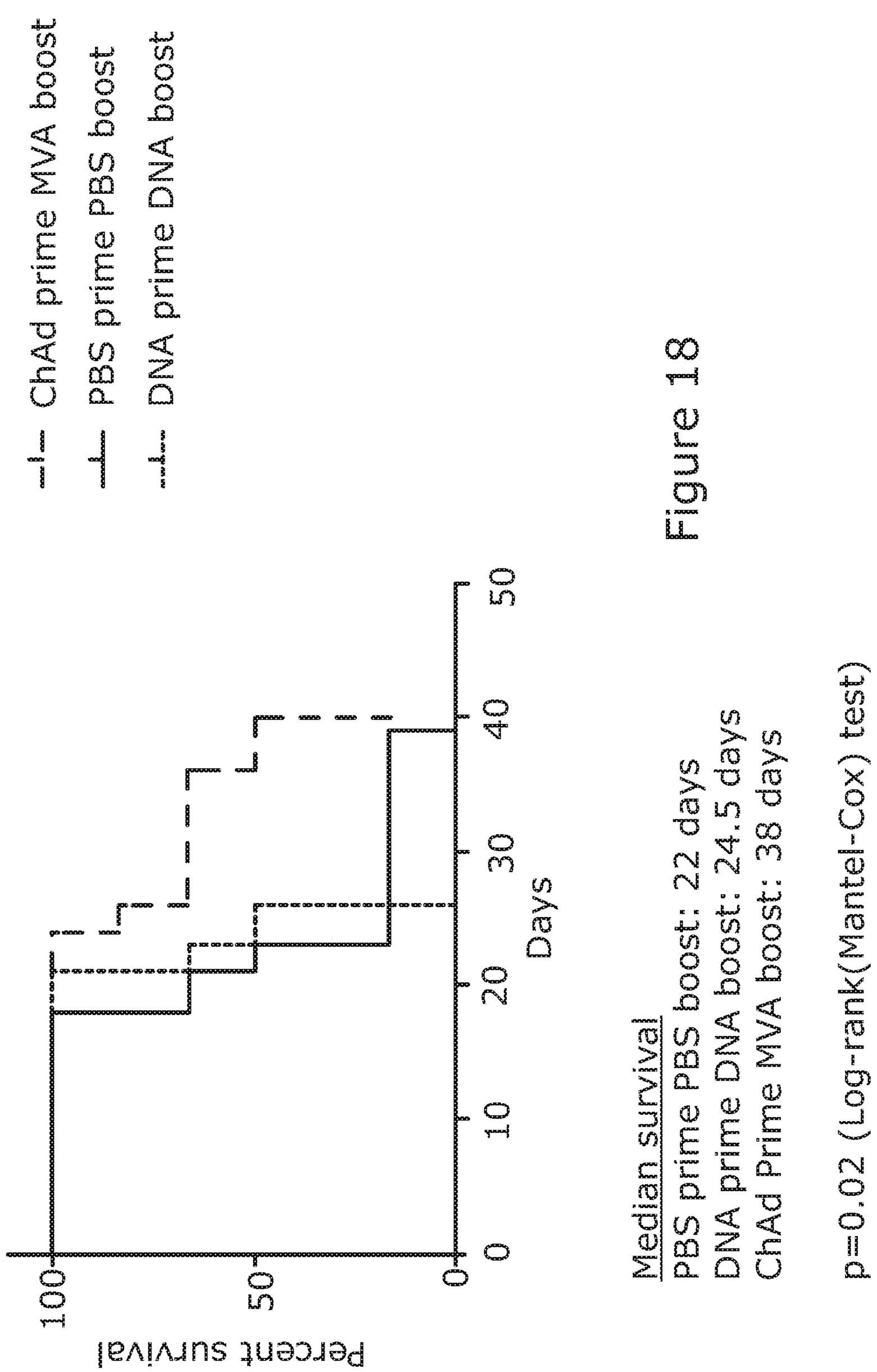

ChAd prime MVA boost
PBS prime PBS boost
DNA prime DNA boost
Percent survival
100
50
0
0
10
20
30
40
50
Days
Median survival
PBS prime PBS boost: 22 days
DNA prime DNA boost: 24.5 days
ChAd Prime MVA boost: 38 days
p=0.02 (Log-rank(Mantel-Cox) test)

