

Quad versus Tibialis

Adam Struck

February 25, 2015

Contents

Identify mis-regulated nonUTRevents	2
Compare delta psi values	2
Scatterplots of delta psi values	2
All events dyesregulated in either quad or tibialis (DM vs Control)	2
Events dyesregulated in either quad or tibialis (DM vs Control) that are also differentially spliced between healthy quad and healthy tibialis	3
Linear model residuals [DM_Tibialis_deltapsi ~ DM_Quad_deltapsi] plotted against healthy quad_vs_tibialis deltapsi	5
Gene ontology term enrichment for dysregulated events in quad OR tibialis (DM vs Control) AND between controls (Quad vs Tib)	6
Ontology term enrichment results	7

```
library(knitr)
library(ggplot2)
library(topGO)
library(org.Hs.eg.db)
library(dplyr)
library(tidyr)
library(data.table)
library(foreach)
library(stringr)
library(Rgraphviz)
```

```
opts_chunk$set(background='gray80', echo = TRUE, tidy=TRUE,
                warning = FALSE, cache=FALSE, comment='', dpi=72)
```

```
find.dm.events <- function(DT) {
  if ("gene_symbol" %in% colnames(DT)) {
    DT %>% filter(abs(delta_psi_mean) >= 0.05, Control_n/max(Control_n,
      na.rm = TRUE) >= 0.75, DM1_n/max(DM1_n, na.rm = TRUE) >= 0.75, DM1_n_sig/DM1_n >=
      0.25) %>% select(gene_symbol, event_name, isoforms, Control_psi_mean,
      Control_psi_sd, Control_n, DM1_psi_mean, DM1_psi_sd, DM1_n, delta_psi_mean,
      DM1_n_sig) %>% arrange(desc(abs(delta_psi_mean)))
  } else {
    DT %>% filter(abs(delta_psi_mean) >= 0.05, Control_n/max(Control_n,
      na.rm = TRUE) >= 0.75, DM1_n/max(DM1_n, na.rm = TRUE) >= 0.75, DM1_n_sig/DM1_n >=
      0.25) %>% select(event_name, isoforms, Control_psi_mean, Control_psi_sd,
      Control_n, DM1_psi_mean, DM1_psi_sd, DM1_n, delta_psi_mean, DM1_n_sig) %>%
      arrange(desc(abs(delta_psi_mean)))
  }
}
```

Identify mis-regulated nonUTRevents

```
event_type <- "nonUTRevents.multi"

## Load healthy quadricep versus tibialis results
allControls_res <- tbl_dt(fread(paste("~/Projects/DMseq/results/allControls/allControls",
  event_type, "results.txt", sep = "_")))
allControls_res <- allControls_res %>% mutate(delta_psi = Quad_psi_mean - Tibialis_psi_mean)

## filter to identify events with differnet splicing patterns between the
## tissues
quad_vs_tibialis <- allControls_res %>% select(gene_symbol, event_name, Quad_psi_mean,
  Quad_n, Tibialis_psi_mean, Tibialis_n, Quad_vs_Tibialis_n_sig, isoforms,
  delta_psi) %>% filter(Quad_n/max(Quad_n, na.rm = TRUE) >= 0.75, Tibialis_n/max(Tibialis_n,
  na.rm = TRUE) >= 0.75, abs(delta_psi) >= 0.05, Quad_vs_Tibialis_n_sig/Quad_n >=
  0.25) %>% arrange(desc(Quad_vs_Tibialis_n_sig))

## Load tibialis & quadricep DM_vs_Control results
tibialis_pdata <- tbl_dt(fread("~/Projects/DMseq/data/DM_tibialis_pdata.txt"))
tibialis_res <- tbl_dt(fread(paste("~/Projects/DMseq/results/tibialis/tibialis",
  event_type, "results.txt", sep = "_")))
tibialis_res <- tibialis_res %>% mutate(frac_sig = DM1_n_sig/DM1_n)

quadricep_pdata <- tbl_dt(fread("~/Projects/DMseq/data/DM_quadricep_pdata.txt"))
quadricep_res <- tbl_dt(fread(paste("~/Projects/DMseq/results/quadricep/quadricep",
  event_type, "results.txt", sep = "_")))
quadricep_res <- quadricep_res %>% mutate(frac_sig = DM1_n_sig/DM1_n)

## Filter to identify events mis-regulated in DM1
dm_tibialis <- find.dm.events(tibialis_res)
dm_quadricep <- find.dm.events(quadricep_res)
```

