

Code ▾

Galaxy 36mo Nakamura et al_06262024 Clinical analysis

```
library(swimplot) library(coxphf) library(grid) library(gtable) library(readr) library(mosaic) library(dplyr) library(survival)
library(survminer) library(ggplot2) library(scales) 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%)
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	943 (42%)
Observation	1,297 (58%)
BRAF.V600E	
BRAF wt	2,062 (92%)
BRAF V600E	178 (7.9%)
RAS	
¹ Median (Range); n (%)	

Characteristic	N = 2,240 ¹
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)

¹ 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	

¹Median (Range); n (%)

Characteristic	N = 2,240 ¹
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	943 (42%)
Observation	1,297 (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")
```

#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	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<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	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<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

```
#ctDNA MRD Detection rate Stage I/II vs III
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"))
circ_data$Stage_Grouped <- factor(ifelse(circ_data$Stage %in% c("I", "II"), "I/II", "II I"))
contingency_table <- table(circ_data$Stage_Grouped, circ_data$ctDNA.MRD)
chi_square_test <- chisq.test(contingency_table)
print(contingency_table)
```

NEGATIVE POSITIVE

I/II	810	47
III	963	290

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 116.96, df = 1, p-value < 2.2e-16
```

#ctDNA Surveillance Detection rate Stage I/II vs III

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIVE", "POSITIVE"))
circ_data$Stage_Grouped <- factor(ifelse(circ_data$Stage %in% c("I", "II"), "I/II", "II I"))
contingency_table <- table(circ_data$Stage_Grouped, circ_data$ctDNA.Surveillance)
chi_square_test <- chisq.test(contingency_table)
print(contingency_table)
```

NEGATIVE POSITIVE

I/II	543	47
III	938	266

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 53.889, df = 1, p-value = 2.122e-13
```

#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)
```

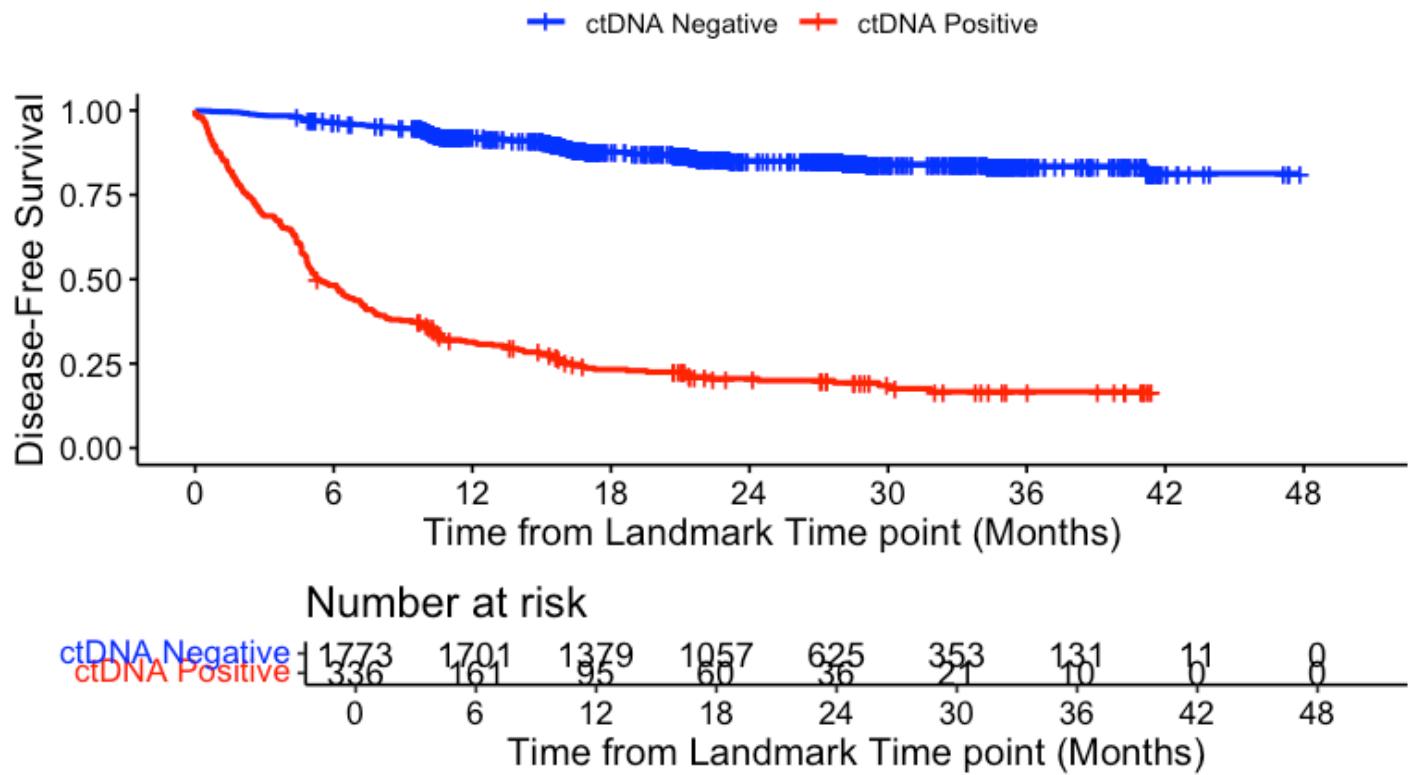
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1773	233	0.1314157	13.14157
POSITIVE	336	263	0.7827381	78.27381
NA	1	NA	NA	NA

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	
30	353	6	0.841	0.01025	0.820		0.860	
36	131	2	0.835	0.01101	0.812		0.856	

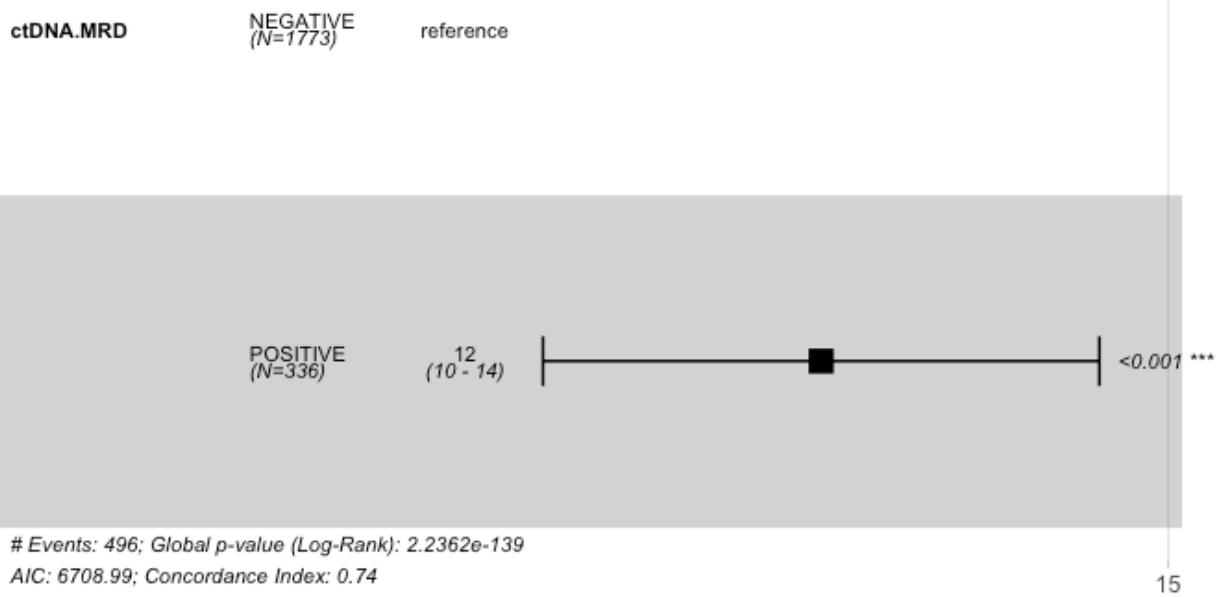
ctDNA.MRD=POSITIVE

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	
30	21	3	0.185	0.0242	0.140		0.234	
36	10	2	0.167	0.0250	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)

# 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"
```

#DFS by ctDNA at the MRD Window - Stage High Risk II/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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="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)

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

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

17 observations deleted due to missingness

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	1158	106	NA	NA	NA
ctDNA.MRD=POSITIVE	204	145	9.46	7.89	11.5

[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)
```

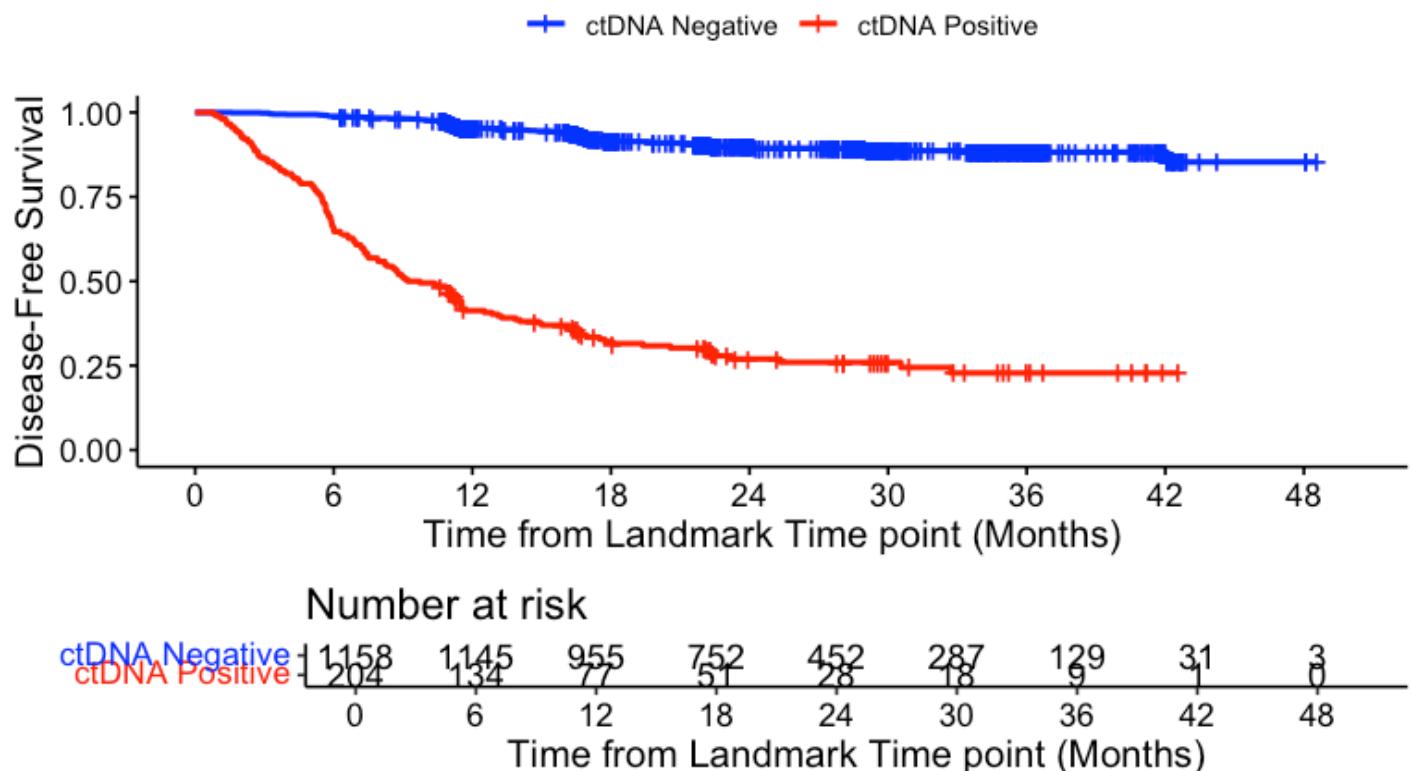
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1158	106	0.09153713	9.153713
POSITIVE	204	145	0.71078431	71.078431
NA	17	NA	NA	NA

3 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$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 | Stages High Risk II-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 | Stages High Risk II-III



```
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")
```

17 observations deleted due to missingness

ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	452	100	0.896	0.0100	0.875		0.914	
30	287	4	0.887	0.0111	0.863		0.906	
36	129	1	0.883	0.0117	0.857		0.904	

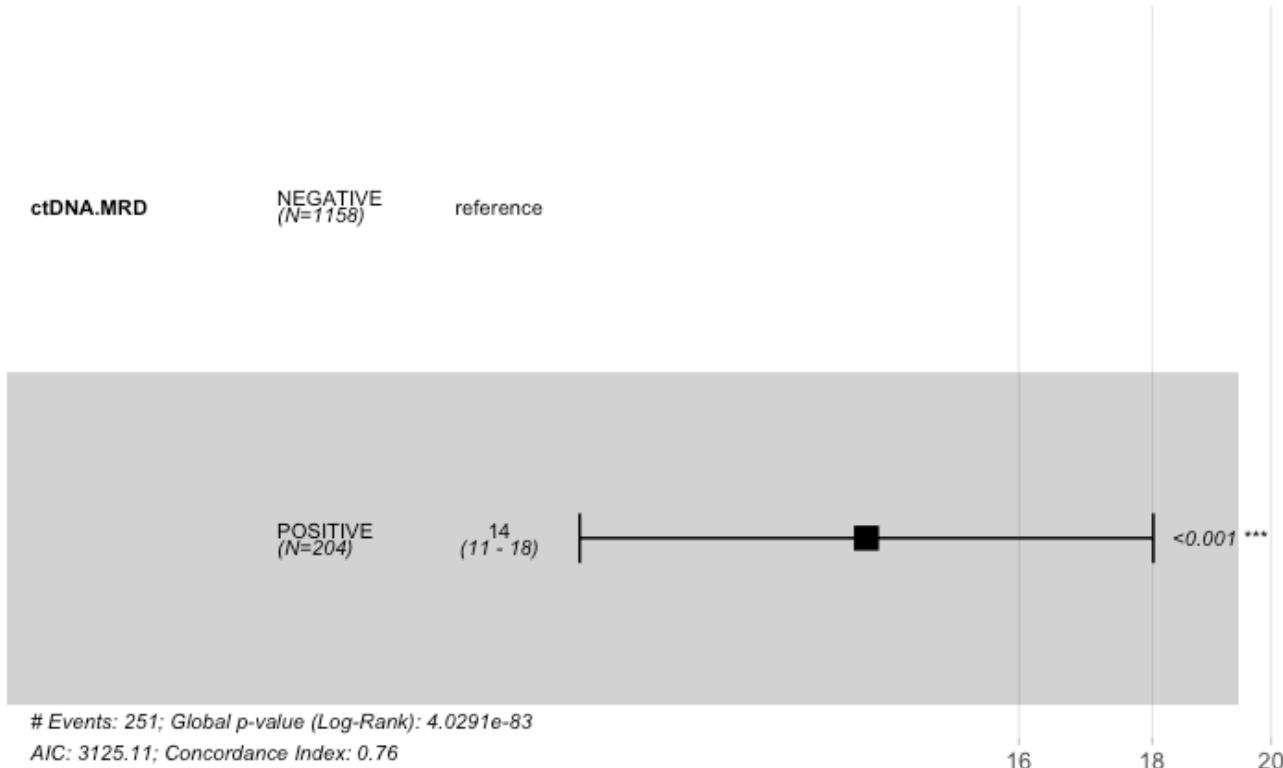
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	28	142	0.270	0.0337	0.206		0.337	
30	18	1	0.260	0.0339	0.196		0.328	
36	9	2	0.228	0.0364	0.161		0.303	

[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= 1362, number of events= 251
(17 observations deleted due to missingness)

            coef exp(coef) se(coef)    z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.6379   13.9843   0.1293 20.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   13.98    0.07151    10.85    18.02

Concordance= 0.761 (se = 0.014 )
Likelihood ratio test= 373.1 on 1 df,  p=<2e-16
Wald test            = 415.9 on 1 df,  p=<2e-16
Score (logrank) test = 703.1 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 = 13.98 (10.85-18.02); p = 0"
```

```
#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](#)

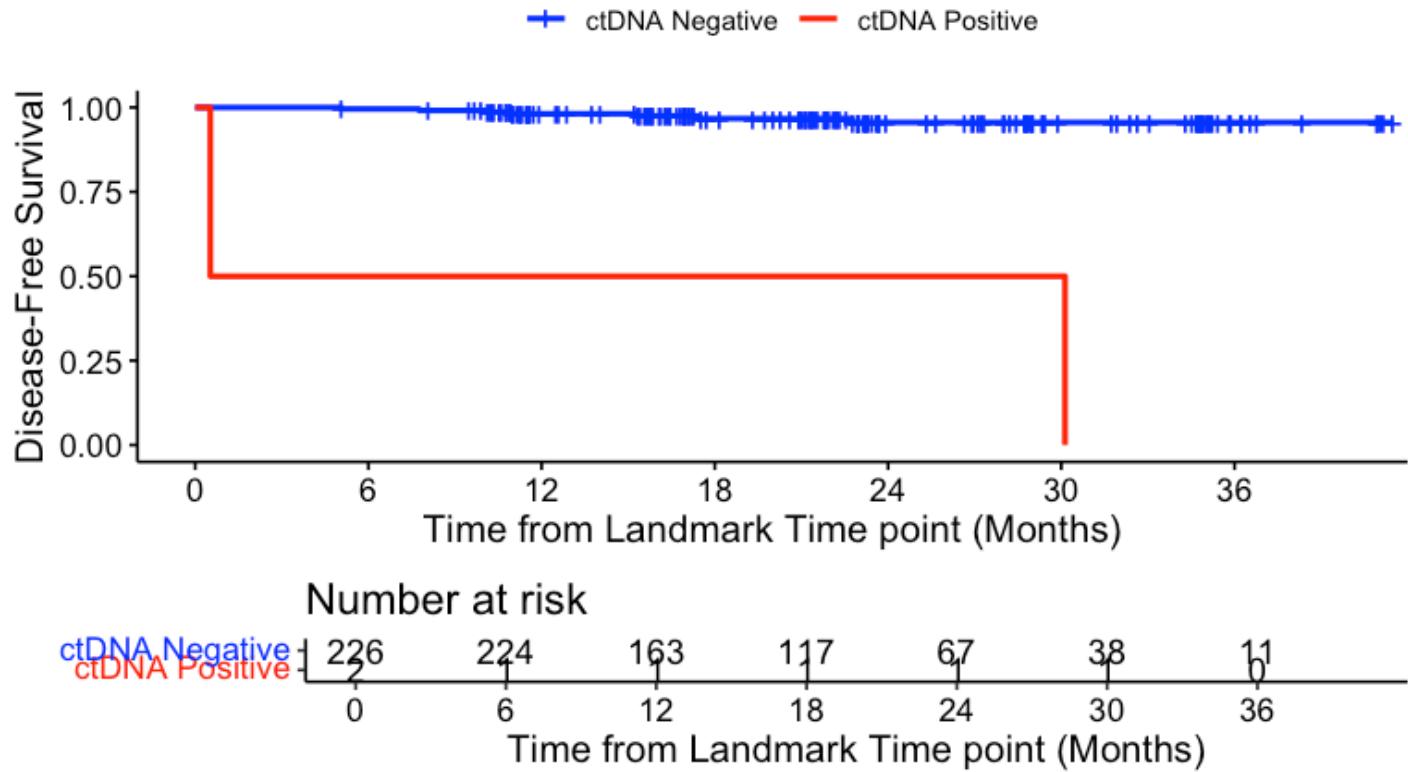
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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=c("blue","red"), title="DFS - ctDNA MRD window | Stage I", y
lab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs
=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage I



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")
```

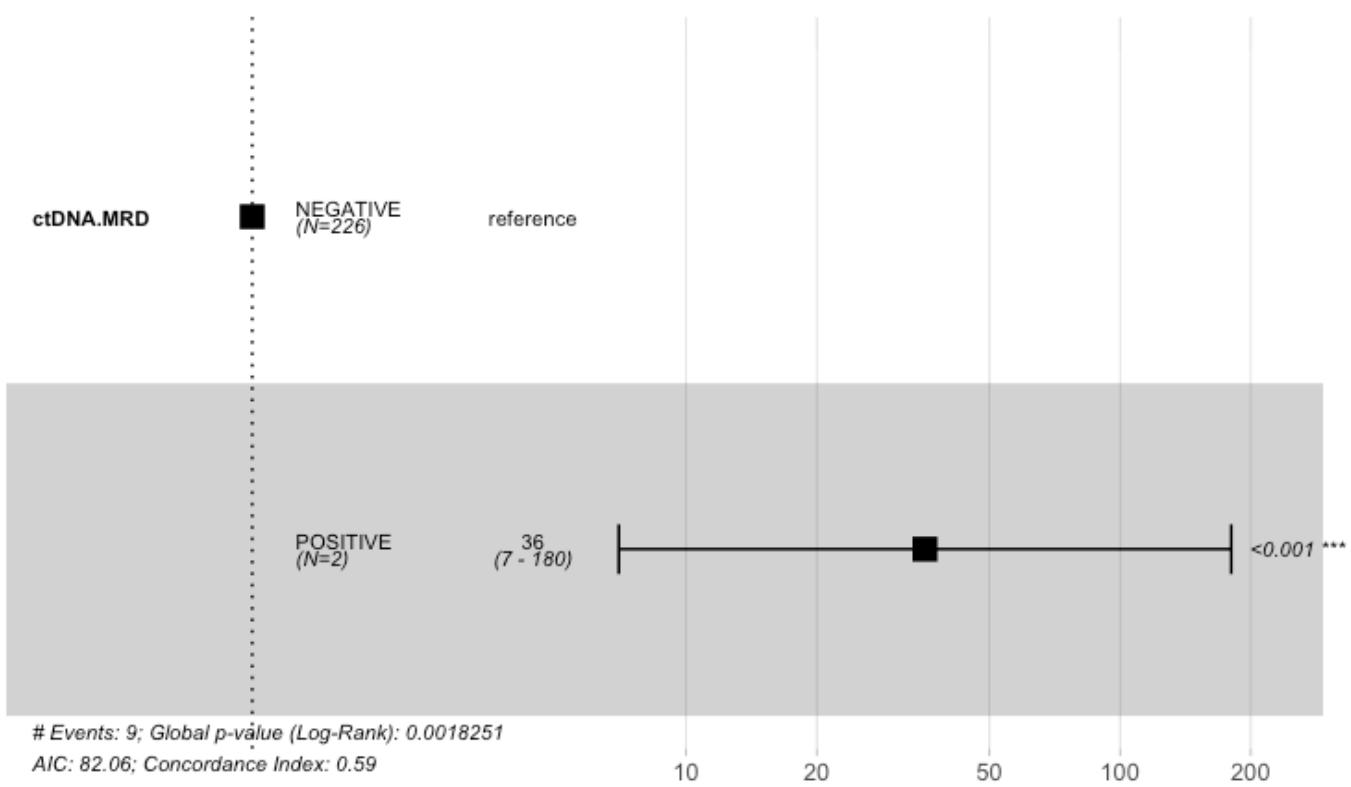
ctDNA.MRD=NEGATIVE							
CI	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
97	24.0000	67.0000	7.0000	0.9556	0.0176	0.9043	0.97

ctDNA.MRD=POSITIVE							
CI	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
41	24.00000	1.00000	1.00000	0.50000	0.35355	0.00598	0.910

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= 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"
```

#DFS by ctDNA at the MRD Window - 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$Stage %in% c("I", "III", "IV")),]
circ_data <- circ_data[circ_data$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

[Hide](#)

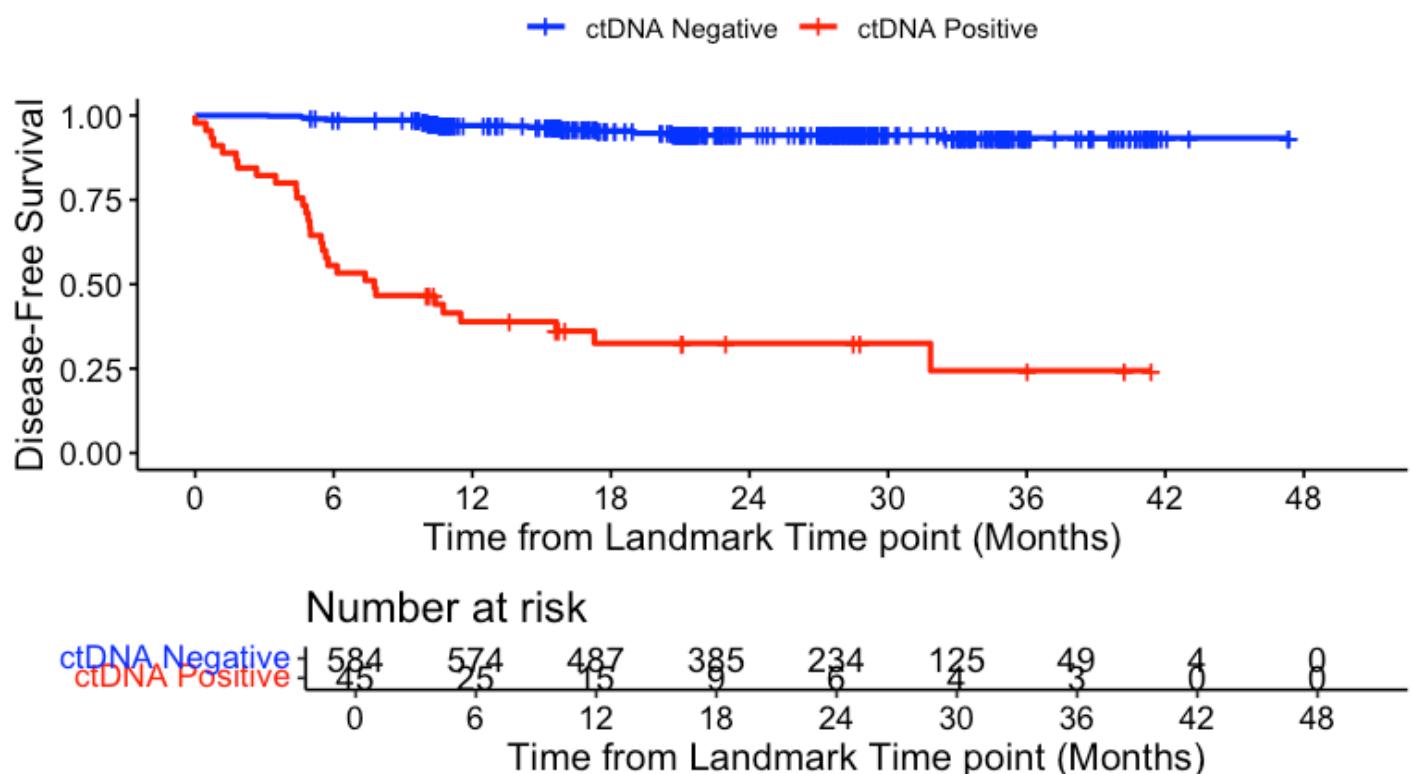
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	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.labels=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")
```

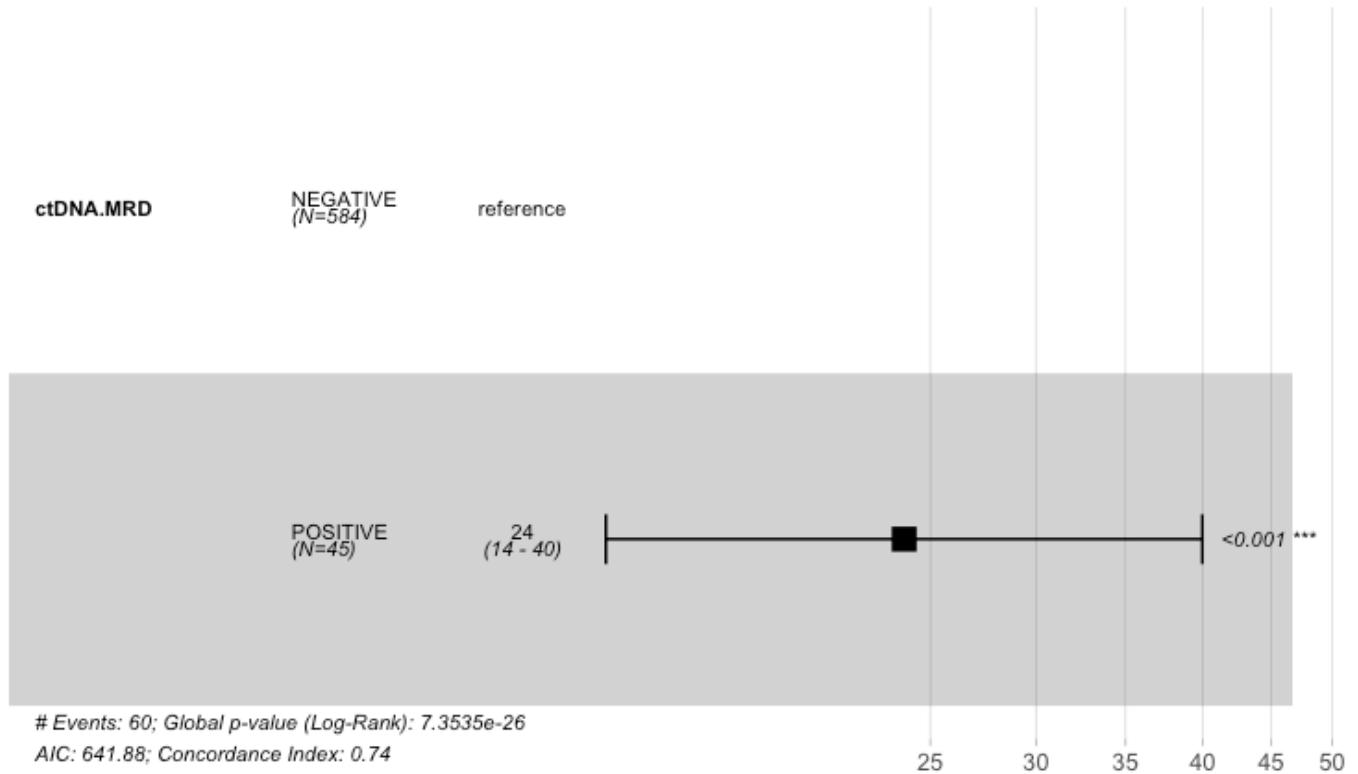
CI	ctDNA.MRD=NEGATIVE						
	time	n.risk	n.event	survival	std.err	lower	95% CI
24.0000	234.0000	29.0000	0.9413	0.0108	0.9159	0.95	

CI	ctDNA.MRD=POSITIVE						
	time	n.risk	n.event	survival	std.err	lower	95% CI
24.0000	6.0000	29.0000	0.3250	0.0749	0.1864	0.47	

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)
```

```
# 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"
```

#DFS by ctDNA at the MRD Window - Stage II & T3N0/T4N0 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$Stage %in% c("I", "III", "IV")),]
circ_data <- circ_data[circ_data$StageII.Group!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Stage.II.Risk <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.II.TNM = case_when(
    ctDNA.MRD == "NEGATIVE" & StageII.Group == "T3N0" ~ 1,
    ctDNA.MRD == "POSITIVE" & StageII.Group == "T3N0" ~ 2,
    ctDNA.MRD == "NEGATIVE" & StageII.Group == "T4N0" ~ 3,
    ctDNA.MRD == "POSITIVE" & StageII.Group == "T4N0" ~ 4
  ))
circ_data <- circ_data[circ_data$ctDNA.Stage.II.TNM!="",]
survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.Stage.II.TNM, data = circ_data)

```

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

17 observations deleted due to missingness

	n	events	median	0.95LCL	0.95UCL
ctDNA.Stage.II.TNM=1	476	18	NA	NA	NA
ctDNA.Stage.II.TNM=2	29	18	10.74	6.14	NA
ctDNA.Stage.II.TNM=3	93	11	NA	NA	NA
ctDNA.Stage.II.TNM=4	14	10	5.22	4.37	NA

Hide

```

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

```

ctDNA.Stage.II.TNM	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	476	18	0.03781513	3.781513
2	29	18	0.62068966	62.068966
3	93	11	0.11827957	11.827957
4	14	10	0.71428571	71.428571
NA	17	NA	NA	NA

5 rows

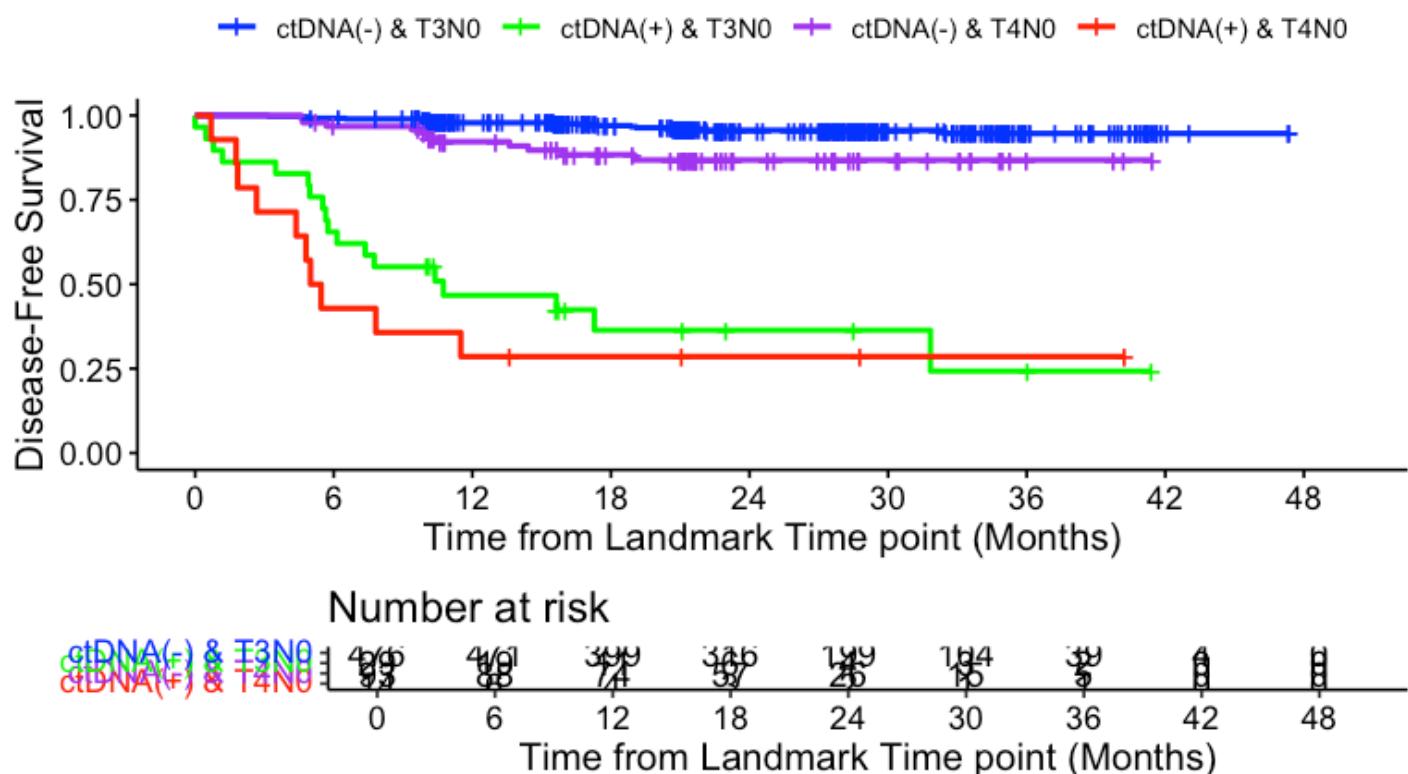
Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.II.TNM, 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","green","purple", "red"), title="DFS - ctDNA MRD & Stage II TNM", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & T3N0", "ctDNA(+) & T3N0", "ctDNA(-) & T4N0", "ctDNA(+) & T4N0"), legend.title="")

```

DFS - ctDNA MRD & Stage II TNM



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

[Hide](#)

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

17 observations deleted due to missingness
  ctDNA.Stage.II.TNM=1
    time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI      24.0000    199.0000    17.0000    0.9561    0.0107    0.9295    0.97
29

  ctDNA.Stage.II.TNM=2
    time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI      24.0000     4.0000    17.0000    0.3638    0.0996    0.1793    0.55
16

  ctDNA.Stage.II.TNM=3
    time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI      24.0000    26.0000    11.0000    0.8680    0.0375    0.7730    0.92
52

  ctDNA.Stage.II.TNM=4
    time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI      24.0000     2.0000    10.0000    0.2857    0.1207    0.0883    0.52
37
```

[Hide](#)

```
circ_data$ctDNA.Stage.II.TNM <- factor(circ_data$ctDNA.Stage.II.TNM, levels=c
  ("1","2","3","4"), labels = c("ctDNA(-) & T3N0", "ctDNA(+) & T3N0", "ctDNA(-) & T4N0",
  "ctDNA(+) & T4N0"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.II.TNM, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.Stage.II.TNM, data = circ_data)

n= 612, number of events= 57
 (17 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.II.TNMctDNA(+) & T3N0	3.3181	27.6092	0.3364	9.864	< 2e-16 ***
ctDNA.Stage.II.TNMctDNA(-) & T4N0	1.1962	3.3077	0.3829	3.124	0.00178 **
ctDNA.Stage.II.TNMctDNA(+) & T4N0	3.6897	40.0340	0.3977	9.277	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Stage.II.TNMctDNA(+) & T3N0	27.609	0.03622	14.279	53.382
ctDNA.Stage.II.TNMctDNA(-) & T4N0	3.308	0.30233	1.562	7.006
ctDNA.Stage.II.TNMctDNA(+) & T4N0	40.034	0.02498	18.360	87.295

Concordance= 0.798 (se = 0.032)
 Likelihood ratio test= 110.8 on 3 df, p=<2e-16
 Wald test = 135 on 3 df, p=<2e-16
 Score (logrank) test = 297.7 on 3 df, p=<2e-16

[Hide](#)

```
#Repeat analysis to compare ctDNA MRD (-) vs (+) in T4N0
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$Stage %in% c("I", "III", "IV")),]
circ_data <- circ_data[circ_data$StageII.Group!="",]
circ_data <- circ_data[circ_data$DFS.MRD.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$ctDNA.Stage.II.Risk <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.II.TNM = case_when(
    ctDNA.MRD == "NEGATIVE" & StageII.Group == "T3N0" ~ 1,
    ctDNA.MRD == "POSITIVE" & StageII.Group == "T3N0" ~ 2,
    ctDNA.MRD == "NEGATIVE" & StageII.Group == "T4N0" ~ 3,
    ctDNA.MRD == "POSITIVE" & StageII.Group == "T4N0" ~ 4
  ))
circ_data <- circ_data[circ_data$ctDNA.Stage.II.TNM!="",]
survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.Stage.II.TNM, data = circ_data)
```

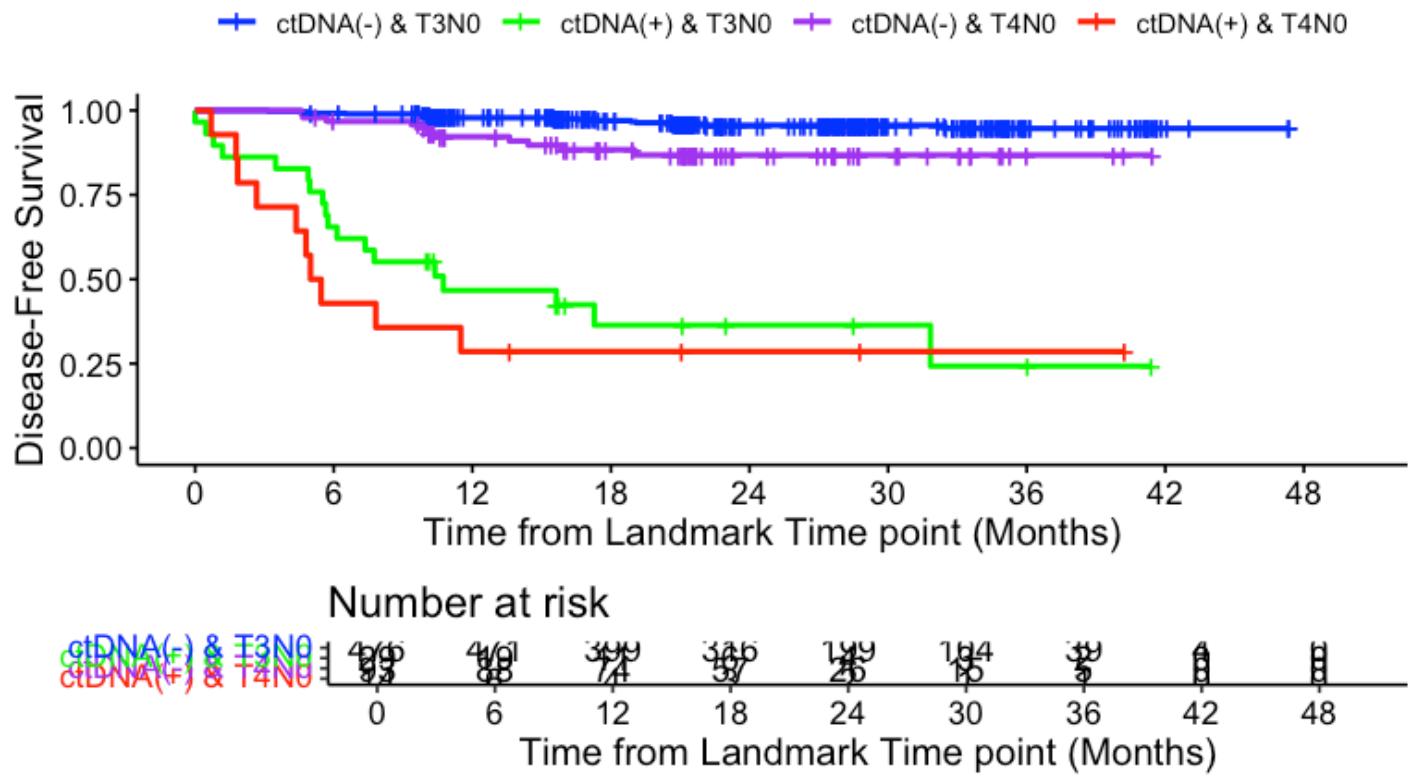
```
Call: survfit(formula = Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event) ~
  ctDNA.Stage.II.TNM, data = circ_data)
```

```
17 observations deleted due to missingness
      n  events median 0.95LCL 0.95UCL
ctDNA.Stage.II.TNM=1 476      18      NA      NA      NA
ctDNA.Stage.II.TNM=2  29      18  10.74     6.14      NA
ctDNA.Stage.II.TNM=3  93      11      NA      NA      NA
ctDNA.Stage.II.TNM=4  14      10     5.22     4.37      NA
```

Hide

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.II.TNM, 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","green","purple", "red"), title="DFS - ctDNA MRD & Stage II TNM", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & T3N0", "ctDNA(+) & T3N0", "ctDNA(-) & T4N0", "ctDNA(+) & T4N0"), legend.title="")
```

DFS - ctDNA MRD & Stage II TNM



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

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

17 observations deleted due to missingness

ctDNA.Stage.II.TNM=1

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
29	24.0000	199.0000	17.0000	0.9561	0.0107	0.9295		0.97	

ctDNA.Stage.II.TNM=2

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
16	24.0000	4.0000	17.0000	0.3638	0.0996	0.1793		0.55	

ctDNA.Stage.II.TNM=3

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
52	24.0000	26.0000	11.0000	0.8680	0.0375	0.7730		0.92	

ctDNA.Stage.II.TNM=4

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
37	24.0000	2.0000	10.0000	0.2857	0.1207	0.0883		0.52	

```
circ_data$ctDNA.Stage.II.TNM <- factor(circ_data$ctDNA.Stage.II.TNM, levels=c
  ("2","4","1","3"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.II.TNM, data=circ_data)
summary(cox_fit)
```

Call:

coxph(formula = surv_object ~ ctDNA.Stage.II.TNM, data = circ_data)

n= 612, number of events= 57

(17 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.II.TNM4	0.37158	1.45003	0.39523	0.940	0.347
ctDNA.Stage.II.TNM1	-3.31815	0.03622	0.33640	-9.864	< 2e-16 ***
ctDNA.Stage.II.TNM3	-2.12190	0.11980	0.38492	-5.513	3.54e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Stage.II.TNM4	1.45003	0.6896	0.66828	3.14625
ctDNA.Stage.II.TNM1	0.03622	27.6092	0.01873	0.07003
ctDNA.Stage.II.TNM3	0.11980	8.3470	0.05634	0.25475

Concordance= 0.798 (se = 0.032)

Likelihood ratio test= 110.8 on 3 df, p=<2e-16

Wald test = 135 on 3 df, p=<2e-16

Score (logrank) test = 297.7 on 3 df, p=<2e-16

#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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NA	1	NA	NA	NA
3 rows				

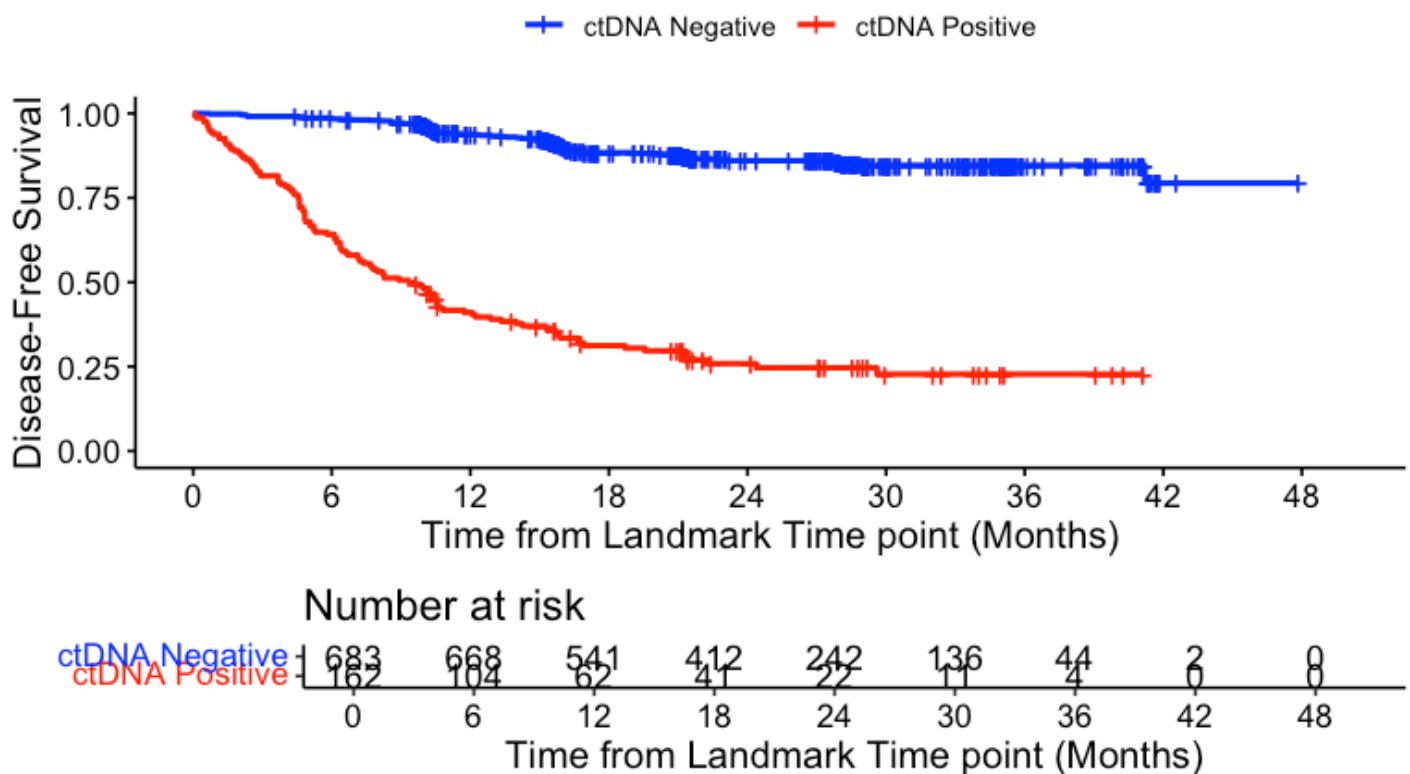
Hide

```

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.lab=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

CI	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
70	24.0000	242.0000	78.0000	0.8600	0.0152	0.8272	0.88

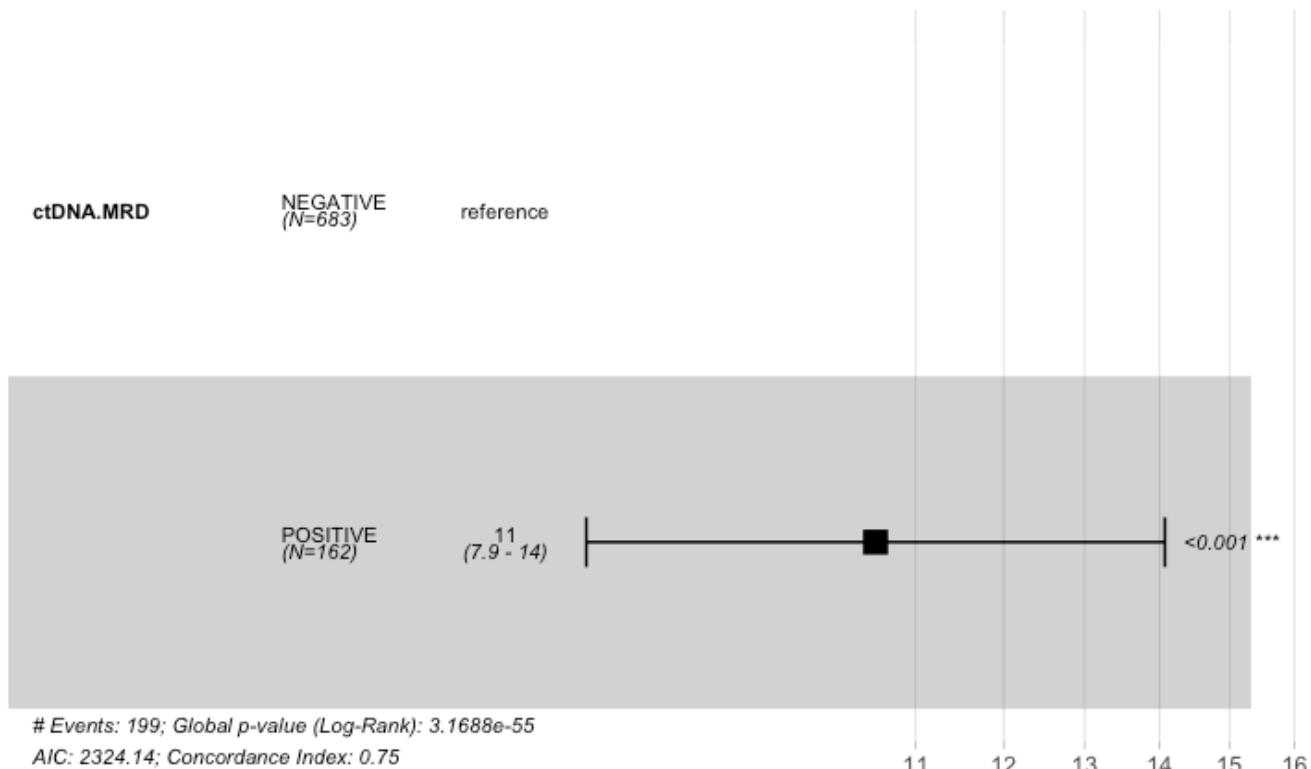
ctDNA.MRD=POSITIVE

CI	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
34	24.000	22.000	115.000	0.259	0.037	0.190	0.3

[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)

# 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"
```

```
#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_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)
```

```

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)
```

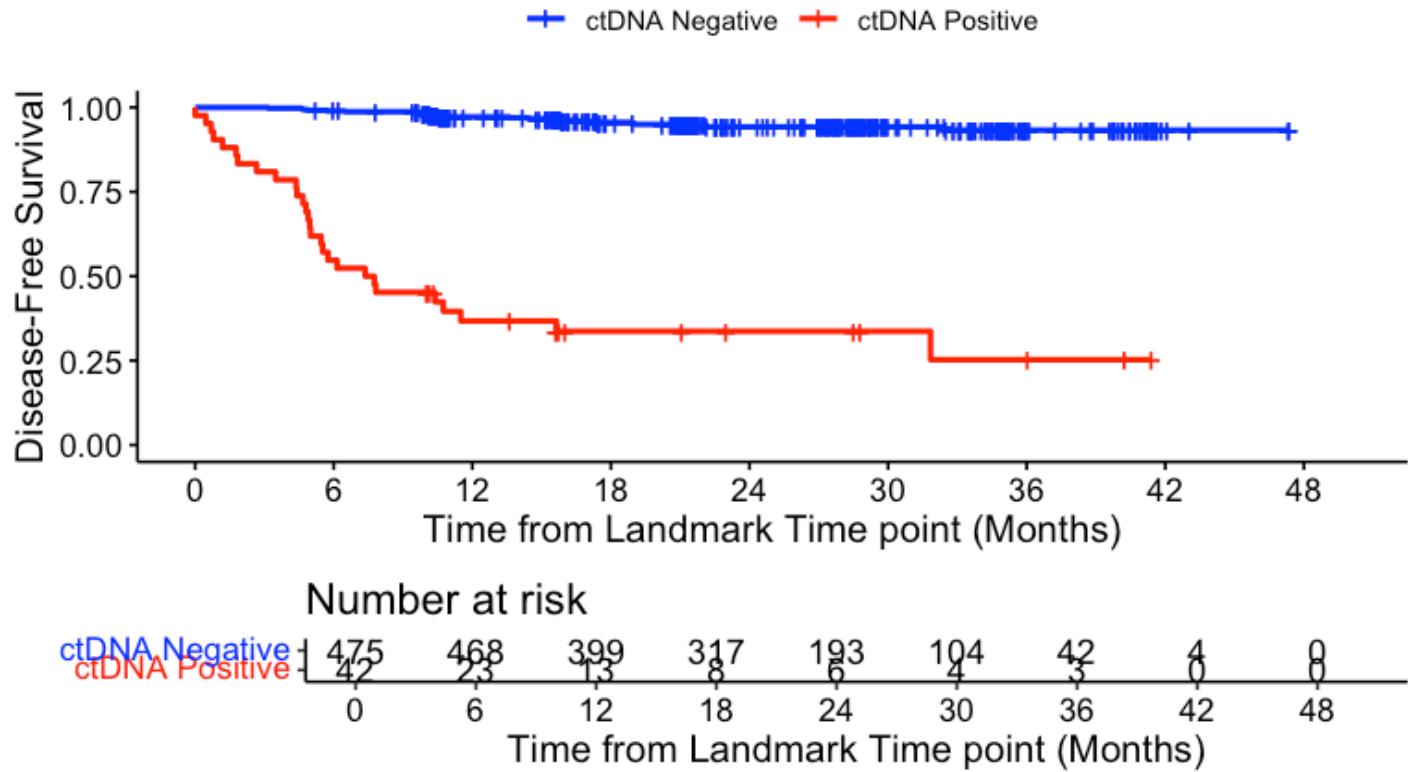
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	475	24	0.05052632	5.052632
POSITIVE	42	28	0.66666667	66.666667
NA	1481	NA	NA	NA

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="")
```

DFS - ctDNA MRD window | High Risk Stage II



```
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

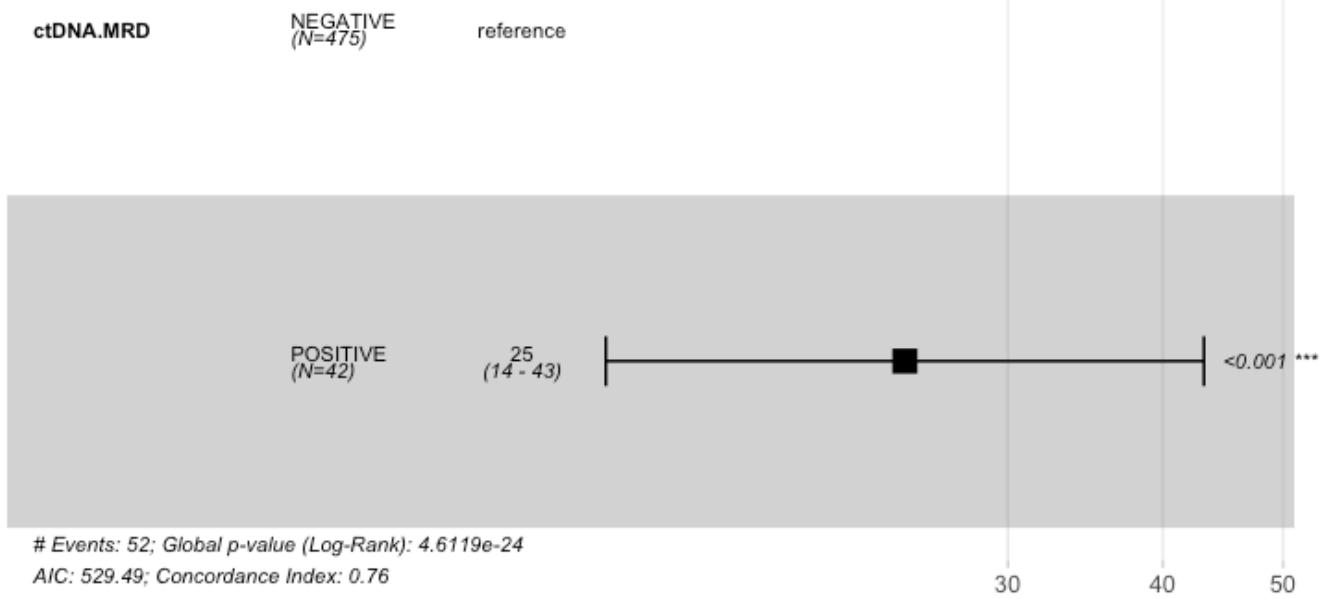
CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
62	24.000	193.000	23.000	0.942	0.012	0.914	0.9		

ctDNA.MRD=POSITIVE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
84	24.000	6.000	27.000	0.337	0.076	0.195	0.4		

