JVI Accepts, published online ahead of print on 7 April 2010
J. Virol. doi:10.1128/JVI.02690-09

Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus.

Joseph G. Victoria[1,2], Chunlin Wang[3], Morris S. Jones[4], Crystal Jaing[5], Kevin McLoughlin[5], Shea Gardner[5], Eric L. Delwart[1,2*]

[1] Blood Systems Research Institute, San Francisco, CA, 94118

[2] Dept of Laboratory Medicine, University of California, San Francisco CA 94118

[3] Stanford Genome Technology Center, Stanford, CA, 94304

[4] Clinical Investigation Facility, David Grant USAF Medical Center, Travis, CA, 94535 USA

[5] Lawrence Livermore National Laboratory, Livermore, CA 94551

Running title: Adventitious virus in live attenuated vaccine

* Corresponding author: BSRI, 270 Masonic Ave, San Francisco CA 94118. delwarte@medicine.ucsf.edu

Abstract

Metagenomics and a pan-microbial microarray were used to examine eight live attenuated viral vaccines. Viral nucleic acids in trivalent oral poliovirus (OPV), rubella, measles, yellow fever, varicella-zoster, multivalent Measles/Mumps/Rubella, and two rotavirus live vaccines were partially purified, randomly amplified, and pyrosequenced. Over half a million sequence reads were generated covering from 20 to 99% of the attenuated viral genomes at depths reaching up to 8000 reads per nucleotides. Mutations and minority variants, relative to vaccine strains, not known to affect attenuation were detected in OPV, mumps virus, and varicella-zoster virus. The anticipated detection of endogeneous retroviral sequences from the producer avian and primate cells was confirmed. Avian leucosis virus (ALV), previously shown to be non-infectious for humans, was present as RNA in viral particles while simian retrovirus (SRV) was present as genetically defective DNA. Rotarix, an orally administered rotavirus vaccine, contained porcine circovirus-1 (PCV1), a highly prevalent non-pathogenic pig virus, which has not been shown to be infectious in human. Hybridization of vaccine nucleic acids to a pan-microbial microarray confirmed the presence of endogenous retroviral and PCV1 nucleic acids. Deep sequencing and microarrays can therefore detect attenuated virus sequence changes, minority variants, and adventitious viruses and help maintain the current safety record of live attenuated viral vaccines.

Introduction:

Highly effective, safe, and relatively inexpensive, live attenuated viruses protect against numerous human and animal viral infections. Attenuation is achieved by genetically adapting viruses for replication in a different host species or under non-physiological conditions such that viruses lose their pathogenic potential in their original hosts while remaining sufficiently antigenic to induce lasting protective immunity. Live attenuated vaccines are highly efficacious due to the physiologic presentation of native antigen to the host's immune system and include the earliest human vaccine developed through serial passages of rabies virus in rabbits. In very rare instances one attenuated viral vaccine, the oral poliovirus vaccine (OPV), can accumulate mutations as well as recombine with other co-infecting enteroviruses and revert to a pathogenic state (18, 24). Attenuated live vaccines also carry a potential risk of contamination with adventitious viruses introduced during the attenuation process, from the cell lines used, and/or from the animal sera or other biologics often used in cell cultures. Very early Theiler's yellow fever attenuated virus was once "stabilized" with human plasma thought to contain hepatitis B virus resulting in many cases of hepatitis (5, 28). Some early Sabin poliovirus vaccines were contaminated with the SV40 polyomavirus from the monkey cells used to amplify polioviruses. While carcinogenic in rodents, SV40 has no epidemiologic association with human cancers (10). Avian leukosis virus (ALV) and endogenous avian virus (AEV) has been reported in attenuated vaccines grown in chicken embryo fibroblasts but extensive testing has also ruled out human infections (14, 15). Vaccine associated ALV and AEV are thought to originate from endogenous retroviruses in the chicken germ line (14, 15, 17).

Because the chemical inactivation used in the manufacture of killed virus vaccines is also likely to inactivate adventitious viruses, we focused on eight live attenuated viruses, OPV, rubella (Meruvax-II), measles (Attenuvax), yellow fever (YF-Vax), human herpes 3 (Varivax), rotavirus (Rotarix and Rotateq) and multivalent Measles/Mumps/Rubella (MMR-II), to re-sequence the attenuated viruses and test for the presence of adventitious viruses after viral particle purification, massively parallel pyrosequencing and viral sequence similarity searches. Vaccine nucleic acids were also analyzed using a pan-microbial microarray.

Methods

Viral particle purification

Lyophilized live attenuated vaccines were resuspended in 200µl of manufacturer appropriate sterile diluent (Merck, LOT#: 4089) or sodium chloride solution (Sanofi Pasteur, LOT#: UF198AB). Rotarix (Glaxo Smith Kline, LOT#: A41XA799A) and Rotateq (Merck, LOT#: 1724X) were resuspended in 1 dose volume of accompanying oral diluent. 200µl of either OPV (Bharat Biotech, LOT #: 63CM9004), Meruvax (Merck, LOT#: 1198X), Attenuvax (Merck, LOT#: 1440X), YF-VAX (Sanofi Pasteur, LOT# UF430AA), MMR-II (Merck: LOT#: 1732X), Rotarix, Rotateq, and Varivax (Merck: LOT# 1526X) was filtered through a 0.45µM filter (Millipore). Filtrate containing viral particles was treated with a mixture of DNase and RNase to remove exogenous, unprotected nucleic acids as described previously (36). Viral nucleic acids were then extracted using QIAamp Viral RNA extraction kit (Qiagen).

Sequence independent amplification of viral nucleic acids

Viral cDNA synthesis and random PCR amplification were performed as previously described (36). Briefly, 100pmol of primer consisting of an arbitrarily designed 20-base oligonucleotide followed by a randomized octamer (8N) sequence at the 3'-end was used in a reverse transcription reaction (Superscript III, Invitrogen). For each live attenuated vaccine, two distinct primers containing different 20-base fixed sequences were used in separate RT reactions. A single round of DNA synthesis was then performed using Klenow fragment polymerase (New England Biolabs; NEB) followed by PCR amplification of double stranded DNA using a primer consisting of only the 20 fixed bases. Independent duplicate PCR reactions were performed for each random primer, generating a total of 4 separate reactions for each vaccine. Random PCR DNA products were separated on an agarose gel, and fragments from 500bp to 1000bp were excised and extracted. DNA was pooled and sequenced on a single pyrosequencing gasket using GS FLX Titanium reagents.

