Exploring chromosome 12

Going to look closer at chromosome 12

```r
library("data.table")
library("plotfunctions")
source("myfunctions.R")
```

```r
genepos <- fread("results/gene_pos.gz")
if (!exists("find.effects_TF")){
  find.effects_TF <- fread("results/findeffects_TF_newparams.gz")
}

# Add chomosome and start position to each gene
causal.pos.A <- merge(find.effects_TF, genepos, by.x="geneA", by.y="gene", all.x=T)
colnames(causal.pos.A) <- c("geneA", "geneB", "eqtl.A", "eqtl.B", "A->B", "B->A", "strand.A", "start.A"
causal.pos.B <- merge(causal.pos.A, genepos, by.x="geneB", by.y="gene", all.x=T)


# Keep olnly columns with genes, start positions and chromosomes
causal.pos.B.2 <- unique(causal.pos.B[,.(geneA, geneB, start.A, chr.A, chr.start, chr.id)])
colnames(causal.pos.B.2) <- c("geneA", "geneB", "start.A", "chr.A", "start.B", "chr.B")

## transform chromosome ids into numbers
# remove "chr" part of the chromosome name
causal.pos.B.2$chr.A <-gsub('chr', '', causal.pos.B.2$chr.A)
causal.pos.B.2$chr.B <-gsub('chr', '', causal.pos.B.2$chr.B)
colnames(causal.pos.B.2) <- c("geneA", "geneB", "start.A", "chr.A", "start.B", "chr.B")

# convert roman chromosome numbers to numbers
causal.pos.B.2$chr.A <- as.numeric(as.roman(causal.pos.B.2$chr.A))
