Soamtic_mutation

Subinoy Biswas

9/23/2019

Taking publicly available dataset from Esophageal carcinoma, acute myeloid leukemia and breast invasive carcinoma the following analysis has been done. I try to include as much results possible as I can.

```
## –Reading
## –Validating
## –Silent variants: 475
## –Summarizing
## –Processing clinical data
## –Finished in 0.752s elapsed (1.067s cpu)
```

```
## An object of class  MAF
##                     ID          summary  Mean Median
## 1:          NCBI_Build               37    NA     NA
## 2:              Center genome.wustl.edu    NA     NA
## 3:             Samples              193    NA     NA
## 4:              nGenes             1241    NA     NA
## 5:      Frame_Shift_Del               52 0.271      0
## 6:      Frame_Shift_Ins               91 0.474      0
## 7:         In_Frame_Del               10 0.052      0
## 8:         In_Frame_Ins               42 0.219      0
## 9:   Missense_Mutation             1342 6.990      7
## 10:  Nonsense_Mutation              103 0.536      0
## 11:         Splice_Site               92 0.479      0
## 12:               total             1732 9.021      9
```

Tumor_Sample_Barcode <fctr>	Frame_Shift_Del <int>	Frame_Shift_Ins <int>	In_Frame_Del <int>	In_Frame_Ins <int>	Mi…
TCGA-AB-3009	0	5	0	1	
TCGA-AB-2807	1	0	1	0	
TCGA-AB-2959	0	0	0	0	
TCGA-AB-3002	0	0	0	0	
TCGA-AB-2849	0	1	0	0	
TCGA-AB-2923	1	1	0	0	
TCGA-AB-2972	0	1	0	0	
TCGA-AB-2912	0	1	0	0	

Tumor_Sample_Barcode <fctr>	Frame_Shift_Del <int>	Frame_Shift_Ins <int>	In_Frame_Del <int>	In_Frame_Ins <int>	Mis
TCGA-AB-2877	0	0	0	1	
TCGA-AB-2927	0	2	0	0	

1-10 of 192 rows | 1-6 of 9 columns   Previous **1** 2 3 4 5 6 … 20 Next

Hugo_Sym... <chr>	Frame_Shift_Del <int>	Frame_Shift_Ins <int>	In_Frame_Del <int>	In_Frame_Ins <int>	Missense_Muta
FLT3	0	0	1	33	
DNMT3A	4	0	0	0	
NPM1	0	33	0	0	
IDH2	0	0	0	0	
IDH1	0	0	0	0	
TET2	10	4	0	0	
RUNX1	1	3	1	0	
TP53	2	2	0	0	
NRAS	0	0	0	0	
CEBPA	6	2	2	6	

1-10 of 1,241 rows | 1-6 of 11 columns   Previous **1** 2 3 4 5 6 … 125 Next

Tumor_Sample_Barcode <fctr>	FAB_classification <fctr>	days_to_last_followup <fctr>	Overall_Survival_Status <fctr>
TCGA-AB-2802	M4	365	1
TCGA-AB-2803	M3	792	1
TCGA-AB-2804	M3	2557	0
TCGA-AB-2805	M0	577	1
TCGA-AB-2806	M1	945	1
TCGA-AB-2807	M1	181	1
TCGA-AB-2808	M2	2861	0
TCGA-AB-2809	M3	62	1
TCGA-AB-2810	M2	31	1
TCGA-AB-2812	M2	366	1

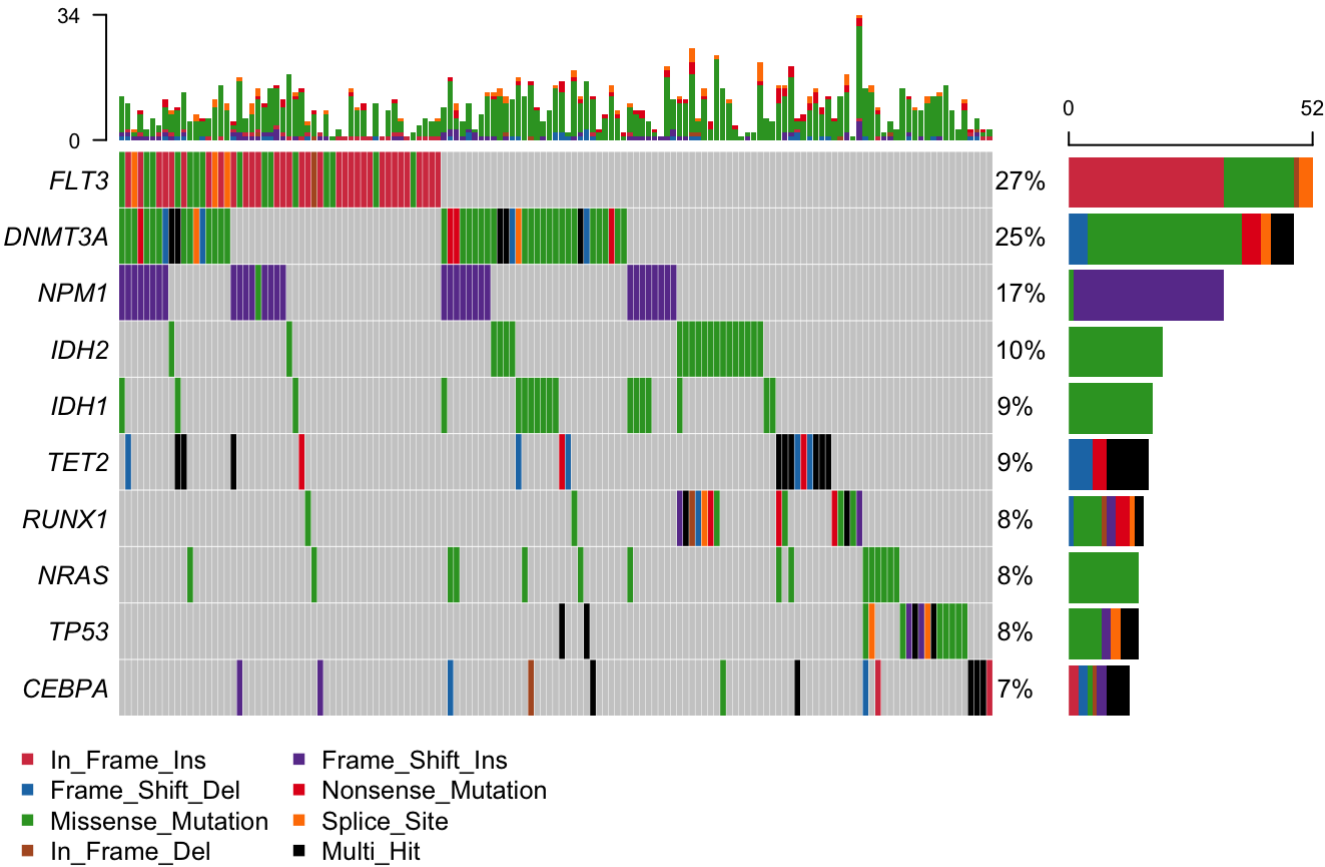1-10 of 193 rows   Previous **1** 2 3 4 5 6 … 20 Next

summary of the maf file, which displays number of variants in each sample as a stacked barplot and variant types as a boxplot summarized by Variant_Classification
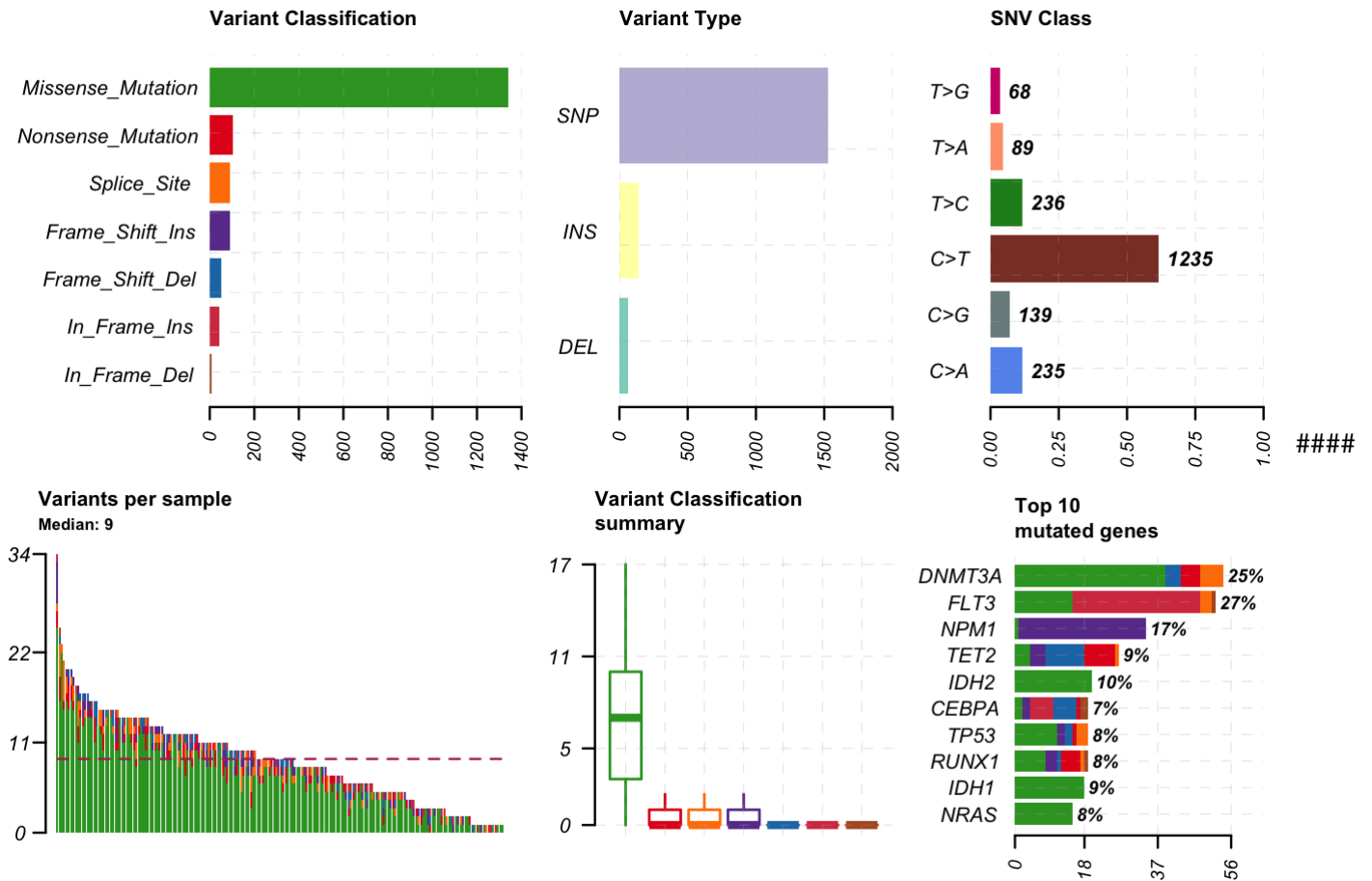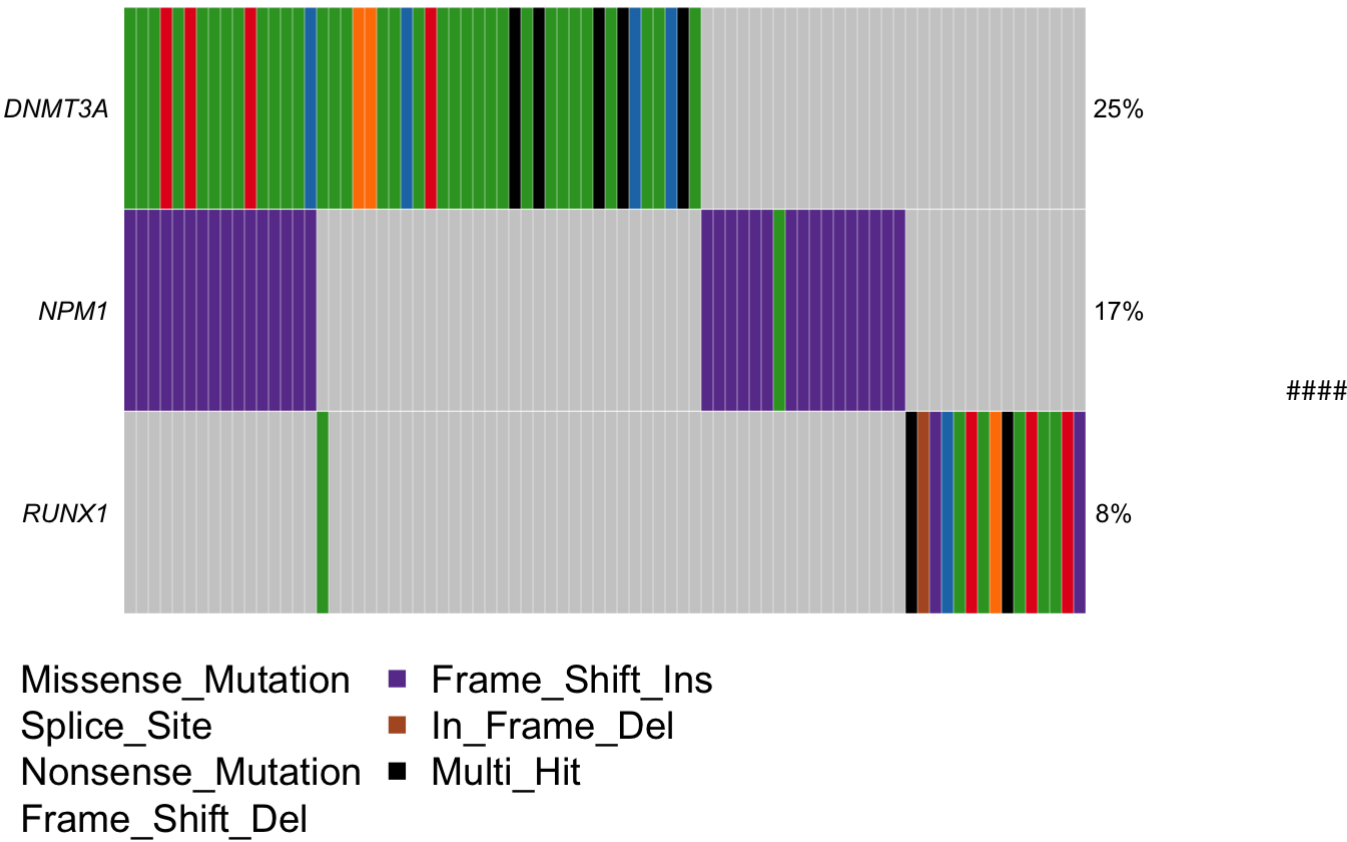

