Chromosomal Instability Index Analysis

Lavinia Carabet

Chromosomal Instability Index (CIN) Analysis

CIN Analysis - Quantitative characterization of genome-wide copy number alterations as a measure of chromosomal instability at chromosome and cytoband levels between groups of patients

Case study: Colorectal Cancer (CRC)

Design: 38 Stage II CRC tissue samples, 15 with and 23 without relapse of tumor

Platform: Agilent Human Genome Comparative Genomic Hybridization (CGH) microarray aCGH - compares patients's genomes against a reference genome, identifies their differences and locates regions of genomic imbalances in patients

Dataset: GEO GSE17181 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17181

Clinical cohort selection

Prior to analysis clinical cohorts are selected based on the desired criteria, here having Event indicator for Disease Free Survival (EVENT_DFS) as outcome.

In other words, creating two groups:

- Relapse patients with Disease free survival as EVENT
- Relapse free patients with Disease free survival as CENSORING

```
clinical.dat <- read.csv('./CRC_BROSEN_2010_01_CLINICAL_DATA.csv', header = T)

relapse.group <- clinical.dat[clinical.dat$EVENT_DFS == "EVENT",]
relapse.group
```

```
##     SAMPLE_ID  AGE CANCER_DEATH DISEASE_FREE_SURVIVAL_MON DISEASE_LOCALIZATION
## 1  GSM426590 90.1          YES                      1.58         Sigmoid colon
## 2  GSM426578 70.9          YES                      2.11      Transverse colon
## 3  GSM426567 77.4          YES                      7.86                Coecum
## 4  GSM426561 90.6          YES                      9.11                Coecum
## 5  GSM426562 84.9           NO                     10.53      Transverse colon
## 6  GSM426581 73.9          YES                     13.72       Hepatic Flexure
## 7  GSM426565 78.8          YES                     14.51                Coecum
## 8  GSM426592 81.1          YES                     16.02                Coecum
## 9  GSM426582 65.2          YES                     16.18       Hepatic Flexure
## 10 GSM426570 77.2           NO                     21.02       Sigma and Rectum
## 11 GSM426560 73.5          YES                     21.32                Coecum
## 12 GSM426577 79.8          YES                     22.99        Ascending colon
## 13 GSM426571 63.3          YES                     36.55         Sigmoid colon
```

```
## 14 GSM426575 75.1          YES                    38.49              Coecum
## 15 GSM426591 63.8          YES                    59.21       Sigma and Rectum
##    EVENT_DFS EVENT_OS GENDER        HIST_TUMOR_GRADE OVERALL_SURVIVAL_MONTHS
## 1      EVENT    EVENT FEMALE      Well differentiated                     1.6
## 2      EVENT    EVENT   MALE    Poorly differentiated                     7.8
## 3      EVENT    EVENT   MALE Moderately differentiated                    10.7
## 4      EVENT    EVENT FEMALE    Poorly differentiated                    14.5
## 5      EVENT    EVENT FEMALE    Poorly differentiated                    17.3
## 6      EVENT    EVENT   MALE Moderately differentiated                    25.6
## 7      EVENT    EVENT   MALE Moderately differentiated                    15.8
## 8      EVENT    EVENT FEMALE Moderately differentiated                    16.0
## 9      EVENT    EVENT   MALE Moderately differentiated                    22.3
## 10     EVENT    EVENT FEMALE Moderately differentiated                    21.0
## 11     EVENT    EVENT FEMALE Moderately differentiated                    22.3
## 12     EVENT    EVENT   MALE Moderately differentiated                    32.0
## 13     EVENT    EVENT FEMALE Moderately differentiated                    37.2
## 14     EVENT    EVENT FEMALE Moderately differentiated                    45.9
## 15     EVENT    EVENT FEMALE Moderately differentiated                    59.2
##                                     PTNM_T TUMOR_STAGE PATIENT_ID
## 1  T3 - Size and/or extent of the primary tumor    Stage II          8
## 2  T3 - Size and/or extent of the primary tumor    Stage II         34
## 3  T3 - Size and/or extent of the primary tumor    Stage II         21
## 4  T3 - Size and/or extent of the primary tumor    Stage II         14
## 5  T3 - Size and/or extent of the primary tumor    Stage II         15
## 6  T2 - Size and/or extent of the primary tumor    Stage II         39
## 7  T3 - Size and/or extent of the primary tumor    Stage II         18
## 8  T3 - Size and/or extent of the primary tumor    Stage II         10
## 9  T3 - Size and/or extent of the primary tumor    Stage II         40
## 10 T3 - Size and/or extent of the primary tumor    Stage II         26
## 11 T3 - Size and/or extent of the primary tumor    Stage II         13
## 12 T3 - Size and/or extent of the primary tumor    Stage II         33
## 13 T3 - Size and/or extent of the primary tumor    Stage II         27
## 14 T3 - Size and/or extent of the primary tumor    Stage II         31
## 15 T3 - Size and/or extent of the primary tumor    Stage II          9
```

```r
relapse.free.group <- clinical.dat[clinical.dat$EVENT_DFS == 'CENSORING',]
relapse.free.group
```

```
