

METHODS FOR ANALYSING VIRUSES USING RAMAN SPECTROSCOPY

Field of the Invention

- 5 The present invention relates to the use of Raman spectroscopy for the monitoring and assessment of viral titre and/or viral component abundance.

Background to the Invention

- 10 Viral vector manufacture is a crucial process step for the production of many cell and gene therapies, and the growth in this industry has resulted in an increased demand for viral vector supply. In view of this, it is important that analytical tools are available to ensure that viral vector production processes can be monitored and optimised, and that viral vector product can be quantified and characterised. Both the demand for viral vector and the cost
- 15 of production places a particular importance on achieving good viral vector titres during production. Currently, physical viral titre measurements are typically carried out by standard analytical techniques, *e.g.* for lentiviral titre, ELISA to assess p24 or qPCR for the measurement of viral RNA, and for AAV, RT qPCR with primers targeting the ITR, are frequently used. However, these methods are time consuming, are often inaccurate and
- 20 require sampling of media. As such, these methods only provide a retrospective measurement of the viral concentration. New methods for measuring viral titre are therefore needed.

- In addition, for an efficient virus or viral particle production process, especially one used to
- 25 create viral particles for medicinal applications, it is advantageous that the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules is monitored to maximise the proportion of functional viral particles produced. Non-functional viral particles, such as empty particles, are generally considered to be a waste in the production process, and can cause problems if administered to a patient, such as undesired immune
- 30 reactions. At present the most accurate method of quantification of the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules is transmission electron microscopy (TEM). An alternative is to carry out ELISA and qPCR experiments

to calculate viral protein and nucleic acid quantities, respectively. However, the TEM, ELISA and qPCR these methods of quantification can only be carried out retrospectively. Thus, there is currently no method available to quantify viral component abundance in real time in order to calculate the proportion of functional viral particles in a sample.

5

Raman spectroscopy is a vibrational form of spectroscopy, which has been shown to have particular utility for process analytical technology (PAT) applications where molecular information is required. The technique is based upon the detection of wavenumber shifts in photons which have been inelastically scattered by molecules present within a sample
10 (where the difference in wavenumber of such photons either relates to the energy lost from photons by altering the vibrational state of particular molecules from ground state to a first excited state or relates to the energy gained by photons from de-exciting molecules from an excited vibrational state to the ground state). This technique has many advantages over previously used methodologies for PAT applications, most notably the relatively weak
15 signal that is generated by water in comparison to other systems, which facilitates bioprocess monitoring, cell culture analysis and protein analysis in solution. Raman spectroscopy has, for example, been used for variance testing and to determine the identity of raw materials and cell culture media; to characterise macromolecular products; to analyse drug formulations, batch to batch variability, contamination, degradation of media,
20 cell densities including viable cell density and total cell density, protein structure, and protein stability; for polymer and fiber analysis; for material ID testing and for the quantification of glucose, glutamine, lactate and ammonia. Buckley and Ryder (2017) provides a review of the applicability of Raman spectroscopy.

25 Conventional Raman spectroscopy may be associated with the production of weak signals which has been reported to impact on its use for sensitive quantitative analysis as is often required in the biological field. Indeed, the concentration limit for detection of glucose using conventional Raman spectroscopy is reported to be 0.6mM and for phenylalanine is reported to be 1.1mM (Buckley and Ryder, 2017, Applied Spectroscopy, 71, p 1085-1116)
30 (estimated to be equivalent to 0.11mg/ml and 0.18mg/ml, respectively). In view of this, conventional Raman spectroscopy has been used primarily in the art to assess cell culture media and cell growth, and other methods have been employed where more sensitive

measurements are required, for example, to detect entities which are below the reported limits of detection for conventional Raman spectroscopy, such as viral particles. In this respect, Lee et al (2015) describes the use of Surface Enhanced Raman Spectroscopy (SERS) for the detection of HIV-1. SERS differs from conventional Raman and provides enhanced signals from molecules which are adhered to roughened metal surfaces at the nanoscale, such as Ag, Cu or Au. In Lee et al, an Au nanodot fabricated indium tin oxide substrate comprising bound anti-gp120 antibody fragments was used for the specific binding of HIV-1 virus-like particles, to ensure the generation of an enhanced signal for HIV-1 virus-like particle detection. SERS however, can be associated with limitations, including the requirement to pre-prepare a substrate with an appropriate immuno-interactive molecule, which limits the generic application of the technology, and the increased cost associated with this. Further, SERS, by the nature of the requirement of binding to the entity of interest, is always invasive within a sample.

Thus, further methods are required to characterise and to assess entities which may be associated with sensitive or weak signals, such as viruses, and to assess such entities *in situ*. In addition, there is a need for methods to monitor viral component abundance in real time, in order to monitor and maximise the proportion of functional viral particles being produced.

20

Summary of the Invention

Surprisingly, in direct contrast to the teaching in the art, the inventors have shown that conventional Raman spectroscopy can be used to accurately and sensitively quantify viral titre in real-time. This can be achieved by directly irradiating the viral culture medium with a light source and obtaining Raman spectroscopy data as further described herein. The methods thus obviate the need for processing of the viral culture medium. The inventors have shown that viral titre as determined by conventional Raman spectra according to the methods described herein is comparable to offline titre measurements (*e.g.* as measured by offline assays such as RT-qPCR and p24 ELISA), and that conventional Raman spectroscopy used in accordance with the methods described herein can provide an alternative rapid and reliable method to assess viral titre. As discussed previously, this

finding is particularly unexpected in view of the prior art where conventional Raman spectroscopy was associated with the production of weak signals, and where the low concentration of virus in solution would have been understood to be beneath the typical lower limits of detection using conventional Raman spectroscopy, *e.g.* when applied to
5 bulk solution systems, especially those with large scattering variation and in the presence of unwanted fluorescent background signals. This finding is also particularly unexpected in view of the prior art which requires the processing of viral culture medium to allow the use of Raman spectroscopy, for example processing of viral culture medium to concentrate viruses prior to analysis using Raman spectroscopy, or to prevent biofilm formation.

10

In addition, the inventors have shown that it is possible to use conventional Raman spectroscopy to monitor viral component abundance in real time, and in particular to quantify viral nucleic acid abundance and viral structural molecule abundance in real time. Thus, Raman spectroscopy can be used to determine the ratio of viral nucleic acids to
15 viruses comprising one or more viral structural molecules in a sample, and thus determine the proportion of functional viral particles produced.

The inventors have identified a series of spectral variables which are important in enabling predictions with models processing the real-time Raman spectroscopy data to achieve the
20 measure of real-time viral titre and/or viral nucleic acid abundance and viral structural molecule abundance. The inventors have established stratified ranges of increasing numbers of variables with different importance thresholds to provide variable ranges for the accurate prediction of viral titre and/or viral nucleic acid abundance and viral structural molecule abundance when using Raman spectroscopy. The inventors have shown that
25 Raman spectroscopy can be used to identify the start, production phase, and end of the viral production process.

The present invention thus relates to the use of Raman spectroscopy for the monitoring and/or assessment of viral nucleic acid abundance and viral structural molecule abundance
30 in a sample. Alternatively viewed, the invention relates to a method for monitoring and/or assessing viral nucleic acid abundance and viral structural molecule abundance in a sample, comprising the steps of analysing a Raman spectrum of a sample comprising virus

using a multivariate model and determining viral nucleic acid abundance and viral structural molecule abundance. Such a method may additionally comprise a step of carrying out Raman spectroscopy on the sample, *i.e.* to obtain the spectrum.

- 5 The present invention further specifically relates to the use of Raman spectroscopy for the monitoring and/or assessment of the start, production phase, and/or end of the viral production process. The inventors have identified that Raman spectroscopy can be used for the determination of viral nucleic acid abundance and viral structural molecule abundance, as discussed above, and particularly the inventors have identified specific
- 10 wavenumber ranges which require assessment for this purpose. The intensity of such peaks may be determined in a method of the invention, where the intensity of such peaks may result in the production of a fingerprint which can be assessed with a multivariate model to determine viral nucleic acid abundance and viral structural molecule abundance. In a preferred embodiment of the invention the viral nucleic acid abundance and viral
- 15 structural molecule abundance which is monitored and/or assessed is adeno associated virus viral nucleic acid abundance and adeno associated virus viral structural molecule abundance .

- One advantage of the present invention is that viral nucleic acid abundance and viral
- 20 structural molecule abundance, and the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules, can be continuously monitored in real-time. There is no need to process samples from the viral culture medium to generate an estimate of viral nucleic acid abundance, viral structural molecule abundance and the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules. Measurements
- 25 may be made *in situ*, if desirable. In other words, measurements may be made directly on the viral culture medium in the growth incubator. Measurements may be made *ex situ*, if desirable. In other words, measurements may be made directly on the viral culture medium in an aliquot of the viral culture medium taken from the growth incubator or separated from the main chamber of the growth incubator. Whether measurements are
- 30 made *in situ* or *ex situ*, measurements may be made directly on the viral culture medium without the need for further processing of the viral culture medium. This type of approach is sometimes described as being 'in-line' or 'at-line' analysis. Thus, more accurate trends

in viral nucleic acid abundance and viral structural molecule abundance can be produced, without risk of contamination. The methods of the present invention are thus faster and simpler than conventional off-line methods which require processing of viral culture medium. Particularly, the production stage of the culture can be much more accurately
 5 measured, leading to a more accurate timing of the end/harvesting stage of viral production, allowing process cessation at an appropriate time point, potentially reducing the cost of the production process.

Therefore, the invention provides: a method of determining in a sample using Raman
 10 spectroscopy the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules, the method comprising the steps of:

- (a) providing a sample and irradiating the sample with a light source;
- (b) (i) measuring the total intensity of Raman scattered light within each
 15 one of a first plurality of wavenumber ranges to obtain a first wavenumber intensity data set for the sample, wherein the first plurality of wavenumber ranges are pre-selected and are characteristic of viral nucleic acids in the sample;
- (ii) performing a first set of mathematical data processing steps on the first wavenumber intensity data set; and
- (iii) determining the viral nucleic acid content of the sample based upon
 20 the output of the first set of mathematical data processing steps;
- (c) (i) measuring the total intensity of Raman scattered light within each
 25 one of a second plurality of wavenumber ranges to obtain a second wavenumber intensity data set for the sample, wherein the second plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample;
- (ii) performing a second set of mathematical data processing steps on the second wavenumber intensity data set; and
- (iii) determining the content of viruses comprising the one or more viral
 30 structural molecules in the sample based upon the output of the second set of mathematical data processing steps; and

- (d) determining the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules in the sample based on the values determined in steps (b)(iii) and (c)(iii).

5 In the above-defined method, steps (b) and (c) may be performed in the order (b) then (c) or in the order (c) then (b). The exact order in which the steps are performed is not essential, provided that the values identified in in steps (b)(iii) and (c)(iii) are determined such that the ratio may be determined in step (d). Moreover, steps (b) and (c) may be performed simultaneously, since mathematical data processing may allow first and second
10 wavenumber intensity data sets to be processed at the same time in order to provide the values identified in in steps (b)(iii) and (c)(iii).

In the above-defined method the steps of performing the first and second sets of mathematical data processing steps on the first and second wavenumber intensity data sets
15 may comprise:

- (i) optionally normalising the wavenumber signal intensity data by pre-processing the signal intensity data using one or more pre-processing analytical methods, such as a first derivative method, a second derivative
20 method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares (airPLS) method;
- 25 (ii) obtaining model parameters by applying to the wavenumber signal intensity data a multivariate regression algorithm, such as a partial least squares (PLS) regression algorithm, optionally wherein the PLS algorithm is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network; and
- 30 (iii) determining the viral nucleic acid content of the sample and determining the content of viruses comprising the one or more viral structural molecules in

the sample using the model parameters obtained by applying the multivariate regression algorithm to the signal intensity data.

In any of the above-defined methods, the light source used to irradiate the sample may be a laser and the sample may be irradiated with light having a wavelength of 785nm.

In any of the above-defined methods, the Raman scattered light may be detected using a charge-coupled device (CCD).

- 10 In one embodiment of the above-defined methods, the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample may comprise 4 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 12 of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 . In any of these methods, the virus may preferably be an adeno-associated virus (AAV).
- 25 The first plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 4 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or the plurality of

wavenumber ranges in the Raman spectrum which are measured may comprise all 10 of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 . In any of these methods, the virus may preferably be an adeno-associated virus (AAV).

- 5 The first plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 4 or more of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges
10 in the Raman spectrum which are measured may comprise all 8 of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 . In any of these methods, the virus may preferably be an adeno-associated virus (AAV).

- In the same or in another embodiment of the above-defined methods, the second plurality
15 of wavenumber ranges in the Raman spectrum which are measured to obtain the second wavenumber intensity data set for the sample may comprise 4 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or the
20 plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum
25 which are measured may comprise 12 or more, 14 or more, 16 or more or 18 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 20 of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 . In any of these methods, the virus may preferably be an adeno-associated virus (AAV).

- 30 The second plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 4 or more of the wavenumber ranges 21 to 33 as listed in Table

2 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber
5 ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more, 11 or more or 12 of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 13 of the wavenumber ranges 21 to 33 as listed in Table 2 and
10 wherein the VIP is ≥ 1.25 . In any of these methods, the virus may preferably be an adeno-associated virus (AAV).

The second plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 4 or more of the wavenumber ranges 34 to 40 as listed in Table
15 2 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 5 or 6 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 7 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 . In any of these methods, the virus
20 may preferably be an adeno-associated virus (AAV).

In another embodiment of the above-defined methods, the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample may comprise 5 or more of wavenumber ranges 1 to 28 as listed in
25 Table 3 and wherein the variable importance projection (VIP) is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 15 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is $\geq$
30 1.00; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 20 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are

measured may comprise 25 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 28 of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 . In any of these methods, the virus may preferably be a
5 lentivirus.

The first plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 5 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the variable importance projection (VIP) is ≥ 1.25 ; or the plurality of wavenumber
10 ranges in the Raman spectrum which are measured may comprise 10 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 15 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured
15 may comprise 20 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 25 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 30 of wavenumber ranges 29 to 59 as listed in
20 Table 3 and wherein the VIP is ≥ 1.25 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 31 of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 . In any of these methods, the virus may preferably be a lentivirus.

The first plurality of wavenumber ranges in the Raman spectrum which are measured may alternatively comprise 5 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the variable importance projection (VIP) is ≥ 1.50 ; or the plurality of wavenumber
25 ranges in the Raman spectrum which are measured may comprise 10 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise
30 15 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured

- may comprise 20 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 22 of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 . In any of these methods, the virus may preferably be a
- 5 lentivirus.

In any of the above-defined methods, the nucleic acid may comprises a viral DNA genome or a viral RNA genome.

- 10 In any of the above-defined methods, the one or more viral structural molecules may comprise one or more viral proteins such as one or more nucleoproteins and/or one or more capsomeres, one or more viral carbohydrates, one or more glycosylated viral molecules such as a glycosylated viral protein and/or one or more viral lipids.
- 15 In any of the above-defined methods, the ratio may provide a measure of functional viral titre.

- In any of the above-defined methods, the sample may be a viral culture. The viral culture may be comprised in a bioreactor. In any such methods, the steps of irradiating the viral
- 20 culture with a light source and measuring the total intensity of Raman scattered light may be performed directly on the medium of the viral culture (*in situ*). Alternatively the steps of irradiating the viral culture with a light source and measuring the total intensity of Raman scattered light may be performed directly on an aliquot of the medium which has been taken from the viral culture (*ex situ*).

- 25 Any of the above-defined methods may comprise a first step of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules at a first time point and one or more further steps of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules at later time points, and wherein
- 30 the method further comprising measuring the change in the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in the sample between time points, wherein each step is performed by a method according to any one of the above-

- defined methods, preferably wherein each step is performed by the same method. In any such method the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules may be determined repeatedly over a time period to provide a measure of the change in the ratio in real time. The change in the ratio in the sample may be used to
- 5 determine the start phase, the production phase and/or the stationary phase of a viral production process. Any such method may be used to determine the optimal conditions for a viral production process. Any such method may be used to assess a process downstream of a viral production process.
- 10 Any of the above-defined methods may comprise a step of comparing the ratio thereby obtained with the ratio obtained from the same sample by an alternative method, optionally wherein the alternative method is qPCR, RT-qPCR, ELISA or by visual determination by transmission electron microscopy.
- 15 The invention also provides a method of determining the extent of viral infection in an individual using Raman spectroscopy, the method comprising determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample by performing the method of any one of the above-defined methods, wherein the sample is a sample which has previously been obtained from the individual. The sample may be a
- 20 sample of blood, saliva, sputum, plasma, serum, cerebrospinal fluid, urine or faeces. In any such method the ratio in the sample from the subject may be compared with one or more ratio measurements which have previously been obtained for the infection in the individual, in order to provide a prognosis of the stage of infection in the individual.
- 25 The invention additionally provides a method of determining in a sample using Raman spectroscopy the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules, the method comprising the steps of:
- (a) (i) providing a first wavenumber intensity data set for the sample, wherein the first data set has been obtained by irradiating the sample
- 30 with a light source and measuring the total intensity of Raman scattered light within each one of a first plurality of wavenumber ranges, wherein the first plurality of wavenumber ranges in the

- Raman spectrum have been selected as characteristic of viral nucleic acids in the sample;
- (ii) performing a first set of mathematical data processing steps on the first wavenumber intensity data set; and
 - 5 (iii) determining the nucleic acid content of the sample based upon the output of the first set of mathematical data processing steps;
 - (b) (i) providing a second wavenumber intensity data set for the sample, wherein the second data set has been obtained by irradiating the sample with a light source and measuring the total intensity of
 - 10 Raman scattered light within each one of a second plurality of wavenumber ranges, wherein the second plurality of wavenumber ranges in the Raman spectrum have been selected as characteristic of one or more viral structural molecules of the viruses in the sample;
 - (ii) performing a second set of mathematical data processing steps on
 - 15 the second wavenumber intensity data set; and
 - (iii) determining the content of viruses comprising the one or more viral structural molecules in the sample based upon the output of the second set of mathematical data processing steps;
 - (c) determining the ratio of viral nucleic acids to viruses comprising the
 - 20 one or more viral structural molecules in the sample based on the values determined in steps (a)(iii) and (b)(iii).

