

Code ▾

CLIA HNSCC UNM Clinical Analysis 07272025

```
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

[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_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 ¹
Sex	
Female	17 (18%)
Male	80 (82%)
Age	66 (29 - 95)
Tobacco.History	63 (65%)
Prim.Location	
Larynx/Hypopharynx	5 (5.2%)
Oral cavity	16 (16%)

¹ n (%); Median (Min - Max)

Characteristic	N = 97 ¹
Oropharynx	67 (69%)
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)
cT	
T0	2 (2.1%)
T1	12 (12%)
T2	31 (32%)
T3	30 (31%)
T4	21 (22%)
TX	1 (1.0%)
cN	
N0	22 (23%)
N1	33 (34%)
N2	33 (34%)
N3	9 (9.3%)
cM	
M0	93 (96%)
M1	4 (4.1%)
Histology	
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%)
Stage	
I/II	49 (51%)
III/IVA/IVB	45 (46%)
IVC	3 (3.1%)
p16.status	
Negative	43 (44%)
Positive	54 (56%)
Treatment.Group	
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%)
PFS.Event	
No Progression	64 (66%)
Progression	33 (34%)
OS.Event	
¹ n (%); Median (Min - Max)	

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

¹ n (%); Median (Min - Max) Hide

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

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

¹n (%); Median (Min - Max)

Characteristic	N = 97 ¹
p16.status	
Negative	43 (44%)
Positive	54 (56%)
Treatment.Group	
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%)
PFS.Event	
No Progression	64 (66%)
Progression	33 (34%)
OS.Event	
Alive	81 (84%)
Deceased	16 (16%)
OS.months	
	22 (2 - 56)

¹n (%); 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 ¹
Sex	
Female	17 (18%)
Male	80 (82%)
Age	
	66 (29 - 95)
Tobacco.History	
	63 (65%)
Prim.Location	
Larynx/Hypopharynx	5 (5.2%)
Oral cavity	16 (16%)
Oropharynx	67 (69%)
Other (paranasal sinus and nasopharyngeal)	9 (9.3%)
cT	
T0	2 (2.1%)
T1	12 (12%)
T2	31 (32%)
T3	30 (31%)
T4	21 (22%)
TX	1 (1.0%)
cN	
N0	22 (23%)
N1	33 (34%)
N2	33 (34%)
N3	9 (9.3%)
cM	
M0	93 (96%)
M1	4 (4.1%)
Histology	
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%)
Stage	
I/II	49 (51%)
III/IVA/IVB	45 (46%)
IVC	3 (3.1%)
p16.status	
Negative	43 (44%)
Positive	54 (56%)

¹ n (%); Median (Min - Max)

Characteristic	N = 97 ¹
Treatment.Group	
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%)
PFS.Event	
No Progression	64 (66%)
Progression	33 (34%)
OS.Event	
Alive	81 (84%)
Deceased	16 (16%)
OS.months	
	22 (2 - 56)

¹ n (%); Median (Min - Max)

```
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.

ByctDNA_MRD

Characteristic	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
Sex			0.10
Female	3 (43%)	8 (15%)	
Male	4 (57%)	47 (85%)	
Age	80 (53 - 95)	65 (37 - 95)	0.081
Tobacco.History	4 (57%)	37 (67%)	0.7
Prim.Location			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%)	
cT			0.050
T0	0 (0%)	2 (3.6%)	
T1	1 (14%)	4 (7.3%)	

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

Characteristic	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
T2	2 (29%)	19 (35%)	
T3	0 (0%)	21 (38%)	
T4	4 (57%)	9 (16%)	
TX	0 (0%)	0 (0%)	
cN			>0.9
N0	1 (14%)	11 (20%)	
N1	3 (43%)	19 (35%)	
N2	3 (43%)	19 (35%)	
N3	0 (0%)	6 (11%)	
cM			>0.9
M0	7 (100%)	53 (96%)	
M1	0 (0%)	2 (3.6%)	
Histology			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%)	
Stage			0.4
I/II	2 (29%)	32 (58%)	
III/IVA/IVB	5 (71%)	21 (38%)	
IVC	0 (0%)	2 (3.6%)	
p16.status			0.090
Negative	5 (71%)	18 (33%)	
Positive	2 (29%)	37 (67%)	
Treatment.Group			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%)	
PFS.Event			0.4
No Progression	6 (86%)	35 (64%)	
Progression	1 (14%)	20 (36%)	
OS.Event			0.3
Alive	7 (100%)	44 (80%)	
Deceased	0 (0%)	11 (20%)	
OS.months	31 (21 - 39)	16 (2 - 45)	0.028

¹ n (%); Median (Min - Max)² 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 ¹	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
Sex				
Female	17 (18%)	3 (43%)	8 (15%)	
Male	80 (82%)	4 (57%)	47 (85%)	
Age	66 (29 - 95)	80 (53 - 95)	65 (37 - 95)	0.081
Tobacco.History	63 (65%)	4 (57%)	37 (67%)	0.7
Prim.Location				
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%)	
cT				
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%)	
cN				
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%)	
cM				
M0	93 (96%)	7 (100%)	53 (96%)	
M1	4 (4.1%)	0 (0%)	2 (3.6%)	
Histology				
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%)	
Stage				
I/II	49 (51%)	2 (29%)	32 (58%)	
III/IV/A/IVB	45 (46%)	5 (71%)	21 (38%)	
IVC	3 (3.1%)	0 (0%)	2 (3.6%)	

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

Characteristic	Table 1		Table 2	
	N = 97 ¹	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
p16.status				
Negative	43 (44%)	5 (71%)	18 (33%)	0.090
Positive	54 (56%)	2 (29%)	37 (67%)	
Treatment.Group				
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%)	
PFS.Event				
No Progression	64 (66%)	6 (86%)	35 (64%)	
Progression	33 (34%)	1 (14%)	20 (36%)	
OS.Event				
Alive	81 (84%)	7 (100%)	44 (80%)	
Deceased	16 (16%)	0 (0%)	11 (20%)	
OS.months	22 (2 - 56)	31 (21 - 39)	16 (2 - 45)	0.028

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

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

Characteristic	Table 1		Table 2	
	N = 97 ¹	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
Sex				
Female	17 (18%)	3 (43%)	8 (15%)	
Male	80 (82%)	4 (57%)	47 (85%)	
Age	66 (29 - 95)	80 (53 - 95)	65 (37 - 95)	0.081
Tobacco.History	63 (65%)	4 (57%)	37 (67%)	0.7
Prim.Location				
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%)	
cT				
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%)	

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

Characteristic	Table 1		Table 2	
	N = 97 ¹	Negative N = 7 ¹	Positive N = 55 ¹	p-value ²
T4	21 (22%)	4 (57%)	9 (16%)	
TX	1 (1.0%)	0 (0%)	0 (0%)	
cN				>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%)	
cM				>0.9
M0	93 (96%)	7 (100%)	53 (96%)	
M1	4 (4.1%)	0 (0%)	2 (3.6%)	
Histology				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%)	
Stage				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%)	
p16.status				0.090
Negative	43 (44%)	5 (71%)	18 (33%)	
Positive	54 (56%)	2 (29%)	37 (67%)	
Treatment.Group				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%)	
PFS.Event				0.4
No Progression	64 (66%)	6 (86%)	35 (64%)	
Progression	33 (34%)	1 (14%)	20 (36%)	
OS.Event				0.3
Alive	81 (84%)	7 (100%)	44 (80%)	
Deceased	16 (16%)	0 (0%)	11 (20%)	
OS.months	22 (2 - 56)	31 (21 - 39)	16 (2 - 45)	0.028

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

Hide

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

#Overview plot by Stage

Hide

```

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'))

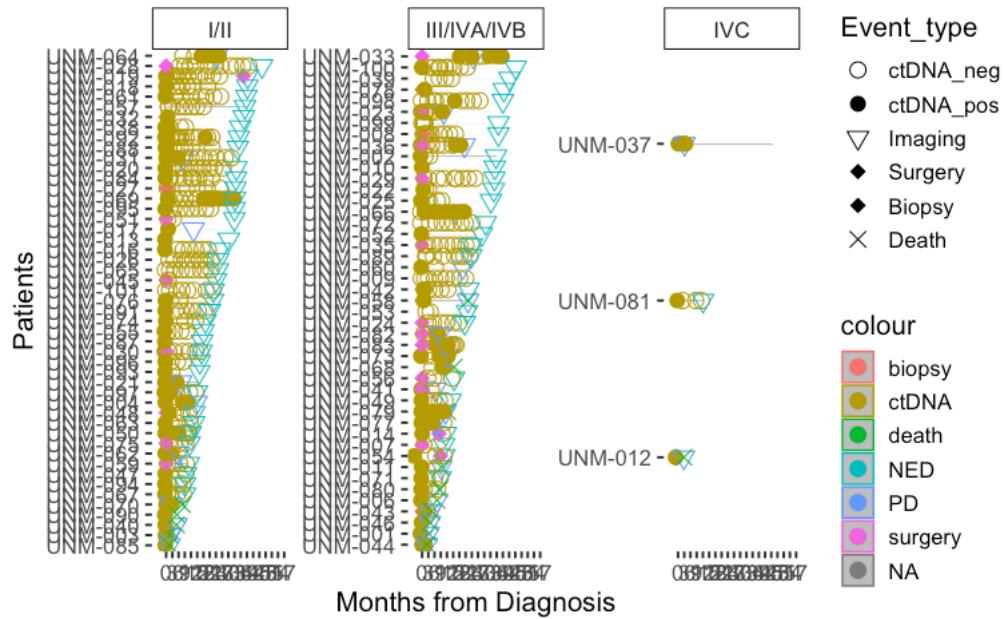
# Adding themes and scales
oplot <- oplot + theme(panel.border = element_blank())
oplot <- oplot + scale_y_continuous(breaks = seq(0, 72, by = 3))
oplot <- oplot + labs(x = "Patients", y="Months from Diagnosis")

# 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

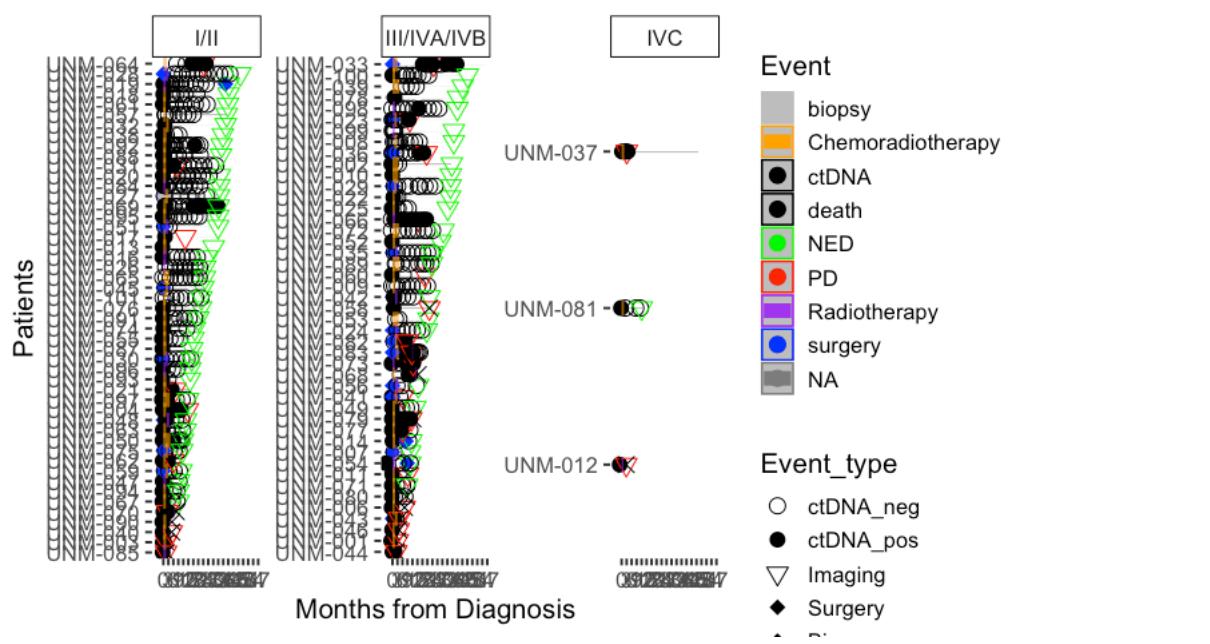
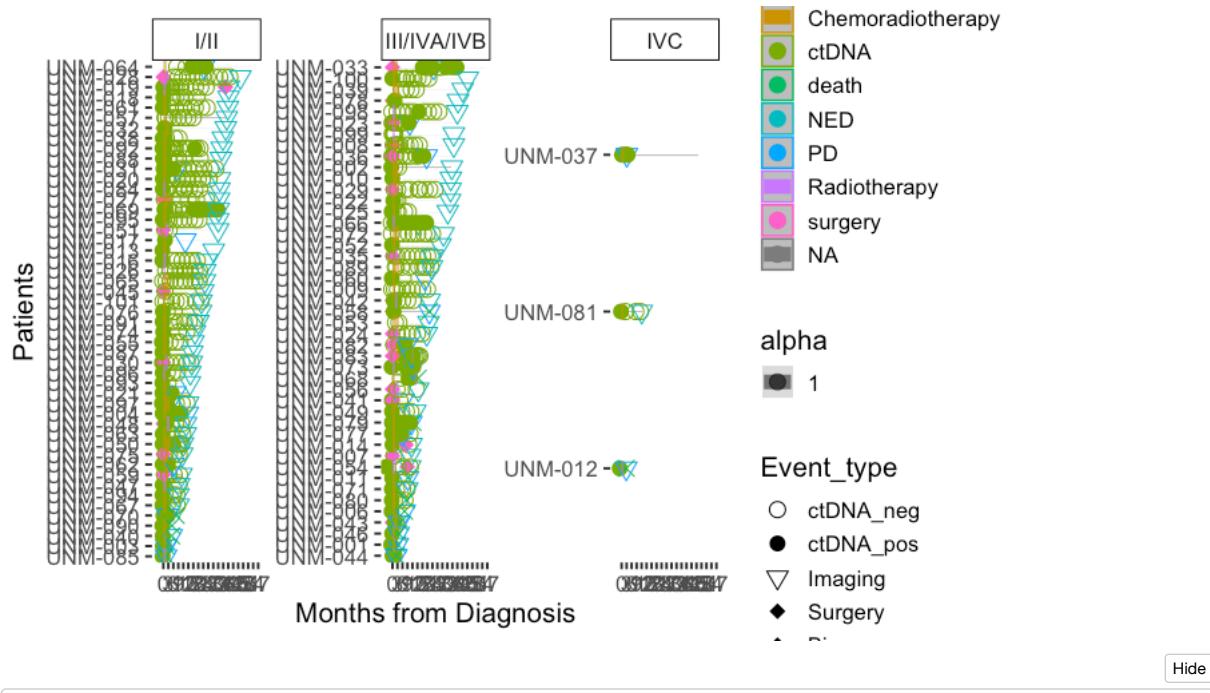
```



```

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

```



#PFS in Complete Cohort (N=97)

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",]

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)
```

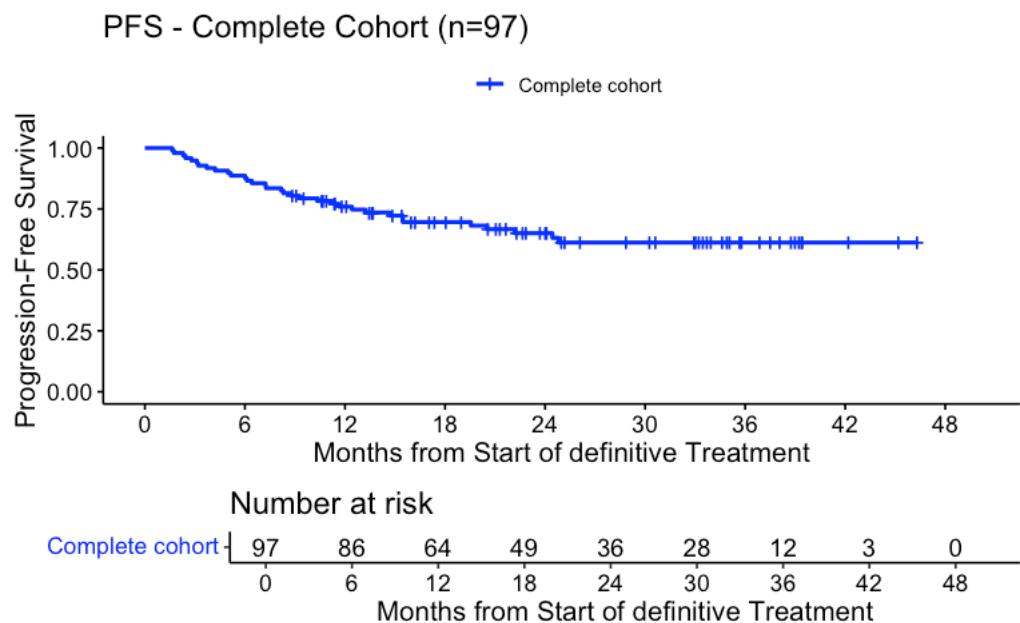
```
n events median 0.95LCL 0.95UCL
[1,] 97      33     NA     NA     NA
```

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=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     64      23    0.760  0.0438      0.660    0.833
  24     36       8    0.651  0.0520      0.538    0.742
  36     12       2    0.612  0.0555      0.494    0.711

```

#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)
```

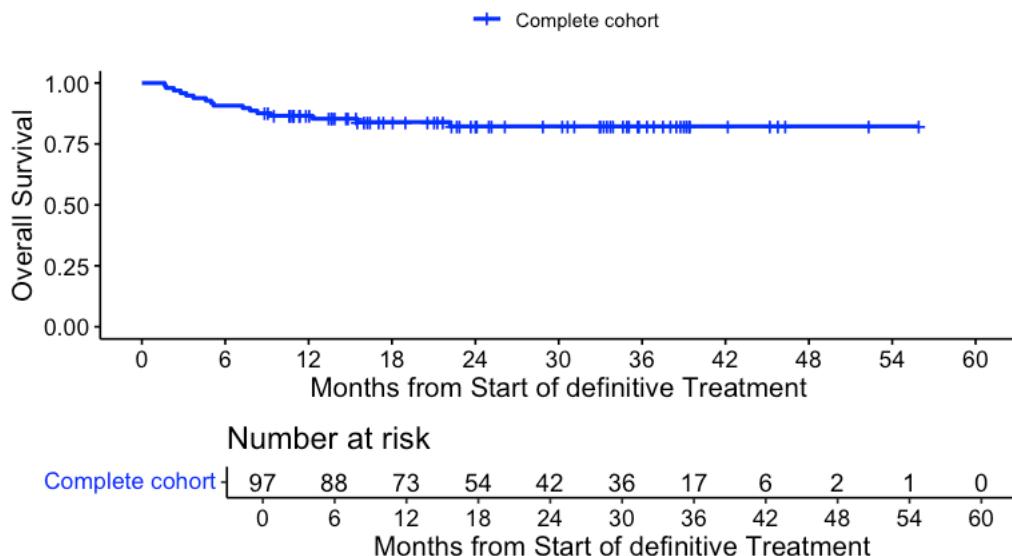
```
Call: survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~
  ctDNA.available, data = circ_data)
```

```
  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="")
```

OS - Complete Cohort (n=97)



[Hide](#)

```
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	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)

