

Sequencing of bivalent Moderna and Pfizer mRNA vaccines reveals nanogram to microgram quantities of expression vector dsDNA per dose

Kevin McKernan, Yvonne Helbert, Liam T. Kane, Stephen McLaughlin
Medicinal Genomics, 100 Cummings Center, Suite 406-L, Beverly Mass, 01915

Several methods were deployed to assess the nucleic acid composition of four expired vials of the Moderna and Pfizer bivalent mRNA vaccines. Two vials from each vendor were evaluated with Illumina sequencing, qPCR, RT-qPCR, Qubit™ 3 fluorometry and Agilent Tape Station™ electrophoresis. Multiple assays support DNA contamination that exceeds the European Medicines Agency (EMA) 330ng/mg requirement and the FDAs 10ng/dose requirements. These data may impact the surveillance of vaccine mRNA in breast milk or plasma as RT-qPCR assays targeting the vaccine mRNA cannot discern DNA from RNA without RNase or DNase nuclease treatments. Likewise, studies evaluating the reverse transcriptase activity of LINE-1 and vaccine mRNA will need to account for the high levels of DNA contamination in the vaccines. The exact ratio of linear fragmented DNA versus intact circular plasmid DNA is still being investigated. Quantitative PCR assays used to track the DNA contamination are described.

Introduction

Several studies have made note of prolonged presence of vaccine mRNA in breast milk and plasma (Bansal et al. 2021; Hanna et al. 2022; Castruita et al. 2023). This could be the result of the stability of N1-methylpseudouridine (m1Ψ) in the mRNA of the vaccine. Nance *et al.* depict a vaccine mRNA synthesis method that utilizes a dsDNA plasmid that is first amplified in *E.coli* prior to an *in-vitro* T7 polymerase synthesis of vaccine mRNA (Nance and Meier 2021). Failure to remove this DNA could result in the injection of spike encoded nucleic acids more stable than the modified RNA. The EMA has stated limits at 330ng/mg of DNA to RNA (Josephson 2020-11-19). The FDA has issued guidance for under 10ng/dose in vaccines (Sheng-Fowler et al. 2009). Residual injected DNA can result in type I interferon responses and can increase the potential for DNA integration (Ulrich-Lewis et al. 2022).

Results

To assess the nucleic acid composition of the vaccines, vaccine DNA was deeply sequenced using two different methods. The first method used a commercially available New England Biolabs RNA-seq method that favored the sequencing of the RNA but still presented over 500X coverage for the unanticipated DNA vectors (Figure 1 and 2). The RNA-seq assemblies had truncated poly A tracts compared to the constructs described by Nance *et al.* The second method eliminated the RNA with RNase A treatment and sequenced only the DNA using a Watchmaker Genomics fragment library kit. The DNA focused assemblies delivered vector assemblies with more intact poly A tracts (Figure 3).

These assemblies were utilized to design multiplex qPCR and RT-qPCR assays that target the spike sequence present in both the vaccine mRNA and the DNA vector while also targeting the origin of replication sequence present only in the DNA vector (Figure 3). The assembly of Pfizer vial 1 contains a 72bp insertion not present in the assembly of Pfizer vial 2. This indel is known

for its enhancement to the SV40 promoter and its nuclear localization signal (Dean et al. 1999) (Moreau et al. 1981).

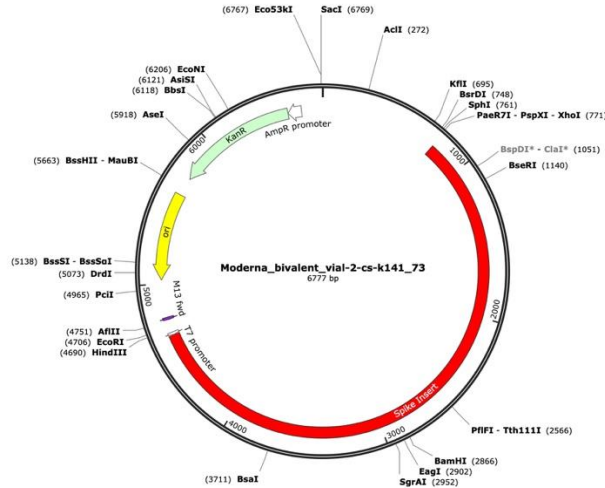


Figure 1. A Moderna vector assembly of an RNA-seq library with a spike insert (red), Kanamycin resistance gene (green) driven by an AmpR promoter and a high copy bacterial origin of replication (yellow).

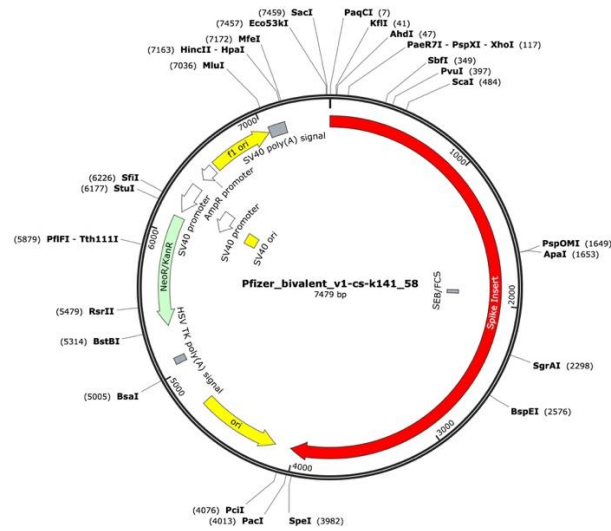


Figure 2. Pfizer bivalent vaccine assembly of the RNA-seq library. Annotated with SEB/FCS, spike insert (red), bacterial origin of replication (yellow), Neo/Kan resistance gene (green), F1 origin (yellow) and an SV40 promoter (yellow and white).

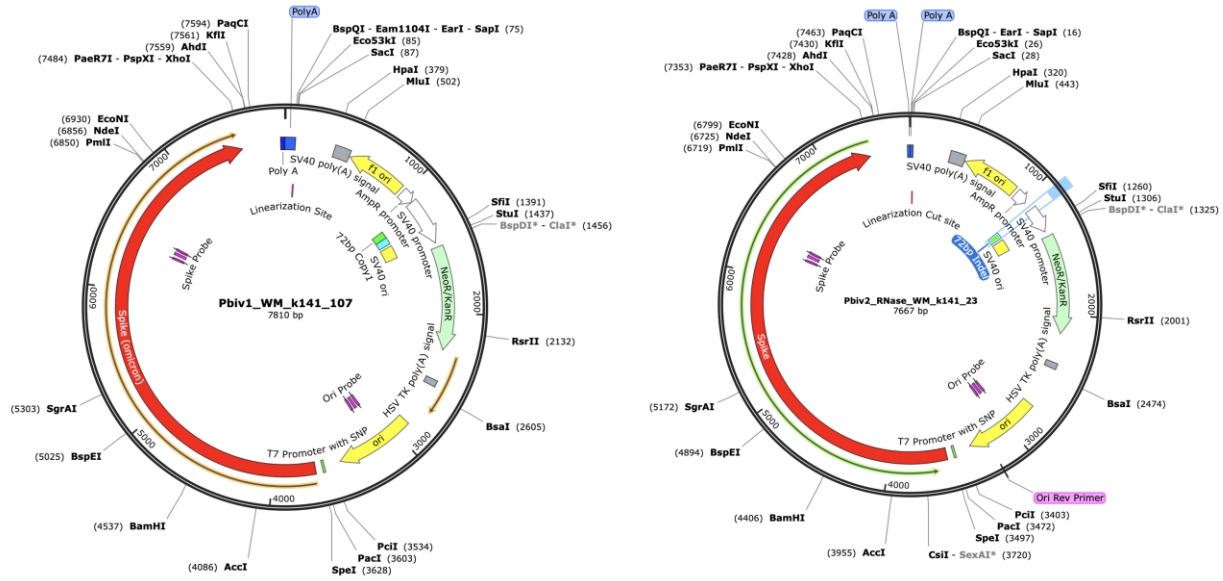


Figure 3. RNase treated vaccines were shotgun sequenced with Illumina (RNase-Seq not RNA-seq). Pfizer vectors from vial 1 (left) and vial 2 (right) contain a 72bp difference in the SV40 promoter (green and light blue annotation). qPCR assays are depicted in pink as Spike probe and Ori probe. The RNase sequencing provided better resolution over the Eam1104i linearization site and the Poly adenylation sequence. The vectors differ in the length of the polyA tail (likely sequencing artifact) and the 72bp indel.

