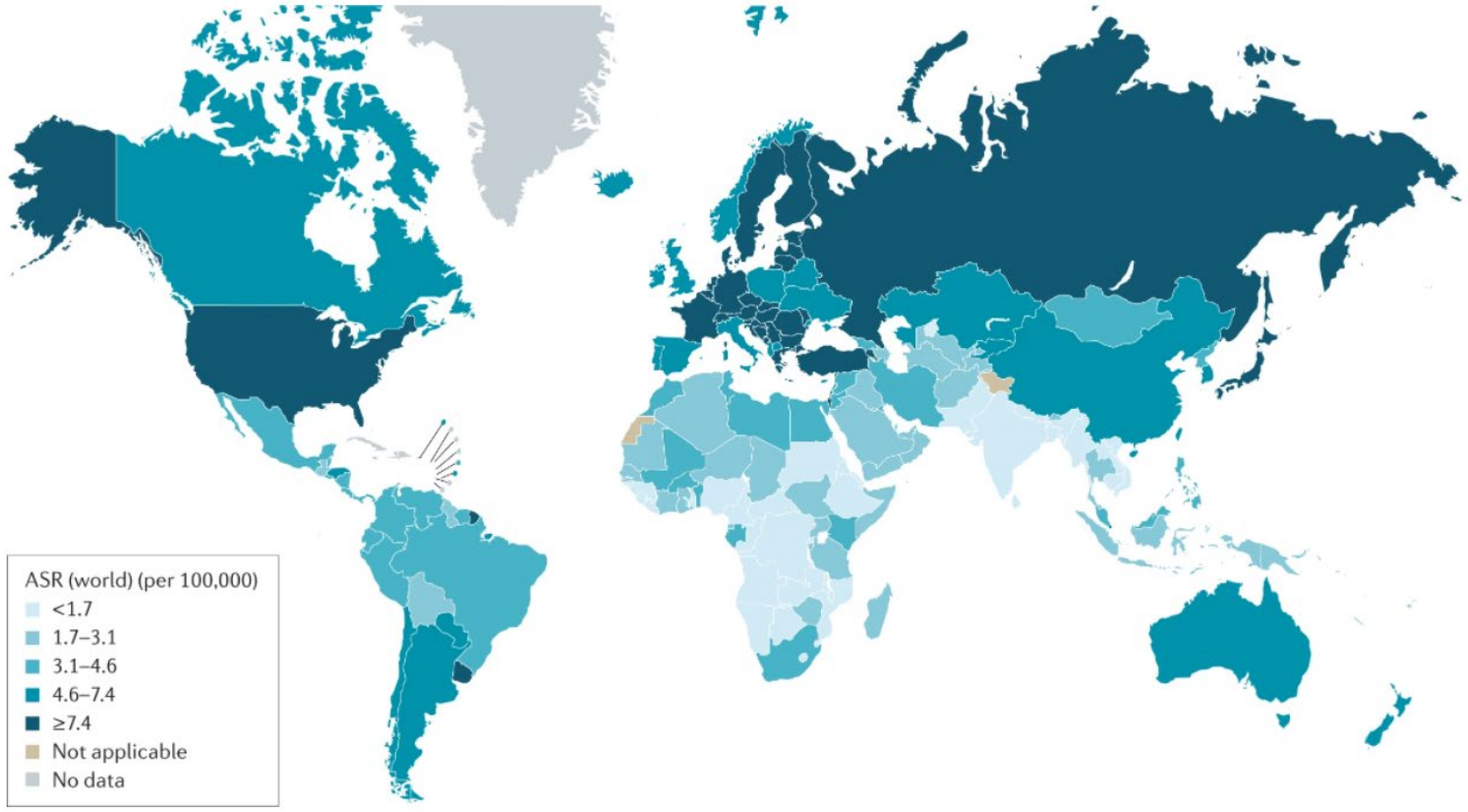
NYU

Analysis of the Pancreatic Adenocarcinoma (PAAD) Dataset

Genome characterization & tumor subclone phylogenies identification

Incidence of pancreatic cancer doubled in the past 20 yrs

- An aggressive disease that typically presents at an advanced stage
- In 2021 an estimated 60,430 Americans are diagnosed with pancreatic cancer

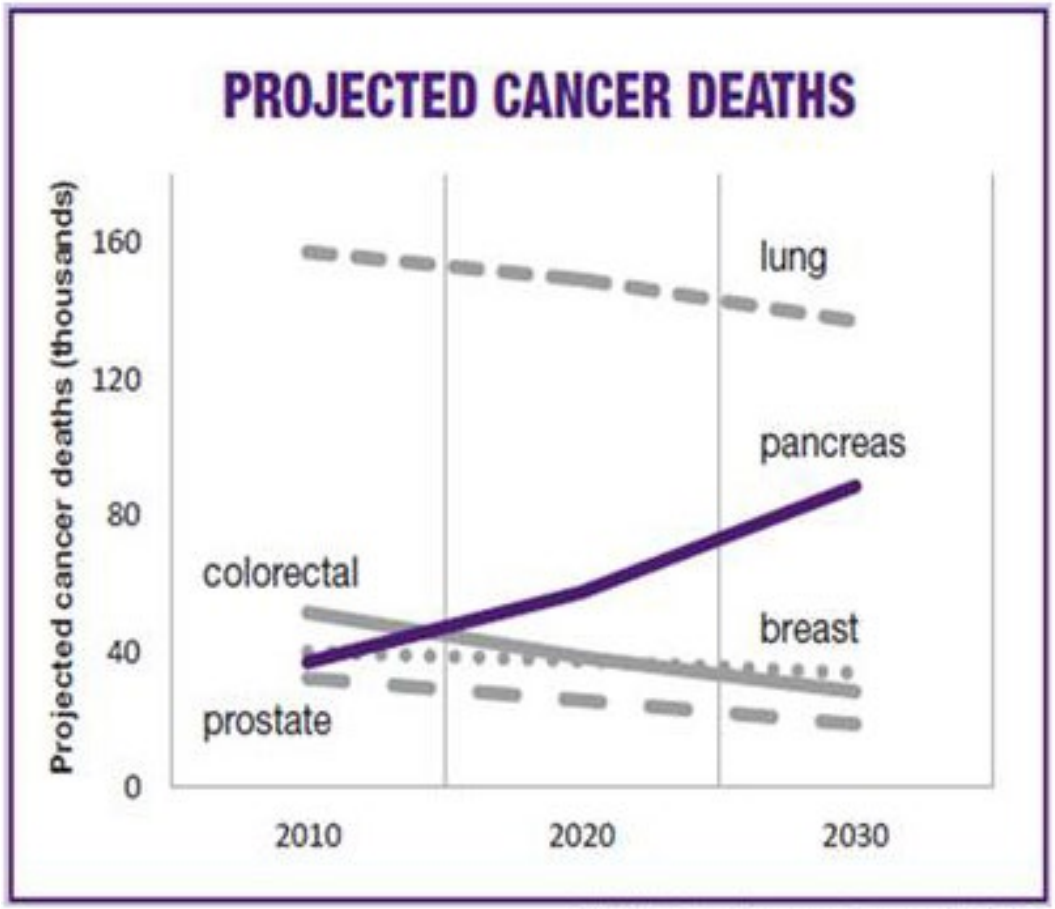


NYU

Ryan DP. N Engl J Med. 2014 Sep 11;371(11):1039-49.
Klein, A.P. Nat Rev Gastroenterol Hepatol 18, 493–502 (2021)

