Profiles of genome complexity identify HORMAD1 as a cause of homologous recombination deficiency in triple-negative breast cancers

Scores of Chromosomal Instability Scarring (SCINS) in TNBC

Johnathan Watkins

****

ABSTRACT Triple-negative breast cancers (TNBCs) are characterised by a wide spectrum of genomic alterations, some of which might be caused by well-studied defects in DNA repair processes such as homologous recombination (HR). Despite this understanding, associating particular patterns of genomic instability with the response to therapy has been challenging. Here, we show that Allelic-imbalanced Copy Number Aberrations (AiCNA) are more prevalent in TNBCs that respond to platinum-based chemotherapy, thus providing a candidate predictive biomarker for this disease. Furthermore, we show that a high level of AiCNA is linked with elevated expression of a meiosis-associated gene HORMAD1. Elevated HORMAD1 expression suppresses RAD51-dependent HR, and drives the use of alternative forms of DNA repair, generation of AiCNAs as well as sensitizing cancer cells to HR targeting therapies. This data thus provides a mechanistic association between HORMAD1 expression, a pattern of genomic instability and its association with response to platinum-based chemotherapy in TNBC.

****

1. FUNCTIONS

```
> #----- CATEGORISATION OF ABERRATIONS -----#
>
>     categoriseSegmentsPerSample <- function(data.f="TNBC111", filenames,
+                                             save=T, output.f,
+                                        pattern=".segmentCN.*",
+                                        pheno.f="phenotypes.BCCL39.txt",
+                                        homd=F) {
+
+     pheno <- read.delim(paste("materials/tables/", pheno.f, sep=""),
+                         sep="\t", header=T, row.names=1, stringsAsFactors=F)
+
+     sampleNames <- gsub(pattern,"",filenames)
+
+     if (!(identical(row.names(pheno), sampleNames) |
```

```
+              identical(pheno[,1], sampleNames) ) )
+         stop("Reconfigure the order of the samples in the phenotypes file")
+
+     sample.scar.list <- lapply(filenames, function(filename) {
+
+         if (is.element(filename, list.files(paste("materials/tables",
+                                                   data.f, "TAPS",
+                                                   sep="/" ) )) ) {
+
+             sampleData <- read.table(paste("materials/tables", data.f,
+                                            "TAPS", filename, sep="/"),
+                                      header=T, sep="\t", stringsAsFactors=F)
+         } else {
+             cat(paste("Missing ", filename, "...", sep=""))
+         }
+
+         cat(gsub(".segmentCN.*","",filename), "is being processed", "\n")
+         sampleData$Chromosome <- gsub("chr","",sampleData$Chromosome)
+         sampleData$Chromosome <- gsub("X",23,sampleData$Chromosome)
+         sampleData <- sampleData[which(sampleData$Chromosome!="Y"),]
+         if (identical(data.f, "BCCL39")) {
+             sampleID <- rep( pheno[ row.names(pheno)%in% gsub(".segmentCN.*",
+                                                               "",filename),
+                                     "CommonName" ], nrow(sampleData) )
+         } else {
+             sampleID <- rep( gsub(".segmentCN.*","",filename) ,
+                              nrow(sampleData) )
+         }
+
+         alleleRatio <- sampleData$Cn/sampleData$mCn
+         copyNum <- sampleData$Cn
+
+         Ai <-
