

CIRCULATE Galaxy Nakamura et al 2024

Code ▾

library(swimplot) library(Rmpfr) library(coxphf) library(grid) library(gtable) library(readr) library(mosaic) library(dplyr) library(survminer) library(gridtext) library(ggplot2) library(scales) library(officer) library(ggthemes) library(tidyverse) library(gtsummary) library(flextable) library(parameters) library(car) library(grid) library(ComplexHeatmap) library(readxl) library(janitor) library(rms) library(DT)

#Demographics Table

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]

circ_data_subset <- circ_data %>%
  select(
    Age,
    Gender,
    ECOG,
    PrimSite,
    pT,
    pN,
    Stage,
    NAC,
    ACT,
    BRAF.V600E,
    RAS,
    MSI,
    RFS.Event,
    OS.months) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    ECOG = factor(ECOG, levels = c(0, 1)),
    PrimSite = factor(PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum")),
    pT = factor(pT, levels = c("T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    Stage = factor(Stage, levels = c("I", "II", "III", "IV")),
    NAC = factor(NAC, levels = c("TRUE", "FALSE"), labels = c("Neoadjuvant Chemotherapy", "Upfront Surgery")),
    ACT = factor(ACT, levels = c("TRUE", "FALSE"), labels = c("Adjuvant Chemotherapy", "Observation")),
    BRAF.V600E = factor(BRAF.V600E, levels = c("WT", "MUT"), labels = c("BRAF wt", "BRAF V600E")),
    RAS = factor(RAS, levels = c("WT", "MUT"), labels = c("RAS wt", "RAS mut")),
    MSI = factor(MSI, levels = c("MSS", "MSI-High")),
    RFS.Event = factor(RFS.Event, levels = c("TRUE", "FALSE"), labels = c("Recurrence", "No Recurrence")),
    OS.months = as.numeric(OS.months))
table1 <- circ_data_subset %>%
 tbl_summary(
  statistic = list(
    all_continuous() ~ "{median} ({min} - {max})",
    all_categorical() ~ "{n} ({p}%)") %>%
  bold_labels()
table1
```

Characteristic	N = 2,240 ¹
Age	69 (28 - 95)
Gender	
Male	1,149 (51%)
Female	1,091 (49%)
ECOG	
0	2,046 (91%)
1	194 (8.7%)
PrimSite	
Right-sided colon	863 (39%)
Left-sided colon	1,377 (61%)

¹ Median (Range); n (%)

Characteristic	N = 2,240 [†]
Rectum	0 (0%)
pT	
T1-T2	317 (16%)
T3-T4	1,630 (84%)
Unknown	293
pN	
N0	922 (47%)
N1-N2	1,025 (53%)
Unknown	293
Stage	
I	234 (10%)
II	652 (29%)
III	936 (42%)
IV	418 (19%)
NAC	
Neoadjuvant Chemotherapy	218 (9.7%)
Upfront Surgery	2,022 (90%)
ACT	
Adjuvant Chemotherapy	946 (42%)
Observation	1,294 (58%)
BRAF.V600E	
BRAF wt	2,062 (92%)
BRAF V600E	178 (7.9%)
RAS	
RAS wt	1,303 (58%)
RAS mut	937 (42%)
MSI	
MSS	2,025 (90%)
MSI-High	215 (9.6%)
RFS.Event	
Recurrence	500 (22%)
No Recurrence	1,740 (78%)
OS.months	23 (2 - 49)

[†] Median (Range); n (%)

Hide

```
fit1 <- as_flex_table(
  table1,
  include = everything(),
  return_calls = FALSE,
  strip_md_bold = TRUE)
```

Warning: The `strip_md_bold` argument of `as_flex_table()` is deprecated as of gtsummary 1.6.0.
 This warning is displayed once every 8 hours.
 Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.

Hide

fit1

Characteristic	N = 2,240 ¹
Age	69 (28 - 95)
Gender	
Male	1,149 (51%)
Female	1,091 (49%)
ECOG	
0	2,046 (91%)
1	194 (8.7%)
PrimSite	
Right-sided colon	863 (39%)
Left-sided colon	1,377 (61%)
Rectum	0 (0%)
pT	
T1-T2	317 (16%)
T3-T4	1,630 (84%)
Unknown	293
pN	
N0	922 (47%)
N1-N2	1,025 (53%)
Unknown	293
Stage	
I	234 (10%)
II	652 (29%)
III	936 (42%)
IV	418 (19%)
NAC	
Neoadjuvant Chemotherapy	218 (9.7%)
Upfront Surgery	2,022 (90%)
ACT	
Adjuvant Chemotherapy	946 (42%)
Observation	1,294 (58%)
BRAF.V600E	
BRAF wt	2,062 (92%)
BRAF V600E	178 (7.9%)
RAS	
RAS wt	1,303 (58%)
RAS mut	937 (42%)
MSI	
MSS	2,025 (90%)
MSI-High	215 (9.6%)
RFS.Event	
Recurrence	500 (22%)
No Recurrence	1,740 (78%)
OS.months	23 (2 - 49)

¹Median (Range); n (%)

Hide

```
save_as_docx(fit1, path= "~/Downloads/table1.docx")
```

#Table of Metastatic organ involvement in Stage IV

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Stage=="IV",]

circ_data_subset <- circ_data %>%
  select(
    Mets.Organ) %>%
  mutate(
    Mets.Organ = factor(Mets.Organ))
table1 <- circ_data_subset %>%
 tbl_summary(
  statistic = list(
    all_continuous() ~ "{median} ({min} - {max})",
    all_categorical() ~ "{n} ({p}%)") %>%
  bold_labels()
table1
```

Characteristic	N = 418 ¹
Mets.Organ	
	8 (1.9%)
Anal canal	1 (0.2%)
Appendix, Adrenal gland	1 (0.2%)
Appendix, Lymph Node	1 (0.2%)
Bone	1 (0.2%)
Cervix, Uterine body, Ovary	1 (0.2%)
Colon, Appendix	1 (0.2%)
Colon, Liver	1 (0.2%)
Colon, Liver, Uterine body, Femoral nerve, Iliopsoas muscle	1 (0.2%)
Colon, Lymph Node	1 (0.2%)
Colon, Peritoneum	2 (0.5%)
Colon, Rectum, Peritoneum	1 (0.2%)
Gallbladder/Bile duct, Pancreas	1 (0.2%)
Liver	214 (51%)
Liver, Gallbladder/Bile duct	38 (9.1%)
Liver, Lung	1 (0.2%)
Liver, Lymph node	2 (0.5%)
Liver, Peritoneum	5 (1.2%)
Liver, Peritoneum, Diaphragm	1 (0.2%)
Liver, Spleen	1 (0.2%)
Lung	73 (17%)
Lymph node	16 (3.8%)
Lymph Node	1 (0.2%)
Ovary	1 (0.2%)
Peritoneum	20 (4.8%)

¹ n (%)

Characteristic	N = 418 ¹
Peritoneum, Ovary	1 (0.2%)
Peritoneum, Pancreas, Spleen	1 (0.2%)
Peritoneum, Uterine Body, Ovary	1 (0.2%)
Rectum	5 (1.2%)
Rectum, Cervix, Uterine body, Ovary, Vagina	1 (0.2%)
Rectum, Liver, Appendix	1 (0.2%)
Rectum, Lymph node	2 (0.5%)
Small intestine	1 (0.2%)
Small intestine, Colon, Liver, Pancreas, Peritoneum	1 (0.2%)
Small intestine, Colon, Pancreas, Peritoneum	1 (0.2%)
Small intestine, Colon, Rectum, Urinary tract	2 (0.5%)
Small intestine, Colon, Urinary tract, Lymph node	1 (0.2%)
Small intestine, Peritoneum, Urinary tract	1 (0.2%)
Small intestine, Uterine body, Urinary tract	1 (0.2%)
Spleen	1 (0.2%)
Transverse colon, Descending colon	1 (0.2%)
Urinary Tract	1 (0.2%)
Urinary Tract, Lymph Node	1 (0.2%)

¹ n (%)[Hide](#)

```
fit1 <- as_flex_table(
  table1,
  include = everything(),
  return_calls = FALSE,
  strip_md_bold = TRUE)
fit1
```

Characteristic	N = 418 ¹
Mets.Organ	
	8 (1.9%)
Anal canal	1 (0.2%)
Appendix, Adrenal gland	1 (0.2%)
Appendix, Lymph Node	1 (0.2%)
Bone	1 (0.2%)
Cervix, Uterine body, Ovary	1 (0.2%)
Colon, Appendix	1 (0.2%)
Colon, Liver	1 (0.2%)
Colon, Liver, Uterine body, Femoral nerve, Iliopsoas muscle	1 (0.2%)
Colon, Lymph Node	1 (0.2%)
Colon, Peritoneum	2 (0.5%)
Colon, Rectum, Peritoneum	1 (0.2%)
Gallbladder/Bile duct, Pancreas	1 (0.2%)
Liver	214 (51%)
Liver, Gallbladder/Bile duct	38 (9.1%)
Liver, Lung	1 (0.2%)
Liver, Lymph node	2 (0.5%)

¹n (%)

Characteristic	N = 418 ¹
Liver, Peritoneum	5 (1.2%)
Liver, Peritoneum, Diaphragm	1 (0.2%)
Liver, Spleen	1 (0.2%)
Lung	73 (17%)
Lymph node	16 (3.8%)
Lymph Node	1 (0.2%)
Ovary	1 (0.2%)
Peritoneum	20 (4.8%)
Peritoneum, Ovary	1 (0.2%)
Peritoneum, Pancreas, Spleen	1 (0.2%)
Peritoneum, Uterine Body, Ovary	1 (0.2%)
Rectum	5 (1.2%)
Rectum, Cervix, Uterine body, Ovary, Vagina	1 (0.2%)
Rectum, Liver, Appendix	1 (0.2%)
Rectum, Lymph node	2 (0.5%)
Small intestine	1 (0.2%)
Small intestine, Colon, Liver, Pancreas, Peritoneum	1 (0.2%)
Small intestine, Colon, Pancreas, Peritoneum	1 (0.2%)
Small intestine, Colon, Rectum, Urinary tract	2 (0.5%)
Small intestine, Colon, Urinary tract, Lymph node	1 (0.2%)
Small intestine, Peritoneum, Urinary tract	1 (0.2%)
Small intestine, Uterine body, Urinary tract	1 (0.2%)
Spleen	1 (0.2%)
Transverse colon, Descending colon	1 (0.2%)
Urinary Tract	1 (0.2%)
Urinary Tract, Lymph Node	1 (0.2%)

¹n (%)[Hide](#)

save_as_docx(fit1, path= "~/Downloads/table1.docx")

#ctDNA Detection Rates by Window and Stages

[Hide](#)

```
#ctDNA at Baseline
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data$ctDNA.Baseline <- factor(circ_data$ctDNA.Baseline, levels=c("NEGATIVE","POSITIVE"))
circ_data <- subset(circ_data, ctDNA.Baseline %in% c("NEGATIVE", "POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I", "II", "III", "IV"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Baseline == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Baseline, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.Baseline == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)
```

Stage	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<chr>
I	228	127	55.70%
II	618	586	94.82%
III	794	759	95.59%
IV	60	59	98.33%
Overall	1700	1531	90.06%

5 rows

Hide

```
#ctDNA at MRD Window
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I", "II", "III", "IV"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.MRD == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.MRD, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.MRD == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)
```

Stage <fctr>	Total_Count <int>	Positive_Count <int>	Rate <chr>
I	228	2	0.88%
II	629	45	7.15%
III	846	163	19.27%
IV	407	127	31.20%
Overall	2110	337	15.97%
5 rows			

Hide

```
#ctDNA at Surveillance Window
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
circ_data <- subset(circ_data, ctDNA.Surveillance %in% c("NEGATIVE", "POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I", "II", "III", "IV"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.Surveillance == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)
```

Stage <fctr>	Total_Count <int>	Positive_Count <int>	Rate <chr>
I	30	0	0.00%
II	560	47	8.39%
III	872	142	16.28%
IV	332	124	37.35%
Overall	1794	313	17.45%
5 rows			

#DFS by ctDNA at the MRD Window - All stages Landmark MRD timepoint

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
  ctDNA.MRD, data = circ_data)

1 observation deleted due to missingness
  n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 1773    233     NA     NA     NA
ctDNA.MRD=POSITIVE  336    263    5.34    4.83    6.7
```

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

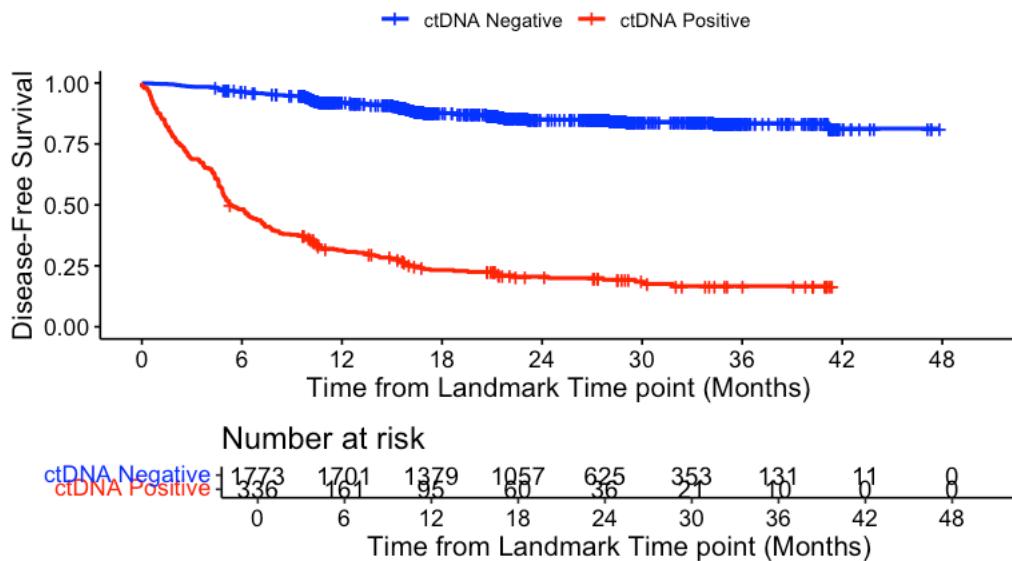
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1773	233	0.1314157	13.14157
POSITIVE	336	263	0.7827381	78.27381
NA	1	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | All stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | All stages



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

1 observation deleted due to missingness
ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	625	224	0.851	0.00949	0.832	0.869	0.860	0.866
30	353	6	0.841	0.01025	0.820	0.860	0.812	0.856
36	131	2	0.835	0.01101	0.812	0.856	0.812	0.856

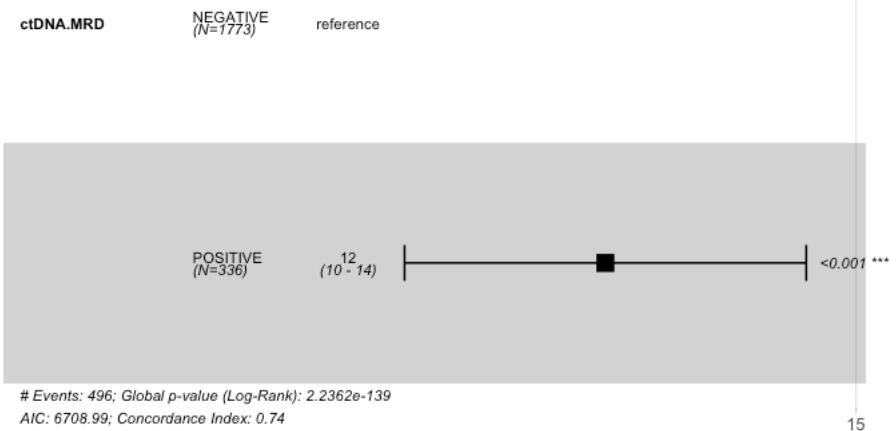
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	36	258	0.206	0.0236	0.161	0.254	0.160	0.254
30	21	3	0.185	0.0242	0.140	0.234	0.121	0.219
36	10	2	0.167	0.0250	0.121	0.219	0.121	0.219

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 2109, number of events= 496
(1 observation deleted due to missingness)

      coef exp(coef)  se(coef)   z Pr(>|z|)
ctDNA.MRDPOSITIVE  2.48392  11.98819  0.09162 27.11  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    11.99     0.08342    10.02     14.35

Concordance= 0.738 (se = 0.01 )
Likelihood ratio test= 631.6 on 1 df,  p=<2e-16
Wald test            = 734.9 on 1 df,  p=<2e-16
Score (logrank) test = 1164 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 1024)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 7.5528013300687189120658297706005503243211205660546529501472236266117489649457
02588559005333873913691374243368415917252509821870794590469056514572824433063799969274721840269882230960709802524
09159977074130441915748770396333861485082783634801646179004987250992843624811754272870226285707769491524942224709
792577e-162"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 11.99 (10.02-14.35); p = 0"
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 662.58, df = 1, p-value < 2.2e-16
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
17.60071 32.35538
sample estimates:
odds ratio
23.75245
```

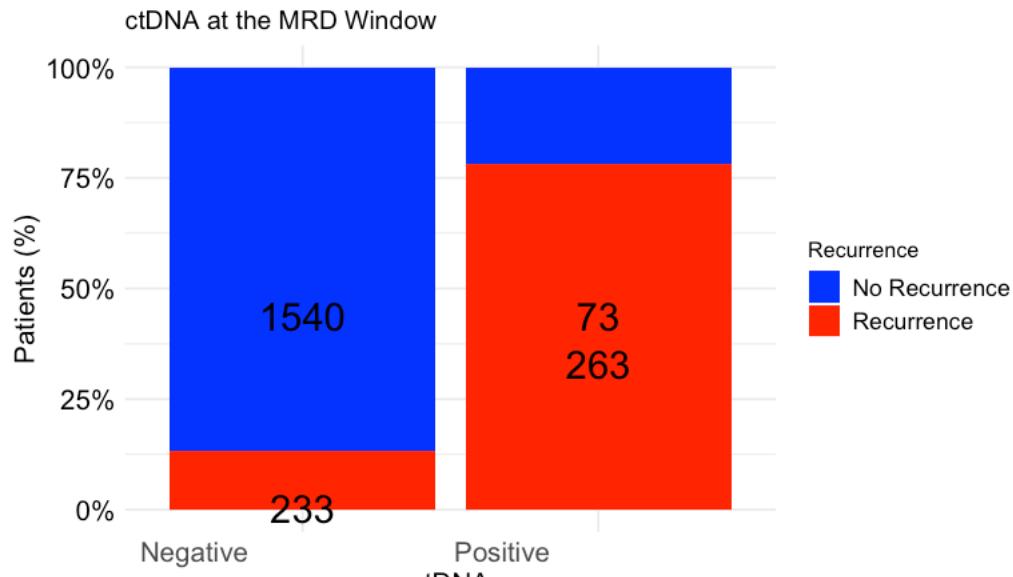
Hide

```
print(contingency_table)
```

		No Recurrence	Recurrence
Negative	1540	233	
Positive	73	263	

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Chi-squared test p-value: < 2.22e-16

Hide

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 4.10912226866353502019019437342893251528010789759394769882243404379480373
6500514e-146"
```

#DFS by ctDNA at the MRD Window - Stage I Landmark MRD timepoint

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("II", "III", "IV"))]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)
```

```
      n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 226      7     NA     NA     NA
ctDNA.MRD=POSITIVE   2      2   15.3   0.526     NA
```

[Hide](#)

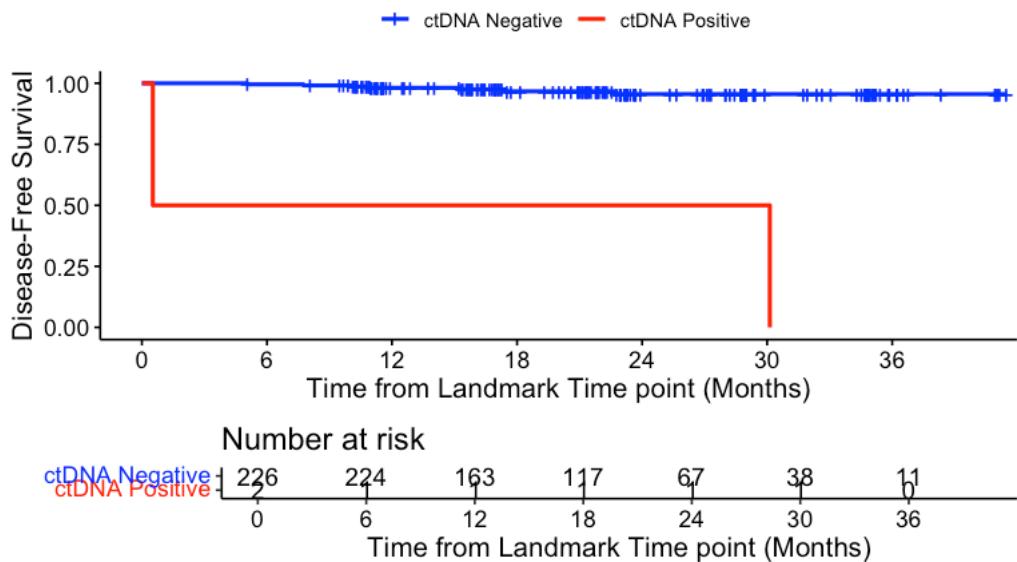
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	226	7	0.03097345	3.097345
POSITIVE	2	2	1.00000000	100.000000
2 rows				

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="DFS - ctDNA MRD window | Stage I", ylab= "Disease-Free Survival", xlab="Time from Landm
ark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage I



summary(KM_curve, times= c(24))

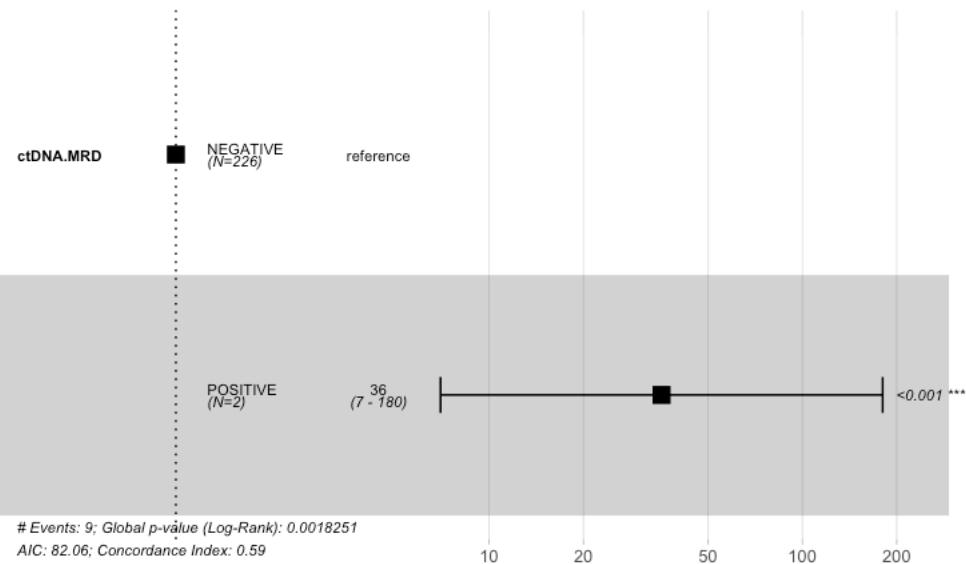
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
 conf.int = 0.95, conf.type = "log-log")

ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	67.0000	7.0000	0.9556	0.0176	0.9043	0.9797

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	1.0000	1.0000	0.50000	0.35355	0.00598	0.91041

circ_data\$ctDNA.MRD <- factor(circ_data\$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
 cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
 ggforest(cox_fit, data = circ_data)

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 228, number of events= 9

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.MRDPOSITIVE 3.5700  35.5148  0.8291  4.306 1.66e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE  35.51    0.02816   6.993    180.4

Concordance= 0.587 (se = 0.069 )
Likelihood ratio test= 9.72 on 1 df,  p=0.002
Wald test            = 18.54 on 1 df,  p=2e-05
Score (logrank) test = 47.16 on 1 df,  p=7e-12
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 35.51 (6.99-180.35); p = 0"
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("II", "III", "IV")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)

```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

[Hide](#)

```
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 26.866, df = 1, p-value = 2.181e-07
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.001391
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
4.884572      Inf
sample estimates:
odds ratio
      Inf
```

[Hide](#)

```
print(contingency_table)
```

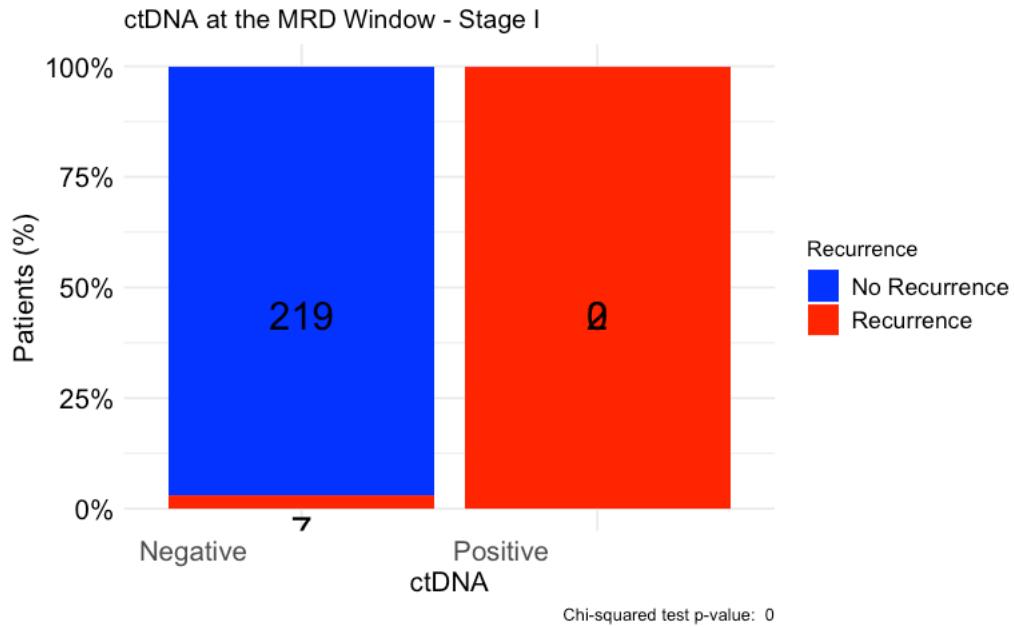
	No Recurrence	Recurrence
Negative	219	7
Positive	0	2

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage I",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



#DFS by ctDNA at the MRD Window - Stage II Landmark MRD timepoint

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "III", "IV")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
  ctDNA.MRD, data = circ_data)

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	584	30	NA	NA	NA
ctDNA.MRD=POSITIVE	45	30	7.75	5.45	NA

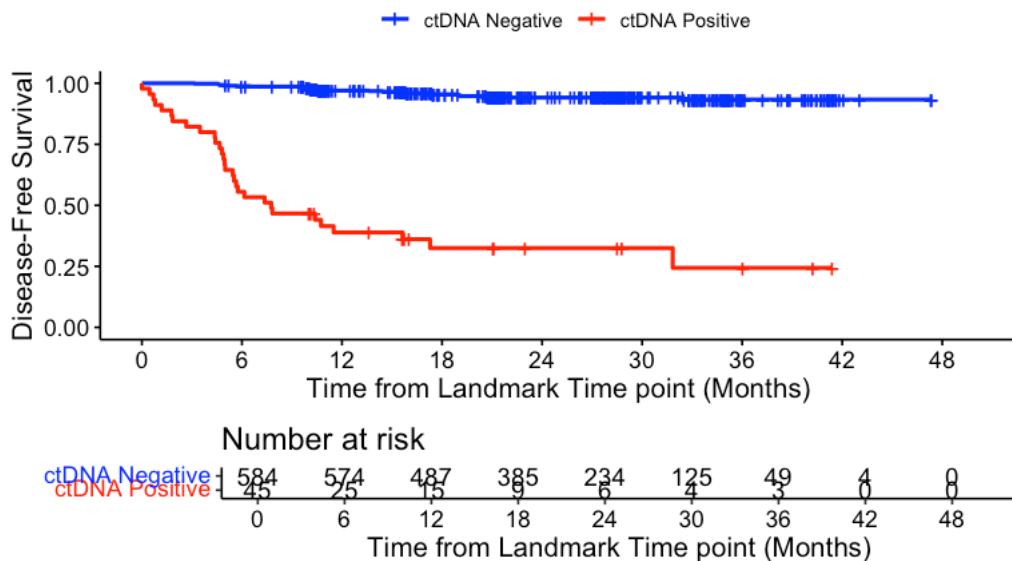
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	584	30	0.05136986	5.136986
POSITIVE	45	30	0.66666667	66.666667
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | Stage II", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage II



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

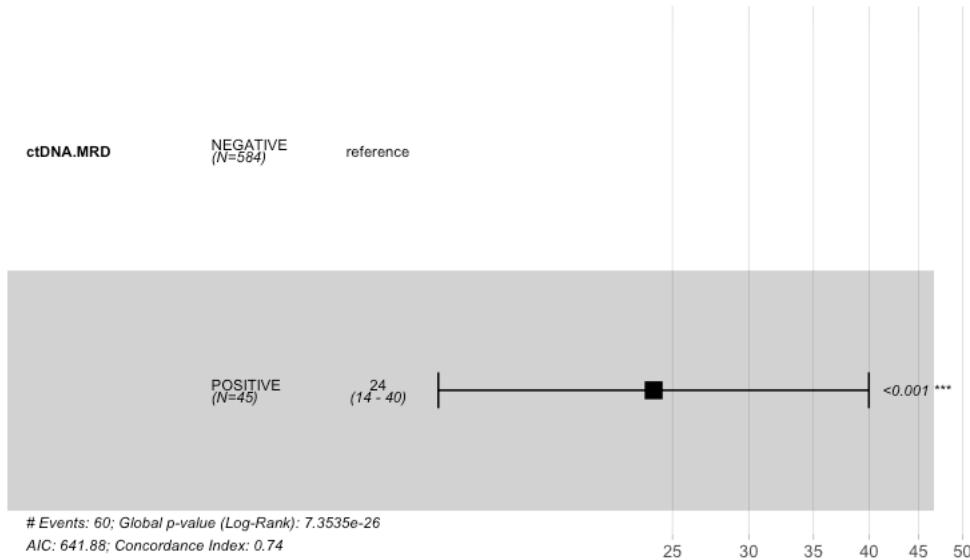
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	234.0000	29.0000	0.9413	0.0108	0.9159	0.9592

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	6.0000	29.0000	0.3250	0.0749	0.1864	0.4714

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 629, number of events= 60

      coef exp(coef)  se(coef)   z Pr(>|z|)
ctDNA.MRDPOSITIVE  3.1738  23.8977  0.2623 12.1  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    23.9    0.04184    14.29    39.96

Concordance= 0.745 (se = 0.031 )
Likelihood ratio test= 110.6 on 1 df,  p=<2e-16
Wald test            = 146.4 on 1 df,  p=<2e-16
Score (logrank) test = 310.6 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 1.04791318986334798932736705009648801906613416483650472938641972665484361424249
1e-33"
```

[Hide](#)

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 23.9 (14.29-39.96); p = 0"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "III", "IV"))]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning in stats::chisq.test(x, y, ...) :
  Chi-squared approximation may be incorrect
```

[Hide](#)

```
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 176.25, df = 1, p-value < 2.2e-16
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data
```

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
16.95374 81.29786
sample estimates:
odds ratio
36.33032
```

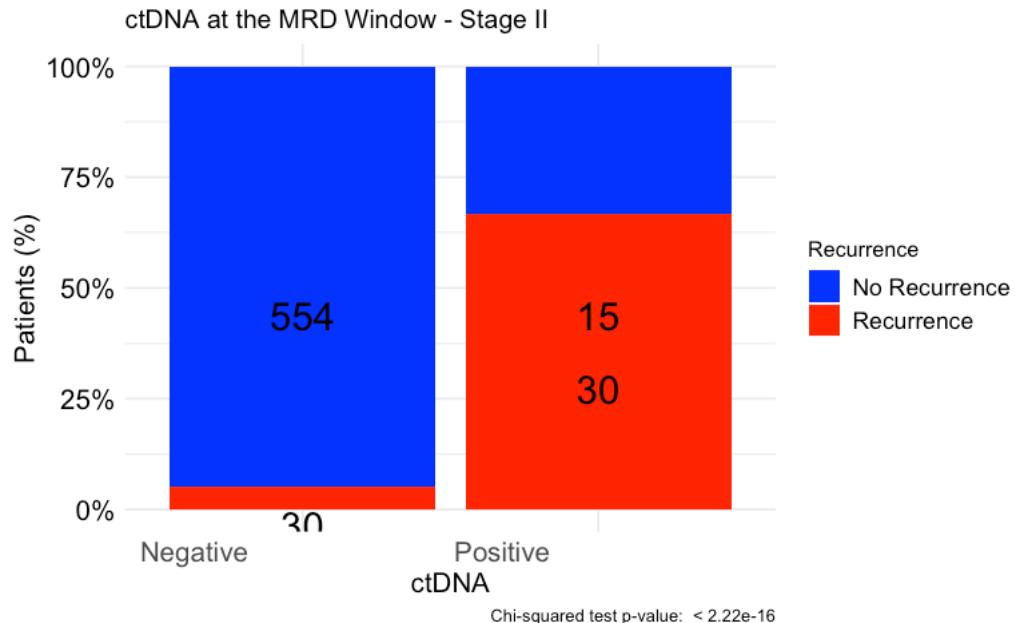
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Negative	554	30
Positive	15	30

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage II",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 3.19820679377280944357342313876631899983532737404256366643891320802423710
6040795e-40"
```

#DFS by ctDNA at the MRD Window - Stage III Landmark MRD timepoint

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)

1 observation deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 683     82     NA     NA     NA
ctDNA.MRD=POSITIVE 162    117    9.48    7.16   11.7

```

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	683	82	0.1200586	12.00586
POSITIVE	162	117	0.7222222	72.22222
NA	1	NA	NA	NA

3 rows

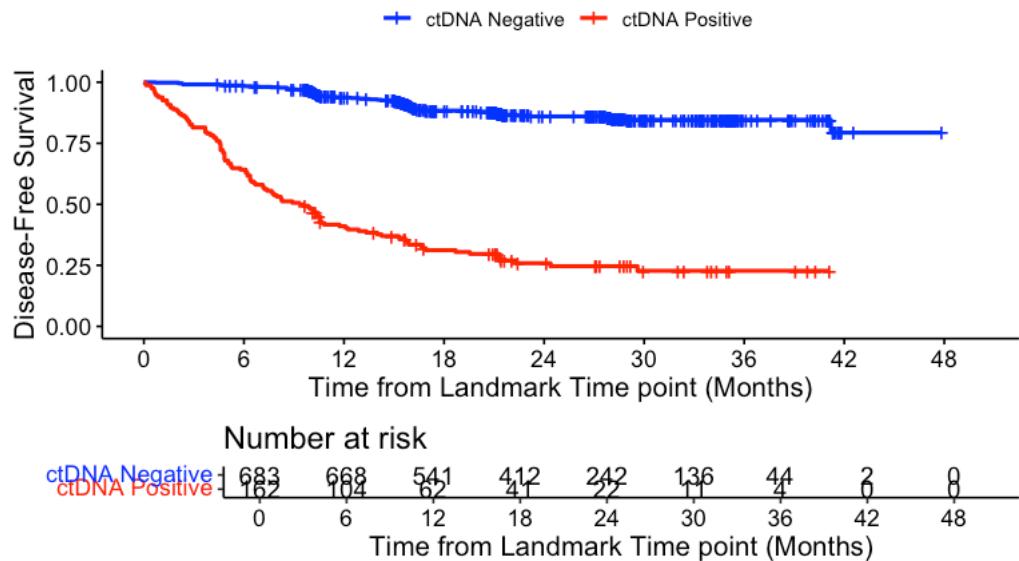
Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```

DFS - ctDNA MRD window | Stage III

[Hide](#)

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

1 observation deleted due to missingness

ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	242.000	78.000	0.8600	0.0152	0.8272	0.8870

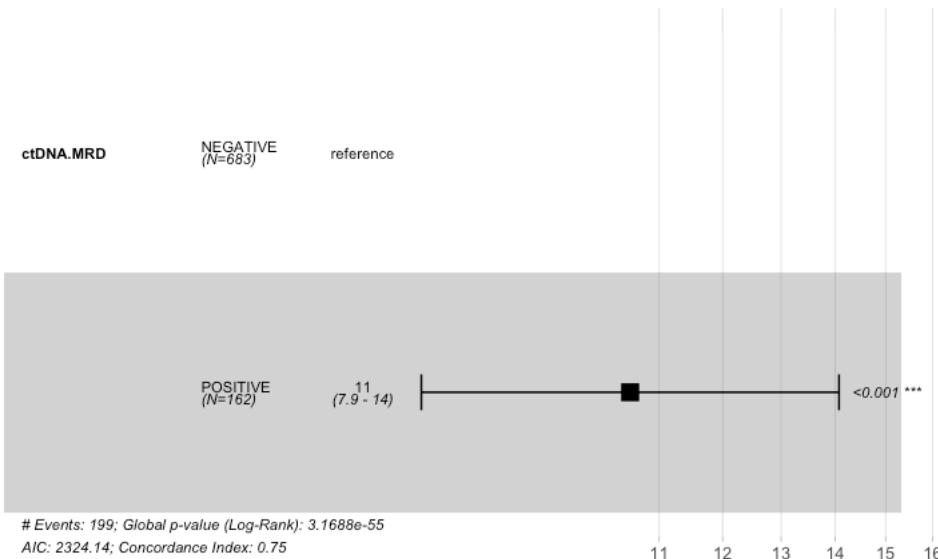
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00	22.00	115.000	0.259	0.037	0.190	0.334

[Hide](#)

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 845, number of events= 199
(1 observation deleted due to missingness)

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE  2.3582   10.5722   0.1459 16.16 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   10.57    0.09459    7.942    14.07

Concordance= 0.752 (se = 0.016 )
Likelihood ratio test= 245 on 1 df,  p=<2e-16
Wald test            = 261.2 on 1 df,  p=<2e-16
Score (logrank) test = 399.4 on 1 df,  p=<2e-16

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))

```

```
[1] "Exact p-value for ctDNA.MRD: 9.55360981371212951558846717784526451006827267181370104385557464531799187996559
5e-59"
```

[Hide](#)

```

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 10.57 (7.94-14.07); p = 0"
```

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data: contingency_table
X-squared = 260.38, df = 1, p-value < 2.2e-16

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
12.35353 29.50193
sample estimates:
odds ratio
18.94366
```

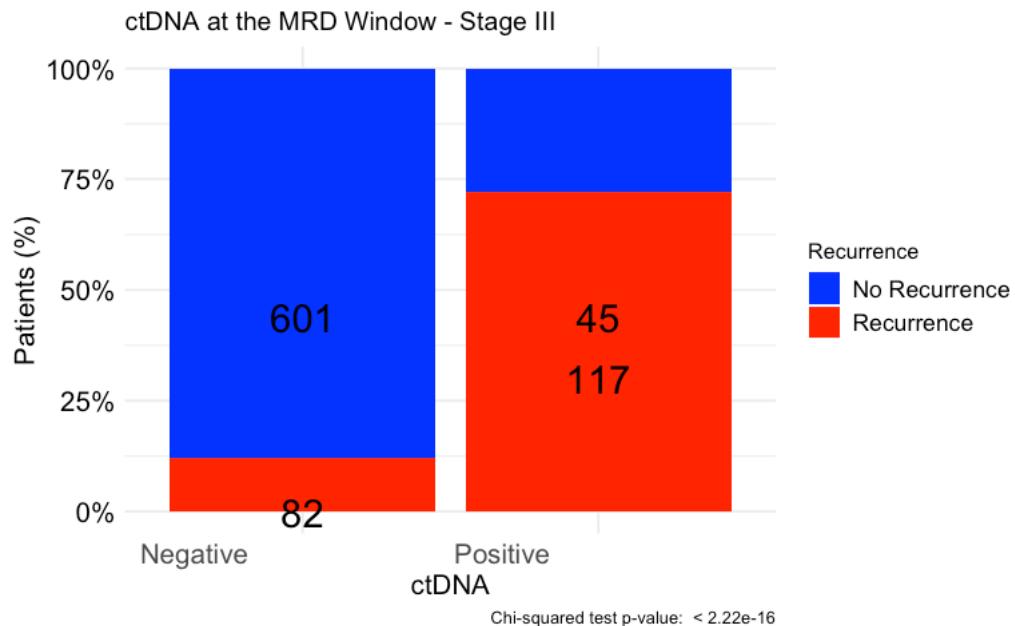
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Negative	601	82
Positive	45	117

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage III",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) # define custom colors
theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
      axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
      axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
      axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
      legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 1.41683853465137144409778446638536062423000782847886558768546058518114554
7173816e-58"
```

#DFS by ctDNA at the MRD Window - High Risk Stage II Landmark MRD timepoint

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)
```

```
1481 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 475      24      NA      NA      NA
ctDNA.MRD=POSITIVE  42       28     7.56     4.99      NA
```

[Hide](#)

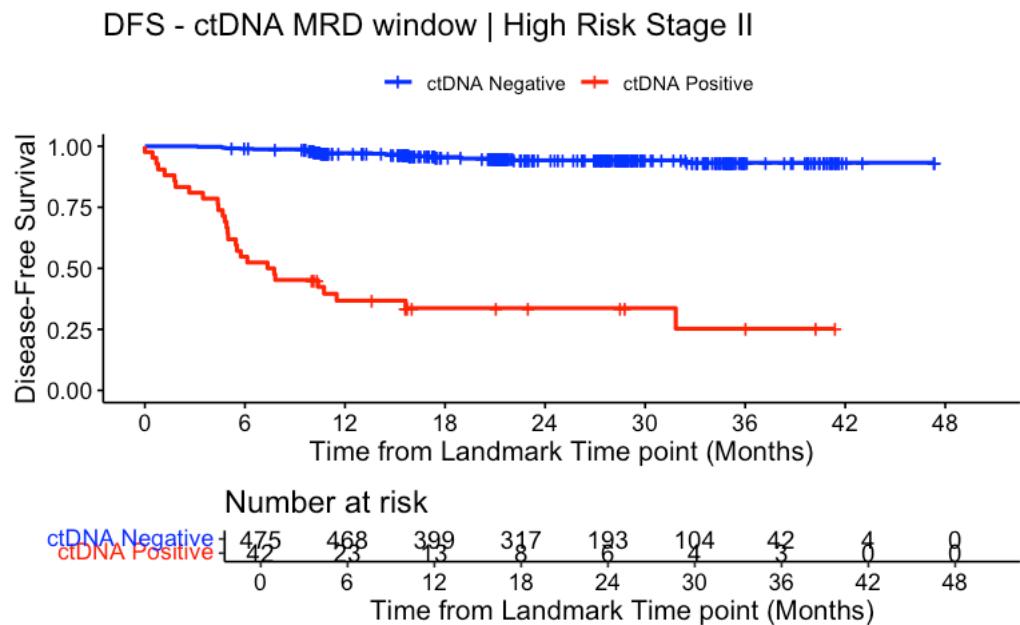
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	475	24	0.05052632	5.052632
POSITIVE	42	28	0.66666667	66.666667
NA	1481	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | High Risk Stage II", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

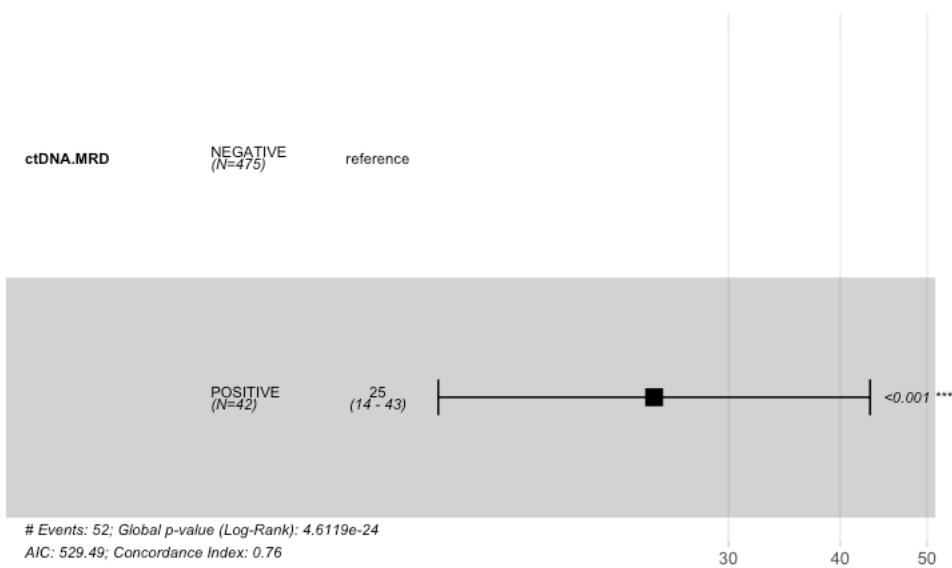
1481 observations deleted due to missingness
  ctDNA.MRD=NEGATIVE
  time      n.risk      n.event      survival      std.err  lower 95% CI upper 95% CI
  24.000    193.000    23.000      0.942      0.012    0.914    0.962

  ctDNA.MRD=POSITIVE
  time      n.risk      n.event      survival      std.err  lower 95% CI upper 95% CI
  24.000     6.000     27.000      0.337      0.076    0.195    0.484
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 517, number of events= 52
(1481 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 3.2102   24.7836   0.2831 11.34 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   24.78    0.04035    14.23    43.16

Concordance= 0.764 (se = 0.033 )
Likelihood ratio test= 102.4 on 1 df,  p=<2e-16
Wald test            = 128.6 on 1 df,  p=<2e-16
Score (logrank) test = 275.5 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 8.29846873012490439010624518128617936562781665334401673588662652406562617664123
0e-30"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 24.78 (14.23-43.16); p = 0"
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)

```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

[Hide](#)

```
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 155.19, df = 1, p-value < 2.2e-16
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
16.43949 86.53043
sample estimates:
odds ratio
36.84362
```

[Hide](#)

```
print(contingency_table)
```

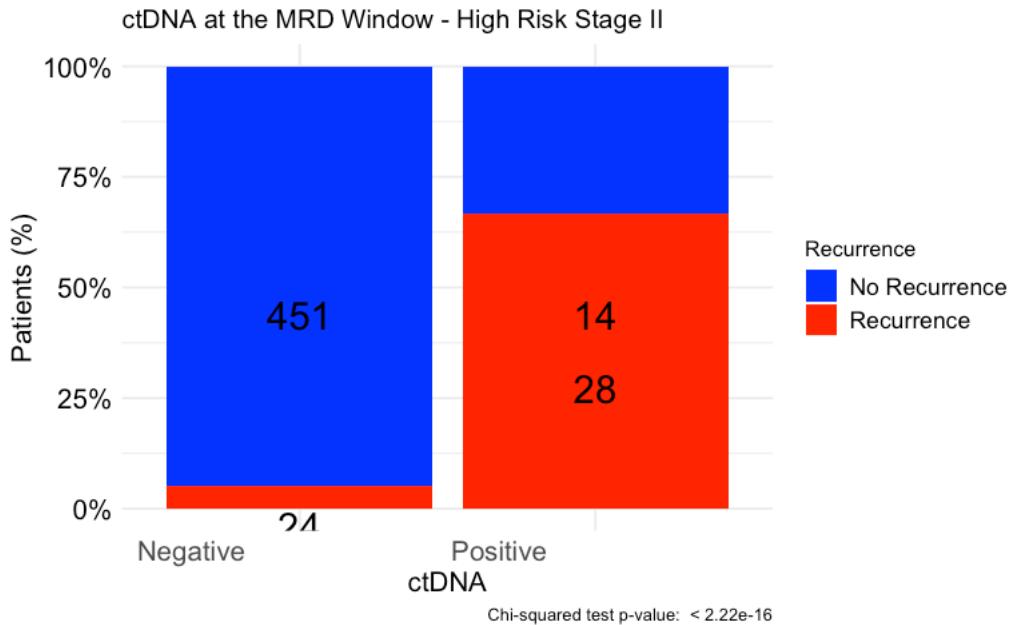
	No Recurrence	Recurrence
Negative	451	24
Positive	14	28

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - High Risk Stage II",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



```

#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value with high precision: 1.26990756673362680456505178129412376000388109755648476849672088686242626
7774383e-35"

```

#DFS by ctDNA at the MRD Window - High Risk Stage III Landmark MRD timepoint

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$Risk.StageIII==TRUE,]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)

```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
  ctDNA.MRD, data = circ_data)

1265 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 383      56      NA      NA      NA
ctDNA.MRD=POSITIVE 105      79   10.1    7.66     14
```

[Hide](#)

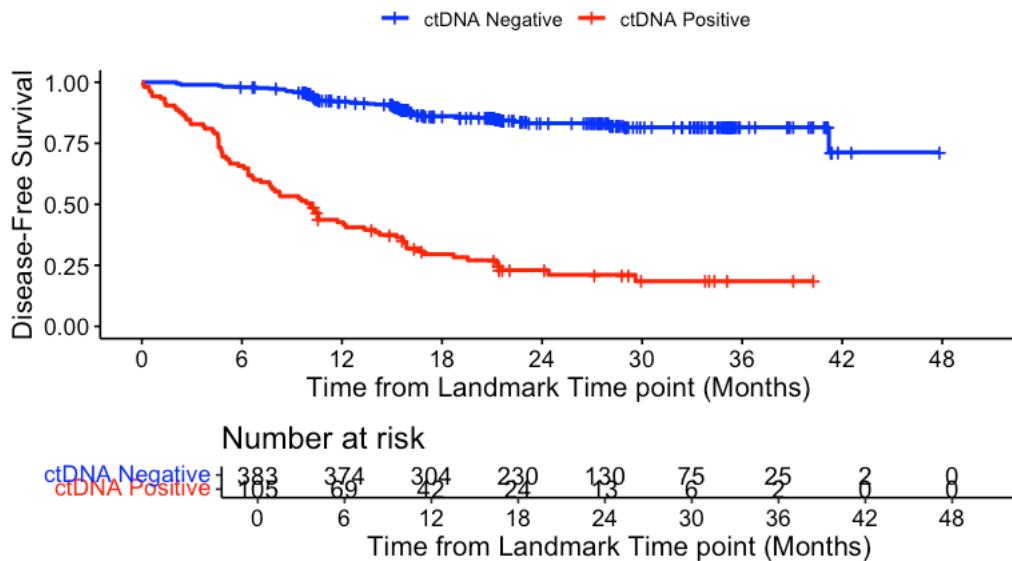
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD <chr>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
NEGATIVE	383	56	0.1462141	14.62141
POSITIVE	105	79	0.7523810	75.23810
NA	1265	NA	NA	NA
3 rows				

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | High Risk Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | High Risk Stage III

[Hide](#)

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

1265 observations deleted due to missingness

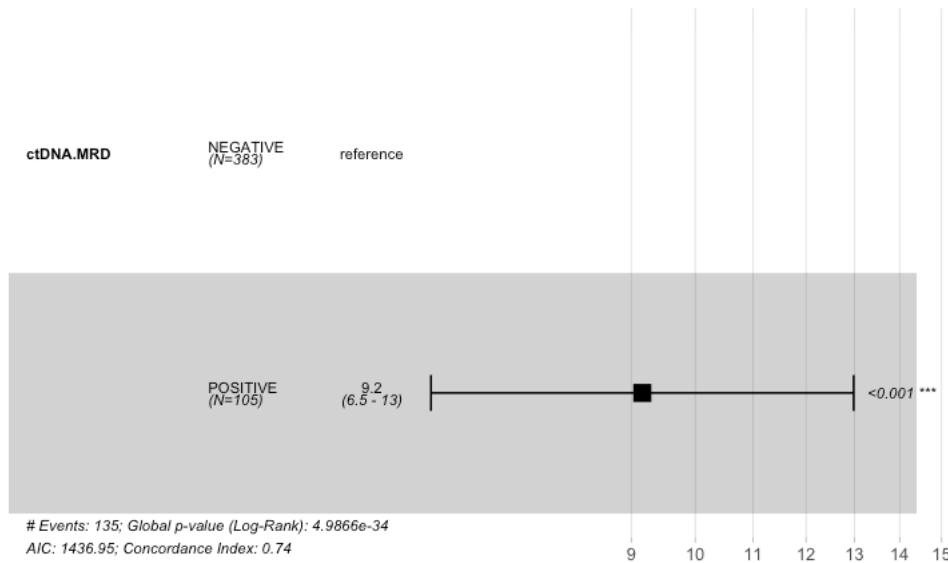
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	130.0000	53.0000	0.8322	0.0219	0.7842	0.8705

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	13.0000	77.0000	0.2305	0.0443	0.1500	0.3214

[Hide](#)

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 488, number of events= 135
(1265 observations deleted due to missingness)

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.2154    9.1654  0.1775 12.48   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    9.165     0.1091    6.472    12.98

Concordance= 0.74  (se = 0.019 )
Likelihood ratio test= 147.9 on 1 df,  p=<2e-16
Wald test            = 155.7 on 1 df,  p=<2e-16
Score (logrank) test = 226.5 on 1 df,  p=<2e-16

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value for ctDNA.MRD: 9.66261771819737748477760609783996451428608465870805945032179694804073694927437
8e-36"

```

[Hide](#)

```

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```

[1] "HR = 9.17 (6.47-12.98); p = 0"

```

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$Risk.StageIII==TRUE,]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data: contingency_table
X-squared = 148.3, df = 1, p-value < 2.2e-16

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
10.18115 31.22155
sample estimates:
odds ratio
17.58359
```

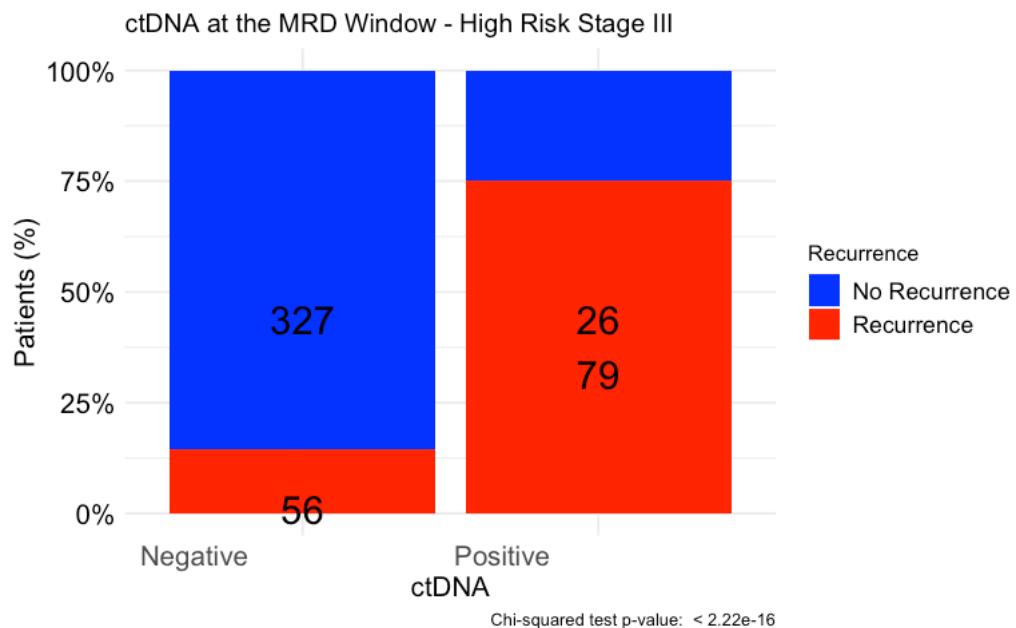
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Negative	327	56
Positive	26	79

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - High Risk Stage III",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) # define custom colors
theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
      axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
      axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
      axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
      legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 4.07566444074808073098139879360067122849218671792285282961239673796757046
5527740e-34"
```

#DFS by ctDNA at the MRD Window - Stage I-III Landmark MRD timepoint

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("IV")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)
```

