

AANB01_504 run analysis

Bioinformatics Development

Sandy

2025.05.01

AANB01_504

- Run analysis plan

Analysis request

- 16 fusion v5 hybrid capture samples

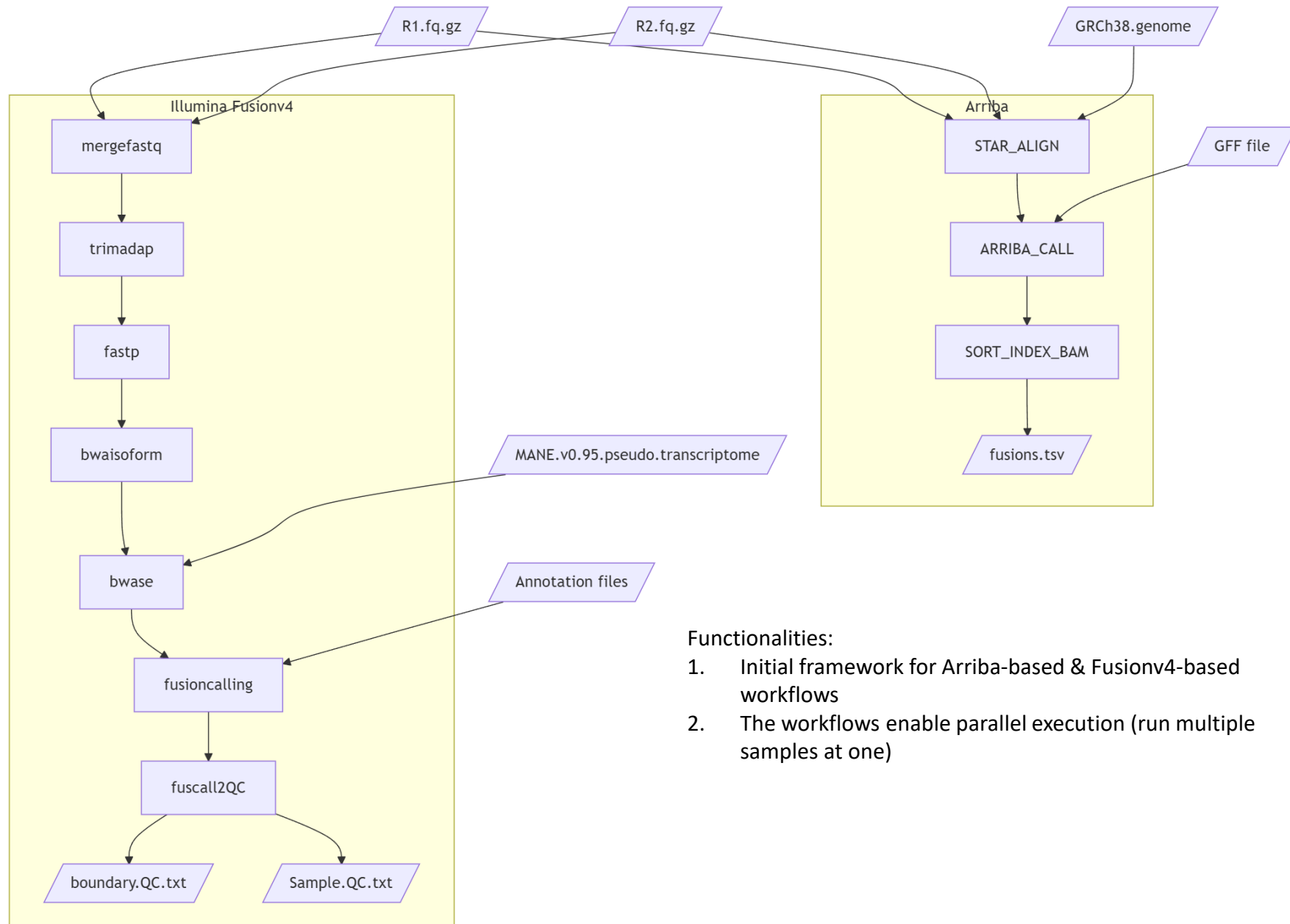
SN	Sample	input amount	Type	Purpose
1	AA-23-07938_R1	10 ng	Clinical FFPE RNA MET e14	Probe feasibility test
2	AA-23-07938_R1	100 ng	Clinical FFPE RNA MET e14	Probe feasibility test
3	ACT1187_R1	10 ng	Normal FFPE RNA	Probe feasibility test
4	ACT1187_R1	100 ng	Normal FFPE RNA	Probe feasibility test
5	ACT1188_R1	10 ng	Normal FFPE RNA	Probe feasibility test
6	ACT1188_R1	100 ng	Normal FFPE RNA	Probe feasibility test
7	RM-25-005_R1	10 ng	GM24385 FFPE RNA	Probe feasibility test
8	RM-25-005_R1	100 ng	GM24385 FFPE RNA	Probe feasibility test
9	RM-23-001_R1	10 ng	NTRK RM FFPE RNA	input amount vs calling
10	RM-23-001_R1	30 ng	NTRK RM FFPE RNA	input amount vs calling
11	RM-23-001_R1	50 ng	NTRK RM FFPE RNA	input amount vs calling
12	RM-23-001_R1	100 ng	NTRK RM FFPE RNA	input amount vs calling
13	RM-23-026_R1	100 ng	GM24385 gRNA	Difference between genomic RNA and FFPE RNA
14	RM-23-027_R1	100 ng	NTRK RM gRNA	Difference between genomic RNA and FFPE RNA
15	ACT2578_D5	50 ng	Normal FFPE DNA lib	Probe feasibility test
16	ACT2704_D2	50 ng	Normal FFPE DNA lib	Probe feasibility test

Analysis plan

- Use Fusion v5 v0.1 pipeline (~~arriba~~ → fusion v4 based) to analyze the 16 samples
- For probe feasibility evaluation, calculate exon-level & gene-level expression.
 - => Target gene expression
 - => Probe-anchored exon quantification
- For “input amount versus calling” test samples (input amount: 10/30/50/100 ng, sample: NTRK RM FFPE RNA)
 - => Calling result (positive fusions) table
- For genomic RNA and FFPE RNA differences comparison, compare the following 2 samples:
 - GM24385 gRNA
 - NTRK RM gRNA
- Supplementary data (nice-to-have)
 - % of probe-anchored reads for the 16 samples

V5 Workflows

- Nextflow repo
(hybridcapture_fusion_pipeline_nextflow)
- Pipeline utility repo
(fusion_pipeline_env)