Microarray analysis

Nucleic acid samples were reverse transcribed to cDNA using tagged random primers as above (36) except that primer A2 (5'-GATGAGGGAAGATGGGGNNNNNNNNN-3') was used (39). The cDNA was synthesized into double-stranded cDNA and then amplified by random PCR using primer B2 (5'-GATGAGGGAAGATGGGG-3'). The PCR product was purified using the Qiaquick PCR Cleanup kit (Qiagen) and labeled with Cy-3 labeled nonamers from Trilink Biotechnologies. Two µg of labeled sample was hybridized to the Microbial Detection Array (MDA) for 16 hours, washed and scanned using the Axon 4000B fluorescent scanner as previously described (16).

Scan data was analyzed as described (11), except that additional stringency criteria were applied to the results. Probes were classified as detected or undetected according to whether their intensity exceeded a threshold, equal to the 99th percentile of intensities for negative control probes. Log-odds scores for presence of targets in a database of 43,705 viral genome sequences were computed using a greedy forward selection method, and targets with scores less than 5 were excluded. The remaining candidate targets were filtered further according to the fraction of expected probes detected for each target. The expectation value of the total target-specific detected probe count was calculated as the sum of the predicted detection probabilities for all probes with BLAST hits to the target sequence with BLAST scores less than 42. The detected probe expected count is the sum of detection probabilities for probes with hits to the target that had intensities above the detection threshold. Candidates for which the detected probe expected count was less than 20% of the total expected count were excluded from the final results. When significant log-odds scores are obtained for a group of closely related targets, scores and probe counts are reported for the highest-scoring target in the group.

Sequence reads classification

The 454 sequence reads were sorted into their vaccine sample of origin according to the unique sequence tag (20 bases of the unique random PCR primer). The sequence reads from each vaccine were then assembled using the program Sequencher (Genecodes) with an overlap set as ≥95% over 35bp to merge two fragments. The

assembled sequences and the singlets greater than 100bp were compared to the NCBI non-redundant nucleotide and protein databases using BLASTn and BLASTx, respectively. Sequences were classified into virus, bacteria, and eukaryota based on the taxonomic origin of the best-hit sequence. Sequences whose best alignment E-value was $>10e^{-5}$ were grouped as unclassifiable.

Virus specific PCR for ALV, SRV and PCV1

RT-PCR conditions for ALV and SRV1 included generating cDNA with MuMLV reverse transcriptase (Promega) using a random hexamer, as per manufacturers protocol. PCR reaction conditions for ALV, SRV, and PCV1 each contained 0.25mM final concentration dNTPs, 1X NEB buffer, 0.2mM of each appropriate primer, and 5U Taq polymerase (NEB). For ALV and SRV, reverse transcriptase reactions were performed on either untreated samples or samples treated for 30 minutes with 2U of rDNase I (Ambion) followed by DNase inactivation according to manufacturers' instructions (Ambion). All PCR cycles included an initial 95°C denaturation for 5 minutes and a post-cycling final extension at 72°C for 7 minutes. All first round PCR included 35 cycles at either: ALV- 95°C for 30s, 56°C for 30s, 72°C for 45s; SRV1 - 95°C for 30s, 50°C for 30s, 72°C for 45s; PCV1 - 95°C for 30s, 55°C for 30s, 72°C for 30s. 1µL of PCR product was used in the second round of PCR which consisted of 30 cycles with identical cycling conditions as round 1 for ALV and SRV but consisting of 95°C for 20s, 55°C for 20s, 72°C for 20s for PCV1. First round primer sequences were as follows: ALV-F1 (5-CTAGCCTGTTGGTCCGTTAAA-3), ALV-R1 (5-CCTACAAGCCTTTTGCAACTTC-3), SRV-F1 (5-GAATCTGTAGCGGAYAATTGGCTT-3), SRV-R1 (5-GGGCGRATKGCTGCYTGACA-3), PCV-F1 (5-TTGGTGTGGGTATTTAAATGGA -3), PCV-R1 (5-GCAGCCATCTTGGAAACAT-3). Second round primers sequences were as follows: ALV-F2 (5'-GGAACATGTCAATAAGATCGGC-3), ALV-R2 (5-ATTCCGTGTGATAGCTGATTGA-3), SRV-F2 (5-ACTTGTTAGGGCAGTCCTYTCWGG-3), SRV-R2 (5-ACAGGCTGGRTTAGCRTTTTCATA-3), PCV-F2 (5-TATAGGGGTCATAGGCCAAGTT-3), PCV-R2 (5-CCCTACCTTTCCAATACTACCG -3).

Sequencing of SRV DNA in Rotateq

A dose of Rotateq vaccine was resuspended in sterile diluent and extracted using QIAamp viral RNA kit (Qiagen). SRV-1 (GenBank accession: U85506) (35) and African Green Monkey (trace ID: AC216616) sequences were used to design primers spanning the length of the genome. Genome sequences were obtained in both the forward and reverse direction.

GenBank accession numbers for the sequence data reported in this paper are in the short read archive under accession number SRA012283. Deposition of PCV1 and SRV sequences are xxxxx.

Results:

Deep sampling of the attenuated viruses: Viral particles in eight live attenuated vaccines were purified through a 450nm filter to eliminate bacteria-sized particles, digested with DNase and RNase enzymes to remove non-particle protected "naked" nucleic acids, amplified by random PCR, and subjected to metagenomic pyrosequencing (4, 7). A total of 501,753 reads above 50bp length (average length 311bp) were generated. Sequence reads were aligned against reference genome(s) of the expected attenuated virus(es). Viral regions sequenced and the depth at which different positions were covered are shown in Figure 1, while Table 1 reports the percentage of viral reads and genome coverage for each vaccine. Remaining sequence reads were assembled to generate contigs and both singlet sequences and contigs were then classified using BLASTx to the GenBank non-redundant database. Sequences were classified as viral, eukaryotic, or bacterial/unclassifiable based on the best BLASTx match with an expectation value (E-value) of $10e^{-5}$ or lower (Fig. 2). Sequences that exhibited a best match with an E-value greater than $10e^{-5}$ were deemed unclassifiable and typically were short reads (less than 150bp) with a weak match to bacterial sequence. The amplification of bacterial and unclassifiable sequences as well as viral (bacteriophage and prophages) and eukaryotic (human and plants) when pure water alone was used as nucleic acid input indicated that low-level contamination of reagents and enzymes may account for some of the non-vaccine sequences detected in vaccines (Fig. 2).