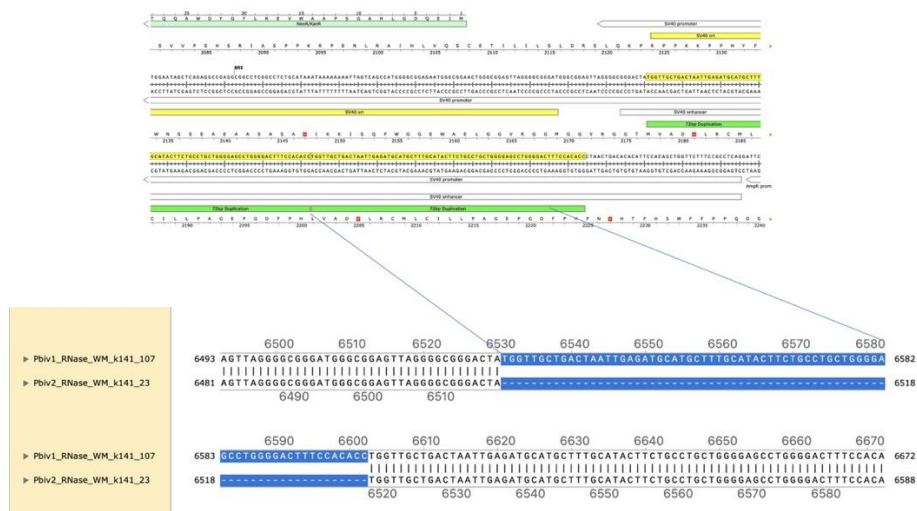


Figure 4. Local alignment of Pfizer vial 1 to Pfizer vial 2 vectors highlights the 72bp tandem duplication in blue.

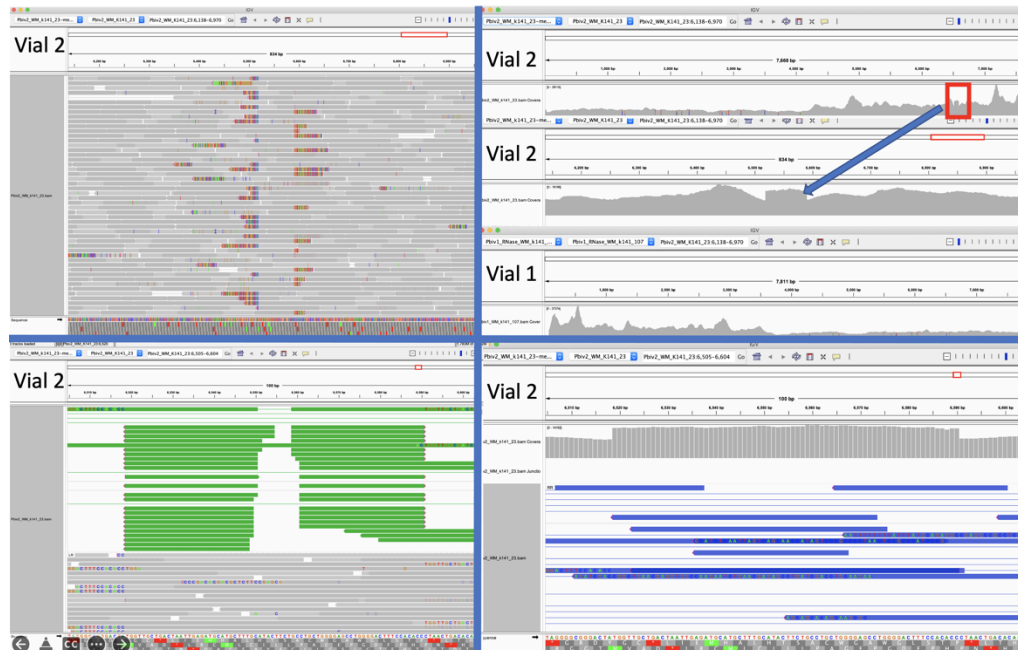


Figure 5A. Close inspection of the Integrative Genome Viewer (IGV) demonstrates the appearance of a 72bp insertion that is heteroplasmic in Pfizer vial 2. The upper left IGV view is a zoomed-out view where the colored marks depict the indel. The lower Left IGV view shows inverted paired reads as the 72bp insertion is a tandem repeat and paired reads shorter than 72bp can be mapped two different ways. Upper Right IGV view demonstrates a read coverage pile up or ‘Plateau’. This occurs when the reference has one copy of the 72bp repeat and the sample has 2 copies. Note- In the upper right IGV depiction, the sequence in Vial 1 is in the opposite orientation in IGV as Vial 2. Lower right IGV view is a zoomed view of the upper right IGV screen.

Since the two Pfizer vials share the same lot number, finding a heterozygous copy number change between the two vials is unexpected. It was hypothesized that the appearance of a heteroplasmic copy number change is instead the result of the Megahit assembler collapsing what is actually two copies of the 72bp sequence into a single copy due to the insert sizes in the sequencing libraries being too short (105bp). It is noteworthy that the longer paired-end reads in the library resolve the 72bp tandem repeat.

When references have a single copy of the 72bp repeat and the sample has two copies of the repeat, reads should pile up to twice the coverage over the single copy 72bp loci as seen in Figure 5A. To test this hypothesis, we added a second 72bp sequence to the shorter plasmid assembly and observed that the reads map without artifact and no evidence of heteroplasmy (Figure 5B).

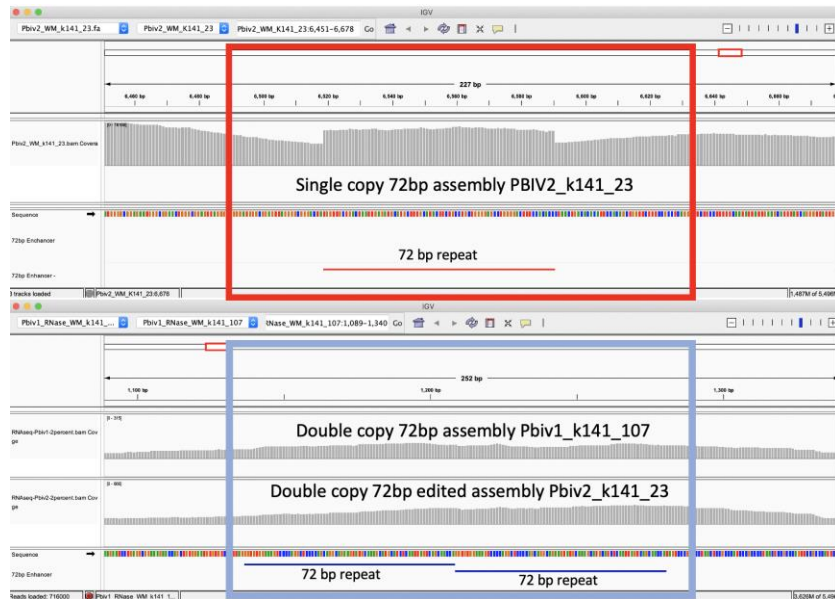


Figure 5B. IGV view of the read coverage over Pbv2_k141_23 shows a discrete 72bp plateau in coverage (red rectangle). Editing the Pbv2_k141_23 reference to include 2 copies of the 72bp sequence, and remapping the sequence data to this corrected sequence shows that the coverage over both vectors is more normal with no coverage plateau in Pfizer vial 2.

These data conclude that all Pfizer vectors contain a homoplasmic 2 copy 72bp SV40 Enhancer associated with more robust expression and nuclear localization. The initial heteroplasmic indel was an artifact of the Megahit assembler and short insert libraries.

To estimate the size of the DNA, the purified vaccines were evaluated on an Agilent Tape Station™ using DNA (genomic DNA screen tapes) and RNA based (high sensitivity RNA tapes) electrophoresis tapes.