##      SAMPLE_ID  AGE CANCER_DEATH DISEASE_FREE_SURVIVAL_MON DISEASE_LOCALIZATION
## 16 GSM426586 81.5           NO                    155.00         Sigmoid colon
## 17 GSM426585 70.7           NO                    143.45      Sigma and Rectum
## 18 GSM426584 48.5           NO                    143.49                Coecum
## 19 GSM426583 73.4           NO                    151.12        Ascending colon
## 20 GSM426580 55.1           NO                     55.66                Coecum
## 21 GSM426579 54.3           NO                    148.65        Hepatic Flexure
## 22 GSM426576 89.9          YES                      0.69                Coecum
## 23 GSM426574 85.9           NO                    109.08      Descending Colon
## 24 GSM426573 75.6           NO                      0.53         Sigmoid colon
## 25 GSM426572 77.0           NO                    131.68        Hepatic Flexure
## 26 GSM426569 74.2           NO                    121.84                Coecum
## 27 GSM426568 62.1           NO                     66.48         Sigmoid colon
## 28 GSM426566 58.3           NO                     90.79      Sigma and Rectum
## 29 GSM426564 80.4           NO                    133.42        Ascending colon
## 30 GSM426563 58.9           NO                     61.05      Transverse colon
```

```
## 31 GSM426594 69.5         NO          119.05    Left Flexure
## 32 GSM426593 81.9         NO           81.81           Coecum
## 33 GSM426589 69.9         NO          124.87  Sigma and Rectum
## 34 GSM426588 69.1         NO          123.78     Sigmoid colon
## 35 GSM426587 69.7         NO           10.46      Left Flexure
## 36 GSM426559 79.8         NO           55.43           Coecum
## 37 GSM426558 72.3         NO          116.02           Coecum
## 38 GSM426557 74.1         NO           38.78  Transverse colon
##     EVENT_DFS  EVENT_OS GENDER       HIST_TUMOR_GRADE OVERALL_SURVIVAL_MONTHS
## 16 CENSORING CENSORING   MALE Moderately differentiated                 155.0
## 17 CENSORING CENSORING   MALE Moderately differentiated                 143.5
## 18 CENSORING CENSORING   MALE Moderately differentiated                 143.5
## 19 CENSORING CENSORING FEMALE       Well differentiated                 151.1
## 20 CENSORING CENSORING   MALE Moderately differentiated                  55.7
## 21 CENSORING CENSORING FEMALE       Well differentiated                 148.7
## 22 CENSORING     EVENT FEMALE Moderately differentiated                   0.7
## 23 CENSORING     EVENT   MALE Moderately differentiated                 109.1
## 24 CENSORING     EVENT   MALE Moderately differentiated                   0.5
## 25 CENSORING CENSORING FEMALE Moderately differentiated                 131.7
## 26 CENSORING CENSORING FEMALE Moderately differentiated                 121.8
## 27 CENSORING CENSORING FEMALE Moderately differentiated                  66.5
## 28 CENSORING CENSORING   MALE Moderately differentiated                  90.8
## 29 CENSORING CENSORING FEMALE     Poorly differentiated                 133.4
## 30 CENSORING CENSORING   MALE Moderately differentiated                  61.1
## 31 CENSORING CENSORING   MALE Moderately differentiated                 119.0
## 32 CENSORING     EVENT FEMALE Moderately differentiated                  81.8
## 33 CENSORING CENSORING   MALE Moderately differentiated                 124.9
## 34 CENSORING     EVENT FEMALE Moderately differentiated                 123.8
## 35 CENSORING     EVENT FEMALE Moderately differentiated                  10.5
## 36 CENSORING CENSORING FEMALE Moderately differentiated                  55.4
## 37 CENSORING CENSORING   MALE Moderately differentiated                 116.0
## 38 CENSORING CENSORING   MALE     Poorly differentiated                  38.8
##                                      PTNM_T TUMOR_STAGE PATIENT_ID
## 16 T3 - Size and/or extent of the primary tumor    Stage II         46
## 17 T3 - Size and/or extent of the primary tumor    Stage II         44
## 18 T3 - Size and/or extent of the primary tumor    Stage II         43
## 19 T3 - Size and/or extent of the primary tumor    Stage II         42
## 20 T3 - Size and/or extent of the primary tumor    Stage II         36
## 21 T2 - Size and/or extent of the primary tumor    Stage II         35
## 22 T3 - Size and/or extent of the primary tumor    Stage II         32
## 23 T3 - Size and/or extent of the primary tumor    Stage II         30
## 24 T3 - Size and/or extent of the primary tumor    Stage II         29
## 25 T4 - Size and/or extent of the primary tumor    Stage II         28
## 26 T2 - Size and/or extent of the primary tumor    Stage II         25
## 27 T2 - Size and/or extent of the primary tumor    Stage II         24
## 28 T3 - Size and/or extent of the primary tumor    Stage II         20
## 29 T3 - Size and/or extent of the primary tumor    Stage II         17
## 30 T3 - Size and/or extent of the primary tumor    Stage II         16
## 31 T3 - Size and/or extent of the primary tumor    Stage II         12
## 32 T3 - Size and/or extent of the primary tumor    Stage II         11
## 33 T3 - Size and/or extent of the primary tumor    Stage II          7
## 34 T3 - Size and/or extent of the primary tumor    Stage II          6
## 35 T4 - Size and/or extent of the primary tumor    Stage II          5
## 36 T3 - Size and/or extent of the primary tumor    Stage II          4
```

```
## 37 T3 - Size and/or extent of the primary tumor      Stage II          2
## 38 T3 - Size and/or extent of the primary tumor      Stage II          1
```

