mainr

August 11, 2021

1 TWAS plot

[1]:
```
suppressMessages({library(data.table)
                  library(dplyr)
                  library(ggplot2)
                  })
```

[2]:
```
pgc2_file = '/ceph/projects/v4_phase3_paper/inputs/sz_gwas/pgc2_clozuk/
 ↪map_phase3/_m/libd_hg38_pgc2sz_snps.tsv'
pgc2_df = fread(pgc2_file)
```

[3]:
```
twas_dat = read.delim(paste0('/ceph/projects/v4_phase3_paper/analysis/twas/',
                             'gene_weights/fusion_pgc2/summary_stats/_m/
 ↪fusion_associations.txt'))
twas_dat = merge(twas_dat, pgc2_df, by.x='BEST.GWAS.ID', by.y='our_snp_id',␣
 ↪suffixes=c('_TWAS', '_PGC3'))

twas_dat$group = 'Up'
twas_dat$group[twas_dat['TWAS.Z'] < 0] = "Down"

for(xx in seq_along(twas_dat$ID)){
    if(twas_dat$ID[xx] == ''){
        twas_dat$New_ID[xx] = as.character(twas_dat$FILE[xx])
    } else {
        twas_dat$New_ID[xx] = as.character(twas_dat$ID[xx])
    }
}
```

[4]:
```
nCHR <- length(unique(twas_dat$CHR_TWAS))
twas_dat$BPcum <- NA

s <- 0
nbp <- c()
for (i in sort(unique(twas_dat$CHR_TWAS))){
    nbp[i] <- max(twas_dat[twas_dat$CHR_TWAS == i,]$hg38pos)
    twas_dat[twas_dat$CHR_TWAS == i,"BPcum"] <- twas_dat[twas_dat$CHR_TWAS ==␣
 ↪i,"hg38pos"] + s
```

```
    s <- s + nbp[i]
}
```

[5]:
```
axis.set <- twas_dat %>%
  group_by(CHR_TWAS) %>%
  summarize(center = (max(BPcum) + min(BPcum)) / 2)

#axis.set
```

1.0.1 All TWAS, Caudate

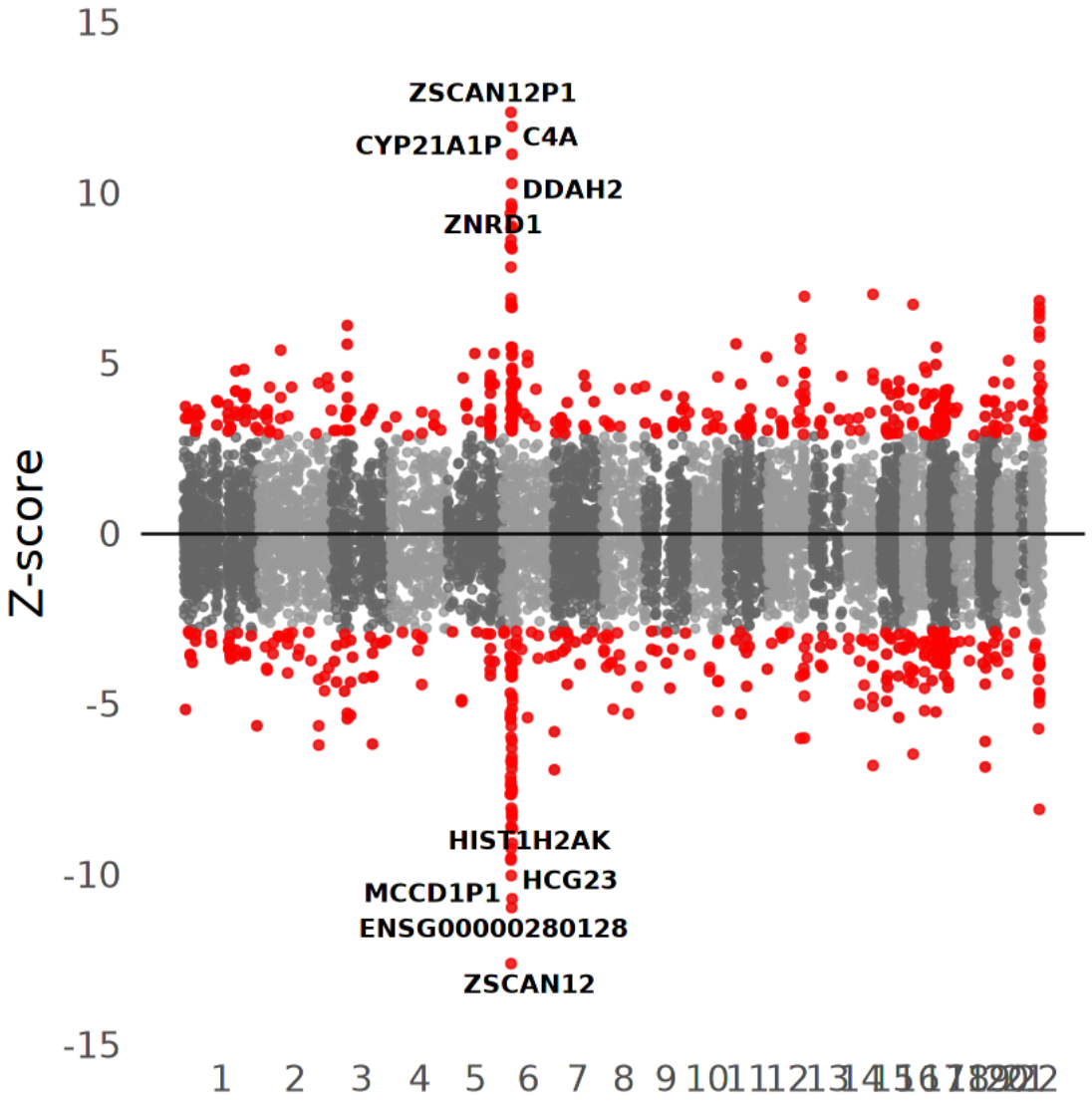
[6]:
```
df = twas_dat[twas_dat$Bonferroni < 0.05, ]
df = head(df[order(df$Bonferroni), ], 10)
df
```