Agilent Tape Station™ electrophoresis reveal 7.5 - 11.3 ng/μl of dsDNA compared to the 23.7 - 55.9ng/μl of mRNA detected in each 300μl sample. Qubit™ 3 fluorometry estimated 1-2.8ng/μl of DNA and 21.8ng - 52.8ng/μl of RNA. There is higher fragmentation seen in the DNA electrophoresis. The total RNA levels are less than the anticipated 30ug (100ng/μl) and 100ug (200ng/μl) doses suggesting a loss of yield in DNA and RNA isolation, manufacturing variance or RNA decay with expired lots.

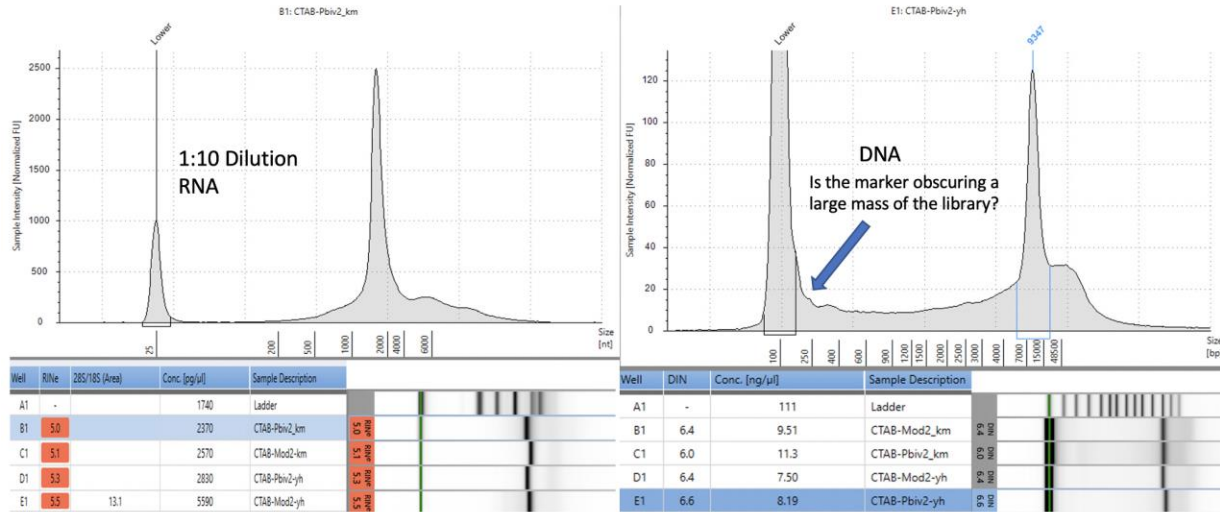
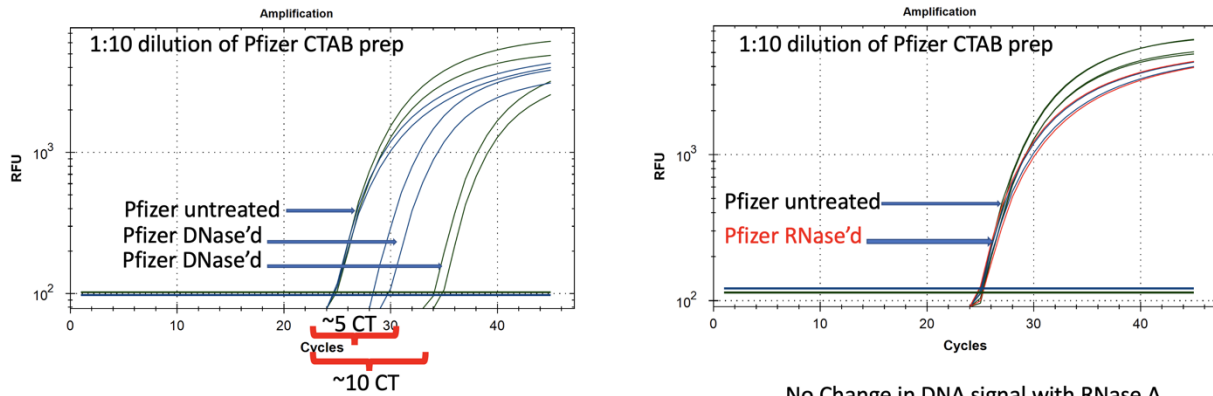


Figure 6. Agilent Tape Station™ electrophoresis demonstrates 23.7ng/μl – 55.9ng/μl of RNA (left). 7.5ng-11.3ng/μl are observed on DNA based Tape Station™. While the DNA electropherogram shows a peak suggestive of a full-length plasmid, this sample is known to have high amounts of N1-methylpseudouridine RNA present. DNA hybrids with N1-methylpseudouridine mRNA may provide enough intercalating dye cross talk to produce a peak. The sizing of the peak on the RNA tape on the left is shorter than expected. This may be the results of N1 methylpseudouridine changing the secondary structure or the mass to charge ratio of the DNA.

Quantitative PCR assays were designed using IDTs Primer Quest software targeting a region in the spike protein that was identical between Moderna and Pfizer spike sequences and a shared sequence in the vectors' origin of replication. This allowed the qPCR and RT-qPCR assessment of the vaccines. qPCR only amplifies DNA while RT-qPCR amplifies both DNA and RNA. Gradient qPCR was utilized to explore conditions where both targets would perform under the same cycling conditions for both RT-qPCR and PCR (gradient PCR data not shown).

Multiplex qPCR targeting Spike (Blue) and Vector Origin (Green)

qPCR Amplifies **ONLY** DNA



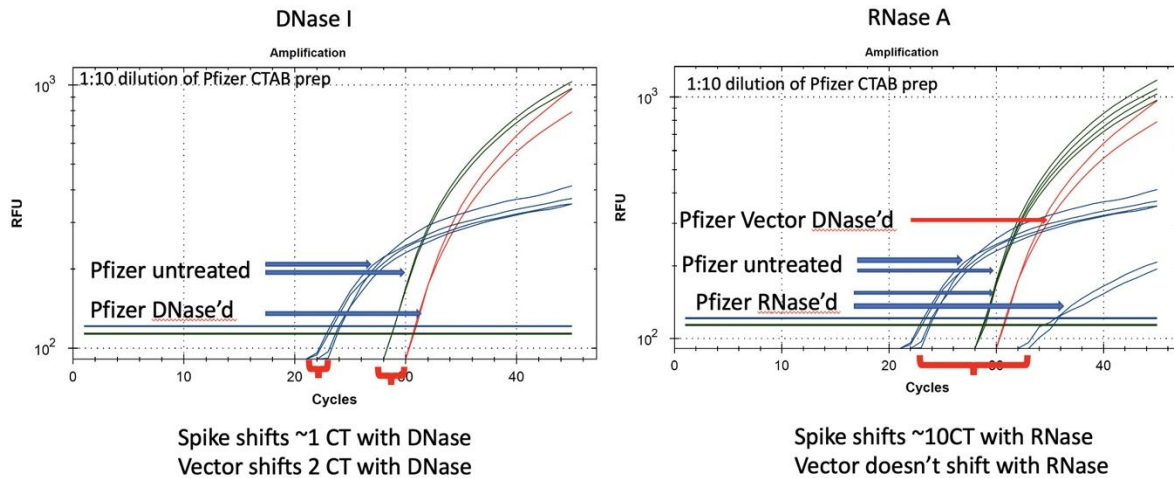
5 CT shift on Spike = 32 fold reduction in DNA with DNase I.
 10CT shift on Vector = 1000 fold reduction in DNA with DNase I
 qPCR does not amplify RNA

No Change in DNA signal with RNase A

Figure 7. qPCR of Pfizer’s bivalent vaccine with and without DNase I (left) and RNase A (right). Untreated mRNA demonstrates equal CTs for Spike and Vector assays as expected. Vector is more DNase I sensitive than the Spike suggesting the modRNA may inhibit nuclease activity of DNase I against complementary DNA targets. RNase A treatment doesn’t alter the qPCR signal.

