

# main

October 18, 2021

## 1 Examine overlaps with published data

```
[1]: import pandas as pd
import numpy as np
```

### 1.1 Public si-eQTL analysis

```
[2]: shen = ["GDAP2", "AIM2", "SLAMF6", "RLF", "ATG4C", "FUT7",
            "TMEM218", "C11orf74", "RAB35", "TMEM5", "HNRNPK",
            "CDCA3", "ERCC5", "GJB6", "SNTB2", "SPNS3",
            "XAF1", "RBBP8", "RUFY4", "CA2", "RAPGEF1"]
print("Shen et al.:")
print(len(shen))

kukurba = ["NOD2", "WDR36", "BSCL2", "MAP7D3", "RHOF1", "DNAH1"]
print("Kukurba et al.:")
print(len(kukurba))

yao = ["NOD2", "HLA-DRB5", "HLA-DRB5", "KIAA0586", "PPP2R5A",
       "TSNAXIP1", "MUT", "GRIK2", "C15orf37", "LIMA1", "IL6ST",
       "HCG8", "BLOC1S3", "NKX3-1", "CXorf23"]
print("Yao et al.:")
print(len(np.unique(yao)))
print("Total of Yao + Kukurba:")
len(set(yao) | set(kukurba))
```

Shen et al.:

21

Kukurba et al.:

6

Yao et al.:

14

Total of Yao + Kukurba:

[2]: 19

## 1.2 Load BrainSeq si-eQTL results

### 1.2.1 Interacting variant-gene pairs

```
[3]: bs0 = pd.read_csv(".././summary_table/_m/
↳BrainSeq_sexGenotypes_4features_3regions.txt.gz", sep='\t')
bs0["ensembl_gene_id"] = bs0.gene_id.str.replace("\\.*", "", regex=True)
biomart = pd.read_csv(".././././residualized_expression/final_plot_eqtls/_h/
↳biomart.csv", index_col=0)
bs = bs0.merge(biomart, on="ensembl_gene_id").drop_duplicates(subset="gene_id")
print(bs.shape)
bs.tail(2)
```

(974, 11)

```
[3]:
```

	Tissue	gene_id	gencodeID	\
15826	Hippocampus	ENSG00000185386.14	ENSG00000185386.14	
15827	Hippocampus	ENSG00000267251.2	ENSG00000267251.2	

  

	variant_id	seqnames	lfsr	Type	ensembl_gene_id	\
15826	chr22:50563003:G:C	chr22	0.049299	Gene	ENSG00000185386	
15827	chr18:80244356:T:C	chr18	0.049375	Gene	ENSG00000267251	

  

	external_gene_name	entrezgene	\
15826	MAPK11	5600.0	
15827	AC139100.1	NaN	

  

	description
15826	mitogen-activated protein kinase 11 [Source:HG...
15827	NaN

```
[4]: bs[(bs['external_gene_name'].isin(shen))].to_csv("siEQTL_Shen_comparison.csv",
↳index=False)
```

```
[5]: bs[(bs['external_gene_name'].isin(kukurba))]
```

```
[5]: Empty DataFrame
Columns: [Tissue, gene_id, gencodeID, variant_id, seqnames, lfsr, Type,
ensembl_gene_id, external_gene_name, entrezgene, description]
Index: []
```

```
[6]: bs[(bs['external_gene_name'].isin(yao))]
```

```
[6]: Empty DataFrame
Columns: [Tissue, gene_id, gencodeID, variant_id, seqnames, lfsr, Type,
ensembl_gene_id, external_gene_name, entrezgene, description]
Index: []
```

```
[7]: bs[(bs['external_gene_name'].isin(shen+kukurba+yao))]
```

```
[7]:      Tissue      gene_id      gencodeID      variant_id \
4670  Caudate  ENSG00000125703.14  ENSG00000125703.14  chr1:63060301:G:A

      seqnames      lfsr  Type  ensembl_gene_id  external_gene_name  entrezgene \
4670      chr1  0.049896  Gene  ENSG00000125703                ATG4C      84938.0

                                description
4670  autophagy related 4C cysteine peptidase [Sourc...
```