Load chromosome-level CIN data matrix for analysis

```r
load("CRC_BROSENS_2010_01_CIN_CHROMOSOMES.Rda")

# dataMatrix - chromosome-level CIN matrix, each row is a chromosome, each column is a sample(patient)
dataMatrix[,1:3]
```

```
##          GSM426590    GSM426591    GSM426592
## 1        0.0000000    0.0000000  111.0234246
## 2        3.5694628    0.0000000    0.0000000
## 3        0.2902253    0.0000000    0.0000000
## 4        1.4771780    0.0000000    0.0000000
## 5        1.5882121    1.7923942   62.1806763
## 6        4.0541977  185.2552754    0.0000000
## 7      545.6952113  210.3718010    0.0000000
## 8        5.1645600  120.1230330   50.1795694
## 9      136.8693085    0.0000000    0.0000000
## 10       0.0000000    0.5088373   66.9445609
## 11       0.6399465    0.0000000    0.0000000
## 12       0.0000000    0.2573348    0.0000000
## 13       0.3501021    2.3154180    0.0000000
## 14       0.0000000    0.0000000    0.0000000
## 15      58.4126823    0.0000000    0.0000000
## 16       0.0000000    1.0009599   45.0412180
## 17       0.0000000    0.3092296    0.0000000
## 18      98.8508383    1.5562982    0.0000000
## 19     122.3744985    0.4671531    0.0000000
## 20      16.8022999  251.7388064    0.3098023
## 21       0.0000000    0.2926180    0.0000000
## 22       0.0000000    0.0000000    0.0000000
```

```r
allids <- dimnames(dataMatrix)[[2]]

# relapse group IDs
grp1ids <-  allids[1:15]
# relapse free group IDs
grp2ids <- allids[16:38]

sampleGrp1Len <- length(grp1ids)
sampleGrp2Len <- length(grp2ids)

## clinical.inf: n*2 matrix, the 1st column is 'sample name', the second is 'label'
clinical.inf <- as.matrix( cbind( dimnames(dataMatrix)[[2]],
                                  c( rep('Relapse Free', sampleGrp2Len),
                                     rep('Relapse',sampleGrp1Len ) ) ) )
clinical.inf
```

```
##          [,1]          [,2]
```

```
##  [1,] "GSM426590" "Relapse Free"
##  [2,] "GSM426591" "Relapse Free"
##  [3,] "GSM426592" "Relapse Free"
##  [4,] "GSM426560" "Relapse Free"
##  [5,] "GSM426561" "Relapse Free"
##  [6,] "GSM426562" "Relapse Free"
##  [7,] "GSM426565" "Relapse Free"
##  [8,] "GSM426567" "Relapse Free"
##  [9,] "GSM426570" "Relapse Free"
## [10,] "GSM426571" "Relapse Free"
## [11,] "GSM426575" "Relapse Free"
## [12,] "GSM426577" "Relapse Free"
## [13,] "GSM426578" "Relapse Free"
## [14,] "GSM426581" "Relapse Free"
## [15,] "GSM426582" "Relapse Free"
## [16,] "GSM426557" "Relapse Free"
## [17,] "GSM426558" "Relapse Free"
## [18,] "GSM426559" "Relapse Free"
## [19,] "GSM426587" "Relapse Free"
## [20,] "GSM426588" "Relapse Free"
## [21,] "GSM426589" "Relapse Free"
## [22,] "GSM426593" "Relapse Free"
## [23,] "GSM426594" "Relapse Free"
## [24,] "GSM426563" "Relapse"
## [25,] "GSM426564" "Relapse"
## [26,] "GSM426566" "Relapse"
## [27,] "GSM426568" "Relapse"
## [28,] "GSM426569" "Relapse"
## [29,] "GSM426572" "Relapse"
## [30,] "GSM426573" "Relapse"
## [31,] "GSM426574" "Relapse"
## [32,] "GSM426576" "Relapse"
## [33,] "GSM426579" "Relapse"
## [34,] "GSM426580" "Relapse"
## [35,] "GSM426583" "Relapse"
## [36,] "GSM426584" "Relapse"
## [37,] "GSM426585" "Relapse"
## [38,] "GSM426586" "Relapse"
```

