

# mainr

March 8, 2022

## 1 TWAS plot

```
[1]: suppressMessages({  
      library(dplyr)  
      library(ggtext)  
      library(ggplot2)  
    })
```

### 1.1 Functions

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

### 1.2 Prepare data

#### 1.2.1 Prepare GWAS

```
[3]: pgc2_file = '/ceph/projects/v4_phase3_paper/inputs/sz_gwas/pgc2_clozuk/  
      ↪map_phase3/_m/libd_hg38_pgc2sz_snps.tsv'  
gwas_data_load = data.table::fread(pgc2_file) %>% janitor::clean_names()  
  
sig_data <- gwas_data_load %>% subset(p < 0.05)  
notsig_data <- gwas_data_load %>% subset(p >= 0.05) %>% group_by(chr) %>%  
      ↪sample_frac(0.1)  
gwas_data <- bind_rows(sig_data, notsig_data) %>%  
      select(chr, our_snp_id, rsid, snp, bp, or, se, p)  
gwas_data %>% head(2)
```

	chr	our_snp_id	rsid	snp	bp	or	se
	<int>	<chr>	<chr>	<chr>	<int>	<dbl>	<dbl>
A data.table: 2 × 8	10	chr10:49722:G:C	rs35849539	rs35849539:95662:G:C	95662	0.97809	0.010970
	10	chr10:50529:G:A	rs4468273	rs4468273:96469:G:A	96469	1.03010	0.010712

## 1.2.2 Preparing TWAS data

```
[4]: twas_file = paste0('/ceph/projects/v4_phase3_paper/analysis/twas_ea/
      ↪gene_weights',
      '/fusion_pgc2/summary_stats/_m/fusion_associations.txt')
twas_data_load <- data.table::fread(twas_file) %>% janitor::clean_names() %>%
  select(id, file, hsq, best_gwas_id, eqtl_id, model, twas_z, twas_p, fdr,
      ↪bonferroni) %>%
  inner_join(gwas_data, by=c("best_gwas_id"="our_snp_id"))
twas_data_load %>% head(2)
```

	id	file	hsq	best_gwas_id	eqtl_id
	<chr>	<chr>	<dbl>	<chr>	<chr>
A data.table: 2 × 17	SAMD11	ENSG00000187634	0.323067	chr1:1357963:T:C	chr1:1095185:C:T
	ENSG00000272512	ENSG00000272512	0.541144	chr1:1357963:T:C	chr1:1118005:C:T

```
[5]: data_cum <- twas_data_load %>% group_by(chr) %>% summarise(max_bp = max(bp)) %>%
  mutate(bp_add = lag(cumsum(as.numeric(max_bp)), default=0)) %>% select(chr,
      ↪bp_add)
twas_data <- twas_data_load %>% inner_join(data_cum, by="chr") %>%
  mutate(bp_cum = bp + bp_add, group = ifelse(twas_z < 0, "Down", "Up"),
  id=ifelse(id == "", as.character(file), as.character(id)))
twas_data %>% head(2)
```

	id	file	hsq	best_gwas_id	eqtl_id
	<chr>	<chr>	<dbl>	<chr>	<chr>
A data.table: 2 × 20	SAMD11	ENSG00000187634	0.323067	chr1:1357963:T:C	chr1:1095185:C:T
	ENSG00000272512	ENSG00000272512	0.541144	chr1:1357963:T:C	chr1:1118005:C:T