```
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)
```

```
# 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"
```

```
#DFS by ctDNA at the MRD Window - High Risk 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$Risk.StageIII==TRUE,]
circ_data <- circ_data[circ_data$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)

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

3 rows

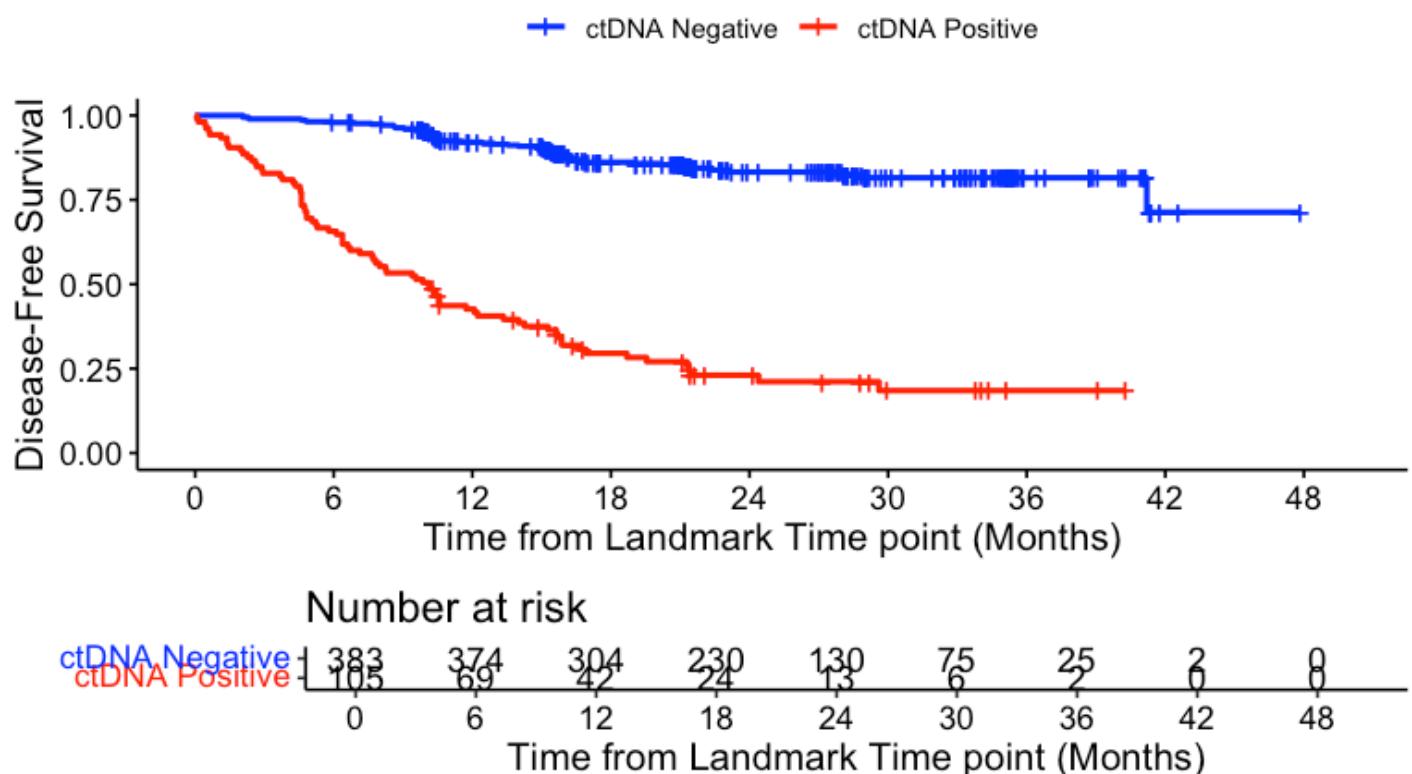
[Hide](#)

```

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



```
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

CI	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
05	24.0000	130.0000	53.0000	0.8322	0.0219	0.7842	0.87

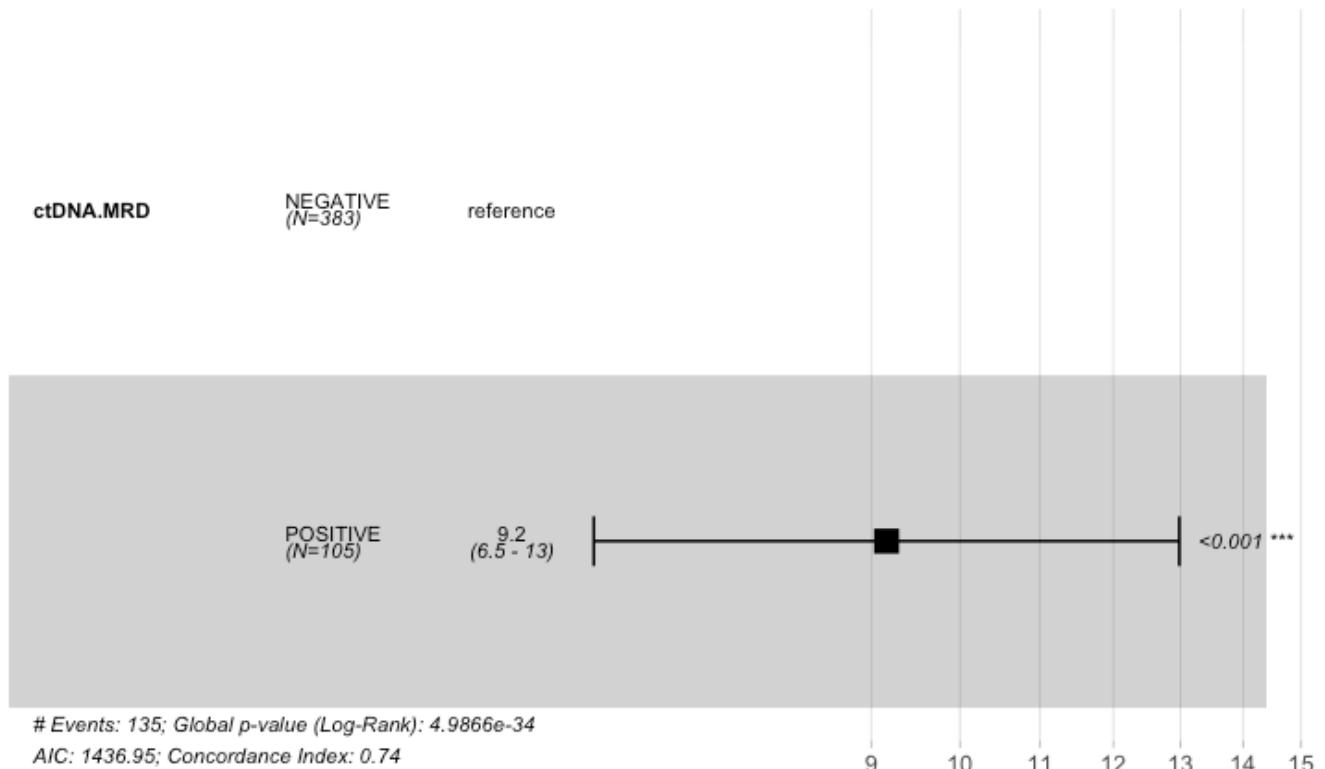
ctDNA.MRD=POSITIVE

CI	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
14	24.0000	13.0000	77.0000	0.2305	0.0443	0.1500	0.32

[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)

# 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"
```

```
#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](#)

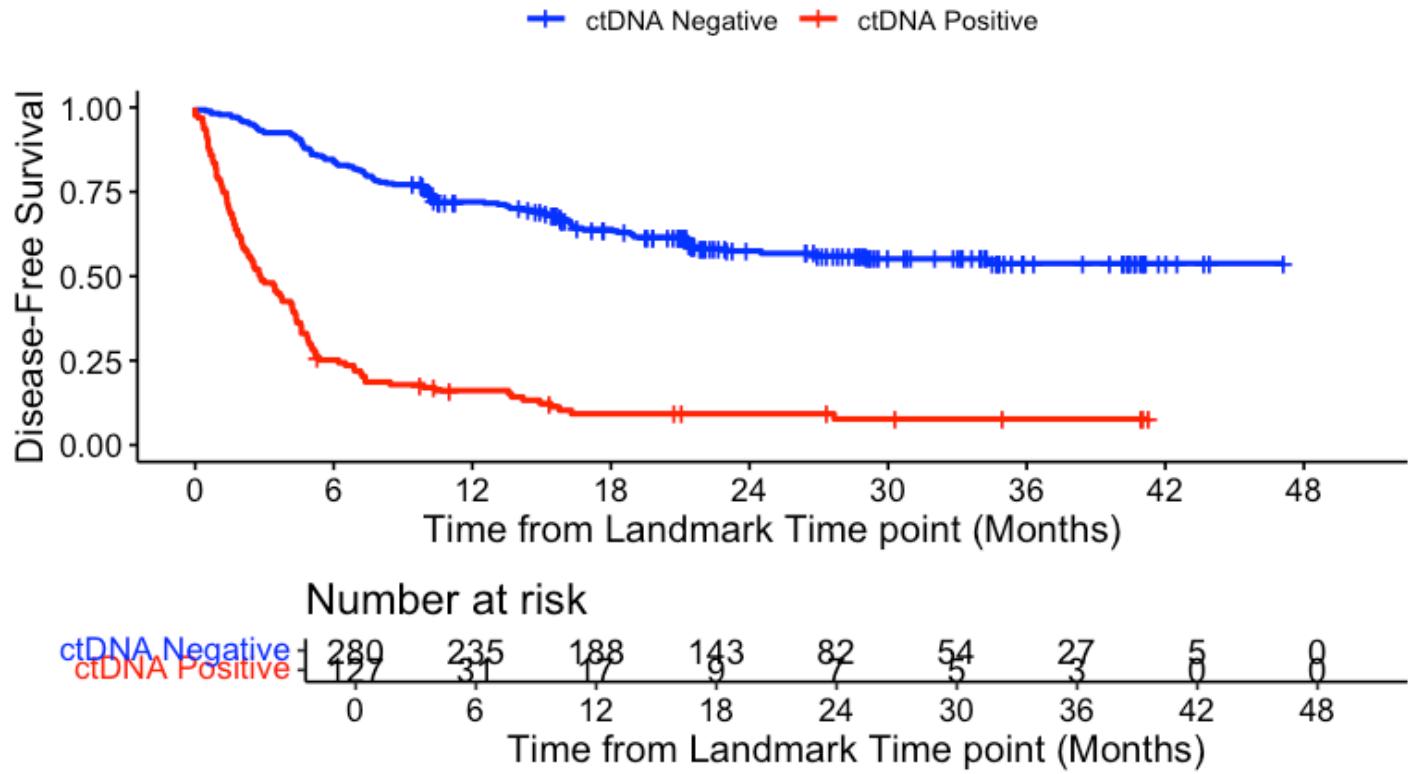
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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.lab
s=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage IV



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")
```

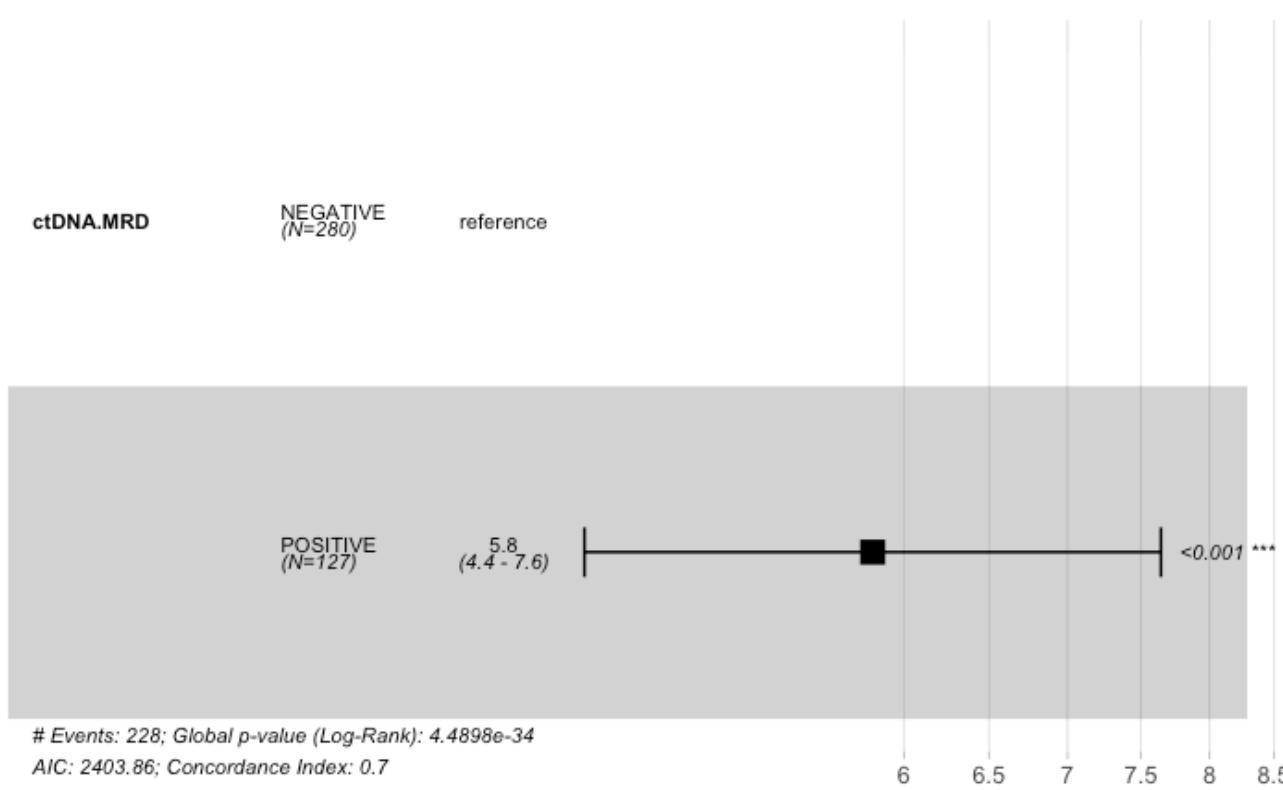
CI	ctDNA.MRD=NEGATIVE							
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper
24.0000	82.0000	110.0000	0.5748	0.0319	0.5097		0.63	

CI	ctDNA.MRD=POSITIVE							
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper
24.0000	7.0000	113.0000	0.0924	0.0274	0.0479		0.15	

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= 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

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 = 5.83 (4.44-7.64); p = 0"
```

```
#DFS by ctDNA at the MRD Window - Stage IV & NAC 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_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Stage.IV.NAC <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.IV.NAC = case_when(
    ctDNA.MRD == "NEGATIVE" & NAC == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & NAC == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & NAC == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & NAC == "FALSE" ~ 4
  ))
  
circ_data <- circ_data[circ_data$ctDNA.Stage.IV.NAC!="",]
survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.Stage.IV.NAC, data = circ_data)

```

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Stage.IV.NAC=1	146	67	29.08	18.07	NA
ctDNA.Stage.IV.NAC=2	61	60	2.00	1.45	2.99
ctDNA.Stage.IV.NAC=3	134	47	NA	34.30	NA
ctDNA.Stage.IV.NAC=4	66	54	4.47	2.76	5.26

Hide

```

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

```

ctDNA.Stage.IV.NAC	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	146	67	0.4589041	45.89041
2	61	60	0.9836066	98.36066
3	134	47	0.3507463	35.07463
4	66	54	0.8181818	81.81818

4 rows

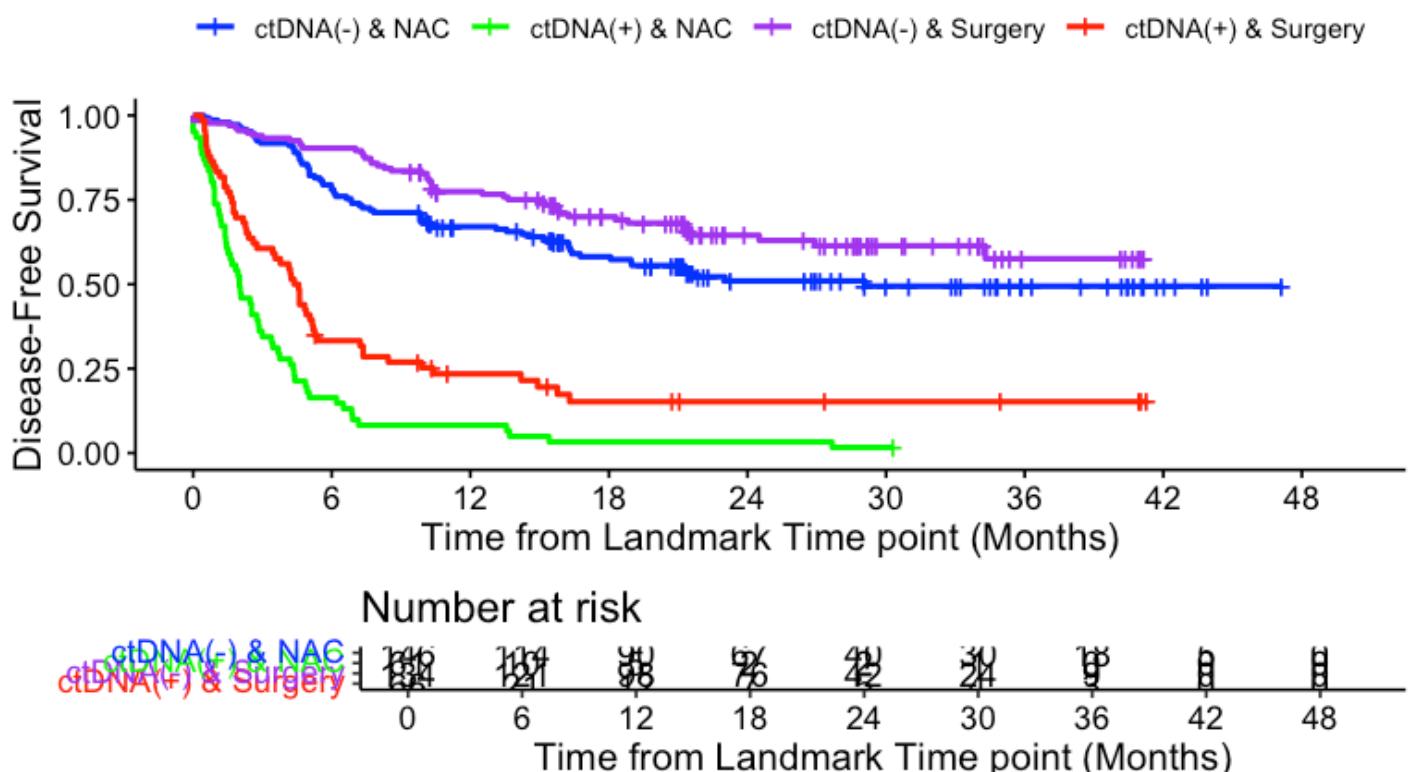
Hide

```

surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.IV.NAC, 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","green","purple", "red"), title="DFS - ctDNA MRD & Stage IV NAC", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & NAC", "ctDNA(+) & NAC", "ctDNA(-) & Surgery", "ctDNA(+) & Surgery"), legend.title="")

```

DFS - ctDNA MRD & Stage IV NAC



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

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

CI	ctDNA.Stage.IV.NAC=1						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
94	24.000	40.000	66.000	0.510	0.045	0.418	0.5
CI	ctDNA.Stage.IV.NAC=2						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
89	24.00000	2.00000	59.00000	0.03279	0.02280	0.00614	0.100
CI	ctDNA.Stage.IV.NAC=3						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
41	24.0000	42.0000	44.0000	0.6450	0.0442	0.5512	0.72
CI	ctDNA.Stage.IV.NAC=4						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
58	24.000	5.000	54.000	0.152	0.048	0.073	0.2

Hide

```
circ_data$ctDNA.Stage.IV.NAC <- factor(circ_data$ctDNA.Stage.IV.NAC, levels=c
("1", "2", "3", "4"), labels = c("ctDNA(-) & NAC", "ctDNA(+) & NAC", "ctDNA(-) & Surgery",
"ctDNA(+) & Surgery"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.IV.NAC, data=circ_data)
summary(cox_fit)
```

Call:

```
coxph(formula = surv_object ~ ctDNA.Stage.IV.NAC, data = circ_data)
```

```
n= 407, number of events= 228
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.IV.NACctDNA(+) & NAC	2.0324	7.6324	0.1860	10.929	< 2e-16 ***
ctDNA.Stage.IV.NACctDNA(-) & Surgery	-0.3717	0.6896	0.1904	-1.952	0.0509 .
ctDNA.Stage.IV.NACctDNA(+) & Surgery	1.2922	3.6409	0.1853	6.975	3.06e-12 ***

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Stage.IV.NACctDNA(+) & NAC	7.6324	0.1310	5.3011	10.989
ctDNA.Stage.IV.NACctDNA(-) & Surgery	0.6896	1.4502	0.4748	1.001
ctDNA.Stage.IV.NACctDNA(+) & Surgery	3.6409	0.2747	2.5323	5.235

```
Concordance= 0.726 (se = 0.016 )
```

```
Likelihood ratio test= 167 on 3 df, p=<2e-16
```

```
Wald test = 187.1 on 3 df, p=<2e-16
```

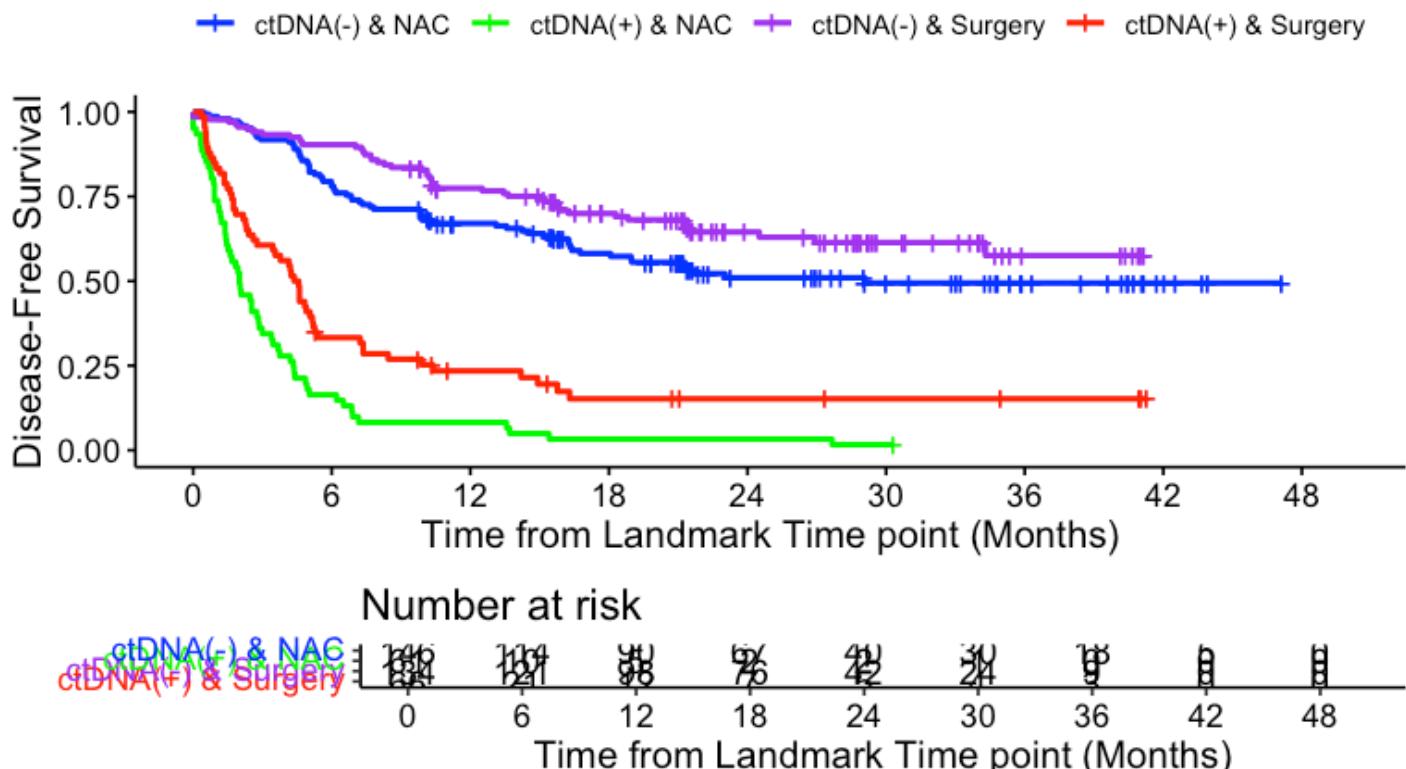
```
Score (logrank) test = 244 on 3 df, p=<2e-16
```

Hide

```
#Repeat analysis to compare ctDNA MRD (-) vs (+) in T4N0
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$ctDNA.Stage.IV.NAC <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.IV.NAC = case_when(
    ctDNA.MRD == "NEGATIVE" & NAC == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & NAC == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & NAC == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & NAC == "FALSE" ~ 4
  ))
circ_data <- circ_data[circ_data$ctDNA.Stage.IV.NAC!="",]
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.IV.NAC, 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","green","purple", "red"), title="DFS - ctDNA MRD & Stage IV NAC", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & NAC", "ctDNA(+) & NAC", "ctDNA(-) & Surgery", "ctDNA(+) & Surgery"), legend.title="")
```

DFS - ctDNA MRD & Stage IV NAC



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

Hide

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

CI	ctDNA.Stage.IV.NAC=1						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
94	24.000	40.000	66.000	0.510	0.045	0.418	0.5
CI	ctDNA.Stage.IV.NAC=2						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
89	24.00000	2.00000	59.00000	0.03279	0.02280	0.00614	0.100
CI	ctDNA.Stage.IV.NAC=3						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
41	24.0000	42.0000	44.0000	0.6450	0.0442	0.5512	0.72
CI	ctDNA.Stage.IV.NAC=4						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
58	24.000	5.000	54.000	0.152	0.048	0.073	0.2

Hide

```
circ_data$ctDNA.Stage.IV.NAC <- factor(circ_data$ctDNA.Stage.IV.NAC, levels=c("2","4","1","3"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.IV.NAC, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.Stage.IV.NAC, data = circ_data)

n= 407, number of events= 228

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.IV.NAC4	-0.74018	0.47703	0.18960	-3.904	9.47e-05 ***
ctDNA.Stage.IV.NAC1	-2.03240	0.13102	0.18596	-10.929	< 2e-16 ***
ctDNA.Stage.IV.NAC3	-2.40410	0.09035	0.20367	-11.804	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Stage.IV.NAC4	0.47703	2.096	0.32897	0.6917
ctDNA.Stage.IV.NAC1	0.13102	7.632	0.09100	0.1886
ctDNA.Stage.IV.NAC3	0.09035	11.068	0.06061	0.1347

Concordance= 0.726 (se = 0.016)

Likelihood ratio test= 167 on 3 df, p=<2e-16

Wald test = 187.1 on 3 df, p=<2e-16

Score (logrank) test = 244 on 3 df, p=<2e-16

#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)
```

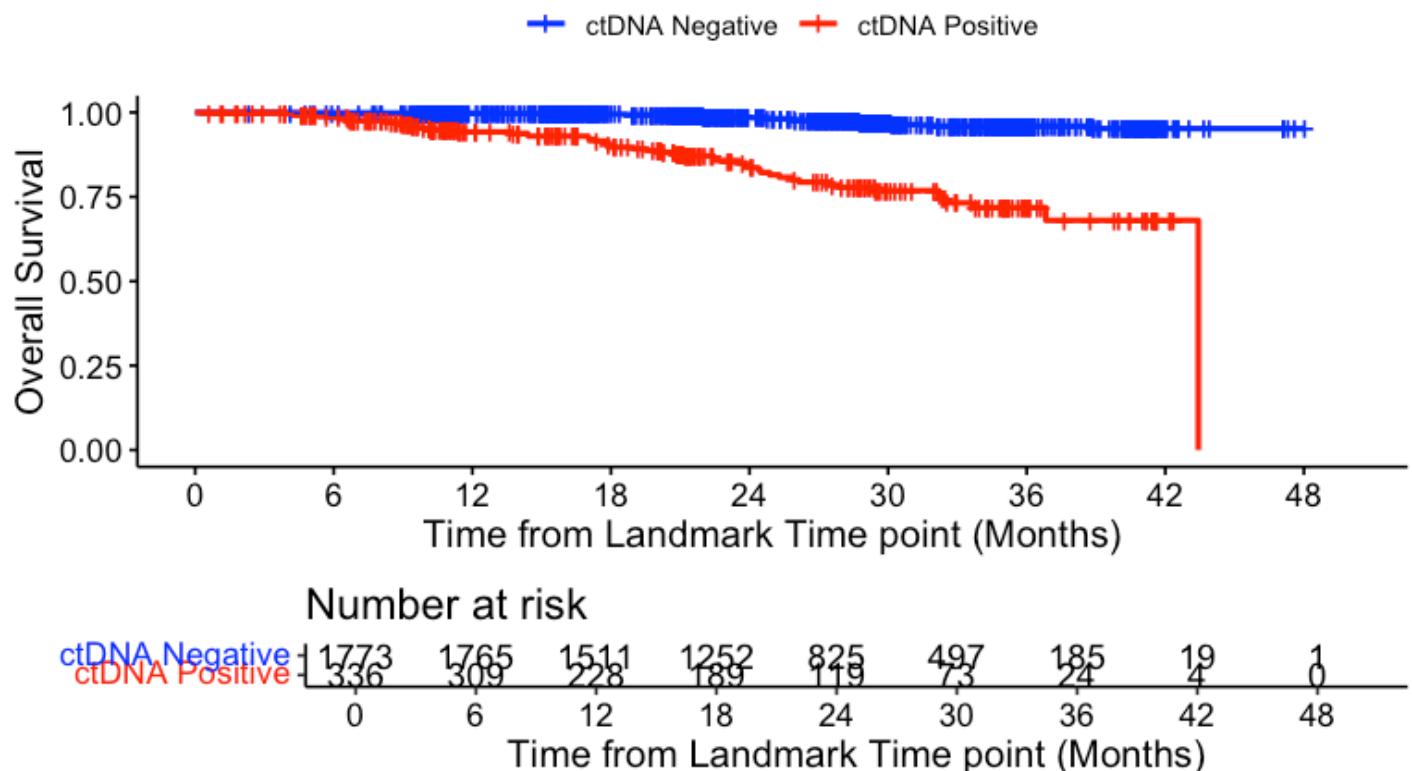
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1773	36	0.02030457	2.030457
POSITIVE	336	52	0.15476190	15.476190
NA	1	NA	NA	NA

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



```
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	

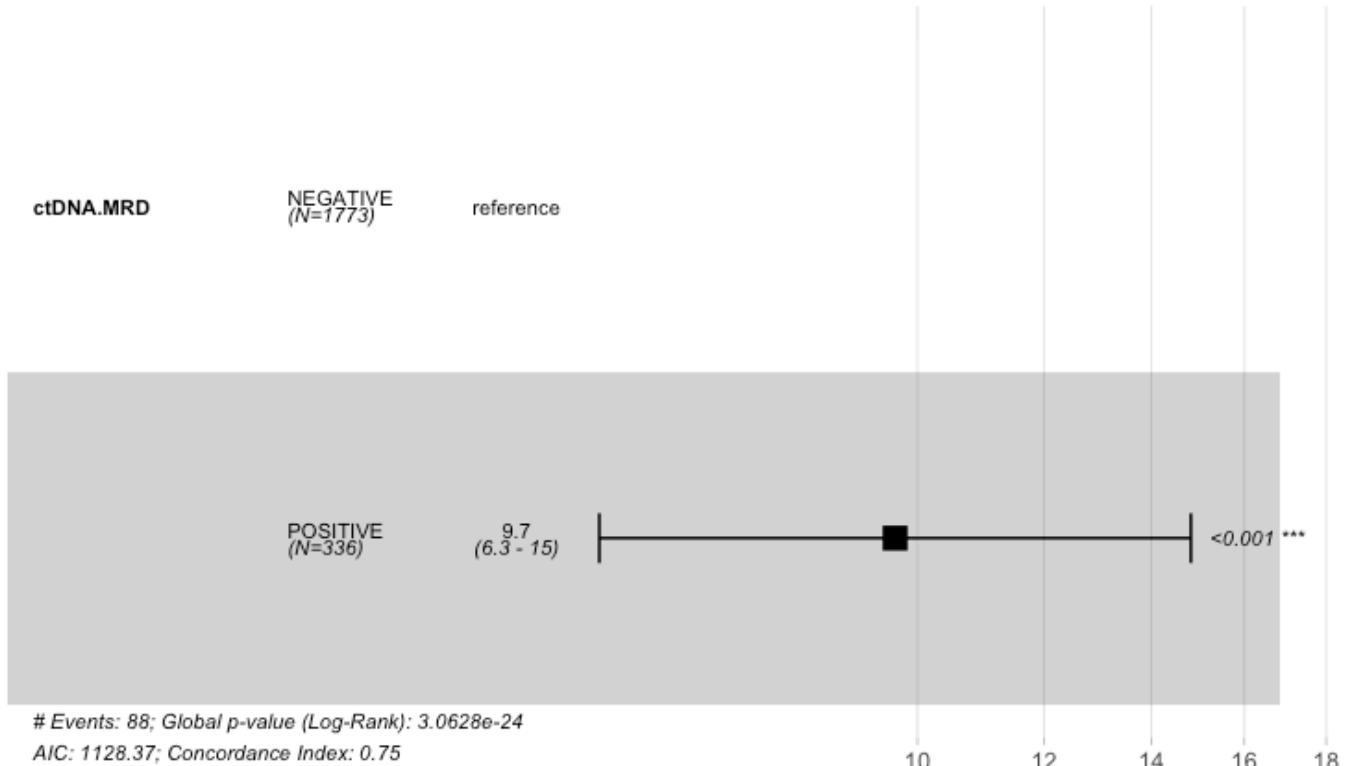
ctDNA.MRD=POSITIVE

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)

# 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"
```

```
#OS by ctDNA at the MRD Window - Stages High Risk II/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$HighRisk.Stage=="TRUE",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_datadf <- 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)

17 observations deleted due to missingness
      n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 1158     19      NA      NA      NA
ctDNA.MRD=POSITIVE  204     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)
```

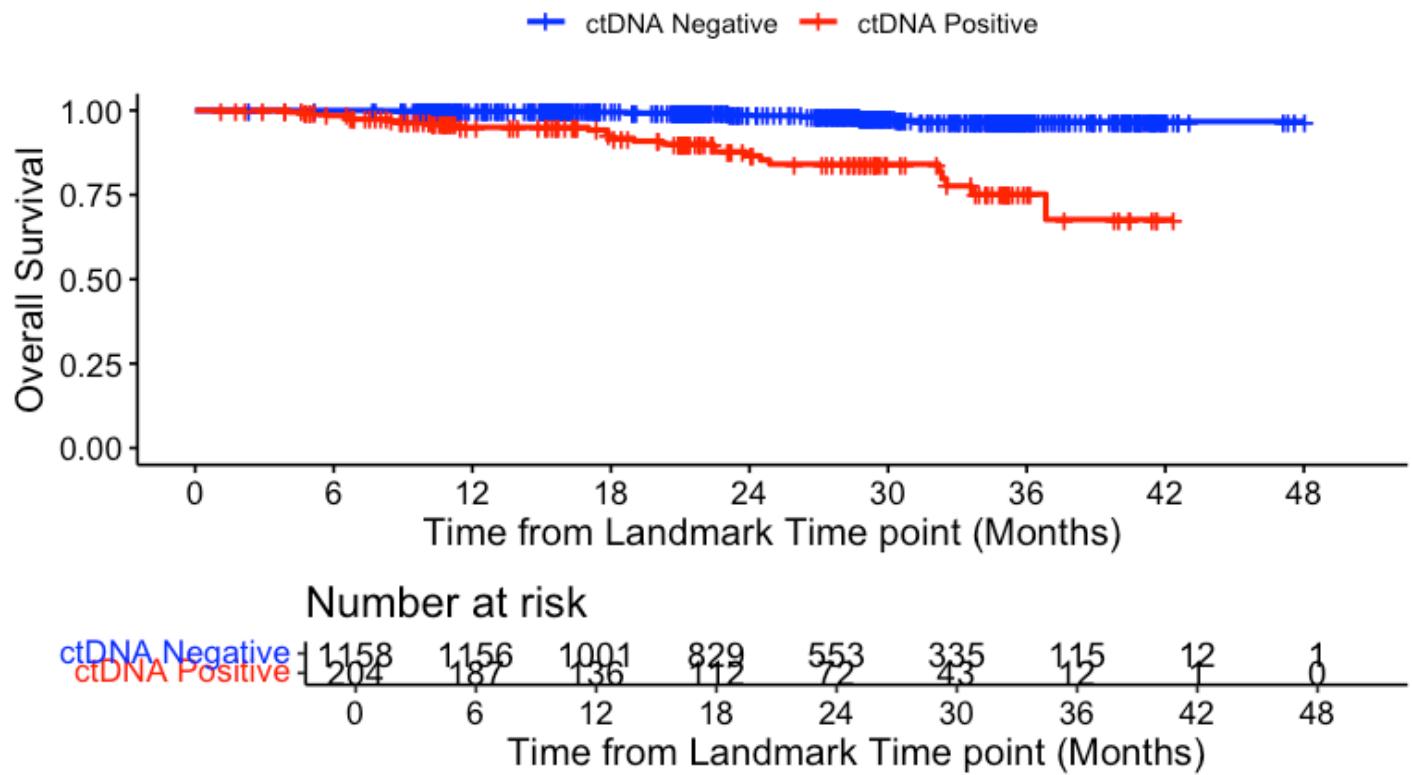
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1158	19	0.0164076	1.64076
POSITIVE	204	25	0.1225490	12.25490
NA	17	NA	NA	NA

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 | High Risk Stage II-III", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - ctDNA MRD window | High Risk Stage II-III



```
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")
```

17 observations deleted due to missingness

ctDNA.MRD=NEGATIVE

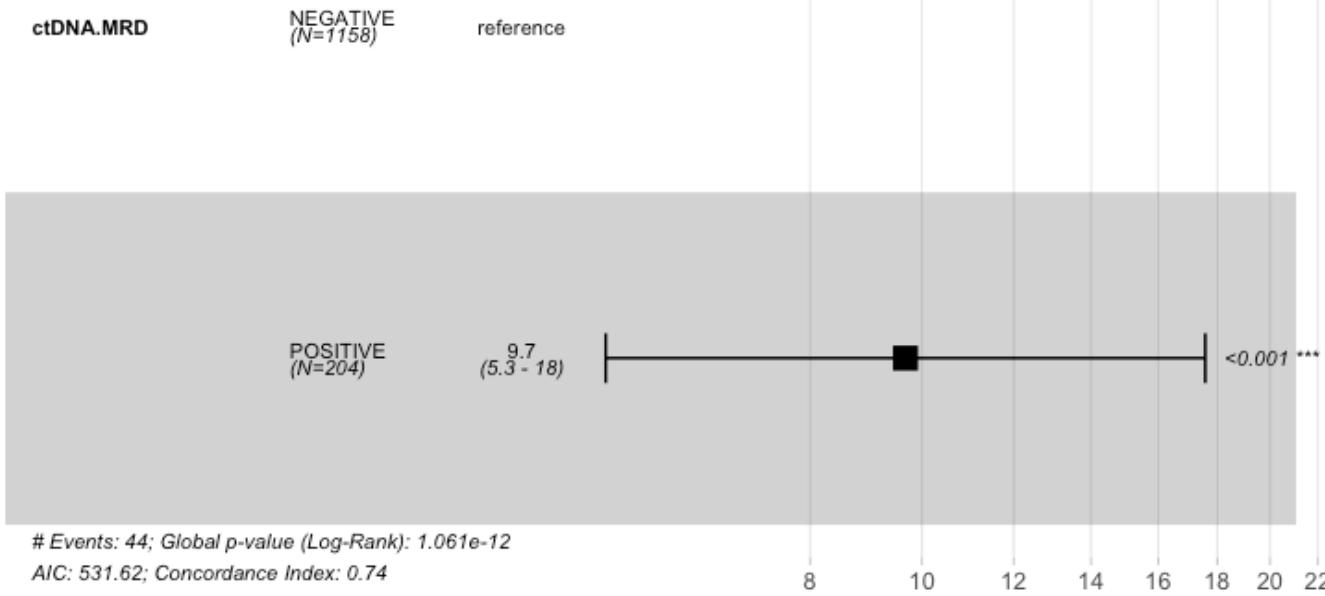
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	553	11	0.986	0.00427	0.975	0.992		
30	335	6	0.974	0.00668	0.957	0.984		
36	115	2	0.967	0.00798	0.947	0.980		

ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	72	18	0.866	0.0311	0.791	0.915		
30	43	2	0.841	0.0348	0.758	0.897		
36	12	4	0.752	0.0524	0.631	0.838		

```
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= 1362, number of events= 44
 (17 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)						
ctDNA.MRDPOSITIVE	2.2696	9.6755	0.3047	7.45	9.37e-14 ***						

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.MRDPOSITIVE	9.675	0.1034	5.325	17.58

Concordance= 0.742 (se = 0.038)
 Likelihood ratio test= 50.73 on 1 df, p=1e-12
 Wald test = 55.5 on 1 df, p=9e-14
 Score (logrank) test = 83.66 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 = 9.68 (5.33-17.58); p = 0"
```

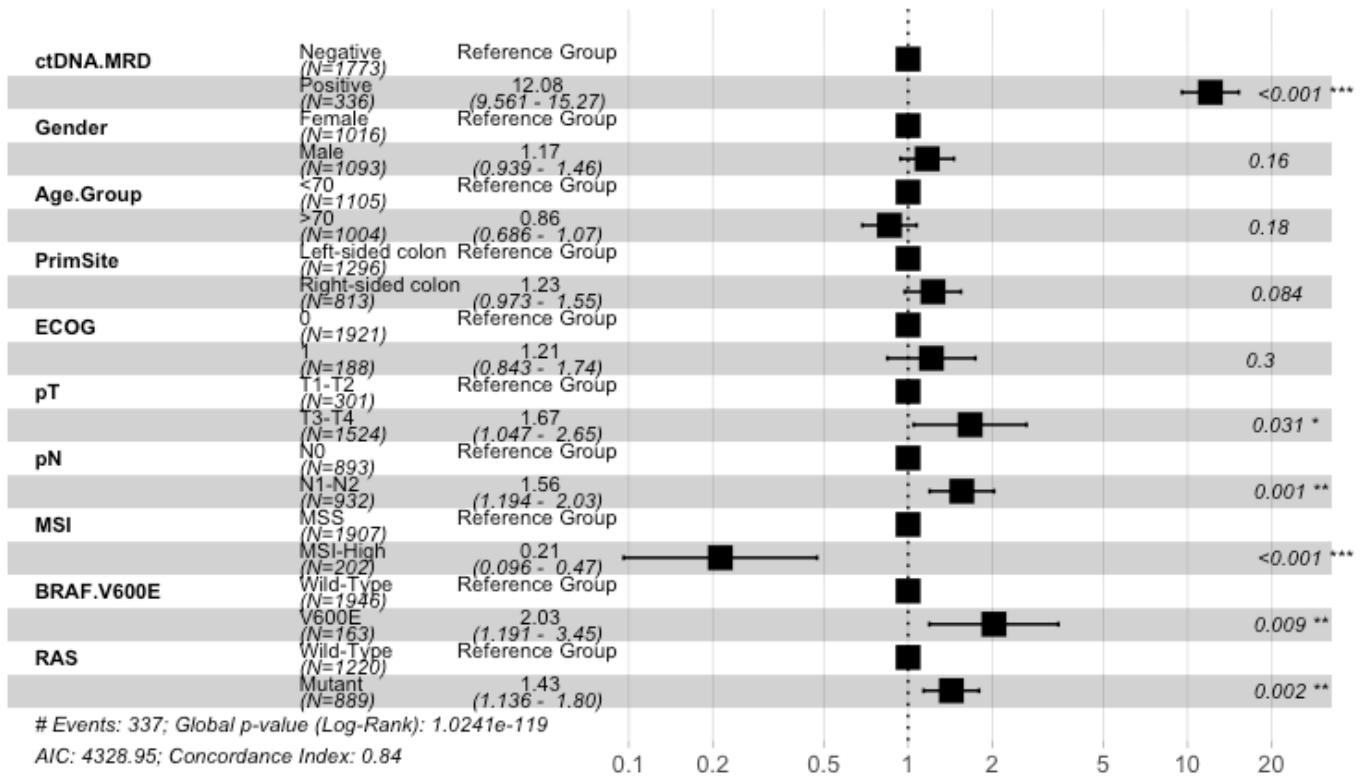
```
#Multivariate cox regression at MRD Window for DFS - All stages Landmark MRD timepoint
```

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



Hide

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

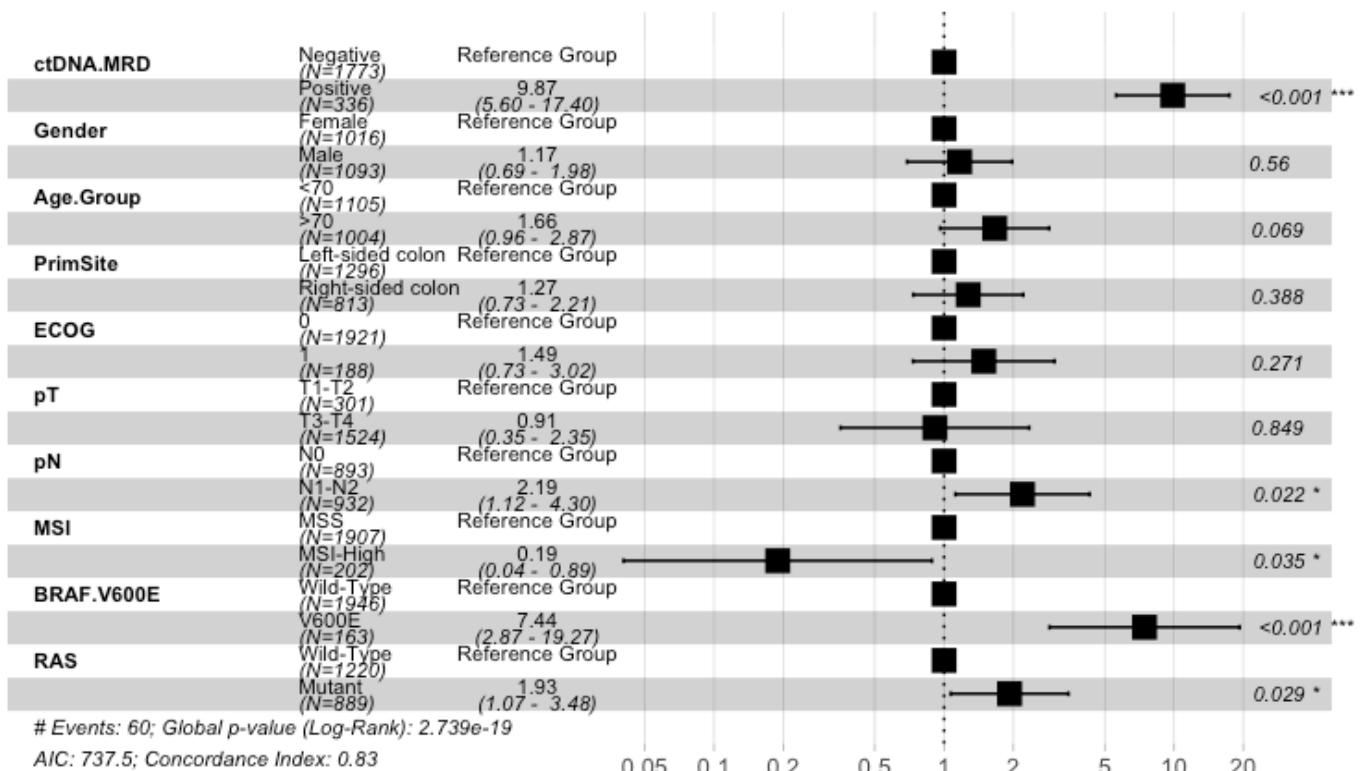
```
#Multivariate cox regression at MRD Window for OS - All stages Landmark MRD timepoint
```

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)
```

```
#MRD Window - Sensitivity and Specificity calculations - All Cohorts
```

Hide

```
#All Patients
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.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

```
[1] "Sensitivity - All pts: 54.6583850931677"
```

[Hide](#)

```
print(paste("Specificity - All pts: ", specificity*100))
```

```
[1] "Specificity - All pts: 95.5132145052243"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - All pts: 78.3382789317507"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - All pts: 87.6480541455161"
```

[Hide](#)

```
#Stage I Patients
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=="I",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage I: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage I: 25"
```

[Hide](#)

```
print(paste("Specificity - Stage I: ", specificity*100))
```

```
[1] "Specificity - Stage I: 100"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage I: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage I: 100"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage I: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage I: 97.3451327433628"
```

[Hide](#)

```
#Stage II Patients
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=="II",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 53.5714285714286"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 97.3821989528796"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 66.6666666666667"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 95.5479452054795"
```

[Hide](#)

```
#Stage III Patients
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=="III",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 60.5128205128205"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 93.0875576036866"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 72.3926380368098"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 88.7262079062958"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 59.8360655737705"
```

[Hide](#)

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 94.7274352100089"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 71.219512195122"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 91.5371329879102"
```

[Hide](#)

```
#Stage IV Patients
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$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

[1] "Sensitivity - Stage IV: 50.8928571428571"

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

[1] "Specificity - Stage IV: 92.896174863388"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - Stage IV: 89.763779527559"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - Stage IV: 60.7142857142857"

#MRD Window - Sensitivity and Specificity calculations - no ACT treated

[Hide](#)

```
#All Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

```
[1] "Sensitivity - All pts: 51.219512195122"
```

[Hide](#)

```
print(paste("Specificity - All pts: ", specificity*100))
```

```
[1] "Specificity - All pts: 99.2864424057085"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - All pts: 95.4545454545455"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - All pts: 87.4326750448833"
```

[Hide](#)

```
#Stage I Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="I",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage I: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage I: 28.5714285714286"
```

[Hide](#)

```
print(paste("Specificity - Stage I: ", specificity*100))
```

```
[1] "Specificity - Stage I: 100"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage I: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage I: 100"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage I: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage I: 97.777777777778"
```

[Hide](#)

```
#Stage II Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="II",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 46.5116279069767"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 99.3406593406593"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 86.9565217391304"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 95.1578947368421"
```

[Hide](#)

```
#Stage III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="III",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 59.0909090909091"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 98.9417989417989"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 95.1219512195122"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 87.3831775700935"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 55.8823529411765"
```

Hide

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 99.2673992673993"
```

Hide

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 93.4426229508197"
```

Hide

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 92.3339011925043"
```

Hide

```
#Stage IV Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="IV",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage IV: 50.2923976608187"
```

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

```
[1] "Specificity - Stage IV: 98.2905982905983"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage IV: 97.7272727272727"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage IV: 57.5"
```

```
#MRD Window - Sensitivity and Specificity calculations - ACT treated
```

[Hide](#)

```
#All Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

```
[1] "Sensitivity - All pts: 59.6938775510204"
```

[Hide](#)

```
print(paste("Specificity - All pts: ", specificity*100))
```

```
[1] "Specificity - All pts: 89.7832817337461"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - All pts: 63.9344262295082"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - All pts: 88.0121396054628"
```

[Hide](#)

```
#Stage II Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="II",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 76.9230769230769"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 89.8305084745763"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 45.4545454545455"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 97.2477064220184"
```

[Hide](#)

```
#Stage III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="III",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 61.2403100775194"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 90.6926406926407"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 64.7540983606557"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 89.3390191897655"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 62.6760563380282"
```

Hide

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 90.4013961605585"
```

Hide

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 61.8055555555556"
```

Hide

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 90.7180385288967"
```

Hide

```
#Stage IV Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="IV",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE","TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.MRD, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

[1] "Sensitivity - Stage IV: 52.8301886792453"

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

[1] "Specificity - Stage IV: 83.3333333333333"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - Stage IV: 71.7948717948718"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - Stage IV: 68.75"

#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)
```

```
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)
```

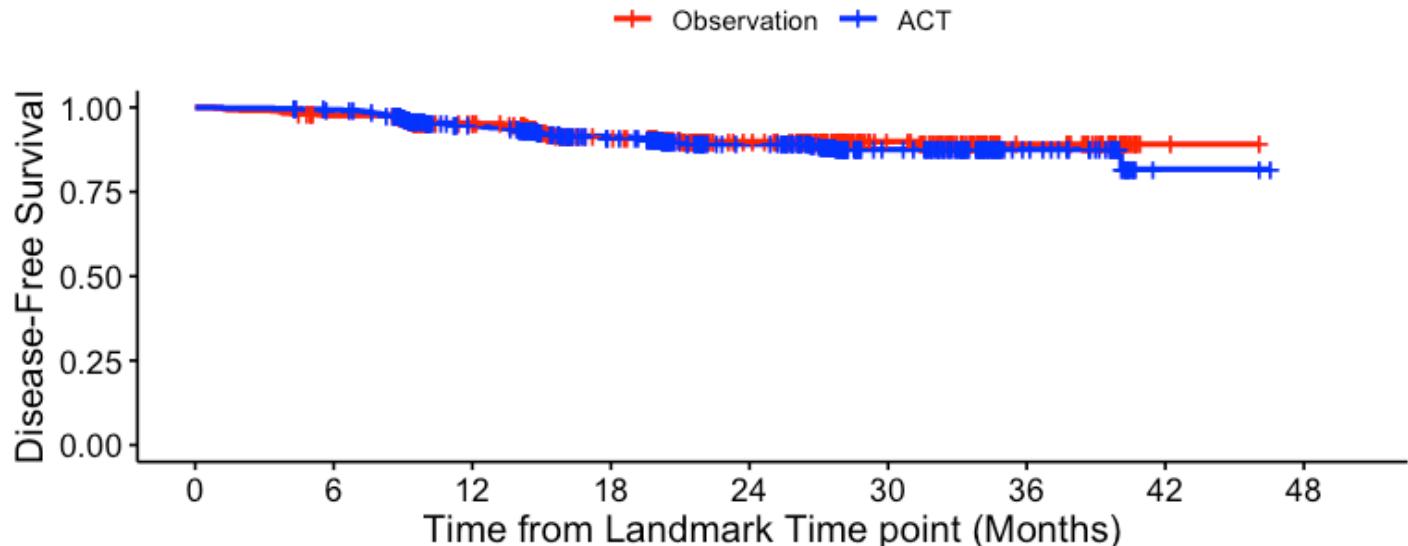
ACT <lgI>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	586	50	0.08532423	8.532423
TRUE	571	55	0.09632224	9.632224
NA	15	NA	NA	NA

3 rows

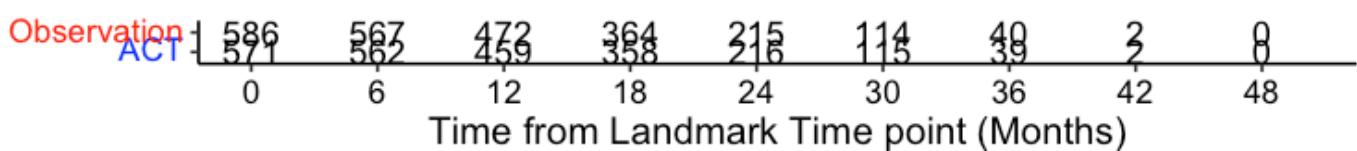
[Hide](#)

```
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



Number at risk



```
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

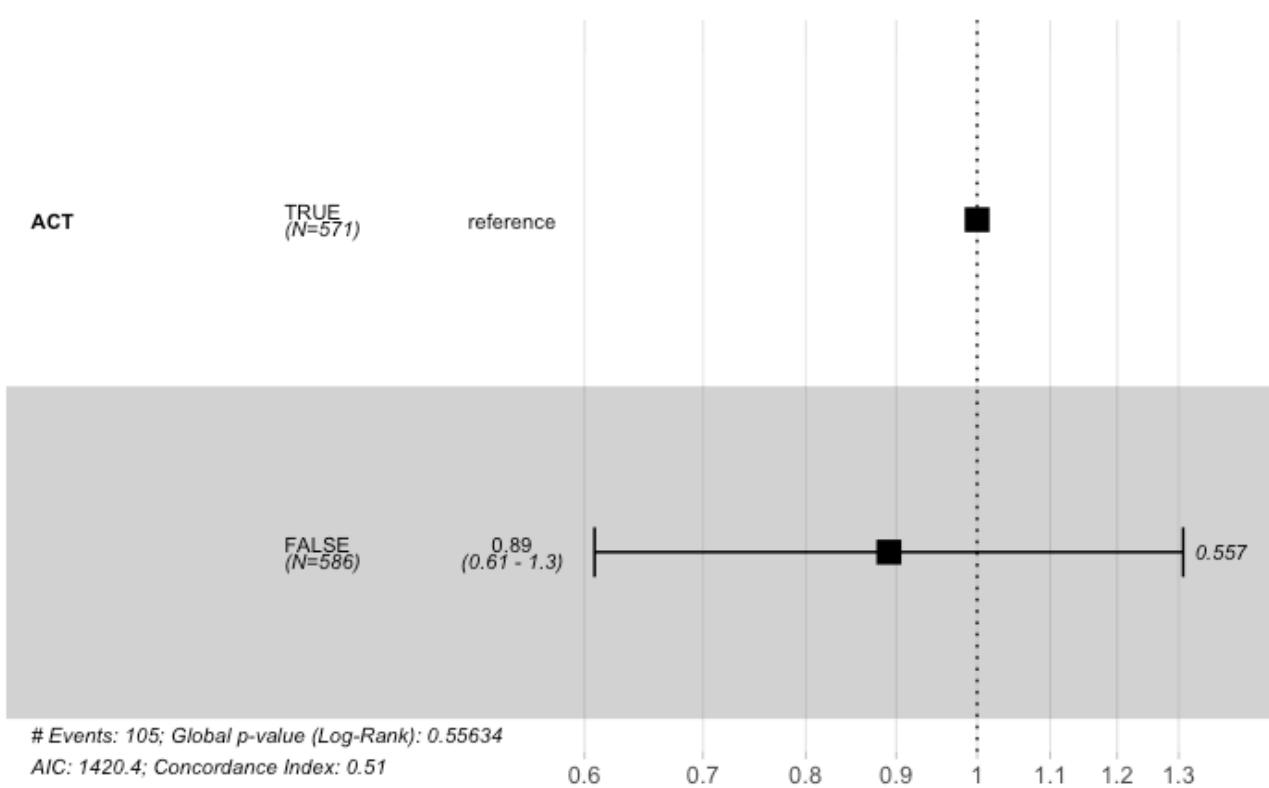
CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
23	24.000	215.000	49.000	0.899	0.014	0.868	0.9		

ACT=TRUE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
68	24.0000	216.0000	51.0000	0.8911	0.0148	0.8581	0.91		