The method described immediately above may be performed according to the steps defined in any one of the methods described and defined herein.

- 25 In the method described immediately above, steps (a) and (b) may be performed in the order (a) then (b) or in the order (b) then (a). The exact order in which the steps are performed is not essential, provided that the values identified in in steps (a)(iii) and (b)(iii) are determined such that the ratio may be determined in step (c). Moreover, steps (a) and
- 30 (b) may be performed simultaneously, since mathematical data processing may allow first and second wavenumber intensity data sets to be processed at the same time in order to provide the values identified in in steps (a)(iii) and (b)(iii).

The invention further provides the use of Raman spectroscopy for determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample. In any such use, the ratio may be determined based upon measurements of the intensity of Raman scattered light obtained from the sample following irradiation of the sample with a light source, wherein the intensity of Raman scattered light is measured from a first plurality of wavenumber ranges in a Raman spectrum which are characteristic of viral nucleic acids in the sample and from a second plurality of wavenumber ranges in a Raman spectrum which are characteristic of the one or more viral structural molecules of the viruses in the sample.

In any such use the sample may be a viral culture, optionally wherein the viral culture is comprised in a bioreactor. The step of measuring the total intensity of Raman scattered light may be performed directly on the medium of the viral culture (*in situ*). Alternatively the step of measuring the total intensity of Raman scattered light may be performed directly on an aliquot of the medium which has been taken from the viral culture (*ex situ*).

In any such use the ratio in the sample may be determined at a first time point and at one or more later time points, and wherein the change in the ratio in the sample between time points is calculated. In any such use the ratio in the sample is quantified repeatedly to provide a measure of the change in the ratio in real time. In any such use the viral titre in the sample may be quantified by performing any of the methods described and defined herein.

In any of the above-defined methods or uses, the viruses in the sample may not be HIV-1 or HIV-1 virus-like particles (HIV-1 VLPs).

In any of the above-defined methods or uses the Raman spectroscopy may not be surface enhanced Raman spectroscopy.

The invention further provides a method of building a multivariate data processing model which is capable of determining the content of viruses comprising one or more viral

structural molecules in a sample from a Raman spectroscopy wavenumber intensity data set obtained for the sample, the method comprising:

- (a) providing the sample and irradiating the sample with a light source;
- (b) measuring the total intensity of the Raman scattered light within each one of a plurality of wavenumber ranges to obtain a wavenumber intensity data set for the sample, wherein the plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample;
- (c) obtaining normalised wavenumber signal intensity data by pre-processing the signal intensity data using a pre-processing analytical method, such as a first derivative method, a second derivative method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares (airPLS) method;
- (d) obtaining model parameters by applying to the pre-processed signal intensity data a multivariate regression algorithm, such as a partial least squares (PLS) regression algorithm, optionally wherein the PLS algorithm is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network, wherein a calibration is performed wherein the pre-processed signal intensity data are compared with viral titre data obtained for the same sample conditions using non-Raman spectroscopy methods such as qPCR, RT-qPCR, ELISA or by visual determination by transmission electron microscopy;
- (e) inferring response values using the model parameters obtained from the pre-processed data; and
- (f) performing variable selection, optionally variable importance projection (VIP), and identifying Raman spectral variables; and
- (g) optionally performing one or more further rounds of modelling by re-applying steps (d) to (f) and wherein unimportant variables are removed; and

wherein the content of viruses comprising one or more viral structural molecules in a sample is determined using the model parameters obtained for the identified Raman spectral variables derived from the multivariate data processing model.

- 5 The invention yet further provides a method of building one or more multivariate data processing models which are capable of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample from a Raman spectroscopy wavenumber intensity data set obtained for the sample, the method comprising:
- 10 (a) providing the sample and irradiating the sample with a light source;
- (b) (i) measuring the total intensity of the Raman scattered light within each one of a first plurality of wavenumber ranges to obtain a first wavenumber intensity data set for the sample wherein the first plurality of wavenumber ranges are pre-selected and are characteristic of viral nucleic acids in the sample;
- 15 (ii) measuring the total intensity of the Raman scattered light within each one of a second plurality of wavenumber ranges to obtain a second wavenumber intensity data set for the sample wherein the first plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample
- 20 (c) obtaining normalised wavenumber signal intensity data for the first and second wavenumber intensity data sets by pre-processing the signal intensity data using a pre-processing analytical method, such as a first derivative method, a second derivative method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares (airPLS) method;
- 25 (d) obtaining model parameters to be applied to the first and second wavenumber intensity data sets by applying to each one of the pre-processed signal intensity data sets a multivariate regression algorithm, such as a partial least squares (PLS) regression algorithm, optionally wherein the
- 30

PLS algorithm is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network, wherein a calibration is performed wherein the pre-processed signal intensity data are compared with viral titre data obtained for the same sample conditions using non-Raman spectroscopy methods such as qPCR, RT-qPCR, ELISA or by visual determination by transmission electron microscopy;

(e) inferring response values using the model parameters obtained from each one of the pre-processed data sets; and

(f) performing variable selection, optionally variable importance projection (VIP), and identifying Raman spectral variables; and

(g) optionally performing one or more further rounds of modelling for any of the data sets by re-applying steps (d) to (f) and wherein unimportant variables are removed; and

wherein the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in the sample is determined using the model parameters obtained for the identified Raman spectral variables derived from the multivariate data processing models.

Brief Description of the Figures

Figure 1: Jablonski diagram showing quantum energy transitions for infrared absorption/emission. The diagram shows Rayleigh (elastic scattering) and Raman (inelastic scattering) with both Stokes and anti-Stokes transitions.

Figure 2: Example pre-processed Raman spectra from a VV Raman Project lentiviral bioreactor run. Inset is a blow up of the 1000 cm^{-1} region. Spectra were acquired for 10 seconds with 75 accumulations, for a total integration time of ~12 mins 30 s, after CCD readout time approximately 15 minutes.

Figure 3: Bar chart showing representative qPCR lentiviral titre results for bioreactor transfection in the viral vector Raman project, cp number = copy number.

Figure 4: A graph showing how the mean squared error of prediction (MSEPCV) after 10 fold cross validation and 20 monte carlo repeats varies as a function of the number of PLS latent variables/components (prior to spectral variable selection).

5 **Figure 5 (A):** A plot of variable importance projection (VIP) calculated from the initial 10 component PLS-R model. The circles indicate spectral variables with $VIP > 1.5$ and this information was used to determine the ranges shown in **Figure 5 (B)**. **Figure 5 (C):** Table of spectral variables with $VIP > 1.5$ with the ranges in order of importance.

10 **Figure 6:** A plot showing the change in the mean squared error of prediction after cross-validation as a function of latent variable or component number. Obtained after conservative spectral variable reduction using $VIP > 1.5$.

Figure 7: A plot showing the Raman copy number / mL PLS-R predictions (10 LV) from
15 lentiviral run 1 bioreactor 4 following spectral variable reduction ($VIP > 1.5$) alongside the offline qPCR data. T means “transfected” and NT means “not transfected”.

Figure 8: A plot showing the Raman copy number PLS-R predictions (10 LV) from
20 lentiviral run 2 bioreactors 1-4 following spectral variable reduction ($VIP > 1.5$) alongside the offline qPCR data.

Figure 9: A plot showing the Raman copy number PLS-R predictions (10 LV) from
25 lentiviral run 3 bioreactors 1- 4 following spectral variable reduction alongside the offline qPCR data.

Figure 10: Comparison RT-qPCR and p24 ELISA Results. The p24 ELISA assay was only used to obtain lentiviral titre on a few of the offline samples.

30 **Figure 11:** Example of application of real-time Raman-derived model of viral titre to identify the start, production phase, and end of the viral production process. From the model (solid black line), a set of 3 indicators are calculated also in real-time. Together with the model estimation of the viral titre, the shape of the indicator curves inform on the

various stages of the virus production (as explained in the text overlaid to the graph). The physical titre, calculated retrospectively using either RT-qPCR and/or ELISA methods, is overlaid to the graph (squares and dashed line) to show the overall agreement between real-time data and retrospective off line data. The inclusion of additional peaks from the Raman spectra can improve the quality of the Raman model, but this plot was generated using the minimum number of regions required to identify the main phases of viral production.

Figure 12: shows an outline schematic of a formula for quantifying viral titre. The formula is applied to the model parameters (in this case regression coefficients) which are obtained from the multivariate regression algorithm which was applied to normalised Raman signal intensity data.

Figure 13: shows a plot of R^2 ($1 - \text{residual sum of squares} / \text{total sum of squares}$) as a function of the number of wavenumber ranges to demonstrate the minimum number of wavenumber ranges which are required to provide an estimate of lentiviral titre.

Figure 14: Example pre-processed Raman spectra from an AAV Raman Project bioreactor run. Inset is a blow up of the $1000\text{-}1200\text{ cm}^{-1}$ region. Spectra were acquired for 10 seconds with 75 accumulations, for a total integration time of $\sim 12\text{ mins } 30\text{ s}$, after CCD readout time approximately 15 minutes.

Figure 15: Bar chart showing representative qPCR AAV viral titre results for bioreactor transfection in the AAV Raman project.

Figure 16: A graph showing how the mean squared error of prediction (MSEPCV) after 10 fold cross validation and 20 monte carlo repeats varies as a function of the number of PLS latent variables/components (prior to spectral variable selection).

Figure 17 (A): A plot of variable importance projection (VIP) calculated from the initial 15 component PLS-R model. The circles indicate spectral variables with $\text{VIP} \geq 1.0$ and

this information was used to determine the ranges shown in **Figure 17 (B)**. **Figure 17 (C):** Table of spectral variables with $VIP \geq 1.0$ with the ranges in order of importance.

Figure 18: A plot showing the change in the mean squared error of prediction after cross-validation as a function of latent variable or component number. Obtained after spectral variable reduction using $VIP \geq 1.0$.

Figure 19: A plot showing the Raman copy number PLS-R predictions (9 LV) from run 4 bioreactors 1- 4 following spectral variable reduction alongside the offline qPCR data.

10

Figure 20: A plot of R^2 ($1 - \text{residual sum of squares} / \text{total sum of squares}$) as a function of the number of wavenumber ranges to demonstrate the minimum number of wavenumber ranges which are required to provide an estimate of AAV viral titre.

Figure 21: Example pre-processed Raman spectra from an AAV Raman Project bioreactor run. Inset is a blow up of the $1000\text{-}1200\text{ cm}^{-1}$ region. Spectra were acquired for 10 seconds with 75 accumulations, for a total integration time of ~12 mins 30 s, after CCD readout time approximately 15 minutes.

Figure 22: Bar chart showing representative qPCR AAV viral titre results for bioreactor transfection in the AAV Raman project.

Figure 23: Bar chart showing representative ELISA AAV viral titre results for bioreactor transfection in the AAV Raman project.

25

Figure 24: A graph showing how the mean squared error of prediction (MSEPCV) for RT-qPCR copy number per ml after 10 fold cross validation and 20 monte carlo repeats varies as a function of the number of PLS latent variables/components (prior to spectral variable selection).

30

Figure 25: A graph showing how the mean squared error of prediction (MSEPCV) for RT-qPCR copy number per ml after 10 fold cross validation and 20 monte carlo repeats varies

as a function of the number of PLS latent variables/components (prior to spectral variable selection).

Figure 26: (A) A plot of variable importance projection (VIP), for the copy number per mL (RT-qPCR based) PLS-R model, calculated from the initial 15 component PLS-R model. The circles indicated spectra variable with $VIP \geq 1.0$ and this information was used to determine the ranges shown in **Figure 26 (B)**. **Figure 26 (C):** Table of spectral variables with $VIP \geq 1.0$ with ranges in order of importance.

Figure 27: (A) A plot of variable importance projection (VIP), for the total particle number per mL (ELISA based) PLS-R model, calculated from the initial 14 component PLS-R model. The circles indicated spectra variable with $VIP \geq 1.0$ and this information was used to determine the ranges shown in **Figure 27 (B)**. **Figure 27 (C):** Table of spectral variables with $VIP \geq 1.0$ with ranges in order of importance.

Figure 28: (A) A plot of variable importance projection (VIP), for the total particle number per mL (ELISA based) PLS-R model, calculated from the initial 14 component PLS-R model. The circles indicated spectra variable with $VIP \geq 1.0$ and this information was used to determine the ranges shown in **Figure 27 (B)**. **Figure 27 (C):** Table of spectral variables with $VIP \geq 1.0$ with ranges in order of importance.

Figure 29: A plot showing the change in the mean squared error of prediction for ELISA total viral particle number per ml after cross-validation as a function of latent variable or component number. Obtained after spectral variable reduction using $VIP \geq 1.0$.

Figure 30: A plot showing the Raman copy number PLS-R predictions (10 LV) from bioreactors 1- 4 following spectral variable reduction alongside the offline qPCR data.

Figure 31: A plot showing the Raman copy number PLS-R predictions (10 LV) from bioreactors 5- 8 following spectral variable reduction alongside the offline RT-qPCR data.

Figure 32: A plot showing the Raman total particle number PLS-R predictions (10 LV) from bioreactors 1- 4 following spectral variable reduction alongside the offline ELISA data.

- 5 **Figure 33:** A plot showing the Raman total particle number PLS-R predictions (10 LV) from bioreactors 5- 8 following spectral variable reduction alongside the offline ELISA data.

- 10 **Figure 34:** A plot showing the calculated Empty-Full Ratio (%) from the Raman PLS-R model predictions of genome copy number (RT-qPCR) and total particle number (ELISA). For bioreactors 2-4 shown from 24 hours post transfection.

- 15 **Figure 35:** A plot showing the calculated Empty-Full Ratio (%) from the Raman PLS-R model predictions of genome copy number (RT-qPCR) and total particle number (ELISA). For bioreactors 5-7 shown from 24 hours post transfection.

Detailed Description of the Invention

- 20 It is to be understood that different applications of the disclosed methods may be tailored to the specific needs in the art. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

- 25 In addition, as used in this specification and the appended claims, the singular forms “a”, “an”, and “the” include plural references unless the content clearly dictates otherwise.

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

Viral analysis using Raman spectroscopy

- 30 The present invention encompasses the use of Raman spectroscopy to monitor and assess viral titre and/or viral component abundance. The present invention encompasses the use

of Raman spectroscopy to monitor and assess viral nucleic acid abundance and viral structural molecule abundance. The present invention encompasses the use of Raman spectroscopy to monitor and assess the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample.

5

“Viral titre” as defined herein refers to the quantity of virus present in a given volume.

Any type of viral titre may be assessed with the present invention, *e.g.* physical viral titre, functional viral titre (also referred to as infectious viral titre) or transducing viral titre, may be assessed. In a particular embodiment, the physical viral titre may be assessed. Physical

10 viral titre is a measure of the concentration of viral particles in a sample, *e.g.* viral culture medium, and is usually based on the presence of a viral protein, such as p24, or viral nucleic acid. Physical titre may be expressed as viral particles per mL (VP/mL), viral genomes per mL (vg/mL), viral copies per mL, or RNA copies per mL and prior art assays to measure physical titre include ELISAs for p24 (*e.g.* Lenti-X p24 Rapid Titer kit
15 (Takara), or Lentivirus-Associated p24 ELISA kit (Cell Biolabs, Inc)), qPCR or ddPCR (*e.g.* AAV real-time PCR titration kit (Takara), or Adeno X qPCR titration kit (Takara)). Physical titre measurements do not always distinguish between empty or defective viral particles and particles capable of infecting a cell. Thus, the physical viral titre can be distinguished from functional titre or infectious titre which determines how many of the
20 particles produced can infect cells, and the transducing viral titre which determines how many of the functional viral particles contain a gene of interest (*e.g.* for the production of a viral vector, the transducing viral titre may be relevant). Thus, a determination of physical titre is not equivalent to a determination of functional titre, unless all particles in a sample are functional. Indeed, functional titre is often 100 to 1000 fold less than physical titre.

