

Code ▾

# CLIA HNSCC UNM Clinical Analysis 02272025

```
library(swimplot) library(grid) library(gttable) library(readr) library(mosaic) library(dplyr) library(survival) library(survminer) library(ggplot2)
library(scales) library(coxphf) library(ggthemes) library(tidyverse) library(gtsummary) library(flextable) library(reshape2) library(parameters)
library(car) library(ComplexHeatmap) library(tidyverse) library(readxl) library(janitor) library(DT) library(pROC) library(rms)
```

```
#ctDNA Detection Rates by Window and Stages
```

Hide

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

Stage <fctr>	Total_Count <int>	Positive_Count <int>	Rate <chr>
I/II	34	32	94.12%
III/IVA/IVB	26	21	80.77%
IVC	2	2	100.00%
Overall	62	55	88.71%
4 rows			

Hide

```

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

```

Stage	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<chr>
I/II	34	5	14.71%
III/IVA/IVB	33	7	21.21%
IVC	2	1	50.00%
Overall	69	13	18.84%
4 rows			

Hide

```

#ctDNA at Surveillance
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I/II","III/IVA/IVB","IVC"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.Surveillance == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)

```

Stage <fctr>	Total_Count <int>	Positive_Count <int>	Rate <chr>
I/II	37	6	16.22%
III/IVA/IVB	30	11	36.67%
IVC	1	0	0.00%
Overall	68	17	25.00%
4 rows			

#Demographics Table

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]

circ_data_subset <- circ_data %>%
  select(
    Sex,
    Age,
    Tobacco.History,
    Prim.Location,
    cT,
    cN,
    cM,
    Histology,
    Stage,
    p16.status,
    Treatment.Group,
    PFS.Event,
    OS.Event,
    OS.months) %>%
  mutate(
    Sex = factor(Sex),
    Age = as.numeric(Age),
    Tobacco.History = factor(Tobacco.History),
    Prim.Location = factor(Prim.Location),
    cT = factor(cT),
    cN = factor(cN),
    cM = factor(cM),
    Histology = factor(Histology),
    Stage = factor(Stage),
    p16.status = factor(p16.status),
    Treatment.Group = factor(Treatment.Group),
    PFS.Event = factor(PFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Progression", "Progression")),
    OS.Event = factor(OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased")),
    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 = 97 <sup>1</sup>
<b>Sex</b>	
Female	17 (18%)
Male	80 (82%)
<b>Age</b>	66 (29 - 95)
<b>Tobacco.History</b>	63 (65%)
<b>Prim.Location</b>	
Larynx/Hypopharynx	5 (5.2%)
Oral cavity	16 (16%)

<sup>1</sup> n (%); Median (Min - Max)

Characteristic	N = 97 <sup>1</sup>
Oropharynx	67 (69%)
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)
<b>cT</b>	
T0	2 (2.1%)
T1	12 (12%)
T2	31 (32%)
T3	30 (31%)
T4	21 (22%)
TX	1 (1.0%)
<b>cN</b>	
N0	22 (23%)
N1	33 (34%)
N2	33 (34%)
N3	9 (9.3%)
<b>cM</b>	
M0	93 (96%)
M1	4 (4.1%)
<b>Histology</b>	
Adenosquamous carcinoma	1 (1.0%)
Basaloid squamous cell carcinoma	6 (6.2%)
Epithelial myoepithelial carcinoma	1 (1.0%)
Squamous cell carcinoma	86 (89%)
Undifferentiated carcinoma	3 (3.1%)
<b>Stage</b>	
I/II	49 (51%)
III/IVA/IVB	45 (46%)
IVC	3 (3.1%)
<b>p16.status</b>	
Negative	43 (44%)
Positive	54 (56%)
<b>Treatment.Group</b>	
Definitive CRT or RT	69 (71%)
None (Declined Treatment)	1 (1.0%)
None (Hospice)	2 (2.1%)
Surgery + CRT or RT	24 (25%)
Surgery only	1 (1.0%)
<b>PFS.Event</b>	
No Progression	65 (67%)
Progression	32 (33%)
<b>OS.Event</b>	
<sup>1</sup> n (%); Median (Min - Max)	

Characteristic	N = 97 <sup>1</sup>
Alive	81 (84%)
Deceased	16 (16%)
<b>OS.months</b>	22 (2 - 56)

<sup>1</sup> n (%); Median (Min - Max)

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

Characteristic	N = 97 <sup>1</sup>
<b>Sex</b>	
Female	17 (18%)
Male	80 (82%)
<b>Age</b>	66 (29 - 95)
<b>Tobacco.History</b>	63 (65%)
<b>Prim.Location</b>	
Larynx/Hypopharynx	5 (5.2%)
Oral cavity	16 (16%)
Oropharynx	67 (69%)
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)
<b>cT</b>	
T0	2 (2.1%)
T1	12 (12%)
T2	31 (32%)
T3	30 (31%)
T4	21 (22%)
TX	1 (1.0%)
<b>cN</b>	
N0	22 (23%)
N1	33 (34%)
N2	33 (34%)
N3	9 (9.3%)
<b>cM</b>	
M0	93 (96%)
M1	4 (4.1%)
<b>Histology</b>	
Adenosquamous carcinoma	1 (1.0%)
Basaloid squamous cell carcinoma	6 (6.2%)
Epithelial myoepithelial carcinoma	1 (1.0%)
Squamous cell carcinoma	86 (89%)
Undifferentiated carcinoma	3 (3.1%)
<b>Stage</b>	
I/II	49 (51%)
III/IVA/IVB	45 (46%)
IVC	3 (3.1%)

<sup>1</sup>n (%); Median (Min - Max)

Characteristic	N = 97 <sup>1</sup>
<b>p16.status</b>	
Negative	43 (44%)
Positive	54 (56%)
<b>Treatment.Group</b>	
Definitive CRT or RT	69 (71%)
None (Declined Treatment)	1 (1.0%)
None (Hospice)	2 (2.1%)
Surgery + CRT or RT	24 (25%)
Surgery only	1 (1.0%)
<b>PFS.Event</b>	
No Progression	65 (67%)
Progression	32 (33%)
<b>OS.Event</b>	
Alive	81 (84%)
Deceased	16 (16%)
<b>OS.months</b>	
	22 (2 - 56)

<sup>1</sup>n (%); Median (Min - Max)

[Hide](#)

```
save_as_docx(fit1, path= "~/Downloads/1. CLIA HNSCC UNM Demographics Table.docx")
```

#Demographics Table by ctDNA at baseline

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]

circ_data_subset1 <- circ_data %>%
  select(
    Sex,
    Age,
    Tobacco.History,
    Prim.Location,
    cT,
    cN,
    cM,
    Histology,
    Stage,
    p16.status,
    Treatment.Group,
    PFS.Event,
    OS.Event,
    OS.months) %>%
  mutate(
    Sex = factor(Sex),
    Age = as.numeric(Age),
    Tobacco.History = factor(Tobacco.History),
    Prim.Location = factor(Prim.Location),
    cT = factor(cT),
    cN = factor(cN),
    cM = factor(cM),
    Histology = factor(Histology),
    Stage = factor(Stage),
    p16.status = factor(p16.status),
    Treatment.Group = factor(Treatment.Group),
    PFS.Event = factor(PFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Progression", "Progression")),
    OS.Event = factor(OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased")),
    OS.months = as.numeric(OS.months))

circ_data1 <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]

circ_data_subset2 <- circ_data1 %>%
  select(
    Sex,
    Age,
    Tobacco.History,
    Prim.Location,
    cT,
    cN,
    cM,
    Histology,
    Stage,
    p16.status,
    Treatment.Group,
    PFS.Event,
    OS.Event,
    OS.months,
    ctDNA.Base) %>%
  mutate(
    Sex = factor(Sex),
    Age = as.numeric(Age),
    Tobacco.History = factor(Tobacco.History),
    Prim.Location = factor(Prim.Location),
    cT = factor(cT),
    cN = factor(cN),
    cM = factor(cM),
    Histology = factor(Histology),
    Stage = factor(Stage),
    p16.status = factor(p16.status),
    Treatment.Group = factor(Treatment.Group),
    PFS.Event = factor(PFS.Event, levels = c("FALSE", "TRUE"), labels = c("No Progression", "Progression")),
    OS.Event = factor(OS.Event, levels = c("FALSE", "TRUE"), labels = c("Alive", "Deceased")),
    OS.months = as.numeric(OS.months),
    ctDNA.Base = factor(ctDNA.Base, levels = c("NEGATIVE", "POSITIVE"), labels = 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 = 97 <sup>1</sup>
<b>Sex</b>	
Female	17 (18%)
Male	80 (82%)
<b>Age</b>	66 (29 - 95)
<b>Tobacco.History</b>	63 (65%)
<b>Prim.Location</b>	
Larynx/Hypopharynx	5 (5.2%)
Oral cavity	16 (16%)
Oropharynx	67 (69%)
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)
<b>cT</b>	
T0	2 (2.1%)
T1	12 (12%)
T2	31 (32%)
T3	30 (31%)
T4	21 (22%)
TX	1 (1.0%)
<b>cN</b>	
N0	22 (23%)
N1	33 (34%)
N2	33 (34%)
N3	9 (9.3%)
<b>cM</b>	
M0	93 (96%)
M1	4 (4.1%)
<b>Histology</b>	
Adenosquamous carcinoma	1 (1.0%)
Basaloid squamous cell carcinoma	6 (6.2%)
Epithelial myoepithelial carcinoma	1 (1.0%)
Squamous cell carcinoma	86 (89%)
Undifferentiated carcinoma	3 (3.1%)
<b>Stage</b>	
I/II	49 (51%)
III/IVA/IVB	45 (46%)
IVC	3 (3.1%)
<b>p16.status</b>	
Negative	43 (44%)
Positive	54 (56%)

<sup>1</sup> n (%); Median (Min - Max)

Characteristic	N = 97 <sup>1</sup>
<b>Treatment.Group</b>	
Definitive CRT or RT	69 (71%)
None (Declined Treatment)	1 (1.0%)
None (Hospice)	2 (2.1%)
Surgery + CRT or RT	24 (25%)
Surgery only	1 (1.0%)
<b>PFS.Event</b>	
No Progression	65 (67%)
Progression	32 (33%)
<b>OS.Event</b>	
Alive	81 (84%)
Deceased	16 (16%)
<b>OS.months</b>	
	22 (2 - 56)

<sup>1</sup> n (%); Median (Min - Max)[Hide](#)

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

36 missing rows in the "ctDNA.Base" column have been removed.

[Hide](#)

ByctDNA\_MRD

Characteristic	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
<b>Sex</b>			0.10
Female	3 (43%)	8 (15%)	
Male	4 (57%)	47 (85%)	
<b>Age</b>	80 (53 - 95)	65 (37 - 95)	0.081
<b>Tobacco.History</b>	4 (57%)	37 (67%)	0.7
<b>Prim.Location</b>			0.037
Larynx/Hypopharynx	0 (0%)	3 (5.5%)	
Oral cavity	3 (43%)	4 (7.3%)	
Oropharynx	3 (43%)	44 (80%)	
Other (paranasal sinus and nasopharyngeal)	1 (14%)	4 (7.3%)	
<b>cT</b>			0.050
T0	0 (0%)	2 (3.6%)	
T1	1 (14%)	4 (7.3%)	

<sup>1</sup> n (%); Median (Min - Max)<sup>2</sup> Fisher's exact test; Wilcoxon rank sum test

Characteristic	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
T2	2 (29%)	19 (35%)	
T3	0 (0%)	21 (38%)	
T4	4 (57%)	9 (16%)	
TX	0 (0%)	0 (0%)	
<b>cN</b>			>0.9
N0	1 (14%)	11 (20%)	
N1	3 (43%)	19 (35%)	
N2	3 (43%)	19 (35%)	
N3	0 (0%)	6 (11%)	
<b>cM</b>			>0.9
M0	7 (100%)	53 (96%)	
M1	0 (0%)	2 (3.6%)	
<b>Histology</b>			0.14
Adenosquamous carcinoma	0 (0%)	0 (0%)	
Basaloid squamous cell carcinoma	0 (0%)	3 (5.5%)	
Epithelial myoepithelial carcinoma	0 (0%)	0 (0%)	
Squamous cell carcinoma	6 (86%)	52 (95%)	
Undifferentiated carcinoma	1 (14%)	0 (0%)	
<b>Stage</b>			0.4
I/II	2 (29%)	32 (58%)	
III/IVA/IVB	5 (71%)	21 (38%)	
IVC	0 (0%)	2 (3.6%)	
<b>p16.status</b>			0.090
Negative	5 (71%)	18 (33%)	
Positive	2 (29%)	37 (67%)	
<b>Treatment.Group</b>			0.2
Definitive CRT or RT	5 (71%)	51 (93%)	
None (Declined Treatment)	0 (0%)	0 (0%)	
None (Hospice)	0 (0%)	0 (0%)	
Surgery + CRT or RT	2 (29%)	3 (5.5%)	
Surgery only	0 (0%)	1 (1.8%)	
<b>PFS.Event</b>			0.4
No Progression	6 (86%)	35 (64%)	
Progression	1 (14%)	20 (36%)	
<b>OS.Event</b>			0.3
Alive	7 (100%)	44 (80%)	
Deceased	0 (0%)	11 (20%)	
<b>OS.months</b>	31 (21 - 39)	16 (2 - 45)	0.028

<sup>1</sup> n (%); Median (Min - Max)<sup>2</sup> Fisher's exact test; Wilcoxon rank sum test

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

Characteristic	Table 1		Table 2	
	N = 97 <sup>1</sup>	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
<b>Sex</b>				0.10
Female	17 (18%)	3 (43%)	8 (15%)	
Male	80 (82%)	4 (57%)	47 (85%)	
<b>Age</b>	66 (29 - 95)	80 (53 - 95)	65 (37 - 95)	0.081
<b>Tobacco.History</b>	63 (65%)	4 (57%)	37 (67%)	0.7
<b>Prim.Location</b>				0.037
Larynx/Hypopharynx	5 (5.2%)	0 (0%)	3 (5.5%)	
Oral cavity	16 (16%)	3 (43%)	4 (7.3%)	
Oropharynx	67 (69%)	3 (43%)	44 (80%)	
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)	1 (14%)	4 (7.3%)	
<b>cT</b>				0.050
T0	2 (2.1%)	0 (0%)	2 (3.6%)	
T1	12 (12%)	1 (14%)	4 (7.3%)	
T2	31 (32%)	2 (29%)	19 (35%)	
T3	30 (31%)	0 (0%)	21 (38%)	
T4	21 (22%)	4 (57%)	9 (16%)	
TX	1 (1.0%)	0 (0%)	0 (0%)	
<b>cN</b>				>0.9
N0	22 (23%)	1 (14%)	11 (20%)	
N1	33 (34%)	3 (43%)	19 (35%)	
N2	33 (34%)	3 (43%)	19 (35%)	
N3	9 (9.3%)	0 (0%)	6 (11%)	
<b>cM</b>				>0.9
M0	93 (96%)	7 (100%)	53 (96%)	
M1	4 (4.1%)	0 (0%)	2 (3.6%)	
<b>Histology</b>				0.14
Adenosquamous carcinoma	1 (1.0%)	0 (0%)	0 (0%)	
Basaloid squamous cell carcinoma	6 (6.2%)	0 (0%)	3 (5.5%)	
Epithelial myoepithelial carcinoma	1 (1.0%)	0 (0%)	0 (0%)	
Squamous cell carcinoma	86 (89%)	6 (86%)	52 (95%)	
Undifferentiated carcinoma	3 (3.1%)	1 (14%)	0 (0%)	
<b>Stage</b>				0.4
I/II	49 (51%)	2 (29%)	32 (58%)	
III/IVA/IVB	45 (46%)	5 (71%)	21 (38%)	
IVC	3 (3.1%)	0 (0%)	2 (3.6%)	

<sup>1</sup> n (%); Median (Min - Max)<sup>2</sup> Fisher's exact test; Wilcoxon rank sum test

Characteristic	Table 1		Table 2	
	N = 97 <sup>1</sup>	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
<b>p16.status</b>				
Negative	43 (44%)	5 (71%)	18 (33%)	0.090
Positive	54 (56%)	2 (29%)	37 (67%)	
<b>Treatment.Group</b>				
Definitive CRT or RT	69 (71%)	5 (71%)	51 (93%)	
None (Declined Treatment)	1 (1.0%)	0 (0%)	0 (0%)	
None (Hospice)	2 (2.1%)	0 (0%)	0 (0%)	
Surgery + CRT or RT	24 (25%)	2 (29%)	3 (5.5%)	
Surgery only	1 (1.0%)	0 (0%)	1 (1.8%)	
<b>PFS.Event</b>				
No Progression	65 (67%)	6 (86%)	35 (64%)	
Progression	32 (33%)	1 (14%)	20 (36%)	
<b>OS.Event</b>				
Alive	81 (84%)	7 (100%)	44 (80%)	
Deceased	16 (16%)	0 (0%)	11 (20%)	
<b>OS.months</b>	22 (2 - 56)	31 (21 - 39)	16 (2 - 45)	0.028

<sup>1</sup>n (%); Median (Min - Max)<sup>2</sup>Fisher's exact test; Wilcoxon rank sum test Hide

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

Characteristic	Table 1		Table 2	
	N = 97 <sup>1</sup>	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
<b>Sex</b>				
Female	17 (18%)	3 (43%)	8 (15%)	
Male	80 (82%)	4 (57%)	47 (85%)	
<b>Age</b>	66 (29 - 95)	80 (53 - 95)	65 (37 - 95)	0.081
<b>Tobacco.History</b>	63 (65%)	4 (57%)	37 (67%)	0.7
<b>Prim.Location</b>				
Larynx/Hypopharynx	5 (5.2%)	0 (0%)	3 (5.5%)	
Oral cavity	16 (16%)	3 (43%)	4 (7.3%)	
Oropharynx	67 (69%)	3 (43%)	44 (80%)	
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)	1 (14%)	4 (7.3%)	
<b>cT</b>				
T0	2 (2.1%)	0 (0%)	2 (3.6%)	
T1	12 (12%)	1 (14%)	4 (7.3%)	
T2	31 (32%)	2 (29%)	19 (35%)	
T3	30 (31%)	0 (0%)	21 (38%)	

<sup>1</sup>n (%); Median (Min - Max)<sup>2</sup>Fisher's exact test; Wilcoxon rank sum test

Characteristic	Table 1		Table 2	
	N = 97 <sup>1</sup>	Negative N = 7 <sup>1</sup>	Positive N = 55 <sup>1</sup>	p-value <sup>2</sup>
T4	21 (22%)	4 (57%)	9 (16%)	
TX	1 (1.0%)	0 (0%)	0 (0%)	
<b>cN</b>				>0.9
N0	22 (23%)	1 (14%)	11 (20%)	
N1	33 (34%)	3 (43%)	19 (35%)	
N2	33 (34%)	3 (43%)	19 (35%)	
N3	9 (9.3%)	0 (0%)	6 (11%)	
<b>cM</b>				>0.9
M0	93 (96%)	7 (100%)	53 (96%)	
M1	4 (4.1%)	0 (0%)	2 (3.6%)	
<b>Histology</b>				0.14
Adenosquamous carcinoma	1 (1.0%)	0 (0%)	0 (0%)	
Basaloid squamous cell carcinoma	6 (6.2%)	0 (0%)	3 (5.5%)	
Epithelial myoepithelial carcinoma	1 (1.0%)	0 (0%)	0 (0%)	
Squamous cell carcinoma	86 (89%)	6 (86%)	52 (95%)	
Undifferentiated carcinoma	3 (3.1%)	1 (14%)	0 (0%)	
<b>Stage</b>				0.4
I/II	49 (51%)	2 (29%)	32 (58%)	
III/IVA/IVB	45 (46%)	5 (71%)	21 (38%)	
IVC	3 (3.1%)	0 (0%)	2 (3.6%)	
<b>p16.status</b>				0.090
Negative	43 (44%)	5 (71%)	18 (33%)	
Positive	54 (56%)	2 (29%)	37 (67%)	
<b>Treatment.Group</b>				0.2
Definitive CRT or RT	69 (71%)	5 (71%)	51 (93%)	
None (Declined Treatment)	1 (1.0%)	0 (0%)	0 (0%)	
None (Hospice)	2 (2.1%)	0 (0%)	0 (0%)	
Surgery + CRT or RT	24 (25%)	2 (29%)	3 (5.5%)	
Surgery only	1 (1.0%)	0 (0%)	1 (1.8%)	
<b>PFS.Event</b>				0.4
No Progression	65 (67%)	6 (86%)	35 (64%)	
Progression	32 (33%)	1 (14%)	20 (36%)	
<b>OS.Event</b>				0.3
Alive	81 (84%)	7 (100%)	44 (80%)	
Deceased	16 (16%)	0 (0%)	11 (20%)	
<b>OS.months</b>	22 (2 - 56)	31 (21 - 39)	16 (2 - 45)	0.028

<sup>1</sup>n (%); Median (Min - Max)<sup>2</sup>Fisher's exact test; Wilcoxon rank sum test

save\_as\_docx(fit1, path = "~/Downloads/1b. CLIA HNSCC UNM Demographics Table by ctDNA.docx")

#Overview plot by Stage

```

setwd("~/Downloads")
clinstage <- read.csv("CLIA HNSCC UNM_OP.csv")
clinstage_df <- as.data.frame(clinstage)

# Creating the basic swimmer plot
oplot <- swimmer_plot(df=clinstage_df,
                      id='PatientName',
                      end='fu.diff.months',
                      fill='gray',
                      width=.01,
                      base_size = 14,
                      stratify= c('Stage'))

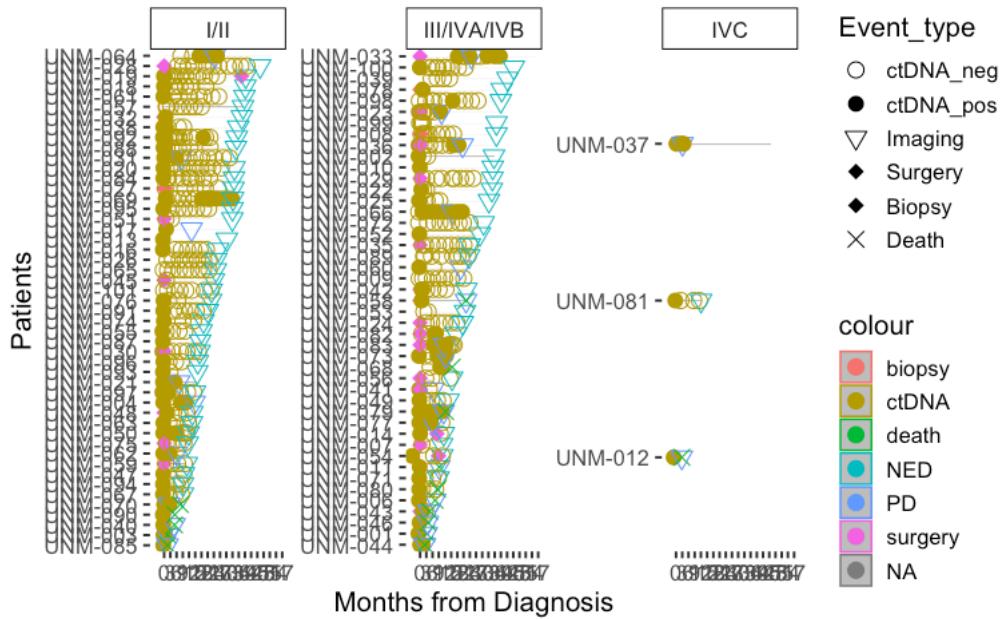
# Adding themes and scales
oplot <- oplot + theme(panel.border = element_blank())
oplot <- oplot + scale_y_continuous(breaks = seq(0, 72, by = 3))
oplot <- oplot + labs(x = "Patients", y="Months from Diagnosis")

# Adding swimmer points
oplot_ev1 <- oplot + swimmer_points(df_points=clinstage_df,
                                      id='PatientName',
                                      time='date.diff.months',
                                      name_shape ='Event_type',
                                      name_col = 'Event',
                                      size=3.5,fill='black')
# Optionally uncomment and use col='darkgreen' if needed

# Adding shape manual scale
oplot_ev1.1 <- oplot_ev1 + ggplot2::scale_shape_manual(name="Event_type",
                                                       values=c(1,16,6,18,18,4),
                                                       breaks=c('ctDNA_neg','ctDNA_pos', 'Imaging','Surgery','Biopsy', 'Death'))

# Display the plot
oplot_ev1.1

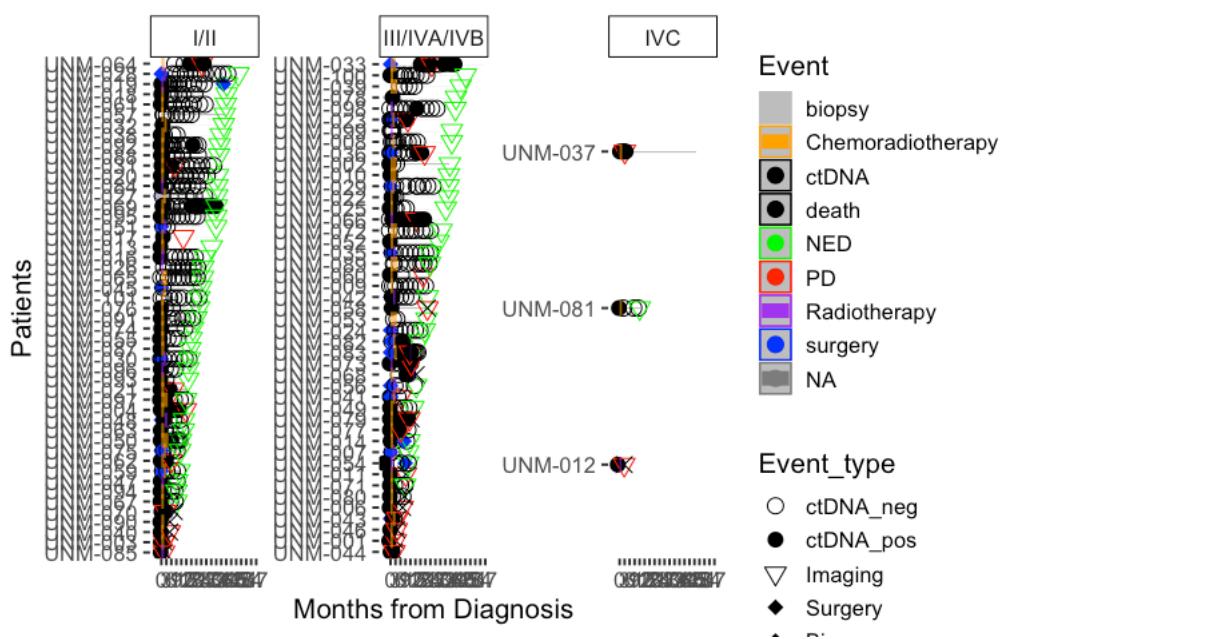
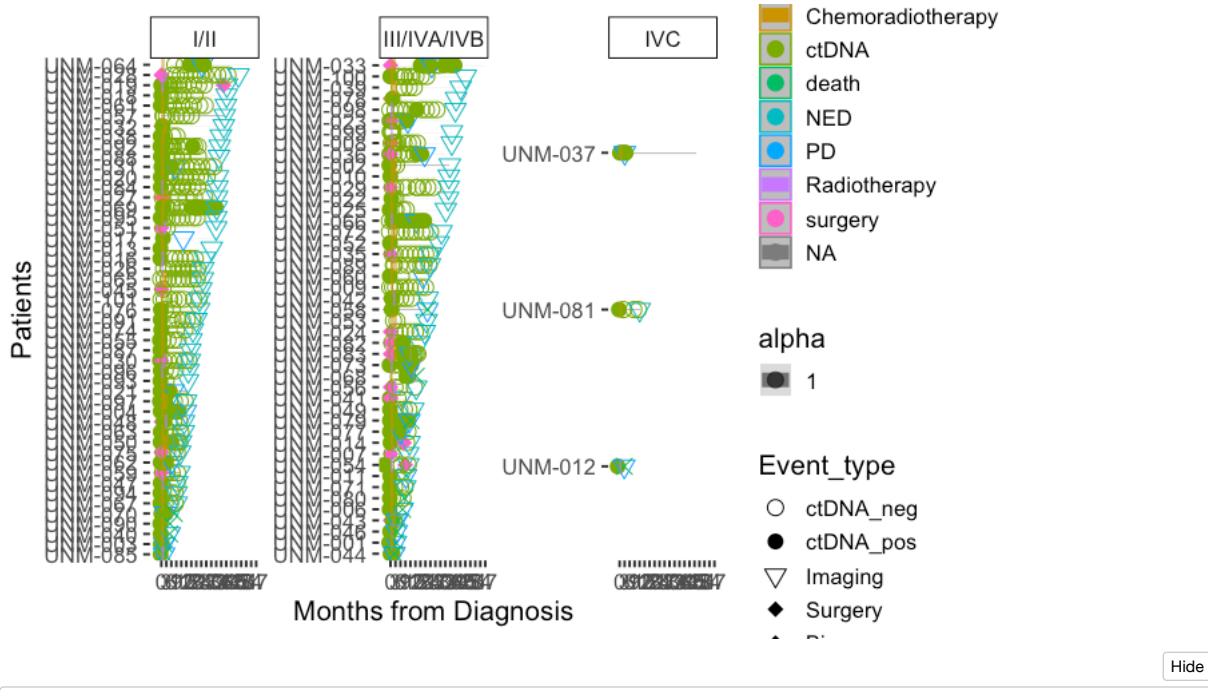
```



```

oplot_ev2 <- oplot_ev1.1 + swimmer_lines(df_lines=clinstage_df,
                                           id='PatientName',
                                           start='Tx_start.months',
                                           end='Tx_end.months',
                                           name_col='Tx_type',
                                           size=3.5,
                                           name_alpha = 1.0)
oplot_ev2 <- oplot_ev2 + guides(linetype = guide_legend(override.aes = list(size = 5, color = "black")))
oplot_ev2