This is oncoplots, also known as waterfall plots.

```
#oncoplot for top ten mutated genes.
oncoplot(maf = laml, top = 10)
```

Altered in 141 (73.06%) of 193 samples.

- ■ In_Frame_Ins   ■ Frame_Shift_Ins
- ■ Frame_Shift_Del   ■ Nonsense_Mutation
- ■ Missense_Mutation   ■ Splice_Site
- ■ In_Frame_Del   ■ Multi_Hit

Mutations in each sample

Altered in 80 (41.45%) of 193 samples.

DNMT3A   25%

NPM1   17%

         ####

RUNX1   8%

- ■ Missense_Mutation
- ■ Splice_Site
- ■ Nonsense_Mutation
- ■ Frame_Shift_Del
- ■ Frame_Shift_Ins
- ■ In_Frame_Del
- ■ Multi_Hit

Classifies SNPs into Transitions and Transversions and returns a list of summarized tables in various ways

boxplot showing overall distribution of six different conversions and as a stacked barplot showing fraction of conversions in each sample.
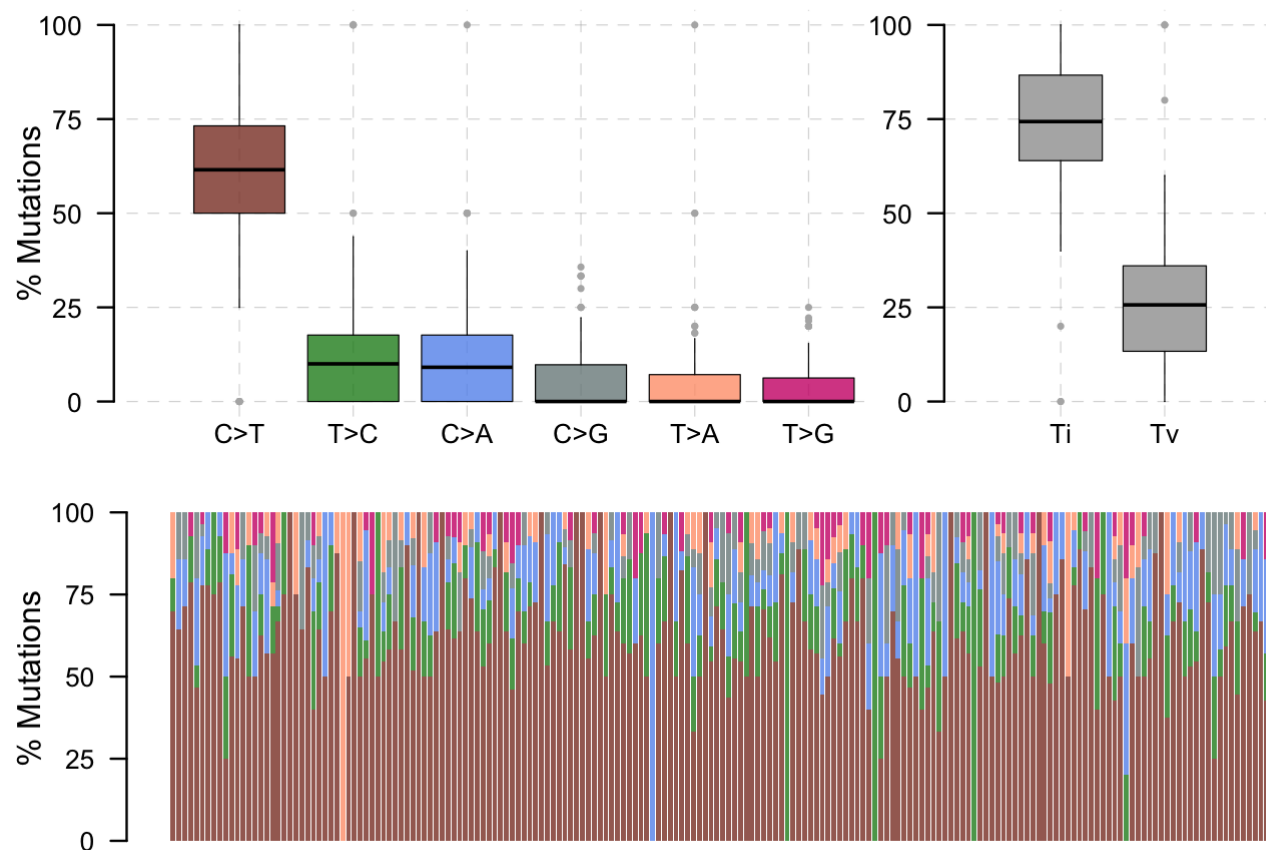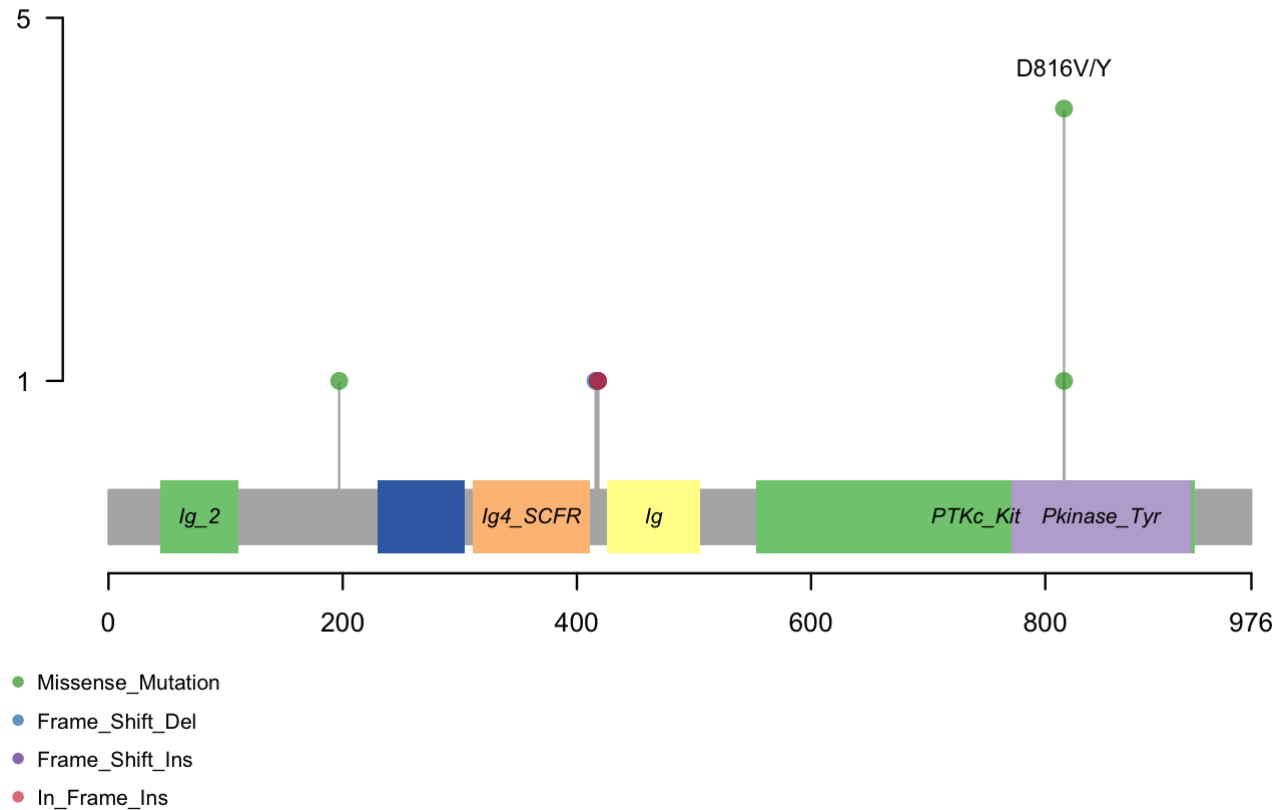
Lollipop plots are simple and most effective way showing mutation spots on protein structure. Many oncogenes have a preferential sites which are mutated more often than any other locus. These spots are considered to be mutational hot-spots and lollipop plots can be used to display them along with rest of the mutations.

```
##        HGNC refseq.ID protein.ID aa.length
## 1: DNMT3A NM_175629  NP_783328        912
## 2: DNMT3A NM_022552  NP_072046        912
## 3: DNMT3A NM_153759  NP_715640        723
```