25

Alternatively, as discussed above, the functional or infectious titre may be measured or assessed with the present invention, where functional or infectious titre is a measure of the amount of viral particles present in a particular volume which are capable of infecting a target cell. Functional titre may be expressed as plaque forming units per mL (pfu/mL) or
30 infectious units per mL (ifu/mL). Off line assays which can be used to measure functional or infective titre include plaque assays, focus forming assays, end point dilution assays or flow cytometry. The transducing titre, as discussed above, is a measure of the amount of

viral particles present in a particular volume which are capable of infecting a target cell and which comprise a gene of interest. Transducing titre may be expressed as transducing units/mL and may be assessed using the assays used to assess functional titre above, together with any known assay which can determine the presence of the gene of interest, 5 *e.g.* PCR. A skilled person will appreciate that functional titre or transducing titre may be determined by scaling down any value obtained for physical titre. As discussed above, the fold differences between physical and functional or transducing titre are well understood in the art. Thus, in one aspect of the invention, functional or transducing titre may be determined indirectly by the methods of the invention (*e.g.* through scaling down a value 10 obtained for physical titre). The methods of the invention may therefore include an additional step of scaling down a determination of physical titre to determine the functional or transducing titre.

The methods of the invention can be used to monitor and assess viral nucleic acid 15 abundance and viral structural molecule abundance in a sample, *e.g.* viral culture medium. The methods of the invention can be used to determine the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules in a sample, *e.g.* viral culture medium. This ratio can be used to determine the proportion of viral particles in a sample that have both nucleic acid and structural components. This proportion can be used as an 20 estimate of functional titre. Thus, the methods of the invention can be used to estimate functional titre in real time, whereas previously known methods for estimating functional titre are retrospective and off-line.

It will be appreciated by a skilled person that the methods of the invention are capable of 25 determining the viral titre and/or viral component abundance of any virus of any serotype, for example, retroviruses such as lentivirus (*e.g.* HIV-1 and HIV-2) and gamma retrovirus; adenovirus and adeno-associated virus (*e.g.* AAV1-11, particularly AAV1, AAV2, AAV5 and AAV8, and self-complementary AAV). Accordingly, the methods of the invention are capable of being applied, as further described and defined herein, to mammalian viruses. 30 The methods of the invention are further capable of being applied, as further described and defined herein, to non-mammalian viruses including plant viruses such as tobacco mosaic virus, algal viruses, yeast viruses and insect viruses including baculoviruses. Although

typically, viral titre and/or viral component abundance of a single virus type may be assessed in the present invention, the methods of the invention would be capable of assessing titre and/or viral component abundance of a mixed virus sample, *e.g.* a sample comprising two or more virus types.

5

In this respect, the Raman spectra produced in accordance with the present invention may either directly detect virus or viral components, indirectly detect virus or viral components or both directly and indirectly detect virus or viral components. Thus, the wavenumber peaks which are generated and shown on a Raman spectrum may be indicative of any one
10 of a number of different compounds or molecules associated with the virus or viral components. A skilled person will appreciate that for the performance of the present invention, it is unnecessary to identify exactly which compounds/molecules each wavenumber peak identified relates to. The spectrum instead acts as a fingerprint under particular conditions (*e.g.* culture conditions, virus, producer cell line and type of viral titre
15 to be measured, etc.) where viral titre and/or viral component abundance can be determined by analysis of the intensity of the signal at each of the wavenumber ranges disclosed herein with a particular multivariate model (typically one that has been produced under the same or similar conditions). Multivariate models are described in more detail herein. Therefore, the peaks which are obtained for a particular Raman spectrum, at
20 particular wavenumbers, may correspond to molecules/compounds which are viral (*e.g.* capsid proteins etc.), or may correspond to molecules/compounds which are non-viral (*e.g.* metabolites in the culture, *e.g.* produced by the virus producing cells). In this way, the Raman spectroscopy used in the present invention may be detecting compounds which are indirectly associated with viral titre and/or viral component abundance as well as, or
25 instead of detecting compounds which are directly associated with viral titre and/or viral component abundance.

It is expected that whilst the intensity of signal at each wavenumber range may be different and may correspond differently to viral titre and/or viral component abundance in Raman
30 spectra obtained under different conditions (*e.g.* for producing different viruses, using different producer cell lines or with different culture media or bioreactors), the wavenumber ranges which may be assessed (*e.g.* which have been determined to be of

- relevance to viral titre and/or viral component abundance), will remain the same. Furthermore, it will be appreciated that for any given set of conditions, a user will be able to create a multivariate model, based on principles described herein and known in the art, and apply this model when analysing different signal intensity data from the same
- 5 wavenumber ranges described herein and which are generated subsequently using the same set of conditions. This means that the user can calculate viral titre and/or viral component abundance via Raman spectroscopy using the wavenumber ranges described herein, wherein data is generated in systems to which different conditions are applied.
- 10 Lentiviral assessment of the wavenumber ranges as used herein, has identified ranges which are important in the assessment of viral titre and/or viral component abundance. As lentivirus is known to be a particularly complex virus in terms of its chemical composition, it is likely that the chemical components of simpler viruses which possess a portion of the components present in lentivirus will produce relevant signals which fall within a portion
- 15 of the identified wavenumber ranges (if any one or more of the wavenumber ranges directly detects the virus), *e.g.* at least 5 of the identified wavenumber ranges. Thus, subsets of the wavenumber ranges provided herein can be used to assess the titre of viruses other than lentivirus. Further, if one or more of the wavenumber ranges identified as correlating to viral titre and/or viral component abundance indirectly detect the virus, then,
- 20 as it is likely that any viral transfection will result in similar metabolomic changes in culture, such wavenumber ranges will likely be useful for the assessment of viral titre and/or viral component abundance in any system.

In a particular embodiment of the invention, lentiviral titre and/or lentiviral component

25 abundance is monitored and assessed. In one aspect, the virus may not be HIV-1 or HIV-1 virus like particles.

A “virus” is typically a small infectious agent (typically smaller than a bacterium) that is only capable of replicating inside the living cell of another organism. Viruses may have

30 RNA or DNA-based genomes. A “virus” as used herein, refers to any virus, modified virus, viral particle, virus-derived particle or viral vector. Thus, although the viral titre and/or viral component abundance of any wild type virus may be assessed in accordance

with the present invention, it will be appreciated that the utility of the invention may particularly extend to the assessment of viral titres and/or viral component abundance of mutant or modified viruses (*i.e.* comprising one or more nucleic acid substitutions, insertions, deletions or translocations as compared to a wildtype or naturally occurring virus, or absent large portions of genetic material encoding for viral proteins) or viral
5 vectors. Mutant or modified viruses or viral particles are often used to produce vaccines, and it is envisaged the methods of the present invention would be particularly effective in monitoring and assessing the efficiency of production of functional viruses or viral particles for use in vaccines.

10

Whilst viral vectors may be based on wild type viruses, they are generally modified as compared to wild type viruses and are commonly used to introduce genetic material into target cells (*e.g.* genes of therapeutic use). Viral vectors therefore have particular utility, *e.g.* for gene therapy, cell therapy or for other molecular applications, and their production
15 is of enormous importance to the gene therapy and cell therapy industries. It will be well understood that for example, modifications may be made to improve safety of viral vectors for gene and/or cell therapy or to improve for example the size of gene which may be carried by the vector. Modifications that may be made to create a viral vector may include the deletion of part of a viral genome which is critical for replication, resulting in a viral
20 vector that is capable of infecting cells but which would require the presence of a helper virus to provide missing proteins which would be required for the production of new virions. Other modifications may include modifications to lower the toxicity of the viral vector on its target cell and/or to improve stability of the virus, *e.g.* to reduce rearrangement of the genome.

25

Viral vectors may typically be produced in packaging cell lines, such as HEK293 cells, by the transduction of the packaging cell line with one or more plasmids encoding viral proteins and carrying the required genetic material. For example, for lentiviral production HEK293 cells may be transduced with one or more plasmids, *e.g.* 3 or 4 plasmids encoding
30 virion proteins, such as the capsid and the reverse transcriptase and carrying the genetic material to be delivered by the vector. This is transcribed to produce the single stranded RNA viral genome and is marked by the presence of the psi sequence which ensures that

the genome is subsequently packaged into the virion. Thus, particularly, lentiviral vectors may be produced by the transformation and expression of three (for second generation systems) or four (for third generation systems) plasmids in a producer cell line. Plasmids for the production of viral vectors are commercially available, *e.g.* Lenti-Pac and AAV Prime (GeneCopoeia).

Particularly, the titre and/or viral component abundance of viral vectors which are produced by packaging cell lines may be monitored or assessed by a method of the invention. As discussed previously, it is particularly important in the gene therapy and cell therapy fields to be able to measure produced titre and/or viral component abundance in a sensitive manner, *e.g.* so that production processes, such as production from a producer cell line, can be accurately monitored and managed, and the proportion of functional vectors can be calculated in real time. A virus does not need to be fully functional or wildtype to be monitored or assessed by a method of the invention.

“Viral components” are considered herein to be any part of the virus, virus particle or viral vector.

A viral particle or “virion” is conventionally understood to consist of: (i) the genetic material of the virus, *i.e.*, molecules of DNA or RNA that encode the structure of the proteins by which the virus acts; (ii) an internal protein coat, referred to as the capsid, formed from capsomeres, which surrounds and protects the genetic material of the virus; and, in some cases, (iii) an outside envelope of lipids which may include envelope proteins.

Viral components include viral nucleic acids and viral structural components (or viral structural molecules). Viral nucleic acids are considered herein to include viral RNA, viral DNA, viral DNA genomes and viral RNA genomes. Viral nucleic acids are packaged within the virion.

A viral structural component or viral structural molecule as used herein is to be understood as any molecule that contributes to the structure of the virus. A viral structural component

or viral structural molecule as used herein may exclude the genetic material of the virus, *i.e.*, molecules of DNA or RNA that encode the structure of the proteins by which the virus acts.

- 5 Viral structural components or viral structural molecules are considered herein to include viral proteins such as nucleoproteins, capsid proteins, protomers, capsid subunits, capsid monomers, combinations of capsid monomers, capsomeres, hexons, pentons, viral coat proteins (VCPs), viral outer surface glycoproteins, viral transmembrane proteins, proteins that are essential for the function of the virus, virus particle or viral vector, viral
- 10 carbohydrates, glycosylated viral molecules such as a glycosylated viral protein and/or viral lipids including viral phospholipids, or combinations thereof.

As discussed above, the methods of the invention can “monitor or assess” viral titre and/or viral component abundance. Thus, the methods of the invention are capable of

15 determining viral titre and/or viral component abundance *e.g.* levels, amounts or concentration of viral nucleic acids and viral structural molecules present in a sample. Particularly, the methods can thus determine whether levels, amounts or concentration of viral nucleic acids and viral structural molecules increase or plateau over time relative to each other (*e.g.* by assaying a sample at different time points), or vary (*e.g.* increase,

20 decrease or are equivalent) compared to different samples (*e.g.* assayed at the same or equivalent time point). In this way, the methods of the invention can be used for example, to assess the efficiency of a production method of the virus *e.g.* where the detection or determination of the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules can be indicative of an efficient method or a sub-optimal production

25 method), or can be used to determine the importance of particular factors in the production method of the virus, for example, by comparison with viral titres and/or viral component abundance measured during other modified production methods (for the same or different virus). Physical titre values which are expected to be detected by the methods described herein are in the range of 1×10^{10} to 1×10^{11} particles/mL. Infectious titre values which

30 are expected to be detected by the methods described herein are in the range of 1×10^8 to 1×10^9 particles/mL. Thus, the modification of a factor which results in a difference in viral titre and/or viral component abundance measured may be determined to be important to the

production method (*e.g.* modification of a factor which results in a difference of at least 5, 10, 20, 30, 40 or 50% in viral titre measured). Such a factor could include incubation temperature, culture media used, % glucose or amino acids used in the media, the presence, absence or amount of agitation used, or the culture flask or volume used, etc.

5

Thus, the methods of the invention could be used to determine optimal conditions of viral production, including an assessment of different systems available for culturing the producer cells which may produce the virus, *e.g.* shaker flasks, Quantum system (Terumo), Ambr systems, *e.g.* Ambr 15 or 250 (TAP Biosystems).

10

The methods of the invention could further be used to assess any process downstream of the viral production process, *e.g.* to determine whether any such process has affected viral titre and/or viral component abundance. Particularly, the methods of the invention could be used to assess purification methods which may be employed, *e.g.* to determine whether such purification methods have had any impact on titre and/or viral component abundance, *e.g.* whether ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules has increased, decreased or remained equivalent after such a purification as compared to the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules which was present in the sample before purification. The methods of the invention may further be used to assess large scale manufacture of virus, *e.g.* of a viral particle for use in a vaccine or a viral vector, which may be particularly important for the manufacture of viral vectors for gene therapy.

15

20

An increase in viral titre and/or viral component abundance as used herein may be an increase of more than 5, 10, 20, 30, 40, 50, 60, 70, 80 or 90% of the viral titre and/or viral component abundance as to which a measurement is being compared, and a decrease in viral titre and/or viral component abundance as used herein may be a decrease or more than 5, 10, 20, 30, 40, 50, 60, 70, 80 or 90% of the viral titre and/or viral component abundance as to which a measurement is being compared. An equivalent viral titre and/or viral component abundance may be within 5% of the viral titre to which a measurement is being compared.

30

- In this regard, it will be appreciated that for some purposes, it may be desirable to assess viral titre and/or viral component abundance prior to carrying out a method and as well as after and/or during a method, in order to determine whether any change or variation in the viral titre and/or viral component abundance has occurred. The methods of the invention
- 5 may further also include a step of comparison of viral titre and/or viral component abundance *e.g.* with the viral titre and/or viral component abundance within a different sample (at an equivalent or different time point), or within the same sample at a different point in time.
- 10 In a further embodiment of the invention, the methods may be used to determine the extent of viral infection in a subject, *e.g.* to determine whether an infection is being successfully treated or reduced. In such a method, it may be desirable to compare the viral titre and/or viral component abundance in a sample, *e.g.* a sample of the same type from a subject at different time points, to determine whether the ratio of viral nucleic acids to viruses
- 15 comprising the one or more viral structural molecules increases, decreases or remains equivalent over time. Alternatively, or additionally, it may be desirable to compare the viral titre and/or viral component abundance in a sample from an individual with viral titre and/or viral component abundance measurements which have been previously obtained for a condition and which for example may be indicative of the stage of infection and/or the
- 20 prognosis.

Alternatively, the methods of the invention may not determine an actual amount, level or concentration of viral component abundance in a sample, but may determine whether the amount, level or concentration is above or below an acceptable threshold, *e.g.* for a

25 production method, the threshold may determine whether there is an acceptable level of functional viral particles within a sample. As discussed above, the methods of the invention may determine whether the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules has increased, decreased or comparable to those of a previously assayed sample, and thus it will be appreciated that for particular applications, it

30 may not be necessary to determine the actual viral titre and/or viral component abundance (*e.g.* amount or concentration of virus present).

The present invention encompasses the use of Raman spectroscopy to monitor and/or assess viral titre and/or viral component abundance in a sample so that any one of the start, production phase, and end of the viral production process can be identified. It will be appreciated that different amounts or concentrations of virus and/or metabolites will be present in the sample at different stages of production. For example, at the beginning of production viral titre, particularly physical viral titre, may be in the range of $0-10^5$, during active virus production viral titre, particularly physical viral titre, may be in the range of 10^5-10^9 and at the end of viral production, viral titre, particularly physical viral titre, may be in the range of 10^9-10^{12} . However, generally, the monitoring of viral titre and/or viral component abundance over time may identify the different phases of production for a particular virus, in for example, a particular packaging cell line, where increased or peak amounts may be associated with the production phase of the process, early low amounts may be associated with the start of the process and a later plateau in amounts may be associated with the end of the process. As previously discussed, this information can be used for a particular process to ensure that cultures are not maintained after production has plateaued, decreased (*e.g.* by at least 50% as compared to the peak production point) or terminated.

The methods of the present invention may also be used to support adaptive manufacturing and further to increase the viral titre and/or viral component abundance production in a system. In this respect, the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules obtained under different conditions and using different systems can be compared to determine optimal conditions for virus production.

The term “sample” as used herein refers to any sample which contains virus (*e.g.* any sample which comprises a viral vector). The “sample” is preferably a viral culture medium, *i.e.* the liquid in which the virus is being incubated. Accordingly the viral culture medium may be directly irradiated to obtain Raman spectroscopy data for use in the present methods as described further herein. In a preferred embodiment of the invention the sample is from an industrial viral production process or a viral vector manufacturing process. Particularly, in the present invention viral titre in a sample may be measured by Raman spectroscopy in real-time, *in situ* or may be carried out on samples *ex situ*.

