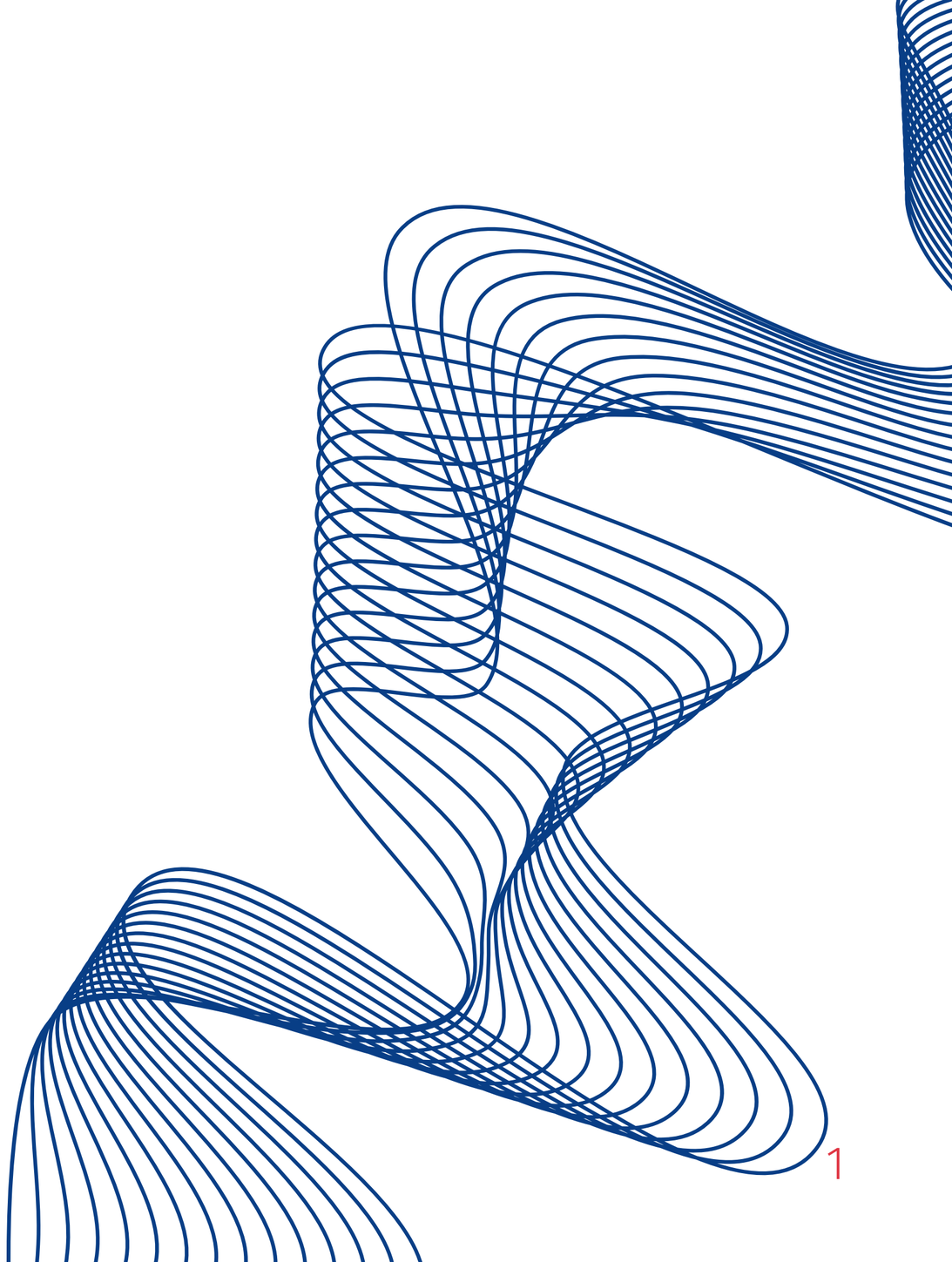
Search for mutations in transcription data
Allele specific expression

Marfa Zakirova
Anastasiya Yudina



Detection of transcription data

Given:

- reference data
- expression data
- two maf files

Determine:

- which file is WES which is RNAseq

Classification problem

WES

- less noisy
- uniform data on most genes

RNAseq

- noise in the data
- data only on expressed genes

Both files will contain information on exons
It is impossible to navigate by file sizes

WES file is more like a reference data
RNA seq is more like expression

Classification problem

WES file is more like a reference data
RNA seq is more like expression

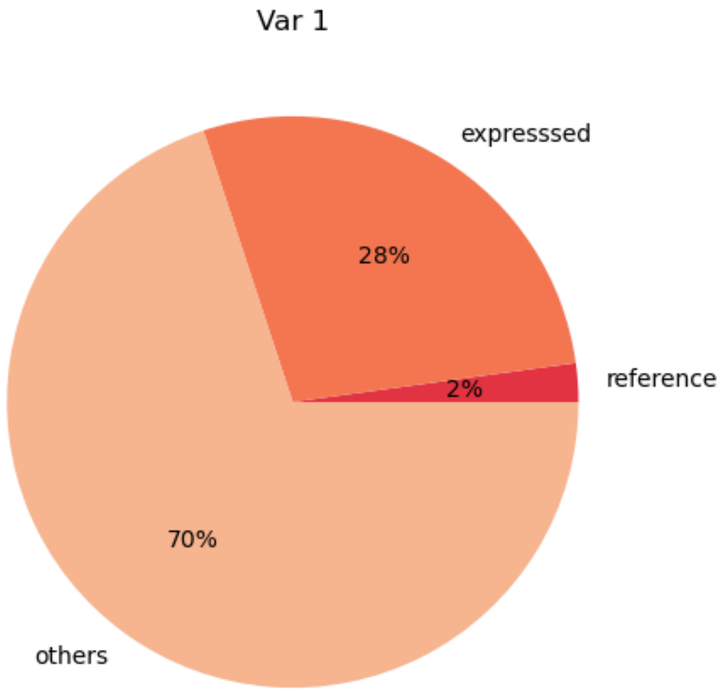
TP - is in the reference, is in the sample
FP - isn't in the reference, is in the sample
FN - in the reference, isn't in the sample

