

Optimization and Implementation of the nf-core/rnafusion Pipeline for fusion gene detection and Interpret the Result

Haymanot Berehan

University of Cape Town

ICGEB Intern

Oct 28, 2025

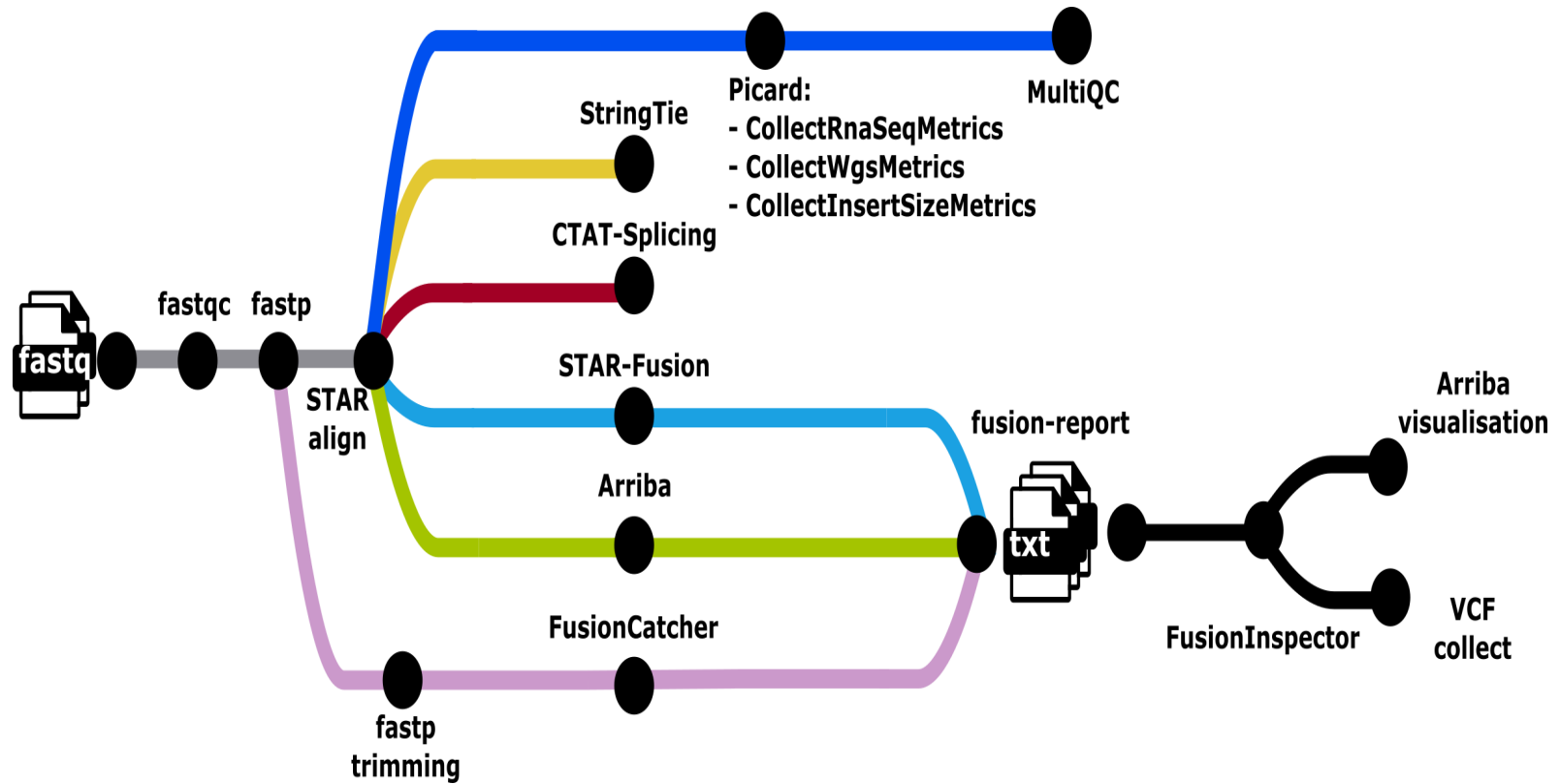
Tasks Performed:

- Setup & Configuration of HPC and Create Conda environment to install AWS tool.
- Build References on the HPC cluster and testing with built in data.
- prepared samplesheet for each patient for pipeline execution
- Extract and merge fusion genes using RStudio.
- Biological interpretation of some fusion genes.

What is Gene Fusion and Why it Matters?

