

Using transcriptome to study relapse in leukaemia

Anna Quaglieri

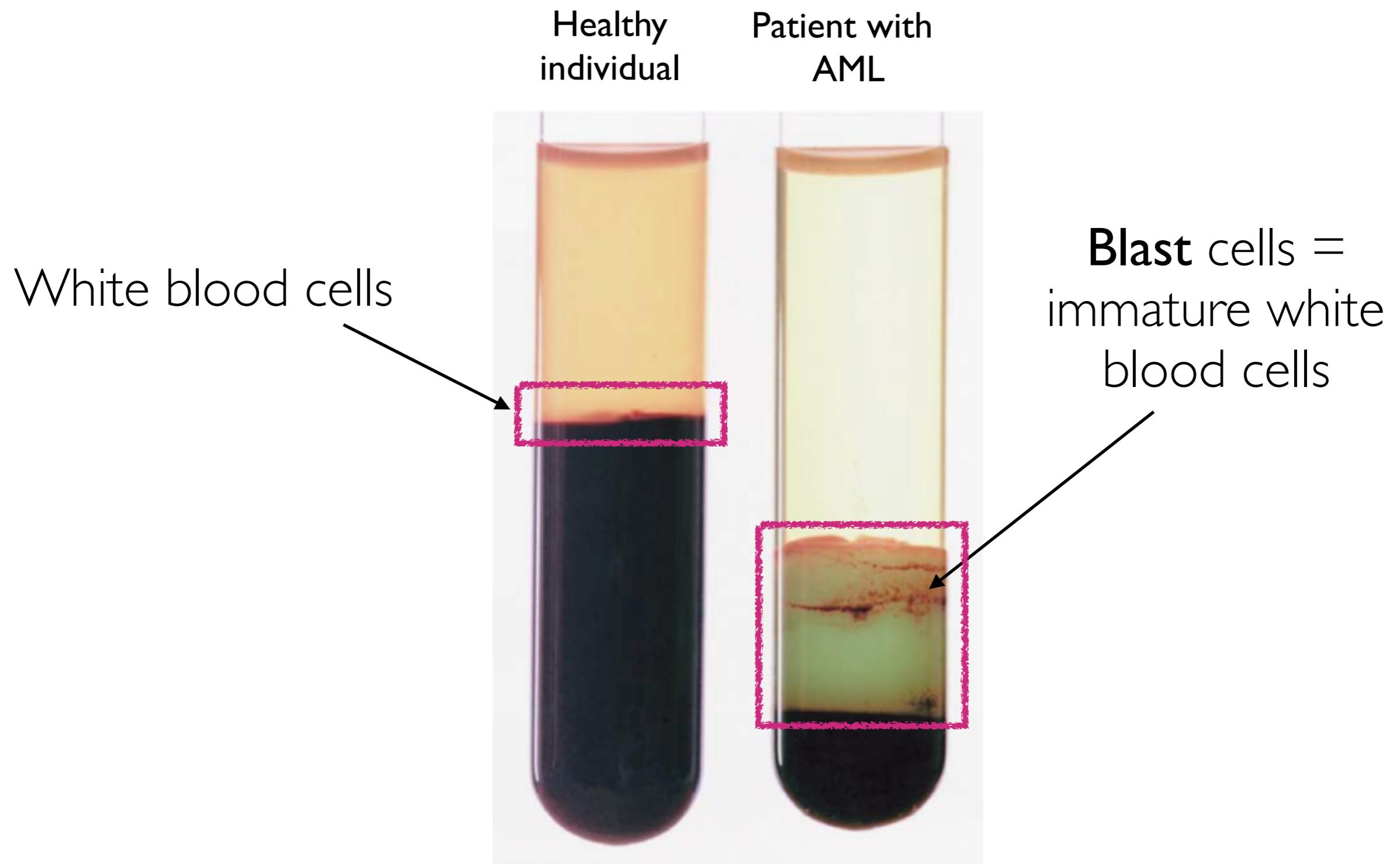
PhD Completion seminar
19 - 08 - 2019

Acute myeloid leukemia (AML)

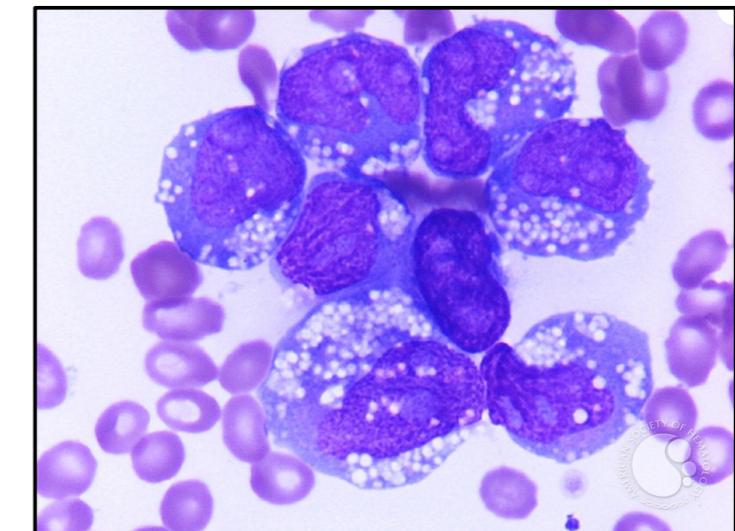
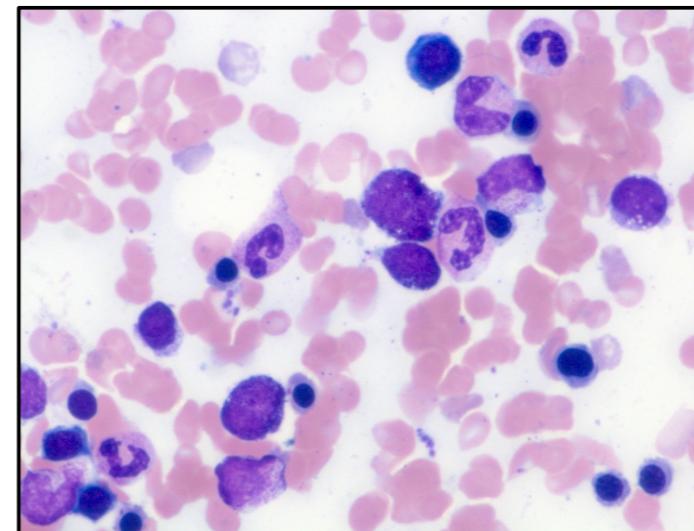
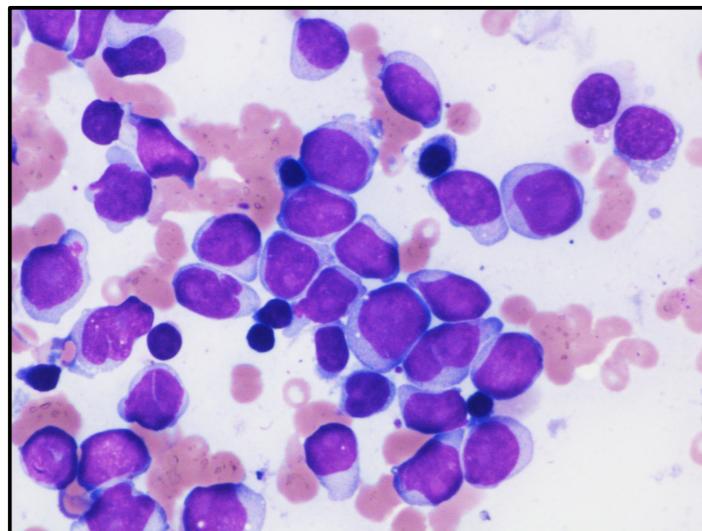
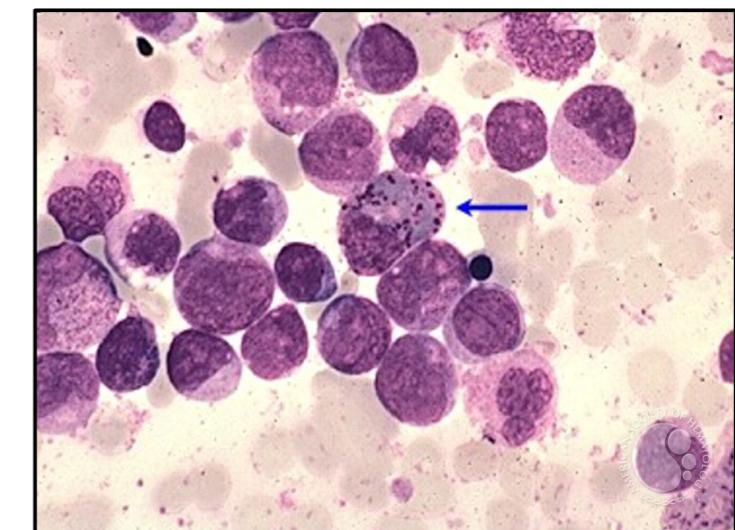
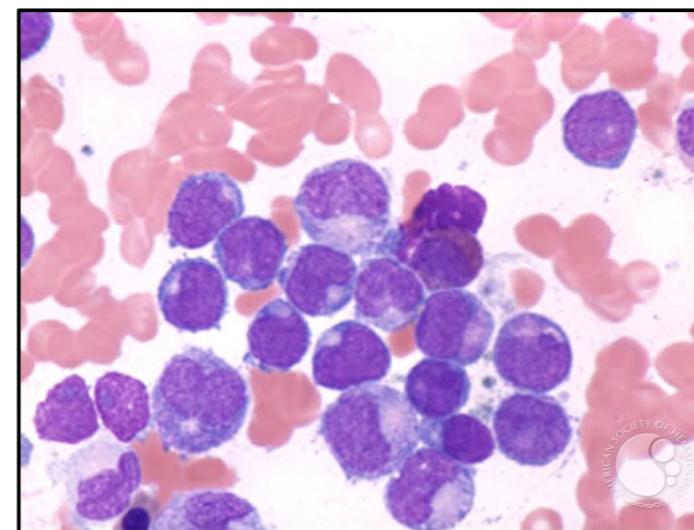
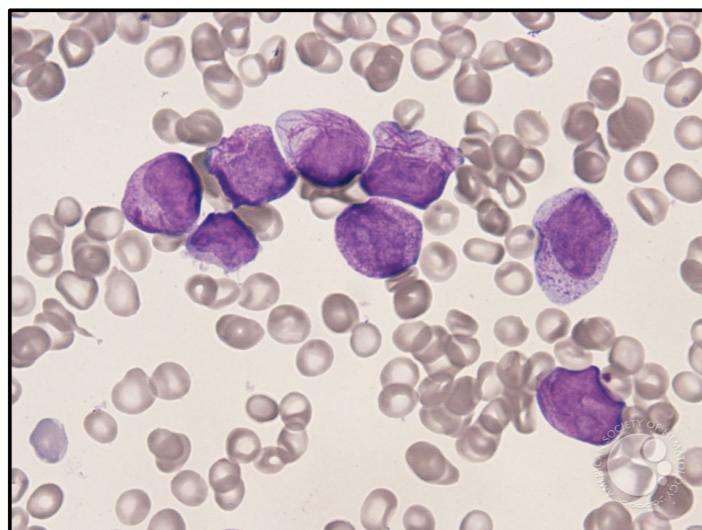
Acute myeloid leukemia (AML)



Acute myeloid leukemia (AML)

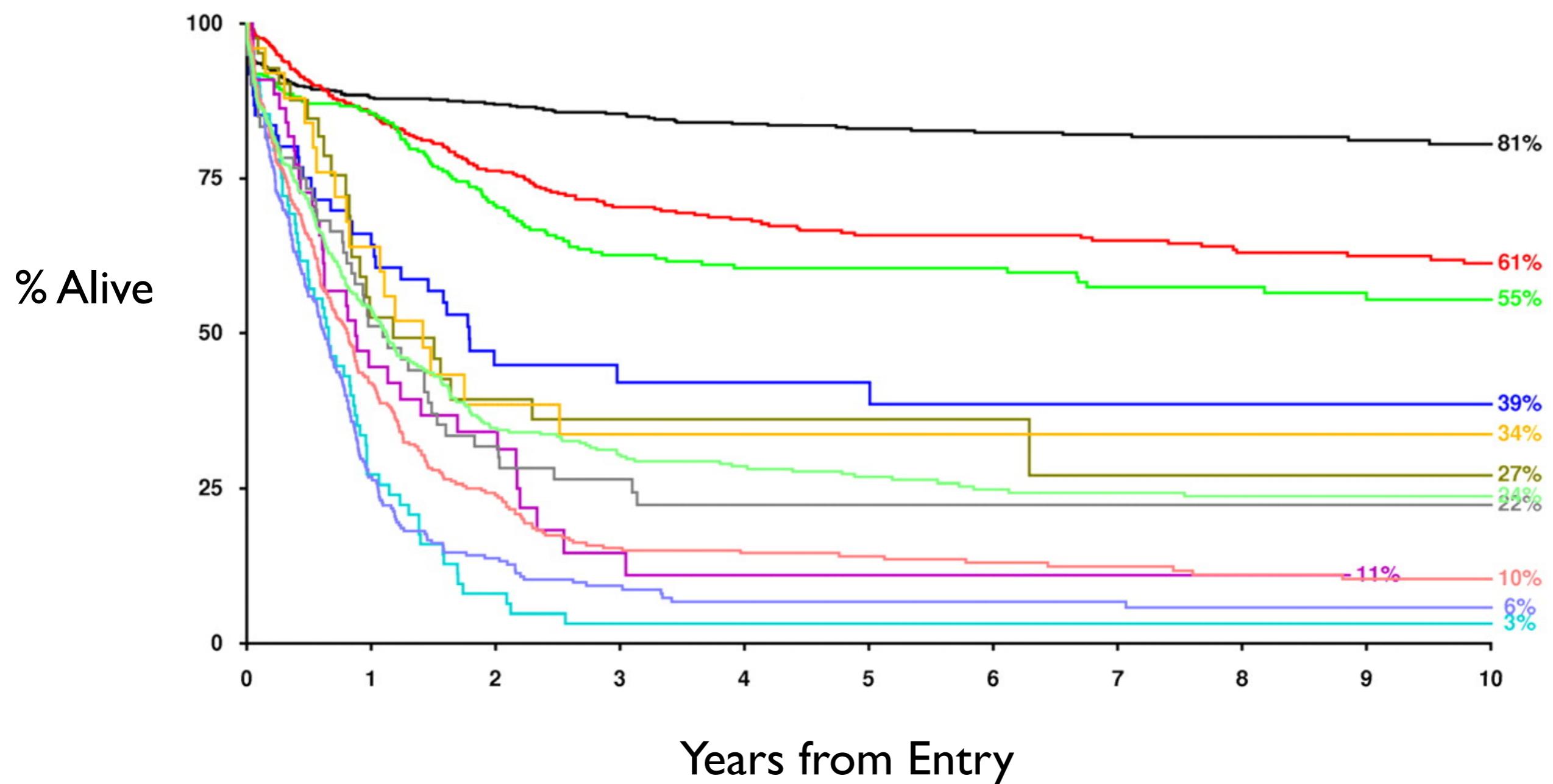


AML is a heterogeneous disease



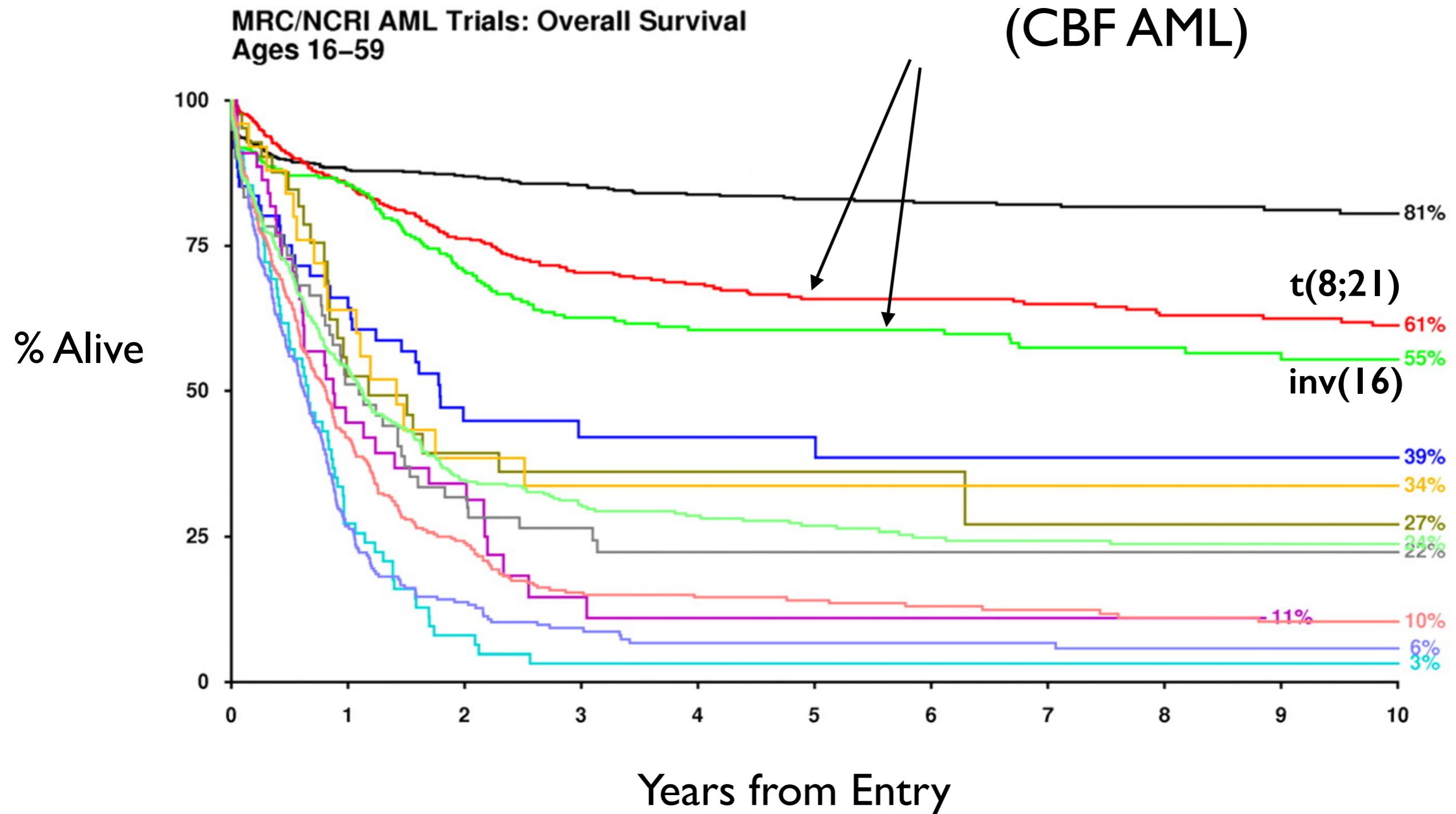
Survival in AML

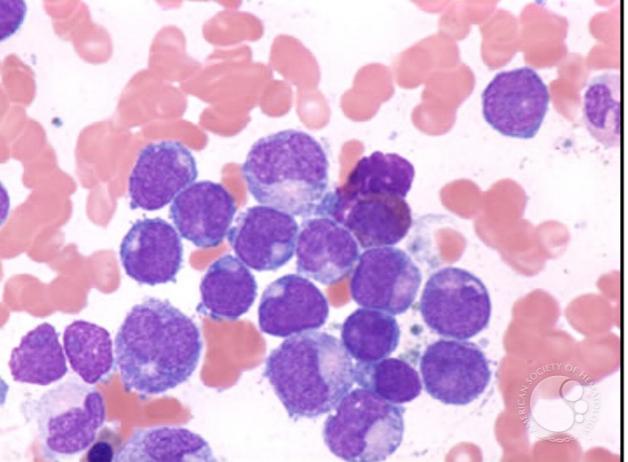
MRC/NCRI AML Trials: Overall Survival
Ages 16–59



Survival in AML

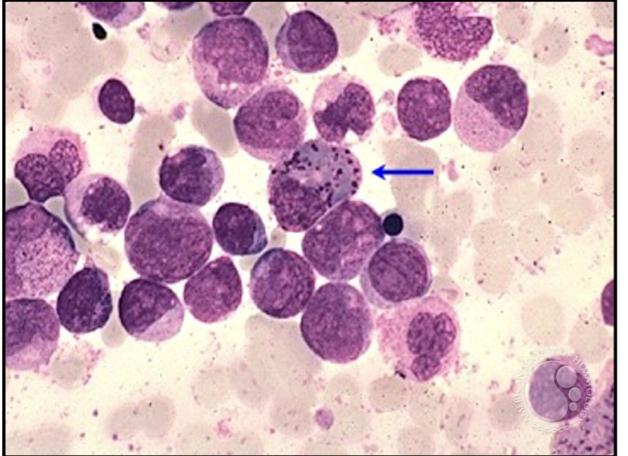
Core Binding Factor AML (CBF AML)



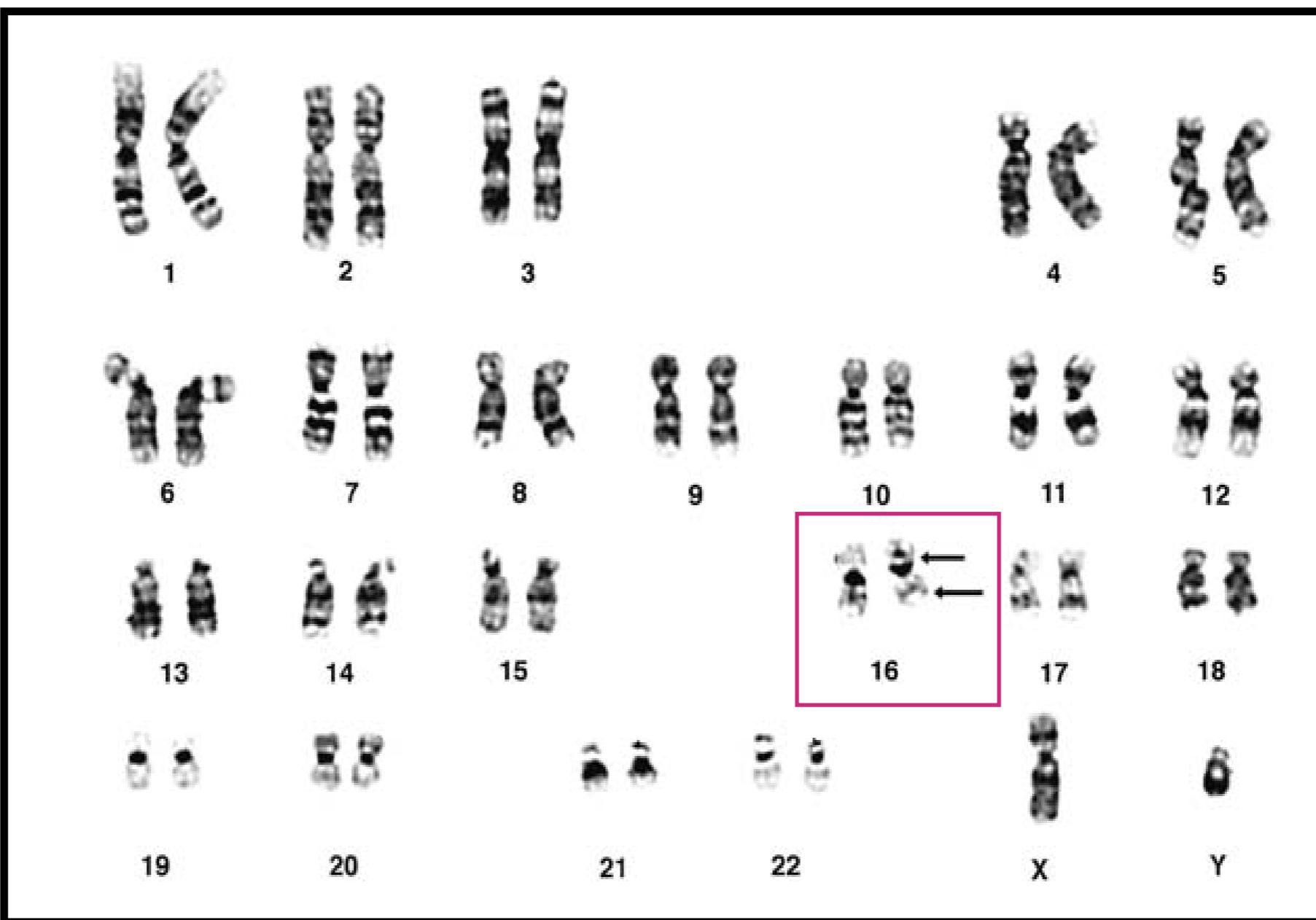


AML with t(8;21)





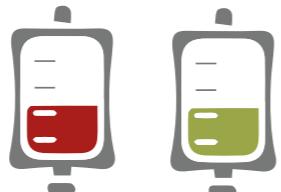
AML with inv(16)



CBF AML treatment

CBF AML treatment

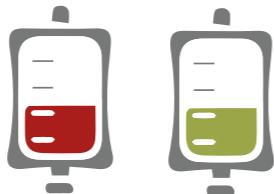
I 1960s



Proven efficacy of Cytarabine
and Combination Chemotherapy

CBF AML treatment

I 1960s



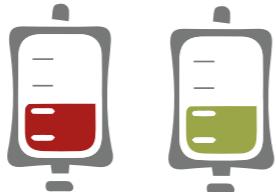
Proven efficacy of Cytarabine
and Combination Chemotherapy

>30 Years

None or little progress in
new therapies

CBF AML treatment

I 1960s



Proven efficacy of Cytarabine
and Combination Chemotherapy

>30 Years

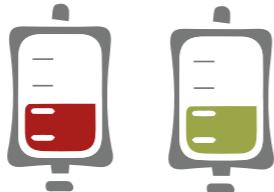
None or little progress in
new therapies