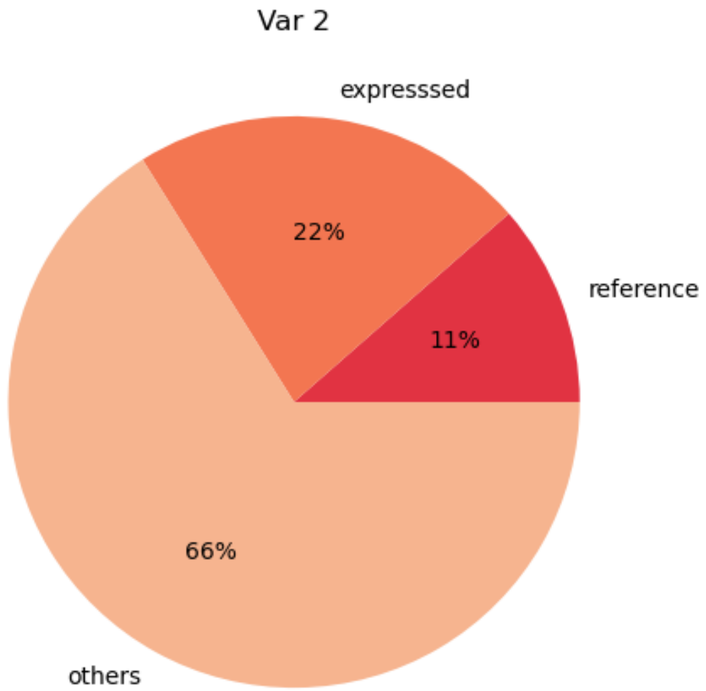
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_\beta = (\beta^2 + 1) \cdot \frac{Recall \cdot Precision}{Recall + \beta^2 \cdot Precision}$$

Classification problem





Var 1 - RNA-seq

Var 2 - WES

Mutations analysis

Given:

- 2 cell lines:

melanoma(COLO), breast cancer(HCC)

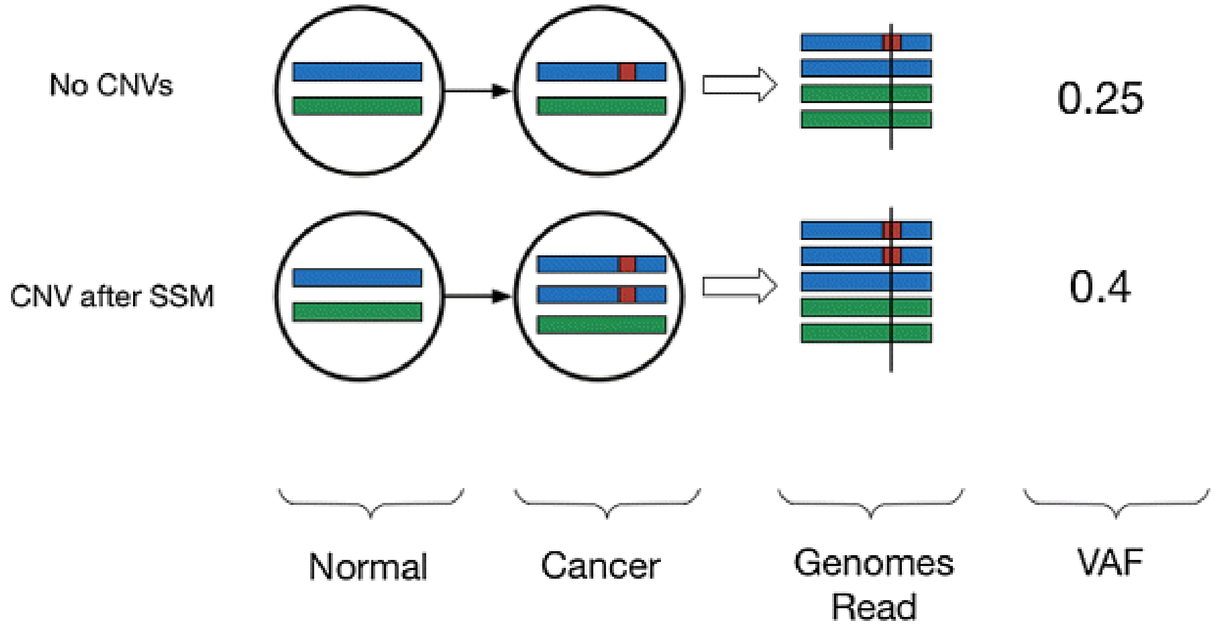
- WES, RNAseq maf files
- different purity

Determine:

- dynamics of mutation frequencies
- groups of allele-specifically

expressed mutations

VAF



$$\mathrm{VAF} = \frac{depth\_alt}{depth\_alt + depth\_norm}$$

Dynamics of mutation frequencies

- Consideration of polymorphism of ordinary cells
- Identification of the type of VAF mutations via purity dependence
- Group selection