```
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

```
#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_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$pN <- factor(circ_data$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_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$pN <- factor(circ_data$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

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$ctDNA.MRD=="POSITIVE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_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)

1 observation deleted due to missingness
 n events median 0.95LCL 0.95UCL
 ACT=FALSE 49 45 3.62 3.26 4.01
 ACT=TRUE 143 88 11.86 9.30 18.57

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	49	45	0.9183673	91.83673
TRUE	143	88	0.6153846	61.53846
NA	1	NA	NA	NA

3 rows

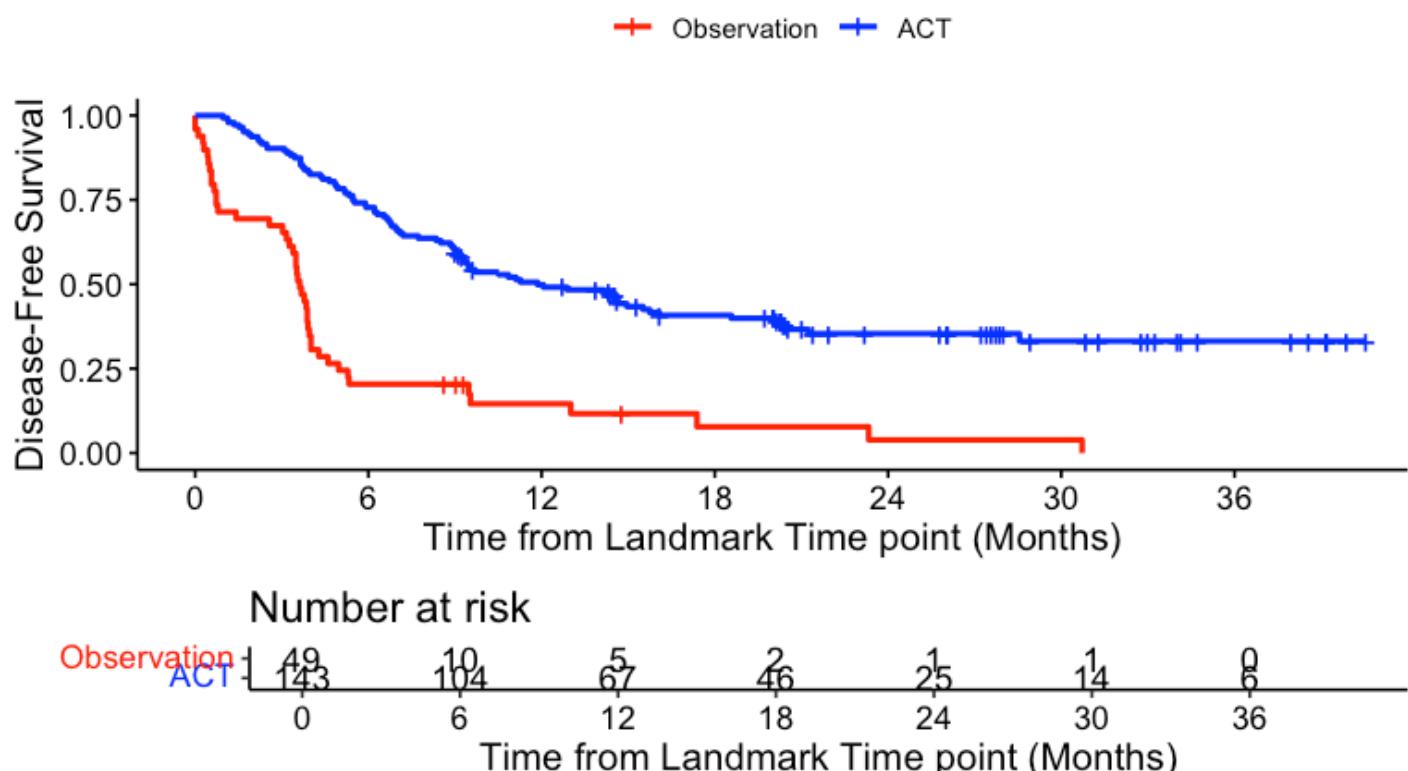
Hide

```

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 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")
```

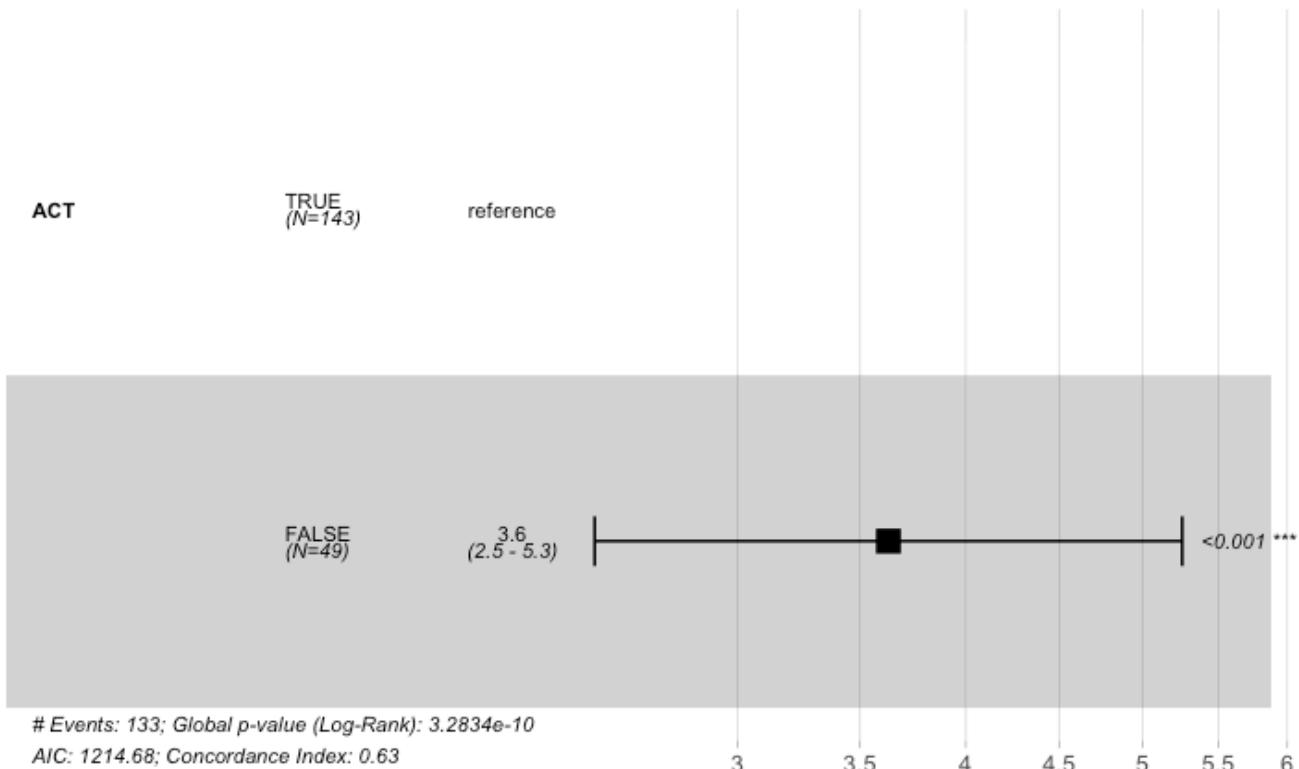
1 observation deleted due to missingness

CI	ACT=FALSE						
	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
74	24.00000	1.00000	44.00000	0.03887	0.03591	0.00344	0.155
CI	ACT=TRUE						
	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
83	24.0000	25.0000	87.0000	0.3536	0.0434	0.2698	0.43

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

      coef exp(coef)  se(coef)    z Pr(>|z|)
ACTFALSE 1.289     3.628    0.189 6.817 9.29e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     3.628     0.2756    2.505    5.255

Concordance= 0.628  (se = 0.019 )
Likelihood ratio test= 39.5  on 1 df,  p=3e-10
Wald test          = 46.47  on 1 df,  p=9e-12
Score (logrank) test = 52.59  on 1 df,  p=4e-13

```

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 = 3.63 (2.5-5.26); p = 0"
```

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-side 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	1.3294241	3.7788665	0.2058337	6.459	1.06e-10 ***
GenderMale	-0.0019182	0.9980836	0.1817332	-0.011	0.992
Age.Group≥70	-0.0005564	0.9994437	0.1873391	-0.003	0.998
StageIII	0.3480780	1.4163427	0.2364217	1.472	0.141
ECOG1	0.1423411	1.1529698	0.2786980	0.511	0.610

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	3.7789	0.2646	2.5244	5.657
GenderMale	0.9981	1.0019	0.6990	1.425
Age.Group≥70	0.9994	1.0006	0.6923	1.443
StageIII	1.4163	0.7060	0.8911	2.251
ECOG1	1.1530	0.8673	0.6677	1.991

Concordance= 0.65 (se = 0.025)

Likelihood ratio test= 42.51 on 5 df, p=5e-08

Wald test = 50.13 on 5 df, p=1e-09

Score (logrank) test = 56.32 on 5 df, p=7e-11

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_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$pN <- factor(circ_data$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	-1.3294241	0.2646296	0.2058337	-6.459	1.06e-10	***
GenderMale	-0.0019182	0.9980836	0.1817332	-0.011	0.992	
Age.Group≥70	-0.0005564	0.9994437	0.1873391	-0.003	0.998	
StageIII	0.3480780	1.4163427	0.2364217	1.472	0.141	
ECOG1	0.1423411	1.1529698	0.2786980	0.511	0.610	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTTRUE	0.2646	3.7789	0.1768	0.3961
GenderMale	0.9981	1.0019	0.6990	1.4251
Age.Group≥70	0.9994	1.0006	0.6923	1.4429
StageIII	1.4163	0.7060	0.8911	2.2512
ECOG1	1.1530	0.8673	0.6677	1.9909

Concordance= 0.65 (se = 0.025)

Likelihood ratio test= 42.51 on 5 df, p=5e-08

Wald test = 50.13 on 5 df, p=1e-09

Score (logrank) test = 56.32 on 5 df, p=7e-11

#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_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)

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 <lgl>	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

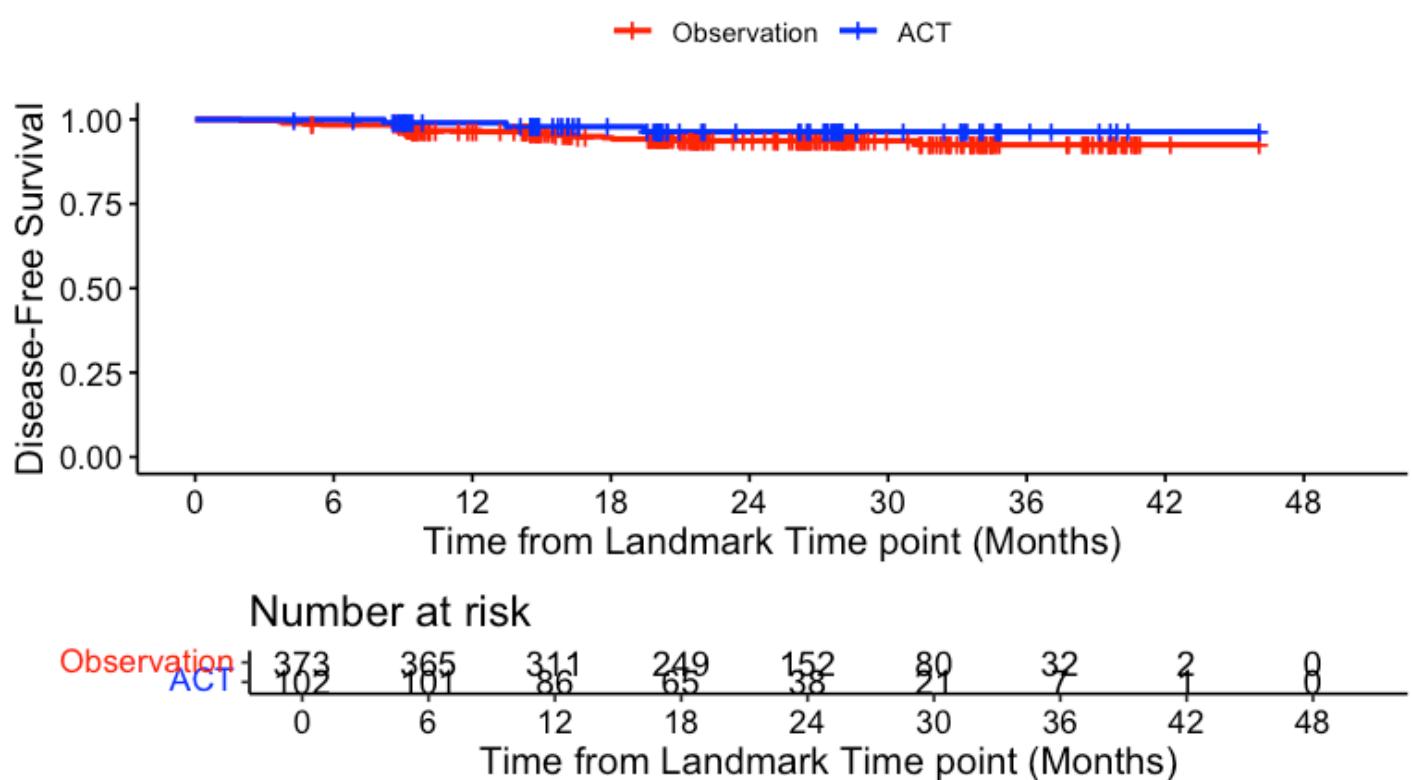
Hide

```

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=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="")

```

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")

1588 observations deleted due to missingness

ACT=FALSE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
59	24.000	152.000	20.000	0.937	0.014	0.903	0.9		

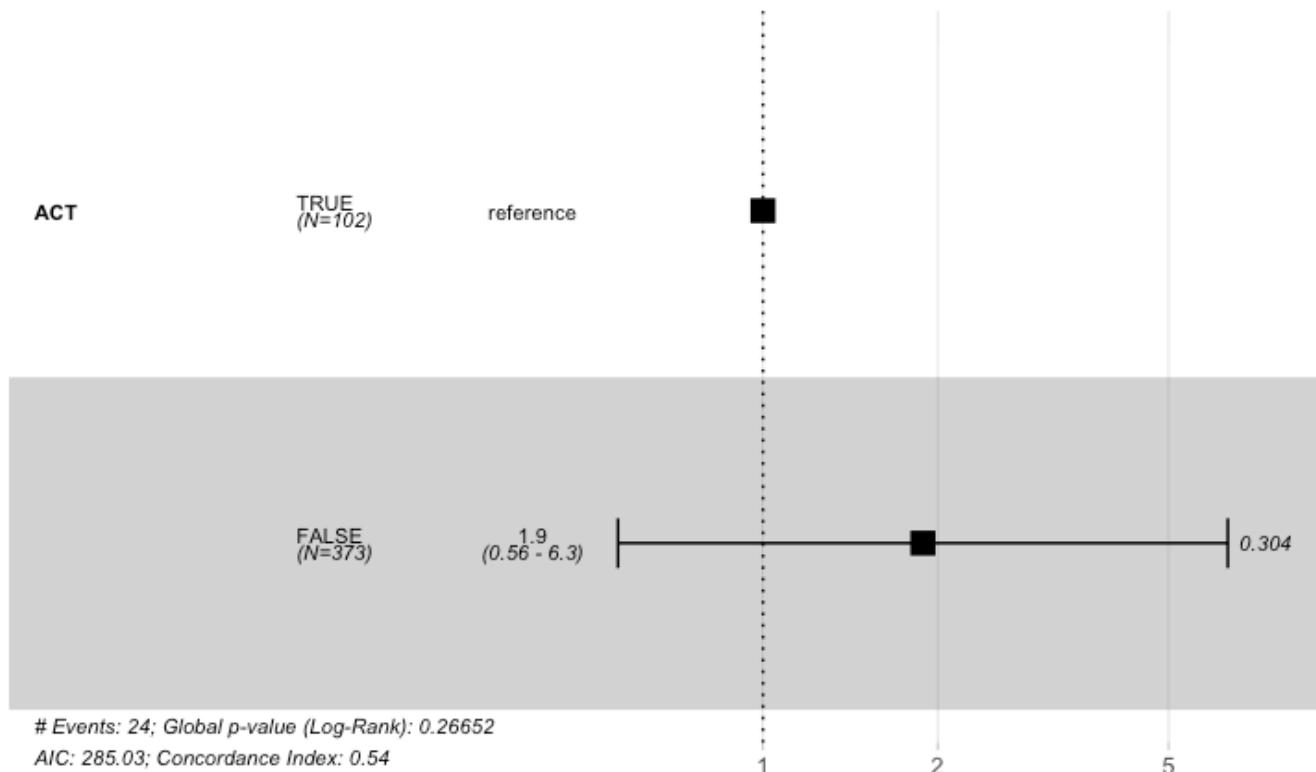
ACT=TRUE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
83	24.0000	38.0000	3.0000	0.9634	0.0211	0.8890	0.98		

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= 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

```
#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_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= 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

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$ctDNA.MRD=="POSITIVE",]
circ_data$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 16 14 3.7 3.39 NA

ACT=TRUE 22 10 NA 9.30 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 <lgl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	16	14	0.8750000	87.50000
TRUE	22	10	0.4545455	45.45455
NA	1588	NA	NA	NA

3 rows

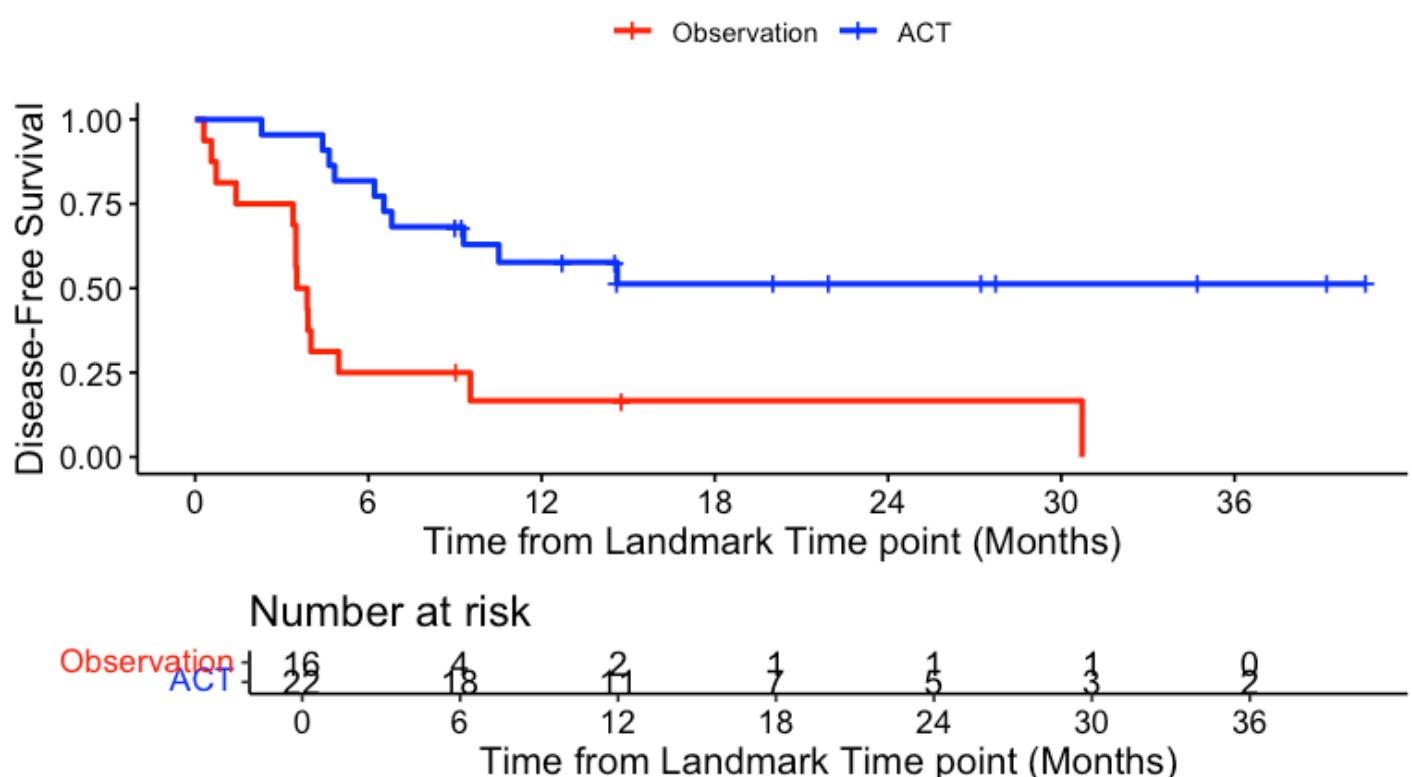
Hide

```

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=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="")

```

DFS - ctDNA MRD Positive 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")

1588 observations deleted due to missingness

ACT=FALSE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
28	24.0000	1.0000	13.0000	0.1667	0.0992	0.0322		0.39	

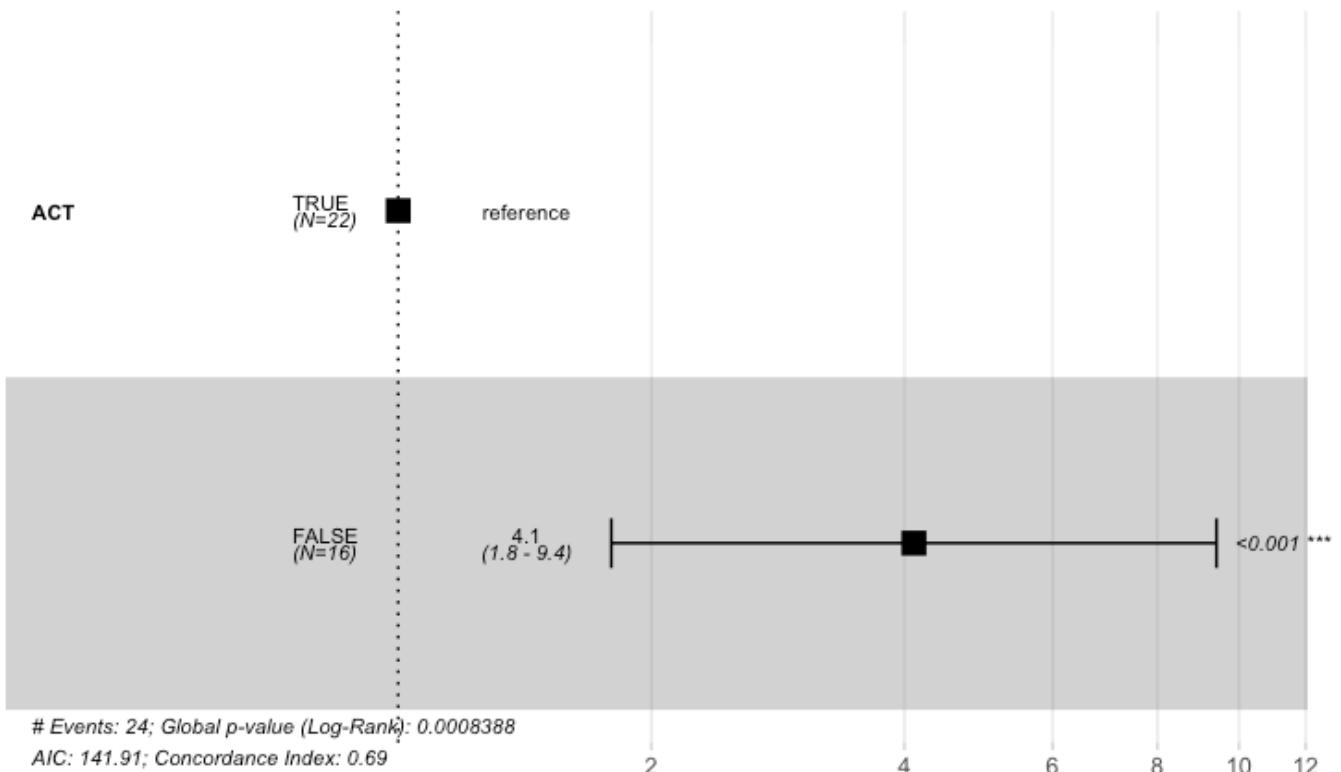
ACT=TRUE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
06	24.00	5.000	10.000	0.513	0.114	0.278		0.7	

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

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 1.4121    4.1047   0.4227 3.341 0.000836 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     4.105     0.2436    1.792      9.4

Concordance= 0.689 (se = 0.043 )
Likelihood ratio test= 11.15 on 1 df,  p=8e-04
Wald test          = 11.16 on 1 df,  p=8e-04
Score (logrank) test = 12.9 on 1 df,  p=3e-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 = 4.1 (1.79-9.4); p = 0.001"
```

[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	1.65875	5.25275	0.47737	3.475	0.000511 ***
GenderMale	-0.05152	0.94978	0.45032	-0.114	0.908913
Age.Group≥70	0.09450	1.09911	0.50749	0.186	0.852275
ECOG1	1.49973	4.48048	0.59745	2.510	0.012065 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	5.2528	0.1904	2.0609	13.388
GenderMale	0.9498	1.0529	0.3929	2.296
Age.Group≥70	1.0991	0.9098	0.4065	2.972
ECOG1	4.4805	0.2232	1.3892	14.450

Concordance= 0.74 (se = 0.054)

Likelihood ratio test= 17.32 on 4 df, p=0.002

Wald test = 15.5 on 4 df, p=0.004

Score (logrank) test = 19.17 on 4 df, p=7e-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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$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-side d 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	-1.65875	0.19038	0.47737	-3.475	0.000511 ***
GenderMale	-0.05152	0.94978	0.45032	-0.114	0.908913
Age.Group≥70	0.09450	1.09911	0.50749	0.186	0.852275
ECOG1	1.49973	4.48048	0.59745	2.510	0.012065 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTTRUE	0.1904	5.2528	0.07469	0.4852
GenderMale	0.9498	1.0529	0.39293	2.2958
Age.Group≥70	1.0991	0.9098	0.40650	2.9718
ECOG1	4.4805	0.2232	1.38924	14.4501

Concordance= 0.74 (se = 0.054)

Likelihood ratio test= 17.32 on 4 df, p=0.002

Wald test = 15.5 on 4 df, p=0.004

Score (logrank) test = 19.17 on 4 df, p=7e-04

#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
<lg>	<int>	<int>	<dbl>	<dbl>
FALSE	400	17	0.04250000	4.250000

ACT <lgl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
TRUE	76	1	0.01315789	1.315789

2 rows

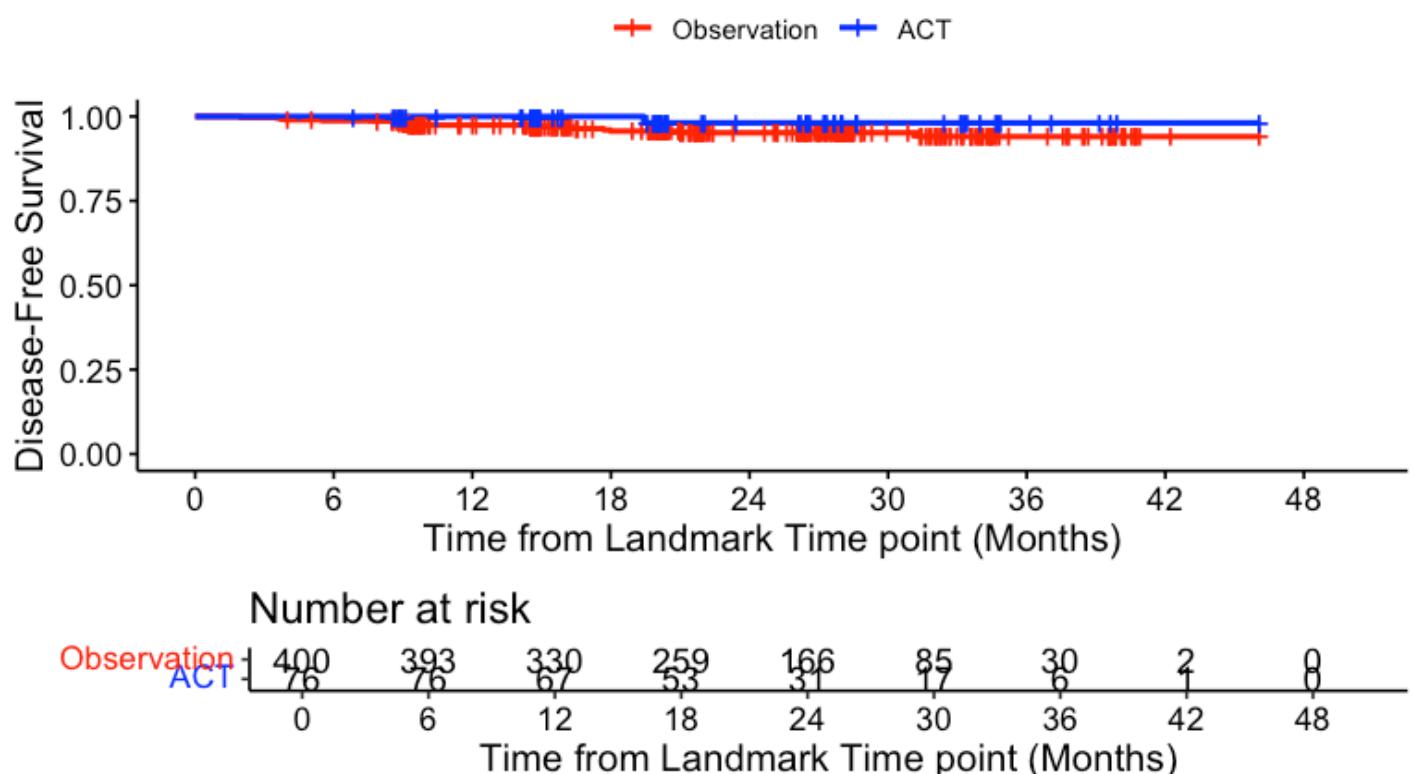
[Hide](#)

```

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

[Hide](#)

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

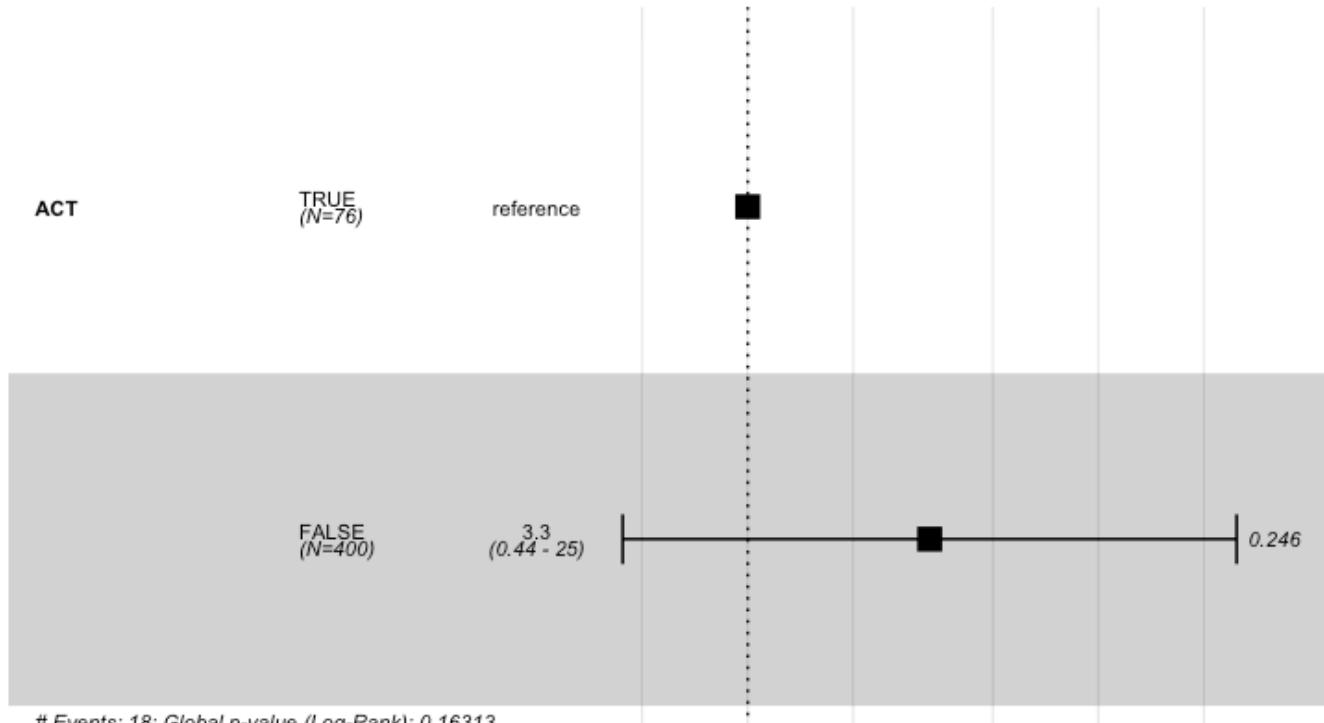
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95, conf.type = "log-log")

CI	ACT=FALSE		survival	std.err	lower	95% CI	upper	95%
	time	n.risk						
04	24.0000	166.0000	16.0000	0.9516	0.0121	0.9212	0.97	
<hr/>								
CI	ACT=TRUE		survival	std.err	lower	95% CI	upper	95%
	time	n.risk						
73	24.0000	31.0000	1.0000	0.9811	0.0187	0.8735	0.99	

[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= 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

```
#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)
```

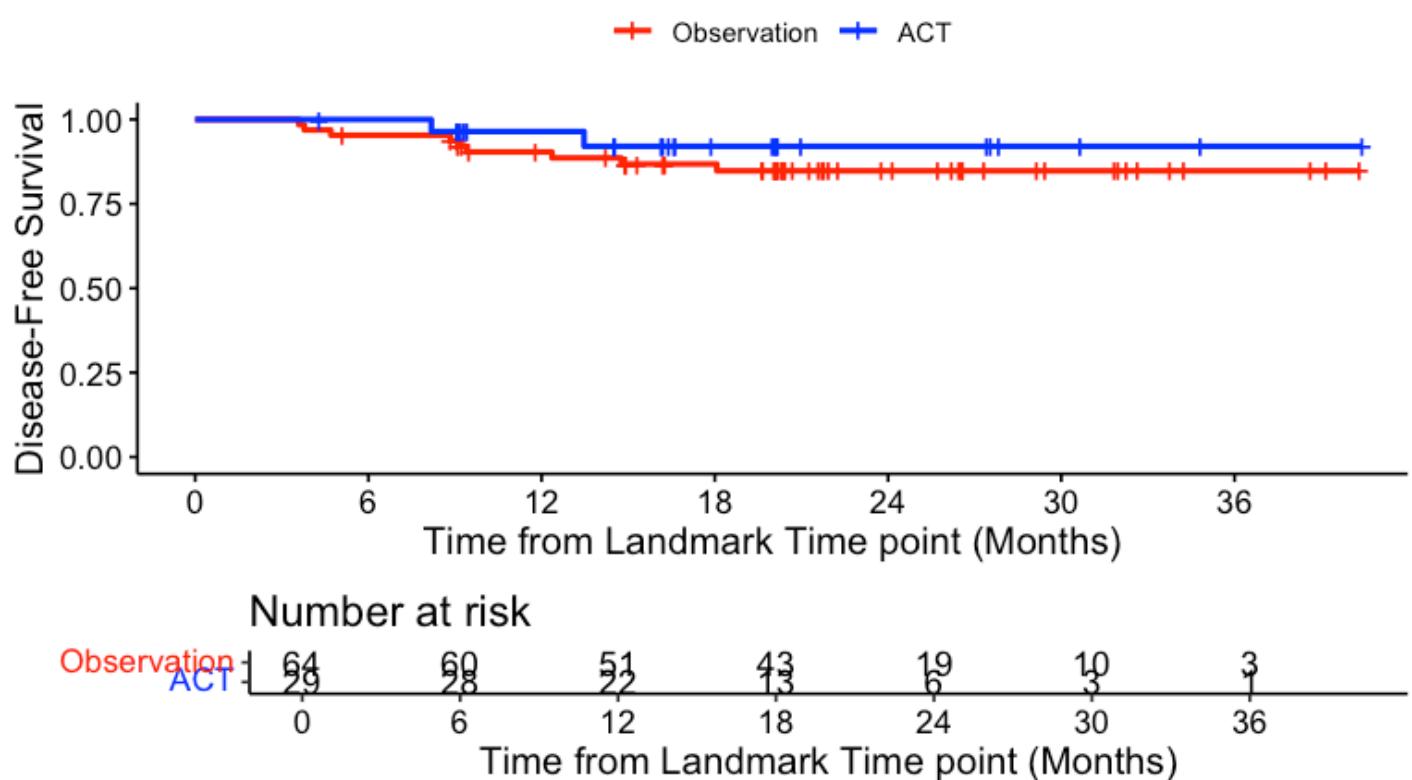
ACT	Total	Events	Fraction	Percentage
<lg1>	<int>	<int>	<dbl>	<dbl>
FALSE	64	9	0.14062500	14.062500
TRUE	29	2	0.06896552	6.896552

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))
```

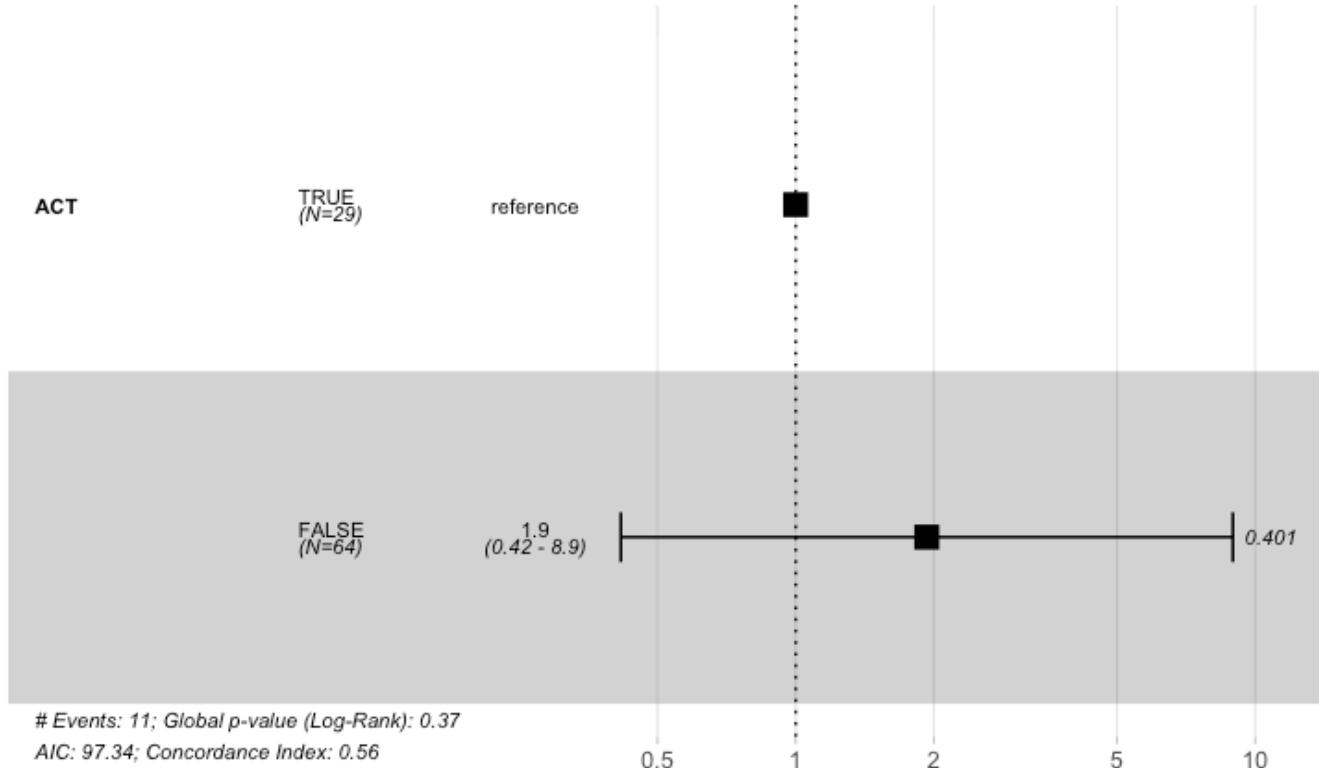
```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

CI	ACT=FALSE							
	time	n.risk	n.event	survival	std.err	lower	95% CI upper	95%
81	24.0000	19.0000	9.0000	0.8478	0.0471	0.7267	0.91	
CI	ACT=TRUE							
	time	n.risk	n.event	survival	std.err	lower	95% CI upper	95%
97	24.0000	6.0000	2.0000	0.9205	0.0544	0.7154	0.97	

[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 0.401
```

```
exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE 1.929 0.5184 0.4162 8.94
```

```
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](#)

```
#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_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$pT <- factor(circ_data$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$ctDNA.MRD=="NEGATIVE",]
circ_data$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$pT <- factor(circ_data$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-side d 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
<lgI>	<int>	<int>	<dbl>	<dbl>
FALSE	213	29	0.1361502	13.61502
TRUE	469	52	0.1108742	11.08742

2 rows

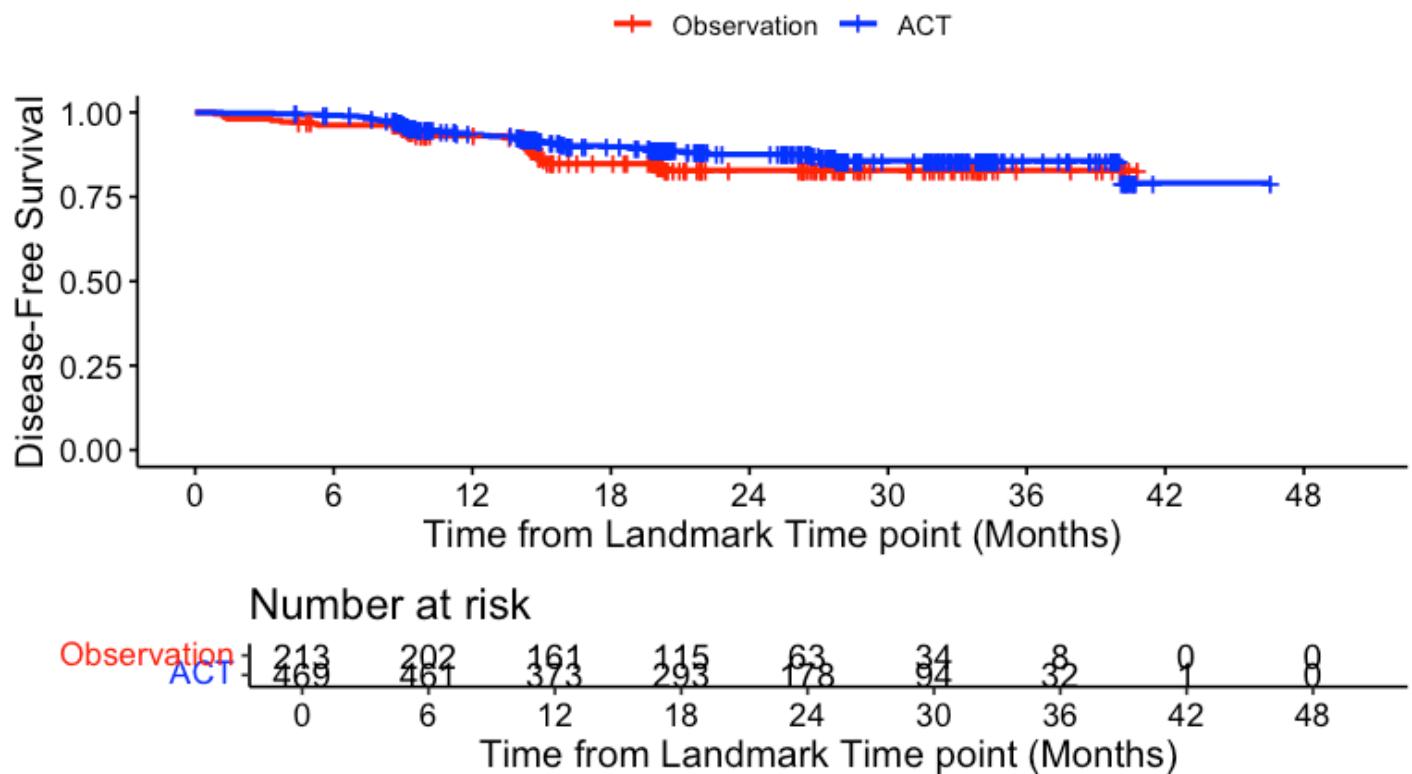
[Hide](#)

```

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



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

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

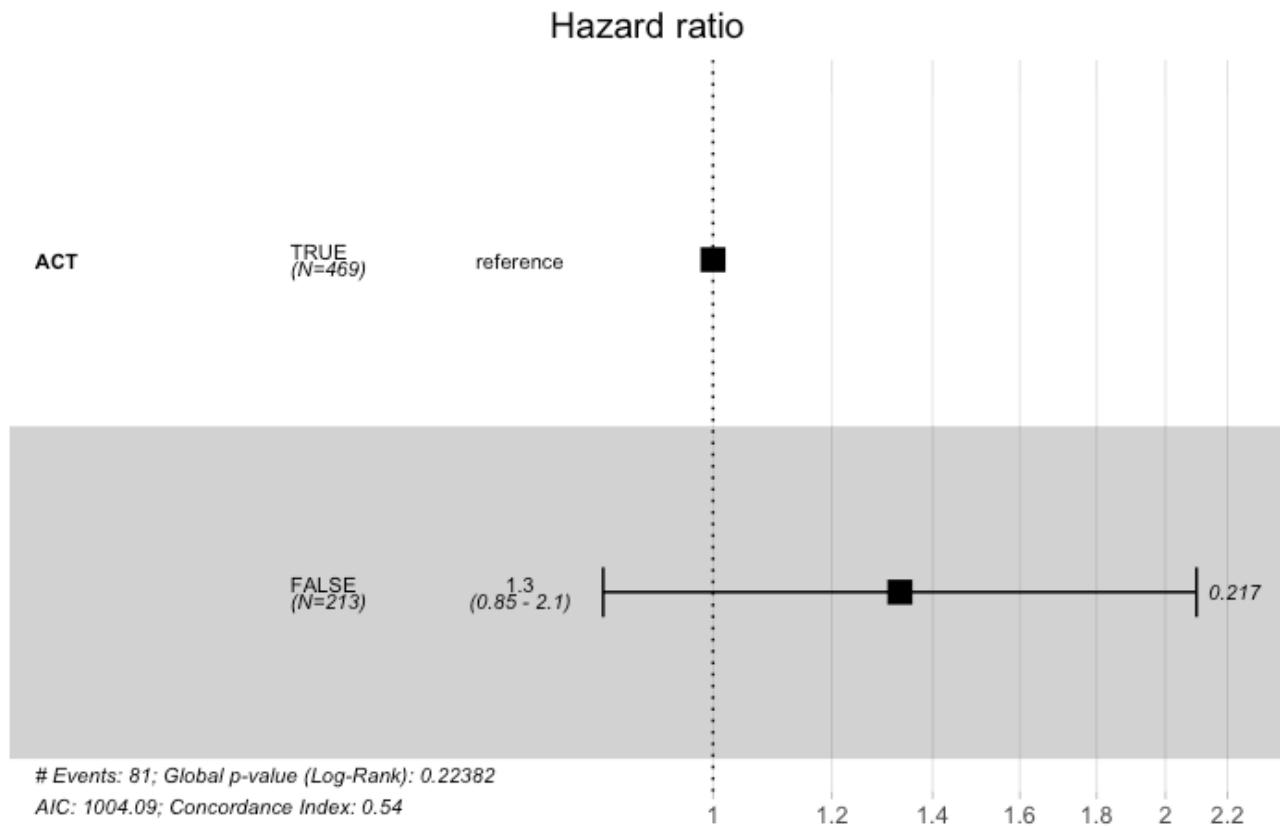
ACT=FALSE

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	0.829	0.924
24	63	2	0.829	0.0300	0.760	0.879	0.799	0.906

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	0.833	0.954
24	178	6	0.876	0.0173	0.837	0.906	0.809	0.925

```
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)
```



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

```
#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$DFS.months-2
circ_data <- circ_data[circ_data$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-side 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-side d 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_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 33      31    3.62    2.57    4.28
ACT=TRUE  121     78   11.27   9.01   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)
```

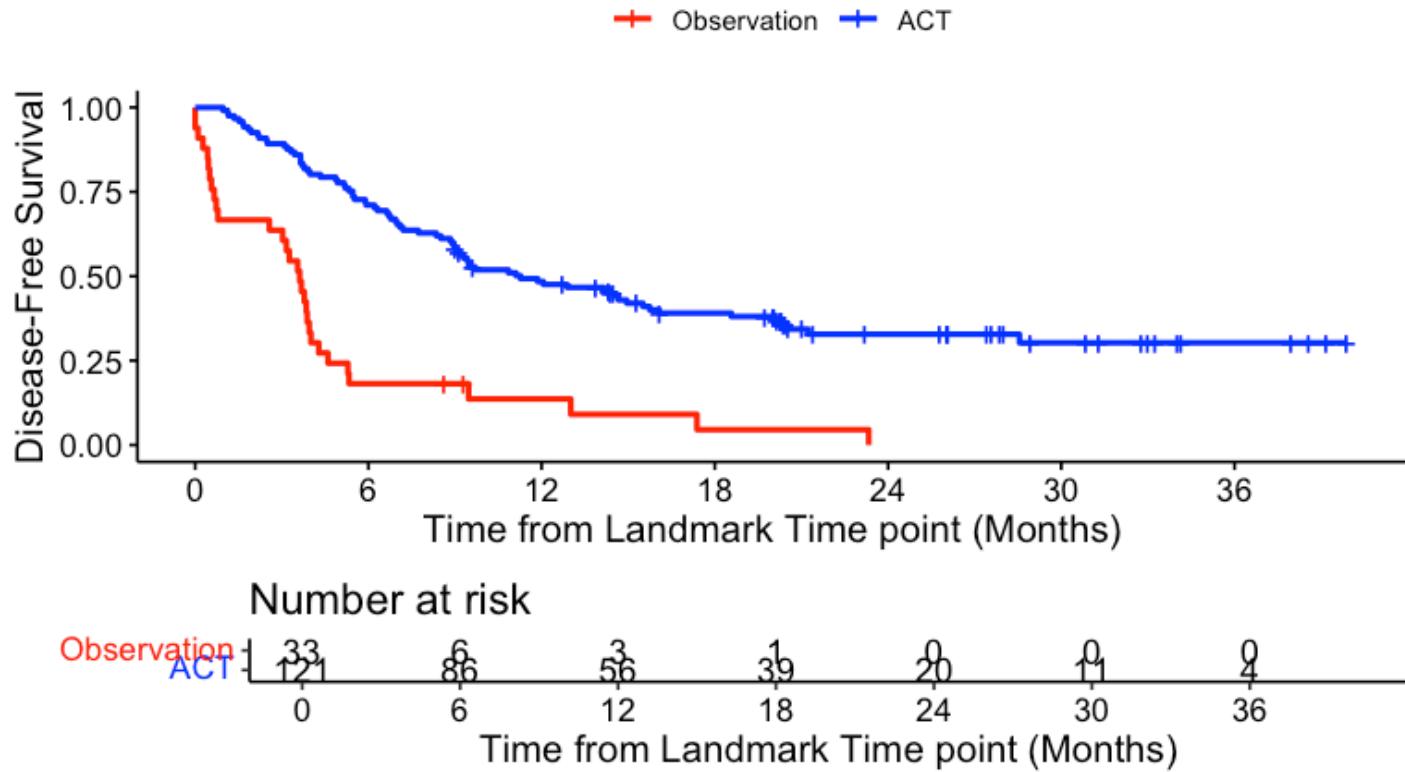
ACT	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
FALSE	33	31	0.9393939	93.93939
TRUE	121	78	0.6446281	64.46281

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")
```

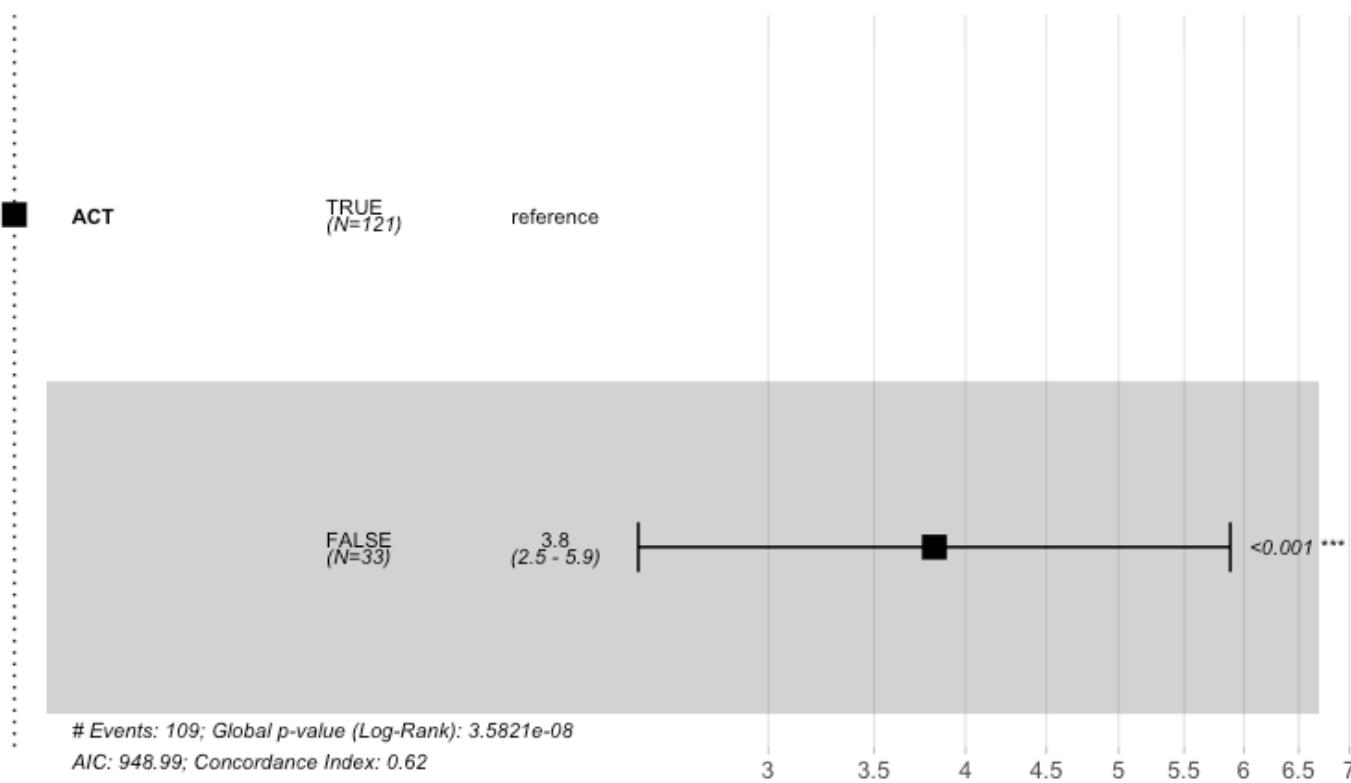
CI	ACT=FALSE						
	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
18.00000	1.00000	30.00000	0.04545	0.04279	0.00364	0.182	

ACT=TRUE							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	
18	39	72	0.391	0.0455	0.302	0.479	
24	20	5	0.328	0.0463	0.240	0.419	

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	1.3407	3.8217	0.2201	6.092	1.12e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	3.822	0.2617	2.483	5.883

Concordance= 0.615 (se = 0.021)

Likelihood ratio test= 30.36 on 1 df, p=4e-08

Wald test = 37.11 on 1 df, p=1e-09

Score (logrank) test = 42.5 on 1 df, p=7e-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 = 3.82 (2.48-5.88); p = 0"
```

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$DFS.months-2
circ_data <- circ_data[circ_data$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-side 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	1.38965	4.01346	0.24012	5.787	7.15e-09 ***
GenderMale	0.04947	1.05071	0.20023	0.247	0.805
Age.Group≥70	-0.04450	0.95648	0.20403	-0.218	0.827
ECOG1	-0.13130	0.87696	0.32956	-0.398	0.690

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	4.0135	0.2492	2.5069	6.425
GenderMale	1.0507	0.9517	0.7097	1.556
Age.Group≥70	0.9565	1.0455	0.6412	1.427
ECOG1	0.8770	1.1403	0.4597	1.673

Concordance= 0.626 (se = 0.027)

Likelihood ratio test= 30.68 on 4 df, p=4e-06

Wald test = 37.33 on 4 df, p=2e-07

Score (logrank) test = 42.8 on 4 df, p=1e-08

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_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-side d 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	-1.38965	0.24916	0.24012	-5.787	7.15e-09 ***
GenderMale	0.04947	1.05071	0.20023	0.247	0.805
Age.Group≥70	-0.04450	0.95648	0.20403	-0.218	0.827
ECOG1	-0.13130	0.87696	0.32956	-0.398	0.690