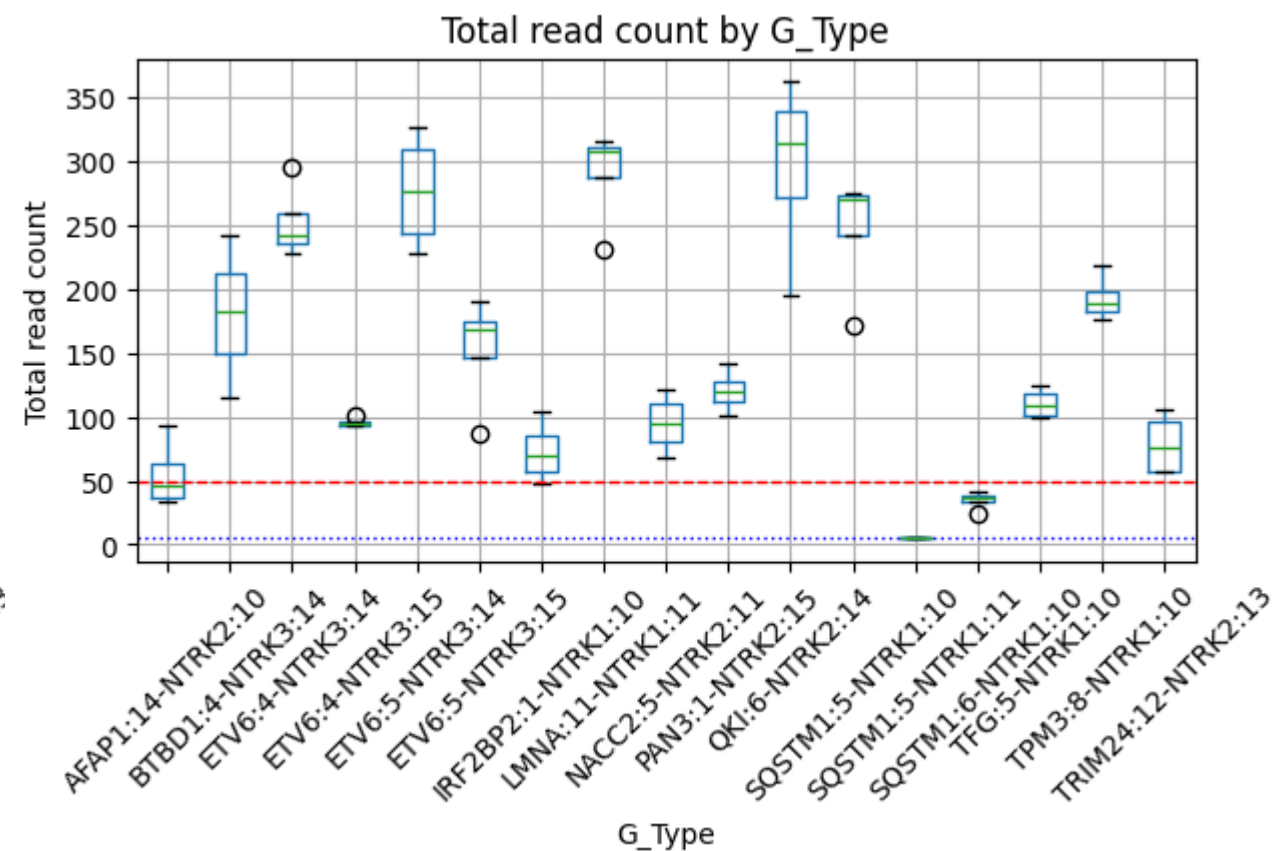
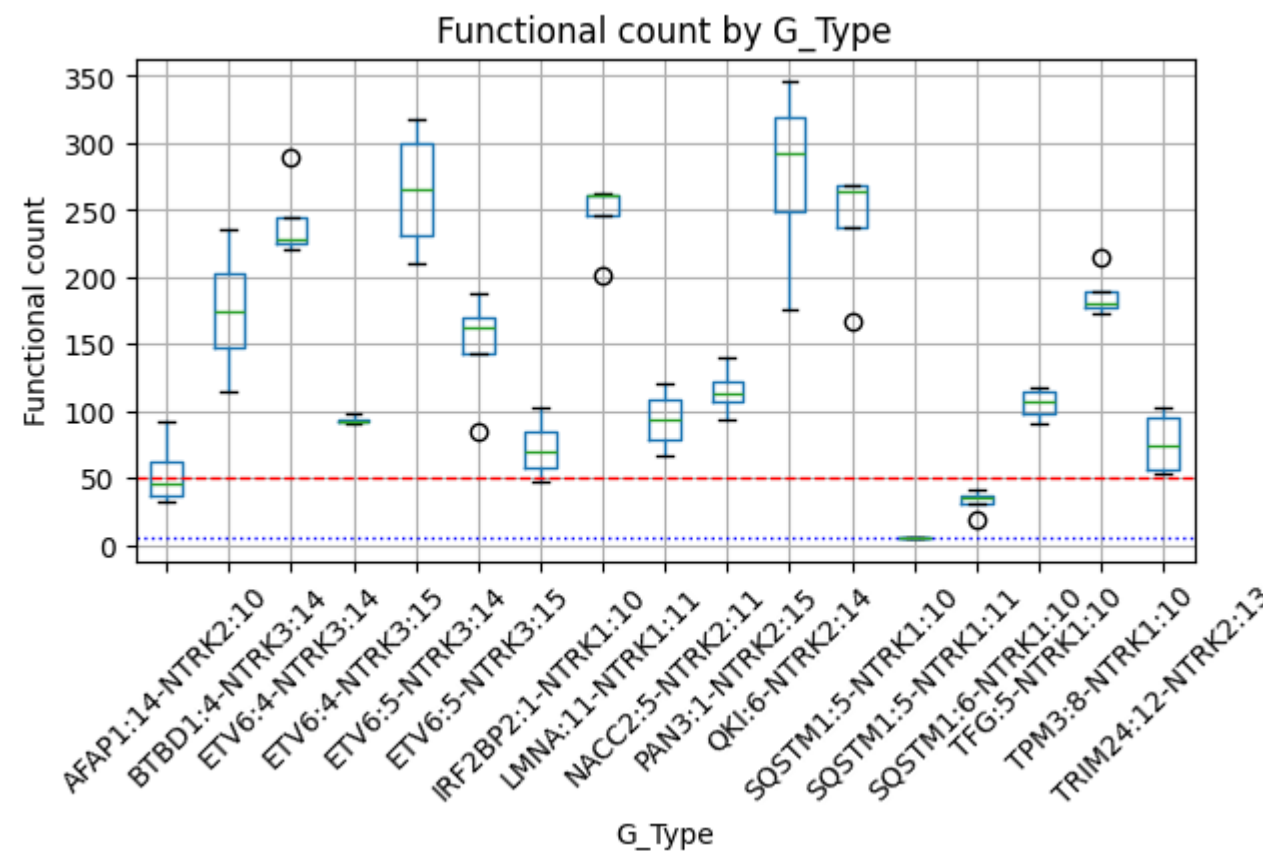


Functionalities:

1. Initial framework for Arriba-based & Fusionv4-based workflows
2. The workflows enable parallel execution (run multiple samples at one)

