CIRCULATE Galaxy cohort B Ando et al 2024

Code ▾

library(swimplot) library(coxphf) library(grid) library(gtable) library(readr) library(mosaic) library(dplyr) library(officer) library(flextable) library(survival) library(survminer) library(gridtext) 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$CohortB=="TRUE",]

circ_data_subset <- circ_data %>%
  select(
    Age,
    Gender,
    ECOG,
    pT,
    pN,
    LVI,
    PNI,
    Grade,
    Location,
    Surg.Type,
    Stage,
    ACT,
    BRAF.V600E,
    RAS,
    MSI,
    RFS.Event,
    OS.Event,
    OS.months) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    ECOG = factor(ECOG, levels = c(0, 1)),
    pT = factor(pT, levels = c("T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    LVI = factor(LVI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    PNI = factor(PNI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    Grade = factor(Grade, levels = c("G1", "G2/G3")),
    Location = factor(Location, levels = c("Upper rectum (Ra)", "Lower rectum (Rb)")),
    Surg.Type = factor(Surg.Type),
    Stage = factor(Stage, levels = c("II", "III")),
    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.Event = factor(OS.Event, levels = c("TRUE", "FALSE"), labels = c("Deceased", "Ali
ve")),
    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 = 250[1]
Age	67 (38 - 92)
Gender	
Male	162 (65%)
Female	88 (35%)
ECOG	
0	228 (91%)
1	22 (8.8%)
pT	
T1-T2	29 (12%)
T3-T4	221 (88%)
pN	
N0	99 (40%)
N1-N2	151 (60%)
LVI	220 (88%)
PNI	109 (44%)
Grade	
G1	144 (58%)
G2/G3	106 (42%)
Location	
Upper rectum (Ra)	127 (51%)
Lower rectum (Rb)	123 (49%)
Surg.Type	
APR	35 (14%)
HAR	8 (3.2%)

[1] Median (Min - Max); n (%)

Characteristic	N = 250[1]
Hartmann procedure	2 (0.8%)
ISR	11 (4.4%)
LAR	189 (76%)
Other	4 (1.6%)
TPE	1 (0.4%)
Stage	
II	100 (40%)
III	150 (60%)
ACT	
Adjuvant Chemotherapy	123 (49%)
Observation	127 (51%)
BRAF.V600E	
BRAF wt	248 (99%)
BRAF V600E	2 (0.8%)
RAS	
RAS wt	122 (49%)
RAS mut	128 (51%)
MSI	
MSS	247 (99%)
MSI-High	3 (1.2%)
RFS.Event	
Recurrence	48 (19%)
No Recurrence	202 (81%)
OS.Event	
Deceased	11 (4.4%)
Alive	239 (96%)

[1] Median (Min - Max); n (%)

Characteristic	N = 250[1]
OS.months	22.0 (1.2 – 35.9)

[1] Median (Min – Max); n (%)

Hide

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

Characteristic	N = 250[1]
Age	67 (38 - 92)
Gender	
Male	162 (65%)
Female	88 (35%)
ECOG	
0	228 (91%)
1	22 (8.8%)
pT	
T1-T2	29 (12%)
T3-T4	221 (88%)
pN	
N0	99 (40%)
N1-N2	151 (60%)
LVI	220 (88%)
PNI	109 (44%)
Grade	
G1	144 (58%)
G2/G3	106 (42%)
Location	
Upper rectum (Ra)	127 (51%)
Lower rectum (Rb)	123 (49%)
Surg.Type	

[1]Median (Min - Max); n (%)

Characteristic	N = 250[1]
APR	35 (14%)
HAR	8 (3.2%)
Hartmann procedure	2 (0.8%)
ISR	11 (4.4%)
LAR	189 (76%)
Other	4 (1.6%)
TPE	1 (0.4%)
Stage	
II	100 (40%)
III	150 (60%)
ACT	
Adjuvant Chemotherapy	123 (49%)
Observation	127 (51%)
BRAF.V600E	
BRAF wt	248 (99%)
BRAF V600E	2 (0.8%)
RAS	
RAS wt	122 (49%)
RAS mut	128 (51%)
MSI	
MSS	247 (99%)
MSI-High	3 (1.2%)
RFS.Event	
Recurrence	48 (19%)
No Recurrence	202 (81%)
OS.Event	
Deceased	11 (4.4%)
Alive	239 (96%)
OS.months	22.0 (1.2 - 35.9)

[1]Median (Min - Max); n (%)

Hide

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

#Demographics Table by MRD ctDNA Status

Hide

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

circ_data_subset1 <- circ_data %>%
  select(
    Age,
    Gender,
    ECOG,
    pT,
    pN,
    LVI,
    PNI,
    Grade,
    Location,
    Surg.Type,
    Stage,
    ACT,
    BRAF.V600E,
    RAS,
    MSI,
    RFS.Event,
    OS.Event,
    OS.months) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    ECOG = factor(ECOG, levels = c(0, 1)),
    pT = factor(pT, levels = c("T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    LVI = factor(LVI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    PNI = factor(PNI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    Grade = factor(Grade, levels = c("G1", "G2/G3")),
    Location = factor(Location, levels = c("Upper rectum (Ra)", "Lower rectum (Rb)")),
    Surg.Type = factor(Surg.Type),
    Stage = factor(Stage, levels = c("II", "III")),
    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.Event = factor(OS.Event, levels = c("TRUE", "FALSE"), labels = c("Deceased", "Ali
ve")),
    OS.months = as.numeric(OS.months))

circ_data1 <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
```

```r
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]

circ_data_subset2 <- circ_data %>%
  select(
    Age,
    Gender,
    ECOG,
    pT,
    pN,
    LVI,
    PNI,
    Grade,
    Location,
    Surg.Type,
    Stage,
    ACT,
    BRAF.V600E,
    RAS,
    MSI,
    RFS.Event,
    OS.Event,
    OS.months,
    ctDNA.MRD) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    ECOG = factor(ECOG, levels = c(0, 1)),
    pT = factor(pT, levels = c("T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    LVI = factor(LVI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    PNI = factor(PNI, levels = c("TRUE", "FALSE"), labels = c("Yes", "No")),
    Grade = factor(Grade, levels = c("G1", "G2/G3")),
    Location = factor(Location, levels = c("Upper rectum (Ra)", "Lower rectum (Rb)")),
    Surg.Type = factor(Surg.Type),
    Stage = factor(Stage, levels = c("II", "III")),
    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.Event = factor(OS.Event, levels = c("TRUE", "FALSE"), labels = c("Deceased", "Ali
ve")),
    OS.months = as.numeric(OS.months),
    ctDNA.MRD = factor(ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE")))
Overall <- circ_data_subset1 %>%
  tbl_summary(
    statistic = list(
      all_continuous() ~ "{median} ({min} - {max})",
      all_categorical() ~ "{n} ({p}%)")) %>%
```

```
    bold_labels()
Overall
```

Characteristic	N = 247[1]
Age	67 (38 - 92)
Gender	
Male	162 (66%)
Female	85 (34%)
ECOG	
0	225 (91%)
1	22 (8.9%)
pT	
T1-T2	29 (12%)
T3-T4	218 (88%)
pN	
N0	98 (40%)
N1-N2	149 (60%)
LVI	218 (88%)
PNI	107 (43%)
Grade	
G1	142 (57%)
G2/G3	105 (43%)
Location	
Upper rectum (Ra)	126 (51%)
Lower rectum (Rb)	121 (49%)
Surg.Type	
APR	35 (14%)
HAR	8 (3.2%)
[1] Median (Min - Max); n (%)	

Characteristic	N = 247[1]
Hartmann procedure	2 (0.8%)
ISR	11 (4.5%)
LAR	186 (75%)
Other	4 (1.6%)
TPE	1 (0.4%)
Stage	
II	99 (40%)
III	148 (60%)
ACT	
Adjuvant Chemotherapy	121 (49%)
Observation	126 (51%)
BRAF.V600E	
BRAF wt	245 (99%)
BRAF V600E	2 (0.8%)
RAS	
RAS wt	121 (49%)
RAS mut	126 (51%)
MSI	
MSS	244 (99%)
MSI-High	3 (1.2%)
RFS.Event	
Recurrence	48 (19%)
No Recurrence	199 (81%)
OS.Event	
Deceased	11 (4.5%)
Alive	236 (96%)

[1] Median (Min - Max); n (%)

Characteristic	N = 247[1]
OS.months	22.0 (1.2 - 35.9)

[1] Median (Min – Max); n (%)

Hide

```
ByctDNA_MRD <- circ_data_subset2 %>%
  tbl_summary(
    by = ctDNA.MRD, # add this line to subgroup by ctDNA.MRD
    statistic = list(
      all_continuous() ~ "{median} ({min} – {max})",
      all_categorical() ~ "{n} ({p}%)")) %>%
  add_p() %>%
  bold_labels()
ByctDNA_MRD
```

Characteristic	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
Age	67 (38 - 92)	67 (43 - 82)	0.6
Gender			0.5
Male	141 (67%)	21 (60%)	
Female	71 (33%)	14 (40%)	
ECOG			0.3
0	191 (90%)	34 (97%)	
1	21 (9.9%)	1 (2.9%)	
pT			0.8
T1-T2	26 (12%)	3 (8.6%)	
T3-T4	186 (88%)	32 (91%)	
pN			<0.001
N0	95 (45%)	3 (8.6%)	
N1–N2	117 (55%)	32 (91%)	
LVI	184 (87%)	34 (97%)	0.092

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Characteristic	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
PNI	86 (41%)	21 (60%)	0.032
Grade			0.7
G1	123 (58%)	19 (54%)	
G2/G3	89 (42%)	16 (46%)	
Location			0.5
Upper rectum (Ra)	110 (52%)	16 (46%)	
Lower rectum (Rb)	102 (48%)	19 (54%)	
Surg.Type			0.5
APR	27 (13%)	8 (23%)	
HAR	7 (3.3%)	1 (2.9%)	
Hartmann procedure	2 (0.9%)	0 (0%)	
ISR	11 (5.2%)	0 (0%)	
LAR	161 (76%)	25 (71%)	
Other	3 (1.4%)	1 (2.9%)	
TPE	1 (0.5%)	0 (0%)	
Stage			<0.001
II	96 (45%)	3 (8.6%)	
III	116 (55%)	32 (91%)	
ACT			0.004
Adjuvant Chemotherapy	96 (45%)	25 (71%)	
Observation	116 (55%)	10 (29%)	
BRAF.V600E			>0.9
BRAF wt	210 (99%)	35 (100%)	
BRAF V600E	2 (0.9%)	0 (0%)	
RAS			0.060

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Characteristic	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
RAS wt	109 (51%)	12 (34%)	
RAS mut	103 (49%)	23 (66%)	
MSI			>0.9
MSS	209 (99%)	35 (100%)	
MSI-High	3 (1.4%)	0 (0%)	
RFS.Event			<0.001
Recurrence	23 (11%)	25 (71%)	
No Recurrence	189 (89%)	10 (29%)	
OS.Event			0.2
Deceased	8 (3.8%)	3 (8.6%)	
Alive	204 (96%)	32 (91%)	
OS.months	22.0 (7.5 - 35.9)	16.4 (1.2 - 34.9)	0.082

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Hide

```
merged_table <- tbl_merge(tbls=list(Overall, ByctDNA_MRD))
merged_table
```

	Table 1	Table 2		
Characteristic	N = 247[1]	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
Age	67 (38 - 92)	67 (38 - 92)	67 (43 - 82)	0.6
Gender				0.5
Male	162 (66%)	141 (67%)	21 (60%)	
Female	85 (34%)	71 (33%)	14 (40%)	
ECOG				0.3

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Characteristic	Table 1	Table 2		
	N = 247[1]	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
0	225 (91%)	191 (90%)	34 (97%)	
1	22 (8.9%)	21 (9.9%)	1 (2.9%)	
pT				0.8
T1-T2	29 (12%)	26 (12%)	3 (8.6%)	
T3-T4	218 (88%)	186 (88%)	32 (91%)	
pN				<0.001
N0	98 (40%)	95 (45%)	3 (8.6%)	
N1-N2	149 (60%)	117 (55%)	32 (91%)	
LVI	218 (88%)	184 (87%)	34 (97%)	0.092
PNI	107 (43%)	86 (41%)	21 (60%)	0.032
Grade				0.7
G1	142 (57%)	123 (58%)	19 (54%)	
G2/G3	105 (43%)	89 (42%)	16 (46%)	
Location				0.5
Upper rectum (Ra)	126 (51%)	110 (52%)	16 (46%)	
Lower rectum (Rb)	121 (49%)	102 (48%)	19 (54%)	
Surg.Type				0.5
APR	35 (14%)	27 (13%)	8 (23%)	
HAR	8 (3.2%)	7 (3.3%)	1 (2.9%)	
Hartmann procedure	2 (0.8%)	2 (0.9%)	0 (0%)	
ISR	11 (4.5%)	11 (5.2%)	0 (0%)	
LAR	186 (75%)	161 (76%)	25 (71%)	
Other	4 (1.6%)	3 (1.4%)	1 (2.9%)	
TPE	1 (0.4%)	1 (0.5%)	0 (0%)	

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

	Table 1	Table 2		
Characteristic	N = 247[1]	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
Stage				<0.001
II	99 (40%)	96 (45%)	3 (8.6%)	
III	148 (60%)	116 (55%)	32 (91%)	
ACT				0.004
Adjuvant Chemotherapy	121 (49%)	96 (45%)	25 (71%)	
Observation	126 (51%)	116 (55%)	10 (29%)	
BRAF.V600E				>0.9
BRAF wt	245 (99%)	210 (99%)	35 (100%)	
BRAF V600E	2 (0.8%)	2 (0.9%)	0 (0%)	
RAS				0.060
RAS wt	121 (49%)	109 (51%)	12 (34%)	
RAS mut	126 (51%)	103 (49%)	23 (66%)	
MSI				>0.9
MSS	244 (99%)	209 (99%)	35 (100%)	
MSI-High	3 (1.2%)	3 (1.4%)	0 (0%)	
RFS.Event				<0.001
Recurrence	48 (19%)	23 (11%)	25 (71%)	
No Recurrence	199 (81%)	189 (89%)	10 (29%)	
OS.Event				0.2
Deceased	11 (4.5%)	8 (3.8%)	3 (8.6%)	
Alive	236 (96%)	204 (96%)	32 (91%)	
OS.months	22.0 (1.2 – 35.9)	22.0 (7.5 – 35.9)	16.4 (1.2 – 34.9)	0.082

[1] Median (Min - Max); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Hide

```
fit1 <- as_flex_table(
  merged_table,
  include = everything(),
  return_calls = FALSE)
fit1
```

| | Table 1 | Table 2 | | |
Characteristic	N = 247[1]	NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
Age	67 (38 - 92)	67 (38 - 92)	67 (43 - 82)	0.6
Gender				0.5
Male	162 (66%)	141 (67%)	21 (60%)	
Female	85 (34%)	71 (33%)	14 (40%)	
ECOG				0.3
0	225 (91%)	191 (90%)	34 (97%)	
1	22 (8.9%)	21 (9.9%)	1 (2.9%)	
pT				0.8
T1-T2	29 (12%)	26 (12%)	3 (8.6%)	
T3-T4	218 (88%)	186 (88%)	32 (91%)	
pN				<0.001
N0	98 (40%)	95 (45%)	3 (8.6%)	
N1-N2	149 (60%)	117 (55%)	32 (91%)	
LVI	218 (88%)	184 (87%)	34 (97%)	0.092
PNI	107 (43%)	86 (41%)	21 (60%)	0.032
Grade				0.7
G1	142 (57%)	123 (58%)	19 (54%)	
G2/G3	105 (43%)	89 (42%)	16 (46%)	
Location				0.5
Upper rectum (Ra)	126 (51%)	110 (52%)	16 (46%)	
Lower rectum (Rb)	121 (49%)	102 (48%)	19 (54%)	
Surg.Type				0.5
APR	35 (14%)	27 (13%)	8 (23%)	

[1]Median (Min - Max); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Characteristic	Table 1 N = 247[1]	Table 2 NEGATIVE N = 212[1]	POSITIVE N = 35[1]	p-value[2]
HAR	8 (3.2%)	7 (3.3%)	1 (2.9%)	
Hartmann procedure	2 (0.8%)	2 (0.9%)	0 (0%)	
ISR	11 (4.5%)	11 (5.2%)	0 (0%)	
LAR	186 (75%)	161 (76%)	25 (71%)	
Other	4 (1.6%)	3 (1.4%)	1 (2.9%)	
TPE	1 (0.4%)	1 (0.5%)	0 (0%)	
Stage				<0.001
II	99 (40%)	96 (45%)	3 (8.6%)	
III	148 (60%)	116 (55%)	32 (91%)	
ACT				0.004
Adjuvant Chemotherapy	121 (49%)	96 (45%)	25 (71%)	
Observation	126 (51%)	116 (55%)	10 (29%)	
BRAF.V600E				>0.9
BRAF wt	245 (99%)	210 (99%)	35 (100%)	
BRAF V600E	2 (0.8%)	2 (0.9%)	0 (0%)	
RAS				0.060
RAS wt	121 (49%)	109 (51%)	12 (34%)	
RAS mut	126 (51%)	103 (49%)	23 (66%)	
MSI				>0.9
MSS	244 (99%)	209 (99%)	35 (100%)	
MSI-High	3 (1.2%)	3 (1.4%)	0 (0%)	
RFS.Event				<0.001
Recurrence	48 (19%)	23 (11%)	25 (71%)	
No Recurrence	199 (81%)	189 (89%)	10 (29%)	
OS.Event				0.2
Deceased	11 (4.5%)	8 (3.8%)	3 (8.6%)	
Alive	236 (96%)	204 (96%)	32 (91%)	
OS.months	22.0 (1.2 - 35.9)	22.0 (7.5 - 35.9)	16.4 (1.2 - 34.9)	0.082

[1]Median (Min - Max); n (%)

[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test

Hide

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

#DFS in Complete Cohort (N=250)

Hide

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

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

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

       n events median 0.95LCL 0.95UCL
[1,] 250     53     NA      NA      NA
```

Hide

```
surv_object <-Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ CohortB, data = circ_data,conf.int=0.95,conf.type="log
-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue"), title="DFS - Complete Cohort (n=250)", ylab= "Pro
gression-Free Survival", xlab="Months from Surgery", legend.labs=c("Complete cohort"), l
egend.title="")
```
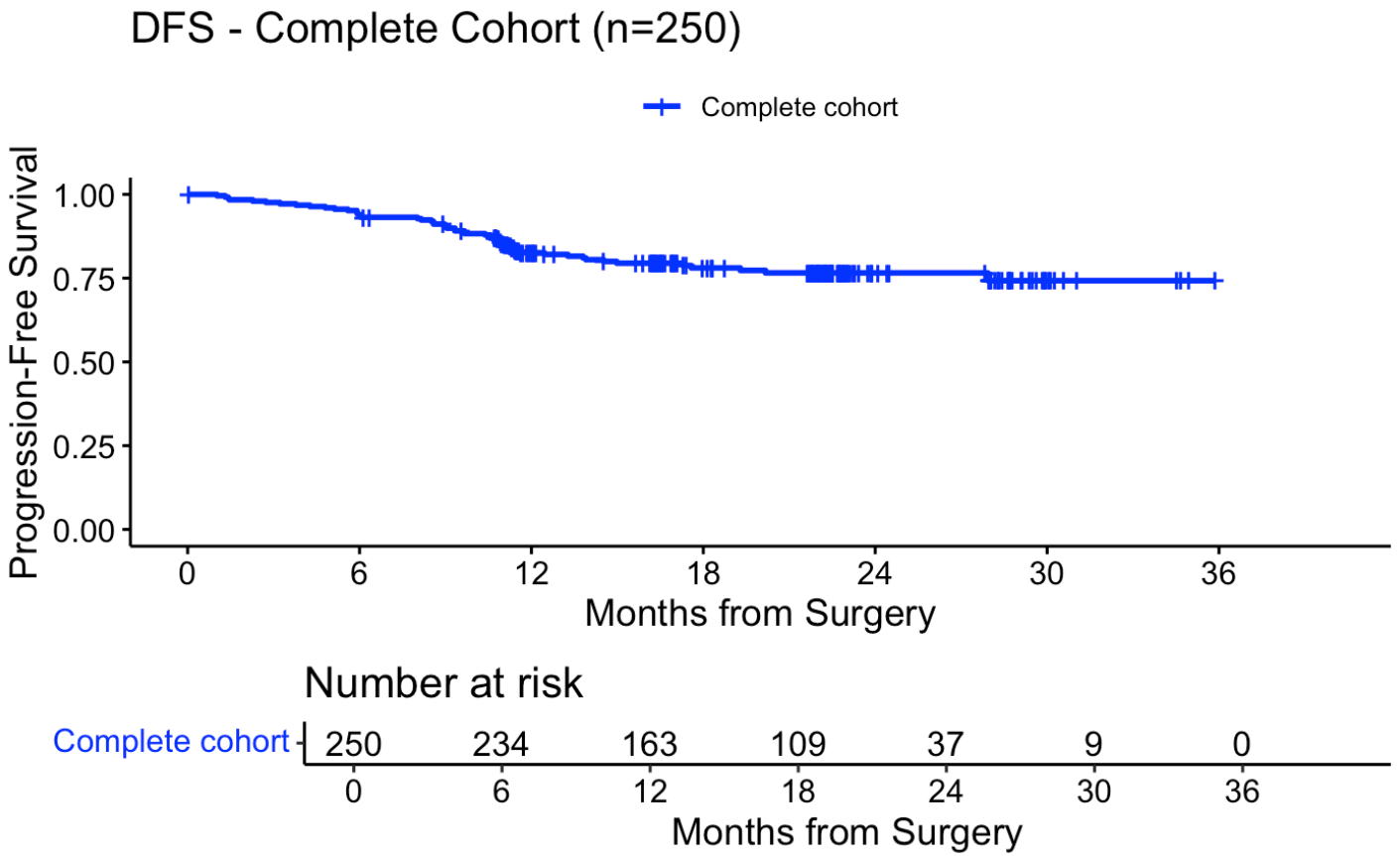
DFS - Complete Cohort (n=250)

Number at risk

Complete cohort	250	234	163	109	37	9	0
	0	6	12	18	24	30	36

Months from Surgery

Hide

```
summary(KM_curve, times= c(12, 24, 36))
```

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

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    163      42    0.826  0.0245        0.772        0.868
   24     37      10    0.766  0.0294        0.702        0.818
```

#OS in Complete Cohort (N=250)

Hide

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

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~CohortB, data = cir
c_data)
```

