Summary

In one aspect, the present disclosure refers 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.

The disclosure 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 example, 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 example, the polypeptide may not comprise a complete gene sequence. The polypeptide may consist essentially of conserved peptide sequences. In another example, the polypeptide may consist essentially of conserved peptide sequences and a peptide adjuvant sequence. In another example, the polypeptide may consist essentially of conserved peptide sequences and one or more linkers therebetween. In another example, the polypeptide may consist essentially of conserved peptide sequences, a peptide adjuvant sequence and one or more linkers therebetween. In one example, 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 disclosure. 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 example, the plurality of conserved peptide sequences may comprise 15 or more conserved peptide sequences. In another example, the plurality of conserved peptide sequences may comprise 20 or more conserved peptide sequences. In another example, the plurality of conserved peptide sequences may comprise 25 or more conserved peptide sequences. In another example, 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 example, 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 example, the plurality of conserved peptide sequences may comprise 50 or more conserved peptide sequences. In another example, the plurality of conserved peptide sequences may comprise 55 or more conserved peptide sequences. In one example, the plurality of conserved peptide sequences consists of about 56 or more conserved peptide sequences. In one example, the plurality of conserved peptide sequences consists of about 57 or more conserved peptide sequences. In one example, the plurality of conserved peptide sequences consists of about 58 or more conserved peptide sequences. In one example, the plurality of conserved peptide sequences consists of about 59 or more conserved peptide sequences. In one example, 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

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 examples 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 examples 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 example, 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 example, the linker may comprise or consist of G and/or P amino acid residues. In one example, 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 example, the linker residues may be between 2 and 8 residues in length. In another example, 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 examples where there are linker/junction residues between the conserved peptide sequences.

In one example, 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 example, 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 example, 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.

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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, 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 example, the polypeptide may comprise or consist of the sequence of SEQ ID NO: 72, 74 or 76, or variants thereof.

In one example, the polypeptide may consist essentially of conserved peptide sequences and a peptide adjuvant. In one example, 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 example, 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 example, 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 example, a variant conserved peptide sequence may have at least 74% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 75% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 79% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 80% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 82% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 83% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 85% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide

sequence may have at least 88% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 90% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 92% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 95% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 98% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, a variant conserved peptide sequence may have at least 99% sequence identity with any one of SEQ ID NOs: 1 to 59. In another example, 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 example, at least one conserved peptide sequence may be at least about 8 amino acids in length. In another example, at least one conserved peptide sequence may be at least about 10 amino acids in length. In another example, at least one conserved peptide sequence may be at least about 15 amino acids in length. In another example, at least one conserved peptide sequence may be at least about 20 amino acids in length. In another example, at least one conserved peptide sequence may be at least about 30 amino acids in length. In one example, at least one conserved peptide sequence may be between about 20 and about 220 amino acids in length. In one example, at least one conserved peptide sequence may be no more than about 300 amino acids in length. In another example, at least one conserved peptide sequence may be no more than about 250 amino acids in length. In another example, at least one conserved peptide sequence may be no more than about 200 amino acids in length. In another example, at least one conserved peptide sequence may be no more than about 150

amino acids in length. In another example, at least one conserved peptide sequence may be no more than about 100 amino acids in length. In another example, at least one conserved peptide sequence may be no more than about 55 amino acids in length. In another example, 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 examples of the disclosure, 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 example, 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 disclosure. 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 examples, 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 examples without the peptide adjuvant, the first linker sequence may not be provided.

Combinations of nucleic acids may encode different polypeptides according to the disclosure 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 disclosure, which may be different relative to each other.

In another aspect, the present disclosure refers to a vaccine comprising a nucleic acid as disclosed herein.

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.

In another aspect, the present disclosure refers to a composition comprising a plurality of different nucleic acids as disclosed herein, optionally wherein the composition is a pharmaceutically acceptable composition.

In another aspect, the present disclosure refers to a polypeptide encoded by the nucleic acid as disclosed herein.

In one example 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.

In another aspect, the present disclosure refers to a viral vector comprising the nucleic acid as disclosed 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 disclosure (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 disclosure herein. In another example, the nucleic acid of the disclosure (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 disclosure 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 disclosure 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 example of the disclosure there is provided a composition comprising one or more of:

-the polypeptide according to the disclosure;

-the nucleic acid according to the disclosure; and

-the viral vector according to the disclosure.

In another aspect, the present disclosure refers to a composition comprising one or more of:

-the nucleic acid as disclosed herein;

-the polypeptide as disclosed herein; and/or

-the viral vector as disclosed herein.

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.

In another aspect, the present disclosure refers to the nucleic acid as disclosed herein; the vaccine as disclosed herein; the polypeptide as disclosed herein; the viral vector as disclosed herein; or the composition as disclosed herein, for use in the prophylaxis or treatment of HPV infection.

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

-the polypeptide according to the disclosure;

-the nucleic acid according to the disclosure;

-the composition according to the disclosure or

-the viral vector according to the disclosure.

In another aspect, the present disclosure refers to use of the nucleic acid as disclosed herein; the vaccine as disclosed herein; the polypeptide as disclosed herein; the viral vector as disclosed herein; or the composition as disclosed herein, in the manufacture of a medicament for the treatment or prophylaxis of HPV infection.

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

According to another example of the disclosure 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 disclosure;

-the composition according to the disclosure;

-the nucleic acid according to the disclosure; or

-the viral vector according to the disclosure.

In another aspect, the present disclosure refers to an agent for use in the prophylaxis or treatment of HPV infection, the agent comprising or consisting of the nucleic acid as disclosed herein; the polypeptide as disclosed herein; the viral vector as disclosed herein.

In one example, 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 example of the disclosure there is provided the polypeptide according to the disclosure; the composition according to the disclosure; the nucleic acid according to the disclosure; or the viral vector according to the disclosure; for use in, or as, a vaccine.

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

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 disclosure herein provides any of the sequences/examples of the disclosure wherein the order

in which conserved peptide sequences are provided may be changed. Such examples 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 example of the disclosure, there is provided a nucleic acid or polypeptide according to the disclosure for use in, or as, a vaccine.

According to another example of the disclosure, there is provided a prime boost vaccination kit comprising

-a prime vaccination according to the disclosure;

-a boost vaccination according to the disclosure.

In another aspect, the present disclosure refers to a prime boost vaccination kit comprising:

-a prime vaccination comprising the nucleic acid as disclosed herein; the vaccine as disclosed herein; the polypeptide as disclosed herein; the viral vector as disclosed herein; or the composition as disclosed herein; and

-a boost vaccination comprising a nucleic acid as disclosed herein; the vaccine as disclosed herein; the polypeptide as disclosed herein; the viral vector as disclosed herein; or the composition as disclosed herein.

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 example of the disclosure, there is provided a composition comprising a nucleic acid according to the disclosure 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 disclosure 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 example or aspect of the disclosure may be applicable, where appropriate, to other examples or aspects of the disclosure.

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

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. Use 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, in the manufacture of a medicament for the treatment or prophylaxis of HPV infection.

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.