```
1 observation deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 1493    119      NA      NA      NA
ctDNA.MRD=POSITIVE  209     149     8.9     7.16    10.7
```

[Hide](#)

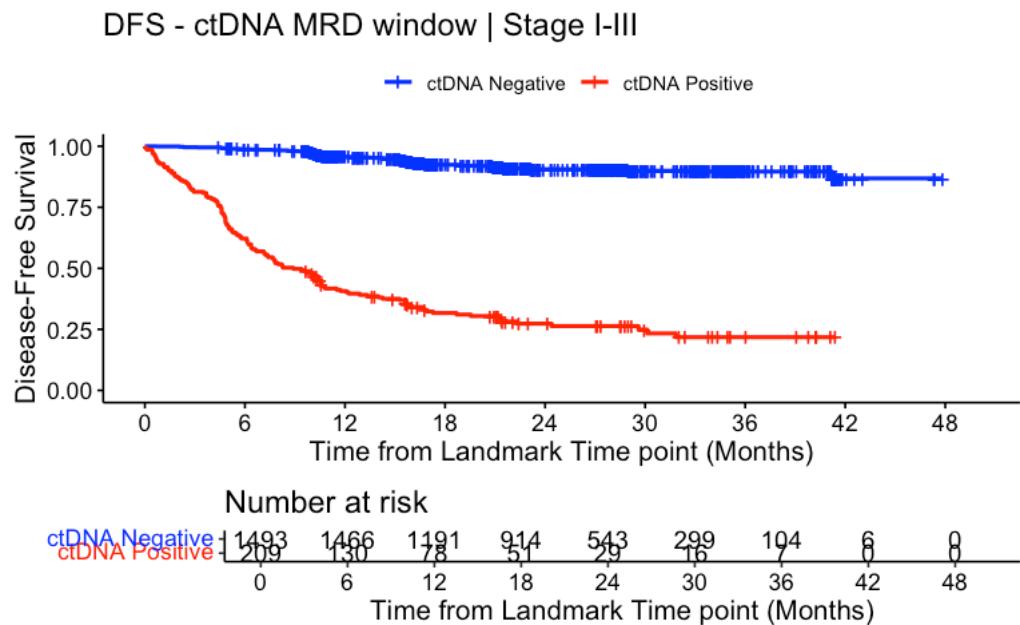
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1493	119	0.07970529	7.970529
POSITIVE	209	149	0.71291866	71.291866
NA	1	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="DFS - ctDNA MRD window | Stage I-III", ylab= "Disease-Free Survival", xlab="Time from L
andmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

1 observation deleted due to missingness
 ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	543	114	0.906	0.00864	0.888		0.922	
30	299	3	0.900	0.00935	0.880		0.917	
36	104	1	0.896	0.00987	0.875		0.914	

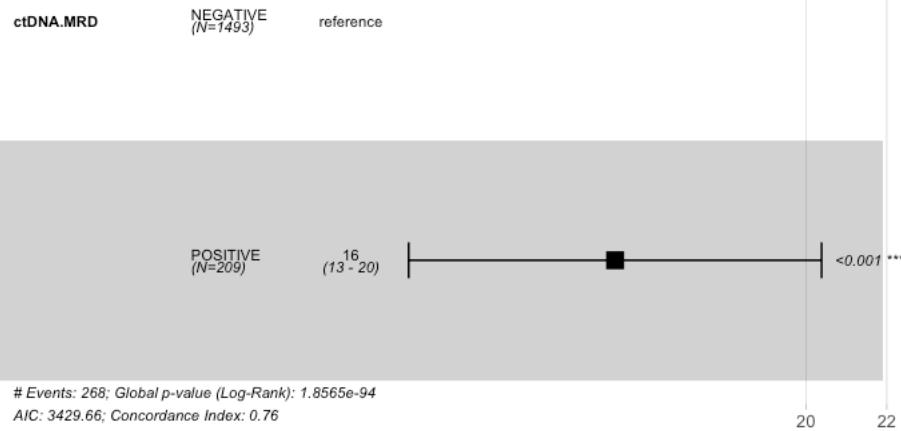
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	29	145	0.273	0.0333	0.210		0.340	
30	16	2	0.249	0.0347	0.184		0.319	
36	7	2	0.218	0.0367	0.151		0.293	

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 1702, number of events= 268
(1 observation deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.7704   15.9658   0.1243 22.29 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   15.97    0.06263    12.51    20.37

Concordance= 0.758 (se = 0.014 )
Likelihood ratio test= 425.1 on 1 df,  p=<2e-16
Wald test            = 496.7 on 1 df,  p=<2e-16
Score (logrank) test = 887.8 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 512)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 4.89310881335822997122369330866796366143543744261804739847188427503089230135526
010889580154112538230913655461810983258781049832056753645199388487490336267101e-110"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 15.97 (12.51-20.37); p = 0"
```

Hide

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

```

Pearson's Chi-squared test with Yates' continuity correction

data: contingency_table
X-squared = 549.33, df = 1, p-value < 2.2e-16

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```

Fisher's Exact Test for Count Data

```

```

data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
19.84311 41.51239
sample estimates:
odds ratio
28.54032

```

[Hide](#)

```
print(contingency_table)
```

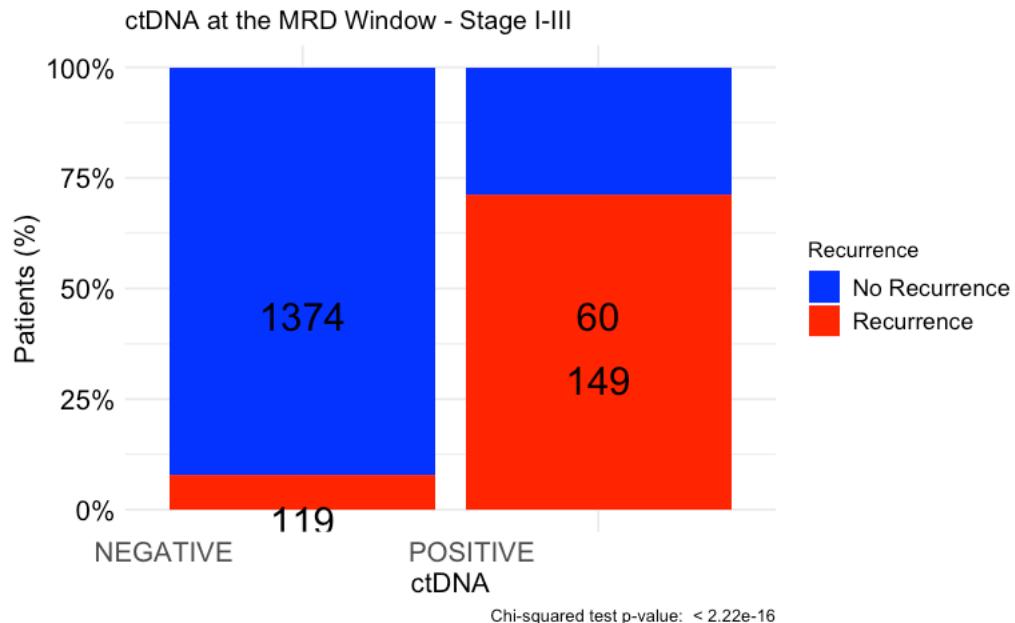
	No Recurrence	Recurrence
NEGATIVE	1374	119
POSITIVE	60	149

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage I-III",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 1.76153595425435273757103079191610411029671413537518423308786846786154308
2396972e-121"
```

#DFS by ctDNA at the MRD Window - Stage IV Landmark MRD timepoint

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	280	114	NA	26.91	NA
ctDNA.MRD=POSITIVE	127	114	2.83	2.17	4.21

[Hide](#)

```
event_summary <- circ_data %>%
group_by(ctDNA.MRD) %>%
summarise(
  Total = n(),
  Events = sum(DFS.Event),
  Fraction = Events / n(),
  Percentage = (Events / n()) * 100
)
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	280	114	0.4071429	40.71429
POSITIVE	127	114	0.8976378	89.76378

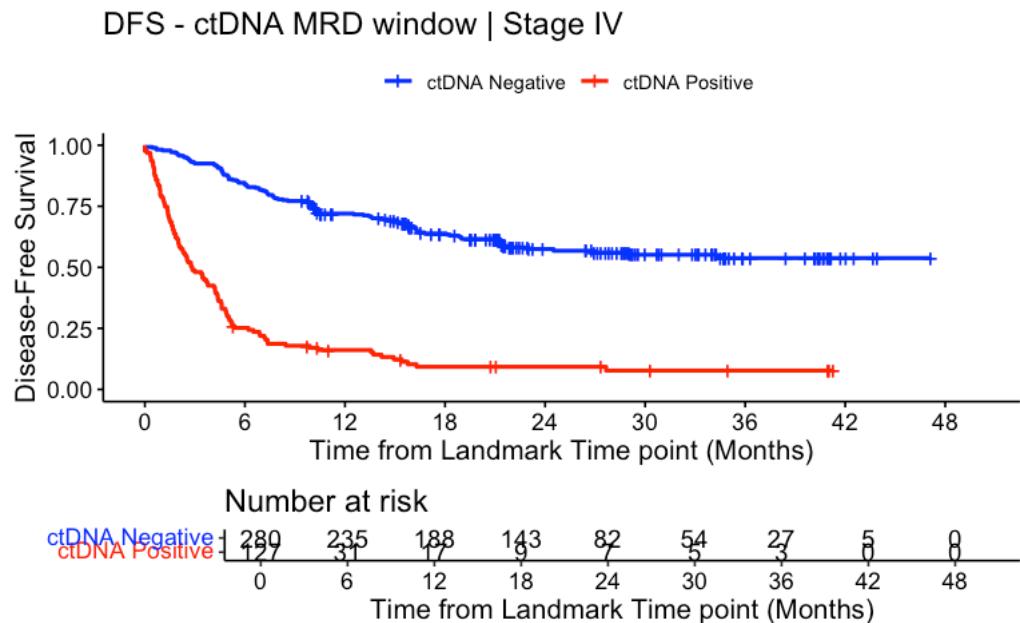
2 rows

Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | Stage IV", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

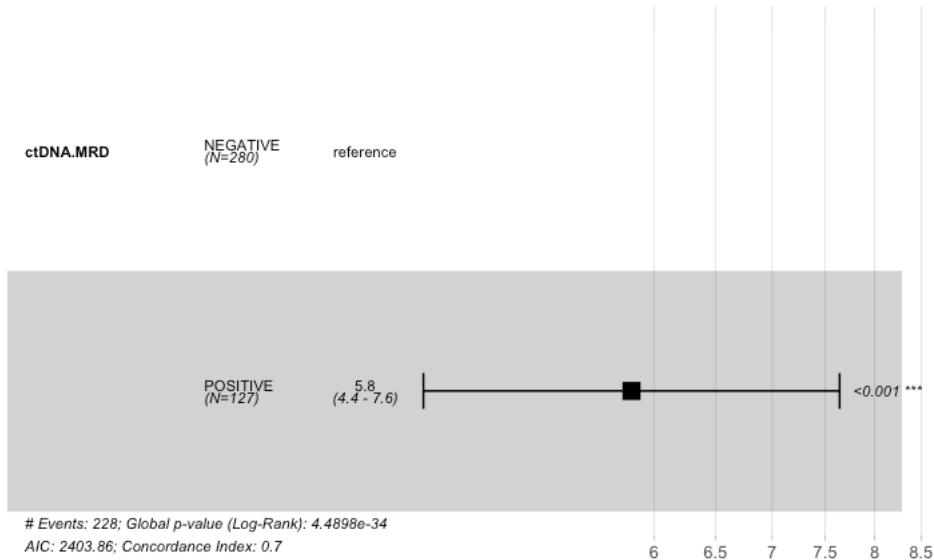
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     82     110    0.575  0.0319     0.510    0.634
  30     54      3    0.551  0.0334     0.484    0.614
  36     27      1    0.538  0.0351     0.467    0.604

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24      7     113    0.0924  0.0274     0.0479    0.155
  30      5      1    0.0770  0.0268     0.0353    0.140
  36      3      0    0.0770  0.0268     0.0353    0.140
```

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 407, number of events= 228

  coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.7624    5.8266   0.1384 12.73  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    5.827     0.1716    4.442     7.642

Concordance= 0.695  (se = 0.013 )
Likelihood ratio test= 148.1 on 1 df,  p=<2e-16
Wald test            = 162.2 on 1 df,  p=<2e-16
Score (logrank) test = 200.2 on 1 df,  p=<2e-16
```

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 3.81954596200233809460687517202118278062311637837365007953018189418730405013811
8e-37"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 5.83 (4.44-7.64); p = 0"
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 83.338, df = 1, p-value < 2.2e-16
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 6.724737 25.779487
sample estimates:
odds ratio
 12.69031
```

Hide

```
print(contingency_table)
```

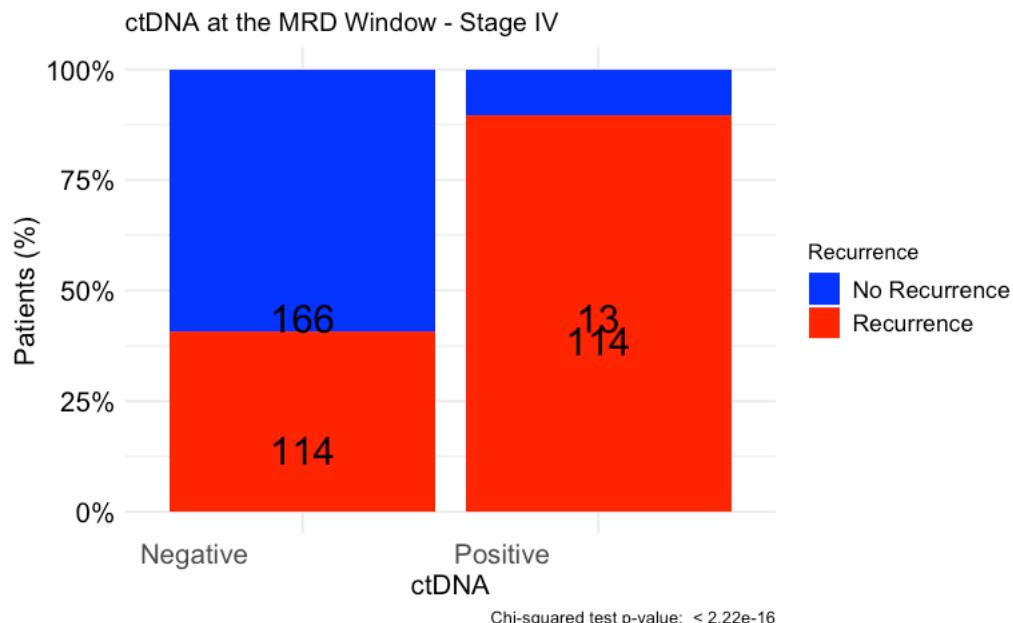
	No Recurrence	Recurrence
Negative	166	114
Positive	13	114

Hide

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage IV",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value with high precision: 6.91529701285629393380690313560733357402852436586246473622374697498571549
6586636e-20"

```

#OS by ctDNA at the MRD Window - All stages Landmark MRD timepoint

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)

```

```
Call: survfit(formula = Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event) ~
  ctDNA.MRD, data = circ_data)

1 observation deleted due to missingness
  n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 1773     36      NA      NA      NA
ctDNA.MRD=POSITIVE  336     52    43.4      NA      NA
```

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

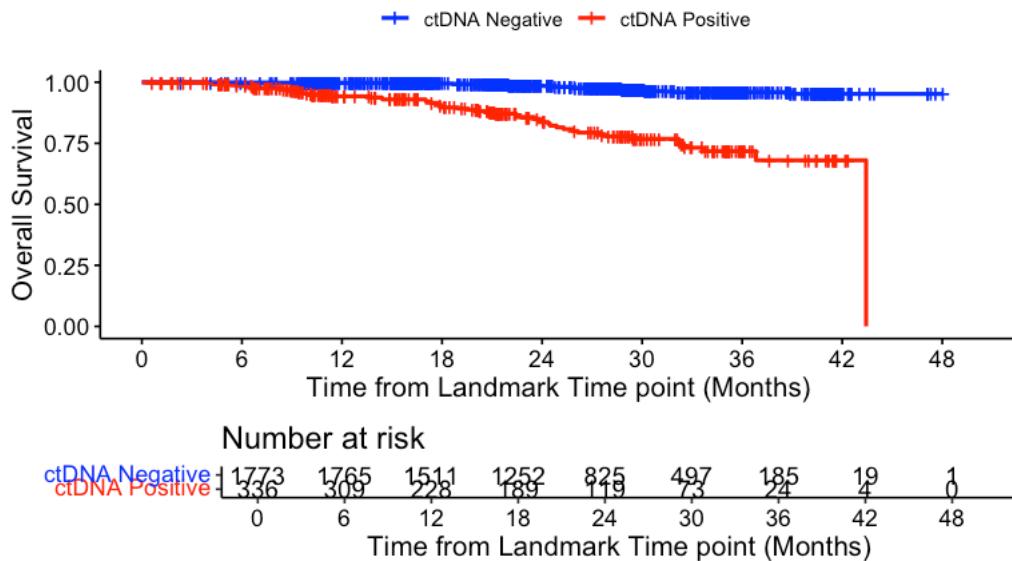
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1773	36	0.02030457	2.030457
POSITIVE	336	52	0.15476190	15.476190
NA	1	NA	NA	NA

3 rows

Hide

```
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="OS - ctDNA MRD window | All stages", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - ctDNA MRD window | All stages

[Hide](#)

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
1 observation deleted due to missingness
  ctDNA.MRD=NEGATIVE
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	825	18	0.985	0.00349	0.977		0.991	
30	497	13	0.968	0.00593	0.954		0.978	
36	185	4	0.960	0.00722	0.943		0.972	

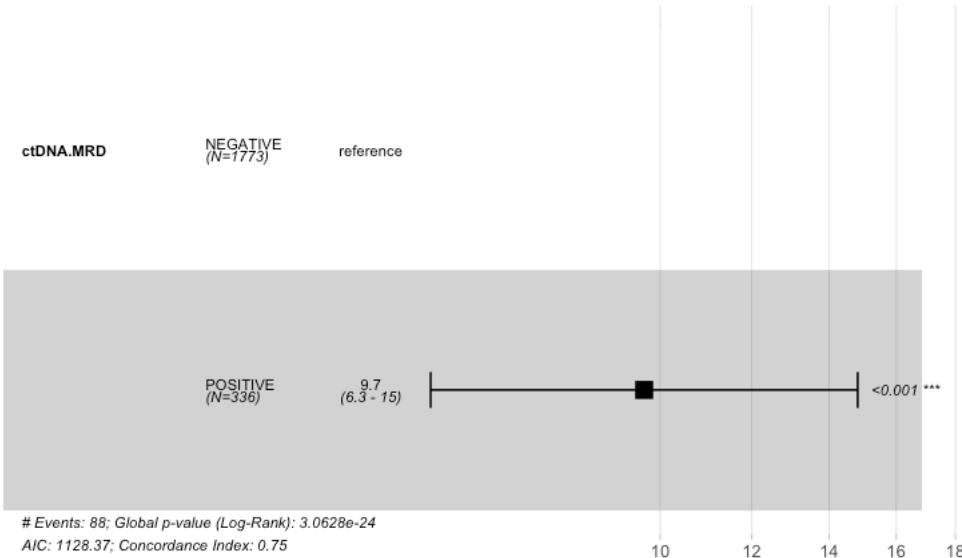
```
  ctDNA.MRD=POSITIVE
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	119	37	0.837	0.0258	0.778		0.881	
30	73	9	0.769	0.0323	0.698		0.825	
36	24	4	0.718	0.0388	0.634		0.786	

[Hide](#)

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 2109, number of events= 88
(1 observation deleted due to missingness)

      coef exp(coef)  se(coef)    z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.271     9.685    0.217 10.46   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE     9.685     0.1033    6.33    14.82

Concordance= 0.754 (se = 0.027 )
Likelihood ratio test= 103.2 on 1 df,  p=<2e-16
Wald test            = 109.5 on 1 df,  p=<2e-16
Score (logrank) test = 165.2 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 1.25377627543380720710249111149144211005682375436423693102191195216641411126895
3e-25"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.68 (6.33-14.82); p = 0"
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 124.38, df = 1, p-value < 2.2e-16
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data

data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 5.545286 14.164077
sample estimates:
odds ratio
 8.818776
```

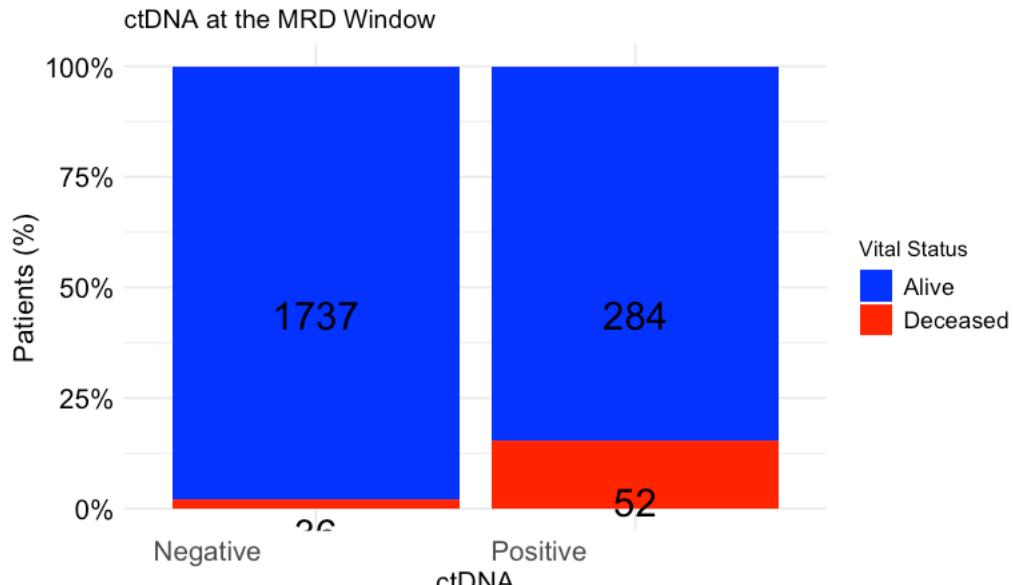
Hide

```
print(contingency_table)
```

	Alive	Deceased
Negative	1737	36
Positive	284	52

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 6.97287201605310937221154477271394463209342519378809784848414868597202716
7695626e-29"
```

#OS by ctDNA at the MRD Window - Stage I-III

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("IV")),]
circ_data$OS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~
ctDNA.MRD, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	1493	23	NA	NA	NA
ctDNA.MRD=POSITIVE	209	25	NA	NA	NA

[Hide](#)

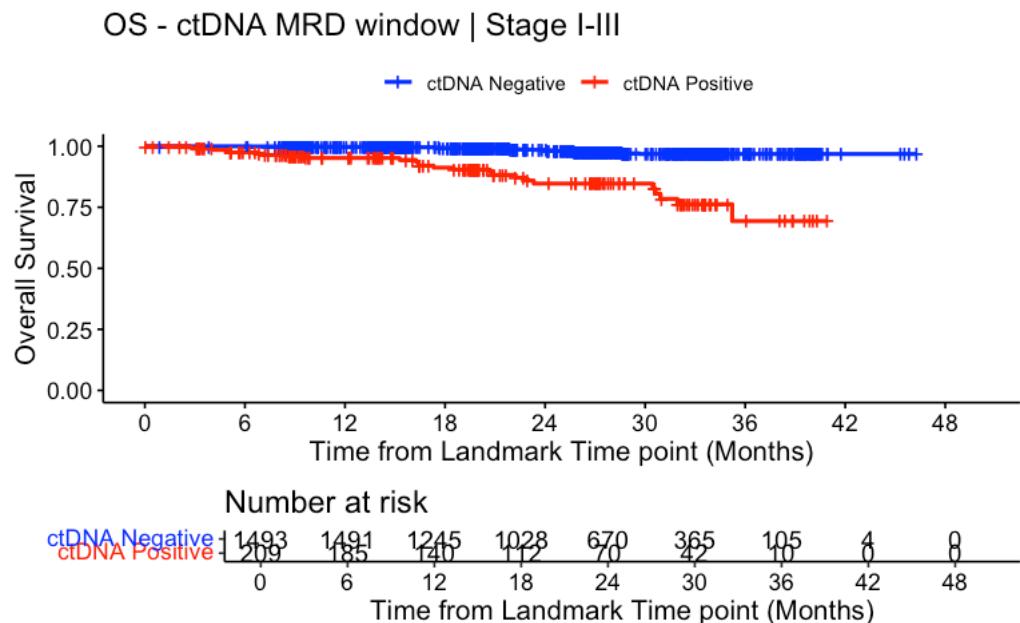
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1493	23	0.01540522	1.540522
POSITIVE	209	25	0.11961722	11.961722

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="OS - ctDNA MRD window | Stage I-III", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

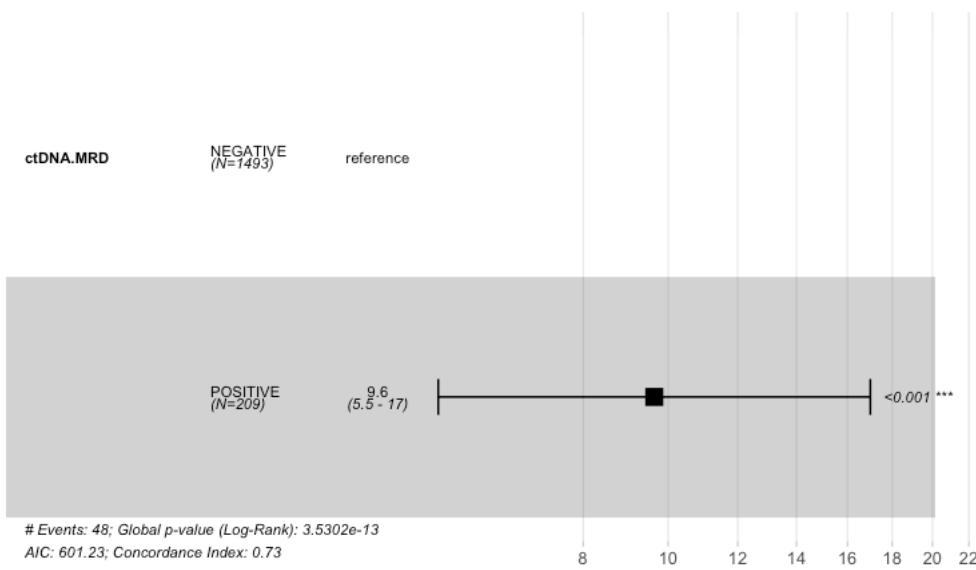
  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     670      14    0.985 0.00409      0.974    0.991
  30     365       9    0.968 0.00701      0.951    0.979
  36     105       0    0.968 0.00701      0.951    0.979

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     70       20    0.847 0.0335      0.767    0.901
  30     42       0    0.847 0.0335      0.767    0.901
  36     10       5    0.693 0.0803      0.506    0.821
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 1702, number of events= 48

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.2661    9.6416   0.2891  7.838 4.57e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    9.642     0.1037    5.471    16.99

Concordance= 0.732  (se = 0.038 )
Likelihood ratio test= 52.89  on 1 df,  p=4e-13
Wald test            = 61.44  on 1 df,  p=5e-15
Score (logrank) test = 92.58  on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 4.56766220051193151791789654154316320414768850978722447502493711275985015532122
6e-15"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.64 (5.47-16.99); p = 0"
```

Hide

```

circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

```

Pearson's Chi-squared test with Yates' continuity correction

data: contingency_table
X-squared = 68.895, df = 1, p-value < 2.2e-16

```

Hide

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```

Fisher's Exact Test for Count Data

data: contingency_table
p-value = 1.07e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
4.614397 16.339262
sample estimates:
odds ratio
8.662461

```

Hide

```
print(contingency_table)
```

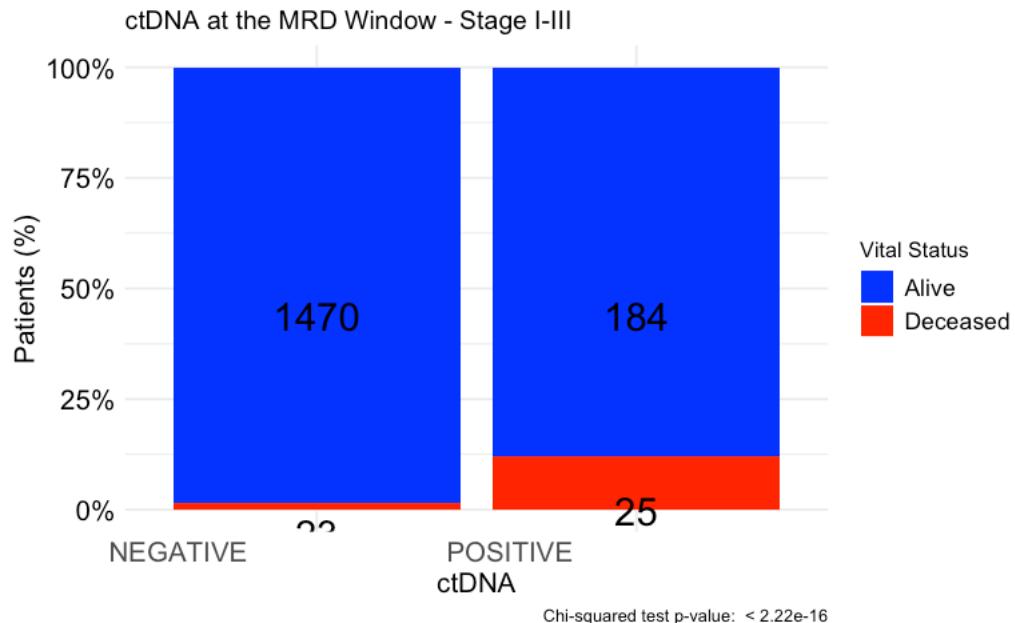
	Alive	Deceased
NEGATIVE	1470	23
POSITIVE	184	25

Hide

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window - Stage I-III",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



```
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 1.03823895795462406519229774044888794352451295019759691662386558164143934
8459244e-16"
```

#OS by ctDNA at the MRD Window - Stage IV

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III"))]
circ_data$OS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~
ctDNA.MRD, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	280	13	NA	NA	NA
ctDNA.MRD=POSITIVE	125	27	41.8	NA	NA

```
event_summary <- circ_data %>%
group_by(ctDNA.MRD) %>%
summarise(
  Total = n(),
  Events = sum(OS.Event),
  Fraction = Events / n(),
  Percentage = (Events / n()) * 100
)
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	280	13	0.04642857	4.642857
POSITIVE	125	27	0.21600000	21.600000

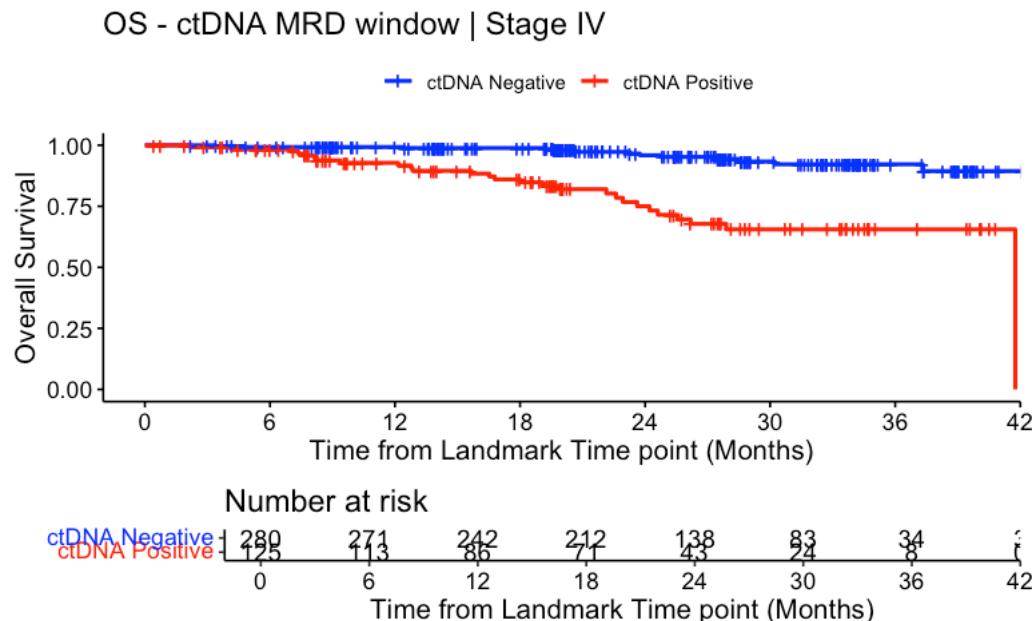
2 rows

Hide

```

surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="OS - ctDNA MRD window | Stage IV", ylab= "Overall Survival", xlab="Time from Landmark T
ime point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```



```
summary(KM_curve, times= c(24, 30, 36))
```

Hide

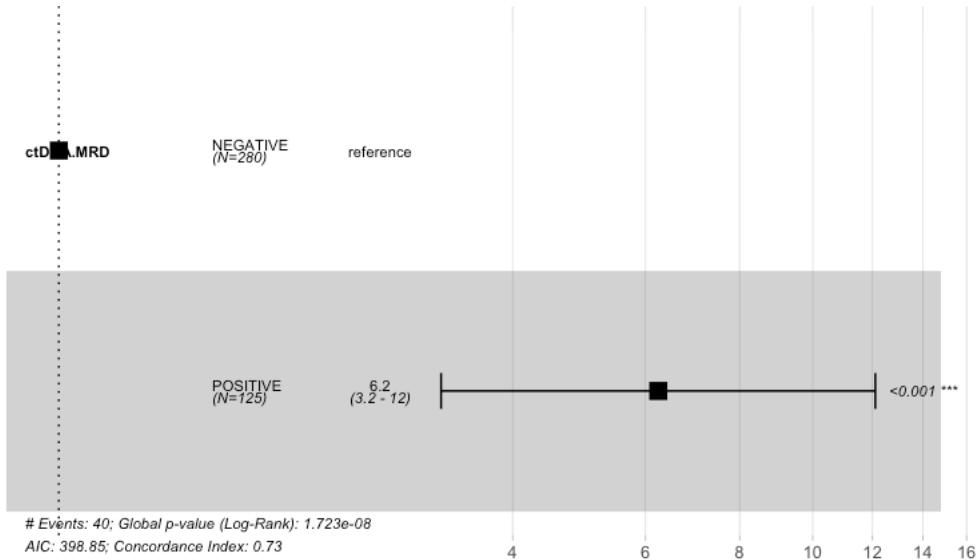
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24    138      8    0.959  0.0147      0.918    0.980
  30     83      3    0.934  0.0205      0.879    0.964
  36     34      1    0.922  0.0232      0.862    0.957

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     43     21    0.750  0.0499      0.636    0.833
  30     24      5    0.655  0.0591      0.526    0.757
  36      8      0    0.655  0.0591      0.526    0.757
```

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 405, number of events= 40

  coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.8308    6.2388  0.3384 5.411 6.28e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    6.239     0.1603    3.214    12.11

Concordance= 0.73  (se = 0.037 )
Likelihood ratio test= 31.78 on 1 df,  p=2e-08
Wald test            = 29.27 on 1 df,  p=6e-08
Score (logrank) test = 38.28 on 1 df,  p=6e-10
```

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

[1] "HR = 6.24 (3.21-12.11); p = 0"

[Hide](#)

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 26.045, df = 1, p-value = 3.336e-07
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 6.728e-07
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.681627 12.393229
sample estimates:
odds ratio
 5.630069
```

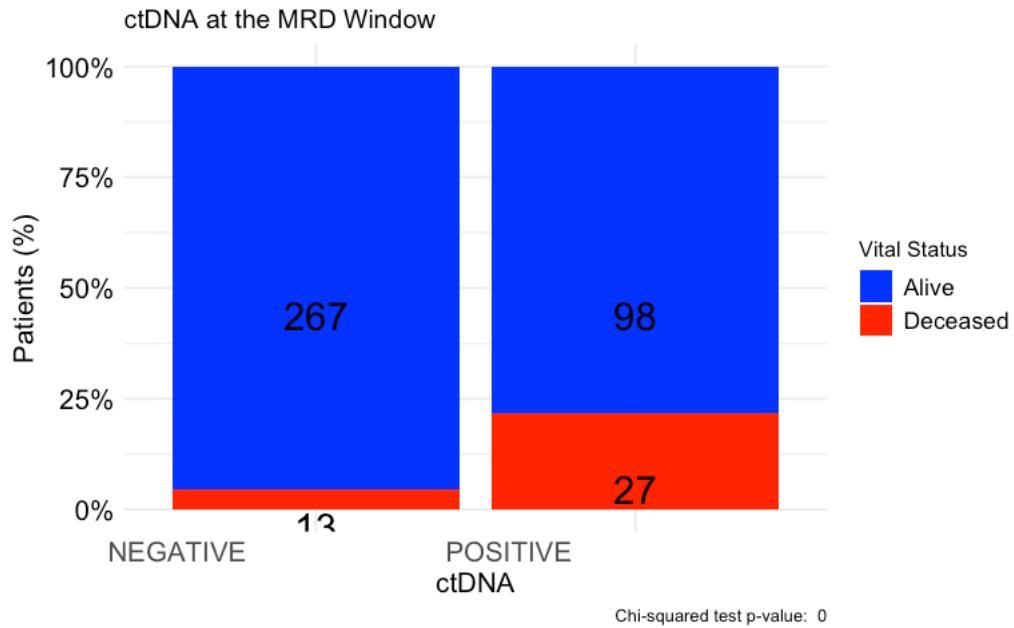
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
NEGATIVE	267	13
POSITIVE	98	27

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#Multivariate cox regression at MRD Window for DFS - All stages Landmark MRD timepoint

Hide

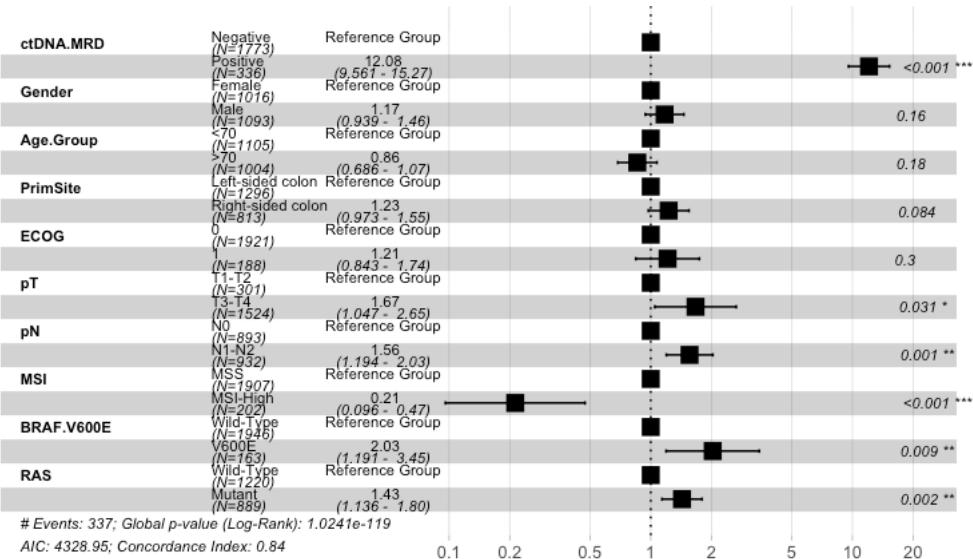
```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"), labels = c("Negative", "Positive"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", ">70"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels = c("0", "1"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"), labels = c("MSS", "MSI-High"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("Wild-Type", "V600E"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("Wild-Type", "Mutant"))
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ctDNA.MRD + Gender + Age.Group + PrimSite + ECOG + pT + pN + MSI + BRAF.V600E + RAS, data=circ_data)
ggforest(cox_fit, data = circ_data, main = "Multivariate Regression Model for DFS - All Stages", refLabel = "Reference Group")

```

Multivariate Regression Model for DFS - All Stages



```
test.ph <- cox.zph(cox_fit)  
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.MRD + Gender + Age.Group +
    PrimSite + ECOG + pT + pN + MSI + BRAF.V600E + RAS, data = circ_data)

n= 1824, number of events= 337
(286 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.MRDPPositive    2.4918   12.0825   0.1194 20.865 < 2e-16 ***
GenderMale             0.1577    1.1708   0.1124  1.403 0.160477
Age.Group>70          -0.1532    0.8580   0.1143 -1.340 0.180165
PrimSiteRight-sided colon 0.2046   1.2270   0.1183  1.729 0.083802 .
ECOG1                  0.1920    1.2117   0.1854  1.036 0.300192
pTT3-T4                0.5113    1.6674   0.2373  2.154 0.031218 *
pNN1-N2                0.4429    1.5572   0.1357  3.264 0.001098 **
MSIMSI-High            -1.5477   0.2127   0.4069 -3.803 0.000143 ***
BRAF.V600EV600E        0.7070    2.0280   0.2717  2.602 0.009261 **
RASMutant              0.3566    1.4284   0.1168  3.053 0.002267 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPPositive    12.0825   0.08276  9.56104 15.2689
GenderMale             1.1708    0.85409  0.93938 1.4593
Age.Group>70          0.8580    1.16552  0.68581 1.0734
PrimSiteRight-sided colon 1.2270   0.81499  0.97305 1.5473
ECOG1                  1.2117    0.82527  0.84260 1.7426
pTT3-T4                1.6674    0.59973  1.04720 2.6550
pNN1-N2                1.5572    0.64220  1.19356 2.0315
MSIMSI-High            0.2127    4.70083  0.09582 0.4723
BRAF.V600EV600E        2.0280    0.49311  1.19067 3.4541
RASMutant              1.4284    0.70008  1.13614 1.7958

Concordance= 0.836 (se = 0.01 )
Likelihood ratio test= 587.1 on 10 df,  p=<2e-16
Wald test               = 609.6 on 10 df,  p=<2e-16
Score (logrank) test = 1076 on 10 df,  p=<2e-16

```

Hide

```

cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPPositive", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 512)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value for ctDNA.MRD: 1.10491338314482535047684564465763026763567054940930072955327678379182870575438
158895286003945965111479324501573041745064343378898409260175641308997605671827e-96"

```

```
#Multivariate cox regression at MRD Window for OS - All stages Landmark MRD timepoint
```

Hide

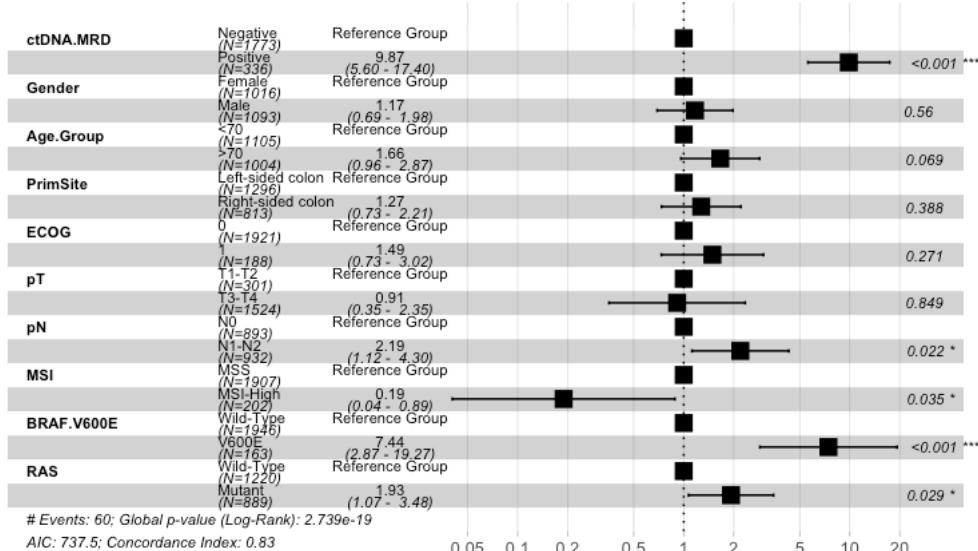
```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", ">70"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels = c("0", "1"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"), labels = c("MSS", "MSI-High"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("Wild-Type", "V600E"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("Wild-Type", "Mutant"))
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
cox_fit <- coxph(surv_object ~ ctDNA.MRD + Gender + Age.Group + PrimSite + ECOG + pT + pN + MSI + BRAF.V600E + RAS, data=circ_data)
ggforest(cox_fit, data = circ_data, main = "Multivariate Regression Model for OS - All Stages", refLabel = "Reference Group")

```

Multivariate Regression Model for OS - All Stages



Hide

```
test.ph <- cox.zph(cox_fit)
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD + Gender + Age.Group +
    PrimSite + ECOG + pT + pN + MSI + BRAF.V600E + RAS, data = circ_data)

n= 1824, number of events= 60
(286 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.MRDPositive 2.2892  9.8668  0.2894  7.911 2.55e-15 ***
GenderMale          0.1567  1.1696  0.2685  0.584  0.5595
Age.Group>70        0.5072  1.6607  0.2787  1.820  0.0688 .
PrimSiteRight-sided colon 0.2426  1.2746  0.2810  0.863  0.3879
ECOG1               0.3982  1.4891  0.3615  1.101  0.2707
pTT3-T4             -0.0915  0.9126  0.4818 -0.190  0.8494
pNN1-N2             0.7855  2.1936  0.3428  2.291  0.0219 *
MSIMSI-High        -1.6650  0.1892  0.7875 -2.114  0.0345 *
BRAF.V600EV600E    2.0072  7.4428  0.4855  4.134  3.56e-05 ***
RASMutant           0.6564  1.9278  0.3007  2.183  0.0290 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPositive  9.8669  0.1013  5.59590 17.3975
GenderMale          1.1696  0.8550  0.69108 1.9795
Age.Group>70        1.6607  0.6021  0.96175 2.8677
PrimSiteRight-sided colon 1.2746  0.7846  0.73478 2.2111
ECOG1               1.4891  0.6716  0.73322 3.0241
pTT3-T4             0.9126  1.0958  0.35492 2.3463
pNN1-N2             2.1936  0.4559  1.12028 4.2953
MSIMSI-High        0.1892  5.2856  0.04041 0.8857
BRAF.V600EV600E    7.4428  0.1344  2.87398 19.2748
RASMutant           1.9278  0.5187  1.06931 3.4757

Concordance= 0.832 (se = 0.033 )
Likelihood ratio test= 111.4 on 10 df,  p=<2e-16
Wald test           = 108.2 on 10 df,  p=<2e-16
Score (logrank) test = 161.4 on 10 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
#DFS by ACT treatment in MRD negative - High Risk Stage II/III
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Hide

```
Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
    ACT, data = circ_data)
```

```
15 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
ACT=FALSE 586    50     NA     NA     NA
ACT=TRUE  571    55     NA     NA     NA
```

Hide

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

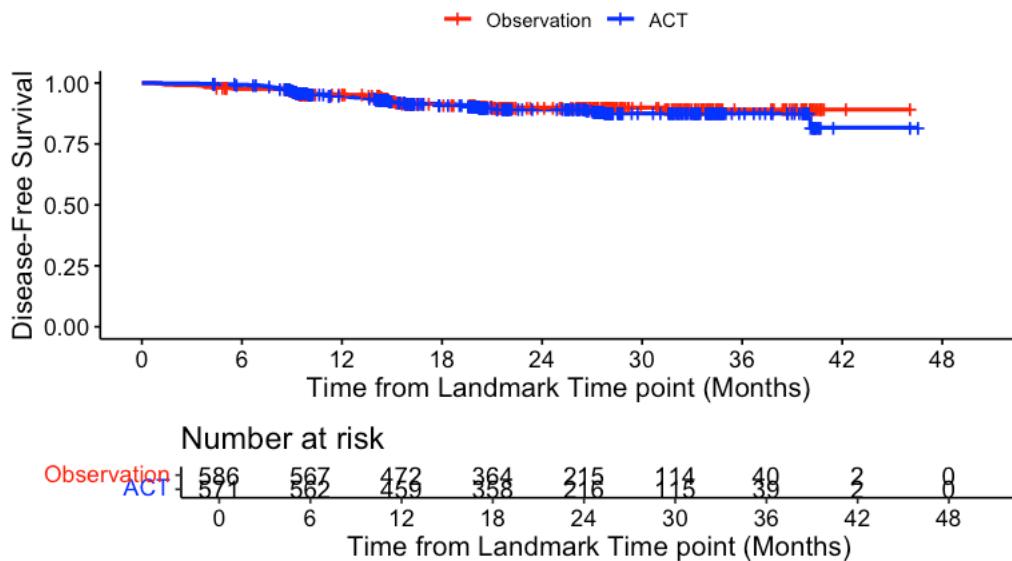
ACT <lgI>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	586	50	0.08532423	8.532423
TRUE	571	55	0.09632224	9.632224
NA	15	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | High Risk Stage II/III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title = "")
```

DFS - ctDNA MRD Negative ACT vs Observation | High Risk Stage



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

15 observations deleted due to missingness

ACT=FALSE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.000	215.000	49.000	0.899	0.014	0.868		0.923	

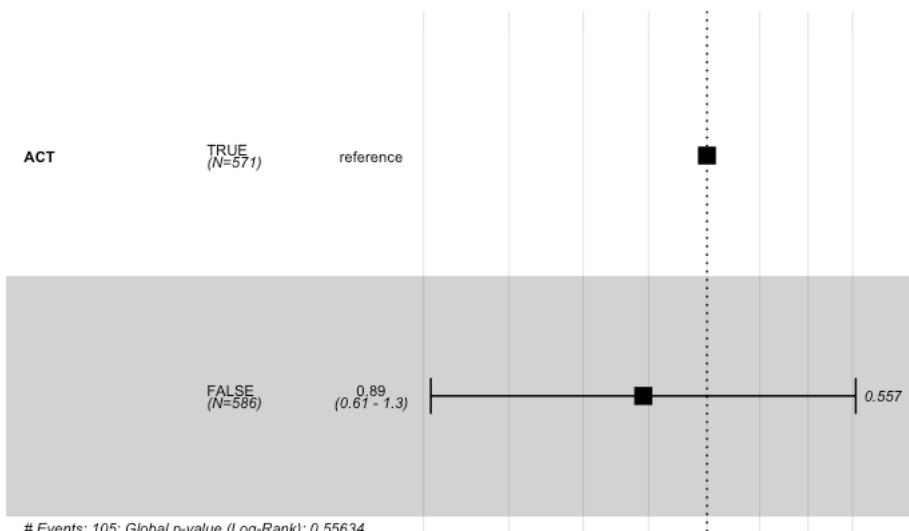
ACT=TRUE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	216.0000	51.0000	0.8911	0.0148	0.8581		0.9168	

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 1157, number of events= 105
(15 observations deleted due to missingness)

  coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE -0.1149    0.8915   0.1954 -0.588    0.557

  exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    0.8915      1.122    0.6078    1.307

Concordance= 0.508 (se = 0.025 )
Likelihood ratio test= 0.35 on 1 df,  p=0.6
Wald test            = 0.35 on 1 df,  p=0.6
Score (logrank) test = 0.35 on 1 df,  p=0.6

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 0.89 (0.61-1.31); p = 0.557"
```

[Hide](#)

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

```

Pearson's Chi-squared test with Yates' continuity correction

data: contingency_table
X-squared = 0.30112, df = 1, p-value = 0.5832

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```

Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.5401
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.5730366 1.3343425
sample estimates:
odds ratio
0.8752713

```

[Hide](#)

```
print(contingency_table)
```

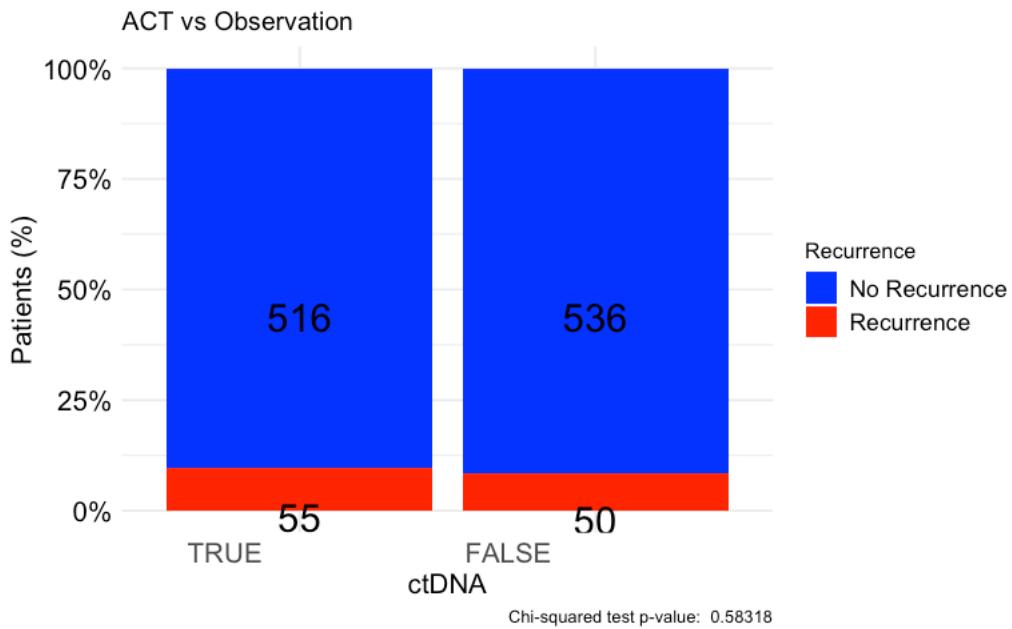
	No Recurrence	Recurrence
TRUE	516	55
FALSE	536	50

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, ECOG and pathological stage
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + Stage + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + Stage +
  ECOG, data = circ_data)

n= 1157, number of events= 105
(15 observations deleted due to missingness)

      coef exp(coef) se(coef)     z Pr(>|z|)    
ACTFALSE    0.3623   1.4367   0.2145  1.689   0.0911 .  
GenderMale    0.1477   1.1591   0.1960  0.753   0.4512  
Age.Group≥70 -0.3075   0.7353   0.2067 -1.487   0.1369  
StageIII     1.0528   2.8656   0.2528  4.164  3.13e-05 ***  
ECOG1        0.2435   1.2756   0.3168  0.769   0.4422  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    1.4367   0.6961   0.9436   2.187
GenderMale    1.1591   0.8627   0.7894   1.702
Age.Group≥70  0.7353   1.3600   0.4903   1.103
StageIII     2.8656   0.3490   1.7458   4.704
ECOG1        1.2756   0.7839   0.6856   2.373

Concordance= 0.629  (se = 0.026 )
Likelihood ratio test= 23.38 on 5 df,  p=3e-04
Wald test      = 21.22 on 5 df,  p=7e-04
Score (logrank) test = 22.35 on 5 df,  p=4e-04

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + Stage + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + Stage +
  ECOG, data = circ_data)

n= 1157, number of events= 105
(15 observations deleted due to missingness)

      coef exp(coef) se(coef)     z Pr(>|z|)    
ACTTRUE     -0.3623   0.6961   0.2145 -1.689   0.0911 .
GenderMale    0.1477   1.1591   0.1960   0.753   0.4512  
Age.Group≥70 -0.3075   0.7353   0.2067 -1.487   0.1369  
StageIII     1.0528   2.8656   0.2528   4.164 3.13e-05 ***
ECOG1        0.2435   1.2756   0.3168   0.769   0.4422  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.6961    1.4367   0.4572    1.060
GenderMale    1.1591    0.8627   0.7894    1.702
Age.Group≥70  0.7353    1.3600   0.4903    1.103
StageIII      2.8656    0.3490   1.7458    4.704
ECOG1        1.2756    0.7839   0.6856    2.373

Concordance= 0.629 (se = 0.026 )
Likelihood ratio test= 23.38 on 5 df,  p=3e-04
Wald test      = 21.22 on 5 df,  p=7e-04
Score (logrank) test = 22.35 on 5 df,  p=4e-04

```

#DFS by ACT treatment in MRD positive - High Risk Stage II/III

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ACT, data = circ_data)

1 observation deleted due to missingness
  n events median 0.95LCL 0.95UCL
ACT=FALSE 47      45    3.55    3.16    3.95
ACT=TRUE  145     88   12.06   9.30   18.57

```

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

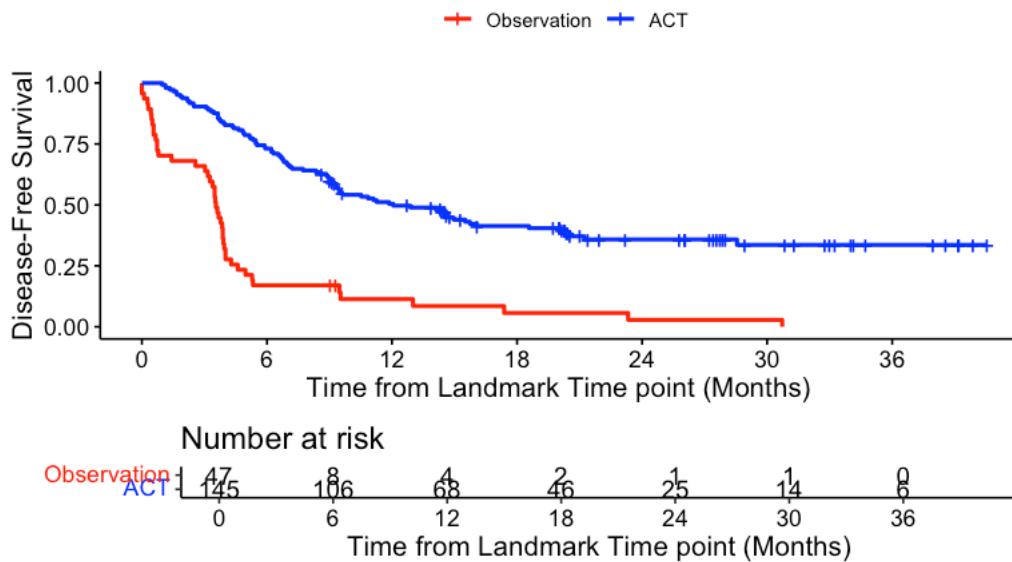
ACT <lgl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	47	45	0.9574468	95.74468
TRUE	145	88	0.6068966	60.68966
NA	1	NA	NA	NA
3 rows				

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palett
e=c("red","blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | High Risk Stage II/III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title
="")

```

DFS - ctDNA MRD Positive ACT vs Observation | High Risk Stage



```
summary(KM_curve, times= c(24))
```

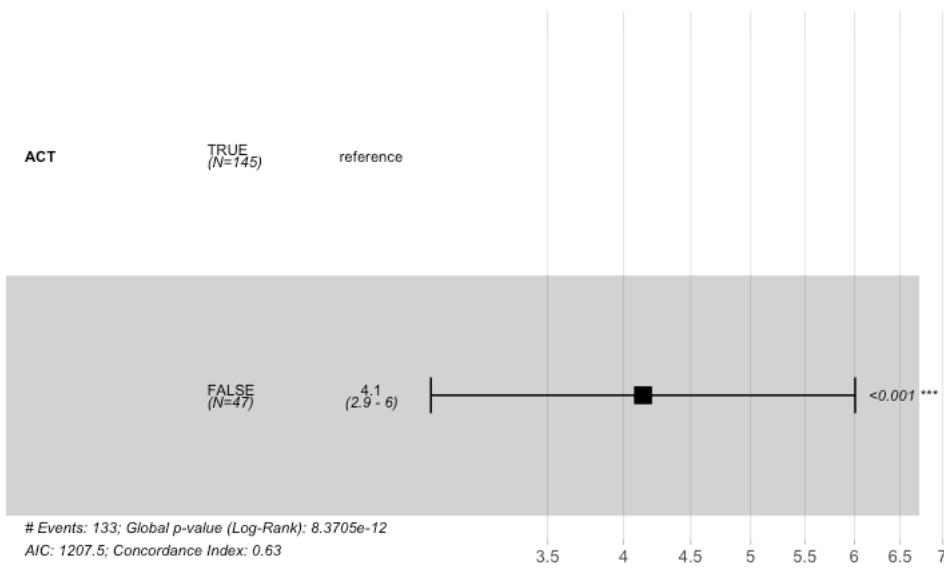
```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
conf.type = "log-log")
```