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

      n events median 0.95LCL 0.95UCL
[1,] 250     11     NA      NA      NA
```

Hide

```
surv_object <-Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ CohortB, data = circ_data,conf.int=0.95,conf.type="log
-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue"), title="OS - Complete Cohort (n=250)", ylab= "Over
all Survival", xlab="Months from Surgery", legend.labs=c("Complete cohort"), legend.titl
e="")
```
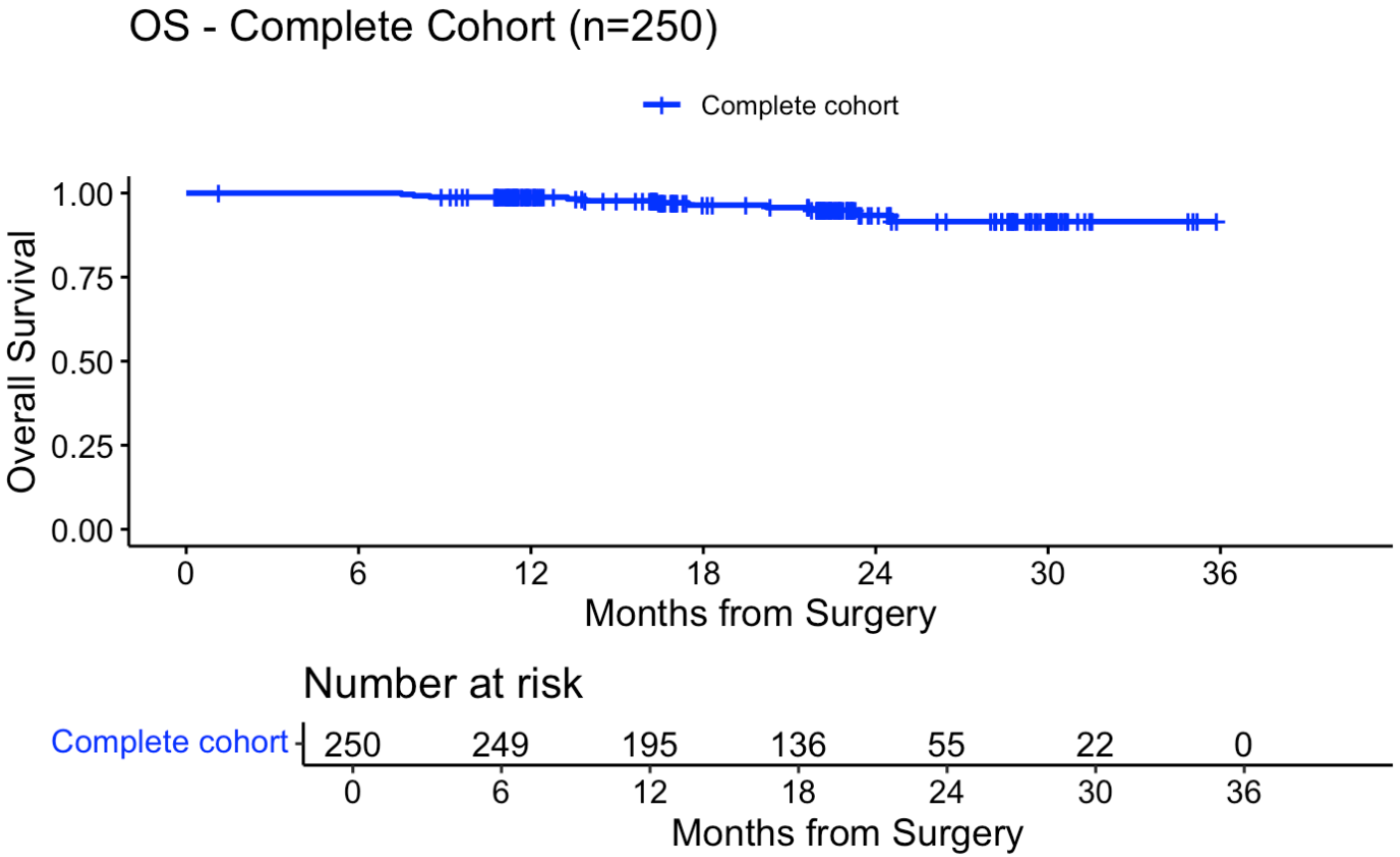
OS - Complete Cohort (n=250)

summary(KM_curve, times= c(12, 24, 36))

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

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    195       3    0.988 0.00691        0.963        0.996
   24     55       7    0.934 0.02271        0.872        0.966
```

#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$CohortB=="TRUE",]
circ_data$ctDNA.Baseline <- factor(circ_data$ctDNA.Baseline, levels=c("NEGATIVE","POSITI
VE"))
circ_data <- subset(circ_data, ctDNA.Baseline %in% c("NEGATIVE", "POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("II", "III"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Baseline == "POSITIVE", by=list(ci
rc_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Baseline, by=list(circ_data$Stage), F
UN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$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 perc
entage
)
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  # Conve
rt 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>
II	97	94	96.91%
III	148	144	97.30%

Stage	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<chr>
Overall	245	238	97.14%
3 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("II", "III"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.MRD == "POSITIVE", by=list(circ_da
ta$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.MRD, by=list(circ_data$Stage), FUN=le
ngth)
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$x) * 100  # Convert to perc
entage
)
combined_data$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  # Conve
rt 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>
II	99	3	3.03%
III	148	32	21.62%
Overall	247	35	14.17%
3 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$CohortB=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIV
E","POSITIVE"))
circ_data <- subset(circ_data, ctDNA.Surveillance %in% c("NEGATIVE", "POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("II", "III"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance == "POSITIVE", by=lis
t(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance, by=list(circ_data$Stag
e), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$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 perc
entage
)
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  # Conve
rt 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>
II	94	8	8.51%
III	139	32	23.02%
Overall	233	40	17.17%

3 rows

#ctDNA MRD Detection rate Stage 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$CohortB=="TRUE",]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"))
circ_data$Stage_Grouped <- factor(ifelse(circ_data$Stage %in% c("II", "II"), "II", "II
I"))
contingency_table <- table(circ_data$Stage_Grouped, circ_data$ctDNA.MRD)
chi_square_test <- chisq.test(contingency_table)
print(contingency_table)
```

```
        NEGATIVE POSITIVE
  II          96        3
  III        116       32
```

Hide

```
print(chi_square_test)
```

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

data:  contingency_table
X-squared = 15.364, df = 1, p-value = 8.865e-05
```

#ctDNA Surveillance Detection rate Stage 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$CohortB=="TRUE",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels = c("NEGATIV
E", "POSITIVE"))
circ_data$Stage_Grouped <- factor(ifelse(circ_data$Stage %in% c("II", "II"), "II", "II
I"))
contingency_table <- table(circ_data$Stage_Grouped, circ_data$ctDNA.Surveillance)
chi_square_test <- chisq.test(contingency_table)
print(contingency_table)
```

```
        NEGATIVE POSITIVE
  II          86        8
  III        107       32
```

Hide

```
print(chi_square_test)
```

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

data:  contingency_table
X-squared = 7.3146, df = 1, p-value = 0.00684
```
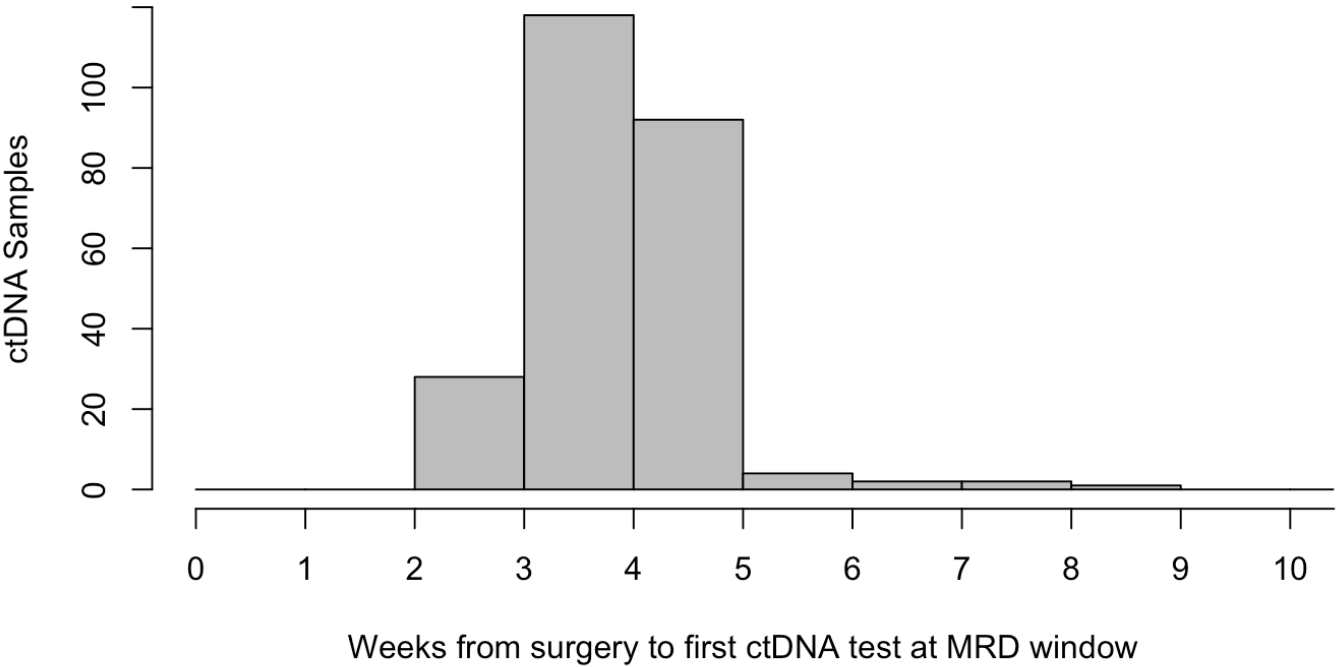
#ctDNA timepoints cadence at the MRD Window

Hide

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

circ_data$MRD_Time_weeks <- circ_data$ctDNA.MRD.Time / 7

# Plot the histogram
hist(circ_data$MRD_Time_weeks,
     col = 'gray',
     main = 'Surgery to first ctDNA testing at MRD window',
     xlab = 'Weeks from surgery to first ctDNA test at MRD window',
     ylab = 'ctDNA Samples',
     ylim = c(0, 120),
     xlim = c(0, 10),
     breaks = seq(0, 24, 1),
     xaxp = c(0, 24, 24))
```

Surgery to first ctDNA testing at MRD window

Weeks from surgery to first ctDNA test at MRD window

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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, da
ta = circ_data)
```

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

                    n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 211     28     NA      NA      NA
ctDNA.MRD=POSITIVE  35     25   9.27    7.62    14.1
```

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	211	28	0.1327014	13.27014
POSITIVE	35	25	0.7142857	71.42857

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="l
og-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, 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)", legen
d.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | All stages
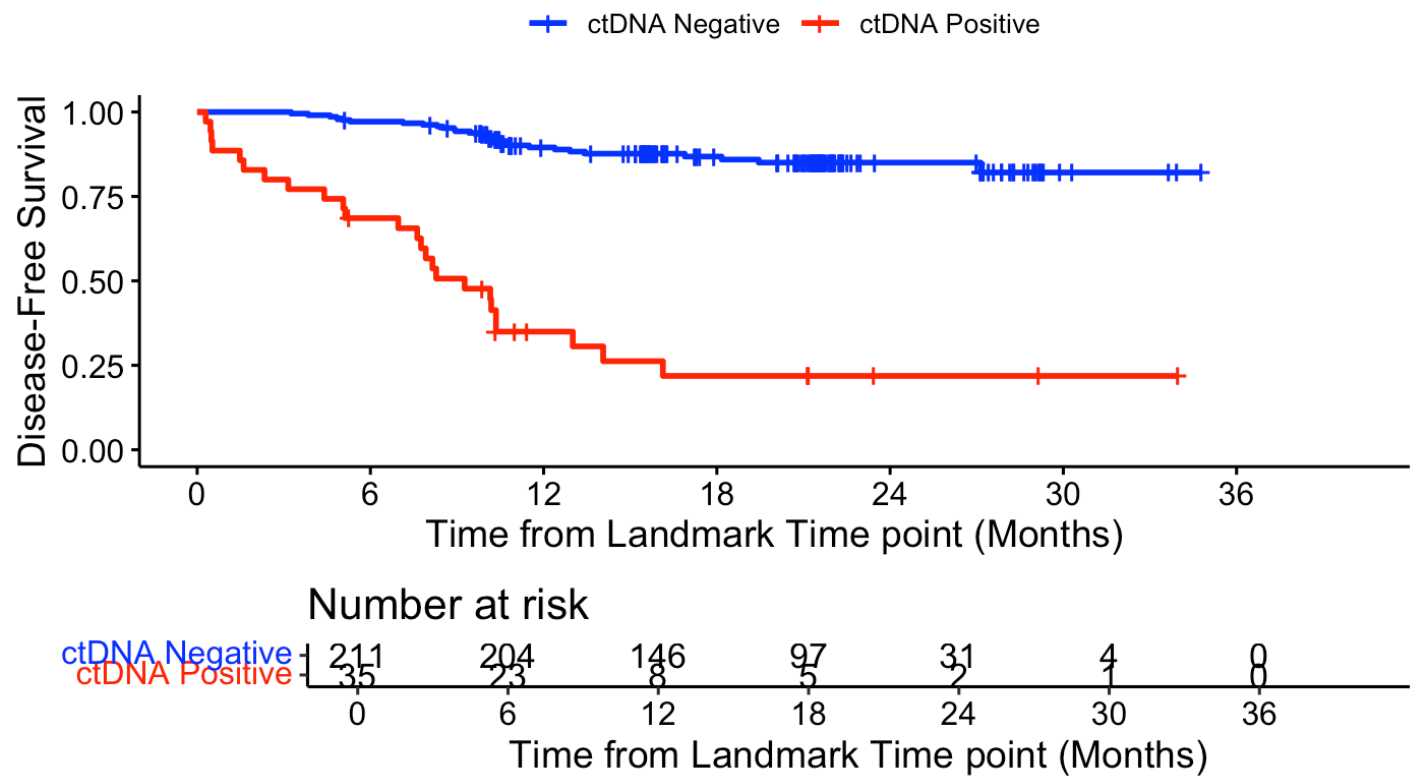

```
summary(KM_curve, times= c(12, 18, 24))
```

```
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
   12    146      21    0.895  0.0218        0.843        0.930
   18     97       4    0.868  0.0250        0.810        0.910
   24     31       2    0.850  0.0275        0.787        0.896


                ctDNA.MRD=POSITIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      8      22    0.350  0.0830       0.1948        0.509
   18      5       3    0.219  0.0792       0.0884        0.386
   24      2       0    0.219  0.0792       0.0884        0.386
```
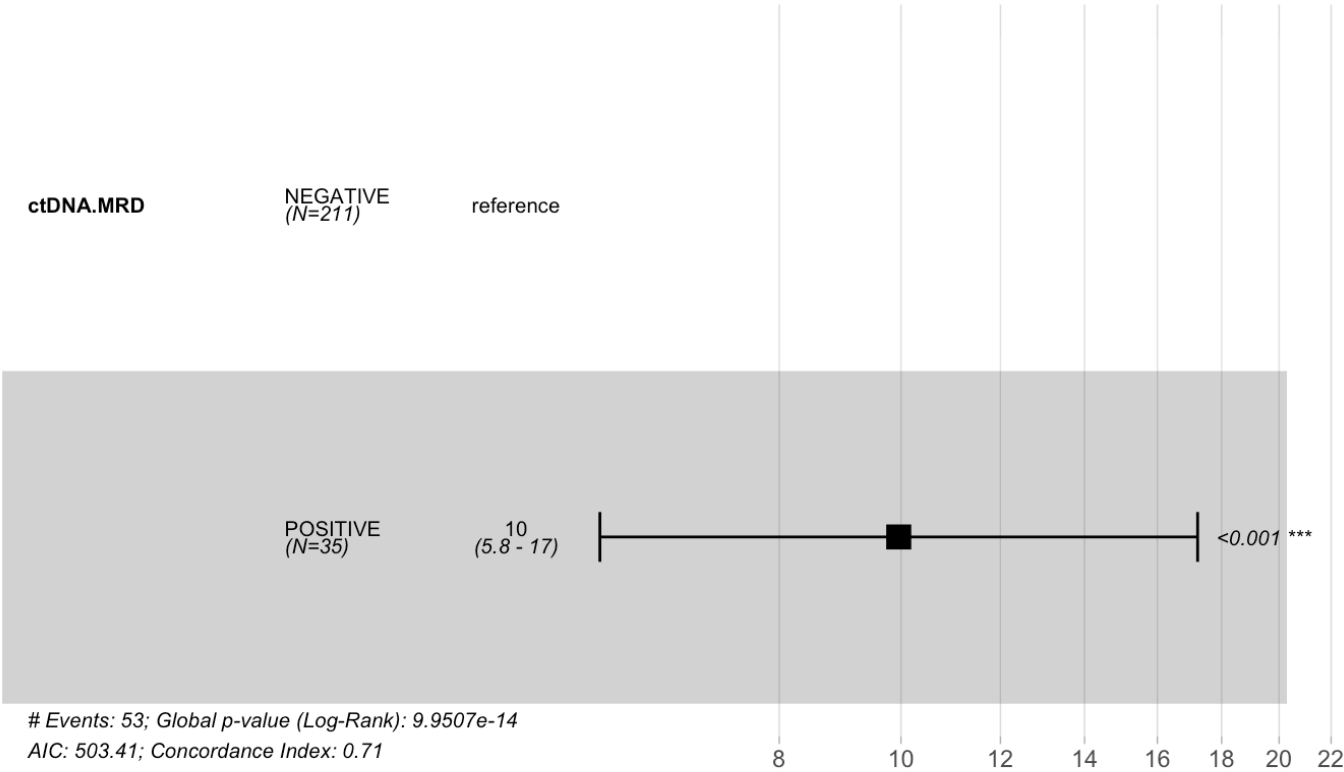
```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio

ctDNA.MRD	NEGATIVE (N=211)	reference
	POSITIVE (N=35)	10 (5.8 - 17) <0.001 ***

 # Events: 53; Global p-value (Log-Rank): 9.9507e-14
AIC: 503.41; Concordance Index: 0.71

(x-axis: 8 10 12 14 16 18 20 22)

Hide

```
summary(cox_fit)
```

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

  n= 246, number of events= 53

                   coef exp(coef) se(coef)     z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.2985   9.9592   0.2795 8.222   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                 exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE     9.959     0.1004     5.758     17.23

Concordance= 0.709  (se = 0.032 )
Likelihood ratio test= 55.38  on 1 df,    p=1e-13
Wald test            = 67.61  on 1 df,    p=<2e-16
Score (logrank) test = 101.2  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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.96 (5.76-17.23); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 12, 18, and 24 months
dfs_times <- c(12, 18, 24)
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.MRD == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.MRD == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 18 months p-value at 24 months
        3.289740e-11         6.618049e-13         2.261958e-12
```
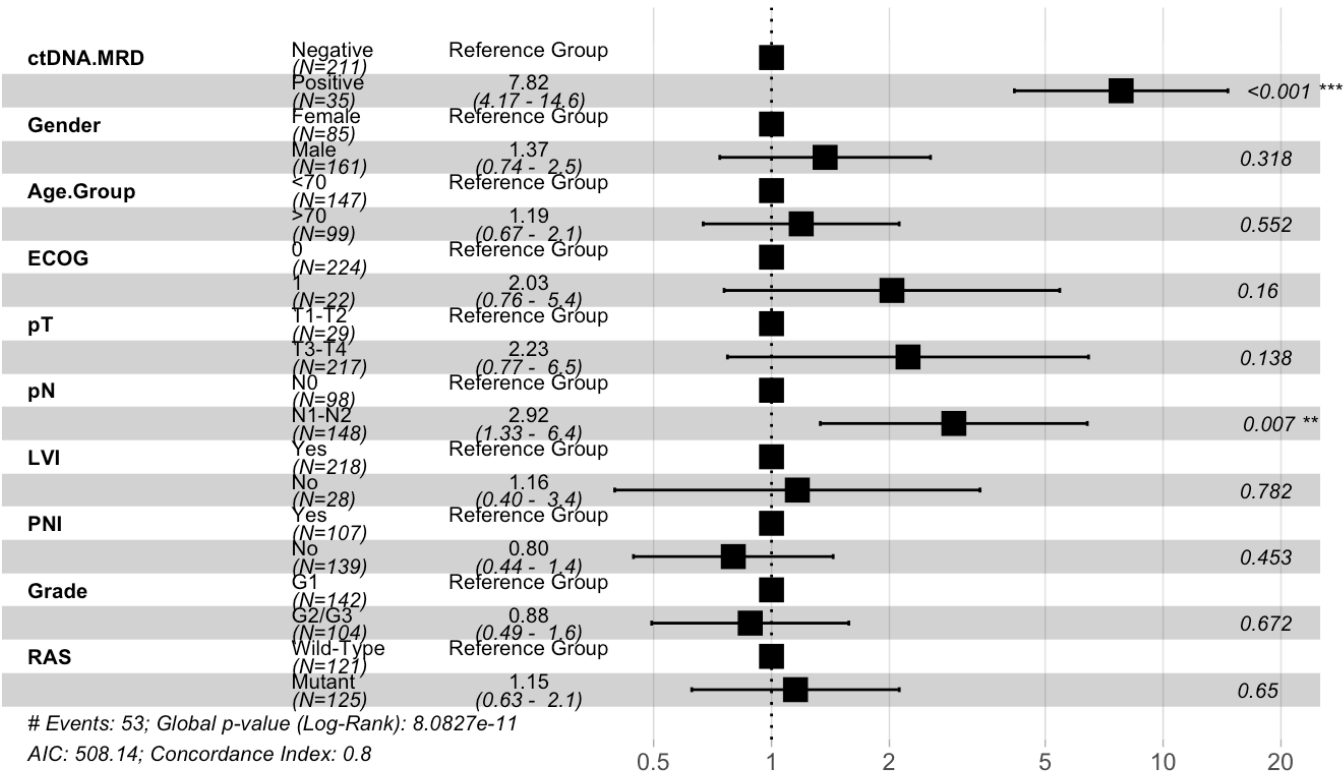
#Multivariate cox regression at MRD Window for DFS - All stages Landmark MRD timepoint

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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"), labe
ls = 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("<7
0", ">70"))
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$LVI <- factor(circ_data$LVI, levels = c("TRUE", "FALSE"), labels = c("Yes", "N
o"))
circ_data$PNI <- factor(circ_data$PNI, levels = c("TRUE", "FALSE"), labels = c("Yes", "N
o"))
circ_data$Grade <- factor(circ_data$Grade, levels = c("G1", "G2/G3"))
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 + ECOG + pT + pN + LVI + P
NI + Grade + 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

ctDNA.MRD	Negative (N=211)	Reference Group		
	Positive (N=35)	7.82 (4.17 - 14.6)		<0.001 ***
Gender	Female (N=85)	Reference Group		
	Male (N=161)	1.37 (0.74 - 2.5)		0.318
Age.Group	<70 (N=147)	Reference Group		
	>70 (N=99)	1.19 (0.67 - 2.1)		0.552
ECOG	0 (N=224)	Reference Group		
	1 (N=22)	2.03 (0.76 - 5.4)		0.16
pT	T1-T2 (N=29)	Reference Group		
	T3-T4 (N=217)	2.23 (0.77 - 6.5)		0.138
pN	N0 (N=98)	Reference Group		
	N1-N2 (N=148)	2.92 (1.33 - 6.4)		0.007 **
LVI	Yes (N=218)	Reference Group		
	No (N=28)	1.16 (0.40 - 3.4)		0.782
PNI	Yes (N=107)	Reference Group		
	No (N=139)	0.80 (0.44 - 1.4)		0.453
Grade	G1 (N=142)	Reference Group		
	G2/G3 (N=104)	0.88 (0.49 - 1.6)		0.672
RAS	Wild-Type (N=121)	Reference Group		
	Mutant (N=125)	1.15 (0.63 - 2.1)		0.65

 # Events: 53; Global p-value (Log-Rank): 8.0827e-11

AIC: 508.14; Concordance Index: 0.8

0.5   1   2   5   10   20

Hide

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

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("III")),]
circ_data <- circ_data[circ_data$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, da
ta = circ_data)
```

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

                   n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 96      6     NA      NA      NA
ctDNA.MRD=POSITIVE  3      3     13   0.493      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	96	6	0.0625	6.25
POSITIVE	3	3	1.0000	100.00

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="l
og-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, 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.lab
s=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage II

Kaplan-Meier curve with legend: ctDNA Negative (blue), ctDNA Positive (red). Y-axis: Disease-Free Survival. X-axis: Time from Landmark Time point (Months).