Multiplex **RT-qPCR** targeting Spike (Blue) and Vector Origin (Green)

RT qPCR Amplifies **BOTH** RNA and DNA



Spike shifts ~1 CT with DNase
 Vector shifts 2 CT with DNase

Spike shifts ~10CT with RNase
 Vector doesn’t shift with RNase

Figure 8. RT-qPCR amplifies both DNA and RNA. The untreated samples show a large CT offset with Pfizer Spike and Vector assays (Left Blue versus Green). This is anticipated as the T7 polymerization should create more mRNA over spike than over the vector. Small 1-2 CT shifts are seen with DNase I treatment. This is expected if the DNA is less than equal concentration of

nucleic acid in RT-PCR. RNase treatment (Right) shows a 10 CT offset but doesn't alter the DNA vector CT.

Pfizer qPCR Results

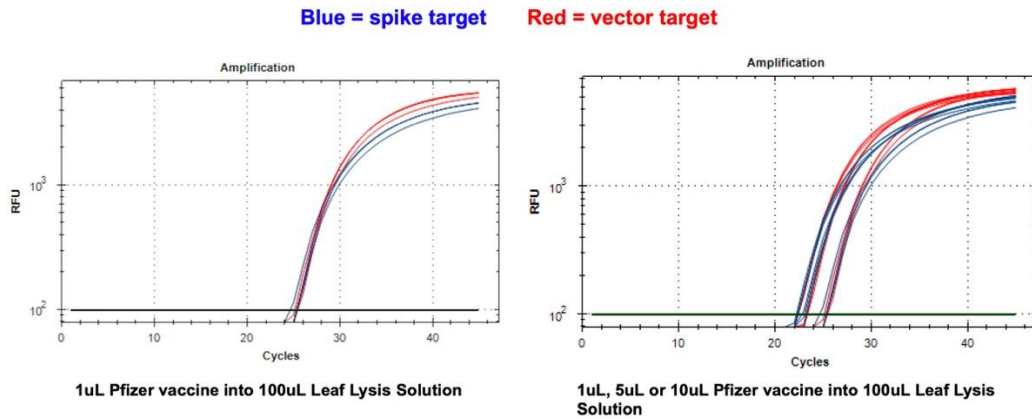


Figure 9. 1 μ l of the Pfizer bivalent vaccine placed in 100 μ l Leaf Lysis buffer for an 8 minute boil step delivers a CT of 24 for both Vector and Spike targets in qPCR (Left). Assay is responsive to 1,5,10 μ l of input (Right).

Pfizer RT-qPCR Results

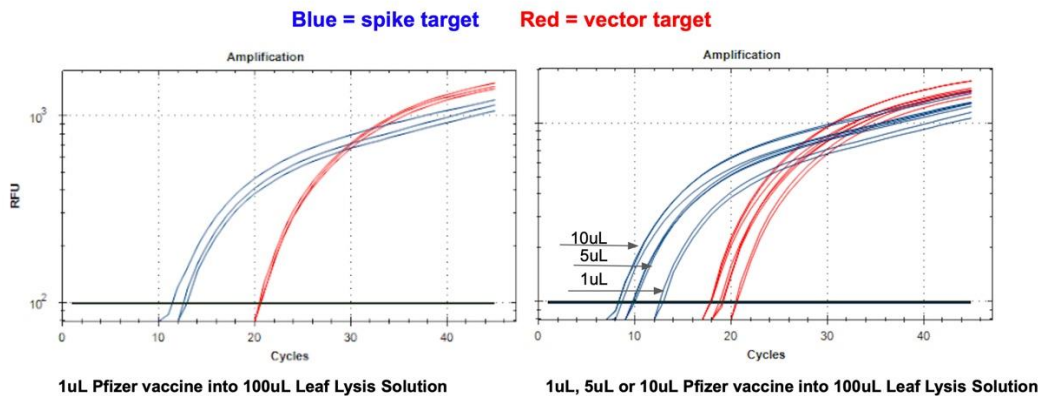


Figure 10. 1 μ l of the Pfizer bivalent vaccine placed in 100 μ l Leaf Lysis buffer for an 8 minute boil step delivers a CT of 20 and 12 for both Vector and Spike targets in RT-qPCR (Left). Assay is responsive to 1,5,10 μ l of input (Right).

Moderna qPCR Results

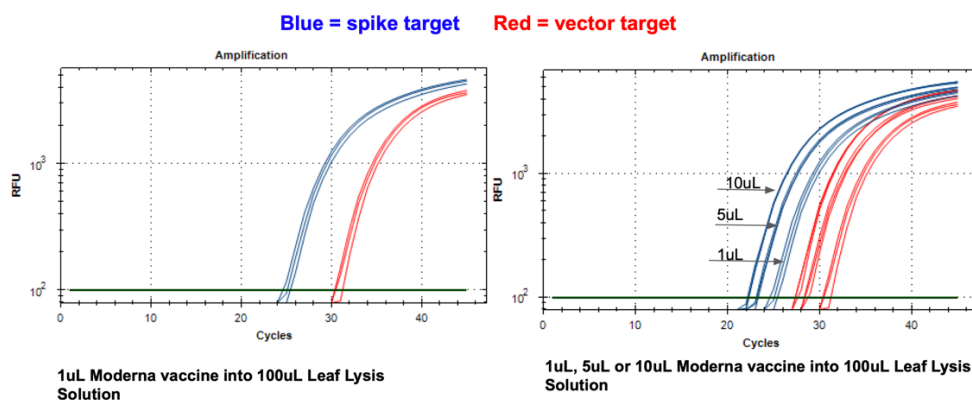


Figure 11. 1µl of the Moderna bivalent vaccine exhibits different CTs values for the spike and the vector targets (Left) with qPCR. This needs to be explored further as the assays provide equal CT scores on Pfizers' vaccines and the sequence of the amplicon is identical between the two vector origins. There are 2 mismatches in the spike amplicons between Moderna and Pfizer but none of the mismatches are under a primer or probe. The assay is responsive to 1,5,10µl of direct boil mRNA (Right).

Moderna RT-qPCR Results

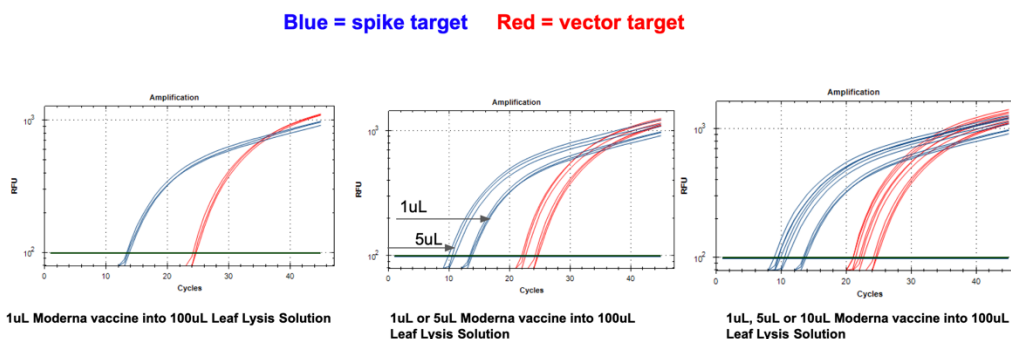


Figure 12. 1µl of the Moderna bivalent vaccine exhibits different CTs values for the spike and the vector targets (Left) with RT-qPCR. The large 10 CT shift between Spike and Vector needs to take into consideration that qPCR control shows a 5 CT offset. The boil preps can tolerate 1-10µl of vaccine (Middle and Right).

	Qubit DNA ng/µl	Qubit RNA ng/µl
Pbiv1	2.81	30.0
Pbiv2	1.47	52.8
Mod1	2.67	21.8
Mod2	1.04	49.0

Table 1. Qubit™ 3 Fluorometry estimates 1.04-2.8 ng/µl of dsDNA in the vaccines and 21.8ng-52.8ng/µl of RNA.

Synthetic templates were synthesized with IDT to build RT-qPCR standard curves to benchmark CTs to the mass of DNA in the reaction. This method uses ideal templates and fails to quantitate DNA molecules smaller than the amplicon size. As expected, this method delivers lower DNA concentration estimates than Qubit™ 3 fluorometry or Agilent Tape Station™. It also represents an ideal environment which doesn't capture the inhibition or primer depletion that can occur when large quantities of mRNA with identical sequence to your DNA target are co-present in a qPCR assay.