## *DNMT3A* : [Somatic Mutation Rate: 24.87%]
NM_175629

- Frame_Shift_Del
- Nonsense_Mutation
- Missense_Mutation
- Splice_Site
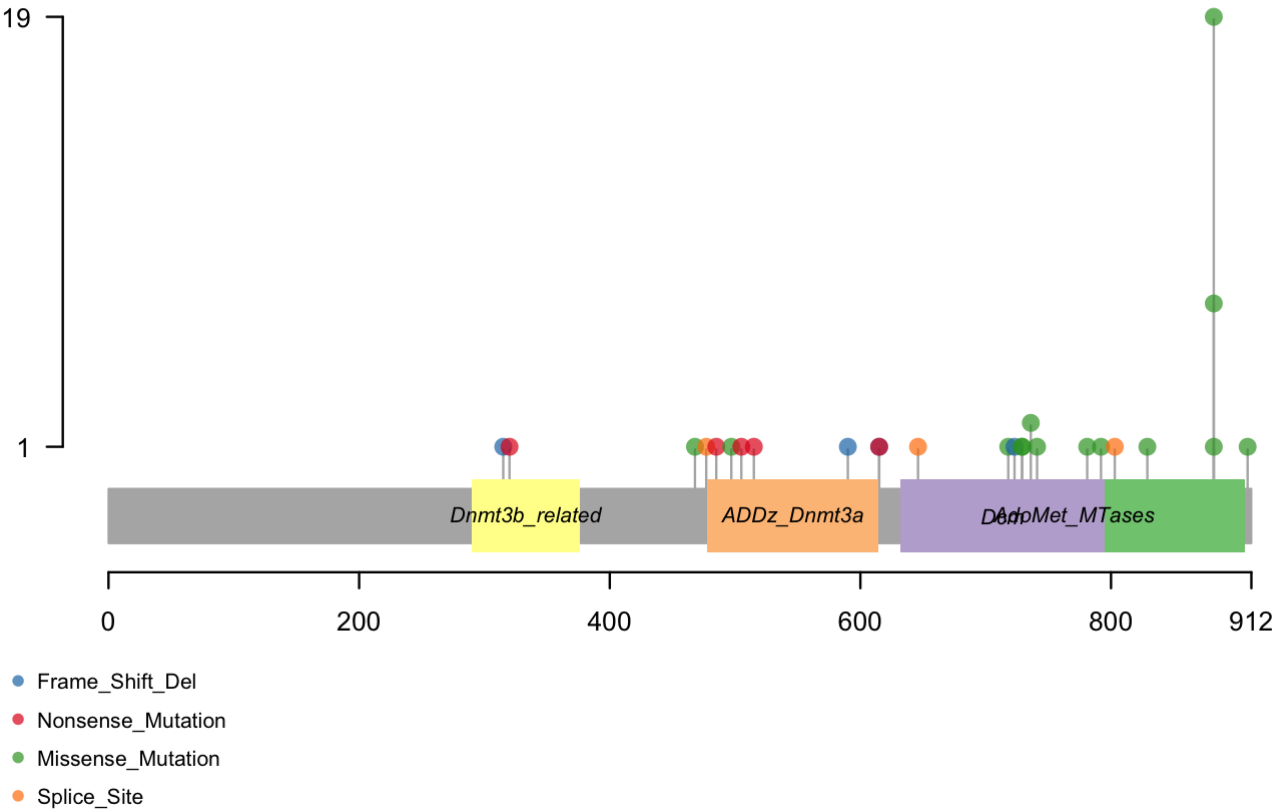
KIT : [Somatic Mutation Rate: 4.15%]
NM_000222

- Missense_Mutation
- Frame_Shift_Del
- Frame_Shift_Ins
- In_Frame_Ins

Cancer genomes, especially solid tumors are characterized by genomic loci with localized hyper-mutations 5. Such hyper mutated genomic regions can be visualized by plotting inter variant distance on a linear genomic scale. These plots generally called rainfall plots

```
##     Chromosome Start_Position End_Position nMuts Avg_intermutation_dist
## 1:          8       98129391      98133560     6                833.8000
## 2:          8       98398603      98403536     8                704.7143
## 3:          8       98453111      98456466     8                479.2857
## 4:          8      124090506     124096810    21                315.2000
## 5:         12       97437934      97439705     6                354.2000
## 6:         17       29332130      29336153     7                670.5000
##     Size Tumor_Sample_Barcode C>G C>T
## 1: 4169          TCGA-A8-A08B   4   2
## 2: 4933          TCGA-A8-A08B   1   7
## 3: 3355          TCGA-A8-A08B  NA   8
## 4: 6304          TCGA-A8-A08B   1  20
## 5: 1771          TCGA-A8-A08B   3   3
## 6: 4023          TCGA-A8-A08B   4   3
```
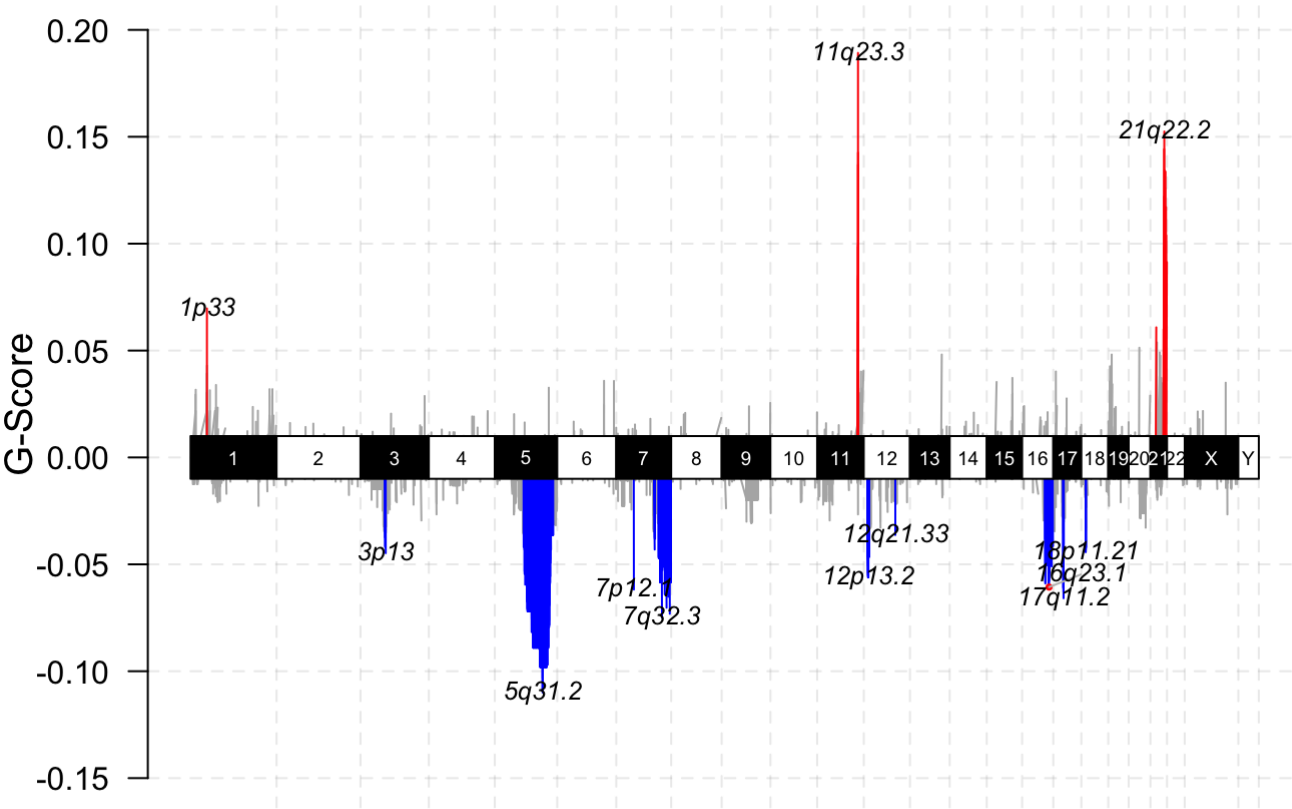

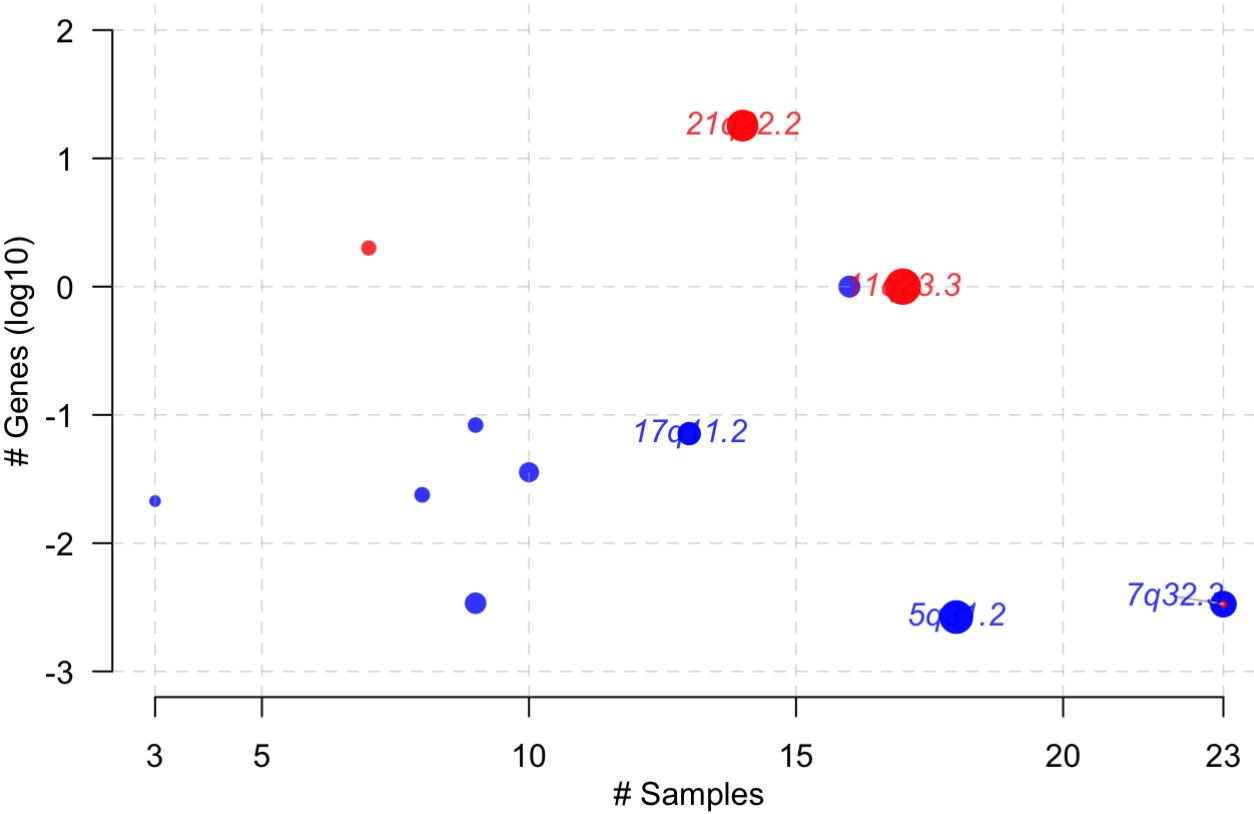TCGA-A8-A08B
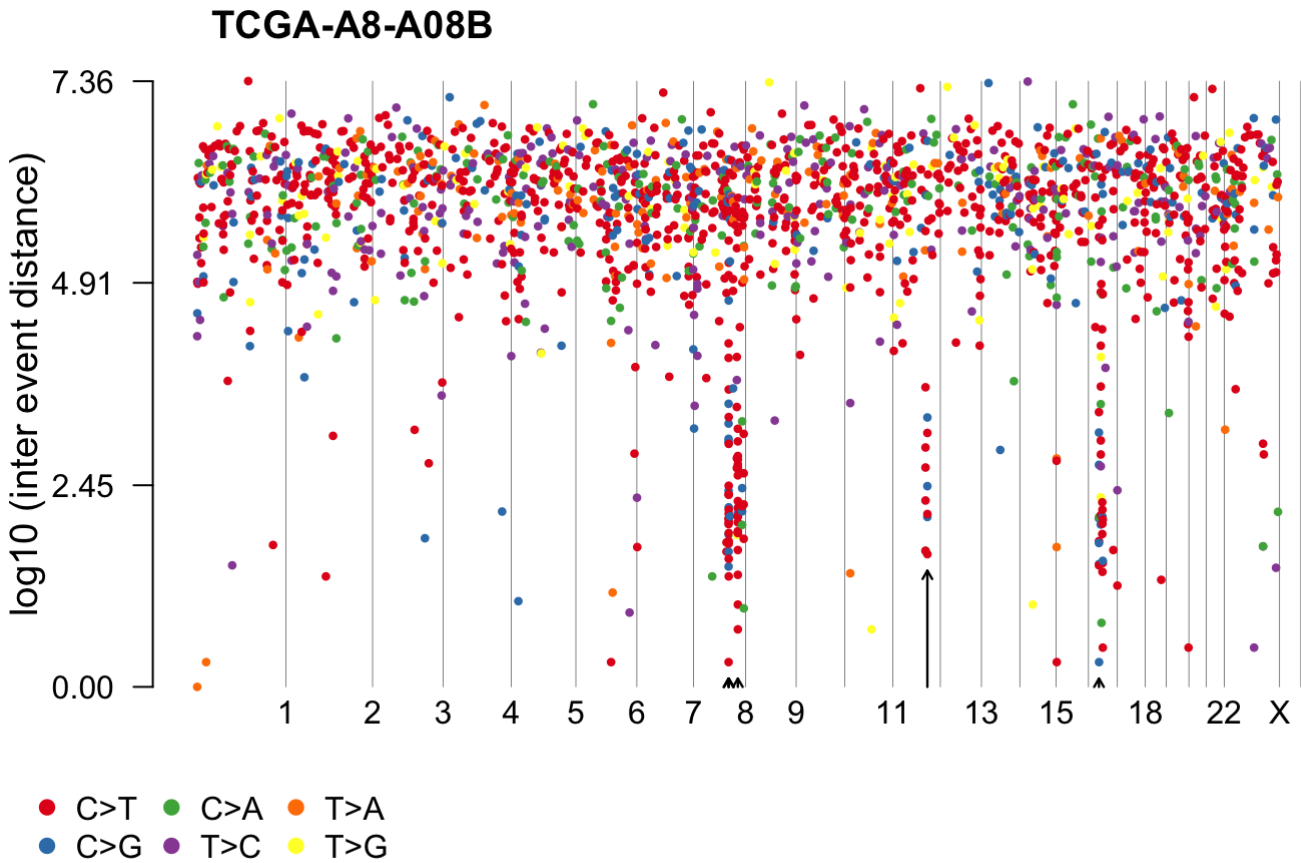
Genome plots

```
## An object of class  GISTIC
##            ID summary
## 1:    Samples     191
## 2:     nGenes    2622
## 3: cytoBands      16
## 4:       Amp     388
## 5:       Del   26481
## 6:     total   26869
```


