main

March 17, 2022

1 TWAS feature summary

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

1.1 Prepare data

[2]:
```
def limiting_features(set_dict, f1, f2):
    xx = len(set_dict[f1] & set_dict[f2]) / len(set_dict[f2]) * 100
    print("Comparing %s with %s: %0.2f%%" % (f1, f2, xx))
    print("Features in common: %d" % len(set_dict[f1] & set_dict[f2]))
```

1.1.1 Load PGC3 GWAS

[3]:
```
pgc3_file = '/ceph/projects/v4_phase3_paper/inputs/sz_gwas/'+\
            'pgc3/map_phase3/_m/libd_hg38_pgc2sz_snps.tsv'
pgc3_df = pd.read_csv(pgc3_file, sep='\t', low_memory=False, index_col=0)
```

```
/home/jbenja13/.local/lib/python3.9/site-packages/numpy/lib/arraysetops.py:583:
FutureWarning: elementwise comparison failed; returning scalar instead, but in
the future will perform elementwise comparison
  mask |= (ar1 == a)
```

1.1.2 With MHC

Genes

[4]:
```
genes = pd.read_csv('/ceph/projects/v4_phase3_paper/analysis/twas_ea/'+\
                    'gene_weights/fusion/summary_stats/_m/fusion_associations.
 ↪txt', sep='\t')
annot = pd.read_csv('../../../differential_expression/_m/genes/
 ↪diffExpr_szVctl_full.txt', sep='\t')
genes = annot[['ensemblID']].merge(genes, left_on='ensemblID', right_on='FILE')
genes = genes[['FILE', 'ensemblID', 'ID', 'HSQ', 'BEST.GWAS.ID', 'EQTL.ID',
               'TWAS.Z', 'TWAS.P', 'FDR', 'Bonferroni']]
genes['Type'] = 'Gene'
genes.rename(columns={'FILE': 'Feature'}, inplace=True)
genes.sort_values('TWAS.P').head(2)
```

[4]:
```
              Feature        ensemblID          ID       HSQ  \
254   ENSG00000244731  ENSG00000244731         C4A  0.330293
4439  ENSG00000219891  ENSG00000219891  ZSCAN12P1  0.219714

           BEST.GWAS.ID            EQTL.ID     TWAS.Z        TWAS.P  \
254   chr6:32219860:G:T  chr6:31902549:G:A  10.971295  5.251485e-28
4439  chr6:27837477:A:C  chr6:27883095:G:A  10.689750  1.136772e-26

                FDR    Bonferroni  Type
254   2.668805e-24  2.668805e-24  Gene
4439  2.888538e-23  5.777075e-23  Gene
```

Transcripts

[5]:
```
trans = pd.read_csv('/ceph/projects/v4_phase3_paper/analysis/twas_ea/'+\
                    'transcript_weights/fusion/summary_stats/_m/
 ↪fusion_associations.txt', sep='\t')
annot = pd.read_csv('../../../differential_expression/_m/transcripts/
 ↪diffExpr_szVctl_full.txt', sep='\t')
annot['ensemblID'] = annot.gene_id.str.replace('\\..*', '', regex=True)
annot['FILE'] = annot.transcript_id.str.replace('\\..*', '', regex=True)
trans = annot[['ensemblID', 'FILE']].merge(trans, on='FILE')
trans = trans[['FILE', 'ensemblID', 'ID', 'HSQ', 'BEST.GWAS.ID', 'EQTL.ID',
               'TWAS.Z', 'TWAS.P', 'FDR', 'Bonferroni']]
trans['Type'] = 'Transcript'
trans.rename(columns={'FILE': 'Feature'}, inplace=True)
trans.sort_values('TWAS.P').head(2)
```