## 1.2.3 All TWAS, Caudate

### 1.2.4 Plotting settings

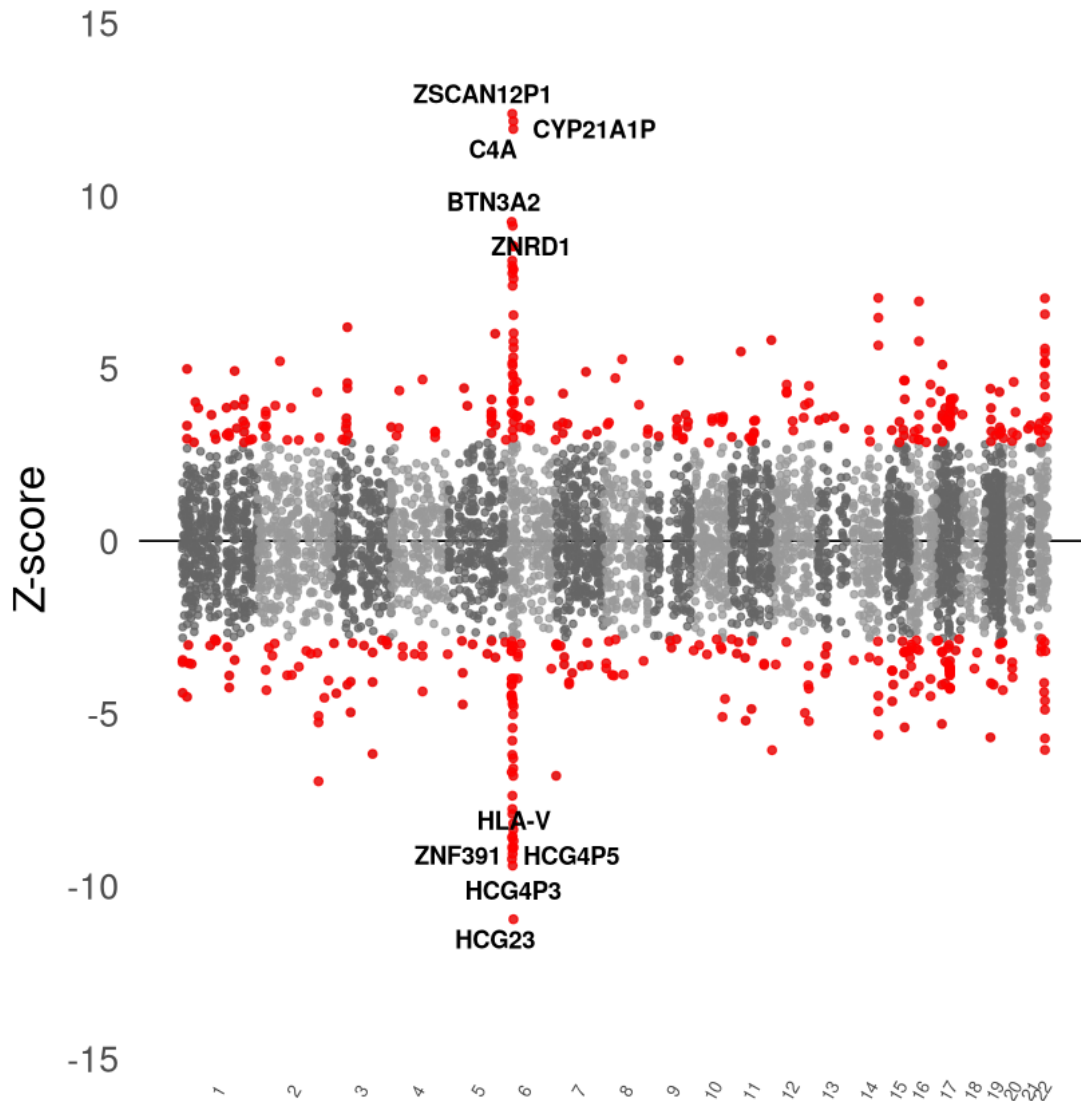
```
[6]: axis_set <- twas_data %>% group_by(chr) %>% summarize(center = mean(bp_cum))
```

```
[7]: df = twas_data[twas_data$bonferroni < 0.05, ]
df = head(df[order(df$bonferroni), ], 10)
df
```

	id <chr>	file <chr>	hsq <dbl>	best_gwas_id <chr>	eqtl_id <chr>
A data.table: 10 × 20	ZSCAN12P1	ENSG00000219891	0.219714	chr6:28426903:C:T	chr6:27883095:G:A
	CYP21A1P	ENSG00000204338	0.318359	chr6:31793436:G:A	chr6:31910656:G:A
	C4A	ENSG00000244731	0.330293	chr6:31793436:G:A	chr6:31902549:G:A
	HCG23	ENSG00000228962	0.464972	chr6:32231204:C:T	chr6:32364487:G:A
	HCG4P3	ENSG00000237669	0.303198	chr6:29516438:G:T	chr6:29852671:G:A
	BTN3A2	ENSG00000186470	0.565614	chr6:26463346:G:T	chr6:26353872:G:T
	ZNF391	ENSG00000124613	0.171150	chr6:27424023:C:T	chr6:27375506:C:A
	ZNRD1	ENSG00000066379	0.213797	chr6:29591461:T:C	chr6:29868964:A:G
	HCG4P5	ENSG00000227766	0.475354	chr6:29445226:C:T	chr6:29850791:T:C
	HLA-V	ENSG00000181126	0.546295	chr6:29315895:C:T	chr6:29785593:C:T

### 1.2.5 Plot

```
[8]: manhplot1 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.75)
  ↪+
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
  scale_color_manual(values = rep(c("grey40", "grey60"),
  ↪unique(length(axis_set$chr)))) +
  labs(x = NULL, y = "Z-score") +
  ggrepel::geom_text_repel(data=df, fontface = "bold", aes(x=bp_cum,
  ↪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(),
    axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
  ↪size=8, vjust=0.5)
  )
manhplot1
```



```
[9]: save_plots(manhplot1, 'caudateOnly_twas_manhattanplot')
```