Bubble plot

Oncoplot sorted according to FAB classification

TCGA-AB-3009
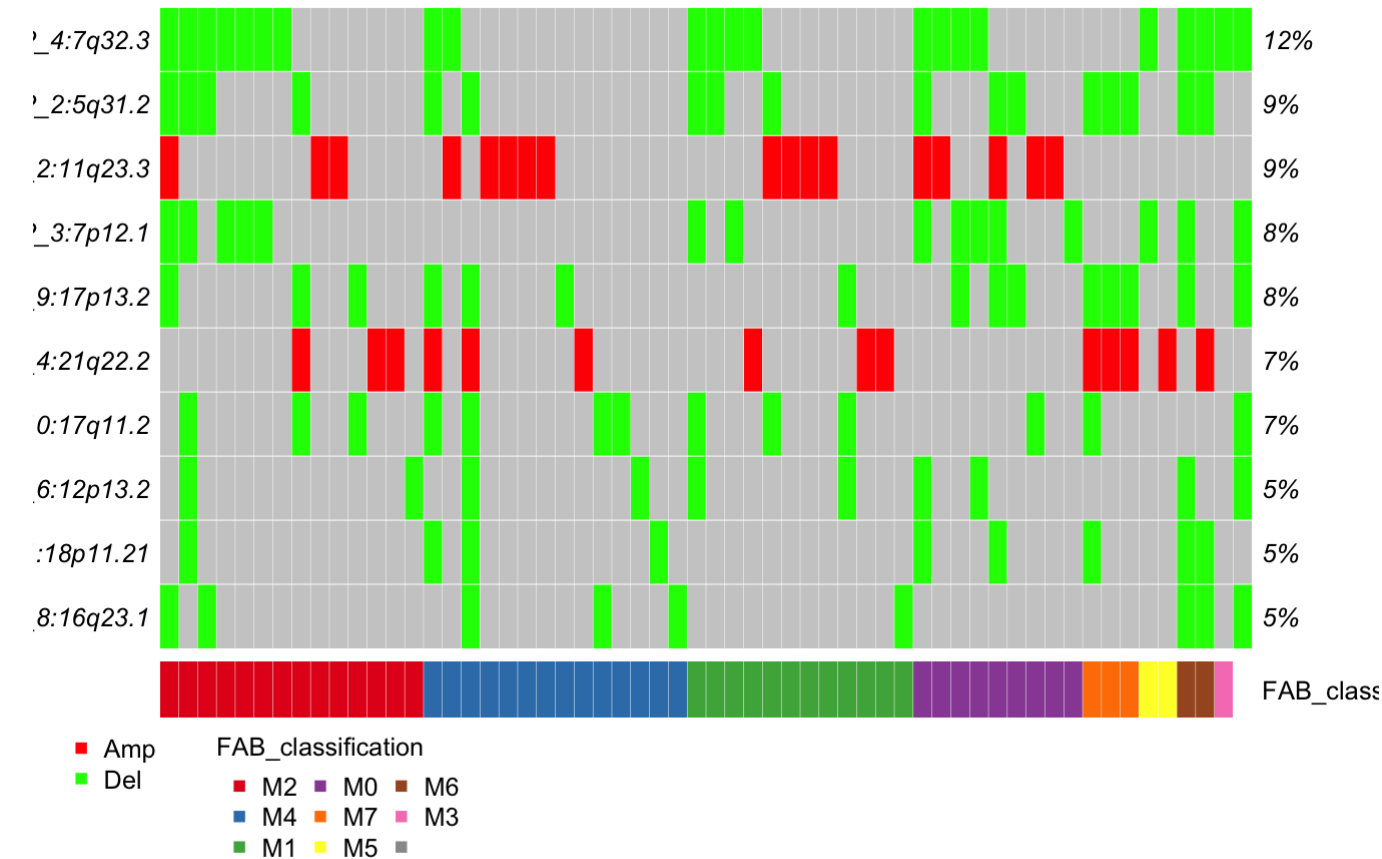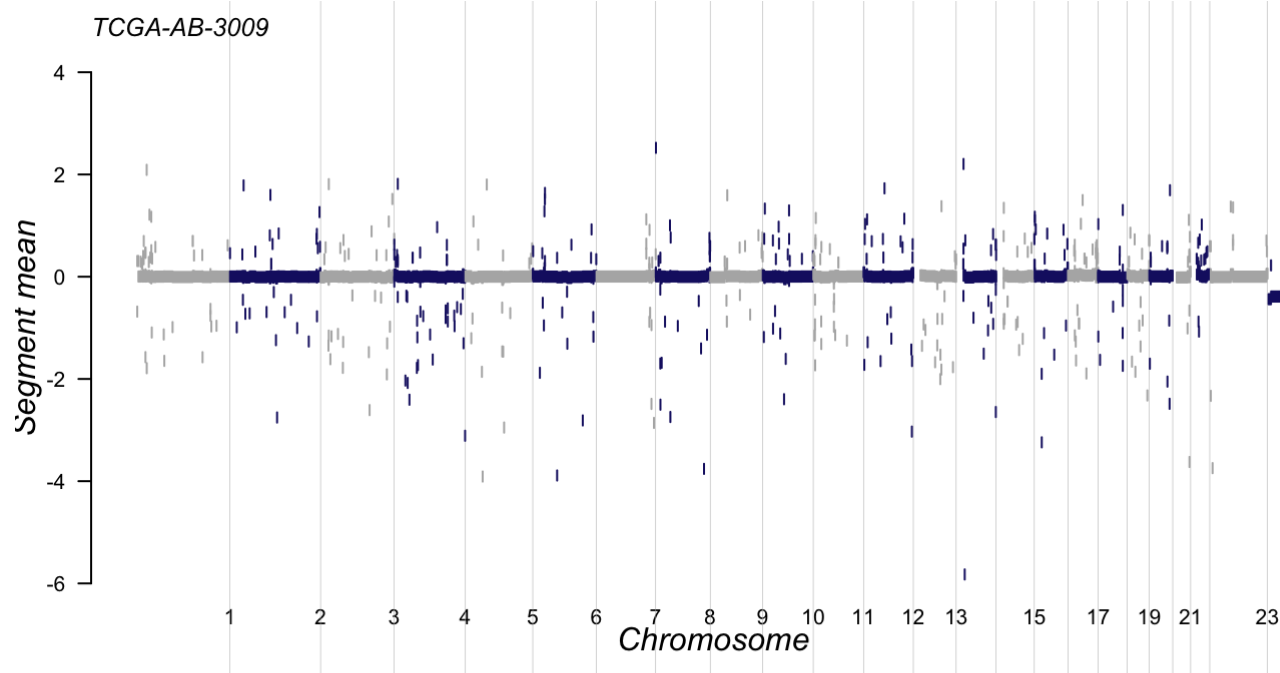
Many disease causing genes in cancer are co-occurring or show strong exclusiveness in their mutation pattern. Such mutually exclusive or co-occurring set of genes were detected by pair-wise Fisher's Exact test
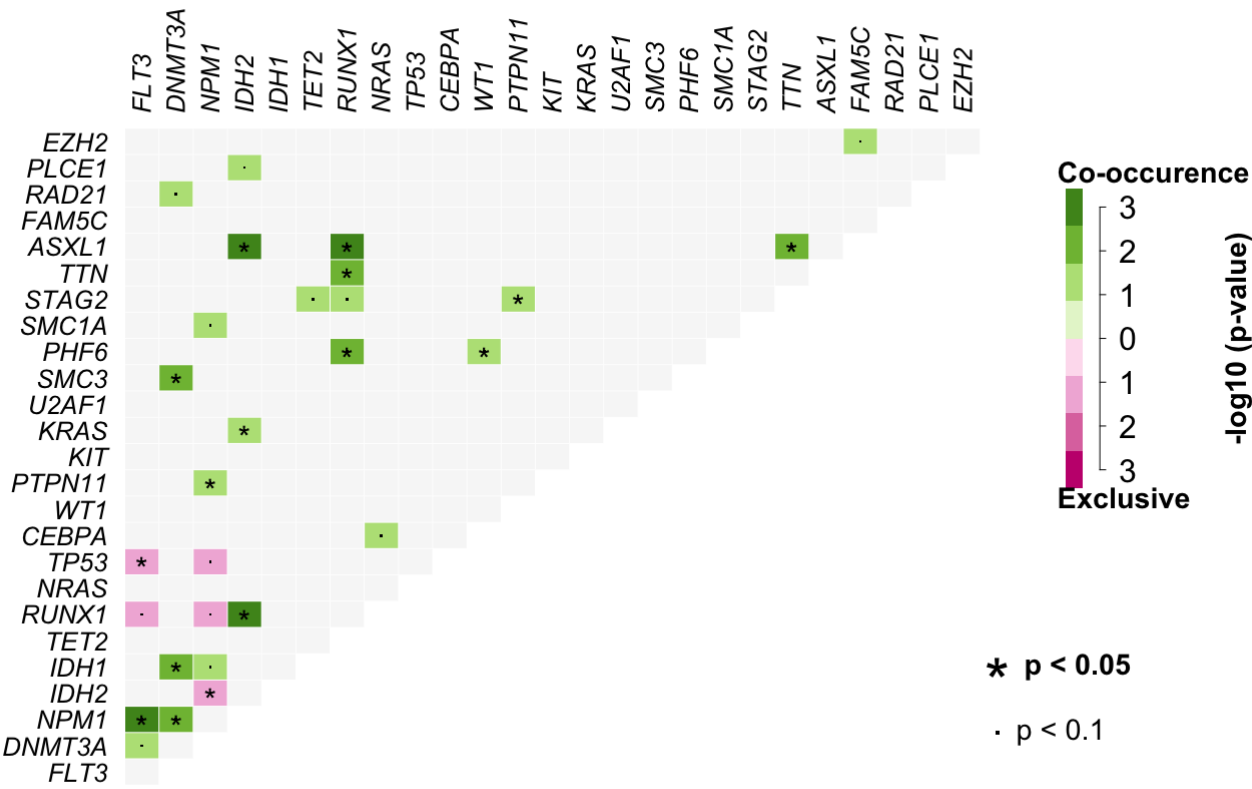
```
## Checking for Gene sets
## -----------------
## genes: 5
## geneset size: 3
## 10 combinations
```

```
## $pairs
##      gene1   gene2        pValue oddsRatio  00 11 01 10               Event
##  1:  ASXL1   RUNX1 0.0001541586 55.215541 176  4 12  1         Co_Occurence
##  2:   IDH2   RUNX1 0.0002809928  9.590877 164  7  9 13         Co_Occurence
##  3:   IDH2   ASXL1 0.0004030636 41.077327 172  4  1 16         Co_Occurence
##  4:   FLT3    NPM1 0.0009929836  3.763161 125 17 16 35         Co_Occurence
##  5:   SMC3   DNMT3A 0.0010451985 20.177713 144  6 42  1        Co_Occurence
##  6: DNMT3A    NPM1 0.0014582861  3.733141 128 16 17 32         Co_Occurence
##  7: DNMT3A    IDH1 0.0033807043  4.462201 137 10  8 38         Co_Occurence
##  8:  ASXL1     TTN 0.0077607658 28.459418 184  2  4  3         Co_Occurence
##  9:   PHF6   RUNX1 0.0081059811 12.967042 174  3 13  3         Co_Occurence
## 10:    TTN   RUNX1 0.0081059811 12.967042 174  3 13  3         Co_Occurence
## 11:   FLT3    TP53 0.0125113481  0.000000 126 NA 15 52 Mutually_Exclusive
## 12:  STAG2  PTPN11 0.0263964643 12.391225 180  2  7  4         