Journal of Virological Methods 169 (2010) 248–252

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Short communication

The impact of PCR-generated recombination on diversity estimation of mixed viral populations by deep sequencing

Irene Görzer[a], Christian Guelly[b], Slave Trajanoski[b], Elisabeth Puchhammer-Stöckl[a,*]

[a] Department of Virology, Medical University of Vienna, Kinderspitalgasse 15, A-1095 Vienna, Austria
[b] Center for Medical Research, Medical University of Graz, Stiftingtalstraße 24, A-8010 Graz, Austria

Article history:
Received 14 June 2010
Received in revised form 26 July 2010
Accepted 28 July 2010
Available online 4 August 2010

Keywords:
PCR-generated chimeric sequences
Human cytomegalovirus
Mixed-genotype populations
Next generation sequencing

A B S T R A C T

Ultra-deep pyrosequencing (UDPS) of targeted amplicons allows to determine a large number of individual sequence reads from a single PCR product, and this approach is thus extremely valuable for analysis of mixed viral populations. A mixture of genetically distinct DNA templates, however, may lead to the generation of recombinant sequences during the initial PCR amplification step. In the present study, the frequency at which these misleading PCR artefacts are formed has been estimated by using artificial mixtures of genetically distinct human cytomegalovirus (HCMV) DNA templates for a given cycling profile. The presence of similar copy numbers of non-identical HCMV target sequences, combined with high levels of input HCMV DNA, as is found in some clinical samples, favored the formation of recombinant PCR products. Thus, to estimate the full natural diversity within mixed viral populations using UDPS, artificial chimeras generated during the PCR step should be taken into account as a potential artefact.
© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Ultra-deep pyrosequencing (UDPS) of amplicons using the 454 technology has proven to be a valuable approach for quantitative analysis of distinct DNA genotypes in clinical samples with mixed-genotype populations (Margulies et al., 2005; Thomas et al., 2006). Specific applications include the highly sensitive and accurate estimation of viral genotype frequencies and variation analysis of diverse viral populations (Eriksson et al., 2008; Görzer et al., 2010; Hoffmann et al., 2007; Rozera et al., 2009; Simen et al., 2009; Tsibris et al., 2009; Wang et al., 2007).

For effective determination of distinct genotypes in samples containing mixed-genotype populations, an initial PCR step is required to amplify each genotype. Sufficient amount of total input DNA and usually up to 40 amplification cycles are required especially when low-abundance genotype templates of 1% or even less have to be amplified efficiently (Simen et al., 2009; Solmone et al., 2009; Tsibris et al., 2009). The subsequent application of UDPS allows a large number of individual amplicon molecules to be sequenced individually (Margulies et al., 2005). The number of DNA sequences, called sequence reads, that are generated from a single PCR product defines the sensitivity of UDPS for low-abundance amplification products and approximately 10,000 sequence reads per amplicon are necessary to detect amplification products present in only 0.1–1% (Eriksson et al., 2008).

When PCR amplification is carried out using a mixture of two or more non-identical but similar genotype target sequences as templates, this can lead potentially to the formation of artificial chimeras between the amplification products (Judo et al., 1998; Lahr and Katz, 2009; Meyerhans et al., 1990). These chimeras are then detected efficiently due to the high sensitivity of UDPS.

The aim of the present study was to explore the degree to which such PCR-generated recombinants are formed when mixed human cytomegalovirus (HCMV) genotype populations are analyzed by this technique. As a model, a previously described UDPS analysis of the HCMV glycoprotein O (gO) gene was used. Here, we demonstrate that artificial gO genotype recombinants are generated by PCR when different genotypes are present in a given sample.