Adventitious viruses in vaccines: In addition to the attenuated virus, we also found other viral sequences. Attenuvax contained 4 reads covering 700 nucleotides of the avian leukosis virus (ALV) genome. Rotateq contained a single 276 nucleotide read with 96% identity to the simian retrovirus (SRV) genome. Rotarix contained 6344 reads with 98% identity to pig circovirus 1 (PCV1), covering the complete circular genome (Fig 1).

Nature of non-vaccine viral sequences: To confirm the presence of ALV, SRV and PCV1, total nucleic acids were directly extracted from all vaccines (without prior filtration and nuclease treatment), and both single round and nested PCR (nPCR) was used to test for the presence of each virus (Table 1). ALV was detected in all three vaccines propagated in chicken embryo fibroblasts, although at lower levels detectable only by nPCR in MMRII and YF-Vax (Table 1 +). The presence of SRV in the Vero cell derived Rotateq vaccine was confirmed with a single round of PCR (Table 1 ++). SRV was also found by nPCR in the other Vero cell derived Rotarix vaccine. However, SRV DNA was not detected in the Vero cells derived OPV. Nucleic acids from all vaccines were tested for PCV1 using nPCR; only Rotarix was PCV1 positive. A second lot of both Rotarix (lot#: A41FA902A) and Rotateq (lot #: 0288Y) were also tested for the presence of SRV and PCV1. Both were positive for SRV by nested PCR. The second lot of Rotarix was also positive in single round PCR for PCV1, while Rotateq remained negative. To ascertain whether PCV1 was found in other GlaxoSmithKline (GSK) vaccines, we obtained a single dose of Pediarix containing non-infectious proteins from diphtheria, tetanus, pertussis bacteria, hepatitis B and killed polioviruses. Pediarix was negative for PCV-1 DNA using nPCR (data not shown). End-point dilution nested PCR was then performed to determine the approximate amount of PCV1 in both Rotarix lots (#A41XA799A and #A41FA902A) and was determined to be $\geq$175,000 and 250,000 PCV1 DNA copies per vaccine dose, respectively.

In order to determine if the retroviral ALV and SRV sequences originated from viral particles, the Rotateq and Attenuvax preparations were spun to pellet viral particles that were then resuspended and extracted. An aliquot of each extraction was also treated with DNase. The DNase-treated and untreated extracts were then amplified by single round or nested PCR, with or without cDNA synthesis using reverse transcriptase. ALV was still detectable after DNase treatment and required cDNA synthesis for its detection indicating that it was in the form of encapsulated RNA. SRV did not require cDNA synthesis for its PCR detection and was no longer detectable after DNase treatment indicating that it was likely present in the form of non-infectious host cell DNA released from Vero cells (Table 1). The presence of endogenous SRV provirus in Vero cell was confirmed by single round PCR using DNA extracted from uninfected Vero cells freshly obtained from ATCC (ATCC:

CCL-81), the parental cell line of the Vero E6 clone used to generate the vaccine (ATCC: CRL-1586). As 41% of the 13,520 pyrosequence reads from the Rotateq were of primate origin, it is likely the SRV sequence detected was the result of leakage of host cell DNA, including that of endogenous (germline) proviral DNA. To further test the endogenous nature of the Rotateq SRV DNA, the entire coding region was sequenced revealing 94% nucleotide identity to SRV-1 (GenBank accession: U85506) (35) . The inactivating mutations of endogenous SRV-1 genomes in baboons (35) were not observed in the Rotateq associated SRV genome. However, a frame-shifting single nucleotide insertion in the polymerase gene was identified at genomic position 3,726 providing evidence that genetically defective SRV proviral DNA is present in Vero cells and in vaccines derived from Vero cells.

Microarray analysis of vaccine samples: The Microbial Detection Array (MDA) developed at Lawrence Livermore National Laboratories contains 388,000 probes designed from all sequenced viral and bacterial organisms (11). The expected attenuated viruses were detected in all vaccines tested (Table 2). OPV was not tested. Two separate total nucleic acid preparations of the MMR-II vaccine yielded different results, identifying measles and mumps viruses in one preparation and rubella in the other. In vaccines grown in chicken embryo fibroblast (CEF) signals to avian endogenous retrovirus (AEV) were detected. The anticipated detection of AEV reflects the presence of multiple loci of these endogenous proviruses in the chicken germline (8, 9) and release of AEV viral particles into CEF derived vaccines (15, 34, 38) as well as possible cross-hybridization with ALV sequences since the more conserved region of related viral groups are used on the microarray. The detection of human endogenous retrovirus K (HERVK) in Varivax, MMR-II, and Meruvax was the expected consequence of their manufacture using human cell lines (Table 2). The origin of the Baboon endogenous retrovirus signal for Rotateq is assumed to be related to the African green monkey derived Vero cell used in its manufacture and cross-hybridization of its endogenous retroviruses to the Baboon endogenous retrovirus probes.

Both nucleic acid extracted directly from the vaccines as well as nucleic acids extracted from vaccines that had first been filtered and nuclease digested (to enrich for viral particle associated nucleic acids) were used

as input to generate the DNA hybridized to the microarrays (Table 2 TOT and VP). The viral particle nucleic acid enrichment step (Table 2 VP) reduced the number of probes yielding signal from producer cell derived endogenous retroviruses to undetectable levels for Attenuvax, Meruvax and Rotateq, while YF-Vax exhibited ~50% reduction when compared to total nucleic acid extraction. No difference was observed between the two extraction methods in the detection of any vaccine virus, with the exception of Rotarix, for which the VP purification procedure resulted in detection of only 7 of the 11 genome segments, yielding a low overall log-odds score for the entire genome.