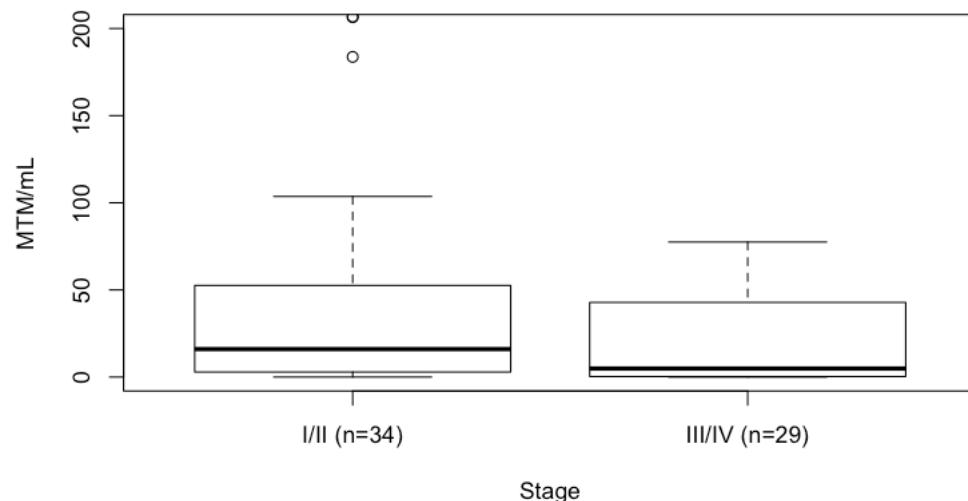
tally(~cStage, data=circ_data, margins = TRUE)
```

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)

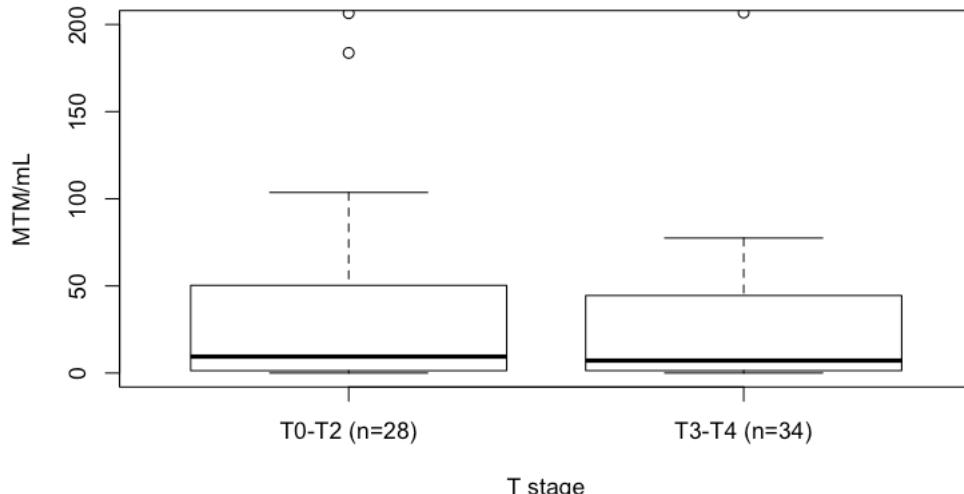
tally(~cT.status, data=circ_data, margins = TRUE)
```

```
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

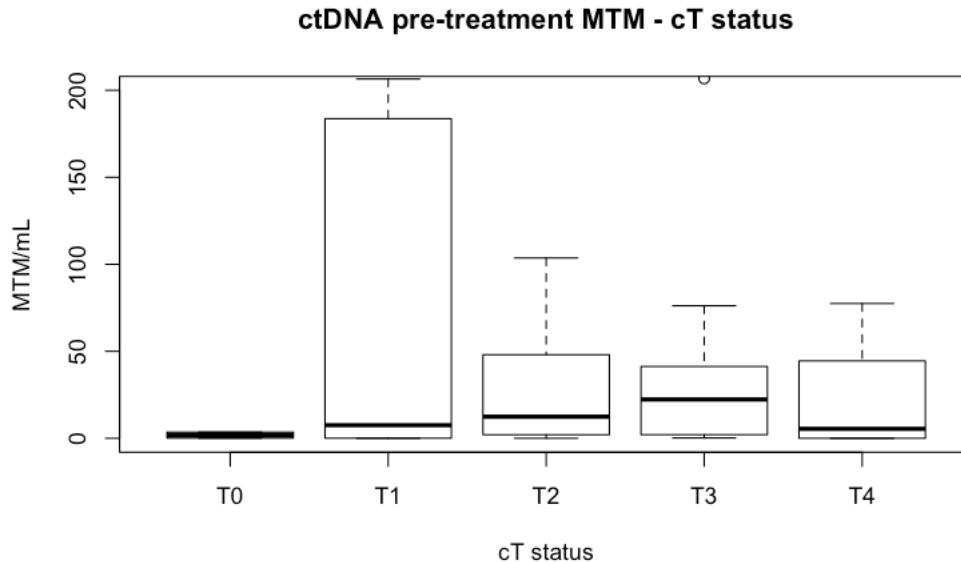
```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_datadf <- as.data.frame(circ_data)

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))
```



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")

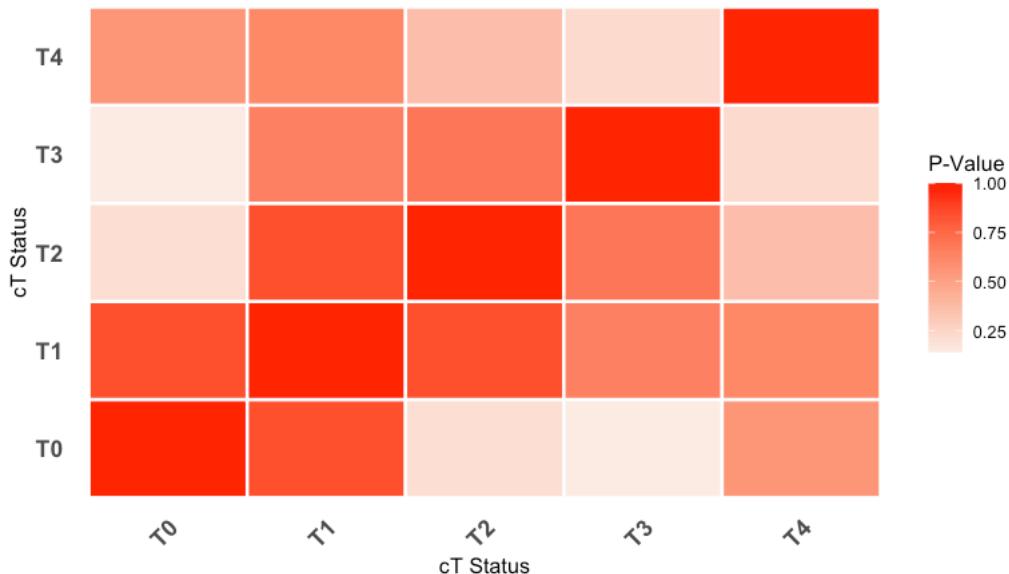
```

```

G2; H2;Warning h: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
: 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 generate
d. g

```

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



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_dataadf <- as.data.frame(circ_data)

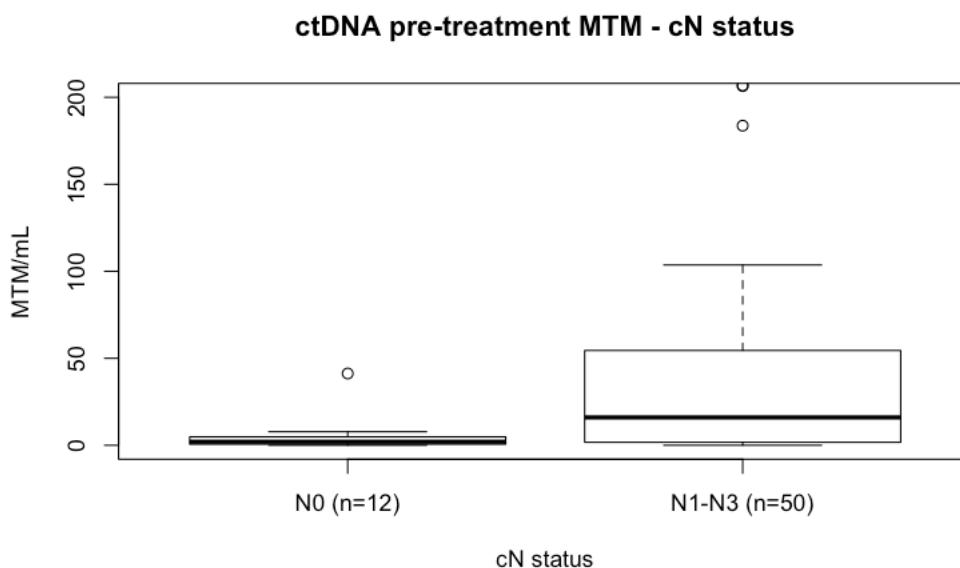
tally(~cN.status, data=circ_data, margins = TRUE)
```

cN.status

	N0	N1-N3	Total
	12	50	62

Hide

```
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))
```



Hide

```
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)

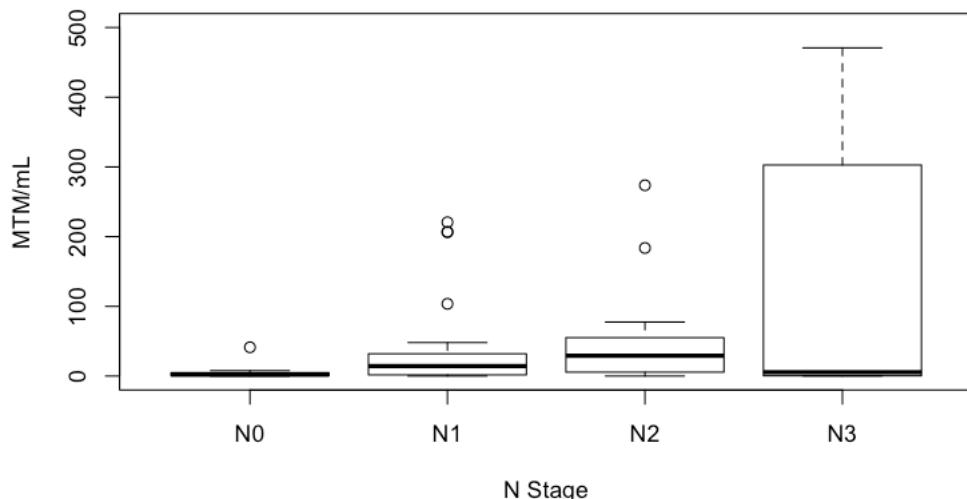
tally(~cN, data=circ_data, margins = TRUE)
```

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

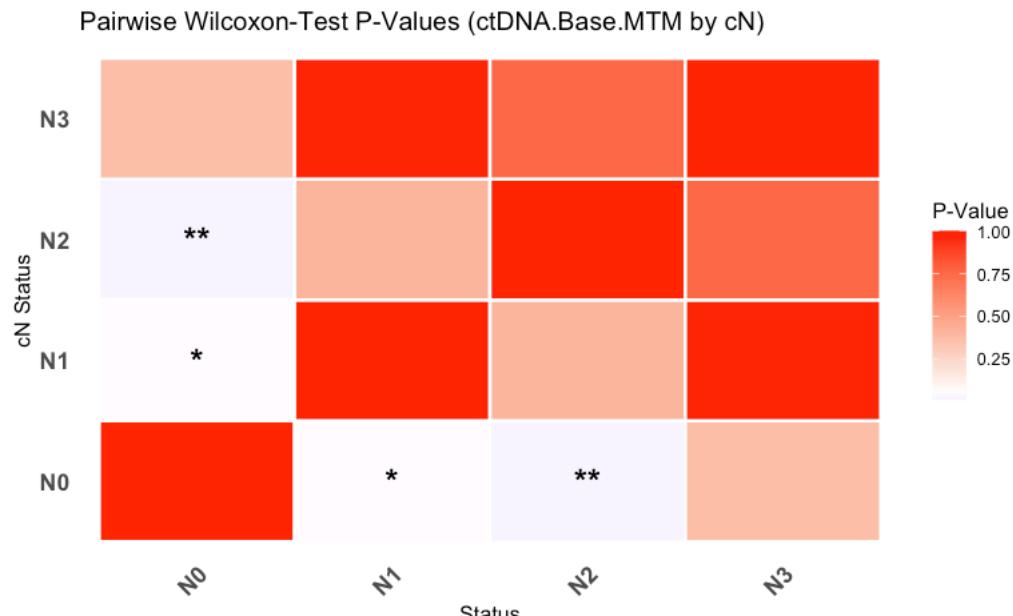
for (i in 1:length(cN_levels)) {
  for (j in i:length(cN_levels)) {
    if (i != j) {
      # Extract data for both groups
      data1 <- circ_data %>% filter(cN == cN_levels[i]) %>% pull(ctDNA.Base.MTM)
      data2 <- circ_data %>% filter(cN == cN_levels[j]) %>% pull(ctDNA.Base.MTM)

      # 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()
```

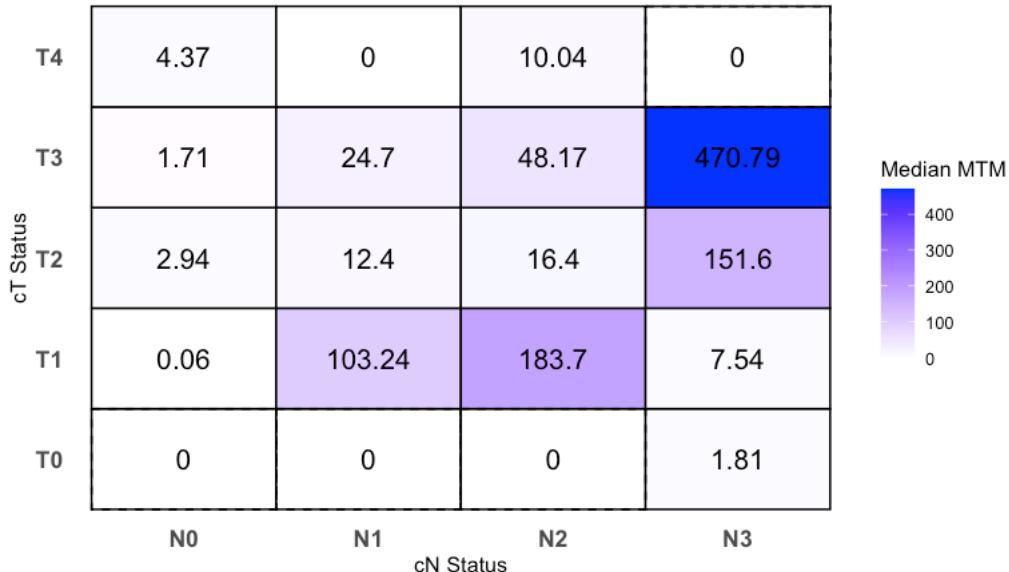
`summarise()` has grouped output by 'cT'. You can override using the ``.groups` argument.

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



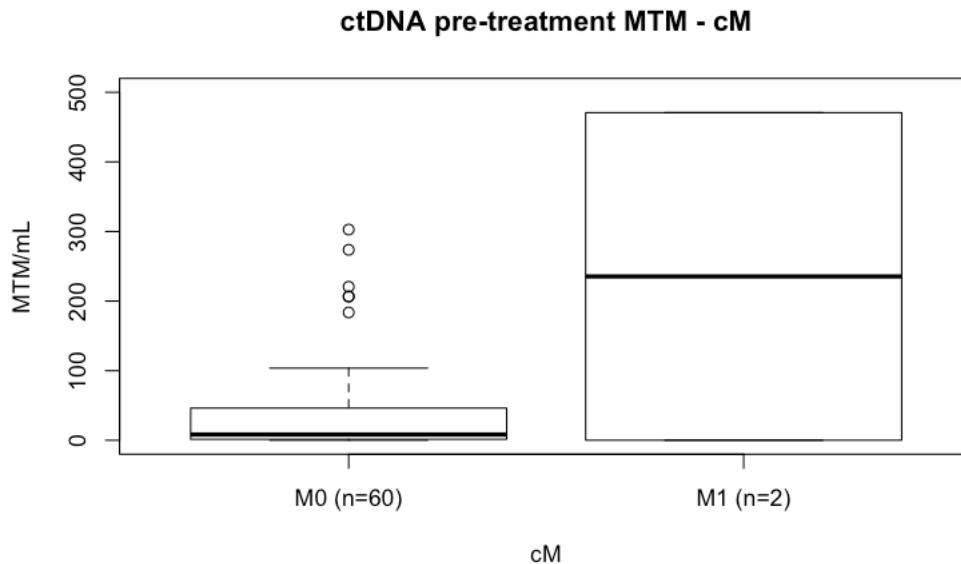
Hide

```
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
```

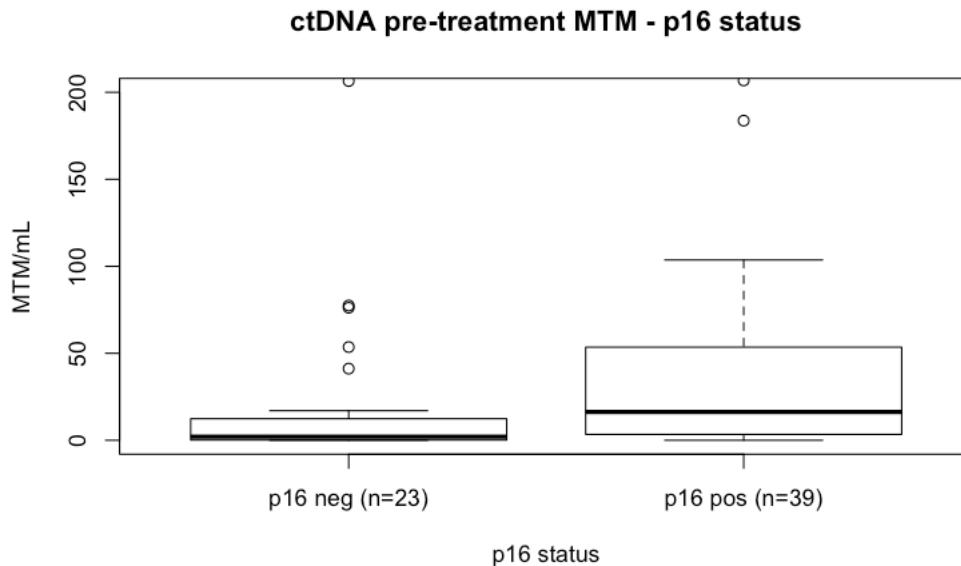
```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA HNSCC Peddada Clinical Data.csv")
circ_data <- circ_data[circ_data$ctDNA.available=="TRUE",]
circ_data <- circ_data[circ_data$ctDNA.Base!=""]
circ_dataadf <- as.data.frame(circ_data)

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))
```



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)

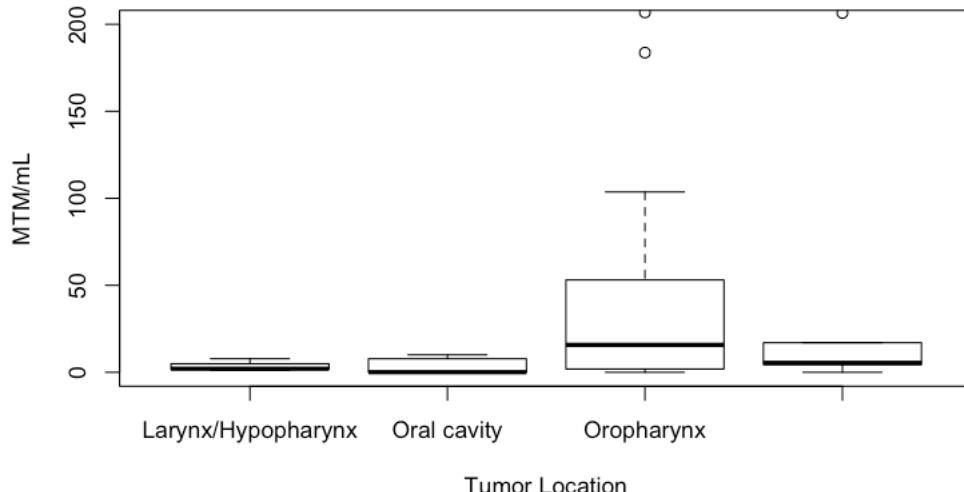
tally(~Prim.Location, data=circ_data, margins = TRUE)
```

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	