+             ifelse( is.na(alleleRatio) | alleleRatio == 2, 0, 1 )
+         AiCna <-
+             ifelse( is.na(alleleRatio) | alleleRatio == 2 | copyNum == 2,
+                     0, 1 )
+         AiGain <-
+             ifelse( is.na(alleleRatio) | alleleRatio == 2 | copyNum <= 2,
+                     0, 1 )
+         AiLoss <-
+             ifelse( is.na(alleleRatio) | alleleRatio == 2 | copyNum >= 2,
+                     0, 1 )
+
+         if(homd){
```

```
+             AbCna <- ifelse(is.na(alleleRatio) & copyNum != 0, 0,
+                             ifelse( copyNum == 0 | (alleleRatio == 2 &
+                                                       copyNum != 2),
+                                     1, 0 ))
+         } else {
+             AbCna <- ifelse(is.na(alleleRatio) & copyNum != 0, 0,
+                             ifelse( (alleleRatio == 2 &
+                                          copyNum > 2), 1, 0 ))
+         }
+
+         AbGain <-
+             ifelse(is.na(alleleRatio) & copyNum != 0, 0,
+                    ifelse( alleleRatio == 2 & copyNum > 2, 1, 0 ))
+         AbLoss <- 0
+         CnLoH <-
+             ifelse(is.na(alleleRatio) | alleleRatio == 2 | copyNum != 2,
+                    0, 1 )
+
+         return(cbind(sampleID, sampleData[,1:5], copyNum, Ai,
+                      AiCna, AiGain, AiLoss, AbCna, AbGain,
+                      AbLoss, CnLoH))#, DipHet))
+
+     })
+
+     sample.scar.df <- sample.scar.list[[1]]
+
+     for (sample in sample.scar.list[2:length(sample.scar.list)]){
+         sample.scar.df <- rbind(sample.scar.df, sample)
+     }
+
+     if (save) save(sample.scar.df, file=paste("results/objects",
+                                               output.f, sep="/"))
+     sample.scar.df
+ }
> #----- MEASUREMENT OF SCINS -----#
>
> measureSCINS <- function( AllAber.sam.l, chromInfo, save=T, dataset ) {
+
+     SCINS.scores <- matrix(rep(NA, 5*length(AllAber.sam.l)),
+                            nrow=length(AllAber.sam.l), ncol=5 )
+     rownames(SCINS.scores) <-
+         sapply(AllAber.sam.l, function(s) s$sampleID[[1]])
+     colnames(SCINS.scores) <-
+         c("AiCna", "AbCna", "CnLoH", "Ai", "Cna")
+     genome.len <- sum(as.numeric(chromInfo$length))/1e6
```

```
+
+     for (i in 1:length(AllAber.sam.l) ) {
+
+         sample <- AllAber.sam.l[[i]]
+
+         # calculates product of proportion of genome with aiCNA and
+         # num of segments with len >=8Mb but less than whole chr
+         SCINS.scores[i,"AiCna"] <-
+             (sum(sample[which(as.numeric(sample$AiCna)==1 &
+                                as.numeric(sample$probes) >= 0 &
+                                !(as.numeric(sample$whChrAber)==1) ),
+                         "lengthMB"] )/genome.len)*
+             (nrow(sample[which(as.numeric(sample$AiCna)==1 &
+                                 as.numeric(sample$lengthMB) >= 8 &
+                                 !(as.numeric(sample$whChrAber)==1) ),] ))
+
+         # calculates # of segments with abCNA and with
+         # len >=8Mb including whole chr
+         SCINS.scores[i,"AbCna"] <-
+             (nrow(sample[which( (as.numeric(sample$AbCna)==1 &
+                                   as.numeric(sample$lengthMB) >= 8) |
+                                  as.numeric(sample$whChrAber)==1 ) ,] ))
+
+         # calculates product of proportion of genome with cnLOH and
+         # num of segments with len >=4Mb
+         SCINS.scores[i,"CnLoH"] <-
+             (sum(sample[which(as.numeric(sample$CnLoH)==1 &