Relapse

is the major cause of
treatment failure

CBF AML treatment

I 1960s



Proven efficacy of Cytarabine
and Combination Chemotherapy

>30 Years

None or little progress in
new therapies

Relapse

is the major cause of
treatment failure

~40% Relapse

Improve understanding
of relapse

Study cohort

37 CBF AML patients from Australia
Standardised treatment - Cytarabine



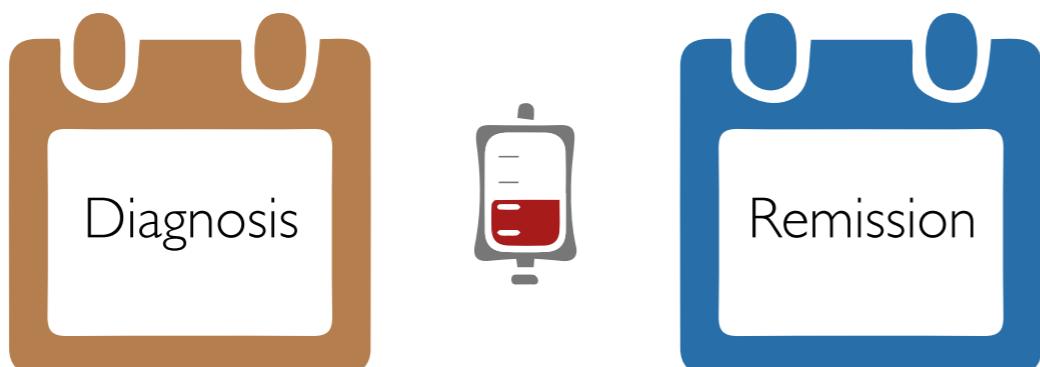
Study cohort

37 CBF AML patients



Study cohort

37 CBF AML patients



Study cohort

37 CBF AML patients

10



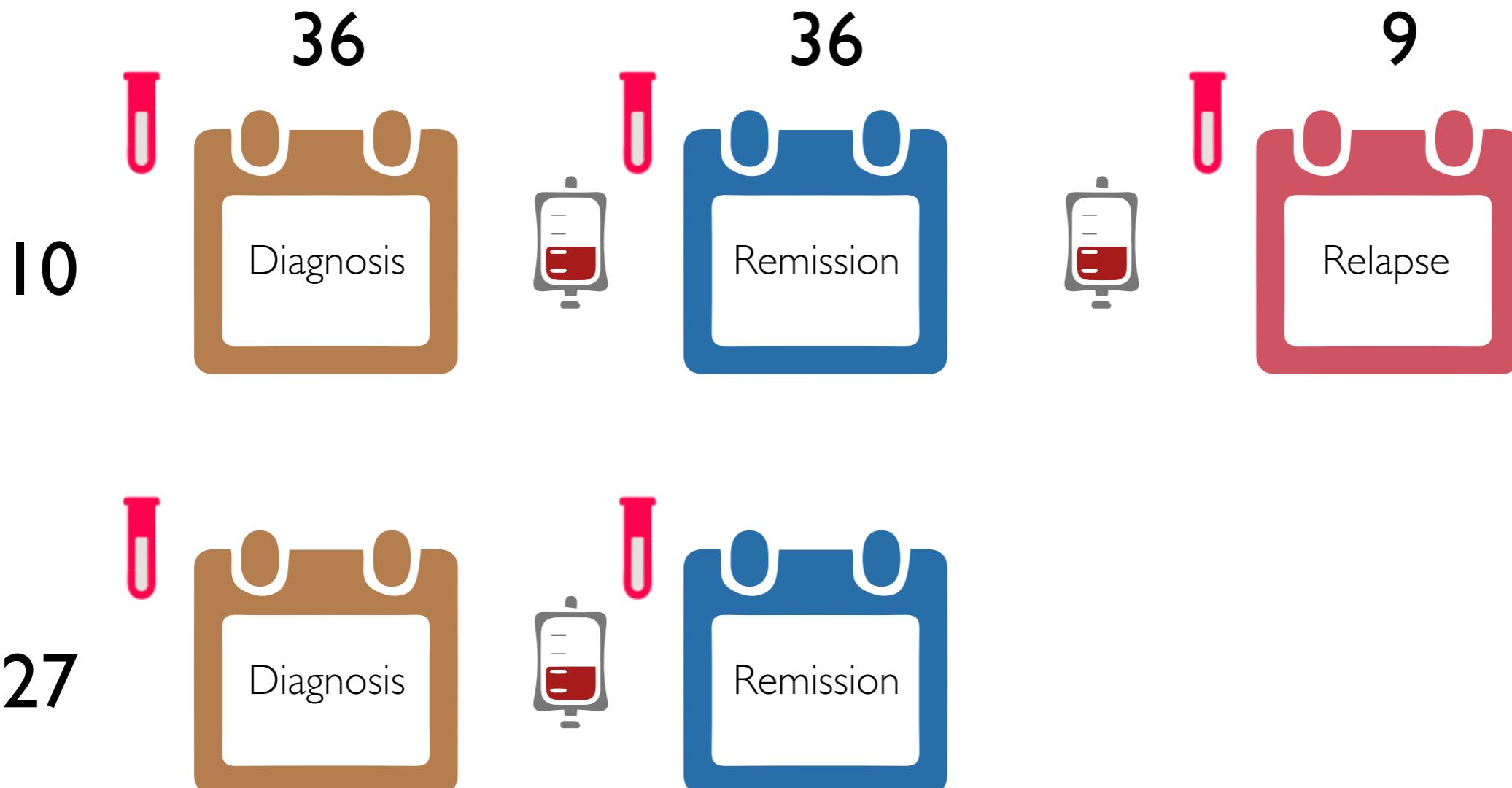
27



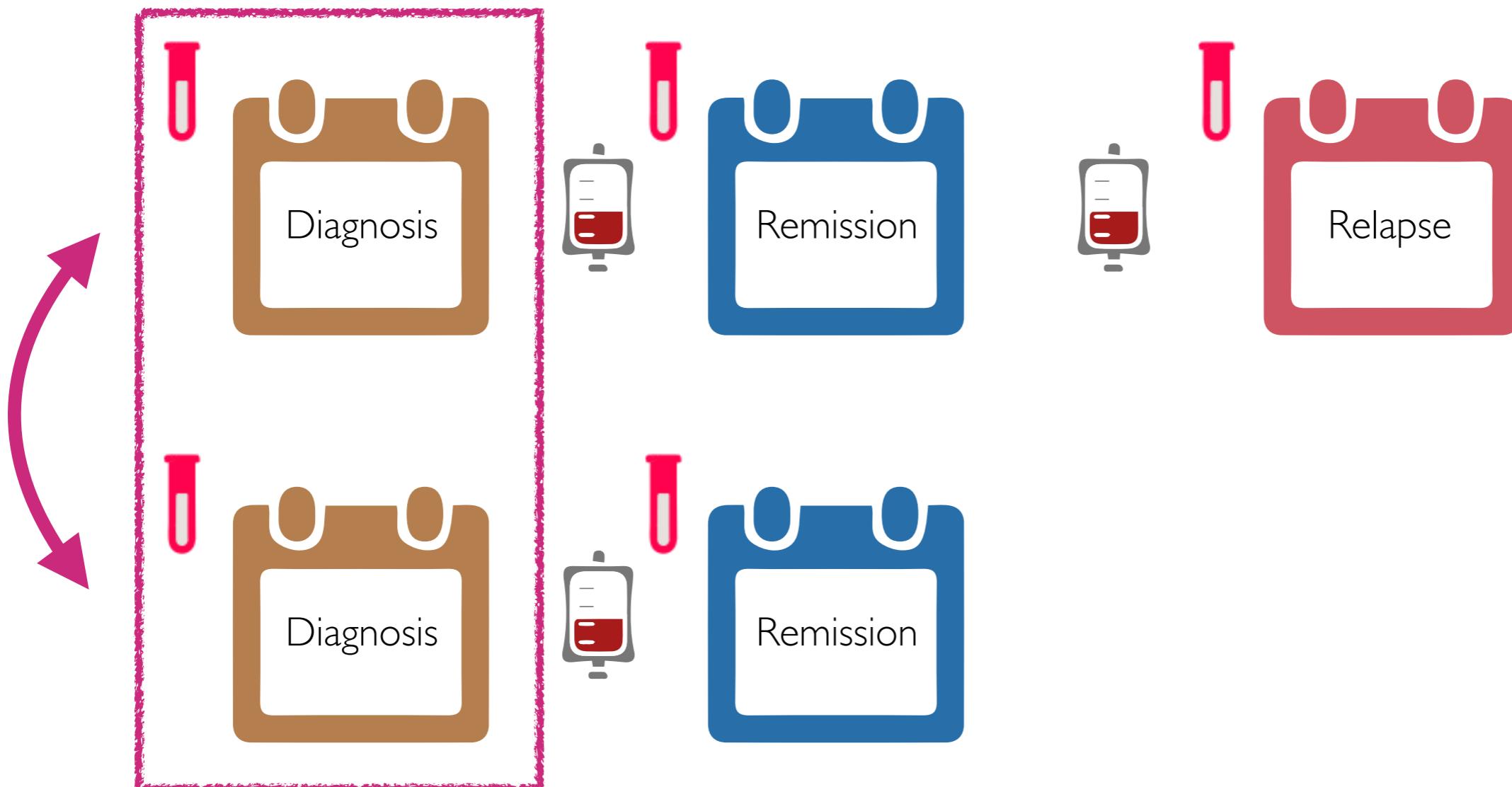
Samples collected Bulk mRNA



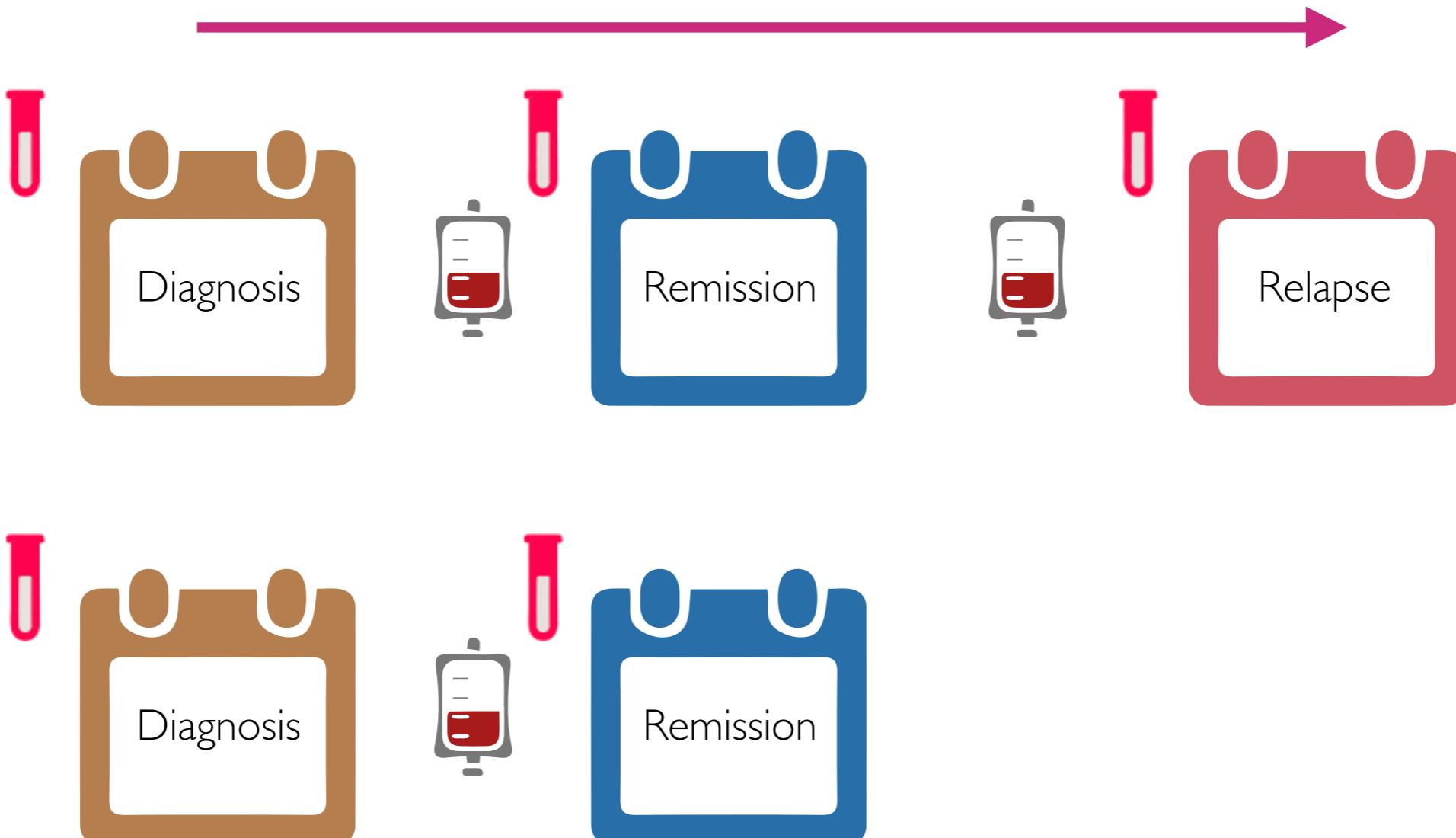
Samples collected Bulk mRNA



Differences at diagnosis

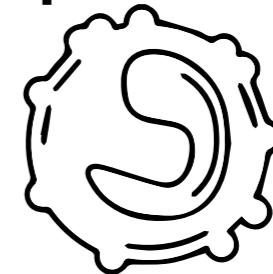


Changes over time

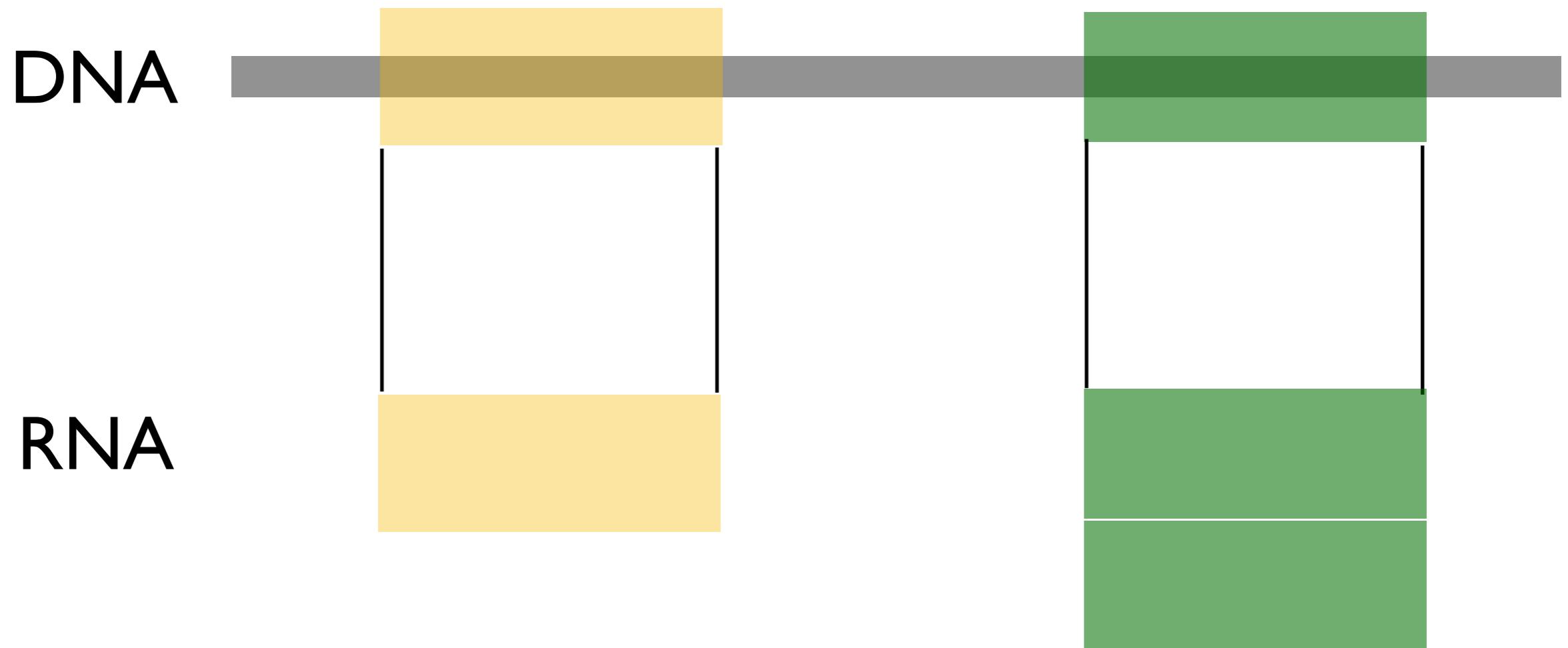
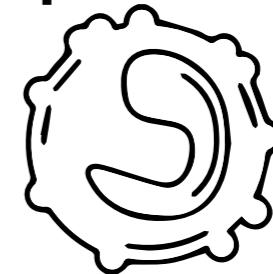


Why RNA?

DNA to RNA transcription in normal cells



DNA to RNA transcription in normal cells

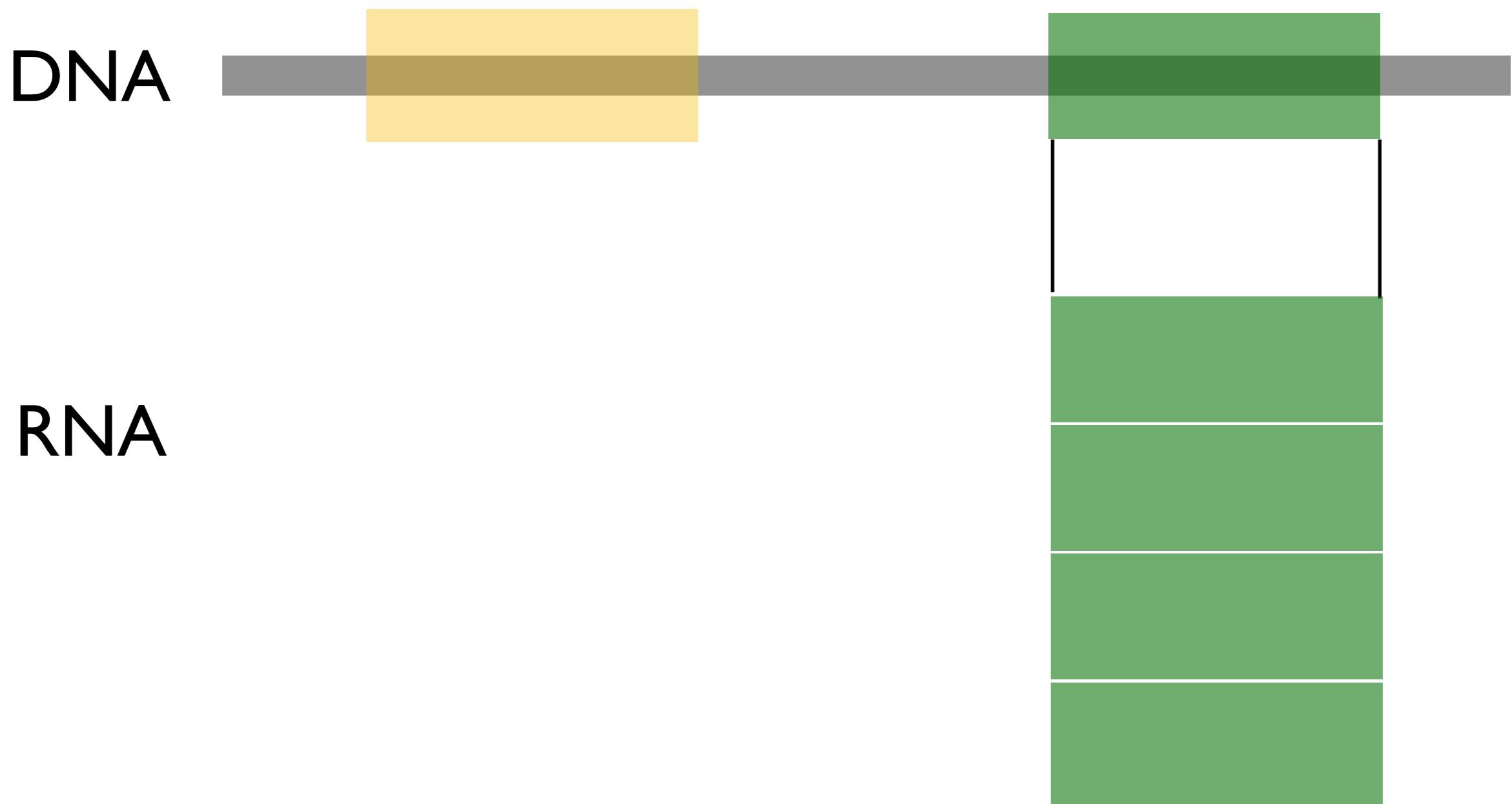
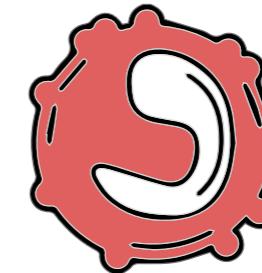


DNA to RNA transcription in cancer cells

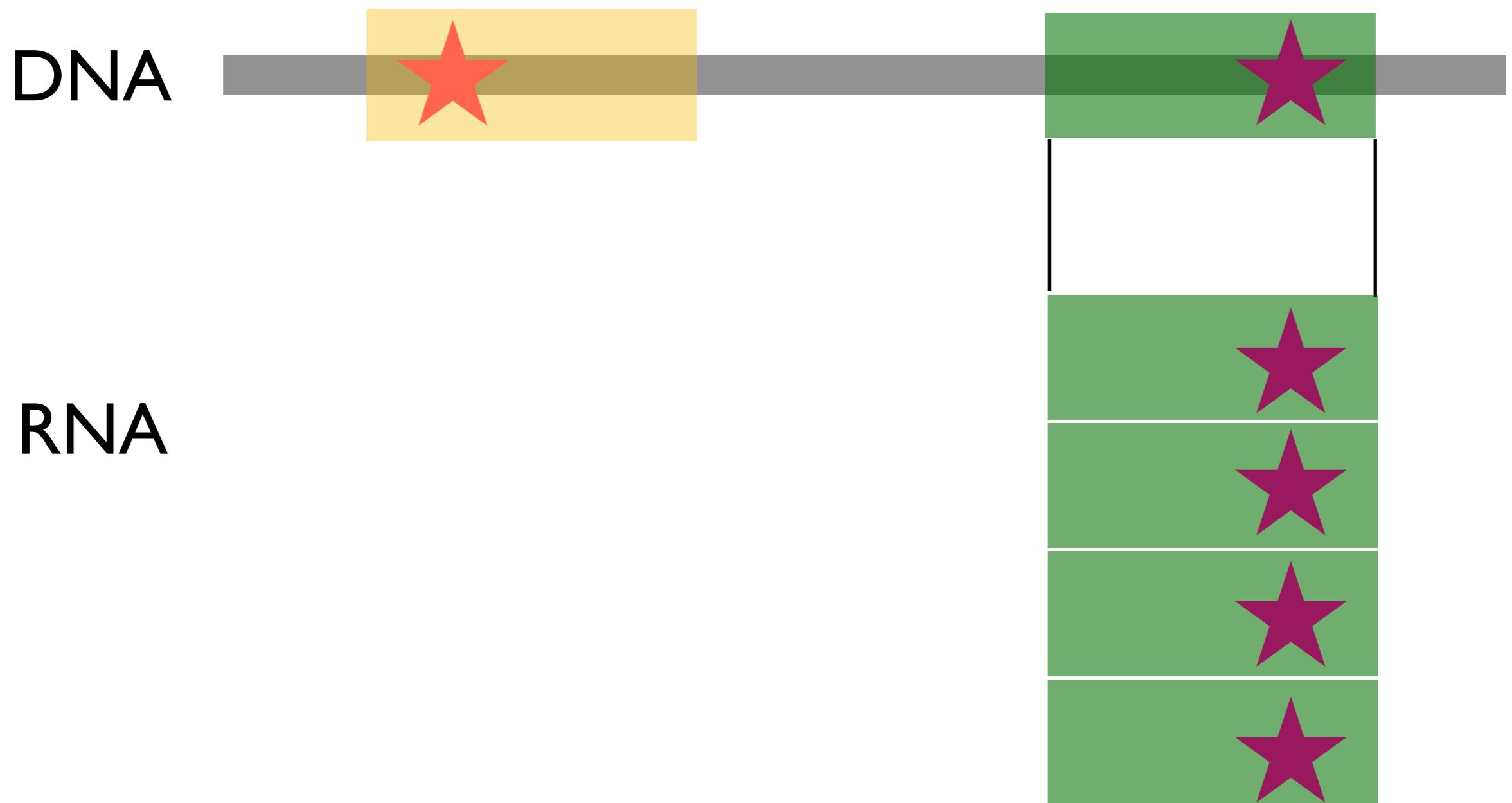
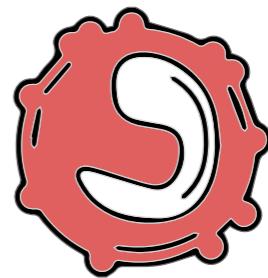


RNA

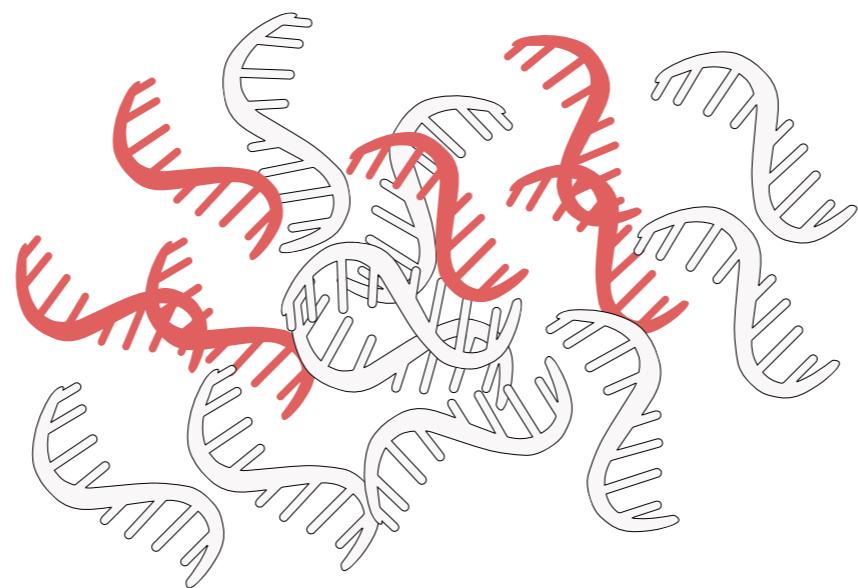
Detect gene expression changes in cancer cells



Detect mutations in expressed genes



RNA as a cost-effective choice

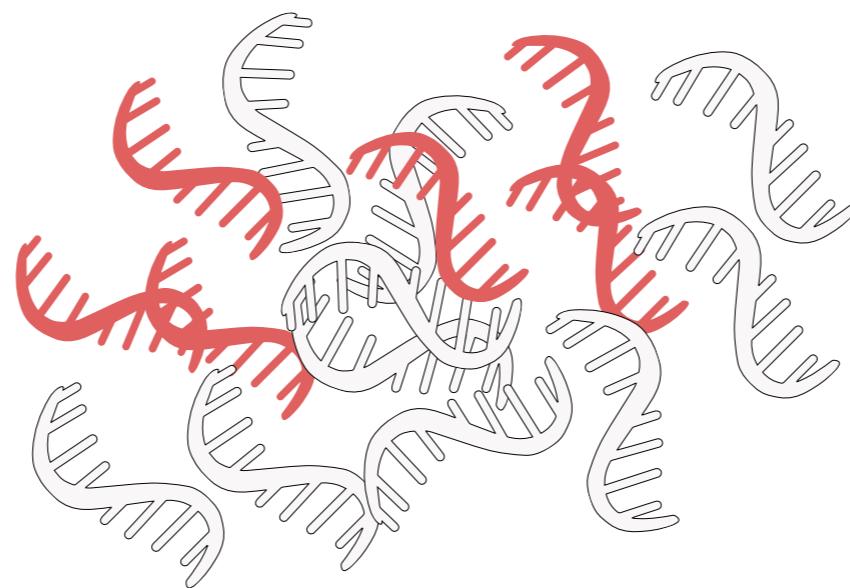


RNA as a cost-effective choice



Extract and compare
gene expression

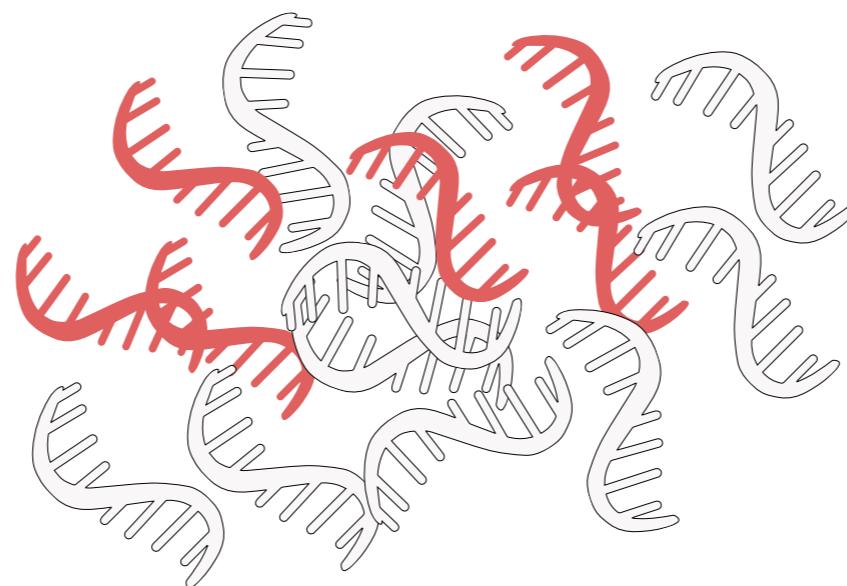
RNA as a cost-effective choice



Extract and compare
gene expression

& Detect mutations
in expressed genes

RNA as a cost-effective choice



Extract and compare
gene expression

&

Detect mutations
in expressed genes

Not commonly investigated using RNA

Workflows to detect mutations in RNA

RNA-Seq variant calling

1 Setup

- 1.1 Overview
- 1.2 Disclaimer

2 Download RNA-Seq data from GEO

- 2.1 Get SRX sample names
- 2.2 Create NCBI query

3 Downsampling FASTQ or BAM files

4 Define files and programs needed for...

5 FASTQC and adapters trimming

- 5.1 Parallelise your FASTQC
- 5.2 Summarise reports with MultiQC
- 5.3 Adapter troubles: STAR vs Subr...

6 Alignment, Read Groups, Mark dupli...

- 6.1 Create STAR index
- 6.2 STAR-1pass
- 6.3 STAR-2pass
- 6.4 Details about post-alignment fu...
- 6.5 Merge bamfiles

7 GATK pre-processing

- 7.1 SplitNCigarReads
- 7.2 Base recalibration

From FASTQ files to Variant Calling for RNA-Seq

Anna Quaglieri

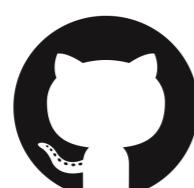
1 Setup

This is an example workflow from `SRR` files to Variant calling using modular functions written in `R` and `bash`.

```
git clone git@github.com:annaquaglieri16/RNA-seq-variant-calling.git  
cd ./RNA-seq-variant-calling
```

All the functions used for the variant calling and downsampling pipeline are inside the `./functions` folder.

- If you want to download sample `FASTQ` files or learn how to download `FASTQ` files from GEO go to Section 2.
- If you already have the `FASTQ` files and YOU WANT TO randomly downsample your samples to a fix number of reads go to Section 3.
- If you already have the `FASTQ` files and you don't need to perform quality control or downsample your files go to Section 6.
- If you already have the `BAM` files and you want to call variants go to Section 7.



rna-mutation-calls.netlify.com/
annaquaglieri16

Back in 2016...



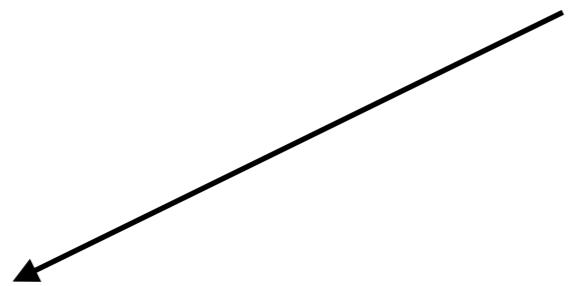
Samples from **blood** and **bone marrow**



Experimental design



Experimental design



Sequencing
design



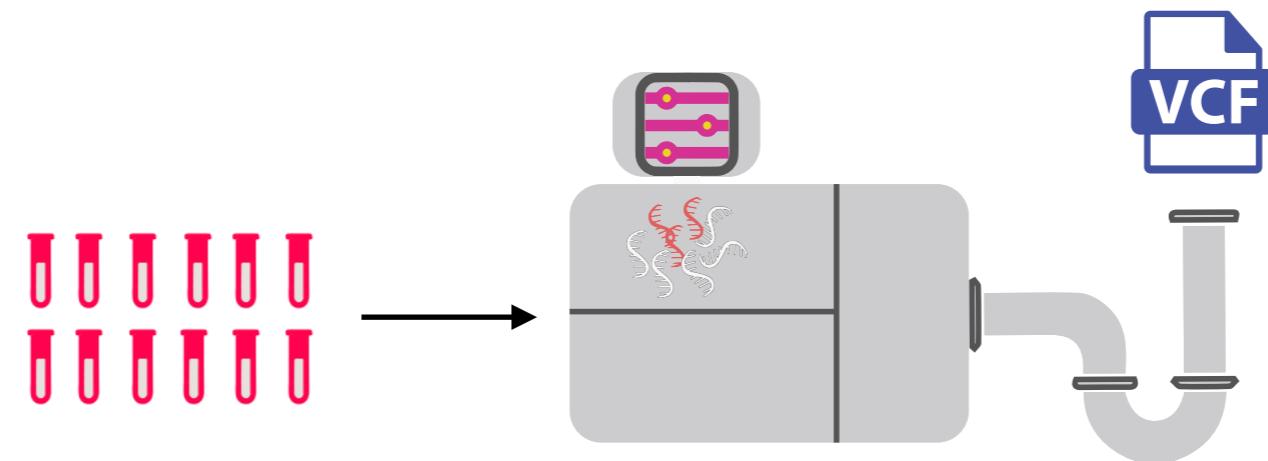
Experimental design

Sequencing
design

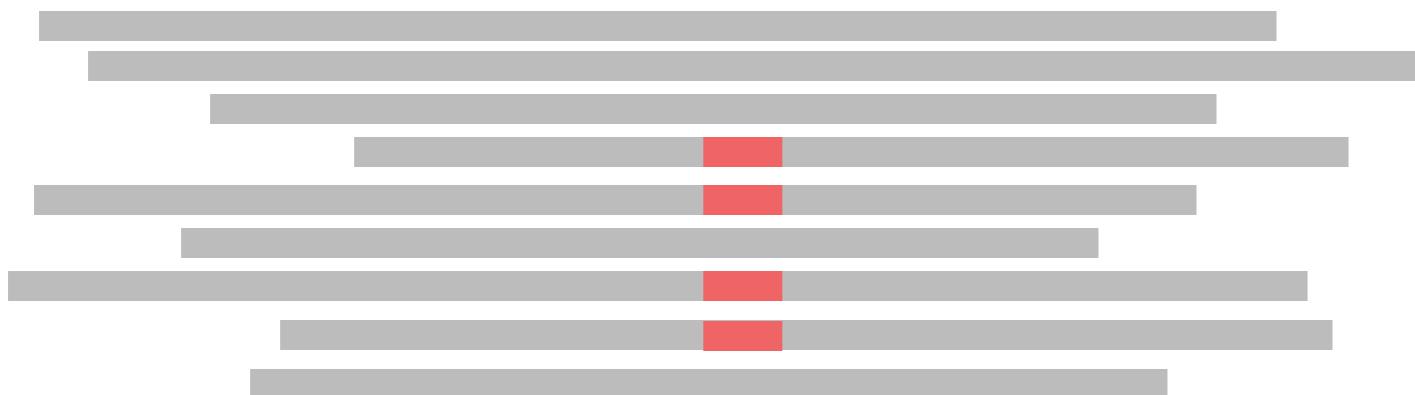
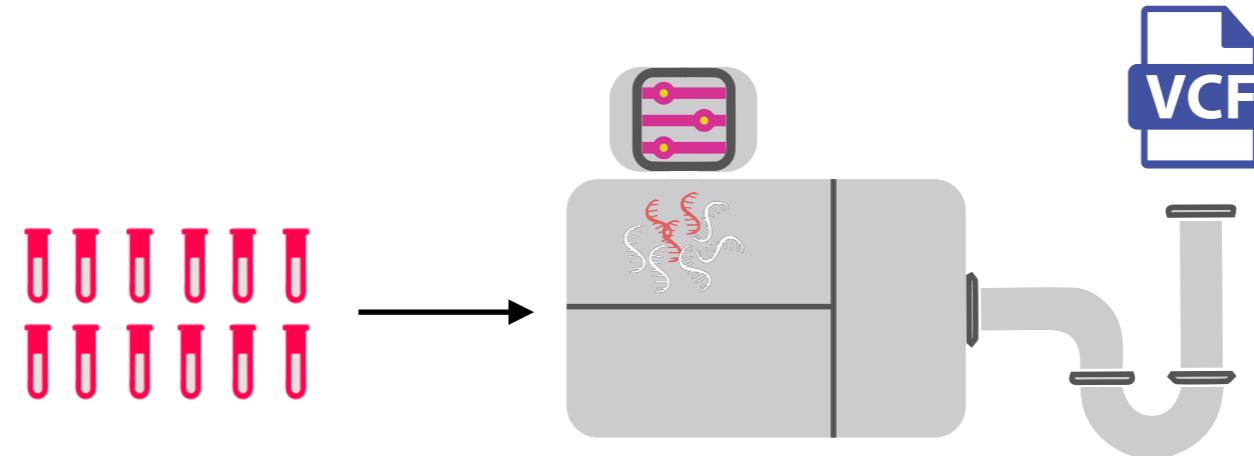


Choice of workflow
to detect mutations
using RNA-Seq

How many fragments to sequence?



How many fragments to sequence?

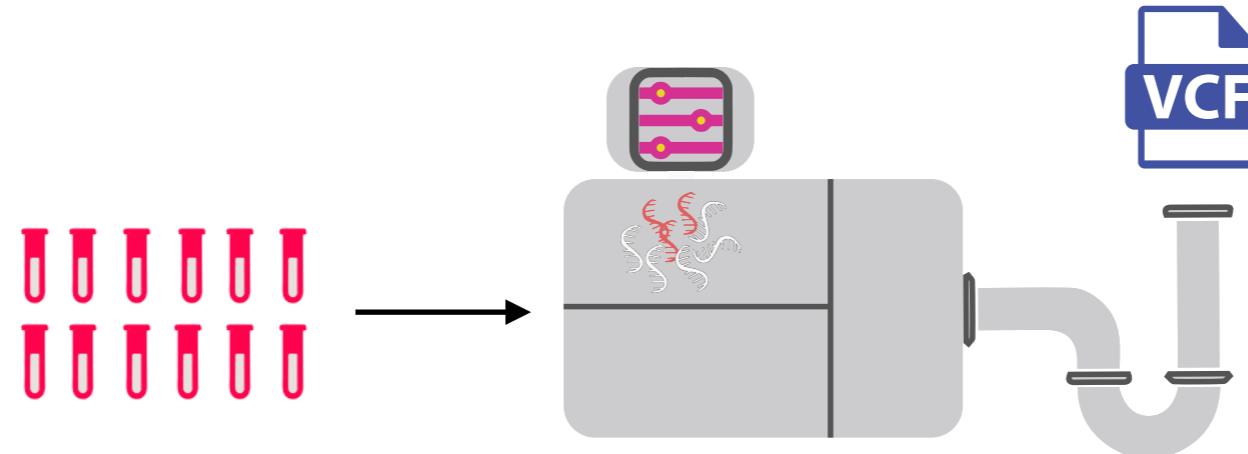


Sequence a lot of RNA fragments



Looks real!

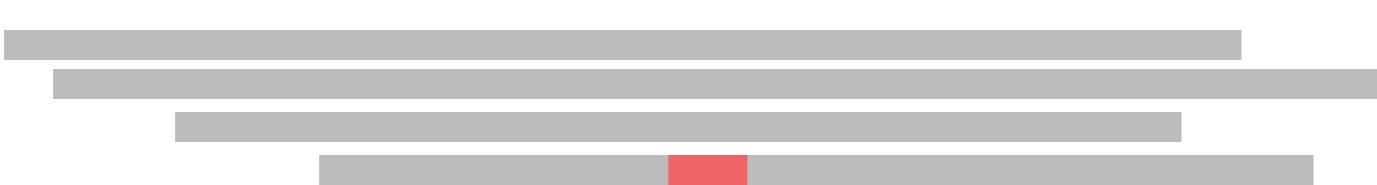
How many fragments to sequence?



Sequence a lot of RNA fragments



Looks real!

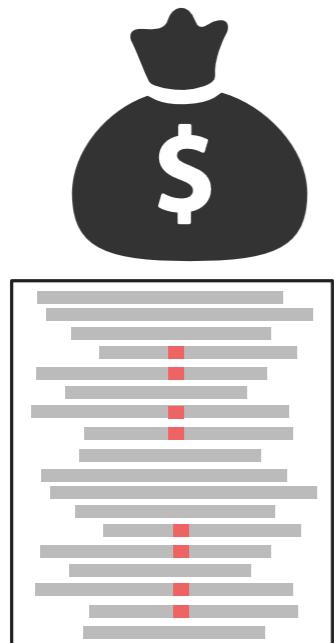


Sequence too few RNA fragments



Probably an error

46
RNA-Seq samples



110M
fragments

46

RNA-Seq samples

“True” mutations
validated in DNA



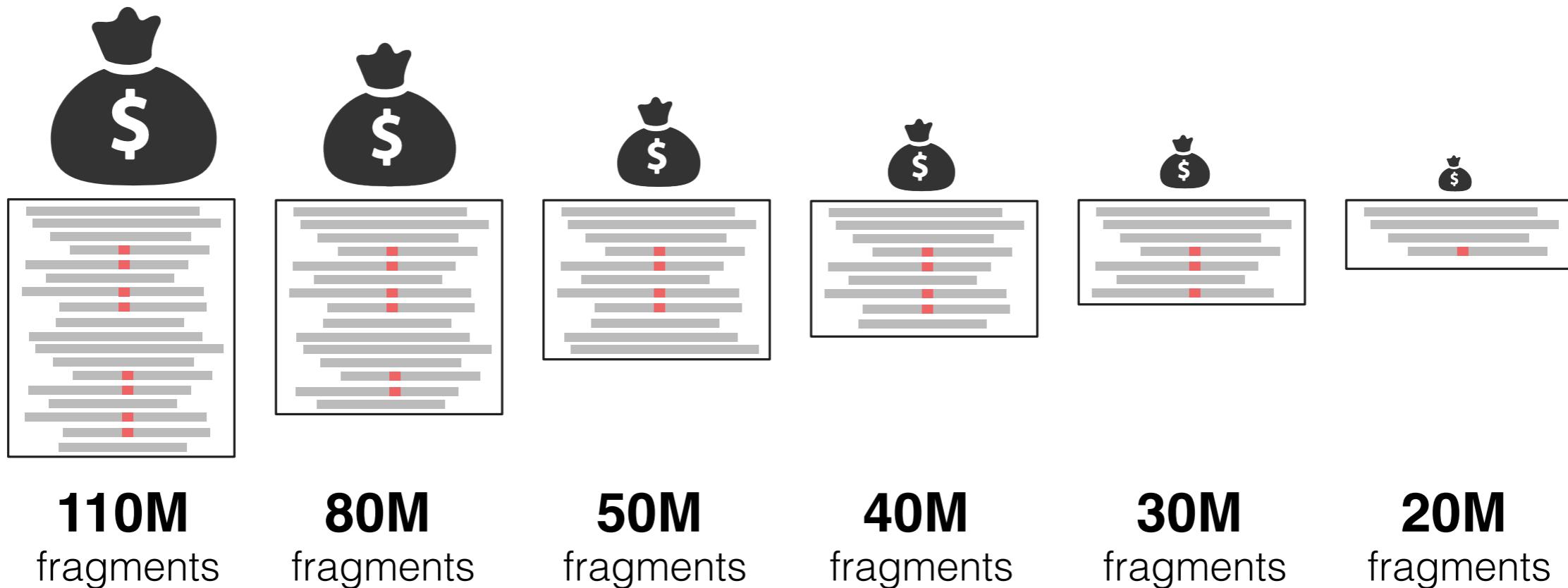
110M
fragments

46

RNA-Seq samples

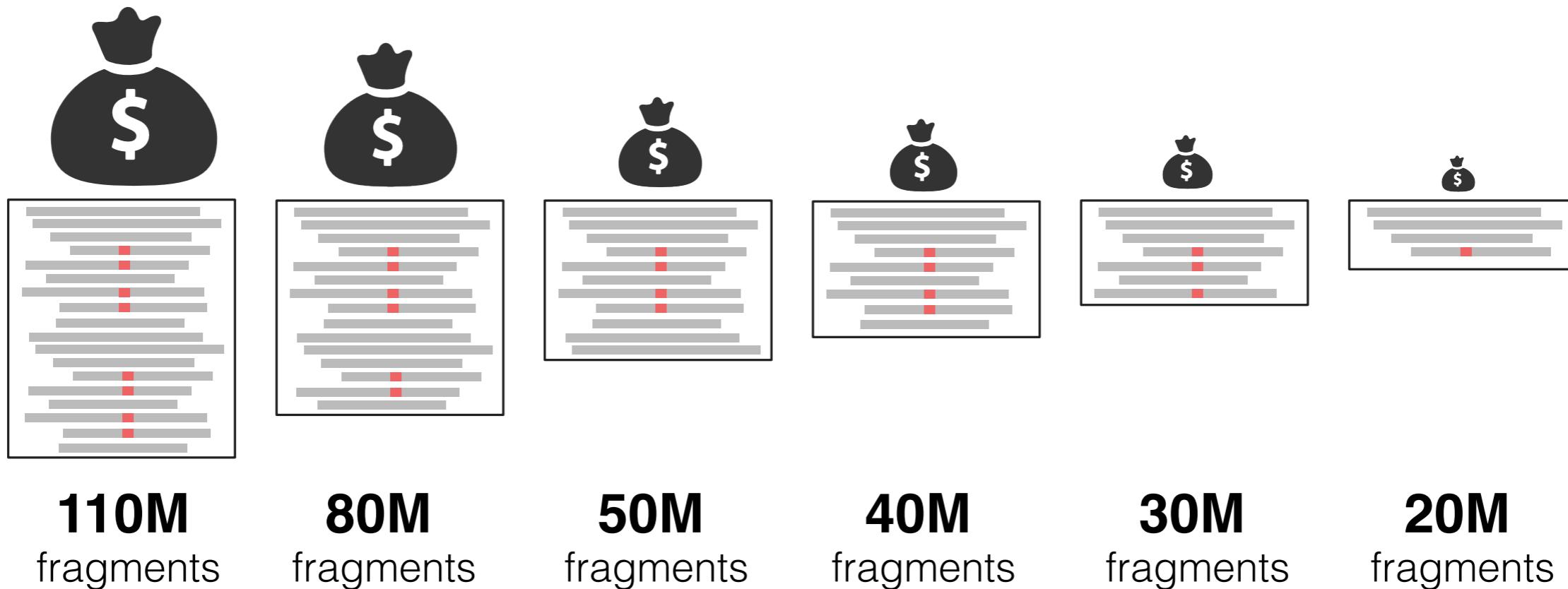


“True” mutations
validated in DNA



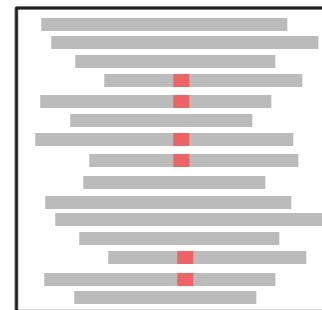
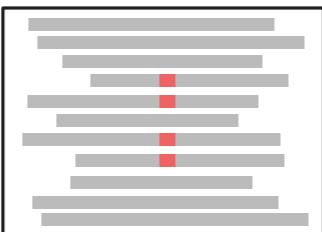
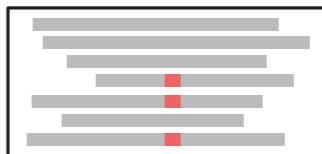
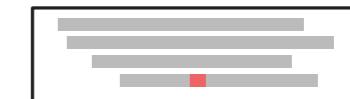
46

RNA-Seq samples

“True” mutations
validated in DNA**6**variant detection
workflows
compared!