[5]:
```
              Feature        ensemblID   ID       HSQ       BEST.GWAS.ID  \
3106  ENST00000496659  ENSG00000244731  C4A  0.208279  chr6:32219860:G:T
291   ENST00000428956  ENSG00000244731  C4A  0.313207  chr6:32219860:G:T

                EQTL.ID     TWAS.Z        TWAS.P           FDR    Bonferroni  \
3106  chr6:31942164:A:G  11.008695  3.469941e-28  3.078185e-24  3.078185e-24
291   chr6:31902549:G:A  10.781964  4.188461e-27  1.857792e-23  3.715584e-23

          Type
3106  Transcript
291   Transcript
```

Exons

[6]:
```
exons = pd.read_csv('/ceph/projects/v4_phase3_paper/analysis/twas_ea/'+\
                    'exon_weights/fusion/summary_stats/_m/fusion_associations.
 ↪txt', sep='\t')
annot = pd.read_csv('../../../differential_expression/_m/exons/
 ↪diffExpr_szVctl_full.txt', sep='\t', index_col=0)
exons = annot[['ensemblID']].merge(exons, left_index=True, right_on='FILE')
exons = exons[['FILE', 'ensemblID', 'ID', 'HSQ', 'BEST.GWAS.ID', 'EQTL.ID',
```

```
                'TWAS.Z', 'TWAS.P', 'FDR', 'Bonferroni']]
exons['Type'] = 'Exon'
exons.rename(columns={'FILE': 'Feature'}, inplace=True)
exons.sort_values('TWAS.P').head(2)
```

[6]:

	Feature	ensemblID	ID	HSQ	BEST.GWAS.ID	EQTL.ID	TWAS.Z	TWAS.P	FDR	Bonferroni	Type
34301	e377846	ENSG00000186470	BTN3A2	0.136590	chr6:26463347:G:T	chr6:26336344:G:A	11.162478	6.223219e-29	2.397931e-24	2.397931e-24	Exon
34737	e384610	ENSG00000244731	C4A	0.387274	chr6:32219860:G:T	chr6:31902549:G:A	11.064041	1.874569e-28	3.611545e-24	7.223090e-24	Exon

Junctions

```
annot = pd.read_csv('jxn_annotation.tsv', sep='\t', index_col=1)
annot["gene_id"] = annot.index
juncs = pd.read_csv('/ceph/projects/v4_phase3_paper/analysis/twas_ea/'+\
                    'junction_weights/fusion/summary_stats/_m/
 ↪fusion_associations.txt', sep='\t')
juncs = pd.merge(annot, juncs, left_on='JxnID', right_on='FILE')
juncs = juncs[['gene_id', 'ensemblID', 'Symbol', 'HSQ', 'BEST.GWAS.ID', 'EQTL.
 ↪ID',
               'TWAS.Z', 'TWAS.P', 'FDR', 'Bonferroni']]
juncs['Type'] = 'Junction'
juncs.rename(columns={'Symbol': 'ID', 'gene_id': 'Feature'}, inplace=True)
juncs.sort_values('TWAS.P').head(2)
```

[7]:

	Feature	ensemblID	ID	HSQ	BEST.GWAS.ID	EQTL.ID	TWAS.Z	TWAS.P	FDR	Bonferroni	Type
10530	chr6:31996601-31996828(+)	ENSG00000244731	C4A	0.375134	chr6:32219860:G:T	chr6:31902549:G:A	10.811784	3.027249e-27	2.030200e-23	3.972356e-23	Junction
10528	chr6:31996112-31996206(+)	ENSG00000244731	C4A	0.410304	chr6:32219860:G:T	chr6:31902549:G:A	10.809773	3.094346e-27	2.030200e-23	4.060401e-23	Junction