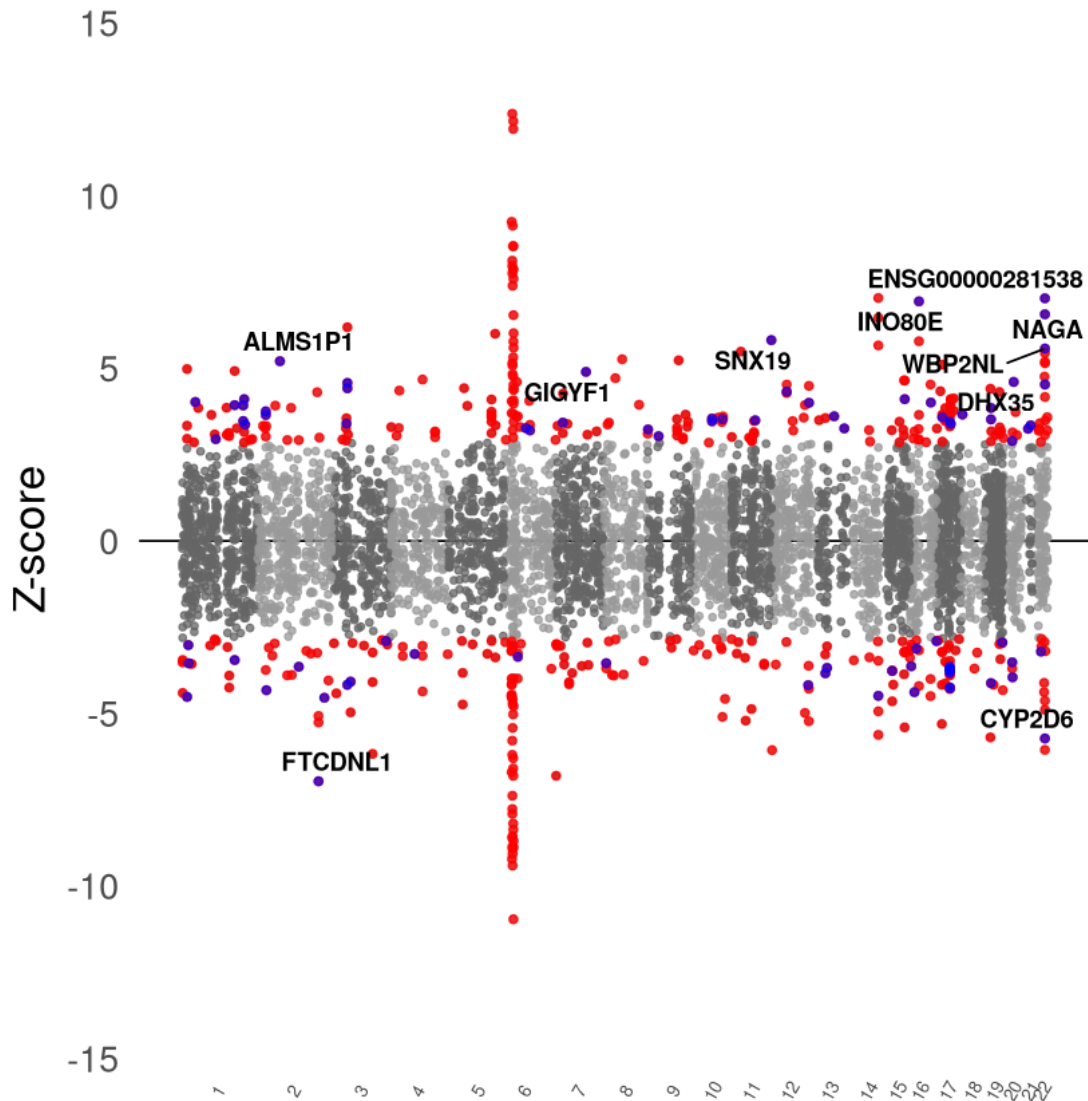
### 1.2.6 Overlapping tissue

```
[10]: overlapping <- data.table::fread("../_m/overlapping_tissue_twasList.txt")
      %>% janitor::clean_names() %>%
      select(id, file, chr_twas, our_snp_id, best_gwas_id, p, twas_z, twas_p,
      %>% fdr) %>%
      arrange(fdr)
      overlapping %>% head(10)
```

	id	file	chr_twas	our_snp_id	best_gwas_id
	<chr>	<chr>	<int>	<chr>	<chr>
A data.table: 10 × 9	ENSG00000281538	ENSG00000281538	22	chr22:42207808:C:T	chr22:42207808
	FTCDNL1	ENSG00000226124	2	chr2:199850665:T:G	chr2:199850665
	INO80E	ENSG00000169592	16	chr16:29971798:G:T	chr16:29971798
	NAGA	ENSG00000198951	22	chr22:42207808:C:T	chr22:42207808
	SNX19	ENSG00000120451	11	chr11:130941461:T:C	chr11:130941461
	CYP2D6	ENSG00000100197	22	chr22:42207808:C:T	chr22:42207808
	WBP2NL	ENSG00000183066	22	chr22:42207808:C:T	chr22:42207808
	ALMS1P1	ENSG00000163016	2	chr2:73404084:G:A	chr2:73404084
	GIGYF1	ENSG00000146830	7	chr7:100717894:A:G	chr7:100717894
	DHX35	ENSG00000101452	20	chr20:38824551:A:G	chr20:38824551

```
[11]: dfx = overlapping %>% select(file) %>% inner_join(twas_data, by='file') %>%
      arrange(fdr)

manhplot2 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.75)
  ↪+
  geom_point(data=dfx, color="blue", alpha=0.6) +
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
  scale_color_manual(values = rep(c("grey40", "grey60"),
  ↪unique(length(axis_set$chr)))) +
  labs(x = NULL, y = "Z-score") +
  ggrepel::geom_text_repel(data=head(dfx, 10), fontface = "bold",
  ↪aes(x=bp_cum, 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(),
    axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
  ↪size=8, vjust=0.5)
  )
manhplot2
```



```
[12]: save_plots(manhplot2, 'twas_manhattanplot')
```