46

RNA-Seq samples

“True” mutations
validated in DNA**6**variant detection
workflows
compared!**110M**
fragments**80M**
fragments**50M**
fragments**40M**
fragments**30M**
fragments**20M**
fragments

Final decision

At least 30M fragments!

Workflow choice

Apply different methods to
call different types of mutations

Normal

A C C T A A G G C T A G G T T A C G A

Point
mutation

SNV

A C C T A A G G C T T G G T T A C G A



Normal

ACCTAAGGCTAGGTTACGA

SNV

Point mutation

A C C T A A G G C T T G G T T A C G A

Normal

ACCTAAGGCTAGGTTACGA

Indels

Deletion

ACCTAA - - - - GGTTACGA

Normal

ACCTAAGGCTAGGTTACGA

Insertion

ACCTAAGGCTAGGTTACGA ACTACT

Normal

ACCTAAGGCTAGGTTACGA

Easier!

Point mutation

ACCTAAGGCTTGGTTACGA

Normal

ACCTAAGGCTAGGTTACGA

Deletion

ACCTAA - - - - GGTTACGA

More headaches!

Normal

ACCTAAGGCTAGGTTACGA

Insertion

ACCTAAGGCTAGGTTACGA ACTACT

Different software

Different software



Huge
headaches!

Different output formats



**Standardise + Combine
variant outputs**



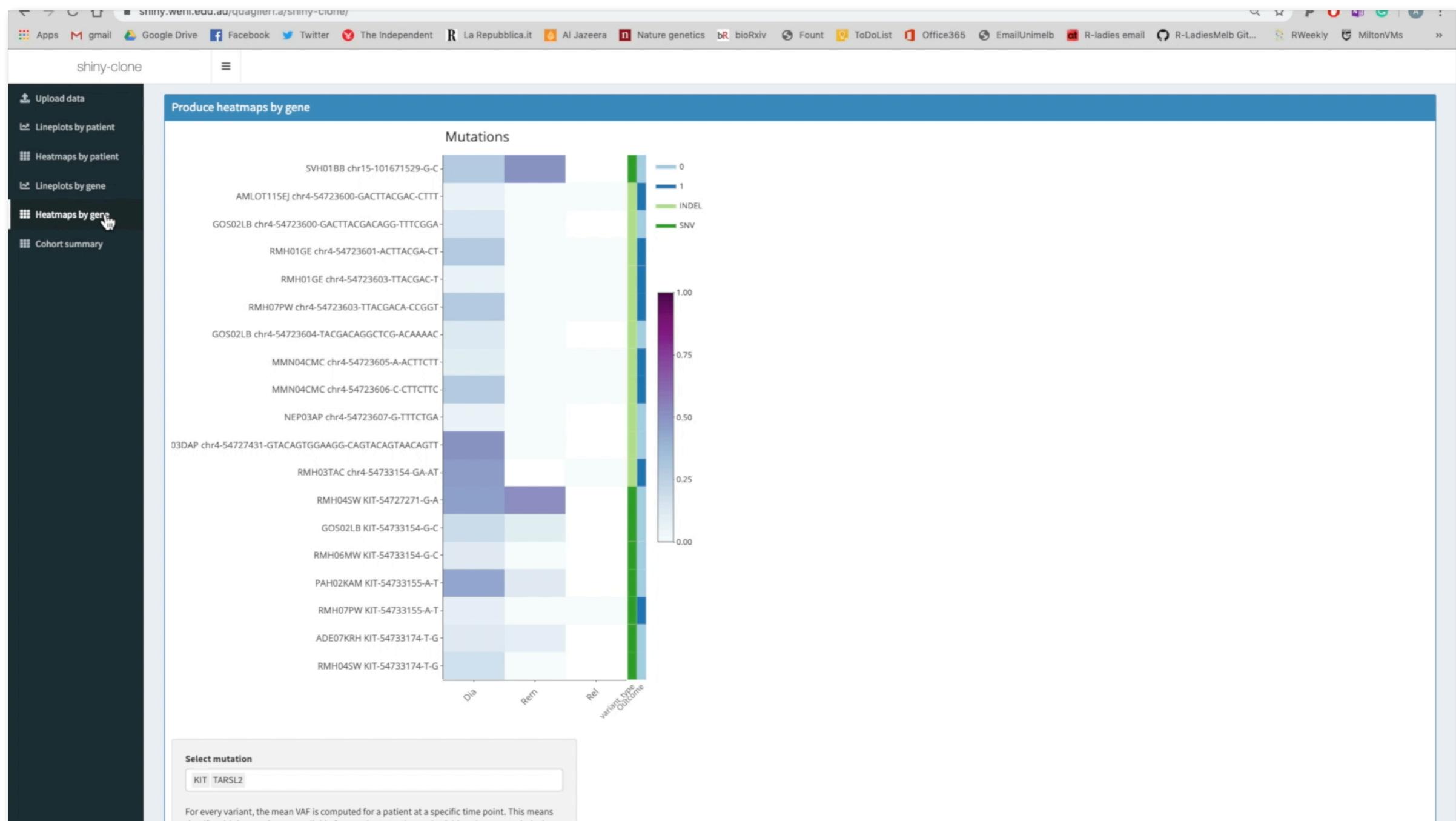
varikondo R

annaquagliari16.github.io/varikondo/





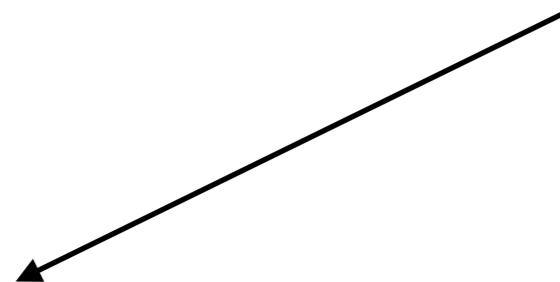
Finally spark joy!



<https://shiny.wehi.edu.au/quagliari.a/shiny-clone/>



Experimental design



30M fragments



Experimental design

30M fragments



Methods for **SNV**
and for **indels**

varikondo R

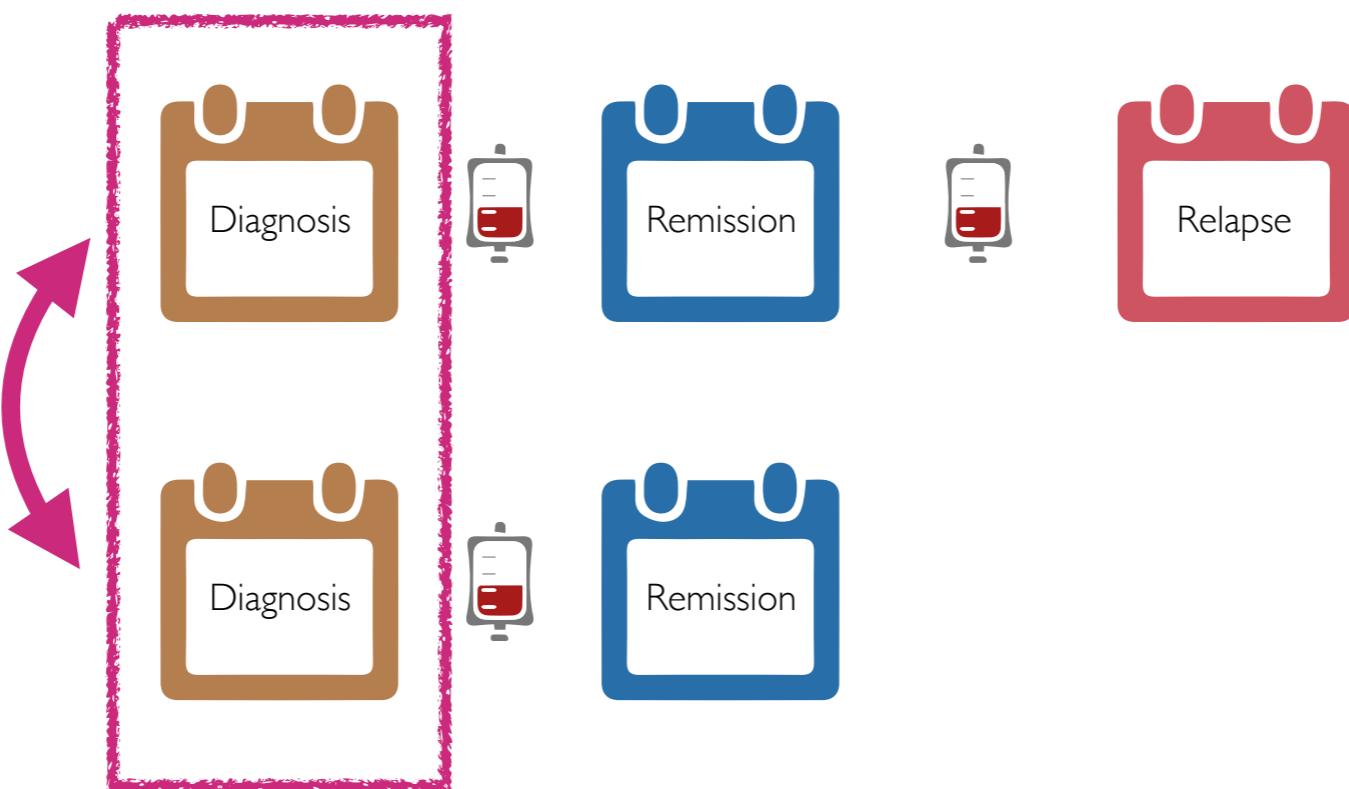
The logo consists of the word 'varikondo' in a black sans-serif font, followed by a blue stylized 'R' character enclosed within a grey oval shape.



Experimental design



Compare patients
at diagnosis





Experimental design



Compare patients
at diagnosis

Mutation



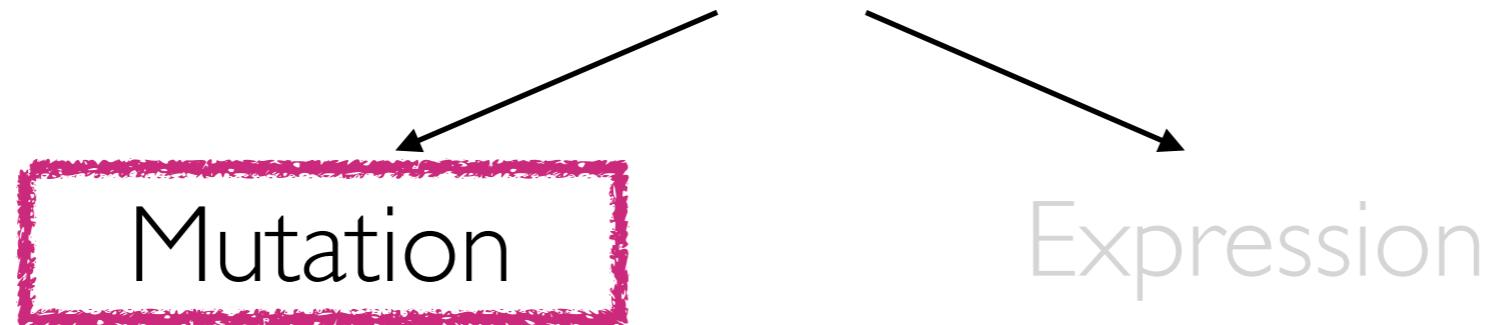
Expression



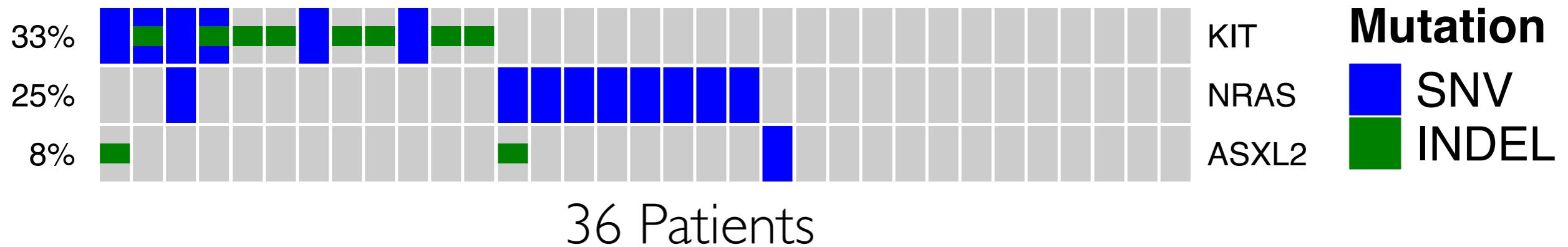
Experimental design



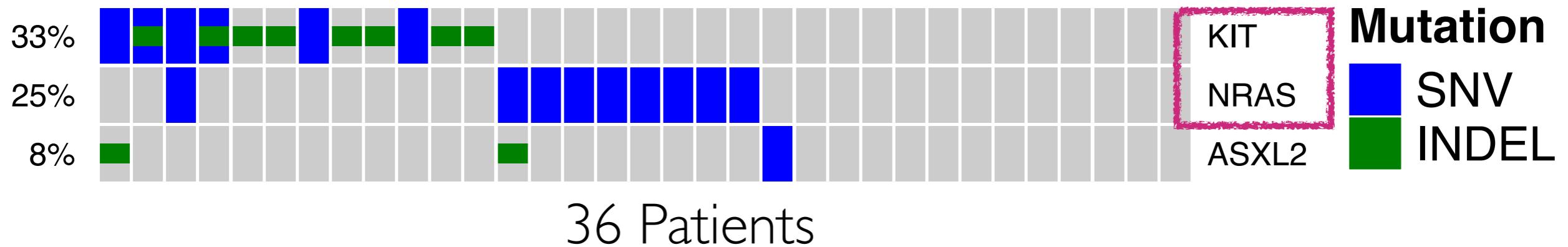
Compare patients
at diagnosis



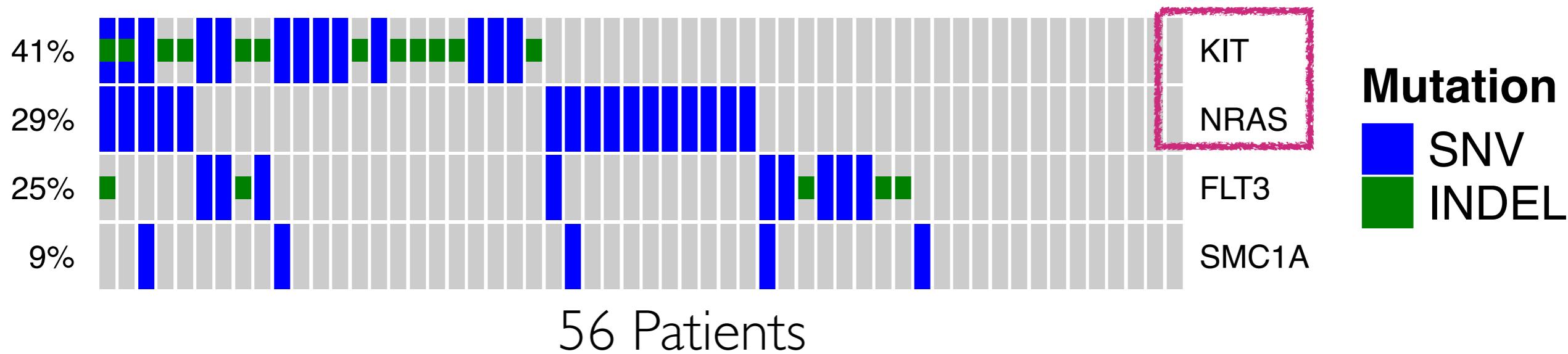
Top mutated AML genes in our data



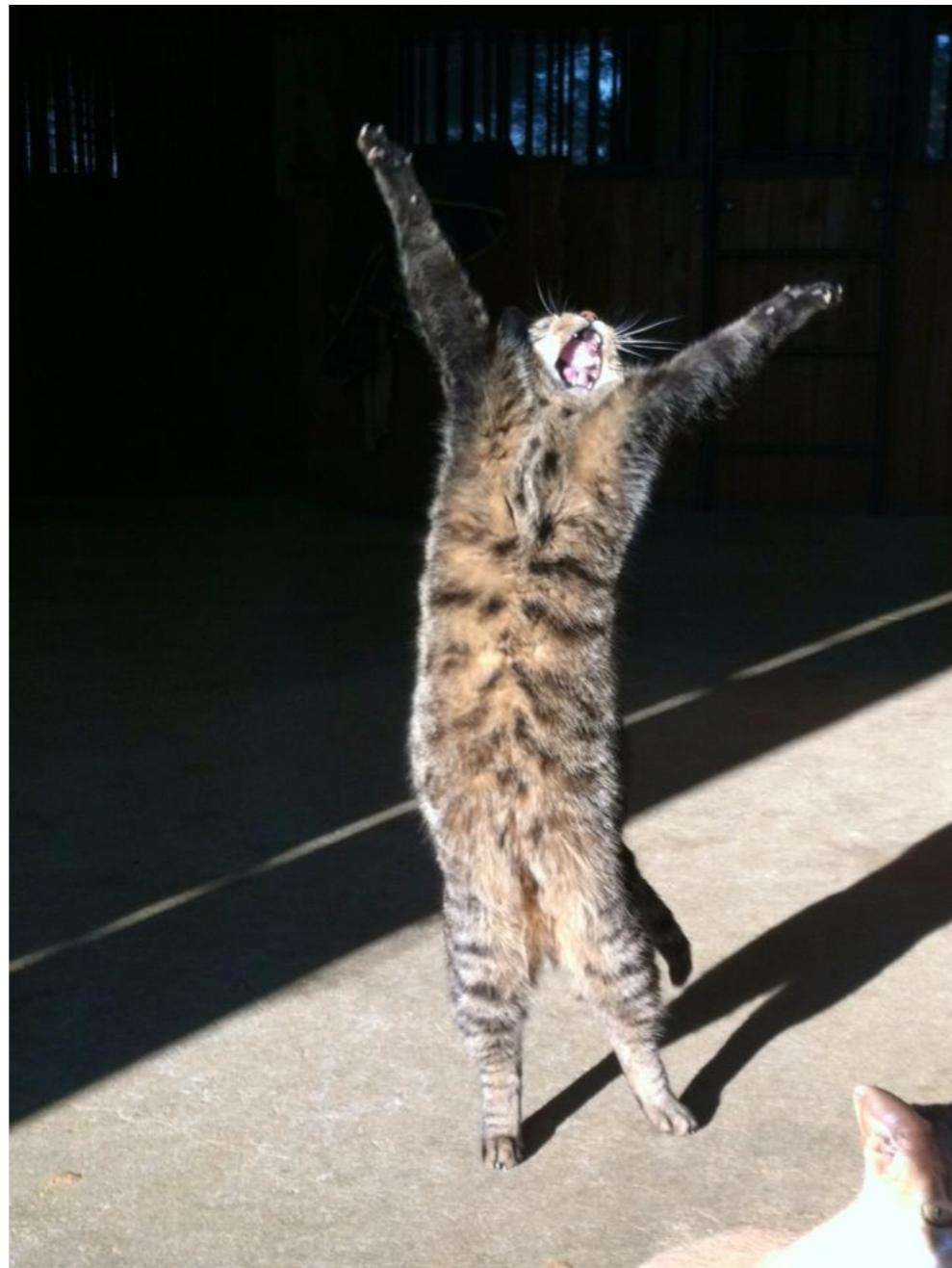
Top mutated AML genes in our data



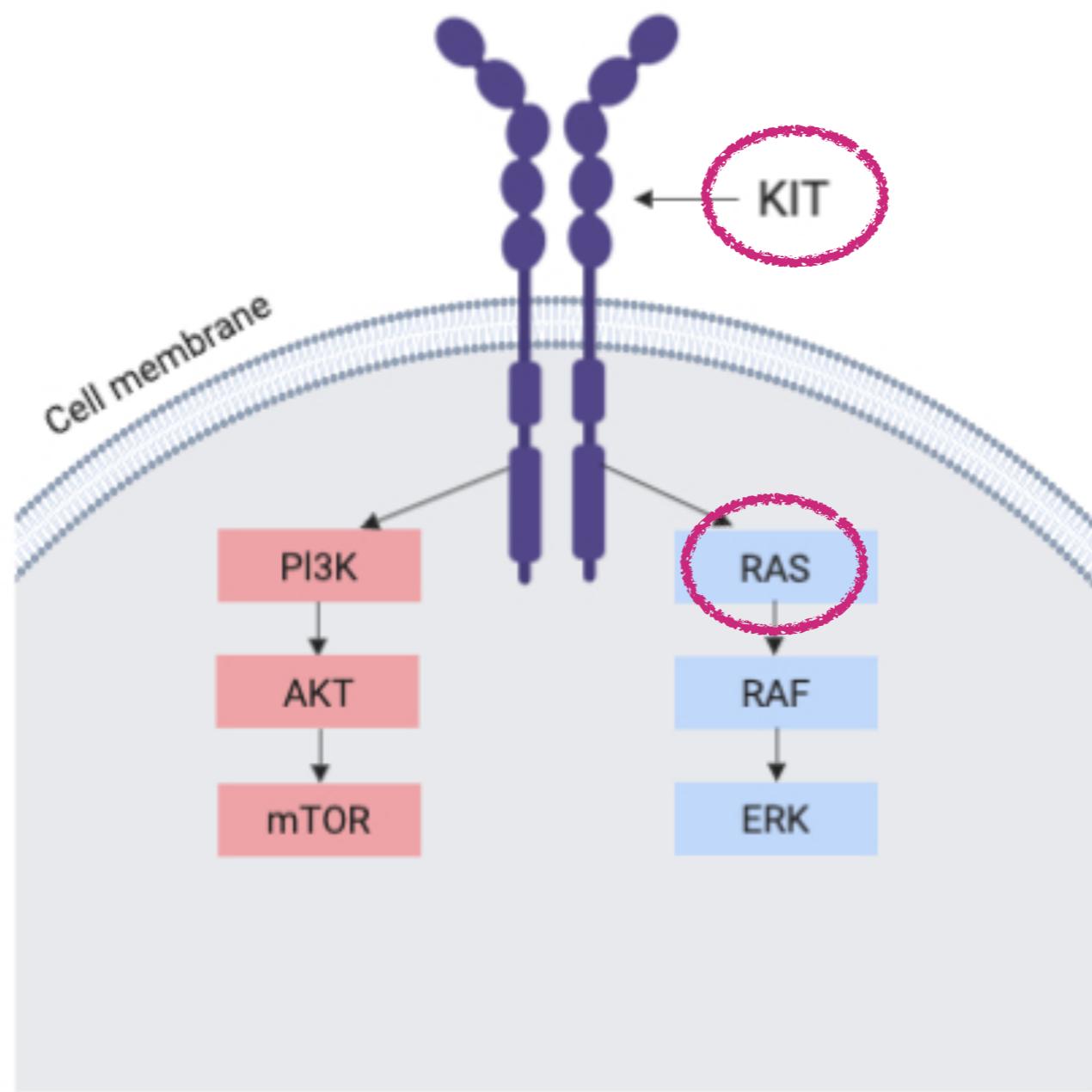
Published cohorts (from DNA)



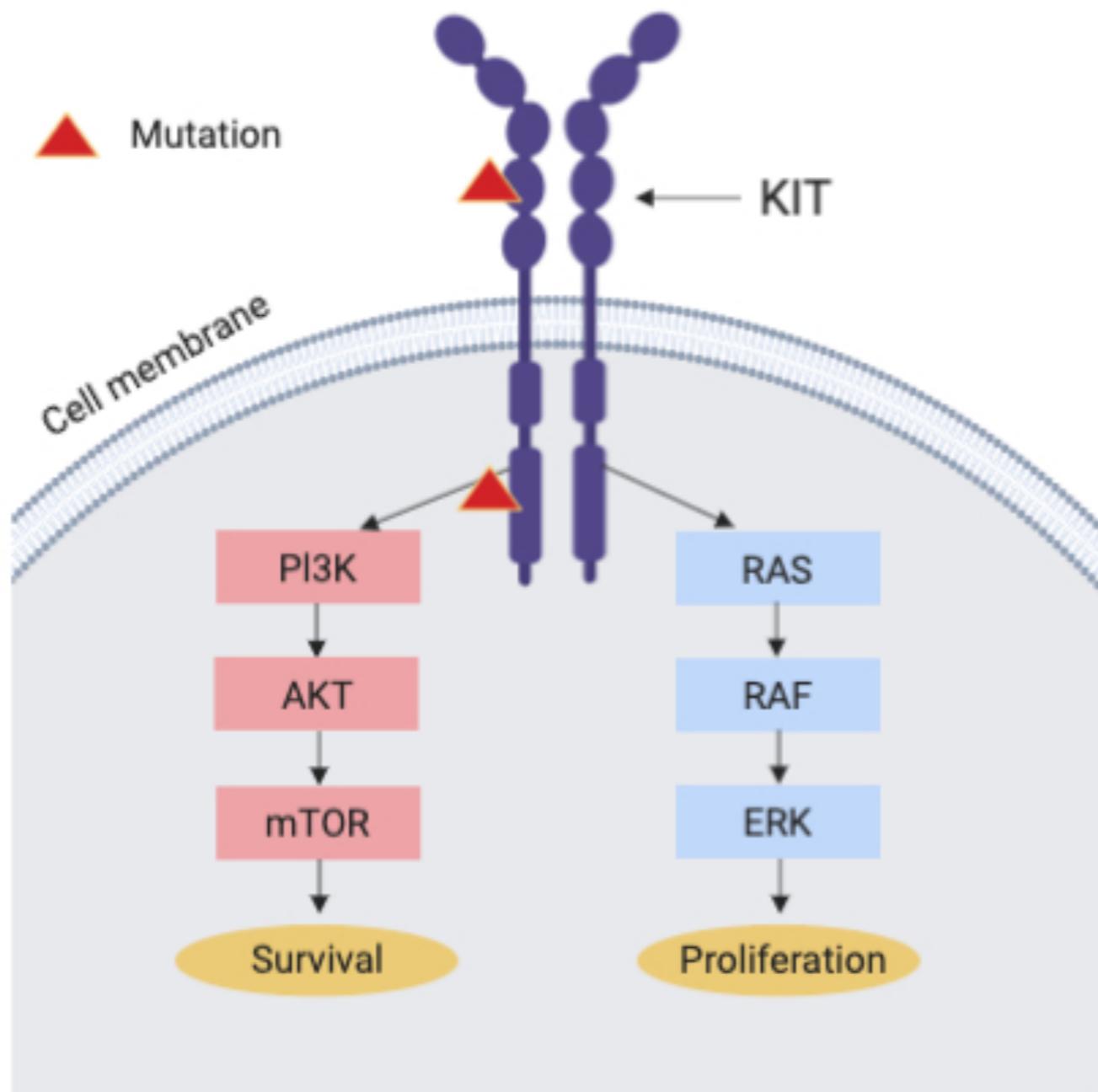
Concordance with published
results using RNA only!