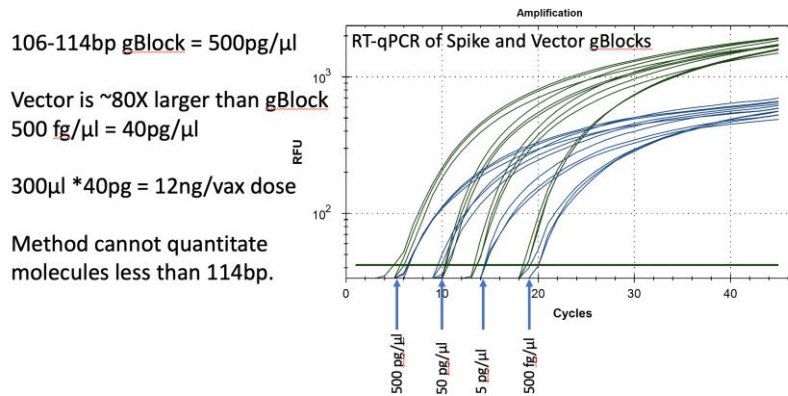


Figure 13. Two gBlocks were synthesized at IDT for Spike and Ori positive control templates used in an RT-qPCR assays. 10-fold serial dilutions were run in triplicate to correlate CT scores with picograms of DNA. The threshold is lowered from 10^2 for review of the background. CT of $\sim 20 = 500\text{fg}/\text{RT-qPCR}$ reaction. Since 100bp targets only represent $1/80^{\text{th}}$ of the vector DNA present as a potential contaminant, 500 fg/μl manifests in 40pg/μl of vector DNA. Any DNA that is DNase I treated and is smaller than the amplicon size cannot amplify or be quantitated with this method. This method will under quantitate DNase I treated samples compared to Qubit™ 3 or Agilent Tape Station™.

This work was further validated by testing 8 unopened Pfizer monovalent vaccines with both qPCR and RT-qPCR.

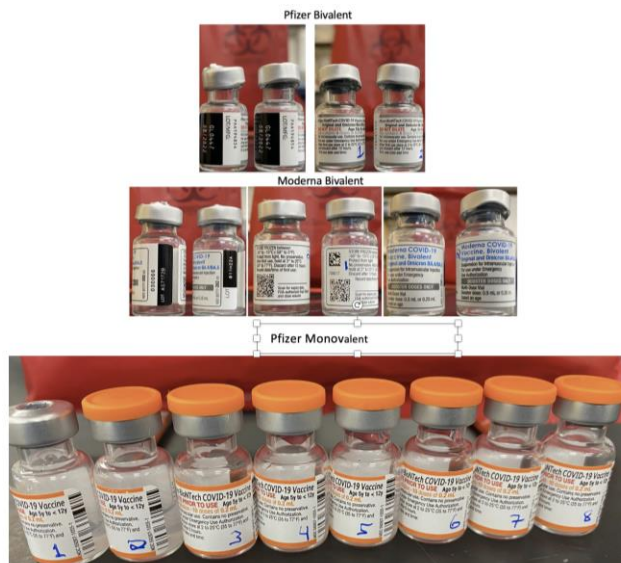


Figure 14. Moderna and Pfizer Bivalent vaccines (Top). 8 Monovalent Pfizer mRNA vaccines. These were unopened but past expiration (Bottom).

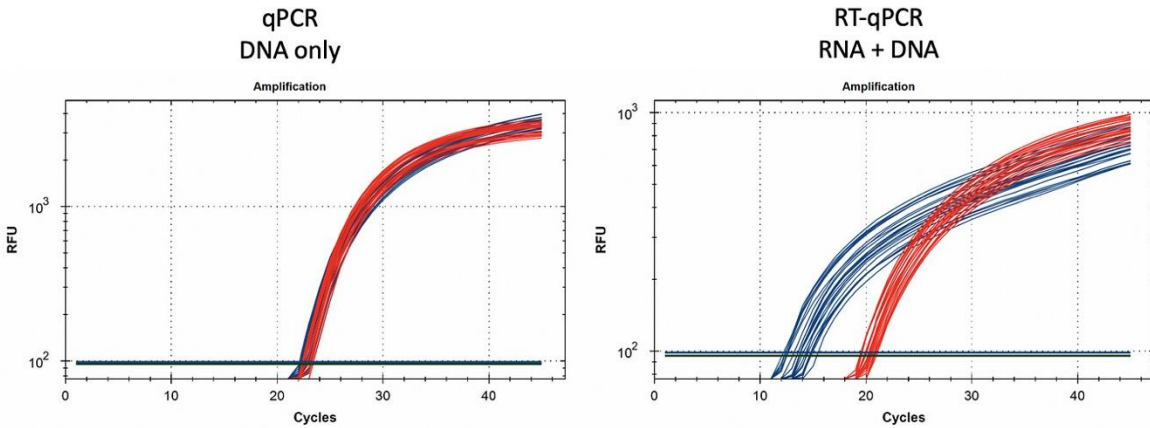


Figure 15. 1µl of vaccine boiled in 100µl of Leaf Lysis buffer was subjected to qPCR (left) and RT-qPCR (right) for Vector (red) and Spike (blue). 8 samples were tested in triplicate.

qPCR-Spike	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	23.12	22.98	22.58	22.33	22.36	22.08	22.20	22.06	0.401
Replicate 2	23.16	22.90	22.70	22.36	22.20	22.16	22.29	22.22	0.373
Replicate 3	23.22	22.84	22.59	22.29	22.44	22.26	22.29	22.11	0.366
STDEV	0.05	0.07	0.07	0.03	0.12	0.09	0.05	0.08	

qPCR: (Vector-Spike)	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	0.20	0.08	0.27	(0.00)	0.18	0.18	0.10	0.24	0.090
Replicate 2	0.16	0.22	0.29	0.11	0.18	0.12	0.03	0.13	0.079
Replicate 3	0.14	0.31	0.20	0.17	0.31	0.19	0.20	0.13	0.069
STDEV	0.03	0.11	0.05	0.09	0.08	0.04	0.08	0.06	

qPCR-Vector	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	23.33	23.06	22.85	22.32	22.54	22.26	22.30	22.30	0.411
Replicate 2	23.32	23.12	23.00	22.47	22.38	22.28	22.32	22.35	0.419
Replicate 3	23.36	23.15	22.79	22.46	22.75	22.46	22.49	22.23	0.383
STDEV	0.02	0.04	0.11	0.08	0.19	0.11	0.11	0.06	

RATIO RNA/DNA	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	1	1	1	1	1	1	1	1	0.068
Replicate 2	1	1	1	1	1	1	1	1	0.062
Replicate 3	1	1	1	1	1	1	1	1	0.056
STDEV	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.1	

Table 2. CT values for Spike and Vector during qPCR (DNA only). Standard deviation for the triplicate measurements run horizontally in black font. Standard deviation for vial to vial run vertically in Red. Delta CT or (Vector CT minus Spike CT) represents the ratio of Spike to Vector DNA and should = 1.

RT-Spike	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	14.05	14.77	13.18	13.77	13.79	12.52	12.62	13.53	0.749
Replicate 2	14.29	14.74	14.38	14.82	13.78	13.82	12.57	12.38	0.925
Replicate 3	14.49	14.91	15.43	13.84	13.74	13.55	12.36	12.19	1.141
STDEV	0.22	0.09	1.12	0.59	0.02	0.69	0.14	0.72	

RT: (Vector-Spike)	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	6.74	5.93	7.20	6.40	6.51	7.31	7.33	5.97	0.570
Replicate 2	6.33	6.06	5.92	5.67	6.34	6.13	6.92	7.06	0.478
Replicate 3	6.33	6.07	5.43	6.39	6.13	6.38	7.09	7.18	0.562
STDEV	0.24	0.07	0.91	0.42	0.19	0.62	0.21	0.67	

RT-Vector	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	20.80	20.71	20.39	20.16	20.30	19.83	19.95	19.50	0.439
Replicate 2	20.62	20.80	20.30	20.49	20.12	19.96	19.49	19.45	0.499
Replicate 3	20.81	20.98	20.86	20.23	19.88	19.93	19.45	19.37	0.638
STDEV	0.11	0.14	0.30	0.17	0.21	0.07	0.28	0.07	

RATIO RNA/DNA	Vial 1	Vial 2	Vial 3	Vial 4	Vial 5	Vial 6	Vial 7	Vial 8	STDEV
Replicate 1	107	61	147	84	91	159	161	63	41.54
Replicate 2	80	67	61	51	81	70	121	134	29.25
Replicate 3	80	67	43	84	70	83	136	145	34.79
STDEV	15.5	3.3	55.8	19.2	10.4	47.9	20.3	44.6	

Table 3. CT values for Spike and Vector during RT-qPCR (RNA+DNA). Ratio of RNA:DNA ranges from 43:1 To 161:1. EMA allowable limit is 3030:1. This is 18-70 fold over the EMA limit.