```



```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]

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

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

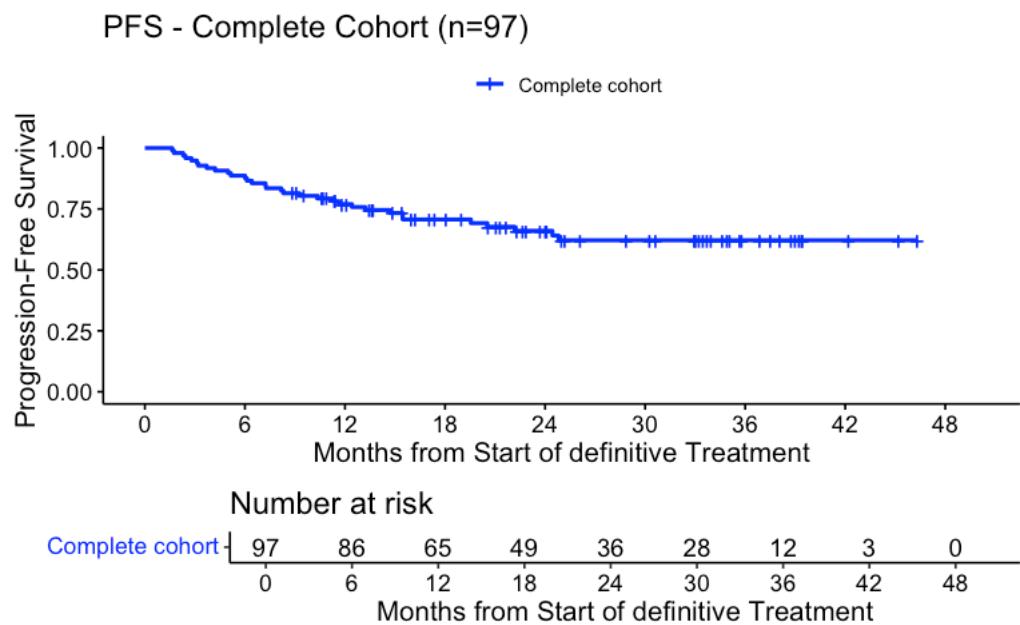
```
n events median 0.95LCL 0.95UCL
[1,] 97 32 NA NA NA
```

Hide

```

surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.available, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue"), title="PFS - Complete Cohort (n=97)", ylab= "Progression-Free Survival", xlab="Months from Start of
definitive Treatment", legend.labs=c("Complete cohort"), legend.title="")

```



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

```

Call: survfit(formula = surv_object ~ ctDNA.available, 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     65      22    0.770  0.0432      0.672    0.842
24     36       8    0.660  0.0518      0.548    0.751
36     12       2    0.622  0.0556      0.503    0.720

```

#OS in Complete Cohort (N=97)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]

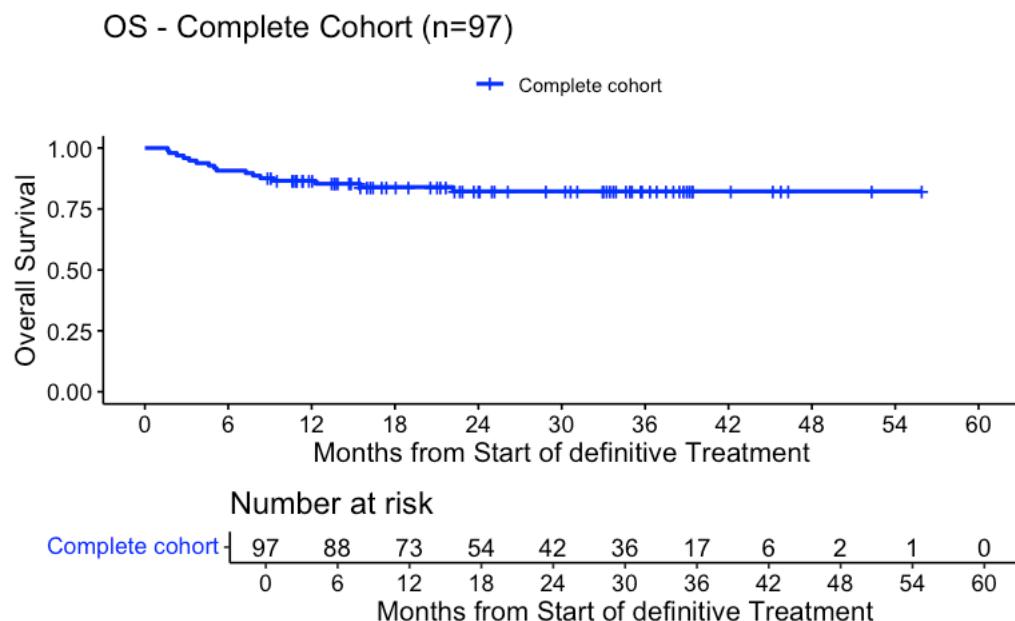
survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.available, data = circ_data)
```

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

```
  n events median 0.95LCL 0.95UCL
[1,] 97      16      NA      NA      NA
```

[Hide](#)

```
surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.available, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue"), title="OS - Complete Cohort (n=97)", ylab= "Overall Survival", xlab="Months from Start of definitive
Treatment", legend.labs=c("Complete cohort"), legend.title="")
```



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

[Hide](#)

```
Call: survfit(formula = surv_object ~ ctDNA.available, 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	73	13	0.866	0.0347	0.780		0.920	
24	42	3	0.822	0.0412	0.724		0.888	
36	17	0	0.822	0.0412	0.724		0.888	

#Association of Baseline ctDNA MTM levels with clinicopathological factors

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_datadf <- as.data.frame(circ_data)

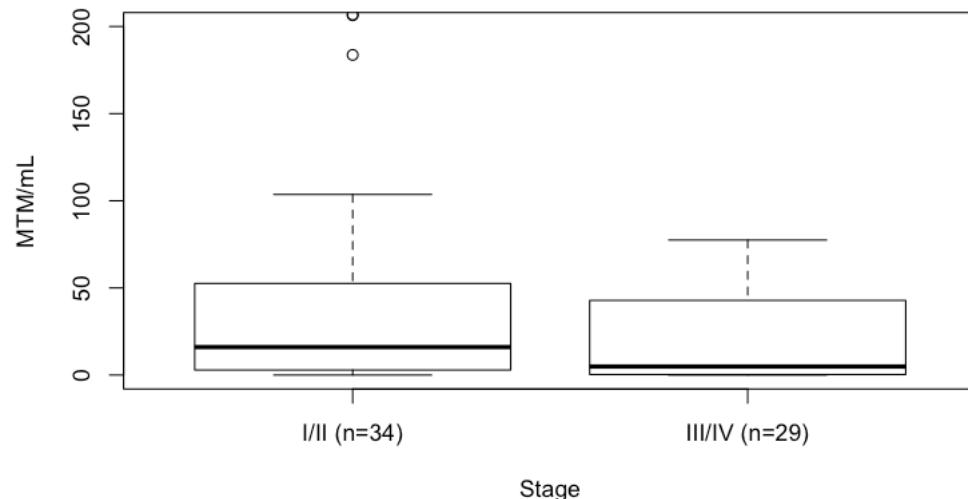
tally(~cStage, data=circ_data, margins = TRUE)
```

cStage	I/II	III/IV	Total
	34	28	62

[Hide](#)

```
circ_data$cStage <- factor(circ_data$cStage, levels = c("I/II","III/IV"), labels = c("I/II (n=34)","III/IV (n=29)"))
boxplot(ctDNA.Base.MTM~cStage, data=circ_data, main="ctDNA pre-treatment MTM - Stage", xlab="Stage", ylab="MTM/mL", col="white",border="black", ylim = c(0, 200))
```

### ctDNA pre-treatment MTM - Stage



```
median_ctDNA.Stage <- circ_data %>%
  group_by(cStage) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.Stage)
```

cStage	median_ctDNA_Base_MTM
<fctr>	<dbl>
I/II (n=34)	16.03
III/IV (n=29)	4.90
2 rows	

[Hide](#)

```
m1<-wilcox.test(ctDNA.Base.MTM ~ cStage, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m1)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: ctDNA.Base.MTM by cStage
W = 590, p-value = 0.1081
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-0.7499638 20.1599876
sample estimates:
difference in location
3.690372
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_datadf <- as.data.frame(circ_data)

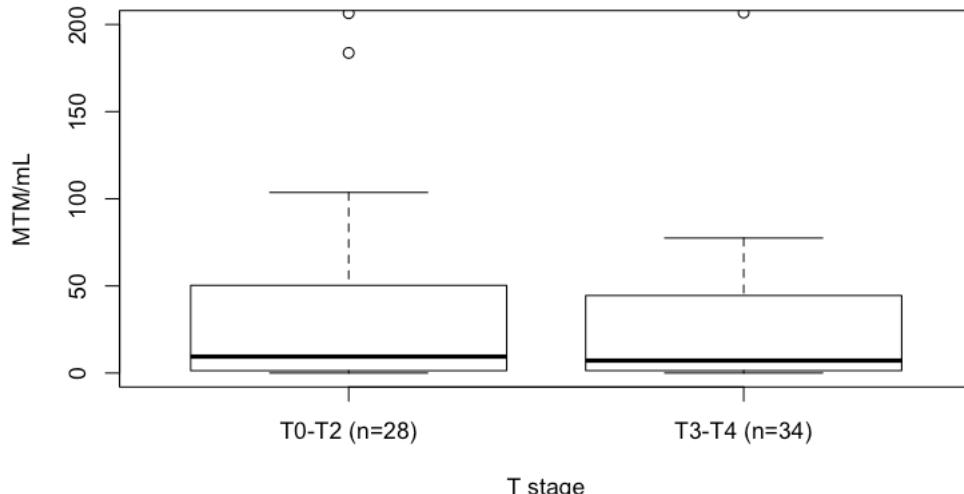
tally(~cT.status, data=circ_data, margins = TRUE)
```

```
cT.status
T0-T2 T3-T4 Total
 28     34    62
```

[Hide](#)

```
circ_data$cT.status <- factor(circ_data$cT.status, levels = c("T0-T2","T3-T4"), labels = c("T0-T2 (n=28)","T3-T4 (n=34)"))
boxplot(ctDNA.Base.MTM~cT.status, data=circ_data, main="ctDNA pre-treatment MTM - T stage", xlab="T stage", ylab ="MTM/mL", col="white",border="black", ylim = c(0, 200))
```

### ctDNA pre-treatment MTM - T stage



[Hide](#)

```
median_ctDNA.cT <- circ_data %>%
  group_by(cT.status) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.cT)
```

cT.status	median_ctDNA_Base_MTM
<fctr>	<dbl>
T0-T2 (n=28)	9.430
T3-T4 (n=34)	7.155
2 rows	

[Hide](#)

```
m2<-wilcox.test(ctDNA.Base.MTM ~ cT.status, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m2)
```

```
Wilcoxon rank sum test with continuity correction

data: ctDNA.Base.MTM by cT.status
W = 466, p-value = 0.893
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-7.789979 7.539938
sample estimates:
difference in location
-0.1111266
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_datadf <- as.data.frame(circ_data)

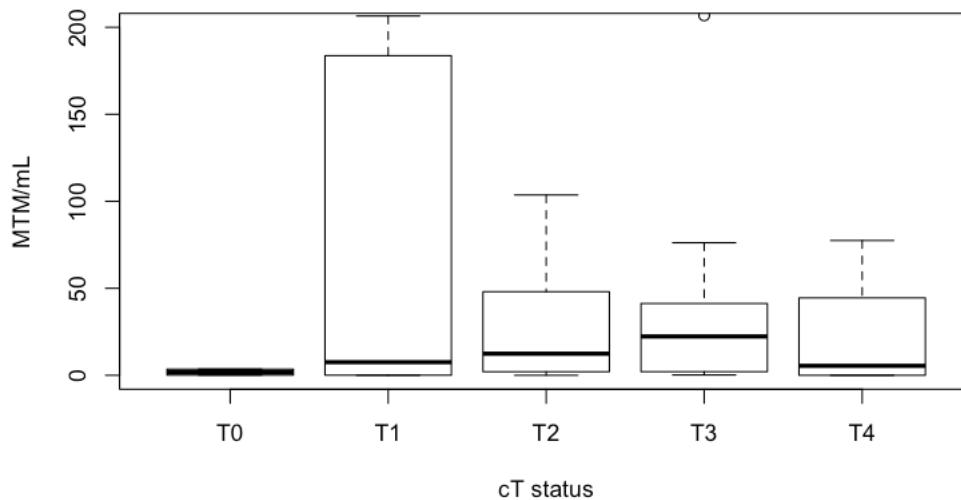
tally(~cT, data=circ_data, margins = TRUE)
```

cT	T0	T1	T2	T3	T4	Total
	2	5	21	21	13	62

Hide

```
circ_data$cT <- factor(circ_data$cT, levels = c("T0","T1","T2","T3","T4"))
boxplot(ctDNA.Base.MTM~cT, data=circ_data, main="ctDNA pre-treatment MTM - cT status", xlab="cT status", ylab="MTM/mL", col="white", border="black", ylim = c(0, 200))
```

### ctDNA pre-treatment MTM - cT status



Hide

```
median_ctDNA.cT <- circ_data %>%
  group_by(cT) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.cT)
```

cT	median_ctDNA_Base_MTM
<fctr>	<dbl>
T0	1.815
T1	7.540
T2	12.400

cT	median_ctDNA_Base_MTM
<fctr>	<dbl>
T3	22.320
T4	5.430

5 rows

```
pairwise_wilcox <- pairwise.wilcox.test(circ_data$ctDNA.Base.MTM, circ_data$cT,
                                         p.adjust.method = "none",
                                         exact = FALSE)
print(pairwise_wilcox)
```

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: circ\_data\$ctDNA.Base.MTM and circ\_data\$cT

	T0	T1	T2	T3
T1	0.85	-	-	-
T2	0.21	0.85	-	-
T3	0.14	0.65	0.69	-
T4	0.55	0.62	0.36	0.23

P value adjustment method: none

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base != "",]
circ_data$cT <- factor(circ_data$cT, levels = c("T0", "T1", "T2", "T3", "T4"))
circ_data$ctDNA.Base.MTM <- as.numeric(circ_data$ctDNA.Base.MTM)
cT_levels <- levels(circ_data$cT)
p_value_matrix <- matrix(NA, nrow = length(cT_levels), ncol = length(cT_levels))
rownames(p_value_matrix) <- cT_levels
colnames(p_value_matrix) <- cT_levels

for (i in 1:length(cT_levels)) {
  for (j in i:length(cT_levels)) {
    if (i != j) {
      # Extract data for both groups
      data1 <- circ_data %>% filter(cT == cT_levels[i]) %>% pull(ctDNA.Base.MTM)
      data2 <- circ_data %>% filter(cT == cT_levels[j]) %>% pull(ctDNA.Base.MTM)

      # Perform Wilcoxon test and store p-value
      test_result <- wilcox.test(data1, data2, exact = FALSE)
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Make symmetric
    } else {
      p_value_matrix[i, j] <- 1 # Self-comparison = 1
    }
  }
}

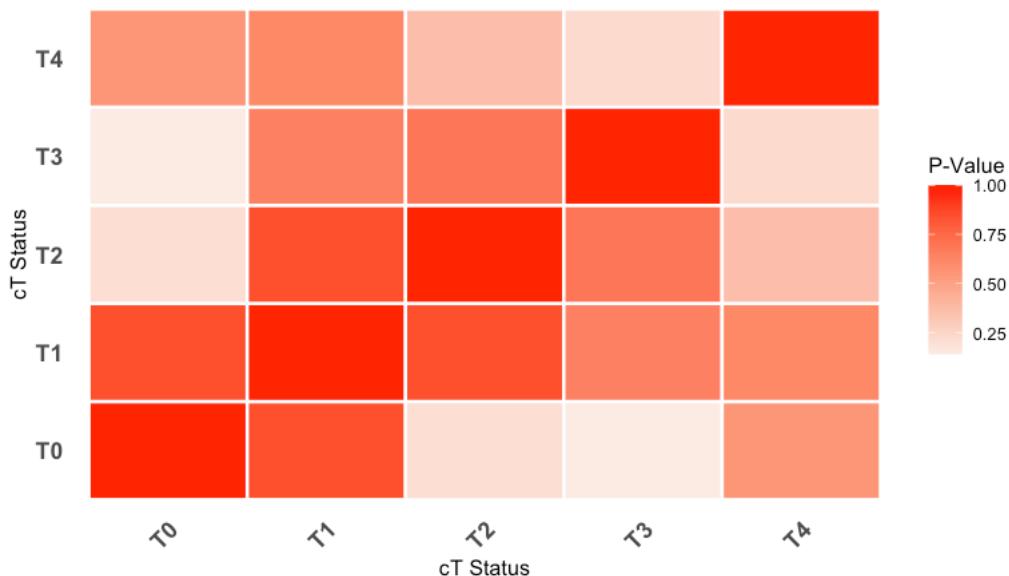
p_value_matrix[is.na(p_value_matrix)] <- 1.00
p_value_data <- melt(p_value_matrix)
colnames(p_value_data) <- c("cT1", "cT2", "p_value")
p_value_data <- p_value_data %>%
  mutate(
    significance = case_when(
      p_value < 0.001 ~ "***",
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    )
  )

ggplot(p_value_data, aes(x = cT1, y = cT2, fill = p_value)) +
  geom_tile(color = "white", size = 0.8) + # Thicker grid lines for separation
  geom_text(aes(label = significance), color = "black", size = 6, fontface = "bold") + # Significance markers
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0.05) + # Gradient colors
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12, face = "bold"),
        axis.text.y = element_text(size = 12, face = "bold"),
        panel.grid = element_blank())
  labs(title = "Pairwise Wilcoxon-Test P-Values (ctDNA.Base.MTM by cT)",
       x = "cT Status", y = "cT Status", fill = "P-Value")

```

G2; H2;Warning h: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.  
 i Please use `linewidth` instead.  
 This warning is displayed once every 8 hours.  
 Call `lifecycle::last\_lifecycle\_`8;`ide:run:warnings() warnings()`8;; ` to see where this warning was generated.  
 d. g

## Pairwise Wilcoxon-Test P-Values (ctDNA.Base.MTM by cT)



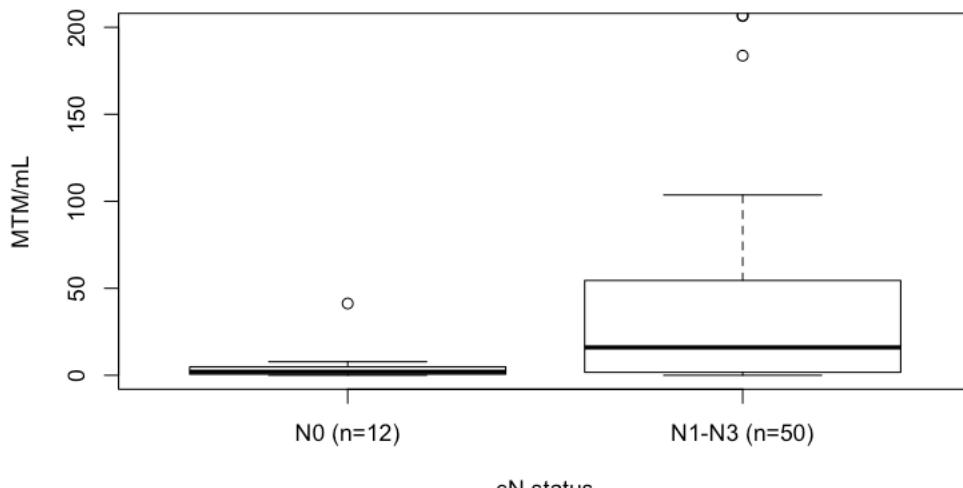
```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_dataadf <- as.data.frame(circ_data)

tally(~cN.status, data=circ_data, margins = TRUE)
```

cN.status	N0	N1-N3	Total
	12	50	62

```
circ_data$cN.status <- factor(circ_data$cN.status, levels = c("N0","N1-N3"), labels = c("N0 (n=12)","N1-N3 (n=50)"))
boxplot(ctDNA.Base.MTM~cN.status, data=circ_data, main="ctDNA pre-treatment MTM - cN status", xlab="cN status", ylab="MTM/mL", col="white",border="black", ylim = c(0, 200))
```

## ctDNA pre-treatment MTM - cN status



```
median_ctDNA.cN <- circ_data %>%
  group_by(cN.status) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.cN)
```

cN.status	median_ctDNA_Base_MTM
<fctr>	<dbl>
N0 (n=12)	2.065
N1-N3 (n=50)	16.030
2 rows	

Hide

```
m3<-wilcox.test(ctDNA.Base.MTM ~ cN.status, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m3)
```

Wilcoxon rank sum test with continuity correction

```
data: ctDNA.Base.MTM by cN.status
W = 162, p-value = 0.01422
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-42.439959 -1.050045
sample estimates:
difference in location
-11.19909
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!="",]
circ_datadf <- as.data.frame(circ_data)

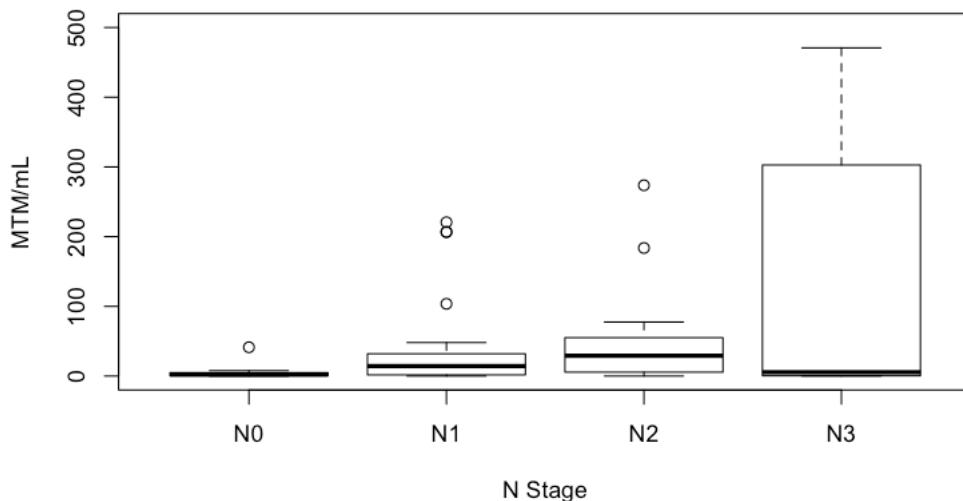
tally(~cN, data=circ_data, margins = TRUE)
```

cN	N0	N1	N2	N3	Total
	12	22	22	6	62

Hide

```
circ_data$cN <- factor(circ_data$cN, levels = c("N0","N1","N2","N3"))
boxplot(ctDNA.Base.MTM~cN, data=circ_data, main="ctDNA pre-treatment MTM - N Stage", xlab="N Stage", ylab="MTM/mL", col="white",border="black", ylim = c(0, 500))
```

### ctDNA pre-treatment MTM - N Stage



Hide

```
median_ctDNA.cN <- circ_data %>%
  group_by(cN) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.cN)
```

cN	median_ctDNA_Base_MTM
<fctr>	<dbl>
N0	2.065
N1	14.340
N2	29.110
N3	5.550
4 rows	

Hide

```
pairwise_wilcox <- pairwise.wilcox.test(circ_data$ctDNA.Base.MTM, circ_data$cN,
                                         p.adjust.method = "none",
                                         exact = FALSE)
print(pairwise_wilcox)
```

```
Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: circ_data$ctDNA.Base.MTM and circ_data$cN

  N0     N1     N2
N1 0.0473 -    
N2 0.0094 0.4108 -
N3 0.3736 0.9777 0.7580

P value adjustment method: none
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base != "",]
circ_data$cN <- factor(circ_data$cN, levels = c("N0", "N1", "N2", "N3"))
circ_data$ctDNA.Base.MTM <- as.numeric(circ_data$ctDNA.Base.MTM)
cN_levels <- levels(circ_data$cN)
p_value_matrix <- matrix(NA, nrow = length(cN_levels), ncol = length(cN_levels))
rownames(p_value_matrix) <- cN_levels
colnames(p_value_matrix) <- cN_levels

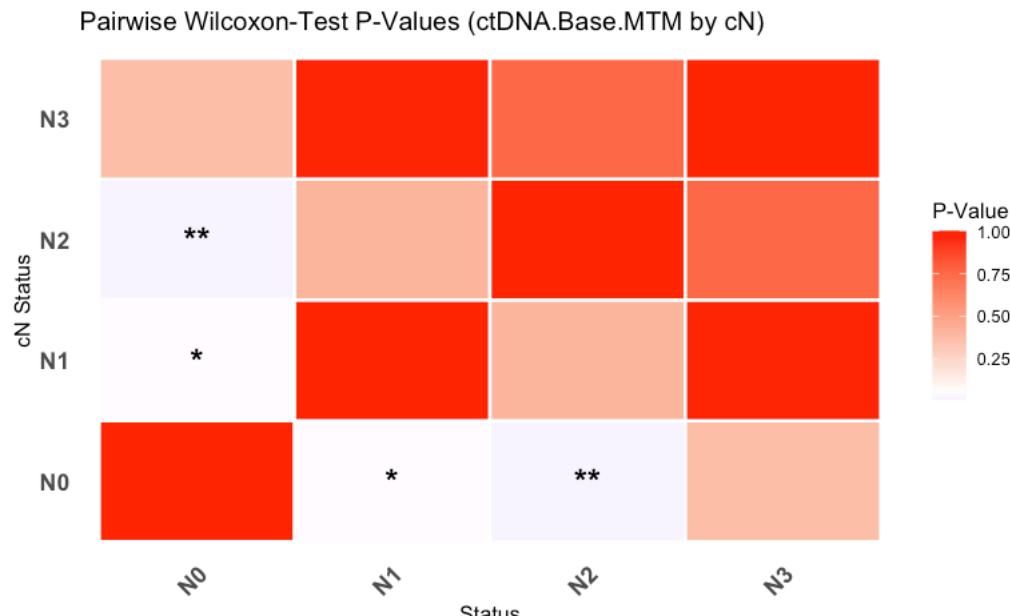
for (i in 1:length(cN_levels)) {
  for (j in i:length(cN_levels)) {
    if (i != j) {
      # Extract data for both groups
      data1 <- circ_data %>% filter(cN == cN_levels[i]) %>% pull(ctDNA.Base.MTM)
      data2 <- circ_data %>% filter(cN == cN_levels[j]) %>% pull(ctDNA.Base.MTM)

      # Perform Wilcoxon test and store p-value
      test_result <- wilcox.test(data1, data2, exact = FALSE)
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Make symmetric
    } else {
      p_value_matrix[i, j] <- 1 # Self-comparison = 1
    }
  }
}

p_value_matrix[is.na(p_value_matrix)] <- 1.00
p_value_data <- melt(p_value_matrix)
colnames(p_value_data) <- c("cN1", "cN2", "p_value")
p_value_data <- p_value_data %>%
  mutate(
    significance = case_when(
      p_value < 0.001 ~ "***",
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    )
  )

ggplot(p_value_data, aes(x = cN1, y = cN2, fill = p_value)) +
  geom_tile(color = "white", size = 0.8) + # Thicker grid lines for separation
  geom_text(aes(label = significance), color = "black", size = 6, fontface = "bold") + # Significance markers
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0.05) + # Gradient colors
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12, face = "bold"),
        axis.text.y = element_text(size = 12, face = "bold"),
        panel.grid = element_blank()) +
  labs(title = "Pairwise Wilcoxon-Test P-Values (ctDNA.Base.MTM by cN)",
       x = "Status", y = "cN Status", fill = "P-Value")

```



Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base != "",]
circ_data$cT <- factor(circ_data$cT, levels = c("T0", "T1", "T2", "T3", "T4"))
circ_data$cN <- factor(circ_data$cN, levels = c("N0", "N1", "N2", "N3"))
circ_data$ctDNA.Base.MTM <- as.numeric(circ_data$ctDNA.Base.MTM)

median_ctDNA <- circ_data %>%
  group_by(cT, cN) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE)) %>%
  ungroup()
```

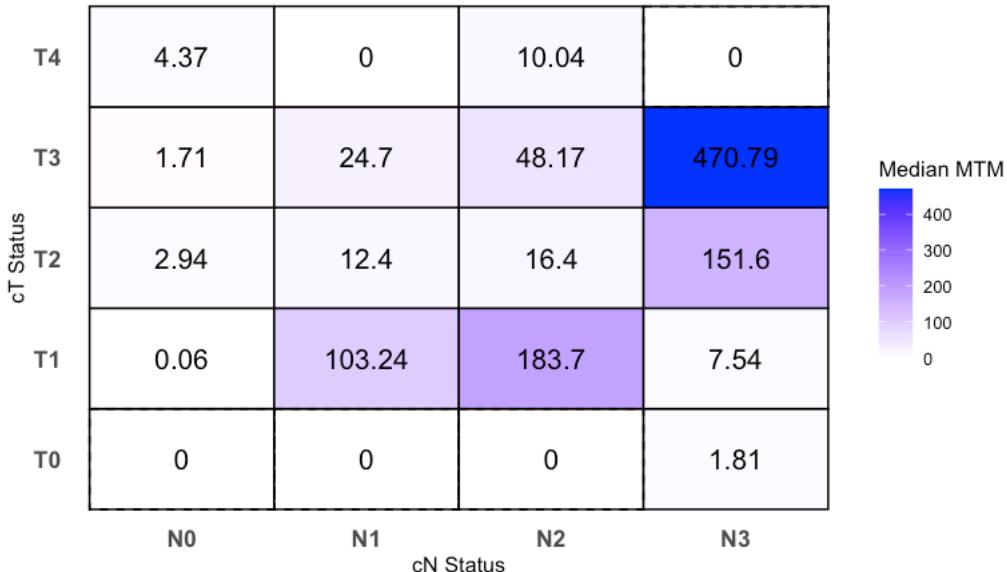
`summarise()` has grouped output by 'cT'. You can override using the ``.groups` argument.

