

fQTLs identified are extremely sensitive to the selections of covariates used

We need to settle on a particular selection of covariates

Include different subsets of covariates
for fQTL identification

$$\binom{4}{0} = 1$$

$$\binom{4}{1} = 4$$

$$\binom{4}{2} = 6$$

$$\binom{4}{3} = 4$$

$$\binom{4}{4} = 1$$

1) Gender

2) Disease status

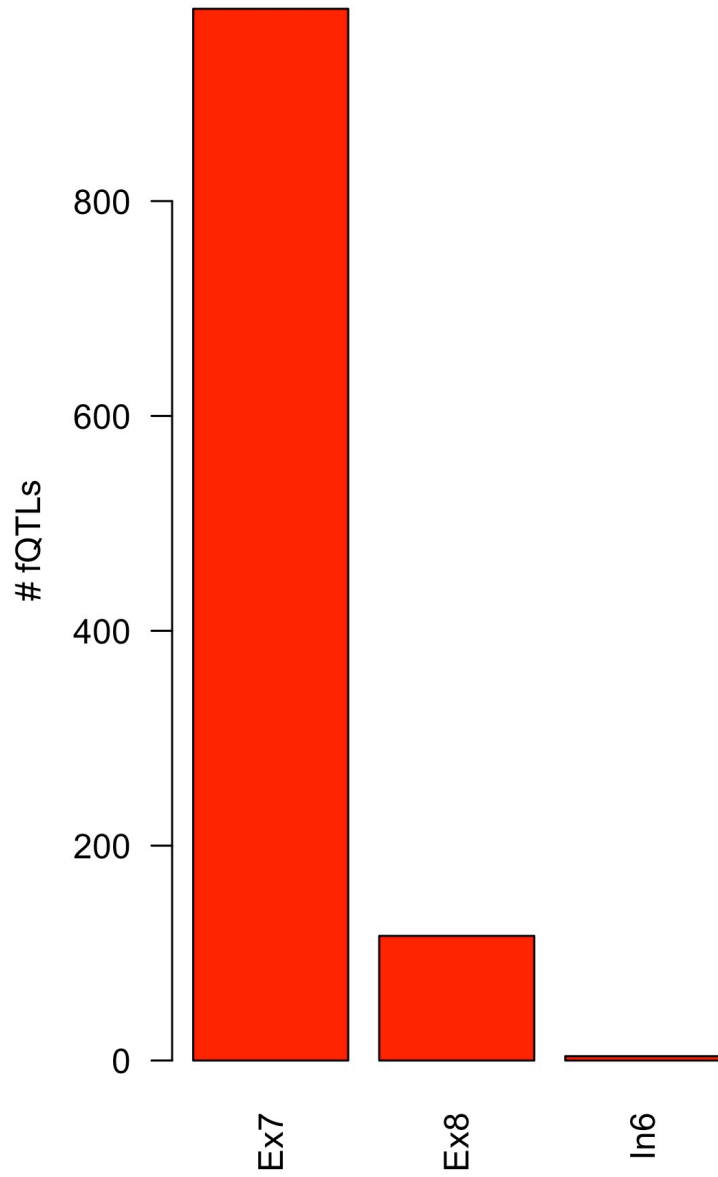
Affective_Disorder
Autism_Spectrum_Disorder
Bipolar_Disorder
Control
Schizophrenia