Figure 18

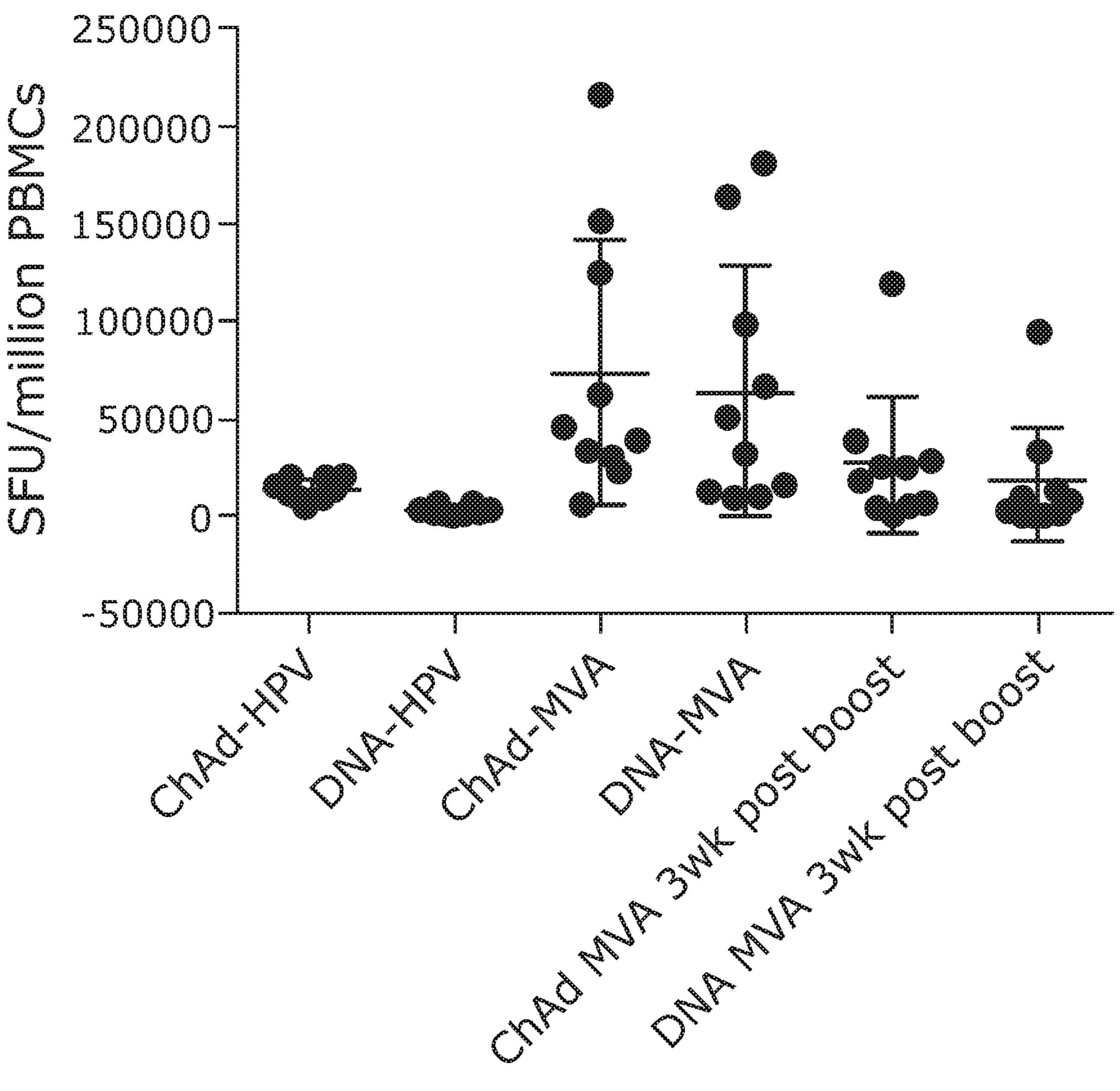

SFU/million PBMCs
250000
200000
150000
100000
50000
0
-50000
ChAd-HPV
DNA-HPV
ChAd-MVA
DNA-MVA
ChAd MVA 3wk post boost
DNA MVA 3wk post boost