By “*in situ*” it is meant that measurements to obtain the intensities of Raman scattered light in a culture capable of producing virus particles are taken from the primary culturing environment in which the virus particles are produced, and not from a sample extracted from the primary culturing environment. Thus, by taking measurements “*in situ*” there are no requirements for liquid handling steps. Thus, removal of a sample from its environment may not be necessary for particular applications of the present invention, and *in situ* measuring of a sample may be preferred. An *in situ* measurement of a sample may allow for regular assessment of viral titre and/or viral component abundance in a sample without the need for an actual sampling step, where a portion of sample, *e.g.* viral culture medium, is removed from the primary culturing environment, *e.g.* the viral growth incubator. Viral titre and/or viral component abundance assessment in this respect can be measured accurately and sensitively in real time without the need for additional steps which could introduce cost and error. As discussed below in further detail, Raman spectroscopy for example provides a probe which can either be placed within or externally to a sample, allowing *in situ* measurements to be taken where desirable. *In situ* measurements are particularly suitable for ‘in line’ process analytical techniques.

Alternatively, the methods of the invention may be carried out on samples *ex situ*. By “*ex situ*” it is meant that measurements to obtain the intensities of Raman scattered light in a culture capable of producing virus particles are taken from aliquots of sample, *e.g.* viral culture medium, extracted from the primary culturing environment in which the virus particles are produced, *e.g.* the viral growth incubator, and are analysed directly. Such *ex situ* measurements are suited to ‘at line’ or retrospective process analytical techniques. Whether measurements are made *in situ* or *ex situ*, measurements may be made directly on the sample, *e.g.* directly on the viral culture medium, without the need for further processing of the sample.

The origin of the sample used in the methods of the invention may be the cell culture in which the virus is being produced. The sample therefore may be one of culture medium (*e.g.* DMEM, MEM or SFII, optionally including serum, L-glutamine and/or other components), which may additionally comprise packaging cells (*e.g.* HEK293 cells), *e.g.* if

taken during a viral production process, or may be a sample of virus for medical use, *e.g.* which requires quality testing, *e.g.* prior to marketing, sale or use. The sample could further be a sample from a subject (*e.g.* a human or mammalian subject) who is suspected of being infected by a virus, *e.g.* a blood, saliva, sputum, plasma, serum, cerebrospinal
5 fluid, urine or faecal sample. Other sources of samples include from open water or public water supplies.

Raman spectroscopy

10 Raman spectroscopy measures changes in the wavenumber of monochromatic light scattered by samples to provide information on their chemical composition, physical state and environment. This is possible because of the way in which the incident light photons interact with the *vibrational modes* that are present in the molecules that comprise the sample. These modes possess specific vibrational frequencies and scattering intensities
15 under a set of given physical conditions and this makes it possible to quantify the amount of a given analyte of interest. Unlike infrared absorption spectroscopy where the absorption of light of different energies from a broadband light source is measured, in Raman spectroscopy the difference in energy of the monochromatic incident light to the scattered light is measured (**Figure 1**); this is known as the *Raman shift*.

20 Typically, the *Stokes* scattered light is monitored as the measured signals are more intense at ambient temperatures. **Figure 2** shows some example spectra; the different peaks represent the presence of different modes of vibration; some bands are overlapped regions of several underlying peaks. For simple mixtures it is possible to identify specific bands
25 that are unique to $n-1$ of the analytes and thus measure changes in their intensity at fixed temperature and pressure to completely quantify the composition; after appropriate calibration. However, for complex systems such as biogenic media, it is not possible to rigorously apply this simple approach, there are too many overlapping bands, and this prevents direct assignment. Instead advanced *chemometric models* that obtain *linear*
30 *combinations* of the variables (wavelengths/wavenumbers) that maximise the covariance with the concentration of the analyte of interest in addition to modelling the original data matrix must be used. The composition of new samples can then be predicted.

Raman spectra provide a “molecular fingerprint”, enabling qualitative and quantitative analysis of samples, for example biological samples. Raman spectra are in general sensitive to changes in physical conditions such as temperature and pH. Often
5 Raman spectra obtained from biological samples can contain background fluorescence signal as frequently such samples contain natural *fluorophores*. In conventional Raman spectroscopy this background should be limited by optimal laser wavenumber selection and any remaining fluorescence removed by using one of several conventionally available algorithms.

10

Raman spectroscopy is a technique known in the art. In the present invention real-time Raman spectroscopy may be used *in-situ* as discussed above, allowing for the continuous measurement of viral titre and/or viral component abundance.

15 “Raman spectroscopy” as used herein may refer to all types of Raman spectroscopy which do not require binding, *e.g.* immuno-interaction, between a substrate and a target molecule of interest (*e.g.* a molecule to be detected by Raman). Binding between a substrate and a target molecule may occur directly, or indirectly using any type of binding molecule, streptavidin/biotin etc., or antibody fragments. An “immuno-interaction” includes the use
20 of antibodies or antibody fragments (*e.g.* scFvs etc) which may be attached to a substrate to specifically bind a molecule of interest in a sample. Particularly, “Raman spectroscopy” as used herein may exclude surface enhanced Raman spectroscopy (SERS), *e.g.* SERS which requires immuno-interaction between a substrate, *e.g.* which may comprise metal nanodots, *e.g.* Au. SERS requires the analyte to be detected to be immobilised on a surface. In one
25 embodiment of the method of the invention, Raman spectroscopy is not carried out to detect analytes in a sample that have been immobilised on a surface. In particular, in one embodiment of the invention Raman spectroscopy is not carried out to detect virus particles in a sample or from a sample that have been immobilised on a surface.

30 SERS is distinct from Raman Spectroscopy according to a preferred embodiment of this invention, in particular SERS requires a specific experimental design to immobilise or bind the analyte of interest to a surface, which leads to an enhanced signal strength using the

SERS methodology. However, such immobilisation of the analyte requires processing of the sample, which may lead to contamination or to interference with the conditions inside a bioreactor, including in situations where the sample is taken from such a system. In general, SERS is more suited to methods involving a ‘simple’ sample comprising the analyte of interest, with few contaminants in the sample, rather than a complex mixture of components such as found in a bioreactor or biological sample for processing viruses according to the methods defined herein. Thus, conventional SERS is not ideally designed for direct monitoring, or in-line or *in situ* monitoring of complex samples, including samples containing viral particles for the assessment of viral titre, but more typically is applicable for the analysis of samples that have been processed and wherein the analyte to be detected has been purified and then immobilised on a surface.

In contrast, the preferred embodiments of the methods of the invention, using Raman spectroscopy as defined herein, can detect analytes, in particular virus particles, in-line/*in situ* in samples without surface attachment of any analyte present in the sample, and in particular without surface attachment or immobilisation of virus particles.

Raman spectroscopy as defined in the present invention particularly includes conventional types of Raman spectroscopy and other types of Raman spectroscopy such as stimulated Raman spectroscopy (SRS), pico Raman, spatially offset Raman (SORS), inverse SORS, see through Raman spectroscopy, coherent anti-Stoke Raman spectroscopy (CARS), coherent Stokes Raman spectroscopy (CSRS), resonance Raman spectroscopy (RR spectroscopy) and total internal reflection Raman spectroscopy (TIR) Raman. Equipment for Raman spectroscopy can be obtained from various suppliers *e.g.* Renishaw, WITec, Horiba, and ThermoFisher Scientific. See also: <http://www.optiqgain.com/>, <https://www.timegate.com/> and <https://www.newport.com/>.

Data processing and multivariate modelling

“Multivariate” data as used herein refers to data where multiple variables are measured for each sample, and a “multivariate model” is a model built using such multivariate data. Raman spectra (*e.g.* generated over a time period from cell culture) comprise multivariate

data, where for each sample or time point measured, intensities at multiple wavenumbers may be recorded.

In the present invention, Raman spectra and the multivariate data that comprise the spectra,
5 resulting from *in situ* monitoring of viral production in culture have been analysed to identify a series of spectral variables which are the most important in enabling model predictions to achieve a measure of real-time viral titre and/or viral component abundance. In particular the model predictions achieve a measure of viral nucleic acid abundance and viral structural molecule abundance in order to assess the ratio of viral nucleic acids to
10 viruses comprising the one or more viral structural molecules in the culture. Plots of variable importance projection (VIP) calculated from a 10 or 12 component multivariate model were created, and the importance of the wavenumber variables established, e.g. as in relation to the data set out in Table 3. Plots of variable importance projection (VIP) calculated from a 8 or 15 component multivariate model were also created, and the
15 importance of the wavenumber variables established e.g. as in relation to the data set out in Table 1. Thus, the inventors used multivariate data from Raman spectral measurements, together with offline data relating to viral titre and viral nucleic acid abundance or viral structural molecule abundance from prior art assays, to identify wavenumber ranges which may be assessed when determining viral titre and/or viral component abundance and to
20 further build a multivariate model which is capable of analysing the intensity of signal at the specified wavenumber ranges from any Raman spectra achieved for a virus containing sample under particular conditions to determine viral titre and/or viral component abundance in order to assess the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules. Any type of viral titre and/or viral component
25 abundance described herein can be determined using the methods of the present invention, provided that the multivariate model used to predict viral component abundance has been built using off-line data relating to the relevant type of viral component abundance.

In complex biological samples the unambiguous assignment of wavenumber ranges to
30 specific analytes is often difficult or impossible. There is also an issue with low concentration of analytes, as signal may be obscured by other high concentration compounds like water or glucose in cell culture media. If the obscuring signals become too

intense the noise associated with them is greater than the variation measured originating from the underlying analyte of interest and the signal intensity of the interesting component falls below the limit of detection. By analysis of Raman spectra, modelling of the data obtained, and calibration of the data with off-line measurement of viral titre and
5 viral nucleic acid abundance or viral structural molecule abundance, the present inventors have identified a series of wavenumber ranges that are correlated with an increase in viral titre and/or viral component abundance in a sample and in particular are correlated with the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules in the sample. As these wavenumber ranges show a consistent and strong correlation over
10 time with viral titre and/or viral component abundance, it is not necessary to assign the ranges to specific analytes. Thus, the problem of a signal being obscured by high concentration compounds is not such an issue for the present methods.

In order to identify the wavenumber ranges for use in the present invention, multivariate
15 model parameters were obtained. These parameters were then used in subsequent analysis to infer response values and to select important variables for use in calculating viral titre and/or viral nucleic acid abundance and viral structural molecule abundance. In this case regression was carried out on Raman data obtained from a virus containing sample.

20 In this particular case the inventors performed regression on pre-processed Raman data. In this regard, it will be appreciated that often raw Raman spectral data acquired from a spectrometer may require correction for several interfering signals, such as background fluorescence and that it is often important to normalise the raw spectra acquired for a sample to correct for gross changes in absolute intensities. Thus, a skilled person would
25 understand that it may be necessary to carry out spectral pre-processing to deal with such issues.

Pre-processing of the Raman spectra obtained may be performed using any one of many algorithms which are available in the scientific literature. Particularly, first derivative,
30 second derivative (Savitzky et al. 1964) and standard normal variate (SNV) normalisation and polynomial background fitting and removal may for example be used. Barnes et al (1989) describe a standard normal variate method. Lieber and Mahadevan-Jansen (2003)

- describe an automated method for fluorescence subtraction from biological Raman spectra, based on a modification to least-squares polynomial curve fitting. Zhao et al. (2007) describe an improved automated algorithm for fluorescence removal based on modified multi-polynomial fitting. Koch et al. (2017) describe a “mollifier”-based baseline
- 5 correction algorithm for pre-processing of Raman spectra. Hu et al. (2018) describe a method for baseline correction based on piecewise polynomial fitting (PPF). Zhang et al. (2010) describe an adaptive iteratively reweighted Penalized Least Squares (airPLS) algorithm method. See also Huang et al (2010).
- 10 Once steps have been taken to pre-process the Raman spectra, if required, parameters for data modelling are obtained. In this regard regression of Raman data, particularly pre-processed Raman data may be carried out based on offline responses obtained using other techniques, such as qPCR and p24 ELISA, plaque assays etc., which can determine viral titre, or CuBiAn and LC-MS which can be used to analyse metabolic markers. Regression
- 15 therefore may involve comparing the pre-processed Raman spectra and the offline data. A typical approach for multivariate regression could employ partial least squares regression (PLS-R).

Specifically, in a standard orthogonal score PLS-R analysis, a linear relationship is sought

20 between the array of Raman spectra X , and a response y (*e.g.* y may be a vector where each element represents a titre or viral component abundance value for each sample:

$$y = \alpha + X\beta + E$$

- 25 where α and β are unknown parameters and E is a matrix of error intensities or residuals). Different basic PLS-R algorithms may be used depending on whether y contains a single response value for each sample or several, where a PLS1 algorithm may be used for a single response value and a PLS2 algorithm may be used for multiple response values. In the case of predicting viral titre and/or viral component abundance from Raman spectra, y
- 30 is typically a univariate parameter for each sample and thus typically PLS1 may be used.

Briefly, a PLS1 algorithm functions as follows. Initially the variables of X and y are mean centred (the mean of each variable may be subtracted from each element in the columns of X and the mean value of y may be subtracted from each element of y. A number of underlying factors, A, may be chosen for the model, which are the factors that can be used in linear combination, to model X. In the first step, X may be projected on y to find the weights; these weights define the direction in the vector/factor space of X that has maximum covariance with y. These weights may then be normalised to have unit length. Subsequently the X scores may be computed by projecting X on these normalised weights. The X-loadings may then be computed by projecting X on the scores. Similarly, the y loadings may be calculated by projecting the transpose of y on these scores. The contribution of the current component may be removed from both X and y by deflating X and y by subtracting the contribution of the given component. This may be carried out by multiplying the component's respective score and loading vectors and subtracting the resulting array or vector from the running X array and y vector respectively. The deflated X and y may then be used again in the same way for each subsequent component in an iterative procedure. *i.e.* the successive determination of weights, scores and loadings until all A components are exhausted and no further deflations are carried out. Reference to this PLS method and NIPALS algorithm can be found, for example, in Wold *et al.* (2001).

During each iteration the calculated weight, score and loading vectors and scalars may be stored sequentially in arrays or vectors of their own *i.e.* for each iteration the relevant vectors or scalars may be placed as new columns or rows in arrays or elements in vectors where the existing vectors or scalars may be those obtained from previous iterations. If regression coefficients are to be used as parameters for subsequent data modelling the regression coefficients, β , may be obtained by multiplying the inverse of the projection of the transpose of the final X-block loading array on the final weights matrix. The optimum number of components may be determined by investigating the prediction error for a test set of pre-processed Raman spectra that were not used in the iterative model building procedure.

Predictions of viral titre and/or viral component abundance may be made as described above using the model parameters, such as regression coefficients, β , and by comparison to offline assay data *e.g.* qPCR/P24 Elisa/Plaque Assay mean squared errors (MSE). The optimal number of underlying components A may be chosen when the MSE of prediction
5 has reached a minimum. For viral nucleic acid abundance off-line assays including qPCR or RT-qPCR can be used. For viral structural molecule abundance off-line assays such as ELISA can be used.

After the building of an initial PLS-R model, *e.g.* using the procedure described above, the
10 most important variables can be selected using one of many available variable selection methods, *e.g.* variable importance projection (VIP) which identifies the variables that may be most important in the prediction of y as well as explaining variance in X . VIP may generate a VIP vector of the same length as the rows of X , *i.e.* the VIP vector contains an element corresponding to each variable of X , where the numerical value of each element
15 may be a measure of the importance of that variable. A common approach is to refine the initial model above, by rebuilding it with only the variables that are most important as determined by VIP or other variable selection method. To determine which variables are important a threshold approach is chosen. Typically, the VIP threshold may be set to 1, as this is the mean value of the VIP parameter, but a skilled person will appreciate that this is
20 intrinsically arbitrary, and that other higher thresholds can be chosen, *e.g.* 1.5.

As discussed below, the wavenumber ranges identified in the present invention as being of importance for the determination of viral titre and/or viral component abundance, are based on setting the VIP parameter to at least 1.00 or higher. Thus, in the present invention,
25 wavenumber ranges which did not generate a peak intensity of greater than 1.00 at this stage were excluded.

A skilled person will appreciate that the PLS1 algorithm may be run again after selection of the VIP parameter but, in this instance, the variables of X below the VIP threshold may
30 be removed, generating a new multivariate model, with shorter loading vectors and a shorter β vector of regression coefficients.

The wavenumber ranges set out below result from conducting Raman spectroscopy with a laser at wavelength 785nm. It is encompassed by the present invention that lasers of different wavelengths can be used, other than 785nm. The wavenumber ranges obtained using lasers at different wavelengths would be the same due to the Raman shift (*i.e.* the difference in the wavelength of the inelastically scattered Raman light from the monochromatic laser beam which is used to induce the Raman scattering) being largely independent of wavelength.

“Wavenumber” ($\tilde{\nu}$) as used herein is the number of wavelengths per unit distance and is measured in cm^{-1} . Typically, $\tilde{\nu}=1/\lambda$, where λ is wavelength.