causal.pos.B.2$chr.B <- as.numeric(as.roman(causal.pos.B.2$chr.B))

# order values
causal.pos.B.2.order <- causal.pos.B.2[order(chr.A, start.A, chr.B, start.B)]


# organize coordinates so that they are ordered by chromosome
# vector of chromosomes
vchr <- 1:16
# how much space will be separating chromosomes
separator <- 1e5

coordinates_plot <- sort_by_chr(vchr = vchr, causal.pos.B.2.order, separator = separator)
```

```r
if (!exists("find.effects")){
  find.effects <- fread("results/findeffects_all_newparams.gz")
}
```

```
#Create a function to generate a continuous color palette
rbPal <- colorRampPalette(c('blue','red'))


coordinates_plot_cor <- merge(coordinates_plot, find.effects[,.(geneA, geneB, cor)], by=c("geneA", "gene
coordinates_plot_cor$cor <- abs(coordinates_plot_cor$cor)
coordinates_plot_cor <- coordinates_plot_cor[order(cor)]
coordinates_plot_cor$col <- rbPal(100)[as.numeric(cut(coordinates_plot_cor$cor,breaks = 100))]

plot_sorted_coordinates(coordinates_plot_cor, separator = separator, col = coordinates_plot_cor$col)
gradientLegend(format(round(coordinates_plot_cor$cor, 3), nsmall = 3), rbPal(100), inside=F, side=3)
```
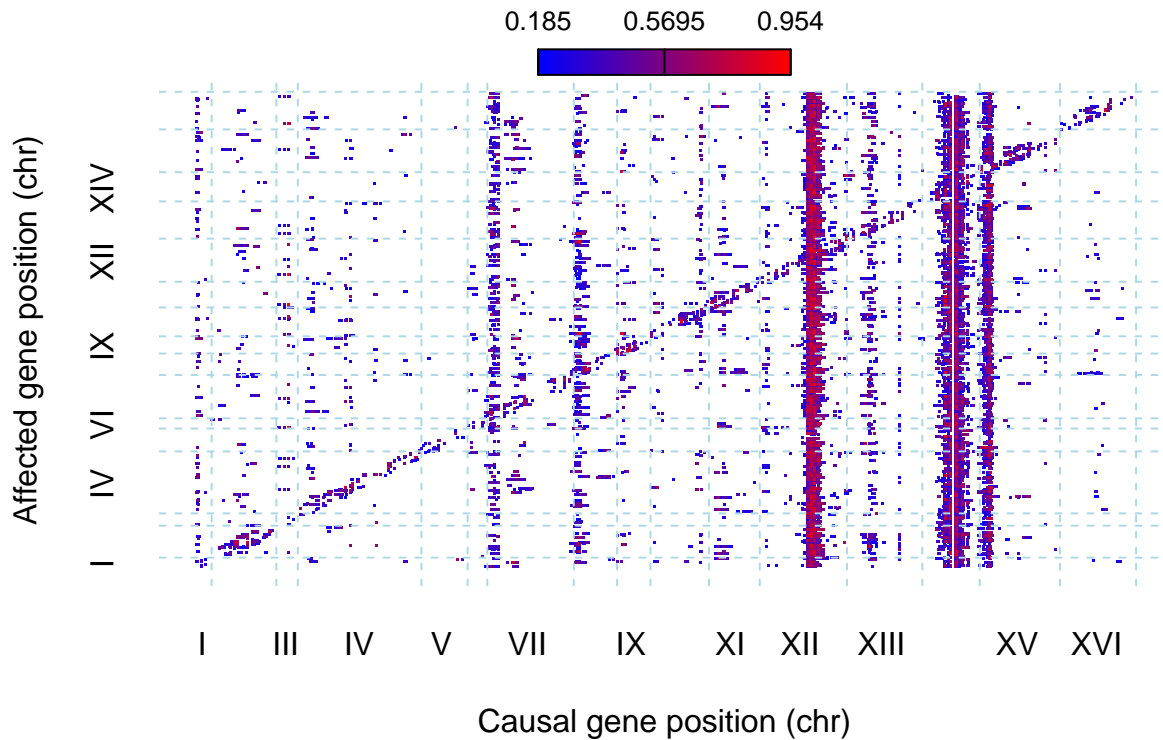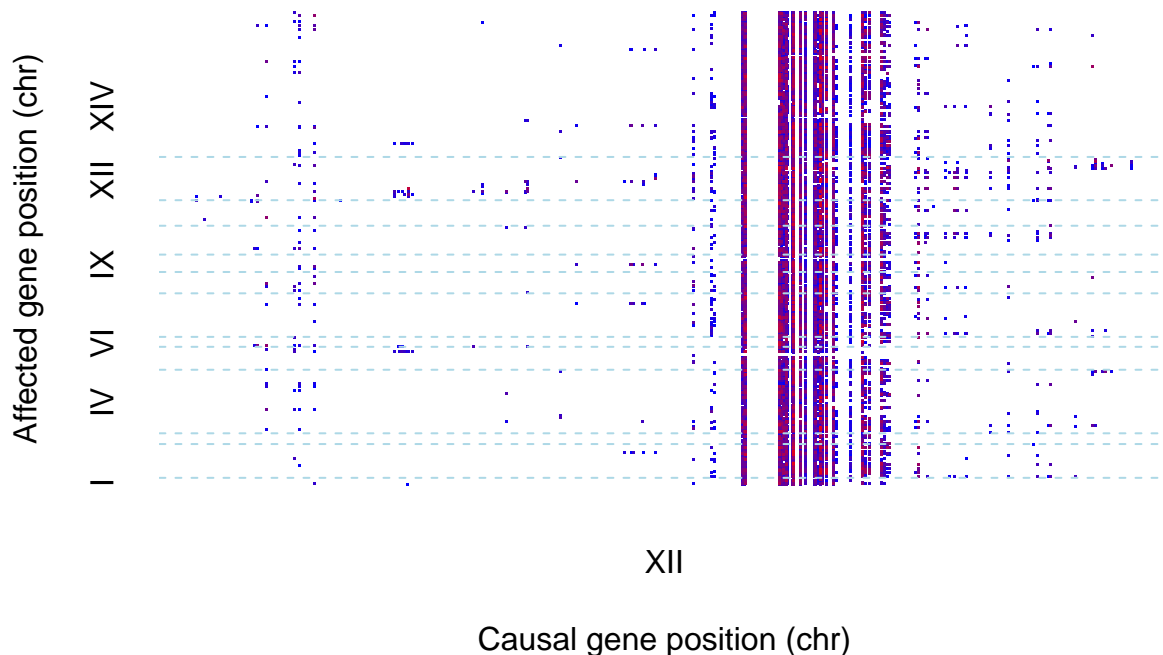