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTTRUE	0.2492	4.0135	0.1556	0.3989
GenderMale	1.0507	0.9517	0.7097	1.5557
Age.Group≥70	0.9565	1.0455	0.6412	1.4267
ECOG1	0.8770	1.1403	0.4597	1.6730

Concordance= 0.626 (se = 0.027)

Likelihood ratio test= 30.68 on 4 df, p=4e-06

Wald test = 37.33 on 4 df, p=2e-07

Score (logrank) test = 42.8 on 4 df, p=1e-08

#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_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	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
<lg1>	<int>	<int>	<dbl>	<dbl>
FALSE	32	32	1.0000000	100.00000
TRUE	14	13	0.9285714	92.85714

2 rows

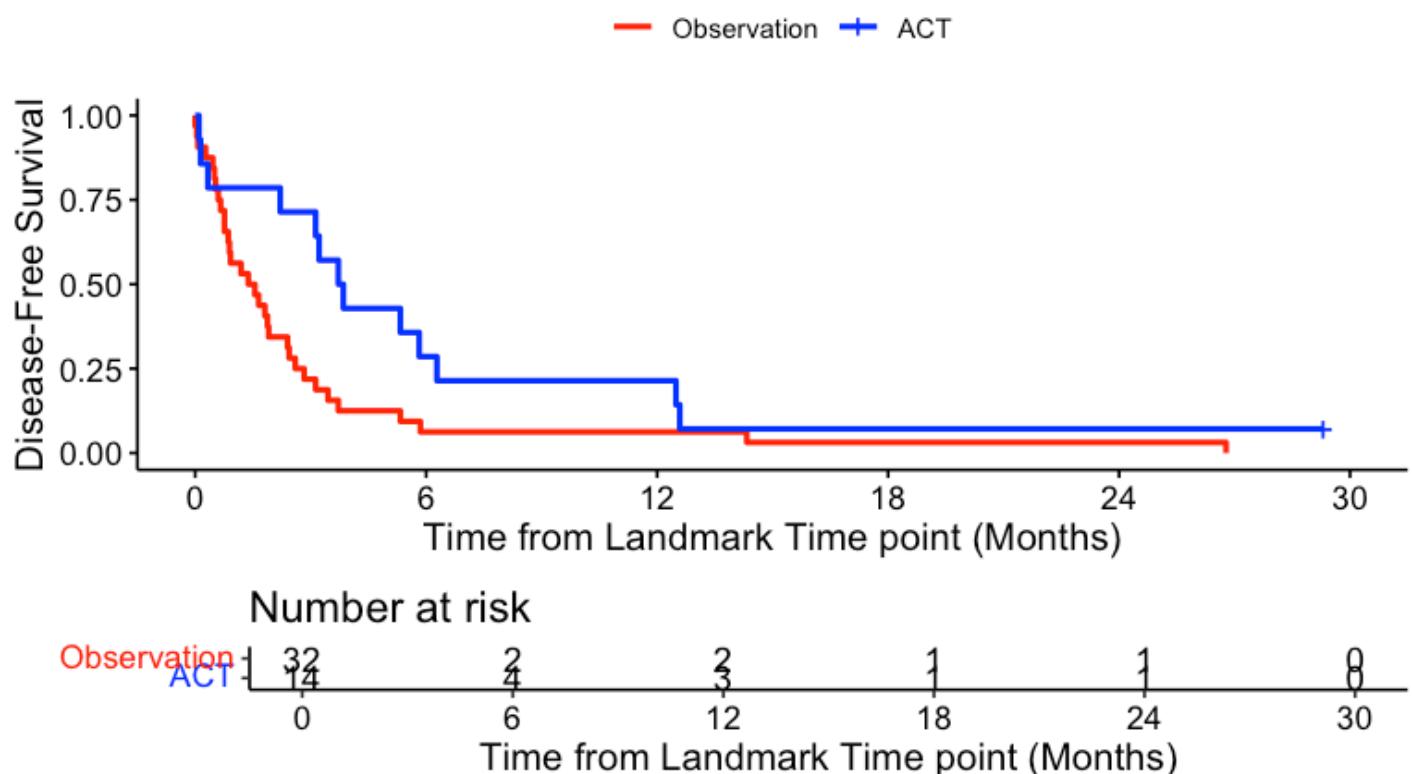
[Hide](#)

```

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-tr€



```
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		

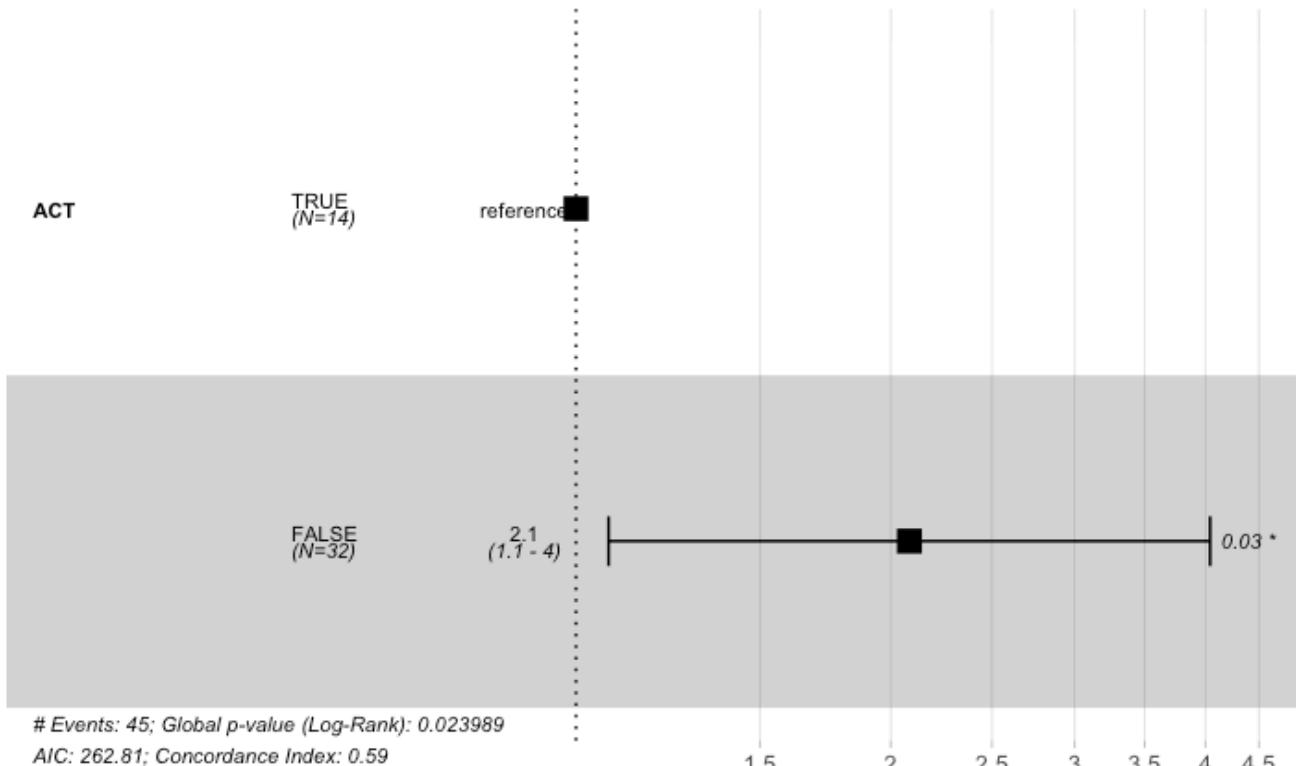
ACT=TRUE

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](#)

```

#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	29	27	3.03	1.29	3.52
ACT=TRUE	25	15	14.16	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
<lg>	<int>	<int>	<dbl>	<dbl>
FALSE	29	27	0.9310345	93.10345

ACT <lgl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
TRUE	25	15	0.6000000	60.00000

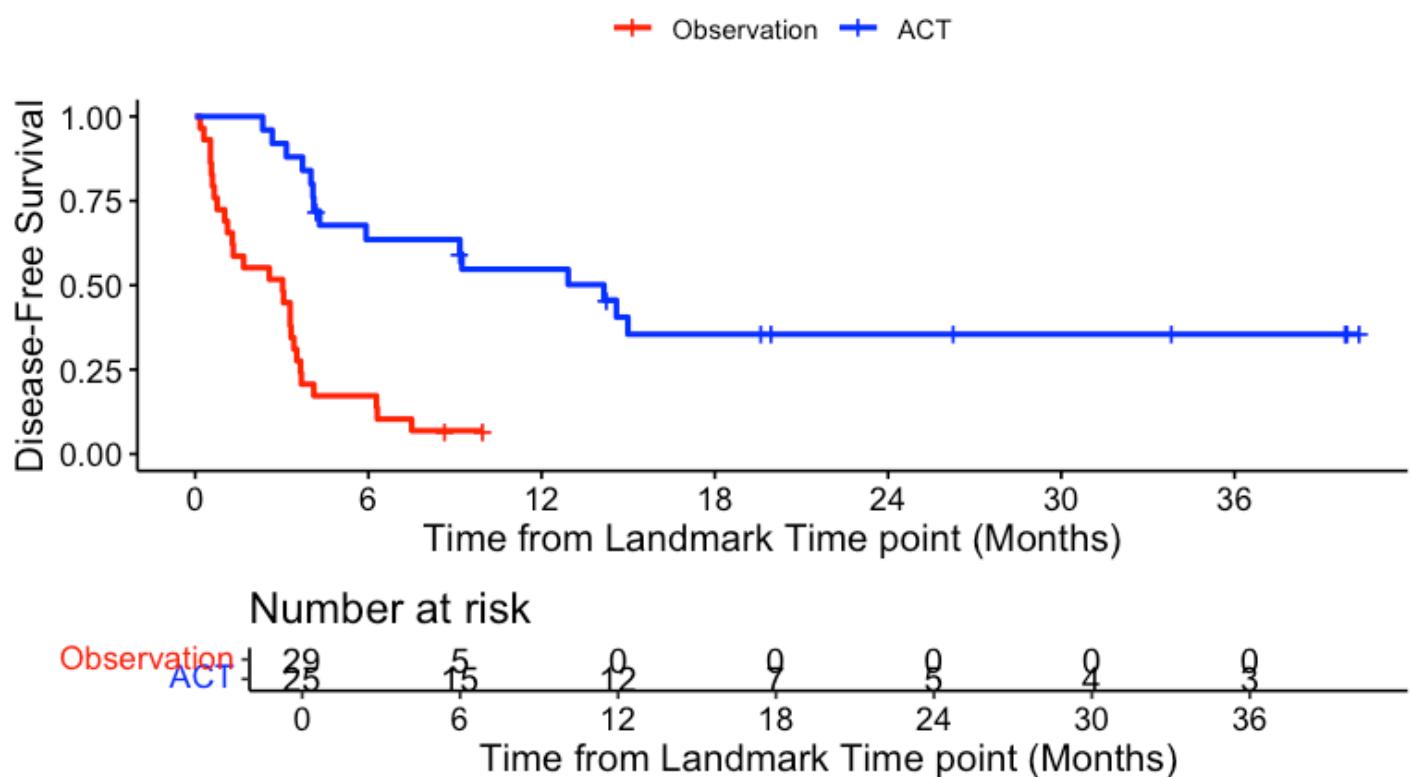
2 rows

```

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=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="")

```

DFS - ctDNA MRD Positive ACT vs Observation | Stage IV No NAC



```
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	15	14	0.517	0.0928	0.3250	0.3250	0.679	0.679
6	5	10	0.172	0.0701	0.0629	0.0629	0.327	0.327

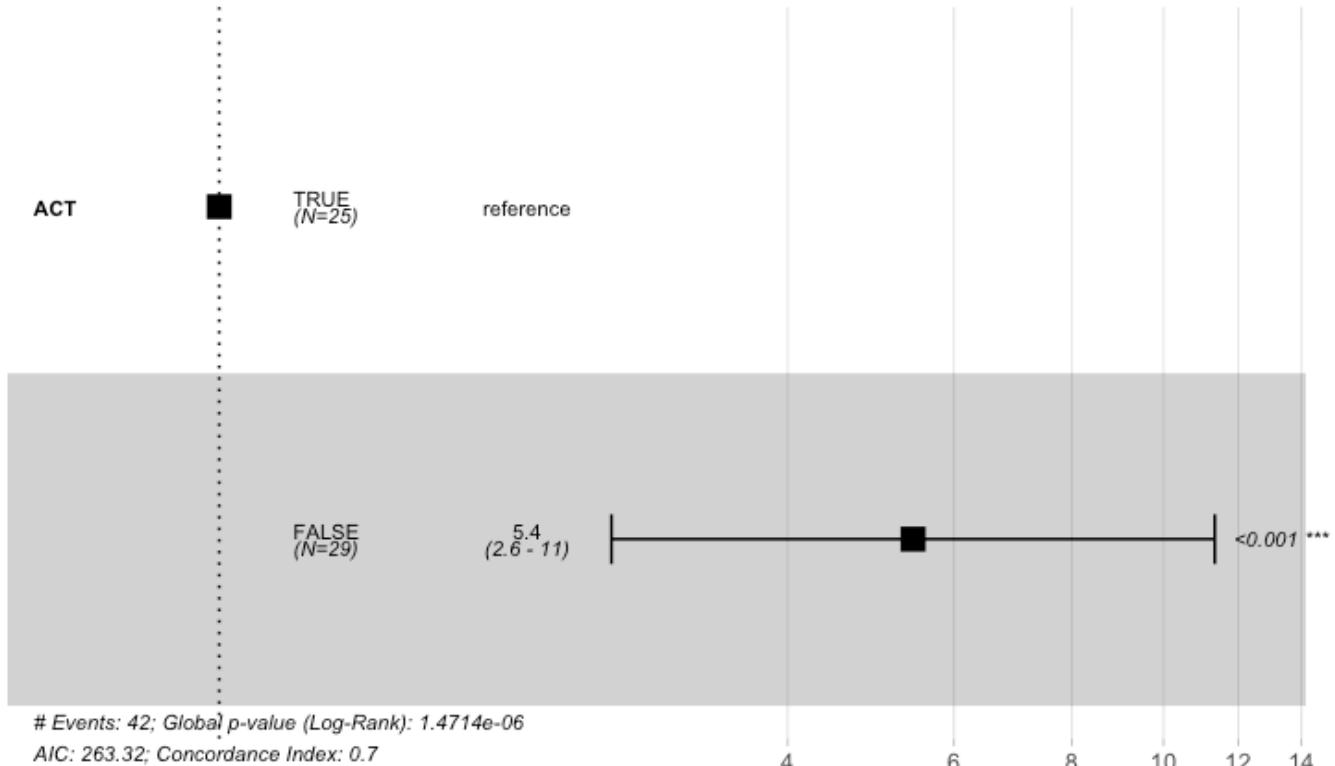
ACT=TRUE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
3	23	2	0.920	0.0543	0.716	0.716	0.979	0.979
6	15	7	0.635	0.0972	0.415	0.415	0.791	0.791
18	7	6	0.355	0.1022	0.167	0.167	0.548	0.548
24	5	0	0.355	0.1022	0.167	0.167	0.548	0.548

[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 1.6921    5.4311   0.3753 4.508 6.53e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE    5.431     0.1841    2.603    11.33

Concordance= 0.7 (se = 0.027 )
Likelihood ratio test= 23.18 on 1 df,  p=1e-06
Wald test          = 20.33 on 1 df,  p=7e-06
Score (logrank) test = 24.48 on 1 df,  p=8e-07

```

[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 = 5.43 (2.6-11.33); p = 0"
```

[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_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= 54, number of events= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)
ACTFALSE	1.773324	5.890400	0.391932	4.525	6.05e-06 ***
GenderMale	-0.238240	0.788014	0.334302	-0.713	0.476
Age.Group≥70	0.097276	1.102165	0.321574	0.303	0.762
ECOG1	-0.004307	0.995703	0.616553	-0.007	0.994

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTFALSE	5.8904	0.1698	2.7323	12.699
GenderMale	0.7880	1.2690	0.4092	1.517
Age.Group≥70	1.1022	0.9073	0.5868	2.070
ECOG1	0.9957	1.0043	0.2974	3.334

Concordance= 0.714 (se = 0.033)

Likelihood ratio test= 23.79 on 4 df, p=9e-05

Wald test = 20.78 on 4 df, p=3e-04

Score (logrank) test = 24.98 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$Eligible=="TRUE",]
circ_data <- circ_data[!(circ_data$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	-1.773324	0.169768	0.391932	-4.525	6.05e-06 ***
GenderMale	-0.238240	0.788014	0.334302	-0.713	0.476
Age.Group≥70	0.097276	1.102165	0.321574	0.303	0.762
ECOG1	-0.004307	0.995703	0.616553	-0.007	0.994

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ACTTRUE	0.1698	5.8904	0.07875	0.366
GenderMale	0.7880	1.2690	0.40924	1.517
Age.Group≥70	1.1022	0.9073	0.58684	2.070
ECOG1	0.9957	1.0043	0.29738	3.334

Concordance= 0.714 (se = 0.033)

Likelihood ratio test= 23.79 on 4 df, p=9e-05

Wald test = 20.78 on 4 df, p=3e-04

Score (logrank) test = 24.98 on 4 df, p=5e-05

#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_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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 98 42 27.53 18.07 NA

ctDNA.Dynamics=2 70 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)
```

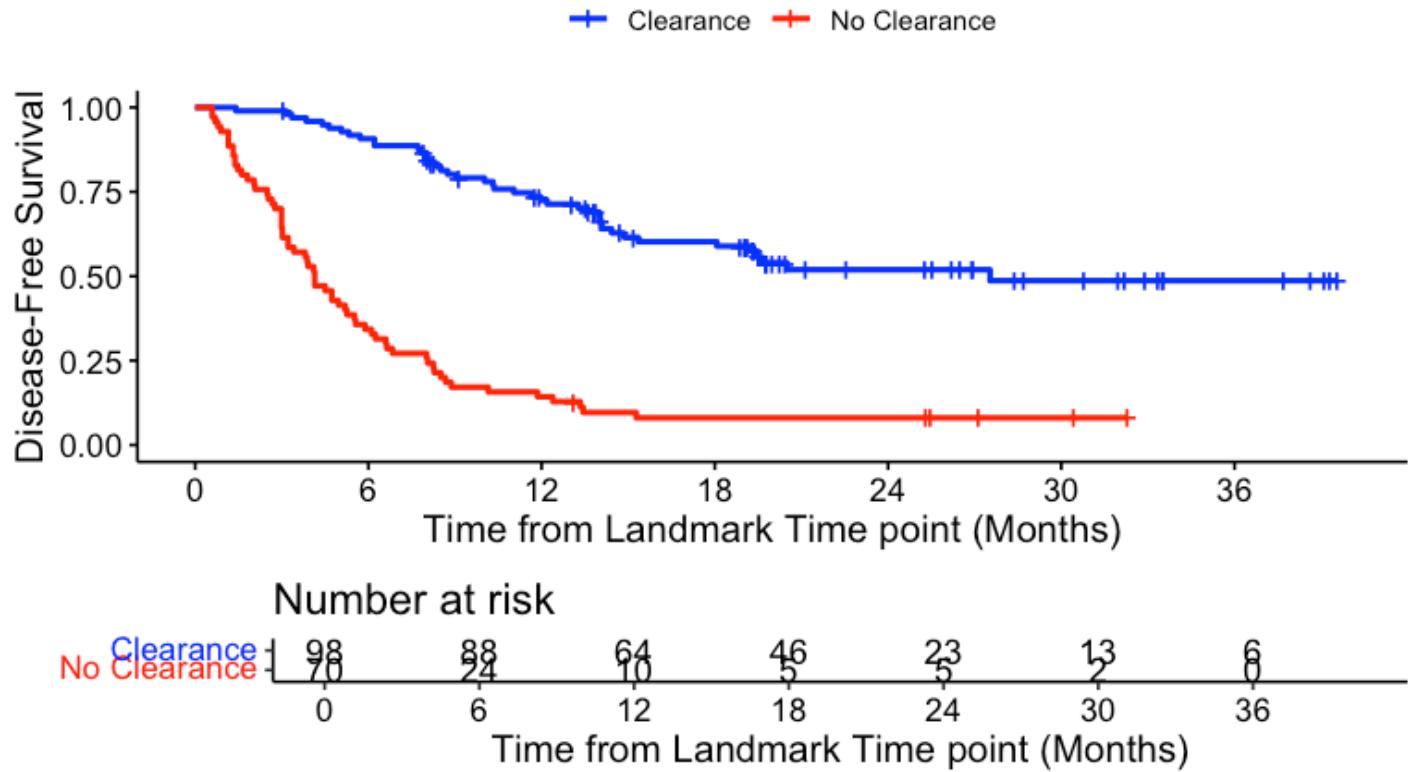
ctDNA.Dynamics	Total	Events	Fraction	Percentage
	<dbl>	<int>	<dbl>	<dbl>
1	98	42	0.4285714	42.85714
2	70	64	0.9142857	91.42857
NA	674	NA	NA	NA

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 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

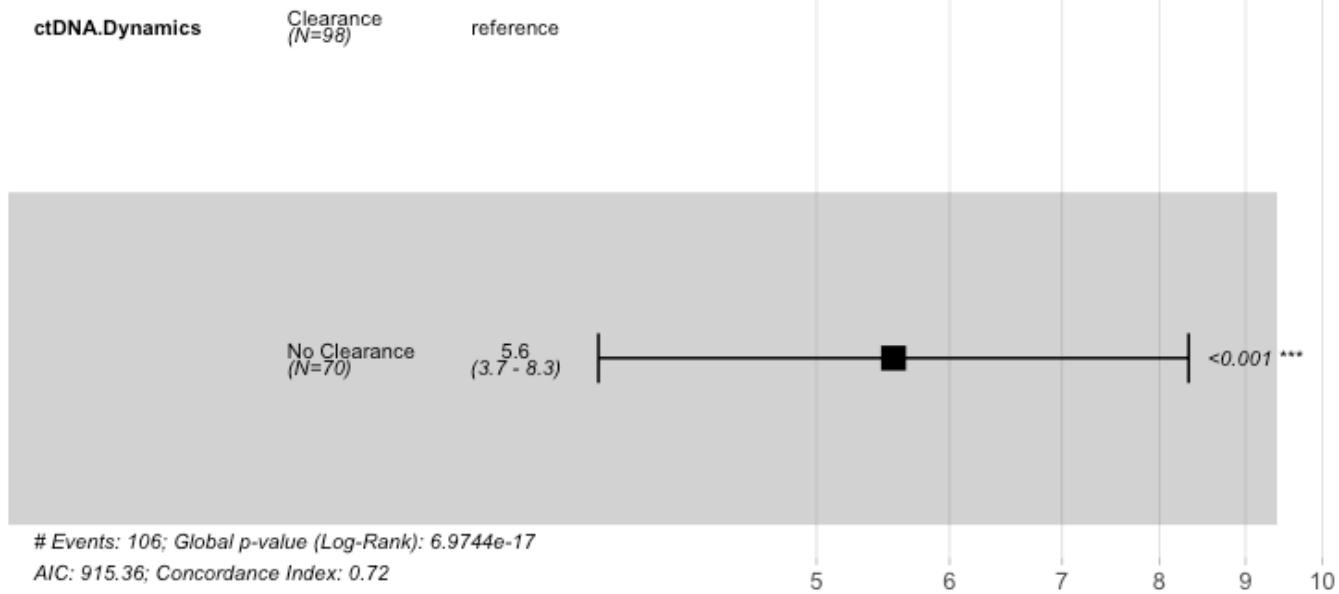
CI	ctDNA.Dynamics=1						
	time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	23.0000	41.0000	0.5193	0.0571	0.4024	0.62	0.39

CI	ctDNA.Dynamics=2						
	time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	5.0000	64.0000	0.0804	0.0333	0.0308	0.16	0.09

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= 168, number of events= 106
(674 observations deleted due to missingness)
```

```
            coef  exp(coef)  se(coef)      z  Pr(>|z|)
ctDNA.DynamicsNo Clearance 1.7150      5.5565  0.2063 8.314 <2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
            exp(coef)  exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance      5.556        0.18     3.709     8.325
```

```
Concordance= 0.717 (se = 0.018 )
```

```
Likelihood ratio test= 69.68 on 1 df,  p=<2e-16
```

```
Wald test             = 69.11 on 1 df,  p=<2e-16
```

```
Score (logrank) test = 84 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 = 5.56 (3.71-8.33); p = 0"
```

```
#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_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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 98      7      NA      NA      NA
ctDNA.Dynamics=2 70      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	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	98	7	0.07142857	7.142857
2	70	16	0.22857143	22.857143
NA	674	NA	NA	NA

3 rows

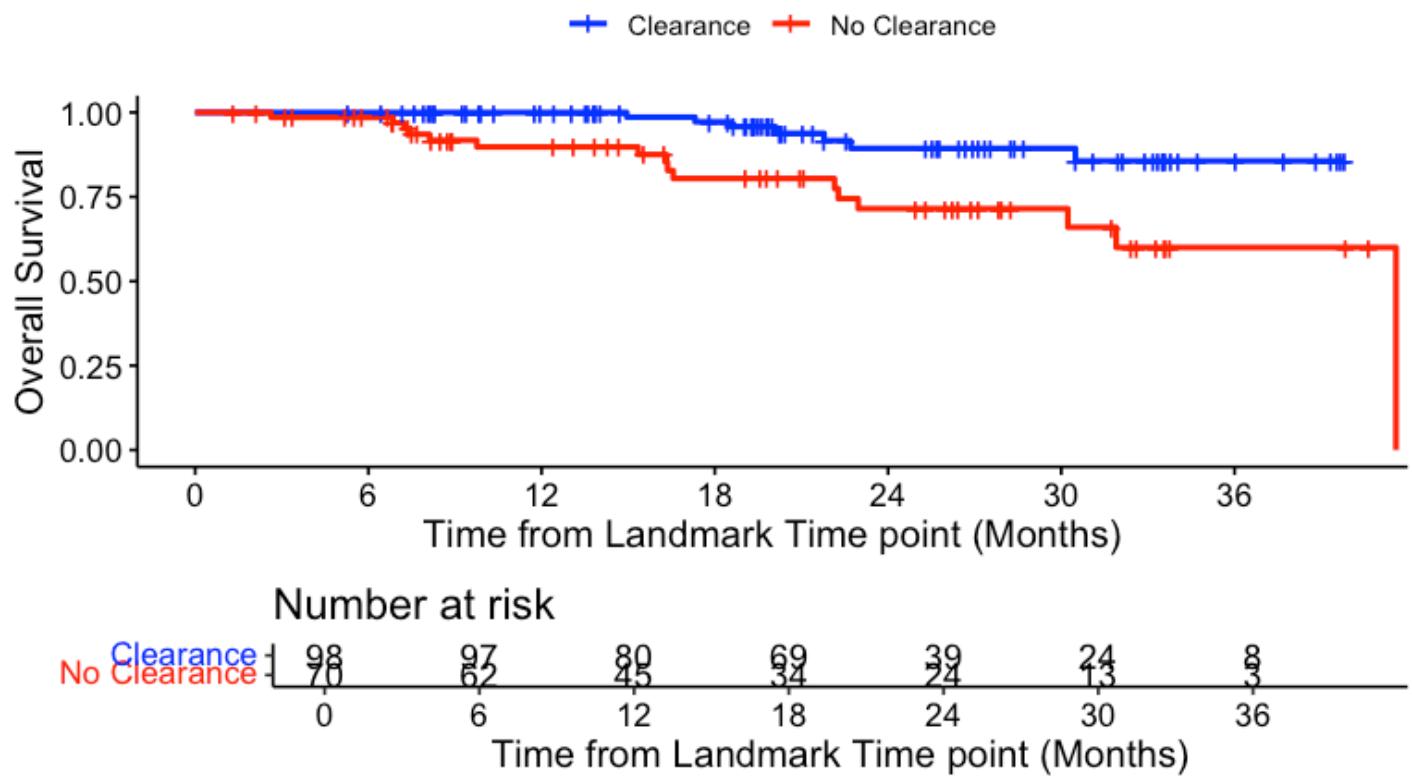
Hide

```

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=c("blue","red"), title="OS - ctDNA Clearance from MRD to 3 months ACT-treated | All Stages", ylab= "Overall Survival", 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 Sta



```
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	0.95
19							

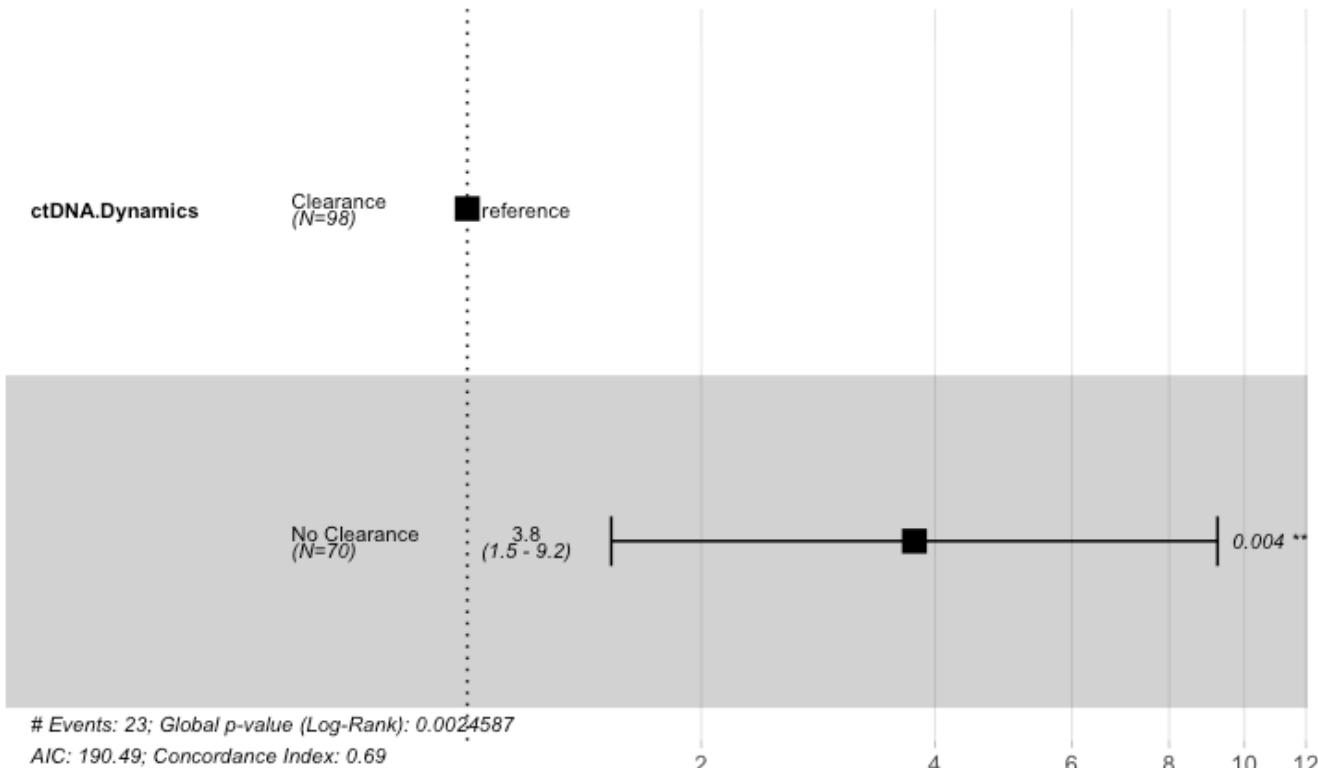
ctDNA.Dynamics=2

	time	n.risk	n.event	survival	std.err	lower 95%	CI upper 95%
CI	24.0000	24.0000	13.0000	0.7150	0.0701	0.5516	0.82
77							

[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= 168, number of events= 23
(674 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance 1.3247    3.7612  0.4583 2.891  0.00385 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     3.761     0.2659     1.532     9.235

Concordance= 0.688  (se = 0.047 )
Likelihood ratio test= 9.17  on 1 df,  p=0.002
Wald test            = 8.36  on 1 df,  p=0.004
Score (logrank) test = 9.64  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 (1.53-9.23); p = 0.004"
```

```
#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_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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)

```

```

729 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)
```

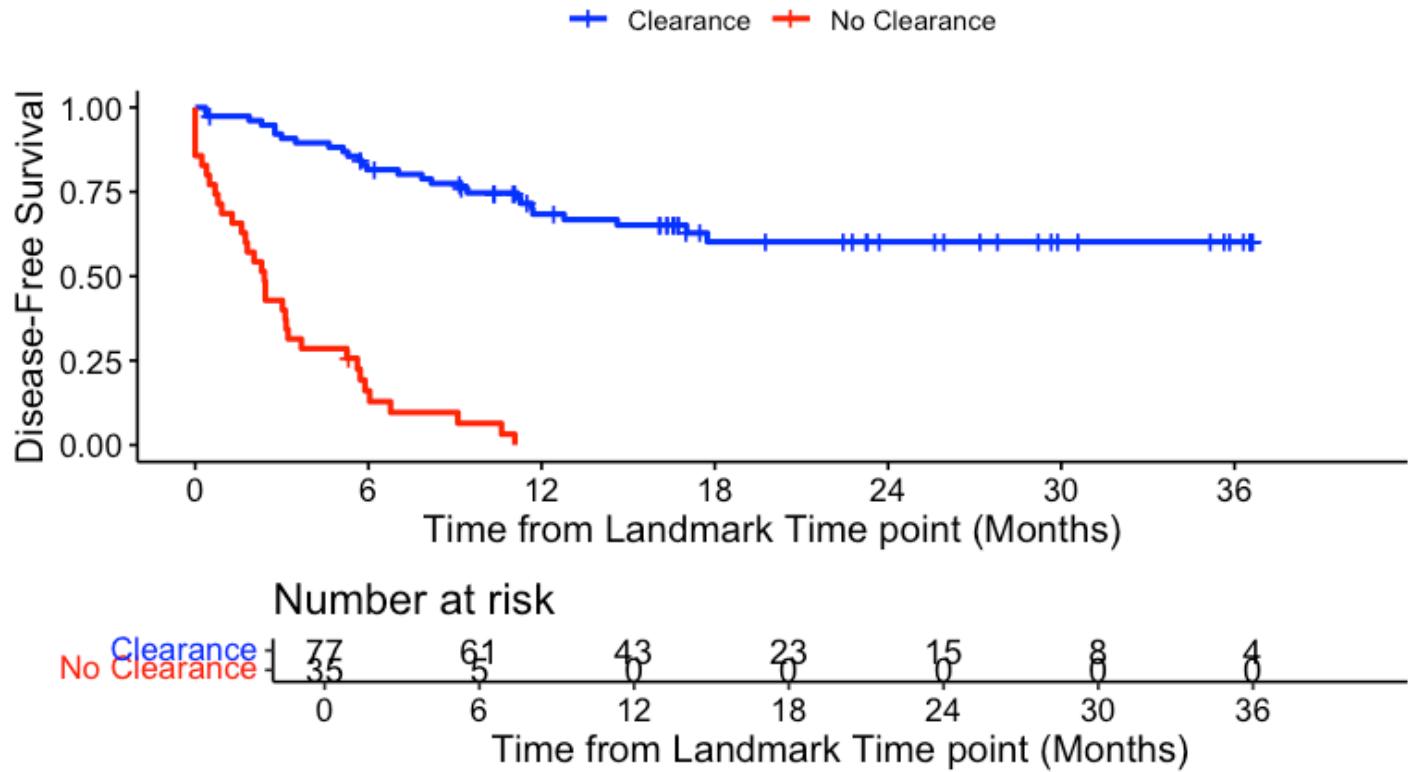
ctDNA.Dynamics	Total	Events	Fraction	Percentage
	<dbl>	<int>	<dbl>	<dbl>
1	77	27	0.3506494	35.06494
2	35	34	0.9714286	97.14286
NA	729	NA	NA	NA

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")
```

729 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	

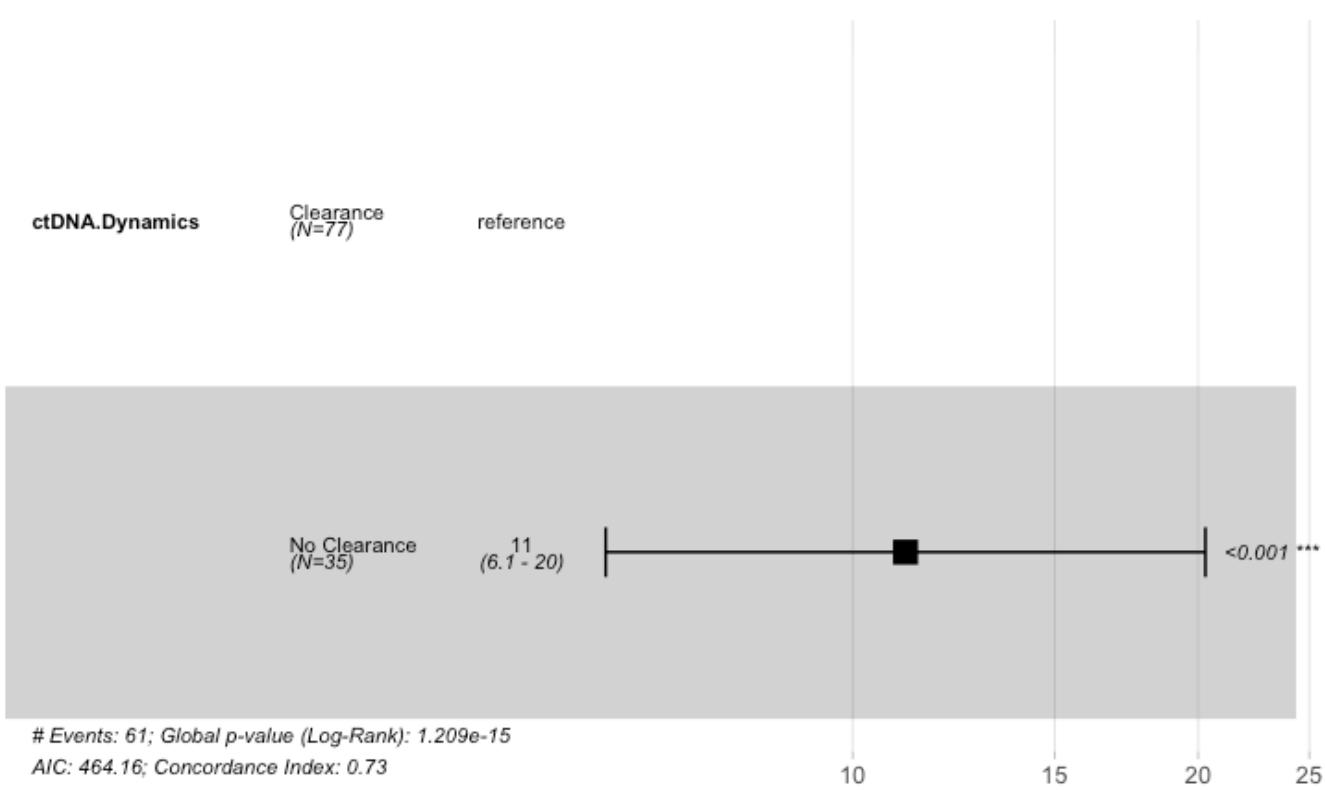
ctDNA.Dynamics=2

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		0.30	
28								

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
(729 observations deleted due to missingness)

		coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Dynamics	No 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"

#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_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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)

729 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	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	77	3	0.03896104	3.896104
2	36	7	0.19444444	19.444444
NA	729	NA	NA	NA

3 rows

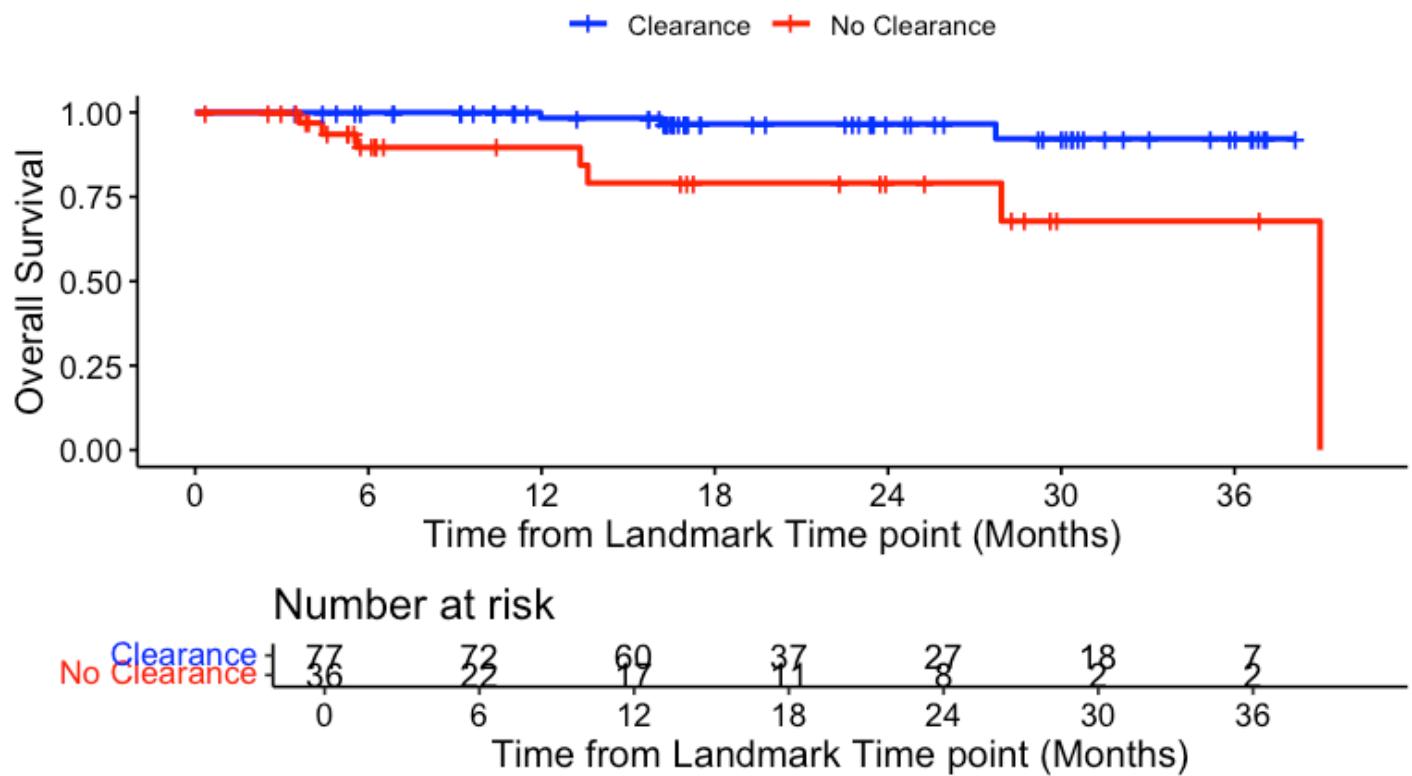
Hide

```

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=c("blue","red"), title="OS - ctDNA Clearance from MRD to 6 months ACT-treated | All Stages", ylab= "Overall Survival", 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 Sta



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")
```

729 observations deleted due to missingness

ctDNA.Dynamics=1

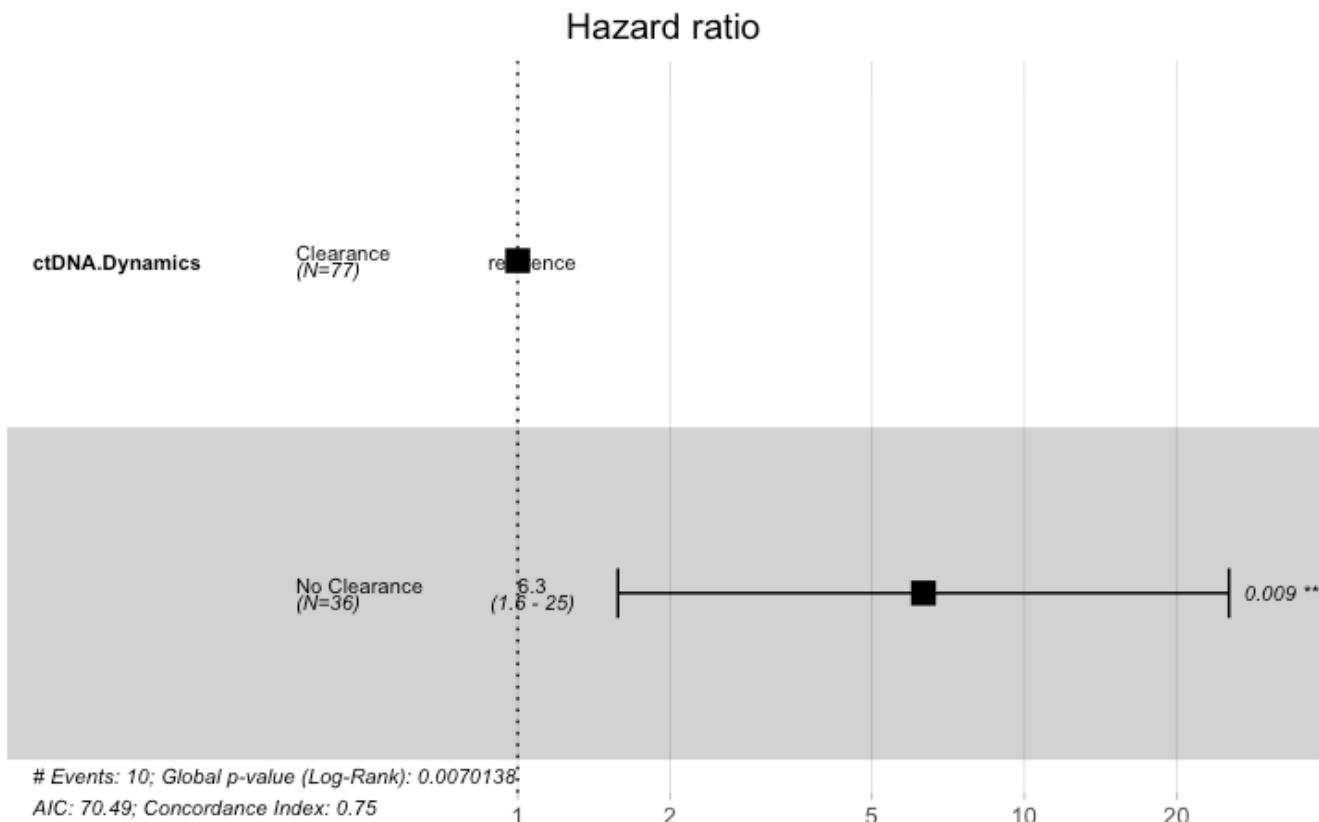
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
6	72	0	1.000	0.0000	NA	NA	NA	NA
24	27	2	0.966	0.0236	0.871	0.991	0.966	0.991

ctDNA.Dynamics=2

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
6	22	3	0.896	0.0571	0.710	0.966	0.966	0.991
24	8	2	0.791	0.0863	0.558	0.910	0.910	0.991

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)
```



Hide

```
summary(cox_fit)
```

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

n= 113, number of events= 10
 (729 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"
```

#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: 182"
```

[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: 54.1666666666667 %"
```

[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: 123"
```

[Hide](#)

```
percentage_clearance <- (clearance_postACT / positive_subset) * 100
print(paste("ctDNA Clearance post-ACT:", percentage_clearance, "%"))
```

```
[1] "ctDNA Clearance post-ACT: 67.5824175824176 %"
```

[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: 123"

[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: 66"

[Hide](#)

```
percentage_sustained_clearance <- (clearance_sustained / clearance_subset) * 100
print(paste("Sustained ctDNA Clearance:", percentage_sustained_clearance, "%"))
```

[1] "Sustained ctDNA Clearance: 53.6585365853659 %"

[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: 57"

[Hide](#)

```
percentage_transient_clearance <- (clearance_transient / clearance_subset) * 100
print(paste("Transient ctDNA Clearance:", percentage_transient_clearance, "%"))
```

[1] "Transient ctDNA Clearance: 46.3414634146341 %"

#Sankey plot for Sustained vs Transient Clearance

[Hide](#)

```
##To run this commands, please visit: https://sankeymatic.com/build/
#ctDNA + MRD window [182] ACT-treated #ADD8E6
#ctDNA + MRD window [154] Not treated #808080
#ACT-treated [123] ctDNA post-MRD Clearance #87EA86
#ACT-treated [5] No Clearance #E67272
#ACT-treated [4] No post-MRD time point #808080
#No Clearance [55] No Clearance analysis #E67272
#ctDNA post-MRD Clearance [123] Available post-MRD Timepoints #ADD8E66
#Available post-MRD Timepoints [66] Sustained Clearance #7393B3
#Available post-MRD Timepoints [57] 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	66	7	NA	NA	NA
ctDNA.Clearance=Transient	57	50	12.09	10.35	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	66	7	0.1060606	10.60606
Transient	57	50	0.8771930	87.71930
NA	131	NA	NA	NA

4 rows

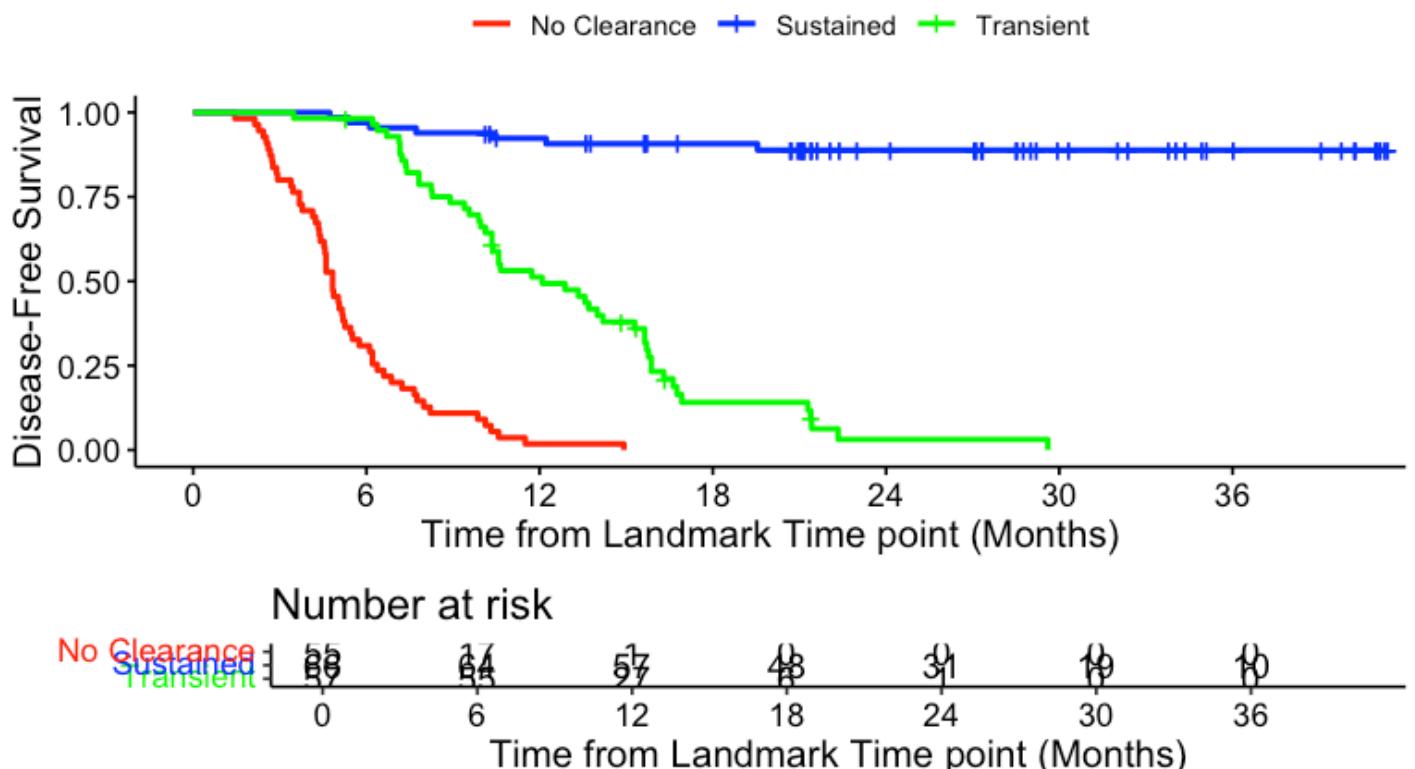
[Hide](#)

```

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=c("red","blue","green"), title="DFS - ctDNA Clearance post-MRD | All Stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("No Clearance", "Sustained", "Transient"), legend.title="")

```

DFS - 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")
```

131 observations deleted due to missingness

ctDNA.Clearance=No Clearance

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95%
12	1.00000	54.00000	0.01818	0.01802	0.00149	0.084
74						

ctDNA.Clearance=Sustained

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	57	5	0.923	0.0329	0.826	0.967
18	48	1	0.907	0.0361	0.805	0.957
24	31	1	0.888	0.0400	0.779	0.945

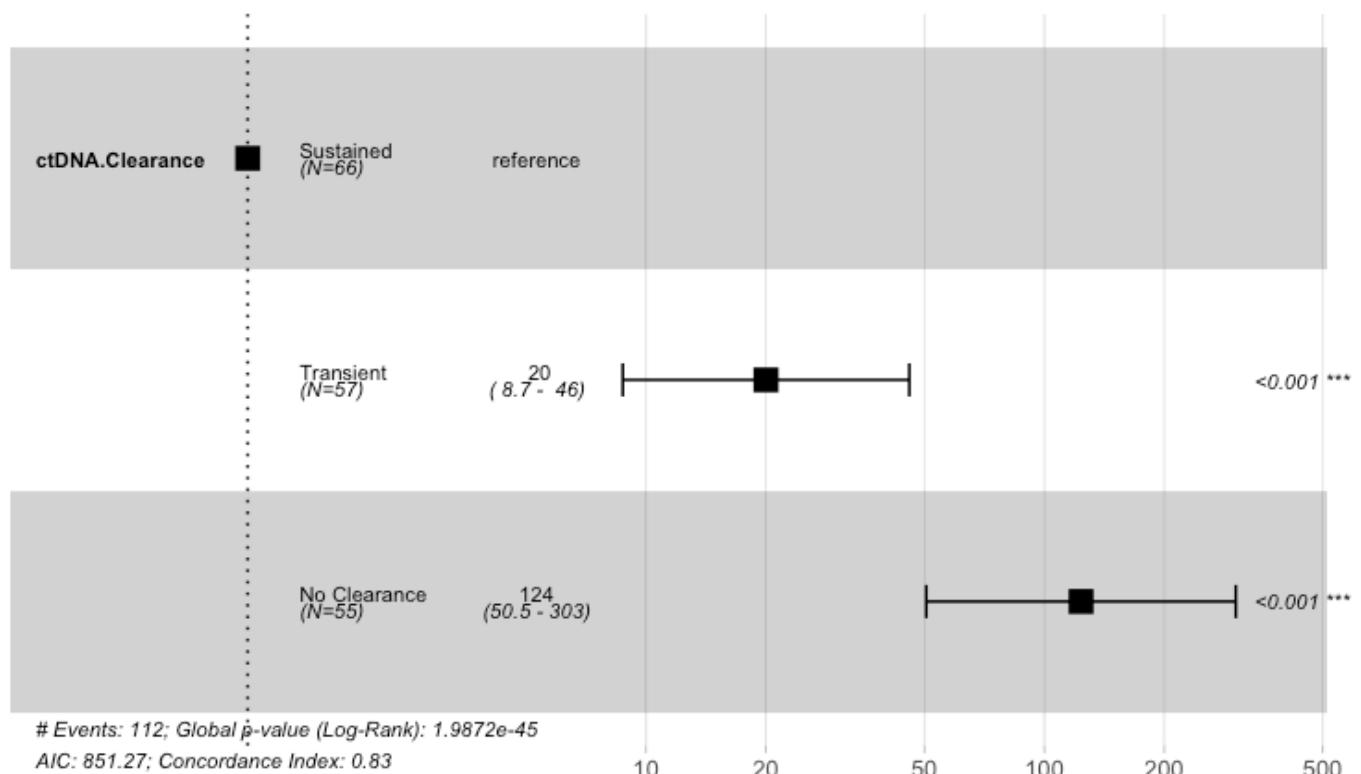
ctDNA.Clearance=Transient

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	27	27	0.5124	0.0675	0.37402	0.635
18	6	18	0.1410	0.0508	0.06023	0.255
24	1	4	0.0313	0.0294	0.00277	0.130

[Hide](#)

```
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= 178, number of events= 112
(131 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.ClearanceTransient    2.9962   20.0103   0.4226  7.09 1.34e-12 ***
ctDNA.ClearanceNo Clearance 4.8175  123.6519   0.4565 10.55 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.ClearanceTransient    20.01   0.049974     8.74    45.81
ctDNA.ClearanceNo Clearance 123.65   0.008087    50.54   302.51

Concordance= 0.829 (se = 0.017 )
Likelihood ratio test= 205.9 on 2 df,  p=<2e-16
Wald test                 = 127.8 on 2 df,  p=<2e-16
Score (logrank) test = 229.9 on 2 df,  p=<2e-16
```