```
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
```

```

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)

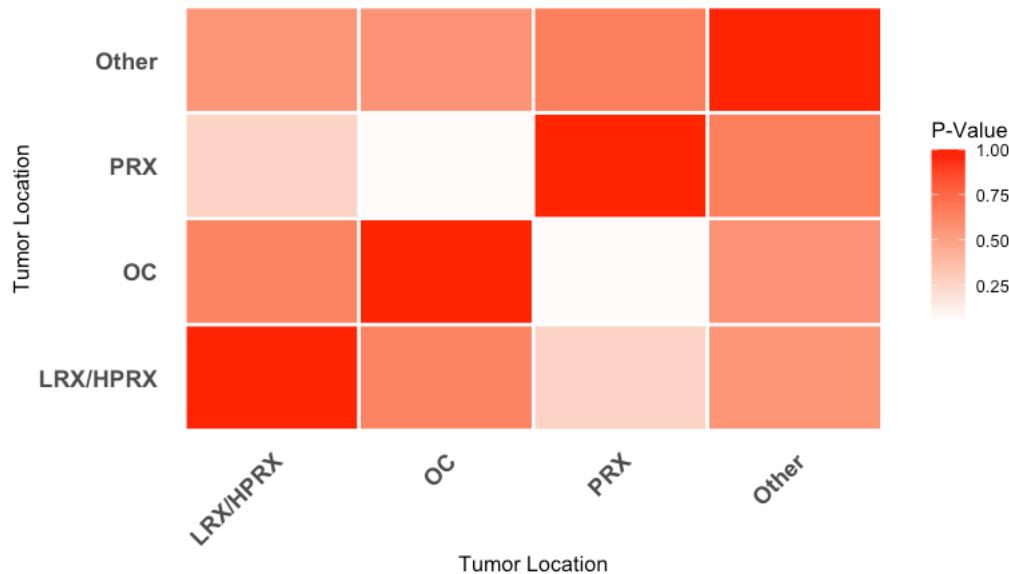
      # Perform Wilcoxon test and store p-value
      test_result <- wilcox.test(data1, data2, exact = FALSE)
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Make symmetric
    } else {
      p_value_matrix[i, j] <- 1 # Self-comparison = 1
    }
  }
}

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	8	NA	NA	NA
ctDNA.MRD=POSITIVE	13	8	15.5	4.21	NA

Hide

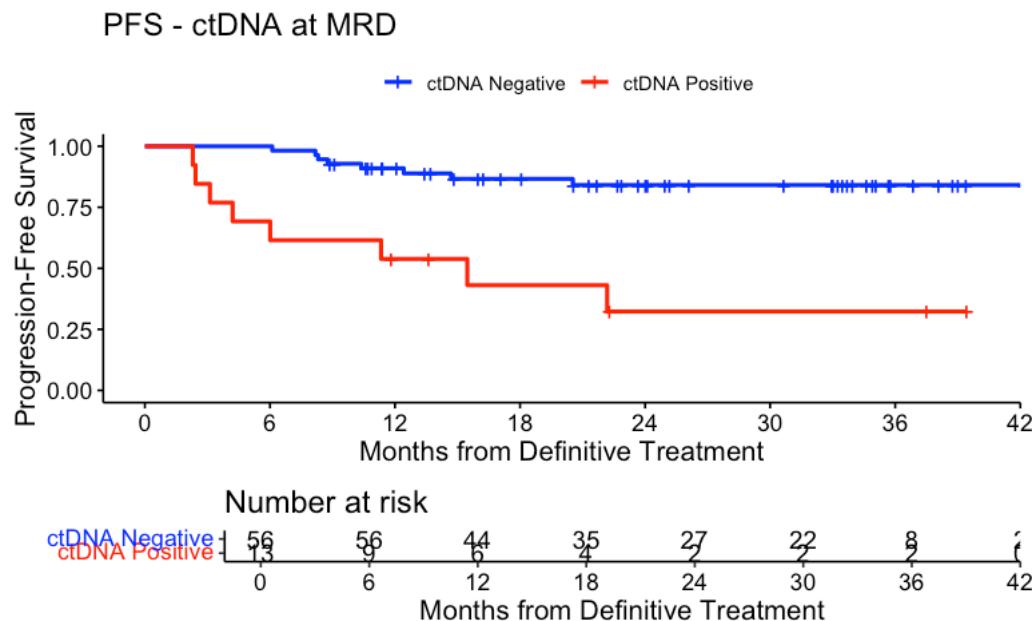
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	56	8	0.1428571	14.28571
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	44	5	0.910	0.0384	0.797		0.962	
24	27	3	0.841	0.0523	0.705		0.918	
36	8	0	0.841	0.0523	0.705		0.918	

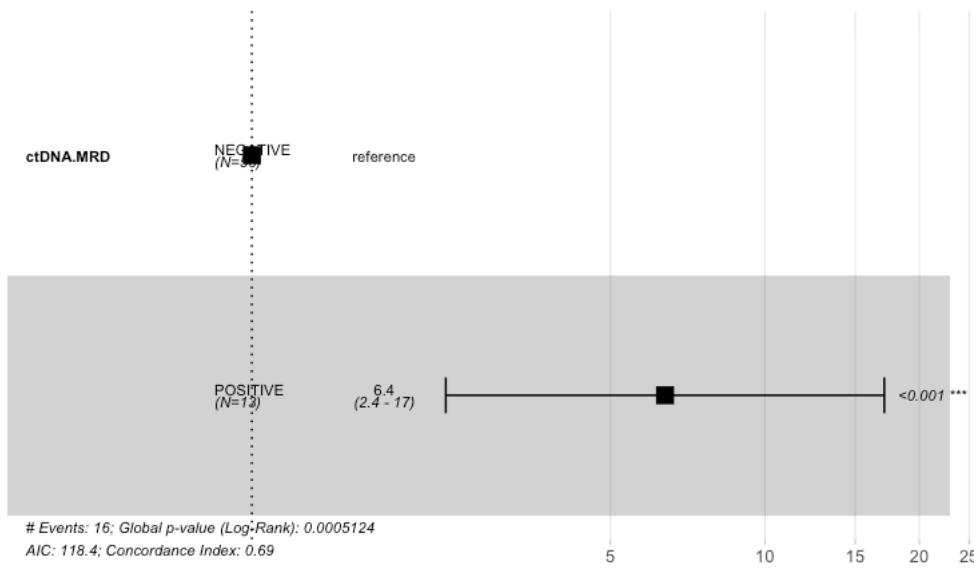
ctDNA.MRD=POSITIVE

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	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= 16

      coef  exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.8545    6.3886  0.5021 3.693 0.000221 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    6.389     0.1565    2.388    17.09

Concordance= 0.69  (se = 0.059 )
Likelihood ratio test= 12.07 on 1 df,  p=5e-04
Wald test            = 13.64 on 1 df,  p=2e-04
Score (logrank) test = 17.94 on 1 df,  p=2e-05
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 6.39 (2.39-17.09); 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 = 10.706, df = 1, p-value = 0.001068
```

[Hide](#)

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

```
Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.001047
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 2.06321 46.02731
sample estimates:
odds ratio
 9.153574
```

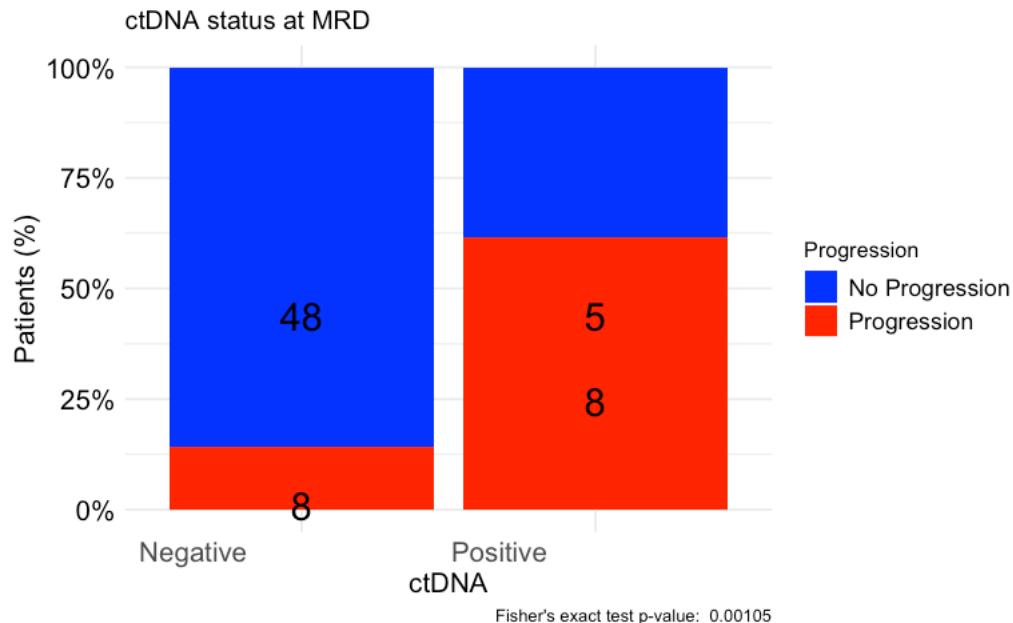
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	48	8
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

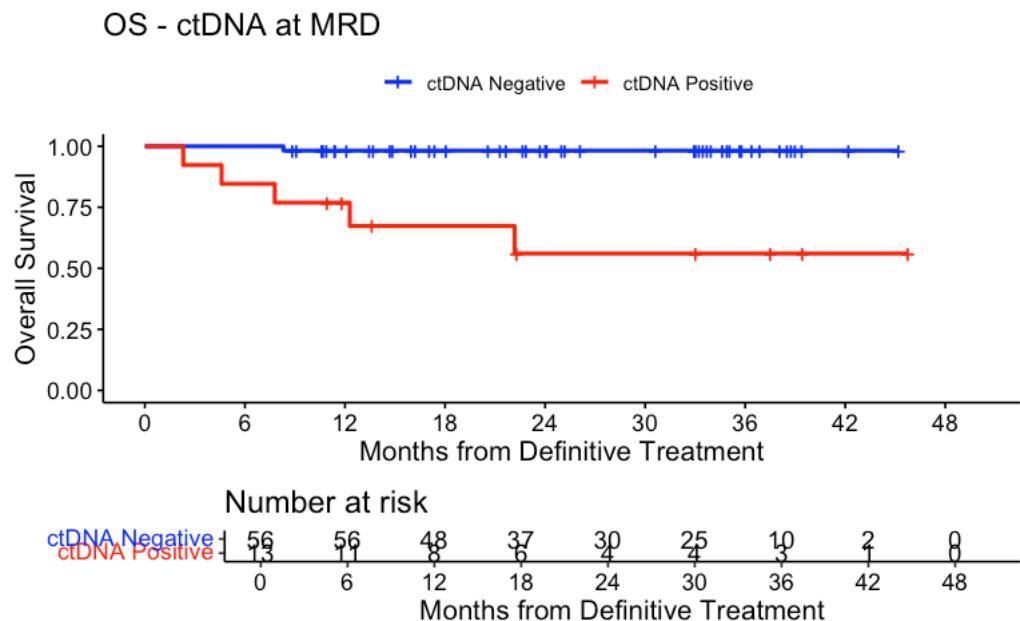
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

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="")
```



Hide

```
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")

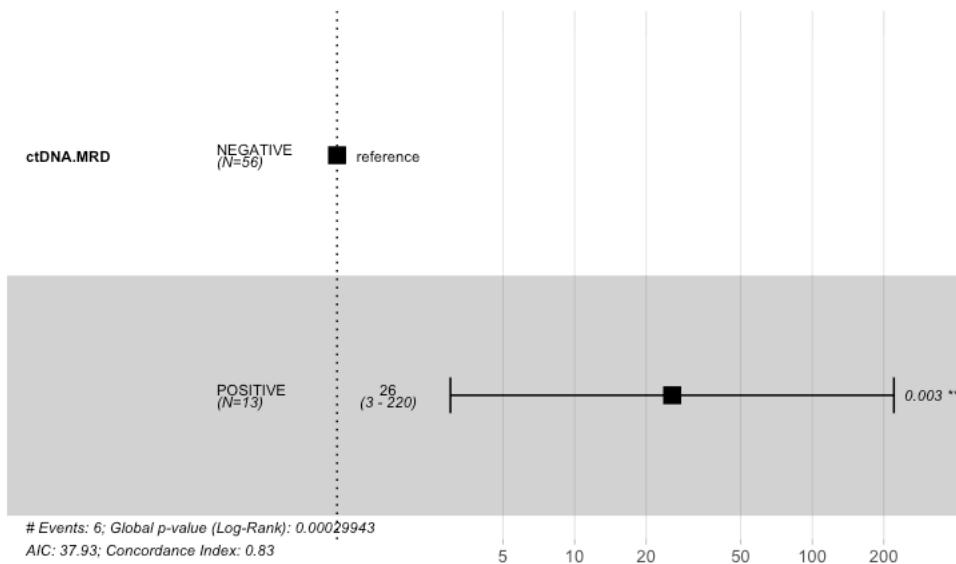
  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12     48      1    0.982  0.0177      0.88    0.997
    24     30      0    0.982  0.0177      0.88    0.997
    36     10      0    0.982  0.0177      0.88    0.997

  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
```

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= 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
```

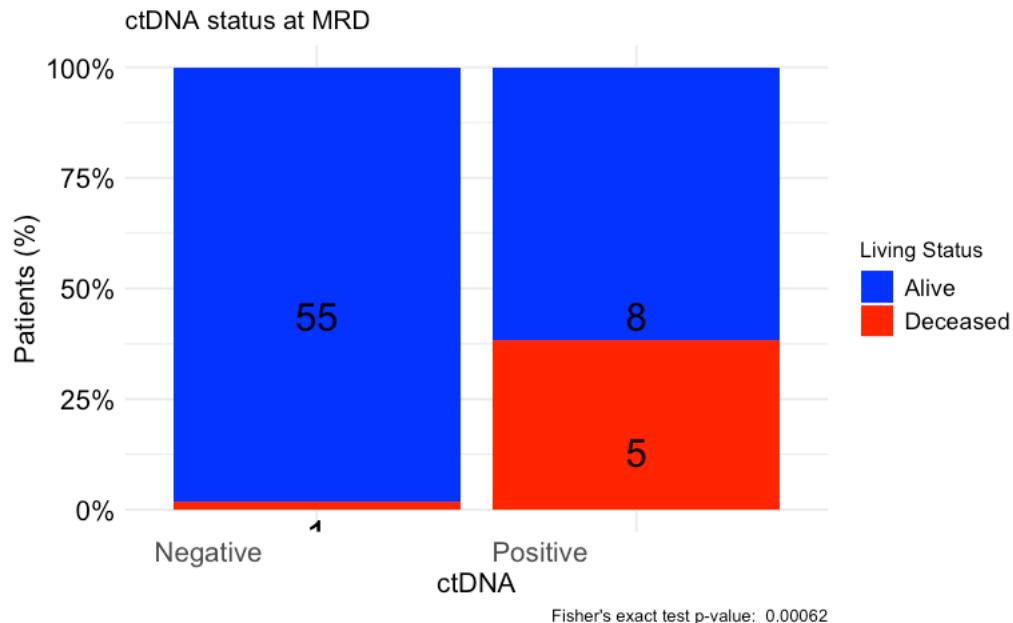
[Hide](#)

```
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 adjuvant treatment post-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!=""]
excluded_ids <- c("UNM-007", "UNM-008", "UNM-023", "UNM-027", "UNM-029",
                  "UNM-030", "UNM-035", "UNM-045", "UNM-051", "UNM-059",
                  "UNM-075", "UNM-082", "UNM-032", "UNM-042", "UNM-043",
                  "UNM-048", "UNM-050", "UNM-070")
circ_data <- circ_data[!circ_data$PatientID %in% excluded_ids, ]

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 44      6      NA      NA      NA
ctDNA.MRD=POSITIVE  7      6    11.3    3.12      NA
```

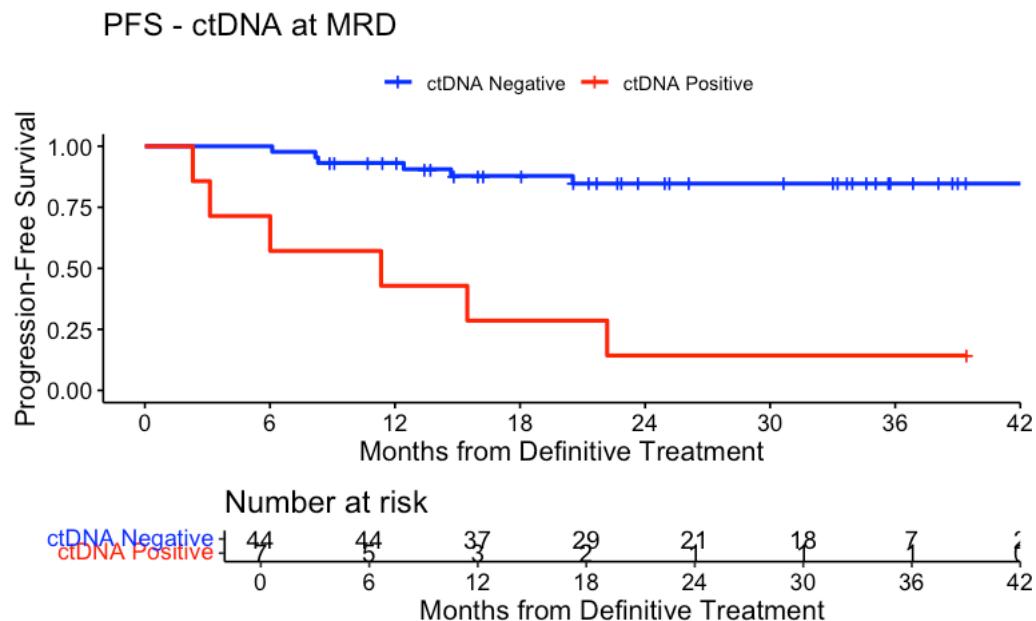
[Hide](#)

```
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	44	6	0.1363636	13.63636
POSITIVE	7	6	0.8571429	85.71429
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	44	0	1.000	0.0000	1.000		1.000	
12	37	3	0.932	0.0380	0.803		0.977	
24	21	3	0.847	0.0582	0.688		0.929	
36	7	0	0.847	0.0582	0.688		0.929	