## 1.2.2 eigenMT corrected p-values

```
[8]: df0 = pd.read_csv("../.../summary_table/_m/
↳Brainseq_sex_interacting_4features_3regions.eFeatures.txt.gz", sep='\t')
df = df0[(df0["Type"] == "Gene")].drop_duplicates(subset="gene_id").
↳drop("Tissue", axis=1)
df["ensembl_gene_id"] = df.gene_id.str.replace("\\.*", "", regex=True)
dft = df.merge(biomart, on="ensembl_gene_id")
print(dft.shape)
dft.head(2)
```

(5177, 14)

```
[8]:      variant_id      gene_id      gencodeID      slope \
0  chr7:42936690:A:C  ENSG000000002746.14  ENSG000000002746.14  0.436502
1  chr17:48075934:C:T  ENSG000000002919.14  ENSG000000002919.14 -0.536903

      statistic  pval_nominal      BF  eigenMT_BH  TESTS  Type \
0    7.820971    0.000023  0.010399    0.445692    445  Gene
1   -7.959488    0.000127  0.037424    0.549367    294  Gene

      ensembl_gene_id  external_gene_name  entrezgene \
0  ENSG000000002746                HECW1    23072.0
1  ENSG000000002919                SNX11    29916.0

                                description
0  HECT, C2 and WW domain containing E3 ubiquitin...
1  sorting nexin 11 [Source:HGNC Symbol;Acc:HGNC:...
```

```
[9]: ## Not all genes significant via mashr are also significant with eigenMT
len(set(bs.gene_id) & set(dft.gene_id))
```

[9]: 568

```
[10]: dft[(dft['external_gene_name'].isin(shen))]
```

```
[10]:      variant_id      gene_id      gencodeID      slope \
486      chr1:63052156:A:G  ENSG00000125703.14  ENSG00000125703.14  0.210967
1029  chr16:68729086:G:C  ENSG00000168807.16  ENSG00000168807.16  0.378104
```

2399	chr13:20258404:TC:T	ENSG00000121742.16	ENSG00000121742.16	0.593176
2804	chr11:36347842:A:G	ENSG00000166352.15	ENSG00000166352.15	0.518875
3861	chr8:84969351:T:C	ENSG00000104267.9	ENSG00000104267.9	-0.230131
4091	chr17:6627584:T:C	ENSG00000132530.16	ENSG00000132530.16	-0.515449

	statistic	pval_nominal	BF	eigenMT_BH	TESTS	Type	\
486	10.344946	7.923880e-05	0.037955	0.551098	479	Gene	
1029	11.451744	1.090640e-04	0.033483	0.541553	307	Gene	
2399	10.551991	8.335770e-07	0.000542	0.317735	650	Gene	
2804	8.165337	1.671190e-05	0.009492	0.539284	568	Gene	
3861	-11.113888	3.085070e-04	0.040723	0.693507	132	Gene	
4091	-14.299219	5.135140e-05	0.044214	0.698000	861	Gene	

	ensembl_gene_id	external_gene_name	entrezgene	\
486	ENSG00000125703	ATG4C	84938.0	
1029	ENSG00000168807	SNTB2	6645.0	
2399	ENSG00000121742	GJB6	10804.0	
2804	ENSG00000166352	C11orf74	119710.0	
3861	ENSG00000104267	CA2	760.0	
4091	ENSG00000132530	XAF1	54739.0	

