Reviewers' comments:

Reviewer #1 (Remarks to the Author):

A. Summary of the key results
The authors describe a new protein candidate (tp0751) for syphilis vaccination. Tp0751 was reported to play an important role in the crossing of vascular endothelium and thus facilitates dissemination of T. pallidum via the bloodstream. A rabbit immunization model was used to investigate efficacy of tp0751 vaccination to block T. pallidum dissemination in vivo.

B. Originality and interest: if not novel, please give reference
The study is a logical follow up of the previous work of the author(s) and complements nicely the work that was done on the characterization of the lipoprotein tp0751 and its function. Three of the authors hold a relatively new patent (WO 2015004604 A1) on immunogenic tp0751 fragments in which some aspects of the study have been described. The search for suitable targets for syphilis vaccination is of great importance to counteract treponematosis. The study is therefore relevant and provides a significant contribution for the development of a vaccine against syphilis.

C. Data & methodology: validity of approach, quality of data, quality of presentation
Overall the MS reads well and the data are presented nicely. However, parts of the Material and Methods need additional information e.g. it is not even introduced which T. pallidum strain has been used to challenge the rabbits. For details see P-2-P comments following point H).

D. Appropriate use of statistics and treatment of uncertainties
Animal numbers and statistics are the major weakness of the study. The authors explanation provided in the Reporting Checklist for Life Sciences Articles "The number of animals included in the protection experiments was reflective of a number sufficient to investigate the protective effect of the vaccine candidate while adhering to budgetary constraints." is not satisfying. (A) with only 2(ctrl)+3 animals results will not be normal distributed. However, throughout the MS parametric tests are applied. This should be seen critical and reanalysis should be made using non-parametric tests. Currently, the risk is high that significance levels provide a false assumption because of a lack of power. (B) I recommend to compute a post hoc power analysis of data to identify the weakness.

E. Conclusion: robustness, validity, reliability
The results are promising and provide a pretty good idea about the usefulness of tp0751 as a component in vaccination against syphilis. However, the presented data are currently based on low animal numbers and must be interpreted carefully.

F. Suggested improvements: experiments, data for possible revision
Ideally, animal numbers need to be upgraded. I would say group size must be at least 4-5/group. It looks like some more animals were used for vaccination against tp0751 (e.g., animal Im26 pops-up all over sudden in Extended Data Fig. 2). I wonder if some analysis can be supported by additional data sets to increase animal numbers. However, in case upgrading animal number is not possible, the statistical analysis needs a thorough revision. At least some discussion should be included that the presented data are limited in that the power is not (always?) sufficient. Having addressed this issue, it would be good to make the raw data available in the SOM.

G. References: appropriate credit to previous work?
Looks OK to me.

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusion
Here and there the text could be more clear. I have addressed some of the critical issues below.

