

9<sup>e</sup> ÉDITION

# JOURNÉES DU **GFCO** 2023

Biomarqueurs et analyses moléculaires en oncologie

Avec la participation  
scientifique du



# *RNA-seq : indications et outils d'analyse*

Franck Tirode

Equipe épigénétique, génétique et biologie des sarcomes

CRCL / CLB



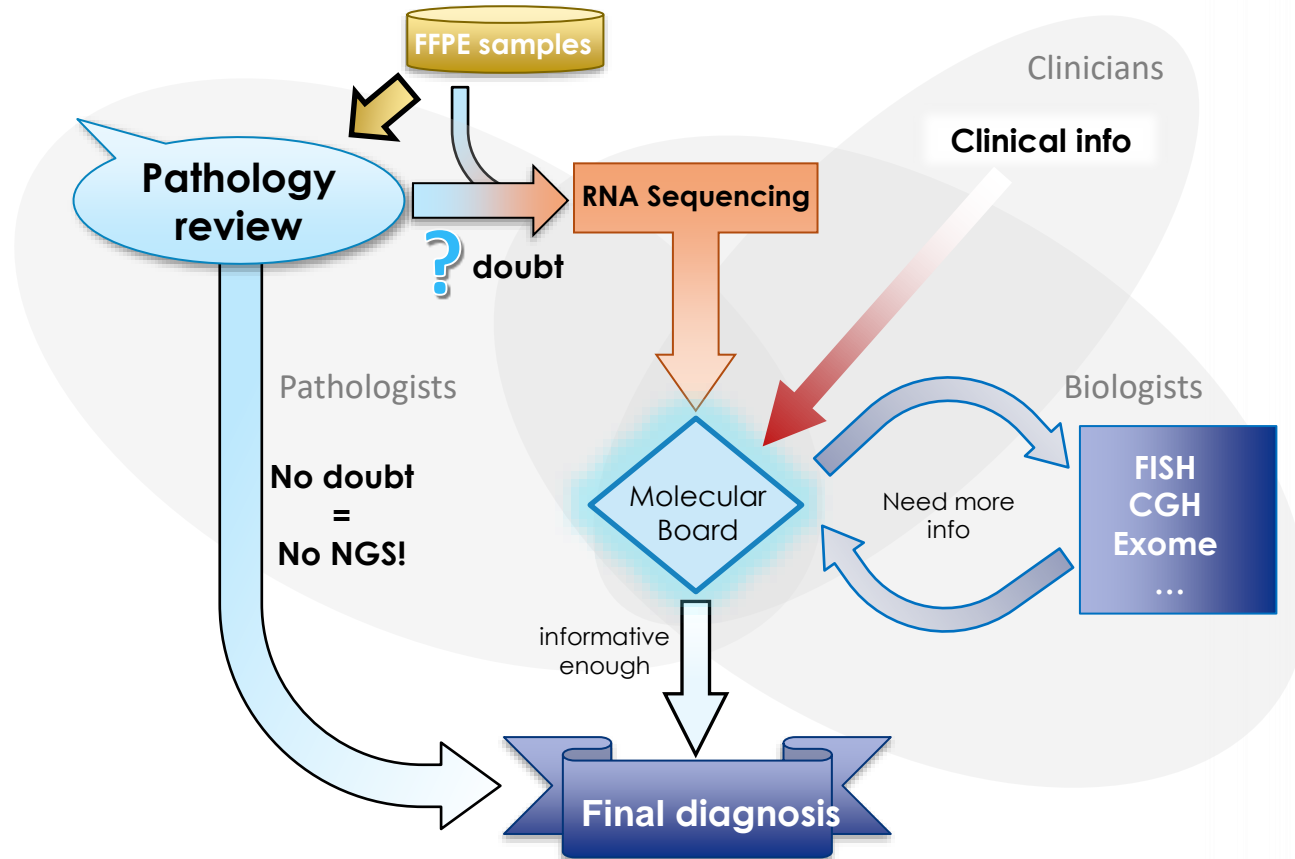
# LIENS D'INTÉRÊT

---

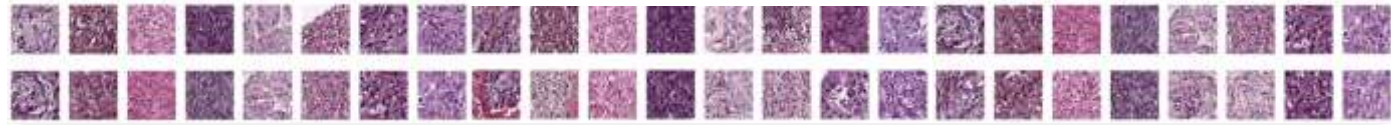
- AstraZeneca



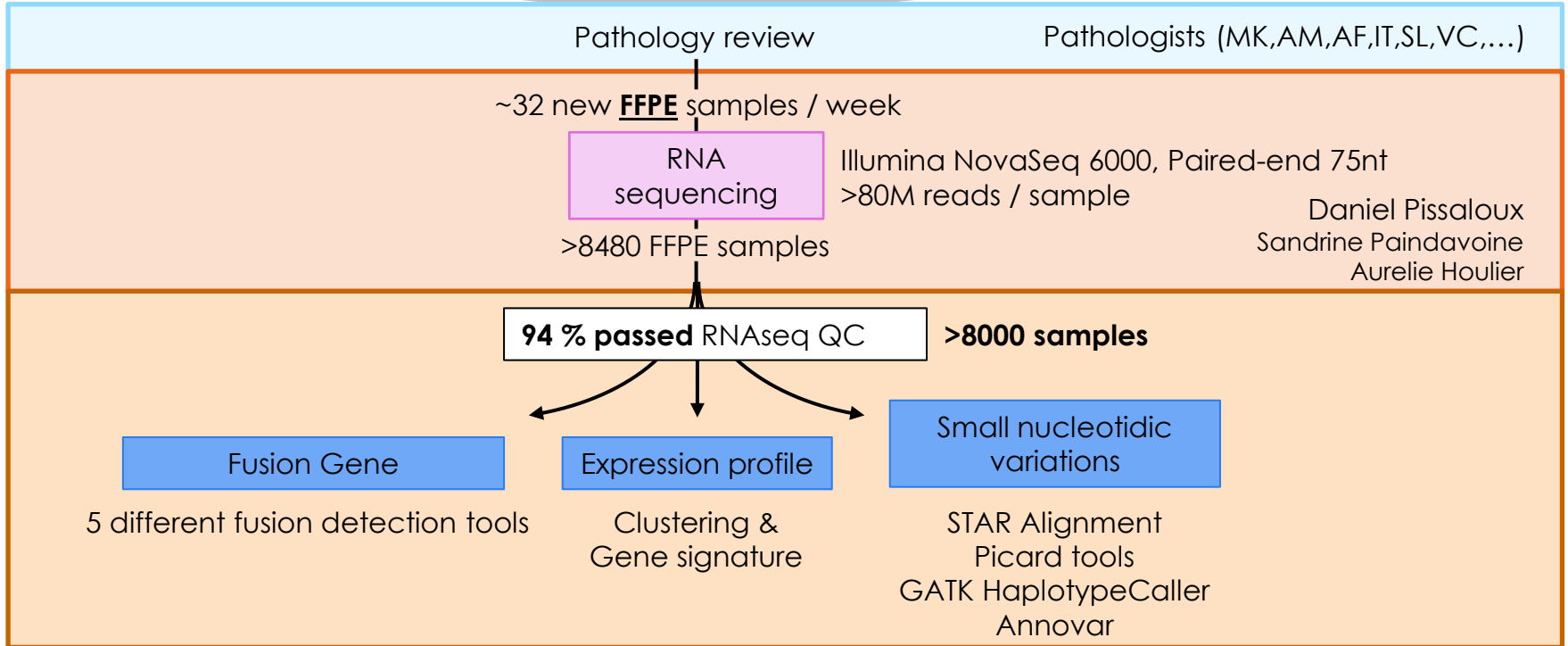
# RNAseq for molecular diagnosis @ CLB



# RNAseq for molecular diagnosis @ CLB



FFPE samples

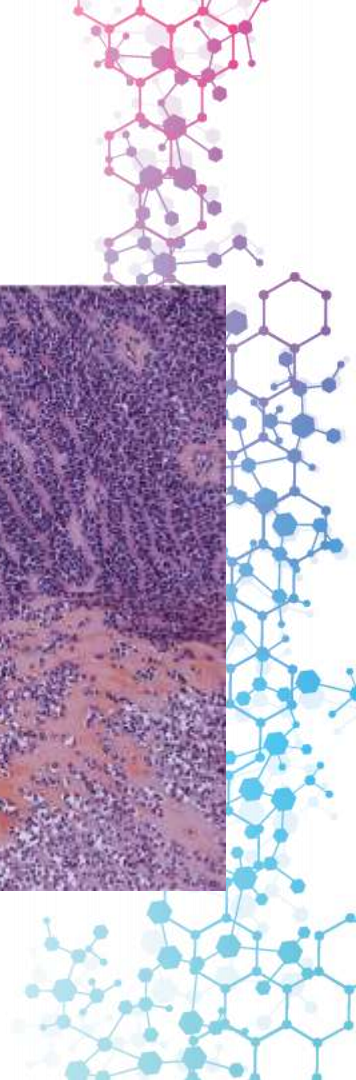
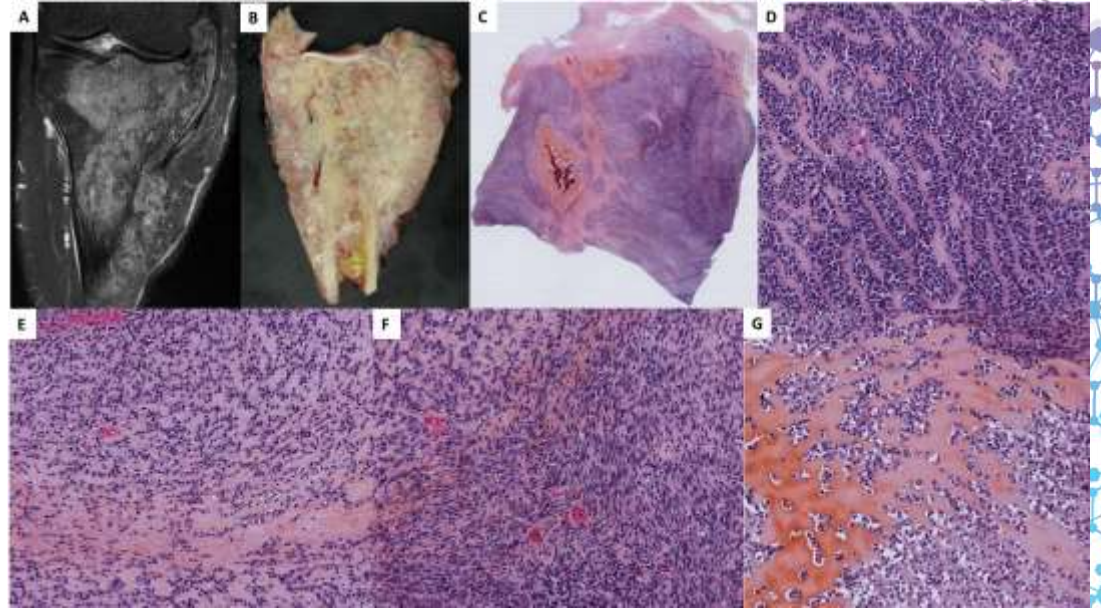


# Case R5997 example

21yo Male with pains for 2 years in the limb

Tumor mass in the in tibia of 115mm

Suspected to be a malignant myoepithelioma



# Case R5997 example

Pathology review

Pathologists (FFL)

RNA  
sequencing

Illumina NovaSeq 6000, Paired-end 75nt  
62 647 092 mapped reads  
Duplicates rate: 0.47

Daniel Pissaloux  
Sandrine Paindavoine  
Aurelie Houlier

Fusion Gene

5 different fusion detection tools

Expression profile

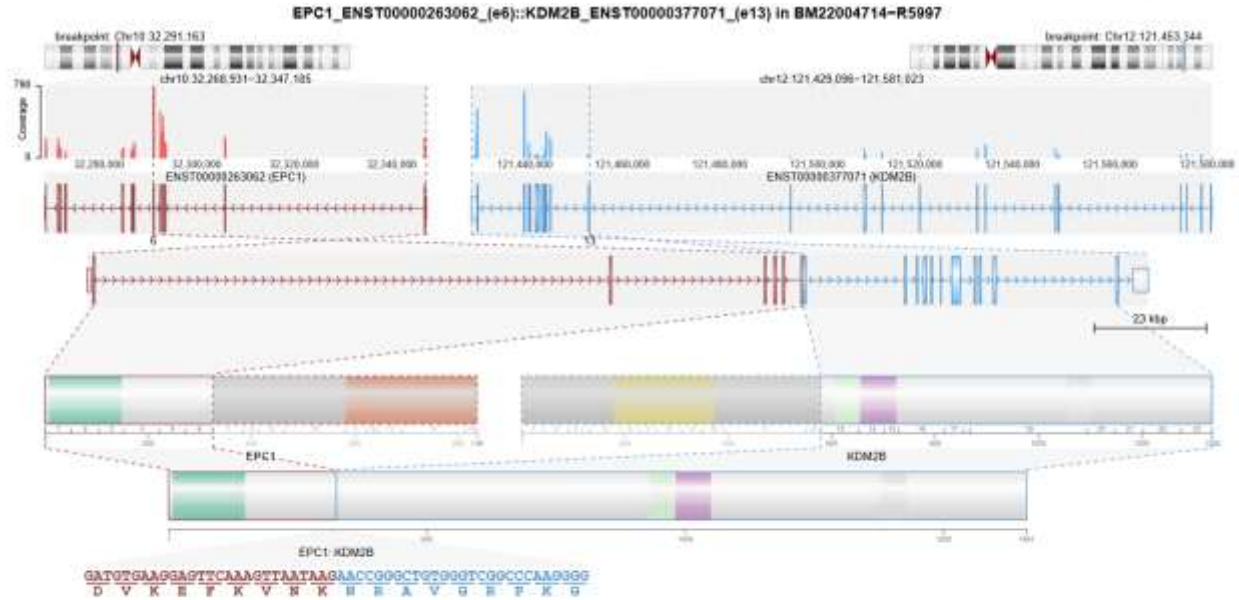
Clustering &  
Gene signature

Small nucleotidic  
variations

STAR Alignment  
Picard tools  
GATK HaplotypeCaller  
Annovar

# Fusion detection: Case R5997 example

FusionGene	Left_break_point	Right_break_point	nb_tools	FM	SF	ER	FC	AR	SplitReads	SpanReads	Frame	Transcripts
<b>EPC1::KDM2B</b>	<b>10:32291163:-</b>	<b>12:121453344:-</b>	<b>5</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>16</b>	<b>21</b>	<b>InFrame</b>	EPC1_ENST00000263062_(e6)::KDM2B_ENST00000377071_(e13)
LMO3::KERA	12:16560413:-	12:91057794:-	5	1	1	1	1	1	88	112	NoFrame	LMO3_ENST00000537304_(e3)::KERA_ENST00000266719_(230bp_INset_e1_five_prime_utr(1))
Known_Fusion:GRIP1::ATP2B1	12:66596847:-	12:89656107:-	3	0	1	1	0	1	1	1	e	GRIP1_ENST00000359742_(e2)::ATP2B1_ENST00000428670_(2_five_prime_utr(0))
Known_Fusion:IL10RB::IFNAR2	21:33241840:-	21:33241840:-	2	0	1	1	0	0	3	4	NoFrame	IL10RB_ENST00000290200_(e2)::IFNAR2_ENST00000342136_(2_five_prime_utr(0))