A data.frame: 10 × 49

	BEST.GWAS.ID <chr>	FILE <chr>	ID <chr>	CHR_TWAS <int>	P0 <i
6842	chr6:28744470:A:G	ENSG00000158691	ZSCAN12	6	28
6831	chr6:28426903:C:T	ENSG00000219891	ZSCAN12P1	6	28
6912	chr6:31793436:G:A	ENSG00000244731	C4A	6	31
6913	chr6:31793436:G:A	ENSG00000204338	CYP21A1P	6	32
6876	chr6:30762705:G:T	ENSG00000280128	ENSG00000280128	6	30
6921	chr6:32231204:C:T	ENSG00000228962	HCG23	6	32
6905	chr6:31348749:T:C	ENSG00000213722	DDAH2	6	31
6853	chr6:29445226:C:T	ENSG00000235963	MCCD1P1	6	29
6861	chr6:29591461:T:C	ENSG00000066379	ZNRD1	6	30
6828	chr6:28071151:G:A	ENSG00000275221	HIST1H2AK	6	27

[7]:
```
manhplot1 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.75, size=1.25)␣
↪+
              geom_point(data=twas_dat[twas_dat$FDR < 0.05, ], color="red",␣
↪alpha=0.75) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
↪set$center) +
              scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
↪= NULL, y = "Z-score") +
              geom_hline(yintercept = 0, color = "black", linetype = "solid") +
              ggrepel::geom_text_repel(data=df, fontface = "bold", #hjust=1,␣
↪vjust=1,
                                       #position=position_jitter(width=1,␣
↪height=1),
                                       aes(x=BPcum, y=TWAS.Z, label=New_ID)) +
              theme_minimal(base_size=20) +
              theme(legend.position="none", panel.border=element_blank(),
```

```
                    panel.grid.major=element_blank(), panel.grid.
↪minor=element_blank())
          )
manhplot1
```