Compare delta psi values

```
event_set <- intersect(union(dm_tibialis$isoforms, dm_quadricep$isoforms), quad_vs_tibialis$isoforms)

deltapsi_data <- Reduce(function(...) merge(..., by = "isoforms", all = TRUE),
  list(select(allControls_res, gene_symbol, event_name, isoforms, delta_psi) %>%
    filter(isoforms %in% event_set), select(tibialis_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set), select(quadricep_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set)))
setnames(deltapsi_data, c("isoforms", "gene_symbol", "event_name", "Quad_vs_Tibialis_deltapsi",
  "Tibialis", "Quad"))
```

Scatterplots of delta psi values

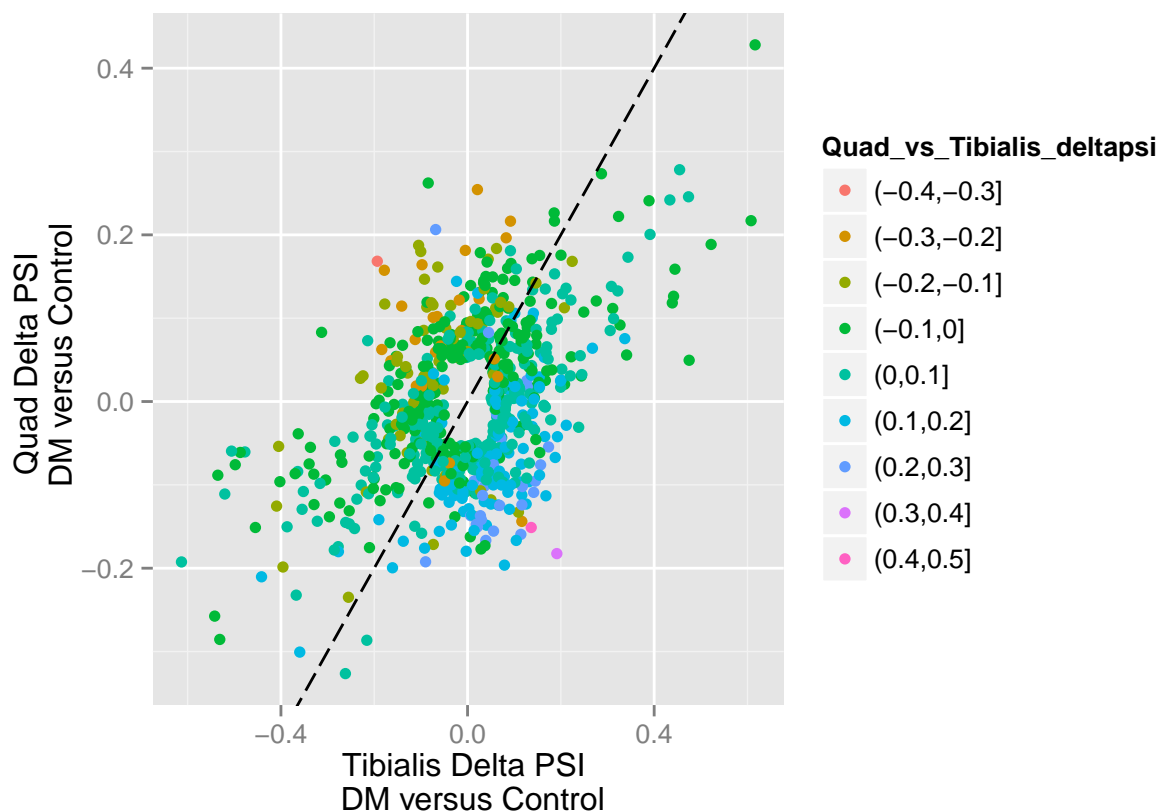
All events dyesregulated in either quad or tibialis (DM vs Control)

```

event_set <- union(dm_tibialis$isoforms, dm_quadricep$isoforms)
deltapsi_data <- Reduce(function(...) merge(..., by = "isoforms", all = TRUE),
  list(select(allControls_res, gene_symbol, event_name, isoforms, delta_psi) %>%
    filter(isoforms %in% event_set), select(tibialis_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set), select(quadricep_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set)))
setnames(deltapsi_data, c("isoforms", "gene_symbol", "event_name", "Quad_vs_Tibialis_deltapsi",
  "Tibialis", "Quad"))

ggplot(deltapsi_data, aes(x = Tibialis, y = Quad, colour = cut(deltapsi_data$Quad_vs_Tibialis_deltapsi,
  seq(-1, 1, 0.1)))) + geom_point() + geom_abline(yintercept = 0, slope = 1,
  linetype = "longdash") + labs(y = "Quad Delta PSI \n DM versus Control",
  x = "Tibialis Delta PSI \n DM versus Control") + scale_color_discrete(name = "Quad_vs_Tibialis_deltapsi")