	description
486	autophagy related 4C cysteine peptidase [Source...
1029	syntrophin beta 2 [Source:HGNC Symbol;Acc:HGNC...
2399	gap junction protein beta 6 [Source:HGNC Symbo...
2804	chromosome 11 open reading frame 74 [Source:HG...
3861	carbonic anhydrase 2 [Source:HGNC Symbol;Acc:H...
4091	XIAP associated factor 1 [Source:HGNC Symbol;A...

```
[11]: dft[(dft['external_gene_name'].isin(shen)) & (dft["eigenMT_BH"] < 0.25)]
```

```
[11]: Empty DataFrame
Columns: [variant_id, gene_id, gencodeID, slope, statistic, pval_nominal, BF,
eigenMT_BH, TESTS, Type, ensembl_gene_id, external_gene_name, entrezgene,
description]
Index: []
```

```
[12]: dft[(dft['external_gene_name'].isin(kukurba)) & (dft["eigenMT_BH"] < 0.25)]
```

```
[12]: Empty DataFrame
Columns: [variant_id, gene_id, gencodeID, slope, statistic, pval_nominal, BF,
eigenMT_BH, TESTS, Type, ensembl_gene_id, external_gene_name, entrezgene,
description]
Index: []
```

```
[13]: dft[(dft['external_gene_name'].isin(yao)) & (dft["eigenMT_BH"] < 0.25)]
```

```
[13]: Empty DataFrame
      Columns: [variant_id, gene_id, gencodeID, slope, statistic, pval_nominal, BF,
      eigenMT_BH, TESTS, Type, ensembl_gene_id, external_gene_name, entrezgene,
      description]
      Index: []
```

### 1.3 GTEx comparison

```
[14]: gtex = pd.read_csv("/ceph/users/jbenja13/projects/sex_sz_ria/input/gtex_v8/"+\
      "_m/GTEx_Analysis_v8_sbeQTLs/GTEx_Analysis_v8_sbeQTLs.txt",
      sep='\t')
      gtex.iloc[0:2, 0:10]
```

```
[14]:      ensembl_gene_id  hugo_gene_id      gene_type \
0  ENSG00000241860.6  RP11-34P13.13  processed_transcript
1  ENSG00000227232.5      WASH7P  unprocessed_pseudogene

      variant_id      rs_id      Tissue      maf \
0  chr1_14677_G_A_b38  rs201327123  Adipose_Subcutaneous  0.051635
1  chr1_64764_C_T_b38  rs769952832  Adipose_Subcutaneous  0.061102

      pval_nominal_sb  slope_sb  slope_se_sb
0      0.847114  0.055080  0.285537
1      0.316881  0.222928  0.222511
```

```
[15]: gtex.iloc[0:2, 10:14]
```

```
[15]:      numtested  pvals.corrected      qval  pval_nominal_f
0           1      0.847114  1.000000      0.022302
1           1      0.316881  0.981254      0.003978
```

```
[16]: ## qual threshold equal to number of published sb-eQTL
      gtex[(gtex['qval'] < 0.25) & (gtex["Tissue"].str.contains("Brain"))]\
      .loc[:, ["ensembl_gene_id", "hugo_gene_id", "Tissue", "pvals.corrected",
      ↪ 'qval']].head(10)
```

```
[16]:      ensembl_gene_id  hugo_gene_id \
62155  ENSG00000026025.15      VIM
116842  ENSG00000160818.16      GPATCH4
121904  ENSG00000141562.17      NARF
122123  ENSG00000267174.5  CTC-510F12.4

      Tissue  pvals.corrected      qval
62155      Brain_Amygdala      0.000004  0.012836
116842  Brain_Nucleus_accumbens_basal_ganglia  0.000088  0.198445
121904  Brain_Nucleus_accumbens_basal_ganglia  0.000056  0.198445
122123  Brain_Nucleus_accumbens_basal_ganglia  0.000083  0.198445
```

```
[17]: ## qual threshold equal to number of published sb-eQTL
gtex[(gtex['qval'] < 0.25) & (gtex["Tissue"].str.contains("Whole"))]\
      .loc[:, ["ensembl_gene_id", "hugo_gene_id", "Tissue", "pvals.corrected", "\
↳ 'qval'"]].head(10)
```

```
[17]:      ensembl_gene_id hugo_gene_id      Tissue  pvals.corrected \
362961  ENSG00000221571.3  RNU6ATAC35P  Whole_Blood      0.000039
365043  ENSG00000196743.8      GM2A  Whole_Blood      0.000011
367164  ENSG00000148459.15      PDSS1  Whole_Blood      0.000027

      qval
362961  0.139762
365043  0.116825
367164  0.139762
```

```
[18]: gtex_sig = gtex[(gtex['qval'] < 0.25)]
gtex_sig.shape
```

```
[18]: (369, 22)
```

```
[19]: gtex_sig.head(10)
```

```
[19]:      ensembl_gene_id hugo_gene_id      gene_type \
1096  ENSG00000076356.6      PLXNA2      protein_coding
5262  ENSG00000170632.13      ARMC10      protein_coding
5644  ENSG00000120907.17      ADRA1A      protein_coding
6414  ENSG00000136830.11      FAM129B      protein_coding
7220  ENSG00000166787.3      SAA3P  transcribed_unprocessed_pseudogene
8540  ENSG00000183463.5      URAD      protein_coding
9191  ENSG00000282651.2  IGHV5-10-1      IG_V_gene
14611  ENSG00000143933.16      CALM2      protein_coding
15082  ENSG00000144410.4      CPD      protein_coding
17452  ENSG00000211698.2      TRGV4      TR_V_gene

      variant_id      rs_id      Tissue \
1096  chr1_208030492_G_A_b38  rs3811383  Adipose_Subcutaneous
5262  chr7_103076937_C_T_b38  rs6958836  Adipose_Subcutaneous
5644  chr8_26839198_G_A_b38  rs117380715  Adipose_Subcutaneous
6414  chr9_127584339_G_A_b38  rs10739693  Adipose_Subcutaneous
7220  chr11_18269355_T_C_b38  rs34068567  Adipose_Subcutaneous
8540  chr13_27990205_T_A_b38  rs7335293  Adipose_Subcutaneous
9191  chr14_106114510_A_G_b38  rs4573838  Adipose_Subcutaneous
14611  chr2_46225349_C_T_b38  rs12477148  Adipose_Visceral_Omentum
15082  chr2_206822186_C_T_b38  rs12470278  Adipose_Visceral_Omentum
17452  chr7_38361995_A_C_b38  rs10233345  Adipose_Visceral_Omentum
```

```
      maf  pval_nominal_sb  slope_sb  slope_se_sb  ...      qval \
```