Number at risk

	0	6	12	18	24	30
ctDNA Negative	96	94	73	51	13	1
ctDNA Positive	3	2	2	0	0	0

Time from Landmark Time point (Months)

Hide

```
summary(KM_curve, times= c(12, 24))
```
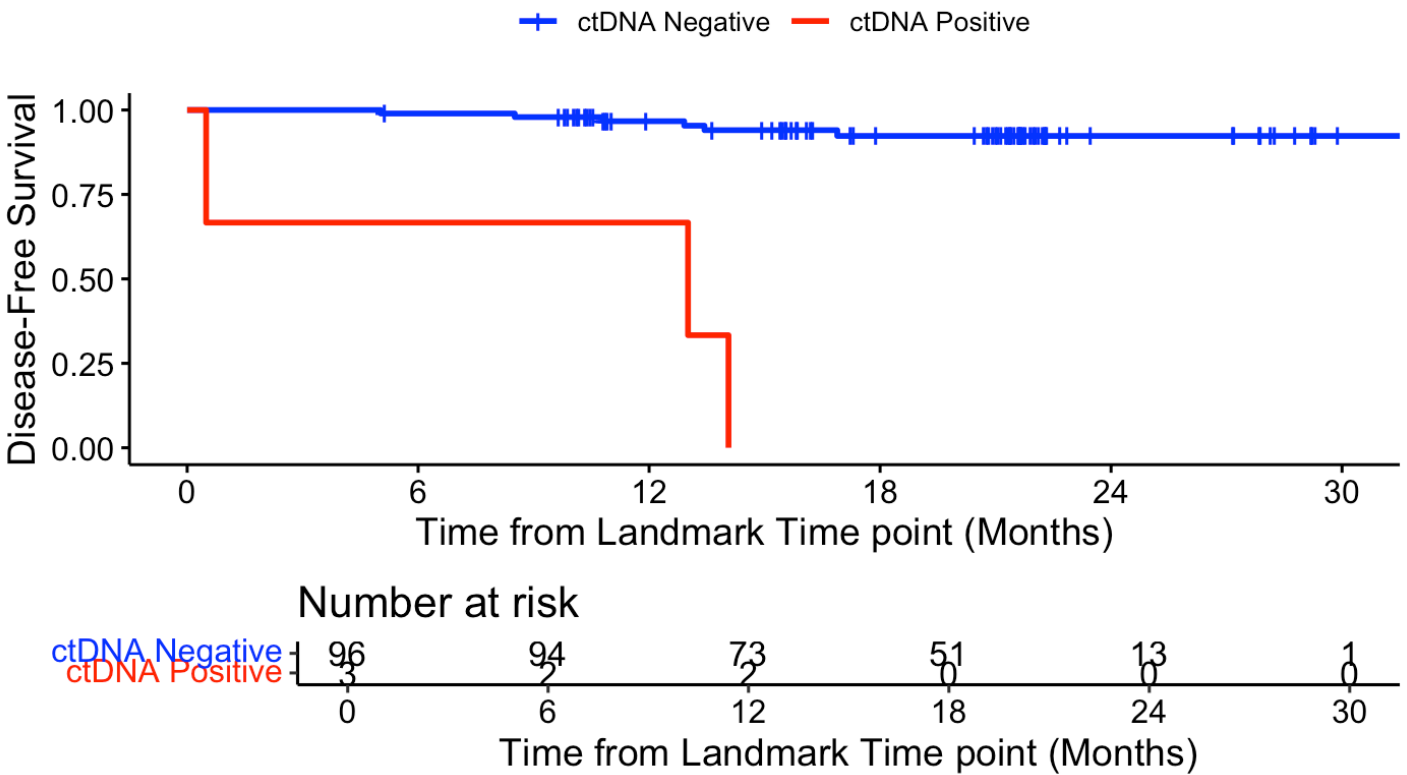
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
    conf.int = 0.95, conf.type = "log-log")

                ctDNA.MRD=NEGATIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     73       3    0.967  0.0189        0.900        0.989
   24     13       3    0.923  0.0307        0.835        0.965

                ctDNA.MRD=POSITIVE
       time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI
    12.0000      2.0000       1.0000        0.6667       0.2722       0.0541       0.94
52
```

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

ctDNA.MRD   NEGATIVE   reference
             (N=96)

             POSITIVE   27
             (N=3)   (6.4 - 115)   ■   <0.001 ***

 # Events: 9; Global p-value (Log-Rank): 0.00039357
AIC: 68.1; Concordance Index: 0.66

                10   20   50   100   200

Hide

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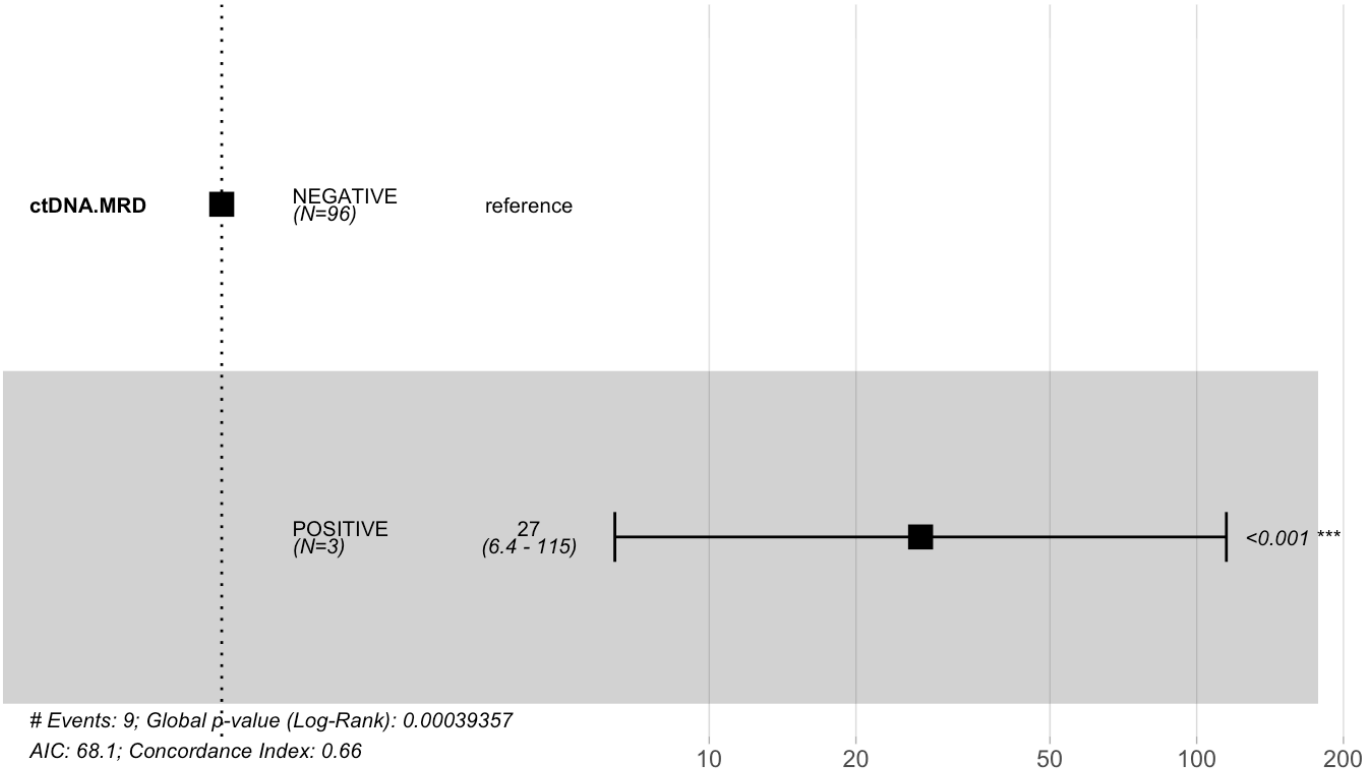
```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

  n= 99, number of events= 9

                    coef exp(coef) se(coef)    z Pr(>|z|)
ctDNA.MRDPOSITIVE  3.3011   27.1423   0.7369 4.48 7.46e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                  exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE     27.14    0.03684     6.404       115

Concordance= 0.66  (se = 0.077 )
Likelihood ratio test= 12.56  on 1 df,   p=4e-04
Wald test            = 20.07  on 1 df,   p=7e-06
Score (logrank) test = 45.99  on 1 df,   p=1e-11
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 27.14 (6.4-115.04); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 12 and 24 months
dfs_times <- c(12, 24)
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.MRD == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.MRD == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
        0.1175270483         0.0005355469
```

#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$Stage %in% c("II")),]
circ_data <- circ_data[circ_data$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, da
ta = circ_data)
```

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

  1 observation deleted due to missingness
                     n events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 1189    203     NA      NA      NA
ctDNA.MRD=POSITIVE  291    233   5.06     4.6    6.51
```

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	1189	203	0.1707317	17.07317
POSITIVE	291	233	0.8006873	80.06873
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="l
og-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, 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
s=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage III
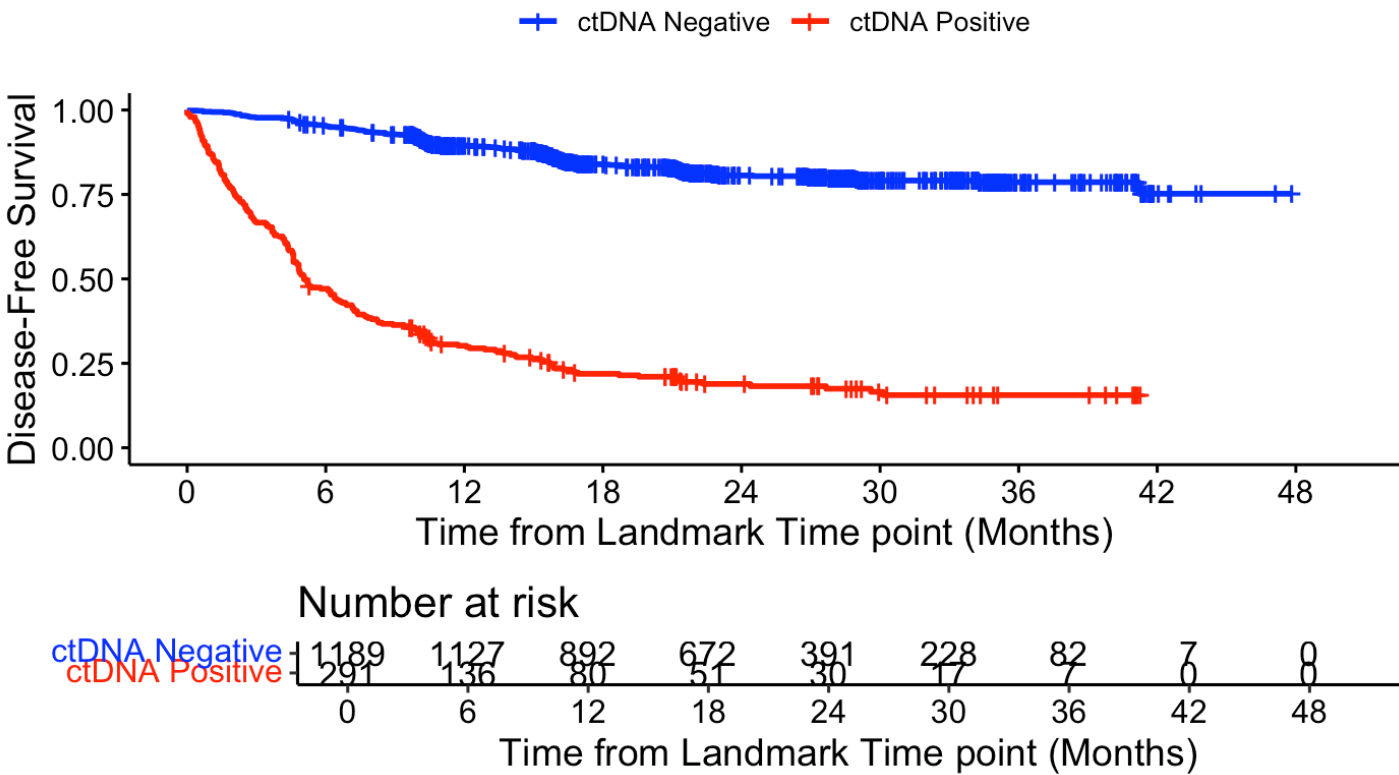

```
summary(KM_curve, times= c(24, 30))
```

```
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    391     195    0.806  0.0130        0.779        0.830
   30    228       6    0.791  0.0141        0.762        0.817


                ctDNA.MRD=POSITIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   24     30     229    0.189  0.0245        0.144        0.239
   30     17       3    0.166  0.0250        0.120        0.218
```

```
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

ctDNA.MRD	NEGATIVE (N=1189)	reference	
	POSITIVE (N=291)	9.4 (7.8 - 11)	<0.001 ***

 # Events: 436; Global p-value (Log-Rank): 9.1856e-104
AIC: 5644.24; Concordance Index: 0.73

11   12

Hide

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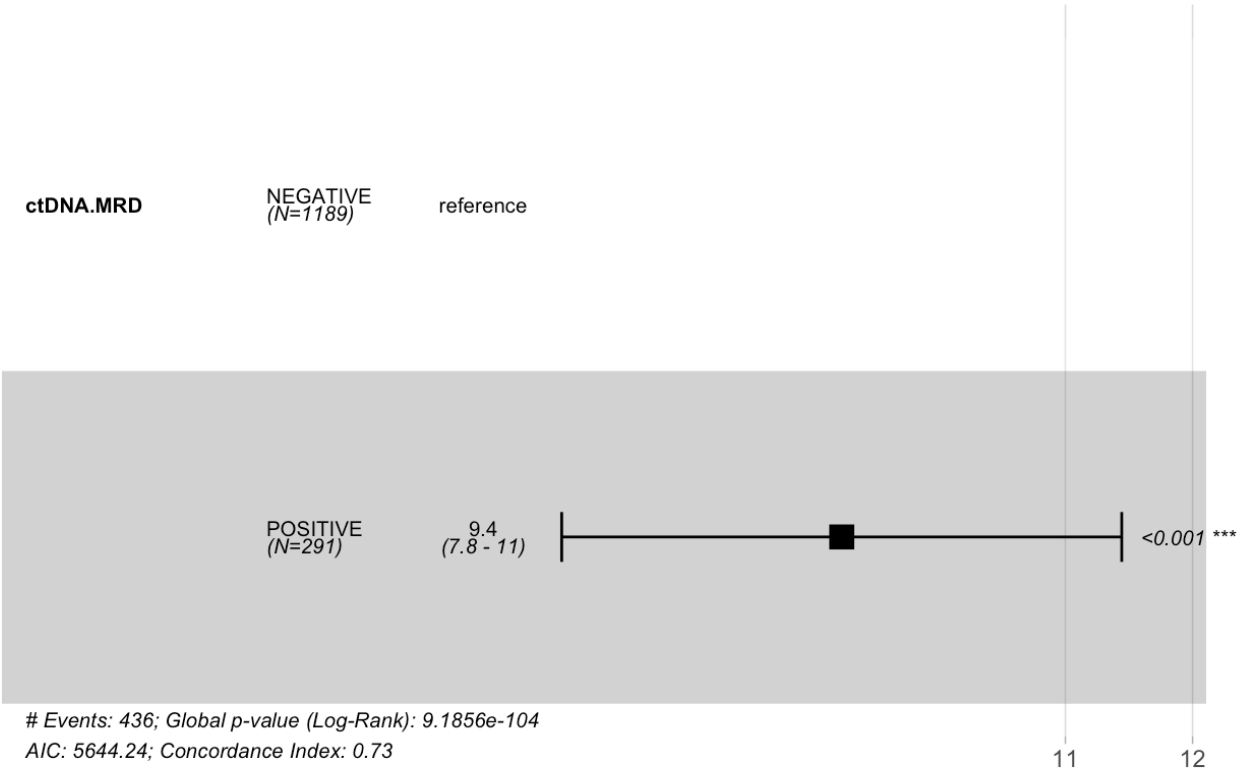
```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

  n= 1480, number of events= 436
   (1 observation deleted due to missingness)

                    coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.24488   9.43924  0.09775 22.97   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                  exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE     9.439     0.1059     7.794     11.43

Concordance= 0.729  (se = 0.011 )
Likelihood ratio test= 467.9  on 1 df,   p=<2e-16
Wald test            = 527.5  on 1 df,   p=<2e-16
Score (logrank) test = 768.2  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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.44 (7.79-11.43); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 24 and 30 months
dfs_times <- c(24, 30)
circ_data <- na.omit(circ_data[, c("ctDNA.MRD", "DFS.MRD.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.MRD.months >= time
& circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.MRD == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.MRD == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.MRD == "NEGATIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.MRD == "POSITIVE" & circ_data$DFS.Event ==
1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 24 months p-value at 30 months
       6.506858e-91         5.794183e-92
```

#DFS by ACT treatment in MRD negative - 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
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 115     16     NA      NA      NA
ACT=TRUE   96     12     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	115	16	0.1391304	13.91304
TRUE	96	12	0.1250000	12.50000

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-lo
g")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Obse
rvation | Stage II/III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time po
int (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

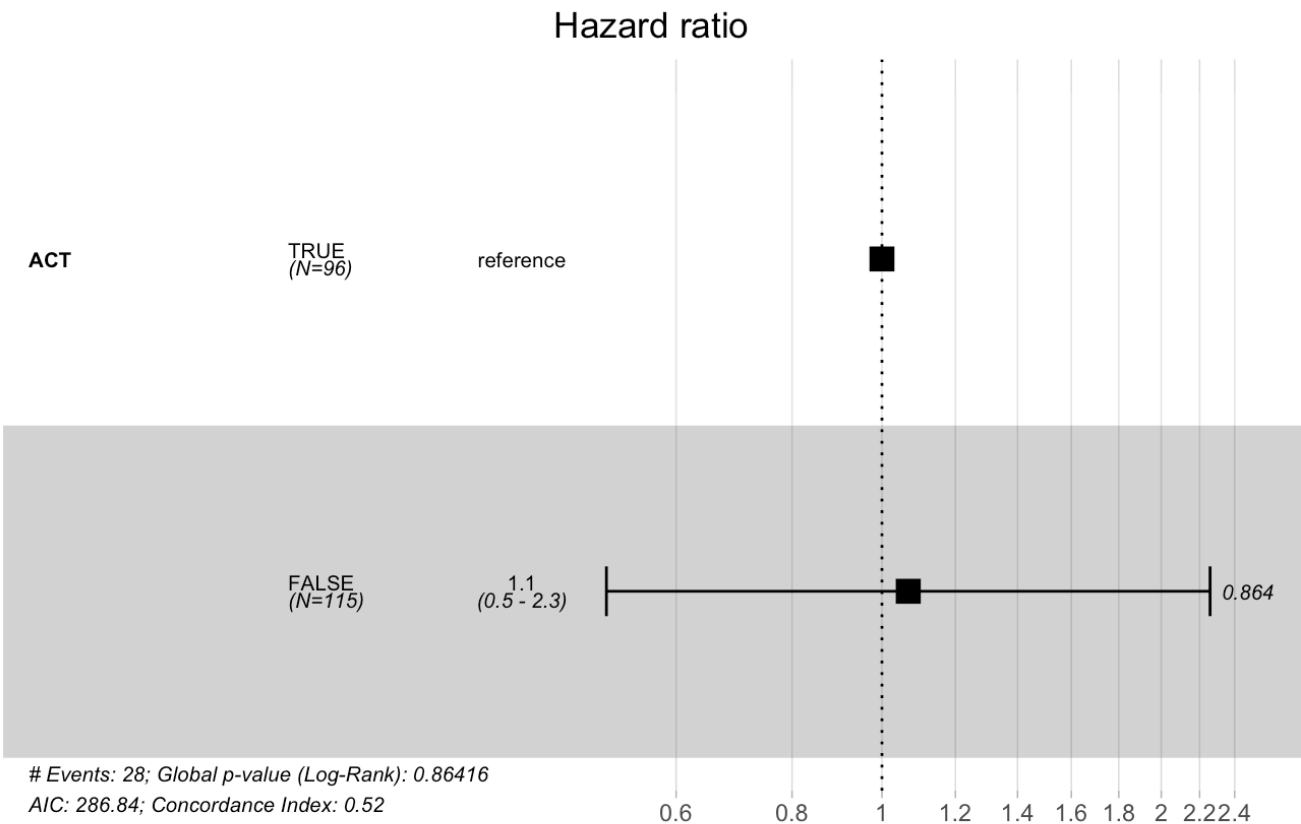# DFS - ctDNA MRD Negative ACT vs Observation | Stage II/III

Number at risk

Observation	115	110	82	58	16	2
ACT	96	94	62	38	15	1

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

Hide

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

                ACT=FALSE
      time        n.risk        n.event      survival       std.err lower 95% CI upper 95%
CI
    24.0000       16.0000       16.0000        0.8419        0.0373        0.7521        0.90
13