2. Materials and methods

2.1. HCMV DNA templates for PCR amplification

For preparation of the three control samples, viral DNA from two genetically distinct HCMV strains, AD169 (Applied Biosystems, Foster City, CA) and TB40E (kindly provided by Christian Sinzger), was used. Each HCMV strain was grown independently in human fibroblasts, and viral DNA was isolated from the cell culture supernatant using a QIAamp Mini Kit (Qiagen, Hilden, Germany). HCMV strain AD169 represents gO genotype 1a (GenBank accession no. X17403.1), and HCMV strain TB40E represents gO genotype 1c

* Corresponding author. Tel.: +43 1 40490 79520; fax: +43 1 40490 9795.
E-mail addresses: irene.goerzer@meduniwien.ac.at (I. Görzer), christian.guelly@klinikum-graz.at (C. Guelly), slave.trajanoski@medunigraz.at (S. Trajanoski), elisabeth.puchhammer@meduniwien.ac.at (E. Puchhammer-Stöckl).

0166-0934/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jviromet.2010.07.040

Illumina Exhibit 1060, Page 1 of 5
Illumina v. Molecular Loop Biosciences - IPR2024-00964

(GenBank accession no. EF999921.1). Purified viral DNA was quantitated using the Roche Amplicor assay (Cobas® Amplicor® CMV Monitor Test, Roche Molecular Systems, Branchburg, USA). Equal copy numbers of the two different HCMV genotypes were mixed at a total concentration of 10^7 copies/ml, serially diluted from 10^7 to 10^5 copies/ml and used as input template DNA for PCR amplification.

Three HCMV-DNA-positive clinical samples, one plasma sample (P1) from patient 1, and 2 independent bronchoalveolar lavage samples (B1 and B2) from patient 2 obtained in the course of routine virological investigation, were included in this study. HCMV DNA was extracted directly from 140 μl of stored clinical sample, eluted with 70 μl of DNase-free water, quantitated with the Roche Amplicor assay and used as input template HCMV DNA for PCR amplification. The total amount of HCMV DNA was 1.3×10^5 copies/ml for P1, 4.8×10^5 copies/ml for B1, and 1.2×10^6 copies/ml for B2. All three clinical samples contained a mixture of HCMV gO1a and gO1c genotypes and the ratios of gO1a to gO1c were 64:34 for P1, 16:84 for B1, and 61:39 for B2, which was determined by UDPS HCMV gO genotyping as described previously (Görzer et al., 2010).

2.2. Conditions for PCR amplification to prepare HCMV gO amplicon library for UDPS

To prepare PCR-derived gO amplicons for ultra-deep pyrosequencing, primers and cycling conditions were used exactly as described previously (Görzer et al., 2010). Briefly, for each PCR, 10 μl of isolated DNA was used as template in a total of 50 μl per well. To amplify the target region within the HCMV gO gene (position 106555–106853 of HCMV strain AD169) two separate PCRs per sample (three control samples and three clinical samples) were performed with AmpliTaq Gold® DNA polymerase (Roche Diagnostics, Basel, Switzerland) using the primer pairs gOforudps/gOMERrevudps and gOforudps/gOADrevudps, respectively. The cycling conditions were 1 cycle of 95 °C for 10 min followed by 40 cycles of denaturation for 1 min at 95 °C, annealing of gOforudps/gOMERrevudps primers for 1 min at 60 °C or of gOforudps/gOADrevudps primers for 1 min at 57 °C, and extension for 1 min with a final 5-min extension at 72 °C. The primers included 5′ extensions, which provided binding sites for ultra-deep pyrosequencing and the forward primer included 6-bp nucleotide barcode sequences, allowing all samples to be tested in a single sequencing experiment (Hoffmann et al., 2007). All amplification reactions yielded amplicons with the expected length of 343 bp.

The two separately amplified PCR products per sample were pooled, purified using Microcon 500 columns (Millipore, Bedford, MA), quality-controlled using an Agilent 2100 Bioanalyzer (Agilent Life Sciences and Chemical Analysis, Santa Clara, CA) and quantified using a Quant-iT PicoGreen Assay Kit (Life Technologies Corporation, Carlsbad, CA) according to the manufacturer's recommendations.