```
1 observation deleted due to missingness
```

ACT=FALSE	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00000	1.00000	44.00000	0.02837	0.02746	0.00232	0.12350	
ACT=TRUE	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00000	25.00000	87.00000	0.3583	0.0435	0.2741	0.4432	

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 192, number of events= 133
(1 observation deleted due to missingness)

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 1.4203    4.1382   0.1901 7.472 7.91e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     4.138      0.2417    2.851     6.006

Concordance= 0.634  (se = 0.019 )
Likelihood ratio test= 46.68  on 1 df,  p=8e-12
Wald test           = 55.83  on 1 df,  p=8e-14
Score (logrank) test = 64.71  on 1 df,  p=9e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 4.14 (2.85-6.01); p = 0"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 18.877, df = 1, p-value = 1.394e-05
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 1.043e-06
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 3.527208 127.775372
sample estimates:
odds ratio
 14.42898
```

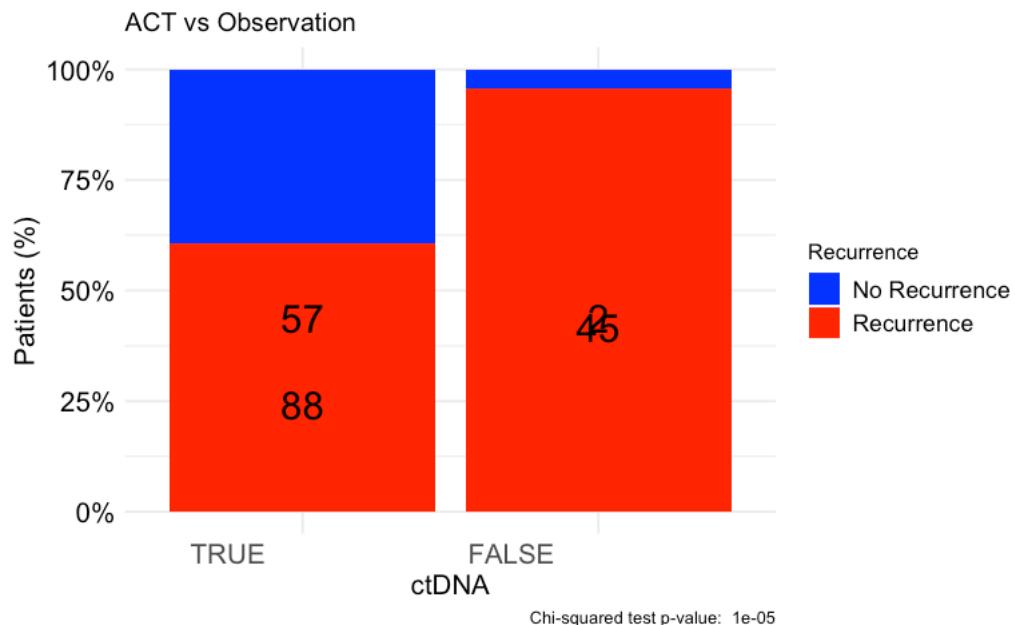
[Hide](#)

```
print(contingency_table)
```

	No	Recurrence
TRUE	57	88
FALSE	2	45

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI and pathological stage
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + Stage + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + Stage +
  ECOG, data = circ_data)

n= 192, number of events= 133
(1 observation deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTFALSE     1.46226  4.31571  0.20651  7.081 1.43e-12 ***
GenderMale   -0.06402  0.93799  0.18183 -0.352   0.725    
Age.Group≥70  0.03736  1.03807  0.18637  0.200   0.841    
StageIII     0.31989  1.37697  0.23571  1.357   0.175    
ECOG1        0.05652  1.05814  0.28089  0.201   0.841    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     4.316    0.2317    2.8792    6.469
GenderMale   0.938    1.0661    0.6568    1.340
Age.Group≥70  1.038    0.9633    0.7204    1.496
StageIII     1.377    0.7262    0.8675    2.186
ECOG1        1.058    0.9451    0.6102    1.835

Concordance= 0.644 (se = 0.026 )
Likelihood ratio test= 49.19 on 5 df,  p=2e-09
Wald test      = 58.77 on 5 df,  p=2e-11
Score (logrank) test = 67.68 on 5 df,  p=3e-13

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + Stage + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + Stage +
  ECOG, data = circ_data)

n= 192, number of events= 133
(1 observation deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTTRUE      -1.46226  0.23171  0.20651 -7.081 1.43e-12 ***
GenderMale   -0.06402  0.93799  0.18183 -0.352   0.725    
Age.Group≥70  0.03736  1.03807  0.18637  0.200   0.841    
StageIII     0.31989  1.37697  0.23571  1.357   0.175    
ECOG1        0.05652  1.05814  0.28089  0.201   0.841    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.2317    4.3157   0.1546   0.3473  
GenderMale   0.9380    1.0661   0.6568   1.3396  
Age.Group≥70  1.0381    0.9633   0.7204   1.4958  
StageIII     1.3770    0.7262   0.8675   2.1855  
ECOG1        1.0581    0.9451   0.6102   1.8350  

Concordance= 0.644 (se = 0.026 )
Likelihood ratio test= 49.19 on 5 df,  p=2e-09
Wald test      = 58.77 on 5 df,  p=2e-11
Score (logrank) test = 67.68 on 5 df,  p=3e-13

```

#DFS by ACT treatment in MRD negative - High Risk Stage II

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ACT, data = circ_data)

1588 observations deleted due to missingness
  n events median 0.95LCL 0.95UCL
ACT=FALSE 373    21    NA    NA    NA
ACT=TRUE  102     3    NA    NA    NA

```

[Hide](#)

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ACT <gl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	373	21	0.05630027	5.630027
TRUE	102	3	0.02941176	2.941176
NA	1588	NA	NA	NA

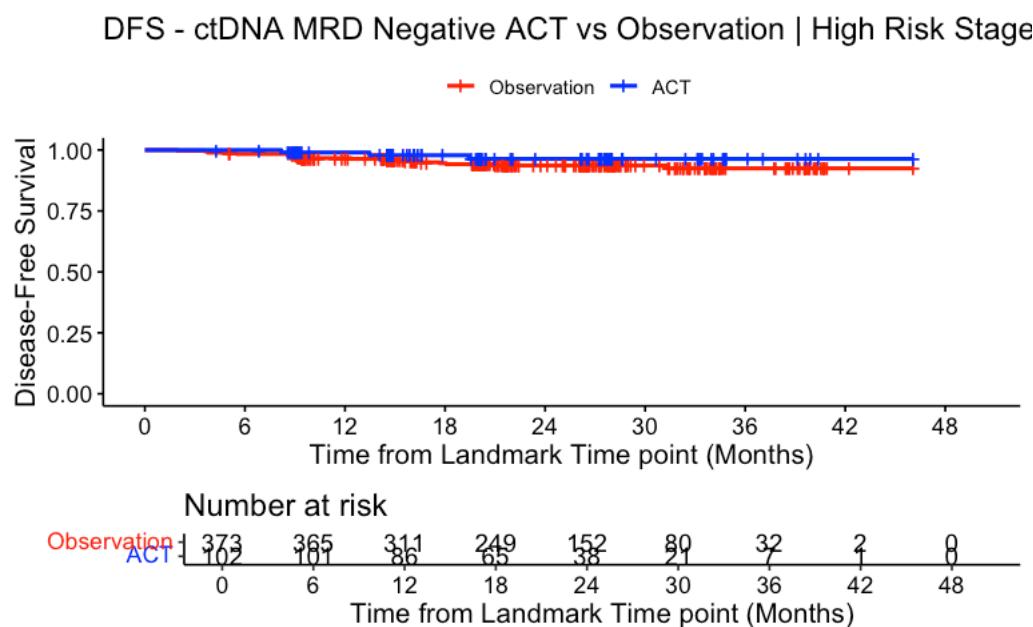
3 rows

[Hide](#)

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | High Risk Stage II", ylab= "Disease-Free
Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")

```



```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
conf.type = "log-log")
```

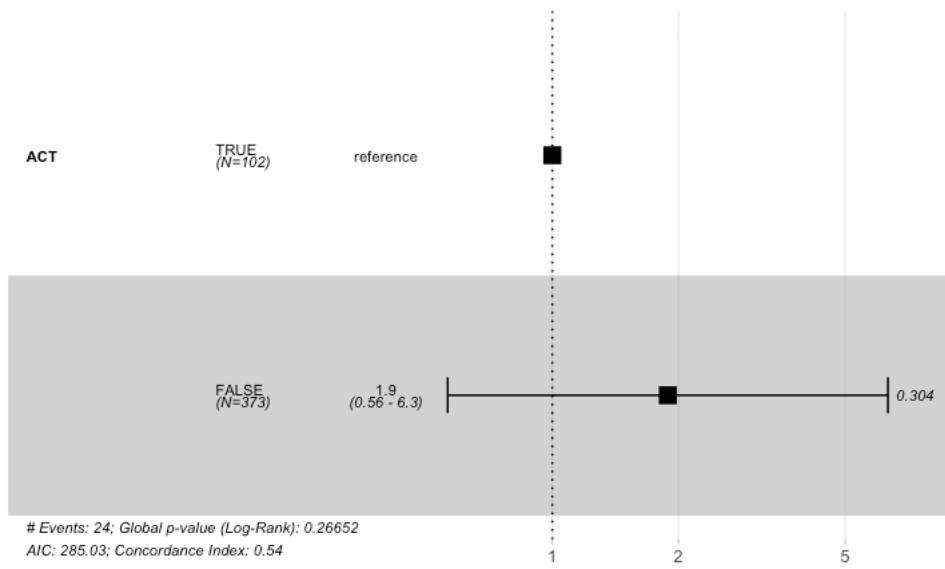
1588 observations deleted due to missingness

```
ACT=FALSE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.000  152.000   20.000     0.937     0.014    0.903    0.959
```

```
ACT=TRUE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.000  38.0000   3.0000     0.9634    0.0211   0.8890   0.9883
```

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 475, number of events= 24
(1588 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ACTFALSE	0.6344	1.8860	0.6173	1.028	0.304
					exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE	1.886	0.5302	0.5625	6.323	

Concordance= 0.544 (se = 0.035)
Likelihood ratio test= 1.23 on 1 df, p=0.3
Wald test = 1.06 on 1 df, p=0.3
Score (logrank) test = 1.09 on 1 df, p=0.3

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.89 (0.56-6.32); p = 0.304"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 0.71169, df = 1, p-value = 0.3989
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.4423
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.5696505 10.5052330
sample estimates:
odds ratio
1.966313
```

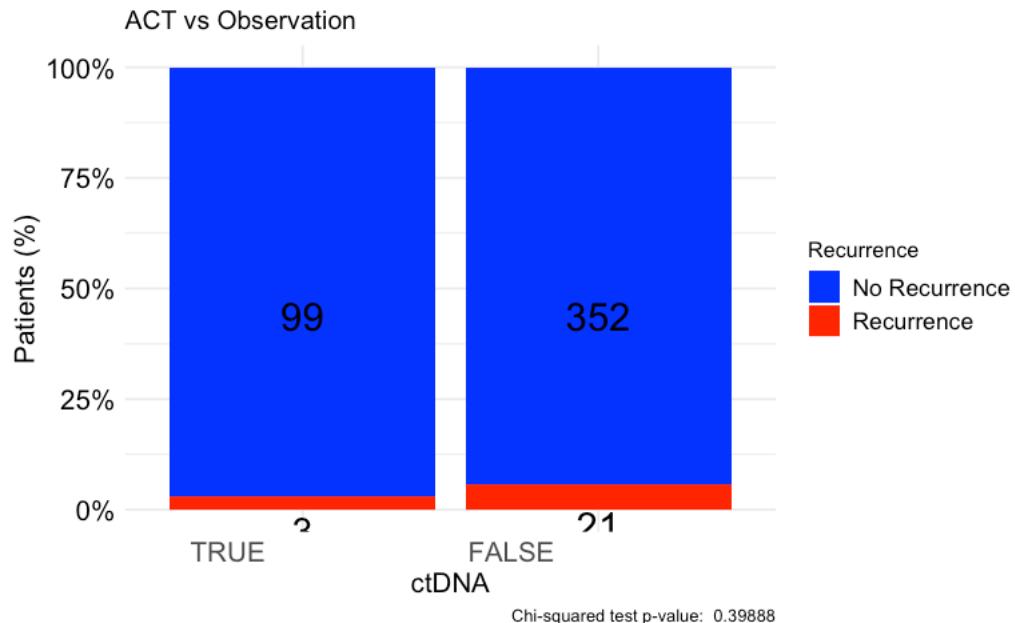
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
TRUE	99	3
FALSE	352	21

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 475, number of events= 24
(1588 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE      0.7519    2.1211   0.6266  1.200   0.2301
GenderMale    -0.1514    0.8595   0.4160 -0.364   0.7159
Age.Group≥70 -0.8105    0.4446   0.4420 -1.834   0.0667 .
ECOG1         0.5506    1.7343   0.5794   0.950   0.3419
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE      2.1211     0.4715   0.6212    7.243
GenderMale    0.8595     1.1634   0.3803    1.943
Age.Group≥70  0.4446     2.2490   0.1870    1.057
ECOG1         1.7343     0.5766   0.5571    5.399

Concordance= 0.629 (se = 0.06 )
Likelihood ratio test= 4.98 on 4 df,  p=0.3
Wald test          = 4.66 on 4 df,  p=0.3
Score (logrank) test = 4.79 on 4 df,  p=0.3

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))

circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 475, number of events= 24
(1588 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE     -0.7519   0.4715   0.6266 -1.200   0.2301
GenderMale   -0.1514   0.8595   0.4160 -0.364   0.7159
Age.Group≥70 -0.8105   0.4446   0.4420 -1.834   0.0667 .
ECOG1        0.5506   1.7343   0.5794   0.950   0.3419
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.4715     2.1211   0.1381    1.610
GenderMale   0.8595     1.1634   0.3803    1.943
Age.Group≥70  0.4446     2.2490   0.1870    1.057
ECOG1        1.7343     0.5766   0.5571    5.399

Concordance= 0.629 (se = 0.06 )
Likelihood ratio test= 4.98 on 4 df,  p=0.3
Wald test          = 4.66 on 4 df,  p=0.3
Score (logrank) test = 4.79 on 4 df,  p=0.3

```

#DFS by ACT treatment in MRD positive - High Risk Stage II

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
ACT, data = circ_data)

1588 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
ACT=FALSE 15      14    3.52    3.39      NA
ACT=TRUE  23      10     NA    9.30      NA

```

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

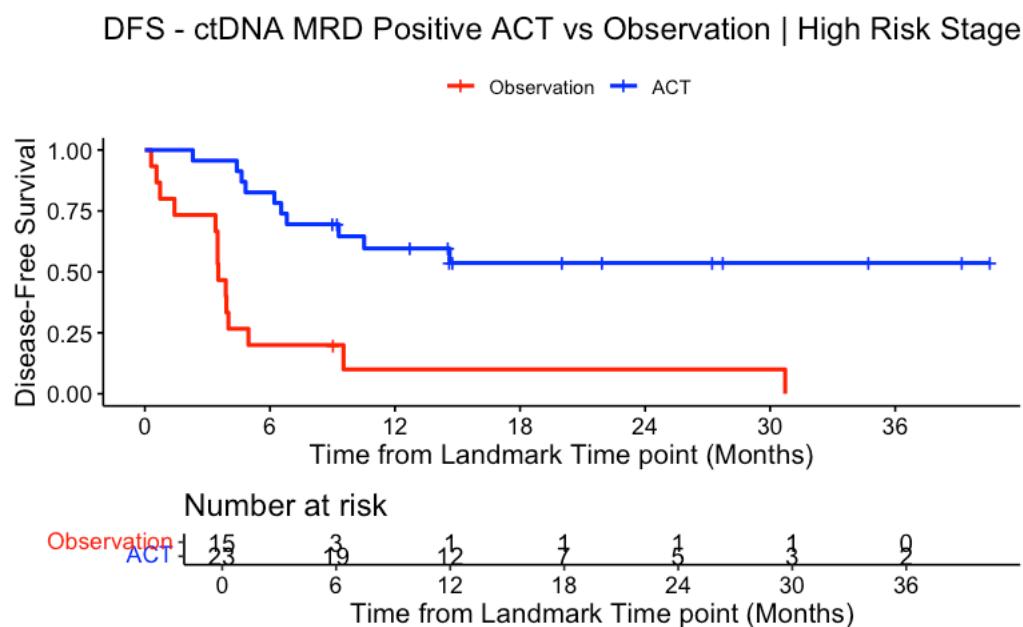
ACT	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
FALSE	15	14	0.9333333	93.33333
TRUE	23	10	0.4347826	43.47826
NA	1588	NA	NA	NA

3 rows

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("red","blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | High Risk Stage II", ylab= "Disease-Free
Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")

```



```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

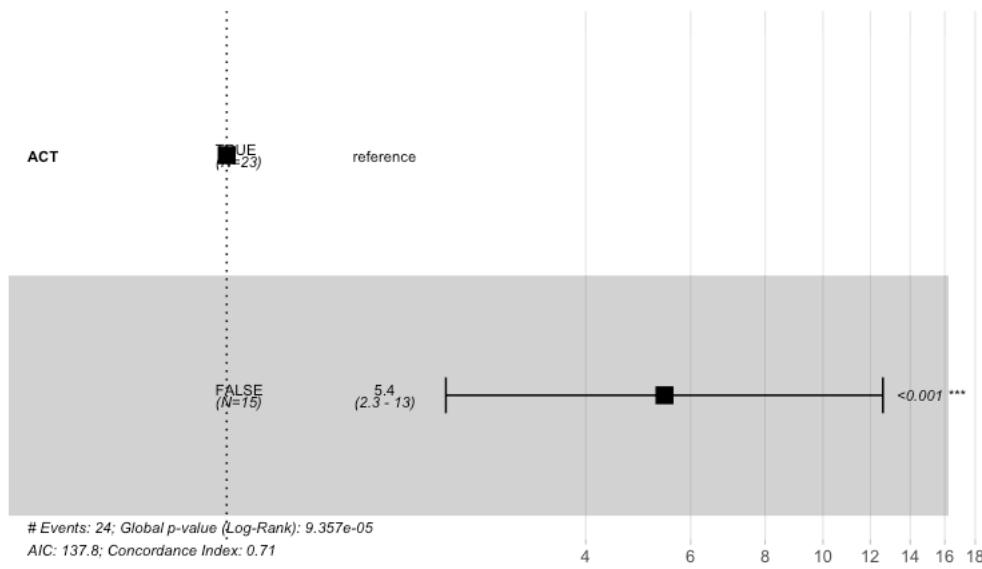
1588 observations deleted due to missingness

```
ACT=FALSE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.00000    1.00000   13.00000    0.10000    0.08756    0.00781    0.33528
```

```
ACT=TRUE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.000    5.000    10.000    0.537      0.110      0.305      0.722
```

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 38, number of events= 24
(1588 observations deleted due to missingness)

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 1.6902    5.4206   0.4305 3.926 8.64e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     5.421     0.1845    2.331     12.6

Concordance= 0.709  (se = 0.039 )
Likelihood ratio test= 15.26  on 1 df,  p=9e-05
Wald test           = 15.41  on 1 df,  p=9e-05
Score (logrank) test = 18.65  on 1 df,  p=2e-05
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 5.42 (2.33-12.6); p = 0"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 7.6738, df = 1, p-value = 0.005603
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.002121
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.954984 823.248016
sample estimates:
odds ratio
 16.90572
```

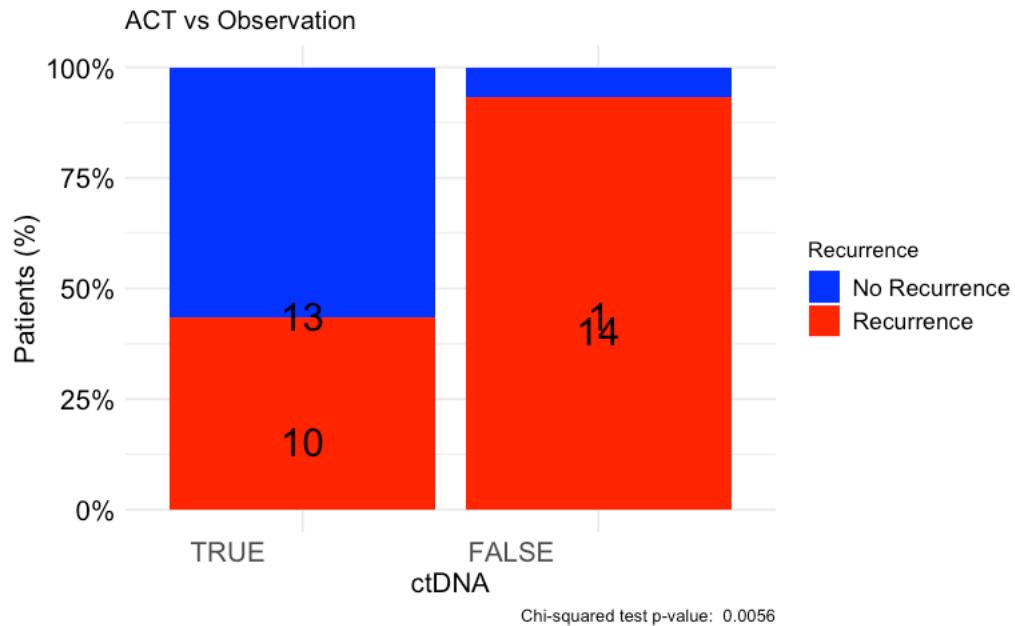
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
TRUE	13	10
FALSE	1	14

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 38, number of events= 24
(1588 observations deleted due to missingness)

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    2.0475    7.7483   0.5070  4.039 5.38e-05 ***
GenderMale   -0.4489   0.6383   0.4772 -0.941   0.3469
Age.Group≥70  0.2109   1.2348   0.4996  0.422   0.6729
ECOG1        1.4282   4.1714   0.5908  2.417   0.0156 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE      7.7483     0.1291    2.8686   20.928
GenderMale     0.6383     1.5666    0.2505    1.626
Age.Group≥70   1.2348     0.8098    0.4638    3.288
ECOG1         4.1714     0.2397    1.3103   13.280

Concordance= 0.759 (se = 0.052 )
Likelihood ratio test= 22.42 on 4 df,  p=2e-04
Wald test       = 19.09 on 4 df,  p=8e-04
Score (logrank) test = 25.15 on 4 df,  p=5e-05

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Risk.StageII==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 38, number of events= 24
(1588 observations deleted due to missingness)

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -2.0475    0.1291   0.5070 -4.039 5.38e-05 ***
GenderMale   -0.4489   0.6383   0.4772 -0.941   0.3469
Age.Group≥70  0.2109   1.2348   0.4996  0.422   0.6729
ECOG1        1.4282   4.1714   0.5908  2.417   0.0156 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.1291    7.7483   0.04778   0.3486
GenderMale     0.6383    1.5666   0.25054   1.6264
Age.Group≥70   1.2348    0.8098   0.46377   3.2878
ECOG1         4.1714    0.2397   1.31029  13.2798

Concordance= 0.759 (se = 0.052 )
Likelihood ratio test= 22.42 on 4 df,  p=2e-04
Wald test       = 19.09 on 4 df,  p=8e-04
Score (logrank) test = 25.15 on 4 df,  p=5e-05

```

#DFS by ACT treatment in MRD negative - Stage II T3N0

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T3N0",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: `survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~ ACT, data = circ_data)`

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	400	17	NA	NA	NA
ACT=TRUE	76	1	NA	NA	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

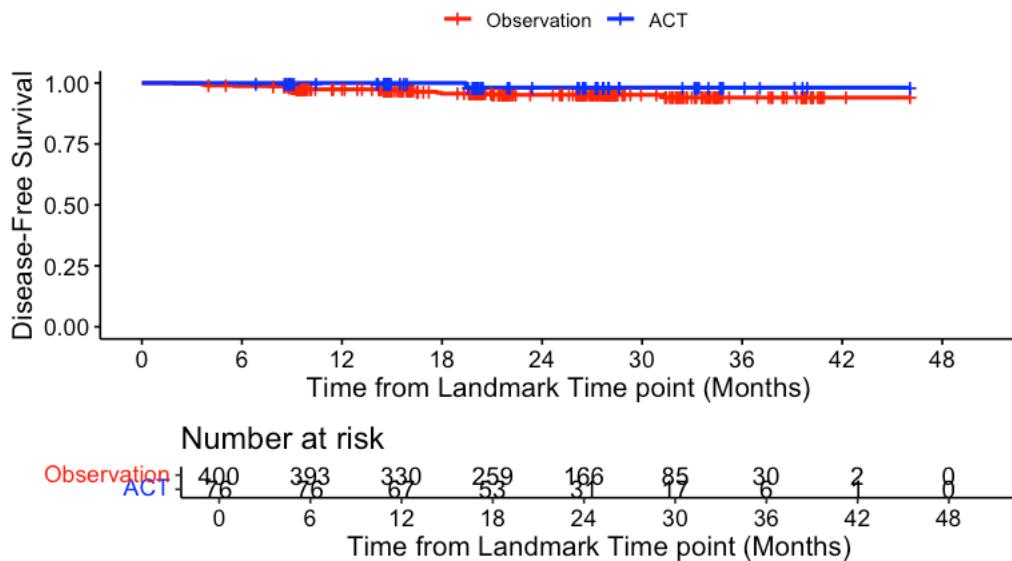
ACT	Total	Events	Fraction	Percentage
FALSE	400	17	0.04250000	4.250000
TRUE	76	1	0.01315789	1.315789

2 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | T3N0", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Negative ACT vs Observation | T3N0



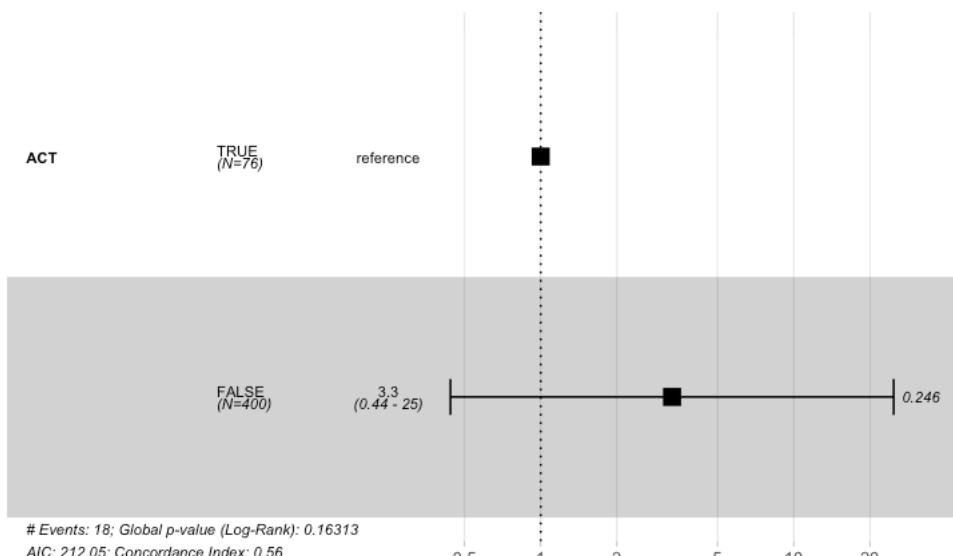
```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

time	ACT=FALSE		ACT=TRUE	
	n.risk	n.event	survival	std.err
24.0000	166.0000	16.0000	0.9516	0.0121
				0.9212
				0.9704
	n.risk	n.event	survival	std.err
24.0000	31.0000	1.0000	0.9811	0.0187
				0.8735
				0.9973

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 476, number of events= 18

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTFALSE 1.195     3.304    1.029  1.161    0.246  
                    exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     3.304     0.3027    0.4396   24.83   

Concordance= 0.559  (se = 0.023 )
Likelihood ratio test= 1.94  on 1 df,  p=0.2
Wald test      = 1.35  on 1 df,  p=0.2
Score (logrank) test = 1.52  on 1 df,  p=0.2

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 3.3 (0.44-24.83); p = 0.246"
```

[Hide](#)

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)

```

```

Warning in stats::chisq.test(x, y, ...):
  Chi-squared approximation may be incorrect

```

[Hide](#)

```
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```

data: contingency_table
X-squared = 0.81236, df = 1, p-value = 0.3674

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```
Fisher's Exact Test for Count Data
```

```

data: contingency_table
p-value = 0.3307
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.5053922 140.8837116
sample estimates:
odds ratio
3.323355

```

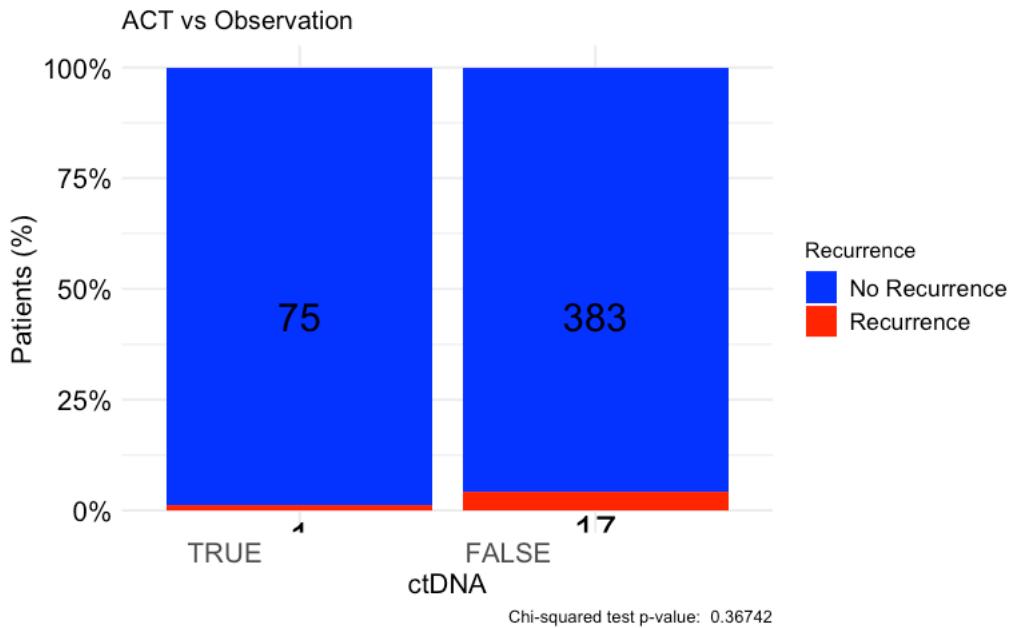
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
TRUE	75	1
FALSE	383	17

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T3N0",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 476, number of events= 18

            coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTFALSE    1.3971    4.0433   1.0319  1.354   0.1758    
GenderMale   0.1738    1.1898   0.4719  0.368   0.7127    
Age.Group≥70 -1.3071    0.2706   0.5576 -2.344   0.0191 *  
ECOG1       0.4088    1.5051   0.7931  0.516   0.6062    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95    
ACTFALSE    4.0433    0.2473   0.53501  30.5566    
GenderMale   1.1898    0.8405   0.47186  3.0000    
Age.Group≥70  0.2706    3.6955   0.09072  0.8072    
ECOG1       1.5051    0.6644   0.31805  7.1221    
Concordance= 0.688 (se = 0.041 )
Likelihood ratio test= 8.51 on 4 df,  p=0.07
Wald test      = 7.13 on 4 df,  p=0.1
Score (logrank) test = 7.92 on 4 df,  p=0.09
```

[Hide](#)

```
#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T3N0",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE", "TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))

circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0", "1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
 data = circ_data)

n= 476, number of events= 18

	coef	exp(coef)	se(coef)	z	Pr(> z)						
ACTTRUE	-1.3971	0.2473	1.0319	-1.354	0.1758						
GenderMale	0.1738	1.1898	0.4719	0.368	0.7127						
Age.Group≥70	-1.3071	0.2706	0.5576	-2.344	0.0191 *						
ECOG1	0.4088	1.5051	0.7931	0.516	0.6062						

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1
	exp(coef)	exp(-coef)	lower .95	upper .95							
ACTTRUE	0.2473	4.0433	0.03273	1.8691							
GenderMale	1.1898	0.8405	0.47186	3.0000							
Age.Group≥70	0.2706	3.6955	0.09072	0.8072							
ECOG1	1.5051	0.6644	0.31805	7.1221							

Concordance= 0.688 (se = 0.041)
 Likelihood ratio test= 8.51 on 4 df, p=0.07
 Wald test = 7.13 on 4 df, p=0.1
 Score (logrank) test = 7.92 on 4 df, p=0.09

#DFS by ACT treatment in MRD negative - Stage II T4N0

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T4N0",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	64	9	NA	NA	NA
ACT=TRUE	29	2	NA	NA	NA

Hide

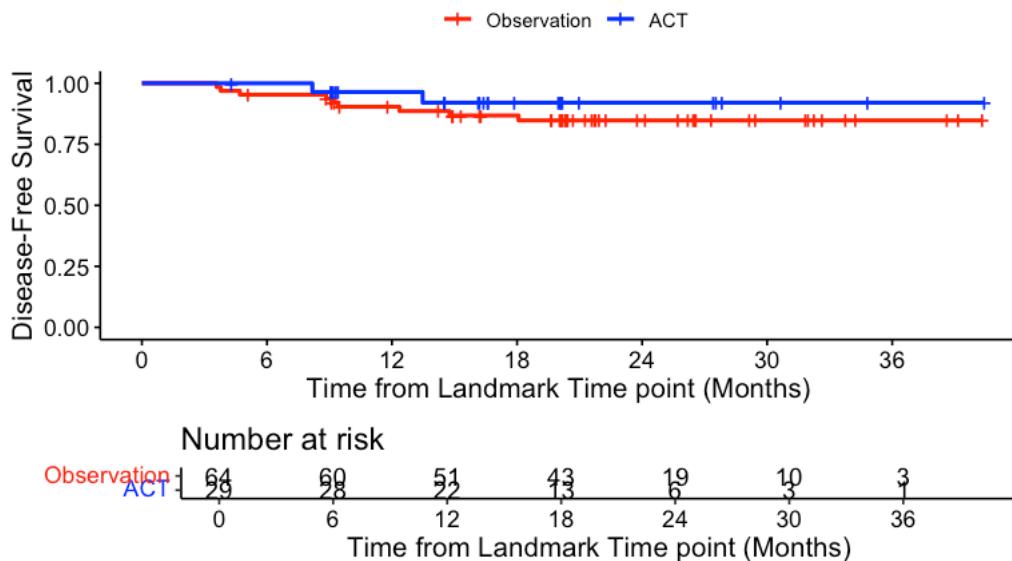
```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ACT	Total	Events	Fraction	Percentage
<lgI>	<int>	<int>	<dbl>	<dbl>
FALSE	64	9	0.14062500	14.062500
TRUE	29	2	0.06896552	6.896552
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | T4N0", ylab= "Disease-Free Survival", xlab = "Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Negative ACT vs Observation | T4N0



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

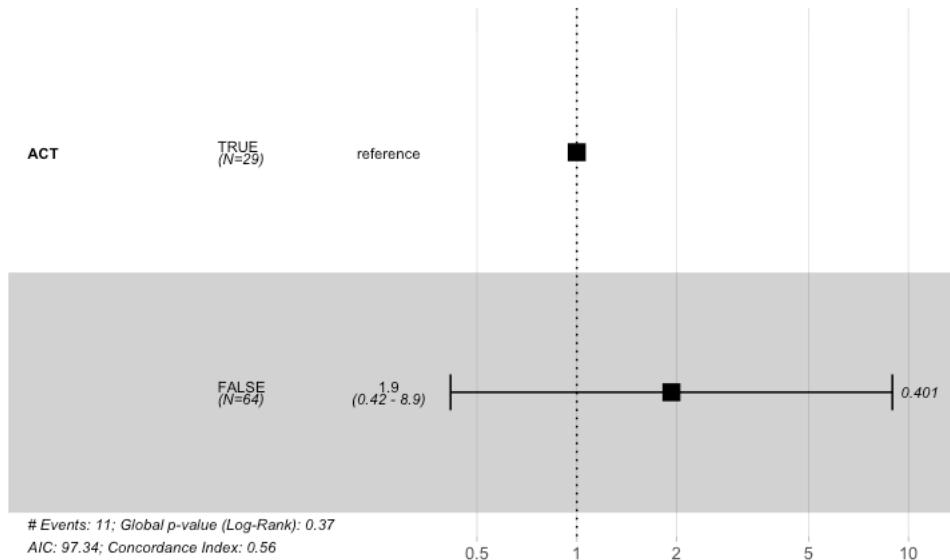
ACT=FALSE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	19.0000	9.0000	0.8478	0.0471	0.7267	0.9181

ACT=TRUE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	6.0000	2.0000	0.9205	0.0544	0.7154	0.9797

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)
```

```
n= 93, number of events= 11
```

coef	exp(coef)	se(coef)	z	Pr(> z)
ACTFALSE	0.6570	1.9290	0.7824	0.84

exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	1.929	0.5184	0.4162

```
Concordance= 0.561 (se = 0.06 )
```

```
Likelihood ratio test= 0.8 on 1 df, p=0.4
```

```
Wald test = 0.71 on 1 df, p=0.4
```

```
Score (logrank) test = 0.73 on 1 df, p=0.4
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

[1] "HR = 1.93 (0.42-8.94); p = 0.401"

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

[Hide](#)

```
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 0.41565, df = 1, p-value = 0.5191
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.4927
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.4118588 22.2431792
sample estimates:
odds ratio
2.192682
```

[Hide](#)

```
print(contingency_table)
```

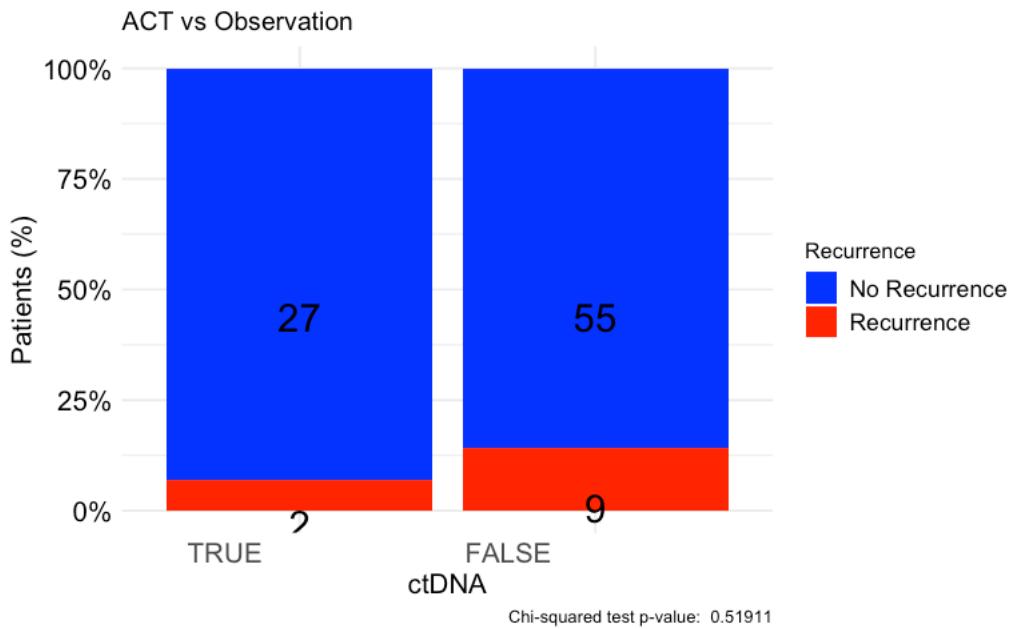
	No Recurrence	Recurrence
TRUE	27	2
FALSE	55	9

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T4N0",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 93, number of events= 11

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    0.6626   1.9399   0.8154  0.813   0.416
GenderMale   -0.1393   0.8700   0.6220 -0.224   0.823
Age.Group≥70 -0.3472   0.7066   0.6563 -0.529   0.597
ECOG1        0.3212   1.3788   0.8365  0.384   0.701

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     1.9399   0.5155   0.3924   9.591
GenderMale    0.8700   1.1494   0.2571   2.944
Age.Group≥70  0.7066   1.4152   0.1952   2.557
ECOG1        1.3788   0.7253   0.2676   7.104

Concordance= 0.588 (se = 0.075 )
Likelihood ratio test= 1.17 on 4 df,  p=0.9
Wald test       = 1.08 on 4 df,  p=0.9
Score (logrank) test = 1.11 on 4 df,  p=0.9

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$StageII.Group=="T4N0",]
circ_data <- circ_data[circ_data$cDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))

circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0", "1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 93, number of events= 11

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -0.6626   0.5155   0.8154 -0.813   0.416
GenderMale  -0.1393   0.8700   0.6220 -0.224   0.823
Age.Group≥70 -0.3472   0.7066   0.6563 -0.529   0.597
ECOG1        0.3212   1.3788   0.8365  0.384   0.701

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE     0.5155   1.9399   0.1043   2.549
GenderMale   0.8700   1.1494   0.2571   2.944
Age.Group≥70  0.7066   1.4152   0.1952   2.557
ECOG1        1.3788   0.7253   0.2676   7.104

Concordance= 0.588 (se = 0.075 )
Likelihood ratio test= 1.17 on 4 df,  p=0.9
Wald test       = 1.08 on 4 df,  p=0.9
Score (logrank) test = 1.11 on 4 df,  p=0.9

```

#DFS by ACT treatment in MRD negative - Stage III

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	213	29	NA	NA	NA
ACT=TRUE	469	52	NA	NA	NA

[Hide](#)

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ACT	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
FALSE	213	29	0.1361502	13.61502
TRUE	469	52	0.1108742	11.08742

2 rows

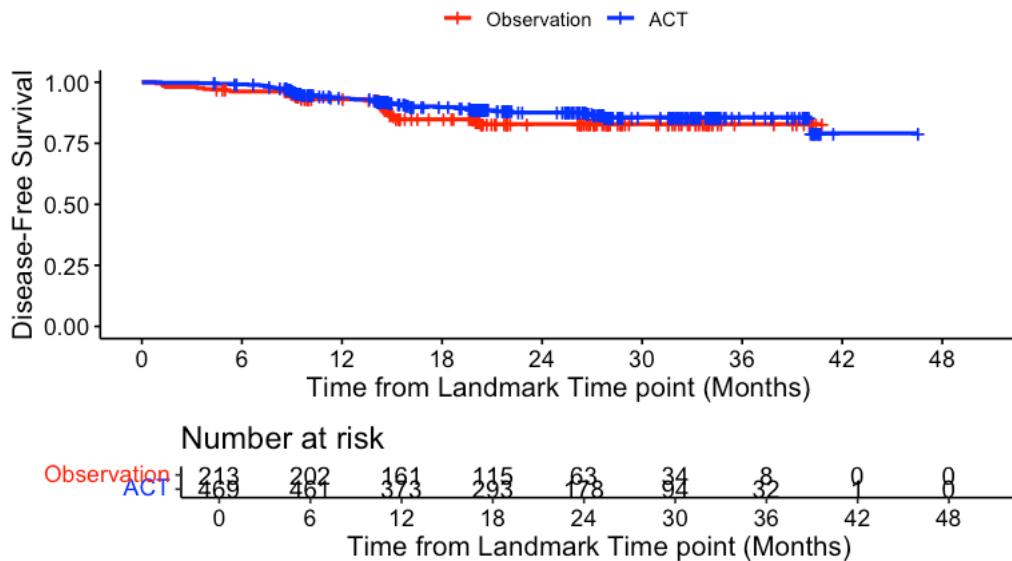
[Hide](#)

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red", "blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")

```

DFS - ctDNA MRD Negative ACT vs Observation | Stage III

[Hide](#)

```
summary(KM_curve, times= c(18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")

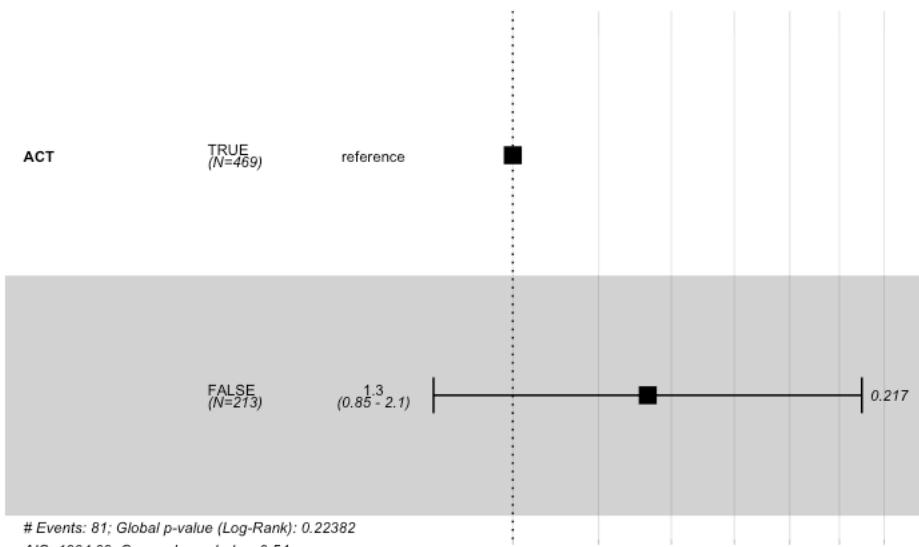
  ACT=FALSE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    18     115      27    0.848  0.0274      0.785    0.894
    24      63       2    0.829  0.0300      0.760    0.879

  ACT=TRUE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    18     293      42    0.898  0.0150      0.864    0.924
    24     178       6    0.876  0.0173      0.837    0.906
```

[Hide](#)

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 682, number of events= 81

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 0.2863    1.3315   0.2319 1.235    0.217

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    1.332      0.751    0.8452    2.098

Concordance= 0.537  (se = 0.028 )
Likelihood ratio test= 1.48  on 1 df,  p=0.2
Wald test            = 1.52  on 1 df,  p=0.2
Score (logrank) test = 1.53  on 1 df,  p=0.2
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.33 (0.85-2.1); p = 0.217"
```

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 0.66893, df = 1, p-value = 0.4134
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.3718
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.7476624 2.1022451
sample estimates:
odds ratio
 1.263412
```

[Hide](#)

```
print(contingency_table)
```

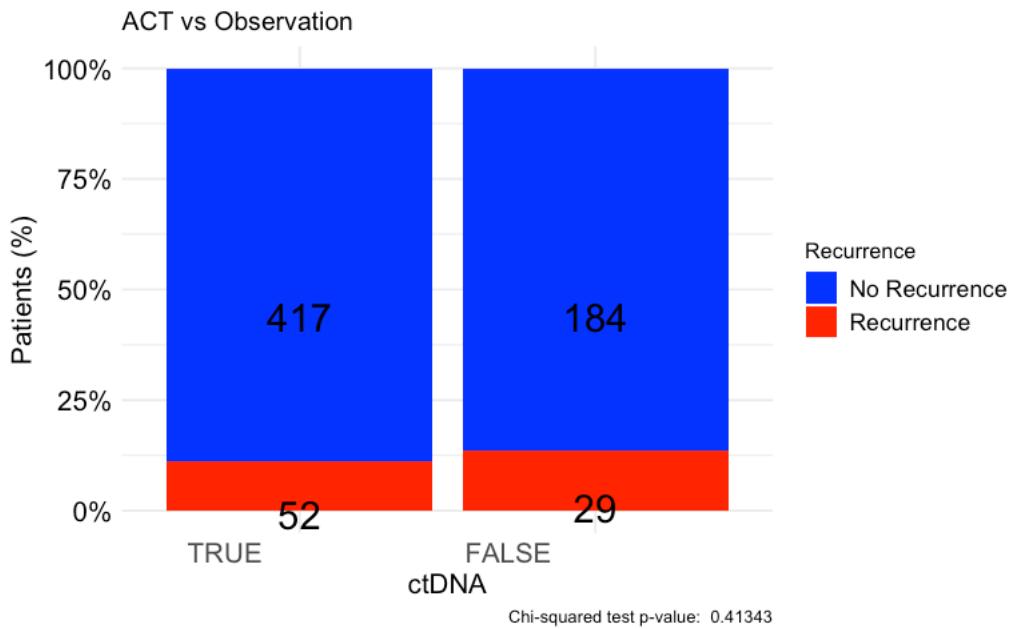
	No Recurrence	Recurrence
TRUE	417	52
FALSE	184	29

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 682, number of events= 81

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    0.3004   1.3505   0.2340  1.284   0.199
GenderMale   0.2382   1.2690   0.2244  1.062   0.288
Age.Group≥70 -0.1732   0.8410   0.2327 -0.744   0.457
ECOG1       0.1347   1.1442   0.3823  0.352   0.725

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     1.350    0.7405   0.8537   2.136
GenderMale   1.269    0.7880   0.8175   1.970
Age.Group≥70  0.841    1.1891   0.5329   1.327
ECOG1       1.144    0.8740   0.5409   2.420

Concordance= 0.553 (se = 0.033 )
Likelihood ratio test= 3.24 on 4 df,  p=0.5
Wald test       = 3.28 on 4 df,  p=0.5
Score (logrank) test = 3.29 on 4 df,  p=0.5

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 682, number of events= 81

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -0.3004   0.7405   0.2340 -1.284   0.199
GenderMale   0.2382   1.2690   0.2244  1.062   0.288
Age.Group≥70 -0.1732   0.8410   0.2327 -0.744   0.457
ECOG1       0.1347   1.1442   0.3823  0.352   0.725

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE     0.7405    1.350    0.4681   1.171
GenderMale  1.2690    0.788    0.8175   1.970
Age.Group≥70  0.8410    1.189    0.5329   1.327
ECOG1       1.1442    0.874    0.5409   2.420

Concordance= 0.553 (se = 0.033 )
Likelihood ratio test= 3.24 on 4 df,  p=0.5
Wald test       = 3.28 on 4 df,  p=0.5
Score (logrank) test = 3.29 on 4 df,  p=0.5

```

#DFS by ACT treatment in MRD positive - Stage III

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	32	31	3.58	2.57	4.01
ACT=TRUE	122	78	11.27	9.10	16.07

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

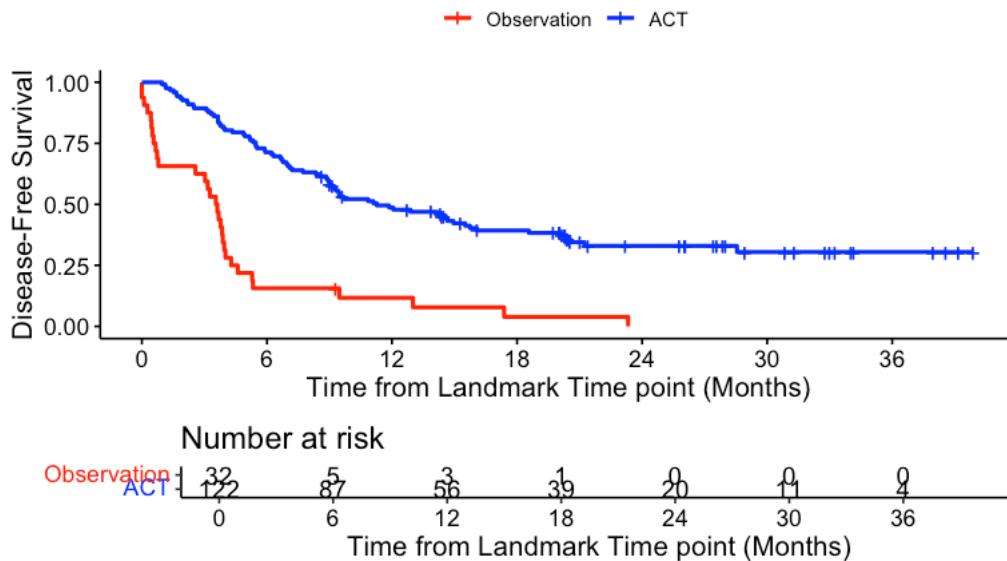
ACT	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
FALSE	32	31	0.9687500	96.87500
TRUE	122	78	0.6393443	63.93443

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red", "blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Positive ACT vs Observation | Stage III



Hide

```
summary(KM_curve, times= c(18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

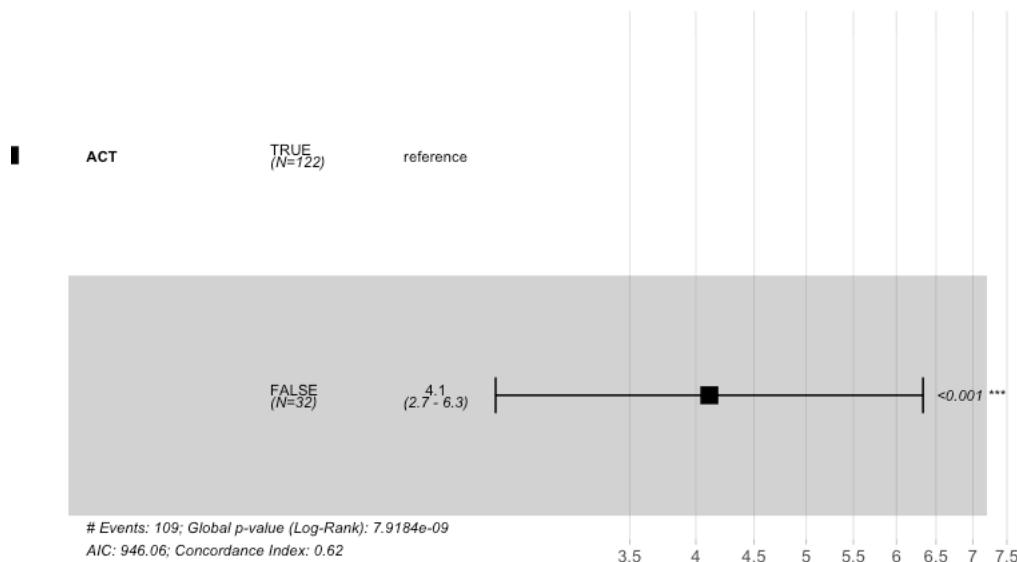
ACT=FALSE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
18.00000	1.00000	30.00000	0.03906	0.03744	0.00306	0.16257

ACT=TRUE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
18	39	72	0.393	0.0455	0.304	0.481
24	20	5	0.330	0.0464	0.241	0.421

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 154, number of events= 109

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTFALSE 1.4135    4.1105   0.2203 6.417 1.39e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
ACTFALSE     4.11      0.2433    2.669     6.33    

Concordance= 0.619  (se = 0.021 )
Likelihood ratio test= 33.3  on 1 df,  p=8e-09
Wald test            = 41.18  on 1 df,  p=1e-10
Score (logrank) test = 47.85  on 1 df,  p=5e-12

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 4.11 (2.67-6.33); p = 0"
```

[Hide](#)

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```

data: contingency_table
X-squared = 11.755, df = 1, p-value = 0.0006068

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```
Fisher's Exact Test for Count Data
```

```

data: contingency_table
p-value = 0.0001195
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.685315 726.194688
sample estimates:
odds ratio
 17.29321

```

[Hide](#)

```
print(contingency_table)
```

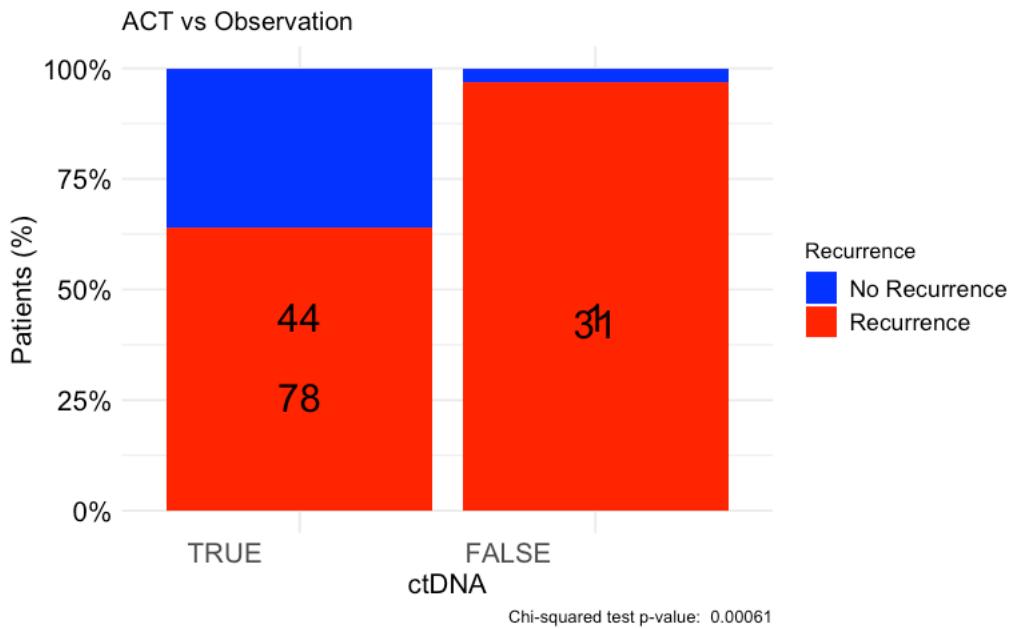
	No Recurrence	Recurrence
TRUE	44	78
FALSE	1	31