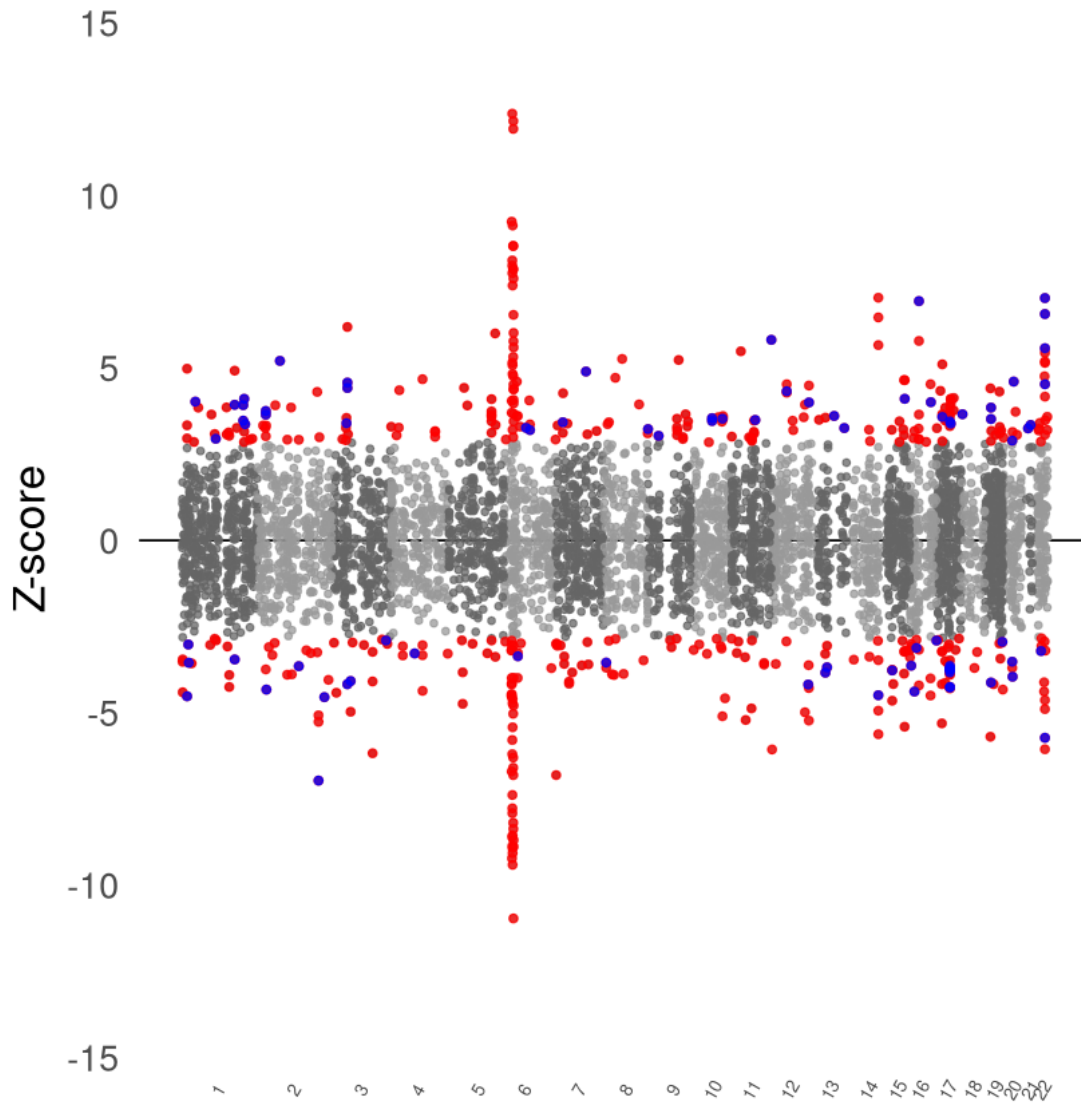
```
[13]: dfx = overlapping %>% select(file) %>% inner_join(twas_data, by='file') %>%
      arrange(fdr)

manhplot2 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.75)
  ↪+
  geom_point(data=dfx, color="blue", alpha=0.75) +
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
```

```

    scale_color_manual(values = rep(c("grey40", "grey60"),
    ↪unique(length(axis_set$chr)))) +
    labs(x = NULL, y = "Z-score") +
    #ggrepel::geom_text_repel(data=head(dfx, 10), fontface = "bold",
    ↪aes(x=bp_cum, 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(),
      axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
    ↪size=8, vjust=0.5)
    )
manhplot2
save_plots(manhplot2, 'twas_manhattanplot_noLabels')