2.3. Ultra-deep pyrosequencing of PCR-derived gO amplicons

The PCR-derived gO amplicons of each of the six samples were pooled and subjected to UDPS, which was carried out on a Genome Sequencer FLX (Roche 454 Life Science, Branford, CT, USA) using a GS FLX emPCR Kit II (GS FLX emPCR Method Manual) according to the manufacturer's GS FLX Sequencing Method Manual.

2.4. Analysis of sequence data and chimera detection

A total of 54,952 sequence reads with a length of up to 255 bp were obtained by UDPS after passing the GS FLX quality filters and each sequence read was then assigned to a sample according to its barcode, ranging from 6621 to 14,171 sequence reads per sample. To scan for and quantify the sequences of genotypes gO1a and gO1c the GS Amplicon Variant Analyzer (AVA) version 2.0.01 software from Roche was used.

For discovery of potential recombinant gO1a/gO1c genotypes, the following strategy was used: To overcome the high pyrosequencing error rate (Huse et al., 2007) all sequence reads were initially pre-processed using RDP's pyrosequencing pipeline (http://pyro.cme.msu.edu/). For the initial processing of the sequences following quality filters were applied: exclusion of ambiguous bases (Ns), inclusion of only sequences with a length of 240 bases or more; and selection of a maximum edit distance of 2 bases for the forward primer to avoid inexact primer matches. Finally, forward primers as well as barcodes were trimmed from the filtered sequence reads. This resulted in a mean of 7567 clonal sequence reads, ranging from 5684 to 9975 per sample. The two genotypes, gO1a and gO1c, differ 9.2% in nucleotide sequence within this gO gene region. Next, the total number of sequence reads per sample was searched for a gO1a sequence pattern within the first 46 bases and a gO1c sequence pattern within the region from bases 196 to 240, and vice versa using self-developed Perl script with regular expressions. Then, for each sample all gO1a–gO1c recombinant sequences and all gO1c–gO1a recombinant sequences, respectively, were aligned to the gO1a and gO1c reference sequences using ClustalX 1.83 (Thompson et al., 1997). All alignments were edited manually using BioEdit 7.0.1 to correct obvious alignment errors. The number of gO1a–gO1c and gO1c–gO1a recombinant sequence reads was determined manually for each sample, and the frequency of occurrence was calculated as a percentage of the total number of sequence reads. Finally, the recombination breakpoints, i.e., the regions where the switch from one gO genotype sequence to another occurred, was analyzed between bases 46 and 196 of the sequence reads. The presence and distribution of the recombination breakpoints could be determined easily by manual inspection of the individual gO recombinant sequence alignments due to the short gO amplicon length. The frequency of occurrence of each of the distinct recombination breakpoints along the sequence was estimated as a percentage of the total number of recombinant sequences.

2.5. Statistical analysis

The Spearman rank test was used to estimate the correlation coefficient between the frequency of recombinant gO genotype sequences and the amount of input HCMV DNA template.

3. Results

The frequency of formation of PCR-generated recombinants between two HCMV gO genotype templates was investigated by applying UDPS HCMV gO genotyping. Therefore, control samples, harbouring two distinct gO genotype DNAs at equal amounts, were used for PCR amplification. To generate the control samples, HCMV DNA was isolated from the two genetically distinct HCMV laboratory strains, AD169 and TB40E, representing gO genotypes 1a and 1c, respectively. Equal amounts of the two HCMV DNA templates were mixed at a total concentration of 10^7 copies/ml. Serial 10-fold dilutions, ranging from 10^7 to 10^5 copies/ml, were used as input template DNA for PCR amplification. The control samples correspond to the viral DNA loads and to the gO genotype mixture present in the three patient samples also included in the study. The defined cycling conditions (see Section 2) included a 40-cycle amplification reaction and the total copy number of the template HCMV DNA mixtures ranged from 1000 to 100,000 per

Illumina Exhibit 1060, Page 2 of 5
Illumina v. Molecular Loop Biosciences - IPR2024-00964