+                                as.numeric(sample$lengthMB) >= 0 &
+                                !(as.numeric(sample$whChrAber)==1) ),
+                         "lengthMB"] )/genome.len)*
+             nrow(sample[which(as.numeric(sample$CnLoH)==1 &
+                                as.numeric(sample$lengthMB) >= 4 &
+                                !(as.numeric(sample$whChrAber)==1) ),] )
+
+         SCINS.scores[i,"Ai"] <-
+             SCINS.scores[i,"AiCna"] + SCINS.scores[i,"CnLoH"]
+         SCINS.scores[i,"Cna"] <-
+             SCINS.scores[i,"AiCna"] + SCINS.scores[i,"AbCna"]
+     }
+
+     rm(sample)
+
+     if (save) save(AllAber.sam.l,
+                    file=paste("results/objects/",dataset,
+                               ".AllAber.sam.RData",sep="") )
```

```
+
+     SCINS.scores
+ }
```

2. EXAMPLE USAGE

```
> #----- KCL111: DETERMINATION OF SAMPLES TO ANALYSE -----#
>
> filenames.TNBC111 <- list.files("materials/tables/TNBC111/TAPS")
> samples.TNBC111 <- gsub("(?i)_segmentCN.txt", "", filenames.TNBC111)
> refs <- read.delim("materials/tables/phenotypes.TNBC111.txt", header=T)
> refs <- gsub("[_.-]","", refs$SampleID)
> samples.TNBC111 <- samples.TNBC111[ gsub("[_.-]", "",
+                                          samples.TNBC111) %in% refs]
> filenames.TNBC111 <- gsub("$","_segmentCN.txt",samples.TNBC111)
> ##### KCL111: CATEGORISATIO OF SEGMENTS BY ABERRATION TYPE #####
>
> sample.scar.df.TNBC111 <-
+     categoriseSegmentsPerSample(data.f="TNBC111", filenames=filenames.TNBC111,
+                                 save=T, output.f="TNBC111.sampleScars.RData",
+                                 pheno.f="phenotypes.TNBC111.txt")
```

```
MAR_A3_20101029 is being processed
MAR_A3_20110315 is being processed
MAR_B1_20110308 is being processed
MAR_B2_20101026 is being processed
MAR_B3_20101029 is being processed
MAR_B3_20110315 is being processed
MAR_B4_20101029 is being processed
MAR_B4_20110315 is being processed
MAR_B5_20110315 is being processed
MAR_C1_20101029 is being processed
MAR_C1_20110308 is being processed
MAR_C2_20101026 is being processed
MAR_C4_20101029 is being processed
MAR_D1_20110308 is being processed
MAR_D2_20101026 is being processed
MAR_D3_20101029 is being processed
MAR_D4_20101029 is being processed
MAR_D4_20110315 is being processed
MAR_D5_20110315 is being processed
MAR_E1_20110308 is being processed
MAR_E2_20101026 is being processed
MAR_E3_20110315 is being processed
MAR_E4_20101029 is being processed
```

```
MAR_E5_20110315 is being processed
MAR_F2_20101026 is being processed
MAR_F4_20101029 is being processed
MAR_F5_20110311 is being processed
MAR_G2_20110315 is being processed
MAR_G3_20101029 is being processed
MAR_G3_20110315 is being processed
MAR_H1_20110308 is being processed
MAR_H2_20110315 is being processed
MAR_H3_20101029 is being processed
MAR_H4_20101029 is being processed
MAR_H4_20110315 is being processed
MAR_P1_A1_20091104 is being processed
MAR_P1_A2_20091210 is being processed
MAR_P1_A3_20091106 is being processed
MAR_P1_A4_20100702 is being processed
MAR_P1_A6_20091106 is being processed
MAR_P1_B1_20091104 is being processed
MAR_P1_B2_20091106 is being processed
MAR_P1_B2_20091210 is being processed
MAR_P1_B2_20100629 is being processed
MAR_P1_B3_20100702 is being processed
MAR_P1_B6_20091106 is being processed
MAR_P1_C1_20091104 is being processed
MAR_P1_C1_20091210 is being processed
MAR_P1_C2_20091106 is being processed
MAR_P1_C2_20091210 is being processed
MAR_P1_C3_20100702 is being processed
MAR_P1_C4_20100702 is being processed