1096	0.123924	5.391600e-05	0.338278	0.083064	...	0.121068
5262	0.169535	5.011130e-05	0.357403	0.087384	...	0.192900
5644	0.216867	1.045890e-05	-0.323552	0.072676	...	0.084548
6414	0.304647	7.387010e-07	-0.283660	0.056579	...	0.004976
7220	0.278830	2.207290e-05	0.323030	0.075427	...	0.074347
8540	0.500000	9.078700e-09	-0.444892	0.076123	...	0.000122
9191	0.419105	2.025150e-05	-0.406760	0.094541	...	0.074347
14611	0.072495	4.497930e-05	-0.480557	0.116471	...	0.161955
15082	0.097015	3.204120e-05	0.682291	0.162191	...	0.115370
17452	0.335821	6.438100e-05	0.427491	0.105837	...	0.139089

	pval_nominal_f	slope_f	slope_se_f	pval_nominal_m	slope_m	\
1096	1.718880e-08	0.456729	0.075705	9.155700e-01	0.009739	
5262	4.933240e-01	-0.054539	0.079379	3.219220e-07	-0.429800	
5644	4.637410e-18	-0.779707	0.076596	3.976660e-10	-0.469672	
6414	1.978000e-06	-0.333315	0.066772	1.653380e-01	-0.082625	
7220	6.409400e-08	0.453034	0.078725	3.138000e-01	0.063002	
8540	9.982650e-21	-0.887723	0.078738	1.892290e-09	-0.457733	
9191	5.248710e-12	-0.682629	0.089412	2.805060e-03	-0.289091	
14611	4.747150e-04	-0.491287	0.134732	9.165740e-01	0.013043	
15082	1.165430e-01	0.280837	0.176978	4.431060e-06	-0.558002	
17452	7.011230e-06	-0.481758	0.100091	8.857670e-15	-1.068840	