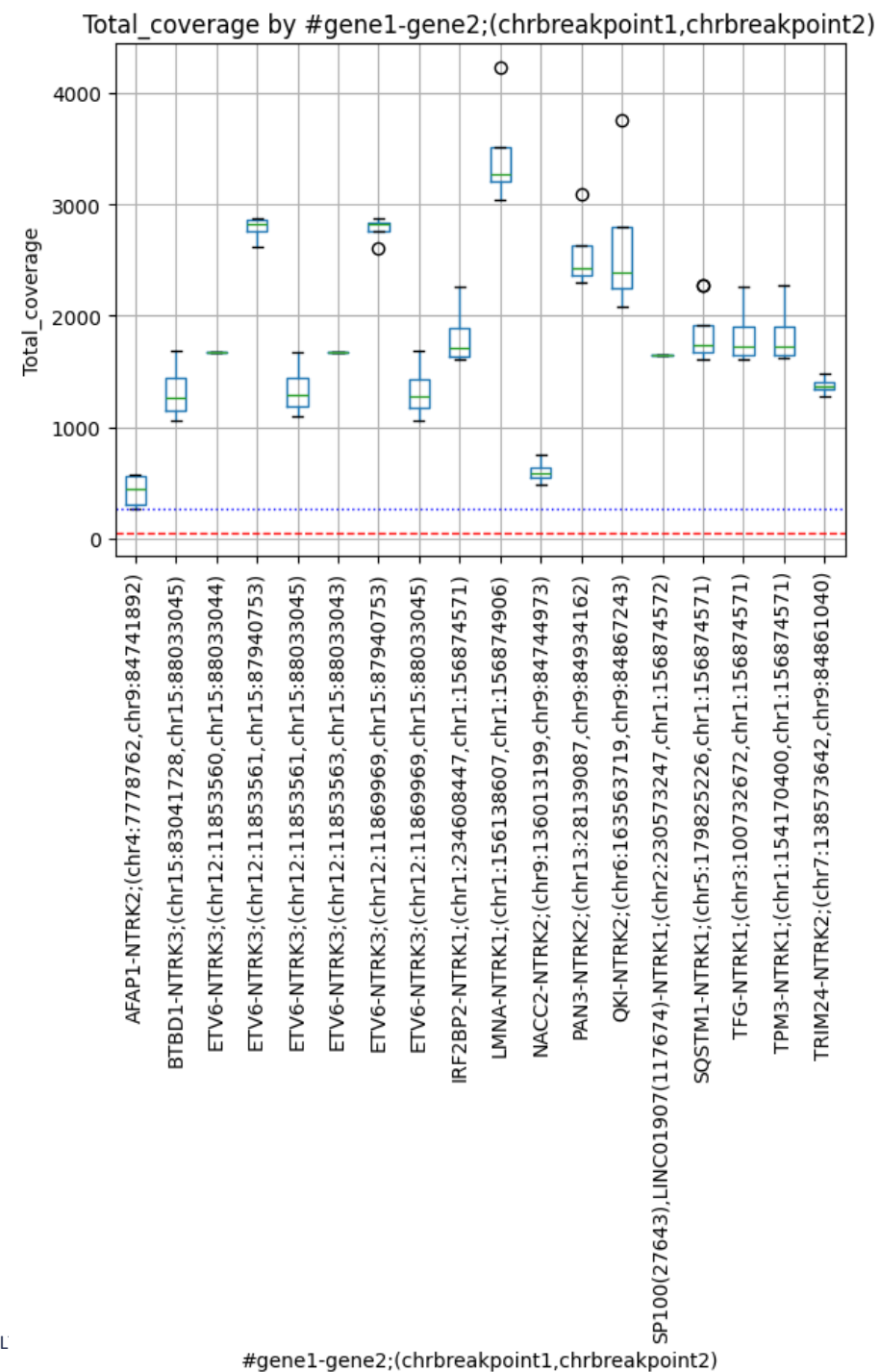


Fusionv4-based calling result



Arriba-based calling result

- $\text{Total_coverage} = \text{coverage1} + \text{coverage2}$



Assay design & Preferred transcripts

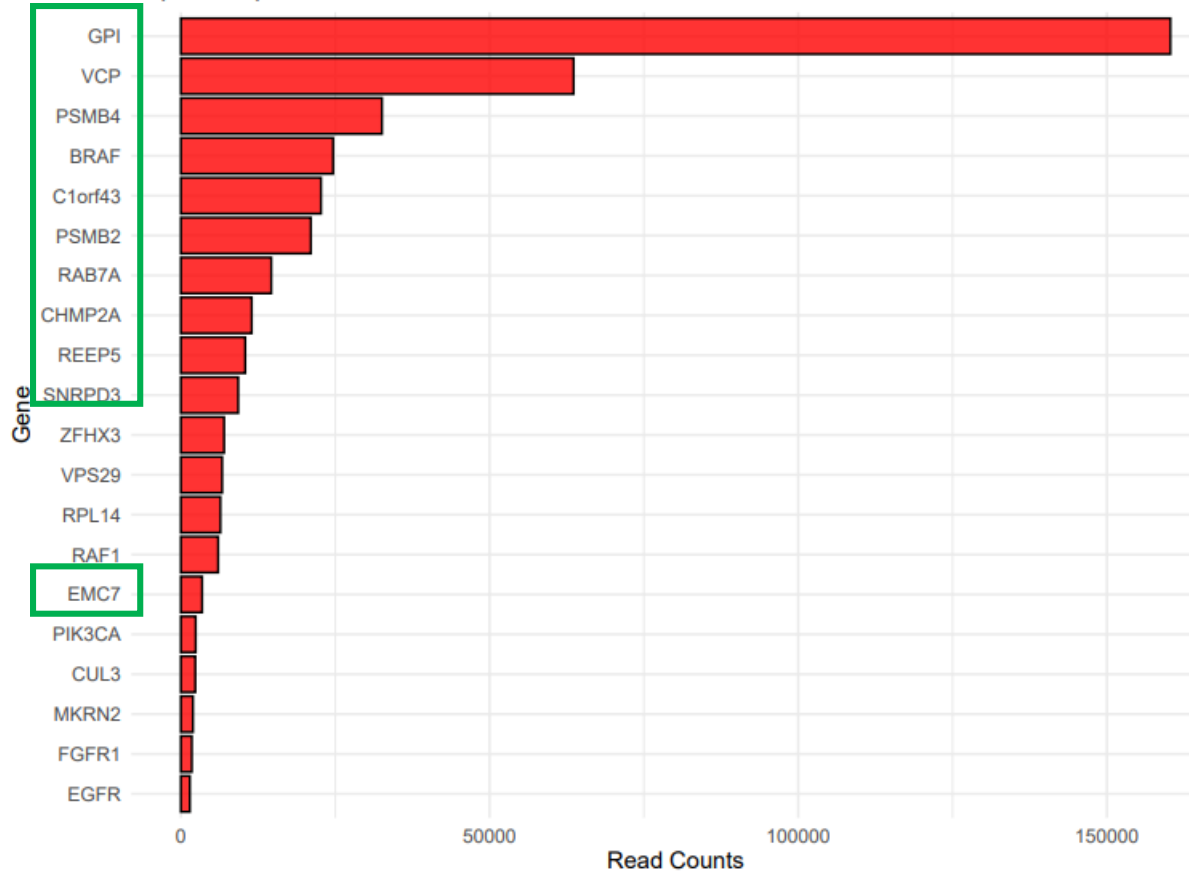
- Fusion V5 assay
 - 26 target genes + 11 internal control genes
- 38 MANE Select v0.95 transcripts for gene/exon quantification
 - ERG
(2 preferred transcripts are included for TMPRSS2-ERG fusion detection)
 - Fusion v4 db is constructed using MANE v0.95 transcripts + 4 manually curated transcripts obtained from GENCODE r38 (NRG1, NTRK3, AR, ERG)