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 154, number of events= 109

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTFALSE    1.48178  4.40077  0.24173  6.130 8.79e-10 *** 
GenderMale   0.02384  1.02413  0.19953  0.119   0.905    
Age.Group≥70 -0.01673  0.98341  0.20368 -0.082   0.935    
ECOG1       -0.20242  0.81675  0.33241 -0.609   0.543    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     4.4008    0.2272    2.7401    7.068
GenderMale    1.0241    0.9764    0.6926    1.514
Age.Group≥70  0.9834    1.0169    0.6597    1.466
ECOG1        0.8168    1.2244    0.4257    1.567

Concordance= 0.63  (se = 0.027 )
Likelihood ratio test= 33.75 on 4 df,  p=8e-07
Wald test        = 41.58 on 4 df,  p=2e-08
Score (logrank) test = 48.35 on 4 df,  p=8e-10

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "IV")),]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE", "TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0", "1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 154, number of events= 109

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ACTTRUE     -1.48178  0.22723  0.24173  -6.130 8.79e-10 *** 
GenderMale   0.02384  1.02413  0.19953  0.119   0.905    
Age.Group≥70 -0.01673  0.98341  0.20368 -0.082   0.935    
ECOG1       -0.20242  0.81675  0.33241 -0.609   0.543    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.2272    4.4008    0.1415    0.3649
GenderMale    1.0241    0.9764    0.6926    1.5142
Age.Group≥70  0.9834    1.0169    0.6597    1.4659
ECOG1        0.8168    1.2244    0.4257    1.5669

Concordance= 0.63  (se = 0.027 )
Likelihood ratio test= 33.75 on 4 df,  p=8e-07
Wald test        = 41.58 on 4 df,  p=2e-08
Score (logrank) test = 48.35 on 4 df,  p=8e-10

```

#DFS by ACT treatment in MRD negative - Stage IV NAC-treated

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	113	53	27.9	15.3	NA
ACT=TRUE	30	11	NA	20.1	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

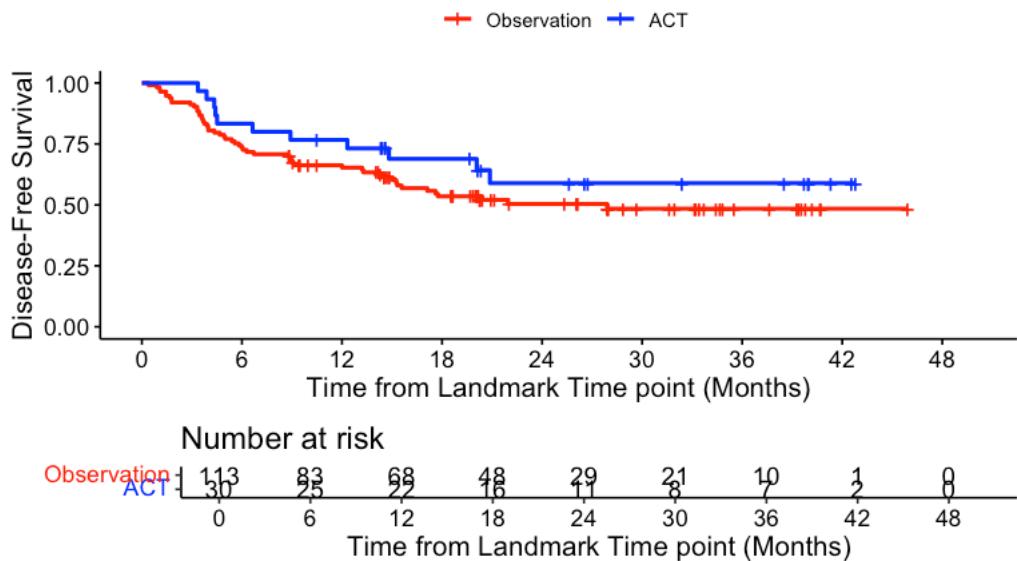
ACT	Total	Events	Fraction	Percentage
<lg1>	<int>	<int>	<dbl>	<dbl>
FALSE	113	53	0.4690265	46.90265
TRUE	30	11	0.3666667	36.66667

2 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red", "blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | Stage IV NAC-treated", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Negative ACT vs Observation | Stage IV NAC-tr



Hide

```
summary(KM_curve, times= c(3, 6, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

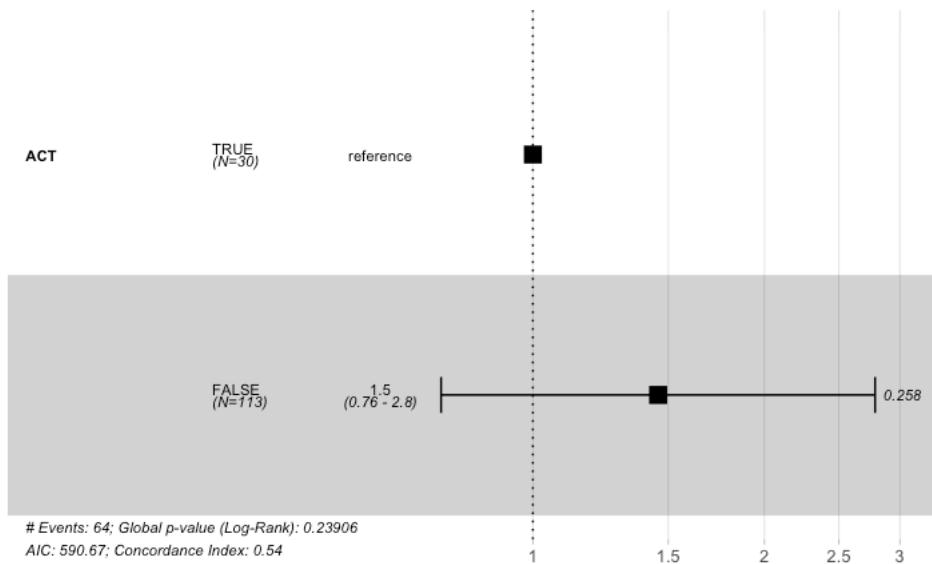
```
ACT=FALSE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  3    103     10    0.912  0.0267    0.842    0.951
  6     83     20    0.735  0.0415    0.643    0.806
 18     48     20    0.535  0.0490    0.435    0.625
 24     29      2    0.504  0.0509    0.400    0.598

ACT=TRUE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  3    30      0    1.000  0.0000    1.000    1.000
  6    25      5    0.833  0.0680    0.645    0.927
 18    16      4    0.689  0.0871    0.484    0.825
 24    11      2    0.589  0.0992    0.373    0.753
```

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 143, number of events= 64

      coef  exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 0.3749    1.4549   0.3314  1.131    0.258

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    1.455     0.6873   0.7598   2.786

Concordance= 0.535  (se = 0.024 )
Likelihood ratio test= 1.39  on 1 df,   p=0.2
Wald test      = 1.28  on 1 df,   p=0.3
Score (logrank) test = 1.29  on 1 df,   p=0.3
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.45 (0.76-2.79); p = 0.258"
```

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 0.63325, df = 1, p-value = 0.4262
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

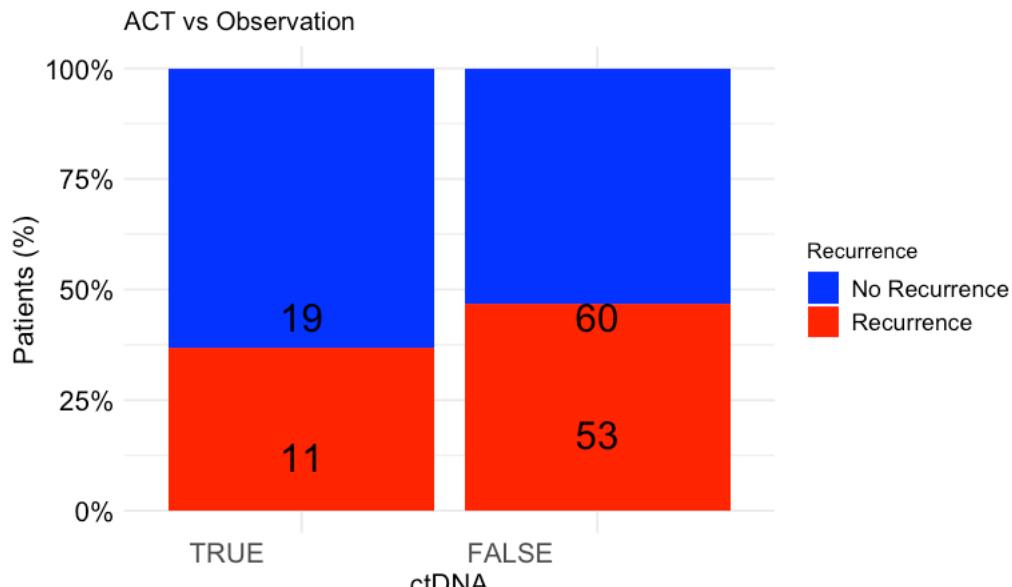
```
Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.4094
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.6212468 3.8850136
sample estimates:
odds ratio
1.521304
```

```
print(contingency_table)
```

		No Recurrence	Recurrence
TRUE	19	11	
FALSE	60	53	

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE", "FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
 data = circ_data)

n= 143, number of events= 64

	coef	exp(coef)	se(coef)	z	Pr(> z)
ACTFALSE	0.3908	1.4781	0.3332	1.173	0.241
GenderMale	0.3629	1.4375	0.2635	1.377	0.168
Age.Group≥70	-0.3175	0.7279	0.2697	-1.178	0.239
ECOG1	-0.5519	0.5759	0.7251	-0.761	0.447

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	1.4781	0.6765	0.7694	2.840
GenderMale	1.4375	0.6956	0.8577	2.409
Age.Group≥70	0.7279	1.3737	0.4291	1.235
ECOG1	0.5759	1.7365	0.1390	2.385

Concordance= 0.574 (se = 0.036)
 Likelihood ratio test= 5.95 on 4 df, p=0.2
 Wald test = 5.5 on 4 df, p=0.2
 Score (logrank) test = 5.6 on 4 df, p=0.2

Hide

```
#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 143, number of events= 64

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE     -0.3908    0.6765   0.3332 -1.173   0.241
GenderMale    0.3629    1.4375   0.2635  1.377   0.168
Age.Group≥70 -0.3175    0.7279   0.2697 -1.178   0.239
ECOG1       -0.5519    0.5759   0.7251 -0.761   0.447

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE      0.6765    1.4781   0.3521   1.300
GenderMale    1.4375    0.6956   0.8577   2.409
Age.Group≥70  0.7279    1.3737   0.4291   1.235
ECOG1       0.5759    1.7365   0.1390   2.385

Concordance= 0.574  (se = 0.036 )
Likelihood ratio test= 5.95  on 4 df,  p=0.2
Wald test          = 5.5  on 4 df,  p=0.2
Score (logrank) test = 5.6  on 4 df,  p=0.2

```

#DFS by ACT treatment in MRD Negative - Stage IV no NAC-treated

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="FALSE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	81	30	NA	33.1	NA
ACT=TRUE	50	14	NA	NA	NA

[Hide](#)

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ACT	Total	Events	Fraction	Percentage
<lg>	<int>	<int>	<dbl>	<dbl>
FALSE	81	30	0.3703704	37.03704
TRUE	50	14	0.2800000	28.00000

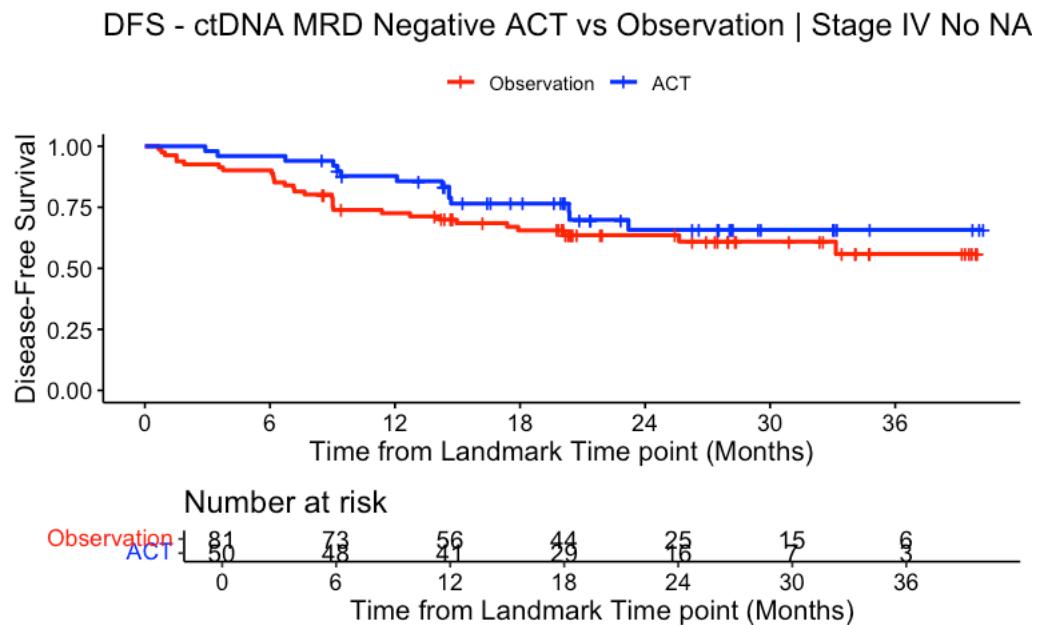
2 rows

[Hide](#)

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("red", "blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | Stage IV No NAC-treated", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title
="")

```



Hide

```
summary(KM_curve, times= c(3, 6, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

```
ACT=FALSE
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
3	75	6	0.926	0.0291	0.843	0.966	0.949	0.994
6	73	2	0.901	0.0331	0.812	0.949	0.750	0.863
18	44	19	0.655	0.0541	0.538	0.750	0.516	0.734
24	25	1	0.636	0.0559	0.479	0.787	0.516	0.734

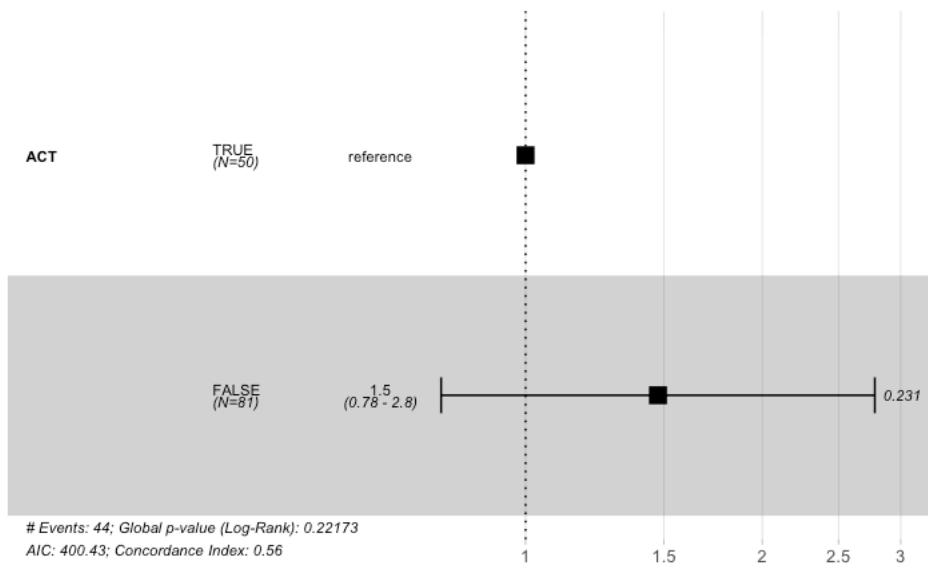
```
ACT=TRUE
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
3	49	1	0.980	0.0198	0.866	0.997	0.990	1.000
6	48	1	0.960	0.0277	0.849	0.990	0.615	0.863
18	29	9	0.765	0.0623	0.615	0.863	0.479	0.787
24	16	3	0.658	0.0790	0.479	0.787	0.516	0.734

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 131, number of events= 44

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ACTFALSE 0.3877    1.4736   0.3240  1.197    0.231    
                    exp(coef) exp(-coef) lower .95 upper .95    
ACTFALSE     1.474      0.6786   0.7809    2.781    

Concordance= 0.56  (se = 0.035 )
Likelihood ratio test= 1.49 on 1 df,   p=0.2
Wald test            = 1.43 on 1 df,   p=0.2
Score (logrank) test = 1.45 on 1 df,   p=0.2
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.47 (0.78-2.78); p = 0.231"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 0.76302, df = 1, p-value = 0.3824
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.343
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.6637884 3.5367388
sample estimates:
odds ratio
1.507881
```

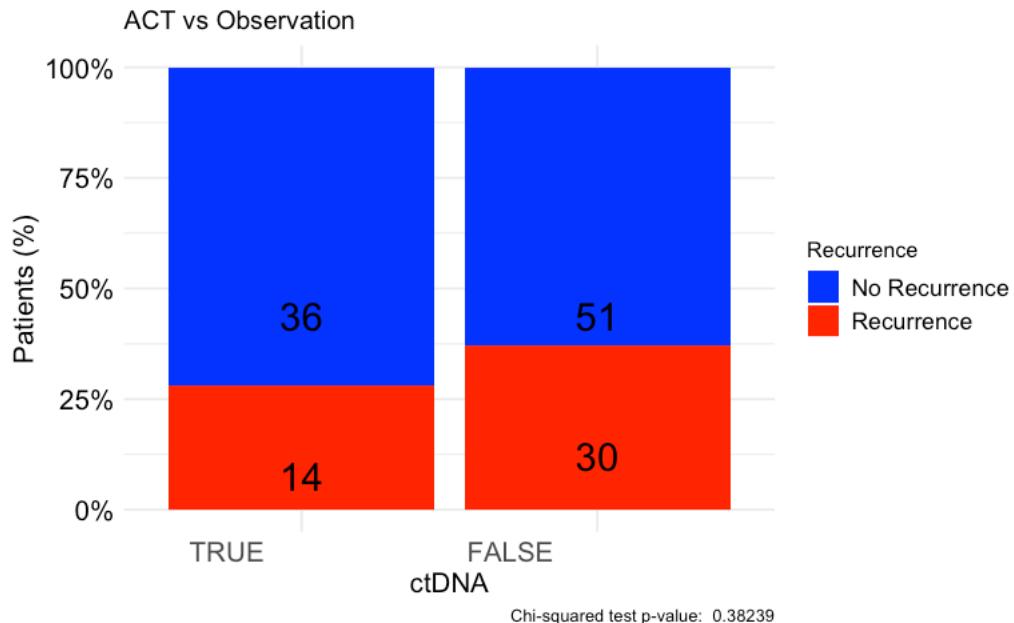
[Hide](#)

```
print(contingency_table)
```

	No	Recurrence
TRUE	36	14
FALSE	51	30

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE"]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III"))]
circ_data <- circ_data[circ_data$NAC=="FALSE"]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE"]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 131, number of events= 44

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    0.2636    1.3016   0.3355  0.786  0.4320
GenderMale   0.3385    1.4029   0.3255  1.040  0.2983
Age.Group≥70  0.7056    2.0251   0.3167  2.228  0.0259 *
ECOG1      -1.5549    0.2112   1.0202 -1.524  0.1275
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     1.3016    0.7683   0.6744   2.512
GenderMale    1.4029    0.7128   0.7412   2.655
Age.Group≥70  2.0251    0.4938   1.0886   3.767
ECOG1        0.2112    4.7344   0.0286   1.560

Concordance= 0.64  (se = 0.042 )
Likelihood ratio test= 10.99 on 4 df,  p=0.03
Wald test       = 10.06 on 4 df,  p=0.04
Score (logrank) test = 10.67 on 4 df,  p=0.03

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="FALSE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 131, number of events= 44

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -0.2636    0.7683   0.3355 -0.786  0.4320
GenderMale   0.3385    1.4029   0.3255  1.040  0.2983
Age.Group≥70  0.7056    2.0251   0.3167  2.228  0.0259 *
ECOG1      -1.5549    0.2112   1.0202 -1.524  0.1275
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE     0.7683    1.3016   0.3981   1.483
GenderMale   1.4029    0.7128   0.7412   2.655
Age.Group≥70  2.0251    0.4938   1.0886   3.767
ECOG1        0.2112    4.7344   0.0286   1.560

Concordance= 0.64  (se = 0.042 )
Likelihood ratio test= 10.99 on 4 df,  p=0.03
Wald test       = 10.06 on 4 df,  p=0.04
Score (logrank) test = 10.67 on 4 df,  p=0.03

```

#DFS by ACT treatment in MRD positive - Stage IV NAC-treated

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	32	32	1.46	0.86	2.44
ACT=TRUE	14	13	3.78	3.13	12.59

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

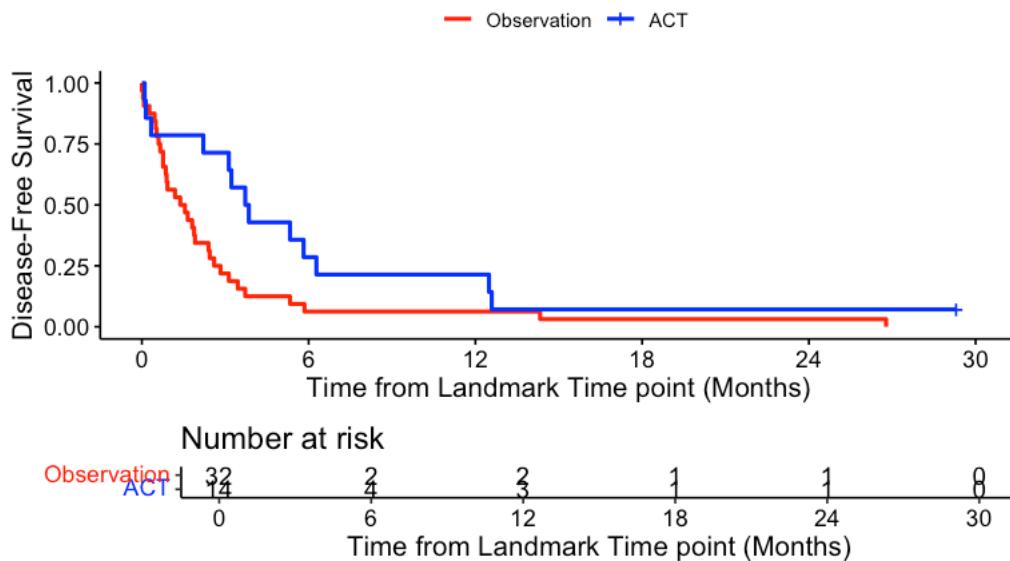
ACT	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
FALSE	32	32	1.0000000	100.00000
TRUE	14	13	0.9285714	92.85714

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red","blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | Stage IV NAC-treated", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Positive ACT vs Observation | Stage IV NAC-tre



Hide

```
summary(KM_curve, times= c(3, 6, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

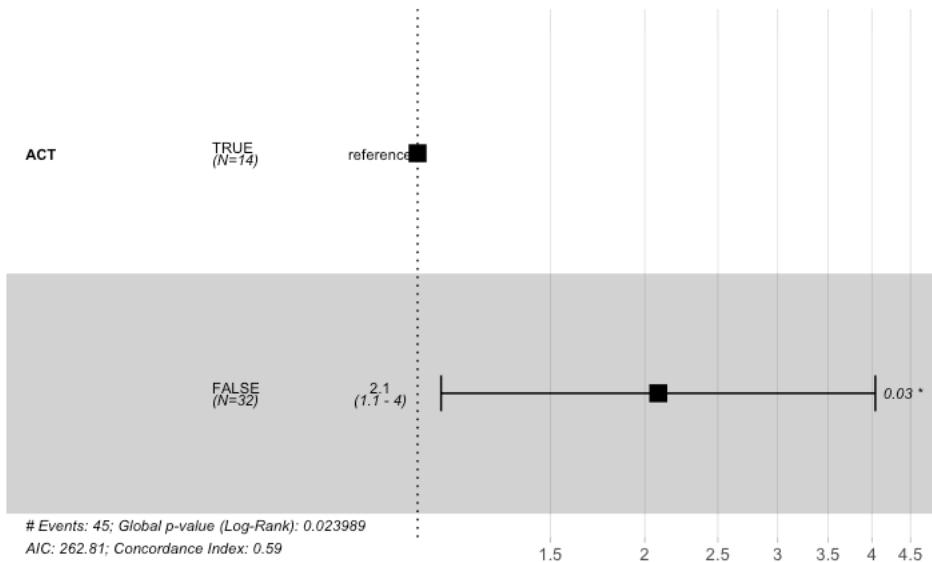
```
ACT=FALSE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  3     7      25  0.2188  0.0731    0.09649  0.372
  6     2      5  0.0625  0.0428    0.01112  0.181
 18     1      1  0.0312  0.0308    0.00237  0.137
 24     1      0  0.0312  0.0308    0.00237  0.137
```

```
ACT=TRUE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  3    10      4  0.7143  0.1207    0.40630  0.882
  6     4      6  0.2857  0.1207    0.08834  0.524
 18     1      3  0.0714  0.0688    0.00452  0.275
 24     1      0  0.0714  0.0688    0.00452  0.275
```

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 46, number of events= 45

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 0.7342   2.0839   0.3380  2.172   0.0298 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    2.084     0.4799    1.074     4.042

Concordance= 0.591  (se = 0.043 )
Likelihood ratio test= 5.1  on 1 df,  p=0.02
Wald test           = 4.72  on 1 df,  p=0.03
Score (logrank) test = 4.9  on 1 df,  p=0.03
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 2.08 (1.07-4.04); p = 0.03"
```

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

[Hide](#)

```
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 0.18482, df = 1, p-value = 0.6673
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data
```

```
data: contingency_table
p-value = 0.3043
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.05860791      Inf
sample estimates:
odds ratio
Inf
```

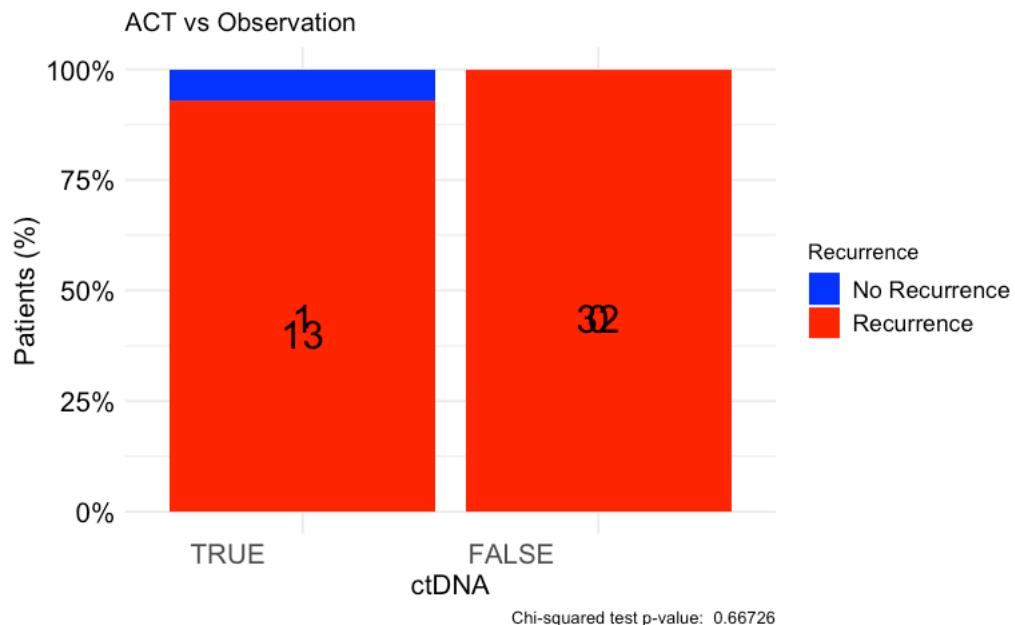
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
TRUE	1	13
FALSE	0	32

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 46, number of events= 45

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    0.9147   2.4961   0.3739  2.447   0.0144 *
GenderMale  -0.4952   0.6095   0.3597 -1.377   0.1686
Age.Group≥70  0.1691   1.1843   0.3357  0.504   0.6145
ECOG1        NA        NA   0.0000   NA        NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     2.4961    0.4006   1.1996   5.194
GenderMale    0.6095    1.6408   0.3011   1.233
Age.Group≥70   1.1843    0.8444   0.6133   2.287
ECOG1        NA        NA        NA        NA

Concordance= 0.637 (se = 0.047 )
Likelihood ratio test= 7 on 3 df,  p=0.07
Wald test      = 6.36 on 3 df,  p=0.1
Score (logrank) test = 6.56 on 3 df,  p=0.09

```

Hide

```

#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
      data = circ_data)

n= 46, number of events= 45

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -0.9147   0.4006   0.3739 -2.447   0.0144 *
GenderMale  -0.4952   0.6095   0.3597 -1.377   0.1686
Age.Group≥70  0.1691   1.1843   0.3357  0.504   0.6145
ECOG1        NA        NA   0.0000   NA        NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE     0.4006    2.4961   0.1925   0.8336
GenderMale   0.6095    1.6408   0.3011   1.2335
Age.Group≥70   1.1843    0.8444   0.6133   2.2868
ECOG1        NA        NA        NA        NA

Concordance= 0.637 (se = 0.047 )
Likelihood ratio test= 7 on 3 df,  p=0.07
Wald test      = 6.36 on 3 df,  p=0.1
Score (logrank) test = 6.56 on 3 df,  p=0.09

```

#DFS by ACT treatment in MRD positive - Stage IV no NAC-treated

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="FALSE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	28	27	2.8	1.12	3.52
ACT=TRUE	26	15	14.2	5.92	NA

[Hide](#)

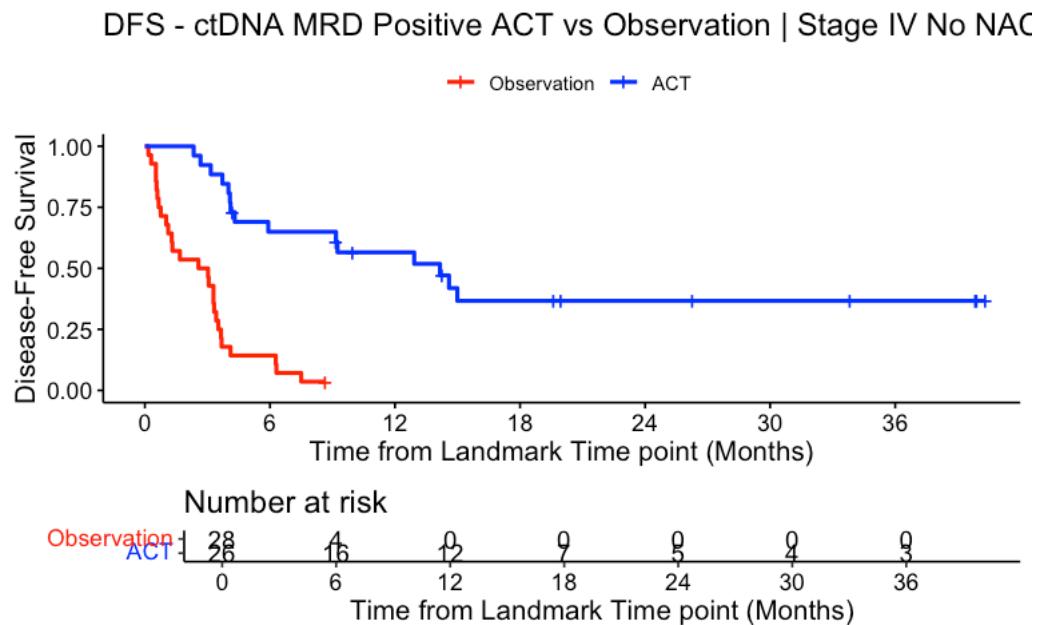
```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ACT	Total	Events	Fraction	Percentage
	<dbl>	<int>	<dbl>	<dbl>
FALSE	28	27	0.9642857	96.42857
TRUE	26	15	0.5769231	57.69231

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red","blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | Stage IV No NAC-treated", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title = "")
```



Hide

```
summary(KM_curve, times= c(3, 6, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")

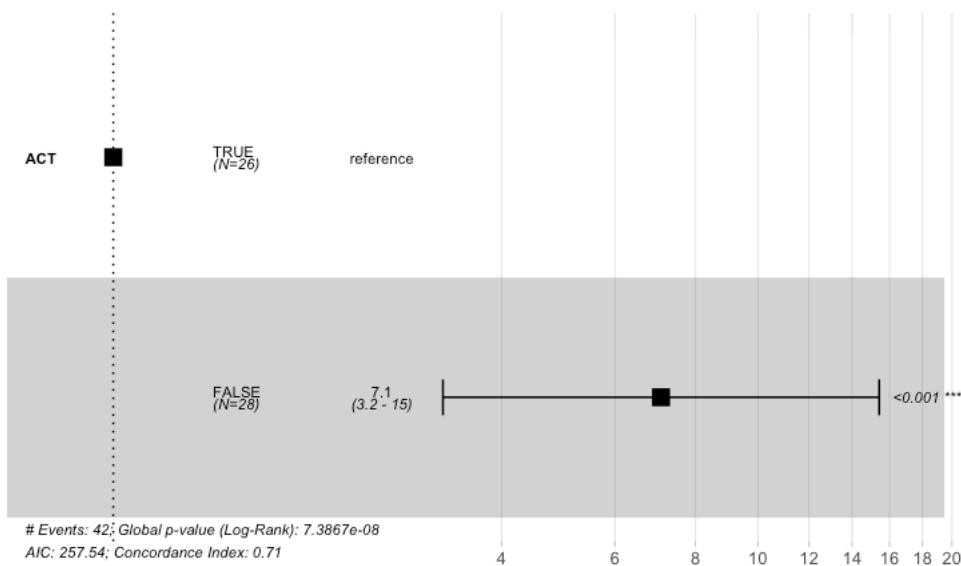
  ACT=FALSE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    3     14     14    0.500  0.0945     0.306    0.666
    6      4     10    0.143  0.0661     0.045    0.295

  ACT=TRUE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    3     24      2    0.923  0.0523     0.726    0.98
    6     16      7    0.650  0.0944     0.434    0.80
   18      7      6    0.367  0.1031     0.176    0.56
   24      5      0    0.367  0.1031     0.176    0.56
```

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 54, number of events= 42

      coef  exp(coef)  se(coef)      z  Pr(>|z|)
ACTFALSE 1.9565    7.0742   0.3975 4.922 8.55e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95
ACTFALSE    7.074      0.1414    3.246     15.42

Concordance= 0.714  (se = 0.025 )
Likelihood ratio test= 28.96  on 1 df,  p=7e-08
Wald test            = 24.23  on 1 df,  p=9e-07
Score (logrank) test = 31.08  on 1 df,  p=2e-08
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 7.07 (3.25-15.42); p = 0"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ACT, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 9.57, df = 1, p-value = 0.001978
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.0007475
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
2.333943 878.542882
sample estimates:
odds ratio
18.79777
```

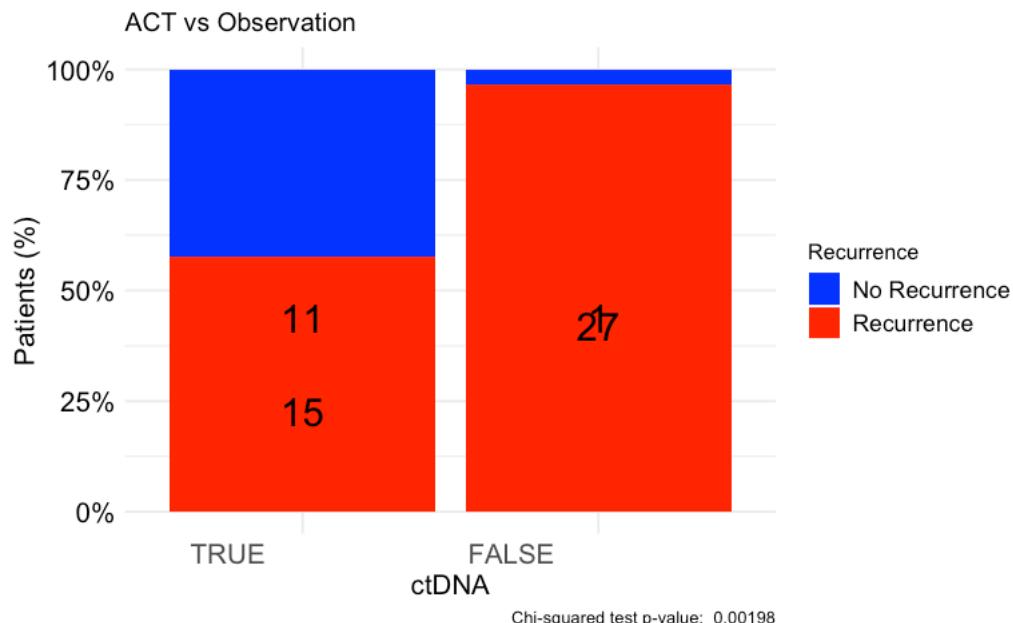
Hide

```
print(contingency_table)
```

	No Recurrence	Recurrence
TRUE	11	15
FALSE	1	27

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ACT vs Observation",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



Hide

```
#Adjusted HR "ACT vs no ACT" - age, gender, MSI, pathological stage, and performance status
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="FALSE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
 data = circ_data)

n= 54, number of events= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
ACTFALSE	1.998062	7.374747	0.405506	4.927	8.34e-07 ***
GenderMale	-0.173869	0.840407	0.330194	-0.527	0.598
Age.Group≥70	0.001981	1.001983	0.319386	0.006	0.995
ECOG1	-0.038826	0.961919	0.615096	-0.063	0.950

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	7.3747	0.1356	3.3310	16.327
GenderMale	0.8404	1.1899	0.4400	1.605
Age.Group≥70	1.0020	0.9980	0.5358	1.874
ECOG1	0.9619	1.0396	0.2881	3.212

Concordance= 0.728 (se = 0.032)

Likelihood ratio test= 29.24 on 4 df, p=7e-06

Wald test = 24.52 on 4 df, p=6e-05

Score (logrank) test = 31.36 on 4 df, p=3e-06

Hide

```
#Same analysis; Non ACT as reference
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data[circ_data$NAC=="FALSE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE", "TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$Colon <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon", "Rectum"))
circ_data$ECOG <- factor(circ_data$ECOG, levels=c("0", "1"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-HIGH"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Gender + Age.Group + ECOG, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ACT + Gender + Age.Group + ECOG,
 data = circ_data)

n= 54, number of events= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
ACTTRUE	-1.998062	0.135598	0.405506	-4.927	8.34e-07 ***
GenderMale	-0.173869	0.840407	0.330194	-0.527	0.598
Age.Group≥70	0.001981	1.001983	0.319386	0.006	0.995
ECOG1	-0.038826	0.961919	0.615096	-0.063	0.950

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTTRUE	0.1356	7.375	0.06125	0.3002
GenderMale	0.8404	1.190	0.43998	1.6053
Age.Group≥70	1.0020	0.998	0.53579	1.8738
ECOG1	0.9619	1.040	0.28812	3.2115

Concordance= 0.728 (se = 0.032)
 Likelihood ratio test= 29.24 on 4 df, p=7e-06
 Wald test = 24.52 on 4 df, p=6e-05
 Score (logrank) test = 31.36 on 4 df, p=3e-06

#DFS by ctDNA Clearance ACT-treated at 3 months - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
#S
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "POSITIVE" & ctDNA.3months == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.3months == "POSITIVE" ~ 2
  ))

circ_data <- circ_data[circ_data$DFS.3mo.months>=0,]
survfit(Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Event) ~
  ctDNA.Dynamics, data = circ_data)

674 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 100      42  27.53  18.07      NA
ctDNA.Dynamics=2  71      64   4.14   3.22   5.55
```

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

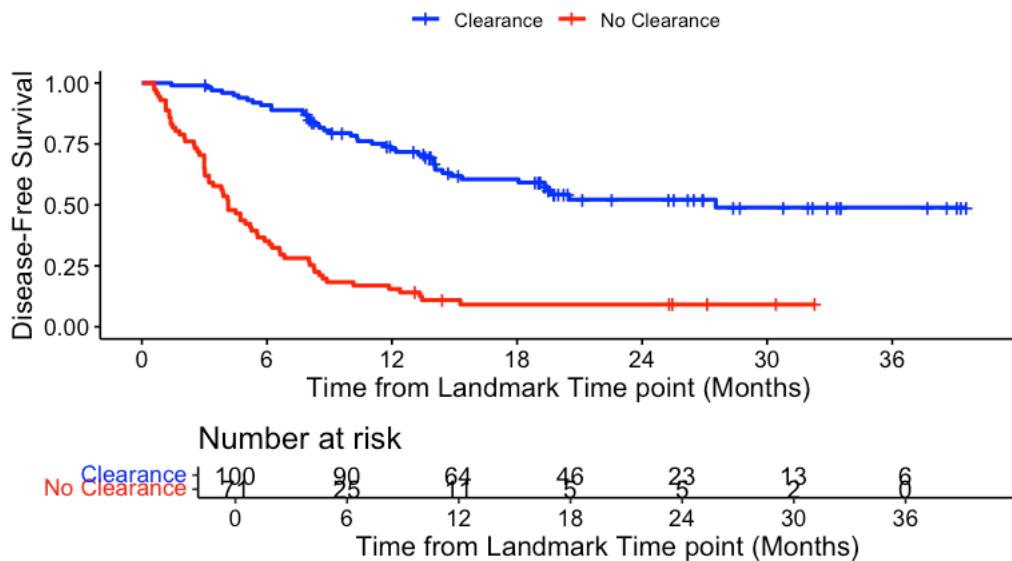
ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	100	42	0.4200000	42.00000
2	71	64	0.9014085	90.14085
NA	674	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA Clearance from MRD to 3 months ACT-treated | All Stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title="")
```

DFS - ctDNA Clearance from MRD to 3 months ACT-treated | All Subjects



Hide

summary(KM_curve, times= c(24))

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

674 observations deleted due to missingness

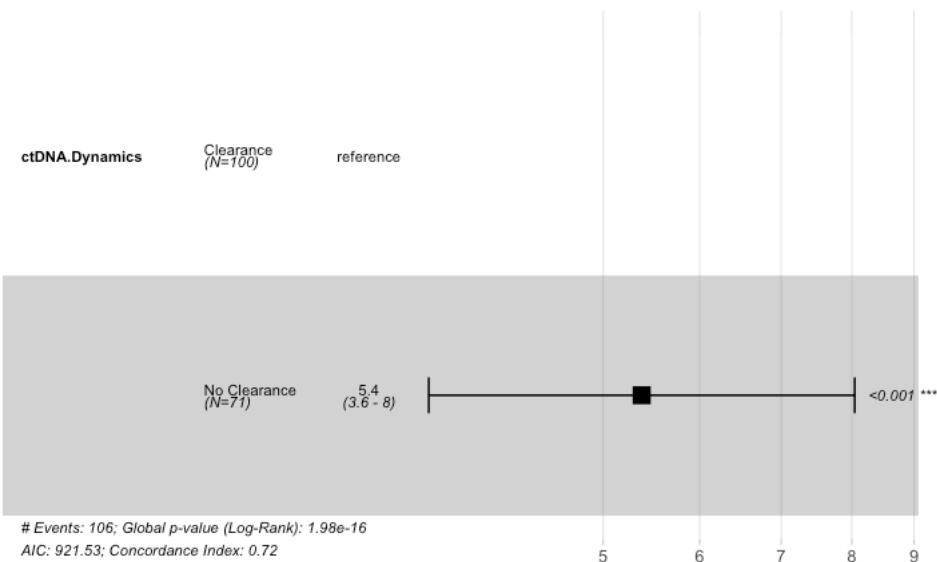
ctDNA.Dynamics=1						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	23.0000	41.0000	0.5217	0.0571	0.4047	0.6263

ctDNA.Dynamics=2						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	5.0000	64.0000	0.0913	0.0355	0.0372	0.1753

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2"), labels = c("Clearance", "No Clearance"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 171, number of events= 106
(674 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance 1.6822    5.3775  0.2055 8.187 2.67e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     5.378      0.186     3.595     8.044

Concordance= 0.716 (se = 0.018 )
Likelihood ratio test= 67.62 on 1 df,  p=<2e-16
Wald test      = 67.03 on 1 df,  p=3e-16
Score (logrank) test = 81.18 on 1 df,  p=<2e-16
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 5.38 (3.59-8.04); p = 0"
```

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 38.82, df = 1, p-value = 4.647e-10
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 4.578e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 5.03289 35.47066
sample estimates:
odds ratio
12.42856
```

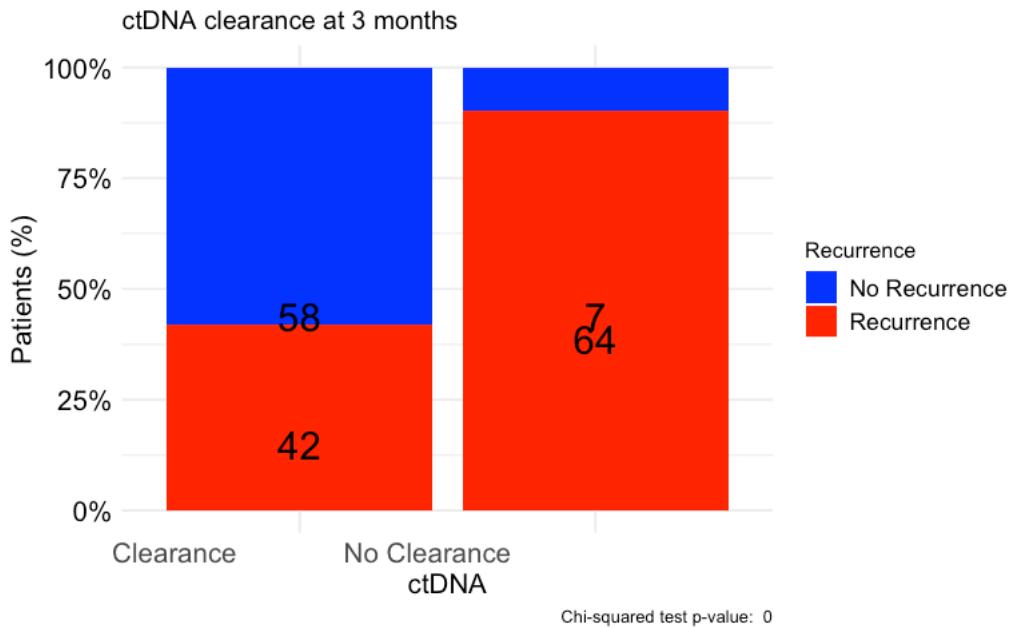
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Clearance	58	42
No Clearance	7	64

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA clearance at 3 months",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#OS by ctDNA Clearance ACT-treated at 3 months - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
#s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "POSITIVE" & ctDNA.3months == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.3months == "POSITIVE" ~ 2
  ))

circ_data <- circ_data[circ_data$OS.3mo.months>=0,]
survfit(Surv(time = circ_data$OS.3mo.months, event = circ_data$OS.Event)~ctDNA.Dynamics, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.3mo.months, event = circ_data$OS.Event) ~
  ctDNA.Dynamics, data = circ_data)

674 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 100      7     NA     NA     NA
ctDNA.Dynamics=2  71     16   41.6   31.9     NA
```

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

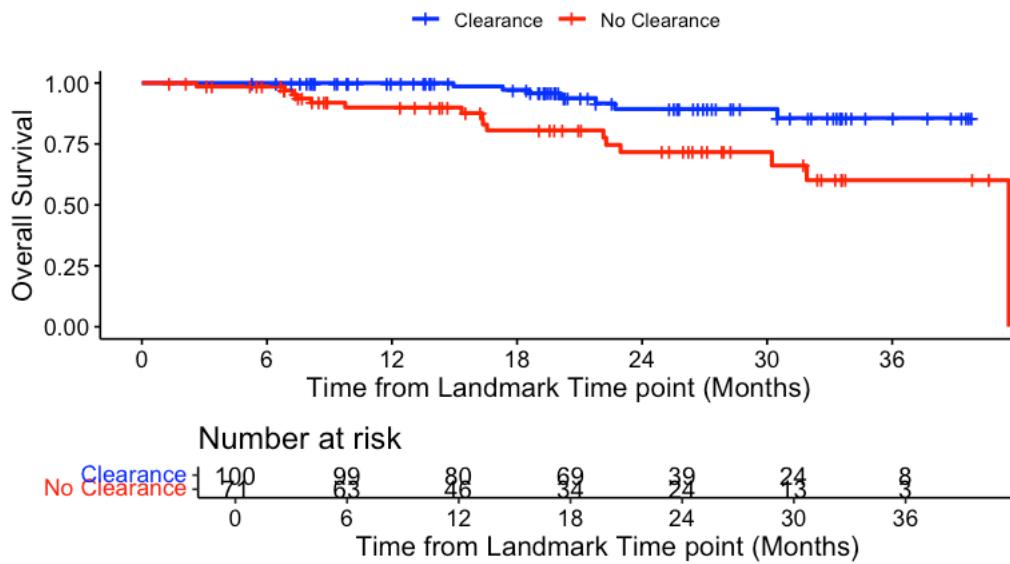
ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	100	7	0.0700000	7.00000
2	71	16	0.2253521	22.53521
NA	674	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.3mo.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="OS - ctDNA Clearance from MRD to 3 months ACT-treated | All Stages", ylab= "Overall Sur
vival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title
="")
```

OS - ctDNA Clearance from MRD to 3 months ACT-treated | All St



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

674 observations deleted due to missingness

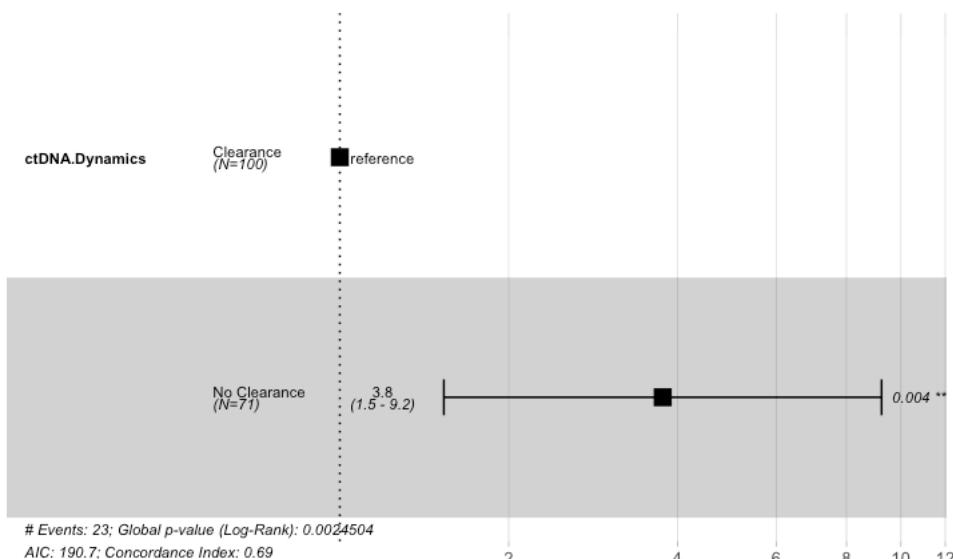
```
ctDNA.Dynamics=1
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
24.0000    39.0000     6.0000     0.8936     0.0423     0.7738     0.9519
```

```
ctDNA.Dynamics=2
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
24.0000    24.0000    13.0000     0.716      0.070      0.553      0.829
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2"), labels = c("Clearance", "No Clearance"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 171, number of events= 23
(674 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance 1.3251    3.7627  0.4583 2.892  0.00383 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     3.763     0.2658    1.533    9.238

Concordance= 0.689 (se = 0.047 )
Likelihood ratio test= 9.18 on 1 df,  p=0.002
Wald test            = 8.36 on 1 df,  p=0.004
Score (logrank) test = 9.65 on 1 df,  p=0.002
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 3.76 (1.53-9.24); p = 0.004"
```

[Hide](#)

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 7.3252, df = 1, p-value = 0.0068
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.005462
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.387371 11.743872
sample estimates:
odds ratio
3.833332
```

[Hide](#)

```
print(contingency_table)
```

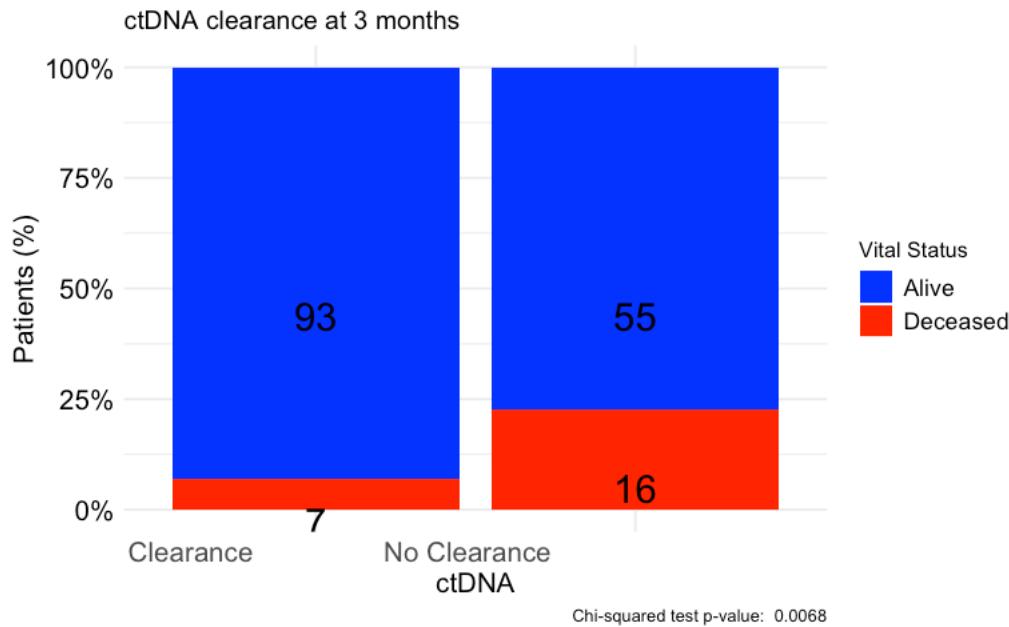
	Alive	Deceased
Clearance	93	7
No Clearance	55	16

Hide

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA clearance at 3 months",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) # define custom colors
theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
      axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
      axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
      axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
      legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



#DFS by ctDNA Clearance ACT-treated at 6 months - all stages

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "POSITIVE" ~ 2
  ))

circ_data <- circ_data[circ_data$DFS.6mo.months>=0,]
survfit(Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

```
Call: survfit(formula = Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Event) ~
  ctDNA.Dynamics, data = circ_data)

732 observations deleted due to missingness
  n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 77      27      NA    17.74      NA
ctDNA.Dynamics=2 35      34      2.4    1.61     3.68
```

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

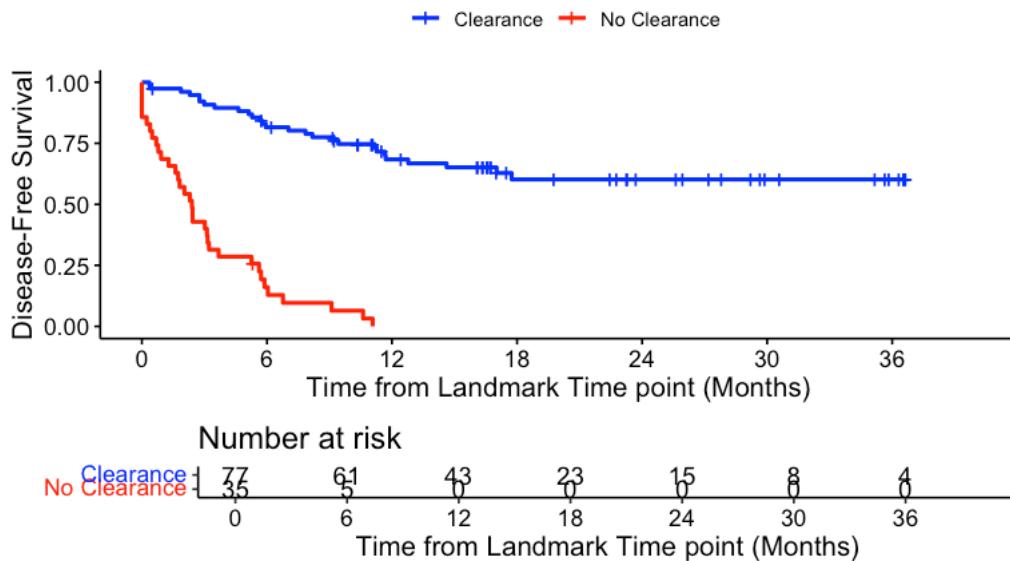
ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	77	27	0.3506494	35.06494
2	35	34	0.9714286	97.14286
NA	732	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA Clearance from MRD to 6 months ACT-treated | All Stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title="")
```

DFS - ctDNA Clearance from MRD to 6 months ACT-treated | All S



Hide

```
summary(KM_curve, times= c(6, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