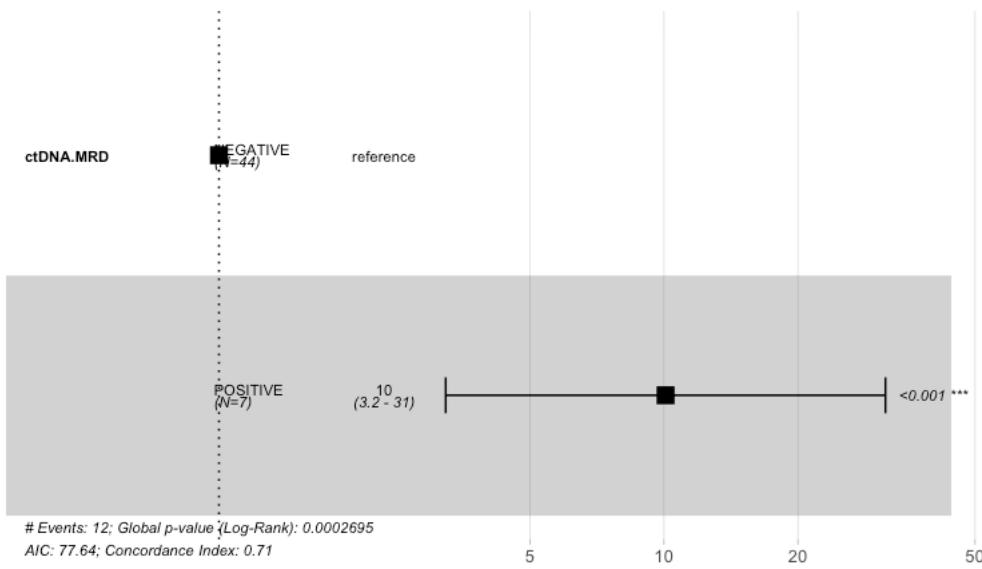
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	7	0	1.000	0.000	1.00000		1.000	
12	3	4	0.429	0.187	0.09775		0.734	
24	1	2	0.143	0.132	0.00712		0.465	
36	1	0	0.143	0.132	0.00712		0.465	

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= 51, number of events= 12

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 2.3109  10.0834  0.5805 3.981 6.87e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE  10.08     0.09917    3.232    31.46

Concordance= 0.711 (se = 0.067 )
Likelihood ratio test= 13.27 on 1 df,  p=3e-04
Wald test           = 15.85 on 1 df,  p=7e-05
Score (logrank) test = 24.07 on 1 df,  p=9e-07
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 10.08 (3.23-31.46); 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 = 13.662, df = 1, p-value = 0.0002189
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 0.0003181
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
3.315922 1768.464983
sample estimates:
odds ratio
33.80814
```

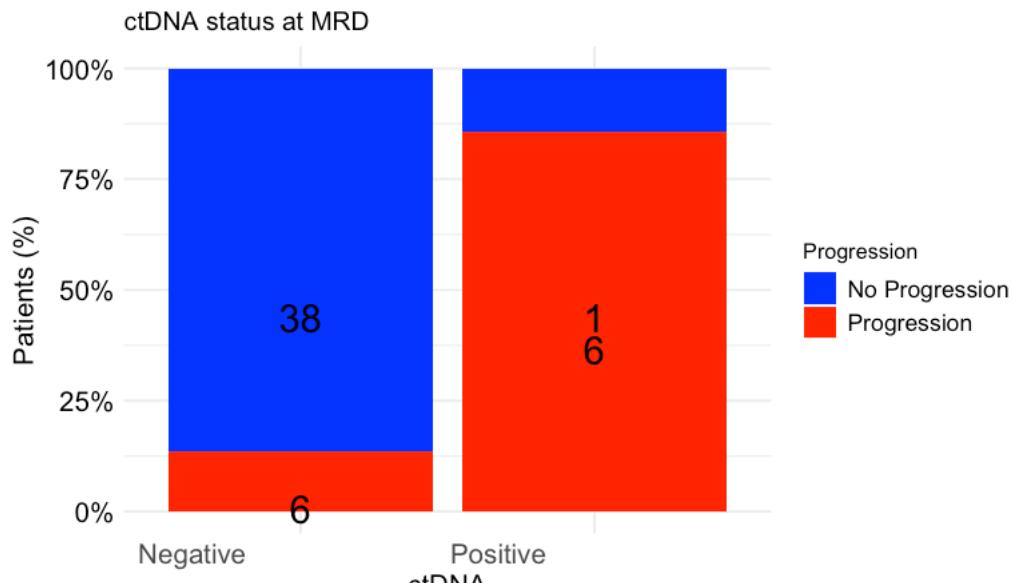
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	38	6
Positive	1	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",
       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 - exclude pts with adjuvant treatment post-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!="",]
excluded_ids <- c("UNM-007", "UNM-008", "UNM-023", "UNM-027", "UNM-029",
                  "UNM-030", "UNM-035", "UNM-045", "UNM-051", "UNM-059",
                  "UNM-075", "UNM-082", "UNM-032", "UNM-042", "UNM-043",
                  "UNM-048", "UNM-050", "UNM-070")
circ_data <- circ_data[!circ_data$PatientID %in% excluded_ids, ]

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	44	1	NA	NA	NA
ctDNA.MRD=POSITIVE	7	3	NA	12.3	NA

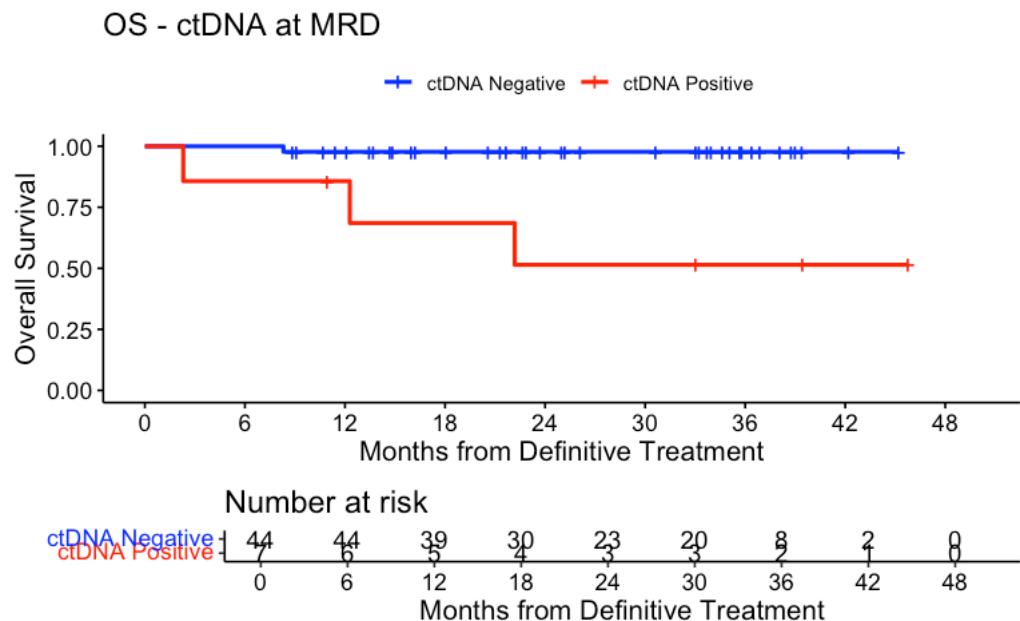
[Hide](#)

```
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	44	1	0.02272727	2.272727
POSITIVE	7	3	0.42857143	42.857143
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
e=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))
```

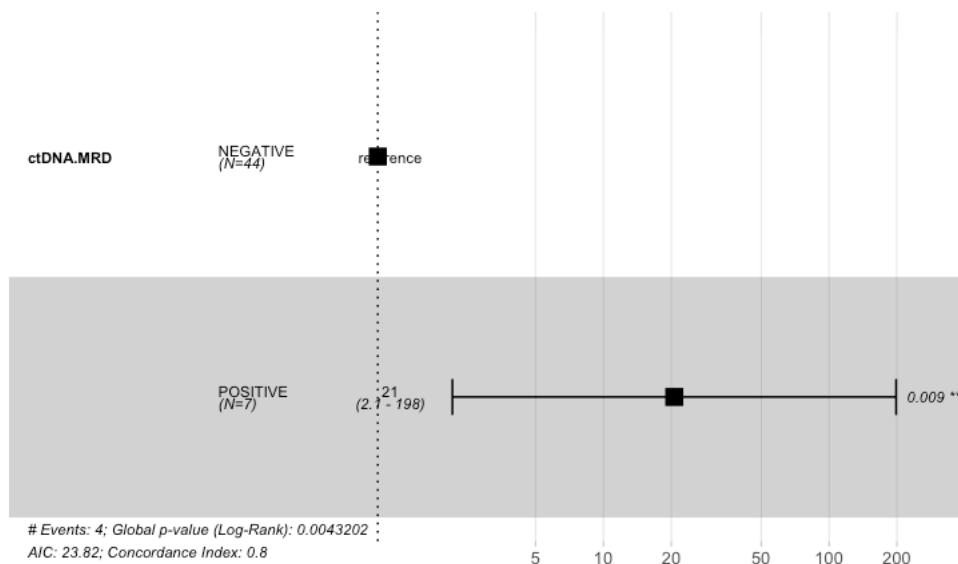
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.MRD=NEGATIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12      39      1    0.977  0.0225      0.849    0.997
    24      23      0    0.977  0.0225      0.849    0.997
    36       8      0    0.977  0.0225      0.849    0.997

  ctDNA.MRD=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
    12      5      1    0.857  0.132      0.334    0.979
    24      3      2    0.514  0.204      0.118    0.813
    36      2      0    0.514  0.204      0.118    0.813
```

```
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= 51, number of events= 4

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 3.026    20.613    1.155 2.619  0.00881 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    20.61     0.04851    2.142    198.4

Concordance= 0.799 (se = 0.119 )
Likelihood ratio test= 8.14 on 1 df,  p=0.004
Wald test            = 6.86 on 1 df,  p=0.009
Score (logrank) test = 13.95 on 1 df,  p=2e-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 = 20.61 (2.14-198.38); p = 0.009"
```

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
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 8.7198, df = 1, p-value = 0.003148
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.006303
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.805221 1704.546058
sample estimates:
odds ratio
27.80596
```

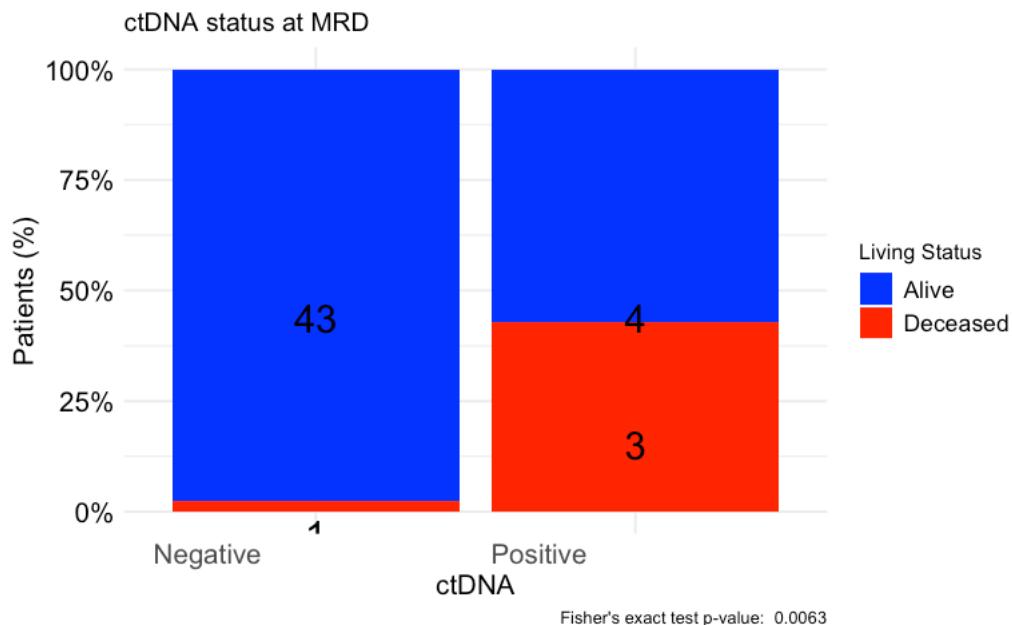
[Hide](#)

```
print(contingency_table)
```

	Alive	Deceased
Negative	43	1
Positive	4	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 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 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
```

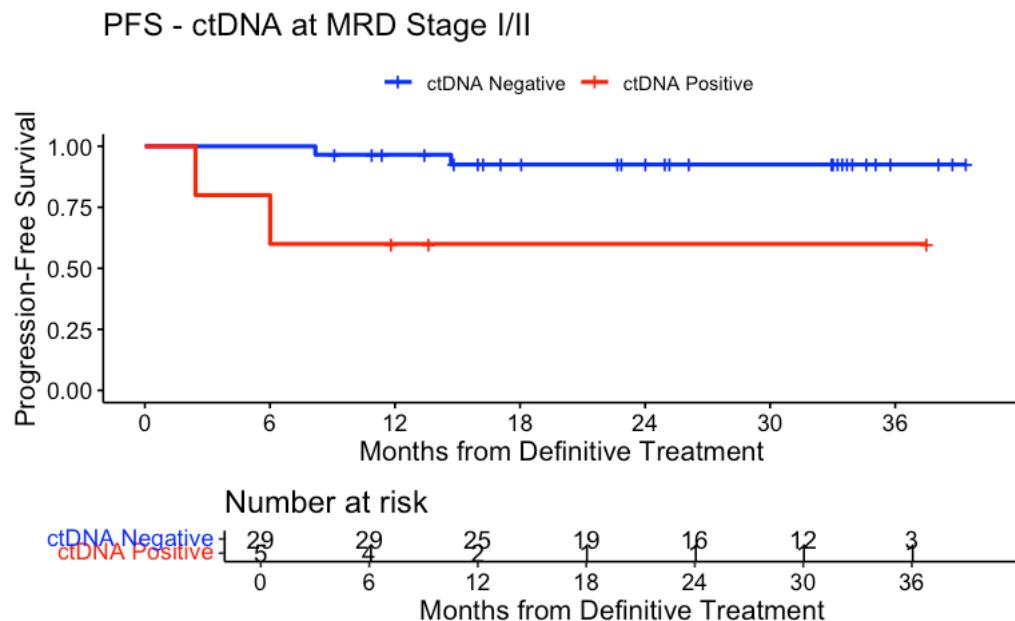
Hide

```
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
e=c("blue","red"), title="PFS - ctDNA at MRD Stage I/II", ylab= "Progression-Free Survival", xlab="Months from De
finitive 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	1.000	1.000
12	25	1	0.966	0.0339	0.779	0.995	0.732	0.981
24	16	1	0.925	0.0510	0.732	0.981	0.732	0.981
36	3	0	0.925	0.0510	0.732	0.981	0.732	0.981

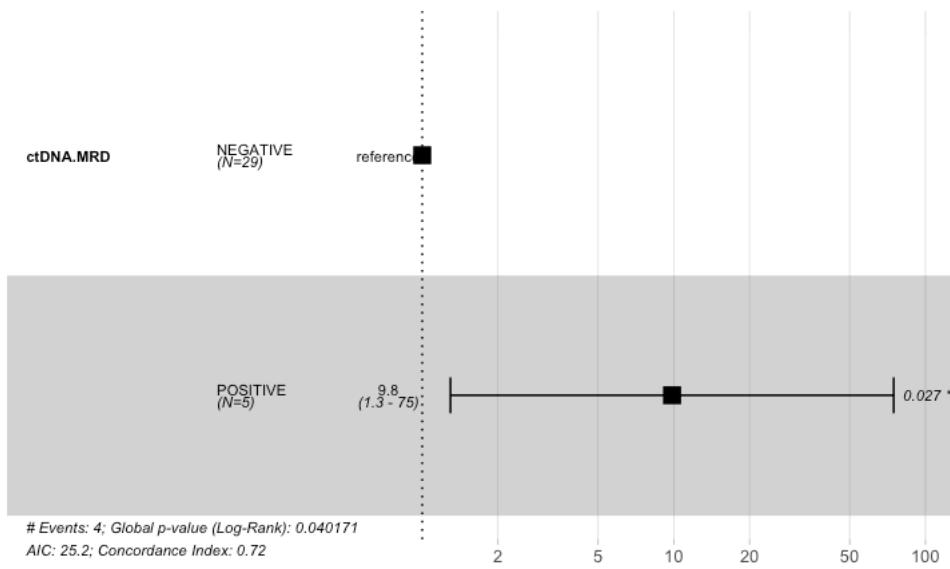
ctDNA.MRD=POSITIVE

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

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= 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 g
```

[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
```

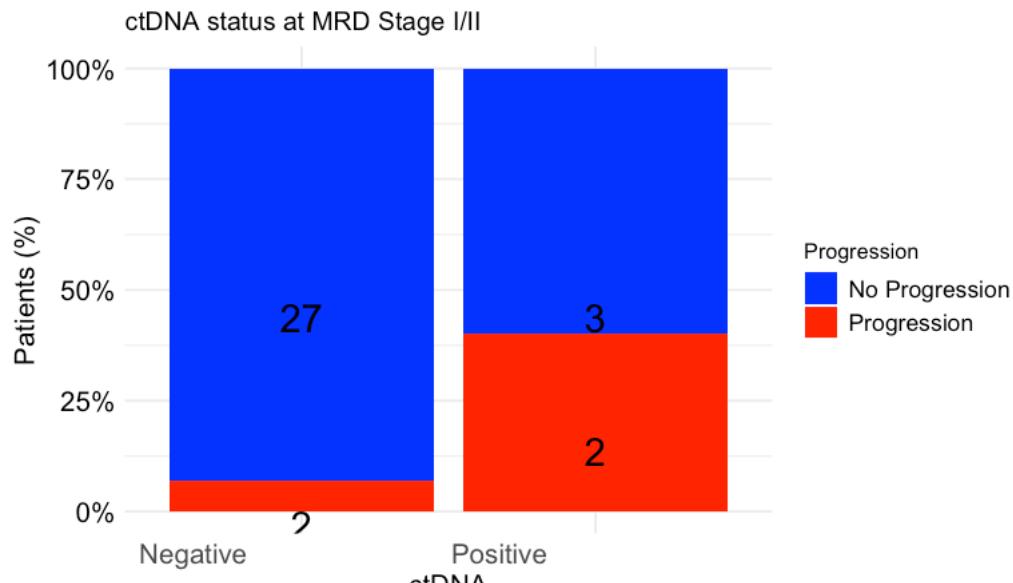
[Hide](#)

```
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      6     NA     NA     NA
ctDNA.MRD=POSITIVE  8      6    13.4    4.21     NA
```

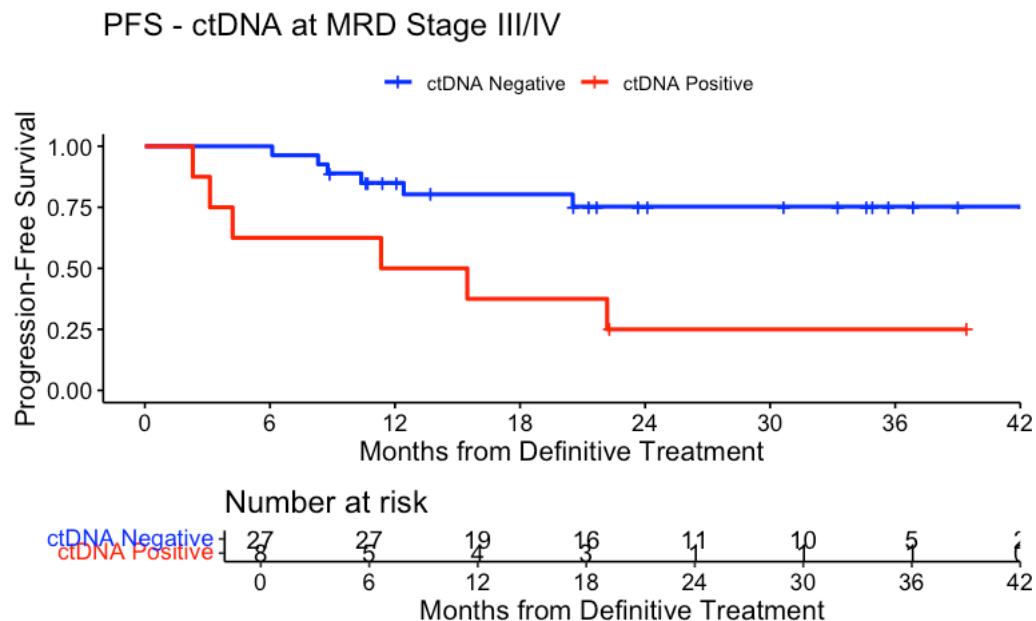
Hide

