We are grateful to the reviewers for their insightful reviews and feedback. We have incorporated fixes to simple issues such as typos and missing references and do not address those issues here.

Reviewer 1, 2, and 4 commented that is that it is not clear how the Renyi divergence bounds in the paper relate/translate to private samplers. An additional issue that Reviewer 2 points out is that while we show that $D_\alpha(P||R)$ is small for P being the distribution of the discrete finite-time dynamics and R being the stationary distribution, we need to also show that $D_\alpha(R||P)$ to get efficient private samplers. We address this below. Assuming that we can bound both $D_\alpha(P||R)$ and $D_\alpha(R||P)$, the conversion to (ϵ, δ)-DP follows from the following argument. Suppose the underlying mechanism guarantees that R, R' satisfy (ϵ, δ)-differential privacy. Fact 9 shows that if for $\alpha = 1 + 2\ln(1/\delta)/\epsilon$, $D_\alpha(P||R)$ and $D_\alpha(R||P)$ are both at most $\epsilon/2$, then P, R also satisfy the (bi-directional) divergence guarantee of (ϵ, δ)-differential privacy (and the same for P', R'). Then by composition theorems, P, P' satisfy $(3\epsilon, 3\delta)$-differential privacy. In our revisions to the paper, we will include a formal theorem/proof for this argument.

Reviewer 2 points out that our result "bypasses" the problem of bounding the bias of the discrete dynamics' stationary distribution. This is indeed the case and we discuss this briefly in our introduction. This can be seen as both a strength and weakness of our approach. We will also add this as an interesting future direction.

Reviewer 2 suggests making the results more rigorous by verifying the "mild conditions" under which Langevin dynamics converges. The result of Vempala and Wibisono suggests that a sufficient condition is strong convexity, which we assume in the paper. This will be made more explicit in the introduction.

Reviewers 3 and 4 suggest further comparisons of our iteration complexity to previous results and clarifying whether the result is optimal or near-optimal. We do currently remark in our concluding section that our dependence on ϵ is worse than that of results for KL-divergence, and also point out that a factor of $1/\epsilon^2$ seems inherent to our analysis. Our revision will this discussion our introduction to make these points clearer to the reader. Unfortunately, we do not know of any lower bounds for the iteration complexity needed for Langevin dynamics to converge in Rényi divergence, so we are unable to comment on the optimality of the main result. We do however note that for Differential Privacy applications, the desired ϵ is usually a small constant (say 0.1 or 0.5).

Reviewer 4 asks the benefit of using the unadjusted Langevin process. Part of our goal is to provide a simple and more accessible analysis, and our analysis is most simple when adapted to this algorithm. However, our approach might be useful to bound the discretization error of algorithms with Metropolis steps.

Reviewer 4 asks to highlight the novelty of the proofs. As mentioned in the introduction, we feel the novelty is the simplicity of the approach, requiring almost no stochastic calculus, making the analysis ideally more accessible to e.g. members of the differential privacy community who are less familiar with the stochastic calculus literature.

We end the rebuttal with a discussion on bounding $D_\alpha(R||P)$, that we overlooked in the submission. Technically, the proof is essentially identical to $D_\alpha(P||R)$ case, and we give details for completeness. The results in the submission bound both $D_\alpha(P||Q)$ and $D_\alpha(Q||P)$, where Q is the distribution of the continuous finite-time dynamics. While we state Theorem 8 as bounding $D_\alpha(P||Q)$, we provide a bi-directional divergence bound in Lemma 5, and so the proof of Theorem 8 easily also bounds $D_\alpha(Q||P)$. What remains is to bound $D_\alpha(R||Q)$, since the cited paper of Vempala and Wibisono (VW19) only immediately bounds $D_\alpha(Q||R)$. However, the techniques in that paper can easily be generalized to also derive a bound $D_\alpha(R||Q)$. Due to space constraints we can only provide a high-level summary of this generalization here, but for completeness our revisions include a full explanation in the appendix.

We need two ingredients from VW19 to bound $D_\alpha(R||Q)$. The first is to show that if our initial distribution Q_0 is a Gaussian with the correct variance then ideally $D_\alpha(Q_0||R), D_\alpha(R||Q_0) \lesssim d$ (as opposed to Lemma 9 in our paper and the cited lemma from VW19, which only shows $D_\alpha(Q_0||R) \lesssim d$). This guarantee is not attainable for large α. However, for $Q_0 = N(0, I)$ and any 1-smooth and L-strongly convex f, both divergences are bounded by d if, say, $\alpha = 1 + 1/2L$. The Hypercontractivity Lemma (Lemma 14) in VW19 shows that after running for continuous time t proportional to $\log(\alpha' L)$ (much smaller than the current continuous time bound we use), we will get that $D_{\alpha'}(R||Q_t), D_{\alpha'}(Q_t||R)$ are both also finite and roughly d. The choice of $Q_0 = N(0, I)$ also satisfies the properties stated in Lemma 5 of our paper.

The second is to show that $D_\alpha(R||Q)$ decreases exponentially. We show that the continuous chain (as opposed to the stationary distribution) always satisfies LSI with constant 1/3, and then we can slightly modify Lemma 5 and 6 in VW19 to show that $D_\alpha(R||Q)$ decays exponentially. Our choice of initial distribution Q_0 satisfies LSI with constant 1. The continuous Langevin dynamics can be viewed as the limit of the discrete chain, which is repeated application of a $(1 - \eta/2)$-Lipschitz gradient step followed by adding Gaussian noise $N(0, 2\eta I)$, as η approaches 0. Lemma 16 in VW19 shows that applying a $(1 - \eta/2)$-contractive map increases the LSI constant by at least a multiplicative factor of $(1 - \eta/2)^{-2}$, and adding $N(0, 2\eta I)$ changes the LSI constant from α to at least $\frac{1}{1/\alpha + 2\eta}$ (see e.g. and "Functional Inequalities for Convolution Probability Measures" by Wang and Wang). Taking the limit as η goes to zero, we get that the continuous Langevin dynamics do not cause the LSI constant to decrease below 1/3.