732 observations deleted due to missingness

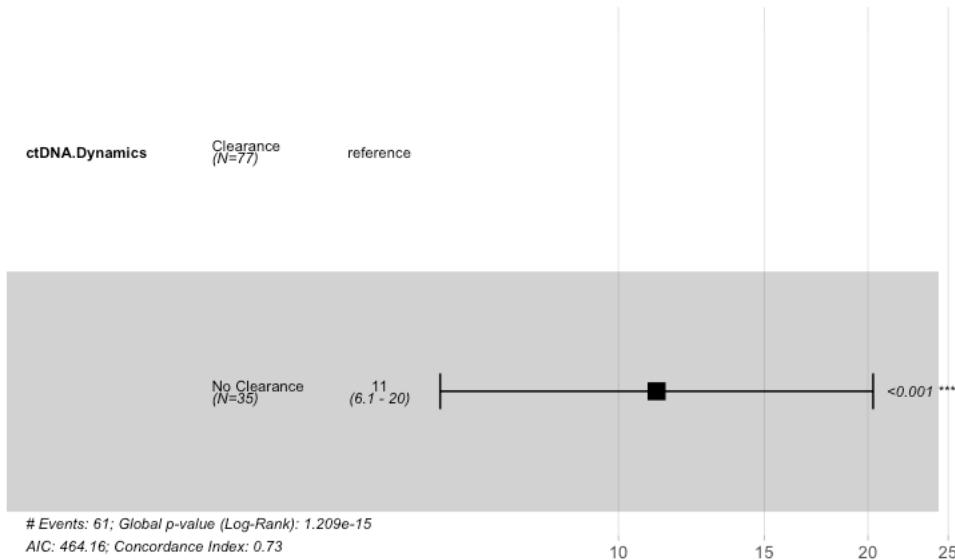
```
ctDNA.Dynamics=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  6      61     14     0.816  0.0445     0.709     0.886
 24      15     13     0.602  0.0625     0.469     0.712
```

```
ctDNA.Dynamics=2
time      n.risk      n.event      survival      std.err      lower 95% CI      upper 95% CI
6.0000  5.0000  29.0000  0.1607  0.0638  0.0609  0.3028
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2"), labels = c("Clearance", "No Clearance"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 112, number of events= 61
(732 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance  2.4088   11.1201   0.3069 7.848 4.24e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     11.12     0.08993    6.093    20.29

Concordance= 0.729 (se = 0.023 )
Likelihood ratio test= 64.06 on 1 df,  p=1e-15
Wald test      = 61.58 on 1 df,  p=4e-15
Score (logrank) test = 88.6 on 1 df,  p=<2e-16
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 11.12 (6.09-20.29); p = 0"
```

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 34.928, df = 1, p-value = 3.42e-09
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 7.575e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 9.188775 2571.010947
sample estimates:
odds ratio
 60.91302
```

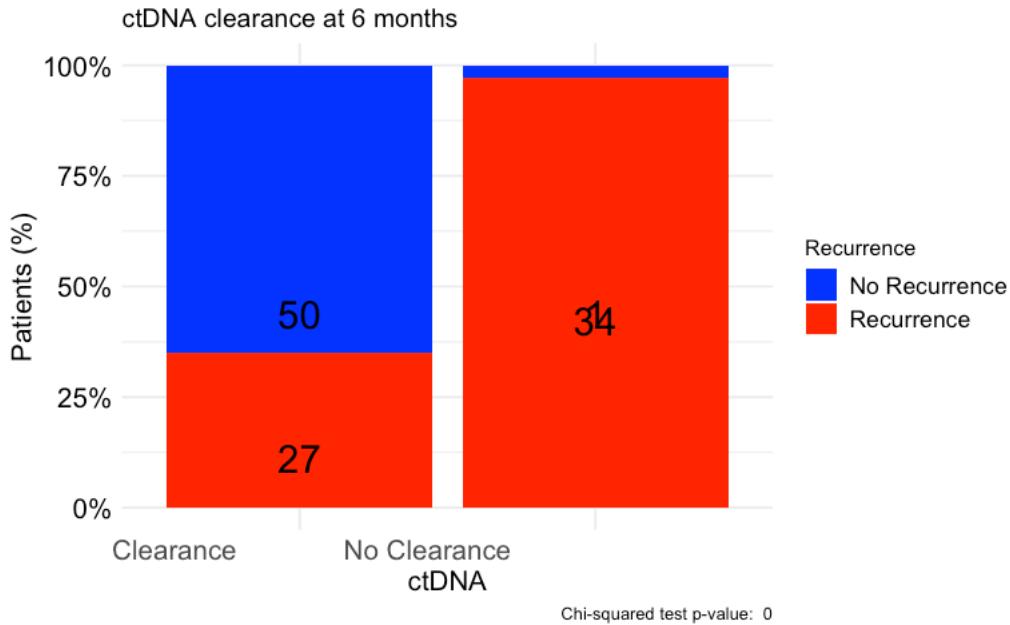
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Clearance	50	27
No Clearance	1	34

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA clearance at 6 months",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#OS by ctDNA Clearance ACT-treated at 6 months - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
#s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "POSITIVE" ~ 2
  ))

circ_data <- circ_data[circ_data$OS.6mo.months>=0,]
survfit(Surv(time = circ_data$OS.6mo.months, event = circ_data$OS.Event)~ctDNA.Dynamics, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.6mo.months, event = circ_data$OS.Event) ~
  ctDNA.Dynamics, data = circ_data)
```

```
732 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 77      3      NA      NA      NA
ctDNA.Dynamics=2 36      7      39     27.9     NA
```

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

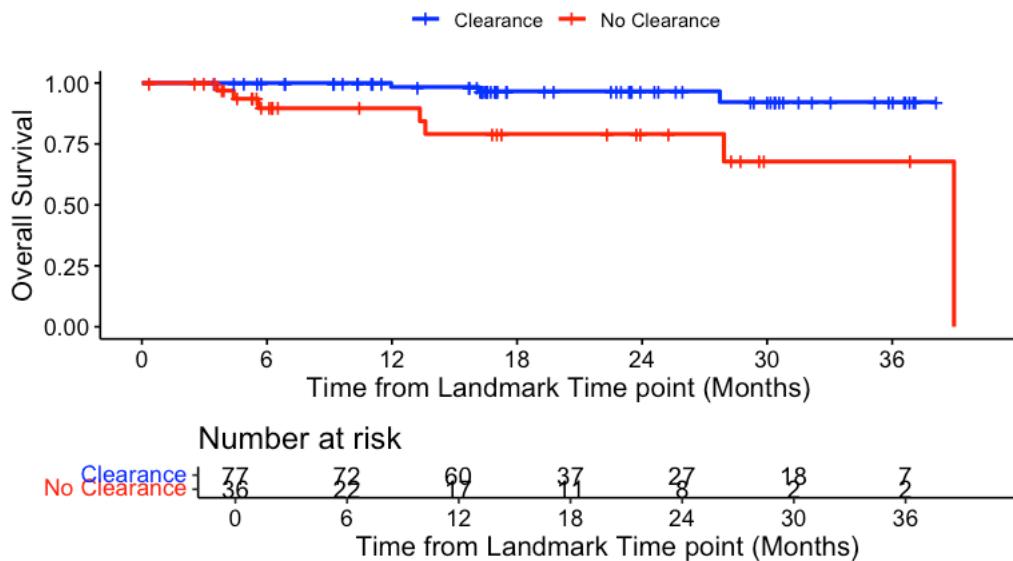
ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	77	3	0.03896104	3.896104
2	36	7	0.19444444	19.444444
NA	732	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.6mo.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="OS - ctDNA Clearance from MRD to 6 months ACT-treated | All Stages", ylab= "Overall Sur
vival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title
="")
```

OS - ctDNA Clearance from MRD to 6 months ACT-treated | All St



summary(KM_curve, times= c(6, 24))

Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int = 0.95, conf.type = "log-log")

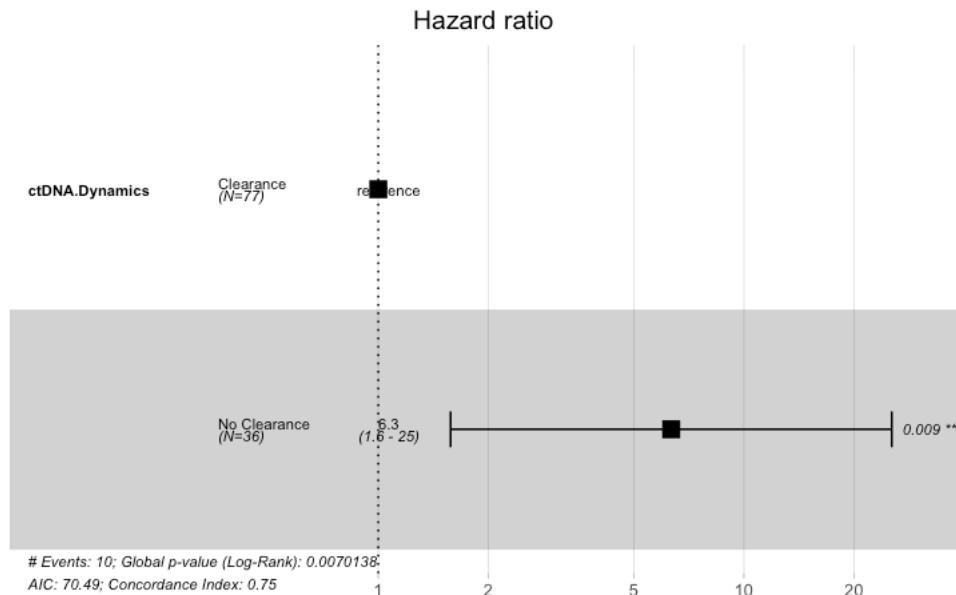
732 observations deleted due to missingness

ctDNA.Dynamics	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
1	6	72	0	1.000	0.0000	NA	NA	NA	NA
1	24	27	2	0.966	0.0236	0.871	0.991	0.991	0.991
2	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
2	6	22	3	0.896	0.0571	0.710	0.966	0.966	0.966
2	24	8	2	0.791	0.0863	0.558	0.910	0.910	0.910

circ_data\$ctDNA.Dynamics <- factor(circ_data\$ctDNA.Dynamics, levels=c("1","2"), labels = c("Clearance", "No Clearance"))

cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)

ggforest(cox_fit, data = circ_data)



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 113, number of events= 10
(732 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance 1.8445    6.3252   0.7088 2.602  0.00926 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance    6.325     0.1581    1.577    25.37

Concordance= 0.747 (se = 0.071 )
Likelihood ratio test= 7.27 on 1 df,  p=0.007
Wald test            = 6.77 on 1 df,  p=0.009
Score (logrank) test = 8.89 on 1 df,  p=0.003
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 6.33 (1.58-25.37); p = 0.009"
```

Hide

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 5.5507, df = 1, p-value = 0.01847
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data
```

```
data: contingency_table
p-value = 0.01138
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.232449 37.417947
sample estimates:
odds ratio
5.846658
```

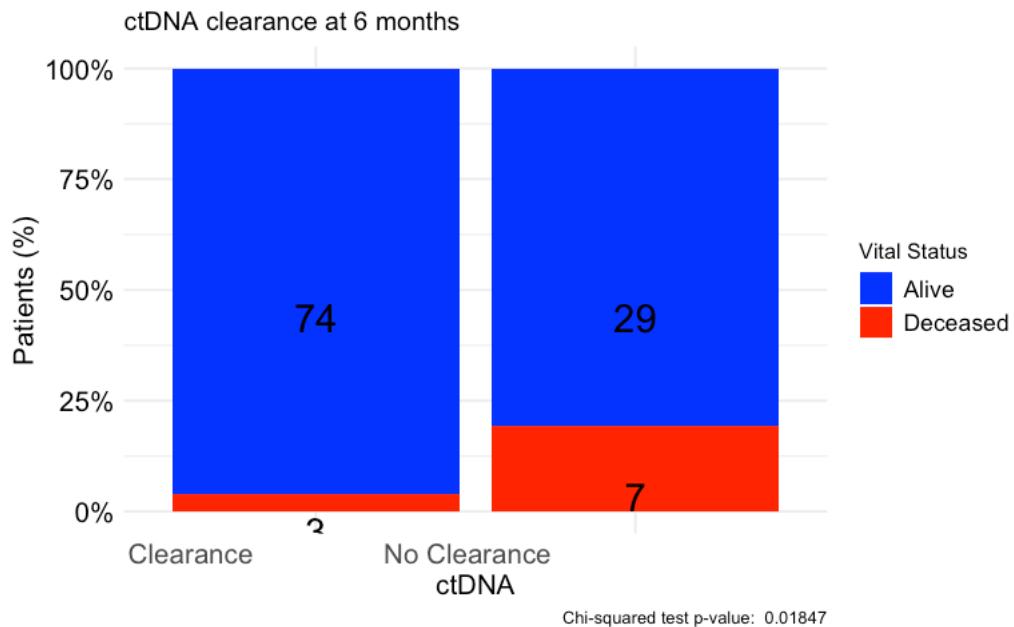
Hide

```
print(contingency_table)
```

	Alive	Deceased
Clearance	74	3
No Clearance	29	7

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA clearance at 6 months",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#Sankey plot for 3 months to 6 months ctDNA clearance

[Hide](#)

```
##To run this commands, please visit: https://sankeymatic.com/build/
#ctDNA + MRD window [185] ACT-treated #ADD8E6
#ctDNA + MRD window [151] Not treated #808080
#ACT-treated [100] ctDNA Clearance at 3 months #87EA86
#ACT-treated [71] No Clearance at 3 months #E67272
#ACT-treated [14] No 3 months time point #808080
#ctDNA Clearance at 3 months [7] ctDNA + at 6 months #E67272
#ctDNA Clearance at 3 months [64] ctDNA - at 6 months #87EA86
#ctDNA Clearance at 3 months [29] No 6 months time point #808080
#No Clearance at 3 months [27] ctDNA + at 6 months #E67272
#No Clearance at 3 months [11] ctDNA - at 6 months #87EA86
#No Clearance at 3 months [33] No 6 months time point #808080
#No 3 months time point [2] ctDNA + at 6 months #E67272
#No 3 months time point [2] ctDNA - at 6 months #87EA86
#No 3 months time point [10] No 6 months time point #808080
```

#Number of MRD positive patients & ctDNA clearance on ACT

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

# Count the number of MRD positive patients
number_of_positive_patients <- sum(circ_datadf$ctDNA.MRD == "POSITIVE", na.rm = TRUE)
print(paste("Number of MRD positive patients:", number_of_positive_patients))
```

[1] "Number of MRD positive patients: 336"

[Hide](#)

```
# Count the number & percentage of MRD positive patients treated with ACT
positive_subset <- sum(circ_datadf$ACT == "TRUE" & circ_datadf$ctDNA.MRD == "POSITIVE", na.rm = TRUE)
print(paste("Number of MRD positive patients treated with ACT:", positive_subset))
```

[1] "Number of MRD positive patients treated with ACT: 185"

[Hide](#)

```
percentage_positive_for_both <- (positive_subset / number_of_positive_patients) * 100
print(paste("Percentage of MRD positive patients treated with ACT:", percentage_positive_for_both, "%"))
```

[1] "Percentage of MRD positive patients treated with ACT: 55.0595238095238 %"

[Hide](#)

```
# Count the number & percentage of patients with ctDNA clearance post-ACT
clearance_postACT <- sum(
  (circ_datadf$ACT == "TRUE") &
  (circ_datadf$ctDNA.MRD == "POSITIVE") &
  (circ_datadf$Clearance.Event == "TRUE"),
  na.rm = TRUE
)
print(paste("Number of patients with ctDNA Clearance post-ACT:", clearance_postACT))
```

[1] "Number of patients with ctDNA Clearance post-ACT: 126"

[Hide](#)

```
percentage_clearance <- (clearance_postACT / positive_subset) * 100
print(paste("ctDNA Clearance post-ACT:", percentage_clearance, "%"))
```

[1] "ctDNA Clearance post-ACT: 68.1081081081081 %"

[Hide](#)

```
# Count the number of patients with subsequent timepoints available
clearance_subset <- sum(
  (circ_datadf$ACT == "TRUE") &
  (circ_datadf$ctDNA.MRD == "POSITIVE") &
  (circ_datadf$Transient.Clearance == "TRUE" | circ_datadf$Transient.Clearance == "FALSE"),
  na.rm = TRUE
)
print(paste("Number of patients with subsequent timepoints available:", clearance_subset))
```

[1] "Number of patients with subsequent timepoints available: 126"

[Hide](#)

```
# Count the number & percentage of patients with sustained clearance
clearance_sustained <- sum(
  (circ_datadf$ACT == "TRUE") &
  (circ_datadf$ctDNA.MRD == "POSITIVE") &
  (circ_datadf$Transient.Clearance == "FALSE"),
  na.rm = TRUE
)
print(paste("Number of patients with sustained clearance:", clearance_sustained))
```

[1] "Number of patients with sustained clearance: 68"

Hide

```
percentage_sustained_clearance <- (clearance_sustained / clearance_subset) * 100
print(paste("Sustained ctDNA Clearance:", percentage_sustained_clearance, "%"))
```

[1] "Sustained ctDNA Clearance: 53.968253968254 %"

Hide

```
# Count the number & percentage of patients with transient clearance
clearance_transient <- sum(
  (circ_datadf$ACT == "TRUE") &
  (circ_datadf$ctDNA.MRD == "POSITIVE") &
  (circ_datadf$Transient.Clearance == "TRUE"),
  na.rm = TRUE
)
print(paste("Number of patients with transient clearance:", clearance_transient))
```

[1] "Number of patients with transient clearance: 58"

Hide

```
percentage_transient_clearance <- (clearance_transient / clearance_subset) * 100
print(paste("Transient ctDNA Clearance:", percentage_transient_clearance, "%"))
```

[1] "Transient ctDNA Clearance: 46.031746031746 %"

#Sankey plot for Sustained vs Transient Clearance

Hide

```
##To run this commands, please visit: https://sankeymatic.com/build/
#ctDNA + MRD window [185] ACT-treated #ADD8E6
#ctDNA + MRD window [151] Not treated #808080
#ACT-treated [126] ctDNA post-MRD Clearance #87EA86
#ACT-treated [55] No Clearance #E67272
#ACT-treated [4] No post-MRD time point #808080
#No Clearance [55] No Clearance analysis #E67272
#ctDNA post-MRD Clearance [126] Available post-MRD Timepoints #ADD8E66
#Available post-MRD Timepoints [68] Sustained Clearance #7393B3
#Available post-MRD Timepoints [58] Transient Clearance #87EA86
```

#DFS by ctDNA Clearance post-MRD - 3 Groups

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_data <- circ_data[circ_data$ctDNA.Clearance!="",]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.Clearance, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.MRD.months, event = circ_data\$DFS.Event) ~ ctDNA.Clearance, data = circ_data)

131 observations deleted due to missingness
 n events median 0.95LCL 0.95UCL
 ctDNA.Clearance=No Clearance 55 55 4.83 4.53 5.45
 ctDNA.Clearance=Sustained 68 7 NA NA NA
 ctDNA.Clearance=Transient 58 50 12.88 10.38 15.64

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.Clearance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

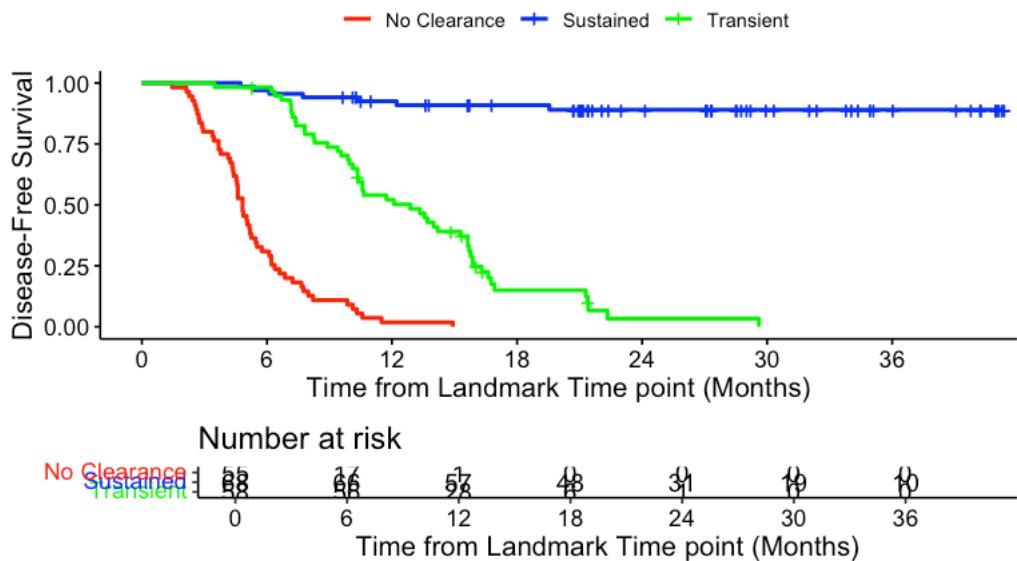
ctDNA.Clearance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
No Clearance	55	55	1.0000000	100.00000
Sustained	68	7	0.1029412	10.29412
Transient	58	50	0.8620690	86.20690
NA	131	NA	NA	NA

4 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Clearance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("red","blue","green"), title="DFS - ctDNA Clearance post-MRD | All Stages", ylab= "Disease-Free Survival", xl
ab="Time from Landmark Time point (Months)", legend.labs=c("No Clearance", "Sustained", "Transient"), legend.titl
e="")
```

DFS - ctDNA Clearance post-MRD | All Stages



```
summary(KM_curve, times= c(12, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Clearance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
131 observations deleted due to missingness
```

```
ctDNA.Clearance=No Clearance
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12.00000	1.00000	54.00000	0.01818	0.01802	0.00149	0.08474

```
ctDNA.Clearance=Sustained
```

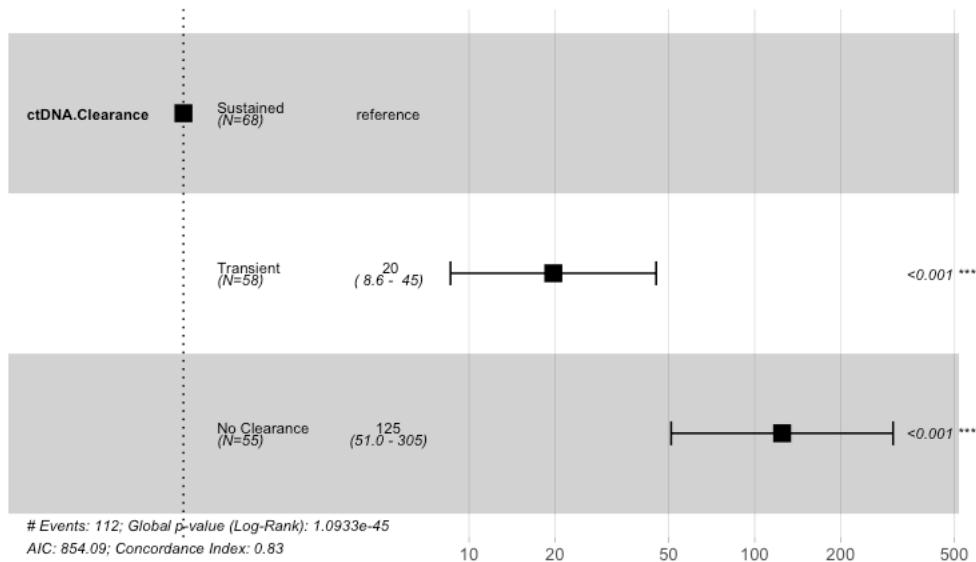
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	57	5	0.925	0.0321	0.830	0.968
18	48	1	0.909	0.0354	0.809	0.958
24	31	1	0.890	0.0394	0.783	0.946

```
ctDNA.Clearance=Transient
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	28	27	0.5212	0.0668	0.38358	0.642
18	6	18	0.1500	0.0527	0.06542	0.267
24	1	4	0.0333	0.0312	0.00294	0.137

```
circ_data$ctDNA.Clearance <- factor(circ_data$ctDNA.Clearance, levels=c("Sustained","Transient", "No Clearance"))
cox_fit <- coxph(surv_object ~ ctDNA.Clearance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
coxph(formula = surv_object ~ ctDNA.Clearance, data = circ_data)

n= 181, number of events= 112
(131 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.ClearanceTransient	2.9815	19.7182	0.4229	7.051	1.78e-12 ***
ctDNA.ClearanceNo Clearance	4.8264	124.7631	0.4565	10.573	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.ClearanceTransient	19.72	0.050715	8.608	45.17
ctDNA.ClearanceNo Clearance	124.76	0.008015	50.996	305.24

Concordance= 0.83 (se = 0.017)

Likelihood ratio test= 207 on 2 df, p=<2e-16

Wald test = 129.1 on 2 df, p=<2e-16

Score (logrank) test = 234.2 on 2 df, p=<2e-16

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
#Calculate the exact p value for No Clearance
z_value <- cox_fit_summary$coefficients["ctDNA.ClearanceNo Clearance", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for No Clearance:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for No Clearance: 3.96544742045303382828672210252136156837230329342884918473580640893060107761
0622e-26"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.Clearance, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test
```

```
data: contingency_table
X-squared = 125.14, df = 2, p-value < 2.2e-16
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data
```

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: two.sided
```

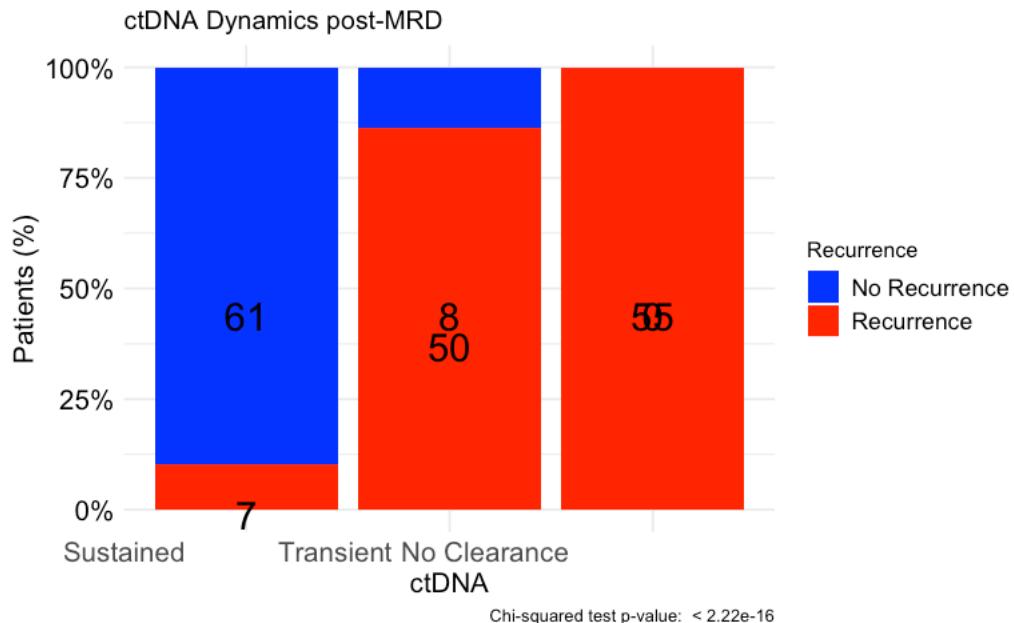
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Sustained	61	7
Transient	8	50
No Clearance	0	55

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA Dynamics post-MRD",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 6.68956602289894660144436130668335008263164621356783259010046013887291777
7164151e-28"
```

#Levels of MRD MTM/mL in Clearance post-MRD log10 transformation

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[!is.na(circ_data$ctDNA.Clearance) & circ_data$ctDNA.Clearance != "",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_data <- as.data.frame(circ_data)

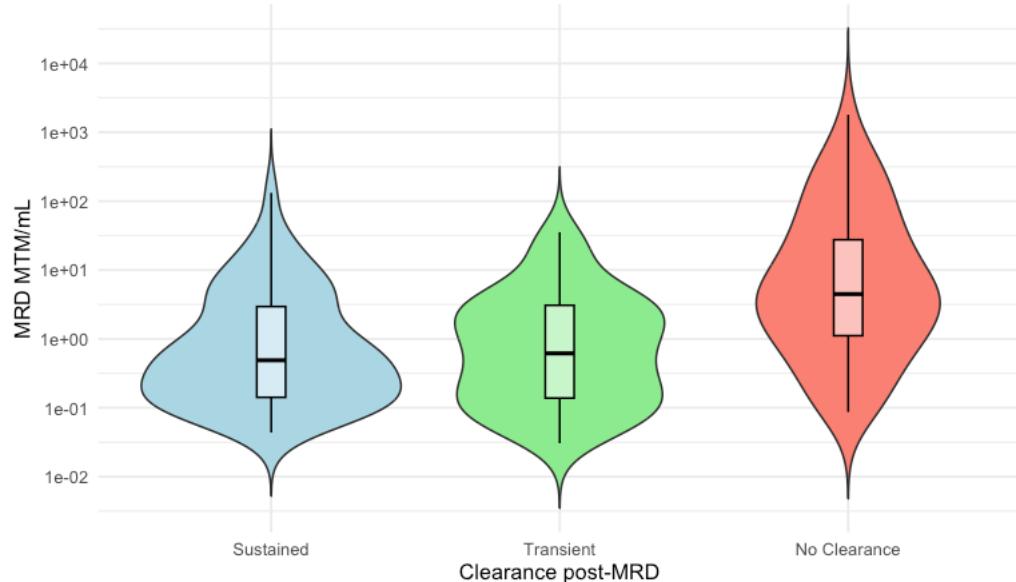
# Transform p_MRDT with log10
circ_data$p_MRDT <- as.numeric(as.character(circ_data$p_MRDT))
circ_data$ctDNA.Clearance <- factor(circ_data$ctDNA.Clearance, levels=c("Sustained", "Transient", "No Clearance"))
median_p_MRDT <- aggregate(p_MRDT ~ ctDNA.Clearance, data = circ_data, FUN = median)
print(median_p_MRDT)
```

ctDNA.Clearance	p_MRDT
<fctr>	<dbl>
Sustained	0.4906143
Transient	0.6186848
No Clearance	4.4903396
3 rows	

[Hide](#)

```
# Create violin plot with log10 scale on y-axis
ggplot(circ_data, aes(x=ctDNA.Clearance, y=p_MRDT, fill=ctDNA.Clearance)) +
  geom_violin(trim=FALSE) +
  scale_fill_manual(values=c("Sustained"="lightblue", "Transient"="lightgreen", "No Clearance"="salmon")) +
  geom_boxplot(width=0.1, fill="white", colour="black", alpha=0.5) +
  scale_y_log10(breaks=c(0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000)) +
  labs(title="MRD MTM/mL | Clearance post-MRD", x="Clearance post-MRD", y="MRD MTM/mL") +
  theme_minimal() +
  theme(legend.position="none")
```

MRD MTM/mL | Clearance post-MRD

[Hide](#)

```
m3_1v2 <- wilcox.test(p_MRDTMT ~ ctDNA.Clearance,
                       data = circ_data[circ_data$ctDNA.Clearance %in% c("Sustained", "Transient"), ],
                       na.rm = TRUE)
print(m3_1v2)
```

```
Wilcoxon rank sum test with continuity correction

data: p_MRDTMT by ctDNA.Clearance
W = 1946, p-value = 0.9007
alternative hypothesis: true location shift is not equal to 0
```

[Hide](#)

```
m3_1v3 <- wilcox.test(p_MRDTMT ~ ctDNA.Clearance,
                       data = circ_data[circ_data$ctDNA.Clearance %in% c("Sustained", "No Clearance"), ],
                       na.rm = TRUE)
print(m3_1v3)
```

```
Wilcoxon rank sum test with continuity correction

data: p_MRDTMT by ctDNA.Clearance
W = 906, p-value = 9.529e-07
alternative hypothesis: true location shift is not equal to 0
```

[Hide](#)

```
m3_2v3 <- wilcox.test(p_MRDTMT ~ ctDNA.Clearance,
                       data = circ_data[circ_data$ctDNA.Clearance %in% c("Transient", "No Clearance"), ],
                       na.rm = TRUE)
print(m3_2v3)
```

```
Wilcoxon rank sum test with continuity correction

data: p_MRDTMT by ctDNA.Clearance
W = 782, p-value = 3.052e-06
alternative hypothesis: true location shift is not equal to 0
```

#Percentages of recurred transient clearance that return positive

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data <- circ_data[circ_data$ACT=="TRUE",]
circ_data <- circ_data[circ_data$Clearance.Event=="TRUE",]
circ_data <- circ_data[circ_data$DFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_data <- subset(circ_data, !is.na(Transient.Clearance))
circ_data <- circ_data[circ_data$Transient.Clearance=="TRUE",]
circ_datadf <- as.data.frame(circ_data)

# Convert days to months
circ_data$p_drelReturned_months <- circ_data$p_drelReturned / 30.437

# Define the intervals: 6-9, 9-12, 12-15, 15-18, 18-21, 21-24, >24 months
breaks <- c(3, 6, 9, 12, 15, 18, 21, 24, 27)
labels <- c("3-6m", "6-9m", "9-12m", "12-15m", "15-18m", "18-21m", "21-24m", ">24m")

# Categorize p_drelReturned_months into intervals
circ_data$p_drelReturned_intervals <- cut(circ_data$p_drelReturned_months, breaks = breaks, labels = labels, right = FALSE)

# Examine the distribution of the intervals
table(circ_data$p_drelReturned_intervals)

```

3-6m	6-9m	9-12m	12-15m	15-18m	18-21m	21-24m	>24m
7	23	8	4	6	0	2	0

Hide

```

# Get the counts for each interval
interval_counts <- table(circ_data$p_drelReturned_intervals)

# Calculate the percentages
interval_percentages <- 100 * interval_counts / sum(interval_counts)

# Combine the counts and percentages for a clearer overview
interval_summary <- data.frame(Counts = interval_counts, Percentages = interval_percentages)

# Print the summary
print(interval_summary)

```

Counts.Var1	Counts.Freq	Percentages.Var1	Percentages.Freq
<fctr>	<int>	<fctr>	<dbl>
3-6m	7	3-6m	14
6-9m	23	6-9m	46
9-12m	8	9-12m	16
12-15m	4	12-15m	8
15-18m	6	15-18m	12
18-21m	0	18-21m	0
21-24m	2	21-24m	4
>24m	0	>24m	0
8 rows			

Hide

```

# Calculate cumulative percentages
cumulative_percentages <- cumsum(interval_percentages)

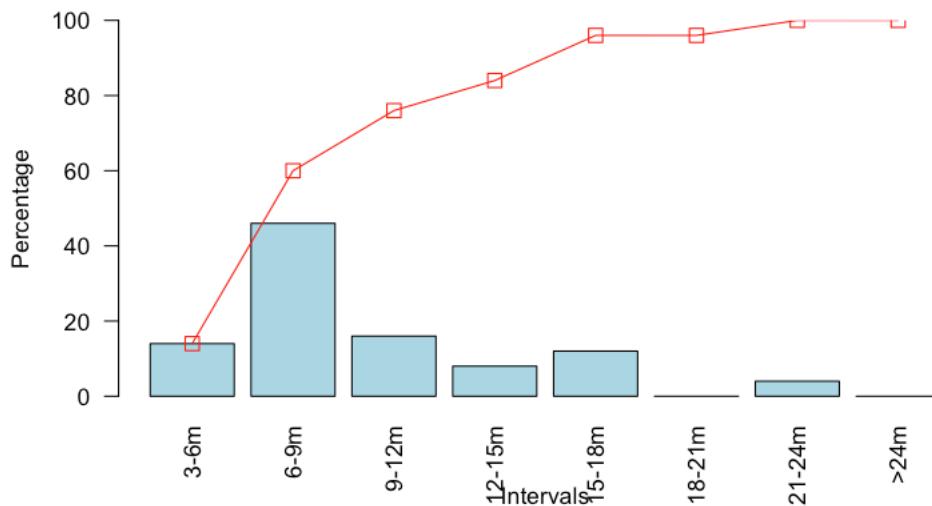
# Combine the counts and percentages for a clearer overview
interval_summary <- data.frame(Counts = interval_counts, Percentages = interval_percentages, CumulativePercentage = cumulative_percentages)

bp <- barplot(interval_percentages,
              main="Distribution of ctDNA Intervals",
              xlab="Intervals",
              ylab="Percentage",
              col="lightblue",
              ylim=c(0, 100),
              las=2) # las=2 makes the axis labels perpendicular to the axis

# Add the cumulative percentages to the plot
points(bp, cumulative_percentages, type="o", pch=22, col="red", cex=1.5)

```

Distribution of ctDNA Intervals



#OS by ctDNA Clearance post-MRD - 3 Groups

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$Clearance.Cohort=="TRUE",]
circ_dataadf <- as.data.frame(circ_data)
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$Event)

survfit(Surv(time = circ_data$OS.MRD.months, event = circ_data$Event)~ctDNA.Clearance, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$OS.MRD.months, event = circ_data\$Event) ~ ctDNA.Clearance, data = circ_data)

ctDNA.Clearance	n	events	median	0.95LCL	0.95UCL
No Clearance	55	17	32.5	23.9	NA
Sustained	68	0	NA	NA	NA
Transient	58	7	NA	NA	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.Clearance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

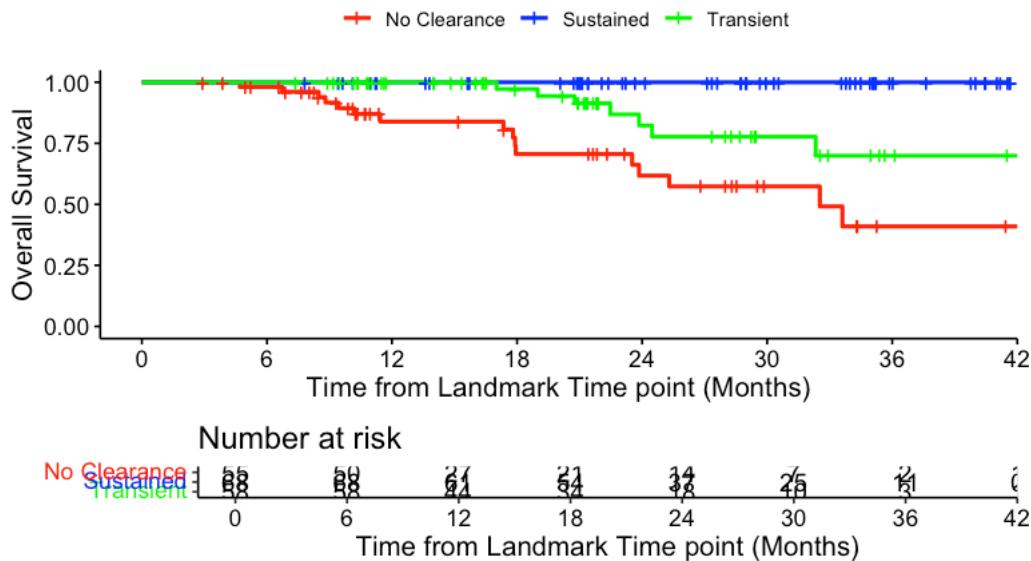
ctDNA.Clearance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
No Clearance	55	17	0.3090909	30.90909
Sustained	68	0	0.0000000	0.00000
Transient	58	7	0.1206897	12.06897

3 rows

Hide

```
KM_curve <- survfit(surv_object ~ ctDNA.Clearance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("red","blue","green"), title="OS - ctDNA Clearance post-MRD | All Stages", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("No Clearance", "Sustained", "Transient"), legend.title="")
```

OS - ctDNA Clearance post-MRD | All Stages



Hide

```
summary(KM_curve, times= c(12, 18, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Clearance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.Clearance=No Clearance
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12    27      7    0.839  0.0570    0.687  0.921
 18    21      4    0.706  0.0776    0.524  0.829
 24    14      2    0.617  0.0895    0.419  0.765
```

```
ctDNA.Clearance=Sustained
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12    61      0      1      0      NA      NA
 18    54      0      1      0      NA      NA
 24    37      0      1      0      NA      NA
```

```
ctDNA.Clearance=Transient
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12    44      0    1.000  0.0000      NA      NA
 18    34      1    0.972  0.0274    0.819  0.996
 24    18      4    0.823  0.0747    0.615  0.925
```

Hide

```
circ_data$ctDNA.Clearance <- as.factor(circ_data$ctDNA.Clearance)
circ_data$ctDNA.Clearance <- factor(circ_data$ctDNA.Clearance, levels=c("Sustained","Transient", "No Clearance"))
cox_fit <- coxphf(surv_object ~ ctDNA.Clearance, data=circ_data)
summary(cox_fit)
```

```
coxphf(formula = surv_object ~ ctDNA.Clearance, data = circ_data)
```

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

	coef	se(coef)	exp(coef)	lower 0.95	upper 0.95	Chisq	p
ctDNA.ClearanceTransient	3.239402	1.510369	25.51846	3.099164	3314.725	11.55743	6.747909e-04
ctDNA.ClearanceNo Clearance	4.325656	1.484378	75.61513	10.218215	9650.929	34.76657	3.717015e-09

Likelihood ratio test=34.78097 on 2 df, p=2.80161e-08, n=181
Wald test = 12.97638 on 2 df, p = 0.001521303

Covariance-Matrix:

	ctDNA.ClearanceTransient	ctDNA.ClearanceNo Clearance
ctDNA.ClearanceTransient	2.281214	2.138730
ctDNA.ClearanceNo Clearance	2.138730	2.203378

Hide

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Clearance, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test

```
data: contingency_table
X-squared = 25.362, df = 2, p-value = 3.109e-06
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 3.67e-07
alternative hypothesis: two.sided
```

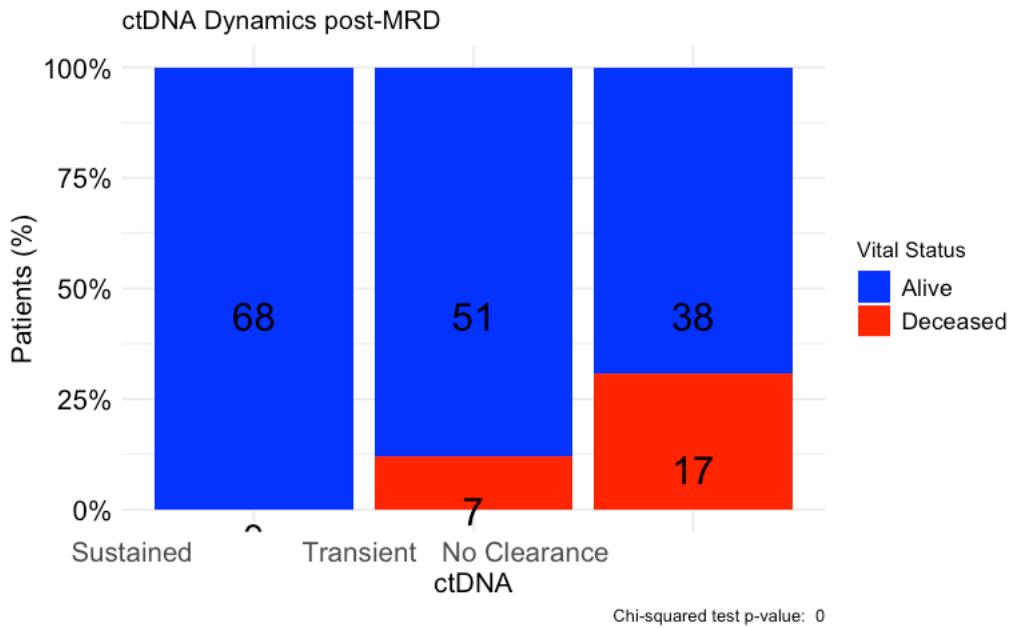
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Sustained	68	0
Transient	51	7
No Clearance	38	17

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA Dynamics post-MRD",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#Percentages of MRD negative with molecular recurrence (returned positive) post-MRD

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "" & circ_data$Lead.Time >= 0, ]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$PostMRDPos.Event=="TRUE",]
circ_dataadf <- as.data.frame(circ_data)

# Convert days to months
#circ_data$PostMRDPos.months <- circ_data$PostMRDPos / 30.437

# Define the intervals: 0-6, 6-9, 9-12, 12-15, 15-18, 18-21, 21-24, >24 months
breaks <- c(0, 6, 9, 12, 15, 18, 21, 24, 48)
labels <- c("0-6m", "6-9m", "9-12m", "12-15m", "15-18m", "18-21m", "21-24m", ">24m")

# Categorize p_drelReturned_months into intervals
circ_data$p_drelReturned_intervals <- cut(circ_data$PostMRDPos.months, breaks = breaks, labels = labels, right = FALSE)

# Examine the distribution of the intervals
table(circ_data$p_drelReturned_intervals)
```

Interval	Count
0-6m	77
6-9m	35
9-12m	23
12-15m	2
15-18m	20
18-21m	1
21-24m	7
>24m	0

[Hide](#)

```
# Get the counts for each interval
interval_counts <- table(circ_data$p_drelReturned_intervals)

# Calculate the percentages
interval_percentages <- 100 * interval_counts / sum(interval_counts)

# Combine the counts and percentages for a clearer overview
interval_summary <- data.frame(Counts = interval_counts, Percentages = interval_percentages)

# Calculate the total number of observations
total_observations <- sum(interval_counts)

# Add the total number of observations to the summary
interval_summary$TotalObservations <- c(rep(NA, length(interval_counts)-1), total_observations)

# Print the summary with total observations
print(interval_summary)
```

Counts.Var1 <fctr>	Counts.Freq <int>	Percentages.Var1 <fctr>	Percentages.Freq <dbl>	TotalObservations <int>
0-6m	77	0-6m	46.6666667	NA
6-9m	35	6-9m	21.2121212	NA
9-12m	23	9-12m	13.9393939	NA
12-15m	2	12-15m	1.2121212	NA
15-18m	20	15-18m	12.1212121	NA
18-21m	1	18-21m	0.6060606	NA
21-24m	7	21-24m	4.2424242	NA
>24m	0	>24m	0.0000000	165

8 rows

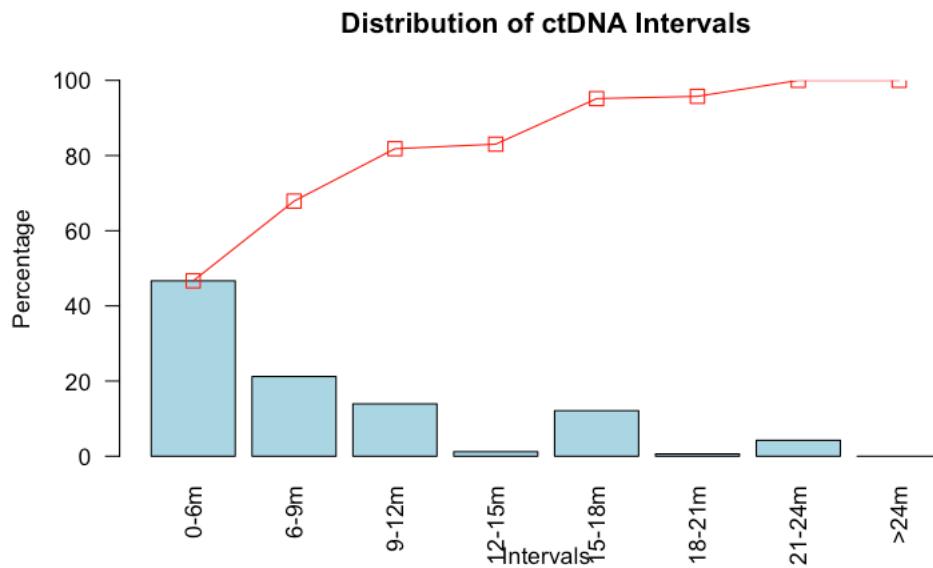
Hide

```
# Calculate cumulative percentages
cumulative_percentages <- cumsum(interval_percentages)

# Combine the counts, percentages, and cumulative percentages for a clearer overview
interval_summary <- data.frame(Counts = interval_counts, Percentages = interval_percentages, CumulativePercentage = cumulative_percentages, TotalObservations = c(rep(NA, length(interval_counts)-1), total_observations))

bp <- barplot(interval_percentages,
              main="Distribution of ctDNA Intervals",
              xlab="Intervals",
              ylab="Percentage",
              col="lightblue",
              ylim=c(0, 100),
              las=2) # las=2 makes the axis labels perpendicular to the axis

# Add the cumulative percentages to the plot
points(bp, cumulative_percentages, type="o", pch=22, col="red", cex=1.5)
```



Hide

print(interval_summary)

Counts.Var1 <fctr>	Counts.Freq <int>	Percentages.Var1 <fctr>	Percentages.Freq <dbl>	CumulativePercentages <dbl>
0-6m	0-6m	77 0-6m	46.666667	46.66667
6-9m	6-9m	35 6-9m	21.212122	67.87879
9-12m	9-12m	23 9-12m	13.9393939	81.81818
12-15m	12-15m	2 12-15m	1.2121212	83.03030

Counts.Var1 <fctr>	Counts.Freq <int>	Percentages.Var1 <fctr>	Percentages.Freq <dbl>	CumulativePercentages <dbl>
15-18m 15-18m	20	15-18m	12.1212121	95.15152
18-21m 18-21m	1	18-21m	0.6060606	95.75758
21-24m 21-24m	7	21-24m	4.2424242	100.00000
>24m >24m	0	>24m	0.0000000	100.00000

8 rows | 1-6 of 6 columns

#OS by ctDNA MRD positive vs ctDNA negative with molecular recurrence at Surveillance - 3 groups

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics
    = case_when(
      ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance=="NEGATIVE" ~ 1,
      ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance=="POSITIVE" ~ 2,
      ctDNA.MRD == "POSITIVE" ~ 3
    ))
  
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
survfit(Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)~ctDNA.Dynamics, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$OS.MRD.months, event = circ_data\$OS.Event) ~
ctDNA.Dynamics, data = circ_data)

321 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 1294 13 NA NA NA
ctDNA.Dynamics=2 159 15 NA NA NA
ctDNA.Dynamics=3 336 52 43.4 NA NA

Hide

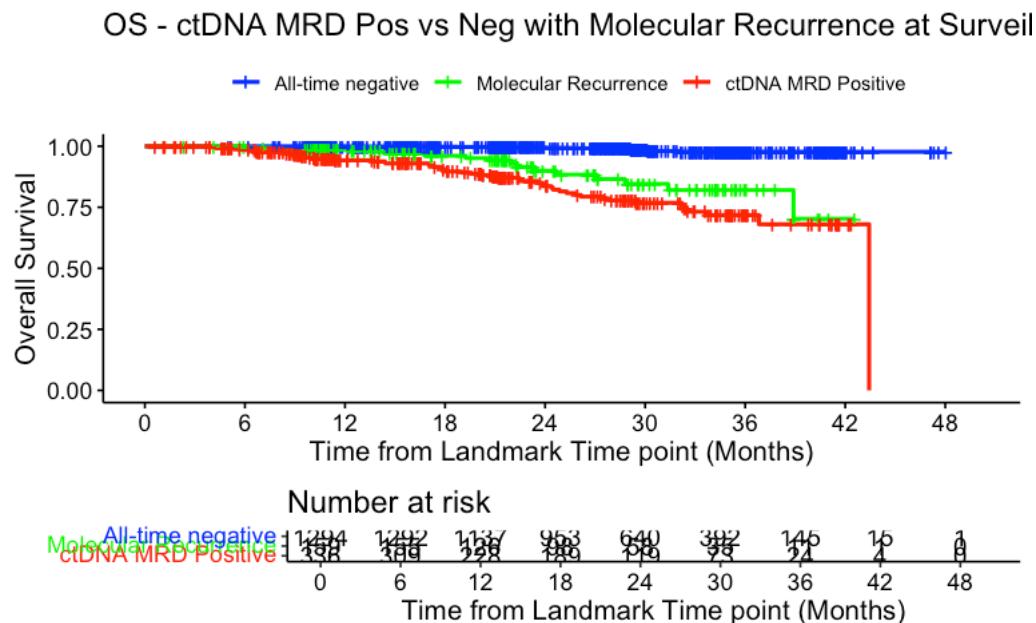
```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	1294	13	0.01004637	1.004637
2	159	15	0.09433962	9.433962
3	336	52	0.15476190	15.476190
NA	321	NA	NA	NA

4 rows

Hide

```
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","red"), title="OS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Surveillance Window", yl
ab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("All-time negative","Molecul
ar Recurrence", "ctDNA MRD Positive"), legend.title="")
```



```
summary(KM_curve, times= c(12, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

321 observations deleted due to missingness

```
ctDNA.Dynamics=1
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
12	1137	0	1.000	0.0000	NA	NA	NA	NA
24	640	5	0.995	0.0023	0.988	0.998	0.998	0.998

```
ctDNA.Dynamics=2
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
12	126	2	0.987	0.00909	0.949	0.997	0.997	0.997
24	58	8	0.900	0.03138	0.817	0.946	0.946	0.946

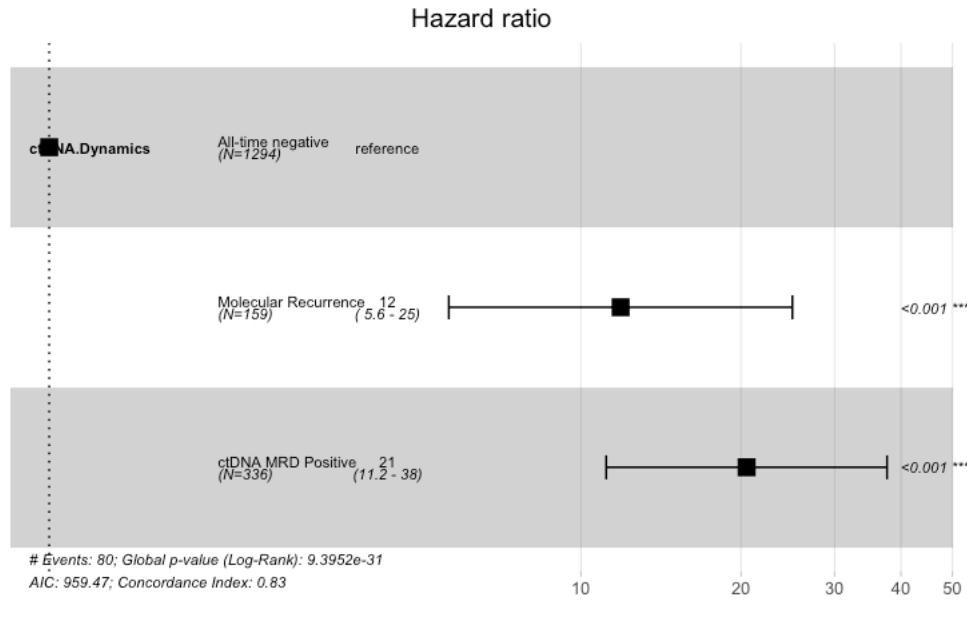
```
ctDNA.Dynamics=3
```

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
12	228	17	0.942	0.0136	0.909	0.964	0.964	0.964
24	119	20	0.837	0.0258	0.778	0.881	0.881	0.881

Hide

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3"), labels = c("All-time negative", "Molecular Recurrence", "ctDNA MRD Positive"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```



Hide

```
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 1789, number of events= 80
 (321 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)						
ctDNA.Dynamics	2.4747	11.8787	0.3796	6.519	7.09e-11 ***						
Molecular Recurrence	3.0205	20.5007	0.3103	9.734	< 2e-16 ***						

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Dynamics	11.88	0.08418	5.644	25.00
Molecular Recurrence	20.50	0.04878	11.160	37.66

Concordance= 0.833 (se = 0.019)
 Likelihood ratio test= 138.3 on 2 df, p=<2e-16
 Wald test = 94.79 on 2 df, p=<2e-16
 Score (logrank) test = 182.9 on 2 df, p=<2e-16

Hide

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test

```
data: contingency_table
X-squared = 140.83, df = 2, p-value < 2.2e-16
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: two.sided
```

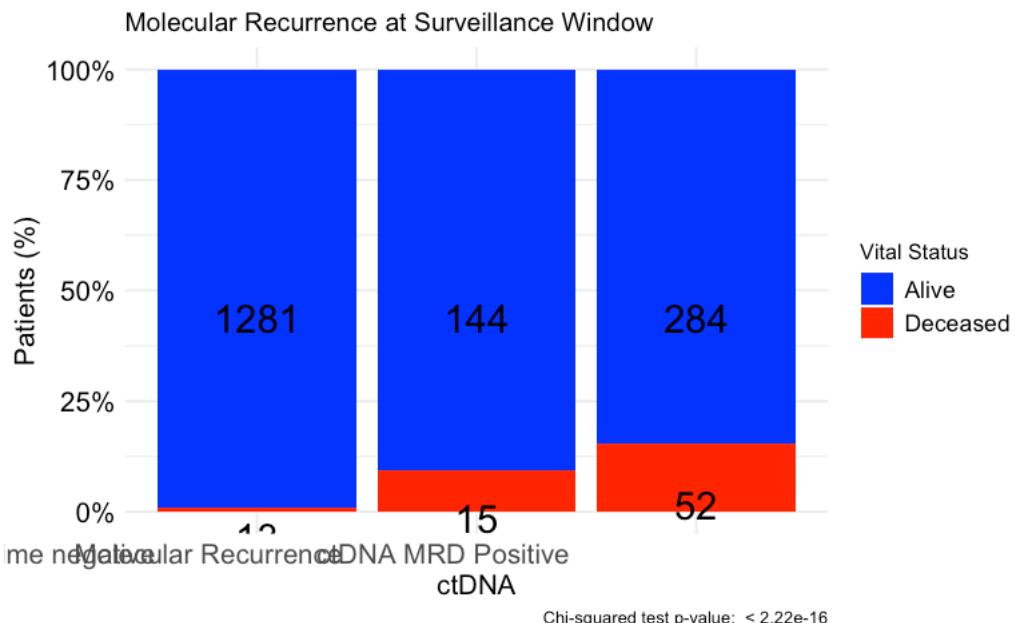
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
All-time negative	1281	13
Molecular Recurrence	144	15
ctDNA MRD Positive	284	52

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "Molecular Recurrence at Surveillance Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 2.6277212849116997866193852202655064971549656459050763069527842062226455
1055174e-31"
```