#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_MR_MTM with log10
circ_data$p_MR_MTM <- as.numeric(as.character(circ_data$p_MR_MTM))
circ_data$ctDNA.Clearance <- factor(circ_data$ctDNA.Clearance, levels=c("Sustained", "Transient", "No Clearance"))
median_p_MR_MTM <- aggregate(p_MR_MTM ~ ctDNA.Clearance, data = circ_data, FUN = median)
print(median_p_MR_MTM)
```

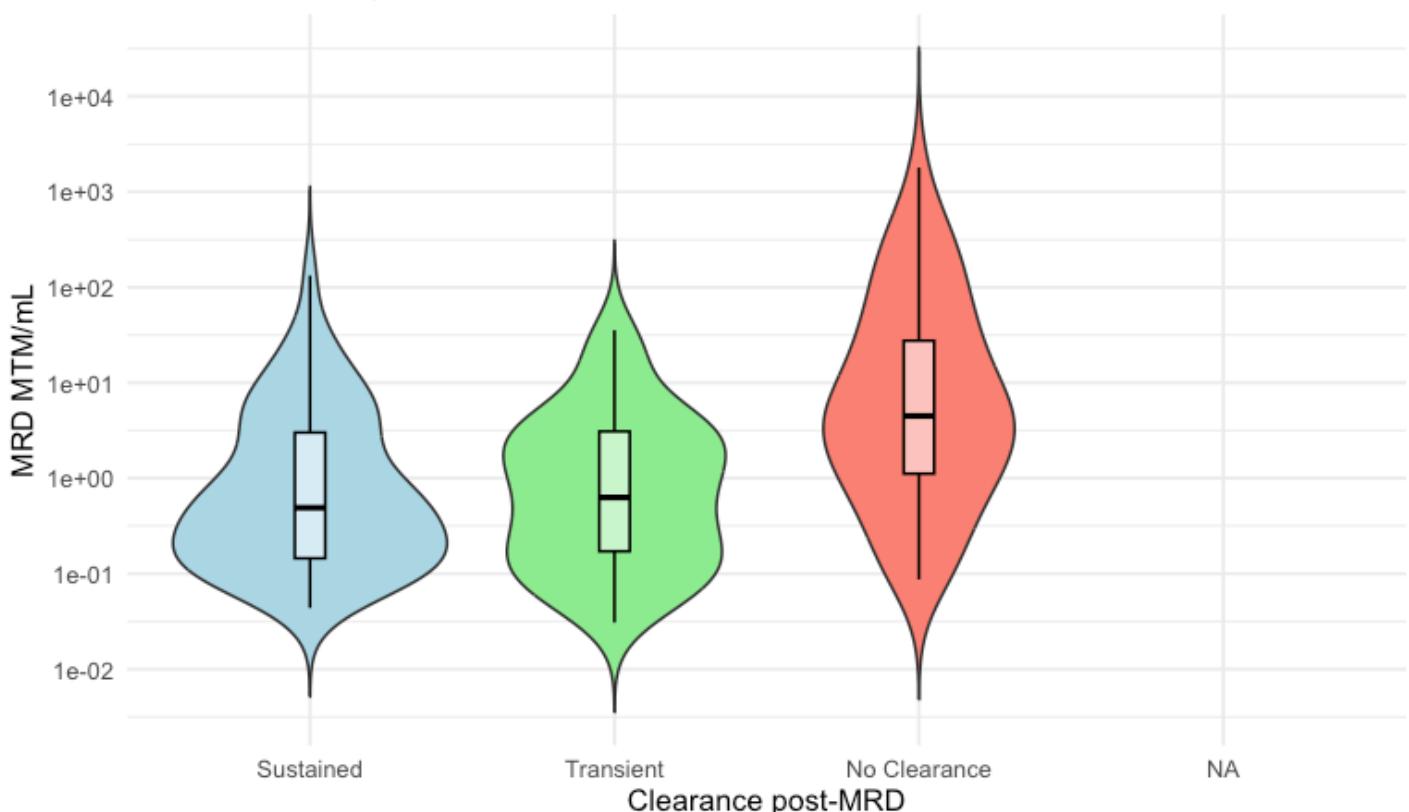
ctDNA.Clearance	p_MR_MTM
<fctr>	<dbl>
Sustained	0.4906143
Transient	0.6286054
No Clearance	4.4903396
3 rows	

Hide

```
# Create violin plot with log10 scale on y-axis
ggplot(circ_data, aes(x=ctDNA.Clearance, y=p_MR_MTM, 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")
```

```
Warning: Removed 1 row containing non-finite outside the scale range (`stat_ydensity()`).
Warning: Removed 1 row containing non-finite outside the scale range (`stat_boxplot()`).
```

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 = 1846, p-value = 0.8611
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 = 894, p-value = 1.655e-06
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_MR_D_MTM by ctDNA.Clearance
W = 782, p-value = 4.905e-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_dataadf <- 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

Counts.Var1	Counts.Freq	Percentages.Var1	Percentages.Freq
<fctr>	<int>	<fctr>	<dbl>
15-18m	6	15-18m	12
18-21m	0	18-21m	0
21-24m	2	21-24m	4
>24m	0	>24m	0
8 rows			

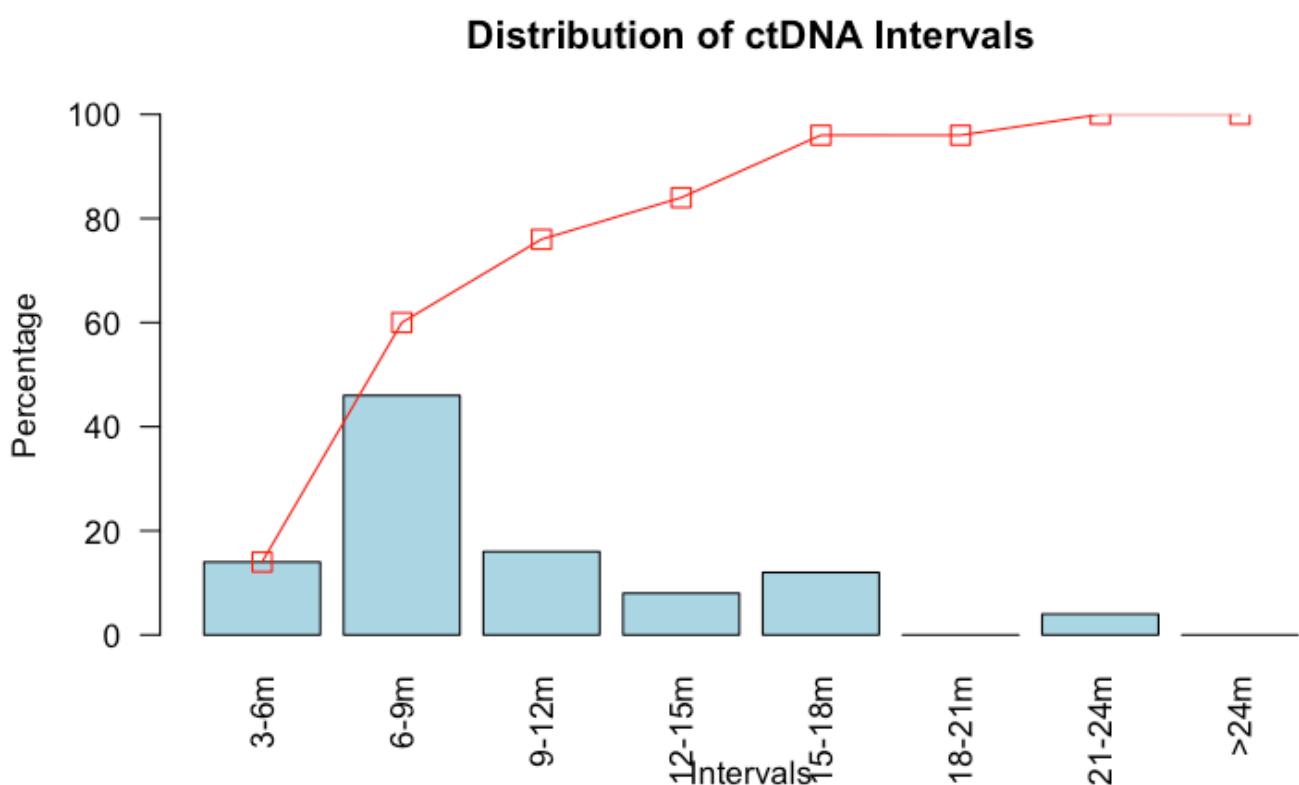
Hide

```
# 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, CumulativePercentages = 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)
```



#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_datadf <- as.data.frame(circ_data)
surv_object <- Surv(time = circ_data$OS.MRD.months, event = circ_data$OS.Event)

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

```

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Clearance=No Clearance	55	17	32.5	23.9	NA
ctDNA.Clearance=Sustained	66	0	NA	NA	NA
ctDNA.Clearance=Transient	57	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	66	0	0.0000000	0.00000
Transient	57	7	0.1228070	12.28070

3 rows

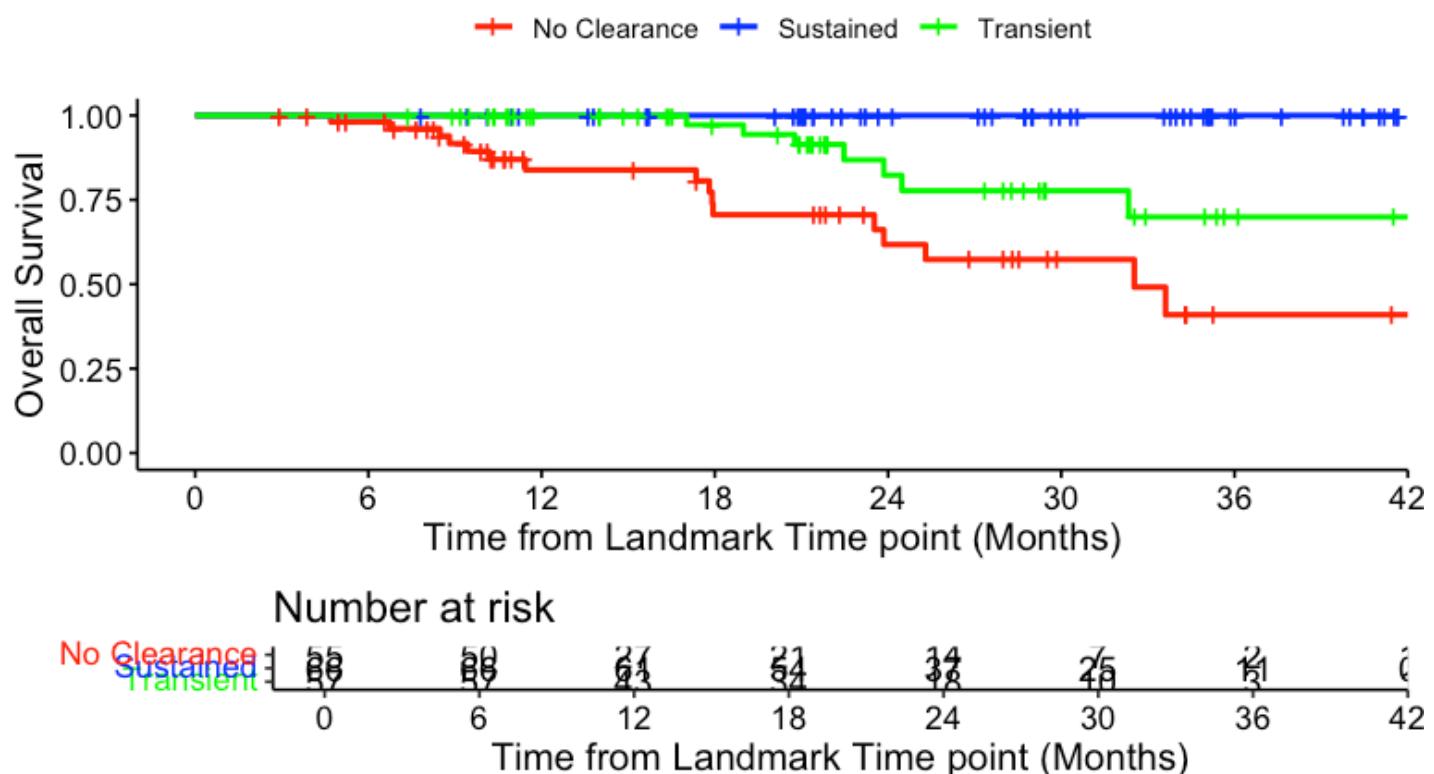
[Hide](#)

```

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



```
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	43	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	3.239767	1.510470	25.52778	3.100228	3315.950	11.56033	6.737400e-04	
ctDNA.ClearanceNo Clearance	4.320421	1.484534	75.22027	10.162345	9600.955	34.62293	4.001653e-09	

Likelihood ratio test=34.63812 on 2 df, p=3.009045e-08, n=178

Wald test = 12.90204 on 2 df, p = 0.001578914

Covariance-Matrix:

	ctDNA.ClearanceTransient	ctDNA.ClearanceNo Clearance
ctDNA.ClearanceTransient	2.281518	2.139113
ctDNA.ClearanceNo Clearance	2.139113	2.203843

#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_datadf <- as.data.frame(circ_data)

# Convert days to months
#circ_data$PostMRDPos.months <- circ_data$PostMRDPos / 30.437

# Define the intervals: 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)

```

0-6m	6-9m	9-12m	12-15m	15-18m	18-21m	21-24m	>24m
77	35	23	2	20	1	7	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	Counts.Freq	Percentages.Var1	Percentages.Freq	TotalObservations
<fctr>	<int>	<fctr>	<dbl>	<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, CumulativePercentages = 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	Counts.Freq	Percentages.Var1	Percentages.Freq	CumulativePercentages
<fctr>	<int>	<fctr>	<dbl>	<dbl>
0-6m	77	0-6m	46.6666667	46.6666667
6-9m	35	6-9m	21.2121212	67.8787878
9-12m	23	9-12m	13.9393939	81.8181818
12-15m	2	12-15m	1.2121212	83.0303030
15-18m	20	15-18m	12.1212121	95.1515151
18-21m	1	18-21m	0.6060606	95.7575757
21-24m	7	21-24m	4.2424242	100.0000000
>24m	0	>24m	0.0000000	100.0000000

8 rows | 1-6 of 6 columns

```
#Percentages of MRD negative with molecular recurrence (returned positive) post-MRD - ACT treated
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD != "" & circ_data$Lead.Time >= 0, ]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data <- circ_data[circ_data$PostMRDPos.Event=="TRUE",]
circ_data <- circ_data[circ_data$ACT=="TRUE",]
circ_datadf <- 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)

```

0–6m	6–9m	9–12m	12–15m	15–18m	18–21m	21–24m	>24m
26	12	15	1	7	0	5	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	Counts.Freq	Percentages.Var1	Percentages.Freq	TotalObservations
<fctr>	<int>	<fctr>	<dbl>	<int>
0–6m	26	0–6m	39.393939	NA
6–9m	12	6–9m	18.181818	NA
9–12m	15	9–12m	22.727273	NA
12–15m	1	12–15m	1.515152	NA
15–18m	7	15–18m	10.606061	NA
18–21m	0	18–21m	0.000000	NA
21–24m	5	21–24m	7.575758	NA
>24m	0	>24m	0.000000	66

8 rows

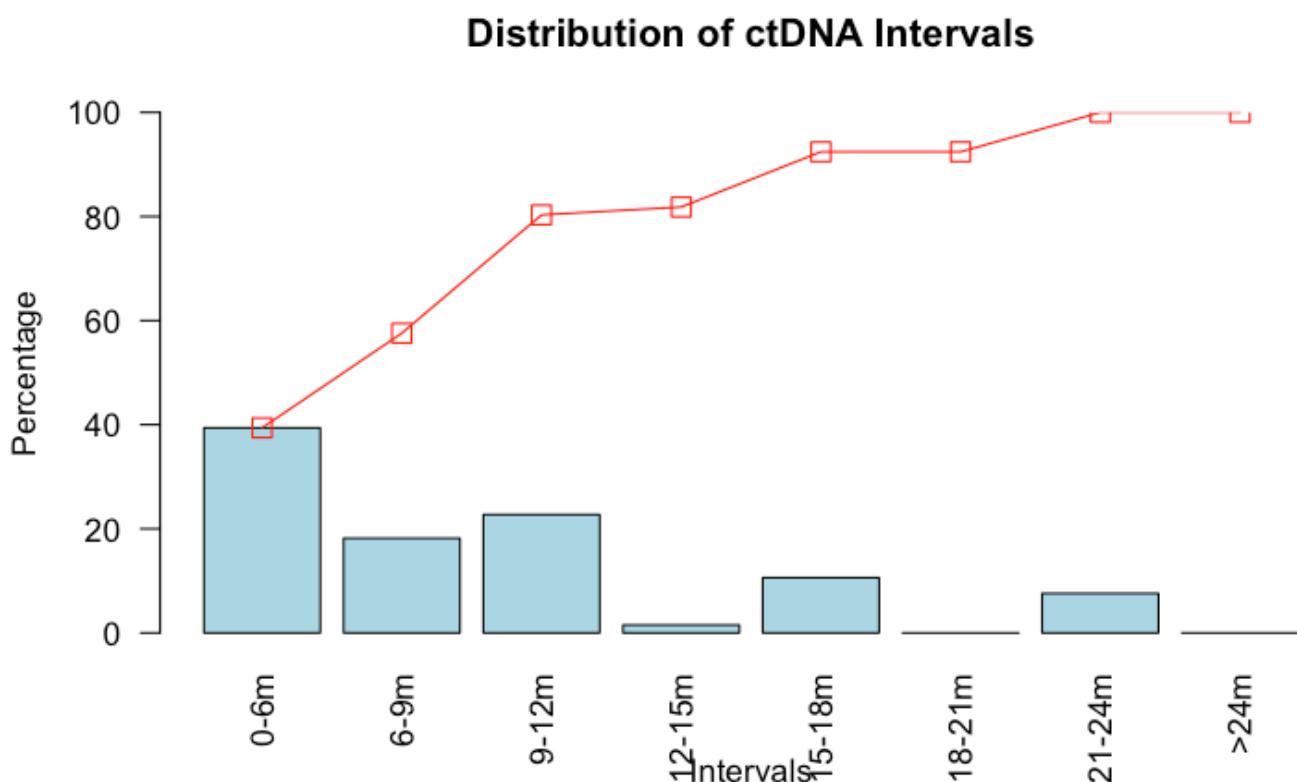
Hide

```
# 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, CumulativePercentages = 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	Counts.Freq	Percentages.Var1	Percentages.Freq	CumulativePercentages
<fctr>	<int>	<fctr>	<dbl>	<dbl>
0-6m	0-6m	26 0-6m	39.393939	39.393939
6-9m	6-9m	12 6-9m	18.181818	57.575758
9-12m	9-12m	15 9-12m	22.727273	80.30303
12-15m	12-15m	1 12-15m	1.515152	81.818182
15-18m	15-18m	7 15-18m	10.606061	92.424242

Counts.Var1 <fctr>	Counts.Freq <int>	Percentages.Var1 <fctr>	Percentages.Freq <dbl>	CumulativePercentage <dbl>
18-21m	0	18-21m	0.000000	92.42
21-24m	5	21-24m	7.575758	100.00
>24m	0	>24m	0.000000	100.00

8 rows | 1-6 of 6 columns

#Percentages of MRD negative with molecular recurrence (returned positive) post-MRD - Observation cohort

[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$PostMRDPos.Event=="TRUE",]
circ_data <- circ_data[circ_data$ACT=="FALSE",]
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)
```

0-6m	6-9m	9-12m	12-15m	15-18m	18-21m	21-24m	>24m
51	23	8	1	13	1	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)

# 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	Counts.Freq	Percentages.Var1	Percentages.Freq	TotalObservations
<fctr>	<int>	<fctr>	<dbl>	<int>
0-6m	51	0-6m	51.515152	NA
6-9m	23	6-9m	23.232323	NA
9-12m	8	9-12m	8.080808	NA
12-15m	1	12-15m	1.010101	NA
15-18m	13	15-18m	13.131313	NA
18-21m	1	18-21m	1.010101	NA
21-24m	2	21-24m	2.020202	NA
>24m	0	>24m	0.000000	99

8 rows

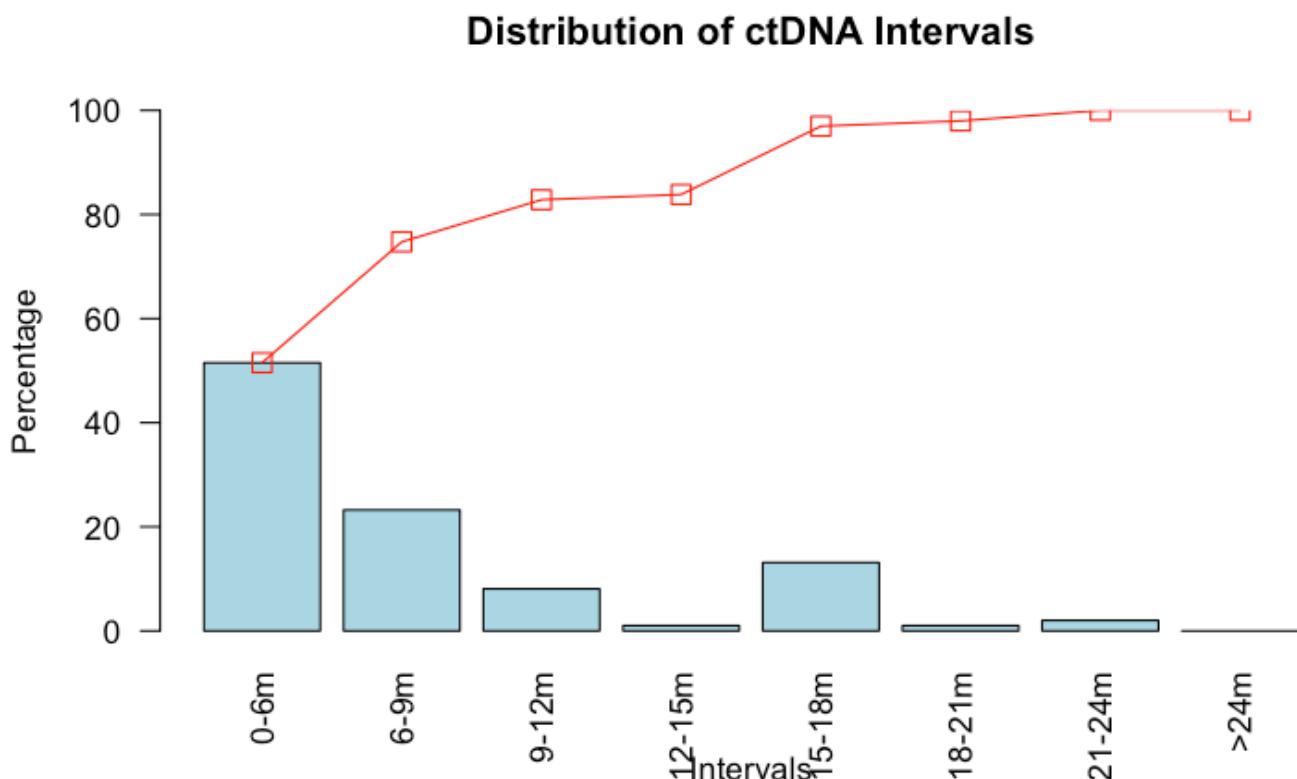
Hide

```
# 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, CumulativePercentages = 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	Counts.Freq	Percentages.Var1	Percentages.Freq	CumulativePercentage
<fctr>	<int>	<fctr>	<dbl>	<dbl>
0-6m	0-6m	51 0-6m	51.515152	51.515152
6-9m	6-9m	23 6-9m	23.232323	74.747474
9-12m	9-12m	8 9-12m	8.080808	82.828282
12-15m	12-15m	1 12-15m	1.010101	83.838383
15-18m	15-18m	13 15-18m	13.131313	96.969696
18-21m	18-21m	1 18-21m	1.010101	97.979797
21-24m	21-24m	2 21-24m	2.020202	100.000000
>24m	>24m	0 >24m	0.000000	100.000000

8 rows | 1-6 of 6 columns

#Statistical analysis (proportions z-test) for Molecular Recurrence Proportions in ACT vs Observation

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible == "TRUE" & circ_data$ctDNA.MRD != "" & circ_data$Lead.Time >= 0,]
circ_data <- circ_data[circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$PostMRDPos.Event == "TRUE",]

# Define the intervals: 0-3, 3-6, 6-9, 9-12, 12-15, 15-18, 18-21, 21-24, >24 months
breaks <- c(0, 3, 6, 9, 12, 15, 18, 21, 24, 48)
labels <- c("10w-3m", "3-6m", "6-9m", "9-12m", "12-15m", "15-18m", "18-21m", "21-24m",
">24m")
circ_data$p_drelReturned_intervals <- cut(circ_data$PostMRDPos.months, breaks = breaks,
labels = labels, right = FALSE)
circ_data$ACT_STATUS <- ifelse(circ_data$ACT == "TRUE", "ACT_TRUE", "ACT_FALSE")
contingency_table <- table(circ_data$p_drelReturned_intervals, circ_data$ACT_STATUS)
cumulative_counts_ACT_TRUE <- cumsum(contingency_table[, "ACT_TRUE"])
cumulative_counts_ACT_FALSE <- cumsum(contingency_table[, "ACT_FALSE"])
total_ACT_TRUE <- sum(contingency_table[, "ACT_TRUE"])
total_ACT_FALSE <- sum(contingency_table[, "ACT_FALSE"])
cumulative_percentages_ACT_TRUE <- cumulative_counts_ACT_TRUE / total_ACT_TRUE * 100
cumulative_percentages_ACT_FALSE <- cumulative_counts_ACT_FALSE / total_ACT_FALSE * 100

# Perform the proportion test for cumulative percentages in each interval
test_results <- lapply(1:length(cumulative_counts_ACT_TRUE), function(i) {
  prop.test(
    x = c(cumulative_counts_ACT_TRUE[i], cumulative_counts_ACT_FALSE[i]),
    n = c(total_ACT_TRUE, total_ACT_FALSE),
    correct = FALSE
  )
})

```

```

Warning in stats::prop.test(x = x, n = n, p = p, alternative = alternative, :
  Chi-squared approximation may be incorrect
Warning in stats::prop.test(x = x, n = n, p = p, alternative = alternative, :
  Chi-squared approximation may be incorrect
Warning in stats::prop.test(x = x, n = n, p = p, alternative = alternative, :
  Chi-squared approximation may be incorrect
Warning in stats::prop.test(x = x, n = n, p = p, alternative = alternative, :
  Chi-squared approximation may be incorrect

```

Hide

```

p_values <- sapply(test_results, function(test) test$p.value)
test_statistics <- sapply(test_results, function(test) test$statistic)
interval_labels <- labels
results_df <- data.frame(
  Interval = interval_labels,
  Cumulative_Counts_ACT_TRUE = cumulative_counts_ACT_TRUE,
  Cumulative_Counts_ACT_FALSE = cumulative_counts_ACT_FALSE,
  Cumulative_Percentages_ACT_TRUE = cumulative_percentages_ACT_TRUE,
  Cumulative_Percentages_ACT_FALSE = cumulative_percentages_ACT_FALSE,
  Test_Statistic = test_statistics,
  P_Value = p_values
)
print(results_df)

```

Interval	Cumulative_Counts_ACT_TRUE	Cumulative_Counts_ACT_FALSE
<chr>	<int>	<int>
10w-3m	13	21
3-6m	26	51
6-9m	38	74
9-12m	53	82
12-15m	54	83
15-18m	61	96
18-21m	61	97
21-24m	66	99
>24m	66	99

9 rows | 1-4 of 7 columns

#DFS 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 values
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$DFS.MRD.months>=0,]
survfit(Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

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

321 observations deleted due to missingness
n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 1294      76      NA      NA      NA
ctDNA.Dynamics=2  159      120  10.74  10.22  15.2
ctDNA.Dynamics=3  336      263   5.34   4.83   6.7
```

[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)
```

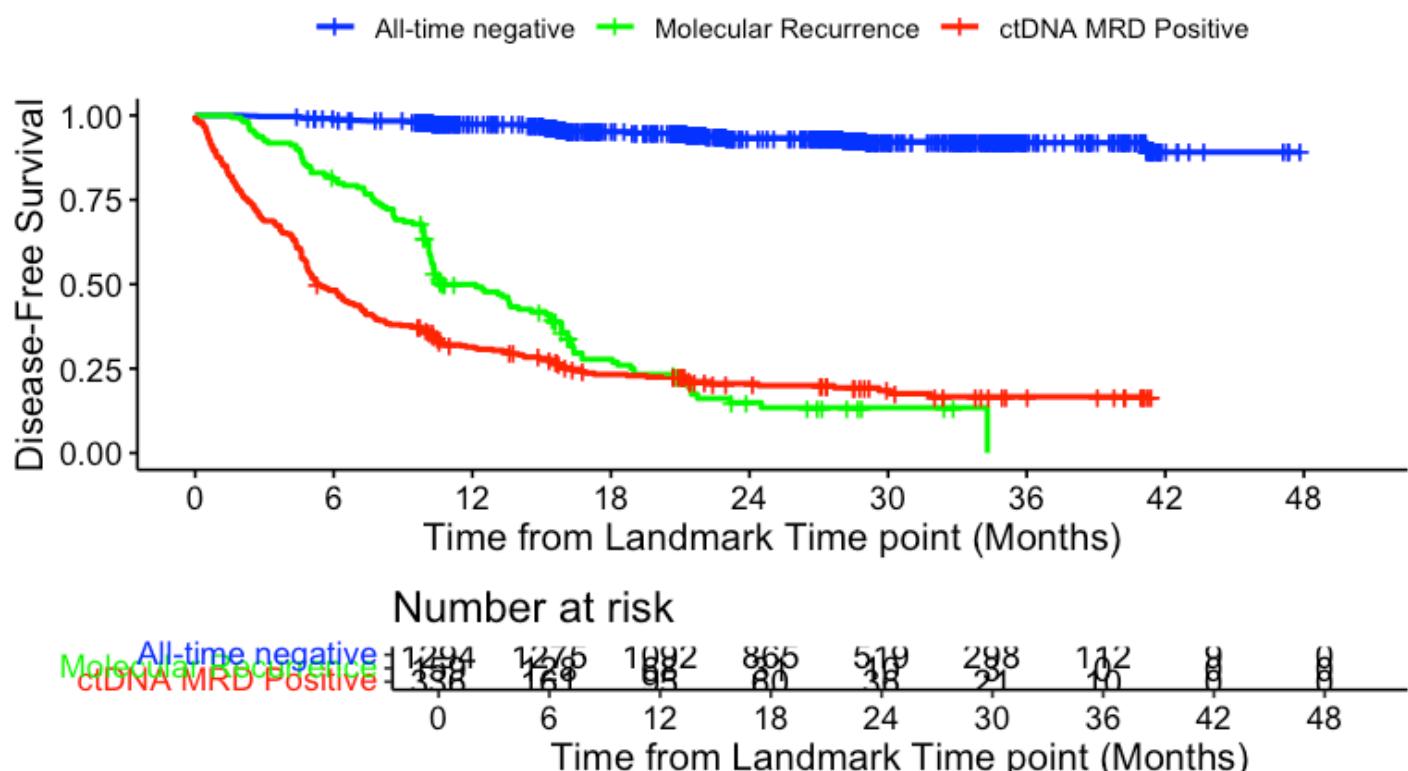
ctDNA.Dynamics	Total	Events	Fraction	Percentage
	<dbl>	<int>	<dbl>	<dbl>
1	1294	76	0.05873261	5.873261
2	159	120	0.75471698	75.471698
3	336	263	0.78273810	78.273810
NA	321	NA	NA	NA

4 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.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","green","red"), title="DFS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Surveillance Window", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("All-time negative","Molecular Recurrence", "ctDNA MRD Positive"), legend.title="")
```

DFS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Survey



```
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    1092      33    0.974 0.00451      0.963      0.981
    24     519      36    0.934 0.00796      0.916      0.948

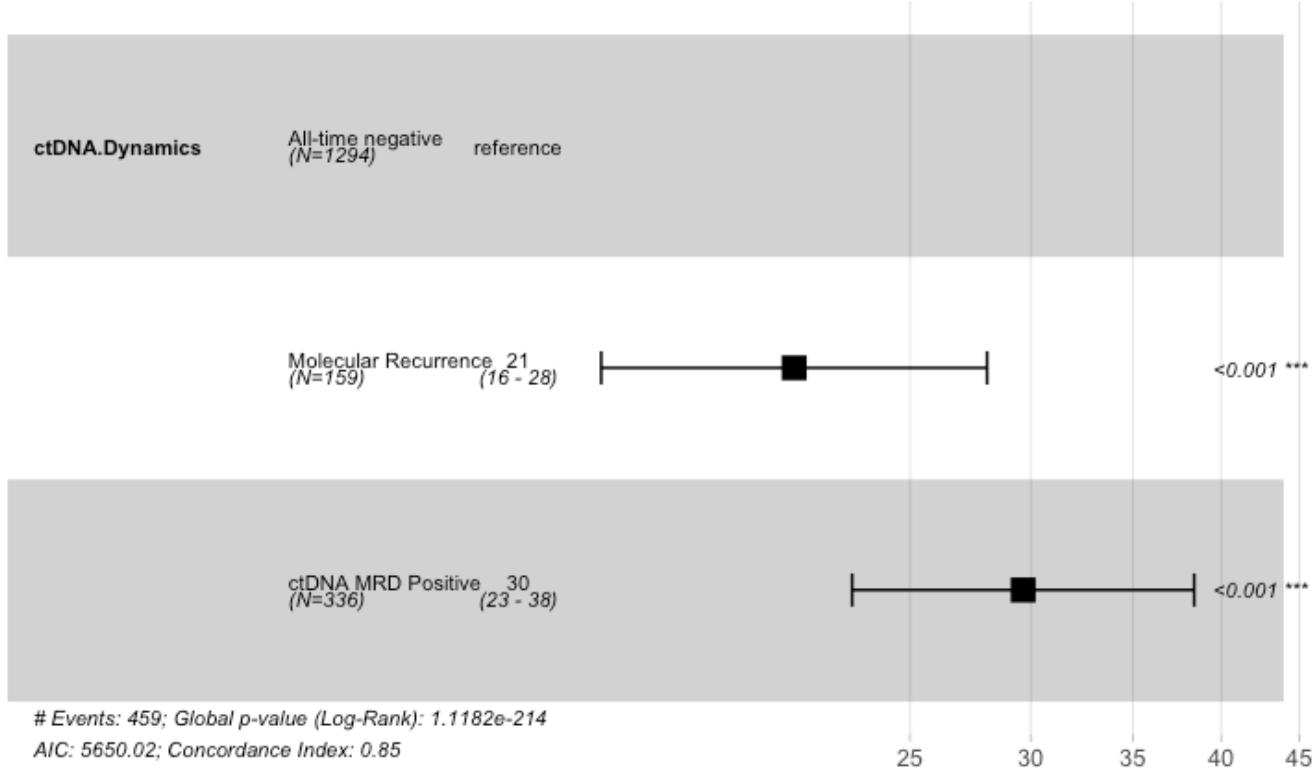
  ctDNA.Dynamics=2
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12      68      78    0.499 0.0403      0.4180      0.575
    24      10      40    0.149 0.0350      0.0886      0.224

  ctDNA.Dynamics=3
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12      95     229    0.314 0.0255      0.265      0.364
    24      36      29    0.206 0.0236      0.161      0.254
```

[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)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

Call:

coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 1789, number of events= 459

(321 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.DynamicsMolecular Recurrence	3.0439	20.9872	0.1487	20.47	<2e-16 ***
ctDNA.DynamicsctDNA MRD Positive	3.3898	29.6590	0.1318	25.72	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.DynamicsMolecular Recurrence	20.99	0.04765	15.68	28.09
ctDNA.DynamicsctDNA MRD Positive	29.66	0.03372	22.91	38.40

Concordance= 0.851 (se = 0.008)

Likelihood ratio test= 985.3 on 2 df, p=<2e-16

Wald test = 674.9 on 2 df, p=<2e-16

Score (logrank) test = 1431 on 2 df, p=<2e-16

Hide

```

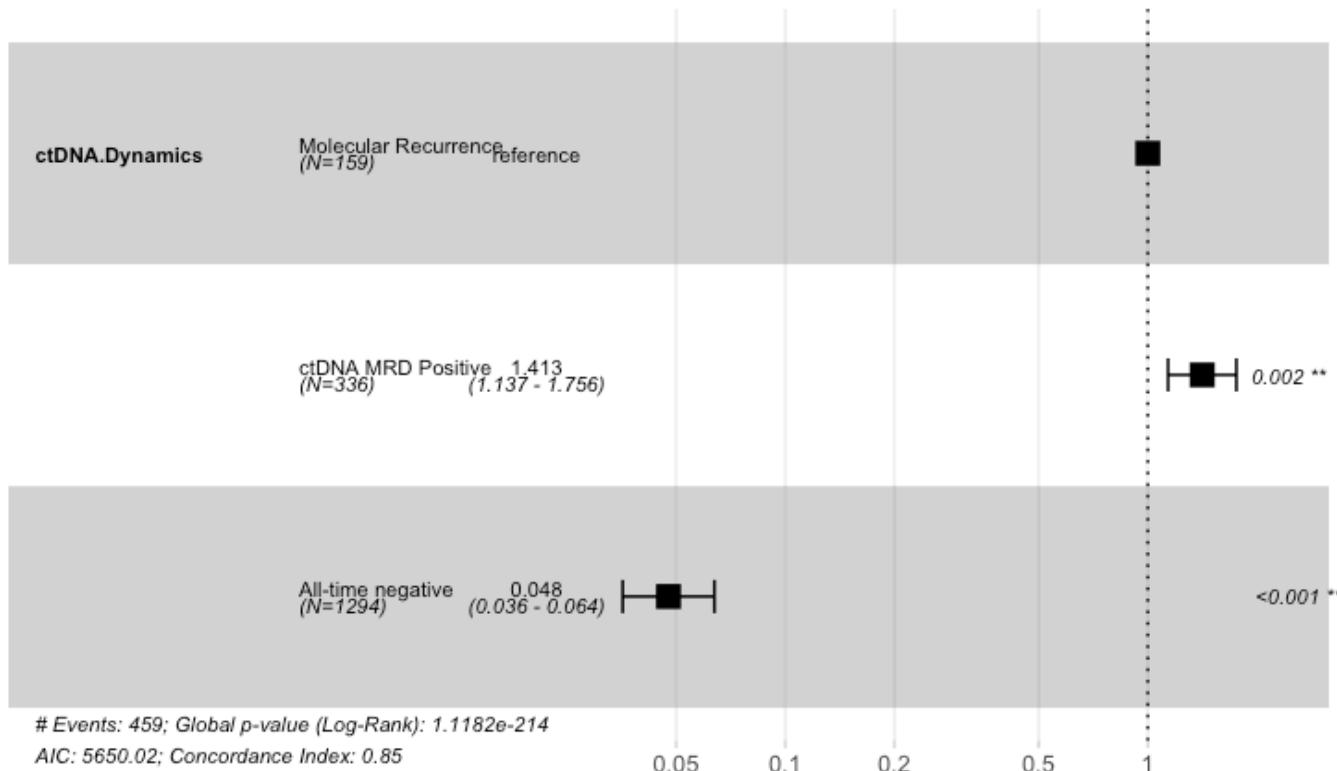
rm(list=ls()) #repeat to compare Molecular Recurrence vs ctDNA MRD positive
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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$DFS.MRD.months>=0,]
surv_object <- Surv(time = circ_data$DFS.MRD.months, event = circ_data$DFS.Event)
circ_data$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= 459  
(321 observations deleted due to missingness)
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.DynamicsctDNA MRD Positive	0.34585	1.41319	0.11077	3.122	0.00179 **
ctDNA.DynamicsAll-time negative	-3.04391	0.04765	0.14867	-20.474	< 2e-16 ***

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.DynamicsctDNA MRD Positive	1.41319	0.7076	1.1374	1.75585
ctDNA.DynamicsAll-time negative	0.04765	20.9872	0.0356	0.06377

```
Concordance= 0.851 (se = 0.008 )
```

```
Likelihood ratio test= 985.3 on 2 df, p=<2e-16
```

```
Wald test = 674.9 on 2 df, p=<2e-16
```

```
Score (logrank) test = 1431 on 2 df, p=<2e-16
```

```
#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 values
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	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<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

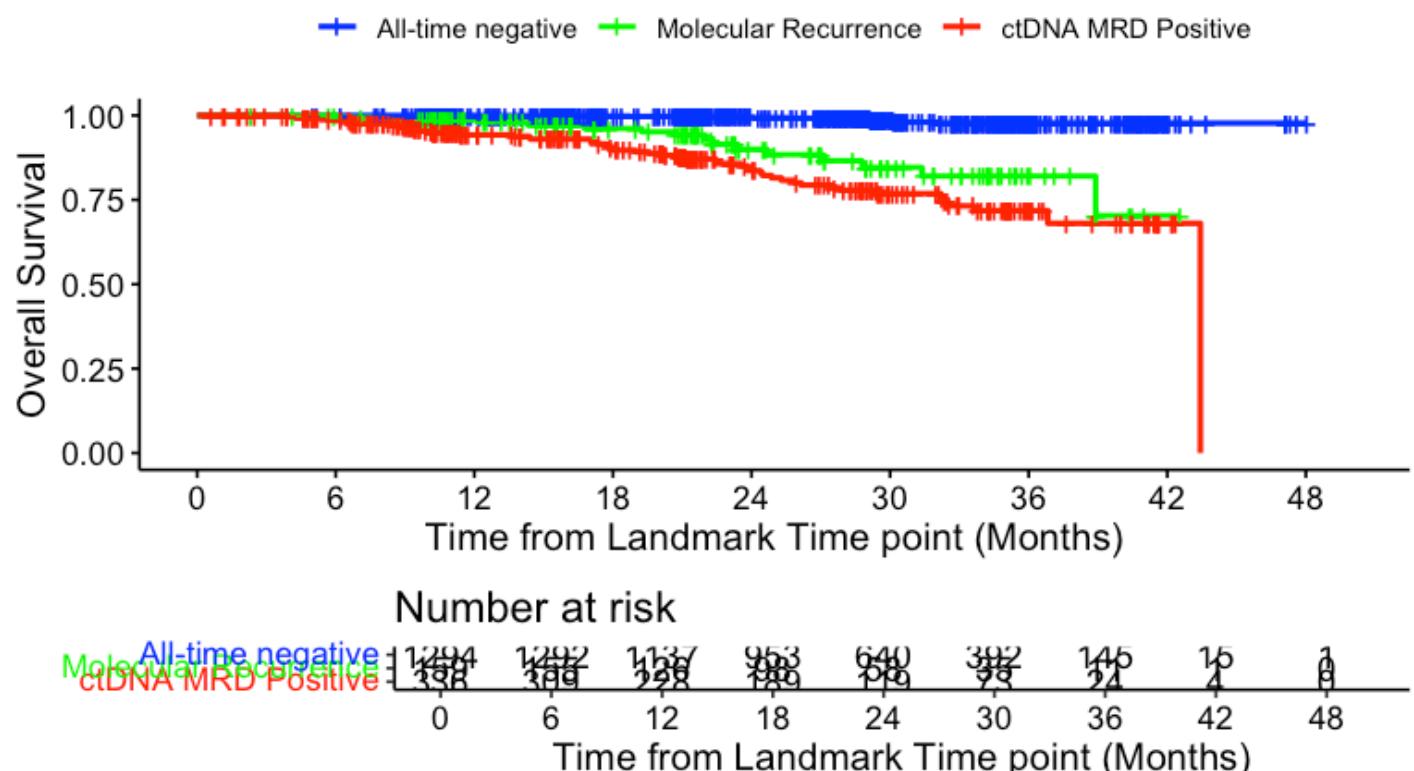
Hide

```

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=c("blue","green","red"), title="OS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Surveillance Window", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("All-time negative","Molecular Recurrence", "ctDNA MRD Positive"), legend.title="")

```

OS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Surveil



```
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")

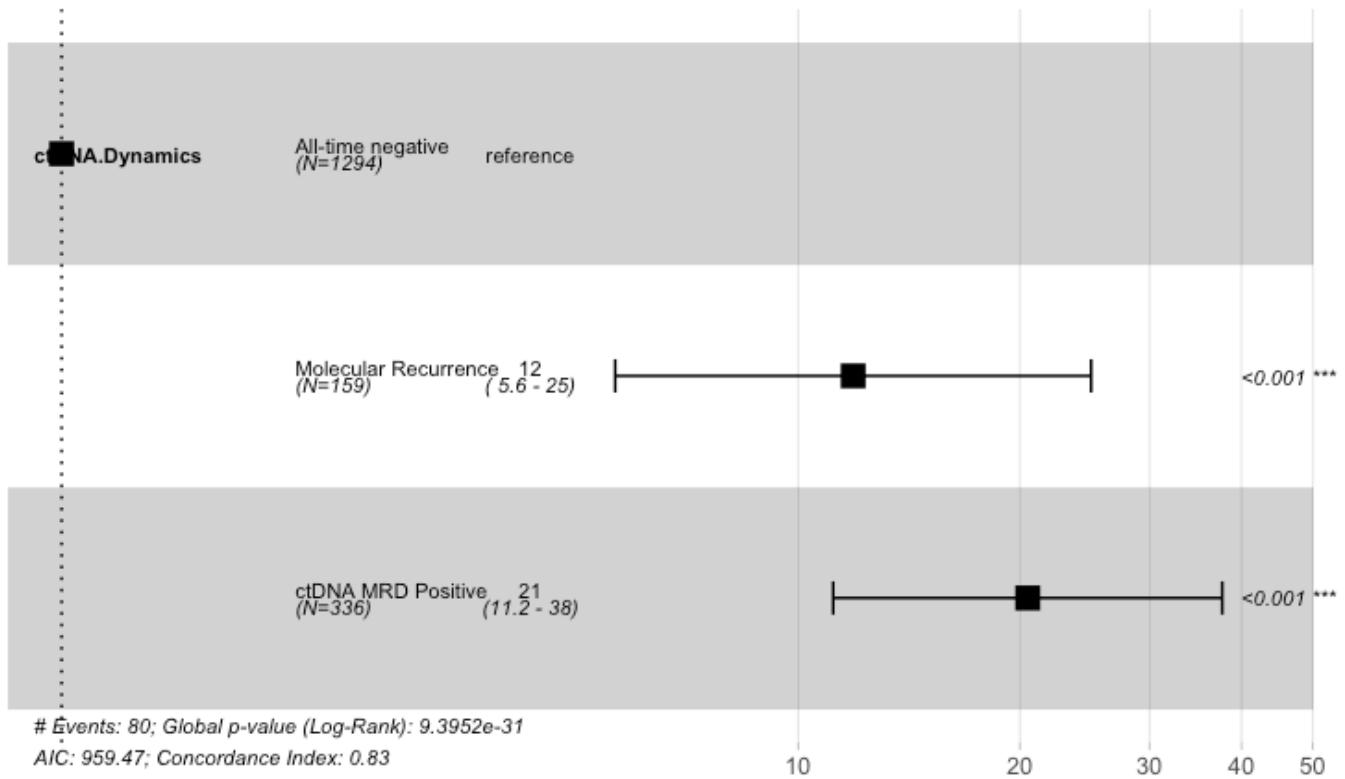
321 observations deleted due to missingness
  ctDNA.Dynamics=1
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12    1137      0    1.000  0.0000          NA          NA
    24     640       5    0.995  0.0023      0.988      0.998

  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
    24     58       8    0.900  0.03138     0.817      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
    24    119      20    0.837  0.0258     0.778      0.881
```

```
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)
```

Hazard ratio



```
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.DynamicsMolecular Recurrence	2.4747	11.8787	0.3796	6.519	7.09e-11 ***
ctDNA.DynamicsctDNA MRD Positive	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.DynamicsMolecular Recurrence	11.88	0.08418	5.644	25.00
ctDNA.DynamicsctDNA MRD Positive	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

```

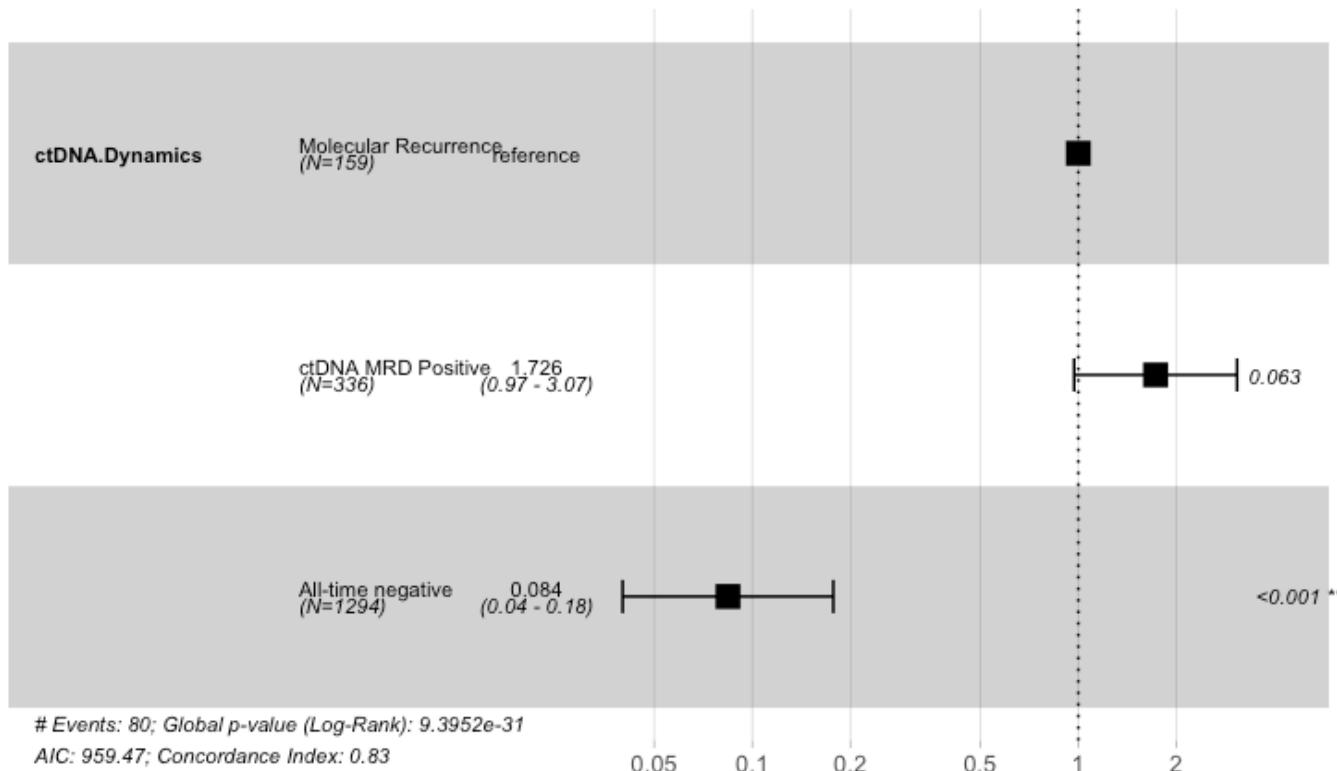
rm(list=ls()) #repeat to compare Molecular Recurrence vs ctDNA MRD positive
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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.Dynamics	ctDNA MRD Positive	0.54572	1.72584	0.29355	1.859	0.063 .
ctDNA.Dynamics	All-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.Dynamics	ctDNA MRD Positive	1.72584	0.5794	0.9708	3.0681
ctDNA.Dynamics	All-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
```

```
#PRS 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="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.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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$OS.MRD.months>=0,]
survfit(Surv(time = circ_data$PRS.months, event = circ_data$OS.Event)~ctDNA.Dynamics, data = circ_data)

```

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

35 observations deleted due to missingness
 n events median 0.95LCL 0.95UCL

ctDNA.Dynamics	1	2	NA	NA	NA
ctDNA.Dynamics=1	65	2	NA	NA	NA
ctDNA.Dynamics=2	120	15	36.3	36.3	NA
ctDNA.Dynamics=3	263	52	38.2	29.2	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	65	2	0.03076923	3.076923
2	120	15	0.12500000	12.500000
3	263	52	0.19771863	19.771863
NA	35	NA	NA	NA

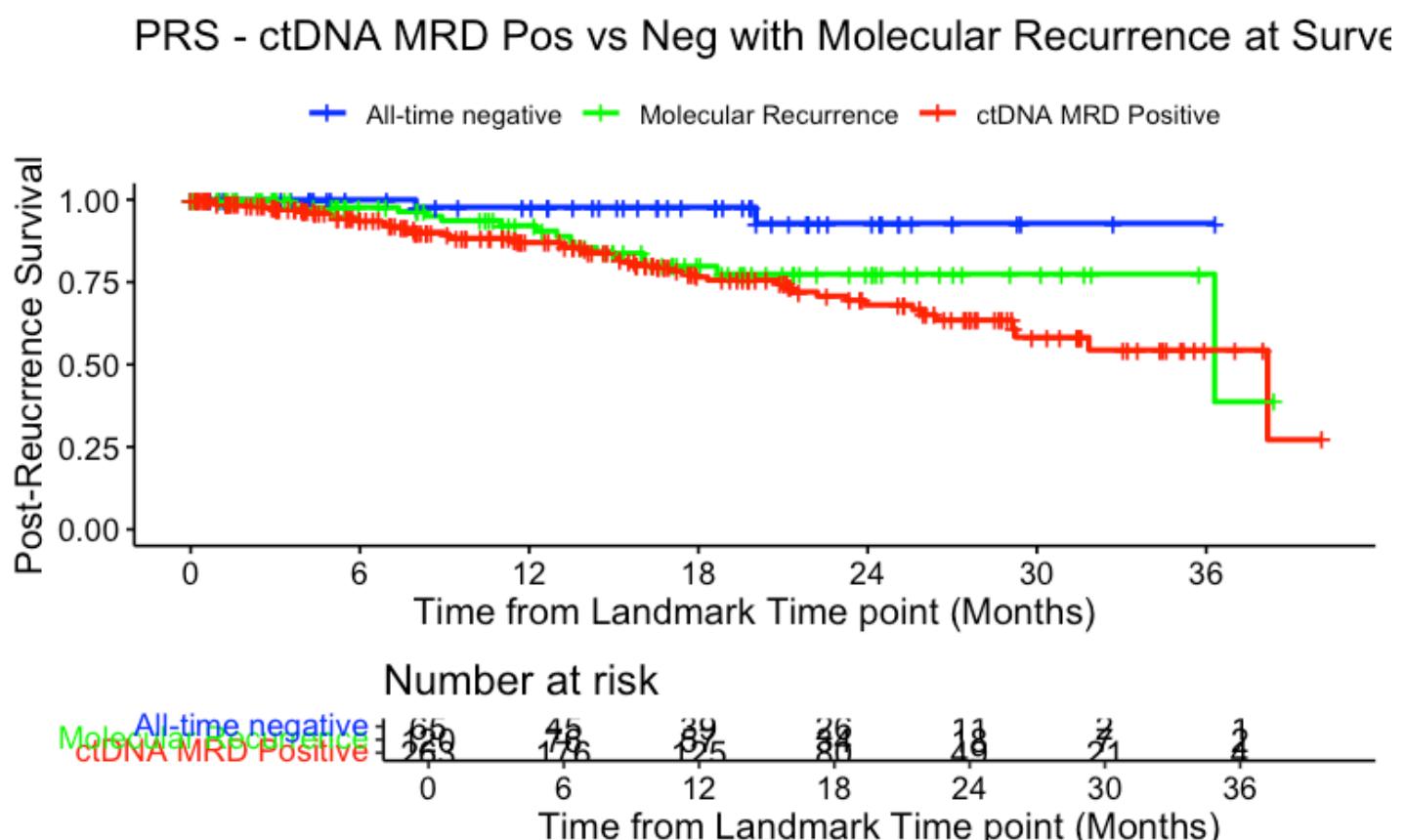
4 rows

Hide

```

surv_object <- Surv(time = circ_data$PRS.months, event = circ_data$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("blue","green","red"), title="PRS - ctDNA MRD Pos vs Neg with Molecular Recurrence at Surveillance Window", ylab= "Post-Reurrence Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("All-time negative","Molecular 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")

35 observations deleted due to missingness
  ctDNA.Dynamics=1
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12     39      1    0.977  0.0225      0.849      0.997
    24     11      1    0.928  0.0522      0.721      0.983

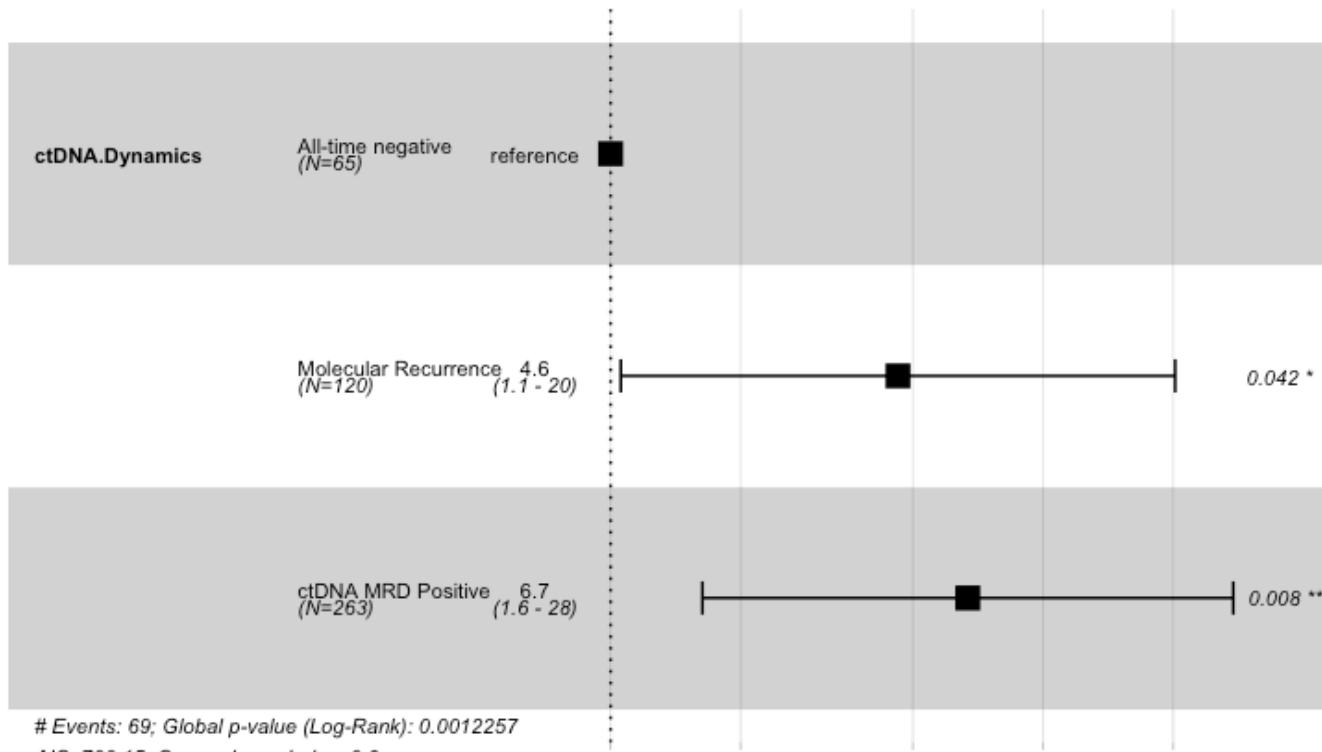
  ctDNA.Dynamics=2
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12     57      6    0.922  0.0309      0.833      0.964
    24     18      8    0.774  0.0550      0.644      0.862

  ctDNA.Dynamics=3
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12    125     25    0.870  0.0246      0.813      0.911
    24     49     20    0.681  0.0435      0.587      0.758
```

[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)
```