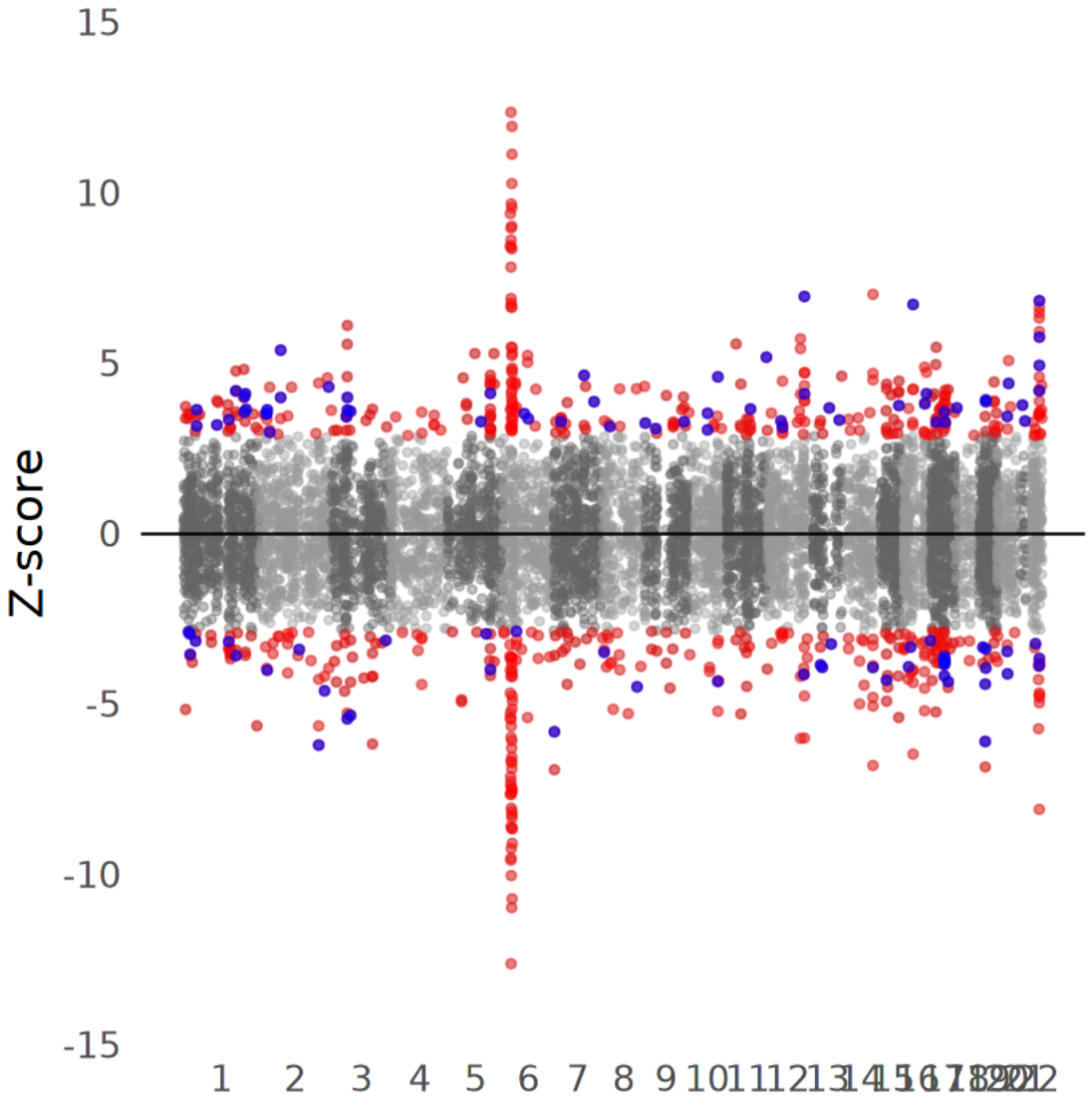
1.0.2 Overlapping tissue

[8]:
```
overlapping = read.delim(paste0('../../_m/overlapping_tissue_twasList.txt'))
overlapping = subset(overlapping, select=c('our_snp_id', 'FILE', 'ID',␣
↪'CHR_TWAS', 'BEST.GWAS.ID', 'P', 'TWAS.Z', 'TWAS.P', 'FDR'))
overlapping = overlapping[order(overlapping$FDR), ]
merge(head(overlapping['FILE'], 10), twas_dat, by='FILE')
```

A data.frame: 10 × 49

FILE <chr>	BEST.GWAS.ID <chr>	ID <chr>	CHR_TWAS <int>	P0 <int>
ENSG00000122687	chr7:1981360:G:A	MRM2	7	223423
ENSG00000129933	chr19:19457850:T:C	MAU2	19	193206
ENSG00000130921	chr12:123148383:G:A	C12orf65	12	123232
ENSG00000163016	chr2:73404084:G:A	ALMS1P1	2	736449
ENSG00000163634	chr3:63881679:A:G	THOC7	3	638338
ENSG00000169592	chr16:29971798:G:T	INO80E	16	299952
ENSG00000198951	chr22:42207808:C:T	NAGA	22	420583
ENSG00000226124	chr2:199850665:T:G	FTCDNL1	2	199760
ENSG00000243696	chr3:53141001:G:A	ENSG00000243696	3	528132
ENSG00000281538	chr22:42207808:C:T	ENSG00000281538	22	421380

[9]:
```
df = merge(overlapping['FILE'], twas_dat, by='FILE')
df = df[order(df$FDR), ]

manhplot2 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.4, size=1.25) +
              geom_point(data=twas_dat[twas_dat$FDR < 0.05, ], color="red",␣
 ↪alpha=0.4) +
              geom_point(data=df, color="blue", alpha=0.6) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
 ↪set$center) +
              scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
 ↪= NULL, y = "Z-score") +
              geom_hline(yintercept = 0, color = "black", linetype = "solid") +
              #geom_text(data=head(df, 10), fontface = "bold", #hjust=1,␣
 ↪vjust=1,
                        #position=position_jitter(width=1, height=1),
              #          aes(x=BPcum, y=TWAS.Z, label=New_ID)) +
              theme_minimal(base_size=20) +
              theme(legend.position="none", panel.border=element_blank(),
                    panel.grid.major=element_blank(),panel.grid.
 ↪minor=element_blank())
            )
manhplot2
```



[10]:
```
save_plots <- function(p, fn){
    for(ext in c('.png', '.pdf', '.svg')){
        ggsave(paste0(fn,ext), p, width=30, height=15, units="cm")
    }
}
```

1.0.3 Save plots

[11]:
```
save_plots(manhplot1, 'caudateOnly_twas_manhattanplot')
save_plots(manhplot2, 'twas_manhattanplot')
```