	slope_se_m	pval_nominal	slope	slope_se
1096	0.091682	2.747400e-05	0.171830	0.040604
5262	0.079545	8.797530e-08	-0.216374	0.039857
5644	0.069091	5.637370e-52	-0.568916	0.033334
6414	0.059205	1.393160e-08	-0.168762	0.029260
7220	0.062292	2.433600e-08	0.211910	0.037395
8540	0.070571	3.077310e-53	-0.640604	0.036976
9191	0.094806	3.458420e-21	-0.445408	0.045073
14611	0.124116	2.197750e-05	-0.246023	0.057281
15082	0.113158	7.896400e-06	-0.320288	0.070745
17452	0.112111	1.630390e-49	-0.838766	0.049090

[10 rows x 22 columns]

### 1.3.1 mashr

```
[20]: gtex_overlap = bs[(bs['gene_id'].isin(gtex_sig.ensembl_gene_id))].
      ↪drop_duplicates()
      print(gtex_overlap.shape)
      gtex_overlap
```

(14, 11)

[20]:

	Tissue	gene_id	gencodeID	variant_id \
2638	Caudate	ENSG00000105695.14	ENSG00000105695.14	chr19:35065453:C:T
3163	Caudate	ENSG00000109775.10	ENSG00000109775.10	chr4:185462033:A:G
4518	Caudate	ENSG00000124406.16	ENSG00000124406.16	chr4:42387723:C:A
5118	Caudate	ENSG00000132199.18	ENSG00000132199.18	chr18:381997:C:G
5635	Caudate	ENSG00000136830.11	ENSG00000136830.11	chr9:127456523:G:A
6605	Caudate	ENSG00000143933.16	ENSG00000143933.16	chr2:47607174:T:C
6687	Caudate	ENSG00000146530.11	ENSG00000146530.11	chr7:12170840:G:A
10405	Caudate	ENSG00000181264.8	ENSG00000181264.8	chr11:120171373:G:A
11623	Caudate	ENSG00000189067.12	ENSG00000189067.12	chr16:11300271:A:G
12455	Caudate	ENSG00000204248.10	ENSG00000204248.10	chr6:33170900:G:A
14958	Caudate	ENSG00000235098.8	ENSG00000235098.8	chr1:1425750:T:C
15443	Caudate	ENSG00000267056.2	ENSG00000267056.2	chr19:16385981:G:A
15478	Caudate	ENSG00000268362.5	ENSG00000268362.5	chr19:23910181:A:G
15565	Caudate	ENSG00000272977.1	ENSG00000272977.1	chr22:25901440:A:C

	seqnames	lfsr	Type	ensembl_gene_id	external_gene_name \
2638	chr19	0.044690	Gene	ENSG00000105695	MAG
3163	chr4	0.034580	Gene	ENSG00000109775	UFSP2
4518	chr4	0.013706	Gene	ENSG00000124406	ATP8A1
5118	chr18	0.044197	Gene	ENSG00000132199	ENOSF1
5635	chr9	0.044628	Gene	ENSG00000136830	FAM129B
6605	chr2	0.022399	Gene	ENSG00000143933	CALM2
6687	chr7	0.023059	Gene	ENSG00000146530	VWDE
10405	chr11	0.002548	Gene	ENSG00000181264	TMEM136
11623	chr16	0.007562	Gene	ENSG00000189067	LITAF
12455	chr6	0.045996	Gene	ENSG00000204248	COL11A2
14958	chr1	0.006797	Gene	ENSG00000235098	ANKRD65
15443	chr19	0.040125	Gene	ENSG00000267056	AC005336.1
15478	chr19	0.042637	Gene	ENSG00000268362	AC092279.1
15565	chr22	0.004803	Gene	ENSG00000272977	AL008721.2