In a preferred embodiment of the invention, peaks over 1.5 (variable importance projection (VIP)) indicate wavenumber ranges likely to be important for the determination of the presence of virus and the determination of viral titre and/or viral component abundance. Determining the signal intensity of specific wavenumber ranges is essential to predict viral titre by the methods of the invention. The 1.5 VIP threshold is used in an embodiment of the invention to determine which wavenumber ranges are important for the prediction of viral titre and/or viral component abundance. Thus, in a preferred embodiment, the present invention encompasses a method of assessing or predicting viral titre and/or viral component abundance using Raman spectroscopy comprising a step of determining from a Raman spectrum obtained from said sample, the intensity of signal at five or more of these wavenumber ranges. The measured pre-processed signals are used for the predictions. Any subsequent spectra *i.e.* those after model building, from which predictions are to be made require pre-processing using the exact same methods as the data used to train and build the PLS model.

Any of the methods of the present invention involve the steps of measuring the total intensity of Raman scattered light within each one of a plurality of wavenumber ranges to obtain a wavenumber intensity data set for the sample, wherein the plurality of wavenumber ranges are pre-selected and are characteristic of the viral components in the sample.

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 1 to 12 as listed in Table 1 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 1 to 12 as listed in Table 1 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 1 to 12 as listed in Table 1 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 1 to 12 as listed in Table 1 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 12 of the wavenumber ranges 1 to 12 as listed in Table 1 below and wherein the VIP is ≥ 1.00 .

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 13 to 22 as listed in Table 1 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 13 to 22 as listed in Table 1 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 13 to 22 as listed in Table 1 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 10 of the wavenumber ranges 13 to 22 as listed in Table 1 below and wherein the VIP is ≥ 1.25 .

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 23 to 30 as listed in Table 1 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 23 to 30 as listed in Table 1 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 8 of the wavenumber ranges 23 to 30 as listed in Table 1 below and wherein the VIP is ≥ 1.50 .

Table 1

	AAV vector production								
	VIP ≥ 1.00			VIP ≥ 1.25			VIP ≥ 1.50		
	#	From:	To:	#	From:	To:	#	From:	To:
Wavenumber Ranges / cm^{-1}	1	420	420	13	512	515	23	848	861
	2	510	517	14	846	862	24	994	1035
	3	844	863	15	993	1036	25	1119	1129
	4	992	1037	16	1060	1066	26	1355	1363
	5	1057	1069	17	1115	1134	27	1425	1431
	6	1112	1137	18	1352	1376	28	1597	1608
	7	1182	1184	19	1415	1455	29	1638	1644
	8	1193	1199	20	1596	1611	30	1652	1658
	9	1333	1380	21	1626	1626			
	10	1410	1461	22	1635	1678			
	11	1583	1586						
	12	1594	1692						

- 5 Viral titre and/or viral component abundance may be measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 1. In a preferred embodiment of the invention, the viral titre and/or viral component abundance measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 1 is adeno associated virus (AAV) titre. In a particularly
- 10 preferred embodiment of the invention, the viral titre and/or viral component abundance measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 1 is adeno associated virus serotype 8 (AAV8) titre.
- In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 1 to 20
- 15 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 6 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 8 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the

VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 12 or more, 14 or more, 16 or
5 more or 18 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 20 of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 .

10 In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein
15 the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 8 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more, 11 or more or 12 of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber
20 ranges in the Raman spectrum which are measured comprises all 13 of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 .

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 4 or more of the wavenumber ranges 34 to 40
25 as listed in Table 2 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 5 or 6 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured may
30 comprise all 7 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 .

Table 2

	AAV vector production CAPSID ELISA								
	VIP >= 1.00			VIP >= 1.25			VIP >= 1.50		
	#	From:	To:	#	From:	To:	#	From:	To:
Wavenumber Ranges / cm⁻¹	1	420	426	21	422	423	34	845	862
	2	447	448	22	843	864	35	995	1010
	3	514	519	23	994	1019	36	1028	1034
	4	824	833	24	1026	1035	37	1060	1066
	5	838	866	25	1057	1069	38	1113	1135
	6	879	884	26	1110	1137	39	1535	1539
	7	993	1037	27	1356	1362	40	1599	1607
	8	1055	1074	28	1415	1420			
	9	1107	1140	29	1450	1452			
	10	1332	1338	30	1530	1543			
	11	1350	1376	31	1598	1607			
	12	1412	1429	32	1675	1675			
	13	1438	1441	33	1689	1690			
	14	1445	1464						
	15	1471	1475						
	16	1486	1506						
	17	1513	1546						
	18	1558	1562						
	19	1597	1609						
	20	1671	1703						

- 5 Viral titre and/or viral component abundance may be measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 2. In a preferred embodiment of the invention, the viral titre and/or viral component abundance measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 2 is adeno associated virus (AAV) titre. In a particularly
- 10 preferred embodiment of the invention, the viral titre and/or viral component abundance measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 2 is adeno associated virus serotype 8 (AAV8) titre.

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 5 or more of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 15 or more of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 20 or more of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 25 or more of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 28 of the wavenumber ranges 1 to 28 as listed in Table 3 below and wherein the VIP is ≥ 1.00 .

In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 5 or more of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 15 or more of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 20 or more of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 25 or more of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 30 of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 31 of the wavenumber ranges 29 to 59 as listed in Table 3 below and wherein the VIP is ≥ 1.25 .

- In any of the methods of the invention, the plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 5 or more of the wavenumber ranges 60 to 81 as listed in Table 3 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 10 or more of the
- 5 wavenumber ranges 60 to 81 as listed in Table 3 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 15 or more of the wavenumber ranges 60 to 81 as listed in Table 3 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise 20 or more of the wavenumber ranges 60 to 81 as listed in Table 3
- 10 below and wherein the VIP is ≥ 1.50 . The plurality of wavenumber ranges in the Raman spectrum which are measured may comprise all 22 of the wavenumber ranges 60 to 81 as listed in Table 3 below and wherein the VIP is ≥ 1.50 .

Table 3

15

Lentiviral vector production									
	VIP ≥ 1.00			VIP ≥ 1.25			VIP ≥ 1.50		
	#	From:	To:	#	From:	To:	#	From:	To:
Wavenumber Ranges / cm^{-1}	1	420	438	29	420	421	60	420	420
	2	457	497	30	426	429	61	467	471
	3	503	552	31	434	436	62	474	481
	4	576	580	32	459	486	63	505	529
	5	588	589	33	490	496	64	537	543
	6	604	608	34	504	549	65	836	884
	7	617	621	35	798	800	66	897	902
	8	796	805	36	834	885	67	919	937
	9	808	809	37	892	907	68	995	1043
	10	824	911	38	919	938	69	1046	1046
	11	918	939	39	973	973	70	1049	1071
	12	971	1168	40	981	983	71	1084	1144
	13	1191	1197	41	990	1145	72	1209	1210
	14	1206	1212	42	1207	1211	73	1271	1273
	15	1234	1237	43	1248	1250	74	1277	1302
	16	1246	1252	44	1270	1322	75	1347	1366
	17	1259	1481	45	1328	1331	76	1386	1433
	18	1497	1500	46	1346	1380	77	1444	1461

19	1526	1540	47	1383	1473	78	1467	1469
20	1545	1550	48	1476	1478	79	1610	1612
21	1584	1591	49	1498	1499	80	1629	1630
22	1598	1685	50	1528	1528	81	1655	1671
23	1699	1699	51	1590	1590			
24	1717	1719	52	1599	1602			
25	1754	1754	53	1609	1613			
26	1768	1771	54	1616	1620			
27	1782	1783	55	1628	1634			
28	1798	1800	56	1640	1672			
			57	1678	1679			
			58	1769	1769			
			59	1800	1800			

Viral titre and/or viral component abundance may be measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 3. In a preferred embodiment of the invention, the viral titre and/or viral component abundance

5 measured using the plurality of wavenumber ranges in the Raman spectrum as described above in relation to Table 3 is lentiviral titre and/or viral component abundance.

In another preferred embodiment of the invention, peaks over 1.5 (variable importance projection (VIP)) indicate wavenumber ranges likely to be important for the determination

10 of the presence of virus and the determination of viral titre and/or viral component abundance. Determining the signal intensity of specific wavenumber ranges is essential to predict viral titre and/or viral component abundance by the methods of the invention. The 1.5 VIP threshold is used in an embodiment of the invention to determine which wavenumber ranges are important for the prediction of viral titre and/or viral component

15 abundance. Thus, in a preferred embodiment, the present invention encompasses a method of assessing or predicting viral titre and/or viral component abundance using Raman spectroscopy comprising a step of determining from a Raman spectrum obtained from said sample, the intensity of signal at four or more of these wavenumber ranges. The measured pre-processed signals are used for the predictions. Any subsequent spectra *i.e.*

20 those after model building, from which predictions are to be made require pre-processing using the exact same methods as the data used to train and build the PLS model.

Any of the methods of the present invention involve the steps of measuring the total intensity of Raman scattered light within each one of a plurality of wavenumber ranges to obtain a wavenumber intensity data set for the sample, wherein the plurality of wavenumber ranges are pre-selected and are characteristic of the viral components in the sample.

The step of determining the intensity of signal at each desired wavenumber range requires a determination of the level of any peak identified within the desired wavenumber range. It will be appreciated by a skilled person that the intensity of signal within any wavenumber range deemed to be associated with viral titre and/or viral component abundance, as set out above, may be at any level, and that the measurement of such intensities when analysed with an appropriate multivariate model will allow the determination of viral titre and/or viral component abundance. As discussed in further detail below, the present invention may therefore include a step of assessing or calculating viral titre and/or viral component abundance by analysing the signal intensities measured using a multivariate model. Such a multivariate model may be prepared in advance of carrying out the present invention or alternatively as part of the methods of the invention, where the methods may additionally comprise a step of building a multivariate model.

The methods of the invention may comprise determining the signal intensity at further wavenumber ranges in addition to the wavelength ranges specified herein.

Neural networks may also be used for classifications based on Raman spectra, for example in analysing diseased tissue vs healthy tissue in pathology. Neural networks may also be used for regression problems, like those faced in applying Raman data for the monitoring of viral production, as described herein.

More sophisticated approaches utilize what are referred to as convolutional neural networks (CNN) (Deep Learning), often using Google's TensorFlow backend and the Keras API for scripting in the object-oriented Python programming language. The advantage of using the convolutional layers is that pre-processing becomes less and less necessary as the network essentially "learns" the perfect way to pre-process the spectra

themselves for optimal titre/concentration predictions. Such neural networks for use in the data processing steps described herein are well known to persons skilled in the art. Further information can be found *e.g.* here:

- 5 <https://www.forbes.com/sites/bernardmarr/2018/09/24/what-are-artificial-neural-networks-a-simple-explanation-for-absolutely-anyone/#2c9442a91245>

<http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>

- 10 As discussed above, multivariate data model parameters may be obtained and used in methods for the quantification of viral titre and/or viral component abundance as defined herein. Such model parameters may also be used in building alternative multivariate data models as defined herein if required. The present inventors have applied a multivariate algorithm to Raman spectral wavenumber signal intensity data to obtain model parameters
- 15 which are then used in the quantification of viral titre and/or viral component abundance. Specifically, the inventors have applied a multivariate algorithm to obtain regression coefficients which are used as the model parameters. The skilled person will however appreciate that alternative model parameters may be obtained and used depending upon the nature of the model selected. Thus, in any of the methods defined herein multivariate data
- 20 model parameters may be appropriately selected and may optionally comprise regression coefficients. Any suitable multivariate algorithm may be applied to Raman spectral wavenumber signal intensity data to obtain model parameters. A multivariate regression algorithm may be used, such as a partial least squares (PLS) regression algorithm, optionally wherein the PLS algorithm is a nonlinear iterative partial least squares
- 25 (NIPALS) regression algorithm. An algorithm involving a neural network may also be used to obtain model parameters.

- The skilled person would also understand that the methods of the invention could encompass additional mathematical data processing and modelling steps to quantify viral
- 30 components of interest.

Model to estimate viral titre

Chemometric modelling of Raman spectra was carried out as described herein to identify a correlation between increases in real-time viral titre and/or viral component abundance and the identity and intensity of wavenumber ranges seen in Raman spectra. The wavenumber ranges identified are described above.

As discussed previously, the present invention requires the assessment of signal intensities at 4 or 5 or more of the wavenumber ranges determined to be of importance in the assessment of viral titre and/or viral component abundance, and the further analysis of the intensities using a multivariate model (either a calibrated or non-calibrated multivariate model) which has been built.

A skilled person will appreciate that different multivariate models may be built depending on the samples to be analysed, and methods for building of multivariate models are well known in the art (*e.g.* see references cited herein). Thus, different multivariate models may be required for the determination of viral titre and/or viral component abundance in samples which comprise different types of virus, different cell culture media or different producer cells, for example.

In one embodiment of the invention, a multivariate model can be built using the following approach:

- i) **Regression of pre-processed Raman data on offline responses** obtained using other techniques such as qPCR and p24 ELISA, Plaque Assay etc, as discussed above, regression may involve comparing pre-processed Raman spectra,
- ii) **Using the regression coefficients obtained** to predict the response values using the pre-processed data, where the quality of these predictions can be optimised by adjusting the underlying number of components/factors used for the multivariate regression,

- iii) **Performing Variable Selection** using any known methods, *e.g.* variable importance projection (VIP) which identifies variables that are powerful/important for predicting Y in addition to explaining X,
- 5 iv) **Building a Refined Model where** following the identification of important spectral variables, a further round of modelling may be performed using the same approach as described in step (ii) but where the variables or columns in the array of pre-processed Raman spectra that were deemed irrelevant by
- 10 variable selection are removed before the model is built. This results in a simpler model built on data with much of the irrelevant variation removed. As in (ii) the number of underlying components may be optimised by selecting the model built with the fewest number of underlying factors with, give component to component variation, the lowest error of prediction.
- 15 v) **Making Future predictions** using the regression coefficients used determined in (iv), where new pre-processed Raman spectra may be multiplied by the regression coefficients obtained to generate the estimate.

It will be appreciated that it may not be necessary to repeat steps (i) to (iii) as set out above once wavenumber ranges for analysis have been identified. Particularly, building a model may only require step iv) in this instance. Furthermore, data modelling parameters other than regression coefficients may be used.

20

In a further aspect of the invention, the methods may comprise an additional step of preparing or building a multivariate model.

25

As set out in step v) above, the regression coefficients from the multivariate models generated may be used to obtain an estimate for viral titre and/or viral component abundance from Raman spectra obtained from one or more samples. Thus, in the present invention, the method may include a step of determining viral titre and/or viral component abundance using the regression coefficients from a multivariate model. The same pre-processing methods used for the training/ building of the model.

30

The present invention is further illustrated by the following examples which should not be construed as further limiting. The contents of all figures and all references, patents and published patent applications cited throughout this application are expressly incorporated
 5 herein by reference.

EXAMPLES

EXAMPLE 1 - CALCULATION OF CONCENTRATION OF LENTIVIRUS AT 10 **DIFFERENT STAGES DURING THE PRODUCTION PROCESS**

Calculations were performed to provide estimated lentiviral concentrations in mg/ml expected at different stages during an example viral production process. The concentrations were calculated using the known buoyant density of lentivirus and the
 15 physical titre. The results are shown in the table below.

<u>Parameter</u>	<u>Value</u>	<u>Units</u>	<u>Comments</u>
Buoyant Density	1.15E+00	g/cm ³	Actual range is 1.15 - 1.19 g/ml using the lower value for conservative estimate.
Particle Radius	7.50E-08	m	
Particle Diameter	1.50E-07	m	Diameter of a lentivirus is approximately 0.15-0.2 microns, using the lower value for a conservative estimate.
Volume of a Particle	1.77E-21	m ³	$\frac{4}{3} \times \pi \times r^3$
Volume of a particle	1.77E-15	cm ³	
Mass of a Particle	2.03E-15	g	

Typical Physical Titre (a)	1.00E+08	particles / ml	physical titres 100-1000 x greater than copy number from qPCR.
Typical Physical Titre (b)	1.00E+07	particles / ml	
Typical Physical Titre (c)	1.00E+09	particles / ml	
Typical Physical Titre (d)	1.00E+10	particles / ml	
Concentration (a)	2.03E-07	g / ml	
Concentration (b)	2.03E-08	g / ml	
Concentration (c)	2.03E-06	g/ ml	
Concentration (d)	2.03E-05	g / ml	
Concentration (a)	2.03E-04	mg/ml	
Concentration (b)	2.03E-05	mg/ml	
Concentration (c)	2.03E-03	mg/ml	
Concentration (d)	2.03E-02	mg/ml	

As shown, estimated lentiviral concentrations were found to be in the range of 2.03E-02 to 2.03E-05 mg/ml.

- 5 For comparison purposes, shown below are some instructive calculations based on the limit of detection for glucose and phenylalanine as taken from Buckley and Ryder (2017, Applied Spectroscopy, 71, p 1085-1116).

Glucose

10

mw = 180.156 g / mol

limit of detection 0.6 mM (see Buckley and Ryder)

moles = volume / 1000 * molarity
1/1000 * 0.0006 - in 1 ml
= 0.0000006 moles
mass = moles * mw
5 = 0.0000006*180.156
= 0.00011 g / ml
Estimated concentration = 0.11 mg / ml.

