

AANB01_504 run analysis

Bioinformatics Development

Sandy

2025.05.09

AANB01_504

- Run analysis plan

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

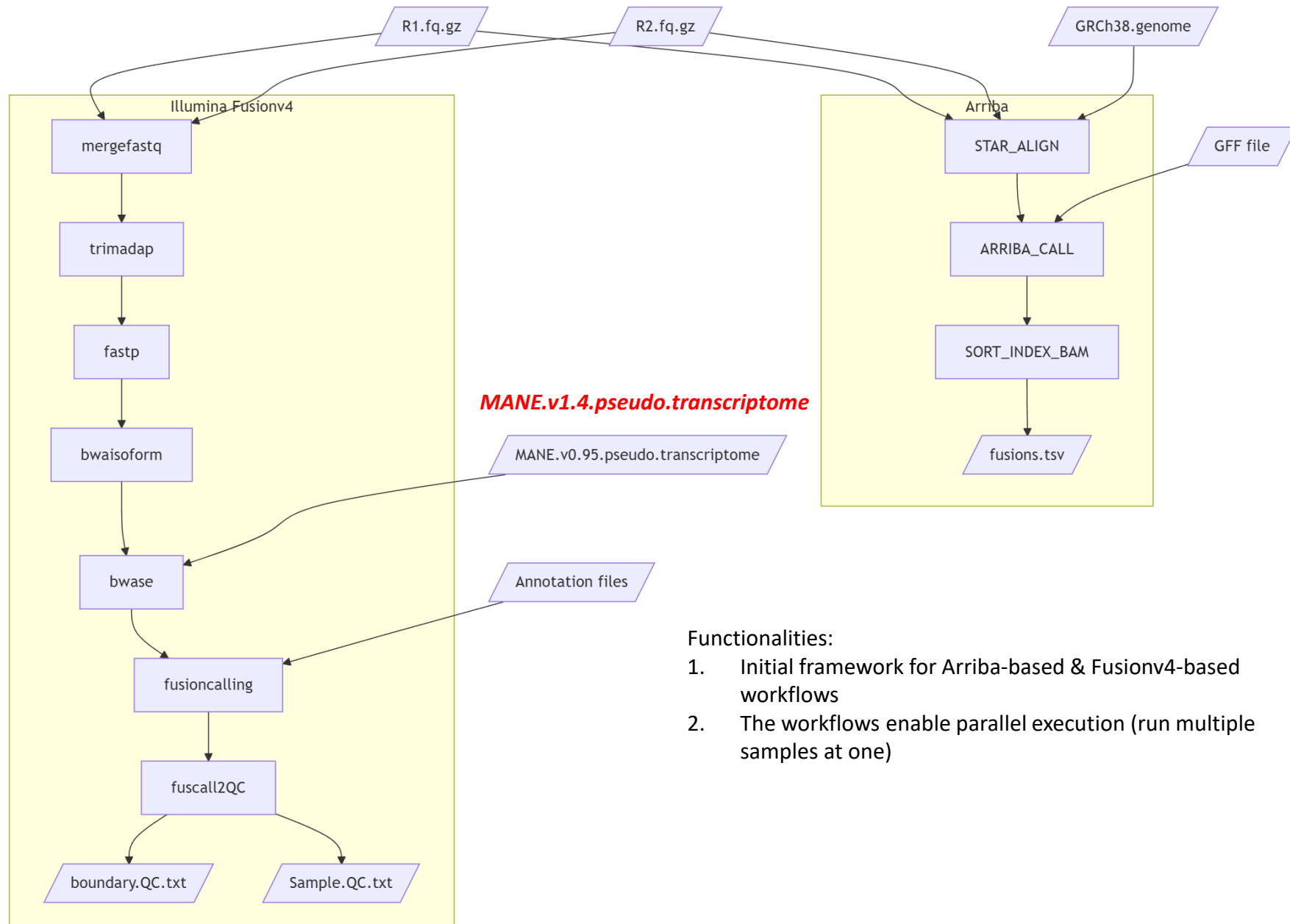
Analysis request

- 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

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)

Overview

- 1,039 fusion v5 probe regions to MANE v1.4 exons check
 - MANE v1.4 exon check
 - 1,039 probes -> 533 exons
 - Missing 32 exons on the INC genes
 - MANE v1.4 exon quantification on the 16 AANB01_504 samples
 - NTRK1 exon 9 and NTRK3 exon 11 are still missing
- NTRK RM sample calling result
 - Arriba
 - Fusion V4 (preferred MANE v0.95, 403 primer annotated pipeline)
 - Fusion V4 (MANE v1.4, 533*2 probe-anchored (exon) annotated pipeline)