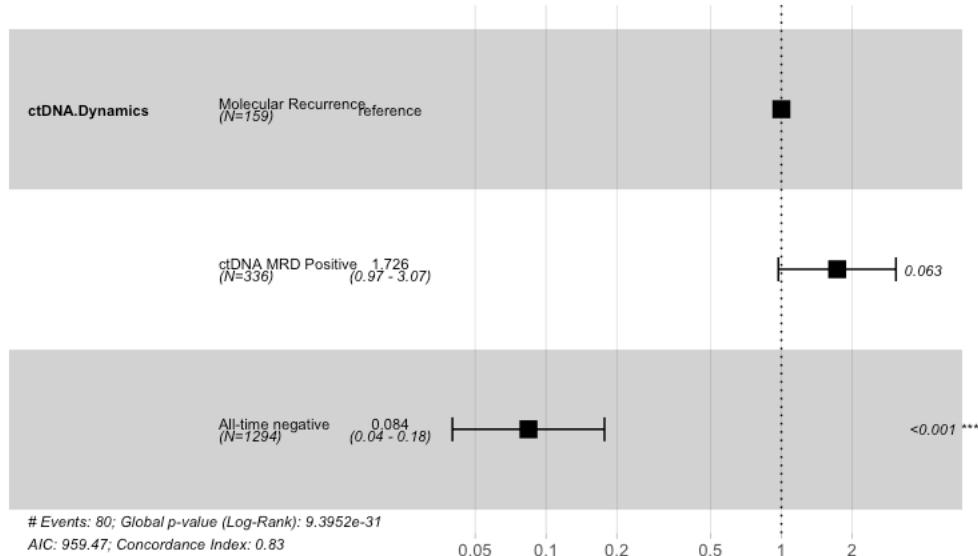
Hide

```
rm(list=ls()) #repeat to compare Molecular Recurrence vs ctDNA MRD positive
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance=="NEGATIVE" ~ 1,
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance=="POSITIVE" ~ 2,
    ctDNA.MRD == "POSITIVE" ~ 3
  ))

circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("2","3","1"), labels = c("Molecular Recurrence",
  "ctDNA MRD Positive", "All-time negative"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 1789, number of events= 80
(321 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsctDNA MRD Positive  0.54572  1.72584  0.29355  1.859   0.063 .
ctDNA.DynamicsAll-time negative -2.47474  0.08418  0.37964 -6.519 7.09e-11 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsctDNA MRD Positive  1.72584   0.5794   0.9708   3.0681
ctDNA.DynamicsAll-time negative  0.08418  11.8787   0.0400   0.1772

Concordance= 0.833 (se = 0.019 )
Likelihood ratio test= 138.3 on 2 df,  p=<2e-16
Wald test            = 94.79 on 2 df,  p=<2e-16
Score (logrank) test = 182.9 on 2 df,  p=<2e-16

```

#Time-dependent analysis - Molecular Recurrence patients Landmark from molecular recurrence with RFS event as outcome

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
dt_final <- read.csv("Galaxy 36mo Time dependent.csv")
dt_final <- dt_final[!is.na(dt_final$tstart4), ]
dt_final$tstart4 <- as.numeric(as.character(dt_final$tstart4))
dt_final$tstop4 <- as.numeric(as.character(dt_final$tstop4))

```

Warning: NAs introduced by coercion

[Hide](#)

```
datatable(dt_final, filter = "top")
```

Show 10 [▼](#) entries

Search:

	pts_id	rfs_time	rfs_event	os_time	os_event	rfs_date	os_date	mrd_date	molrec_date	tp_date	tstar
		A	All	A	All	A	A	All	All	All	
4	CIR-0009	265	true	355	false	2/17/21	5/18/21	6/19/20	2/3/2021	2/3/2021	
8	CIR-0013	350	true	1322	false	5/27/21	1/24/24	7/8/2020	3/3/2021	3/3/2021	
9	CIR-0013	350	true	1322	false	5/27/21	1/24/24	7/8/2020	3/3/2021	5/12/2021	
13	CIR-0046	347	true	1082	false	7/9/21	7/14/23	8/21/2020	4/23/2021	4/23/2021	
14	CIR-0046	347	true	1082	false	7/9/21	7/14/23	8/21/2020	4/23/2021	7/2/2021	
18	CIR-0049	448	true	1281	false	10/13/21	1/24/24	8/24/2020	3/29/2021	3/29/2021	
19	CIR-0049	448	true	1281	false	10/13/21	1/24/24	8/24/2020	3/29/2021	6/21/2021	
23	CIR-0050	664	false	664	false	5/23/22	5/23/22	8/26/2020	4/5/2021	4/5/2021	
24	CIR-0050	664	false	664	false	5/23/22	5/23/22	8/26/2020	4/5/2021	6/28/2021	
25	CIR-0050	664	false	664	false	5/23/22	5/23/22	8/26/2020	4/5/2021	12/13/2021	

Showing 1 to 10 of 335 entries

Previous

1

2

3

4

5

...

34

Next

[Hide](#)

```
fit <- coxph(Surv(tstart4, tstop4, rfs_event) ~ biomarker_status,
  data = dt_final)
```

Warning in Surv(tstart4, tstop4, rfs_event) :
Stop time must be > start time, NA created

[Hide](#)

```
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart4, tstop4, rfs_event) ~ biomarker_status,
  data = dt_final)

n= 272, number of events= 169
(63 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE 1.1143    3.0474   0.2638 4.224  2.4e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE    3.047      0.3281    1.817    5.111

Concordance= 0.543 (se = 0.01 )
Likelihood ratio test= 21.91 on 1 df,  p=3e-06
Wald test            = 17.84 on 1 df,  p=2e-05
Score (logrank) test = 19.48 on 1 df,  p=1e-05
```

[Hide](#)

```
summary_fit <- summary(fit)
hr <- summary_fit$coef[1, "exp(coef)"]
ci_lower <- summary_fit$conf.int[1, "lower .95"]
ci_upper <- summary_fit$conf.int[1, "upper .95"]
p_value <- summary_fit$coef[1, "Pr(>|z|)"]
formatted_p_value <- ifelse(p_value < 0.0001, "<0.0001", sprintf("%.3f", p_value))
result_line <- sprintf("HR = %.2f (%.2f-%.2f); P %s", hr, ci_lower, ci_upper, formatted_p_value)
print(result_line)
```

```
[1] "HR = 3.05 (1.82-5.11); P <0.0001"
```

#OS by timing of molecular recurrence in ctDNA MRD negative - 3 groups

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE"]
circ_data <- circ_data[circ_data$ctDNA.MRD != "" & circ_data$Lead.Time >= 0, ]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE"]
circ_data <- circ_data[circ_data$PostMRDPos.Event=="TRUE"]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    PostMRDPos.months >= 0 & PostMRDPos.months < 6 ~ 1,
    PostMRDPos.months >= 6 & PostMRDPos.months < 12 ~ 2,
    PostMRDPos.months >= 12 & PostMRDPos.months < 24 ~ 3
  ))

circ_data <- circ_data[!is.na(circ_data$ctDNA.Dynamics),]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
survfit(Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)~ctDNA.Dynamics, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event) ~
  ctDNA.Dynamics, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Dynamics=1	77	11	NA	38.9	NA
ctDNA.Dynamics=2	58	5	NA	NA	NA
ctDNA.Dynamics=3	30	0	NA	NA	NA

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

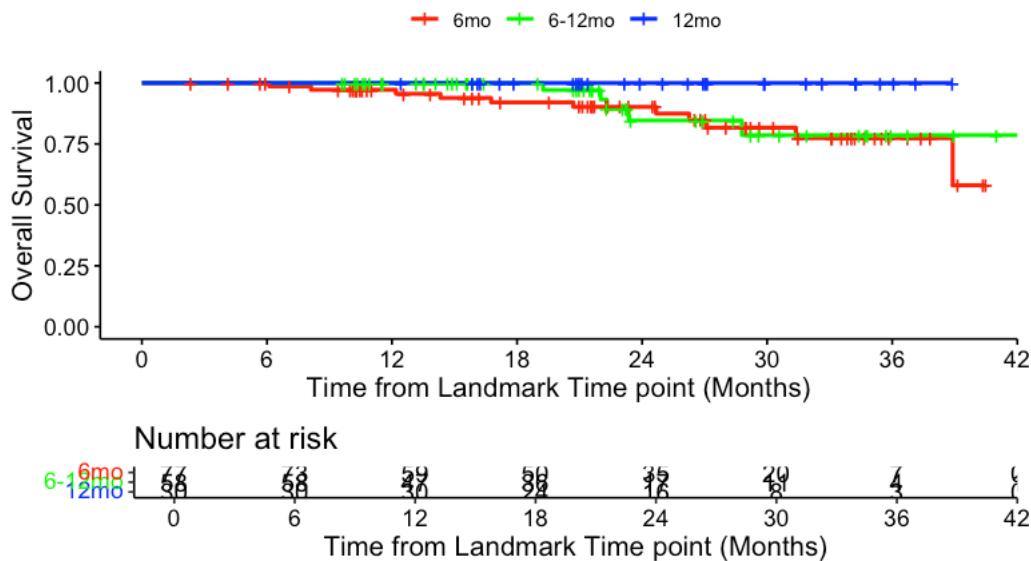
ctDNA.Dynamics	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	77	11	0.1428571	14.28571
2	58	5	0.0862069	8.62069
3	30	0	0.0000000	0.00000

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
  =c("red","green","blue"), title="OS - ctDNA MRD Neg with Molecular Recurrence", ylab= "Overall Survival", xlab
  ="Time from Landmark Time point (Months)", legend.labs=c("6mo","6-12mo", "12mo"), legend.title="")
```

OS - ctDNA MRD Neg with Molecular Recurrence



```
summary(KM_curve, times= c(12, 24, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.Dynamics=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     59      2     0.972  0.0192     0.894     0.993
 24     35      4     0.902  0.0384     0.793     0.955
 36      7      4     0.774  0.0693     0.602     0.879
```

```
ctDNA.Dynamics=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     47      0     1.000  0.0000      NA      NA
 24     17      4     0.847  0.0721     0.635     0.941
 36      4      1     0.786  0.0887     0.547     0.909
```

```
ctDNA.Dynamics=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     30      0      1      0      1      1
 24     16      0      1      0      NA      NA
 36      3      0      1      0      NA      NA
```

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("3","2","1"), labels = c(">12 months","6-12 months", "<6 months"))
cox_fit <- coxphf(surv_object ~ ctDNA.Dynamics, data = circ_data, maxstep = 0.5, maxit = 100)
summary(cox_fit)
```

```
coxphf(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  maxit = 100, maxstep = 0.5)
```

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

	coef	se(coef)	exp(coef)	lower	0.95	upper	0.95	Chisq	p
ctDNA.Dynamics6-12 months	2.025644	1.546344	7.580989	0.8572701	995.9189	3.209661	0.07320463		
ctDNA.Dynamics<6 months	2.334475	1.511804	10.324038	1.3459559	1325.4489	5.531177	0.01868054		

Likelihood ratio test=5.532375 on 2 df, p=0.06290137, n=165
Wald test = 2.538202 on 2 df, p = 0.2810842

Covariance-Matrix:

	ctDNA.Dynamics6-12 months	ctDNA.Dynamics<6 months
ctDNA.Dynamics6-12 months	2.391179	2.190536
ctDNA.Dynamics<6 months	2.190536	2.285550

```
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Dynamics, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

Hide

```
print(chi_square_test)
```

Pearson's Chi-squared test

```
data: contingency_table
X-squared = 5.1498, df = 2, p-value = 0.07616
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.05623
alternative hypothesis: two.sided
```

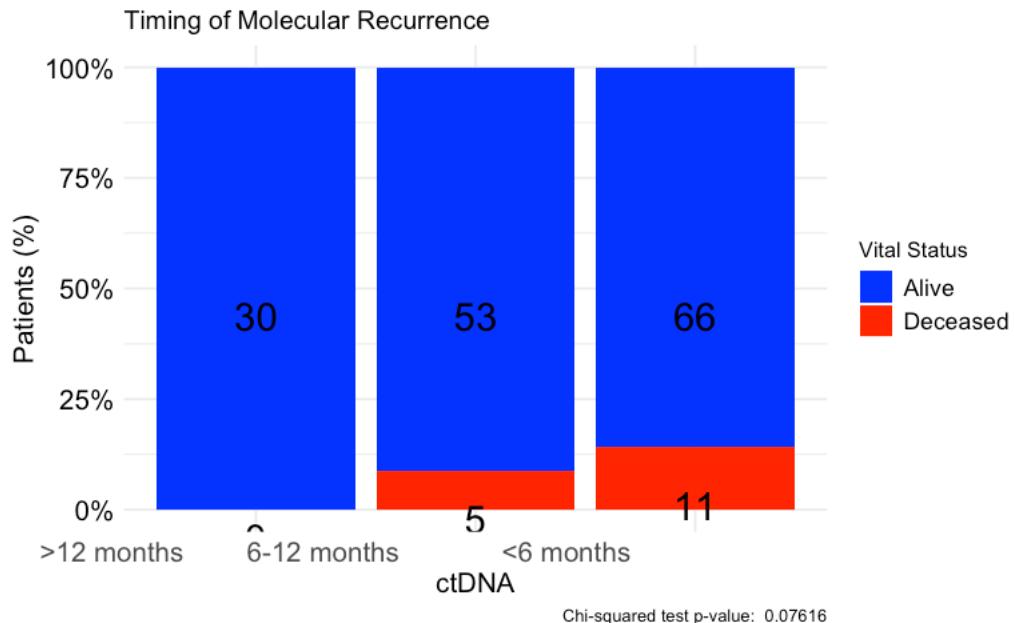
Hide

```
print(contingency_table)
```

	Alive	Deceased
>12 months	30	0
6-12 months	53	5
<6 months	66	11

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "Timing of Molecular Recurrence",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "" & circ_data$Lead.Time >= 0, ]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$PostMRDPos.Event=="TRUE",]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign value
s
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    PostMRDPos.months >= 0 & PostMRDPos.months < 6 ~ 1,
    PostMRDPos.months >= 6 & PostMRDPos.months < 12 ~ 2,
    PostMRDPos.months >= 12 & PostMRDPos.months < 24 ~ 3
  ))
  
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("2","1"), labels = c("6-12 months", "<6 mon
ths"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
summary(cox_fit)

```

Call:
 coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 135, number of events= 16
 (1729 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Dynamics<6 months	0.3594	1.4325	0.5409	0.664	0.506
ctDNA.Dynamics<6 months		exp(coef)	exp(-coef)	lower .95	upper .95
		1.432	0.6981	0.4962	4.135

Concordance= 0.569 (se = 0.061)
 Likelihood ratio test= 0.46 on 1 df, p=0.5
 Wald test = 0.44 on 1 df, p=0.5
 Score (logrank) test = 0.45 on 1 df, p=0.5

[Hide](#)

#DFS by ctDNA at the Surveillance Window - All stages Landmark 10 weeks

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Surveillance, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ctDNA.Surveillance, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	1481	89	NA	NA	NA
ctDNA.Surveillance=POSITIVE	310	261	8.47	7.09	8.74

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

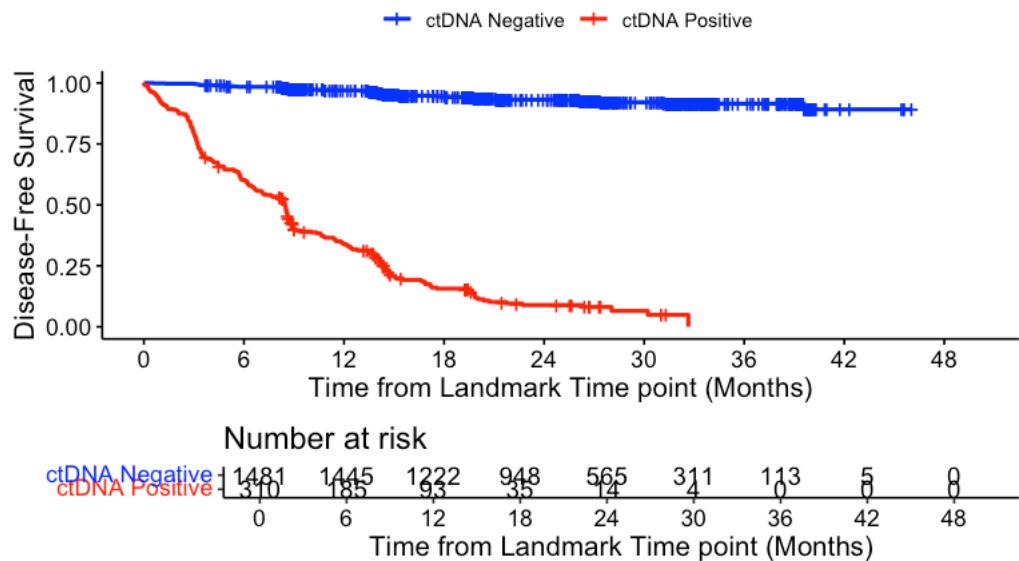
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1481	89	0.06009453	6.009453
POSITIVE	310	261	0.84193548	84.193548

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA Surveillance window | All stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA Surveillance window | All stages



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

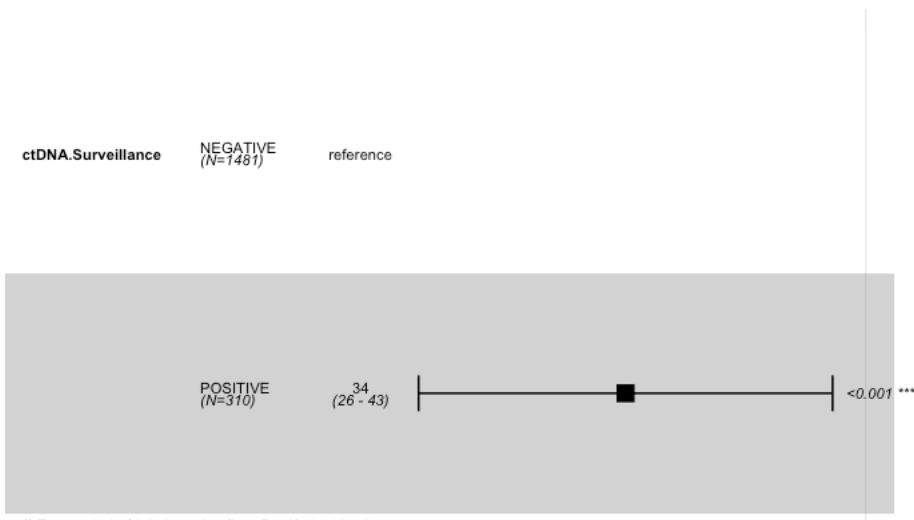
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24     565     81    0.932 0.00756    0.915    0.945
 30     311      5    0.922 0.00878    0.902    0.937
 36     113      2    0.915 0.00975    0.894    0.933
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24     14     257    0.0893 0.0197    0.0556    0.133
 30      4      2    0.0649 0.0213    0.0314    0.115
```

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



45

Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 1791, number of events= 350

            coef exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.SurveillancePOSITIVE  3.5133   33.5603   0.1289  27.26   <2e-16 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE     33.56      0.0298    26.07     43.2   

Concordance= 0.835  (se = 0.01 )
Likelihood ratio test= 875  on 1 df,  p=<2e-16
Wald test      = 743.2  on 1 df,  p=<2e-16
Score (logrank) test = 1682  on 1 df,  p=<2e-16
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.SurveillancePOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 1024)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 1.21455152569113404287108598813568845503449926364311064686637227450052747862851
76839045866292831851914122634086253819655700402714973046179380675057375543912507735674490463906893046610775756047
64646557664393992986262331141367686739919656664044919499358928083657617110519896493480065999595002164834207317118
429762e-163"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.Surveillance, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 991.63, df = 1, p-value < 2.2e-16
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 56.55705 123.45382
sample estimates:
odds ratio
82.94443
```

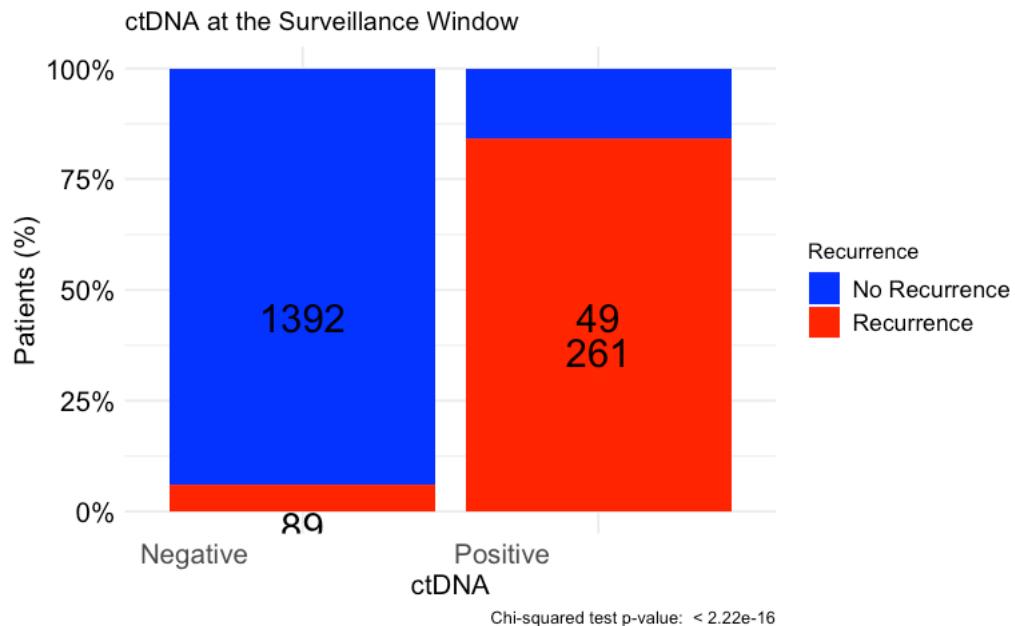
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
Negative	1392	89
Positive	49	261

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the Surveillance Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

[Hide](#)

```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value with high precision: 1.18761884963146710490280427320818799056356353568376464020764076563194185
0017620e-217"
```

#OS by ctDNA at the Surveillance Window - All stages Landmark 10 weeks

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$OS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.Surveillance, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~
ctDNA.Surveillance, data = circ_data)
```

```
           n  events median 0.95LCL 0.95UCL
ctDNA.Surveillance=NEGATIVE 1481      13     NA     NA     NA
ctDNA.Surveillance=POSITIVE  313      41    41.8    37.3     NA
```

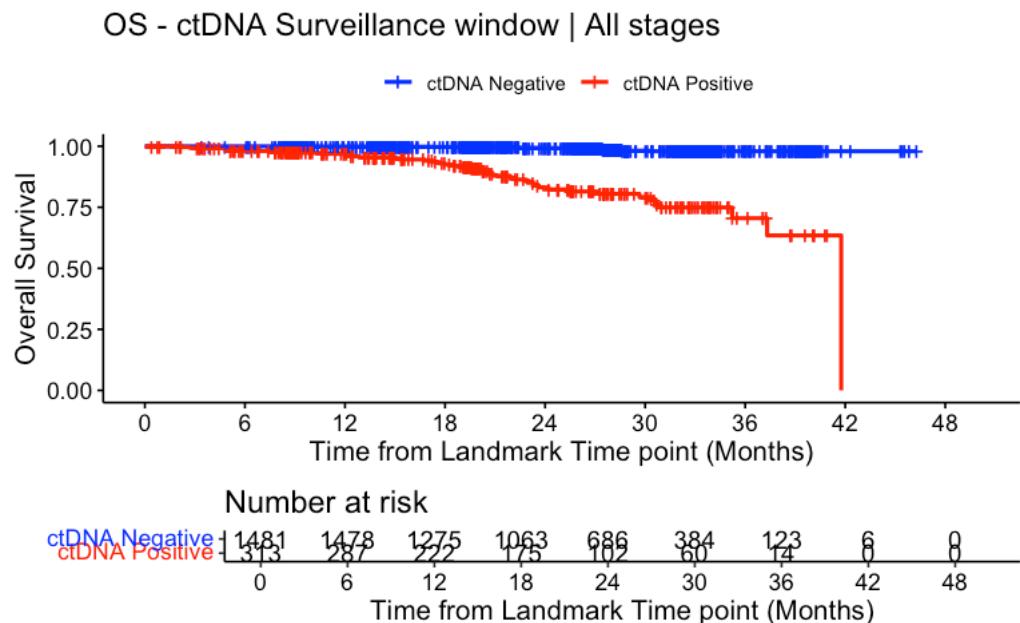
Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1481	13	0.008777853	0.8777853
POSITIVE	313	41	0.130990415	13.0990415
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="OS - ctDNA Surveillance window | All stages", ylab= "Overall Survival", xlab="Time from
Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

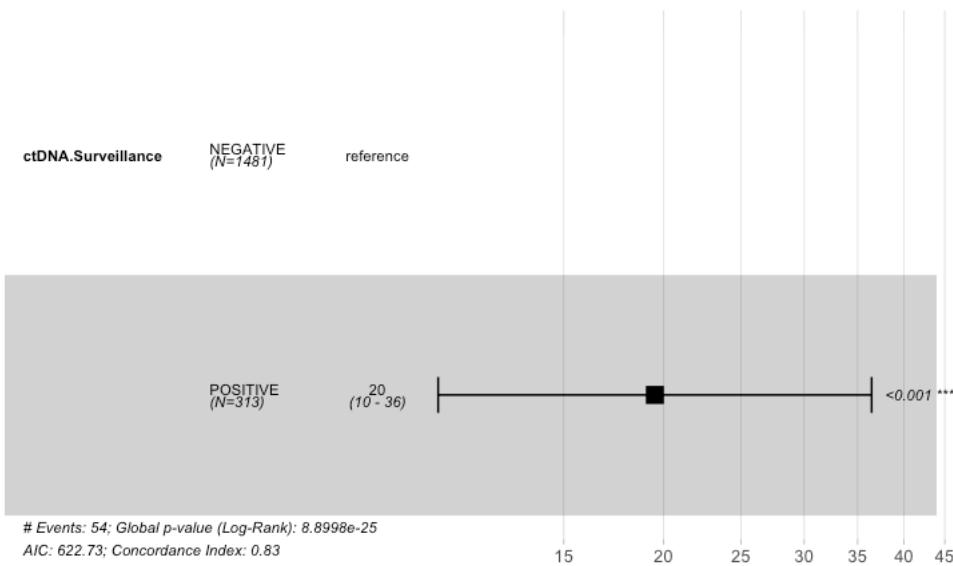
  ctDNA.Surveillance=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     686      7    0.993 0.00288      0.984    0.997
  30     384      5    0.982 0.00552      0.967    0.990
  36     123      1    0.979 0.00608      0.963    0.989

  ctDNA.Surveillance=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     102      31    0.832 0.0294      0.765    0.881
  30      60      4    0.792 0.0343      0.715    0.850
  36      14      4    0.705 0.0571      0.577    0.801
```

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 1794, number of events= 54

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 2.9708   19.5075   0.3189 9.317 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE   19.51     0.05126   10.44    36.44

Concordance= 0.825 (se = 0.028 )
Likelihood ratio test= 105.6 on 1 df,  p=<2e-16
Wald test                 = 86.8 on 1 df,  p=<2e-16
Score (logrank) test = 171.6 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.SurveillancePOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 1.19782832883743298608690836231842628092321329960317230758022587738695815928788
5e-20"
```

Hide

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 19.51 (10.44-36.44); p = 0"
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$OS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Surveillance, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data: contingency_table
X-squared = 128.04, df = 1, p-value < 2.2e-16

```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

Fisher's Exact Test for Count Data

```

data: contingency_table
p-value < 2.2e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 8.778118 35.028967
sample estimates:
odds ratio
16.97861

```

[Hide](#)

```
print(contingency_table)
```

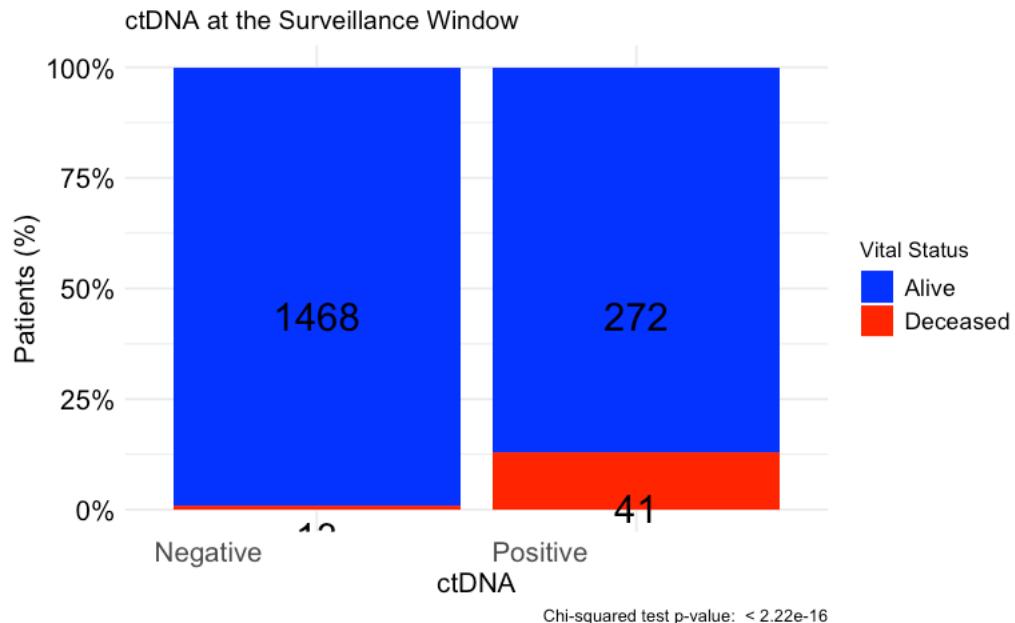
	Alive	Deceased
Negative	1468	13
Positive	272	41

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA at the Surveillance Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



```
#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))
```

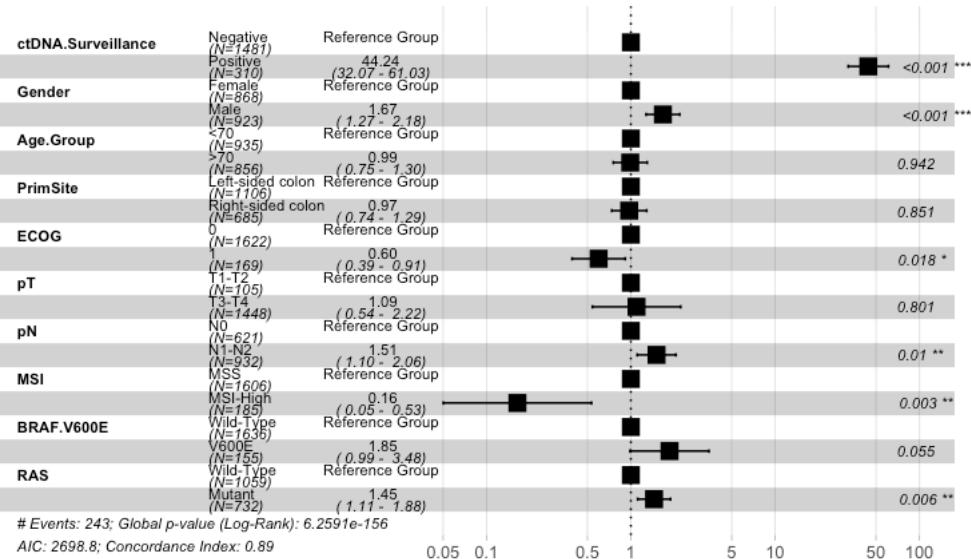
```
[1] "Exact p-value with high precision: 1.09970654621414493584316391973493547011939636884663797418605340339939357
0475089e-29"
```

#Multivariate cox regression at Surveillance Window for DFS - All stages Landmark 10 weeks

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"), labels = c
("Negative", "Positive"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", ">70"))
circ_data$PrimSite <- factor(circ_data$PrimSite, levels = c("Left-sided colon", "Right-sided colon"))
circ_data$ECOG <- factor(circ_data$ECOG, levels = c("0", "1"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"), labels = c("MSS", "MSI-High"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("Wild-Type", "V600E"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("Wild-Type", "Mutant"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance + Gender + Age.Group + PrimSite + ECOG + pT + pN + MSI + BRAF.V
600E + RAS, data=circ_data)
ggforest(cox_fit, data = circ_data, main = "Multivariate Regression Model for DFS - All Stages", refLabel = "Refe
rence Group")
```

Multivariate Regression Model for DFS - All Stages



Hide

```
test.ph <- cox.zph(cox_fit)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.Surveillance + Gender + Age.Group +
    PrimSite + ECOG + pT + pN + MSI + BRAF.V600E + RAS, data = circ_data)

n= 1552, number of events= 243
(239 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.SurveillancePositive 3.78959 44.23832 0.16417 23.083 < 2e-16 ***
GenderMale                  0.51108 1.66710 0.13686  3.734 0.000188 ***
Age.Group>70                -0.01009 0.98996 0.13840 -0.073 0.941857
PrimSiteRight-sided colon  -0.02687 0.97349 0.14265 -0.188 0.850585
ECOG1                       -0.51425 0.59795 0.21677 -2.372 0.017679 *
pTT3-T4                     0.09072 1.09497 0.35985  0.252 0.800956
pNN1-N2                     0.40998 1.50678 0.15843  2.588 0.009662 **
MSIMSI-High                 -1.81099 0.16349 0.60363 -3.000 0.002698 **
BRAF.V600EV600E             0.61736 1.85403 0.32144  1.921 0.054784 .
RASMutant                   0.37032 1.44820 0.13383  2.767 0.005654 **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePositive 44.2383 0.0226 32.06694 61.0295
GenderMale                  1.6671 0.5998 1.27486 2.1800
Age.Group>70                0.9900 1.0101 0.75476 1.2984
PrimSiteRight-sided colon  0.9735 1.0272 0.73605 1.2875
ECOG1                       0.5979 1.6724 0.39097 0.9145
pTT3-T4                     1.0950 0.9133 0.54087 2.2167
pNN1-N2                     1.5068 0.6637 1.10457 2.0555
MSIMSI-High                 0.1635 6.1165 0.05008 0.5337
BRAF.V600EV600E             1.8540 0.5394 0.98742 3.4812
RASMutant                   1.4482 0.6905 1.11408 1.8825

Concordance= 0.886 (se = 0.012 )
Likelihood ratio test= 755.9 on 10 df,  p=<2e-16
Wald test                  = 584 on 10 df,  p=<2e-16
Score (logrank) test = 1553 on 10 df,  p=<2e-16

```

Hide

```

cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.SurveillancePositive", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 512)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value for ctDNA.MRD: 6.79046242980828407778347722517594258672299437962476219045684884610635807239003
942154501178488999766155012921415351152445782542204539645762093295358348123462e-118"

```

#OS by ctDNA at the MRD Window - pts with Radiological Recurrence

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event) ~
    ctDNA.MRD, data = circ_data)

1 observation deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 219    22     NA     NA     NA
ctDNA.MRD=POSITIVE 263    52    43.4    36.8     NA

```

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

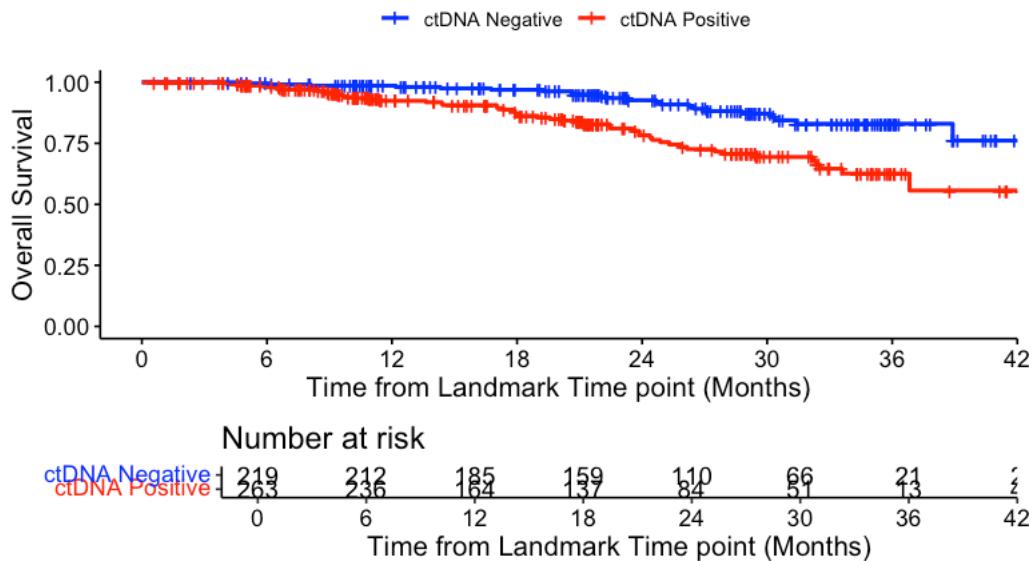
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	219	22	0.1004566	10.04566
POSITIVE	263	52	0.1977186	19.77186
NA	1	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="OS - Radiological Recurrence | ctDNA MRD window", ylab= "Overall Survival", xlab="Time
from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - Radiological Recurrence | ctDNA MRD window



Hide

```
summary(KM_curve, times= c(24, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

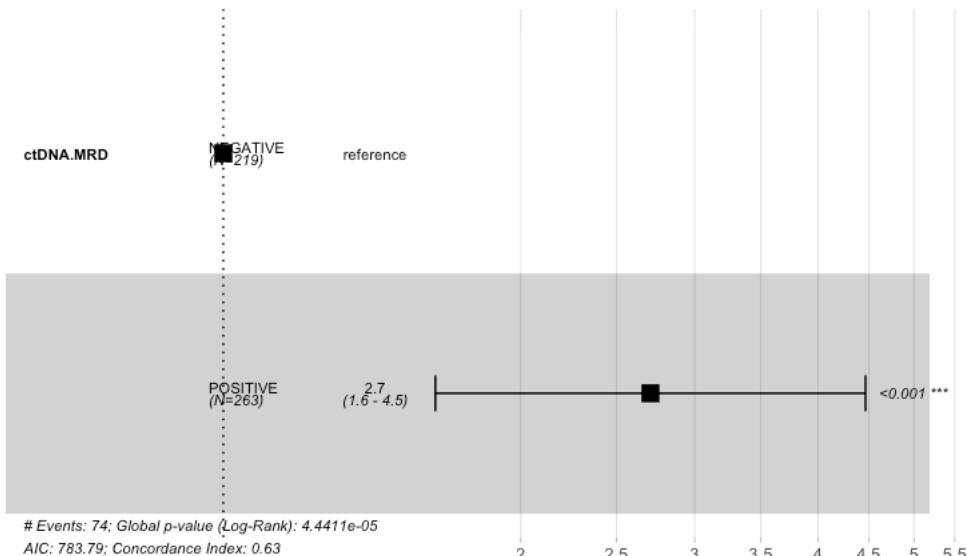
1 observation deleted due to missingness
  ctDNA.MRD=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24     110      12    0.926  0.0209      0.873    0.958
  36      21       9    0.830  0.0364      0.744    0.889

  ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  24      84      37    0.783  0.0334      0.708    0.840
  36      13      13    0.626  0.0490      0.522    0.714
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 482, number of events= 74
(1 observation deleted due to missingness)

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 0.9954    2.7059   0.2557 3.893 9.89e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    2.706     0.3696    1.639     4.466

Concordance= 0.631 (se = 0.027 )
Likelihood ratio test= 16.67 on 1 df,  p=4e-05
Wald test            = 15.16 on 1 df,  p=1e-04
Score (logrank) test = 16.43 on 1 df,  p=5e-05
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 2.71 (1.64-4.47); p = 0"
```

[Hide](#)

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 7.9661, df = 1, p-value = 0.004766
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

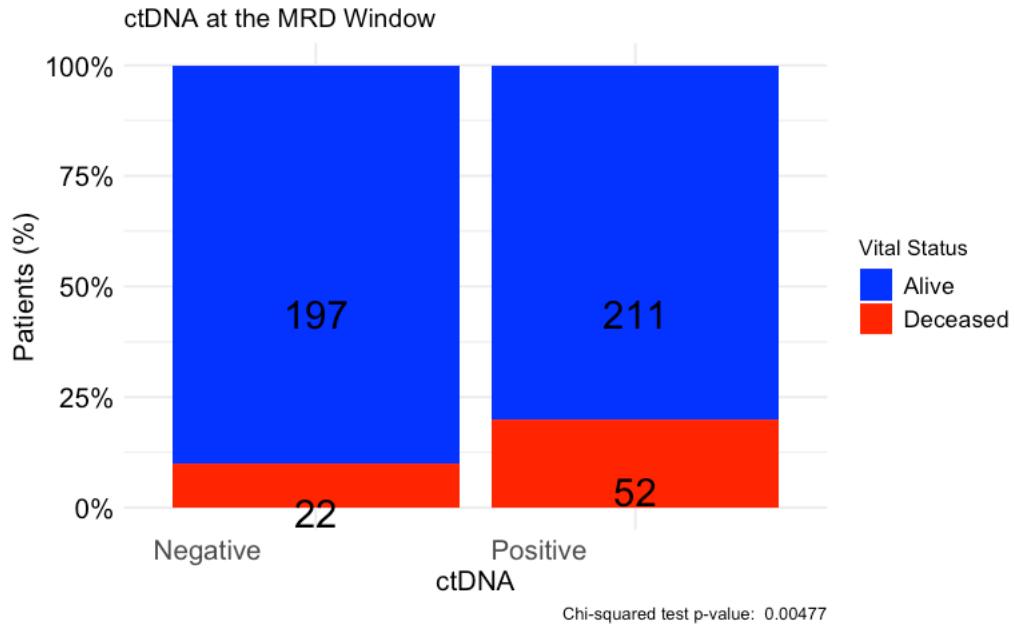
```
data: contingency_table
p-value = 0.00342
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.260726 3.958921
sample estimates:
odds ratio
2.203332
```

[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	197	22
Positive	211	52

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#OS by ctDNA at the MRD Window - pts with Radiological Recurrence Sites

```

# Define the function to analyze each recurrence site and extract HR values
analyze_site <- function(site) {
  circ_data_site <- circ_data %>% filter(grepl(site, RelSite, ignore.case = TRUE))
  circ_data_site <- circ_data_site[circ_data_site$ctDNA.MRD != "",]
  circ_data <- circ_data[circ_data$OS.MRD.months >= 0,]

  surv_object <- Surv(time = circ_data_site$OS.MRD.months, event = circ_data_site$OS.Event)
  cox_fit <- coxph(surv_object ~ ctDNA.MRD, data = circ_data_site)
  cox_fit_summary <- summary(cox_fit)

  HR <- cox_fit_summary$coefficients[2]
  lower_CI <- cox_fit_summary$conf.int[3]
  upper_CI <- cox_fit_summary$conf.int[4]
  p_value <- cox_fit_summary$coefficients[5]

  label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", format.pval(p_value, digits = 3))
  return(list(HR = HR, lower_CI = lower_CI, upper_CI = upper_CI, p_value = p_value, site = site, label_text = label_text))
}

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
circ_data <- circ_data[circ_data$RFS.Event == "TRUE",]
recurrence_sites <- c("liver", "lung", "peritoneum", "lymph node")
results <- lapply(recurrence_sites, analyze_site)
forest_data <- do.call(rbind, lapply(results, function(res) {
  data.frame(
    site = res$site,
    HR = res$HR,
    lower_CI = res$lower_CI,
    upper_CI = res$upper_CI,
    label_text = res$label_text
  )
}))

forest_data$site <- factor(forest_data$site, levels = c("liver", "lung", "peritoneum", "lymph node"))
forest_plot <- ggplot(forest_data, aes(x = site, y = HR, ymin = lower_CI, ymax = upper_CI)) +
  geom_pointrange() +
  geom_text(aes(label = label_text), hjust = -0.1, vjust = -0.5) +
  geom_hline(yintercept = 1, linetype = "dashed") +
  coord_flip() +
  scale_y_continuous(breaks = seq(1, max(forest_data$upper_CI) + 1, by = 2), expand = c(0, 0), limits = c(0, max(forest_data$upper_CI) + 1)) +
  labs(x = "Recurrence Site", y = "HR for OS between ctDNA MRD positive vs negative") +
  theme_minimal()

# Define the function to analyze each recurrence site and extract HR values
analyze_site <- function(site) {
  circ_data_site <- circ_data %>% filter(grepl(site, RelSite, ignore.case = TRUE))
  circ_data_site <- circ_data_site[circ_data_site$ctDNA.MRD != "",]
  circ_data <- circ_data[circ_data$OS.MRD.months >= 0,]

  site_count <- nrow(circ_data_site)
  print(paste("Number of observations for site", site, ":", site_count))

  surv_object <- Surv(time = circ_data_site$OS.months, event = circ_data_site$OS.Event)
  cox_fit <- coxph(surv_object ~ ctDNA.MRD, data = circ_data_site)
  cox_fit_summary <- summary(cox_fit)

  HR <- cox_fit_summary$coefficients[2]
  lower_CI <- cox_fit_summary$conf.int[3]
  upper_CI <- cox_fit_summary$conf.int[4]
  p_value <- cox_fit_summary$coefficients[5]

  label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", format.pval(p_value, digits = 3))
  return(list(HR = HR, lower_CI = lower_CI, upper_CI = upper_CI, p_value = p_value, site = site, label_text = label_text))
}

# Set working directory and load data
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
circ_data <- circ_data[circ_data$RFS.Event == "TRUE",]
circ_data <- circ_data[circ_data$OS.MRD.months >= 0,]

```

```
# Recurrence sites to analyze
recurrence_sites <- c("liver", "lung", "peritoneum", "lymph node")

# Perform analysis for each site
results <- lapply(recurrence_sites, analyze_site)

[1] "Number of observations for site liver : 209"
[1] "Number of observations for site lung : 191"
[1] "Number of observations for site peritoneum : 115"
[1] "Number of observations for site lymph node : 66"
```

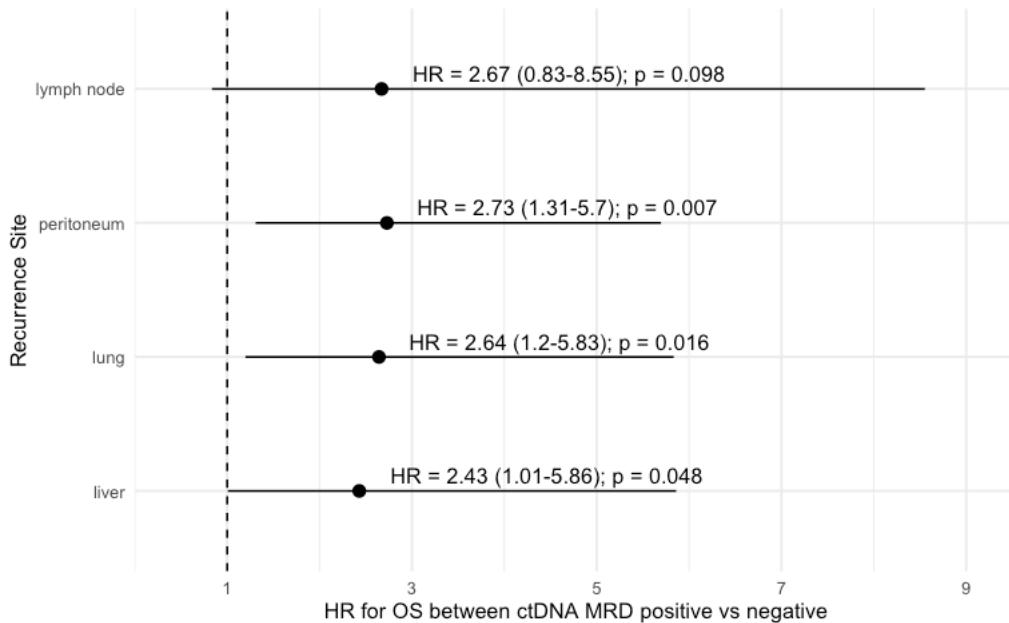
Hide

```
# Create data frame for forest plot
forest_data <- do.call(rbind, lapply(results, function(res) {
  data.frame(
    site = res$site,
    HR = res$HR,
    lower_CI = res$lower_CI,
    upper_CI = res$upper_CI,
    label_text = res$label_text
  )
}))

# Set the order of the levels for the 'site' factor
forest_data$site <- factor(forest_data$site, levels = c("liver", "lung", "peritoneum", "lymph node"))

# Create forest plot
forest_plot <- ggplot(forest_data, aes(x = site, y = HR, ymin = lower_CI, ymax = upper_CI)) +
  geom_pointrange() +
  geom_text(aes(label = label_text), hjust = -0.1, vjust = -0.5) +
  geom_hline(yintercept = 1, linetype = "dashed") +
  coord_flip() +
  scale_y_continuous(breaks = seq(1, max(forest_data$upper_CI) + 1, by = 2), expand = c(0, 0), limits = c(0, max(forest_data$upper_CI) + 1)) +
  labs(x = "Recurrence Site", y = "HR for OS between ctDNA MRD positive vs negative") +
  theme_minimal()

print(forest_plot)
```



Hide

```
for (res in results) {
  print(res$label_text)
}
```

```
[1] "HR = 2.43 (1.01-5.86); p = 0.048"
[1] "HR = 2.64 (1.2-5.83); p = 0.016"
[1] "HR = 2.73 (1.31-5.7); p = 0.007"
[1] "HR = 2.67 (0.83-8.55); p = 0.098"
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$OS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.Surveillance, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$OS.months, event = circ_data\$OS.Event) ~
ctDNA.Surveillance, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	78	2	NA	NA	NA
ctDNA.Surveillance=POSITIVE	264	41	41.8	37.3	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

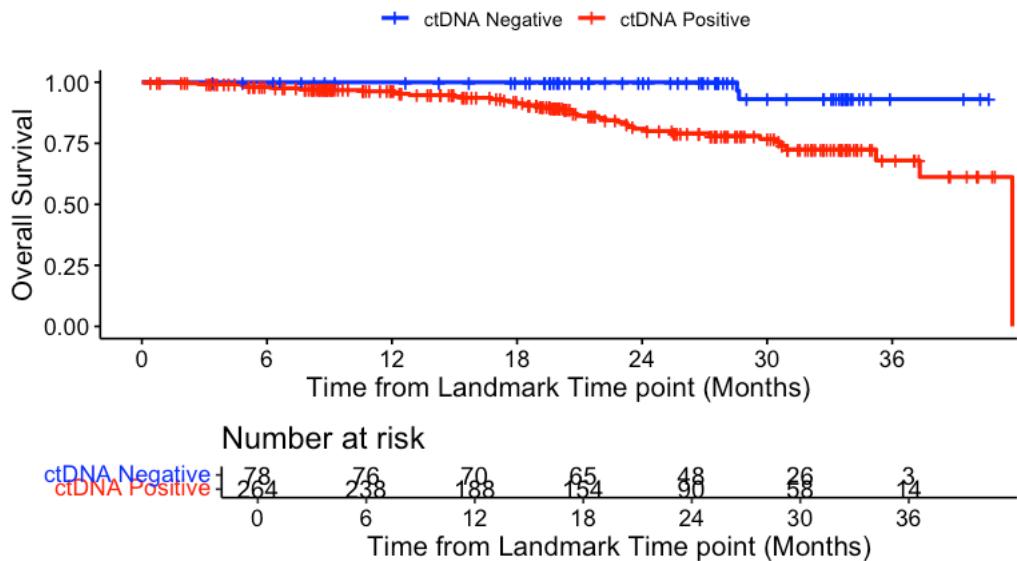
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	78	2	0.02564103	2.564103
POSITIVE	264	41	0.15530303	15.530303

2 rows

Hide

```
surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue","red"), title="OS - Radiological Recurrence | ctDNA Surveillance window", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - Radiological Recurrence | ctDNA Surveillance window



Hide

```
summary(KM_curve, times= c(24, 36))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

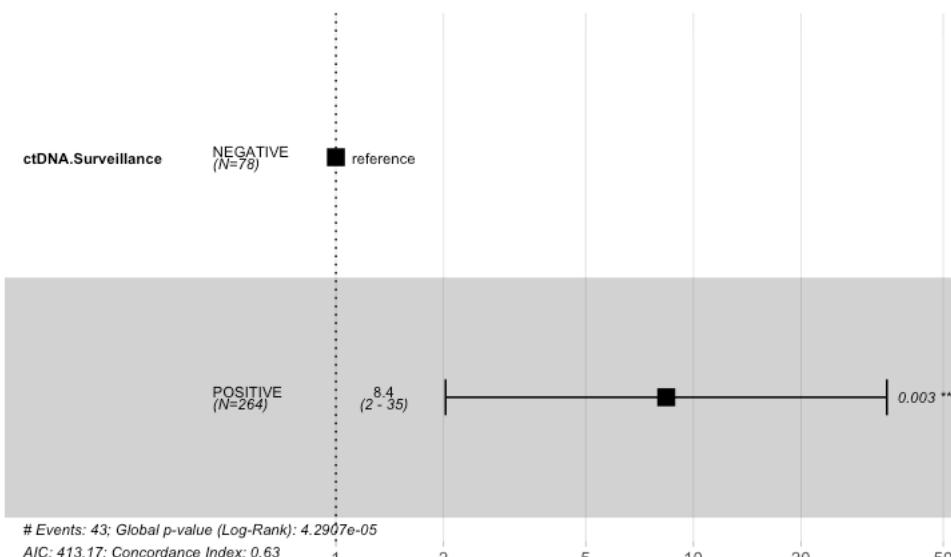
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24     48      0    1.000  0.0000      NA      NA
 36      3      2    0.931  0.0471      0.751    0.982
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24     90     31    0.809  0.0325      0.736    0.864
 36     14      8    0.680  0.0592      0.548    0.780
```

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 342, number of events= 43

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 2.1278    8.3962  0.7252 2.934  0.00334 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    8.396      0.1191    2.027    34.78

Concordance= 0.631 (se = 0.015 )
Likelihood ratio test= 16.74 on 1 df,  p=4e-05
Wald test      = 8.61 on 1 df,  p=0.003
Score (logrank) test = 12.36 on 1 df,  p=4e-04
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 8.4 (2.03-34.78); p = 0.003"
```

[Hide](#)

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Surveillance, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 8.0672, df = 1, p-value = 0.004507
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

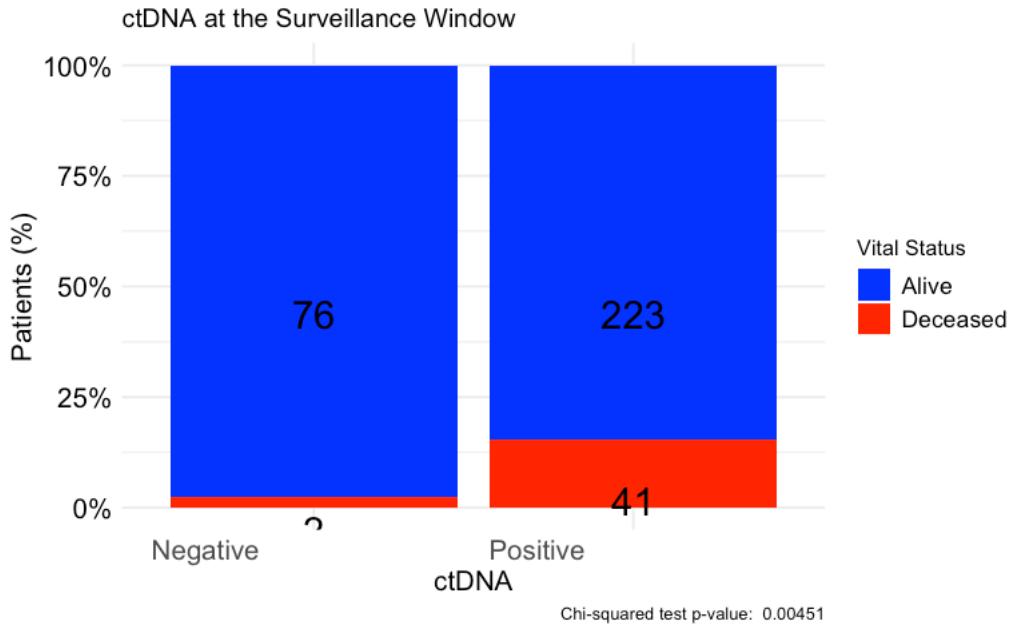
```
data: contingency_table
p-value = 0.001475
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.735104 60.847334
sample estimates:
odds ratio
6.962255
```

[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	76	2
Positive	223	41

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the Surveillance Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#Sankey plot for MRD and Surveillance dynamics

```
##To run this commands, please visit: https://sankeymatic.com/build/
#ctDNA at the MRD Window [336] ctDNA + MRD window #E67272
#ctDNA at the MRD Window [1773] ctDNA - MRD window #87EA86
#ctDNA at the MRD Window [131] Not available ctDNA MRD window #808080
#ctDNA + MRD window [141] ctDNA + Surveillance window #E67272
#ctDNA + MRD window [70] ctDNA - Surveillance window #87EA86
#ctDNA + MRD window [125] Not available ctDNA Surveillance window #808080
#ctDNA - MRD window [159] ctDNA + Surveillance window #E67272
#ctDNA - MRD window [1294] ctDNA - Surveillance window #87EA86
#ctDNA - MRD window [320] Not available ctDNA Surveillance window #808080
#Not available ctDNA MRD window [13] ctDNA + Surveillance window #E67272
#Not available ctDNA MRD window [117] ctDNA - Surveillance window #87EA86
#Not available ctDNA MRD window [1] Not available ctDNA Surveillance window #808080
#ctDNA + Surveillance window [264] Radiological Recurrence #E67272
#ctDNA + Surveillance window [49] No Recurrence #87EA86
#ctDNA - Surveillance window [78] Radiological Recurrence #E67272
#ctDNA - Surveillance window [1403] No Recurrence #87EA86
#Not available ctDNA Surveillance window [158] Radiological Recurrence #E67272
#Not available ctDNA Surveillance window [288] No Recurrence #87EA86
```

#Percentage of ctDNA MRD Window positivity in pts undergoing post-recurrence curative surgery