Phenylalanine
10
mw = 165.19 g / mol
LoD = 1.1 mM (See Buckley and Ryder paper)
moles = volume / 1000 * molarity
1/1000 * 0.0011 - in 1 ml
15 = 0.0000011 moles
mass = moles * mw
= 0.0000011 moles * 165.19 g / moles
= 0.00018 g
Estimated concentration = 0.18 mg/ml.

20
The concentrations in mg/ml for the limits of detection for glucose and phenylalanine are
5-10 times higher than the optimistic estimated concentrations commensurate with the
conservative physical titres.

25 Based on the above, it would be expected that the approximate concentration in mg/mL of
lentivirus in the culture medium would be below what would be considered the limit of
detection using Raman spectroscopy.

EXAMPLE 2 - LENTIVIRAL PRODUCTION FROM A HEK 293 TRANSIENT PROCESS

Experimental Methods

5

Cell culture and Transient Transfection

HEK293 cultures were expanded in Eppendorf DASbox BioBLU 300 bioreactors in FreeStyle 293 expression medium (ThermoFisher) with no additional supplements at 37°C.

- 10 The cells were agitated and were expanded for 2 days prior to transient transfection to produce lentivirus. The cells were transfected with gag-pol, vsv-g and genome encoding eGFP to produce LV particles using PEIPro from Polyplus.

- Throughout the process 10 or 12 samples were acquired from each bioreactor to measure
15 viral titre using qRT-PCR and confirmed by P24 ELISA. Raman spectra were acquired throughout the expansion and viral production phases.

RT-qPCR

- 20 PCR kit Used: Lenti-X™ qRT-PCR Titration Kit (by Takara, Cat # 631235)
Suppliers: Clontech
Method of action: the kit is a one-step reverse transcription and PCR amplification kit. The primers of this kit target a conserved region of the HIV-1 genome adjacent to the packaging signal. Amplicons are detected by SYBR green fluorescence and the final titre
25 determined from a ssRNA standard used to generate the standard curve. Final quantification of virus titre is provided as viral genomes/ml.

P24-ELISA

- 30 ELISA kit: QuickTiter™ Lentivirus Titer Kit (Lentivirus-Associated HIV p24) (Cat # VPK-107)
Suppliers: Cell Biolabs, Inc

Method of action: the kit is an enzyme immunoassay developed for detection and quantification of the lentivirus associated HIV-1 p24 core proteins only. Virus associated p24 can be quantified as p24 titre (ng/ml) or as particles/ml with the assumption there are approximately 2000 molecules of p24 per lentiviral particle.

5

Raman Spectroscopy

Raman measurements were performed using a Kasier Optics RxN2 Raman spectrometer. This spectrometer has the capacity to monitor 4 probe channels sequentially. The RxN2
10 excitation source was a 785 nm near infrared diode laser with a nominal power output of ~270 mW at each probe head. The samples comprised the contents of four Eppendorf, dasBox BioBLU single use systems. The beam was delivered to each sample bioreactor using four Kaiser Optics filtered fibre optic MR probes and BioOptic 220's – one set for each bioreactor. Prior to in-process measurements, the RxN2 system was stabilised for 1
15 hour and then each of the 4 probe channels was calibrated using the RxN2's internal auto-calibration standards, in addition, a CCD sensitivity correction was performed on each probe channel using a National Institute of Standards and Technology (NIST) certified light source (HCA). The scattered light was collected using the same BioOptic 220's and MR probes as those used for beam delivery. Within each MR probe the scattered light was
20 delivered via a second fibre optic to the RxN2 f\1.8 imaging spectrograph. After filtering Rayleigh scattered light using a holographic notch filter, the Raman scattered light was directed to a Kaiser Optics holographic transmission grating and then imaged onto the thermoelectrically cooled 1024 pixel CCD detector. The system has an effective bandwidth of 100 -3425 cm^{-1} and resolution of 4 cm^{-1} . Raman spectra were acquired from 100-3425
25 cm^{-1} with an integration time of ~15 minutes/channel including CCD readout time, 10 s acquisitions were averaged over 75 accumulations to generate each measured spectrum. Each channel was measured in turn. At different times throughout the processes, liquid samples were obtained from each bioreactor and the time point noted to enable the *post hoc* matching of the offline assay data to the commensurate Raman spectra.

30

Raman Data Analysis

All data analysis was performed in MATLAB (The MathWorks, MA, USA) version R2017b. Raw Raman spectra were pre-processed by normalising the entire spectrum to the peak intensity of the water band at $\sim 3000\text{ cm}^{-1}$. The moderate fluorescence background signal was removed for the region of $420\text{--}1800\text{ cm}^{-1}$. The low end of this range was selected to avoid Raman bands that could originate from the sapphire window of the BioOptic-220 or be artefacts of the optical design of the Raman instrument and probes. The reduced normalised spectra were then inspected for obvious outliers and artefacts. The spectra associated with the offline sampling time points were identified and a model training subset of pre-processed spectra created. The training set of pre-processed spectra were then used for chemometric modelling. The spectra were mean-centered prior to chemometric modelling. Several initial projections to latent structures - regression (PLS-R) models for critical analytes and viral titre were built. These models allow you to regress multivariate Raman spectra against samples containing known concentrations of interesting analytes (viral titre). Based on these calculations the concentration of the analytes can be predicted in future. The models were prepared using a 10-fold cross validation procedure on the training data, *i.e.* $1/10^{\text{th}}$ of the data was randomly selected and removed from the training data and used to assess model performance, this was done 10 times and the error values, model accuracy/performance statistics are the averages obtained for each of the 10-fold training sets. Choosing the number of underlying components or basis vectors is an important step in building supervised linear models such as PLS-R. In this work the optimal number of underlying components was identified by examining plots of the mean squared error of prediction after cross-validation (MSECV) as a function of component number; a minimum identifies the optimal number of PLS components. A second stage of variable selection is required to optimise the models built by choosing only wavenumbers/variables that are most significant for prediction. This was carried out using the Variable Importance Projection (VIP) method. However, many methods of conducting variables selection exist. A typical VIP plot is shown in **Figure 5**. Typically, variables with VIP values greater than 1 are used for the final model. However, here we have built and assessed models using several VIP thresholds to identify the minimum number of spectral variables required to make good predictions and the threshold at which one ceases to be

able to model the offline RT-qPCR data. As the VIP threshold is increased the number of spectral variables identified decreases. Once the significant variables were determined for each VIP threshold, final models were built. Subsequently these models were used to predict the intermediate viral titre values, *i.e.* those between each offline data point for all
5 available runs.

Preliminary viral titre model evaluation and range selection

Example, pre-processed Raman spectra as used for chemometric modelling are shown in
10 **Figure 2.**

Viral titre was monitored throughout the project, a representative titre obtained by RT-qPCR is summarised in **Figure 3.**

15 A Plot of the mean squared error of prediction after cross-validation for the initial PLS-R model is shown in **Figure 4.** When using all spectral variables or channels, the minimal effective prediction error was found to occur when 12 PLS components were used.

From this plot (**Figure 4**) it therefore can be concluded that the best compromise between
20 prediction error minimization and model simplicity lies in a 12-component model. A small improvement in predictive power could be obtained by increasing the number of components but this is likely just the result of incorporating noise and overfitting the model to the training set. That is, building a model that is only predictive for the training data and not new unseen data and is based on spurious correlations between the measured variables
25 and the dependent/response variables.

Following preparation of the initial 12 component model different variable selection methods were evaluated to select the optimal/most predictive spectral variables for the final model. The aim here was to remove unnecessary spectral channels/variables from the
30 model to enhance its parsimony and only include physically meaningful information. The variable importance projection (VIP) was finally calculated to determine which spectral variables have the greatest importance in predicting the viral copy number (**Figure 5A**). To

assess and identify the minimum number of spectral variables required to make acceptable physical titre predictions, several variable importance thresholds were investigated as the criterion for retained variables; generally, a VIP threshold of 1 is used – thresholds of 1.00 – 1.75 were investigated. **Figure 5B** shows variable or wavenumber ranges that the VIP
5 algorithm identifies as regions considered most important *i.e.* those greater than a selected threshold, in this case 1.5, **Figure 5C** shows these wavenumber ranges in order of importance.

After the number of spectral variables was reduced a further assessment of the number of
10 underlying latent variables was carried out. The optimal number of PLS components can vary with the number of spectral variables or wavenumbers that are used in the model. After spectral variable reduction using a threshold of 1.5 the optimum number of PLS latent variables was found to be 10. **Figure 6** shows the MSECv plot for the refined models with different numbers of underlying components. The fact that the mean squared
15 error of prediction increased with larger numbers of underlying components indicates that where more than 10 PLS components were included the models produced were overfitting the training set.

Model predictions of RT-qPCR viral copy number for each of the 3 runs estimated using
20 the regression coefficients obtained from the 10-latent variable and $VIP \geq 1.5$ selected spectral variable (conservative) model are shown below in **Figures 7, 8 and 9** respectively. The results show that the model using the Raman spectroscopy data is consistent with offline measurements of viral titre over time. A comparison of the titres obtained from the RT-qPCR assay and P24 ELISA are shown in **Figure 10**.

25 **Figure 11** describes how the methods of the present invention can be used to monitor the stage of a viral production culture. As Raman spectroscopy is used in real-time, in a continuous manner, in the methods of the invention changes in the production rate of virus can be accurately followed. Thus, the change from start to production phase, and
30 production phase to end phase, can be identified.

EXAMPLE 3 - REFINED VIRAL TITRE MODEL EVALUATION AND RANGE SELECTION

Further studies using additional samples were performed using the same methodologies as described above for Example 2 in order to further refine the wavenumber range selection for viral titre evaluation. Using this approach various wavenumber ranges were identified for use in calculating viral titre by applying a variable importance projection (VIP) threshold of ≥ 1.00 . Additional wavenumber ranges were identified by applying a variable importance projection (VIP) threshold of ≥ 1.25 and further wavenumber ranges were identified for by applying a variable importance projection (VIP) threshold of ≥ 1.50 . The results are presented in Table 3 above.

Figure 12 shows an outline schematic of a formula for quantifying viral titre. The formula is applied to the regression coefficients which are obtained from the multivariate regression algorithm which was applied to normalised Raman signal intensity data.

A further analysis was performed to analyse the number of wavenumber ranges which can be used to provide an accurate estimate of viral titre.

The ranges identified as important for viral vector production, *i.e.* the ranges identified as important by variable importance projection (VIP) ≥ 1.00 after initial PLS modelling using the extended spectral range ($\sim 420 - 1800 \text{ cm}^{-1}$), were identified (*i.e.* wavenumber ranges 1 to 28 as listed in Table 3 above) and further analysis was performed.

The data were split into randomly selected paired blocks of training and test data in a 4:1 ratio, that is Raman spectra and their associated offline viral titre data for model building (80%) and model testing (20%).

Different combinations of the ranges deemed important by VIP were evaluated stochastically for the different training and test pairs, *i.e.* for each r total number of ranges 1-28, many combinations were evaluated based on the model performance R^2 statistic (n.b. $R^2 = 1 - \text{residual sum of squares} / \text{total sum of squares}$) and the standard deviations of the

different models' performances was evaluated to generate the confidence intervals. The minimum number of ranges was identified by choosing the number of ranges where the mean of mean R^2 values for several training/test pairs of data was approximately 0.5.

Figure 13 shows a plot of R^2 as a function of the number of wavenumber ranges.

5

This analysis identified five as being the minimum number of wavenumber ranges which are required to provide an estimate of viral titre.

Thus, in any of the methods of the invention, 5 or more of wavenumber ranges 1 to 28 as presented in Table 3 identified at a VIP threshold of ≥ 1.00 may be used to calculate viral titre, as described in more detail herein. In any of the methods of the invention, preferably 5 or more of wavenumber ranges 29 to 59 as presented in Table 3 identified at a VIP threshold of ≥ 1.25 may be used to calculate viral titre, as described in more detail herein. In any of the methods of the invention, more preferably 5 or more of wavenumber ranges 60 to 81 as presented in Table 3 identified at a VIP threshold of ≥ 1.50 may be used to calculate viral titre, as described in more detail herein. In any of these methods, preferably 10 or more of the wavenumber ranges may be used to calculate viral titre as described in more detail herein, more preferably 15 or more, or yet more preferably 20 or more.

20 **EXAMPLE 4 – AAV8 PRODUCTION FROM A HEK 293 TRANSIENT PROCESS**

Experimental Methods

Cell culture and Transient Transfection

25

HEK 293F cultures were expanded in Eppendorf DASbox BioBLU 300 bioreactors in BalanCD media (Irvine Scientific) with 4 mM GlutaMAX (Fisher) at 37°C. The cells were agitated and were expanded for 24 hours prior to transient transfection to produce AAV8. The cells were transfected with rep, cap, genome encoded eGFP plasmids and helper plasmid (E2A, E4) in serum-free Opti-MEM (Gibco) to produce AAV8 particles with PEIPro (Polyplus transfection).

30

Throughout the process 12 samples were acquired from each bioreactor to measure viral titre using qRT-PCR. Raman spectra were acquired throughout the expansion and viral production phases.

5 *RT-qPCR*

Viral titre of AAV8 samples was measured using TaqMan™ based real-time qPCR, with final quantification provided as viral genome/mL (VG/mL). The primers of the assay targeted the *ITR2* sequences in the AAV8 viral genome. Amplicons were detected by
10 TaqMan™ fluorogenic probe. Viral titre was determined from a standard curve generated from a linearised plasmid.

Raman Spectroscopy

15 Raman measurements were performed using a Kasier Optics RxN2 Raman spectrometer. This spectrometer has the capacity to monitor 4 probe channels sequentially. The RxN2 excitation source was a 785 nm near infrared diode laser with a nominal power output of ~270 mW at each probe head. The samples comprised the contents of four Eppendorf, dasBox BioBLU single use systems. The beam was delivered to each sample bioreactor
20 using four Kaiser Optics filtered fibre optic MR probes and BioOptic 220's – one set for each bioreactor. Prior to in-process measurements, the RxN2 system was stabilised for 1 hour and then each of the 4 probe channels was calibrated using the RxN2's internal auto-calibration standards, in addition, a CCD sensitivity correction was performed on each
25 probe channel using a National Institute of Standards and Technology (NIST) certified light source (HCA). The scattered light was collected using the same BioOptic 220's and MR probes as those used for beam delivery. Within each MR probe the scattered light was delivered via a second fibre optic to the RxN2 f1.8 imaging spectrograph. After filtering Rayleigh scattered light using a holographic notch filter, the Raman scattered light was directed to a Kaiser Optics holographic transmission grating and then imaged onto the
30 thermoelectrically cooled 1024 pixel CCD detector. The system has an effective bandwidth of 100 -3425 cm⁻¹ and resolution of 4 cm⁻¹. Raman spectra were acquired from 100-3425 cm⁻¹ with an integration time of ~15 minutes/channel including CCD readout time, 10 s acquisitions were averaged over 75 accumulations to generate each measured spectrum.

Each channel was measured in turn. At different times throughout the processes, liquid samples were obtained from each bioreactor and the time point noted to enable the *post hoc* matching of the offline assay data to the commensurate Raman spectra.

5 *Raman Data Analysis*

All data analysis was performed in MATLAB (The MathWorks, MA, USA) version R2019b. Raw Raman spectra were pre-processed by normalising the entire spectrum to the peak intensity of the water band at $\sim 3000\text{ cm}^{-1}$. The moderate fluorescence background signal was removed for the region of $420\text{--}1800\text{ cm}^{-1}$. The low end of this range was selected to avoid Raman bands that could originate from the sapphire window of the BioOptic-220 or be artefacts of the optical design of the Raman instrument and probes. The reduced normalised spectra were then inspected for obvious outliers and artefacts. The spectra associated with the offline sampling time points were identified and a model training subset of pre-processed spectra created. The training set of pre-processed spectra were then used for chemometric modelling. The spectra were mean-centered prior to chemometric modelling. Several initial projections to latent structures - regression (PLS-R) models for critical analytes and viral titre were built. These models allow you to regress multivariate Raman spectra against samples containing known concentrations of interesting analytes (viral titre). Based on these calculations the concentration of the analytes can be predicted in future. The models were prepared using a 10-fold cross validation procedure on the training data, *i.e.* $1/10^{\text{th}}$ of the data was randomly selected and removed from the training data and used to assess model performance, this was done 10 times and the error values, model accuracy/performance statistics are the averages obtained for each of the 10-fold training sets. Choosing the number of underlying components or basis vectors is an important step in building supervised linear models such as PLS-R. In this work the optimal number of underlying components was identified by examining plots of the mean squared error of prediction after cross-validation (MSECV) as a function of component number; a minimum identifies the optimal number of PLS components. A second stage of variable selection is required to optimise the models built by choosing only wavenumbers/variables that are most significant for prediction. This was carried out using the Variable Importance Projection (VIP) method. However, many methods of conducting

variables selection exist. A typical VIP plot is shown in **Figure 18**. Typically, variables with VIP values greater than 1 are used for the final model. However, here we have built and assessed models using several VIP thresholds to identify the minimum number of spectral variables required to make good predictions and the threshold at which one ceases to be able to model the offline RT-qPCR data. As the VIP threshold is increased the number of spectral variables identified decreases. Once the significant variables were determined for each VIP threshold, final models were built. Subsequently these models were used to predict the intermediate viral titre values, *i.e.* those between each offline data point for all available runs.