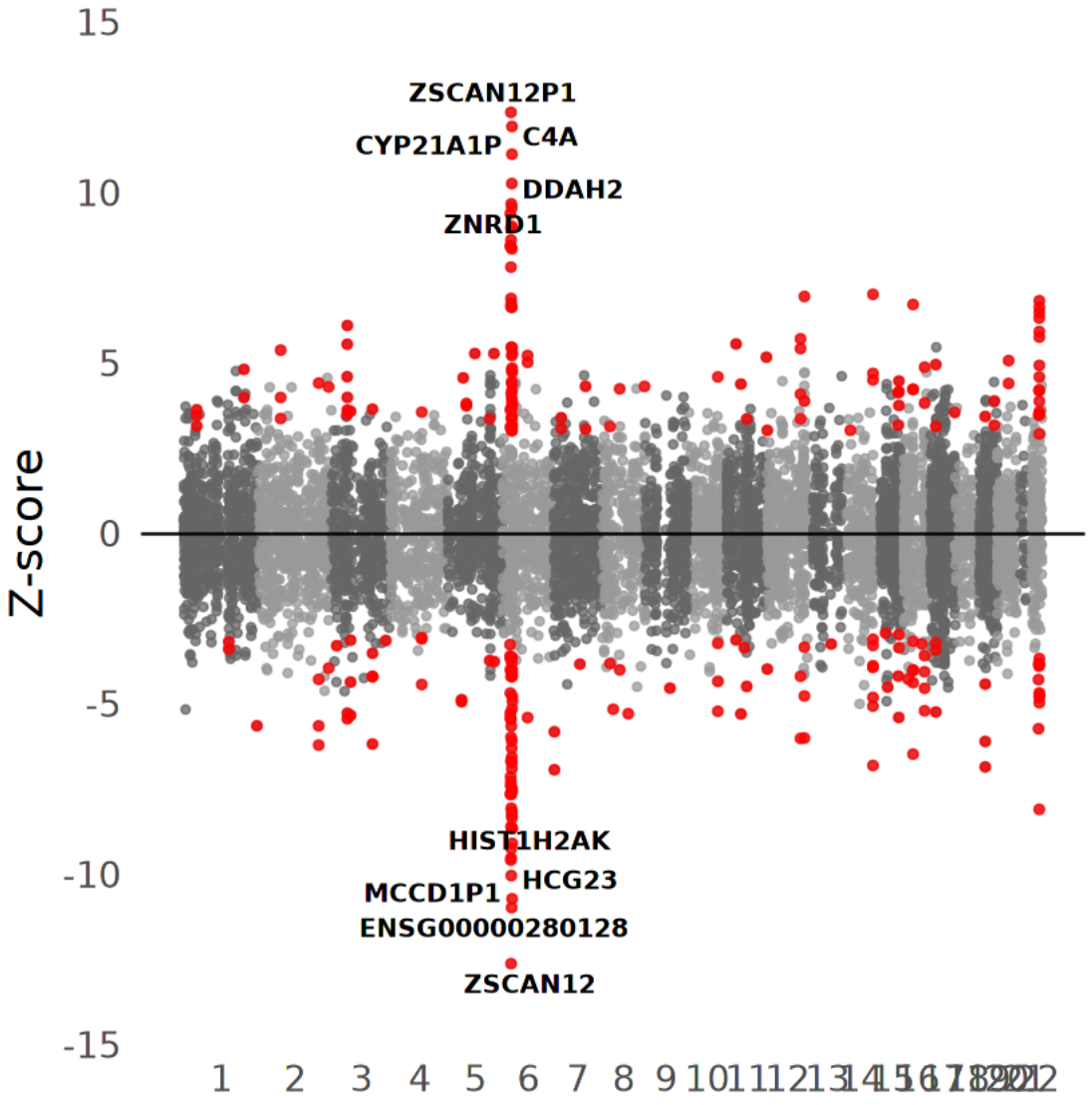
1.0.4 TWAS in GWAS loci, caudate

[12]:
```
df = twas_dat[twas_dat$Bonferroni < 0.05 & twas_dat$P<5e-8, ]
df = head(df[order(df$FDR), ], 10)
df
```

A data.frame: 10 × 49

	BEST.GWAS.ID <chr>	FILE <chr>	ID <chr>	CHR_TWAS <int>	P0 <i
6842	chr6:28744470:A:G	ENSG00000158691	ZSCAN12	6	28
6831	chr6:28426903:C:T	ENSG00000219891	ZSCAN12P1	6	28
6912	chr6:31793436:G:A	ENSG00000244731	C4A	6	31
6913	chr6:31793436:G:A	ENSG00000204338	CYP21A1P	6	32
6876	chr6:30762705:G:T	ENSG00000280128	ENSG00000280128	6	30
6921	chr6:32231204:C:T	ENSG00000228962	HCG23	6	32
6905	chr6:31348749:T:C	ENSG00000213722	DDAH2	6	31
6853	chr6:29445226:C:T	ENSG00000235963	MCCD1P1	6	29
6861	chr6:29591461:T:C	ENSG00000066379	ZNRD1	6	30
6828	chr6:28071151:G:A	ENSG00000275221	HIST1H2AK	6	27