CIN Chromosome level analysis

```
source("getSubmatrix.R")
cinInputMatrix <- getSubmatrix.twogrps(dataMatrix, grp2ids, grp1ids)

source("heatmap.draw.R")
heatmap.draw(cinInputMatrix, clinical.inf)
```

heatmap chromosome CIN
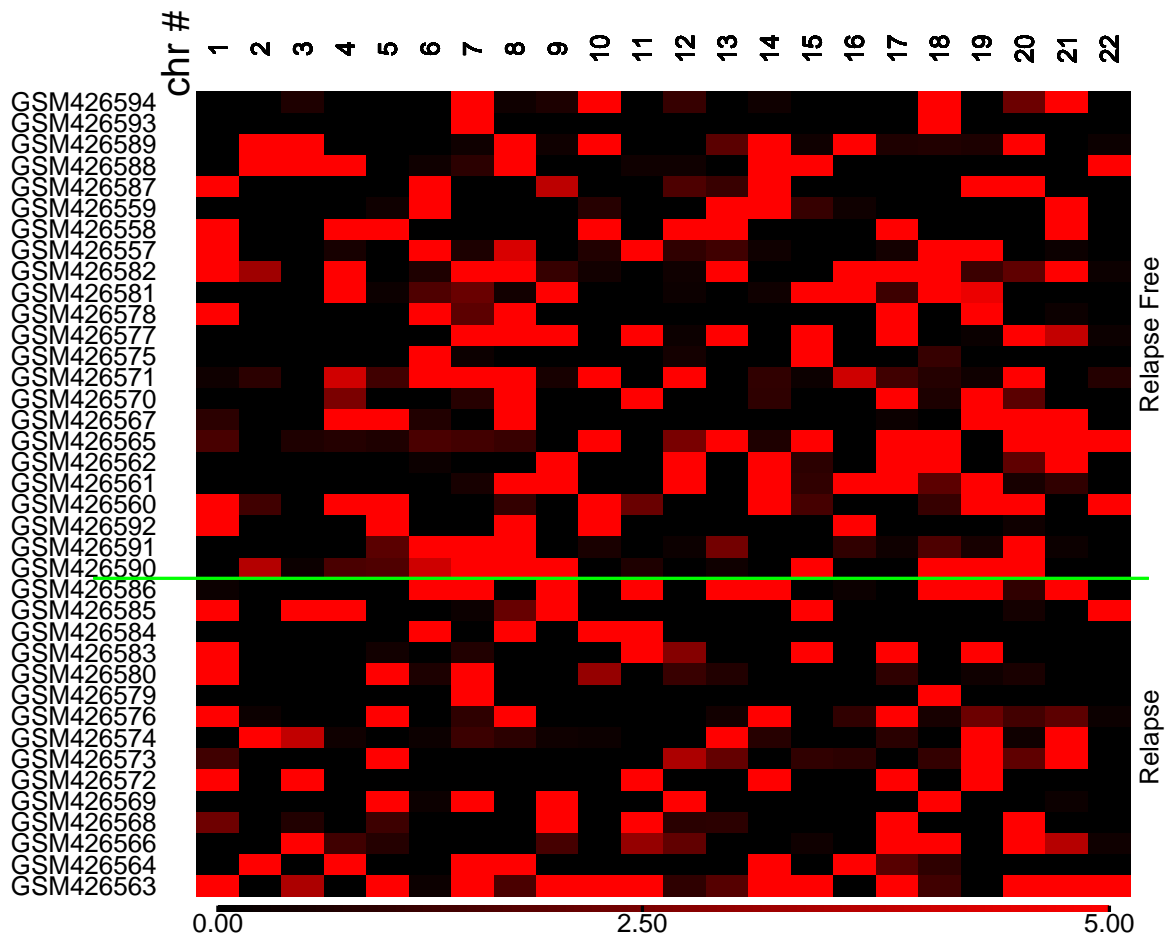
CIN Cytoband level analysis

```
load("CRC_BROSENS_2010_01_CIN_CYTOBANDS.Rda")

# dataMatrix - cytoband-level CIN matrix, each row is a cytoband, each column is a sample(patient)
head(dataMatrix[,1:3],20)
```

```
##          GSM426590 GSM426591 GSM426592
## 1p36.33          0         0 0.7079186
## 1p36.32          0         0 2.5380271
## 1p36.31          0         0 0.8655009
```

```
## 1p36.23           0           0 0.6251955
## 1p36.22           0           0 1.5305669
## 1p36.21           0           0 2.9463504
## 1p36.13           0           0 2.0100286
## 1p36.12           0           0 2.6730605
## 1p36.11           0           0 2.4018286
## 1p35.3            0           0 1.5246383
## 1p35.2            0           0 0.0000000
## 1p35.1            0           0 0.9812331
## 1p34.3            0           0 4.3940856
## 1p34.2            0           0 2.4563252
## 1p34.1            0           0 1.0316332
## 1p33             0           0 2.7105703
## 1p32.3            0           0 1.1876797
## 1p32.2            0           0 0.3472964
## 1p32.1            0           0 1.4151507
## 1p31.3            0           0 3.6456482
```