```



## 1.2.7 TWAS in GWAS loci, caudate

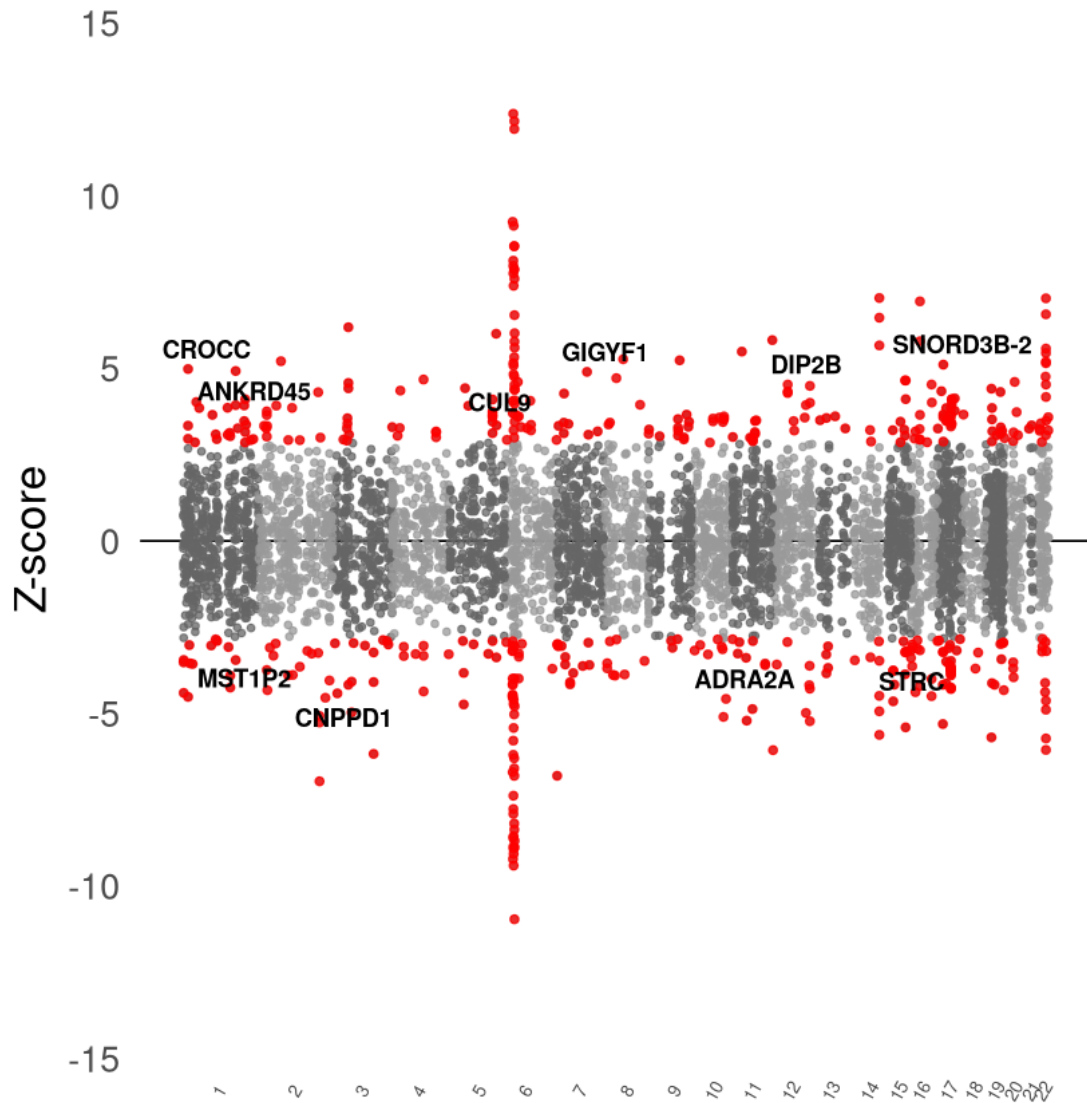
```
[14]: df = twas_data %>% filter(bonferroni < 0.05, p > 5e-8) %>%
      arrange(fdr) %>% head(10)
df
```

A data.table: 10 × 20

	id <chr>	file <chr>	hsq <dbl>	best_gwas_id <chr>	eql_id <chr>
	SNORD3B-2	ENSG00000262074	0.419966	chr17:19250104:T:C	chr17:19340319:C:T
	CROCC	ENSG00000058453	0.289504	chr1:16497972:T:C	chr1:16497972:T:C
	ANKRD45	ENSG00000183831	0.186577	chr1:173743105:T:C	chr1:173740560:G:T
	GIGYF1	ENSG00000146830	0.257250	chr7:100717894:A:G	chr7:100683273:A:G
	STRC	ENSG00000242866	0.187533	chr15:43782086:A:G	chr15:43631757:A:G
	CUL9	ENSG00000112659	0.408106	chr6:43217995:A:G	chr6:43183242:T:C
	ADRA2A	ENSG00000150594	0.228912	chr10:111513602:C:T	chr10:111189236:G:T
	CNPPD1	ENSG00000115649	0.451437	chr2:219196879:T:C	chr2:219185985:G:A
	DIP2B	ENSG00000066084	0.109400	chr12:50258497:G:A	chr12:50489104:T:C
	MST1P2	ENSG00000186301	0.365544	chr1:16497972:T:C	chr1:16996954:T:C

```
[15]: manhplot1 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
      geom_hline(yintercept = 0, color = "black", linetype = "solid") +
      geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
      geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.75)
      ↪+
      scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
      scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
      scale_color_manual(values = rep(c("grey40", "grey60"),
      ↪unique(length(axis_set$chr)))) +
      labs(x = NULL, y = "Z-score") +
      ggrepel::geom_text_repel(data=df, fontface = "bold", aes(x=bp_cum,
      ↪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(),
        axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
      ↪size=8, vjust=0.5)
      )
manhplot1
```