3) Study

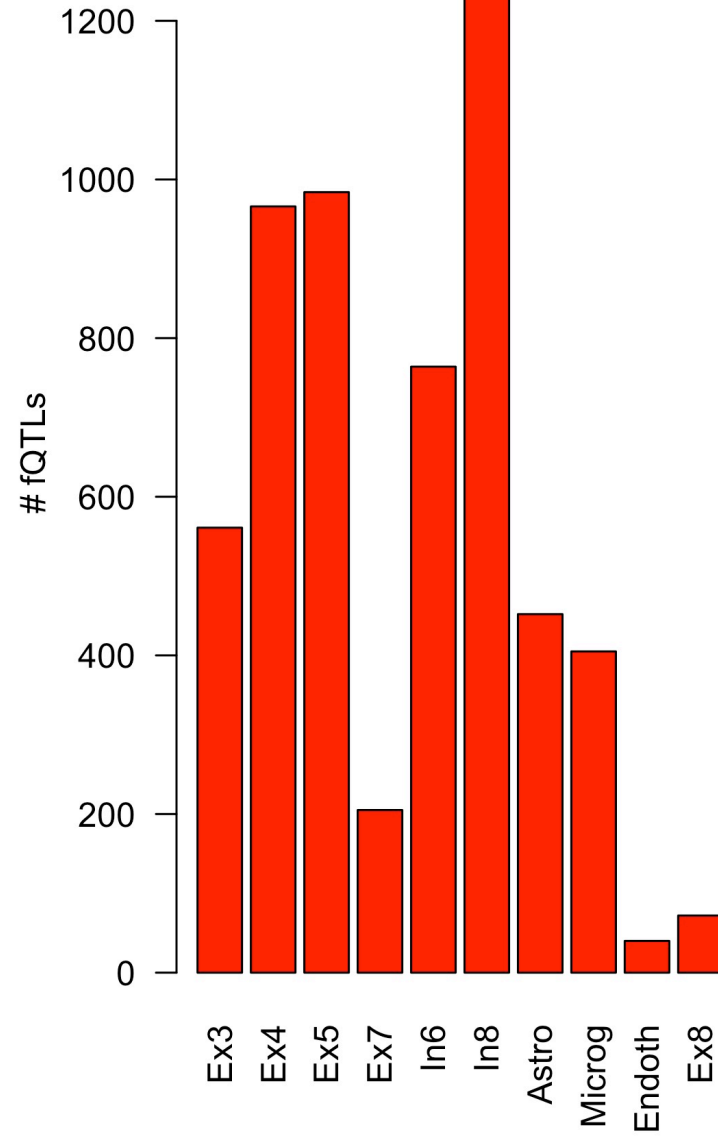
Bipseq
BrainGVEX
Brainspan
CMC
CMC_HBCC
LIBD
UCLA-ASD_DFC
Yale-ASD_DFC
GTEx_DFC

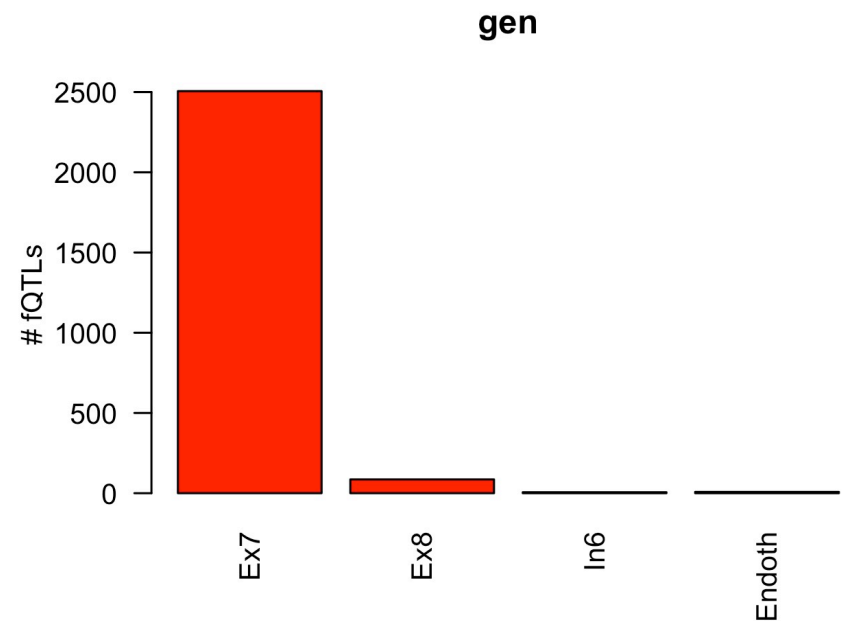
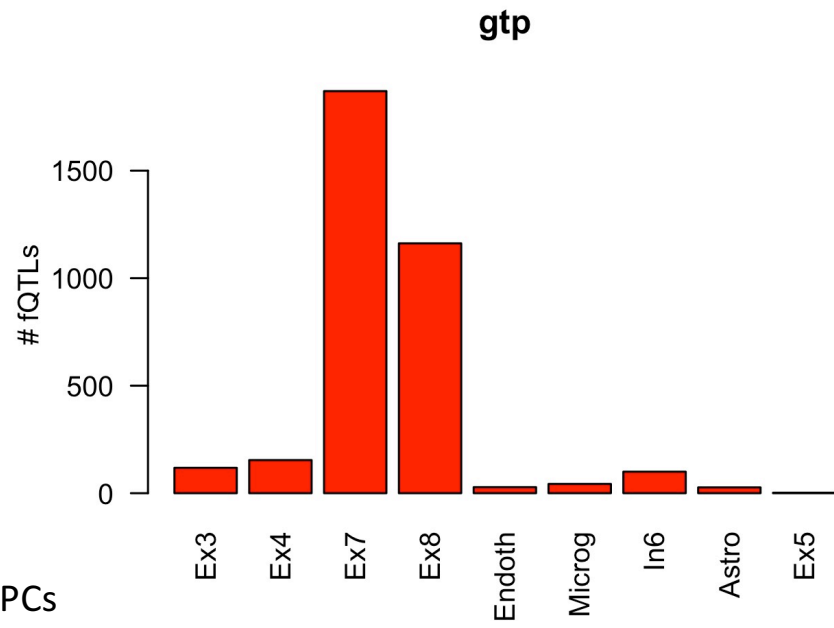
4) Genotype PCs (all 3)