L. 31: Normally tp0751 should be written in italics (pl. check w/ Journal stile)
L. 46 and 47: Suggest to add (%)
L. 61: Is 'desensitization' a real issue in the clinic? It is often recommended by CDC and others, but how often does it really occur. I think it is not necessary to make this point here. There are enough plausible reasons why a vaccine is needed.
L. 66 and 71: The authors should be more accurate here. Talking about T. pallidum would include ssp. pertenue and endemicum. Not all subspecies are reported to cause congenital infection etc.
L. 103 and 349ff. and 357 ff.: Citation of submitted work. Is this conform to the Journal guidelines? It seems weird that some authors are listed with full names, others are just provided as initials. I guess "Parker et al., unpublished data" or similar should do the job.
Fig 1.: Should be improved by the addition of details about the immunization process. Also, here and throughout the MS, the subsp. and strain that was used to challenge animals is not addressed.
Fig. 2: Although raw data are not available, the power of these analysis seems to be quite low. It is not lege artis to present non-parametric data as mean values since this assumes Gaussian distribution.
L. 139: Which subsp./strain?
Extended Data Table 1: Time points do not become clear; values are mean(?) but no SD/SEM is presented? It would be helpful to get a more detailed dataset.
L. 153: Was analysis also supported by qPCR?
Fig. 3: I suggest to change the ordinate in B. to D. into scientific scaling similar to graph A.
I may have missed it, but Im1 has a huge variation. It would be good to explain this in the Discussion as well as the reason why only Im2 was tested positive in bone tissue.
L. 184 ff.: Why was analysis on the RNA level not included? This should provide reliable data on viability.
L. 242: TP DNA was detected at least in spleen tissue of Im3 (Fig. 3). The statement seems to be incorrect.
Fig. 4: Abscissa > Average cell count ..per what?
L 250 ff.: "Field of view" is not defined (which magnification; staining?)
L. 285: Again, RNA based methods would have been useful here.
L. 95 ff.: There is no information on the T. pallidum strain that was used to challenge the animals.
L. 327: homing
L. 329: Again, when outbreed animals and low animal numbers come together it is challenging to achieve adequate statistical power. This should be reflected in the study design.
L. 361: The discussion is missing any aspects of multicomponent vaccination for TP (e.g. tprK + tp0571).
L. 369: "protects against Treponema dissemination" this is not shown by the data. Better "significantly reduces dissemination…"
L. 398: FTA-ABS serology IgG or IgM or both? Which test was used (manufacturer?)
L. 402: Do I understand it right that one immunization was 4x0.1 ml SC plus 2x 0,04 ml IM? What was the reason for so many injections?
L. 408: Again, which subsp. and strain?
L. 413: Fig. 1 does not illustrate what is written here. Why were only the ctrl animals bled and where do the 10 days come from?
L. 423: RIT > first time use in the text (apart from Fig. legend)
L. 484: Why was the 84°C step included?
L. 494 ff: Important details are missing about stainings that were used. Also, why was IHC not conducted to depict TP?
L. 499: Field of view must be defined. TP often causes perivascular-inflammation and thus inflammatory cells are not normal distributed in the tissue. How were the fields of view standardized? Which area of the skin was analyzed?
L. 507 ff: see the initial statement on statistics.
L. 609: T(reponema) must be all italic
L. 687: It does not become clear why titers are displayed relative to serum from an animal that is not part of the study (Im26).

Extended Data Tabel 2/L. 712: If FTA-ABS is never determined you may want to exclude it from the table. The information is useless.
Extended Data Table 3: Why are d147-185 missing for Im3?

Reviewer #2 (Remarks to the Author):

In this manuscript the authors show that rabbits immunised with the antigen, Tp0751, resulted in attenuated lesion development (moderately convincing), inhibition of T.pallidum dissemination (convincing) and increased cellular infiltration at lesion sites (convincing at the tissue level). There is still an urgent need for a better vaccine for syphilis and this manuscript does make a significant contribution in this area.

However, while some of the findings were convincing at the tissue level, others were less so, as evidenced by comments such as " …difficult to reach a definitive conclusion concerning the true lesion burden …" or "It is likely that both of these proposed mechanisms contributed to the proposed mechanisms …".
One aspect that would significantly strengthen the manuscript would be if the authors could link their whole animal and tissue level observations with some underlying immunological mechanisms. While they did observe the broad types of cells at the primary lesion sites, it would have been more useful if they had analysed specific Tp0751 antigen responses in defined T cell types.

The authors also suggest that their positive observation may have been " …suggestive of local production of Tp0751-specific antibodies". Why didn't they measure these antibody levels and perhaps even test for their functional role; neutralisation or binding. By including these additional levels of analyses the study results would be more likely to have interest and impact in the field.

Reviewer #3 (Remarks to the Author):

Summary of the key results
The authors have undertaken an experimental model to investigate immunological, clinical, and microbiological correlates of protection against syphilis conferred by a newly developed vaccine. They have found that immunized rabbits have a reduced bacterial load measured by PCR in the organs as compared to unimmunized rabbits, but the vaccine apparently did not provide sterile protection which would be ideal. In the second part of the study, however, the rabbit infectivity testing negative results point towards the desired sterile protection.

Originality and interest:
The claims of the paper are novel and the vaccine may have an important role in mitigation of syphilis complications, but would not work to prevent the disease and ultimately stop transmission. Although the data seems not to be conclusive, the results are most interesting and timely for the development of a fully protective syphilis vaccine.