ENST	ENSG	Gene	RefSeq ID
ENST00000389048.8	ENSG00000171094.18	ALK	NM_004304.5
ENST00000644969.2	ENSG00000157764.14	BRAF	NM_001374258.1
ENST00000275493.7	ENSG00000146648.20	EGFR	NM_005228.5
ENST00000269571.10	ENSG00000141736.14	ERBB2	NM_004448.4
ENST00000342788.9	ENSG00000178568.15	ERBB4	NM_005235.3
ENST00000288319.12	ENSG00000157554.20	ERG	NM_182918.4
ENST00000206249.8	ENSG00000091831.25	ESR1	NM_000125.4
ENST00000447712.7	ENSG00000077782.22	FGFR1	NM_023110.3
ENST00000440486.8	ENSG00000068078.20	FGFR3	NM_000142.5
ENST00000358487.10	ENSG00000066468.24	FGFR2	NM_000141.5
ENST00000356487.11	ENSG00000105220.17	GP1	NM_000175.5
ENST00000288135.6	ENSG00000157404.17	KIT	NM_000222.3
ENST00000397752.8	ENSG00000105976.16	MET	NM_000245.4
ENST00000524377.7	ENSG00000198400.13	NTRK1	NM_002529.4
ENST00000277120.8	ENSG00000148053.18	NTRK2	NM_006180.6
ENST00000263967.4	ENSG00000121879.6	PIK3CA	NM_006218.4
ENST00000373237.4	ENSG00000126067.12	PSMB2	NM_002794.5
ENST00000290541.7	ENSG00000159377.11	PSMB4	NM_002796.3
ENST00000442415.7	ENSG00000132155.14	RAF1	NM_001354689.3
ENST00000355710.8	ENSG00000165731.21	RET	NM_020975.6
ENST00000368507.8	ENSG00000047936.11	ROS1	NM_001378902.1
ENST00000215829.8	ENSG00000100028.12	SNRPD3	NM_004175.5
ENST00000332149.10	ENSG00000184012.14	TMPRSS2	NM_005656.4
ENST00000358901.11	ENSG00000165280.18	VCP	NM_007126.5
ENST00000265062.8	ENSG00000075785.14	RAB7A	NM_004637.6
ENST00000379638.9	ENSG00000129625.13	REEP5	NM_005669.5
ENST00000361474.6	ENSG00000158458.21	NRG2	NM_004883.3
ENST00000673466.1	ENSG00000117020.19	AKT3	NM_005465.7
ENST00000368521.10	ENSG00000143612.21	C1orf43	NM_001098616.3
ENST00000312547.7	ENSG00000130724.10	CHMP2A	NM_014453.4
ENST00000549578.6	ENSG00000111237.19	VPS29	NM_016226.5
ENST00000256545.9	ENSG00000134153.10	EMC7	NM_020154.3
ENST00000537011.6	ENSG00000184507.16	NUTM1	NM_001284292.2
ENST00000276659.10	ENSG00000147655.12	RSPO2	NM_178565.5
ENST00000356819.7	ENSG00000157168.20	NRG1	NM_013957.5
ENST00000394480.6	ENSG00000140538.16	NTRK3	NM_002530.4
ENST00000442448.5	ENSG00000157554.19	ERG	NM_004449.4
ENST00000504326.5	ENSG00000169083.18	AR	NM_001348061.1

Gene/Exon quantification

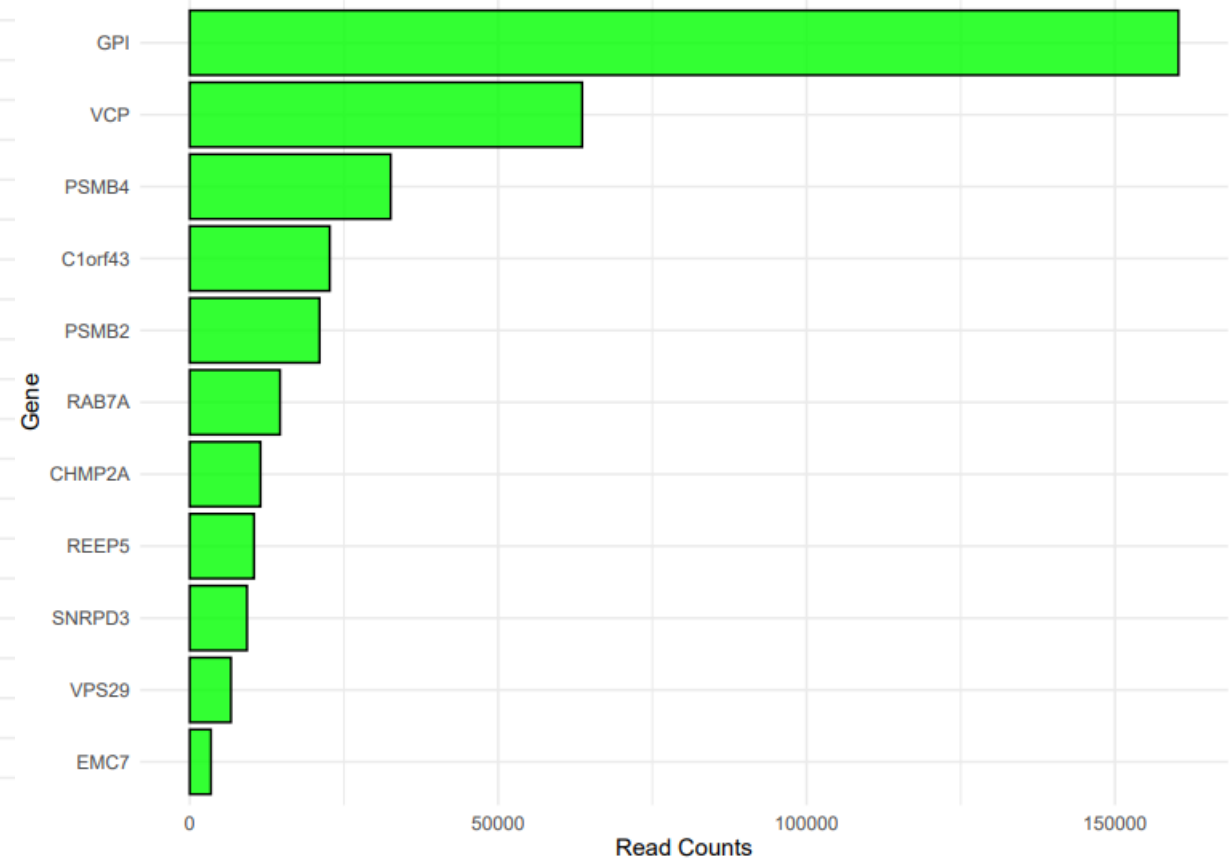
- AANB01_504_A88_RM-25-005

Top 20 Expressed Genes



Expression of Specified Genes

11 INC genes



Exon Quantification

- AANB01_504_A83_AA-23-07938_R1
(FFPE RNA with relatively higher mapped exon count)
- AANB01_504_A53_ACT2578_D5
(DNA as starting material)

DNA (AANB01_504_A53_ACT2578_D5)