Hide

```
p_value_matrix <- dcast(median_ctDNA, cT ~ cN, value.var = "median_ctDNA_Base_MTM")
p_value_data <- melt(p_value_matrix, id.vars = "cT", variable.name = "cN", value.name = "median_value")
p_value_data$missing <- ifelse(is.na(p_value_data$median_value), "Missing", "Present")
p_value_data$median_value[is.na(p_value_data$median_value)] <- 0

ggplot(p_value_data, aes(x = cN, y = cT, fill = median_value)) +
  geom_tile(color = "black", size = 0.5) + # Black gridlines for separation
  geom_text(aes(label = round(median_value, 2)), color = "black", size = 5) + # Display median values
  scale_fill_gradient(low = "white", high = "blue") + # Color gradient similar to the reference image
  theme_minimal() +
  theme(axis.text.x = element_text(size = 12, face = "bold"),
        axis.text.y = element_text(size = 12, face = "bold"),
        panel.grid = element_blank()) +
  labs(title = "Median ctDNA.Base.MTM by cT and cN",
       x = "cN Status", y = "cT Status", fill = "Median MTM") +
  geom_tile(data = subset(p_value_data, missing == "Missing"),
            aes(x = cN, y = cT), color = "black", fill = NA, size = 0.5, linetype = "dashed") # Add diagonal crosses for missing values
```

Median ctDNA.Base.MTM by cT and cN



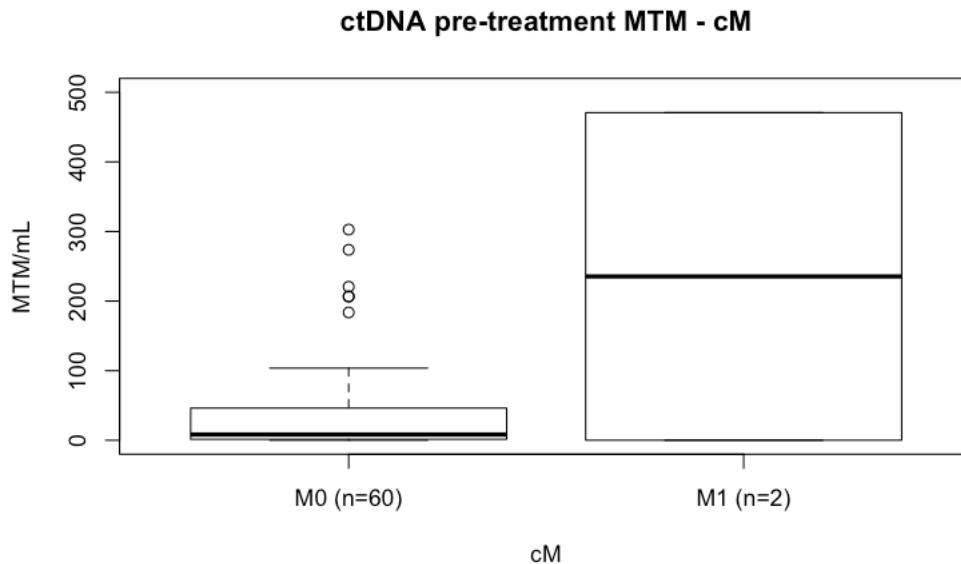
Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!="",]
circ_datadf <- as.data.frame(circ_data)

tally(~cM, data=circ_data, margins = TRUE)
```

cM	M0	M1	Total
	60	2	62

```
circ_data$cM <- factor(circ_data$cM, levels = c("M0","M1"), labels = c("M0 (n=60)","M1 (n=2)"))
boxplot(ctDNA.Base.MTM~cM, data=circ_data, main="ctDNA pre-treatment MTM - cM", xlab="cM", ylab="MTM/mL", col="white", border="black", ylim = c(0, 500))
```



```
median_ctDNA.cM <- circ_data %>%
  group_by(cM) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.cM)
```

cM	median_ctDNA_Base_MTM
<fctr>	<dbl>
M0 (n=60)	8.315
M1 (n=2)	235.430
2 rows	

```
m4<-wilcox.test(ctDNA.Base.MTM ~ cM, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m4)
```

```
Wilcoxon rank sum test with continuity correction

data: ctDNA.Base.MTM by cM
W = 53, p-value = 0.7955
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-469.98001 76.11002
sample estimates:
difference in location
-0.07005722
```

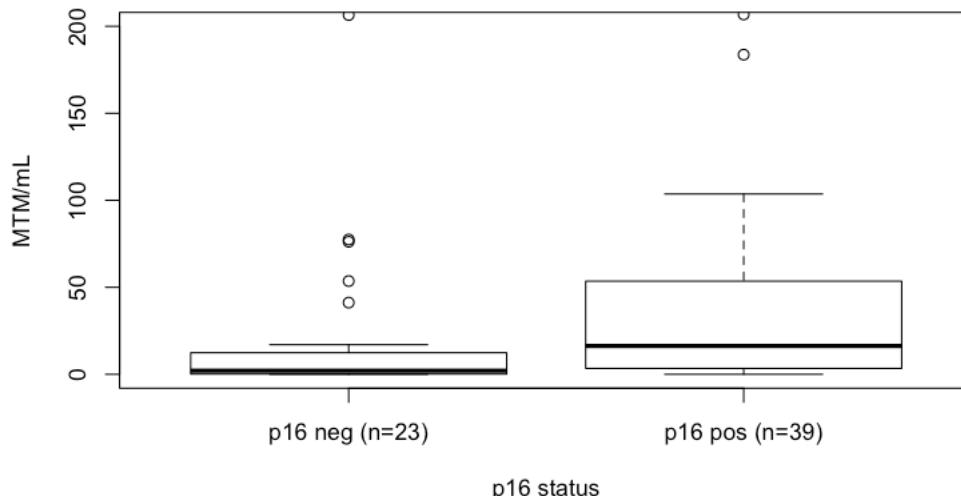
```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_dataadf <- as.data.frame(circ_data)

tally(~p16.status, data=circ_data, margins = TRUE)
```

p16.status	Negative	Positive	Total
	23	39	62

Hide

```
circ_data$p16.status <- factor(circ_data$p16.status, levels = c("Negative","Positive"), labels = c("p16 neg (n=23)","p16 pos (n=39)"))
boxplot(ctDNA.Base.MTM~p16.status, data=circ_data, main="ctDNA pre-treatment MTM - p16 status", xlab="p16 status", ylab="MTM/mL", col="white",border="black", ylim = c(0, 200))
```

**ctDNA pre-treatment MTM - p16 status**

Hide

```
median_ctDNA.p16 <- circ_data %>%
  group_by(p16.status) %>%
  summarise(median_ctDNA_Base_MTM = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.p16)
```

p16.status	median_ctDNA_Base_MTM
<fctr>	<dbl>
p16 neg (n=23)	2.08
p16 pos (n=39)	16.28
2 rows	

Hide

```
m5<-wilcox.test(ctDNA.Base.MTM ~ p16.status, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m5)
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: ctDNA.Base.MTM by p16.status
W = 269, p-value = 0.009047
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
-27.889950 -1.040095
sample estimates:
difference in location
-8.739937
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_datadf <- as.data.frame(circ_data)

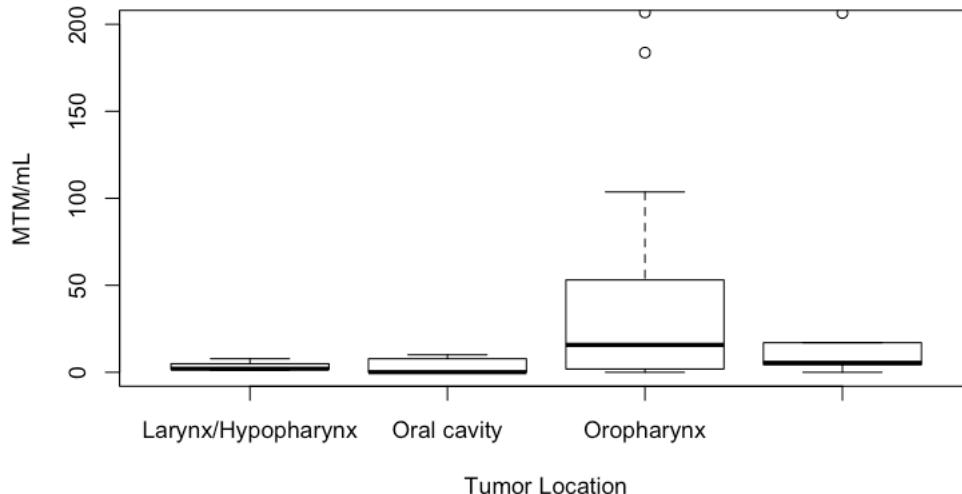
tally(~Prim.Location, data=circ_data, margins = TRUE)
```

Prim.Location	Larynx/Hypopharynx	Oral cavity
Oropharynx	3	7
Other (paranasal sinus and nasopharyngeal)	5	
47	62	
Total		

[Hide](#)

```
circ_data$Prim.Location <- factor(circ_data$Prim.Location, levels = c("Larynx/Hypopharynx", "Oral cavity", "Oropharynx", "Other (paranasal sinus and nasopharyngeal)"))
boxplot(ctDNA.Base.MTM~Prim.Location, data=circ_data, main="ctDNA pre-treatment MTM - Tumor Location", xlab="Tumor Location", ylab="MTM/mL", col="white", border="black", ylim = c(0, 200))
```

### ctDNA pre-treatment MTM - Tumor Location



[Hide](#)

```
median_ctDNA.loc <- circ_data %>%
  group_by(Prim.Location) %>%
  summarise(median_ctDNA_Base_MTm = median(ctDNA.Base.MTM, na.rm = TRUE))
print(median_ctDNA.loc)
```

Prim.Location	median_ctDNA_Base_MTm
<fctr>	<dbl>
Larynx/Hypopharynx	2.08
Oral cavity	0.06

Prim.Location	median_ctDNA_Base_MTM
<fctr>	<dbl>
Oropharynx	15.78
Other (paranasal sinus and nasopharyngeal)	5.43
4 rows	

[Hide](#)

```
pairwise_wilcox <- pairwise.wilcox.test(circ_data$ctDNA.Base.MTM, circ_data$Prim.Location,
                                         p.adjust.method = "none",
                                         exact = FALSE)

print(pairwise_wilcox)
```

```
Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: circ_data$ctDNA.Base.MTM and circ_data$Prim.Location

Larynx/Hypopharynx Oral cavity Oropharynx
Oral cavity 0.644 - -
Oropharynx 0.253 0.065 -
Other (paranasal sinus and nasopharyngeal) 0.551 0.563 0.653

P value adjustment method: none
```

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available == "TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base != "",]
circ_data$Prim.Location <- factor(circ_data$Prim.Location, levels = c("Larynx/Hypopharynx", "Oral cavity", "Oropharynx", "Other (paranasal sinus and nasopharyngeal)"), labels = c("LRX/HPRX", "OC", "PRX", "Other"))
circ_data$ctDNA.Base.MTM <- as.numeric(circ_data$ctDNA.Base.MTM)
pl_levels <- levels(circ_data$Prim.Location)
p_value_matrix <- matrix(NA, nrow = length(pl_levels), ncol = length(pl_levels))
rownames(p_value_matrix) <- pl_levels
colnames(p_value_matrix) <- pl_levels

for (i in 1:length(pl_levels)) {
  for (j in i:length(pl_levels)) {
    if (i != j) {
      # Extract data for both groups
      data1 <- circ_data %>% filter(Prim.Location == pl_levels[i]) %>% pull(ctDNA.Base.MTM)
      data2 <- circ_data %>% filter(Prim.Location == pl_levels[j]) %>% pull(ctDNA.Base.MTM)

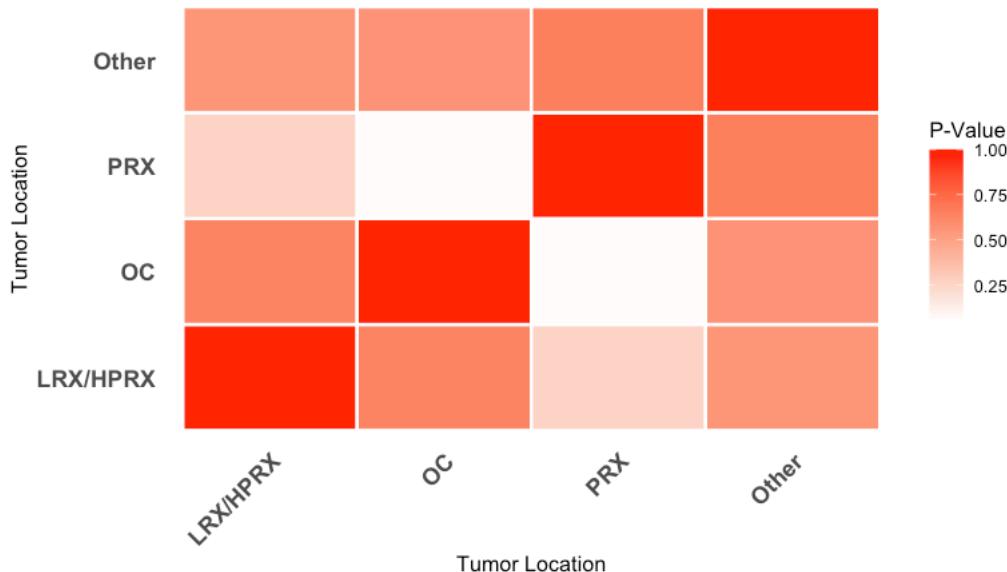
      # Perform Wilcoxon test and store p-value
      test_result <- wilcox.test(data1, data2, exact = FALSE)
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Make symmetric
    } else {
      p_value_matrix[i, j] <- 1 # Self-comparison = 1
    }
  }
}

p_value_matrix[is.na(p_value_matrix)] <- 1.00
p_value_data <- melt(p_value_matrix)
colnames(p_value_data) <- c("pl1", "pl2", "p_value")
p_value_data <- p_value_data %>%
  mutate(
    significance = case_when(
      p_value < 0.001 ~ "***",
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    )
  )

ggplot(p_value_data, aes(x = pl1, y = pl2, fill = p_value)) +
  geom_tile(color = "white", size = 0.8) + # Thicker grid lines for separation
  geom_text(aes(label = significance), color = "black", size = 6, fontface = "bold") + # Significance markers
  scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0.05) + # Gradient colors
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 12, face = "bold"),
        axis.text.y = element_text(size = 12, face = "bold"),
        panel.grid = element_blank()) +
  labs(title = "Pairwise Wilcoxon-Test P-Values (ctDNA.Base.MTM by Tumor Location)",
       x = "Tumor Location", y = "Tumor Location", fill = "P-Value")

```

## Pairwise Wilcoxon-Test P-Values (ctDNA.Base.MTM by Tumor Location)



#PFS by ctDNA status at MRD

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	56	7	NA	NA	NA
ctDNA.MRD=POSITIVE	13	8	15.5	4.21	NA

Hide

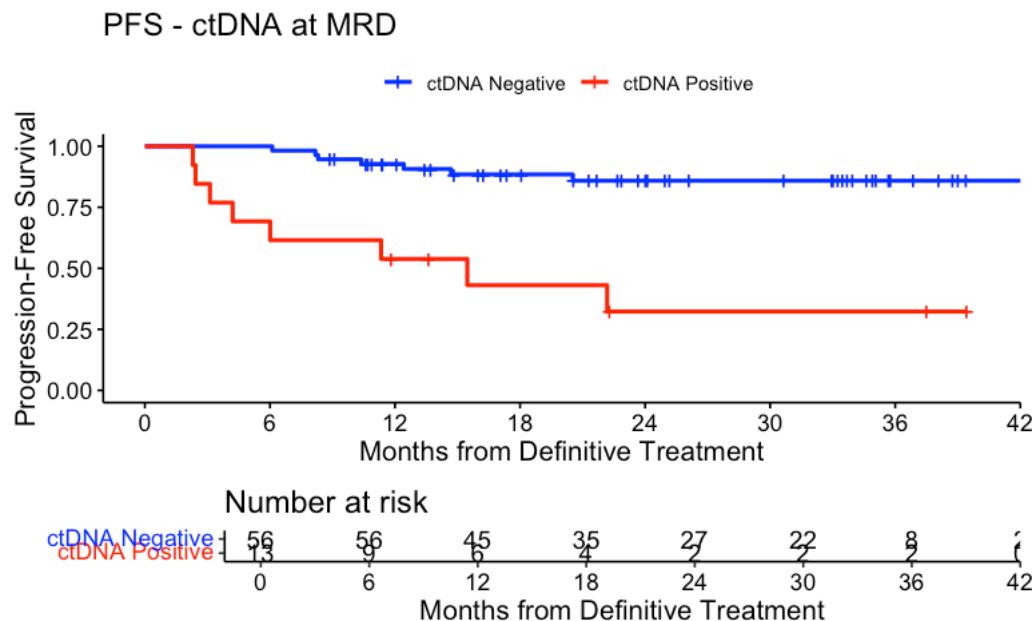
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	56	7	0.1250000	12.50000
POSITIVE	13	8	0.6153846	61.53846

2 rows

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="PFS - ctDNA at MRD", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

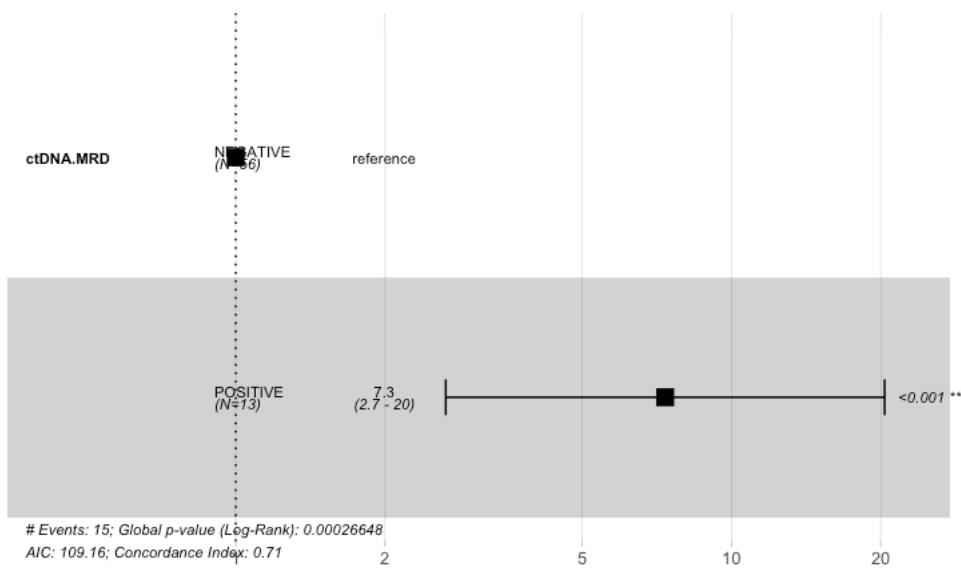
```
ctDNA.MRD=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    56      0    1.000  0.0000    1.000    1.000
 12    45      4    0.928  0.0348    0.819    0.972
 24    27      3    0.859  0.0503    0.723    0.931
 36     8      0    0.859  0.0503    0.723    0.931
```

```
ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    13      0    1.000  0.0000    1.0000    1.000
 12     6      6    0.538  0.138    0.2477    0.760
 24     2      2    0.323  0.144    0.0862    0.594
 36     2      0    0.323  0.144    0.0862    0.594
```

Hide

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 69, number of events= 15

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE 1.995     7.349    0.520  3.836  0.000125 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE     7.349      0.1361    2.652    20.36

Concordance= 0.709  (se = 0.061 )
Likelihood ratio test= 13.29  on 1 df,  p=3e-04
Wald test             = 14.71  on 1 df,  p=1e-04
Score (logrank) test = 20.16  on 1 df,  p=7e-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_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 7.35 (2.65-20.36); p = 0"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 12.17, df = 1, p-value = 0.0004856
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.0005695
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.336708 55.181495
sample estimates:
odds ratio
 10.61815
```

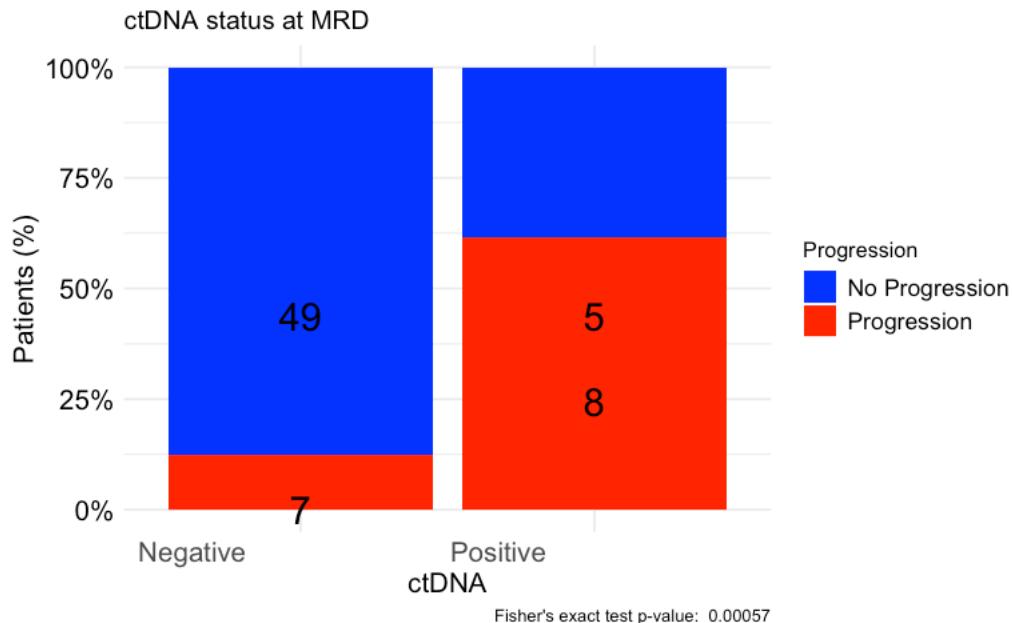
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	49	7
Positive	5	8

[Hide](#)

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



#OS by ctDNA status at MRD

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	56	1	NA	NA	NA
ctDNA.MRD=POSITIVE	13	5	NA	12.3	NA

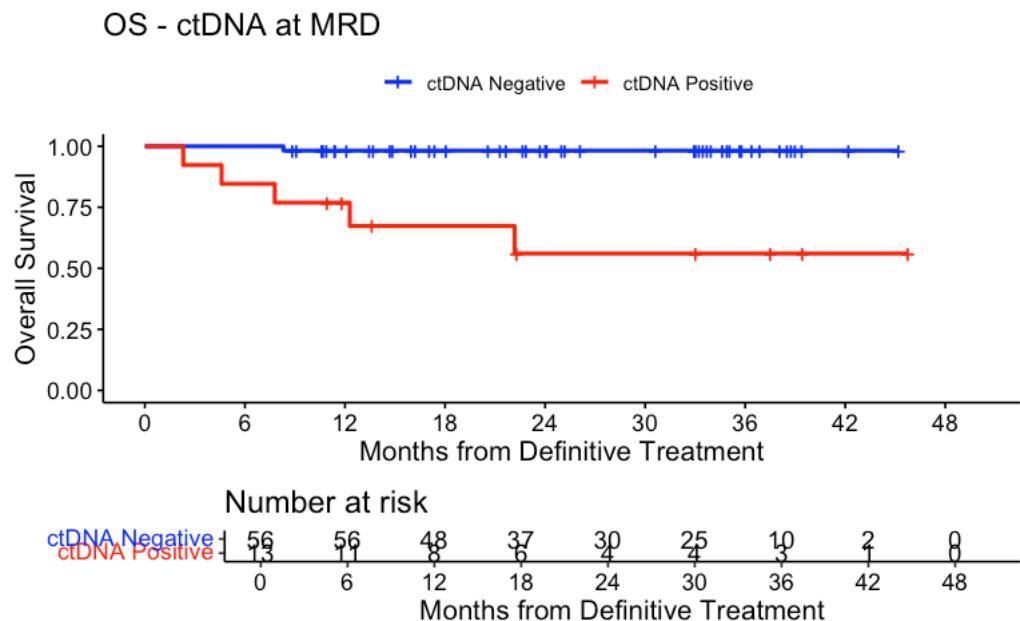
Hide

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	56	1	0.01785714	1.785714
POSITIVE	13	5	0.38461538	38.461538
2 rows				

Hide

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



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

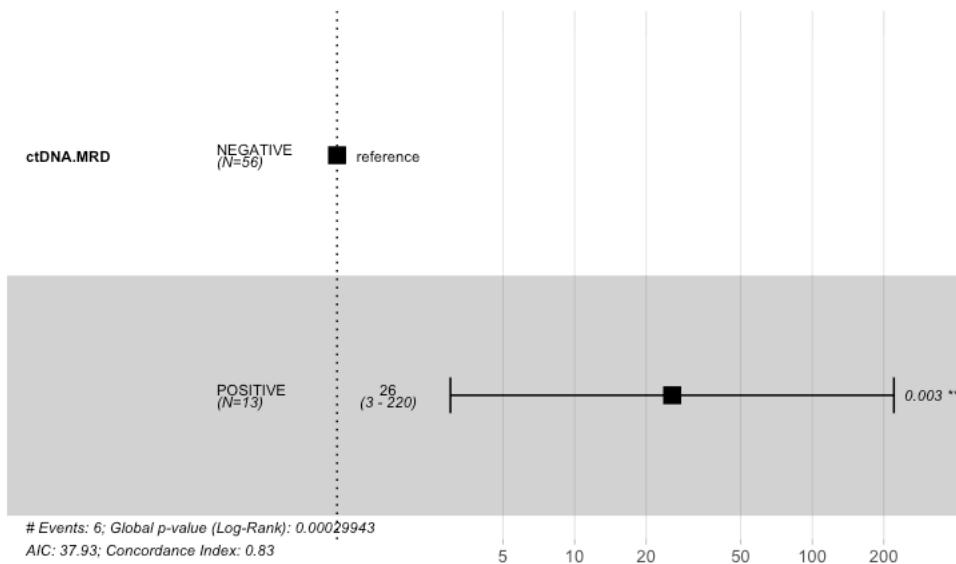
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12     48      1    0.982  0.0177      0.88    0.997
    24     30      0    0.982  0.0177      0.88    0.997
    36     10      0    0.982  0.0177      0.88    0.997

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12      8      3    0.769  0.117     0.442    0.919
    24      4      2    0.561  0.153     0.233    0.795
    36      3      0    0.561  0.153     0.233    0.795
```

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

## Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 69, number of events= 6

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 3.247    25.718    1.096  2.962  0.00306 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    25.72     0.03888      3     220.5

Concordance= 0.832  (se = 0.081 )
Likelihood ratio test= 13.07  on 1 df,  p=3e-04
Wald test            = 8.77  on 1 df,  p=0.003
Score (logrank) test = 19.67  on 1 df,  p=9e-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_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 25.72 (3-220.48); p = 0.003"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 13.554, df = 1, p-value = 0.0002318
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.0006155
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 3.015475 1634.641331
sample estimates:
odds ratio
 31.44433
```

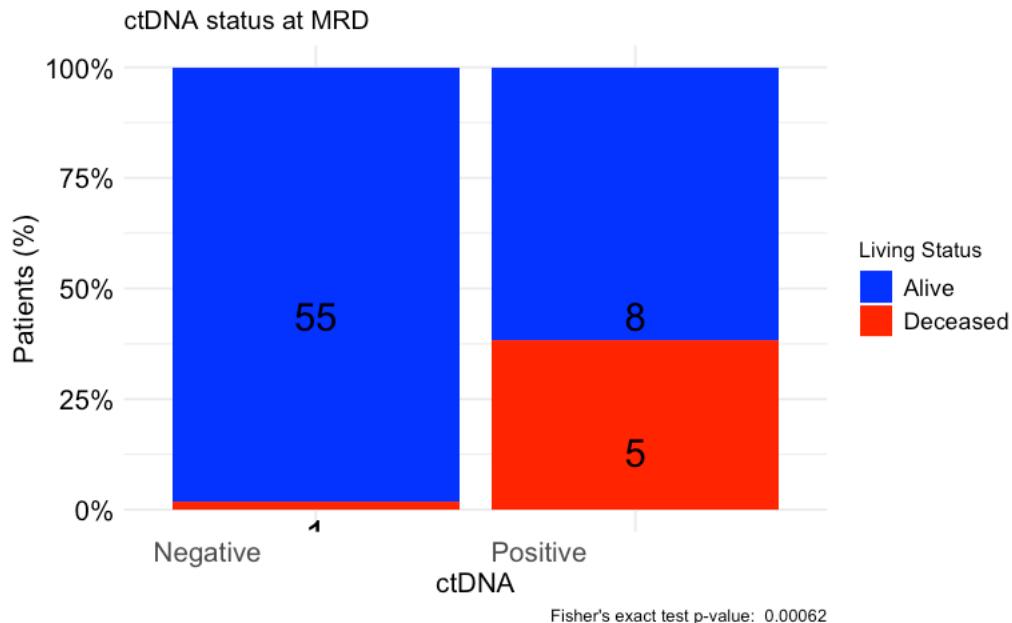
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	55	1
Positive	8	5

[Hide](#)

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



#PFS by ctDNA status at MRD - exclude pts with no subsequent adj. treatment

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[!(circ_data$Surgery == TRUE & circ_data$Chemotherapy == FALSE), ]
circ_datadf <- as.data.frame(circ_data)

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

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

```
      n  events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 48      6     NA     NA     NA
ctDNA.MRD=POSITIVE 10      7  8.67    3.12     NA
```

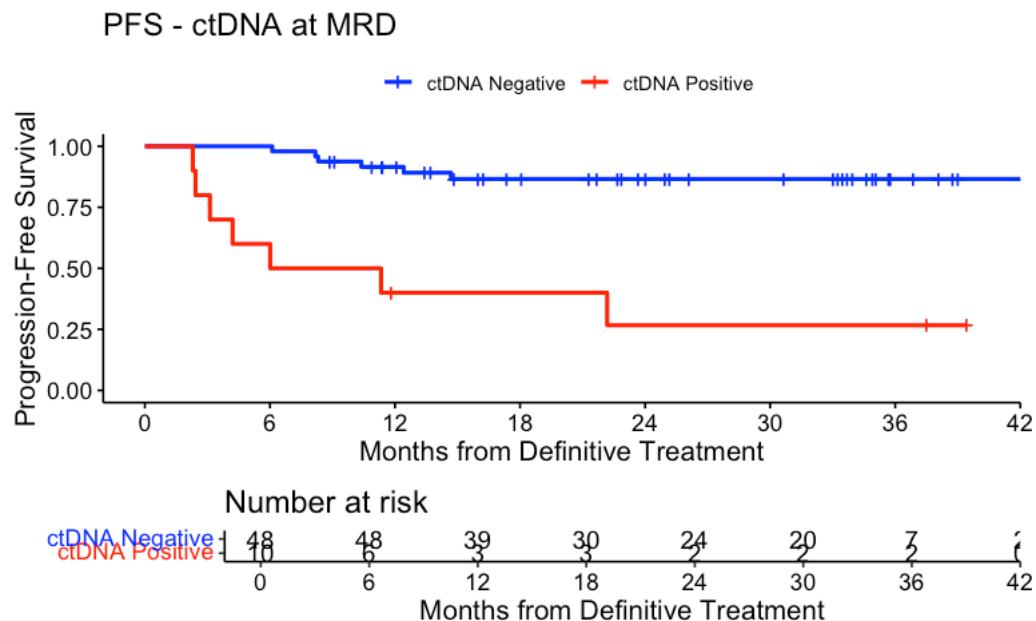
Hide

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	48	6	0.125	12.5
POSITIVE	10	7	0.700	70.0
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="PFS - ctDNA at MRD", ylab= "Progression-Free Survival", xlab="Months from Definitive Tr
eatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

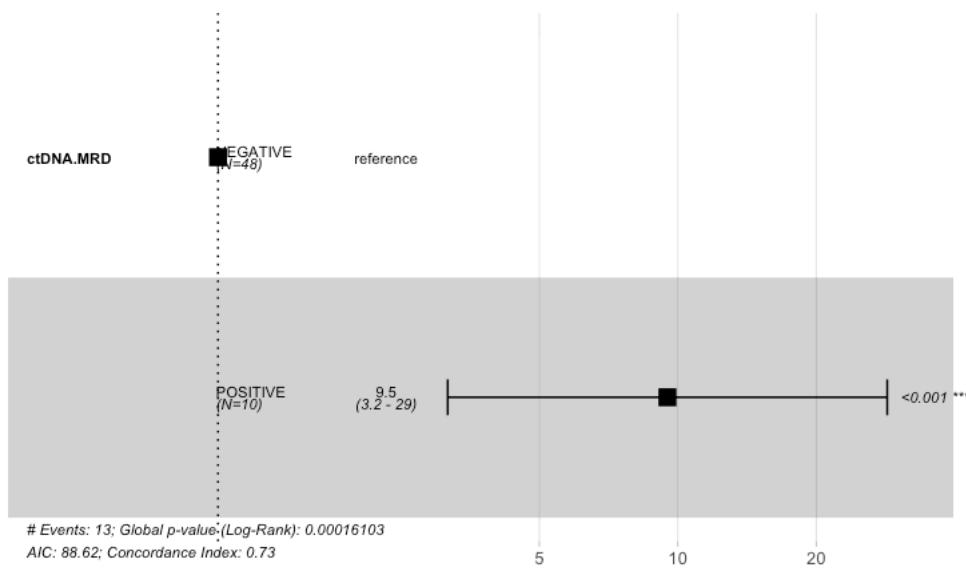
```
ctDNA.MRD=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    48      0    1.000  0.0000    1.000    1.000
 12    39      4    0.916  0.0404    0.791    0.968
 24    24      2    0.866  0.0512    0.725    0.938
 36     7      0    0.866  0.0512    0.725    0.938
```