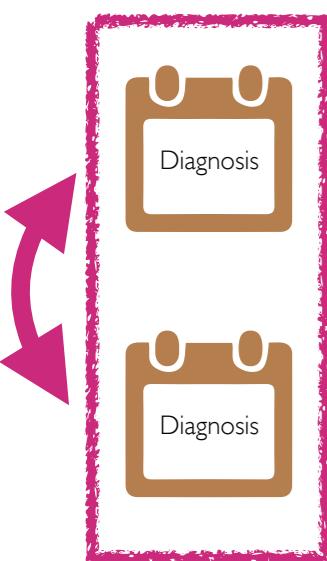


RAS pathway

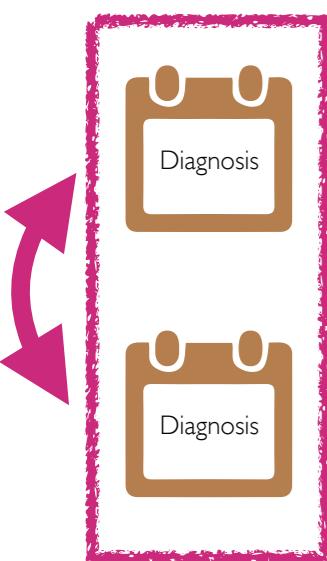


KIT activation





Compare mutation frequency

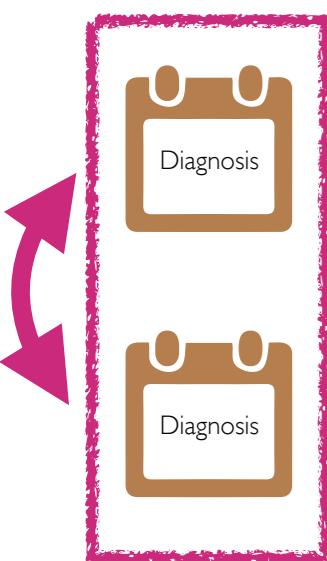


Compare mutation frequency

Relapse (n=9) No relapse (n=27)

KIT

NRAS



Compare mutation frequency

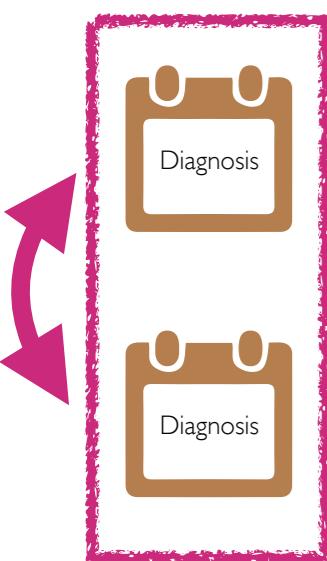
Relapse (n=9) No relapse (n=27)

KIT

5 (55%)

7 (26%)

NRAS



Compare mutation frequency

Relapse (n=9) No relapse (n=27)

KIT

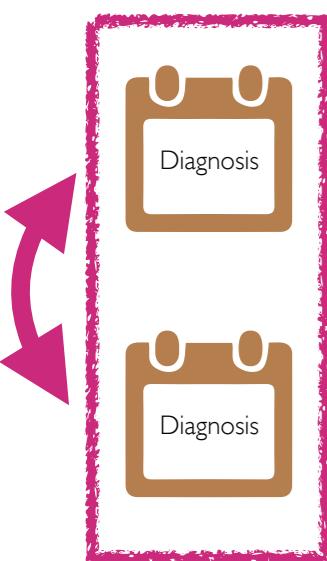
5 (55%)

7 (26%)

NRAS

1 (11%)

8 (30%)



Compare mutation frequency

Relapse (n=9) No relapse (n=27)

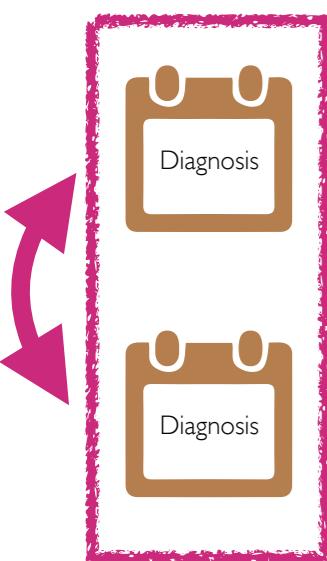
KIT

5 (55%)

7 (26%)



Only inv(16)



Compare mutation frequency

Relapse (n=9) No relapse (n=27)

KIT

5 (55%)

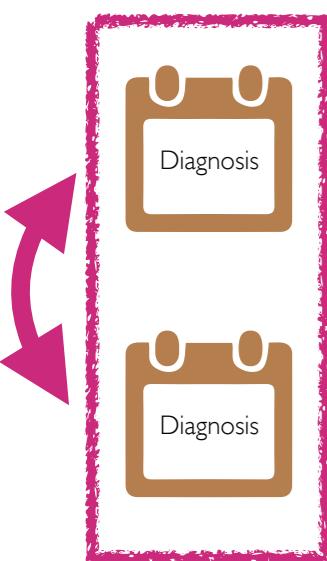
7 (26%)



Only inv(16)



exon 8



Compare mutation frequency

Relapse (n=9) No relapse (n=27)

KIT

5 (55%)

7 (26%)



Only inv(16)



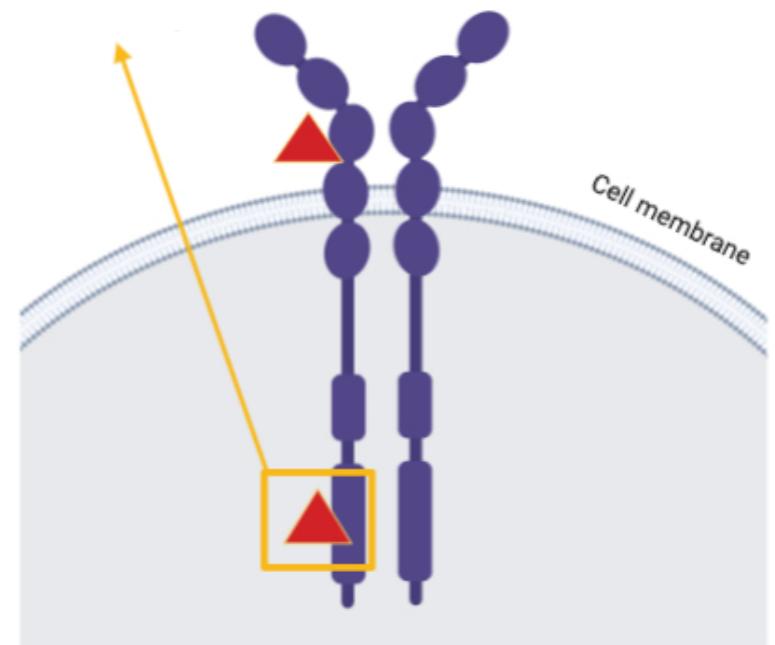
exon 8



exon 17 (5)
exon 8 (2)

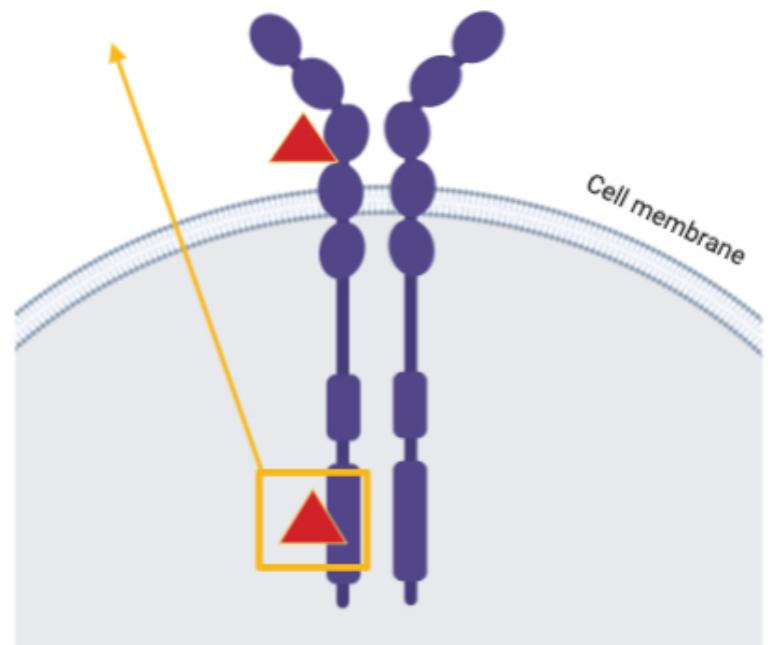
exon 17
Common KIT hot spot

Exon 17 SNVs



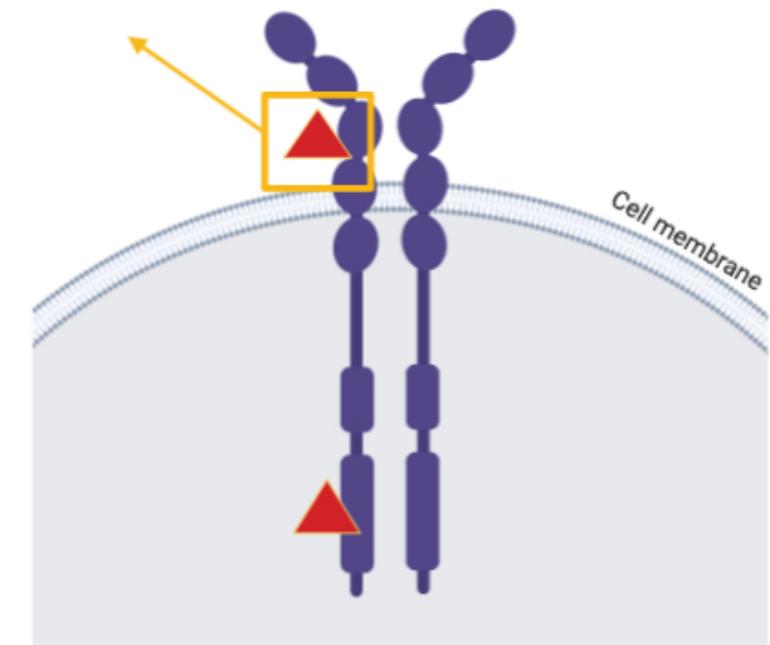
exon 17
Common KIT hot spot

Exon 17 SNVs



exon 8
Not usually screened

Exon 8 indels



AML with CBFB–MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions

C Haferlach , F Dicker, A Kohlmann, S Schindela, T Weiss, W Kern, S Schnittger & T Haferlach

Leukemia **24**, 1065–1069 (2010) | [Download Citation](#) 

- ~ 90% of inv(16) patients harbour mutations in the RAS pathway

AML with CBFB–MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions

C Haferlach , F Dicker, A Kohlmann, S Schindela, T Weiss, W Kern, S Schnittger & T Haferlach

Leukemia **24**, 1065–1069 (2010) | [Download Citation](#) 

- ~ 90% of inv(16) patients harbour mutations in the RAS pathway
- NRAS mutant has been associated with better survival and lower risk of relapse

AML with CBFB–MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions

C Haferlach , F Dicker, A Kohlmann, S Schindela, T Weiss, W Kern, S Schnittger & T Haferlach

Leukemia **24**, 1065–1069 (2010) | [Download Citation](#) 

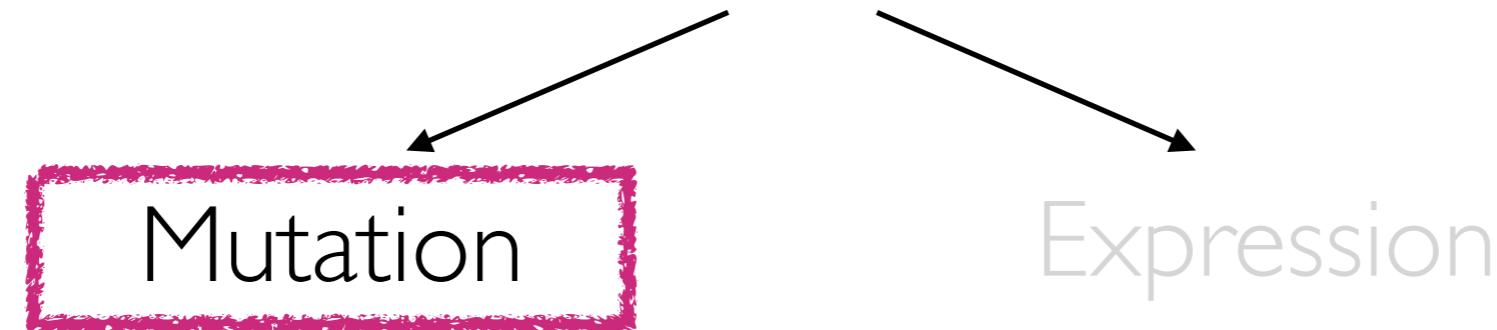
- ~ 90% of inv(16) patients harbour **mutations** in the **RAS pathway**
- **NRAS mutant** has been associated with **better survival** and **lower risk of relapse**
- **KIT mutant** patients have a **higher risk of relapse**, even though findings are **controversial**



Experimental design



Compare patients
at diagnosis



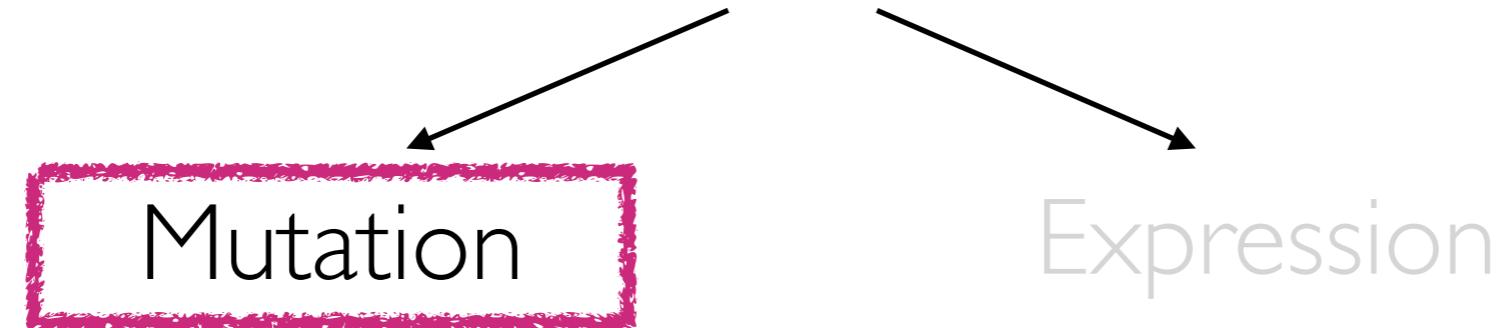
- KIT more mutated in patients who relapse



Experimental design



Compare patients
at diagnosis



- KIT more mutated in patients who relapse
- All KITmut patients who relapse are mutated on exon 8



Experimental design



Compare patients
at diagnosis

Mutation

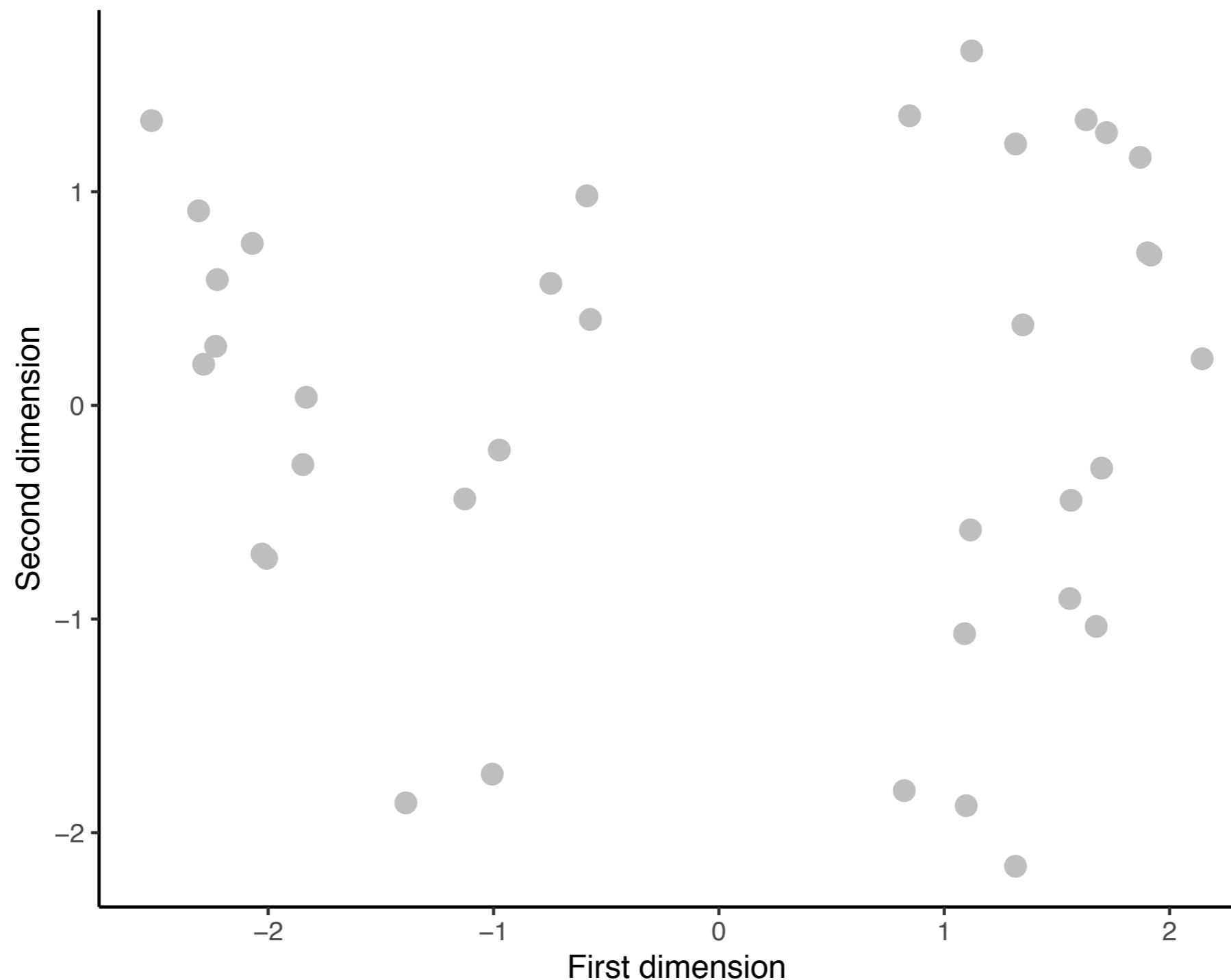
Expression



- KIT more mutated in patients who relapse
- All KITmut patients who relapse are mutated on exon 8

Summarise gene information

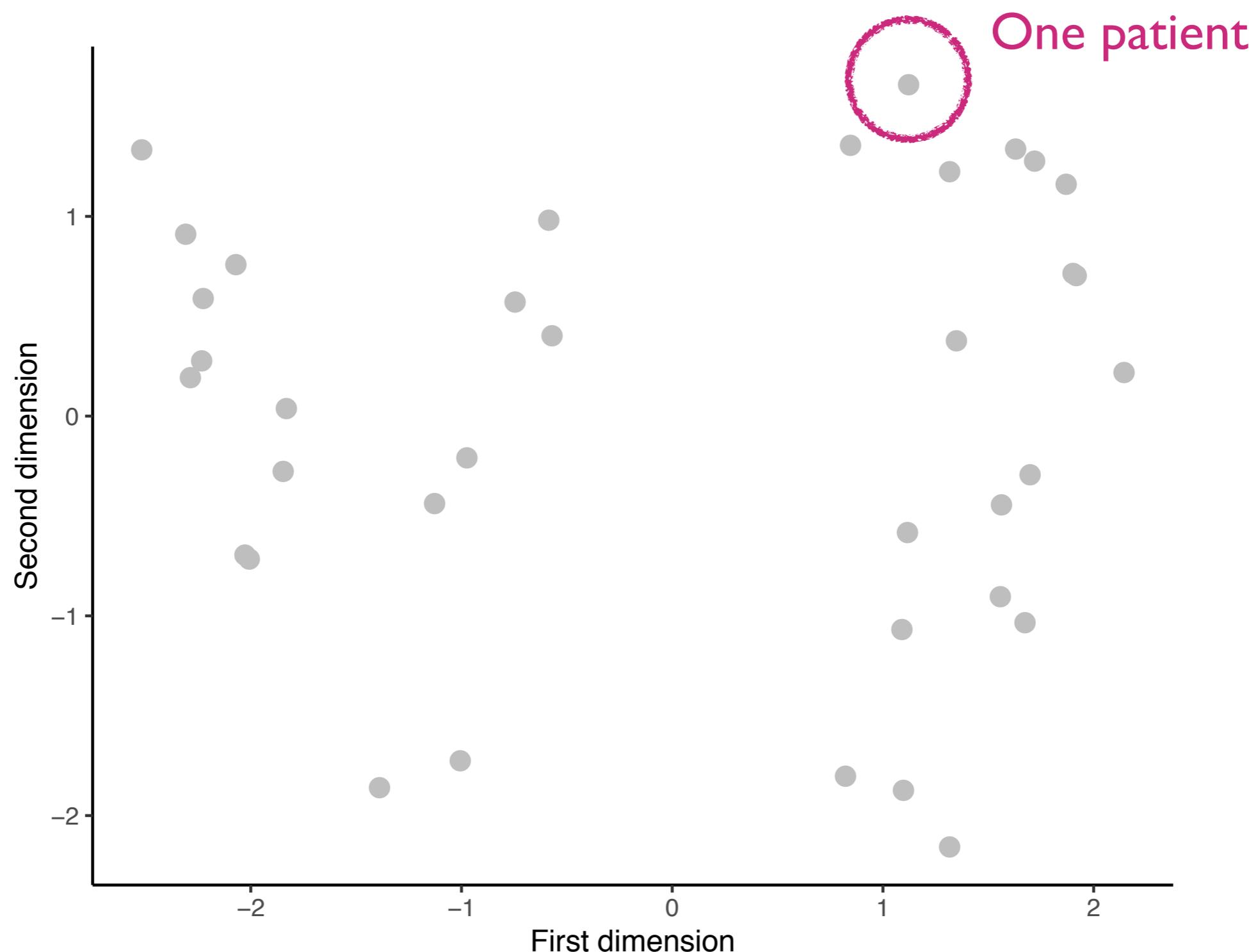
MDS plots



Diagnosis

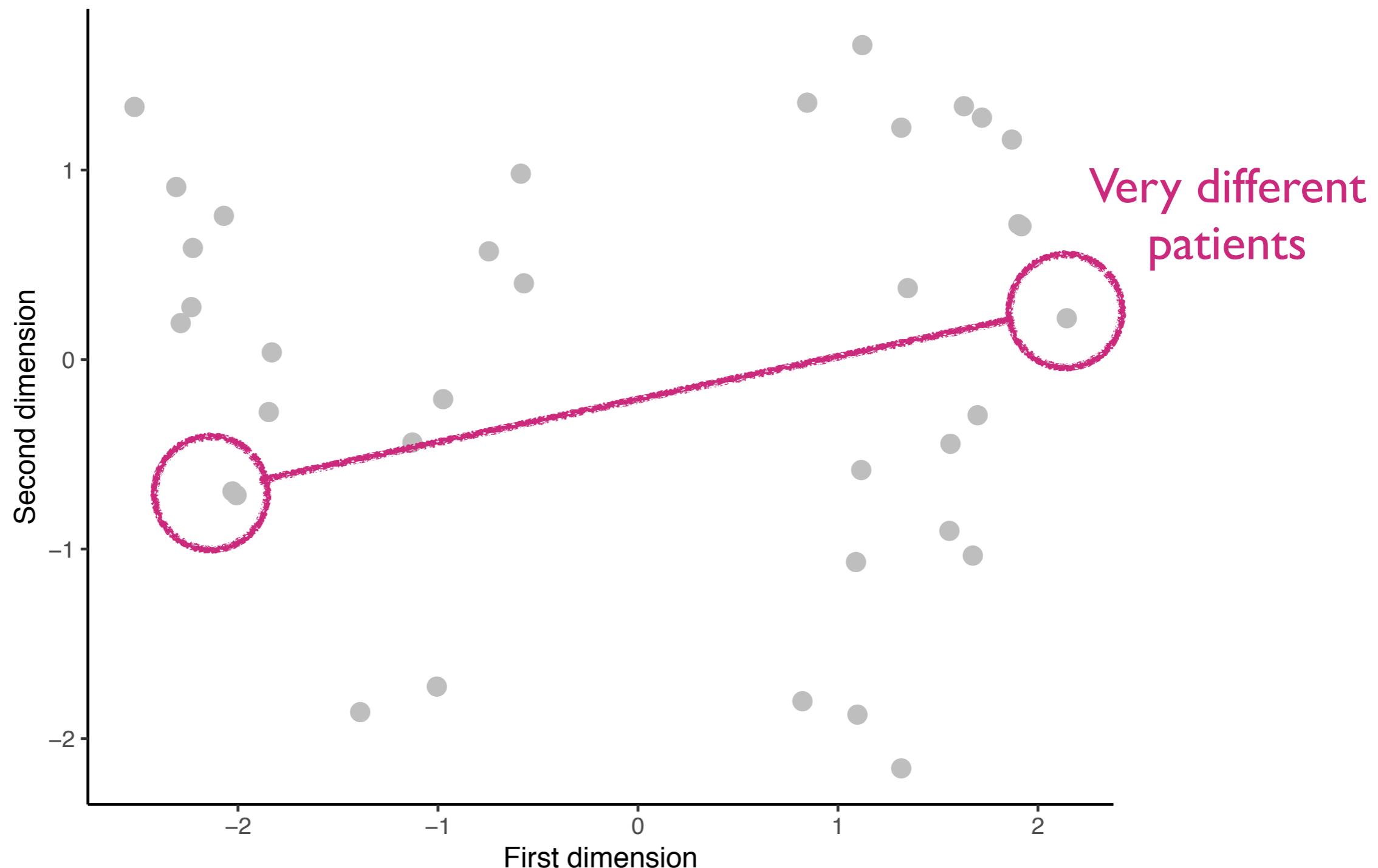
Summarise gene information

MDS plots



Summarise gene information

MDS plots

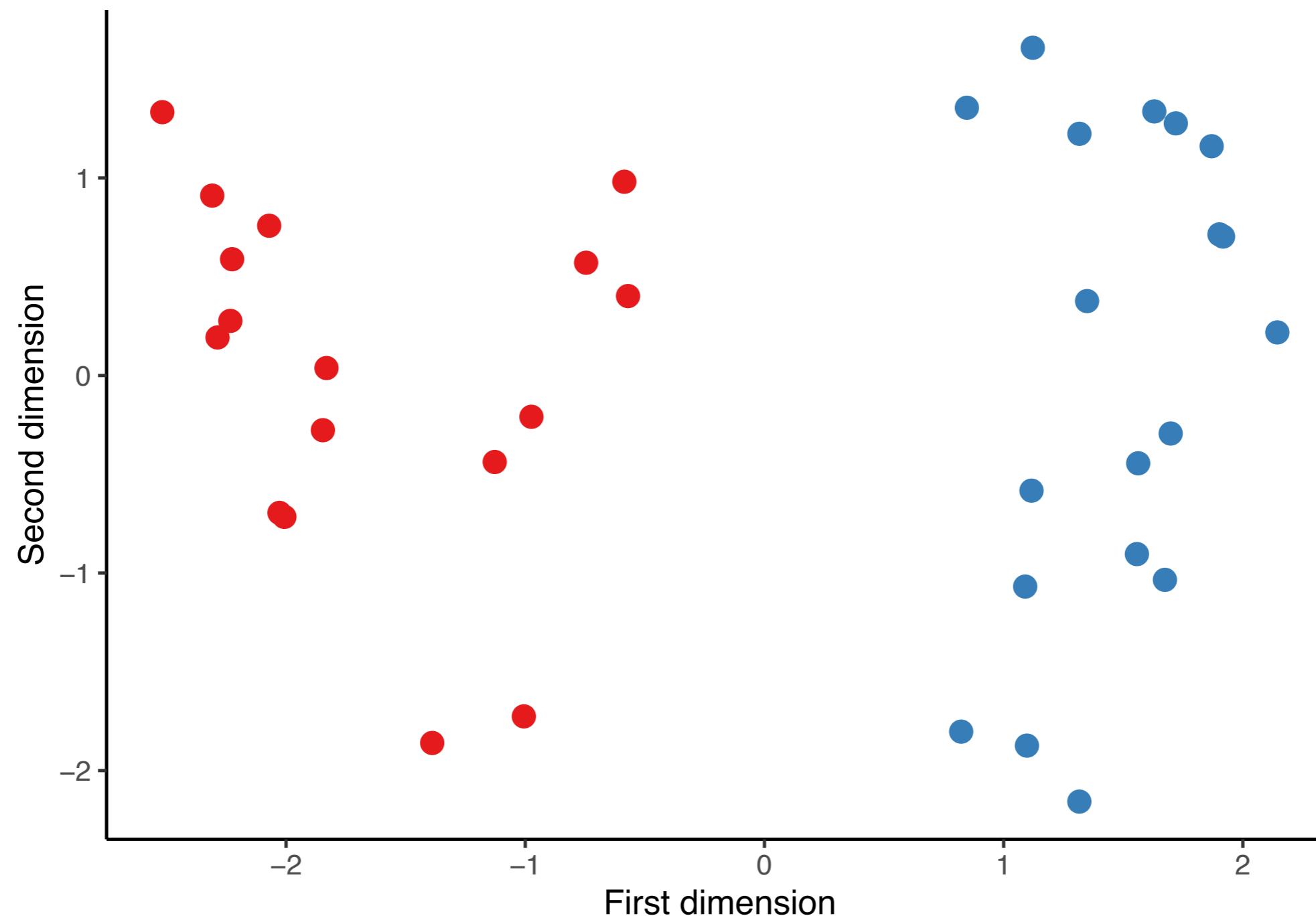


Diagnosis

CBF AML Type

inv(16)

t(8;21)



Diagnosis

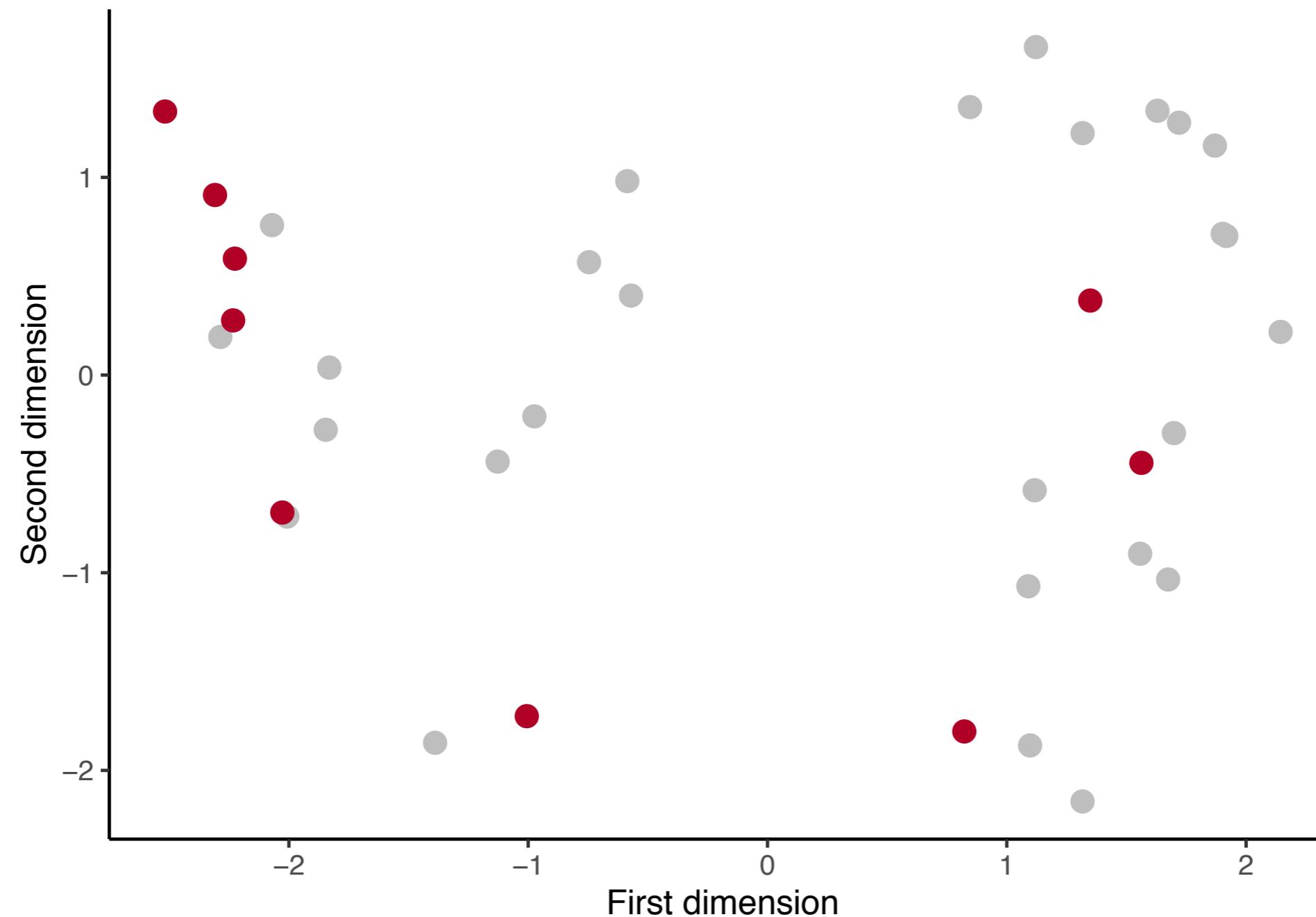
Outcome

No relapse

Relapse

Inv(16)

t(8;21)

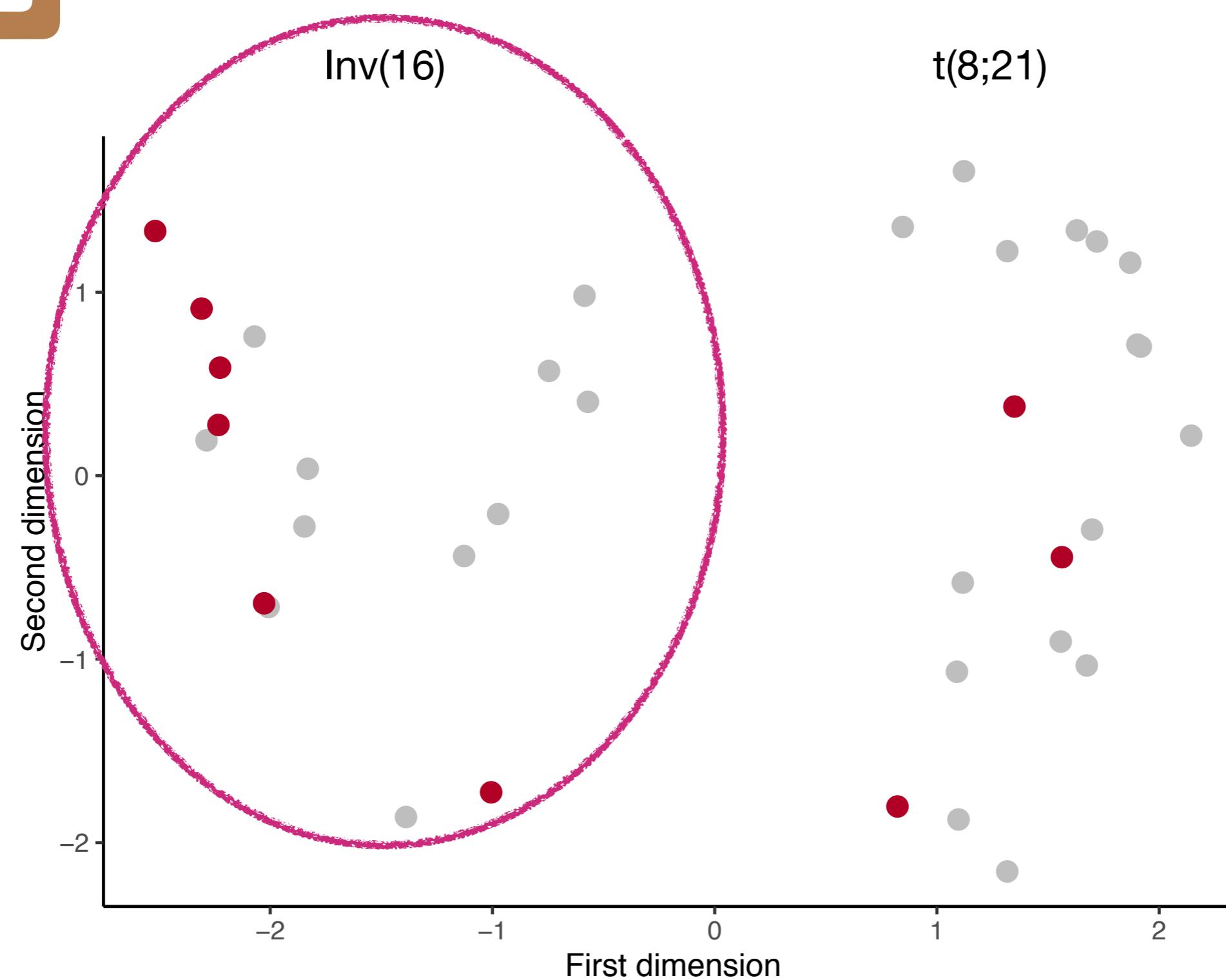




Outcome

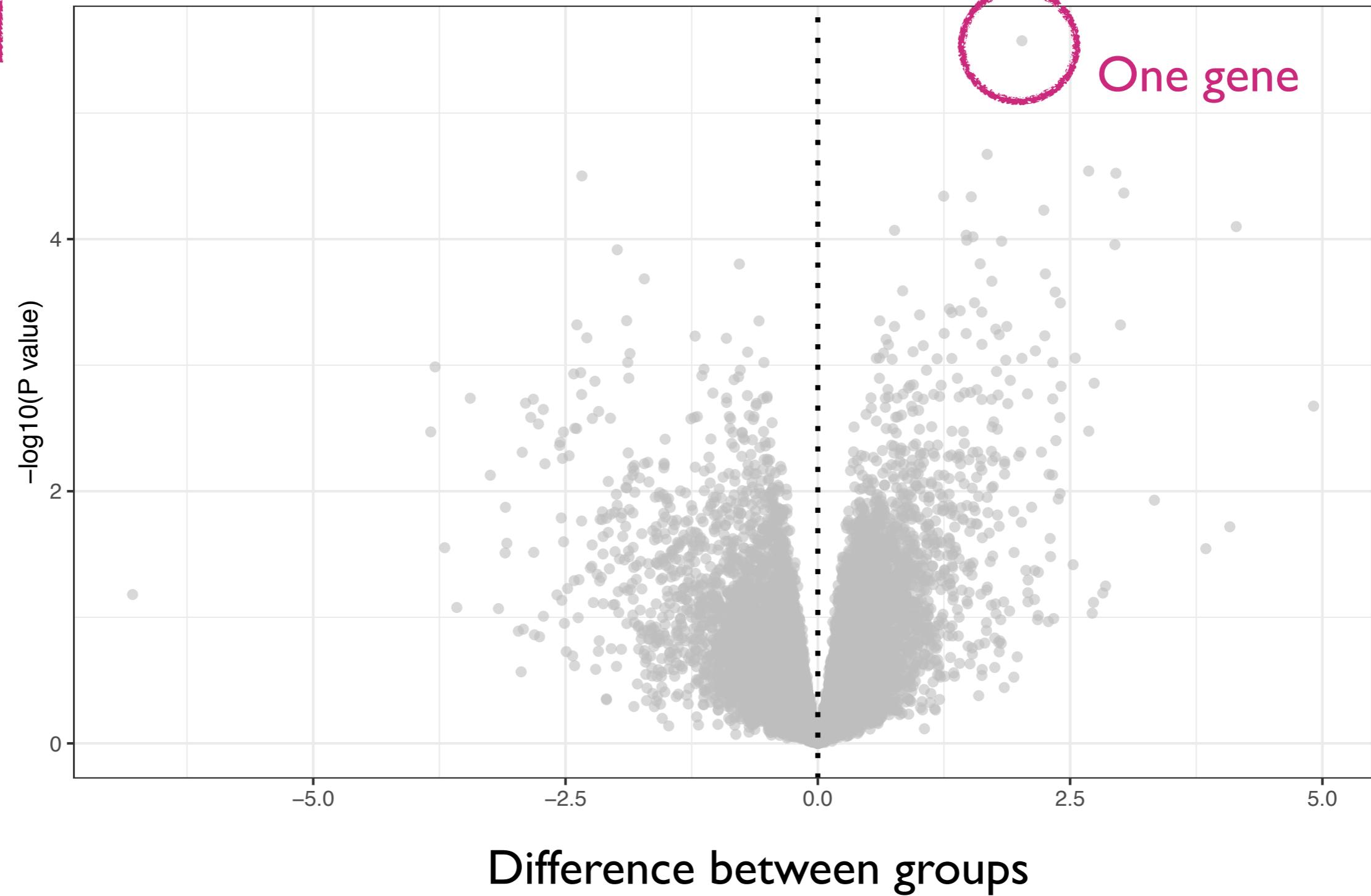
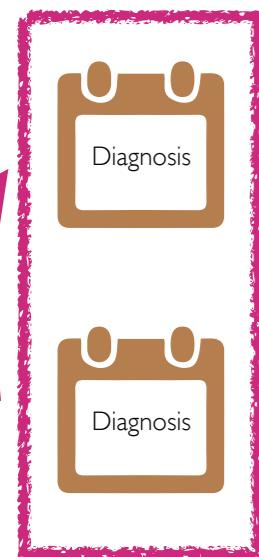
No relapse

