When cells be friends: A proposal for playing with Venessa's data

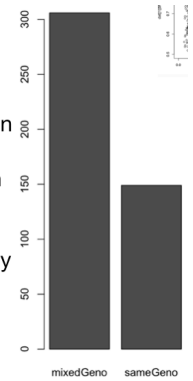
1) Overall question: Can doublets be useful?

- Is there a clever strategy for delineating "accidental" (fluidic) doublets from "actual" (tissue) doublets
- Can celltypes be teased out?
- Can these tell you about the tissue?

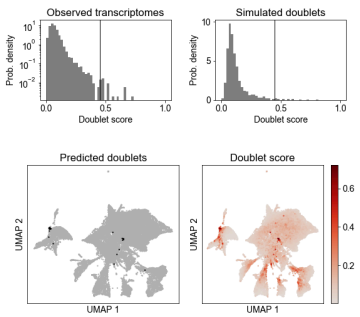
2) Prelim data

- I have used demuxlet to assign genotypes and genotype doublets in mixed cell population
  - Am also trouble shooting FreeBayes methods to call variants direct from scRNA
- I have used scrublet to assign expression based doublets in a solid tumour
  - There are a bunch of tools, this one is fine (sticky cells are more commonly ID)
- I can use expression profiles of ligand/ receptors to tell me about whether doublet profiles are compatible
  - But much work is put in to prevent doublets getting into samples

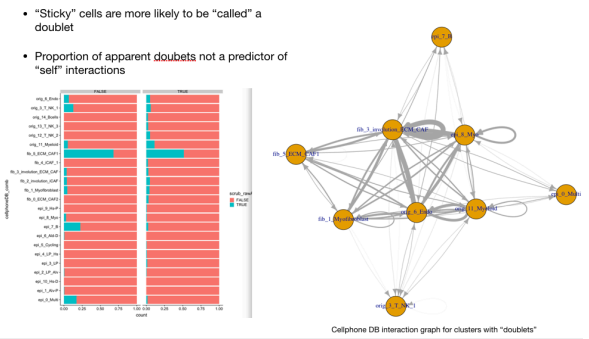
- VCF guided duplicate score
- Tuneable on relative proportion
- Solid estimation of population (51%) for sample 2
- Cutoff 75% doublet probability give 7.2% doublet rate



- Coz its a dataset im familiar with
- Coz we have a decent handle on whats/what
- Scrublet calls border cells as doublets (surprise)



- "Sticky" cells are more likely to be "called" a doublet
- Proportion of apparent doubets not a predictor of "self" interactions



3) Proposal

- A complex mixture of cell types with an expected spread of cell dissociation would be ideal for putting this all together
  - Mechanical separation is more likely to retain biologically relevant duplets
- I want to use Vanessa's data to develop an end-to end doublet ID (genotype and profile) and reconstruction of physical TME and tumour interactions
  - If I can make one that makes sense to Venessa, id be confident rolling it out elsewhere