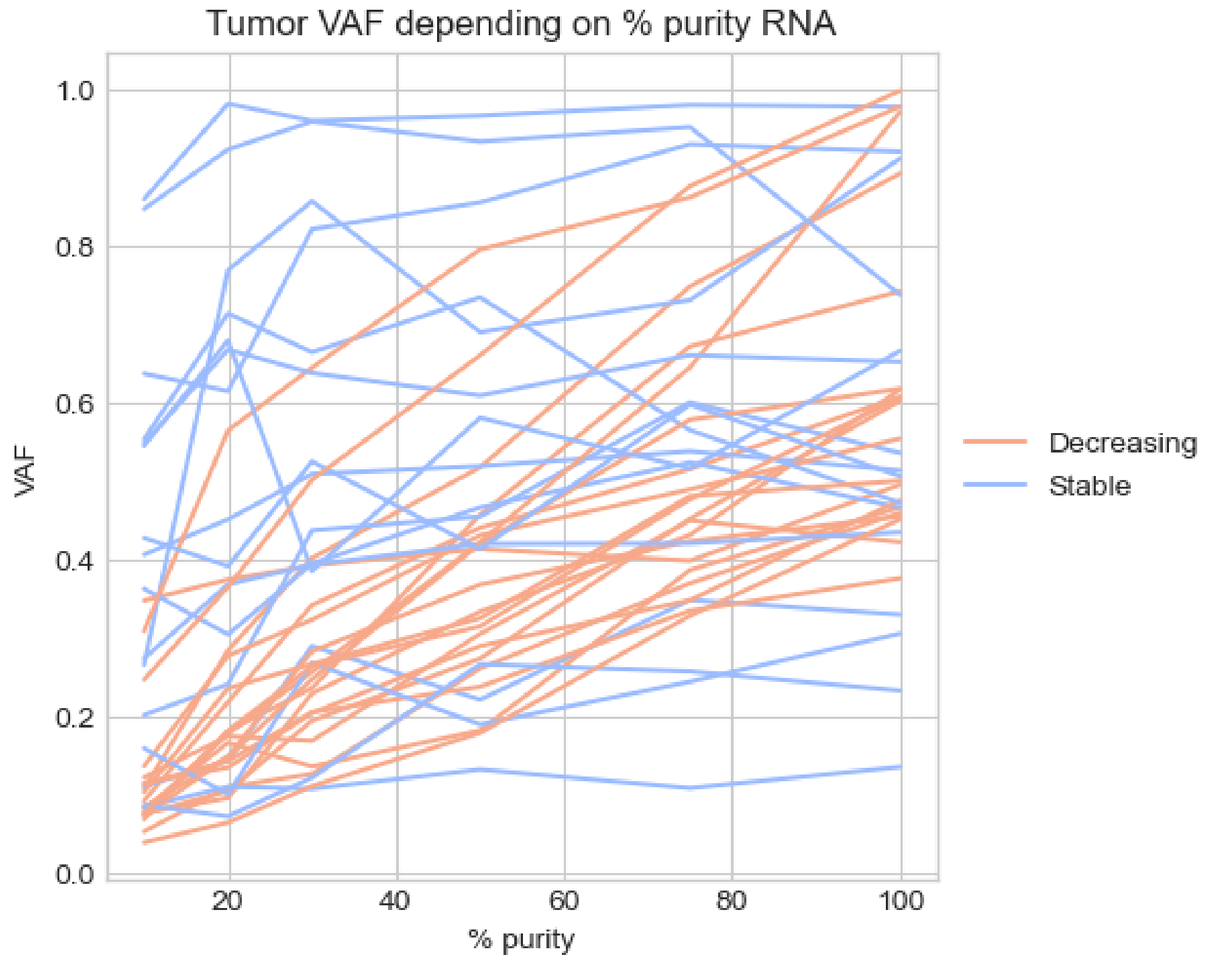
Purity: 10%, 20%, 30%, 50%, 75%, 100%

100% == 100% Tumor

COLO RNA VAF plot

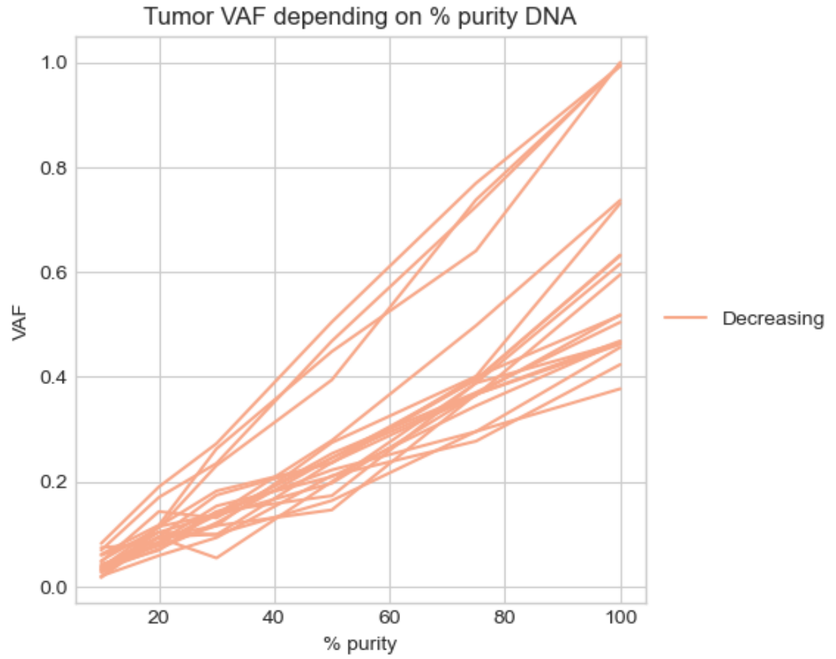
- The presence of VAF correlation from purity - "Decreasing" label
- No correlation - "Stable"