Co_Occurence
## 13:   IDH2    NPM1 0.0277733049  0.000000 140 NA 33 20 Mutually_Exclusive
## 14:   IDH2    KRAS 0.0382620610  5.832674 168  3  5 17         Co_Occurence
## 15:    WT1    PHF6 0.0463612252  8.623360 177  2  4 10         Co_Occurence
## 16:   NPM1  PTPN11 0.0479288542  4.230142 155  4  5 29         Co_Occurence
## 17:   IDH2   PLCE1 0.0540565743  9.280043 171  2  2 18         Co_Occurence
## 18: DNMT3A   FLT3 0.0630630121  1.951476 111 18 34 30         Co_Occurence
## 19:   NPM1   SMC1A 0.0635083207  5.167266 157  3  3 30         Co_Occurence
## 20:  CEBPA    NRAS 0.0678045968  4.149259 168  3 12 10         Co_Occurence
## 21:  RUNX1    FLT3 0.0740850163  0.165692 126  1 51 15 Mutually_Exclusive
## 22:   EZH2   FAM5C 0.0761095136 21.701827 186  1  4  2         Co_Occurence
## 23:   TP53    NPM1 0.0785739379  0.000000 145 NA 33 15 Mutually_Exclusive
## 24:   NPM1   RUNX1 0.0787933378  0.000000 144 NA 16 33 Mutually_Exclusive
## 25:  RUNX1   STAG2 0.0795738073  6.066792 173  2  4 14         Co_Occurence
## 26:   TET2   STAG2 0.0888875052  5.638415 172  2  4 15         Co_Occurence
## 27:   IDH1    NPM1 0.0914621351  2.722302 148  6 27 12         Co_Occurence
## 28:  RAD21   DNMT3A 0.0993846041  4.719039 143  3 45  2        Co_Occurence
##      gene1   gene2        pValue oddsRatio  00 11 01 10               Event
##
## $gene_sets
## Empty data.table (0 rows and 2 cols): gene_set,pvalue
```
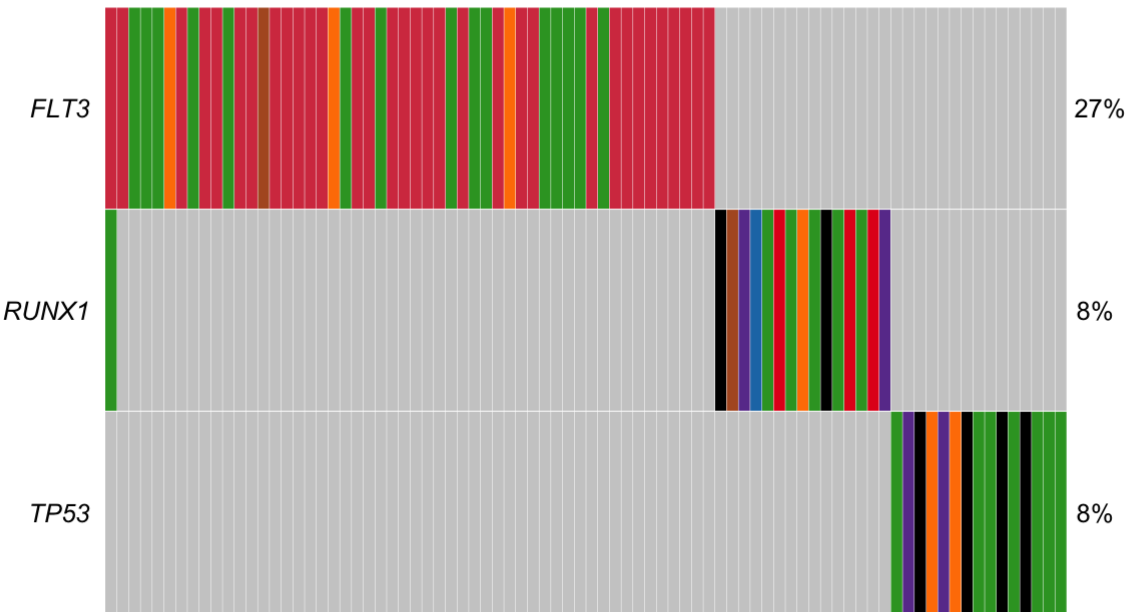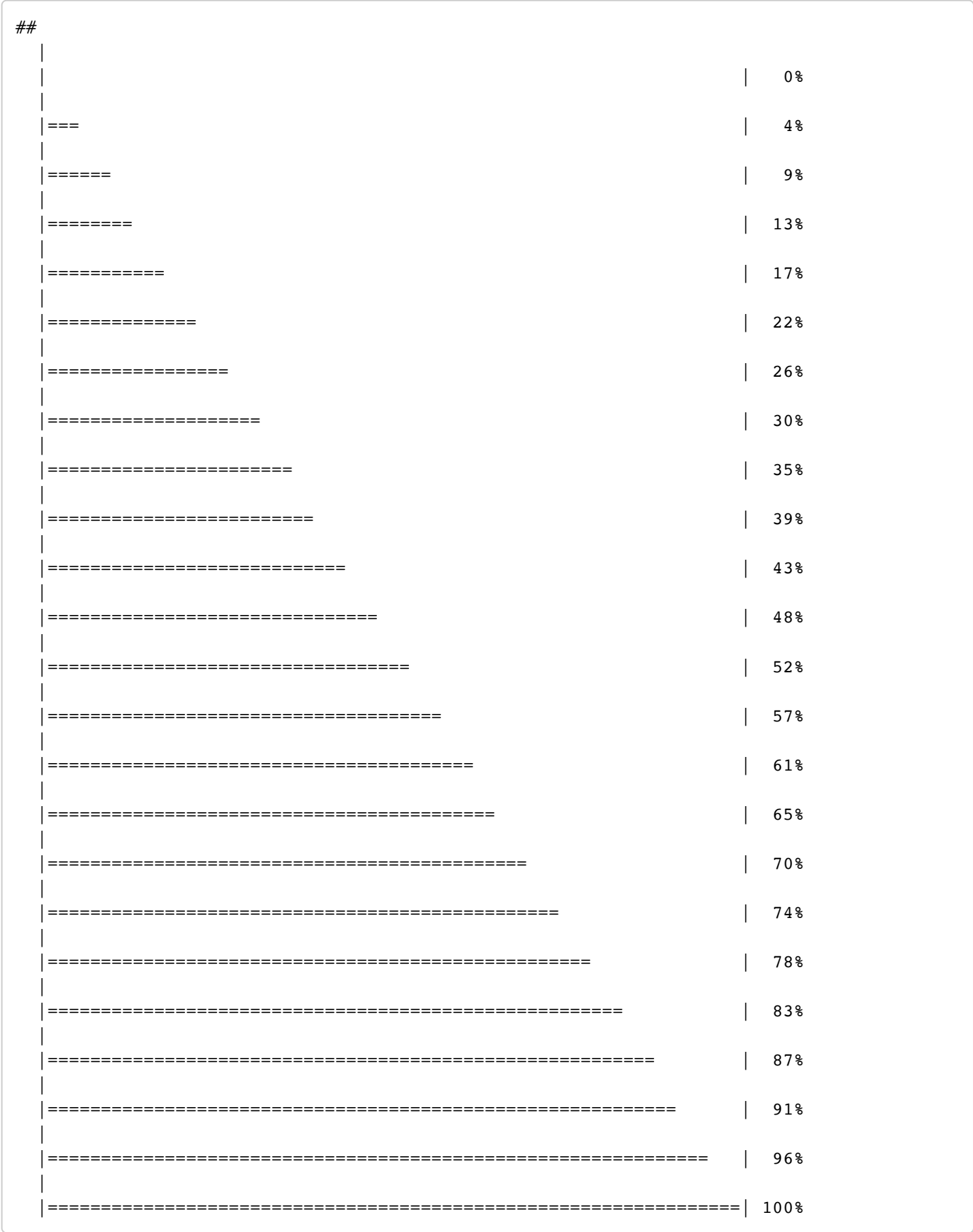
Altered in 82 (42.49%) of 193 samples.