Focus on causal gene 12

```
# coordinates_plot_cor[chr.A==12]
plot_sorted_coordinates(coordinates_plot_cor[chr.A==12], separator = separator, col = coordinates_plot_
```

Affected gene position (chr)

I   IV VI   IX   XII XIV
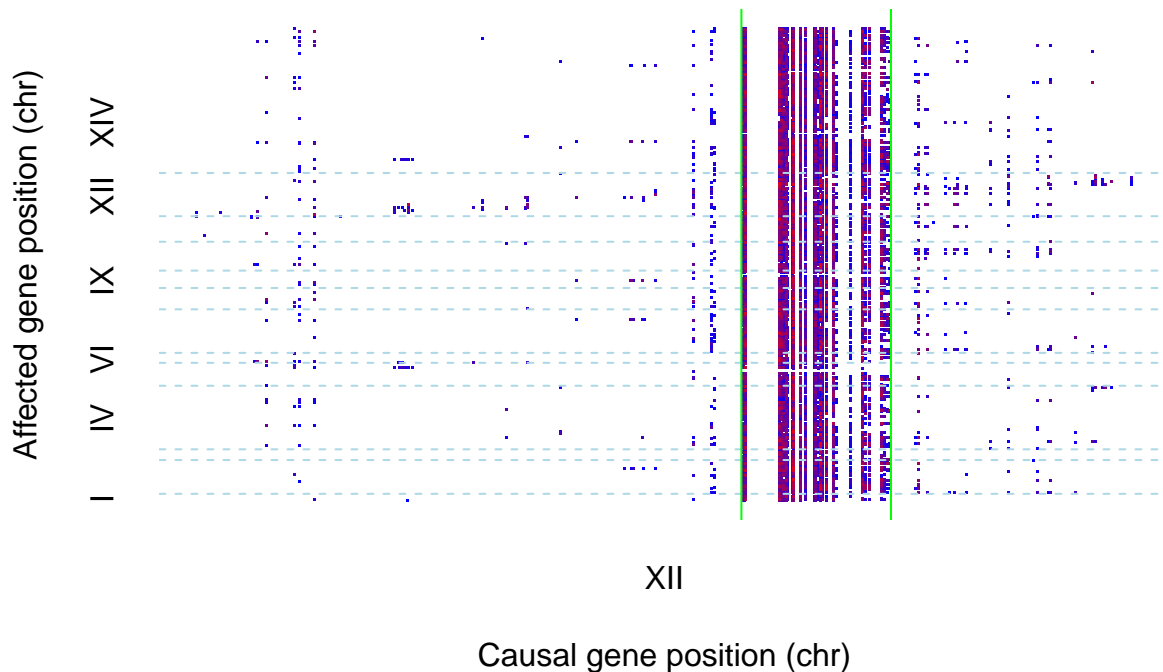
XII

Causal gene position (chr)

Focus on the group of vertical bands

Chromosome 3 seems to mostly be affected by the genes in the vertical bands so I'm going to look closer at chromosome 3

```r
# the first gene of chromosome 12 that affects chromosome 3
lower_limit <- coordinates_plot_cor[chr.A==12 & chr.B==3][order(start.A)][1]$start.A
# the last gene on chromosome 12 that affects chromosome 3 and that's in the band (there's an extra gen
top_limit <- coordinates_plot_cor[chr.A==12 & chr.B==3][order(start.A)][nrow(coordinates_plot_cor[chr.A=

# plot chromosome 12 - the green lines are the limits of the group of vertical bands
plot_sorted_coordinates(coordinates_plot_cor[chr.A==12], separator = separator, col = coordinates_plot_c
abline(v = c(lower_limit-1500, top_limit+1500), col="green")
```

Affected gene position (chr)

XIV XII IX VI IV I

XII

Causal gene position (chr)

there are 30 genes between the two green bands

get gene names for genes on chromosome 12

```r
if (!exists("genes_GO.bio")){
  genes_GO.table <- fread("results/genelistwithGOterm.txt")
  genes_GO.table <- unique(genes_GO.table)
  genes_GO.bio   <- unique(genes_GO.table[GO.namespace=="biological_process"])
}
```

```r
genesA_start_order <- unique(coordinates_plot_cor[chr.A == 12 &chr.B == 3][order(start.A)][, .(geneA, s
chr12_genenames <- merge(genesA_start_order, unique(genes_GO.bio[,.(gene, symbol, gene.name, GO.term)]))
```

