1. **Introduction**
   1. Drug-induced gene expression profiles are useful
   2. Due to lack of data, haven’t yet been able to delve into cell-specificity
   3. LINCS L1000 data has opened up this possibility, but still so many data gaps
   4. We propose to impute this data using the existing data
   5. There is related yet different work
   6. Outline of the paper
2. **Methods**
   1. Notation and terminology
   2. Data processing/filtering and tensor construction
   3. Cross-validation setup
   4. Averaging baselines
   5. KNN baseline
   6. Benchmarking tensor completion algorithms
   7. FaLRTC
   8. Ensemble approach \*\*Need to develop this!
   9. Calling DEGs
   10. Predicting side effects \*\*Need to build this!
3. **Results**
   1. Accuracy evaluation:
      1. Overall
      2. DEGs
   2. Tradeoffs in accuracy with amount of observed data:
      1. Look at the sampling in the drug-cell space- super heterogenous.
      2. Look at where KNN does better, and where Tensor does better
      3. If we split into these four tensors, you can see it more clearly
      4. We can also look at how the accuracy changes with observation density, within a fixed set of drugs and cells.
   3. Prediction accuracy for each drug, gene and cell
   4. Analysis of cell-specificity
   5. Utility of cell-specific signatures for predicting side effects
      1. Tissue-matched profiles have most signal for side effect prediction
      2. Imputed dataset adds value for tissue-specific side effect prediction
4. **Conclusion**
   1. (Reiterate that) cell-specific gene expression is useful, but data gaps have limited cell-specific utility
   2. We’ve demonstrated that it’s feasible to impute within this space and get predictions that are much more accurate than simply averaging. We demonstrate that the imputed data adds value to the observed data for predicting side effects.
   3. Several results are consistent with prior expectations..
   4. There are some tradeoffs between KNN and the tensor completion approach, and the ensemble method is able to leverage this.
   5. Any comments about predicting side effects
   6. We produce testable and usable predictions, although you have to be careful since we’ve averaged over dose and timepoint (or reconsider this!)
   7. There’s lots of low quality signatures in L1000, so be careful!
   8. Lack of established baselines. We’re kind of solving a new problem..
   9. Potential biases
   10. Interesting future directions
   11. 1-2 sentence summary of contributions

**Changes from previous submission:**

1. Add ensemble approach (combination of Tensor and KNN)
2. More compelling demonstration that imputed dataset adds value:
   1. Predicting cell-specific side effect (e.g. a case study with hepatotoxicity)
   2. Cell viability prediction?