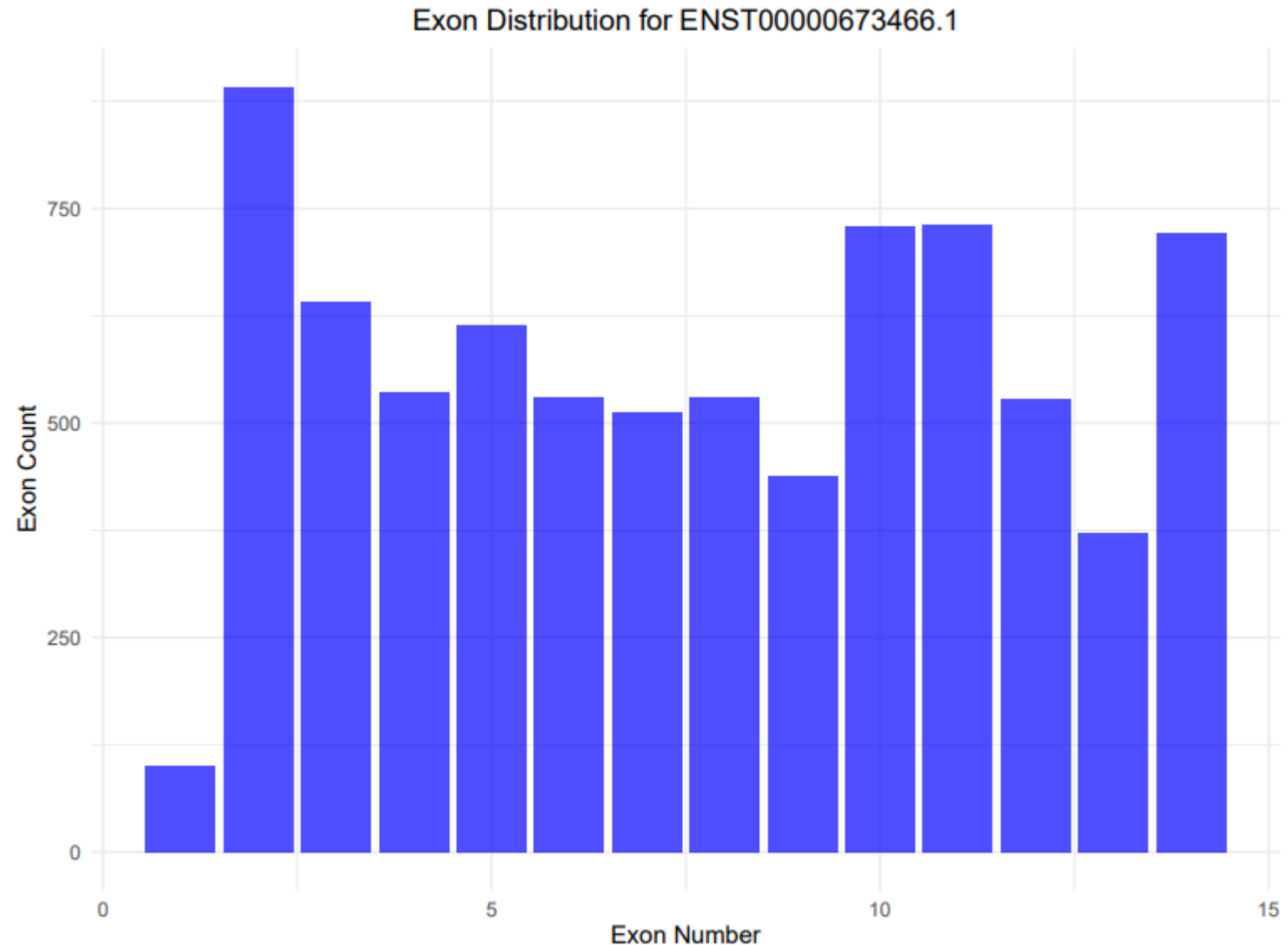
EGFR 27647
 ROS1 23374
 FGFR3 21116
 MET 21026
 ALK 19305
 ERBB2 19254
 BRAF 18798
 ERBB4 17250
 RET 15727
 PIK3CA 14667
 KIT 13968
 FGFR1 13946
 NTRK1 13327
 NTRK2 13308
 VCP 11987
 RAF1 11810
 NTRK3 11368
 GPI 10860
 AR 10030
 FGFR2 9947
 TMPRSS2 9493
 AKT3 7876
 ERG 7792
 NRG2 7578
 ESR1 7130
 C1orf43 5560
 NRG1 5378
 NUTM1 4773
 REEP5 4444
 PSMB4 4436
 RSPO2 4037
 PSMB2 3544
 RAB7A 2866
 CHMP2A 2649
 EMC7 1934
 SNRPD3 1836
 VPS29 1569

FFPE RNA (AANB01_504_A83_AA-23-07938_R1)

EGFR 460626
 VCP 151316
 MET 125731
 PSMB4 97687
 RAB7A 93879
 GPI 91581
 PSMB2 64943
 C1orf43 59189
 BRAF 25770
 CHMP2A 24711
 SNRPD3 24526
 VPS29 22448
 REEP5 21781
 EMC7 16149
 RAF1 8736
 ERBB2 5583
 PIK3CA 5157
 NRG1 3952
 ALK 3363
 FGFR1 2859
 AKT3 2634
 ERG 2588
 ROS1 1444
 AR 1390
 FGFR3 1253
 FGFR2 1128
 TMPRSS2 523
 RSPO2 479
 KIT 387
 NRG2 214
 ESR1 173
 NTRK2 144
 ERBB4 128
 NTRK1 125
 RET 104
 NUTM1 56
 NTRK3 31