Figure 19

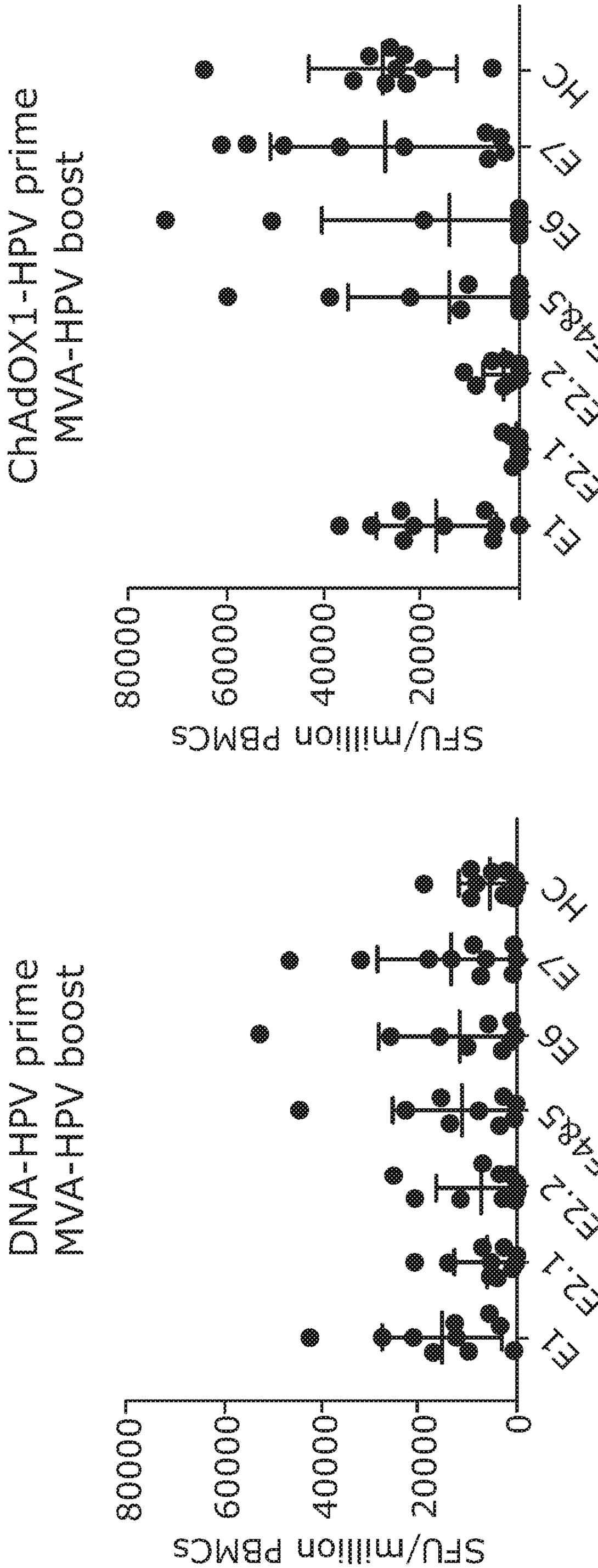