■ Enhancer of polycomb-like, N-terminal
 ■ Enhancer of polycomb, C-terminal

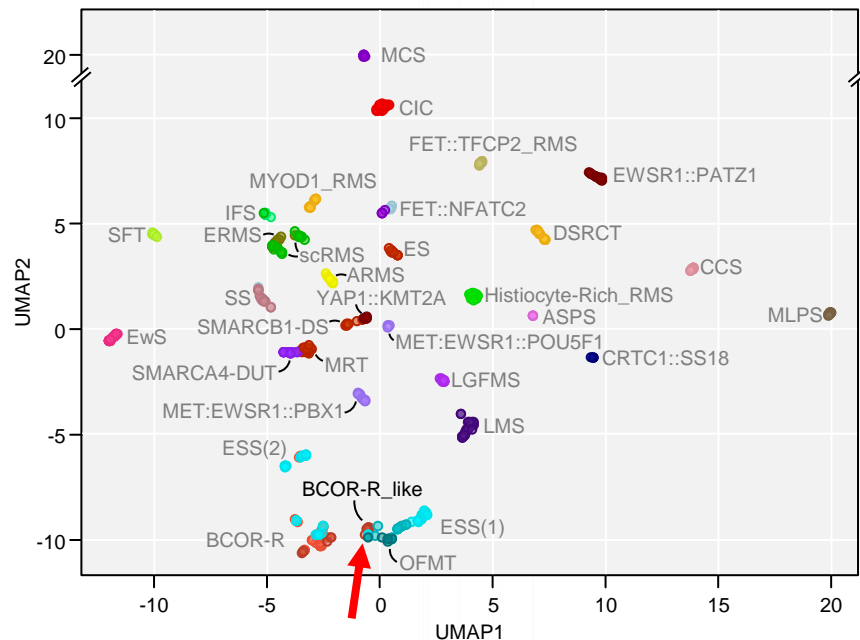
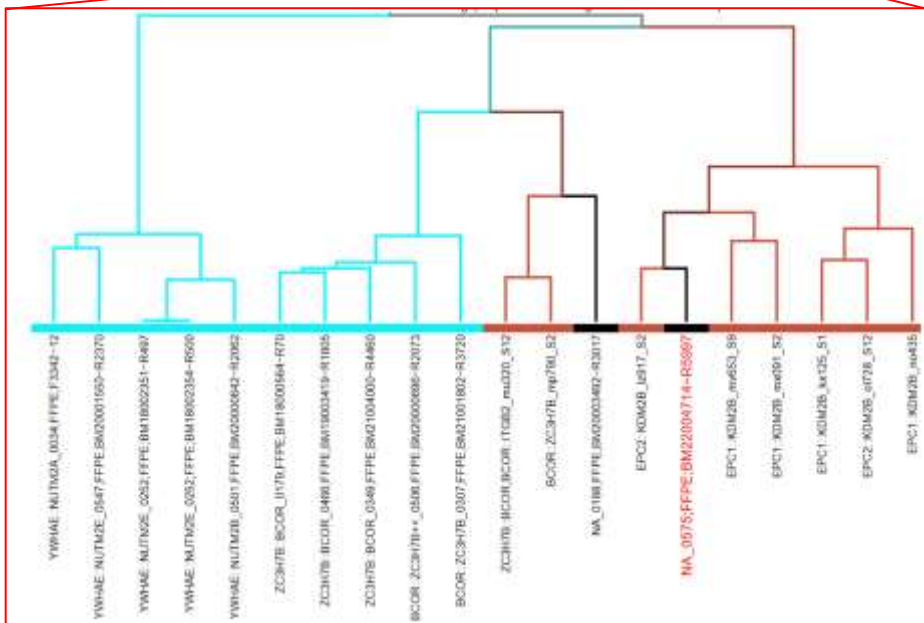
■ JmjC domains
 ■ Zinc finger, PHD-finger  
■ Zinc finger, CXXC-type
 ■ F-box domain



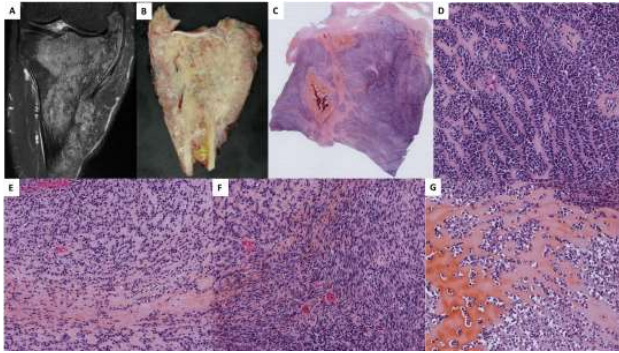
# Variation detection: Case R5997 example

Selected_AAChange_refGenes	SNV_abbreviation	ExonicFunc.refGene	cosmic_coding96	CLINSIG_20220320	AltCov	AAR	Chr	Start	End	Ref	Alt
FAT1:NM_005245:exon10:c.G6764A:p.R2255H	FAT1.pR2255H	nonsynonymous SNV	NA	NA	169	0.565217	chr4	186619822	186619822	C	T
ARNT:NM_001668:exon16:c.T1551G:p.D517E	ARNT.pD517E	nonsynonymous SNV	NA	NA	144	0.537313	chr1	150817388	150817388	A	C
HLA-A:NM_002116:exon5:c.TG899_900CA:p.L300P	HLA-A.pL300P	nonframeshift substitution	NA	NA	108	0.473684	chr6	29944503	29944504	TG	CA
BIRC6:NM_016252:exon55:c.C10807G:p.L3603V	BIRC6.pL3603V	nonsynonymous SNV	NA	Benign	102	0.435897	chr2	32515228	32515228	C	G
RABEP1:NM_004703:exon7:c.A815G:p.Q272R	RABEP1.pQ272R	nonsynonymous SNV	NA	NA	74	0.471338	chr17	5350481	5350481	A	G
KDM5A:NM_001042603:exon2:c.G206A:p.R69H	KDM5A.pR69H	nonsynonymous SNV	NA	NA	46	0.380165	chr12	385934	385934	C	T
TET1:NM_030625:exon2:c.A1828G:p.K610E	TET1.pK610E	nonsynonymous SNV	COSV65385743	NA	27	0.421875	chr10	68574166	68574166	A	G
RFWD3:NM_018124:exon2:c.G258T:p.L86F	RFWD3.pL86F	nonsynonymous SNV	NA	NA	22	0.5	chr16	74661192	74661192	C	A
NSD3:NM_023034:exon13:c.C2261T:p.S754L	NSD3.pS754L	nonsynonymous SNV	NA	Benign	15	0.483871	chr8	38305427	38305427	G	A
SPEN:NM_015001:exon11:c.T6021A:p.D2007E	SPEN.pD2007E	nonsynonymous SNV	COSV65367003	NA	11	0.5	chr1	15932261	15932261	T	A
IKZF3:NM_012481:exon3:c.A104C:p.H35P	IKZF3.pH35P	nonsynonymous SNV	NA	NA	8	0.421053	chr17	39829446	39829446	T	G
PDGFB:NM_002608:exon6:c.C670T:p.R224W	PDGFB.pR224W	nonsynonymous SNV	NA	Benign/Likely_benign	8	0.347826	chr22	39225779	39225779	G	A
BCL11A:NM_022893:exon4:c.1381_1383del;p.S461del	BCL11A.pS461del	nonframeshift deletion	NA	NA	4	0.4	chr2	60461529	60461531	GCT	-

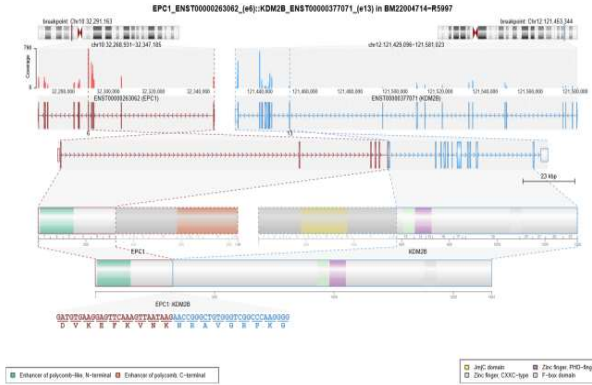
# Expression profiles : Case R5997 example



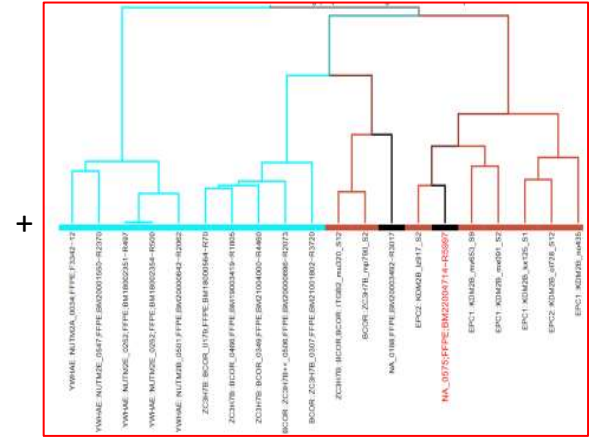
# Molecular diagnosis conclusion: Case R5997 example



Pathology



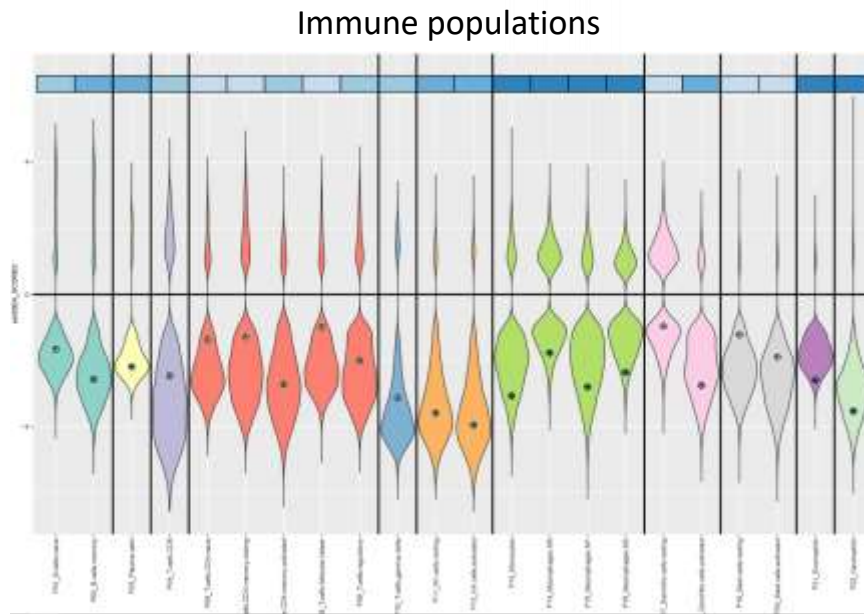
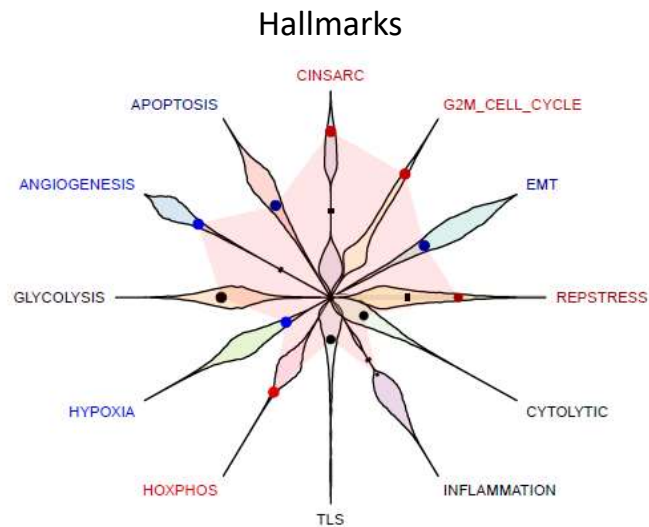
EPC1::KDM2B fusion



Clustering with BCOR-rearranged-like

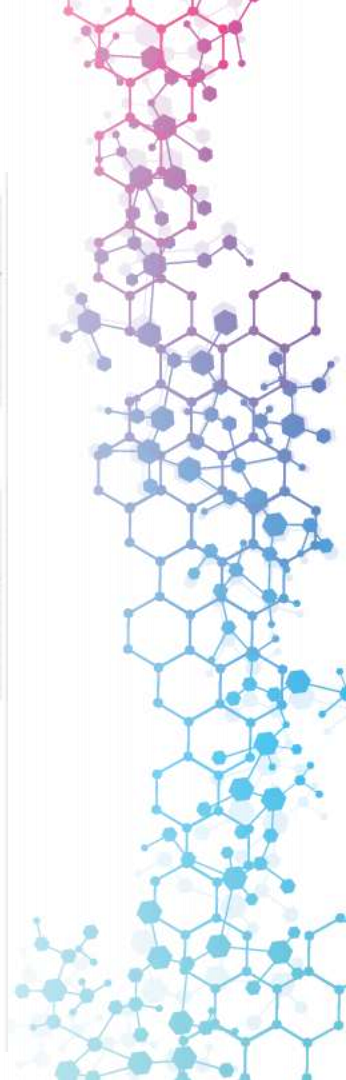
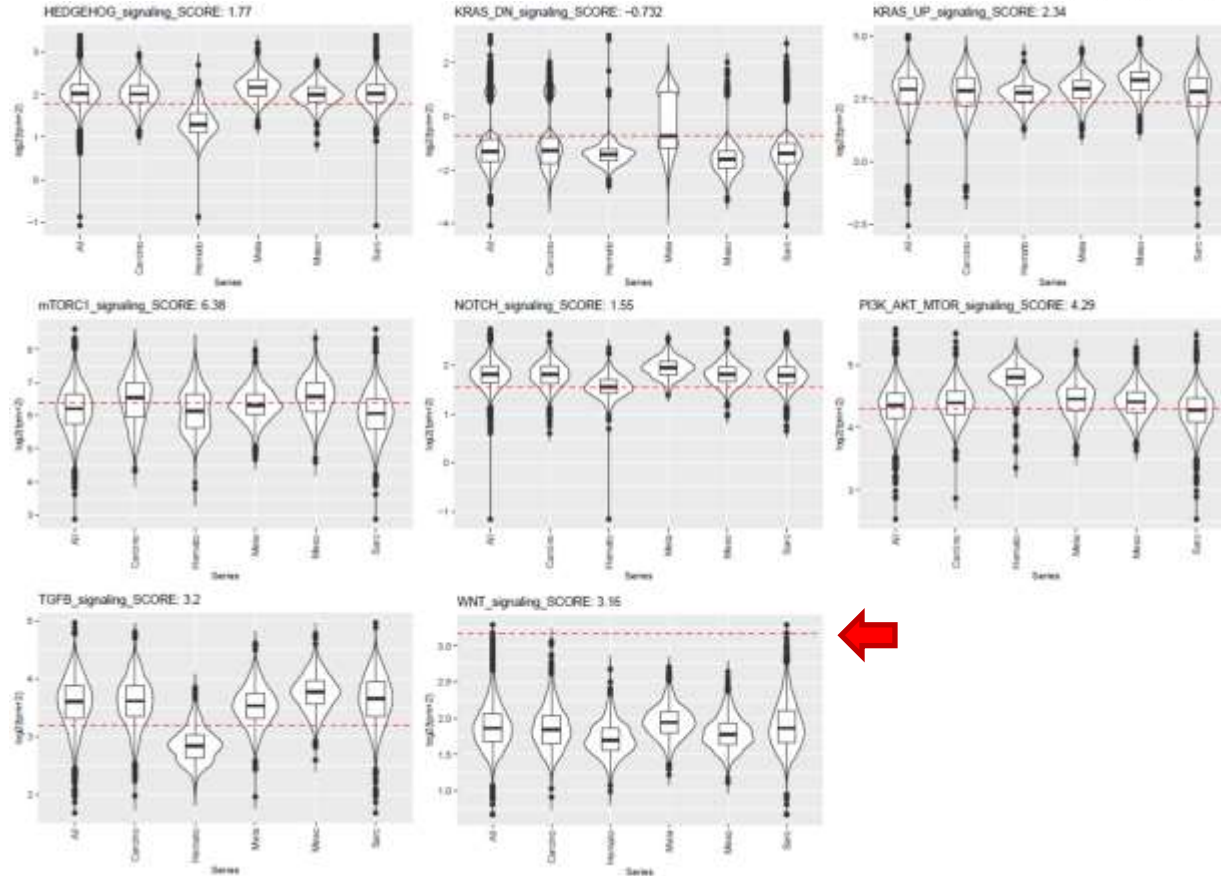
= BCOR-rearranged-like sarcoma