1.2 Heritable features

1.2.1 Feature summary

[8]:
```
gg = len(set(genes['Feature']))
tt = len(set(trans['Feature']))
ee = len(set(exons['Feature']))
jj = len(set(juncs['Feature']))

print("===Unique Features===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes['ensemblID']))
tt = len(set(trans['ensemblID']))
ee = len(set(exons['ensemblID']))
jj = len(set(juncs['ensemblID']))

print("===Unique Ensembl Gene===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes['ID']))
tt = len(set(trans['ID']))
ee = len(set(exons['ID']))
jj = len(set(juncs['ID']))

print("===Unique Gene Name===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique Features===
Gene:           5082
Transcript:     8871
Exon:           38532
Junction:       13122

===Unique Ensembl Gene===
Gene:           5082
Transcript:     5975
Exon:           7711
Junction:       4346

===Unique Gene Name===
Gene:           5082
Transcript:     5971
Exon:           8789
Junction:       4342
```

1.2.2 Overlap

[9]:
```
features = {
    'Genes': set(genes['ensemblID']),
    'Transcripts': set(trans['ensemblID']),
    'Exons': set(exons['ensemblID']),
    'Junctions': set(juncs['ensemblID']),
}

limiting_features(features, 'Genes', 'Transcripts')
limiting_features(features, 'Genes', 'Junctions')
limiting_features(features, 'Exons', 'Genes')
print("\n")
limiting_features(features, 'Transcripts', 'Junctions')
limiting_features(features, 'Exons', 'Transcripts')
limiting_features(features, 'Exons', 'Junctions')
```

```
Comparing Genes with Transcripts: 55.65%
Features in common: 3325
Comparing Genes with Junctions: 49.47%
Features in common: 2150
Comparing Exons with Genes: 84.83%
Features in common: 4311


Comparing Transcripts with Junctions: 59.99%
Features in common: 2607
Comparing Exons with Transcripts: 72.95%
Features in common: 4359
Comparing Exons with Junctions: 80.76%
Features in common: 3510
```

[10]:
```
len(features['Genes'] & features['Transcripts'] & features['Exons'] &␣
↪features['Junctions'])
```

[10]: 1714

[11]:
```
len(features['Genes'] | features['Transcripts'] | features['Exons'] |␣
↪features['Junctions'])
```

[11]: 10469

1.2.3 SNPs not in significant PGC2+COLUZK GWAS

[12]:
```
new_genes = pd.merge(genes, pgc3_df, left_on='BEST.GWAS.ID',␣
↪right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_trans = pd.merge(trans, pgc3_df, left_on='BEST.GWAS.ID',␣
↪right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
```

```
new_exons = pd.merge(exons, pgc3_df, left_on='BEST.GWAS.ID',␣
 ↪right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_juncs = pd.merge(juncs, pgc3_df, left_on='BEST.GWAS.ID',␣
 ↪right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])

new_genes = new_genes[(new_genes['P'] > 5e-8)].copy()
new_trans = new_trans[(new_trans['P'] > 5e-8)].copy()
new_exons = new_exons[(new_exons['P'] > 5e-8)].copy()
new_juncs = new_juncs[(new_juncs['P'] > 5e-8)].copy()

gg = len(set(new_genes['BEST.GWAS.ID']))
tt = len(set(new_trans['BEST.GWAS.ID']))
ee = len(set(new_exons['BEST.GWAS.ID']))
jj = len(set(new_juncs['BEST.GWAS.ID']))

print("===Unique novel SNPs===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique novel SNPs===
Gene:           2057
Transcript:     2275
Exon:           2907
Junction:       2241
```

[13]:
```
len(set(new_genes['BEST.GWAS.ID']) | set(new_trans['BEST.GWAS.ID']) |
    set(new_exons['BEST.GWAS.ID']) | set(new_juncs['BEST.GWAS.ID']))
```