No pathogenic revertant sequences in the oral poliovirus vaccine. Since the year 2000, 426 cases of acute flaccid paralysis in 11 countries have been attributed to 9 outbreaks of OPV revertants resulting in pathogenic vaccine-derived poliovirus (VDPV) (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5836a3.htm). Recombination of attenuated poliovirus with other human enterovirus C viruses together with several reversion mutations to the wild-type poliovirus is the most common route in the development of VDPV (http://www.polioeradication.org/content/fixed/VDPV_background.asp). In-depth sequencing of OPV allowed for detection of low-level polymorphisms generated by culture of the vaccine strains. No mutations associated with reversion to poliovirus pathogenicity were observed as minority variants (Table 3). Mutations at levels lower than 2%, a conservative cutoff exceeding published rates of pyrosequencing error (22, 37), were observed at sites known to be polymorphic (26). For example PV3 exhibited a C to T change at position 2,493 in 0.5% (38/8250) of sequence reads. The detection of high-frequency frame-shifting mutations also provided evidence of a mixture of both attenuated and defective PV2 (6%) and PV3 (30%) genomes (Table 4). Coverage depths for Varivax and Mumps (in the MMR-II vaccine) were also sufficient to examine variation from reported vaccine strains and the presence of nucleotide polymorphisms. The percentage of nucleotide position for the human herpes 3 in Varivax differing from reported Oka strain (31) was less than 0.1% with 110 mixed bases at both previously described and novel positions (31). No mutation reached 100% of the HHV3 population.

Mumps virus in MMRII exhibited only a 0.30% mutation frequency from a published vaccine strain (Accession number: FJ211586) (32) with 4 nucleotide substitutions and 31 mixed base polymporphisms.

Conclusion:

Using viral metagenomics we identified nucleic acids from endogenous retroviruses and one adventitious virus in attenuated viral vaccines. The detection of ALV and EAV particles in live vaccines derived from chicken embryo fibroblast has been previously reported and extensively investigated (14, 15). Sero-conversion to ALV and EAV was not detected among vaccinated individuals indicating these viruses were non-infectious to human (14, 15). *In vitro* inoculation studies in a variety of human cell lines and peripheral blood mononuclear cells have also failed to show replication of CEF derived retroviral particles nor was there evidence of ALV proviral integration (19, 27, 29). The detection of ALV and EAV in attenuated viral vaccines is therefore not of concern as an infectious agent.

The detection of SRV DNA in Vero cells reflected the presence of endogenous SRV in the germline of African green monkeys from which this cell line was derived. SRV is thought to be a highly prevalent infection of old-world monkeys and endogenized at multiple copies in the genomes of many non-human primate species (23, 30, 35). The removal of SRV nucleic acid from the Rotateq vaccine by DNase indicated that it was present as naked DNA released from Vero cells. The detection of a retroviral inactivating mutation in the Rotateq SRV DNA pol gene indicated a defective nature for this endogenous retrovirus. A prior study showed two cases of SRV sero-conversion in people with extensive blood and saliva contacts with non-human primates presumably following infection with replication competent SRV (20). Neither associated clinical symptoms nor ongoing viremia were observed in these sero-positive individuals (20). The detection of SRV DNA (or other retroviral sequences) in live attenuated vaccines therefore simply reflects their generation in cells, with their very large genome load of endogenous retroviruses, rather than the presence of adventitious viruses of concern.

A large fraction of the sequence reads from the Rotarix vaccines (41.6%) consisted of PCV1 sequences. Because serum free media is used in the manufacture of Rotarix a possible source for PCV1 is porcine pancreas derived trypsin used for passaging Vero cells. After introduction of PCV1 a chronic infection may have become established in Vero cells (1). Alternatively, since the porcine derived trypsin was irradiated before use to inactivate adventitious viruses, the PCV1 DNA may simply reflect carryover of non-infectious, lethally

irradiated, PCV1 (GSK personal communications). PCV1 viral loads were measured in two lots at $>10^5$ PCV1 DNA molecules per vaccine dose and the ratio of PCV1 to rotavirus pyrosequencing reads was >10. Because the relative efficiency of viral particle purification and random PCR for PCV1 (a circular DNA viruses) and rotavirus (a segmented RNA virus) is not known the ratio of their viral particle numbers in Rotarix is currently unknown. Whether the PCV1 in Rotarix is infectious in various human or porcine cell lines is currently unknown. PCV1 is a highly common and non-pathogenic pig infection usually transmitted among pigs through the fecal-oral route. PCV2, a close relative of PCV1, is a major pig pathogen with a high economic impact on the swine industry. Since both PCV species are highly prevalent in healthy pigs, human dietary and respiratory exposure to this virus is common through pork consumption or inhalation of particles from pig feces in the swine industry. Both PCV species as well as other members of the *Circoviridae* family are commonly found in human stools (21). Whether PCV1 or PCV2 can actually replicate in humans is controversial, with most data weighing against human tropism. A single human gut biopsy has been reported to contain PCV2 DNA but contamination from human feces containing PCV2 from consumed pork cannot be excluded (6). A large PCR screen of human plasma and tissues did not detect any PCV2 DNA and inoculations of various human cell lines with PCV2 were non-productive (12) (13). Serological testing for PCV exposure has yielded ambivalent results with one group reporting a high rate of human sero-reactivity to PCV1 although with altered binding profile relative to pig sera indicating the possible presence of a related but distinct virus (33). Other studies reported no detectable human anti-PCV1 (3) or anti-PCV2 antibodies (2). While PCV2 transmission can occur between pigs through eating of pig flesh (25), the absence of robust and specific human sero-conversion to pig circoviruses, despite frequent consumption of infected pork, further supports an inability of pig circoviruses to replicate in humans. Therefore the detection of viral DNA of unknown infectivity in a live attenuated oral vaccine, from a virus not shown to infect humans, may not be of concern. If contamination of the vaccine producer Vero cells with PCV1 did occur their replacement with fresh Vero cells should remove the source of PCV1.

Production of live attenuated vaccines, requiring viral amplification, may results in reversion mutations.