Hazard ratio

[Hide](#)

```
summary(cox_fit)
```

Call:

coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 448, number of events= 69

(35 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)	
ctDNA.Dynamics	Molecular Recurrence	1.5303	4.6196	0.7531	2.032	0.04216 *
ctDNA.Dynamics	ctDNA MRD Positive	1.9024	6.7020	0.7211	2.638	0.00834 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95	
ctDNA.Dynamics	Molecular Recurrence	4.620	0.2165	1.056	20.21
ctDNA.Dynamics	ctDNA MRD Positive	6.702	0.1492	1.631	27.54

Concordance= 0.598 (se = 0.027)

Likelihood ratio test= 13.41 on 2 df, p=0.001

Wald test = 8.02 on 2 df, p=0.02

Score (logrank) test = 10.1 on 2 df, p=0.006

Hide

```

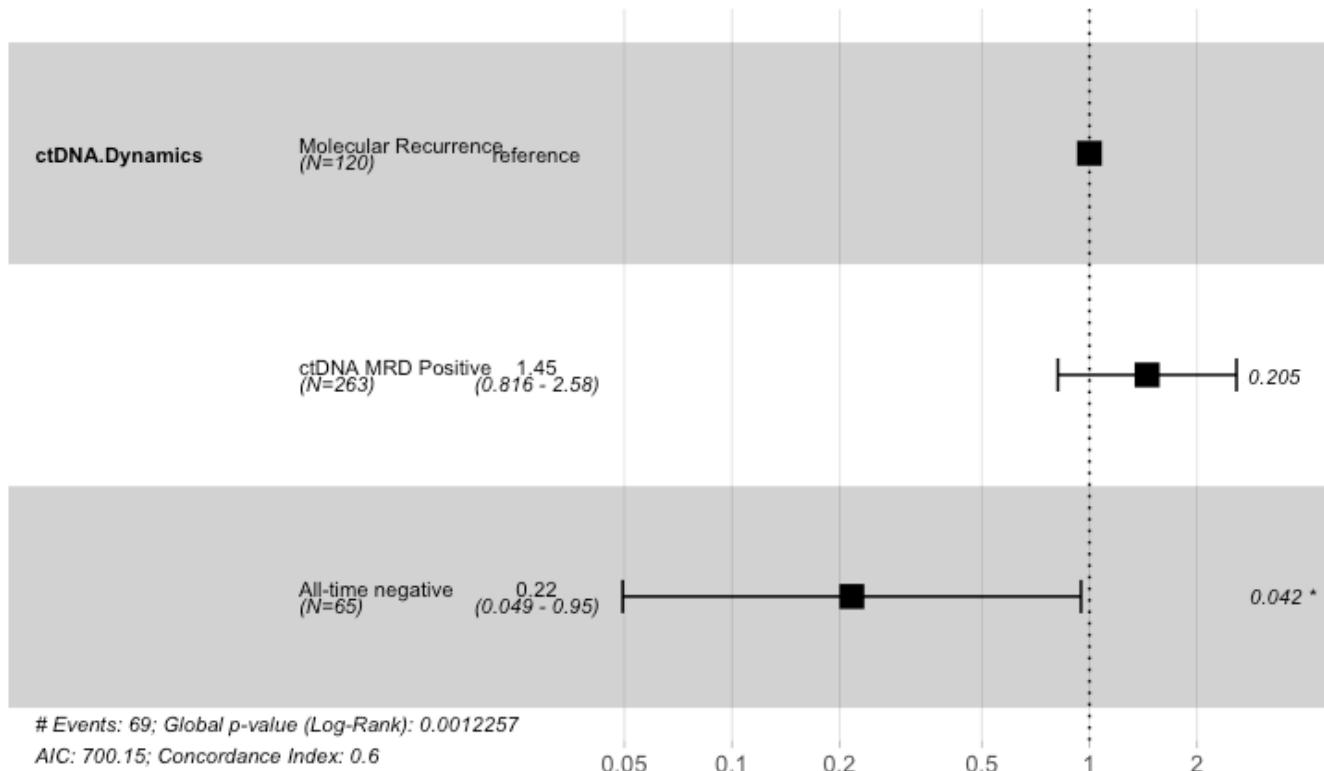
rm(list=ls()) #repeat to compare Molecular Recurrence vs ctDNA MRD positive
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$OS.MRD.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
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
  ))

surv_object <- Surv(time = circ_data$PRS.months, event = circ_data$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= 448, number of events= 69

(35 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.DynamicsctDNA MRD Positive	0.3721	1.4508	0.2935	1.268	0.2048
ctDNA.DynamicsAll-time negative	-1.5303	0.2165	0.7531	-2.032	0.0422 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.DynamicsctDNA MRD Positive	1.4508	0.6893	0.81623	2.5787
ctDNA.DynamicsAll-time negative	0.2165	4.6196	0.04947	0.9472

Concordance= 0.598 (se = 0.027)

Likelihood ratio test= 13.41 on 2 df, p=0.001

Wald test = 8.02 on 2 df, p=0.02

Score (logrank) test = 10.1 on 2 df, p=0.006

#DFS 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$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_data$df <- 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](#)

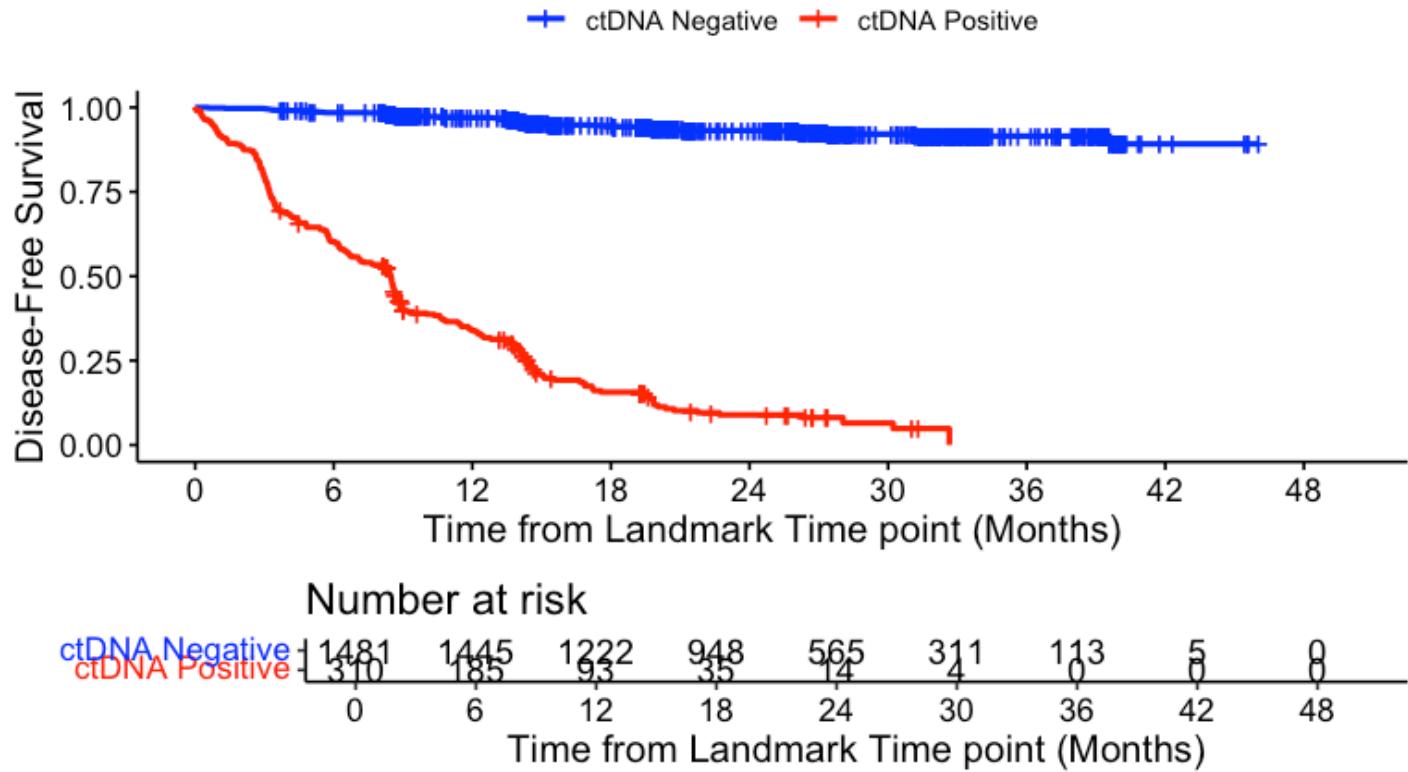
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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



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

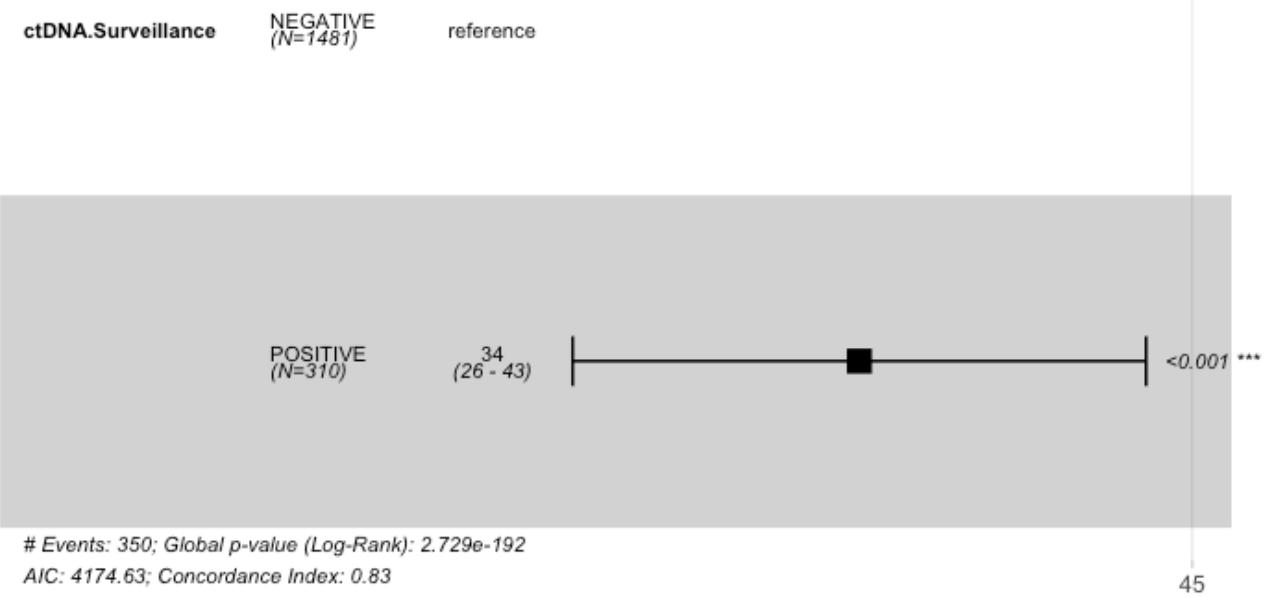
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	565	81	0.932	0.00756	0.915	0.945		
30	311	5	0.922	0.00878	0.902	0.937		
36	113	2	0.915	0.00975	0.894	0.933		

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		

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= 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)
```

```
# 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 = 33.56 (26.07-43.2); p = 0"
```

```
#DFS by ctDNA at the Surveillance Window - High Risk Stages II/III 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 <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

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

```

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

3 observations deleted due to missingness

		n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	1200	49	NA	NA	NA	NA
ctDNA.Surveillance=POSITIVE	186	153	8.8	8.51	10.3	

[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	1200	49	0.04083333	4.083333
POSITIVE	186	153	0.82258065	82.258065
NA	3	NA	NA	NA

3 rows

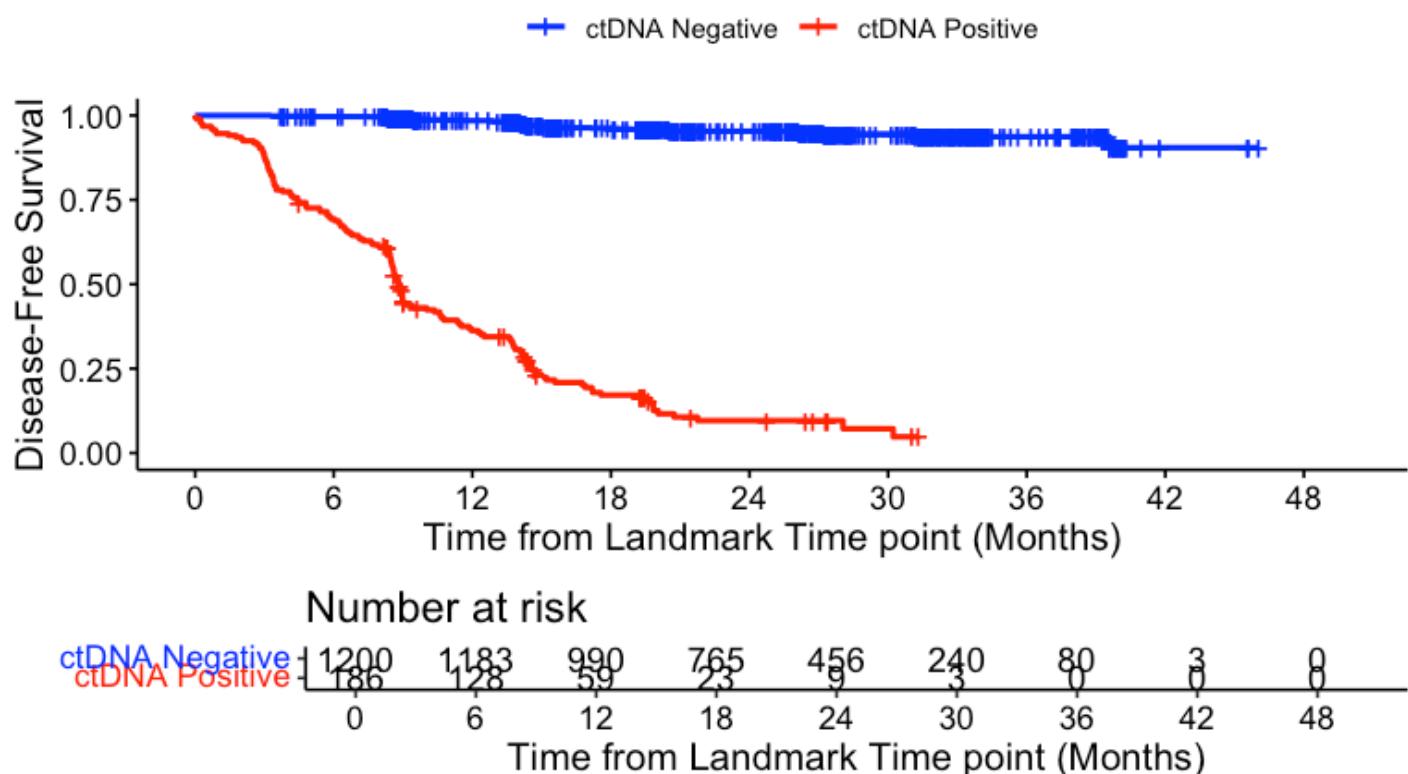
[Hide](#)

```

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 | High Risk Stage II-III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```

DFS - ctDNA Surveillance window | High Risk Stage II-III



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")
```

3 observations deleted due to missingness

ctDNA.Surveillance=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	456	43	0.953	0.00711	0.937		0.965	
30	240	3	0.945	0.00845	0.926		0.960	
36	80	2	0.937	0.01018	0.914		0.954	

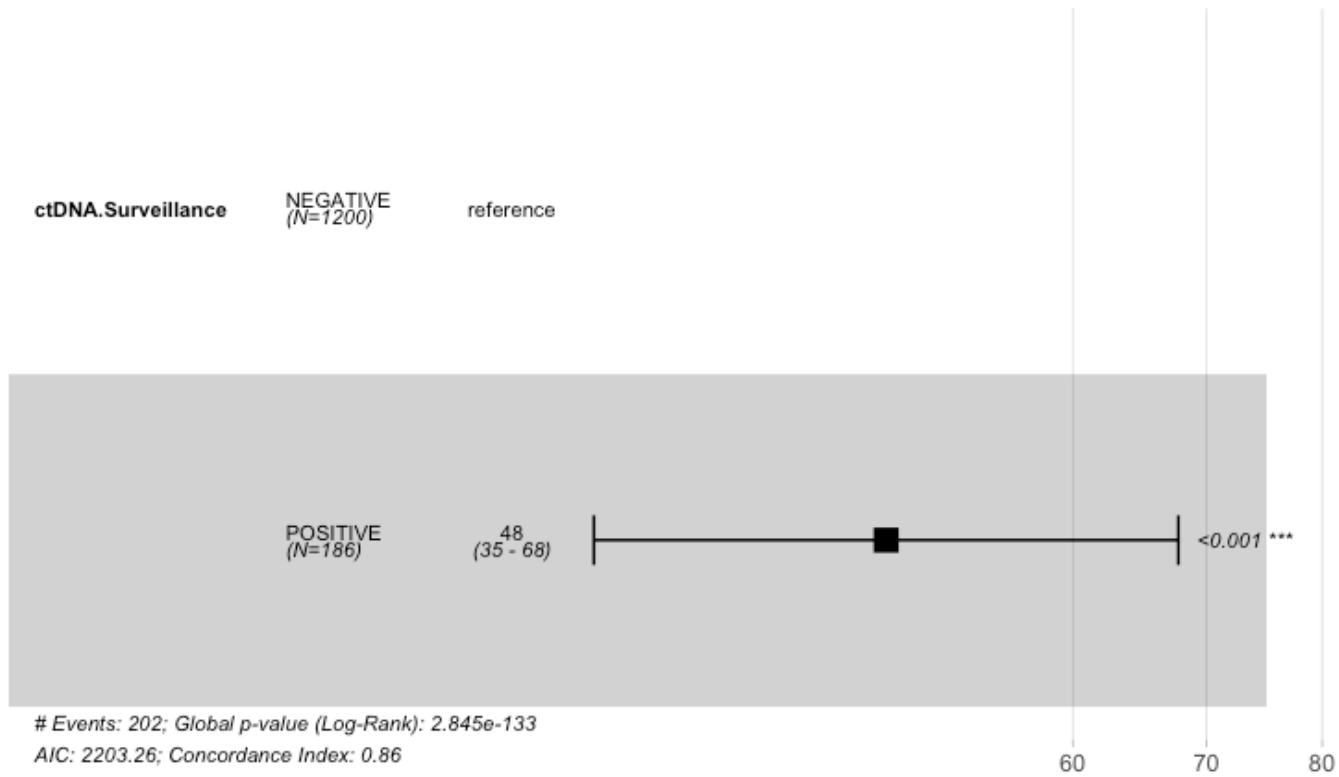
ctDNA.Surveillance=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	9	151	0.0959	0.0265	0.0521		0.156	
30	3	1	0.0719	0.0288	0.0289		0.142	

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= 1386, number of events= 202
(3 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.SurveillancePOSITIVE	3.8788	48.3678	0.1721	22.54	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.SurveillancePOSITIVE	48.37	0.02067	34.52	67.77

Concordance= 0.859 (se = 0.013)
Likelihood ratio test= 603.5 on 1 df, p=<2e-16
Wald test = 508 on 1 df, p=<2e-16
Score (logrank) test = 1376 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 = 48.37 (34.52-67.77); p = 0"
```

#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_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	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)
```

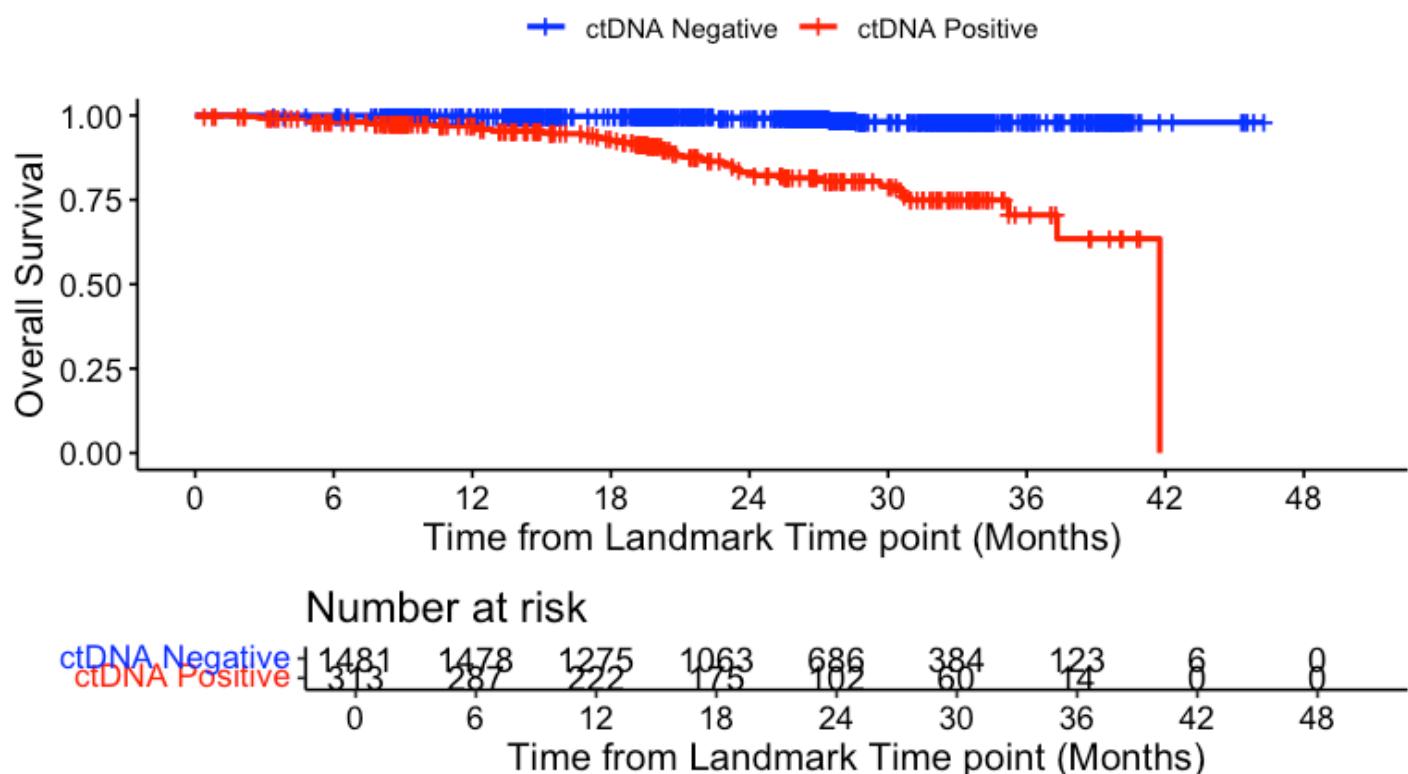
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1481	13	0.008777853	0.8777853
POSITIVE	313	41	0.130990415	13.0990415

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=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="")
```

OS - 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	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		

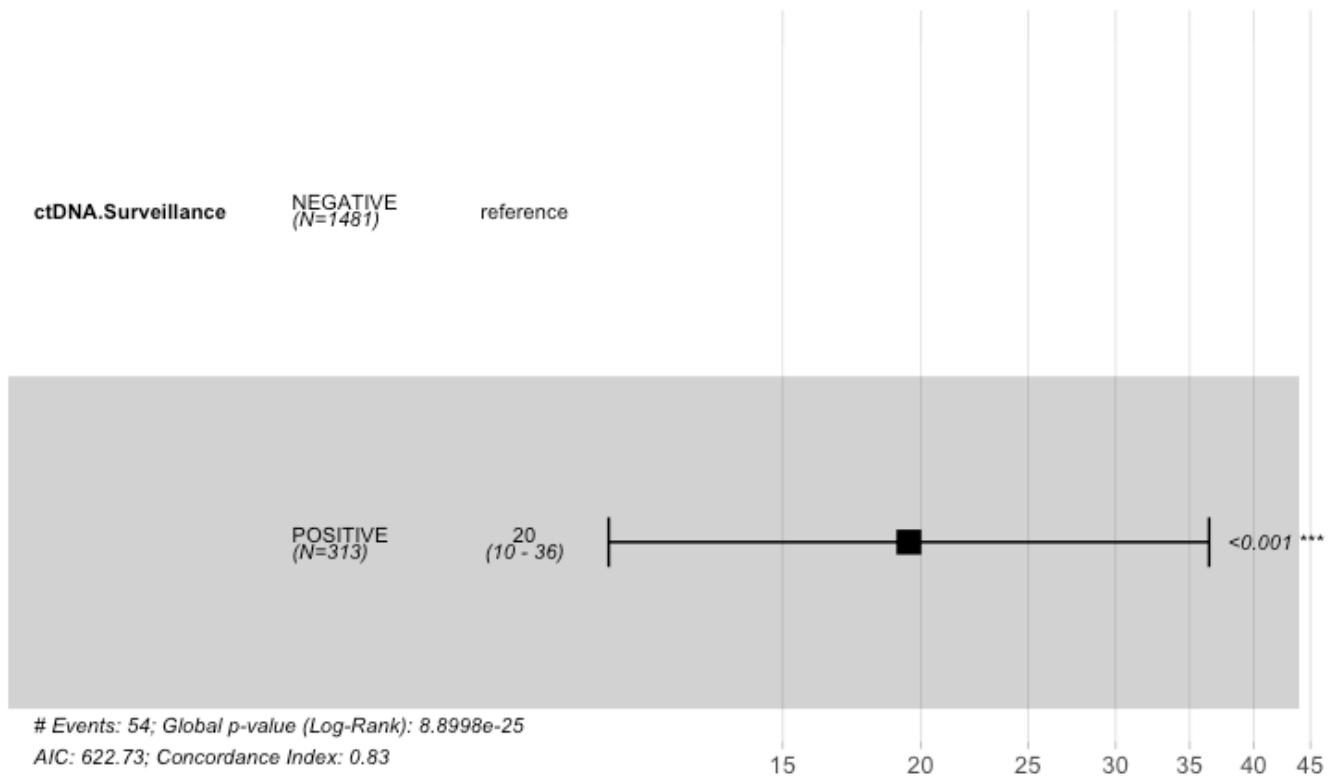
ctDNA.Surveillance=POSITIVE

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)

# 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"
```

#OS by ctDNA at the Surveillance Window - High Risk Stages II/III 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 <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$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)

3 observations deleted due to missingness

n events median 0.95LCL 0.95UCL

ctDNA.Surveillance	1200	8	NA	NA	NA
ctDNA.Surveillance=POSITIVE	186	24	NA	35.2	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)
```

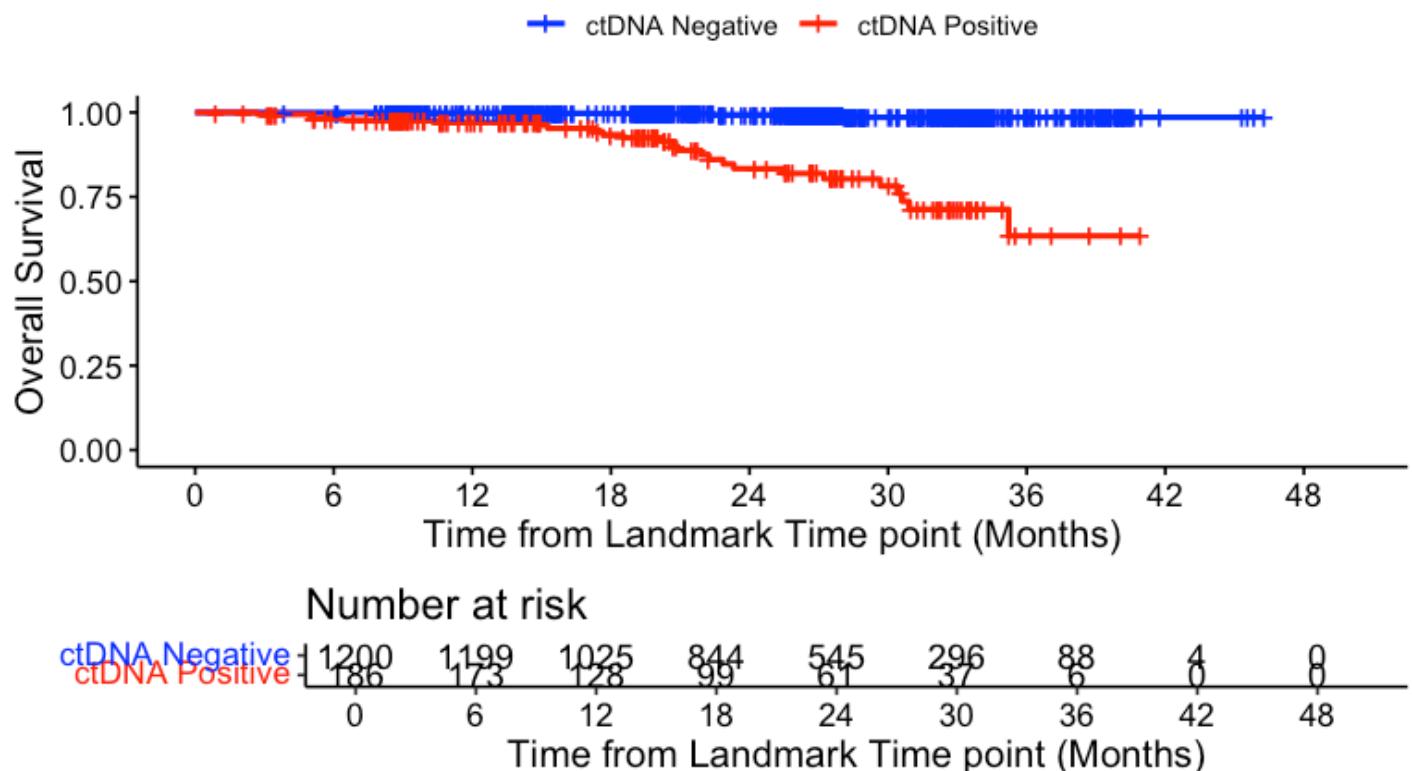
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	1200	8	0.006666667	0.6666667
POSITIVE	186	24	0.129032258	12.9032258
NA	3	NA	NA	NA

3 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 - ctDNA Surveillance window | High Risk Stage II-III", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - ctDNA Surveillance window | High Risk Stage II-III



```
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")
```

3 observations deleted due to missingness

ctDNA.Surveillance=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	545	5	0.994	0.00291	0.984	0.997	0.997	0.997
30	296	3	0.986	0.00548	0.970	0.993	0.993	0.993
36	88	0	0.986	0.00548	0.970	0.993	0.993	0.993

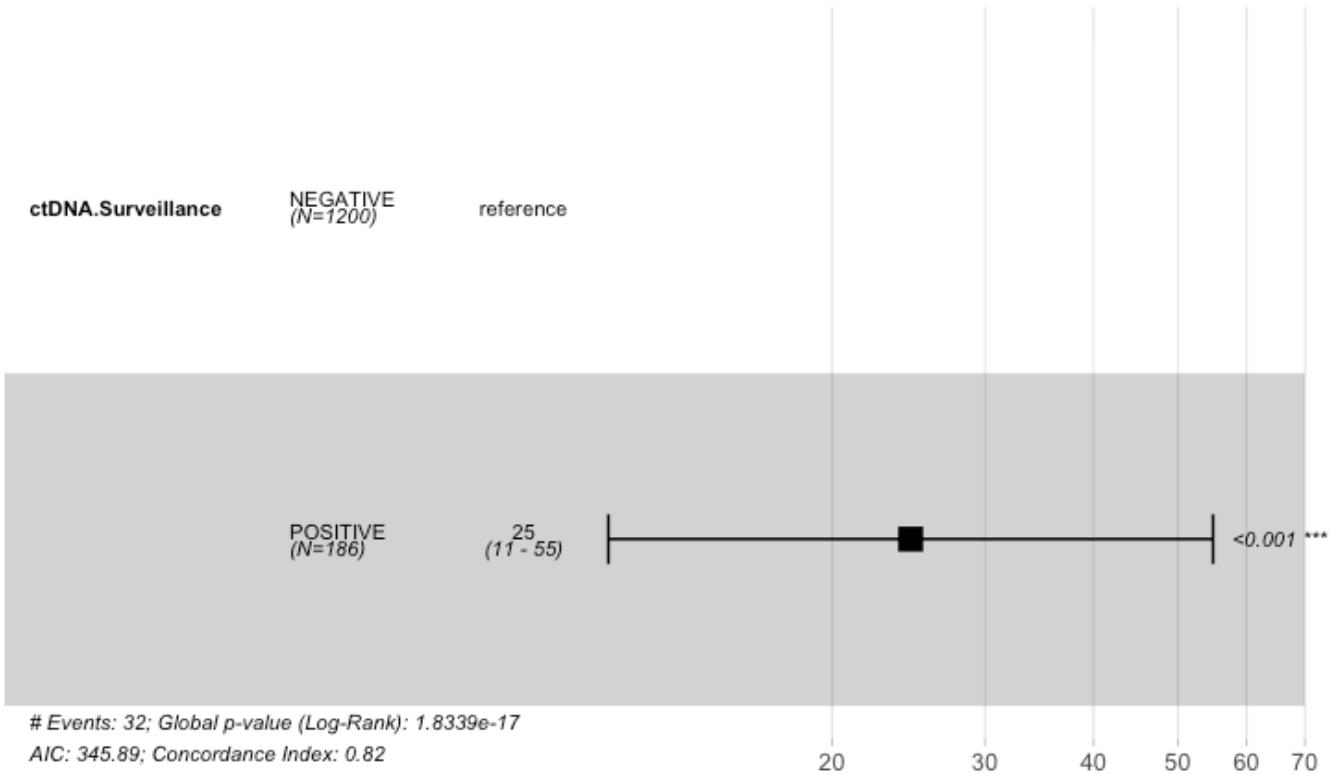
ctDNA.Surveillance=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	61	17	0.834	0.0391	0.740	0.896	0.896	0.896
30	37	3	0.782	0.0470	0.672	0.859	0.859	0.859
36	6	4	0.634	0.0904	0.431	0.781	0.781	0.781

[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= 1386, number of events= 32
(3 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE  3.2044   24.6401   0.4086 7.842 4.43e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE     24.64     0.04058     11.06     54.88

Concordance= 0.823 (se = 0.04 )
Likelihood ratio test= 72.32 on 1 df,  p=<2e-16
Wald test             = 61.5 on 1 df,  p=4e-15
Score (logrank) test = 135.3 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 = 24.64 (11.06-54.88); p = 0"
```

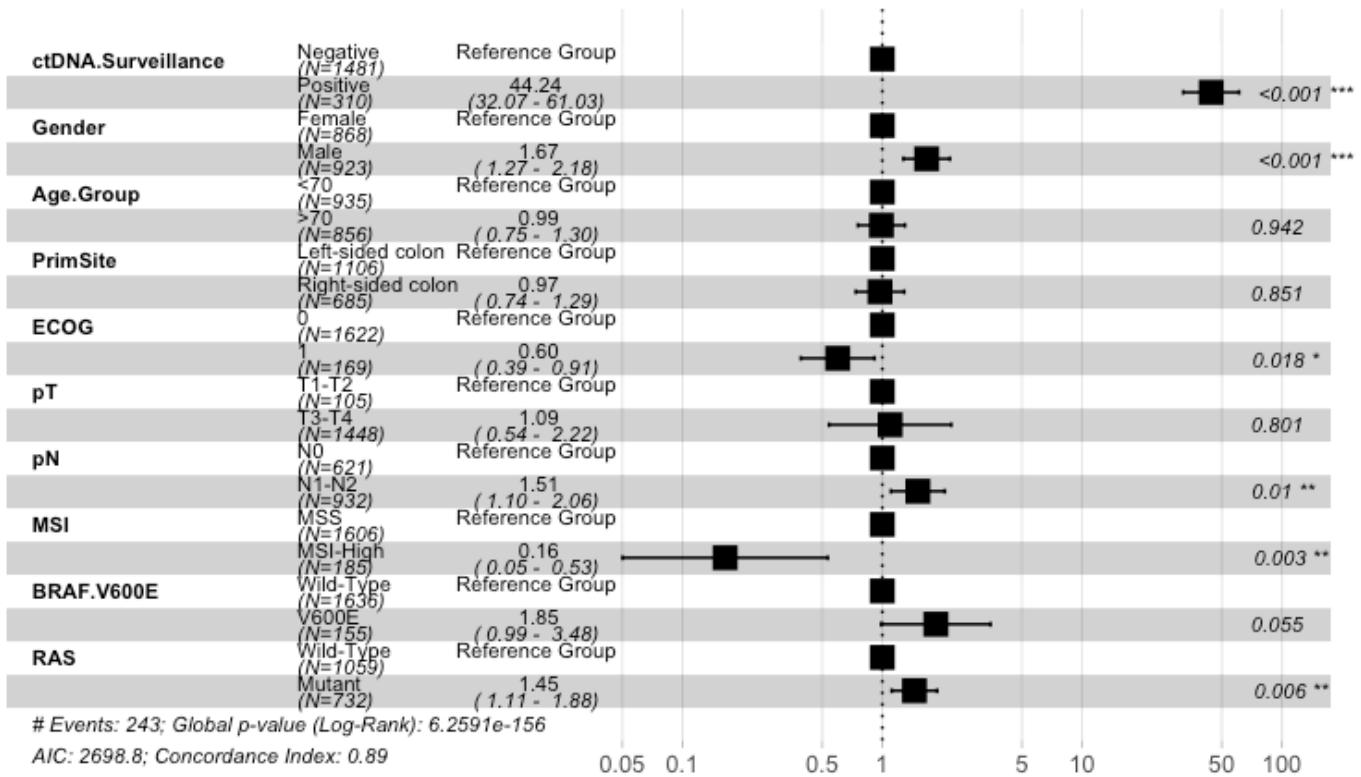
#Multivariate cox regression at Surveillance Window for DFS - All stages Landmark 10 weeks

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$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- 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.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



Hide

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

```
#Surveillance Window - Sensitivity and Specificity calculations - All Cohorts
```

Hide

```
#All Patients
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$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

```
[1] "Sensitivity - All pts: 77.1929824561403"
```

Hide

```
print(paste("Specificity - All pts: ", specificity*100))
```

```
[1] "Specificity - All pts: 96.6253443526171"
```

Hide

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - All pts: 84.3450479233227"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - All pts: 94.7332883187036"
```

[Hide](#)

```
#Stage I Patients
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=="I",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage I: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage I: 0"
```

[Hide](#)

```
print(paste("Specificity - Stage I: ", specificity*100))
```

```
[1] "Specificity - Stage I: 100"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage I: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage I: NaN"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage I: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage I: 96.6666666666667"
```

[Hide](#)

```
#Stage II Patients
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=="II",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 70.7317073170732"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 96.5317919075144"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 61.7021276595745"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 97.6608187134503"
```

[Hide](#)

```
#Stage III Patients
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=="III",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 80"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 97.489539748954"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 87.3239436619718"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 95.7534246575343"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 78.4615384615385"
```

[Hide](#)

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 97.2292191435768"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 82.258064516129"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 96.5"
```

[Hide](#)

```
#Stage IV Patients
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$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

[1] "Sensitivity - Stage IV: 76.551724137931"

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

[1] "Specificity - Stage IV: 93.048128342246"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - Stage IV: 89.5161290322581"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - Stage IV: 83.6538461538462"

#Surveillance Window - Sensitivity and Specificity calculations - non ACT treated

[Hide](#)

```
#All Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

```
[1] "Sensitivity - All pts: 78.894472361809"
```

[Hide](#)

```
print(paste("Specificity - All pts: ", specificity*100))
```

```
[1] "Specificity - All pts: 95.8960328317373"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - All pts: 83.9572192513369"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - All pts: 94.3472409152086"
```

[Hide](#)

```
#Stage I Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="I",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage I: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage I:  0"
```

[Hide](#)

```
print(paste("Specificity - Stage I: ", specificity*100))
```

```
[1] "Specificity - Stage I:  100"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage I: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage I:  NaN"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage I: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage I:  96.666666666667"
```

[Hide](#)

```
#Stage II Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="II",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 73.3333333333333"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 96.401028277635"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 61.111111111111"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 97.911227154047"
```

[Hide](#)

```
#Stage III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="III",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 90.1960784313726"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 95.8333333333333"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 85.1851851851852"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 97.3544973544974"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 85"
```

[Hide](#)

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 96.4944649446495"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 78.1609195402299"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 97.7570093457944"
```

[Hide](#)

```
#Stage IV Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==FALSE,]
circ_data <- circ_data[circ_data$Stage=="IV",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

[1] "Sensitivity - Stage IV: 76.0683760683761"

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

[1] "Specificity - Stage IV: 93.3884297520661"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - Stage IV: 91.7525773195876"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - Stage IV: 80.1418439716312"

#Surveillance Window - Sensitivity and Specificity calculations - ACT treated

[Hide](#)

```
#All Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - All pts: ", sensitivity*100))
```

[1] "Sensitivity - All pts: 74.8251748251748"

[Hide](#)

```
print(paste("Specificity - All pts: ", specificity*100))
```

[1] "Specificity - All pts: 97.3647711511789"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - All pts: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - All pts: 84.9206349206349"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - All pts: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - All pts: 95.1219512195122"

[Hide](#)

```
# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage I: ", sensitivity*100))
```

[1] "Sensitivity - Stage I: 74.8251748251748"

[Hide](#)

```
print(paste("Specificity - Stage I: ", specificity*100))
```

```
[1] "Specificity - Stage I: 97.3647711511789"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage I: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage I: 84.9206349206349"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage I: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage I: 95.1219512195122"
```

[Hide](#)

```
#Stage II Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="II",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage II: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage II: 63.6363636363636"
```

[Hide](#)

```
print(paste("Specificity - Stage II: ", specificity*100))
```

```
[1] "Specificity - Stage II: 96.9230769230769"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage II: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage II: 63.6363636363636"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage II: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage II: 96.9230769230769"
```

[Hide](#)

```
#Stage III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="III",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage III: ", sensitivity*100))
```

```
[1] "Sensitivity - Stage III: 75"
```

[Hide](#)

```
print(paste("Specificity - Stage III: ", specificity*100))
```

```
[1] "Specificity - Stage III: 98.0952380952381"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - Stage III: 88.6363636363636"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - Stage III: 95.1940850277264"
```

[Hide](#)

```
#High-risk Stage II/III Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$HighRisk.Stage=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - High-risk Stage II/III: ", sensitivity*100))
```

```
[1] "Sensitivity - High-risk Stage II/III: 73.9130434782609"
```

[Hide](#)

```
print(paste("Specificity - High-risk Stage II/III: ", specificity*100))
```

```
[1] "Specificity - High-risk Stage II/III: 97.8428351309707"
```

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - High-risk Stage II/III: ", ppv * 100))
```

```
[1] "Positive Predictive Value (PPV) - High-risk Stage II/III: 85.8585858585859"
```

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - High-risk Stage II/III: ", npv * 100))
```

```
[1] "Negative Predictive Value (NPV) - High-risk Stage II/III: 95.4887218045113"
```

[Hide](#)

```
#Stage IV Patients
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ACT==TRUE,]
circ_data <- circ_data[circ_data$Stage=="IV",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE", "POSITIVE"))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels=c("FALSE", "TRUE"))
circ_data$RFS.Event <- factor(ifelse(circ_data$RFS.Event == "FALSE", "NEGATIVE", "POSITIVE"))
conf_matrix <- table(circ_data$ctDNA.Surveillance, circ_data$RFS.Event)

# Calculate sensitivity and specificity manually
true_positives <- conf_matrix["POSITIVE", "POSITIVE"]
false_negatives <- conf_matrix["NEGATIVE", "POSITIVE"]
true_negatives <- conf_matrix["NEGATIVE", "NEGATIVE"]
false_positives <- conf_matrix["POSITIVE", "NEGATIVE"]
sensitivity <- true_positives / (true_positives + false_negatives)
specificity <- true_negatives / (true_negatives + false_positives)
ppv <- true_positives / (true_positives + false_positives)
npv <- true_negatives / (true_negatives + false_negatives)
print(paste("Sensitivity - Stage IV: ", sensitivity*100))
```

[1] "Sensitivity - Stage IV: 78.5714285714286"

[Hide](#)

```
print(paste("Specificity - Stage IV: ", specificity*100))
```

[1] "Specificity - Stage IV: 92.4242424242424"

[Hide](#)

```
print(paste("Positive Predictive Value (PPV) - Stage IV: ", ppv * 100))
```

[1] "Positive Predictive Value (PPV) - Stage IV: 81.4814814814815"

[Hide](#)

```
print(paste("Negative Predictive Value (NPV) - Stage IV: ", npv * 100))
```

[1] "Negative Predictive Value (NPV) - Stage IV: 91.044776119403"

#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)
```

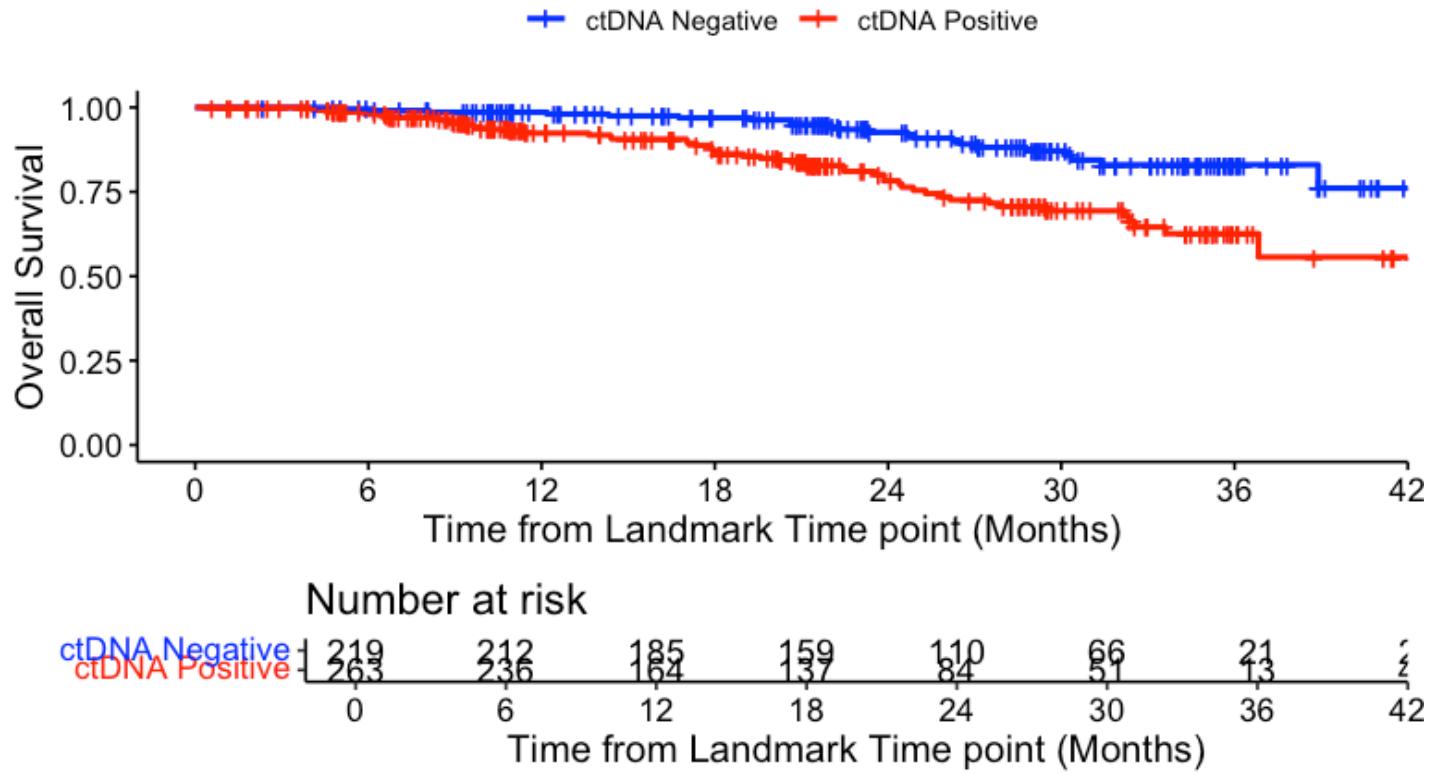
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	219	22	0.1004566	10.04566
POSITIVE	263	52	0.1977186	19.77186
NA	1	NA	NA	NA

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=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



```
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

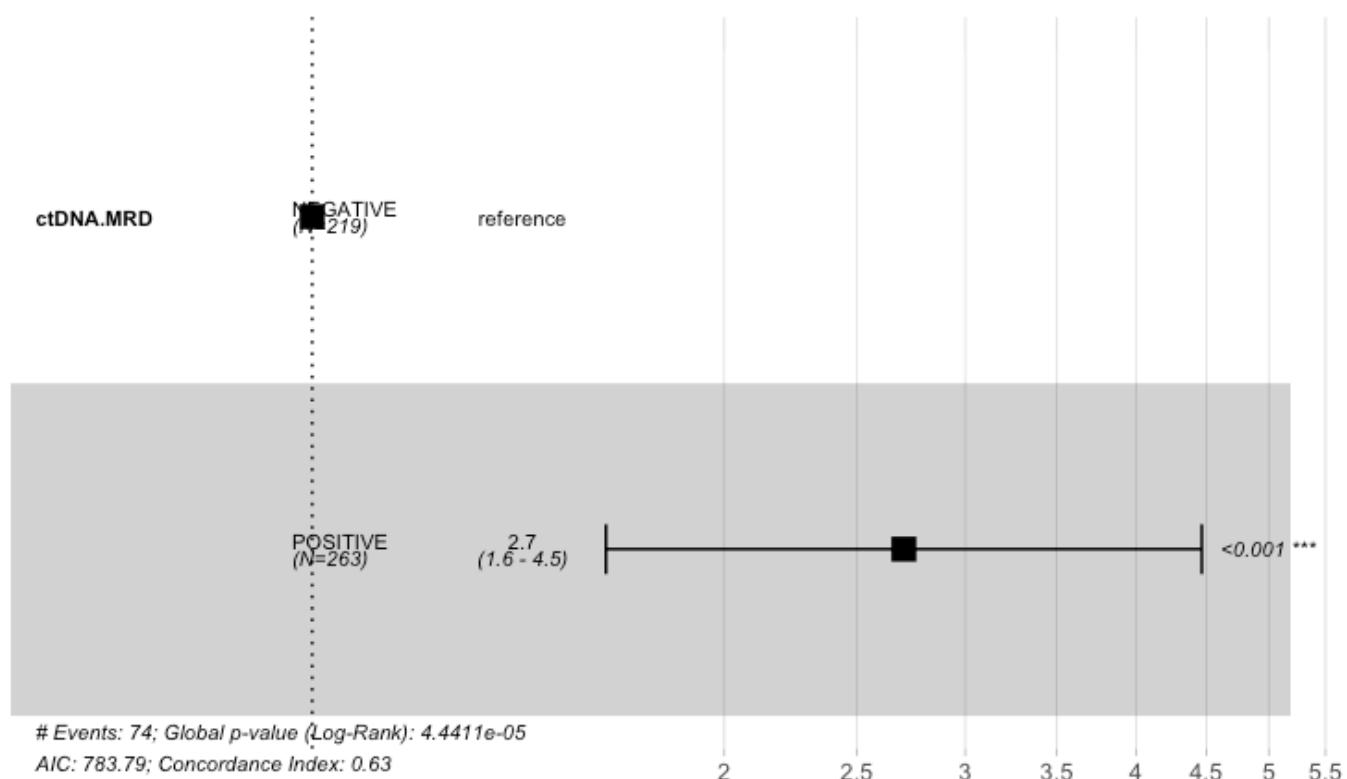
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	110	12	0.926	0.0209	0.873		0.958	
36	21	9	0.830	0.0364	0.744		0.889	

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	

```
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"
```

```
#OS by ctDNA at the MRD Window - pts with Radiological Recurrence Sites
```

Hide

```

# 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,]

  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")

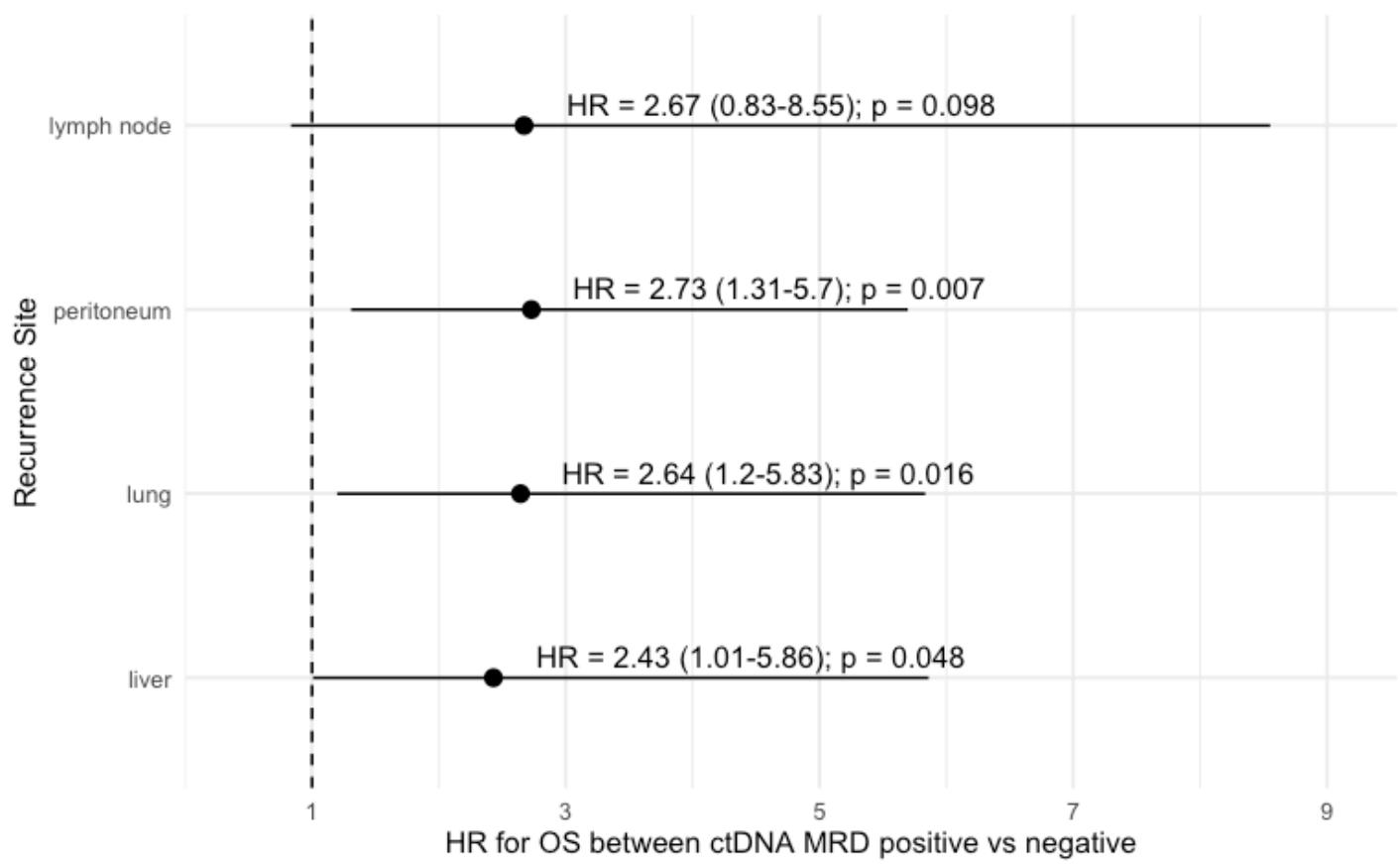
# Perform analysis for each site
results <- lapply(recurrence_sites, analyze_site)