MAR_P1_C5_20091106 is being processed
MAR_P1_D1_20091104 is being processed
MAR_P1_D2_20091106 is being processed
MAR_P1_D5_20091106 is being processed
MAR_P1_D6_20091106 is being processed
MAR_P1_D6_20100702 is being processed
MAR_P1_E1_20091210 is being processed
MAR_P1_E2_20091210 is being processed
MAR_P1_E2_20100629 is being processed
MAR_P1_E3_20100702 is being processed
MAR_P1_E4_20091106 is being processed
MAR_P1_E5_20091106 is being processed
MAR_P1_E6_20091106 is being processed
MAR_P1_F1_20091210 is being processed
MAR_P1_F2_20091210 is being processed
MAR_P1_F2_20100629 is being processed
```

```
MAR_P1_F4_20100702 is being processed
MAR_P1_F6_20091106 is being processed
MAR_P1_G1_20091210 is being processed
MAR_P1_G1_20100624 is being processed
MAR_P1_G2_20091210 is being processed
MAR_P1_G2_20100629 is being processed
MAR_P1_G4_20091106 is being processed
MAR_P1_G5_20100702 is being processed
MAR_P1_H1_20091104 is being processed
MAR_P1_H1_20091210 is being processed
MAR_P1_H2_20100629 is being processed
MAR_P1_H3_20100702 is being processed
MAR_P1_H4_20091106 is being processed
MAR_P1_H4_20100702 is being processed
MAR_P2_A1_20091106 is being processed
MAR_P2_A2_20091104 is being processed
MAR_P2_A4_20091106 is being processed
MAR_P2_A5_20091106 is being processed
MAR_P2_A6_20091106 is being processed
MAR_P2_B4_20091106 is being processed
MAR_P2_B6_20091106 is being processed
MAR_P2_C2_20091104 is being processed
MAR_P2_C6_20091106 is being processed
MAR_P2_D2_20091104 is being processed
MAR_P2_D4_20091106 is being processed
MAR_P2_D6_20091106 is being processed
MAR_P2_E4_20091106 is being processed
MAR_P2_E6_20091106 is being processed
MAR_P2_F4_20091106 is being processed
MAR_P2_F6_20091106 is being processed
MAR_P2_G1_20091106 is being processed
MAR_P2_G2_20091104 is being processed
MAR_P2_G3_20091106 is being processed
MAR_P2_G4_20091106 is being processed
MAR_P2_G5_20091106 is being processed
MAR_P2_H1_20091106 is being processed
MAR_P2_H3_20091106 is being processed
MAR_P2_H4_20091106 is being processed
MAR_P3_A1_20091105 is being processed
MAR_P3_B1_20091105 is being processed
MAR_P3_B3_20091105 is being processed
MAR_P3_C1_20091105 is being processed
MAR_P3_C3_20091105 is being processed
```

```
> sample.scar.df.TNBC111[,"whChrAber"] <- 0
```

```
> #----- KCL111: CREATION OF A GENOME LOOKUP TABLE -----#
>
> chromInfo <- read.delim("materials/tables/chromInfo.build36.txt")
> chromInfo$chrom <- gsub("chr","",chromInfo$chrom)
> chromInfo$chrom <- gsub("X",23,chromInfo$chrom)
> chromInfo <- chromInfo[!(chromInfo$chrom=="Y"),]
> chromInfo <- chromInfo[with(chromInfo,
+                             order(as.numeric(chromInfo$chrom),
+                                   as.numeric(chromInfo$chromStart))),]
> chromInfo.centromere <- chromInfo[chromInfo$type=="centromere",]
> chromInfo.centromere <- chromInfo.centromere[,c("chrom",rep("chromStart",2),
+                                                 "chromEnd")]
> chrom.lens <- sapply(unique(chromInfo$chrom),
+                      function(chr) tail(chromInfo[chromInfo$chrom==chr,
+                                                   "chromEnd"],n=1))
> chromInfo.centromere[,2] <- chrom.lens
> colnames(chromInfo.centromere) <- c("chrom","length",
+                                     "centromere_start","centromere_end")
> chromInfo <- chromInfo.centromere; rm(chromInfo.centromere)
> #----- KCL111: FLANKING ABERRATION DISTANCE CALCULATION -----#
>
> AllAber.sam.l.TNBC111 <- lapply(unique(sample.scar.df.TNBC111$sampleID),