	entrezgene	description
2638	4099.0	myelin associated glycoprotein [Source:HGNC Sy...
3163	55325.0	UFM1 specific peptidase 2 [Source:HGNC Symbol;...
4518	10396.0	ATPase phospholipid transporting 8A1 [Source:H...
5118	55556.0	enolase superfamily member 1 [Source:HGNC Symb...
5635	64855.0	family with sequence similarity 129 member B [...
6605	805.0	calmodulin 2 [Source:HGNC Symbol;Acc:HGNC:1445]
6687	221806.0	von Willebrand factor D and EGF domains [Sourc...
10405	219902.0	transmembrane protein 136 [Source:HGNC Symbol;...
11623	9516.0	lipopolysaccharide induced TNF factor [Source:...
12455	1302.0	collagen type XI alpha 2 chain [Source:HGNC Sy...
14958	441869.0	ankyrin repeat domain 65 [Source:HGNC Symbol;A...
15443	NaN	NaN
15478	NaN	NaN
15565	NaN	NaN



```
[21]: gtex_overlap.shape[0]/bs.shape[0] * 100
```

```
[21]: 1.4373716632443532
```

```
[22]: gtex_sig[(gtex_sig['ensembl_gene_id'].isin(bs.gene_id))]
```

```
[22]:
```

	ensembl_gene_id	hugo_gene_id	gene_type	\
6414	ENSG00000136830.11	FAM129B	protein_coding	
14611	ENSG00000143933.16	CALM2	protein_coding	
45230	ENSG00000181264.8	TMEM136	protein_coding	
135656	ENSG00000235098.8	ANKRD65	protein_coding	
137724	ENSG00000124406.16	ATP8A1	protein_coding	
137946	ENSG00000109775.10	UFSP2	protein_coding	
138621	ENSG00000204248.10	COL11A2	protein_coding	
138995	ENSG00000146530.11	VWDE	protein_coding	
142471	ENSG00000189067.12	LITAF	protein_coding	
143690	ENSG00000267056.2	AC005336.4	processed_pseudogene	
143783	ENSG00000268362.5	CTD-2017D11.1	lincRNA	
143823	ENSG00000105695.14	MAG	protein_coding	
263405	ENSG00000132199.18	ENOSF1	protein_coding	
338770	ENSG00000272977.1	CTA-390C10.10	sense_intronic	

  

	variant_id	rs_id	Tissue	\
6414	chr9_127584339_G_A_b38	rs10739693	Adipose_Subcutaneous	
14611	chr2_46225349_C_T_b38	rs12477148	Adipose_Visceral_Omentum	
45230	chr11_120493228_T_C_b38	rs4938809	Artery_Coronary	
135656	chr1_1419214_A_G_b38	rs3766165	Breast_Mammary_Tissue	
137724	chr4_42653682_A_T_b38	rs17448575	Breast_Mammary_Tissue	
137946	chr4_185425919_G_C_b38	rs11132303	Breast_Mammary_Tissue	
138621	chr6_33081200_G_GA_b38	rs113353922	Breast_Mammary_Tissue	
138995	chr7_12312811_T_C_b38	rs4721084	Breast_Mammary_Tissue	
142471	chr16_11551157_A_T_b38	rs11644920	Breast_Mammary_Tissue	
143690	chr19_15934939_G_A_b38	rs12985091	Breast_Mammary_Tissue	
143783	chr19_24059854_C_T_b38	rs17206087	Breast_Mammary_Tissue	
143823	chr19_35309759_G_T_b38	rs10411704	Breast_Mammary_Tissue	
263405	chr18_712037_G_A_b38	rs2612103	Muscle_Skeletal	
338770	chr22_25459662_G_A_b38	rs6004655	Spleen	