                ACT=TRUE
      time        n.risk        n.event      survival       std.err lower 95% CI upper 95%
CI
    24.0000       15.0000       11.0000        0.8612        0.0407        0.7574        0.92
28
```

[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

ACT   TRUE (N=96)   reference

FALSE (N=115)   1.1 (0.5 - 2.3)   0.864

Events: 28; Global p-value (Log-Rank): 0.86416
AIC: 286.84; Concordance Index: 0.52

0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4

[Hide]

```
summary(cox_fit)
```

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

  n= 211, number of events= 28

            coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE 0.06529   1.06747  0.38220 0.171    0.864

         exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     1.067     0.9368    0.5047     2.258

Concordance= 0.517  (se = 0.048 )
Likelihood ratio test= 0.03  on 1 df,    p=0.9
Wald test            = 0.03  on 1 df,    p=0.9
Score (logrank) test = 0.03  on 1 df,    p=0.9
```

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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.07 (0.5-2.26); p = 0.864"
```

Hide

```
#Fisher test for DFS percentages at 24
dfs_times <- c(24)
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ACT == "TRUE" & circ_data$DFS.months >= time & circ_data$DF
S.Event == 0)
  pos_count <- sum(circ_data$ACT == "FALSE" & circ_data$DFS.months >= time & circ_data$D
FS.Event == 0)
  neg_total <- sum(circ_data$ACT == "TRUE")
  pos_total <- sum(circ_data$ACT == "FALSE")

  neg_surv <- neg_total - sum(circ_data$ACT == "TRUE" & circ_data$DFS.Event == 1 & circ_
data$DFS.months < time)
  pos_surv <- pos_total - sum(circ_data$ACT == "FALSE" & circ_data$DFS.Event == 1 & circ
_data$DFS.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 24 months
          0.6811626
```

Hide

```
#Adjusted HR "ACT vs no ACT" - 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
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("<7
0", "≥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$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
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= 211, number of events= 28


               coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE     -0.5269    0.5904   0.4211 -1.251  0.21080
GenderMale   0.1145    1.1213   0.4081  0.280  0.77910
Age.Group≥70 0.3280    1.3883   0.4022  0.816  0.41472
StageIII     1.5211    4.5774   0.4886  3.113  0.00185 **
ECOG1        0.5215    1.6846   0.5509  0.947  0.34383
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


            exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE        0.5904     1.6937    0.2587     1.348
GenderMale     1.1213     0.8918    0.5039     2.495
Age.Group≥70   1.3883     0.7203    0.6311     3.054
StageIII       4.5774     0.2185    1.7569    11.926
ECOG1          1.6846     0.5936    0.5722     4.960


Concordance= 0.71  (se = 0.043 )
Likelihood ratio test= 12.82  on 5 df,   p=0.03
Wald test            = 11.73  on 5 df,   p=0.04
Score (logrank) test = 12.73  on 5 df,   p=0.03
```

#DFS by ACT treatment in MRD positive - 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
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  6      5   5.62    1.22      NA
ACT=TRUE  25     16   9.33    6.97      NA
```
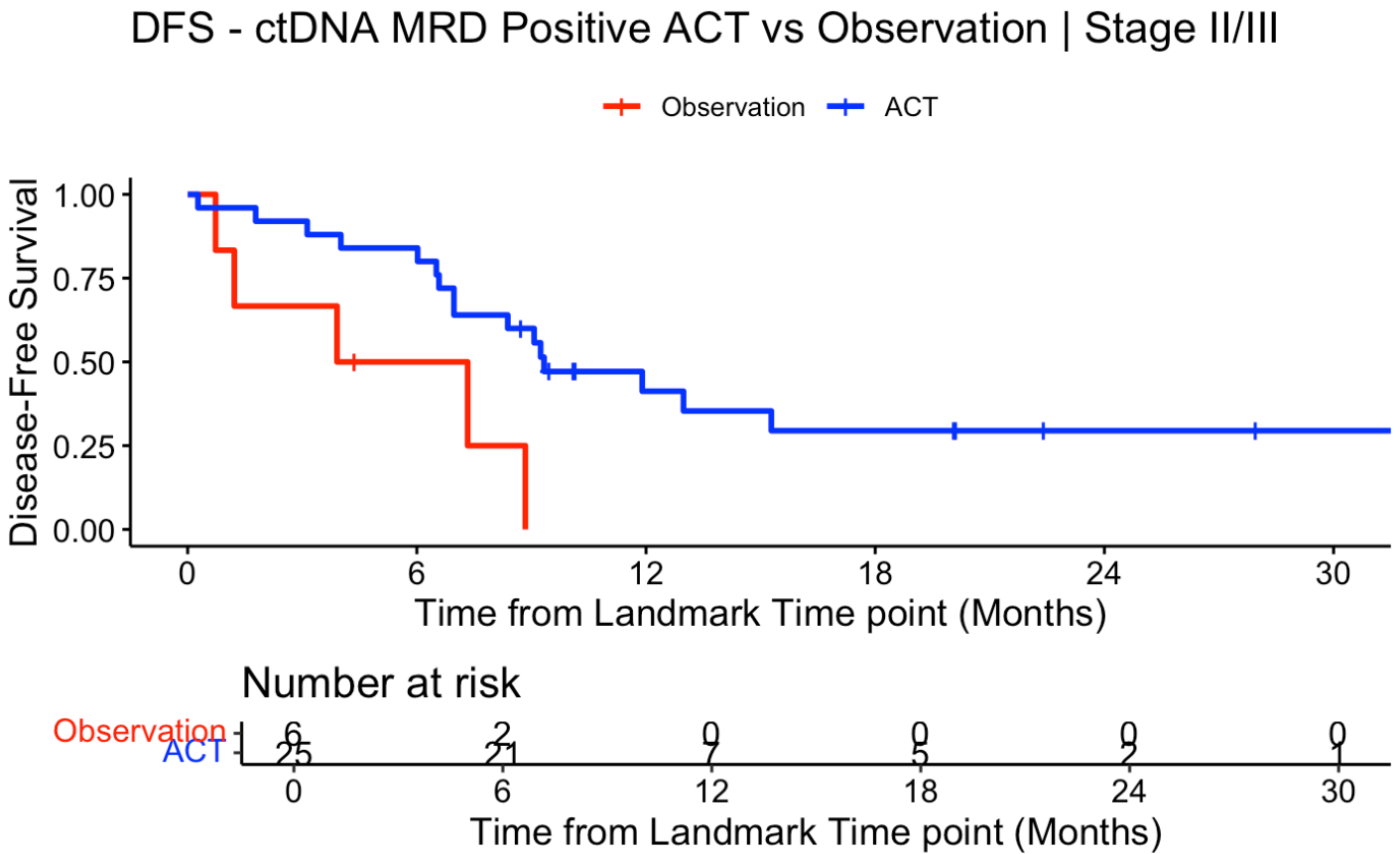
Hide

```
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	6	5	0.8333333	83.33333
TRUE	25	16	0.6400000	64.00000

2 rows

Hide

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

DFS - ctDNA MRD Positive ACT vs Observation | Stage II/III


```
summary(KM_curve, times= c(6, 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
       6.000        2.000        3.000         0.500        0.204        0.111         0.8
04


                 ACT=TRUE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    6     21       4    0.840  0.0733        0.628        0.937
   24      2      12    0.295  0.1026        0.118        0.497
```

[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)
```


[Hide]

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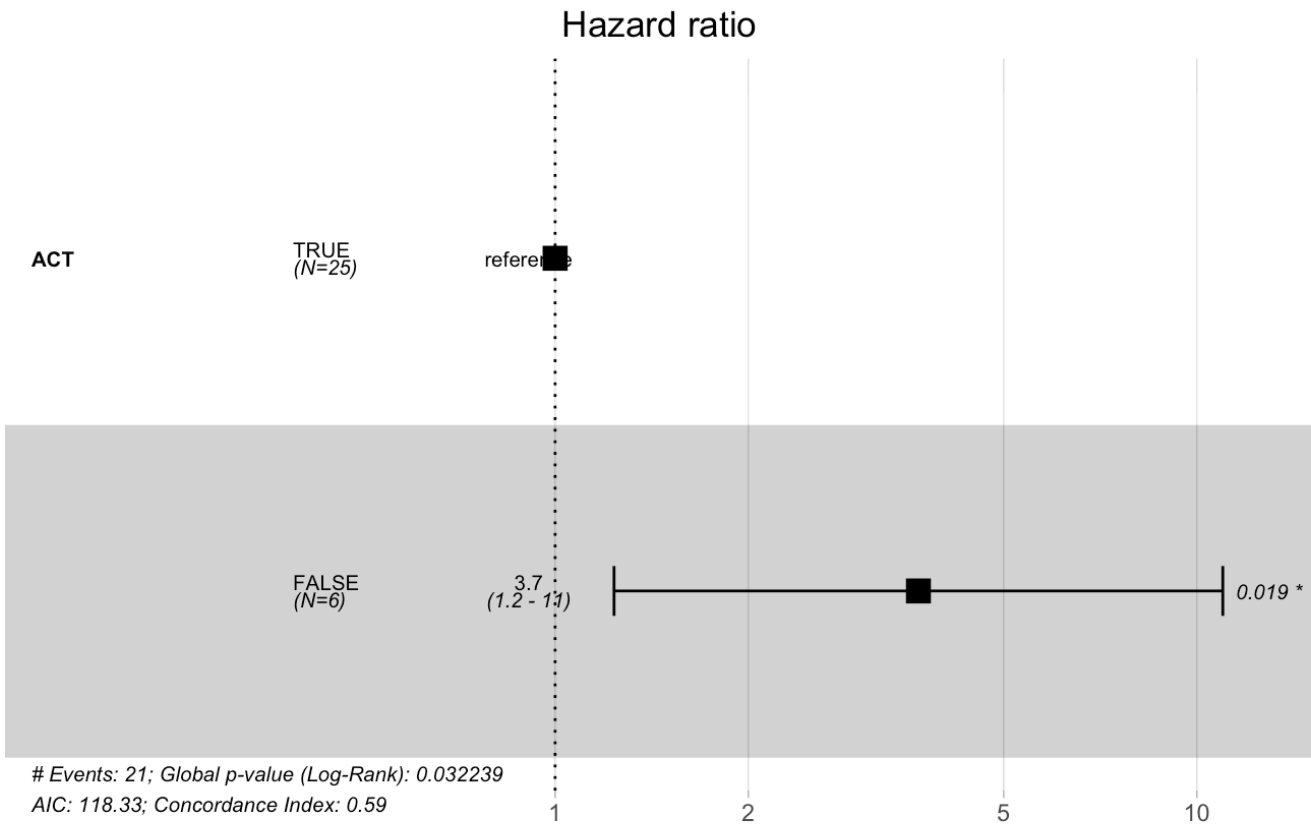
```
Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

  n= 31, number of events= 21

           coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE 1.3036    3.6825   0.5572 2.339   0.0193 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

         exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     3.682     0.2716     1.235     10.98

Concordance= 0.593  (se = 0.045 )
Likelihood ratio test= 4.59  on 1 df,   p=0.03
Wald test            = 5.47  on 1 df,   p=0.02
Score (logrank) test = 6.25  on 1 df,   p=0.01
```

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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 3.68 (1.24-10.98); p = 0.019"
```

Hide

```
#Adjusted HR "ACT vs no ACT" – 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ACT <- factor(circ_data$ACT, levels=c("FALSE","TRUE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<7
0", "≥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$ECOG <- factor(circ_data$ECOG, levels=c("0","1"))
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= 31, number of events= 21


              coef exp(coef) se(coef)      z Pr(>|z|)
ACTTRUE    -1.27835   0.27850  0.59142 -2.162   0.0307 *
GenderMale  0.35495   1.42611  0.51244  0.693   0.4885
Age.Group≥70 0.12911  1.13782  0.45951  0.281   0.7787
StageIII    0.08804   1.09203  0.81274  0.108   0.9137
ECOG1       1.40597   4.07950  1.13540  1.238   0.2156
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


           exp(coef) exp(-coef) lower .95 upper .95
ACTTRUE       0.2785     3.5907   0.08738    0.8876
GenderMale    1.4261     0.7012   0.52236    3.8935
Age.Group≥70  1.1378     0.8789   0.46231    2.8003
StageIII      1.0920     0.9157   0.22204    5.3708
ECOG1         4.0795     0.2451   0.44071   37.7622


Concordance= 0.632  (se = 0.066 )
Likelihood ratio test= 6.01  on 5 df,   p=0.3
Wald test             = 6.81  on 5 df,   p=0.2
Score (logrank) test = 7.76  on 5 df,   p=0.2
```

#DFS by ctDNA at 3 months - All stages Landmark 3 months timepoint

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.3months!="",]
circ_data <- circ_data[circ_data$DFS.3mo.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Event)~ctDNA.3month
s, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Even
t) ~
    ctDNA.3months, data = circ_data)

                          n events median 0.95LCL 0.95UCL
ctDNA.3months=NEGATIVE 204     31     NA      NA      NA
ctDNA.3months=POSITIVE  19     14   8.28    5.59      NA
```

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.3months) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.3months	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	204	31	0.1519608	15.19608
POSITIVE	19	14	0.7368421	73.68421

2 rows

Hide

```
surv_object <-Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.3months, data = circ_data,conf.int=0.95,conf.typ
e="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA 3 months | All stages",
ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.lab
s=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA 3 months | All stages
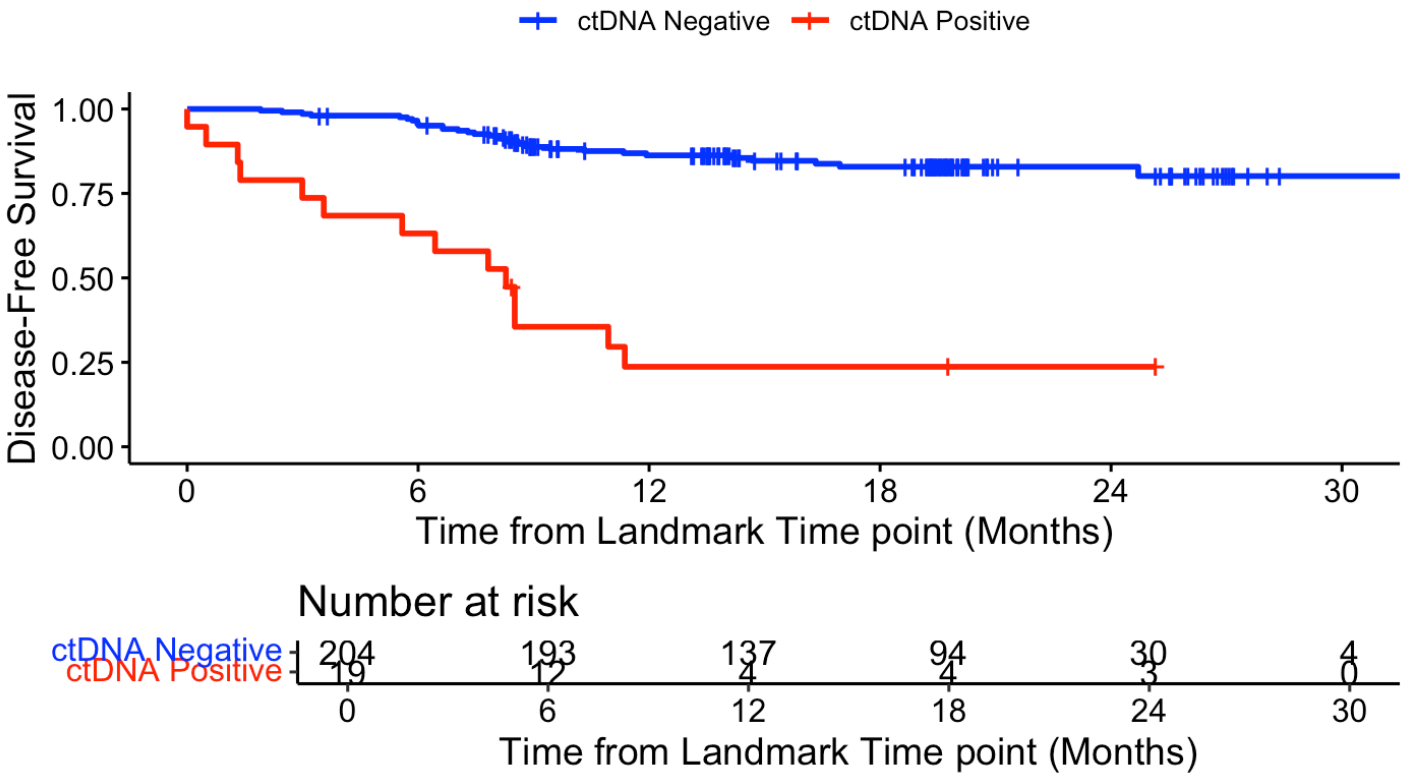

```
summary(KM_curve, times= c(12, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.3months, data = circ_data,
    conf.int = 0.95, conf.type = "log-log")

                ctDNA.3months=NEGATIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    137      26    0.863  0.0252        0.805        0.905
   24     30       4    0.829  0.0293        0.762        0.879

                ctDNA.3months=POSITIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      4      14    0.237   0.101       0.0758        0.447
   24      3       0    0.237   0.101       0.0758        0.447
```
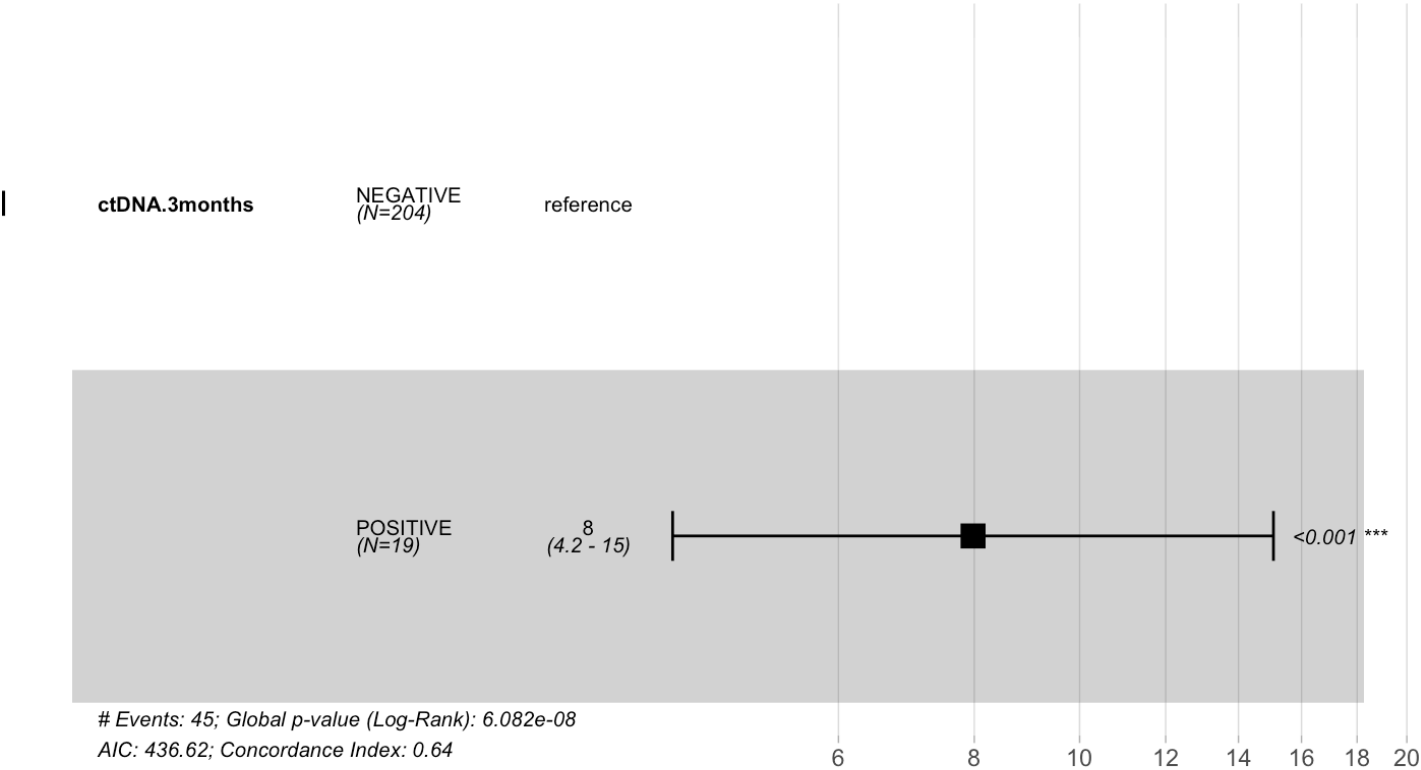
Hide

```
circ_data$ctDNA.3months <- factor(circ_data$ctDNA.3months, levels=c("NEGATIVE","POSITIV
E"))
cox_fit <- coxph(surv_object ~ ctDNA.3months, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio

ctDNA.3months   NEGATIVE (N=204)   reference

POSITIVE (N=19)   8 (4.2 - 15)   <0.001 ***

Events: 45; Global p-value (Log-Rank): 6.082e-08
AIC: 436.62; Concordance Index: 0.64

6   8   10   12   14   16   18   20

Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.3months, data = circ_data)

  n= 223, number of events= 45

                        coef exp(coef) se(coef)     z Pr(>|z|)
ctDNA.3monthsPOSITIVE 2.0768    7.9792   0.3247 6.397 1.58e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                   exp(coef) exp(-coef) lower .95 upper .95
ctDNA.3monthsPOSITIVE    7.979     0.1253     4.223     15.08

Concordance= 0.641  (se = 0.034 )
Likelihood ratio test= 29.34  on 1 df,    p=6e-08
Wald test             = 40.92  on 1 df,    p=2e-10
Score (logrank) test = 57.62  on 1 df,    p=3e-14
```

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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 7.98 (4.22-15.08); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 12 and 24 months
dfs_times <- c(12, 24)
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.3months == "NEGATIVE" & circ_data$DFS.3mo.months >= t
ime & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.3months == "POSITIVE" & circ_data$DFS.3mo.months >= t
ime & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.3months == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.3months == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.3months == "NEGATIVE" & circ_data$DFS.Even
t == 1 & circ_data$DFS.3mo.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.3months == "POSITIVE" & circ_data$DFS.Even
t == 1 & circ_data$DFS.3mo.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
        2.546518e-08         1.138180e-07
```

#DFS by ctDNA at 6 months - All stages Landmark 6 months timepoint

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.6months!="",]
circ_data <- circ_data[circ_data$DFS.6mo.months>=0,]
circ_datadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Event)~ctDNA.6month
s, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Even
t) ~
    ctDNA.6months, data = circ_data)