Fusion	Digital PCR Fusion Copies/ng of total RNA	note	Arriba	Fusion v4 (primer, v0.95 MANE)	Fusion v4 (probe, v1.4 MANE)
TPM3-NTRK1	259		*	*	*
LMNA-NTRK1	139		*	*	*
IRF2BP2-NTRK1	113		*	*	*
SQSTM1-NTRK1	180		*	*	*
TFG-NTRK1	129		*	*	*
AFAP1-NTRK2	636		*	*	*
NACC2-NTRK2	332		* (Different NTRK2 IDs for different input amount)	*	*
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		*	*	*

MANE v1.4 exon check

- 559 exons (on MANE v1.4) for the 37 genes
=>
527 annotated exons +
32 exons not annotated

Gene	Annotated exon count	Expected exon count	Missing count
C1orf43	5	7	2
CHMP2A	3	6	3
EMC7	3	5	2
GPI	12	18	6
PSMB2	3	6	3
PSMB4	4	7	3
RAB7A	3	6	3
REEP5	3	5	2
SNRPD3	2	4	2
VCP	13	17	4
VPS29	2	4	2

Remark:

The 32 exons are on internal control genes.

=>

Won't affect variant calling.

Non-target mapping

- 6 probe regions/exons => genomic locations

ENST00000366541.8;SDCCAG8-exon-fusionv4-18

ENST00000368195.4;INSRR-exon-fusionv4-19

ENST00000170447.12;MKRN2-exon-fusionv4-8

ENST00000652575.1;LOC128092250-exon-fusionv4-1

ENST00000300658.9;PGAP3-exon-fusionv4-1

ENST00000394231.8;MIEN1-exon-fusionv4-4

Exons with 0 read in DNA samples

- 559 exons (on MANE v1.4, 37 genes) * 16 samples

=> similar result for NTRK1 exon 9 and NTRK3 exon 11 (same preferred transcripts)

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	12674	25147	186	804	0	0	5604	13740	7703	10053	10504	14630	14918	23770
ENST00000524377.7;NTRK1-exon-fusionv4-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ENST00000251849.9;RAF1-exon-fusionv4-8	0	0	36	117	0	1	0	0	166	366	126	181	159	235	337	379
ENST00000265062.8;RAB7A-exon-fusionv4-2	0	0	346	349	1	5	0	0	114	181	164	158	172	247	197	341
ENST00000368507.8;ROS1-exon-fusionv4-4	0	0	17	25	0	0	0	0	0	0	0	0	0	0	1	0
ENST00000256545.9;EMC7-exon-fusionv4-2	0	0	1736	2201	15	103	0	0	600	874	1206	1337	1516	1996	998	1687
ENST00000629765.3;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	