In the case of OPV these mutations may match those found in VDPV. No virulence-associated mutations were detected as minority variants following deep sequencing of the three poliovirus serotypes in the trivalent OPV viral quasispecies tested. As is well documented, the rare case of reversion of attenuated OPV to VDPV therefore likely occur during the multiple rounds of replication occurring in vivo following vaccination, typically in immunocompromised vaccinees or their infected contacts (18, 24). Beside polioviruses the attenuated mumps virus and VZV were also sequenced at depth sufficient to derive accurate genome sequences and identify minority variants. Mumps virus and VSV both showed very low rates of mutations when compared to the published genomes of attenuated vaccine strains. There is no evidence that any of the sequence change detected affects virus attenuation.

The use of metagenomics and microarray technologies for the detection of adventitious viral contamination can detect a more diverse range of viruses than possible with the currently mandated methods of cell cultures and viral species specific PCR. PCR is more sensitive than either metagenomics or microarrays when screening for specific viruses but the sheer number of potential contaminating viruses limits its use to a small number of suspected viruses. The use of viral metagenomics and microarrays tests therefore seems warranted for the surveillance of products that are derived from cell cultures, plasma pools, or other biological sources and which may contain and transmit adventitious viruses.

Despite an extensive record of safety and efficacy, common misconceptions remain regarding vaccine safety resulting in reduced childhood vaccination and the resurgence of vaccine-preventable infections. Given that live attenuated viral vaccines are safe, effective, and relatively inexpensive their use against human and animal pathogens should be encouraged. The application of high throughput sequencing and microarrays provides more effective means to interrogate current and future vaccines for genetic variants of the attenuated viruses and the presence of adventitious viruses. The wider range of sequences detectable by these newer methods (endogenous retroviruses, bacterial and other nucleic acids whose taxonomic origin cannot not be determined, and adventitious viruses such as PCV1) is an expected outcome of closer scrutiny to the nucleic acids present in vaccines and not necessarily a reflection of unsafe products. In view of the demonstrated

benefit and safety of Rotarix the implications (if any) on current immunization policies of the detection of PCV1 DNA of unknown infectivity for humans needs to be carefully considered.

Added note: Recent testing by GSK indicated that PCV1 was also present in the early lots of Rotarix used in the extensive clinical trials that demonstrated the safety and efficacy of this vaccine. These trials indicate a lack of detectable pathogenic effects from PCV1 DNA on vaccinees.

Acknowledgements

We thank for support NHLBI R01HL083254 grant to ELD, Blood Systems Research Institute, Laboratory Directed Research and Development Program at the Lawrence Livermore National Laboratory (Project LLNL 02-SI-008 under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344), and U.S. Air Force Surgeon General-approved Clinical Investigation no. FDG20040024E. We thank Dr Michael P. Busch for encouragements, James B. Thissen for technical assistance in microarray experiments, and T. N. Dhole Dept of Microbiology, SGPGIMS, Lucknow, India for providing OPV vaccine. The views expressed in this material are those of the authors and do not reflect the official policy or position of the U.S. Government or the Departments of Energy, Defense, or the Air Force.

Author Contributions

JGV designed the study, performed lab work and bioinformatics analyses, and wrote the manuscript; CW performed the metagenomics bioinformatics analysis; MSJ assisted with the study design; CJ performed the microarray lab work; KM performed the microarray bioinformatics analyses; SG designed the probes on the microarray; ED designed the study and wrote the manuscript.

Figure Legends

FIG. 1: Genomic Coverage: Depth of sequence coverage at each genomic positions for OPV, Attenuvax, YF-Vax, Varivax, MMR-II, Rotarix (PCV1), and Rotateq. For vaccines containing multiple genomes (MMRII) or multiple segments (Rotateq) of variable lengths, coverage is normalized to genetic position as percentage of fragment length. Depth of coverage for rotavirus and rubella virus from Rotarix and Meruvax-II respectively are not shown due to low coverage. PCV1 coverage depth in Rotarix is shown.

FIG. 2: Taxonomic distribution of sequences in 8 live attenuated viral vaccines (A-H) and water (I): reads were classified using BLASTx as either viral (V), eukaryotic (E), or bacterial/unclassifiable (E/U)(see materials and methods). Vaccine viral input for nuclease treatment and viral particle extraction is listed based on manufacturers' reported titers as either 50% tissue culture infectious dose ($TCID_{50}$), plaque forming units (PFU), or infectious units (IU).