Lethal prognosis

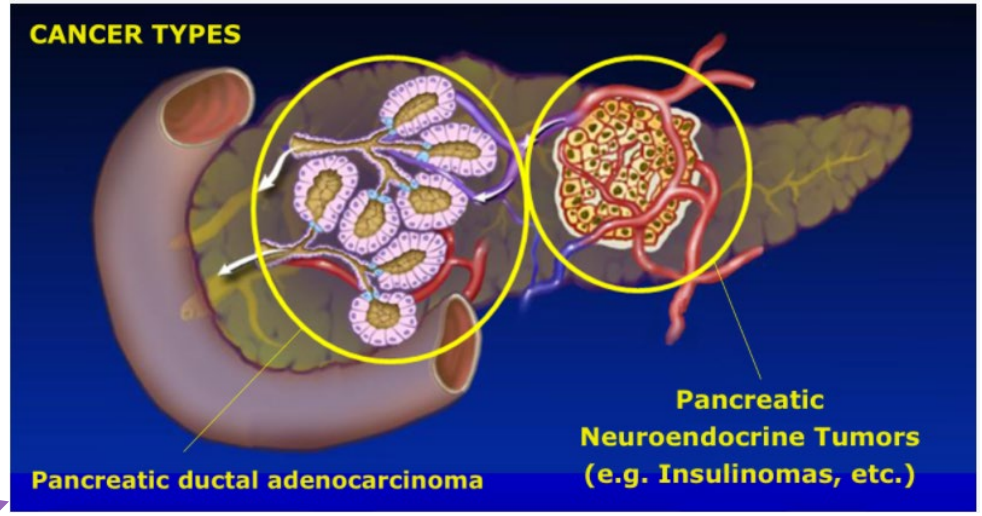
- Pancreatic cancer is currently the third leading cause of cancer death in the United States
- Epidemiologic projections indicate that it will be second only to lung cancer in its lethality by 2025