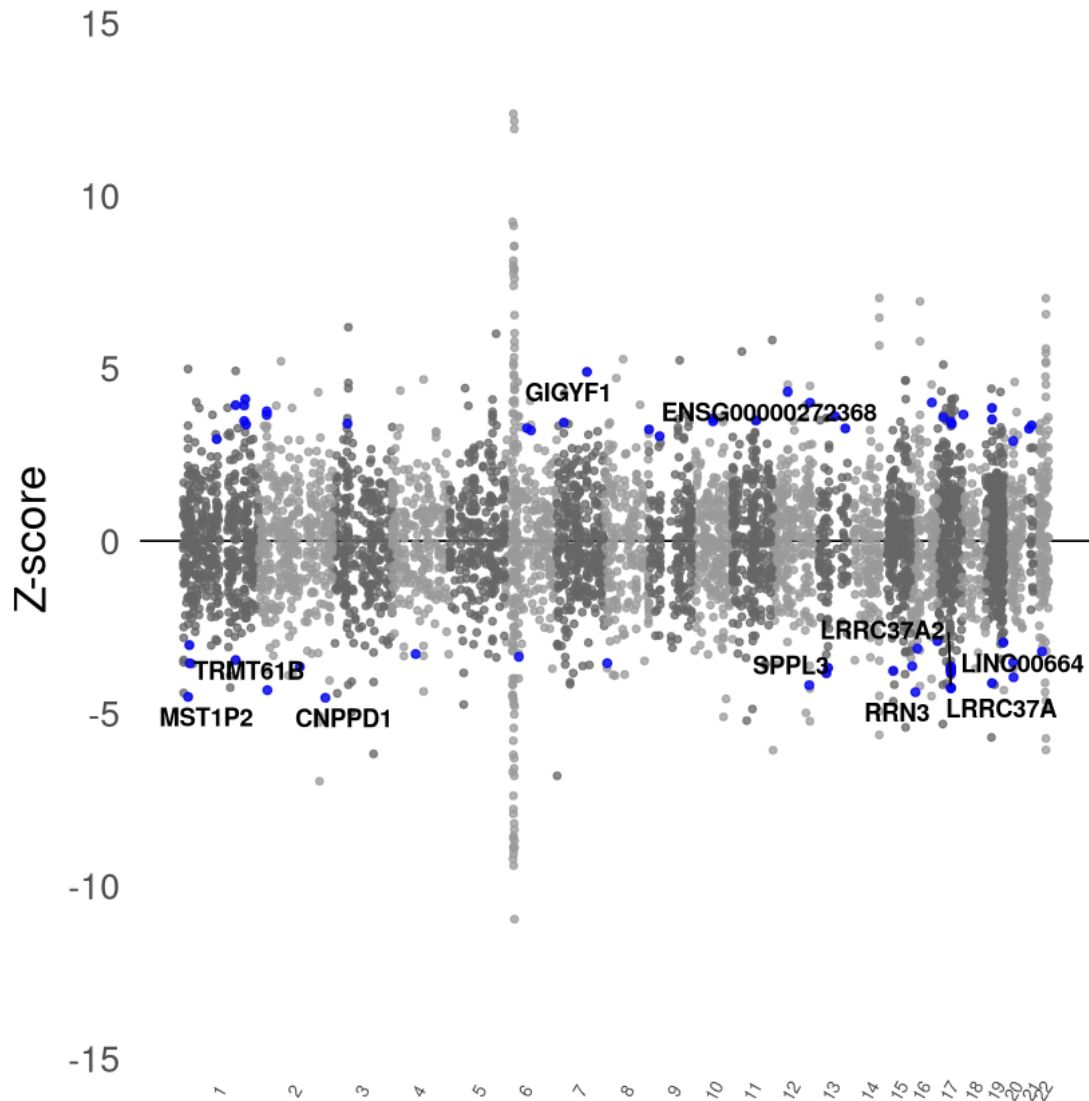
### 1.2.8 Overlapping tissue in significant GWAS loci

```
[16]: overlapping = overlapping %>% filter(p > 5e-8) %>% arrange(fdr)
      overlapping %>% head(10)
```

	id	file	chr_twas	our_snp_id	best_gwas_id
	<chr>	<chr>	<int>	<chr>	<chr>
A data.table: 10 × 9	GIGYF1	ENSG00000146830	7	chr7:100717894:A:G	chr7:100717894:A:G
	CNPPD1	ENSG00000115649	2	chr2:219196879:T:C	chr2:219196879:T:C
	MST1P2	ENSG00000186301	1	chr1:16497972:T:C	chr1:16497972:T:C
	RRN3	ENSG00000085721	16	chr16:15034559:A:G	chr16:15034559:A:G
	ENSG00000272368	ENSG00000272368	12	chr12:50258497:G:A	chr12:50258497:G:A
	TRMT61B	ENSG00000171103	2	chr2:28777421:C:T	chr2:28777421:C:T
	LRRC37A	ENSG00000176681	17	chr17:46288008:C:T	chr17:46288008:C:T
	LRRC37A2	ENSG00000238083	17	chr17:46288008:C:T	chr17:46288008:C:T
	SPPL3	ENSG00000157837	12	chr12:121244248:T:C	chr12:121244248:T:C
	LINC00664	ENSG00000268658	19	chr19:21282807:T:C	chr19:21282807:T:C

```
[17]: dfx = overlapping %>% select(file) %>% inner_join(twas_data, by='file') %>%
  arrange(fdr)

manhplot2 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  #geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.
  ↪75) +
  geom_point(data=dfx, color="blue", alpha=0.75) +
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
  scale_color_manual(values = rep(c("grey40", "grey60"),
  ↪unique(length(axis_set$chr)))) +
  labs(x = NULL, y = "Z-score") +
  ggrepel::geom_text_repel(data=head(dfx, 10), fontface = "bold",
  ↪aes(x=bp_cum, 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(),
    axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
    ↪size=8, vjust=0.5)
  )
manhplot2
```



