First off, thanks to all our reviewers for taking the time to look over our submission and provide thoughtful feedback. Individual responses are below.

Response to Reviewer 1:

Oracle Knowledge of k: In our superset method, you are correct that oracle knowledge of k is assumed both in determining the Bernoulli probability of the matrix entries and in determining the number of rows of the matrix. However, in the absence of exact knowledge of k, we can instead treat these quantities as hyperparameters to be optimized; for the Bernoulli probability, we give some experimental evidence in Appendix B that this parameter is not overly sensitive to tuning, and likely any reasonable estimate of k could be used in place of the exact value. Similar assumptions have been made in previous work such as the BIHT algorithm, but finding new algorithms that work independently of the value of k is an interesting problem.

Noise: We can extend our methods to handle the case of noisy measurements by using "error-correcting list disjunct matrices" (see "Efficiently decodable error-correcting list disjunct matrices and applications" Ngo, Porat and Rudra 2011) in place of standard list disjunct matrices in order to recover the superset of the support, at the cost of slightly more measurements. Since the measurement outcomes are only bits, the *bit-flip* error is the best noise model in this setting, which can be taken care of by the error-correcting list disjunct matrices. Running some further experiments with these matrices and measurement noise would be a good addition.

Comparison with Other Methods: The main reason we have not attempted to give experimental comparison with these methods is that they are designed specifically for the universal setting. In the experimental setting we consider, the measurement matrices have far fewer measurements than what is required for them to be universal; our superset method gives a fairly smooth degradation of performance even when there is a small measurement budget, but this is not the case for other methods. The method of Acharya et al., for instance, with less measurements than what is needed for universality, will have false negatives in the support recovery phase, which results in large error in the second stage. We could provide this comparison experimentally if it would be beneficial, though.

Response to Reviewer 2:

Robustness: While badly misidentifying the superset would indeed lead to poor performance in the second stage, this is only a serious issue if we have false negatives (i.e. we miss coordinates that were in fact in the support). The way our decoding works, this will never happen, there will be only false positives. So in the worst case, the superset just ends up being slightly larger, and somewhat more work has to be done in the second stage. For robustness to measurement noise, see the comments to Reviewer 1 ("Noise"). Handling approximately sparse signals is also an interesting direction – we believe this could be dealt with by slightly modifying the quantization scheme, so that measurements of very small magnitude are treated as having zero magnitude. You are correct that as is, our method will not be robust to approximately sparse signals.

Knowledge of k: Please see comments to Reviewer 1 ("Oracle Knowledge of k"). A reasonable estimate of k will suffice in place of exact knowledge.

Experiments: The idea of using the superset matrix directly with BIHT is a very interesting one, that we will consider in the future version (we believe getting a theoretical result may be challenging). Perhaps using some other matrices such as an all 0/1 matrix with Bernoulli entries would be worth checking as well.

Response to Reviewer 3:

Wording of "optimal number of measurements": You are correct, in both the binary and nonnegative cases it is open whether or not the $k^{3/2}$ term is necessary, so it is unclear whether the number of measurements we obtain is optimal; thanks for the catch. The lower bound of $k\log(n/k) + (k/\epsilon)$ from Acharya et al. does not obviously apply in either of these restricted settings, but it is included in Table 1 for the case of real vectors.

Proof of Theorem 10: The last step follows from substituting $q = O(k/\alpha)$ as you mention, and also that $1/(1 - H_q((k-\alpha)/k)) = \Theta((\alpha/k)\log(\alpha/k))$, which can be seen by expanding out the terms (using the definition $H_q(p) = -p\log_q(p) - (1-p)\log_q(1-p) + p\log_q(q-1)$). We will include this definition and the details of this step for the future.

Novelty: We believe the primary novelty in the submission is in the use of superset recovery in the 1bCS setting, and also in developing the relationships between some of the tools and methods used in group testing and those used in 1bCS.

Thanks also for the further references.