# Other information that we are taking into account: Hallmarks and immune signatures

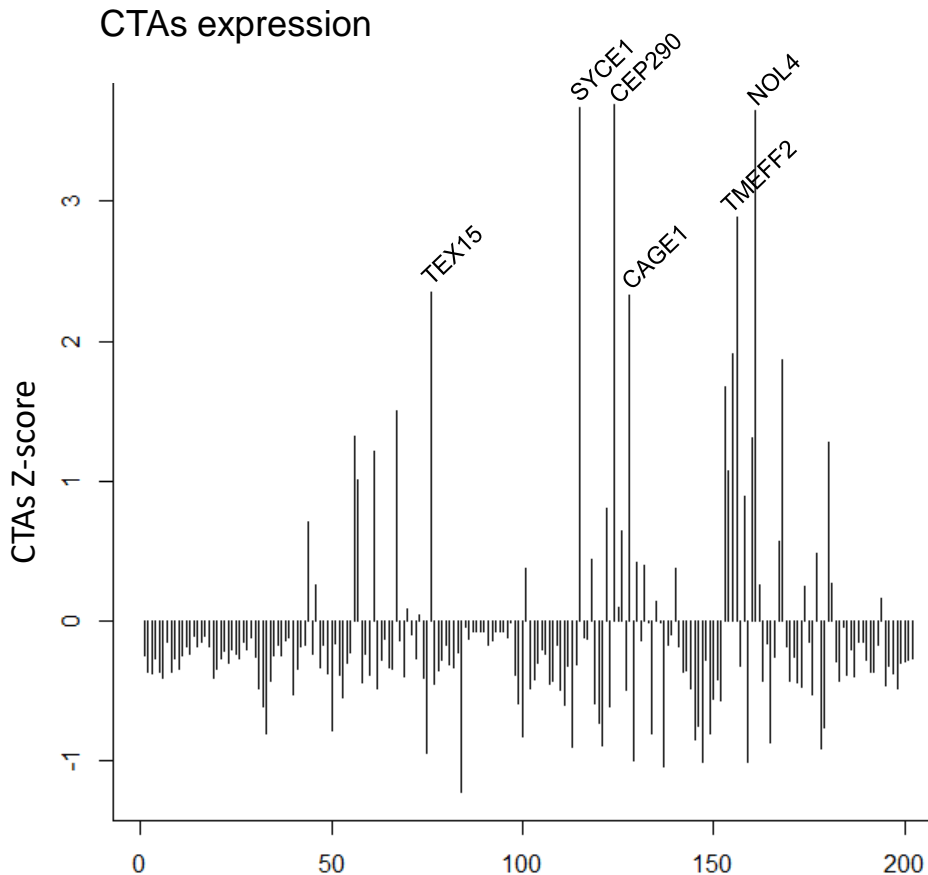


TMB: 3.82 (low)

# Other information that we are taking into account: Pathways signatures

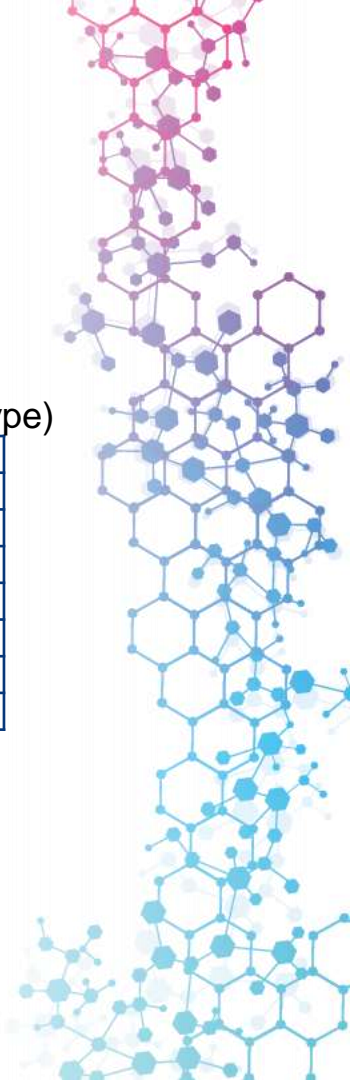


# Other information that we are taking into account

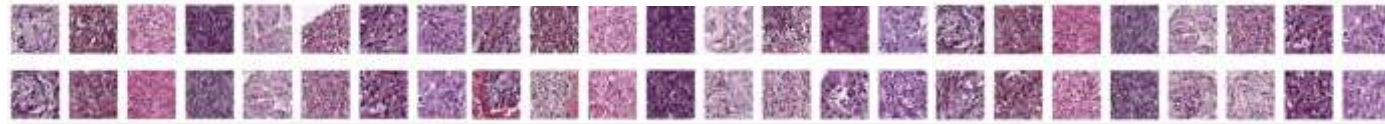


## HLA typing (optitype)

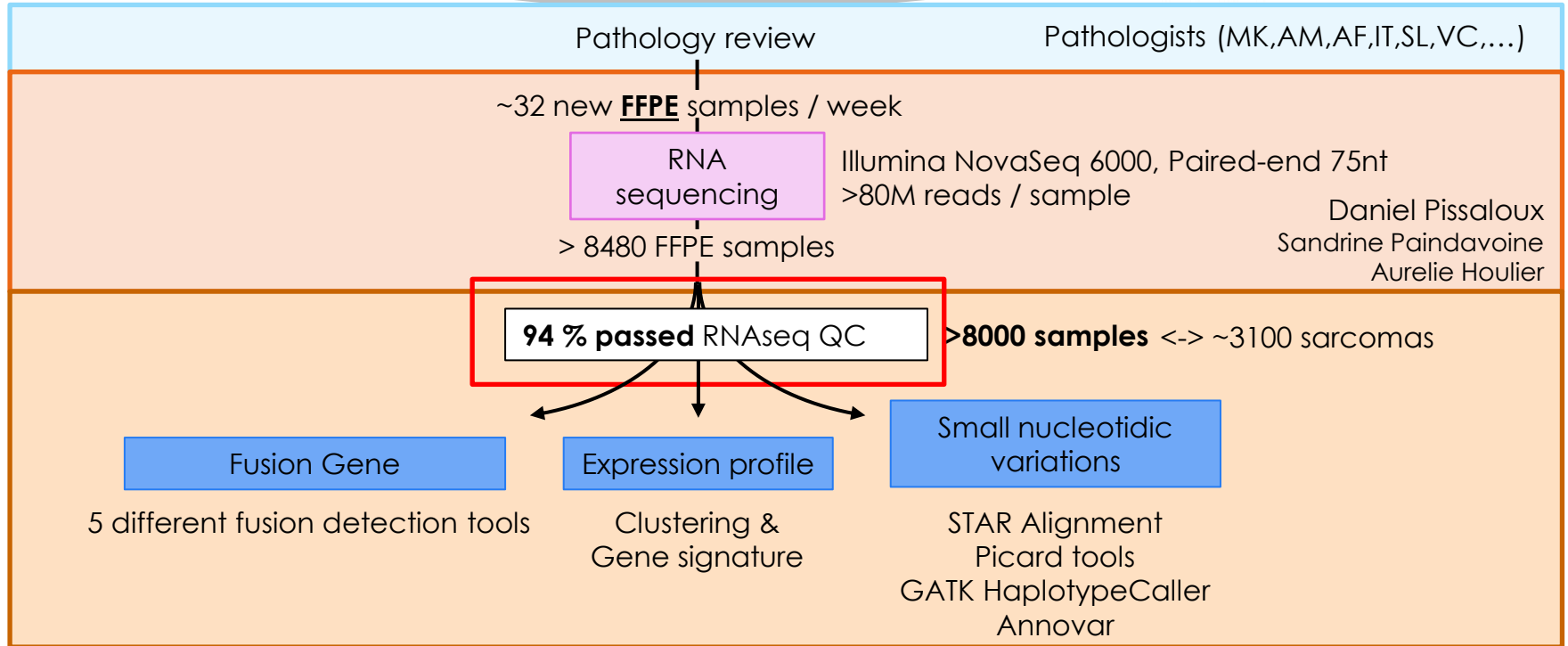
A1	A*24:02
A2	A*30:02
B1	B*44:05
B2	B*49:01
C1	C*02:02
C2	C*07:01
Reads	8054
Objective	7836.542



# RNAseq for molecular diagnosis @ CLB

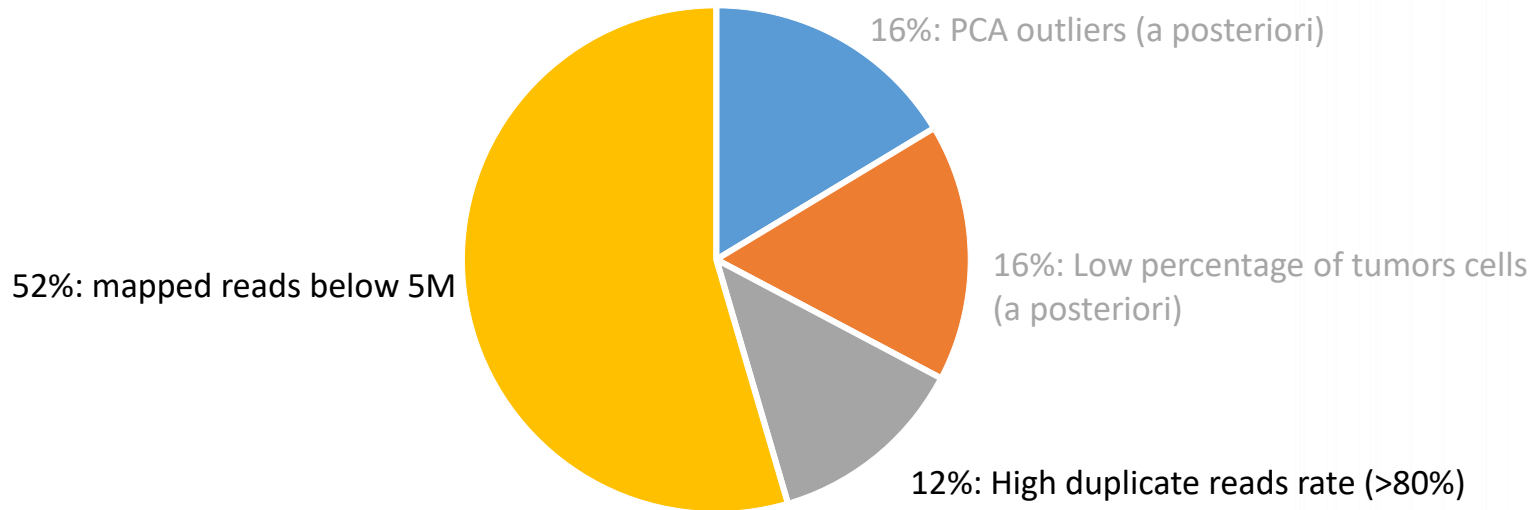


FFPE samples



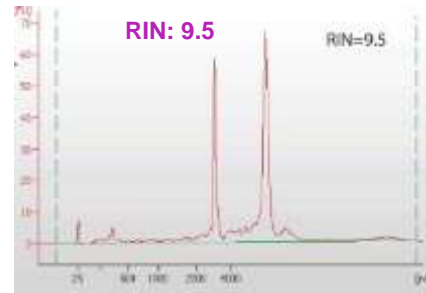
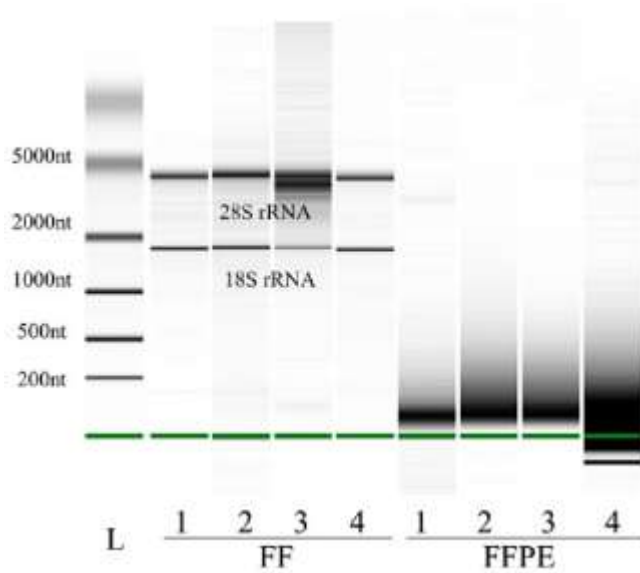
# QC controls

Out of 500 samples that did not pass QC:

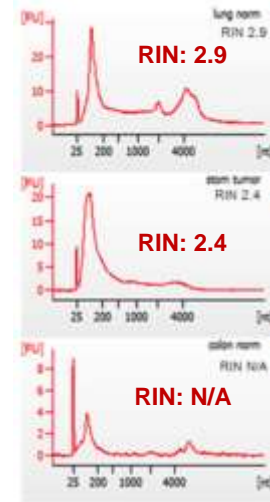




# Difference between FF and FFPE extracted RNA: 1) Degradation



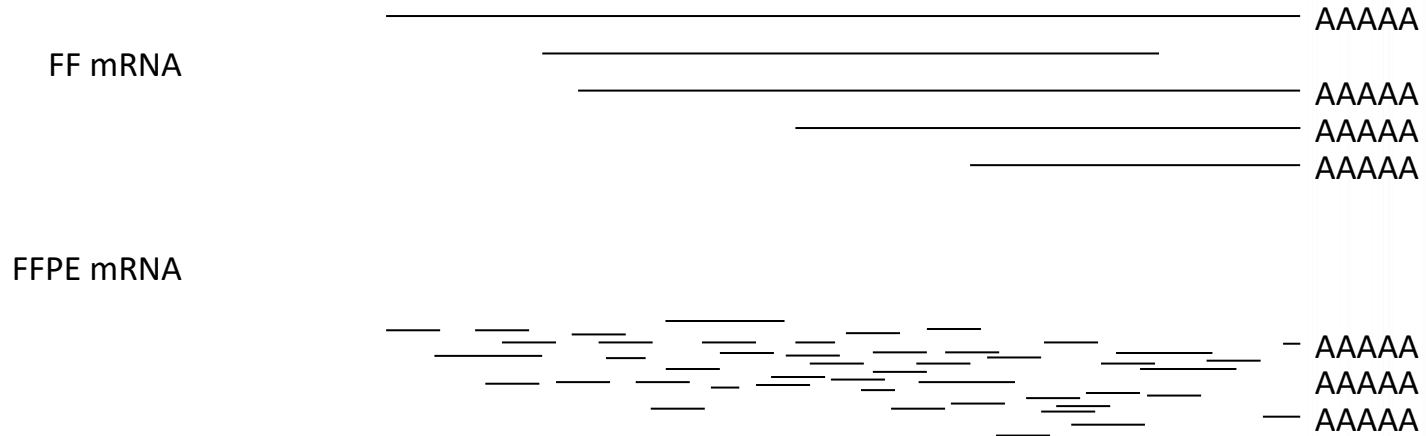
FF



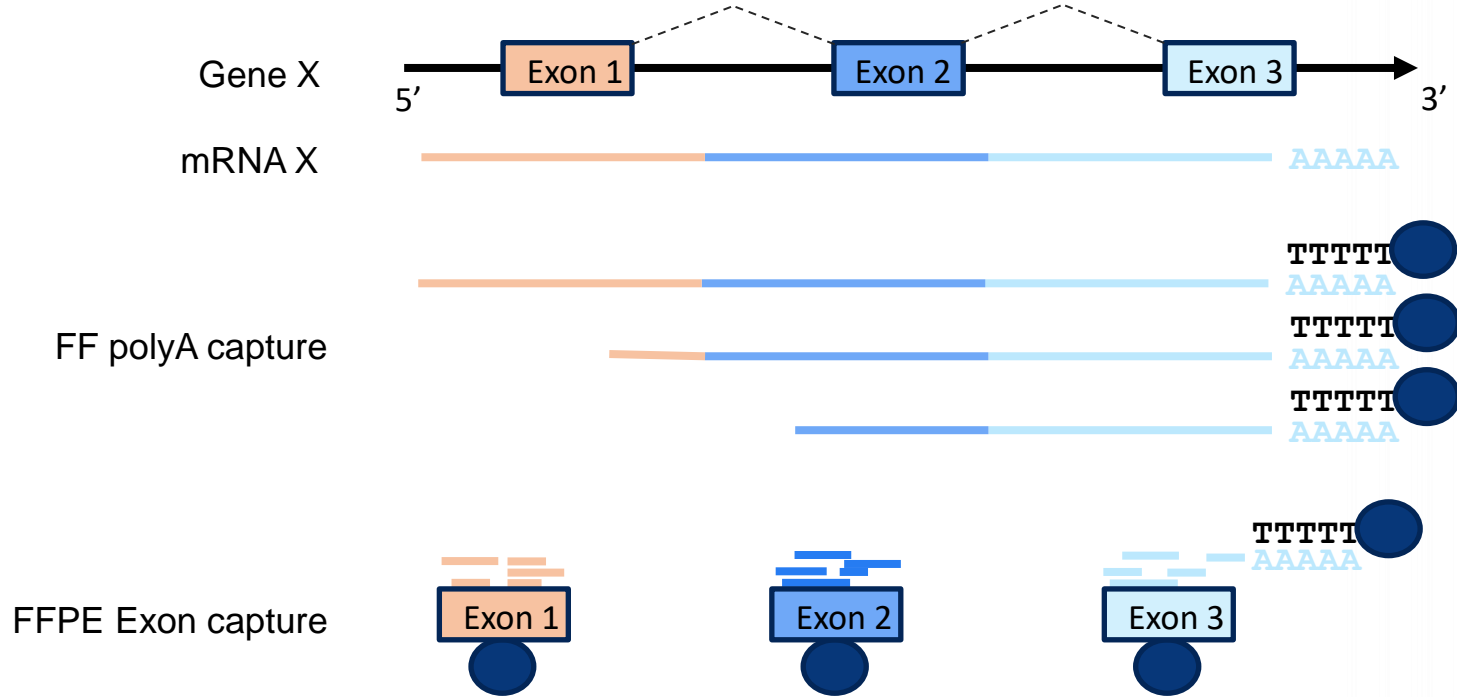
FFPE



# Difference between FF and FFPE extracted RNA: 1) Degradation



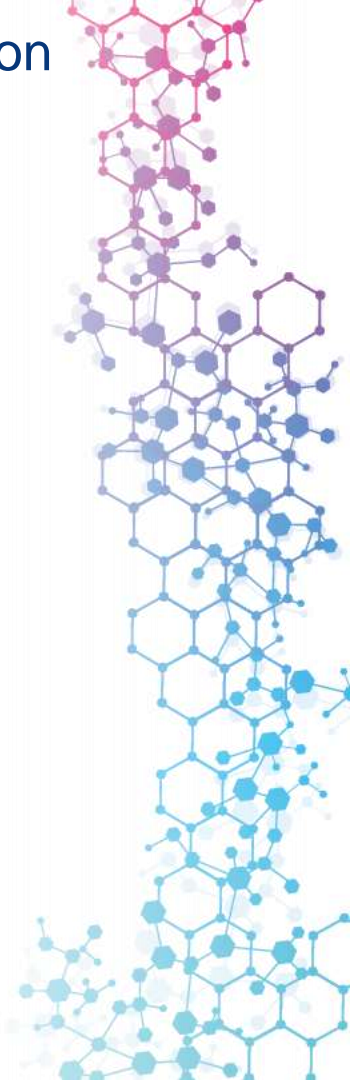
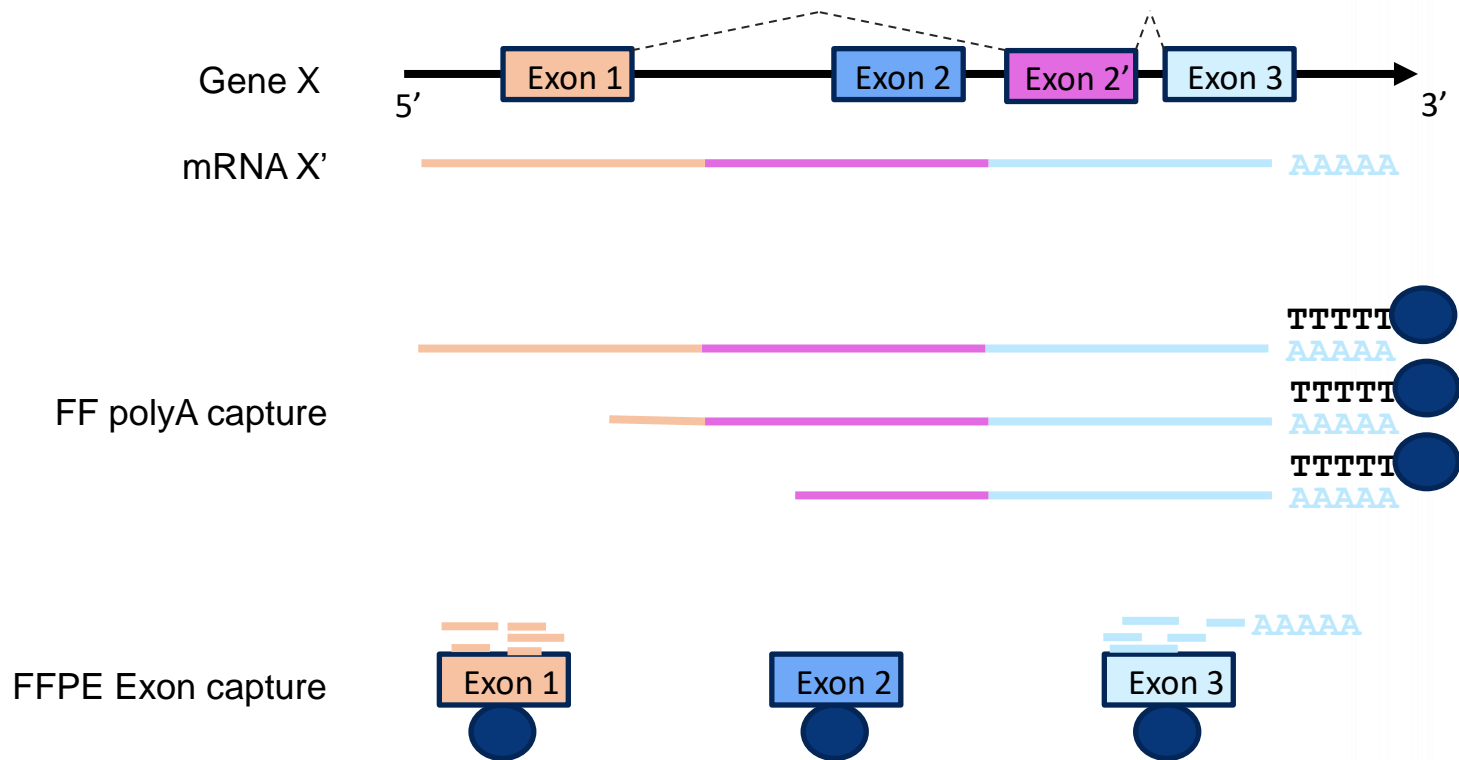
# Difference between FF and FFPE extracted RNA: 1) Degradation → Exon capture strategy



Limitation: You get only what you capture!



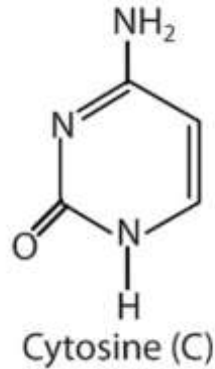
# Difference between FF and FFPE extracted RNA: 1) Degradation → Unknown genes or isoforms issue



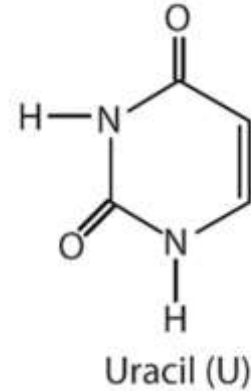
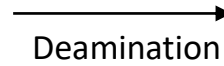
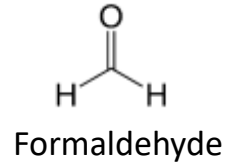
# Difference between FF and FFPE extracted RNA: 1) Degradation → Unknown genes or isoforms issue



# Difference between FF and FFPE extracted RNA: 2) Formol-induced deamination of C

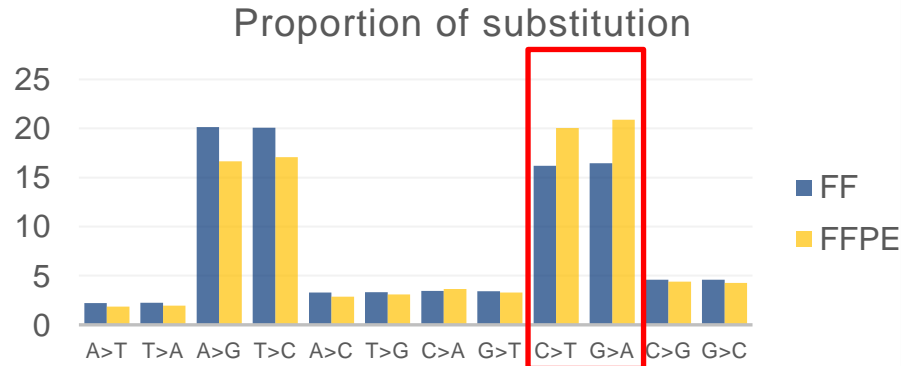


+



# Difference between FF and FFPE extracted RNA:

## 2) Formol-induced deamination of C



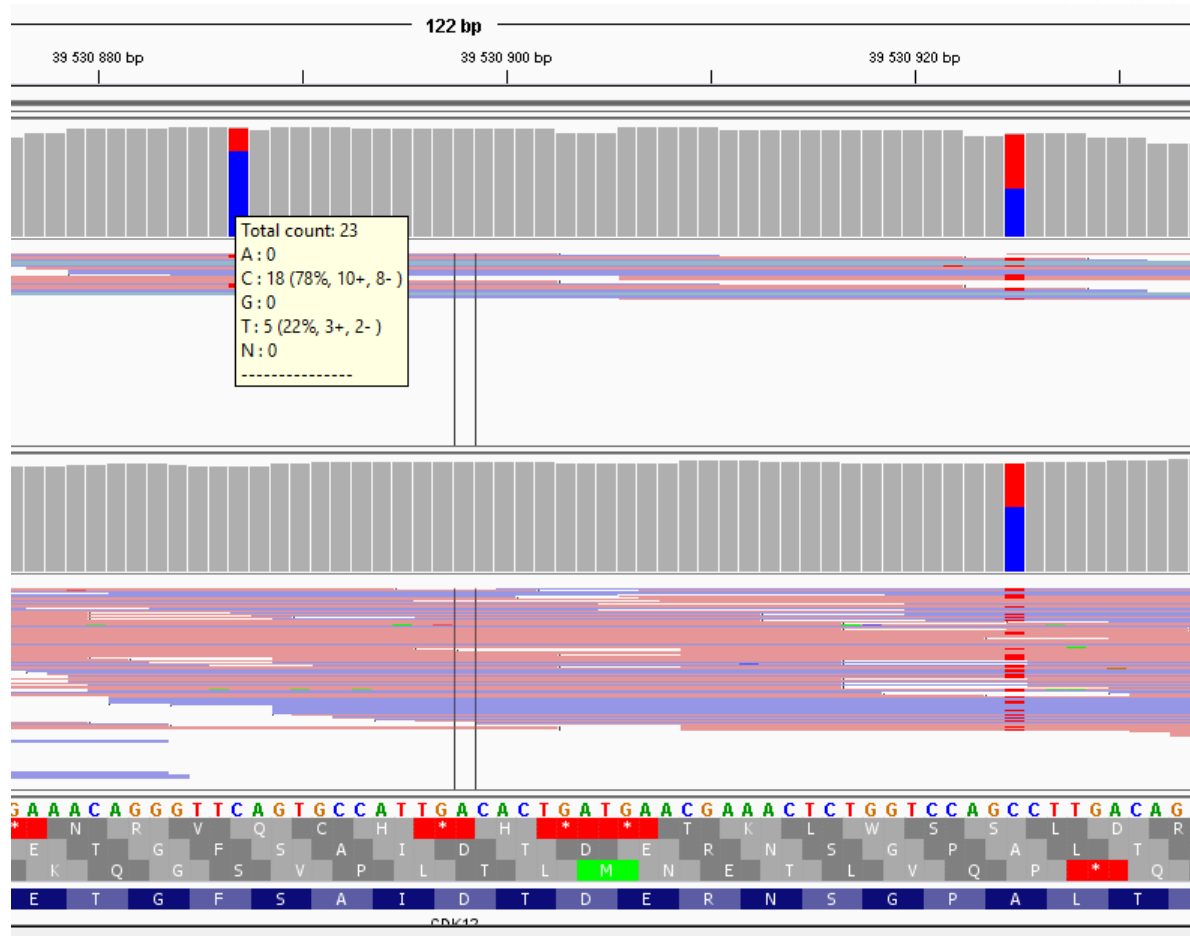
Higher mismatch number in reads = poorer quality of reads