Data & methodology:
- Would the authors consider having used a group of only three animals as a limitation given the discrepant results between animals?
- Despite the authors' argument that PCR may have detected dead organisms may be true, the discrepant result for PCR vs microscopic assessment of primary lesions (Rabbit 1) needs to be interpreted with caution.
- Could the lower number of Treponema in the inoculation site of Rabbit 3 be related to a

decreased proliferation? The current explanation (i.e. greater number of cells migration) seems contradictory to the greater immune cell infiltration in the skin of Rabbit 3.
- qPCR results in Extended Data Table 1 do not seem to correspond to Figure 3. There might be a typo for the Liver and Spleen Ct1 and Ct2 results: where it says "10 squared" should read "103". Also Liver Im2 seems to have the same mistake. Please confirm since these are the most critical findings of the study.
- I would recommend that Extended Table 1 is fitted into the manuscript to facilitate reading comprehension.
- Could the popliteal lymph node for Rabbit 3 have been tested using PCR to confirm the hypothesis of greater load of dead treponemes draining to the lymph node?

Appropriate use of statistics:
- Authors may want to revise the statistical analysis of lesions progression (i.e. ulceration and diameter). Currently individual lesions on each rabbit are analyzed as independent items. However, the lesion sites in one rabbit may have correlated results in terms of lesion progression compared to other rabbits. The reason is that lesions in a single rabbit are exposed to the same protective immunity; therefore these are not fully independent units and may need specific statistical methods for comparison (e.g. multilevel model).

Conclusions:
The overall results are very promising but with some contradictions that seem difficult to overcome with a small sample (n=3) study like this, the authors may want to discuss what are the next steps.

It would be good that the researchers explain the path towards development of a syphilis vaccine that can be trialed in humans and possibly to acknowledge certain gaps, including that the current vaccine would not be effective to prevent chancre development and that the current study does not ensure the vaccine is broadly protective against different T pallidum strains.

Clarity and context:
- The researchers should more clearly explain why it is important that Tp0751 belongs to the same protein family as the meningococcus b vaccine. If one reads the abstract, as it is currently explained, some people could think that the meningococcus vaccine is protective for syphilis.
- Discrepant results of experiments could be more clearly written
- Readers outside the discipline would benefit of a schematic of the main result to accompany publication.

Reviewer #1

1. "…parts of the Material and Methods need additional information e.g. it is not even introduced which *T. pallidum* strain has been used to challenge the rabbits." *We thank the Reviewer for bringing this omission to our attention. We have now fixed this error, please see Lines 414 and 424-425.*
2. Animal numbers and statistics are the major weakness of the study. The authors explanation provided in the Reporting Checklist for Life Sciences Articles "The number of animals included in the protection experiments was reflective of a number sufficient to investigate the protective effect of the vaccine candidate while adhering to budgetary constraints." is not satisfying. (A) with only 2(ctrl)+3 animals results will not be normal distributed. However, throughout the MS parametric tests are applied. This should be seen critical and reanalysis should be made using non-parametric tests. Currently, the risk is high that significance levels provide a false assumption because of a lack of power. (B) I recommend to compute a post hoc power analysis of data to identify the weakness. *We agree with the Reviewer and recognize the limitations of our animal numbers. (A) Statistical analyses have been re-done using non-parametric tests for situations where n>2 (non-parametric tests cannot be performed with n=2). For Fig. 2a-b, we performed non-linear regression analysis with extra sum of squares F-test, which is less dependent upon normal distribution than two-way ANOVA. We were unable to apply the appropriate non-parametric tests to data displayed in Figs. 3 and 4 due to low sample size (Mann-Whitney requires a combined n=8). To evaluate the overall difference in T. pallidum burden in disseminated lesion sites (Fig. 3b-d) between control and immunized rabbits, flaA copy number/µg rabbit DNA was normalized for each tissue by expressing as a percentage relative to the mean (median) from the control animals, which increased our sample sizes (n=9 immunized, n=6 control) permitting us to calculate normality and use a non-parametric test. An extra figure panel has been added (Fig. 3e) to show this normalized data and this has been explained in the Data Analysis section of the materials and methods section (Lines 539-554). This same approach was used for analysis of cellular infiltrates in primary lesions (Figure 4), and an additional panel was added to display this normalized data (Figure 4e). Modifications have been made throughout the results and discussion sections to reflect these new analyses. (B) It is our understanding the power post-hoc analysis is useful to determine the required n value for a future experiment based upon a pilot study, but not to evaluate the statistical significance of a collected dataset. Furthermore, p-values and power are directly related to one another. Since we have re-analyzed our data using non-parametric comparisons, we do not believe post-hoc power analysis is necessary for this study. However, we will be using such calculations to design our future experiments, and this has been mentioned in the Discussion section (please see Lines 312-314).*
3. The results are promising and provide a pretty good idea about the usefulness of tp0751 as a component in vaccination against syphilis. However, the presented data are currently based on low animal numbers and must be interpreted carefully. *We agree with the Reviewer and have included a statement in the Discussion explaining this, see Lines 312-314.*