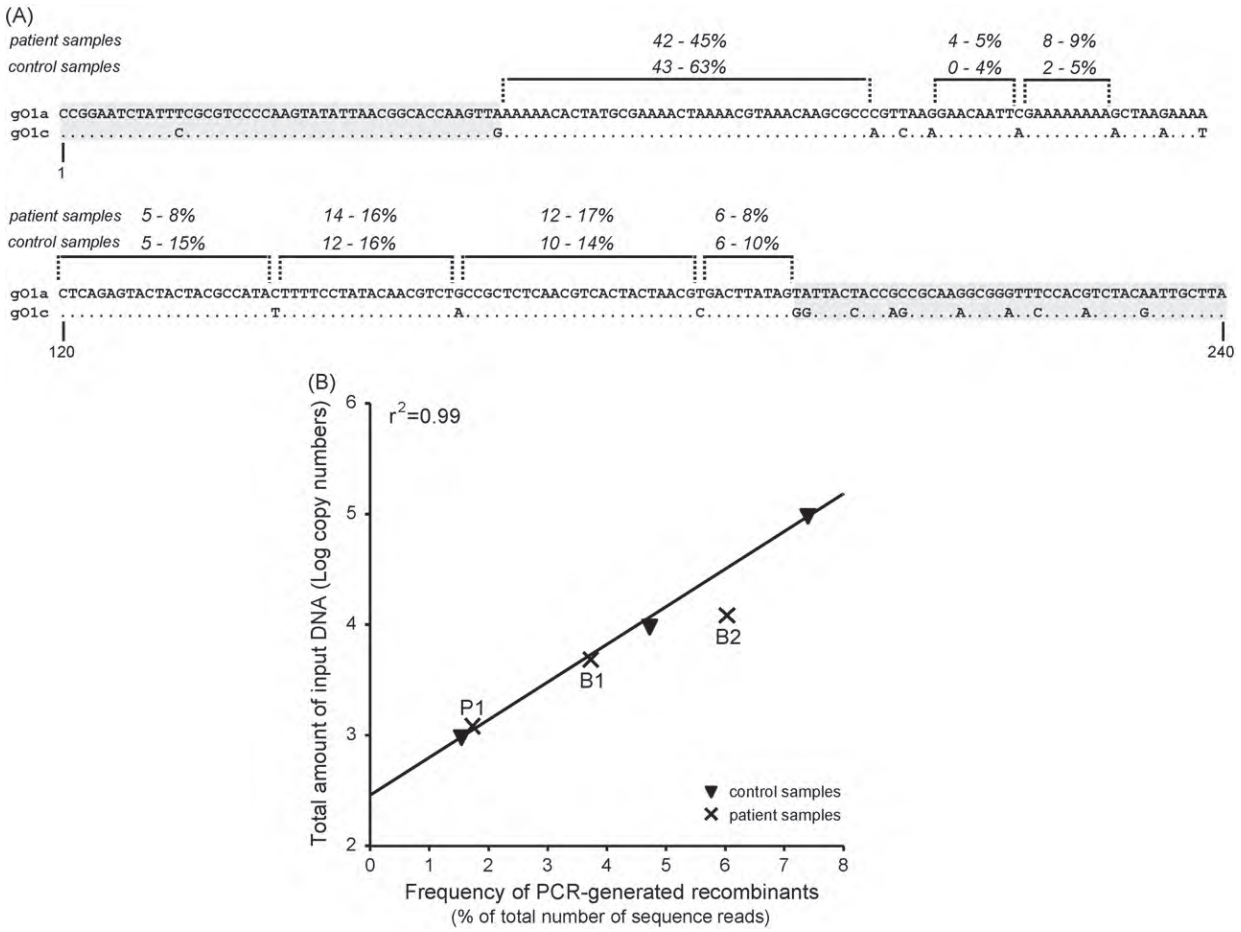

Fig. 1. (A) Nucleotide sequence similarity (90.8%) shared by human cytomegalovirus (HCMV) gO genotypes gO1a and gO1c within the sequenced region. Genotype-specific sequence patterns, as indicated by gray shading (nucleotides 1–46 and 196–240), were used to determine the original gO genotypes from which the recombinant sequences were derived. The frequency and distribution of the breakpoints along the recombinant sequence reads detected in the control samples and in the patient samples are shown above the alignment as a percentage of the total number of sequence reads of the recombinant sequence. (B) The frequency of PCR-generated gO recombinants which were derived from the control samples (▼) correlated significantly with the copy number of the input HCMV DNA used for the initial PCR amplification step of UDPS HCMV gO genotyping. The frequency of gO recombinants derived from the patient samples (×) are indicated and are as follows: P1, plasma sample from patient 1; B1 and B2, two independent bronchoalveolar lavage samples from patient 2.

reaction well. The sequenced length of the resulting gO amplicons comprised 240 bp, and the overall nucleotide sequence identity between the two distinct gO genotype sequences was 90.8%. As shown in Fig. 1A, sequence diversity between the two genotypes in the sequenced gO region is limited to distinct areas, which are interspersed between long identical or nearly identical stretches. Deep amplicon sequencing resulted in 6450–9957 sequencing reads per sample. Analysis of the sequence reads showed that PCR-generated gO genotype recombinants were obtained at all dilutions, and their frequency ranged from 1.5% to 7.4% of the total number of sequence reads. It was found that the number of recombinants correlated significantly ($r^2 = 0.99$) with the initial amount of HCMV DNA that was used for the PCR amplification step (Fig. 1B).

To obtain further information about the molecular basis of formation of recombinant gO sequences by PCR, the recombination regions for each of the recombinant sequence reads were determined. Almost all chimeric gO genotype sequences had a single breakpoint, and only sporadically (fewer than 5 sequence reads per sample) were chimeric sequences with 2 or 3 breakpoints found. This could be due to the short amplicon length. Although the breakpoints were distributed along the amplified sequence, about 50% of the recombination events were located within the longest stretch of identity (38 nucleotides) that was shared between the two distinct template DNAs (Fig. 1A).