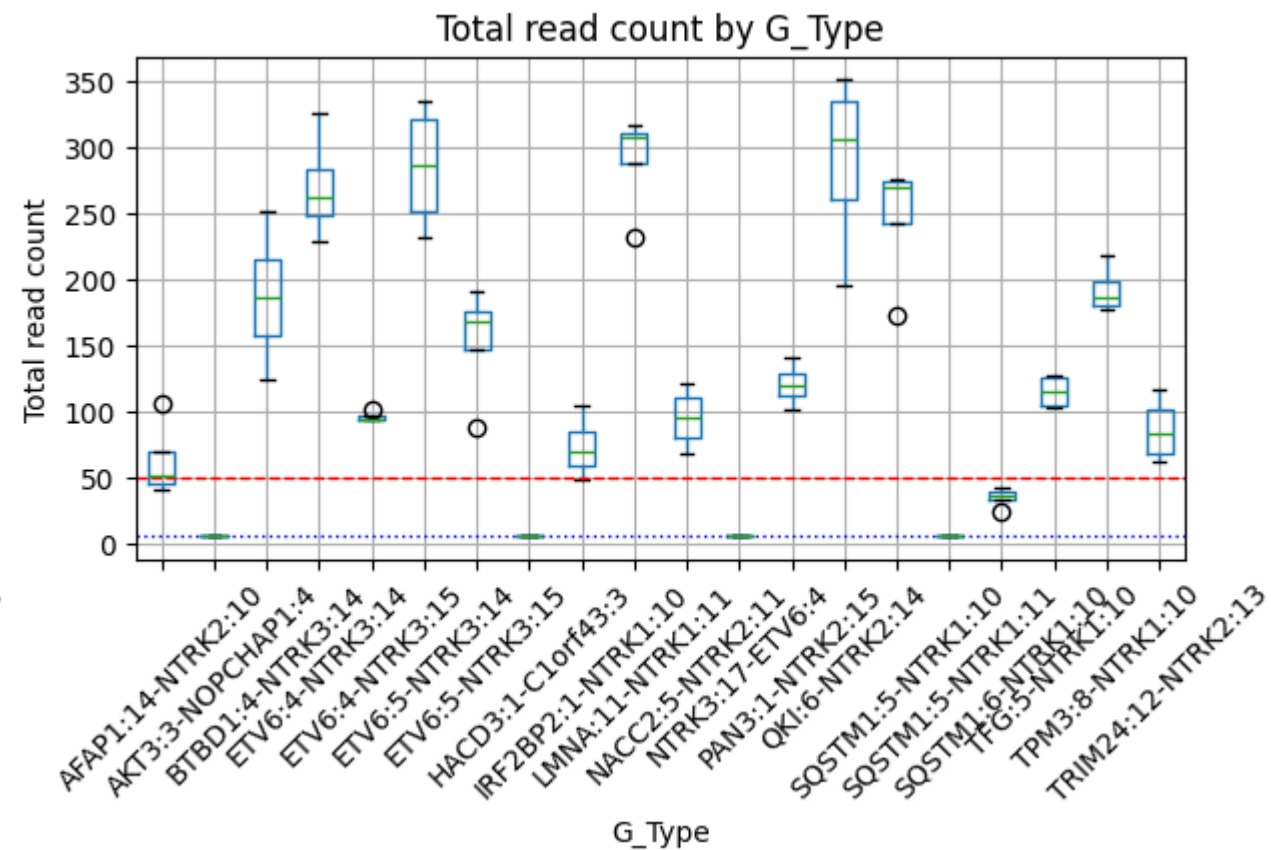
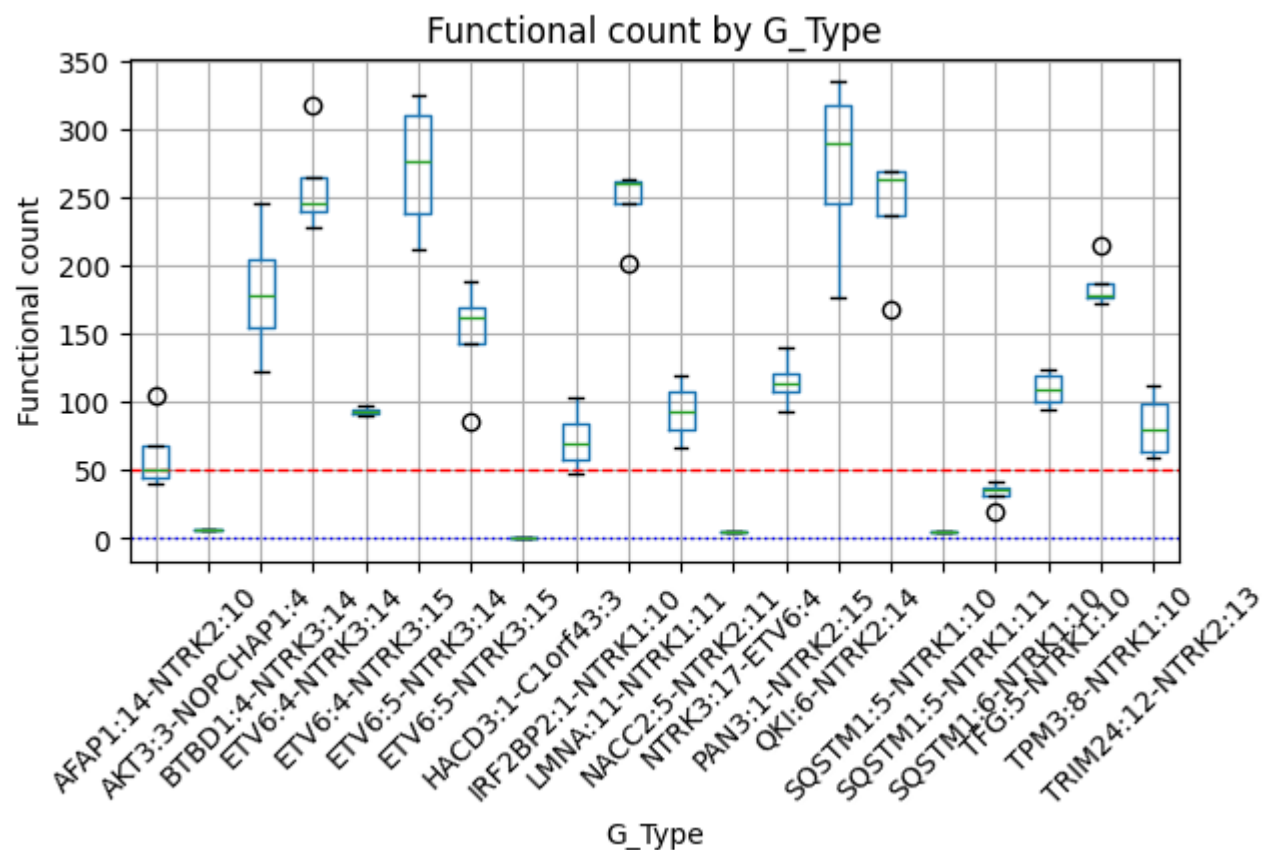
(Input amount vs calling)