Look at the first group

```r
# genes that are affecting chromosome 3 + positions
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim = c(min(genesA_start_order$start.A), top_limit)
)
abline(v=c(min(genesA_start_order$start.A)-1500, 7425907+1500), col="green")
```
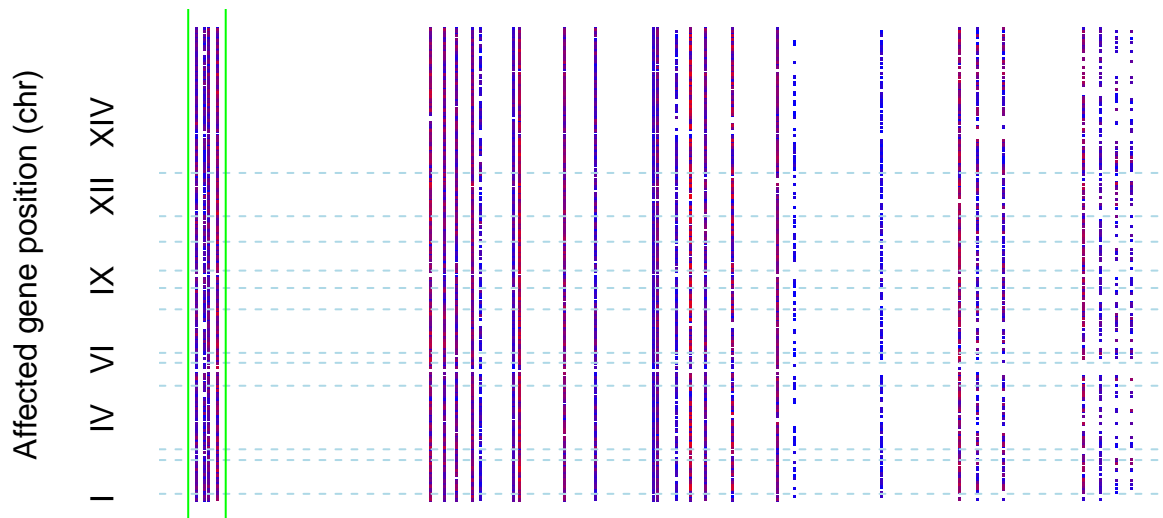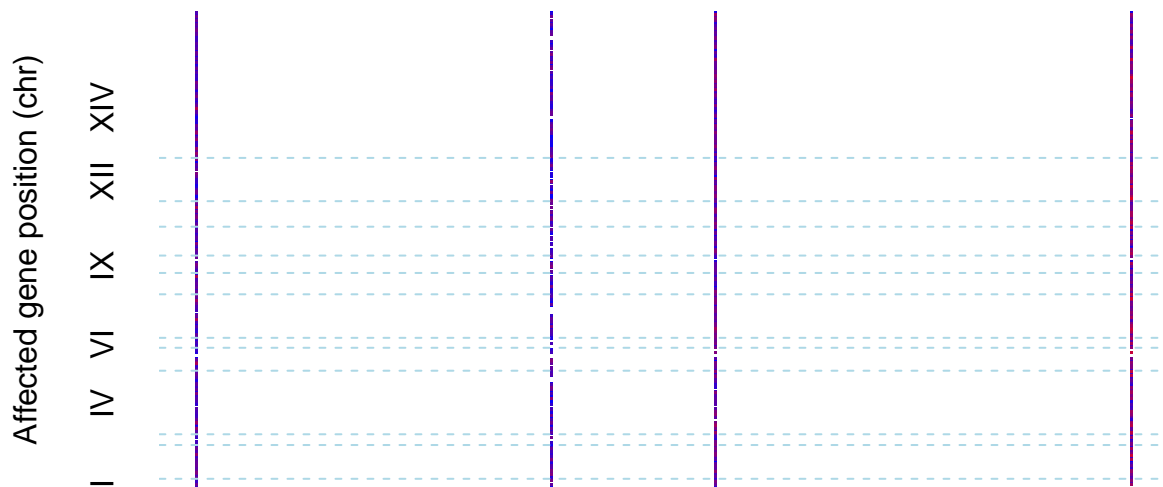
Causal gene position (chr)

There are 4 between the green lines

```r
# genes that are affecting chromosome 3 + positions
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim=c(min(genesA_start_order$start.A), 7425907)
)
abline(v=c(min(genesA_start_order$start.A)-1500, 7425907+1500), col="green")
```


Causal gene position (chr)

genes on the first "band" of chromosome 12

```
# gene names
unique(chr12_genenames[between(start.A,7422391, 7425907)]$gene.name)
```

```
## [1] "Methionine AminoPeptidase"      "CytiDine Deaminase"
## [3] "Effect on Ras Function"         "Increased Recombination Centers"
```

```
# GO terms present
unique(chr12_genenames[between(start.A,7422391-1500, 7425907+1500)]$GO.term)
```

```
##  [1] "proteolysis"
##  [2] "negative regulation of gene expression"
##  [3] "protein initiator methionine removal involved in protein maturation"
##  [4] "protein initiator methionine removal"
##  [5] "cytidine catabolic process"
##  [6] "deoxycytidine catabolic process"
##  [7] "pyrimidine-containing compound salvage"
##  [8] "cytidine deamination"
##  [9] "protein targeting to membrane"
## [10] "peptidyl-L-cysteine S-palmitoylation"
## [11] "protein palmitoylation"
## [12] "double-strand break repair via nonhomologous end joining"
## [13] "protein ubiquitination"
## [14] "double-strand break repair via synthesis-dependent strand annealing"
```

