Experimental outline for side effect prediction:

Q1: Do tissue-matched signatures contain more signal for tissue-specific side effects?

Q2: Does the imputed data add value for side effect prediction?

Side effect prediction. Binary labels. Map it to a tissue. Map cell lines to a tissue. Choose three to start with, or OMOP side effects (?).

First can see: Does expression correlation correspond to side effect association?

Can use this to select side effects! And can show that

Can use cell-specific signatures. KNN. Weighted/unweighted vote.

The VariancePartition package is intended to identify which factors contribute to the variance in gene expression. But I want to see whether I can use gene expression to predict things….

What if I did a t-SNE embedding and colored them based on certain side effects? hmm too passive.

Maybe just intersect as many side effects as I can with the 2130 drugs. Then for each cell type, do a (hierarchical?) clustering. And show this with the labels. Can judge