[13]: 3518

1.3 TWAS P-value < 0.05

1.3.1 Feature summary

[14]:
```
gg = len(set(genes[(genes['TWAS.P'] <= 0.05)].loc[:, 'Feature']))
tt = len(set(trans[(trans['TWAS.P'] <= 0.05)].loc[:, 'Feature']))
ee = len(set(exons[(exons['TWAS.P'] <= 0.05)].loc[:, 'Feature']))
jj = len(set(juncs[(juncs['TWAS.P'] <= 0.05)].loc[:, 'Feature']))

print("===Unique Features===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']))
tt = len(set(trans[(trans['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']))
ee = len(set(exons[(exons['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']))
jj = len(set(juncs[(juncs['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']))
```

```
print("===Unique Ensembl Gene===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['TWAS.P'] <= 0.05)].loc[:, 'ID']))
tt = len(set(trans[(trans['TWAS.P'] <= 0.05)].loc[:, 'ID']))
ee = len(set(exons[(exons['TWAS.P'] <= 0.05)].loc[:, 'ID']))
jj = len(set(juncs[(juncs['TWAS.P'] <= 0.05)].loc[:, 'ID']))

print("===Unique Gene Names===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique Features===
Gene:           1182
Transcript:     2152
Exon:           9560
Junction:       3241

===Unique Ensembl Gene===
Gene:           1182
Transcript:     1667
Exon:           2685
Junction:       1456

===Unique Gene Names===
Gene:           1182
Transcript:     1665
Exon:           2858
Junction:       1453
```

1.3.2 Overlap

[15]:
```
features = {
    'Genes': set(genes[(genes['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']),
    'Transcripts': set(trans[(trans['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']),
    'Exons': set(exons[(exons['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']),
    'Junctions': set(juncs[(juncs['TWAS.P'] <= 0.05)].loc[:, 'ensemblID']),
}

limiting_features(features, 'Genes', 'Transcripts')
limiting_features(features, 'Genes', 'Junctions')
limiting_features(features, 'Exons', 'Genes')
print("\n")
limiting_features(features, 'Transcripts', 'Junctions')
limiting_features(features, 'Exons', 'Transcripts')
limiting_features(features, 'Exons', 'Junctions')
```

```
Comparing Genes with Transcripts: 37.01%
```

```
Features in common: 617
Comparing Genes with Junctions: 30.15%
Features in common: 439
Comparing Exons with Genes: 76.31%
Features in common: 902


Comparing Transcripts with Junctions: 40.59%
Features in common: 591
Comparing Exons with Transcripts: 59.63%
Features in common: 994
Comparing Exons with Junctions: 68.20%
Features in common: 993
```

[16]:
```
len(features['Genes'] & features['Transcripts'] & features['Exons'] &␣
↪features['Junctions'])
```

[16]: 292

[17]:
```
len(features['Genes'] | features['Transcripts'] | features['Exons'] |␣
↪features['Junctions'])
```

[17]: 3943

1.3.3 SNPs not in significant PGC2+COLUZK GWAS

[18]:
```
new_genes = pd.merge(genes[(genes['TWAS.P'] <= 0.05)], pgc3_df, left_on='BEST.
↪GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_trans = pd.merge(trans[(trans['TWAS.P'] <= 0.05)], pgc3_df, left_on='BEST.
↪GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_exons = pd.merge(exons[(exons['TWAS.P'] <= 0.05)], pgc3_df, left_on='BEST.
↪GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_juncs = pd.merge(juncs[(juncs['TWAS.P'] <= 0.05)], pgc3_df, left_on='BEST.
↪GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])

new_genes = new_genes[(new_genes['P'] > 5e-8)].copy()
new_trans = new_trans[(new_trans['P'] > 5e-8)].copy()
new_exons = new_exons[(new_exons['P'] > 5e-8)].copy()
new_juncs = new_juncs[(new_juncs['P'] > 5e-8)].copy()

gg = len(set(new_genes['BEST.GWAS.ID']))
tt = len(set(new_trans['BEST.GWAS.ID']))
ee = len(set(new_exons['BEST.GWAS.ID']))
```

```
jj = len(set(new_juncs['BEST.GWAS.ID']))

print("===Unique novel SNPs===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique novel SNPs===
Gene:           571
Transcript:     767
Exon:           1194
Junction:       798
```