- Gene fusion is the process where two distinct genes are joined into a single gene.
- This fusion creates a new, hybrid gene and a fusion protein with the combined functions of the original two genes.
- Many gene fusions are associated with different types of cancers. For example, the BCR-ABL fusion gene is linked to chronic myeloid leukemia.
- Understanding these fusions can help in developing targeted therapies.

nf-core/rnafusion v4.0.0 Overview



Cont'd...

```
!! Only displaying parameters that differ from the pipeline defaults !!

* The nf-core framework
  https://doi.org/10.1038/s41587-020-0439-x

* Software dependencies
  https://github.com/nf-core/rnafusion/blob/master/CITATIONS.md

executor > local (4)
[-] process > RNAFUSION:SAMTOOLS_CONVERT -
executor > local (6)
[-] process > RNAFUSION:SAMTOOLS_CONVERT -
executor > local (7)
executor > local (9)
[-] process > RNAFUSION:SAMTOOLS_CONVERT -
[e4/b3104a] process > RNAFUSION:FASTQ_FASTQC_UMITTOOLS_FASTP:FASTQC_RAW (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[df/01e697] process > RNAFUSION:FASTQ_FASTQC_UMITTOOLS_FASTP:FASTP (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[d2/838370] process > RNAFUSION:FASTQ_FASTQC_UMITTOOLS_FASTP:FASTQC_TRIM (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[b7/6b0376] process > RNAFUSION:SALMON_QUANT (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[d8/e74fdf] process > RNAFUSION:FASTQ_ALIGN_STAR:STAR_ALIGN (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[1e/f62ff0] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_GENOME:SAMTOOLS_SORT (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[f1/c6cfb2] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_GENOME:SAMTOOLS_INDEX (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[42/602200] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_GENOME:BAM_STATS_SAMTOOLS:SAMTOOLS_STATS (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[3f/028959] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_GENOME:BAM_STATS_SAMTOOLS:SAMTOOLS_FLAGSTAT (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[99/51889c] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_GENOME:BAM_STATS_SAMTOOLS:SAMTOOLS_IDXSTATS (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[-] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_TRANSCRIPTOME:SAMTOOLS_SORT -
[-] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_TRANSCRIPTOME:SAMTOOLS_INDEX -
[-] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_TRANSCRIPTOME:BAM_STATS_SAMTOOLS:SAMTOOLS_STATS -
[-] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_TRANSCRIPTOME:BAM_STATS_SAMTOOLS:SAMTOOLS_FLAGSTAT -
[-] process > RNAFUSION:FASTQ_ALIGN_STAR:BAM_SORT_STATS_SAMTOOLS_TRANSCRIPTOME:BAM_STATS_SAMTOOLS:SAMTOOLS_IDXSTATS -
[07/f6f9a5] process > RNAFUSION:SAMTOOLS_INDEX (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[c9/6c5fb6] process > RNAFUSION:CTATSPlicing_STARTCANCERINTRONS (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[8d/ea911f] process > RNAFUSION:ARRIBA_ARRIBA (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[d8/964818] process > RNAFUSION:STARFUSION_DETECT (CONTROL_REP1) [100%] 1 of 1 ✓
[0a/ab612a] process > RNAFUSION:FUSIONCATCHER_WORKFLOW:FUSIONCATCHER_FUSIONCATCHER (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[f6/aec984] process > RNAFUSION:BAM_STRINGTIE_MERGE:STRINGTIE_STRINGTIE (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[ea/7f2f16] process > RNAFUSION:BAM_STRINGTIE_MERGE:STRINGTIE_MERGE (1) [100%] 1 of 1, cached: 1 ✓
[42/1dcc6f] process > RNAFUSION:FUSIONREPORT_DETECT (CONTROL_REP1) [100%] 1 of 1 ✓
[bf/440bb1] process > RNAFUSION:FUSIONINSPECTOR_WORKFLOW:FUSIONINSPECTOR (CONTROL_REP1) [100%] 1 of 1 ✓
[f1/5fe729] process > RNAFUSION:FUSIONINSPECTOR_WORKFLOW:AGAT_CONVERTSPGFF2TSV (CONTROL_REP1) [100%] 1 of 1 ✓
[a8/e46efb] process > RNAFUSION:FUSIONINSPECTOR_WORKFLOW:VCF_COLLECT (CONTROL_REP1) [100%] 1 of 1 ✓
[e8/2da778] process > RNAFUSION:FUSIONINSPECTOR_WORKFLOW:ARRIBA_VISUALISATION (CONTROL_REP1) [100%] 1 of 1 ✓
[60/81ee93] process > RNAFUSION:QC_WORKFLOW:PICARD_COLLECTRNASQMTRICS (CONTROL_REP1) [100%] 1 of 1 ✓
[c2/385034] process > RNAFUSION:QC_WORKFLOW:GATK4_MARKDUPLICATES (CONTROL_REP1) [100%] 1 of 1 ✓
[40/85ddfa] process > RNAFUSION:QC_WORKFLOW:PICARD_COLLECTINSERTSIZEMTRICS (CONTROL_REP1) [100%] 1 of 1, cached: 1 ✓
[d9/404d3f] process > RNAFUSION:MULTIQC [100%] 1 of 1 ✓

-[nf-core/rnafusion] Pipeline completed successfully-
Completed at: 01-Oct-2025 20:47:04
Duration : 38m 6s
CPU hours : 13.6 (68.5% cached)
Succeeded : 9
Cached : 17

haymanot@compute-237:/cbio/users/haymanot$
```