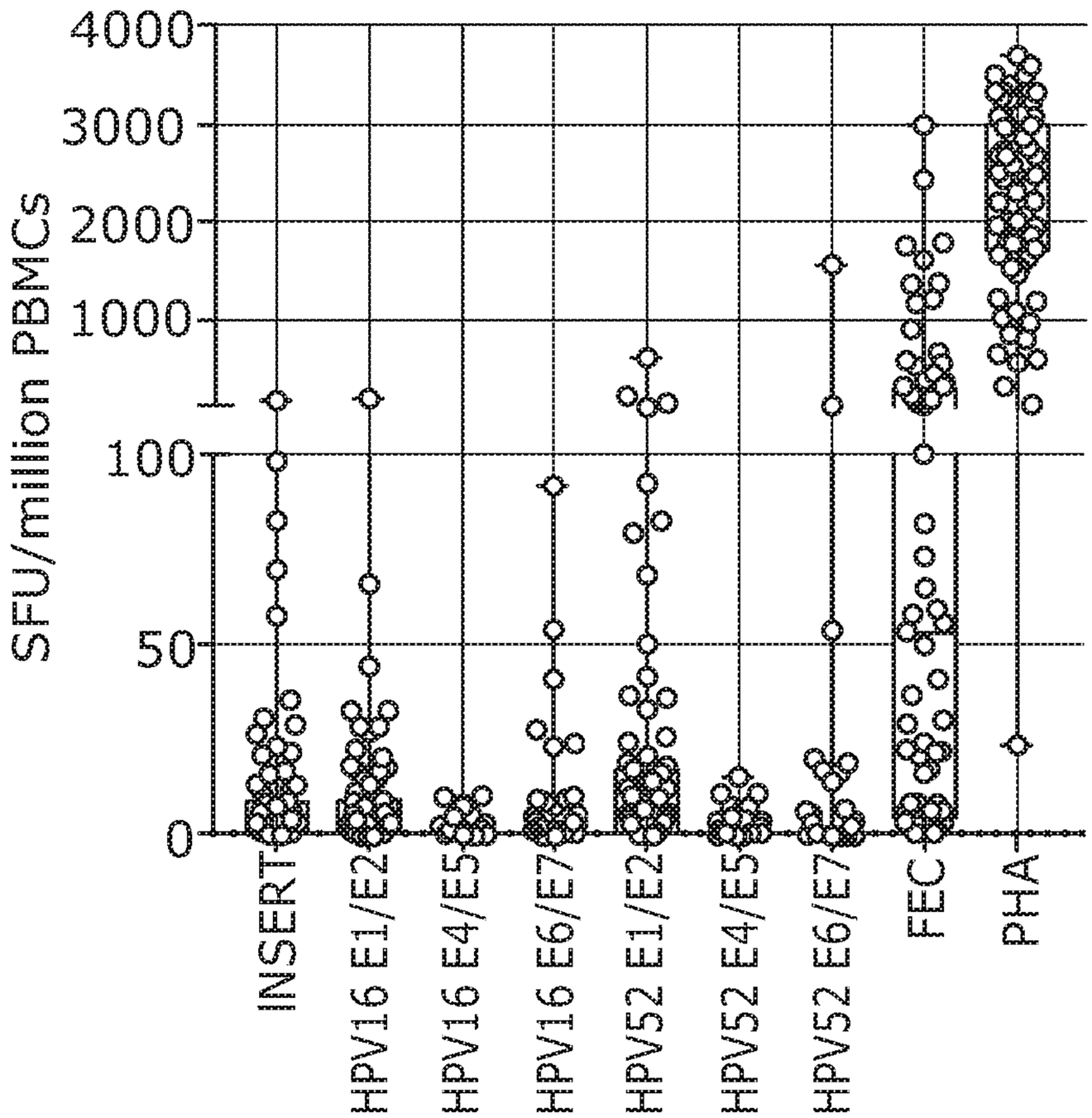
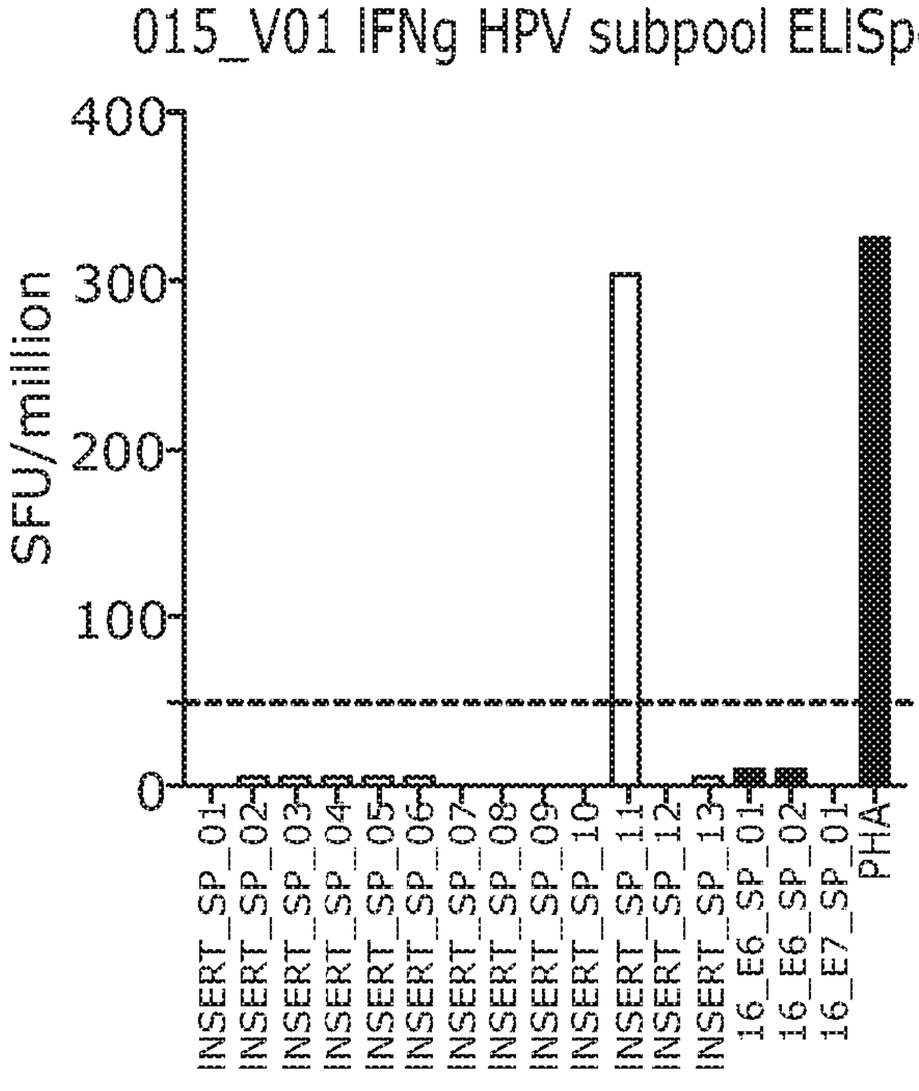