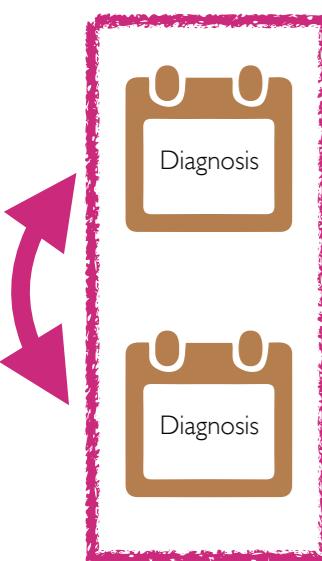
Relapse



Compare expression - $\text{inv}(16)$

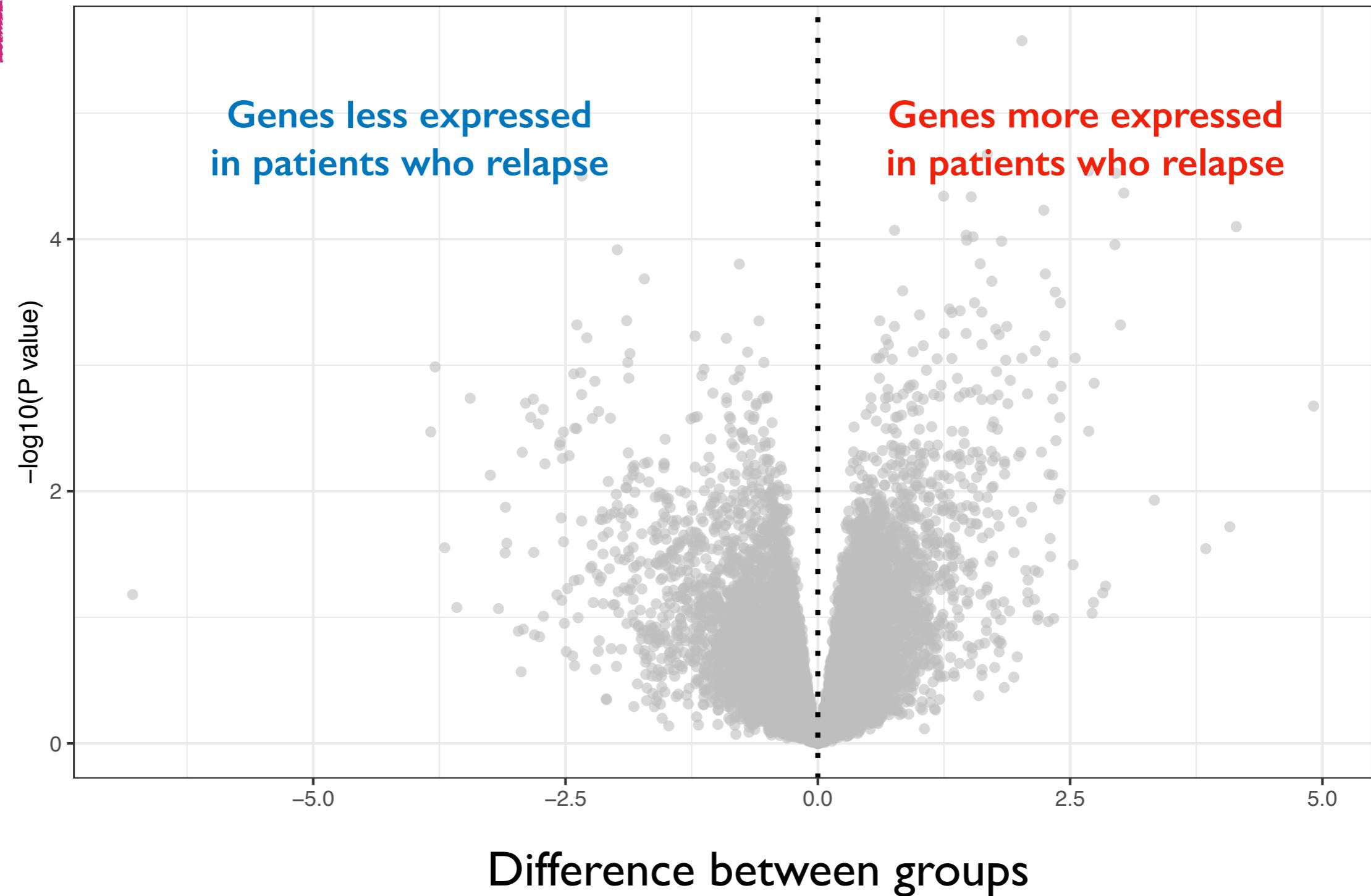
Relapse ($n=6$) vs No relapse ($n=11$)

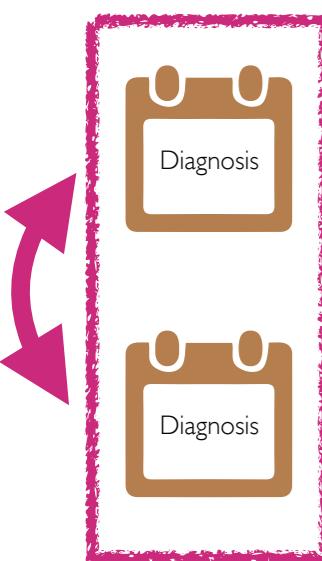




Compare expression - $\text{inv}(16)$

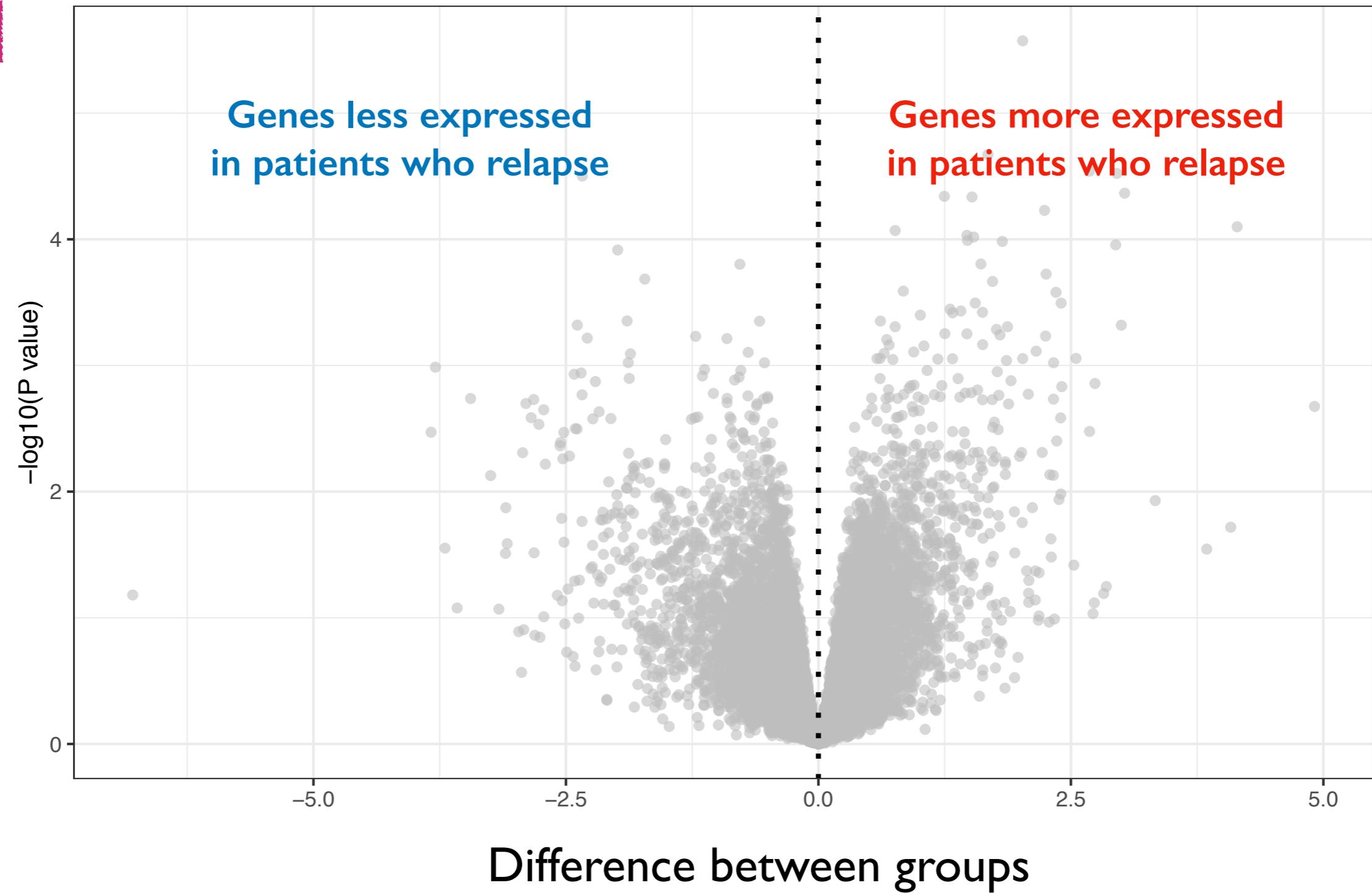
Relapse ($n=6$) vs No relapse ($n=11$)





No genes significantly different

FDR 5%

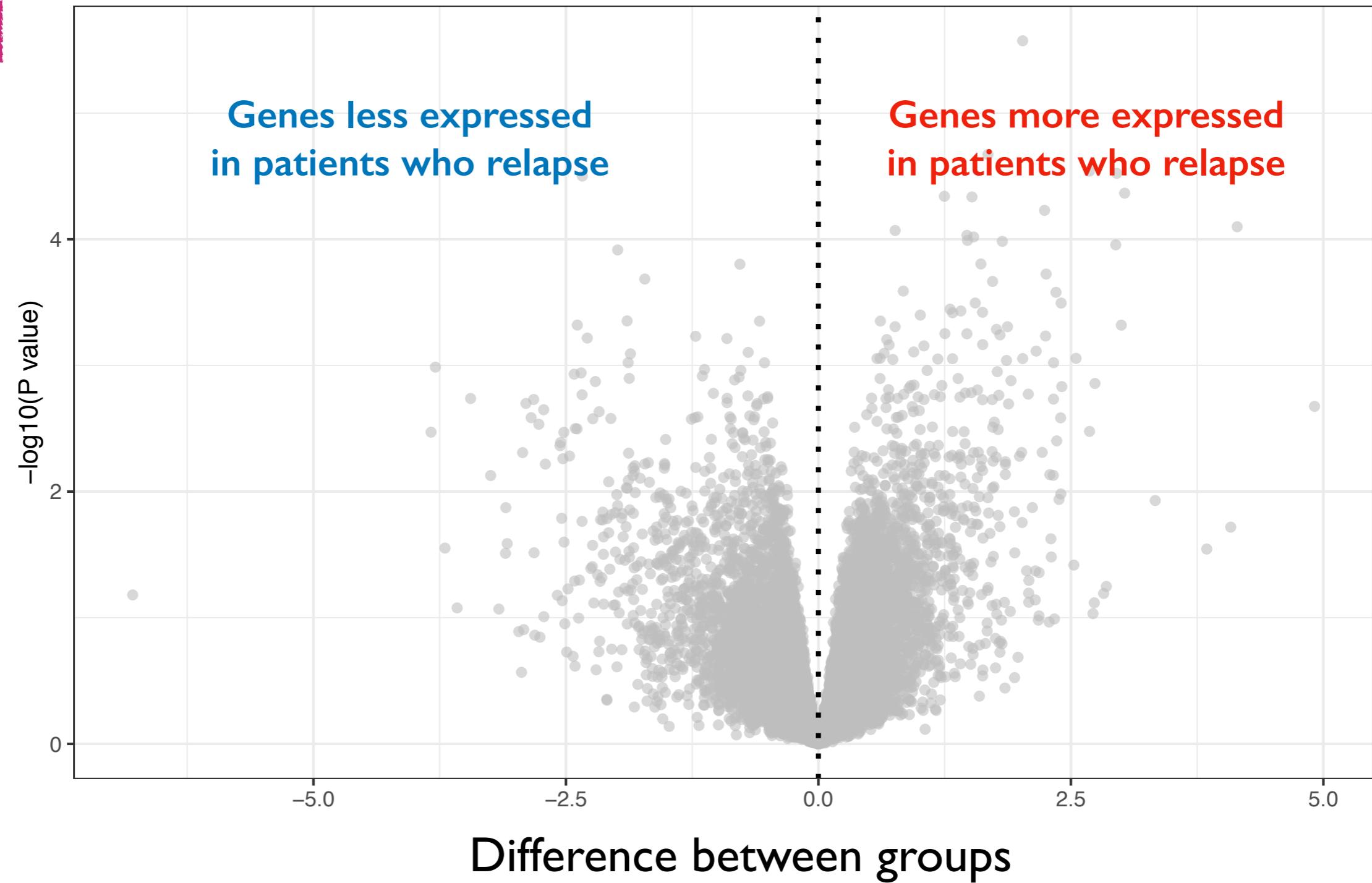


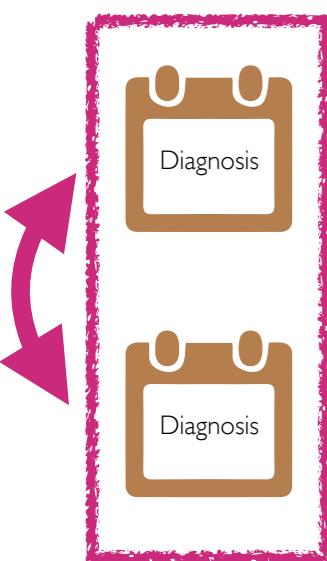
Published CBF AML gene set

Lück SC, 2010
British Journal of Haematology

Upregulated when KIT is mutated

Downregulated when KIT is mutated





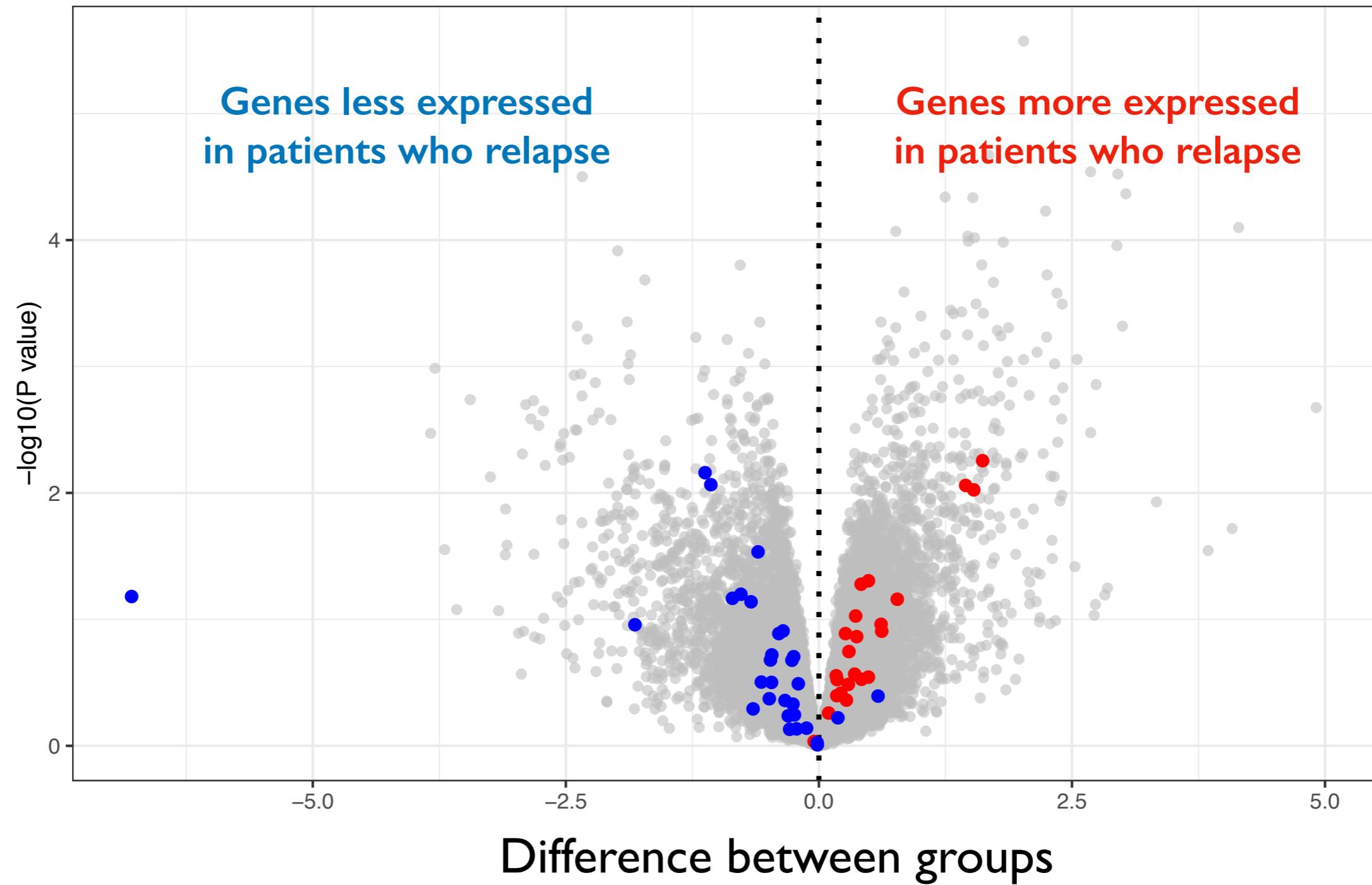
Published CBF AML gene set

Lück SC, 2010
British Journal of Haematology



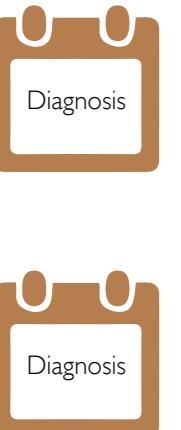
Upregulated when KIT is mutated

Downregulated when KIT is mutated



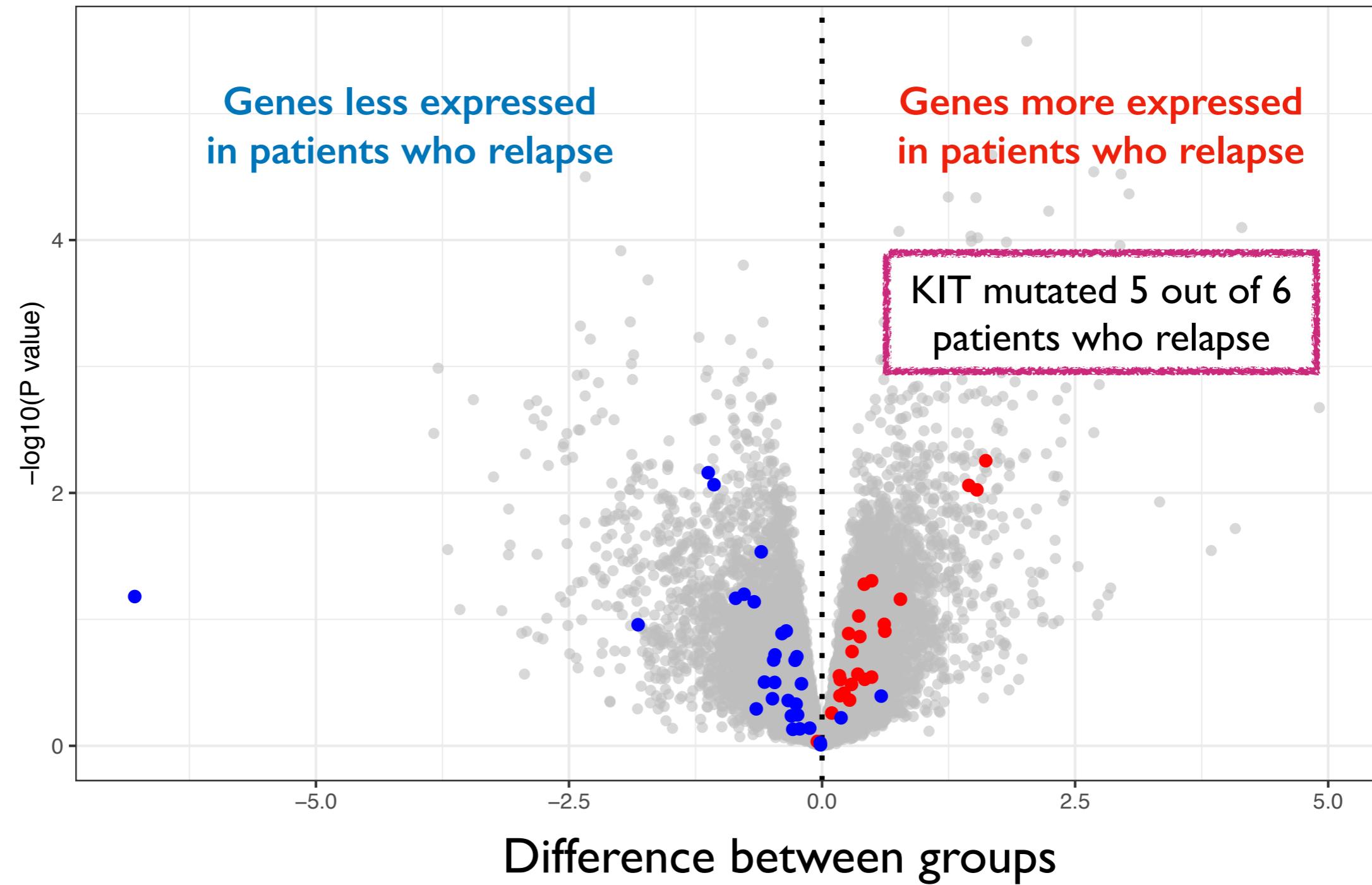
Published CBF AML gene set

Lück SC, 2010
British Journal of Haematology



Upregulated when KIT is mutated

Downregulated when KIT is mutated





Experimental design



Compare patients
at diagnosis

Mutation



Expression

- KIT more mutated in patients who relapse
- All KITmut patients who relapse are mutated on exon 8

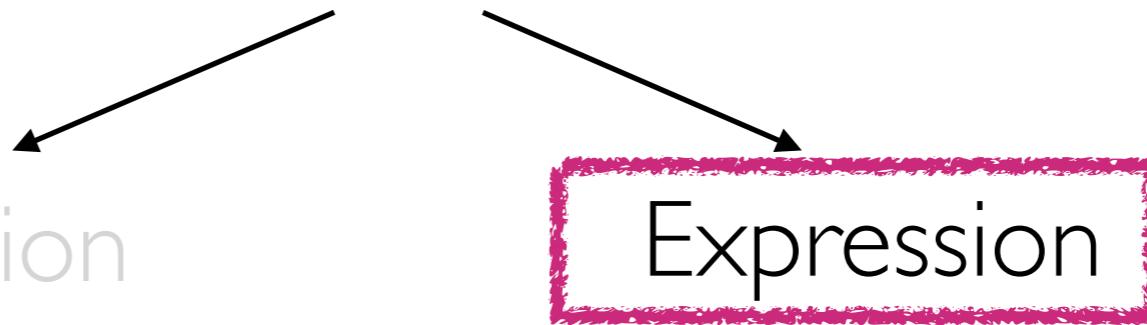
- No significantly changing genes - inv(16)



Experimental design



Compare patients at diagnosis



- KIT more mutated in patients who relapse
- All KITmut patients who relapse are mutated on exon 8

- No significantly changing genes - inv(16)
- KIT mutation affects gene expression - inv(16)



Experimental design



Compare patients at diagnosis

Mutation



Expression

- KIT more mutated in patients who relapse
- All KITmut patients who relapse are mutated on exon 8

- No significantly changing genes - inv(16)
- KIT mutation affects gene expression - inv(16)
- No significantly changing genes combined CBF type



Experimental design



Compare patients
at diagnosis



Changes over time





Experimental design



Compare patients
at diagnosis

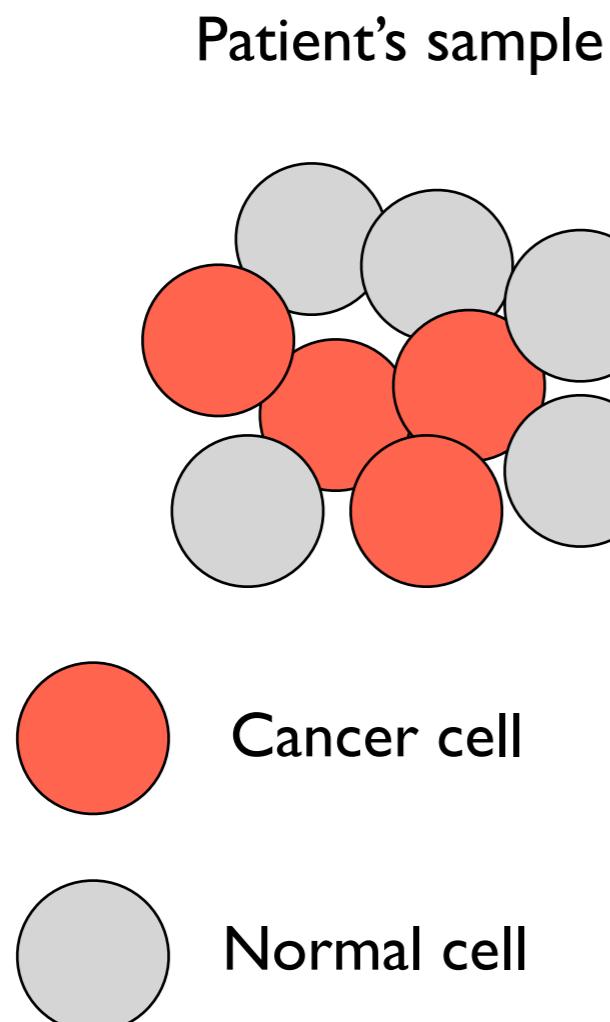


Changes over time

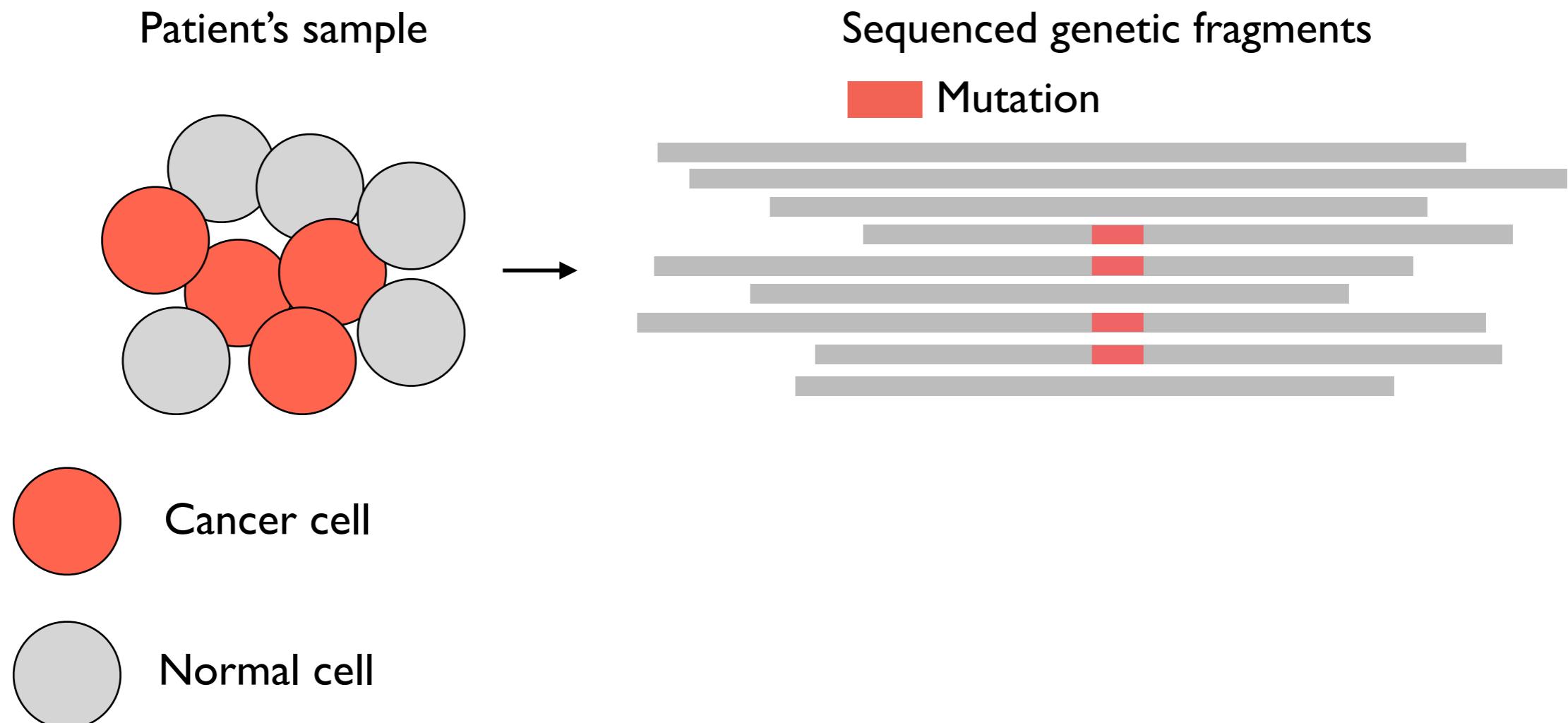
Mutation

Variant Allele Frequency (VAF)

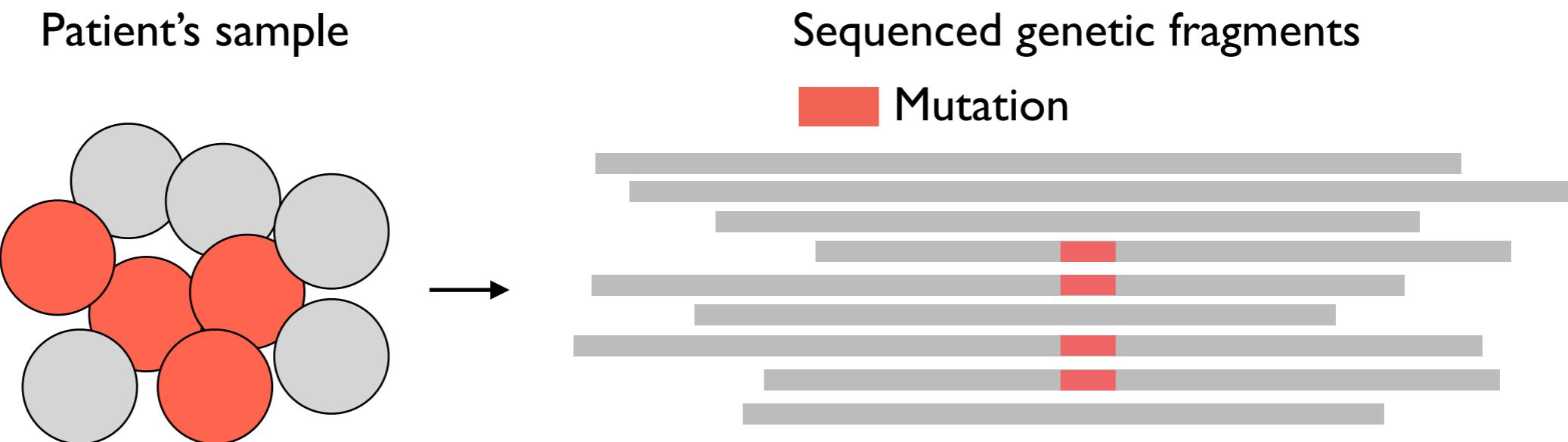
Variant Allele Frequency (VAF)



Variant Allele Frequency (VAF)

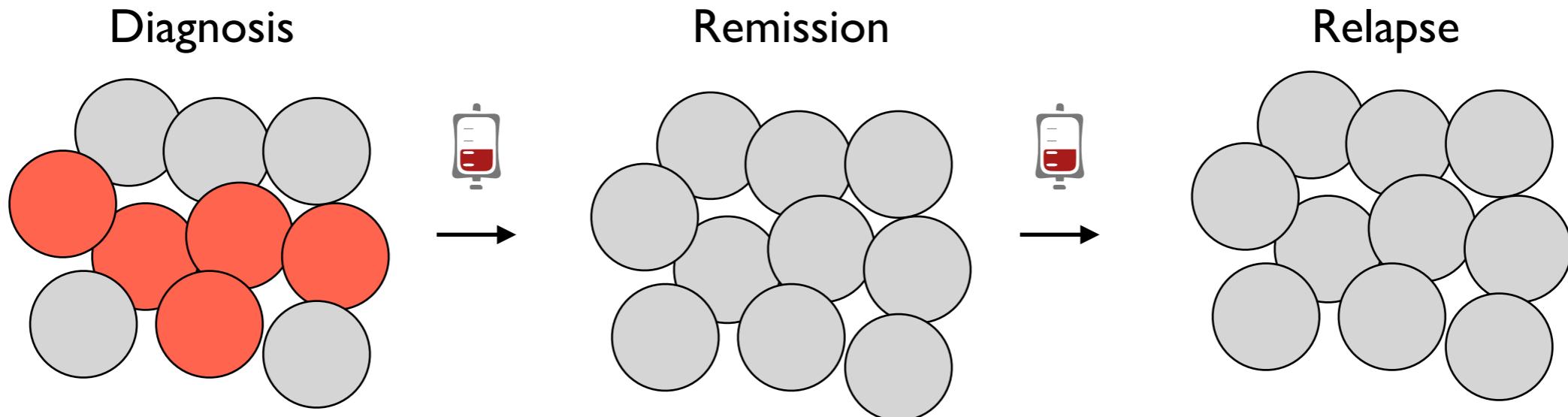


Variant Allele Frequency (VAF)