```
tail(dataMatrix[,1:3],20)
```

```
##            GSM426590 GSM426591 GSM426592
## 21q22.12           0  0.292618           0
## 21q22.13           0  0.292618           0
## 21q22.2            0  0.292618           0
## 21q22.3            0  0.292618           0
## 22p13             0  0.000000           0
## 22p12             0  0.000000           0
## 22p11.2            0  0.000000           0
## 22p11.1            0  0.000000           0
## 22q11.1            0  0.000000           0
## 22q11.21           0  0.000000           0
## 22q11.22           0  0.000000           0
## 22q11.23           0  0.000000           0
## 22q12.1            0  0.000000           0
## 22q12.2            0  0.000000           0
## 22q12.3            0  0.000000           0
## 22q13.1            0  0.000000           0
## 22q13.2            0  0.000000           0
## 22q13.31           0  0.000000           0
## 22q13.32           0  0.000000           0
## 22q13.33           0  0.000000           0
```

```
allids <- dimnames(dataMatrix)[[2]]

grp1ids <-  allids[1:15]
grp2ids <- allids[16:38]

sampleGrp1Len <- length(grp1ids)
sampleGrp2Len <- length(grp2ids)

## clinical.inf: n*2 matrix, the 1st column is 'sample name', the second is 'label'
clinical.inf <- as.matrix( cbind( dimnames(cinInputMatrix)[[2]],
                                   c( rep('Relapse Free', sampleGrp2Len),
```

```
                                rep('Relapse',sampleGrp1Len ) ) ) )
clinical.inf
```

```
##         [,1]        [,2]
##  [1,] "GSM426557" "Relapse Free"
##  [2,] "GSM426558" "Relapse Free"
##  [3,] "GSM426559" "Relapse Free"
##  [4,] "GSM426587" "Relapse Free"
##  [5,] "GSM426588" "Relapse Free"
##  [6,] "GSM426589" "Relapse Free"
##  [7,] "GSM426593" "Relapse Free"
##  [8,] "GSM426594" "Relapse Free"
##  [9,] "GSM426563" "Relapse Free"
## [10,] "GSM426564" "Relapse Free"
## [11,] "GSM426566" "Relapse Free"
## [12,] "GSM426568" "Relapse Free"
## [13,] "GSM426569" "Relapse Free"
## [14,] "GSM426572" "Relapse Free"
## [15,] "GSM426573" "Relapse Free"
## [16,] "GSM426574" "Relapse Free"
## [17,] "GSM426576" "Relapse Free"
## [18,] "GSM426579" "Relapse Free"
## [19,] "GSM426580" "Relapse Free"
## [20,] "GSM426583" "Relapse Free"
## [21,] "GSM426584" "Relapse Free"
## [22,] "GSM426585" "Relapse Free"
## [23,] "GSM426586" "Relapse Free"
## [24,] "GSM426590" "Relapse"
## [25,] "GSM426591" "Relapse"
## [26,] "GSM426592" "Relapse"
## [27,] "GSM426560" "Relapse"
## [28,] "GSM426561" "Relapse"