```
ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    10      0    1.000  0.0000    1.0000    1.000
 12     3      6    0.400  0.155    0.1227    0.670
 24     2      1    0.267  0.150    0.0476    0.563
 36     2      0    0.267  0.150    0.0476    0.563
```

Hide

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 58, number of events= 13

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE 2.2507    9.4944   0.5614  4.009  6.1e-05 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE    9.494     0.1053    3.159    28.53    

Concordance= 0.725  (se = 0.064 )
Likelihood ratio test= 14.24  on 1 df,  p=2e-04
Wald test             = 16.07  on 1 df,  p=6e-05
Score (logrank) test = 23.76  on 1 df,  p=1e-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_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.49 (3.16-28.53); p = 0"
```

```
#PFS by ctDNA status at MRD Stage I/II
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$cStage=="I/II",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	29	2	NA	NA	NA
ctDNA.MRD=POSITIVE	5	2	NA	6.01	NA

[Hide](#)

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

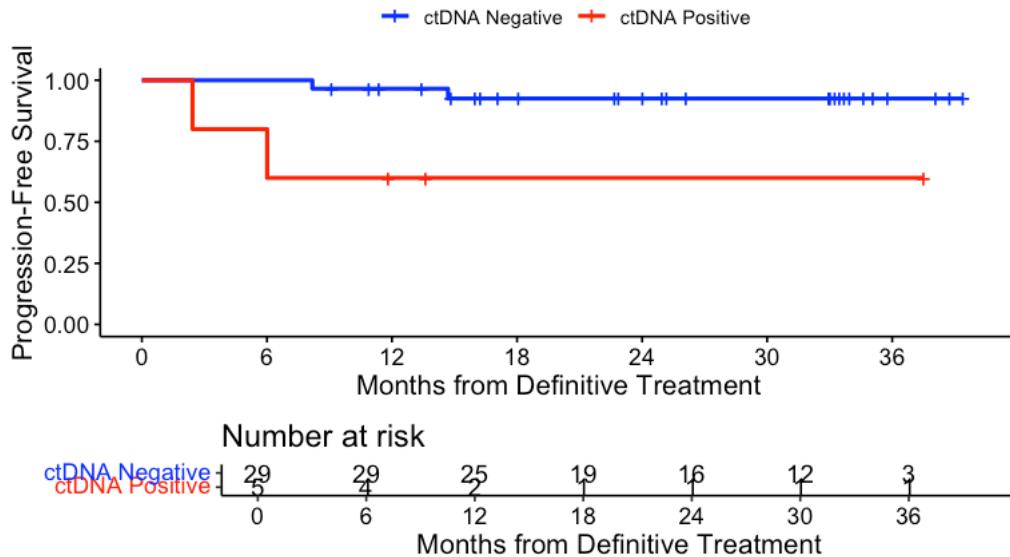
ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	29	2	0.06896552	6.896552
POSITIVE	5	2	0.40000000	40.000000

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="PFS - ctDNA at MRD Stage I/II", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

## PFS - ctDNA at MRD Stage I/II



```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

ctDNA.MRD=NEGATIVE

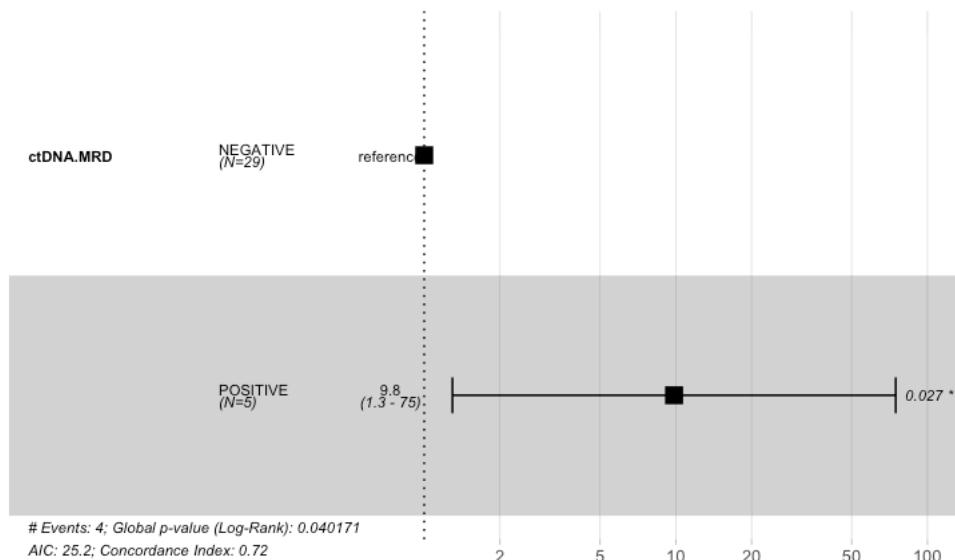
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	29	0	1.000	0.0000	1.000	1.000	1.000	1.000
12	25	1	0.966	0.0339	0.779	0.995	0.925	0.951
24	16	1	0.925	0.0510	0.732	0.981	0.875	0.981
36	3	0	0.925	0.0510	0.732	0.981	0.875	0.981

ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	5	0	1.0	0.000	1.000	1.000	1.000	1.000
12	2	2	0.6	0.219	0.126	0.882	0.126	0.882
24	1	0	0.6	0.219	0.126	0.882	0.126	0.882
36	1	0	0.6	0.219	0.126	0.882	0.126	0.882

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

## Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 34, number of events= 4

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE 2.286      9.838    1.035  2.209    0.0272 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE      9.838      0.1016     1.294     74.78

Concordance= 0.725  (se = 0.118 )
Likelihood ratio test= 4.21  on 1 df,  p=0.04
Wald test             = 4.88  on 1 df,  p=0.03
Score (logrank) test = 7.19  on 1 df,  p=0.007
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 9.84 (1.29-74.78); p = 0.027"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 1.8778, df = 1, p-value = 0.1706
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.09391
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.4412663 153.9655852
sample estimates:
odds ratio
 8.070894
```

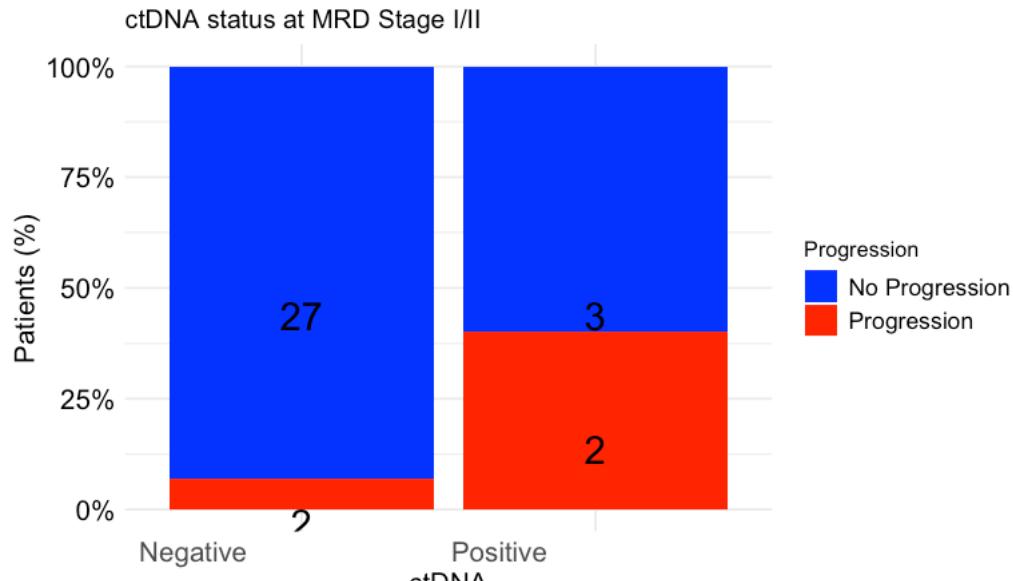
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	27	2
Positive	3	2

[Hide](#)

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



#PFS by ctDNA status at MRD Stage III/IV

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$cStage=="III/IV",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

```
      n  events median 0.95LCL 0.95UCL
ctDNA.MRD=NEGATIVE 27      5     NA     NA     NA
ctDNA.MRD=POSITIVE  8      6  13.4    4.21     NA
```

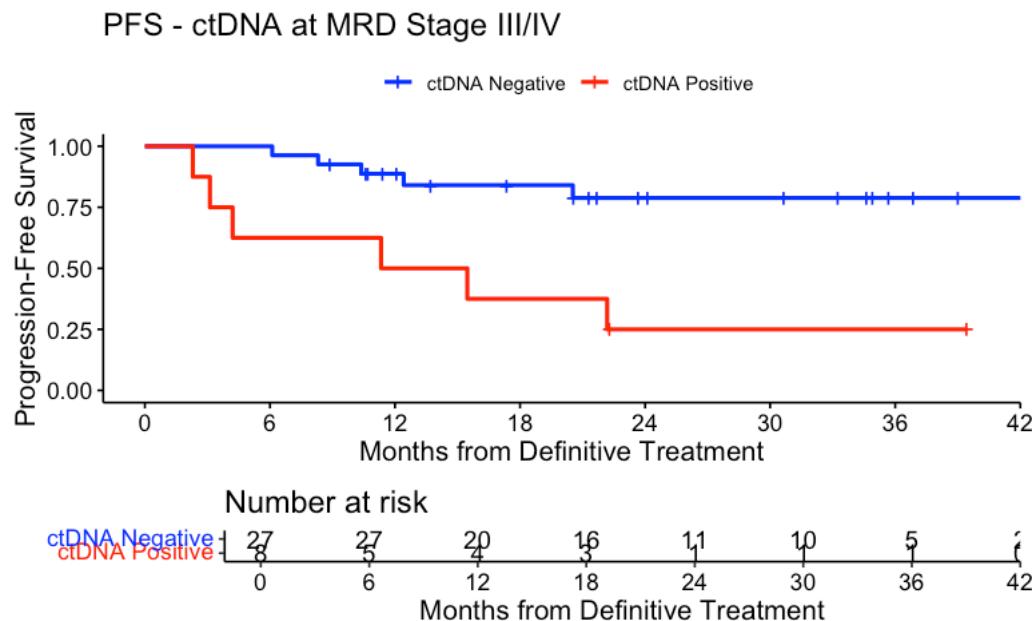
Hide

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	27	5	0.1851852	18.51852
POSITIVE	8	6	0.7500000	75.00000
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="PFS - ctDNA at MRD Stage III/IV", ylab= "Progression-Free Survival", xlab="Months from
Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

ctDNA.MRD=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	27	0	1.000	0.0000	1.000	1.000	1.000	1.000
12	20	3	0.887	0.0613	0.690	0.962	0.558	0.907
24	11	2	0.788	0.0858	0.558	0.907	0.558	0.907
36	5	0	0.788	0.0858	0.558	0.907	0.558	0.907

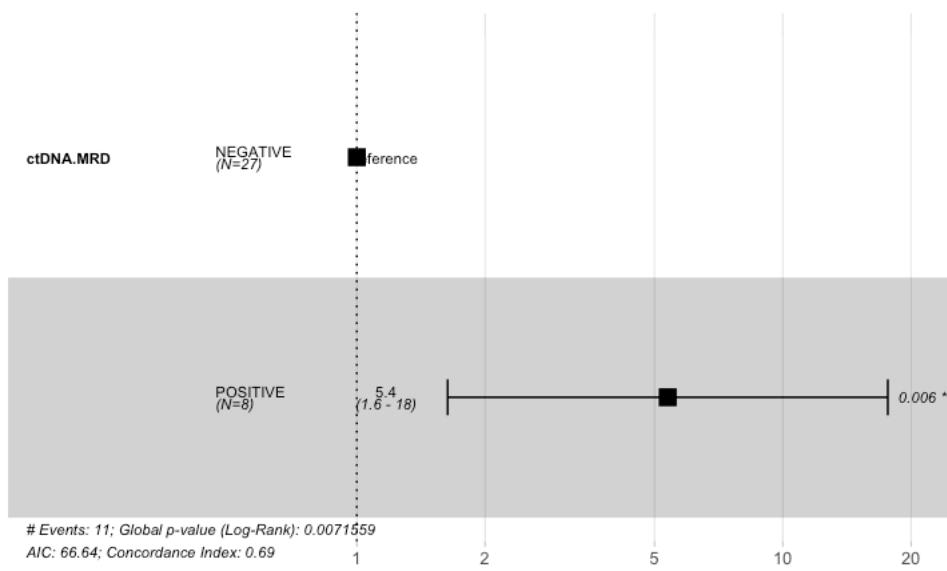
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	8	0	1.00	0.000	1.0000	1.000	1.000	1.000
12	4	4	0.50	0.177	0.1520	0.775	0.25	0.558
24	1	2	0.25	0.153	0.0371	0.558	0.25	0.558
36	1	0	0.25	0.153	0.0371	0.558	0.25	0.558

Hide

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 35, number of events= 11

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE 1.681      5.371     0.607  2.769  0.00562 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE      5.371      0.1862     1.634     17.65 

Concordance= 0.69  (se = 0.072 )
Likelihood ratio test= 7.23  on 1 df,  p=0.007
Wald test             = 7.67  on 1 df,  p=0.006
Score (logrank) test = 9.63  on 1 df,  p=0.002
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 5.37 (1.63-17.65); p = 0.006"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 6.7026, df = 1, p-value = 0.009627
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.005761
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.572192 155.593667
sample estimates:
odds ratio
11.93406
```

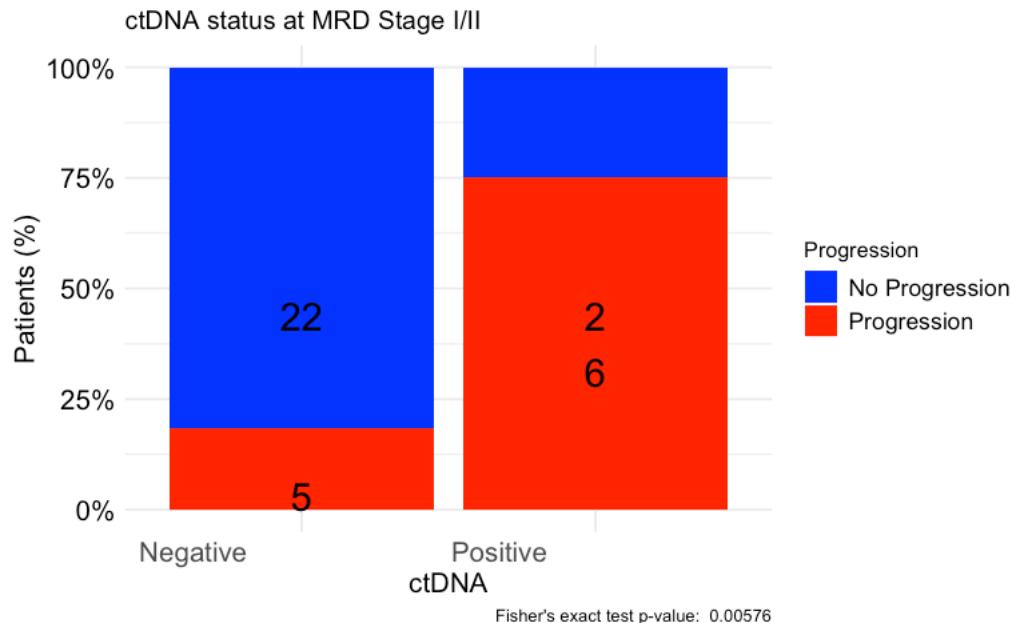
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	22	5
Positive	2	6

[Hide](#)

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



#PFS by ctDNA at MRD p16(+)

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$p16.status=="Positive",]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	29	2	NA	NA	NA
ctDNA.MRD=POSITIVE	8	4	22.2	6.01	NA

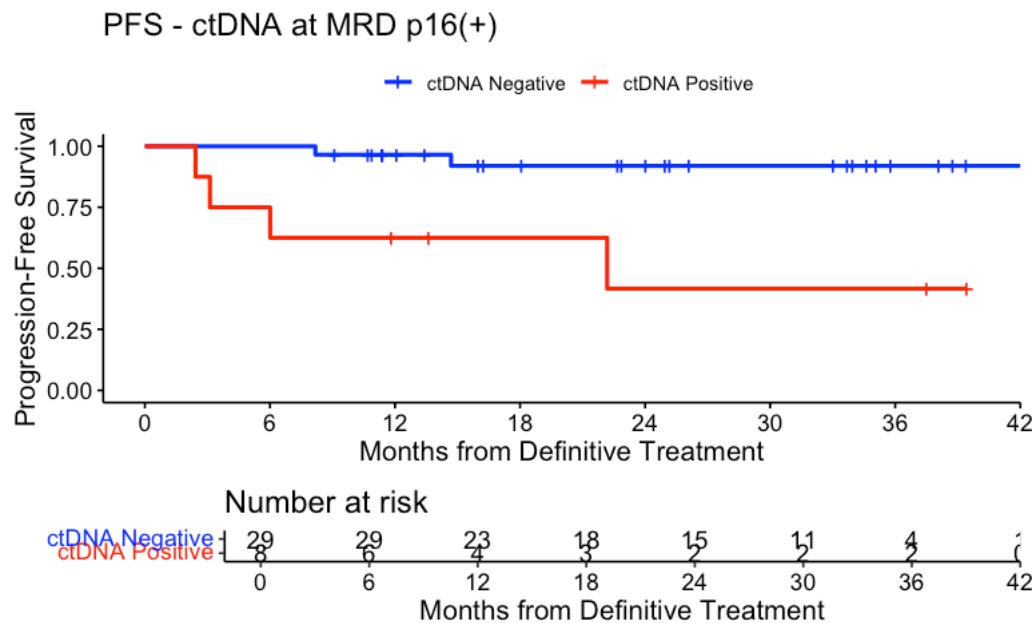
Hide

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	29	2	0.06896552	6.896552
POSITIVE	8	4	0.50000000	50.000000
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "red"), title="PFS - ctDNA at MRD p16(+)", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

```
ctDNA.MRD=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    29      0    1.000  0.0000    1.000    1.000
 12    23      1    0.966  0.0339    0.779    0.995
 24    15      1    0.920  0.0553    0.711    0.980
 36     4      0    0.920  0.0553    0.711    0.980
```

```
ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     8      0    1.000  0.0000    1.000    1.000
 12     4      3    0.625  0.171    0.229    0.861
 24     2      1    0.417  0.205    0.072    0.747
 36     2      0    0.417  0.205    0.072    0.747
```

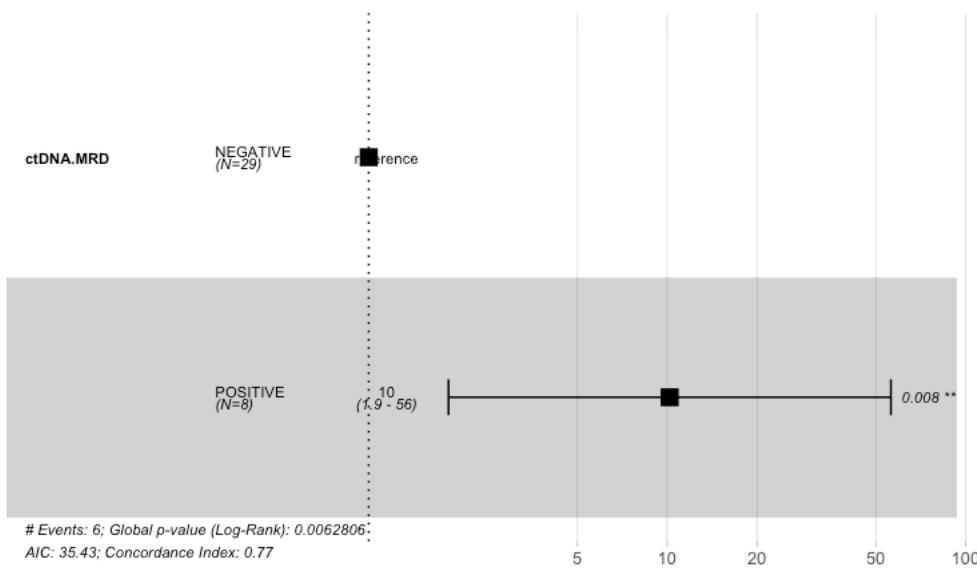
Hide

```

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

```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

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

n= 37, number of events= 6

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE  2.3218   10.1936   0.8706  2.667  0.00766 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE   10.19      0.0981     1.85     56.16

Concordance= 0.767  (se = 0.09 )
Likelihood ratio test= 7.47  on 1 df,  p=0.006
Wald test            = 7.11  on 1 df,  p=0.008
Score (logrank) test = 10.81  on 1 df,  p=0.001

```

Hide

```

cox_fit_summary <- summary(cox_fit)

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

```

```
[1] "HR = 10.19 (1.85-56.16); p = 0.008"
```

Hide

```

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

```

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 5.6953, df = 1, p-value = 0.01701
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.01294
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.281882 176.017338
sample estimates:
odds ratio
12.07276
```

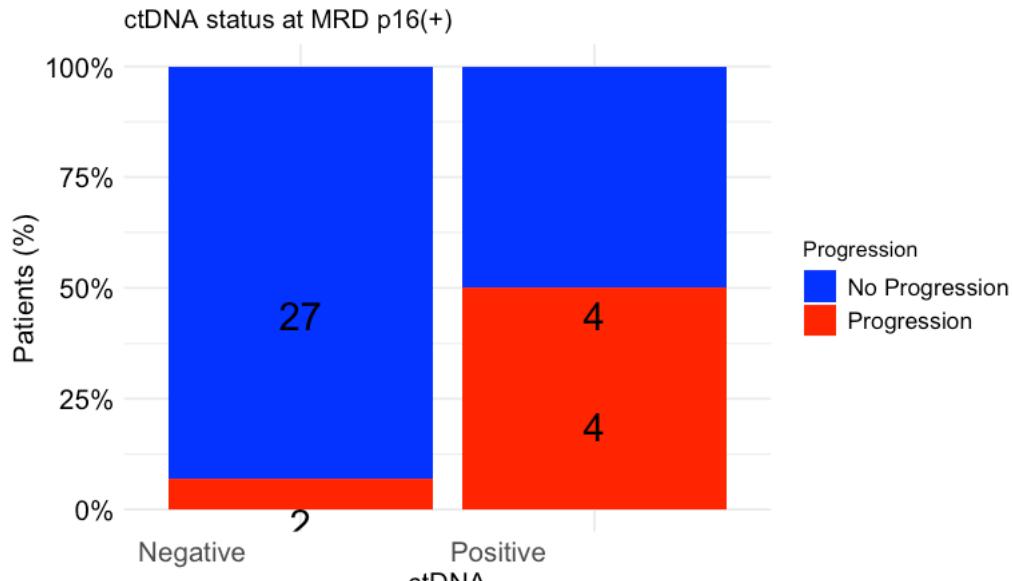
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	27	2
Positive	4	4

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA status at MRD p16(+)",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Progression",
       caption = paste("Fisher's exact test p-value: ", format.pval(fisher_exact_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Progression" = "blue", "Progression" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Progression label size
```



#PFS by ctDNA at MRD p16(-)

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$p16.status=="Negative",]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	27	5	NA	NA	NA
ctDNA.MRD=POSITIVE	5	4	11.3	4.21	NA

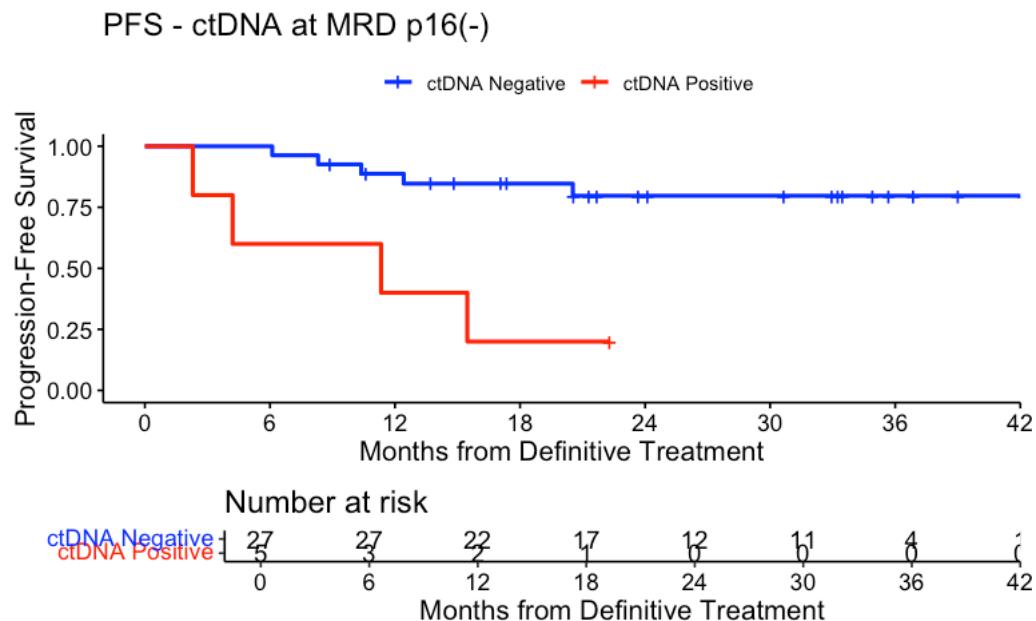
Hide

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

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	27	5	0.1851852	18.51852
POSITIVE	5	4	0.8000000	80.00000
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "red"), title="PFS - ctDNA at MRD p16(-)", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

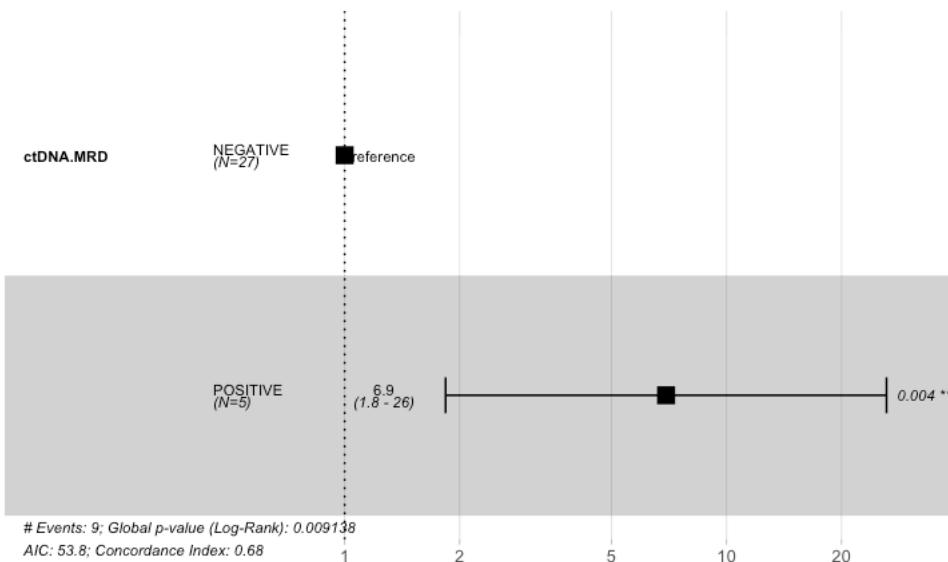
  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0     27      0     1.000  0.0000     1.000     1.000
   12     22      3     0.887  0.0613     0.690     0.962
   24     12      2     0.797  0.0821     0.576     0.911
   36      4      0     0.797  0.0821     0.576     0.911

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    0      5      0     1.0     0.000     1.000     1.000
   12      2      3     0.4     0.219     0.052     0.753
```

Hide

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

## Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 32, number of events= 9

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.9377    6.9431  0.6781 2.857  0.00427 **

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

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    6.943     0.144    1.838    26.23