```



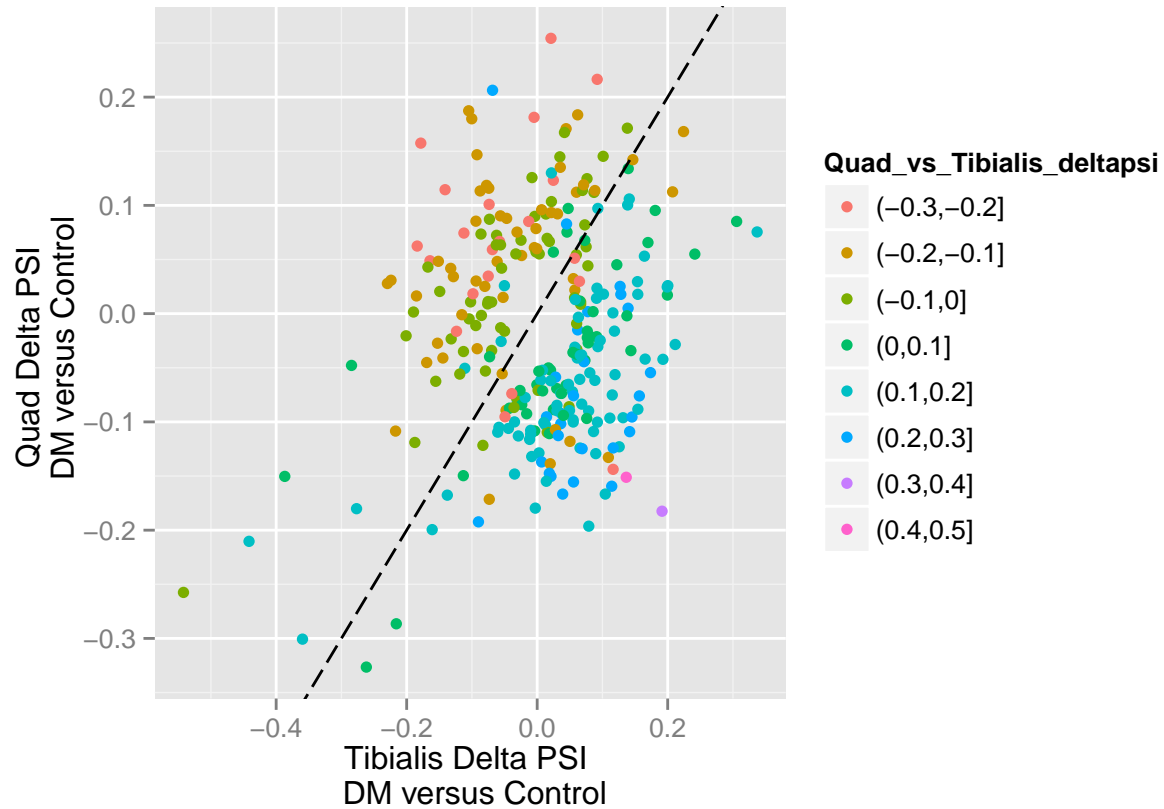
Events dysregulated in either quad or tibialis (DM vs Control) that are also differentially spliced between healthy quad and healthy tibialis

```

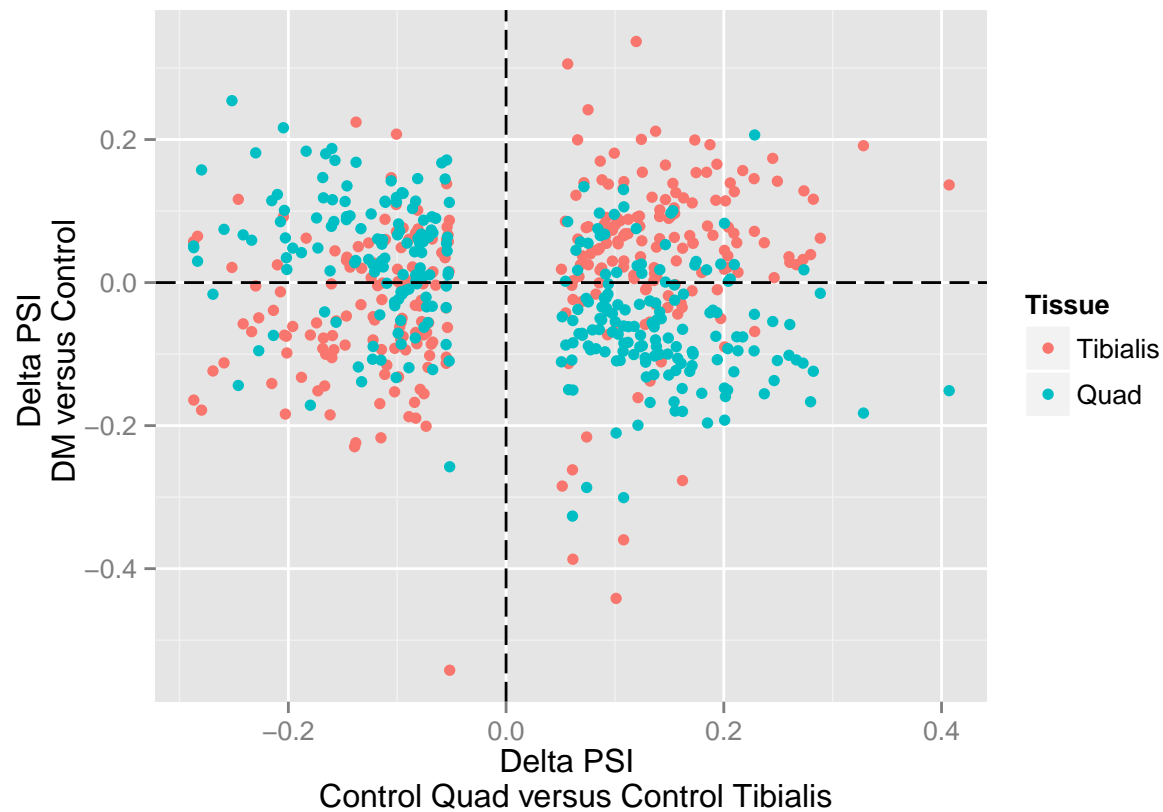
event_set <- intersect(union(dm_tibialis$isoforms, dm_quadricep$isoforms), quad_vs_tibialis$isoforms)
deltapsi_data <- Reduce(function(...) merge(..., by = "isoforms", all = TRUE),
  list(select(allControls_res, gene_symbol, event_name, isoforms, delta_psi) %>%
    filter(isoforms %in% event_set), select(tibialis_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set), select(quadricep_res, isoforms, delta_psi_mean) %>%
    filter(isoforms %in% event_set)))
setnames(deltapsi_data, c("isoforms", "gene_symbol", "event_name", "Quad_vs_Tibialis_deltapsi",
  "Tibialis", "Quad"))