## Look at the second group

```
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim = c(min(genesA_start_order$start.A), top_limit)
)
abline(v=c(7463067-1500, 7526178+1500), col="green")
```

Causal gene position (chr)
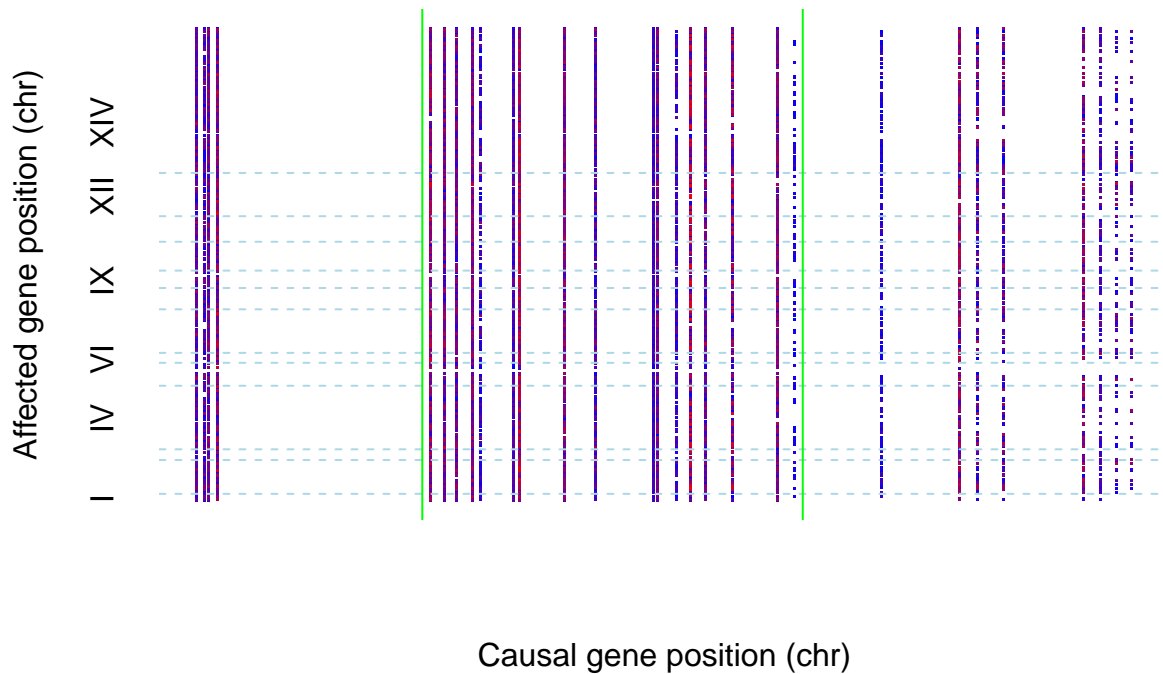
There are 18 genes between the green lines

Looking closer

```
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim = c(7463067, 7526178)
)
abline(v=c(7463067-1500, 7526178+1500), col="green")
```