4. Ideally, animal numbers need to be upgraded. I would say group size must be at least 4-5/group. It looks like some more animals were used for vaccination against tp0751 (e.g., animal Im26 pops-up all over sudden in Supplementary Fig. 2). I wonder if some analysis can be supported by additional data sets to increase animal numbers. However, in case upgrading animal number is not possible, the statistical analysis needs a thorough revision. *We agree that the ideal situation would be to perform additional experiments with more animals. However, considering that the experiment took approximately 40 months to complete, this addition is not possible within the timeline of this manuscript submission or the confines of the budget that supports this research. The proposed plan is to perform a scaled-up experiment within the next one to two years using larger animal group sizes and a longer duration to test the extent of immunity generated. No additional animals were included in the experiment reported in this manuscript; lm26 represents an animal from a prior experiment where a single animal was immunized with Tp0751 to generate antiserum. For the original manuscript submission, we included individual graphs showing Im26 as a comparator for the different titers of the immunized animals. For the revised manuscript submission, we have simplified Supplementary Fig. 2 and now only show a single panel which includes all the sera titers, in comparison to Im26. Please refer to Lines 464-466 within the revised manuscript for an explanation.*
5. At least some discussion should be included that the presented data are limited in that the power is not (always?) sufficient. Having addressed this issue, it would be good to make the raw data available in the SOM. *In the revised manuscript we have included a statement in the discussion about the limitations of the study due to the small animal numbers; please see Lines 312-314. The mean values of the technical replicates for individual rabbits can now be found in Table 1.*

Additional points:

- L. 31: Normally tp0751 should be written in italics (pl. check w/ Journal stile). *The standard convention is for genes to be italicized and proteins to remain in standard font. This is also the standard for articles published in Nature Communications, and we have followed that convention in this manuscript.*
- L. 46 and 47: Suggest to add (%). *This has been added, please see Lines 46 and 47.*
- L. 61: Is ‘desensitization’ a real issue in the clinic? It is often recommended by CDC and others, but how often does it really occur. I think it is not necessary to make this point here. There are enough plausible reasons why a vaccine is needed. *While we appreciate the Reviewer’s comment, the important point here is that pregnant women cannot receive treatment with any antibiotic other than penicillin because of the possibility of failed treatment due to antibiotic resistance, and the significant public health consequence of congenital syphilis. Penicillin allergies are a true phenomenon, and the desensitization process (which does, unfortunately, occur in the clinic) presents health risks to the pregnant female and the fetus. We respectfully feel this is an important point in favour of preventative vaccination as opposed to therapeutic, potentially dangerous, antibiotic treatment.*