Finally, the number of gO genotype recombinants was calculated when natural HCMV DNA templates were used. For this purpose, template HCMV DNA was isolated directly from three individual patient samples. All three samples contained a mixture of the genotypes gO1a and gO1c. The input HCMV DNA copy numbers of the sample-derived template HCMV DNA were 1300, 4800, and 12,000, and the ratios of gO1a to gO1c were 64:34, 16:84, and 61:39, respectively, as determined by UDPS HCMV gO genotyping. PCR amplification and downstream deep sequencing were performed as for the control samples, and this resulted in 5684–8330 sequence reads. As shown in Fig. 1A and B, the frequency of gO genotype recombinants, ranging from 1.25% to 5.4%, and the distribution of the recombination breakpoints along the sequence were similar to those obtained with the control samples.

4. Discussion

The findings of the present study show that the formation of PCR-generated recombinants must be taken into account when analyzing mixed viral populations by ultra-deep pyrosequencing. Under the defined experimental conditions used in this study, artificial chimeras were formed at a frequency ranging from 1.5% to 7.4%, depending on the total amount of input DNA. This is in agreement with chimera formation rates reported in previous studies

Illumina Exhibit 1060, Page 3 of 5
Illumina v. Molecular Loop Biosciences - IPR2024-00964

(Judo et al., 1998; Meyerhans et al., 1990; Shafikhani, 2002). Higher rates of PCR-generated recombination have been observed when the starting material included a higher number of genotypes (Lahr and Katz, 2009). In the present study, an equimolar mixture of the two initial gO genotypes, gO1a and gO1c, were chosen in order to mimic the natural HCMV genotype mixtures, observed in three clinical samples. Besides, it has been shown previously that naturally occurring HCMV populations are composed frequently of two or even more gO genotypes (Görzer et al., 2010; Stanton et al., 2005).

