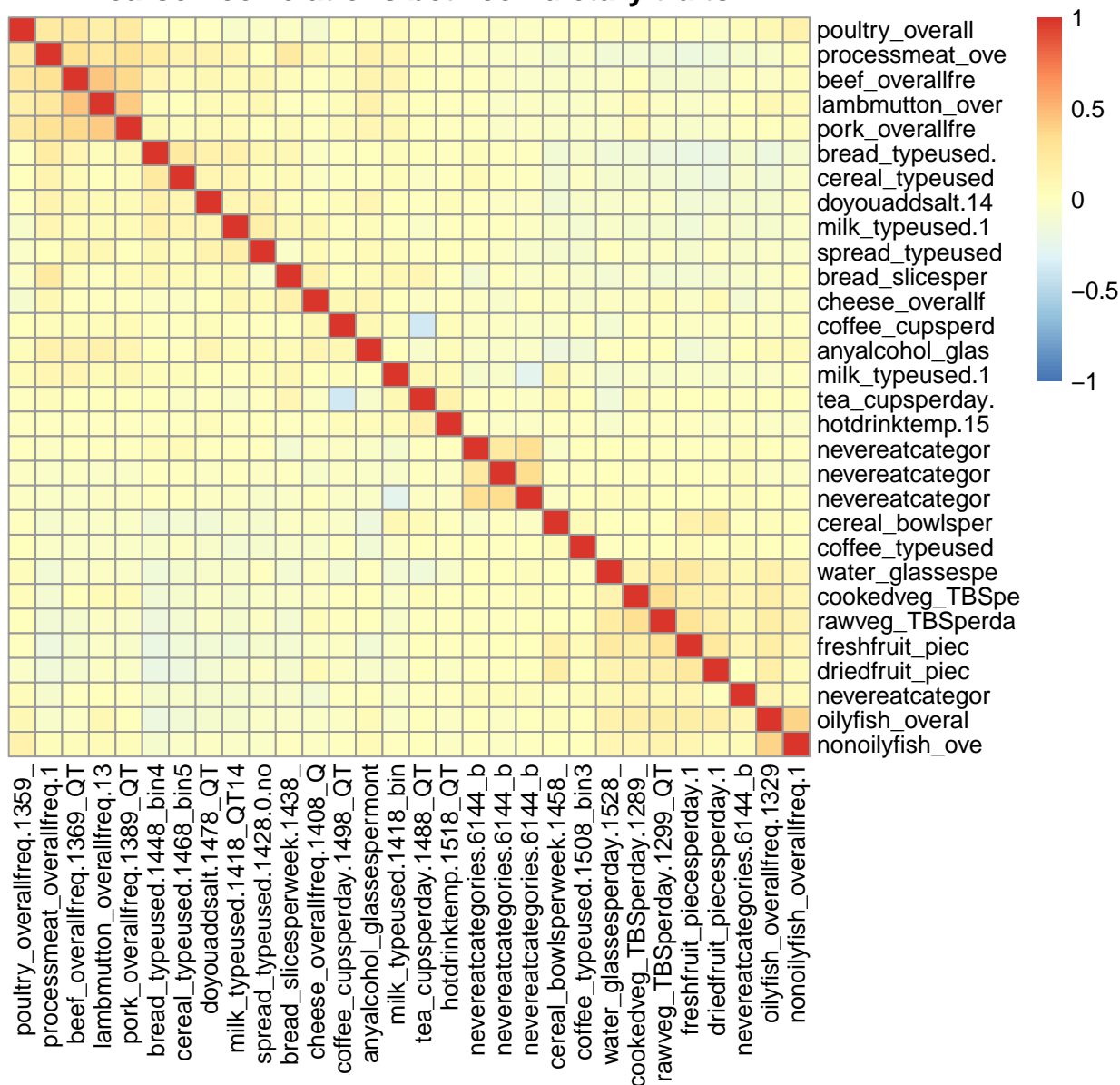


Dietary trait QC for UKB gene-diet interaction analysis

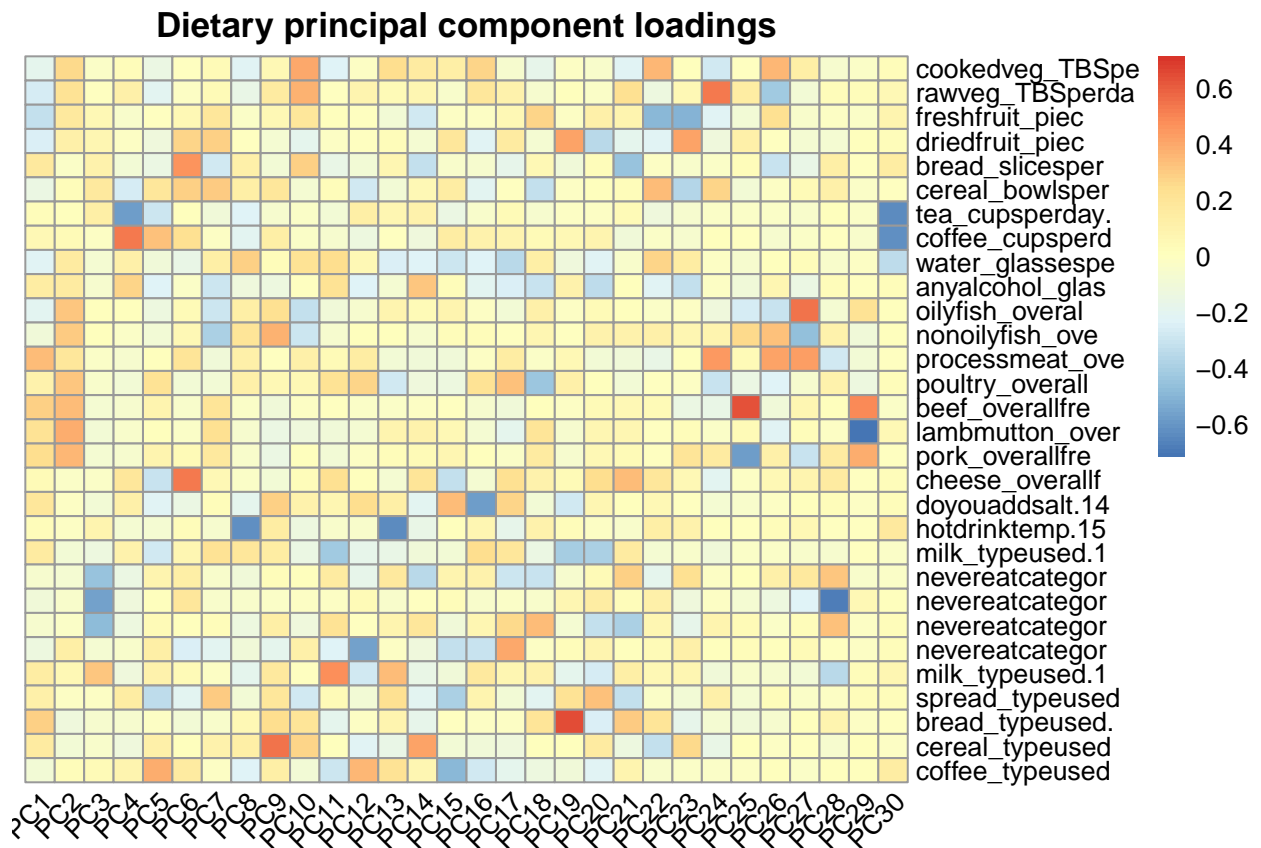
Input dietary variables

- These correlations are based on the final PCA input dataset (i.e. after INT and median imputation)