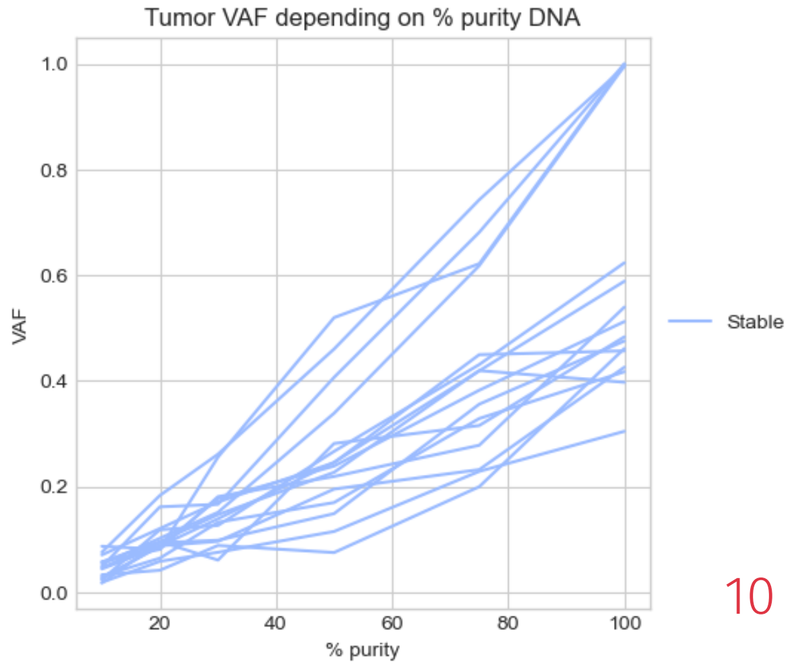
FDR control, alpha = 0.1



COLO DNA VAF plots

- The color of VAF DNA is the same as that of the same mutations in RNA
- Mutations that are stable in RNA decline in DNA
- Most likely these are cancer mutations



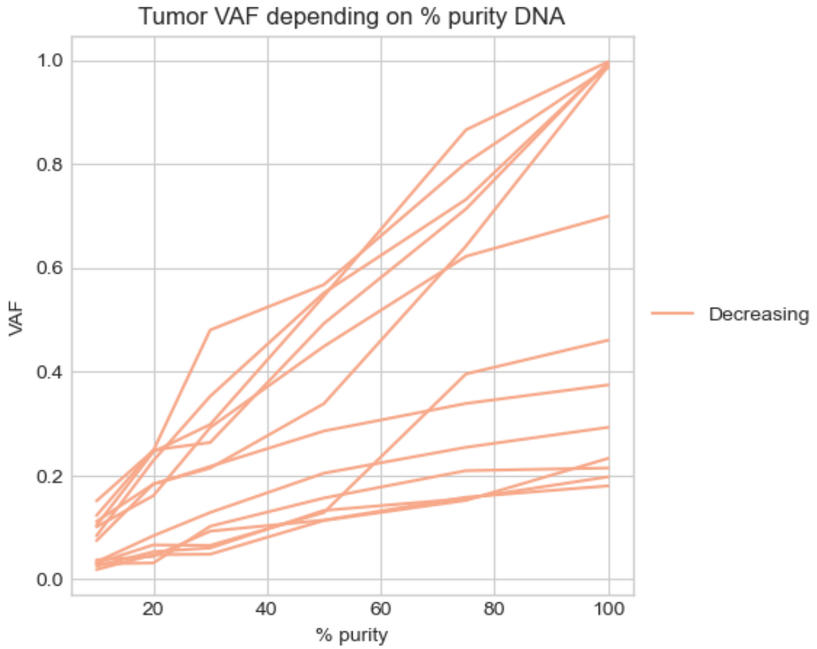


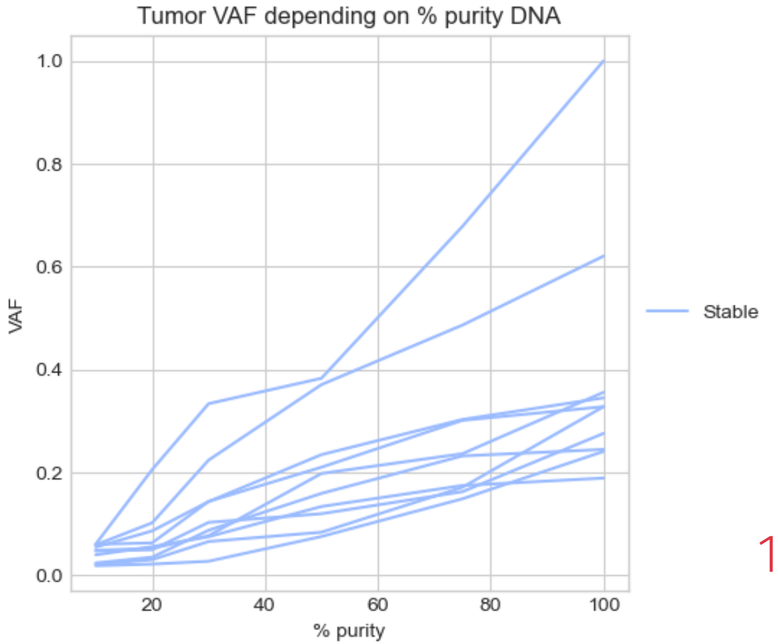
HCC RNA VAF plot

FDR control, alpha = 0.1

HCC DNA VAF plots

- The color of VAF DNA is the same as that of the same mutations in RNA
- Mutations that are stable in RNA decline in DNA
- Most likely these are cancer mutations





COLO results annotation

MGAT4A, ALS2, ETV5, CRACD, VCAN, ABCB5, PLOD3, MTUS1, MPPED2, TNKS1BP1, CPB2, ARHGAP5, PLD3, TPTE, CECR2, FGD1

- Most of them have cancerous impact
- Their GO correlates with membrane, Golgi functions