```
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	6	0.2222222	22.22222
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))
```

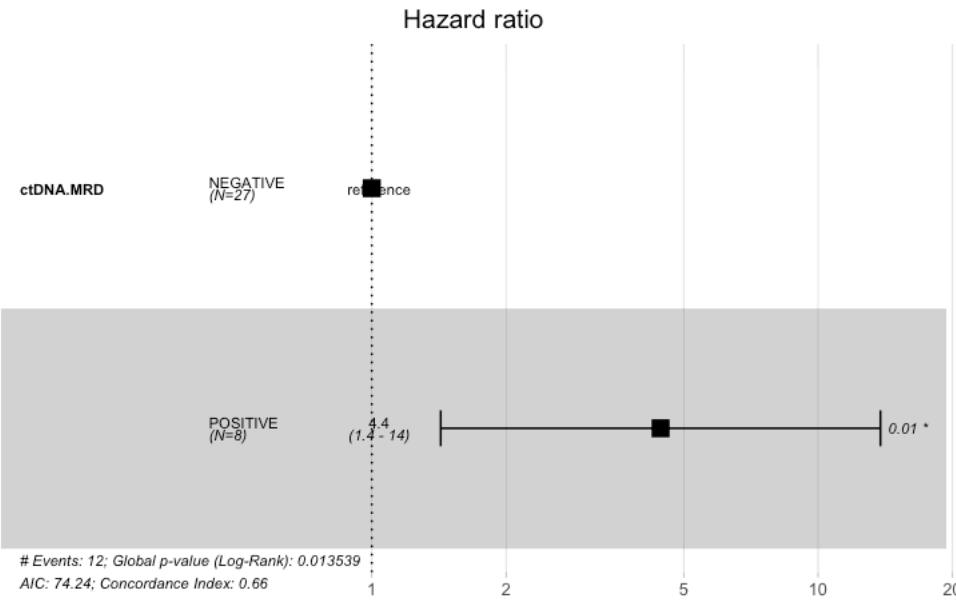
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
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
 12    19      4    0.850  0.0691    0.649    0.941
 24    11      2    0.753  0.0892    0.526    0.882
 36     5      0    0.753  0.0892    0.526    0.882
```

```
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
 12     4      4    0.50  0.177    0.1520    0.775
 24     1      2    0.25  0.153    0.0371    0.558
 36     1      0    0.25  0.153    0.0371    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)
```



Hide

```
summary(cox_fit)
```

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

n= 35, number of events= 12

      coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.MRDPOSITIVE 1.4899    4.4365   0.5788  2.574    0.01 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE    4.437     0.2254    1.427    13.79    

Concordance= 0.663  (se = 0.07 )
Likelihood ratio test= 6.1  on 1 df,  p=0.01
Wald test            = 6.63  on 1 df,  p=0.01
Score (logrank) test = 7.93  on 1 df,  p=0.005
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 4.44 (1.43-13.79); p = 0.01"
```

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 = 5.4671, df = 1, p-value = 0.01938
```

[Hide](#)

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

```
Fisher's Exact Test for Count Data

data: contingency_table
p-value = 0.01073
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
1.308188 121.976548
sample estimates:
odds ratio
9.642373
```

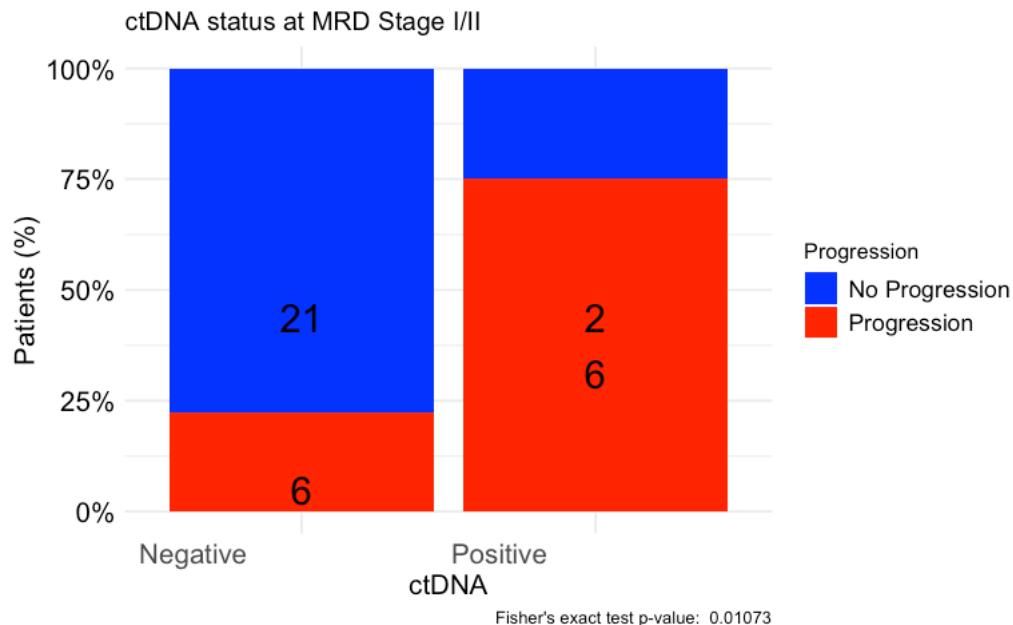
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	21	6
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

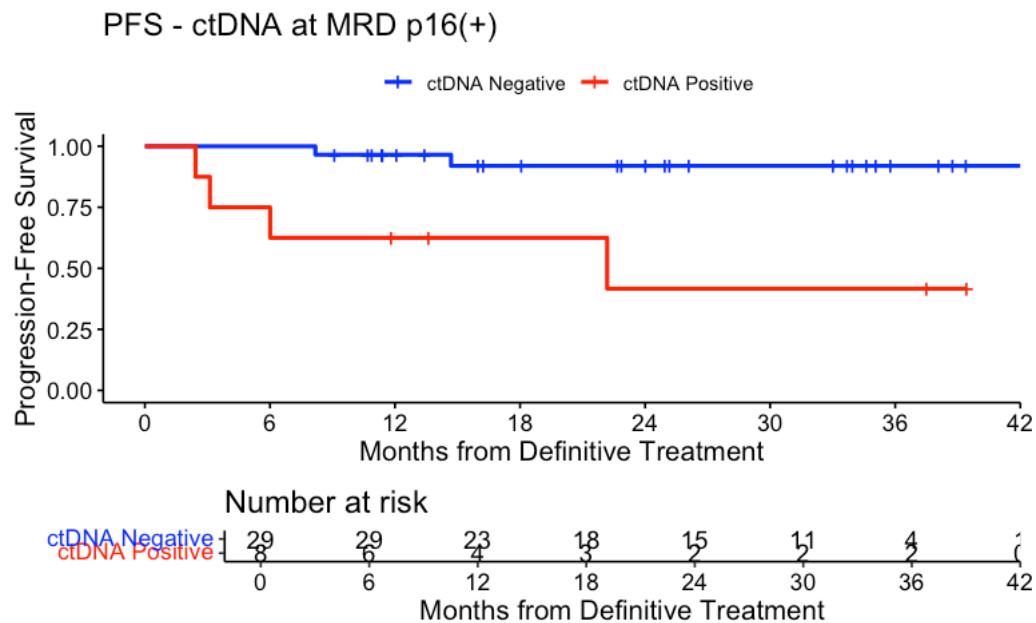
Hide

```
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	1.000	1.000
12	23	1	0.966	0.0339	0.779	0.995	0.711	0.980
24	15	1	0.920	0.0553	0.711	0.980	0.711	0.980
36	4	0	0.920	0.0553	0.711	0.980	0.711	0.980

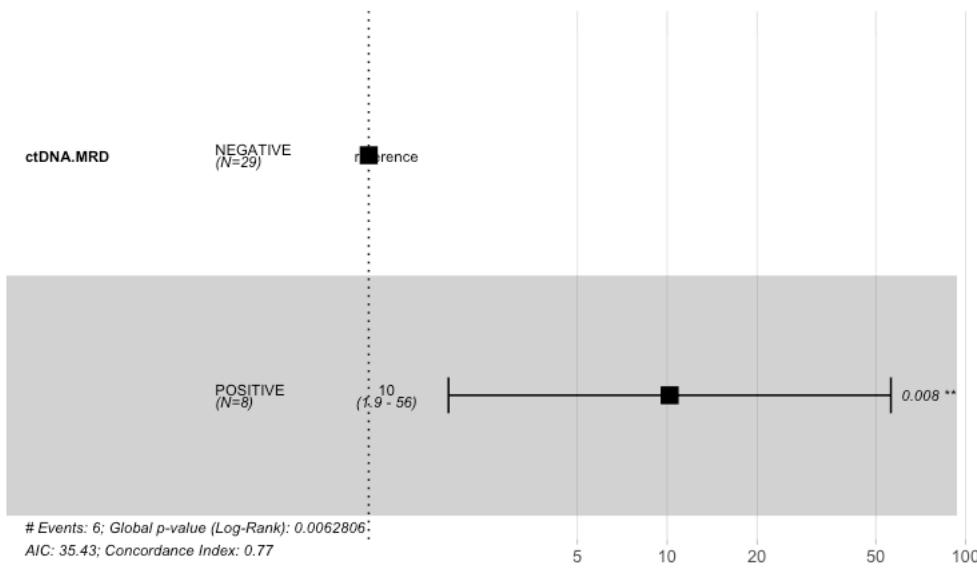
ctDNA.MRD=POSITIVE

time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
0	8	0	1.000	0.000	1.000	1.000	1.000	1.000
12	4	3	0.625	0.171	0.229	0.861	0.072	0.747
24	2	1	0.417	0.205	0.072	0.747	0.072	0.747
36	2	0	0.417	0.205	0.072	0.747	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
```

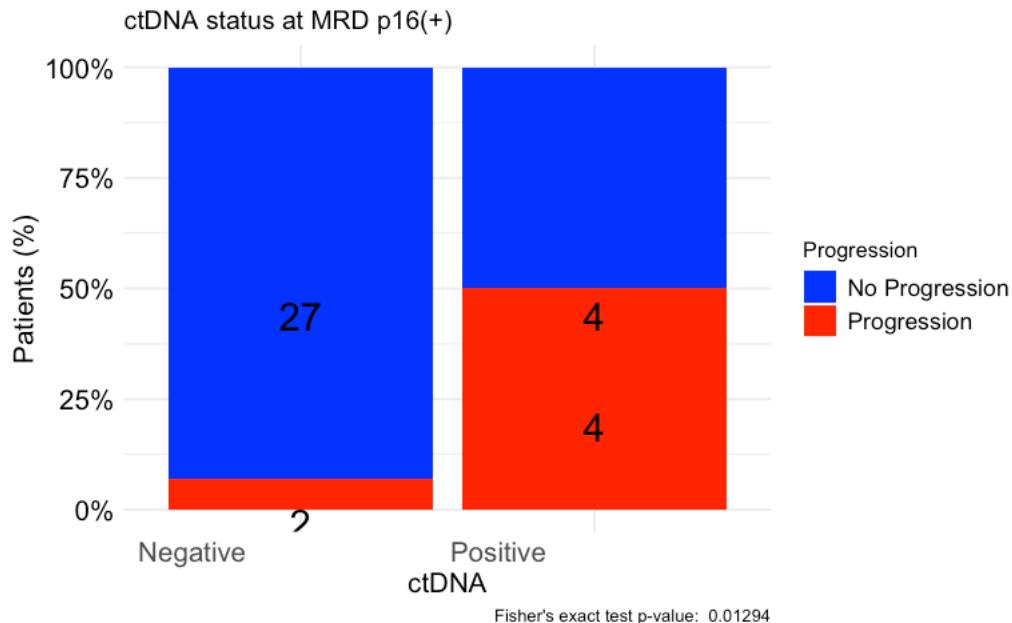
[Hide](#)

```
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	6	NA	NA	NA
ctDNA.MRD=POSITIVE	5	4	11.3	4.21	NA

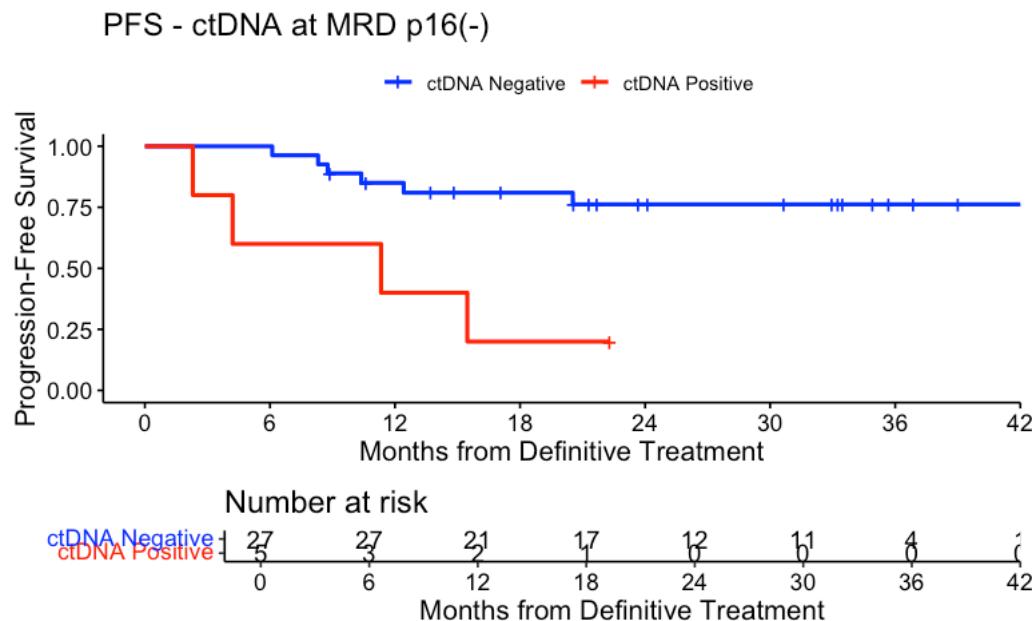
Hide

```
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	6	0.2222222	22.22222
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")
```

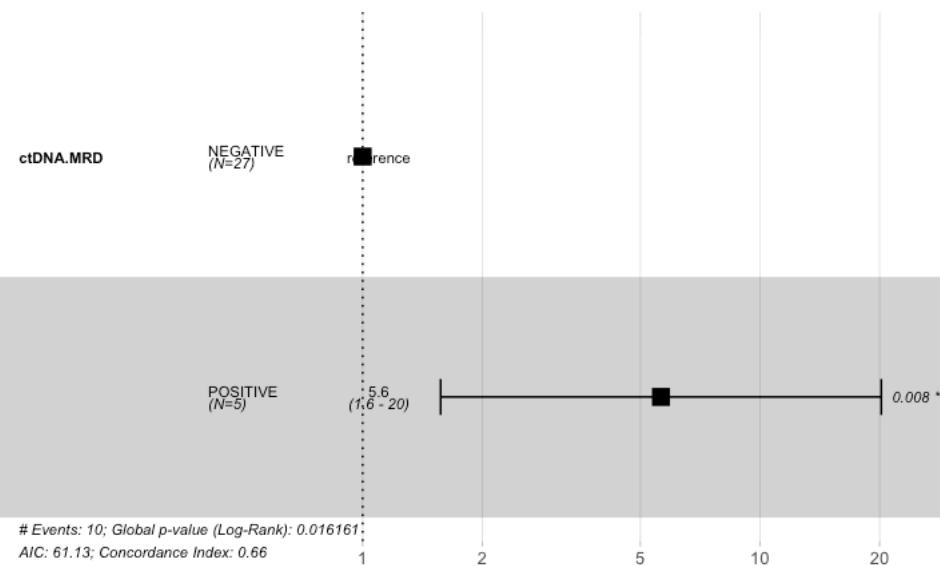
```
ctDNA.MRD=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     27      0     1.000  0.0000     1.000     1.000
 12     21      4     0.850  0.0691     0.649     0.941
 24     12      2     0.762  0.0858     0.542     0.886
 36      4      0     0.762  0.0858     0.542     0.886
```

```
ctDNA.MRD=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0      5      0     1.000  0.0000     1.000     1.000
 12      2      3     0.400  0.2190     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= 10

      coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.7286    5.6328  0.6515 2.653  0.00797 **

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

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    5.633     0.1775    1.571    20.2

Concordance= 0.655  (se = 0.072 )
Likelihood ratio test= 5.79  on 1 df,  p=0.02
Wald test            = 7.04  on 1 df,  p=0.008
Score (logrank) test = 8.93  on 1 df,  p=0.003
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 5.63 (1.57-20.2); 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
```

Hide

```
print(chi_square_test)
```

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

data: contingency_table
X-squared = 4.1417, df = 1, p-value = 0.04184
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 0.02419
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.013292 717.346329
sample estimates:
odds ratio
 12.61534
```

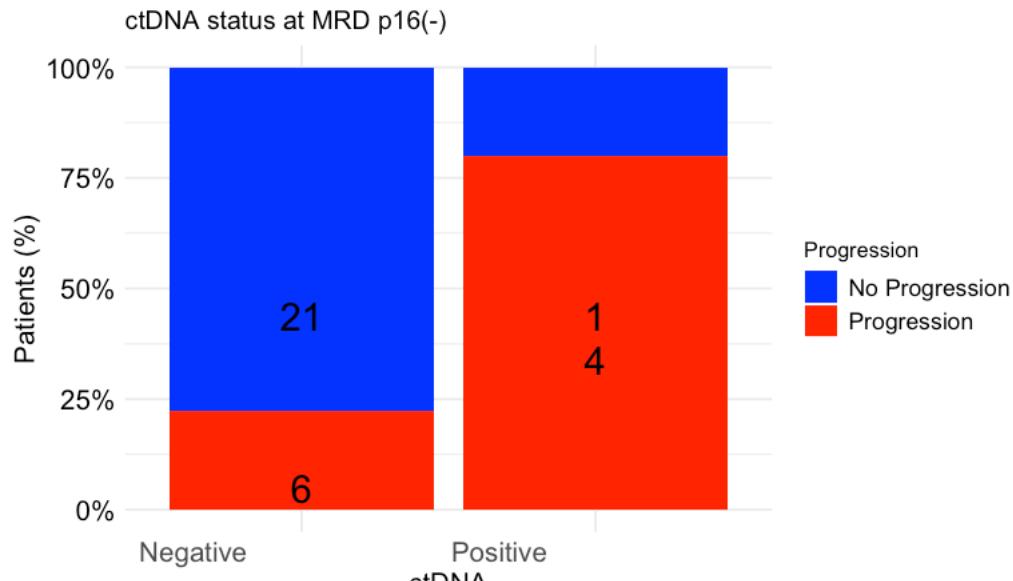
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	21	6
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	14	14.7	11.3	24.8