Discussion

Multiple methods highlight high levels of DNA contamination in the both the monovalent and bivalent vaccines. While the Qubit™ 3 and Agilent Tape Station™ differ on their absolute quantification, both methods demonstrate it is orders of magnitude higher than the EMAs limit of 330ng DNA/ 1mg RNA. qPCR and RT-qPCR confirms the relative RNA to DNA ratio. An 11-12 CT offset should be seen between Spike and Vector RT-qPCR signals to represent a 1:3030

contamination limit ($2^{11.6} = 3100$). Instead, we observe much smaller CT offsets (5-7 CTs) when looking at qPCR and RT-qPCR data with these vaccines. It should be noted that Qubit™ 3 and Agilent methods stain all DNA in solution while qPCR measures only amplifiable molecules without DNase I cut sites between the primers. The further apart you space the qPCR primers, the fewer Qubit™ 3 and Agilent detectable molecules will amplify. The primers used in this study are 106bp and 114bp apart, thus any molecules that are DNase I cut below this length will be undercounted with the qPCR methods relative to more general dsDNA measurements from Qubit™ 3 or Agilent Tape Station™.

This also implies that qPCR standard curves using 100% intact synthetic DNA standards will amplify more efficiently and thus undercount the total digested DNA contamination. For example, standard curves with 106-114bp synthetic templates provide CTs under 20 in the picogram range (not low nanogram range) suggesting large portions of the library are smaller than the minimum amplifiable size. Pure standards also do not contain high concentrations of modified mRNA with identical sequence which could serve as a competitive primer sink or inhibitor to qPCR methods.

Alternatively, the Qubit™ 3 and the Agilent Tape Station™ could be inflating the DNA quantification due to intercalating dye cross talk with N1-methylpseudouridine RNA. For this reason, we believe the ratio we observed when these molecules are more scrupulously interrogated with polymerases specific for each template type in qPCR and RT-qPCR is a more relevant metric. The EMA metric is also stated as such a ratio.

This also brings into focus if these EMA limits took into consideration the nature of the DNA contaminants. Replication competent DNA should arguably have a more stringent limit. DNA with mammalian promoters or antibiotic resistance genes may also be of more concern than just random background *E.coli* genomic DNA from a plasmid preparation (Sheng-Fowler et al. 2009). Background *E.coli* DNA was measured with qPCR and had CT over 35.

There has been a healthy debate about the capacity for SARs-CoV-2 to integrate into the human genome (Zhang et al. 2021). This work has inspired questions regarding the capacity for the mRNA vaccines to also genome integrate. Such an event would require LINE-1 driven reverse transcription of the mRNA into DNA as described by Alden *et al.* (Alden et al. 2022). dsDNA contamination of sequence encoding the spike protein wouldn't require LINE-1 for Reverse Transcription and the presence of an SV40 nuclear localization signal in Pfizer's vaccine vector would further increase the odds of integration. This work does not present evidence of genome integration but does underscore that LINE-1 activity is not required given the dsDNA levels in these vaccines. The nuclear localization of these vectors should also be verified.

Prior sequencing of the monovalent vaccines from Jeong *et al.* only published the consensus sequence (Dae-Eun Jeong 2021). The raw reads for this project are not available and should be scrutinized for the presence of vector sequence.

Given these vaccines exceed the EMA limits (330ng/mg DNA/RNA) with the Qubit™ 3 and Agilent data and these data also exceed the FDA limit (10ng/dose) with the more conservative qPCR standard curves, we should revisit the lipopolysaccharide (LPS) levels. Plasmid contamination from *E.coli* preps are often co-contaminated with LPS. Endotoxins contamination can lead to anaphylaxis upon injection (Zheng et al. 2021).

A limitation of this study is the unknown provenance of the vaccine vials under study. These vials were sent to us anonymously in the mail without cold packs. RNA is known to degrade faster than DNA and it is possible poor storage could result in faster degradation of RNA than DNA. RNA as a molecule is very stable but in the presence of metals and heat or background ubiquitous RNases, it can degrade very quickly. All of the vaccines in this study are past the expiration date listed on the vial suggesting more work is required to understand the DNA to RNA ratios in fresh lots. The publication of these qPCR primers may assist in surveying additional lots with more controlled supply chains. Studies evaluating vaccine longevity in breast milk or plasma may benefit from vector DNA surveillance as this sequence is unique to the vaccine and may persist longer than mRNA.

While the sequencing delivered full coverage of the plasmid backbones, it is customary to assemble plasmids from DNase I fragmented libraries. These methods have not discerned the ratio of linear versus circular DNA in the vials. While plasmid DNA is more competent and stable, linear DNA may have higher genome integration risks.

The intercalating dyes used in the Qubit™ 3 and Agilent systems are known to have low fluorescent cross talk with DNA and RNA but it is unknown to what degree N1-methylpseudouridine alters the specificity of these intercalating dyes. As a result, we have relied on the CT offsets between RT-qPCR and qPCR with the vector and spike sequence as the best relative assessment of the EMA ratio-metric regulation. These qPCR and RT-qPCR reagents may be useful in tracking these contaminants in vaccines, blood banks or patient tissues in the future.

Methods

Purifying the mRNA from the LNPs

LiDs/SPRI purification

100µl of each vial was sampled (1/3rd to 1/5th of a dose)

- 5µl of 2% LiDs was added to 100µl of Vaccine to dissolve LNPs
- 100µl of 100% Isopropanol
- 233µl of Ampure (Beckman Genomics)
- 25µl of 25mM MgCl₂ (New England Biolabs)

Samples were tip mixed 10X and incubated for 5 minutes for magnetic bead binding. Magnetic Beads were separated on a 96-well magnet plate for 10 minutes and washed twice with 200µl of 80% EtOH. The beads were left to air dry for 3 minutes and eluted in 100µl of ddH₂O. 2µl of eluted sample was run on an Agilent Tape Station™.

CTAB/Chloroform/SPRI purification of Vaccines

Some variability in qPCR performance was noted with our LiDs/SPRI purification method of the vaccines. This left some samples opaque and may represent residual LNPs in the purification. A CTAB/Chloroform/SPRI isolation was optimized to address this and used for further qPCR and Agilent electrophoresis. Briefly, 300µl of Vaccine was added to 500µl of CTAB (MGC solution A in SenSATIVax MIP purification kit. #420004). The sample was then vortexed and heated for 5 minutes at 37°C. 800µl of chloroform was added, vortexed and spun at 19,000 rpms for 3 minutes. The top 250µl of aqueous phase was collected and added to 250µl of solution B and 1ml of magnetic binding buffer. Samples were vortexed and incubated for 5 minutes and magnetically separated. The supernatant was removed and the beads washed with 70% Ethanol two times. Samples were finally eluted in 300µl of MGC elution buffer.

Simple boil preparation for evaluating vaccine qPCR.

This boil prep process simply takes 1-10µl of the vaccine and dilutes it into a PCR compatible leaf lysis buffer and heats it (Medicinal Genomics part number 420208).

- 65°C for 6 minutes
- 95°C for 2 minutes

Library Construction for Sequencing

50µl of each 100µl sample was converted into RNA-Seq libraries for Illumina sequencing using the NEB NEBNext Ultrall Directional RNA library Kit for Illumina (NEB#E7760S).

To enrich for longer insert libraries the fragmentation time was reduced from 15 minutes to 10 minutes and the First strand synthesis time was extended at 42°C to 50 minutes per the long insert recommendations in the protocol.