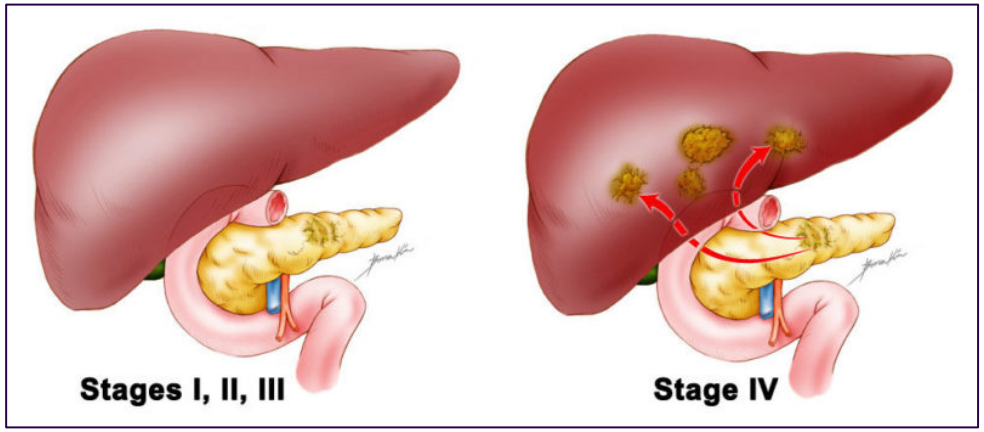
Rahib et al, Cancer Research, 2014

https://www.med.upenn.edu/pcrc/aboutpancreaticcancer.html

Cancer types & stages



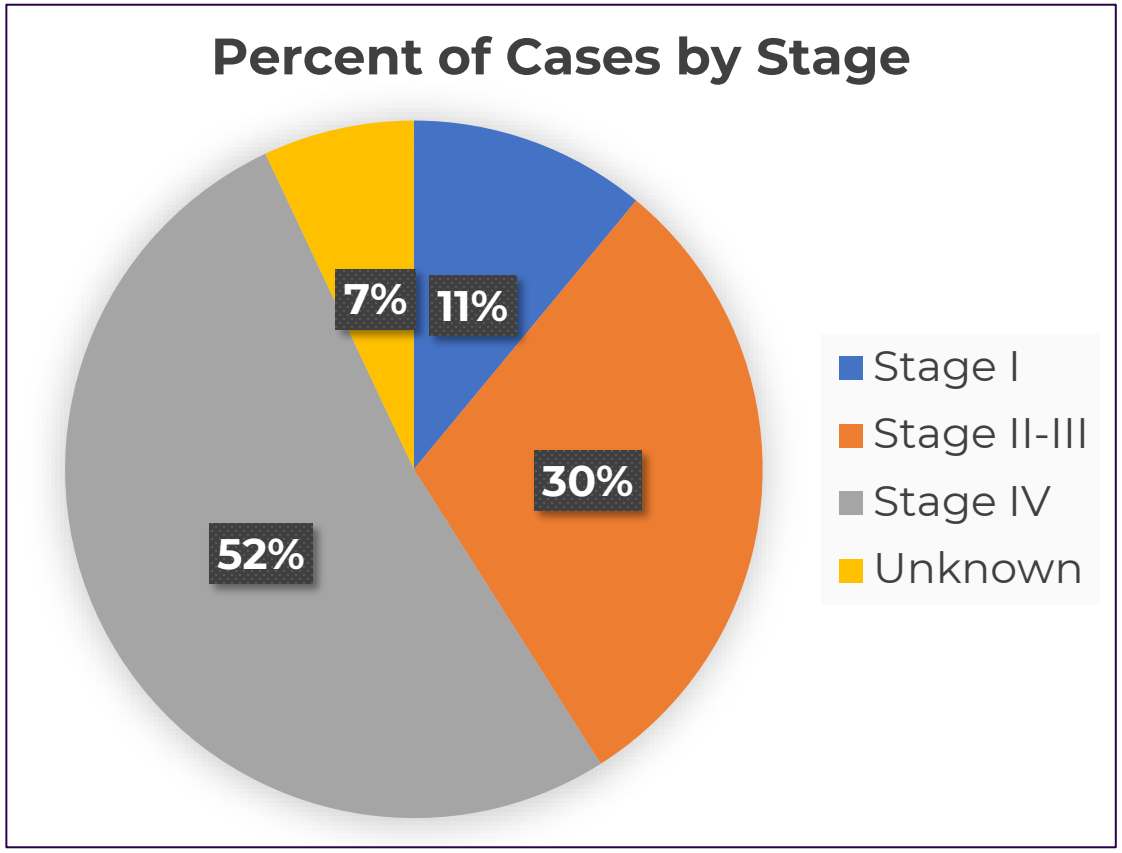
PDAC: 90%

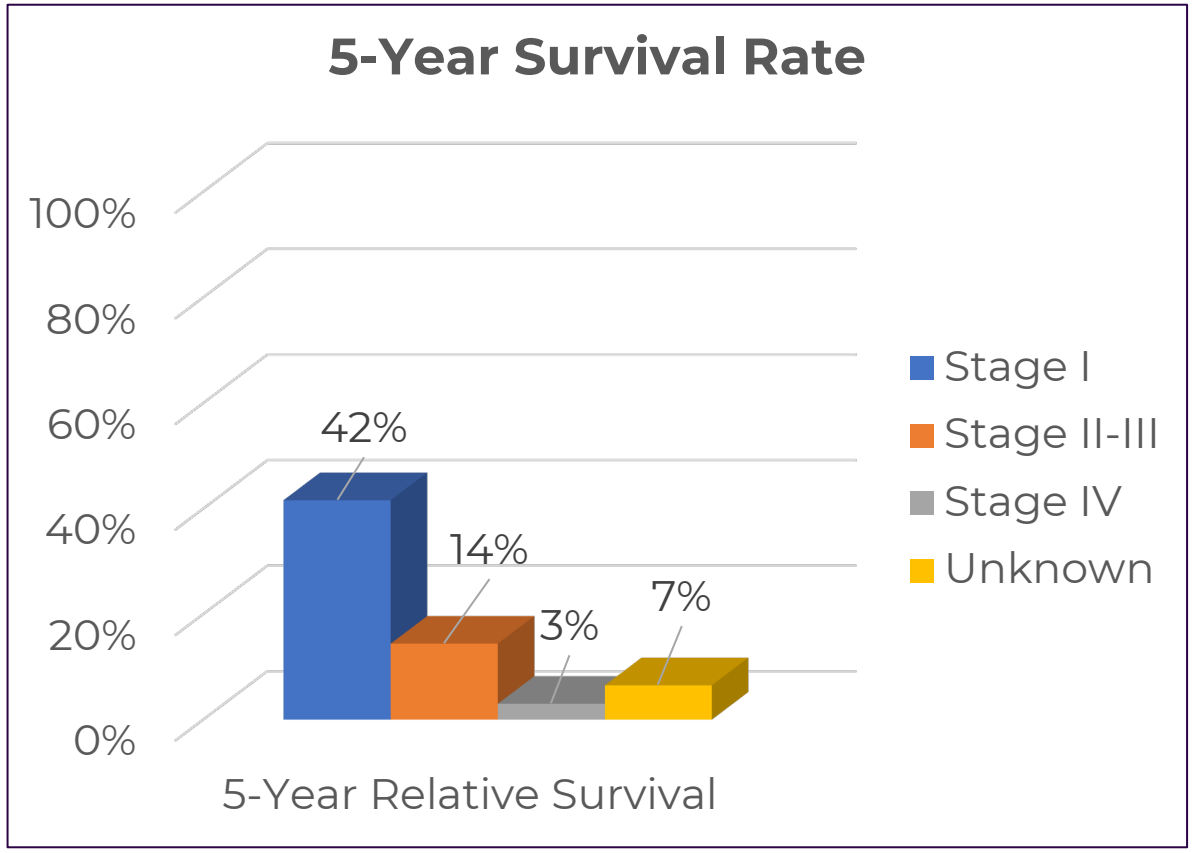


- **Stage I**: The tumor is found only in the pancreas.
- **Stage II**: The tumor has invaded nearby tissue but not nearby blood vessels. The cancer may have spread to the lymph nodes.
- **Stage III**: The tumor has invaded nearby blood vessels.
- **Stage IV**: The cancer has spread to a distant organ, such as the liver or lungs.

https://pancreatic.org/pancreatic-cancer

Percent of cases & 5-year survival by stage at diagnosis: PDAC





NYU

Revised from: Cancer Stat Facts

Workflow

1. Download **TCGA-PAAD** dataset
2. Landscape of top 10 mutated genes --
3. Genomic loci with hyper-mutations
4. Identify onco-driver genes --
5. Identify genes associated with pathological stages --
6. Identify affected pathways

Gene list

→ 10 genes

→ 6 genes

→ 4 genes

CAPRI

CAPRESE

The last universal common ancestor of the selected mutated genes

- **KRAS**
- TTN
- FLG

7. Target treatment

NYU

Download dataset

- Download ->

```
library(RTCGAToolbox)

# get data sets ==

# Valid aliases
getFirehoseDatasets()
getFirehoseRunningDates(last = 3)
getFirehoseAnalyzeDates(last=3)

# READ mutation data and clinical data
brcaData <- getFirehoseData(dataset="PAAD", runDate="20160128",
                            forceDownload=TRUE, clinical=TRUE, Mutation=TRUE)
brcaData

paad_maf2 <- getData(brcaData,'Mutation')
paad_clinic2 <- getData(brcaData,'clinical')
```

- Import ->

```
paad_maf <- read.csv("paad_maf.csv")
paad_maf <- paad_maf[,c(-1)]
paad_dataset_maf = import.MAF(paad_maf,
                        merge.mutation.types = FALSE)
```