[13]:
```
manhplot1 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.75, size=1.25)␣
 ↪+
              geom_point(data=twas_dat[twas_dat$FDR < 0.05 & twas_dat$P<5e-8,␣
 ↪], color="red", alpha=0.75) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
 ↪set$center) +
              scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
 ↪= NULL, y = "Z-score") +
              geom_hline(yintercept = 0, color = "black", linetype = "solid") +
              ggrepel::geom_text_repel(data=df, fontface = "bold", #hjust=1,␣
 ↪vjust=1,
                                       #position=position_jitter(width=1,␣
 ↪height=1),
                                       aes(x=BPcum, y=TWAS.Z, label=New_ID)) +
              theme_minimal(base_size=20) +
              theme(legend.position="none", panel.border=element_blank(),
                    panel.grid.major=element_blank(),panel.grid.
 ↪minor=element_blank())
            )
manhplot1
```



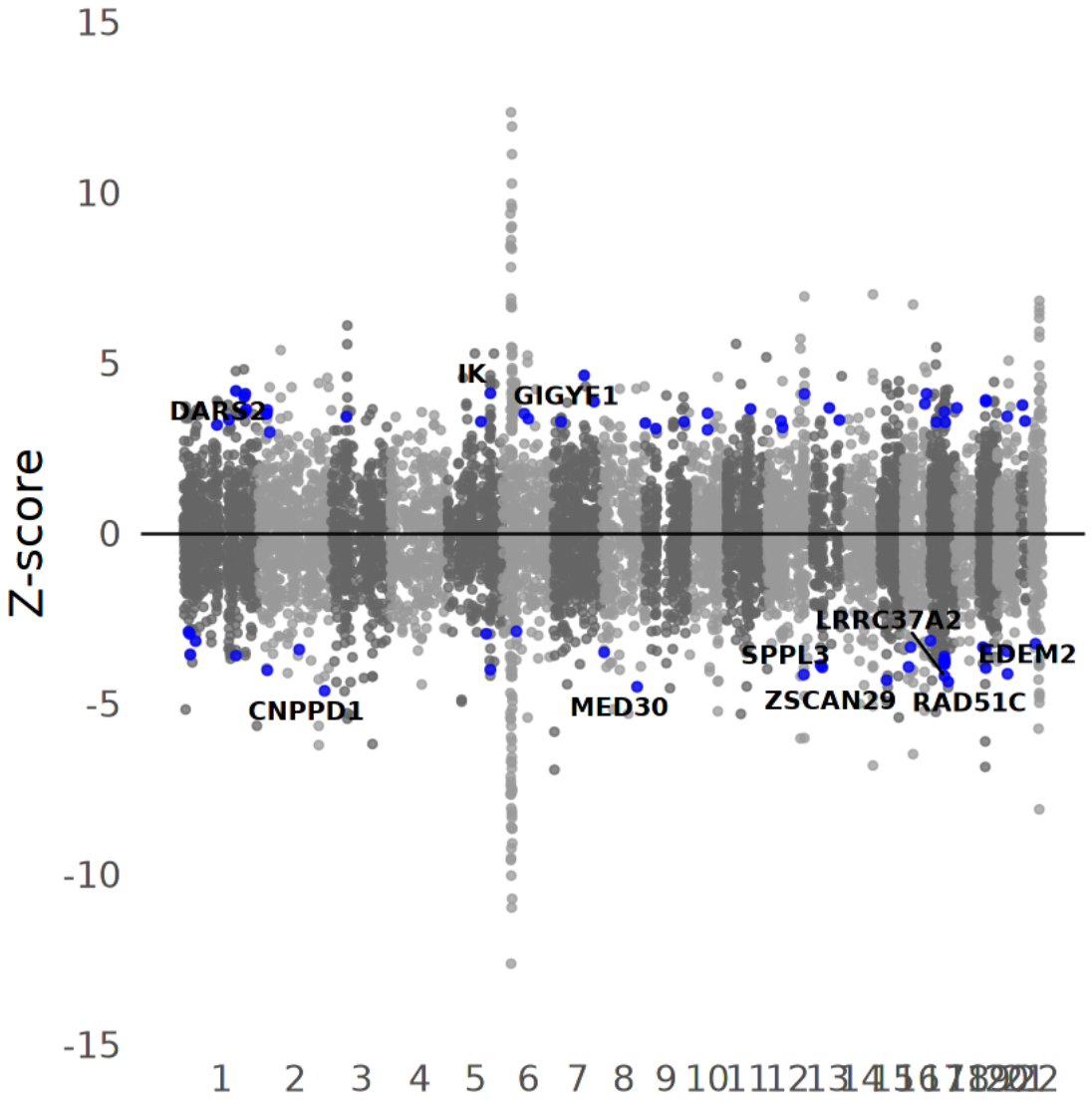
1.0.5 Overlapping tissue in significant GWAS loci