- L. 66 and 71: The authors should be more accurate here. Talking about T. pallidum would include ssp. pertenue and endemicum. Not all subspecies are reported to cause congenital infection etc. *We thank the Reviewer for alerting us to this omission. We have now included the correct subspecies (Line 66).*
- L. 103 and 349ff. and 357 ff.: Citation of submitted work. Is this conform to the Journal guidelines? It seems weird that some authors are listed with full names, others are just provided as initials. I guess "Parker et al., unpublished data" or similar should do the job. *This manuscript has now been published. Please refer to Reference 34.*
- Fig 1.: Should be improved by the addition of details about the immunization process. Also, here and throughout the MS, the subsp. and strain that was used to challenge animals is not addressed. *We thank the Reviewer for mentioning these points. We have now included details regarding the immunization process within the legend for Figure 1. We have also included the subspecies in the abstract (pallidum; see Line 29), have included the statement "…with the spirochete Treponema pallidum subsp. pallidum (hereafter referred to as T. pallidum)" on Lines 41-42, and have included the strain used for challenge (Nichols), as shown in Lines 125 and 414.*
- Fig. 2: Although raw data are not available, the power of these analysis seems to be quite low. It is not lege artis to present non-parametric data as mean values since this assumes Gaussian distribution. *We thank the reviewer for noticing this error. Since we are unable to assume Gaussian distribution for figures throughout the manuscript, we have adjusted all figures to display median values +/- interquartile range or 95% confidence interval.*
- L. 139: Which subsp./strain? *The subspecies and strain have now been indicated.*
- Extended Data Table 1: Time points do not become clear; values are mean(?) but no SD/SEM is presented? It would be helpful to get a more detailed dataset. *Extended Data Table 1 (now Table 1) has been altered to include the time points for the collection of each dataset and, where possible, values are now presented as mean +/- SEM. We hope that the dataset now includes enough detail.*
- L. 153: Was analysis also supported by qPCR? *qPCR was also performed on primary lesions, however, a different result was observed. This is explored in the discussion Lines 242-252.*
- Fig. 3: I suggest to change the ordinate in B. to D. into scientific scaling similar to graph A. *The ordinate has been changed to scientific scaling for Fig. 3b-d.*
- I may have missed it, but Im1 has a huge variation. It would be good to explain this in the Discussion as well as the reason why only Im2 was tested positive in bone tissue. *Please refer to Table 1, which shows the mean and SE for each individual rabbit. While Im1 displays higher variation than the other immunized rabbits, the variation is similar to that of the control rabbits. The reason underlying the positive bone result for Im2 is unknown, but this finding relates back to the difficulties of working with a small sample size and an outbred animal model. This has been discussed in Lines 312-314.*

- L. 184 ff.: Why was analysis on the RNA level not included? This should provide reliable data on viability. *At the time that the experiment was performed we did not collect samples in a manner that would facilitate analysis at the RNA level. This is an excellent suggestion and is a methodology that we are including in the future, scaled up experiments. This comment is included in the revised manuscript, please refer to Lines 170-172.*
- L. 242: TP DNA was detected at least in spleen tissue of Im3 (Fig. 3). The statement seems to be incorrect. *We thank the Reviewer for noticing this error. We have revised this statement to fix this error (please refer to Lines 176-183).*
- Fig. 4: Abscissa > Average cell count ..per what? *The data now presents the median cell count per field of view, please refer to the y-axes in Fig 4.*
- L 250 ff.: “Field of view” is not defined (which magnification; staining?) *The magnification (400X) and staining procedures have now been included in the legend for Fig 4 and in the materials and methods section (please see Line 529 and Lines 525-529, respectively).*
- L. 285: Again, RNA based methods would have been useful here. *We agree, please see our response above.*
- L. 95 ff.: There is no information on the T. pallidum strain that was used to challenge the animals. *This has now been included. Please see Line 414-415.*
- L. 327: homing. *We thank the Reviewer for noticing this error, which has now been fixed (Line 305).*
- L. 329: Again, when outbreed animals and low animal numbers come together it is challenging to achieve adequate statistical power. This should be reflected in the study design. *We agree, and have added a comment discussing this limitation into the discussion (see Lines 312-314).*
- L. 361: The discussion is missing any aspects of multicomponent vaccination for TP (e.g. tprK + tp0571). *We have now included a discussion regarding the potential for inclusion of this protein in a multicomponent vaccine (see Line 325-335).*
- L. 369: “protects against Treponema dissemination” this is not shown by the data. Better “significantly reduces dissemination…” *This revision has been made, see Line 368.*
- L. 398: FTA-ABS serology IgG or IgM or both? Which test was used (manufacturer?) *FTA-ABS serology was performed using IgG only*. *Details regarding this methodology have now been included, please see Lines 451-453.*
- L. 402: Do I understand it right that one immunization was 4x0.1 ml SC plus 2x 0,04 ml IM? What was the reason for so many injections? *This is correct. This is the immunization regimen recommended for the adjuvant, TiterMax Gold, that was used in this study, and administration of this many injections is designed to generate a robust antibody response. This clarification has now been added in the Materials and Methods section, see Lines 400-411.*
- L. 408: Again, which subsp. and strain? *This information has now been included, see Lines 424-425.*

