**Step 1: Run Qualitative GWAS.R with required input files**

**Input:**

“mGWASqual*.*txt*”*, this is a tab delimited file in which the first column contains LC-MS features (“compound\_ID”), followed by the columns containing the normalized abundance recorded for each ecotype (represented by the ecotype number)

“ecotypesALL*.*txt”, this is a tab delimited file containing the ecotypes (represented by numbers following the mGWASqual.txt column order)

“snps**\_**sel**.**RData”, RData file in which the columns contain SNP data per ecotype (rows), 0 = SNP is absent, 1 = SNP is present

“kinship**\_**sel**.**Rdata”, RData file with pairwise genetic relatedness matrix (as for mGWAS – see Methods)

**Output:**

**“**fulldataset.txt”: tab delimited file with columns: trt\_cha (LC-MS features), snp\_cha (SNP represented by chromosome number and location in bp in format chromosome\_bp: e.g 4\_1266038), prob\_num (p-value of the association of the corresponding LC-MS feature and SNP as calculated by Fisher’s exact tests), odds\_cha (the odds calculated by Fisher’s exact tests), trtXkinAVE\_num (calculated average kinship), trtXkinSD\_num (standard deviation of the calculated kinship)

**See requirements.txt file for versions of used modules for all subsequent python scripts**

**Step 2: Run ‘qualtraits\_conttable2.py’**

**Input :**

“fulldataset.txt” generated in Step 1

**Output:**

**“**conttable2\_fulldataset.txt” format as for fulldataset.txt with an additional column contingency tables: The contingency tables can be interpreted in the following way: for SNP “4\_1266038” and trait “M172.24\_RT5.89” the contingency table is [[7, 121], [47, 8]]. This corresponds to:

snp/ 0 1

trait

0 7 121

1 47 8

This means in this particular case that when the SNP is absent (= 0) the trait is present (= 1) for 47, and absent (= 0) for 8 ecotypes and when the SNP is present (= 1) the trait is absent ( = 0) for 121 ecotypes and present ( = 1) for 7

**Step 3: Run regions2\_fulldataset.py**

**Input**:

“conttable2\_fulldataset.txt”: generated in Step 2

**Output:**

“regions\_fulldata\_nosnps.txt”, this is a tab delimited file with format 'association number', 'chromosome', 'traitnumber', 'Trait', 'Trait', 'Trait', 'start', 'stop', 'folder', to match the format required for the “analyze\_regions4.pl” script originally created in the mGWAS pipeline

**Step 4: Run analyze\_regions4.pl**

**Input:**

“regions\_fulldata\_nosnps.txt”, save the file with .regions extension

“gene\_ontology\_ext2015.obo"

“ATH\_GO\_GOSLIM2015.txt"

"TAIR10\_GFF3\_genes.gff";

**Output:**

“fulldata\_genes.txt”

**Step 5: Run** **getpvals\_allfiles.py**

**Input:**

“fulldata\_genes.txt”, generated in step 4

“conttable2\_fulldataset.txt”: generated in Step 2

**Output:**

**“**fulldata\_pvals.txt”

**Step 6: Run fulldataset\_analyses.py**

**Input:**

**“**fulldata\_pvals.txt” generated in step 5

**Output:**

“filtered\_fulldataset.txt”

**Step 7: Run ST\_converter.py**

**Input:**

“filtered\_fulldataset.txt” generated in Step 6

**Output:**

“filtered\_fulldataset\_reformat.txt”

**Step 8: Run add\_processing\_and\_triv\_names.py**

**Input:**

“conversion\_nodenames.txt”, contains both XCMS name and processing name formats of all features

“charcomps.txt”, contains information of all characterized compounds (= Supplemental Table 2)

“filtered\_fulldataset\_reformat.txt”, generated in step 7

**Output:**

“filtered\_fulldataset\_reformat\_trivnames.txt”

**Step 9: run GetGOs.py**

**Input:**

“GO\_bulk\_tair.csv”, data obtained from TAIR: contains GOterms for each gene in the QT-GWAS dataset, separated by ';’

“filtered\_fulldataset\_reformat\_trivnames.txt”, generated in step 8

**Output:**

“SD2\_QTGWAS\_GOterms.txt”, Supplemental Dataset 2