[14]:
```
overlapping = read.delim('../../_m/overlapping_tissue_twasList.txt')
overlapping = subset(overlapping, P>5e-8, select=c('our_snp_id', 'FILE', 'ID',␣
↪'CHR_TWAS',
                                                   'BEST.GWAS.ID', 'P', 'TWAS.
↪Z', 'TWAS.P', 'FDR'))
overlapping = overlapping[order(overlapping$FDR), ]
merge(head(overlapping['FILE'], 10), twas_dat, by='FILE')
```

A data.frame: 10 × 49

FILE <chr>	BEST.GWAS.ID <chr>	ID <chr>	CHR_TWAS <int>	P0 <int>	P1 <int>
ENSG00000088298	chr20:35145275:T:C	EDEM2	20	35115357	35
ENSG00000108384	chr17:58681018:A:G	RAD51C	17	58692573	58
ENSG00000113141	chr5:140841554:G:A	IK	5	140647058	14
ENSG00000115649	chr2:219196879:T:C	CNPPD1	2	219171897	21
ENSG00000117593	chr1:173743105:T:C	DARS2	1	173824503	17
ENSG00000140265	chr15:43782086:A:G	ZSCAN29	15	43358172	43
ENSG00000146830	chr7:100717894:A:G	GIGYF1	7	100679507	10
ENSG00000157837	chr12:121244248:T:C	SPPL3	12	120762510	12
ENSG00000164758	chr8:117545237:G:A	MED30	8	117520713	11
ENSG00000238083	chr17:46288008:C:T	LRRC37A2	17	46511511	46

[15]:
```
df = merge(overlapping['FILE'], twas_dat, by='FILE')
df = df[order(df$FDR), ]

manhplot2 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.75, size=1.25)␣
↪+
              #geom_point(data=twas_dat[twas_dat$FDR < 0.05 & twas_dat$P<5e-8,␣
↪], color="red", alpha=0.75) +
              geom_point(data=df, color="blue", alpha=0.75) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
↪set$center) +
              scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
↪= NULL, y = "Z-score") +
              geom_hline(yintercept = 0, color = "black", linetype = "solid") +
              ggrepel::geom_text_repel(data=head(df, 10), fontface = "bold",␣
↪#hjust=1, vjust=1,
                                       #position=position_jitter(width=1,␣
↪height=1),
                                       aes(x=BPcum, y=TWAS.Z, label=New_ID)) +
              theme_minimal(base_size=20) +
              theme(legend.position="none", panel.border=element_blank(),
                    panel.grid.major=element_blank(),panel.grid.
↪minor=element_blank())
            )
manhplot2
```



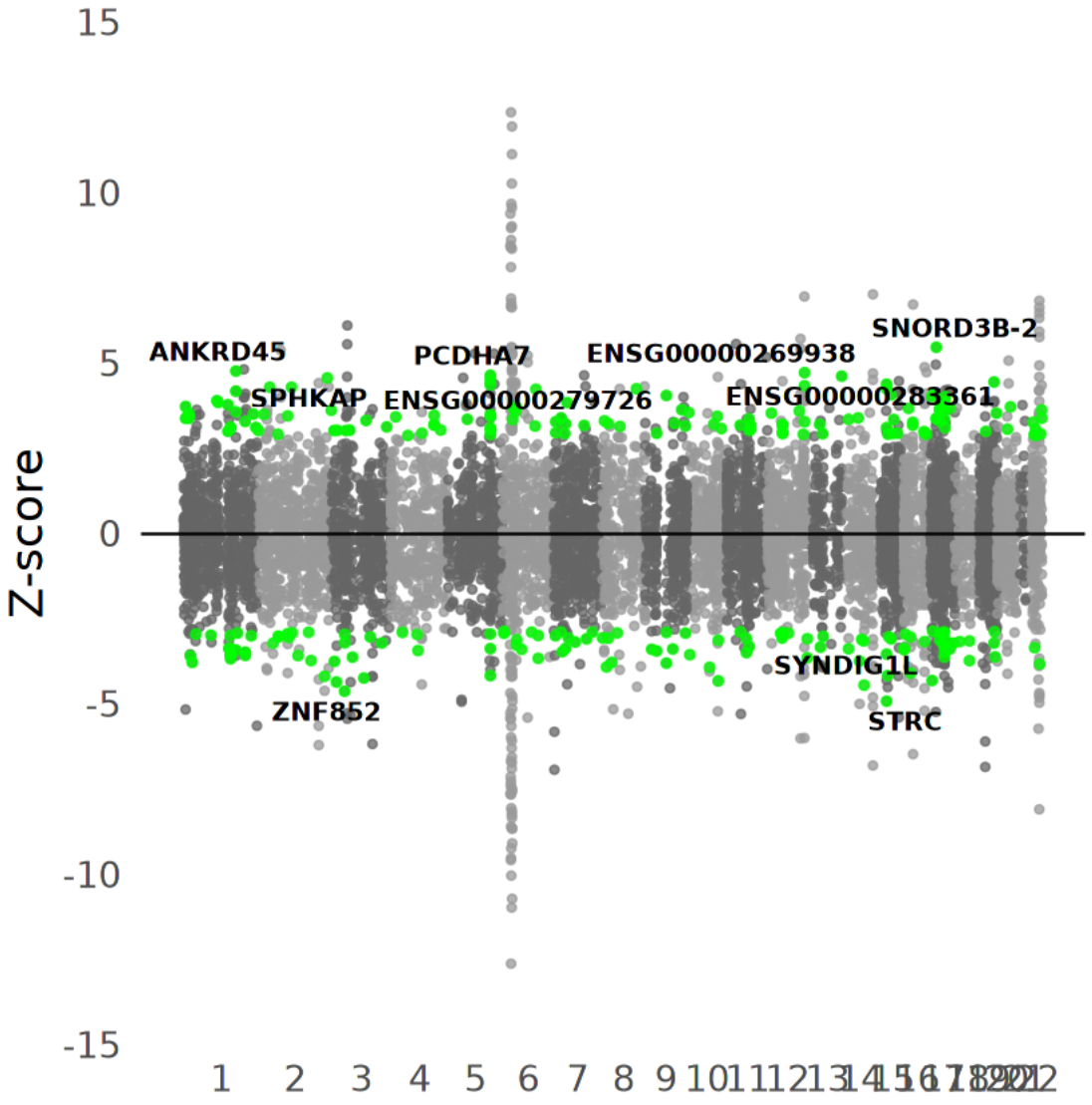
1.0.6 Caudate only in significant GWAS loci