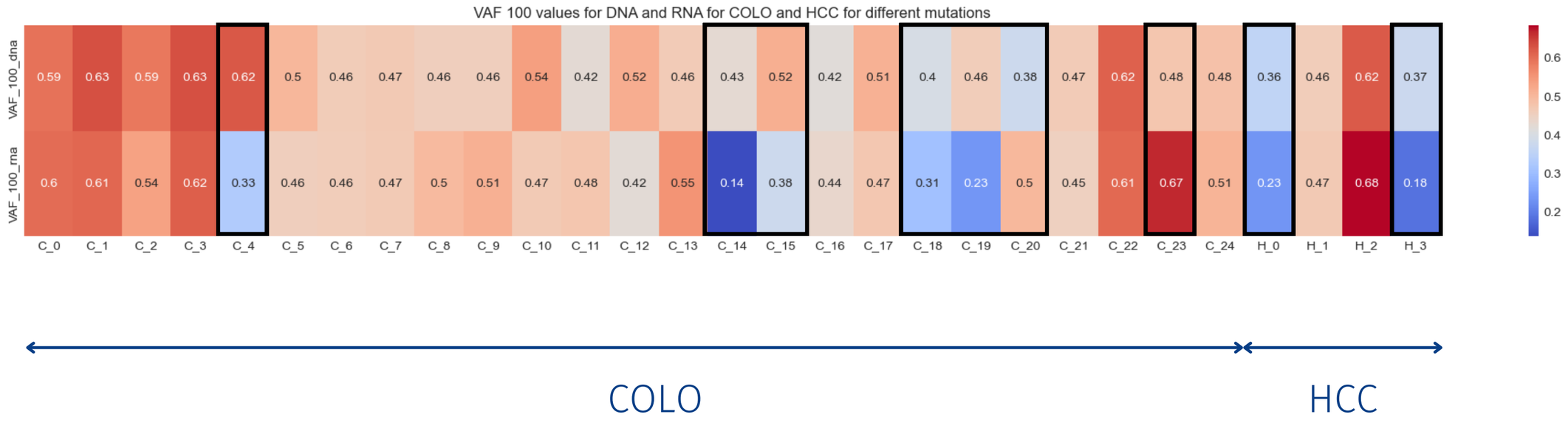
HCC results annotation

HCN3, NOS1AP, RAPH1, SERPINE2, CC2D2A, GABRP, NET1, TLE6, HOOK2, OPHN1

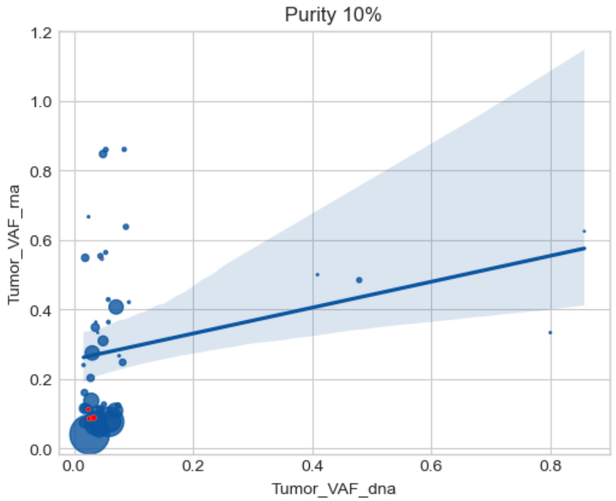
- Most of them have cancerous impact
- Their GO correlates with cytoplasm

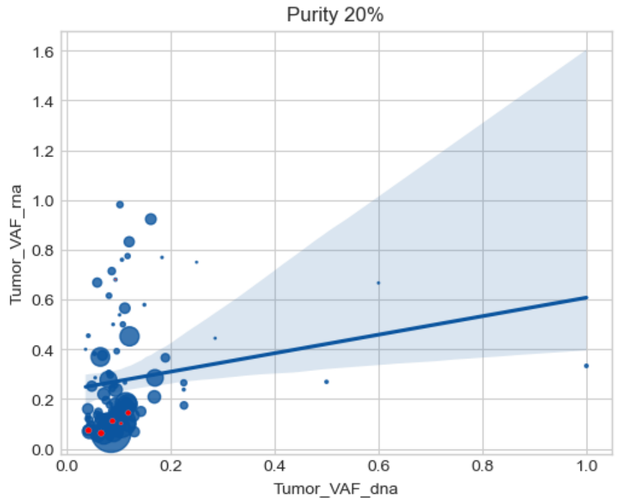
cell projection, signal transduction, synapse, nucleus,