ALL

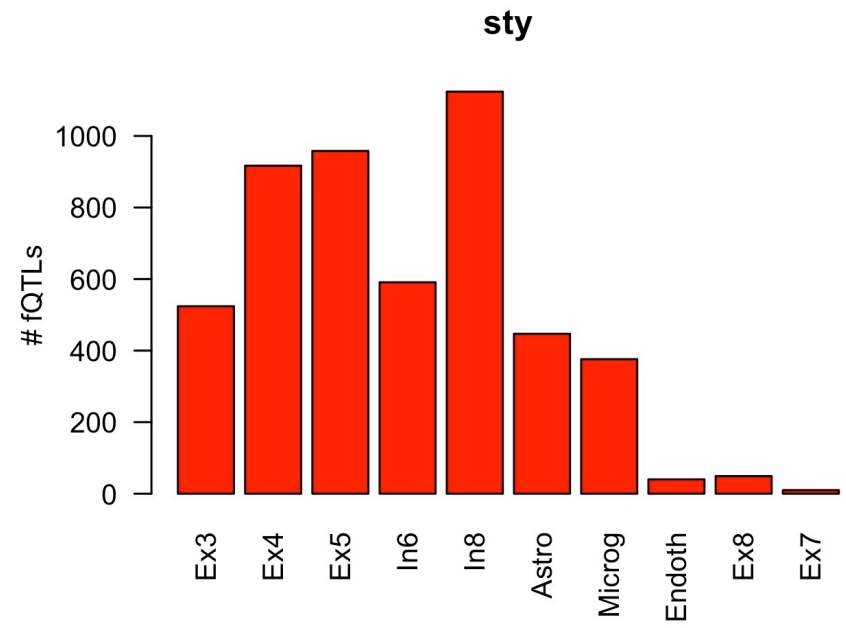
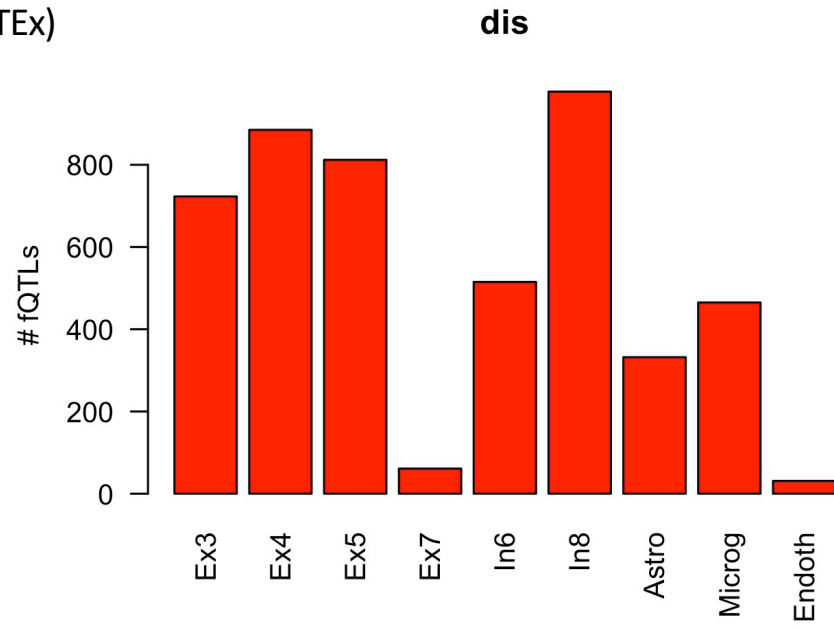