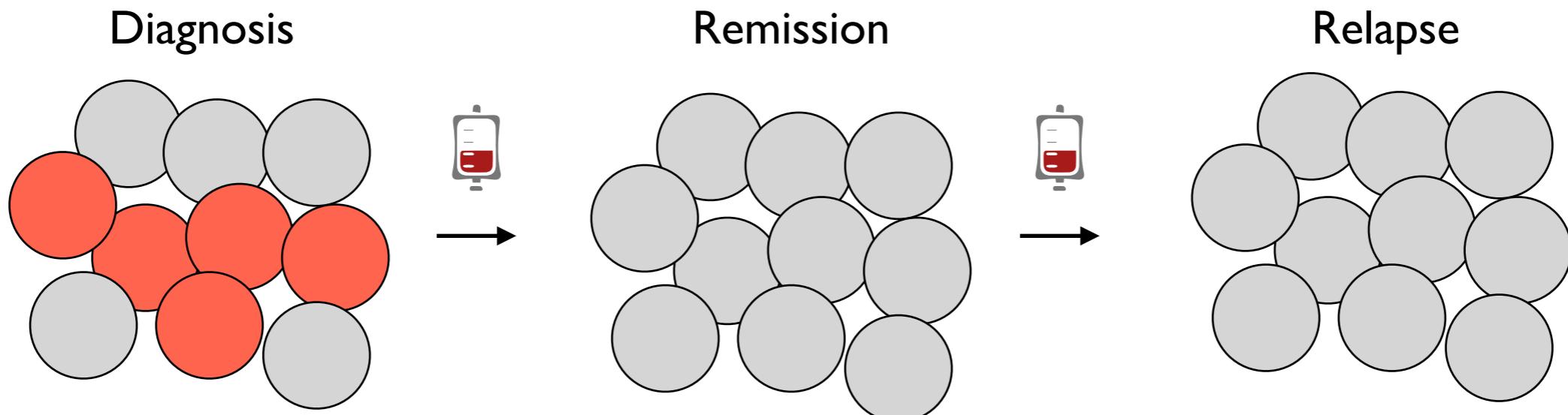


$$\text{VAF} = \frac{\text{Mutation}}{\text{Mutation} + \text{Normal}} = 4/9 = 0.44$$

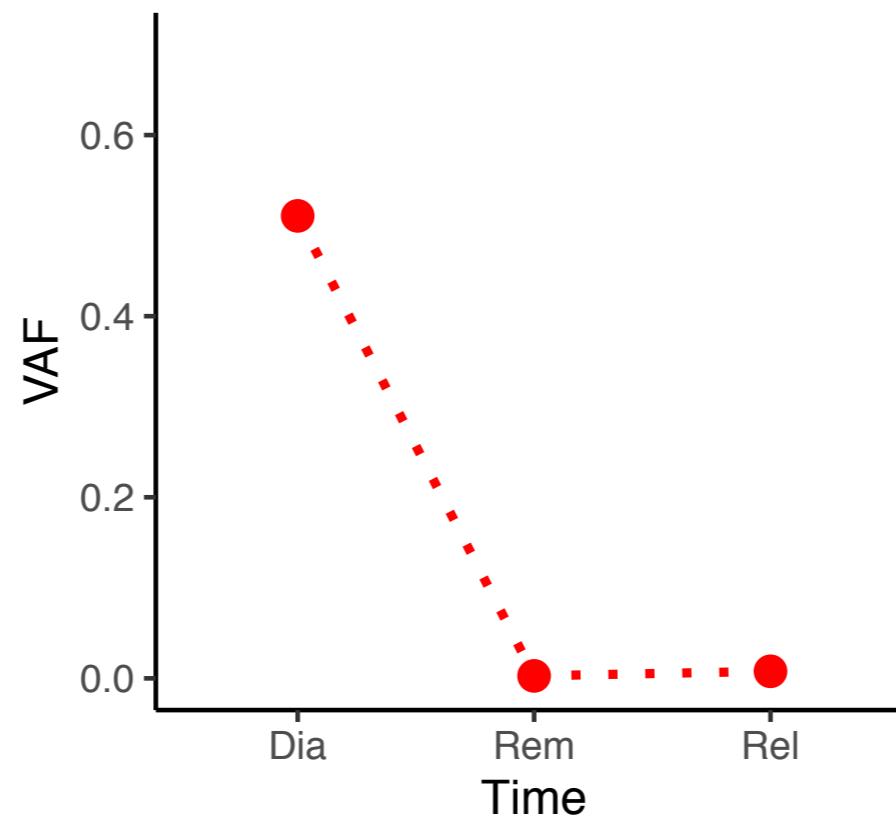
● Respondent mutations



● Respondent mutations

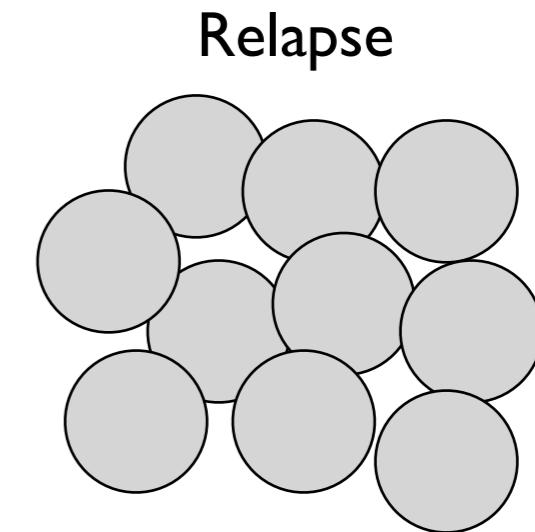
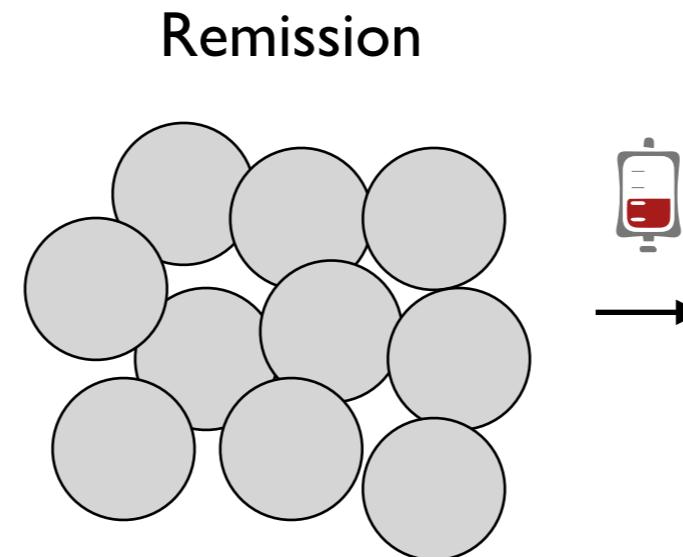
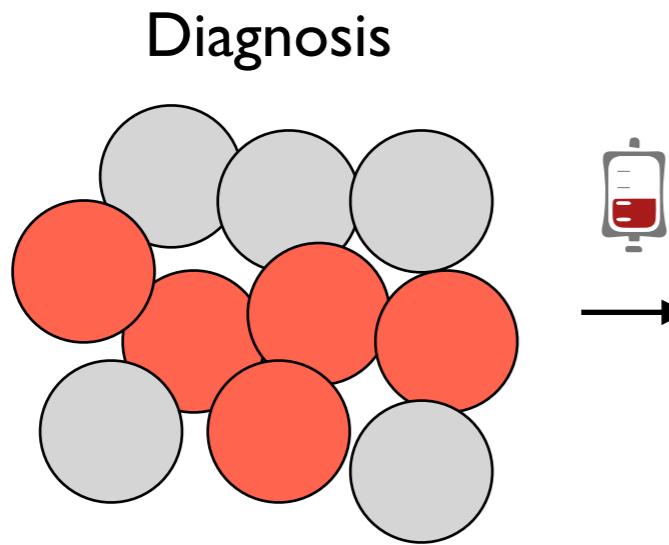


Followed over time

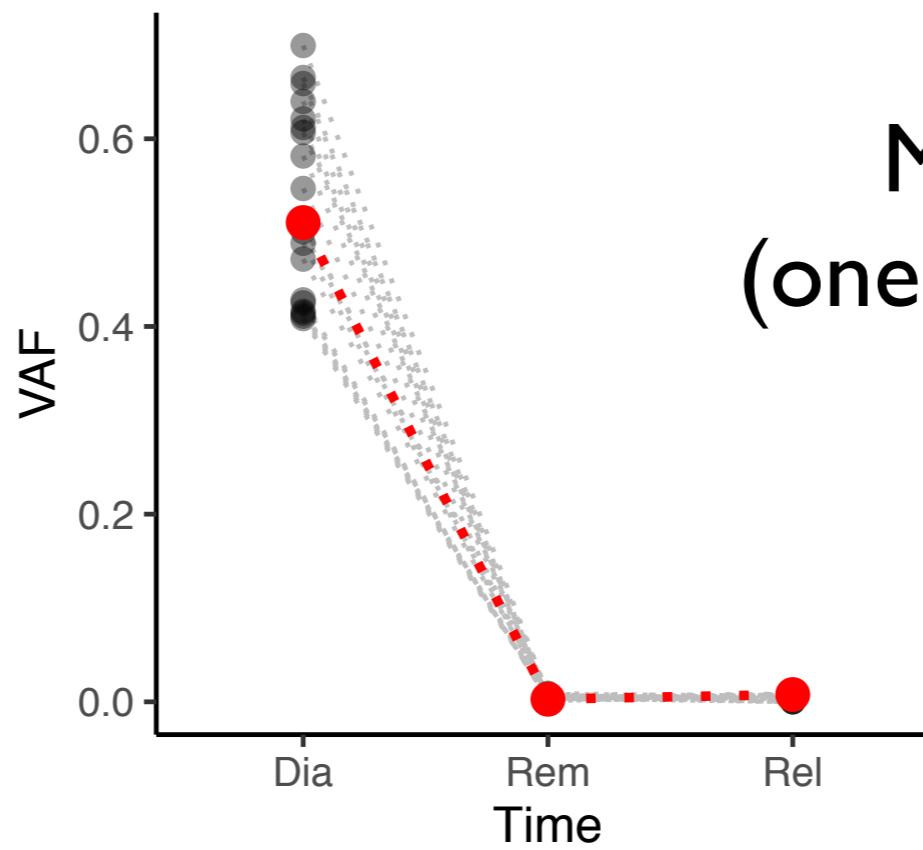


One patient
One mutation

● Respondent mutations



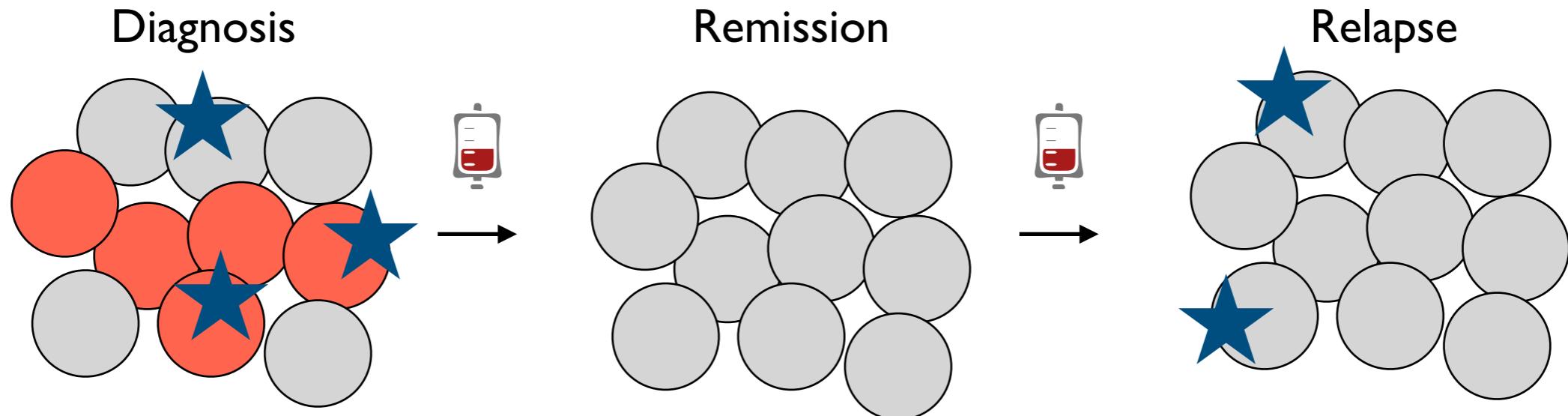
Followed over time



Many mutations
(one or more patients)

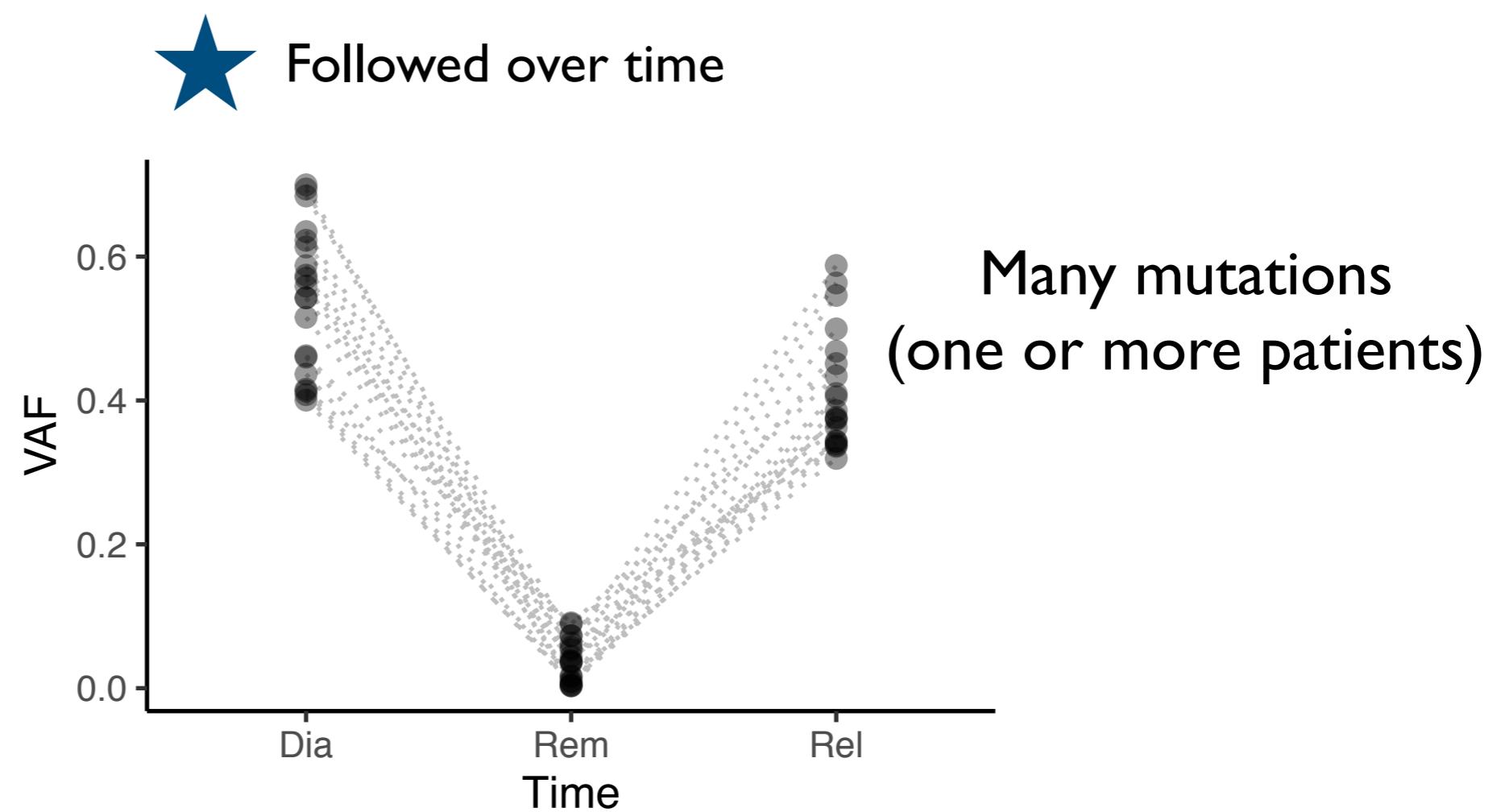
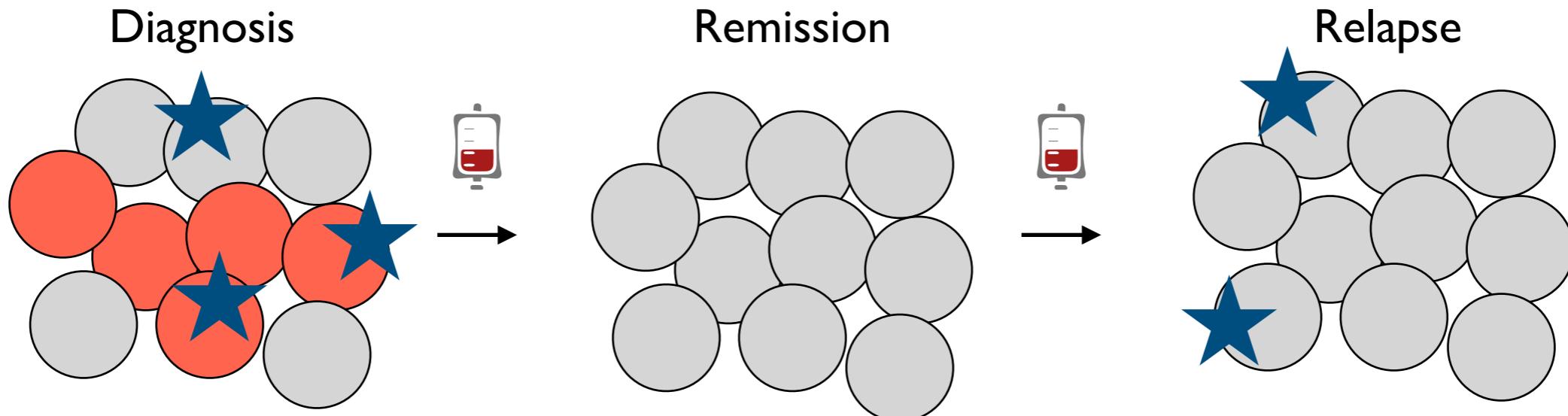


Recurrent mutations

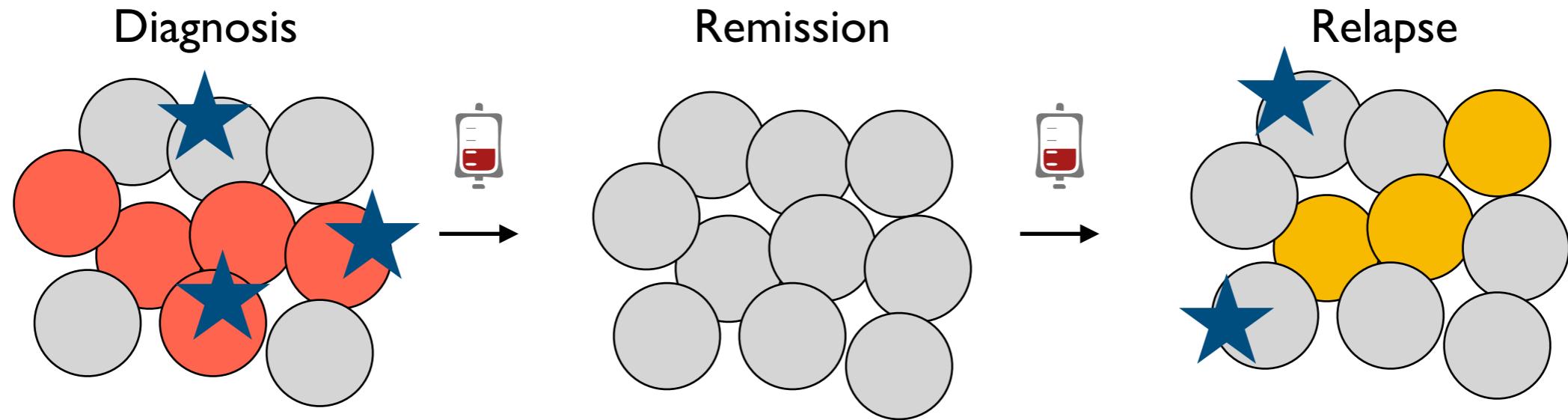




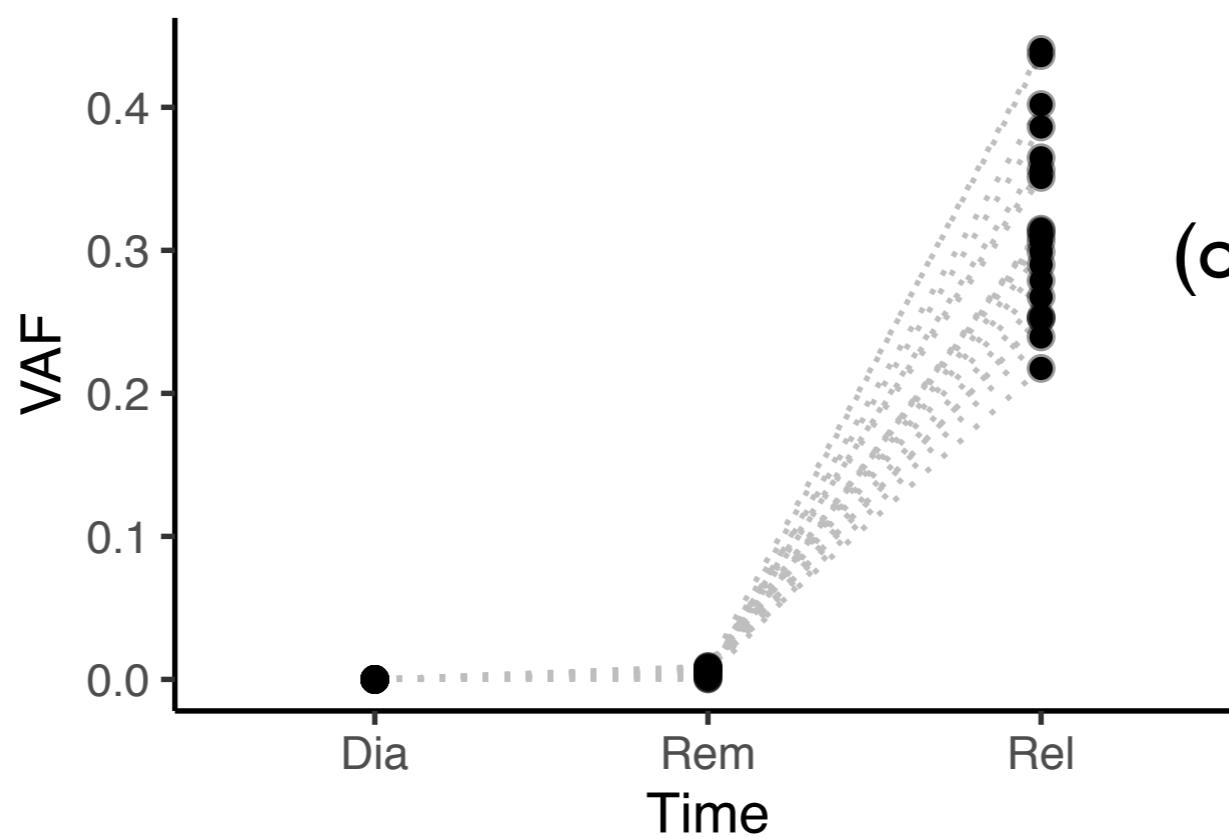
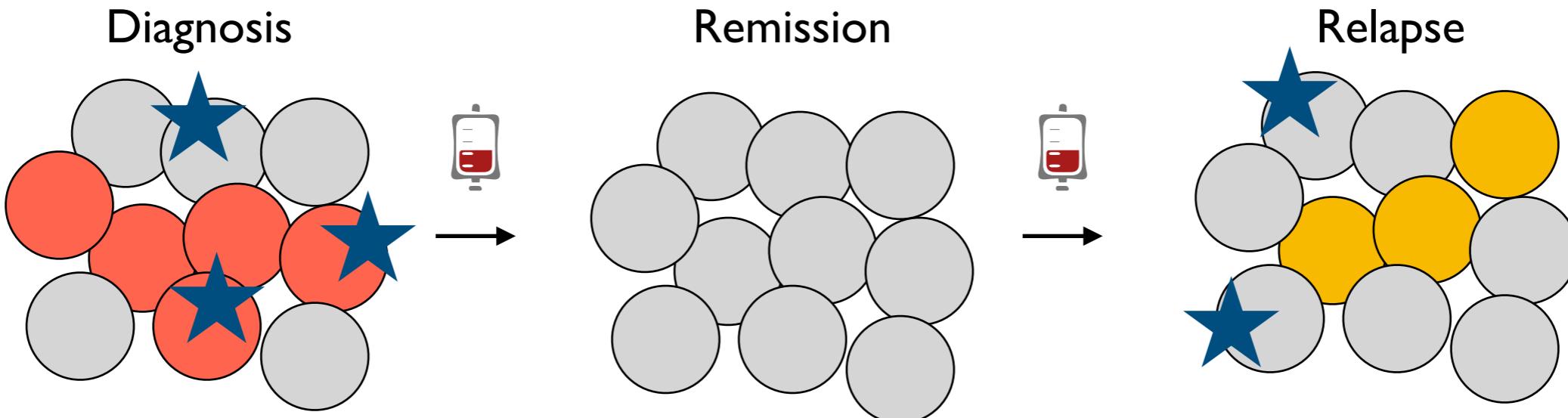
Recurrent mutations



● Relapse specific mutations

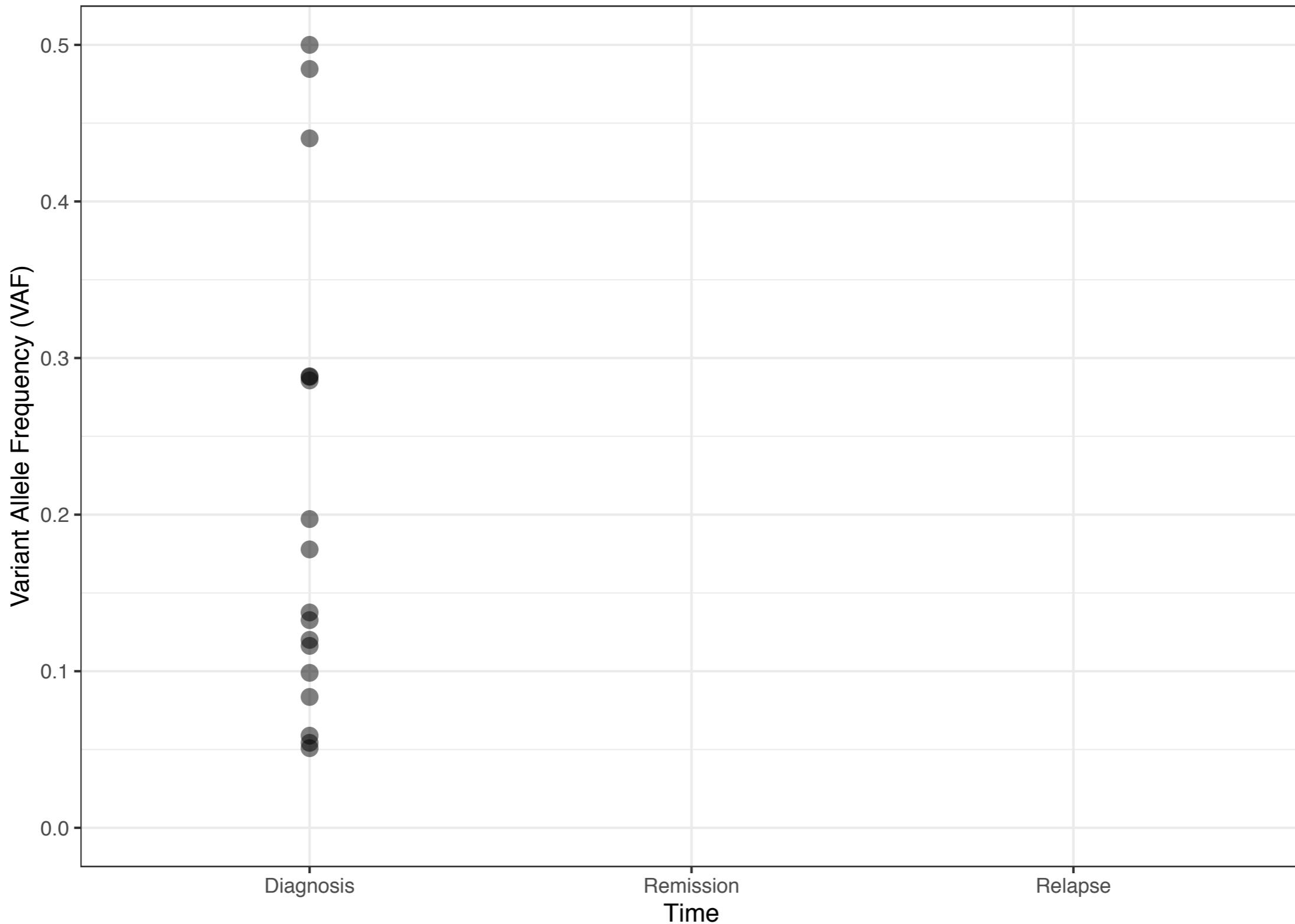


● Relapse specific mutations

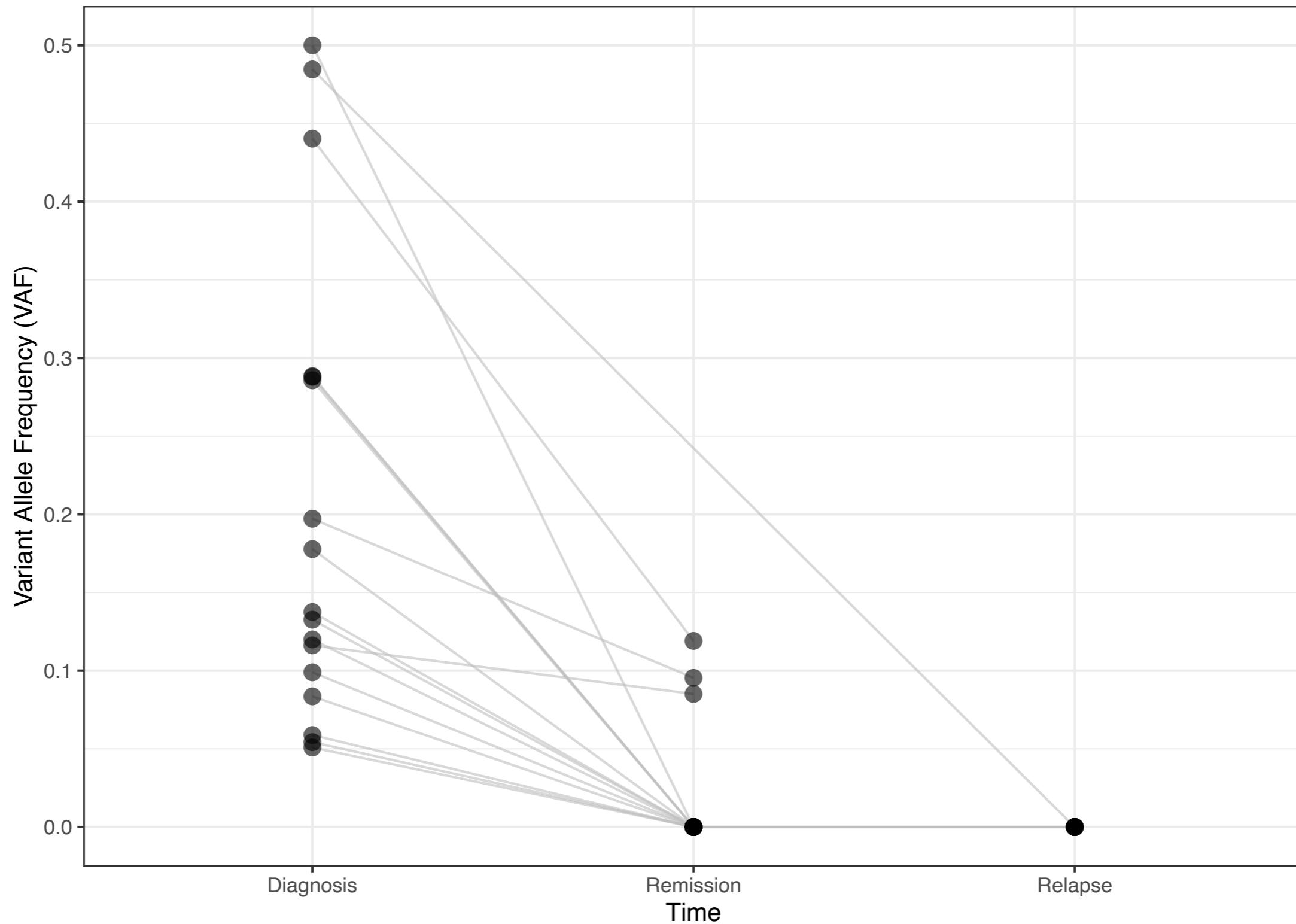


KIT mutations over time

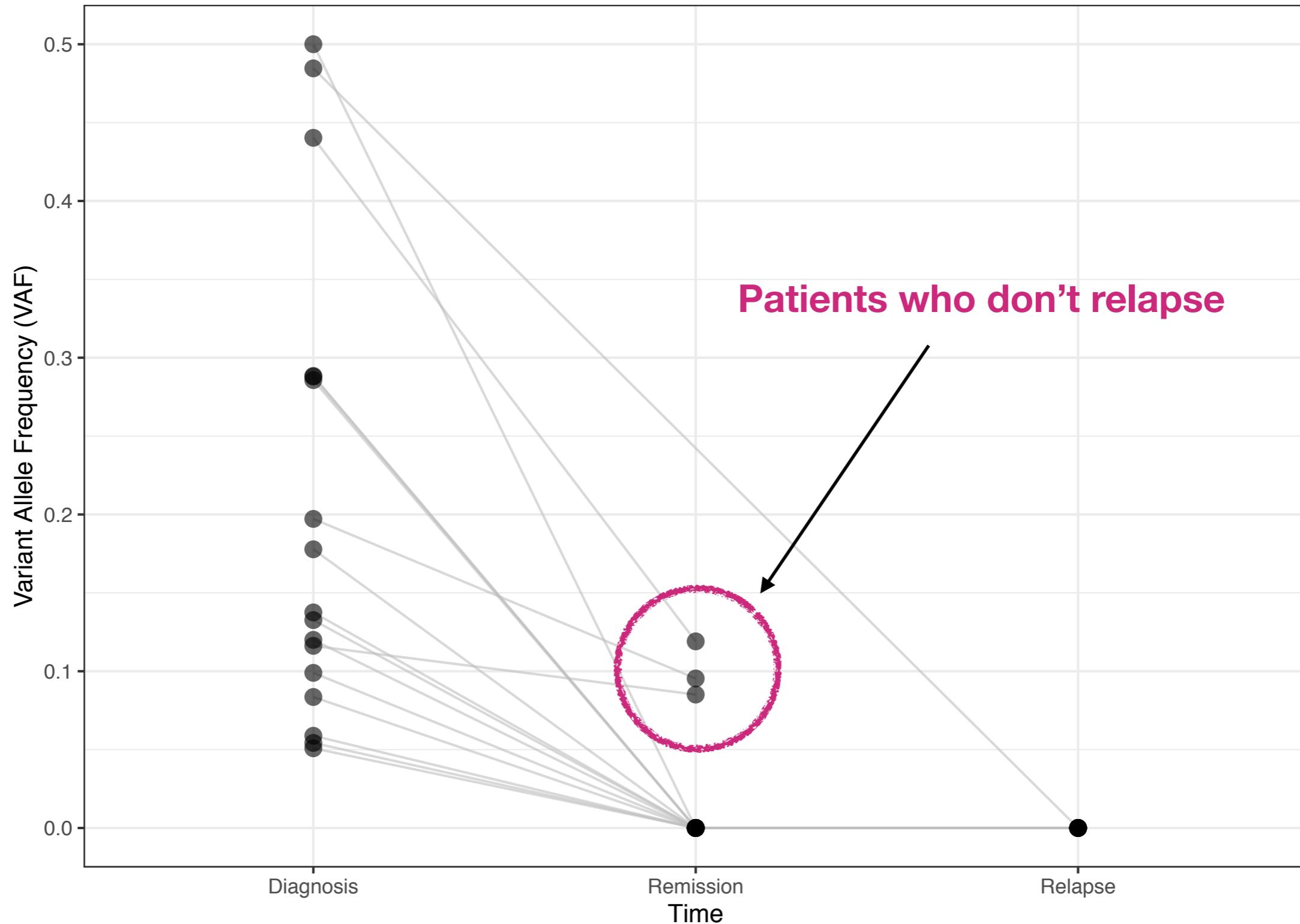
KIT mutations over time



All respondent mutations!