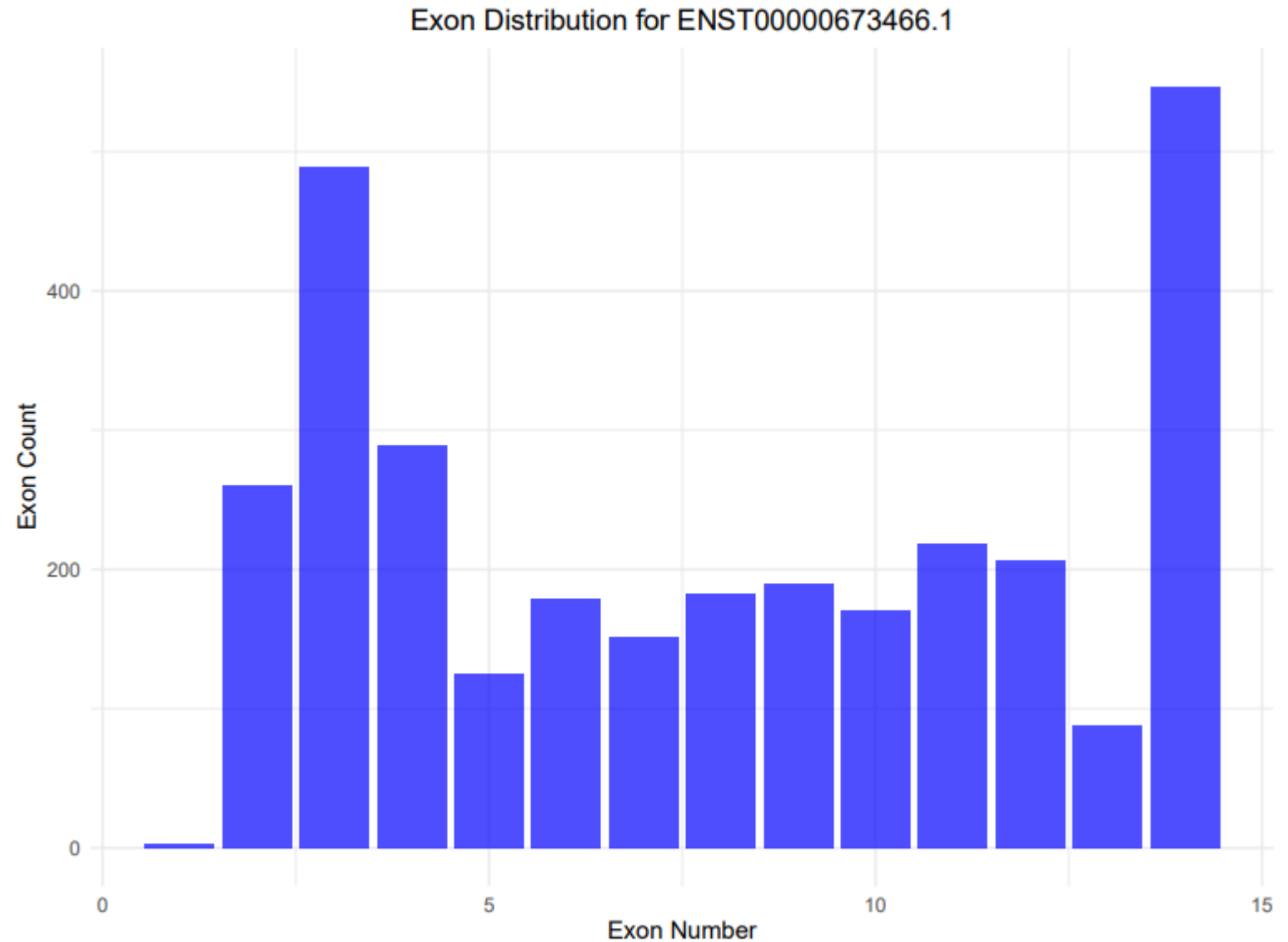
AKT3

- DNA
 - AANB01_504_A53_ACT2578_D5



AKT3

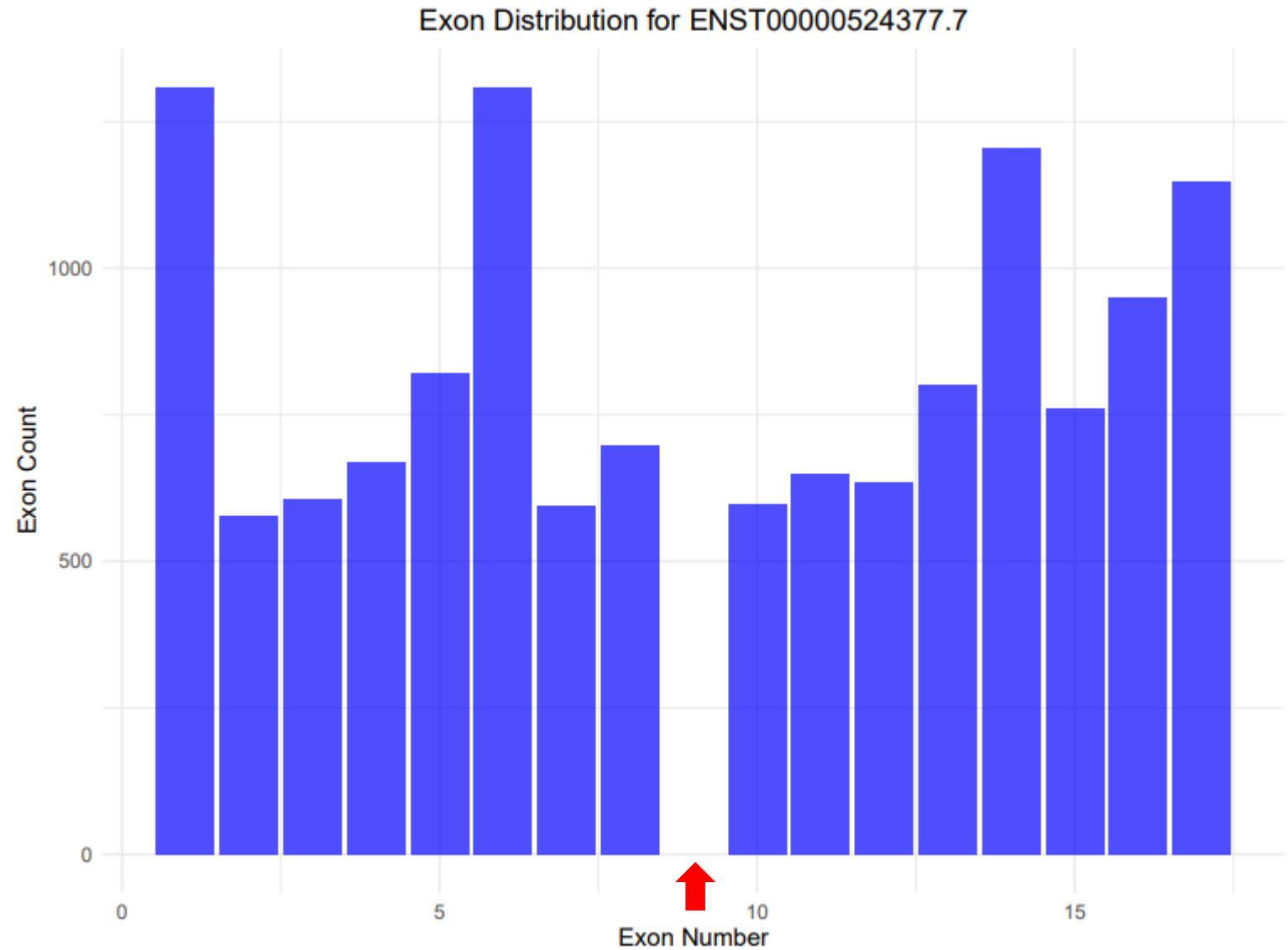
- FFPE RNA
 - AANB01_504_A83_AA-23-07938_R1



NTRK1

- DNA
 - AANB01_504_A53_ACT2578_D5

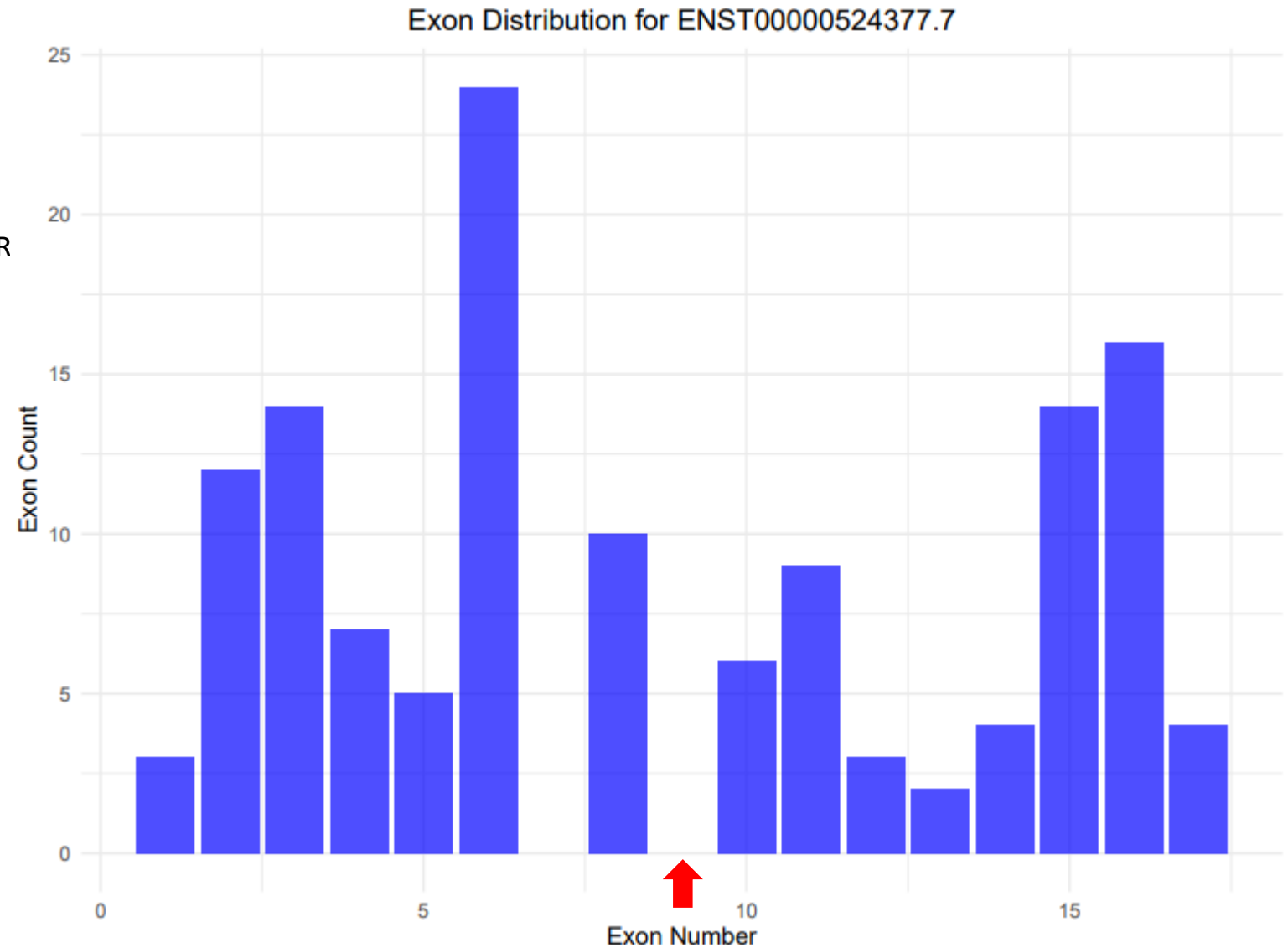
NTRK1 exon 9 is missing.