References

1. **Allan G. M., D. P. Mackie, J. McNair, B. M. Adair, and M. S. McNulty.** 1994. Production, preliminary characterisation and applications of monoclonal antibodies to porcine circovirus. Vet Immunol Immunopathol **43**:357-371.
2. **Allan G. M., F. McNeilly, I. McNair, M. D. Curran, I. Walker, J. Ellis, C. Konoby, S. Kennedy, and B. Meehan.** 2000. Absence of evidence for porcine circovirus type 2 in cattle and humans, and lack of seroconversion or lesions in experimentally infected sheep. Arch Virol **145**:853-857.
3. **Allan G. M., K. V. Phenix, D. Todd, and M. S. McNulty.** 1994. Some biological and physico-chemical properties of porcine circovirus. Zentralbl Veterinarmed B **41**:17-26.
4. **Allander T., S. U. Emerson, R. E. Engle, R. H. Purcell, and J. Bukh.** 2001. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. Proc Natl Acad Sci U S A **98**:11609-11614.
5. **Allen A.** 2007. Vaccine : the controversial story of medicine's greatest lifesaver . W.W. Norton, New York.
6. **Bernstein C. N., G. Nayar, A. Hamel, and J. F. Blanchard.** 2003. Study of animal-borne infections in the mucosas of patients with inflammatory bowel disease and population-based controls. J Clin Microbiol **41**:4986-4990.
7. **Delwart E. L.** 2007. Viral metagenomics. Rev Med Virol **17**:115-131.
8. **Dunwiddie C. T., R. Resnick, M. Boyce-Jacino, J. N. Alegre, and A. J. Faras.** 1986. Molecular cloning and characterization of gag-, pol-, and env-related gene sequences in the ev- chicken. J Virol **59**:669-675.
9. **Dunwiddie C., and A. J. Faras.** 1985. Presence of retrovirus reverse transcriptase-related gene sequences in avian cells lacking endogenous avian leukosis viruses. Proc Natl Acad Sci U S A **82**:5097-5101.
10. **Engels E. A.** 2005. Cancer risk associated with receipt of vaccines contaminated with simian virus 40: epidemiologic research. Expert Rev Vaccines **4**:197-206.
11. **Gardner S. N., A. L. Hiddessen, P. L. Williams, C. Hara, M. C. Wagner, and B. W. Colston.** 2009. Multiplex primer prediction software for divergent targets. Nucleic Acids Res **37**:6291-6304.
12. **Hattermann K., A. Maerz, H. Slanina, C. Schmitt, and A. Mankertz.** 2004. Assessing the risk potential of porcine circoviruses for xenotransplantation: consensus primer-PCR-based search for a human circovirus. Xenotransplantation **11**:547-550.
13. **Hattermann K., C. Roedner, C. Schmitt, T. Finsterbusch, T. Steinfeldt, and A. Mankertz.** 2004. Infection studies on human cell lines with porcine circovirus type 1 and porcine circovirus type 2. Xenotransplantation **11**:284-294.
14. **Hussain A. I., J. A. Johnson, M. Da Silva Freire, and W. Heneine.** 2003. Identification and characterization of avian retroviruses in chicken embryo-derived yellow fever vaccines: investigation of transmission to vaccine recipients. J Virol **77**:1105-1111.
15. **Hussain A. I., V. Shanmugam, W. M. Switzer, S. X. Tsang, A. Fadly, D. Thea, R. Helfand, W. J. Bellini, T. M. Folks, and W. Heneine.** 2001. Lack of evidence of endogenous avian leukosis virus and endogenous avian retrovirus transmission to measles, mumps, and rubella vaccine recipients. Emerg Infect Dis **7**:66-72.
16. **Jaing C., S. Gardner, K. McLoughlin, N. Mulakken, M. Alegria-Hartman, P. Banda, P. Williams, P. Gu, M. Wagner, C. Manohar, and T. Slezak.** 2008. A functional gene array for detection of bacterial virulence elements. PLoS ONE **3**:e2163.
17. **Johnson J. A., and W. Heneine.** 2001. Characterization of endogenous avian leukosis viruses in chicken embryonic fibroblast substrates used in production of measles and mumps vaccines. J Virol **75**:3605-3612.
18. **Kew O. M., R. W. Sutter, E. M. de Gourville, W. R. Dowdle, and M. A. Pallansch.** 2005. Vaccine-derived polioviruses and the endgame strategy for global polio eradication. Annu Rev Microbiol **59**:587-635.
19. **Khan A. S., T. Maudru, A. Thompson, J. Muller, J. F. Sears, and K. W. Peden.** 1998. The reverse transcriptase activity in cell-free medium of chicken embryo fibroblast cultures is not associated with a replication-competent retrovirus. J Clin Virol **11**:7-18.

20. **Lerche N. W., W. M. Switzer, J. L. Yee, V. Shanmugam, A. N. Rosenthal, L. E. Chapman, T. M. Folks, and W. Heneine.** 2001. Evidence of infection with simian type D retrovirus in persons occupationally exposed to nonhuman primates. J Virol **75**:1783-1789.
21. **Li L., A. Kapoor, B. Slikas, O. S. Bamidele, C. Wang, S. Shaukat, M. A. Masroor, M. L. Wilson, J. B. Ndjango, M. Peeters, N. D. Gross-Camp, M. N. Muller, B. H. Hahn, N. D. Wolfe, H. Triki, J. Bartkus, S. Z. Zaidi, and E. Delwart.** 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. J Virol **84**:1674-1682.
22. **Margeridon-Thermet S., N. S. Shulman, A. Ahmed, R. Shahriar, T. Liu, C. Wang, S. P. Holmes, F. Babrzadeh, B. Gharizadeh, B. Hanczaruk, B. B. Simen, M. Egholm, and R. W. Shafer.** 2009. Ultra-deep pyrosequencing of hepatitis B virus quasispecies from nucleoside and nucleotide reverse-transcriptase inhibitor (NRTI)-treated patients and NRTI-naive patients. J Infect Dis **199**:1275-1285.
23. **Marx P. A., and A. F. Voevodin.** 2009. Simian Virology. Wiley-Blackwell.
24. **Minor P. D.** 1993. Attenuation and reversion of the Sabin vaccine strains of poliovirus. Dev Biol Stand **78**:17-26.
25. **Opriessnig T., A. R. Patterson, X. J. Meng, and P. G. Halbur.** 2009. Porcine circovirus type 2 in muscle and bone marrow is infectious and transmissible to naïve pigs by oral consumption. Vet Microbiol **133**:54-64.
26. **Rezapkin G. V., L. P. Norwood, R. E. Taffs, E. M. Dragunsky, I. S. Levenbook, and K. M. Chumakov.** 1995. Microevolution of type 3 Sabin strain of poliovirus in cell cultures and its implications for oral poliovirus vaccine quality control. Virology **211**:377-384.
27. **Robertson J. S., C. Nicolson, A. M. Riley, M. Bentley, G. Dunn, T. Corcoran, G. C. Schild, and P. Minor.** 1997. Assessing the significance of reverse transcriptase activity in chick cell-derived vaccines. Biologicals **25**:403-414.
28. **Seeff L. B., G. W. Beebe, J. H. Hoofnagle, J. E. Norman, Z. Buskell-Bales, J. G. Waggoner, N. Kaplowitz, R. S. Koff, J. L. Petrini, and E. R. Schiff.** 1987. A serologic follow-up of the 1942 epidemic of post-vaccination hepatitis in the United States Army. N Engl J Med **316**:965-970.
29. **Shahabuddin M., J. F. Sears, and A. S. Khan.** 2001. No evidence of infectious retroviruses in measles virus vaccines produced in chicken embryo cell cultures. J Clin Microbiol **39**:675-684.
30. **Sommerfelt M. A., N. Harkestad, and E. Hunter.** 2003. The endogenous langur type D retrovirus PO-1-Lu and its exogenous counterparts in macaque and langur monkeys. Virology **315**:275-282.
31. **Tillieux S. L., W. S. Halsey, E. S. Thomas, J. J. Voycik, G. M. Sathe, and V. Vassilev.** 2008. Complete DNA sequences of two oka strain varicella-zoster virus genomes. J Virol **82**:11023-11044.
32. **Tillieux S. L., W. S. Halsey, G. M. Sathe, and V. Vassilev.** 2009. Comparative analysis of the complete nucleotide sequences of measles, mumps, and rubella strain genomes contained in Priorix-Tetra and ProQuad live attenuated combined vaccines. Vaccine **27**:2265-2273.
33. **Tischer I., L. Bode, J. Apodaca, H. Timm, D. Peters, R. Rasch, S. Pociuli, and E. Gerike.** 1995. Presence of antibodies reacting with porcine circovirus in sera of humans, mice, and cattle. Arch Virol **140**:1427-1439.
34. **Tsang S. X., W. M. Switzer, V. Shanmugam, J. A. Johnson, C. Goldsmith, A. Wright, A. Fadly, D. Thea, H. Jaffe, T. M. Folks, and W. Heneine.** 1999. Evidence of avian leukosis virus subgroup E and endogenous avian virus in measles and mumps vaccines derived from chicken cells: investigation of transmission to vaccine recipients. J Virol **73**:5843-5851.
35. **van der Kuyl A. C., R. Mang, J. T. Dekker, and J. Goudsmit.** 1997. Complete nucleotide sequence of simian endogenous type D retrovirus with intact genome organization: evidence for ancestry to simian retrovirus and baboon endogenous virus. J Virol **71**:3666-3676.
36. **Victoria J. G., A. Kapoor, L. Li, O. Blinkova, B. Slikas, C. Wang, A. Naeem, S. Zaidi, and E. Delwart.** 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol **83**:4642-4651.
37. **Wang C., Y. Mitsuya, B. Gharizadeh, M. Ronaghi, and R. W. Shafer.** 2007. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res **17**:1195-1201.