- L. 413: Fig. 1 does not illustrate what is written here. Why were only the ctrl animals bled and where do the 10 days come from? *Control animals were bled 10 days prior to challenge to establish negative VDRL serology. We have explained the procedure in more detail in the Materials and Methods section, Lines 421-423.*
- L. 423: RIT > first time use in the text (apart from Fig. legend). *This has now been defined, please refer to Line 433.*
- L. 484: Why was the 84°C step included? *Through optimization the 84ºC step was included as an additional, optimized extension step. An explanation has now been added, please refer to Lines 511-512.*
- L. 494 ff: Important details are missing about stainings that were used. Also, why was IHC not conducted to depict TP? *Staining details have now been included in the materials and methods (please refer to Lines 526-529). Steiner silver staining was used to stain for Tp instead of IHC; this has now been included in the manuscript (Lines 530-531).*
- L. 499: Field of view must be defined. TP often causes perivascular-inflammation and thus inflammatory cells are not normal distributed in the tissue. How were the fields of view standardized? Which area of skin was analyzed? *The field of view has now been defined, please see Lines 529-531. Punch biopsies were vertically sectioned (4 µm) and therefore encompassed all layers between the epidermis and muscular layer. Fields of view from each section were randomly selected. These details have now been included, please see Lines 530-531.*
- L. 507 ff: see the initial statement on statistics. *Please refer to our comment above regarding the statistics.*
- L. 609: T(reponema) must be all italic. *We thank the Reviewer for noticing this error, which has now been fixed.*
- L. 687: It does not become clear why titers are displayed relative to serum from an animal that is not part of the study (Im26). *Titers were displayed relative to Im26 to account for inter-experiment variability that was observed in separate ELISAs for the serum from each rabbit.*
- Extended Data Tabel 2/L. 712: If FTA-ABS is never determined you may want to exclude it from the table. The information is useless. *We thank the Reviewer for noticing this; FTA-ABS has been removed from this table (now Supplementary Table 1).*
- Extended Data Table 3: Why are d147-185 missing for Im3? *The rabbit was euthanized on day 134. This has been clarified in the Results section (please refer to Lines 443-445).*

<u>Reviewer #2</u>

1. "…while some of the findings were convincing at the tissue level, others were less so, as evidenced by comments such as " …difficult to reach a definitive conclusion concerning the true lesion burden …" or "It is likely that both of these proposed mechanisms contributed to

the proposed mechanisms …". *We agree with the Reviewer, but included this wording because we wanted to accurately represent our results. The finer mechanistic details of the protection that is generated via Tp0751 immunization will be the subject of a future, larger-scale protection experiment that is beyond the scope of the current manuscript.*

2. One aspect that would significantly strengthen the manuscript would be if the authors could link their whole animal and tissue level observations with some underlying immunological mechanisms. While they did observe the broad types of cells at the primary lesion sites, it would have been more useful if they had analysed specific Tp0751 antigen responses in defined T cell types. *We agree with the Reviewer, and thank them for this suggestion, which would be a good addition to the future larger scale protection experiment.*
3. The authors also suggest that their positive observation may have been " …suggestive of local production of Tp0751-specific antibodies". Why didn't they measure these antibody levels and perhaps even test for their functional role; neutralisation or binding. By including these additional levels of analyses the study results would be more likely to have interest and impact in the field. *We would like to respectfully point out that we did measure anti-Tp0751 antibody levels; these results are shown in Supplementary Fig. 2. We did not, however, measure their functional role, as the Reviewer correctly points out. Due to the complexities associated with working with T. pallidum, including the inability to culture the bacterium and the rapidity with which it dies outside of a rabbit or human host, these experiments are technically very difficult to perform. The context in which the statement was made was that the higher B cell numbers that were observed via histology were likely indicative of cells secreting Tp0751-specific antibodies. This statement has now been clarified (please refer to Lines 281-283).*