Hide

```

rm(list = ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data %>%
  filter(Eligible == "TRUE" & RFS.Event == "TRUE" & ctDNA.MRD != "")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
positive_rate <- sum(circ_data$ctDNA.MRD == "Positive" & circ_data$PostRecurrenceSurgery == "TRUE") / sum(circ_data$ctDNA.MRD == "Positive") * 100
positive_ci <- binconf(sum(circ_data$ctDNA.MRD == "Positive" & circ_data$PostRecurrenceSurgery == "TRUE"),
                       sum(circ_data$ctDNA.MRD == "Positive"),
                       alpha = 0.05)[c(2, 3)] * 100
negative_rate <- sum(circ_data$ctDNA.MRD == "Negative" & circ_data$PostRecurrenceSurgery == "TRUE") / sum(circ_data$ctDNA.MRD == "Negative") * 100
negative_ci <- binconf(sum(circ_data$ctDNA.MRD == "Negative" & circ_data$PostRecurrenceSurgery == "TRUE"),
                       sum(circ_data$ctDNA.MRD == "Negative"),
                       alpha = 0.05)[c(2, 3)] * 100
data <- data.frame(
  ctDNA.MRD = c("Positive", "Negative"),
  percentage = c(positive_rate, negative_rate),
  lower_ci = c(positive_ci[1], negative_ci[1]),
  upper_ci = c(positive_ci[2], negative_ci[2])
)
cross_tab <- table(circ_data$ctDNA.MRD, circ_data$PostRecurrenceSurgery)
chi_test <- chisq.test(cross_tab)
p_value <- format.pval(chi_test$p.value, digits = 3)
print(data)

```

ctDNA.MRD	percentage	lower_ci	upper_ci
<chr>	<dbl>	<dbl>	<dbl>
Positive	29.92424	24.72053	35.70382
Negative	41.09589	34.78772	47.71105
2 rows			

Hide

print(cross_tab)

	FALSE	TRUE
Negative	129	90
Positive	185	79

Hide

print(chi_test)

Pearson's Chi-squared test with Yates' continuity correction

```

data: cross_tab
X-squared = 6.0858, df = 1, p-value = 0.01363

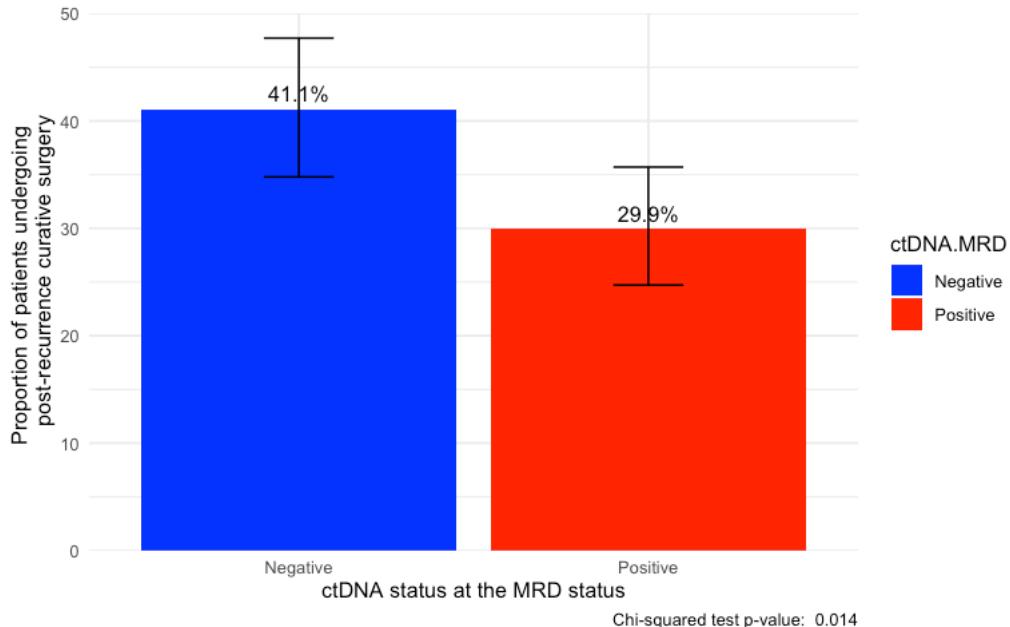
```

Hide

```

barplot <- ggplot(data, aes(x = ctDNA.MRD, y = percentage, fill = ctDNA.MRD)) +
  geom_bar(stat = "identity") +
  geom_errorbar(aes(ymin = lower_ci, ymax = upper_ci), width = 0.2) +
  geom_text(aes(label = paste0(round(percentage, 1), "%")), vjust = -0.5) +
  labs(
    x = "ctDNA status at the MRD status",
    y = "Proportion of patients undergoing
    post-recurrence curative surgery",
    caption = paste("Chi-squared test p-value: ", p_value)
  ) +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 50)) +
  scale_fill_manual(values = c("Negative" = "blue", "Positive" = "red")) +
  theme_minimal()
print(barplot)

```



#PRS by ctDNA at the MRD Window - pts with Radiological Recurrence

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]

survfit(Surv(time = circ_data$PRS.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$PRS.months, event = circ_data\$OS.Event) ~ ctDNA.MRD, data = circ_data)

18 observations deleted due to missingness

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	219	22	NA	36.3	NA
ctDNA.MRD=POSITIVE	263	52	38.2	29.2	NA

Hide

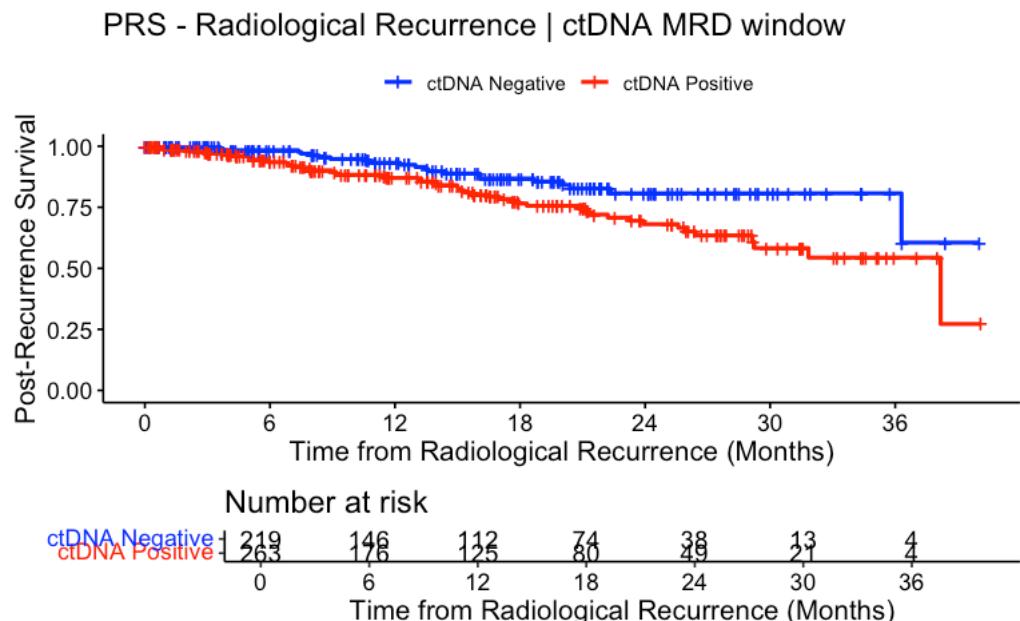
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	219	22	0.1004566	10.04566
POSITIVE	263	52	0.1977186	19.77186
NA	18	NA	NA	NA

3 rows

Hide

```
surv_object <- Surv(time = circ_data$PRS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="PRS - Radiological Recurrence | ctDNA MRD window", ylab = "Post-Recurrence Survival", xlab = "Time from Radiological Recurrence (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title = "")
```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

18 observations deleted due to missingness
  ctDNA.MRD=NEGATIVE
  time      n.risk      n.event      survival      std.err  lower 95% CI upper 95% CI
 24.0000    38.0000    21.0000      0.8073      0.0412    0.7105    0.8745

  ctDNA.MRD=POSITIVE
  time      n.risk      n.event      survival      std.err  lower 95% CI upper 95% CI
 24.0000    49.0000    45.0000      0.6809      0.0435    0.5872    0.7577
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 482, number of events= 74
(18 observations deleted due to missingness)

      coef exp(coef)  se(coef)    z Pr(>|z|)
ctDNA.MRDPOSITIVE 0.6772    1.9683  0.2546 2.66  0.00782 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    1.968     0.5081    1.195    3.242

Concordance= 0.579 (se = 0.03 )
Likelihood ratio test= 7.63 on 1 df,  p=0.006
Wald test            = 7.08 on 1 df,  p=0.008
Score (logrank) test = 7.35 on 1 df,  p=0.007

```

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 1.97 (1.2-3.24); p = 0.008"
```

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```

data: contingency_table
X-squared = 7.9661, df = 1, p-value = 0.004766

```

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

```
Fisher's Exact Test for Count Data
```

```

data: contingency_table
p-value = 0.00342
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.260726 3.958921
sample estimates:
odds ratio
2.203332

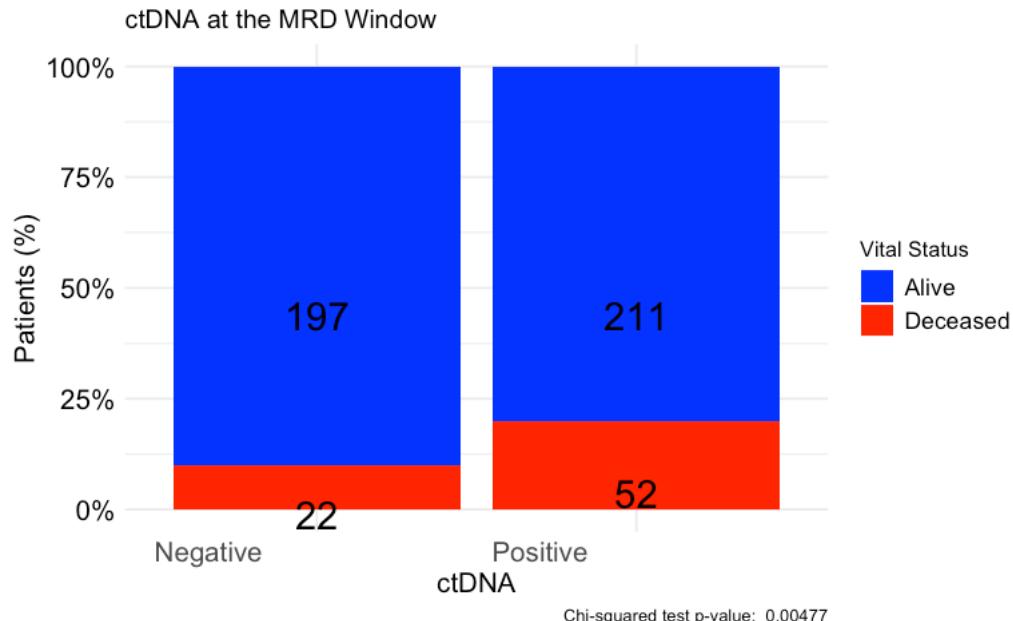
```

```
print(contingency_table)
```

	Alive	Deceased
Negative	197	22
Positive	211	52

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#PRS by ctDNA at the Surveillance Window - pts with Radiological Recurrence

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]

survfit(Surv(time = circ_data$PRS.months, event = circ_data$OS.Event)~ctDNA.Surveillance, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$PRS.months, event = circ_data$OS.Event) ~
  ctDNA.Surveillance, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	78	2	NA	NA	NA
ctDNA.Surveillance=POSITIVE	264	41	38.2	36.3	NA

Hide

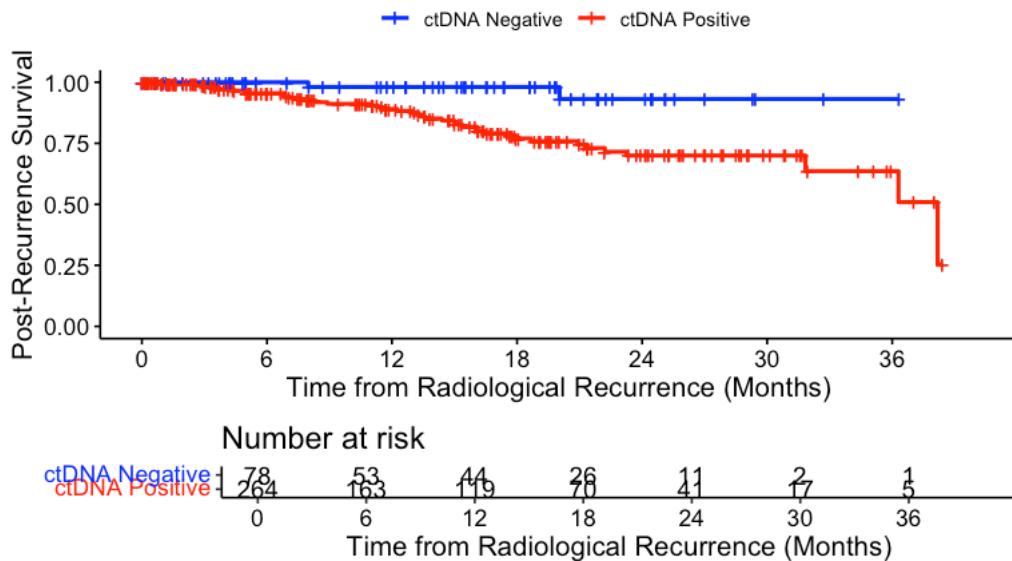
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	78	2	0.02564103	2.564103
POSITIVE	264	41	0.15530303	15.530303
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PRS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="PRS - Radiological Recurrence | ctDNA Surveillance window", ylab= "Post-Recurrence Survival", xlab="Time from Radiological Recurrence (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

PRS - Radiological Recurrence | ctDNA Surveillance window



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

ctDNA.Surveillance=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	11.0000	2.0000	0.9317	0.0511	0.7237	0.9847

ctDNA.Surveillance=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00	41.000	38.000	0.700	0.045	0.602	0.778

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)
```

```
n= 342, number of events= 43
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.SurveillancePOSITIVE	1.8831	6.5739	0.7248	2.598	0.00938 **

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.SurveillancePOSITIVE	6.574	0.1521	1.588	27.21

```
Concordance= 0.606 (se = 0.02 )
```

```
Likelihood ratio test= 12.21 on 1 df, p=5e-04
```

```
Wald test = 6.75 on 1 df, p=0.009
```

```
Score (logrank) test = 8.99 on 1 df, p=0.003
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
```

```
HR <- cox_fit_summary$coefficients[2]
```

```
lower_CI <- cox_fit_summary$conf.int[3]
```

```
upper_CI <- cox_fit_summary$conf.int[4]
```

```
p_value <- cox_fit_summary$coefficients[5]
```

```
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
```

```
print(label_text)
```

```
[1] "HR = 6.57 (1.59-27.21); p = 0.009"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]

circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$OS.Event <- factor(circ_data$OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased"))
contingency_table <- table(circ_data$ctDNA.Surveillance, circ_data$OS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 8.0672, df = 1, p-value = 0.004507
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.001475
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.735104 60.847334
sample estimates:
odds ratio
 6.962255
```

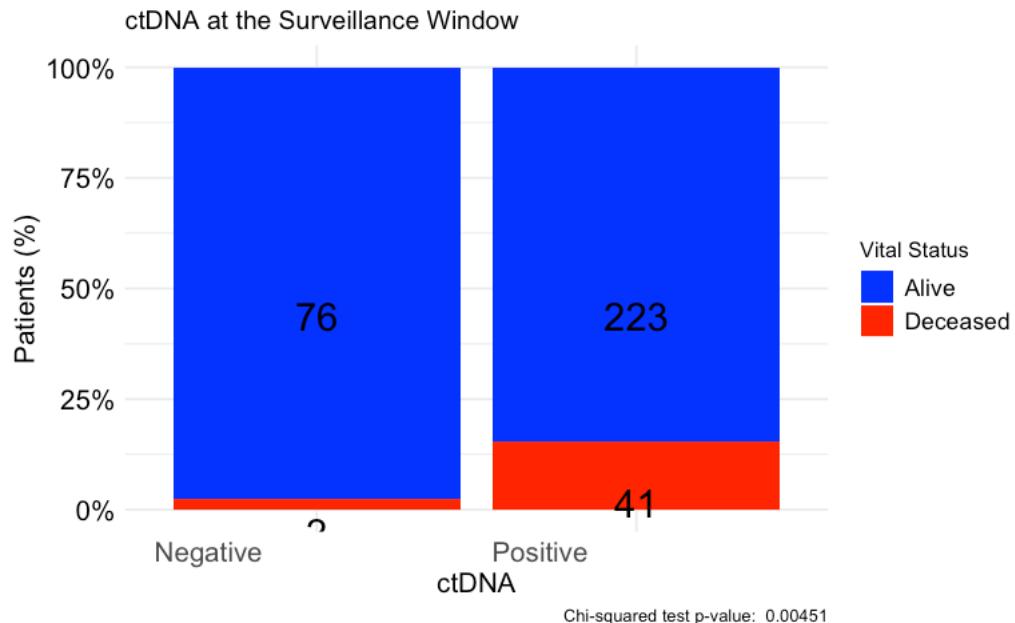
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	76	2
Positive	223	41

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the Surveillance Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Vital Status",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Alive" = "blue", "Deceased" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#Detection ctDNA rates based on sites of relapse

Hide

```
# Remove existing objects and set the working directory
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]

# Create a table of counts for the "Rec.Site" variable
relsite_counts <- table(circ_data$Rec.Site)
relsite_df <- as.data.frame(relsite_counts)
names(relsite_df) <- c("RelSite", "Count")
circ_data_pos_mrd <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data_pos_anytime <- circ_data[circ_data$ctDNA.anytime=="POSITIVE",]
pos_counts_mrd <- table(circ_data_pos_mrd$Rec.Site)
pos_counts_anytime <- table(circ_data_pos_anytime$Rec.Site)
relsite_df$MRDPos_Count <- ifelse(is.na(match(relsite_df$RelSite, names(pos_counts_mrd))), 0, pos_counts_mrd[match(h(relsite_df$RelSite, names(pos_counts_mrd)))]
relsite_df$MRDPos_Count[is.na(relsite_df$MRDPos_Count)] <- 0
relsite_df$AnytimePos_Count <- ifelse(is.na(match(relsite_df$RelSite, names(pos_counts_anytime))), 0, pos_counts_anytime[match(relsite_df$RelSite, names(pos_counts_anytime))])
relsite_df$AnytimePos_Count[is.na(relsite_df$AnytimePos_Count)] <- 0
relsite_df$Percent <- (relsite_df$Count / sum(relsite_df$Count)) * 100
relsite_df$MRDPos_Percent <- (relsite_df$MRDPos_Count / relsite_df$Count) * 100
relsite_df$AnytimePos_Percent <- (relsite_df$AnytimePos_Count / relsite_df$Count) * 100
total_observations <- sum(relsite_df$Count)
total_pos_mrd <- sum(relsite_df$MRDPos_Count)
total_pos_anytime <- sum(relsite_df$AnytimePos_Count)
total_row <- data.frame(RelSite = "Total", Count = total_observations, MRDPos_Count = total_pos_mrd, AnytimePos_Count = total_pos_anytime, Percent = 100, MRDPos_Percent = (total_pos_mrd / total_observations) * 100, AnytimePos_Percent = (total_pos_anytime / total_observations) * 100)
relsite_df <- rbind(relsite_df, total_row)
print(relsite_df)
```

RelSite	Co... <fctr>	MRDPos_Count <int>	AnytimePos_Count <dbl>	Percent <dbl>	MRDPos_Percent <dbl>	AnytimePos_Percent <dbl>
	230	118	186	46.0	51.30435	80.86957
Brain	2	2	2	0.4	100.00000	100.00000
Liver	107	77	102	21.4	71.96262	95.32710
Local/LN	25	17	21	5.0	68.00000	84.00000
Lung	73	18	36	14.6	24.65753	49.31507
Others	3	3	3	0.6	100.00000	100.00000
Peritoneum	24	9	19	4.8	37.50000	79.16667
Peritoneum & Others	36	20	32	7.2	55.55556	88.88889

RelSite <fctr>	Co... <int>	MRDPos_Count <dbl>	AnytimePos_Count <dbl>	Percent <dbl>	MRDPos_Percent <dbl>	AnytimePos_Percent <dbl>
Total	500	264	401	100.0	52.80000	80.20000
9 rows						

Hide

```
ft <- flextable(relsite_df)
doc <- read_docx() %>%
  body_add_flextable(value = ft)
print(doc, target = "relsite_df.docx")
```

#Heatmap for Biomarkers factors

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data %>% arrange(RAS.BRAF)
circ_data$RAS <- factor(circ_data$RAS.BRAF, levels = c("TRUE", "FALSE"))
circ_datadf <- as.data.frame(circ_data)

ha <- HeatmapAnnotation(
  RAS.BRAF = circ_data$RAS.BRAF,
  TMB = circ_data$TMB,
  MSI = circ_data$MSI,
  BRAF.V600E = circ_data$BRAF.V600E,
  KRAS.G12C = circ_data$KRAS.G12C,
  ERBB2 = circ_data$ERBB2,
  TP53.Y220C = circ_data$TP53.Y220C,
  NTRK = circ_data$NTRK,
  RET = circ_data$RET,

  col = list(RAS.BRAF = c("TRUE" = "blue", "FALSE" = "grey"),
  TMB = c("TMB-High" = "blue", "TMB-Low" = "grey"),
  MSI = c("MSI-High" = "blue", "MSS" = "grey"),
  BRAF.V600E = c("MUT" = "blue", "WT" = "grey"),
  KRAS.G12C = c("MUT" = "blue", "WT" = "grey"),
  ERBB2 = c("MUT" = "blue", "WT" = "grey"),
  TP53.Y220C = c("MUT" = "blue", "WT" = "grey"),
  NTRK = c("MUT" = "blue", "WT" = "grey"),
  RET = c("MUT" = "blue", "WT" = "grey")))
ht <- Heatmap(matrix(nrow = 0, ncol = length(circ_data$RAS.BRAF)), show_row_names = FALSE, cluster_rows = F, cluster_columns = FALSE, top_annotation = ha)
pdf("heatmap.pdf", width = 7, height = 7)
draw(ht, annotation_legend_side = "bottom")
dev.off()
```

#Calculate the % altered variables

Hide

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
conditions <- list(
  RAS.BRAF = "TRUE",
  TMB = "TMB-High",
  MSI = "MSI-High",
  BRAF.V600E = "MUT",
  KRAS.G12C = "MUT",
  ERBB2 = "MUT",
  TP53.Y220C = "MUT",
  NTRK = "MUT",
  RET = "MUT"
)
total_observations <- nrow(circ_data)
condition_counts <- list()
for (var in names(conditions)) {
  condition_value <- conditions[[var]]
  condition_count <- sum(circ_data[[var]] == condition_value, na.rm = TRUE)
  condition_percentage <- (condition_count / total_observations) * 100
  condition_counts[[var]] <- list('Count' = condition_count, 'Percentage' = condition_percentage)
}
condition_counts_df <- do.call(rbind, lapply(names(condition_counts), function(x) {
  data.frame(Variable = x,
             Count = condition_counts[[x]]$Count,
             Percentage = condition_counts[[x]]$Percentage)
)))
print(condition_counts_df)

```

Variable	Count	Percentage
<chr>	<int>	<dbl>
RAS.BRAF	1125	50.22321429
TMB	230	10.26785714
MSI	215	9.59821429
BRAF.V600E	178	7.94642857
KRAS.G12C	49	2.18750000
ERBB2	36	1.60714286
TP53.Y220C	24	1.07142857
NTRK	2	0.08928571
RET	1	0.04464286
9 rows		

#DFS by Biomarkers

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data %>% filter(Eligible == "TRUE")
circ_data <- circ_data %>%
  mutate(
    RAS.BRAF = ifelse(RAS.BRAF == "TRUE", "RAS/BRAF WT", NA),
    TMB = ifelse(TMB == "TMB-High", "TMB High", NA),
    MSI = ifelse(MSI == "MSI-High", "MSI High", NA),
    BRAF.V600E = ifelse(BRAF.V600E == "MUT", "BRAF V600E", NA),
    KRAS.G12C = ifelse(KRAS.G12C == "MUT", "KRAS G12C", NA),
    ERBB2 = ifelse(ERBB2 == "MUT", "ERBB2", NA),
    TP53.Y220C = ifelse(TP53.Y220C == "MUT", "TP53 Y220C", NA)
  )
circ_data_long <- circ_data %>%
  gather(key = "group", value = "value", RAS.BRAF, TMB, MSI, BRAF.V600E, KRAS.G12C, ERBB2, TP53.Y220C) %>%
  filter(!is.na(value))
circ_data_long$value <- factor(circ_data_long$value, levels = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"))

survfit(Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)~value, data = circ_data_long)

```

```
Call: survfit(formula = Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event) ~
  value, data = circ_data_long)
```

	n	events	median	0.95LCL	0.95UCL
value=RAS/BRAF WT	1125	233	NA	NA	NA
value=TMB High	230	10	NA	NA	NA
value=MSI High	215	8	NA	NA	NA
value=BRAF V600E	178	25	NA	NA	NA
value=KRAS G12C	49	19	33.7	22.1	NA
value=ERBB2	36	12	NA	23.2	NA
value=TP53 Y220C	24	6	NA	NA	NA

```
event_summary <- circ_data_long %>%
  group_by(value) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

value	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
RAS/BRAF WT	1125	233	0.20711111	20.711111
TMB High	230	10	0.04347826	4.347826
MSI High	215	8	0.03720930	3.720930
BRAF V600E	178	25	0.14044944	14.044944
KRAS G12C	49	19	0.38775510	38.775510
ERBB2	36	12	0.33333333	33.333333
TP53 Y220C	24	6	0.25000000	25.000000

7 rows

```
surv_obj <- Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)
cox_model <- coxph(surv_obj ~ value, data = circ_data_long)
summary(cox_model)
```

```
Call:
coxph(formula = surv_obj ~ value, data = circ_data_long)
```

n= 1857, number of events= 313

	coef	exp(coef)	se(coef)	z	Pr(> z)
valueTMB High	-1.6745	0.1874	0.3230	-5.184	2.17e-07 ***
valueMSI High	-1.8298	0.1605	0.3596	-5.088	3.62e-07 ***
valueBRAF V600E	-0.4366	0.6462	0.2105	-2.074	0.03806 *
valueKRAS G12C	0.7798	2.1810	0.2387	3.267	0.00109 **
valueERBB2	0.5571	1.7456	0.2961	1.882	0.05987 .
valueTP53 Y220C	0.2368	1.2671	0.4135	0.573	0.56693

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
valueTMB High	0.1874	5.3362	0.09950	0.3529
valueMSI High	0.1605	6.2324	0.07929	0.3247
valueBRAF V600E	0.6462	1.5474	0.42779	0.9762
valueKRAS G12C	2.1810	0.4585	1.36608	3.4821
valueERBB2	1.7456	0.5729	0.97711	3.1185
valueTP53 Y220C	1.2671	0.7892	0.56344	2.8497

Concordance= 0.635 (se = 0.012)

Likelihood ratio test= 107 on 6 df, p=<2e-16

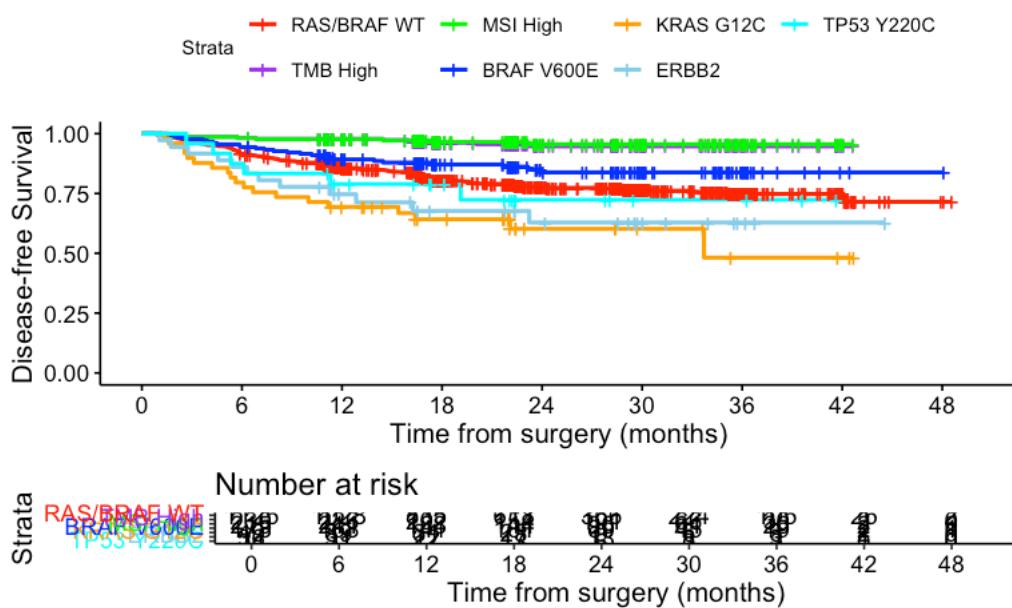
Wald test = 73.9 on 6 df, p=6e-14

Score (logrank) test = 93.74 on 6 df, p=<2e-16

```

KM_curve <- survfit(surv_obj ~ value, data = circ_data_long)
ggsurvplot(
  KM_curve,
  data = circ_data_long,
  risk.table = TRUE,
  pval = FALSE,
  conf.int = FALSE,
  break.time.by = 6,
  xlab = "Time from surgery (months)",
  ylab = "Disease-free Survival",
  legend.labs = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"),
  palette = c("red", "purple", "green", "blue", "orange", "skyblue", "cyan")
)

```



```
summary(KM_curve, times = c(24))
```

Hide

```
Call: survfit(formula = surv_obj ~ value, data = circ_data_long)
```

value=RAS/BRAF WT								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	366.0000	224.0000	0.7755	0.0137	0.7491	0.8028		
value=TMB High								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	101.0000	10.0000	0.9471	0.0169	0.9146	0.9807		
value=MSI High								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	96.0000	8.0000	0.9558	0.0159	0.9252	0.9874		
value=BRAF V600E								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	69.0000	25.0000	0.8382	0.0311	0.7793	0.9015		
value=KRAS G12C								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	9.0000	18.0000	0.6023	0.0765	0.4696	0.7726		
value=ERBB2								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	13.0000	12.0000	0.6287	0.0887	0.4769	0.8289		
value=TP53 Y220C								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000	6.0000	6.0000	0.7237	0.0993	0.5530	0.9470		

Hide

```
circ_data_long$DFS.Event <- factor(circ_data_long$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data_long$value, circ_data_long$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

Hide

```
print(chi_square_test)
```

Pearson's Chi-squared test

```
data: contingency_table
X-squared = 89.99, df = 6, p-value < 2.2e-16
```

Hide

```
print(contingency_table)
```

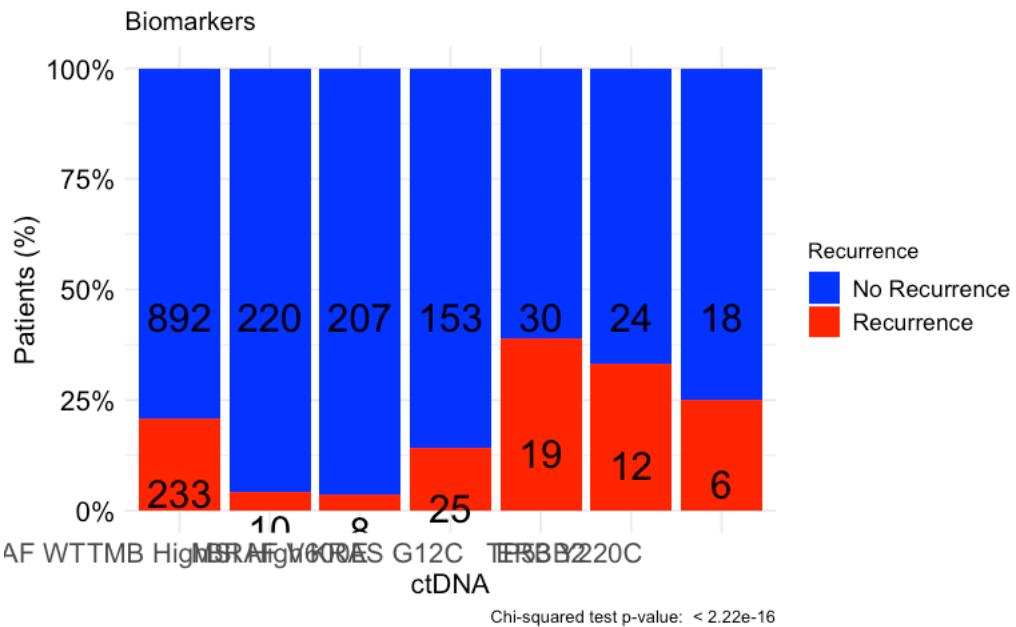
	No Recurrence	Recurrence
RAS/BRAF WT	892	233
TMB High	220	10
MSI High	207	8
BRAF V600E	153	25
KRAS G12C	30	19
ERBB2	24	12
TP53 Y220C	18	6

Hide

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "Biomarkers",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) # define custom colors
theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
      axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
      axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
      axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
      legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



```

#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value with high precision: 3.04483818727436627873963430781181117538223803364524910897159770684083923
6974716e-17"

```

#DFS by Biomarkers - Stage I-III

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data %>% filter(Eligible == "TRUE")
circ_data <- circ_data[!(circ_data$Stage %in% c("IV")),]
circ_data <- circ_data %>%
  mutate(
    RAS.BRAF = ifelse(RAS.BRAF == "TRUE", "RAS/BRAF WT", NA),
    TMB = ifelse(TMB == "TMB-High", "TMB High", NA),
    MSI = ifelse(MSI == "MSI-High", "MSI High", NA),
    BRAF.V600E = ifelse(BRAF.V600E == "MUT", "BRAF V600E", NA),
    KRAS.G12C = ifelse(KRAS.G12C == "MUT", "KRAS G12C", NA),
    ERBB2 = ifelse(ERBB2 == "MUT", "ERBB2", NA),
    TP53.Y220C = ifelse(TP53.Y220C == "MUT", "TP53 Y220C", NA)
  )
circ_data_long <- circ_data %>%
  gather(key = "group", value = "value", RAS.BRAF, TMB, MSI, BRAF.V600E, KRAS.G12C, ERBB2, TP53.Y220C) %>%
  filter(!is.na(value))
circ_data_long$value <- factor(circ_data_long$value, levels = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"))

survfit(Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)~value, data = circ_data_long)

```

Call: survfit(formula = Surv(time = circ_data_long\$DFS.months, event = circ_data_long\$DFS.Event) ~ value, data = circ_data_long)

	n	events	median	0.95LCL	0.95UCL
value=RAS/BRAF WT	919	121	NA	NA	NA
value=TMB High	226	9	NA	NA	NA
value=MSI High	211	7	NA	NA	NA
value=BRAF V600E	167	19	NA	NA	NA
value=KRAS G12C	34	8	NA	33.7	NA
value=ERBB2	24	6	NA	NA	NA
value=TP53 Y220C	19	3	NA	NA	NA

Hide

```

event_summary <- circ_data_long %>%
  group_by(value) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

value	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
RAS/BRAF WT	919	121	0.13166485	13.166485
TMB High	226	9	0.03982301	3.982301
MSI High	211	7	0.03317536	3.317536
BRAF V600E	167	19	0.11377246	11.377246
KRAS G12C	34	8	0.23529412	23.529412
ERBB2	24	6	0.25000000	25.000000
TP53 Y220C	19	3	0.15789474	15.789474

7 rows

Hide

```

surv_obj <- Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)
cox_model <- coxph(surv_obj ~ value, data = circ_data_long)
summary(cox_model)

```

```

Call:
coxph(formula = surv_obj ~ value, data = circ_data_long)

n= 1600, number of events= 173

      coef exp(coef) se(coef)      z Pr(>|z|)    
valueTMB High -1.2824    0.2774   0.3456 -3.711 0.000207 ***
valueMSI High -1.4652    0.2310   0.3888 -3.768 0.000164 ***
valueBRAF V600E -0.1646    0.8482   0.2468 -0.667 0.504805    
valueKRAS G12C  0.6159    1.8514   0.3655  1.685 0.091909 .  
valueERBB2      0.6998    2.0133   0.4183  1.673 0.094374 .  
valueTP53 Y220C 0.2591    1.2958   0.5845  0.443 0.657578    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
valueTMB High    0.2774    3.6052   0.1409   0.5461    
valueMSI High    0.2310    4.3284   0.1078   0.4950    
valueBRAF V600E   0.8482    1.1789   0.5229   1.3759    
valueKRAS G12C   1.8514    0.5401   0.9045   3.7894    
valueERBB2       2.0133    0.4967   0.8868   4.5707    
valueTP53 Y220C   1.2958    0.7717   0.4121   4.0746    

Concordance= 0.615 (se = 0.017 )
Likelihood ratio test= 46.29 on 6 df,  p=3e-08
Wald test      = 35.17 on 6 df,  p=4e-06
Score (logrank) test = 41.44 on 6 df,  p=2e-07

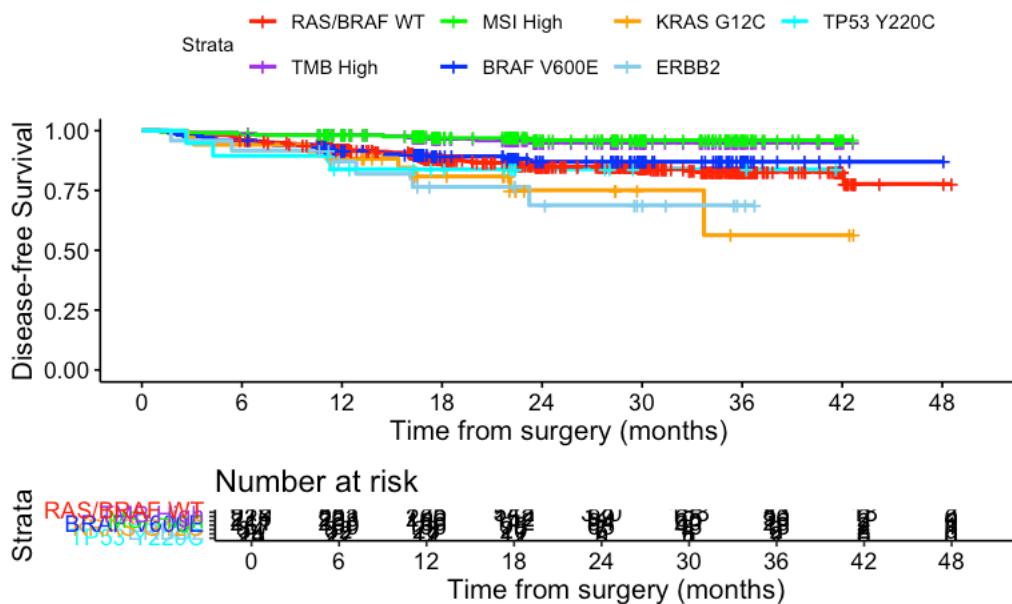
```

Hide

```

KM_curve <- survfit(surv_obj ~ value, data = circ_data_long)
ggsurvplot(
  KM_curve,
  data = circ_data_long,
  risk.table = TRUE,
  pval = FALSE,
  conf.int = FALSE,
  break.time.by = 6,
  xlab = "Time from surgery (months)",
  ylab = "Disease-free Survival",
  legend.labs = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"),
  palette = c("red", "purple", "green", "blue", "orange", "skyblue", "cyan")
)

```



Hide

```
summary(KM_curve, times = c(24))
```

```
Call: survfit(formula = surv_obj ~ value, data = circ_data_long)
```

	value=RAS/BRAF WT	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		320.0000	115.0000		0.8519	0.0132	0.8264	0.8783		
	value=TMB High	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		99.0000	9.0000		0.9505	0.0166	0.9185	0.9837		
	value=MSI High	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		94.0000	7.0000		0.9597	0.0155	0.9298	0.9906		
	value=BRAF V600E	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		65.0000	19.0000		0.8700	0.0289	0.8150	0.9286		
	value=KRAS G12C	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		7.0000	7.0000		0.7511	0.0863	0.5996	0.9408		
	value=ERBB2	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		9.000	6.000		0.688	0.111	0.502	0.944		
	value=TP53 Y220C	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		5.0000	3.0000		0.8388	0.0854	0.6871	1.0000		

Hide

```
circ_data_long$DFS.Event <- factor(circ_data_long$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data_long$value, circ_data_long$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning in stats::chisq.test(x, y, ...):
  Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

```
Pearson's Chi-squared test
```

```
data: contingency_table
X-squared = 39.76, df = 6, p-value = 5.079e-07
```

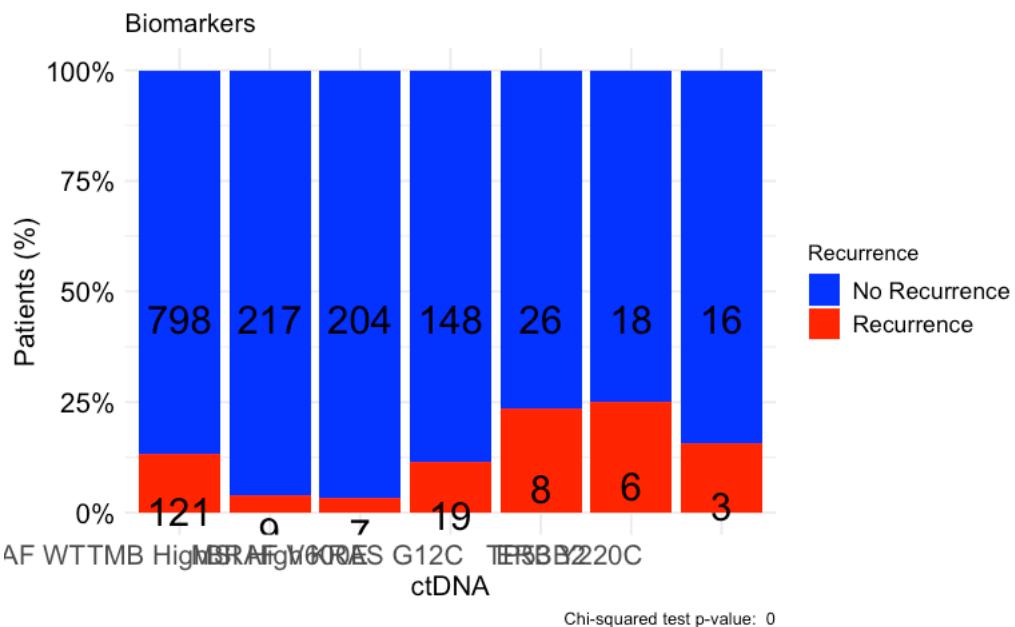
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
RAS/BRAF WT	798	121
TMB High	217	9
MSI High	204	7
BRAF V600E	148	19
KRAS G12C	26	8
ERBB2	18	6
TP53 Y220C	16	3

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "Biomarkers",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#DFS by Biomarkers - Stage IV

[Hide](#)

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data %>% filter(Eligible == "TRUE")
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II", "III")),]
circ_data <- circ_data %>%
  mutate(
    RAS.BRAF = ifelse(RAS.BRAF == "TRUE", "RAS/BRAF WT", NA),
    TMB = ifelse(TMB == "TMB-High", "TMB High", NA),
    MSI = ifelse(MSI == "MSI-High", "MSI High", NA),
    BRAF.V600E = ifelse(BRAF.V600E == "MUT", "BRAF V600E", NA),
    KRAS.G12C = ifelse(KRAS.G12C == "MUT", "KRAS G12C", NA),
    ERBB2 = ifelse(ERBB2 == "MUT", "ERBB2", NA),
    TP53.Y220C = ifelse(TP53.Y220C == "MUT", "TP53 Y220C", NA)
  )
circ_data_long <- circ_data %>%
  gather(key = "group", value = "value", RAS.BRAF, TMB, MSI, BRAF.V600E, KRAS.G12C, ERBB2, TP53.Y220C) %>%
  filter(!is.na(value))
circ_data_long$value <- factor(circ_data_long$value, levels = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"))

survfit(Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)~value, data = circ_data_long)

```

```
Call: survfit(formula = Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event) ~ value, data = circ_data_long)
```

	n	events	median	0.95LCL	0.95UCL
value=RAS/BRAF WT	206	112	16.95	14.49	35.1
value=TMB High	4	1	NA	1.38	NA
value=MSI High	4	1	NA	1.38	NA
value=BRAF V600E	11	6	23.95	8.71	NA
value=KRAS G12C	15	11	5.62	4.21	NA
value=ERBB2	12	6	8.28	6.31	NA
value=TP53 Y220C	5	3	19.12	6.11	NA

Hide

```

event_summary <- circ_data_long %>%
  group_by(value) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

value	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
RAS/BRAF WT	206	112	0.5436893	54.36893
TMB High	4	1	0.2500000	25.00000
MSI High	4	1	0.2500000	25.00000
BRAF V600E	11	6	0.5454545	54.54545
KRAS G12C	15	11	0.7333333	73.33333
ERBB2	12	6	0.5000000	50.00000
TP53 Y220C	5	3	0.6000000	60.00000

7 rows

Hide

```

surv_obj <- Surv(time = circ_data_long$DFS.months, event = circ_data_long$DFS.Event)
cox_model <- coxph(surv_obj ~ value, data = circ_data_long)
summary(cox_model)

```

```

Call:
coxph(formula = surv_obj ~ value, data = circ_data_long)

n= 257, number of events= 140

      coef exp(coef)  se(coef)      z Pr(>|z|)    
valueTMB High -0.81504  0.44262  1.00462 -0.811  0.4172    
valueMSI High -0.81504  0.44262  1.00462 -0.811  0.4172    
valueBRAF V600E -0.10339  0.90177  0.41918 -0.247  0.8052    
valueKRAS G12C  0.67999  1.97385  0.31697  2.145  0.0319 *  
valueERBB2      -0.05252  0.94883  0.41916 -0.125  0.9003    
valueTP53 Y220C  0.01708  1.01723  0.58519  0.029  0.9767    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
valueTMB High    0.4426    2.2593  0.06179  3.171    
valueMSI High    0.4426    2.2593  0.06179  3.171    
valueBRAF V600E   0.9018    1.1089  0.39654  2.051    
valueKRAS G12C   1.9739    0.5066  1.06049  3.674    
valueERBB2       0.9488    1.0539  0.41725  2.158    
valueTP53 Y220C   1.0172    0.9831  0.32308  3.203    

Concordance= 0.53  (se = 0.018 )
Likelihood ratio test= 5.91 on 6 df,  p=0.4
Wald test          = 6.28 on 6 df,  p=0.4
Score (logrank) test = 6.6 on 6 df,  p=0.4

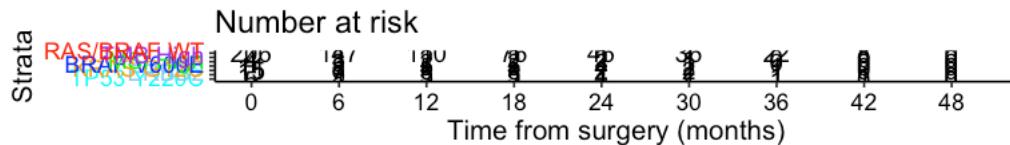
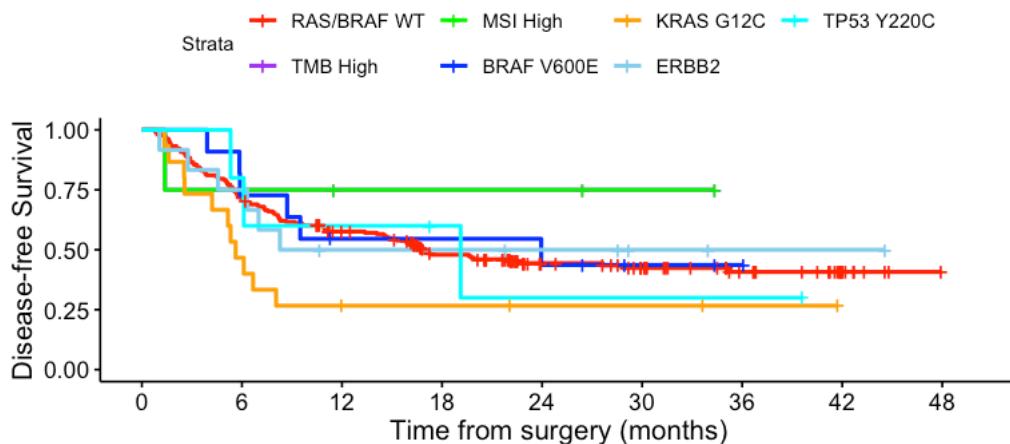
```

Hide

```

KM_curve <- survfit(surv_obj ~ value, data = circ_data_long)
ggsurvplot(
  KM_curve,
  data = circ_data_long,
  risk.table = TRUE,
  pval = FALSE,
  conf.int = FALSE,
  break.time.by = 6,
  xlab = "Time from surgery (months)",
  ylab = "Disease-free Survival",
  legend.labs = c("RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"),
  palette = c("red", "purple", "green", "blue", "orange", "skyblue", "cyan")
)

```



Hide

```
summary(KM_curve, times = c(24))
```

```
Call: survfit(formula = surv_obj ~ value, data = circ_data_long)
```

	value=RAS/BRAF WT	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		24.0000	46.0000	109.0000	0.4441	0.0366	0.3779	0.5220		
<hr/>										
	value=TMB High	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.00		24.00	2.000	1.000	0.750	0.217	0.426	1.000		
	value=MSI High	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.00		24.00	2.000	1.000	0.750	0.217	0.426	1.000		
	value=BRAF V600E	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.00		24.00	4.000	6.000	0.436	0.155	0.218	0.874		
	value=KRAS G12C	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.00		24.00	2.000	11.000	0.267	0.114	0.115	0.617		
	value=ERBB2	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.00		24.00	4.000	6.000	0.500	0.144	0.284	0.880		
	value=TP53 Y220C	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24.0000		24.0000	1.0000	3.0000	0.3000	0.2387	0.0631	1.0000		

Hide

```
circ_data_long$DFS.Event <- factor(circ_data_long$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data_long$value, circ_data_long$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning in stats::chisq.test(x, y, ...):
  Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

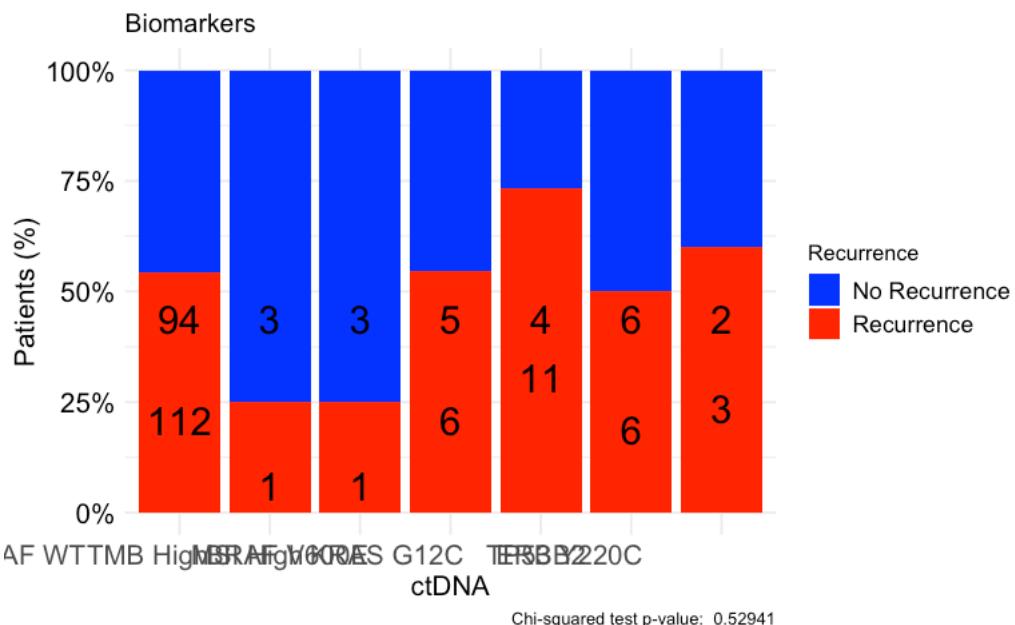
Pearson's Chi-squared test

```
data: contingency_table
X-squared = 5.113, df = 6, p-value = 0.5294
```

```
print(contingency_table)
```

	No Recurrence	Recurrence
RAS/BRAF WT	94	112
TMB High	3	1
MSI High	3	1
BRAF V600E	5	6
KRAS G12C	4	11
ERBB2	6	6
TP53 Y220C	2	3

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "Biomarkers",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



```
#DFS by MSI status - All stages
```

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_datadf <- as.data.frame(circ_data)
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"), labels = c("MSS", "MSI-High"))

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~MSI, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
MSI, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
MSI=MSS	2025	506	NA	NA	NA
MSI=MSI-High	215	8	NA	NA	NA

[Hide](#)

```
event_summary <- circ_data %>%
  group_by(MSI) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

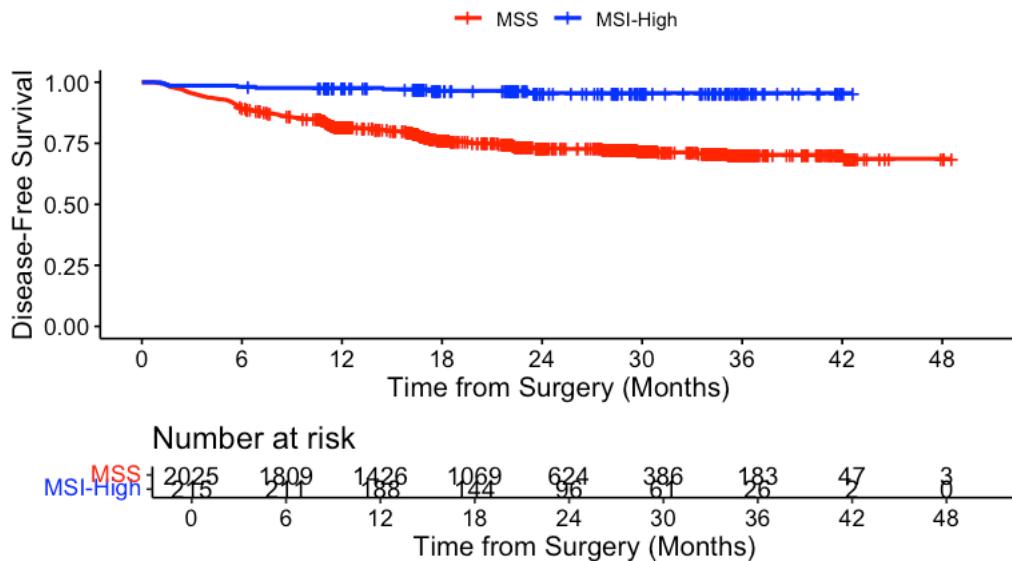
MSI	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
MSS	2025	506	0.2498765	24.98765
MSI-High	215	8	0.0372093	3.72093

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ MSI, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("red", "blue"), title="DFS - MSI Status | All stages", ylab= "Disease-Free Survival", xlab="Time from Surgery (Months)", legend.labs=c("MSS", "MSI-High"), legend.title="")
```

DFS - MSI Status | All stages



```
summary(KM_curve, times= c(24, 30, 36))
```

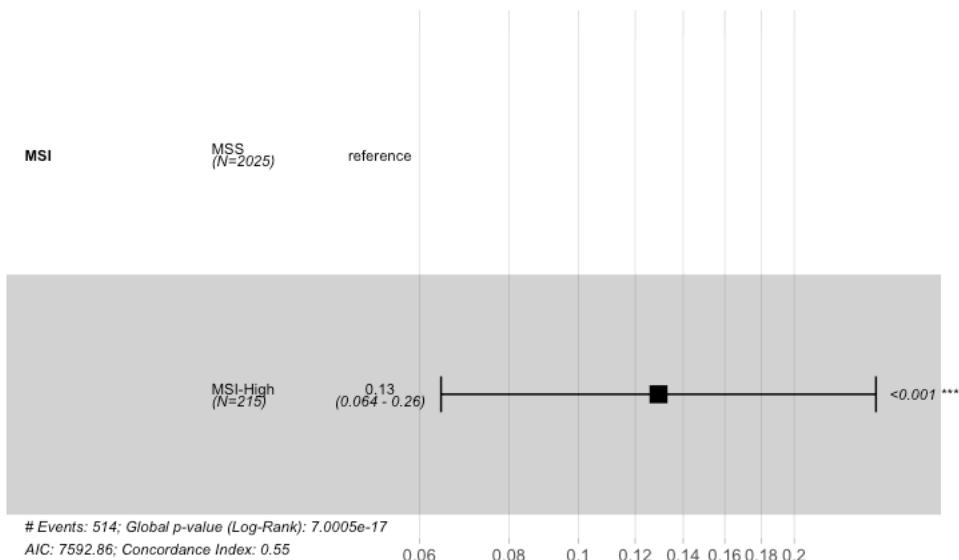
```
Call: survfit(formula = surv_object ~ MSI, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

```
MSI=MSS
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24     624     490    0.730  0.0108     0.709    0.751
 30     386      9    0.717  0.0114     0.694    0.739
 36     183      6    0.703  0.0126     0.678    0.727
```

```
MSI=MSI-High
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24      96      8    0.956  0.0159     0.911    0.978
 30      61      0    0.956  0.0159     0.911    0.978
 36      26      0    0.956  0.0159     0.911    0.978
```

```
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"), labels = c("MSS", "MSI-High"))
cox_fit <- coxph(surv_object ~ MSI, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ MSI, data = circ_data)

n= 2240, number of events= 514

      coef exp(coef)  se(coef)      z Pr(>|z|)    
MSIMSI-High -2.0456    0.1293   0.3564 -5.74 9.47e-09 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
MSIMSI-High    0.1293     7.734   0.06431     0.26    

Concordance= 0.546  (se = 0.004 )
Likelihood ratio test= 69.67 on 1 df,  p=<2e-16
Wald test           = 32.95 on 1 df,  p=9e-09
Score (logrank) test = 46.16 on 1 df,  p=1e-11
```

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 0.13 (0.06-0.26); p = 0"
```