[Hide](#)

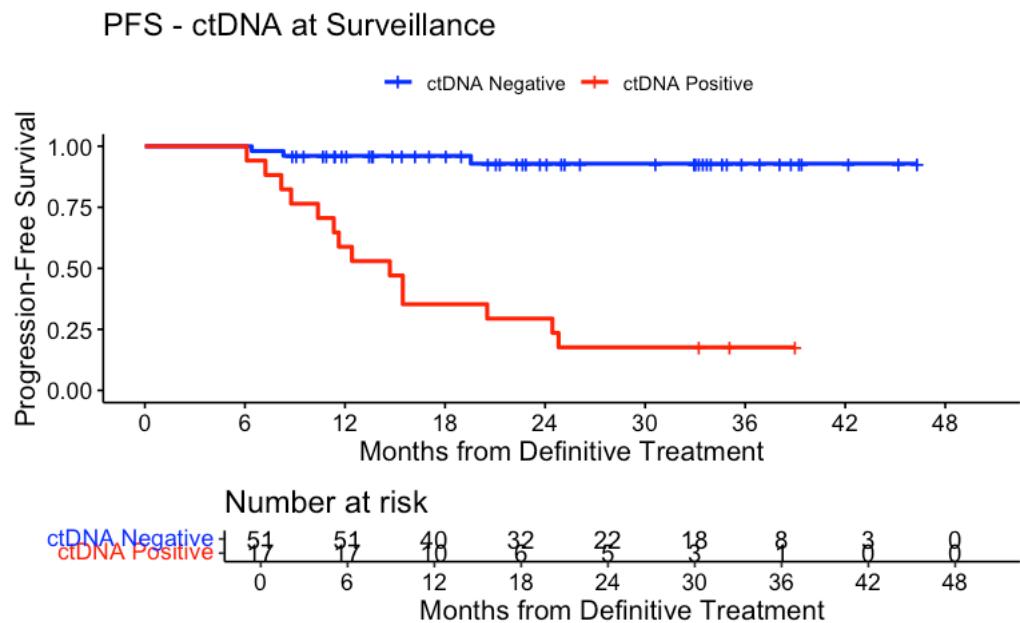
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	51	3	0.05882353	5.882353
POSITIVE	17	14	0.82352941	82.352941

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="")
```



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

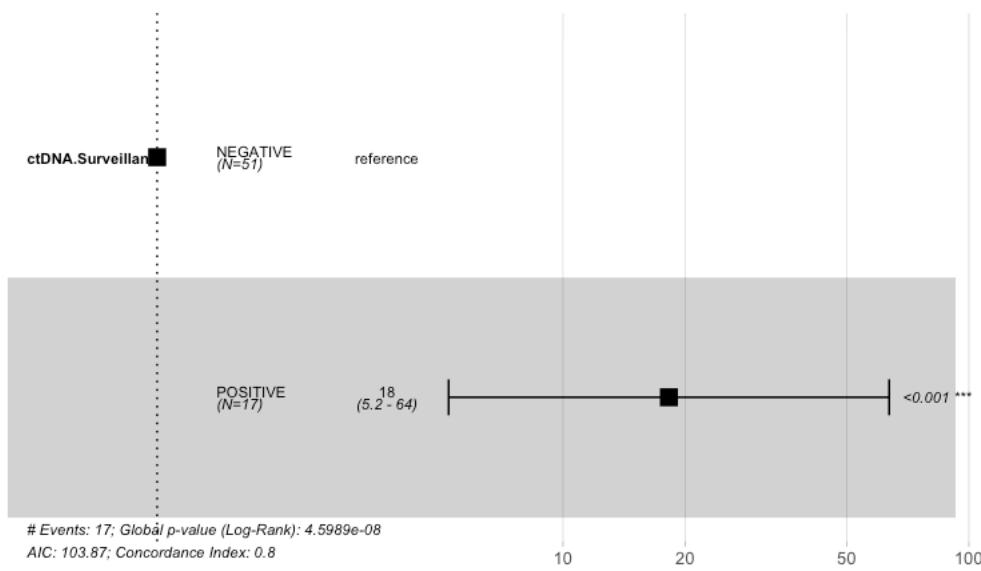
```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

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	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
0	17	0	1.000	0.0000	1.0000	1.000
12	10	7	0.588	0.1194	0.3254	0.778
24	5	5	0.294	0.1105	0.1071	0.511
36	1	2	0.176	0.0925	0.0435	0.383

```
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= 17

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE  2.9034    18.2359   0.6376 4.554 5.27e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    18.24     0.05484    5.227    63.62

Concordance= 0.802 (se = 0.052 )
Likelihood ratio test= 29.88 on 1 df,  p=5e-08
Wald test                 = 20.74 on 1 df,  p=5e-06
Score (logrank) test = 39.64 on 1 df,  p=3e-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 = 18.24 (5.23-63.62); 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 g
```

[Hide](#)

```
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 35.791, df = 1, p-value = 2.197e-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 = 3.189e-09
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
11.08872 588.54005
sample estimates:
odds ratio
64.48451
```

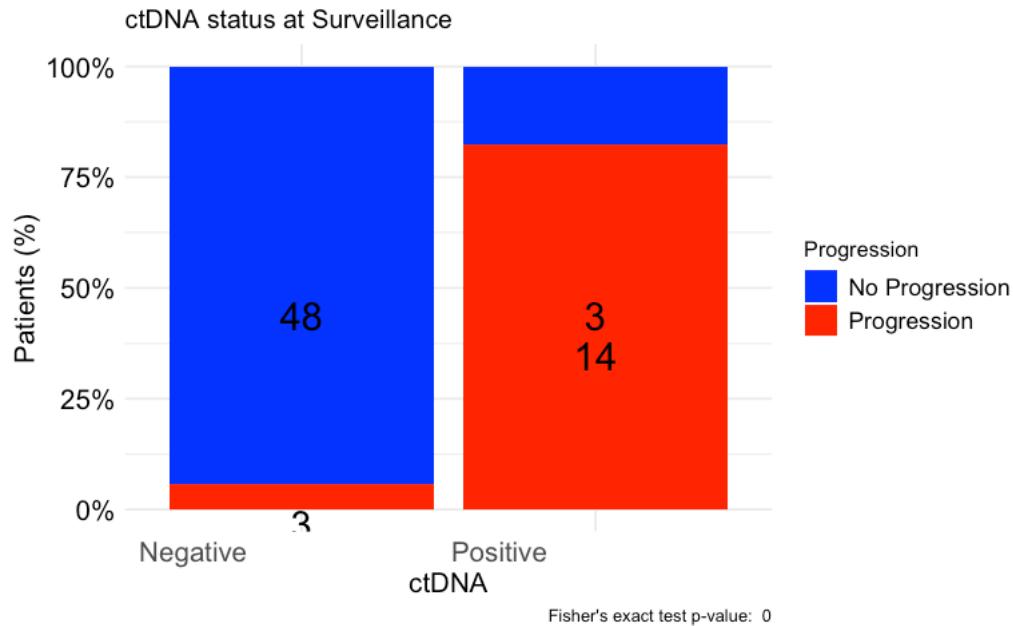
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	48	3
Positive	3	14

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7)
+
  theme_minimal() +
  labs(title = "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
```



#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 4.44 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, 257...
$ biomarker_status <chr> "NEGATIVE", "POSITIVE", "POSITIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", NA,
"NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NE...

```

```
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(days(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)))

```

```
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= 17
(66 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE 3.0934  22.0511  0.5178 5.974 2.32e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE 22.05    0.04535    7.992   60.84

Concordance= 0.718 (se = 0.06 )
Likelihood ratio test= 30.03 on 1 df,  p=4e-08
Wald test            = 35.68 on 1 df,  p=2e-09
Score (logrank) test = 69.77 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 = 22.05 (7.99-60.84); p = 0"
```

#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)
```

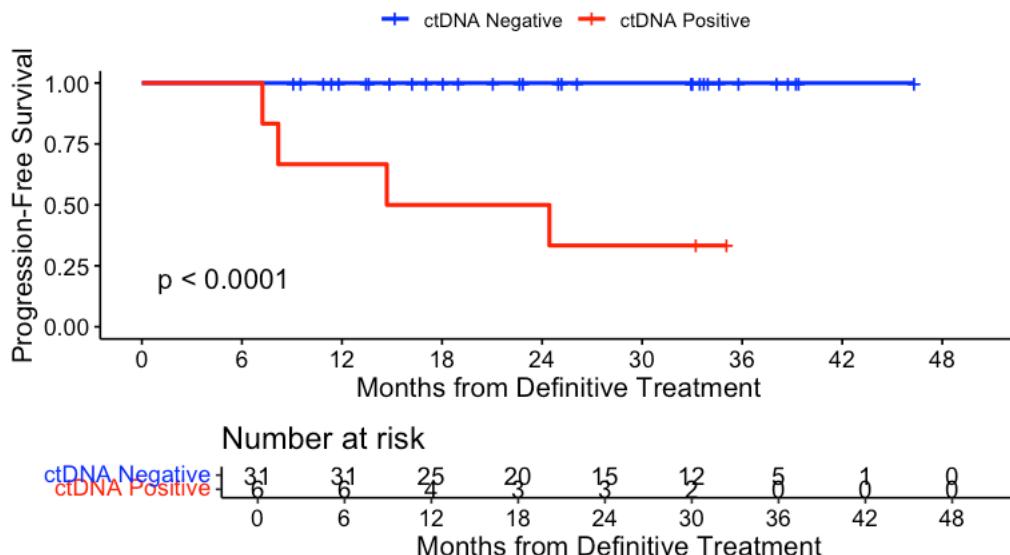
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	31	0	0.0000000	0.00000
POSITIVE	6	4	0.6666667	66.66667

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 <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
summary(cox_fit)
```

```
coxph(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
```

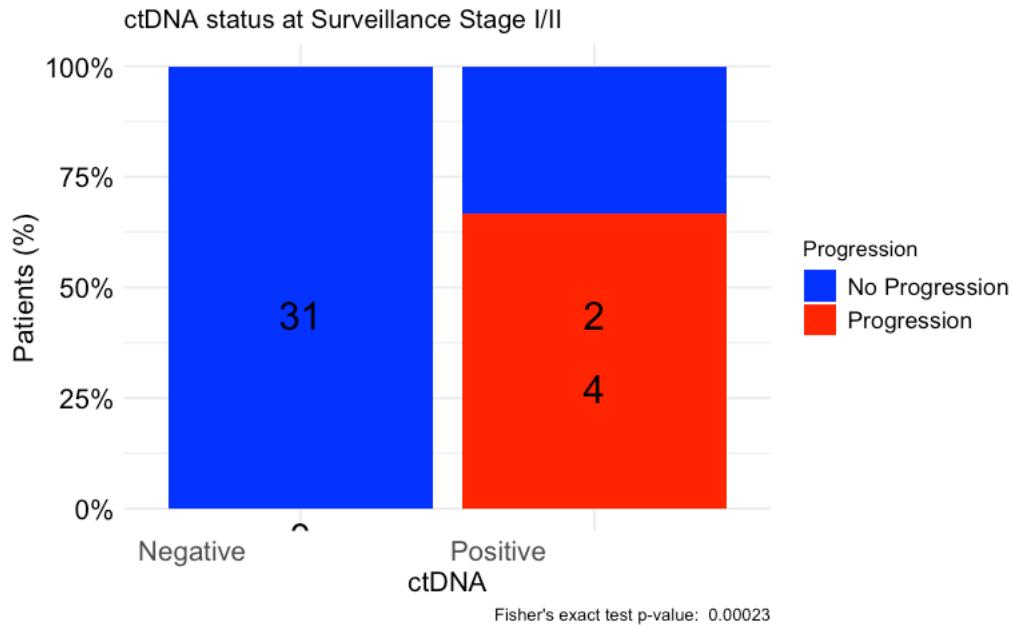
```
print(contingency_table)
```

	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	10	12.4	11.3	NA

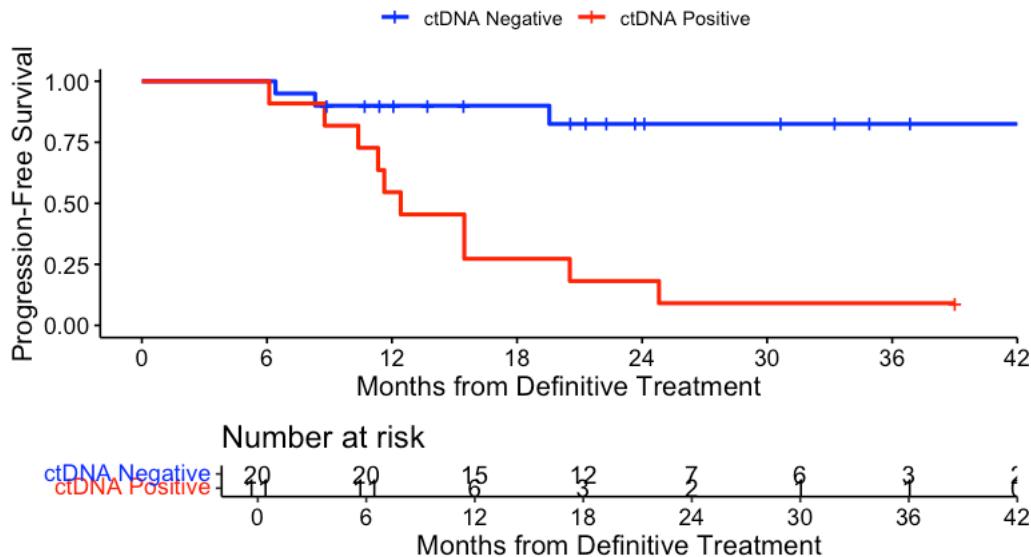
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(PFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	20	3	0.1500000	15.00000
POSITIVE	11	10	0.9090909	90.90909
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")
```

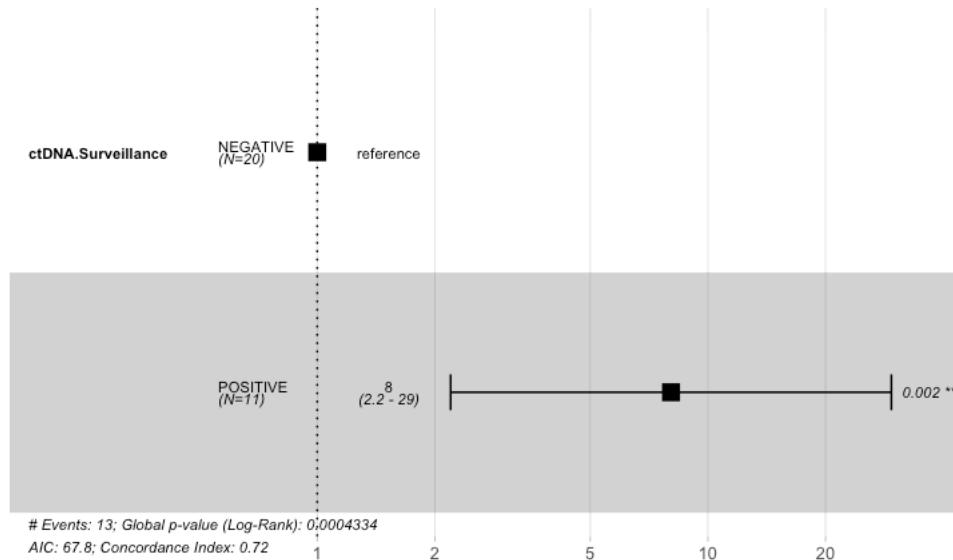
```
ctDNA.Surveillance=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     20      0  1.000  0.0000      1.000  1.000
 12    15      2  0.900  0.0671      0.656  0.974
 24     7      1  0.825  0.0945      0.539  0.942
 36     3      0  0.825  0.0945      0.539  0.942
```

```
ctDNA.Surveillance=POSITIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0     11      0  1.0000  0.0000      1.00000  1.000
 12     6      5  0.5455  0.1501      0.22854  0.780
 24     2      4  0.1818  0.1163      0.02854  0.442
 36     1      1  0.0909  0.0867      0.00537  0.333
```

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= 13

      coef exp(coef) se(coef)   z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 2.0850    8.0447   0.6627 3.146  0.00165 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    8.045     0.1243    2.195    29.48

Concordance= 0.722 (se = 0.07 )
Likelihood ratio test= 12.38 on 1 df,  p=4e-04
Wald test                 = 9.9 on 1 df,  p=0.002
Score (logrank) test = 13.83 on 1 df,  p=2e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

[1] "HR = 8.04 (2.2-29.48); p = 0.002"

[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 = 13.821, df = 1, p-value = 0.000201
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 6.172e-05
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 4.285476 2542.711782
sample estimates:
odds ratio
 45.74172
```

[Hide](#)

```
print(contingency_table)
```

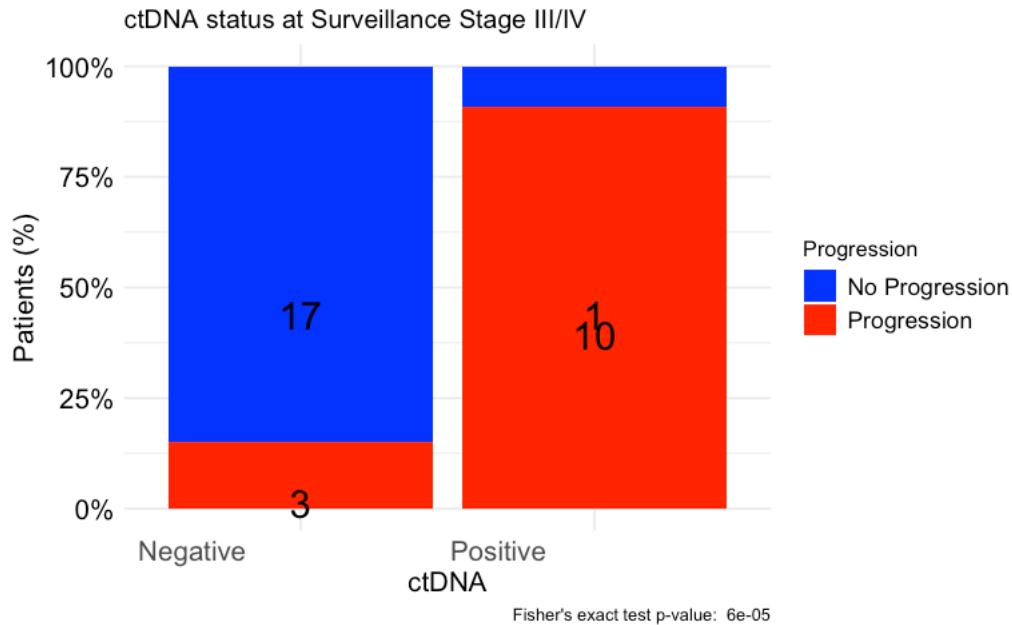
	No Progression	Progression
Negative	17	3
Positive	1	10

[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 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(+)

[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.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