Causal gene position (chr)
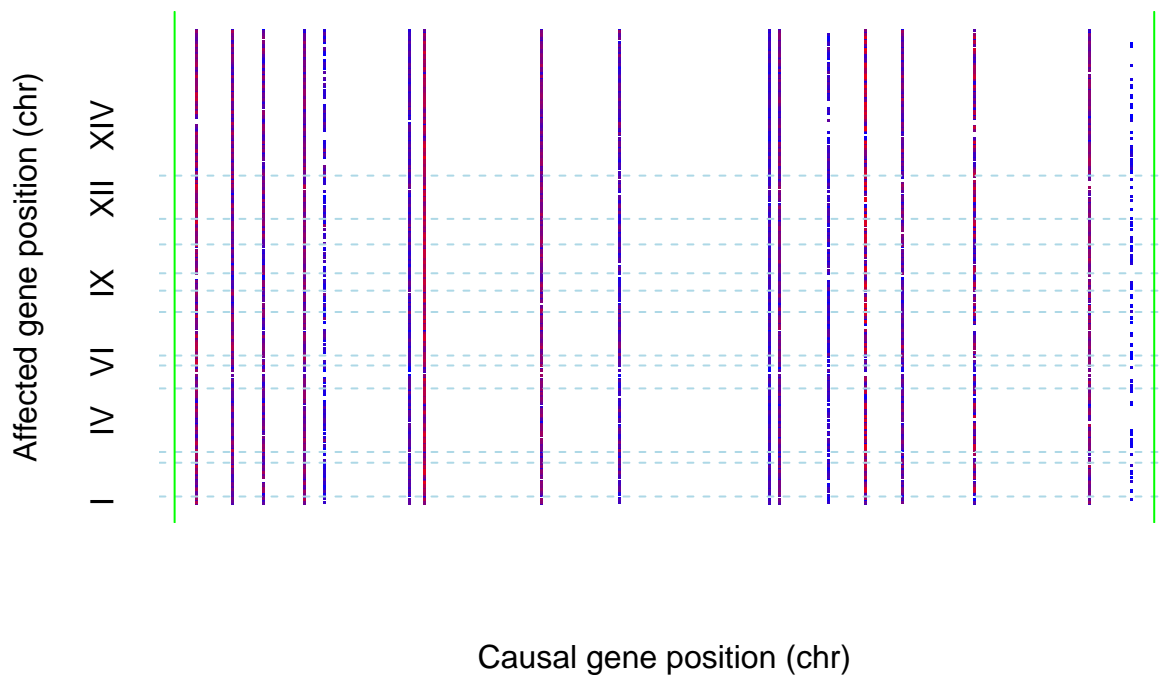
genes on the second band of chromsome 12

```r
# gene names
unique(chr12_genenames[between(start.A,7463067-1500, 7526178+1500)]$gene.name)
```

```
##  [1] "Long-Chain Base"
##  [2] "Vacuolar Protein Sorting"
##  [3] "Yeast Protein Two"
##  [4] "REDuctional division"
##  [5] "Ribosomal Protein of the Small subunit"
##  [6] "Nonhomologous End-Joining defective"
##  [7] "SECretory"
##  [8] "DeCapping Scavenger"
##  [9] "Protein Interacting with Gsy2p"
## [10] NA
## [11] "mitochondrial protein Related to Spastic paraplegia with Optic atrophy and neuropathy SPG55"
## [12] "ChiTinaSe"
## [13] "Mitosis Entry Checkpoint"
## [14] "General Control Derepressed"
## [15] "ExtraCellular Mutant"
## [16] "EXo-1,3-beta-Glucanase"
```

```r
# GO terms present
unique(chr12_genenames[between(start.A,7463067-1500, 7526178+1500)]$GO.term)
```

```
##  [1] "lipid metabolic process"
##  [2] "sphingolipid metabolic process"
##  [3] "response to heat"
##  [4] "phosphorylation"
##  [5] "calcium-mediated signaling"
##  [6] "lipid phosphorylation"
##  [7] "protein targeting to vacuole"
##  [8] "retrograde transport, vesicle recycling within Golgi"
##  [9] "intracellular protein transport"
## [10] "retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum"
## [11] "intra-Golgi vesicle-mediated transport"
## [12] "Golgi to endosome transport"
## [13] "protein transport"
## [14] "macroautophagy"
## [15] "cytoplasm to vacuole transport by the Cvt pathway"
## [16] "Rab protein signal transduction"
## [17] "protein localization to phagophore assembly site"
## [18] "cellular protein-containing complex localization"
## [19] "retrograde transport, endosome to Golgi"
## [20] "synaptonemal complex assembly"
## [21] "reciprocal meiotic recombination"
## [22] "sporulation resulting in formation of a cellular spore"
## [23] "positive regulation of catalytic activity"
## [24] "meiotic cell cycle"
## [25] "meiotic recombination checkpoint"
## [26] "ribosomal small subunit assembly"
## [27] "cytoplasmic translation"
## [28] "rRNA export from nucleus"
## [29] "translation"
## [30] "maturation of SSU-rRNA"
## [31] "positive regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent deca
```

```
## [32] "DNA repair"
## [33] "double-strand break repair"
## [34] "double-strand break repair via nonhomologous end joining"
## [35] "cellular response to DNA damage stimulus"
## [36] "homologous recombination"
## [37] "double-strand break repair via single-strand annealing"
## [38] "endoplasmic reticulum to Golgi vesicle-mediated transport"
## [39] "vesicle fusion"
## [40] "vesicle-mediated transport"
## [41] "vesicle fusion with endoplasmic reticulum"
## [42] "vesicle fusion with Golgi apparatus"
## [43] "deadenylation-dependent decapping of nuclear-transcribed mRNA"
## [44] "cellular response to starvation"
## [45] "nuclear-transcribed mRNA catabolic process, deadenylation-independent decay"
## [46] "positive regulation of exoribonuclease activity"
## [47] "glycogen biosynthetic process"
## [48] "regulation of glycogen biosynthetic process"
## [49] "regulation of phosphoprotein phosphatase activity"
## [50] "spliceosomal snRNP assembly"
## [51] "mRNA processing"
## [52] "biological_process"
## [53] "RNA splicing"
## [54] "translational termination"
## [55] "polysaccharide catabolic process"
## [56] "septum digestion after cytokinesis"
## [57] "carbohydrate metabolic process"
## [58] "chitin catabolic process"
## [59] "metabolic process"
## [60] "cell wall organization"
## [61] "regulation of cell cycle"
## [62] "DNA damage checkpoint"
## [63] "telomere maintenance via recombination"
## [64] "telomere maintenance"
## [65] "double-strand break repair via homologous recombination"
## [66] "nucleotide-excision repair"
## [67] "chromatin silencing at telomere"
## [68] "cell cycle"
## [69] "intra-S DNA damage checkpoint"
## [70] "mitotic DNA replication checkpoint"
## [71] "meiotic DNA integrity checkpoint"
## [72] "translational initiation"
## [73] "regulation of translation"
## [74] "regulation of translational initiation"
## [75] "cellular metabolic process"
## [76] "regulation of catalytic activity"
## [77] "proteolysis"
## [78] "glutathione metabolic process"
## [79] "glutathione catabolic process"
## [80] "xenobiotic metabolic process"
## [81] "cellular glucan metabolic process"
## [82] "glucan catabolic process"
## [83] "fungal-type cell wall organization"
```