In the present study, it was observed that the rate of recombination and the distribution of the recombination sites were similar in both, the control and the clinical samples. Hence, it is very likely that most of the gO genotype recombinants that were found after amplification of the clinical samples were PCR-generated artefacts that were not present originally in the natural virus population. However, there is evidence that genotype recombinants may emerge in natural HCMV populations (Faure-Della Corte et al., 2009; Garrigue et al., 2008; Grosjean et al., 2009; Haberland et al., 1999; Yan et al., 2008). Thus, it is of high importance to distinguish natural recombinants from artificially generated chimeras when both of the parental genotypes are present in the sample, in order to capture and reflect the full natural diversity within a mixed virus population. One possibility is to diminish chimera formation by reducing the number of PCR cycles and the template concentration, the two main critical factors that affect PCR-generated recombination (Lahr and Katz, 2009). However, since it has been shown previously that low-abundance HCMV genotypes at relative frequencies of 1% or even less can be present in clinical samples with mixed HCMV populations, a high initial copy number of HCMV DNA templates and a 40-cycle amplification reaction are essential to capture low-abundance gO genotypes (Görzer et al., 2010). In this regard, the PCR conditions needed to utilize the full potential of the ultra-deep pyrosequencing technology for detection of low-abundance populations simultaneously favor the generation of such PCR-generated recombinants. Amplification from single genomes generated by limiting dilution PCR analysis followed by Sanger sequencing could be another possibility to confirm natural recombinants. However, it will almost be unfeasible to reach sensitivity levels of 1% or even less for reliable detection of low-abundance variants.

Detailed inspection of the distribution of the recombination sites showed that the recombination events occurred more frequently in a long stretch of sequence identity (38 nucleotides) shared between the two genotypes. Thus, it is likely that the degree and distribution of diversity along the whole target region also affects the formation of chimeric sequences during PCR, and these PCR artefacts might therefore only occur when using templates with a high overall sequence similarity. A careful selection of the target region enabling the efficient discrimination of genotypes and avoiding long runs of sequence identity might be an option to diminish the formation of PCR-derived recombinants. Therefore, the formation of PCR-derived recombinants with a defined template mixture, specified by the target sequence and its variable regions, has to be determined before any conclusions can be drawn in respect to the presence of naturally occurring virus recombinants.

Funding

No funding.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank Barbara Dalmatiner and Daniela Schmidt for excellent technical assistance with UDPS amplicon library preparation, and we are grateful to Theresa Maierhofer and Gabriele Michelitsch for help with the Ultra Deep 454 Sequencing.

References

Eriksson, N., Pachter, L., Mitsuya, Y., Rhee, S.Y., Wang, C., Gharizadeh, B., Ronaghi, M., Shafer, R.W., Beerenwinkel, N., 2008. Viral population estimation using pyrosequencing. PLoS Comput. Biol. 4, e1000074.

Faure-Della Corte, M., Samot, J., Garrigue, I., Magnin, N., Reigadas, S., Couzi, L., Dromer, C., Velly, J.F., Dechanet-Merville, J., Fleury, H.J., Lafon, M.E., 2009. Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. J. Clin. Virol. 47, 161–169.

Garrigue, I., Faure-Della Corte, M., Magnin, N., Recordon-Pinson, P., Couzi, L., Lebrette, M.E., Schrive, M.H., Roncin, L., Taupin, J.L., Dechanet-Merville, J., Fleury, H., Lafon, M.E., 2008. UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. Transplantation 86, 826–835.

Görzer, I., Guelly, C., Trajanoski, S., Puchhammer-Stöckl, E., 2010. Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. J. Virol. 84, 7195–7203.

Grosjean, J., Hantz, S., Cotin, S., Baclet, M.C., Mengelle, C., Trapes, L., Virey, B., Undreiner, F., Brosset, P., Pasquier, C., Denis, F., Alain, S., 2009. Direct genotyping of cytomegalovirus envelope glycoproteins from toddler's saliva samples. J. Clin. Virol. 46 (Suppl. 4), S43–S48.