+                                 function(sample) {
+
+     whChr <-
+         unlist(lapply(unique(sample.scar.df.TNBC111$Chromosome),
+               function(chrom)
+                    if (!nrow(sample.scar.df.TNBC111[which(
+                        sample.scar.df.TNBC111$Chromosome==chrom &
+                            sample.scar.df.TNBC111$sampleID%in%sample),
+                                                     ])==1) F else T ))
+
+     if (T %in% whChr) {
+
+         whChr.i <-
+             unlist(lapply(which(whChr),
+                           function(chrom)
+                               which(sample.scar.df.TNBC111$Chromosome==chrom &
+                                         (sample.scar.df.TNBC111$copyNum!= 2|
+                                              sample.scar.df.TNBC111$CnLoH==1) &
+                                         sample.scar.df.TNBC111$sampleID==sample)))
+
+         sample.scar.df.TNBC111[whChr.i,"whChrAber"] <- 1
+     }
+     sample.scar.df.TNBC111[which( (sample.scar.df.TNBC111$copyNum != 2 |
```

```
+                                          sample.scar.df.TNBC111$CnLoH == 1) &
+                                         sample.scar.df.TNBC111$sampleID==sample ),
+                             ]
+
+ })
> names(AllAber.sam.l.TNBC111) <-
+     sapply(AllAber.sam.l.TNBC111, function(x) x[1,1])
> for (s.no in 1:length(AllAber.sam.l.TNBC111)) {
+
+     sample <- AllAber.sam.l.TNBC111[[s.no]]
+     sample.tmp <- sample[,1:7]
+     sample.tmp <- cbind( sample.tmp,
+                          "Upstream.dist"=rep(NA, nrow(sample.tmp)),
+                          "Downstream.dist"=rep(NA, nrow(sample.tmp)) )
+
+     for (chrom in unique(sample.tmp$Chromosome)) {
+         sample.chrom <- sample.tmp[which(sample.tmp$Chromosome==chrom),]
+         for (row in 1:nrow(sample.chrom)) {
+             if (row > 1) sample.chrom[row, "Upstream.dist"] <-
+                 as.numeric(sample.chrom[row,"Start"]) -
+                 as.numeric(sample.chrom[row-1,"End"])
+             if (row < nrow(sample.chrom))
+                 sample.chrom[row, "Downstream.dist"] <-
+                 as.numeric(sample.chrom[row+1,"Start"]) -
+                 as.numeric(sample.chrom[row,"End"])
+         }
+         sample.tmp[which(sample.tmp$Chromosome==chrom),] <- sample.chrom
+     }
+     AllAber.sam.l.TNBC111[[s.no]] <-
+         cbind(sample.tmp, sample[,8:ncol(sample)])
+     cat(paste("Completed sample", names(AllAber.sam.l.TNBC111[s.no]),
+               "\n", sep=" "))
+ }

Completed sample MAR_A3_20101029
Completed sample MAR_A3_20110315
Completed sample MAR_B1_20110308
Completed sample MAR_B2_20101026
Completed sample MAR_B3_20101029
Completed sample MAR_B3_20110315
Completed sample MAR_B4_20101029
Completed sample MAR_B4_20110315
Completed sample MAR_B5_20110315
Completed sample MAR_C1_20101029
Completed sample MAR_C1_20110308
```

```
Completed sample MAR_C2_20101026
Completed sample MAR_C4_20101029
Completed sample MAR_D1_20110308
Completed sample MAR_D2_20101026
Completed sample MAR_D3_20101029
Completed sample MAR_D4_20101029
Completed sample MAR_D4_20110315
Completed sample MAR_D5_20110315
Completed sample MAR_E1_20110308
Completed sample MAR_E2_20101026
Completed sample MAR_E3_20110315
Completed sample MAR_E4_20101029
Completed sample MAR_E5_20110315
Completed sample MAR_F2_20101026
Completed sample MAR_F4_20101029
Completed sample MAR_F5_20110311
Completed sample MAR_G2_20110315
Completed sample MAR_G3_20101029
Completed sample MAR_G3_20110315