                         n events median 0.95LCL 0.95UCL
ctDNA.6months=NEGATIVE 173     21     NA      NA      NA
ctDNA.6months=POSITIVE   9      7   3.68     2.6      NA
```

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.6months) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```
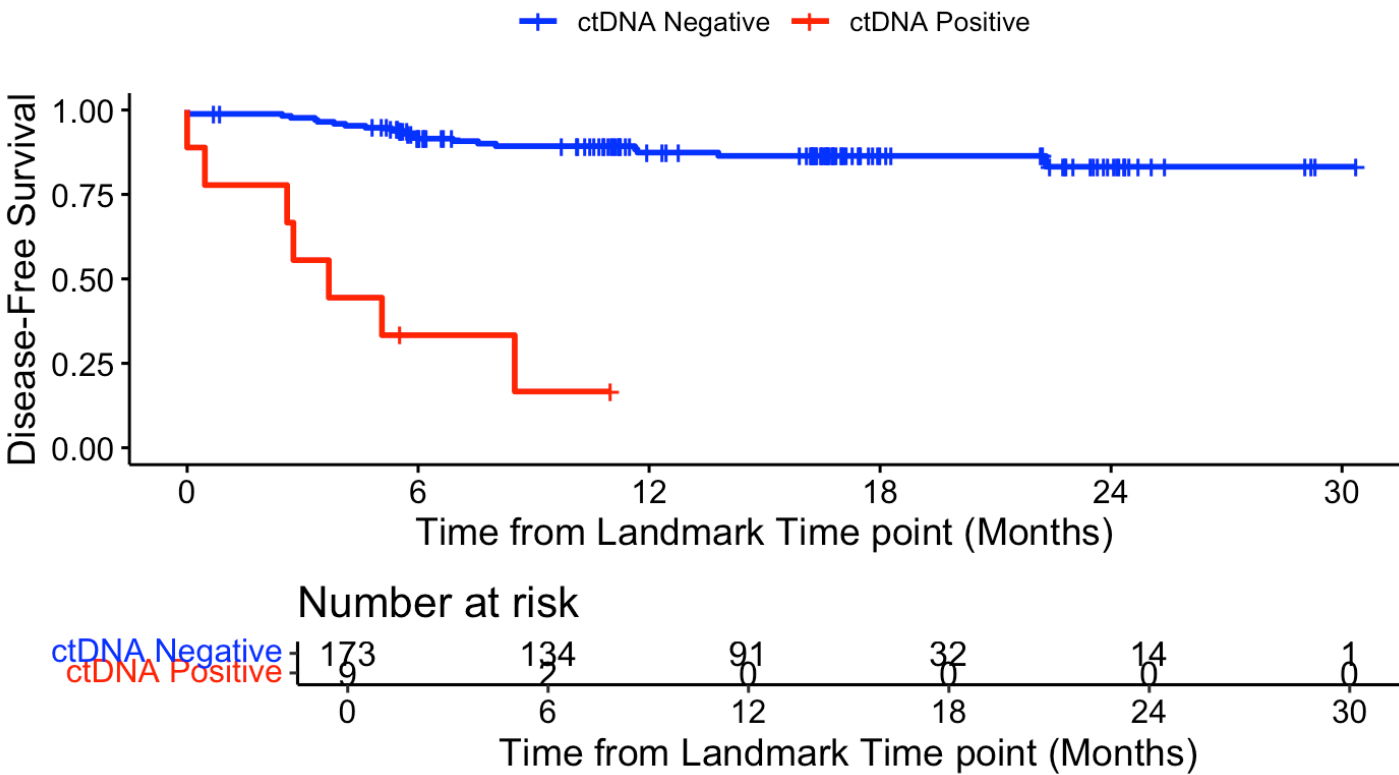
ctDNA.6months	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	173	21	0.1213873	12.13873
POSITIVE	9	7	0.7777778	77.77778

2 rows

Hide

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

DFS - ctDNA 6 months | All stages

ctDNA Negative + ctDNA Positive +

Number at risk

ctDNA Negative	173	134	91	32	14	1
ctDNA Positive	9	2	0	0	0	0
	0	6	12	18	24	30

Time from Landmark Time point (Months)

Hide

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

```
Call: survfit(formula = surv_object ~ ctDNA.6months, data = circ_data,
    conf.int = 0.95, conf.type = "log-log")

                ctDNA.6months=NEGATIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    6    134      14    0.915  0.0217        0.861        0.949
   24     14       7    0.832  0.0420        0.730        0.898

                ctDNA.6months=POSITIVE
      time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI
    6.0000      2.0000      6.0000        0.3333       0.1571       0.0783         0.62
26
```
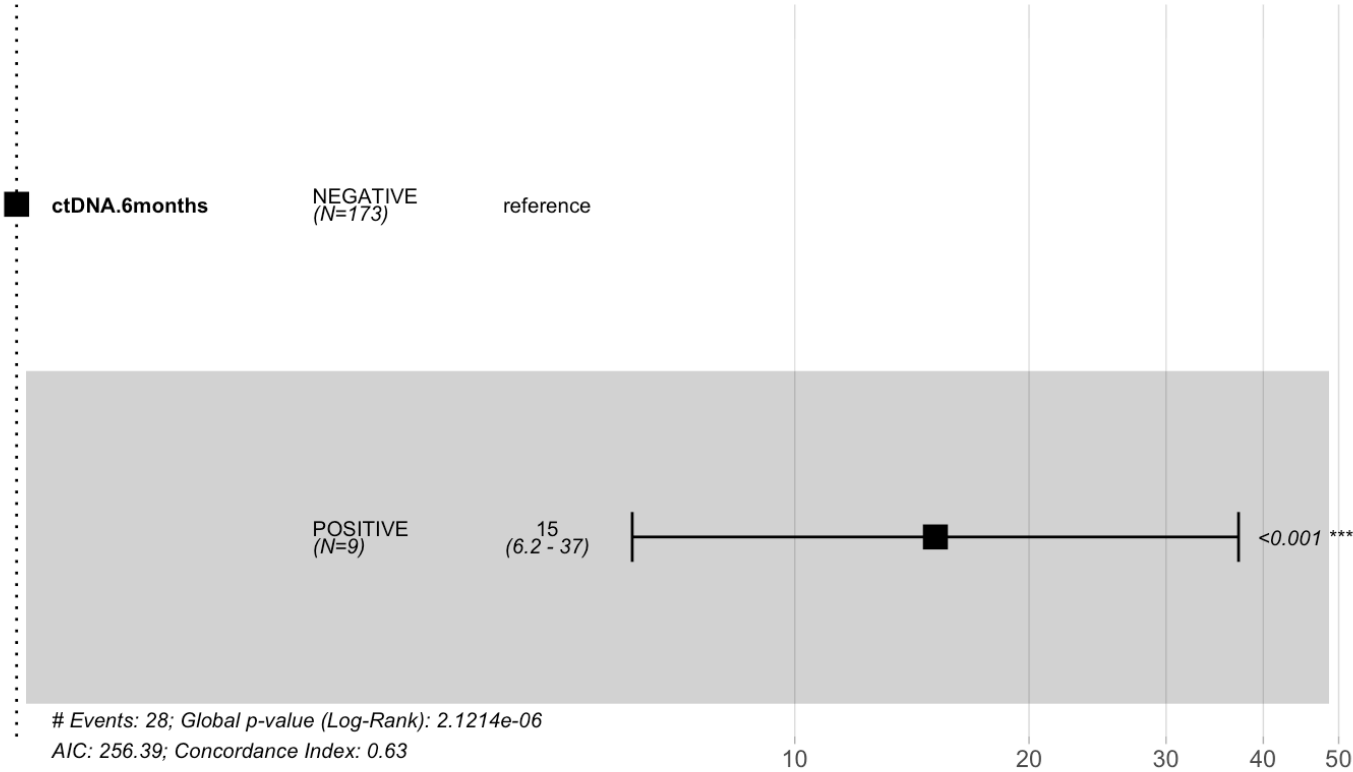
Hide

```
circ_data$ctDNA.6months <- factor(circ_data$ctDNA.6months, levels=c("NEGATIVE","POSITIV
E"))
cox_fit <- coxph(surv_object ~ ctDNA.6months, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio

ctDNA.6months	NEGATIVE (N=173)	reference
	POSITIVE (N=9)	15 (6.2 - 37) <0.001 ***

Events: 28; Global p-value (Log-Rank): 2.1214e-06
AIC: 256.39; Concordance Index: 0.63

10   20   30   40 50

Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.6months, data = circ_data)

  n= 182, number of events= 28

                          coef exp(coef) se(coef)    z Pr(>|z|)
ctDNA.6monthsPOSITIVE  2.7187   15.1598   0.4577 5.94 2.85e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.6monthsPOSITIVE     15.16    0.06596     6.182     37.18

Concordance= 0.628  (se = 0.042 )
Likelihood ratio test= 22.48  on 1 df,   p=2e-06
Wald test            = 35.29  on 1 df,   p=3e-09
Score (logrank) test = 62.3  on 1 df,    p=3e-15
```

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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 15.16 (6.18-37.18); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 6 and 24 months
dfs_times <- c(6, 24)
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.6months == "NEGATIVE" & circ_data$DFS.6mo.months >= t
ime & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.6months == "POSITIVE" & circ_data$DFS.6mo.months >= t
ime & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.6months == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.6months == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.6months == "NEGATIVE" & circ_data$DFS.Even
t == 1 & circ_data$DFS.6mo.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.6months == "POSITIVE" & circ_data$DFS.Even
t == 1 & circ_data$DFS.6mo.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
 p-value at 6 months p-value at 24 months
        5.667547e-05         2.922826e-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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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 combina
tion, 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.Dynamic
s, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.3mo.months, event = circ_data$DFS.Even
t) ~
    ctDNA.Dynamics, data = circ_data)

   97 observations deleted due to missingness
                  n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 16      8  14.16    6.01      NA
ctDNA.Dynamics=2  8      7   8.05    5.59      NA
```

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)
```
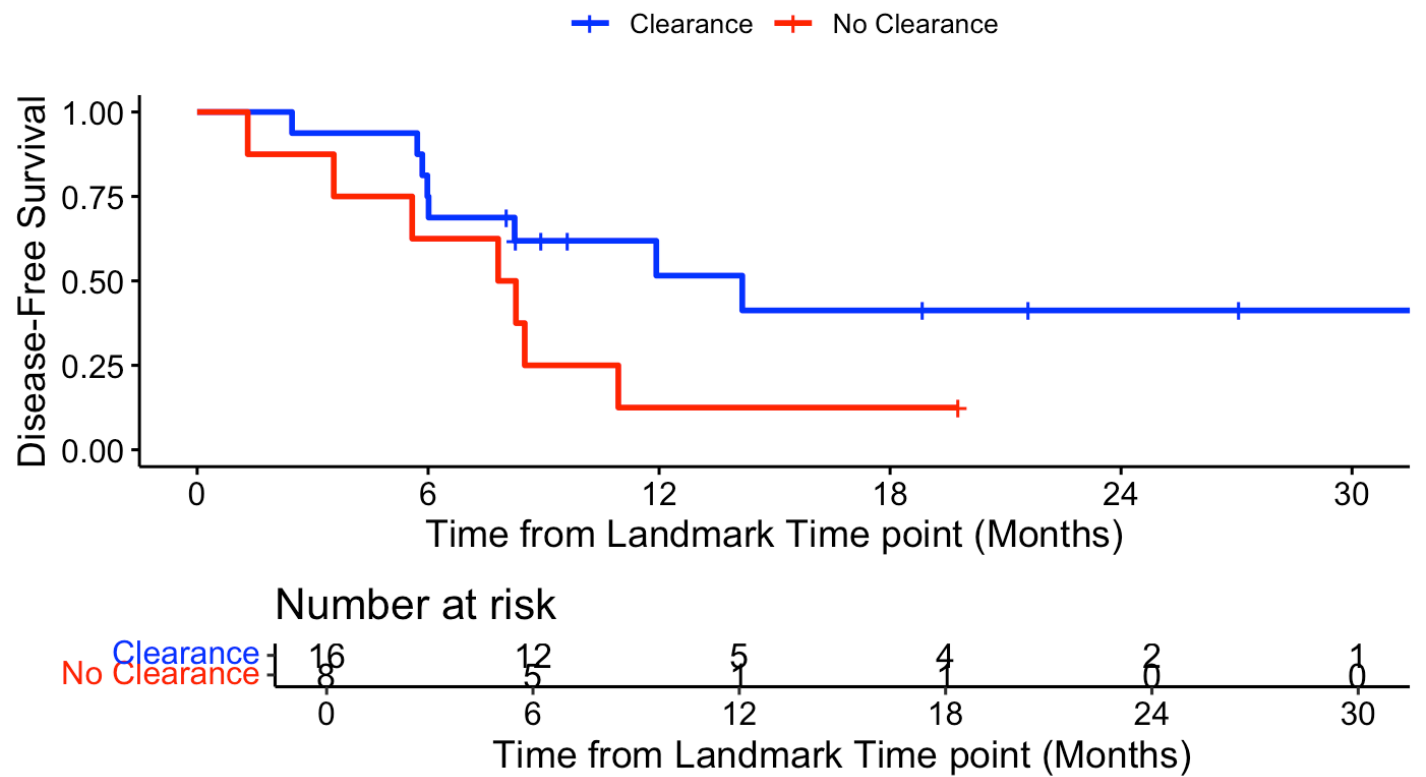
ctDNA.Dynamics	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	16	8	0.500	50.0
2	8	7	0.875	87.5
NA	97	NA	NA	NA

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.ty
pe="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, 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 Landmar
k Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title="")
```

DFS - ctDNA Clearance from MRD to 3 months ACT-treated | All St

Number at risk

Clearance	16	12	5	4	2	1
No Clearance	8	5	1	1	0	0

Time from Landmark Time point (Months)

Hide

```
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")

97 observations deleted due to missingness
              ctDNA.Dynamics=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      5       7    0.516   0.139        0.229        0.742
   24      2       1    0.413   0.145        0.146        0.665

              ctDNA.Dynamics=2
      time      n.risk      n.event      survival      std.err lower 95% CI upper 95%
CI
   12.00000     1.00000      7.00000       0.12500      0.11693      0.00659        0.422
71
```

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

ctDNA.Dynamics   Clearance   reference
                 (N=16)

                 No Clearance   2.4
                 (N=8)   (0.86 - 6.7)   0.096

 # Events: 15; Global p-value (Log-Rank): 0.10352
AIC: 79.96; Concordance Index: 0.6

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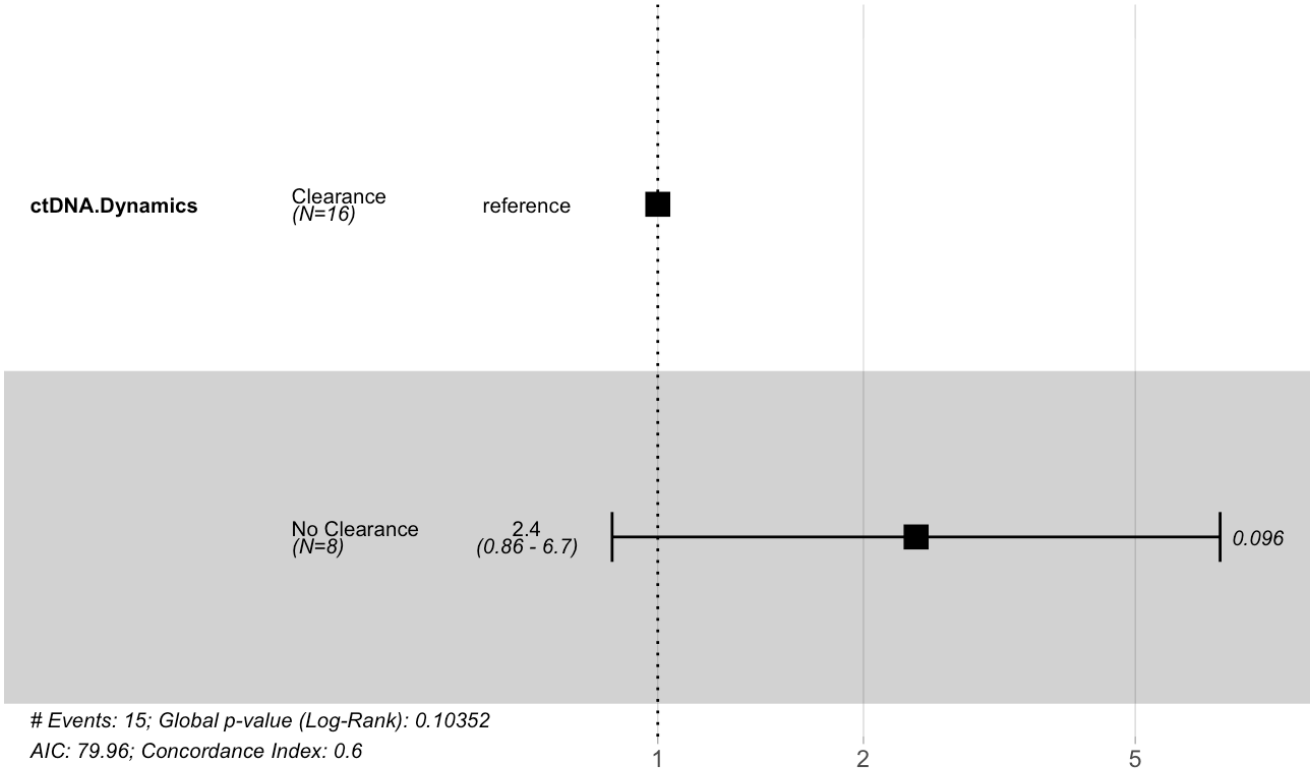
```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

  n= 24, number of events= 15
   (97 observations deleted due to missingness)


                             coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsNo Clearance 0.8707    2.3885   0.5227 1.666   0.0958 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                           exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     2.389     0.4187    0.8574     6.654

Concordance= 0.605  (se = 0.066 )
Likelihood ratio test= 2.65  on 1 df,    p=0.1
Wald test            = 2.77  on 1 df,    p=0.1
Score (logrank) test = 2.95  on 1 df,    p=0.09
```

```
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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 2.39 (0.86-6.65); p = 0.096"
```

Hide

```
#Fisher test for DFS percentages at 12 and 24 months
dfs_times <- c(12)
circ_data <- na.omit(circ_data[, c("ctDNA.Dynamics", "DFS.3mo.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.Dynamics == "Clearance" & circ_data$DFS.3mo.months >=
time & circ_data$DFS.Event == FALSE)
  pos_count <- sum(circ_data$ctDNA.Dynamics == "No Clearance" & circ_data$DFS.3mo.months
>= time & circ_data$DFS.Event == FALSE)
  neg_total <- sum(circ_data$ctDNA.Dynamics == "Clearance")
  pos_total <- sum(circ_data$ctDNA.Dynamics == "No Clearance")

  neg_surv <- neg_total - sum(circ_data$ctDNA.Dynamics == "Clearance" & circ_data$DFS.Ev
ent == TRUE & circ_data$DFS.3mo.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.Dynamics == "No Clearance" & circ_data$DF
S.Event == TRUE & circ_data$DFS.3mo.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months
          0.07907586
```

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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 combina
tion, 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.Dynamic
s, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Even
t) ~
    ctDNA.Dynamics, data = circ_data)

   106 observations deleted due to missingness
                 n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 13      5     NA    5.52      NA
ctDNA.Dynamics=2  2      2   2.68    2.60      NA
```

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)
```
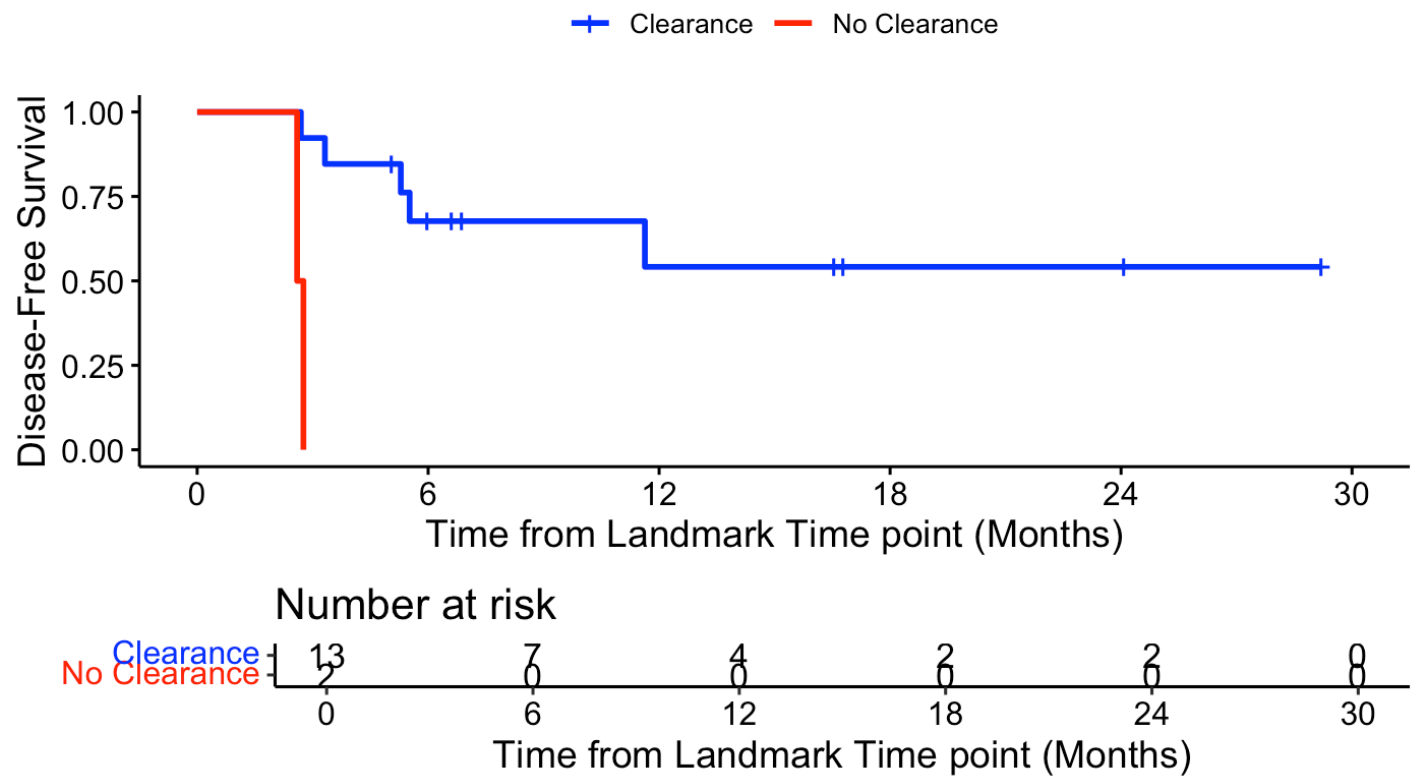
ctDNA.Dynamics	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	13	5	0.3846154	38.46154
2	2	2	1.0000000	100.00000
NA	106	NA	NA	NA