  

	maf	pval_nominal_sb	slope_sb	slope_se_sb	...	qval	\
6414	0.304647	7.387010e-07	-0.283660	0.056579	...	0.004976	
14611	0.072495	4.497930e-05	-0.480557	0.116471	...	0.161955	
45230	0.403756	1.832690e-05	0.332900	0.075523	...	0.098269	
135656	0.131313	7.063370e-08	0.499559	0.090555	...	0.000111	
137724	0.354430	2.577750e-04	0.246240	0.066638	...	0.090512	
137946	0.309343	4.674540e-07	-0.348749	0.067806	...	0.000399	
138621	0.118687	6.199050e-04	0.418594	0.121102	...	0.102197	
138995	0.452020	1.741340e-06	-0.301523	0.061908	...	0.000744	

142471	0.339646	1.196880e-04	0.247997	0.063687	...	0.035705
143690	0.478535	7.228630e-05	-0.373013	0.092781	...	0.015797
143783	0.170455	2.481440e-03	-0.260946	0.085573	...	0.155454
143823	0.179293	3.726460e-06	-0.621945	0.132132	...	0.001446
263405	0.443343	2.579330e-05	-0.266917	0.062968	...	0.123976
338770	0.167401	3.787520e-05	0.413128	0.097856	...	0.212883

	pval_nominal_f	slope_f	slope_se_f	pval_nominal_m	slope_m	\
6414	1.978000e-06	-0.333315	0.066772	1.653380e-01	-0.082625	
14611	4.747150e-04	-0.491287	0.134732	9.165740e-01	0.013043	
45230	6.159840e-08	0.485968	0.074741	1.990100e-01	0.113737	
135656	7.568940e-06	-0.439928	0.091937	1.443540e-11	-0.879777	
137724	2.989190e-07	-0.323359	0.057862	1.095640e-13	-0.613169	
137946	9.942900e-10	0.432381	0.062550	7.750080e-18	0.758610	
138621	2.304030e-01	-0.125417	0.103786	2.720980e-04	-0.553034	
138995	5.408710e-12	-0.446355	0.055294	1.501580e-03	-0.237816	
142471	3.307190e-04	-0.171901	0.045841	2.925070e-07	-0.385154	
143690	1.961980e-07	-0.587505	0.103261	1.935340e-03	-0.245596	
143783	1.501750e-17	0.858155	0.078738	8.474450e-17	0.981955	
143823	8.536830e-14	-1.030720	0.114726	1.809540e-04	-0.520457	
263405	1.058820e-29	-0.787918	0.056652	4.149110e-16	-0.508209	
338770	7.931210e-14	-1.129030	0.107327	5.155260e-16	-1.276590	

	slope_se_m	pval_nominal	slope	slope_se
6414	0.059205	1.393160e-08	-0.168762	0.029260
14611	0.124116	2.197750e-05	-0.246023	0.057281
45230	0.087221	8.491350e-08	0.231096	0.041304
135656	0.111990	3.281960e-30	-0.726153	0.057266
137724	0.068668	1.462140e-26	-0.370315	0.031672
137946	0.068668	2.014910e-57	0.706155	0.035862
138621	0.145261	4.482050e-06	-0.293079	0.062803
138995	0.072364	2.765010e-20	-0.319694	0.032375
142471	0.068855	9.495520e-22	-0.323697	0.031398
143690	0.076638	1.176010e-12	-0.331468	0.044801
143783	0.093415	1.509120e-67	0.972900	0.043631
143823	0.132593	8.545960e-22	-0.693095	0.067157
263405	0.056303	5.746020e-65	-0.613437	0.031989
338770	0.104690	4.896640e-51	-1.193670	0.056659

[14 rows x 22 columns]

```
[23]: gtex_sig[(gtex_sig['ensembl_gene_id'].isin(bs.gene_id))]\
      .to_csv("siEQTl_gtex_comparison.csv", index=False)
```

### 1.3.2 eigenMT corrected

```
[24]: gtex_overlap = dft[(dft['gene_id'].isin(gtex_sig.ensembl_gene_id)) & (dft["BF"]_
    ↳ < 0.05)].drop_duplicates()
print(gtex_overlap.shape)
gtex_overlap.sort_values("BF").head()
```