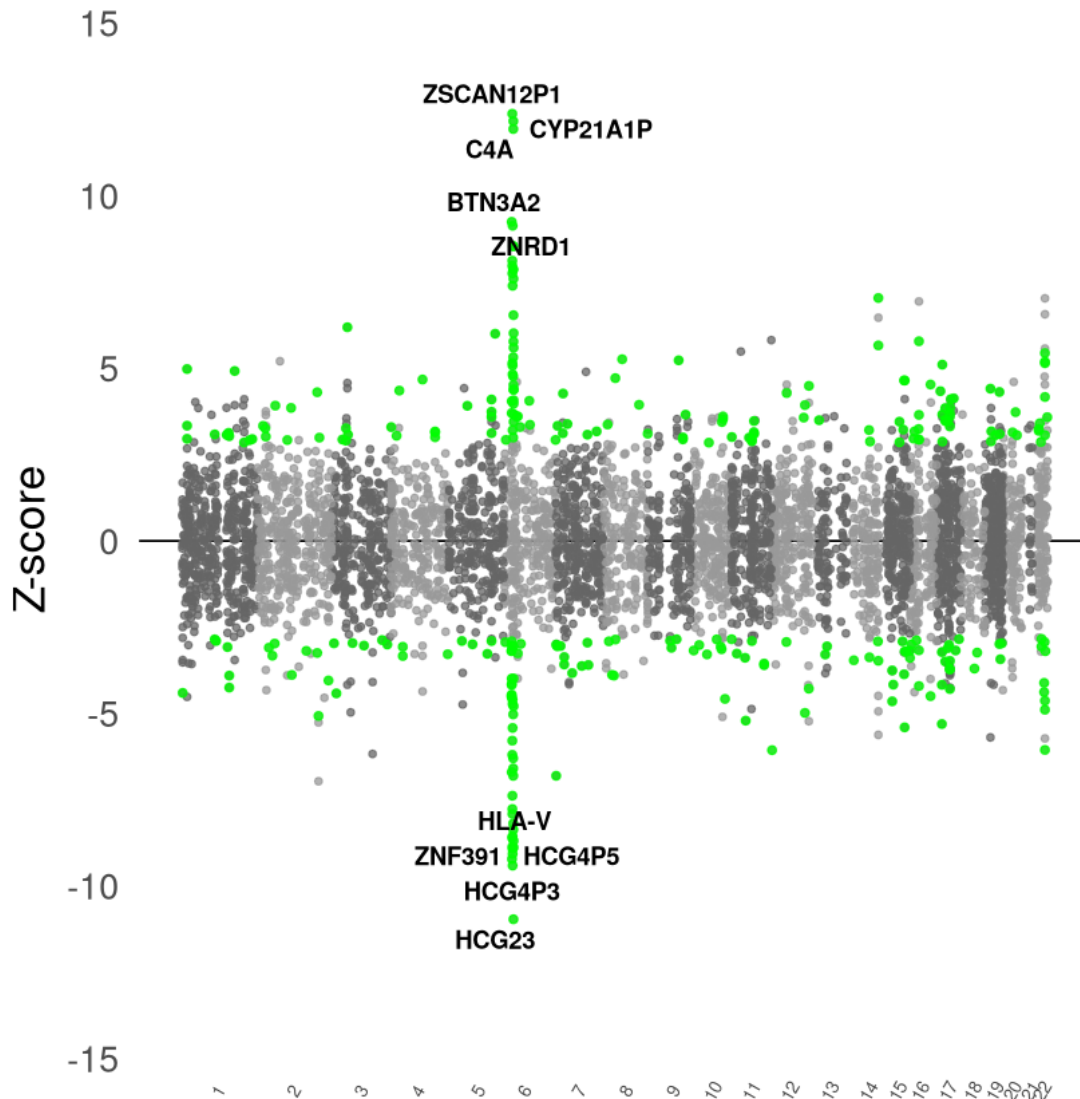
### 1.2.9 Caudate only in significant GWAS loci

```
[18]: caudate_only <- data.table::fread("../_m/caudate_only_twasList_genes.txt")
      ↪ %>% janitor::clean_names() %>%
      select(id, file, chr_twas, our_snp_id, best_gwas_id, p, twas_z, twas_p,
      ↪ fdr) %>% arrange(fdr)
      caudate_only %>% head(10)
```

	id	file	chr_twas	our_snp_id	best_gwas_id
	<chr>	<chr>	<int>	<chr>	<chr>
A data.table: 10 × 9	ZSCAN12P1	ENSG00000219891	6	chr6:28426903:C:T	chr6:28426903:C:T
	CYP21A1P	ENSG00000204338	6	chr6:31793436:G:A	chr6:31793436:G:A
	C4A	ENSG00000244731	6	chr6:31793436:G:A	chr6:31793436:G:A
	HCG23	ENSG00000228962	6	chr6:32231204:C:T	chr6:32231204:C:T
	HCG4P3	ENSG00000237669	6	chr6:29516438:G:T	chr6:29516438:G:T
	BTN3A2	ENSG00000186470	6	chr6:26463346:G:T	chr6:26463346:G:T
	ZNF391	ENSG00000124613	6	chr6:27424023:C:T	chr6:27424023:C:T
	ZNRD1	ENSG00000066379	6	chr6:29591461:T:C	chr6:29591461:T:C
	HCG4P5	ENSG00000227766	6	chr6:29445226:C:T	chr6:29445226:C:T
	HLA-V	ENSG00000181126	6	chr6:29315895:C:T	chr6:29315895:C:T

```
[19]: df2 = caudate_only %>% select(file) %>% inner_join(twas_data, by="file") %>%
  ↪ arrange(fdr)

manhplot3 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  #geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.
  ↪ 75) +
  #geom_point(data=dfx, color="blue", alpha=0.75) +
  geom_point(data=df2, color="green", alpha=0.75) +
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
  scale_color_manual(values = rep(c("grey40", "grey60"),
  ↪ unique(length(axis_set$chr)))) +
  labs(x = NULL, y = "Z-score") +
  ggrepel::geom_text_repel(data=head(df2, 10), fontface = "bold",
  ↪ aes(x=bp_cum, 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(),
    axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
  ↪ size=8, vjust=0.5)
  )
manhplot3
```