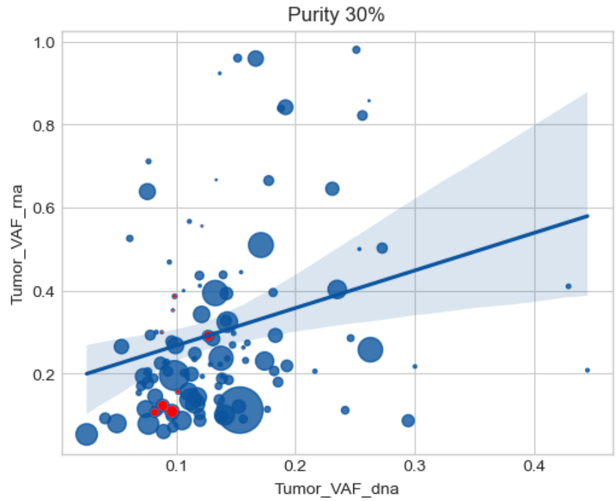
Mutations analysis

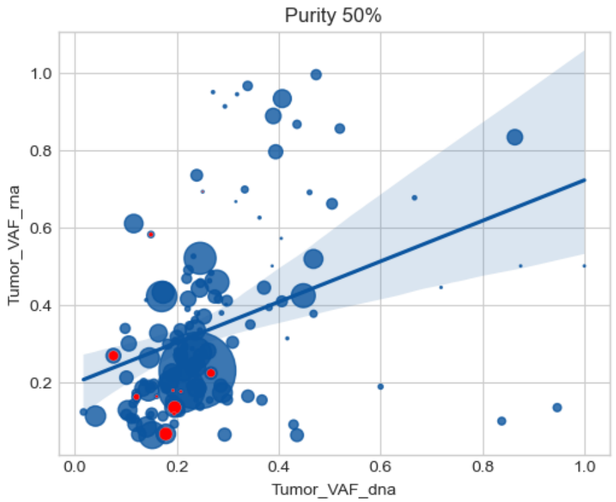


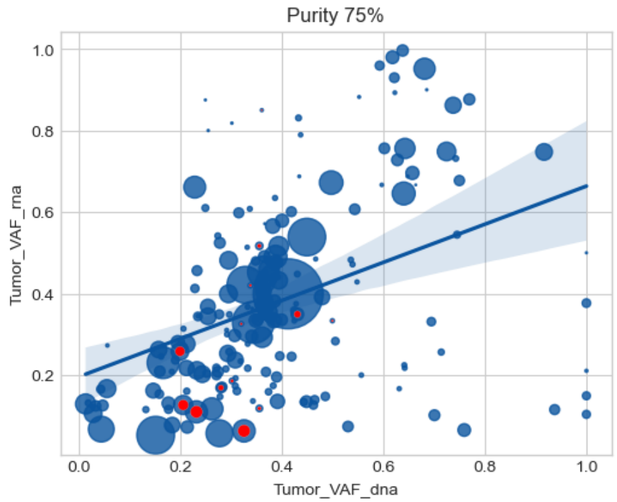
Mutations analysis COLO

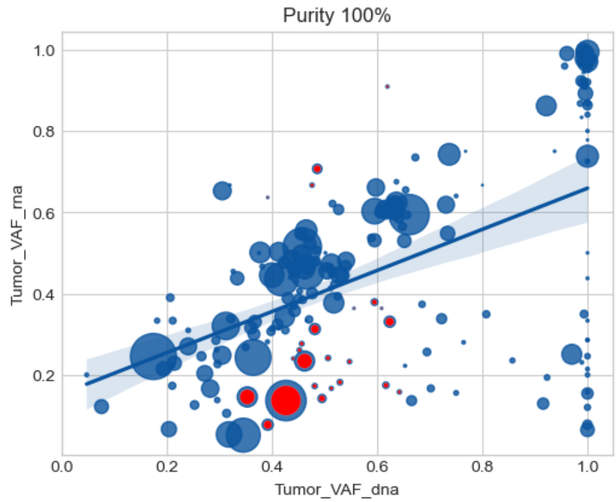






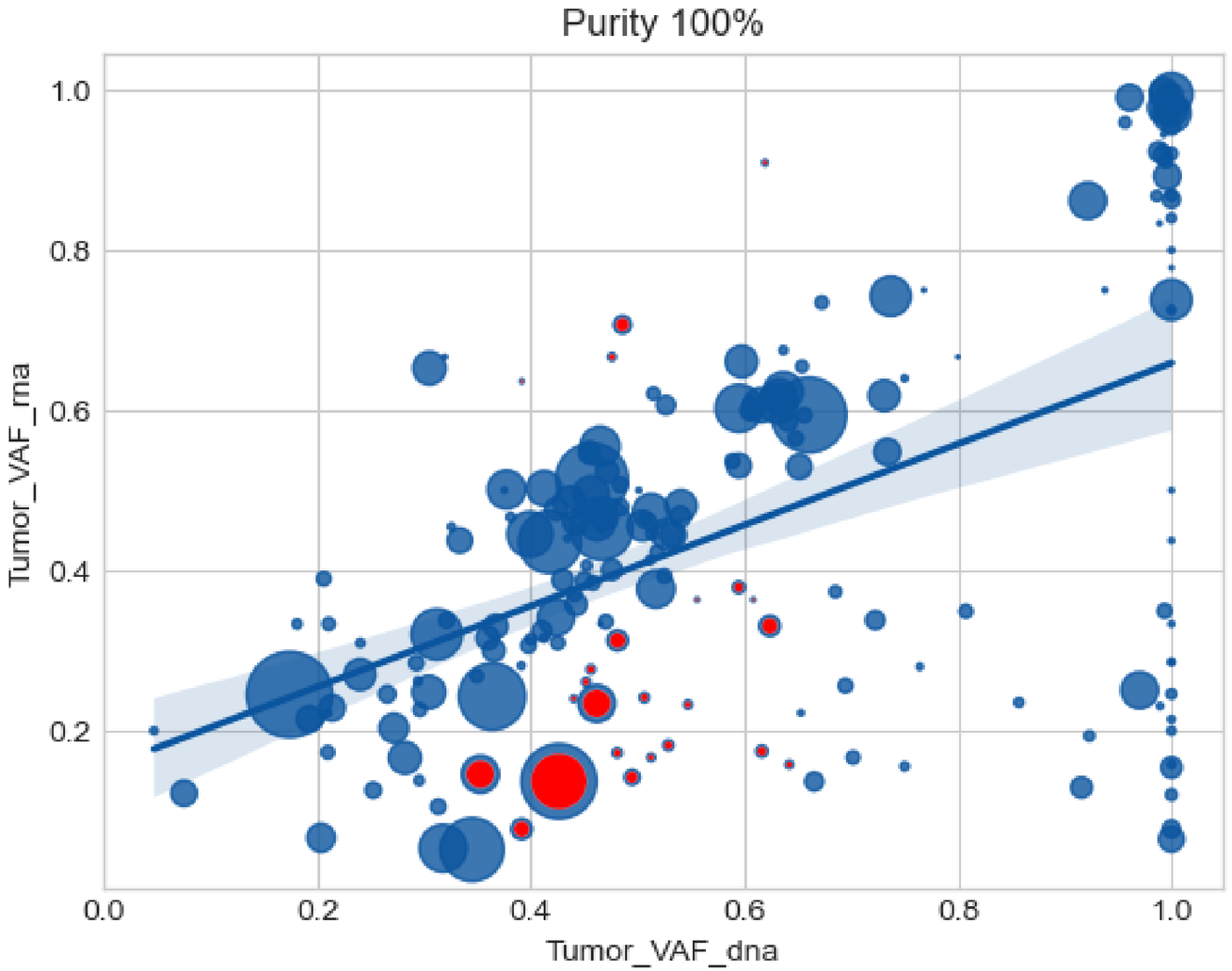






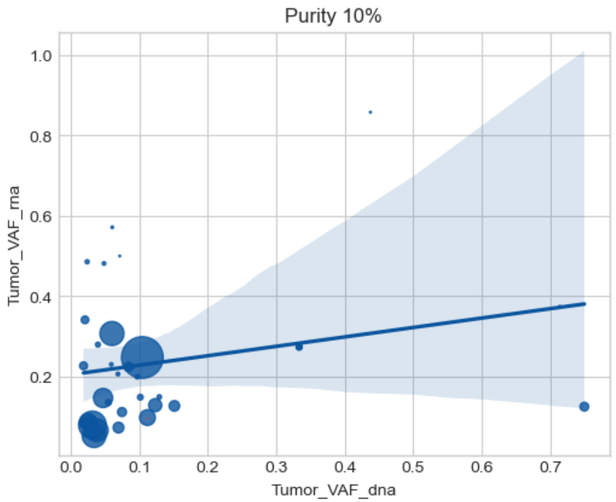