- View ->

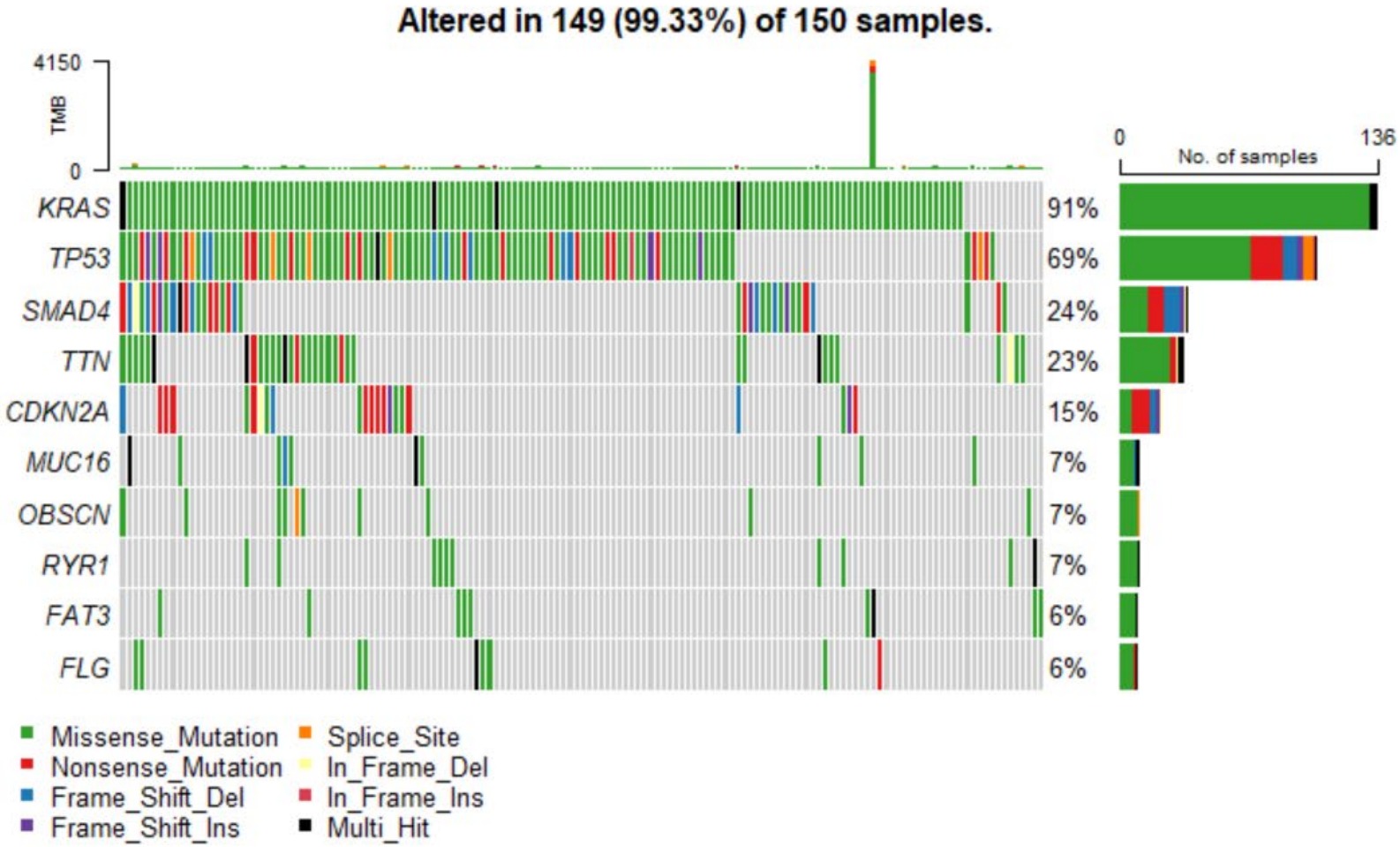
```
view(paad_dataset_maf)
```

```
-- TRONCO Dataset: n=126, m=20866, |G|=12939, patterns=0.
Events (types): Missense_Mutation, Silent, Nonsense_Mutation, Splice_Site, Frame_Shift_Del, 5'Flank, Frame_Shift_Ins, Intro
n, RNA, In_Frame_Del, IGR, 5'UTR, In_Frame_Ins, 3'UTR, Nonstop_Mutation, De_novo_Start_OutOfFrame, De_novo_Start_InFrame, St
art_Codon_Del, Start_Codon_Ins.
Colors (plot): #7FC97F, #97BEA0, #B0B4C1, #C8B1C7, #E1B8A8, #F9BF8A, #FDD58C, #FEED93, #E8EE9B, #9BB5A4, #4E7CAD, #6B4EA2, #
B2258F, #ED0679, #DA2950, #C74C28, #AB5D28, #886147, #666666.
Events (5 shown):
     G1 : Missense_Mutation A1BG
     G2 : Silent A1BG
     G3 : Missense_Mutation A1CF
     G4 : Missense_Mutation A2BP1
     G5 : Silent A2BP1
Genotypes (5 shown):
```

NYU

Landscape for top mutated genes

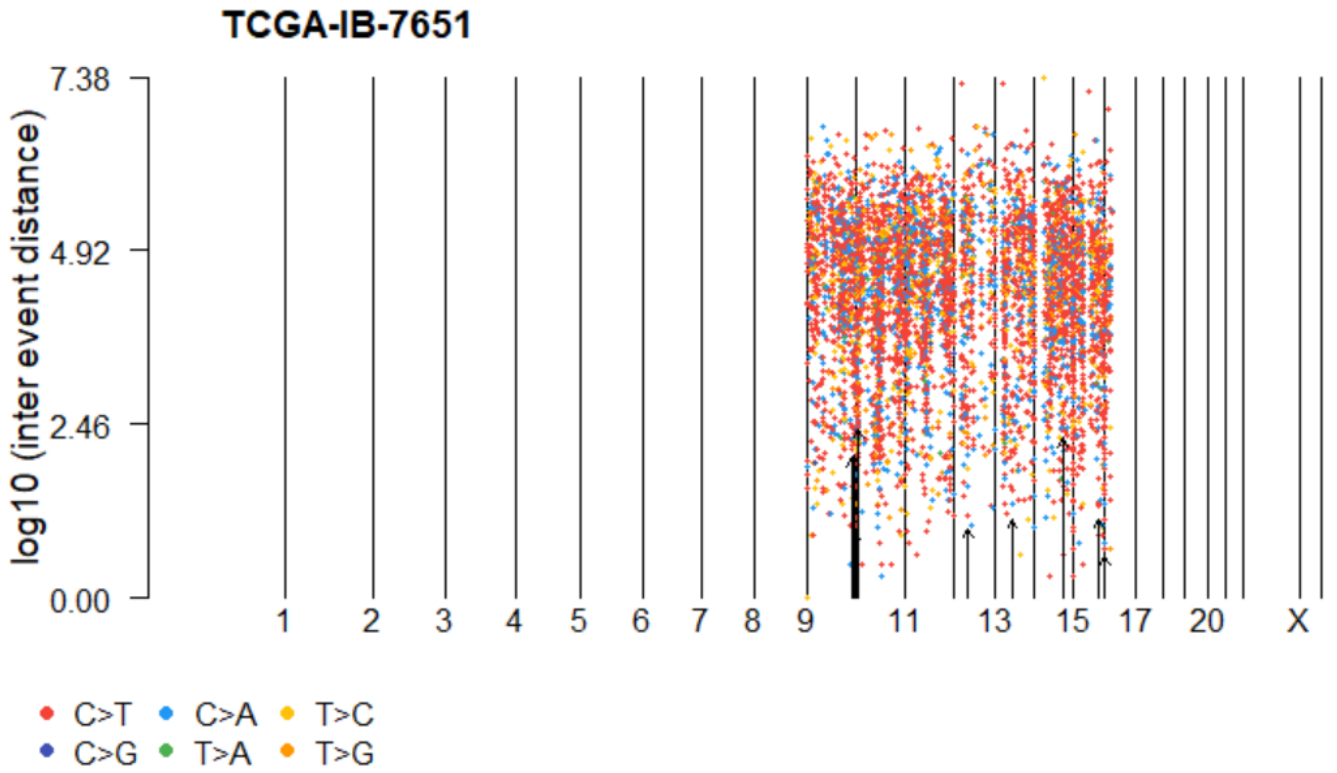
```
# Drawing oncoplots for top ten mutated genes
oncoplot(maf = paad, top = 10)
```