NTRK1

- FFPE RNA
 - AANB01_504_A83_AA-23-07938_R

NTRK1 exon 9 is missing.



Exons with 0 read in DNA samples

- 560 exons (on the 37 genes) * 16 samples

Transcript_exon_id	AANB01_504_A53_ACT2578_D5	AANB01_504_A54_ACT2704_D2	AANB01_504_A82_AA-23-07938_R1	AANB01_504_A83_AA-23-07938_R1	AANB01_504_A84_ACT1187_R1	AANB01_504_A85_ACT1187_R1	AANB01_504_A86_ACT1188_R1	AANB01_504_A87_ACT1188_R1	AANB01_504_A88_RM-25-005_R1	AANB01_504_A89_RM-25-005_R1	AANB01_504_A90_RM-23-001_R1	AANB01_504_A91_RM-23-001_R1	AANB01_504_A92_RM-23-001_R1	AANB01_504_A93_RM-23-001_R1	AANB01_504_A94_RM-23-026_R1	AANB01_504_A95_RM-22-014_R1
ENST00000290541.7;PSMB4-exon-fusionv4-7	0	0	12131	24105	207	811	0	0	5312	13046	7071	9534	10007	14040	14867	22807
ENST00000524377.7;NTRK1-exon-fusionv4-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENST00000442415.7;RAF1-exon-fusionv4-9	0	0	20	47	0	1	0	0	16	16	0	5	15	21	34	28
ENST00000265062.8;RAB7A-exon-fusionv4-2	0	0	347	353	1	5	0	0	114	181	164	160	172	250	198	345
ENST00000368507.8;ROS1-exon-fusionv4-4	0	0	21	22	0	0	0	0	0	0	0	0	0	0	1	0
ENST00000256545.9;EMC7-exon-fusionv4-2	0	0	1736	2214	15	103	0	0	600	881	1206	1338	1525	2011	1007	1691
ENST00000394480.6;NTRK3-exon-fusionv4-11	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0

X-NTRK1/2/3 (positive) RM sample

- RM sample (RM-23-001_R1)

=>

**Need to confirm the expected fusion boundary
(e.g., which transcript version is preferred?)**

Fusion	Digital PCR Fusion Copies/ng of total RNA	note
TPM3-NTRK1	259	
LMNA-NTRK1	139	
IRF2BP2-NTRK1	113	
SQSTM1-NTRK1	180	
TFG-NTRK1	129	
AFAP1-NTRK2	636	
NACC2-NTRK2	332	
QKI-NTRK2	150	
TRIM24-NTRK2	302	
PAN3-NTRK2	224	
ETV6-NTRK3	275	(E5N14)
ETV6-NTRK3	371	(E5N15)
ETV6-NTRK3	388	(E4N15)
ETV6-NTRK3	390	(E4N14)
BTBD1-NTRK3	251	

Fusion v4-based workflow

- AANB01_504_A90_RM-23-001_R1 (10 ng, RM-23-001_R1 sample)
- Decision = '+'

T_Type	Functional count	Total read count (filtered)	Total read count	Fusion Type	G_Type	Boundary status	5'UTR	Decision
ENST00000651641.1:8-ENST00000524377.7:10	215	218	218	FUSION	TPM3:8-NTRK1:10	PASS	-	+
ENST00000368300.9:11-ENST00000524377.7:11	201	231	231	FUSION	LMNA:11-NTRK1:11	PASS	-	+
ENST00000389805.9:5-ENST00000524377.7:10	269	273	273	FUSION	SQSTM1:5-NTRK1:10	PASS	-	+
ENST00000366609.4:1-ENST00000524377.7:10	60	61	61	FUSION	IRF2BP2:1-NTRK1:10	PASS	+	+
ENST00000240851.9:5-ENST00000524377.7:10	100	102	102	FUSION	TFG:5-NTRK1:10	PASS	-	+
ENST00000389805.9:6-ENST00000524377.7:10	41	42	42	FUSION	SQSTM1:6-NTRK1:10	PASS	-	+
ENST00000420658.6:14-ENST00000277120.8:10	39	39	39	FUSION	AFAP1:14-NTRK2:10	PASS	-	+
ENST00000361752.8:6-ENST00000277120.8:14	176	195	195	FUSION	QKI:6-NTRK2:14	PASS	-	+
ENST00000343526.9:12-ENST00000277120.8:13	53	57	57	FUSION	TRIM24:12-NTRK2:13	PASS	-	+
ENST00000277554.4:5-ENST00000277120.8:11	104	106	106	FUSION	NACC2:5-NTRK2:11	PASS	-	+
ENST00000380958.8:1-ENST00000277120.8:15	115	123	123	FUSION	PAN3:1-NTRK2:15	PASS	+	+
ENST00000396373.9:4-ENST00000394480.6:14	226	226	228	FUSION	ETV6:4-NTRK3:14	PASS	-	+
ENST00000396373.9:4-ENST00000394480.6:15	93	94	94	FUSION	ETV6:4-NTRK3:15	PASS	-	+
ENST00000396373.9:5-ENST00000394480.6:14	210	227	227	FUSION	ETV6:5-NTRK3:14	PASS	-	+
ENST00000396373.9:5-ENST00000394480.6:15	188	190	190	FUSION	ETV6:5-NTRK3:15	PASS	-	+
ENST00000261721.9:4-ENST00000394480.6:14	115	115	115	FUSION	BTBD1:4-NTRK3:14	PASS	-	+

Arriba-based workflow

- AANB01_504_A90_RM-23-001_R1 (10 ng, RM-23-001_R1 sample)
- Fusions.tsv file

NTRK2

=>

NTRK2(NM_001369532)

NTRK2(NM_006180)