Look at the last group

```
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12 & chr.B==3],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim = c(7541374, 7744476)
)
abline(v=c(7541374-1500, 7584887+1500), col="green")
```
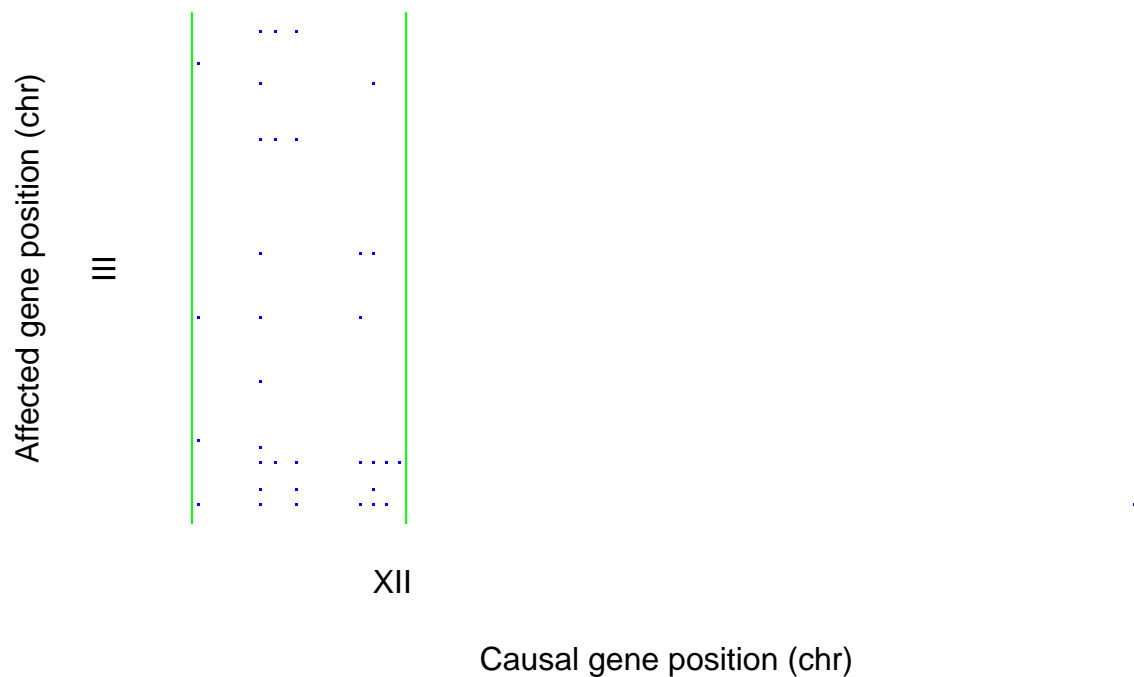

There's one extra gene that does not belong to the "band"

```
plot_sorted_coordinates(
  coordinates_plot_cor[chr.A == 12],
  separator = separator,
  col = coordinates_plot_cor[chr.A==12]$col,
  xlim = c(7541374, 7584887)
)
abline(v=c(7541374-1500, 7584887+1500), col="green")
```