Detecting cancer driver genes based on positional clustering

```
 ##
  |
  |                                                                          |    0%
  |
  |===                                                                       |    4%
  |
  |======                                                                    |    9%
  |
  |========                                                                  |   13%
  |
  |==========                                                                |   17%
  |
  |=============                                                             |   22%
  |
  |================                                                          |   26%
  |
  |===================                                                       |   30%
  |
  |=====================                                                     |   35%
  |
  |========================                                                  |   39%
  |
  |===========================                                               |   43%
  |
  |==============================                                            |   48%
  |
  |================================                                          |   52%
  |
  |===================================                                       |   57%
  |
  |======================================                                    |   61%
  |
  |=========================================                                 |   65%
  |
  |============================================                              |   70%
  |
  |===============================================                           |   74%
  |
  |==================================================                        |   78%
  |
  |====================================================                      |   83%
  |
  |=======================================================                   |   87%
  |
  |==========================================================                |   91%
  |
  |=============================================================             |   96%
  |
  |==============================================================|  100%
```

Hugo_Sym…	Frame_Shift_Del	Frame_Shift_Ins	In_Frame_Del	In_Frame_Ins	Missense_Muta
<fctr>	<int>	<int>	<int>	<int>	

Hugo_Sym...	Frame_Shift_Del	Frame_Shift_Ins	In_Frame_Del	In_Frame_Ins	Missense_Muta
<fctr>	<int>	<int>	<int>	<int>	
IDH1	0	0	0	0	
IDH2	0	0	0	0	
NPM1	0	33	0	0	
NRAS	0	0	0	0	
U2AF1	0	0	0	0	
KIT	1	1	0	1	

6 rows | 1-6 of 19 columns
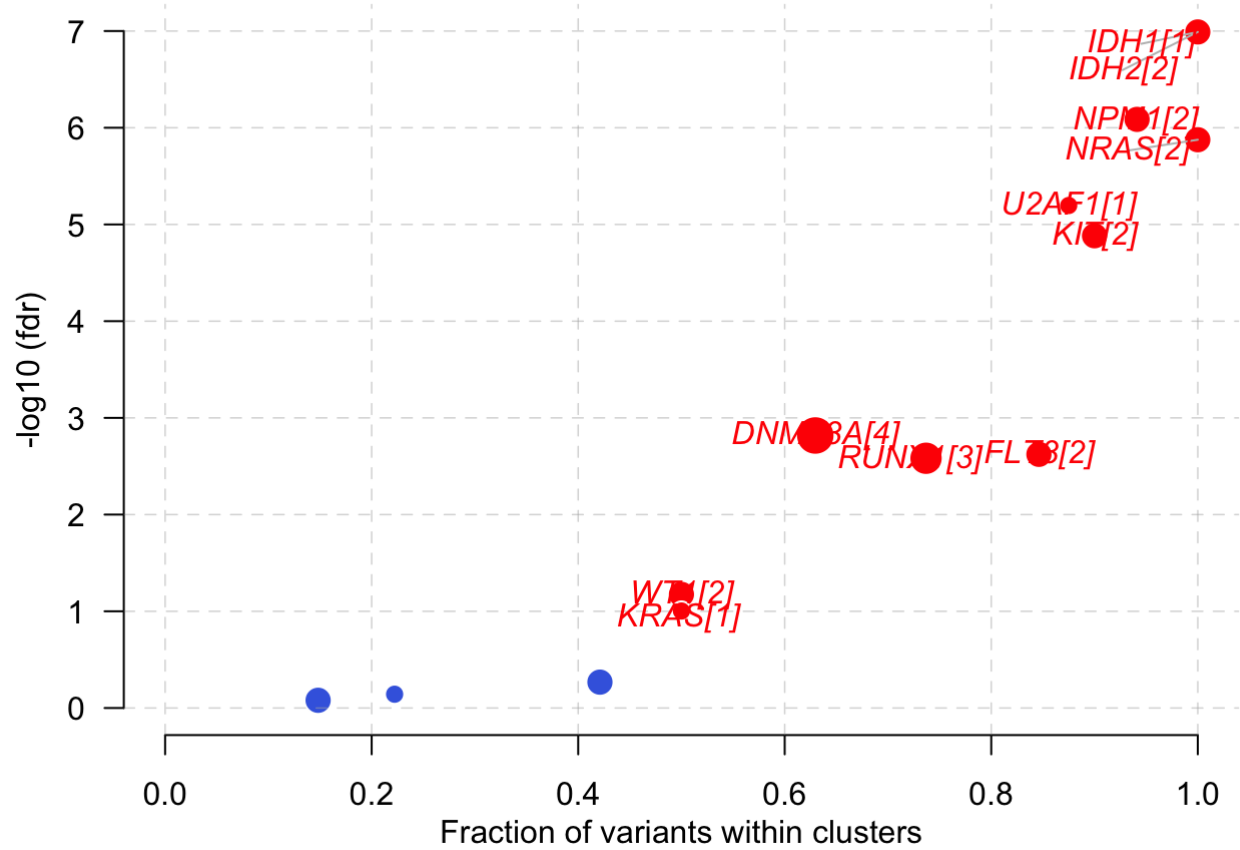
Adding and summarizing pfam domains

```
## Warning in pfamDomains(maf = laml, AACol = "Protein_Change", top = 10):
## Removed 50 mutations for which AA position was not available
```
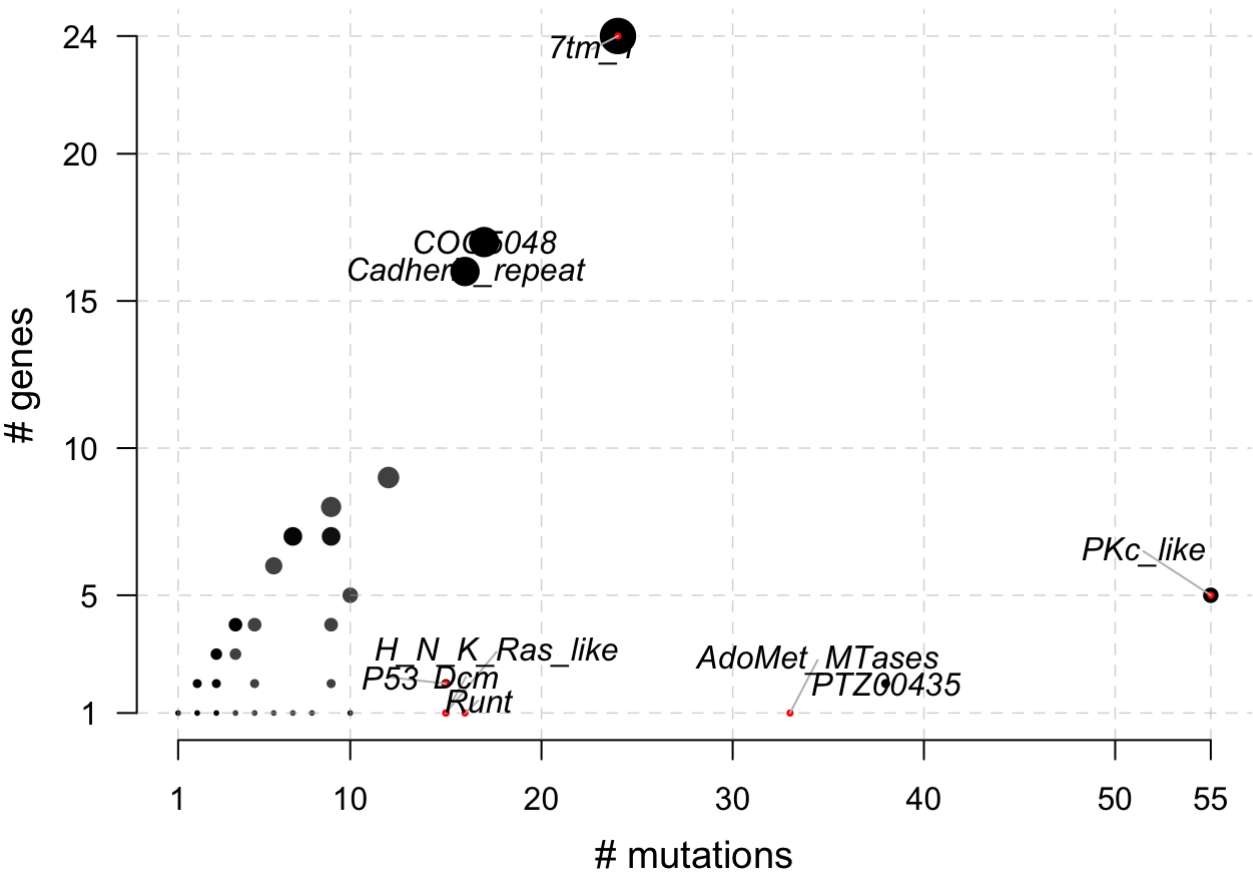
HGNC	AA…	Variant_Classification	N	total	fraction	DomainLabel
<chr>	<dbl>	<fctr>	<int>	<dbl>	<dbl>	<chr>
DNMT3A	882	Missense_Mutation	27	54	0.50000000	AdoMet_MTases
IDH1	132	Missense_Mutation	18	18	1.00000000	PTZ00435
IDH2	140	Missense_Mutation	17	20	0.85000000	PTZ00435
FLT3	835	Missense_Mutation	14	52	0.26923077	PKc_like
FLT3	599	In_Frame_Ins	10	52	0.19230769	PKc_like
U2AF1	34	Missense_Mutation	7	8	0.87500000	zf-CCCH
NRAS	61	Missense_Mutation	6	15	0.40000000	H_N_K_Ras_like
KIT	816	Missense_Mutation	5	10	0.50000000	PTKc_Kit
NRAS	13	Missense_Mutation	5	15	0.33333333	H_N_K_Ras_like
FLT3	601	In_Frame_Ins	4	52	0.07692308	PKc_like

1-10 of 1,516 rows   Previous **1** 2 3 4 5 6 … 152 Next

DomainLabel	nMuts	nGenes
<chr>	<int>	<int>

DomainLabel	nMuts	nGenes
<chr>	<int>	<int>
PKc_like	55	5
PTZ00435	38	2
AdoMet_MTases	33	1
7tm_1	24	24
COG5048	17	17
Cadherin_repeat	16	16
Runt	16	1
Dcm	15	1
H_N_K_Ras_like	15	1
P53	15	2

1-10 of 503 rows   Previous **1** 2 3 4 5 6 … 51 Next

Pan-Cancer comparison

```
##       gene pancan           q nMut log_q_pancan      log_q
##  1:  CEBPA  1.000 3.500301e-12   13   0.00000000 11.455895
##  2:   EZH2  1.000 7.463546e-05    3   0.00000000  4.127055
##  3: GIGYF2  1.000 6.378338e-03    2   0.00000000  2.195292
##  4:    KIT  0.509 1.137517e-05    8   0.29328222  4.944042
##  5:   PHF6  0.783 6.457555e-09    6   0.10623824  8.189932
##  6: PTPN11  0.286 7.664584e-03    9   0.54363397  2.115511
##  7:  RAD21  0.929 1.137517e-05    5   0.03198429  4.944042
##  8:  SMC1A  0.801 2.961696e-03    6   0.09636748  2.528460
##  9:   TET2  0.907 2.281625e-13   17   0.04239271 12.641756
## 10:    WT1  1.000 2.281625e-13   12   0.00000000 12.641756
```