[16]:
```
caudate_only = read.delim('../../_m/caudate_only_twasList_genes.txt')
caudate_only = subset(caudate_only, P>5e-8, select=c('our_snp_id', 'FILE',␣
↪'ID', 'CHR_TWAS',
                                                     'BEST.GWAS.ID', 'P', 'TWAS.
↪Z', 'TWAS.P', 'FDR'))
caudate_only = caudate_only[order(caudate_only$FDR), ]
merge(head(caudate_only['FILE'], 10), twas_dat, by='FILE')
```

A data.frame: 10 × 49

FILE <chr>	BEST.GWAS.ID <chr>	ID <chr>	CHR_TWAS <int>	P0 <int>
ENSG00000153820	chr2:228452776:C:T	SPHKAP	2	227979
ENSG00000178917	chr3:44034110:G:A	ZNF852	3	444989
ENSG00000183379	chr14:74416653:A:C	SYNDIG1L	14	744058
ENSG00000183831	chr1:173743105:T:C	ANKRD45	1	173609
ENSG00000204963	chr5:140841554:G:A	PCDHA7	5	140834
ENSG00000242866	chr15:43782086:A:G	STRC	15	435993
ENSG00000262074	chr17:19250104:T:C	SNORD3B-2	17	190633
ENSG00000269938	chr12:123996254:A:G	ENSG00000269938	12	123968
ENSG00000279726	chr5:140841554:G:A	ENSG00000279726	5	140788
ENSG00000283361	chr13:114134675:T:C	ENSG00000283361	13	114179