➡ Increase total number of reads: 80M/samples

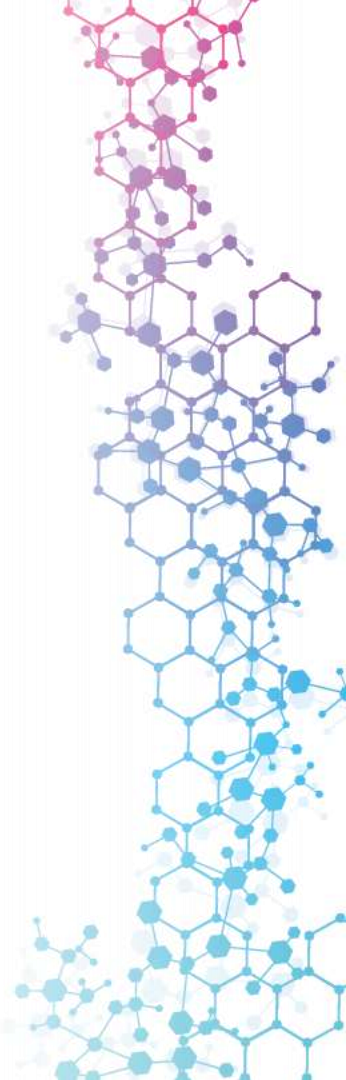


# Formol-induced deamination of C: True detection or artifact?

FFPE

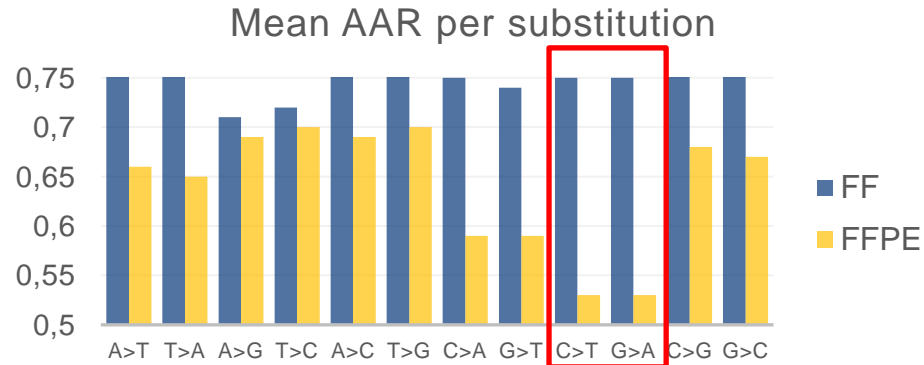
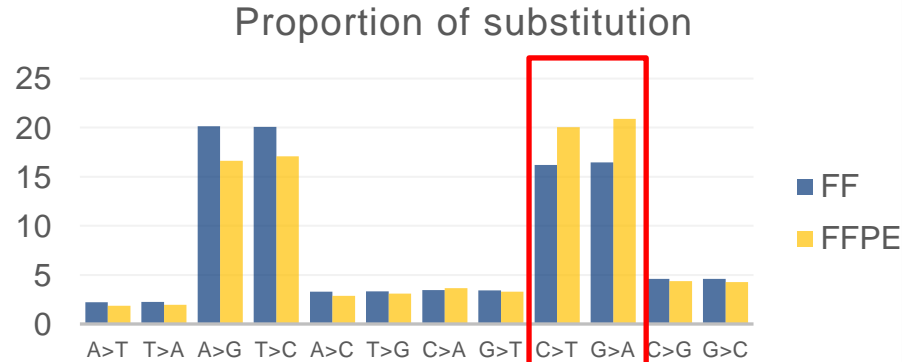


FF



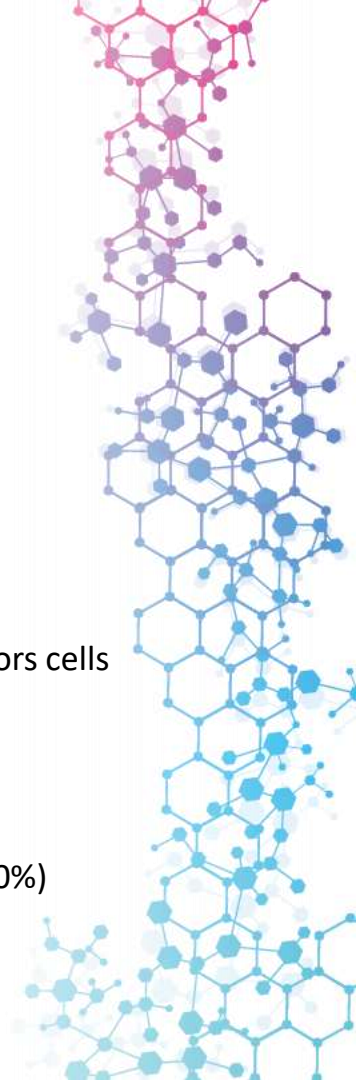
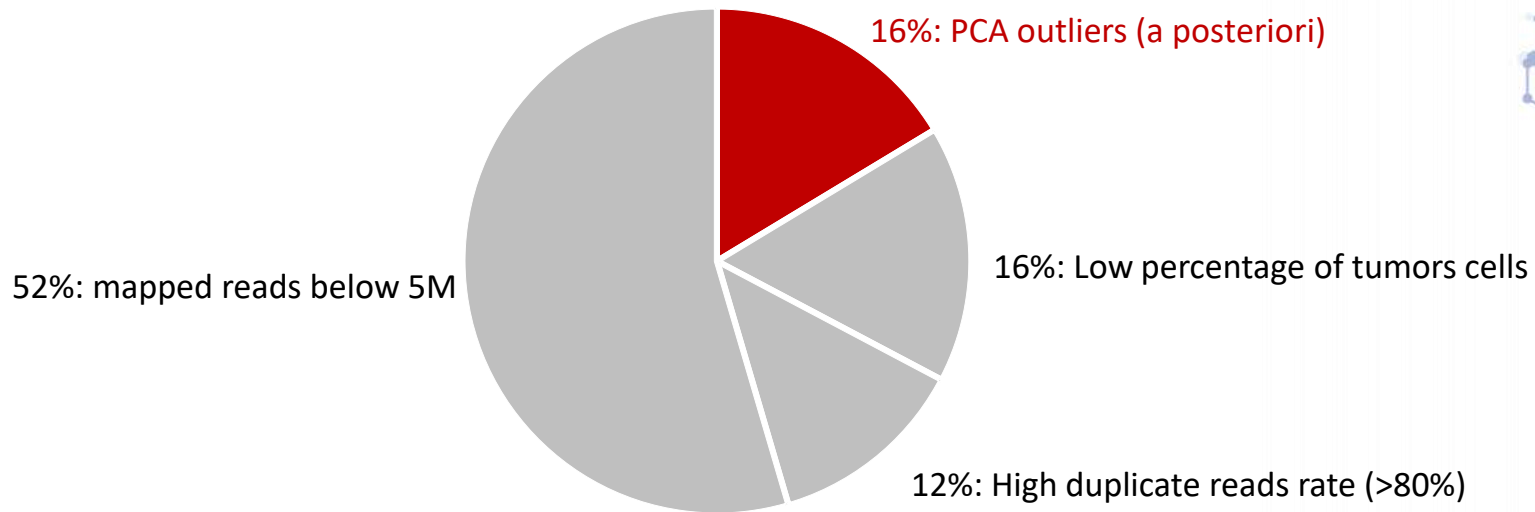


# Formol-induced deamination of C: True detection or artifact?



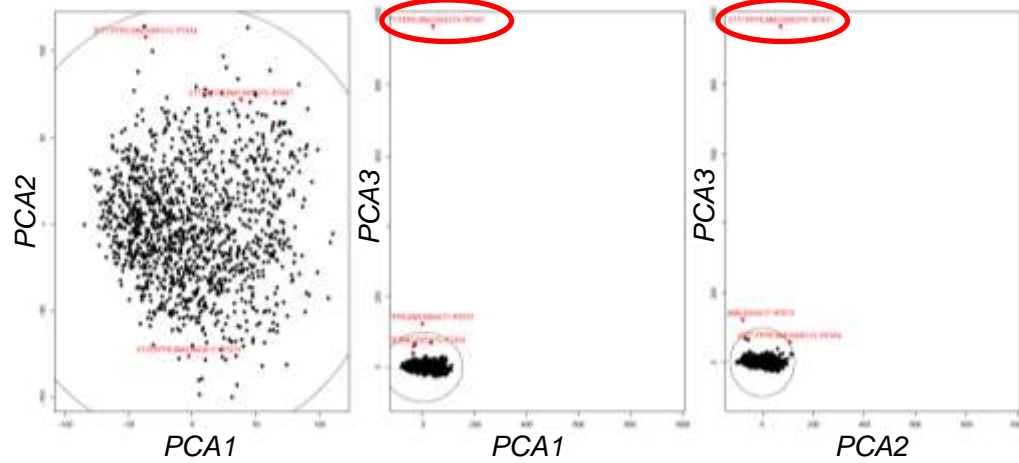
# QC controls

Out of 500 samples that did not pass QC:

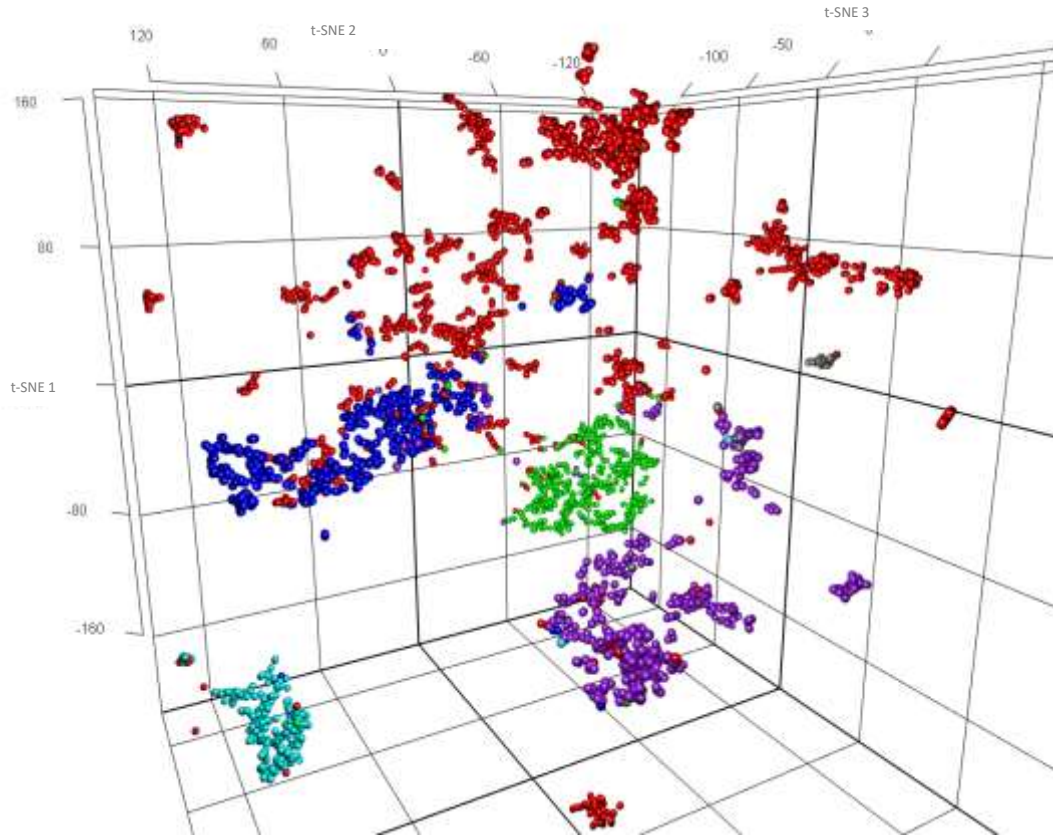


# QC controls: 3D PCA to identify outliers

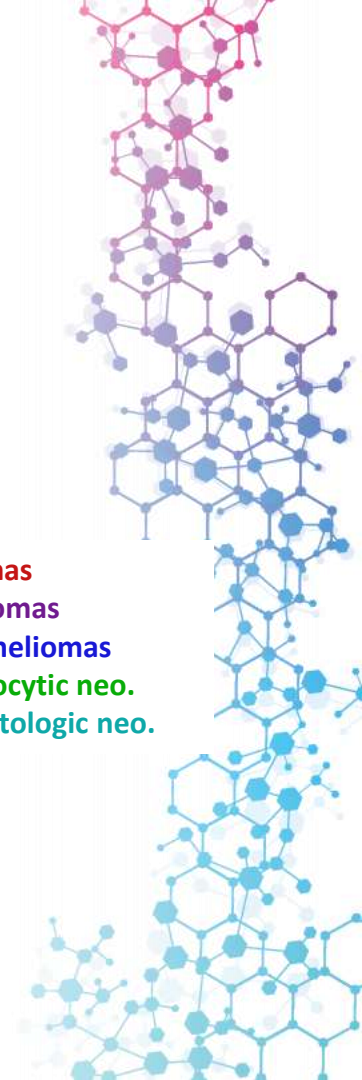
PCA: Sarc.