dots size - tumor depth
red dots - probably allel - specific mutations

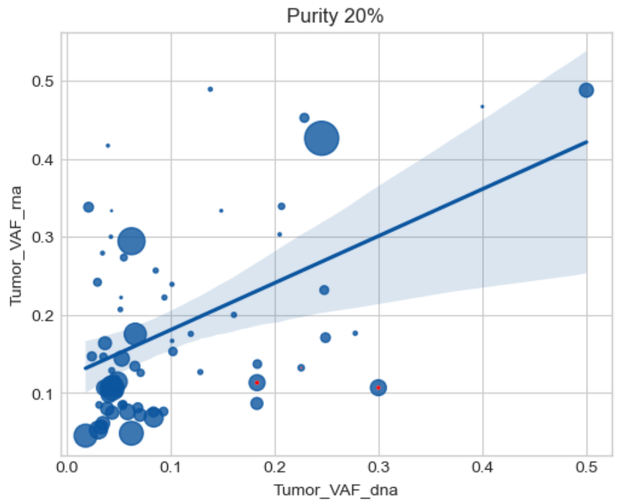
COLO plot for 100% purity

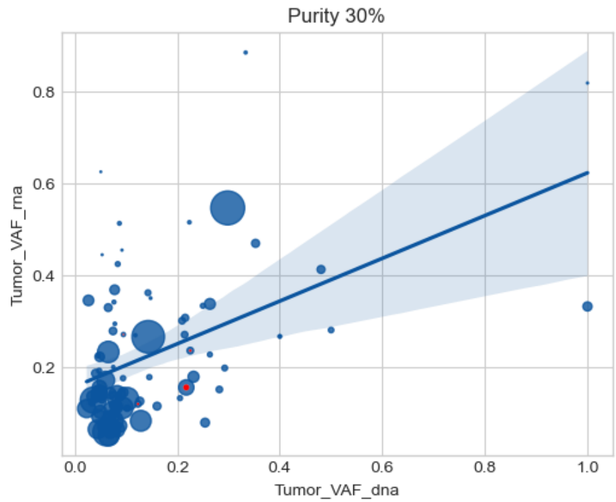


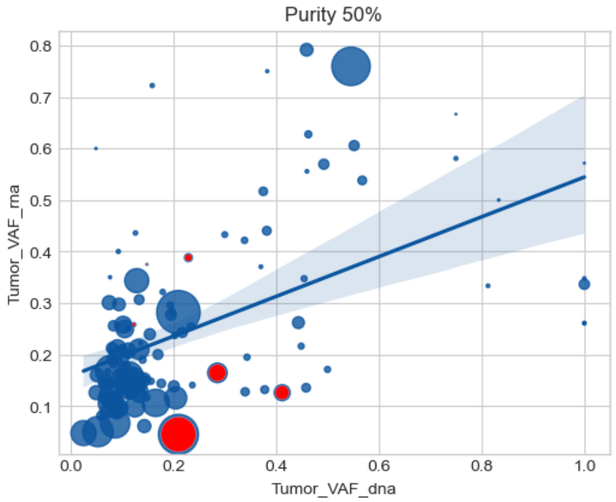
dots size - tumor depth
red dots - probably allel
- specific mutations