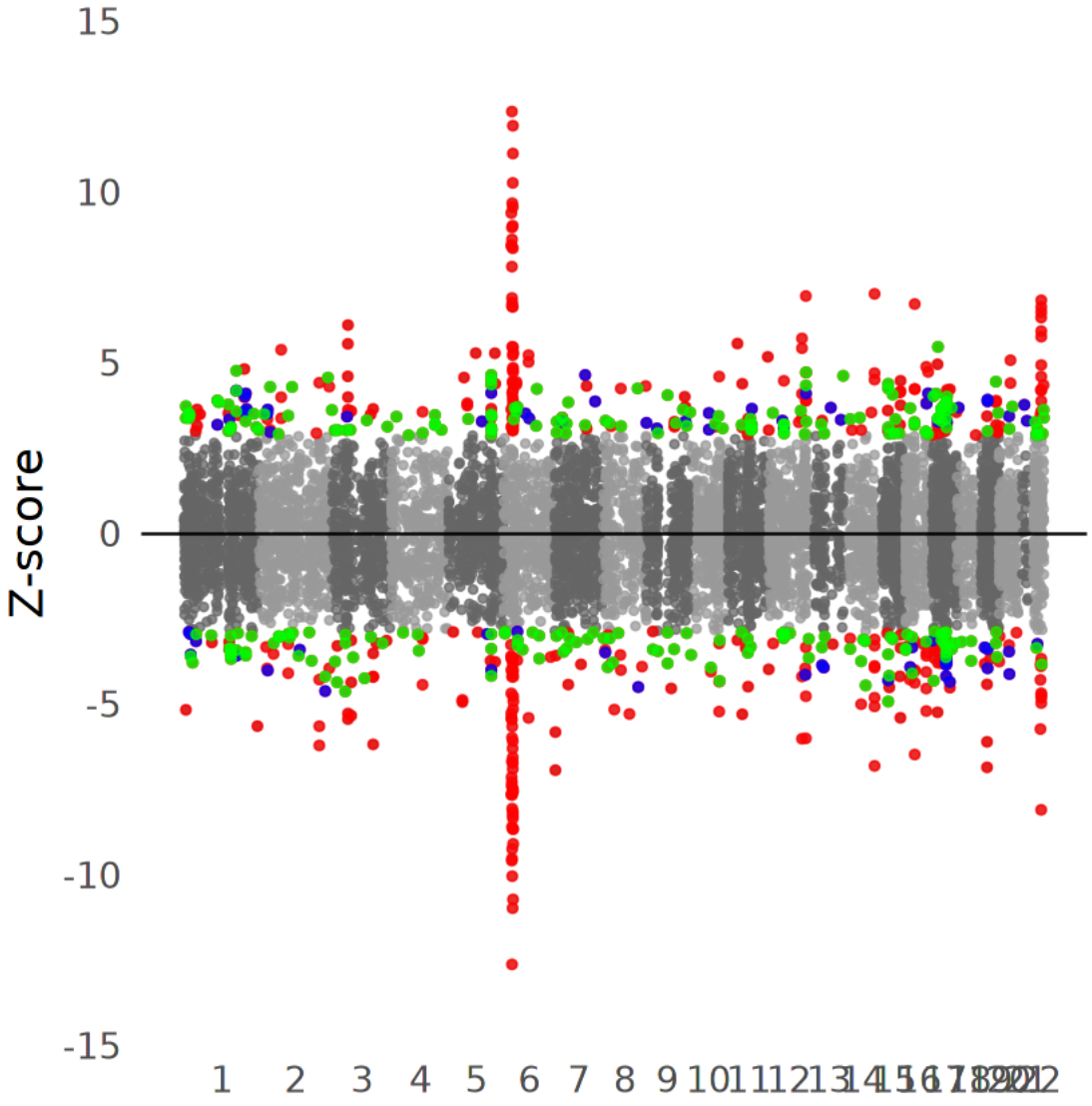
[17]:
```
df2 = merge(caudate_only['FILE'], twas_dat, by='FILE')
df2 = df2[order(df2$FDR), ]

manhplot3 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.75, size=1.25)␣
 ↪+
              #geom_point(data=twas_dat[twas_dat$FDR < 0.05 & twas_dat$P<5e-8,␣
 ↪], alpha=0.75, color="red") +
              #geom_point(data=df, color="blue", alpha=0.75) +
              geom_point(data=df2, color="green", alpha=0.75) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
 ↪set$center) +
              scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
 ↪= NULL, y = "Z-score") +
              geom_hline(yintercept = 0, color = "black", linetype = "solid") +
              ggrepel::geom_text_repel(data=head(df2, 10), fontface = "bold",␣
 ↪#hjust=1, vjust=1,
                                       #position=position_jitter(width=1,␣
 ↪height=1),
                                       aes(x=BPcum, y=TWAS.Z, label=New_ID)) +
              theme_minimal(base_size=20) +
              theme(legend.position="none", panel.border=element_blank(),
                    panel.grid.major=element_blank(),panel.grid.
 ↪minor=element_blank())
            )
manhplot3
```



1.0.7 No labels

[18]:
```
manhplot0 <- (ggplot(twas_dat, aes(x=BPcum, y=TWAS.Z)) +
              geom_point(aes(color=as.factor(CHR_TWAS)), alpha=0.75, size=1.25)␣
↪+
              geom_point(data=twas_dat[twas_dat$FDR < 0.05, ], color="red",␣
↪alpha=0.75) +
              geom_point(data=df, color="blue", alpha=0.75) +
              geom_point(data=df2, color="green", alpha=0.75) +
              scale_color_manual(values = rep(c("grey40","grey60"), nCHR)) +
              scale_x_continuous(label = axis.set$CHR_TWAS, breaks = axis.
↪set$center) +
```

```
            scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) + labs(x␣
↪= NULL, y = "Z-score") +
            geom_hline(yintercept = 0, color = "black", linetype = "solid") +
            #geom_text(data=subset(twas_dat, FDR < 0.05), hjust=-0.2,␣
↪aes(x=BPcum, y=TWAS.Z, label=ID)) +
            theme_minimal(base_size=20) +
            theme(legend.position="none", panel.border=element_blank(),
                  panel.grid.major=element_blank(),panel.grid.
↪minor=element_blank())
          )
manhplot0
```



1.1 Save plots

[19]:
```
save_plots(manhplot0, 'noLabel_twas_manhattanplot')
save_plots(manhplot1, 'topSig_twas_manhattanplot')
save_plots(manhplot2, 'sharedTop_twas_manhattanplot')
save_plots(manhplot3, 'caudateTop_twas_manhattanplot')
```