Completed sample MAR_H1_20110308
Completed sample MAR_H2_20110315
Completed sample MAR_H3_20101029
Completed sample MAR_H4_20101029
Completed sample MAR_H4_20110315
Completed sample MAR_P1_A1_20091104
Completed sample MAR_P1_A2_20091210
Completed sample MAR_P1_A3_20091106
Completed sample MAR_P1_A4_20100702
Completed sample MAR_P1_A6_20091106
Completed sample MAR_P1_B1_20091104
Completed sample MAR_P1_B2_20091106
Completed sample MAR_P1_B2_20091210
Completed sample MAR_P1_B2_20100629
Completed sample MAR_P1_B3_20100702
Completed sample MAR_P1_B6_20091106
Completed sample MAR_P1_C1_20091104
Completed sample MAR_P1_C1_20091210
Completed sample MAR_P1_C2_20091106
Completed sample MAR_P1_C2_20091210
Completed sample MAR_P1_C3_20100702
Completed sample MAR_P1_C4_20100702
Completed sample MAR_P1_C5_20091106
Completed sample MAR_P1_D1_20091104
Completed sample MAR_P1_D2_20091106
Completed sample MAR_P1_D5_20091106
```

```
Completed sample MAR_P1_D6_20091106
Completed sample MAR_P1_D6_20100702
Completed sample MAR_P1_E1_20091210
Completed sample MAR_P1_E2_20091210
Completed sample MAR_P1_E2_20100629
Completed sample MAR_P1_E3_20100702
Completed sample MAR_P1_E4_20091106
Completed sample MAR_P1_E5_20091106
Completed sample MAR_P1_E6_20091106
Completed sample MAR_P1_F1_20091210
Completed sample MAR_P1_F2_20091210
Completed sample MAR_P1_F2_20100629
Completed sample MAR_P1_F4_20100702
Completed sample MAR_P1_F6_20091106
Completed sample MAR_P1_G1_20091210
Completed sample MAR_P1_G1_20100624
Completed sample MAR_P1_G2_20091210
Completed sample MAR_P1_G2_20100629
Completed sample MAR_P1_G4_20091106
Completed sample MAR_P1_G5_20100702
Completed sample MAR_P1_H1_20091104
Completed sample MAR_P1_H1_20091210
Completed sample MAR_P1_H2_20100629
Completed sample MAR_P1_H3_20100702
Completed sample MAR_P1_H4_20091106
Completed sample MAR_P1_H4_20100702
Completed sample MAR_P2_A1_20091106
Completed sample MAR_P2_A2_20091104
Completed sample MAR_P2_A4_20091106
Completed sample MAR_P2_A5_20091106
Completed sample MAR_P2_A6_20091106
Completed sample MAR_P2_B4_20091106
Completed sample MAR_P2_B6_20091106
Completed sample MAR_P2_C2_20091104
Completed sample MAR_P2_C6_20091106
Completed sample MAR_P2_D2_20091104
Completed sample MAR_P2_D4_20091106
Completed sample MAR_P2_D6_20091106
Completed sample MAR_P2_E4_20091106
Completed sample MAR_P2_E6_20091106
Completed sample MAR_P2_F4_20091106
Completed sample MAR_P2_F6_20091106
Completed sample MAR_P2_G1_20091106
Completed sample MAR_P2_G2_20091104
Completed sample MAR_P2_G3_20091106
```

```
Completed sample MAR_P2_G4_20091106
Completed sample MAR_P2_G5_20091106
Completed sample MAR_P2_H1_20091106
Completed sample MAR_P2_H3_20091106
Completed sample MAR_P2_H4_20091106
Completed sample MAR_P3_A1_20091105
Completed sample MAR_P3_B1_20091105
Completed sample MAR_P3_B3_20091105
Completed sample MAR_P3_C1_20091105
Completed sample MAR_P3_C3_20091105

> #----- KCL111: MEASUREMENT OF SCINS -----#
>
> SCINS.scores.TNBC111 <-
+     measureSCINS(AllAber.sam.l=AllAber.sam.l.TNBC111,
+                  dataset="TNBC111", chromInfo=chromInfo)
> head(signif(SCINS.scores.TNBC111,3))

                   AiCna AbCna  CnLoH     Ai      Cna
MAR_A3_20101029 11.50000    23 24.100 35.600 34.50000
MAR_A3_20110315  0.00576     0  0.501  0.507  0.00576
MAR_B1_20110308 20.00000     6 24.300 44.300 26.00000
MAR_B2_20101026 18.80000     9 23.300 42.100 27.80000
MAR_B3_20101029 40.00000     9  3.050 43.000 49.00000
MAR_B3_20110315 13.30000     0  0.000 13.300 13.30000
```