Causal gene position (chr)
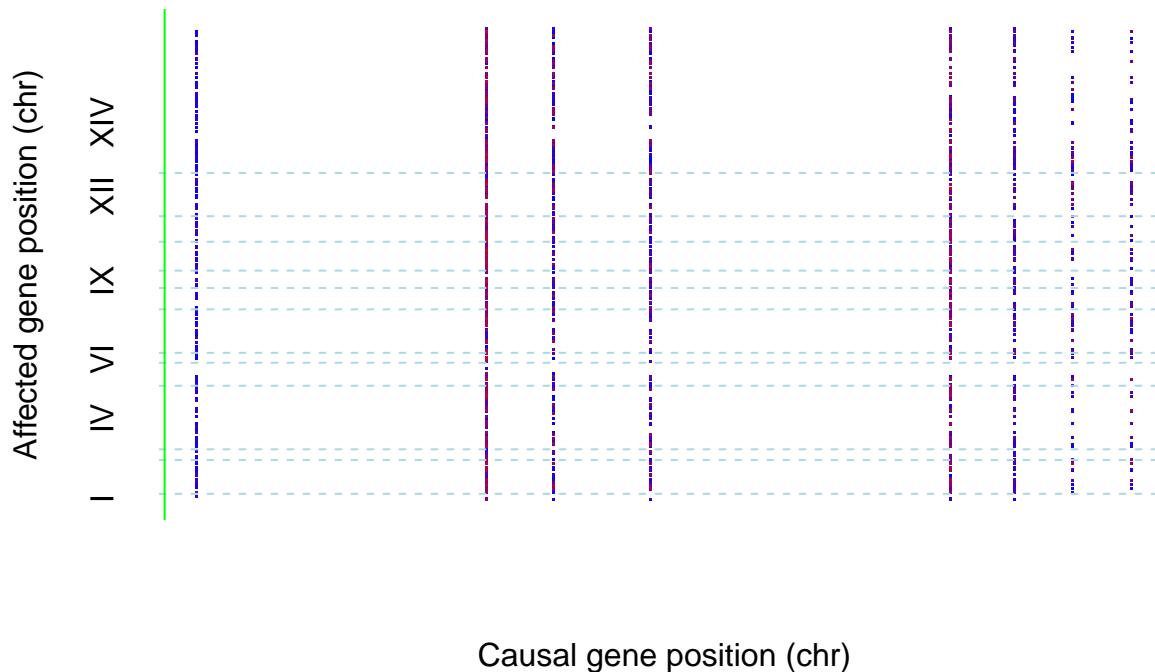
genes on the last band of chromsome 12

```
# gene names
unique(chr12_genenames[between(start.A,7541374-1500, 7584887+1500)]$gene.name)
```

```
## [1] "UBiquitin-Conjugating"
## [2] "AuTophaGy related"
## [3] "SPa2 Homolog"
## [4] "tRNA-specific Adenosine Deaminase"
## [5] "PEroXisome related"
## [6] NA
## [7] "Nicotinamide Mononucleotide Adenylyltransferase"
## [8] "CHitin Synthase-related"
```

```
# GO terms present
unique(chr12_genenames[between(start.A,7541374-1500, 7584887+1500)]$GO.term)
```

```
##  [1] "protein ubiquitination"
##  [2] "protein neddylation"
##  [3] "autophagy"
##  [4] "autophagy of nucleus"
##  [5] "reticulophagy"
##  [6] "conjugation"
##  [7] "bipolar cellular bud site selection"
##  [8] "pseudohyphal growth"
##  [9] "regulation of cell shape"
## [10] "mating projection formation"
## [11] "invasive filamentous growth"
## [12] "positive regulation of MAPK cascade"
## [13] "tRNA wobble adenosine to inosine editing"
## [14] "tRNA modification"
## [15] "tRNA processing"
## [16] "peroxisome organization"
```

```
## [17] "ER-dependent peroxisome organization"
## [18] "membrane tubulation"
## [19] "regulation of peroxisome organization"
## [20] "biological_process"
## [21] "biosynthetic process"
## [22] "NAD biosynthetic process"
## [23] "pyridine nucleotide biosynthetic process"
## [24] "cellular bud site selection"
## [25] "conjugation with cellular fusion"
## [26] "cell wall chitin catabolic process"
## [27] "regulation of transcription, DNA-templated"
## [28] "Golgi to plasma membrane transport"
## [29] "protein transport"
## [30] "ascospore wall assembly"
```