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.ty
pe="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, 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 Landmar
k Time point (Months)", legend.labs=c("Clearance", "No Clearance"), legend.title="")
```

DFS - ctDNA Clearance from MRD to 6 months ACT-treated | All St


```
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")

106 observations deleted due to missingness
               ctDNA.Dynamics=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      4       5    0.542   0.162        0.204        0.789
   24      2       0    0.542   0.162        0.204        0.789

               ctDNA.Dynamics=2
     time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
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

Events: 7; Global p-value (Log-Rank): 0.014782
AIC: 29.07; Concordance Index: 0.67

Hide

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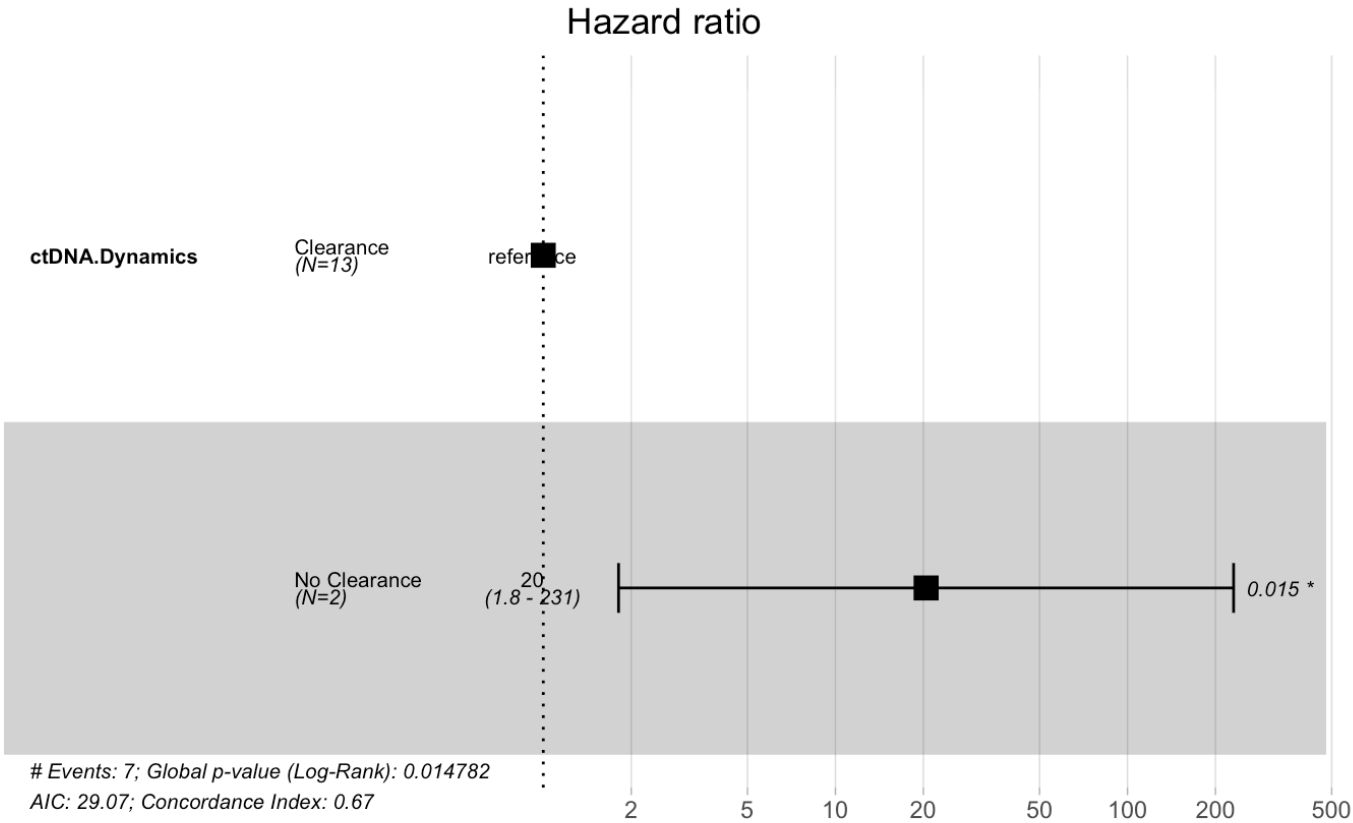
```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

  n= 15, number of events= 7
   (106 observations deleted due to missingness)


                             coef exp(coef) se(coef)     z Pr(>|z|)
ctDNA.DynamicsNo Clearance  3.017    20.438    1.237 2.44   0.0147 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                           exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsNo Clearance     20.44    0.04893      1.81     230.8

Concordance= 0.669  (se = 0.087 )
Likelihood ratio test= 5.94  on 1 df,   p=0.01
Wald test            = 5.95  on 1 df,   p=0.01
Score (logrank) test = 11.68  on 1 df,   p=6e-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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 20.44 (1.81-230.77); p = 0.015"
```

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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: 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
      9       5       7       1       4       0       0       0
```

Hide

```r
# Get the counts for each interval
interval_counts <- table(circ_data$p_drelReturned_intervals)

# Calculate the percentages
interval_percentages <- 100 * interval_counts / sum(interval_counts)

# Combine the counts and percentages for a clearer overview
interval_summary <- data.frame(Counts = interval_counts, Percentages = interval_percentages)

# Calculate the total number of observations
total_observations <- sum(interval_counts)

# Add the total number of observations to the summary
interval_summary$TotalObservations <- c(rep(NA, length(interval_counts)-1), total_observations)

# Print the summary with total observations
print(interval_summary)
```

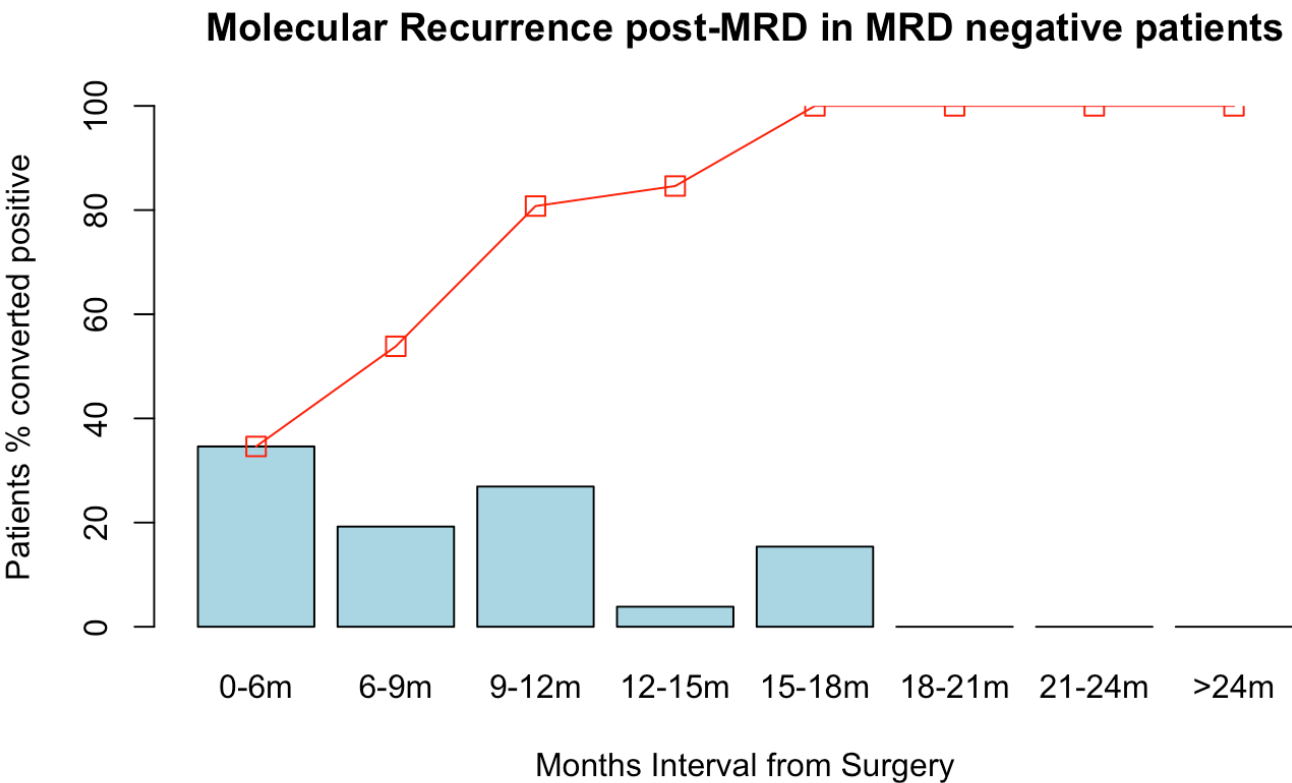| Counts.Var1 <fctr>	Counts.Freq <int>	Percentages.Var1 <fctr>	Percentages.Freq <dbl>	TotalObservations <int>
0-6m	9	0-6m	34.615385	NA
6-9m	5	6-9m	19.230769	NA
9-12m	7	9-12m	26.923077	NA
12-15m	1	12-15m	3.846154	NA
15-18m	4	15-18m	15.384615	NA
18-21m	0	18-21m	0.000000	NA
21-24m	0	21-24m	0.000000	NA
>24m	0	>24m	0.000000	26

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_percenta
ges, CumulativePercentages = cumulative_percentages, TotalObservations = c(rep(NA, lengt
h(interval_counts)-1), total_observations))

bp <- barplot(interval_percentages,
              main="Molecular Recurrence post-MRD in MRD negative patients",
              xlab="Months Interval from Surgery",
              ylab="Patients % converted positive",
              col="lightblue",
              ylim=c(0, 100))

# Add the cumulative percentages to the plot
points(bp, cumulative_percentages, type="o", pch=22, col="red", cex=1.5)
```


Molecular Recurrence post-MRD in MRD negative patients

Hide

```
print(interval_summary)
```

	Counts.Var1	Counts.Freq	Percentages.Var1	Percentages.Freq	CumulativePercenta
	<fctr>	<int>	<fctr>	<dbl>	<d
0-6m	0-6m	9	0-6m	34.615385	34.61

	Counts.Var1	Counts.Freq	Percentages.Var1	Percentages.Freq	CumulativePercenta
	<fctr>	<int>	<fctr>	<dbl>	<c
6-9m	6-9m	5	6-9m	19.230769	53.84
9-12m	9-12m	7	9-12m	26.923077	80.76!
12-15m	12-15m	1	12-15m	3.846154	84.61
15-18m	15-18m	4	15-18m	15.384615	100.00
18-21m	18-21m	0	18-21m	0.000000	100.00
21-24m	21-24m	0	21-24m	0.000000	100.00
>24m	>24m	0	>24m	0.000000	100.00

8 rows | 1-6 of 6 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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combina
tion, 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.Dynamic
s, data = circ_data)
```

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

   7 observations deleted due to missingness
                  n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 180     10     NA      NA      NA
ctDNA.Dynamics=2  24     14  13.44   10.55      NA
ctDNA.Dynamics=3  35     25   9.27    7.62    14.1
```
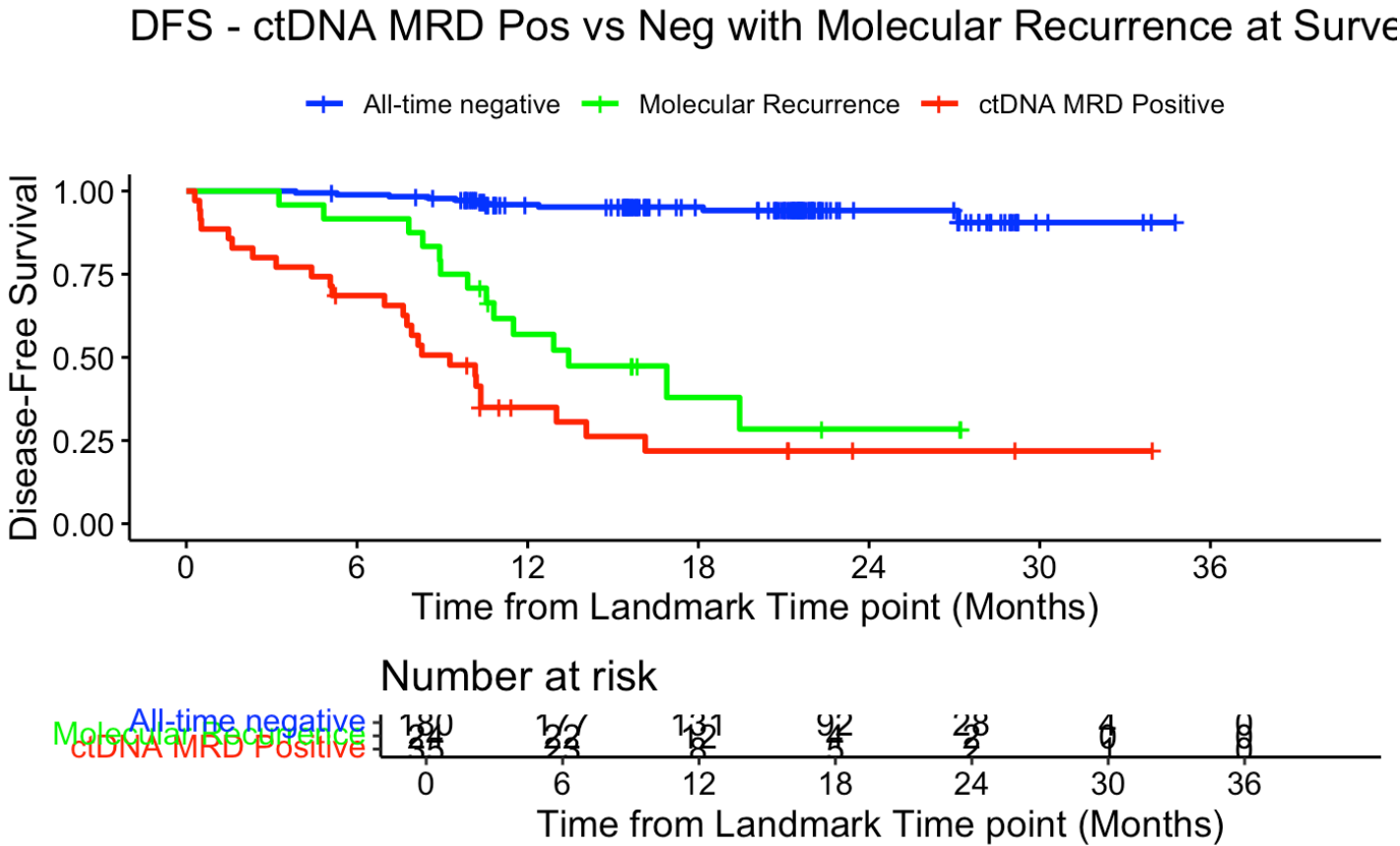
Hide

```
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>	<int>	<dbl>	<dbl>
1	180	10	0.05555556	5.555556
2	24	14	0.58333333	58.333333
3	35	25	0.71428571	71.428571
NA	7	4	0.57142857	57.142857

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.ty
pe="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue","green","red"), title="DFS - ctDNA MRD Pos vs Neg w
ith Molecular Recurrence at Surveillance Window", ylab= "Disease-Free Survival", xlab="T
ime from Landmark Time point (Months)", legend.labs=c("All-time negative","Molecular Rec
urrence", "ctDNA MRD Positive"), legend.title="")
```

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

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")

7 observations deleted due to missingness
                ctDNA.Dynamics=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    131       7    0.959  0.0151        0.916         0.98
   24     28       2    0.942  0.0194        0.889         0.97

                ctDNA.Dynamics=2
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     12      10    0.569   0.104       0.3457        0.742
   24      2       4    0.285   0.122       0.0863        0.525

                ctDNA.Dynamics=3
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      8      22    0.350  0.0830       0.1948        0.509
   24      2       3    0.219  0.0792       0.0884        0.386
```
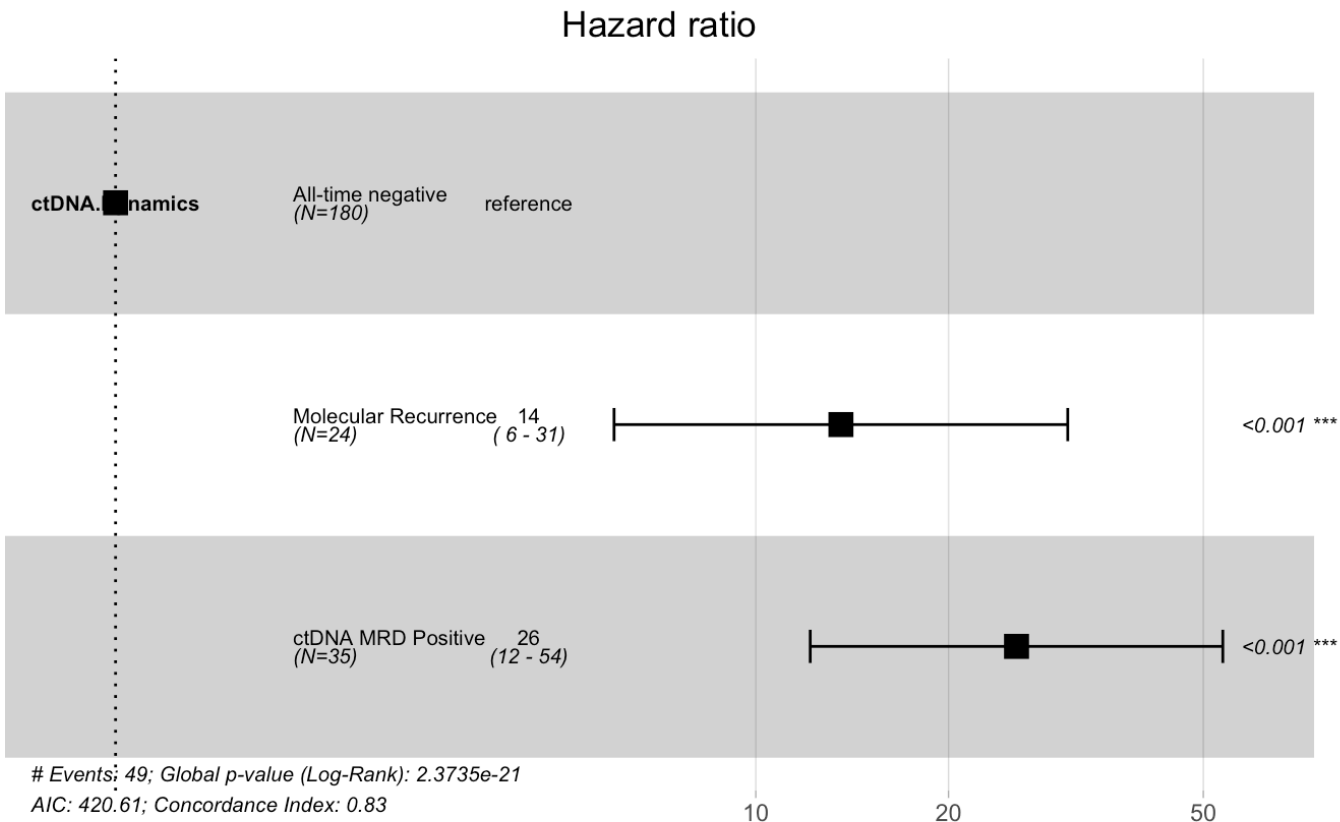
Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3"), labe
ls = 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

ctDNA.Dynamics   All-time negative (N=180)   reference

Molecular Recurrence (N=24)   14 (6 - 31)   <0.001 ***

ctDNA MRD Positive (N=35)   26 (12 - 54)   <0.001 ***

Events: 49; Global p-value (Log-Rank): 2.3735e-21
AIC: 420.61; Concordance Index: 0.83

Hide

```
summary(cox_fit)
```

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

  n= 239, number of events= 49
   (7 observations deleted due to missingness)

                                       coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsMolecular Recurrence   2.6086   13.5800   0.4162 6.267 3.68e-10 ***
ctDNA.DynamicsctDNA MRD Positive     3.2402   25.5396   0.3786 8.559  < 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     13.58    0.07364     6.006     30.70
ctDNA.DynamicsctDNA MRD Positive       25.54    0.03915    12.162     53.63


Concordance= 0.834  (se = 0.029 )
Likelihood ratio test= 94.98  on 2 df,    p=<2e-16
Wald test            = 74.17  on 2 df,    p=<2e-16
Score (logrank) test = 146.1  on 2 df,    p=<2e-16
```

Hide

```
#Fisher test for DFS percentages at 12 and 24 months - All time negative vs Molecular Re
currence
dfs_times <- c(12, 24)
circ_data <- na.omit(circ_data[, c("ctDNA.Dynamics", "DFS.MRD.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.Dynamics == "All-time negative" & circ_data$DFS.MRD.m
onths >= time & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence" & circ_data$DFS.MR
D.months >= time & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.Dynamics == "All-time negative")
  pos_total <- sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence")

  neg_surv <- neg_total - sum(circ_data$ctDNA.Dynamics == "All-time negative" & circ_dat
a$DFS.Event == 1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence" & circ_
data$DFS.Event == 1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
        8.575063e-07          6.424306e-10
```

Hide

```
#Repeat analysis to run Fisher test for All time negative vs ctDNA Positive
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combina
tion, 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)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data,conf.int=0.95,conf.ty
pe="log-log")
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")

7 observations deleted due to missingness
                ctDNA.Dynamics=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    131       7    0.959  0.0151        0.916         0.98
   24     28       2    0.942  0.0194        0.889         0.97

                ctDNA.Dynamics=2
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     12      10    0.569   0.104       0.3457        0.742
   24      2       4    0.285   0.122       0.0863        0.525

                ctDNA.Dynamics=3
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      8      22    0.350  0.0830       0.1948        0.509
   24      2       3    0.219  0.0792       0.0884        0.386
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3"), labe
ls = c("All-time negative","Molecular Recurrence", "ctDNA MRD Positive"))

#Fisher test for DFS percentages at 12 and 24 months - All time negative vs ctDNA Positi
ve
dfs_times <- c(12, 24)
circ_data <- na.omit(circ_data[, c("ctDNA.Dynamics", "DFS.MRD.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.Dynamics == "All-time negative" & circ_data$DFS.MRD.m
onths >= time & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive" & circ_data$DFS.MRD.
months >= time & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.Dynamics == "All-time negative")
  pos_total <- sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive")

  neg_surv <- neg_total - sum(circ_data$ctDNA.Dynamics == "All-time negative" & circ_dat
a$DFS.Event == 1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive" & circ_da
ta$DFS.Event == 1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
       2.382798e-15         1.933852e-17
```