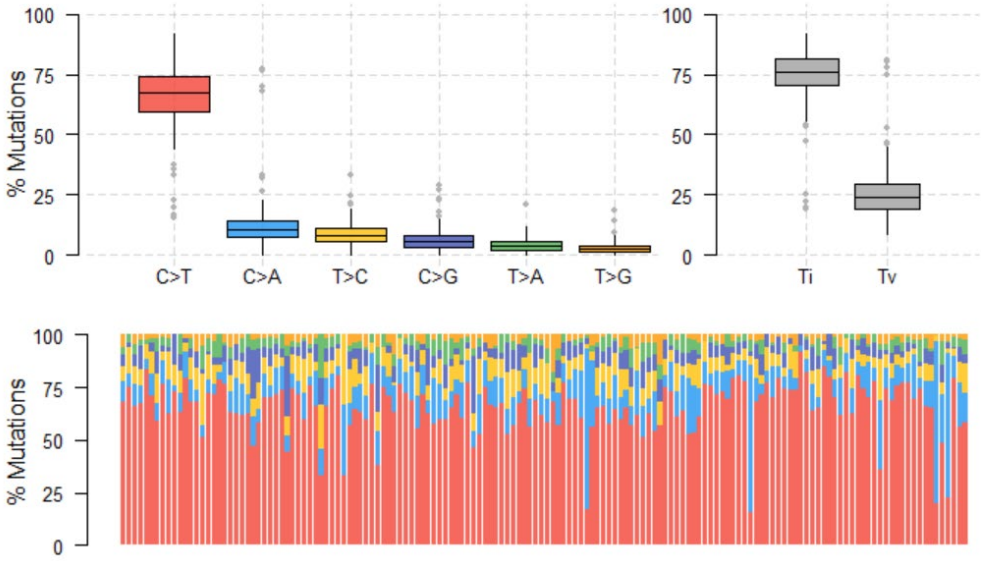
NYU

Genomic loci with hyper-mutations

```
# rainfall plot to show genomic loci with localized hyper-mutati ons
rainfallPlot(maf = paad, detectChangePoints = TRUE, pointSize = 0.4)
```



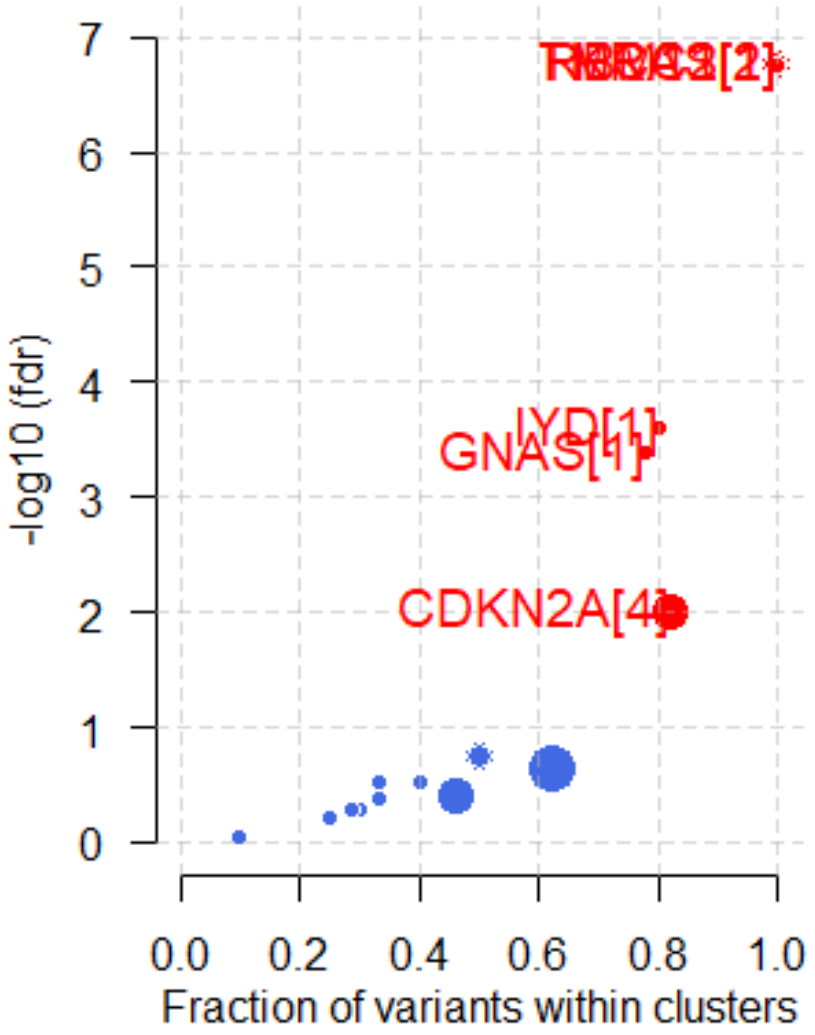

```
# transversion versus transition
paad_titv = titv(maf = paad, plot = FALSE, useSyn = TRUE)
#plot titv summary
plotTiTv(res = paad_titv)
```



NYU

Identify Onco-driver genes

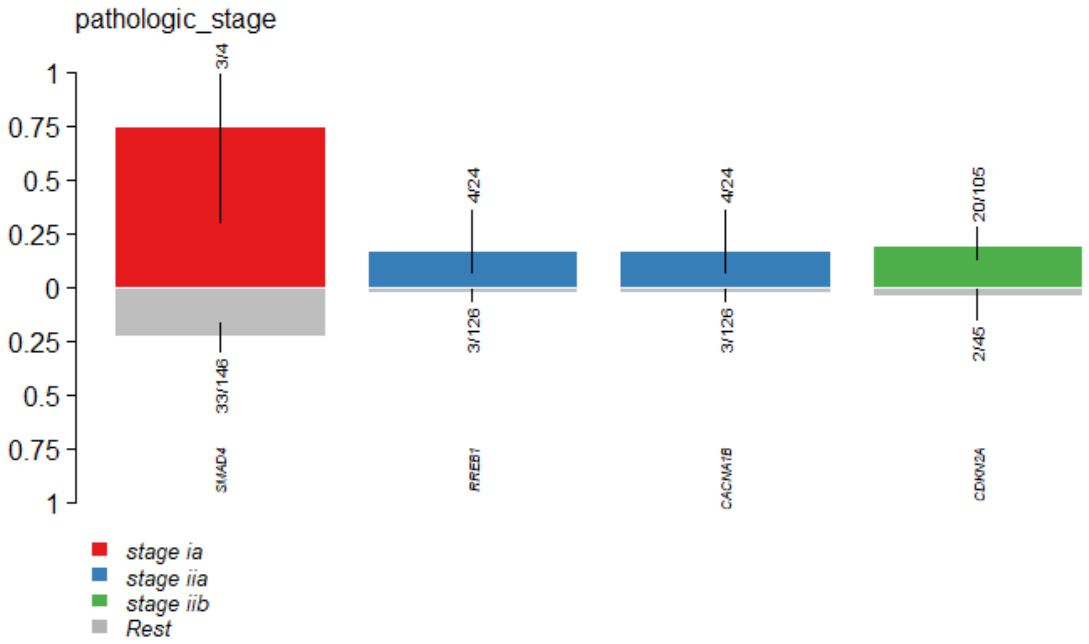
Hugo_Symbol <chr>	clusterScores <dbl>	protLen <int>	zscore <dbl>	pval <dbl>	fdr <dbl>
KRAS	0.9979079	189	5.5300608	1.600599e-08	1.760659e-07
RBM12	1.0000000	932	5.5461538	1.460110e-08	1.760659e-07
TMCC1	1.0000000	653	5.5461538	1.460110e-08	1.760659e-07
IYD	0.8000000	293	4.0076923	3.065747e-05	2.529241e-04
GNAS	0.7777778	1037	3.8367521	6.233609e-05	4.114182e-04
CDKN2A	0.6579488	167	2.9149908	1.778495e-03	9.781723e-03
PSG6	0.5000000	435	1.7000000	4.456546e-02	1.838325e-01
RET	0.5000000	1114	1.7000000	4.456546e-02	1.838325e-01
TP53	0.4805812	393	1.5506249	6.049580e-02	2.218179e-01
DDX10	0.4000000	875	0.9307692	1.759865e-01	3.056607e-01