10

Preliminary viral titre model evaluation and range selection.

Example, pre-processed Raman spectra as used for the chemometric modelling of AAV are shown in **Figure 14**.

15

AAV titre was monitored throughout the project, a representative titre obtained by RT-qPCR is summarised in **Figure 15**.

A plot of the mean squared error of prediction after cross-validation for the initial PLS-R model is shown in **Figure 16**. When using all spectral variables or channels, the minimal effective prediction error was found to occur when 15 PLS components were used.

From this plot (**Figure 16**) it therefore can be concluded that the best compromise between prediction error minimization and model simplicity lies in a 15-component model.

Following preparation of the initial 15 component model different variable selection methods were evaluated to select the optimal/most predictive spectral variables for the final model. The aim here was to remove unnecessary spectral channels/variables from the model to enhance its parsimony and only include physically meaningful information. The variable importance projection (VIP) was finally calculated to determine which spectral variables have the greatest importance in predicting the viral copy number (**Figure 17A**). To assess and identify the minimum number of spectral variables required to make acceptable physical titre predictions, several variable importance thresholds were

investigated as the criterion for retained variables; generally, a VIP threshold of 1 is used – thresholds of 1.00 – 1.75 were investigated. **Figure 17B** shows variable or wavenumber ranges that the VIP algorithm identifies as regions considered most important *i.e.* those greater than a selected threshold, in this case 1.0, **Figure 17C** shows these wavenumber ranges in order of importance.

After the number of spectral variables was reduced a further assessment of the number of underlying latent variables was carried out. The optimal number of PLS components can vary with the number of spectral variables or wavenumbers that are used in the model.

10 After spectral variable reduction using a threshold of 1.0 the optimum number of PLS latent variables was found to be 9. **Figure 18** shows the MSECv plot for the refined models with different numbers of underlying components. The fact that the mean squared error of prediction increased with larger numbers of underlying components indicates that where more than 9 PLS components were included the models produced were overfitting

15 the training set.

Model predictions of RT-qPCR viral copy number for the example run of 4 bioreactors estimated using the regression coefficients obtained from the 9-latent variable and VIP > 1.0 selected spectral variable (conservative) model are shown below in **Figure 19**. The results show that the model using the Raman spectroscopy data is consistent with offline measurements of viral titre over time.

EXAMPLE 5 - REFINED VIRAL TITRE MODEL EVALUATION AND RANGE SELECTION FOR AAV

25 A further analysis to that described in example 4 was performed to analyse the number of wavenumber ranges which can be used to provide an accurate estimate of AAV viral titre.

The ranges identified as important for viral vector production, *i.e.* the ranges identified as important by variable importance projection (VIP) ≥ 1.00 after initial PLS modelling using the extended spectral range ($\sim 420 - 1800 \text{ cm}^{-1}$), were identified (*i.e.* wavenumber ranges 1 to 12 as listed in Table 1 above) and further analysis was performed.

The data were split into randomly selected paired blocks of training and test data in a 4:1 ratio, that is Raman spectra and their associated offline viral titre data for model building (80%) and model testing (20%).

5

Different combinations of the ranges deemed important by VIP were evaluated stochastically for the different training and test pairs, *i.e.* for each r total number of ranges 1-(12), many combinations were evaluated based on the model performance R^2 statistic (n.b. $R^2 = 1 - \text{residual sum of squares} / \text{total sum of squares}$) and the standard deviations of the different models' performances was evaluated to generate the confidence intervals. The minimum number of ranges was identified by choosing the number of ranges where the mean of mean R^2 values for several training/test pairs of data was approximately 0.5. **Figure 20** shows a plot of R^2 as a function of the number of wavenumber ranges.

15 This analysis identified four as being the minimum number of wavenumber ranges which are required to provide an estimate of AAV viral titre.

Thus, in any of the methods of the invention, 4 or more of wavenumber ranges 1 to 12 as presented in Table 1 identified at a VIP threshold of ≥ 1.00 may be used to calculate viral titre, as described in more detail herein. In any of these methods, preferably 6 or more of the wavenumber ranges may be used to calculate viral titre as described in more detail herein, more preferably 8 or more, or yet more preferably 10 or more, or most preferably all 12. In any of the methods of the invention, 4 or more of wavenumber ranges 13 to 22 as presented in Table 1 identified at a VIP threshold of ≥ 1.25 may be used to calculate viral titre, as described in more detail herein. In any of these methods, preferably 6 or more of the wavenumber ranges may be used to calculate viral titre as described in more detail herein, more preferably 8 or more, or most preferably all 10. In any of the methods of the invention, 4 or more of wavenumber ranges 23 to 30 as presented in Table 1 identified at a VIP threshold of ≥ 1.50 may be used to calculate viral titre, as described in more detail herein. In any of these methods, preferably 6 or more of the wavenumber ranges may be used to calculate viral titre as described in more detail herein, or most preferably all 8.

**EXAMPLE 7 – AAV8 PRODUCTION FROM A HEK 293 TRANSIENT PROCESS,
DETERMINATION OF EMPTY-VS-FULL RATIO**

Experimental Methods

5

Cell culture and Transient Transfection

HEK 293F cultures were expanded in Eppendorf DASbox BioBLU 300 bioreactors in BalanCD media (Irvine Scientific) with 4 mM GlutaMAX (Fisher) at 37°C. The cells were
10 agitated and were expanded for 24 hours prior to transient transfection to produce AAV8. The cells were transfected with rep, cap, genome encoded eGFP plasmids and helper plasmid (E2A, E4) in serum-free Opti-MEM (Gibco) to produce AAV8 particles with PEIPro (Polyplus transfection).

15 Throughout the process 11 samples were acquired from each bioreactor to measure viral titre using RT-qPCR (genome copies per ml) and 5 of these samples were additionally used for ELISA (total particles per ml). Raman spectra were acquired throughout the expansion and viral production phases.

20 *RT-qPCR*

Viral titre of AAV8 samples was measured using TaqMan™ based real-time qPCR, with final quantification provided as viral genome/mL (VG/mL). The primers of the assay targeted the *IIR2* sequences in the AAV8 viral genome. Amplicons were detected by
25 TaqMan™ fluorogenic probe. Viral titre was determined from a standard curve generated from a linearised plasmid.

ELISA

30 Total AAV8 capsid titers were determined in the extracellular AAV8 samples by ELISA, with final quantification provided as total particles/mL (TP/mL). To accurately quantify the TP/mL in each sample a reconstituted AAV8 standard of known particle concentration was used to generate a standard curve. To perform the ELISA, a mouse monoclonal antibody

specific for a conformational epitope on assembled AAV8 capsids (clone ADK8) was coated onto strips of a microtiter plate and used to capture AAV8 particles within the sample. Captured AAV8 particles were detected using two steps 1) a biotin- conjugated anti-AAV8 antibody was bound to the immune complex 2) a streptavidin peroxidase conjugate reacts
5 with the biotin molecules. Addition of the tetramethylbenzidine (TMB) substrate solution resulted in a colour reaction, which is proportional to the amount of specifically bound viral particles. The absorbance is then measured photometrically at 450nm.

10 *Raman Spectroscopy*

Raman measurements were performed using a Kaiser Optics RxN2 Raman spectrometer. This spectrometer has the capacity to monitor 4 probe channels sequentially. The RxN2 excitation source was a 785 nm near infrared diode laser with a nominal power output of
15 ~270 mW at each probe head. The samples comprised the contents of four Eppendorf, dasBox BioBLU single use systems. The beam was delivered to each sample bioreactor using four Kaiser Optics filtered fibre optic MR probes and BioOptic 220's – one set for each bioreactor. Prior to in-process measurements, the RxN2 system was stabilised for 1 hour and then each of the 4 probe channels was calibrated using the RxN2's internal auto-
20 calibration standards, in addition, a CCD sensitivity correction was performed on each probe channel using a National Institute of Standards and Technology (NIST) certified light source (HCA). The scattered light was collected using the same BioOptic 220's and MR probes as those used for beam delivery. Within each MR probe the scattered light was delivered via a second fibre optic to the RxN2 f1.8 imaging spectrograph. After filtering
25 Rayleigh scattered light using a holographic notch filter, the Raman scattered light was directed to a Kaiser Optics holographic transmission grating and then imaged onto the thermoelectrically cooled 1024 pixel CCD detector. The system has an effective bandwidth of 100 -3425 cm^{-1} and resolution of 4 cm^{-1} . Raman spectra were acquired from 100-3425 cm^{-1} with an integration time of ~15 minutes/channel including CCD readout time, 10 s
30 acquisitions were averaged over 75 accumulations to generate each measured spectrum. Each channel was measured in turn. At different times throughout the processes, liquid

samples were obtained from each bioreactor and the time point noted to enable the *post hoc* matching of the offline assay data to the commensurate Raman spectra.

Raman Data Analysis

5 All data analysis was performed in MATLAB (The MathWorks, MA, USA) version R2019b. Raw Raman spectra were pre-processed by normalising the entire spectrum to the peak intensity of the water band at $\sim 3000\text{ cm}^{-1}$. The moderate fluorescence background signal was removed for the region of $420\text{--}1800\text{ cm}^{-1}$. The low end of this range was
10 selected to avoid Raman bands that could originate from the sapphire window of the BioOptic-220 or be artefacts of the optical design of the Raman instrument and probes. The reduced normalised spectra were then inspected for obvious outliers and artefacts. The spectra associated with the offline sampling time points were identified and a model training subset of pre-processed spectra created. The training set of pre-processed spectra
15 were then used for chemometric modelling. The spectra were mean-centered prior to chemometric modelling. Several initial projections to latent structures - regression (PLS-R) models for critical analytes and viral titre (one viral titre model based on RT-qPCR, calibrated to genome copies per ml and one model viral titre model based on AAV8 ELISA calibrated to total particles per ml) were built. These models allow you to regress
20 multivariate Raman spectra against samples containing known concentrations of interesting analytes (viral titre determined from different assays, in this example RT-qPCR and ELISA). Based on these calculations the concentration of the analytes can be predicted in future. The models were prepared using a 10-fold cross validation procedure on the training data, *i.e.* $1/10^{\text{th}}$ of the data was randomly selected and removed from the training
25 data and used to assess model performance, this was done 10 times and the error values, model accuracy/performance statistics are the averages obtained for each of the 10-fold training sets. Choosing the number of underlying components or basis vectors is an important step in building supervised linear models such as PLS-R. In this work the optimal number of underlying components was identified for each model by examining
30 plots of the mean squared error of prediction after cross-validation (MSECV) as a function of component number; a minimum identifies the optimal number of PLS components for a given model. A second stage of variable selection is required to optimise the models built

by choosing only wavenumbers/variables that are most significant for prediction. This was carried out using the Variable Importance Projection (VIP) method. However, many methods of conducting variables selection exist. A typical VIP plot is shown in **Figure 26A**. Typically, variables with VIP values greater than 1 are used for the final model.

5 However, here we have built and assessed models using several VIP thresholds to identify the minimum number of spectral variables required to make good predictions and the threshold at which one ceases to be able to model the offline data of interest. As the VIP threshold is increased the number of spectral variables identified decreases. Once the significant variables were determined for each VIP threshold, final models were built.

10 Subsequently these models were used to predict the intermediate viral titre values as both genome copies per ml and total particles per ml from the Raman spectra, *i.e.* those between each offline data point for all available runs.

Preliminary viral titre model evaluation and range selection.

15 Example, pre-processed Raman spectra as used for the chemometric modelling of AAV titre both genome copies per ml and total particles per ml are shown in **Figure 21**.

AAV titre was monitored throughout the project, a representative titre (genome copies per ml) obtained by RT-qPCR is summarised in **Figure 22** and representative total particles per ml obtained by ELISA are shown in **Figure 23**.

20

A plot of the mean squared error of prediction after cross-validation for the initial PLS-R model for genome copies per ml is shown in **Figure 24**. When using all spectral variables or channels, the minimal effective prediction error was found to occur when 15 PLS components were used. Another plot of the mean squared error of predictions after cross-validation for the initial PLS-R model for total particles per ml as calibrated from ELISA data is shown in **Figure 25**. When using all spectral variables or channels, the minimal effective prediction error was found to occur when 14 PLS components were used. These choices of numbers of components offer a good compromise between prediction error minimization and model simplicity.

25

30

Following preparation of the initial 15 and 14 component models, different variable selection methods were evaluated to select the optimal/most predictive spectral variables for the final RT-qPCR and ELISA calibrated models, respectively. The aim here was to remove unnecessary spectral channels/variables from the two models to enhance their parsimony and only include physically meaningful information. The variable importance projection (VIP) was finally calculated to determine which spectral variables have the greatest importance in predicting the viral copy number (**Figure 26A**). To assess and identify the minimum number of spectral variables required to make acceptable physical titre predictions, several variable importance thresholds were investigated as the criterion for retained variables; generally, a VIP threshold of 1 is used – thresholds of 1.00 – 1.75 were investigated. **Figure 26B** shows variable or wavenumber ranges that the VIP algorithm identifies as regions considered most important for predicting genome copies per ml *i.e.* those greater than a selected threshold, in this case 1.0, **Figure 26C** shows these wavenumber ranges in order of importance.

In addition, a similar analysis was performed for identifying the most important spectral variables for predicting the particle number per ml using the above described ELISA. The variable importance projection (VIP) was calculated to determine which spectral variables have the greatest importance in predicting the viral particle number (**Figure 27A**). To assess and identify the minimum number of spectral variables required to make acceptable physical titre predictions, several variable importance thresholds were investigated as the criterion for retained variables; generally, a VIP threshold of 1 is used – thresholds of 1.00 – 1.75 were investigated. **Figure 27B** shows variable or wavenumber ranges that the VIP algorithm identifies as regions considered most important *i.e.* those greater than a selected threshold, in this case 1.0, **Figure 27C** shows these wavenumber ranges in order of importance.

After the number of spectral variables was reduced a further assessment of the number of underlying latent variables was carried out for both the viral copy number and viral particle number models, respectively. The optimal number of PLS components can vary with the number of spectral variables or wavenumbers that are used in the model. After spectral variable reduction using a threshold of 1.0 the optimum number of PLS latent variables

was found to be 10 for both genome copies per ml and particle number per ml models.

Figures 28 & 29 show the MSEC plot for the refined models with different numbers of underlying components.

- 5 Model predictions of RT-qPCR viral copy number for the example run of 8 bioreactors estimated using the regression coefficients obtained from the 10-latent variable and VIP > 1.0 selected spectral variable (conservative) model are shown below in **Figure 30** and **Figure 31**. The results show that the model using the Raman spectroscopy data is consistent with offline measurements of viral titre (genome copies per ml) over time.
- 10 Similar predictions from the ELISA total particle number for the example run of 8 bioreactors estimated using the regression coefficients obtained from the 10-latent variables and VIP > 1.0 selected spectral variable (conservative) model are shown below in **Figure 32** and **Figure 33**. The results show that the model using the Raman spectroscopy data is consistent with offline measurements of viral titre (particle number per ml) over
- 15 time.

- A method to estimate the empty-vs-full ratio for individual AAV samples as a percentage is to divide the genome copies per ml (RT-qPCR) by the total particles per ml (ELISA), and to multiply this number by 100. A similar calculation can be performed using the outputs from the predictive models developed above (**Figures 30– 33**) based on both these
- 20 methods of viral titre determination. The results of these calculations are shown in **Figure 34** and **Figure 35**, as can be seen from these transfected cultures, the empty vs full ratio matches well the estimates made using the offline RT-qPCR data and the ELISA data.

25 **EXAMPLE 8 - REFINED ELISA VIRAL TITRE MODEL EVALUATION AND RANGE SELECTION FOR AAV**

- A further analysis to the AAV8 ELISA model training such as that described in examples 3 and 5 above could be performed to calculate the number of wavenumber ranges which are necessary to provide an estimate of AAV viral titre, specifically total particles per ml.

30

The ranges identified as important for the AAV8 ELISA , *i.e.* the ranges identified as important by variable importance projection (VIP) ≥ 1.00 after initial PLS modelling using

the extended spectral range ($\sim 420 - 1800 \text{ cm}^{-1}$), would be used (*i.e.* wavenumber ranges 1 to 20 as shown in **Figure 27B** and **Figure 27C** above) and further analysis would be performed.

- 5 The data would be split into randomly selected paired blocks of training and test data in a 4:1 ratio, that is Raman spectra and their associated offline viral titre data for model building (80%) and model testing (20%).