Hide

```
#Repeat analysis to run Fisher test for Molecular Recurrence vs ctDNA Positive
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combina
tion, 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)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data,conf.int=0.95,conf.ty
pe="log-log")
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")

7 observations deleted due to missingness
                ctDNA.Dynamics=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    131       7    0.959  0.0151        0.916         0.98
   24     28       2    0.942  0.0194        0.889         0.97

                ctDNA.Dynamics=2
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     12      10    0.569   0.104       0.3457        0.742
   24      2       4    0.285   0.122       0.0863        0.525

                ctDNA.Dynamics=3
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12      8      22    0.350  0.0830       0.1948        0.509
   24      2       3    0.219  0.0792       0.0884        0.386
```

Hide

```r
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3"), labe
ls = c("All-time negative","Molecular Recurrence", "ctDNA MRD Positive"))

#Fisher test for DFS percentages at 12 and 24 months - Molecular Recurrence vs ctDNA Pos
itive
dfs_times <- c(12, 24)
circ_data <- na.omit(circ_data[, c("ctDNA.Dynamics", "DFS.MRD.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence" & circ_data$DFS.MR
D.months >= time & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive" & circ_data$DFS.MRD.
months >= time & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence")
  pos_total <- sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive")

  neg_surv <- neg_total - sum(circ_data$ctDNA.Dynamics == "Molecular Recurrence" & circ_
data$DFS.Event == 1 & circ_data$DFS.MRD.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.Dynamics == "ctDNA MRD Positive" & circ_da
ta$DFS.Event == 1 & circ_data$DFS.MRD.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
           0.1214065            0.4023373
```

Hide

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

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combina
tion, 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"), labels =
c("Molecular Recurrence", "ctDNA MRD Positive"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit,data = circ_data)
```



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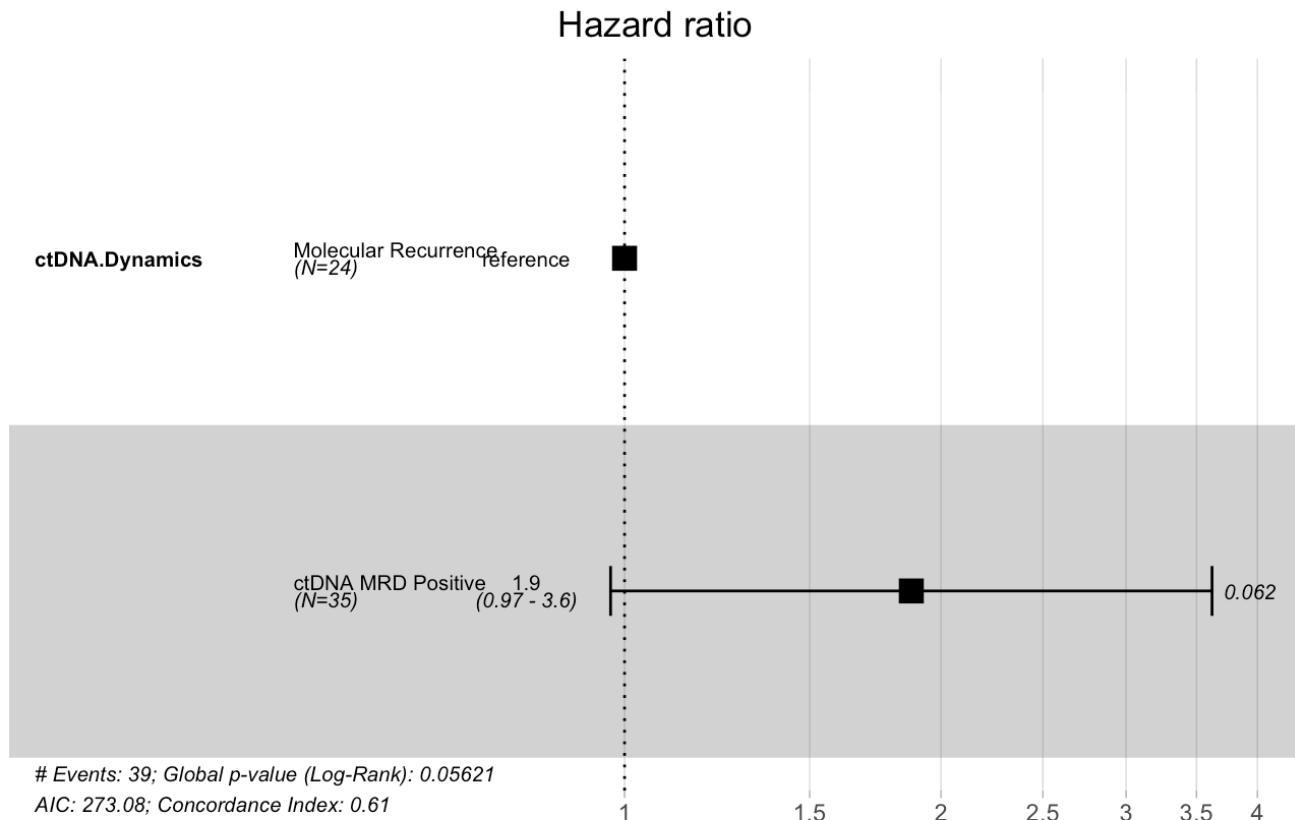
```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

  n= 59, number of events= 39
    (187 observations deleted due to missingness)

                                  coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsctDNA MRD Positive 0.6281    1.8741   0.3361 1.869   0.0616 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                                 exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsctDNA MRD Positive     1.874     0.5336    0.9699     3.621

Concordance= 0.608  (se = 0.037 )
Likelihood ratio test= 3.65  on 1 df,    p=0.06
Wald test            = 3.49  on 1 df,    p=0.06
Score (logrank) test = 3.6  on 1 df,    p=0.06
```

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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.87 (0.97-3.62); p = 0.062"
```

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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)

survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Surveillanc
e, 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 193    10     NA      NA      NA
ctDNA.Surveillance=POSITIVE  40    30   8.79    7.16      12
```

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	193	10	0.05181347	5.181347
POSITIVE	40	30	0.75000000	75.000000

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,con
f.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA Surveillance window | Al
l 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
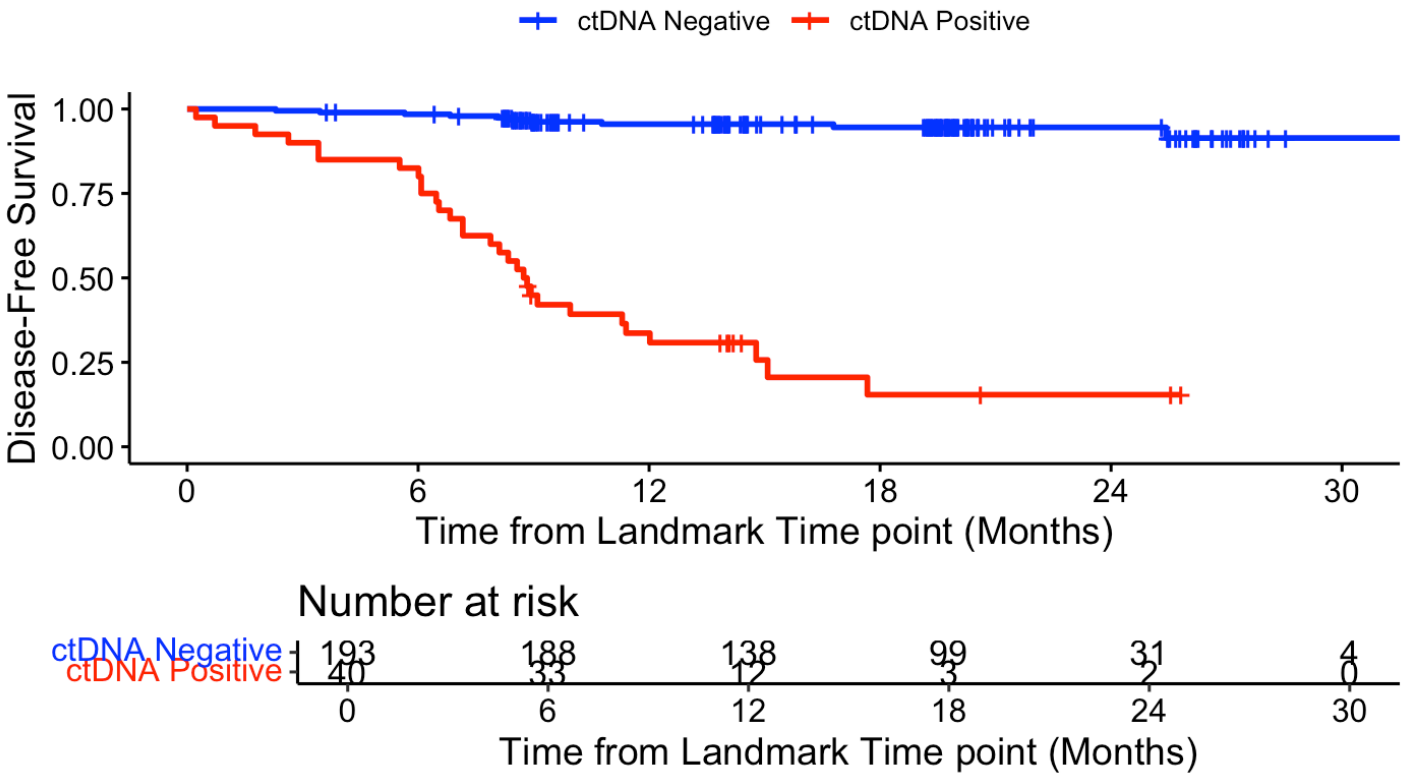

```
summary(KM_curve, times= c(12, 24))
```

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

                ctDNA.Surveillance=NEGATIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12    138       8    0.955  0.0156        0.912        0.977
   24     31       1    0.946  0.0181        0.896        0.972


                ctDNA.Surveillance=POSITIVE
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   12     12      26    0.336  0.0765       0.1940        0.485
   24      2       4    0.154  0.0733       0.0461        0.321
```
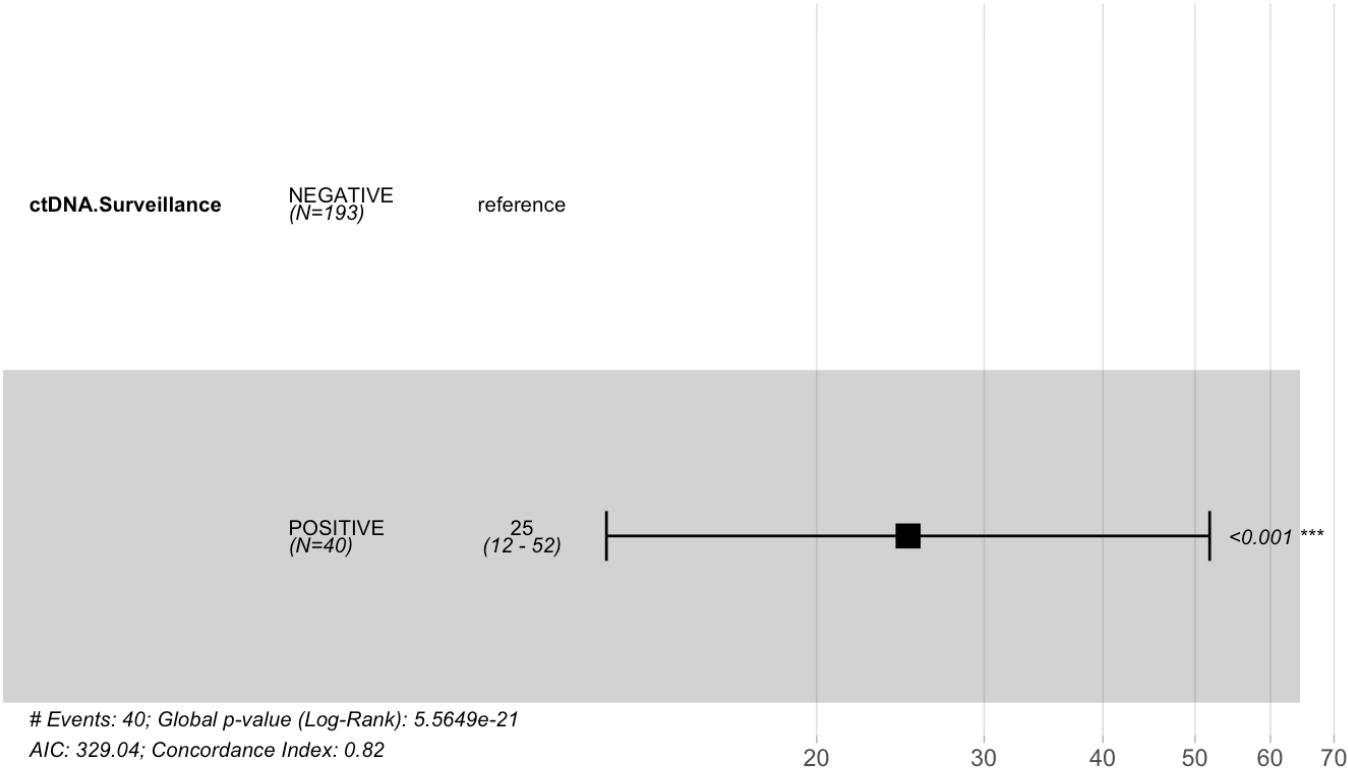
Hide

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

Hazard ratio

ctDNA.Surveillance   NEGATIVE (N=193)   reference

POSITIVE (N=40)   25 (12 - 52)   <0.001 ***

Events: 40; Global p-value (Log-Rank): 5.5649e-21
AIC: 329.04; Concordance Index: 0.82

20   30   40   50   60   70

Hide

```
summary(cox_fit)
```

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

  n= 233, number of events= 40


                            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 3.2170   24.9530   0.3726 8.633   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                           exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE     24.95    0.04008     12.02      51.8

Concordance= 0.822  (se = 0.033 )
Likelihood ratio test= 88.32  on 1 df,   p=<2e-16
Wald test            = 74.53  on 1 df,   p=<2e-16
Score (logrank) test = 156.3  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_C
I, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 24.95 (12.02-51.8); p = 0"
```

Hide

```
#Fisher test for DFS percentages at 12 and 24 months
dfs_times <- c(12, 24)
circ_data <- na.omit(circ_data[, c("ctDNA.Surveillance", "DFS.months", "DFS.Event")])
p_values <- sapply(dfs_times, function(time) {
  neg_count <- sum(circ_data$ctDNA.Surveillance == "NEGATIVE" & circ_data$DFS.months >=
time & circ_data$DFS.Event == 0)
  pos_count <- sum(circ_data$ctDNA.Surveillance == "POSITIVE" & circ_data$DFS.months >=
time & circ_data$DFS.Event == 0)
  neg_total <- sum(circ_data$ctDNA.Surveillance == "NEGATIVE")
  pos_total <- sum(circ_data$ctDNA.Surveillance == "POSITIVE")

  neg_surv <- neg_total - sum(circ_data$ctDNA.Surveillance == "NEGATIVE" & circ_data$DF
S.Event == 1 & circ_data$DFS.months < time)
  pos_surv <- pos_total - sum(circ_data$ctDNA.Surveillance == "POSITIVE" & circ_data$DF
S.Event == 1 & circ_data$DFS.months < time)

  surv_matrix <- matrix(c(neg_surv, pos_surv, neg_total - neg_surv, pos_total - pos_sur
v), nrow = 2)
  test_result <- fisher.test(surv_matrix)
  return(test_result$p.value)
})
names(p_values) <- paste0("p-value at ", dfs_times, " months")
print(p_values)
```

```
p-value at 12 months p-value at 24 months
        1.175567e-17         2.041854e-21
```
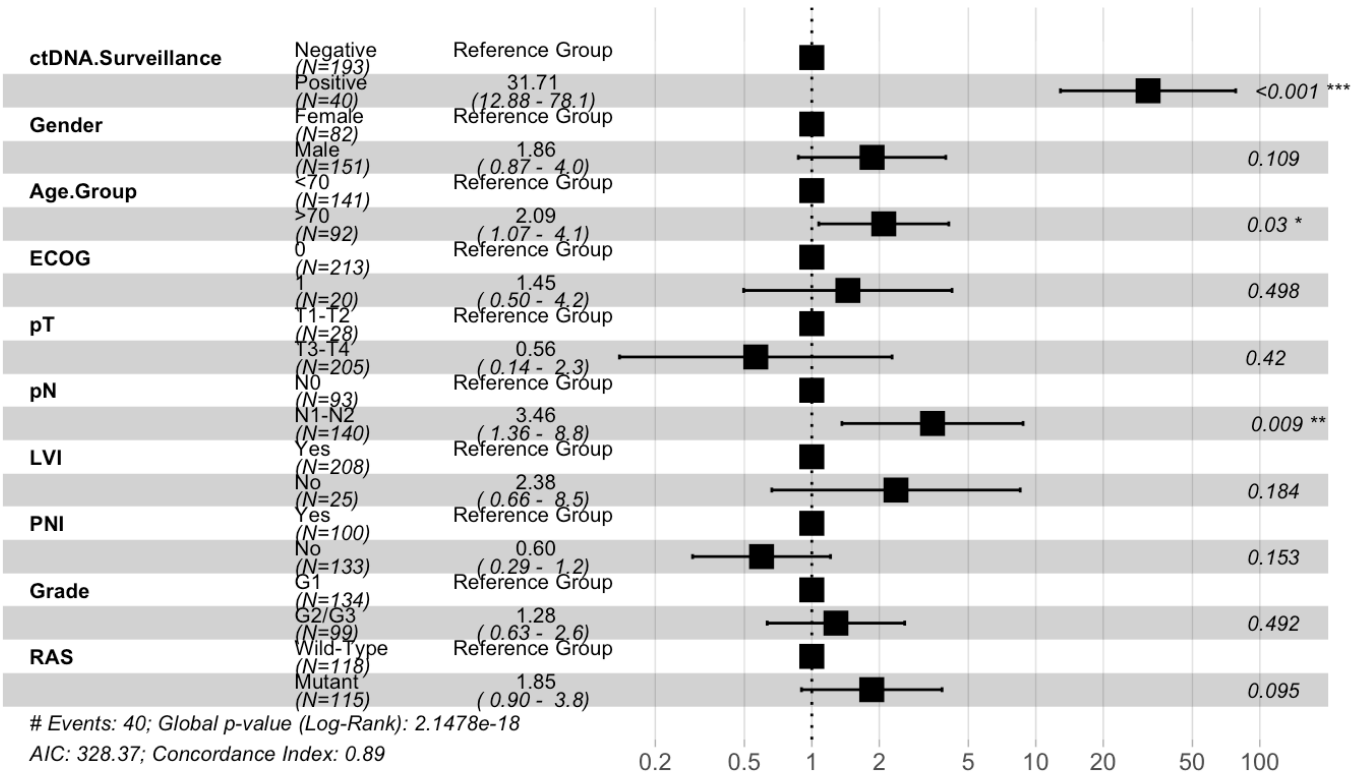
#Multivariate cox regression at Surveillance Window for DFS - All stages Landmark 10 weeks

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$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("NEGATIV
E","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("<7
0", ">70"))
circ_data$PrimSite <- factor(circ_data$PrimSite, levels = c("Left-sided colon", "Right-s
ided 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$LVI <- factor(circ_data$LVI, levels = c("TRUE", "FALSE"), labels = c("Yes", "N
o"))
circ_data$PNI <- factor(circ_data$PNI, levels = c("TRUE", "FALSE"), labels = c("Yes", "N
o"))
circ_data$Grade <- factor(circ_data$Grade, levels = c("G1", "G2/G3"))
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 + ECOG + pT + pN
+ LVI + PNI + Grade + 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

ctDNA.Surveillance	Negative (N=193)	Reference Group		
	Positive (N=40)	31.71 (12.88 - 78.1)		<0.001 ***
Gender	Female (N=82)	Reference Group		
	Male (N=151)	1.86 (0.87 - 4.0)		0.109
Age.Group	<70 (N=141)	Reference Group		
	>70 (N=92)	2.09 (1.07 - 4.1)		0.03 *
ECOG	0 (N=213)	Reference Group		
	1 (N=20)	1.45 (0.50 - 4.2)		0.498
pT	T1-T2 (N=28)	Reference Group		
	T3-T4 (N=205)	0.56 (0.14 - 2.3)		0.42
pN	N0 (N=93)	Reference Group		
	N1-N2 (N=140)	3.46 (1.36 - 8.8)		0.009 **
LVI	Yes (N=208)	Reference Group		
	No (N=25)	2.38 (0.66 - 8.5)		0.184
PNI	Yes (N=100)	Reference Group		
	No (N=133)	0.60 (0.29 - 1.2)		0.153
Grade	G1 (N=134)	Reference Group		
	G2/G3 (N=99)	1.28 (0.63 - 2.6)		0.492
RAS	Wild-Type (N=118)	Reference Group		
	Mutant (N=115)	1.85 (0.90 - 3.8)		0.095

 # Events: 40; Global p-value (Log-Rank): 2.1478e-18

AIC: 328.37; Concordance Index: 0.89

Hide

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

#DFS by ctDNA Dynamics from MRD to 6 months - all stages

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("Galaxy Data_20240603 Complete Dataset.csv")
circ_data <- circ_data[circ_data$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$DFS.6mo.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$ctDNA.Dynamics <- NA #first we create the variable for the ctDNA & NAC combina
tion, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.6months == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "NEGATIVE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ctDNA.6months == "POSITIVE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ctDNA.6months == "POSITIVE" ~ 4
  ))

survfit(Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Event)~ctDNA.Dynamic
s, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.6mo.months, event = circ_data$DFS.Even
t) ~
    ctDNA.Dynamics, data = circ_data)

   66 observations deleted due to missingness
                   n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 157     16     NA      NA      NA
ctDNA.Dynamics=2  14      5     NA    5.52      NA
ctDNA.Dynamics=3   4      2   8.51    0.00      NA
ctDNA.Dynamics=4   5      5   2.76    2.60      NA
```

```
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>	<int>	<dbl>	<dbl>
1	157	16	0.1019108	10.19108
2	14	5	0.3571429	35.71429
3	4	2	0.5000000	50.00000
4	5	5	1.0000000	100.00000
NA	66	NA	NA	NA

5 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.ty
pe="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRU
E, break.time.by=6, palette=c("blue","green","purple", "red"), title="DFS - ctDNA Dynami
cs from MRD to 6 months | All Stages", ylab= "Disease-Free Survival", xlab="Time from La
ndmark Time point (Months)", legend.labs=c("Persistently Negative", "Converted Negativ
e","Converted Positive", "Persistently Positive"), legend.title="")
summary(KM_curve, times= c(24))
```

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

66 observations deleted due to missingness
            ctDNA.Dynamics=1
     time     n.risk    n.event    survival    std.err lower 95% CI upper 95%
CI
    24.000    12.000    16.000      0.851       0.045       0.736        0.9
19

            ctDNA.Dynamics=2
     time     n.risk    n.event    survival    std.err lower 95% CI upper 95%
CI
    24.000     2.000     5.000      0.542       0.162       0.204        0.7
89

            ctDNA.Dynamics=3
    time n.risk n.event survival std.err lower 95% CI upper 95% CI

            ctDNA.Dynamics=4
    time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3", "4"),
labels = c("Persistently Negative", "Converted Negative","Converted Positive", "Persiste
ntly Positive"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hide

```
summary(cox_fit)
```