Hide

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$MSI, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 48.522, df = 1, p-value = 3.266e-12
```

Hide

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

```
Fisher's Exact Test for Count Data
```

```
data: contingency_table
p-value = 5.283e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.04912506 0.23532131
sample estimates:
odds ratio
 0.1160797
```

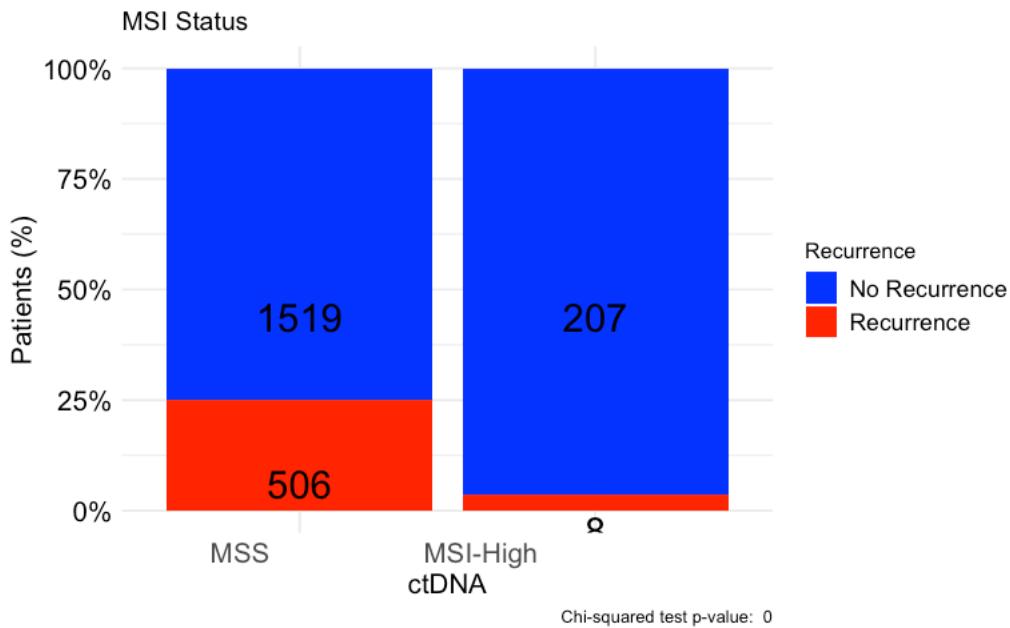
Hide

```
print(contingency_table)
```

	No Recurrence	Recurrence
MSS	1519	506
MSI-High	207	8

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "MSI Status",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#DFS by TMB status - All stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data$TMB <- factor(circ_data$TMB, levels = c("TMB-Low", "TMB-High"), labels = c("TMB-Low", "TMB-High"))

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~MSI, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
MSI, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
MSI=MSI-High	215	8	NA	NA	NA
MSI=MSS	2025	506	NA	NA	NA

Hide

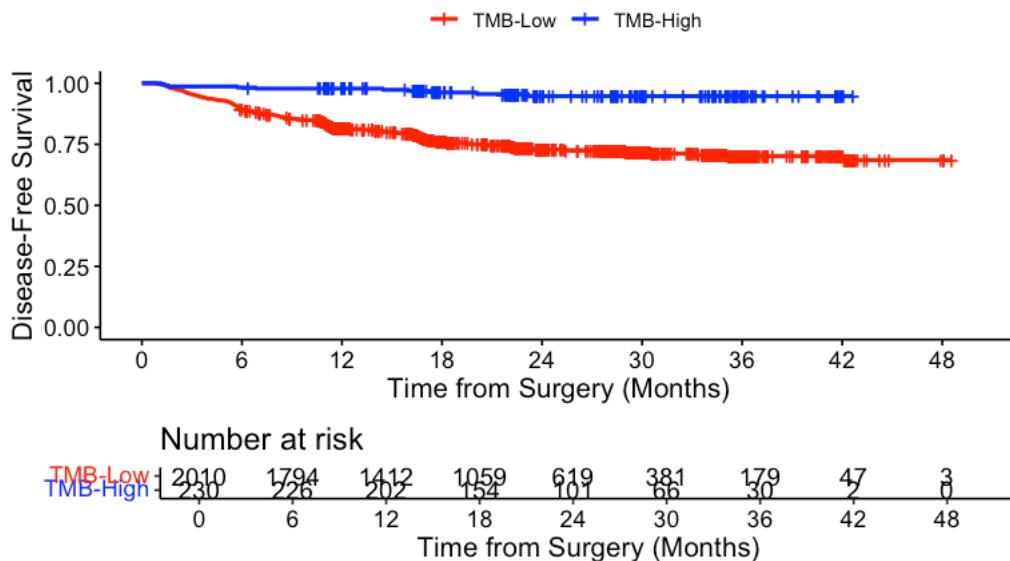
```
event_summary <- circ_data %>%
  group_by(TMB) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

TMB	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
TMB-Low	2010	504	0.25074627	25.074627
TMB-High	230	10	0.04347826	4.347826
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ TMB, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("red","blue"), title="DFS - TMB Status | All stages", ylab = "Disease-Free Survival", xlab="Time from Surgery (Months)", legend.labs=c("TMB-Low", "TMB-High"), legend.title="")
```

DFS - TMB Status | All stages



Hide

```
summary(KM_curve, times= c(24, 30, 36))
```

```
Call: survfit(formula = surv_object ~ TMB, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

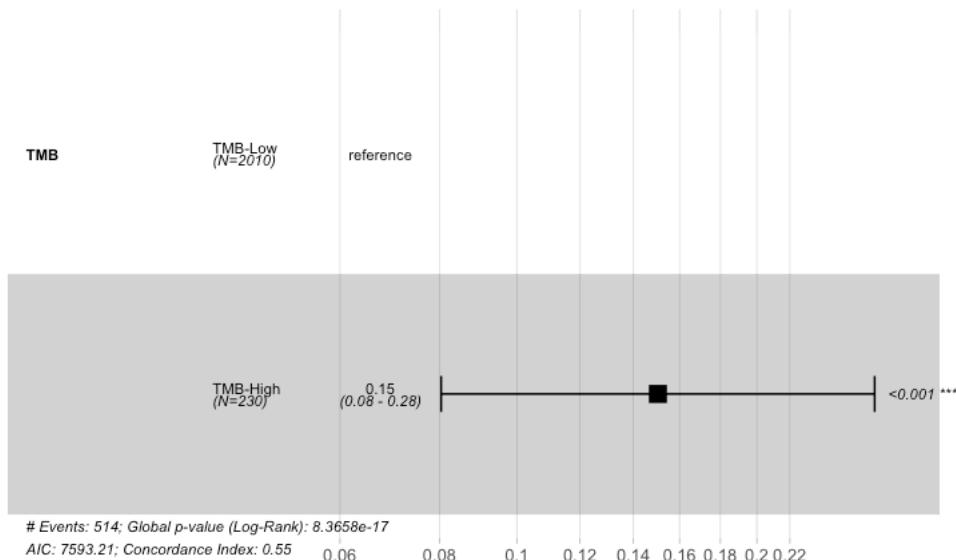
```
TMB=TMB-Low
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24    619     488    0.730  0.0108     0.708     0.750
 30    381      9    0.717  0.0115     0.693     0.738
 36    179      6    0.702  0.0127     0.676     0.726
```

```
TMB=TMB-High
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 24    101      10    0.947  0.0169     0.902     0.972
 30     66       0    0.947  0.0169     0.902     0.972
 36     30       0    0.947  0.0169     0.902     0.972
```

Hide

```
circ_data$TMB <- factor(circ_data$TMB, levels = c("TMB-Low", "TMB-High"), labels = c("TMB-Low", "TMB-High"))
cox_fit <- coxph(surv_object ~ TMB, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ TMB, data = circ_data)

n= 2240, number of events= 514

      coef exp(coef)  se(coef)      z Pr(>|z|)
TMBTMB-High -1.8953    0.1503    0.3194 -5.934 2.96e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
TMBTMB-High    0.1503      6.654   0.08036     0.281

Concordance= 0.548  (se = 0.004 )
Likelihood ratio test= 69.32  on 1 df,  p=<2e-16
Wald test      = 35.21  on 1 df,  p=3e-09
Score (logrank) test = 47.09  on 1 df,  p=7e-12
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

[1] "HR = 0.15 (0.08-0.28); p = 0"

[Hide](#)

```
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$TMB, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 48.98, df = 1, p-value = 2.586e-12
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 6.3e-16
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.06377761 0.25743811
sample estimates:
odds ratio
0.1358906
```

[Hide](#)

```
print(contingency_table)
```

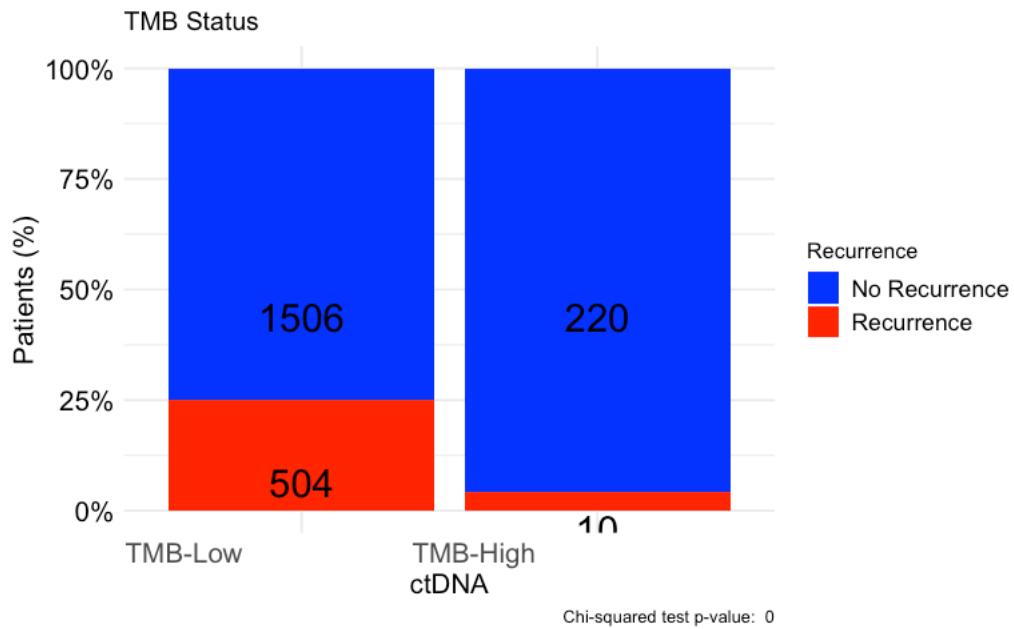
	No Recurrence	Recurrence
TMB-Low	1506	504
TMB-High	220	10

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "TMB Status",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



Hide

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data %>% filter(Eligible == "TRUE")
circ_data <- circ_data %>%
  mutate(
    RAS.BRAF = ifelse(RAS.BRAF == "TRUE", "RAS/BRAF WT", NA),
    TMB = ifelse(TMB == "TMB-High", "TMB High", NA),
    MSI = ifelse(MSI == "MSI-High", "MSI High", NA),
    BRAF.V600E = ifelse(BRAF.V600E == "MUT", "BRAF V600E", NA),
    KRAS.G12C = ifelse(KRAS.G12C == "MUT", "KRAS G12C", NA),
    ERBB2 = ifelse(ERBB2 == "MUT", "ERBB2", NA),
    TP53.Y220C = ifelse(TP53.Y220C == "MUT", "TP53 Y220C", NA)
  )
circ_data_long <- circ_data %>%
  gather(key = "group", value = "value", RAS.BRAF, TMB, MSI, BRAF.V600E, KRAS.G12C, ERBB2, TP53.Y220C) %>%
  filter(!is.na(value))

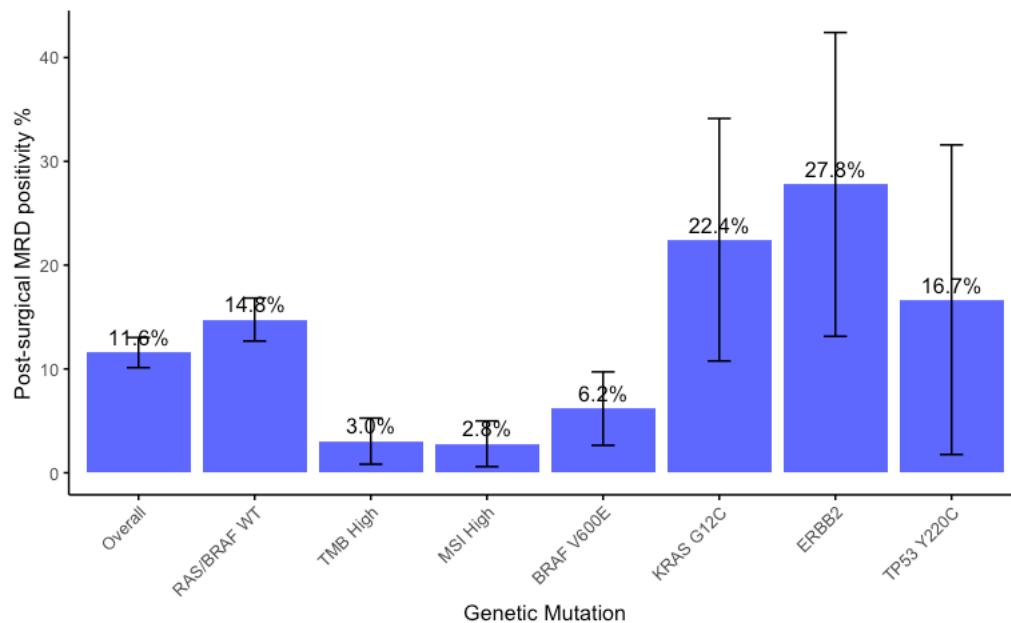
summary_data <- circ_data_long %>%
  group_by(value) %>%
  summarise(
    n = n(),
    positive = sum(ctDNA.MRD == "POSITIVE"),
    pct_positive = (positive / n) * 100,
    se = sqrt((pct_positive / 100) * (1 - pct_positive / 100) / n),
    ci_low = pct_positive - 1.96 * se * 100,
    ci_high = pct_positive + 1.96 * se * 100
  )

overall_summary <- circ_data_long %>%
  summarise(
    value = "Overall",
    n = n(),
    positive = sum(ctDNA.MRD == "POSITIVE"),
    pct_positive = (positive / n) * 100,
    se = sqrt((pct_positive / 100) * (1 - pct_positive / 100) / n),
    ci_low = pct_positive - 1.96 * se * 100,
    ci_high = pct_positive + 1.96 * se * 100
  )

summary_data <- bind_rows(overall_summary, summary_data)

summary_data$value <- factor(summary_data$value, levels = c("Overall", "RAS/BRAF WT", "TMB High", "MSI High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"))
ggplot(summary_data, aes(x = value, y = pct_positive)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.7) +
  geom_errorbar(aes(ymin = ci_low, ymax = ci_high), width = 0.2) +
  geom_text(aes(label = sprintf("%.1f%", pct_positive)), vjust = -0.5, color = "black") +
  labs(
    x = "Genetic Mutation",
    y = "Post-surgical MRD positivity %"
  ) +
  theme(
    panel.background = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.line = element_line(color = "black"),
    axis.ticks = element_line(color = "black"),
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.background = element_blank()
  )

```



#DFS by ctDNA at the MRD Window - BRAF V600E Landmark MRD timepoint

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$BRAF.V600E=="MUT",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_data$df <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.MRD, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.MRD.months, event = circ_data\$DFS.Event) ~
ctDNA.MRD, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	152	12	NA	NA	NA
ctDNA.MRD=POSITIVE	11	11	2.89	1.38	NA

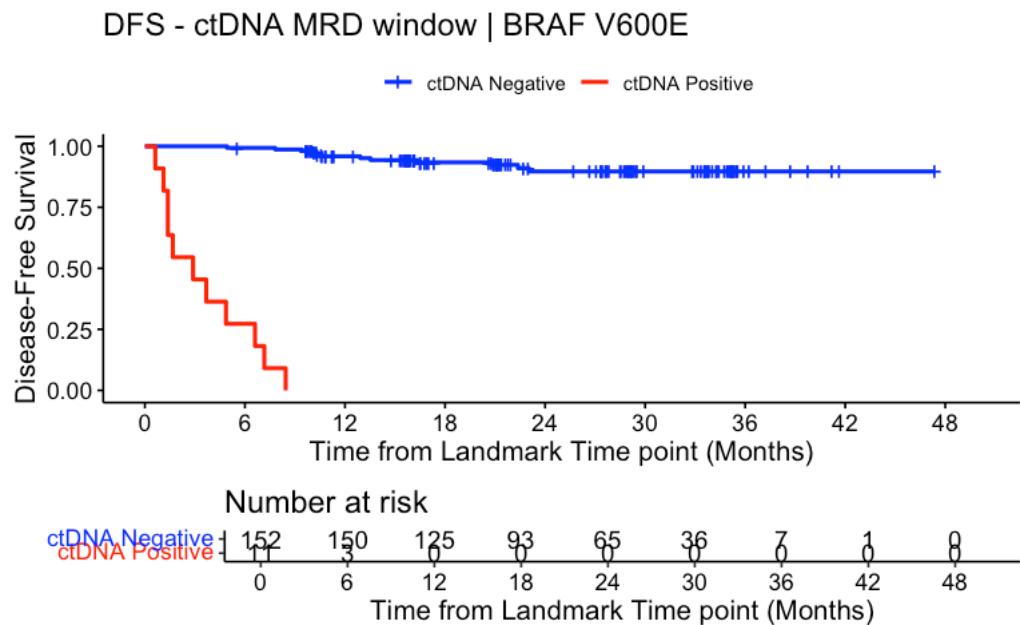
[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	152	12	0.07894737	7.894737
POSITIVE	11	11	1.00000000	100.000000
2 rows				

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "red"), title="DFS - ctDNA MRD window | BRAF V600E", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

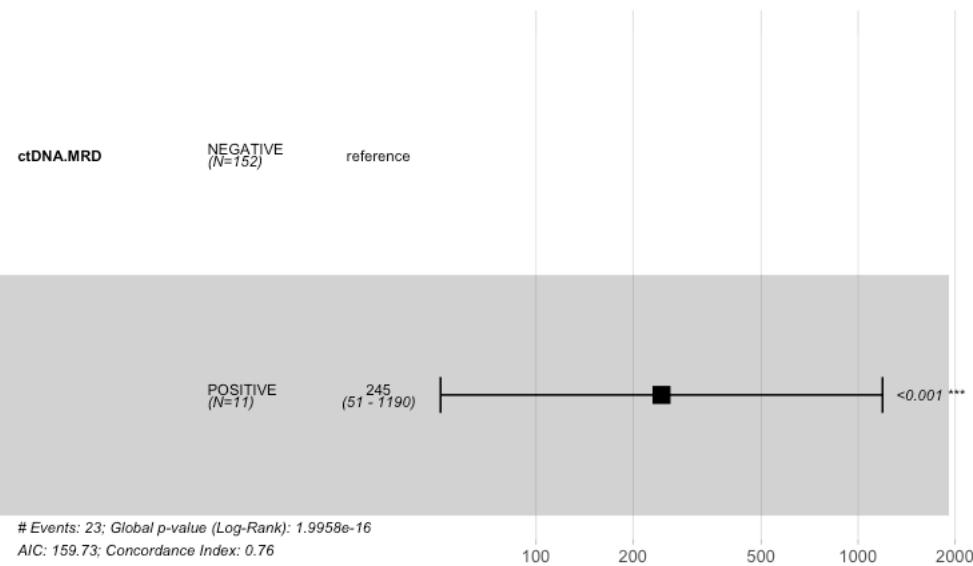
  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     152      0     1.000  0.0000     1.000     1.000
   24     65      12     0.897  0.0296     0.821     0.942

  ctDNA.MRD=POSITIVE
  time      n.risk      n.event      survival      std.err      lower 95% CI      upper 95% CI
    0          11          0          1          0          1          1
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 163, number of events= 23

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE  5.5020  245.1912  0.8061  6.826  8.75e-12 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE  245.2   0.004078   50.51    1190 

Concordance= 0.764 (se = 0.049 )
Likelihood ratio test= 67.61 on 1 df,  p=<2e-16
Wald test           = 46.59 on 1 df,  p=9e-12
Score (logrank) test = 265.5 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 245.19 (50.51-1190.25); p = 0"
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$BRAF.V600E=="MUT",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative", "Positive"))
circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$ctDNA.MRD, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)

```

Warning in stats::chisq.test(x, y, ...) :
Chi-squared approximation may be incorrect

[Hide](#)

```
print(chi_square_test)
```

Pearson's Chi-squared test with Yates' continuity correction

```
data: contingency_table
X-squared = 64.403, df = 1, p-value = 1.014e-15
```

[Hide](#)

```
fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)
```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 3.531e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
24.33026      Inf
sample estimates:
odds ratio
      Inf
```

[Hide](#)

```
print(contingency_table)
```

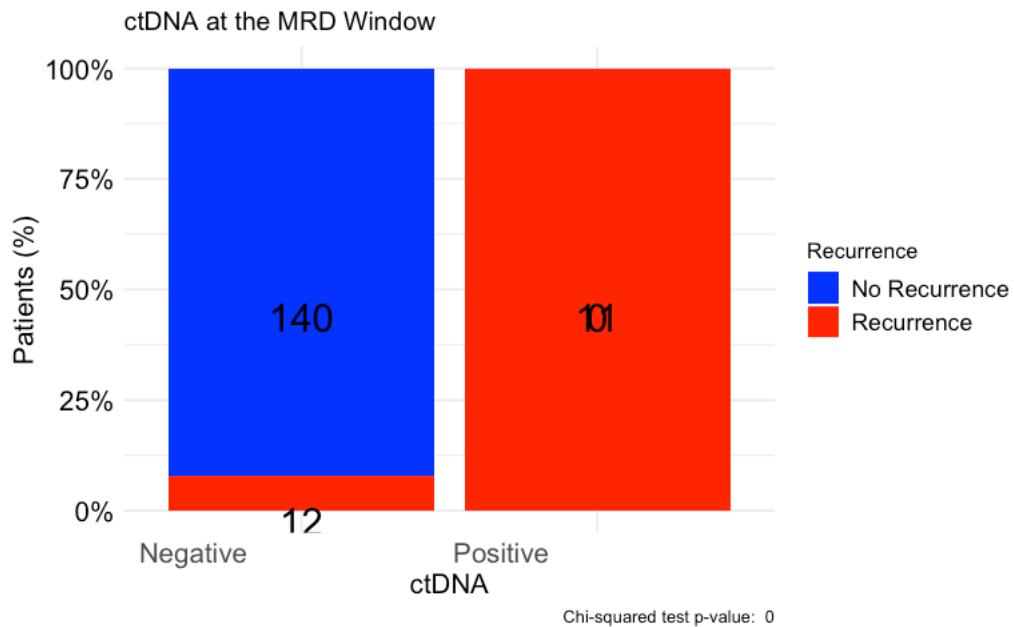
	No Recurrence	Recurrence
Negative	140	12
Positive	0	11

[Hide](#)

```

table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size

```



```

#Calculate the exact p-value from Chi-square test
p_value <- chi_square_test$p.value
p_value_mpfr <- mpfr(p_value, precBits = 256)
print(paste("Exact p-value with high precision:", format(p_value_mpfr, scientific = TRUE)))

```

```

[1] "Exact p-value with high precision: 1.01386630245057068611337444083588489579568644137486099765510516590438783
1687927e-15"

```

#DFS by ctDNA at the MRD Window - Forest plot with all subgroups of biomarkers

```

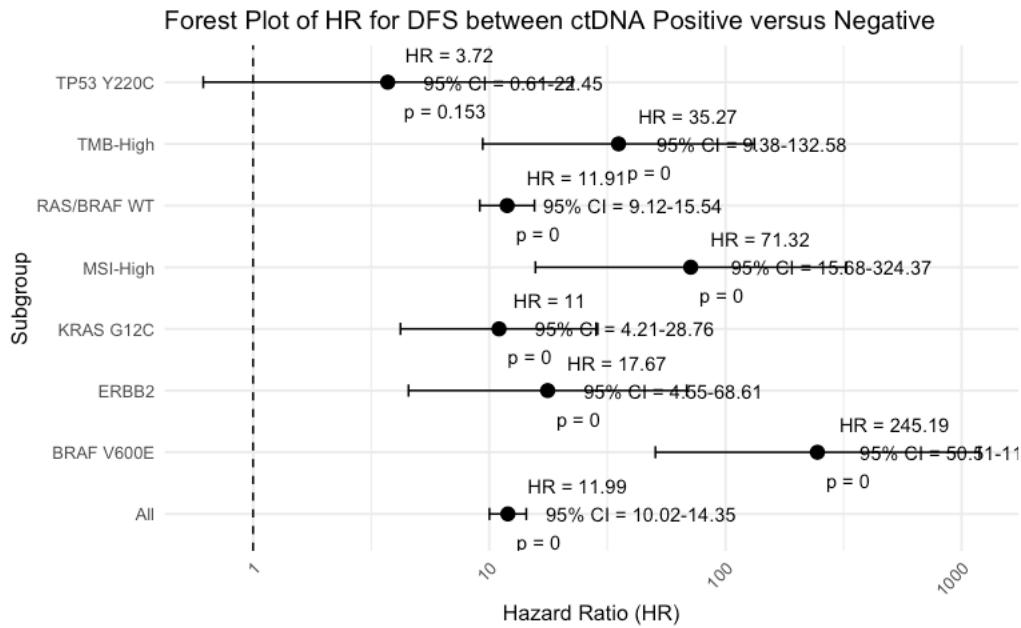
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "",]
circ_data <- circ_data[circ_data$DFS.MRD.months >= 0,]
perform_cox <- function(data, filter_col = NULL, filter_val = NULL) {
  if (!is.null(filter_col) & !is.null(filter_val)) {
    data <- data[data[[filter_col]] == filter_val,]
  }
  surv_object <- Surv(time = data$DFS.MRD.months, event = data$DFS.Event)
  cox_fit <- coxph(surv_object ~ ctDNA.MRD, data = data)
  cox_fit_summary <- summary(cox_fit)
  HR <- cox_fit_summary$coefficients[2]
  lower_CI <- cox_fit_summary$conf.int[3]
  upper_CI <- cox_fit_summary$conf.int[4]
  p_value <- cox_fit_summary$coefficients[5]
  return(c(HR, lower_CI, upper_CI, p_value))
}

results <- data.frame(
  Subgroup = c("All", "RAS/BRAF WT", "TMB-High", "MSI-High", "BRAF V600E", "KRAS G12C", "ERBB2", "TP53 Y220C"),
  HR = rep(NA, 8),
  lower_CI = rep(NA, 8),
  upper_CI = rep(NA, 8),
  p_value = rep(NA, 8)
)

results[1, 2:5] <- perform_cox(circ_data)
results[2, 2:5] <- perform_cox(circ_data, "RAS.BRAF", "TRUE")
results[3, 2:5] <- perform_cox(circ_data, "TMB", "TMB-High")
results[4, 2:5] <- perform_cox(circ_data, "MSI", "MSI-High")
results[5, 2:5] <- perform_cox(circ_data, "BRAF.V600E", "MUT")
results[6, 2:5] <- perform_cox(circ_data, "KRAS.G12C", "MUT")
results[7, 2:5] <- perform_cox(circ_data, "ERBB2", "MUT")
results[8, 2:5] <- perform_cox(circ_data, "TP53.Y220C", "MUT")

results$HR <- as.numeric(results$HR)
results$lower_CI <- as.numeric(results$lower_CI)
results$upper_CI <- as.numeric(results$upper_CI)
results$p_value <- as.numeric(results$p_value)
results$label_text <- paste0(
  "HR = ", round(results$HR, 2),
  "\n95% CI = ", round(results$lower_CI, 2), "-", round(results$upper_CI, 2),
  "\nnp = ", round(results$p_value, 3)
)
ggplot(results, aes(x = Subgroup, y = HR)) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = lower_CI, ymax = upper_CI), width = 0.2) +
  geom_text(aes(label = label_text), hjust = -0.2, vjust = 0.5, size = 3.5) +
  scale_y_log10() +
  geom_hline(yintercept = 1, linetype = "dashed") +
  labs(title = "Forest Plot of HR for DFS between ctDNA Positive versus Negative",
       x = "Subgroup",
       y = "Hazard Ratio (HR)") +
  coord_flip() +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

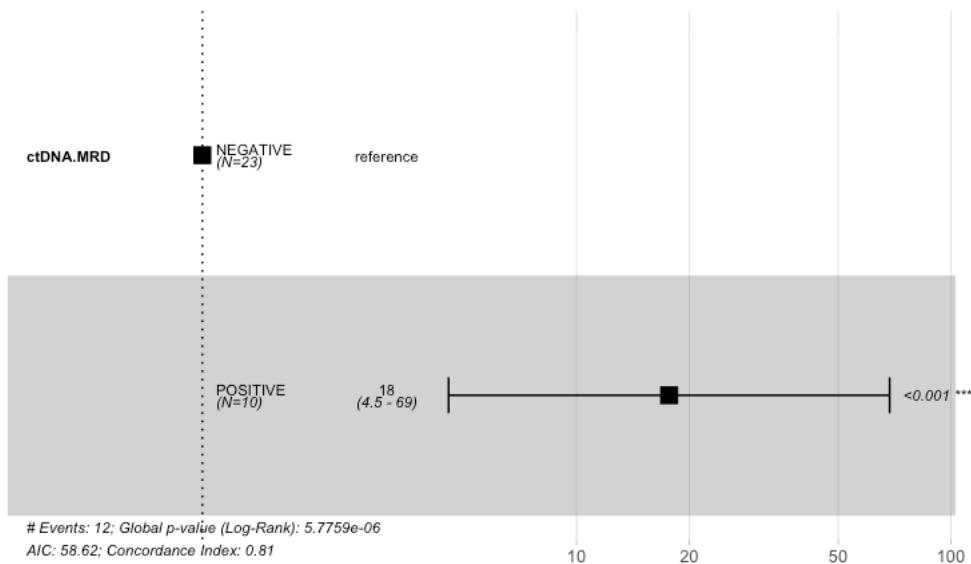


#DFS by ctDNA at the MRD Window - individual cox regression models for each biomarker to extract the exact p value

```
#ERBB2 Amplification
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ERBB2=="MUT",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 33, number of events= 12

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.8717  17.6668  0.6922 4.148 3.35e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE 17.67     0.0566    4.549    68.61

Concordance= 0.808 (se = 0.046 )
Likelihood ratio test= 20.56 on 1 df,  p=6e-06
Wald test           = 17.21 on 1 df,  p=3e-05
Score (logrank) test = 29.02 on 1 df,  p=7e-08
```

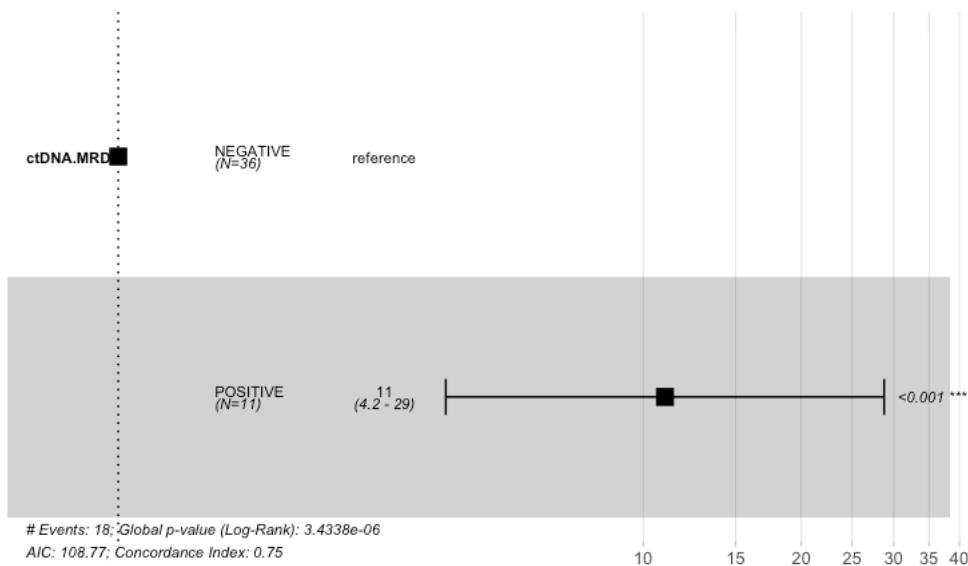
Hide

```
cox_fit_summary <- summary(cox_fit)

#KRAS G12C Amplification
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$KRAS.G12C=="MUT",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 47, number of events= 18

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.3978  10.9994  0.4904 4.889 1.01e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE      11     0.09091     4.206     28.76

Concordance= 0.746 (se = 0.048 )
Likelihood ratio test= 21.56 on 1 df,  p=3e-06
Wald test             = 23.9 on 1 df,  p=1e-06
Score (logrank) test = 35.4 on 1 df,  p=3e-09
```

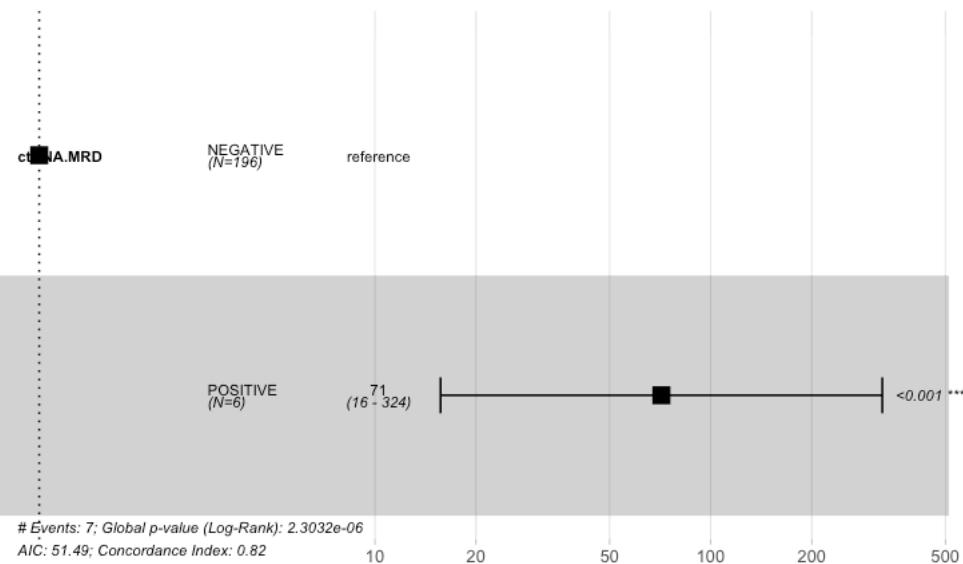
Hide

```
cox_fit_summary <- summary(cox_fit)

#MSI-High
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$MSI=="MSI-High",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 202, number of events= 7

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE  4.2671   71.3153   0.7729  5.521 3.37e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   71.32    0.01402    15.68    324.4

Concordance= 0.822  (se = 0.086 )
Likelihood ratio test= 22.32  on 1 df,  p=2e-06
Wald test            = 30.48  on 1 df,  p=3e-08
Score (logrank) test = 112.6  on 1 df,  p=<2e-16
```

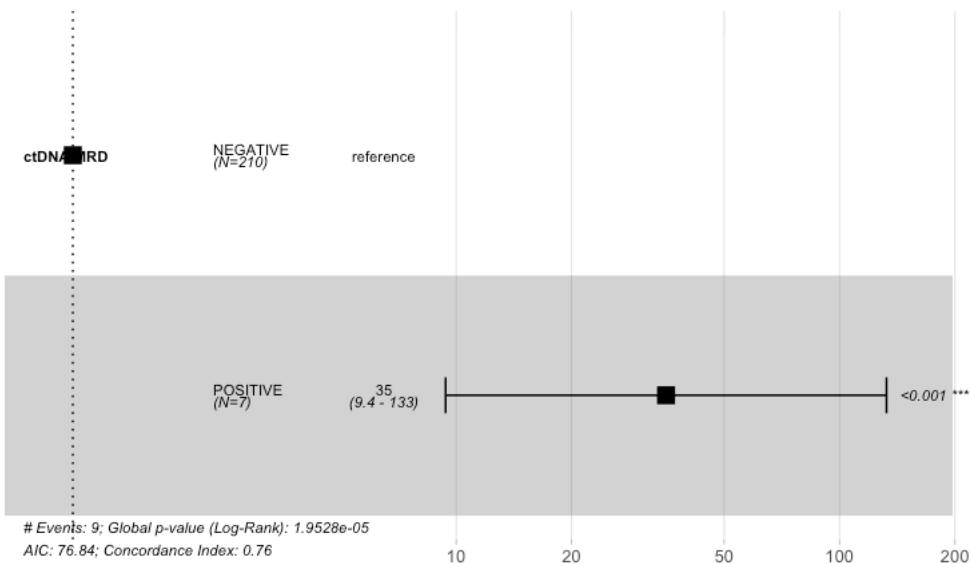
Hide

```
cox_fit_summary <- summary(cox_fit)

#TMB-High
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$TMB=="TMB-High",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 217, number of events= 9

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 3.5631   35.2728  0.6756 5.274 1.33e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   35.27    0.02835   9.384    132.6

Concordance= 0.755 (se = 0.083 )
Likelihood ratio test= 18.23 on 1 df,  p=2e-05
Wald test            = 27.82 on 1 df,  p=1e-07
Score (logrank) test = 72.18 on 1 df,  p=<2e-16
```

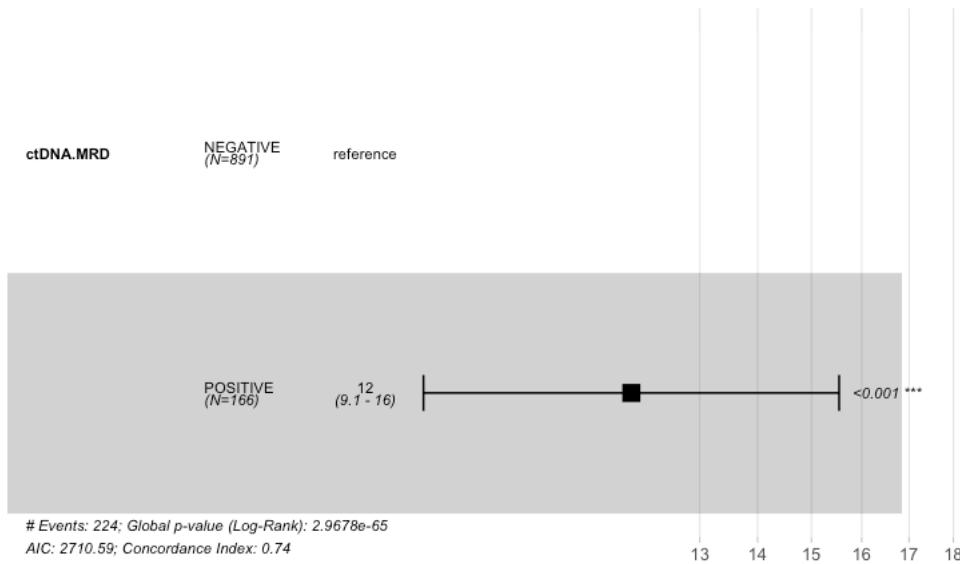
Hide

```
cox_fit_summary <- summary(cox_fit)

#RAS/BRAF WT
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RAS.BRAF=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 1057, number of events= 224

      coef exp(coef)  se(coef)    z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.4770  11.9052  0.1361 18.2  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE 11.91      0.084    9.118    15.54

Concordance= 0.742 (se = 0.015 )
Likelihood ratio test= 291 on 1 df,  p=<2e-16
Wald test           = 331.4 on 1 df,  p=<2e-16
Score (logrank) test = 527.1 on 1 df,  p=<2e-16
```

```
cox_fit_summary <- summary(cox_fit)

#Calculate the exact p value for ctDNA.MRD
z_value <- cox_fit_summary$coefficients["ctDNA.MRDPOSITIVE", "z"]
z_value_mpfr <- mpfr(abs(z_value), precBits = 256)
p_value_mpfr <- 2 * (1 - pnorm(z_value_mpfr))
print(paste("Exact p-value for ctDNA.MRD:", format(p_value_mpfr, scientific = TRUE)))
```

```
[1] "Exact p-value for ctDNA.MRD: 4.74816547159092565503741625416765968419957700036706750548603216789869177337870
2e-74"
```

```
#DFS by BRAF & MSI - ctDNA Positive Landmark MRD timepoint
```

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "",]
circ_data <- circ_data[circ_data$ctDNA.MRD == "POSITIVE",]
circ_data <- circ_data[circ_data$DFS.MRD.months >= 0,]

# Create the BRAF.MSI variable
circ_data$BRAF.MSI <- NA
circ_data <- circ_data %>%
  mutate(BRAF.MSI = case_when(
    BRAF.V600E == "WT" & MSI == "MSS" ~ 1,
    BRAF.V600E == "WT" & MSI == "MSI-High" ~ 2,
    BRAF.V600E == "MUT" & MSI == "MSI-High" ~ 3,
    BRAF.V600E == "MUT" & MSI == "MSS" ~ 4
  ))

circ_data$BRAF.MSI <- factor(circ_data$BRAF.MSI, levels = c(1, 2, 3, 4),
                             labels = c("BRAF WT & MSS", "BRAF WT & MSI-High",
                                       "BRAF V600E & MSI-High", "BRAF V600E & MSS"))

print(table(circ_data$BRAF.MSI, useNA = "ifany"))

```

BRAF WT & MSS	BRAF WT & MSI-High	BRAF V600E & MSI-High	BRAF V600E & MSS	<NA>
320	5	1	10	1

Hide

```

circ_data <- circ_data[!is.na(circ_data$BRAF.MSI),]
if(nrow(circ_data) == 0) {
  stop("No non-missing observations in the dataset after filtering.")
}
survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~BRAF.MSI, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.MRD.months, event = circ_data\$DFS.Event) ~ BRAF.MSI, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
BRAF.MSI=BRAF WT & MSS	320	249	5.520	4.895	7.16
BRAF.MSI=BRAF WT & MSI-High	5	3	4.731	0.559	NA
BRAF.MSI=BRAF V600E & MSI-High	1	1	0.624	NA	NA
BRAF.MSI=BRAF V600E & MSS	10	10	3.285	1.380	NA

Hide

```

event_summary <- circ_data %>%
  group_by(BRAF.MSI) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

BRAF.MSI	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
BRAF WT & MSS	320	249	0.778125	77.8125
BRAF WT & MSI-High	5	3	0.600000	60.0000
BRAF V600E & MSI-High	1	1	1.000000	100.0000
BRAF V600E & MSS	10	10	1.000000	100.0000

4 rows

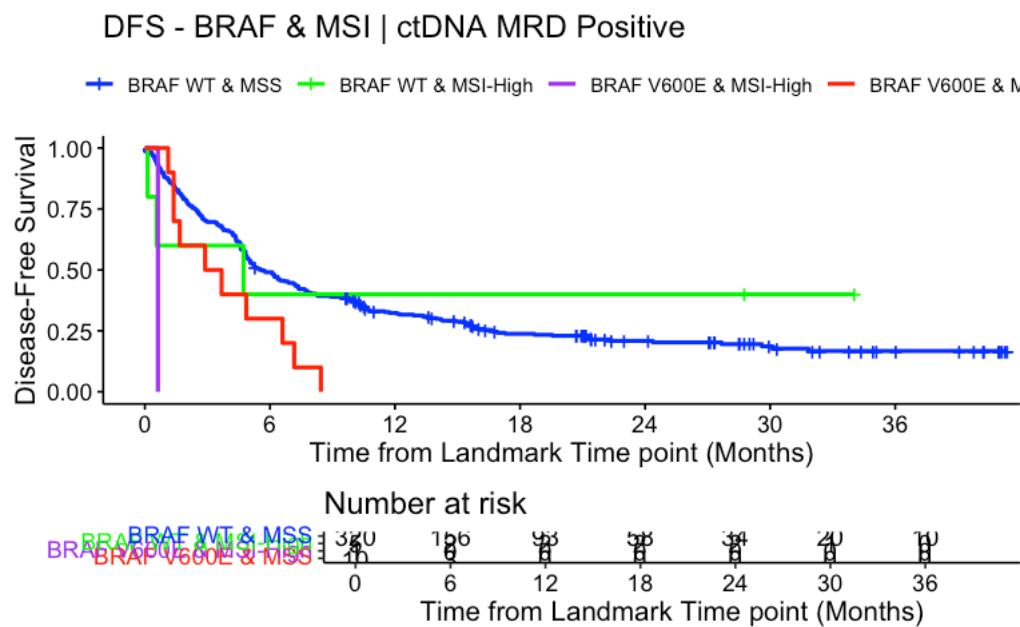
Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ BRAF.MSI, data = circ_data, conf.int = 0.95, conf.type = "log-log")

# Plot the Kaplan-Meier curve
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE,
           break.time.by = 6, palette = c("blue", "green", "purple", "red"),
           title = "DFS - BRAF & MSI | ctDNA MRD Positive", ylab = "Disease-Free Survival",
           xlab = "Time from Landmark Time point (Months)",
           legend.labs = c("BRAF WT & MSS", "BRAF WT & MSI-High",
                         "BRAF V600E & MSI-High", "BRAF V600E & MSS"),
           legend.title = "")

```



```
summary(KM_curve, times = c(0, 24))
```

Hide

```

Call: survfit(formula = surv_object ~ BRAF.MSI, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")

      BRAF.MSI=BRAF WT & MSS
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     320      4     0.988 0.00621     0.967     0.995
   24     34     240     0.209 0.02448     0.163     0.259

      BRAF.MSI=BRAF WT & MSI-High
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0      5      0     1.0  0.000     1.000     1.000
   24      2      3     0.4  0.219     0.052     0.753

      BRAF.MSI=BRAF V600E & MSI-High
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0      1      0      1      0      1      1

      BRAF.MSI=BRAF V600E & MSS
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     10      0      1      0      1      1

```

Hide

```

cox_fit <- coxph(surv_object ~ BRAF.MSI, data = circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ BRAF.MSI, data = circ_data)

n= 336, number of events= 263

```

```

      coef exp(coef) se(coef)      z Pr(>|z|)
BRAF.MSIBRAF WT & MSI-High -0.2883    0.7495  0.5818 -0.496  0.62018
BRAF.MSIBRAF V600E & MSI-High  2.6324   13.9073  1.0209  2.579  0.00992 **
BRAF.MSIBRAF V600E & MSS     0.7860    2.1947  0.3250  2.419  0.01557 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
BRAF.MSIBRAF WT & MSI-High    0.7495    1.3342  0.2397  2.344
BRAF.MSIBRAF V600E & MSI-High 13.9073    0.0719  1.8805 102.851
BRAF.MSIBRAF V600E & MSS      2.1947    0.4556  1.1608  4.149

Concordance= 0.511 (se = 0.008 )
Likelihood ratio test= 8.29 on 3 df,  p=0.04
Wald test            = 12.54 on 3 df,  p=0.006
Score (logrank) test = 17.48 on 3 df,  p=6e-04

```

Hide

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$BRAF.MSI, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)

```

```

Warning in stats::chisq.test(x, y, ...) :
  Chi-squared approximation may be incorrect

```

Hide

```

print(chi_square_test)

```

```

Pearson's Chi-squared test

data:  contingency_table
X-squared = 4.0751, df = 3, p-value = 0.2535

```

Hide

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.232
alternative hypothesis: two.sided
```

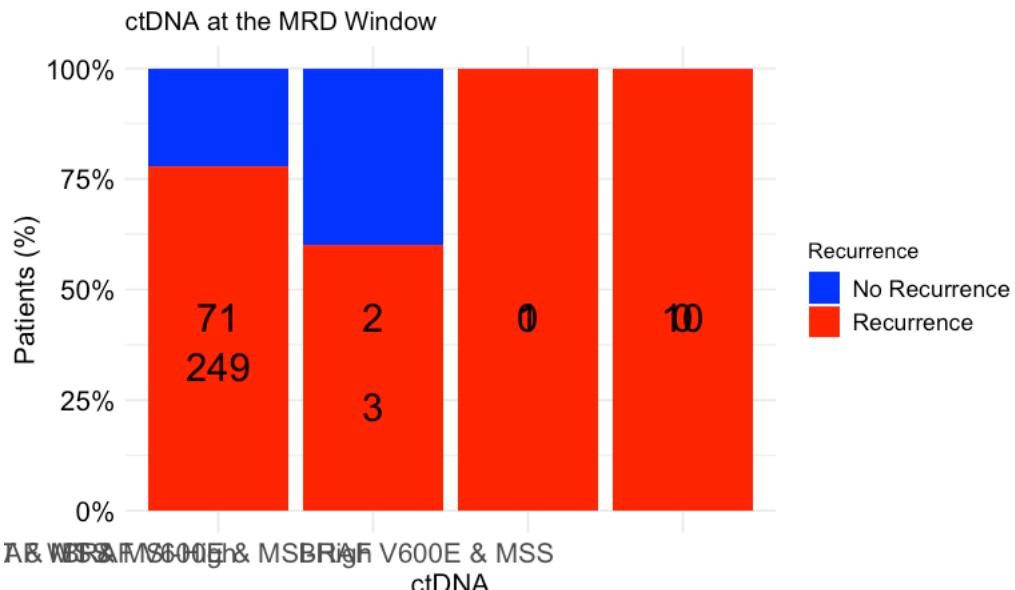
[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
BRAF WT & MSS	71	249
BRAF WT & MSI-High	2	3
BRAF V600E & MSI-High	0	1
BRAF V600E & MSS	0	10

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```



#DFS by BRAF & MSI - ctDNA Negative Landmark MRD timepoint

[Hide](#)

```

setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "",]
circ_data <- circ_data[circ_data$ctDNA.MRD == "NEGATIVE",]
circ_data <- circ_data[circ_data$DFS.MRD.months > 0,]

circ_data$BRAF.MSI <- NA
circ_data <- circ_data %>%
  mutate(BRAF.MSI = case_when(
    BRAF.V600E == "WT" & MSI == "MSS" ~ 1,
    BRAF.V600E == "WT" & MSI == "MSI-High" ~ 2,
    BRAF.V600E == "MUT" & MSI == "MSI-High" ~ 3,
    BRAF.V600E == "MUT" & MSI == "MSS" ~ 4
  ))
circ_data$BRAF.MSI <- factor(circ_data$BRAF.MSI, levels = c(1, 2, 3, 4),
                             labels = c("BRAF WT & MSS", "BRAF WT & MSI-High",
                                       "BRAF V600E & MSI-High", "BRAF V600E & MSS"))
print(table(circ_data$BRAF.MSI, useNA = "ifany"))

```

BRAF WT & MSS	BRAF WT & MSI-High	BRAF V600E & MSI-High	BRAF V600E & MSS
1526	93	103	49

Hide

```

circ_data <- circ_data[!is.na(circ_data$BRAF.MSI),]
if (any(!is.finite(circ_data$DFS.MRD.months)) || any(!is.finite(circ_data$DFS.Event))) {
  stop("Data contains non-finite values.")
}
if (nrow(circ_data) == 0) {
  stop("No non-missing observations in the dataset after filtering.")
}

survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~BRAF.MSI, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.MRD.months, event = circ_data\$DFS.Event) ~ BRAF.MSI, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
BRAF.MSI=BRAF WT & MSS	1526	219	NA	NA	NA
BRAF.MSI=BRAF WT & MSI-High	93	0	NA	NA	NA
BRAF.MSI=BRAF V600E & MSI-High	103	3	NA	NA	NA
BRAF.MSI=BRAF V600E & MSS	49	9	NA	NA	NA

Hide

```

event_summary <- circ_data %>%
  group_by(BRAF.MSI) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

BRAF.MSI	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
BRAF WT & MSS	1526	219	0.14351245	14.351245
BRAF WT & MSI-High	93	0	0.00000000	0.000000
BRAF V600E & MSI-High	103	3	0.02912621	2.912621
BRAF V600E & MSS	49	9	0.18367347	18.367347
4 rows				

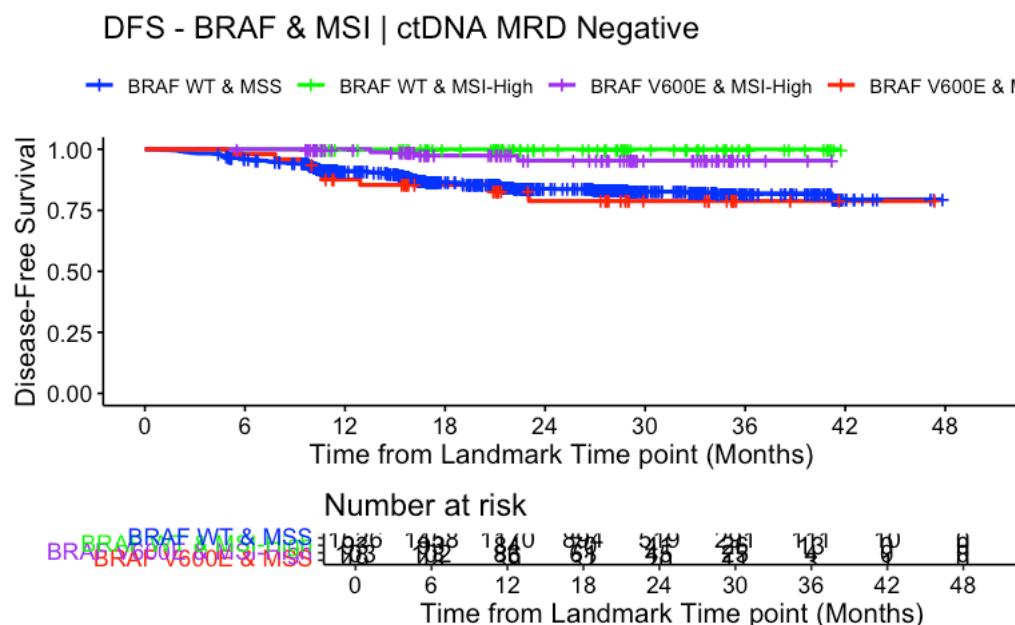
Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ BRAF.MSI, data = circ_data, conf.int = 0.95, conf.type = "log-log")

# Plot the Kaplan-Meier curve
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE,
           break.time.by = 6, palette = c("blue", "green", "purple", "red"),
           title = "DFS - BRAF & MSI | ctDNA MRD Negative", ylab = "Disease-Free Survival",
           xlab = "Time from Landmark Time point (Months)",
           legend.labs = c("BRAF WT & MSS", "BRAF WT & MSI-High",
                         "BRAF V600E & MSI-High", "BRAF V600E & MSS"),
           legend.title = "")

```



```
summary(KM_curve, times = c(0, 24))
```

Hide

```

Call: survfit(formula = surv_object ~ BRAF.MSI, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")

          BRAF.MSI=BRAF WT & MSS
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0    1526      0    1.000  0.0000    1.000    1.000
  24     519     210    0.838  0.0106    0.816    0.858

          BRAF.MSI=BRAF WT & MSI-High
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     93      0      1      0      1      1
  24     41      0      1      0      NA      NA

          BRAF.MSI=BRAF V600E & MSI-High
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0    103      0    1.000  0.0000    1.000    1.000
  24     45      3    0.954  0.0269    0.859    0.985

          BRAF.MSI=BRAF V600E & MSS
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     49      0    1.000  0.0000    1.000    1.000
  24     20      9    0.788  0.0658    0.622    0.887

```

[Hide](#)

```

cox_fit <- coxphf(surv_object ~ BRAF.MSI, data = circ_data)
summary(cox_fit)

```

```
coxphf(formula = surv_object ~ BRAF.MSI, data = circ_data)
```

Model fitted by Penalized ML

Confidence intervals and p-values by Profile Likelihood

	coef	se(coef)	exp(coef)	lower 0.95	upper 0.95	Chisq	p
BRAF.MSIBRAF WT & MSI-High	-3.4234200	1.4219914	0.03260075	0.0002591613	0.2215540	25.2847247	4.946103e-07
BRAF.MSIBRAF V600E & MSI-High	-1.5067027	0.5411239	0.22163959	0.0620647658	0.5473653	13.3792192	2.544276e-04
BRAF.MSIBRAF V600E & MSS	0.2475077	0.3328541	1.28082917	0.6222123751	2.3148209	0.5176429	4.718489e-01

Likelihood ratio test=38.29511 on 3 df, p=2.44771e-08, n=1771

Wald test = 14.17091 on 3 df, p = 0.002681504

Covariance-Matrix:

	BRAF.MSIBRAF WT & MSI-High	BRAF.MSIBRAF V600E & MSI-High	BRAF.MSIBRAF V600E & MSS
BRAF.MSIBRAF WT & MSI-High	2.022059448	0.004612586	0.004606403
BRAF.MSIBRAF V600E & MSI-High	0.004612586	0.292815067	0.004594608
BRAF.MSIBRAF V600E & MSS	0.004606403	0.004594608	0.110791854

[Hide](#)

```

circ_data$DFS.Event <- factor(circ_data$DFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Recurrence", "Recurrence"))
contingency_table <- table(circ_data$BRAF.MSI, circ_data$DFS.Event)
chi_square_test <- chisq.test(contingency_table)
print(chi_square_test)

```

Pearson's Chi-squared test

```
data: contingency_table
X-squared = 26.796, df = 3, p-value = 6.497e-06
```

[Hide](#)

```

fisher_exact_test <- fisher.test(contingency_table)
print(fisher_exact_test)

```

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 2.589e-08
alternative hypothesis: two.sided
```

[Hide](#)

```
print(contingency_table)
```

	No Recurrence	Recurrence
BRAF WT & MSS	1307	219
BRAF WT & MSI-High	93	0
BRAF V600E & MSI-High	100	3
BRAF V600E & MSS	40	9

Hide

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA at the MRD Window",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Recurrence",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Recurrence" = "blue", "Recurrence" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Recurrence label size
```