- Different combinations of the ranges deemed important by VIP would be evaluated
10 stochastically for the different training and test pairs, *i.e.* for each r total number of ranges 1-(20), many combinations would be evaluated based on the model performance statistics such as the R^2 statistic (n.b. $R^2 = 1 - \text{residual sum of squares} / \text{total sum of squares}$) and the standard deviations of the different models' performances would be evaluated to generate confidence intervals. The minimum number of ranges would be identified by choosing the
15 number of ranges where the mean of mean R^2 values for several training/test pairs of data was approximately 0.5. Other performance statistics than R^2 could also be used for this approach.

References

- 20 Barnes, R. J., Dhanoa, M. S. and Lister, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*. 1989, 43(5): pp772-777.
- 25 Buckley, K. and Ryder, A. G. Applications of Raman Spectroscopy in Biopharmaceutical Manufacturing: A Short Review. *Applied Spectroscopy*. 2017, 71(6): pp1085-1116.
- Hu H., Bai, J. Xia, G. Zhang, W. Ma, Y. Improved Baseline Correction Method Based on Polynomial Fitting for Raman Spectroscopy. *Photonic Sensors*. 2018, 8(4): pp332-340.
- 30 Huang, J., Romero-Torres, S. and Moshgbar, M. Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review*. 2010.

- Koch, M., Suhr, C., Roth, B. and Meinhardt-Wollweber, M. Iterative morphological and mollifier-based baseline correction for Raman spectra. *Journal of Raman Spectroscopy*. 2017, 48(2): pp336-342.
- 5
- Lee, J. H., Kim, B. C., Oh, B. K. and Choi, J. W. Rapid and Sensitive Determination of HIV-1 Virus Based on Surface Enhanced Raman Spectroscopy. *J. Biomed. Nanotechnol.* 2015, 11(12): pp2223-2230.
- 10
- Lieber, C. A. and Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Applied Spectroscopy*. 2003, 57(11): pp1363-1367.
- Savitzky, A. and Golay, M. J. E. Smoothing and Differentiation of Data by Simplified
- 15
- Least Squares Procedures. *Analytical Chemistry*. 1964, 36 (8): pp1627–1639.
- Wold, S., Sjostrom, M and Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*. 2001, 58: pp109-130.
- 20
- Zhang, Z. M., Chen, S. and Liang, Y. Z. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*. 2010, 135(5): pp1138-1146.
- Zhao, J., Lui, H., McLean, D. I. and Zeng, H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Applied Spectroscopy*. 2007,
- 25
- 61(11): pp1225-32.

CLAIMS

1. A method of determining in a sample using Raman spectroscopy the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules, the method
- 5 comprising the steps of:
- (a) providing a sample and irradiating the sample with a light source;
 - (b) (i) measuring the total intensity of Raman scattered light within each one of a first plurality of wavenumber ranges to obtain a first wavenumber intensity data set for the sample, wherein the first
 - 10 plurality of wavenumber ranges are pre-selected and are characteristic of viral nucleic acids in the sample;
 - (ii) performing a first set of mathematical data processing steps on the first wavenumber intensity data set; and
 - (iii) determining the viral nucleic acid content of the sample based upon
 - 15 the output of the first set of mathematical data processing steps;
 - (c) (i) measuring the total intensity of Raman scattered light within each one of a second plurality of wavenumber ranges to obtain a second wavenumber intensity data set for the sample, wherein the second
 - 20 plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample;
 - (ii) performing a second set of mathematical data processing steps on the second wavenumber intensity data set; and
 - (iii) determining the content of viruses comprising the one or more viral
 - 25 structural molecules in the sample based upon the output of the second set of mathematical data processing steps; and
 - (d) determining the ratio of viral nucleic acids to viruses comprising the one or more viral structural molecules in the sample based on the values determined in steps (b)(iii) and (c)(iii).
- 30

2. A method according to claim 1, wherein the steps of performing the first and second sets of mathematical data processing steps on the first and second wavenumber intensity data sets comprise:
- 5 (i) optionally normalising the wavenumber signal intensity data by pre-processing the signal intensity data using one or more pre-processing analytical methods, such as a first derivative method, a second derivative method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise
- 10 polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares (airPLS) method;
- (ii) obtaining model parameters by applying to the wavenumber signal intensity data a multivariate regression algorithm, such as a partial least squares
- 15 (PLS) regression algorithm, optionally wherein the PLS algorithm is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network; and
- (iii) determining the viral nucleic acid content of the sample and determining the
- 20 content of viruses comprising the one or more viral structural molecules in the sample using the model parameters obtained by applying the multivariate regression algorithm to the signal intensity data.
3. A method according to claim 1 or claim 2, wherein the light source used to irradiate
- 25 the sample is a laser and the sample is irradiated with light having a wavelength of 785nm.
4. A method according to any one of the preceding claims, wherein the Raman scattered light is detected using a charge-coupled device (CCD).
- 30 5. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 4 or more of the wavenumber

ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 8 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 12 of the wavenumber ranges 1 to 12 as listed in Table 1 and wherein the VIP is ≥ 1.00 .

6. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 4 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 8 or more of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 10 of the wavenumber ranges 13 to 22 as listed in Table 1 and wherein the VIP is ≥ 1.25 .

7. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 4 or more of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 8 of the wavenumber ranges 23 to 30 as listed in Table 1 and wherein the VIP is ≥ 1.50 .

8. A method according to any one of claims 1 to 7, wherein the second plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the second wavenumber intensity data set for the sample comprises 4 or more of the wavenumber
5 ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 8 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the
10 VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 12 or more, 14 or more, 16 or more or 18 or more of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is $\geq$
15 1.00; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 20 of the wavenumber ranges 1 to 20 as listed in Table 2 and wherein the VIP is ≥ 1.00 .

9. A method according to any one of claims 1 to 7, wherein the second plurality of
20 wavenumber ranges in the Raman spectrum which are measured to obtain the second wavenumber intensity data set for the sample comprises 4 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 6 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or
25 wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 8 or more of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more, 11 or more or 12 of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of
30 wavenumber ranges in the Raman spectrum which are measured comprises all 13 of the wavenumber ranges 21 to 33 as listed in Table 2 and wherein the VIP is ≥ 1.25 .

10. A method according to any one of claims 1 to 7, wherein the second plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the second wavenumber intensity data set for the sample comprises 4 or more of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 5 or 6 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 7 of the wavenumber ranges 34 to 40 as listed in Table 2 and wherein the VIP is ≥ 1.50 .
11. A method according to any one of the preceding claims, wherein the viruses in the sample are adeno-associated viruses (AAV).
12. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 5 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the variable importance projection (VIP) is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 15 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 20 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 25 or more of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 28 of wavenumber ranges 1 to 28 as listed in Table 3 and wherein the VIP is ≥ 1.00 .
13. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 5 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the variable importance projection (VIP) is $\geq$

- 1.25; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 15 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 20 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 25 or more of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 30 of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 31 of wavenumber ranges 29 to 59 as listed in Table 3 and wherein the VIP is ≥ 1.25 .
14. A method according to any one of claims 1 to 4, wherein the first plurality of wavenumber ranges in the Raman spectrum which are measured to obtain the first wavenumber intensity data set for the sample comprises 5 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the variable importance projection (VIP) is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 10 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 15 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises 20 or more of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 ; or wherein the plurality of wavenumber ranges in the Raman spectrum which are measured comprises all 22 of wavenumber ranges 60 to 81 as listed in Table 3 and wherein the VIP is ≥ 1.50 .
15. A method according to any one of claims 12 to 14, wherein the viruses in the sample are lentiviruses.

16. A method according to any one of the preceding claims, wherein the nucleic acid comprises a viral DNA genome or wherein the nucleic acid comprises a viral RNA genome.
- 5 17. A method according to any one of the preceding claims, wherein the one or more viral structural molecules comprises one or more viral proteins such as one or more nucleoproteins and/or one or more capsomeres, one or more viral carbohydrates, one or more glycosylated viral molecules such as a glycosylated viral protein and/or one or more viral lipids.
- 10 18. A method according to any one of the preceding claims, wherein the ratio provides a measure of functional viral titre.
- 15 19. A method according to any one of the preceding claims, wherein the sample is a viral culture, optionally wherein the viral culture is comprised in a bioreactor.
- 20 20. A method according to claim 19, wherein the steps of irradiating the viral culture with a light source and measuring the total intensity of Raman scattered light is performed directly on the medium of the viral culture (*in situ*).
21. A method according to claim 19, wherein the steps of irradiating the viral culture with a light source and measuring the total intensity of Raman scattered light is performed directly on an aliquot of the medium which has been taken from the viral culture (*ex situ*).
- 25 22. A method according to any one of the preceding claims, the method comprising a first step of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules at a first time point and one or more further steps of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules at later time points, and further comprising measuring the change in the ratio of viral
- 30 nucleic acids to viruses comprising one or more viral structural molecules in the sample between time points, wherein each step is performed by a method according to any one of the preceding claims, preferably wherein each step is performed by the same method.

23. A method according to claim 22, wherein the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules is determined repeatedly over a time period to provide a measure of the change in the ratio in real time.
- 5
24. A method according to claim 22 or claim 23, wherein the change in the ratio in the sample is used to determine the start phase, the production phase and/or the stationary phase of a viral production process.
- 10
25. A method according to any one of the preceding claims, wherein the method is used to determine the optimal conditions for a viral production process.
26. A method according to any one of the preceding claims, wherein the method is used to assess a process downstream of a viral production process.
- 15
27. A method according to any one of the preceding claims, wherein the method comprises a step of comparing the ratio thereby obtained with the ratio obtained from the same sample by an alternative method, optionally wherein the alternative method is qPCR, RT-qPCR, ELISA or by visual determination by transmission electron microscopy.
- 20
28. A method of determining the extent of viral infection in an individual using Raman spectroscopy, the method comprising determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample by performing the method of any one of claims 1 to 23, wherein the sample is a sample which has previously been
- 25
- obtained from the individual.
29. A method according to claim 28, wherein the sample is a sample of blood, saliva, sputum, plasma, serum, cerebrospinal fluid, urine or faeces.
- 30
30. A method according to claim 28 or claim 29, wherein the ratio in the sample from the subject is compared with one or more ratio measurements which have previously been

obtained for the infection in the individual, in order to provide a prognosis of the stage of infection in the individual.

31. A method of determining in a sample using Raman spectroscopy the ratio of viral
5 nucleic acids to viruses comprising one or more viral structural molecules, the method comprising the steps of:

- (a) (i) providing a first wavenumber intensity data set for the sample,
wherein the first data set has been obtained by irradiating the sample
with a light source and measuring the total intensity of Raman
10 scattered light within each one of a first plurality of wavenumber
ranges, wherein the first plurality of wavenumber ranges in the
Raman spectrum have been selected as characteristic of viral nucleic
acids in the sample;
- (ii) performing a first set of mathematical data processing steps on the
15 first wavenumber intensity data set; and
- (iii) determining the nucleic acid content of the sample based upon the
output of the first set of mathematical data processing steps;
- (b) (i) providing a second wavenumber intensity data set for the sample,
wherein the second data set has been obtained by irradiating the
20 sample with a light source and measuring the total intensity of
Raman scattered light within each one of a second plurality of
wavenumber ranges, wherein the second plurality of wavenumber
ranges in the Raman spectrum have been selected as characteristic of
one or more viral structural molecules of the viruses in the sample;
- (ii) performing a second set of mathematical data processing steps on
25 the second wavenumber intensity data set; and
- (iii) determining the content of viruses comprising the one or more viral
structural molecules in the sample based upon the output of the
second set of mathematical data processing steps;
- (c) 30 determining the ratio of viral nucleic acids to viruses comprising the
one or more viral structural molecules in the sample based on the
values determined in steps (a)(iii) and (b)(iii).

32. A method according to claim 31, wherein the method is performed according to the steps defined in any one of claims 2 to 30.
- 5 33. The use of Raman spectroscopy for determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample.
34. The use according to claim 33, wherein the ratio is determined based upon measurements of the intensity of Raman scattered light obtained from the sample following
10 irradiation of the sample with a light source, wherein the intensity of Raman scattered light is measured from a first plurality of wavenumber ranges in a Raman spectrum which are characteristic of viral nucleic acids in the sample and from a second plurality of wavenumber ranges in a Raman spectrum which are characteristic of the one or more viral structural molecules of the viruses in the sample.
- 15 35. The use according to claim 34, wherein the sample is a viral culture, optionally wherein the viral culture is comprised in a bioreactor.
36. The use according to claim 35, wherein the step of measuring the total intensity of
20 Raman scattered light is performed directly on the medium of the viral culture (*in situ*).
37. The use according to claim 35, wherein the step of measuring the total intensity of Raman scattered light is performed directly on an aliquot of the medium which has been taken from the viral culture (*ex situ*).
- 25 38. The use according to any one of claims 34 to 37, wherein the ratio in the sample is determined at a first time point and at one or more later time points, and wherein the change in the ratio in the sample between time points is calculated.
- 30 39. The use according to claim 38, wherein the ratio in the sample is quantified repeatedly to provide a measure of the change in the ratio in real time.

40. The use according to any one of claims 33 to 39, wherein the ratio in the sample is determined by performing a method according to any one of claims 1 to 18.
41. A method or use according to any one of the preceding claims, wherein the viruses
5 in the sample are not HIV-1 or HIV-1 virus-like particles (HIV-1 VLPs).
42. A method or use according to any one of the preceding claims, wherein the Raman spectroscopy is not surface enhanced Raman spectroscopy.
- 10 43. A method of building a multivariate data processing model which is capable of determining the content of viruses comprising one or more viral structural molecules in a sample from a Raman spectroscopy wavenumber intensity data set obtained for the sample, the method comprising:
- (a) providing the sample and irradiating the sample with a light source;
 - 15 (b) measuring the total intensity of the Raman scattered light within each one of a plurality of wavenumber ranges to obtain a wavenumber intensity data set for the sample, wherein the plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample;
 - 20 (c) obtaining normalised wavenumber signal intensity data by pre-processing the signal intensity data using a pre-processing analytical method, such as a first derivative method, a second derivative method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares
25 (airPLS) method;
 - (d) obtaining model parameters by applying to the pre-processed signal intensity data a multivariate regression algorithm, such as a partial least squares (PLS) regression algorithm, optionally wherein the PLS algorithm
30 is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network, wherein a calibration is performed wherein the pre-processed signal intensity data are compared with viral titre data obtained

for the same sample conditions using non-Raman spectroscopy methods such as qPCR, RT-qPCR, ELISA or by visual determination by transmission electron microscopy;

- 5 (e) inferring response values using the model parameters obtained from the pre-processed data; and
- (f) performing variable selection, optionally variable importance projection (VIP), and identifying Raman spectral variables; and
- (g) optionally performing one or more further rounds of modelling by re-applying steps (d) to (f) and wherein unimportant variables are removed;
- 10 and

wherein the content of viruses comprising one or more viral structural molecules in a sample is determined using the model parameters obtained for the identified Raman spectral variables derived from the multivariate data processing model.

- 15 44. A method of building one or more multivariate data processing models which are capable of determining the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in a sample from a Raman spectroscopy wavenumber intensity data set obtained for the sample, the method comprising:

- (a) providing the sample and irradiating the sample with a light source;
- 20 (b) (i) measuring the total intensity of the Raman scattered light within each one of a first plurality of wavenumber ranges to obtain a first wavenumber intensity data set for the sample wherein the first plurality of wavenumber ranges are pre-selected and are characteristic of viral nucleic acids in the sample;
- 25 (ii) measuring the total intensity of the Raman scattered light within each one of a second plurality of wavenumber ranges to obtain a second wavenumber intensity data set for the sample wherein the first plurality of wavenumber ranges are pre-selected and are characteristic of the one or more viral structural molecules of the viruses in the sample
- 30 (c) obtaining normalised wavenumber signal intensity data for the first and second wavenumber intensity data sets by pre-processing the signal intensity data using a pre-processing analytical method, such as a first

- derivative method, a second derivative method, a standard normal variate (SNV) method, a polynomial fitting method, a multi-polynomial fitting method, a mollifier method, a piecewise polynomial fitting (PPF) method or an adaptive iteratively reweighted Penalized Least Squares (airPLS) method;
- 5 (d) obtaining model parameters to be applied to the first and second wavenumber intensity data sets by applying to each one of the pre-processed signal intensity data sets a multivariate regression algorithm, such as a partial least squares (PLS) regression algorithm, optionally wherein the
- 10 PLS algorithm is a nonlinear iterative partial least squares (NIPALS) regression algorithm or a neural network, wherein a calibration is performed wherein the pre-processed signal intensity data are compared with viral titre data obtained for the same sample conditions using non-Raman spectroscopy methods such as qPCR, RT-qPCR, ELISA or by visual
- 15 determination by transmission electron microscopy;
- (e) inferring response values using the model parameters obtained from each one of the pre-processed data sets; and
- (f) performing variable selection, optionally variable importance projection (VIP), and identifying Raman spectral variables; and
- 20 (g) optionally performing one or more further rounds of modelling for any of the data sets by re-applying steps (d) to (f) and wherein unimportant variables are removed; and
- wherein the ratio of viral nucleic acids to viruses comprising one or more viral structural molecules in the sample is determined using the model parameters obtained for the
- 25 identified Raman spectral variables derived from the multivariate data processing models.