All respondent mutations!

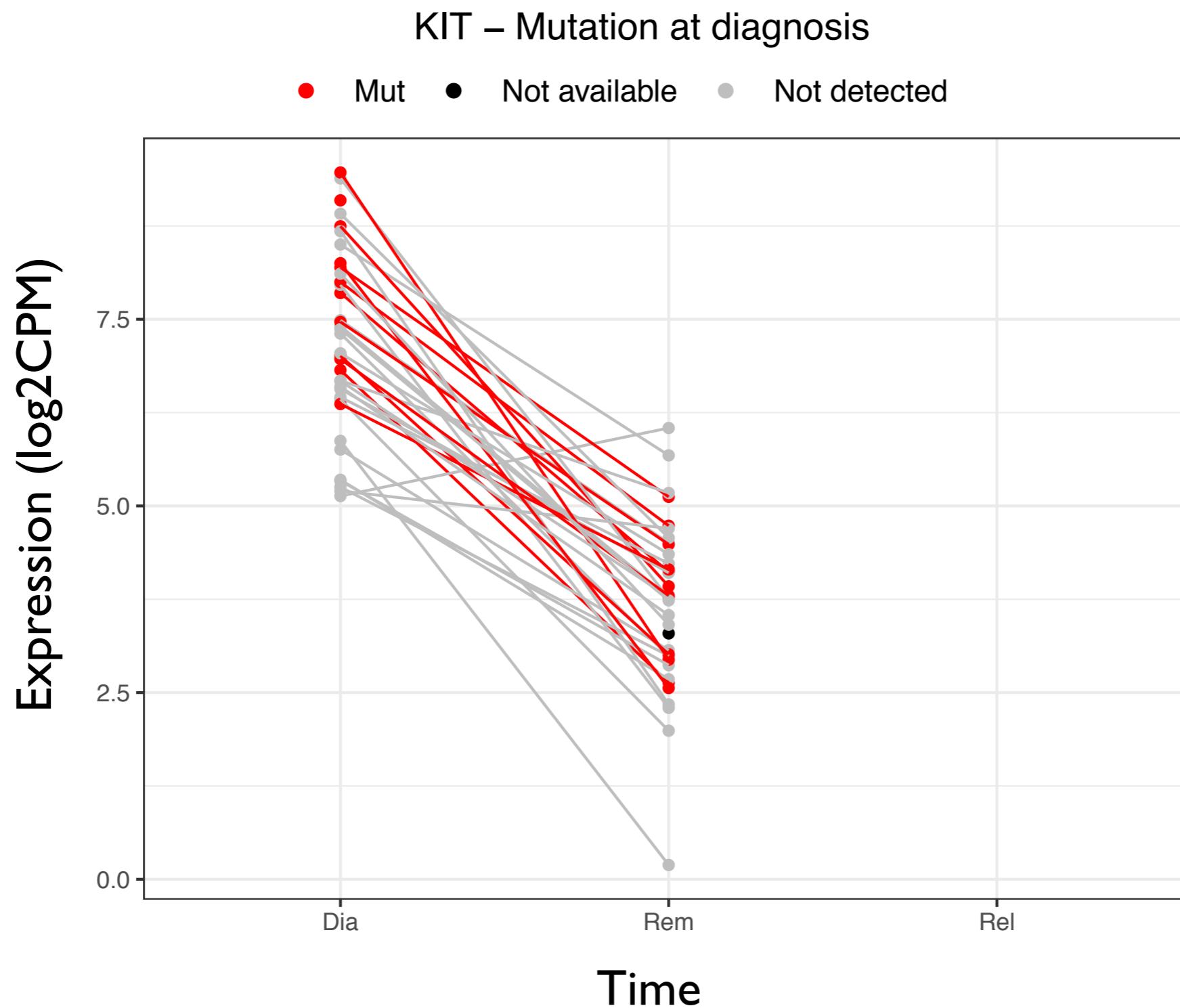


Is KIT expressed at relapse?

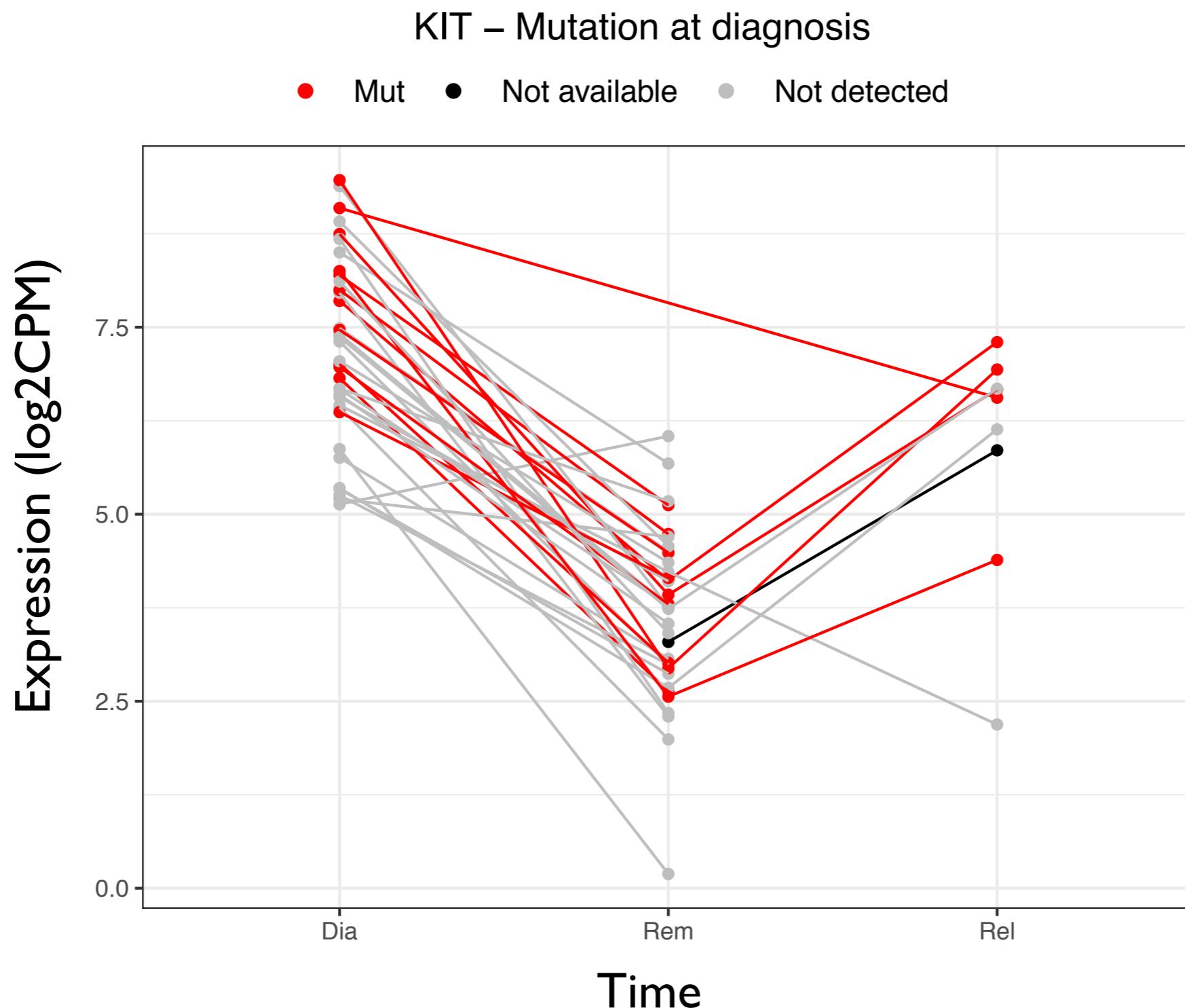


RNA ?

KIT expression decreases at remission



KIT is re-expressed at relapse



Unexpected result!

Unexpected result!

KITmut associated with relapse
at diagnosis

Unexpected result!

KITmut associated with relapse
at diagnosis



KITmut not detected at relapse

Unexpected result!

KITmut associated with relapse
at diagnosis

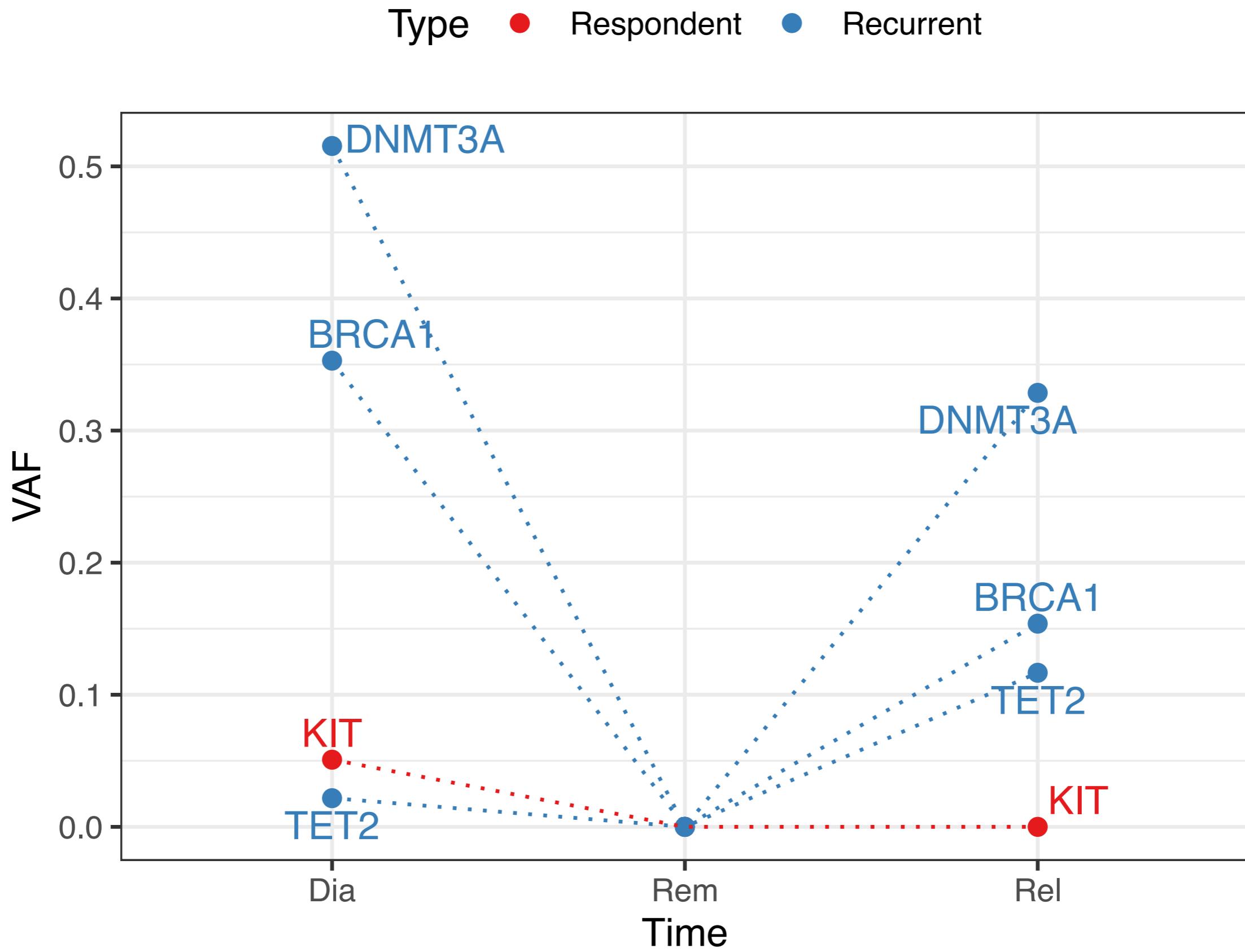


KITmut not detected at relapse



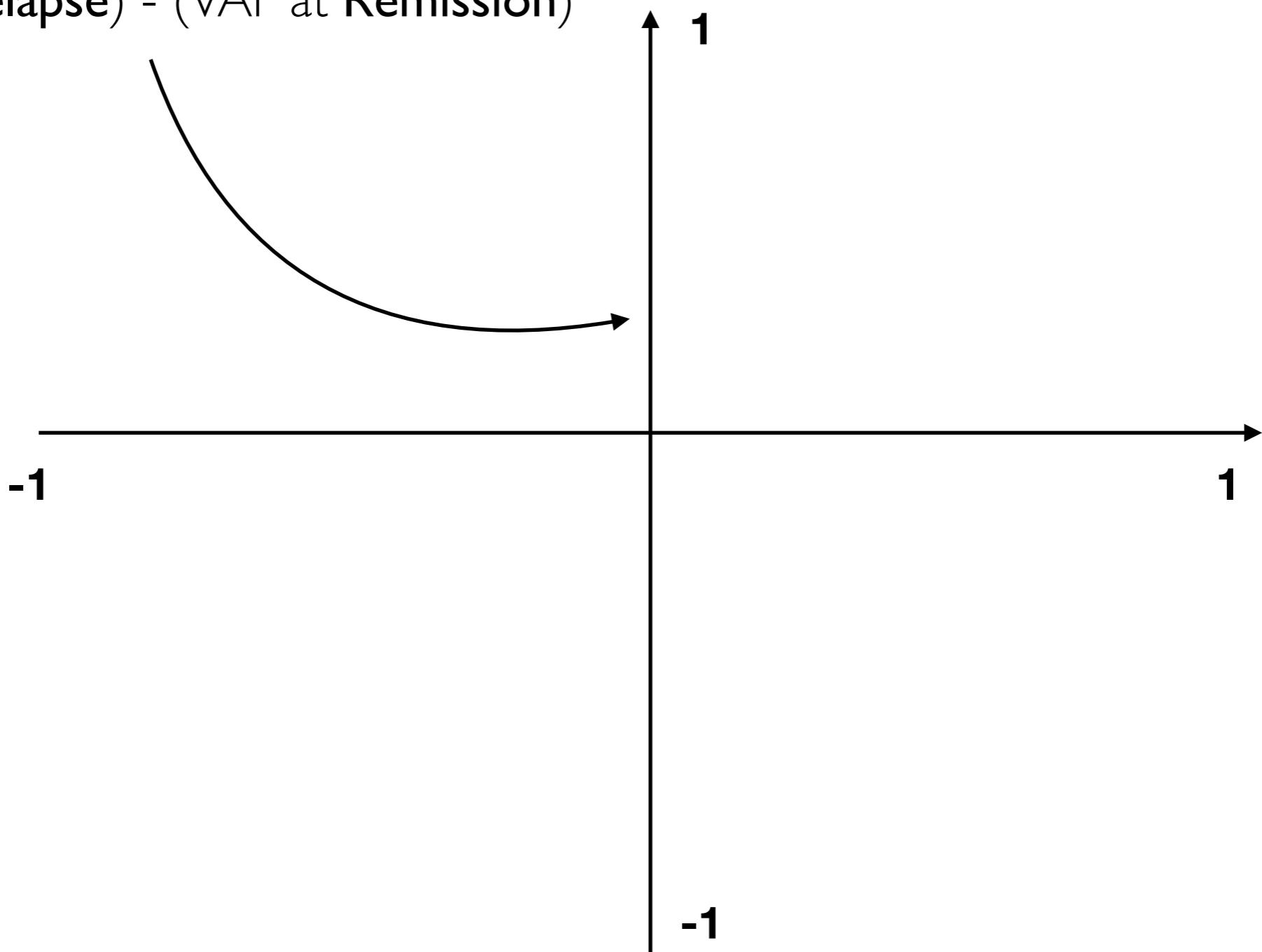
Are there other genes at relapse?

Patient by patient...



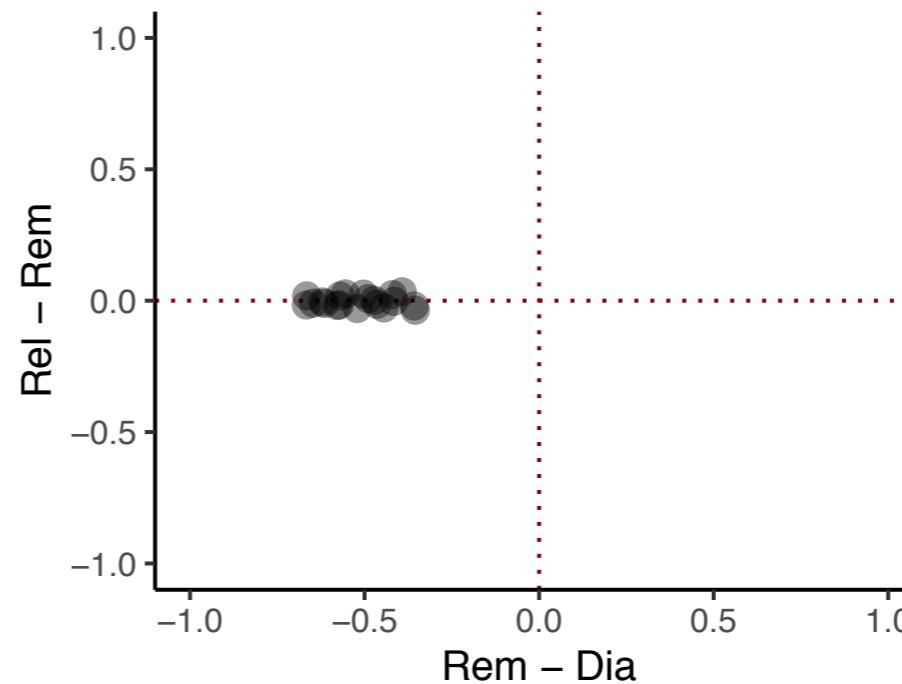
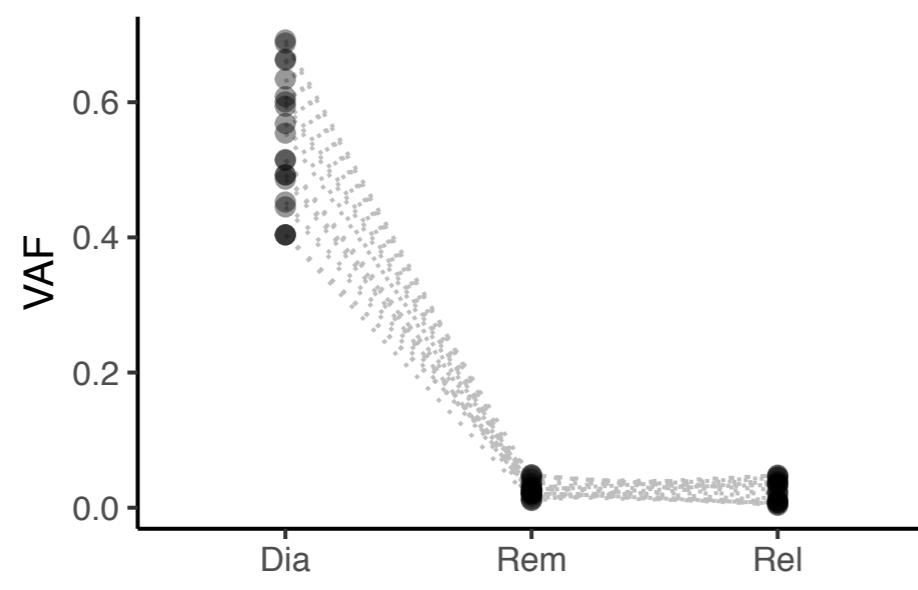
Summarise VAF shifts over time

$(\text{VAF at Relapse}) - (\text{VAF at Remission})$

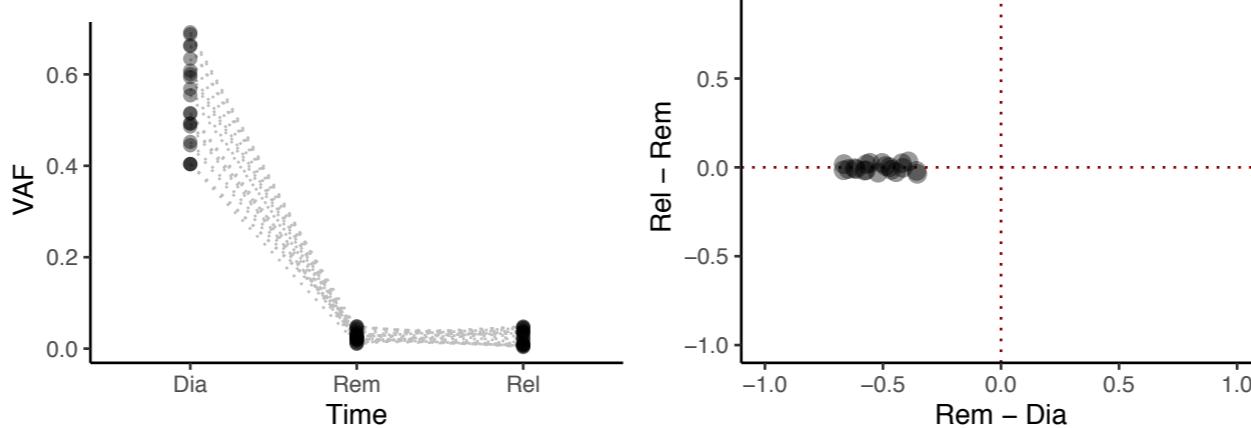


$(\text{VAF at Remission}) - (\text{VAF at Diagnosis})$

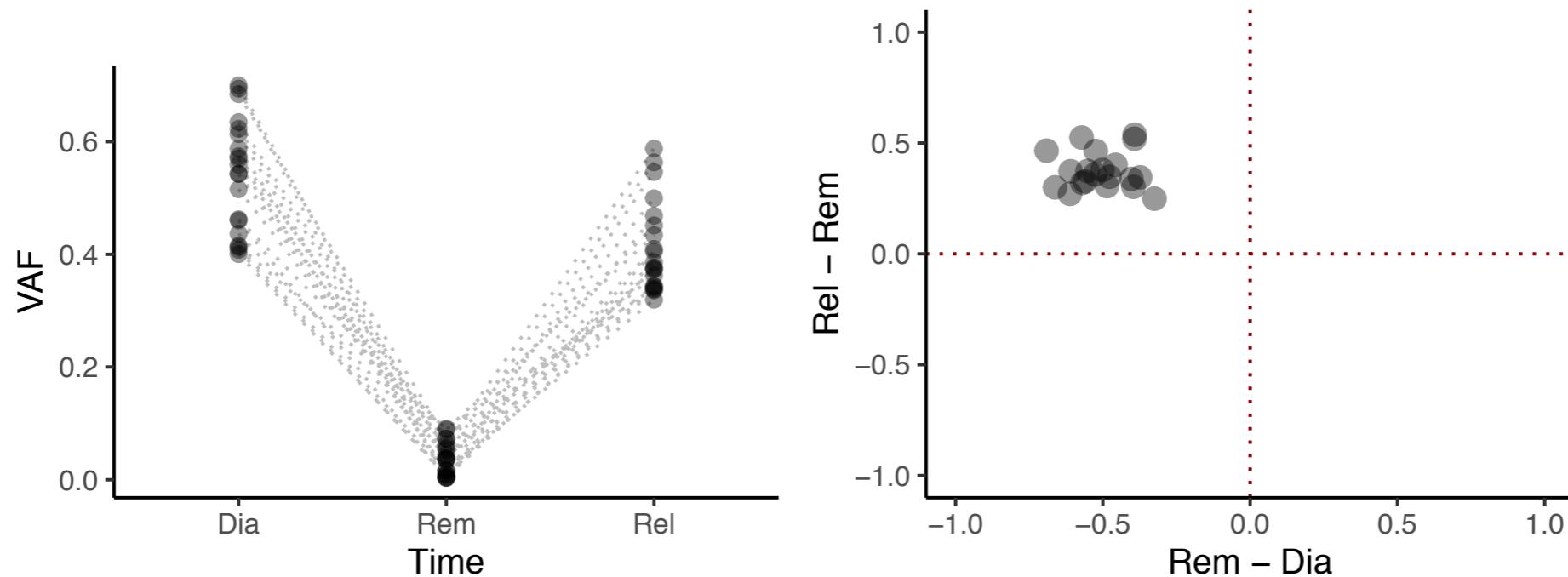
Respondent mutations



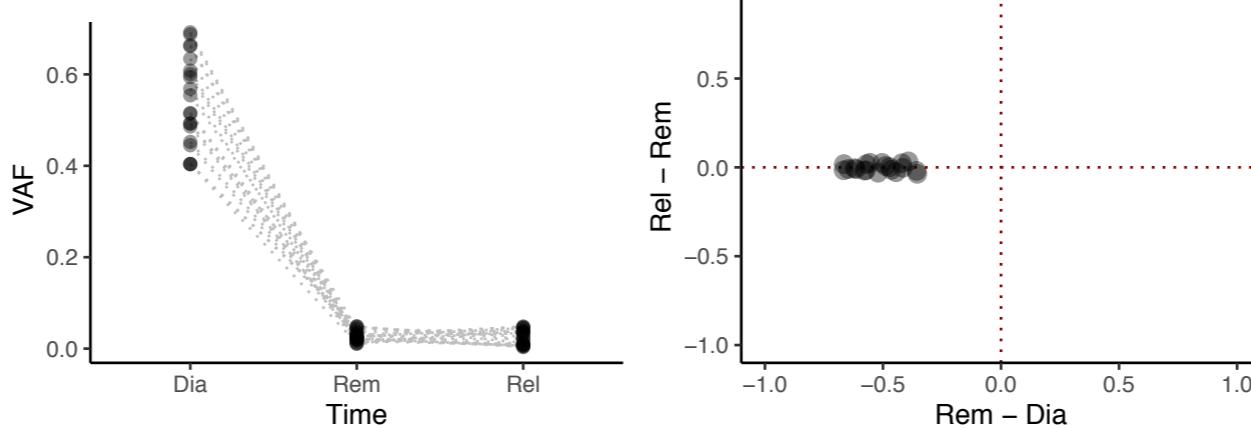
Respondent mutations



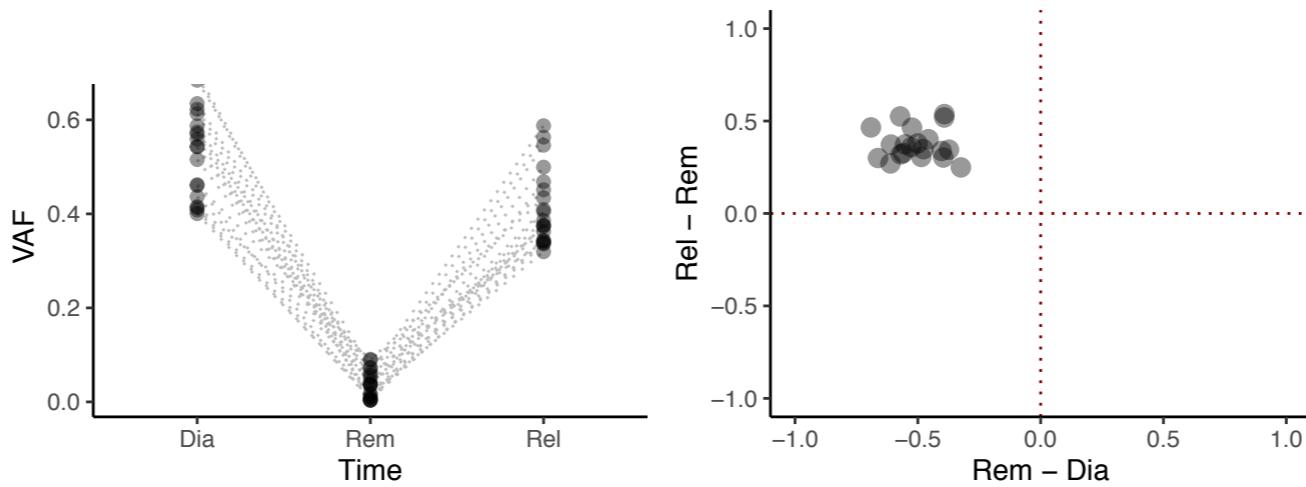
Recurrent mutations



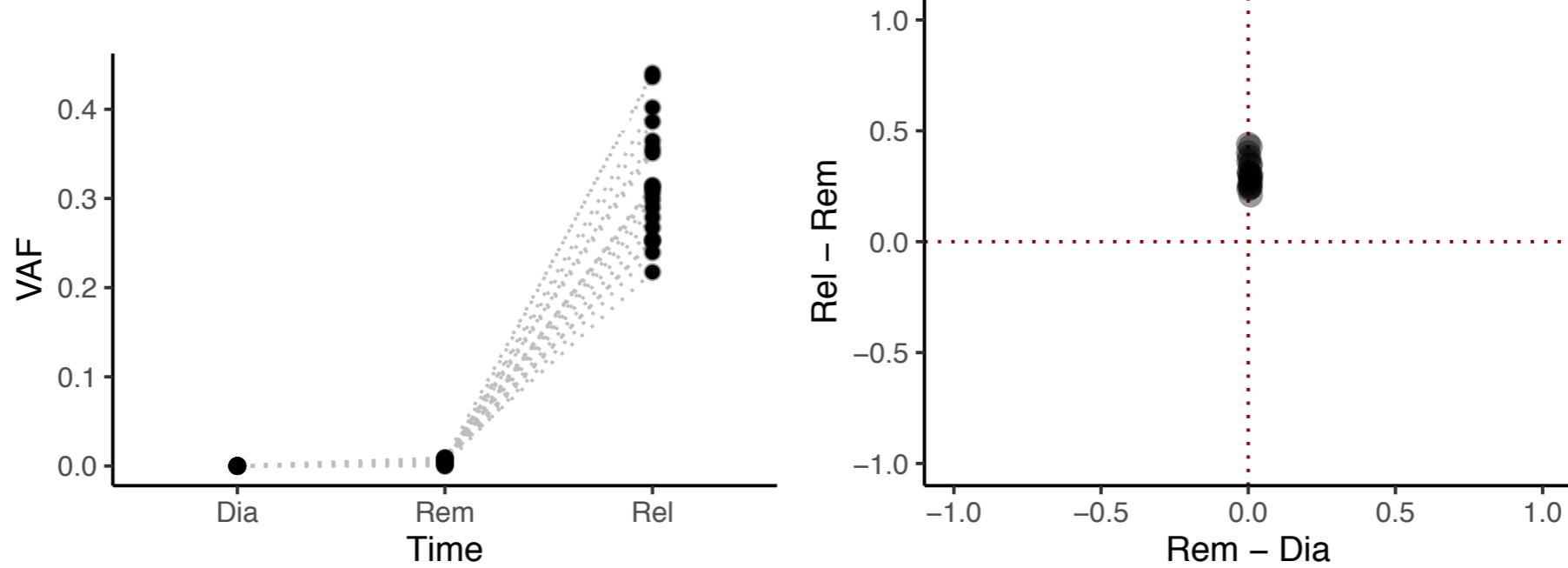
Respondent mutations



Recurrent mutations

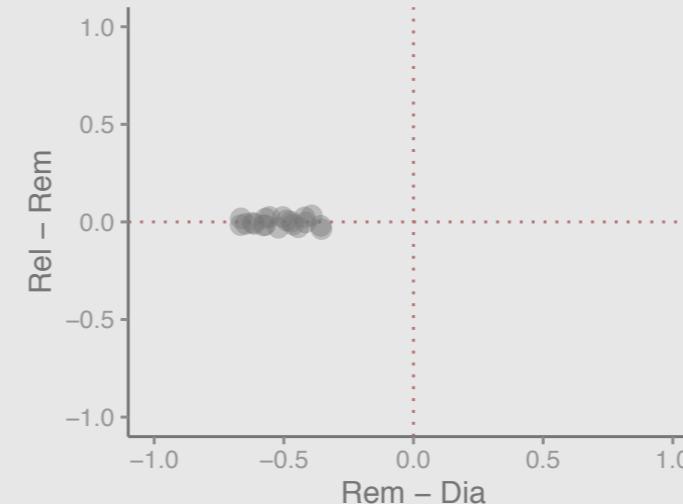
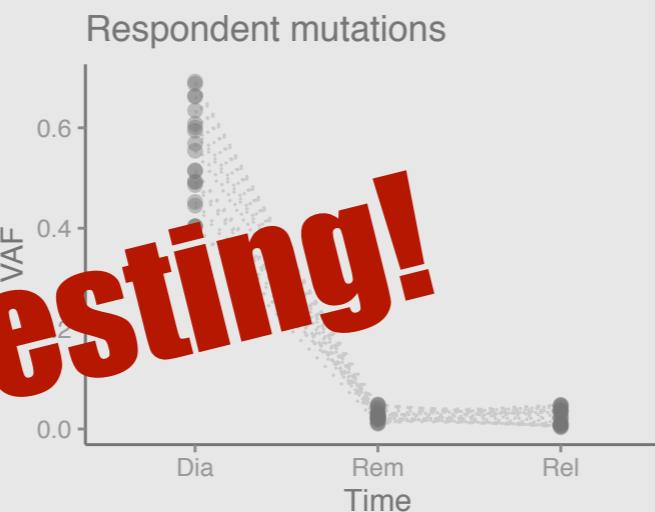


Relapse specific mutations

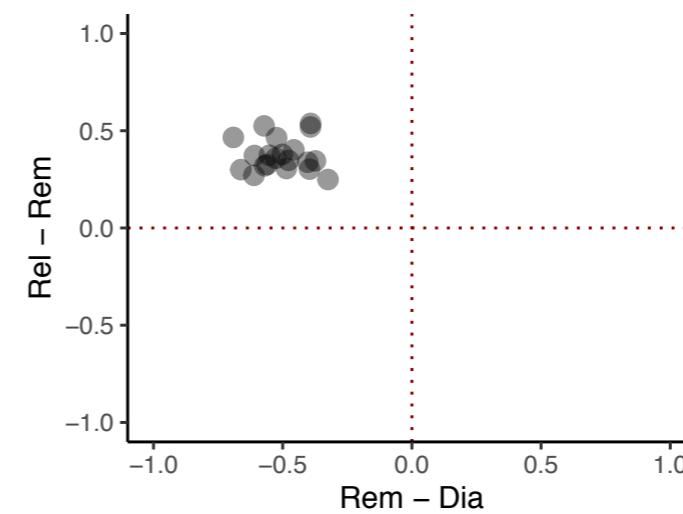
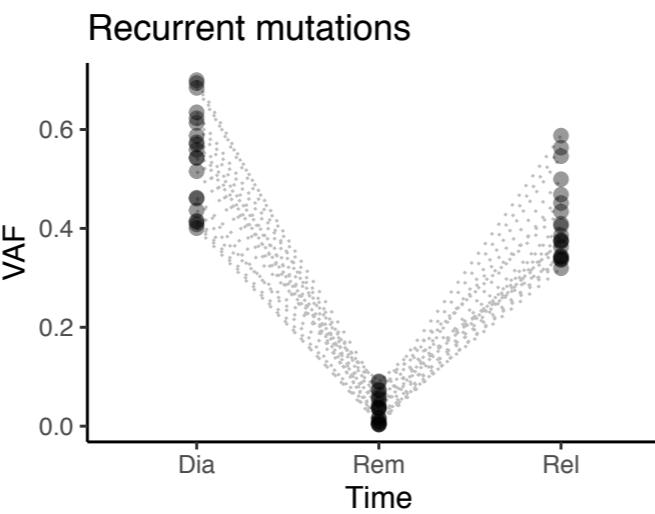


Respondent mutations

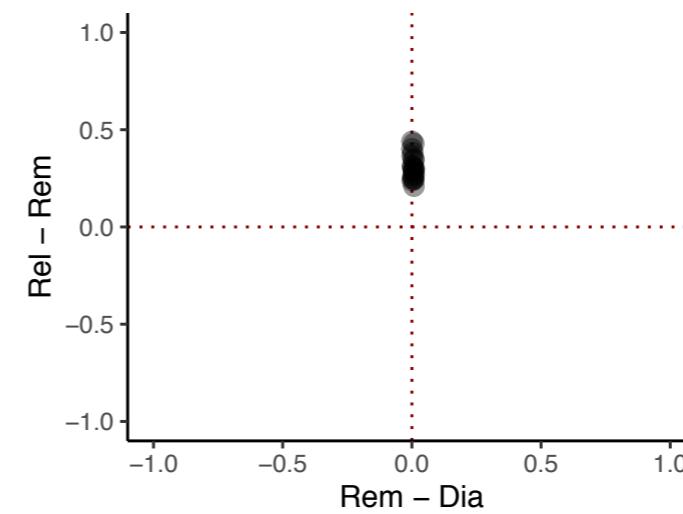
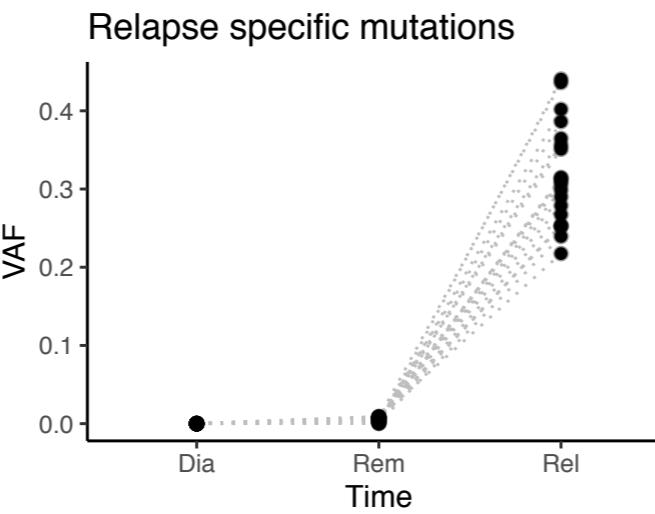
Interesting!