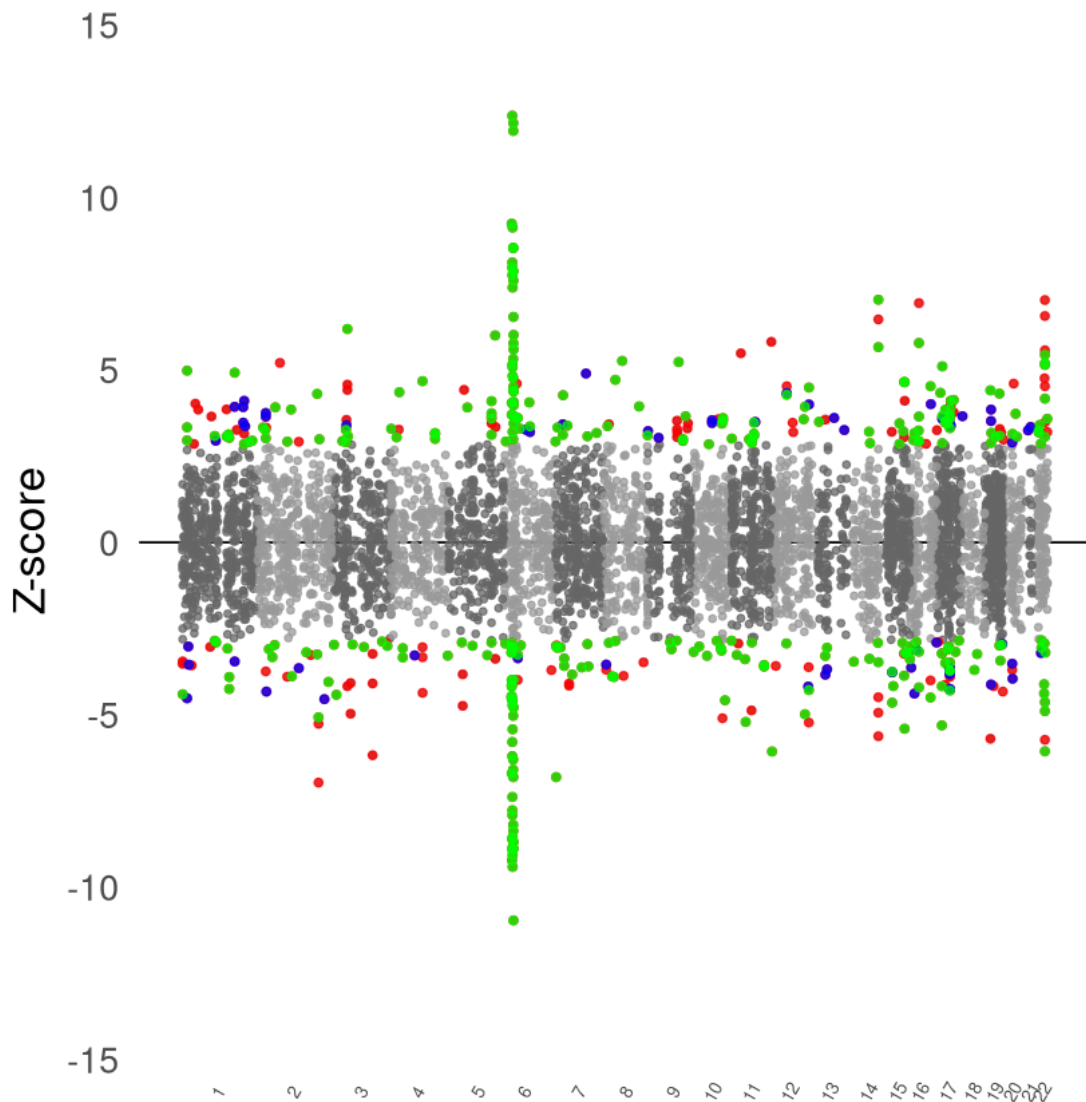
### 1.2.10 No labels

```
[20]: manhplot0 <- ggplot(twas_data, aes(x=bp_cum, y=twas_z)) +
  geom_hline(yintercept = 0, color = "black", linetype = "solid") +
  geom_point(alpha=0.75, size=1.25, aes(color=forcats::as_factor(chr))) +
  geom_point(data=twas_data[twas_data$fdr < 0.05, ], color="red", alpha=0.75)
  ↪+
  geom_point(data=dfx, color="blue", alpha=0.75) +
  geom_point(data=df2, color="green", alpha=0.75) +
  scale_x_continuous(label = axis_set$chr, breaks = axis_set$center) +
  scale_y_continuous(expand = c(0,0), limits = c(-15, 15)) +
  scale_color_manual(values = rep(c("grey40", "grey60"),
  ↪unique(length(axis_set$chr)))) +
```

```

labs(x = NULL, y = "Z-score") +
  #ggrepel::geom_text_repel(data=head(df2, 10), fontface = "bold",
  ↪aes(x=bp_cum, 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(),
    axis.title.y=element_markdown(), axis.text.x=element_text(angle=60,
  ↪size=8, vjust=0.5)
  )
manhplot0

```



### 1.3 Save plots

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