type	#gene1(transcript_id1)	gene2(transcript_id2)	breakpoint1	breakpoint2	split_reads1	split_reads2	discordant_mates	filters	coverage1	coverage2	confidence
translocation	SQSTM1(NM_001142298_2)	NTRK1(NM_001007792)	5:179825226	1:156874571	9	20		0duplicates(89),inconsistently_clipped(2),mismatches(1)	20	1773	high
translocation	SQSTM1(NM_001142299)	NTRK1(NM_001007792)	5:179825226	1:156874571	9	20		0duplicates(89),inconsistently_clipped(2),mismatches(1)	20	1773	high
translocation	QKI(NM_001301085)	NTRK2(NM_001369532)	6:163563719	9:84867243	5	19		0duplicates(175)	18	2280	high
deletion	LMNA(NM_001282626)	NTRK1(NM_001007792)	1:156138607	1:156874906	13	10		1duplicates(75)	82	3196	high
translocation	ETV6(NM_001987)	NTRK3(NM_002530)	12:11853561	15:88033045	12	10		0duplicates(96)	77	1015	high
translocation	ETV6(NM_001987)	NTRK3(NM_002530)	12:11869969	15:88033045	9	12		0duplicates(117),inconsistently_clipped(3)	44	1015	high
translocation	ETV6(NM_001987)	NTRK3(NM_002530)	12:11869969	15:87940753	2	16		0duplicates(145),mismatches(1)	44	2778	high
translocation	ETV6(NM_001987)	NTRK3(NM_002530)	12:11853561	15:87940753	9	8		0duplicates(77)	77	2778	high
duplication	BTBD1(NM_001011885)	NTRK3(NM_002530)	15:83041728	15:88033045	10	9		0duplicates(87)	45	1015	high
translocation	PAN3(NM_175854)	NTRK2(NM_001369532)	13:28139087	9:84934162	2	17		0duplicates(126)	7	2470	high
inversion	TPM3(NM_001364679)	NTRK1(NM_001007792)	1:154170400	1:156874571	3	16		0duplicates(92),mismatches(3)	14	1773	high
translocation	TFG(NM_001007565)	NTRK1(NM_001007792)	3:100732672	1:156874571	3	8		0duplicates(59),mismatches(1)	15	1773	high
translocation	TRIM24(NM_003852)	NTRK2(NM_006180)	7:138573642	9:84861040	1	9		0duplicates(27)	1	1275	high
inversion	NACC2(NM_144653)	NTRK2(NM_001369532)	9:136013199	9:84744973	5	4		0duplicates(34)	22	582	high
inversion	IRF2BP2(NM_182972)	NTRK1(NM_001007792)	1:234608447	1:156874571	0	7		0duplicates(47),mismatches(1)	0	1773	medium
translocation	AFAP1(NM_001134647)	NTRK2(NM_001369532)	4:7778762	9:84741892	0	5		0duplicates(28)	0	262	medium

RM-23-001_R1 sample (merged on “genomic coordinates”)

- Arriba/fusionv4 * 10/30/50/100 ng
 - Filtering criteria
 - Fusion v4: Decision == '+'
 - Arriba: All the rows within "fusions.tsv" (=> considered positive by Arriba)

[illegible]

NTRK2 (different RefSeq IDs for different input amount)

- NTRK2 fusions among different conditions (input amount)
 - Observations
 - NACC2-NTRK2 fusions: same genomic break points => different NTRK2 transcripts
 - 2 transcripts for “NTRK2” fusions: NTRK2(NM_001369532) & NTRK2(NM_006180)

#gene1-gene2;(chrbreakpoint1,chrbreakpoint2)	type	#gene1(transcript_id1)	gene2(transcript_id2)	coverage1	coverage2	confidence	file_name	input_amount
AFAP1-NTRK2;(chr4:7778762,chr9:84741892)	translocation	AFAP1(NM_001134647)	NTRK2(NM_001369532)	0	262	medium	AANB01_504_A90_RM-23-001_R1.fusions.keycols.tsv	10
AFAP1-NTRK2;(chr4:7778762,chr9:84741892)	translocation	AFAP1(NM_001134647)	NTRK2(NM_001369532)	0	563	medium	AANB01_504_A91_RM-23-001_R1.fusions.keycols.tsv	30
AFAP1-NTRK2;(chr4:7778762,chr9:84741892)	translocation	AFAP1(NM_001134647)	NTRK2(NM_001369532)	2	317	high	AANB01_504_A92_RM-23-001_R1.fusions.keycols.tsv	50
AFAP1-NTRK2;(chr4:7778762,chr9:84741892)	translocation	AFAP1(NM_001134647)	NTRK2(NM_001369532)	1	572	high	AANB01_504_A93_RM-23-001_R1.fusions.keycols.tsv	100
NACC2-NTRK2;(chr9:136013199,chr9:84744973)	inversion	NACC2(NM_144653)	NTRK2(NM_001369532)	22	582	high	AANB01_504_A90_RM-23-001_R1.fusions.keycols.tsv	10
NACC2-NTRK2;(chr9:136013199,chr9:84744973)	inversion	NACC2(NM_144653)	NTRK2(NM_001369532)	20	552	high	AANB01_504_A91_RM-23-001_R1.fusions.keycols.tsv	30
NACC2-NTRK2;(chr9:136013199,chr9:84744973)	inversion	NACC2(NM_144653)	NTRK2(NM_006180)	20	469	high	AANB01_504_A92_RM-23-001_R1.fusions.keycols.tsv	50
NACC2-NTRK2;(chr9:136013199,chr9:84744973)	inversion	NACC2(NM_144653)	NTRK2(NM_001369532)	40	712	high	AANB01_504_A93_RM-23-001_R1.fusions.keycols.tsv	100
PAN3-NTRK2;(chr13:28139087,chr9:84934162)	translocation	PAN3(NM_175854)	NTRK2(NM_001369532)	7	2470	high	AANB01_504_A90_RM-23-001_R1.fusions.keycols.tsv	10
PAN3-NTRK2;(chr13:28139087,chr9:84934162)	translocation	PAN3(NM_175854)	NTRK2(NM_001369532)	6	2376	high	AANB01_504_A91_RM-23-001_R1.fusions.keycols.tsv	30
PAN3-NTRK2;(chr13:28139087,chr9:84934162)	translocation	PAN3(NM_175854)	NTRK2(NM_001369532)	3	2298	medium	AANB01_504_A92_RM-23-001_R1.fusions.keycols.tsv	50
PAN3-NTRK2;(chr13:28139087,chr9:84934162)	translocation	PAN3(NM_175854)	NTRK2(NM_001369532)	8	3078	high	AANB01_504_A93_RM-23-001_R1.fusions.keycols.tsv	100
QKI-NTRK2;(chr6:163563719,chr9:84867243)	translocation	QKI(NM_001301085)	NTRK2(NM_001369532)	18	2280	high	AANB01_504_A90_RM-23-001_R1.fusions.keycols.tsv	10
QKI-NTRK2;(chr6:163563719,chr9:84867243)	translocation	QKI(NM_001301085)	NTRK2(NM_001369532)	51	2025	high	AANB01_504_A91_RM-23-001_R1.fusions.keycols.tsv	30
QKI-NTRK2;(chr6:163563719,chr9:84867243)	translocation	QKI(NM_001301085)	NTRK2(NM_001369532)	23	2447	high	AANB01_504_A92_RM-23-001_R1.fusions.keycols.tsv	50
QKI-NTRK2;(chr6:163563719,chr9:84867243)	translocation	QKI(NM_001301085)	NTRK2(NM_001369532)	34	3725	high	AANB01_504_A93_RM-23-001_R1.fusions.keycols.tsv	100
TRIM24-NTRK2;(chr7:138573642,chr9:84861040)	translocation	TRIM24(NM_003852)	NTRK2(NM_006180)	1	1275	high	AANB01_504_A90_RM-23-001_R1.fusions.keycols.tsv	10
TRIM24-NTRK2;(chr7:138573642,chr9:84861040)	translocation	TRIM24(NM_003852)	NTRK2(NM_006180)	1	1363	high	AANB01_504_A91_RM-23-001_R1.fusions.keycols.tsv	30
TRIM24-NTRK2;(chr7:138573642,chr9:84861040)	translocation	TRIM24(NM_003852)	NTRK2(NM_006180)	3	1372	high	AANB01_504_A92_RM-23-001_R1.fusions.keycols.tsv	50
TRIM24-NTRK2;(chr7:138573642,chr9:84861040)	translocation	TRIM24(NM_003852)	NTRK2(NM_006180)	4	1479	high	AANB01_504_A93_RM-23-001_R1.fusions.keycols.tsv	100

Make
Personalized
Medicine
Accessible to All