ChAdOX1-HPV prime
MVA-HPV boost
SFU/million PBMCs
80000
60000
40000
20000
0
E
E2.1
E2.2
EA85
E6
E7
HC
DNA-HPV prime
MVA-HPV boost
SFU/million PBMCs
80000
60000
40000
20000
0
E
E2.1
E2.2
EA85
E6
E7
HC

Figure 20


4000
3000
2000
1000
100
50
0
SFU/million PBMCs
INSERT
HPV16 E1/E2
HPV16 E4/E5
HPV16 E6/E7
HPV52 E1/E2
HPV52 E4/E5
HPV52 E6/E7
FEC
PHA


015_V01 IFNg HPV subpool ELISpot
400
300
200
100
0
SFU/million
INSERT_SP_01
INSERT_SP_02
INSERT_SP_03
INSERT_SP_04
INSERT_SP_05
INSERT_SP_06
INSERT_SP_07
INSERT_SP_08
INSERT_SP_09
INSERT_SP_10
INSERT_SP_11
INSERT_SP_12
INSERT_SP_13
16_E6_SP_01
16_E6_SP_02
16_E7_SP_01
PHA

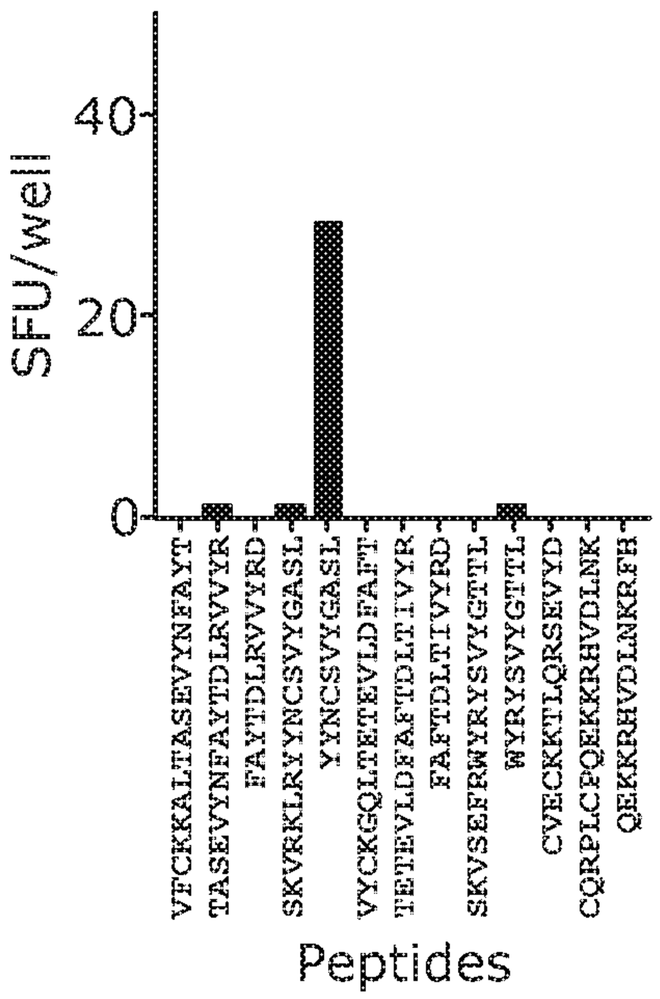

HPV_15_V01_INSERT_SP11
40
20
0
SFU/well
VFCKKALTASEVYNFAYTDLRVVYRD
TASEVYNFAYTDLRVVYRD
FAYTDLRVVYRD
SKVRKLRYYNCSVYGASL
YYNCSVYGASL
VYCKGQLTETEVLDFAFTDLRVVYRD
TETEVLDFAFTDLTIVYRD
FAFTDLTIVYRD
SKVSEFRWYRYSVYGTTL
WYRYSVYGTTL
CVECKKTLQRSEVYLSTTL
CQRPLCPQEKKRHVDLNKRFYVYD
QEKKRHVDLNKRFH
Peptides