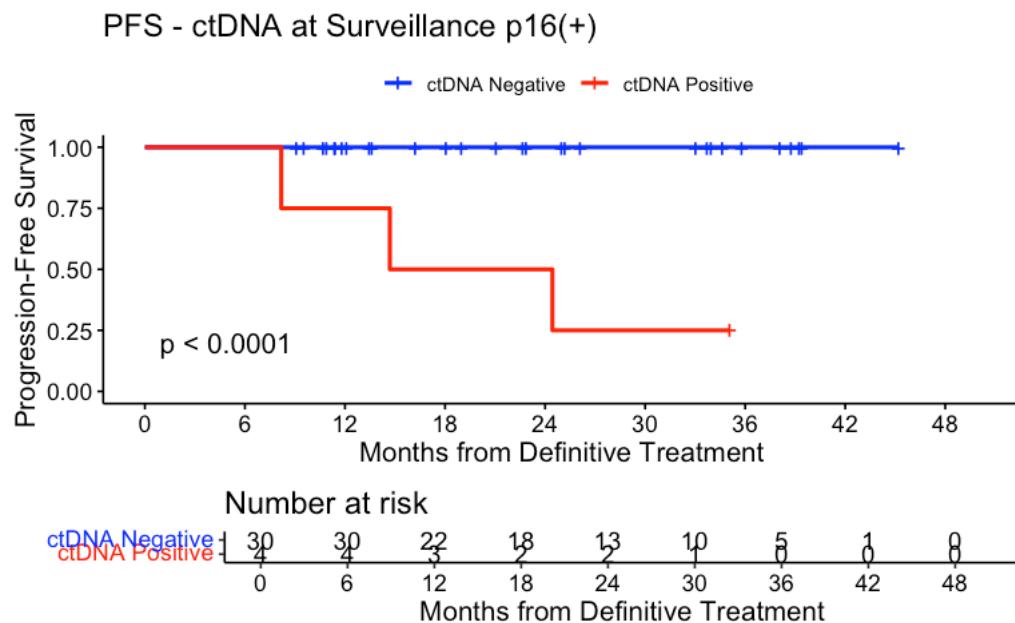
[Hide](#)

```
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	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="")
```



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
```

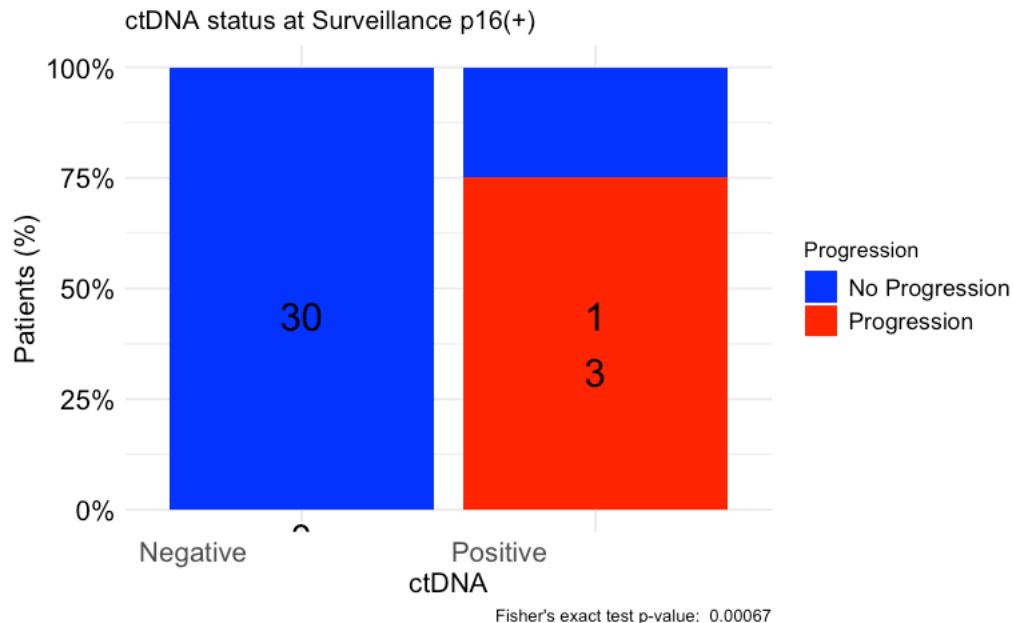
[Hide](#)

```
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	11	12.4	10.4	NA

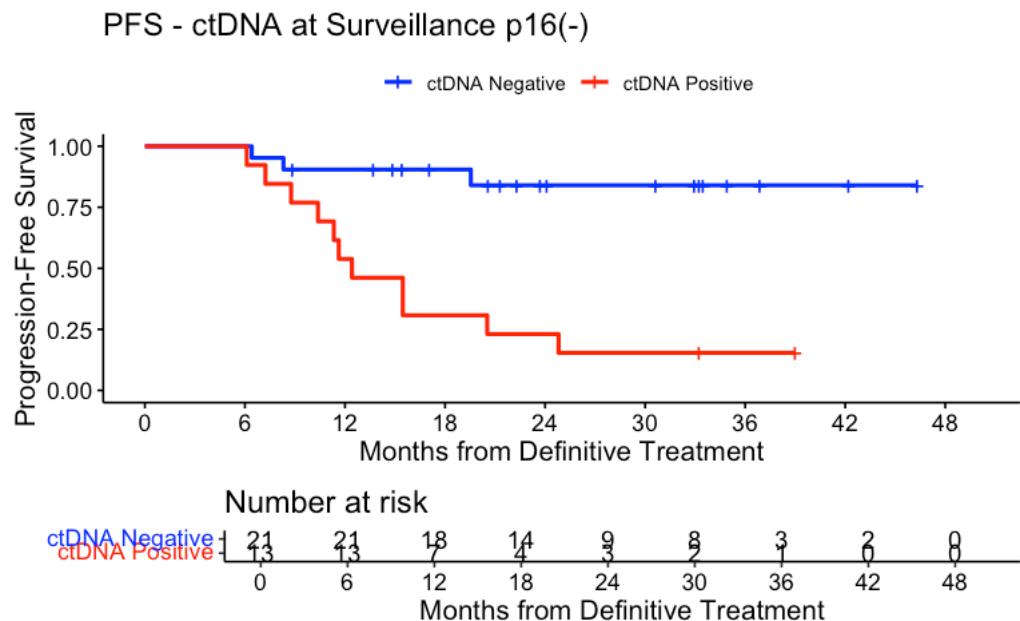
Hide

```
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	11	0.8461538	84.61538
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))
```

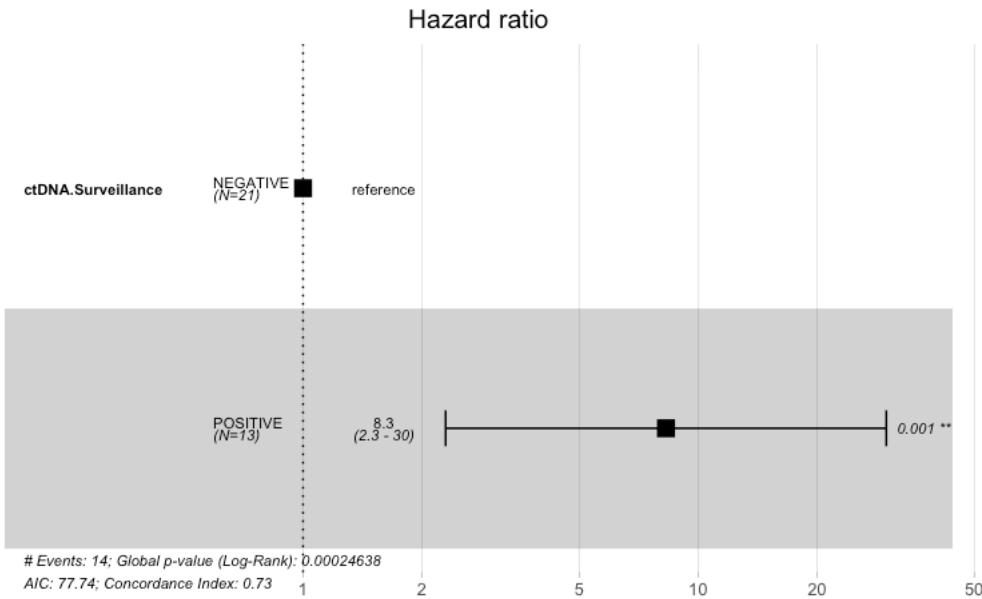
```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
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     7      6    0.538  0.138    0.2477    0.760
 24     3      4    0.231  0.117    0.0558    0.475
 36     1      1    0.154  0.100    0.0248    0.388
```

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)
```



Hide

```
summary(cox_fit)
```

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

n= 34, number of events= 14

            coef  exp(coef)  se(coef)      z  Pr(>|z|)    
ctDNA.SurveillancePOSITIVE 2.1148    8.2877  0.6551  3.228  0.00125 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.SurveillancePOSITIVE    8.288     0.1207    2.295    29.92 

Concordance= 0.731  (se = 0.064 )
Likelihood ratio test= 13.44 on 1 df,  p=2e-04
Wald test             = 10.42 on 1 df,  p=0.001
Score (logrank) test = 14.72 on 1 df,  p=1e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 8.29 (2.3-29.92); 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)
print(chi_square_test)
```

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

```
data: contingency_table
X-squared = 13.622, df = 1, p-value = 0.0002236
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 7.65e-05
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 3.750528 386.659717
sample estimates:
odds ratio
27.99066
```

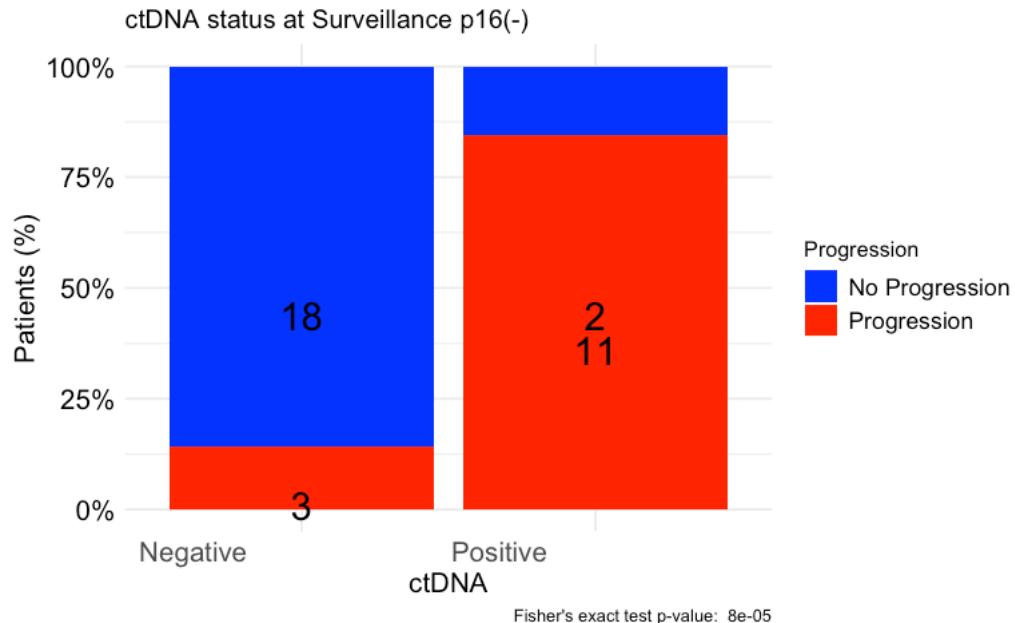
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	18	3
Positive	2	11

[Hide](#)

```
table_df <- as.data.frame(contingency_table)
table_df$Total <- ave(table_df$Freq, table_df$Var1, FUN = sum)
table_df$Percentage <- table_df$Freq / table_df$Total
table_df$MiddlePercentage <- table_df$Percentage / 2
ggplot(table_df, aes(x = Var1, y = Percentage, fill = Var2)) +
  geom_bar(stat = "identity") +
  geom_text(aes(y = MiddlePercentage, label = Freq), position = "stack", color = "black", vjust = 1.5, size = 7) +
  theme_minimal() +
  labs(title = "ctDNA 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
```



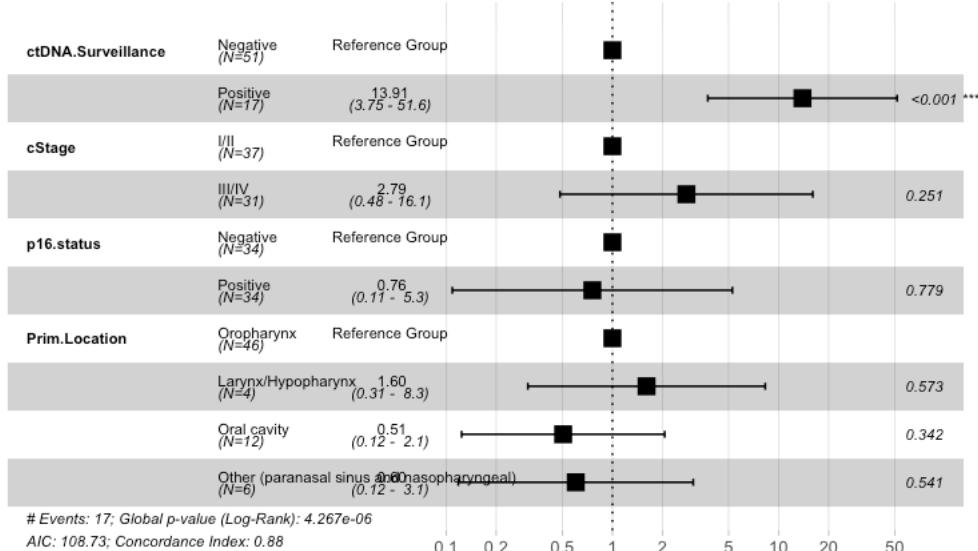
#Multivariate cox regression for PFS ctDNA status at surveillance

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")
```

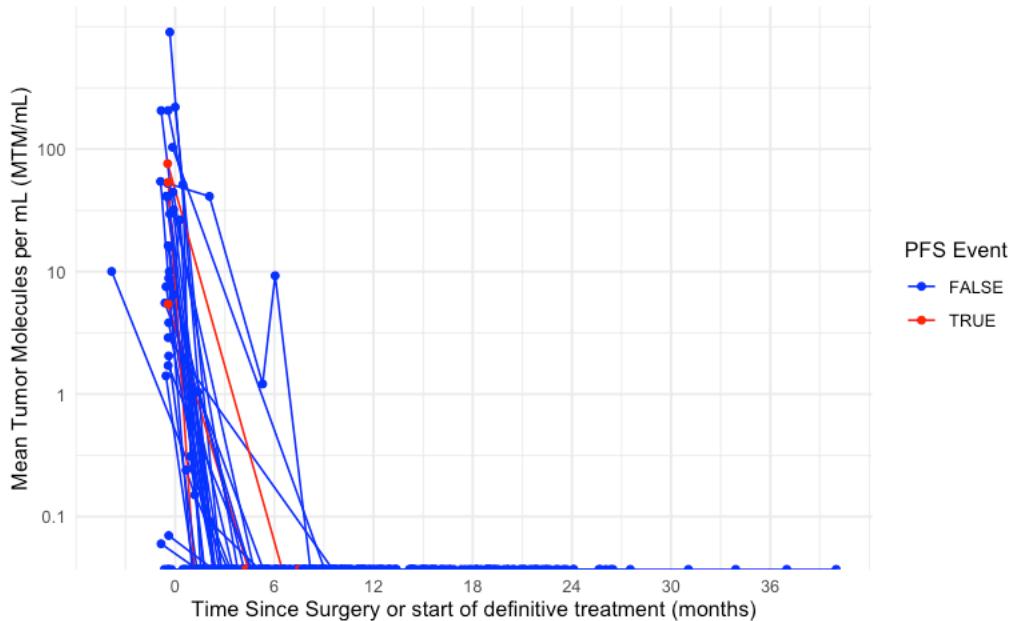
Number of unique patients with No Event: 48

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: 14

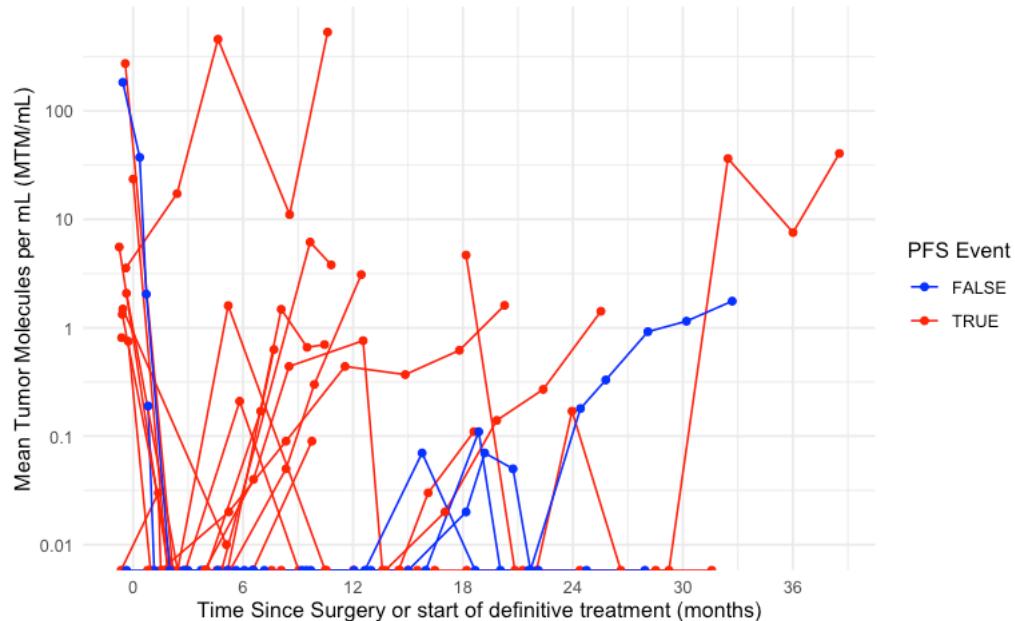
[Hide](#)

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

Number of unique patients with No Event: 3

[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	9	13.6	10.4	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)
```

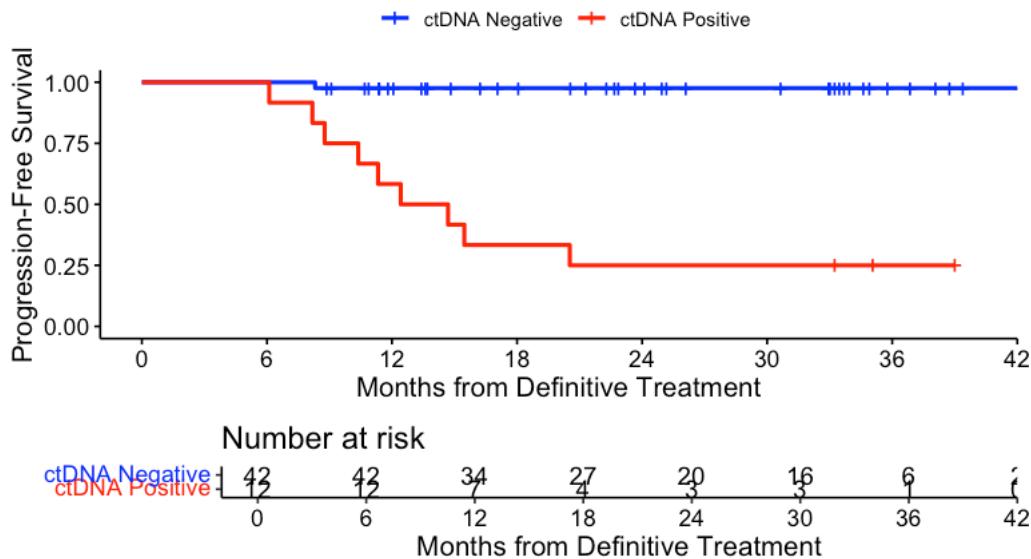
ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	42	1	0.02380952	2.380952
POSITIVE	12	9	0.75000000	75.000000