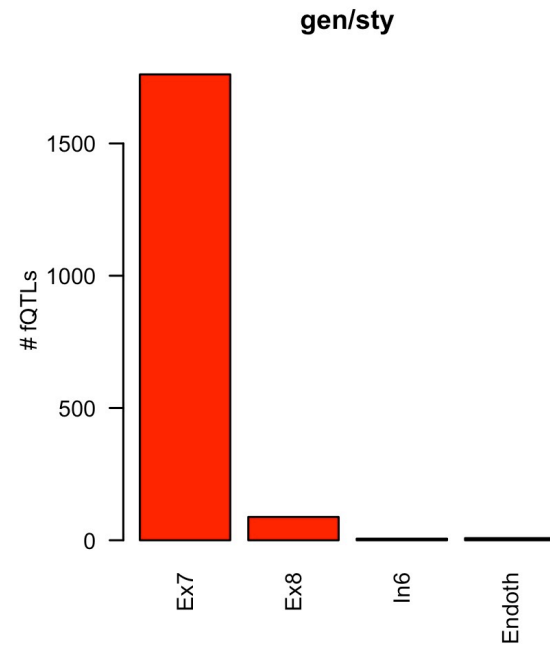
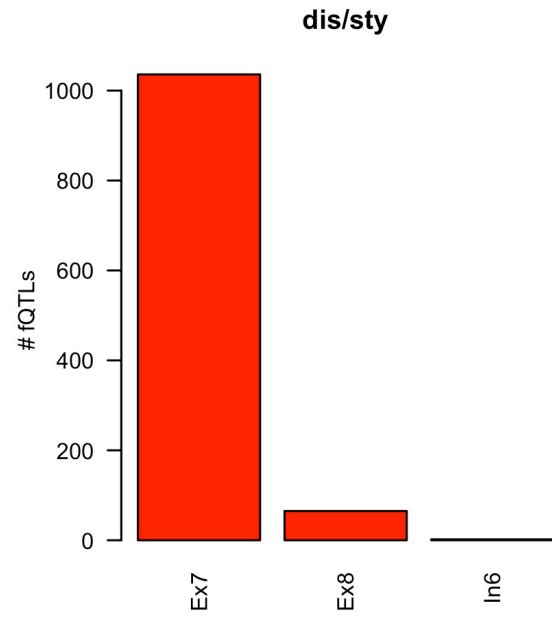
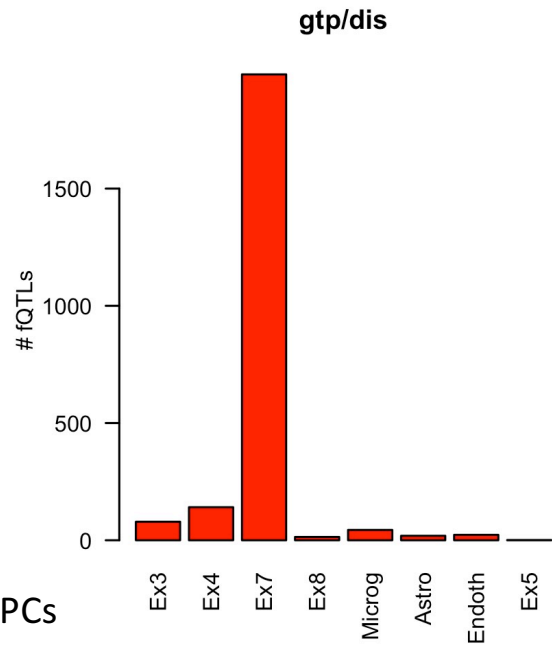
NO covariates