(57, 14)

```
[24]:
```

	variant_id	gene_id	gencodeID	slope \
473	chr4:42392391:T:G	ENSG00000124406.16	ENSG00000124406.16	-0.291917
4271	chr8:11603602:A:G	ENSG00000154328.15	ENSG00000154328.15	0.380078
406	chr1:89159687:G:A	ENSG00000117226.11	ENSG00000117226.11	0.431197
2556	chr8:63463308:T:C	ENSG00000137563.11	ENSG00000137563.11	0.527946
3627	chr22:25059120:A:C	ENSG00000272977.1	ENSG00000272977.1	0.626299

	statistic	pval_nominal	BF	eigenMT_BH	TESTS	Type \
473	-13.197929	0.000004	0.002091	0.362373	502	Gene
4271	14.183350	0.000002	0.002130	0.458802	888	Gene
406	15.848961	0.000009	0.002374	0.362373	266	Gene
2556	10.192544	0.000007	0.002460	0.438136	335	Gene
3627	16.616479	0.000003	0.002482	0.438136	744	Gene

	ensembl_gene_id	external_gene_name	entrezgene \
473	ENSG00000124406	ATP8A1	10396.0
4271	ENSG00000154328	NEIL2	252969.0
406	ENSG00000117226	GBP3	2635.0
2556	ENSG00000137563	GGH	8836.0
3627	ENSG00000272977	AL008721.2	NaN

	description
473	ATPase phospholipid transporting 8A1 [Source:H...
4271	nei like DNA glycosylase 2 [Source:HGNC Symbol...
406	guanylate binding protein 3 [Source:HGNC Symbo...
2556	gamma-glutamyl hydrolase [Source:HGNC Symbol;A...
3627	NaN

```
[25]: gtex_overlap = dft[(dft['gene_id'].isin(gtex_sig.ensembl_gene_id))].
    ↳ drop_duplicates()
print(gtex_overlap.shape)
gtex_overlap.sort_values("BF").head()
```

(57, 14)

```
[25]:
```

	variant_id	gene_id	gencodeID	slope \
473	chr4:42392391:T:G	ENSG00000124406.16	ENSG00000124406.16	-0.291917
4271	chr8:11603602:A:G	ENSG00000154328.15	ENSG00000154328.15	0.380078
406	chr1:89159687:G:A	ENSG00000117226.11	ENSG00000117226.11	0.431197

```

2556 chr8:63463308:T:C ENSG00000137563.11 ENSG00000137563.11 0.527946
3627 chr22:25059120:A:C ENSG00000272977.1 ENSG00000272977.1 0.626299

```

	statistic	pval_nominal	BF	eigenMT_BH	TESTS	Type	\
473	-13.197929	0.000004	0.002091	0.362373	502	Gene	
4271	14.183350	0.000002	0.002130	0.458802	888	Gene	
406	15.848961	0.000009	0.002374	0.362373	266	Gene	
2556	10.192544	0.000007	0.002460	0.438136	335	Gene	
3627	16.616479	0.000003	0.002482	0.438136	744	Gene	

	ensembl_gene_id	external_gene_name	entrezgene	\
473	ENSG00000124406	ATP8A1	10396.0	
4271	ENSG00000154328	NEIL2	252969.0	
406	ENSG00000117226	GBP3	2635.0	
2556	ENSG00000137563	GGH	8836.0	
3627	ENSG00000272977	AL008721.2	NaN	

	description
473	ATPase phospholipid transporting 8A1 [Source:H...
4271	nei like DNA glycosylase 2 [Source:HGNC Symbol...
406	guanylate binding protein 3 [Source:HGNC Symbo...
2556	gamma-glutamyl hydrolase [Source:HGNC Symbol;A...
3627	NaN

```
[26]: gtex_overlap.shape[0]/dft.shape[0] * 100
```

```
[26]: 1.1010237589337455
```

```
[ ]:
```