Results

```
# Check combined table
print(dim(merged_fusion_table))
```

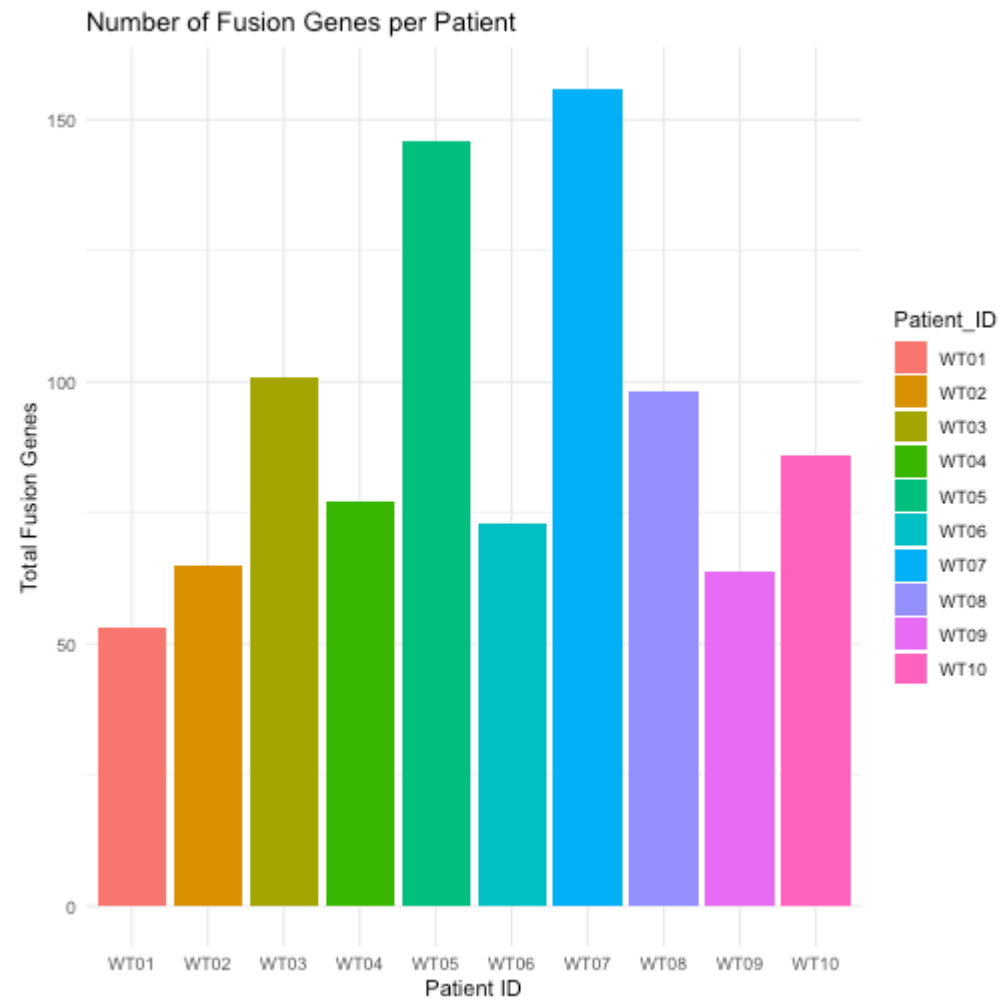
[1] 919 8

Preview of Merged Fusion Table							
Patient_ID	Fusion_gene	Found_in_DB	arriba	fusioncatcher	starfusion	Tools	Fusion_Inc
WT01	CTSC--RAB38	["Mitelman","FusionGDB2"]	false	true	false	1	0.417
WT01	CTBS--GNG5	["Mitelman"]	false	true	false	1	0.417
WT01	SIDT2--TAGLN	["Mitelman"]	false	true	false	1	0.417
WT01	CTNNBIP1--CLSTN1	["Mitelman","FusionGDB2"]	false	true	false	1	0.417
WT01	NCOR2--UBC	["Mitelman","FusionGDB2"]	false	true	false	1	0.417
WT01	UBE2E1--HNRNPC	[]	false	true	true	2	0.333
WT01	VSTM5--SNORC	[]	false	true	true	2	0.333
WT01	ENSG00000286129-IGLL5	[]	false	true	false	1	0.167
WT01	IGH@--IGHV4-4	[]	false	true	false	1	0.167
WT01	BOLA2-SMG1P6--SMG1	[]	false	true	false	1	0.167