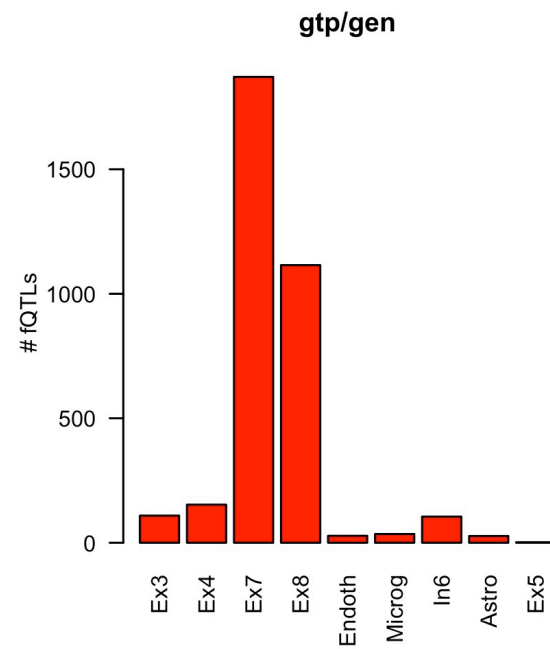
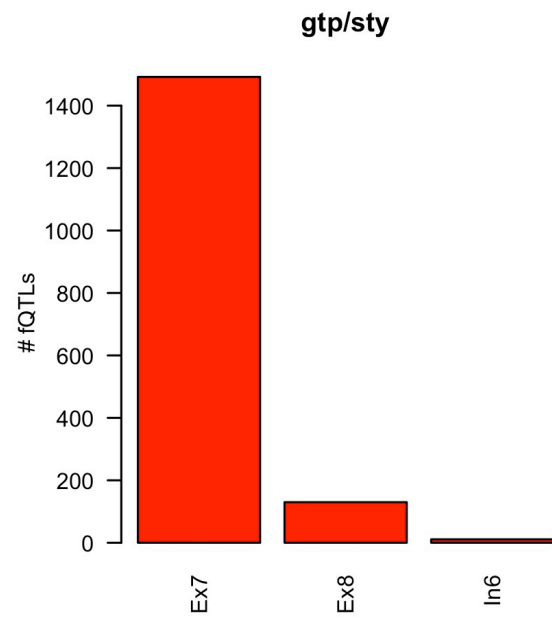
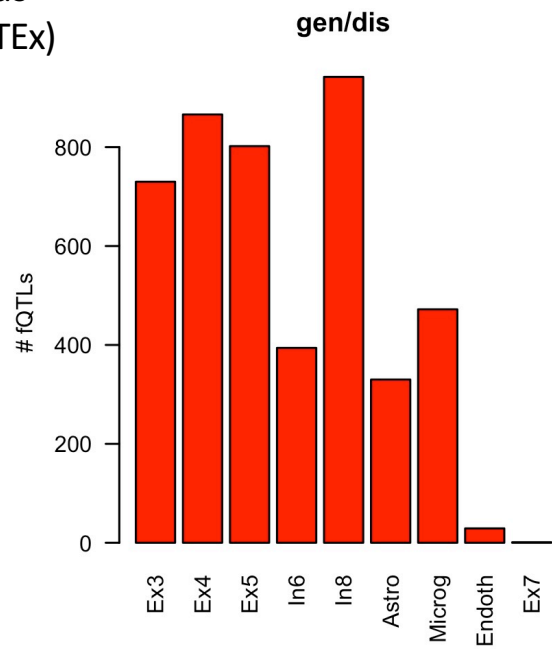


"gtp" - 3 genotype PCs
 "gen" - gender
 "dis" - disease status
 "sty" - study (ex: GTEx)