[19]:
```
len(set(new_genes['BEST.GWAS.ID']) | set(new_trans['BEST.GWAS.ID']) |
    set(new_exons['BEST.GWAS.ID']) | set(new_juncs['BEST.GWAS.ID']))
```

[19]:
```
1570
```

1.4 TWAS FDR < 0.05

1.4.1 Feature summary

[20]:
```
gg = len(set(genes[(genes['FDR'] <= 0.05)].loc[:, 'Feature']))
tt = len(set(trans[(trans['FDR'] <= 0.05)].loc[:, 'Feature']))
ee = len(set(exons[(exons['FDR'] <= 0.05)].loc[:, 'Feature']))
jj = len(set(juncs[(juncs['FDR'] <= 0.05)].loc[:, 'Feature']))

print("===Unique Features===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['FDR'] <= 0.05)].loc[:, 'ensemblID']))
tt = len(set(trans[(trans['FDR'] <= 0.05)].loc[:, 'ensemblID']))
ee = len(set(exons[(exons['FDR'] <= 0.05)].loc[:, 'ensemblID']))
jj = len(set(juncs[(juncs['FDR'] <= 0.05)].loc[:, 'ensemblID']))

print("===Unique Ensembl Gene===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['FDR'] <= 0.05)].loc[:, 'ID']))
tt = len(set(trans[(trans['FDR'] <= 0.05)].loc[:, 'ID']))
ee = len(set(exons[(exons['FDR'] <= 0.05)].loc[:, 'ID']))
jj = len(set(juncs[(juncs['FDR'] <= 0.05)].loc[:, 'ID']))

print("===Unique Gene Name===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique Features===
Gene:           553
Transcript:     1117
Exon:           4779
```

```
Junction:       1558

===Unique Ensembl Gene===
Gene:           553
Transcript:     856
Exon:           1413
Junction:       722

===Unique Gene Name===
Gene:           553
Transcript:     856
Exon:           1477
Junction:       724
```

1.4.2 Overlap

[21]:
```
features = {
    'Genes': set(genes[(genes['FDR'] <= 0.05)].loc[:, 'ensemblID']),
    'Transcripts': set(trans[(trans['FDR'] <= 0.05)].loc[:, 'ensemblID']),
    'Exons': set(exons[(exons['FDR'] <= 0.05)].loc[:, 'ensemblID']),
    'Junctions': set(juncs[(juncs['FDR'] <= 0.05)].loc[:, 'ensemblID']),
}

limiting_features(features, 'Genes', 'Transcripts')
limiting_features(features, 'Genes', 'Junctions')
limiting_features(features, 'Exons', 'Genes')
print("\n")
limiting_features(features, 'Transcripts', 'Junctions')
limiting_features(features, 'Exons', 'Transcripts')
limiting_features(features, 'Exons', 'Junctions')
```

```
Comparing Genes with Transcripts: 33.06%
Features in common: 283
Comparing Genes with Junctions: 28.53%
Features in common: 206
Comparing Exons with Genes: 79.39%
Features in common: 439


Comparing Transcripts with Junctions: 39.75%
Features in common: 287
Comparing Exons with Transcripts: 57.94%
Features in common: 496
Comparing Exons with Junctions: 66.34%
Features in common: 479
```

[22]:
```
len(features['Genes'] & features['Transcripts'] & features['Exons'] &␣
↪features['Junctions'])
```

[22]:
```
125
```

[23]:
```
len(features['Genes'] | features['Transcripts'] | features['Exons'] |␣
↪features['Junctions'])
```