```

```
ggplot(deltapsi_data, aes(x = Tibialis, y = Quad, colour = cut(deltapsi_data$Quad_vs_Tibialis_deltapsi,
  seq(-1, 1, 0.1)))) + geom_point() + geom_abline(yintercept = 0, slope = 1,
  linetype = "longdash") + labs(y = "Quad Delta PSI \n DM versus Control",
  x = "Tibialis Delta PSI \n DM versus Control") + scale_color_discrete(name = "Quad_vs_Tibialis_deltapsi")
```



```
deltapsi_data_long <- gather(deltapsi_data, key = "Tissue", value = "DM_vs_Control_deltapsi",
  5:6)
ggplot(deltapsi_data_long, aes(x = Quad_vs_Tibialis_deltapsi, y = DM_vs_Control_deltapsi,
  colour = Tissue)) + geom_point() + geom_hline(yintercept = 0, linetype = "longdash") +
  geom_vline(xintercept = 0, linetype = "longdash") + labs(y = "Delta PSI \n DM versus Control",
  x = "Delta PSI \n Control Quad versus Control Tibialis")
```



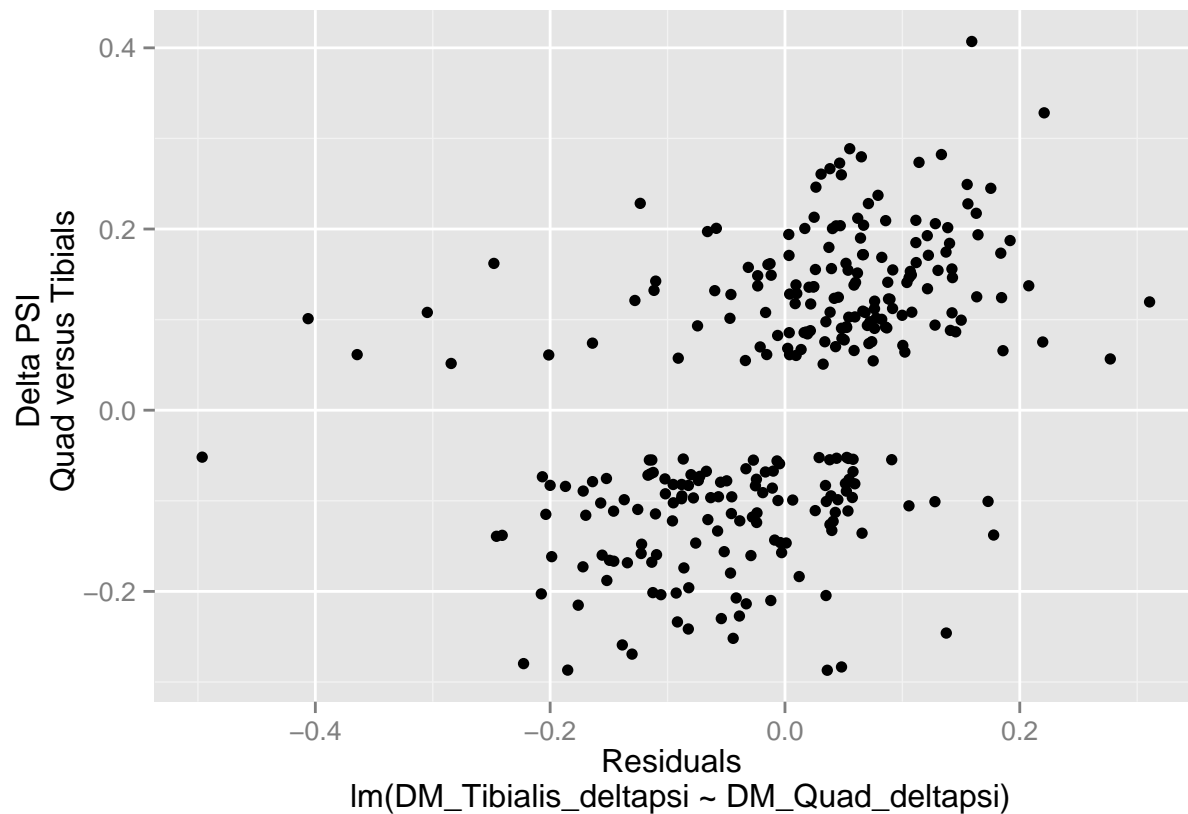
Linear model residuals [DM_Tibialis_deltapsi ~ DM_Quad_deltapsi] plotted against healthy quad_vs_tibialis deltapsi

```
fit <- lm(Tibialis ~ Quad, data = deltapsi_data)
cor.test(fit$residuals, deltapsi_data$Quad_vs_Tibialis_deltapsi)
```

Pearson's product-moment correlation

```
data: fit$residuals and deltapsi_data$Quad_vs_Tibialis_deltapsi
t = 9.4909, df = 284, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3973429 0.5740249
sample estimates:
      cor
0.4907118
```

```
ggplot() + geom_point(aes(y = deltapsi_data$Quad_vs_Tibialis_deltapsi, x = fit$residuals)) +
  labs(x = "Residuals \n lm(DM_Tibialis_deltapsi ~ DM_Quad_deltapsi)", y = "Delta PSI \n Quad versus Tibialis")
```