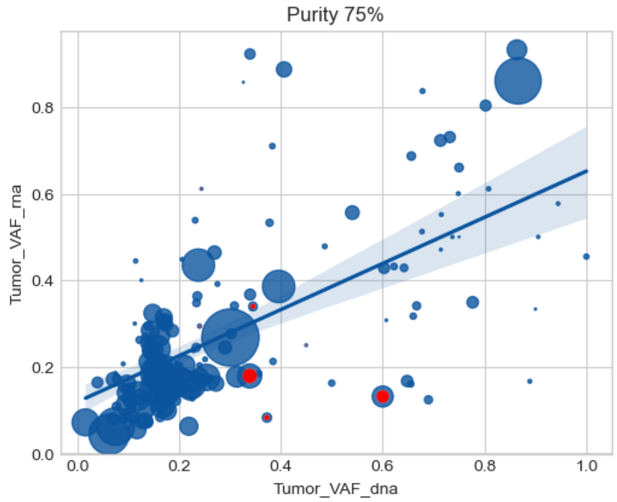
Mutations analysis HCC

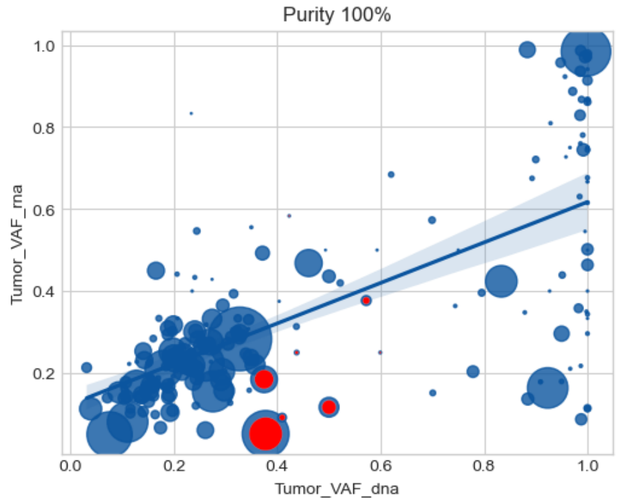






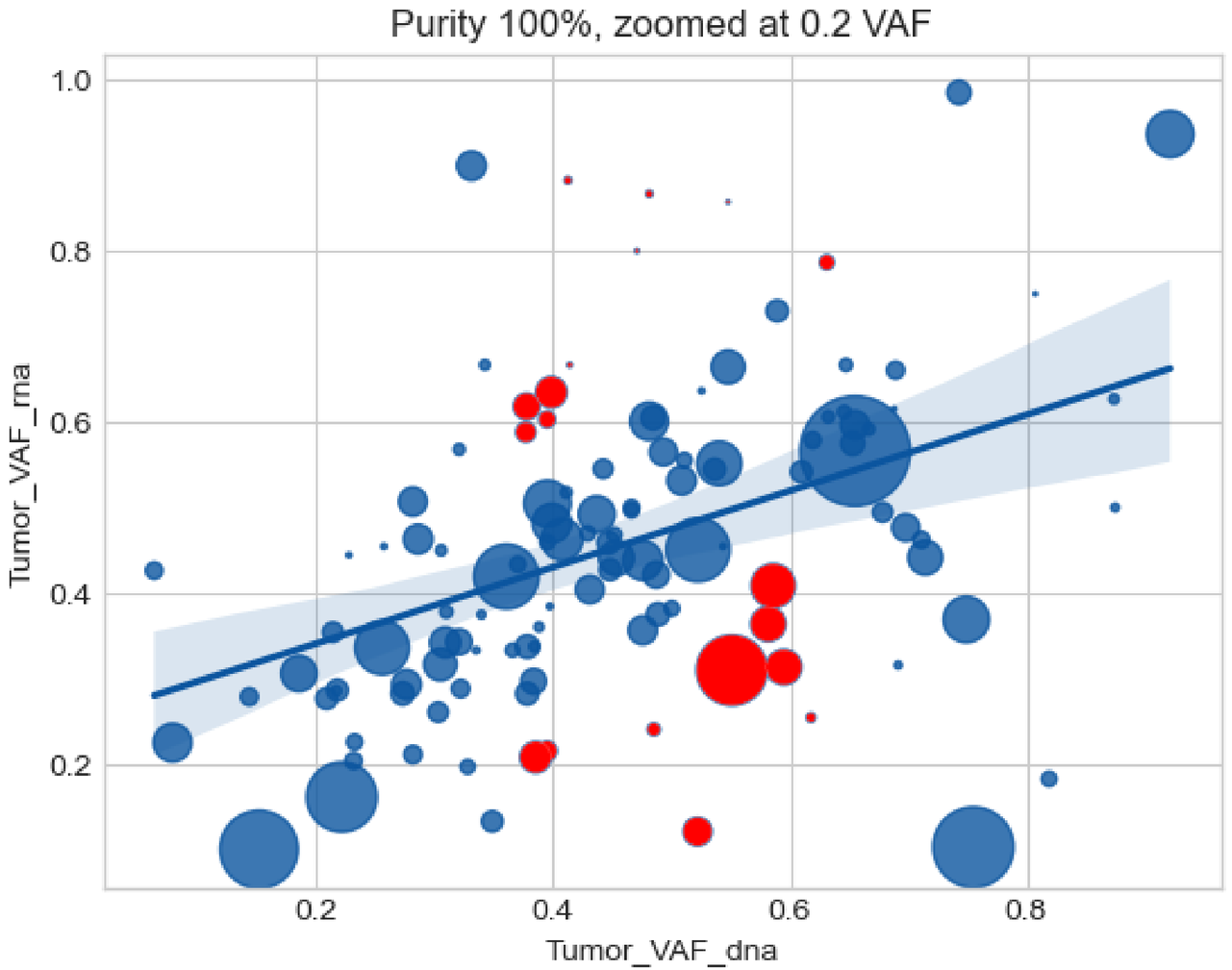






dots size - tumor depth
red dots - probably allel - specific mutations

HCC plot for 100% purity



dots size - tumor depth
red dots - probably allel
- specific mutations

Conclusions

- At a qualitative level, it turned out to distinguish between RNA and WES files
- The dynamics of mutation frequencies depending on the breeding of cell lines has been studied
- Allele-specific mutations have been identified

COLO: ALS2, ABCB5, PLD3, CECR2

HCC: NET1, OPHN1