[23]:
```
2042
```

1.4.3 SNPs not in significant PGC2+CLOZUK GWAS

[24]:
```
new_genes = pd.merge(genes[(genes['FDR'] <= 0.05)], pgc3_df, left_on='BEST.GWAS.
↪ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_trans = pd.merge(trans[(trans['FDR'] <= 0.05)], pgc3_df, left_on='BEST.GWAS.
↪ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_exons = pd.merge(exons[(exons['FDR'] <= 0.05)], pgc3_df, left_on='BEST.GWAS.
↪ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_juncs = pd.merge(juncs[(juncs['FDR'] <= 0.05)], pgc3_df, left_on='BEST.GWAS.
↪ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])

new_genes = new_genes[(new_genes['P'] > 5e-8)].copy()
new_trans = new_trans[(new_trans['P'] > 5e-8)].copy()
new_exons = new_exons[(new_exons['P'] > 5e-8)].copy()
new_juncs = new_juncs[(new_juncs['P'] > 5e-8)].copy()

gg = len(set(new_genes['BEST.GWAS.ID']))
tt = len(set(new_trans['BEST.GWAS.ID']))
ee = len(set(new_exons['BEST.GWAS.ID']))
jj = len(set(new_juncs['BEST.GWAS.ID']))

print("===Unique novel SNPs===\nGene:\t\t%d\nTranscript:\t%d\nExon:
↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique novel SNPs===
Gene:           214
Transcript:     361
Exon:           580
Junction:       366
```

[25]:
```
len(set(new_genes['BEST.GWAS.ID']) | set(new_trans['BEST.GWAS.ID']) |
    set(new_exons['BEST.GWAS.ID']) | set(new_juncs['BEST.GWAS.ID']))
```

[25]: 784

1.5 TWAS Bonferroni < 0.05

1.5.1 Feature summary

[26]:
```
gg = len(set(genes[(genes['Bonferroni'] <= 0.05)].loc[:, 'Feature']))
tt = len(set(trans[(trans['Bonferroni'] <= 0.05)].loc[:, 'Feature']))
ee = len(set(exons[(exons['Bonferroni'] <= 0.05)].loc[:, 'Feature']))
jj = len(set(juncs[(juncs['Bonferroni'] <= 0.05)].loc[:, 'Feature']))

print("===Unique Features===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']))
tt = len(set(trans[(trans['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']))
ee = len(set(exons[(exons['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']))
jj = len(set(juncs[(juncs['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']))

print("===Unique Ensembl Gene===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))

gg = len(set(genes[(genes['Bonferroni'] <= 0.05)].loc[:, 'ID']))
tt = len(set(trans[(trans['Bonferroni'] <= 0.05)].loc[:, 'ID']))
ee = len(set(exons[(exons['Bonferroni'] <= 0.05)].loc[:, 'ID']))
jj = len(set(juncs[(juncs['Bonferroni'] <= 0.05)].loc[:, 'ID']))

print("===Unique Gene Name===\nGene:\t\t%d\nTranscript:\t%d\nExon:
 ↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique Features===
Gene:           160
Transcript:     272
Exon:           849
Junction:       402

===Unique Ensembl Gene===
Gene:           160
Transcript:     215
Exon:           272
Junction:       161

===Unique Gene Name===
Gene:           160
Transcript:     215
Exon:           278
Junction:       162
```

1.5.2 Overlap

[27]:
```
features = {
    'Genes': set(genes[(genes['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']),
    'Transcripts': set(trans[(trans['Bonferroni'] <= 0.05)].loc[:,␣
 ↪'ensemblID']),
    'Exons': set(exons[(exons['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']),
    'Junctions': set(juncs[(juncs['Bonferroni'] <= 0.05)].loc[:, 'ensemblID']),
}

limiting_features(features, 'Genes', 'Transcripts')
limiting_features(features, 'Genes', 'Junctions')
limiting_features(features, 'Exons', 'Genes')
print("\n")
limiting_features(features, 'Transcripts', 'Junctions')
limiting_features(features, 'Exons', 'Transcripts')
limiting_features(features, 'Exons', 'Junctions')
```

```
Comparing Genes with Transcripts: 38.60%
Features in common: 83
Comparing Genes with Junctions: 28.57%
Features in common: 46
Comparing Exons with Genes: 69.38%
Features in common: 111


Comparing Transcripts with Junctions: 42.24%
Features in common: 68
Comparing Exons with Transcripts: 51.16%
Features in common: 110
Comparing Exons with Junctions: 61.49%
Features in common: 99
```