Figure 21

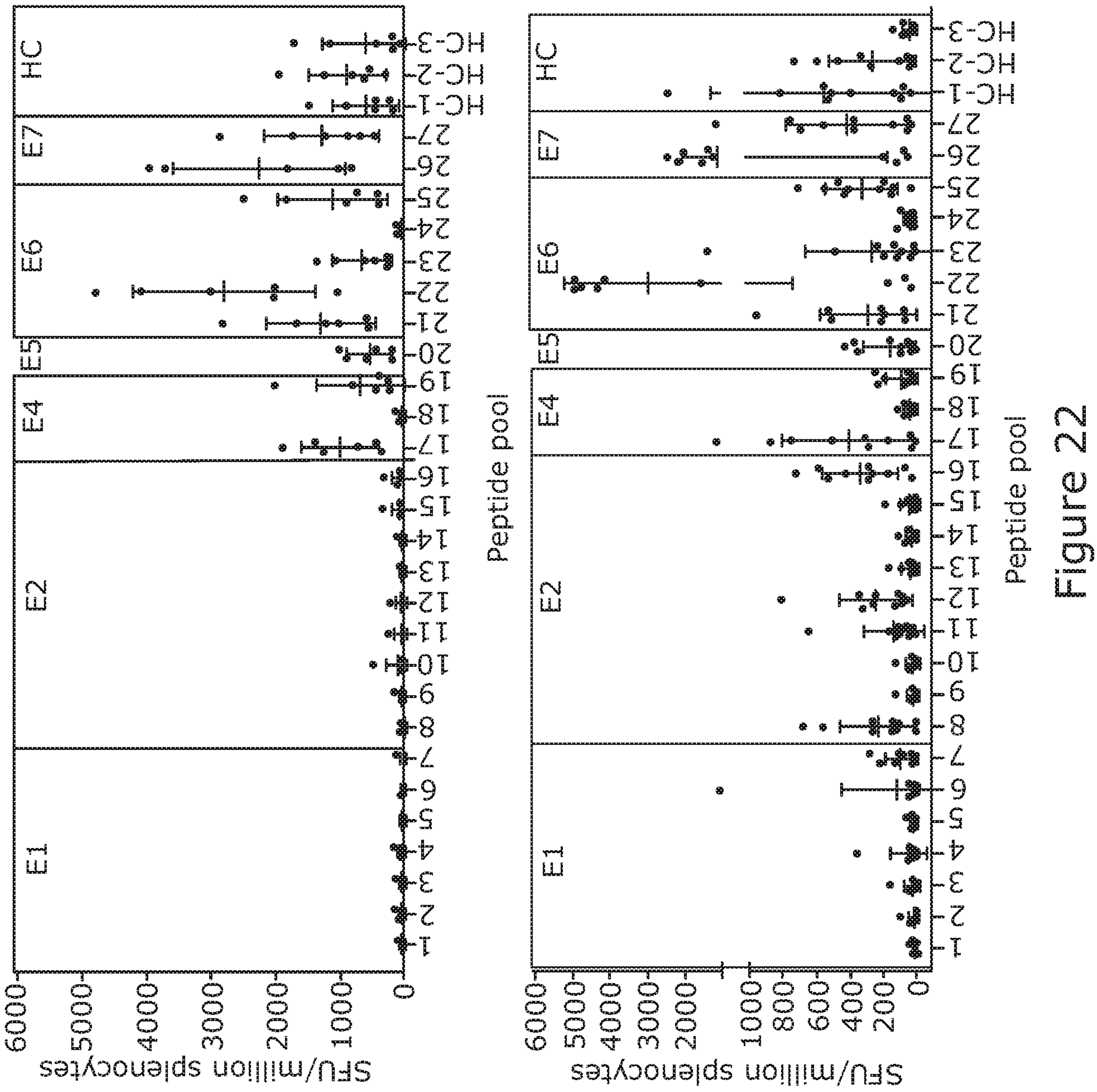

HC
E7
E6
E5
E4
E2
E1
HC-3
HC-2
HC-1
Peptide pool