Concordance= 0.682  (se = 0.077 )
Likelihood ratio test= 6.8  on 1 df,  p=0.009
Wald test            = 8.16  on 1 df,  p=0.004
Score (logrank) test = 10.96  on 1 df,  p=9e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 6.94 (1.84-26.23); p = 0.004"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 5.1404, df = 1, p-value = 0.02338
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.01502
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.221317 898.357859
sample estimates:
odds ratio
15.55541
```

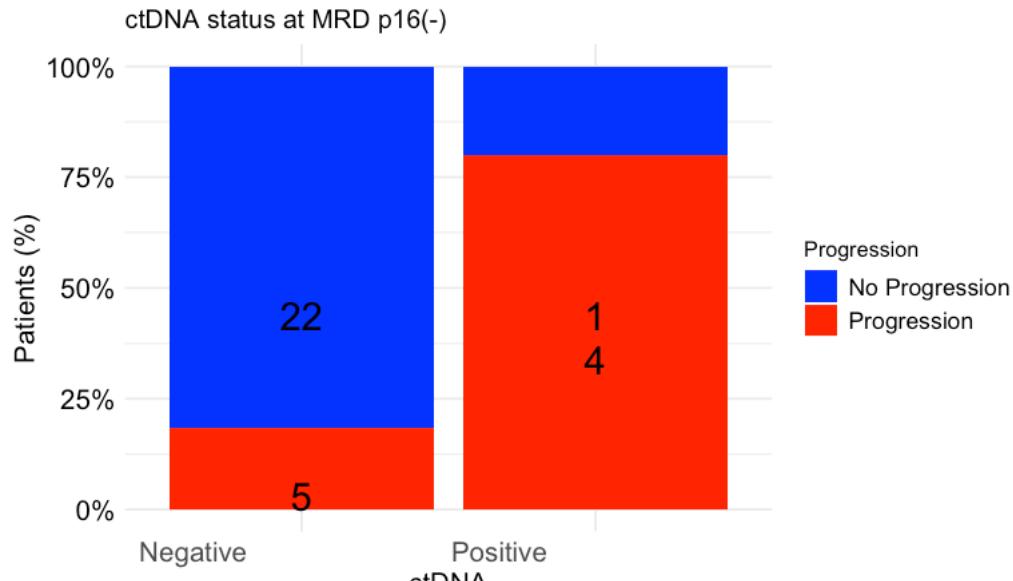
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	22	5
Positive	1	4

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA status at MRD p16(-)",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Progression",
       caption = paste("Fisher's exact test p-value: ", format.pval(fisher_exact_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Progression" = "blue", "Progression" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Progression label size
```



#PFS by ctDNA status at surveillance

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	51	3	NA	NA	NA
ctDNA.Surveillance=POSITIVE	17	13	15.5	11.6	NA

Hide

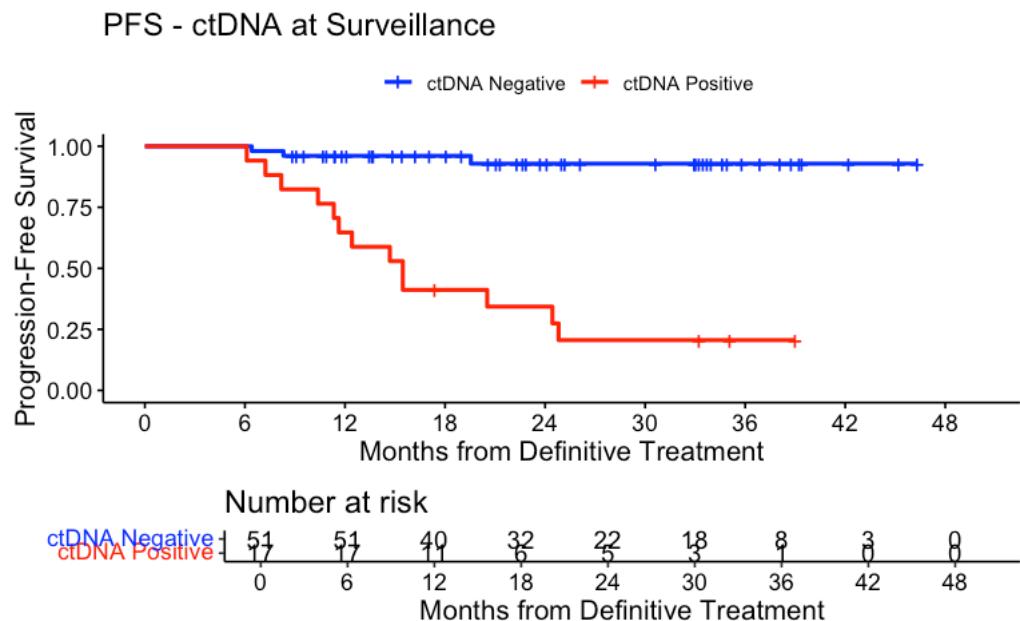
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	51	3	0.05882353	5.882353
POSITIVE	17	13	0.76470588	76.470588

2 rows

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "red"), title="PFS - ctDNA at Surveillance", ylab = "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

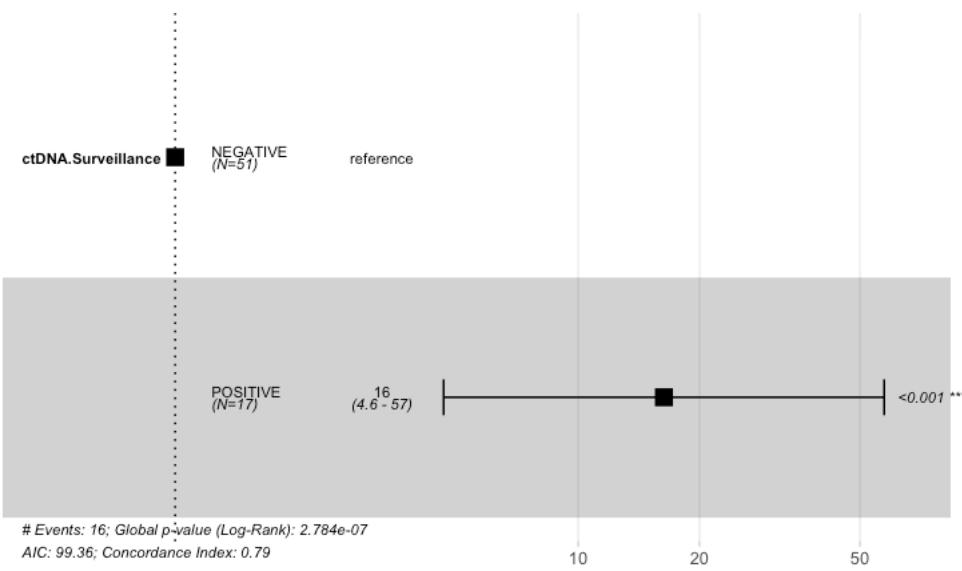
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     51      0    1.000  0.0000    1.000    1.000
 12     40      2    0.961  0.0272    0.852    0.990
 24     22      1    0.929  0.0410    0.788    0.977
 36      8      0    0.929  0.0410    0.788    0.977
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     17      0    1.000  0.0000    1.0000    1.000
 12     11      6    0.647  0.116    0.3771    0.823
 24      5      5    0.343  0.118    0.1348    0.565
 36      1      2    0.206  0.103    0.0528    0.428
```

Hide

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 68, number of events= 16

            coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.SurveillancePOSITIVE  2.792    16.317    0.642  4.349 1.37e-05 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE    16.32     0.06129    4.636    57.43

Concordance= 0.792  (se = 0.055 )
Likelihood ratio test= 26.39  on 1 df,  p=3e-07
Wald test             = 18.91  on 1 df,  p=1e-05
Score (logrank) test = 34.52  on 1 df,  p=4e-09
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 16.32 (4.64-57.43); p = 0"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 31.494, df = 1, p-value = 2.001e-08
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 3.432e-08
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 8.523283 366.804741
sample estimates:
odds ratio
 46.11116
```

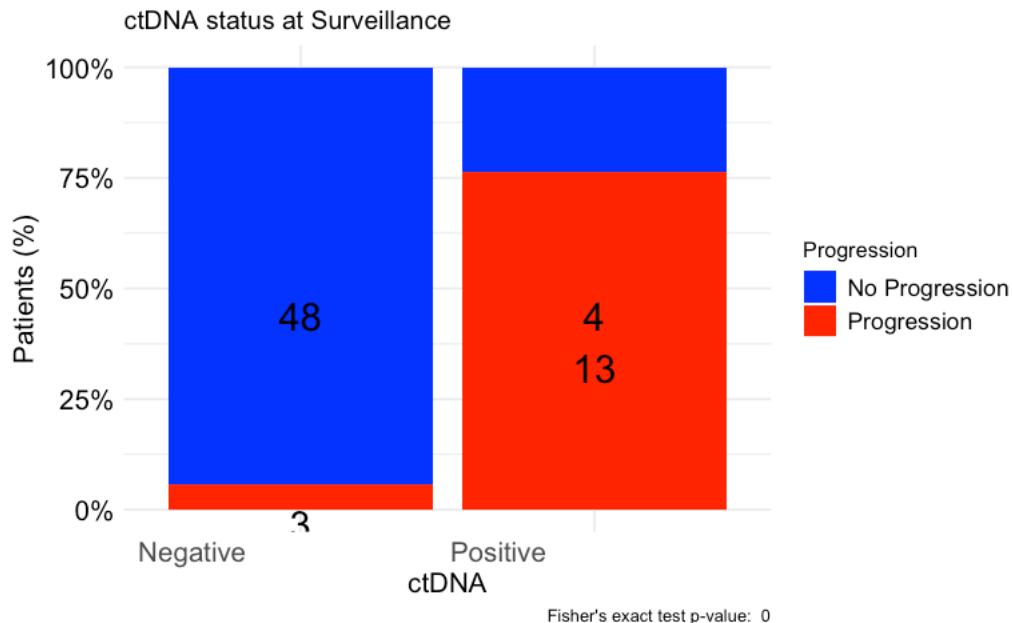
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	48	3
Positive	4	13

[Hide](#)

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



#OS by ctDNA status at surveillance

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	51	1	NA	NA	NA
ctDNA.Surveillance=POSITIVE	17	2	NA	NA	NA

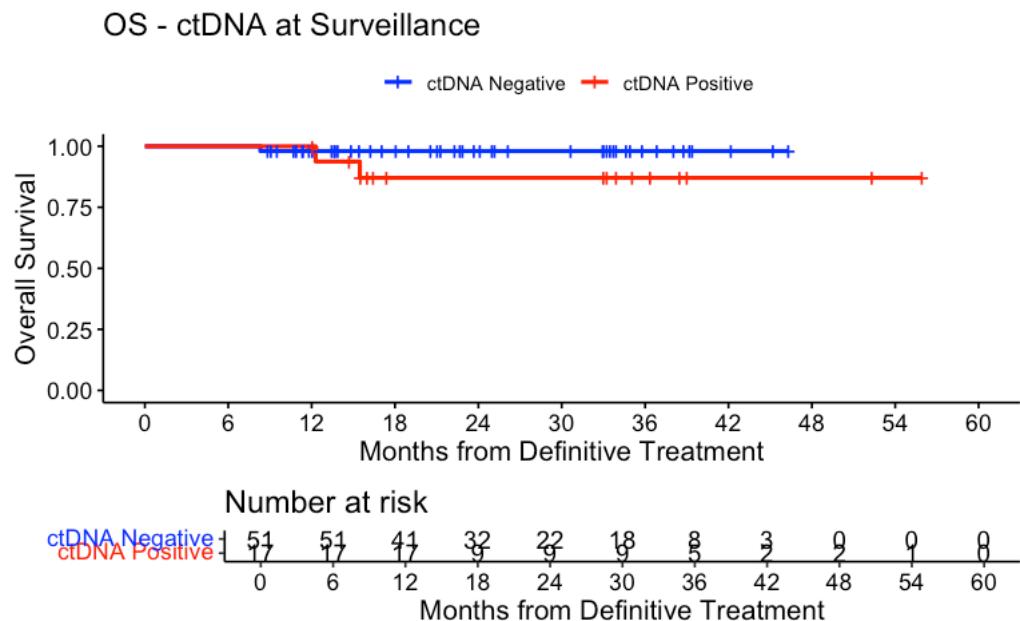
[Hide](#)

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

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	51	1	0.01960784	1.960784
POSITIVE	17	2	0.11764706	11.764706
2 rows				

[Hide](#)

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



Hide

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

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

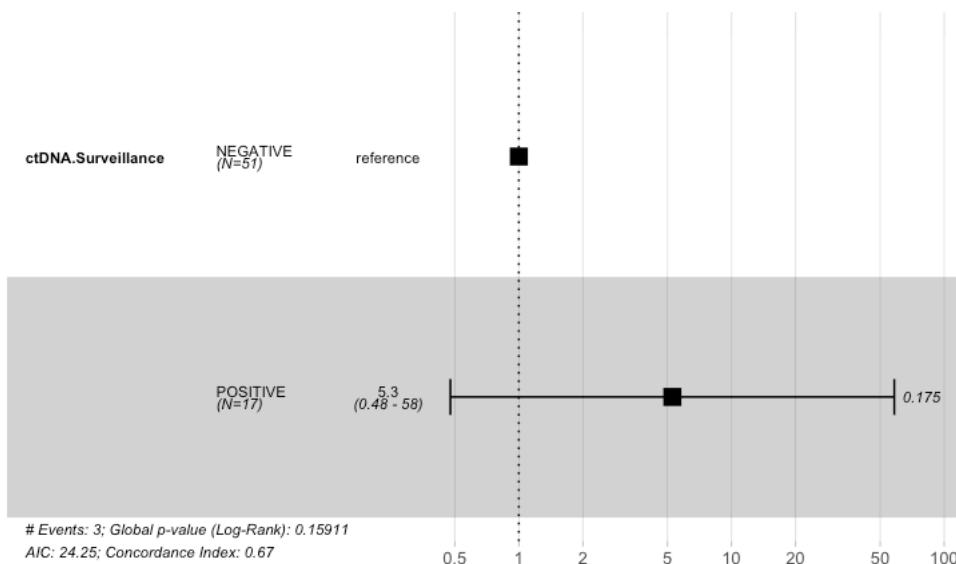
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     41      1     0.98  0.0194     0.869     0.997
 24     22      0     0.98  0.0194     0.869     0.997
 36      8      0     0.98  0.0194     0.869     0.997
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     17      0     1.000  0.0000     1.000     1.000
 24      9      2     0.871  0.0856     0.573     0.966
 36      5      0     0.871  0.0856     0.573     0.966
```

Hide

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

## Hazard ratio



```
summary(cox_fit)
```

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

n= 68, number of events= 3

            coef exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.SurveillancePOSITIVE 1.662      5.270    1.226 1.356    0.175

            exp(coef) exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE      5.27      0.1898    0.4767    58.25

Concordance= 0.669 (se = 0.143 )
Likelihood ratio test= 1.98 on 1 df,  p=0.2
Wald test      = 1.84 on 1 df,  p=0.2
Score (logrank) test = 2.3 on 1 df,  p=0.1
```

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 5.27 (0.48-58.25); p = 0.175"
```

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 1.0462, df = 1, p-value = 0.3064
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.152
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.3148908 399.7731665
sample estimates:
odds ratio
6.433899
```

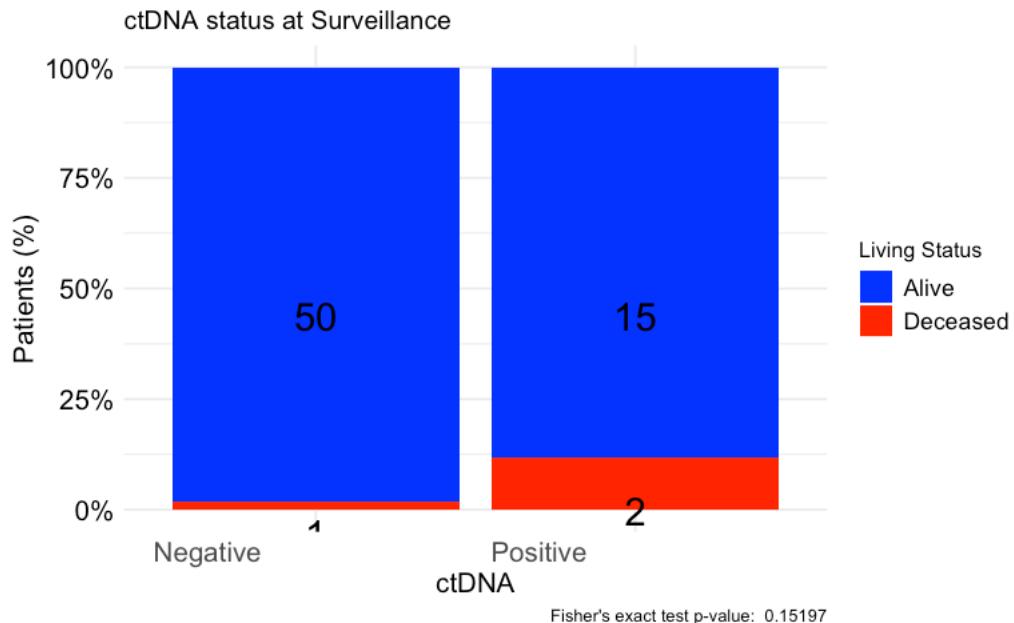
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	50	1
Positive	15	2

[Hide](#)

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



#Median numbers of time points and lead time in the longitudinal setting

[Hide](#)

```
# Load the dataset
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_datadf <- as.data.frame(circ_data)

median_Nsurvtps <- median(circ_datadf$Nsurvtps, na.rm = TRUE)
min_Nsurvtps <- min(circ_datadf$Nsurvtps, na.rm = TRUE)
max_Nsurvtps <- max(circ_datadf$Nsurvtps, na.rm = TRUE)

cat(sprintf("Median # of surveillance time points: %d (%d-%d)\n",
            median_Nsurvtps, min_Nsurvtps, max_Nsurvtps))
```

Median # of surveillance time points: 4 (1-13)

[Hide](#)

```
circ_datadf$LeadTime_Months <- circ_datadf$LeadTime / 30.437
median_LeadTime <- median(circ_datadf$LeadTime_Months, na.rm = TRUE)
min_LeadTime <- min(circ_datadf$LeadTime_Months, na.rm = TRUE)
max_LeadTime <- max(circ_datadf$LeadTime_Months, na.rm = TRUE)
cat(sprintf("Longitudinally, ctDNA positivity preceded progression by a median of %.2f mo (%.2f-%.2f)\n",
            median_LeadTime, min_LeadTime, max_LeadTime))
```

Longitudinally, ctDNA positivity preceded progression by a median of 4.75 mo (0.00-13.96)

#Time-dependent analysis for PFS in longitudinal time points

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
dt <- read_xlsx("CLIA HNSCC Peddada Clinical Data_Time dependent.xlsx") |>
  clean_names() |>
  mutate(across(.cols = c(window_start_date,dfs_date,
                         surveillance_1_date:surveillance_12_date),
               .fns = ~ as_date(as.Date(.x, format = "%Y-%m-%d"))))
```

```

G2; H2;Warning h: Expecting numeric in Z5 / R5C26: got a date g
G2; H2;Warning h: Expecting numeric in Z8 / R8C26: got a date g
G2; H2;Warning h: Expecting numeric in Z9 / R9C26: got a date g
G2; H2;Warning h: Expecting numeric in Z15 / R15C26: got a date g
G2; H2;Warning h: Expecting numeric in Z17 / R17C26: got a date g
G2; H2;Warning h: Expecting numeric in Z18 / R18C26: got a date g
G2; H2;Warning h: Expecting numeric in Z19 / R19C26: got a date g
G2; H2;Warning h: Expecting numeric in Z22 / R22C26: got a date g
G2; H2;Warning h: Expecting numeric in Z23 / R23C26: got a date g
G2; H2;Warning h: Expecting numeric in Z25 / R25C26: got a date g
G2; H2;Warning h: Expecting numeric in Z26 / R26C26: got a date g
G2; H2;Warning h: Expecting numeric in Z27 / R27C26: got a date g
G2; H2;Warning h: Expecting numeric in Z28 / R28C26: got a date g
G2; H2;Warning h: Expecting numeric in Z29 / R29C26: got a date g
G2; H2;Warning h: Expecting numeric in Z32 / R32C26: got a date g
G2; H2;Warning h: Expecting numeric in Z34 / R34C26: got a date g
G2; H2;Warning h: Expecting numeric in Z35 / R35C26: got a date g
G2; H2;Warning h: Expecting numeric in Z38 / R38C26: got a date g
G2; H2;Warning h: Expecting numeric in Z41 / R41C26: got a date g
G2; H2;Warning h: Expecting numeric in Z47 / R47C26: got a date g
G2; H2;Warning h: Expecting numeric in Z48 / R48C26: got a date g
G2; H2;Warning h: Expecting numeric in Z49 / R49C26: got a date g
G2; H2;Warning h: Expecting numeric in Z56 / R56C26: got a date g
G2; H2;Warning h: Expecting numeric in Z58 / R58C26: got a date g
G2; H2;Warning h: Expecting numeric in Z60 / R60C26: got a date g
G2; H2;Warning h: Expecting numeric in Z63 / R63C26: got a date g
G2; H2;Warning h: Expecting numeric in Z64 / R64C26: got a date g
G2; H2;Warning h: Expecting numeric in Z65 / R65C26: got a date g
G2; H2;Warning h: Expecting numeric in Z67 / R67C26: got a date g
G2; H2;Warning h: Expecting numeric in Z68 / R68C26: got a date g
G2; H2;Warning h: Expecting numeric in Z71 / R71C26: got a date g
G2; H2;Warning h: Expecting numeric in Z72 / R72C26: got a date g
G2; H2;Warning h: Expecting numeric in Z74 / R74C26: got a date g
G2; H2;Warning h: Expecting numeric in Z78 / R78C26: got a date g
G2; H2;Warning h: Expecting numeric in Z81 / R81C26: got a date g
G2; H2;Warning h: Expecting numeric in Z82 / R82C26: got a date g
G2; H2;Warning h: Expecting numeric in Z83 / R83C26: got a date g
G2; H2;Warning h: Expecting numeric in Z86 / R86C26: got a date g
G2; H2;Warning h: Expecting numeric in Z87 / R87C26: got a date g
G2; H2;Warning h: Expecting numeric in Z89 / R89C26: got a date g
G2; H2;Warning h: Expecting numeric in Z90 / R90C26: got a date g
G2; H2;Warning h: Expecting numeric in Z93 / R93C26: got a date g
G2; H2;Warning h: Expecting numeric in Z95 / R95C26: got a date g
G2; H2;Warning h: Expecting numeric in Z96 / R96C26: got a date g
G2; H2;Warning h: Expecting numeric in Z98 / R98C26: got a date g
G2; H2;Warning h: Expecting numeric in Z99 / R99C26: got a date g

```

```

dt_biomarker <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date,
         surveillance_1_status:surveillance_12_date) |>
  filter(ct_dna_surveillance_available) |>
  pivot_longer(cols = surveillance_1_status:surveillance_12_date,
               names_to = c("visit_number", ".value"),
               names_pattern = "surveillance_(.)_(.*)") |>
  mutate(biomarker_time = day(date - window_start_date)) |>
  select(pts_id, biomarker_time, biomarker_status = status) |>
  filter(!is.na(biomarker_time))

glimpse(dt_biomarker)

```

```

Rows: 219
Columns: 3
$ pts_id      <chr> "UNM-004", "UNM-004", "UNM-004", "UNM-008", "UNM-008", "UNM-008", "UNM-008",
"UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-014", "UNM-016", ...
$ biomarker_time <dbl> 18, 25719, 179, -75, 25647, 154, 236, 322, 46, 25792, 327, 418, 507, 156, 112, 25865, 38
7, 481, 9, 25746, 265, 28, 477, 19, 25756, 273, 361, 454, 550, 649, 32, 2578...
$ biomarker_status <chr> "NEGATIVE", "POSITIVE", "POSITIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", NA,
"NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEG...

```

```

dt_survival <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date:dfs_date, dfs_event) |> # Added dfs_event here
  filter(ct_dna_surveillance_available) |>
  mutate(dfs_time = (dfs_date - window_start_date),
         dfs_time = day(dfs_time)),
         dfs_event = as.numeric(dfs_event)) |>
  select(pts_id, dfs_time, dfs_event)

glimpse(dt_survival)

```

Hide

```
aux <- dt_survival %>%
  filter(dfs_time <= 0)

tab <- left_join(aux, dt) |>
  select(pts_id, window_start_date, dfs_time, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  select(pts_id, window_start_date, dfs_date, dfs_time)
```

Joining with `by = join\_by(pts\_id, dfs\_event)`

Hide

```
datatable(tab, filter = "top")
```

Show 10 entries

**Search:**

pts_id	window_start_date	dfs_date	dfs_time
All	All	All	All

No data available in table

Showing 0 to 0 of 0 entries

Previous Next

```

dt_survival <- dt_survival |>
  filter(dfs_time > 0)

aux <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x < 0)) |>
  rowwise() |>
  mutate(sum_neg =
        sum(c_across(surveillance_1_date:surveillance_12_date),
            na.rm = TRUE)) |>
  select(pts_id, sum_neg)

tab <- left_join(aux, dt) |>
  filter(sum_neg > 0) |>
  select(pts_id, sum_neg, window_start_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = window_start_date:surveillance_12_date,
               .fns = ~ as_date(.x)))

```

Hide

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search: 

	pts_id	sum_neg	window_start_date	surveillance_1_date	surveillance_2_date	surveillance_3_date	surveillance_4_date
1	UNM-008	1	2022-06-30	2022-04-16		2092-09-17	2022-12-01
2	UNM-059	1	2024-09-04	2024-06-24		2095-01-04	

Showing 1 to 2 of 2 entries

Previous

1

Next

```
aux <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_2_date:surveillance_12_date,
               .fns = ~ dfs_date < .x)) |>
  rowwise() |>
  mutate(n_biomarker_after_event = sum(c_across(surveillance_2_date:
                                                surveillance_12_date),
                                       na.rm = TRUE)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ !is.na(.x))) |>
  mutate(total_biomarker = sum(c_across(surveillance_2_date:
                                         surveillance_12_date),
                               na.rm = TRUE)) |>
  select(pts_id, n_biomarker_after_event, total_biomarker)

temp <- aux |>
  select(-pts_id) |>
  group_by(n_biomarker_after_event, total_biomarker) |> # Direct grouping
  summarise(freq = n(), .groups = "drop") # Drop groups after summarization
```

```
tab <- left_join(aux, dt) |>
  select(pts_id, n_biomarker_after_event, total_biomarker,
         dfs_date,
         surveillance_2_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  filter(n_biomarker_after_event > 0)
```

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search:

pts_id	n_biomarker_after_event	total_biomarker	dfs_date	surveillance_2_date	surveillance_3_date	surveillance_4_da
1 UNM-004	1	2	2025-03-17		2094-10-12	2024-11-08
2 UNM-008	1	4	2025-02-07		2092-09-17	2022-12-01
3 UNM-009	1	4	2025-02-07		2094-04-25	2024-08-05
4 UNM-016	1	3	2025-02-07		2094-02-15	2024-05-15
5 UNM-018	1	4	2025-02-07		2092-09-11	2022-12-07
6 UNM-019	1	6	2025-02-07		2092-09-14	2022-12-08
7 UNM-020	1	6	2025-02-07		2093-03-28	2023-06-28
8 UNM-023	1	1	2022-10-05		2092-09-17	
9 UNM-024	1	2	2025-02-07		2094-08-23	2024-11-14
10 UNM-026	1	4	2025-02-07		2094-02-01	2024-05-06

Showing 1 to 10 of 46 entries

Previous 1 2 3 4 5 Next[Hide](#)

```

aux <- tmerge(data1 = dt_survival,
               data2 = dt_survival,
               id = pts_id,
               dfs_event = event(dfs_time, dfs_event))
dt_final <- tmerge(data1 = aux,
                     data2 = dt_biomarker,
                     id = pts_id,
                     biomarker_status =
                     tdc(biomarker_time, biomarker_status))

datatable(dt_final, filter = "top")

```

Show 10 entriesSearch: 

pts_id	dfs_time	dfs_event	tstart	tstop	biomarker_status
1 UNM-004	308	0	0	18	
2 UNM-004	308	0	18	179	NEGATIVE
3 UNM-004	308	1	179	308	POSITIVE
4 UNM-008	953	0	0	154	NEGATIVE
5 UNM-008	953	0	154	236	NEGATIVE
6 UNM-008	953	0	236	953	NEGATIVE
7 UNM-009	513	0	0	46	
8 UNM-009	513	0	46	327	NEGATIVE
9 UNM-009	513	0	327	418	NEGATIVE

pts_id	dfs_time	dfs_event	tstart	tstop	biomarker_status
10 UNM-009	513	0	418	507	NEGATIVE

Showing 1 to 10 of 235 entries

Previous 1 2 3 4 5 ... 24 Next

Hide

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(dfs_time, dfs_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, dfs_event) ~ biomarker_status,
              data = dt_final)
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, dfs_event) ~ biomarker_status,
      data = dt_final)

n= 169, number of events= 16
(66 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE  3.2652   26.1856   0.5423  6.021 1.73e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE   26.19    0.03819    9.046    75.8

Concordance= 0.737  (se = 0.062 )
Likelihood ratio test= 31.78  on 1 df,  p=2e-08
Wald test             = 36.25  on 1 df,  p=2e-09
Score (logrank) test = 75.43  on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(fit)

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

```
[1] "HR = 26.19 (9.05-75.8); p = 0"
```

#Time-dependent analysis for OS in longitudinal time points

Hide

```
rm(list=ls())
setwd("~/Downloads")
dt <- read_xlsx("CLIA HNSCC Peddada Clinical Data_Time dependent.xlsx") |>
  clean_names() |>
  mutate(across(.cols = c(window_start_date,os_date,
                         surveillance_1_date:surveillance_12_date),
               .fns = ~ as_date(as.Date(.x, format = "%Y-%m-%d"))))
```