LAML v/s Pan-cancer
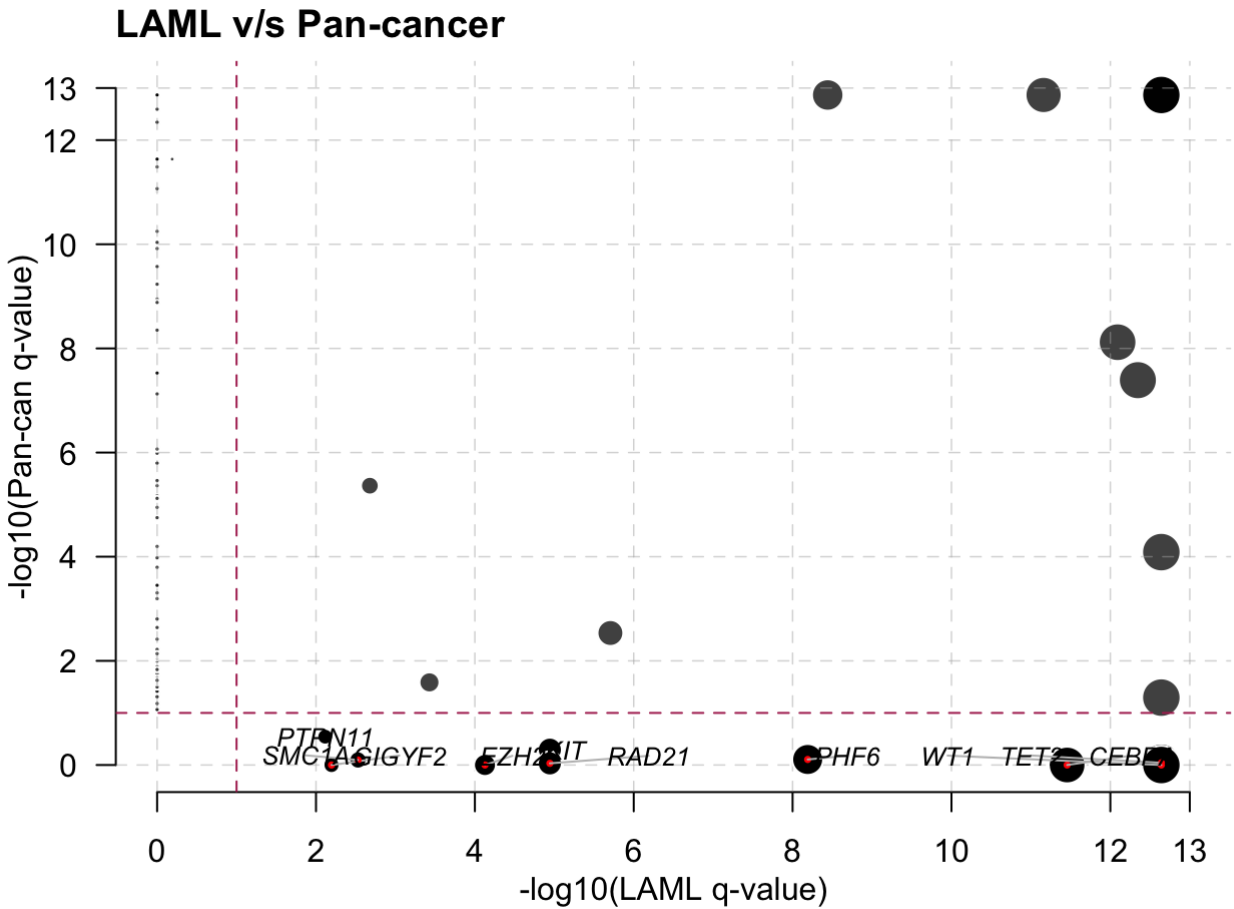
gene	pancan	q	nMut	log_q_pancan	log_q
<chr>	<dbl>	<dbl>	<int>	<dbl>	<dbl>
ACVR1B	6.11e-02	1.000000e+00	0	1.21395879	0.0000000
AKT1	2.68e-10	1.000000e+00	0	9.57186521	0.0000000
APC	1.36e-13	1.000000e+00	0	12.86646109	0.0000000
APOL2	7.96e-03	1.000000e+00	0	2.09908693	0.0000000
ARHGAP35	2.32e-12	1.000000e+00	1	11.63451202	0.0000000
ARID1A	2.32e-12	1.000000e+00	1	11.63451202	0.0000000
ARID2	1.03e-06	1.000000e+00	1	5.98716278	0.0000000
ASXL1	4.33e-06	2.098493e-03	5	5.36351210	2.6780925
ASXL2	8.18e-02	1.000000e+00	1	1.08724670	0.0000000
ATM	5.85e-10	1.000000e+00	0	9.23284413	0.0000000

1-10 of 124 rows   Previous **1** 2 3 4 5 6 … 13 Next

Survival analysis

```
## DNMT3A
##      48
##     Group medianTime   N
## 1: Mutant        243  48
## 2:     WT        366 145
```
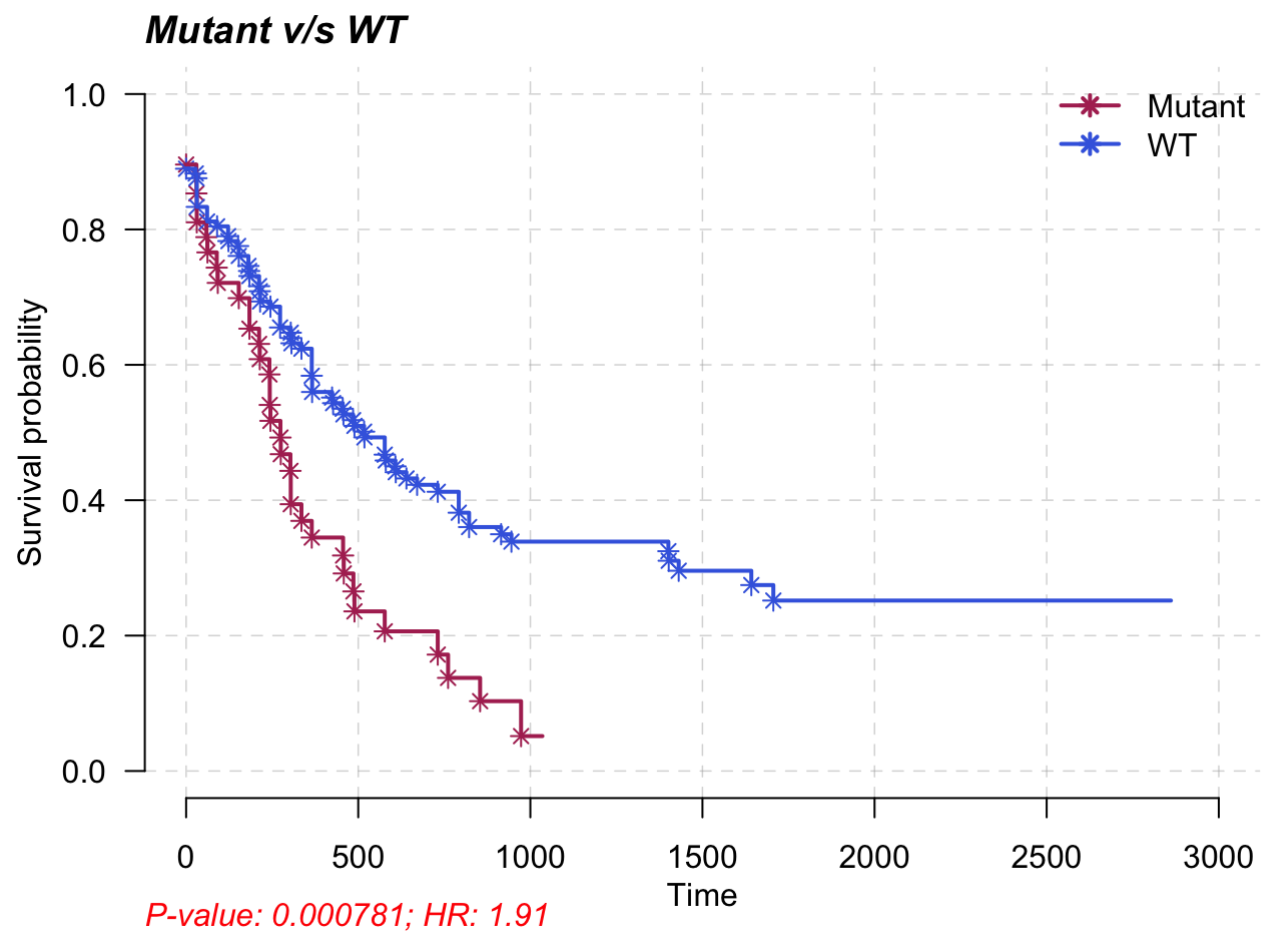
Mutant v/s WT

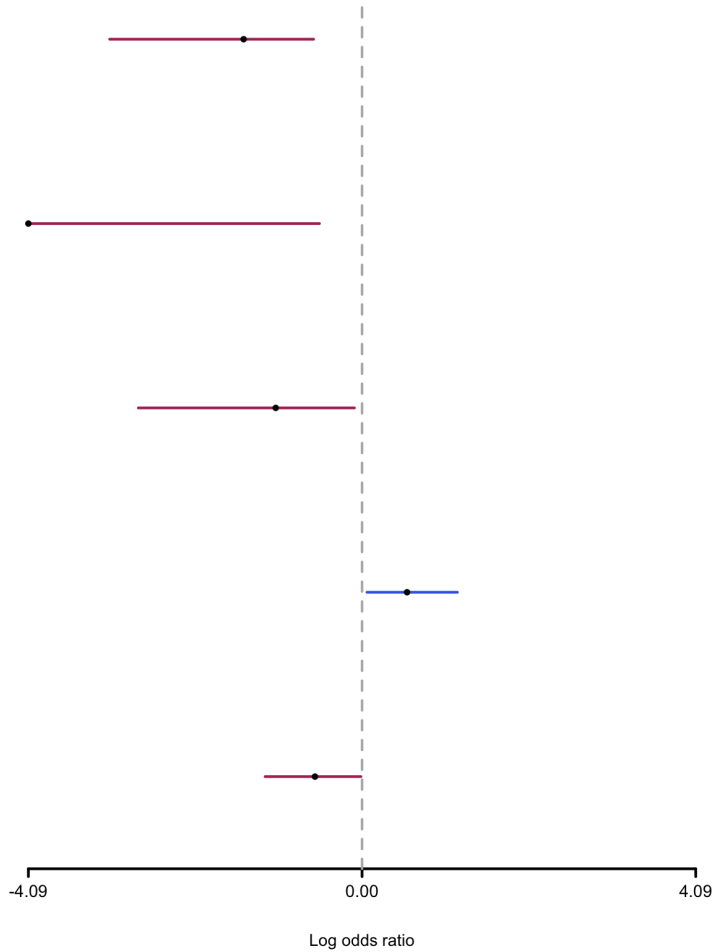
P-value: 0.000781; HR: 1.91

```
## -Reading
## -Validating
## --Non MAF specific values in Variant_Classification column:
##    ITD
## -Silent variants: 45
## -Summarizing
## -Processing clinical data
## --Missing clinical data
## -Finished in 0.085s elapsed (0.153s cpu)
```

```
## –Reading
## –Validating
## --Non MAF specific values in Variant_Classification column:
##    ITD
## –Silent variants: 19
## –Summarizing
## –Processing clinical data
## --Missing clinical data
## –Finished in 0.073s elapsed (0.127s cpu)
```

```
## $results
##     Hugo_Symbol Primary Relapse         pval          or        ci.up
## 1:          PML       1      11 1.529935e-05 0.03537381    0.2552937
## 2:         RARA       0       7 2.574810e-04 0.00000000    0.3006159
## 3:        RUNX1       1       5 1.310500e-02 0.08740567    0.8076265
## 4:         FLT3      26       4 1.812779e-02 3.56086275   14.7701728
## 5:       ARID1B       5       8 2.758396e-02 0.26480490    0.9698686
## 6:          WT1      20      14 2.229087e-01 0.60619329    1.4223101
## 7:         KRAS       6       1 4.334067e-01 2.88486293  135.5393108
## 8:         NRAS      15       4 4.353567e-01 1.85209500    8.0373994
## 9:       ARID1A       7       4 7.457274e-01 0.80869223    3.9297309
##         ci.low       adjPval
## 1: 0.000806034 0.0001376942
## 2: 0.000000000 0.0011586643
## 3: 0.001813280 0.0393149868
## 4: 1.149009169 0.0407875250
## 5: 0.064804160 0.0496511201
## 6: 0.263440988 0.3343630535
## 7: 0.337679367 0.4897762916
## 8: 0.553883512 0.4897762916
## 9: 0.195710173 0.7457273717
##
## $SampleSummary
##     Cohort SampleSize
## 1: Primary        124
## 2: Relapse         58
```

Relapse (n = 58) v/s Primary (n = 124)

	Primary	Relapse	p-value
PML	1	11	***
RARA	0	7	***
RUNX1	1	5	*
FLT3	26	4	*
ARID1B	5	8	*