Gene ontology term enrichment for dysregulated events in quad OR tibialis (DM vs Control) AND between controls (Quad vs Tib)

```
all_muscle_genes <- filter(allControls_res, Quad_n/max(Quad_n, na.rm = TRUE) >=
  0.75, Tibialis_n/max(Tibialis_n, na.rm = TRUE) >= 0.75)$gene_symbol %>%
  unique()

## different in quad OR tibialis (DM vs Control) AND between controls (Quad
## vs Tib)
sig_genes <- intersect(union(dm_tibialis$gene_symbol, dm_quadricep$gene_symbol),
  quad_vs_tibialis$gene_symbol)

ontology_class <- "BP"

myGO2genes <- AnnotationDbi::select(org.Hs.eg.db, keys = all_muscle_genes, columns = c("ENSEMBL",
  "GO"), keytype = "SYMBOL")
myGO2genes <- myGO2genes %>% filter(!is.na(ENSEMBL)) %>% tbl_df
myGO2genesList <- tapply(filter(myGO2genes, ONTOLOGY == ontology_class)$ENSEMBL,
  filter(myGO2genes, ONTOLOGY == ontology_class)$GO, FUN = c)

ensemblIDs <- myGO2genes$ENSEMBL[match(all_muscle_genes, myGO2genes$SYMBOL)]
geneList <- factor(as.integer(all_muscle_genes %in% sig_genes))
names(geneList) <- ensemblIDs

GOdata <- new("topGOdata", description = "GO analysis of genes with differential splicing",
```

```

ontology = ontology_class, allGenes = geneList, nodeSize = 5, annot = annFUN.GO2genes,
GO2genes = myGO2genesList)

resultFisher <- runTest(GOdata, algorithm = "classic", statistic = "fisher")
resultFisher.elim <- runTest(GOdata, algorithm = "elim", statistic = "fisher")

allRes <- GenTable(GOdata, classicFisher = resultFisher, elimFisher = resultFisher.elim,
  orderBy = "elimFisher", ranksOf = "elimFisher", topNodes = 20)

head(sig_genes, n = 50)

```

```

[1] "PDLIM3"          "ABLM1"          "ARHGEF10L"
[4] "MBNL1"          "MYBPC1"         "TACC2"
[7] "NFIX"           "BEST3"          "MEF2C"
[10] "SEMA6C"         "NUMA1"          "GOLGA4"
[13] "PHKA1"          "BIN1"           "SORBS1"
[16] "KIAA1191"       "TTN"            "UBE2D3"
[19] "NDUFV3"         "RYR1"           "PPP1R12B"
[22] "DTNA"           "HP1BP3"         "HDAC9"
[25] "KIF1B"          "EI24"           "PREPL"
[28] "WNK1"           "CALU"           "NEB"
[31] "ITGA7"          "HFE2"           "YBX3"
[34] "SCARB1"         "ATP6V1G2-DDX39B" "PQBP1"
[37] "CORO6"          "SPEG"           "TOR1AIP1"
[40] "PI4KB"          "IMMT"           "UGP2"
[43] "DCAF6"          "ANKRD10"        "RPS3"
[46] "CELF1"          "EIF4A1"         "TMEM159"
[49] "RPS3A"          "IDH3A"