# 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"
```

#OS 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$ctDNA.Surveillance!="",]
circ_data$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

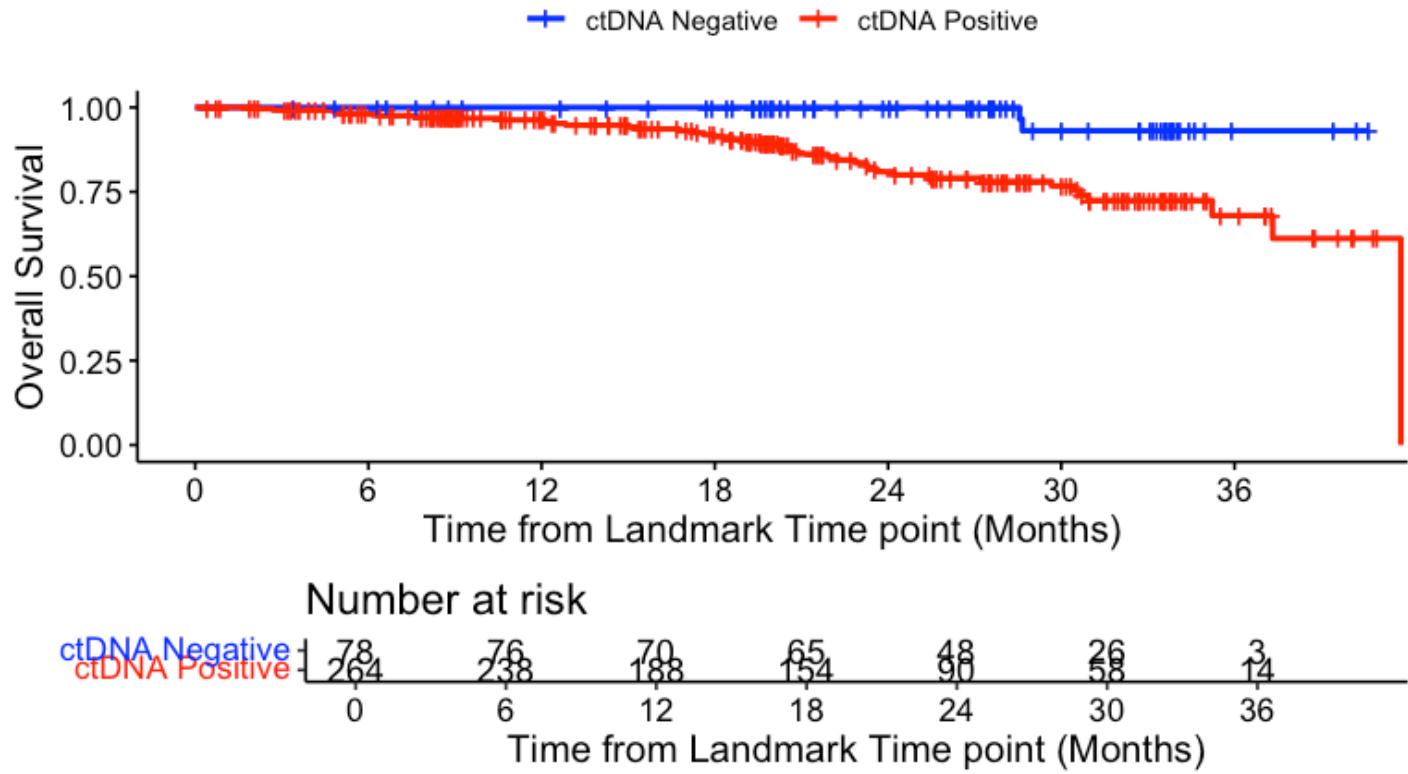
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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



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

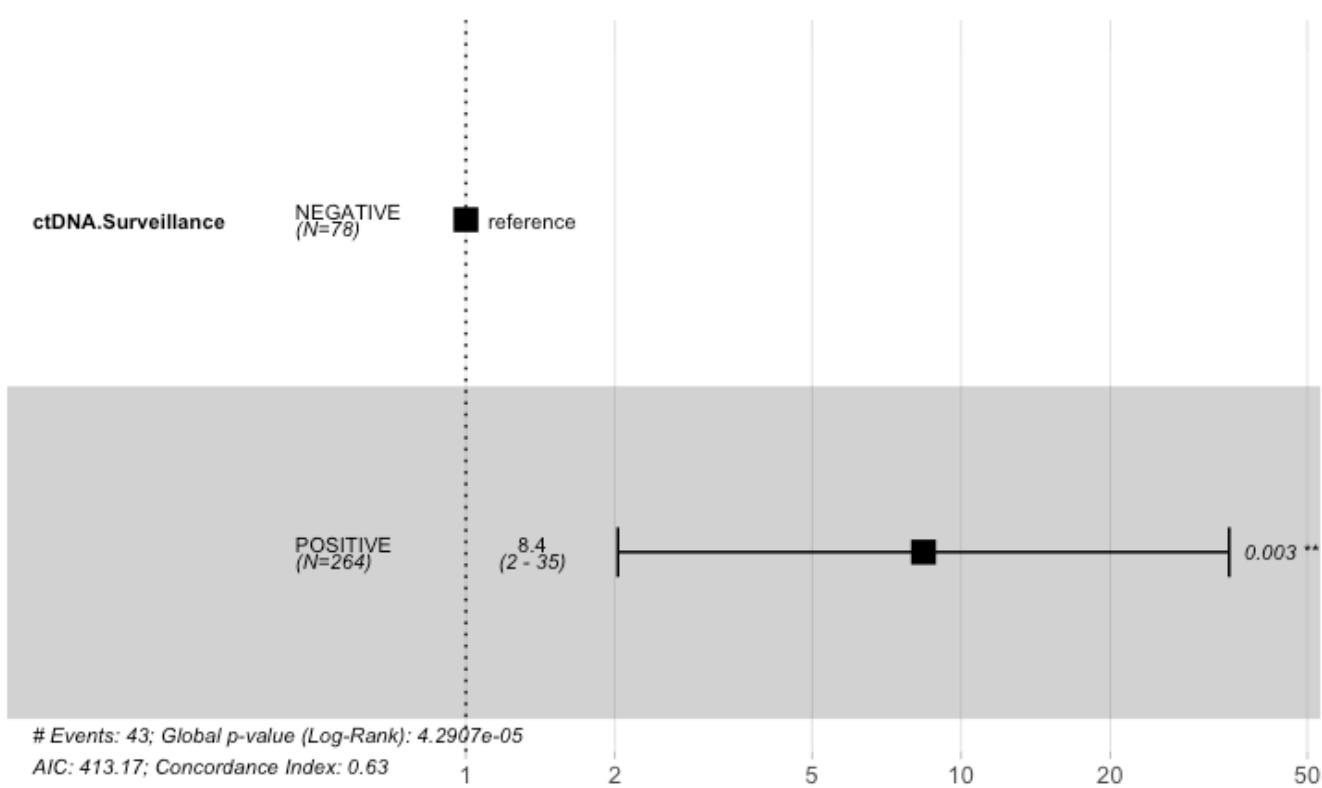
```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
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
```

```
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"
```

```
#OS by ctDNA at the Surveillance Window - pts with Lung 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 %>% filter(grepl("lung", RelSite, ignore.case = 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,]

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	59	2	NA	NA	NA
ctDNA.Surveillance=POSITIVE	83	16	41.8	NA	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	59	2	0.03389831	3.389831
POSITIVE	83	16	0.19277108	19.277108

2 rows

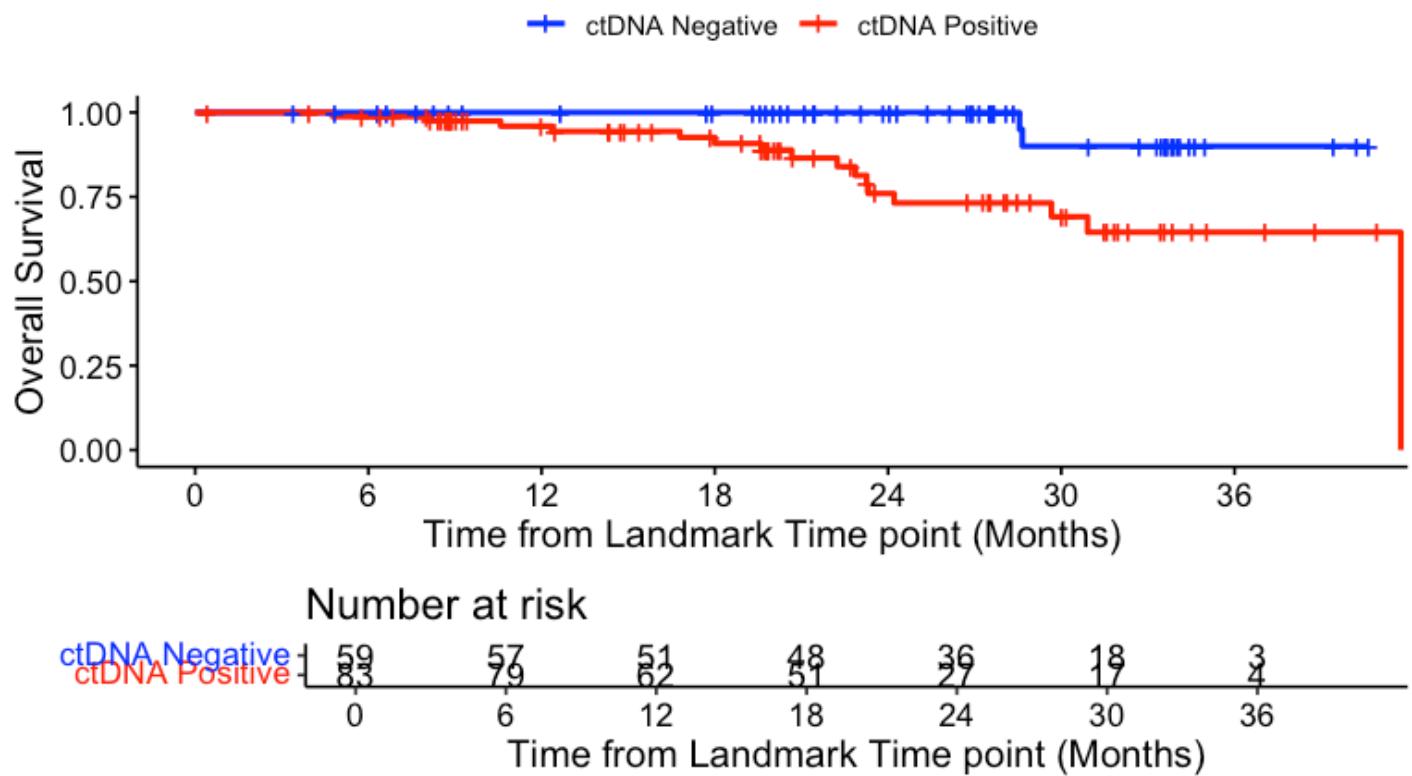
[Hide](#)

```

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 - Lung 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 - Lung Radiological Recurrence | ctDNA Surveillance window



```
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	36	0	1.0	0.0000	NA	NA	NA	NA
36	3	2	0.9	0.0671	0.656	0.974	0.656	0.974

ctDNA.Surveillance=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
24	27	12	0.760	0.0636	0.607	0.860	0.607	0.860
36	4	3	0.645	0.0827	0.459	0.781	0.459	0.781

```
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= 142, number of events= 18

	coef	exp(coef)	se(coef)	z	Pr(> z)						
ctDNA.SurveillancePOSITIVE	2.033	7.638	0.754	2.696	0.00701 **						

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.SurveillancePOSITIVE	7.638	0.1309	1.742	33.48

Concordance= 0.718 (se = 0.029)

Likelihood ratio test= 11.41 on 1 df, p=7e-04

Wald test = 7.27 on 1 df, p=0.007

Score (logrank) test = 10.1 on 1 df, p=0.001

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 = 7.64 (1.74-33.48); p = 0.007"

#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

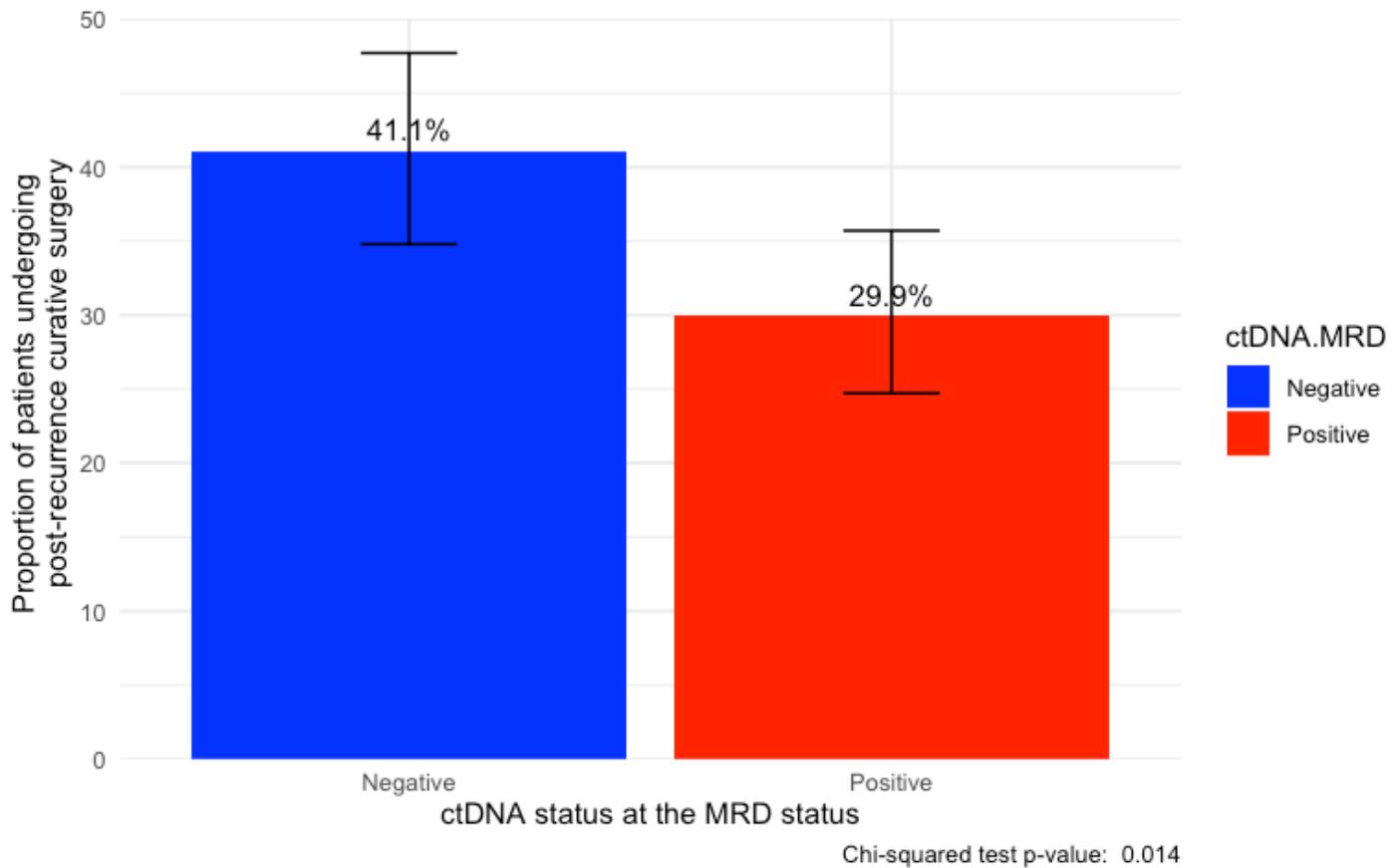
```
print(chi_test)
```

```
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)
```

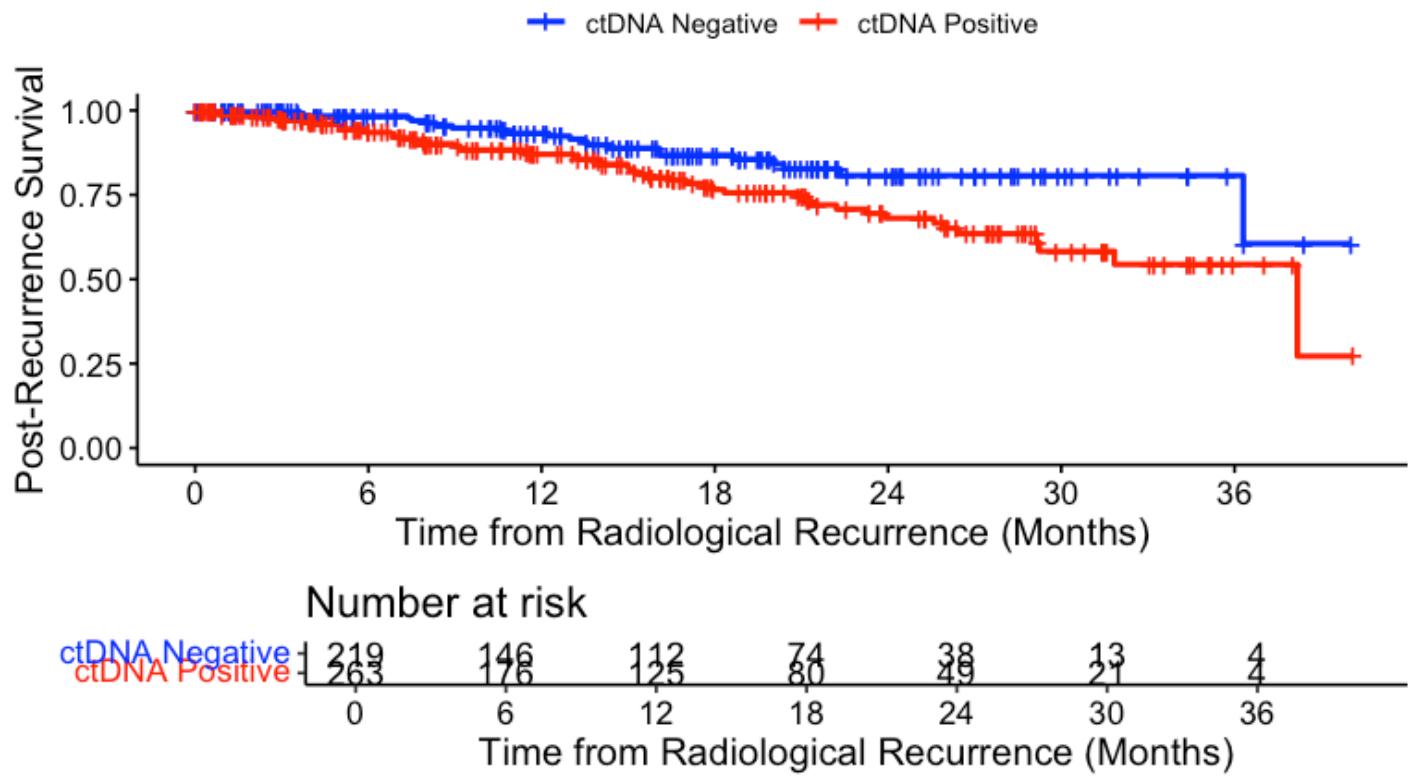
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	219	22	0.1004566	10.04566
POSITIVE	263	52	0.1977186	19.77186
NA	18	NA	NA	NA

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="")
```

PRS - Radiological Recurrence | ctDNA MRD window



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

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
45	24.0000	38.0000	21.0000	0.8073	0.0412	0.7105		0.87	

ctDNA.MRD=POSITIVE

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
77	24.0000	49.0000	45.0000	0.6809	0.0435	0.5872		0.75	

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
```

[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 = 1.97 (1.2-3.24); p = 0.008"
```

#PRS by ctDNA at the MRD Window - pts with Radiological Recurrence Sites

[Hide](#)

```

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,]

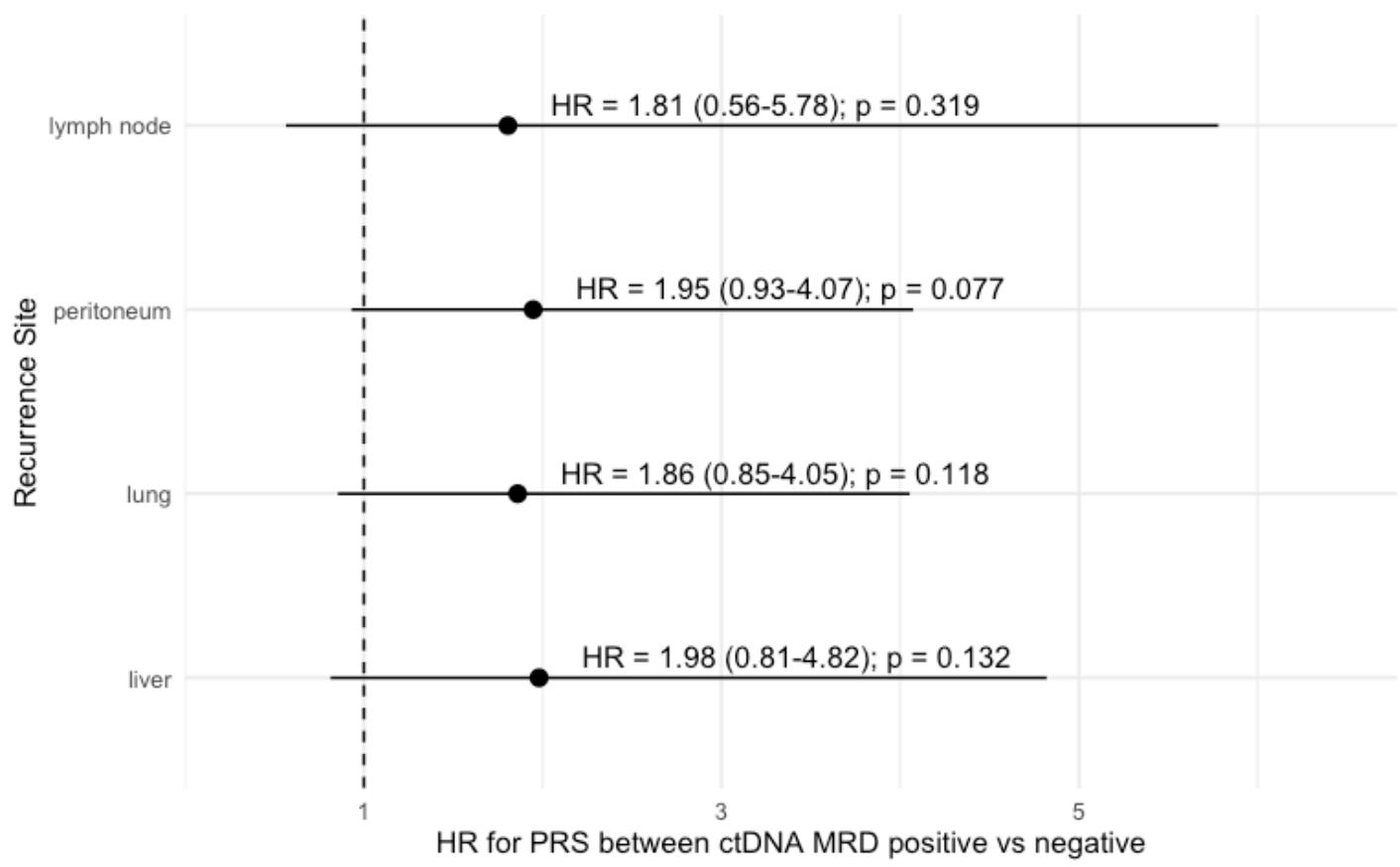
  surv_object <- Surv(time = circ_data_site$PRS.months, event = circ_data_site$OS.Event)
  cox_fit <- coxph(surv_object ~ ctDNA.MRD, data = circ_data_site)
  cox_fit_summary <- summary(cox_fit)

  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",]
circ_data <- circ_data[circ_data$OS.MRD.months >= 0,]
recurrence_sites <- c("liver", "lung", "peritoneum", "lymph node")
results <- lapply(recurrence_sites, analyze_site)
forest_data <- do.call(rbind, lapply(results, function(res) {
  data.frame(
    site = res$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 PRS between ctDNA MRD positive vs negative") +
 theme_minimal()
print(forest_plot)



Hide

```
for (res in results) {
  print(res$label_text)
}
```

```
[1] "HR = 1.98 (0.81-4.82); p = 0.132"
[1] "HR = 1.86 (0.85-4.05); p = 0.118"
[1] "HR = 1.95 (0.93-4.07); p = 0.077"
[1] "HR = 1.81 (0.56-5.78); p = 0.319"
```

#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

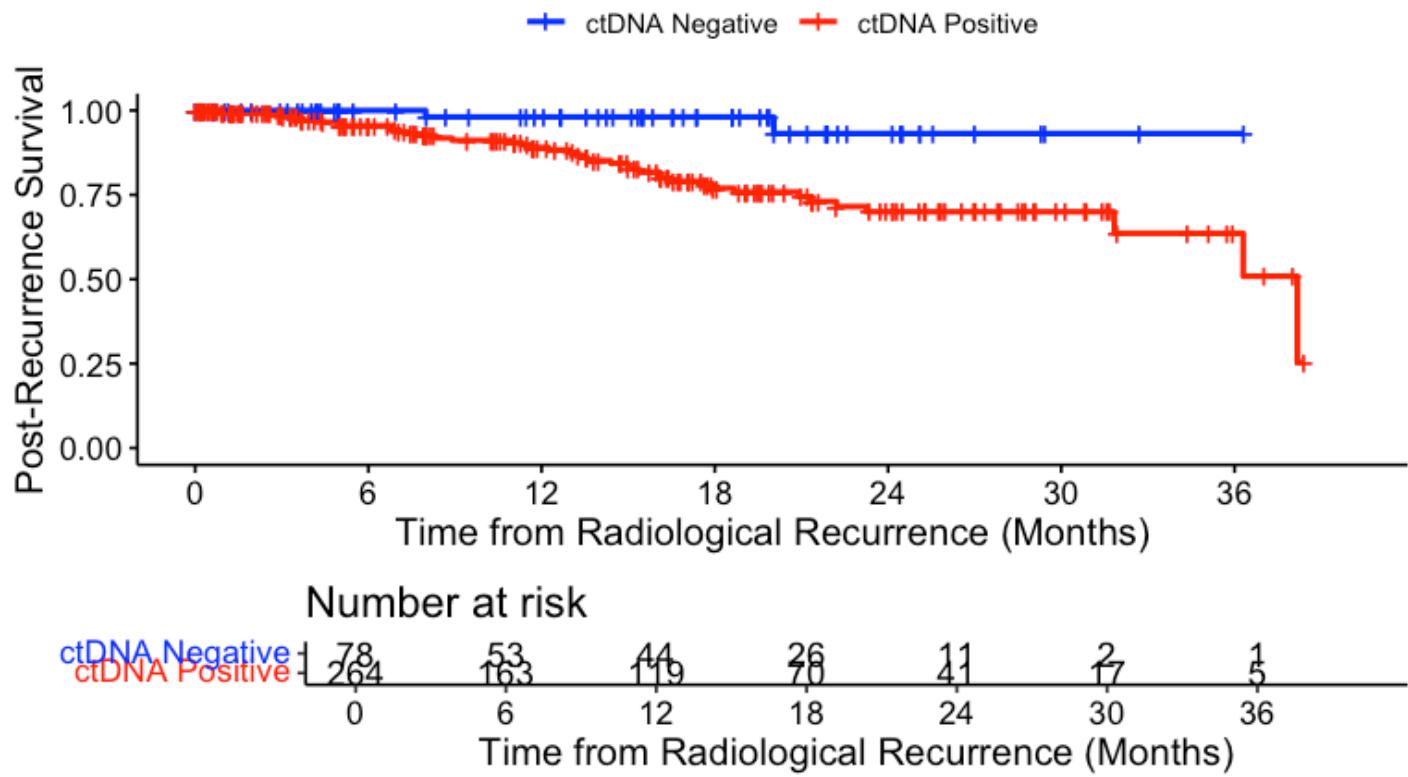
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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



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

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

CI	ctDNA.Surveillance=NEGATIVE						
	time	n.risk	n.event	survival	std.err	lower	95% CI upper
47	24.0000	11.0000	2.0000	0.9317	0.0511	0.7237	0.98

CI	ctDNA.Surveillance=POSITIVE						
	time	n.risk	n.event	survival	std.err	lower	95% CI upper
78	24.000	41.000	38.000	0.700	0.045	0.602	0.7

```
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"
```

#PRS by ctDNA at the Surveillance Window - pts with Lung 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 %>% filter(grepl("lung", RelSite, ignore.case = 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	59	2	NA	NA	NA
ctDNA.Surveillance=POSITIVE	83	16	38.2	23.3	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	59	2	0.03389831	3.389831
POSITIVE	83	16	0.19277108	19.277108
2 rows				

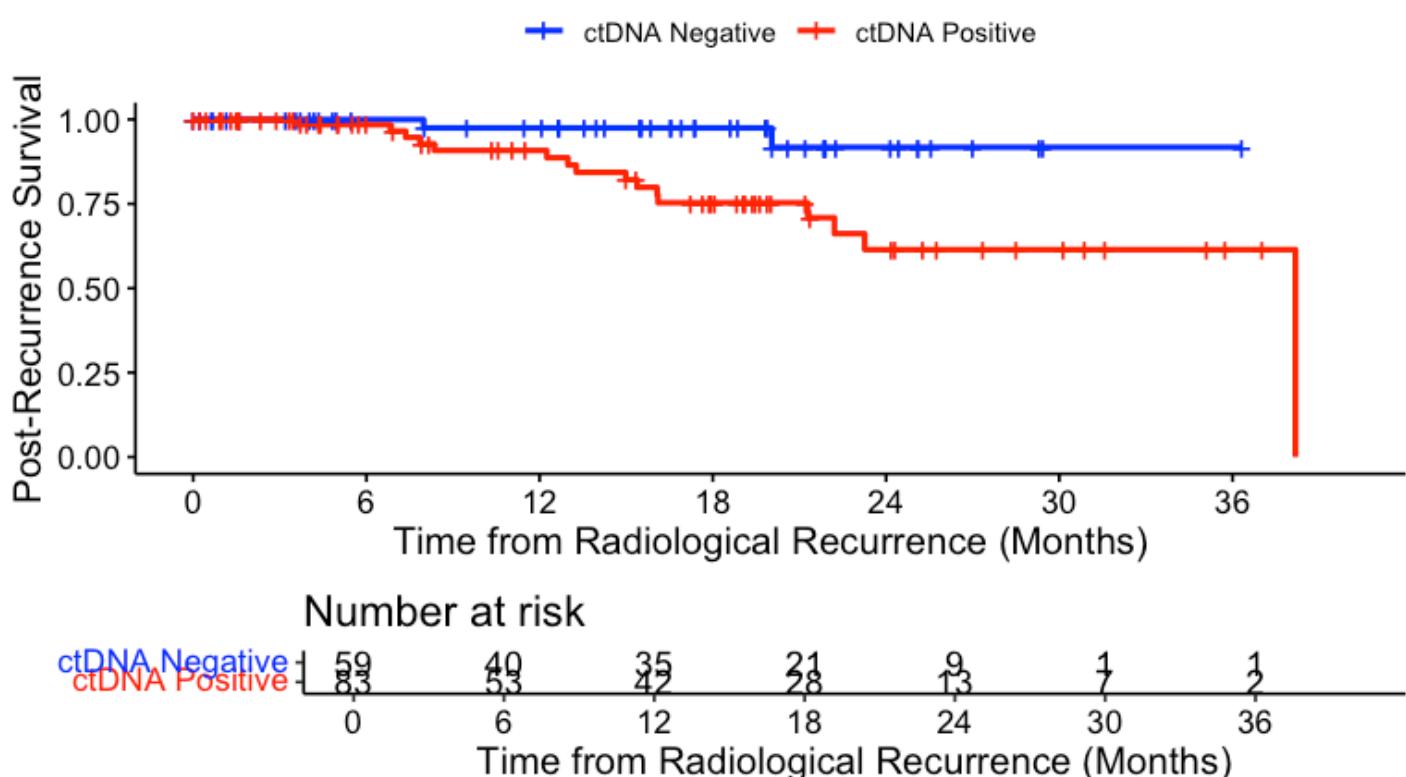
Hide

```

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 - Lung 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 - Lung 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")

CI	ctDNA.Surveillance=NEGATIVE						
	time	n.risk	n.event	survival	std.err	lower	95% CI upper
10	24.0000	9.0000	2.0000	0.9176	0.0603	0.6807	0.98
15	24.0000	13.0000	15.0000	0.6144	0.0887	0.4186	0.76

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= 142, number of events= 18

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.SurveillancePOSITIVE	1.758	5.801	0.753	2.335	0.0196 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.SurveillancePOSITIVE	5.801	0.1724	1.326	25.38

Concordance= 0.668 (se = 0.044)

Likelihood ratio test= 8.18 on 1 df, p=0.004

Wald test = 5.45 on 1 df, p=0.02

Score (logrank) test = 7 on 1 df, p=0.008

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 = 5.8 (1.33-25.38); p = 0.02"
```

#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
relsight_counts <- table(circ_data$Rec.Site)
relsight_df <- as.data.frame(relsight_counts)
names(relsight_df) <- c("RelSite", "Count")
circ_data_pos_mrd <- circ_data[circ_data$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$Rec.Site)
relsight_df$MRDPos_Count <- ifelse(is.na(match(relsight_df$RelSite, names(pos_counts_mrd))), 0, pos_counts_mrd[match(relsight_df$RelSite, names(pos_counts_mrd))])
relsight_df$MRDPos_Count[is.na(relsight_df$MRDPos_Count)] <- 0
relsight_df$AnytimePos_Count <- ifelse(is.na(match(relsight_df$RelSite, names(pos_counts_anytime))), 0, pos_counts_anytime[match(relsight_df$RelSite, names(pos_counts_anytime))])
relsight_df$AnytimePos_Count[is.na(relsight_df$AnytimePos_Count)] <- 0
relsight_df$Percent <- (relsight_df$Count / sum(relsight_df$Count)) * 100
relsight_df$MRDPos_Percent <- (relsight_df$MRDPos_Count / relsight_df$Count) * 100
relsight_df$AnytimePos_Percent <- (relsight_df$AnytimePos_Count / relsight_df$Count) * 100
total_observations <- sum(relsight_df$Count)
total_pos_mrd <- sum(relsight_df$MRDPos_Count)
total_pos_anytime <- sum(relsight_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)
relsight_df <- rbind(relsight_df, total_row)
print(relsight_df)

```

RelSite	Co...	MRDPos_C...	AnytimePos_Count	Percent	MRDPos_Per...	Any...
<fctr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Brain	2	2	0	0.4	100.00000	
Liver	191	138	0	38.2	72.25131	
Local/LN	35	20	0	7.0	57.14286	
Lung	142	39	0	28.4	27.46479	
Others	12	8	0	2.4	66.66667	
Peritoneum	55	23	0	11.0	41.81818	
Peritoneum & Others	63	34	0	12.6	53.96825	
Total	500	264	0	100.0	52.80000	
8 rows						

#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 = FALSE, cluster_columns = FALSE, top_annotation = ha)
pdf("heatmap.pdf", width = 7, height = 7)
draw(ht, annotation_legend_side = "bottom")
dev.off()
```

```
null device
1
```

#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

[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))
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

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	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

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= 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)

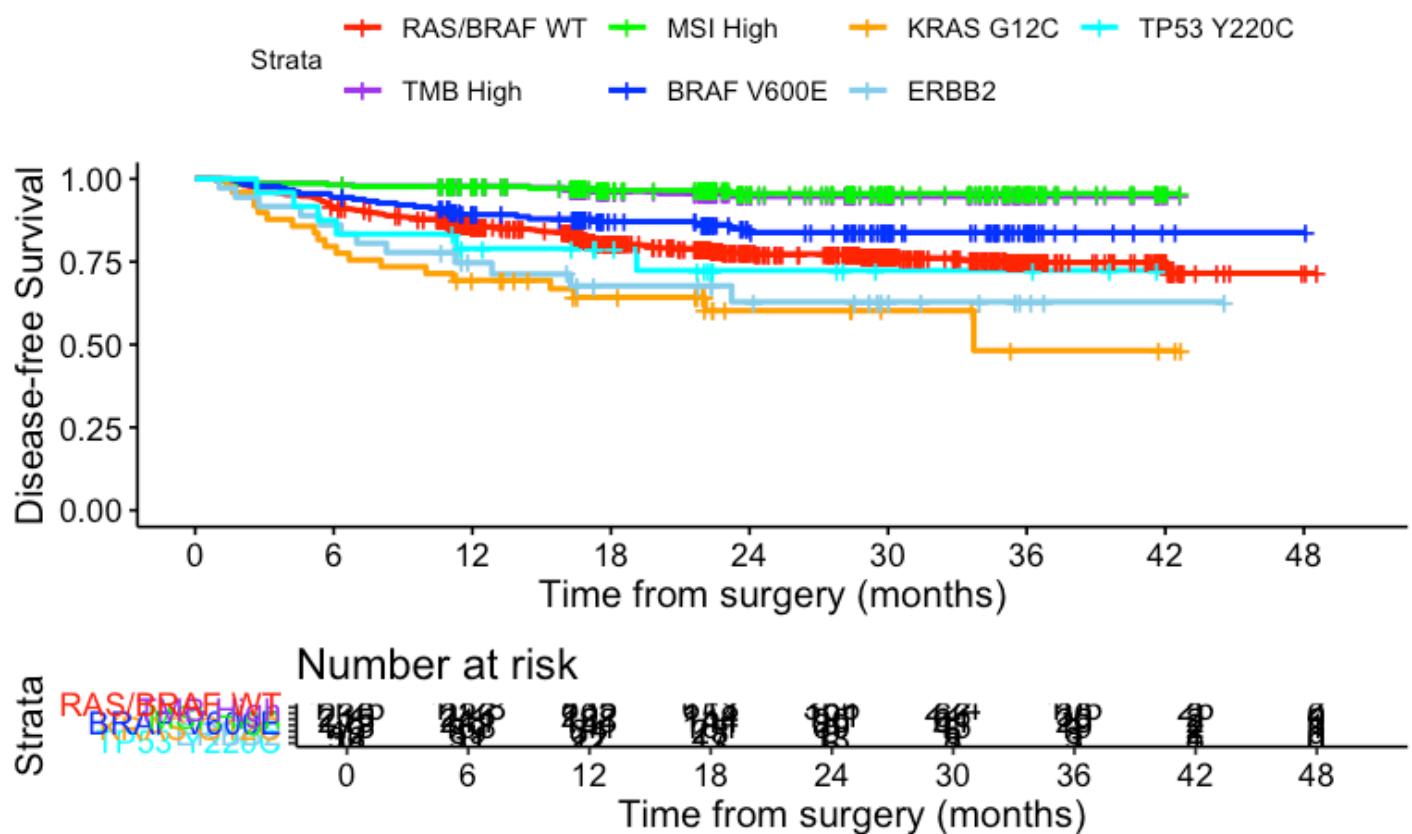
Likelihood ratio test= 107 on 6 df, p=<2e-16

Wald test = 73.9 on 6 df, p=6e-14

Score (logrank) test = 93.74 on 6 df, p=<2e-16

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", "ERB B2", "TP53 Y220C"),
  palette = c("red", "purple", "green", "blue", "orange", "skyblue", "cyan")
)
```



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

Call: `survfit(formula = surv_obj ~ value, data = circ_data_long)`

CI	value=RAS/BRAF WT								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
28	24.0000	366.0000	224.0000	0.7755	0.0137	0.7491		0.80	
CI	value=TMB High								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
07	24.0000	101.0000	10.0000	0.9471	0.0169	0.9146		0.98	
CI	value=MSI High								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
74	24.0000	96.0000	8.0000	0.9558	0.0159	0.9252		0.98	
CI	value=BRAF V600E								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
15	24.0000	69.0000	25.0000	0.8382	0.0311	0.7793		0.90	
CI	value=KRAS G12C								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
26	24.0000	9.0000	18.0000	0.6023	0.0765	0.4696		0.77	
CI	value=ERBB2								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
89	24.0000	13.0000	12.0000	0.6287	0.0887	0.4769		0.82	
CI	value=TP53 Y220C								
	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95%
70	24.0000	6.0000	6.0000	0.7237	0.0993	0.5530		0.94	

#Percentage of ctDNA MRD Window positivity in biomarker groups

[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))

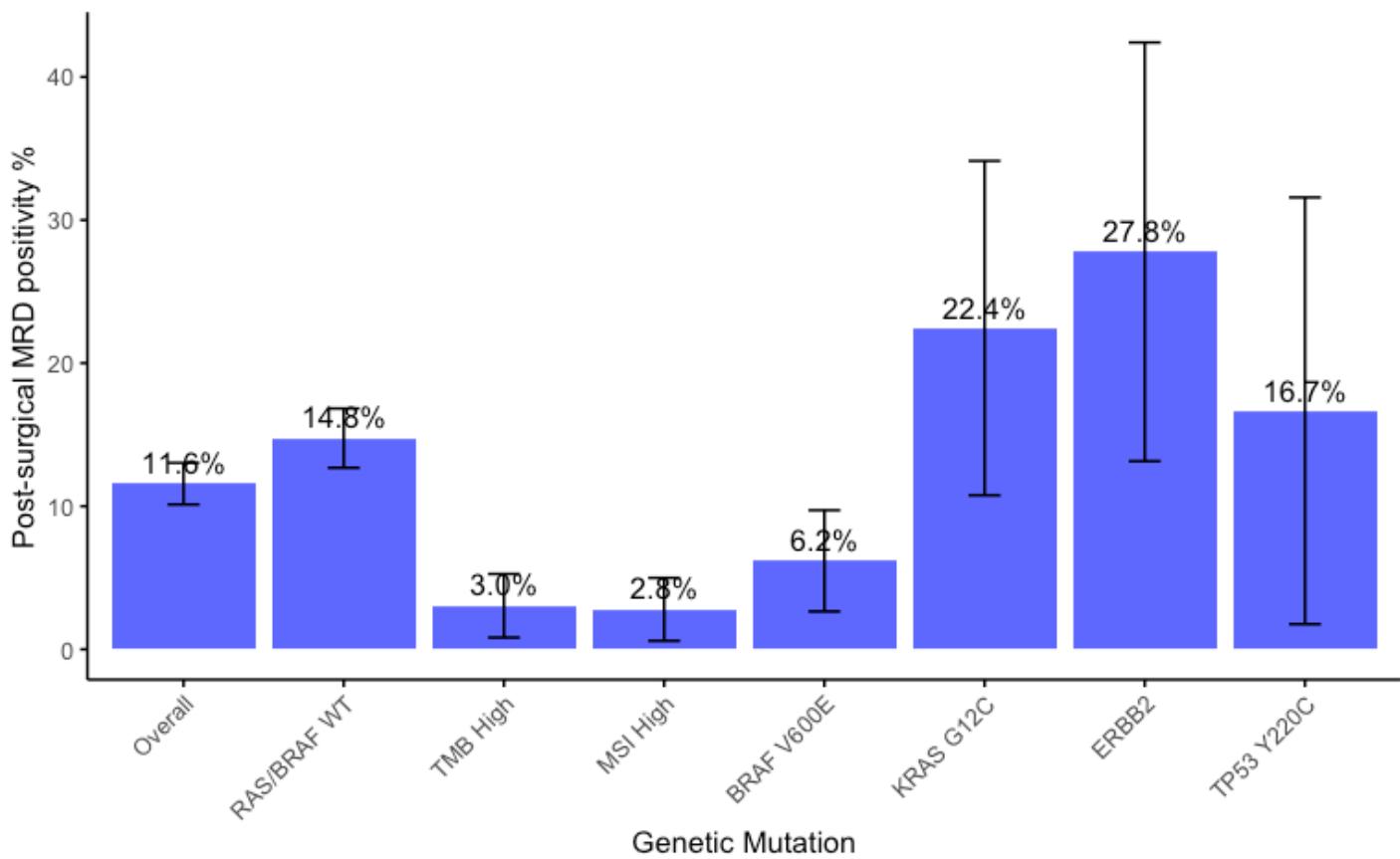
summary_data <- circ_data_long %>%
  group_by(value) %>%
  summarise(
    n = n(),
    positive = sum(ctDNA.MRD == "POSITIVE"),
    pct_positive = (positive / n) * 100,
    se = sqrt((pct_positive / 100) * (1 - pct_positive / 100) / n),
    ci_low = pct_positive - 1.96 * se * 100,
    ci_high = pct_positive + 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 - All pts 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)

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NA	1	NA	NA	NA
3 rows				

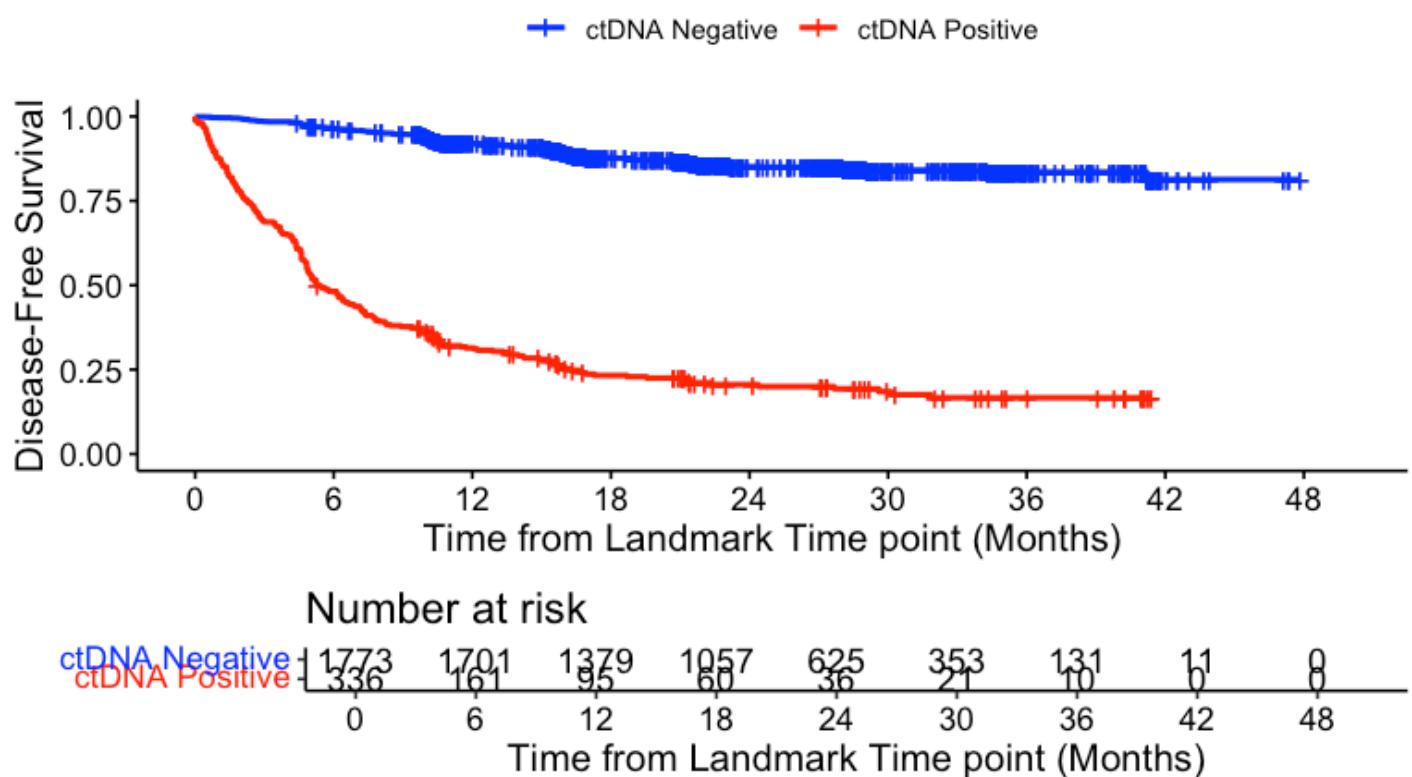
Hide

```

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 Stage s",
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(0, 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
0	1773	2	0.999	0.000797	0.995		1.000	
24	625	222	0.851	0.009494	0.832		0.869	

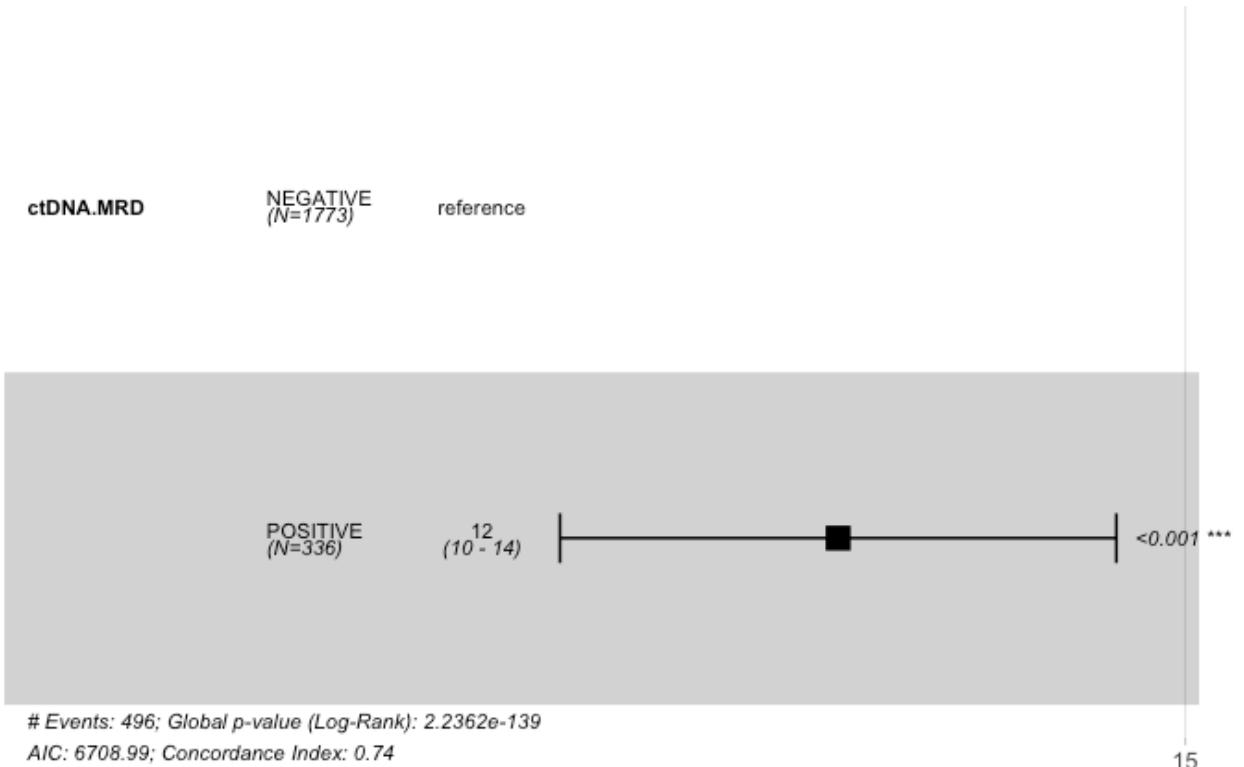
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	336	4	0.988	0.00592	0.969		0.996	
24	36	254	0.206	0.02364	0.161		0.254	

[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)

# 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"
```

```
#DFS by ctDNA at the MRD Window - RAS/BRAF WT 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$RAS.BRAF=="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)

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	891	101	NA	NA	NA
ctDNA.MRD=POSITIVE	166	123	6.37	5.06	10.3

[Hide](#)

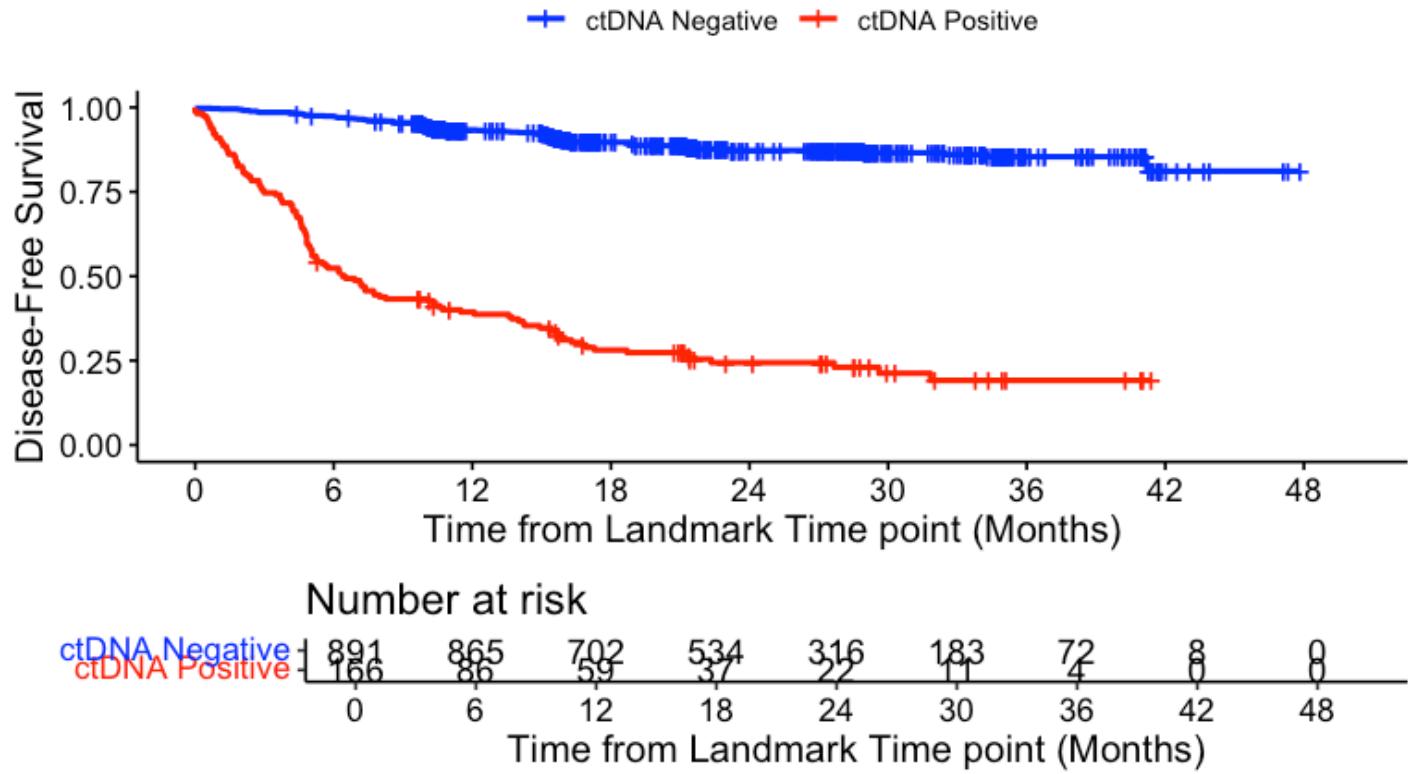
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	891	101	0.1133558	11.33558
POSITIVE	166	123	0.7409639	74.09639
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 | RAS/BRAF WT",
ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | RAS/BRAF WT



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	891	2	0.998	0.00159	0.991	0.999	0.999	0.999
24	316	94	0.873	0.01258	0.846	0.895	0.895	0.895

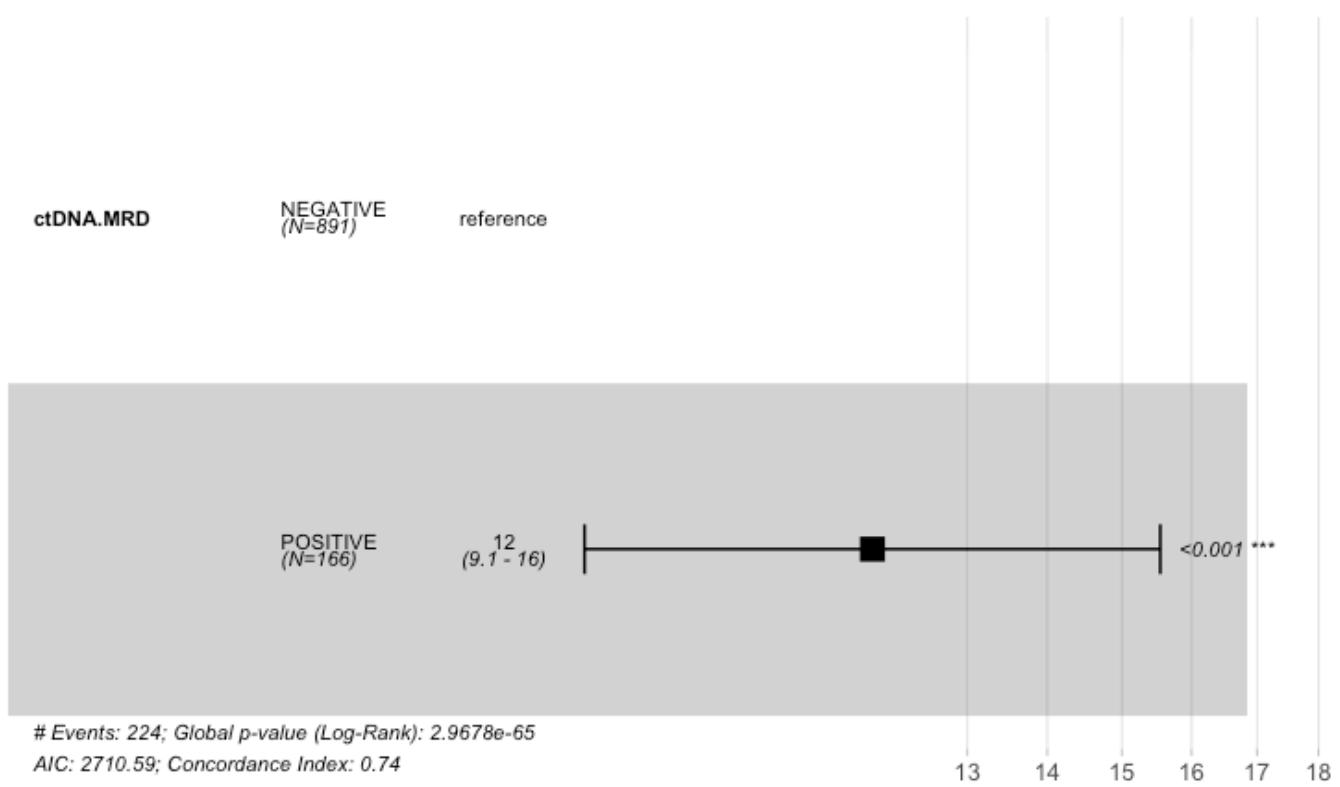
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	166	2	0.988	0.00847	0.953	0.997	0.997	0.997
24	22	118	0.244	0.03605	0.177	0.317	0.317	0.317

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= 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
```

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 = 11.91 (9.12-15.54); p = 0"
```

```
#DFS by ctDNA at the MRD Window - TMB High 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$TMB=="TMB-High",]
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)