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		

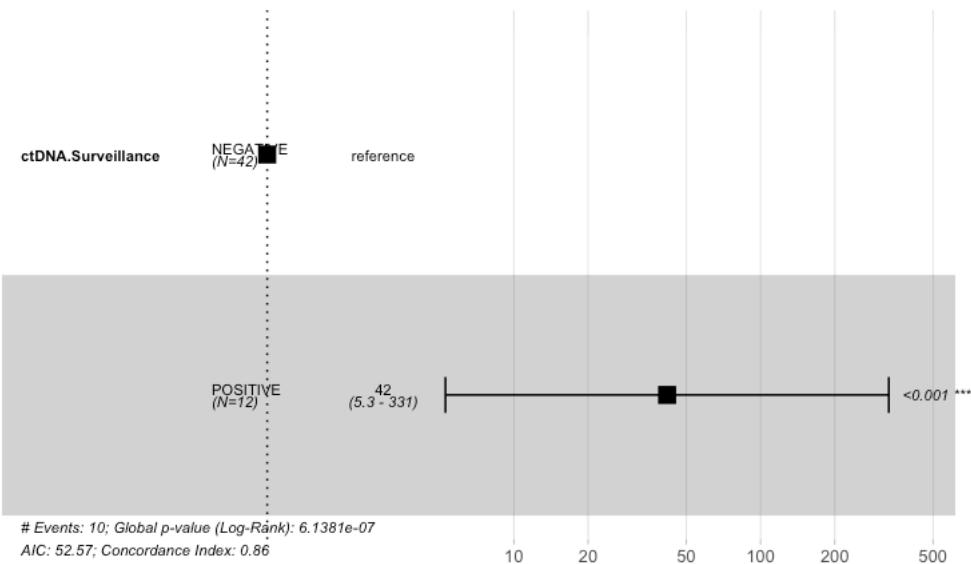
ctDNA.Surveillance=POSITIVE

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	7	5	0.583	0.142	0.2701	0.801		
24	3	4	0.250	0.125	0.0601	0.505		
36	1	0	0.250	0.125	0.0601	0.505		

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= 10

            coef  exp(coef)  se(coef)      z  Pr(>|z|)
ctDNA.SurveillancePOSITIVE  3.733    41.789    1.056  3.536 0.000407 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef)  exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    41.79     0.02393    5.278   330.9

Concordance= 0.857 (se = 0.057 )
Likelihood ratio test= 24.87 on 1 df,  p=6e-07
Wald test             = 12.5 on 1 df,  p=4e-04
Score (logrank) test = 35.14 on 1 df,  p=3e-09
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 41.79 (5.28-330.88); 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 = 27.984, df = 1, p-value = 1.223e-07
```

[Hide](#)

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

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

```
data: contingency_table
p-value = 3.889e-07
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 9.555849 5259.013678
sample estimates:
odds ratio
 98.99253
```

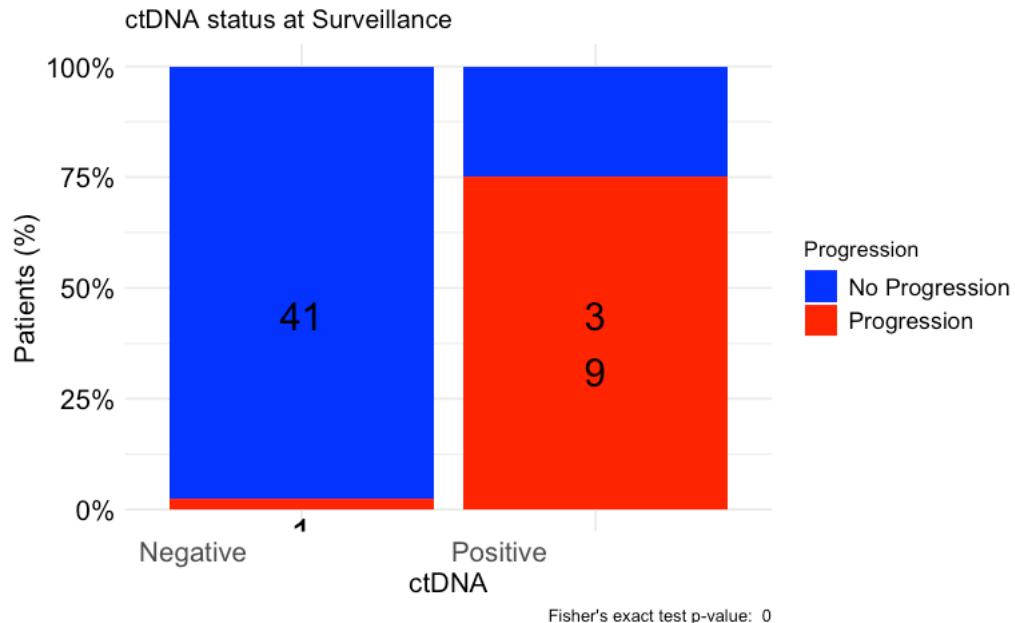
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	41	1
Positive	3	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",
       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"))))
```

```

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_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(days(date - window_start_date))) |>
  select(pts_id, biomarker_time, biomarker_status = status) |>
  filter(!is.na(biomarker_time))

glimpse(dt_biomarker)

```

```

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, ...
$ biomarker_status <chr> "NEGATIVE", "POSITIVE", "POSITIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", NA,
"NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NEGATIVE", "NE...

```

```
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

```

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)))

```

file:///Users/georgelaliotis/Downloads/CLIA HNSCC UNM Clinical Analysis 07272025.nb.html

```
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= 10
(52 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE  3.2868   26.7577   0.6568 5.004 5.62e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE    26.76     0.03737    7.385    96.95

Concordance= 0.756 (se = 0.079 )
Likelihood ratio test= 21.73 on 1 df,  p=3e-06
Wald test             = 25.04 on 1 df,  p=6e-07
Score (logrank) test = 54.47 on 1 df,  p=2e-13
```

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 = 26.76 (7.38-96.95); 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 4.07 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	22	13.2	10.4	24.4

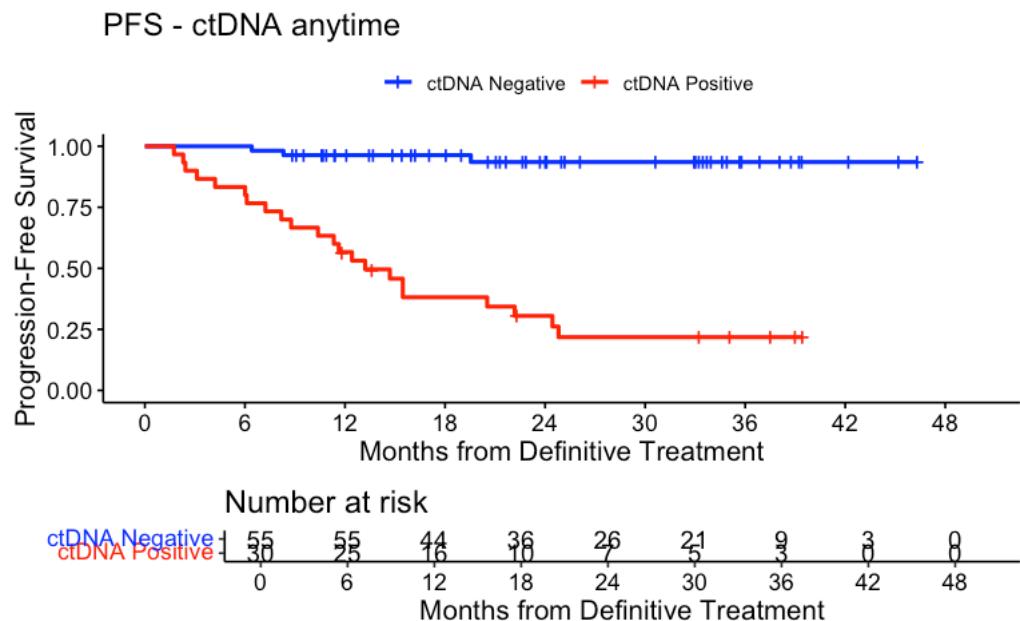
[Hide](#)

```
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	22	0.73333333	73.333333
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 T 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.anytime, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

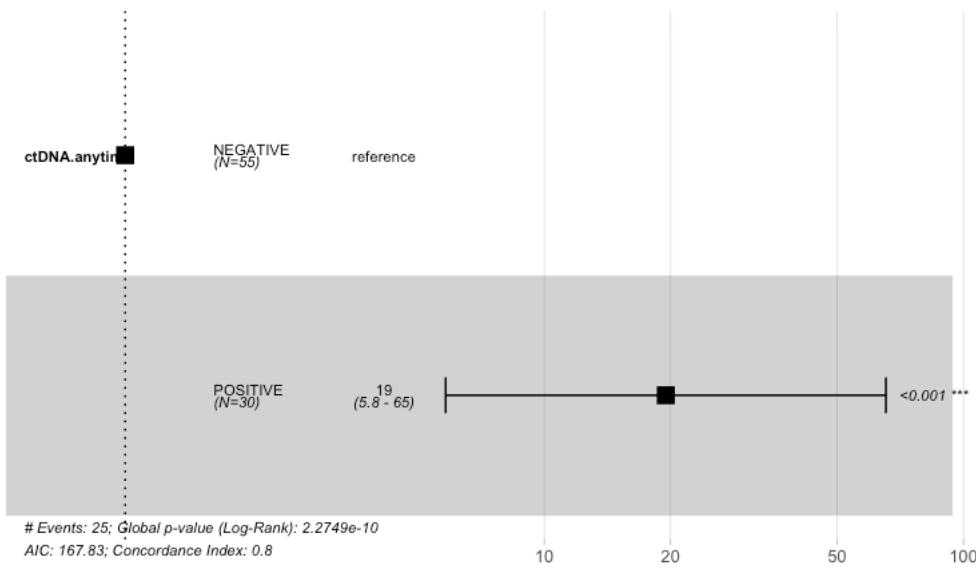
```
ctDNA.anytime=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0    55      0    1.000  0.0000    1.000    1.000
 12    44      2    0.964  0.0252    0.862    0.991
 24    26      1    0.935  0.0371    0.807    0.979
 36     9      0    0.935  0.0371    0.807    0.979
```

```
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    16      13    0.567  0.0905    0.3733    0.721
 24     7      7    0.305  0.0876    0.1485    0.478
 36     3      2    0.218  0.0814    0.0851    0.390
```

Hide

```
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= 25

      coef exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.anytimePOSITIVE  2.9697   19.4860   0.6169  4.814 1.48e-06 *** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95    
ctDNA.anytimePOSITIVE    19.49     0.05132     5.815    65.29    

Concordance= 0.799  (se = 0.037 )
Likelihood ratio test= 40.22 on 1 df,  p=2e-10
Wald test             = 23.17 on 1 df,  p=1e-06
Score (logrank) test = 45.33 on 1 df,  p=2e-11
```

Hide

```
cox_fit_summary <- summary(cox_fit)

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

```
[1] "HR = 19.49 (5.82-65.29); 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 = 39.873, df = 1, p-value = 2.71e-10
```

[Hide](#)

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

Fisher's Exact Test for Count Data

```
data: contingency_table
p-value = 7.175e-11
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 10.20553 282.29143
sample estimates:
odds ratio
 44.11145
```

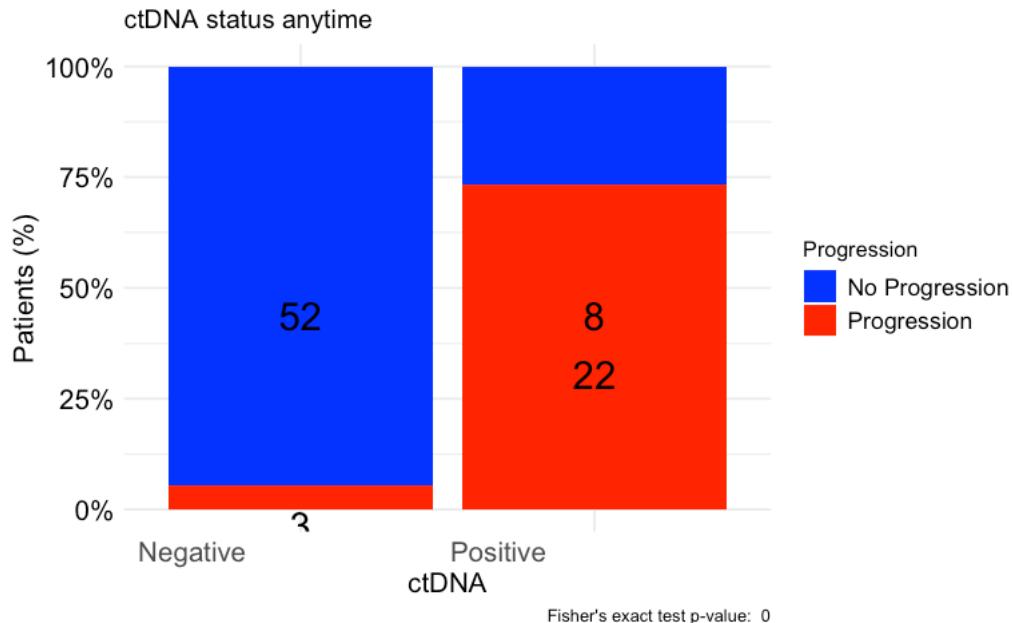
[Hide](#)

```
print(contingency_table)
```

	No Progression	Progression
Negative	52	3
Positive	8	22

[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
```

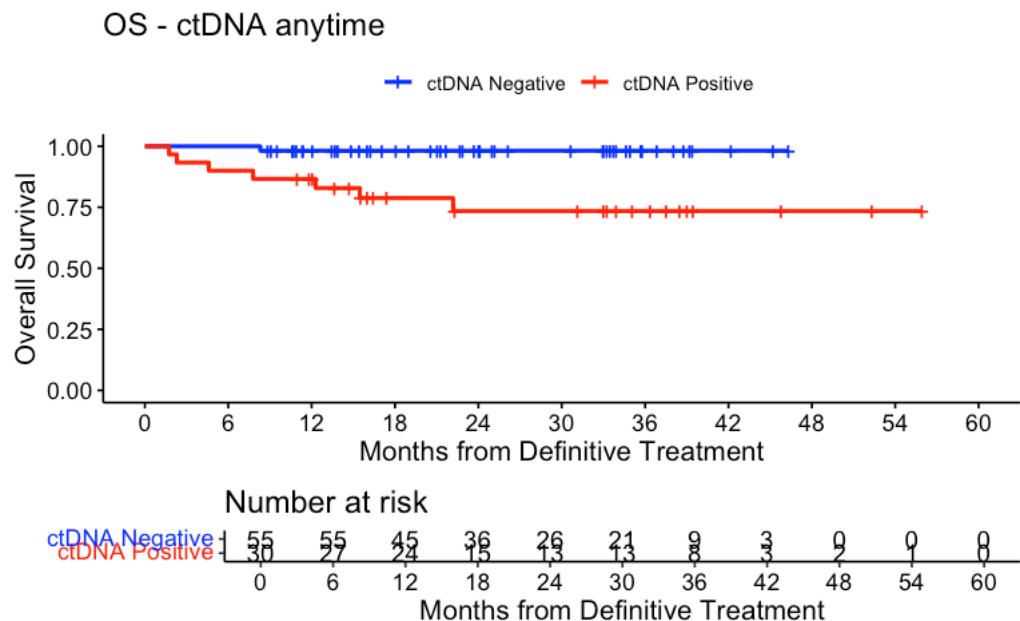
Hide

```
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="")
```



Hide

```
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")
```

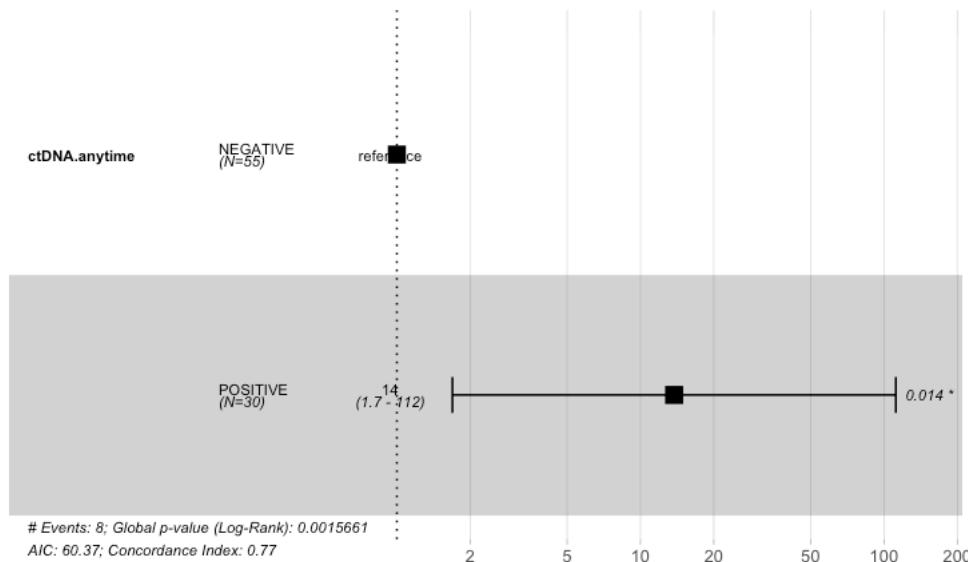
```
ctDNA.anytime=NEGATIVE
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 12     45      1    0.982   0.018     0.878    0.997
 24     26      0    0.982   0.018     0.878    0.997
 36      9      0    0.982   0.018     0.878    0.997
```

```
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
```

Hide

```
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
```

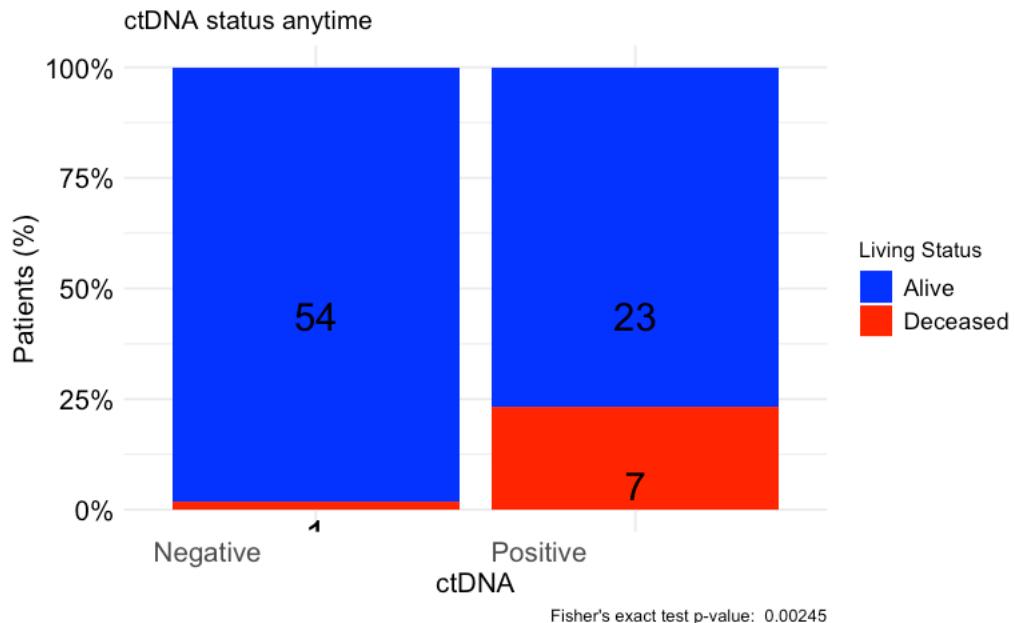
[Hide](#)

```
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 3.12 mo (0.00-21.49)