SFU/million splenocytes

Figure 22

Figure 1

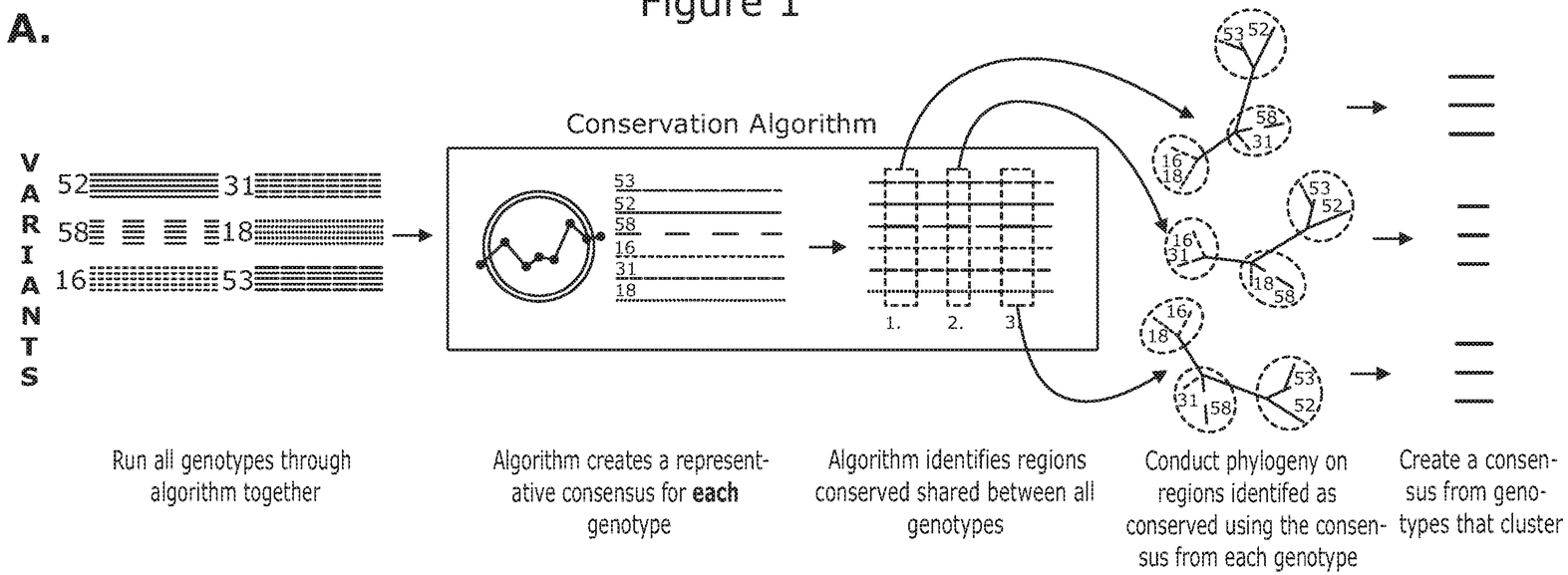


Figure 1 continued