NYU

Methods from: David Tamborero 2013
OncodriveCLUST article (http://bioinformati cs.oxfordjournals.org/content/early/2013/07/31/bioinformatics.btt 395.full)

Associations with pathological stages




```
Sample size per factor in pathologic_stage:
```

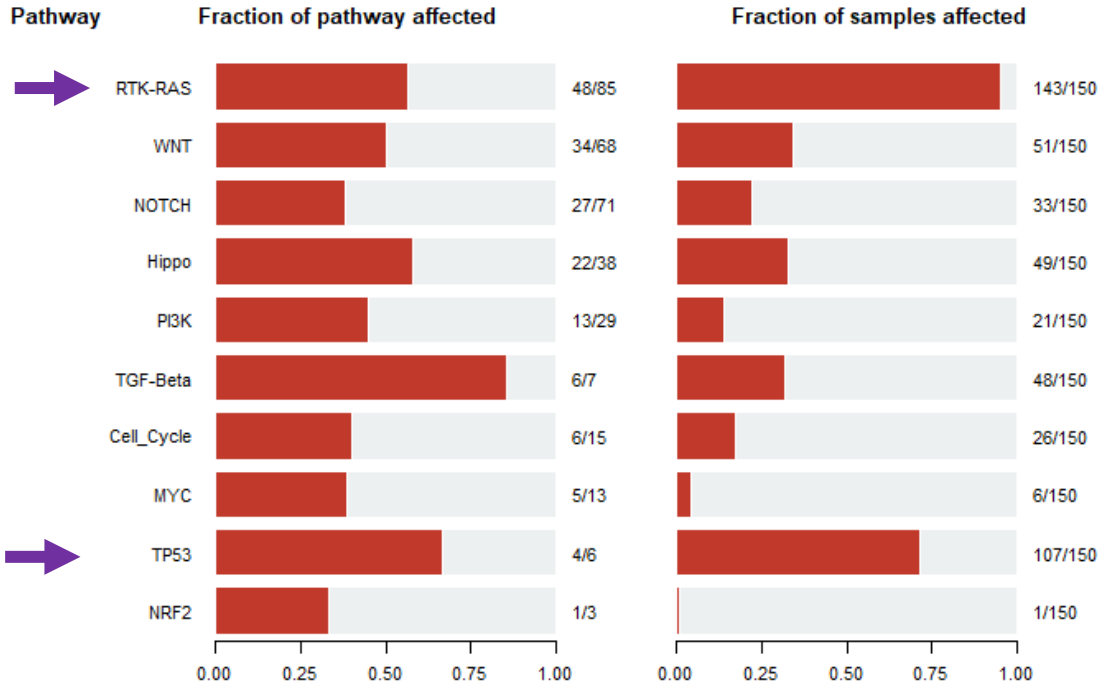
```
stage ia  stage ib stage iia stage iib stage iii  stage iv
       4         8        24       105         4         4
```

Hugo_Symbol <chr>	Group1 <chr>	Group2 <chr>	n_mutated_group1 <chr>	n_mutated_group2 <chr>	p_value <dbl>	OR <dbl>
RREB1	stage iia	Rest	4 of 24	3 of 126	0.01295703	8.0148346
CACNA1B	stage iia	Rest	4 of 24	3 of 126	0.01295703	8.0148346
CDKN2A	stage iib	Rest	20 of 105	2 of 45	0.02259533	5.0182634
RREB1	stage iib	Rest	2 of 105	5 of 45	0.02574350	0.1576112
SMAD4	stage ia	Rest	3 of 4	33 of 146	0.04308258	10.0683643

NYU

Methods from: David Tamborero 2013

Oncogene pathways



Pathway <chr>	N <int>	n_affected_genes <int>	fraction_affected <dbl>	Mutated_samples <int>
NRF2	3	1	0.3333333	1
TP53	6	4	0.6666667	107
MYC	13	5	0.3846154	6
Cell_Cycle	15	6	0.4000000	26
TGF-Beta	7	6	0.8571429	48
PI3K	29	13	0.4482759	21
Hippo	38	22	0.5789474	49
NOTCH	71	27	0.3802817	33
WNT	68	34	0.5000000	51
RTK-RAS	85	48	0.5647059	143

NYU

Methods from: David Tamborero 2013

Pre-process for CAPRI/CAPRESE

	Events	Genes	Samples
	20,866	12,939	126
1.	1,291	435	126
2.	246	111	126
3.	57	14	126

1.

```
# select only the genes mutated at least in the 5% of the patients
alterations <- events.selection(as.alterations(paad_dataset_maf), filter.freq = .05)
paad_clean <- events.selection(paad_dataset_maf,
                               filter.in.names=c(as.genes(alterations)))
```

2.

```
# data consolidation
paad_consolid <- consolidate.data(paad_clean)
paad_consolid
```

3.