```

G2; H2;Warning h: Expecting numeric in Z5 / R5C26: got a date g
G2; H2;Warning h: Expecting numeric in Z8 / R8C26: got a date g
G2; H2;Warning h: Expecting numeric in Z9 / R9C26: got a date g
G2; H2;Warning h: Expecting numeric in Z15 / R15C26: got a date g
G2; H2;Warning h: Expecting numeric in Z17 / R17C26: got a date g
G2; H2;Warning h: Expecting numeric in Z18 / R18C26: got a date g
G2; H2;Warning h: Expecting numeric in Z19 / R19C26: got a date g
G2; H2;Warning h: Expecting numeric in Z22 / R22C26: got a date g
G2; H2;Warning h: Expecting numeric in Z23 / R23C26: got a date g
G2; H2;Warning h: Expecting numeric in Z25 / R25C26: got a date g
G2; H2;Warning h: Expecting numeric in Z26 / R26C26: got a date g
G2; H2;Warning h: Expecting numeric in Z27 / R27C26: got a date g
G2; H2;Warning h: Expecting numeric in Z28 / R28C26: got a date g
G2; H2;Warning h: Expecting numeric in Z29 / R29C26: got a date g
G2; H2;Warning h: Expecting numeric in Z32 / R32C26: got a date g
G2; H2;Warning h: Expecting numeric in Z34 / R34C26: got a date g
G2; H2;Warning h: Expecting numeric in Z35 / R35C26: got a date g
G2; H2;Warning h: Expecting numeric in Z38 / R38C26: got a date g
G2; H2;Warning h: Expecting numeric in Z41 / R41C26: got a date g
G2; H2;Warning h: Expecting numeric in Z47 / R47C26: got a date g
G2; H2;Warning h: Expecting numeric in Z48 / R48C26: got a date g
G2; H2;Warning h: Expecting numeric in Z49 / R49C26: got a date g
G2; H2;Warning h: Expecting numeric in Z56 / R56C26: got a date g
G2; H2;Warning h: Expecting numeric in Z58 / R58C26: got a date g
G2; H2;Warning h: Expecting numeric in Z60 / R60C26: got a date g
G2; H2;Warning h: Expecting numeric in Z63 / R63C26: got a date g
G2; H2;Warning h: Expecting numeric in Z64 / R64C26: got a date g
G2; H2;Warning h: Expecting numeric in Z65 / R65C26: got a date g
G2; H2;Warning h: Expecting numeric in Z67 / R67C26: got a date g
G2; H2;Warning h: Expecting numeric in Z68 / R68C26: got a date g
G2; H2;Warning h: Expecting numeric in Z71 / R71C26: got a date g
G2; H2;Warning h: Expecting numeric in Z72 / R72C26: got a date g
G2; H2;Warning h: Expecting numeric in Z74 / R74C26: got a date g
G2; H2;Warning h: Expecting numeric in Z78 / R78C26: got a date g
G2; H2;Warning h: Expecting numeric in Z81 / R81C26: got a date g
G2; H2;Warning h: Expecting numeric in Z82 / R82C26: got a date g
G2; H2;Warning h: Expecting numeric in Z83 / R83C26: got a date g
G2; H2;Warning h: Expecting numeric in Z86 / R86C26: got a date g
G2; H2;Warning h: Expecting numeric in Z87 / R87C26: got a date g
G2; H2;Warning h: Expecting numeric in Z89 / R89C26: got a date g
G2; H2;Warning h: Expecting numeric in Z90 / R90C26: got a date g
G2; H2;Warning h: Expecting numeric in Z93 / R93C26: got a date g
G2; H2;Warning h: Expecting numeric in Z95 / R95C26: got a date g
G2; H2;Warning h: Expecting numeric in Z96 / R96C26: got a date g
G2; H2;Warning h: Expecting numeric in Z98 / R98C26: got a date g
G2; H2;Warning h: Expecting numeric in Z99 / R99C26: got a date g

```

```

dt_biomarker <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date,
         surveillance_1_status:surveillance_12_date) |>
  filter(ct_dna_surveillance_available) |>
  pivot_longer(cols = surveillance_1_status:surveillance_12_date,
               names_to = c("visit_number", ".value"),
               names_pattern = "surveillance_(.)_(.*)" ) |>
  mutate(biomarker_time = day(date - window_start_date)) |>
  select(pts_id, biomarker_time, biomarker_status = status) |>
  filter(!is.na(biomarker_time))

glimpse(dt_biomarker)

```

```

Rows: 219
Columns: 3
$ pts_id      <chr> "UNM-004", "UNM-004", "UNM-004", "UNM-008", "UNM-008", "UNM-008", "UNM-008",
"UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-014", "UNM-016", ...
$ biomarker_time <dbl> 18, 25719, 179, -75, 25647, 154, 236, 322, 46, 25792, 327, 418, 507, 156, 112, 25865, 38
7, 481, 9, 25746, 265, 28, 477, 19, 25756, 273, 361, 454, 550, 649, 32, 2578...
$ biomarker_status <chr> "NEGATIVE", "POSITIVE", "POSITIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", NA,
"NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE"

```

```
dt_survival <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date:os_date, os_event) |> # Added os_event here
  filter(ct_dna_surveillance_available) |>
  mutate(os_time = (os_date - window_start_date),
         os_time = day(days(os_time)),
         os_event = as.numeric(os_event)) |>
  select(pts_id, os_time, os_event)

glimpse(dt_survival)
```

Hide

```

aux <- dt_survival %>%
  filter(os_time <= 0)

tab <- left_join(aux, dt) |>
  select(pts_id, window_start_date, os_time, os_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = os_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  select(pts_id, window_start_date, os_date, os_time)

```

Joining with `by = join\_by(pts\_id, os\_event)`

Hide

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

**Search:**

pts_id	window_start_date	os_date	os_time
All	All	All	All

No data available in table

Showing 0 to 0 of 0 entries

Previous Next

```

dt_survival <- dt_survival |>
  filter(os_time > 0)

aux <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date, os_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x < 0)) |>
  rowwise() |>
  mutate(sum_neg =
        sum(c_across(surveillance_1_date:surveillance_12_date),
            na.rm = TRUE)) |>
  select(pts_id, sum_neg)

tab <- left_join(aux, dt) |>
  filter(sum_neg > 0) |>
  select(pts_id, sum_neg, window_start_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = window_start_date:surveillance_12_date,
               .fns = ~ as_date(.x)))

```

Hide

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search: 

	pts_id	sum_neg	window_start_date	surveillance_1_date	surveillance_2_date	surveillance_3_date	surveillance_4_date
1	UNM-008	1	2022-06-30	2022-04-16		2092-09-17	2022-12-01
2	UNM-059	1	2024-09-04	2024-06-24		2095-01-04	

Showing 1 to 2 of 2 entries

Previous

1

Next

```
aux <- dt |>
  select(pts_id, ct_dna_surveillance_available,
         window_start_date, os_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = os_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_2_date:surveillance_12_date,
               .fns = ~ os_date < .x)) |>
  rowwise() |>
  mutate(n_biomarker_after_event = sum(c_across(surveillance_2_date:
                                                surveillance_12_date),
                                       na.rm = TRUE)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ !is.na(.x))) |>
  mutate(total_biomarker = sum(c_across(surveillance_2_date:
                                         surveillance_12_date),
                               na.rm = TRUE)) |>
  select(pts_id, n_biomarker_after_event, total_biomarker)

temp <- aux |>
  select(-pts_id) |>
  group_by(n_biomarker_after_event, total_biomarker) |> # Direct grouping
  summarise(freq = n(), .groups = "drop") # Drop groups after summarization
```

```
tab <- left_join(aux, dt) |>
  select(pts_id, n_biomarker_after_event, total_biomarker,
         os_date,
         surveillance_2_date:surveillance_12_date) |>
  mutate(across(.cols = os_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  filter(n_biomarker_after_event > 0)
```

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search:

pts_id	n_biomarker_after_event	total_biomarker	os_date	surveillance_2_date	surveillance_3_date	surveillance_4_date
1 UNM-004	1	2	2025-03-17		2094-10-12	2024-11-08
2 UNM-008	1	4	2025-02-07		2092-09-17	2022-12-01
3 UNM-009	1	4	2025-02-07		2094-04-25	2024-08-05
4 UNM-016	1	3	2025-02-07		2094-02-15	2024-05-15
5 UNM-018	1	4	2025-02-07		2092-09-11	2022-12-07
6 UNM-019	1	6	2025-02-07		2092-09-14	2022-12-08
7 UNM-020	1	6	2025-02-07		2093-03-28	2023-06-28
8 UNM-023	1	1	2025-02-07		2092-09-17	
9 UNM-024	1	2	2025-02-07		2094-08-23	2024-11-14
10 UNM-026	1	4	2025-02-07		2094-02-01	2024-05-06

Showing 1 to 10 of 46 entries

Previous 1 2 3 4 5 NextHide

```

aux <- tmerge(data1 = dt_survival,
               data2 = dt_survival,
               id = pts_id,
               os_event = event(os_time, os_event))
dt_final <- tmerge(data1 = aux,
                     data2 = dt_biomarker,
                     id = pts_id,
                     biomarker_status =
                     tdc(biomarker_time, biomarker_status))

datatable(dt_final, filter = "top")

```

Show 10 entriesSearch: 

pts_id	os_time	os_event	tstart	tstop	biomarker_status
1 UNM-004	308	0	0	18	
2 UNM-004	308	0	18	179	NEGATIVE
3 UNM-004	308	0	179	308	POSITIVE
4 UNM-008	953	0	0	154	NEGATIVE
5 UNM-008	953	0	154	236	NEGATIVE
6 UNM-008	953	0	236	953	NEGATIVE
7 UNM-009	513	0	0	46	
8 UNM-009	513	0	46	327	NEGATIVE
9 UNM-009	513	0	327	418	NEGATIVE

pts_id	os_time	os_event	tstart	tstop	biomarker_status
10 UNM-009	513	0	418	507	NEGATIVE

Showing 1 to 10 of 235 entries

Previous 1 2 3 4 5 ... 24 Next

[Hide](#)

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(os_time, os_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, os_event) ~ biomarker_status,
              data = dt_final)
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, os_event) ~ biomarker_status,
      data = dt_final)

n= 169, number of events= 3
(66 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE 1.902      6.698    1.240 1.534    0.125
                                         exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE      6.698      0.1493    0.5893    76.13

Concordance= 0.637  (se = 0.135 )
Likelihood ratio test= 1.78 on 1 df,  p=0.2
Wald test            = 2.35 on 1 df,  p=0.1
Score (logrank) test = 3.13 on 1 df,  p=0.08
```

[Hide](#)

```
cox_fit_summary <- summary(fit)

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

```
[1] "HR = 6.7 (0.59-76.13); p = 0.125"
```

#PFS by ctDNA status at surveillance Stage I/II

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$cStage=="I/II",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	31	0	NA	NA	NA
ctDNA.Surveillance=POSITIVE	6	4	19.6	8.18	NA

Hide

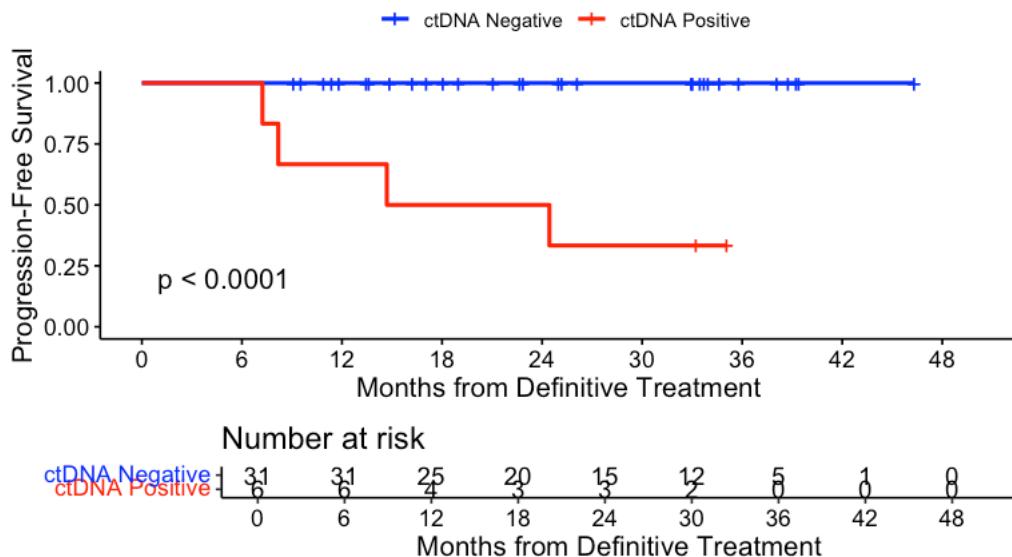
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	31	0	0.0000000	0.00000
POSITIVE	6	4	0.6666667	66.66667
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = TRUE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
=c("blue","red"), title="PFS - ctDNA at Surveillance Stage I/II", ylab= "Progression-Free Survival", xlab="Months
from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

## PFS - ctDNA at Surveillance Stage I/II



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     31      0      1      0      1      1
 12     25      0      1      0      NA      NA
 24     15      0      1      0      NA      NA
 36      5      0      1      0      NA      NA
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     6      0    1.000  0.000      1.000      1.000
 12     4      2    0.667  0.192      0.195      0.904
 24     3      1    0.500  0.204      0.111      0.804
```

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxphf(surv_object ~ ctDNA.Surveillance, data=circ_data)
summary(cox_fit)
```

```
coxphf(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)
```

```
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood
```

```
coef se(coef) exp(coef) lower 0.95 upper 0.95      Chisq      p
ctDNA.SurveillancePOSITIVE 3.907704 1.668013 49.78452  5.304948 6600.525 13.77645 0.0002059016
```

```
Likelihood ratio test=13.77645 on 1 df, p=0.0002059016, n=37
Wald test = 5.488381 on 1 df, p = 0.01914326
```

```
Covariance-Matrix:
```

```
           ctDNA.SurveillancePOSITIVE
ctDNA.SurveillancePOSITIVE          2.782269
```

Hide

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

```
coxphf(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

      coef se(coef) exp(coef) lower 0.95 upper 0.95   Chisq      p
ctDNA.SurveillancePOSITIVE 3.907704 1.668013  49.78452  5.304948 6600.525 13.77645 0.0002059016

Likelihood ratio test=13.77645 on 1 df, p=0.0002059016, n=37
Wald test = 5.488381 on 1 df, p = 0.01914326

Covariance-Matrix:
            ctDNA.SurveillancePOSITIVE
ctDNA.SurveillancePOSITIVE            2.782269
```

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

G2; H2;Warning h in stats::chisq.test(x, y, ...):  
Chi-squared approximation may be incorrect g

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 16.773, df = 1, p-value = 4.212e-05
```

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.0002271
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 5.305224      Inf
sample estimates:
odds ratio
      Inf
```

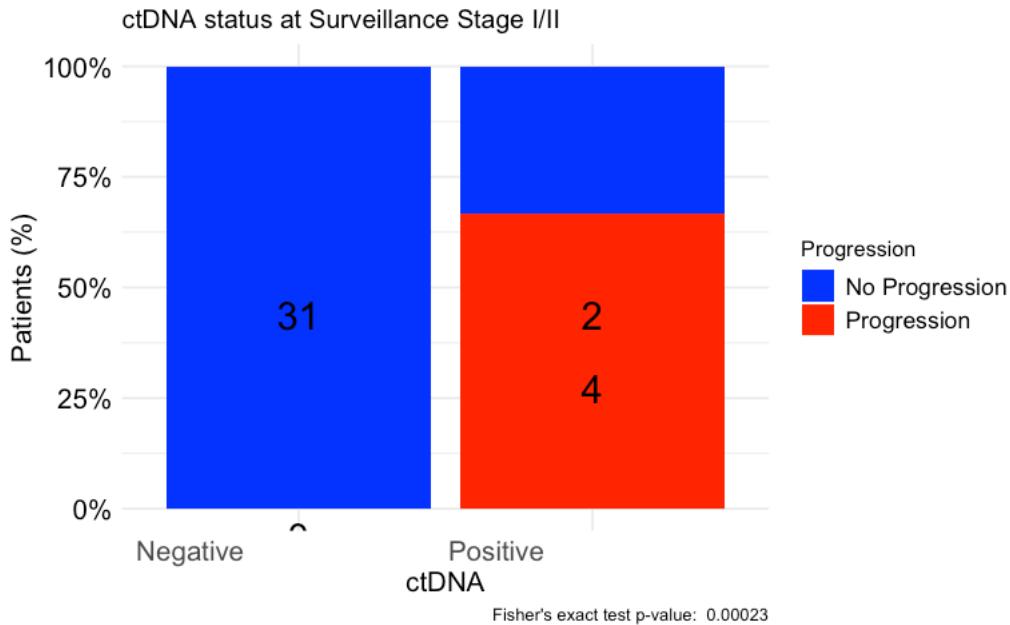
```
print(contingency_table)
```

	No Progression	Progression
Negative	31	0
Positive	2	4

```

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

```



#PFS by ctDNA status at surveillance Stage III/IV

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$cStage=="III/IV",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_dataadf <- as.data.frame(circ_data)

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

```

```

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

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	20	3	NA	NA	NA
ctDNA.Surveillance=POSITIVE	11	9	15.5	11.6	NA

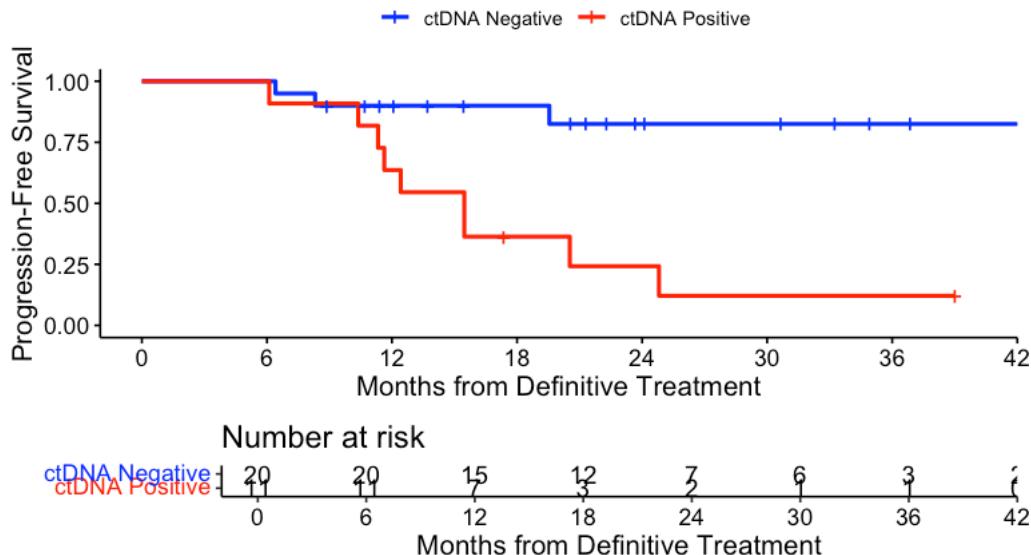
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	20	3	0.1500000	15.00000
POSITIVE	11	9	0.8181818	81.81818
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="PFS - ctDNA at Surveillance Stage III/IV", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

### PFS - ctDNA at Surveillance Stage III/IV



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

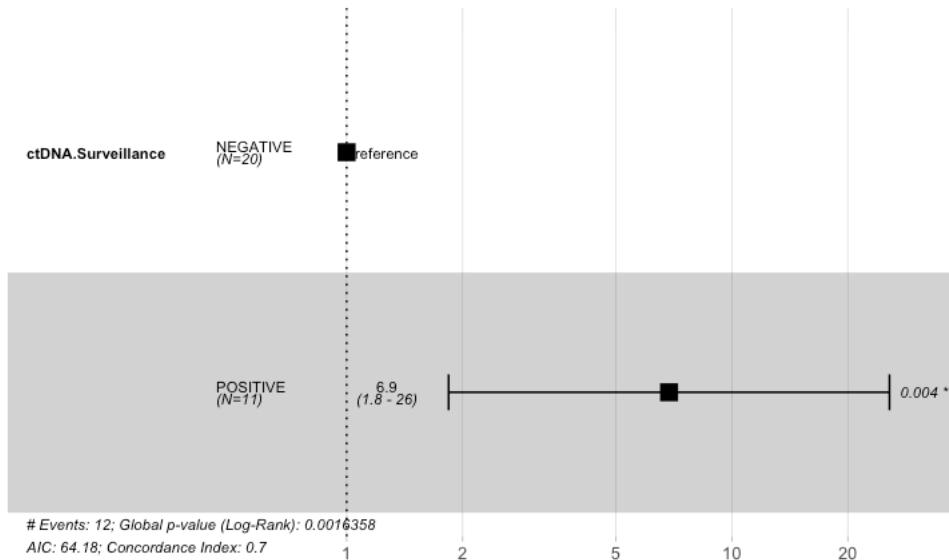
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     20      0    1.000  0.0000    1.000    1.000
 12     15      2    0.900  0.0671    0.656    0.974
 24      7      1    0.825  0.0945    0.539    0.942
 36      3      0    0.825  0.0945    0.539    0.942
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     11      0    1.000  0.0000    1.00000    1.000
 12      7      4    0.636  0.145    0.29689    0.845
 24      2      4    0.242  0.138    0.04413    0.525
 36      1      1    0.121  0.110    0.00739    0.404
```

Hide

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

### Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 31, number of events= 12

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.SurveillancePOSITIVE 1.9270    6.8686   0.6722  2.867  0.00415 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE    6.869     0.1456    1.839    25.65 

Concordance= 0.704 (se = 0.074 )
Likelihood ratio test= 9.92 on 1 df,  p=0.002
Wald test             = 8.22 on 1 df,  p=0.004
Score (logrank) test = 10.93 on 1 df,  p=9e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

[1] "HR = 6.87 (1.84-25.65); p = 0.004"

[Hide](#)

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

G2; H2;Warning h in stats::chisq.test(x, y, ...) :  
Chi-squared approximation may be incorrect g

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 10.687, df = 1, p-value = 0.001079
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.0004593
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.787876 307.695962
sample estimates:
odds ratio
 21.73942
```

[Hide](#)

```
print(contingency_table)
```

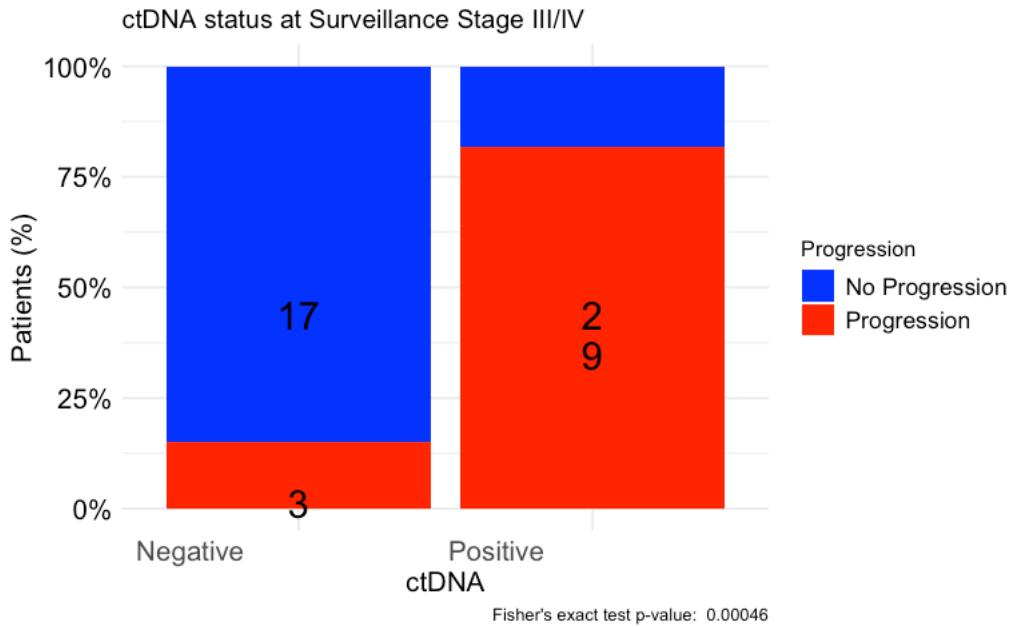
	No Progression	Progression
Negative	17	3
Positive	2	9

[Hide](#)

```

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

```



#PFS by ctDNA status at surveillance p16(+)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$p16.status=="Positive",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_dataadf <- as.data.frame(circ_data)

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

```

```

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

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	30	0	NA	NA	NA
ctDNA.Surveillance=POSITIVE	4	3	19.6	8.18	NA

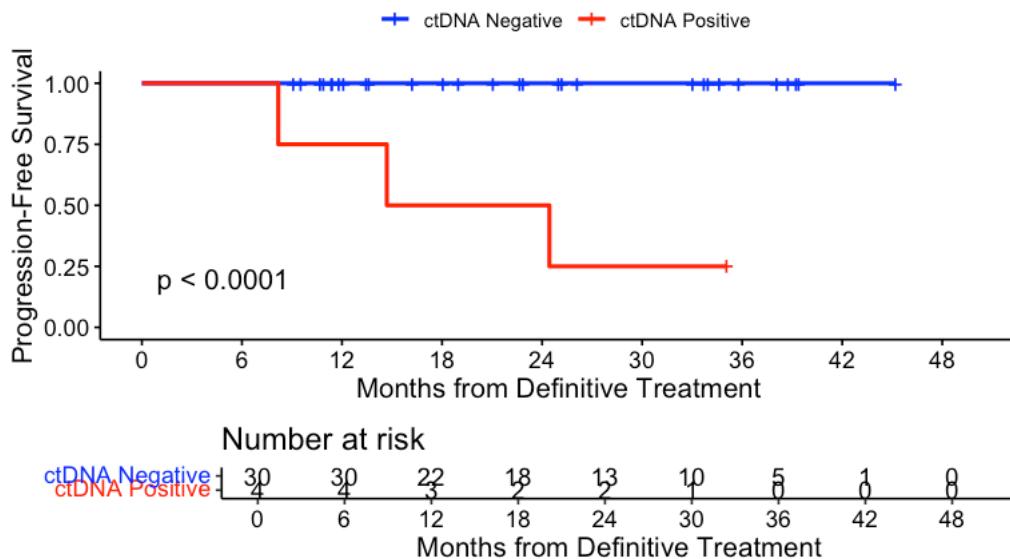
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	30	0	0.00	0
POSITIVE	4	3	0.75	75
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = TRUE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("blue","red"), title="PFS - ctDNA at Surveillance p16(+)", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

### PFS - ctDNA at Surveillance p16(+)



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     30      0       1      0       1       1
 12    22      0       1      0      NA      NA
 24    13      0       1      0      NA      NA
 36     5      0       1      0      NA      NA
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0      4      0     1.00  0.000    1.0000  1.000
 12      3      1     0.75  0.217    0.1279  0.961
 24      2      1     0.50  0.250    0.0578  0.845
```

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxphf(surv_object ~ ctDNA.Surveillance, data=circ_data)
summary(cox_fit)
```

```
coxphf(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)
```

```
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood
```

```
coef se(coef) exp(coef) lower 0.95 upper 0.95      Chisq      p
ctDNA.SurveillancePOSITIVE 3.859906 1.746986 47.46087 4.594352 6384.72 11.46564 0.0007089475
```

```
Likelihood ratio test=11.46564 on 1 df, p=0.0007089475, n=34
Wald test = 4.881741 on 1 df, p = 0.02714223
```

```
Covariance-Matrix:
```

```
           ctDNA.SurveillancePOSITIVE
ctDNA.SurveillancePOSITIVE           3.051959
```

Hide

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

```
coxphf(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)
```

```
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood
```

```
coef se(coef) exp(coef) lower 0.95 upper 0.95      Chisq      p
ctDNA.SurveillancePOSITIVE 3.859906 1.746986 47.46087 4.594352 6384.72 11.46564 0.0007089475
```

```
Likelihood ratio test=11.46564 on 1 df, p=0.0007089475, n=34
Wald test = 4.881741 on 1 df, p = 0.02714223
```

```
Covariance-Matrix:
```

```
           ctDNA.SurveillancePOSITIVE
ctDNA.SurveillancePOSITIVE           3.051959
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

Hide

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 16.235, df = 1, p-value = 5.594e-05
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.0006684
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
4.703171      Inf
sample estimates:
odds ratio
      Inf
```

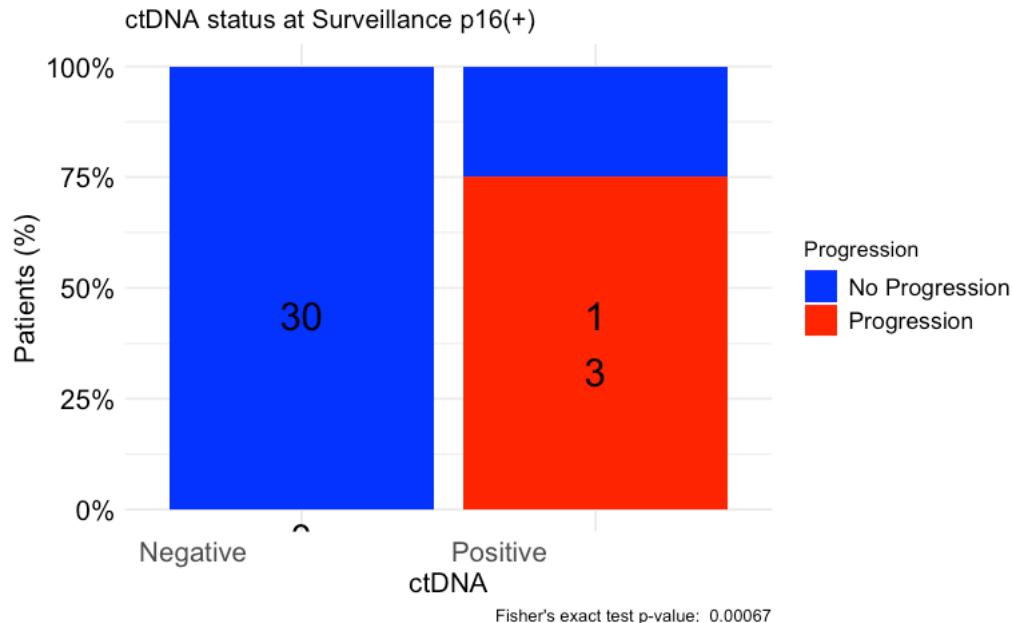
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	30	0
Positive	1	3

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA status at Surveillance p16(+)",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Progression",
       caption = paste("Fisher's exact test p-value: ", format.pval(fisher_exact_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Progression" = "blue", "Progression" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Progression label size
```



#PFS by ctDNA status at surveillance p16(-)

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$p16.status=="Negative",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_datadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	21	3	NA	NA	NA
ctDNA.Surveillance=POSITIVE	13	10	15.5	11.3	NA

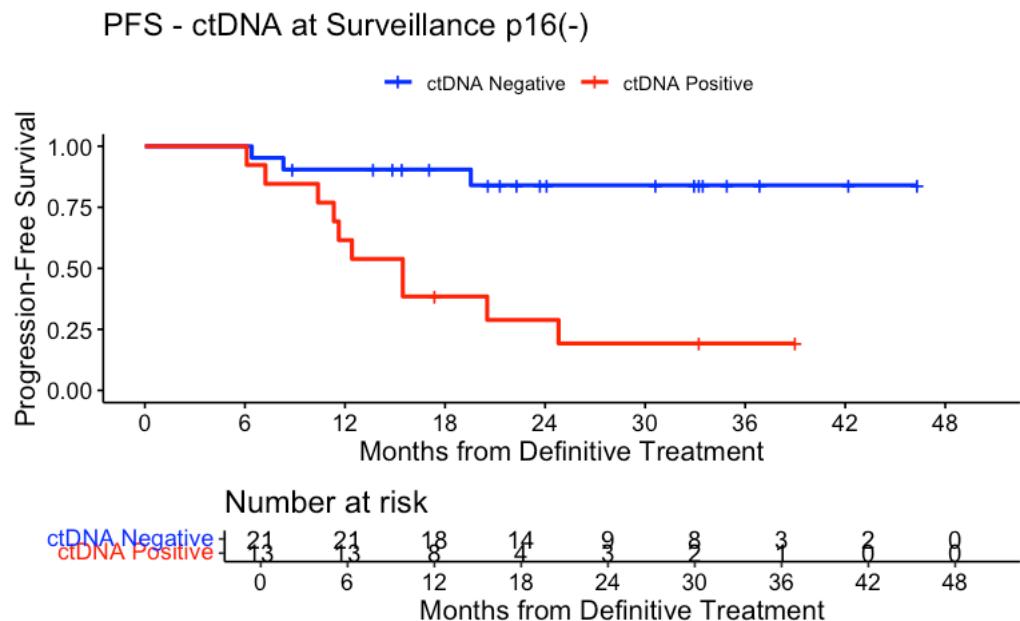
Hide

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

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	21	3	0.1428571	14.28571
POSITIVE	13	10	0.7692308	76.92308
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="PFS - ctDNA at Surveillance p16(-)", ylab= "Progression-Free Survival", xlab="Months fr
om Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