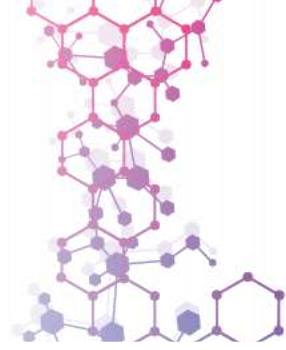
# QC controls: 3D PCA to identify outliers



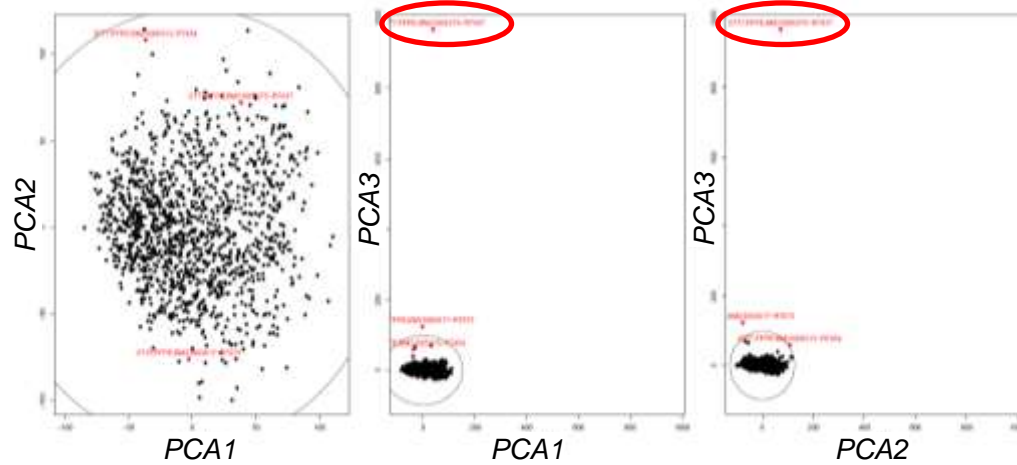
- Sarcomas
- Carcinomas
- Mesotheliomas
- Melanocytic neo.
- Haematologic neo.



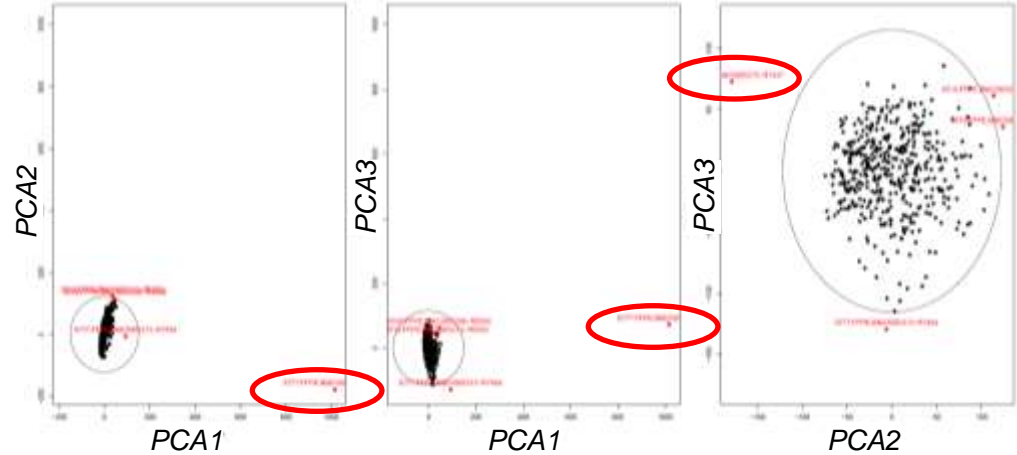
# QC controls: 3D PCA to identify outliers



PCA: Sarc.



PCA: Melan.



##### New Samples PCA Outliers - Distance from confidence #####

SAMPLEID	SARC	MELA	HEMATO	MESO	CARCINO
R7437	1000	600	200	300	400
R7454	1	5.5	NA	2	NA
R7440	NA	NA	2	4.5	NA
R7445	NA	NA	2	2	NA
R7458	NA	NA	3	2.5	NA
R7451	NA	NA	3.5	5.5	NA
R7460	NA	NA	1	4	NA



# QC controls: Number of reads thresholds issue ?

**Below 5M** unique mapped reads : **non-contributory**

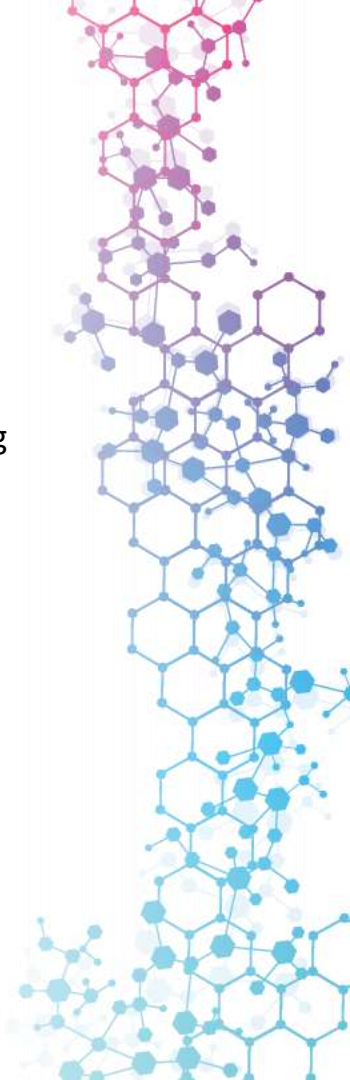
**Starting from 5M** unique mapped reads : suitable **only for expression** data & clustering

**Starting from 10M** unique mapped reads: **fusion** detection (but the more the better)

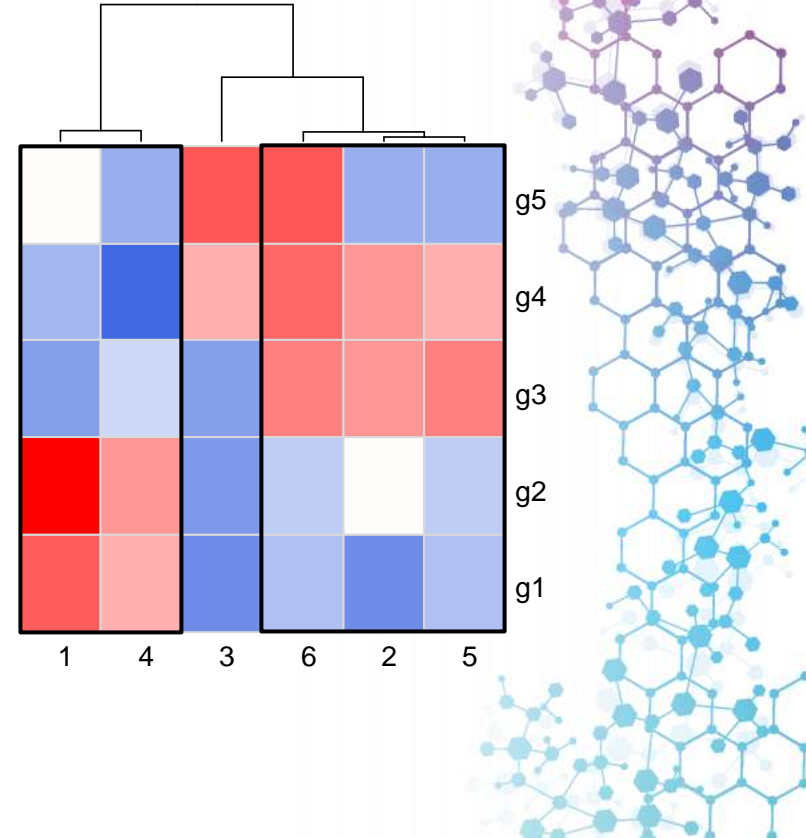
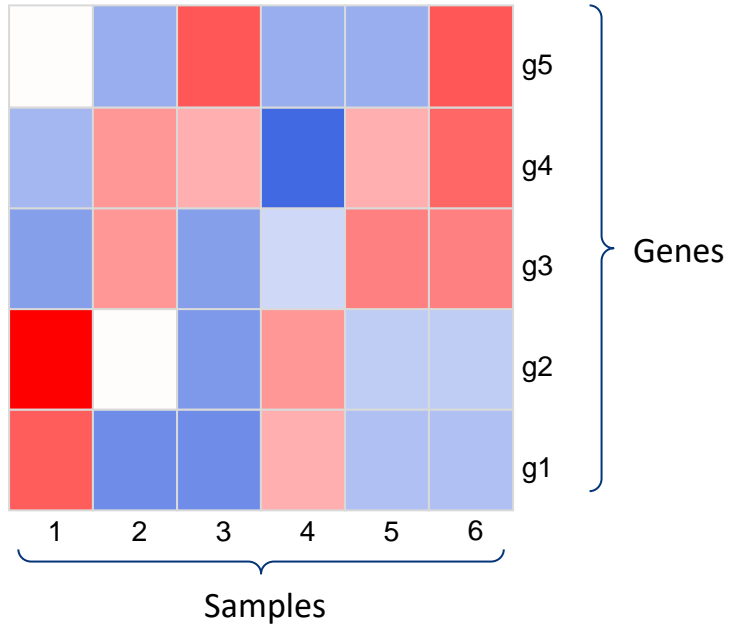
**SNV** detection should not be done **below 20M** unique mapped reads:

False negative because of low detection level

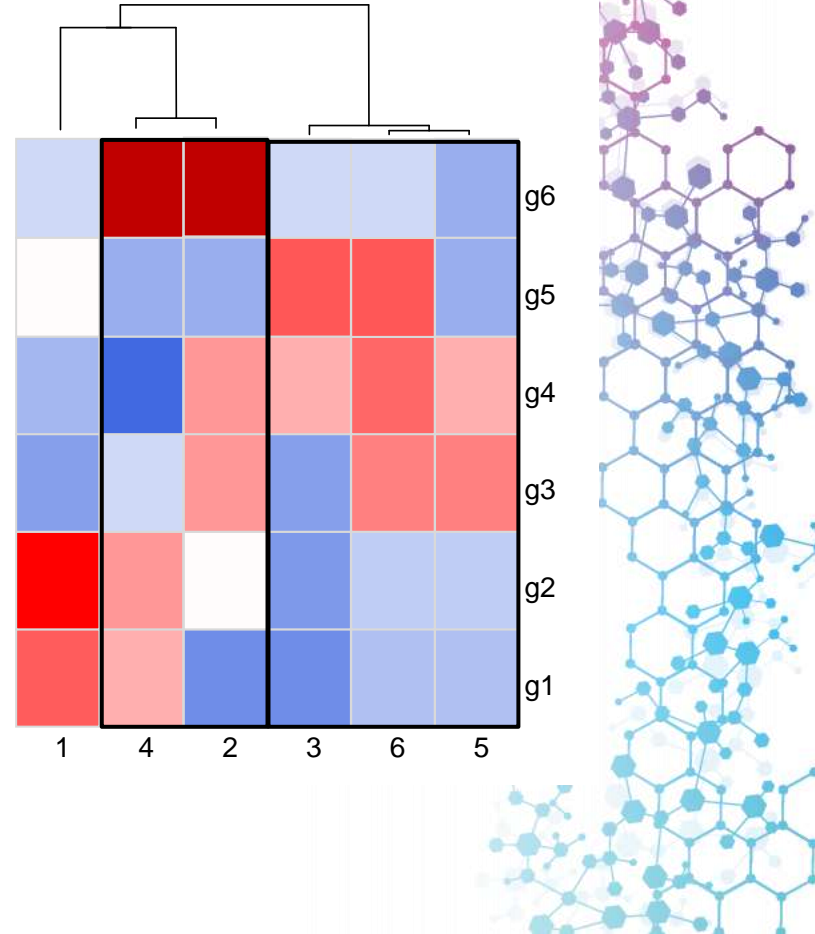
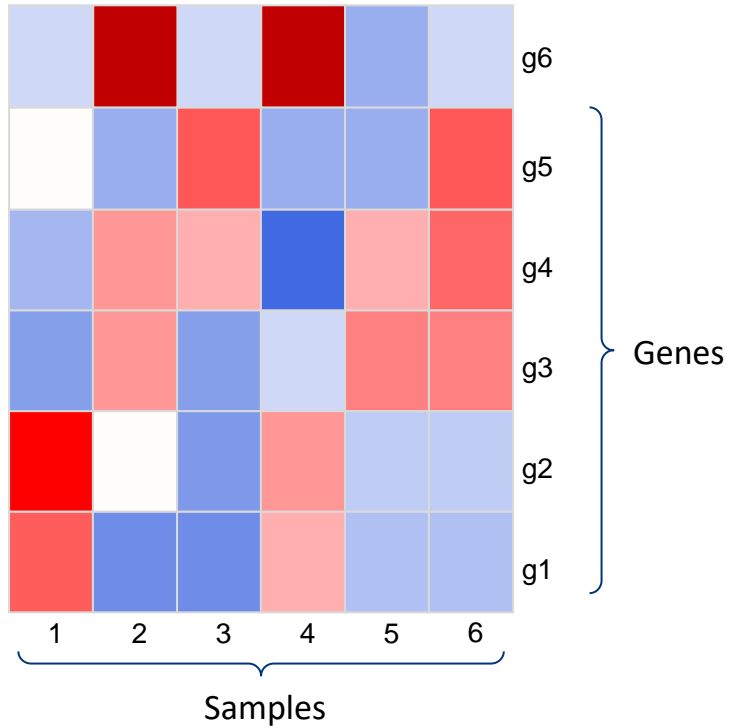
False positive because of higher AAR



# Expression profiles clustering issues

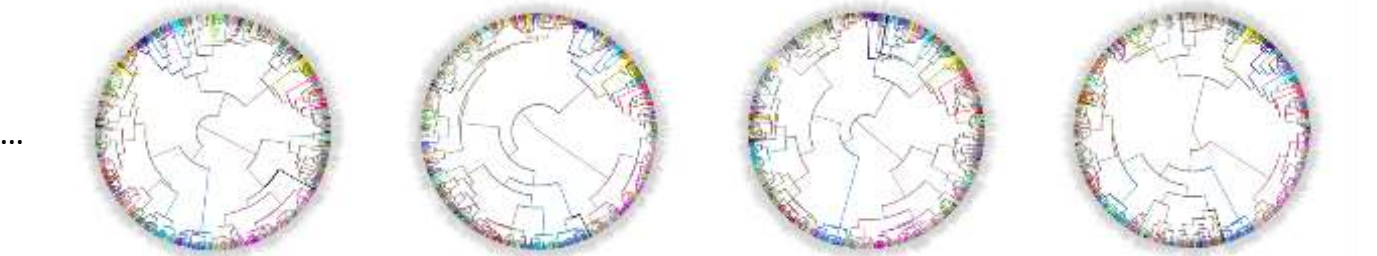


# Expression profiles clustering issues



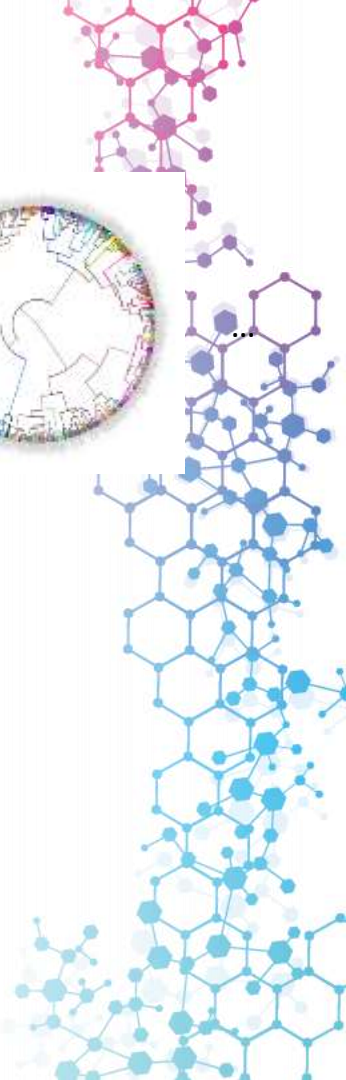
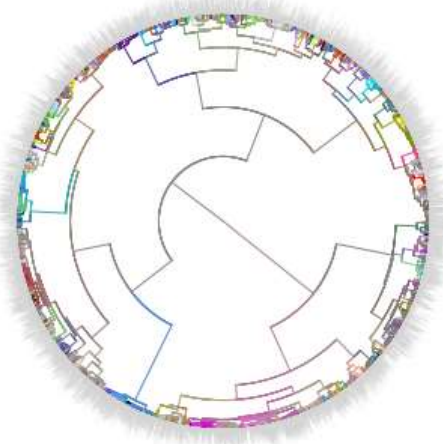


