We thank the reviewers for their attentive and thorough reviews! We first respond to common concerns before moving to reviewer-specific ones below. As a general rule, we will fix all minor points/typos.

Common concerns: 1. *The lower bound takes up a lot of space for a result that is subsumed by Duchi-Rogers 19.* As noted in the introduction, we agree that the lower bounds of Duchi-Rogers 19 generalize and subsume our lower bound. However, we released the original preprint of our results (complete with the first adaptation of the SDPI lower bound from Braverman et al.) a few months before the preprint of Duchi-Rogers 19, so we feel it is still an important contribution. We therefore included the lower bound statement in the main body. However, we will add material in the introduction clarifying that Duchi-Rogers subsumes our work, and will move the bulk of the lower bound section to the appendix in the final version of the paper. As a side effect, this will free up space to add the changes promised below.

2. *The proof sketches in the main body are somewhat unclear and disconnected from the proofs in the Supplement.* We will revisit the sketches and add explicit references connecting sketch claims and their precise claims in the Supplement.

Reviewer 1: 1. *Remove the definition of central differential privacy.* We will remove this.

2. *The inclusion/omission of different routines and subroutines is confusing.* We note that all pseudocode for subroutines appears in the Supplement. Given space constraints, we felt it was clearest to include pseudocode for the main routines (to be explicit about how the subroutines fit together) and descriptions for the simpler subroutines. We will revisit and expand our descriptions of the subroutines and move pseudocode to the main body where appropriate.

3. *Do these techniques work for any subgaussian/near-gaussian distribution?* The bulk of our techniques should generalize to sub/gaussian/near-gaussian distributions. A possible exception is the comparison to the Gaussian CDF (via erf) used in the known-variance case, which may suffer given CDFs that are actually not close to Gaussian. However, the slightly worse (but less delicate) Laplace noise method used in the unknown-variance case should perform better.

Reviewer 2: 1. *Where is $\sigma_{\max}/\sigma_{\min}$ in part one of Theorem 1.2?* In the corresponding formal statement, Theorem 4.1, we assume $n = \Omega(\log(\sigma_{\max}/\sigma_{\min}))$. This is the "sufficiently large n" mentioned in the informal Theorem 1.2.

2. *If $R \geq \mu$, but σ can be exponential in n, doesn't Gaboardi et al's bound become vacuous?* When μ (but not σ) is exponential in n, their claim is not vacuous, but their round complexity is still $\Omega(n)$. We will clarify this.

3. *The lower bound only holds when the mean is of an order of the accuracy.* We agree that instance optimal lower bounds are better. However, we think that the worst case optimality shown by our current lower bound is still useful.

4. *Where is the dependence on ε in the centrally private result?* The centrally private upper bound is $O(\sigma\sqrt{\log(1/\beta)/n} + \text{poly}\log(1/\beta)[\varepsilon n])$. Ignoring $\log$ factors, the additional cost of central privacy over the non-private case is the second term $O(1/[\varepsilon n])$ (lower-order in n) whereas local privacy costs $O(1/(\varepsilon\sqrt{n}))$ (not lower-order).

5. *Isn't DJW13 the first local privacy paper to use an SDPI approach?* DJW13 use the term SDPI in a different way. Informally, their SDPI only controls the mutual information between a sample and its locally private output. In contrast, our SDPI also takes advantage of a bound on the mutual information between the distribution-specifying parameter and a sample (and is the first to do so). It is this latter interpretation that is typically the focus of the strong data processing inequality literature. Our SDPI therefore provides strictly stronger mutual information bounds overall.

6. *I am confused by the induction in the proof of Claim 1. Also, how can you use Claim 1 in the proof of Claim 2, which has a different assumption?* Claim 1 is a statement about interval I_j assuming that a property holds for successive supersets of I_j, $I_{j+1}, I_{j+2}, \ldots, I_{L_{\max}}$. Since $u \in I_{L_{\max}}$, the property holds for $j = L_{\max}$, the base case. We will clarify this. For the second point, note that Claim 2 is about the maximum j with $\hat{H}_1^j(M_1(j)) < 0.52k + \psi$. Thus for $j' > j$, $\hat{H}_1^j(M_1(j')) \geq 0.52k + \psi$. This requirement on $j' > j$ is used in Claim 1 (the $M_1(j)$ should be $M_1(j')$ in the second line of Claim 1; we will fix this). Thus we can apply Claim 1 to get $\mu \in I_j$, as written.

8. *Where is privacy used in proving Lemmas 1.4 and 2.4?* In the proof of Lemma 1.4, privacy enters when analyzing $|T - \hat{T}|$ (lines 90-96). In the proof of Lemma 2.4, this happens in lines 150-157. We will better highlight this.

9. *In Equation (3) of the Supplement, why is it the expectation of y_i?* $E[y_i] = E[y]$.

Reviewer 3: 1. *More comparison to existing approaches would be helpful.* We will add more comparison. For example: Gabordi et al. use a similar "binary search" approach, but they do not use our modular single round trick (which dramatically reduces round complexity). They also use a Gaussian noise approach for known variance (less accurate than our use of Braverman et al's CDF-based approach). Karwa and Vadhan employ an approach based on the histogram that crucially relies on central access to the raw data, and is therefore qualitatively different from ours.

2. *What happens in higher dimensions?* If the covariance matrix is diagonal, parallel invocations of our algorithms (for each dimension) would work. In other cases, it is not obvious how to scale the binary search approach used here without incurring penalty exponential in the dimension due to covariance.