Haberland, M., Meyer-Konig, U., Hufert, F.T., 1999. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. J. Gen. Virol. 80, 1495–1500.

Hoffmann, C., Minkah, N., Leipzig, J., Wang, G., Arens, M.Q., Tebas, P., Bushman, F.D., 2007. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 35, e91.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 8, R143.

Judo, M.S., Wedel, A.B., Wilson, C., 1998. Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Res. 26, 1819–1825.

Lahr, D.J., Katz, L.A., 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. Biotechniques 47, 857–866.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Meyerhans, A., Vartanian, J.P., Wain-Hobson, S., 1990. DNA recombination during PCR. Nucleic Acids Res. 18, 1687–1691.

Rozera, G., Abbate, I., Bruselles, A., Vlassi, C., D'Offizi, G., Narciso, P., Chillemi, G., Prosperi, M., Ippolito, G., Capobianchi, M.R., 2009. Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. Retrovirology 6, 15.

Shafikhani, S., 2002. Factors affecting PCR-mediated recombination. Environ. Microbiol. 4, 482–486.

Simen, B.B., Simons, J.F., Hullsiek, K.H., Novak, R.M., Macarthur, R.D., Baxter, J.D., Huang, C., Lubeski, C., Turenchalk, G.S., Braverman, M.S., Desany, B., Rothberg, J.M., Egholm, M., Kozal, M.J., 2009. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J. Infect. Dis. 199, 693–701.

Solmone, M., Vincenti, D., Prosperi, M.C., Bruselles, A., Ippolito, G., Capobianchi, M.R., 2009. Use of massively parallel ultradeep pyrosequencing to characterize the genetic diversity of hepatitis B virus in drug-resistant and drug-naive patients and to detect minor variants in reverse transcriptase and hepatitis B S antigen. J. Virol. 83, 1718–1726.

Stanton, R., Westmoreland, D., Fox, J.D., Davison, A.J., Wilkinson, G.W., 2005. Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. J. Med. Virol. 75, 42–46.

Thomas, R.K., Nickerson, E., Simons, J.F., Janne, P.A., Tengs, T., Yuza, Y., Garraway, L.A., LaFramboise, T., Lee, J.C., Shah, K., O'Neill, K., Sasaki, H., Lindeman, N., Wong, K.K., Borras, A.M., Gutmann, E.J., Dragnev, K.H., DeBiasi, R., Chen, T.H., Glatt, K.A., Greulich, H., Desany, B., Lubeski, C.K., Brockman, W., Alvarez, P., Hutchison, S.K., Leamon, J.H., Ronan, M.T., Turenchalk, G.S., Egholm, M., Sellers, W.R., Rothberg, J.M., Meyerson, M., 2006. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat. Med. 12, 852–855.

Illumina Exhibit 1060, Page 4 of 5
Illumina v. Molecular Loop Biosciences - IPR2024-00964

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tsibris, A.M., Korber, B., Arnaout, R., Russ, C., Lo, C.C., Leitner, T., Gaschen, B., Theiler, J., Paredes, R., Su, Z., Hughes, M.D., Gulick, R.M., Greaves, W., Coakley, E., Flexner, C., Nusbaum, C., Kuritzkes, D.R., 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One 4, e5683.

Wang, C., Mitsuya, Y., Gharizadeh, B., Ronaghi, M., Shafer, R.W., 2007. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 17, 1195–1201.

Yan, H., Koyano, S., Inami, Y., Yamamoto, Y., Suzutani, T., Mizuguchi, M., Ushijima, H., Kurane, I., Inoue, N., 2008. Genetic linkage among human cytomegalovirus glycoprotein N (gN) and gO genes, with evidence for recombination from congenitally and post-natally infected Japanese infants. J. Gen. Virol. 89, 2275–2279.