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	210	5	NA	NA	NA
ctDNA.MRD=POSITIVE	7	4	4.73	0.559	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)
```

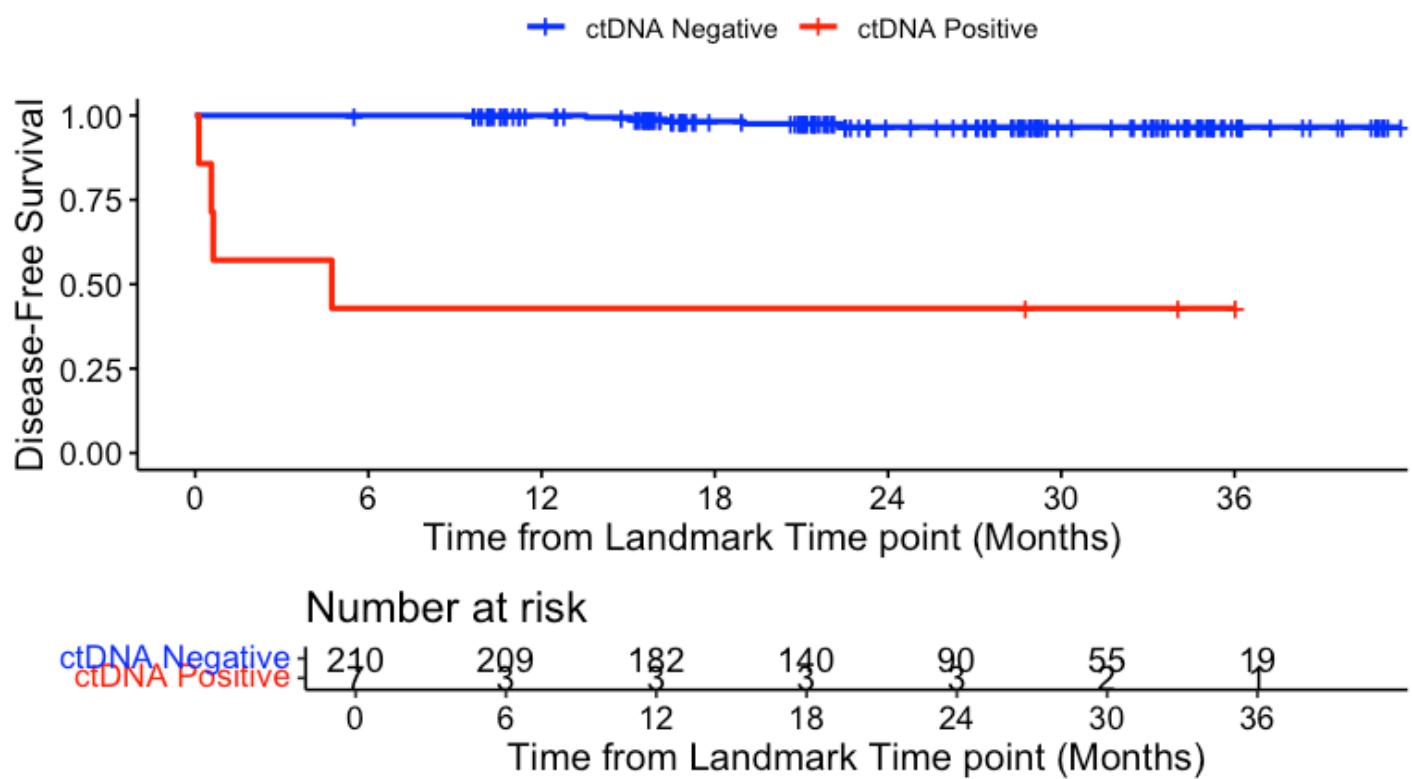
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	210	5	0.02380952	2.380952
POSITIVE	7	4	0.57142857	57.142857

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 | TMB-High", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labels=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | TMB-High



```
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")
```

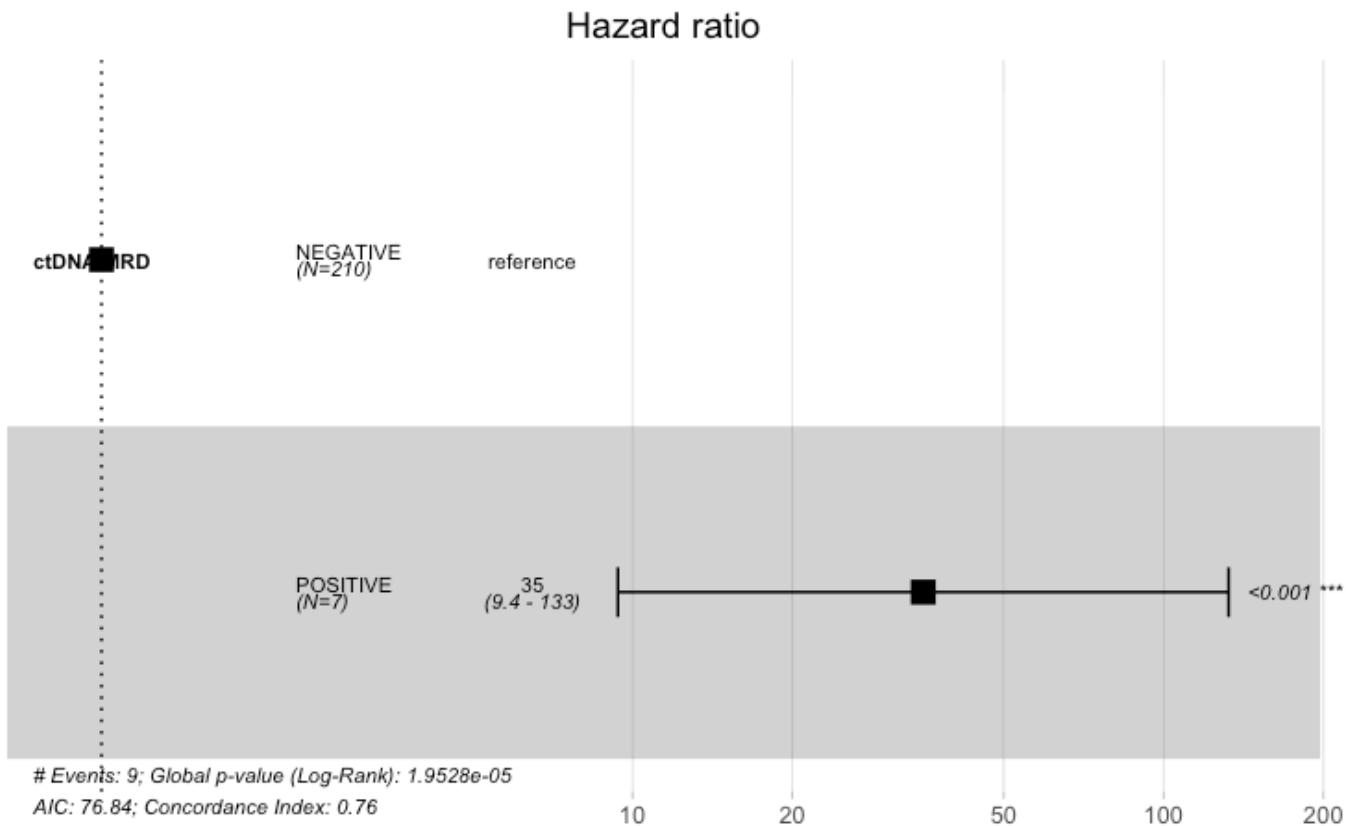
ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	210	0	1.000	0.0000	1.000		1.000	
24	90	5	0.966	0.0155	0.917		0.986	

ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	7	0	1.000	0.000	1.0000		1.000	
24	3	4	0.429	0.187	0.0978		0.734	

```
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)
```



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)

# 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 = 35.27 (9.38-132.58); p = 0"
```

#DFS by ctDNA at the MRD Window - MSI High 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$MSI=="MSI-High",]
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)

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	196	3	NA	NA	NA
ctDNA.MRD=POSITIVE	6	4	2.68	0.559	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)
```

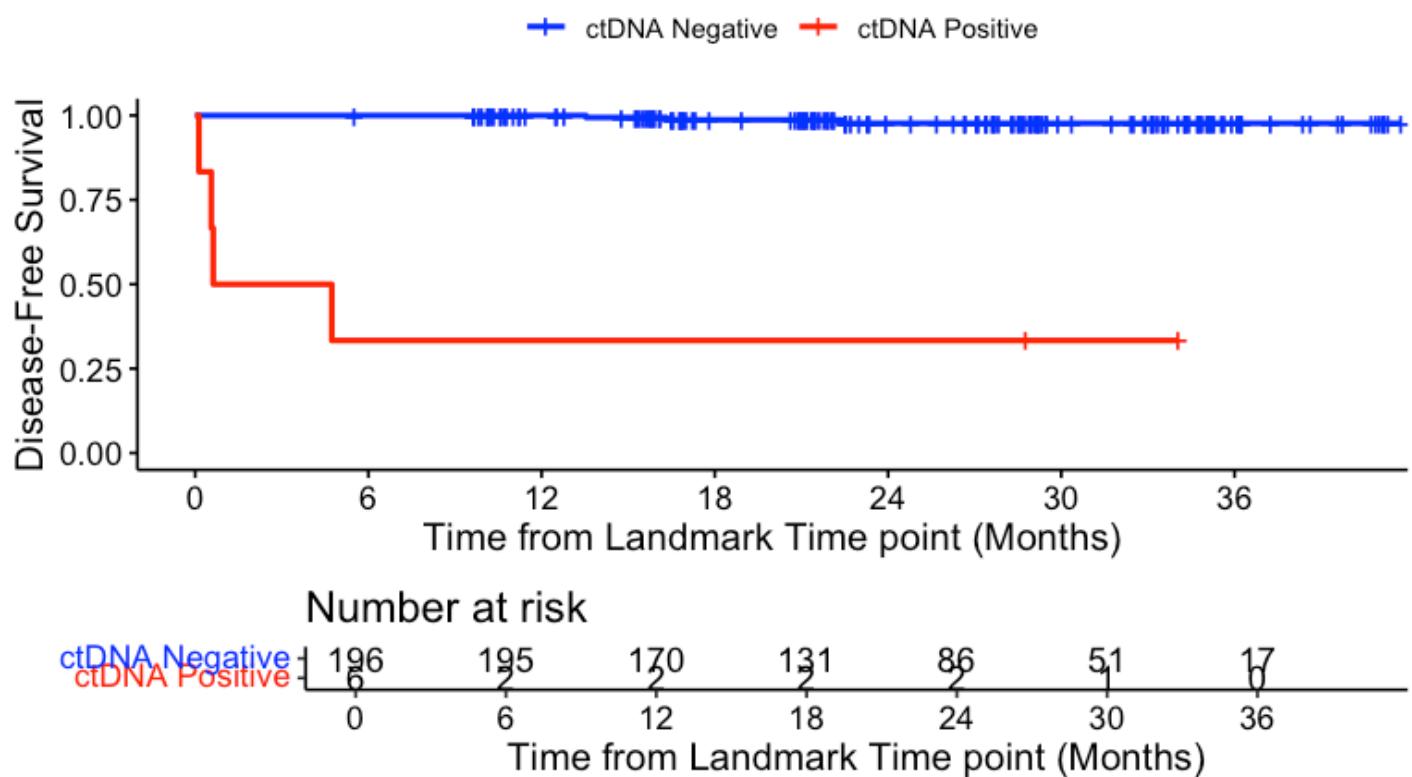
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	196	3	0.01530612	1.530612
POSITIVE	6	4	0.66666667	66.666667

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 | MSI-High", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.lab=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | MSI-High



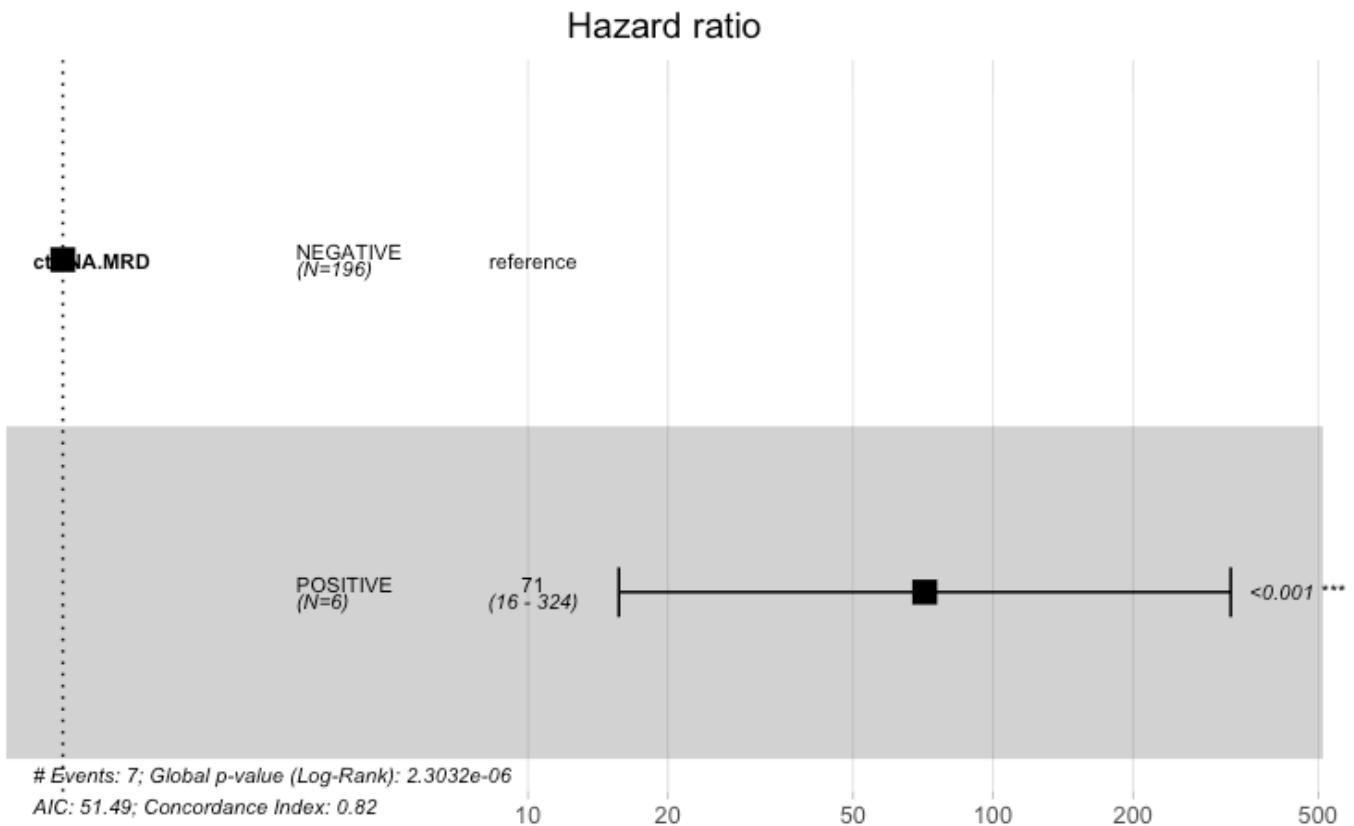
```
summary(KM_curve, times= c(0, 24))
```

```
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     196      0     1.000  0.0000     1.000     1.000
 24     86      3     0.977  0.0136     0.928     0.993
```

```
ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     6      0     1.000  0.000     1.0000     1.000
 24     2      4     0.333  0.192     0.0461     0.676
```

```
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)
```



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)
```

```
# 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 = 71.32 (15.68-324.37); p = 0"
```

#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_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	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)
```

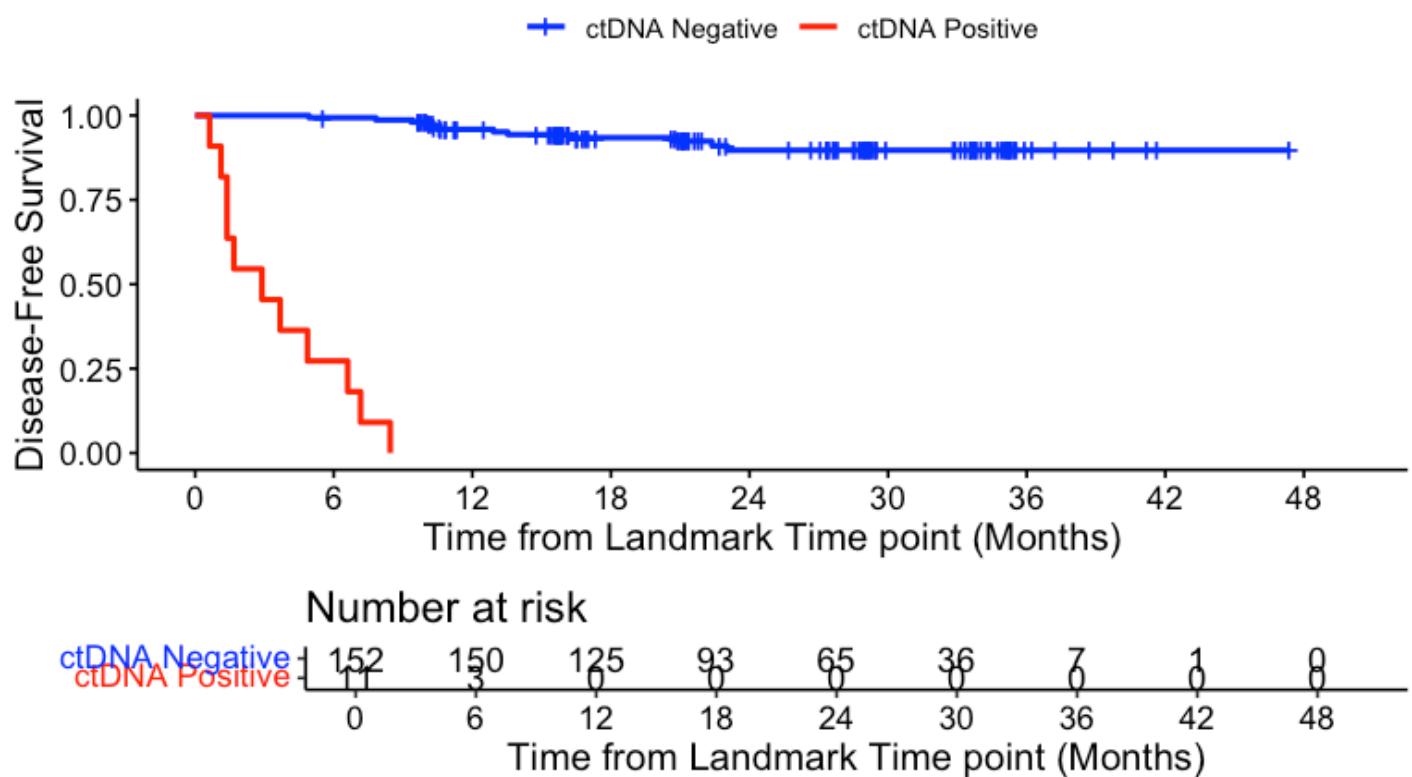
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	152	12	0.07894737	7.894737
POSITIVE	11	11	1.00000000	100.000000

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="")
```

DFS - ctDNA MRD window | BRAF V600E



```
summary(KM_curve, times= c(0, 24))
```

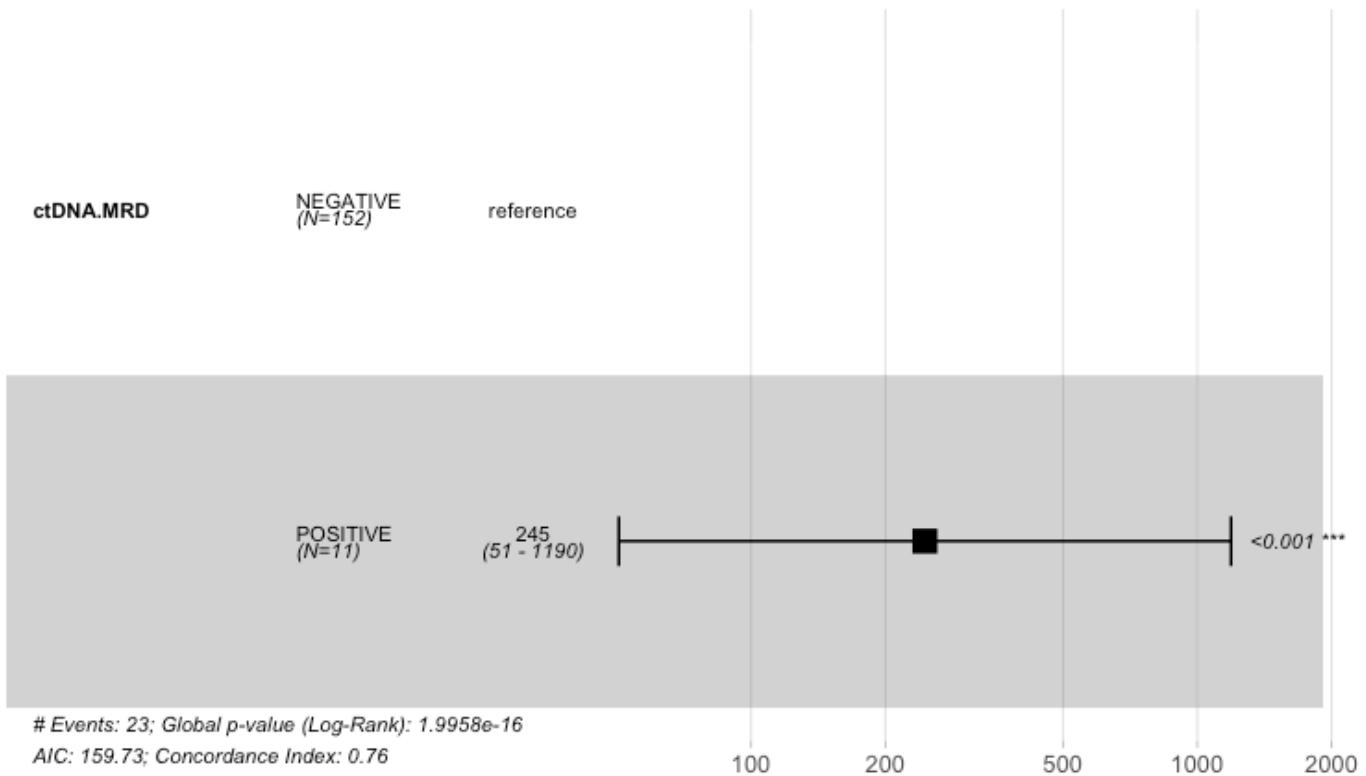
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
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
0	11	0	1	0	0	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"
```

#DFS by ctDNA at the MRD Window - KRAS G12C 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$KRAS.G12C=="MUT",]
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)

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	36	8	NA	NA	NA
ctDNA.MRD=POSITIVE	11	10	2.14	1.61	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)
```

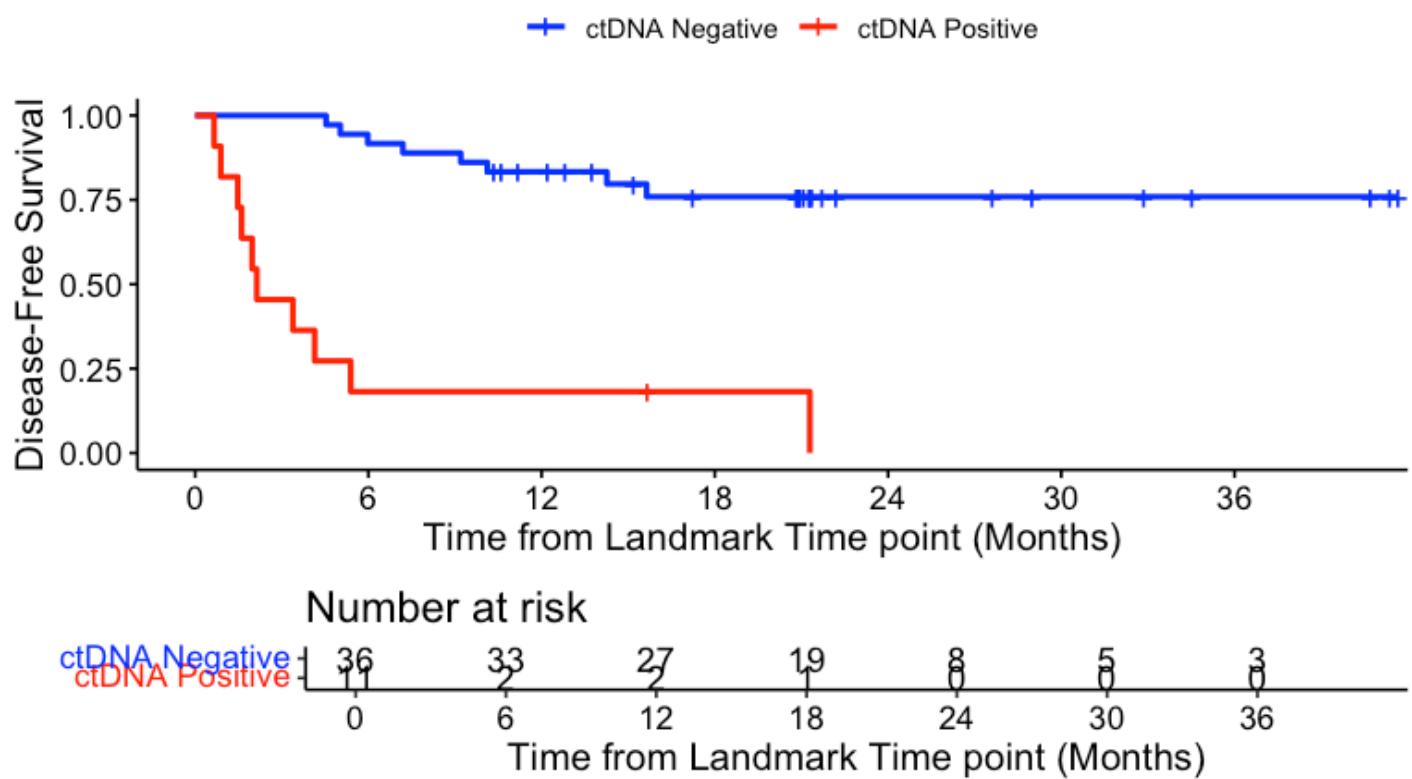
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	36	8	0.2222222	22.22222
POSITIVE	11	10	0.9090909	90.90909

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 | KRAS G12C", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.lab=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | KRAS G12C



```
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	36	0	1.000	0.0000	1.000	1.000	1.000	1.000
24	8	8	0.759	0.0756	0.572	0.873	0.873	0.873

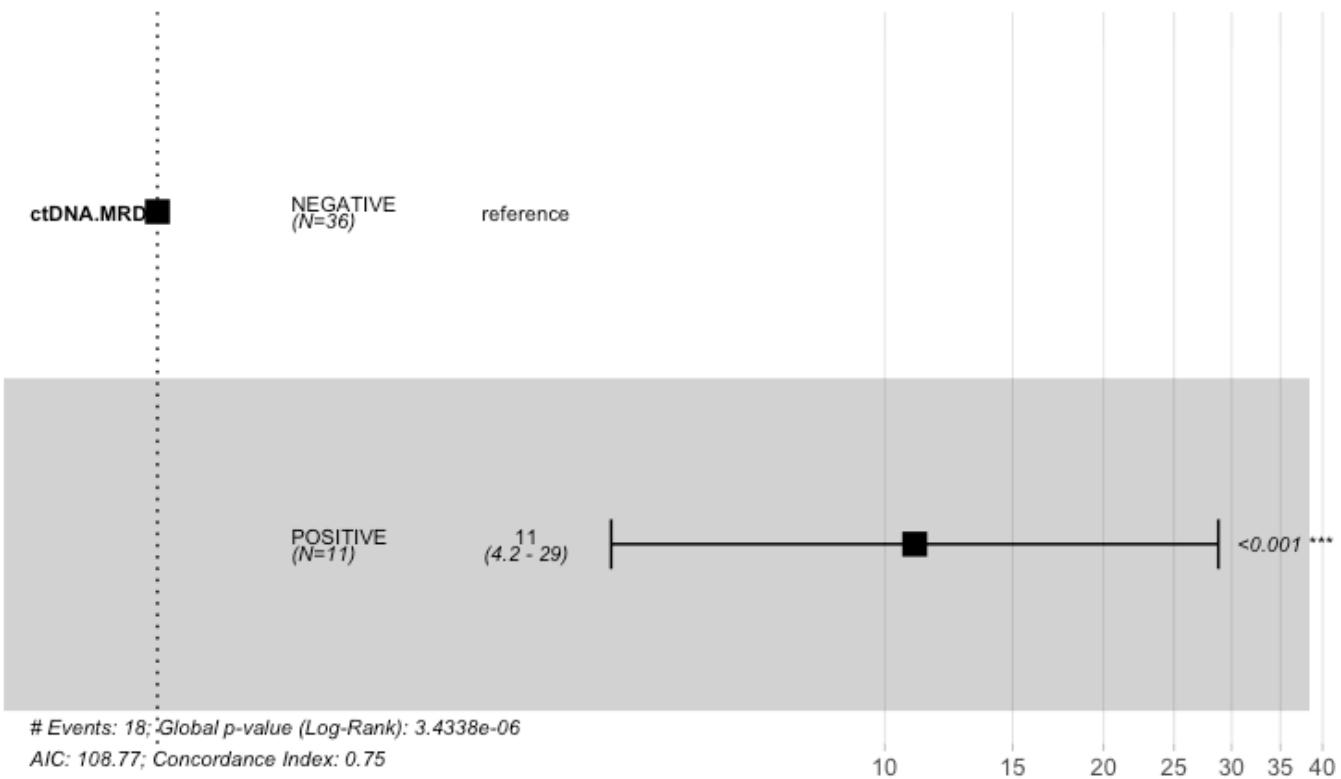
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	11	0	1	0	0	1	1	1
1	0	0	0	0	0	0	0	0

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= 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)
```

```
# 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 (4.21-28.76); p = 0"
```

#DFS by ctDNA at the MRD Window - ERBB2 Amplification 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$ERBB2=="MUT",]
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)

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	23	3	NA	NA	NA
ctDNA.MRD=POSITIVE	10	9	4.8	1.84	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)
```

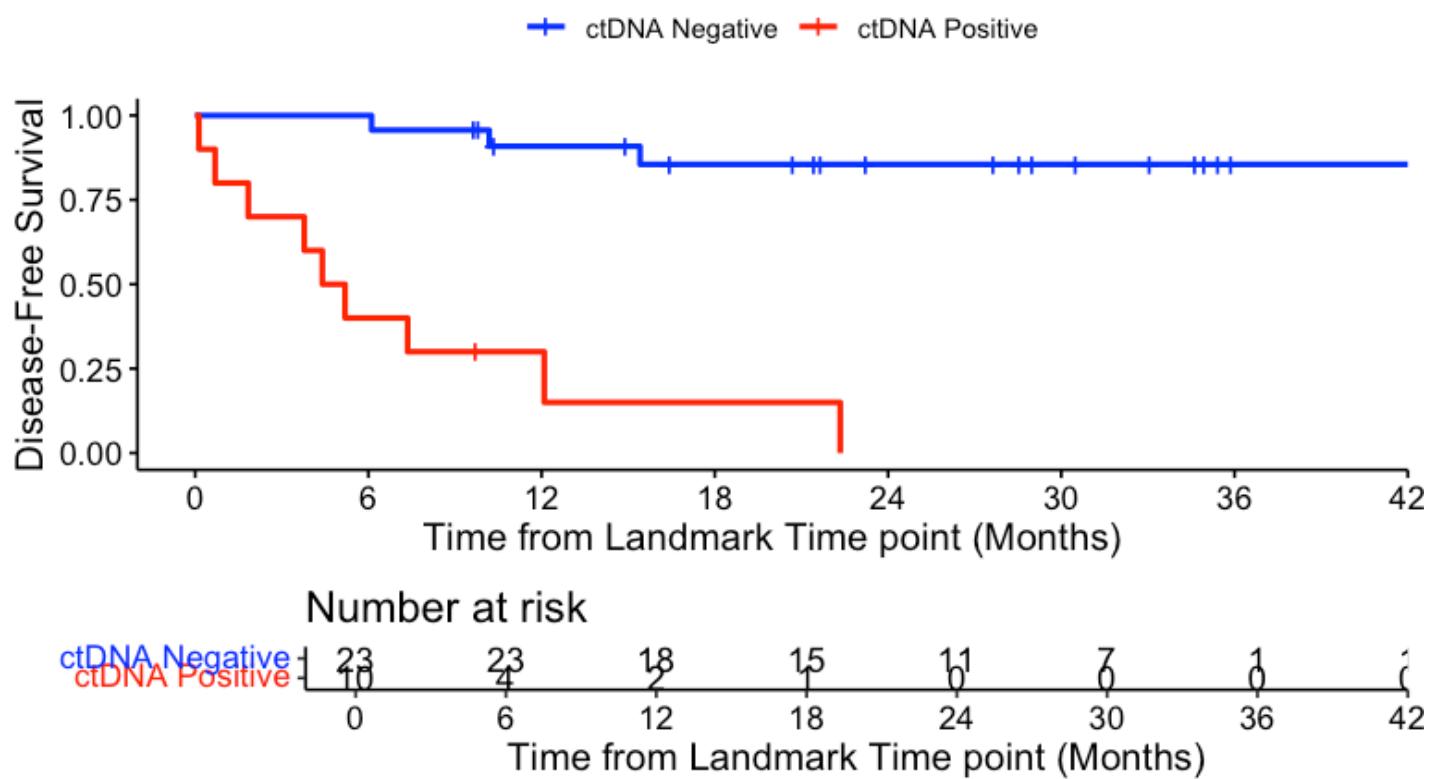
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	23	3	0.1304348	13.04348
POSITIVE	10	9	0.9000000	90.00000

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 | ERBB2 Amplification", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | ERBB2 Amplification



```
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")
```

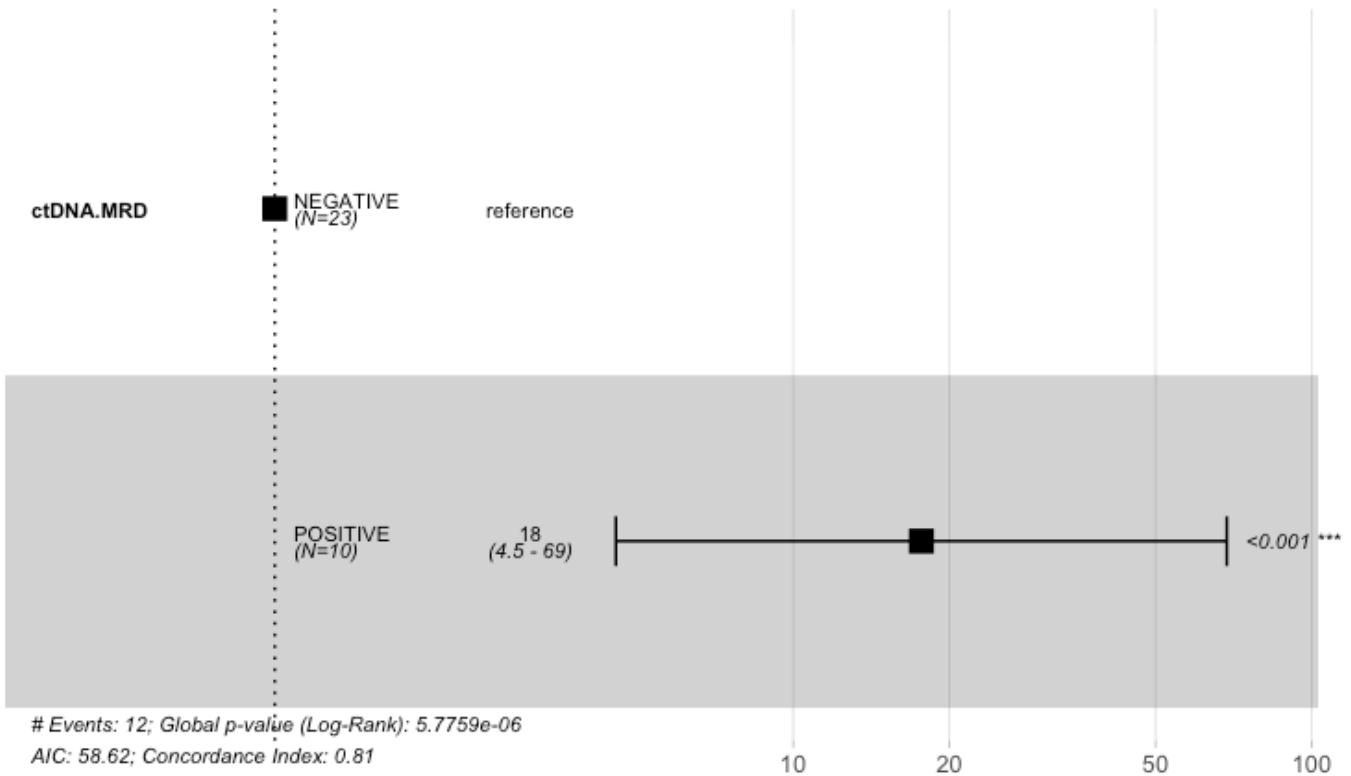
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI
0	23	0	1.000	0.0000	1.000	1.000
24	11	3	0.855	0.0778	0.613	0.951

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI
0	10	0	1	0	0	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= 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)
```

```
# 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 = 17.67 (4.55-68.61); p = 0"
```

#DFS by ctDNA at the MRD Window - TP53 Y220C 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$Eligible=="TRUE",]
circ_data <- circ_data[circ_data$TP53.Y220C=="MUT",]
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)

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	18	3	NA	NA	NA
ctDNA.MRD=POSITIVE	4	2	5.39	1.77	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)
```

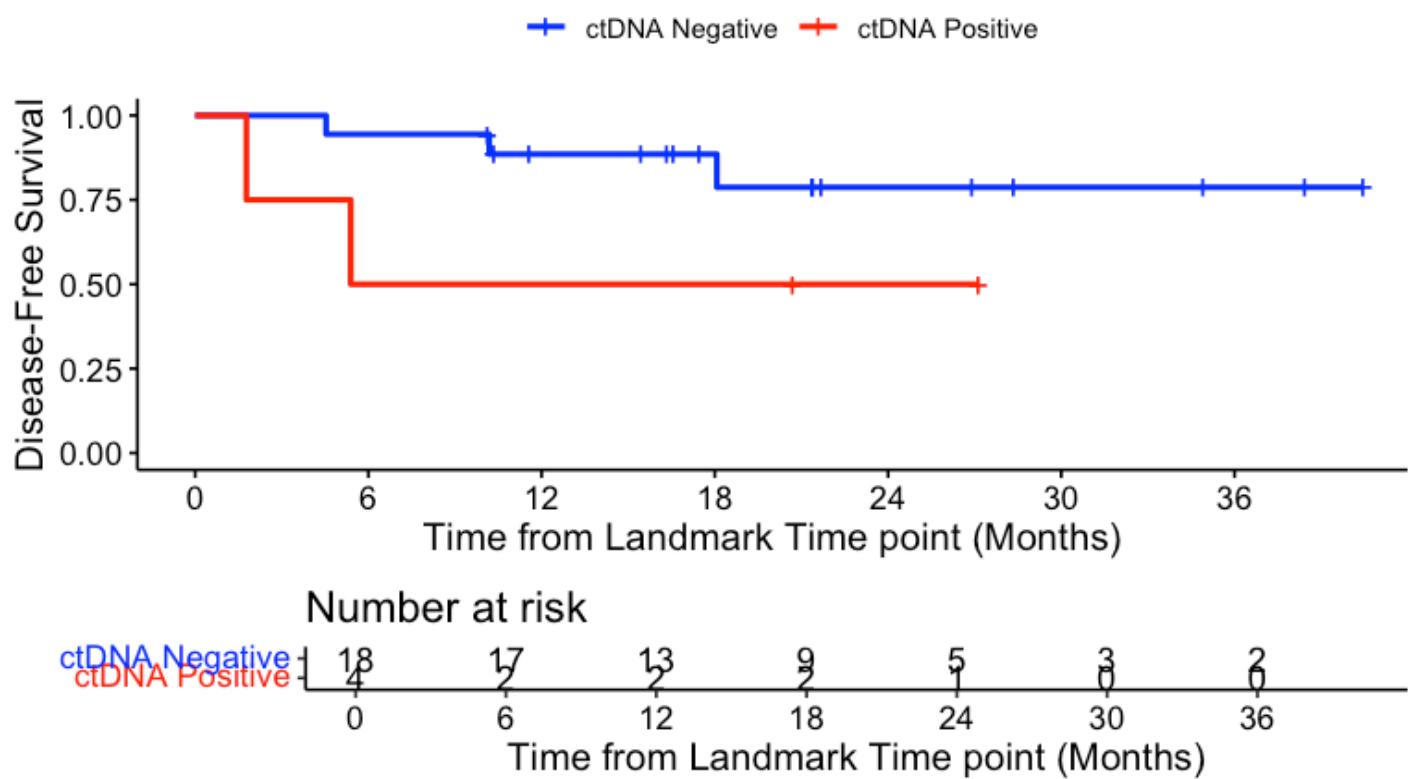
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	18	3	0.1666667	16.66667
POSITIVE	4	2	0.5000000	50.00000

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 | TP53 Y220C", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | TP53 Y220C



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

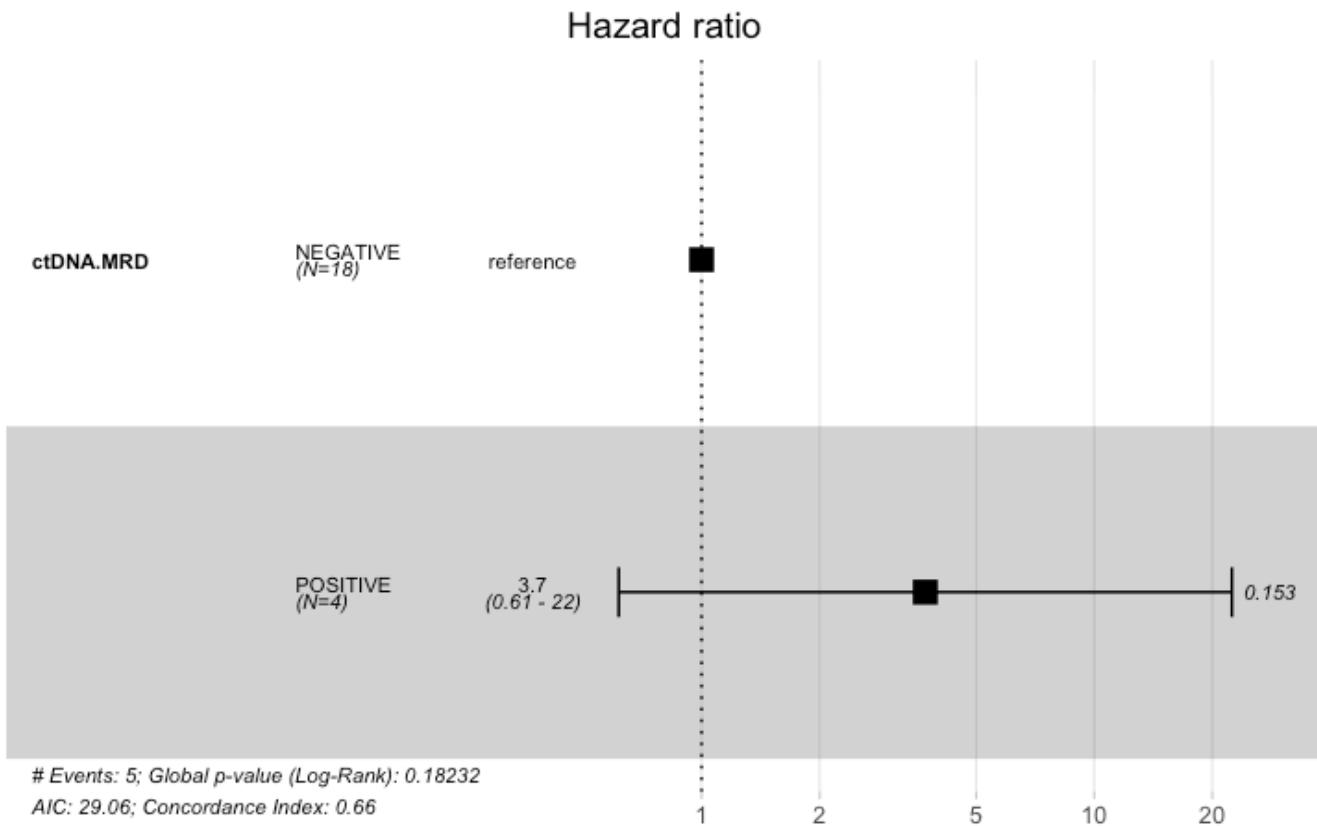
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	18	0	1.000	0.000	1.000		1.00	
24	5	3	0.787	0.115	0.453		0.93	

ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	4	0	1.0	0.00	1.0000		1.000	
24	1	2	0.5	0.25	0.0578		0.845	

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)
```



Hide

```
summary(cox_fit)
```

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

n= 22, number of events= 5

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.MRDPOSITIVE	1.3124	3.7152	0.9177	1.43	0.153

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.MRDPOSITIVE	3.715	0.2692	0.6149	22.45

Concordance= 0.661 (se = 0.115)
 Likelihood ratio test= 1.78 on 1 df, p=0.2
 Wald test = 2.05 on 1 df, p=0.2
 Score (logrank) test = 2.35 on 1 df, p=0.1

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 = 3.72 (0.61-22.45); p = 0.153"

#DFS by ctDNA at the MRD Window - Forest plot with all subgroups of biomarkers

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$DFS.MRD.months >= 0,]
perform_cox <- function(data, filter_col = NULL, filter_val = NULL) {
  if (!is.null(filter_col) & !is.null(filter_val)) {
    data <- data[data[[filter_col]] == filter_val,]
  }
  surv_object <- Surv(time = data$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))
}

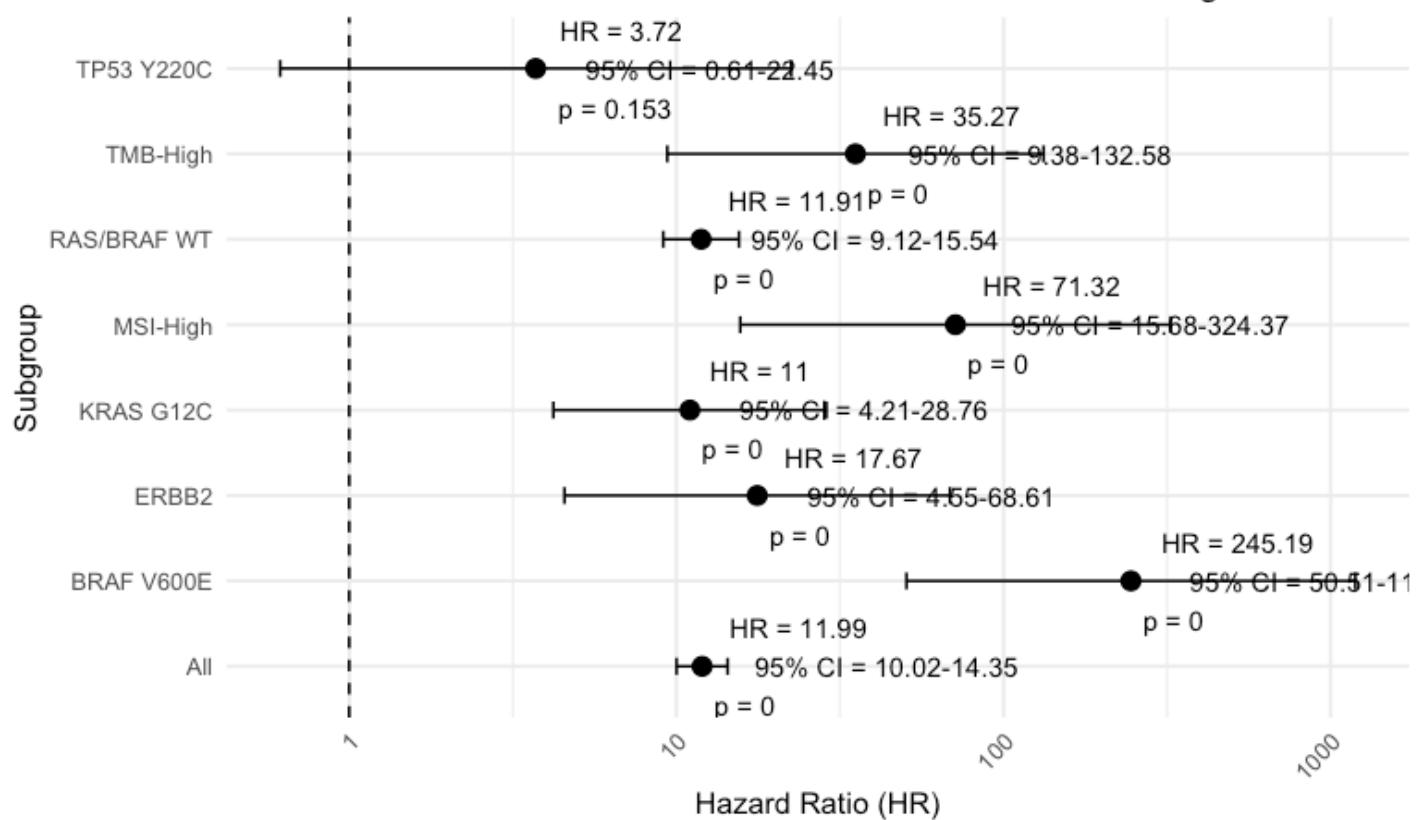
results <- data.frame(
  Subgroup = c("All", "RAS/BRAF WT", "TMB-High", "MSI-High", "BRAF V600E", "KRAS G12C",
  "ERBB2", "TP53 Y220C"),
  HR = rep(NA, 8),
  lower_CI = rep(NA, 8),
  upper_CI = rep(NA, 8),
  p_value = rep(NA, 8)
)

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))

```

Forest Plot of HR for DFS between ctDNA Positive versus Negative



#DFS by BRAF & MSI - ctDNA Positive 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$ctDNA.MRD == "POSITIVE",]
circ_data <- circ_data[circ_data$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
	<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

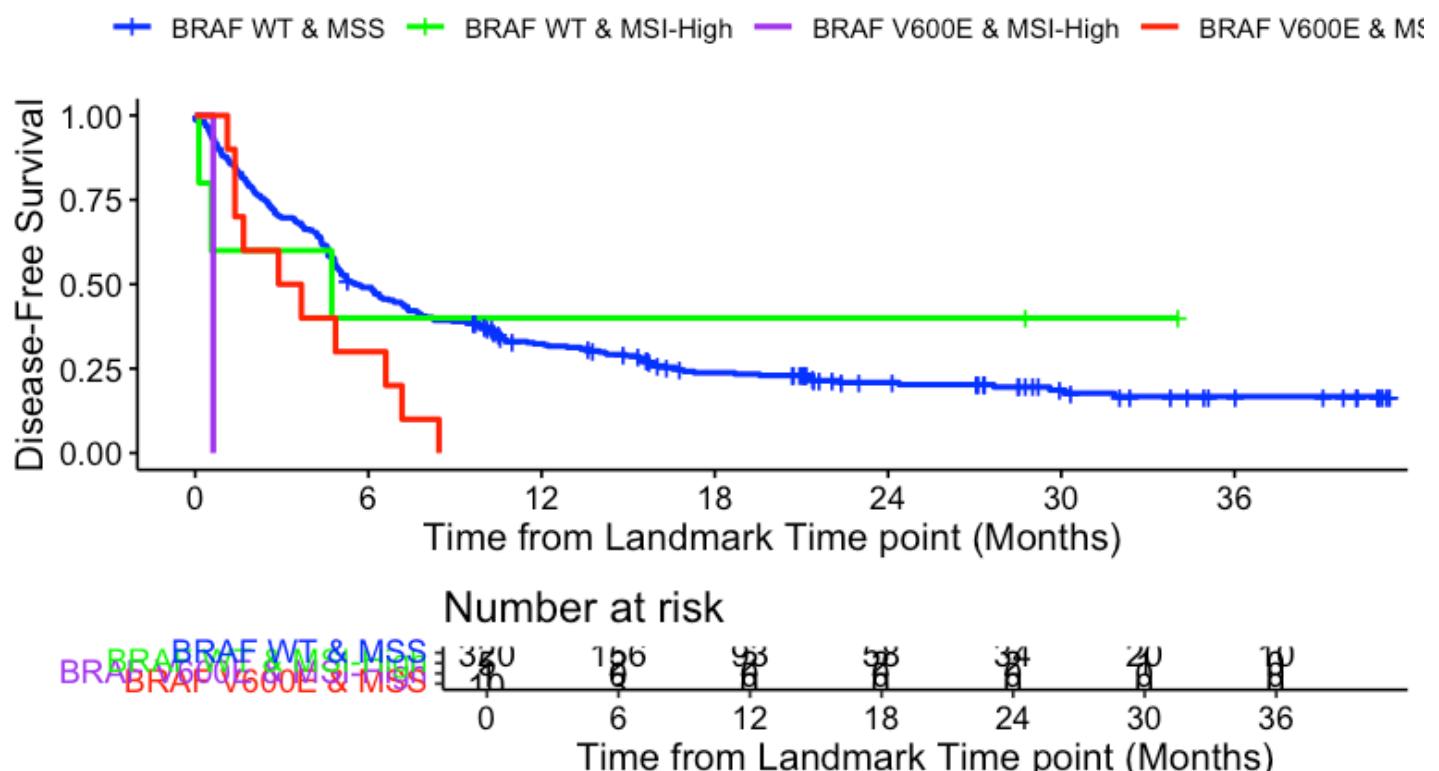
4 rows

[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 = "")
```

DFS - BRAF & MSI | ctDNA MRD Positive



```
summary(KM_curve, times = c(0, 24))
```

```
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

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
1	0	1	0	1	0	0		1	

BRAF.MSI=BRAF V600E & MSS

CI	time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
1	0	10	0	1	0	0		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

#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.MSI	Total	Events	Fraction	Percentage
<fctr>	<int>	<int>	<dbl>	<dbl>
BRAF V600E & MSI-High	103	3	0.02912621	2.912621
BRAF V600E & MSS	49	9	0.18367347	18.367347
4 rows				

Hide

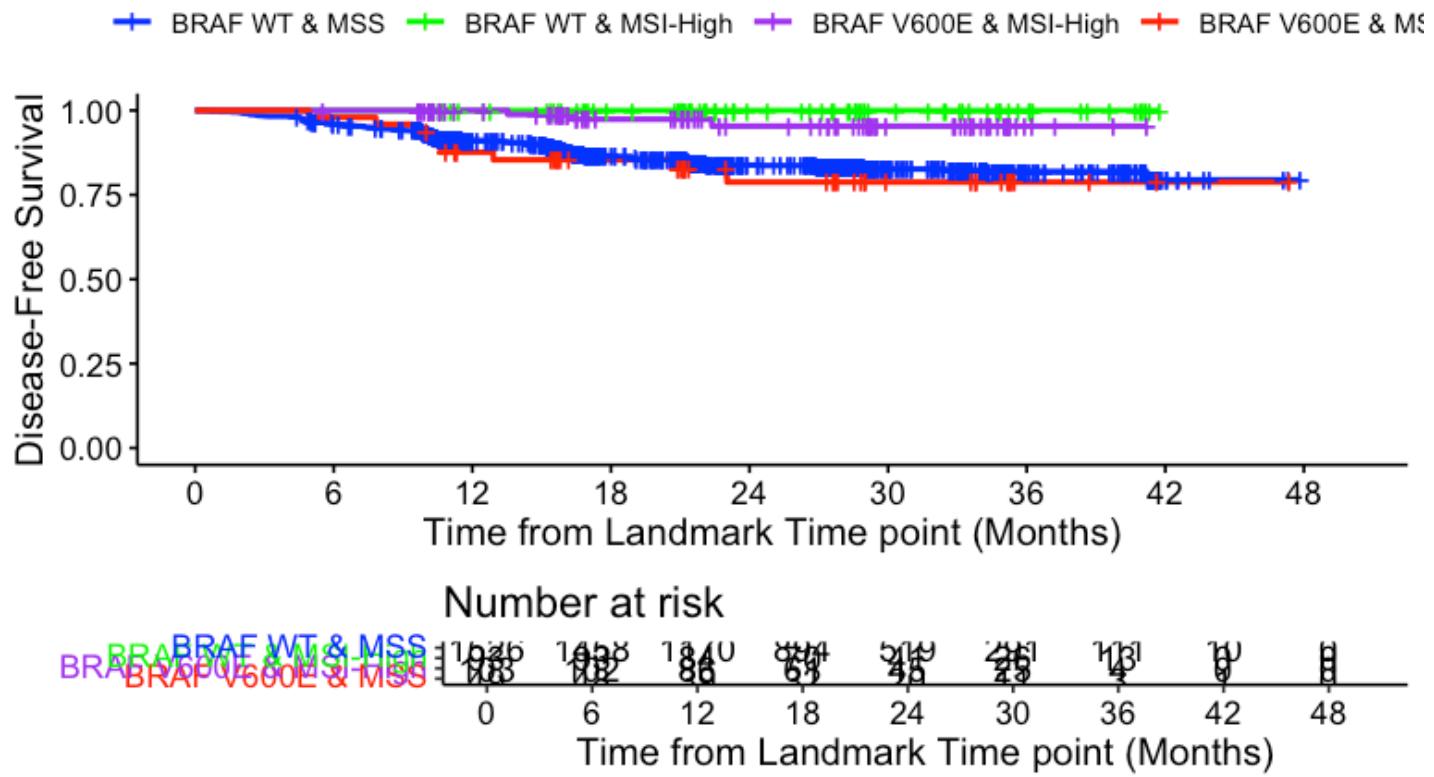
```

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 = "")

```

DFS - BRAF & MSI | ctDNA MRD Negative



```
summary(KM_curve, times = c(0, 24))
```

```
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    1.000  0.0000    1.000    1.000
 24     41      0    1.000  0.0000    1.000    1.000
```

```
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
```

```
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
5.2847247	4.946103e-07				
BRAF.MSIBRAF V600E & MSI-High	-1.5067027	0.5411239	0.22163959	0.0620647658	0.5473653
3.3792192	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	B
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			