38. **Weissmahr R. N., J. Schüpbach, and J. Böni.** 1997. Reverse transcriptase activity in chicken embryo fibroblast culture supernatants is associated with particles containing endogenous avian retrovirus EAV-0 RNA. J Virol **71**:3005-3012.
39. **Wong C. W., C. L. Heng, L. Wan Yee, S. W. Soh, C. B. Kartasasmita, E. A. Simoes, M. L. Hibberd, W. K. Sung, and L. D. Miller.** 2007. Optimization and clinical validation of a pathogen detection microarray. Genome Biol **8**:R93.

8

TABLE 1 **Pyrosequencing coverage of viral genomes and PCR detection of endogenous retroviruses and PCV.**

Vaccine	Virus	Total Reads	Viral Sequences	Genomic Coverage	Avg. Coverage	Non-vaccine Viruses	PCR Test			RT-PCR Test*	
							ALV	SRV	PCV1	ALV	SRV
	Polio 1	150899	40.25%	99.5%	2009X						
OPV	Polio 2	150899	14.08%	99.6%	639X	None	-	-	-	-	-
	Polio 3	150899	39.72%	97.0%	2425X						
Meruvax-II	Rubella	14389	0.97%	41.8%	<1X	None	-	-	-	-	-
Attenuvax	Measles	9863	11.96%	95.9%	21X	ALV	++	-	-	++	-
YF-Vax	Yellow Fever	30770	1.43%	78.3%	11X	None	+	-	-	+	-
Varivax	Human Herpes 3	116905	89.00%	99.0%	287X	None	-	-	-	-	-
	Measles	39646	1.19%	95.8%	9X						
MMR-II	Mumps	39646	6.95%	99.9%	55X	None	+	-	-	+	-
	Rubella	39646	>0.01%	28.4%	<1X						
Rotarix	Rotavirus	15249	>0.01%	20.6%	<1X	PCV-1	-	+	+	-	-
Rotateq	Rotavirus	18671	2.45%	92.6%	9X	SRV-1	-	++	-	-	-

*Samples were treated with DNase prior to reverse transcription to eliminate proviral DNA

9

0

Table 2

Detection of viral sequences using pan-microbial microarray

Vaccine	Processing	Virus	Log-Odds Score	Probe Hybridization
Attenuvax CEF	TOT[1]	Measles Virus	679.9	107/155
		Avian Leukosis Virus	492.8	103/169
		Avian Endogenous Retrovirus	276.2	43/59
	VP[2]	Measles	673.6	105/154
YF-Vax CEF	TOT	Yellow Fever Virus	621.4	88/120
		Avian Endogenous Retrovirus	437.6	57/60
	VP	Yellow Fever Virus	628.3	89/120
		Avian Endogenous Retrovirus	111.7	26/59
Varivax MRC-5 (human)	TOT	Human Herpes 3	608.6	103/130
		Human Endogenous Retrovirus K	230.5	26/28
	VP	Human Herpes 3	614.7	104/130
		Human Endogenous Retrovirus K	199.7	26/28
Rotarix Vero E6 (AGM)	TOT	Rotavirus A	482.3	232/802
		Porcine Circovirus 2	533.7	105/120
	VP	Rotavirus A	2.4	203/802
		Porcine Circovirus 2	505.2	96/103
Rotateq Vero (AGM)	TOT	Rotavirus A	3559.8	490/802
		Baboon Endogenous Retrovirus	256.5	44/82
	VP	Rotavirus A	2954.7	489/802
MMR-II CEF and WI-38 (human)	TOT (1st prep)	Measles Virus	688.8	104/154
		Mumps Virus	555.7	75/85
		Human Endogenous Retrovirus K	222.3	22/28
		Avian Endogenous Retrovirus	159.7	31/60
	TOT (2nd prep)	Rubella Virus	58.1	26/74
		Human Endogenous Retrovirus K	229.6	26/28
	VP	Not done		n/a
Meruvax-II WI-38 (human)	TOT	Rubella Virus	260.4	49/74
		Human Endogenous Retrovirus K	247.3	26/28
	VP	Rubella Virus	57.8	28/74

1. TOT refers to total nucleic acids directly extracted from the vaccine suspensions.
2. VP refers to nucleic acids extracted from the vaccine suspensions following viral particles purification.