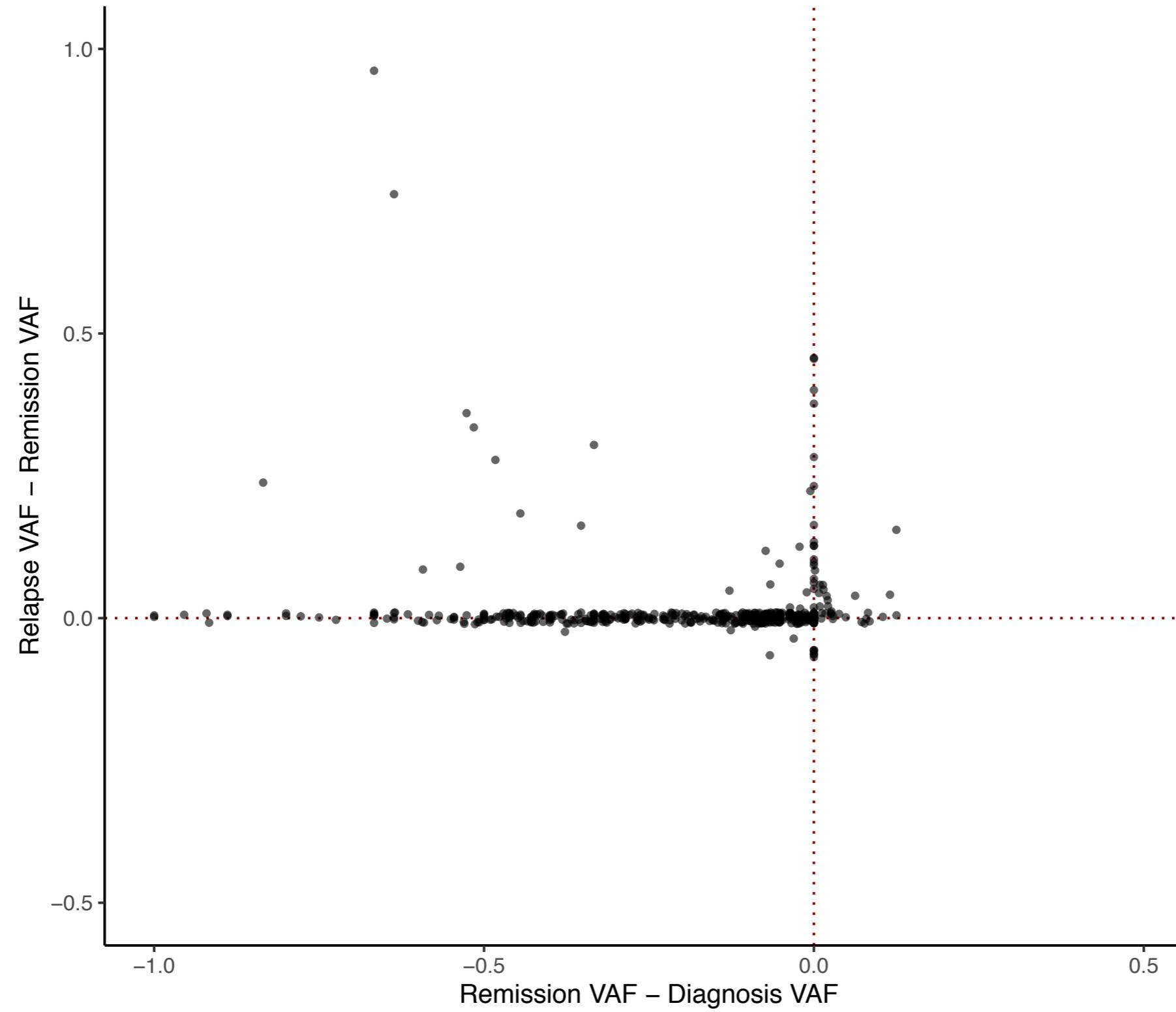
Recurrent mutations



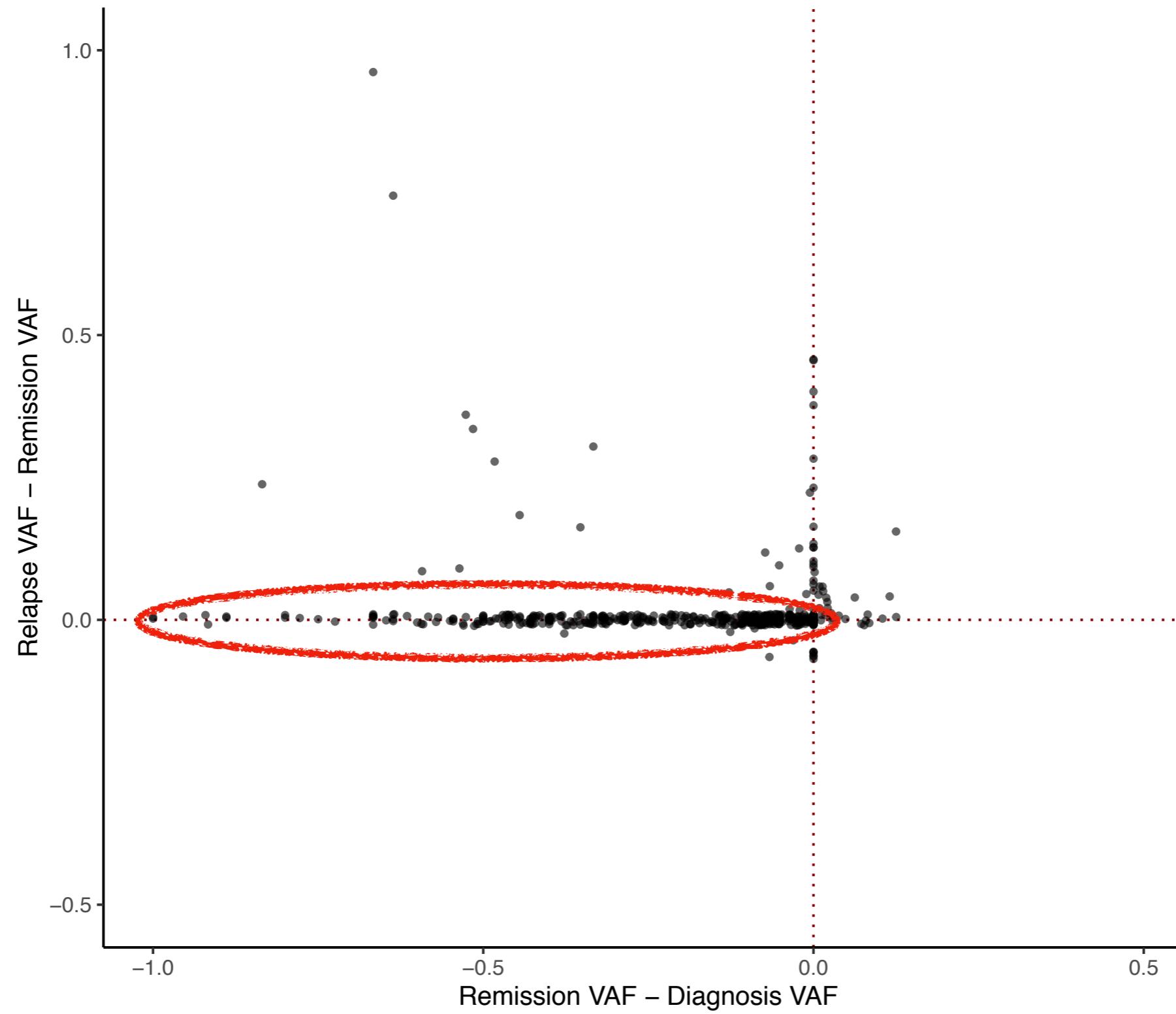
Relapse specific mutations



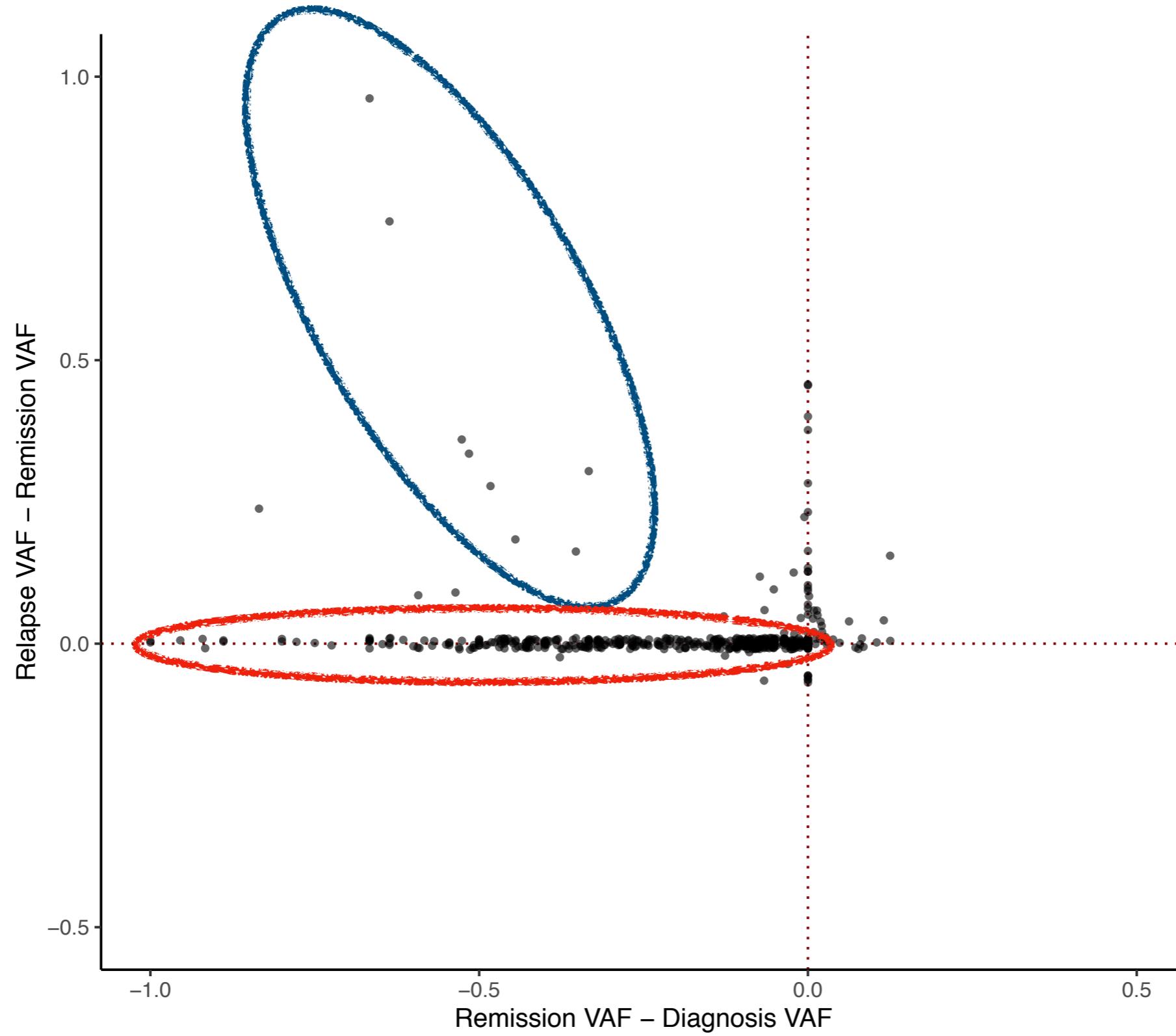
All mutations across patients



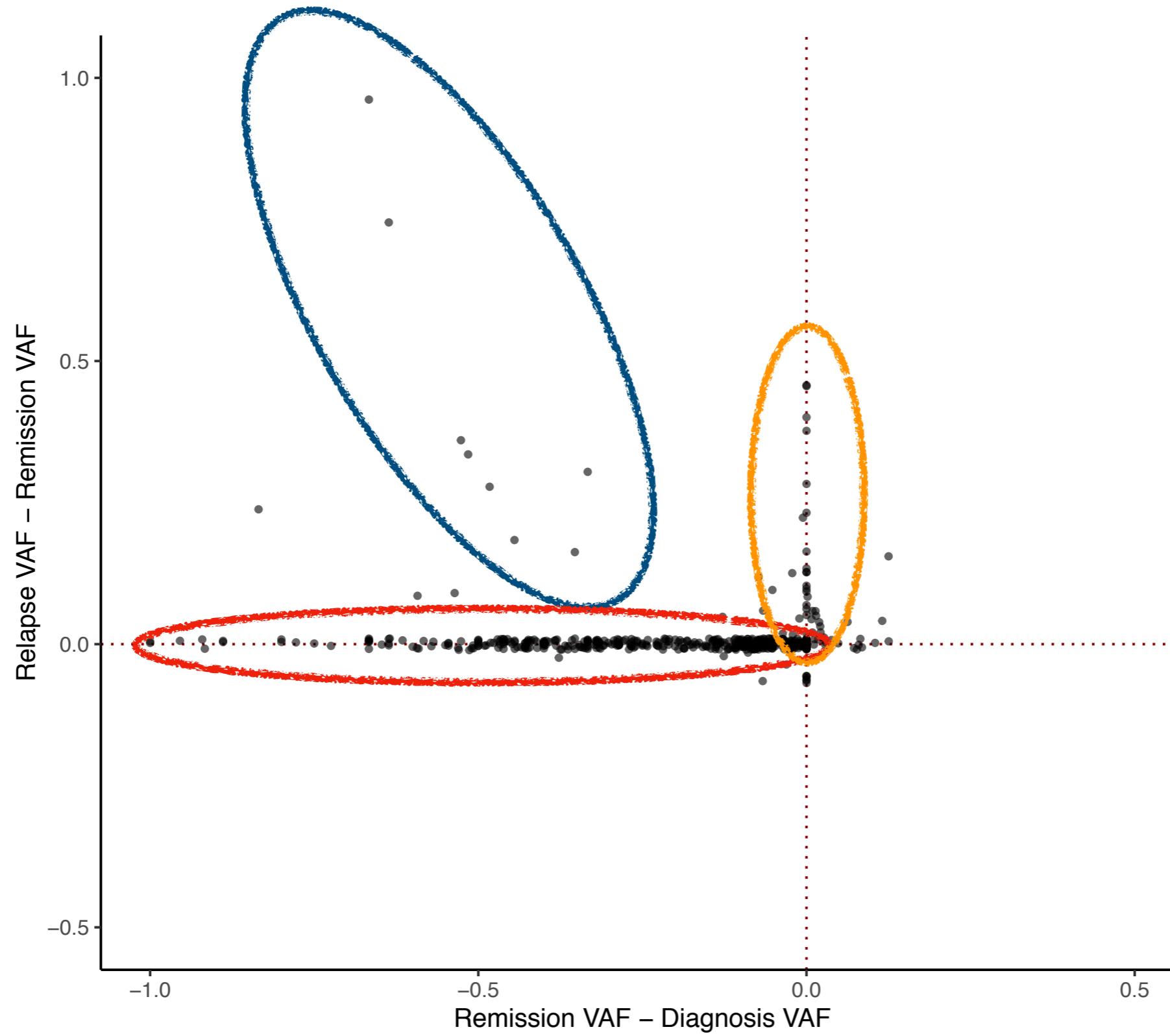
Respondent mutations



Recurrent mutations

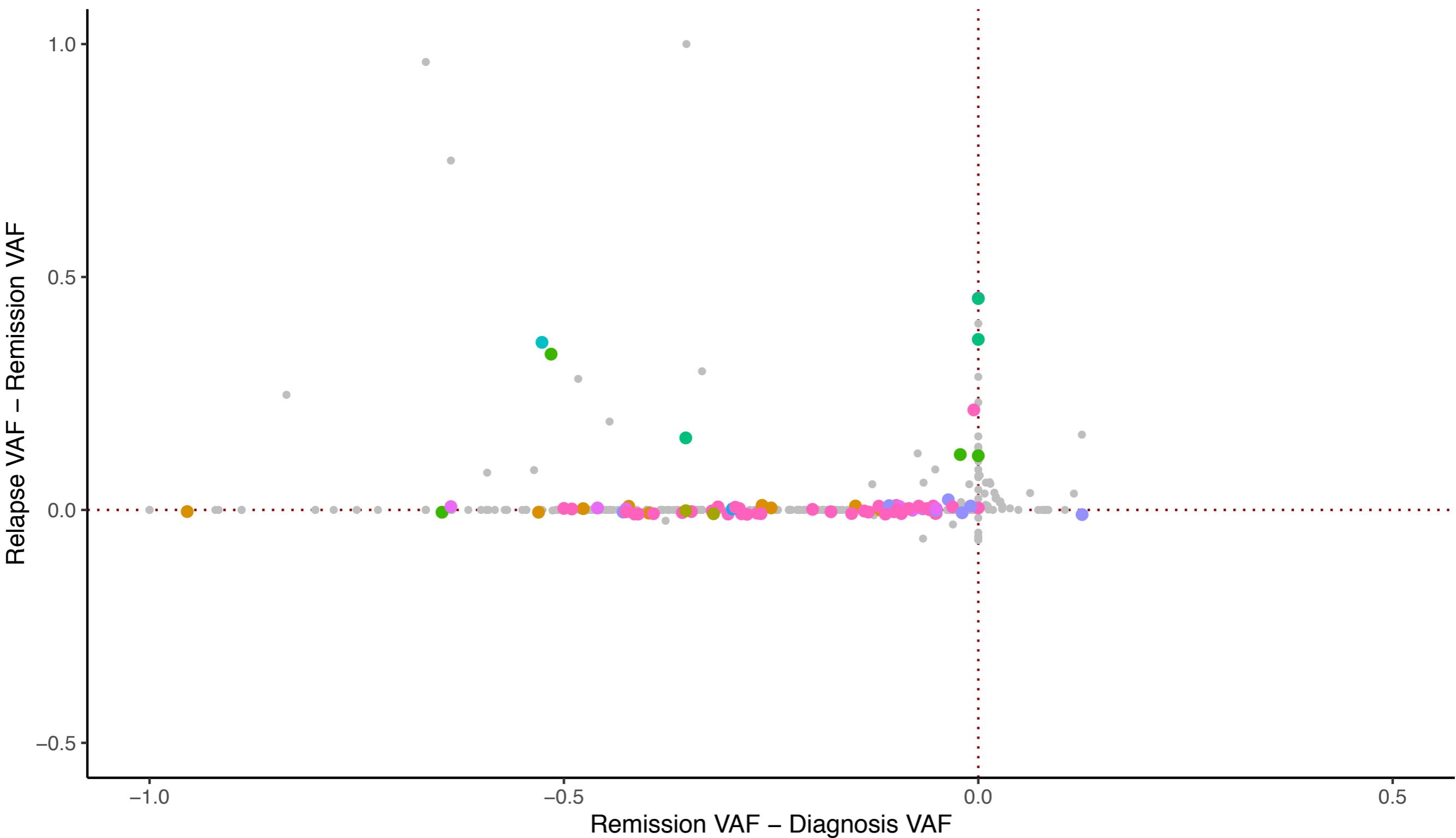


Relapse specific mutations



FUNCTION

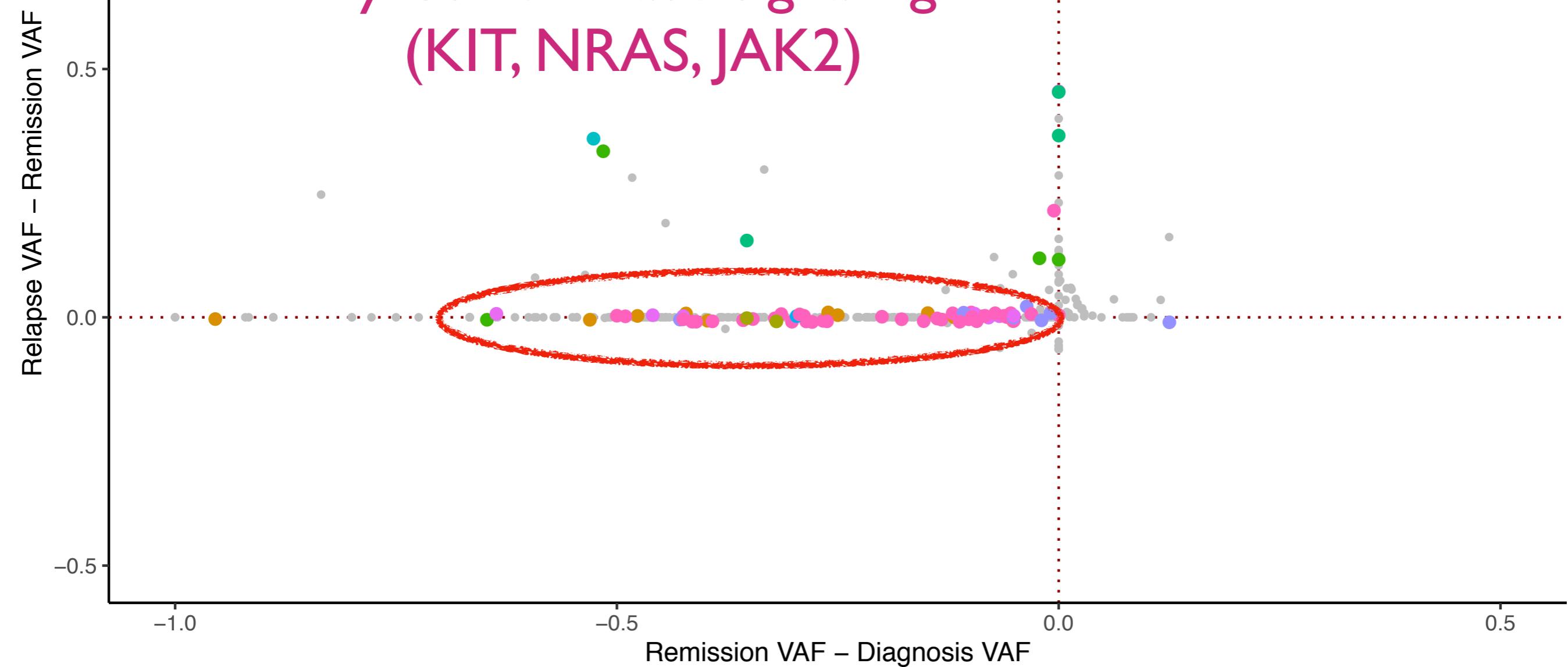
- Cell Cycle
- Chromatin remodeling
- Cohesin complex
- DNA methylation
- DNA repair
- Multidrug resistance
- Other genes
- Signalling
- Transcription regulator
- Tyrosine kinase signaling



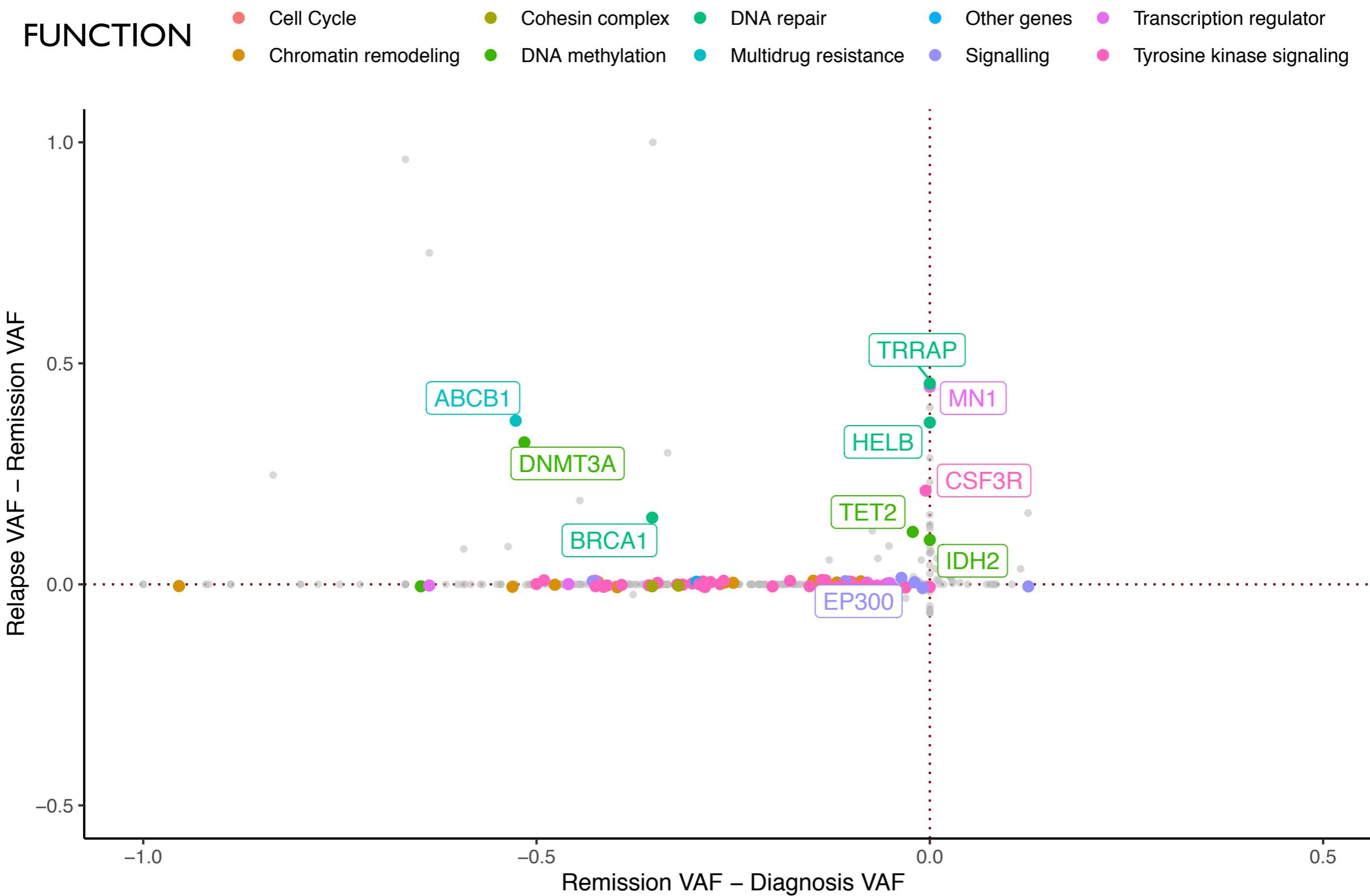
FUNCTION

- Cell Cycle
- Cohesin complex
- DNA repair
- Other genes
- Transcription regulator
- Chromatin remodeling
- DNA methylation
- Multidrug resistance
- Signalling
- Tyrosine kinase signaling

Tyrosine kinase signaling (KIT, NRAS, JAK2)

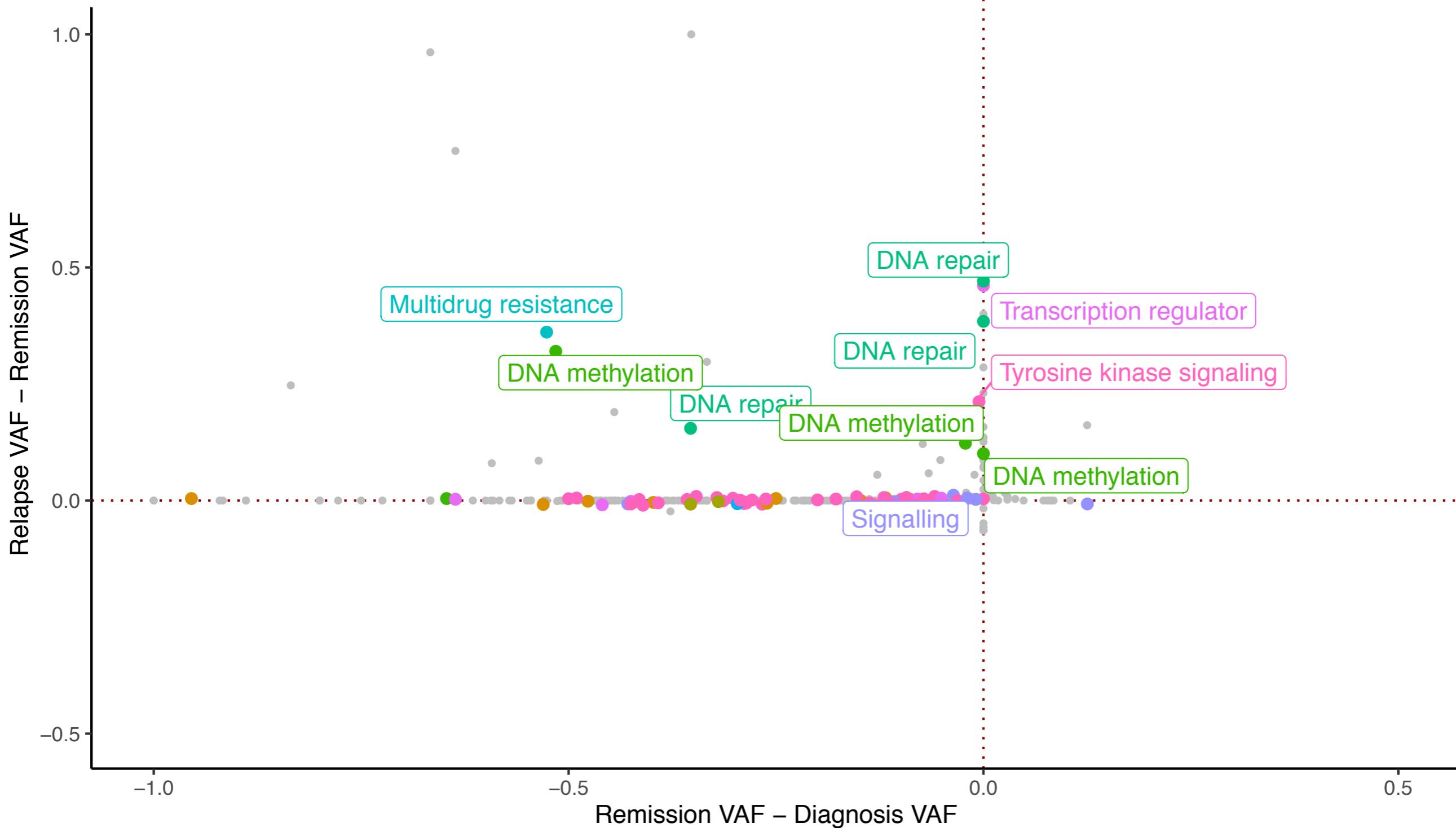


Recurrent/Relapse specific



FUNCTION

- Cell Cycle
- Chromatin remodeling
- Cohesin complex
- DNA methylation
- DNA repair
- Multidrug resistance
- Other genes
- Signalling
- Transcription regulator
- Tyrosine kinase signaling



- ✓ Experimental design
- ✓ Compare patients at diagnosis



Changes over time

Mutation

- KIT (mysteriously) not detected at relapse

- ✓ Experimental design
- ✓ Compare patients at diagnosis



Changes over time

Mutation

- KIT (mysteriously) not detected at relapse
- DNA methylation & repair genes mutated at relapse

✓ Experimental design

✓ Compare patients
at diagnosis



Changes over time

Mutation

- KIT (mysteriously) not detected at relapse
- DNA methylation & repair genes mutated at relapse
- No strong patterns in gene expression over time (not shown)

What next?

What next?

- Validate KIT mutation in DNA at relapse

What next?

- Validate KIT mutation in DNA at relapse
- Evaluate clinical importance of pathways/genes mutated at relapse

What next?

- Validate KIT mutation in DNA at relapse
- Evaluate clinical importance of pathways/genes mutated at relapse
- Collecting more CBF AML patients who relapse from the Royal Melbourne Hospital to validate our observations

Methods summary

- Workflows to detect mutations in RNA-Seq



Methods summary



- Workflows to detect mutations in RNA-Seq
 - **varikondo** to standardise and combine variants





Methods summary



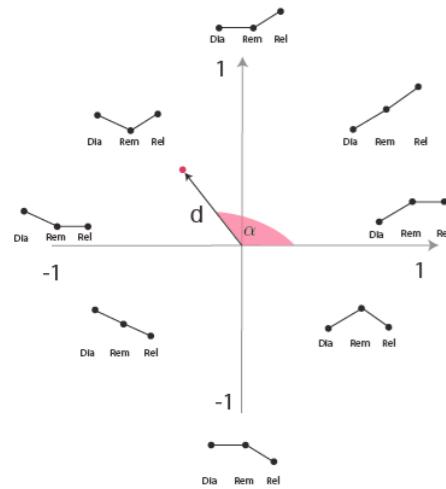
- Workflows to detect mutations in RNA-Seq
 - **varikondo** to standardise and combine variants
 - R Shiny application to explore mutations interactively <https://shiny.wehi.edu.au/quagliari.a/shiny-clone/>



Methods summary



- Workflows to detect mutations in RNA-Seq
 - **varikondo** to standardise and combine variants
 - R Shiny application to explore mutations interactively <https://shiny.wehi.edu.au/quagliari.a/shiny-clone/>
- Not shown - **RUV** methods to adjust for unwanted variation in gene expression



Methods summary



- Workflows to detect mutations in RNA-Seq
 - **varikondo** to standardise and combine variants
 - R Shiny application to explore mutations interactively <https://shiny.wehi.edu.au/quagliari.a/shiny-clone/>
- Not shown - **RUV** methods to adjust for unwanted variation in gene expression
- **VAF shift plots** to spot mutations patterns over time

Thanks to...