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

  n= 180, number of events= 28
   (66 observations deleted due to missingness)

                                        coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.DynamicsConverted Negative      1.5689    4.8013   0.5147 3.048  0.00230 **
ctDNA.DynamicsConverted Positive      2.1062    8.2170   0.7582 2.778  0.00547 **
ctDNA.DynamicsPersistently Positive   3.8173   45.4828   0.5900 6.470  9.8e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

                                     exp(coef) exp(-coef) lower .95 upper .95
ctDNA.DynamicsConverted Negative         4.801    0.20828     1.751     13.17
ctDNA.DynamicsConverted Positive         8.217    0.12170     1.859     36.32
ctDNA.DynamicsPersistently Positive     45.483    0.02199    14.310    144.57

Concordance= 0.697  (se = 0.048 )
Likelihood ratio test= 33.4  on 3 df,   p=3e-07
Wald test            = 46.4  on 3 df,   p=5e-10
Score (logrank) test = 106.2  on 3 df,    p=<2e-16
```

#Table with recurrence sites by ctDNA at the MRD Window

Hide

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

circ_data_subset1 <- circ_data %>%
  select(
    Rel.Site) %>%
  mutate(
    Rel.Site = factor(Rel.Site, levels = c("Local", "Lymph Node", "Liver", "Lung", "Peri
toneum", "Brain")))

circ_data_subset2 <- circ_data %>%
  select(
    Rel.Site,
    ctDNA.MRD) %>%
  mutate(
    Rel.Site = factor(Rel.Site, levels = c("Local", "Lymph Node", "Liver", "Lung", "Peri
toneum", "Brain")),
    ctDNA.MRD = factor(ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE")))
Overall <- circ_data_subset1 %>%
  tbl_summary(
    statistic = list(
      all_continuous() ~ "{median} ({min} – {max})",
      all_categorical() ~ "{n} ({p}%)")) %>%
  bold_labels()
Overall
```

Characteristic	N = 48[1]
Rel.Site	
Local	4 (8.3%)
Lymph Node	5 (10%)
Liver	17 (35%)
Lung	18 (38%)
Peritoneum	3 (6.3%)
Brain	1 (2.1%)
[1] n (%)	

Hide

```
ByctDNA_MRD <- circ_data_subset2 %>%
  tbl_summary(
    by = ctDNA.MRD, # add this line to subgroup by ctDNA.MRD
    statistic = list(
      all_continuous() ~ "{median} ({min} — {max})",
      all_categorical() ~ "{n} ({p}%)")) %>%
  add_p() %>%
  bold_labels()
ByctDNA_MRD
```

Characteristic	NEGATIVE N = 23[1]	POSITIVE N = 25[1]	p-value[2]
Rel.Site			<0.001
Local	2 (8.7%)	2 (8.0%)	
Lymph Node	1 (4.3%)	4 (16%)	
Liver	2 (8.7%)	15 (60%)	
Lung	15 (65%)	3 (12%)	
Peritoneum	2 (8.7%)	1 (4.0%)	
Brain	1 (4.3%)	0 (0%)	

[1] n (%)

[2] Fisher's exact test

Hide

```
merged_table <- tbl_merge(tbls=list(Overall, ByctDNA_MRD))
merged_table
```

Characteristic	Table 1 N = 48[1]	Table 2 NEGATIVE N = 23[1]	POSITIVE N = 25[1]	p-value[2]
Rel.Site				<0.001
Local	4 (8.3%)	2 (8.7%)	2 (8.0%)	
Lymph Node	5 (10%)	1 (4.3%)	4 (16%)	
Liver	17 (35%)	2 (8.7%)	15 (60%)	

[1] n (%)

[2] Fisher's exact test

Characteristic	Table 1	Table 2		p-value[2]
	N = 48[1]	NEGATIVE N = 23[1]	POSITIVE N = 25[1]	
Lung	18 (38%)	15 (65%)	3 (12%)	
Peritoneum	3 (6.3%)	2 (8.7%)	1 (4.0%)	
Brain	1 (2.1%)	1 (4.3%)	0 (0%)	

[1] n (%)

[2] Fisher's exact test

```
fit1 <- as_flex_table(
  merged_table,
  include = everything(),
  return_calls = FALSE)
fit1
```

Characteristic	Table 1	Table 2		p-value[2]
	N = 48[1]	NEGATIVE N = 23[1]	POSITIVE N = 25[1]	
Rel.Site				<0.001
Local	4 (8.3%)	2 (8.7%)	2 (8.0%)	
Lymph Node	5 (10%)	1 (4.3%)	4 (16%)	
Liver	17 (35%)	2 (8.7%)	15 (60%)	
Lung	18 (38%)	15 (65%)	3 (12%)	
Peritoneum	3 (6.3%)	2 (8.7%)	1 (4.0%)	
Brain	1 (2.1%)	1 (4.3%)	0 (0%)	

[1]n (%)

[2]Fisher's exact test

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

#MTM/mL levels at the MRD Window by Radiological Recurrence

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

# Transform p_MRD_MTM with log10
circ_data$p_MRD_MTM <- as.numeric(as.character(circ_data$p_MRD_MTM))
circ_data$RFS.Event <- factor(circ_data$RFS.Event, levels = c("FALSE", "TRUE"), labels =
c("No Recurrence", "Recurrence"))
summary_stats <- circ_data %>%
  group_by(RFS.Event) %>%
  summarise(
    median_p_MRD_MTM = median(p_MRD_MTM, na.rm = TRUE),
    range_p_MRD_MTM = paste0(min(p_MRD_MTM, na.rm = TRUE), " - ", max(p_MRD_MTM, na.rm =
TRUE))
  )
print(summary_stats)
```

RFS.Event <fctr>	median_p_MRD_MTM <dbl>	range_p_MRD_MTM <chr>
No Recurrence	0.00000000	0 - 9.694491335
Recurrence	0.06305439	0 - 645.7568074

2 rows

Hide

```
m3_1v2 <- wilcox.test(p_MRD_MTM ~ RFS.Event,
                  data = circ_data[circ_data$RFS.Event %in% c("No Recurrence", "Recu
rrence"), ],
                  na.rm = TRUE)
print(m3_1v2)
```
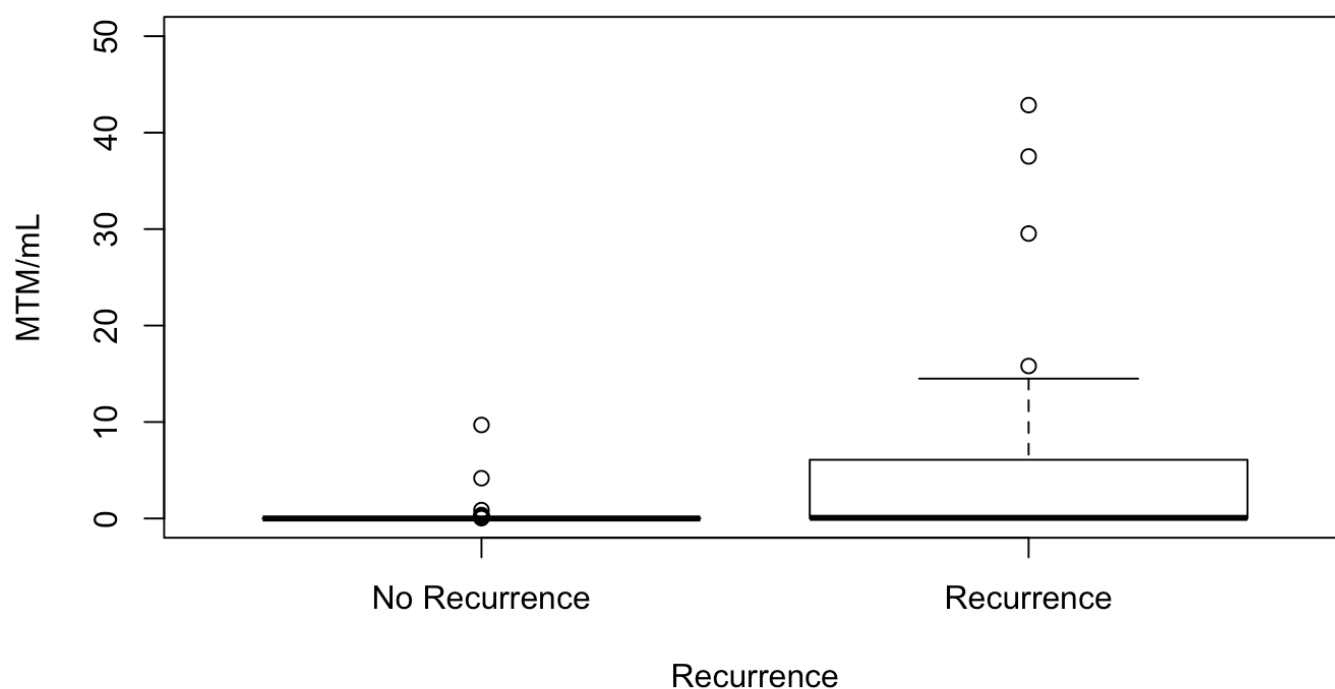
```
    Wilcoxon rank sum test with continuity correction

data:  p_MRD_MTM by RFS.Event
W = 2458, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Hide

```
boxplot(p_MRD_MTM~RFS.Event, data=circ_data, main="ctDNA MRD Window MTM - Recurrence", x
lab="Recurrence", ylab="MTM/mL", col="white",border="black", ylim = c(0, 50))
```

ctDNA MRD Window MTM - Recurrence

#MTM/mL levels at the MRD Window by Radiological Recurrence Sites

Hide

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

# Transform p_MRD_MTM with log10
circ_data$p_MRD_MTM <- as.numeric(as.character(circ_data$p_MRD_MTM))
circ_data$Rel.Site <- factor(circ_data$Rel.Site, levels = c("Liver", "Lymph Node", "Loca
l", "Peritoneum", "Lung", "Brain"))
summary_stats <- circ_data %>%
  group_by(Rel.Site) %>%
  summarise(
    median_p_MRD_MTM = median(p_MRD_MTM, na.rm = TRUE),
    range_p_MRD_MTM = paste0(min(p_MRD_MTM, na.rm = TRUE), " - ", max(p_MRD_MTM, na.rm =
TRUE))
    )
print(summary_stats)
```

Rel.Site	median_p_MRD_MTM	range_p_MRD_MTM
<fctr>	<dbl>	<chr>
Liver	1.66880485	0 - 645.7568074
Lymph Node	4.10239125	0 - 42.85542855
Local	0.07993897	0 - 0.2238338905
Peritoneum	0.00000000	0 - 15.80708422
Lung	0.00000000	0 - 11.24115557
Brain	0.00000000	0 - 0

6 rows

Hide

```
kruskal_test <- kruskal.test(p_MRD_MTM ~ Rel.Site, data = circ_data)
print(kruskal_test)
```
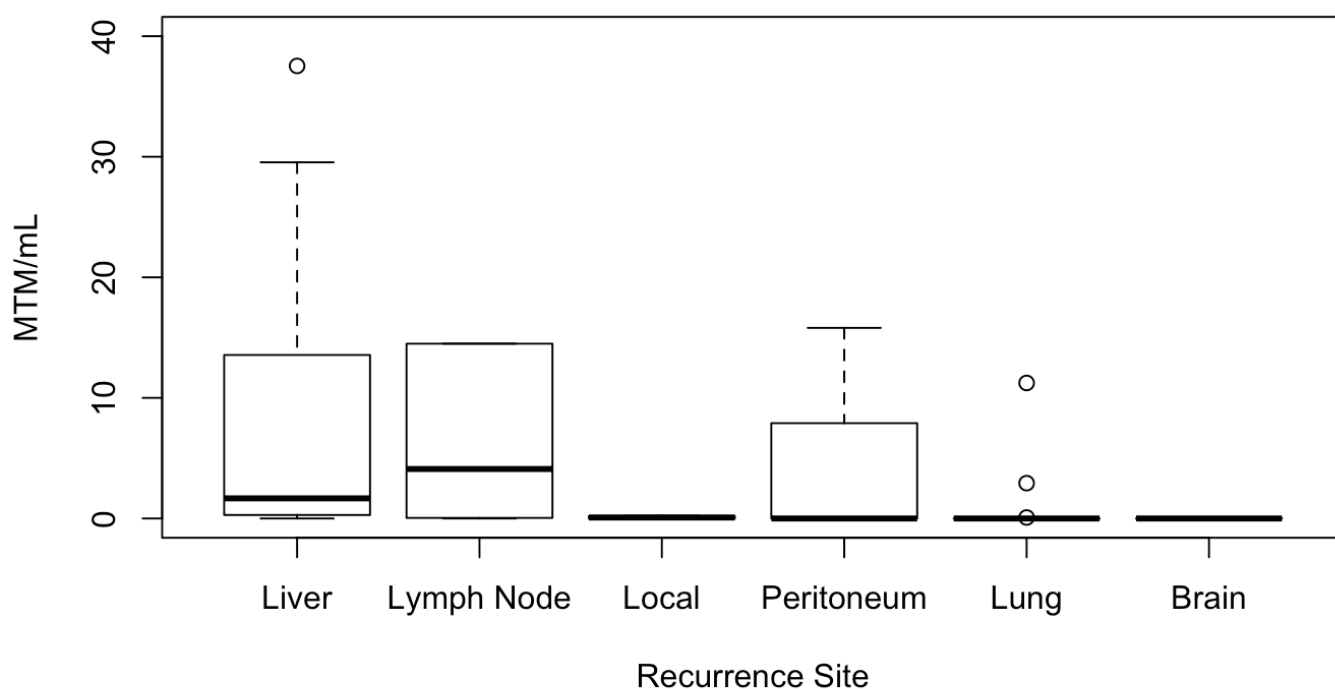
```
	Kruskal-Wallis rank sum test

data:  p_MRD_MTM by Rel.Site
Kruskal-Wallis chi-squared = 20.366, df = 5, p-value = 0.001067
```

Hide

```
boxplot(p_MRD_MTM~Rel.Site, data=circ_data, main="ctDNA MRD Window MTM - Recurrence Site", xlab="Recurrence Site", ylab="MTM/mL", col="white",border="black", ylim = c(0, 40))
```

ctDNA MRD Window MTM - Recurrence Site

#Barplot with Recurrence Sites (Liver vs Others) by ctDNA at the MRD Window

Hide

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

circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"), l
abels = c("Negative", "Positive"))
circ_data$Rel.Site <- factor(circ_data$Rel.Site, levels = c("Liver", "Lymph Node", "Loca
l", "Peritoneum", "Lung", "Brain"))
contingency_table <- table(circ_data$Rel.Site, circ_data$ctDNA.MRD)
chi_square_test <- chisq.test(contingency_table)
```

```
Warning: Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

```
    Pearson's Chi-squared test

data:  contingency_table
X-squared = 21.028, df = 5, p-value = 0.0008004
```

Hide

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

```
    Fisher's Exact Test for Count Data

data:  contingency_table
p-value = 5.02e-05
alternative hypothesis: two.sided
```

Hide

```
print(contingency_table)
```

```
            Negative Positive
  Liver            2       15
  Lymph Node       1        4
  Local            2        2
  Peritoneum       2        1
  Lung            15        3
  Brain            1        0
```
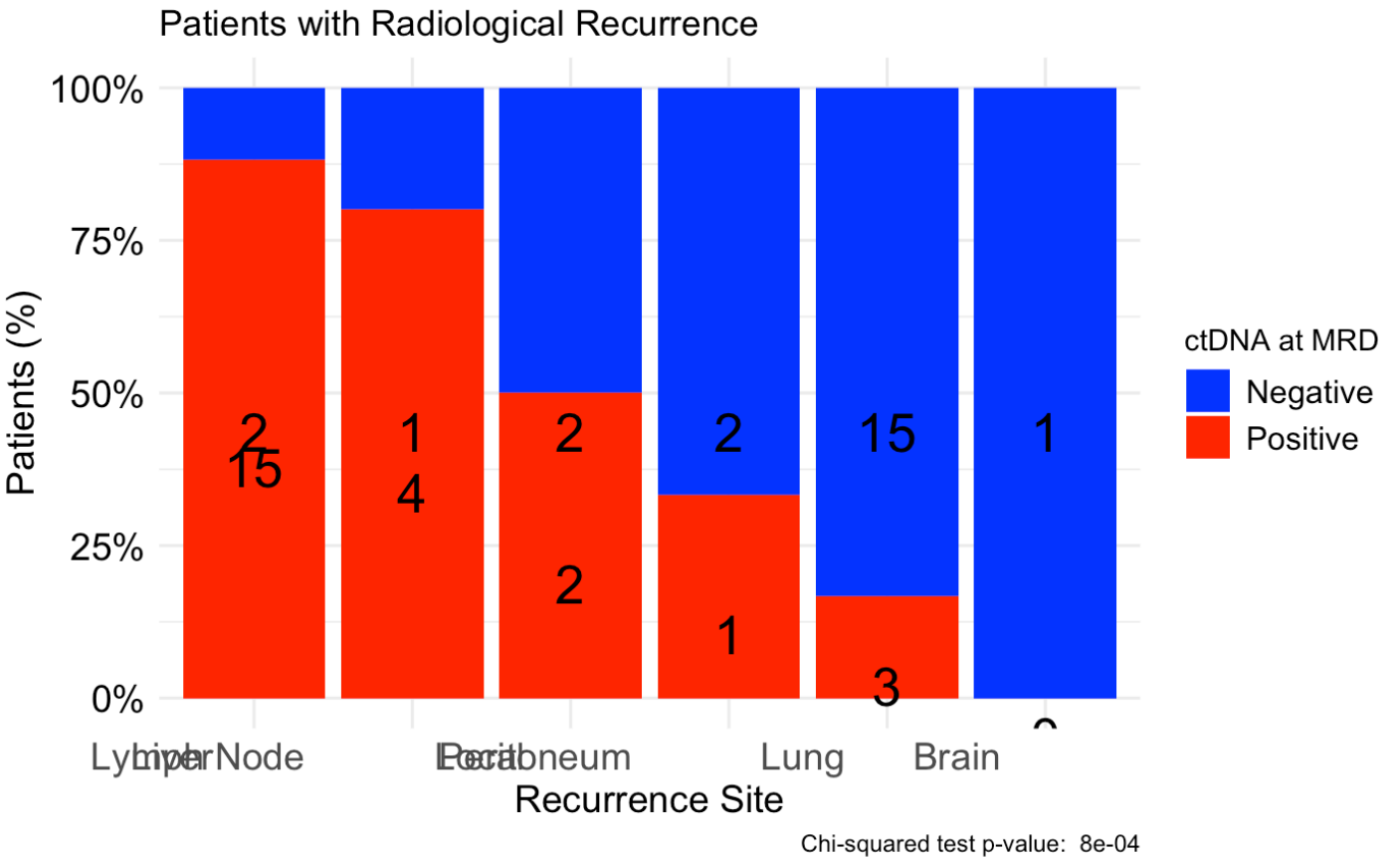
Hide

```r
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "blac
k", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "Patients with Radiological Recurrence",
       x = "Recurrence Site",
       y = "Patients (%)",
       fill = "ctDNA at MRD",
       caption = paste("Chi-squared test p-value: ", format.pval(chi_square_test$p.valu
e))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("Negative" = "blue", "Positive" = "red")) + # define cust
om colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis
text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text s
ize
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label
size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label
size
        legend.text = element_text(size = 12, color = "black"))  # increase Recurrence l
abel size
```

Patients with Radiological Recurrence

Chi-squared test p-value: 8e-04

#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$CohortB=="TRUE",]
circ_data <- circ_data[circ_data$RFS.Event=="TRUE",]

# Create a table of counts for the "Rel.Site" variable
relsite_counts <- table(circ_data$Rel.Site)
relsite_df <- as.data.frame(relsite_counts)
names(relsite_df) <- c("Rel.Site", "Count")
circ_data_pos_mrd <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data_pos_anytime <- circ_data[circ_data$ctDNA.anytime=="POSITIVE",]
pos_counts_mrd <- table(circ_data_pos_mrd$Rel.Site)
pos_counts_anytime <- table(circ_data_pos_anytime$Rel.Site)
relsite_df$MRDPos_Count <- ifelse(is.na(match(relsite_df$Rel.Site, names(pos_counts_mr
d))), 0, pos_counts_mrd[match(relsite_df$Rel.Site, names(pos_counts_mrd))])
relsite_df$MRDPos_Count[is.na(relsite_df$MRDPos_Count)] <- 0
relsite_df$AnytimePos_Count <- ifelse(is.na(match(relsite_df$Rel.Site, names(pos_counts_
anytime))), 0, pos_counts_anytime[match(relsite_df$Rel.Site, names(pos_counts_anytim
e))])
relsite_df$AnytimePos_Count[is.na(relsite_df$AnytimePos_Count)] <- 0
relsite_df$Percent <- (relsite_df$Count / sum(relsite_df$Count)) * 100
relsite_df$MRDPos_Percent <- (relsite_df$MRDPos_Count / relsite_df$Count) * 100
relsite_df$AnytimePos_Percent <- (relsite_df$AnytimePos_Count / relsite_df$Count) * 100
total_observations <- sum(relsite_df$Count)
total_pos_mrd <- sum(relsite_df$MRDPos_Count)
total_pos_anytime <- sum(relsite_df$AnytimePos_Count)
total_row <- data.frame(Rel.Site = "Total", Count = total_observations, MRDPos_Count = t
otal_pos_mrd, AnytimePos_Count = total_pos_anytime, Percent = 100, MRDPos_Percent = (tot
al_pos_mrd / total_observations) * 100, AnytimePos_Percent = (total_pos_anytime / total_
observations) * 100)
relsite_df <- rbind(relsite_df, total_row)
print(relsite_df)
```

Rel.Site	Co...	MRDPos_C...	AnytimePos_Count	Percent	MRDPos_Perc...	AnytimePos
<fctr>	<int>	<dbl>	<dbl>	<dbl>	<dbl>	
Brain	1	0	0	2.083333	0.00000	
Liver	17	15	17	35.416667	88.23529	1
Local	4	2	4	8.333333	50.00000	1
Lung	18	3	10	37.500000	16.66667	
Lymph Node	5	4	5	10.416667	80.00000	1
Peritoneum	3	1	3	6.250000	33.33333	1
Total	48	25	39	100.000000	52.08333	

7 rows

Hide

```
ft <- flextable(relsite_df)
doc <- read_docx() %>%
  body_add_flextable(value = ft)
print(doc, target = "relsite_df.docx")
```