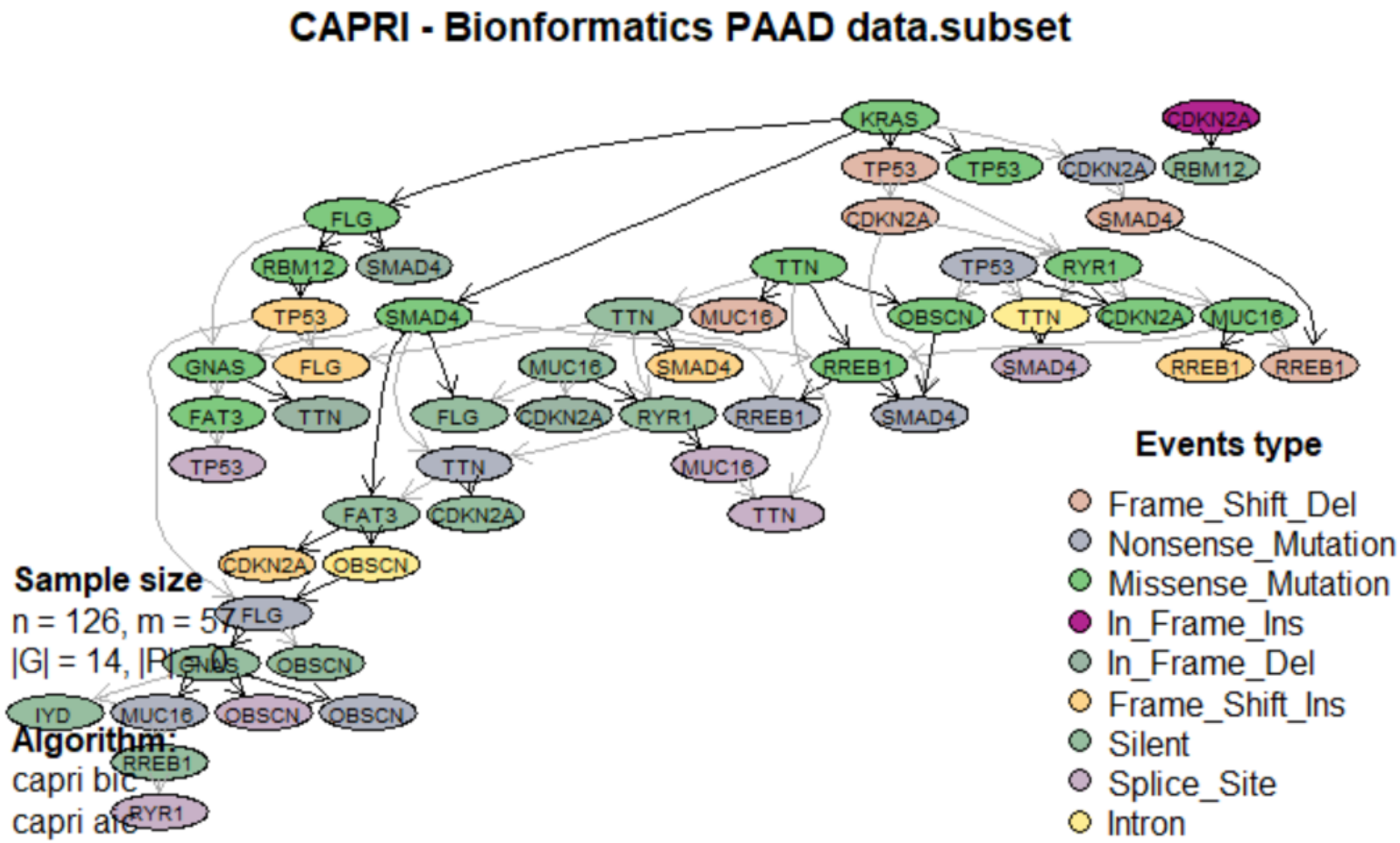
```
# subset according to previous analysis===
paad_clean_consolid_subset <- events.selection(paad_clean,
                                               filter.in.names = c('KRAS','TP53','TTN'
,'SMAD4','CDKN2A','FAT3','MUC16','RYR1','FLG','OBSCN','RBM12','TMCC1','IYD','GNAS'
,'RREB1','CACNA1B' ) )
```

NYU

CAPRI

```
tronco.plot(paad_subset_model_capri)
```

```
*** Expanding hypotheses syntax as graph nodes:
*** Rendering graphics
Nodes with no incoming/outgoing edges will not be displayed.
RGraphviz object prepared.
Plotting graph and adding legends.
```

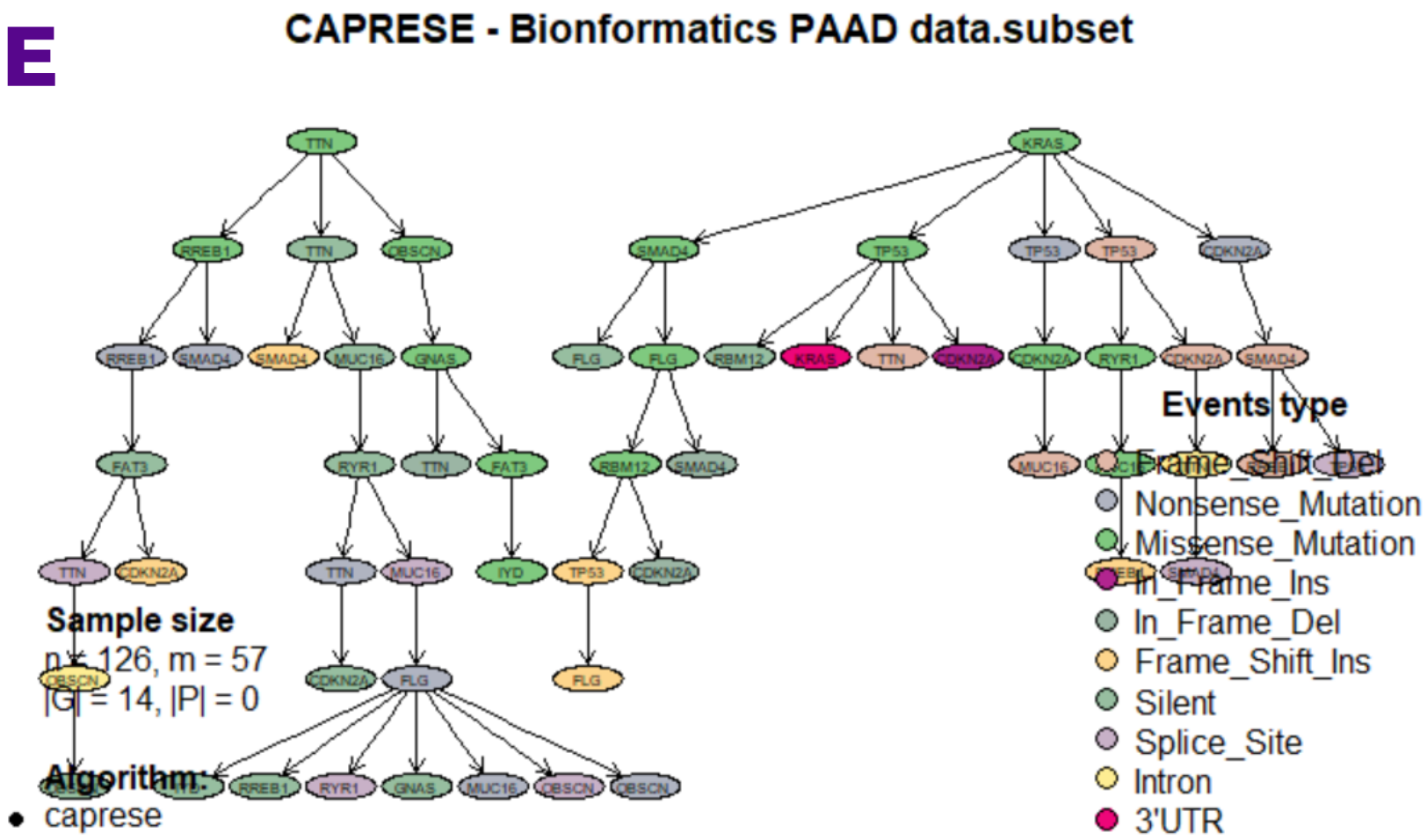


NYU

```
tronco.plot(paad_subset_model_caprese)
```

```
*** Expanding hypotheses syntax as graph nodes:
*** Rendering graphics
Nodes with no incoming/outgoing edges will not be displayed.
RGraphviz object prepared.
Plotting graph and adding legends.
```