# Do not trust one cluster but trust clusterS



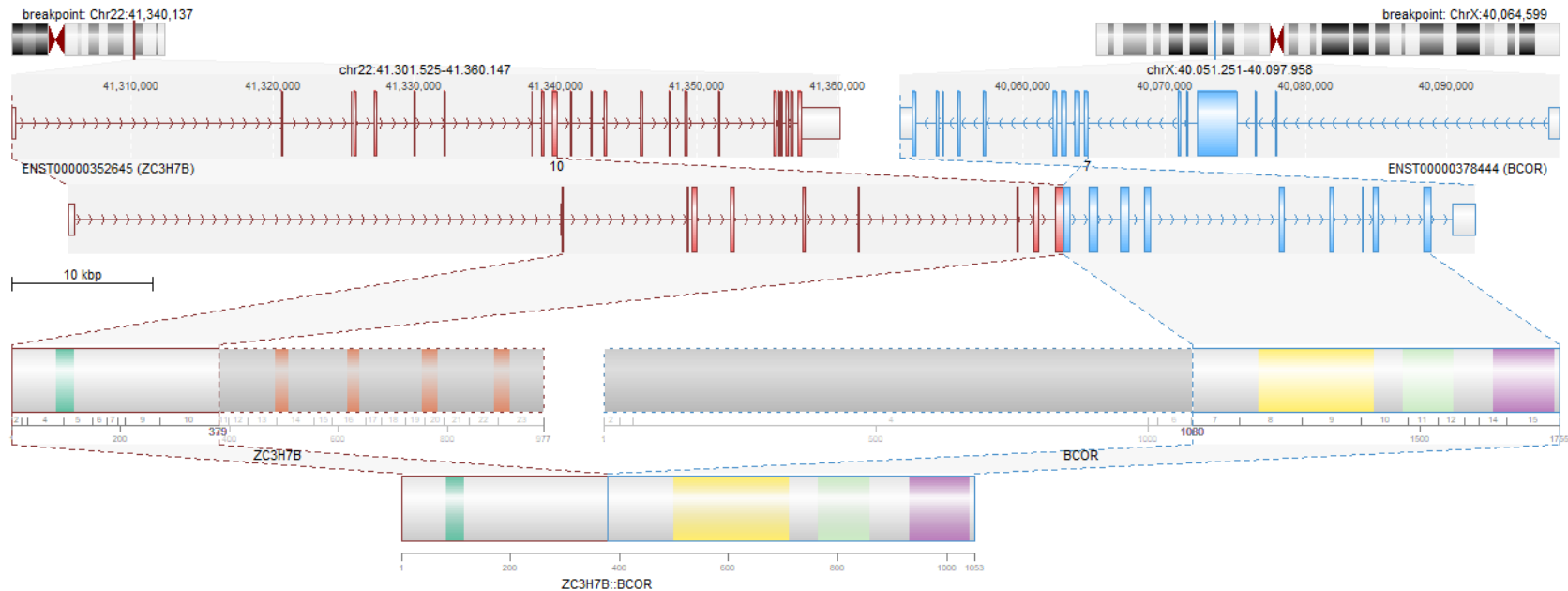
1000-5000 clusters omitting 20% of the samples and 20% of the genes

Consensus cluster →



# How to determine if a fusion is relevant?

ZC3H7B\_ENST00000352645\_(e10)::BCOR\_ENST00000378444\_(e7) in BCOR::ZC3H7B

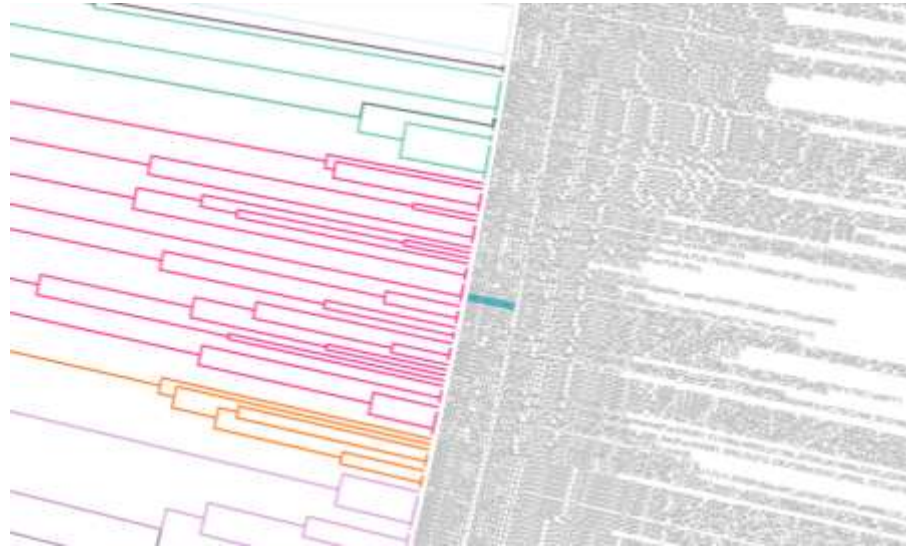


■ Tetratricopeptide repeat ■ Zinc finger, CCCH-type

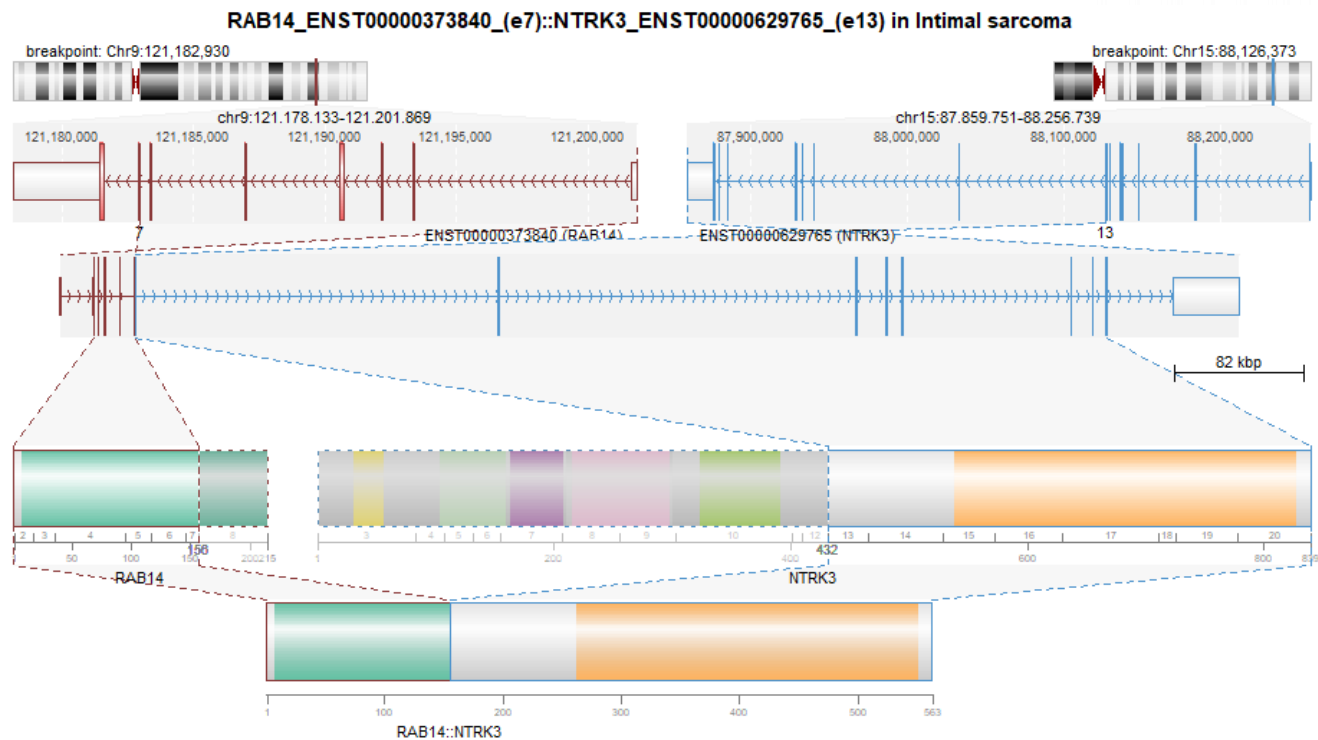
■ BCL-6 corepressor, non-ankyrin-repeat domain ■ BCL-6 corepressor, PCGF1 binding domain  
■ Ankyrin repeat

# How to determine if a fusion is relevant?

SampleID	FusionGene	Left_break_point	Right_break_point	nb_tools	FusionMap	STARFusion	ERICScript	FUSIONcatcher	ARRIBA	SplitReads	SpanReads	Frame	Transcripts	Known_in	Protein1_Domains	Protein2_Domains	
CLB_RNA_1594	EWSR1 :: FLI1	chr22 29287134 +	chr11 128781958 +	1	0	0	0	1	0	4	5	1	In Frame	EWSR1_ENST00000397938_(e7)::FLI1_ENST00000527786_(e5)	Ewing Sarcoma	264/656AA retained in fusion. RETAINED:None   LOST:{RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)}[Full] & {Zn-finger in Ran binding protein and others}[Full]	255/452AA retained in fusion. RETAINED:{Ets-domain}[Full] & {Sterile alpha motif (SAM)/Pointed domain}[0%]   LOST:{Sterile alpha motif (SAM)/Pointed domain}[Full] & {Sterile alpha motif (SAM)/Pointed domain}[100%]



# How to determine if a fusion is relevant?



■ Small GTPase

- |   |                         |
|---|-------------------------|
| ■ Leucine-rich repeat N-terminal domain                       | ■ Immunoglobulin        |
| ■ Leucine-rich repeat   | ■ Immunoglobulin I-set  |
| ■ Growth factor receptor NTRK, leucine rich repeat C-terminal | ■ Protein kinase domain |
| ■ Immunoglobulin-like domain                                  |                         |