Count the number of fusion genes per patient

Patient_ID	Total_Fusions
WT01	53
WT02	65
WT03	101
WT04	77
WT05	146
WT06	73
WT07	156
WT08	98
WT09	64
WT10	86

Diagrammatic Illustration



Count number of fusion genes detected by each tool per patient

Patient_ID	Arriba_only	FusionCatcher_only	STARFusion_only	Arriba+FusionCatcher	Fusion
WT01	5	35	11	0	
WT02	2	59	2	2	
WT03	8	89	1	1	
WT04	0	76	1	0	
WT05	14	127	5	0	
WT06	5	58	9	0	
WT07	14	137	3	1	
WT08	4	78	16	0	
WT09	7	56	1	0	
WT10	1	84	1	0	

How many of these gene fusions found in any database for each patient?

Known vs Novel Fusion Genes per Patient		
Patient_ID	Known	Novel
WT01	6	47
WT02	8	57
WT03	13	88
WT04	14	63
WT05	20	126
WT06	11	62
WT07	20	136
WT08	11	87
WT09	12	52
WT10	10	76

- known (In DB): Known fusions reported in fusion databases (e.g., Mitelman, FusionGDB2).
- Novel (Not in DB): Potentially novel or unreported fusions.

How many gene fusions are shared between more than one patient?

Total fusion genes found in more than one patient: 162

Fusion Genes Shared Between Multiple Patients		
Fusion_gene	Patients	Patient_List
BOLA2-SMG1P6--SMG1	10	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
CROCC--ENSG00000291072	10	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
CTBS--GNG5	10	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
ENSG00000291078--CIDEA	10	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
HERC2--ENSG00000291003	10	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
ENSG00000290928--SUZ12	9	WT01, WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT10
ENSG00000291117--PLEKHM1	9	WT02, WT03, WT04, WT05, WT06, WT07, WT08, WT09, WT10
ENSG00000291136--RARGEE1	9	WT01, WT02, WT03, WT04, WT05, WT07,

```
fusion_multi_tools <- merged_fusion_table %>%
  filter(Tools >= 2) %>%
  select(-Found_in_DB, -Fusion_Indication_Index)

# View result
print(fusion_multi_tools)
```

```
## # A tibble: 10 × 6
##   Patient_ID Fusion_gene      arriba fusioncatcher starfusion Tools
##   <chr>      <chr>      <chr>    <chr>      <chr>    <dbl>
## 1 WT01      UBE2E1--HNRNPC    false   true        true      2
## 2 WT01      VSTM5--SNORC      false   true        true      2
## 3 WT02      NUTM2A-AS1--MINPP1 true    true        false     2
## 4 WT02      DNTTIP2--C1orf21  true    true        false     2
## 5 WT03      VPS33A--EIF3J     true    true        true      3
## 6 WT03      GOLM2--TRAF7      true    true        true      3
## 7 WT03      ENSG00000256861--EIF3J true    true        false     2
## 8 WT06      B4GALT1--BAG1     true    true        true      3
## 9 WT07      CPSF6--AOPEP      false   true        true      2
## 10 WT07     BBS9--BMPER       true    true        false     2
```

chromosomal breakpoint information for some of common fusion genes

fusion genes	position
CTBS--GNG5	1:84563257:-#1:84501970:
BOLA2-SMG1P6--SMG1	16:29454016:-#16:18879719:
CROCC--ENSG00000291072	14:106276942:-#14:106737420:
ENSG00000291078--CIDEA	3:10026041:-#3:9870069:
HERC2--ENSG00000291003	15:28196215:-#15:22587344:
ENSG00000290928--SUZ12	17:30709811:+#17:31947617:
ENSG00000291136--RABGEF1	7:66592257:-#7:66775227:

Biological interpretation of some fusion genes.

CTBS--GNG5: is a *chimeric gene* (fusion transcript) formed by the fusion of the CTBS and GNG5 genes.

Gene Components

Gene	Function
CTBS	Encodes <i>chitobiase</i> , a lysosomal glycosidase that breaks down sugar chains on glycoproteins.
GNG5	Encodes the <i>gamma 5 subunit</i> of G-proteins, involved in intracellular signaling pathways regulating growth and metabolism.

Biological Role and Implications

- **Cancer Association:** Promotes cell growth, invasion, **stemness**, and **glycolysis** in glioblastoma cells.
- **Potential Biomarker:** Its presence in tumors but not normal tissues makes it a promising **diagnostic and therapeutic target**.
- **Functional Impact:** Suggests reprogramming of cellular metabolism and enhanced tumor aggressiveness.

Thank you!