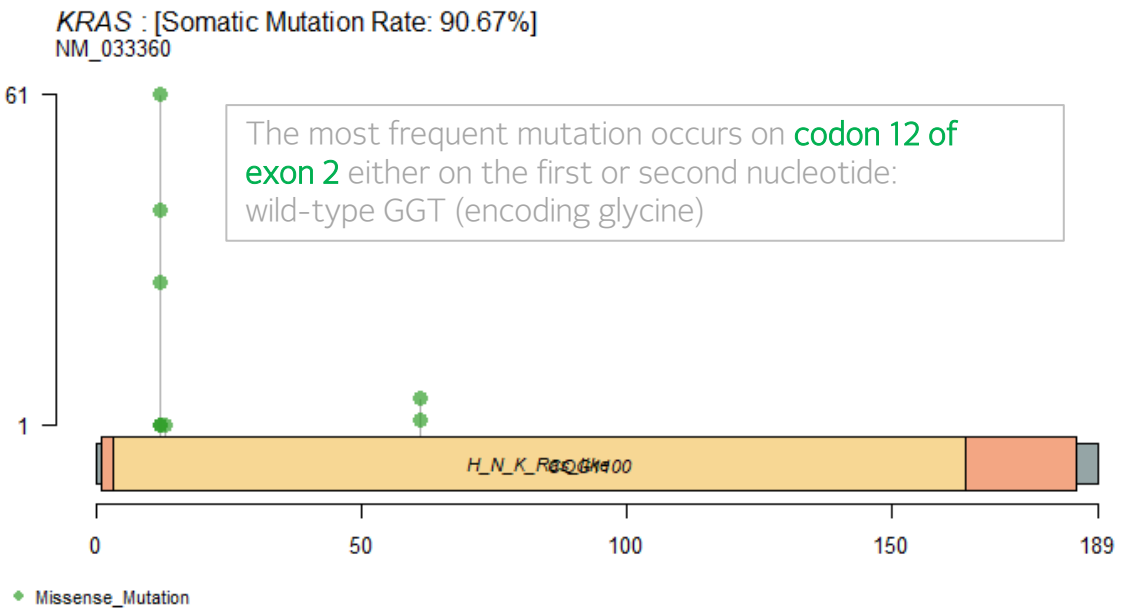
CAPRESE



NYU

Treatment against KRAS

```
#lollipop plot for KRAS, which is one of the most frequent mutated gene in PAAD.
lollipopPlot(
  maf = paad,
  gene = 'KRAS',
  AACol = 'Protein_Change',
  showMutationRate = TRUE)
```



The most frequent mutation occurs on codon 12 of exon 2 either on the first or second nucleotide: wild-type GGT (encoding glycine)

Genome

```
kras_dgi = drugInteractions(genes = "KRAS", drugs = TRUE)
```

```
Number of claimed drugs for given genes:
```

Gene <chr>	N <int>
KRAS	295

1 row

```
Warning message:
package 'maftools' was built under R version 4.1.1
```

Hide

```
#Printing selected columns.
kras_dgi[,.(Gene, interaction_types, drug_name, drug_claim_name)]
```

295 drug candidates

Gene <chr>	interaction_types <chr>	drug_name <chr>	drug_claim_name <chr>
KRAS		LENALIDOMIDE	lenalidomide
KRAS		IMGATUZUMAB	GA201
KRAS			3144
KRAS		SELUMETINIB	Selumetinib
KRAS		BUPARLISIB	BKM120
KRAS		GEFITINIB	Gefitinib
KRAS			CGM097
KRAS		IRX-4204	IRX4204
KRAS		MK-2206	MK2206
KRAS			Radiotherapy

1-10 of 295 rows Previous 1 2 3 4 5 6 … 30 Next

Drug

NYU

Conclusion

- Obtained TCGA-PAAD dataset and performed genomic analysis
- our result suggests
  - Oncogenic driver role of **KRAS, TP53**
  - **RTK-RASS** and **TP53 pathway** enrichment

- Screen potential treatments
  - genome target
  - drug candidates are created for further explorations.

Session Information

```
sessionInfo()
```

```
R version 4.1.0 (2021-05-18)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19044)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252 LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] TRONCO_2.24.0       forcats_0.5.1       dplyr_1.0.7         purrr_0.3.4         readr_2.1.1
 [6] tidyr_1.1.4         tibble_3.1.6        ggplot2_3.3.5       tidyverse_1.3.1     stringr_1.4.0
[11] maftools_2.8.05     RTCGAToolbox_2.22.1

loaded via a namespace (and not attached):
  [1] bitops_1.0-7                matrixStats_0.61.0          fs_1.5.1                    lubridate_1.8.0
  [5] doParallel_1.0.16           RColorBrewer_1.1-2          httr_1.4.2                  GenomeInfoDb_1.28.4
  [9] Rgraphviz_2.36.0            tools_4.1.0                 backports_1.3.0             bslib_0.3.1
 [13] utf8_1.2.2                  R6_2.5.1                    DBI_1.1.1                   BiocGenerics_0.38.0
 [17] colorspace_2.0-2            withr_2.4.3                 gridExtra_2.3               tidyselect_1.1.1
 [21] compiler_4.1.0              graph_1.70.0                cli_3.1.0                   rvest_1.0.2
 [25] Biobase_2.52.0              xml2_1.3.3                  DelayedArray_0.18.0         sass_0.4.0
 [29] scales_1.1.1                RCircos_1.2.1               digest_0.6.28               rmarkdown_2.11
 [33] R.utils_2.11.0              XVector_0.32.0              pkgconfig_2.0.3             htmltools_0.5.2
 [37] MatrixGenerics_1.4.3        dbplyr_2.1.1                fastmap_1.1.0               limma_3.48.3
 [41] GlobalOptions_0.1.2         rlang_0.4.12                readxl_1.3.1                rstudioapi_0.13
 [45] farver_2.1.0                shape_1.4.6                 jquerylib_0.1.4             generics_0.1.1
 [49] jsonlite_1.7.2              gtools_3.9.2                R.oo_1.24.0                 RCurl_1.98-1.5
 [53] magrittr_2.0.1              GenomeInfoDbData_1.2.6      R.matlab_3.6.2              Matrix_1.3-4
 [57] Rcpp_1.0.7                  munsell_0.5.0               S4Vectors_0.30.2            fansi_0.5.0
 [61] lifecycle_1.0.1             R.methodsS3_1.8.1           bnlearn_4.7                 stringi_1.7.6
 [65] yaml_2.2.1                  RaggedExperiment_1.16.0     RJSONIO_1.3-1.6             SummarizedExperiment_1.22.0
 [69] zlibbioc_1.38.0             grid_4.1.0                  parallel_4.1.0              crayon_1.4.2
 [73] lattice_0.20-45             haven_2.4.3                 splines_4.1.0               circlize_0.4.13
 [77] hms_1.1.1                   knitr_1.36                  pillar_1.6.4                igraph_1.2.9
 [81] GenomicRanges_1.44.0        cgdsr_1.3.0                 codetools_0.2-18            stats4_4.1.0
 [85] reprex_2.0.1                XML_3.99-0.8                glue_1.5.1                  evaluate_0.14
 [89] data.table_1.14.2           BiocManager_1.30.16         modelr_0.1.8                foreach_1.5.1
 [93] vctrs_0.3.8                 tzdb_0.2.0                  cellranger_1.1.0            gtable_0.3.0
 [97] assertthat_0.2.1            xfun_0.28                   xtable_1.8-4                broom_0.7.10
[101] survival_3.2-13             pheatmap_1.0.12             iterators_1.0.13            IRanges_2.26.0
[105] ellipsis_0.3.2
```

NYU