[28]:
```
len(features['Genes'] & features['Transcripts'] & features['Exons'] &␣
 ↪features['Junctions'])
```

[28]:
```
32
```

[29]:
```
len(features['Genes'] | features['Transcripts'] | features['Exons'] |␣
 ↪features['Junctions'])
```

[29]:
```
454
```

1.5.3 SNPs not in significant PGC2+CLOZUK GWAS

[30]:
```
new_genes = pd.merge(genes[(genes['Bonferroni'] <= 0.05)], pgc3_df,␣
↪left_on='BEST.GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_trans = pd.merge(trans[(trans['Bonferroni'] <= 0.05)], pgc3_df,␣
↪left_on='BEST.GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_exons = pd.merge(exons[(exons['Bonferroni'] <= 0.05)], pgc3_df,␣
↪left_on='BEST.GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])
new_juncs = pd.merge(juncs[(juncs['Bonferroni'] <= 0.05)], pgc3_df,␣
↪left_on='BEST.GWAS.ID',
                     right_on='our_snp_id', suffixes=['_TWAS', '_PGC2'])

new_genes = new_genes[(new_genes['P'] > 5e-8)].copy()
new_trans = new_trans[(new_trans['P'] > 5e-8)].copy()
new_exons = new_exons[(new_exons['P'] > 5e-8)].copy()
new_juncs = new_juncs[(new_juncs['P'] > 5e-8)].copy()

gg = len(set(new_genes['BEST.GWAS.ID']))
tt = len(set(new_trans['BEST.GWAS.ID']))
ee = len(set(new_exons['BEST.GWAS.ID']))
jj = len(set(new_juncs['BEST.GWAS.ID']))

print("===Unique novel SNPs===\nGene:\t\t%d\nTranscript:\t%d\nExon:
↪\t\t%d\nJunction:\t%d\n" % (gg, tt, ee, jj))
```

```
===Unique novel SNPs===
Gene:           12
Transcript:     24
Exon:           22
Junction:       24
```

[31]:
```
len(set(new_genes['BEST.GWAS.ID']) | set(new_trans['BEST.GWAS.ID']) |
    set(new_exons['BEST.GWAS.ID']) | set(new_juncs['BEST.GWAS.ID']))
```

[31]: 51

1.6 Session Information

[32]:
```
import types
from IPython import sys_info

def imports():
    for name, val in globals().items():
        if isinstance(val, types.ModuleType):
```

```
            yield val.__name__

#exclude all modules not listed by `!pip freeze`
excludes = ['__builtin__', 'types', 'IPython.core.shadowns', 'sys', 'os']
function_modules = []
imported_modules = [module for module in imports() if module not in excludes] +␣
 ↪function_modules
pip_modules = !pip freeze #you could also use `!conda list` with anaconda
```

[33]:
```
print(sys_info())
```

```
{'commit_hash': '3813660de',
 'commit_source': 'installation',
 'default_encoding': 'utf-8',
 'ipython_path': '/usr/lib/python3.9/site-packages/IPython',
 'ipython_version': '7.29.0',
 'os_name': 'posix',
 'platform': 'Linux-5.15.5-arch1-1-x86_64-with-glibc2.33',
 'sys_executable': '/usr/bin/python3',
 'sys_platform': 'linux',
 'sys_version': '3.9.7 (default, Oct 10 2021, 15:13:22) \n[GCC 11.1.0]'}
```