Analysis summary (RM-23-001_R1 samples)

- Data inspection:
 - 16 stable fusions (identified across the 4 different input amount samples)
 - X-NTRK1/NTRK2/NTRK3 => 16 fusions
 - The 4 forms of ETV6-NTRK3 fusions are detected
- Result summary
 - The 16 fusions are also detectable using v0.95 db files (i.e. the boundary are identical across v0.95 and v1.4 in terms of stably expressed fusion types)
 - As compared to the v0.95 version pipeline, the v1.4 version will report more positive fusions.
 - Additional (unstable) fusions (non-target fusions)
 - AKT3:3-NOPCHAP1:4 (only detected in 50ng sample)
 - HACD3:1-C1orf43:3 (only detected in 50ng sample)
 - NTRK3:17-ETV6:4 (only detected in 10ng sample)
 - SQSTM1:5-NTRK1:11 (only detected in 50ng sample) => detected via the v0.95 db version pipeline as well

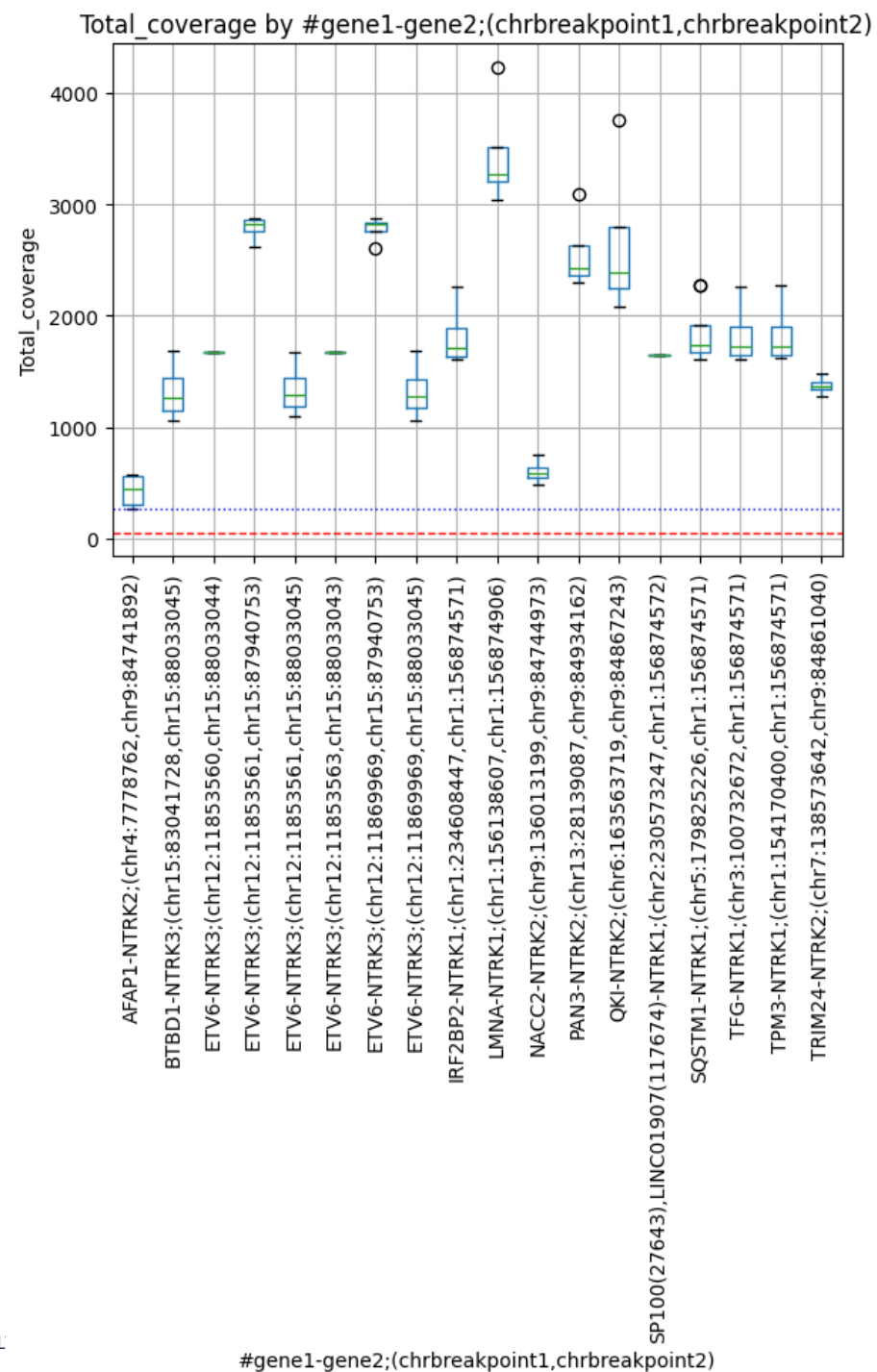
Fusionv4-based calling result (v1.4)