## [29,] "GSM426562" "Relapse"
## [30,] "GSM426565" "Relapse"
## [31,] "GSM426567" "Relapse"
## [32,] "GSM426570" "Relapse"
## [33,] "GSM426571" "Relapse"
## [34,] "GSM426575" "Relapse"
## [35,] "GSM426577" "Relapse"
## [36,] "GSM426578" "Relapse"
## [37,] "GSM426581" "Relapse"
## [38,] "GSM426582" "Relapse"
```

```
cinInputMatrix <- getSubmatrix.twogrps(dataMatrix, grp2ids, grp1ids)

load("hg18_annot.Rda")                # load reference genome, human genome 18 data frame
annotInputMatrix <- hg18_annot        # each row corresponds to a cytoband
head(annotInputMatrix,5)              # columns: "chrom", "start", "end", "name", and "stain"
```

```
##   chrom   start      end   name   stain
## 1  chr1       0  2300000 p36.33    gneg
## 2  chr1 2300000  5300000 p36.32  gpos25
## 3  chr1 5300000  7100000 p36.31    gneg
```
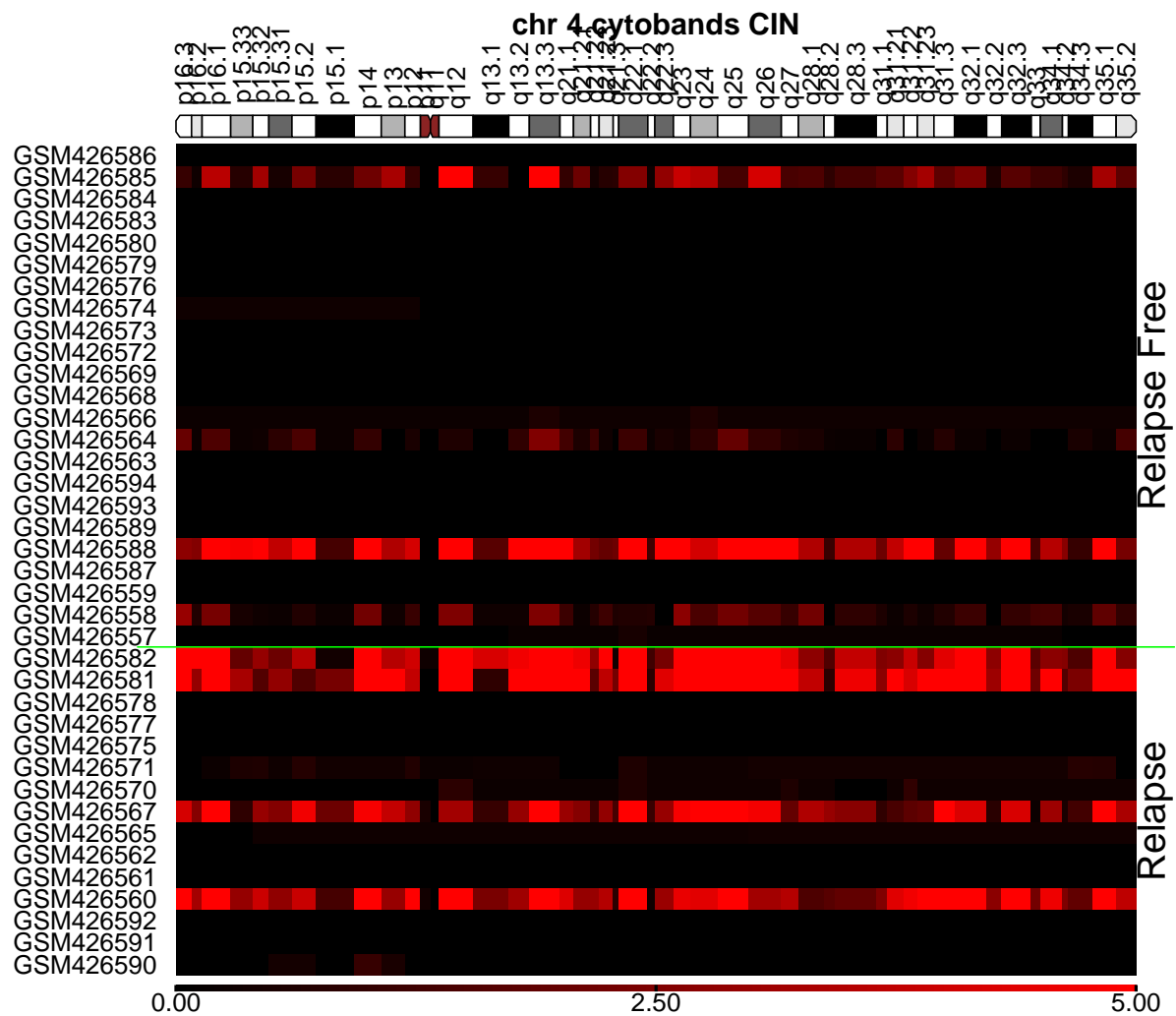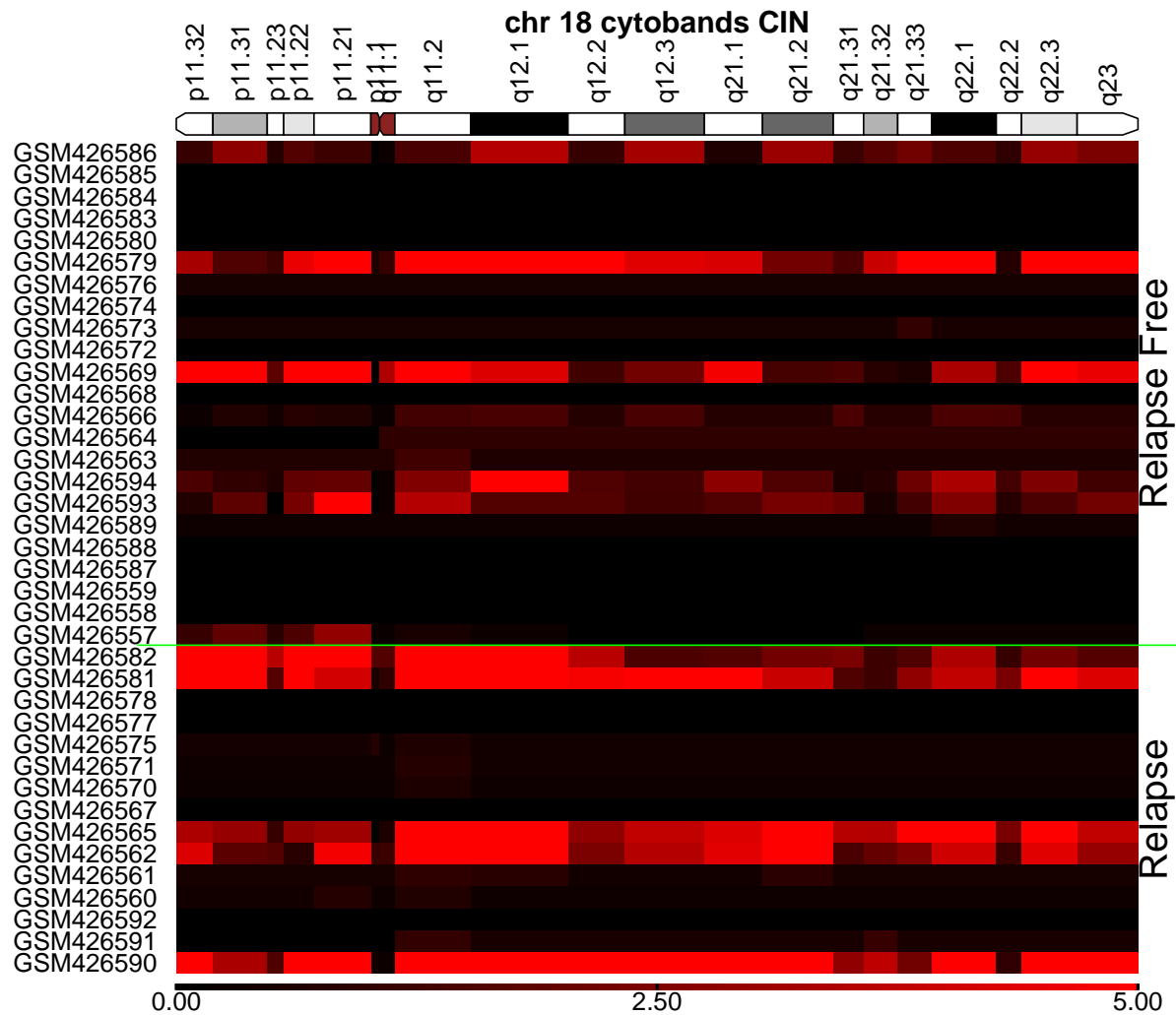
```
## 4   chr1 7100000   9200000 p36.23 gpos25
## 5   chr1 9200000 12600000 p36.22    gneg
```

```
source("cytobands_cin.draw.R")
#for (i in 1:22) {
for (chr in c(4,18)) {
  cytobands_cin.draw(cinInputMatrix, clinical.inf, chr, annotInputMatrix,
                     title_text = paste('chromosome',chr,'cytobands CIN overview'))
}
```

```
## Loading required package: bitops
```