<u>Reviewer #3</u>

1. They have found that immunized rabbits have a reduced bacterial load measured by PCR in the organs as compared to unimmunized rabbits, but the vaccine apparently did not provide sterile protection which would be ideal. In the second part of the study, however, the rabbit infectivity testing negative results point towards the desired sterile protection. *We expect that the vaccine administration can be optimized, in terms of antigen dose, adjuvant, number of immunizations etc. Through these refinements we are optimistic that we can achieve sterile immunity in the original, immunized animals; however, these refinements are beyond the scope of the current manuscript. We have included a statement along these lines in the discussion (see Lines 316-321).*
2. The claims of the paper are novel and the vaccine may have an important role in mitigation of syphilis complications, but would not work to prevent the disease and ultimately stop transmission. Although the data seems not to be conclusive, the results are most interesting and timely for the development of a fully protective syphilis vaccine. *We thank the Reviewer for their positive comments. The current formulation of the vaccine is predicted to prevent transmission at the secondary stage (due to the inability of the treponemes to disseminate from the local site of infection to establish the highly infectious rash characteristic of secondary infection) but it is unclear if it would prevent the transmission at the primary stage of infection; the treponemal burden at the site of challenge may represent dead treponemes*

that would not be able to transmit infection. These experiments will be performed in future studies that are beyond the scope of the current manuscript.

3. Would the authors consider having used a group of only three animals as a limitation given the discrepant results between animals? *Yes, using this small of a number of animals is clearly a limitation of the study, this is now discussed in Lines 312-314. However, due to current animal care guidelines the emphasis is on the 3 R's (replace, reduce, and refine) and our mandate, in this initial experiment, was to "reduce" the number of animals used for the initial experiment, in case the vaccinogen was ineffective at providing protection. We now know it is worthwhile to move forward with larger animal groups in a future experiment; however, this is beyond the scope of the current manuscript.*
4. Despite the authors' argument that PCR may have detected dead organisms may be true, the discrepant result for PCR vs microscopic assessment of primary lesions (Rabbit 1) needs to be interpreted with caution. *We agree with the Reviewer, that this needs to be treated with caution, and we have included a statement in the revised manuscript to this effect. Please refer to Lines 249-252.*
5. Could the lower number of Treponema in the inoculation site of Rabbit 3 be related to a decreased proliferation? The current explanation (i.e. greater number of cells migration) seems contradictory to the greater immune cell infiltration in the skin of Rabbit 3. *Yes, this a possibility, and this theory has now been added to the results (see Lines 205-207).*
6. qPCR results in Extended Data Table 1 do not seem to correspond to Figure 3. There might be a typo for the Liver and Spleen Ct1 and Ct2 results: where it says "10 squared" should read "103". Also Liver Im2 seems to have the same mistake. Please confirm since these are the most critical findings of the study. *We have double-checked these values, in both Fig. 3 and Supplementary Table 1, and they are accurate as listed. The values listed in Extended Data Table 1 (now Table 1) represent the average flaA copy number per µg of rabbit DNA from three technical replicates for each rabbit, whereas each value plotted in Fig. 3 represents one technical replicate and the median and IQR are displayed between rabbits. The average values for the liver and spleen from Ct1 are 5.2 x 10exp3 (liver) and 1.91 x 10exp3 (spleen) and for Ct2 are 2.47 x 10exp3 (liver) and 1.55 x 10exp3 (spleen), compared to 5.85 x 10exp2 (liver) and 7.11 x 10exp2 (spleen) for Im1; 1.11 x 10exp3 (liver) and 9.16 x 10exp2 (spleen) for Im2; and 1.81 x 10exp2 (liver) and 2.34 x 10exp2 (spleen) for Im3. This does match the results shown in Fig. 3; there are consistently fewer treponemes found in the organs distant from the site of challenge in the immunized rabbits compared to the control rabbits. .*
7. I would recommend that Extended Table 1 is fitted into the manuscript to facilitate reading comprehension. *We agree with this excellent suggestion by the Reviewer. This Table has now been moved into the main manuscript, and comprises Table 1.*
8. Could the popliteal lymph node for Rabbit 3 have been tested using PCR to confirm the hypothesis of greater load of dead treponemes draining to the lymph node? *We had the same thought as the Reviewer when we were performing this experiment. However, we did not want to compromise the integrity of the lymph node when we were performing the transfer to the naïve rabbit, and instead wanted to transfer the <u>entire</u> lymph node from each of the control and immunized animals to avoid any inadvertent introduction of bias. For example, if a portion of the lymph node was collected to extract RNA or DNA, would we be*

able to take a consistent portion of each lymph node from the control and immunized animals? In future experiments we have plans to address this challenge, since we agree the viability of the treponemes in the lymph nodes used for transfer needs to be assessed.