- 20 fusions



Arriba-based calling result

- 18 fusions
(Total_coverage = coverage1+coverage2)



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

- Arriba/fusionv4 (v0.95) * 10/30/50/100 ng
 - Filtering criteria
 - Fusion v4: Decision == '+'
 - Arriba: All the rows within "fusions.tsv" (=> considered positive by Arriba)

[illegible]

Arriba's algorithm

- Fusion candidates must satisfy the following criteria (artifact filtering):
 - Support read count > Estimated read count (e-value, background noise estimation)
 - Not included in the blacklist
 - Not a read-through break point
- Fusion break points refseq ID annotation
 - Annotate RefSeq ID nearby the fusion breakpoint within the gtf files

NTRK2 rows (in gtf):

...									
9	RefSeq	CDS	84741892	84741927	.	+	1	gene_id "NTRK2"; transcript_id "NM_001369532"; exon_number "11"; exon_id "NM_001369532.11"; gene_name "NTRK2";	
9	RefSeq	exon	84741892	84741927	.	+	.	gene_id "NTRK2"; transcript_id "NM_001369532"; exon_number "11"; exon_id "NM_001369532.11"; gene_name "NTRK2";	
9	RefSeq	CDS	84744973	84745073	.	+	1	gene_id "NTRK2"; transcript_id "NM_001369532"; exon_number "12"; exon_id "NM_001369532.12"; gene_name "NTRK2";	
9	RefSeq	exon	84744973	84745073	.	+	.	gene_id "NTRK2"; transcript_id "NM_001369532"; exon_number "12"; exon_id "NM_001369532.12"; gene_name "NTRK2";	
...									
9	RefSeq	CDS	84744973	84745073	.	+	1	gene_id "NTRK2"; transcript_id "NM_006180"; exon_number "11"; exon_id "NM_006180.11"; gene_name "NTRK2";	
9	RefSeq	exon	84744973	84745073	.	+	.	gene_id "NTRK2"; transcript_id "NM_006180"; exon_number "11"; exon_id "NM_006180.11"; gene_name "NTRK2";	
...									

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
NACC2-5-NTRK2:11 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
NACC2-5-NTRK2:12 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

Data (folder description)

- Probe_check
 - Probe2Exon.check_v1.4.xlsx (1039 probe regions => v1.4 exon map)
- Exon_count
 - Exon_tables.v0.95_37genes.xlsx
 - Exon_tables.v1.4_37genes.xlsx (exon tables)
- Calling_summary
 - AANB01_504_16samples-v0.95.boundaryQC.summary.xlsx (v0.95 fusionv4-based workflow result)
 - AANB01_504_16samples-v1.4.boundaryQC.summary.xlsx (v1.4 fusionv4-based workflow result)
 - AANB01_504.Calling.tables.xlsx (v0.95 & Arriba result summary for the 16 samples)
- DNA_arriba_raw_output
 - AANB01_504_A53_ACT2578_D5.fusions.tsv
 - AANB01_504_A54_ACT2704_D2.fusions.tsv

名稱	修改日期	類型	大小
Calling_summary	2025/5/6 下午 03:11	檔案資料夾	
DNA_arriba_raw_output	2025/5/5 上午 08:56	檔案資料夾	
Exon_count	2025/5/6 下午 02:59	檔案資料夾	
Probe_check	2025/5/7 下午 03:02	檔案資料夾	
fusionv5.37.genes.IDmap.xlsx	2025/5/6 下午 04:40	Microsoft Excel ...	14 KB

Make
Personalized
Medicine
Accessible to All