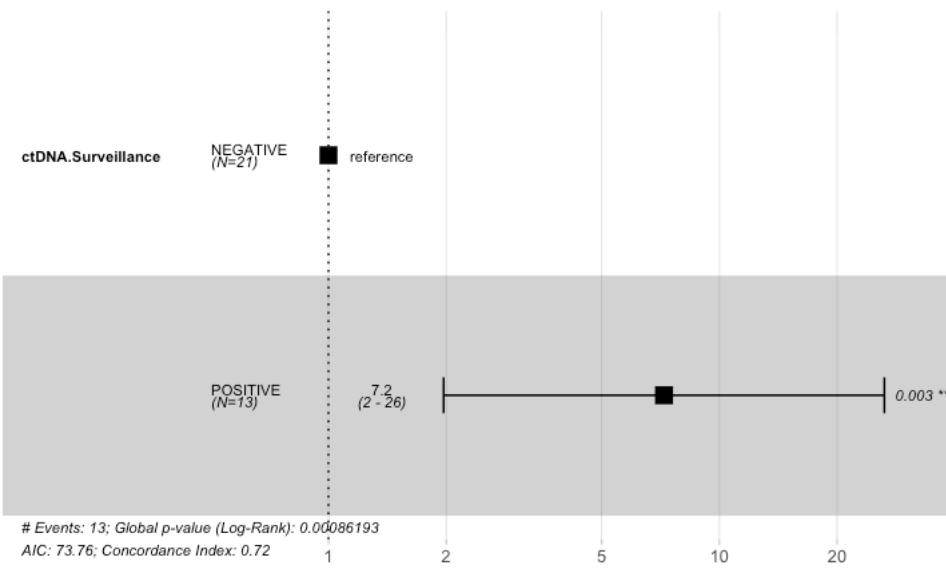
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    21      0    1.000  0.0000    1.000    1.000
 12    18      2    0.905  0.0641    0.670    0.975
 24     9      1    0.840  0.0861    0.576    0.947
 36     3      0    0.840  0.0861    0.576    0.947
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    13      0    1.000  0.000    1.0000    1.000
 12    8      5    0.615  0.135    0.3083    0.818
 24    3      4    0.288  0.131    0.0785    0.545
 36    1      1    0.192  0.117    0.0331    0.450
```

Hide

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 34, number of events= 13

            coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.SurveillancePOSITIVE 1.9761     7.2142   0.6622  2.984  0.00285 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE     7.214      0.1386    1.97     26.42 

Concordance= 0.717 (se = 0.067 )
Likelihood ratio test= 11.1 on 1 df,  p=9e-04
Wald test             = 8.9 on 1 df,  p=0.003
Score (logrank) test = 12.05 on 1 df,  p=5e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 7.21 (1.97-26.42); p = 0.003"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect g
```

[Hide](#)

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 10.819, df = 1, p-value = 0.001005
```

[Hide](#)

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

```
Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.0006471
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.675136 170.607840
sample estimates:
odds ratio
 17.58424
```

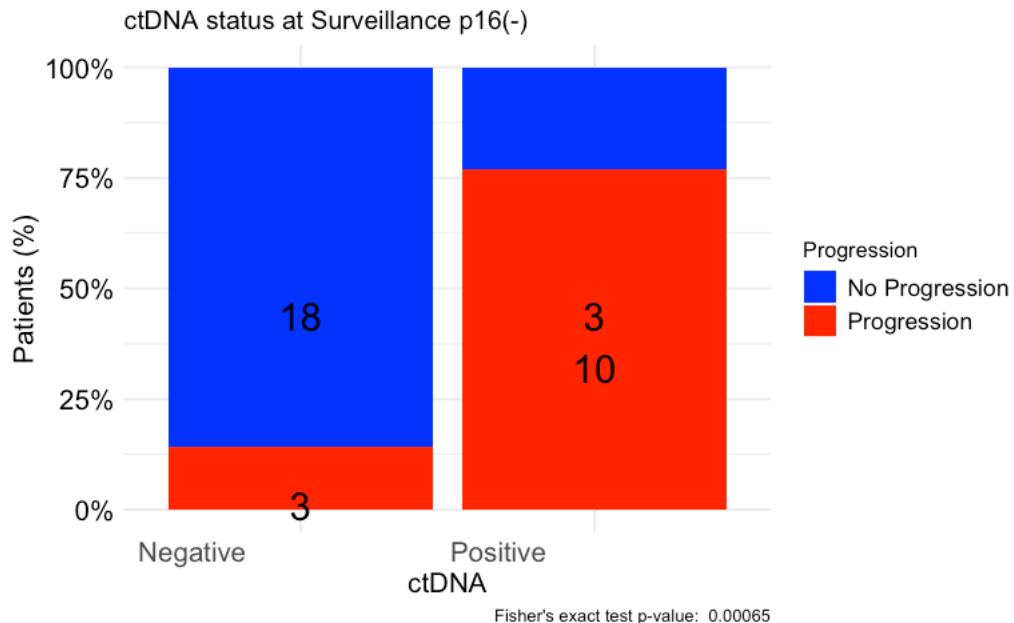
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	18	3
Positive	3	10

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "ctDNA status at Surveillance p16(-)",
       x = "ctDNA",
       y = "Patients (%)",
       fill = "Progression",
       caption = paste("Fisher's exact test p-value: ", format.pval(fisher_exact_test$p.value))) +
  scale_y_continuous(labels = scales::percent_format()) +
  scale_fill_manual(values = c("No Progression" = "blue", "Progression" = "red")) + # define custom colors
  theme(axis.text.x = element_text(angle = 0, hjust = 1.5, size = 14), # increase x-axis text size
        axis.text.y = element_text(size = 14, color = "black"), # increase y-axis text size
        axis.title.x = element_text(size = 14, color = "black"), # increase x-axis label size
        axis.title.y = element_text(size = 14, color = "black"), # increase y-axis label size
        legend.text = element_text(size = 12, color = "black")) # increase Progression label size
```



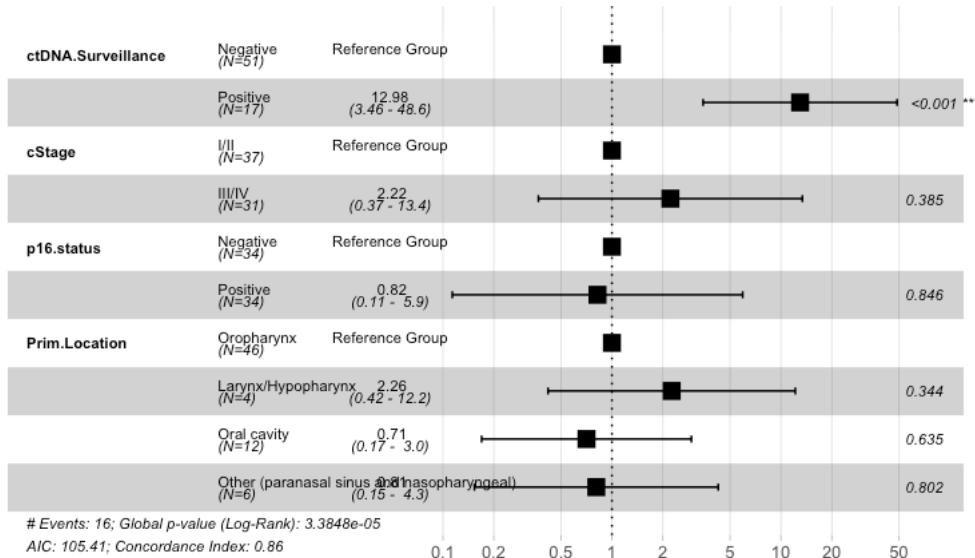
#Multivariate cox regression for PFS ctDNA status at surveillance

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]

circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"), labels = c("Negative", "Positive"))
circ_data$cStage <- factor(circ_data$cStage, levels = c("I/II", "III/IV"))
circ_data$p16.status <- factor(circ_data$p16.status, levels = c("Negative", "Positive"))
circ_data$Prim.Location <- factor(circ_data$Prim.Location, levels = c("Oropharynx", "Larynx/Hypopharynx", "Oral cavity", "Other (paranasal sinus and nasopharyngeal)"))
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance + cStage + p16.status + Prim.Location, data=circ_data)
ggforest(cox_fit, data = circ_data, main = "Multivariate Regression Model for PFS", refLabel = "Reference Group")
```

## Multivariate Regression Model for PFS



Hide

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

#ctDNA and MTM/mL Dynamics for pts at surveillance window

Hide

```
#Dynamics and MTM/mL plots for patients with ctDNA negative at surveillance
rm(list=ls())
setwd("~/Downloads")
df <- read.csv("CLIA HNSCC ctDNA MTM.csv", stringsAsFactors = FALSE)
df <- df[df$ctDNA.Surveillance=="NEGATIVE",]

df$PFS.Event <- ifelse(df$PFS.Event %in% c("No", "no", "FALSE", "False", "0"), FALSE,
                        ifelse(df$PFS.Event %in% c("Yes", "yes", "TRUE", "True", "1"), TRUE, NA))
df$PFS.Event <- factor(df$PFS.Event, levels = c(FALSE, TRUE))
df <- df %>%
  group_by(PatientName) %>%
  filter(n() >= 2) %>% #keep only pts with at least 2 post-surgery time points
  ungroup()

num_unique <- length(unique(df$PatientName))
cat("Number of unique patients:", num_unique, "\n")
```

Number of unique patients: 51

Hide

```
df_patient_pfs <- df %>%
  group_by(PatientName) %>%
  dplyr::summarize(
    PFS_True = any(PFS.Event == TRUE, na.rm = TRUE),
    PFS_False = all(PFS.Event == FALSE, na.rm = TRUE)
  )

num_true <- sum(df_patient_pfs$PFS_True)
num_false <- sum(df_patient_pfs$PFS_False)

cat("Number of unique patients with Event:", num_true, "\n")
```

Number of unique patients with Event: 3

Hide

cat("Number of unique patients with No Event:", num\_false, "\n")

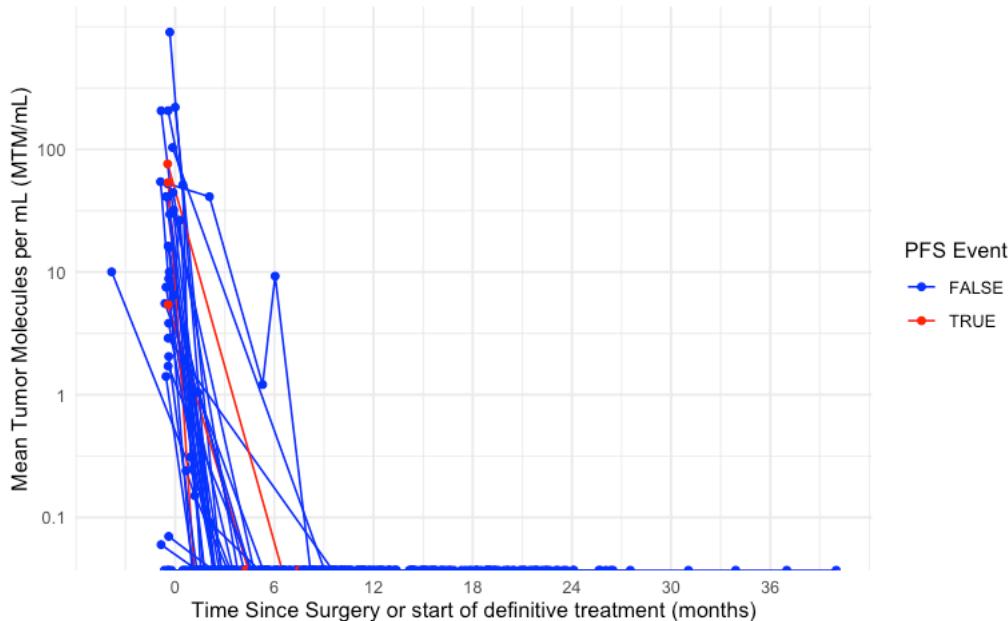
Number of unique patients with No Event: 48

Hide

```

p <- ggplot(df, aes(x = date.diff.months,
                     y = MTM.mL,
                     group = PatientName,
                     color = PFS.Event)) +
  geom_line() +      # Connect timepoints for each patient
  geom_point() +     # Add points for each timepoint
  # Use a log10 scale for the y-axis with specified breaks
  scale_y_log10(breaks = c(0.01, 0.1, 1, 10, 100),
                 labels = c("0.01", "0.1", "1", "10", "100")) +
  scale_x_continuous(breaks = seq(0, max(df$date.diff.months, na.rm = TRUE), by = 6)) +
  scale_color_manual(values = c("FALSE" = "blue", "TRUE" = "red")) +
  labs(
    x = "Time Since Surgery or start of definitive treatment (months)",
    y = "Mean Tumor Molecules per mL (MTM/mL)",
    color = "PFS Event"
  ) +
  theme_minimal()
print(p)

```



Hide

```

#Dynamics and MTM/mL plots for patients with ctDNA positive at surveillance
rm(list=ls())
setwd("~/Downloads")
df <- read.csv("CLIA HNSCC ctDNA MTM.csv", stringsAsFactors = FALSE)
df <- df[df$ctDNA.Surveillance=="POSITIVE",]

df$PFS.Event <- ifelse(df$PFS.Event %in% c("No", "no", "FALSE", "False", "0"), FALSE,
                        ifelse(df$PFS.Event %in% c("Yes", "yes", "TRUE", "True", "1"), TRUE, NA))
df$PFS.Event <- factor(df$PFS.Event, levels = c(FALSE, TRUE))
df <- df %>%
  group_by(PatientName) %>%
  filter(n() >= 2) %>% #keep only pts with at least 2 post-surgery time points
  ungroup()

num_unique <- length(unique(df$PatientName))
cat("Number of unique patients:", num_unique, "\n")

```

Number of unique patients: 17

Hide

```
df_patient_pfs <- df %>%
  group_by(PatientName) %>%
  dplyr::summarize(
    PFS_True = any(PFS.Event == TRUE, na.rm = TRUE),
    PFS_False = all(PFS.Event == FALSE, na.rm = TRUE)
  )

num_true <- sum(df_patient_pfs$PFS_True)
num_false <- sum(df_patient_pfs$PFS_False)

cat("Number of unique patients with Event:", num_true, "\n")
```

Number of unique patients with Event: 13

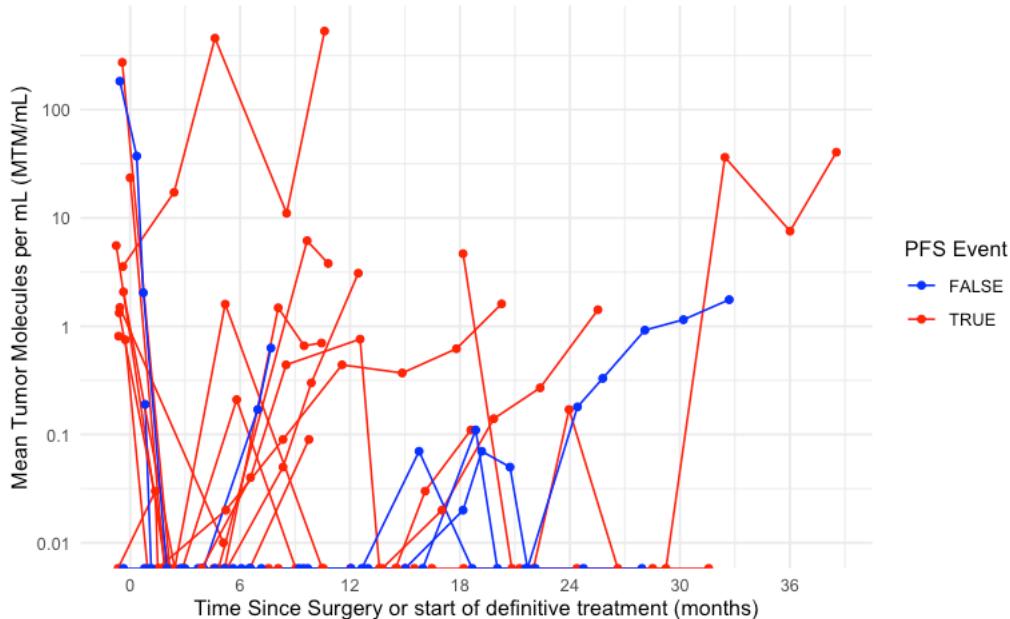
[Hide](#)

```
cat("Number of unique patients with No Event:", num_false, "\n")
```

Number of unique patients with No Event: 4

[Hide](#)

```
p <- ggplot(df, aes(x = date.diff.months,
                      y = MTM.mL,
                      group = PatientName,
                      color = PFS.Event)) +
  geom_line() +      # Connect timepoints for each patient
  geom_point() +     # Add points for each timepoint
  # Use a log10 scale for the y-axis with specified breaks
  scale_y_log10(breaks = c(0.01, 0.1, 1, 10, 100),
                 labels = c("0.01", "0.1", "1", "10", "100")) +
  scale_x_continuous(breaks = seq(0, max(df$date.diff.months, na.rm = TRUE), by = 6)) +
  scale_color_manual(values = c("FALSE" = "blue", "TRUE" = "red")) +
  labs(
    x = "Time Since Surgery or start of definitive treatment (months)",
    y = "Mean Tumor Molecules per mL (MTM/mL)",
    color = "PFS Event"
  ) +
  theme_minimal()
print(p)
```



#ctDNA and MTM/mL Dynamics for pts at surveillance window (excluding baseline & post-progression samples)

[Hide](#)

```
#Dynamics and MTM/mL plots for patients with ctDNA negative at surveillance
rm(list=ls())
setwd("~/Downloads")
df <- read.csv("CLIA HNSCC ctDNA MTM.csv", stringsAsFactors = FALSE)
df <- df[!(df$ctDNA.Window %in% c("Baseline", "Post-PD")), ]
df <- df[df$ctDNA.Surveillance=="NEGATIVE",]

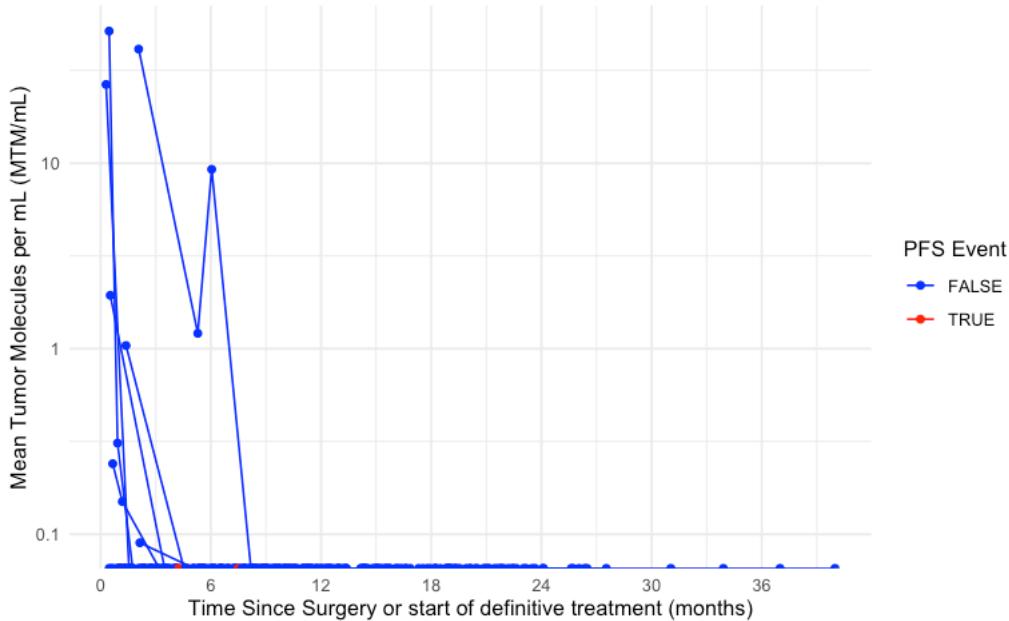
df$PFS.Event <- ifelse(df$PFS.Event %in% c("No", "no", "FALSE", "False", "0"), FALSE,
                        ifelse(df$PFS.Event %in% c("Yes", "yes", "TRUE", "True", "1"), TRUE, NA))
df$PFS.Event <- factor(df$PFS.Event, levels = c(FALSE, TRUE))
df <- df %>%
  group_by(PatientName) %>%
  filter(n() >= 2) %>% #keep only pts with at least 2 post-surgery time points
  ungroup()

num_unique <- length(unique(df$PatientName))
cat("Number of unique patients:", num_unique, "\n")
```

Number of unique patients: 48

[Hide](#)

```
p <- ggplot(df, aes(x = date.diff.months,
                      y = MTM.mL,
                      group = PatientName,
                      color = PFS.Event)) +
  geom_line() +      # Connect timepoints for each patient
  geom_point() +     # Add points for each timepoint
  # Use a log10 scale for the y-axis with specified breaks
  scale_y_log10(breaks = c(0.01, 0.1, 1, 10, 100),
                 labels = c("0.01", "0.1", "1", "10", "100")) +
  scale_x_continuous(breaks = seq(0, max(df$date.diff.months, na.rm = TRUE), by = 6)) +
  scale_color_manual(values = c("FALSE" = "blue", "TRUE" = "red")) +
  labs(
    x = "Time Since Surgery or start of definitive treatment (months)",
    y = "Mean Tumor Molecules per mL (MTM/mL)",
    color = "PFS Event"
  ) +
  theme_minimal()
print(p)
```



[Hide](#)

```
#Dynamics and MTM/mL plots for patients with ctDNA positive at surveillance
rm(list=ls())
setwd("~/Downloads")
df <- read.csv("CLIA HNSCC ctDNA MTM.csv", stringsAsFactors = FALSE)
df <- df[!(df$ctDNA.Window %in% c("Baseline", "Post-PD")), ]
df <- df[df$ctDNA.Surveillance=="POSITIVE",]

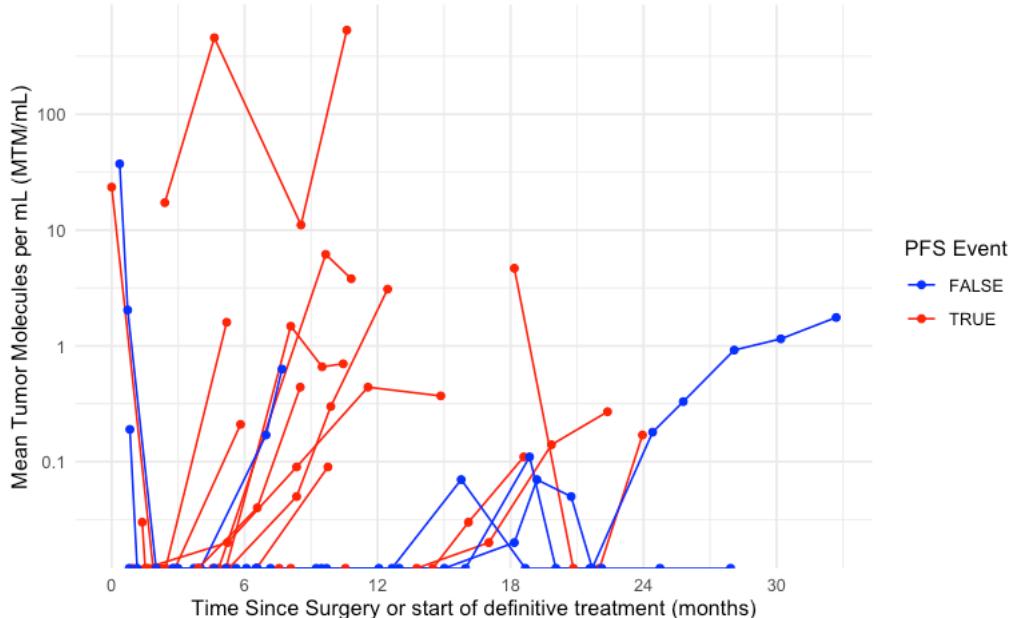
df$PFS.Event <- ifelse(df$PFS.Event %in% c("No", "no", "FALSE", "False", "0"), FALSE,
                        ifelse(df$PFS.Event %in% c("Yes", "yes", "TRUE", "True", "1"), TRUE, NA))
df$PFS.Event <- factor(df$PFS.Event, levels = c(FALSE, TRUE))
df <- df %>%
  group_by(PatientName) %>%
  filter(n() >= 2) %>% #keep only pts with at least 2 post-surgery time points
  ungroup()

num_unique <- length(unique(df$PatientName))
cat("Number of unique patients:", num_unique, "\n")
```

Number of unique patients: 16

[Hide](#)

```
p <- ggplot(df, aes(x = date.diff.months,
                      y = MTM.mL,
                      group = PatientName,
                      color = PFS.Event)) +
  geom_line() +      # Connect timepoints for each patient
  geom_point() +     # Add points for each timepoint
  # Use a log10 scale for the y-axis with specified breaks
  scale_y_log10(breaks = c(0.01, 0.1, 1, 10, 100),
                 labels = c("0.01", "0.1", "1", "10", "100")) +
  scale_x_continuous(breaks = seq(0, max(df$date.diff.months, na.rm = TRUE), by = 6)) +
  scale_color_manual(values = c("FALSE" = "blue", "TRUE" = "red")) +
  labs(
    x = "Time Since Surgery or start of definitive treatment (months)",
    y = "Mean Tumor Molecules per mL (MTM/mL)",
    color = "PFS Event"
  ) +
  theme_minimal()
print(p)
```



#PFS by ctDNA status at surveillance for pts with MRD & Surveillance time points available

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.complete=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_dataadf <- as.data.frame(circ_data)

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

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

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	42	1	NA	NA	NA
ctDNA.Surveillance=POSITIVE	12	8	15.1	11.3	NA

[Hide](#)

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

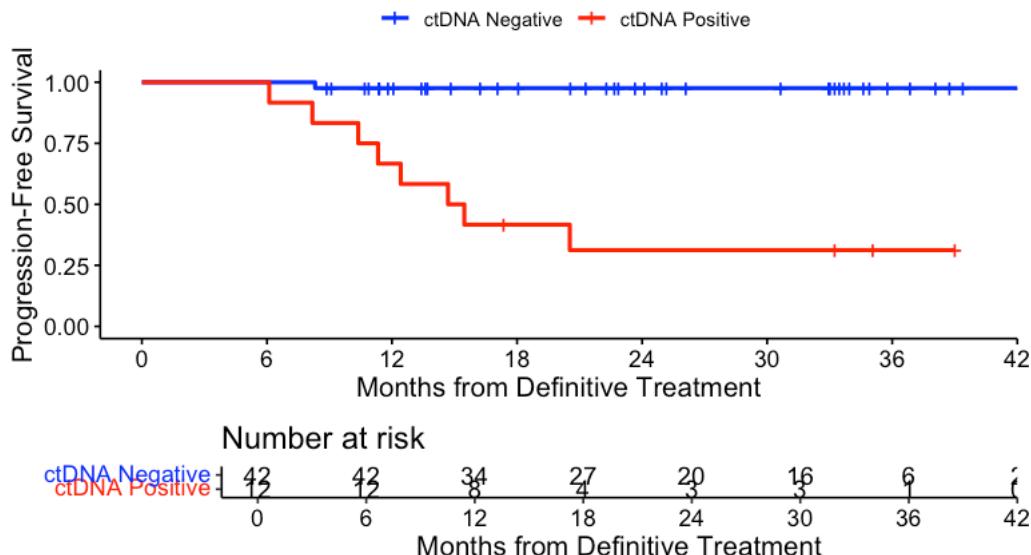
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	42	1	0.02380952	2.380952
POSITIVE	12	8	0.66666667	66.666667

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "red"), title="PFS - ctDNA at Surveillance", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

## PFS - ctDNA at Surveillance



Hide

```
summary(KM_curve, times= c(0, 12, 24, 36))
```

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

ctDNA.Surveillance=NEGATIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	42	0	1.000	0.0000	1.000	1.000	1.000	1.000
12	34	1	0.976	0.0235	0.843	0.997		
24	20	0	0.976	0.0235	0.843	0.997		
36	6	0	0.976	0.0235	0.843	0.997		

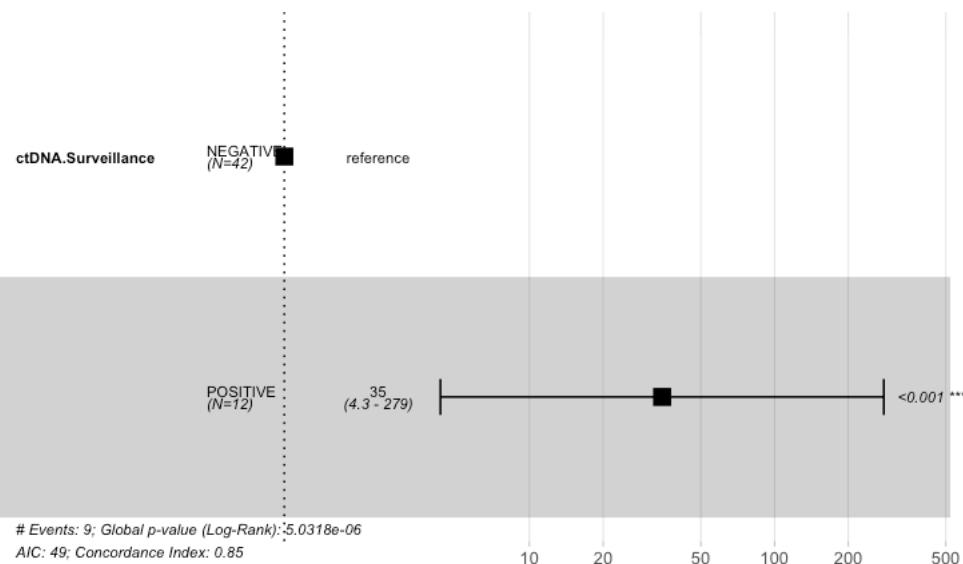
ctDNA.Surveillance=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	12	0	1.000	0.000	1.0000	1.000	1.000	1.000
12	8	4	0.667	0.136	0.3370	0.860		
24	3	4	0.312	0.140	0.0845	0.578		
36	1	0	0.312	0.140	0.0845	0.578		

Hide

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

## Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 54, number of events= 9

            coef exp(coef) se(coef)   z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 3.549    34.774    1.062 3.34 0.000837 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    34.77     0.02876     4.334     279

Concordance= 0.845 (se = 0.063 )
Likelihood ratio test= 20.83 on 1 df,  p=5e-06
Wald test             = 11.16 on 1 df,  p=8e-04
Score (logrank) test = 28.65 on 1 df,  p=9e-08
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 34.77 (4.33-279); p = 0.001"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 23.336, df = 1, p-value = 1.361e-06
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 3.951e-06
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 6.898669 3612.916842
sample estimates:
odds ratio
 69.02378
```