```

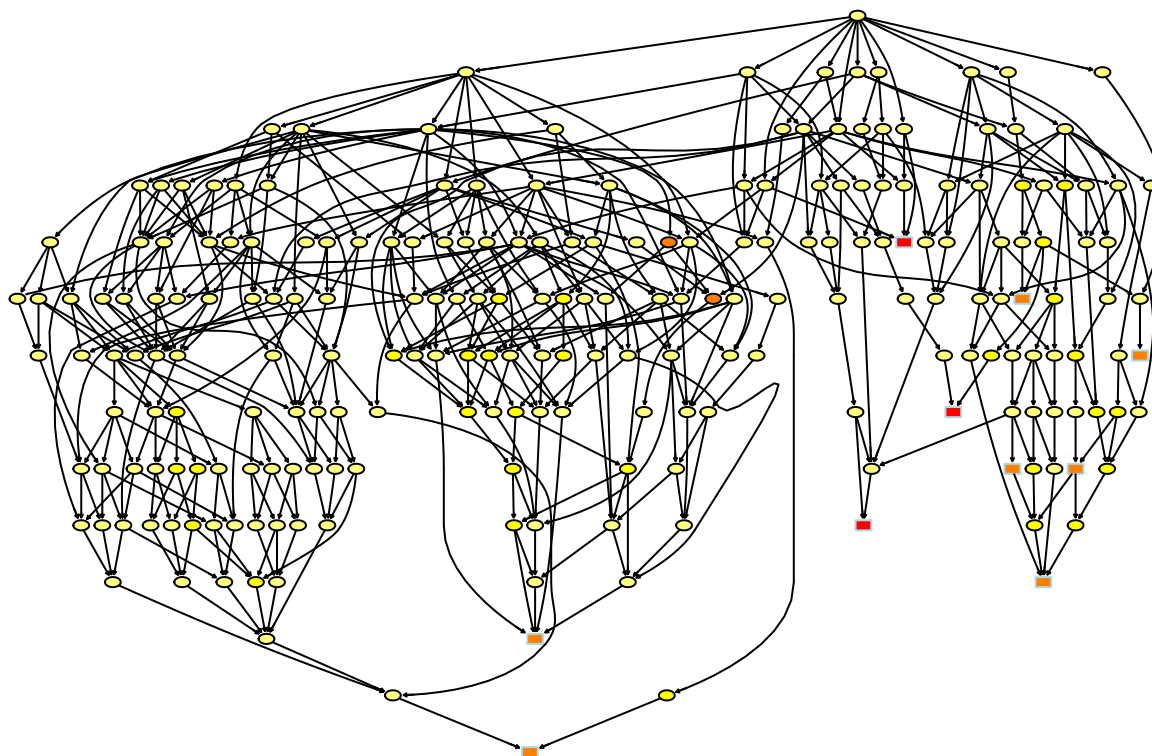
Ontology term enrichment results

```
kable(allRes)
```

GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFish
GO:0030049	muscle filament sliding	17	6	1.05	0.00036	0.00036
GO:0045927	positive regulation of growth	43	9	2.65	0.00100	0.00097
GO:0045214	sarcomere organization	14	5	0.86	0.00110	0.00107
GO:0007519	skeletal muscle tissue development	48	12	2.96	2.2e-05	0.00133
GO:0090257	regulation of muscle system process	54	10	3.33	0.00144	0.00139
GO:0045725	positive regulation of glycogen biosynth...	5	3	0.31	0.00213	0.00210
GO:0043462	regulation of ATPase activity	16	5	0.99	0.00216	0.00211
GO:0048747	muscle fiber development	22	8	1.36	2.8e-05	0.00309
GO:0009409	response to cold	11	4	0.68	0.00334	0.00328
GO:0014888	striated muscle adaptation	11	4	0.68	0.00334	0.00328
GO:0048742	regulation of skeletal muscle fiber deve...	11	4	0.68	0.00334	0.00328
GO:0006107	oxaloacetate metabolic process	6	3	0.37	0.00407	0.00402
GO:0010524	positive regulation of calcium ion trans...	6	3	0.37	0.00407	0.00402
GO:2000114	regulation of establishment of cell pola...	6	3	0.37	0.00407	0.00402
GO:0070296	sarcoplasmic reticulum calcium ion trans...	19	5	1.17	0.00494	0.00483
GO:0043624	cellular protein complex disassembly	85	12	5.24	0.00546	0.00526

GO.ID	Term	Annotated	Significant	Expected	classicFisher	elimFisher
GO:0043484	regulation of RNA splicing	36	7	2.22	0.00564	0.00549
GO:0002181	cytoplasmic translation	7	3	0.43	0.00680	0.00671
GO:0003208	cardiac ventricle morphogenesis	7	3	0.43	0.00680	0.00671
GO:0043536	positive regulation of blood vessel endo...	7	3	0.43	0.00680	0.00671

```
showSigOfNodes(GOdata, score(resultFisher.elim), firstSigNodes = 10, useInfo = "all")
```



```
$dag
```

```
A graphNEL graph with directed edges
```

```
Number of Nodes = 226
```

```
Number of Edges = 488
```

```
$complete.dag
```

```
[1] "A graph with 226 nodes."
```