No Ribo depletion or PolyA enrichment was performed as to provide the most unbiased assessment of all fragments in the library. The library was amplified for 16 cycles according to the manufacturers protocol. A directional library construction method was used to evaluate the single stranded nature of the mRNA. This is an important quality metric in the EMA and TGA disclosure documents as dsRNA (>0.5%) can induce an innate immune response. dsRNA content is often estimated using an ELISA. Directional DNA sequencing offers a more comprehensive method for its estimation and was previously measured and 99.99% in Jeong et al. It is unclear how this may vary lot to lot or within the new manufacturing process for the newer bivalent vaccines.

RNase A treatment of the Vaccines

RNase A cleaves both uracils and cytosines. N1-methylpseudouridine is known to be RNase-L resistant but RNase A will cleave cytosines which still exist in the mRNAs. This leaves predominantly DNA for sequencing. Vaccine mRNA that was previously sequenced and discussed here, was treated at 37°C for 30 minutes with 10µl of 20 Units/µl Monarch RNase A from NEB. The RNase reaction was purified using 1.5X of SenSATIVax (Medicinal Genomics #420001). Sample were eluted in 20µl ddH2O after DNA purification. 15µl was used for DNA sequencing.

DNase treatment of the vaccines

50µl of CTAB purified vaccine was treated at 37°C for 30 minutes with 2µl DNase I and 6µl of DNase I buffer (Grim reefer MGC#420143). 2.5µl of LiDs Lysis buffer was added to stop the DNase reaction. Reactions were purified using 60µl 100% Isopropanol, 140µl Ampure, 15µl MgCl₂. Magnetic beads were tip mixed 10 times, left for 5 minutes to incubate, magnetically separated and then washed twice with 80% EtOH.

Whole genome shotgun of RNase'd Vaccines.

15µl of the DNA was converted into sequence ready libraries using Watchmakers Genomics WGS library construction kit. This kit further fragments the DNA to smaller sizes making fragment length in the vaccines difficult to predict.

Qubit™ 3 Fluorometry

Qubit™ 3 fluorometry was performed using Biotum AccuBlue RNA Broad Range kit (#31073) and Biotum AccuGreen High Sensitivity dsDNA Quantitation Kit (#31066) according to the manufacturers instructions.

E.coli qPCR

Medicinal Genomics PathoSEEK™ E.coli Detection assay (#420102) was utilized according to the manufacturers instructions.

qPCR and RT-qPCR Spike Assay

- MedGen-Moderna_Pfizer_Janssen_Vax-Spike_Forward
- >AGATGGCCTACCGTTCA
- MedGen-Moderna_Pfizer_Janssen_Vax-Spike_Reverse
- >TCAGGCTGTCCTGGATCTT
- MedGen-Moderna_Pfizer_Janssen_Vax-Spike_Probe
- >/56-FAM/CGAGAACCA/ZEN/GAAGCTGATCGCCAA/3IABkFQ/

qPCR and RT-qPCR Vector Origin Assay

- MedGen_Vax-vector_Ori_Forward
- >CTACATACCTCGCTCTGCTAATC
- MedGen_Vax-vector_Ori_Reverse
- GCGCCTTATCCGGTAACTATC
- MedGen_Vax-vector_Ori_Probe
- /5HEX/AAGACACGA/ZEN/CTTATCGCCACTGGC/3IABkFQ/

Elute primer to 100uM according to IDT instructions.

Make 50X primer-probe mix.

1. 25µl 100uM Forward Primer
2. 25µl 100uM Reverse Primer
3. 12.5µl 100uM Probe
4. 37.5µl nuclease free ddH₂O.

Use 15µl of this mixture in the **qPCR master mix** setup seen below. (0.5µl primer/probe per reaction)

Use 10µl of this mixture in the **RT-qPCR master mix** setup seen below.

Medicinal Genomics Master Mix kits used

1. <https://store.medicinalgenomics.com/qPCR-Master-Kit-v3-200-rxns>
2. <https://store.medicinalgenomics.com/pathoseek-rt-qpcr-master-kit>

Reaction setup for 30 reactions of qPCR

- 114µl Enzyme Mix (green tube)
- 24µl Reaction Buffer (blue tube)
- 246µl nuclease free ddH₂O
- 15µl of Primer-Probe set Spike
- 15µl of Primer-Probe set Ori

Use 13.8µl of above MasterMix and 5µl of purified sample (1µl Vax DNA/RNA + 4µl ddH₂O if CT <15)

Reaction setup for 34 reactions of RT-qPCR

- 200µl Enzyme mix
- 96µl nuclease free ddH₂O
- 20µl RNase Inhibitor (purple tube)
- 4µl DTT (green tube)
- 10µl Primer-Probe set Spike
- 10µl Primer-Probe set Ori

10µl of MasterMix and 1µl of Vax DNA/RNA

Medicinal Genomics MIP DNA Purification Kit used

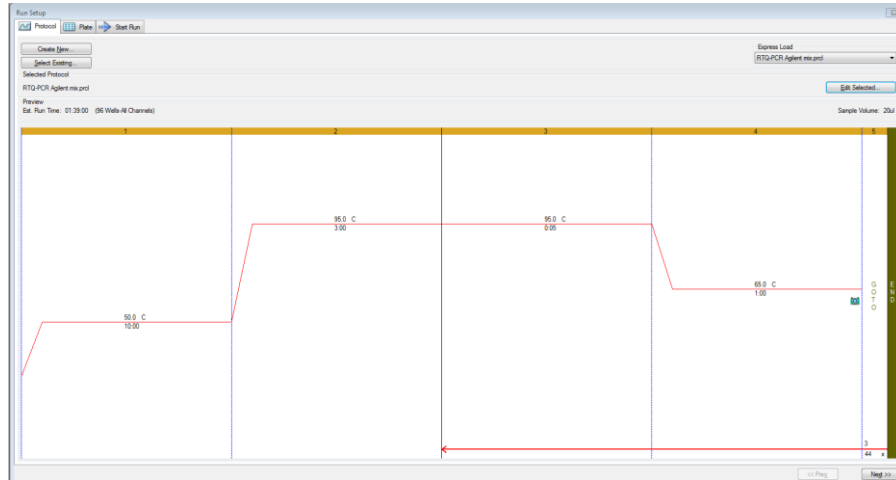
1. https://store.medicinalgenomics.com/SenSATIVAx-DNA-Extraction-Kit-200-reactions_2

he CTAB/Chloroform/SPRI based DNA/RNA isolation methods are described above.

Cycling conditions

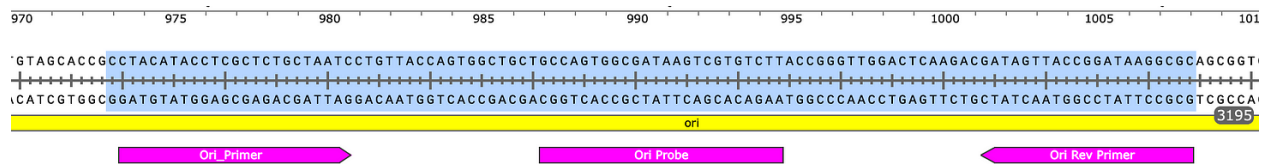
These conditions work for both qPCR and RT-qPCR. Note: The 50°C RT step can be skipped with qPCR. The MGC qPCR MasterMix kits used have a hot start enzyme which are unaffected by this 50°C step. For the sake of controlling RNA to DNA comparisons, we have put qPCR and RT-qPCR assays on the same plate and run the below program with the RT step included for all samples.

Cycling Conditions used for qPCR and RT-qPCR



Sequences of amplicons for gBlock Positive Controls. Ori = 106bp, Spike = 114bp.

Ori target



Spike target



Sequencing Data

Raw Illumina Reads RNA-seq

- [Pfizer Bivalent Vial 1 Forward reads](#)
- [Pfizer Bivalent Vial 1 Reverse reads](#)
- [Pfizer Bivalent Vial 2 Forward reads](#)
- [Pfizer Bivalent Vial 2 Reverse reads](#)
- [Moderna Vial 1 Forward reads](#)
- [Moderna Vial 1 Reverse reads](#)
- [Moderna Vial 2 Forward reads](#)
- [Moderna Vial 2 Reverse reads](#)

Read files are run through sha256 (Hash and stash) and etched onto the DASH blockchain. The sha256 hash of the read file is spent into the OP_RETURN of an immutable ledger. If the hash of the file doesn't match the hash in these transactions, the file has been tampered with.