9. Authors may want to revise the statistical analysis of lesions progression (i.e. ulceration and diameter). Currently individual lesions on each rabbit are analyzed as independent items. However, the lesion sites in one rabbit may have correlated results in terms of lesion progression compared to other rabbits. The reason is that lesions in a single rabbit are exposed to the same protective immunity; therefore these are not fully independent units and may need specific statistical methods for comparison (e.g. multilevel model). *We agree that lesions cannot be assessed as independent units on a single rabbit*. *In the analysis individual lesions are regarded as technical replicates for individual rabbit, and the average of these technical replicates is compared between rabbits. We have altered Fig. 2a-b to utilize non-linear regression and an extra sum of squares F-test to compare the slope of the lesion diameter or ulceration over the 14 day post-challenge period between the control and immunized groups using the average between all 10 lesions for each rabbit. This analysis has been explained more clearly in the Fig 2. legend and in the materials and methods section, see Lines 542-552.*
10. The overall results are very promising but with some contradictions that seem difficult to overcome with a small sample (n=3) study like this, the authors may want to discuss what are the next steps. *We have now included a paragraph in the discussion to address the proposed next steps. Please refer to Lines 314-324.*
11. It would be good that the researchers explain the path towards development of a syphilis vaccine that can be trialed in humans and possibly to acknowledge certain gaps, including that the current vaccine would not be effective to prevent chancre development and that the current study does not ensure the vaccine is broadly protective against different T pallidum strains. *We agree, and have included a paragraph within the discussion addressing these important points (please refer to our response to comment 10 above).*
12. The researchers should more clearly explain why it is important that Tp0751 belongs to the same protein family as the meningococcus b vaccine. If one reads the abstract, as it is currently explained, some people could think that the meningococcus vaccine is protective for syphilis. *We have clarified the text surrounding the mention of the meningococcus vaccinogens and the relevance that the T. pallidum protein belongs to the same structural protein family. Please refer to Lines 33-35 and Lines 348-365.*
13. Discrepant results of experiments could be more clearly written. *We have attempted to write the results in as clear a manner as possible. However, it should be mentioned that we do not have all the answers regarding these discrepant results; instead a complete understanding must await further study.*
14. Readers outside the discipline would benefit of a schematic of the main result to accompany publication. *We thank the Reviewer for this excellent suggestion. In response to the Reviewer's comment, we have moved Extended Data Figure 3 into the main text of the manuscript (now Fig. 5b) and have added a panel (Fig. 5a) reporting the main result of the experiment to enhance understanding of the significance of the results.*

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have addressed the reviewer's comments and the MS has improved significantly. Weaknesses (e.g. animal numbers) and challenges of the current work are now appropriately addressed in the MS as well as the statistical analysis has become more accurate.

To me the MS is suitable for publication.

Reviewer #2 (Remarks to the Author):

Overall, the authors have done a good job in addressing the majority of the suggestions and criticisms raised by the reviewers. However, one aspect remains that, for this reviewer, is still not adequately addressed. The number of animals in each treatment group is very low and hence the strength of the claims made from this data is perhaps still not sufficiently supported. The authors have now expanded their statistical analyses, which has certainly helped, however the main finding of the manuscript still hinges on the interpretations from the small number of animals per group. The authors even suggested that the principle of the "3Rs (replace, reduce and refine)" is a reason for the small group sizes. However, any reduction in animal usage must still result in a statistically significant finding. While I appreciate that the animal experiments are complex and of long duration, making them difficult to perform, the lack of strong statistical significance of the main result is still somewhat unfortunate.

Reviewer #3 (Remarks to the Author):

I am satisfied with the authors answers and corrections. I would like to congratulate the authors on this great work.