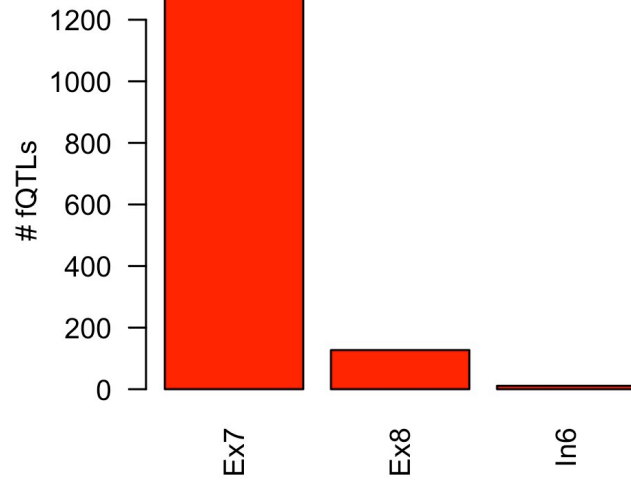




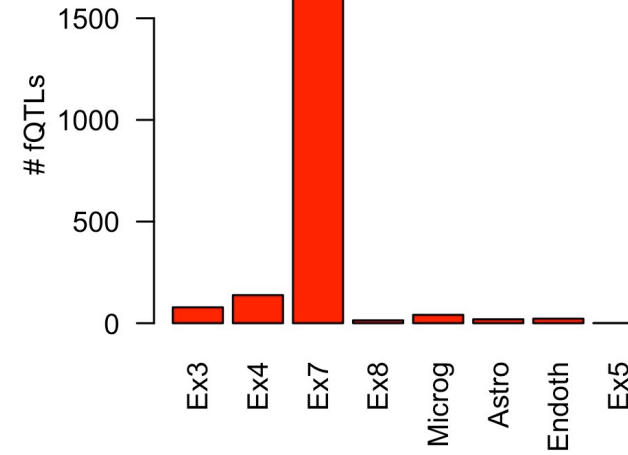
"gtp" - 3 genotype PCs
 "gen" - gender
 "dis" - disease status
 "sty" - study (ex: GTEx)



gtp/gen/sty

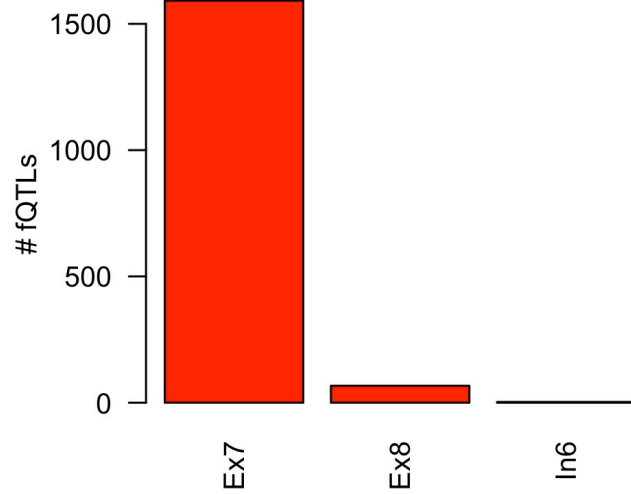


gtp/gen/dis

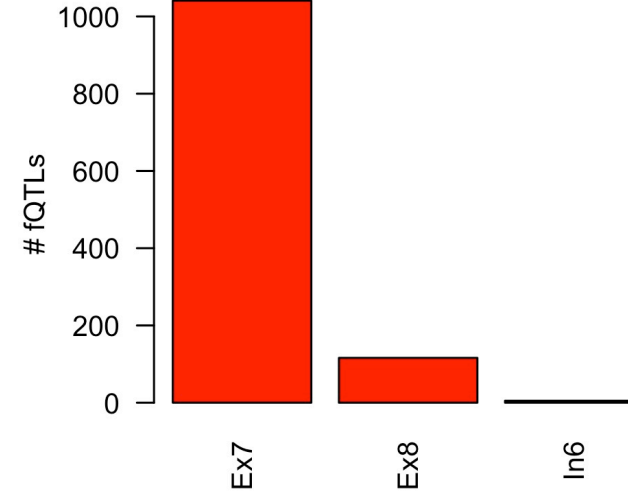


"gtp" - 3 genotype PCs
"gen" - gender
"dis" - disease status
"sty" - study (ex: GTEx)