- [Pfizer Vial 1 Forward hash](#)
- [Pfizer Vial 1 Reverse hash](#)
- [Pfizer Vial 2 Forward hash](#)
- [Pfizer Vial 2 Reverse hash](#)
- [Moderna Vial 1 Forward hash](#)
- [Moderna Vial 1 Reverse hash](#)
- [Moderna Vial 2 Forward hash](#)
- [Moderna Vial 2 Reverse hash](#)

Megahit Assemblies

- [Pfizer Vial 1](#)
- [Pfizer Vial 2](#)
- [Moderna Vial 1](#)
- [Moderna Vial 2](#)

Illumina Reads mapped back to Megahit Assemblies

- [Pfizer Vial 1 BAM File](#). [Index File](#)
 - [Pfizer Vial 2 BAM File](#). [Index File](#)
 - [Moderna Vial 1 BAM File](#). [Index File](#)
 - [Moderna Vial 2 BAM File](#). [Index File](#)
-

Q30 Filtered Illumina Reads (use these for transcriptional error rate estimates)

[FastQ-Filter download](#): usage> fastq-filter -e 0.001 -o output.fastq input.fastq

- [Pfizer bivalent Vial 1 Forward Reads](#)
- [Pfizer bivalent Vial 1 Reverse Reads](#)
- [Pfizer bivalent Vial 2 Forward Reads](#)
- [Pfizer bivalent Vial 2 Reverse Reads](#)
- [Moderna bivalent Vial 1 Forward Reads](#)
- [Moderna bivalent Vial 1 Reverse Reads](#)
- [Moderna bivalent Vial 2 Forward Reads](#)
- [Moderna bivalent Vial 2 Reverse Reads](#)

Q30 BAM files. Q30 Reads mapped against Megahit assemblies

- [Pfizer Vial 1 q30-BAM file](#). [Index File](#)
 - [Pfizer Vial 2 q30-BAM file](#). [Index File](#)
 - [Moderna Vial 1 q30-BAM file](#). [Index File](#)
 - [Moderna Vial 2 q30-BAM file](#). [Index File](#)
-

IGVtools error by base on q30 reads

Fields = Position in contig, Positive stand (+)A, +C, +G, +T, +N, +Deletion, +Insertion, Negative strand -A, -C, -G, -T, -N, -Deletion, -Insertion

- [Moderna Vial 1](#)

- [Moderna Vial 2](#)
- [Pfizer Vial 1](#)
- [Pfizer Vial 2](#)

Analysis pipeline

Reads were demultiplexed and processed with

- [Trimgalore](#) - Removes Illumina Sequencing adaptors.
- [Megahit](#)- assembles reads into contigs.
- [Megahit for SARs-CoV-2](#)
- [Samtools](#)- generates BAM files for viewing in IGV.
- Samtools stats used to calculate outie reads.
- [BWA-mem](#)- Short read mapper used to align reads back to the assembled references.
- SnapGene software- (www.snapgene.com)- Used to visualize and annotate expression vectors
- [IGV](#)- Integrated Genome Viewer used to visualize Illumina sequencing reads.

RNase Treated Libraries-BAM files

contig specific BAM files were created using samtools

```
samtools view -h input.bam contig_name -O BAM > contig.bam; samtools index contig.bam;
```

Samtools stats run on a each contig in each assembly.

```
for out_prefix in `ls *.sort.bam | perl -pe "s/.sort.bam//"; do mkdir -p ${out_prefix}-samtools-
stats; for contig in `samtools view -H ${out_prefix}.sort.bam | grep "^@SQ" | cut -f 2 | perl -pe
"s/SN\://"; do echo "Now calculating stats for ${contig}/${out_prefix}..."; samtools stats
${out_prefix}.sort.bam $contig > ${out_prefix}-samtools-stats/${contig}-samtools-stats.txt;
done; done
```

- [Pbiv1 RNase WM k141 107.fa](#)
- [Pbiv1 RNase WM k141 107.bam](#)
- [Pbiv1 RNase WM k141 107.bam.bai](#)
- [Pbiv2 RNase WM k141 23.fa](#)
- [Pbiv2 RNase WM k141 23.bam](#)

- [Pbiv2 RNase WM k141 23.bam.bai](#)

Author contributions

KJM- constructed the sequencing libraries, designed the qPCR assays, ran Qubit™ 3s and Agilent Tape Station™ and performed the analysis, drafted the manuscript.

YH-Optimized DNA isolations, Tape Station™ and qPCR results.

SM, LTK- assisted in demultiplexing and trimming the reads and assembly troubleshooting

Conflicts of interest- Authors of this paper are employees of Medicinal Genomics which manufacturers some of the qPCR and DNA isolation kits used in this study.

Acknowledgements.

We'd like to thank Lei Zhang for organizing the sequencing outsourcing. We'd like to thank Jessica Rose, Sabine Hazan, Jikkyleaks, Pathogenetics, Steve Massey, Valentine Bruttel, Lynn Fynn, Sasha Letypova, David Wiseman and Pharmacoconuts for helpful comments on this topic.

References

- Alden M, Olofsson Falla F, Yang D, Barghouth M, Luan C, Rasmussen M, De Marinis Y. 2022. Intracellular Reverse Transcription of Pfizer BioNTech COVID-19 mRNA Vaccine BNT162b2 In Vitro in Human Liver Cell Line. *Curr Issues Mol Biol* **44**: 1115-1126.
- Bansal S, Perincheri S, Fleming T, Poulson C, Tiffany B, Bremner RM, Mohanakumar T. 2021. Cutting Edge: Circulating Exosomes with COVID Spike Protein Are Induced by BNT162b2 (Pfizer-BioNTech) Vaccination prior to Development of Antibodies: A Novel Mechanism for Immune Activation by mRNA Vaccines. *J Immunol* **207**: 2405-2410.
- Castruita JAS, Schneider UV, Mollerup S, Leineweber TD, Weis N, Bukh J, Pedersen MS, Westh H. 2023. SARS-CoV-2 spike mRNA vaccine sequences circulate in blood up to 28 days after COVID-19 vaccination. *APMIS* **131**: 128-132.
- Dae-Eun Jeong MM, Karen Artiles, Orkan Ilbay, Andrew Fire*, Kari Nadeau, Helen Park, Brooke Betts, Scott Boyd, Ramona Hoh, and Massa Shoura*. 2021. Assemblies of putative SARS-CoV2-spike-encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273. *GitHub*.
- Dean DA, Dean BS, Muller S, Smith LC. 1999. Sequence requirements for plasmid nuclear import. *Exp Cell Res* **253**: 713-722.
- Hanna N, Heffes-Doon A, Lin X, Manzano De Mejia C, Botros B, Gurzenda E, Nayak A. 2022. Detection of Messenger RNA COVID-19 Vaccines in Human Breast Milk. *JAMA Pediatr* **176**: 1268-1270.
- Josephson F. 2020-11-19. Rapporteur's Rolling Review assessment report. *Committee for Medicinal Products for Human Use EMEA/H/C/005735/RR*.
- Moreau P, Hen R, Wasylyk B, Everett R, Gaub MP, Chambon P. 1981. The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic acids research* **9**: 6047-6068.

- Nance KD, Meier JL. 2021. Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. *ACS Cent Sci* **7**: 748-756.
- Sheng-Fowler L, Lewis AM, Jr., Peden K. 2009. Issues associated with residual cell-substrate DNA in viral vaccines. *Biologicals* **37**: 190-195.
- Ulrich-Lewis JT, Draves KE, Roe K, O'Connor MA, Clark EA, Fuller DH. 2022. STING Is Required in Conventional Dendritic Cells for DNA Vaccine Induction of Type I T Helper Cell-Dependent Antibody Responses. *Front Immunol* **13**: 861710.
- Zhang L, Richards A, Barrasa MI, Hughes SH, Young RA, Jaenisch R. 2021. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proceedings of the National Academy of Sciences of the United States of America* **118**.
- Zheng Q, Wang T, Zhu X, Tian X, Zhong C, Chang G, Ran G, Xie Y, Zhao B, Zhu L et al. 2021. Low endotoxin E. coli strain-derived plasmids reduce rAAV vector-mediated immune responses both in vitro and in vivo. *Mol Ther Methods Clin Dev* **22**: 293-303.