chr 18 cytobands CIN

Above are the results of CIN analysis showing cumulative instabilities (losses and gains) at chromosome and cytoband levels with more instabilities in the relapsed group compared to relapse-free in chromosomes 4 and 18

Downstream T-test Cytoband-level Differential Analysis

CRC_BROSENS_2010_01_CIN_CYTOBANDS.Rda cytoband CIN data and groups above

```
# function to calculate Student's two-sample t-test on all cytobands
# function returns the p-value for the test
# NAs are removed for each test

t.test.all.cyt <- function(x,s1,s2) {
```

```
    x1 <- x[s1]
    x2 <- x[s2]
    x1 <- as.numeric(x1)
    x2 <- as.numeric(x2)
    t.out <- t.test(x1, x2, alternative='two.sided', var.equal = TRUE)
    out <- as.numeric(t.out$p.value)
    return(out)
}

t.test.run <- apply(cinInputMatrix, 1, t.test.all.cyt, s1=grp1ids, s2=grp2ids)

# calculate means of the groups
grp1.mean <- apply(cinInputMatrix[,grp1ids], 1, mean, na.rm = T)
grp2.mean <- apply(cinInputMatrix[,grp2ids], 1, mean, na.rm = T)

# calculate fold change
fold.change <- grp1.mean - grp2.mean
#range(fold.change)

#Build data frame with p-value and fold information
ds.featured.cyt <- cbind(t.test.run, fold.change)

# Select cytobands with significance alpha=0.05 and fold.change = 1
featured.sel.cyt <- (t.test.run < 0.05) & ((fold.change < -log2(1)) | (fold.change > log2(1)))

#Filters data frame with featured selection cytobands found
diff.changed.cyt <- ds.featured.cyt[featured.sel.cyt,]

#Orders data frame by fold change
diff.changed.cyt <- diff.changed.cyt[order(diff.changed.cyt[,2], decreasing = T),]
diff.changed.cyt
```

```
##          t.test.run fold.change
## 4q35.1   0.04676346   1.8600425
## 4q31.3   0.02349399   1.5587130
## 18q21.2  0.03646747   1.4041709
## 4q21.23  0.03601223   0.8813700
## 4q34.3   0.02179689   0.4522813
## 4q28.2   0.04116379   0.4266848
## 7q11.1   0.02660328   0.3730372
## 18p11.1  0.03432688   0.2485463
## 7p11.1   0.03929041   0.2259242
## 4p11     0.01196546   0.1118742
## 3q27.2   0.03152071  -0.1457290
## 3q12.3   0.04296355  -0.2730306
## 3q13.11  0.04094549  -0.3431846
## 3p12.3   0.04953014  -0.5410803
```

The follow-up T-test differential analysis at cytoband-level showed that the most differentially changed cytobands are in chromosomes 4 and 18 (as well as 3), correlating with results from the literature.

Chromosomal instability in 4q arm (4q22.2-35.2) has been previously reported to be associated with local recurrence in colon cancer and worst outcomes in stage II CRC patients.

In addition, the cytoband region 18q21.2 contains Deleted in Colorectal Carcinoma (DCC) gene where frequent loss of heterozygosity (LOH) events in colon cancer occur, as per OMIM (Online Mendelian Inheritance in man).