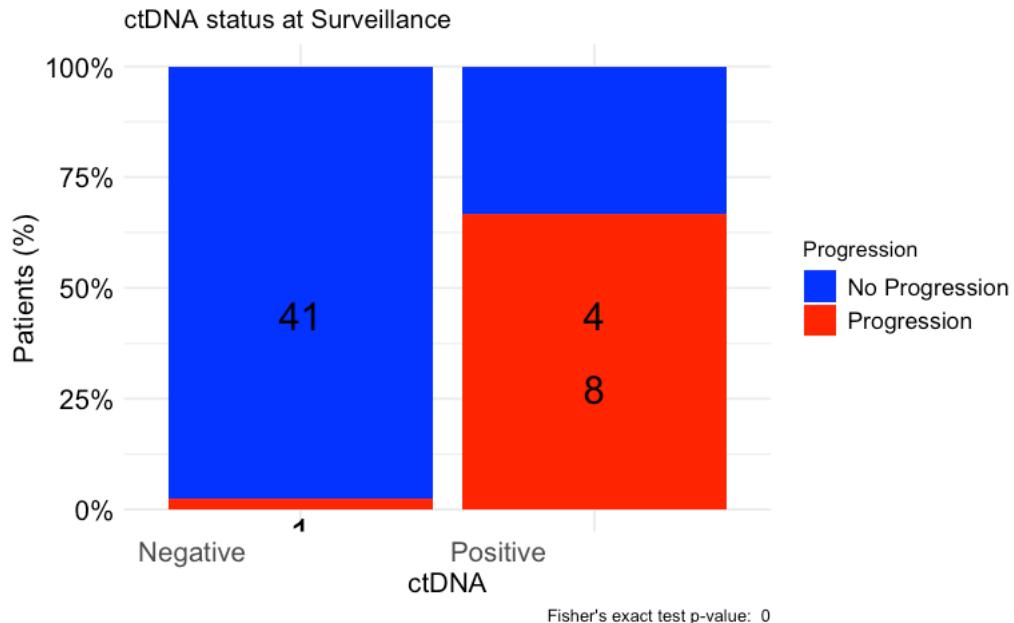
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	41	1
Positive	4	8

[Hide](#)

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



#Time-dependent analysis for PFS in longitudinal time points for pts with MRD & Surveillance time points available

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
dt <- read_xlsx("CLIA HNSCC Peddada Clinical Data_Time dependent.xlsx") |>
  clean_names() |>
  mutate(across(.cols = c(window_start_date,dfs_date,
                         surveillance_1_date:surveillance_12_date),
               .fns = ~ as_date(as.Date(.x, format = "%Y-%m-%d"))))
```

Hide

```

dt biomarker <- dt |>
  select(pts_id, ct_dna_complete,
         window_start_date,
         surveillance_1_status:surveillance_12_date) |>
  filter(ct_dna_complete) |>
  pivot_longer(cols = surveillance_1_status:surveillance_12_date,
               names_to = c("visit_number", ".value"),
               names_pattern = "surveillance_(.)_(.*)") |>
  mutate( biomarker_time = day(dates(date - window_start_date))) |>
  select(pts_id, biomarker_time, biomarker_status = status) |>
  filter(!is.na( biomarker_time))

glimpse(dt biomarker)

```

Hide

```
Rows: 183
Columns: 3
#> #> #> pts_id      <chr> "UNM-004", "UNM-004", "UNM-004", "UNM-008", "UNM-008", "UNM-008", "UNM-008", "UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-009", "UNM-014", "UNM-016", ...
#> #> #> biomarker_time <dbl> 18, 25719, 179, -75, 25647, 154, 236, 322, 46, 25792, 327, 418, 507, 156, 112, 25865, 38
#> #> #> 7, 481, 19, 25756, 273, 361, 454, 550, 649, 32, 25783, 307, 398, 502, 579, 7...
#> #> #> biomarker_status <chr> "NEGATIVE", "POSITIVE", "POSITIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", NA,
#> #> #> "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEG...
```

```
dt_survival <- dt |>
  select(pts_id, ct_dna_complete,
         window_start_date:dfs_date, dfs_event) |> # Added dfs_event here
  filter(ct_dna_complete) |>
  mutate(dfs_time = (dfs_date - window_start_date),
         dfs_time = day(dfs_time)),
         dfs_event = as.numeric(dfs_event)) |>
  select(pts_id, dfs_time, dfs_event)

glimpse(dt_survival)
```

Hide

```

aux <- dt_survival %>%
  filter(dfs_time <= 0)

tab <- left_join(aux, dt) |>
  select(pts_id, window_start_date, dfs_time, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  select(pts_id, window_start_date, dfs_date, dfs_time)

```

Joining with `by = join\_by(pts\_id, dfs\_event)`

Hide

```
datatable(tab, filter = "top")
```

Show 10 entries

**Search:**

pts_id	window_start_date	dfs_date	dfs_time
All	All	All	All

No data available in table

Showing 0 to 0 of 0 entries

Previous Next

Hide

```

dt_survival <- dt_survival |>
  filter(dfs_time > 0)

aux <- dt |>
  select(pts_id, ct_dna_complete,
         window_start_date, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ .x < 0)) |>
  rowwise() |>
  mutate(sum_neg =
        sum(c_across(surveillance_1_date:surveillance_12_date),
            na.rm = TRUE)) |>
  select(pts_id, sum_neg)

tab <- left_join(aux, dt) |>
  filter(sum_neg > 0) |>
  select(pts_id, sum_neg, window_start_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = window_start_date:surveillance_12_date,
               .fns = ~ as_date(.x)))

```

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search: 

	pts_id	sum_neg	window_start_date	surveillance_1_date	surveillance_2_date	surveillance_3_date	surveillance_4_date
1	UNM-008	1	2022-06-30	2022-04-16		2092-09-17	2022-12-01
2	UNM-059	1	2024-09-04	2024-06-24		2095-01-04	

Showing 1 to 2 of 2 entries

Previous

1

Next

```
aux <- dt |>
  select(pts_id, ct_dna_complete,
         window_start_date, dfs_date,
         surveillance_1_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ .x - window_start_date)) |>
  mutate(across(.cols = surveillance_2_date:surveillance_12_date,
               .fns = ~ dfs_date < .x)) |>
  rowwise() |>
  mutate(n_biomarker_after_event = sum(c_across(surveillance_2_date:
                                                surveillance_12_date),
                                       na.rm = TRUE)) |>
  mutate(across(.cols = surveillance_1_date:surveillance_12_date,
               .fns = ~ !is.na(.x))) |>
  mutate(total_biomarker = sum(c_across(surveillance_2_date:
                                         surveillance_12_date),
                               na.rm = TRUE)) |>
  select(pts_id, n_biomarker_after_event, total_biomarker)

temp <- aux |>
  select(-pts_id) |>
  group_by(n_biomarker_after_event, total_biomarker) |> # Direct grouping
  summarise(freq = n(), .groups = "drop") # Drop groups after summarization
```

```
tab <- left_join(aux, dt) |>
  select(pts_id, n_biomarker_after_event, total_biomarker,
         dfs_date,
         surveillance_2_date:surveillance_12_date) |>
  mutate(across(.cols = dfs_date:surveillance_12_date,
               .fns = ~ as_date(.x))) |>
  filter(n_biomarker_after_event > 0)
```

```
Joining with `by = join_by(pts_id)`
G2; H2;Warning h in left_join(aux, dt) :
  Detected an unexpected many-to-many relationship between `x` and `y`.
  i Row 99 of `x` matches multiple rows in `y`.
  i Row 99 of `y` matches multiple rows in `x`.
  i If a many-to-many relationship is expected, set `relationship = "many-to-many"` to silence this warning. g
```

```
datatable(tab, filter = "top")
```

Show 10 ▾ entries

Search:

pts_id	n_biomarker_after_event	total_biomarker	dfs_date	surveillance_2_date	surveillance_3_date	surveillance_4_da
1 UNM-004	1	2	2025-03-17		2094-10-12	2024-11-08
2 UNM-008	1	4	2025-02-07		2092-09-17	2022-12-01
3 UNM-009	1	4	2025-02-07		2094-04-25	2024-08-05
4 UNM-016	1	3	2025-02-07		2094-02-15	2024-05-15
5 UNM-018	1	4	2025-02-07		2092-09-11	2022-12-07
6 UNM-019	1	6	2025-02-07		2092-09-14	2022-12-08
7 UNM-020	1	6	2025-02-07		2093-03-28	2023-06-28
8 UNM-023	1	1	2022-10-05		2092-09-17	
9 UNM-024	1	2	2025-02-07		2094-08-23	2024-11-14
10 UNM-026	1	4	2025-02-07		2094-02-01	2024-05-06

Showing 1 to 10 of 46 entries

Previous 1 2 3 4 5 Next[Hide](#)

```

aux <- tmerge(data1 = dt_survival,
               data2 = dt_survival,
               id = pts_id,
               dfs_event = event(dfs_time, dfs_event))
dt_final <- tmerge(data1 = aux,
                     data2 = dt_biomarker,
                     id = pts_id,
                     biomarker_status =
                     tdc(biomarker_time, biomarker_status))

datatable(dt_final, filter = "top")

```

Show 10 entriesSearch: 

pts_id	dfs_time	dfs_event	tstart	tstop	biomarker_status
1 UNM-004	308	0	0	18	
2 UNM-004	308	0	18	179	NEGATIVE
3 UNM-004	308	1	179	308	POSITIVE
4 UNM-008	953	0	0	154	NEGATIVE
5 UNM-008	953	0	154	236	NEGATIVE
6 UNM-008	953	0	236	953	NEGATIVE
7 UNM-009	513	0	0	46	
8 UNM-009	513	0	46	327	NEGATIVE
9 UNM-009	513	0	327	418	NEGATIVE

pts_id	dfs_time	dfs_event	tstart	tstop	biomarker_status
10 UNM-009	513	0	418	507	NEGATIVE

Showing 1 to 10 of 193 entries

Previous 1 2 3 4 5 ... 20 Next

[Hide](#)

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(dfs_time, dfs_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, dfs_event) ~ biomarker_status,
              data = dt_final)
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, dfs_event) ~ biomarker_status,
      data = dt_final)

n= 141, number of events= 9
(52 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE  3.6249   37.5215   0.7191 5.041 4.63e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE   37.52    0.02665    9.166   153.6

Concordance= 0.795 (se = 0.082 )
Likelihood ratio test= 24.07 on 1 df,  p=9e-07
Wald test             = 25.41 on 1 df,  p=5e-07
Score (logrank) test = 63.8 on 1 df,  p=1e-15
```

[Hide](#)

```
cox_fit_summary <- summary(fit)

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

```
[1] "HR = 37.52 (9.17-153.6); p = 0"
```

#Median numbers of time points and lead time in the longitudinal setting for pts with MRD & Surveillance time points available

[Hide](#)

```
# Load the dataset
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.complete=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!=""]
circ_datadf <- as.data.frame(circ_data)

median_Nsurvtps <- median(circ_datadf$Nsurvtps, na.rm = TRUE)
min_Nsurvtps <- min(circ_datadf$Nsurvtps, na.rm = TRUE)
max_Nsurvtps <- max(circ_datadf$Nsurvtps, na.rm = TRUE)

cat(sprintf("Median # of surveillance time points: %d (%d-%d)\n",
            median_Nsurvtps, min_Nsurvtps, max_Nsurvtps))
```

Median # of surveillance time points: 4 (1-13)

[Hide](#)

```
circ_datadf$LeadTime_Months <- circ_datadf$LeadTime / 30.437
median_LeadTime <- median(circ_datadf$LeadTime_Months, na.rm = TRUE)
min_LeadTime <- min(circ_datadf$LeadTime_Months, na.rm = TRUE)
max_LeadTime <- max(circ_datadf$LeadTime_Months, na.rm = TRUE)
cat(sprintf("Longitudinally, ctDNA positivity preceded progression by a median of %.2f mo (%.2f-%.2f)\n",
            median_LeadTime, min_LeadTime, max_LeadTime))
```

Longitudinally, ctDNA positivity preceded progression by a median of 4.25 mo (0.62-13.96)

#PFS by ctDNA status anytime

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.anytime!=""]
circ_datadf <- as.data.frame(circ_data)

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

Call: survfit(formula = Surv(time = circ\_data\$PFS.months, event = circ\_data\$PFS.Event) ~ ctDNA.anytime, data = circ\_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.anytime=NEGATIVE	55	3	NA	NA	NA
ctDNA.anytime=POSITIVE	30	21	14.7	11.3	24.8

[Hide](#)

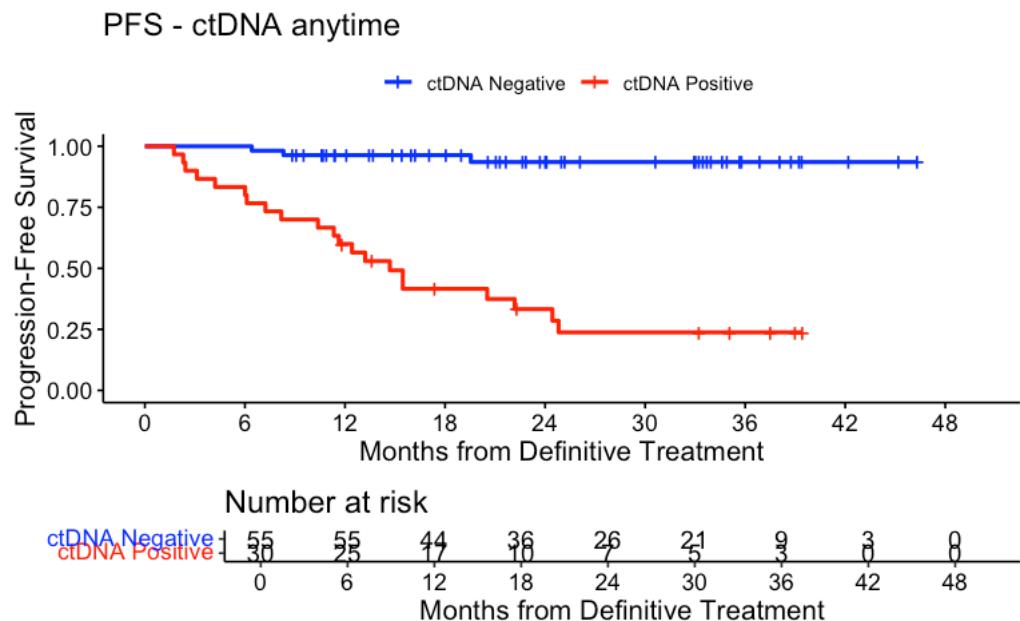
```
event_summary <- circ_data %>%
  group_by(ctDNA.anytime) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.anytime	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	55	3	0.05454545	5.454545
POSITIVE	30	21	0.70000000	70.000000

2 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$PFS.months, event = circ_data$PFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.anytime, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue","red"), title="PFS - ctDNA anytime", ylab= "Progression-Free Survival", xlab="Months from Definitive Treatment", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



```
summary(KM_curve, times= c(0, 12, 24, 36))
```

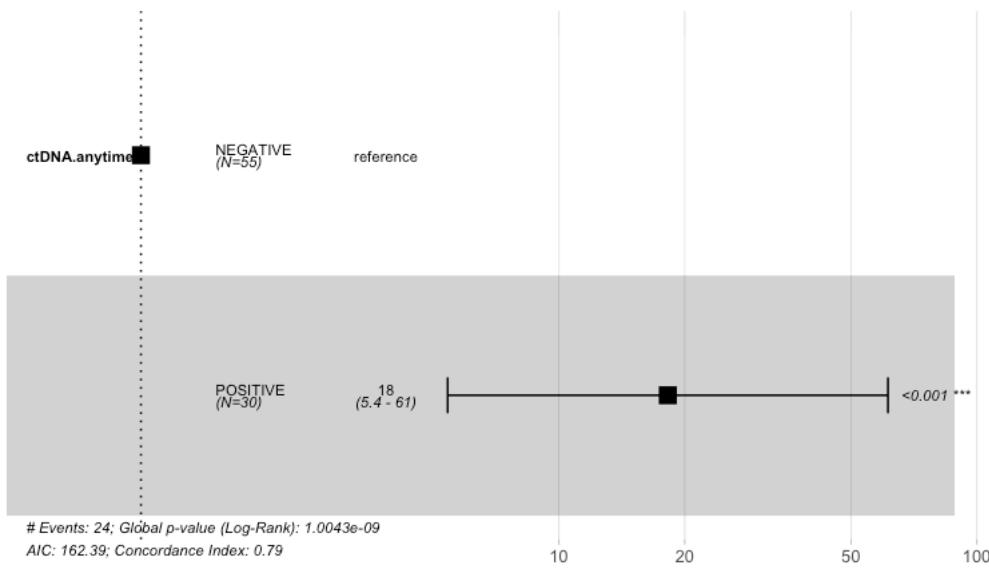
```
Call: survfit(formula = surv_object ~ ctDNA.anytime, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.anytime=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    55      0    1.000  0.0000    1.000    1.000
 12    44      2    0.964  0.0252    0.862    0.991
 24    26      1    0.935  0.0371    0.807    0.979
 36     9      0    0.935  0.0371    0.807    0.979
```

```
ctDNA.anytime=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    30      0    1.000  0.0000    1.0000    1.000
 12    17     12    0.600  0.0894    0.4045    0.750
 24     7      7    0.333  0.0909    0.1670    0.508
 36     3      2    0.238  0.0863    0.0947    0.417
```

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

Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 85, number of events= 24

            coef exp(coef)  se(coef)    z Pr(>|z|)
ctDNA.anytimePOSITIVE  2.9021    18.2128   0.6187 4.69 2.73e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.anytimePOSITIVE    18.21     0.05491     5.416    61.24

Concordance= 0.794  (se = 0.038 )
Likelihood ratio test= 37.32 on 1 df,  p=1e-09
Wald test             = 22 on 1 df,  p=3e-06
Score (logrank) test = 41.85 on 1 df,  p=1e-10
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 18.21 (5.42-61.24); p = 0"
```

Hide

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

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

```
data: contingency_table
X-squared = 36.789, df = 1, p-value = 1.316e-09
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 4.294e-10
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 8.850524 237.891004
sample estimates:
odds ratio
37.70824
```

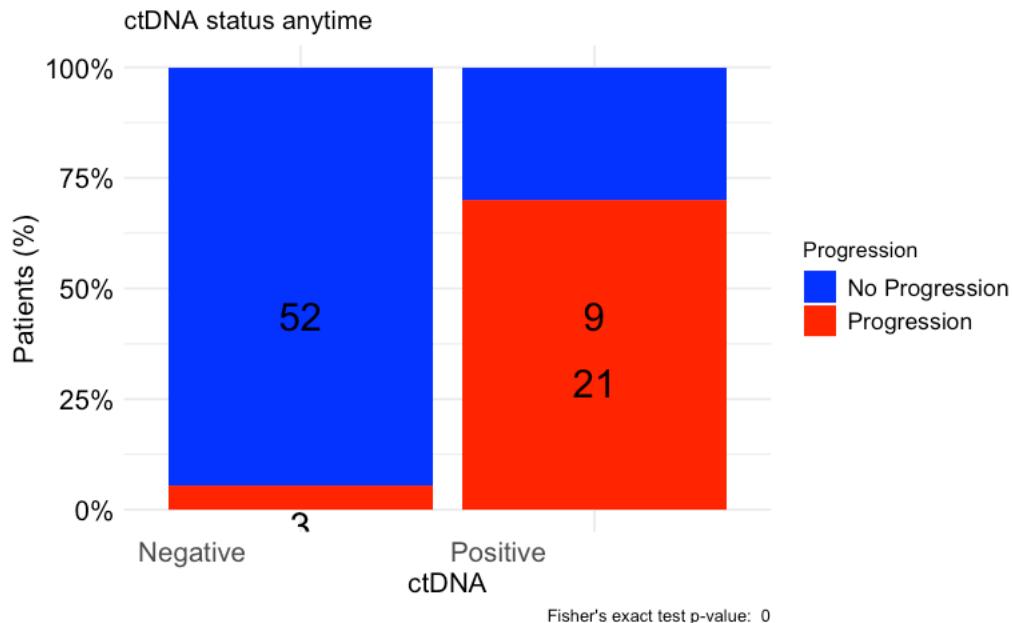
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	52	3
Positive	9	21

[Hide](#)

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



#OS by ctDNA status anytime

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.anytime!=""]
circ_datadf <- as.data.frame(circ_data)

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

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

```
      n  events  median 0.95LCL 0.95UCL
ctDNA.anytime=NEGATIVE 55      1      NA      NA      NA
ctDNA.anytime=POSITIVE 30      7      NA      NA      NA
```

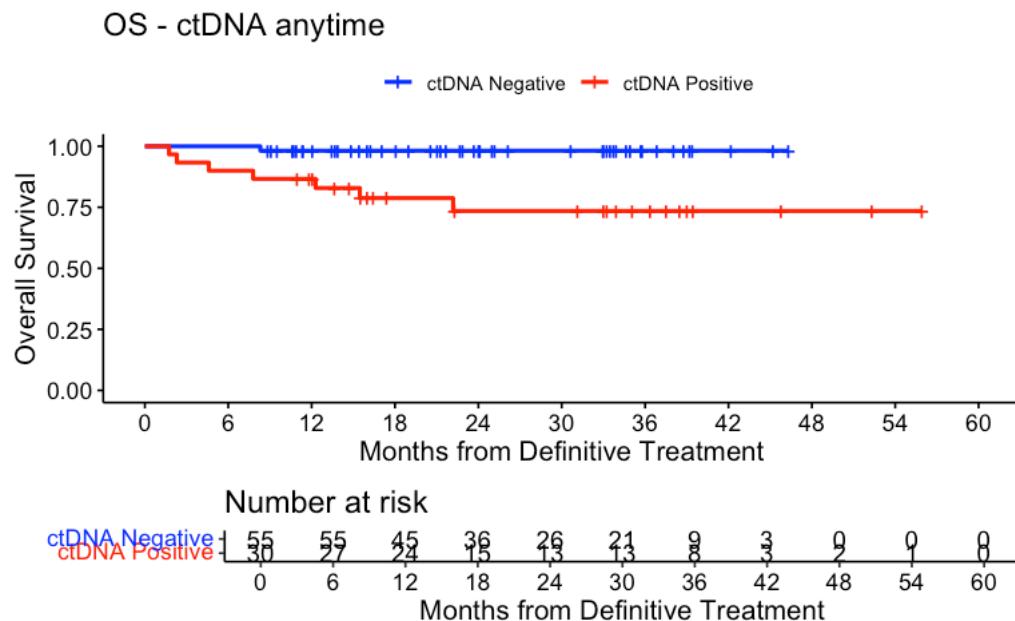
Hide

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

ctDNA.anytime	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	55	1	0.01818182	1.818182
POSITIVE	30	7	0.23333333	23.333333
2 rows				

Hide

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



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

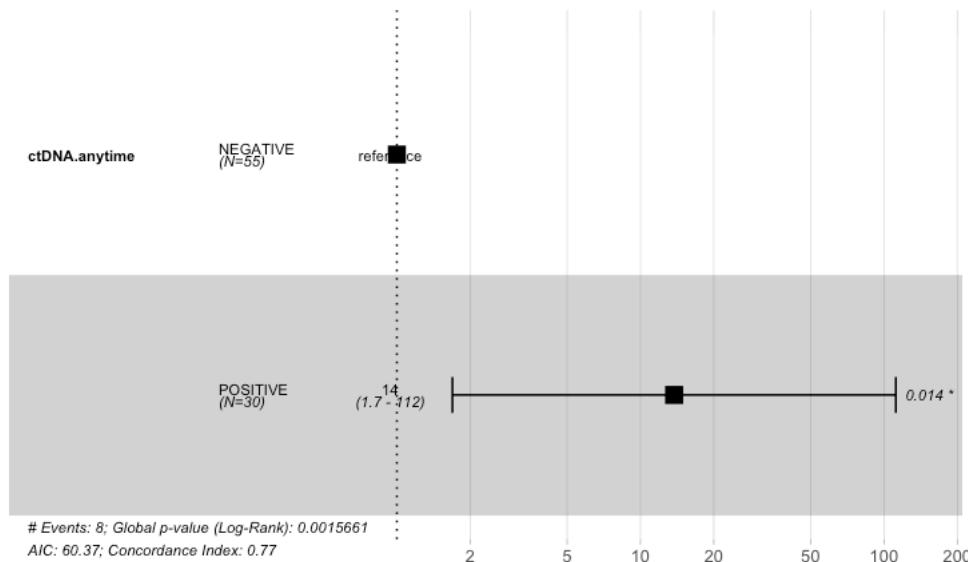
```
Call: survfit(formula = surv_object ~ ctDNA.anytime, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.anytime=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     45      1    0.982   0.018     0.878    0.997
 24     26      0    0.982   0.018     0.878    0.997
 36      9      0    0.982   0.018     0.878    0.997
```

```
ctDNA.anytime=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     24      4    0.867   0.0621    0.683    0.948
 24     13      3    0.735   0.0885    0.516    0.867
 36      8      0    0.735   0.0885    0.516    0.867
```

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

## Hazard ratio



Hide

```
summary(cox_fit)
```

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

n= 85, number of events= 8

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.anytimePOSITIVE 2.621    13.743    1.069 2.451  0.0142 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.anytimePOSITIVE    13.74     0.07276    1.691    111.7

Concordance= 0.766 (se = 0.066 )
Likelihood ratio test= 10  on 1 df,  p=0.002
Wald test             = 6.01  on 1 df,  p=0.01
Score (logrank) test = 10.33  on 1 df,  p=0.001
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 13.74 (1.69-111.72); p = 0.014"
```

Hide

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

```
G2; H2;Warning h in stats::chisq.test(x, y, ...):
Chi-squared approximation may be incorrect
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 8.1668, df = 1, p-value = 0.004266
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.002448
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.874902 750.814710
sample estimates:
odds ratio
 15.89819
```

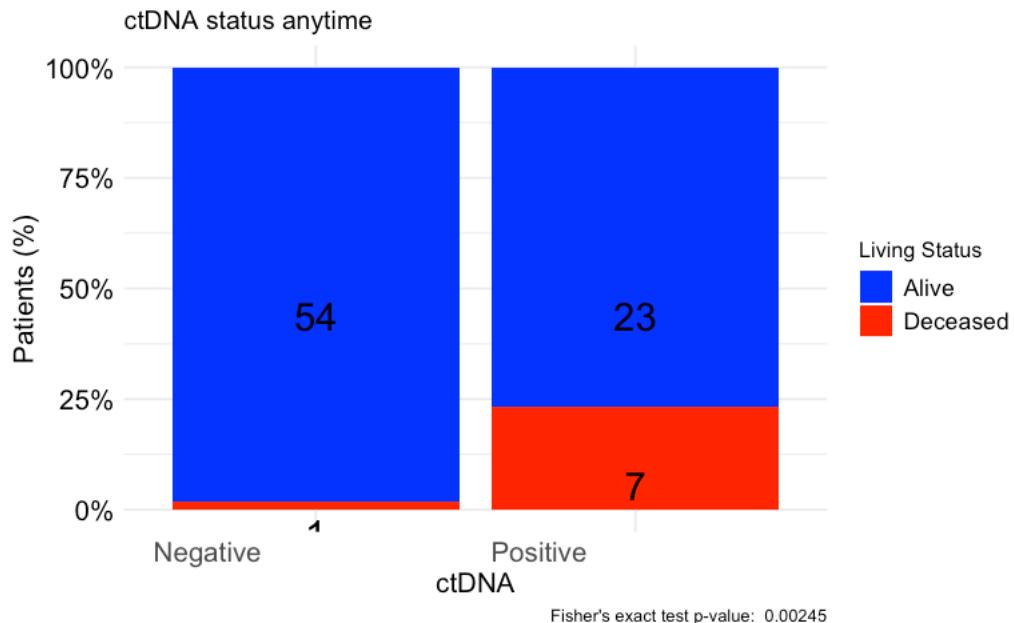
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	54	1
Positive	23	7

[Hide](#)

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



#Median numbers of time points and lead time anytime post-surgery or definitive treatment

[Hide](#)

```
# Load the dataset
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.anytime!=""]
circ_datadf <- as.data.frame(circ_data)

median_Nsurvtps <- median(circ_datadf$Ntotaltps, na.rm = TRUE)
min_Nsurvtps <- min(circ_datadf$Ntotaltps, na.rm = TRUE)
max_Nsurvtps <- max(circ_datadf$Ntotaltps, na.rm = TRUE)

cat(sprintf("Median # of time points anytimes: %d (%d-%d)\n",
            median_Nsurvtps, min_Nsurvtps, max_Nsurvtps))
```

Median # of time points anytimes: 4 (1-16)

[Hide](#)

```
circ_datadf$LeadTime_Months <- circ_datadf$Anytime.LeadTime / 30.437
median_LeadTime <- median(circ_datadf$LeadTime_Months, na.rm = TRUE)
min_LeadTime <- min(circ_datadf$LeadTime_Months, na.rm = TRUE)
max_LeadTime <- max(circ_datadf$LeadTime_Months, na.rm = TRUE)
cat(sprintf("Anytime post-surgery or start of definitive treatment, ctDNA positivity preceded progression by a median of %.2f mo (%.2f-%.2f)\n",
            median_LeadTime, min_LeadTime, max_LeadTime))
```

Anytime post-surgery or start of definitive treatment, ctDNA positivity preceded progression by a median of 3.25 mo (0.00-21.49)