Pearson correlations between dietary traits



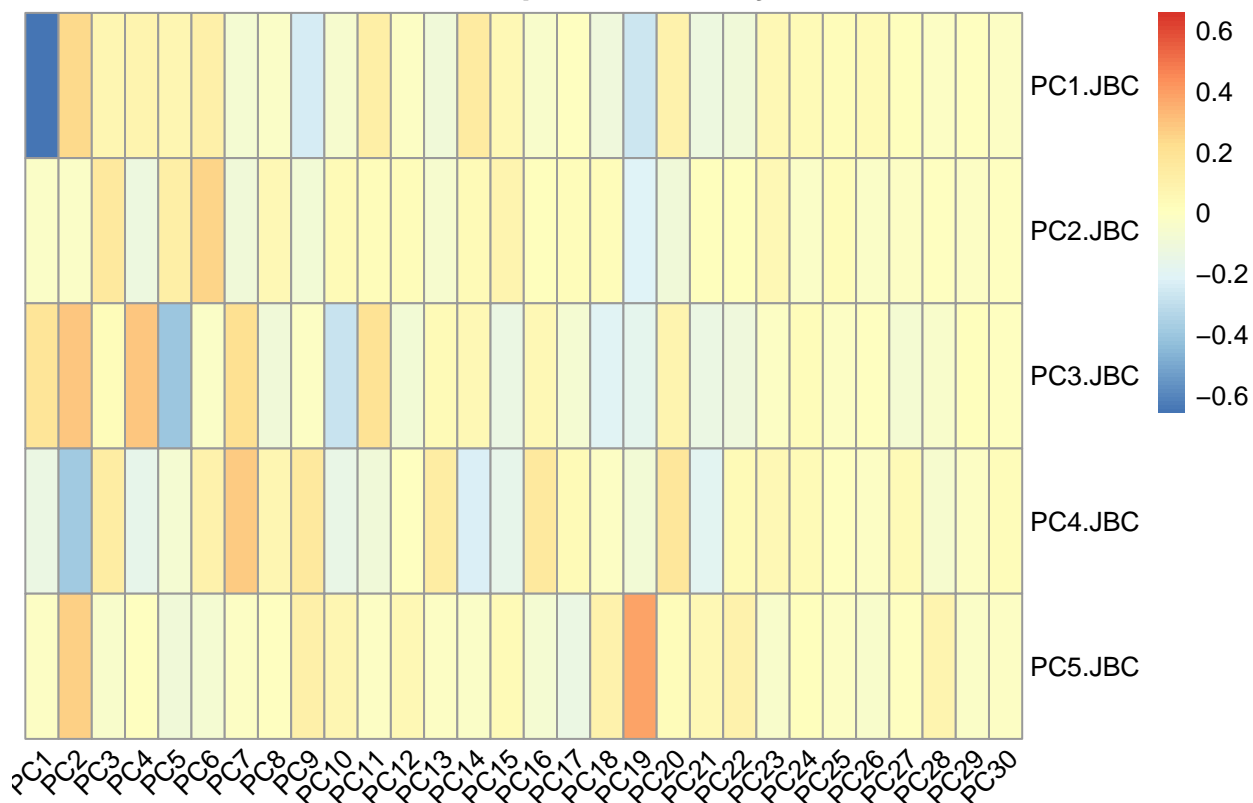
Dietary PCs



Comparison of KW and JC PCs

- Differences may have arisen due to any of the following changes:
 - Sample inclusion:
 - * Only unrelated Europeans
 - * Inclusion of disease & pregnancy
 - Baseline FFQ responses only
 - Alcohol imputation – JC imputed values in $\{-1, -3\}$ as the median of the full set of non-missing values, versus KW as the median of the values not in $\{-1, -3\}$
 - No residualizing on age + sex
 - Trimming to a much smaller set of binary variables to reduce heavy collinearity

Correlations with top 5 JBC dietary PCs



Rough QC of KW preprocessing based on missingness

- Uses the “raw” dietary dataset (i.e. missingness calculated prior to imputation)

Table 1: Fractions non-missing in KW vs. JC diet traits

trait	frac_kw	frac_jc
cookedveg_TBSpersperday.1289_QT	0.98	0.99
rawveg_TBSpersperday.1299_QT	0.98	0.99
freshfruit_piecesperday.1309_QT	0.99	1.00
driedfruit_piecesperday.1319_QT	0.99	0.99
bread_slicesperweek.1438_QT	0.98	0.99
cereal_bowlsperweek.1458_QT	0.99	1.00
tea_cupsperday.1488_QT	1.00	1.00
coffee_cupsperday.1498_QT	1.00	1.00
water_glassesperday.1528_QT	0.99	0.99
anyalcohol_glassespermonth.derived_QT	1.00	1.00
oilyfish_overallfreq.1329_QT	0.99	0.99
nonoilyfish_overallfreq.1339_QT	0.99	1.00
processmeat_overallfreq.1349_QT	1.00	1.00
poultry_overallfreq.1359_QT	1.00	1.00
beef_overallfreq.1369_QT	0.99	1.00
lambmutton_overallfreq.1379_QT	0.99	0.99
pork_overallfreq.1389_QT	0.99	0.99
cheese_overallfreq.1408_QT	0.97	0.98
doyouaddsalt.1478_QT	1.00	1.00
hotdrinktemp.1518_QT	1.00	1.00
milk_typeused.1418_QT14	0.91	NA
nevereatcategories.6144_bin2	0.99	1.00
nevereatcategories.6144_bin4	0.99	1.00
nevereatcategories.6144_bin6	0.99	1.00
nevereatcategories.6144_bin8	0.99	1.00
milk_typeused.1418_bin	0.98	0.95
spread_typeused.1428.0.nonbutterspread_typeused.2654.combined_bin7	0.99	1.00
bread_typeused.1448_bin4	0.91	0.93
cereal_typeused.1468_bin5	0.82	0.83
coffee_typeused.1508_bin3	0.76	0.79

Histograms for each trait (prior to median imputation & INT)