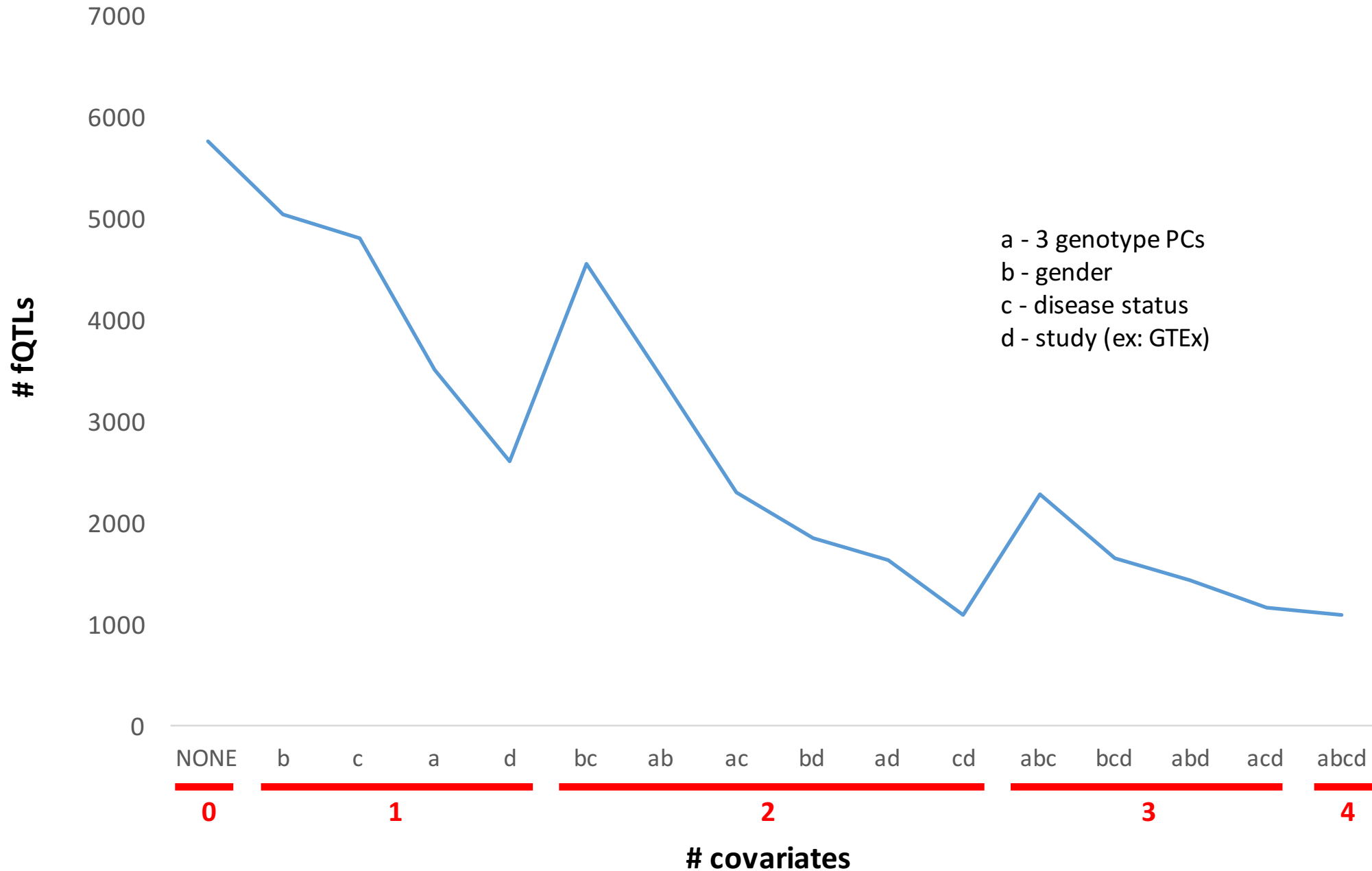
gen/dis/sty



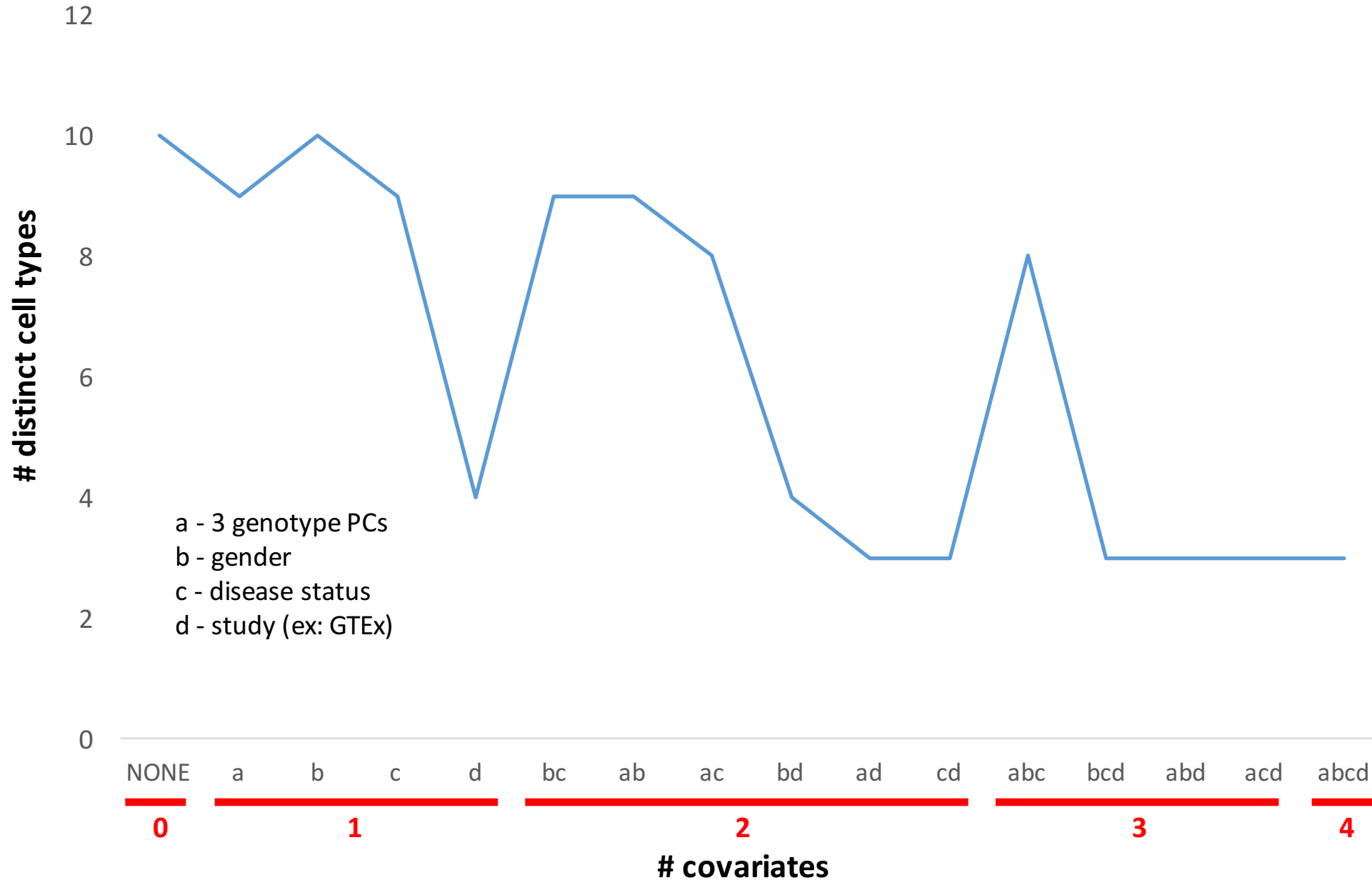
gtp/dis/sty



fQTLs as a function of the number of covariates used



distinct cell types as a function of the number of covariates used



Main points

- 1) The fQTLs identified are extremely sensitive to the particular choice of covariates. Some highly abundant cell types in one set of covariates can completely disappear in another set of covariates (even if the same number of covariates are used).
- 2) Increasing the number of covariates tends to diminish both the number of fQTLs and the number distinct cell types (though these relationships are not strictly monotonic).
- 3) The choice of covariates to use is not obvious – there's a tradeoff between the degree of correction (i.e., # covariates) and potentially interesting biology (diversity and number in fQTLs).
- 4) fQTLs were calculated for all 16 combinations of the 4 covariates -- each with 10 different Bonferroni-corrected p-values [we have 160 different lists of fQTLs]

Misc Post-notes

“This model assumes that phenotypes are normally distributed.”

# fQTLs		
	normalized	not normalized
with covariates	1099	325
w/o covariates	5774	5590

# distinct cell types		
	normalized	not normalized
w/covariates	3	6
w/o covariates	10	10

Note – overlaps w/cis-eQTLs: ~1000 fQTL SNVs overlap w/~100 cis-eQTL SNVs (all covariates, p-val 0.05)