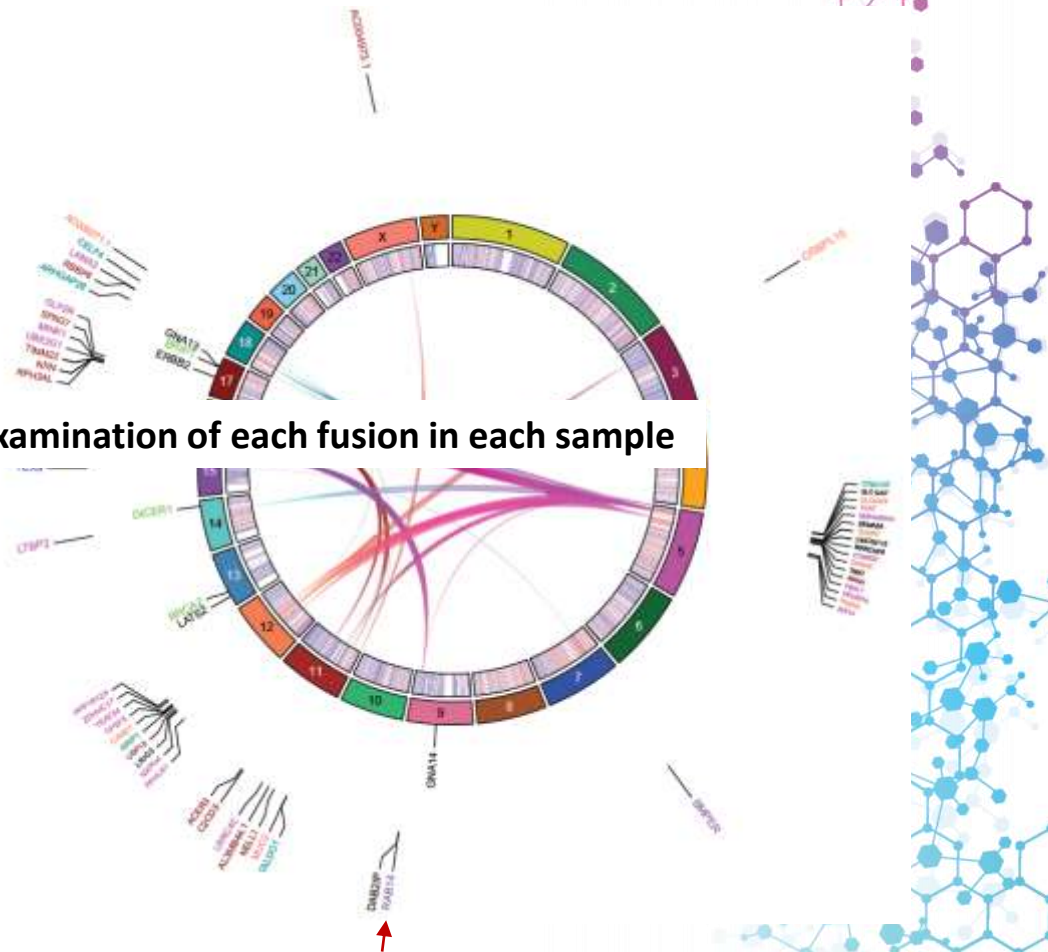


# How to determine if a fusion is relevant?

## Complex genomics

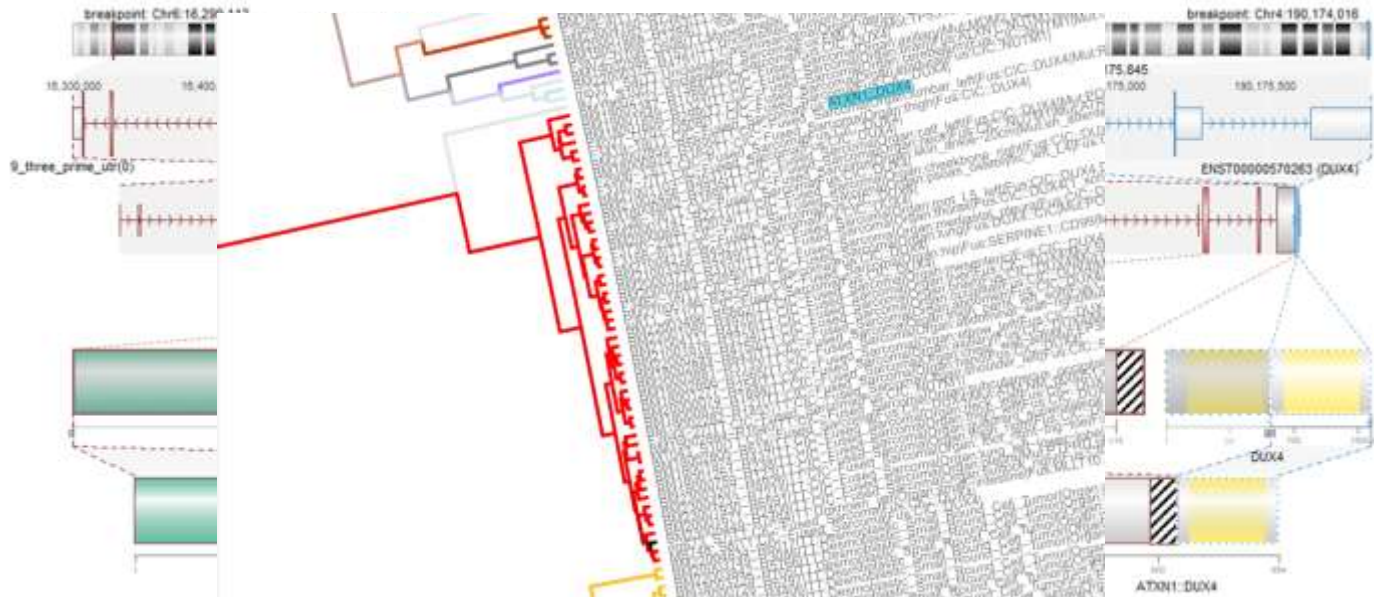
PPHLN1::DNAH5	12:42398994:+	5:13839528:-
OSBPL10::USP15	3:31980899:-	12:62294179:+
LRIG3::TAS2R1	12:58889996:-	5:9629992:-
PDZD2::ZDHC17	5:31855168:+	12:76797434:+
DROSHA::CTNND2	5:31511035:-	5:11199661:-
TRIO::GLP2R	5:14492814:+	17:9854495:+
NXP4::SLC12A7	12:57217026:+	5:1094248:-
CCDC127::LAMA3	5:216729:-	18:23819841:+
<b>RAB14::NTRK3</b>	<b>9:121182930:-</b>	<b>15:88126373:-</b>
LTBP2::SLC12A7	14:74535926:-	5:1094248:-
NTRK3::DAB2IP	15:88126271:-	9:121678678:+
NTRK3::RAB14	15:88126271:-	9:121193419:-
SEMA5A::LRRC4C	5:9224674:-	11:40933723:-
ANKH::UBE2G1	5:14704773:-	17:4271123:-
MARCHF6::MINK1		
FBXL7::SEMA5A		
SPAG7::C2CD3	17:4507720:-	17:4507720:-
TRIO::SEMA5A	5:14420021:+	5:9063105:-
RPH3AL::ACER3	17:219623:-	11:76926557:+
FBXL7::TRIO	5:15616072:+	5:14270825:+
LRIG3::CAND1	12:58913982:-	12:67281910:+
MARCHF6::MIR4458HG	5:10353917:+	5:8459933:+
LRIG3::LINC02112	12:58889996:-	5:9641941:-
RAI14::LINC02112	5:34803776:+	5:9714239:-
DAB2IP::TERT	9:121763879:+	5:1282625:-
NXN::NELL1	17:979319:-	11:21370853:+
YEATS4::SLC6A18	12:69370975:+	5:1232219:+
BMPER::TEX9	7:34086092:+	15:56373441:+
AC004973.1::MUC2	X:119416757:+[FC]	11:1094734:+
CPSF6::MARCHF6	12:69262562:+	5:10377798:+
PPP1R12A::PDZD2	12:79845302:-	5:31798889:+
TIMM22::AL358944.1	17:997380:+	11:30044455:-
AC009271.1::GRIP1	18:55893442:+	12:66465422:-
ARHGAP28::CELFA	18:6837414:+	18:37485607:-
UBE2G1::ANKH	17:4307021:-	5:14758598:-
TALDO1::RBBP8	11:747578:+	18:22997620:+

**There is no rule but a careful examination of each fusion in each sample**

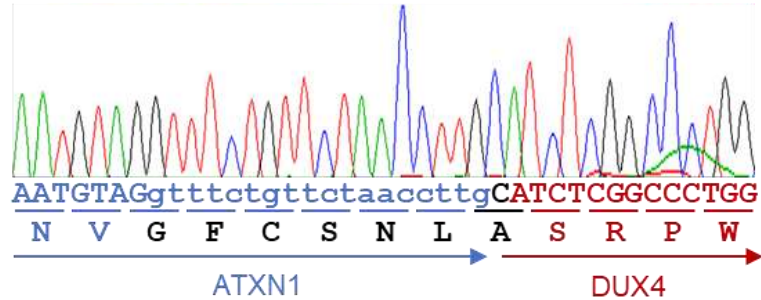
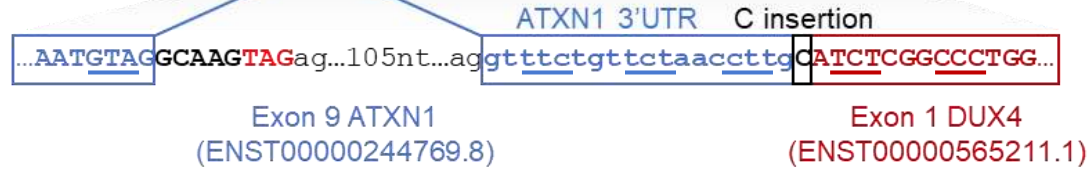


# How to determine if a fusion is relevant?

SampleID	FusionGene	Left_break_point	Right_break_point	nb_tools	FusionMap	STARFusion	ERICScript	FUSIONcatcher	ARRIBA	SplitReads	SpanReads	Frame	Transcripts	Known_in	Protein1_Domains	Protein2_Domains
CLB_RNA_0250	ATXN1 :: DUX4	chr6 16306201 -	chr4 190174015 +	1	0	0	1	0	0	176	39	No Frame	ATXN1_ENST0000024476 9_(three_prime_utr)- ::DUX4_ENST000005702 63_(242pb_INset_e1)	NA	Fusion point after CDS: Protein might be whole	342/424AA might be retained in fusion.

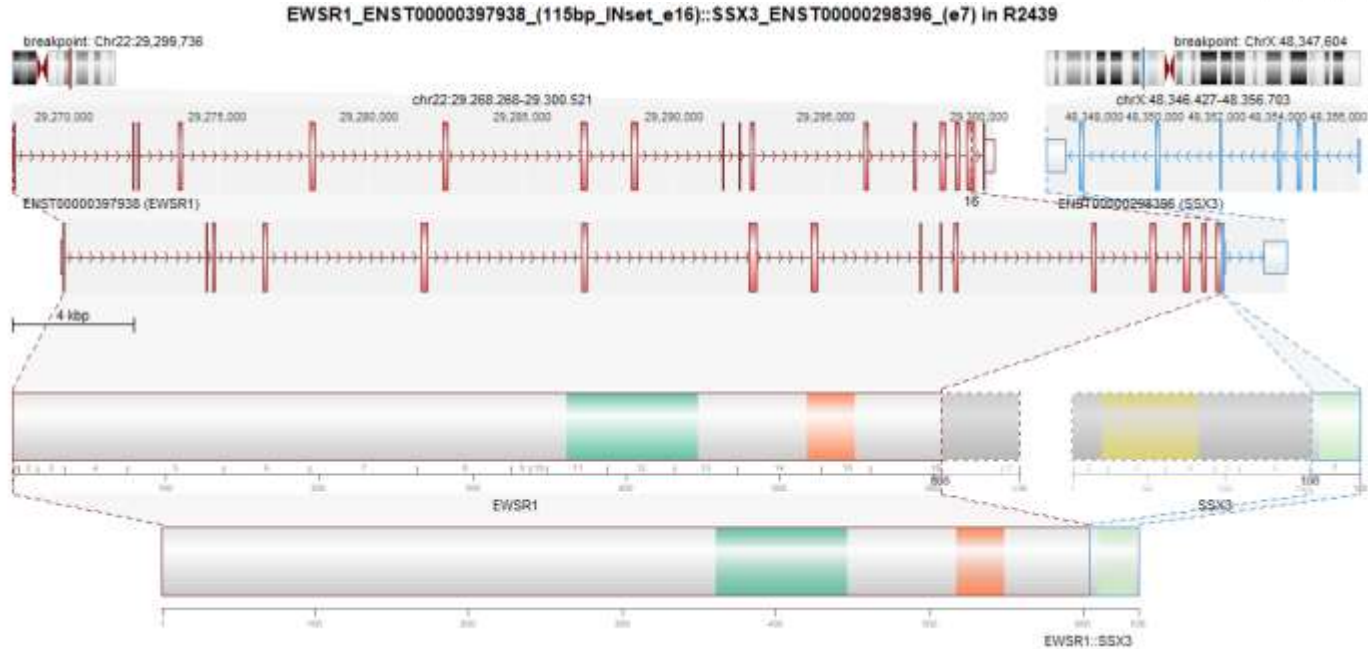


# How to determine if a fusion is relevant?



# Clustering might be (very) helpful for unknown fusions

## EWSR1::SSX3 fusion



- Ewing sarcoma?
- Synovial sarcoma ?
- Other?





Clustering might be (very) helpful for unknown fusions

**Expanding the molecular landscape of undifferentiated sarcomas of bone with a novel *EWSR1-SSX3* gene fusion**

(published online 05/12/2023)

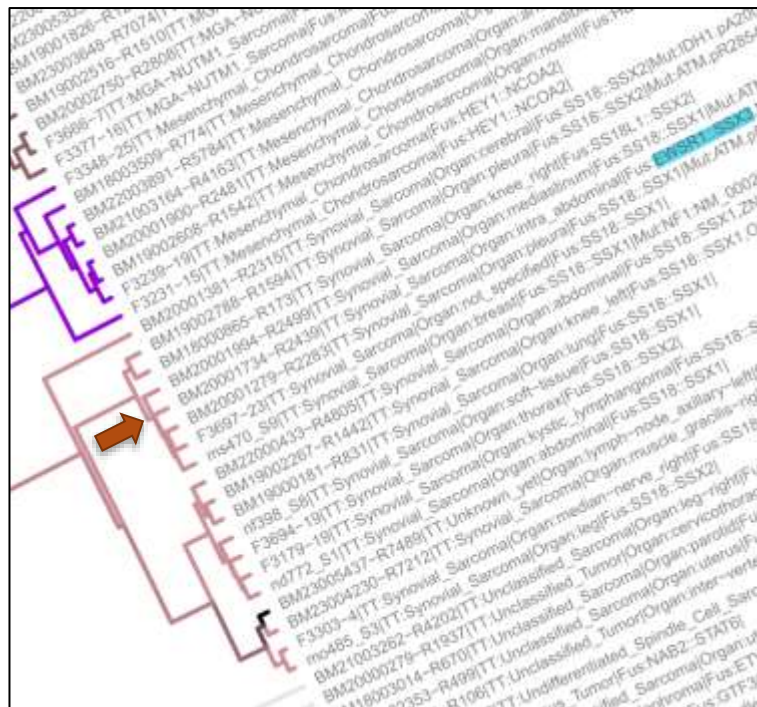
~~Ewing sarcoma?~~

~~Synovial sarcoma?~~

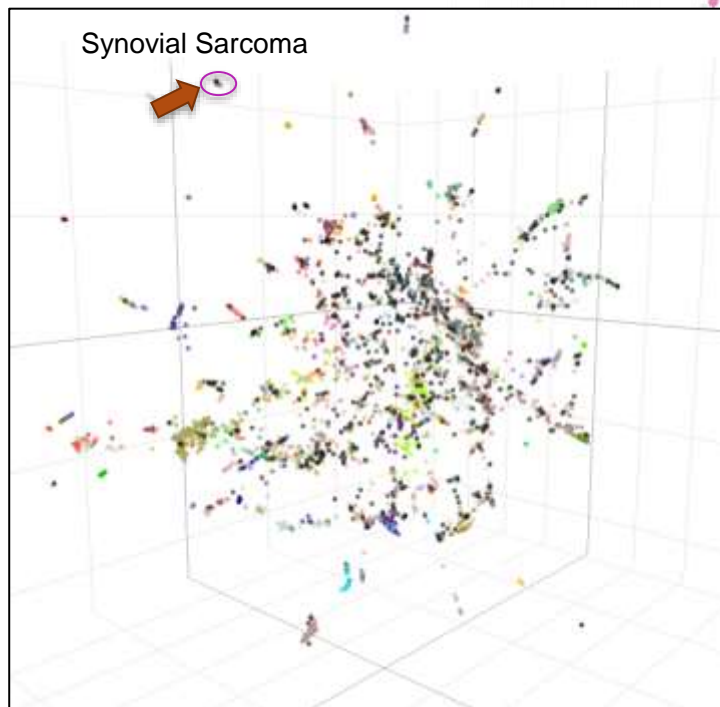
**Other**



# Clustering might be (very) helpful for unknown fusions



Consensus Clustering

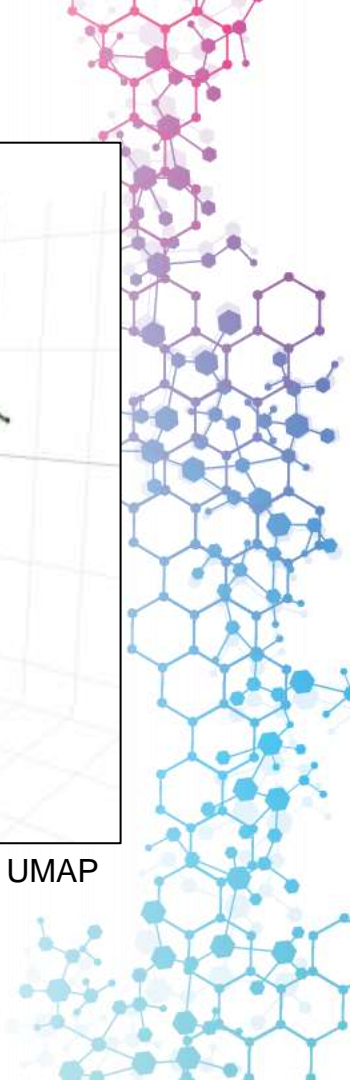


3D UMAP

~~Ewing sarcoma?~~

→ **Synovial sarcoma**

Other?





# Questions time!



- Franck Tirode, DR2 INSERM
- Arnaud Vigneron, MCU Lyon1
- Samia Ruby, CRHC INSERM
- Waisse Waissi MD, PhD
- Mehdi Brahmi, MD, PhD
- Sylvie Lantuejoul, MD, PhD
- Hélène Vanacker, MD, PhD student
- Enzo Amoroso, PhD student
- Clara Pambruni, PhD student
- Thomas Barre, Msc
- Virginie Perrin, Msc
- Adeline Duc-Gallet, Bsc



MERCI DE VOTRE ATTENTION