1

Table 3 **Absence of mutations associated with reversion to virulence**

	Genomic Region	Nucleotide Position	Wild Type	Vaccine Strain	Amino acid change	454	Coverage
Poliovirus 1	5'-UTR	480	A	G	N/A	G	663X
	VP3	1944	C	A	K → T	A	1161X
	VP1	2775	C	A	L → T	A	1827x
	VP1	2795	G	A	T → A	A	1870X
Poliovirus 2	5'-UTR	481	G	A	N/A	A	634X
	VP1	2908	G	A	I → V	A	435X
Poliovirus 3	5'-UTR	472	C	U	N/A	U	1645X
	VP3	2034	C	U	F → S	U	530X
	VP1	2636	A	G	A → T	G	843X

2

3

3

Table 4 **Detection of poliovirus minority variants**

	Genomic Region	Nucleotide Position	Mutation	Frequency (%)	Coding
Poliovirus 1	5'-UTR	442-443	GAG insertion	38/50 (76%)	non-coding
	5'-UTR	565	T → C	43/487 (9%)	non-coding
	3C	5983	T → C	106/420 (25%)	Synonymous
Poliovirus 2	5'-UTR	619	G → A	2/89 (2%)	non-coding
	2C	4722	T → C	85/491 (17%)	Synonymous
	2C	4965	T → C	24/209 (11%)	Synonymous
	3D	6084	G → DEL	9/143 (6%)	Frameshift
	3D	6087	T → DEL	9/143 (6%)	Frameshift
Poliovirus 3	3D	6902	T → DEL	461/1495(31%)	Frameshift
	3D	6903	T → DEL	462/1557(30%)	Frameshift

4
5

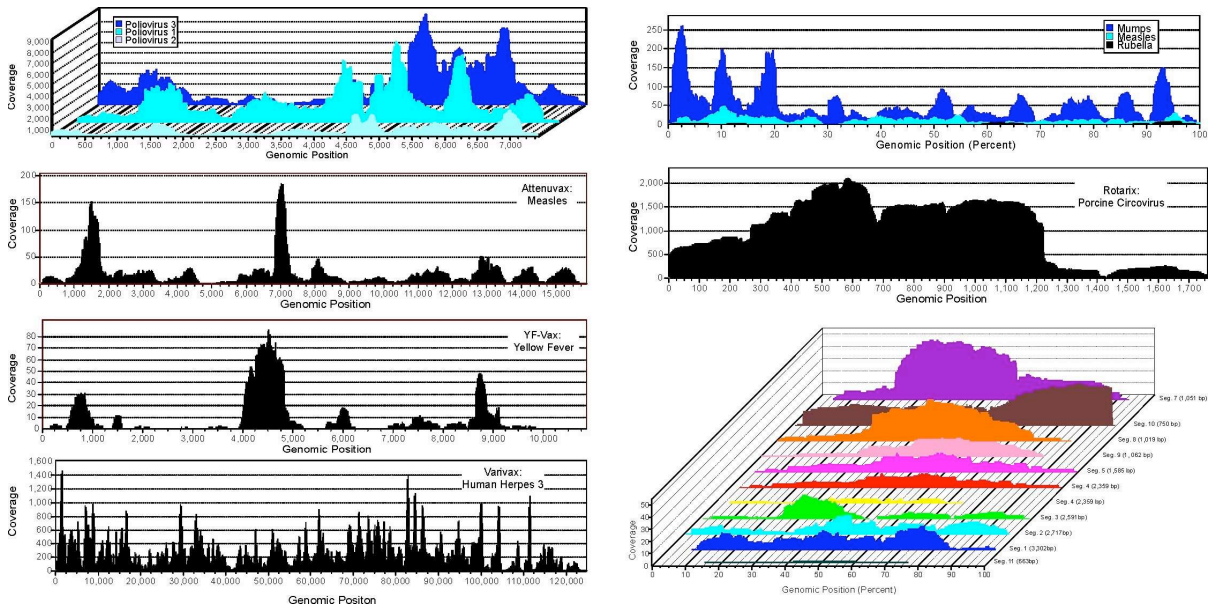

Figure 1

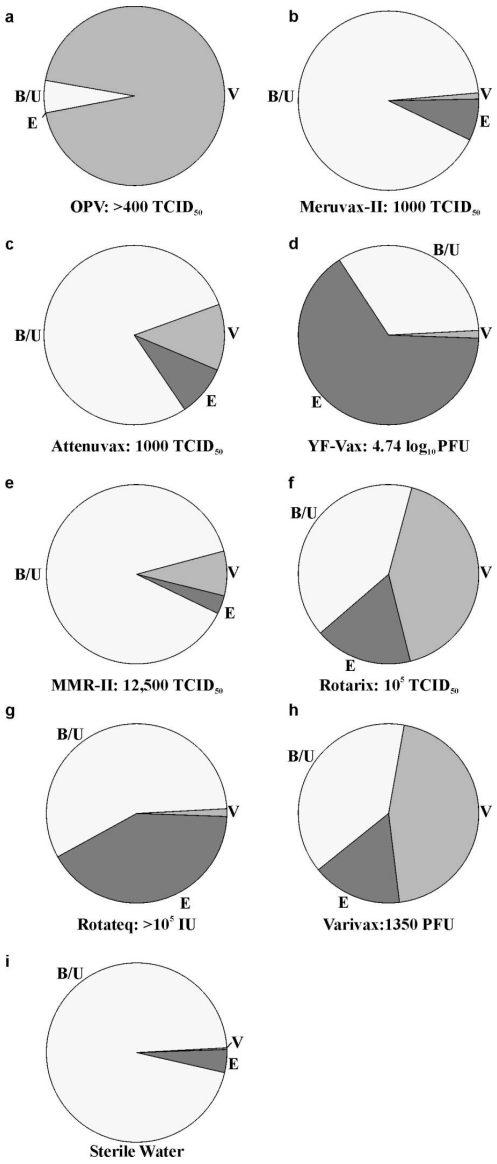

Figure 2