-4.09   0.00   4.09

Log odds ratio

Primary   Relapse

PrimaryAPL (N = 124)   **RelapseAPL (N = 58)**

1%	*PML*	19%
0%	*RARA*	12%
1%	*RUNX1*	9%
4%	*ARID1B*	14%

21%

FLT3

7%

- ■ Missense_Mutation
- ■ In_Frame_Del
- ■ Frame_Shift_Ins
- ■ Frame_Shift_Del
- ■ Multi_Hit
- ■ Nonsense_Mutation

```
##      HGNC refseq.ID protein.ID aa.length
## 1:   PML NM_033238  NP_150241        882
## 2:   PML NM_002675  NP_002666        633
## 3:   PML NM_033249  NP_150252        585
## 4:   PML NM_033247  NP_150250        435
## 5:   PML NM_033239  NP_150242        829
## 6:   PML NM_033250  NP_150253        781
## 7:   PML NM_033240  NP_150243        611
## 8:   PML NM_033244  NP_150247        560
## 9:   PML NM_033246  NP_150249        423
##      HGNC refseq.ID protein.ID aa.length
## 1:   PML NM_033238  NP_150241        882
## 2:   PML NM_002675  NP_002666        633
## 3:   PML NM_033249  NP_150252        585
## 4:   PML NM_033247  NP_150250        435
## 5:   PML NM_033239  NP_150242        829
## 6:   PML NM_033250  NP_150253        781
## 7:   PML NM_033240  NP_150243        611
## 8:   PML NM_033244  NP_150247        560
## 9:   PML NM_033246  NP_150249        423
```
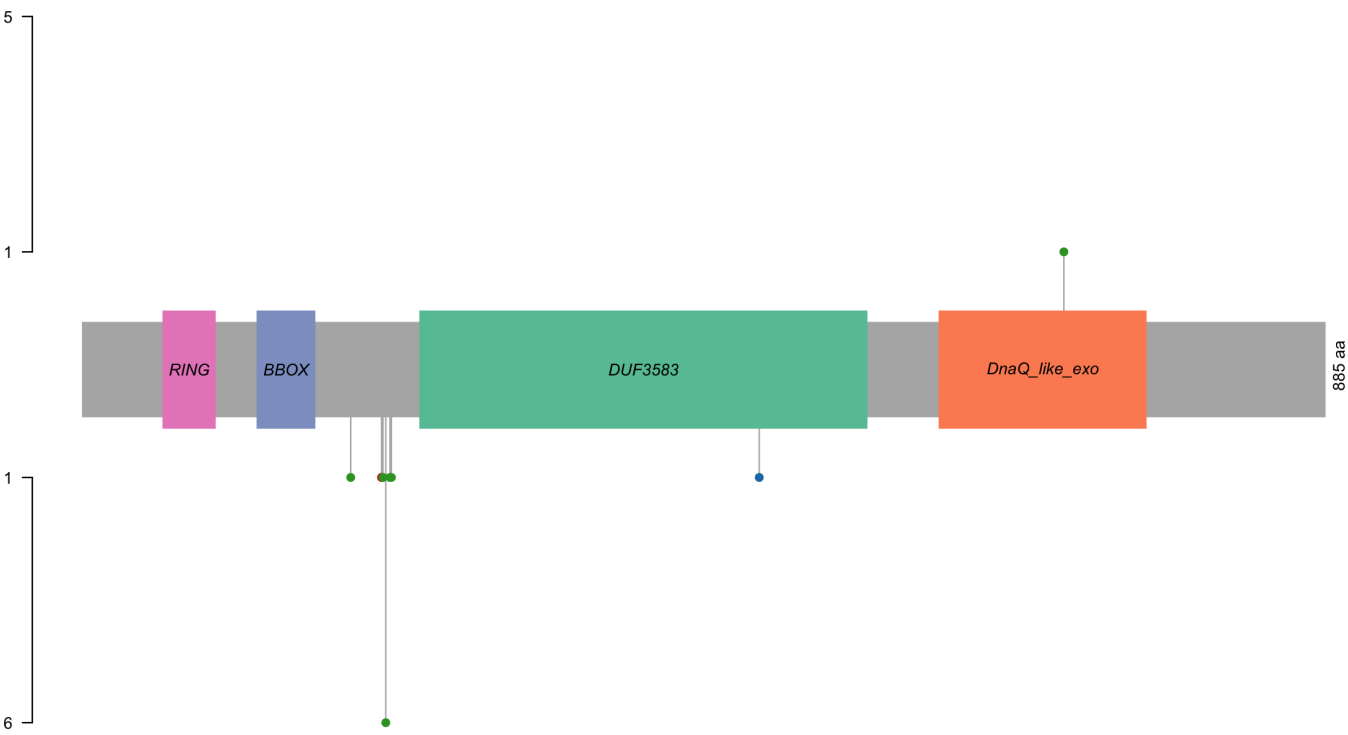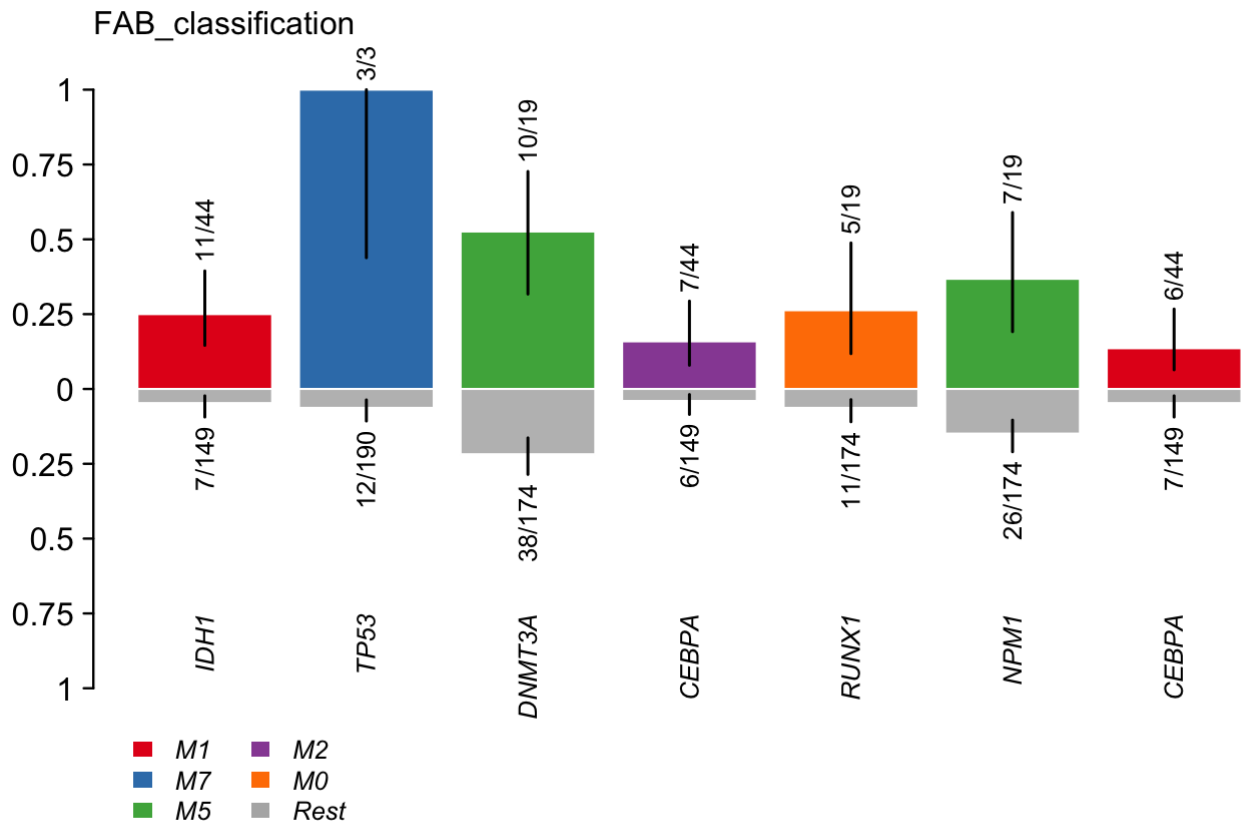
Primary [0.81%; N = 124]   PML: NM_033238

Relapse [18.97%; N = 58]

- Missense_Mutation
- Nonsense_Mutation
- Frame_Shift_Del

```
##
## M0 M1 M2 M3 M4 M5 M6 M7
## 19 44 44 21 39 19  3  3
```

Hugo_Sym...	Gro...	Gro...	n_mutated_group1	n_mutated_group2	p_value	OR_...	OR_
<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<
IDH1	M1	Rest	11 of 44	7 of 149	0.0002597371	0	0.392
TP53	M7	Rest	3 of 3	12 of 190	0.0003857187	0	0.131
DNMT3A	M5	Rest	10 of 19	38 of 174	0.0057610493	0	0.640
CEBPA	M2	Rest	7 of 44	6 of 149	0.0117352110	0	0.687
RUNX1	M0	Rest	5 of 19	11 of 174	0.0117436825	0	0.646
NPM1	M5	Rest	7 of 19	26 of 174	0.0248582372	0	0.834
CEBPA	M1	Rest	6 of 44	7 of 149	0.0478737468	0	0.986

7 rows

FAB_classification

Legend:
- M1 (red)
- M7 (blue)
- M5 (green)
- M2 (purple)
- M0 (orange)
- Rest (gray)

Bar values:
- IDH1: 11/44, 7/149
- TP53: 3/3, 12/190
- DNMT3A: 10/19, 38/174
- CEBPA: 7/44, 6/149
- RUNX1: 5/19, 11/174
- NPM1: 7/19, 26/174
- CEBPA: 6/44, 7/149

Druggable categories

RNA DIRECTED DNA POLYMERASE [TP53]
PROTEIN PHOSPHATASE [PTPN11]
PHOSPHOLIPASE [KIT]
EXTERNAL SIDE OF PLASMA MEMBRANE [KIT]
CELL SURFACE [KIT]
TRANSCRIPTION FACTOR COMPLEX [CEBPA,TP53]
LIPID KINASE [FLT3,KIT]
HISTONE MODIFICATION [TET2,TP53]
DRUG RESISTANCE [DNMT3A,TP53]
TRANSCRIPTION FACTOR BINDING [CEBPA,NPM1,TP53]
TYROSINE KINASE [FLT3,KIT,NPM1,TTN]
DNA REPAIR [NPM1,SMC1A,SMC3,TP53]
SERINE THREONINE KINASE [CEBPA,FLT3,KIT,KRAS,PTPN11]
TUMOR SUPPRESSOR [CEBPA,NPM1,PTPN11,SMC1A,SMC3]
KINASE [CEBPA,FLT3,KIT,KRAS,NPM1]
DRUGGABLE GENOME [DNMT3A,FLT3,KIT,NPM1,PTPN11]
CLINICALLY ACTIONABLE [CEBPA,DNMT3A,FLT3,IDH1,IDH2]

0   # Genes   16

```
## Number of claimed drugs for given genes:
##       Gene N
## 1: DNMT3A 7
```

Gene <chr>	interaction_types <chr>	drug_name <chr>	drug_claim_name <chr>
DNMT3A			N/A
DNMT3A		DAUNORUBICIN	Daunorubicin
DNMT3A		DECITABINE	Decitabine
DNMT3A		IDARUBICIN	IDARUBICIN
DNMT3A		DECITABINE	DECITABINE
DNMT3A	inhibitor	DECITABINE	CHEMBL1201129
DNMT3A	inhibitor	AZACITIDINE	CHEMBL1489

7 rows

```
##          Pathway  N n_affected_genes fraction_affected
##  1:      RTK-RAS 85               18        0.21176471
##  2:        Hippo 38                7        0.18421053
##  3:        NOTCH 71                6        0.08450704
##  4:          MYC 13                3        0.23076923
##  5:          WNT 68                3        0.04411765
##  6:         TP53  6                2        0.33333333
##  7:         NRF2  3                1        0.33333333
##  8:         PI3K 29                1        0.03448276
##  9: Cell_Cycle 15                0        0.00000000
## 10:     TGF-Beta  7                0        0.00000000
```

RTK-RAS pathway